UCLA School of Law

Public Law & Legal Theory Research Paper No. 23-11

APPLYING MILITANT DEMOCRACY TO DEFEND AGAINST SOCIAL MEDIA HARMS

BY

NEIL NETANEL

PETE KAMERON PROFESSOR OF LAW

---

48 Cardozo Law Review (forthcoming 2023)

APPLYING MILITANT DEMOCRACY TO DEFEND AGAINST SOCIAL MEDIA HARMS

Neil Netanel†

† Pete Kameron Professor of Law, UCLA School of Law. For their helpful comments, I thank Jack Balkin, Dave Fagundes, Richard Hasen, Claudia Haupt, Aziz Huq, Shalev Netanel, João Pedro Quintais, Lea Rabe, Pamela Samuelson, Alexander Tsesis, and participants in the Yale Information Society Project Conference on Deceptive Technologies, UCLA School of Law faculty summer workshop, Tel Aviv University Law and Technology workshop, and Tel Aviv University Shamgar Center for Digital Law and Innovation Conference on Freedom of Expression and Digital Platforms. I also thank Rachel Green, Max Ellis, and Arkadi de Proft for their research assistance.

INTRODUCTION

Social media inflict multiple harms on liberal democracy. Online platforms thrive on propagating emotionally inflammatory content that maximizes user engagement.[1] Too often that entails amplifying disinformation, hate speech, online extremism, and deep-seated partisan animosity. Tellingly, as documented in testimony before the House Select Committee to Investigate the January 6th Attack, in the weeks following the 2020 presidential election, Facebook, Twitter, YouTube, and Reddit knowingly enabled a firestorm of vitriolic far-right election denial on

[1] *See* SAMUEL WOOLLEY, MANUFACTURING CONSENSUS: UNDERSTANDING PROPAGANDA IN THE ERA OF AUTOMATION AND ANONYMITY 120–22 (2023); *see also* Steve Rathje, Jay J. Van Bavel & Sander van der Linden, *Out-Group Animosity Drives Engagement on Social Media*, 118 PNAS, June 23, 2021, at 1, 1, https://doi.org/10.1073/pnas.2024292118 [https://perma.cc/XXP9-YP6Y]; Keach Hagey & Jeff Horwitz, *The Facebook Files: Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead.*, WALL ST. J. (Sept. 15, 2021, 9:26 AM), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [https://perma.cc/N6DU-2SX9].

their platforms.[2] In so doing, a Select Committee staff report concludes, the platforms "helped to facilitate the attack on January 6th."[3]

More broadly, social media undermine the critical linchpins of democracy. While scholars debate exactly how and to what extent,[4] the most comprehensive social science literature review to date concludes that social media are a significant factor in emergent autocratic populism, dwindling political and social trust, and growing polarization in established democracies.[5] Another meta-analysis finds that social media use generally hampers gaining the political knowledge that is critical for effective democracy.[6] Among other factors, social media foster a

2 *See* H. SELECT COMM. TO INVESTIGATE THE JAN. 6TH ATTACK, 117TH CONG., SOCIAL MEDIA & THE JANUARY 6TH ATTACK ON THE U.S. CAPITOL: SUMMARY OF INVESTIGATIVE FINDINGS 8-10 (Draft 2022), https://techpolicy.press/wp-content/uploads/2023/01/J6-Committee-Draft-Social-Media-Report-TPP.pdf [https://perma.cc/Y3X9-KDJZ] [hereinafter DRAFT REPORT]; *see also* Cat Zakrzewski, Cristiano Lima & Drew Harwell, *What the Jan. 6 Probe Found Out About Social Media, but Didn't Report*, WASH. POST (Jan. 17, 2023, 5:43 PM), https://www.washingtonpost.com/technology/2023/01/17/jan6-committee-report-social-media [https://perma.cc/P2MT-HZQV] (summarizing investigators' 122-page memo and opining that it was not included in the Committee's final report because Committee members wished to focus on Donald Trump's role and avoid raising technology companies' ire).

3 DRAFT REPORT, *supra* note 2, at 8.

4 *See, e.g.*, Jonathan Haidt & Chris Bail, Social Media and Political Dysfunction: A Collaborative Review (Aug. 28, 2023) (unpublished manuscript), https://docs.google.com/document/d/1vVAtMCQnz8WVxtSNQev_e1cGmY9rnY96ecYuAj6C548/edit [https://perma.cc/R6LB-YW54] (collecting citations, links, and abstracts of published scholarly articles addressing various facets of the question: "Is social media a major contributor to the rise of political dysfunction seen in the USA and some other democracies since the early 2010s?"). *Compare* Jonathan Haidt, *Yes, Social Media Really Is Undermining Democracy*, THE ATLANTIC (July 28, 2022) https://www.theatlantic.com/ideas/archive/2022/07/social-media-harm-facebook-meta-response/670975 [https://perma.cc/9PKY-CJQB] (concluding on the basis of collected studies that social media is a likely causal factor in growing affective polarization (i.e., partisan animosity), information homophily, and echo chambers), *and* Ludovic Terren & Rosa Borge, *Echo Chambers on Social Media: A Systematic Review of the Literature*, 9 REV. COMMC'N RSCH. 99 (2021) (reviewing fifty-five studies and noting that only five studies found no evidence of echo chambers on social media, close to half found clear evidence of echo chambers, and some found echo chambers around political topics and controversial issues but not other issues), *with* Gideon Lewis-Kraus, *How Harmful is Social Media?*, NEW YORKER (June 3, 2022), https://www.newyorker.com/culture/annals-of-inquiry/we-know-less-about-social-media-than-we-think [https://perma.cc/JP4Q-G7UU] (contending that there is as yet inadequate evidence to support arguments that social media causes echo chambers, widespread susceptibility to disinformation, and increased radicalization due to personalized feed and recommendations).

5 Philipp Lorenz-Spreen, Lisa Oswald, Stephan Lewandowsky & Ralph Hertwig, *A Systematic Review of Worldwide Causal and Correlational Evidence on Digital Media and Democracy*, 7 NATURE HUM. BEHAV. 74 (2023) (examining digital media, including websites and general internet access, as well as social media platforms).

6 *See* Eran Amsalem & Alon Zoizner, *Do People Learn About Politics on Social Media? A Meta-Analysis of 76 Studies*, 73 J. COMMC'N 3 (2022). The Lorenz-Spreen et al. literature review, *supra* note 5, concludes that digital media consumption overall likely increases political knowledge. But that finding includes reading traditional news media websites—and studies show

misperception that the "news finds me," "that all the news I need to know will appear in my feed."[7]

Relatedly, social media have corrosive effects on democratic institutions. Democratic government cannot function without broadly accepted, legitimate political authority, some basic consensus regarding how to distinguish truth from falsity, and a sense that even ardent political opponents are part of the same polity, bound by a common fate.[8] Yet online platforms radically undermine those pillars, challenging democratic political authority, fueling the disintegration of traditional and stable political parties, empowering free agent politicians who are not beholden to party leadership, heightening partisan animosity, and rendering effective government based on compromise exceedingly difficult.[9] As election law scholar Richard Hasen notes: "Rather than improving our politics, cheap speech [through social media] makes

that reading traditional news, particularly public service news media, increases political knowledge. Lorenz-Spreen, Oswald, Lewandowsky & Hertwig, *supra* note 5, at 78; Toril Aalberg, *Does Public Media Enhance Citizen Knowledge*?, *in* THE DEATH OF KNOWLEDGE? (Aeron Davis ed., 2019) (finding that public service media enhances political knowledge to a greater extent than does commercial media). The findings that social media stifle political knowledge, relative to traditional news media, are especially worrisome given that social media have rapidly become a dominant source of news consumption. According to a recent Pew Research Center study, roughly half of Americans get news on social media "often" or "sometimes." Mason Walker & Katerina Eva Matsa, *News Consumption Across Social Media in 2021*, PEW RSCH. CTR. (Sept. 20, 2021), https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021 [https://perma.cc/K78S-5C4J]. On the importance of a well-informed public for effective democracy, see HENRY MILNER, CIVIC LITERACY: HOW INFORMED CITIZENS MAKE DEMOCRACY WORK (2002); and Ilya Somin, *Voter Ignorance and the Democratic Ideal*, 12 CRITICAL REV. 413 (1998).

7 *See* Chang Sup Park, *Reading a Snippet on a News Aggregator vs. Clicking Through the Full Story: Roles of Perceived News Importance, News Efficacy, and News-Finds-Me Perception*, 23 JOURNALISM STUD. 1350, 1357–58, 1369 (2022). See also Homero Gil de Zúñiga & Zichen Cheng, *Origin and Evolution of the News Finds Me Perception: Review of Theory and Effects*, 30 PROFESSIONAL DE LA INFORMACIÓN, May 2021, at 1, 1; Nadine Strauß, Brigitte Huber & Homero Gil de Zúñiga, *Structural Influences on the News Finds Me Perception: Why People Believe They Don't Have to Actively Seek News Anymore*, 7 SOC. MEDIA + SOC'Y, Apr.–June 2021, at 1. Researchers have identified a number of possible additional reasons for social media's stultifying effect, including social media recommender systems' propensity to limit exposure to diverse information, the prevalence of misinformation on social media, and a learning-impeding feeling of overload arising from the surfeit of information on social media. Amsalem & Zoizner, *supra* note 6.

8 *See* Robert Post, *The Unfortunate Consequences of a Misguided Free Speech Principle*, DAEDALUS (forthcoming 2023) (manuscript at 17), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4255938 [https://perma.cc/U4TG-PFTN] ("Politics is possible only when diverse persons agree to be bound by a common fate. Lacking that fundamental commitment, politics can easily slide into an existential struggle for survival that is the equivalent of war." (footnote omitted)).

9 *See* Richard H. Pildes, *Democracies in the Age of Fragmentation*, 110 CALIF. L. REV. 2051, 2052-56, 2059-68 (2022).

political parties increasingly irrelevant by allowing demagogues to appeal directly and repeatedly at virtually no cost to voters for financial and electoral support, with incendiary appeals and often with lies."[10] As Hasen aptly concludes, the greatest danger facing democracy today "is a public that cannot determine truth or make voting decisions that are based on accurate information, and a public susceptible to political manipulation through repeatedly amplified, data-targeted, election-related content, some it false or misleading."[11]

The current state of affairs is untenable. Surely, it is incumbent upon the democratic state to combat social media's palpable threats to democracy, even while carving out space for the diversity of voice and civic engagement that social media can offer.[12] In that vein, recent years have seen a number of federal and state legislative proposals designed, among other things, to impose regulatory oversight over digital platforms, require transparency in content moderation, promote civic discourse, defend election integrity by prohibiting social media from carrying micro-targeted political ads and banning the use of bots in political advertising, and holding social media companies accountable for targeted harassment, terrorist recruiting, and violations of federal civil rights laws on their platforms.[13] Yet, as this Article enumerates, the neoliberal techno-utopianism and First Amendment jurisprudence that dominate American law, policy, and political thought have presented nigh-insurmountable obstacles to any serious consideration of these and other proposed initiatives designed to combat social media's multivalent harms to American democracy.[14]

The political theory and practice of "militant democracy," I argue, provides a superior policy framework for defending democratic institutions against social media harms.[15] Militant democracy arose as a

10 RICHARD L. HASEN, CHEAP SPEECH: HOW DISINFORMATION POISONS OUR POLITICS—AND HOW TO CURE IT 21 (2022).

11 *Id.* at 24.

12 As Renée DiResta observes, social media have helped to shunt aside traditional elites' top-down control of political narratives. In so doing, they have enabled the much-lauded flourishing of new, previously silenced voices and communities. But, in so doing, social media have also fueled the ominous power of "whoever manages to wield the affordances of social networks most adeptly to solidify online factions and command public attention." Renée DiResta, *Algorithms, Affordances, and Agency*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY 121, 128 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022).

13 *See infra* notes 158–160 and accompanying text.

14 On neoliberal technology utopianism, see Paul Starr, *How Neoliberal Policy Shaped the Internet—and What to Do About It Now*, AM. PROSPECT (Oct. 2, 2019), https://prospect.org/power/how-neoliberal-policy-shaped-internet-surveillance-monopoly [https://perma.cc/N5XU-44NF].

15 In similar vein, Aziz Z. Huq draws on the scholarship regarding militant democracy for broad conceptual insights, unbounded "by the familiar intellectual orthodoxies of the First Amendment,"

central feature of post-war constitutionalism in Europe. It views democracy as inherently precarious, at risk of being undone at the hands of antidemocratic forces that, like Nazism in the Weimar Republic, exploit democratic freedoms to undermine democracy. Democracies, it posits, must resolutely defend themselves against avowedly antidemocratic forces and the use of manipulative propaganda to prey upon democracies' weaknesses. No less importantly, democratic states must actively foster the basic political trust, social cohesion, equality of voice, and respect for diversity upon which enduring liberal democratic governance depends.[16] At bottom, militant democracy counsels that enduring liberal democracy must rest on some approximation of the ideal, Habermasian public sphere in which citizens exercise collective self-determination through a discursive exchange of informed, reason-based views among equal participants, free of coercion, manipulative propaganda, and the undue influence of wealth and power.[17]

Militant democracy encompasses a range of strategies designed to underwrite a robust, enduring liberal democracy. Most basically, European countries outlaw antidemocratic political parties, private militias, and terrorist incitement that pose a palpable threat to democratic governance.[18] Concomitantly, they bolster inclusive, egalitarian participation in public discourse and protect minorities against effective disenfranchisement by forbidding group libel, hate speech, and Holocaust denial.[19] To promote trustworthy information and fact-based democratic debate, they also generously fund independent public service media and sharply restrict potentially manipulative political advertising.[20]

---

to meet the challenge that digital platforms pose to contemporary democracy. *See* Aziz Z. Huq, *Militant Democracy Comes to the Metaverse?*, 72 EMORY L.J. 1105, 1112 (2023).

16 *See infra* notes 266—277 and accompanying text.

17 *See* Emilie Prattico, *Introduction*, *in* HABERMAS AND THE CRISIS OF DEMOCRACY: INTERVIEWS WITH LEADING THINKERS 1, 16–26, 33–40 (Emilie Prattico ed., 2022) (describing Habermas's vision of an inclusive, reason-based public sphere as the linchpin of democracy); *see also* Peter Stone, *Democratic Equality and Militant Democracy*, *in* MILITANT DEMOCRACY AND ITS CRITICS: POPULISM, PARTIES, EXTREMISM 38, 45–50 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2021) [hereinafter MILITANT DEMOCRACY AND ITS CRITICS] (maintaining that democracies must provide mechanisms for the electorate to engage in reasoned deliberation about public policy and in electing representatives).

18 *See infra* notes 280–283 and accompanying text; *see also* Samuel Issacharoff, *Fragile Democracies*, 120 HARV. L. REV. 1405, 1421–51 (2007) (presenting a typology of methods that democracies in Europe and elsewhere use to suppress antidemocratic political mobilizations).

19 *See* ALEXANDER TSESIS, DESTRUCTIVE MESSAGES: HOW HATE SPEECH PAVES THE WAY FOR HARMFUL SOCIAL MOVEMENTS 180–92 (2002) (surveying laws penalizing hate speech in several countries); Jeffrey W. Howard, *Free Speech and Hate Speech*, 22 ANN. REV. POL. SCI. 93, 94 (2019)(observing that incitement of racial or religious hatred is illegal in the preponderance of developed democracies, both in Europe and elsewhere).

20 *See infra* notes 285–291 and accompanying text.

Unlike the neoliberal and classical liberal models that have long dominated First Amendment jurisprudence, the constitutional principle of militant democracy rejects the notion that the only available remedy for antidemocratic speech, hate speech, and manipulative propaganda is more speech.[21] It recognizes, rather, that limits on speech may sometimes be critical to defending democracy. Speech restrictions must be proportionate and narrowly targeted as needed to defend against palpable threats to democracy, taking account of identifiable vulnerabilities in existing democratic institutions. From the militant democracy perspective, however, it is not a valid criticism of a measure that aims to counter serious harms to civic discourse merely to say that the measure constrains speech or, for that matter, constrains the amplification of speech through social media. As we shall see, such strictures find support in international human rights jurisprudence as well as national constitutional law.[22]

Social media have become a primary platform for authoritarian propaganda and ethno-nationalist extremism. Yet, social media generally threaten democracy in ways that are more diffuse than the antidemocratic political movements that were the traditional, core concerns of militant democracy. I argue in this Article that militant democracy nevertheless provides a fruitful conceptual framework for countering the threats that social media pose to enduring democratic governance.

Most basically, a militant democracy framework would support democratic countries' initiatives to induce online platforms to cease the rampant propagation of the types of speech that legacy news media committed to basic journalistic norms of fairness and accuracy would not publish—hate speech, incitement to violence, disinformation, conspiracy theories, hyped-up partisan vitriol, and coordinated personal attacks designed to silence victims through intimidation. Importantly, however, militant democracy principles would not merely target social media content curation and moderation practices that amplify harmful speech. They would also require online platforms affirmatively to give prominence to trustworthy information and fact-based democratic debate. As such, militant democracy stands in sharp contrast to dominant schools

---

21 American scholars have also questioned whether the assumption, often repeated in First Amendment jurisprudence, that counterspeech is the remedy for falsity and extremism holds water in the age of social media, if it ever did before. *See, e.g.*, Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J. L. & TECH. 387, 406–07, 418–19 (2020); Philip M. Napoli, *What if More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMMC'NS L.J. 55 (2018); Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547 (2018).

22 *See infra* notes 305–313 and accompanying text.

of First Amendment jurisprudence that impose significant barriers to government intervention,

That broad militant democracy approach to countering social media harms finds expression in the European Commission's 2020 European Democracy Action Plan.[23] The Action Plan encompasses several regulatory initiatives for bolstering democratic institutions in the face of authoritarian populists' and foreign operatives' exploitation of online platforms. This Article summarizes the Action Plan, with particular focus on the EU Digital Services Act (DSA),[24] adopted in October 2022, and related measures designed to target social media's subversion of democracy while minimizing restrictions on users' freedom of expression.[25] As we shall see, the DSA and its related measures present a potentially far-reaching, useful model for combating social media harms to democracy.

Importantly, given the much touted "Brussels Effect," the European initiatives will likely inform social media practice not just in Europe but also in the United States. Indeed, that trans-Atlantic influence will likely be felt even if neoliberal technology policy and First Amendment strictures prevent U.S. regulators from acting to counter social media harms to democracy.[26] For an online platform to build and maintain

---

23 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European Democracy Action Plan, COM (2020) 790 final (March 12, 2020), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A790%3AFIN&qid=1607079662423 [https://perma.cc/QR6T-KVVB][hereinafter Democracy Action Plan].

24 Council Regulation 2022/2065, of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1 [hereinafter DSA].

25 While not the focus of this Article, the Democracy Action Plan also includes the proposed European Media Freedom Act and related European Commission Recommendations, which aim to protect commercial news media's editorial independence and pluralism, primarily by increasing transparency of media ownership and promoting standards of journalistic ethics that empower journalists to resist undue political and commercial pressure. *See Proposal for a Regulation of the European Parliament and of the Council Establishing a Common Framework for Media Services in the Internal Market (European Media Freedom Act) and Amending Directive 2010/13/EU*, COM (2022) 457 final (Sept. 16, 2022) [hereinafter Proposed Media Freedom Act]; Commission Recommendation 2022/1634 of 16 September 2022 on Internal Safeguards for Editorial Independence and Ownership Transparency in the Media Sector, 2022 O.J. (L 245) 56.

26 *See* Anu Bradford, *Hey, US Tech: Here Comes the Brussels Effect*, COLUM. BUS. SCH. INSIGHTS (Dec. 16, 2020), https://leading.business.columbia.edu/media-tech/chazen-global-insights/hey-us-tech-here-comes-brussels-effect **[**https://perma.cc/8JUF-GC48**]** (noting that "thanks to their global reach, EU regulations have brought significant benefits to American Internet users, many of whom welcome enhanced privacy protections and less rampant online hate speech"); Charis Papaevangelou, *Digital Services Act, Brussels Effect and the Future of the Internet*, JOLT (Dec. 8, 2020), http://joltetn.eu/digital-services-act-brussels-effect-and-the-future-of-the-internet **[**https://perma.cc/25LJ-38KS]. *See generally* ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD (2020).

different systems and sets of rules in different countries is difficult, expensive, inefficient, and, given communication between users from across the globe, a significant logistical challenge.[27] Hence, even for major platforms, "global compliance is easier and causes less legal or political grief." In that connection, the European initiatives that we will examine are particularly likely to impact social media practice in the United States given the substantial size of the EU social media market—larger than that of the United States—and the penchant of European courts to order extraterritorial compliance.[28]

My argument proceeds in five Parts. Part I details the principal harms that social media inflict on democratic governance. Part II chronicles the neoliberal technology utopianism and First Amendment models that have dominated U.S. policy toward social media. It also highlights the inadequacy of that approach to meet the serious threats to democracy that social media pose, and it briefly considers how militant democracy might serve as a regulatory ideal for a reimagined First Amendment doctrine. Part III recounts the origins of militant democracy and traces its expression in European and international human rights law. Part IV expounds upon how militant democracy can and should be applied to meet the threats posed by social media, while generally supporting informed citizens' robust exchange of views in the digital public sphere. In so doing, I critically assess the recent European regulatory initiatives, with a focus on the DSA. In conjunction with other facets of the European Democracy Action Plan, the DSA requires large online platforms to (1) remove illegal antidemocratic speech; (2) assess, report, and mitigate systemic risks arising from the design, function, or use of their platforms to civic discourse, electoral processes, or the exercise of fundamental rights; and (3) account for the fundamental rights of users and others impacted by platform content moderation. I examine each of those requirements in turn. Part V examines the European constitutional framework for militant democracy regulation of social media. The Article then concludes.

---

[27] *See* Daphne Keller, *Who Do You Sue? State and Platform Hybrid Power over Online Speech* 8 (Hoover Inst., Aegis Series Paper No. 1902, 2019), https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf [https://perma.cc/3TQG-V2GT].

[28] *Id.* at 8. On social media use in Europe, the United States, and elsewhere, see Daniel Ruby, *Social Media Users 2023 (Global Demographics)*, DEMANDSAGE, (July 26, 2023), https://www.demandsage.com/social-media-users/#:~:text=USA%2DSpecific%20Social%20Media%20Statistics&text=The%20USA%20has%20302.35%20million,74.2%25%20of%20adults%20using%20it [https://perma.cc/TSG6-2UT8].

I. HOW SOCIAL MEDIA UNDERMINE DEMOCRACY

In recent years, a chorus of voices proclaim that something has gone profoundly wrong with democracy in the United States and other developed countries.[29] With good reason. Advanced economy democracies face a tidal wave of illiberal populism, virulent political polarization, epistemic crisis, government paralysis, and popular mistrust in traditional institutions of democratic governance, including legislatures, political parties, and the press. According to a recent study, more than two-thirds of Americans perceive "a serious threat to our democracy" and nearly half believe that "in the next few years, there will be civil war in the United States."[30] At the same time, while 88.8% believe that it is very or extremely important "for the United States to remain a democracy," over 40% agree that "having a strong leader for America is more important than having a democracy."[31] More than one in five agree with the QAnon myth that U.S. institutions are "controlled by a group of Satan-worshipping pedophiles" who traffic children for sex; one in six believe that political violence is justified to "save our American way of life," which is "disappearing"; and nearly a third endorse the statement that "the 2020 election was stolen from Donald Trump, and Joe Biden is an illegitimate president."[32]

There are, no doubt, deep structural causes for our crisis in democracy. They include the abject failure of neoliberalism and globalization to bring growth and broad prosperity to developed countries. Indeed, deregulation, cuts in government services, the acceleration of the knowledge economy, and massive wealth transfers from the middle and working classes of the West to China and other rising developing countries over the last couple of decades have fueled a dramatic increase in inequality and growing economic insecurity in advanced-economy democracies.[33] In tandem, mobility and resulting

29 *See, e.g.*, ANNE APPLEBAUM, TWILIGHT OF DEMOCRACY: THE SEDUCTIVE LURE OF AUTHORITARIANISM (2020); RICHARD L. HASEN, ELECTION MELTDOWN: DIRTY TRICKS, DISTRUST, AND THE THREAT TO AMERICAN DEMOCRACY (2020); YASCHA MOUNK, THE PEOPLE VS. DEMOCRACY: WHY OUR FREEDOM IS IN DANGER AND HOW TO SAVE IT (2018); STEVEN LEVITSKY & DANIEL ZIBLATT, HOW DEMOCRACIES DIE (2018).

30 Garen J. Wintemute et al., Views of American Democracy and Society and Support for Political Violence: First Report from a Nationwide Population-Representative Survey 11–12 (Jul. 15, 2022) (unpublished manuscript), https://doi.org/10.1101/2022.07.15.22277693 [https://perma.cc/C4WY-RGA5].

31 *Id.* at 11.

32 *Id.* at 12–13.

33 Yochai Benkler, *Cautionary Notes on Disinformation and the Origins of Distrust*, MEDIA WELL (Oct. 22, 2019), https://mediawell.ssrc.org/expert-reflections/cautionary-notes-on-disinformation-benkler [https://perma.cc/TG3S-KMBJ]**;** Pildes, *supra* note 9, at 2057–59. Rising

demographic changes in previously homogeneous communities have fueled a turn to ethno-nationalist authoritarianism among longtime residents who perceive a profound threat to their status and way of life.[34] As a result, traditional political alliances have fractured. Relatively affluent, college-educated voters increasingly join with racial minorities to support left-wing parties, while white working-class voters increasingly turn to right-wing, anti-immigrant populist parties. Outside the United States, various electorates have also come to support small fringe parties, ranging from neofascist to self-styled "anti-parties."[35]

Hate speech, conspiracy theories, violent incitement, and pumped-up outrage on social media are expressions of widespread disillusionment with traditional liberal democratic values and institutions. But they also fuel that disillusionment and further destabilize democratic governance. As noted above, surveys of the scientific literature conclude that social media contribute significantly to emergent authoritarian populism, declining political and social trust, growing polarization, and ignorance about the pressing issues of the day.[36] To say the least, as one meta-literature review observes: "One need not share Habermas' conception of 'deliberate democracy' to see that current platforms fail to produce an information ecosystem that empowers citizens to make political choices that are as rationally motivated as possible."[37] Indeed, whatever the laudable opportunities social media offer for giving voice to marginalized communities, social media's overall corrosive effects amount to "clear evidence of serious threats to democracy."[38]

To some degree, those threats are inherent in any online platform for users' relatively uninhibited expression. As netizens have long lamented, unmoderated online discussion that is open to the public at large almost inevitably degenerates into trolling, bullying, and flame wars.[39] Scholars advance various hypotheses for why this is so. Theories center on an "online disinhibition effect" applicable to human psychology generally;[40]

economic inequality is part of a long-term trend. *See* THOMAS PIKETTY, CAPITAL IN THE TWENTY-FIRST CENTURY (Arthur Goldhammer trans., 2014).

34 Michael H. Keller & David D. Kirkpatrick, *Their America Is Vanishing. Like Trump, They Insist They Were Cheated*, N.Y. TIMES (Oct. 27, 2022), https://www.nytimes.com/2022/10/23/us/politics/republican-election-objectors-demographics.html **[**https://perma.cc/6VC9-AEVT**]** **(**reporting that congressional districts represented by Trumpist election deniers tend to be low-income, low-education areas where whites have recently lost their majority status).

35 Pildes, *supra* note 9, at 2053–55.

36 *See supra* notes 5–7.

37 Lorenz-Spreen, Oswald, Lewandowsky & Hertwig, *supra* note 5, at 85.

38 *Id.* at 83.

39 *See* FLAME WARS: THE DISCOURSE OF CYBERCULTURE (Mark Dery ed., 1994).

