The Law of Facebook

Ashutosh Bhagwat[*]

Twenty-six years ago, Eugene Volokh published his seminal article Cheap Speech and What It Will Do, *predicting many of the consequences of the then-brand-new Internet. On the whole, Volokh's tone was optimistic. While many of his predictions have indeed come true, many would argue that his optimism was overstated. To the contrary, in recent years Internet giants generally, social media firms specifically, and Facebook and its CEO Mark Zuckerberg more specifically, have come under sharp and extensive criticism. Among other things, Facebook has been accused of violating its users' privacy, of failing to remove content that constitutes stalking or personal harassment, of permitting domestic and foreign actors (notably Russia) to use fake accounts to manipulate American voters by disseminating false and misleading political speech, of failing to remove content that incites violence, and of excessive censorship of harmless content. Inevitably, critics of Facebook have proposed a number of regulatory solutions to Facebook's alleged problems, ranging from regulating the firm's use of personal data, imposing liability on Facebook for harm caused by content on its platform, treating Facebook as a utility, to even breaking up the company. Given the importance of Facebook, with over two billion users worldwide and a valuation of well over half a trillion dollars, these proposals raise serious questions.*

This Essay will argue that while Facebook is certainly not free of fault, many of the criticisms directed at it are overstated or confused. Furthermore, the criticisms contradict one another, because some of the solutions proposed to solve one set of problems — notably privacy — would undermine our ability to respond to other problems such as harassment,

 Martin Luther King, Jr. Professor of Law and Boochever and Bird Endowed Chair for the Study and Teaching of Freedom and Equality, University of California, Davis School of Law. B.A. 1986 Yale University. J.D. 1990 The University of Chicago. Contact: aabhagwat@ucdavis.edu. Thanks to Jane Bambauer, Alan Chen, Chris Elmendorf, Greg Magarian, Dawn Nunziato, Ann Ravel, Eugene Volokh, Jim Weinstein, and participants in the UC Davis School of Law "Schmooze" for extremely helpful comments. Thanks also to the students at the UC Davis Law Review for putting together this excellent symposium.

incitement and falsehood, and vice versa. More fundamentally, critics fail to confront the simple fact that Facebook and other Internet firms (notably Google) provide, without financial charge, services such as social media, searches, email, and mapping, which people want and value but whose provision entails costs. To propose regulatory "solutions" which would completely undermine the business model that permits these free services without proposing alternatives and without taking into account the preferences and interests of Facebook users, especially in poor and autocratic countries where, for all of its conceded problems, Facebook provides important and even essential services, is problematic at best. Finally, the failure of critics to seriously consider whether the First Amendment would even permit many of the regulatory approaches they propose, all in the name of preserving democracy and civil dialogue, raises questions about the seriousness of some of these critics.

Ultimately, this Essay argues that aside from some limited regulatory initiatives, we should probably embrace humility. This means, first, that unthinkingly importing old approaches such as a utility or publisher model to social media is wrong-headed, and will surely do harm without accomplishing their goals. Other proposals, on the other hand, might "solve" some problems, but at the cost of killing the goose that lays the golden egg. For now, the best path might well be the one we are on: supporting sensible, narrow reforms, but otherwise muddling along with a light regulatory touch, while encouraging/pressuring companies to adopt voluntary policies such as Twitter's recent ban on political advertising, Google's restrictions on microtargeted political ads, and Facebook's prohibitions on electoral manipulation. Before we take potentially dangerous and unconstitutional legislative action, perhaps we should first see how these experiments evolve and work out. After all, social media is less than two decades old and there is still much we need to learn before thoughtful and effective regulation is plausible.

TABLE OF CONTENTS

INTRODUCTION

Twenty-six years ago, Eugene Volokh published his seminal article *Cheap Speech and What It Will Do*.[1] Even today, the article remains an astonishing read. It predicted the practical extinction of physical media for music,[2] the explosion of video streaming,[3] and more broadly the democratization of speech on public affairs.[4] It also predicted the decline of information intermediaries such as the mass media,[5] the possibility of increased extremist speech,[6] and of filter bubbles.[7] On the whole, however, the tone of the article was relentlessly upbeat. From today's perspective, while an extraordinary number of Volokh's practical predictions have come true, his optimism seems more questionable. To the contrary, if there is one topic on which people around the globe appear to have a consensus, it is that the Internet, and social media in particular, is the source of enormous societal problems, including the loss of privacy and the rise of "surveillance capitalism,"[8] harassment and worse of women and minorities,[9] and the systematic manipulation of our democracy.[10] Social media has also been linked to

[1] Eugene Volokh, *Cheap Speech and What It Will Do*, 104 YALE L.J. 1805 (1995).

[2] *Id.* at 1808-14.

[3] *Id.* at 1831-33.

[4] *Id.* at 1833-38.

[5] *Id.* at 1834-36.

[6] *See id.* at 1848.

[7] *See id.* at 1849-50.

[8] SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 9 (2019).

[9] *See* DANIELLE CITRON, HATE CRIMES IN CYBERSPACE 14 (2014).

[10] *See* Alexis C. Madrigal, *What Facebook Did to American Democracy: And Why It Was So Hard to See It Coming*, ATLANTIC (Oct. 12, 2017), https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/ [https://perma.cc/3DHR-HDMY].

the incitement and encouragement of violence in places such as Myanmar.[11] In addition to these sins mainly of omission, social media firms have also been charged with over-censoring harmless and even important material, such as the iconic "Napalm Girl" photograph from the Vietnam War.[12] Far from constituting a vast forum for open, democratic debate, as the Supreme Court recently described social media,[13] social media is widely viewed today as a cesspool.

Furthermore, if social media is the broad culprit, there seems to be a similar consensus that Facebook, and its founder and CEO Mark Zuckerberg, are at the heart of the problem.[14] The Facebook/Cambridge Analytica scandal is the public face of the failure of social media firms to protect user privacy.[15] Facebook was the primary platform used by Russia to manipulate the 2016 presidential election,[16] and Facebook has been under sustained attack by many, including President Joe Biden and Facebook's own employees, for refusing to fact-check political advertisements and take down violent posts by prominent politicians.[17] Facebook, and its subsidiary WhatsApp, were the primary mediums for inciting the pogrom against Rohingya in Myanmar as well as violence

[11] Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. TIMES (Nov. 6, 2018), https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html [https://perma.cc/LWD2-5M2R].

[12] Aarti Shahani, *With 'Napalm Girl,' Facebook Humans (Not Algorithms) Struggle to Be Editor*, NPR (Sept. 10, 2016, 11:12 PM), https://www.npr.org/sections/alltechconsidered/2016/09/10/493454256/with-napalm-girl-facebook-humans-not-algorithms-struggle-to-be-editor [https://perma.cc/2UF5-9UZY].

[13] *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1735-36 (2017).

[14] *See, e.g.*, ROGER MCNAMEE, ZUCKED: WAKING UP TO THE FACEBOOK CATASTROPHE 2 (2019) (accusing Facebook and its leadership of undermining democracy); Kara Swisher, *Zuckerberg Never Fails to Disappoint*, N.Y. TIMES (July 10, 2020), https://www.nytimes.com/2020/07/10/opinion/facebook-zuckerberg.html [https://perma.cc/799Z-4DMF].

[15] *See* Kevin Granville, *Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens*, N.Y. TIMES (Mar. 19, 2018), https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html [https://perma.cc/VBT6-UQQQ].

[16] Mike Isaac & Daisuke Wakabayashi, *Russian Influence Reached 126 Million Through Facebook Alone*, N.Y. TIMES (Oct. 30, 2017), https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html [https://perma.cc/PY2L-APGB]; Madrigal, *supra* note 10.

[17] Sheera Frenkel, Mike Isaac, Cecilia Kang & Gabriel J.X. Dance, *Facebook Employees Stage Virtual Walkout to Protest Trump Posts*, N.Y. TIMES (June 1, 2020), https://www.nytimes.com/2020/06/01/technology/facebook-employee-protest-trump.html [https://perma.cc/Q2GZ-KLYJ]; Cecilia Kang, *Biden Prepares Attack on Facebook's Speech Policies*, N.Y. TIMES (June 11, 2020), https://www.nytimes.com/2020/06/11/technology/biden-facebook-misinformation.html [https://perma.cc/2V7Q-4K7F].

elsewhere.[18] And it was Facebook that deleted the "Napalm Girl" photograph.[19]

Given the ills attributed to Facebook and social media, there have unsurprisingly been many strong calls to impose legal regulations on social media firms to address these problems.[20] At first glance, this seems a sensible reaction, and indeed one that in some places (notably the European Union) have already been heeded to some extent.[21] It is the thesis of this Essay, however, that most of these calls for regulation are a mistake. Indeed, not only are they a mistake, they are also self-contradictory. Many if not most efforts to address one set of problems on social media will exacerbate others. Many are entirely impractical. And worst of all, many are likely unconstitutional. This Essay proposes that instead of jumping in and imposing expansive regulatory regimes on what is after all a very new technology, perhaps humility is the better course. There is certainly room for narrow, carefully considered regulatory reforms. But beyond that, instead of causing unintended harms it might be better to sit back and see how voluntary steps undertaken by social media work out.

Part I examines criticisms of social media, and Facebook in particular.[22] Part II summarizes some of the prominent regulatory proposals recently advanced in response to these problems.[23] Part III identifies the practical and constitutional problems with these proposals.[24] The Essay concludes by identifying some narrow regulatory

[18] Prateek Raj, *We Need to Deal with WhatsApp's Misinformation Problem*, PROMARKET (July 8, 2019), https://promarket.org/2019/07/08/we-need-to-deal-with-whatsapp-misinformation-problem/ [https://perma.cc/4KSG-YDJS]; Stevenson, *supra* note 11.

[19] Shahani, *supra* note 12.

[20] *E.g.*, Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 UC DAVIS L. REV. 1183, 1185-87 (2016) [hereinafter *Information Fiduciaries*]; Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, U. CHI. LEGAL F. 45, 69-74 (2020); Dipayan Ghosh, *Don't Break Up Facebook — Treat It Like a Utility*, HARV. BUS. REV. (May 30, 2019), https://hbr.org/2019/05/dont-break-up-facebook-treat-it-like-a-utility [https://perma.cc/2PQZ-CZKB]; Chris Hughes, *It's Time to Break Up Facebook*, N.Y. TIMES (May 9, 2019), https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html [https://perma.cc/J9GV-R8FH].

[21] Notably through the adoption in 2018 of the General Data Protection Regulation, or GDPR. For the text of the GDPR, see *General Data Protection Regulation: GDPR*, INTERSOFT CONSULTING, https://gdpr-info.eu/ (last visited Jan. 24, 2021) [https://perma.cc/YV7K-FUAR].

[22] *See infra* Part I.

[23] *See infra* Part II.

[24] *See infra* Part III.

reforms that can and should be implemented at this juncture, noting voluntary steps that social media firms are undertaking, and suggesting that for the present, the best course might be to maintain sustained public pressure on tech companies to continue to improve their own content moderation policies.

I. THE ILLS OF SOCIAL MEDIA

Facebook, social media, and other Internet firms have been accused of an enormous range of wrongdoing and mistakes. In this Part, I identify some of the major critiques and their bases. In addition to the intrinsic value of identifying these critiques, there is also value in seeing, in one place, the sheer range of misdeeds attributed to social media, and identifying the inherent tensions between some of these arguments.

A. Privacy/Big Data/The Surveillance State

One of the most consistent critiques of large Internet firms such as Google and Amazon, which also and especially have been leveled against social media giants such as Facebook, target these firms' collection of personal data about their users (and others), and use of that data to control and manipulate our choices. Such data collection and processing practices intrinsically raise serious privacy concerns, but those concerns are exacerbated when Internet firms share personal data with third parties, including the government.

First, the data collection. As Jack Balkin has famously pointed out, Internet firms collect a lot of data about their users.[25] Every time we buy something on Amazon, the firm keeps a record of that purchase. Every time we engage in a Google search, Google tracks the subject matter. Every time we use Gmail to send an email, Google scans and records the content. And every time we post on Facebook, Facebook records the content. Especially for ubiquitous companies such as Google, the information recorded about individuals can be so extensive as to permit the firm to create a robust picture of the lives of particular people. Furthermore, if firms share data with each other, as they sometimes do,[26] they can develop even more extensive pictures of individual lives.

---

[25] Balkin, *Information Fiduciaries*, *supra* note 20, at 1187-94; *see also* Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497, 498-502 (2019) (agreeing with Balkin about the reality of data practices and privacy concerns, but raising doubts about Balkin's proposed solution to the problem).

[26] *Your Data is Shared and Sold . . . What's Being Done About It?*, KNOWLEDGE@WHARTON (Oct. 28, 2019), https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/ [https://perma.cc/8V7G-BS8U].

Furthermore, data collection is not a minor part of these firms' practices. It is rather fundamental to their business model. Especially for advertising-driven firms like Google and Facebook, the ability to sell targeted advertising, where the targeting is based on data collected about individual users, is their business model.[27] Without personal data they would have nothing to sell. And even for firms that sell goods or services such as Amazon, customer data helps them target recommendations and so generate new sales, an enormously valuable feature. There is thus little chance that these firms will cease or reduce their data collecting practices voluntarily.

The harms that are associated with data collection and use are both obvious and complex. To start with is the obvious risk that firms will inadvertently release embarrassing personal information such as infidelity. Even more concerning, and as Balkin points out entirely plausible, is the possibility that firms will threaten to or actually release such information in order to blackmail or discredit critics.[28] And finally, the very fact that these firms (and their employees) know so much about individuals' lives is extremely troubling and frankly "creepy," even if firms do not habitually sell personal data (which they do not, because the data is too valuable to share with potential competitors).

But privacy, as in the release of private information, is not the only concern raised by data collection. Consider targeted advertising.[29] At first, targeted advertising of goods and services seems at most annoying, and sometimes useful. But personal information can be used to influence and manipulate choices beyond the commercial sphere. As Balkin and Jonathan Zittrain recount, during the 2010 election Facebook conducted an experiment in which it added graphics to some users' news feeds that were designed to encourage them to vote.[30] The

---

[27] *See* Mark Zuckerberg, *Understanding Facebook's Business Model*, FACEBOOK (Jan. 24, 2019), https://about.fb.com/news/2019/01/understanding-facebooks-business-model/ [https://perma.cc/Q8DH-99LB].

[28] *See* Balkin, *Information Fiduciaries*, *supra* note 20, at 1187-88. Note that while actual blackmail would likely be subject to criminal prosecution, the simple release of discrediting information is probably not.

