

Fordham Law School

## FLASH: The Fordham Law Archive of Scholarship and History

---

Faculty Scholarship

---

2019

### Recovering Tech's Humanity

Olivier Sylvain

*Fordham University School of Law*, [sylvain@law.fordham.edu](mailto:sylvain@law.fordham.edu)

Follow this and additional works at: [https://ir.lawnet.fordham.edu/faculty\\_scholarship](https://ir.lawnet.fordham.edu/faculty_scholarship)



Part of the [Science and Technology Law Commons](#)

---

#### Recommended Citation

Olivier Sylvain, *Recovering Tech's Humanity*, 119 Colum. L. Rev. F. 252 (2019)

Available at: [https://ir.lawnet.fordham.edu/faculty\\_scholarship/1088](https://ir.lawnet.fordham.edu/faculty_scholarship/1088)

This Article is brought to you for free and open access by FLASH: The Fordham Law Archive of Scholarship and History. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of FLASH: The Fordham Law Archive of Scholarship and History. For more information, please contact [tmelnick@law.fordham.edu](mailto:tmelnick@law.fordham.edu).

## RECOVERING TECH'S HUMANITY

*Olivier Sylvain\**

## INTRODUCTION

Tim Wu's essay, *Will Artificial Intelligence Eat the Law?*, posits that automated decisionmaking systems may be taking the place of human adjudication in social media content moderation. Conventional adjudicative processes, he explains, are too slow or clumsy to keep up with the speed and scale of online information flows. Their eclipse is imminent, inevitable, and, he concludes, just as well.<sup>1</sup>

Wu's essay does not really indulge in the romantic tropes about cyborg robot overlords, nor does he seem to hold a conceit about the superiority of networked technologies. He does not promise, for example, anything similar to Mark Zuckerberg's prophecy to Congress in spring 2018 that artificial intelligence would soon cure Facebook of its failings in content moderation.<sup>2</sup> To the contrary, Wu here is sober about the private administration of consumer information markets. After all, he has been among the most articulate proponents of positive government regulation in this area for almost two decades. The best we can do, Wu argues, is create hybrid approaches that carefully integrate artificial intelligence into the content moderation process.<sup>3</sup>

But in at least two important ways, Wu's essay masks important challenges. First, by presuming the inevitability of automated decisionmaking systems in online companies' distribution of user-generated content and data, Wu obscures the indispensable role that human managers at the Big Tech companies have in developing and selecting their business designs, algorithms, and operational techniques for managing content distribution.<sup>4</sup> These companies deploy these resources to further their

---

\* Professor of Law, Fordham Law School.

1. Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 *Colum. L. Rev.* 2001, 2001–02 (2019) [hereinafter Wu, *AI Eat the Law*].

2. Sarah Jeong, *AI Is an Excuse for Facebook to Keep Messing Up*, *The Verge* (Apr. 13, 2018), <https://www.theverge.com/2018/4/13/17235042/facebook-mark-zuckerberg-ai-artificial-intelligence-excuse-congress-hearings> [<https://perma.cc/6JB6-NJSG>].

3. Wu, *AI Eat the Law*, supra note 1, at 2001–05.

4. By Big Tech companies, I refer to the dozen or so internet companies that dominate the networked information economy, but especially the “big five”: Facebook, Alphabet (the owner of Google), Microsoft, Amazon, and Apple. For the purposes of this Response, under this coinage, I also include Twitter, the social media company which, after Facebook-owned entities, has the second-largest U.S. user base. See J. Clement, *Most*

bottom-line interests in enlarging user engagement and dominating markets.<sup>5</sup> In this way, social media content moderation is really only a tool for achieving these companies' central objectives. Wu's essay also says close to nothing about the various resources at work "behind the screens" that support this commercial mission.<sup>6</sup> While he recognizes that tens of thousands of human reviewers exist, for example, Wu downplays the companies' role as managers of massive transnational production lines and employers of global labor forces. These workers and the proprietary infrastructure with which they engage are invaluable to the distribution of user-generated content and data.

Second, the claim that artificial intelligence is eclipsing law is premature, if not just a little misleading. There is nothing inevitable about the private governance of online information flows when we do not yet know what law can do in this area. This is because courts have abjured their constitutional authority to impose legal duties on online intermediaries' administration of third-party content. The prevailing judicial doctrine under section 230 of the Communications Act (as amended by the Communications Decency Act)<sup>7</sup> (section 230) allows courts to adjudicate the question of intermediary liability for user-generated content when the service at issue "contributes materially" to that content.<sup>8</sup> This is to say that the common law has not had a meaningful hand in shaping intermediaries' moderation of user-generated content because courts, citing section 230, have foresworn the law's application. Defamation, fraud, and consumer protection law, for example, generally hold parties legally responsible for disseminating unlawful information that originates with third parties. But, under the prevailing section 230 doctrine, powerful companies like Facebook, Google, and Amazon do not have any legal obligation to block or remove user-generated content that they have no hand in "creat[ing]" or "develop[ing]."<sup>9</sup> This is a standard that requires

---

Popular Social Networks in the United States in October 2018, Based on Active Monthly Users (in Millions), Statista, <https://www.statista.com/statistics/247597/global-traffic-to-leading-us-social-networking-sites/> [<https://perma.cc/3Z7L-R5FX>] (last visited Sept. 26, 2019).

5. See Sarah T. Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* 33–34 (2019) (discussing how "moderation and screening are crucial steps that protect [internet companies'] corporate or platform brand . . . and contribute positively to maintaining an audience of users willing to upload and view content on their sites").

6. See *id.* at 34–35 (noting that while some content moderation is well suited for "machine-automated filtering," the majority of such work requires human screeners that are "called upon to employ an array of high-level cognitive functions and cultural competencies to make decisions about their appropriateness for a site or platform").

7. 47 U.S.C. § 230 (2012).

8. *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1168 (9th Cir. 2008); accord *Jones v. Dirty World Entm't Recordings LLC*, 755 F.3d 398, 410 (6th Cir. 2014) (describing *Roommates.com* as the "leading case" and applying the "contributes materially" standard); *FTC v. Accusearch Inc.*, 570 F.3d 1187, 1200 (10th Cir. 2009) (applying the "contributes materially" standard).

9. 47 U.S.C. § 230(f)(3).

a substantial amount of involvement on the part of online companies to justify liability. This is why it is not quite right to say, as Wu does here, that we are witnessing the retreat of judicial decisionmaking in this setting. There has never been the chance to see what even modest run-of-the-mill judicial adjudication of content moderation decisions looks like since Congress enacted section 230 over twenty years ago.

The view of online content moderation that Wu advances here is pristine. Its exclusive focus on the ideal Platonic form of speech moderation resonates with the view that the internet can be an open and free forum for civic republican deliberation.<sup>10</sup> In this vein, he appeals to the healthy constitutional skepticism in the United States about government regulation of expressive conduct. One might associate his arguments here with other luminaries who have proposed that we use communication technologies to create opportunities for discovery and progress.<sup>11</sup>

In any case, by presenting the issue of content moderation as a battle between human adjudication and artificial intelligence, Wu's essay here fails to identify the industrial designs, regulatory arrangements, and human labor that have put the Big Tech companies in their position of control. It does not really engage the political economy and structural arrangements that constitute and condition online content moderation.

I generally admire and subscribe to Wu's various accounts and critiques of the networked information economy. He is a clear and eloquent spokesperson for why positive procompetitive regulation and consumer protection in communications markets are vital to the operation of democracy. I, therefore, take his recent essay, and its relatively light touch

---

10. Compare Robert D. Putnam, *Bowling Alone: The Collapse and Revival of American Community* 245–46 (2000) (noting that “the rise of electronic communications and entertainment,” especially television, coincided with “the national decline in social connectedness” and civic disengagement among younger generations, although the evidence was not conclusive on causality), with Cass Sunstein, *Republic.com* 8–9, 167–70 (2001) (suggesting that while technology has given more power to consumers “to filter what they see,” a “widely publicized deliberative domain[] on the Internet, ensuring opportunities for discussion among people with diverse views,” would aid in maintaining a “well-functioning system of free expression”). This approach also recalls far more theoretical treatments of “discourse ethics.” See generally Jurgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (1996) (arguing for a shift from communicative to strategic action in which parties to a dispute spend less time debating validity claims and more time bargaining with threats and promises).

11. See Yochai Benkler, *Wealth of Networks: How Social Production Transforms Markets and Freedom* 2 (2006) (describing a “new information environment” in which users play “a more active role” and its potential to “serve as a platform for better democratic participation” and as a “mechanism to achieve improvements in human development everywhere”); Lawrence Lessig, *Code and Other Laws of Cyberspace* 7–8 (1999) (arguing that cyberspace as an open commons is key to checking government control and advocating for open code); Ithiel de Sola Pool, *Technologies of Freedom* 10 (1983) (“The onus is on us to determine whether free societies in the twenty-first century will conduct electronic communication under the conditions of freedom established for the domain of print through centuries of struggle, or whether that great achievement will become lost in a confusion about new technologies.”).

on the Big Tech companies' content moderation choices, as being addressed to whom he says it is addressed: the designers of these new hybrid processes. In contrast, this Response is addressed to policymakers and reformers: the very people whom Wu has inspired with his other writing. I offer this caveat to say that Wu and I may not actually have a disagreement as a matter of substance. I will just use this generous opportunity to respond to his essay by identifying the reasons we cannot afford to turn away from the lived political economy that shapes our networked world.

### I. CODIFYING CONTENT MODERATION

Communications law, Wu explains, has over the past two centuries calibrated the “implicit competition” between “private and public decisionmaking” and “norms and law.”<sup>12</sup> In the United States, this balance has been informed in large part by the constitutional skepticism about government regulation of information flows.<sup>13</sup> But the recent “arrival of software and artificial intelligence,” Wu explains, presents a “new” challenge to the old public–private equilibrium in communications policy.<sup>14</sup> Today, a handful of powerful private online platforms employ technologies to harvest and redistribute an extraordinary amount of consumer data.<sup>15</sup> Common law adjudication, he concludes, cannot compete.

This was not always the case. For the first decade or so after the commercial deployment of the internet (say, from 1995 to 2010), online intermediaries were avowedly *laissez faire* about user-generated content. Twitter, for example, presented itself as the “free speech wing of the free

---

12. Wu, *AI Eat the Law*, *supra* note 1, at 2006.

13. See, e.g., *Reno v. ACLU*, 521 U.S. 844, 874 (1997) (holding that provisions of the Communications Decency Act of 1996 were unconstitutional because they lacked “the precision that the First Amendment requires when a statute regulates the content of speech”); *N.Y. Times v. Sullivan*, 376 U.S. 254, 283 (1964) (requiring “proof of actual malice for an award of punitive damages” in libel actions “brought by public officials against critics of their official conduct”); see also *Abrams v. United States*, 250 U.S. 616, 628 (1919) (Holmes, J., dissenting) (noting that speech that does not pose an “immediate danger” or “hinder the success of the government arms” should be protected by free speech rights and that such principles do not change during times of war).