40 *See* John Suler, *The Online Disinhibition Effect*, 7 CYBERPSYCHOLOGY & BEHAV. 321 (2004).

uniquely hostile individuals' intensive and highly visible participation in online debate;[41] and affective polarization——the proclivity of global peer communication to funnel multiple "identities, beliefs, and cultural preferences . . . into an all-encompassing societal division," in contrast to local, face-to-face interaction in which neighbors often find some common ground on shared interests even if they virulently disagree on many issues.[42]

Whatever the reason, unmediated online networks readily splinter into a sea of uncivil, manipulative free-for-all zones. That cacophony is highly unlikely to secure democratic self-government or lead to greater understanding of our social or physical world. If social media are to have any value for furthering those vital objectives, they must act as public-regarding institutions governed by something akin to professional journalistic norms for curating information, producing knowledge, and providing a space where disparate views can constructively converge.[43] They must build on the example of the world's leading newspapers, which aim to enable a reasoned exchange of views and information on their websites and thus typically moderate online user comments, deleting or blocking hate speech, defamation, political propaganda, personal vitriol, abusive attacks, and spam.[44]

However, social media algorithmic content curation and recommender systems have just the opposite effect. As Shoshana Zuboff has detailed, social media firms have embraced a "surveillance capitalism" business model to meet the market imperative of global growth.[45] They fuel emotionally appealing—and manipulative—content designed to maximize user engagement, amass vast quantities of user data, and enable targeted behavioral advertising. In so doing, they amplify anger, perceived threat, epistemic uncertainty, disinformation,

41 *See* Alexander Bor & Michael Bang Petersen, *The Psychology of Online Political Hostility: A Comprehensive Cross-National Test of the Mismatch Hypothesis*, 116 AM. POL. SCI. REV. 1 (2022); *see also* CHRIS BAIL, BREAKING THE SOCIAL MEDIA PRISM: HOW TO MAKE OUR PLATFORMS LESS POLARIZING 58–61 (2021) (providing examples of a combination of the online disinhibition effect and unique hostility—or at least delight in causing chaos—in which online trolls are lonely, marginalized people who express frustration and find community online).

42 *See* Petter Törnberg, *How Digital Media Drive Affective Polarization Through Partisan Sorting*, 119 PNAS, Oct. 18, 2022, at 1.

43 *See* Jack M. Balkin, *To Reform Social Media, Reform Informational Capitalism*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY, *supra* note 12, at 240–43 (arguing that a healthy public sphere requires such public-regarding institutions but that large social media companies regularly fail to act in the public interest).

44 *See* EMMA GOODMAN, WORLD EDS. F., ONLINE COMMENT MODERATION: EMERGING BEST PRACTICES 14, 61 (2013) https://netino.fr/wp-content/uploads/2013/10/WAN-IFRA_Online_Commenting.pdf [https://perma.cc/G7ZP-LP4Y].

45 SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

and lack of trust. As computational social science scholar Chris Bail puts it: "The social media prism fuels status-seeking extremists, mutes moderates who think there is little to be gained by discussing politics on social media, and leaves most of us with profound misgivings about those on the other side. . . . ."[46]

Social media platforms do engage in extensive content moderation. Certainly, they attempt to extinguish public relations "fires," blocking pornography and removing the most egregious instances of terrorist recruiting and graphic glorification of violence that would raise the ire of advertisers, government regulators, and many users.[47] But that whack-a-mole strategy, aimed at preventing the worst of the worst, fails to address systemic harms arising from the platforms' fundamental design for promoting and distributing information.[48] The strategy is also applied inconsistently. As noted above, Facebook, Twitter, YouTube, and Reddit knowingly served as havens for election-denial right-wing extremism and calls for violence in the weeks before the January 6th riot.[49] A September 2022 NYU study warns, indeed, that, given social media's flawed policies and inconsistent enforcement, the platforms continue to serve as ready vehicles for spewing election denialism, including threats of violence against local election officials.[50]

Social media's failure to implement content moderation at a level and consistency that would prevent the propagation of toxic vitriol and conspiracy theories is partly due to the vast scale, complexity, and unprecedented speed at which social media must operate.[51] But social

46 BAIL, *supra* note 42, at 10.

47 *See* Alex Barker & Hannah Murphy, *Advertisers Strike Deal with Facebook and YouTube on Harmful Content*, FIN. TIMES (Sept. 23, 2020), https://www.ft.com/content/d7957f86-760b-468b-88ec-aead6a558902 [https://perma.cc/LA7B-7K54] (reporting on agreement between World Federation of Advertisers and leading social media platforms requiring common reporting standards for social media regarding harmful content and greater advertiser control over where ads are placed).

48 *See* Karen Hao, *Troll Farms Reached 140 Million Americans a Month on Facebook Before 2020 Election, Internal Report Shows*, MIT TECH. REV. (Sept. 16, 2021), https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election [https://perma.cc/GGN6-23GL] (reporting on an internal Facebook report).

49 *See* DRAFT REPORT, *supra* note 2, at 8–10. They did so partly to avoid Republican charges of anti-conservative bias. *Id.*

50 PAUL M. BARNETT, NYU STERN CTR. FOR BUS. & HUM. RTS., SPREADING THE BIG LIE: HOW SOCIAL MEDIA SITES HAVE AMPLIFIED FALSE CLAIMS OF U.S. ELECTION FRAUD (2022), https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/6321e2ca392ecd06e5d60c4c/1663165130817/NYU+Stern+Center+report-+Spreading+the+Big+Lie_FINAL.pdf [https://perma.cc/QMD2-UVTW].

51 *See* Tarleton Gillespie, *Content Moderation, AI, and the Question of Scale*, 7 BIG DATA & SOC'Y, July–Dec. 2020, at 1; *see also* GIOVANNI PITRUZZELLA & ORESTE POLLICINO, DISINFORMATION AND HATE SPEECH: A EUROPEAN CONSTITUTIONAL PERSPECTIVE 1 (2020) (noting that while hate speech and disinformation are hardly new phenomena, "the potentially

media firms also lack an overriding institutional commitment to reasoned, fact-based discussion. Indeed, just the opposite: social media recommender systems and user interfaces are designed to serve the surveillance capitalism business model, and in so doing, amplify outrageous, emotionally captivating content.[52] Social media content moderation policies remain deeply intertwined with that algorithmic business model and the platforms' overriding commercial interests.[53] Further, social media provide user affordances, including rapid, viral sharing of user posts and tallying how many times other users have shared and liked one's posts, that give strong incentives for users to propagate the kind of bullying, harassment, and rage that is likely to go viral and provoke the response of other users.[54]

Those features of social media design and function harm democracy on multiple fronts. I briefly survey some of those harmful social media effects.

A. *Extremism and Affective Polarization*

Social media fuel extremism and ethnic hatred. As the recent Facebook revelations highlight, the machine learning models that Facebook uses to maximize engagement on the platform favor controversy, misinformation, extremism, and other "outrageous stuff."[55] There is mixed evidence regarding the extent to which such recommender systems radicalize users by funneling ever more extreme content based

global scale of their reach, and the unprecedented speed of their dissemination, raise concerns that are specific to our digital age").

52 *See* ZUBOFF, *supra* note 45; *see also* Luke Thorburn, Priyanjana Bengani & Jonathan Stray, *How Platform Recommenders Work*, MEDIUM (Jan. 20, 2022), https://medium.com/understanding-recommenders/how-platform-recommenders-work-15e260d9a15a [https://perma.cc/7KL8-2FWS] (explaining that platform recommender system algorithms optimize for user engagement in various ways, but have also come to incorporate "integrity signals" that, to some extent, block or demote illegal, violent, or low quality news content).

53 *See* Niva Elkin-Koren, *Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence*, 7 BIG DATA & SOC'Y, July–Dec. 2020, at 1, 4–5.

54 William J. Brady, Killian McLoughlin, Tuan N. Doan & Molly J. Crockett, *How Social Learning Amplifies Moral Outrage Expression in Online Social Networks*, 7 SCI. ADVANCES, Aug. 13, 2021, at 1, 4, https://www.science.org/doi/10.1126/sciadv.abe5641 [https://web.archive.org/web/20230523021429/https://www.science.org/doi/10.1126/sciadv.abe5641]; *see also* Drew Harwell & Taylor Lorenz, *Sorry You Went Viral*, WASH. POST (Oct. 21, 2022, 5:00 AM), https://www.washingtonpost.com/technology/interactive/2022/tiktok-viral-fame-harassment [https://perma.cc/4V86-67DL] (reporting that on TikTok, nothing goes viral as much as bullying, harassment, and rage).

55 Karen Hao, *How Facebook Got Addicted to Spreading Misinformation*, MIT TECH. REV. (March 11, 2021), https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation [https://perma.cc/6HN3-CTXM]; Hagey & Horwitz, *supra* note 1.

on the user's initial forays in that direction. It may differ from platform to platform. An internal Facebook study in 2016 found that "64% of all extremist group joins are due to our recommendation tools."[56] However, another study shows that YouTube recommends extremist videos largely to those who already hold those views——a small group of people who exhibit a high level of gender and racial resentment and who often subscribe to extremist channels.[57] Yet other studies show that rabid partisans who attract an online audience are often egged on to post ever more ardently hostile expression, attacking both leaders of the opposing party and those deemed to be too moderate in their own party.[58] As Chris Bail observes, such vitriolic trolls find status and support in a community of like-minded social media users, and then take increasingly extreme, vindictive positions to prove their loyalty to the cause——and to garner more likes and followers.[59]

In short, social media amplify extremist speech and serve as a catalyst for those with extremist proclivities to further retrench their views and to organize campaigns, both online and off, aimed at furthering their radical, authoritarian ideology. That phenomenon poses dangers for democracy in and of itself. Further, even if social media recommender systems primarily radicalize users who are already highly receptive to extremist messages, the incentives, built into human psychology and social media design, to maximize engagement of other users lead partisans to express ever more extremist positions. In any case, social media tend to normalize one side of partisan extremism, for those who share those views, while exaggerating the extremism and aggression of the other side by amplifying posts that express partisan outrage.[60]

As such, social media have the broad effect of driving affective polarization—the tendency to intensely dislike and distrust those with opposing views.[61] Today, most Americans agree that voters of the opposing party are "ignorant, narrow minded, and ideologically

---

56 Hao, *id.*

57 Annie Y. Chen, Brendan Nyhan, Jason Reifler, Ronald E. Robertson & Christo Wilson, Subscriptions and External Links Help Drive Resentful Users to Alternative and Extremist YouTube Channels (Apr. 2, 2023) (unpublished manuscript), https://arxiv.org/abs/2204.10921 [https://perma.cc/LPD5-GC9J]; *see also* Megan A. Brown et al., Echo Chambers, Rabbit Holes, and Algorithmic Bias: How YouTube Recommends Content to Real Users (Nov. 11, 2022) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4114905 [https://perma.cc/A8N4-MK9J].

58 See BAIL, *supra* note 41, at 63–66.

59 *Id.* at 65–66.

60 *Id.* at 67.

61 *See generally* JAIME E. SETTLE, FRENEMIES: HOW SOCIAL MEDIA POLARIZES AMERICA (2018); Törnberg, *supra* note 42.

driven."[62] Affective polarization results, in part, because people encounter a surfeit of extremist speech on social media, where extremists post far more often than moderates.[63] As a result, people who get their news primarily from social media tend to exaggerate the ideological extremism of people from the other political party significantly more than those who get their news from traditional media.[64]

The merger of social and political identity further contributes to affective polarization.[65] And, as Jaime Settle details, various design features of Facebook's Feed facilitate that merger.[66] These include the public intermixing of political and social communications, express quantification of feedback on users' political expression from those in one's online social network, and encountering short statements expressing the political viewpoints of a person's weakest social connections without the broader context of ongoing face-to-face interaction. These features also crystallize apparent differences between people who share one's views and those who do not, thus facilitating abstraction-based stereotyping.[67] In turn, the resultant partisan animus—as fused with individuals' social identify—tends to push users towards more intolerant and extreme positions.[68]

B. *Hate Speech*

Hate speech is speech that willfully promotes violence or hatred on the basis of targeted individuals' race, gender, religious affiliation, sexual orientation, gender identity, ethnicity, national origin, age, disability, or disease.[69] Hate speech is ubiquitous on social media even though a small minority of users post it. While studies of the prevalence and impact of online hate speech are far from definitive, it appears that less than one

62 SETTLE, *id.*, at 6.

63 Christopher A. Bail et al., *Exposure to Opposing Views on Social Media Can Increase Political Polarization*, 115 PNAS 9216 (2018).

64 BAIL, *supra* note 41, at 75–76.

65 SETTLE, *supra* note 61, at 7.

66 *Id.* at 78–80.

67 *Id.*; *see also* Törnberg, *supra* note 42, at 1 (finding that "digital media polarize through partisan sorting, creating a maelstrom in which more and more identities, beliefs, and cultural preferences become drawn into an all-encompassing societal division").

68 *See* James N. Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky & John Barry Ryan, *Affective Polarization, Local Contexts and Public Opinion in America*, 5 NATURE HUM. BEHAV. 28 (2021) (discussing impact of affective polarization).

69 That said, there is no single agreed-upon definition of hate speech. *See* Alexandra A. Siegel, *Online Hate Speech*, *in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM 56, 56–58 (Nathaniel Persily & Joshua A. Tucker eds., 2020) [hereinafter SOCIAL MEDIA AND DEMOCRACY] (discussing various definitions of hate speech).

percent of social media posts consist of hate speech.[70] However, given that total social media posts number in the billions every day, even a small fraction would amount to a very large quantity of posts. Further, those who generate hate speech online are densely connected to one another and tend to share hateful posts at a high rate. One study of hate speech on Gab, a platform that touts minimal content moderation, found that hate speech tends to spread faster, farther, and reach a much wider audience than do other types of posts.[71]

Not surprisingly, therefore, reported exposure to online hate speech is quite common. In a cross-national survey of teenage and young adult internet users, published in 2017, fifty-three percent of American respondents reported being exposed to hate speech online, while forty-eight percent of Finnish, thirty-nine percent of British, and thirty-one percent of German respondents reported such exposure.[72] Another study found that social media users report statistically significantly greater exposure to hate speech than do nonusers.[73]

Online hate speech has a deleterious psychological impact on those who are targeted, including, most obviously, increased fear and anxiety.[74] Further, of particular import to democracy, studies find that exposure to online hate speech may push people to withdraw from public debate, both online and offline, thus diminishing civic engagement.[75] In that light, coordinated hate speech campaigns are regularly used as a tool to harass, intimidate, and silence journalists; artists; bloggers; high-profile, politically engaged social media users; local officials; and other public figures.[76] Finally, there is convincing evidence that online hate speech

70 *Id.* at 66.

71 Binny Mathew, Ritam Dutt, Pawam Goyal & Animesh Mukherjee, *Spread of Hate Speech in Online Social Media*, WEBSCI '19: PROC. 10TH ACM CONFERENCE ON WEB SCI., June 2019, at 173, 181; *see also* Manoel Horta Ribeiro, Pedro H. Calais, Yuri A. Santos, Virgílio A. F. Almeida & Wagner Meira Jr., Characterizing and Detecting Hateful Users on Twitter (Mar. 23, 2018) (unpublished manuscript), https://arxiv.org/pdf/1803.08977.pdf [https://perma.cc/D8ZN-39LT].

72 James Hawdon, Atte Oksanen & Pekka Räsänen, *Exposure to Online Hate in Four Nations: A Cross-National Consideration*, 38 DEVIANT BEHAV. 254, 260 (2017).

73 Matthew Barnidge, Bumsoo Kim, Lindsey A. Sherrill, Žiga Luknar & Jiehua Zhang, *Perceived Exposure to and Avoidance of Hate Speech in Various Communication Settings*, 44 TELEMATICS & INFORMATICS, July 2019, at 1.

74 Siegel, *supra* note 69, at 68; Magdalena Obermaier & Desirée Schmuck, *Youths as Targets: Factors of Online Hate Speech Victimization Among Adolescents and Young Adults*, 27 J. COMPUT.-MEDIATED COMMC'N, July 2022, at 1, 1 (presenting results of a survey of young adults in Germany).

75 Siegel, *supra* note 69, at 68.

76 *Id.* at 64–65. As Daniel E. Rauch cogently argues, operatives' use of social media for coordinated defamation campaigns against public officials, civil servants, and political opponents similarly aims to harass, intimidate, and silence. *See* Daniel E. Rauch, *Defamation as Democracy Tort*, 172 U. PA. L. REV. (forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4423892 [https://perma.cc/AFF6-2YZL].

normalizes hatred of the targeted groups and triggers violent hate crimes offline.[77]

In fueling hate speech, social media thus undermines foundational principles of democratic governance.[78] Most basically, rampant hate speech deprives its targets of human dignity and of recognition as free and equal citizens. Beyond that, hate speech both intimidates victims and conveys the message that the views of targeted individuals and groups are not worthy of consideration. In so doing, hate speech silences targeted individuals and groups, denying them a voice in public discourse.[79]

C. *Disinformation and Epistemic Fog*

Disinformation is the deliberate propagation of false or misleading content with the intent to deceive.[80] Democracy need not demand the suppression of all such falsehood.[81] But democracy does depend on institutions dedicated to capturing "reality and distinguish[ing] reliable knowledge about it from falsehoods, errors, bullshit, or even just unproven belief."[82] Social media are, at best, indifferent to such truthfulness. They aim to propagate whatever content maximizes user engagement and enables the platform to collect data that can be used for targeted behavioral advertising. As such, social media undermine the fundamental epistemic predicate of pluralist democracy.

Social media recommender systems amplify conspiracy theories and other disinformation because such content plays to users' emotions, particularly by igniting partisan outrage.[83] In addition, trolls are adept at using bots, fake "sock puppet" accounts, and coordinated propaganda campaigns to bolster false content's apparent popularity, which in turn manipulates recommender systems into identifying that false content as

77 Siegel, *supra* note 69, at 67, 69–70.

78 *See generally* Steven J. Heyman, *Hate Speech, Public Discourse, and the First Amendment*, *in* EXTREME SPEECH AND DEMOCRACY 158 (Ivan Hare & James Weinstein eds., 2009) (arguing that public hate speech should not be protected under the First Amendment).

79 *See* OWEN M. FISS, THE IRONY OF FREE SPEECH 25 (1996); *see also* sources cited *infra* note 284.

80 Andrew M. Guess & Benjamin A. Lyons, *Misinformation, Disinformation, and Online Propaganda*, *in* SOCIAL MEDIA AND DEMOCRACY, *supra* note 69, at 9–10.

81 This paragraph draws on Huq, *supra* note 15, at 1115.

82 Huq, *supra* note 15, at 1115 (quoting SOPHIA ROSENFELD, DEMOCRACY AND TRUTH: A SHORT HISTORY 19–20 (2019)).

83 Cameron Martel, Gordon Pennycook & David G. Rand, *Reliance on Emotion Promotes Belief in Fake News*, 5 COGNITIVE RSCH., Oct 7, 2020, at 1; Guess & Lyons, *supra* note 80, at 16–18.

highly engaging to users and thus to amplifying it further.[84] Such tools are regularly deployed by foreign and domestic actors, with motivations ranging from political to financial, to fuel outrage, suppress voter turnout, intimidate opponents, drown out opposing views, and manufacture consensus.[85] Operatives also use such tools to instill epistemic uncertainty by simply overwhelming the public with an avalanche of competing stories, what Steven Bannon notoriously called "to flood the zone with shit."[86]

In that regard, studies have found that social bots account for much of the traffic surrounding disinformation.[87] Bots first amplify false content in the early stages of dissemination. Once the false content spreads organically, bots single out influential social media accounts, aiming to leverage their influence by gaining attention through replies and mentions. Among countless documented examples, at the height of the fabricated Pizzagate conspiracy that linked Hillary Clinton to an alleged child abuse ring during the 2016 presidential campaign, automated Twitter shell accounts, many originating in Cyprus, the Czech Republic, and Vietnam, rampantly circulated memes pushing the rumor, helping the story to spread like wildfire.[88] As scholar of computational propaganda Samuel Woolley notes, "bots are "growing increasingly sophisticated and difficult to detect."[89] Moreover, there is good reason to fear that the coming integration of AI into social media will greatly

84 *See generally* WOOLLEY, *supra* note 1 (detailing the nature and use of such tools for political manipulation). As Woolley explains, a bot is "a software tool built to do autonomous tasks, including communicate with other users online. Bots are often core mechanisms for spreading computational propaganda." *Id.* at 11.

85 *See generally* COMPUTATIONAL PROPAGANDA: POLITICAL PARTIES, POLITICIANS, AND POLITICAL MANIPULATION ON SOCIAL MEDIA (Samuel C. Woolley & Philip N. Howard eds., 2019) [hereinafter COMPUTATIONAL PROPAGANDA] (presenting country-specific case studies of uses of such tools to manipulate public opinion).

86 Sean Illing, "*Flood the Zone with Shit": How Misinformation Overwhelmed Our Democracy*, VOX (Feb. 6, 2020, 9:27 AM), https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation [https://perma.cc/S4R2-NSFL]; *see also* Mona Elswah & Philip N. Howard, *"Anything that Causes Chaos": The Organizational Behavior of Russia Today (RT)*, 70 J. COMMC'N 623, 623 (2020); Tommaso Venturini, *From Fake to Junk News: The Data Politics of Online Virality*, *in* DATA POLITICS: WORLDS, SUBJECTS, RIGHTS 123, 126 (Didier Bigo, Engin Isin & Evelyn Ruppert eds., 2019) ("'Junk news' is dangerous not because it is false, but because it saturates public debate, leaving little space to other discussions . . . .").

87 Guess & Lyons, *supra* note 80, at 20–21.

88 WOOLLEY, *supra* note 1, at 142–43; Samuel C. Woolley & Douglas Guilbeault, *United States: Manufacturing Consensus Online*, *in* COMPUTATIONAL PROPAGANDA, *supra* note 85, at 185, 193.

89 WOOLLEY, *supra* note 1, at 12. Woolley also notes that sock puppet accounts, in which real people assume false identities online, have become almost equally common vectors for political manipulation. *Id.*

magnify opportunities to manufacture and propagate fully credible audiovisual deepfakes and textual disinformation.[90]

Whatever the causes and motivations, coordinated disinformation campaigns appear to be highly successful. Facebook's internal documents show, for example, that in the run-up to the 2020 election, troll farms reached 140 million U.S. users per month. Seventy-five percent of those users had never previously followed any of the troll farm pages. They were seeing the content only because Facebook's content recommender system pushed it to their news feeds.[91]

Social media algorithms also propagate disinformation by promoting the rapid, viral spread of novel stories and other "breaking news" within and among social media user clusters.[92] Indeed, novelty appears to be a significant driver of misinformation. Social media users tend both to favor sharing spicy, novel stories and to place the highest trust in posts shared by their close friends, as opposed to independently assessing the trustworthiness of the original source of the story. That helps explain why, research shows, conspiracy theories and lies spread faster and more widely on Twitter than does truth.[93] Fueling those user proclivities, social media recommender systems tend to surface close friends' posts. In so doing, the systems prioritize popular, novel content, including disinformation, over trustworthy content in users' feeds.

Significantly, social psychology and cognitive research supports the conclusion that the onslaught of disinformation on social media impacts the beliefs of many who are repeatedly exposed to it. Cognition researchers' experimental data, from studies spanning over four decades, confirm that propaganda—including the repeated exposure to lies——tends to convince recipients that the repeated statements are true.[94] That "repetition-induced truth effect" seems robust to individual differences in cognitive ability and it persists even when participants are warned to

90 Jonathan Haidt & Eric Schmidt, *AI Is About to Make Social Media (Much) More Toxic*, THE ATLANTIC (May 5, 2023), https://www.theatlantic.com/technology/archive/2023/05/generative-ai-social-media-integration-dangers-disinformation-addiction/673940 [https://perma.cc/V8VS-NTHT].

91 Hao, *supra* note 48.

92 This paragraph draws on Guess & Lyons, *supra* note 80, at 20–23.

93 *See* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCIENCE, Mar. 9, 2018, at 1146, 1147. (finding in study of rumors on Twitter that "falsehood diffused significantly farther, faster, deeper, and more broadly than truth in a categories of information").

94 Emma L. Henderson, Daniel J. Simons & Dale J. Barr, *The Trajectory of Truth: A Longitudinal Study of the Illusory Truth Effect*, 4 J. COGNITION, June 8, 2021, at 1.

avoid it, possess knowledge about the facts, or are explicitly informed about which statements are true and which are false.[95]

Beliefs in dubious claims that speak to ideological issues seem also to reflect the motivated reasoning of those who are predisposed to believe claims that support their ideological views.[96] Surveys conducted as late as 2016 showed, for example, that some forty percent of registered Republicans believed that then-President Barack Obama was not born in the United States despite Obama's release of his Hawaii birth certificate. Further, there was little difference in the percentage of believers in the oft-repeated "birther" tale among Republicans with higher and lower political knowledge.[97]

In sum, by propagating disinformation, platforms' recommender systems and the third-party tools that exploit them greatly magnify the force of democracy-destabilizing speech. Importantly, rampant and coordinated disinformation campaigns can harm democracy simply by heightening general epistemic uncertainty and reinforcing beliefs of those who are predisposed to accept the dubious claims, even if most social media users do not believe or even come across particular disinformation content.

D. *Targeted Amplification*

Merely documenting vast amounts of online disinformation, hate speech, extremist incitement, and other "outrageous stuff" does not necessarily establish that most users view that content, that the content influences most users who do view it, or, more generally, that the content causes widespread polarization, epistemic uncertainty, and disillusionment with democracy.[98] Some studies suggest, for example,

95 *Id.* at 2; *see also* Lisa K. Fazio, David G. Rand & Gordon Pennycook, *Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements*, 26 PSYCHONOMIC BULL. & REV. 1705 (2019); Lisa K. Fazio, Nadia M. Brashier, B. Keith Payne & Elizabeth J. Marsh, *Knowledge Does Not Protect Against Illusory Truth*, 144 J. OF EXPERIMENTAL PSYCH.: GEN. 993 (2015).

96 Adam M. Enders & Joseph E. Uscinski, *Are Misinformation, Antiscientific Claims, and Conspiracy Theories for Political Extremists?*, 24 GRP. PROCESSES & INTERGROUP RELS. 583 (2020) (discussing motivated reasoning stemming from preexisting ideological beliefs).

97 Josh Clinton & Carrie Roush, *Poll: Persistent Partisan Divide Over "Birther" Question*, NBC NEWS (Aug. 10, 2016, 2:19 PM), https://www.nbcnews.com/politics/2016-election/poll-persistent-partisan-divide-over-birther-question-n627446 [https://perma.cc/5BH8-Y85W]; *see also* Kaleigh Rogers, *The Birther Myth Stuck Around for Years. The Election Fraud Myth Might Too*, FIVETHIRTYEIGHT, (Nov. 23, 2020, 10:10 AM), https://fivethirtyeight.com/features/the-birther-myth-stuck-around-for-years-the-election-fraud-myth-might-too [https://perma.cc/673X-PQC4].

98 Benkler, *supra* note 33.

that a small percentage of both American and European social media users share and consume blatant disinformation, and that sharing blatant disinformation is concentrated among politically engaged right-wing users who are already highly receptive to its populist, anti-immigrant, and Islamophobic themes.[99] Yet, propaganda that magnifies out-group animosity and further radicalizes the views even of a discrete minority of users may well be sufficient to destabilize democracy. Witness the cascade of falsehoods regarding the "stolen" 2020 presidential election that helped to incite the January 6 Capitol insurrection.[100] Even if social media is relatively benign for the average use base, its exploitation to construct and promote parallel epistemic universes—rejecting science, mainstream journalism, and all other widely accepted epistemic authority—among particularly susceptible communities may wreak havoc on democracy.

In that regard, political operatives' data-driven advertising and influence campaigns are widely used to identify and target vulnerable users and to render users more pliable to such influence.[101] To do so, they draw on online platforms' vast infrastructures for user surveillance, profiling, responsiveness testing, and targeted communications. Digital

99 Richard Fletcher, Alessio Cornia, Lucas Graves & Rasmus Kleis Nielsen, *Measuring the Reach of "Fake News" and Online Disinformation in Europe*, REUTERS INST. (Feb. 2018), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/Measuring%20the%20reach%20of%20fake%20news%20and%20online%20distribution%20in%20Europe%20CORRECT%20FLAG.pdf [https://perma.cc/ZB8A-44NS]; *see also* Benkler, *supra* note 33 (describing Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson & David Lazer, *Fake News on Twitter During the 2016 U.S. Presidential Election*, 363 SCIENCE, Jan. 25, 2019, at 374); Andrew Guess, Jonathan Nagler & Joshua Tucker, *Less Than You Think: Prevalence and Predicators of Fake News Dissemination on Facebook*, 5 SCI. ADVANCES, Jan. 9, 2019, at 1, https://www.science.org/doi/10.1126/sciadv.aau4586 [https://web.archive.org/web/20230727235230/https://www.science.org/doi/10.1126/sciadv.aau4586]. The studies cited by Benkler focus on fake news sites—those that masquerade as genuine news sites but feature largely fabricated content. They do not address sharing and consumption of hyper-partisan sites like *Breitbart*, disinformation presented in user posts and comments, or communications within Facebook groups. Recent studies show that, "compared to liberals, politically conservative [Facebook] users are far more siloed in their news sources, driven in part by algorithmic processes, and especially apparent on Facebook's Pages and Groups." *See* Jeff Horwitz, *Does Facebook Polarize Users? Meta Disagrees with Partners over Research Conclusions*, WALL ST. J. (July 27, 2023, 3:53 PM) (quoting Meagan Phelan, spokesperson for the journal Science), https://www.wsj.com/articles/does-facebook-polarize-users-meta-disagrees-with-partners-over-research-conclusions-24fde67a [https://perma.cc/S5FC-4VZY].

