

FACEBOOK'S SPEECH CODE AND POLICIES: HOW THEY SUPPRESS SPEECH AND DISTORT DEMOCRATIC DELIBERATION

JOSEPH THAI*

With nearly two and a half billion users—a third of the world's population—Facebook far and away hosts the largest speech platform in the history of humanity. In the United States, seven out of ten adults use it, and nearly half get news from it. It is therefore no exaggeration to observe that Facebook's self-promulgated rules for "what is and is not allowed on Facebook," its content-based Community Standards, rival if not exceed the First Amendment's importance in shaping discourse in the United States. Yet, unlike the First Amendment, which only protects against government censorship, no scholarship has scrutinized the Community Standards' private regulation of speech on Facebook's far-reaching social media platform. As a result, basic questions about Facebook's sprawling speech code remain unexamined. For instance, at a minimum, do they offer enough clarity to ordinary users and prevent arbitrary or discriminatory moderation? Are they coherent and consistent as a whole? Furthermore, what do they reveal about the nature of Facebook's professed "commitment to expression"? And how do the Community Standards compare—for better or for worse—with modern First Amendment principles?

This Article inaugurates scholarly consideration of these basic questions about the Community Standards in the critical context of the spread of deliberate falsehoods and other disinformation on the platform to influence the electorate.

* Watson Centennial Chair in Law and Presidential Professor, *The University of Oklahoma*. I am grateful to the organizers and editors of this timely symposium for their exceptional work, to my copanelists Evelyn Aswad, David Hudson, and Stephen Wermiel for their insightful presentations and feedback, to my family—human, canine, and feline—for their tolerance and support, and to Stephen Colbert and Trevor Noah for their late night comedy breaks during the drafting of this Article.

Spoiler alert: Facebook’s speech code suffers from a basic lack of clarity and do not embody a coherent or consistent commitment to expression. Indeed, in many respects, the Community Standards suppress significantly more speech—including a wide range of political speech—than the First Amendment would permit of government censors.

In addition, Facebook categorically exempts “politicians” from its Community Standards as well as its fact-checking policy out of “respect for the democratic process.” But this selective hands-off approach perversely skews public debate by amplifying the expressive power of already dominant speakers in our society. Politicians enjoy an unrestricted license to exploit Facebook’s vast reach and highly effective ad targeting tools to spread expedient falsehoods among the most receptive users.

Making matters worse, Facebook by design does not broaden or balance public discussion but maximizes user engagement through algorithmically feeding users content tailored to their interests and likes. Many users thereby end up in politically imbalanced, self-reinforcing content bubbles where sensational falsehoods can gain credibility and velocity. While Facebook’s special solicitude for politicians and polarizing personalized feed might advance the company’s political and business interests, they also promote a race to the bottom among willing politicians and further divide the electorate into balkanized realities based on “alternative facts.”

Ultimately, Facebook’s speech code and policies for its social media platform present serious shortcomings, particularly in the electoral context. There is much room—and need—for further scrutiny.

TABLE OF CONTENTS

Introduction	1643
I. Facebook’s Community Standards	1646
A. A Qualified “Commitment to Expression,” Not to the First Amendment.....	1646
B. The Rules for “What Is and Is Not Allowed”	1650
1. Violence and incitement.....	1652
2. Dangerous individuals and organizations.....	1657
3. Hate speech	1662
4. Adult nudity and sexual activity.....	1665
II. Facebook’s Policies on Politicians, Political Speech, and Falsehoods	1671
A. Controversies.....	1671
B. Policies	1675
C. Critiques	1679
Conclusion.....	1687

INTRODUCTION

Facebook's user growth has exploded since its launch from a Harvard dormitory in 2004. By the end of its first year, the social media platform had attracted more than one million monthly users.¹ Four years after launch, Facebook hit 100 million users.² Another four years later, in 2012, it surpassed one billion users.³ Now, in its fifteenth year, nearly a third of humanity—2.45 billion out of 7.6 billion people—regularly uses Facebook.⁴ By comparison, Facebook has acquired more users in less than two decades than Christianity has followers after two millennia.⁵ No other platform for sharing or spreading speech in our digital age—or any age—comes close to matching Facebook's planetwide reach.⁶

As a result, Facebook's self-promulgated rules for what content can or cannot be published on its platform regulate more speakers than any other speech regime in the history of humanity. Those sprawling

1. See Kurt Wagner & Rani Molla, *Facebook's First 15 Years Were Defined by User Growth*, VOX (Feb. 5, 2019, 11:15 AM), <https://www.vox.com/2019/2/4/18203992/facebook-15-year-anniversary-user-growth> [<https://perma.cc/2JRA-3U75>].

2. *Id.*

3. *Id.*

4. See Josh Constine, *Facebook Shares Rise on Strong Q3, Users Up 2% to 2.45B*, TECHCRUNCH (Oct. 30, 2019, 4:15 PM), <https://techcrunch.com/2019/10/30/facebook-earnings-q3-2019> [<https://perma.cc/7FEU-8C7V>]; *U.S. and World Population Clock*, U.S. CENSUS BUREAU (Mar. 27, 2020, 1:32 PM), <https://www.census.gov/popclock> [<https://perma.cc/FNU5-2HTU>].

5. See Conrad Hackett & David McClendon, *Christians Remain World's Largest Religious Group, But They Are Declining in Europe*, PEW RES. CTR. (Apr. 5, 2017), <https://www.pewresearch.org/fact-tank/2017/04/05/christians-remain-worlds-largest-religious-group-but-they-are-declining-in-europe> [<https://perma.cc/TVE7-AXAE>] (reporting a population of 2.3 billion Christians worldwide in 2015).

6. Two out of the next four largest digital speech platforms are owned by Facebook: WhatsApp, a messaging service with over 1.5 billion users, and Instagram, a photo and video sharing service with around a billion users. See Josh Constine, *WhatsApp Hits 1.5 Billion Monthly Users. \$19B? Not So Bad*, TECHCRUNCH (Jan. 31, 2018, 5:15 PM), <https://techcrunch.com/2018/01/31/whatsapp-hits-1-5-billion-monthly-users-19b-not-so-bad> [<https://perma.cc/Z7K2-L6MR>]; Esteban Ortiz-Ospina, *The Rise of Social Media*, OUR WORLD DATA (Sept. 18, 2019), <https://ourworldindata.org/rise-of-social-media> [<https://perma.cc/3X3Y-83MC>]. YouTube, owned by Google, has a user base of nearly two billion, and WeChat, a Chinese messaging, social media, and payment platform, has around one billion users. *Id.* As part of its "Great Firewall," China blocks Facebook, WhatsApp, and Instagram, among other platforms. See Ariel Hochstadt, *The Complete List of Blocked Websites in China & How to Access Them*, VPNMENTOR <https://www.vpnmentor.com/blog/the-complete-list-of-blocked-websites-in-china-how-to-access-them> [<https://perma.cc/EFJ6-UPDR>].

rules, published as Facebook's Community Standards,⁷ presently regulate the speech of well over two billion more speakers than the First Amendment.⁸ Moreover, while the First Amendment protects around 330 million people in the United States from government censorship, it does not restrict private individuals or companies from censoring speech on their own property, including their social media platforms.⁹ Consequently, the seven in ten U.S. adults and half of U.S. teens who use Facebook subject their speech to its Community Standards without any First Amendment protection.¹⁰ And yet, while countless works of scholarship have been devoted to the freedom of speech protected by the First Amendment, none to date have engaged in a sustained examination or evaluation of Facebook's speech code despite its unprecedented reach.¹¹ This Article inaugurates that endeavor.

7. See *Community Standards*, FACEBOOK <https://www.facebook.com/communitystandards> [<https://perma.cc/VZJ7-UQR8>].

8. The population of the United States in July of 2019 exceeded 331 million. See *U.S. and World Population Clock*, *supra* note 4.

9. See Dawn C. Nunziato, *The Death of the Public Forum in Cyberspace*, 20 BERKELEY TECH. L.J. 1115, 1116–17 (2005).

10. See Monica Anderson & Jingjing Jiang, *Teens, Social Media & Technology 2018*, PEW RES. CTR. (May 31, 2018), <https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018> [<https://perma.cc/2WZK-N2L7>]; Andrew Perrin & Monica Anderson, *Share of U.S. Adults Using Social Media, Including Facebook, Is Mostly Unchanged Since 2018*, PEW RES. CTR. (Apr. 10, 2019), <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018> [<https://perma.cc/CF2S-X4RN>].

11. To be sure, a growing body of legal scholarship has examined the free-speech implications of the rise of online services such as social media platforms as fora for speech. For example, many have argued for the extension of First Amendment, legislative, or other legal protections to speech on such private platforms. See, e.g., Benjamin F. Jackson, *Censorship and Freedom of Expression in the Age of Facebook*, 44 N.M. L. REV. 121 (2014); Jonathan Peters, *The "Sovereigns of Cyberspace" and State Action: The First Amendment's Application—or Lack Thereof—to Third-Party Platforms*, 32 BERKELEY TECH. L.J. 989 (2017); Colby M. Everett, Comment, *Free Speech on Privately-Owned Fora: A Discussion on Speech Freedoms and Policy for Social Media*, 28 KAN. J.L. & PUB. POL'Y 113 (2018); Mason C. Shefa, Comment, *First Amendment 2.0: Revisiting Marsh and the Quasi-Public Forum in the Age of Social Media*, 41 U. HAW. L. REV. 159 (2018). Few have cautioned against doing so. See, e.g., Julian N. Eule & Jonathan D. Varat, *Transporting First Amendment Norms to the Private Sector: With Every Wish There Comes A Curse*, 45 UCLA L. REV. 1537, 1539 (1998). Others have decried uncertainty over whether popular social media platforms might qualify as protected public fora. See *Leading Cases, Packingham v. North Carolina*, 131 HARV. L. REV. 233, 233 (2017). However, no scholarly work has engaged in a sustained examination of Facebook's Community Standards as a governing set of speech rules, either on their own merits or in comparison to First Amendment principles and norms. Cf. Sarah C. Haan, *Facebook's*

Part I of this Article presents an overview of the nature of the Community Standards and examines content restrictions in four representative subject areas—violence and incitement, dangerous individuals or organizations, hate speech, and adult nudity and sexual health. To illuminate the contours and value choices of these content restrictions, this Part compares them with familiar First Amendment principles in related areas. This comparison reveals that Facebook's professed "commitment to expression"—touted as "paramount" at the outset of its Community Standards¹²—often conflicts with a speech code that in critical parts is significantly less clear, less coherent, and more restrictive than the First Amendment. Most notably, through problematic do-not-post rules, the Community Standards broadly prohibit content that many community members may find offensive in order to promote user satisfaction and retention.

Part II then examines how, beyond the Community Standards, Facebook's policies distort democratic deliberation. Foremost, Facebook privileges the speech of "politicians" through categorically exempting them from its Community Standards as well as its fact-checking policy. This approach perversely skews public debate in favor of an already dominant class of speakers in our society by empowering them alone to speak freely and falsely on the far-reaching platform. Furthermore, Facebook facilitates the viral spread of falsehoods by feeding users personalized content algorithmically narrowed to their interests and likes. This personalized feed migrates many users into politically imbalanced, self-reinforcing content bubbles and enables willing politicians to leverage powerful ad targeting that pushes disinformation to the multitudes most receptive to them. Thus, while Facebook's policies might advance its business interests, they divide and distort public discussion.

In the end, this Article will hopefully demonstrate the pressing need for further study of Facebook's unprecedented speech code and policies that shape the speech of billions of people, including disinformation that may undermine democratic self-governance.

Alternative Facts, 105 VA. L. REV. ONLINE 18, 20 (2019) (arguing that "Facebook's adoption of the alternative-facts frame" through the use of "related articles" and other responses to misinformation "potentially contributes to the divisiveness that has made social media misinformation a powerful digital tool").

12. See *Community Standards*, *supra* note 7.

I. FACEBOOK'S COMMUNITY STANDARDS

According to Facebook, its Community Standards set out “the rules for what content is and isn’t permitted on Facebook.”¹³ This Part discusses the avowed goals and values of Facebook’s speech code and scrutinizes its problematic content restrictions in four representative areas.

A. A Qualified “Commitment to Expression,” Not to the First Amendment

The Introduction to the Community Standards states that the goal “has always been to create a place for expression and give people a voice,” so they can “talk openly about the issues that matter to them, even if some may disagree or find them objectionable.”¹⁴ This “commitment to expression”—which Facebook also brands in Silicon-valley speak as “commitment to voice”—assertedly serves not only to enable users to “connect with friends and family” but also to promote the utopian aims of “[b]uilding community and bringing the world closer together.”¹⁵ Indeed, because Facebook’s “commitment to expression is paramount,” the company professes to permit some content that “would otherwise go against our Community Standards” if “it is newsworthy and in the public interest.”¹⁶ However, apart from this limited newsworthiness exception, the Community Standards do not set out any explicit protection of “expression” or “voice” that would give users some affirmative right to free speech on Facebook.

On the other hand, noting that “the internet creates new and increased opportunities for abuse,” the Community Standards explicitly “limit expression” in furtherance of certain “values”:

- “Authenticity,” which “creates a better environment for sharing” and opposes users “misrepresent[ing] who they are or what they’re doing”;
- “Safety,” which excludes content that “has the potential to intimidate, exclude or silence others”;

13. *Writing Facebook’s Rulebook*, FACEBOOK (Apr. 10, 2019), <https://about.fb.com/news/2019/04/insidefeed-community-standards-development-process> [<https://perma.cc/366J-KT37>]; see also *Community Standards*, *supra* note 7 (explaining Community Standards “outline what is and is not allowed on Facebook”).

14. *Community Standards*, *supra* note 7.

15. *Id.*

16. *Id.* This exception is discussed *infra* Section II.B.

- “Privacy,” which includes “personal privacy and information” and permits users to “choose how and when to share on Facebook”; and
- “Dignity,” which calls on users to “respect” and “not harass or degrade others.”¹⁷

In the entirety of the Community Standards, there is not a single reference to the text, principles, or values of the First Amendment, despite more than a century’s worth of precedents and scholarship on its protections of the fundamental right of free speech.¹⁸ For example, the Community Standards’ public avowal of a “commitment to expression” and “commitment to voice” does not reference the most widely accepted justifications for the robust protection of speech under the First Amendment—promoting truth-seeking, democratic self-governance, autonomy, and self-expression.¹⁹ Nor do the Community Standards purport to adopt, adapt, or otherwise rely on the “core postulate of free speech law” forbidding censorship based on disagreement with the message or viewpoint.²⁰ Indeed, as discussed in Section I.B, the Community Standards directly contravene this core postulate in many ways, including in censoring political speech based on potential offense to user sensitivities. Accordingly, the conspicuous absence of a Facebook “like” or “tag,” so to speak, to the most prominent and celebrated speech protection in the United States does not seem accidental. And it is not.