14. Wu, *AI Eat the Law*, *supra* note 1, at 2007.

15. I here want to associate myself with Tarleton Gillespie’s very helpful way of writing about platforms, now that the term has been divorced from its connotation among network engineers. See Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* 18–23 (2018).

speech party.”<sup>16</sup> Reddit purported to be the uncensored “front page of the Internet,” warts and all.<sup>17</sup>

It was an exciting time for entrepreneurs and free speech advocates. The most outspoken proponents of the new technology were bullish about the ways in which the new information technology would disrupt markets, governments, and all kinds of centralized power.<sup>18</sup> They generally believed that the internet was a democratizing communication platform that would empower “users” to articulate themselves in ways that incumbent telecommunications and media companies made impossible.<sup>19</sup> This strong argument for unfettered networked communications resonated with the deregulatory, free-market worldview among mainstream politicians in both parties and policy elites from the 1980s into the mid-2000s. The language its proponents used could easily be associated, for example, with President Bill Clinton’s declaration in his 1996 State of the Union address that “the era of big government is over.”<sup>20</sup> Writers and scholars often refer to John Perry Barlow’s powerful *Declaration of the Independence of Cyberspace*—a magnificent artifact of the Davos neoliberal consensus of the day.<sup>21</sup> Also significant is *Cyberspace and*

---

16. Josh Halliday, Twitter’s Tony Wang: ‘We Are the Free Speech Wing of the Free Speech Party,’ *Guardian* (Mar. 22, 2012), <https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech> [<https://perma.cc/7GBW-RKX3>].

17. John-Michael Bond, Understanding Reddit: A Beginner’s Guide to the Front Page of the Internet, *Daily Dot* (Sept. 2, 2018), <https://www.dailydot.com/debug/what-is-reddit/> [<https://perma.cc/4NX2-9DPA>]; Jake Widman & Will Nicol, What Is Reddit? A Beginner’s Guide to the Front Page of the Internet, *Dig. Trends* (May 22, 2019), <https://www.digitaltrends.com/web/what-is-reddit/> [<https://perma.cc/9WSE-ZBEL>].

18. See, e.g., Benkler, *supra* note 11, at 460–73 (concluding that the economic impact of massive changes in information technology attenuates the power of traditional media institutions to mediate the public sphere, providing the opportunity to enhance democracy and individual autonomy); Clay Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations* 21–24 (2008) (“Now that there is competition to traditional institutional forms for getting things done, those institutions will continue to exist, but their purchase on modern life will weaken as novel alternatives for group action arise.”).

19. See, e.g., Benkler, *supra* note 11, at 465 (“The emergence of a networked public sphere is attenuating, or even solving, the most basic failings of the mass-mediated public sphere. . . . The views of many more individuals and communities can be heard.”).

20. See Alison Mitchell, State of the Union: The Overview; Clinton Offers Challenge to Nation, Declaring, ‘Era of Big Government Is Over,’ *N.Y. Times* (Jan. 24, 1996), <https://www.nytimes.com/1996/01/24/us/state-union-overview-clinton-offers-challenge-nation-declaring-era-big.html> (on file with the *Columbia Law Review*).

21. John Perry Barlow, A Declaration of the Independence of Cyberspace, Elec. Frontier Found. (Feb. 8, 1996), <https://www.eff.org/cyberspace-independence> [<https://perma.cc/2W4F-RXSA>] (declaring that obsolete governments of the world will fail to control a new cyberspace civilization); see also Andy Greenberg, It’s Been 20 Years Since This Man Declared Cyberspace Independence, *WIRED* (Feb. 8, 2016), <https://www.wired.com/2016/02/its-been-20-years-since-this-man-declared-cyberspace-independence/> [<https://perma.cc/3LE7-DZDJ>] (reflecting on how Barlow’s early manifesto has aged and reporting that “in recent years, Barlow admits his ideas have become less commonly used as a call to arms than as a political punching bag”).

*the American Dream*, authored by celebrated futurists and entrepreneurs of the time, which claimed that the popular adoption of the internet would make governments less relevant to public life.<sup>22</sup> This, they wrote, “is an inevitable implication of the transition from the centralized power structures of the industrial age to the dispersed, decentralized institutions” of the networked world.<sup>23</sup>

During this period, courts, too, were generally pretty solicitous of the new technology.<sup>24</sup> In *Reno v. ACLU*, the Supreme Court remarked that the internet was a “dynamic, multifaceted category of communication” that exceeded anything humanity had experienced before.<sup>25</sup> It, Justice John Paul Stevens wrote for the Court, could transform “any person with a phone line” into “a town crier with a voice that resonates farther than it could from any soapbox.”<sup>26</sup>

But things have evidently changed over the past decade or so. Above all, as Wu powerfully explains in his recent book, only a handful of firms control the ways in which the vast majority of information flows to users around the world.<sup>27</sup> Amazon, Microsoft, Apple, Alphabet (the owner of Google), and Facebook sit atop the list of such companies based on market capitalization, along with Netflix and Chinese giants Alibaba and Tencent.<sup>28</sup> Facebook (the owner of Instagram and WhatsApp) generates about half the market value of what Microsoft or Amazon do, but its reach to users around the world is unrivaled.<sup>29</sup> I have been using the shorthand “Big Tech companies” to identify them.<sup>30</sup> As a result, these companies, as well as other popular moderators of user-generated content like Twitter and Reddit, have become the de facto arbiters of content

---

22. Esther Dyson, George Gilder, George Keyworth & Alvin Toffler, *Cyberspace and the American Dream: A Magna Carta for the Knowledge Age*, 12 *Info. Soc’y* 295, 297 (1996).

23. *Id.* at 303.

24. See, e.g., *Reno v. ACLU*, 521 U.S. 844, 870, 874 (1997) (holding that provisions of the 1996 Communications Decency Act restricting indecent content on the internet violated the First Amendment); *Zeran v. AOL, Inc.*, 129 F.3d 327, 330 (4th Cir. 1997) (“By its plain language, § 230 [of the Communications Decency Act] creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service.”).

25. 521 U.S. at 870.

26. *Id.*

27. Tim Wu, *The Curse of Bigness: Antitrust in the New Gilded Age* 11 (2018) [hereinafter Wu, *Curse of Bigness*].

28. J. Clement, *Market Capitalization of the Biggest Internet Companies Worldwide as of June 2019* (in Billion U.S. Dollars), Statista, <https://www.statista.com/statistics/277483/market-value-of-the-largest-internet-companies-worldwide/> [https://perma.cc/669Y-DJ6M] (last visited Aug. 28, 2019).

29. J. Clement, *Most Popular Social Networks Worldwide as of July 2019, Ranked by Number of Active Users* (in Millions), Statista (July 22, 2019), <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [https://perma.cc/64EB-SBTQ] (last visited Sept. 18, 2019).

30. See *supra* note 4.

online. Society has come a long way from the democratic view of the internet that prevailed in its nascent period.

And, as concentrated as the social media market has become, the Big Tech companies have become more competitive about the size of the audiences they reach. It is largely in service of this market pressure that they have sought to create “healthy” environments for speech.<sup>31</sup> To Wu, there are three reasons that online platforms have become more attentive to content.<sup>32</sup> First, activists of the past decade have raised general awareness about the material harms that some kinds of expressive conduct exact on historically subordinated groups.<sup>33</sup> Second, the major platforms have done more to limit or block the dissemination of “hateful speech, foreign interference with elections, atrocity propaganda, and hoaxes.”<sup>34</sup> This second concern does not arise from wariness about harm to marginalized groups but rather from the damage that users believe social media distribution of “disinformation” has wrought on democracies across the globe.<sup>35</sup>

The third and final reason for the current state of affairs, according to Wu, has been the aggressive consolidation of the online “speech platform market” by Big Tech.<sup>36</sup> It is little surprise that Wu asserts as much here. He has been among the most articulate advocates of aggressive antitrust regulation of communications markets. These powerful intermediaries, he observes here, now decide in the first instance which kinds of content get distributed to subscribers.<sup>37</sup> They do this at remarkable speeds and with increasing precision. Thus, as Wu puts it, “[W]hat might have been thought to be important public decisions have either been displaced or are beginning to be displaced by software, in whole or in part,” by the sheer force of this enlarged technological capacity.<sup>38</sup> In this way, online platforms, especially those administered by the Big Tech companies, have effectively supplanted the role of judges and legislatures

---

31. See Isaac Chotiner, *The Underworld of Online Content Moderation*, *New Yorker* (July 5, 2019), <https://www.newyorker.com/news/q-and-a/the-underworld-of-online-content-moderation> [<https://perma.cc/34R7-YCMS>] (interviewing Sarah T. Roberts, author of *Behind the Screen*, *supra* note 5, who argues that the purpose of content moderation is for brand management of social media platforms and not a genuine interest in healthier online environments for speech).

32. Wu, *AI Eat the Law*, *supra* note 1, at 2010.

33. *Id.* at 2011–12.

34. *Id.* at 2010.

35. See Siva Vaidhyanathan, *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy* 5 (2018) (“The dominance of Facebook on our screens, in our lives, and over our minds has many dangerous aspects. . . . We’ve seen this in post-election stories about the flurry of ‘fake news,’ which is actually just one form of information ‘pollution.’”).

36. Wu, *AI Eat the Law*, *supra* note 1, at 2010–11.

37. *Id.* at 2013–18.

38. *Id.* at 2007.



as policymakers.<sup>39</sup> Today, the Big Tech companies are increasingly relying on “proactive filters which prevent certain forms of content from being posted at all.”<sup>40</sup> They have employed screening systems for the detection of the dissemination of child pornography, terrorist conspiracies, and the unauthorized distribution of copyright-protected works.<sup>41</sup> In Wu’s account, algorithms now govern online information flows on social media.

## II. HUMAN–TECH HYBRIDS

There can be little doubt that, today, automated decisionmaking systems are everywhere and deeply affecting. Some have been around for decades in, say, air-flight and traffic-light management.<sup>42</sup> Others, like algorithms for pretrial risk assessment and secondary school placement, are of relatively recent vintage, even if not brand new.<sup>43</sup> Today, most of us rely on automated decisionmaking systems for our quotidian online existence—through recommendations, news feeds, stories, and advertisements.

The business models that sustain the most successful platforms rely on such systems to solicit, harvest, collect, analyze, sort, and repurpose massive amounts of consumer data.<sup>44</sup> These same companies rely on similar automated decisionmaking systems to filter out and take down illegal, unauthorized, or offensive material. Last year, Twitter publicly announced its initiative to promote “healthier” dialogue on its site by using screening technologies to block or substantially reduce election med-

---

39. These developments, Wu observes, vindicate Lawrence Lessig’s prediction twenty years ago that the internet would reveal the indispensable role that “code” (shorthand for what we now colloquially refer to as “algorithms”) would have in regulating people’s online conduct. See *id.*

40. *Id.* at 2016–17.

41. *Id.*

42. See, e.g., Clay McShane, *The Origins and Globalization of Traffic Control Signals*, 25 *J. Urb. Hist.* 379, 385 (1999) (describing the rapid adoption of automatic traffic lights in American cities and London during the 1920s); Now—The Automatic Pilot, *Popular Sci. Monthly*, Feb. 1930, at 22 (reporting on an early automatic flight mechanism that kept a plane flying for three hours without human control on a 1930 flight from Ohio to Washington, D.C.).

43. See Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* 3 (2018) (“[W]e have ceded much of [our] decision-making power to sophisticated machines. Automated eligibility systems, ranking algorithms, and predictive risk models control which neighborhoods get policed, which families attain needed resources, who is short-listed for employment, and who is investigated for fraud.”).

44. Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* 3 (2015) (“Everything we do online is recorded . . . Surveillance cameras, data brokers, sensor networks, and ‘supercookies’ record how fast we drive, what pills we take, what books we read, what websites we visit.”); Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* 211 (2016) [hereinafter Wu, *Attention Merchants*] (“Years before Facebook or Google undertook a similar mission, AOL’s business team also began coming up with ways to cash in on the ‘big data’ they had collected: that is, the addresses, phone numbers, and credit card numbers of millions of users.”).

dling, harassment, user-bot activity, and general scams.<sup>45</sup> Similarly, last March, in response to lawsuits, Facebook announced that it would rely on new filters to block the distribution of advertisements in housing, employment, and credit markets that violate civil rights laws.<sup>46</sup> This came within months of the social media giant's announcement that it would use surveys to tweak its News Feed and promote more meaningful engagement among users.<sup>47</sup> And, just this summer, Facebook and its sister social media service, Instagram, announced that they would use new screening tools to limit the advertisement and sale of alcohol and tobacco products, including e-cigarettes.<sup>48</sup>

As pervasive as automated screening technologies are, however, we have not been completely overtaken by them just yet. Human resources remain vitally important to when and how the major platforms publicly distribute user-generated content because, as of now, automated screening technologies are just not good enough to make sense of the massive amount of content that flows through their servers. As information studies scholar Sarah Roberts puts it in her account of the political economy of content moderation, “[T]he complex process of sorting user-uploaded material into either the acceptable or the rejected pile is far beyond the capabilities of software or algorithms alone.”<sup>49</sup> This is why, in all of the instances identified above, the companies also rely on people to correct oversights, redress algorithmic biases, and clean up other mistakes the screening technologies make.<sup>50</sup>

In fact, the Big Tech companies have relied on human resources to moderate user-generated content since the mid-2000s. The largest platforms, for example, have long called on their own users to flag or report

---

45. Louise Matsakis, What Would a ‘Healthy’ Twitter Even Look Like?, WIREd (Mar. 1, 2018), <https://www.wired.com/story/what-would-healthy-twitter-look-like/> [<https://perma.cc/PS6E-3R4P>].

46. Katie Paul, Facebook Agrees to Advertising Overhaul to Settle U.S. Discrimination Suits, Reuters (Mar. 19, 2019), <https://www.reuters.com/article/us-facebook-advertisers/facebook-agrees-to-advertising-overhaul-to-settle-u-s-discrimination-suits-idUSKCN1R02CO> [<https://perma.cc/T2Q2-PCKU>].

47. Ramya Sethuraman, Jordi Vallmitjana & Jon Levin, Using Surveys to Make News Feeds More Personal, Facebook Newsroom (May 16, 2019), <https://newsroom.fb.com/news/2019/05/more-personalized-experiences/> [<https://perma.cc/ZA6L-679H>].

48. David Cohen, People Can No Longer Sell or Exchange Alcohol and Tobacco Products on Facebook and Instagram, Adweek (July 24, 2019), <https://www.adweek.com/digital/people-can-no-longer-sell-or-exchange-alcohol-and-tobacco-products-on-facebook-and-instagram/> (on file with the *Columbia Law Review*).

49. Roberts, *supra* note 5, at 34.

50. Arman Azad, First on CNN: Facebook and Instagram to Restrict Content Related to Alcohol, Tobacco and e-Cigarettes, CNN (July 24, 2019), <https://amp.cnn.com/cnn/2019/07/24/health/facebook-instagram-alcohol-tobacco-bn/index.html> [<https://perma.cc/3U73-W5HM>] (explaining that, in order to implement its new alcohol and tobacco rule, Facebook will “use a combination of technology, human review and reports from our community to find and remove any content that violates these policies”).

content that violates content guidelines.<sup>51</sup> The companies then employ vast global workforces to check on this flagged content.<sup>52</sup> Today, Facebook and Google each employ tens of thousands of salaried workers and contractors around the world who review user-reported material.<sup>53</sup> Depending on the complexity or difficulty of knowing whether any given content should be blocked (think of the live stream of the Christchurch massacre or the doctored video clips of the Speaker of the House of Representatives Nancy Pelosi), these reviewers escalate evaluation through internal appellate processes that, in the hardest cases, might even culminate in a final decision at the highest levels of the company.<sup>54</sup>

Human review is essential today because it confers a degree of legitimacy on the platforms' moderation choices. For Wu, Facebook's plans for an Oversight Board help to illustrate the point. Late last year, months after Facebook head Mark Zuckerberg hinted at it, the social media giant announced its plans for a transnational "Supreme Court" of sorts to rule on whether any especially controversial user-generated content should be taken down.<sup>55</sup> Soon after, it released a draft charter and framework for the plan and sought out feedback from policymakers, activists, and users around the world.<sup>56</sup> The company published the final charter in September of this year.<sup>57</sup> While the fine details of its implementation remain uncertain, Facebook concluded that the Board is to be composed of forty independent experts from around the world.<sup>58</sup> Each member will serve three-year terms and may serve a total of three terms if re-appointed.<sup>59</sup> The Board will develop its procedures.<sup>60</sup> It will publish deci-

---

51. See, e.g., How to Report Things on Facebook, Facebook, <https://www.facebook.com/help/181495968648557> [<https://perma.cc/744Y-D3FM>] (last visited Aug. 28, 2019); Report Abusive Behavior, Twitter, <https://help.twitter.com/en/safety-and-security/report-abusive-behavior> [<https://perma.cc/7GUJ-ZZDD>] (last visited Aug. 28, 2019).

52. Gillespie, *supra* note 15, at 117–24.

53. Roberts, *supra* note 5, at 1–2.

54. See Kate Klonick, Inside the Team at Facebook that Dealt with the Christchurch Shooting, *New Yorker* (Apr. 25, 2019), <https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting> [<https://perma.cc/EW6L-KZVV>].

55. The idea has generally been attributed to Harvard Law professor Noah Feldman. See Mark Sullivan, Exclusive: The Harvard Professor Behind Facebook's Oversight Board Defends Its Role, *Fast Company* (July 8, 2019), <https://www.fastcompany.com/90373102/exclusive-the-harvard-professor-behind-facebooks-oversight-board-defends-its-role> [<https://perma.cc/8UVH-RPJ6>].

56. The company published a summary of the feedback received on the draft charter after conducting workshops and roundtables around the world during the first half of 2019. See Brent Harris, Global Feedback and Input on the Facebook Oversight Board for Content Decisions, Facebook Newsroom (June 27, 2019), <https://newsroom.fb.com/news/2019/06/global-feedback-on-oversight-board/> [<https://perma.cc/VQL9-CNBM>].

57. Facebook, Oversight Board Charter (2019), [https://fbnewsroomus.files.wordpress.com/2019/09/oversight\\_board\\_charter.pdf](https://fbnewsroomus.files.wordpress.com/2019/09/oversight_board_charter.pdf) [<https://perma.cc/6PHQ-HM34>].

58. *Id.* art. I, §§ 1–2.

59. *Id.* art. I, § 3.

sions on content, which must be supported by “plain language” explanations.<sup>61</sup> Over time, these decisions would ostensibly comprise a body of something like “law” for future Board decisionmaking.

Observers have highlighted a variety of potential problems with the Board’s framework. Some are administrative: As in, how quickly will the Board be able to adjudicate especially inflammatory content like the livestream of the Christchurch massacre or the doctored Pelosi videos?<sup>62</sup> Waiting just minutes after the public release of such content is all a person with just a few followers or connections needs to (innocently, mischievously, or malevolently) disseminate especially inflammatory content.<sup>63</sup> Relatedly, will the Board have the capacity to redress different controversies that simultaneously arise in different parts of the world? Other questions are related, but more substantive: As in, how could a forty-member board be capable of authoritatively resolving regionally or culturally idiosyncratic controversies around the world?<sup>64</sup>

These are altogether different from the existential question of what role Facebook (or any online platform) will in fact have in the Board’s governance.<sup>65</sup> According to the Charter, the company will contract for the Board’s services,<sup>66</sup> select the Board’s co-chairs,<sup>67</sup> and establish an irrevocable trust for the compensation of its members and the provision of other financial support.<sup>68</sup> The Charter provides that Facebook will be bound by the Board’s decisions in individual cases, but the company will

---

60. Id. art. II, § 1.

61. Id. art. III, §§ 4, 6.

62. See, e.g., Melanie Ehrenkranz, Facebook Shares More Details About the Content Oversight Board You’ll Be Getting Pissed Off About in the Future, Gizmodo (Jan. 28, 2019), <https://gizmodo.com/facebook-shares-more-details-about-the-content-oversight-1832134512> [<https://perma.cc/8UBD-AL2Y>] (expressing concern that the Board will not be able to tackle the pressing day-to-day content monitoring issues on Facebook).

63. See, e.g., Klönick, *supra* note 54 (“While obvious bad actors were pushing the [Christchurch shooting] video on the site to spread extremist content or to thumb their noses at authority, many more posted it to condemn the attacks, to express sympathy for the victims, or because of the video’s newsworthiness.”).

64. See, e.g., Max Read, Facebook Is Going to Have a Supreme Court. Will It Work?, N.Y. Mag.: Intelligencer (Jan. 30, 2019), <http://nymag.com/intelligencer/2019/01/facebooks-new-oversight-board-is-a-supreme-court.html> [<https://perma.cc/M2KX-J26Q>] (criticizing the proposed board as being “ill-equipped” to address complex issues like Facebook’s role in ethnic cleansing in Myanmar).

65. See generally Glob. Partners Dig., A Rights-Respecting Model of Online Content Regulation by Platforms (2018), <https://www.gp-digital.org/wp-content/uploads/2018/05/A-rights-respecting-model-of-online-content-regulation-by-platforms.pdf> [<https://perma.cc/TF47-L5UB>] (“[T]here is a middle ground between the current purely self-regulatory approach and the development of national-level regulatory or oversight mechanisms.”).

66. Facebook, *supra* note 57, at art. V, § 1.

67. Id. art. I, § 8.

68. Id. art. V, § 1–2.

not be bound by the Board's advisory opinions about Facebook's content policy.<sup>69</sup>

We do not yet know, of course, how all of this will actually play out, but for Wu, the case of the Oversight Board illustrates the emergent “hybrid” approach to online content moderation; Facebook relies on automated screening technologies, user reports, and an elaborate process for human adjudication inside (and potentially outside) of the company to decide which user-generated content to distribute.<sup>70</sup> This kind of overt human involvement, he explains, engenders “deeper acceptance and greater public satisfaction.”<sup>71</sup> Perhaps this is because algorithmic decisionmaking systems typically do not explain their final decisions in language that most people understand.<sup>72</sup> Or perhaps general ambivalence about automated systems arises from artificial intelligence's inability (at least for now) to read social cues that, in real time among real people, elicit trust.<sup>73</sup> In any event, Wu observes, human oversight plainly pre-

---

69. *Id.* art. IV.

70. See Wu, *AI Eat the Law*, *supra* note 1, at 2018–20.

71. *Id.* at 2002–03. One might associate this sentiment with the notion of trust. See generally Ari Ezra Waldman, *Privacy as Trust: Information Privacy for an Information Age* (2018) (exploring how people are more likely to share information on social media in “contexts of trust”).