100 *See* DRAFT REPORT, *supra* note 2. For that matter, as I will discuss, Nazi propaganda destabilized the Weimar Republic as much by mobilizing and radicalizing potential supporters as by converting the mass public to support the Nazi Party.

101 *See generally* ANTHONY NADLER, MATTHEW CRAIN & JOAN DONOVAN, DATA & SOC'Y, WEAPONIZING THE DIGITAL INFLUENCE MACHINE: THE POLITICAL PERILS OF ONLINE AD TECH (2018), https://datasociety.net/wp-content/uploads/2018/10/DS_Digital_Influence_Machine.pdf [https://perma.cc/AK4F-242P].

influencers have shown particular interest in predicting users' "underlying psychological profiles," based in large part on data regarding users' personal attributes, including data derived from users' social media consumption and communication.[102] Operatives use such data to create psychologically customized, and highly effective, influence campaigns. As a *Data & Society* report warns:

> With mass consumer surveillance, political advertisers can maximize the potential influence of their nudges by sifting through data to identify who is most likely to be influenced, what kinds of nudges or triggers they may be most affected by, or even factors like at what moments or in what moods a target may be most receptive.[103]

Such tools have been deployed by both domestic and foreign operatives to fuel affective polarization, raise fears of perceived threats, and sow confusion and division among political opponents.[104] Like the propagation of disinformation, the coming integration of AI into social media will vastly enhance both platforms' and operatives' capacity to generate such an emotionally compelling, user-customized feeds.[105]

E. *Tabloid and Mainstream Media*

The destabilizing impact of rampant, targeted disinformation extends beyond direct persuasion on social media. Studies show that online disinformation campaigns spill over to mass media.[106] Most conspicuously, online disinformation reverberates within the hyper-partisan, right-wing tabloid news ecosystem in the United States.[107] But

---

102 *Id.* at 13–14.

103 *Id.* at 38.

104 *Id.* at 27–35.

105 *See* Haidt & Schmidt, *supra* note 90; Josh A. Goldstein, Jason Chao, Shelby Grossman, Alex Stamos & Michael Tomz, Can AI Write Persuasive Propaganda? (Feb. 21, 2023) (unpublished manuscript), https://osf.io/preprints/socarxiv/fp87b [https://perma.cc/Z52G-GM85] (finding that, even using the now outmoded ChatGPT-3, it is possible, with limited effort, to produce propaganda that is nearly as persuasive as that generated by foreign actors).

106 For a comprehensive early study, see ALICE MARWICK & REBECCA LEWIS, DATA &SOC'Y, MEDIA MANIPULATION AND DISINFORMATION ONLINE (2017), https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf [https://perma.cc/J3WH-G3HX].

107 *See* Emily Bazelon, *The Disinformation Dilemma*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY, *supra* note 12, at 44 (noting that "[r]ight-wing media and social media have a symbiotic relationship"); YOCHAI BENKLER, ROBERT FARIS & HAL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS 75–100 (2018) (finding that the right-wing media ecosystem has long been

online disinformation campaigns are also adept at manipulating general news coverage.[108] Indeed, social media trolls strategically employ bots and sock puppet accounts to manipulate recommender algorithms to push a given topic as "trending" and thus to capture the attention of mainstream media.[109] In that way, trolls garner media coverage even if their disinformation campaign is not truly a popular, trending topic among actual social media users.[110]

Social media also serve to amplify expressions of hyper-partisan outrage that originate on right-wing broadcast and digital media. In the United States, the right-wing media ecosystem, encompassing talk-radio, white evangelical Christian broadcasters, cable television outlets, and digital media, is a significant force for stoking intolerant, autocratic populism and rejecting the core epistemological foundations of modern liberal democracy, including fact-based science and expertise.[111] Right-wing media clearly predate social media and serve to reinforce the hyper-partisan identity and outrage of its core audience independently of social media impact.[112] Nonetheless, studies show that social media news sharing tends to amplify right-wing media content whether because outrage and deep resentment capture users' attention and motivate users to share further or because social media recommender systems give such emotive content greater prominence in user feeds.[113] Further, far right

insular and focused on ideological purity rather than fact, and is thus particularly receptive to foreign and domestic propaganda operations, whether originating online or offline). Far right media outlets are also leading spreaders of disinformation on social media. *See* Ian Kennedy et al., *Repeat Spreaders and Election Delegitimization: A Comprehensive Dataset of Misinformation Tweets from the 2020 U.S. Election*, 2 J. QUANTITATIVE DESCRIPTION: DIGIT. MEDIA 1, 28–29 (2022) (study of spreaders of election misinformation on Twitter).

108 Yariv Tsfati et al., *Causes and Consequences of Mainstream Media Dissemination of Fake News: Literature Review and Synthesis*, 44 ANNALS INT'L COMMC'N ASS'N 157, 160 (2020); Chris J. Vargo, Lei Guo & Michelle A. Amazeen, *The Agenda-Setting Power of Fake News: A Big Data Analysis of the Online Media Landscape from 2014 to 2016*, 20 NEW MEDIA & SOC'Y 2028 (2018). With regard to manipulation of unsuspecting journalists, see WHITNEY PHILLIPS, DATA & SOC'Y, THE OXYGEN OF AMPLIFICATION: BETTER PRACTICES FOR REPORTING ON EXTREMISTS, ANTAGONISTS, AND MANIPULATORS (2018), https://datasociety.net/wp-content/uploads/2018/05/1_PART_1_Oxygen_of_Amplification_DS.pdf [https://perma.cc/4E35-LAJP].

109 *See* WOOLLEY, *supra* note 1, at 132–33.

110 *Id.* at 121 (discussing operatives' use of bots to create the illusion of popularity and thereby to "get[] their pet ideas into the mainstream media").

111 *See* Yochai Benkler, *A Political Economy of the Origins of Asymmetric Propaganda in American Media*, *in* THE DISINFORMATION AGE: POLITICS, TECHNOLOGY, AND DISRUPTIVE COMMUNICATION IN THE UNITED STATES 43, 50–54 (W. Lance Bennett & Steven Livingston eds., 2021); JACOB S. HACKER & PAUL PIERSON, LET THEM EAT TWEETS: HOW THE RIGHT RULES IN AN AGE OF EXTREME INEQUALITY 100–07 (2020).

112 *See* BENKLER, FARIS & ROBERTS, *supra* note 107, at 311–39.

113 *See* Sandra González-Bailón, Valeria d'Andra, Deen Freelon & Manlio De Domenico, *The Advantage of the Right in Social Media News Sharing*, 1 PNAS NEXUS, no. 3, July 2022, at 1,

media outlets are themselves leading spreaders of disinformation on social media.[114] Concomitantly, news items that Meta fact-checkers have identified as false are viewed and reshared almost entirely by right-wing users.[115]

At bottom, social media and news media depend on each other for audience attention. As a result, content likely to capture user and audience attention flows dynamically between social media and news media. In that feedback loop, conspiracy theories, disinformation campaigns, and populist right-wing outrage tend to move up the chain from social media, to right-wing hyper-partisan media, to coverage on mainstream news media, and back to social media, where it is further amplified with user likes and shares.[116]

In turn, news media effects research suggests that repeated exposure to brazen disinformation increases cynicism, apathy, and epistemic uncertainty, as audiences are overwhelmed by a "firehose of falsehood."[117] Ultimately, that second-order effect may contribute to reduced trust in legitimate news media and is likely to feed extremism and affective polarization.[118]

F. *Political Instability*

Finally, by fueling outrage, affective polarization, and the rapid, viral spread of disinformation and conspiracy theories, social media radically undermine the capacity for shared epistemic understanding and broadly accepted, legitimate authority upon which democratic governance depends. As Rick Pildes documents, wild, unsubstantiated charges in a YouTube video or Facebook post that go viral shortly before an election—and which are often then reported on by mass media—can

https://doi.org/10.1093/pnasnexus/pgac137 [https://perma.cc/9RGH-6EMN]; *see also* Carsten Schwemmer, *The Limited Influence of Right-Wing Movements on Social Media User Engagement*, 7 SOC. MEDIA + SOC'Y, July–Sept. 2021, at 1 (finding that the German right-wing movement Pegida attains user engagement by posting ever more extreme xenophobic content rather than by simply creating more posts).

114 *See* Kennedy et al., *supra* note 107, at 28–29 (study of spreaders of election misinformation on Twitter).

115 Sandra González-Bailón et al., *Asymmetric Ideological Segregation in Exposure to Political News on Facebook*, 381 SCIENCE, July 28, 2023, at 392.

116 *See* Yini Zhang, Zhiying Yue, Xiyu Yang, Fan Chen & Nojin Kwak, *How a Peripheral Ideology Becomes Mainstream: Strategic Performance, Audience Reaction, and News Media Amplification in the Case of QAnon Twitter Accounts*, NEW MEDIA & SOC'Y, Nov. 26, 2022.

117 Guess & Lyons, *supra* note 80, at 24–25.

118 MARWICK & LEWIS, *supra* note 106, at 44–47.

affect enough voters to alter the results.[119] Further, social media contributes substantially to the weakening and disintegration of established, broad-coalition political parties that are essential for effective governance.

In European proportional representation political systems, fringe movements are able to destabilize democracy by employing viral, targeted marketing to create the illusion of an organic, bottom-up popular uprising, which then cascades on social media, garners mass media attention, and attracts real followers.[120] Organizers use a combination of fake accounts, social bots, social media engagement algorithms that promote outrage, dark money contributions, Facebook's recommendation to join ever more extremist fringe Facebook groups, and Facebook's magnification of feeds that originate in those groups. Similarly, the small, fringe group of Canadian truckers protesting COVID vaccine mandates was transformed by social media, right-wing conspiracy theorists, grifter opportunists, and then right-wing broadcasters into an ersatz popular uprising, which snowballed as it garnered dark money contributions and mobilized support in other countries.[121]

Likewise, Vox, once a fringe far-right-wing political party in Spain touted its "Make Spain Great Again" slogan and disseminated videos of rallies on social media as part of a carefully orchestrated marketing campaign to "make anyone following Vox feel as if they were part of something big, exciting, growing—and homogenous."[122] Like the radical right-wing AfD in Germany, Vox supporters also made extensive use of bots to pump out right-wing disinformation and conspiracy theories and to amplify their presence on social media.[123] Using that strategy, Vox

119 *See* Pildes, *supra* note 9, at 2059–60.

120 *See* WOOLLEY, *supra* note 1, at 8 (describing the leveraging of social media to create an appearance of popularity in an attempt to generate bandwagon support).

121 *See* Ryan Broderick, *How Facebook Twisted Canada's Trucker Convoy into an International Movement: A Labyrinth of Facebook Groups and Right-Wing Media*, THE VERGE (Feb. 19, 2022, 9:00 AM), https://www.theverge.com/2022/2/19/22941291/facebook-canada-trucker-convoy-gofundme-groups-viral-sharing [https://perma.cc/5F49-ZNLC]; Zack Beauchamp, *The Canadian Trucker Convey Is an Unpopular Uprising*, VOX (Feb. 11, 2022, 7:50 AM), https://www.vox.com/policy-and-politics/22926134/canada-trucker-freedom-convoy-protest-ottawa [https://perma.cc/ZP32-7KNX].

122 APPLEBAUM, *supra* note 29, at 123.

123 *Id.* at 132–33 (reporting on findings of Alto Data Analytics and the Institute for Strategic Dialogue). On the use of bots by AfD supporters to boost AfD's message online, see Juan Carlos Medina Serrano, Morteza Shahrezaye, Orestis Papakyriakopoulos & Simon Hegelich, *The Rise of Germany's AfD: A Social Media Analysis*, SMSOCIETY '19: PROC. 10TH INT'L CONF. ON SOC. MEDIA & SOC., Jul. 2019, at 214; *see also* Jacob van de Kerkhof & Catalina Goanta, *Fakeness in Political Popularity*, VERFASSUNGSBLOG (Oct. 28, 2022), https://verfassungsblog.de/fakeness

successfully built itself into a significant political force, like other right-wing populist parties in Europe. Such strategic employment of deceptive technologies, including gaming the algorithms that platforms use to identify popular content and surface it to users, fuels the delegitimization of all forms of constituted political authority, the emergence of spontaneous, non-organized pop-up groups like the Yellow Vests in France, and the debilitating fragmentation of political parties.[124]

In the two-party system in the United States, social media has spawned internal fragmentation within the two major parties by fueling the rise of free-agent politicians on the extreme margins of each party. Congressional party leaders used to enjoy considerable leverage over members through party fundraising and by allocating high-profile committee assignments, which typically went to senior members who had loyally served the party for many years.[125] As such, leaders were able to assert the party discipline needed for effective governance, including reaching compromises with the opposing party.

In the age of social media, however, newly elected backbenchers are able to garner national attention and millions of dollars in donations from individual donors through peddling outrage and extreme partisan positions on social media. A prime example: newly elected House Representative Marjorie Taylor Greene was removed from all House committees, by a bipartisan vote, because of her expressed support for political violence and her peddling of far-right, anti-Semitic, and white supremacist conspiracy theories, including QAnon and Pizzagate, on social media. She then quickly raised over $3.2 million in a single calendar quarter, from over 100,000 individual donors.[126] In January 2023, the House GOP Steering Committee agreed to place Greene on the Homeland Security Committee.[127]

G. Sum

As meta-analyses of the scientific literature conclude, social media contribute significantly to emergent authoritarian populism, declining political and social trust, growing polarization, and citizen ignorance

[https://perma.cc/5BUC-SH6W] (noting that some established political parties apparently also use fake accounts and fake messages on social media to give the appearance of greater popularity).

124 Pildes, *supra* note 9, at 2060–66.

125 *Id.* at 2066–67.

126 *Id.* at 2067.

127 Melanie Zanona & Manu Raju, *Marjorie Taylor Greene and Paul Gosar Get Committee Assignments*, CNN POLITICS (Jan. 17, 2023, 6:40 PM), https://www.cnn.com/2023/01/17/politics/marjorie-taylor-greene-paul-gosar-committee-assignments/index.html [https://perma.cc/R8GE-G3YU].

about the pressing issues of the day.[128] Studies point to several contributing factors to that assault on sustainable democratic self-government. Social media business models and platform design amplify extremism, affective polarization, hate speech, disinformation, and propaganda. In so doing, they fuel the disintegration of widely shared epistemic understandings and undermine democratic political authority. Online platforms' vast capacity for user surveillance, profiling, responsiveness testing, and customized communications serve both the platforms themselves and third-party operatives in targeting and manipulating susceptible users. They also enable operatives to manufacture an illusion of a groundswell of interest in particular topics or support for political movements, thus influencing legacy news media coverage and popular support.

At bottom, social media's surveillance capitalism business model and design are fundamentally antithetical to the democratic ideal of reason-based collective self-determination among equal participants. The platforms privilege raw emotion over reasoned argument based on verifiable fact. They also foster imbalances in discursive power by favoring speakers with the wherewithal to manipulate content curation algorithms and user affordances.[129] Finally, social media platforms fundamentally deceive their users. The platforms hide the fact that it is the logic of surveillance capitalism—and not anything approximating journalistic integrity--that determines the mix of content that appears in users' feeds.[130]

Granted, social media are far from the only cause of the democratic backsliding of recent years. In the United States, moreover, right-wing legacy media have, for decades, stoked autocratic populism, racial and ethnic hatred, and a rejection of basic journalistic and scientific norms for validating truth. But social media both fuel and amplify those corrosive forces. As such, the harms social media inflict on democracy are independently palpable and severe. Under principles of militant democracy, it is incumbent on democratic states to defend against those harms while ensuring that social media continue to serve as vibrant fora for diverse voices, undistorted by the platforms' surveillance capitalism business model, to the extent that is possible.

Yet, the United States has failed to mount a regulatory response to social media harms despite the growing body of evidence that social media undermine democracy on a number of fronts. The next Part investigates the underlying reasons for that inaction. We then turn to the

128 *See supra* notes 5–7 (citing meta-analysis studies).

129 DiResta, *supra* note 12, at 128.

130 I thank Lea Rabe for highlighting these points in noting that social media function diametrically opposed to the Habermasian concept of ideal communication.

militant democracy framework that informs the European Union's multi-faceted regulatory blueprint for defending democracy against the social media assault.

II. IDEOLOGICAL AND DOCTRINAL BARRIERS TO ADDRESSING SOCIAL MEDIA HARMS IN THE UNITED STATES

A. *The Road Not Taken in the United States*

As of this writing, the European Union has assumed global leadership in addressing social media's palpable threats to democracy. In 2020, the European Commission adopted the European Democracy Action Plan, which aspires to defend democratic institutions against social media's corrosive impact.[131] As the Action Plan underscores, democratic governments cannot remain passive in the face of autocratic populist exploitation of online platforms to undermine election integrity, funnel disinformation, manipulate voters, and intimidate journalists, women, and minority speakers through targeted harassment and hate speech.[132]

Accordingly, as further described in Part IV below, the Action Plan advances a potentially far-reaching multi-pronged regulatory framework to obligate large online platforms to identify, report to regulators, and mitigate any systemic risk that the design, functioning, or use of their services will propagate hate speech, terrorist incitement, or other illegal content or will otherwise undermine civic discourse, the electoral process, or the exercise of fundamental rights. As set out in the Digital Services Act and related measures, large online platforms will also be required to enhance the prominence of authoritative, trustworthy information, while demoting disinformation in their users' feeds.[133] Further, while platforms need not proactively monitor user posts for illegal content, they are obligated to expeditiously remove illegal content of which they do have knowledge.[134] In that regard, moreover, platforms must establish procedures for state-credentialed "trusted flaggers" and others to notify them of hosted content that is illegal under the law of any EU country to which the hosting platform is subject.[135] Finally, EU regulations impose on large platforms various transparency and due process obligations aimed at minimizing errors and biases in the

131 Democracy Action Plan, *supra* note 23.

132 Democracy Action Plan, *supra* note 23, at 1–3, 10.

133 *See infra* notes 329–331 and accompanying text.

134 *See infra* note 337 and accompanying text.

135 *See infra* note 337 and accompanying text.

platforms' content moderation algorithms, particularly those that might systematically suppress minority viewpoints and thus stifle democratic debate.[136]

The EU framework combines various regulatory approaches. It includes direct government command and control regulation; "co-regulation," in which the platforms are required to participate in implementing broad regulatory goals; and prodding platforms to engage in more effective self-regulation by removing harmful content and mitigating systemic risks that their systems will be used to undermine democracy.[137] All in all, the framework aims to require and/or induce large online platforms to employ their recommender system and content moderation algorithms, in conjunction with notices from government officials and trusted flaggers, to prevent the operation and exploitation of their systems for the propagation of content that poses substantial risks to democracy.[138] At the same time, the EU framework aims to enlist large platforms to serve as vital public forums for diverse, trustworthy, and fact-based information and opinion.

As such, and as I enumerate in Parts III and IV, the European Democracy Action Plan and framework for regulating online platforms draw upon foundational dictates of militant democracy. As embedded in European constitutions, human rights law, and regulatory policy, the concept of militant democracy posits that democratic states must actively defend themselves against palpable threats to their continued existence as democracies. That means, most narrowly, that democracies must have the constitutional prerogative to ban avowedly antidemocratic incitement and political movements. More broadly, it requires that democratic states affirmatively underwrite an inclusive, pluralist, and broadly egalitarian public sphere and promote reason-based public deliberation, free from

136 *See infra* notes 340–343 and accompanying text.

137 For discussion and comparison of these regulatory approaches, see Michèle Finck, *Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy*, 43 EUR. L. REV. 47 (2018). *Cf.* Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117, 153–61 (2021) (describing a variety of "collaborative governance" models, employed in a wide range of regulatory frameworks in the United States, that "aim to chart a course between top-down, state-centered approaches on the one side and total deregulation on the other").

138 Jack Balkin characterizes such regulation somewhat less charitably: "Europe has taken advantage of the [platforms'] algorithmic administrative system by employing platform companies as private bureaucracies for speech governance." Jack M. Balkin, *Free Speech Versus the First Amendment*, 70 UCLA L. REV. 1206, 1245 (2023). Yet, as Professor Balkin well recognizes, but for state intervention, the platforms employ those algorithmic systems to fuel emotionally captivating (and thus often divisive, socially harmful) content in a manner designed to maximize user engagement and conduct pervasive surveillance of user activity to better target and market advertising. *Id.* at 1243-44.

the undue influence of manipulative propaganda, whether propagated by market forces, political actors, or the state.

For its part, the United States has yet to enact any federal legislative or regulatory measures targeting social media's harms. That inaction lies in part in the flatly contradictory proposals from across the partisan divide about how, if at all, online platforms should be regulated.[139] More broadly, U.S. regulatory passivity flows from the neoliberal techno-utopianism that has dominated both U.S. technology policy and First Amendment law in recent decades. I consider each in turn.

B. *American Technology Policy: Neoliberal and Techno-Utopian*

American technology policy and current First Amendment doctrine exhibit a far-reaching neoliberal aversion to regulating private power. With roots in the writings of Fredrich Hayek, neoliberalism gained force in the early 1980s—the era of Ronald Reagan and Margaret Thatcher——as a counter to Keynesian macroeconomic planning and social democracy.[140] Neoliberalism posits that human welfare is best advanced by strong private property rights, free markets, and free trade. In this view, the state should regulate industry and finance only as necessary to secure the proper functioning of markets. Indeed, aside from regulations required to underwrite market competition, all public goods, including parks, water, education, telecommunications, health care, social security, prisons, and protecting the environment, should be privatized and subject to market forces and incentives.

Significantly, neoliberalism constitutes an all-encompassing political theory, not just a laissez-faire economic policy. Following Hayek, neoliberals posit that government's sole legitimate function is to enhance market competition.[141] Concomitantly, neoliberals admit no role

---

139 At this point in time, Democrats typically want to require platforms to take greater responsibility for policing harmful online content, while Republicans want to transform platforms into quasi-common carriers, with minimal prerogative to block third-party content on their sites. *See* Mark A. Lemley, *The Contradictions of Platform Regulation*, 1 J. FREE SPEECH L. 303, 306–11 (2021); *see also* Evelyn Douek & Genevieve Lakier, *First Amendment Politics Gets Weird: Public and Private Platform Reform and the Breakdown of the Laissez-Faire Free Speech Consensus*, U. CHI. L. REV. ONLINE (June 6, 2022), https://lawreviewblog.uchicago.edu/2022/06/06/douek-lakier-first-amendment [https://perma.cc/26PF-5UXT] (discussing convoluted, temporally inconsistent views about regulating speech intermediaries among conservatives and liberals).

140 *See* WENDY BROWN, UNDOING THE DEMOS: NEOLIBERALISM'S STEALTH REVOLUTION 20–21 (2015); DAVID HARVEY, A BRIEF HISTORY OF NEOLIBERALISM 2–3 (2005).

141 This paragraph draws on Annabel Herzog, *The Attack on Sovereignty: Liberalism and Democracy in Hayek, Foucault, and Lefort*, 49 POL. THEORY 662, 664–66 (2021), and David Singh

for the democratic pursuit of common policy goals characterized by reason-based debate about what those goals should be and how government should seek to attain them. According to neoliberal thought, there are too many things that we cannot know, and collective efforts to pursue rational, government-directed supply of social goods will inevitably lead to error, concentrations of corrupt power, and repression. Rather, neoliberalism posits, well-being, freedom, and knowledge are best secured by the "morals of the market," the "set of individualistic, commercial values that [prioritize] the pursuit of self-interest above the development of common purposes."[142]

The end result is that, while democratic politics might demand distributive fairness, workplace security, civic equality, state provision of basic welfare goods, and the nurturing of social solidarity, neoliberalism presses against those democratic claims in the service of capital accumulation and market imperatives.[143] For neoliberals, only the spontaneous operation of market forces—the "automatic mechanism of adjustment" fueled by unhindered market competition among rationally self-interested actors and represented in the price system—yields the optimal, objectively accurate measure of human well-being.[144] As we will see, that understanding underlies current conceptions of the marketplace of ideas no less than markets for non-speech goods.

Of particular import for social media, neoliberalism has both drawn upon and celebrated information technology on a number of fronts. First, the neoliberal project of bringing vast swaths of human activity into the domain of the market looks to information technology to accumulate, transfer, analyze, and exploit massive databases capable of amassing wealth and guiding decisions in the global marketplace.[145] Second, for neoliberals, the transformation of information searches and interpersonal communications into commercially exploitable data and the rise of profit-seeking social media influencers represent yet another celebrated penetration of the market into social life.[146] Third, neoliberalism trumpets

Grewal & Jedediah Purdy, Introduction, *Law and Neoliberalism*, 77 LAW & CONTEMP. PROBS., no. 4, 2014, at 1.

142 Herzog, *supra* note 141, at 666 (alteration in original) (quoting JESSICA WHYTE, THE MORALS OF THE MARKET: HUMAN RIGHTS AND THE RISE OF NEOLIBERALISM 11 (2019)).

143 *See* Grewal & Purdy, *supra* note 141, at 3–4.

144 Stephen Metcalf, *Neoliberalism: The Idea that Swallowed the World*, THE GUARDIAN, (Aug. 18, 2017, 1:00 AM) https://www.theguardian.com/news/2017/aug/18/neoliberalism-the-idea-that-changed-the-world [https://perma.cc/P6B7-7J9J] (describing the neoliberal philosophy of Friedrich Hayek and its influence on current understandings and policies).

145 HARVEY, *supra* note 140, at 3.

146 *See* Mike Berry, *Neoliberalism and the Media*, *in* MEDIA AND SOCIETY 57, 69, 73 (James Curran & David Hesmondhalgh eds., 6th ed. 2019); Sarah Manavis, *"Social Media Is Sentient Neoliberalism": Symeon Brown on the Exploitative Influencer Economy*, NEW STATESMAN (Mar.

the global online economy and seemingly free flows of information as the epitome of an ideal, unregulated market.[147] As Paul Starr has aptly noted: "The explosive growth of the online economy in the 1990s and early 2000s appeared to validate the idea that markets were best left to themselves. The internet of that era was neoliberalism's greatest triumph."[148]

Alongside neoliberalism, early internet evangelists imagined that digital networks and information technology were "technologies of freedom" that would promote personal autonomy, foster nonhierarchical bottom-up decision-making, enable unprecedented sharing of dispersed information. and engender new forms of creative production.[149] Many techno-utopians also shared the neoliberal belief in free markets and a profound distrust of government regulation.[150] They advanced the notion of internet exceptionalism, the belief that the unique nature of the internet renders government regulation inappropriate and, indeed, given the internet's global reach, untenable.[151] As Anu Bradford has recently described, "the American market-driven model . . . extends beyond traditional neoliberal thinking," bringing together "the cultural bohemianism of San Francisco with the high-tech industries invented in Silicon Valley . . . under a shared rubric of profound techno-optimism."[152] Yet, such laissez-faire information technology capitalism actually fits firmly within the broader neoliberal framework. It is yet another instance in which neoliberals favor enacting "not just laws that enable markets but also laws that protect [markets] from democratic majorities that might remake them."[153]

Indeed, as buttressed by the twin forces of neoliberalism and techno-utopianism, national policy in the United States has long stood clear of any meaningful regulation of the online sphere. Section 230 of the

7, 2022), https://www.newstatesman.com/the-culture-interview/2022/03/symeon-brown-on-the-exploitative-influencer-economy [https://perma.cc/5J89-6SDA].