17. *Community Standards*, *supra* note 7.

18. The modern interpretation of the First Amendment is commonly recognized to have begun with the Supreme Court’s review of free speech challenges to Espionage Act convictions during World War I. *See, e.g.*, *Schenck v. United States*, 249 U.S. 47, 52 (1919) (describing the “clear and present danger” test); *Abrams v. United States*, 250 U.S. 616, 628 (1919) (Holmes, J., dissenting) (arguing that only “the present danger of immediate evil or an intent to bring it about” warrants limitations on expression).

19. *See, e.g.*, ERWIN CHERMERINSKY, *CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES* 1006–10 (6th ed. 2019); KATHLEEN M. SULLIVAN & NOAH FELDMAN, *FIRST AMENDMENT LAW* 5–9 (6th ed. 2016); Justin Marceau & Alan K. Chen, *Free Speech and Democracy in the Video Age*, 116 *COLUM. L. REV.* 991, 999 (2016).

20. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2297, 2299 (2019) (invalidating the Lanham Act’s prohibition on the registration of “‘immoral or scandalous’ trademarks”); *see R.A.V. v. City of St. Paul, Minn.* 505 U.S. 377, 430 (1992) (“If there is a bedrock principle underlying the First Amendment, it is that the Government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable. Viewpoint discrimination is censorship in its purest form” (internal citations and quotation marks omitted)).

For one, in the Introduction to its Community Standards, Facebook touts how it “take[s] great care” to write and regularly revise its speech rules in light of “feedback from our community” as well as consultation with experts “in fields such as technology, public safety and human rights.”²¹ Notably absent from this primary list of consulted fields is either free speech, generally, or the First Amendment, specifically. Other Facebook articles that discuss how the company’s “policy,” “product,” and “operations” teams continually evolve the Community Standards to “account for a range of perspectives and opinions across the globe” also fail to acknowledge consultation with scholars on the freedom of expression.²² It is only at the end of its Community Standards that Facebook sets out a longer list of “stakeholders” with whom it engages, including “civil society organizations, activist groups, and thought leaders” in areas such as “digital and civil rights, anti-discrimination, free speech, and human rights.”²³ In this laundry list of stakeholders, free speech experts do not occupy a “paramount” place, and mention of the First Amendment is again noticeably avoided.

This absence of acknowledgment of First Amendment principles, norms, or expert consultation may seem unsurprising in one respect, as the company’s speech platform serves billions of international users not protected by nor likely familiar with the constitutional provision. However, in publicly defending its content moderation policies in the face of domestic criticism, Facebook’s executives have not shied from invoking robust free speech rhetoric that echoes modern First Amendment principles and norms.²⁴ Furthermore, in leaked audio from an internal employee forum, Facebook founder, CEO, and chair Mark Zuckerberg related essentially domestic political considerations for avoiding any modeling of the Community Standards after the First Amendment. Despite plenty of scholars and commentators calling for Facebook and other social media platforms to adopt First Amendment rules or norms,²⁵

21. *Community Standards*, *supra* note 7.

22. *Promoting Safety with Policy, Product and Operations*, FACEBOOK (Nov. 15, 2018), <https://about.fb.com/news/2018/11/inside-feed-community-integrity-keeping-people-safe> [<https://perma.cc/XSA6-K4TA>]; *Writing Facebook’s Rulebook*, *supra* note 13; *see also Standing Against Hate*, *infra* note 76 (discussing how Facebook did not mention consulting with free speech or First Amendment experts in developing policy against white nationalism).

23. *Stakeholder Engagement, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/stakeholder_engagement [<https://perma.cc/4K7T-UX9J>].

24. *See infra* notes 147–48, 192.

25. *See supra* note 11 and accompanying text; *see also* David L. Hudson, Jr., *In the Age of Social Media, Expand the Reach of the First Amendment*, ABA, <https://>

Zuckerberg responded to an employee asking whether it should do so by asserting that “I don’t think anyone says . . . that we should [follow the] First Amendment.”²⁶ Instead, Zuckerberg explained, “a lot of people think that we need to be more aggressive in moderating content that is offensive or basically would make certain groups of people feel unsafe.”²⁷ However, with respect to “political discourse,” Zuckerberg asserted that a “lot of people” feel that those “arbitrating what is misinformation and doing fact-checking tend to be left of center” and are “getting in the way of an ability to express something that they feel is real and that matches their lived experience.”²⁸ As a result, Facebook’s content moderation gets pressure from “different sides” and “tr[ies] to navigate this as well as possible.”²⁹

Zuckerberg’s views matter greatly not only because he occupies the very top executive positions within Facebook, but also because he has always possessed controlling shares of the company³⁰ and lately has been the most vocal defender of its speech rules.³¹ Moreover, the two sentiments he expressed above—needing to police “offensive” or “unsafe” content aggressively, while “giv[ing] people a voice” to express “their lived experience” without fact checking “getting in the way”—align more neatly with how the Community Standards censor some kinds of speech while countenancing others than with any coherent or consistent “commitment to expression.” In other words, as discussed in Part II, the Community Standards in critical parts seem to reflect Facebook’s attempt to politically “navigate” the pressure it receives from “different sides” to promote user satisfaction and otherwise protect its business interests.

www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/in-the-age-of-social-media-first-amendment [<https://perma.cc/FBV6-AZN5>].

26. See Casey Newton, *Why Facebook Doesn’t Follow the First Amendment*, VERGE (Oct. 10, 2019, 6:00 AM), <https://www.theverge.com/facebook/2019/10/10/20906304/facebook-first-amendment-political-ads-leaked-audio> [<https://perma.cc/A3BB-WDPH>].

27. *Id.*

28. *Id.*

29. *Id.*

30. See Emily Stewart, *Facebook Will Never Strip Away Mark Zuckerberg’s Power*, VOX (May 30, 2019, 2:38 PM), <https://www.vox.com/recode/2019/5/30/18644755/facebook-stock-shareholder-meeting-mark-zuckerberg-vote> [<https://perma.cc/74AJ-UCTB>].

31. See *Mark Zuckerberg Stands for Voice and Free Expression*, FACEBOOK (Oct. 17, 2019), <https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression> [<https://perma.cc/D8DZ-2YQT>].

B. The Rules for “What Is and Is Not Allowed”

Facebook divides its rules for “what is and is not allowed” into five major content categories: “Violence and Criminal Behavior,” “Safety,” “Objectionable Content,” “Integrity and Authenticity,” and “Respecting Intellectual Property.”³² These primary categories are further divided into subtopics, for a total of over twenty subject areas in which Facebook significantly restricts speech. Subtopics usually open with a broad “Policy Rationale” for censorship within that subject area and follow with a list of “Do not post” rules of varying length and breadth.

Given the multitude and diversity of speech restrictions across the sprawling Community Standards, it is not possible to scrutinize them all in this Article. Instead, this Section examines four representative subject areas in which Facebook restricts speech—violence and incitement, dangerous individuals or organizations, hate speech, and adult nudity and sexual health. This survey exposes the considerable lack of clarity and coherence that characterizes the Community Standards as well as the Standards’ notably wide differences from First Amendment principles—differences that often result in considerably less speech protection.

The “Violence and Criminal Behavior” category of the Community Standards hosts the rules restricting speech in five subject areas. To give context to an examination of two of those areas, all five are listed in the table below alongside the main policy rationales offered by the Community Standards for censorship within them.³³

32. *Community Standards*, *supra* note 7.

33. The quoted policy rationales are usually the first sentences introducing each subtopic.

Table 1: Violence and Criminal Behavior

Subtopic	Policy Rationale
Violence and Incitement	“We aim to prevent potential offline harm that may be related to content on Facebook. While we understand that people commonly express disdain or disagreement by threatening or calling for violence in non-serious ways, we remove language that incites or facilitates serious violence.” ³⁴
Dangerous Individuals and Organizations	“In an effort to prevent and disrupt real-world harm, we do not allow any organizations or individuals that proclaim a violent mission or are engaged in violence to have a presence on Facebook.” ³⁵
Coordinating Harm and Publicizing Crime	“In an effort to prevent and disrupt offline harm and copycat behavior, we prohibit people from facilitating, organizing, promoting, or admitting to certain criminal or harmful activities targeted at people, businesses, property or animals.” ³⁶
Regulated Goods	“To encourage safety and compliance with common legal restrictions, we prohibit

34. *Violence and Incitement, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/credible_violence [<https://perma.cc/8CY5-S5XB>].

35. *Dangerous Individuals and Organizations, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/dangerous_individuals_organizations [<https://perma.cc/A42G-RLT5>].

36. *Coordinating Harm and Publicizing Crime, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/coordinating_harm_publicizing_crime [<https://perma.cc/GX7U-EP5H>].

	attempts by individuals, manufacturers, and retailers to purchase, sell, or trade non-medical drugs, pharmaceutical drugs, and marijuana. We also prohibit the purchase, sale, gifting, exchange, and transfer of firearms, including firearm parts or ammunition, between private individuals on Facebook.” ³⁷
Fraud and Deception	“In an effort to prevent and disrupt harmful or fraudulent activity, we remove content aimed at deliberately deceiving people to gain an unfair advantage or deprive another of money, property, or legal right.” ³⁸

1. *Violence and incitement*

Whether or not by design, the first listed subtopic, “Violence and Incitement,” exemplifies how and why the Community Standards aggressively censor speech. The subtopic opens with a “Policy Rationale” that generally declares the aim to “prevent potential offline harm that may be related to content on Facebook” without specifically justifying the substance or reach of any of the restrictions in this area.³⁹ The text also does not articulate an overarching rule, standard, or test for what constitutes forbidden speech under this subtopic. Instead, the Community Standards generally describe the forbidden speech as “language that incites or facilitates serious violence,” “direct threats to public safety,” or speech that otherwise raises “a genuine risk of physical harm” without defining any of those terms.⁴⁰ The Community Standards further note some content that would not fall

37. *Regulated Goods, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/regulated_goods [<https://perma.cc/9FBD-8MES>].

38. *Fraud and Deception, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/fraud_deception [<https://perma.cc/R3FS-65AE>].

39. *Violence and Incitement, supra* note 34.

40. *Id.*

within this subtopic—“casual statements,” “threatening or calling for violence in non-serious ways,” or non-credible “aspirational or conditional threats directed at terrorists and other violent actors.”⁴¹ More significantly, under a “Do not post” header, the Community Standards list dozens of kinds of forbidden content that do not seem organized in any overall order.⁴²

The laundry list of forbidden content includes three classes of threats associated with different levels of physical severity and different targets:

- “[t]hreats that could lead to death (and other forms of high-severity violence)” against “any target(s)”;
- “[t]hreats that lead to serious injury (mid-severity violence) towards private individuals, minor public figures, high risk persons, or high risk groups”; and
- “[t]hreats that lead to physical harm (or other forms of lower-severity violence) towards private individuals (self-reporting required) or minor public figures.”⁴³

Accompanying these threat proscriptions are bans against advocacy of corresponding levels of violence: “[s]tatements advocating for high-severity violence,” “[c]alls for mid-severity violence,” and “calls for . . . low-severity violence.”⁴⁴

In several respects, this cluster of proscribed threats and incitements showcases the frequent lack of clarity and coherence in the Community Standards as well as their significant differences from First Amendment protections.

First, key terms often are not defined, illustrated, or otherwise elucidated, leaving their application unclear at best. For example, while “death” does not need further explanation, it is hardly obvious what constitutes high-severity violence as opposed to mid-severity violence, as neither expression comes from common parlance or First Amendment doctrine. Would a punch in the nose count as one or the other? Alternatively, would a slap on the cheek amount to mid-severity violence

41. *Id.*

42. The haphazard nature of this “Do not post” list is punctuated by inconsistencies or errors in punctuation. For example, while all the listed items are sentence fragments, some end in periods while others do not, and spaces sometimes are not inserted where needed (e.g., after a comma) and inserted where not needed (e.g., after an opening parenthesis). *See id.* These observations may seem like nit-picking, but they reinforce the impression that the Community Standards are not especially refined in substance or presentation.

43. *Id.*

44. *Id.*

or lower-severity violence, or would it not rise to any level of “serious violence” that could result in censorship? Furthermore, if a threat or advocacy of mid-severity violence is made against a minority group, such as the Black Student Association, would it qualify as a prohibited threat against a “high risk group”? Relatedly, could threats or advocacy of lower-severity violence freely target minority groups, since this tier expressly protects only “private individuals” and “minor public figures”?⁴⁵ In addition, since these do-not-post rules address “minor public figures” (undefined) but do not mention other kinds of public figures, how would threats or incitements against non-minor (major?) public figures fare? While threats or advocacy of death or high-severity violence are banned against any target, would those involving mid-severity violence or lower-severity violence—say, threats or calls to punch, slap, or milkshake a major celebrity or politician⁴⁶—escape Facebook censorship? Again, answers to these questions do not emerge from everyday understanding or First Amendment doctrine.

Second, from the standpoint of either “prevent[ing] potential offline harm” or upholding a “commitment to expression,”⁴⁷ it is difficult to discern a coherent explanation for various lines drawn in this list of prohibited threats, and the Community Standards provide none. As highlighted, one line that they implicitly—if perhaps unintentionally—draw is between threats or advocacy of death or high-severity violence against non-minor public figures, which are banned, and threats or advocacy of mid-severity or low-severity violence against the same figures, which are not. It is not apparent why the balance between expression and safety tips in favor of the former when major public figures (Oprah Winfrey? Bernie Sanders?) are targeted only with mid-severity violence or lower-severity violence. Nor is it apparent why threats or calls for violence against “high risk groups” or “high risk persons” (whoever they are) are specifically forbidden for mid-severity violence but implicitly allowed for lower-severity violence.⁴⁸ It is also not obvious why private individuals are required to self-report threats or advocacy of lower-severity violence, but minor public figures are not.⁴⁹

45. *Cf. Tenn. Valley Auth. v. Hill*, 437 U.S. 153, 188 (1978) (applying the maxim “*expressio unius est exclusio alterius*,” i.e., “the expression of one thing is the exclusion of the other”).

46. See Iliana Magra, *Why Are Milkshakes Being Thrown at Right-Wing Politicians Like Nigel Farage?*, N.Y. TIMES (May 21, 2019), <https://www.nytimes.com/2019/05/21/world/europe/milkshake-nigel-farage.html>.

47. *Community Standards*, *supra* note 7; *Violence and Incitement*, *supra* note 34.

48. *Violence and Incitement*, *supra* note 34.