72. See Wu, *AI Eat the Law*, *supra* note 1, at 2021–22 (“Software can often explain *how* it reached a decision, but not *why*. That may be fine for a thermostat, but is a limitation for a system that is supposed both to satisfy those subjected to it and to prompt acceptance of an adverse ruling.”). Wu's focus here on “procedural fairness,” *id.* at 2–5, recalls what Professor Danielle Citron has called “technological due process.” See Danielle Keats Citron, *Technological Due Process*, 85 *Wash. U. L. Rev.* 1249, 1258 (2008); see also Danielle Keats Citron, *Big Data Should Be Regulated by ‘Technological Due Process,’* *N.Y. Times: Room for Debate* (July 29, 2016), <https://www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/big-data-should-be-regulated-by-technological-due-process> [<https://perma.cc/67PR-646L>] (arguing for greater transparency into automated decisionmaking). For Citron, technological due process requires democratic oversight of algorithmic decisionmaking, including transparency and routine auditing by government agencies. See *id.* She argues, among other things, that such oversight would engender procedural regularity. See *id.*

73. Professors Brett Frischmann and Evan Selinger speculate about the variety of sociological and psychological hurdles that stand in the way of accepting the wide deployment of “techno-social systems,” including the difficulty for laypeople to understand how the technologies work. There are also general worries about the “tipping point” after which automated decisionmaking systems effectively colonize and substantially diminish human agency and freedom. See Brett Frischmann & Evan Selinger, *Re-Engineering Humanity* 10–11 (2018). Anyway, facial recognition technology is years away from helping on that front. See Woodrow Hartzog & Evan Selinger, *Facial Recognition Is the Perfect Tool for Oppression*, *Medium* (Aug. 2, 2018), <https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66> [<https://perma.cc/8TH6-33UG>]. Local governments have also been banning its use. See, e.g., J.E.F., *Why San Francisco Banned the Use of Facial-Recognition Technology*, *Economist* (May 16, 2019), <https://www.economist.com/democracy-in-america/2019/05/16/why-san-francisco-banned-the-use-of-facial-recognition-technology> [<https://perma.cc/9ZFS-RQ2F>]; Sara Ravani, *Oakland Bans Use of Facial Recognition Technology, Citing Bias Concerns*, *S.F. Chron.* (July 17, 2019),

sents the distinctive advantage of “highly trained human judgment,” a feature that is sometimes at odds with or just orthogonal to the decisions that automated systems make in reliance on *ex ante* rules.<sup>74</sup> This kind of reasonable common sense puts the relative equities in any given case into perspective in ways that hard-and-fast rules sometimes do not get right, but that good human judgment does.

### III. CONTENT MODERATION IS INDUSTRIAL DESIGN

Wu’s account here is useful, but it significantly downplays the scope and depth of human involvement in the day-to-day political economy of content moderation at the Big Tech platforms. His essay is principally concerned with explaining the ways in which automated decisionmaking systems (ought to) interact with human judgment. Even on these terms, however, the current law does not create incentives that might influence the Big Tech’s industrial designs. Wu asks here: “Will Artificial Intelligence Eat the Law?” The inquiry is an alluring one for our times. But it obscures who and what is actually at work.

Wu has been among the most effective chroniclers of the networked information economy for almost two decades.<sup>75</sup> He and others, including Professors Siva Vaidhyanathan and Frank Pasquale, for example, have written compellingly about the ways in which the Big Tech companies are driven chiefly by a business model that relies on harvesting as much consumer data as possible and using that data to optimize consumer engagement.<sup>76</sup> They do this in the name of increasing shareholder value.<sup>77</sup>

This remains an important account. So, it is worth stating here plainly what Wu’s essay does not: The ambition to foster “healthy” online engagement, while more than an afterthought, is hardly the Big Tech companies’ main priority. These companies are not (and do not see themselves as) chiefly in the business of calibrating the right balance between human moderators and screening algorithms. Rather, their aim is to hold and expand their dominion over networked information flows. They protect this position by, among other things, developing products

---

<https://www.sfchronicle.com/bayarea/article/Oakland-bans-use-of-facial-recognition-14101253.php> [<https://perma.cc/3D6Q-SYVW>].

74. Wu, *AI Eat the Law*, *supra* note 1, at 2005–06.

75. Wu, *Attention Merchants*, *supra* note 44; Wu, *Curse of Bigness*, *supra* note 27; Tim Wu, *The Master Switch: The Rise and Fall of Information Empires* (2011) [hereinafter Wu, *Master Switch*]. I have engaged with it directly with great appreciation and admiration for as long as I have been writing about information law and policy. Some of my efforts have been more successful than others. E.g., Olivier Sylvain, *Network Equality*, 67 *Hastings L.J.* 443 (2016); Olivier Sylvain, *Wireless Localism: Beyond the Shroud of Objectivity in Federal Spectrum Administration*, 20 *Mich. Telecomm. & Tech. L. Rev.* 121 (2013).

76. See Pasquale, *supra* note 44, at 20; Vaidhyanathan, *supra* note 35, at 83–84; Wu, *Attention Merchants*, *supra* note 44, at 325.

77. See Pasquale, *supra* note 44, at 65–66 (explaining that shareholder demand drives technology companies like Google to use data as a revenue source).

that users cannot resist or acquiring rivals' emergent services and products to avoid any contraction of their market share.<sup>78</sup> They also routinely rely on intellectual property law (such as patent and trade secrets laws), worker contracts (such as nondisclosure and noncompete agreements), and a wide assortment of other legal tools to preserve control over their internal decisionmaking processes.<sup>79</sup> And, besides all of this, they also appeal to the variety of governments and jurisdictions around the world where their users reside, including authoritarian regimes with no compunction about imposing restrictions on political speech.<sup>80</sup> Artificial intelligence surely matters for content moderation. But, in light of all else that is at work, it is more likely an incident of these companies' overt industrial designs on the control and consolidation of the distribution of user information.

The idea that private corporations in the United States have incentives for profit and control is not news. The political economy of media and communications markets over the past two centuries has generally turned on companies' power to provide advertisers access to audiences.<sup>81</sup> As Wu himself has observed elsewhere, since at least the penny press, media companies have traded on their gatekeeper position for advertisers who want to reach users.<sup>82</sup> One difference today is that online platforms have at their disposal powerful tools that track and predict each consumer's idiosyncratic behavior and preferences. They do this with much more precision and at greater scale than ever before. What is more, the platforms, especially the Big Tech companies, exploit the user data that they harvest to develop new services and products that keep their users engaged.<sup>83</sup> Online platforms' capacity for large-scale automated "micro-targeting" paired with their ability to keep consumers' attention has triggered a paradigmatic shift in the media and communications ecosystem. Advertisers have flocked to online platforms (really, affiliate-advertising networks that track consumers' behavior across sites on the web) at the expense of traditional brick-and-mortar media companies whose capacity to reach audiences is more speculative and far less fine-tuned.<sup>84</sup>

---

78. See Lina M. Khan, Note, Amazon's Antitrust Paradox, 126 Yale L.J. 710, 752–55 (2017) (describing Amazon's business model, which relies on the indispensability of its product and its central role in the infrastructure of the internet economy).

79. Pasquale, *supra* note 44, at 3, 193.

80. Adrian Shahbaz, Freedom House, Freedom on the Net: The Rise of Digital Authoritarianism 11–15 (2018), [https://freedomhouse.org/sites/default/files/FOTN\\_2018\\_Final%20Booklet\\_11\\_1\\_2018.pdf](https://freedomhouse.org/sites/default/files/FOTN_2018_Final%20Booklet_11_1_2018.pdf) [<https://perma.cc/88RG-47PT>].

81. See Michael Schudson, *Discovering the News: A Social History of American Newspapers* 93–94 (1978) (“Newspapers became brokers of their own columns, selling their space and the readership it represented to advertisers.”).

82. Wu, *Attention Merchants*, *supra* note 44, at 12.

83. *Id.* at 53–54.

84. *Id.* at 264.

This is why it is not exactly right to suggest, as Wu does in his essay, that *algorithms* are suppressing human adjudication.<sup>85</sup> Automated screening technologies know nothing about it, as they are what Lewis Mumford a century ago called the mere “technics” that firm managers have chosen to deploy to enlarge their commercial reach and capacity to administer online user experiences.<sup>86</sup> In this vein, policymakers ought to be eyeing, above all, the ways in which the Big Tech companies manipulate these tools to exploit their powerful market position between users and advertisers. Artificial intelligence and automated decisionmaking are not to blame. To be sure, those are important aspects of the current state of affairs. But the focus of policymakers and regulators should be on the Big Tech companies’ pecuniary objectives and the designs that they have chosen to use in service of those aims.

These (very human) incentives to control and consolidate information flows explain only a part of the political economy of social media content distribution. The vast infrastructure of hardware and labor “behind the screen” are the indispensable resources on which these companies’ executives organize the content review process. They are assets that must be managed, and they represent costs that must be offset. It may be that the Big Tech companies have not been as deliberate as Frederick Taylor and other proponents of “scientific management” a century ago were about every detail in the production process.<sup>87</sup> But the sequence of human resources that platforms devote to user-generated content review at the Big Tech companies—from automated screening, to users who flag objectionable content, to off-site content moderation contractors and review boards, to in-house reviewers—resembles a factory-floor production line, albeit a transnational and virtual one.<sup>88</sup> The ways in which the Big Tech companies organize all of these resources is central to understanding their distribution of user-generated content.

For the past several years, a handful of intrepid scholars and journalists have reported on the industrial design of the content review process at the Big Tech companies.<sup>89</sup> The pace and quantity of critical writing on

---

85. See Wu, *AI Eat the Law*, *supra* note 1, at 2013–18.

86. See Lewis Mumford, *Technics and Civilization* 426–27 (1934) (“All [technic] mechanisms are dependent upon human aims and desires.”).

87. See Frederick Winslow Taylor, *The Principles of Scientific Management* 7 (1911) (describing the need for “systematic management” to root out inefficiency in the production process).

88. See e.g., *Community Standards, Facebook*, <https://www.facebook.com/communitystandards/> [<https://perma.cc/BT5C-3N3F>] (last visited Aug. 28, 2019) (providing a thorough look at Facebook’s user-generated content review process).

89. See, e.g., Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, *The Verge* (Apr. 13, 2016), <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech> [<https://perma.cc/LXA9-UKCG>]; Adrian Chen, *The Laborers Who Keep Dick Pics and Beheadings out of Your Facebook Feed*, *WIRED* (Oct. 23, 2014), <https://www.wired.com/2014/10/content-moderation/> [<https://perma.cc/8NH5-AZ7N>].



this once-overlooked area have increased substantially over the past few months, signaling an important shift in the popular mind about the online platforms.<sup>90</sup> This, by the way, is not an emergent field of inquiry that is limited to social media content moderation. Kate Crawford and Vladan Joler, for example, have mapped the resources and industrial processes that power Amazon's online retail business.<sup>91</sup> In *Anatomy of an AI System*, which is part narrative and part graphic visualization, Crawford and Joler vividly describe the full range of infrastructures and processes that deliver products to users: from the Echo unit that sits in a living room, to the mines from which companies extract the rare earth elements of which fiber optic cables and satellites are built, to the working conditions at Amazon's distribution centers.<sup>92</sup>

More pertinently, in her new book *Behind the Screen*, Sarah Roberts provides perhaps the most comprehensive account of "industrial-scale" organized and professional content moderation.<sup>93</sup> Key to her account is the human labor involved.<sup>94</sup> And this is vitally important in light of Wu's thesis. This workforce, she explains, brings "an array of high-level cognitive functions and cultural competencies" that continues to elude automated decisionmaking systems.<sup>95</sup> But, Roberts continues, this specific kind of "knowledge work" is difficult to the point of being traumatizing.<sup>96</sup> After all, these are the people who watch, categorize, take down, and block the worst kinds of content: graphic abuse of children, acts of self-harm, gory war zone footage.<sup>97</sup> This is unequivocally ugly and isolating, she explains.<sup>98</sup> But someone has to do it if the internet is to be inviting for the vast majority of users.