147 See Jürgen Habermas, *Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere*, 39 THEORY CULTURE & SOC'Y, July 2022, at 145, 167 (contending that the emergence of Silicon Valley has been instrumental in the global spread of neoliberal economic policy).

148 Starr, *supra* note 14.

149 *See* Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L.J. 1460, 1492-96 (2020) (criticizing that techno-utopian view).

150 *See* ANU BRADFORD, DIGITAL EMPIRES: THE GLOBAL BATTLE TO REGULATE TECHNOLOGY 33–42 (2023) (describing the joinder of techno-utopianism and neoliberalism).

151 *See, e.g.*, David R. Johnson & David Post, *Law and Borders—The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367 (1996); A. Michael Froomkin, *The Internet as a Source of Regulatory Arbitrage*, *in* BORDERS IN CYBERSPACE: INFORMATION POLICY AND THE GLOBAL INFORMATION INFRASTRUCTURE (Brian Kahin & Charles Nesson eds., 1997).

152 BRADFORD, *supra* note 150, at 34.

153 Kapczynski, *supra* note 149, at 1466.

Communications Decency Act is the notable, but quite telling, exception to that regulatory silence. Section 230 removes legal obstacles to technology platforms' rapid growth by according them far-reaching immunity from liability for both user-posted content and their own content moderation policies, including, arguably, their algorithmic recommender systems.[154]

In so doing, the statute vests private parties—online platforms—with the authority and considerable leeway to regulate harmful online content as they see fit, rather than authorizing federal regulators to define and police such content.[155] As William Kennard, chairman of the Federal Communications Commission during the Clinton Administration proclaimed, the best approach to telecommunications policy is to allow the "marketplace to find business solutions . . . as an alternative to intervention by government."[156] The result has been the U.S.-based, global digital economy, featuring the rise of platform monopolies and "surveillance capitalism," in which users are manipulated to generate ever more personal data used to target advertising.[157]

In a sharp reversal, the "techlash" of recent years has brought numerous calls to regulate online platforms in the United States. Some proposed legislation aims to defend election integrity or other democratic institutions.[158] Other proposed and arguably related regulatory

154 47 U.S.C. § 230. The immunity does not extend to platform-authored content, intellectual property infringement, certain federal crimes, or content removal in bad faith. *Id. See generally* Alan Z. Rozenshtein, *Silicon Valley's Speech: Technology Giants and the Deregulatory First Amendment*, 1 J. FREE SPEECH L. 337, 338 (2021); Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986 (2008). Lower courts have ruled that platform recommender systems that channel otherwise illegal or tortious content do not constitute platform-authored speech and thus fall within § 230 immunity. *See* Gonzalez v. Google LLC, 2 F.4th 871, 892–93 (9th Cir. 2021), *vacated on other grounds*, 598 U.S. 631 (2023); Force v. Facebook, Inc., 934 F.3d 53, 57 (2d Cir. 2019).

155 BRADFORD, *supra* note 150, at 42–43 (describing how § 230 underpins the deregulatory architectural of today's digital economy in the United States).

156 Starr, *supra* note 14.

157 *See generally* ZUBOFF, *supra* note 45 (coining the term and detailing the workings of "surveillance capitalism"). On the role of U.S. law, particularly immunity from intermediary liability and weak data privacy protections for individuals, in underwriting the growth of Silicon Valley, see Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639 (2014).

158 *See, e.g.*, The Digital Platform Commission Act of 2023, S. 4201, 118th Cong. (2023) (proposing to establish an independent federal agency regulator for the technology sector in order to enhance competition, protect consumers, and promote civic discourse and democracy); Protecting Americans from Dangerous Algorithms Act, H.R. 8636, 116th Cong. (2020) (proposing to make social media platforms subject to civil liability for violations of federal civil rights and antiterrorism laws when the platform's recommender system amplifies such tortious content); Banning Microtargeted Political Ads Act, H.R. 7014, 116th Cong. (2020) (proposing to prohibit social media from carrying political ads targeted to users' individual characteristics other than place of residency); Bot Disclosure and Accountability Act of 2019, S. 2125, 116th Cong. (2019)

interventions are designed to protect individuals—particularly women and minorities——against exploitation, discrimination, and harassment.[159] Thus far, no proposed federal measures targeting social media harms to democracy have been enacted. On the other hand, the Department of Justice and the Federal Trade Commission have brought antitrust claims against Google and Facebook to promote competition in markets for digital search, communication, advertising, and other services.[160]

C. *First Amendment*

Even if U.S. lawmakers' were inclined to enact legislation to address social media harms, they would be severely constrained by neoliberal and classical liberal understandings of freedom of speech. Under the neoliberal model, social media firms' surveillance capitalist business model—their targeted amplification of content designed to maximize user engagement and amass user data for behavior advertising—would qualify as fully protected First Amendment speech. Further, the neoliberal model rests on the laissez-faire, classical liberal understanding that rights to free speech—whether they be held by social media platforms or users——must be protected against state interference even if the speech in question "is likely to poison the public sphere, create long-term civic unrest, provoke random acts of violence, slowly undermine democracy, or deter the spread and acceptance of truth."[161]

We consider the neoliberal and classical liberal models in turn.

(proposing to require that bots that impersonate or replicate human activity on social media be clearly identified as such and to prohibit the use of such bots in online political advertising.

159 *See, e.g.*, Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms Act, H.R. 3421, 117th Cong. (2021) (proposing to amend § 230 to allow social media companies to be civilly liable to enabling cyber-stalking, targeted harassment, and discrimination on their platforms).

160 *See* Herbert Hovenkamp, *Antitrust and Platform Monopoly*, 130 YALE L.J. 1952 (2021); *see also* Sara Morrison & Shirin Ghaffary, *The Case Against Big Tech*, VOX (Dec. 8, 2021, 5:30 AM), https://www.vox.com/recode/22822916/big-tech-antitrust-monopoly-regulation [https://perma.cc/H28K-VK8Z]; Daisuke Wakabayashi, *The Antitrust Case Against Big Tech, Shaped by Tech Industry Exiles*, N.Y. TIMES (June 28, 2021), https://www.nytimes.com/2020/12/20/technology/antitrust-case-google-facebook.html [https://perma.cc/YY77-CM2D].

161 Balkin, *supra* note 138, at 62–63.

1. The Neoliberal First Amendment Model

During the decades in which neoliberalism and techno-utopianism have informed U.S. technology policy, U.S. courts have increasingly interpreted First Amendment law in line with the neoliberal vision that equates markets with speech, elevates property owner rights over asserted speech rights of non–property owners, and posits that elections and other public institutions should come to resemble markets as much as possible.[162] In a series of decisions, the Supreme Court has ruled that advertising, campaign spending, and selling data constitute constitutionally protected speech. As Justice Kennedy wrote for the Court in *Sorrell v. IMS Health*, striking down a state's ban on providing doctors' prescription records to pharmaceutical companies for use in drug marketing: "The State may not burden the speech of others in order to tilt public debate in a preferred direction. 'The commercial marketplace, like other spheres of our social and cultural life, provides a forum where ideas and information flourish.'"[163]

Likewise, in *Citizens United v. FEC*, the Court struck down federal limitations on the use of corporate funds to support or oppose candidates for federal government office.[164] In so holding, the Court reiterated that campaign expenditures are speech and that the corporate nature of the speaker makes no difference to the analysis in free speech cases. More broadly, *Citizens United* also explicitly overruled Supreme Court precedent that held that government may combat the corrosive and distorting impact on public debate wrought by untrammeled corporate election spending that draws upon "immense aggregations of wealth that are accumulated with the help of the corporate form and that have little or no correlation to the public's support for the corporation's political ideas."[165] The *Citizens United* Court insisted that the goals of fostering political equality and preventing the distortion of public debate resulting

162 *See* Jedediah Purdy, *Neoliberal Constitutionalism: Lochnerism for a New Economy*, 77 LAW & CONTEMP. PROBS. no. 4, 2014, at 195, 198–203; *see also* JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 7 (2019) (explaining that "the neoliberal political orientation emphasizes not only market liberties but also a market-based approach to structuring political and social participation").

163 Sorrell v. IMS Health, Inc., 564 U.S. 552, 578–79 (2011) (quoting Edenfield v. Fane, 507 U.S. 761, 767 (1993)).

164 558 U.S. 310 (2010).

165 Austin v. Mich. Chamber of Com., 494 U.S. 652, 660 (1990), *overruled by Citizens United*, 558 U.S. at 365–66; *see also* Richard L. Hasen, Citizens United *and the Orphaned Antidistortion Rationale*, 27 GA. ST. U. L. REV. 989, 990 (2011) (faulting the Stevens dissent in *Citizens United* for failing to expressly defend corporate spending limits on political equality grounds).

from vastly unequal economic resources are "wholly foreign to the First Amendment."[166]

In sum, the neoliberal turn in First Amendment jurisprudence—what scholars and dissenting judges have derisively labeled "First Amendment Lochnerism"——obstructs economic regulation, measures designed to produce a more egalitarian distribution of social goods, and legislation that aims to further political equality by reducing the influence of wealth on the democratic process.[167] Of particular relevance to social media, courts have done so in large part by extending "the status of speech to what are essentially the commercial operations of firms in the information economy."[168] As such, First Amendment Lochnerism could well provide grounds for blocking a wide range of regulations aimed at countering social media harms to democracy.

Much like providing doctors' prescription records to pharmaceutical companies for drug marketing, social media's content curation practices overwhelmingly serve commercial ends, not that of disseminating the platforms' own expressive message. As we have seen, social media platforms' algorithmic recommender systems are designed to maximize user engagement, collect data on users' content preferences, and microtarget advertising based on data gleaned about users.[169] Social media community standards and terms of use do typically prohibit user postings of violent incitement, hate speech, misinformation, and other harmful and controversial content. And social media content moderation algorithms aim to block or demote such content.[170] Yet social media have hardly adopted and instituted content moderation to convey their own expressive message. Platforms, rather, took on content moderation reluctantly, as required to "present their best face to new users, to their

166 *Citizens United*, 558 U.S. at 349–50 (quoting Buckley v. Valeo, 424 U.S. 1, 48–49 (1976)).

167 *See, e.g.*, Genevieve Lakier, *The First Amendment's Real Lochner Problem*, 87 U. CHI. L. REV. 1241 (2020) (chronicling claims of First Amendment Lochnerism and arguing that current First Amendment doctrine repeats the errors of the Lochner Court not by protecting commercial speech but by relying upon an almost wholly negative notion of freedom of speech and by assuming that the only relevant constitutional interest at stake in free speech cases is the autonomy interest of the speaker); Nelson Tebbe, *A Democratic Political Economy for the First Amendment*, 105 CORNELL L. REV. 959, 960 (2020) (using the term and noting that it has been deployed by Supreme Court Justices Elena Kagan and Stephen Breyer in dissent); Amy Kapczynski, *The Lochnerized First Amendment and the FDA: Toward a More Democratic Political Economy*, 118 COLUM. L. REV. ONLINE 179, 181–82 (2018); Amanda Shanor, *The New* Lochner, 2016 WIS. L. REV. 133.

168 Yochai Benkler, *Through the Looking Glass: Alice and the Constitutional Foundations of the Public Domain*, 66 LAW & CONTEMP. PROBS., Winter/Spring 2003, at 173, 223.

169 *See* ZUBOFF, *supra* note 45; Balkin, *supra* note 138, at 39–43; Jeff Gary & Ashkan Soltani, *First Things First: Online Advertising Practices and Their Effects on Platform Speech*, KNIGHT FIRST AMEND. INST. (Aug. 21, 2019), https://knightcolumbia.org/content/first-things-first-online-advertising-practices-and-their-effects-on-platform-speech [https://perma.cc/VH3C-SS78].

170 *See* Thorburn, Bengani & Stray, *supra* note 52.

advertisers and partners, and to the public at large."[171] Like algorithmic recommender systems, social media content moderation is part and parcel of the platform's commercial operations, aimed at satisfying advertisers, promoting a reputation for brand safety, and avoiding government regulation.[172]

Put another way, content curation, content moderation, user affordances, and user data comprise social media firms' menu of products. In designing those commodities, social media firms must be attentive to consumer demand, product safety, and firm reputation, like any market entity. And it is overwhelmingly those market dictates, not any desire to convey an expressive message, that determines how social media shape the mix of content that appears on their platforms.

Nonetheless, as some commentators and lower courts have posited, predominant neoliberal First Amendment doctrine might well insist that social media recommender systems and content curation constitute fully protected platform speech.[173] In that view social media recommender and content curation algorithmic systems are akin to traditional news media's First Amendment–protected editorial discretion regarding which content to disseminate to the public. Neoliberal proponents apply that analogy even though, unlike legacy news media, social media platforms almost entirely fuel third-party speech rather than their own and treat users' posts

171 TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 5 (2018).

172 *See* Yi Liu, Pinar Yildirim & Z. John Zhang, Implications of Revenue Models and Technology for Content Moderation Strategies (Nov. 23, 2021) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3969938 [https://perma.cc/KYJ7-J8BF] (explaining how optimal content modification strategies differ for platforms with different revenue models); Leonardo Madio & Martin Quinn, *Content Moderation and Advertising in Social Media Platforms* (Univ. of Padova, Marco Fanno Working Paper No. 297, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3551103 [https://perma.cc/Z7B3-CY2B].

173 *See, e.g.*, NetChoice, LLC v. Att'y Gen., 34 F.4th 1196, 1210 (11th Cir. 2022) (explaining that when "[s]ocial media platforms . . . 'disclos[e],' 'publish[],' or 'disseminat[e]' information, they engage in 'speech within the meaning of the First Amendment'" (quoting Sorrell v. IMS Health Inc., 564 U.S. 552, 570 (2011))); O'Handley v. Padilla, 579 F. Supp. 3d 1163, 1186–87 (N.D. Cal. 2022) (holding that "[l]ike a newspaper or a news network," Twitter's decisions "about what content to include, exclude, moderate, filter, label, restrict, or promote . . . are protected by the First Amendment"); La'Tiejira v. Facebook, Inc., 272 F. Supp. 3d 981, 991 (S.D. Tex. 2017) (acknowledging "Facebook's First Amendment right to decide what to publish and what not to publish on its platform"); Jian Zhang v. Baidu.com Inc., 10 F. Supp. 3d 433, 440 (S.D.N.Y. 2014) (holding that suit "to hold [a website] liable for . . . a conscious decision to design its search-engine algorithms to favor certain expression on core political subjects over other expression on those same political subjects" would "violate[] the fundamental rule of protection under the First Amendment, that a speaker has the autonomy to choose the content of his own message" (alteration in original) (quoting Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., 515 U.S. 557, 573 (1995))); *see also* Eric Goldman, *The Constitutionality of Mandating Editorial Transparency*, 73 HASTINGS L.J. 1203, 1220 (2022); Eugene Volokh & Donald M. Falk, *Google: First Amendment Protections for Search Engine Search Results*, 8 J.L. ECON. & POL'Y 883, 886 (2012).

as instrumental data for spurring targeted advertising, not as speech that conveys the platform's own carefully curated message.[174]

2. Classical Liberal Model

Some commentators have argued that even if social media's content curation can be analogized to newspapers' editorial discretion, social media should enjoy more limited First Amendment protection, perhaps targeted at whether and to what extent platform recommender systems and content moderation contribute to democratic discourse.[175] However, ascendant neoliberal First Amendment jurisprudence does not merely redefine information economy commercial operations as speech. Rather, the neoliberal approach also expands upon a venerable tradition in First Amendment jurisprudence that manifests profound, overriding resistance to any state interference in public discourse. Under this classical liberal, laissez-faire model, the state must be prevented from interfering with free speech rights even when speech consists of manipulative propaganda, threatens serious social harm, and undermines liberal democracy.[176] Taken together, the neoliberal and classical liberal models thus raise a nigh insurmountable obstacle to any meaningful regulatory response to social media harms.

---

174 *See, e.g.*, *NetChoice*, 34 F.4th at 1218. *But see* Adam Candeub, *Editorial Decision-Making and the First Amendment*, 2 J. FREE SPEECH L. 157 (2022) (arguing that platform content moderation decisions are mostly non-expressive editorial decisions that do not enjoy First Amendment protection); Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1528 (2013) (arguing that Google is not like a newspaper that selects and endorses the articles that appear on its pages—Google's search engine merely "helps its users find websites, but it does not sponsor or publish those websites").

175 *See, e.g.*, Evelyn Douek & Genevieve Lakier, *Rereading "Editorial Discretion"*, KNIGHT FIRST AMEND. INST.: REREADING THE FIRST AMEND. (Oct. 24, 2022), https://knightcolumbia.org/blog/rereading-editorial-discretion [https://perma.cc/VTP4-PFWC] (arguing that the extent to which platforms enjoy First Amendment protection for content curation and moderation as editorial discretion akin to that of newspapers should depend on the extent to which that platform's practices contribute to democratic discourse); Ashutosh Bhagwat, *Do Platforms Have Editorial Rights?*, 1 J. FREE SPEECH L. 97 (2021) (arguing that platforms' algorithmic content curation and moderation amount to value-based choices regarding what third-party content to permit on their platforms and thus should enjoy First Amendment protection as editorial discretion, even if lesser protection than that accorded to newspapers); Jameel Jaffer & Scott Wilkens, *Social Media Companies Want to Co-opt the First Amendment. Courts Shouldn't Let Them.*, N.Y. TIMES (Dec. 9, 2021), https://www.nytimes.com/2021/12/09/opinion/social-media-first-amendment.html (highlighting the differences between social media and legacy media and arguing that when social media platforms add their own voices to public discourse, such as when they attach warning labels to users' posts, they are exercising a kind of editorial discretion that should be protected by the First Amendment).

176 Balkin, *supra* note 138, at 62–63.

The classical liberal model hearkens back to the foundational free speech dissents of Justices Holmes and Brandeis of the early twentieth century. As Justice Holmes famously proclaimed: "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."[177] Less fatalistically, Justice Brandeis expressed faith that, in the end, the benefits of freedom of speech for self-government and individual liberty will prevail over possible harms:

> To courageous, selfreliant men, with confidence in the power of free and fearless reasoning applied through the processes of popular government, no danger flowing from speech can be deemed clear and present, unless the incidence of the evil apprehended is so imminent that it may befall before there is opportunity for full discussion. 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.[178]

That view, initially in dissent, eventually came to dominate First Amendment doctrine, reaching fruition with civil liberties rulings of the Warren Court and continuing in recent decades.[179] Sharply discordant with European militant democracy understandings, the Supreme Court has held to be protected First Amendment speech: a march by neo-Nazis in a town populated by Holocaust survivors,[180] a protest against the military's tolerance of homosexuality at the funeral of a veteran who was killed in the line of duty,[181] a public official's blatant lie that he had been awarded the Congressional Medal of Honor for exceptional valor in the military,[182] violent video games marketed to minors,[183] and racially derogatory trademarks.[184] As the Supreme Court has insisted with regard to hate speech: "Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express 'the thought that we hate.'"[185] Highlighting that

177 Gitlow v. New York, 268 U.S. 652, 673 (1925) (Holmes, J., dissenting).

178 Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring).

179 *See* Balkin, *supra* note 138, at 64–65.

180 Nat'l Socialist Party of Am. v. Vill. of Skokie, 432 U.S. 43 (1977).

181 Snyder v. Phelps, 562 U.S. 443 (2011).

182 United States v. Alvarez, 567 U.S. 709 (2012) (where defendant falsely claimed receiving Congressional Medal of Honor, in violation of Stolen Valor Act of 2005).

183 Brown v. Ent. Merchs. Ass'n, 564 U.S. 786 (2011).

184 Matal v. Tam, 582 U.S. 218 (2017).

185 *Id.* at 246 (quoting United States v. Schwimmer, 279 U.S. 644, 655 (1929) (Holmes, J., dissenting)).

statement, a lower court has struck down as a First Amendment violation even a state law requiring social media platforms to devise some policy—completely of their own choosing—for responding to user complaints about hate speech and informing users of that policy on their website.[186]

To be certain, the Supreme Court has long emphasized that the First Amendment serves as the "guardian of our democracy."[187] Much of First Amendment doctrine, indeed, purports to be grounded in the political imperatives of democratic self-government.[188] As the Court reiterated in *Citizens United*: "Speech is an essential mechanism of democracy, for it is the means to hold officials accountable to the people. The right of citizens to inquire, to hear, to speak, and to use information to reach consensus is a precondition to enlightened self-government and a necessary means to protect it."[189] As such, freedom of speech is essential both to the democratic process——serving as the "foundation of free government by free men"—and to disseminating the information, ideas, and knowledge that the people need to govern.[190]

Yet even the Court's repeated reference to democratic self-government as a central purpose and justification for freedom of speech carries the imprint of a neoliberal faith in unhindered markets and longstanding, overriding resistance to government regulation. As Erin Miller has underscored, "[n]o notion has a firmer grip on First Amendment doctrine than the marketplace of ideas," a metaphor that "analogizes public discourse to an economic marketplace."[191] And in repeatedly incantating the marketplace metaphor, the Court has embraced a decidedly laissez-faire vision, evincing a profound distrust of virtually

186 Volokh v. James, --F.Supp.3d--, No. 22-CV-10195, 2023 WL 1991435 (S.D.N.Y. Feb. 14, 2023).

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

188 *See generally* Robert Post, *Participatory Democracy and Free Speech*, 97 VA. L. REV. 477 (2011); James Weinstein, *Participatory Democracy as the Central Value of American Free Speech Doctrine*, 97 VA. L. REV. 491 (2011); Erin L. Miller, *Amplified Speech*, 43 CARDOZO L. REV. 1, 27–29 (2021).

189 Citizens United v. FEC, 558 U.S. 310, 339 (2010) (citations omitted).

190 Schneider v. New Jersey, 308 U.S. 147, 161 (1939) (freedom of speech and of the press "reflects the belief of the framers of the Constitution that exercise of the rights lies at the foundation of free government by free men").

191 Miller, *supra* note 188, at 30–31; *see also* Claudia E. Haupt, *Regulating Speech Online: Free Speech Values in Constitutional Frames*, 99 WASH. U. L. REV. 751, 754, 756, 781–83 (2012) (arguing that American free speech scholarship has "canonized" the marketplace of ideas rationale and concomitantly marginalized the view that freedom of speech must serve the needs of democratic governance (quoting William P. Marshall, *In Defense of the Search for Truth as a First Amendment Justification*, 30 GA. L. REV. 1, 1 (1995))).

any government intervention to foster egalitarian participation in public discourse, reasoned debate, and trustworthy information.[192]

In that view, just as government is the great nemesis of efficient markets for goods and services, so does the marketplace of ideas best function to produce truth and to underwrite the democratic process when government is banned from intervening in public discourse. Predominant First Amendment doctrine looks overwhelmingly to the private sector and free market to provide opportunities for speech, both to foster individual autonomy and to underwrite democratic self-governance. The state, by contrast, is the actor that may not interfere.[193]

3. Sum

To be certain, there are very good reasons to be wary of government regulation of speech.[194] The state has unparalleled capacity to suppress speech. Government officials face the ever-present temptation to censor the speech of their political opponents. Democratic governments might also serve as a vehicle for majorities to silence nonconforming views that challenge the dominant consensus.

Yet, some carefully measured and targeted government intervention is vital to fostering the pluralist, reasoned, fact-based public discourse upon which democratic governance depends.[195] The laissez-faire marketplace of ideas is highly vulnerable to manipulation and distortion. Those with substantially greater resources have the power to drown out

192 *See* Miller, *supra* note 188, at 31-33. Indeed, the marketplace of ideas metaphor envisions an idealized realm of speech that is presumptively immune from government regulation to a far greater extent than in the economic sphere. *See* Kathleen M. Sullivan, *Free Speech and Unfree Markets*, 42 UCLA L. REV. 949, 949-51 (1995).

193 Morgan N. Weiland, *Expanding the Periphery and Threatening the Core: The Ascendant Libertarian Speech Tradition*, 69 STAN. L. REV. 1389, 1407 (2017).

194 *See* David A. Strauss, *Social Media and First Amendment Fault Lines*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY, *supra* note 12, at 4.

195 *See* Miller, *supra* note 188, at 36–38 (advocating government intervention to ensure epistemic competition among diverse and antagonistic speakers). Indeed, as John Fabian Witt has elucidated, even as Justices Holmes and Brandeis first gave voice to the proposition that the marketplace of ideas will yield truth and safeguard democracy, leading progressive thinkers expressed grave concern that, while free speech might serve as an indispensable foundation for democratic self-government, unrestricted communication can also be readily employed to subvert democracy. John Fabian Witt, Weaponized from the Beginning (Sept. 19, 2022) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4208158 [https://perma.cc/9EQ9-JH6D]; *cf.* Bazelon, supra note 107, at 49 (calling for government investment to increase the delivery of reliable information even if First Amendment doctrine prohibits speech regulation for that purpose).

less well-funded voices.[196] Hate speech, trolling, and violent incitement intimidate and silence vulnerable targets. Bold-faced lies can be every bit as convincing as accurate statements of fact.[197] Bombastic propaganda drowns out the basic civility norms and epistemic understandings upon which democracy rests.

As such, social media burst open the fault lines on which classical liberal First Amendment doctrine has always rested.[198] They fuel the viral proliferation of falsehood, emotive propaganda, and virulent hatred, and do so at an unprecedented rapidity and scale that belie the notion that evil speech can simply be proven wrong by education and counterspeech. As we have seen, social media enables bad actors strategically—and effectively——to flood our digital public sphere with disinformation, conspiracy theories, and emotionally manipulative outrage designed to instill epistemic fog and maximize outgroup animosity. Indeed, platform algorithms actively funnel divisive, corrosive content to those most likely to be receptive to it.

Under those conditions, manipulation and distortion repeatedly drown out fact-based, reasoned discourse. Our public sphere becomes a space where the speech that prevails is that which is best able to seize our scarce attention through repetition, virality, and by triggering anger, ridicule, out-group hostility, a craving to feel in control, and other potent emotions.[199] As Robert Post bemoans, it is a public sphere governed by a kind of mob mentality, characterized by tribalism and self-reinforcing gossip, rather than public reflection on information filtered through commonly respected epistemological authorities.[200]

In that space, it is increasingly apparent that courts' repeated invocation of the archetypical marketplace of ideas as a reasoned, educative forum for democratic self-governance rests on an entirely unrealistic assessment of the digital public sphere. It expresses

196 As Frederick Schauer reminds us, "the marketplace of ideas is less metaphor than description, and . . . in the marketplace of ideas, like in most other markets, one can compete much more successfully if one has greater resources." Frederick Schauer, *The Political Incidence of the Free Speech Principle*, 64 U. COLO. L. REV. 935, 949 (1993).

197 Remarkably, as Frederick Schauer rightly notes, "we have . . . arrived at a point in history in which an extremely important social issue about the proliferation of demonstrable factual falsity in public debate is one as to which the venerable and inspiring history of freedom of expression has virtually nothing to say." Frederick Schauer, *Facts and the First Amendment*, 57 UCLA L. REV. 897, 908 (2010)

198 *See generally* Strauss, *supra* note 194 (discussing First Amendment fault lines and how social media requires us to confront them).

199 *See* Balkin, *supra* note 138, at 46–49 (contrasting the archetypical marketplace of ideas with the digital age's ecology of memes).

200 *See* Robert Post, *Democracy and the Internet*, BALKINIZATION (Jan. 28, 2023), https://balkin.blogspot.com/2023/01/democracy-and-internet.html [https://perma.cc/S2ZC-PXGN].

anachronistic blind faith in the traditional laissez-faire First Amendment model, with undercurrents of the neoliberal belief that untrammeled market competition, not reasoned democratic discourse, yields socially useful knowledge and optimal human welfare.[201]

D. *First Amendment and Social Media Regulation*

Given the dearth of federal legislation designed to address social media harms to democracy, the Supreme Court has yet to weigh in specifically on how First Amendment law might apply to such measures. Indeed, as of this writing, it remains uncertain whether the law views social media platforms as (1) speakers with capacious First Amendment protection for their "editorial" decisions, (2) speakers with First Amendment protection of somewhat more limited scope than that accorded to newspapers, (3) common carriers that are forbidden to "censor" their users' expression based on "viewpoint," or (4) something else.[202]

However, as of this writing, the Supreme Court has just granted certiorari to resolve a circuit court split regarding the constitutionality of state statutes that impose quasi-common carrier obligations on major platforms. The Eleventh Circuit held that when social media platforms remove or deprioritize users or posts in accordance with platform content moderation policies, they are making editorial decisions that are subject to First Amendment protection akin to that accorded to newspapers.[203] It did so in striking down a Florida statute that provides that large social media companies must employ content moderation, including "censorship, deplatforming, and shadow banning," in "a consistent manner," meaning that the companies may not target particular viewpoints, political candidates, or news media.[204] But the Eleventh Circuit's rationales might apply equally to government regulations that do just the opposite, i.e., requiring platforms to cease amplifying disinformation, political manipulation, and hate speech.