49. Presumably, “self-reporting” means reporting threats against oneself, not threats one has made. If so, then another line needs explaining: why require self-

The violence and incitement subtopic also notably bars “statements of intent, calls for action, or advocacy for high or mid-severity violence due to voting, voter registration, or the outcome of an election.”⁵⁰ By omission, the Community Standards seemingly allow statements of intent or advocacy of lower-severity violence with respect to voting, voter registration, or the outcome of an election. But if there is an overriding interest in protecting the exercise of the fundamental right to vote from violence, it is neither apparent nor explained why that interest gives way to speech threatening or advocating less severe forms of violence. This line-drawing question is compounded by the lack of clarity over what constitutes lower-severity violence.

Third, the Community Standards differ substantially from First Amendment standards. For one, the Supreme Court’s “true threats” jurisprudence has not made fine distinctions between different severity levels of violence.⁵¹ So far, it has simply described constitutionally-proscribable threats as “those statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals.”⁵² The Court has broadly justified the lack of First Amendment protection for this undifferentiated category of speech based on the classic rationale advanced in *Chaplinsky v. New Hampshire*⁵³ for censorship of certain historical categories of unprotected speech—they are “of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.”⁵⁴ While perhaps a case may be made for drawing finer distinctions in levels of protection based on the severity of the threat or the prominence of the target,⁵⁵ the Community Standards have not taken this route.

reporting rather than take down threats against private individuals reported by third-parties or discovered by Facebook moderators?

50. *Violence and Incitement*, *supra* note 34.

51. *See generally* *Watts v. United States*, 394 U.S. 705, 707–08 (1969) (per curiam) (describing factors to be used in interpreting whether a statement is a true threat).

52. *Virginia v. Black*, 538 U.S. 343, 359 (2003); *see also id.* at 360 (“Intimidation in the constitutionally proscribable sense of the word is a type of true threat, where a speaker directs a threat to a person or group of persons with the intent of placing the victim in fear of bodily harm or death.”).

53. 315 U.S. 568 (1942).

54. *Black*, 538 U.S. at 358–59 (quoting *R.A.V. v. City of St. Paul*, 505 U.S. 377, 382–83 (1992) (quoting *Chaplinsky*, 315 U.S. at 572)).

55. *Cf. FCC v. Pacifica Found.*, 438 U.S. 726, 750 (1978) (“[C]ontext is all-important.”).

More significantly, the Community Standards' ban on advocacy of violence sweeps much more broadly than incitement that the First Amendment permits governments to censor. Under the modern test set forth in *Brandenburg v. Ohio*,⁵⁶ advocacy of lawless (including violent) action receives First Amendment protection "except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action."⁵⁷ This stringent, highly speech-protective test requires both the specific intent ("directed") to produce "imminent lawless action" and a showing that such imminent lawlessness is "likely" to materialize.⁵⁸ Under this test, "mere advocacy" of lawless (even violent) action receives First Amendment protection.⁵⁹ Accordingly, aspirational or conditional calls for lawless action at some indeterminate point in the future, or which may not occur at all, enjoy constitutional protection.⁶⁰ As Justice Brandeis argued in an influential opinion that prefigured the *Brandenburg* test, "If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence."⁶¹

By contrast, the Community Standards forbid calls for high, mid, and lower-severity violence without any requirement of imminence or any other limiting timeframe.⁶² In addition, the Community Standards target "credible" threats and incitements as opposed to "non-serious" ones but do not impose any requirement that the called-for action be likely to come about.⁶³ Thus, "aspirational or conditional statements" may run afoul of

56. 395 U.S. 444 (1969) (per curiam).

57. *Id.* at 447.

58. See Gerald Gunther, *Learned Hand and the Origins of Modern First Amendment Doctrine: Some Fragments of History*, 27 STAN. L. REV. 719, 755 (1975) (characterizing the *Brandenburg* test as the "most speech-protective standard yet evolved by the Supreme Court").

59. *Brandenburg*, 395 U.S. at 448–49; cf. *Gitlow v. New York*, 268 U.S. 652, 673 (1925) (Holmes, J., dissenting) ("Every idea is an incitement If in the long run the beliefs expressed in proletarian dictatorship are destined to be accepted by the dominant forces of the community, the only meaning of free speech is that they should be given their chance and have their way.").

60. See *Brandenburg*, 395 U.S. at 446, 449.

61. *Whitney v. California*, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring); see Vincent Blasi, *The First Amendment and the Ideal of Civic Courage: The Brandeis Opinion in Whitney v. California*, 29 WM. & MARY L. REV. 653, 682 (1988) (discussing how Brandeis's concurrence is frequently invoked by Supreme Court Justices who have "contributed significantly to the modern elaboration of the free speech principle").

62. *Violence and Incitement*, *supra* note 34.

63. *Id.*

the Community Standards⁶⁴ even if the prospect of violence is unlikely or there is time “to avert the evil” through counterspeech.⁶⁵ Consequently, a call for protesters at a rally months hence to resist counterprotesters with violence if necessary would likely receive First Amendment protection in the United States because of its conditionality and lack of imminence but would likely draw Facebook censorship.

Curiously, the Community Standards *do* explicitly require a short fuse for “[m]isinformation that contributes to the risk of *imminent* violence or physical harm.”⁶⁶ Facebook thus would not forbid posting misinformation about a rally that could contribute to violence or physical harm, so long as the timeframe is not imminent. While the Community Standards do not explain this special solicitude for misinformation over public safety in this context, the policy does align with Zuckerberg’s private acknowledgment of the need to “navigate” external pressure to “give people a voice to express their lived experience” without fact-checking “getting in the way.”⁶⁷ However, this special solicitude comes at a potential cost to the democratic process that Facebook elsewhere claims to respect.⁶⁸ The Standards permit, if not encourage, speakers to spread misinformation that, for example, could chill voting and other electoral activities by suggesting that non-imminent violence or physical harm could result.

2. *Dangerous individuals and organizations*

Under the category of “Violence and Criminal Behavior,” the Community Standards not only forbid threats and incitements of physical violence but also ban “any organizations or individuals that proclaim a violent mission or are engaged in violence, from having a presence on Facebook.”⁶⁹ The classes of banned speakers include organizations or individuals involved in “terrorist activity,” “organized hate,” “mass murder” or “multiple murder,” “human trafficking,” or “organized violence or criminal activity.”⁷⁰ Facebook also broadly bans “content that expresses support or praise for groups, leaders, or individuals involved in these activities.”⁷¹

64. *Id.*

65. *Whitney*, 274 U.S. at 377.

66. *Violence and Incitement*, *supra* note 34 (emphasis added).

67. *See supra* notes 30–31.

68. *See infra* note 147; *see also infra* Part II.

69. *Dangerous Individuals and Organizations*, *supra* note 35.

70. *Id.*

71. *Id.*

Notably, with respect to terrorism, the Community Standards ban both “terrorist organizations” and individual “terrorists” without regard to their role in any affiliated organization; however, with respect to hate, they ban only “[h]ate organizations and their leaders and prominent members.”⁷² Apparently, the policy aim of “prevent[ing] and disrupt[ing] real-world harm” does not extend to rank-and-file members of hate organizations or lone wolves.⁷³

It is easy to appreciate why Facebook would boot terrorist organizations, hate groups, and their ilk off its platform. In the wake of the violent Unite the Right rally in Charlottesville⁷⁴ and resurgence of white supremacists online, as well as on streets, Facebook and other online platforms faced tremendous public pressure to do so, including from lawmakers and civil rights groups.⁷⁵ Facebook responded with a statement proclaiming that, after “conversations with members of civil society and academics who are experts in race relations,” it would continue to allow people to “demonstrate pride in their ethnic heritage” but “will not tolerate praise or support for white nationalism and white separatism” in applying its “Dangerous Individuals and Organizations” standards.⁷⁶ However, Facebook’s efforts have had decidedly mixed success—in part because rank-and-file members of banned hate groups can simply create new pages.⁷⁷

72. *Id.*

73. *Id.* One such lone wolf live streamed his mass shooting at a New Zealand mosque on Facebook. See Meagan Flynn, *No One Who Watched New Zealand Shooter’s Video Live Reported It to Facebook, Company Says*, WASH. POST (Mar. 19, 2019, 5:04 AM), <https://www.washingtonpost.com/nation/2019/03/19/new-zealand-mosque-shooters-facebook-live-stream-was-viewed-thousands-times-before-being-removed/>.

74. See *What Charlottesville Changed*, POLITICO MAG. (Aug. 12, 2018), <https://www.politico.com/magazine/story/2018/08/12/charlottesville-anniversary-supremacists-protests-dc-virginia-219353> [<https://perma.cc/3YMZ-W27J>].

75. See David Ingram & Ben Collins, *Facebook Bans White Nationalism from Platform After Pressure from Civil Rights Groups*, NBC NEWS (Mar. 27, 2019, 1:06 PM), <https://www.nbcnews.com/tech/tech-news/facebook-bans-white-nationalism-after-pressure-civil-rights-groups-n987991> [<https://perma.cc/C4Y7-WLM4>]; Tony Romm, *A Flood of Online Hate Speech Greets Lawmakers Probing Facebook, Google About White Nationalism*, WASH. POST (Apr. 9, 2019, 10:08 AM), <https://www.washingtonpost.com/technology/2019/04/09/flood-online-hate-speech-greets-lawmakers-probing-facebook-google-about-white-nationalism/>.

76. *Standing Against Hate*, FACEBOOK (Mar. 27, 2019), <https://about.fb.com/news/2019/03/standing-against-hate> [<https://perma.cc/7H8J-BV5N>]. Again, tellingly, Facebook did not mention consulting free speech experts.

77. See Casey Newton, *How White Supremacists Evade Facebook Bans*, VERGE (May 31, 2019, 6:00 AM), <https://www.theverge.com/interface/2019/5/31/18646525/>

Putting aside the efficacy of Facebook's efforts to ban hate groups, terrorist groups, and other dangerous organizations and individuals from its social media platform, the breadth of its censorship of booted speakers is sweeping. Banned organizations and individuals cannot speak on Facebook at all, on any topic, whether personal or political, benign or harmful. They cannot share baby photos or endorse mainstream candidates any more than they can incite violence or spread hate.

In stark contrast, in cases such as *Citizens United v. Federal Election Commission*,⁷⁸ the modern Supreme Court has made clear that, just as "the First Amendment stands against attempts to disfavor certain subjects or viewpoints," it also prohibits "restrictions distinguishing among different speakers, allowing speech by some but not others."⁷⁹ The Court views restrictions against speakers skeptically because it is "all too often simply a means to control content."⁸⁰ Consequently, the Court has held that this general principle applies with special force to "the suppression of political speech based on the speaker's identity."⁸¹ Of course, *Citizens United* has drawn considerable criticism for applying this general principle to corporate campaign speech,⁸² but neither Court decision nor constitutional amendment has overruled it.⁸³

Indeed, of particular relevance, the Court has since applied the First Amendment's robust speaker-based protection to invalidate a state's categorical exclusion of registered sex offenders from social media.⁸⁴

facebook-white-supremacist-ban-evasion-proud-boys-name-change [https://perma.cc/MBG7-FQNJ]; see also Julia Carrie Wong, *White Nationalists Are Openly Operating on Facebook. The Company Won't Act*, GUARDIAN (Nov. 21, 2019, 6:00 AM), https://www.theguardian.com/technology/2019/nov/21/facebook-white-nationalists-banned-red-ice [https://perma.cc/WSW9-HWX6].

78. 558 U.S. 310 (2010).

79. *Id.* at 340; see also *id.* at 341 ("The Government may not . . . deprive the public of the right and privilege to determine for itself what speech and speakers are worthy of consideration.").

80. *Id.* at 340–41.

81. *Id.* at 350.

82. See, e.g., *id.* at 393–95 (Stevens, J., concurring in part and dissenting in part) (criticizing the majority opinion as "misguided" and "a dramatic break from our past"); Barack Obama, *Address Before a Joint Session of the Congress on the State of the Union*, AM. PRESIDENCY PROJ. (Jan. 27, 2010), https://www.presidency.ucsb.edu/documents/address-before-joint-session-the-congress-the-state-the-union [https://perma.cc/BYQ6-NSSP].

83. Cf. JOHN PAUL STEVENS, SIX AMENDMENTS: HOW AND WHY WE SHOULD CHANGE THE CONSTITUTION 57–79 (1st ed. 2014) (arguing for and proposing text of constitutional amendment to overrule *Citizens United*).

84. See *Packingham v. North Carolina*, 137 S. Ct. 1730, 1733 (2017).

The challenger there had violated the state ban by using Facebook, to which the majority opinion of Justice Kennedy (the author of the *Citizens United* majority) gave a judicial “like” as one of “the most important places . . . for the exchange of views” today.⁸⁵ While the Court assumed that the state could enact “specific, narrowly tailored laws” to ban expression that facilitates a sex crime, such as contacting a minor through social media, the Court declared that the state could not “with one broad stroke bar[] access to what for many are the principal sources for knowing current events, checking ads for employment, speaking and listening in the modern public square, and otherwise exploring the vast realms of human thought and knowledge.”⁸⁶

That is precisely the broad stroke Facebook has applied against the “dangerous individuals and organizations” it has banned. Given the constraints of space and focus, the point here is not to assess Facebook’s decision to do so as a private platform owner,⁸⁷ or to question yet again the clarity or line-drawing of its rules,⁸⁸ but to highlight another area in which its asserted “commitment to expression” differs from and falls far short of the First Amendment’s commitment.⁸⁹

Another primary content category in which the Community Standards restrict speech is “Objectionable Content.” This category is in turn divided into five subject areas. To give context to an examination of two of these areas, all five of them and their main policy rationales are set forth in the table below.

85. *Id.* at 1734–35.

86. *Id.* at 1737.

87. *But see supra* note 11 (highlighting articles arguing that Facebook and other social media platforms should be treated as public fora subject to the First Amendment).

88. It would be easy to question the line-drawing in this area, including the decision to allow rank-and-file members of hate organizations to post on the social media platform, as well as perpetrators of single but not multiple murders, and perpetrators of other heinous crimes such as serial child rapists, to use Facebook.

89. *Cf.* T.S. Eliot, *The Hollow Men*, in *SELECTED POEMS* 77, 79 (1st ed. 1930) (“Between the idea/And the reality . . . Falls the Shadow.”).