Meanwhile, the managers at the Big Tech companies responsible for managing this workforce are making "a business decision on the part of

---

90. See, e.g., Elizabeth Dwoskin, Jeanne Whalen & Regine Cabato, Content Moderators at YouTube, Facebook and Twitter See the Worst of the Web—And Suffer Silently, *Wash. Post* (July 25, 2019), <https://www.washingtonpost.com/technology/2019/07/25/social-media-companies-are-outsourcing-their-dirty-work-philippines-generation-workers-is-paying-price/> (on file with the *Columbia Law Review*); Casey Newton, The Trauma Floor, *The Verge* (Feb. 25, 2019), <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> [<https://perma.cc/35QE-LXSW>].

91. Kate Crawford & Vladan Joler, *Anatomy of an AI System*, *Anatomy of an AI Sys.* (Sept. 8, 2018), <https://anatomyof.ai/> [<https://perma.cc/R24M-SSWP>].

92. *Id.*

93. See Roberts, *supra* note 5, at 1–2.

94. See *id.* at 34–35 (“[T]he vast majority of social media content uploaded by its users requires human intervention for it to be appropriately screened . . .”).

95. *Id.* at 34–35.

96. *Id.* at 104–05.

97. *Id.* at 105.

98. See *id.* at 111–23 (outlining the psychological effects of continual exposure to traumatic videos on commercial content moderators).

the social media companies themselves.”<sup>99</sup> The inclusion of human reviewers has never been, she explains, “a foregone conclusion based on technological necessity or other factors.”<sup>100</sup> To the extent the workers committed to this are in-house, for example, executives at these companies will presumably consider resiliency trainings, in addition to infrastructures that support and guide workers who click through so much graphic content.<sup>101</sup> The same kind of calculus guides managerial decisions about how many reviewers should be outside contractors, as opposed to in-house employees. Indeed, for managers of the content moderation process at the Big Tech companies, the choice to consign this difficult work to outside contractors is suggestive.<sup>102</sup> First, as Roberts explains, by entering into short-term itinerant contracts with these workers, the companies have tacitly acknowledged that any given person cannot stay in the position for more than a couple years because of the trauma it causes;<sup>103</sup> to include such workers within the salaried confines of the Big Tech firms would arguably alter those firms’ obligations to these workers.<sup>104</sup> Second, the cost of labor for such work is simply cheaper in other parts of the world, largely because the host countries, mainly in East Asia, create incentives for transnational firms to locate their workers in industrial zones.<sup>105</sup> At these sites, the companies enjoy “tax exemptions and other sweetheart economic terms that may also include relaxed labor laws or other incentives that leave workers and other citizens at a deficit.”<sup>106</sup>

Wu’s essay needn’t delve into this detail. Again, his essay’s aim is to puzzle through whether and how human adjudication and artificial intelligence can interact.<sup>107</sup> Thus, a reader might think that my point here about industrial designs is fussier or more rhetorical than illuminative, particularly since Wu has elsewhere written extensively about the political economy of information markets.<sup>108</sup> Or perhaps my intervention here just reflects a dispute over emphasis rather than substance.

I insist that it does not. Wu’s essay makes the case that conventional forms of human decisionmaking are sometimes not as effective or

---

99. *Id.* at 35.

100. *Id.*

101. See *id.* at 209–10 (describing early steps being taken by tech companies to support employee wellness).

102. See *id.* at 123.

103. See *id.* at 123–25 (“Assuredly, burnout due to the constant viewing of troubling content was a factor among the commercial content moderation workers at MegaTech and elsewhere.”).

104. *Id.* at 125–37 (outlining the ways in which a Big Tech firm could arguably use employees’ contractor status to “distan[c]e itself from the results of the job on the employees”).

105. *Id.* at 62.

106. *Id.*

107. See Wu, *AI Eat the Law*, *supra* note 1, at 2002–05.

108. See, e.g., Wu, *Master Switch*, *supra* note 75, at 5–7, 10–12.

responsive as platforms' automated screening techniques.<sup>109</sup> But there is nothing inevitable or necessary about artificial intelligence's place in the decisions that platforms make. Nor, to be plain, do such technologies matter in fact as much as he avers here. What matters most today are deeply human concerns: the political economy, geography, incentives, regulatory arrangements, and industrial designs that guide platforms' decisions about when and how to employ different screening techniques. To put the point a little more starkly: Screening technologies do not make content distribution decisions. As rule-bound as they are, they do not eat anything they are told to leave alone. The question of how to structure online content moderation review processes is an industrial design question, not one for algorithms.

#### IV. BIG TECH'S DESIGNS UNFETTERED BY LAW

Focusing on the relative efficacy of artificial intelligence in content moderation (at the expense of human adjudication) is mistaken for another important reason: It clouds extant obstacles to assigning legal responsibility to online intermediaries. The rise of automated decisionmaking is not the main reason judicial adjudication has had little to no hand in regulating Big Tech's publication and distribution of information. Section 230 doctrine is the reason that the courts have read the provision as blocking them from doing so.

In 1996, Congress sought to ensure that "interactive computer services" were "unfettered" by any legal obligation to publish user-generated content with section 230.<sup>110</sup> The courts have read this charge to effectively immunize online platforms from liability for distributing any material to which they do not "materially contribute."<sup>111</sup> So, if there is anything that has eaten the law in this area, it is section 230 doctrine—a creation of legislators and judges, not artificial intelligence.<sup>112</sup>

---

109. See Wu, *AI Eat the Law*, *supra* note 1, at 2021.

110. 47 U.S.C. § 230(b)(2) (2012). See generally Jeff Kosseff, *The Twenty-Six Words that Created the Internet 9–76* (2019) (providing an illuminating account of the statute's enactment).

111. See, e.g., *Fair Hous. Council v. Roommates.com, LLC*, 521 F.3d 1157, 1167–68 (9th Cir. 2008) (distinguishing between "passive conduits" and "co-developers" in determining liability for online websites).

112. See generally Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 *Fordham L. Rev.* 401, 402 (2017) ("Courts applying § 230's immunity provision have dismissed efforts to hold [online platforms] responsible for defamatory posts that have damaged lives and careers."); Olivier Sylvain, *Intermediary Design Duties*, 50 *Conn. L. Rev.* 203, 215 (2018) [hereinafter Sylvain, *Design Duties*] (arguing that "online intermediary designs implicate online services and applications far more than courts have recognized to this point"); Benjamin C. Zipursky, *The Monsanto Lecture: Online Defamation, Legal Concepts, and the Good Samaritan*, 51 *Val. U. L. Rev.* 1, 1–2 (2018) (arguing that the current section 230 doctrine "is alien from what Congress intended").

Wu devotes one paragraph of his essay to section 230.<sup>113</sup> He could be forgiven for this relative inattention. The debate that the statute's mere mention generates in some circles these days is often hyperbolic and sometimes dispiriting and vitriolic—it tends to emit more heat than light. The debate in vogue in Washington, D.C., particularly among right-wing conservatives, is that section 230 requires political neutrality or that it immunizes social media companies from discriminating against conservative viewpoints.<sup>114</sup> Neither is a correct reading of the statute or the doctrine.<sup>115</sup> But, as challenging as discussion about the topic may often be, the doctrine is essential to understanding the basic contours of the current state of affairs. No matter one's view of it, section 230 doctrine should be part of any serious consideration of whether judicial adjudication in this area is receding.

It is not obvious here whether Wu's silence on current section 230 doctrine reflects his support for or indifference about it. Anyway, his relative inattention to the immunity here is glaring in light of the essay's premise about human adjudication. If I were asked to provide a revision to Wu's essay title (and no one has asked me, of course), it would be to offer a qualification: AI is eating the law, and "Congress and jurists have let it."

Section 230's stated purpose is "to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation."<sup>116</sup> Congress set out the workings of the statute in section 230(c). First, under section 230(c)(1), Congress sought to ensure that courts would not "treat[]" online intermediaries as "publishers or speakers" of their users' content.<sup>117</sup> This language was meant to head off judicial application of defamation law to online platforms.<sup>118</sup> Under traditional defamation principles, publishers are as liable for disseminating unlawful speech as the original speaker.<sup>119</sup> Section 230 removes that duty for "interactive computer services" on the theory that the massive amounts of content that flowed through online services (back then, think electronic bulletin boards, online chatrooms, and newsgroups) are impossible to monitor without substantially curtailing the distribution of lawful user-generated content—the application of defamation law would have the effect of

---

113. See Wu, *AI Eat the Law*, *supra* note 1, at 2009.

114. See Sarah Jeong, *Politicians Want to Change the Internet's Most Important Law. They Should Read It First*, *N.Y. Times* (July 26, 2019), <https://www.nytimes.com/2019/07/26/opinion/section-230-political-neutrality.html> (on file with the *Columbia Law Review*).

115. *Id.*

116. 47 U.S.C. § 230(b)(2) (2012).

117. *Id.* § 230(c)(1).

118. See *Zeran v. AOL*, 129 F.3d 327, 330–31 (4th Cir. 1997) (holding that section 230 was intended to shield computer service providers from publisher liability).

119. See *id.* at 331 (outlining "the strict liability standard normally applied to original publishers of defamatory statements").

“chilling” lawful online speech.<sup>120</sup> That is why, the courts have explained, under the statute, an intermediary may only be held liable if it somehow “creat[es] or develop[s]” the offending content.<sup>121</sup>

Congress signaled the second notable feature of the statute in the header of section 230(c): “Protection for ‘Good Samaritan’ Blocking and Screening of Offensive Material.”<sup>122</sup> Here, legislators created a safe harbor that was to shield from liability “interactive computer services” that voluntarily take good faith efforts to block or take down objectionable content.<sup>123</sup> One straightforward reading of the statute would presume that this subsection, section 230(c)(2), is actually the operative provision, and that section 230(c)(1), described above, only removes the presumption that publishing torts apply as usual to online intermediaries.<sup>124</sup> But the courts have since read section 230 so broadly as to effectively write the Good Samaritan safe harbor at section 230(c)(2) out of the statute. The U.S. Court of Appeals for the Fourth Circuit in *Zeran v. AOL* in 1997 was the first federal appellate court to apply the provision.<sup>125</sup> There, the panel concluded that the statute protected AOL from being held liable for its repeated distribution of an anonymous person’s defamatory posts about the plaintiff.<sup>126</sup> In an opinion that would set the tone for the doctrine for the next two decades, Judge Harvey Wilkinson held that, under section 230, America Online had no obligation to block or take down the unlawful content, even when, acting as a distributor, it had knowledge that the inflammatory material at issue was defamatory.<sup>127</sup>

---

120. Felix T. Wu, Collateral Censorship and the Limits of Intermediary Immunity, 87 Notre Dame L. Rev. 293, 303–04 (2011) (“[T]he chilling effects on intermediaries are even greater [than for publishers], and the law ought to account for that difference.”).

121. 47 U.S.C. § 230(f)(3). See generally *Fair Hous. Council v. Roommates.com, LLC*, 521 F.3d 1157 (9th Cir. 2008) (outlining the distinction between a nonliable “service provider” that “passively displays content that is created entirely by third parties” and a liable “content provider” that creates content itself); *Chi. Lawyers’ Comm. for Civil Rights Under Law, Inc. v. Craigslist*, 519 F.3d 666 (7th Cir. 2008) (finding no liability for Craigslist in discriminatory postings on the site, as the service did not “cause” or “induce” such postings); *Zeran*, 129 F.3d 327 (holding section 230 to preclude liability on the part of the online service provider for the comments made by third-party participants); *Blumenthal v. Drudge*, 992 F. Supp. 44 (D.D.C. 1998) (“[Congress] made the legislative judgment to effectively immunize providers of interactive computer services . . . with respect to material disseminated by them but created by others.”).