---

201 *See* Metcalf, *supra* note 144 (discussing Hayek's grand epistemological claim that that market is superior to public discourse to revealing useful knowledge).

202 *See infra* notes 173-175 and accompanying text and *supra* notes 203-206 and accompanying text.

203 NetChoice, LLC v. Att'y Gen., 34 F.4th 1196, 1210–14 (11th Cir. 2022), *cert. granted in part sub nom.* Moody v. Netchoice, LLC, No. 22-277, 2023 WL 6319654 (Sept. 29, 2023).

204 The statute also generally forbids deplatforming political candidates or "journalistic enterprises" or blocking or demoting their posts. FLA. STAT. § 501.2041(2)(j) (2023) (journalistic enterprises); *id.* § 106.072(2) (2023) (deplatform candidate); *id.* § 501.2041(2)(h) (2023) (posts by or about candidate).

By contrast, the Fifth Circuit upheld a Texas statute that effectively treats social media platforms as common carriers by barring them from blocking, removing, or demonetizing user content based on the user's views.[205] As the court stated: "[W]e reject the idea that corporations have a freewheeling First Amendment right to censor what people say."[206] While the two appellate courts disagree about whether social media platforms or their users should be treated as the First Amendment-protected speakers, the common rationale underlying both approaches is that such speakers must enjoy wide latitude to propagate online whatever speech they want, whatever the harms to democratic self-government.

American commentators writing about social media regulation and free speech operate within that framework. In a recent article, for example, leading First Amendment scholar, Jack Balkin, presents a thoughtful, careful analysis of "how to regulate (and not regulate) social media."[207] Balkin understands the American free speech principle as one that "allows the state to impose only a very limited set of civility norms on public discourse."[208] In lieu of state regulation, he explains, the free speech principle leaves a diverse array of intermediate institutions, including the press and social media, to impose their own civility norms, some of which may be stricter than the minima that the state is entitled to impose, but some which might not.[209] As such, Balkin posits, freedom of speech serves democratic values because it helps to enable (although, Balkin points out, it does not guarantee) individuals' participation in forming public opinion, state responsiveness to public opinion, and an informed public.[210]

Balkin admonishes, however, that freedom of speech serves democratic values only if intermediate institutions are generally trustworthy and trusted by the public. As Balkin puts it:

> Without these trusted institutions . . . the practices of free expression become a rhetorical war of all against all. Such a war undermines the values of political democracy, cultural democracy, and the growth and spread of knowledge that free expression is supposed to serve.

205 NetChoice, L.L.C. v. Paxton, 49 F.4th 439, 445 (5th Cir. 2022), *cert. granted in part sub nom.* Netchoice, LLC v. Paxton, No. 22-555, 2023 WL 6319650 (Sept. 29, 2023).

206 *Id.*

207 Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media*, 1 J. FREE SPEECH L. 71 (2021).

208 *Id.* at 76.

209 *Id.* at 76–77. Of course, if the Supreme Court upholds the recent Fifth Circuit ruling that states are entitled to regulate social media as quasi-common carriers, social media platforms may be barred from imposing meaningful civility norms and prohibitions on blatant falsehood and deception. *See* NetChoice, L.L.C. v. Paxton, 49 F.4th at 445.

210 *Id.* at 77–78.

> Protection of the formal right to speak is necessary to a well-functioning public sphere. It is just not sufficient.[211]

Balkin then acknowledges that our current digital public sphere, abound in amped-up disinformation, tribalism, and outrage, exemplifies that rhetorical war of all against all. As he notes:

> We have moved into a new kind of public sphere—a digital public sphere—without the connective tissue of the kinds of institutions necessary to safeguard the underlying values of free speech. We lack trusted digital institutions guided by public-regarding professional norms. . . .
>
> . . . Never has access to the means of communication been so inexpensive and so widely distributed. But without the connective tissue of trusted and trustworthy intermediate institutions guided by professional and public-regarding norms, the values that freedom of speech is designed to serve are increasingly at risk. Antagonistic sources of information do not serve the values of free expression when people don't trust anyone and professional norms dissolve.[212]

In short, Balkin recognizes that untrammeled freedom of speech——a public sphere that is not actively filtered through trustworthy and trusted professional gateway institutions such as the traditional press——is deleterious to the democratic values that freedom of speech is ideally supposed to serve. Yet, to avoid violating "free speech values or the First Amendment,"[213] he proffers policy levers that only tangentially—and, I would argue, ineffectually—address the problem. He aims first to increase the number and diversity of social media platforms through antitrust law.[214] But even if network effects would not soon lead to a reconstituted Facebook or something like it, it is hard to see how a balkanized public sphere of multiple, largely coequal social media platforms and unregulated content moderation regimes, including the likes of Gab, Gettr, Parler, Rumble, and Donald Trump's Truth Social, would end tribalism and engender constitutive democratic dialogue.[215] Balkin also proposes giving social media companies incentives to take responsibility for the health of the public sphere. To that end, he would make them fiduciaries for their users' personal data, impose on them

211 *Id.* at 79.

212 *Id.*

213 *Id.* at 90.

214 *Id.* at 91–92.

215 *See* Daniel Karell, Andrew Linke, Edward Holland & Edward Hendrickson, *"Born for a Storm": Hard-Right Social Media and Civil Unrest*, 88 AM. SOCIO. REV. 322 (2023) (concluding that hard-right social media platforms that are insulated from both opposing views and legacy media gatekeepers further radicalize participants, shift users' perceptions of social norms, and, consequently, lead to greater civil unrest).

liability for defamatory advertising and privacy violations like revenge porn, and encourage them to hire more content moderators.[216]

Assuming that social media platforms are speakers rather than common carriers, Balkin's proffered policy levers comfortably comport with current First Amendment doctrine and his understanding of American free speech values, which abhor state interference in public discourse. For that reason, Balkin's proposals lack the teeth and targeted impact of the European regulations discussed below.

For her part, Daphne Keller, another leading commentator on free speech and social media regulation, is forthright about the contrast between absolutist First Amendment doctrine and "European legal cultures' greater willingness to trust regulators and greater tolerance for restrictions on expression."[217] Like Balkin, she readily recognizes that social media algorithmic recommender systems that push sensationalist, extremist content cause serious social harm, even if as Keller rightly notes, "electoral disinformation and other harmful untruths [also] go dangerously viral on platforms like WhatsApp, which has no ads and no platform-initiated ranking" of content.[218]

Yet Keller also concedes that the First Amendment imposes significant constraints on social media speech regulation. As she puts it: "Congress cannot go too far in requiring or incentivizing platforms to take down legal speech."[219] Indeed, Keller adds, "[t]he same constitutional limits apply if Congress wants platforms to demote or cease amplifying that speech."[220] As Keller notes, in *United States v. Playboy*, a case in which the U.S. Supreme Court struck down requirements for cable operators to limit access to Playboy's pornographic content, the Court stated unequivocally that "[t]he Government's content-based burdens must satisfy the same rigorous scrutiny as its content-based bans."[221] Accordingly, like Balkin, Keller turns to user privacy rights, consumer protection, and antitrust law, with the hope that increased competition and user empowerment will "open up space to argue for healthier intellectual fare as a matter of user autonomy, rather than as top-down restriction on speech and information."[222]

For his part, Eric Goldman has argued that even transparency requirements—regulation that compels internet services to disclose

---

216 Balkin, *supra* note 207 at 92–96.

217 Daphne Keller, *Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard*, 1 J. FREE SPEECH L. 227, 251 (2021).

218 *Id.* at 264.

219 *Id.* at 271.

220 *Id.* at 271.

221 *Id.* at 237–38 (quoting United States v. Playboy Ent. Grp., Inc., 529 U.S. 803, 812 (2000)).

222 Keller, *supra* note 217, at 265.

information about their recommender systems, content curation policies, and content moderation practices—run afoul of the First Amendment.[223] According to Goldman, it is axiomatic that internet services' decisions about what content to include, exclude, filter, label, restrict, or promote are First Amendment-protected "editorial decisions," no less than those of a newspaper or news network.[224] Goldman contends further that regulators' attempts to confirm the accuracy of internet services' disclosures would motivate services to alter their policies in order to please regulators—thus having the same effect on speech as more direct, and obviously unconstitutional, speech regulations.[225] Goldman concludes, therefore, that, at least in the United States, such mandatory transparency regulations are "another policy dead-end in regulators' quest to control online speech."[226] As we shall see, the type of transparency and reporting requirements that, Goldman explains, might well violate the First Amendment lie at the core of the EU regulation of very large online platforms and search engines. And, like Balkin and Keller, Goldman proffers First Amendment–compliant—and, I would argue, fatally feeble—alternatives, including, in his case, private auditing, funding for research into content moderation, and digital citizenship education.[227]

Finally, the EU transparency and reporting requirements that, following Goldman's argument, might fail to meet First Amendment muster if enacted in the United States are but part of a multifaceted EU approach that applies formal and informal pressure, including the threat of direct regulatory intervention in the future, to induce platforms to act "voluntarily" to mitigate systemic democracy-harming risks on their systems.[228] The question of whether and when government officials run

223 Eric Goldman, *The Constitutionality of Mandating Editorial Transparency*, 73 HASTINGS L.J. 1203, 1219–21 (2022). Goldman cites the "social media censorship" laws enacted by Florida and Texas as classic examples of this type of regulation. *Id.* at 1209–12.

224 Goldman, *supra* note 223, at 1220 (citing O'Handley v. Padilla, 579 F. Supp. 3d 1163, 1186–87 (N.D. Cal. 2022)).

225 Keller likewise argues that transparency requirements raise First Amendment concerns because they might impact platforms' ability to implement their "editorial policies" and, thus, internet users' ability to seek and impart information online. She notes that disclosure rules targeted at types of data that have public importance to the functioning of democracy might or might not survive First Amendment scrutiny depending on evolving case law and how the transparency requirements are structured. Daphne Keller, Platform Transparency and the First Amendment (Mar. 7, 2023) (unpublished manuscript), https://ssrn.com/abstract=4377578 [https://perma.cc/PQU9-T49T].

226 Goldman, *supra* note 223, at 1206.

227 *Id.* at 1232.

228 *See* Tarlach McGonagle, *Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation*, *in* OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY 467, 481–85 (Giancarlo Frosio, ed., 2020) (noting that the EU's ostensibly voluntary codes of

afoul of the First Amendment when they seek to convince, induce, or assist social media to cease amplifying harmful speech is a matter of judicial uncertainty and scholarly debate in the United States.[229] Scholars generally agree that public officials' mere encouragement of platforms to remove disinformation does not generally rise to the level of more heavy-handed, impermissible jawboning.[230] Yet, the Supreme Court has recently granted a writ of certiorari to consider whether Biden administration officials crossed the First Amendment line in urging the major social media platforms to remove or demote misinformation on a variety of pressing issues and upbraiding the platforms when they failed to do so. It is anybody's guess how the Court will rule on which government officials' uses of their bully pulpit exceed mere encouragement and rise to the level of impermissible jawboning.[231] In any event, jawboning of the type contemplated under the EU regulation, in which regulators put considerable pressure on social media to mitigate democracy-harming risks and aggressively monitor social media compliance, would almost certainly run afoul of First Amendment strictures.

E. *Coda*

In opposition to the neoliberal and classical liberal models of free speech jurisprudence, there is a rich, venerable tradition of First Amendment thought that grounds freedom of speech in the collective

platform conduct regarding terrorist content, hate speech, and disinformation have a coercive undertone and have been adopted under the threat of more pervasive direct regulation); *see also* Rachel Griffin & Carl Vander Maelen, Codes of Conduct in the Digital Services Act: Exploring the Opportunities and Challenges (May 30, 2023) (unpublished manuscript), https://ssrn.com/abstract=4463874 [https://perma.cc/F4TN-8UBD] (characterizing the putatively voluntary codes of conduct as "de facto legal obligations").

229 *See* Genevieve Lakier, *Informal Government Coercion and The Problem of "Jawboning*", LAWFARE, (July 26, 2021, 3:52 PM), https://www.lawfaremedia.org/article/informal-government-coercion-and-problem-jawboning [https://perma.cc/2THF-D267].

230 *See, e.g.*, Leah Litman & Laurence H. Tribe, *Restricting the Government from Speaking to Tech Companies Will Spread Disinformation and Harm Democracy*, JUST SEC. (July 5, 2023), https://www.justsecurity.org/87155/restricting-the-government-from-speaking-to-tech-companies-will-spread-disinformation-and-harm-democracy [https://perma.cc/PUA4-MV8R]. *But see* Derek E. Bambauer, *Against Jawboning*, 100 MINN. L. REV. 51 (2015) (arguing that jawboning that targets platforms over information they carry is normatively illegitimate, even if the Supreme Court has been inconsistent in defining which types of government pressure raise First Amendment concerns).

231 Murthy v. Missouri, No. 23-411, 2023 WL 6935337 (U.S. Oct. 20, 2023).

dictates of democracy.[232] That tradition focuses on the need for credible information, a vital watchdog press, a knowledgeable electorate, egalitarian participation in public discourse, and reasoned debate as linchpins for functioning democratic governance. Scholars have called for a revival of a free speech tradition centered in constitutive democratic values rather than individual autonomy and antiregulatory economics to counteract the neoliberal, laissez-faire understanding that has come to dominate First Amendment doctrine.[233]

The constitutional framework of militant democracy that arose in postwar Europe has received no express recognition in U.S. First Amendment jurisprudence. Yet, early to mid-twentieth century U.S. case law repeatedly voiced militant democracy style arguments for blocking speech that was understood to pose a clear, palpable danger to the democratic order.[234] Most pointedly, in an often cited passage, Justice Robert Jackson admonished that government complacence in the face of antidemocratic mass demonstrations and incitement to racial violence would convert' the "Bill of Rights into a suicide pact," a warning he backed with references to the Nazis' exploitation of democratic liberties to overthrow democracy.[235]

Today, militant democracy might serve as a regulatory ideal for a reimagined First Amendment doctrine that views public discourse fundamentally as a critical instrument for underwriting constitutive democratic values and sees the democratic state as a potential guarantor of the robustness and diversity of public debate.[236] Certainly, militant democracy's support for preventing the undue influence of wealth and power on democratic debate would resonate with Justice Stevens's

232 *See, e.g.*, OWEN M. FISS, THE IRONY OF FREE SPEECH (1996); CASS R. SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREEDOM OF SPEECH (1993); ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT (1948).

233 *See, e.g.*, Haupt, *supra* note 191, at 782–83 (favoring an understanding of the First Amendment that foregrounds the role of speech in democratic self-government and on participants in public discourse); Genevieve Lakier, *The Non–First Amendment Law of Freedom of Speech*, 134 HARV. L. REV. 2299, 2342–70 (2021) (describing the "pluralistic free speech tradition"); Lakier, *supra* note 167; Nelson Tebbe, *A Democratic Political Economy for the First Amendment*, 105 CORNELL L. REV. 959 (2020); Jeremy K. Kessler & David E. Pozen, *The Search for an Egalitarian First Amendment*, 118 COLUM. L. REV. 1953 (2018).

234 *See* Thomas M. Keck, *Erosion, Backsliding, or Abuse: Three Metaphors for Democratic Decline*, LAW & SOC. INQUIRY 314, 334–35 (2023).

235 Terminiello v. City of Chicago, 337 U.S. 1, 13–37 (1949) (Jackson, J., dissenting). Justice Jackson had previously served as a prosecutor of Nazi war criminals at Nuremberg. Keck, *supra* note 234, at 334.

236 Thomas Keck argues, for example, that the current "hyper-libertarian conception of the First Amendment is a barrier to multiple, common-sense democracy preservation reforms" and that "European-style militant democracy arguments" might serve as a source of inspiration for the consequent need for a "broad rethinking of contemporary First Amendment doctrine." Keck, *supra* note 234, at 332.

insistence in his dissenting opinion in *Citizens United* that campaign finance restrictions actually serve First Amendment values by reducing the risk that corporations will "distort public debate" and stymie the search for truth by "cow[ing]" politicians "into silence," "drowning out . . . noncorporate voices," and "diminish[ing] citizens' willingness and capacity to participate in the democratic process."[237] In that understanding, the First Amendment would also enable the democratic state to counter the weaponization of online "cheap speech"——whether at the hands of private actors, political organizations, public officials, foreign governments, or social media platforms driven by surveillance capitalism——to undermine public discourse and democratic institutions, even as it would continue to guard against abuses of the state's censorial power. In particular, as suggested by some American commentators and the European Union's regulatory framework detailed below, a First Amendment that draws upon militant democracy principles might enable the state to prod social media to act as trusted intermediaries for reason-based public discourse and to condition First Amendment protection for their content curation practices on whether those practices subvert or contribute to democratic self-government.

Finally, even under the predominant neoliberal First Amendment framework, European online platform regulation informed by militant democracy will likely color social media design, content moderation, and recommender system practice in the United States, not just in the European Union. As discussed above, as the European Union asserts global leadership in regulating social media and other online platforms, we would expect to see a "Brussels Effect," in which platforms conform their operation to EU requirements throughout the world.[238] In that regard, U.S. social media users would have no valid First Amendment objection if platforms, as private corporations, tighten their content moderation policies in the United States in complying with European speech restrictions.[239]

With those considerations in mind, we now turn to explore militant democracy and the European online platform regulation that it has engendered.

---

237 Citizens United v. FEC, 558 U.S. 310, 469–72 (Stevens, J., concurring in part and dissenting in part).

238 See *supra* notes 26–28 and accompanying text.

239 *See* Keller, *supra* note 27, at 8–10. Pursuant to the U.S. "state-action doctrine," "the Free Speech Clause [of the First Amendment] prohibits only *governmental* abridgment of speech. The Free Speech Clause does not prohibit *private* abridgment of speech." Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1928 (2019).

III. MILITANT DEMOCRACY

A. *What is Militant Democracy?*

1. Basic Tenets

Militant democracy hearkens back to John Locke's defense of popular resistance to would-be tyrants. As Locke argued, "Men can never be secure from Tyranny, if there be no means to escape it, till they are perfectly under it."[240] Individuals thus "have not only a right to get out of [tyranny], but to prevent it."[241]

Yet militant democracy came to the fore as a core principle of constitutional law and international human rights only following World War II. As those who drafted postwar constitutions and human rights instruments were keenly aware, Hitler rose to power by exploiting the political liberties and structural vulnerabilities of the Weimar Republic's constitutional democracy.[242] As Joseph Goebbels later taunted: "It will always be one of the best jokes of democracy . . . that it gave its deadly enemies the means to destroy it."[243]

Militant democracy posits that democracies should not have "to wait until the very moment a totalitarian take-over is imminent . . . to protect themselves."[244] Rather, democracies must be able "to act preemptively against" forces bent on their destruction.[245] Underlying militant democracy is the idea that, at a minimum, any true democracy must rest on an enduring, indissoluble system of periodic, fair, free, multiparty elections. A democracy cannot countenance its own destruction, even at the hands of a political party that attains power through an election or, for that matter, by the people in general.[246] Rather, the need to maintain and protect an enduring representative democracy must trump any purported political authority to overthrow democracy at any given time.[247]

---

240 ALEXANDER S. KIRSHNER, A THEORY OF MILITANT DEMOCRACY: THE ETHICS OF COMBATTING POLITICAL EXTREMISM 8 (2014) (quoting JOHN LOCKE, TWO TREATISES OF GOVERNMENT 411 (Peter Laslett ed., Cambridge Univ. Press 1988) (1689)).

241 *Id.* (quoting LOCKE, *supra* note 240, at 411).

242 For a brief historical account, see Gregory H. Fox & Georg Nolte, *Intolerant Democracies*, 36 HARV. INT'L L.J. 1, 10–11 (1995).

243 KIRSHNER, *supra* note 240, at 23.

244 Fox & Nolte, *supra* note 242, at 41.

245 *Id.*

246 Gregory H. Fox & Georg Nolte, Response, *Fox and Nolte Response*, 37 HARV. INT'L L.J. 231, 238–39 (1996).

247 *See* Udi Greenberg, *Militant Democracy and Human Rights*, 42 NEW GERMAN CRITIQUE 169, 175–79 (2015) (describing the debate among legal scholars in Weimar Germany about whether

As such, militant democracy is often narrowly portrayed as the right of a democratic state to target the enemies of democracy and, in particular, to suppress antidemocratic political parties and subversive political movements.[248] Yet, a more robust conception of militant democracy emphasizes that the democratic state is not merely the site for formal elections. Rather, the state stands as active guardian of the integrity and vitality of the democratic process, including the polity's commitment to democratic equality.[249] In this understanding, each person has a fundamental right of democratic participation and "all citizens are to be regarded as equal partners in . . . governing the polity in which they live."[250] Further, representative democracy is the source and, ultimately, the sole guarantor of fundamental human rights.[251] And given that fundamental rights flow from the democratic state, they must be defined and enforced in basic conformity with the democratic state's norms and principles.

2. Militant Democracy Versus Authoritarian Propaganda

Karl Loewenstein crafted the pillars of militant democracy in 1937, some three years after Hitler came to power. For Loewenstein, the legitimacy of democratic government rests on pragmatic, rational administration and appeal to human reason.[252] By contrast, totalitarianism is rooted in mobilizing emotionalism, typically grounded in high-pitched nationalism, hatred, tribalism, and devotion to a charismatic leader.[253]

---

the German people had the authority to replace the republic with a different kind of regime and Karl Lowenstein's insistence that representative democracy trumped such a right).

248 Somewhat more capaciously, Svetlana Tyulkina conceives of militant democracy as a principle that supports democratic states in safeguarding democracy against threats not only of antidemocratic political parties, but also of other antidemocratic actors, including terrorists and intolerant, totalitarian political ideologies and movements. Svetlana Tyulkina, *Militant Democracy as an Inherent Democratic Quality*, *in* MILITANT DEMOCRACY AND ITS CRITICS, *supra* note 17, at 207, 207.

249 Anthoula Malkopoulou & Ludvig Norman, *Three Models of Democratic Self-Defence: Militant Democracy and its Alternatives*, 66 POL. STUD. 442, 449 (2018); Issacharoff, *supra* note 18, at 1414.

250 *See* Stone, *supra* note 17, at 40–42; *see also* Malkopoulou & Norman, *supra* note 249, at 450-55 (advocating a social democratic model of democratic self-defense).

251 *See* Greenberg, *supra* note 247, at 177–78 (describing Karl Lowenstein's views).

252 Karl Loewenstein, *Militant Democracy and Fundamental Rights, I*, 31 AM. POL. SCI. REV. 417, 417–18 (1937).

253 *See* András Sajó, *Militant Democracy and Emotional Politics*, 19 CONSTELLATIONS 562, 571 (2012) (suggesting that Lowenstein's criticism of emotional politics related specifically to fascists' mobilization of hatred, fear, and the desire for identification with a charismatic leader).

Loewenstein focused particularly on Nazi propaganda as an instrument in mobilizing emotionalism and subverting democracy. He warned:

> Perhaps the thorniest problem of democratic states still upholding fundamental rights is that of curbing the freedom of public opinion, speech, and press in order to check the unlawful use thereof by revolutionary and subversive propaganda, when attack presents itself in the guise of lawful political criticism of existing institutions. Overt acts of incitement to armed sedition can easily be squashed, but the vast armory of fascist technique includes the more subtle weapons of vilifying, defaming, slandering, and last but not least, ridiculing, the democratic state itself, its political institutions and leading personalities.[254]

The Nazis viewed propaganda as a central tool for seizing and consolidating power. Hitler devoted two chapters in *Mein Kampf* to the importance of propaganda.[255] As such, Hitler candidly embraced emotionalism as key to attaining and maintaining political power: "The art of propaganda lies in understanding the emotional ideas of the great masses and finding, through a psychologically correct form, the way to the attention and thence to the heart of the broad masses."[256] To reach the masses, Hitler understood, propaganda had to be simple and repetitive. Given the masses' limited intelligence, receptivity, and attention span, "all effective propaganda must be limited to a very few points and must harp on these in slogans until the last member of the public understands what you want him to understand by your slogan."[257] Putting Hitler's propaganda theories into practice, under Goebbels' leadership, the Nazis projected Hitler as a charismatic leader, the "national community" as an alternative to the partisan and class divisiveness that plagued the Weimar Republic, and Jews as the personification of cultural decadence, weakness, racial impurity, and political immorality.[258]

Historians debate whether Nazi propaganda was as effective in converting the masses as Hitler and Goebbels believed. Some argue that Nazi propaganda primarily appealed to preexisting beliefs and values of those regions and population sectors who were predisposed to be

254 Karl Loewenstein, *Militant Democracy and Fundamental Rights, II*, 31 AM. POL. SCI. REV. 638, 652 (1937).

255 ADOLF HITLER, MEIN KAMPF 176–186, 579–595 (Ralph Manheim trans. 1943) (chapters titled "War Propaganda" and "Propaganda and Organization").

256 *Id.* at 180.

257 *Id.*

258 DAVID WELCH, THE THIRD REICH: POLITICS AND PROPAGANDA 52–110 (2d ed. 2002).

receptive.[259] In regions that were historically anti-Semitic, for example, the Nazi's anti-Semitic propaganda was apparently seen as a welcome cue that the Nazi regime was on their side, and thus that they could freely express and act upon their prejudices.[260] By contrast, some studies suggest that Nazi propaganda was generally effective at key points. One such study concludes that Hitler's speeches were instrumental in garnering new voters during the 1932 presidential runoff election.[261] And another finds that Nazi radio propaganda succeeded in gaining and consolidating electoral support for the Nazi Party after the Nazis gained control over radio in January 1933.[262] Nazis also made effective, systematic use of defamation, character assassination, and group libel to disempower political opponents and bring widespread contempt upon Jews and other marginalized communities.[263] All in all, Nazi propaganda served as a highly potent weapon, albeit certainly not the single causal factor in undermining Weimar democracy and solidifying Nazi rule.[264]

In calling for militant democracy, Loewenstein was acutely aware of democracy's vulnerability to the propaganda machinery that the Nazis and other fascist parties deployed to foment invective and outrage:

> Fascism simply wants to rule. The vagueness of the fascist offerings hardens into concrete invective only if manifest deficiencies of the democratic system are singled out for attack. Leadership, order, and discipline are set over against parliamentary corruption, chaos, and selfishness; while a cryptic corporativism is substituted for political representation. General discontent is focussed on palpable objectives (Jews, freemasons, bankers, chain stores). Colossal propaganda is launched against what appears as the most conspicuously vulnerable targets. A technique of incessant repetition, of over-statements and over-simplifications, is evolved and applied. The different sections of

259 David Welch, *Nazi Propaganda and the Volksgemeinschaft: Constructing a People's Community*, 39 J. CONTEMP. HIST. 213, 213–14 (2004); Peter Selb & Simon Munzert, *Examining a Mostly Likely Case for Strong Campaign Effects: Hitler's Speeches and the Rise of the Nazi Party, 1927–1933*, 112 AM. POL. SCI. REV. 1050 (2018).

260 HUGO MERCIER, NOT BORN YESTERDAY: THE SCIENCE OF WHO WE TRUST AND WHAT WE BELIEVE 129–30 (2020); Maja Adena, Ruben Enikolopov, Maria Petrova, Veronica Santarosa & Ekaterina Zhuravskaya, *Radio and the Rise of the Nazis in Prewar Germany*, 130 Q.J. ECON. 1885, 1889–90 (2015).

261 Selb & Munzert, *supra* note 259, at 1060.

262 Adena, Enikolopov, Petrova, Santarosa & Zhuravskaya, *supra* note 260, at 1889.

263 *See* Rauch, *supra* note 76, at 14–15 (summarizing David Riesman's *Columbia Law Review* study of the failure of defamation law to prevent such attacks, David Riesman, *Democracy and Defamation: Fair Game and Fair Comment I*, 42 COLUM. L. REV. 1085 (1942)).