[29] For a sense of just how precisely Facebook can target advertising based on individuals' characteristics, see *Ad Targeting: Help Your Ads Finds the People Who Will Love Your Business*, FACEBOOK FOR BUSINESS, https://www.facebook.com/business/ads/ad-targeting (last visited Jan. 24, 2021) [https://perma.cc/4WYR-G8HY].

[30] Balkin, *Information Fiduciaries*, *supra* note 20, at 1188-89 (internal citation omitted); *see* Jonathan Zittrain, *Facebook Could Decide an Election Without Anyone Ever Finding Out*, NEW REPUBLIC (June 1, 2014), http://www.newrepublic.com/article/117878/information-fiduciary-solution-facebook-digital-gerrymandering [https://perma.cc/F3QV-C44C] [hereinafter *Facebook*].

impact of this post was small (targeted users were 0.39 percent more likely to vote), but given Facebook's enormous user base, that can translate into a lot of votes, potentially enough to swing a close election.[31] The risk, of course, is that because Facebook can pretty reliably predict users' political inclinations based on their personal data, it could manipulate election results by encouraging turnout only of voters of a particular political persuasion.[32] This possibility poses a rather more serious problem than a consumer being convinced to buy a pair of shoes they do not need.

Ultimately, the basic social risk posed by the collecting and processing of massive amounts of personal data is the creation of a society based on what Shoshana Zuboff has labeled "surveillance capitalism."[33] The basic idea here is that in the modern digital economy, information about human experience (i.e., personal data) is the key input into most economic activity. The result is that those who possess and control that data, primarily the major technology companies such as Google, Facebook, and Amazon, have the power to predict and manipulate a huge range of human choices. Their primary motivations in doing so are, of course, commercial; surveillance capitalism is, after all, *capitalism*. But as noted earlier, the power to extend such manipulation and control into social and political spheres certainly exists. Furthermore, because the tech sector is highly concentrated and (unlike the manufacturing firms that dominated earlier versions of capitalism) tends to employ very small numbers of highly educated people, the concentration of power entailed by this system is far more dramatic than in earlier systems.[34] As such, the rise of Big Data has fundamentally altered the structure of our societies, making them less democratic and in some sense less free.

B. Cyberstalking, Trolling, and Harm to the Vulnerable

Another important harm associated with the rise of the Internet, and especially social media, is the use of such technology by bad actors to cause serious, personal harm, often to vulnerable members of society such as women, people of color, and LGBTQ individuals. As Danielle Citron has persuasively (and prophetically) argued since at least 2009, the Internet and especially social media have enabled cyber mobs to

[31] Jonathan Zittrain, Response, *Engineering an Election*, 127 HARV. L. REV. F. 335, 336 (2014) [hereinafter *Engineering an Election*].

[32] *Id.*

[33] ZUBOFF, *supra* note 8, at 9.

[34] *Id.* at 500-01.

harass, threaten, and ultimately silence individuals with impunity because of the extraordinary ease of organizing such behavior.[35] Internet platforms have also become deluged with vicious hate speech, often accompanied with calls for violence.[36] One particularly grotesque pattern that has emerged is "revenge porn": individuals (usually men) posting voluntarily-shared intimate pictures of their former partners (usually women) for the sole purpose of causing them social and psychological damage.[37] These behaviors cause serious, real harm to real, vulnerable individuals, and constitute some of the worst individual misuses of the Internet and social media. Yet they are ubiquitous.

An important factor in the kind of misuse of social media described above is Section 230 of the Communications Decency Act, enacted by Congress in 1996.[38] Section 230 famously provides that Internet platform providers shall not be treated as "the publisher or speaker" of information posted by third parties, thereby relieving them of any liability for such speech and the harm it causes.[39] Less famously, Section 230 also relieves such platforms of liability "for good-faith filtering or blocking of user-generated content."[40] The impact of Section 230 is, of course, that platform providers such as Facebook have no direct, economic incentive to block harmful content, since they are not liable for it.

On the other hand, Section 230 does, through the second provision described above, affirmatively *encourage* platforms to block harmful content, without fear of liability if they over-block unintentionally. And in fact, over the years the major social media platforms such as Facebook and Twitter have developed complex machineries for filtering and blocking bad content such as hate speech, incitements to violence, pornography (whether consensual or not), and the like, presumably because some combination of public pressure and their own business interests in retaining users, drive them to do so. As Kate Klonick has extensively described, these processes are extraordinarily complex and

---

[35] Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61, 62 (2009); Danielle Keats Citron & Neil M. Richards, *Four Principles for Digital Expression (You Won't Believe #3!)*, 95 WASH. U. L. REV. 1353, 1364-66 (2018).

[36] Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1460-63 (2011).

[37] Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 346 (2014).

[38] 47 U.S.C. § 230 (2018).

[39] Citron & Franks, *supra* note 20, at 5-6.

[40] *Id.* at 5.

detailed.[41] This is not to say that Facebook and other social media do not face criticism for how they go about filtering, for example hate speech — they certainly do.[42] Indeed, in an audit commissioned by Facebook itself and published in July 2020, the company was sharply criticized for its failures in these areas.[43] The policies of social media companies regarding control of harmful speech thus remain a work in progress, and one whose efficacy is heavily disputed.[44] What is clear is that while much harmful content is blocked by such companies, the efforts are incomplete, and have been criticized as insufficient.

C. Election Manipulation and False Political Speech

As everyone in the United States knows, during the 2016 presidential campaign social media platforms were used extensively by a number of actors, including foreign governments, to spread "fake news" and otherwise manipulate American voters.[45] Their goals were varied, as were their methods. As Abby Wood and Ann Ravel discuss, the tools used to manipulate voters in 2016 ranged from outright disinformation — "fake news" — to more subtle attempts to increase social divisions on hot button political issues by using bots and other devices to spread stories quickly.[46] And motivations varied from the mercenary, to the nihilistic, to the Russian goals of electing Donald Trump and discrediting American democracy.[47] The effectiveness of these

[41] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1625-30 (2018); *see also* Simon Adler, *Post No Evil*, RADIOLAB (Aug. 17, 2018), https://www.wnycstudios.org/podcasts/radiolab/articles/post-no-evil [https://perma.cc/8WU5-3NFC].

[42] Charlie Warzel, *When a Critic Met Facebook: 'What They're Doing is Gaslighting*,' N.Y. TIMES (July 9, 2020), https://www.nytimes.com/2020/07/09/opinion/facebook-civil-rights-robinson.html [https://perma.cc/JYD2-PE6G].

[43] FACEBOOK'S CIVIL RIGHTS AUDIT — FINAL REPORT 42-58 (July 8, 2020), https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf [https://perma.cc/D97A-RLMJ].

[44] *See* Davey Alba, *Facebook Must Better Police Online Hate, State Attorneys General Say*, N.Y. TIMES (Aug. 5, 2020), https://www.nytimes.com/2020/08/05/technology/facebook-online-hate.html [https://perma.cc/EX7K-GKJN].

[45] Robert Yablon, *Political Advertising, Digital Platforms, and the Democratic Deficiencies of Self-Regulation*, 104 MINN. L. REV. HEADNOTES 13, 14 & n.5 (2020) (citing Nathaniel Persily, *Can Democracy Survive the Internet?*, 28 J. DEMOCRACY 63, 67-71 (2017); Abby K. Wood & Ann M. Ravel, *Fool Me Once: Regulating "Fake News" and Other Online Advertising*, 91 S. CAL. L. REV. 1223, 1229-34 (2018)).

[46] Abby K. Wood & Ann M. Ravel, *Fool Me Once: Regulating "Fake News" and Other Online Advertising*, 91 S. CAL. L. REV. 1223, 1229-32 (2018).

[47] *Id.*; *see also* Derek E. Bambauer, *Information Hacking*, 2020 UTAH L. REV. 987, 987-94 (summarizing various Russia-backed disinformation campaigns in 2016).

techniques depended crucially on being able to microtarget specific audiences deemed likely to be receptive to particular messages.[48] In particular, there is strong evidence that Russian manipulation was especially targeted at particular segments of African American voters, seeking to discourage them from voting for Hilary Clinton.[49] In many ways, the criticism of social media companies, especially Facebook, for permitting their platforms to be hijacked in this way are the most serious of any, because of the very real (if unprovable) possibility that these efforts influenced the result of the 2016 election.

Since 2016, social media companies have undoubtedly made significant efforts to forestall similar abuses in the future.[50] Most strikingly, Twitter (as well as a number of less significant platforms) has banned all political advertising on its platform on the grounds that misleading or fake political advertising threatens democracy, and that political speech should obtain audiences based on its popularity, not the purchasers' deep pockets.[51] Notably, however, Facebook and Google, who between them control the lion's share of the online advertising market, have not followed Twitter's example.[52] Instead, they have focused on restricting rather than banning political advertising. Google in particular applies its standard content rules to political advertising, including prohibitions on incitement of violence, on hate speech, and on profanity, *and* fact-checking ads.[53] Google has also placed limits on microtargeting of election advertising, permitting only a few factors (age, gender, and zip code),[54] on the (widely shared) belief that microtargeting is an especially effective way to spread disinformation and sow political divisions.

Facebook, however, has not gone nearly as far as its rivals. It does impose most of its content-based rules, including bans on incitement and hate speech, on political ads, and indeed recently announced a significant tightening of these restrictions, with a particular focus on preventing voter suppression.[55] However, Facebook decided *not* to

[48] Wood & Ravel, *supra* note 46, at 1231.

[49] Scott Shane & Sheera Frenkel, *Russian 2016 Influence Operation Targeted African-Americans on Social Media*, N.Y. TIMES (Dec. 17, 2018), https://www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html [https://perma.cc/4LFD-YRZ6].

[50] Yablon, *supra* note 45, at 17-30.

[51] *Id.* at 17-19.

[52] *Id.* at 19-21.

[53] *Id.* at 21-22.

[54] *Id.* at 27.

[55] Mark Zuckerberg, FACEBOOK (June 26, 2020, 11:25 AM), https://www.facebook.com/zuck/posts/10112048980882521 [https://perma.cc/J4QQ-PMWF].

restrict targeted political advertising or to fact-check political ads (as opposed to commercial ads);[56] and in May 2020 Mark Zuckerberg confirmed that Facebook would continue not to fact-check political posts by politicians, in the face of Twitter's decision to begin fact-checking and labeling such posts.[57] Instead, Facebook has largely limited itself to policies (adopted by other tech firms as well) that seek to block foreign purchases of political ads, and to transparently disclose the content, purchasers, and targets of such ads.[58] Facebook did, however, on September 3, 2020 announce a significant set of restrictions designed specifically to protect the integrity of the 2020 election, including a flat ban on new political ads in the week before the election, further strengthening of its rules against voter suppression, and a commitment to blocking efforts by candidates to falsely claim electoral victory.[59] Finally, on October 7, 2020 Facebook announced that it would prohibit all political and issue advertising once polls close on Election Day.[60] These steps move Facebook more in the direction of its rivals such as Twitter; but notably, they are temporary measures only, focused specifically on the period immediately prior to and following November 3, 2020, Election Day.

In short, despite substantial progress since 2016, tech companies remain subject to intense criticism and pressure regarding their efforts to prevent electoral manipulation. Facebook in particular has been sharply attacked by many, including presidential candidate Senator Elizabeth Warren, for its refusal to fact-check ads by politicians.[61]

---

[56] Tony Romm, Isaac Stanley-Becker & Craig Timberg, *Facebook Won't Limit Political Ad Targeting or Stop False Claims Under New Ad Rules*, WASH. POST (Jan. 9, 2020, 11:24 AM), https://www.washingtonpost.com/technology/2020/01/09/facebook-wont-limit-political-ad-targeting-or-stop-pols-lying/ [https://perma.cc/SC5B-CCC7].

[57] Yael Halon, *Zuckerberg Knocks Twitter for Fact-Checking Trump, Says Private Companies Shouldn't Be 'the Arbiter of Truth*,' FOX NEWS (May 27, 2020), https://www.foxnews.com/media/facebook-mark-zuckerberg-twitter-fact-checking-trump [https://perma.cc/ZY3T-SERB].

[58] Yablon, *supra* note 45, at 24-26, 28-30.

[59] Mike Isaac, *Facebook Moves to Limit Election Chaos in November*, N.Y. TIMES (Sept. 3, 2020), https://www.nytimes.com/2020/09/03/technology/facebook-election-chaos-november.html [https://perma.cc/HPF9-WP46]; Press Release, Facebook, New Steps to Protect the US Elections (Sept. 3, 2020), https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/ [https://perma.cc/R5ZT-A9TU].

[60] Mike Isaac, *Facebook Widens Ban on Political Ads as Alarm Rises Over Election*, N.Y. TIMES (Oct. 7, 2020), https://www.nytimes.com/2020/10/07/technology/facebook-political-ads-ban.html [https://perma.cc/UVN2-P3KK].

[61] Cecelia Kang & Thomas Kaplan, *Warren Dares Facebook with Intentionally False Political Ad*, N.Y. TIMES (Oct. 12, 2019), https://www.nytimes.com/2019/10/12/technology/elizabeth-warren-facebook-ad.html [https://perma.cc/CG7F-T3MK].

Furthermore, even aside from outright falsehoods, there remains plenty of potential even in the face of existing policies for bad actors to use social media to sow political divisions and deceptions, given the enormous difficulties in defining what exactly constitutes problematic speech seeking to sow disagreement, as opposed to core political speech criticizing ideological opponents. There is every reason to expect, therefore, that the problem of false and divisive political speech on social media will remain a serious point of contention.

D. Incitement of Violence

Despite the fact that Facebook, and all other major social media platforms,[62] prohibit posts inciting or glorifying violence, there is also clear evidence that in the past, social media has been used to incite actual violence. In particular, Facebook itself has conceded that its platform was used to incite unspeakable violence against the Muslim, Rohingya minority in Myanmar,[63] and there is clear evidence that Facebook posts have contributed to violence in other countries, including Sri Lanka and the Philippines.[64] Indeed, in Myanmar there is evidence that the violence against the Rohingya was orchestrated by the military through a systematic campaign on Facebook.[65] The result was a humanitarian disaster, forcing 700,000 Rohingya to flee Myanmar in what the United Nations has called "a textbook example of ethnic cleansing."[66]

Given Facebook's Terms of Service banning incitement of violence,[67] how could these things have happened? The answer is simple: Facebook proved unable to enforce its own policies. Nor is this surprising. Facebook operates in an enormous number of countries, and in an enormous number of languages. Reuters reported in 2019 that Facebook was available in 111 languages supported by Facebook, and

---

62 "Major" is an important modifier, as there are clearly more niche internet platforms that do not block violent content.

63 Stevenson, *supra* note 11.

64 *Id.*; Amanda Taub & Max Fisher, *Where Countries Are Tinderboxes and Facebook Is a Match*, N.Y. TIMES (Apr. 21, 2018), https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html [https://perma.cc/R3QH-YPPH].