Table 2: Objectionable Content

Subtopic	Policy Rationale
Hate Speech	“We do not allow hate speech on Facebook because it creates an environment of intimidation and exclusion and in some cases may promote real-world violence.” ⁹⁰
Violent and Graphic Content	“We remove content that glorifies violence or celebrates the suffering or humiliation of others because it may create an environment that discourages participation.” ⁹¹
Adult Nudity and Sexual Activity	“We restrict the display of nudity or sexual activity because some people in our community may be sensitive to this type of content. Additionally, we default to removing sexual imagery to prevent the sharing of non-consensual or underage content.” ⁹²
Sexual Solicitation	“We draw the line . . . when content facilitates, encourages or coordinates sexual encounters between adults. We also restrict sexually explicit language that may lead to solicitation because some audiences within our global community may be sensitive to this type of content and it may impede the ability for

90. *Hate Speech, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/objectionable_content [<https://perma.cc/K9QT-UBQE>].

91. *Violent and Graphic Content, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/graphic_violence [<https://perma.cc/WA9V-GLGD>].

92. *Adult Nudity and Sexual Activity, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity [<https://perma.cc/PG7N-R522>].

	people to connect with their friends and the broader community.” ⁹³
Cruel and Insensitive	“We believe that people share and connect more freely when they do not feel targeted based on their vulnerabilities. As such, we have higher expectations for content that we call cruel and insensitive, which we define as content that targets victims of serious physical or emotional harm.” ⁹⁴

3. *Hate speech*

In addition to banning hate groups, their leaders, and prominent members under the “Violence and Criminal Behavior” category of the Community Standards, Facebook forbids “hate speech” under the “Objectionable Content” category.⁹⁵ The Community Standards define “hate speech” as “a direct attack on people” based on “protected characteristics—race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability.”⁹⁶ The term “attack” is defined in turn as “violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.”⁹⁷

The Community Standards proffer the policy rationale that hate speech “creates an environment of intimidation and exclusion and in some cases may promote real-world violence.”⁹⁸ Censorship in this area also aligns with Zuckerberg’s assertion to Facebook employees that “a lot of people think that we need to be more aggressive in moderating content that is offensive or basically would make certain groups of people feel unsafe.”⁹⁹

93. *Sexual Solicitation, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/sexual_solicitation [<https://perma.cc/5WYP-CSWH>].

94. *Cruel and Insensitive, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/cruel_insensitive [<https://perma.cc/5MJ4-MX8U>].

95. *Hate Speech*, *supra* note 90.

96. *Id.*

97. *Id.*

98. *Id.*

99. Newton, *supra* note 26.

The “Do-not-post” section lists dozens of examples of the kinds of “dehumanizing speech,” “statements of inferiority,” or “calls for exclusion or segregation” that cross the line. They are ostensibly divided into “three tiers of severity,” although the degree of censorship—total—is the same. It is not possible to excerpt a representative sampling of the hodgepodge of verboten hate speech, but the extensive list includes:

- “Tier 1”: various forms of “violent,” “dehumanizing,” or “mocking” speech, such as comparisons of a protected characteristic to insects or other animals perceived as inferior (e.g., “Black people and apes,” “Jewish people and rats,” “Muslim people and pigs”), “unqualified behavioral statements” (e.g., referring to those of a protected characteristic as “violent and sexual criminals” or “bank robbers”), or referring to transgender or non-binary people as “it”;
- “Tier 2”: generalizations regarding the “physical,” “mental,” or “moral” deficiencies of a protected characteristic (e.g., “filthy,” “ignorant,” or “useless”), “expressions of contempt,” “dismissal,” or “disgust” (e.g., saying you “despise,” “hate,” “don’t respect,” “don’t like,” or “don’t care for” a protected characteristic, or that it is “vile” or “disgusting”), or “cursing” with “the intent to insult”;
- “Tier 3”: negative “slurs” regarding a protected characteristic, or “calls for segregation” or “exclusion” (including “not allowed”), except for “criticism of immigration policies and arguments for restricting those policies,” which are expressly permitted.¹⁰⁰

It would be difficult to find a starting—or stopping—point for questioning the clarity of these motley rules or the justifications for their line-drawing. For present purposes, it suffices to raise the problematic nature of these rules and instead focus on the extent to which they differ from established First Amendment principles and censor an extraordinary amount of constitutionally protected speech.

For starters, the Court recently reaffirmed “the ‘bedrock First Amendment principle’ that the government cannot discriminate against ‘ideas that offend.’”¹⁰¹ Today, this bedrock principle undoubtedly applies with full force to hate speech, including speech that “disparage[s] . . . or bring[s] . . . contemp[t] or disrepute” to any

100. *Hate Speech*, *supra* note 90.

101. *Iancu v. Brunetti*, 139 S. Ct. 2294, 2299 (2019) (quoting *Matal v. Tam*, 137 S. Ct. 1744, 1751 (2017)); *see supra* note 20.

persons or classes of people,¹⁰² as well as speech that could “arouse[] anger, alarm or resentment in others on the basis of race, color, creed, religion or gender.”¹⁰³ Indeed, First Amendment protection of speech that offends—as part of its broader protection against the government “discriminat[ing] on the basis of viewpoint” or “disfavor[ing] certain ideas”—is so fundamental that the Court has applied it to bar censorship of speech that otherwise falls outside the protection of the First Amendment, such as censorship of “fighting words” on the basis of their offensiveness rather than their potential to provoke violence.¹⁰⁴

As a result, most if not all of the Community Standards’ far-reaching (if not entirely fathomable) suppression of offensive speech would lie beyond the censorial power of government.¹⁰⁵ For instance, the Community Standards would forbid users from declaring that “men are pigs,” that they “don’t like straight people,” or that “blacks shouldn’t be allowed to vote,” while permitting users to express the opposing views that “men are not pigs,” that they “like straight people,” and that “blacks should be allowed to vote.” In light of these examples—and many more that easily come to mind—Facebook might credibly claim that it restricts speech that “creates an environment of intimidation and exclusion,”¹⁰⁶ but not that it allows users to “talk openly about the issues that matter to them, even if some may disagree or find them objectionable.”¹⁰⁷

102. *Tam*, 137 S. Ct. at 1751 (quoting the Lanham Act, 15 U.S.C. § 1052(a) (2012)).

103. *R.A.V. v. City of St. Paul*, 505 U.S. 377, 380 (1992) (quoting St. Paul Bias-Motivated Crime Ordinance, St. Paul, Minn., Legis. Code § 292.02 (1990)). Of course, *R.A.V.* has received its share of scholarly criticism. See, e.g., Elena Kagan, *Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine*, 63 U. CHI. L. REV. 413, 416 (1996) (discussing some of the “hotly debated” questions that *R.A.V.* has raised regarding “what counts, or should count, as the core concern of the First Amendment”); Steven H. Shiffrin, *Racist Speech, Outsider Jurisprudence, and the Meaning of America*, 80 CORNELL L. REV. 43, 46 (1994) (arguing that “the Court simply bungled the first amendment job in its most important encounter with racist speech in the last forty years”).

104. *Iancu*, 139 S. Ct. at 2297; see *R.A.V.*, 505 U.S. at 386; cf. *Wisconsin v. Mitchell*, 508 U.S. 476, 487–88 (1993), *rev’d sub nom.* *State v. Mitchell*, 504 N.W.2d 610 (Wis. 1993) (upholding sentencing enhancement for racially-motivated violence because it punishes “conduct [that] is thought to inflict greater individual and societal harm” rather than, as in *R.A.V.*, “a class of ‘fighting words’ deemed particularly offensive by the city”).

105. Of course, though the Community Standards’ rules against offensive speech may be extraordinarily overbroad on their face, they may be applied to expression that, in context, may constitute unprotected speech, such as a serious threat of violence against members of a certain race or sexual orientation.

106. *Hate Speech*, *supra* note 90.

107. *Community Standards*, *supra* note 7.

4. *Adult nudity and sexual activity*

Another notable subject of expansive Facebook censorship within the “Objectionable Content” category of the Community Standards is “Adult Nudity and Sexual Activity.”¹⁰⁸ The primary rationale for restricting the display of nudity or sexual activity, including images of real people as well as digitally-created content, is that “some people . . . may be sensitive to this type of content.”¹⁰⁹ A secondary rationale is “to prevent the sharing of non-consensual or underage content.”¹¹⁰

The “Do-not-post” rules ban images in three different areas: “Real nude adults,” “Sexual activity,” and “Digital content.”¹¹¹ The complicated and qualified list of banned imagery includes:

- “Real nude adults” with “visible genitalia” (except in “health-related situations,” such as “birth giving and after-birth moments,” or “visible anus and/or fully nude close-ups of buttocks unless photoshopped on a public figure”); and “uncovered female nipples” (except in “health-related situations,” such as “post-mastectomy, breast cancer awareness or gender confirmation surgery,” or in the context of “breastfeeding,” “birth giving and after-birth moments,” or “an act of protest”);
- “Sexual activity” with “explicit sexual intercourse” where at least one person’s genitals are nude, “implied sexual intercourse” or “implied stimulation of genitalia/anus” even when not directly visible (except in “a sexual health context,” “advertisements,” or “fiction”), other sexual activities (e.g., “erections,” “by-products of sexual activity,” using “sex toys,” or “squeezing female breast”), and “fetish content” (e.g., “dismemberment,” “cannibalism,” or “feces, urine, spit, snot, menstruation, or vomit”);
- “Digital content” that meets any of the above criteria *unless* “sexual activity . . . is not directly visible,” content is “satirical,” “humorous,” “educational,” or “scientific” in context, or “only body shapes or contours are visible.”¹¹²

108. *Adult Nudity and Sexual Activity*, *supra* note 92.

109. *Id.*

110. *Id.* The “Safety” category of the Community Standards contains rules on “Child Nudity and Sexual Exploitation of Children,” which already prohibits a host of content that “sexually exploits or endangers children.” *Child Nudity and Sexual Exploitation of Children, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/child_nudity_sexual_exploitation [https://perma.cc/75FY-VP4L].

111. *Adult Nudity and Sexual Activity*, *supra* note 92.

112. *Id.*

In addition, under the “Policy Rationale,” Facebook states that it allows “photographs of paintings, sculptures, and other art that depicts nude figures.”¹¹³

Facebook relates that its nudity policies have become “more nuanced over time,” now recognizing that nudity can be shared to “raise awareness about a cause,” for “educational or medical reasons,” or for “protest.”¹¹⁴ The evolution of these policies often followed public protests and negative press. For example, policies regarding the display of female breasts have evolved to become less restrictive but more complicated and not much more coherent. Prior to 2014, Facebook barred photos of uncovered female nipples categorically, including breastfeeding photos that show “an exposed breast that is not being used for feeding.”¹¹⁵ However, following years of protests that included an organized social media campaign, “nurse-ins” outside of Facebook offices, and press coverage of the controversy,¹¹⁶ Facebook updated its Community Standards to allow uncovered female nipples “in the context of breastfeeding.”¹¹⁷ Similarly, Facebook’s no-uncovered-female-nipples rule once barred post-mastectomy photos that showed the nipple of a breast unaffected by surgery,¹¹⁸ but after years of public criticism and negative press, Facebook relented to allow them and other exposed nipple photos “in health-related situations.”¹¹⁹

113. *Id.*

114. *Id.*

115. Emil Protalinski, *Facebook Clarifies Breastfeeding Photo Policy*, ZDNET (Feb. 7, 2012, 11:54 PM), <https://www.zdnet.com/article/facebook-clarifies-breastfeeding-photo-policy> [<https://perma.cc/CV6T-82F4>].

116. *Id.*; see Mikaela Conley, *Breastfeeding Advocates Protest Facebook with Nurse-In*, ABC NEWS (Feb. 7, 2012, 11:38 AM), <https://abcnews.go.com/Health/breastfeeding-advocates-hold-facebook-protest/story?id=15530012> [<https://perma.cc/8FJ3-6U4K>].

117. *Adult Nudity and Sexual Activity*, *supra* note 92; see *Does Facebook Allow Photos of Mothers Breastfeeding?*, FACEBOOK, <https://www.facebook.com/help/340974655932193> [<https://perma.cc/U6DR-8MAC>]; see also Soraya Chemaly, *#FreeTheNipple: Facebook Changes Breastfeeding Mothers Photo Policy*, HUFFPOST (June 9, 2014, 6:48 PM), https://www.huffpost.com/entry/freethenipple-facebook-changes_b_5473467 [<https://perma.cc/SPG2-NAG9>].

118. See Susan Donaldson James, *Facebook Allows Mastectomy Photos, Not Nudity*, ABC NEWS (June 12, 2013), <https://abcnews.go.com/blogs/health/2013/06/12/facebook-launches-new-policy-to-allow-mastectomy-photos> [<https://perma.cc/NPB2-2AY7>].

119. *Adult Nudity and Sexual Activity*, *supra* note 92; see Jessica Firger, *Breast Cancer Survivor Battles Facebook over Mastectomy Photos*, CBS NEWS (May 9, 2014, 4:16 PM), <https://www.cbsnews.com/news/breast-cancer-survivor-battles-facebook-over-mastectomy-photos> [<https://perma.cc/MGT7-FLVW>]; Margaret Wheeler Johnson, *Facebook Revises Wording of Policy on Post-Mastectomy Photos*, HUFFPOST (June 12, 2013,

After these and other modest changes to Facebook's nudity policies,¹²⁰ the Community Standards still draw a sharp censorial line between the display of nude male breasts, which generally are allowed, and nude female breasts, which generally are not, with a handful of exceptions. Indeed, the censorial line between female and male breasts extends beyond nudity, as the Community Standards forbid any display of “[s]queezing female breast,” even if clothed, whereas it allows images of male breasts being squeezed in any state of (un)dress.¹²¹ This differential treatment has led to continued criticism of Facebook for maintaining content policies that reflect and reinforce gender discrimination and sexualized objectification of the female body.¹²² For instance, to put this differential treatment on public display, nude female protestors laid in front of Facebook headquarters with their nipples covered by stickers of male nipples and their genitalia covered by larger prints of male nipples, thereby absurdly ensuring that posted photos of their protests would comply with Facebook's nudity policies.¹²³

Much more could be written about the lack of clarity, coherence, or justification for the many questionable lines drawn in the “Adult Nudity and Sexual Activity” subject area beyond its fixation with female

4:59 PM), https://www.huffpost.com/entry/facebook-mastectomy-photos-allowed_n_3428602 [<https://perma.cc/US65-HUEV>]; see also *Does Facebook Allow Post-Mastectomy Photos?*, FACEBOOK, https://www.facebook.com/help/318612434939348?helpref=uf_permalink [<https://perma.cc/H95X-SUEW>].

120. The Community Standards now also exempt displays of uncovered female nipples in the context of “gender confirmation surgery” or “an act of protest.” *Adult Nudity and Sexual Activity*, *supra* note 92; see *supra* note 112 and accompanying text.