264 WELCH, *supra* note 258, at 18.

> the people are played off against one another. In brief, to arouse, to guide, and to use emotionalism in its crudest and its most refined forms is the essence of the fascist technique for which movement and emotion are not only linguistically identical. It is a peculiar feature of the emotional technique that those who are brought into play as the instruments, i.e., the masses, should not be aware of the rational calculations by which the wire-pullers direct it. Fascism is the true child of the age of technical wonders and of the emotional masses.
>
> This technique could be victorious only under the extraordinary conditions offered by democratic institutions. Its success is based on its perfect adjustment to democracy. Democracy and democratic tolerance have been used for their own destruction. Under cover of fundamental rights and the rule of law, the anti-democratic machine could be built up and set in motion legally. Calculating adroitly that democracy could not, without self-abnegation, deny to any body of public opinion the full use of the free institutions of speech, press, assembly, and parliamentary participation, fascist exponents systematically discredit the democratic order and make it unworkable by paralyzing its functions until chaos reigns. They exploit the tolerant confidence of democratic ideology that in the long run truth is stronger than falsehood, that the spirit asserts itself against force.[265]

Loewenstein placed primary emphasis on preventing fascists and other totalitarian forces from exploiting liberal democratic rights of speech, association, and political participation to subvert democracy. By contrast with Brandeis, Loewenstein had little faith that time to expose falsehood and fallacies through discussion would necessarily avert evil. In Lowenstein's experience, large segments of the population are fatally susceptible to emotional manipulation at the hands of totalitarian forces with the political liberty and organizational capacity to deploy a machine of propaganda. It is thus incumbent on democratic states to confront and stifle antidemocratic subversive propaganda before it has mass effect.

3. Militant Democracy and Democratic Citizenship

[265] Loewenstein, *supra* note 252, at 423–24.

Loewenstein's theory of militant democracy has been criticized by some as being elitist given his belief that democracy must be protected against the masses' inherent emotionalism and susceptibility to propaganda.[266] Yet other thinkers have emphasized that militant democracy also posits that the democratic state must guarantee the requisite conditions for a robust participatory democracy. They devote particular attention to the need for democratic equality and to countering the undue influence of market forces on democratic institutions and debate.

In his book, published in 1938, *It is Later than You Think: The Need for a Militant Democracy*, Max Lerner argued that unbridled capitalism undermines democracy.[267] A staunch supporter of the New Deal, Lerner advocated a kind of social democracy, replete with state-led economic planning and the socialization of the banking and credit system. He also contended that capitalism is prone to the formation of powerful antidemocratic oligarchies, as demonstrated by the affinity of conservative business magnates with European fascism.[268] Similarly, the Austro-Hungarian economic anthropologist Karl Polyani maintained that authoritarian market liberalism, with its fierce resistance to social democracy, "hollowed out democracy" and "weaken[ed] its ability to respond to" fascism.[269] Echoing that view, Karl Mannheim, the Hungarian sociologist who, like Lowenstein, fled the Nazi regime, distinguished militant democracy from laissez-faire liberalism. He argued that if democracy is to survive, the state must militantly pursue social justice through progressive taxation, control of investment, public works, and the radical extension of social services, while still leaving ample room for individual liberty and choice.[270] Building on such insights, contemporary scholars Anthoula Malkopoulou and Ludvig Norman similarly argue that social justice is a "precondition for political

266 *See, e.g.*, Sajó, *supra* note 253, at 570.

267 MAX LERNER, IT IS LATER THAN YOU THINK: THE NEED FOR A MILITANT DEMOCRACY 9–19, 24 (1943).

268 This paragraph draws on Graham Maddox, *Karl Loewenstein, Max Lerner, and Militant Democracy: An Appeal to 'Strong Democracy,'* 54 AUSTRALIAN J. POL. SCI. 490, 493–96 (2019). By contrast to Lerner, Loewenstein exhibited a certain tendency toward neoliberalism. Criticizing Lerner, he argued that democracy and capitalism together "need peace and safety of investment more than anything else." Loewenstein, *supra* note 252, at 422; *see also* Karl Loewenstein, *Reviewed Work: It Is Later Than You Think; The Need for a Militant Democracy by Max Lerner*, 33 AM. POL. SCI. REV. 519, 519–21 (1939).

269 Michael A. Wilkinson, *Authoritarian Liberalism in Europe: A Common Critique of Neoliberalism and Ordoliberalism*, 45 CRITICAL SOCIO. 1023, 1025 (2019).

270 KARL MANNHEIM, DIAGNOSIS OF OUR TIME: WARTIME ESSAYS OF A SOCIOLOGIST 6–8 (1943).

participation," countering authoritarian extremism, and "stabilising democracy."[271]

Finally, Jürgen Habermas supports post-war Germany's adoption of Loewenstein's call for a "militant" democracy that is "prepared to defend itself" against avowed enemies of the liberal democratic polity embodied in … [Germany's] constitution——whether that enemy be a secular totalitarian "political ideologist who combats the liberal state, or . . . [a] religious . . . fundamentalist who violently attacks the modern way of life per se."[272] In those narrow cases, Habermas contends, democracies must, paradoxically, jettison their guarantees of tolerance and political freedom, taking preventative measures against the enemies of those core constitutional principles, whether by "bringing to bear the means afforded by political criminal law or by decreeing the prohibition of particular political parties (Article 21.2 of the German Constitution) and the forfeiture of basic rights (Article 18 and Article 9.2 of the same)."[273]

Importantly, Habermas also highlights "the egalitarian and universalistic standards of democratic citizenship, something that calls for the equal treatment of the 'other' and mutual recognition of all as 'full' members of the political community."[274] In that vein, he further posits, the democratic state must foster a political public sphere that is "an inclusive space for possible discursive clarification of competing claims to truth and the generalisation of interests."[275] Critically for Habermas, public discourse, which is the essence of democratic governance, must encompass speech embodying rationality, reliability, and general relevance to promoting mutual understanding about common and different interests, while still enabling citizens to make their own considered judgments about the relevant issues for political decision-making.[276] Of relevance to both traditional and social media, therefore, "maintaining a media structure that ensures the inclusive character of the public sphere and the deliberative character of the formation of public opinion and political will is not a matter of political preference but a constitutional imperative."[277]

271 Malkopoulou & Norman, *supra* note 249, at 454.

272 Jürgen Habermas, *Religious Tolerance: The Pacemaker for Cultural Rights*, 79 PHIL. 5, 8 (2004) (citing Loewenstein, *supra* note 252).

273 Habermas, *supra* note 272, at 8.

274 *Id.* at 10.

275 Habermas, *supra* note 147, at 166.

276 *Id.* at 165.

277 *Id.* at 168.

4. Sum

In sum, militant democracy encompasses both negative and positive features. At its core, as set out by Lowenstein, militant democracy insists that democratic states may and, indeed, must curtail democratic liberties when needed to prevent authoritarian forces from exploiting them to subvert democracy. Yet, as Habermas and others have emphasized, militant democracy also calls upon democratic states affirmatively to foster the conditions for a robust liberal democracy predicated upon an egalitarian, inclusive and rationally deliberative public sphere.

In that view, particularly in our era of growing illiberal populism, militant democracy cannot rest merely on countering avowedly antidemocratic parties who would seize power and abolish further elections. Rather, to defend democracy, democratic states must actively underwrite a robust liberal democratic order predicated on minority rights, pluralism, social justice, the rule of law, and a political process undistorted by concentrations of wealth.[278] An enduring democracy also depends upon independent news media and other institutions dedicated to providing citizens with dependable, truthful information and a range of reason-based opinion.[279] In arguing for applying militant democracy to counter social media's corrosive impact on democracy, I mean to invoke that broad understanding of how democratic states must assertively foster robust, liberal democratic self-governance, not just militant democracy's core mission of defending against avowedly antidemocratic political forces.

B. *Principal Legal-Constitutional Measures of Militant Democracy*

Militant democracy finds expression in a variety of strategies designed to protect and promote a robust liberal democratic order. Traditional strategies most commonly defined as manifestations of militant democracy include measures that prohibit or contain political parties and activities that aim to subvert democratic governance, as well as measures prohibiting speech and political associations that threaten to

278 See Angela K. Bourne & Bastiaan Rijpkema, *Militant Democracy, Populism, Illiberalism: New Challengers and New Challenges*, 18 EUR. CONST. L. REV. 375 (2022) (introducing symposium that surveys challenges to militant democracy posed by purportedly "democratic" illiberal populism).

279 *See* Huq, *supra* note 15, at 1115 (arguing that democracy needs to provide citizens with dependable mechanisms for distinguishing reliable knowledge from out-and-out falsehood and unproven belief).

undermine the social fabric of a pluralist, democratic civil society by fomenting racial, ethnic, or national hatred. For example, the Federal Republic of Germany's Basic Law provides for the possibility of disbanding antidemocratic parties and dissolving antidemocratic associations.[280] Relying on those provisions, Germany has dissolved various antidemocratic political organizations and banned the neo-Nazi and Communist parties in the 1950s. Other democratic countries have also banned and/or required the disbandment of extremist political parties and associations regarded as a threat to the democratic order or to the defining principles of the particular democratic state.[281]

Applying principles of militant democracy, democratic countries also commonly prohibit speech that foments racial, ethnic, or national hatred, as well as hate speech that targets individuals based on their gender, religious belief, or sexual orientation.[282] In addition, Germany and other countries criminalize Holocaust denial as a form of hate speech.[283] Importantly, in the militant democracy perspective hate speech is not merely denigrating and highly offensive to individuals. Rather, it is understood to undermine the democratic equality and social cohesion upon which enduring democracy depends. In particular, as commentators have highlighted, hate speech can silence and sharply diminish the civic

280 *See* Grundgesetz [GG] [Basic Law], art. 21.2 (disbanding parties), 9.2 (dissolving associations), 18 (individuals who abuse rights may lose freedoms), translation at http://www.gesetze-im-internet.de/englisch_gg/index.html [https://perma.cc/EGC7-B9E8].

281 For discussion and examples, see Giovanni Capoccia, *Militant Democracy: The Institutional Bases of Democratic Self-Preservation*, 9 ANN. REV. L. & SOC. SCI. 207 (2013); Patrick Macklem, *Militant Democracy, Legal Pluralism, and the Paradox of Self-Determination*, 4 INT'L J. CONST. L. 488, 488, 493–94 (2006); and Issacharoff, *supra* note 18. Somewhat similarly, the U.S. Constitution provides for the possibility of disqualifying individuals from federal office, including following their impeachment or participation in insurrection. However, those provisions generally target individuals for their prior egregious conduct rather than political movements deemed to pose an ongoing threat to democracy. *See generally* Tom Ginsburg, Aziz Z. Huq & David Landau, *The Law of Democratic Disqualification*, 111 CAL. L. REV. (forthcoming 2023).

282 *See* Alexander Tsesis, *Democratic Values and the Regulation of Hate Speech*, *in* MINORITIES, FREE SPEECH AND THE INTERNET 19, 23–33 (Oscar Pérez de la Fuente, Alexander Tsesis & Jędrzej Skrzypczak eds., 2023) (surveying hate speech prohibitions in democratic countries); Erik Bleich & Sylvia Al-Mateen, *Hate Speech and the European Court of Human Rights: Ideas and Judicial Decision-Making*, 29 MICH. ST. INT'L L. REV. 179 (2021); Claudia E. Haupt, *Regulating Hate Speech—Damned If You Do and Damned If You Don't: Lessons Learned from Comparing the German and U.S. Approaches*, 23 B.U. INT'L L.J. 299 (2005) (comparing German and U.S. approaches); *cf.* JEREMY WALDRON, THE HARM IN HATE SPEECH (2012) (defending hate speech bans on the grounds of protecting the human dignity of vulnerable minorities).

283 *See generally* Paolo Lobba, *Holocaust Denial Before the European Court of Human Rights: Evolution of an Exceptional Regime*, 26 EUR. J. INT'L L. 237 (2015); John C. Knechtle, *Holocaust Denial and the Concept of Dignity in the European Union*, 36 FLA. ST. U. L. REV. 41 (2008).

engagement of its victims.[284] In sum, militant democracy favors limiting individuals' right to express such hatred in order to promote the capacity of all citizens to participate in democratic governance and debate.

In addition to those core features of militant democracy, European countries defend democracy by forbidding or sharply restricting paid political advertising on certain media, typically radio and television.[285] Political advertising restrictions apply to issue advertising as well as to partisan advertising in the context of election campaigns. Such restrictions are understood to protect the integrity of democratic debate by preventing those with greater resources from dominating the most important media affecting public discourse. As the U.K. Parliament concluded in banning paid political advertising on television, the prohibition is required to "avoid the unacceptable risk that the public debate would be distorted in favor of deep pockets funding advertising in the most potent and expensive media."[286] Political advertising restrictions also aim to ensure broadcasters' impartiality and independence from powerful sponsors in making editorial judgments. Finally, political advertising restrictions are seen to improve the quality of public debate given that paid political ads are conveyed without any immediate opposition or critical journalistic filter. Such ads would, therefore, paint a "manufactured picture," akin in those "found in propaganda in totalitarian regimes."[287]

More generally, European countries also bolster democracy by fostering independent media dedicated to providing reliable information and, in particular, heavily subsidizing independent public service media.[288] As the German Constitutional Court has declared, the dissemination of information and opinion through broadcasting is so central to democratic governance that, pursuant to Germany's Basic Law,

284 This was an important observation of foundational critical race theory and feminist scholars in the United States. *See, e.g.*, Charles R. Lawrence III, *If He Hollers Let Him Go: Regulating Racist Speech on Campus*, 1990 DUKE L.J. 431, 452–53, 470–71; Catherine A. MacKinnon, *Pornography, Civil Rights, and Speech*, 20 HARV. C.R.-C.L. L. REV. 1, 3–4, 7–8 (1985); Richard Delgado, *Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 HARV. C.R.-C.L. L. REV. 133, 136–49 (1982); *see also* Owen M. Fiss, *The Supreme Court and the Problem of Hate Speech*, 24 CAP. U. L. REV. 281, 287–88 (1995). For empirical findings, see Siegel, *supra* note 69, at 64–69.

285 *See* Richard H. Pildes, *The Law of Democracy and the European Court of Human Rights*, *in* JUDICIAL POWER: HOW CONSTITUTIONAL COURTS AFFECT POLITICAL TRANSFORMATIONS 109, 115–22 (Christine Landfried ed., 2020).

286 Animal Defs. Int'l v. United Kingdom, 2013-II Eur. Ct. H.R. 203, 226.

287 TV Vest AS v. Norway, 2008-V Eur. Ct. H.R. 265, 283.

288 The United States lags far behind other democratic countries in supporting public broadcasting. *See* Neil Weinstock Netanel, *Mandating Digital Platform Support for Quality Journalism*, 34 HARV. J.L. & TECH. 473, 514–16 (2021).

it must be independent of both state control and market forces.[289] In that regard, the Council of Europe has similarly reiterated that politically independent public service media are "a vital element of democracy in Europe."[290] As the Council has emphasized, public service media plays an important role "in upholding the fundamental right to freedom of expression . . . , enabling people to seek and receive information, and promoting the values of democracy, diversity, and social cohesion."[291]

C. *Militant Democracy and International Human Rights*

Militant democracy posits that, even as fundamental human rights are intricately bound up with representative democracy, democratic states must defend themselves against those who would use democratic rights of free association, free assembly, free expression, and participation in elections and representative government to subvert democracy. As such, militant democracies impose limitations on rights that are otherwise protected under both domestic constitutions and international human rights instruments, including the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the European Convention on Human Rights. Reflecting the principles of militant democracy, international human rights jurisprudence supports limitations on individual rights to the extent such limitations are truly necessary to protect democracy.

First, international human rights instruments contain "abuse of rights" provisions that expressly withhold protection from antidemocratic actors. For example, Article 5(1) of the International Covenant on Civil

289 *See* Christopher Witteman, *Constitutionalizing Communications: The German Constitutional Court's Jurisprudence of Communications Freedom*, 33 HASTINGS INT'L & COMPAR. L. REV. 95, 114–15 (2010).

290 *Recommendation on Public Service Broadcasting*, COUNCIL OF EUR., (Jan. 27, 2007), http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=17177 [https://perma.cc/AD3F-REU6], *quoted in* KAREN DONDERS, PUBLIC SERVICE MEDIA IN EUROPE: LAW, THEORY AND PRACTICE 136 (2021).

291 *Public Service Media*, COUNCIL OF EUR., www.coe.int/en/web/freedom-expression/public-service-media [https://perma.cc/8K4Z-MYDF], *quoted in* DONDERS, *supra* note 290, at 138–39. The traditional European model of public service media as a cornerstone of democratic discourse does face challenges, ranging from competition from commercial broadcasters to political interference in some countries. Nonetheless, the view that it is incumbent on democratic states to proactively underwrite media that are independent of party and business interests and that will provide news, information, and diverse perspectives remains central to countries that bear other features of militant democracy. *See* DONDERS, *supra* note 290, at 136–38; John O'Hagan & Michael Jennings, *Public Broadcasting in Europe: Rationale, Licence Fee and Other Issues*, 27 J. CULTURAL ECON. 31 (2003).

and Political Rights provides: "Nothing in the present Covenant may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms recognized herein . . . ."[292] Essentially identical abuse of rights provisions are set out in Article 30 of the Universal Declaration of Human Rights,[293] Article 17 of the European Convention on Human Rights,[294] and Article 54 of the Charter of Fundamental Rights of the European Union.[295]

Those limitations on the rights of antidemocratic actors emerged from the postwar drafters' vivid memories of European fascist parties' exploitation of democratic political rights to seize power.[296] The drafters also faced the palpable threat that Stalinist parties would soon attempt to do the same.[297] As the French delegate to the U.N. Human Rights Commission stated: "The edifice of liberty which was erected in the Covenant must not be capable of being used against liberty itself."[298] Democracies, in other words, should be not forced to wait until the very moment a totalitarian seizure of power is imminent to protect themselves. Rather, they must be able to act preemptively against their demise.[299]

Further, Article 20 of the International Covenant on Civil and Political Rights (ICCPR) requires states to prohibit the advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility or violence.[300] The International Convention on the Elimination of all Forms of Racial Discrimination similarly obligates states to outlaw "all dissemination of ideas based on racial superiority or hatred" and all "organizations . . . and all other propaganda activities,

292 International Covenant on Civil and Political Rights art. 5(1), Dec. 16, 1966, 999 U.N.T.S 171 [hereinafter ICCPR].

293 G.A. Res. 217 (III) A, Universal Declaration of Human Rights, U.N. Doc. A/RES/217, art. 30 (Dec. 10, 1948).

294 European Convention on Human Rights [ECHR], art. 17, Nov. 4, 1950, 213 U.N.T.S. 221.

295 Charter of Fundamental Rights of the European Union [EU Fundamental Rights Charter], art. 54, 2010 O.J. (C 83) 403.

296 On Karl Lowenstein's central role in drafting the ALI's Statement of Essential Human Rights, issued in 1944, which later served as a defining text in crafting international human rights instruments as well as the post-war German constitution, see Greenberg, *supra* note 247, at 184–91.

297 *See* Fox & Nolte, *supra* note 242, at 1, 40–41; *see also* BERNADETTE RAINEY, PAMELA MCCORMICK & CLARE OVEY, THE EUROPEAN CONVENTION ON HUMAN RIGHTS 3–4 (8th ed. 2021).

298 Fox & Nolte, *supra* note 242, at 41 (quoting United Nations, Economic and Social Council, Commission on Human Rights, 5th Sess., 123d Mtg. at 8, U.N. Doc. E/CN.4/SR.123 (1949)).

299 On the abuse of rights provisions of Article 17 of the European Convention and their application by the European Court on Human Rights as reflections of militant democracy, see PITRUZZELLA & POLLICINO, *supra* note 51, at 76.

300 ICCPR, *supra* note 292, art. 20.

which promote and incite racial discrimination."[301] Such provisions require states to take preemptive action by making the advocacy of racial hatred illegal. They stand in opposition to the notion that the sole remedy for hate speech in democratic society is wide-open debate to discredit insidious ideas.

Finally, all comprehensive human rights instruments provide that key rights that would normally be deemed essential to effective political participation may be restricted when "necessary in a democratic society." For example, Article 22(2) of ICCPR, protecting the freedom of association, provides that:

> No restrictions may be placed on the exercise of this right other than those which are prescribed by law and which are necessary in a democratic society in the interests of national security or public safety, public order (*ordre public*), the protection of public health or morals or the protection of the rights and freedoms of others. [302]

The European Convention on Human Rights contains a similar provision with respect to the freedom of assembly and association.[303] It also provides that the right of freedom of expression

> . . . may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[304]

In applying the provisions of the European Convention, the European Court on Human Rights has expressed keen awareness of the tension between, on one hand, protecting fundamental rights of freedom of expression, association, and political participation, and, on the other, the need for democratic states to defend themselves against antidemocratic subversion. As the court affirmed in *Klass v. Germany*:

---

301 International Convention on the Elimination of all Forms of Racial Discrimination art. 4, Dec. 21, 1965, 660 U.N.T.S. 195. The United States is party to the Convention but ratified the treaty subject to a reservation under which the United States "does not accept any obligation . . . to restrict those [extensive protections of individual freedom of speech, expression and association contained in the Constitution and laws of the United States], through the adoption of legislation or any other measures, to the extent that they are protected by the Constitution and laws of the United States." Reservations to International Convention on the Elimination of All Forms of Racial Discrimination, Dec. 21, 1965, 660 U.N.T.S. 195.

302 ICCPR, *supra* note 292, art. 22(2).

303 *See* ECHR, *supra* note 294, art. 11(2).

304 *Id.* art. 10(2).

"[S]ome compromise between the requirements for defending democratic society and individual rights is inherent in the system of the Convention . . . ."[305] Indeed, as the court noted, the Convention's Preamble states that "Fundamental Freedoms . . . are best maintained on the one hand by an effective political democracy and on the other by a common understanding and observance of the Human Rights upon which (the Contracting States) depend."[306]

The European Court on Human Rights requires that limitations on fundamental freedoms are prescribed by law and are strictly proportionate to a clear, serious threat to the democratic polity. Nonetheless, the Court has upheld the banning of antidemocratic political parties and prohibitions on a wide range of speech and activities, including political advertising; hate speech against racial, ethnic, or religious groups, or against homosexuals; distributing neo-Nazi or Stalinist communist pamphlets; advocating political violence; certain expressions of Islamic fundamentalism; and Holocaust denial.[307] In that regard, the Court has broadly held that "speech that is incompatible with the values proclaimed and guaranteed by the Convention is not protected by Article 10 [setting out the right of freedom of speech] by virtue of Article 17 of the Convention."[308]

In so doing, the Court has pointed to the continuing relevance of the principle of militant democracy in defining the Convention's compromise between defending democracy and human rights.[309] For example, in a 2001 ruling upholding the decision of the Turkish Constitutional Court to ban an Islamist party, the European Court on Human Rights accepted the Turkish government's argument that the party advocated state implementation of a version of Sha'aria that would be irreconcilable with democracy as conceived by the Convention and that there was a pressing need to order that the Islamist party be dissolved. As the Court explained: "In view of the very clear link between the Convention and democracy . . . , no one must be authorised to rely on the Convention's provisions in order to weaken or destroy the ideas and values of a democratic society."[310] Further, "a State cannot be required to wait, before intervening, until a political party has seized power and

305 Klass v. Germany, App. No. 5029/71, 28 Eur. Ct. H.R. (ser. A) ¶ 59 (1978).

306 *Id*. (quoting ECHR, *supra* note 294).

307 *See* DAVID HARRIS, MICHAEL O'BOYLE, ED BATES & CARLA M. BUCKLEY, LAW OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS 601 (4th ed. 2018).

308 Delfi AS v. Estonia, App. No. 64569/09, ¶ 136 (June 16, 2015), https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-155105%22]} [https://perma.cc/FF57-RYPH].

309 *See generally* Paul Harvey, *Militant Democracy and the European Convention on Human Rights*, 29 EUROPEAN L. REV. 407 (2004); Rory O'Connell, *Militant Democracy and Human Rights Principles*, 1 CONST. L. REV. 84 (2009).

310 Refah Partisi v. Turkey, 2003-II Eur. Ct. H.R. 267, 303–04.

begun to take concrete steps to implement a policy incompatible with the standards of the Convention and democracy, even though the danger of that policy for democracy is sufficiently established and imminent."[311]

Likewise, in its rulings on restrictions on paid political advertising, the Court has reiterated that, as the "ultimate guarantor" of "free and pluralist debate," states are entitled to enact proportionate, targeted regulations to "protect the democratic debate and process from distortion by powerful financial groups with advantageous access to [the] media."[312] Indeed, to permit powerful financial groups to obtain competitive advantages in commercial political advertising and, possibly, to exploit that market power "to curtail the freedom of[] the radio and television stations broadcasting the commercials," would "undermine[] the fundamental role of freedom of expression in a democratic society as enshrined in Article 10 of the Convention, in particular where it serves to impart information and ideas of general interest, which the public is moreover entitled to receive . . . ."[313]

IV. MILITANT DEMOCRACY PRINCIPLES APPLIED TO REGULATING ONLINE PLATFORMS

Militant democracy is not without its critics. Some question whether government officials can be trusted to accurately and dispassionately identify which ideological or religious groups actually pose a serious threat to the democratic order.[314] Others contend that militant democracy entrenches elites even against democratic populism.[315]

I cannot fully engage such criticisms within these limited pages.[316] I will only observe, very briefly, that politically independent courts and administrative agencies, including supranational bodies such as the

311 *Id.* at 305.

312 Animal Defs. Int'l v. United Kingdom, 2013-II Eur. Ct. H.R. 203, 235 (upholding paid political advertising ban against Article 10 challenge).

313 Verein gegen Tierfabriken (Vgt) v. Switzerland, No. 24699/94, ECHR 2009-VI, ¶ 73 (ruling that paid political advertising ban as applied to the particular case violated Article 10).

314 *See, e.g.*, Carlo Invernizzi Accetti & Ian Zuckerman, *What's Wrong with Militant Democracy?*, 65 POL. STUD. 182 (2016); *see also* UDI GREENBERG, THE WEIMAR CENTURY: GERMAN ÉMIGRÉS AND THE IDEOLOGICAL FOUNDATIONS OF THE COLD WAR 203–09 (2014) (contending that militant democracy achieved prominence in postwar West Germany and then in Western Europe as part of the virulent, and arguably excessive, anticommunism of the Cold War).

315 *See*, Angela K. Bourne, *From Militant Democracy to Normal Politics? How European Democracies Respond to Populist Parties*, 18 EUR. CONST. L. REV. 488, 490–93 (2022) (discussing views of critics and defenders).

316 For a thoughtful response to some critics, see Alexander S. Kirschner, *Militant Democracy Defended*, *in* MILITANT DEMOCRACY AND ITS CRITICS, *supra* note 17, at 56, 56–71.

European Court on Human Rights, can provide significant institutional checks on misuses of militant democracy prerogatives.[317] Further, the argument that militant democracy entrenches elites seems aimed primarily at Loewenstein's original conception of militant democracy, with its singular focus on suppressing political forces that are determined by those in power to constitute intolerable threats to democracy. The criticism largely ignores more robust interpretations of militant democracy that incorporate a Habermasian social democratic framework designed to promote broad, pluralist, egalitarian participation in democratic debate. Finally, whatever the criticism, militant democracy remains a fundamental, accepted feature of human rights law and postwar constitutionalism in many democratic states. In that light, while constitutional safeguards and judicial review cannot eliminate possibilities for abuse of power even by nominally democratic states, it is perhaps telling that European democracies that incorporate features of militant democracy score significantly higher than the United States on indices of democracy and freedom published by independent think tanks, including Freedom House, Economist Intelligence Unit, and V-Dem Institute.[318]

In any event, as I will presently discuss, the application of militant democracy principles to social media harms raises some different issues—and challenges—than those subsumed within critics' focus on militant democracy's original core pursuit: heading off threats from avowed, ideological enemies of democracy. Militant democracy principles undergird the European Commission's 2020 European Democracy Action Plan to bolster democratic institutions in the face of authoritarian populist and foreign exploitation of online platforms to undermine election integrity, engage in coordinated disinformation campaigns, manipulate voters, and intimidate journalists, women, and minority speakers through targeted harassment and hate speech.[319] As the Commission Action Plan recognizes, democracy, "a core European

317 Aziz Huq comes to a similar conclusion. *See* Huq, *supra* note 15,, at 1132–33.

318 *See* FREEDOM HOUSE, FREEDOM IN THE WORLD 2023, at 12 (2023), https://freedomhouse.org/sites/default/files/2023-03/FIW_World_2023_DigtalPDF.pdf [https://perma.cc/549S-FHUA]; ECONOMIST INTELLIGENCE UNIT, DEMOCRACY INDEX 2021, at 12 tbl.2 (2022), https://www.eiu.com/n/campaigns/democracy-index-2021 [https://perma.cc/7DHG-ASME]; V-DEM INST., DEMOCRACY REPORT 2022: AUTOCRATIZATION CHANGING NATURE? 10 fig.1 (2022), https://v-dem.net/media/publications/dr_2022.pdf [https://perma.cc/V2WC-5QB7] (countries by Score on V-Dem's Liberal Democracy Index (LDI) 2011 Compared to 2021). There might be a wide array of explanations for why European democracies score higher on those indices. But the scores do suggest that, at the very least, militant democracy features do not correlate with abuses of power that substantially impair democracy, as defined and measured by the three leading indices.