65 Paul Mozur, *A Genocide Incited on Facebook, with Posts from Myanmar's Military*, N.Y. TIMES (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html [https://perma.cc/S9CY-ZJ6B].

66 *Id.*

67 *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/credible_violence (last updated 2021) [https://perma.cc/5U8J-Y3K7].

another thirty-one without support.[68] To be able to track content at this scale is a huge task, and one that inevitably will sometimes fail. In Myanmar in particular, Facebook's problems appear to have been caused by a lack of Burmese-speaking content-moderators and the fact that Facebook does not have local offices in Myanmar.[69] In Sri Lanka Facebook's problems similarly appear to be linked to a dearth of Sinhalese-speaking moderators.[70] As a consequence, while actual incitement of violence via Facebook appears to be relatively rare in the United States, where content is closely monitored, that is not true in other countries, especially ones which do not speak widely-spoken languages.

Furthermore, incitement of violence is not limited to Facebook's primary platform, the Facebook app; it also has proven an issue on WhatsApp, Facebook's messaging platform. In India in particular — a place where, this author can testify, WhatsApp is ubiquitous — the platform has been used by politicians, especially those associated with the ruling Bharatiya Janata Party ("BJP"), to spread falsehoods among supporters, to rile up ethnic and religious divisions, and occasionally trigger violence by spreading false stories (especially about made-up atrocities by Muslims).[71] WhatsApp is particularly difficult to police because it is end-to-end encrypted, which means that no one outside of a messaging group, including law enforcement and Facebook itself, can monitor messages. WhatsApp also permits effective anonymity by allowing messages to be sent and received based only on cell phone numbers, which further complicates efforts to find out who has started or spread false information.[72]

In short, there is strong evidence that both Facebook itself, and its messaging app WhatsApp, have in recent years been used to incite

---

[68] Maggie Fick & Paresh Dave, *Facebook's Flood of Languages Leave It Struggling to Monitor Content*, REUTERS (Apr. 23, 2019, 12:01 AM), https://www.reuters.com/article/us-facebook-languages-insight/facebooks-flood-of-languages-leave-it-struggling-to-monitor-content-idUSKCN1RZ0DW [https://perma.cc/4V8Y-KEEM].

[69] Steve Stecklow, *Why Facebook Is Losing the War on Hate Speech in Myanmar*, REUTERS (Aug. 15, 2018, 3:00 PM), https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ [https://perma.cc/6P2A-DFQM].

[70] Taub & Fisher, *supra* note 64.

[71] Vindu Goel, *In India, Facebook's WhatsApp Plays Central Role in Elections*, N.Y. TIMES (May 14, 2018), https://www.nytimes.com/2018/05/14/technology/whatsapp-india-elections.html [https://perma.cc/8A75-6QPJ]; Elyse Samuels, *How Misinformation on WhatsApp Led to a Mob Killing in India*, WASH. POST (Feb. 21, 2020, 12:00 AM), https://www.washingtonpost.com/politics/2020/02/21/how-misinformation-whatsapp-led-deathly-mob-lynching-india/ [https://perma.cc/3R76-GVD9].

[72] Goel, *supra* note 68.

violence. Regarding Facebook, it would be at least theoretically possible for the company to block such posts, but the enormous number of languages used on Facebook makes such policing extremely challenging. With respect to WhatsApp, on the other hand, its encryption (which is a feature, not a bug, it should be noted) makes such policing effectively impossible, though as we shall see, WhatsApp has taken steps to at least slow the speed at which posts circulate.

E. Excessive Censorship

Each of the criticisms of Facebook discussed until now come down to accusations that Facebook censors too little, by permitting false, harmful, or dangerous posts to remain on its platform. Facebook has also, however, been subject to the opposite criticism, that it censors too much. As discussed in the Introduction, perhaps the most famous example of this occurred in 2016, when a Norwegian writer posted on his Facebook account the famous "Napalm Girl" photograph (which won a Pulitzer Prize), showing terrified children fleeing a napalm attack during the Vietnam War. Facebook blocked the post and suspended the writer's account because one of the children, a nine-year-old girl, was naked and therefore violated Facebook's prohibitions on nudity and child pornography.[73] Ultimately, under strong public pressure, Facebook reversed its decision, but did not truly apologize.[74]

While a particularly striking example, the deletion of "Napalm Girl" is not an isolated incident. To the contrary, as discussed in detail in a thoughtful episode of the podcast Radiolab, Facebook content moderators face such difficult questions constantly.[75] Particularly fascinating is the description of Facebook's early struggles with how to handle pictures of mothers breastfeeding their children, and the huge number of semi-arbitrary rules this one issue generated.[76] As with the Napalm Girl controversy, the underlying problem here is that while few question Facebook's general policy of prohibiting nudity, specific applications can raise very difficult line-drawing questions and Facebook (perhaps understandably given the under-censorship criticisms described earlier) has a tendency to err on the side of censorship. Nor is this problem limited to nudity. Similar issues arise in the context of violent speech, when Facebook pulls down posts by

[73] Shahani, *supra* note 12.

[74] *Id.*

[75] *See* Adler, *supra* note 41.

[76] *Id.*

nonprofit organizations documenting atrocities, because the posts (naturally) contain violent content.[77]

In addition to the line-drawing issues just discussed, over-censorship issues also arise because today most content moderation on Facebook is automated, creating inevitable problems because algorithms have a hard time evaluating context. Thus, the nudity and violence problems just discussed are often the product of automated content moderation. There have also been innumerable instances of Facebook blocking posts containing phrases like "COVID is a hoax," even when the phrase is used for satire.[78] Indeed, the whole problem of distinguishing actual falsehood or hate speech from satire is one rife with problems for automated systems; but given Facebook's likely tendency to err on the side of caution, the result is predictably to over-censor.

In recent years, the over-censorship attacks on Facebook have also taken on a political tinge. As far back as the 2016 presidential campaign, conservative commentators began accusing Facebook of a left-leaning bias in its selection of "trending" news articles.[79] More recently, during Mark Zuckerberg's testimony before Congress in July 2020, Republican members of Congress repeatedly accused Facebook of disproportionately targeting conservative content for blocking, echoing longstanding similar claims made by a number of prominent Republican political leaders.[80] Soon thereafter, in early August, Facebook deleted a post by President Trump's campaign linking to a video in which Trump

[77] *See "Video Unavailable": Social Media Platforms Remove Evidence of War Crimes*, HUMAN RIGHTS WATCH (Sept. 10, 2020), https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes# [https://perma.cc/TR9U-RDMA]; *see also Social Media Platforms Remove War Crimes Evidence*, HUMAN RIGHTS WATCH (Sept. 10, 2020, 1:00 AM), https://www.hrw.org/news/2020/09/10/social-media-platforms-remove-war-crimes-evidence# [https://perma.cc/V9U3-7BUF].

[78] *See* James Vincent, *AI Won't Relieve the Misery of Facebook's Human Moderators*, VERGE (Feb. 27, 2019, 12:41 PM), https://www.theverge.com/2019/2/27/18242724/facebook-moderation-ai-artificial-intelligence-platforms [https://perma.cc/PMZ3-RGBG].

[79] John Herrman & Mike Isaac, *Conservatives Accuse Facebook of Political Bias*, N.Y. TIMES (May 9, 2016), https://www.nytimes.com/2016/05/10/technology/conservatives-accuse-facebook-of-political-bias.html [https://perma.cc/6ZHX-2UQN]. Ironically, Facebook faces recent allegations that it acted in exactly the opposite fashion, tweaking its algorithms to reduce the visibility of left-wing publications such as Mother Jones, in order to appease Republicans. *See* Monika Bauerlein & Clara Jeffery, *Facebook Manipulated the News You See to Appease Republicans, Insiders Say*, MOTHER JONES (Oct. 21, 2020), https://www.motherjones.com/media/2020/10/facebook-mother-jones/ [https://perma.cc/8F44-DJYE].

[80] David McCabe & Cecelia Kang, *Lawmakers from Both Sides Take Aim at Big Tech Executives*, N.Y. TIMES (July 29, 2020, 6:44 PM), https://www.nytimes.com/live/2020/07/29/technology/tech-ceos-hearing-testimony#republicans-focused-on-bias-concerns-about-platforms [https://perma.cc/S8VW-9HYC].

had said that children were "virtually immune" from COVID-19, on the grounds that the post violated its policies against COVID misinformation (soon after this Twitter blocked the Trump campaign's account for linking to the same video).[81] In response, the White House deputy national press secretary accused Facebook and other Silicon Valley firms of "flagrant bias against this president, where the rules are only enforced in one direction."[82] In short, despite the lack of any empirical evidence supporting this claim, Facebook and other social media firms are regularly accused by conservative politicians and commentators of over-censoring conservative content.

The final over-censorship controversy discussed here is deeply ironic, but also deeply telling, as it results directly from Facebook's own content moderation rules. Facebook has long exempted posts by public figures from some of its content moderation rules, including in the past permitting some forms of what would otherwise be treated as hate speech. And while Facebook has recently substantially tightened those rules with respect to posts by politicians,[83] Facebook continues to refuse to fact-check posts by political figures, including notably President Trump.[84] The result of this dual approach, however, is that if a private individual reposts language used by a political figure, even for satirical purposes, the private individual's post might be blocked while the politician's post stays up. Stanford Law Professor Mark Lemley reports that this is precisely what happened to him, when he ironically quoted President Trump's comment about "N*STY women."[85] And Newsweek similarly reports that Facebook blocked an account that simply copied President Trump's social media posts word-for-word, as violating Facebook Terms of Service.[86] The criticism of Facebook in this regard is two-fold. First, it seems intrinsically unfair to apply different rules to different speakers, and worse, to favor powerful over less powerful speakers. And second, during the 2020 presidential campaign, this policy seemed to directly benefit President Trump over his critics (yet

[81] Cecilia Kang & Sheera Frenkel, *Facebook Removes Trump Campaign's Misleading Coronavirus Video*, N.Y. TIMES (Aug. 5, 2020), https://www.nytimes.com/2020/08/05/technology/trump-facebook-coronavirus-video.html [https://perma.cc/C8DG-4RAT].

[82] *Id.*

[83] *See* Zuckerberg, *supra* note 55.

[84] Halon, *supra* note 57.

[85] Private Correspondence with Mark Lemley (on file with the author).

[86] Jason Murdock, *Facebook Account Copying Trump's Posts Word-for-Word Gets Flagged for Inciting Violence*, NEWSWEEK (June 12, 2020, 6:48 AM), https://www.newsweek.com/facebook-donald-trump-suspendthepres-experiement-post-tweet-flagged-inciting-violence-1510418 [https://perma.cc/MV8D-B79H].

another level of irony, given Trump's accusation that social media firms are biased *against* him).

II. PROPOSED SOLUTIONS TO THE ILLS OF SOCIAL MEDIA

Given the broad and consistent criticisms that social media firms and other Internet giants have face in recent years, there have unsurprisingly been a range of legal responses that have been proposed, and in some instances enacted, to curb perceived problems. In this Part, I will lay out some of the major such proposals and responses.

A. Information Fiduciaries, Data Protection, and the Nuclear Option

In response to the privacy issues posed by the collection, retention, and processing of person data by tech companies discussed in Part I.A above, three prominent responses have emerged. One is Jack Balkin's proposal that such companies be treated as "information fiduciaries" with ethical obligations to the subjects of data.[87] The second is the adoption of data collection and processing regulations, most famously by the European Union and the State of California.[88] Finally, and most radically, one of the co-founders of Facebook has argued that the only solution to these issues is to "break up" the firm.[89] Each of these will be discussed briefly in turn.

Since about 2014,[90] and in more detail in a series of articles beginning in 2016 (two of which were published in this journal),[91] Professor Jack Balkin of the Yale Law School has argued that large tech companies who collect and control sensitive personal information should be treated by the law as "information fiduciaries." The idea here, in Balkin's own words, is that when a firm "collects sensitive information about the client that might be used to the client's disadvantage . . . [and] the client is not well-equipped to understand and monitor the [firm's]

[87] Balkin, *Information Fiduciaries*, *supra* note 20, at 1186.

[88] *See supra* notes 21–25 and accompanying text.

[89] Chris Hughes, *It's Time to Break Up Facebook*, N.Y. TIMES (May 9, 2019), https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html [https://perma.cc/W7MV-2FFD].

[90] Jack M. Balkin, *Information Fiduciaries in the Digital Age*, BALKINIZATION (Mar. 5, 2014, 4:50 PM), https://balkin.blogspot.com/2014/03/information-fiduciaries-in-digital-age.html [https://perma.cc/L56D-GQUL].

[91] Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 UC DAVIS L. REV. 1149, 1160-63 (2018) [hereinafter *Free Speech in the Algorithmic Society*]; Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2047-54 (2018); Balkin, *Information Fiduciaries*, *supra* note 20, at 1186.

operation,"[92] the firm should be treated as a fiduciary. This in turn means that the fiduciary has an ethical and legal obligation to "act in good faith towards their clients, particularly with respect to the information they learn about their clients in the course of the relationship."[93] In essence, Balkin is analogizing the relationship between tech firms and their users to doctor/patient and lawyer/client relationships.[94] Because tech firms such as search engine providers and social media firms collect sensitive user information in similar ways to traditional professionals, their relationships with their users should be subject to similar regulatory restrictions.[95] In particular, Balkin asserts (borrowing from Robert Post) that because communications between professionals and patients/clients, like interactions between tech firms and users, are not part of "public discourse," they generally fall outside robust First Amendment protections.[96]

It should be noted, however, that Balkin concedes that his analogy is imperfect. In particular, he agrees that the obligations of tech companies are clearly less onerous than those of traditional professionals since tech companies rely on monetizing personal data, unlike doctors and lawyers, and because the level of expert service and protection people expect from professionals has no analogy in the tech world.[97] Nonetheless, he insists that the basic principles that fiduciaries act in good faith, and do not betray their users/patients/clients, apply to tech firms in the same way as traditional professionals.

There are many things to be said, positive and negative, about Balkin's "information fiduciary" approach. The most obvious positive aspect of his proposal is that it acknowledges the need to preserve tech firms' business models, which require monetizing data through advertising if free services such as search, mapping, and social media are to be preserved. But at the same time, it seeks to redress the potential for misuse of information threatened by the fact that tech firms' monetary customers — advertisers — do not have the same interests as their users. There are, however, objections, both legal and practical, to Balkin's proposal, which I will discuss in the next Part. For now, the key is to recognize that the treatment of tech firms as information fiduciaries opens the door to some regulation of their data practices,

[92] Balkin, *Free Speech in the Algorithmic Society*, *supra* note 91, at 1160.
[93] *Id.* at 1160-61.
[94] *Id.* at 1161.
[95] *Id.* at 1162.
[96] *Id.* at 1161 (citing Balkin, *Information Fiduciaries*, *supra* note 20, at 1210-11).
[97] *Id.* at 1162-63.

especially disclosure, without seeking to undermine their ability to exist.