121. See *Adult Nudity and Sexual Activity*, *supra* note 92; *supra* note 112 and accompanying text.

122. See, e.g., Kari Paul, *Naked Protesters Condemn Nipple Censorship at Facebook Headquarters*, GUARDIAN (June 3, 2019, 3:09 PM), <https://www.theguardian.com/technology/2019/jun/03/facebook-nude-nipple-protest-wethenipple> [<https://perma.cc/2AV7-DS2G>].

123. *Id.* Displays of topless females (sans stickers) in this context might also fall within the exemption for “an act of protest.” However, it is not clear that the exemption applies to protests of Facebook's policies as opposed to (other) political protests. See Megan Farokhmanesh, *Facebook Sure Has Been Thinking a Lot About Nipples*, VERGE (Apr. 24, 2018, 1:23 PM), <https://www.theverge.com/2018/4/24/17275114/facebook-community-guidelines-nipples-nudity> [<https://perma.cc/E53V-E247>]; Sophie Roberts, *The Breast Explanation: What Is Free the Nipple, What Celebrities Are Involved, Is It Illegal to Go Naked in Public, and What Counts as Graphic Content on Facebook?*, SUN (Mar. 26, 2019, 9:00 PM), <https://www.thesun.co.uk/fabulous/2778411/free-the-nipple-campaign-celebrity-supporters-laws-illegal-facebook-instagram> [<https://perma.cc/QR2S-MT69>].

breasts. For example, ordinary users cannot post images of “[i]mplied sexual intercourse” or “[i]mplied stimulation of genitalia,” even when “the activity is not directly visible,” despite the significance of sexuality in expression and Facebook’s “commitment to voice.”¹²⁴ Yet the Community Standards nakedly exempt advertisers from these rules, despite Facebook’s overall policy rationale for censoring speech in this area to protect the sensitivities of community members.¹²⁵ That communitarian concern apparently takes a back seat to advertising dollars.¹²⁶

Furthermore, while Facebook updated its policies in 2018 to allow “photographs of paintings, sculptures, and other art that depicts nude figures,” photographs of real-life nude models still apparently do not qualify as “other art,” no matter how creative or acclaimed.¹²⁷ Thus, incongruously, users may post photos of artistic nude figures but not artistic nude photos.¹²⁸

The Community Standards also offer no guidance on what kinds of imagery of sexual activity count as exempted “cases of a sexual health context” (a video on the sex positivity of masturbation?), what qualifies as “educational” digital content (a fetish video could be eye-opening to the uninitiated), or how to assess the many other lines that the Community Standards attempt to draw in this area.¹²⁹ And there is no mention of verbal, as opposed to visual, imagery of sexual activity,

124. See Andrew Koppelman, *Is Pornography “Speech”?*, 14 LEGAL THEORY 71 (2008) (arguing that pornography cannot be distinguished from art as protected speech); cf. *Cohen v. California*, 403 U.S. 15, 25 (1971) (“For, while the particular four-letter word being litigated here is perhaps more distasteful than most others of its genre, it is nevertheless often true that one man’s vulgarity is another’s lyric.”).

125. *Adult Nudity and Sexual Activity*, *supra* note 92; Community Standards, *supra* note 7.

126. Cf. RUSH, *The Big Money*, on POWER WINDOWS (Mercury Records 1985) (“Big money got a mighty voice”). Rest in peace, Neil Peart.

127. Matthew Robinson, *Facebook to Meet #WeTheNipple Campaigners Amid Nudity Censorship Row*, CNN (June 6, 2019), <https://www.cnn.com/style/article/facebook-wethenipple-protesters-meeting-intl-scli/index.html> [<https://perma.cc/QR2-RSVP>]; Michael Zhang, *Facebook to Reconsider Nudity Policy After Photographers Protest*, PETAPIXEL (June 6, 2019), <https://petapixel.com/2019/06/06/facebook-to-reconsider-nudity-policy-after-photographers-protest> [<https://perma.cc/AC8S-2T87>]. These articles report that Facebook promised to hold a meeting with photographers and other protestors to discuss its nudity policies, but expressly declined to promise any policy changes.

128. Would a photo of an artistic nude photo (perhaps hanging in a museum) qualify, or only a photo of nudity in non-photographic mediums such as paintings and sculptures? That this question can be asked—but not answered—again illustrates the lack of clarity and coherence that pervades the Community Standards.

129. *Adult Nudity and Sexual Activity*, *supra* note 92.

which of course can be quite graphic and offensive to the sensibilities of “some people” in the vast Facebook community.¹³⁰

Regardless of where precisely (or not) these various lines might fall, they individually and collectively fall far from the line of unprotected obscenity under the First Amendment. After all, the multipronged test from *Miller v. California*¹³¹ for obscenity narrowly requires that the average person applying contemporary community standards finds that the work “appeal[s] to the prurient interest,” that it depicts state-defined “sexual conduct in a patently offensive way,” and that, taken as a whole, the work lacks “serious literary, artistic, political, or scientific value.”¹³² While this First Amendment test is not without its critics¹³³ or ambiguities,¹³⁴ none of Facebook’s broad do-not-post rules even attempt to limit their application to materials that are “prurient” or “patently offensive” (case-in-point: the ban on uncovered female nipples). Nor do any of the rules against adult nudity contain a safety valve for works with “literary, artistic, political, or scientific value” (beyond a peculiarly limited allowance for photoshopped buttocks on public figures).¹³⁵ Indeed, it is difficult to identify a single do-not-post rule in this area that does not sweep beyond obscenity.

Facebook has justified its wide-ranging censorship of nudity and sexual imagery by asserting that its rules “closely mirror[] the policy that governs broadcast television.”¹³⁶ But this assertion is inaccurate. The Federal Communications Commission’s (FCC) prohibition of “indecent” content on broadcast television and radio, while broader than obscenity under the *Miller* test, nevertheless requires that a depiction of “sexual or excretory organs” be “patently offensive as measured by contemporary community standards for the broadcast medium.”¹³⁷ Furthermore, what the FCC considers “patently offensive” is informed by three “significant” factors: the “explicitness or graphic

130. *Id.*

131. 413 U.S. 15 (1973).

132. *Id.* at 24.

133. See, e.g., David A.J. Richards, *Free Speech and Obscenity Law: Toward a Moral Theory of the First Amendment*, 123 U. PA. L. REV. 45, 72 (1974) (arguing that “[a]n understanding of the moral function of the first amendment compels a conclusion” that the *Miller* decision finding obscenity outside first amendment protection is wrong).

134. See CHEMERINSKY, *supra* note 19, at 1113–14 (critiquing the ambiguity of the “prurient interest” and “patently offensive” prongs).

135. See *Adult Nudity and Sexual Activity*, *supra* note 92.

136. Protalinski, *supra* note 115.

137. *FCC v. Fox Television Stations, Inc.*, 567 U.S. 239, 246 (2012); see *Miller*, 413 U.S. at 24.

nature” of the depiction, the extent to which the material “dwells on or repeats at length” the sexual or excretory organs, and whether the content appears to “pander,” “titillate,” or is presented for “shock value.”¹³⁸ By contrast, none of the Community Standards’ do-not-post rules limit their applicability to content that is “patently offensive” or incorporate any of the factors that inform the FCC’s analysis. Moreover, the Supreme Court has upheld the FCC’s regulation of indecency in broadcast television and radio because of broadcasting’s “uniquely pervasive presence” and easy accessibility to children.¹³⁹ These rationales do not fit as neatly in the social media context, where children’s access depends on the more affirmative (and, in theory, adult-assisted) steps of obtaining an internet connection, opening a social media account, and signing in.

In the end, Facebook’s extensive censorship in the areas of violence and incitement, dangerous organizations and individuals, hate speech, and adult nudity and sexual activity illustrate not only the lack of clarity, consistency, and coherence of its speech code, but moreover the extent to which the company’s touted “commitment to expression”¹⁴⁰ is substantially qualified by its asserted interests in promoting safety, dignity, and inclusion, as well as in avoiding offense to user sensitivities.¹⁴¹ This aggressive censorship is consistent with Zuckerberg’s internal explanation to employees that “a lot of people think that we need to be more aggressive in moderating content that is offensive.”¹⁴² Such censorship promotes Facebook’s recently stated aim of improving user satisfaction so that they “feel positive, rather than negative, after visiting,” thereby increasing their likelihood of coming back.¹⁴³

138. *Fox*, 567 U.S. at 246. In *Fox*, the FCC unsuccessfully sought to apply this standard to the television broadcast of fleeting expletives and a scene showing the nude buttocks of a female character. The Court held that the FCC’s application of the “indecency” standard to those instances violated the Due Process Clause for lack of fair notice and, therefore, declined to address the First Amendment implications of the agency’s policy. *See id.* at 258.

139. *Id.* at 244 (quoting *FCC v. Pacifica Found.*, 438 U.S. 726, 748–49 (1978)).

140. *See* Community Standards, *supra* note 7.

141. *See supra* notes 17, 109 and accompanying text.

142. *See supra* notes 26–27 and accompanying text.

143. Mike Isaac, *Facebook Overhauls News Feed to Focus on What Friends and Family Share*, N.Y. TIMES (Jan. 11, 2018) <https://www.nytimes.com/2018/01/11/technology/facebook-news-feed.html>.

II. FACEBOOK'S POLICIES ON POLITICIANS, POLITICAL SPEECH, AND FALSEHOODS

As the foregoing Part illustrates, the Community Standards' extensive censorship of the speech of billions of users worldwide makes Facebook's speech code one of the most impactful in human history. But just as important as *what* Facebook censors through its Community Standards is *who* Facebook exempts from that censorship: politicians. This Part examines how Facebook's special treatment of politicians and their falsehoods interacts with its personalized news feed and powerful ad targeting tools to optimize the platform for the viral spread of disinformation that promotes the company's business interests but grossly distorts democratic deliberation.

A. Controversies

At first blush, Facebook's "commitment to expression" seems most fully realized in its highly public resistance to suppressing political speech on its platform. Indeed, Facebook unfailingly invokes the robustness of that commitment in justifying its refusal to take down false or misleading political speech—including deliberate lies to influence the electorate—despite mounting criticism in the face of high-profile controversies.

For example, Facebook refused to take down an ad by President Donald Trump's 2020 reelection campaign asserting that his potential democratic rival, former Vice President Joe Biden, "promised Ukraine \$1 billion if they fired the prosecutor investigating his son's company."¹⁴⁴ Though CNN refused to air the ad because it made "assertions that have been proven demonstrably false" by fact-checkers at various news outlets, including CNN itself,¹⁴⁵ Facebook rejected the Biden campaign's request to take the ad off its platform.¹⁴⁶ Facebook explained that "when a politician speaks or makes an ad," the company does not subject the content to its fact-checking policies because of its "fundamental belief in free expression, respect for the democratic

144. Donald J. Trump, *Biden Corruption*, YOUTUBE (Sept. 27, 2019), <https://youtu.be/bbixdV2F6Ts> [<https://perma.cc/G7AN-WNVH>]; see Michael M. Grynbaum & Tiffany Hsu, *CNN Rejects 2 Trump Campaign Ads, Citing Inaccuracies*, N.Y. TIMES (Oct. 3, 2019), <https://www.nytimes.com/2019/10/03/business/media/cnn-trump-campaign-ad.html>; Cecilia Kang, *Facebook's Hands-Off Approach to Political Speech Gets Impeachment Test*, N.Y. TIMES, (Oct. 8, 2019), <https://www.nytimes.com/2019/10/08/technology/facebook-trump-biden-ad.html?auth=login-email&login=email>.

145. Grynbaum & Hsu, *supra* note 144.

146. Kang, *supra* note 144.

process, and the belief that, in mature democracies with a free press, political speech is already arguably the most scrutinized speech there is.”¹⁴⁷ Likewise, this refusal aligns with a fall 2019 policy address by Zuckerberg at Georgetown University, “Standing For Voice and Free Expression.”¹⁴⁸ He asserted that the company neither fact checks nor takes down speech by politicians “even if it would otherwise conflict with many of our standards,” because “people should be able to see for themselves what politicians are saying,” and “I don’t think it’s right for a private company to censor politicians or the news in a democracy.”¹⁴⁹

Similarly, the company refused to take down a doctored video of House Speaker Nancy Pelosi that had drawn millions of views on a conservative group’s Facebook page and spread virally across Twitter and YouTube.¹⁵⁰ The video had been edited to slow Pelosi’s speech while retaining its pitch so that her words sounded drunkenly slurred.¹⁵¹ As it spread, the video drew comments highlighting and mocking her apparent drunkenness.¹⁵² News organizations and independent fact-checking groups deemed the video fake, and YouTube removed it.¹⁵³ However, Facebook explained that “[w]e don’t have a policy that stipulates that the information you post on Facebook must be true.”¹⁵⁴ Instead, the company stated that it would reduce the video’s appearance in users’ news feeds and surface links alongside it to fact checking and additional reporting on the video.¹⁵⁵

147. Donie O’Sullivan (@donie), TWITTER (Oct. 8, 2019, 11:58 PM), <https://twitter.com/donie/status/1181780966806839297> [<https://perma.cc/V2BA-EW36>]. This is the same explanation the company gives for exempting posts and ads from politicians from its fact-checking policy. *See infra* notes 187–88 and accompanying text.

148. Mark Zuckerberg *Stands for Voice and Free Expression*, *supra* note 31.

149. *Id.*

150. See Drew Harwell, *Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media*, WASH. POST (May 24, 2019, 4:41 PM), <https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media>.

151. *See id.*

152. *See id.*

153. Drew Harwell, *Facebook Acknowledges Pelosi Video Is Faked But Declines to Delete It*, WASH. POST (May 24, 2019, 4:17 PM), <https://www.washingtonpost.com/technology/2019/05/24/facebook-acknowledges-pelosi-video-is-faked-declines-delete-it>; *see* Miriam Valverde, *Viral Video of Nancy Pelosi Slowed Down Her Speech*, POLITIFACT (May 24, 2019), <https://www.politifact.com/truth-o-meter/statements/2019/may/24/politics-watchdog/viral-video-nancy-pelosi-speech-was-manipulated> [<https://perma.cc/H3JL-86JX>].