319 *See* Democracy Action Plan, *supra* note 23.

value," "cannot be taken for granted—it needs to be actively nurtured and defended."[320]

To that end, the Commission celebrates the digital revolution's affordance of "new opportunities for civic engagement, making it easier for some groups . . . to access information and participate in public life and democratic debate."[321] The Commission recognizes, however, that "the rapid growth of online campaigning and online platforms has also opened up new vulnerabilities and made it more difficult to maintain the integrity of elections, ensure a free and plural media, and protect the democratic process from disinformation and other manipulation."[322] As the Commission warns: "Our democratic systems and institutions have come increasingly under attack in recent years . . . The very freedoms we strive to uphold, like the freedom of expression, have been used in some cases to deceive and manipulate."[323]

The European Democracy Action Plan advances a broad, multi-pronged framework to regulate online political advertising, ensure a free and plural media, and protect the democratic process from disinformation and manipulation. The Action Plan finds concrete expression in an interlocking matrix of recent European initiatives. The initiatives, some of which have been adopted and others of which are still under consideration, including, among others, the DSA,[324] the Code of Conduct on Countering Illegal Hate Speech Online,[325] the Strengthened Code of Practice on Disinformation 2022,[326] the Artificial Intelligence Act,[327] the European Media Freedom Act,[328] and the Regulation on the Transparency and Targeting of Political Advertising.[329] The principle that

320 *Id.* at 1.

321 *Id.* at 2.

322 *Id.*

323 *Id.* at 1.

324 DSA, *supra* note 24.

325 *The EU Code of Conduct on Countering Illegal Hate Speech Online*, EUROPEAN COMM'N, https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en [https://perma.cc/X47D-X4G2] [hereinafter Hate Speech Code].

326 2022 Strengthened Code of Practice on Disinformation, EUROPEAN COMM'N 1 (June 16, 2022), https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation [https://perma.cc/VJ8D-ETNJ] [hereinafter Disinformation Code].

327 *Proposal for a Regulation of a European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 4, 2021).

328 Proposed Media Freedom Act, *supra* note 25.

329 *Proposal for a Regulation of the European Parliament and of the Council on the Transparency and Targeting of Political Advertising*, COM (2021) 731 final (Nov. 25, 2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0731 [https://perma.cc/9T4L-TFRW].

democratic states must actively defend against threats to democracy informs that regulatory framework, in sharp contrast to the neoliberal, market-centered understandings that generally animate U.S. policy.[330]

As applied specifically to social media harms to democracy, the Action Plan initiatives generally aim to prod large social media companies to act as actively engaged, trusted intermediaries in providing online platforms for public discourse that present a broad range of information and opinion comporting with basic standards for democratic debate. The Action Plan fully recognizes the fundamental differences between social media and legacy media. Indeed, it applauds the vast new opportunities for bottom-up civic engagement that social media offers. Nonetheless, the Action Plan insists that large social media companies assert gatekeeper responsibilities for promoting trustworthy, fact-based information and for preventing the systematic exploitation of their platforms to propagate disinformation, emotionally manipulative propaganda, hate speech, and violent incitement.

The Action Plan pursues its objectives through a combination of regulatory mandate, co-regulation, and strong encouragement of social media self-regulation prompted by due diligence requirements for reporting, external audit, and transparency. In particular, as provided under the DSA and related measures, large digital platforms must (1) remove illegal antidemocratic speech; (2) assess, report, and mitigate systemic risks arising from the design, function, or use of their platforms to civic discourse, electoral processes, or the exercise of fundamental rights; and (3) account for the fundamental rights of users and others impacted by platform content moderation.

We consider each in turn.

A. *Illegal Antidemocratic Speech*

At its core, the European militant democracy framework requires that online services disable access to the types of antidemocratic speech that are typically illegal in European countries. These include hate speech, terrorist propaganda, incitement to serious violent offenses, Holocaust denial, and recruitment initiatives of banned antidemocratic political parties and associations.[331] The basic principle that democratic

330 Similarly, as Claudia Haupt argues, Germany's Network Enforcement Act (known as the NetzDG), which requires online platforms to expeditiously remove hate speech, terrorist recruitment, and other illegal speech, is of a piece with the broad militant democracy aim "to protect democratic public discourse and defend democracy itself." Haupt, *supra* note 191, at 780.

331 For example, among the categories of illegal content enumerated in Germany's Network Enforcement Act (known as the NetzDG) are propaganda material of unconstitutional organizations, symbols of unconstitutional organizations, preparation of a serious violent offense

states must defend themselves against antidemocratic speech and subversion applies no less to speech on social media platforms than to offline speech.[332] As the European Council iterated in support of the then-proposed European Union Digital Services Act: "[W]hat is illegal offline should also be illegal online."[333]

Under EU law, platforms that host user-posted content are not liable for illegal content of which they have no actual knowledge.[334] Nor, under the DSA, are they obligated to actively monitor user posts to identify illegal content.[335] However, the DSA provides that: (1) national

endangering the state, forming terrorist organizations, incitement to hatred, dissemination of depictions of violence, and defamation of religions or of religious and ideological associations. Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz) [NetzDG] [Network Enforcement Act], Sept. 1, 2017, BUNDESGESETZBLATT [BGBL] § 1(3) (Ger.); *see* Patrick Zurth, *The German NetzDG as Role Model or Cautionary Tale? Implications for the Debate on Social Media Liability*, 31 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1084, 1110 (2021) (describing what content is illegal under German law and referenced in the NetzDG); *see also* Erik Bleich & Sylvia AL-Mateen, *Hate Speech and the European Court of Human Rights: Ideas and Judicial Decision-Making*, 29 MICH. ST. INT'L L. REV. 179 (2021) (discussing national law prohibitions to hate speech that have been brought before the court); Paolo Lobba, *Holocaust Denial Before the European Court of Human Rights: Evolution of an Exception Regime*, 26 EUR. J. INT'L L. 237 (2015).

332 In that regard, the European Court of Human Rights rejected a claimed violation of the right to free expression brought by the leader of the organization "Sharia4Belgium," who had been convicted of inciting hatred by posting YouTube videos expressing contempt for non-Muslims and calling on Muslims to dominate them, an instance, the court held, of attempting to use the right of freedom of expression for purposes manifestly contrary to the ECHR values of tolerance, peace, and non-discrimination. Belkacem v. Belgium, App. No. 34367/14, (June 27, 2017), https://hudoc.echr.coe.int/eng?i=001-175941 [https://perma.cc/A6Y5-QBBH].

333 European Council Press Release, What Is Illegal Offline Should Be Illegal Online: Council Agrees Position on the Digital Services Act (Nov. 25, 2021), https://www.consilium.europa.eu/en/press/press-releases/2021/11/25/what-is-illegal-offline-should-be-illegal-online-council-agrees-on-position-on-the-digital-services-act/#:~:text=The%20Council%20agreed%20its%20position,protect%20their%20fundamental%20rights%20online [https://perma.cc/EE3S-Z85A] .

334 That principle is expressly stated in the DSA, *supra* note 24, art. 6, at 45. *See also* Folkert Wilman, *Between Preservation and Clarification: The Evolution of the DSA's Liability Rules in Light of the CJEU's Case Law*, *in* PUTTING THE DSA INTO PRACTICE 35, 37–39 (Joris van Hoboken, et al. eds., 2023) [hereinafter PUTTING THE DSA INTO PRACTICE] (explaining that the DSA expressly follows existing EU law on internet intermediary liability for hosting illegal content); McGonagle, *supra* note 228, at 481–85 (noting some movement in EU law towards requiring intermediaries to take greater initiative to proactively prevent illegal speech).

335 DSA, *supra* note 24, art. 8, at 45. The EU Regulation to Address the Dissemination of Terrorist Content Online requires hosting service providers that have been exposed to terrorist content to take specific measures to protect their services against the further dissemination of terrorist content to the public. But the Regulation leaves the service provider considerable discretion regarding which measures to take, providing the measures are effective in mitigating exposure to terrorist content, are targeted and proportional, and include safeguards to avoid removing material that is not terrorist content. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on Addressing the Dissemination of Terrorist Content Online, 2021 O.J. (L 172) 79, art. 5, at 92–93.

authorities may order platforms to remove specific items of illegal content; (2) platforms must put in place user-friendly mechanisms for individuals or civil society organizations to notify them of allegedly illegal content and that such notices may subject the platform to liability for knowingly hosting illegal content where the illegality is apparent without a detailed legal examination; and (3) platforms must suspend users who frequently post manifestly illegal content.[336] Further, the Act provides that online platforms must expeditiously assess and act upon notices of illegal content submitted by "trusted flaggers": civil society organizations that national authorities have certified as having expertise in detecting and identifying illegal content.[337] Finally, the DSA provides a strong incentive for "very large" platforms to proactively remove illegal content.[338] It requires that such platforms carry out and submit to the EU Digital Services Coordinator an annual risk assessment that, among other matters, must evaluate the systemic risk of dissemination of illegal content through their services.[339]

Additionally, the leading social media platforms have adopted and agreed to abide by the European Union's Code of Conduct on Countering Illegal Hate Speech Online.[340] Following EU precedent, the Code of Conduct describes hate speech as "all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or

336 DSA, *supra* note 24, art. 9, at 46–47 (national authority orders); *id.* art. 16, at 50–51 (notice and action mechanisms); *id.* art. 23, at 57 (user suspension). National legislation currently goes further than the DSA. Germany's Network Enforcement Act requires social media platforms with over 2 million users to remove "clearly illegal" content within 24 hours and all illegal content within 7 days of receiving a complaint about such content. NetzDG, *supra* note 331, § 1(2).

337 DSA, *supra* note 24, art. 22, at 56–57.

338 Pursuant to the DSA, "very large online platforms" and "very large search engines" are those that have 45 million or more active users in the EU. DSA, *supra* note 24, art. 33, ¶ 1, at 63. As of April 2023, the seventeen very large online platforms include, among others, YouTube, Facebook, Instagram, Amazon Store, Apple Store, TikTok, LinkedIn, Pinterest, Snapchat, Twitter, and Wikipedia, and the two very large online search engines include Google Search and Bing. European Commission Press Release, Digital Services Act: Commission Designates First Set of Very Large Online Platforms and Search Engines (Apr. 25, 2023), https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_23_2413/IP_23_2413_EN.pdf [https://perma.cc/3LWW-YLSK].

339 DSA, *supra* note 24, art. 34, ¶ 1(a), at 64. In addition to the DSA, the EU Audiovisual Media Services Directive, as updated in 2018, gives national media regulators authority over video platforms like YouTube. *See* Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018: Amending Directive 2010/13/EU on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States Concerning the Provision of Audiovisual Media Services (Audiovisual Media Services Directive) in View of Changing Market Realities, 2018 O.J. (L 303) 69, 70.

340 Hate Speech Code, *supra* note 325. Online platform participants in the Code of Conduct include Facebook, Microsoft, Twitter, YouTube, Instagram, Snapchat, Dailymotion, Jeuxvideo.com, TikTok, Linked, Rakuten Viber, and Twitch.

national or ethnic origin."[341] It declares that "[t]he spread of illegal hate speech online not only negatively affects the groups or individuals that it targets, it also negatively impacts those who speak out for freedom, tolerance and non-discrimination in our open societies and has a chilling effect on the democratic discourse on online platforms."[342] The Code of Conduct expresses the platforms' public commitment "to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary."[343]

The Code is putatively voluntary, but it provides for annual reporting and European Commission monitoring of platform actions taken to remove hate speech. The latest Commission evaluation, released in November 2022, reports that the platforms received a total of 3,634 notifications of hate speech from civil society organizations, trusted flaggers, and the general public during the sample period of March 28 to May 13, 2022.[344] The platforms assessed 64.4% of those complaints within less than twenty-four hours and removed 63.6% of the notified content, percentages that, while meeting the platform's basic commitment under the Code, are lower than the previous two years.[345]

It is expected that the DSA will lend further force to the Code of Conduct, and thus that the Code will effectively augment the DSA's provisions regarding illegal content. The DSA provides that the European Commission and the newly formed European Board for Digital Services shall regularly monitor and evaluate whether the Code of Conduct has achieved its objectives.[346] The two bodies are also required to ensure that the participating online platforms report to national as well as EU authorities on whether the measures they have implemented have met those objectives.[347]

---

341 *Id.* at 1 (citing Council Framework Decision 2008/913/JHA of 28 November 2008 on Combatting Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law, 2008 O.J. (L 328) 55).

342 *Id.*

343 *Id.* at 2. By comparison, the EU Regulation to Address the Dissemination of Terrorist Content Online requires platforms to remove "terrorist content" within one hour after receiving a removal order from the applicable national authority. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on Addressing the Dissemination of Terrorist Content Online, 2021 O.J. (L 172) 79, 90–91.

344 DIDIER REYNDERS, EUROPEAN COMMISSION, COUNTERING ILLEGAL HATE SPEECH ONLINE: 7TH EVALUATION OF THE CODE OF CONDUCT 1 (2022), https://commission.europa.eu/system/files/2022-12/Factsheet%20-%207th%20monitoring%20round%20of%20the%20Code%20of%20Conduct.pdf [https://perma.cc/36C2-2PEW].

345 *Id.*

346 DSA, *supra* note 24, art. 45, ¶ 4, at 76.

347 *Id.* art. 45, ¶ 3, at 76.

B. *Legal But Harmful Speech*

From the European militant democracy perspective, laws that require platforms to remove illegal content of which they have knowledge are largely uncontroversial, provided they are narrowly targeted to avoid unduly burdening the dissemination of legal content and provide sufficient due process to those accused of posting illegal speech.[348] The more complicated issue is whether and how militant democracy principles should apply to the design or use of social media that generates, amplifies, and channels vast amounts of disinformation and emotionally manipulative propaganda—speech that might not be illegal offline but that, given its propagation on social media, fuels antidemocratic extremism, polarization, general epistemic uncertainty, and profound disillusionment with democracy.

Militant democracy arose to thwart political parties bent on using democratic liberties to overthrow democracy. By contrast, social media have a broad range of uses. Many who propagate disinformation and emotive content are not necessarily opposed to democracy per se. They might merely be populists who aim to undermine established elites. Or they might be hardcore opportunists using whatever tools are available to win democratic elections. Even those who employ bots, fake accounts, and other deceptive technologies have a broad range of motives, ranging from political to financial.

Nonetheless, militant democracy provides a useful framework for addressing such diffuse social media harms. First, as I have discussed, militant democracy reaches beyond defending against avowedly autocratic political organizations. It also insists that the democratic state actively promote the integrity and vitality of the democratic process—and that includes the polity's commitment to respecting all citizens as equal partners in democratic governance. As such, militant democracy supports prohibitions on hate speech not just because hate speech is a tool for antidemocratic parties to seize power, but because hate speech undermines democratic equality and chills democratic discourse. Likewise, militant democracy supports democratic states' measures to ensure that public debate about political issues furthers democratic equality and is not distorted by financial advantage, market forces, ruling party interference, or the absence of critical filters. Those values

348 A French law was ruled unconstitutional because it did not provide online platforms sufficient time to assess whether posted speech is, in fact, illegal before removing it and provided inadequate judicial review for such removals. *See* Conseil constitutionnel [CC] [Constitutional Court] June 18, 2020, decision No. 2020-801DC, 1 (Fr.).

undergird European countries' restrictions on political advertising and their subsidizing of independent public service media.

Second, social media endanger democracy even when not marshalled by autocratic forces intent on seizing power by subverting democracy. Whatever the motive, when social media algorithms amplify disinformation and emotive outrage or are exploited to "flood the zone with shit," they pose a serious threat to democratic governance and equality.[349] In that connection, the rise of extremist right-wing movements in democratic countries are closely linked to the business models of Facebook, YouTube, TikTok, and other large digital platforms, which profit on spewing hatred, anger, and fear.[350] In that regard, militant democracy teaches that democracies need not and should not wait until their imminent demise to defend themselves against antidemocratic forces. While that lesson was initially grounded in prewar fascist takeovers of democratic government, it holds equally well for more diffuse threats posed by online platforms, provided that regulation is narrowly targeted to defend against palpable, serious harms.

Notably, the need to counter threats to democracy arising from the amplification of out-group hatred, political manipulation, and disinformation online expressly informs recent European regulatory initiatives, no less than requiring platforms to block illegal content. At the same time, the initiatives recognize that barring platforms and their users from promoting legal, but harmful speech, is a considerably more complex undertaking than prohibiting the prescribed categories of content that are also illegal offline. In particular, there remain many unknowns about how such otherwise legal speech harms democracy when propagated on social media. Further, social media platforms vary significantly in design and function. The platform recommender systems that amplify harmful speech involve an intricate combination of human and machine behavior, differ from platform to platform, and are subject to periodic platform modifications.[351]

Accordingly, European regulation in this area rightly focuses on (1) inducing the major platforms with the greatest impact and financial wherewithal to self-regulate by identifying and mitigating systemic risks of harms to democracy in their design and use, and (2) requiring those platforms to make their workings transparent to regulators, both as a

349 *See supra* notes 61–87 and accompanying text.

350 *See* Alexandra Geese, *Why the DSA Could Save Us From the Rise of Authoritarian Regimes*, *in* PUTTING THE DSA INTO PRACTICE, *supra* note 334, at 63, 65.

351 *See* Thorburn, Bengani & Stray, *supra* note 52; Alex Heath, *Facebook Is Changing Its Algorithm to Take on TikTok, Leaked Memo Reveals*, THE VERGE (Jun. 15, 2022, 12:46 PM), https://www.theverge.com/2022/6/15/23168887/facebook-discovery-engine-redesign-tiktok [https://perma.cc/TA8K-YHPP].

means to induce effective self-regulation and to educate regulators about how social media propagate harmful speech and, thus, whether there might be a need for further regulatory intervention. The regulations aim to further those objectives in several, interrelated ways.

To begin with, as with hate speech, the European Union puts considerable pressure on large platforms to enter into a co-regulatory code of practice that obligates signatories to address the harmful influence of online disinformation on the democratic process.[352] The Strengthened Code of Practice on Disinformation 2022, which forty-four leading platforms and technology companies have signed at the behest of the European Commission, proclaims that the

> . . . [s]ignatories recognise and agree with the European Commission's conclusions that "[t]he exposure of citizens to large scale Disinformation, including misleading or outright false information, is a major challenge for Europe," and that "[o]ur open democratic societies depend on public debates that allow well-informed citizens to express their will through free and fair political processes."[353]

Pursuant to the Code, the signatories undertake to employ measures to mitigate the risk that their services will fuel the viral spread of harmful disinformation. These include, among other measures, employing "recommender systems designed to improve the prominence of authoritative information and reduce the prominence of Disinformation based on clear and transparent methods and approaches for defining the criteria for authoritative information."[354] The Code likewise requires recommender system transparency for users, including the use of independent fact-checkers, enabling users to access indicators of trustworthiness developed by independent third parties in collaboration with news media professionals, and giving users the option of having the

352 For an assessment of the move from command-and-control regulation to self- and co-regulation in EU governance of the platform economy, see Michèle Finck, *Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy*, 43 EUR. L. REV. 47 (2018).

353 Disinformation Code, *supra* note 326. Signatories include Meta, Google, Microsoft, Vimeo, Twitch, and TikTok, as well as fact-checkers, advertisers, news media, and other organizations. *Signatories of the 2022 Strengthened Code of Practice on Disinformation*, EUROPEAN COMM'N (June 16, 2022), https://digital-strategy.ec.europa.eu/en/library/signatories-2022-strengthened-code-practice-disinformation [https://perma.cc/LV5B-Y68A]. Notably, Twitter was initially a signatory but withdrew from the Code of Practice following Elon Musk's takeover of the company. *See* Christopher Pitchers, *Twitter Has Chosen 'Confrontation' With Brussels Over Disinformation Code of Conduct*, EURONEWS (June 5, 2023, 2:10 PM), https://www.euronews.com/my-europe/2023/06/05/twitter-has-chosen-confrontation-with-brussels-over-disinformation-code-of-conduct [https://perma.cc/7BMV-G3AX].

354 Disinformation Code, *supra* note326, at 20.

recommender system reflect signals related to the trustworthiness of media sources. Code signatories also commit to reporting on how their recommender systems account for the trustworthiness of media sources and, in that connection, to reviewing information sources in a transparent, apolitical, unbiased, and independent manner.[355]

The Strengthened Code of Practice on Disinformation states that it aims to "complement and be aligned with the regulatory requirements and overall objectives in the Digital Services Act."[356] In turn, DSA Article 34 provides that "very large" online platforms and search engines, some of which are signatories to the Strengthened Code, must conduct an annual assessment of any systemic risks stemming from the design or functioning of their service, including related algorithmic systems, or from uses made of their services.[357] Per Article 34, systemic risks include, among other things, any actual or foreseeable negative effects on civic discourse, electoral processes, or the exercise of fundamental rights, including rights to human dignity, freedom of expression and information (including media freedom and pluralism), and non-discrimination.[358] Article 34 further provides that the risk assessment must account for how the online service's recommender systems, content moderation practice, data management, and advertising might impact any of the identified systematic risks.[359] The platform must also assess the possible impact of any "intentional manipulation of their service, including by inauthentic use or automated exploitation," such as the deployment of fake accounts, coordinated propaganda, and bots.[360] The risk assessments and supporting documentation must be filed with European regulators upon request.[361]

Nor are the DSA's very large platform and search engine obligations limited to filing annual reports. Article 35 requires that those platforms and search engines must put in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks they have identified, with particular consideration to the impacts of such measures on fundamental rights. As set forth in Article 35, such measures may include modifying the online service's recommender systems,

355 *Id.* at 8-9.

356 *Id.* at 2.

357 The services must also carry out a risk assessment "prior to deploying functionalities that are likely to have a critical impact" on the enumerated systemic risks. DSA, *supra* note 24, art. 34, ¶ 1, at 64.

358 *Id.* art. 34, ¶ 1(b), at 64.

359 *Id.* art. 34, ¶ 2, at 64–65.

360 *Id.* art. 34, ¶ 2, at 64.

361 The European regulators include the European Commission and national Digital Service Coordinators in the EU country where the "main establishment" of the platform or search engine is located. *Id.* art. 34, ¶ 3, at 65.

content moderation process, and advertising systems, among other measures. Mitigation measures regarding recommender systems would likely include some version of the measures set out in the Strengthened Code of Practice on Disinformation. As enumerated above, those measures include recommender system modifications to improve the prominence of authoritative information, including news media content that independent third parties have identified as trustworthy, and reducing the prominence of disinformation.[362] They might also include measures to reduce virality and false signals of popularity of user postings.[363]

In addition, pursuant to Article 37, very large platforms and search engines must fund an independent external audit, to take place at least annually, to assess the online service's compliance with various obligations that the DSA imposes. Among other obligations, these include the obligations on very large platforms and search engines set out in Articles 34 and 35 and any commitments pursuant to agreed-upon codes of conduct identified in Article 45, including the Code of Conduct to Counter Illegal Hate Speech Online and the Code of Practice on Disinformation, as well as obligations related to content moderation and removing illegal content.[364]

Finally, DSA Article 40 provides that very large online platforms and search engines must provide European regulators, upon request, access to data that are necessary to monitor and assess compliance with the DSA. In turn, European regulators may give access to that data to vetted researchers for the sole purpose of conducting research that contributes to the detection, identification and understanding of the systemic risks enumerated in Article 34, including negative effects on civic discourse, electoral process and fundamental rights, and the adequacy, efficiency and impacts of the platform's or search engine's risk mitigation measures pursuant to Article 35.[365] Those provisions recognize that further research is needed to determine how to algorithmically identify and combat hate speech, violent incitement, and

362 *See* Netanel, *supra* note 288, at 526–32 (proposing that platforms should be required to favor original reporting in their recommender system generated feed and to enable news sources to include third party ratings of trustworthiness); Martin Husovec, *Trusted Content Creators* (LSE L. Sch., Policy Briefing No. 52, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4290917 [https://perma.cc/H34F-H7PH] (proposing that platforms would be entitled to rely on regulator-certified trusted content creators, consisting of associations of journalists, fact-checkers, product reviewers, activists, or academics, who would police and certify the content of their members).

363 *See* Ellen P. Goodman, *Digital Information Fidelity and Friction*, KNIGHT FIRST AMEND. INST. (Feb. 26, 2020), https://knightcolumbia.org/content/digital-fidelity-and-friction [https://perma.cc/9D5P-TYJA] (advocating introducing friction to slow social media message transmission).

364 DSA, *supra* note 24, art. 37, ¶¶ 1–2, 4, at 67.

365 *Id.* art. 40, ¶ 4, at 70.

misinformation, as well as to determine whether and how platform content subverts the foundations of democratic governance.[366]

In short, while DSA regulation depends in the first instance on very large platforms' and search engines' self-assessment and nominally voluntary acceptance of codes of conduct, the requirements of mitigation measures, reporting, independent audits, and regulator access to data appear to lend the regulations considerable, far-reaching teeth. Much will depend on the rigor of regulatory enforcement, as well as the active involvement of trusted flaggers, researchers, and civil society organizations.[367] The regulatory framework will also be shaped by how platforms and regulators define, far more precisely, what is meant by "civic discourse" and "electoral processes," and what counts as a harm to them.[368] Also critical will be the extent to which platforms demote disinformation and give prominence to trustworthy news content, including that of independent public service media. Finally, platforms and regulators will have to draw a delicate balance between enabling freedom of expression and suppressing content deemed to endanger democracy, a subject to which we now turn.

C. *State-Imposed Digital Constitutionalism*

Alongside the EU's regulatory framework to counter the propagation of democracy-harming speech on online platforms, the DSA adopts and applies to the platforms what scholars have called "digital

366 *See* Ioanna Tourkochoriti, *The Digital Services Act and the EU as the Global Regulator of the Internet*, 24 CHI. J. INT'L L. 129, 131–32, 140–42 (2023) (noting need for further research).

367 *See generally* Martin Husovec, *Will the DSA Work?*, *in* PUTTING THE DSA INTO PRACTICE, *supra* note 334, at 19, 21; Julian Jaursch, *Platform Oversight: Here Is What a Strong Digital Services Coordinator Should Look Like*, *in* PUTTING THE DSA INTO PRACTICE, *supra* note 334, at 91, 93 ("The DSA's success depends on how well it is enforced."). A recent oversight report concludes that the overall quality of the baseline reports published by the largest signatories to the putatively voluntary Strengthened Code of Practice on Disinformation in February 2023 was less than adequate. *See* KIRSTY PARK & STEPHAN MÜNDGES, EDMO IR. & GERMAN-AUSTRIAN DIGIT. MEDIA OBSERVATORY, COP MONITOR: BASELINE REPORTS (2023), https://www.politico.eu/wp-content/uploads/2023/09/05/CoP-Monitor-Report.pdf [https://perma.cc/KK6Q-L85S]. The oversight report suggests that platform compliance might improve if and when the Code is effectively incorporated into the DSA, but that remains to be seen. *Id.* at 13. For a sobering view of co-regulatory approaches generally, see Selbst, *supra* note 137, at 162–68 (describing tendency within co-regulation, collaborative governance frameworks for firms to interpret and apply flexible and/or ambiguous regulatory directives to comport with the firm's preexisting organizational, business goals).