While Balkin's "information fiduciary" proposal seeks to justify data regulation, particularly within the United States, against a First Amendment challenge, in recent years important steps have been taken to actually implement such rules. Most significantly, in May 2018 the European Union's General Data Protection Regulation ("GDPR") came into effect.[98] As Meg Leta Jones and Margot Kaminski discuss, while the GDPR has been described in the United States as a law focused on consumer consent, this is in fact not entirely accurate.[99] Rather than being a traditional data privacy law on the American model (which does typically focus exclusively on consent), the GDPR regulates *data* and data processing.[100] The core of the GDPR, contained in Article 6, is that holders of data may personally process it *only* for one of six listed reasons — all other processing is illegal.[101] And this restriction applies to all holders of data, not just the original collector or the firm with which the subject of the data has a relationship.[102] Furthermore, while consent is indeed the first justification for data processing listed in Article 6, it is far from the only one. To the contrary, Jones and Kaminski argue that the sixth justification — that the "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party"[103] — is the one that most firms are likely to rely upon, at least in part because consent requirements in the GDPR are far more onerous than in the typical American privacy law.[104] Finally, while this provision would appear to permit extensive data processing by tech firms in the course of selling advertising or other business processes, it is important to note that the ability to process data under this justification may be "overridden by the interests of fundamental rights and freedoms of the data subject."[105]

In addition to restricting data processing, the GDPR also grants important rights to the individual subjects of personal data. While a full

[98] General Data Protection Regulation 2016/679, 2016 O.J. (L 119), https://gdpr-info.eu/ [https://perma.cc/3ZZ3-2AF4] [hereinafter GDPR].

[99] Meg Leta Jones & Margot E. Kaminski, *An American's Guide to the GDPR*, 98 DENVER L. REV. 93, 95 (2021).

[100] *Id.* at 106-08.

[101] GDPR, *supra* note 98, art. 6(1); Jones & Kaminski, *supra* note 99, at 108.

[102] Jones & Kaminski, *supra* note 99, at 107.

[103] GDPR, *supra* note 98, art. 6(1)(f).

[104] *See* Jones & Kaminski, *supra* note 99, at 108-09. Whether this prediction is correct, or alternatively whether firms end up relying more on consent, as in the past, remains to be seen.

[105] GDPR, *supra* note 98, art. 6(1)(f).

description of those rights is not possible in this space, important elements include extensive rights of detailed notification regarding data collection, storage and processing;[106] a right to access stored and/or processed data;[107] a right to correct inaccurate data;[108] and a right to object to continued processing of data by government entities or private entities under the "legitimate interest" justification, though importantly, the right to object is *not* absolute, and can be outweighed by "compelling legitimate grounds for the processing."[109] Most famously, the GDPR also codifies the "right to be forgotten,"[110] which had been recognized earlier by the Court of Justice of the European Union.[111] This provision effectively permits data subjects to demand the erasure of data no longer needed for processing — though as with many GDPR "rights," this one is limited and can be overridden by, inter alia, "exercising the freedom of expression and information."[112]

There is little doubt that the GDPR, through these and many other provisions, establishes one of the most comprehensive data regulation regimes in the world. It also creates important personal rights that, if invoked, could return substantial power to individual data subjects. It should also be noted, however, that there is absolutely no evidence that the GDPR has substantially restricted or interfered with the data practices of the major tech companies since it became effective in May of 2018. As such, while the GDPR remains a work in progress, especially at the level of national implementation, its actual impact remains unclear (though the European Court of Justice's recent *Schrems II* decision has raised the possibility that the GDPR will significantly impact/disrupt current data practices[113]).

Perhaps the most significant restriction data privacy regulation other than the GDPR is the California Consumer Privacy Act ("CCPA"),

106 *Id.* art. 13, 14.

107 *Id.* art. 15.

108 *Id.* art. 16.

109 *Id.* art. 21.

110 *Id.* art. 17.

111 Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos, 2014 E.C.R. 317. For an excellent discussion of the Google Spain decision and its problems, see Robert C. Post, *Data Privacy and Dignitary Privacy:* Google Spain*, the Right to be Forgotten, and the Construction of the Public Sphere*, 67 DUKE L.J. 981 (2018).

112 GDPR, *supra* note 98, art. 17(3)(a).

113 *See* Joshua P. Meltzer, *Why Schrems II Requires US-EU Agreement on Surveillance and Privacy*, BROOKINGS (Dec. 8, 2020), https://www.brookings.edu/techstream/why-schrems-ii-requires-us-eu-agreement-on-surveillance-and-privacy/ [https://perma.cc/GDU8-NUWX].

which became effective on January 1, 2020.[114] The CCPA has obvious parallels to the GDPR, though it is clearly more limited in scope. In particular, the CCPA's primary provisions give California consumers the right to request information about a firm's data collection and retention practices, a qualified right to have personal data deleted, and a right to opt out of sale of personal information.[115] The CCPA does not, however, contain the sorts of broad restrictions on data processing, or scope of consumer rights, contained in the GDPR. Nonetheless, because of the sheer size of California, the fact that it is home to most of the major American tech giants, and because many tech firms seem inclined to apply the CCPA's rules to all U.S. consumers, not just Californians, the impact of the law is likely to be significant.

Finally, we come to Facebook co-founder Chris Hughes's argument, first enunciated in a New York Times opinion piece on May 9, 2019,[116] that the only solution to the data, privacy, and other problems at Facebook is to break the firm up, because at bottom the problem lies in the fact that Facebook — and Mark Zuckerberg in particular — simply have too much power. On its surface, this approach — what I call the "nuclear option" — looks groundbreaking. If indeed the government were to break up Facebook (as well as Google and Amazon, as Hughes hints) into much smaller firms, that would indeed change the face of the tech industry. In fact, however, there is much less to Hughes's suggestion than meets the eye. In practice, all he is advocating is that the Federal Trade Commission ("FTC") and Justice Department require Facebook to reverse, and spin off, its acquisitions of Instagram (in 2012) and WhatsApp (in 2014) — an invitation which the FTC accepted when it filed an antitrust suit against Facebook on December 9, 2020, seeking a judicial order require divestitures by Facebook of Instagram and WhatsApp.[117] Whatever the merits of this lawsuit (a topic I will return to in the next Part), it should be noted that it would have no impact on the Facebook *platform* itself, which as Hughes himself acknowledges is used by two-thirds of current social media users.

---

114 CAL. CIV. CODE § 1798.100 (2020).

115 State of California Department of Justice, *California Consumer Privacy Act (CCPA)*, https://oag.ca.gov/privacy/ccpa#sectione (last visited Jan. 23, 2021) [https://perma.cc/7U4R-UXGB].

116 Hughes, *supra* note 20.

117 *See FTC Sues Facebook for Illegal Monopolization*, FED. TRADE COMM'N (Dec. 9, 2020), https://www.ftc.gov/news-events/press-releases/2020/12/ftc-sues-facebook-illegal-monopolization [https://perma.cc/EQT6-YDUT].

B. Censorship and Section 230

There can be no doubt that the Internet, including especially social media, contains a lot of what Eric Goldman and Jess Miers call "awful content."[118] Depictions of violence, cyberstalking, trolling, hate speech, and outright threats are extraordinarily common — though to be fair, it is not clear that such speech has become *more* common over time, as many assume it has.[119] But such speech does exist, and causes substantial harm, both psychological and sometimes physical, which disproportionately targets vulnerable segments of society such as women, racial minorities, and sexual minorities. Furthermore, it is as clear as such things can be that the rapid dissemination of information that the Internet facilitates exacerbates those harms. It is also clear that the major Internet companies have not successfully blocked such content, despite the fact that their terms of service typically prohibit such speech. It is therefore perfectly understandable that many people believe that the law should respond. The question, of course, is how.

Though complaints and proposals are myriad, they fall into two primary groups. First are calls to directly require Internet companies to block specific content, on pain of punishment (typically fines). The second is to reduce or eliminate platforms' Section 230 immunity, so as to incentivize them to act. Both approaches have serious advocates, and so deserve some attention.

Consider first direct regulation. On August 5, 2020, a group of twenty state Attorneys General sent a joint letter to Facebook, calling on the firm to make greater efforts to block harmful content.[120] And although the letter identified a number of different kinds of harmful speech, including disinformation, cyberstalking, doxing (publishing private information), and swatting (filing false police reports), the primary focus of the letter was hate speech — which is to say, speech that vilifies specific minority groups.[121] And while the letter itself does not go beyond calling on Facebook to take voluntary action, in an interview with the New York Times, Attorney General Gurbir S. Grewal of New

---

[118] Eric Goldman & Jess Miers, *Why Can't Internet Companies Stop Awful Content?*, ARS TECHNICA (Nov. 27, 2019, 5:45 AM), https://arstechnica.com/tech-policy/2019/11/why-cant-internet-companies-stop-awful-content/ [https://perma.cc/EDC5-9EE8].

[119] *See id.*

[120] Alba, *supra* note 44.

[121] Letter from Karl A. Racine, Attorney General, District of Columbia, Kwame Raoul, Attorney General, State of Illinois, Gurbir S. Grewal, Attorney General, State of New Jersey, et. al., to Mark Zuckerberg, Chairman & Chief Executive Officer, Sheryl Sandberg, Chief Operating Officer (Aug. 5, 2020).

Jersey (one of the signatories), threatened that if Facebook did not act, "we always have a variety of legal tools at our disposal."[122] In other words, Grewal appeared to be suggesting that if Facebook failed to do a better job of blocking hate speech and other harmful content, state prosecutors would seek legal remedies against it, thus opening the door to direct legal regulation of Facebook's content moderation policies.

Aside from (admittedly vague) threats of direct regulation, the primary proposed remedy for "awful content" on the Internet, supported by politicians across the spectrum, is Section 230 reform. President Joe Biden, for example, said in an interview with the New York Times in December 2019 that "Section 230 should be revoked, immediately should be revoked."[123] The reason he gave was Facebook's failure to block harmful speech (with a particular focus on falsehoods), though House Speaker Nancy Pelosi cited harassment and abuse as the reason to eliminate immunity.[124] In May 2020 President Trump joined the bandwagon, tweeting "REVOKE 230!" in response to a dispute with Twitter,[125] a position he reiterated in December 2020 in the course of vetoing a major defense appropriation bill;[126] though admittedly, Trump's calls to revoke Section 230 were triggered not by awful content (of which he is a regular source), but rather by social media firms' alleged anti-conservative bias. Nevertheless, it is noteworthy that both major-party candidates in the 2020 presidential election shared this view, given how little else they agreed upon. Nor are Trump and Biden alone in calling for repeal of Section 230.[127]

Aside from calls to entirely eliminate the platform immunity that Section 230 provides, there have also been calls for more restrained revisions to that provision. For example, Republican Senator Josh Hawley, also responding to alleged political bias, has proposed

---

122 Alba, *supra* note 44.

123 Editorial Board, *Joe Biden*, N.Y. TIMES (Jan. 17, 2020), https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html [https://perma.cc/ZKF8-QHAD].

124 Bobby Allyn, *As Trump Targets Twitter's Legal Shield, Experts Have a Warning*, NPR (May 30, 2020, 11:36 AM), https://www.npr.org/2020/05/30/865813960/as-trump-targets-twitters-legal-shield-experts-have-a-warning [https://perma.cc/KF5Z-JRY6].

125 *Id.*

126 *Presidential Veto Message to the House of Representatives for H.R. 6395*, WHITE HOUSE (Dec. 23, 2020), https://www.whitehouse.gov/briefings-statements/presidential-veto-message-house-representatives-h-r-6395/ [https://perma.cc/3XP3-EC27].

127 *See, e.g.*, Steve Randy Waldman, *The 1996 Law That Ruined the Internet: Why I Changed My Mind About Section 230*, ATLANTIC (Jan. 3, 2021), https://www.theatlantic.com/ideas/archive/2021/01/trump-fighting-section-230-wrong-reason/617497/ [https://perma.cc/GKF7-V668].

legislation that would condition immunity on tech platforms passing an independent audit that confirmed the platforms were not politically biased.[128] Law professors Danielle Keats Citron and Mary Anne Franks have proposed a set of more thoughtful, and more limited, reforms of Section 230. One would limit Section 230 immunity to *speech*, thereby clarifying that online commercial transactions and the like would not fall within the provision.[129] They would also deny immunity to truly bad actors, meaning websites that knowingly keep up illegal content, encourage illegality, principally host illegal content, or solicit illegal content.[130] Such a revision would presumably have no impact on the major platforms such as Facebook and Twitter, but would permit action against the seediest parts of the Internet. Finally, they propose language that would condition Section 230 immunity on platforms taking "reasonable steps to address unlawful uses of its service that create serious harm to others."[131] Such a provision would, of course, require courts to determine what constitutes "reasonable steps" in a world in which, all acknowledge, content moderation will necessarily be imperfect. Citron and Franks argue, however, that courts have proven capable of making such judgments in the past, and that over time best practices will emerge.[132]

These various proposals culminated, in September 2020, with the Trump Administration Justice Department sending legislation to Congress proposing substantial revisions to Section 230.[133] Part of the proposal, evidently drawing on the Citron/Franks reforms discussed above, would deny immunity to "Bad Samaritans," meaning platforms that knowingly facilitate criminal behavior, or knowingly failed to remove material that violated criminal law. Other parts, however, go much farther. Notably, the Department of Justice ("DOJ") proposal would narrow immunity under Section 230(2) to decisions to remove material that are based on "an objectively reasonable belief" that the material is harmful, and it would eliminate platforms' ability to remove content that it believes to be "otherwise objectionable," thereby sharply narrowing platform owners' discretion regarding what sorts of material is harmful (largely limiting it to sexual, violent, or unlawful materials).

---

128 Allyn, *supra* note 124.

129 Citron & Franks, *supra* note 20.

130 *Id.* at 70-71.

131 *Id.* at 71.

132 *Id.* at 71-73.

133 *The Justice Department Unveils Proposed Section 230 Legislation*, U.S. DEP'T OF JUST. (Sept. 23, 2020), https://www.justice.gov/opa/pr/justice-department-unveils-proposed-section-230-legislation [https://perma.cc/VY8H-M3UG].

Finally, the proposal would eliminate Section 230 liability for actions brought under a wide swath of laws, including laws regulating terrorism, child sex abuse, cyber-stalking, and antitrust. Some of these proposals, notably the Bad Samaritan exception, respond to widely shared concerns. However, it seems fairly clear, given President Trump's disputes with social media firms, that a major objective of some of these proposals, especially the removal of the "otherwise objectionable" language, was to restrict social media platforms' ability to block politically charged posts by conservative politicians.