122. 47 U.S.C. § 230(c).

123. *Id.* § 230(c)(2).

124. See *Sylvain, Design Duties*, supra note 112, at 258 n.298 (discussing *Doe v. GTE Corp.*, 347 F.3d 655 (7th Cir. 2003)).

125. 129 F.3d 327; see also Davis R. Sheridan, *Zeran v. AOL* and the Effect of Section 230 of the Communications Decency Act upon Liability for Defamation on the Internet, 61 Alb. L. Rev. 147, 151 (1997) (“In *Zeran*, the first reported case to consider the scope of the immunity conferred by § 230, the court held that § 230 also confers upon interactive computer services immunity from liability as distributors.”).

126. *Zeran*, 129 F.3d at 332–33.

127. *Id.* at 333.

The *Zeran* opinion and those that have followed have set out an exceptional regulatory arrangement. Courts have applied *Zeran* to a variety of cases involving online material, reaching conclusions that they would never reach in cases involving broadcasters, newspapers, or other traditional media. For example, they have immunized an openly misogynistic blog that encourages visitors to post and comment on defamatory and otherwise degrading material about specific young women.<sup>128</sup> They have immunized an online dating service that did nothing to forbid a user from creating false and lurid profiles of an ex-lover, even after the victim, a nonuser, had sent repeated requests to have the site take the material down.<sup>129</sup> Courts also have applied the immunity well outside of the defamation setting, granting it, for example, to a classifieds page that knowingly facilitated sex trafficking<sup>130</sup> and an online marketplace that is an overt platform for the purchase and sale of unregistered automatic rifles on its site.<sup>131</sup>

We might assume that these “interactive computer services” are exactly what the drafters of the immunity meant to shield. The premise of the broad protection, after all, was to allow online intermediaries to be “unfettered” conduits for the public distribution of user-generated content—that consumers would be the best adjudicators of the content they want.<sup>132</sup> In the late 1990s and early 2000s, when the federal courts forged the doctrine’s contours, it seemed sensible to insulate the new technology from legal regulation if doing so would ensure that the user-generated content could flow freely.<sup>133</sup>

Today, however, intermediaries are far more than simple publishers or distributors of third-party content. And users appear to have few constraints on what they are willing to post or consume. Online platforms have been able to experiment with all kinds of business designs that elicit the worst sorts of expressive conduct and transactions.<sup>134</sup> They can do this, in spite of the Good Samaritan language in the statute, without ever even pretending a “good faith” interest in moderating such conduct and transactions.<sup>135</sup> Today, as I explain in Part III, they carefully engineer and curate our online experiences; they elicit, collect, harvest, sort, analyze, and repurpose consumer data in service of their own business objectives and the prerogatives of the advertisers who have come to rely on them.<sup>136</sup>

---

128. See *Jones v. Dirty World Entm’t Recordings*, 765 F.3d 398, 414–15 (6th Cir. 2014).

129. See *Herrick v. Grindr*, 756 F. App’x 586, 588–89, 591 (2d Cir. 2019).

130. See *Jane Doe No. 1 v. Backpage.com*, 817 F.3d 12, 16–17, 24 (1st Cir. 2016).

131. See *Daniel v. Armslist, LLC*, 926 N.W.2d 710, 714–15, 727 (Wis. 2019).

132. 47 U.S.C. § 230(b)(2) (2012).

133. See *Zeran v. AOL, Inc.*, 129 F.3d 327, 330 (4th Cir. 1997).

134. See *supra* notes 128–131 and accompanying text.

135. 47 U.S.C. § 230(c).

136. Sylvain, *Design Duties*, *supra* note 112, at 205–07 (describing how online platforms have become more than neutral conduits for information and now are commercial enterprises profiting from users’ conduct and information).

This political economy of networked information flows looks nothing like the internet of twenty years ago. And now that just half a dozen or so companies administer most of our online activity, we are a far cry from the days of a user-powered internet populated by a diversity of online forums, news groups, and electronic bulletin boards.

It is also not at all clear that a *laissez-faire* approach to online platform moderation was ever a good idea to begin with. The logic for open and robust debate does not require that every possible thought find expression.<sup>137</sup> The problem with the blanket immunity is that, pursuant to section 230 immunity, platforms need not mind laws that forbid harmful expressive conduct, even as they are the best situated to do so.<sup>138</sup> Never mind the generic safety of online platforms for all users. An unregulated speech environment generally has devastating effects on the people and groups for whom legal protections exist. There are, of course, the people for whom judicial recourse is effectively a dead letter as a general matter under the current doctrine.<sup>139</sup> But, to the extent public laws exist to remedy extant information asymmetries, the people whom such laws are supposed to protect—those who are likeliest to be harmed because of inequality and other background sociopolitical and economic structures—become more exposed to harm if intermediaries presume no native legal or ethical duty to mind third-party content.<sup>140</sup> Online, these harms metastasize rapidly.<sup>141</sup> Individual retail consumers are more vulner-

---

137. See Raymond Geuss, *A Republic of Discussion: Habermas at Ninety*, Point, <https://thepointmag.com/2019/politics/republic-of-discussion-habermas-at-ninety> [<https://perma.cc/8UCK-KY6Y>] (last visited Aug. 28, 2019) (discussing philosophical arguments for why completely unfettered free speech may not be ideal).

138. See Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 *Geo. Wash. L. Rev.* 986, 987–88 (2008) (“[I]ntermediary liability for users’ speech is largely uncoupled from intermediary control over such speech: intermediaries possess power over individual speakers, but they have little or no corresponding responsibility to individuals for the use or abuse of that power.”). See generally Guido Calabresi, *The Costs of Accidents: A Legal and Economic Analysis* (1970) (suggesting that a fault system of accident law may be efficiently tailored to directly assign costs to the least-cost avoiders).

139. Zipursky, *supra* note 112, at 5–6, 14 (discussing how courts’ interpretation of section 230 has made it difficult to hold internet service providers liable for promulgating defamatory content posted by a third party).

140. Cf. Rohit Chopra, Comm’r, FTC Tech Platforms, Content Creators, and Immunity, Prepared Remarks at the American Bar Association Section of Antitrust Law Annual Spring Meeting (Mar. 28, 2019), [https://www.ftc.gov/system/files/documents/public\\_statements/1510713/chopra\\_-\\_aba\\_spring\\_meeting\\_3-28-19\\_0.pdf](https://www.ftc.gov/system/files/documents/public_statements/1510713/chopra_-_aba_spring_meeting_3-28-19_0.pdf) [<https://perma.cc/HZQ8-7UHS>] (“The deployment of online behavioral advertising distorts the incentives of technology companies that might ordinarily be seen as neutral intermediaries.”).

141. Olivier Sylvain, *Discriminatory Designs on User Data*, Knight First Amendment Inst. (Apr. 1, 2018), <https://knightcolumbia.org/content/discriminatory-designs-user-data> [<https://perma.cc/PM6K-PJ96>] [hereinafter *Discriminatory Designs*] (“The harm that these users experience is made worse by the way in which illicit or inflammatory content, once distributed, can spread across the internet at a speed and scale that is hard, if not impossible, to control.”).

able to deceptive advertisements.<sup>142</sup> Voters are misinformed about candidates.<sup>143</sup> Investors are likelier to get duped into purchasing stocks.<sup>144</sup> Members of historically subordinated groups—women, people of color, disabled people, members of the LGBTQ community—are more vulnerable to harassment and discrimination.<sup>145</sup> The ways in which the Big Tech companies design their services, if unchecked, could compound these conditions of inequality.<sup>146</sup>

These critics of section 230 doctrine do not propose, as some skeptics have supposed,<sup>147</sup> that platforms favor certain kinds of speech over others, which would violate the First Amendment if state action were involved (which isn't the case here). Rather, this argument recommends that section 230 doctrine incorporate laws and regulations that protect against harms that worsen extant inequalities. It is in this vein that, in recent years, scholars have written about the perils of the sweeping protection under section 230. Mary Anne Franks, Danielle Citron, Benjamin Wittes, and Ann Bartow in particular have described the ways in which the current doctrine exposes young women, who are the likeliest victims of nonconsensual pornography, to greater injuries.<sup>148</sup> I have

---

142. See, e.g., *Goddard v. Google*, 640 F. Supp. 2d 1193, 1201–02 (N.D. Cal. 2009) (“Plaintiff’s claims would treat Google as the publisher or speaker of third-party content. Yet Plaintiff has failed to allege facts that plausibly would support a conclusion that Google created or developed, in whole or in part, any of the allegedly fraudulent AdWords advertisements.”).

143. See, e.g., Joshua Gillin, *How Pizzagate Went from Fake News to a Real Problem for a D.C. Business*, *PolitiFact* (Dec. 5, 2016), <https://www.politifact.com/truth-o-meter/article/2016/dec/05/how-pizzagate-went-fake-news-real-problem-dc-busin/> [<https://perma.cc/W6UV-G24V>] (“Fake news became all too real over the weekend after a North Carolina man entered a Washington pizzeria with an assault rifle in an attempt to ‘self-investigate’ a false but persistent conspiracy theory about Hillary Clinton.”).

144. See, e.g., Kirsten Korosec, *Elon Musk, SEC Agree to Guidelines on Twitter Use*, *TechCrunch* (Apr. 26, 2019), <https://techcrunch.com/2019/04/26/elon-musk-sec-agree-to-guidelines-on-twitter-use/> [<https://perma.cc/3W6E-DB43>] (discussing the agreement between the SEC and Elon Musk to restrain Musk’s tweets about Tesla to avoid possible securities fraud).

145. See Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* 1–14 (2018) (“The insights about sexist or racist biases that I convey here are important because information organizations, from libraries to schools and universities to governmental agencies, are increasingly reliant on or being displaced by a variety of web-based ‘tools’ . . .”).

146. Separately, I have written about the ways in which the blanket immunity has allowed Facebook and other platforms to develop products (for example, Facebook’s Ad Manager) that analyze and sort harvested user data to generate classifications for advertisers that overtly violate civil rights laws in housing, employment, and credit markets and then target and deliver such content to consumers on behalf of advertisers. Sylvain, *Design Duties*, *supra* note 112, at 208–10.

147. See, e.g., James Grimmelman, *To Err Is Platform*, *Knight First Amendment Inst.* (Apr. 6, 2018), <https://knightcolumbia.org/content/err-platform> [<https://perma.cc/NZ6H-CN4U>].

148. Mary Anne Franks, *The Cult of the Constitution* 127–28 (2019) (arguing that current First Amendment doctrine artificially entrenches inequality because it limits



written, moreover, about the ways in which the doctrine should not be so forgiving of platform designs that facilitate discrimination against historically subordinated or vulnerable users.<sup>149</sup>

Online content moderation, an aspect of Big Tech companies' industrial design, should be treated for what it is: a managerial choice about how these companies control and consolidate information flows to achieve their commercial objectives. But, today, the blanket immunity has effectively given online platforms license to disregard expressive conduct that perpetuates or deepens harms for which policymakers have drafted legal protections.<sup>150</sup>

#### V. WHEN LAW RULES, PLATFORMS WILL FOLLOW

Several developments of the past couple of years suggest that meaningful reform of section 230 doctrine is afoot. In 2018, Congress enacted exceptions to the immunity for intermediaries that knowingly facilitate sex trafficking.<sup>151</sup> Legislators acted in response to news about online marketplaces and classifieds (namely, Backpage.com) that had been engaged in the practice.<sup>152</sup> The statute's wording is vulnerable to constitutional challenge, but it demonstrates that legislators are prepared today to narrow the protection under the statute in ways that did not seem possible just five years ago.