368 *See* Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 598–99 (2022) (highlighting the impossibility of quantifying risks to nebulous concepts like "civic discourse").

constitutionalism."[369] Digital constitutionalism recognizes that major online platforms, while nominally private actors, perform quasi-public tasks, including providing an essential forum for individual expression, democratic debate, and the open exchange of information and ideas. As such, the democratic state must impose on platforms duties to respect fundamental human rights in their relations with their users and others.[370]

Significantly, platforms' deployment of artificial intelligence tools trained to detect and automatically suppress harmful content lies at the forefront of concerns that platforms could trample on freedom of expression and other fundamental rights while attempting to comply with EU mandates to remove content that undermines democracy.[371] Social media already rely heavily on AI tools to flag hate speech, terrorist propaganda, extremist incitement, and pornography. Those tools remain far from perfectly accurate. They are also subject to bias, especially given that they are integrated within an overall platform design geared towards furthering the platform's commercial interests, primarily including business models predicated on maximizing user engagement and selling targeted, data-based advertising. While platforms generally thrive on propagating user-engaging outrage, they have incentives to block or demote items that might be especially abhorrent to advertisers, government officials, or most users. Researchers charge that due to those incentives and AI's still limited capacity to grasp context, platforms' automated content moderation systems disproportionately silence minority groups' speech and have mistakenly blocked legitimate protests

369 *See, e.g.*, GIOVANNI DE GREGORIO, DIGITAL CONSTITUTIONALISM IN EUROPE: REFRAMING RIGHTS AND POWERS IN THE ALGORITHMIC SOCIETY (2022); João Pedro Quintais, Naomi Appleman & Ronan Ó Fathaigh, *Using Terms and Conditions to Apply Fundamental Rights to Content Moderation*, 24 GER. L.J. 881, 907 (2022) (noting that "the push for the protection of fundamental rights in the content moderation process can be seen as part of a 'digital constitutionalism' response to the relative dominance of the big social media platforms"); Edoardo Celeste, *Digital Constitutionalism: A New Systematic Theorisation*, 33 INT'L REV. L., COMPUTS. & TECH. 76 (2019); Nicolas Suzor, *Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms*, 4 SOC. MEDIA + SOC'Y, July–Sept. 2018, at 1.

370 The term "digital constitutionalism" has sometimes been used, in a very different sense, to refer to digital platforms' privately defined norms and procedures that roughly parallel state constitutional structures but operate independently of the state or even aim to displace state institutions. The Facebook Oversight Board, sometimes called the Facebook Supreme Court, is a prominent example. That use of the term carries a certain cyberlibertarian theme in that it tends to view digital spaces as distinct jurisdictional realms, external to state law. *See* Róisín Á Costello, Faux Ami? *Interrogating the Normative Coherence of 'Digital Constitutionalism'*, 12 GLOB. CONSTITUTIONALISM 326 (2023) (criticizing the use of the digital constitutionalism label in that cyberlibertarian sense, given the misappropriation of the term constitutionalism to refer to private ordering). I follow, rather, those commentators who use the term "digital constitutionalism" to mean state-imposed due process and individual rights obligations, not platform-directed private ordering.

371 This paragraph draws upon Elkin-Koren, *supra* note 53, at 4–7.

against police violence as well as videos documenting human rights violations.[372]

In that light, the DSA obligates platforms to give "due regard" in the platform's content moderation practices to the fundamental rights of their users and others.[373] In addition, as we have seen, the DSA requires very large platforms to assess and mitigate any negative systematic risks that their recommender system, content moderation practices, or terms of service impede the exercise of fundamental rights.[374] In both contexts, fundamental rights encompass the rights to "freedom of expression and information," "the freedom and pluralism of the media," "non-discrimination," "human dignity," and other rights enshrined in the Charter of Fundamental Rights of the European Union.[375]

Relatedly, the DSA imposes due process obligations on platforms. First, platforms have an obligation of transparency. They must clearly set forth in their terms of service "any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review."[376] They must also set forth, "in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters."[377] Second, platforms must provide certain rights to contest their content moderation decisions. In that regard, platforms that remove, disable, or demote user content, or that suspend a user's account, must provide the affected user with a clear and specific statement of reasons for doing so.[378] The platforms must also establish a user-friendly internal complaint-handling system for lodging complaints, electronically and free of charge, against

372 *See* Daphne Keller, *Facebook Filters, Fundamental Rights, and the CJEU's* Glawischnig-Piesczek *Ruling*, 69 GRUR INT'L 616, 617 (2020) (discussing minority speech).

373 DSA, *supra* note 22, art. 14, ¶¶ 1, 4 at 49.

374 *Id.* arts. 34–35, at 64–66.

375 *Id.* art. 34, ¶ 1(b), at 64–65. For the enumeration of rights, see EU Fundamental Rights Charter, *supra* note 295.

376 DSA, *supra* note 24, art. 14, ¶ 1, at 49.

377 *Id.* art. 27, ¶ 1, at 59.

378 *Id.* art. 17, ¶¶ 1, 3(f), at 51–52. With respect to demotion and other visibility reduction, see Paddy Leerssen, *An End to Shadow Banning? Transparency Rights in the Digital Services Act Between Content Moderation and Curation*, 48 COMPUT. L. & SEC. REV., Apr. 2023, at 1. By empowering users to challenge platforms' demoting of content, not just blocking content, the DSA effects a significant, unprecedented expansion of user rights and may well impact platform content moderation practices. *See* Niklas Eder, Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation 6 (June 29, 2023) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4491365 [https://perma.cc/4SML-C32M].

content moderation decisions taken by the platform.[379] Finally, the platforms must provide a procedure for appeal before an out-of-court dispute settlement body, although decisions of that body are not binding.[380] The complaint system and appeal must be available for affected platform users, as well as for trusted flaggers and others who have submitted a notice of hate speech or other illegal content on the platform that the platform fails to remove.

As Giovanni De Gregorio highlights, such state-imposed digital constitutionalism represents a rejection of the U.S. neoliberal paradigm in which online platforms, as private market actors, are free to determine relations with their users through contract.[381] Digital constitutionalism does not necessarily deny that platforms are private market actors with their own fundamental rights to conduct business in accordance with the EU and national law.[382] For that matter, the European Court of Human Rights has recognized that online platforms have their own right of freedom of expression.[383] But, as expressed in the DSA and other EU regulations, digital constitutionalism views online platforms essentially as quasi-public governing bodies, with obligations to users and others that flow from the centrality of online platforms for the exchange of information and the formation of democratic public opinion.

---

379 DSA, *supra* note 22, art. 20, ¶¶ 1, 3, 4 at 53–54. Small platforms are exempt from this requirement. *Id.* art. 19, ¶ 1, at 53. The Proposed European Media Freedom Act would require very large platforms to accord additional due process rights to media service providers that meet requirements for editorial standards and independence. Of particular import, Article 17 of the proposed Act provides that very large platforms must give such media service providers notice and a statement of reasons prior to removing media content unless that content contributes to a systemic risk of harmful content under the DSA. Proposed Media Freedom Act, *supra* note 25, art. 17.

380 DSA, *supra* note 24, art. 21, ¶ 1, at 54.

381 DE GREGORIO, *supra* note 369, at 278–79. Of note, however, Texas and Florida statutes that impose quasi-common carrier obligations on large social media companies also require those companies to provide users with an individualized explanation for why the social media company is removing the user's content. *See* Brief for United States as Amicus Curiae at 5, 9, NetChoice v. Paxton, No. 22-555 (Aug. 14, 2023), 2023 WL 5280330, at *5, *9. The Texas statute further provides that the social media platform must allow the user to appeal a decision removing the user's content and must address such appeals within 14 days. *Id.* at 9. The Eleventh Circuit struck down Florida's individualized explanation requirement on the grounds that it abridged the social media platforms' First Amendment rights, but the Fifth Circuit upheld Texas' similar requirement. NetChoice, LLC v. Attorney General, Florida, 34 F.4th 1196 (11th Cir. 2022); NetChoice LLC v. Paxton, 49 F.4th 439 (5th Cir.2022). The Supreme Court has granted certiorari to consider whether the laws' individualized-explanation requirements, as well as the laws' content-moderation restrictions, comply with the First Amendment. Moody v. NetChoice, LLC, No. 22-277, 2023 WL 6319654 (Sept. 29, 2023); *see supra* notes 204–205 and accompanying text.

382 Fundamental Rights Charter, *supra* note 295, art. 16, at 12. *See* Eduardo Gill-Pedro, *Whose Freedom Is It Anyway? The Fundamental Rights of Companies in EU Law*, 18 EUR. CONST. L. REV. 183, 183–84 (2022).

383 *See supra* notes 291–**Error! Bookmark not defined.**.

As we have seen, the DSA transposes that understanding into a matrix of legal obligations. It requires that major platforms block illegal hate speech and mitigate harms to democracy on their systems, while aiming to minimize errors and biases in their largely algorithmic content moderation systems that might systematically suppress minority viewpoints and thus distort democratic debate. As such, under the European regulations—unlike under the two competing extremes of current American free speech thought—social media are neither speakers with the absolute prerogative to make algorithmic "editorial decisions" about which user-posted content to amplify nor common carriers with minimal rights to block or demote user-posted speech.

Having said that, the precise parameters of the DSA's digital constitutionalism requirements are far from certain. In particular, under European human rights jurisprudence, fundamental rights are balanced against one another, as well as balanced against and informed by the need to defend and affirmatively bolster democracy. For example, the right to be free from discrimination, including a person's right to be free from disenfranchising hate speech, is commonly understood to stand on par with other persons' or entities' rights of freedom of expression.[384] Likewise, the right to receive factually accurate information from a diversity of sources, as required to exercise effective participation in the democratic process, might limit others' right to disseminate falsehood.[385] It is of yet unclear how platforms are supposed to define fundamental rights—let alone balance them against one another—in fulfilling their obligations to take account of the fundamental rights of their users and others. Nor is it clear whether, under the European doctrine of "horizontal effect," European courts might entertain complaints from platform users—or, for that matter, nonusers who claim to be harmed by content on social media—that the platform has itself violated that person's fundamental rights.[386]

In sum, as commentators have noted, the DSA "mark[s] a 'procedural turn' in European lawmaking: rather than setting forth any bright-line substantive rule on the limits of online freedom of expression, the new Regulation creates a series of procedural obligations and redress

384 *See* PITRUZZELLA & POLLICINO, *supra* note 51, at 56–59, 69.

385 *Id.*

386 Under the European doctrine of horizontal effect, private persons can sometimes be directly required to abstain from interfering with another's fundamental rights, or the State is held to have a positive obligation to protect fundamental rights even against interference by private persons. *See generally* Stephen Gardbaum, *The "Horizontal Effect" of Constitutional Rights*, 102 MICH. L. REV. 387 (2003). For discussion in the context of the DSA, see Quintas, Appelman & Fathaigh, *supra* note 369, at 901–02.

avenues."[387] What is more, the extent to which the DSA requires labor-intensive individualized due process in response to complaints about platform content moderation decisions is unclear. The DSA provides that platforms must issue reasoned decisions in response to such complaints and that such decisions must be "taken under the supervision of appropriately qualified staff, and not solely on the basis of automated means."[388] But as Evelyn Douek has underscored, the vast scale and speed of social media user postings and thus of automated content moderation decisions—amounting to millions of such cases every day—far exceed even the largest platforms' capacity to provide individualized procedural rights that require human review for each disputed case.[389] Daphne Keller provides this telling example: "YouTube . . . currently allows appeals for the roughly 9 million videos it removes every three months. But it does not yet do what the DSA will require: offering appeals for the additional *billion* comments it removes in the same time period."[390] Professor Robert Post, currently serving as Trustee of Meta's Oversight Board, further explains: "During the first quarter of 2022 . . . Facebook alone took down some 151,900,000 pieces of content. These removals resulted in some 2,614,400 appeals. No court has the capacity to oversee this volume of business. No human judgment can operate at this scale."[391]

Hence, the human supervision that the DSA requires simply cannot entail more than overseeing the design and functioning of automated complaint-handling systems or, at most, reviewing a small statistical sample of complaints about particular content moderation decisions. It would be impossible to conduct human review of anything approaching every case. Insistence on a formalist, quasi-judicial, individual human review of each content moderation decision would simply be doomed to failure. Even worse, such insistence would undermine efforts to achieve

387 Pietro Ortolani, *If You Build It, They Will Come: The DSA "Procedure Before Substance" Approach*, *in* PUTTING THE DSA INTO PRACTICE, *supra* note 334 at 151, 154 (quoting Christoph Busch & Vanessa Mak, *Putting the Digital Services Act into Context: Bridging the Gap Between EU Consumer Law and Platform Regulation*, 10 J. EUROPEAN CONSUMER & MKT. L. 109 (2021)); *see also* Eder, *supra* note 378, at 10–17 (pointing out that it is still largely unclear how European fundamental rights will apply to platform content moderation and how the DSA requirement of major platform systemic risk assessments will be implemented).

388 DSA, *supra* note 22, art. 20, ¶¶ 5–6, at 54.

389 Douek, *supra* note 368, at 535–38, 568–84.

390 Daphne Keller, *The European Union's New DSA and the Rest of the World*, *in* PUTTING THE DSA INTO PRACTICE, *supra* note 334, at 227, 231.

391 Post, *supra* note 200.

a more realistic and flexible systemic approach to content moderation, focusing on content moderation algorithms and general trends.[392]

All in all, the DSA requirement that very large platforms assess, mitigate, and render transparent to regulators any systemic risks to civic discourse, electoral processes, and the exercise of fundamental rights, including risks flowing from the platform's content moderation system, offers a far more promising tool for identifying and correcting untoward error and bias than does an individualized complaint procedure involving human review.[393] In that regard, the DSA provides that platforms that restrict user content must submit to the European Commission a description of each such content moderation decision and the statement of reasons for restricting content that the platform provided to the user, typically by automated means.[394] In turn, the Commission operates a publicly available online "DSA Transparency Database," featuring information regarding all reported platforms' content moderation decisions.[395]

At present, that information includes, among other items, keywords and categories of the moderated content, statements of reasons for moderation, the nature of moderation, whether the moderated content was identified by the platform itself or by a trusted flagger, whether the content was identified by an algorithm, and whether the moderation decision was automated. As such, the database is useful for revealing general and comparative platform trends, including, for example, that, as of this writing, the overwhelming majority of restricted content is adult sexual material, with hate speech and risk to public health far behind in second and third place.[396] But, since the data does not include specific descriptions of moderated content or, to protect privacy, any information that could identify the user who posted it, it is far less useful for

392 *See generally* Evelyn Douek, *The Siren Call of Moderation Formalism*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY, *supra* note 12, at 139.

393 *Cf.* Eder, *supra* note 378, at 6–7 (contending that while the new individual remedy under the DSA is an important step forward for user rights, it cannot match the scale of automated decisions and puts the burden of holding platforms accountable on the individual).

394 DSA, *supra* note 24, art. 24, ¶¶ 5, at 58.

395 *DSA Transparency Database*, EUR. COMM'N, https://transparency.dsa.ec.europa.eu [https://perma.cc/PBH8-NHGT].

396 *DSA Transparency Database: Analytics*, EUR. COMM'N, https://transparency.dsa.ec.europa.eu/analytics/keywords [https://perma.cc/7CLB-T8V7]. Alternatively, as Niva Elkin-Koren has proposed, perhaps platforms could be required to run their algorithmic content moderation decisions through an adversarial public AI system, designed to reflect a tradeoff——informed entirely by non-commercial, public values——of protecting democratic discourse based on trustworthy information and promoting freedom of expression. Elkin-Koren, *supra* 53, at 8–11. While normatively attractive, it is unclear how Professor Elkin-Koren's proposal would operate within the regulatory requirements of the DSA and the proposed EU Artificial Intelligence Act. I thank João Pedro Quintais for this observation.

uncovering any algorithmic bias regarding particular messages or speakers.

V. EUROPEAN CONSTITUTIONAL FRAMEWORK FOR SOCIAL MEDIA REGULATION

The European Democracy Act Plan framework for social media regulation would be highly unlikely to pass First Amendment muster if enacted in the United States today. The European regulations obligate platforms to remove hate speech and other speech that is illegal in many European countries but protected by the First Amendment in the United States. Further, although much of the regulation regarding legal but harmful speech is styled as co-regulation, in which platforms are merely encouraged to mitigate systemic risks to civic discourse and electoral processes, the platforms must do so under the watchful eye of European regulators, as a possible predicate to further regulation. As noted above, that considerable government involvement and pressure would almost certainly exceed the kind of informal jawboning that some courts have held constitutional in the United States.[397]

The regulations also raise myriad constitutional issues in Europe. In that regard, the Charter of Fundamental Rights of the European Union, the human rights instrument to which the DSA refers, enshrines in binding EU law both all the fundamental rights set forth in the European Convention on Human Rights and a parallel abuse of rights provision that denies protection to antidemocratic actors.[398] The Charter provides that its enumerated rights shall have the same meaning and scope as the corresponding rights set out in the ECHR.[399] In particular, the Court of Justice of the European Union (CJEU) has ruled that the protection of freedom of expression under Article 11 of The Charter of Fundamental Rights of the European Union shall "be given the same meaning and the same scope" as Article 10 of the ECHR, "as interpreted by the case-law of the European Court of Human Rights."[400]

Within that constitutional framework, European courts have thus far only considered state imposition of liability on online platforms for failing to remove user-posted content that is illegal under European nations' law. Courts have yet to consider regulatory requirements that

397 *See*, *supra* notes 229-231 and accompanying text.

398 *See generally* Stephen Brittain, *The Relationship Between the EU Charter of Fundamental Rights and the European Convention on Human Rights: An Originalist Analysis*, 11 EUROPEAN CONST. L. REV. 482, 482–83 (2015) (discussing the relationship between the two instruments).

399 EU Fundamental Rights Charter, *supra* note 295375, art. 52, ¶ 3, at 21.

400 Case C-345/17, Sergejs Buivids v. Datu Valsts Inspekcija, 2019 ECLI:EU:C:2019:122, ¶ 65.

platforms demote or block access to content that is legal offline but that may be harmful to democracy when amplified on social media.

While far from definitive, decisions of national courts, the European Court of Human Rights, and the CJEU strongly suggest that regulatory bodies may require social media platforms to remove antidemocratic speech that is illegal under European nations' laws. For example, France's so-called "Avia Law" required that online platforms remove clearly illegal content, including content inciting hatred, violence, or discrimination on the basis of ethnicity, religion, gender, sexual orientation, or disability, within twenty-four hours of being notified of it, and content promoting terrorism with one hour of notification.[401] In hearing a challenge to the law before it took effect, France's Constitutional Council acknowledged that, in principle, platforms may be required to remove speech that incites hatred or promotes terrorism.[402] The court held, however, that the obligation to remove illegal speech on the exceedingly short notice set out in the Avia Law would impose an unconstitutional burden on freedom of expression, given it that would not provide enough time for online platforms, which would likely be flooded by user notifications, to adequately assess the legality of the content.[403]

For its part, the European Court of Human Rights has observed that the internet, at one and the same time, "provides an unprecedented platform for the exercise of freedom of expression" and brings the danger that "clearly unlawful speech, including hate speech and speech inciting violence, can be disseminated like never before, worldwide, in a matter of seconds, and sometimes remain persistently available online."[404] Weighing that benefit and danger, the Court ruled in the landmark case, *Delfi AS v. Estonia*, that a commercial online news portal could be held strictly liable for failing to block hate speech and incitement to violence in user comments to its published news articles.[405] The Court noted, however, that the case before it did not necessarily pertain to other types

401 Loi 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet [Law 2020-766 of 24 June 2020 Aimed at Combating Hate Content on the Internet] [Avia Law], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], June 25, 2020, p. 11.

402 Conseil Constitutionnel [CC] [Constitutional Court] June 18, 2020, decision No. 2020-801DC (Fr.); *see* Nicolas Boring, *France: Constitutional Court Strikes Down Key Provisions of Bill on Hate Speech*, LIBR. OF CONG. (June 29, 2020), https://www.loc.gov/item/global-legal-monitor/2020-06-29/france-constitutional-court-strikes-down-key-provisions-of-bill-on-hate-speech/#:~:text=The%20Conseil%20constitutionnel%20agreed%20that,harmful%20to%20freedom%20of%20expression [https://perma.cc/6P7T-79TA].

403 Conseil Constitutionnel [CC] [Constitutional Court] June 18, 2020, decision No. 2020-801DC (Fr.).

404 Delfi AS v. Estonia, App. No. 64569/09, ¶ 110 (June 16, 2015), https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-155105%22]} [https://perma.cc/59GT-UFKN].

405 *Id.* ¶¶ 153, 159.

of online sites, including "a social media platform where [unlike a news portal] the platform provider does not offer any [of its own] content . . . ."[406]

Moreover, in two other cases, *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary* and *Tamiz v. United Kingdom,*[407] the Court found that an imposition of strict liability against a news portal or a social media platform (the Google Blogger Platform) would run contrary to the Article 10 protection of free expression where the content in question was at most trivially defamatory and the platform had promptly removed the offensive posts upon receiving the injured party's complaint. In *Magyar*, the Court noted "that if accompanied by effective procedures allowing for rapid response, [a] notice-and-take-down-system could function in many cases as an appropriate tool for balancing the" plaintiff's reputational interests and the platform's right to free expression.[408] Likewise, in *Tamiz*, the Court noted that European institutions and the United Nations "have all indicated that [Internet platforms] should not be held responsible for content emanating from third parties unless they failed to act expeditiously in removing or disabling access to it once they became aware of its illegality."[409] In sum, social media platforms may be required to remove illegal content of which they have knowledge but they must generally be accorded sufficient time to assess whether putatively illegal content is, in fact, illegal.[410]

Granted, some commentators express concern that European laws that require social media to remove illegal content induce social media to over-police.[411] The CJEU has also warned that social media deployment of automatic content filtering "might not distinguish adequately between unlawful content and lawful content, with the result that its introduction

406 *Id.* ¶ 116.

407 Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, App. No. 22947/13, (Feb. 2, 2016), https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-160314%22]} [https://perma.cc/4WT4-B6GP]; Tamiz v. United Kingdom, App. No. 3877/14, ¶¶ 70, 75–77, 84–85, 89, 91 (Sept. 9, 2017), https://hudoc.echr.coe.int/eng#{%22appno%22:[%223877/14%22],%22itemid%22:[%22001-178106%22]} [https://perma.cc/7DZG-MR9C].

408 *Magyar Tartalomszolgáltatók*, App. No. 22947/13, ¶ 91.

409 *Tamiz*, App. No. 3877/14, ¶ 84.

410 The requisite time that must be accorded to the platform to assess the content might vary depending on the potential danger posed by such content and whether the platform receives notice of the allegedly illegal content from a public authority. Of note, the EU Regulation on Addressing the Dissemination of Terrorist Content Online requires platforms to remove "terrorist content" within one hour after receiving a removal order from the applicable national authority. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on Addressing the Dissemination of Terrorist Content Online, 2021 O.J. (L 172), 79, art. 3, ¶ 3, at 90.

411 *See, e.g.*, Keller, *supra* note 372, at 616–18.

could lead to blocking of lawful communications."[412] Indeed, in a recent case involving Article 17 of the Directive on Copyright in the Digital Single Market, the CJEU ruled that while platforms may be obligated to remove and filter infringing content, they may not be required to implement automatic content filtering systems unless the filtering system narrowly targets identified infringing content and provides adequate safeguards to ensure that users may post lawful content.[413]

The concerns over automatic content filtering and excessive policing inform the DSA requirement that platforms must provide those adversely impacted by content moderation with an opportunity to object. Similarly, Germany's Federal Court of Justice has ruled that Facebook must accord its users a right to file an objection to Facebook's deletion of their posts or suspension of their account.[414]

As discussed above, however, given the massive scale of social media communications, there is no mechanism for carefully, individually assessing each and every communication, certainly if that review must be conducted by humans.[415] At the very least, therefore, the rule cannot be that all communications must remain on the platform unless and until proven to be illegal, following a full and fair hearing in which the user may raise objections to deletion. If that were the case, illegal and antidemocratic communications would rapidly disseminate on social media and the harm they cause would be done before they are deleted. Hence, to the extent a procedure for user objection to content moderation decisions, whether the review is conducted by AI or a human, is constitutionally required, the default should be that the post is blocked—or, as the case may be, that it is rendered substantially less visible—pending resolution of the user's objection.[416] As has famously been observed, "free speech does not mean free reach."[417] Nor, given the

412 Case C-360/10, SABAM v. Netlog NV, ECLI:EU:C:2012:85, ¶ 50 (Feb. 16, 2012).

413 Case C-401/19, Poland v. Parliament & Council, ECLI:EU:C:2022:297, ¶¶ 85-98 (Apr. 26, 2022).

414 Press Release, Bundesgerichtshof, Bundesgerichtshof zu Ansprüchen gegen die Anbieterin eines sozialen Netzwerks, die unter dem Vorwurf der "Hassrede" Beiträge gelöscht und Konten gesperrt hat [Federal Court of Justice on Claims Against the Provider of a Social Network that has Deleted Posts and Blocked Accounts on Charges of "Hate Speech"] (July 29, 2021), https://www.bundesgerichtshof.de/SharedDocs/Pressemitteilungen/DE/2021/2021149.html [https://perma.cc/CH39-3XGJ] [hereinafter Bundesgerichtshof].

415 *See supra* notes 389–391 and accompanying text.

416 That ex post notification and right of redress is the procedure required by the DSA. The German Federal Court of Justice required only ex post notification for removing posts but advance notification for suspending a user's account. *See* Bundesgerichtshof, *supra* note 414.

417 Renée DiResta, *Free Speech is Not the Same as Free Reach*, WIRED (Aug. 30, 2018, 4:00 PM), https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach [https://perma.cc/9XJM-M99B] (emphasis omitted).

serious harms inflicted by antidemocratic speech, can free speech be the right to immediate viral amplification.

Finally, it seems doubtful that European courts would void regulations that require platforms to demote or block speech that is lawful offline but that European legislators and regulators have determined is harmful to democracy when amplified on social media. As noted above, in its rulings on nations' restrictions on paid political advertising, the European Court of Human Rights has reiterated that, as the "ultimate guarantor" of "free and pluralist debate," states are entitled to enact proportionate, targeted regulations to "protect the democratic debate and process from distortion."[418] In the case of paid political advertising, that distortion might stem from "powerful financial groups with advantageous access to [the] media."[419] But the same principle should apply where the negative effects on civic discourse, the electoral process, and, ultimately, fundamental rights stem from amplified disinformation campaigns or bot-propelled, antidemocratic extremism. Further, the Court's statement that "no one must be authorized to rely on the Convention's provisions in order to weaken or destroy the ideas and values of a democratic society" should apply to those who would "flood the zone with shit" no less than to antidemocratic political parties.[420]

VI. CONCLUSION

By contrast with the neoliberal values that animate First Amendment doctrine and American technology policy, militant democracy insists that democratic states both defend against antidemocratic subversion and actively promote reason-based, egalitarian democratic discourse. As such, constitutional courts in European countries and the European Court on Human Rights countenance proportionate limitations on individual rights when necessary to defend democracy. Indeed, the Preamble to the European Convention on Human Rights recognizes that an "effective political democracy" is an essential prerequisite for individuals' human rights.[421]

While many details remain to be fleshed out, the militant democracy-informed regulation of online platforms in the European Union represents the better overall framework for protecting the democratic process against manipulation, disinformation, divisive and debilitating hate speech, violent incitement, and authoritarian

418 *See supra* notes 312–313 and accompanying text.

419 *See supra* notes 312–313 and accompanying text.

420 Refah Partisi v. Turkey, 2003-II Eur. Ct. H.R. 267, 303–04.

421 *See supra* notes 305–306 and accompanying text.

extremism—and, concomitantly, for protecting individual freedoms. To that end, the EU framework strongly induces, if not effectively requires, social media to give prominence to trustworthy sources of information, expressions of broadly inclusive civic solidarity, and reason-based discourse rather than funneling emotionally inflammatory content to maximize user engagement.

Militant democracy principles might serve as a useful regulatory ideal for a reinvigorated, democracy-centered First Amendment jurisprudence. Yet even under neoliberal First Amendment doctrine as it currently stands, EU regulations are likely to inform major social media design and practice in the United States as well as in Europe. To the extent that the EU initiatives succeed in fundamentally redirecting social media to serve as a vital public sphere for reasoned, fact-based democratic debate, three cheers for the Brussels Effect.