C. Fact-checking and Limits on Microtargeting

In the battle over false political speech, social media and other Internet firms disagree sharply among themselves regarding appropriate responses. Twitter has simply banned political advertising on its platform, unlike Google and Facebook.[134] But the major disagreement has been over fact-checking of political ads *and* posts by politicians. Google does fact-check political ads, and as of April 2020 announced that it would add third-party generated information panels to some searches on YouTube (which Google owns).[135] Furthermore, Twitter, as noted earlier, has begun to fact-check and label political posts (since Twitter does not accept political ads, fact-checking them is moot).[136] Facebook, however, continues to hold to its policy of fact-checking neither.[137] Public pressure continues to be applied against Facebook from a number of sources, including commentators, politicians (including Joe Biden, as noted earlier), and its own employees to reconsider this policy; but to date Mark Zuckerberg has not budged.

Assuming Facebook retains its current stance, it is possible that proposals will emerge to either require Facebook to block false statements, or to impose liability on Facebook for such statements by amending Section 230. For reasons discussed in the next Part, however, neither option is likely to work, and in any event both are almost certainly unconstitutional. More realistic are proposals to restrict political ads microtargeting voters based on specific criteria, on the theory that such conduct poses a particularly dangerous threat to democracy by suppressing specific communities' votes, and by

---

134 *See* Romm, Stanley-Becker & Timberg, supra note 56.

135 *See Fact Checks in YouTube Search Results*, YOUTUBE HELP, https://support.google.com/youtube/answer/9229632?hl=en (last visited Jan. 23, 2021) [https://perma.cc/FT83-E4T4].

136 *See* Halon, supra note 57.

137 *See supra* notes 46–48 and accompanying text.

undermining public discourse because secret microtargeting makes it difficult to engage in "counter speech." Spencer Overton in particular urges states to consider legislation imposing liability on social media firms that permit microtargeting audiences of color.[138] Google, as noted earlier, has imposed substantial restrictions on microtargeted political ads.[139] Facebook, however, has limited itself to transparency, giving the public access to a database of political ads and their targeted audiences in order to enable counter speech, without restricting speech directly. The controversy, therefore, is unlikely to go away.

D. Regulating Incitement, and Reconsidering Brandenburg

In response to the problem of violent incitement, many commentators propose that Facebook and other social media firms be required to speedily take down posts that, directly or indirectly, incite violence, on pain of substantial penalties. And indeed Germany and Australia, among other countries, have both enacted precisely such legislation. Australia, for example, enacted legislation in April 2019 that threatens jail time for employees of social media companies who fail to remove "abhorrent violent material" — defined as audio or video depicting extreme violence that is posted by someone involved in the violence — "expeditiously."[140] The legislation was enacted in response to the terrorist attack the previous month on two mosques in Christ Church, New Zealand by a white supremacist, who livestreamed the attacks on Facebook. Given the narrowness of the law and the draconian punishments it envisions, however, it is unclear what impact it will have.

It is Germany, rather, that in practice has been the world leader in this area. Under Germany's NetzDG law, effective January 1, 2018, websites that do not, within twenty-four hours, remove hate speech that is "obviously illegal" under German law are subject to fines of up to fifty million euros.[141] In an attempt to comply with this law (and enforce its

---

[138] Spencer Overton, *State Power to Regulate Social Media Companies to Prevent Voter Suppression*, 53 UC DAVIS L. REV. 1793, 1799-1801 (2020). Overton's primary argument is that such state provisions would not conflict with section 230, a question beyond the scope of this Essay.

[139] *See supra* note 54.

[140] *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) (Austl.); Damien Cave, *Australia Passes Law to Punish Social Media Companies for Violent Posts*, N.Y. TIMES (Apr. 3, 2019, 10:40 AM), https://www.nytimes.com/2019/04/03/world/australia/social-media-law.html [https://perma.cc/57K4-QMDE].

[141] *Germany Starts Enforcing Hate Speech Law*, BBC (Jan. 1, 2018), https://www.bbc.com/news/technology-42510868 [https://perma.cc/EF8A-DY56].

own Terms of Service), Facebook established a deletion center outside of Berlin staffed by over 1200 content moderators.[142] And there are at least some indications that Facebook has increased the amount and speed of deletions in response to NetzDG.[143] On the other hand, there have also been complaints that Facebook is, out of caution triggered by the risk of large fines, deleting legitimate posts, and that the law is chilling political speech.[144] And worse, nations with less liberal agendas than Germany's have adopted copycat laws with the predictable result of significantly chilling or silencing legitimate speech the government disapproves of.[145] The merits of Germany's approach to hate speech and incitement in the NetzDG, therefore, remains highly disputed.

Not coincidentally, the examples of official efforts to suppress hate speech and incitement all come from abroad. Within the United States, the First Amendment as currently interpreted by the Supreme Court poses an essentially insurmountable barrier to similar efforts. Hate speech is fully protected by the First Amendment under almost all circumstances, and therefore cannot be banned by the government.[146] Even direct incitement of violence is protected by the First Amendment unless it can meet the extremely strict standard, established in the *Brandenburg* case in 1969, that the speech advocating violence "is directed to inciting or producing imminent lawless action and is likely to incite or produce such action."[147] It is possible that some of the speech discussed above,[148] inciting violence in Myanmar and elsewhere, might satisfy the *Brandenburg* test; but the vast majority of speech on social media attacking others and calling for violence is far too abstract and indefinite to qualify as unprotected incitement under U.S. law. Finally, it should be obvious that just as the government cannot directly

142 Katrin Bennhold, *Germany Acts to Tame Facebook, Learning from Its Own History of Hate*, N.Y. TIMES (May 19, 2018, 10:45 AM), https://www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html [https://perma.cc/D6BV-69R8].

143 *See* Rebecca Zipursky, Note, *Nuts About NETZ: The Network Enforcement Act and Freedom of Expression*, 42 FORDHAM INT'L L.J. 1325, 1353 (2019).

144 *Id.* at 1359-60; *see* Linda Kinstler, *Germany's Attempt to Fix Facebook Is Backfiring*, ATLANTIC (May 18, 2018), https://www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/ [https://perma.cc/AS5Z-GGGY].

145 *See* Zipursky, *supra* note 143, at 1360-62.

146 *See* Matal v. Tam, 137 S. Ct. 1744, 1764 (2017); R.A.V. v. City of St. Paul, 505 U.S. 377, 395-96 (1992). The narrow circumstances in which hate speech can be banned is when it constitutes a "true threat," Virginia v. Black, 538 U.S. 343, 359 (2003), or when it is directed at an individual, in-person, in a way that makes the speech "fighting words." Chaplinsky v. New Hampshire, 315 U.S. 568, 571-72 (1942). Since the fighting words doctrine is limited to in-person speech, it is of course irrelevant on the Internet.

147 Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).

148 *See* discussion *supra* Part I.D.

punish such speech or speakers, it also cannot punish Facebook for circulating it. That is the enduring lesson of *New York Times v. Sullivan*.[149]

It must be noted, however, that most countries are far less protective of speech inciting violence than the United States, and that in any event even if incitement is constitutionally protected, it violates Facebook's Terms of Service (which is of course permissible because Facebook is not a state actor subject to the strictures of the First Amendment). The real problem, as noted earlier, is that with respect to its Facebook platform, lack of translators makes tracking incitement difficult for Facebook, and with respect to its WhatsApp platform encryption makes it impossible. There is no solution to the latter barrier except for the creation of backdoors in WhatsApp's encryption scheme (on which more in the next Part). The translation problem, however, does in theory have a solution, which is to throw resources at it by hiring more native language speakers as content moderators for every country in which Facebook operates. In response to specific incidents, Facebook unsurprisingly often commits to hiring more local-language moderators, as in Myanmar[150] and Sri Lanka.[151] The problem, of course, is that to a substantial extent this is an example of closing the barn door after the horse has bolted. To be truly effective, Facebook would have to hire large numbers of local-language translators in every single country in which Facebook is available, and in which there is a risk of violent incitement (which is to say, every country).

E. State Actors, Utilities, and Due Process of Censorship

In proposing solutions to the problems of excessive, and allegedly discriminatory content moderation (i.e., censorship) by social media firms such as Facebook, unsurprisingly commentators have moved in very different directions from those addressing insufficient censorship. Perhaps the most radical proposals have argued that social media firms should be treated as state actors, so that their content moderation practices would be subject to First Amendment scrutiny.[152] These

149 376 U.S. 254 (1964).

150 *See* Stevenson, *supra* note 11.

151 *See* Taub & Fisher, *supra* note 64.

152 *See, e.g.*, Benjamin F. Jackson, *Censorship and Freedom of Expression in the Age of Facebook*, 44 N.M. L. REV. 121, 142 (2014); Colby M. Everett, Note, *Free Speech on Privately-Owned Fora: A Discussion on Speech Freedoms and Policy for Social Media*, 28 KAN. J.L. & PUB. POL'Y 113, 125-26 (2018); Daniel Rudofsky, Note, *Modern State Action Doctrine in the Age of Big Data*, 71 N.Y.U. ANN. SURV. AM. L. 741, 774-77 (2017). *But see* Matthew P. Hooker, *Censorship, Free Speech & Facebook: Applying the First Amendment*

commentators typically rely on the Supreme Court's decision in *Marsh v. Alabama*,[153] in which the Court held that the owner of a company town (a private corporation) was a state actor, subject to the First Amendment, because it performed a "public function." The argument is that by operating what is in effect (and what the Supreme Court has called) a "public forum,"[154] social media firms are similarly performing a public function and so should be considered state actors. (One commentator has, for similar reasons, argued that Internet Service Providers should also be considered state actors.)[155]

A seemingly more modest proposal regarding social media is to treat social media firms as "utilities," and subject them to the kind of regulation we apply to traditional utilities such as electric and phone companies. In particular Dipayan Ghosh, a former official in the Obama White House (and a former advisor to Facebook), argued in the Harvard Business Review that Facebook's social media functions constitute a natural monopoly, and so should be regulated as a utility, just as we do other natural monopolies.[156] Ghosh is quite vague about what utility treatment of Facebook would entail, except greater regulation (and he entirely ignores the First Amendment), but his proposal has received some attention. Earlier, Sabeel Rahman had advanced a similar argument that because large Internet firms such as Facebook and Google (and Amazon) control critical infrastructure, they should be treated as public utilities.[157] Like Ghosh, Rahman is somewhat vague about what utility regulation entails, but again emphasizes tighter scrutiny of tech mergers, and preventing tech firms from expanding into adjacent markets.

The final set of proposals to address over- or biased censorship revolve not around substance but rather process. In effect, they would require social media firms to increase transparency, and to create structural mechanisms that govern their content moderation practices in which users who object to particular decisions would have participation and appeals rights. Thus Jack Balkin, explicitly drawing

to Social Media Platforms via the Public Function Exception, 15 WASH. J.L. TECH. & ARTS 36, 62-67 (2019).

153 326 U.S. 501 (1946).

154 *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017).

155 *See* Eric Sirota, *Can the First Amendment Save Net Neutrality?*, 70 BAYLOR L. REV. 781, 784-87 (2018).

156 Dipayan Ghosh, *Don't Break Up Facebook – Treat It Like a Utility*, HARV. BUS. REV. (May 30, 2019), https://hbr.org/2019/05/dont-break-up-facebook-treat-it-like-a-utility [https://perma.cc/AQ6Q-43UU].

157 K. Sabeel Rahman, *Regulating Informational Infrastructure: Internet Platforms as the New Public Utilities*, 2 GEO. L. TECH. REV. 234, 235-36 (2018).

on the analogy to due process,[158] suggests that social media firms should abide by "(1) obligations of transparency, notice, and fair procedures; (2) the offer of reasoned explanations for decisions or changes of policy; (3) the ability of end-users to complain about the conduct of the institution and demand reforms; and (4) the ability of end-users to participate, even in the most limited ways, in the governance of the institution."[159] Similarly Kate Klonick, in her pathbreaking study of content moderation at Facebook, identified concerns with Facebook's governance practices rooted in the lack of access and participation rights of users, and the lack of accountability on the part of Facebook.[160] One proposal Klonick (drawing on earlier, pathbreaking work by Danielle Citron) advances is that Facebook voluntarily commit to "technological due process," including some limits on automated content moderation.[161] More recently, Klonick has expressed (admittedly cautious) hope that Facebook's new "Oversight Board" might alleviate some of these concerns, by increasing transparency and user participation rights.[162] What these proposals have in common is that they accept the reality that content moderation by social media firms is here to stay, and so try to tame it. The ultimate goals are to reduce errors, to protect the real — if probably not legally enforceable — rights of users, and ultimately to mimic some version of democratic accountability, even if not through direct popular control.

III. OBJECTIONS AND CONTRADICTIONS

This Part sets out some of the difficulties, both practical and constitutional, raised by the reform proposals discussed in the previous Part. At the outset, I will emphasize that it is not my position that no legal reforms are possible. To the contrary, some of the more modest steps, such as the FTC lawsuit seeking divestiture of Instagram and the Citron/Franks proposal to amend Section 230, show great promise. The reality is, though, that many of the more far-reaching regulatory initiatives set forth above are probably unconstitutional, in any event they are deeply inconsistent with one another.

---

158 As with the First Amendment, social media firms are not subject to constitutional Due Process provisions because they are not state actors.

159 Balkin, *Free Speech in the Algorithmic Society*, *supra* note 91, at 1198.

160 Klonick, *supra* note 41, at 1662-69.

161 *Id.* at 1668-69 & nn.483–84 (citing Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1301-13 (2008)).

162 *See* Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2491-92 (2020).

A. Information Fiduciaries and Data Privacy

Let us begin with Jack Balkin's proposal to treat tech giants as "Information Fiduciaries." As a starter, it should be noted that Lina Khan and David Pozen have published a long, thoughtful critique of the "Information Fiduciary" proposal,[163] which some of my objections parallel. In this short space, however, the most I hope to accomplish is to demonstrate that Balkin's proposal is subject to substantial practical and constitutional objections. To begin with, the entire notion of a "fiduciary" seems a poor fit to describe the relationship between social media firms and their users. Merriam-Webster defines fiduciary as "of, relating to, or involving a confidence or trust."[164] Yet the very idea that Facebook users "trust" either Facebook, the firm, or CEO Mark Zuckerberg, its public face, seems quite extraordinary. Thus to convert tech giants into legal fiduciaries is to impose, by law, a relationship of trust where none existed before, a piece of social engineering that seems unlikely to succeed.