Other recent opinions and litigation suggest that, looking forward, platforms' respective designs on user content and data may not remain as immune from judicial scrutiny as they have been.<sup>153</sup> The civil rights cases

---

regulation of nonconsensual pornography, which "disproportionately affect[s] women and minorities"); Ann Bartow, *Copyright Law and Pornography*, 91 Or. L. Rev. 1, 45 (2012) (suggesting alterations in copyright law as a solution to the broad protections afforded by section 230); Citron & Wittes, *supra* note 112, at 413–14 nn.94, 102 ("[T]he broad reading of § 230 is why revenge porn operators have been so brazen about their business model.").

149. Sylvain, *Design Duties*, *supra* note 112, at 54–55 (arguing that current invocations of section 230 contradict congressional intent to protect vulnerable users); Sylvain, *Discriminatory Designs*, *supra* note 141 ("[T]he current Section 230 doctrine . . . foreclose[s] liability analysis for companies whose service designs routinely facilitate or even encourage illicit content.").

150. Citron and Wittes have proposed a sensible way of reforming the law. They recommend a legislative amendment that would condition immunity on intermediaries' exercise of a reasonable standard of care. Citron & Wittes, *supra* note 112, at 419.

151. See *Allow States and Victims to Fight Online Sex Trafficking Act of 2017*, Pub. L. No. 115–164, 132 Stat. 1253 (2018) (codified as amended in 18 U.S.C.A. §§ 1591(e) (West 2018), 1595, 2421A, 47 U.S.C.A. § 230(e) (West 2018)).

152. Tom Jackman, *House Passes Anti-Online Sex Trafficking Bill, Allows Targeting of Websites like Backpage.com*, Wash. Post (Feb. 27, 2018), <https://www.washingtonpost.com/news/true-crime/wp/2018/02/27/house-passes-anti-online-sex-trafficking-bill-allows-targeting-of-websites-like-backpage-com/> (on file with the *Columbia Law Review*).

153. This Response does not address the European Union's aggressive push under the General Data Protection Directive against the extractive designs of "data controllers." See Anmar Frangoul, *Data Protection in Europe Is About to Transform: Here's How*, CNBC (Mar. 8, 2018), <https://www.cnbc.com/2018/03/08/data-protection-in-europe-is-about-to->

mentioned above against Facebook's administration of its Ad Manager are among them. Civil rights groups and aggrieved users filed five separate lawsuits between late 2016 and 2018, in the wake of excellent reporting by Julia Angwin and Terry Paris Jr. at ProPublica about "machine bias" in Facebook's advertising service.<sup>154</sup> The cases ended in a settlement in March 2019, just a few weeks after oral arguments in federal district courts in New York and California, so there is not yet a full judicial opinion on point.<sup>155</sup>

Plaintiffs' complaints generally alleged that Facebook had developed unlawful content within the meaning of section 230 by creating suspect classifications (such as racial "affinity" groups and age) that the company would, in turn, present to advertisers as characteristics on which those advertisers could find prospective buyers.<sup>156</sup> This allegation would be enough, one would think, to raise eyebrows because civil rights laws flatly prohibit advertisements that hint at such classifications or proxies for them.<sup>157</sup> The complaints invoked some of the same logic on which the U.S. Court of Appeals for the Ninth Circuit held that a section 230 defense did not block the trial court's consideration of dropdown menus

---

transform-heres-how.html [https://perma.cc/Z924-WBHK]. See generally Regulation (EU) 2016/679, 2016 O.J. (L 119) (limiting the processing and movement of personal data by controllers). It is enough to mention here that, on July 29, 2019, the European Court of Justice determined that Belgian data protection authorities were right to sanction a fashion company that used the Facebook "like" plugin on its site because the company became jointly responsible for the collection of user data. See Stephanie Bodoni, Facebook's Like Button Makes Websites Liable, Top EU Court Rules, Bloomberg (July 29, 2019), <https://www.bloomberg.com/news/articles/2019-07-29/facebook-s-like-button-makes-websites-liable-top-eu-court-rules> (on file with the *Columbia Law Review*).

154. Julia Angwin & Terry Paris Jr., Facebook Lets Advertisers Exclude Users by Race, ProPublica (Oct. 28, 2016), <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race> [https://perma.cc/TP9U-BNPN]; Kaya Yurieff, Facebook Settles Lawsuits Alleging Discriminatory Ads, CNN (Mar. 19, 2019), <https://www.cnn.com/2019/03/19/tech/facebook-discriminatory-ads-settlement/index.html> [https://perma.cc/S2PQ-NNWQ].

155. See Sheryl Sandberg, Doing More to Protect Against Discrimination in Housing, Employment, and Credit Advising, Facebook Newsroom (Mar. 19, 2019), <https://newsroom.fb.com/news/2019/03/protecting-against-discrimination-in-ads/> [https://perma.cc/N5ZR-B5WV]; Facebook Settlement, Nat'l Fair Hous. All., <https://nationalfairhousing.org/facebook-settlement/> [https://perma.cc/GRJ9-D73X] (last visited Aug. 28, 2019). For the settlement stipulations, see Stipulation and Order at 1, *Nat'l Fair Hous. All. v. Facebook, Inc.*, No. 1:18-cv-02689-JGK (S.D.N.Y. Mar. 29, 2019); Plaintiffs' Notice of Voluntary Dismissal at 2, *Mobley v. Facebook, Inc.*, No. 5:16-cv-06440-EJD (N.D. Cal. Mar. 19, 2019); Voluntary Dismissal Pursuant to Federal Rule of Civil Procedure 41 at 2, *Riddick v. Facebook, Inc.*, No. 3:18-cv-04529-LB (N.D. Cal. Mar. 19, 2019). The other two cases were disputes filed with the Equal Employment Opportunity Commission. See ACLU, Summary of Settlements Between Civil Rights Advocates and Facebook 2-3 (2019), <https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook> [https://perma.cc/JD4A-KHT8].

156. See Complaint ¶ 20, *Mobley*, No. 5:16-cv-06440-EJD; Third Amended Class Action Complaint ¶ 1, *Riddick*, No. 3:18-cv-04529-LB; Complaint ¶¶ 2, 4, *Nat'l Fair Hous. All.*, No. 1:18-cv-02689 [hereinafter *Nat'l Fair Hous. All. Complaint*].

157. See 42 U.S.C. § 3604(a), (f) (2012).

on Roommates.com that elicited users' gender, sexual orientation, and family status, as well as their preferences for roommates on those dimensions.<sup>158</sup>

Plaintiffs in the most recent cases alleged more, however. Among other things, they argued that Facebook distributed advertisements to “lookalike” demographic groups that its algorithms predicted would be interested.<sup>159</sup> That is, Facebook took steps to reach prospective buyers and renters who were otherwise unknown or unavailable to advertisers in order to better discriminate between buyers.<sup>160</sup> The March 2019 settlement effectively ended all of this.<sup>161</sup> And, while we will never know for sure, it is reasonable to suspect that Facebook settled the case because, quite unlike typical section 230 motions to dismiss that social media companies successfully file,<sup>162</sup> this was a close call. The U.S. Department of Housing and Urban Development filed its own suit against Facebook days after the company settled with the civil rights plaintiffs.<sup>163</sup> It, too, is focused on the ways in which Facebook facilitates racial discrimination in housing markets in the design of its Ad Manager.<sup>164</sup> This suit makes sense, since, no matter whether Facebook abides by the March 2019 settlement terms, their algorithms might still revert to discriminatory advertisement distribution patterns.<sup>165</sup> Anyway, the reviewing court might very well have to entertain the section 230 defense anew and, this time, reach a decision.

Other relatively recent cases that did reach judicial opinion are even more suggestive that courts' solicitude for online intermediaries under section 230 doctrine is waning. This past summer, for example, the U.S. Court of Appeals for the Third Circuit in *Oberdorf v. Amazon* rejected

---

158. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1161, 1169 (9th Cir. 2008).

159. See Third Amended Class Action Complaint, supra note 156, ¶¶ 34–35, *Riddick*, No. 3:18-cv-04529-LB.

160. See *id.* ¶¶ 38–39.

161. See Noam Scheiber & Mike Isaac, Facebook Halts Ad Targeting Cited in Bias Complaints, *N.Y. Times* (Mar. 19, 2019), <https://www.nytimes.com/2019/03/19/technology/facebook-discrimination-ads.html> (on file with the *Columbia Law Review*).

162. See, e.g., *Klayman v. Zuckerberg*, 753 F.3d 1354, 1355 (D.C. Cir. 2014) (affirming the dismissal of a section 230 claim against Facebook and its owner Mark Zuckerberg); *Fields v. Twitter, Inc.*, 217 F. Supp. 3d 1116, 1118 (N.D. Cal. 2016) (granting Twitter's motion to dismiss).

163. Ariana Tobin, HUD Sues Facebook over Housing Discrimination and Says the Company's Algorithms Have Made the Problem Worse, *ProPublica* (Mar. 28, 2019), <https://www.propublica.org/article/hud-sues-facebook-housing-discrimination-advertising-algorithms> [<https://perma.cc/6RWX-ANZP>].

164. *Id.*

165. Olivier Sylvain, Opinion, A Watchful Eye on Facebook's Advertising Practices, *N.Y. Times* (Mar. 28, 2019), <https://www.nytimes.com/2019/03/28/opinion/facebook-ad-discrimination-race.html> [<https://perma.cc/6FQH-P4H4>] (“[E]ven if Facebook substantially narrows the scope of targeting, there is no reason to believe that its algorithms might not revert to discriminatory ad distribution patterns.”).

Amazon's section 230 defense to a product liability claim.<sup>166</sup> Plaintiff, an Amazon user who blamed the online retailer for a defective leash manufactured by someone else, alleged that Amazon was not entitled to the immunity because the injuries she suffered arose from Amazon's actions as a seller, not as a publisher.<sup>167</sup> The panel agreed, explaining that Amazon "plays a large role in the actual sales process. This includes receiving customer shipping information, processing customer payments, relaying funds and information to third-party vendors, and collecting the fees it charges for providing these services."<sup>168</sup> The panel remanded the case to the lower court to identify the claims in plaintiff's complaint that alleged "actions or failures in the sales or distribution processes," rather than the editorial "failure to warn."<sup>169</sup> That is, plaintiff's complaint would survive Amazon's motion to dismiss if the trial court finds that its allegations are addressed "to selling, inspecting, marketing, distributing, failing to test, or designing."<sup>170</sup> Just weeks later, a federal trial court in the Western District of Wisconsin cited the Third Circuit's opinion to reject Amazon's section 230 defense in another strict liability claim under a similar state law theory of product liability.<sup>171</sup>

The Amazon opinions align with an emergent line of cases, starting with *Barnes v. Yahoo*, an opinion seemingly written for law school teaching, in which the court rejected a section 230 defense.<sup>172</sup> This was another case involving an aggrieved ex-lover who created false and lurid accounts of a former lover.<sup>173</sup> Yahoo failed to take the material down after hearing repeatedly from plaintiff.<sup>174</sup> What is worse, on one occasion, it failed to take down the material even after one of its representatives told plaintiff that the company would do so.<sup>175</sup> Plaintiff alleged two claims: negligence for failure to block or take down the profiles, and promissory estoppel for lulling plaintiff into thinking Yahoo would take the profile

---

166. *Oberdorf v. Amazon.com Inc.*, 930 F.3d 136, 151 (3d Cir. 2019) (concluding that some of plaintiff's claims were barred by section 230).