154. Harwell, *supra* note 153.

155. *Id.*

The company's refusal to remove false and misleading political content on its platform has drawn fierce public criticism from many quarters, including political and technology writers,¹⁵⁶ prominent opinion pages,¹⁵⁷ Facebook's own employees,¹⁵⁸ and lawmakers.¹⁵⁹ Most of the latter have been Democratic politicians, who have frequently found themselves targets of misinformation spread by President Trump or other right-wing critics.¹⁶⁰ One notable response came from Senator Elizabeth Warren, a Democratic candidate in the 2020 presidential election, who posted a fake ad claiming that "Mark Zuckerberg and Facebook just endorsed Donald Trump for re-election."¹⁶¹ The ad went on to disclose that the claim was false and criticize Zuckerberg for giving Trump "free rein to lie on his platform—and then to pay Facebook gobs of money to push out their lies to American voters."¹⁶²

156. See, e.g., Sue Halpern, *Facebook's False Standards for Not Removing a Fake Nancy Pelosi Video*, NEW YORKER (May 28, 2019), <https://www.newyorker.com/tech/annals-of-technology/facebooks-false-standards-for-not-removing-a-fake-nancy-pelosi-video> [<https://perma.cc/S6J7-7D3J>]; Emily Stewart, *Facebook's Political Ads Policy Is Predictably Turning Out to Be a Disaster*, VOX (Oct. 30, 2019, 4:57 PM), <https://www.vox.com/recode/2019/10/30/20939830/facebook-false-ads-california-adriel-hampton-elizabeth-warren-aoc> [<https://perma.cc/8EHD-W4K3>].

157. See, e.g., Aaron Sorkin, *Aaron Sorkin: An Open Letter to Mark Zuckerberg*, N.Y. TIMES (Oct. 31, 2019), <https://www.nytimes.com/2019/10/31/opinion/aaron-sorkin-mark-zuckerberg-facebook.html>.

158. Mike Isaac, *Dissent Erupts at Facebook over Hands-Off Stance on Political Ads*, N.Y. TIMES (Oct. 28, 2019), <https://www.nytimes.com/2019/10/28/technology/facebook-mark-zuckerberg-political-ads.html>.

159. See *id.*

160. See, e.g., Jeff Horwitz, *Pelosi Slams Facebook over Altered Video*, WALL ST. J. (May 29, 2019, 6:38 PM), https://www.wsj.com/articles/pelosi-slams-facebook-over-altered-video-11559164773?mod=article_inline; Cecilia Kang & Mike Isaac, *Biden Escalates Attack on Facebook over False Political Ads*, N.Y. TIMES (Oct. 18, 2019), <https://www.nytimes.com/2019/10/18/technology/biden-facebook-ad.html>; Sarah Mervosh, *Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump*, N.Y. TIMES (May 24, 2019), <https://www.nytimes.com/2019/05/24/us/politics/pelosi-doctored-video.html>; Craig Timberg et al., *A Facebook Policy Lets Politicians Lie in Ads, Leaving Democrats Fearing What Trump Will Do*, WASH. POST (Oct. 10, 2019, 4:31 PM), <https://www.washingtonpost.com/technology/2019/10/10/facebook-policy-political-speech-lets-politicians-lie-ads>.

161. Dan MacGuill, *Did Elizabeth Warren's Campaign Say Mark Zuckerberg Had Endorsed Trump?*, SNOPE (Nov. 8, 2019), <https://www.snopes.com/fact-check/liz-warren-zuckerberg-trump> [<https://perma.cc/DVG4-E8B4>]; see Madeleine Carlisle, *Elizabeth Warren Targets Facebook's Ad Policy with a Fake Ad Claiming Mark Zuckerberg Endorsed Trump's Re-Election*, TIME (Oct. 12, 2019), <https://time.com/5699122/elizabeth-warren-facebook-fake-ad-mark-zuckerberg> [<https://perma.cc/5GDH-Z46U>].

162. Carlisle, *supra* note 161.

On the other side, conservatives from Trump on down have criticized Facebook and other online platforms for what they perceive to be liberal bias in content moderation.¹⁶³ Even though no credible evidence of systemic bias exists,¹⁶⁴ they have pointed to instances such as Facebook fact-checking as “inaccurate” an anti-abortion video claiming that “abortion is never medically necessary.”¹⁶⁵ Facebook ended up removing the fact-checking label and supporting links after Republican senators sent it a letter of complaint.¹⁶⁶ Episodes like this undoubtedly informed Zuckerberg’s assertion to employees that, in the arena of “political discourse,” a “lot of people feel” that those “arbitrating what is misinformation and doing fact-checking tend to be left of center” and are “getting in the way of an ability to express something that they feel is real.”¹⁶⁷

In answer to the growing criticism of its platform from “different sides” leading up to the 2020 elections,¹⁶⁸ Facebook has reaffirmed its basic policies regarding political speech but also adopted some consequential and controversial new ones, including its general exemption of politicians from its Community Standards.

163. See, e.g., David Shepardson, *Facebook, Google Accused of Anti-Conservative Bias at U.S. Senate Hearing*, REUTERS (Apr. 10, 2019, 5:35 PM), <https://www.reuters.com/article/us-usa-congress-socialmedia/facebook-google-accused-of-anti-conservative-bias-at-u-s-senate-hearing-idUSKCN1RM2SJ> [<https://perma.cc/E888-L2SC>].

164. See Mathew Ingram, *The Myth of Social Media Anti-Conservative Bias Refuses to Die*, COLUM. JOURNALISM REV. (Aug. 8, 2019), https://www.cjr.org/the_media_today/platform-bias.php [<https://perma.cc/ATG8-9X44>].

165. Isobel Asher Hamilton, *Facebook Removed a Fact-Check on an Anti-Abortion Video After Republican Senators Including Ted Cruz Complained*, BUS. INSIDER (Sept. 13, 2019, 6:27 AM), <https://www.businessinsider.com/facebook-removes-anti-abortion-factcheck-after-senators-complain-2019-9> [<https://perma.cc/E4GY-G55Z>].

166. *Id.*

167. See Newton, *supra* note 26. Facebook also faced conservative criticism in 2016 when it used to serve “Trending” topics at the top of its News Feed, which excluded stories covered by conservative outlets such as Breitbart unless the same stories were also covered by mainstream news sites such as the *New York Times*. See Haan, *supra* note 11, at 23. Facebook’s response to this criticism—the elimination of human editors and reliance on algorithm-made selections based on popularity among users—amplified false news stories during the 2016 election cycle. See *id.* at 23–24.

168. See Newton, *supra* note 26.

B. Policies

Just as no provision of the Community Standards guarantees users any general right of expression on Facebook,¹⁶⁹ no provision affirmatively protects political speech on the platform. Indeed, with one notable exception, political speech does not appear to receive any special protection in the Community Standards. It is permitted to the same extent as speech on any other subject, including personal ones such as vacation photos or cat videos. As the Introduction to the Community Standards states, users can “talk openly about the issues that matter to them”—political, personal, or otherwise—so long as they do not violate any of the content-moderation rules.¹⁷⁰

The notable exception lies in the Introduction to the Community Standards, which does not protect political speech per se, but states, “In some cases, we allow content which would otherwise go against our Community Standards—if it is newsworthy and in the public interest.”¹⁷¹ Critically, Facebook does not define what is “newsworthy” or “in the public interest.” However, the company does hedge that it makes allowances “only after weighing the public interest value against the risk of harm,” and it “look[s] to international human rights standards”—not First Amendment standards—“to make these judgments.”¹⁷²

For ordinary users, it appears remarkably difficult even for exceptionally newsworthy content to qualify. For instance, Facebook took down the Pulitzer Prize-winning photograph “Napalm Girl” from the Vietnam War, shared by a Norwegian writer posting about photos that “changed the history of warfare,” for violating the Community Standards’ rules against nudity generally as well as child nudity.¹⁷³ Facebook confirmed that the platform’s human editors, rather than algorithms, made the call to take down the photo.¹⁷⁴ It was only after

169. See *supra* Section I.A.

170. See *Community Standards*, *supra* note 7.

171. *Id.*

172. *Id.* The Community Standards Introduction does not specify which international human rights standards it consults. Cf. Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 26 (2018) (exploring what online platforms would need to do to comply with Article 19 of the International Covenant on Civil and Political Rights, “the key international standard for protecting freedom of expression”).

173. Aarti Shahani, *With ‘Napalm Girl,’ Facebook Humans (Not Algorithms) Struggle to Be Editor*, NPR (Sept. 10, 2016, 11:12 PM), <https://www.npr.org/sections/alltechconsidered/2016/09/10/493454256> [<https://perma.cc/KWU4-Z7CU>].

174. See *id.*

global criticism from news organizations, media experts, and even the Norwegian prime minister, who reposted the photo on Facebook, that the company reversed itself.¹⁷⁵ But not every newsworthy post by an ordinary user can claim the public interest value of a Pulitzer Prize or inspire a worldwide campaign in support of its newsworthiness.

By contrast, the posts or ads of politicians would not need any such backing. Facebook publicly declared in the run-up to the 2020 election that it would apply this newsworthiness exemption categorically to the posts and ads of “politicians.” It announced that “from now on we will treat speech from politicians as newsworthy content that should, as a general rule, be seen and heard.”¹⁷⁶ Additionally, Facebook affirmed that the platform categorically “exempts politicians from our third-party fact-checking program.”¹⁷⁷

As background, the company started fact-checking news in the spring of 2017, following revelations that the platform amplified false news stories and significantly promoted the spread of Russian disinformation in the 2016 election.¹⁷⁸ Even though Zuckerberg initially dismissed criticism that Facebook “influenced the election in any way” as “a pretty crazy idea,”¹⁷⁹ the company subsequently declared

175. See *id.*; see also Sam Levin et al., *Facebook Backs Down from ‘Napalm Girl’ Censorship and Reinstates Photo*, GUARDIAN (Sept. 9, 2016, 1:44 PM), <https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo> [<https://perma.cc/87AP-7JY4>].

176. Nick Clegg, *Facebook, Elections and Political Speech*, FACEBOOK (Sept. 24, 2019), <https://about.fb.com/news/2019/09/elections-and-political-speech> [<https://perma.cc/END3-8LAN>].

177. *Id.*

178. See Joseph Thai, *The Right to Receive Foreign Speech*, 71 OKLA. L. REV. 269, 270–72 (2018).

179. Casey Newton, *Zuckerberg: The Idea that Fake News on Facebook Influenced the Election Is ‘Crazy’*, VERGE (Nov. 10, 2016, 9:10 PM), <https://www.theverge.com/2016/11/10/13594558/mark-zuckerberg-election-fake-news-trump> [<https://perma.cc/8K36-4MN6>]. But see Margaret Sullivan, *Facebook’s Role in Trump’s Win Is Clear. No Matter What Mark Zuckerberg Says.*, WASH. POST (Sept. 7, 2017), <https://www.washingtonpost.com/lifestyle/style/facebooks-role-in-trumps-win-is-clear-no-matter-what-mark-zuckerberg-says/2017/09/07/b5006c1c-93c7-11e7-89fa-bb822a46da5b>. Zuckerberg since apologized for not doing enough to prevent the spread of “fake news” and “foreign interference in elections.” See Danielle Kurtzleben, *Did Fake News on Facebook Help Elect Trump? Here’s What We Know*, NPR (Apr. 11, 2018, 7:00 AM), <https://www.npr.org/2018/04/11/601323233/6-facts-we-know-about-fake-news-in-the-2016-election> [<https://perma.cc/78YF-N2T7>].

that it is “committed to fighting the spread of false news”¹⁸⁰ as “a tool for economic or political gains” that is “bad for people and bad for Facebook.”¹⁸¹ Accordingly, the company adopted a “Fact-Checking on Facebook” policy.¹⁸² In partnership with third-party fact-checkers, the company now attempts to identify false news content (defined as “factually inaccurate” or “misleading”), reduce its distribution and lower its appearance in users’ feeds, flag it with a warning about its falsehood, and notify users trying to share it of additional reporting.¹⁸³

So far, Facebook has not disclosed enough data to assess the extent to which its fact checking has stemmed the proliferation of falsehoods on its platform.¹⁸⁴ However, critics have contended that Facebook has not devoted nearly enough resources to address the escalating problem.¹⁸⁵ Furthermore, several of its fact-checking partners have quit in frustration, concerned that their work was “not a priority” and the company was “more interested in making themselves look good and passing the buck.”¹⁸⁶

In any case, Facebook’s refusal to fact-check “[p]osts and ads from politicians” effectively gives them a license to lie on the platform.¹⁸⁷ This license applies to current officeholders and candidates “at every level”—

180. *Fact-Checking on Facebook: What Publishers Should Know*, FACEBOOK, <https://www.facebook.com/help/publisher/18222309230722?ref=MisinformationPolicyPage> [<https://perma.cc/D25T-MLBQ>].

181. Tessa Lyons, *Hard Questions: What’s Facebook’s Strategy for Stopping False News?*, FACEBOOK (May 23, 2018), <https://about.fb.com/news/2018/05/hard-questions-false-news> [<https://perma.cc/5USS-SJ2L>].

182. *Fact-Checking on Facebook*, *supra* note 180.

183. *Id.* The fact-checking policy indicates that repeat offenders are subject to additional steps, such as having the distribution of their pages and domain reduced or (after further violations) revoked. *Id.*

184. Mathew Ingram, *Facebook’s Fact-Checking Program Falls Short*, COLUM. JOURNALISM REV., Aug. 2, 2019, https://www.cjr.org/the_media_today/facebook-fact-checking.php [<https://perma.cc/E95F-UFYG>].

185. See Chris Mills Rodrigo, *Critics Fear Facebook Fact-Checkers Losing Misinformation Fight*, THE HILL (Jan. 20, 2020, 7:30 AM), <https://thehill.com/policy/technology/478896-critics-fear-facebook-fact-checkers-losing-misinformation-fight> [<https://perma.cc/VX4Z-ZKMQ>].

186. Sam Levin, *‘They Don’t Care’: Facebook Factchecking in Disarray as Journalists Push to Cut Ties*, GUARDIAN (Dec. 13, 2018, 3:00 AM), <https://www.theguardian.com/technology/2018/dec/13/they-dont-care-facebook-fact-checking-in-disarray-as-journalists-push-to-cut-ties> [<https://perma.cc/KC7R-KWJB>].

187. *Fact-Checking on Facebook*, *supra* note 180. Facebook qualifies that this exemption only applies to the “direct speech” of politicians, rather than to content that they pass on, such as links to an article or video “created by someone else.” *Id.* But a politician can easily skirt this qualification by repeating or endorsing a false claim made by someone else.

presumably from President of the United States to dog catcher of Duxbury, Vermont—but does not apply to former candidates, former officials, or unaffiliated super political action committees (so-called super PACs).¹⁸⁸ Consequently, campaigns with the financial resources can leverage Facebook’s powerful ad targeting tools and deep user knowledge to spread politically expedient disinformation to the multitude of voters most receptive to them.¹⁸⁹ Facebook does not deny this.