Furthermore, as Khan and Pozen point out, the fiduciary concept is inconsistent with the underlying business model of social media and tech firms, which relies on selling targeted advertising based on user data; data which the firms obtain in exchange for providing free services such as social media accounts and searches.[165] The idea that firms should be treating the very commodity that they sell — user data — as something to be used for the benefit of users seems bizarre. It would be like telling GM that it should treat cars as existing for the benefit of steel makers. The objection here is not that regulation would undermine Facebook's business model (thought that objection might also be raised), but rather that the concept of a fiduciary is inconsistent with that model.

Finally — and inevitably — there is a serious constitutional problem with the Balkin proposal, rooted in the First Amendment. The problem, in short, is that the current Supreme Court has strongly suggested that it considers the transfer and sale of data to be speech.[166] As such,

---

[163] *See* Khan & Pozen, *supra* note 25, at 501-02; *see also* Jane R. Bambauer, *The Relationships Between Speech and Conduct*, 49 UC Davis L. Rev. 1941, 1944 (2016) (arguing that Balkin's theory on its own terms, does not reach many tech giants).

[164] *Fiduciary*, MERRIAM WEBSTER, https://www.merriam-webster.com/dictionary/fiduciary (last visited Jan. 24, 2021) [https://perma.cc/G3RN-TU8Q].

[165] *See* Khan & Pozen, *supra* note 25, at 510-16.

[166] *See* Sorrell v. IMS Health Inc., 564 U.S. 552, 570 (2011); *see also* Jane Bambauer, *Is Data Speech?*, 66 STAN. L. REV. 57, 60-61 (2014); Ashutosh Bhagwat, Sorrell v. IMS Health*: Details, Detailing, and the Death of Privacy*, 36 VT. L. REV. 855, 855-56 (2012); *cf.* Neil M. Richards, *Reconciling Data Privacy and the First Amendment*, 52 UCLA L. REV.

restrictions on the sale of data (especially if anonymized, as in effect it is when social media firms sell advertising) poses serious First Amendment challenges. Balkin's answer to this is that "[t]he First Amendment does not prevent the state from regulating how professionals interact with their clients and how they use their clients' information . . . because professionals have a fiduciary relationship with their clients."[167] The difficulties with this argument are two-fold. First, there seems an inherent circularity in defending the designation of an entity as a fiduciary from constitutional attack based on the idea that fiduciaries lack First Amendment rights. Second, and more basically, in 2018, in a case in which the Supreme Court struck down California's attempts to regulate so-called "Crisis Pregnancy Centers," a majority of the Court quite firmly rejected the argument that "professional speech" between professionals and patients/clients fell outside the First Amendment, or was subject to lower First Amendment standards.[168] The implication of all of this is that any attempt to impose fiduciary obligations on social media firms would have to survive stringent First Amendment scrutiny, a most unlikely outcome.

Many of the objections to the "information fiduciary" approach also apply to the GDPR and similar data protection regimes. Most fundamentally, if such regimes were applied so strictly that they prevented tech firms from monetizing data via the sale of targeted advertising, they would completely undermine the business model of those firms. The result would then be either that the services these firms offer would disappear — an outcome it is hard to believe anyone seriously supports — or would require firms to charge users for services such as search and social media. It seems questionable, however, whether being forced to pay for services previously provided for free is something most users would desire, surely a relevant consideration given that it is those users that data protection laws purport to protect. Furthermore, there can be little doubt that moving to a business model based on charging users will substantially exacerbate the existing digital divide, especially between users in the global North and South. It should be noted in this regard that Facebook and WhatsApp provide essential services in many poor and authoritarian countries, including access to uncensored news and uncensored forms of electronic communication, that would not otherwise be available. Supporters of legal proposals that

1149, 1151 (2005) (disagreeing with the view that the First Amendment is a bar to data privacy laws).

167 Balkin, *Free Speech in the Algorithmic Society*, *supra* note 91, at 1161.

168 Nat'l Inst. for Family and Life Advocates v. Becerra, 138 S. Ct. 2361, 2374-75 (2018).

blithely ignores these important benefits of social media should at least be forced to explain why the benefits they provide outweigh these substantial harms.

Given the unattractiveness of these outcomes, it seems likely that data protection regimes such as the GDPR will not be enforced in ways that fundamentally interfere with current tech business models. Instead, as appears to be the case with the GDPR, tech firms will likely be permitted to continue current practices under exceptions such as the consent or the "legitimate interest" GDPR exceptions, so long as they take additional steps to ensure transparency and data integrity. This seems like a positive outcome, but it too has a significant downside: such data protection steps are expensive. Obviously, few would shed tears over behemoths such as Facebook and Google having to expend some of their seemingly limitless funds on data protection; but it must be recognized that expensive regulatory obligations inevitably act as barriers to entry, preventing startups and small firms entering into these markets because they cannot afford the same levels of protections.[169] And yet, aside from privacy, one of the prime complaints about the tech giants is their market power. Adopting regulatory regimes that accentuate that market power seems questionable policy.

One short word, also, on the "right to be forgotten." In 2014 the European Court of Justice held, in *Google Spain SL v. Agencia Española de Protección de Datos*,[170] that the then-existing European Union data Directive (which dated from 1995) entitled individuals to require search engines such as Google to remove from search results links to pages containing true and embarrassing, but no longer relevant, information about the individual. As Robert Post has brilliantly demonstrated, *Google Spain* and the GDPR, which later codified the right to be forgotten, are deeply problematic because they confuse an instrumental right to data protection, which is the focus of the GDPR, with a dignitary right to move on from the past, which has to underlie any plausible claim for a right to be forgotten.[171] As a result, Post points out, the *Google Spain* decision oddly restricts Google's ability to link to a newspaper website containing the offending information, but does *not* restrict the newspaper website itself despite the fact that *it* is the source

[169] One option might be to exempt small firms from data protection rules; but that would have the perverse effect of incentivizing consumers to migrate to platforms that do not protect their data.

[170] Case C-131/12, 2014 E.C.R. 317.

[171] Post, *supra* note 111.

of the information, and the actual communicative actor.[172] In addition, within the American context there are deep, unresolved questions about whether a "right to be forgotten" could be reconciled with the First Amendment. At a minimum, it seems clear that the right could not be applied to information "on matters of public concern," defined extremely broadly to include even highly offensive content.[173] And given the Supreme Court's increasingly expansive approach to speech protections in recent years, even when the speech is not even arguably on matters of public concern,[174] there are reasons to doubt whether the Court would sustain even a narrow right to be forgotten.

Finally, we come to Chris Hughes's "nuclear option," to break up Facebook (and other tech giants, perhaps). For starters, it should be remembered that there is less to Hughes's proposal than meets the eye. Hughes does not propose dividing up the Facebook platform itself, but only requiring Facebook to reverse its acquisitions of Instagram and WhatsApp.[175] As noted earlier, in December 2020 the FTC filed an antitrust action seeking precisely this result.[176] If extended beyond Facebook, presumably that could result in legal action in the future requiring Google to divest itself of YouTube. First of all, while such divestitures would reduce Facebook and Google's market power vis-à-vis advertisers, even if the FTC is successful the impact on Facebook would be fairly minimal since Facebook's Facebook platform has far more users than Instagram or WhatsApp (and WhatsApp, being encrypted, doesn't even sell targeted advertising).[177] Second, it is truly hard to see what benefit users would derive from such a breakup. Facebook would remain Facebook, Instagram would remain Instagram, and YouTube would remain YouTube, with all of the same control over user data that they enjoy today. And if the next question is why not just

[172] *See id.* at 1062-67. One is left to speculate whether it is a coincidence that the import of the European Court's decision was to impose burdens on an American search engine, while shielding European websites.

[173] *See* Snyder v. Phelps, 562 U.S. 443, 452-55 (2011).

[174] *See, e.g.*, United States v. Alvarez, 567 U.S. 709, 729 (2012) (extending full First Amendment protection to knowing falsehoods); Brown v. Entm't Merchs. Ass'n, 564 U.S. 786, 799 (2011) (extending full First Amendment protection to violent video games sold to minors); United States v. Stevens, 559 U.S. 460, 472 (2010) (extending full First Amendment protection to "depictions of animal cruelty").

[175] *See* Hughes, *supra* note 20 and accompanying text.

[176] *See FTC Sues Facebook for Illegal Monopolization*, *supra* note 117 and accompanying text.

[177] Given Google's dominance in areas such as search and email, which surely attract far more users than YouTube, one suspects that spinning off YouTube similarly will have a minor effect on Google's market power over online advertising.

break up those platforms, the short answer is that it would almost certainly not work. Communications platforms such as Facebook famously are characterized by network effects, where the value of the platform grows directly with the number of users it connects. In such as world, one platform will inevitably, over time, come to dominate a specific social media niche. Of course, as Twitter, YouTube, and, yes, Instagram demonstrate, there is space in the world for multiple social media niches. But, as Facebook's sheer size and reach demonstrate, a single niche is likely to be dominant for precisely the same network-effect reasons. So, while the FTC's efforts to reverse Facebook's acquisition of Instagram as a forbidden horizontal merger probably do make sense (assuming such a divestiture could actually be accomplished), it is doubtful if this step will accomplish much if, as seems likely, the Facebook platform retains its dominance. As for WhatsApp, given that it is a messaging platform rather than true social media, and that Facebook has not monetized it, it remains far from clear what benefit anyone would derive from separating it from Facebook (though it might end WhatsApp as a free service).

B. Awful Content and Section 230

Moving on now from privacy to the problem of "awful content," the barriers facing proposed reforms are once again both constitutional and practical — though with narrow reforms not insurmountable. With respect to the worst sorts of speech on the Internet, such as personal threats or stalking, there are no constitutional barriers to regulation because such speech is unprotected under the First Amendment[178] (though the practical barriers to regulation discussed earlier apply here as well given the sheer scale of speech on social media). The primary focus of criticisms of social media, however, have been on hate speech, meaning speech that demeans groups on the basis of characteristics such as race, national origin, sex, religion, or sexual orientation. The difficulty is that such speech, while undoubtedly vile and harmful, is fully protected by the First Amendment.[179] Indeed, because such speech is considered political speech on matters of public concern, it receives the very highest level of First Amendment protection.[180] And to cap things off, the Court has consistently in recent years treated efforts to

[178] *See* Virginia v. Black, 538 U.S. 343, 359-60 (2003) (holding that "true threats" constitute unprotected speech).

[179] *See* Matal v. Tam, 137 S. Ct. 1744, 1764 (2017) (plurality opinion); *id.* at 1766-67 (Kennedy, J., concurring in part and concurring in the judgment).

[180] *See* Snyder v. Phelps, 562 U.S. 443, 458 (2011).

suppress hate speech as almost *per se* unconstitutional viewpoint-based regulations.[181] The implication of all of this is that New Jersey Attorney General Grewal's threat to deploy "legal tools" against Facebook if it does not do a better job of blocking hate speech[182] is so much hot air, and that in the United States at least, any efforts to punish Facebook for failing to suppress hate speech would clearly be unconstitutional. The same is not true in other countries, where hate speech typically receives far less or no constitutional protections, admittedly, though there too practical barriers are likely to prove significant.

Turning to those practical barriers, the most obvious one is definitional. Facebook operates in a lot of different countries, and many users (this author included) have Facebook friends spread out through different countries. Yet it is clear that there is nothing close to an international consensus regarding what constitutes "awful content" or "hate speech." Perhaps most people agree that outright racist speech is awful (though recent political rhetoric in the United States demonstrates how underdeveloped even that consensus is). But there is clearly no such consensus on misogynist speech, much less homophobic speech. Facebook's Community Standards do define hate speech, and do so broadly (generally following American and European understandings);[183] but it is a fair question to ask if it is appropriate to extend these cultural norms to other cultures. That Facebook has done so is to its credit in my view, but to legally require such imposition of cultural norms strikes me as highly problematic.

The problem of definition, indeed, may well be unsolvable. Even such a seemingly unproblematic rule such as a ban on terrorist propaganda (short of incitement of violence, which is discussed below) can cause problems. Surely most people (other than terrorists) would agree that Facebook's ban on postings by or on behalf of groups such as ISIS makes sense.[184] But given the propensity of many regimes around the world (including a certain recent American President) to label political enemies and activists as "terrorists," even this rule faces major pitfalls. And while Facebook has voluntarily taken on itself the task of

---

[181] *See, e.g.*, *Matal*, 137 S. Ct. at 1766-67 (Kennedy, J., concurring in part and concurring in the judgment); R.A.V. v. City of St. Paul, 505 U.S. 377, 391-92 (1992).

[182] *Supra* note 122 and accompanying text.

[183] *See Hate Speech*, *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/hate_speech (last visited Jan. 23, 2021) [https://perma.cc/C8Y5-V4GN].

[184] *See Dangerous Individuals and Organizations, Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/dangerous_individuals_organizations (last visited Jan. 23, 2021) [https://perma.cc/NK6D-CYNG].

navigating those pitfalls, a legal regime that seeks to do so again seems like a very steep climb.

On the flip side of attempts to make Facebook censor more harmful content, consider the fact that Facebook follows local legal standards and so *does* suppress in other countries speech, such as blasphemy or criticisms of authoritarian governments, which in western democracies would be considered highly protected speech. Unsurprisingly, Facebook and other tech firms are targets of domestic criticism for such actions; but oddly, few seem to question Facebook's compliance with arguably problematic American laws such as intellectual property regimes that many consider excessively restrictive. The truth is that to expect private firms to do more than comply with local law seems quixotic, unless that local law is so horrific (think Nazi Germany) that there is a moral imperative to withdraw entirely from that jurisdiction.

Finally, a major problem with legal restrictions on bad content is that they can easily result in what Danielle Citron has called "mission creep,"[185] in which definitions of disfavored speech are extended from "core" to more contested examples, often with a particular political valence. Thus, it would not be hard to imagine an authoritarian regime seeking to label video of unruly protests as depictions of violence. And even in democratic regimes there have been instances in which expression of religious views has been punished as hate speech.[186] The thought of such an approach being extended to online media through legal restrictions on both platforms and users raises obvious concerns about both censorship and chilling effects. Indeed, even a well-intentioned legal regime, such as Germany's NetzDG law,[187] if it combines large fines with short quick deletion requirements (as NetzDG does), will inevitably result in excessive self-censorship by social media platforms, in order to minimize the risk of liability.[188]

Section 230 reform proposals offer more hope than blunt efforts to directly regulate speech on social media platforms; but even there, the more radical proposals are entirely impractical, and in many instances unconstitutional. First of all, to simply eliminate Section 230 immunity for social media, as Joe Biden and Donald Trump have both proposed,[189]

---

[185] Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1050-51 (2018).

[186] *See, e.g.*, James Weinstein, *Hate Speech Bans, Democracy, and Political Legitimacy*, 32 CONST. COMMENT. 527, 555-61 (2017).