167. *Id.* at 153 (discussing *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096 (9th Cir. 2009)).

168. *Id.* This is to say nothing of the way in which Amazon assumes an even greater role as a seller under its fulfillment service, not at issue in this Third Circuit case, where it actually takes possession of the seller's goods and manages the delivery. *Id.* at 141–42.

169. *Id.* at 153–54.

170. *Id.*; cf. *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 851 (9th Cir. 2016) (denying preemption of duty to warn relating to defendant's online practices).

171. *State Farm Fire & Cas. Co. v. Amazon.com, Inc.*, No. 18-cv-261-jdp, 2019 WL 3304887, at \*7 (W.D. Wis. July 23, 2019) ("Courts that have considered whether the CDA applies to strict liability claims against Amazon have held that it does not.").

172. *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096, 1105–06 (9th Cir. 2009).

173. *Id.* at 1098 ("Barnes did not authorize her now former boyfriend to post the profiles, which . . . contained nude photographs . . . and some kind of open solicitation . . . to engage in sexual intercourse.").

174. *Id.* at 1098–99.

175. *Id.*

down.<sup>176</sup> The court rejected plaintiff's first claim on section 230 grounds, holding that it was a defamation claim masquerading as a negligence claim.<sup>177</sup> But it rejected the motion to dismiss as to the second claim.<sup>178</sup> Plaintiff's alleged injury, the court explained, arose from Yahoo's promise that it would take the material down, which was an intervening decision to "de-publish."<sup>179</sup>

Courts have repeatedly cited the *Barnes* case to distinguish between claims addressed to publishing and those arising from nonpublishing events. The Tenth Circuit cited *Barnes* to reject a section 230 defense in *Federal Trade Commission v. Accusearch*.<sup>180</sup> There, the defendant company operated an elaborate business that provided paying consumers with the private confidential telephone records of anyone by actively enlisting third-party researchers to find such records in violation of privacy law.<sup>181</sup> Similarly, in *Federal Trade Commission v. LeadClick Media*, the Second Circuit, citing *Barnes*, rejected the defendant's section 230 defense because the defendant orchestrated the development of "fake news" on sites across the web about weight loss products.<sup>182</sup> The *LeadClick Media* panel held that the section 230 defense did not apply because the FTC's allegations arose from the defendant's deceptive actions as the manager of an advertising network, not as a publisher.<sup>183</sup> And, in *Airbnb v. San Francisco*, the Northern District of California found that the city's ordinance barring booking services for short-term rentals that are not registered with the city does not violate section 230 doctrine.<sup>184</sup> Citing *Barnes*, the court explained that the law imposes requirements that are unrelated to Airbnb's editorial decision about which units to make available to guests.<sup>185</sup>

*Barnes* and its progeny have added important texture to section 230 doctrine. They, along with the Facebook Ad Manager settlements, are also more than suggestive that the courts' laissez-faire approach to platform moderation is not inevitable. Indeed, by the looks of it, judges' earlier reticence may be giving way to serious but ordinary judicial scrutiny. The systems and business designs that online companies develop and operate are being seen for what they are: tools and techniques for

---

176. *Id.* at 1099.

177. *Id.* at 1105–06.

178. *Id.* at 1107–09.

179. *Id.*

180. 570 F.3d 1187, 1205 (10th Cir. 2009) (Tymkovich, J., concurring).

181. *Id.* at 1206.

182. 838 F.3d 158, 177 (2d Cir. 2014) ("Accordingly, because LeadClick's Section 5 liability is not derived from its status as a publisher or speaker, . . . Section 230 immunity should not apply.")

183. *Id.*

184. *Airbnb, Inc. v. City & County of San Francisco*, 217 F. Supp. 3d 1066, 1074 (N.D. Cal. 2016).

185. *Id.* at 1072–73.

advancing online companies' own commercial mission, not a pristine public-regarding exercise to promote "safe" or "healthy" conversations.

What would the world look like if section 230 doctrine were not the obstacle to litigation that it has been for the past two decades? At a minimum, judges would be less likely to grant section 230 motions to dismiss (or, as in *Airbnb v. San Francisco*, declaratory judgment claims brought pursuant to that provision). This means more discovery which, in turn, will give courts the occasion to scrutinize the ways in which platforms develop content review procedures and the incentives for doing so, as well as the design and operation of screening technologies and the guidelines that moderators must implement. In these, courts might find evidence of overt collaboration between platforms and "content developers."<sup>186</sup> Or they might find that platform managers set out to repurpose user information for their own commercial gain, at the expense of those very users.<sup>187</sup>

Courts would probably also seriously consider the merits of any given plaintiff's legal claims and theories about causation and harm. That is, they will draw on time-tested concepts of legal responsibility that the prevailing section 230 doctrine has shut out of consideration. In such a world, courts would no doubt begin to help reach an equilibrium between the industrial interest in innovation on the one hand and legal protections for consumers and historically subordinated groups on the other. The very notion that such interventions would be available is surely worrying to the champions of unfettered internet freedom.<sup>188</sup> But, at least then, courts would begin to deliberate at trial and on appeal about how far platforms may go in their designs on user-generated content and data.

We might learn, for example, that, to best protect consumers, online platforms ought to consider the regulatory impacts of any update or new service before its public release. Over the past couple of years, scholars

---

186. See, e.g., *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 17 (1st Cir. 2016) (describing how Backpage provided a platform for online escort services to advertise underage girls); see *Accusearch*, 570 F.3d at 1191–92 (discussing the process by which Accusearch collaborated with third-party researchers to provide confidential telephone records to paying customers).

187. See, e.g., *Nat'l Fair Hous. All. Complaint*, supra note 156, ¶ 122 ("Facebook excludes members of the public from access to housing advertising and information about housing opportunities based on their sex, familial status, disability, race, national origin, or other protected characteristics."); cf. Adam Sternbergh, *A Dystopian Cocktail, Served Chilled with a Twist*, N.Y. Times (July 26, 2019), <https://www.nytimes.com/2019/07/26/books/review/empty-hearts-julie-zeh.html> (on file with the *Columbia Law Review*) (reviewing Juli Zeh, *Empty Hearts* (2019), describing one character as "Britta, 'a wife, mother and successful businesswoman, [who] runs a start-up called, innocuously, The Bridge, which algorithmically scours the internet in search of despondent people, then matches them up with terrorist organizations to act as suicide bombers'" (internal quotation marks omitted) (quoting Zeh, supra)).

188. See, e.g., supra notes 16–23 and accompanying text.

and advocates have proposed, for example, “algorithmic impact assessments.”<sup>189</sup> One such proposal, although addressed to the specific case of government procurement of such systems (say, again, for pretrial detention determinations or school placement), sets out procedures for public notice and evaluation of any new system’s impact on different communities.<sup>190</sup> Other similar proposals recommend that government administrators of automated decisionmaking systems explain the processes that produce the system’s design, as well as its outcome.<sup>191</sup> While, again, these proposals generally address government administration of such systems, such an approach could also apply to platforms in areas where courts and policymakers have already imposed constitutionally valid limits on information flow, including consumer protection,<sup>192</sup> civil rights,<sup>193</sup> and securities law.<sup>194</sup> In all of these circumstances, courts would not refrain from scrutinizing the inner workings—what I have been calling the “designs”—of the respective services.

Critics might raise doubts about the courts’ institutional competence to fully understand this dynamic line of business. But this challenge would not be a new jurisprudential problem. To the contrary, this challenge really just revisits one of the most enduring questions about generalist judges deciding complex cases across substantive areas.<sup>195</sup> In any event, there are too many cases in the recent past that suggest the opposite.<sup>196</sup> And, even if courts did not have the sophistication to understand the specific workings of platform moderation tools and techniques, there are institutional mechanisms and judicial doctrines to redress this chal-

---

189. See, e.g., Dillon Reisman, Jason Schultz, Kate Crawford & Meredith Whittaker, *AI Now Inst., Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* 4 (2018), <https://ainowinstitute.org/aiareport2018.pdf> [<https://perma.cc/6AX9-Q78H>]; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 *Fordham L. Rev.* 1085, 1133–35 (2018).

190. Reisman et al., *supra* note 189, at 13–14.

191. Selbst & Barocas, *supra* note 189, at 1136–37.

192. See, e.g., *FTC v. LeadClick Media, LLC*, 838 F.3d 158, 176–77 (2d Cir. 2014).

193. See, e.g., *Nat’l Fair Hous. All. Complaint*, *supra* note 156, ¶ 123.

194. See, e.g., *SEC v. Todd*, 642 F.3d 1207, 1222–23 (9th Cir. 2011).

195. See *Ass’n for Molecular Pathology v. Myriad Genetics, Inc.*, 569 U.S. 576, 596 (2013) (Scalia, J., concurring in part and concurring in the judgment) (refusing to join the majority opinion portions that went into fine detail about molecular biology because of his lack of expertise in the field).

196. See, e.g., *ABC, Inc. v. Aereo, Inc.*, 134 S. Ct. 2498, 2503 (2014); *MGM Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913, 919–20 (2005); *Reno v. ACLU*, 521 U.S. 844, 849–53 (1997); *Turner Broad. Sys., Inc. v. FCC*, 512 U.S. 622, 627–29 (1994); *MCI Telecomms. Corp. v. AT&T Co.*, 512 U.S. 218, 220–21 (1994); *NBC, Inc. v. United States*, 319 U.S. 190, 194 (1943); *Comcast Corp. v. FCC*, 526 F.3d 763, 764–65 (D.C. Cir. 2008); *WPIX, Inc. v. ivi, Inc.*, 765 F. Supp. 2d 594, 598–99 (S.D.N.Y. 2011); *United States v. AT&T Co.*, 552 F. Supp. 131, 197–200 (D.D.C. 1982).

lenge.<sup>197</sup> If agencies start playing a greater role in taking on online platform moderation of content, without fear of blanket immunity, courts would abide by longstanding deference norms in administrative law doctrine.<sup>198</sup>

#### CONCLUSION

The big tech firms, policymakers, and courts are poised to reform the ways in which to govern social media content. The proposed changes, as varied as they are, will one way or another allocate responsibility between the stakeholders in new ways. This Response has argued, among other things, that any such reforms should not treat social media applications as pristine speech platforms for earnest democratic debate. Rather, in recognition of the work of many scholars, including Wu's own impressive work, I assert that reform should address social media companies as commercial enterprises whose priority is to maximize user attention and engagement for advertisers. It is for this reason, I assert, that policymakers should reject efforts to defer to the companies' own automated decisionmaking systems for moderating content. These technologies are just one asset in the political economy and industrial designs of social media. They have not displaced the vital role that human labor and human incentives play. It is for this reason, I argue, that it is far too soon to write off untested but extant resources in law.

---

197. E.g., *Whitman v. Am. Trucking Ass'ns*, 531 U.S. 457 (2001); *Motor Vehicle Mfrs. Ass'n of the U.S. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29 (1983); *Vt. Yankee Nuclear Power Corp. v. NRDC*, 435 U.S. 519 (1978); *SEC v. Chenery Corp.*, 332 U.S. 194 (1947).

198. See Ronald J. Krotoszynski, Jr., "History Belongs to the Winners": The Bazon-Leventhal Debate and the Continuing Relevance of the Process/Substance Dichotomy in Judicial Review of Agency Action, 58 *Admin. L. Rev.* 995, 998 (2006) ("[A]n agency that scrupulously observes fundamentally fair processes will receive a higher measure of deference from a reviewing court."). That a growing faction on the current Supreme Court has expressed some doubts about some deference doctrines does not recommend their discontinuance.