Indeed, Twitter’s CEO Jack Dorsey acknowledged this very danger when his company banned all political ads on its platform. He warned that “machine learning-based optimization of messaging and microtargeting” can spread “unchecked misleading information” at “increasing velocity, sophistication, and overwhelming scale” that “today’s democratic infrastructure may not be prepared to handle.”¹⁹⁰ Similarly, in recommending that online platforms ban microtargeting of political ads, Federal Election Commission Chair Ellen Weintraub sounded the alarm against algorithms that make it “easy to single out susceptible groups and direct political misinformation to them with little accountability, because the public at large never sees the ad.”¹⁹¹

188. *Id.*; see Scott Detrow, ‘You Couldn’t Get Elected Dogcatcher!’ No, Seriously, NPR (Apr. 7, 2018, 8:45 AM), <https://www.npr.org/2018/04/07/600482792/you-couldn-t-get-elected-dogcatcher-no-seriously> [<https://perma.cc/YSP5-TU8S>]. The policy indicates that “politicians” exempt from its fact-checking at least include “candidates running for office, current office holders—and, by extension, many of their cabinet appointees—along with political parties and their leaders.” *Fact-Checking on Facebook*, *supra* note 180.

189. See Sue Halpern, *The Problem of Political Advertising on Social Media*, NEW YORKER (Oct. 24, 2019), <https://www.newyorker.com/tech/annals-of-technology/the-problem-of-political-advertising-on-social-media> [<https://perma.cc/M3Z6-HWJQ>] (“Facebook’s tools, and its unprecedented cache of data, allows marketers—both commercial and political—to test various approaches and identify users who are most susceptible to their message.”); Kevin Roose et al., *Don’t Tilt Scales Against Trump, Facebook Executive Warns*, N.Y. TIMES (Jan. 7, 2020), <https://www.nytimes.com/2020/01/07/technology/facebook-trump-2020.html> (reporting on a leaked internal memo of a Facebook executive crediting Trump’s effective use of the platform’s advertising tools for his 2016 victory); *supra* notes 156–62 and accompanying text.

190. Jack Dorsey (@jack), TWITTER (Oct. 30, 2019, 4:05 PM), <https://twitter.com/jack/status/1189634360472829952?s=20> [<https://perma.cc/C4S9-VX9R>]; see Kate Conger, *Twitter Will Ban All Political Ads, C.E.O. Jack Dorsey Says*, N.Y. TIMES (Oct. 30, 2019), <https://www.nytimes.com/2019/10/30/technology/twitter-political-ads-ban.html>.

191. Ellen L. Weintraub, *Don’t Abolish Political Ads on Social Media. Stop Microtargeting*, WASH. POST (Nov. 1, 2019, 6:51 PM), <https://www.washingtonpost.com/opinions/2019/11/01/dont-abolish-political-ads-social-media-stop-microtargeting>. Following Commissioner Weintraub’s recommendation in part, Google announced that it will not allow political ad targeting on the basis of public voter records (e.g., party affiliation) or political characteristics such as “left-leaning,” “right-leaning,” or

By contrast, Facebook continues to justify the free reign it gives politicians to spread disinformation in terms of its “fundamental belief in free expression” and “respect for the democratic process.”¹⁹² Expanding on this justification, the company asserts that “by limiting political speech we would leave people less informed about what their elected officials are saying and leave politicians less accountable for their words.”¹⁹³ Likewise, echoing Zuckerberg, other Facebook executives have contended that it would be inappropriate for Facebook to be “the arbiter of what’s true and what’s false”¹⁹⁴ or otherwise “referee political debates and prevent a politician’s speech from reaching its audience and being subject to public debate and scrutiny.”¹⁹⁵

C. Critiques

Facebook’s assertedly hands-off approach to the speech of politicians may lend credence to its ostensibly principled defense of “expression,” “voice,” and non-interference with political discussion out of “respect for the democratic process.” But Facebook’s free speech and democratic process justifications for singularly exempting politicians from its Community Standards and fact-checking policy do not withstand scrutiny. Simply put, Facebook is not a neutral platform that promotes an open marketplace of ideas or facilitates robust public debate across the political spectrum. Quite the opposite. The company takes an aggressive hands-on approach to censoring and shaping political speech that divides and distorts democratic deliberation in a number of harmful ways.

First, Facebook heavily filters and skews what every user sees based on content and viewpoint. The platform employs machine-learning algorithms that continually adapt and narrow each user’s news feed “to connect people to the stories they care about most,” based on their “connections and activity on Facebook.”¹⁹⁶ The predictable, widely

“independent.” Daisuke Wakabayashi & Shane Goldmacher, *Google Policy Change Upends Online Plans for 2020 Campaigns*, N.Y. TIMES (Nov. 20, 2019), <https://www.nytimes.com/2019/11/20/technology/google-political-ads-targeting.html>. However, advertisers may still target users based on a host of other personal characteristics, such as age, gender, location, searches, visited websites, and watched videos. *See id.*

192. *Fact-Checking on Facebook*, *supra* note 180.

193. *Fact-Checking on Facebook*, *supra* note 180.

194. Foreign Influence Operations’ Use of Social Media Platforms: Hearing Before the S. Select Comm. on Intelligence, 115th Cong. 43 (2018) [hereinafter Sandberg Senate Testimony] (testimony of Sheryl Sandberg, Chief Operating Officer, Facebook).

195. Clegg, *supra* note 176.

196. *Publishing News Feed*, FACEBOOK BUS., <https://www.facebook.com/help/publisher/275495073088478> [<https://perma.cc/3LDH-4S5T>]; *How News Feed Works*,

reported result of this unrelenting drive to maximize user engagement and retention is the rise of social media bubbles that largely echo and reinforce users' political interests and biases.¹⁹⁷ As one commentator observed, "Rarely will our Facebook comfort zones expose us to opposing views," "we eventually become victims to our own biases," and "a critical eye grows less keen the more it is ambushed with one-sided propaganda."¹⁹⁸

Facebook's politically skewed, source-narrowing feeds—what its former vice president for user growth described as "dopamine-driven feedback loops"¹⁹⁹—are all the more concerning considering that nearly half of Americans get news through Facebook.²⁰⁰ While these personalized feeds may keep users coming back for more pleasingly-bespoke content, they do not broaden, balance, or otherwise promote open and informed public discussion. Instead, they divide the electorate into polarized realities based on disparate streams of "alternative facts."²⁰¹

FACEBOOK HELP CTR., https://www.facebook.com/help/1155510281178725/?helpref=hc_fnav [<https://perma.cc/95GF-QJMT>]; see Isaac, *supra* note 143.

197. Isaac, *supra* note 143; see Callum Borchers, *Facebook Invites You to Live in a Bubble Where You Are Always Right*, WASH. POST (Jan. 14, 2019, 8:00 AM), <https://www.washingtonpost.com/news/the-fix/wp/2018/01/14/facebook-invites-you-to-live-in-a-bubble-where-you-are-always-right> (arguing that Facebook's newsfeed "changes seem to make it easier than ever to create filter bubbles that block out opinions that don't match your own"); Alex Hern, *How Social Media Filter Bubbles and Algorithms Influence the Election*, GUARDIAN (May 22, 2017, 9:14 AM), <https://www.theguardian.com/technology/2017/may/22/social-media-election-facebook-filter-bubbles> [<https://perma.cc/4UCS-72MQ>] (discussing the outsized impact Facebook had in the 2018 parliamentary election in the United Kingdom and noting that "more than 60% of Facebook users are entirely unaware of any curation on Facebook at all"); Alexis C. Madrigal, *What Facebook Did to American Democracy*, ATLANTIC (Oct. 12, 2017), <https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502> (arguing that Facebook had a profound impact on the 2016 election and that Facebook's primary "draw is its ability to give you what you want" in terms of news and other media).

198. Mostafa M. El-Bermawy, *Your Filter Bubble Is Destroying Democracy*, WIRED (Nov. 18, 2016, 5:45 AM), <https://www.wired.com/2016/11/filter-bubble-destroying-democracy> [<https://perma.cc/M3RG-D9GW>].

199. See Amy B. Wang, *Former Facebook VP Says Social Media Is Destroying Society with Dopamine-Driven Feedback Loops*, WASH. POST (Dec. 12, 2017, 1:37 PM), <https://www.washingtonpost.com/news/the-switch/wp/2017/12/12/former-facebook-vp-says-social-media-is-destroying-society-with-dopamine-driven-feedback-loops>.

200. See John Gramlich, *10 Facts About Americans and Facebook*, PEW RES. CTR. (May 16, 2019), <https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook> [<https://perma.cc/93CU-D7Z7>].

201. In 2017, Kellyanne Conway, Counsel to President Trump, coined the term "alternative facts" during a television interview to describe what her interviewer rebuked in response as "not facts" but "falsehoods," and the phrase has been derisively

Second, as scrutinized in Part I, the Community Standards do “referee political debates” among its billions of users, censoring an extensive amount of speech that might offend.²⁰² To add examples to those discussed above,²⁰³ none of Facebook’s ordinary users could, hypothetically speaking, characterize Mexicans as “rapists” or African nations as “shithole countries”; describe white supremacists as “very fine people”; call Christians “bigoted” or white people “racist”; or argue that women “should not be allowed to work.” Each of these sentiments would violate at least one do-not-post rule.²⁰⁴ However, posts denouncing such hate speech (e.g., “white supremacists are *not* very fine people”) would survive censorship, even though they would be just as political. Hence, Facebook hardly serves as a politically neutral marketplace of ideas where users can “talk openly about the issues that matter to them, even if some may disagree or find them objectionable,”²⁰⁵ “inform themselves” of what others are saying, and hold others “accountable for their words.”²⁰⁶

Third, at least in theory, Facebook further referees the political speech of ordinary users by labeling and throttling the distribution of content that its fact-checking partners deem false or misleading.²⁰⁷ Setting aside the desirability or effectiveness of this fact-checking policy,²⁰⁸ its existence on top of the feed curation and Community Standards undercuts Zuckerberg’s assertion that the “private company” should not “censor politicians or the news.”²⁰⁹ It censors the latter quite extensively.

used as code for just that. Haan, *supra* note 11, at 18. Facebook Chief Operating Officer Sheryl Sandberg, apparently without irony, touted the company’s efforts to present users with “alternative facts” as part of its fact-checking efforts. *See id.* at 19.

202. *See* Clegg, *supra* note 176; *Sandberg Senate Testimony*, *supra* note 194, at 43.

203. *See supra* note 106 and accompanying text.

204. *See supra* notes 71, 100 and accompanying text. One qualification to the above examples is that Facebook apparently allows otherwise forbidden hate speech for “criticism of immigration policies and arguments for restricting those policies.” *Hate Speech*, *supra* note 90. So, while a non-politician could not generally call Haiti, El Salvador, or African countries “shithole countries,” they could if arguing that the United States should not accept more immigrants from those nations. *Cf.* Ibram X. Kendi, *The Day Shithole Entered the Presidential Lexicon*, ATLANTIC (Jan. 13, 2019), <https://www.theatlantic.com/politics/archive/2019/01/shithole-countries/580054>. This qualification does little to mitigate the amount of censorship of political speech on Facebook beyond licensing hate speech in the immigration context.

205. *Community Standards*, *supra* note 7.

206. *Fact-Checking on Facebook*, *supra* note 180.

207. *See supra* notes 182–83 and accompanying text.

208. *See supra* notes 184–86 and accompanying text.

209. *See Mark Zuckerberg Stands for Voice and Free Expression*, *supra* note 31; *supra* note 149 and accompanying text.

Fourth, Facebook's exemption of politicians from its speech code and fact checking creates a two-tier speech platform that, as one commentator put it, "treats people who aren't politicians as second-class citizens."²¹⁰ Of course, politicians at the national level already enjoy considerable access to news coverage and have an outsized voice on social media by virtue of their position and prominence.²¹¹ Facebook amplifies these powerful advantages by inviting them to speak without restraint *and* selling them access to the targeting tools that can optimize the spread of their speech, including, if they choose, hate or lies.²¹² While an opposing politician may offer a fact-check, if not "alternative facts," Facebook's algorithmic feed works to keep such displeasing counterspeech out of partisan bubbles.

Consequently, Facebook's special solicitude toward politicians promotes a race to the bottom in which willing ones may take the low road of spreading lies that they perceive to be politically advantageous.²¹³ A recent study published in *Science* by researchers at the Massachusetts Institute of Technology generally supports such a perception.²¹⁴ Using a dataset of approximately 126,000 verified true and false news stories spread on Twitter from 2006 to 2017, the study found that "[f]alsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information."²¹⁵ Furthermore, it found that "the

210. Gilad Edelman, *How Facebook Gets the First Amendment Backward*, WIRED (Nov. 7, 2019, 5:28 PM), <https://www.wired.com/story/facebook-first-amendment-backwards> [<https://perma.cc/53L5-KX9L>].

211. See *U.S. Political Leaders Have a Sizable Following Among U.S. Adults on Twitter*, PEW RESEARCH CTR. (July 12, 2019), <https://www.pewresearch.org/fact-tank/2019/07/15/about-one-in-five-adult-twitter-users-in-the-u-s-follow-trump> [<https://perma.cc/DE2M-BR29>] (noting that one-in-five adults on Twitter follow Trump).

212. It is not clear whether the Community Standards would allow non-politicians to express agreement with a hateful sentiment by a politician which would otherwise violate the rules. Facebook does not say one way or another, and either way would raise questions regarding the consistency and coherence of the company's line-drawing.

213. In the 2020 Presidential election cycle, Warren and several other Democratic candidates pledged not to spread disinformation online. See Cat Zakrzewski, *The Technology 202: Several Presidential Candidates Have Yet to Promise Not to Spread Disinformation*, WASH. POST (Jan. 30, 2020), <https://www.washingtonpost.com/news/poweporst/paloma/the-technology-202/2020/01/30/the-technology-202-several-presidential-candidates-have-yet-to-promise-not-to-spread-disinformation/5e31bdbb88e0fa6ea99d6ff61>. So far, Trump has made no such commitment.

214. See Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCIENCE 1146, 1148 (2018); see also Robinson Meyer, *The Grim Conclusions of the Largest-Ever Study of Fake News*, ATLANTIC (Mar. 8, 2018), <https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104>.