[187] *See Germany Starts Enforcing Hate Speech Law*, *supra* note 141 and accompanying text.

[188] Citron, *supra* note 185, at 1055.

[189] *See supra* notes 123–26 and accompanying text.

is a nonstarter. Section 230 immunity is essential for social media platforms to operate, because if the alternative is publisher liability on the model of traditional media, social media platforms would face essentially unlimited liability. The reality is that because social media platforms do not primarily generate their own content but rather post third-party content, mainly created by their users (in the case of Facebook, numbering in the billions), platforms simply cannot police the truth, falsity, or otherwise harmfulness of every post. Algorithms are of course of some use here, but they are necessarily imperfect, especially so in rooting out falsehoods about non-public figures. Thus, without Section 230, defamation liability alone would shut down social media as we know it. Indeed, it was the realization that even traditional media could not perfectly police falsehoods about even public figures that lead the Supreme Court, in the *New York Times v. Sullivan* case and its progeny,[190] to interpret the First Amendment to limit defamation and other liability of the media. With social media, that problem is hugely magnified, arguing for a constitutional command of even greater protection from liability. Leaving aside questions of constitutionality, proposing action that would destroy a multi-billion-dollar industry that provides services that billions of people around the world evidently value seems the height of Luddism. Furthermore, even if a behemoth such as Facebook could afford to undertake the enormous amounts of content moderation that repeal of Section 230 would require, several of its smaller competitors have already expressed concerns about their ability to do so.[191] Finally, some commentators have pointed out that eliminating or severely restricting Section 230 immunity will inevitably lead social media firms to over-filter borderline speech on topics such as sexuality, which could work to the detriment of marginal groups such as LGBTQ youth.[192]

Some of the more "narrow" reform proposals, such as Senator Hawley's idea to condition Section 230 immunity on political neutrality on the part of platforms, and the elements of the Trump DOJ proposal

[190] New York Times v. Sullivan, 376 U.S. 254, 283 (1964) (holding that public officials may not sue for defamation without demonstrating "actual malice"); *see also* Hustler Magazine v. Falwell, 485 U.S. 46, 56 (1988) (extending *Sullivan* holding to claims for intentional infliction of emotional distress); Curtis Publ'g Co. v. Butts, 388 U.S. 130, 155 (1967) (extending *Sullivan* holding to public figures).

[191] Todd Shields & Ben Brody, *Facebook Worries Smaller Rivals with Openness on Liability*, YAHOO! FIN. (Dec. 23, 2020), https://finance.yahoo.com/news/facebook-support-liability-reform-little-070000635.html [https://perma.cc/NSS3-EZUE].

[192] *See* Bill Easley, *Revising the Law That Lets Platforms Moderate Content Will Silence Marginalized Voices*, SLATE (Oct. 29, 2020, 5:43 PM), https://slate.com/technology/2020/10/section-230-marignalized-groups-speech.html [https://perma.cc/D9JR-C5QZ].

seeking to limit platform authority to block content they object to, run directly into the First Amendment. The basic problem is that the very idea that social media platforms have some sort of an obligation of political neutrality is simply wrong, and if enforced by law, unconstitutional. To the contrary, as quasi-media entities, social media platform owners almost certainly have a First Amendment right to exercise editorial control over content on their platforms. The Supreme Court has recognized such an editorial right on the part of cable television operators' choice of channels to carry, even though (like social media) cable television firms carry third-party content.[193] Indeed, Supreme Court Justice Brett Kavanaugh, when on the D.C. Circuit, took the position that even Internet Service Providers — far more passive conduits of content than social media firms — enjoy First Amendment editorial rights.[194] In this legal landscape, it is very difficult to imagine an argument that social media firms do not enjoy some editorial rights. And once such a right is recognized, firms *must* possess the right to control the political "slant" of content it hosts, no less than Fox News. Any other approach raises core First Amendment concerns about political bias in regulation. Finally, the fact that most major social firms, for business reasons, claim (convincingly, in my view) not to exercise such politically driven editorial control, is beside the point. Choosing not to exercise a First Amendment right is not the same as waiving it. The only strong contrary argument would appear to be that the Hawley/Trump proposals do not directly regulate platforms' editorial control, they merely condition Section 230 immunity on giving up such control. But given the enormous importance of Section 230 liability, and the strong First Amendment policy against governmental interference with editorial control,[195] it seems extremely unlikely that the courts would permit the government to leverage Section 230 for political ends in this way.

On the other hand, not all Section 230 reform proposals are unreasonable. The Citron/Franks argument that Section 230 immunity should not apply to truly bad actors, or those who knowingly ignore illegal activity, and the aspect of the Trump DOJ proposal that builds on these ideas, seem entirely reasonable reforms that should be adopted as soon as possible. It is very hard to imagine how such narrow, targeted exceptions to Section 230 could possibly affect or harm most social

193 *See* Turner Broad. Sys. v. FCC, 512 U.S. 622, 667-68 (1994).

194 U.S. Telecom Ass'n v. FCC, 855 F.3d 381, 418 (D.C. Cir. 2017) (Kavanaugh, J., dissenting), *denying reh'g en banc*.

195 *See, e.g.*, Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 258 (1974) (striking down a "right of reply" requirement imposed on a newspaper).

media platforms; and the good that they would accomplish is well worth any minor risks. This does NOT, however, mean that broader exceptions, such as the Trump DOJ's proposal to carve out a huge number of substantive laws from Section 230, are necessarily a good idea. If the implications of these carve-outs is to impose strict or negligence-based liability on platforms used to carry out violations of the listed laws, that might well incentivize platforms to over-censor speech at the margins of these laws. On the other hand, if platform liability is limited to knowing violations of these laws — i.e., if platforms are susceptible to civil damages only if they knowingly refuse to take down illegal material after having been given meaningful notice that the materials violate federal law — then the carve-outs seem within the range of the perfectly reasonable Citron/Franks proposals, and pose little threat to free expression. The devil is, of course, in the details.

C. Fact-checking and Microtargeting

When we come to proposals to require social media platforms (notably Facebook) to increase fact-checking of political ads and posts by politicians, as well as to restrict microtargeting of political ads, at least within the United States the First Amendment is almost certainly a bar to both proposals. To start with, it is blackletter law under Supreme Court precedent that the First Amendment protects even intentional falsehoods.[196] Furthermore, courts appear to give particularly strong protection to falsehoods in campaign materials and other political speech.[197] A law seeking to directly regulate false political statements on Facebook would therefore almost certainly violate the First Amendment, and a law requiring Facebook to do the same surely would as well because the First Amendment bars the government from requiring a private party to censor content that the government could not directly restrict.[198] A narrower law, targeting only false statements designed to interfere with voting such as lies about poll locations or timings, might survive strict scrutiny; however, since Facebook already blocks such content, such a law would accomplish little. But fundamentally, a law that generally targets false political speech is a constitutional nonstarter. That Facebook or other social media firms, as

[196] United States v. Alvarez, 567 U.S. 709, 723 (2012).

[197] *See, e.g.*, Susan B. Anthony List v. Driehaus, 814 F.3d 466, 473 (6th Cir. 2016); 281 Care Comm. v. Arneson, 766 F.3d 774, 783-84 (8th Cir. 2014).

[198] *See* Denver Area Educ. Telecomms. Consortium v. FCC, 518 U.S. 727, 753-60 (1996) (striking down law requiring cable television operators to restrict access to sexually oriented programming on leased channels).

non-state actors, may voluntarily choose to restrict falsehoods, most assuredly does not mean that the government may force them to do so.

A prohibition or restriction on microtargeting of political advertising, while less clear cut, would also probably violate the First Amendment. To begin with, the speech being restricted — political advertising — lies at the core of the First Amendment, and receives the most robust protection.[199] Furthermore, the right to communicate political speech *must* include an attendant right to choose the audience for the speech. Any other rule would permit the government to hobble the effectiveness of an unpopular speaker by restricting their audience either to those already convinced of the message, or to those adamantly hostile to the speaker's messages. As a consequence, prohibitions or significant restrictions on microtargeted political advertising are almost certainly unconstitutional. On the other hand, precedent supports the imposition of disclosure requirements,[200] suggesting that it would be constitutionally permissible to legally require social media firms to disclose the content of and audience for microtargeted ads, in order to permit "counter speech" to alleviate the effects of deceptive microtargeting. Facebook, the primary vehicle for microtargeted ads, already maintains such a public database, but if (as some claim) Facebook's library is insufficiently transparent,[201] regulation to cure such deficiencies seems on its face permissible (though at least one recent and striking appellate decision raises constitutional questions about such laws, at least as directed at platforms themselves rather than at speakers[202]).

D. Incitement

We now come to perhaps the most intractable, and certainly most troubling problem posed by social media, which is its use to incite widespread, often ethnically based violence. No one — least of all social media firms — question that inciting violence is a serious problem that needs to be addressed. And there is no question that in recent years social media platforms, notably Facebook and WhatsApp, have been used to incite violence. But a pause is also necessary here. If one reads

---

199 *See* Citizens United v. FEC, 558 U.S. 310, 319 (2010); Buckley v, Valeo, 424 U.S. 1, 57-58 (1976).

200 *See Citizens United*, 558 U.S. at 366-71.

201 Edward Ongweso, Jr., *This Tool Lets You See Facebook's Targeted Political Ads All Over the World*, VICE (Aug. 1, 2019, 8:40 AM), https://www.vice.com/en/article/pa7edb/this-tool-lets-you-see-facebooks-targeted-political-ads-all-over-the-world [https://perma.cc/R9F8-2E8B].

202 Wash. Post v. McManus, 944 F.3d 506, 515-17 (4th Cir. 2019).

some recent press commentary,[203] one might understandably get the impression that social media created the problem of mass, ethnic-based violence. This is, of course, nonsense. The countries in which social media platforms have been used to incite such violence have very long histories of ethnic tensions and violence that obviously have no connection to American tech companies. Nor is it completely clear that the Internet has contributed to a rise in such violence. The calls for genocide against Tutsis in Rwanda were, after all, spread via *radio*;[204] yet no one called for restricting radio in response to this unspeakable abuse of the medium.

That said, it seems reasonable to believe that the extraordinary speed with which messages, good and bad, can be spread via social media does increase the risk that calls for violence will spread more quickly via that medium. And even in the United States under the strict *Brandenburg* standard, much less the rest of the world, calls for violence in a volatile situation may legally be suppressed or punished. But the major social media firms do not disagree with that conclusion. The problem, however, is how to achieve that goal as a practical matter given the enormous number of languages in which Facebook — the prime target for criticism — operates. Given the sheer scale of Facebook's operations, there are reasons to doubt if completely blocking incitement is possible, especially because in many countries multiple languages are spoken (India alone has twenty-two official languages and countless other minor ones, often with no relation to each other[205]). Of course, wealthy social media firms can and should be pushed to devote more resources to the task; but again, resources are never unlimited, nor is the pool of potential local-language monitors.

Using legal sanctions to press the issue also threatens to create two major unintended consequences. First, it would create huge barriers to entry for potential competitors to existing social media giants, since few startups could afford to maintain the kind of content moderation network that the incumbents support (and it would seem bizarre to

---

[203] Jamelle Bouie, *Facebook Has Been a Disaster for the World*, N.Y. TIMES (Sept. 18, 2020), https://www.nytimes.com/2020/09/18/opinion/facebook-democracy.html [https://perma.cc/DCL2-JTQL].

[204] *See* Kennedy Ndahiro, *In Rwanda, We Know All About Dehumanizing Language*, ATLANTIC (Apr. 13, 2019), https://www.theatlantic.com/ideas/archive/2019/04/rwanda-shows-how-hateful-speech-leads-violence/587041/ [https://perma.cc/7VXF-933Z].

[205] GOV'T OF INDIA: MINISTRY OF EDUC., LANGUAGE, https://www.education.gov.in/sites/upload_files/mhrd/files/upload_document/languagebr.pdf (last visited Jan. 24, 2021) [https://perma.cc/4VAL-TTWD].

exempt startups from laws restricting incitement of violence).[206] Second, it would create strong incentives for social media firms to simply exit nations and areas in which ethnic tensions exist, and the local languages are not widely spoken. But that would deny access to the dominant communications platform of our time to millions if not billions of blameless individuals, almost all of whom live in the Global South. This seems an odd outcome for progressive reformers to champion.

Related to the problem of direct incitement is speech depicting or glorifying violence, which as noted earlier Australia has adopted strong restrictions on in response to the Christ Church mass shooting.[207] Leaving aside the question of what policy goals this in fact advances — which is not obvious[208] — there is little doubt that such a law would be unconstitutional in the United States. Depictions of violence would not qualify as unprotected incitement under *Brandenburg* since the key element of language "directed to inciting or producing imminent lawless action" is missing;[209] and otherwise there is no question that the First Amendment fully protects depictions of even gratuitous violence.[210] So while other countries might pursue such policies (though again it is unclear what they accomplish), in the United States it is at present impossible.

So perhaps the solution is to reconsider *Brandenburg* and remove constitutional protections for speech calling for, or glorifying violence. Certainly there have been reasonable and thoughtful calls in that direction.[211] It should be noted, though, that Facebook's Community Standards already ban content that glorifies violence,[212] so a change in law will have little impact on the practical problem of how to enforce

---

206 *See* Shields & Brody, *supra* note 192 (documenting concerns raised by smaller firms about their capacity to conduct content moderation).

207 *See supra* note 140 and accompanying text.

208 I suppose the theory is that being unable to live-stream their actions will disincentivize future mass shooters; but the empirical basis for this theory seems rather thin.

209 *See* Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).

210 *See* Brown v. Entm't Merchs. Ass'n, 564 U.S. 786, 842 (2011); United States v. Stevens, 559 U.S. 460, 472 (2010).

211 *See* Eric Posner, *ISIS Gives Us No Choice but to Consider Limits on Speech*, SLATE (Dec. 15, 2015, 5:37 PM), https://slate.com/news-and-politics/2015/12/isiss-online-radicalization-efforts-present-an-unprecedented-danger.html [https://perma.cc/3U4G-UHJ5].

212 *Community Standards: Violent and Graphic Content*, FACEBOOK, https://www.facebook.com/communitystandards/graphic_violence (last visited Jan. 24, 2021) [https://perma.cc/VUL3-ZQ2P].

restrictions promptly. Perhaps the answer is to weaken *Brandenburg*, then to incentivize platforms to address violent speech through the threat of legal sanctions. The assumption underlying this proposal, however, that the only reason tech firms do not better monitor content is lack of incentives to do so, is not evident. More troublingly, overly strict enforcement of rules against depictions of violence can and does have the perverse consequence of suppressing content intended to bring to public attention, and condemn, acts of violence[213] — content which Facebook's Community Standards expressly permits[214] — thereby exacerbating the excessive censorship problem noted earlier. Yet, if social media firms faced legal sanctions for failure to adequately block violent content, they will undoubtedly adopt precisely such an overinclusive strategy. Again, in a world in which perfectly calibrated content moderation is impossible, it is important to bear in mind the potentially perverse consequences of requiring greater content moderation.