215. Vosoughi et al., *supra* note 214, at 1147.

effects were more pronounced for false political news” than other kinds of news, perhaps because of the sensational novelty.²¹⁶

Since Twitter’s timeline serves content based on whom users follow and how users and their social network previously interacted with content,²¹⁷ the study’s findings would apply to similar “social technologies,” such as Facebook’s personalized feed, that “facilitate rapid information sharing” and “can enable the spread of misinformation.”²¹⁸ Indeed, the viral spread of the doctored Pelosi video among millions of users on Facebook and Twitter illustrates the potency of sensational false political news.²¹⁹ In addition, user comments responding to the video on conservative Facebook pages (e.g., mocking Pelosi as “drunk” or “a babbling mess”) illustrate the confirmation bias fostered by algorithms designed to serve users content that, in Zuckerberg’s words, “matches their lived experience” without “getting in the way of an ability to express something that they feel is real.”²²⁰

To be sure, the doctored Pelosi speech fell under Facebook’s fact-checking policy because it was not the “direct speech” of a politician but the creation of an anonymous user.²²¹ However, Facebook only flagged the content as “false” and demoted it in users’ feeds after it had gone viral and drawn national media attention.²²² Zuckerberg himself conceded, in response to criticism from Pelosi and others, that “[i]t took a while for our systems to flag that,” during which time “it got more distribution than our policies should have allowed.”²²³ But not only was Facebook’s fact-checking concededly too-little-too-late. As Professor Sarah Haan has argued, Facebook’s fact checking itself may reinforce lies as truths. In a “post-truth” world, the “binary construction” of a fact check may ingrain a “two-sided view of information that may increase polarization and partisanship, not diffuse it,” by presenting users with “alternative facts” that match their political preferences.²²⁴

216. *Id.* at 1146.

217. *See About Your Twitter Timeline*, TWITTER HELP CTR. <https://help.twitter.com/en/using-twitter/twitter-timeline> [<https://perma.cc/T65A-AC3F>].

218. Vosoughi et al., *supra* note 214, at 1146.

219. *See* Harwell, *supra* note 150.

220. *Id.*; Newton, *supra* note 26.

221. *See supra* note 187.

222. *See* Kurt Wagner, *Facebook CEO: Company Was Too Slow to Respond to Pelosi Deepfake*, BLOOMBERG (June 26, 2019, 6:25 PM), <https://www.bloomberg.com/news/articles/2019-06-26/facebook-ceo-company-was-too-slow-to-respond-to-pelosi-deepfake> [<https://perma.cc/6HU2-EYZ6>].

223. *Id.*

224. *See* Haan, *supra* note 11, at 34–35.

Fifth, a race to the bottom by politicians willing to spread disinformation could spiral down further with the rise of deepfakes. These are highly manipulated media, such as photos, videos, or audio, that could make someone appear to say or do something that they did not say or do (e.g., making a racist remark or participating in a sex video), or distort something that someone said or did to make them seem to say or do something else (e.g., speaking with a drunken slur or touching someone inappropriately).²²⁵ Following controversies over the spread of the doctored Pelosi video and other manipulated media on the platform, in January 2020, Facebook added to the Community Standards a deepfake ban against “[v]ideo”—but not other media—that “has been edited or synthesized” in ways that “would likely mislead an average person to believe that a subject of the video said words that they did not say,” and “is the product of artificial intelligence or machine learning.”²²⁶ Ironically, this peculiarly limited ban would not seem to apply to the Pelosi video, as the manipulation slowed her words but did not insert different ones, and it was made with low-tech video editing software rather than AI or machine learning.²²⁷ Moreover, the ban does not appear to apply to politicians, as nothing in this addition to the Community Standards states or suggests that it overrides the categorical exemptions for politicians that Facebook has so vigorously defended.²²⁸ Thus, Facebook’s extremely narrow deepfake ban invites evasion and leaves politicians free to exploit another powerful tool to deceive the electorate.

225. See Ian Sample, *What Are Deepfakes—And How Can You Spot Them?*, GUARDIAN (Jan. 13, 2020, 5:00 AM), <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them> [<https://perma.cc/G9CZ-LDP9>].

226. *Manipulated Media, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/manipulated_media [<https://perma.cc/3YS2-7MKE>]; see Monika Bickert, *Enforcing Against Manipulated Media*, FACEBOOK (Jan. 6, 2020), <https://about.fb.com/news/2020/01/enforcing-against-manipulated-media> [<https://perma.cc/5XMA-DZQF>].

227. See Tony Romm et al., *Facebook Bans Deepfakes, But New Policy May Not Cover Controversial Pelosi Video*, WASH. POST (Jan. 7, 2020, 3:56 PM), <https://www.washingtonpost.com/technology/2020/01/06/facebook-ban-deepfakes-sources-say-new-policy-may-not-cover-controversial-pelosi-video>. Additionally, the deepfake ban exempts “parody or satire” or manipulations that “omit words that were said or change the order,” which gives users plenty of license to make politicians appear to say something entirely different from what they said. *Manipulated Media*, *supra* note 226.

228. Facebook has not claimed that the ban applies to politicians, but a spokesperson has issued conflicting statements about whether the ban applies to political ads. See Romm, *supra* note 227 (explaining that Facebook has declined to send political ads with manipulated media to third-party fact-checkers or label the ads as false but subsequently claimed that it may do so with respect to “highly manipulated video”).

By contrast, Twitter recently adopted a broader policy that governs “deceptively . . . synthetic and manipulated media . . . [that] are likely to cause harm.”²²⁹ The policy applies to politicians and non-politicians alike and extends to media that has been “significantly and deceptively altered or fabricated” and “shared in a deceptive manner,” *regardless* of the technology used to produce it.²³⁰ Twitter states that it will remove the media if it is “likely to cause harm,” defined to include non-violent harms such as “voter suppression or intimidation.”²³¹ Otherwise, Twitter will label the media as deceptively manipulated, reduce its distribution, and “provide additional context.”²³² Under this deepfake policy, the Pelosi video would likely draw at least the remedial steps for deceptively manipulated media. Thus, as with its complete ban on political ads, Twitter has gone much farther than Facebook in adopting policies to suppress disinformation.

Finally, the privileged class whom Facebook frees from the constraints of its speech code and fact checking, and to whom it sells the keys to its ad targeting kingdom, possesses regulatory power over the company and could wield that authority in ways that threaten its existence. Indeed, some politicians, including Warren and other prominent Democrats, have called for a break-up of Facebook for antitrust, privacy, and electoral integrity reasons.²³³ On the other end of the political spectrum, prominent Republican politicians such as Trump have threatened to regulate the platform to combat alleged anti-conservative bias.²³⁴ Freeing politicians

229. Yoel Roth & Ashita Achuthan, *Building Rules in Public: Our Approach to Synthetic & Manipulated Media*, TWITTER BLOG (Feb. 4, 2020), https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html [<https://perma.cc/JH7M-SG3E>].

230. *Id.*

231. *Id.*

232. *Id.*

233. See Elizabeth Warren, *Here's How We Can Break Up Big Tech*, MEDIUM (Mar. 8, 2019), <https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c> [<https://perma.cc/Y8X3-Y767>]; cf. Marcy Gordon, *US Senator Asks Facebook CEO Mark Zuckerberg to Sell Instagram and WhatsApp*, BUS. INSIDER (Sept. 20, 2019, 5:52 AM), <https://www.businessinsider.com/us-senator-josh-hawley-asks-mark-zuckerberg-break-up-facebook-instagram-whatsapp-2019-9> [<https://perma.cc/7Q3T-EX6N>] (discussing Republican Senator Josh Hawley's suggestion to Zuckerberg to sell off WhatsApp and Instagram). Biden has also advocated making Facebook at least civilly liable for the falsehoods that it posts. See Editorial Board, Opinion, *Joe Biden: Former Vice President of the United States*, N.Y. TIMES (Jan. 17, 2020), <https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html>.

234. Emily Stewart, *Silicon Valley Should Take Josh Hawley's Big War on Big Tech Seriously*, VOX (Oct. 29, 2019, 6:30 AM), <https://www.vox.com/recode/2019/10/29/20932064/senator-josh-hawley-tech-facebook-google-mark-zuckerberg-missouri>

from Facebook's content moderation and fact checking may not appease Democratic critics,²³⁵ but it does mitigate the pressure Facebook faces from powerful politicians such as Trump who want to speak freely (including falsely) on Facebook.²³⁶ Moreover, Trump's reelection campaign is currently the platform's top spender on political ads.²³⁷ Thus, Facebook's exemption of politicians from its speech restrictions and fact checking furthers both its political interests and its business interests.

Of course, it might be risky in the long run to appease some politicians, including the current White House occupant, at the cost of antagonizing others, including possibly the next one. But it makes some near-term political and business sense to adopt policies that align with the political opponent of the person calling for the break-up of the company. Along these lines, it is notable, though perhaps coincidental, that Zuckerberg had two meetings with Trump in 2019, the first a week before Facebook announced its exemptions for the speech of politicians, and the second a month later with Facebook board member and prominent Trump supporter Peter Thiel.²³⁸ Democratic critics panned these meetings as "corruption, plain and simple," and accused Facebook of "potentially tailoring its ad policy in order to appease the Trump administration and the right so they do not get regulated."²³⁹

[<https://perma.cc/LN5R-3QNR>]. As noted, such claims of political bias lack empirical support. See Ingram, *supra* note 164.

235. See *supra* notes 146, 162 and accompanying text.

236. See *supra* note 144 and accompanying text. It is well documented that Trump has lied prolifically as a politician. See Glenn Kessler et al., *President Trump Made 16,241 False or Misleading Claims in His First Three Years*, WASH. POST (Jan. 20, 2020, 3:00 AM), <https://www.washingtonpost.com/politics/2020/01/20/president-trump-made-16241-false-or-misleading-claims-his-first-three-years>.

237. *Facebook Ad Library Report*, FACEBOOK, <https://www.facebook.com/ads/library/report> [<https://perma.cc/U7GK-EFQG>] (providing totals for political ad campaign spending on Facebook from January 1, 2019 through March 14, 2020).

238. See Lizette Chapman & Gerrit De Vynck, *Peter Thiel Pledges to Support Trump's 2020 Campaign*, BLOOMBERG (Nov. 1, 2018, 7:23 PM), <https://www.bloomberg.com/news/articles/2018-11-01/peter-thiel-praises-trump-pledges-to-support-his-2020-campaign> [<https://perma.cc/4E9U-X6UD>]; Eric Lutz, *Mark Zuckerberg Had a Secret Dinner with Trump and Peter Thiel*, VANITY FAIR (Nov. 21, 2019), <https://www.vanityfair.com/news/2019/11/mark-zuckerberg-donald-trump-secret-dinner>; Jordan Novet, *Facebook's Zuckerberg Met with President Trump at the White House*, CNBC (Sept. 19, 2019, 7:27 PM), <https://www.cnn.com/2019/09/19/facebook-ceo-mark-zuckerberg-met-president-trump-at-the-white-house.html> [<https://perma.cc/J8WQ-NSG5>].

239. Cristiano Lima, *Democrats Pound Facebook over Zuckerberg's Secret Dinner with Trump*, POLITICO (Nov. 21, 2019, 10:02 AM), <https://www.politico.com/news/>

CONCLUSION

Though the length of this Article may try even the proverbial Patient Reader,²⁴⁰ it has only scrutinized a small selection of Facebook's sprawling speech restrictions in its Community Standards.²⁴¹ There is much more that free speech scholars can study and should critique about the Community Standards.²⁴² After all, in extensively restricting the speech of billions worldwide and hundreds of millions in the United States, Facebook's speech code is the most far-reaching and among the most consequential in history.

2019/11/21/elizabeth-warren-zuckerbergs-secret-dinner-with-trump-072410
[<https://perma.cc/6EZT-7HTJ>].

240. Cf. Samuel Johnson, *A Study of Milton's Paradise Lost*, in 1 LIVES OF THE ENGLISH POETS, at 183 (George Birkbeck Hill ed., 1905) (writing of Milton's epic poem, "None ever wished it longer than it is").

241. In addition, this Article has not discussed the outside board that Facebook has proposed to resolve content disputes, which Zuckerberg likened to "a Supreme Court" for Facebook. Ezra Klein, *Mark Zuckerberg on Facebook's Hardest Year, and What Comes Next*, VOX (Apr. 2, 2018, 6:00 AM), <https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge> [<https://perma.cc/UT3A-TKFP>]. As currently conceived, this external board will be comprised of "independent experts" in many "relevant disciplines," from "content" and "civil rights" to "safety" and (thank goodness) "free expression." FACEBOOK, DRAFT CHARTER: AN OVERSIGHT BOARD FOR CONTENT DECISIONS 1, <https://about.fb.com/wp-content/uploads/2019/01/draft-charter-oversight-board-for-content-decisions-2-1.pdf> [<https://perma.cc/2TKS-B6U2>]. Its primary responsibility will be to "review Facebook's most challenging content decisions" in "enforcing our Community Standards," and it will have the power to reverse them. *Id.* at 3. However, it is questionable whether the board can be truly independent of Facebook. For one, its initial members will be selected by Facebook, and though the board will select future members, Facebook can continue to nominate candidates. *See id.* at 2. Furthermore, funding for the board's operations will come from Facebook, and there is no guarantee that Facebook will fund or even maintain the board in the long term. *See id.* at 4. More fundamentally, the board will only interpret and apply the Community Standards that Facebook has promulgated and alone has power to revise—including in response to board decisions. Thus, as the top executive and controlling shareholder of Facebook, Zuckerberg retains the ultimate power to dictate what is or is not said on Facebook, as well as whether any speakers get favorable treatment. This is a far cry from an independent supreme court in a constitutional democracy that enforces fundamental rights, including robust free speech and due process rights, that cannot be rewritten by the government. *See* Daphne Keller, *Facebook Restricts Speech by Popular Demand*, ATLANTIC (Sept. 22, 2019), <https://www.theatlantic.com/ideas/archive/2019/09/facebook-restricts-free-speech-popular-demand/598462>. For these and other reasons, it is highly unlikely that the proposed board can ameliorate the main problems with Facebook's speech code and policies, except perhaps to mitigate some of their vagueness.

242. Cf. Andrew Marvell, *To His Coy Mistress*, in MISCELLANEOUS POEMS 19 (1681) ("Had we but World enough, and Time . . .").

To be clear, this Article is not advocating that Facebook stop operating as a neutral platform that refuses to censor political speech and lets users decide for themselves what is true or false. Facebook cannot stop being what it is not. In addition to censoring a wide range of political speech through its Community Standards, Facebook effectively censors the content that users see through leveraging machine learning to narrow their feeds to match their political interests and biases rather than to broaden or balance their exposure. While Facebook might thereby maximize user satisfaction and retention, it also facilitates a ruinous race to the bottom in which potent disinformation by politicians spread virally and gain credibility among millions of primed supporters in polarized social media bubbles. Facebook reaps the gains, but our democracy reaps the costs.

It is difficult to imagine a speech platform more optimized to distort democratic discussion and divide the electorate than Facebook. It is not too early—and, hopefully, it is not too late—to scrutinize Facebook's speech platform as it is, not as the company passes it off to be.