Finally, we come to criticisms directed at the use of WhatsApp to incite violence. The reason why WhatsApp is such an effective means to avoid content moderation is because, as noted earlier, it is an end-to-end encrypted platform so that content transmitted on it is invisible to everyone other than the participants in the relevant "chat," including the owner of the platform (Facebook). Facebook did in fact respond to the problem of the spread of incitement and misinformation in April 2020 by placing strict limits on mass-forwarding of viral messages.[215] However, the only way to fully prevent such uses of WhatsApp would require eliminating encryption, or at a minimum providing a back door into the encryption for the platform and local law enforcement officials. But, of course, such a "solution" to the incitement problem comes at the

---

[213] *See Social Media Platforms Remove War Crimes Evidence: Archives Needed to Preserve Content Deemed Dangerous*, HUM. RTS. WATCH (Sept. 10, 2020), https://www.hrw.org/news/2020/09/10/social-media-platforms-remove-war-crimes-evidence [https://perma.cc/H9YA-CLZ4]; *"Video Unavailable": Social Media Platforms Remove Evidence of War Crimes*, *supra* note 77.

[214] *Community Standards: Violent and Graphic Content*, *supra* note 212 ("We allow graphic content (with some limitations) to help people raise awareness about issues.").

[215] Alex Hern, *WhatsApp to Impose New Limits on Forwarding to Fight Fake News*, GUARDIAN (Apr. 7, 2020, 3:00 PM), https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news [https://perma.cc/89B8-FPY4]. In September, Facebook extended this policy to its Messenger platform. Bulbul Dhawan, *Facebook Brings WhatsApp-Like Limit to Forwarding Messages on Messenger; Here is What It Means for Users*, FIN. EXPRESS (Sept. 4, 2020, 2:17 PM), https://www.financialexpress.com/industry/technology/facebook-brings-whatsapp-like-limit-to-forwarding-messages-on-messenger-here-is-what-it-means-for-users/2074749/ [https://perma.cc/S73K-C4HN].

price of eliminating the privacy that social media critics elsewhere strongly support (and that is the very function of encryption to protect). Indeed, providing law enforcement officials with tools to invade privacy seems far more worrisome than current concerns about the use of data to sell targeted advertising, especially in ethnically divided or authoritarian nations (unless, that is, privacy advocates do not believe that the government poses a threat to privacy). In short, it is possible to criticize Facebook for permitting its WhatsApp platform to be used to incite violence, and it is possible to criticize Facebook for giving insufficient protection to privacy interests. But to level both criticisms at the same time is flatly self-contradictory.

E. Excessive Censorship

As noted earlier, in addition to broad attacks on social media for insufficiently moderating harmful content, another set of critics accuse the firms of blocking harmless and socially valuable content. And inevitably, these critics propose a different set of reforms designed to alleviate that problem.

Perhaps the most extreme proposal, as noted earlier, is to treat social media firms as state actors subject to First Amendment restraints. In truth, however, the legal reasoning behind these arguments is very weak. The *Marsh v. Alabama* precedent upon which proponents of this theory rely is widely recognized as an outlier in the Supreme Court's state-action jurisprudence, and its "public function" theory has not been followed in more recent cases.[216] Furthermore, the very idea that operating a communications platform constitutes a public function is an exceptionally weak one, especially in the United States. After all, in the past when new communications technologies became ubiquitous they were typically dominated by one or a few *private* firms. Think in this regard of telegraphs (Western Union), telephony (the Bell System), broadcast television (the three major networks) and cable television (Comcast and Charter (formerly Time Warner)). Yet at no time were any of these entities treated as state actors. The most prominent legal effort in that direction (with respect to broadcast television) was indeed soundly rejected by the Supreme Court.[217] For all of these reasons, it

[216] *See* S.F. Arts & Athletics, Inc. v. U.S. Olympic Comm., 483 U.S. 522, 549-56 (1987); Rendell-Baker v. Kohn, 457 U.S. 830, 842 (1982); Flagg Bros. v. Brooks, 436 U.S. 149, 158-61 (1978); Jackson v. Metro. Edison Co., 419 U.S. 345, 352-53, 371-72 (1974).

[217] *See* Columbia Broad. Sys., Inc. v. Democratic Nat'l Comm., 412 U.S. 94, 133-41 (1973).

seems most unlikely that a court today would classify Facebook as a state actor subject to the First Amendment. To the contrary, as noted earlier, social media firms not only are not subject to First Amendment constraints, they probably enjoy positive editorial rights under the First Amendment, including the right to block content they do not wish to carry.[218]

Regardless of its legal tenability, moreover, the implications of the state-action argument are profound and troubling. If subject to the First Amendment, social media firms would only be permitted to flatly ban those narrow categories of speech that the Supreme Court has identified as unprotected.[219] As noted above, neither hate speech nor most calls for violence constitute unprotected speech. Nor do depictions of violence, even if the audience includes children,[220] factual falsehoods (except under narrow circumstances),[221] curse words,[222] or non-obscene nudity and pornography.[223] If treated as state actors, under well-established law social media firms could suppress such (and other) speech based on its content — as content moderation definitionally does — only if it can show that its actions survive "strict scrutiny," meaning that they are "narrowly tailored to promote a compelling Government interest."[224] This standard, however, has proven almost impossible to satisfy in the First Amendment context except in the context of national security. This means that almost all content moderation in the United States would be found unconstitutional under this approach. Radical indeed.

The argument that social media firms should be treated as public utilities, also discussed above, suffers from many of the same problems as the state-actor argument. Part of the problem, however, is that the primary proponents of this approach, Dipayan Ghosh and Sabeel Rahman, are quite vague about what exactly they mean by utility regulation. If, as they sometimes suggest, all it means is closer scrutiny of mergers and horizontal expansion by tech firms, that seems entirely unproblematic but is a call for actually enforcing existing antitrust laws,

---

[218] *See supra* notes 189–92 and accompanying text.

[219] *See* United States v. Stevens, 559 U.S. 460, 468-71 (2010) (listing unprotected categories, and holding that such categories must be rooted in history).

[220] *See* Brown v. Entm't Merchs. Ass'n, 564 U.S. 786, 842 (2011).

[221] *See* United States v. Alvarez, 567 U.S. 709, 718 (2012).

[222] *See* Cohen v. California, 403 U.S. 15, 18 (1971).

[223] *See* United States v. Playboy Entm't Grp., Inc., 529 U.S. 803, 816 (2000); Erznoznik v. City of Jacksonville, 422 U.S. 205, 213 (1975).

[224] *See* Reed v. Town of Gilbert, 576 U.S. 155, 163-64 (2015); *Playboy Entm't Grp., Inc.*, 529 U.S. at 813.

rather than adopting a "utility" approach. But if more is meant, then the question becomes what. In particular, what is decided odd about both Ghosh's and Rahman's arguments is that they fail to address what have historically been the core obligations of utilities and common carriers: nondiscrimination, and the duty to serve at just and reasonable rates.[225] Yet if Facebook were treated as a utility, then presumably it would be subject to such requirements. That would entail, for one, price regulation, which to my knowledge no one advocates (especially since for users, social media is free). But in addition, a nondiscrimination requirement would appear to doom *all* content moderation by social media firms, except perhaps regarding unprotected speech (though even that is unclear). Once again, radical indeed. Finally, if social media firms truly do enjoy First Amendment editorial rights, as I have argued earlier, then I am fairly confident that those rights cannot simply be stripped away by legislatively labeling the firms "utilities." So, not only is a utility model very ill-suited to the problems posed by social media, it is also probably unconstitutional.

Finally, we come to the proposals, such as Jack Balkin's, that can be corralled within the rubric of "due process of content moderation" — requirements that social media and other tech firms are more transparent about their content moderation practices, and provide users with procedural avenues to challenge moderation decisions. At first glance, this seems like an entirely unproblematic approach. Indeed, on October 22, 2020, Facebook took a significant, voluntary step in that direction when its "Oversight Board" went into operation.[226] The Board is intended to provide an independent forum before which individuals can object to Facebook's decision to take down content (and eventually to other content moderation decisions) and, if their appeal is heard, receive a reasoned explanation for why the decision was affirmed or reversed.[227] Furthermore, if the Oversight Board is perceived as inadequate, or if other social media firms do not adopt similar measures, there seems no reason why similar review processes cannot be legally imposed on social media firms. So long as it is implemented in a

[225] *See, e.g.*, 16 U.S.C. § 824d (2018) (imposing similar obligations on electric utilities); Communications Act of 1934, 47 U.S.C. §§ 201, 202 (1934) (imposing such obligations on communications firms); CAL. PUB. UTIL. CODE § 453 (2008) (prohibiting public utilities from granting preferences or discriminating).

[226] Brian Fung, *Facebook's Oversight Board Is Finally Hearing Cases, Two Years After It Was First Announced*, CNN (Oct. 22, 2020, 12:45 PM ET), https://www.cnn.com/2020/10/22/tech/facebook-oversight-board/index.html [https://perma.cc/D3YZ-LZZ3].

[227] OVERSIGHT BOARD, https://www.oversightboard.com/ (last visited Jan. 24, 2021) [https://perma.cc/7N56-HMZ5].

content-neutral manner, imposing such a legal requirement is likely to survive constitutional scrutiny since it does not trample on social media firms' ultimate authority to make content moderation decisions.

One objection to *requiring* due process of content moderation should be recognized, however, which is that the process is resource- and time-consuming. Certainly Facebook, Twitter, and YouTube can afford to spend the resources needed to create a robust content-moderation process. But what about startups and potential entrants? Like any other regulatory requirement, this one might well raise barriers to entry, thereby bolstering the very big-tech dominance that so many critics decry. Of course, one might exempt smaller firms from such rules, and so avoid creating barriers to entry. Whether such an exemption would face First Amendment challenges is difficult to say — though there is certainly precedent for the proposition that discriminatory regulations of the press, which impose selective burdens on large actors, are constitutionally suspect.[228] Furthermore, one might question the social value of encouraging perhaps artificial entry by social media firms who are not required to comply with principles of access and transparency. But ultimately, the latter is a policy judgment on which, First Amendment concerns aside, reasonable minds can differ.

Finally, it is likely that any procedural restrictions on content moderation are likely to slow decision-making, especially if as Klonick proposes limits are placed on automated content moderation. And while there is often something to be said for a slow and careful approach to making important decisions, in the context of social media where information, disinformation, defamation, and incitement can spread like wildfire, deliberation carries with it a very high price. Given these concerns, and given that tech firms undoubtedly have a better understanding of these tradeoffs than either legislators or outside critics, there are again reasons to question whether legal intervention makes sense here, as opposed to continuing to place public pressure on tech firms to improve their transparency and accept greater public input into their decisions.

CONCLUSION: HUMILITY

Much of this Essay has sounded like a libertarian jeremiad against efforts to regulate tech firms, especially social media. That, however, is not my intent. There are many perfectly reasonable regulatory steps that should find broad support, including tweaking Section 230 as Citron

[228] *See* Minneapolis Star & Tribune Co. v. Minn. Comm'r of Revenue, 460 U.S. 575, 583-84 (1983); Grosjean v. Am. Press Co., 297 U.S. 233, 240-43 (1936).

and Franks propose; enforcing antitrust laws by, for example, forcing Facebook to spin off Instagram (and perhaps WhatsApp) as the FTC is seeking to do; expanding disclosure obligations on platforms regarding political ads they host, including detailed information regarding microtargeting (assuming the courts would permit such a step[229]); and requiring firms to follow their own privacy and content moderation policies via the law of contracts. Prohibitions on political advertising (regarding U.S. elections) paid for by foreign entities, including foreign governments, is also entirely unproblematic.[230]

Another highly desirable reform, proposed by Abby Wood and Ann Ravel, would require *political advertisers* on social media platforms to keep a record of, and disclose in a publicly accessible database, all political advertising they have placed.[231] Such a requirement would provide more transparency into microtargeted advertising, and would thus enable counter speech to such advertising. Furthermore, given the Supreme Court's openness to disclosure and transparency measures even when it is otherwise most hostile to speech regulation,[232] and because recent judicial resistance to disclosure rules have involved obligations imposed on platforms, not advertisers themselves,[233] the Wood/Ravel proposal is also very likely to survive constitutional attack.

Beyond such limited steps, however, it is important to embrace humility. As this Essay has shown, efforts to reign in social media are myriad, but they have serious flaws, risk major unintended consequences, and suffer from internal contradictions. Efforts to solve one problem (incitement of violence) could easily exacerbate other problems (privacy and over-censorship). And many proposals threaten the basic business model of social media, thereby throwing out the baby with the bathwater and denying important services to billions of people worldwide. Social media is still a very young medium — Facebook only became available to the general public in September of 2006[234] — and is still evolving rapidly. Furthermore, despite criticism and setbacks, it is clear that social media firms are taking active steps to try and combat the misuse of their networks in a multiplicity of ways, as described here

---

229 *See supra* note 202 and accompanying text.

230 *See* Agency for Int'l Dev. v. All. for Open Soc'y Int'l, Inc., 140 S. Ct. 2082, 2086 (2020) ("[I]t is long settled as a matter of American constitutional law that foreign citizens outside U.S. territory do not possess rights under the U. S. Constitution.").

231 Wood & Ravel, *supra* note 46, at 1256-63.

232 *See* Citizens United v. FEC, 558 U.S. 310, 366-71 (2010).

233 *See* Wash. Post v. McManus, 944 F.3d 506, 514-15 (4th Cir. 2019).

234 *Company Info: About Facebook*, FACEBOOK, https://about.fb.com/company-info/ (last visited Jan. 24, 2021) [https://perma.cc/JH2P-WTZE].

and elsewhere.[235] And under continuing public and media pressure, I would expect those firms to continue to take further steps, while of course preserving their business model like any other for-profit entity.

I personally suspect that much of the flurry of immoderate criticism of social media over the past several years has been fueled by the (in my mind dubious) proposition that manipulation of social media contributed meaningfully to Donald Trump's election to the Presidency in 2016. With the Trump era now (mercifully) behind us, and (hopefully) unlikely to recur, perhaps it is time to accept that we do not yet know how to "cure" social media, and so leave well enough alone. And if, along the way, critics and the media want to continue to castigate Mark Zuckerberg, by all means continue. Just leave the law out of it.

[235] *See* Yablon, *supra* note 45, at 17-30.