GEORGETOWN LAW TECHNOLOGY REVIEW

PLATFORMS ARE NOT INTERMEDIARIES

Tarleton Gillespie[*]

CITE AS: 2 GEO. L. TECH. REV. 198 (2018)

Content moderation is such a complex and laborious undertaking, it is amazing that it works at all and as well as it does. Moderation is hard. This should be obvious, but it is easily forgotten. Policing a major platform turns out to be a resource intensive and relentless undertaking; it requires making difficult and often untenable distinctions between the acceptable and the unacceptable; it is wholly unclear what the standards for moderation should be, especially on a global scale; and one failure can incur enough public outrage to overshadow a million quiet successes. And we as a society are partly to blame for having put platforms in this untenable situation by asking way too much of them. We sometimes decry the intrusions of moderators and sometimes decry their absence. Users probably should not expect platforms to be hands-off *and* expect them to solve problems perfectly *and* expect them to get with the times *and* expect them to be impartial and automatic.

Even so, we have handed over the power to set and enforce the boundaries of appropriate public speech to private companies. This enormous cultural power is held by a few deeply invested stakeholders, and it is being wielded behind closed doors, making it difficult for anyone else to inspect or challenge their decisions. Platforms frequently, and conspicuously, fail to live up to our expectations. In fact, given the enormity of the undertaking, most platforms' own definition of success includes failing users on a regular basis.[1]

Social media companies have profited by selling the promises of the web and participatory culture back to us: open participation, free information, expression for all, a community right for you waiting to be

[*] Parts of this article have been excerpted from TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA (Yale U. Press 2018). Tarleton Gillespie is a principal researcher at Microsoft Research and an affiliated associate professor in the Department of Communication and Department of Information Science at Cornell University.

[1] As an example, Del Harvey explained Twitter's dilemma in this way: "Given the context of the scale we're dealing with, if you're talking about a billion tweets, and everything goes perfectly right 99.999% of the time, then you're still talking about 10,000 tweets where everything might not have gone right." *Content Moderation at Scale Summit – Under the Hood: UGC Moderation*, YOUTUBE (May 7, 2018), https://www.youtube.com/watch?v=stB23tNBl2o [https://perma.cc/AMU7-G55B].

found. But as those promises have begun to sour, and the reality of these platforms' impact on public life has become more obvious and complicated, these companies are beginning to actually grapple with how best to be stewards of public culture, a responsibility that was not evident to them at the start.

It is time for the discussion about content moderation to expand beyond the harms users face—and the missteps platforms sometimes make in response—to a more expansive examination of the platforms' responsibilities. For more than a decade, social media platforms have portrayed themselves as mere conduits, obscuring and disavowing their active role in content moderation.[2] When they acknowledge moderation at all, platforms generally frame themselves as open, impartial, and noninterventionist—in part because their founders fundamentally believe them to be so and in part to avoid obligation or liability. Their instincts have been to dodge, dissemble, or deny every time it becomes clear that, in fact, they powerfully shape and censor public discourse.[3]

In other words, tools matter. Our public culture is, in important ways, a product of their design and oversight. Platforms do not just mediate public discourse: they constitute it.[4] While we cannot hold

[2] *See generally* Philip Napoli & Robyn Caplan, *Why Media Companies Insist They're Not Media Companies, Why They're Wrong, and Why It Matters*, 22 FIRST MONDAY (2017), http://firstmonday.org/ojs/index.php/fm/article/view/7051/6124 [https://perma.cc/LC7N-8JX6]; Frank Pasquale, *Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power*, 17 THEORETICAL INQUIRIES IN L. 487 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2779270 [https://perma.cc/9X5W-793A].

[3] JEAN BURGESS & JOSHUA GREEN, YOUTUBE: ONLINE VIDEO AND PARTICIPATORY CULTURE (Polity Press 2009); Carolyn Gerlitz & Anne Helmond, *The Like Economy: Social Buttons and the Data-Intensive Web*, 15 NEW MEDIA & SOC'Y 1348 (2013), https://pdfs.semanticscholar.org/fbd2/67f56b61934918e7ccefdf7f4d0c2e1d96e5.pdf [https://perma.cc/KN2X-RW9D]; James Grimmelmann, *Speech Engines*, 98 MINN. L. REV., 868 (2014), https://james.grimmelmann.net/files/articles/speech-engines.pdf [https://perma.cc/2A8Z-PWDH]; Gaenele Langlois, *Participatory Culture and the New Governance of Communication: The Paradox of Participatory Media*, 14 TELEVISION & NEW MEDIA 91 (2013), http://journals.sagepub.com/doi/pdf/10.1177/1527476411433519 [https://perma.cc/G25L-BF4P];

[4] *See generally* Nancy Baym & danah boyd, *Socially Mediated Publicness: An Introduction*, 56:3 J. BROADCASTING & ELECTRONIC MEDIA 320 (2012), https://www.tandfonline.com/doi/pdf/10.1080/08838151.2012.705200 [https://perma.cc/7MM7-3V9X]; David Beer & Roger Burrows, *Sociology and, of, and in Web 2.0: Some Initial Considerations*, 12 SOC. RES. ONLINE (Sept. 30, 2007), http://www.socresonline.org.uk/12/5/17.html [https://perma.cc/5SPR-KKC2]; Tarleton Gillespie, *The Politics of 'Platforms'*, 12 NEW MEDIA & SOC'Y 347 (2010), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1601487 [https://perma.cc/6BCG-YZAZ]; T. L. Taylor, *The Social Design of Virtual Worlds: Constructing the User and Community through Code*, *in* INTERNET RESEARCH ANNUAL, VOL. 1:,SELECTED PAPERS

platforms responsible for the fact that some people want to post pornography, or mislead, or be hateful to others, we are now painfully aware of the ways in which platforms invite, facilitate, amplify, and exacerbate those, and other, negative tendencies. The public problems we now face are old information challenges paired with the affordances of the platforms they exploit: networked, coordinated harassment; misinformation buoyed by its algorithmically-calculated popularity; polarization as a side effect of personalization; bots speaking as humans, humans speaking as bots; the tactical gaming of platforms in order to simulate genuine cultural participation and value. In all of these ways, and others, platforms invoke and amplify particular forms of discourse, while moderating away others, all in the name of being impartial conduits of open participation. The controversies around content moderation over the last half decade have spurred the slow recognition that platforms now constitute a powerful infrastructure for knowledge, participation, and public expression. This infrastructure includes what is prohibited and how that prohibition is enforced.[5]

FROM THE ASSOCIATION OF INTERNET RESEARCHERS CONFERENCES 2000–2002 260 (2004), http://tltaylor.com/wp-content/uploads/2009/07/Taylor-SocialDesign.pdf [https://perma.cc/VSV7-DM5B]; SIVA VAIDHYANATHAN, THE GOOGLIZATION OF EVERYTHING (AND WHY WE SHOULD WORRY) (U.C. Press 2012); JOSÉ VAN DIJCK, THE CULTURE OF CONNECTIVITY: A CRITICAL HISTORY OF SOCIAL MEDIA (Oxford U. Press 2013); José van Dijck & Thomas Poell, *Understanding Social Media Logic*, 1 MEDIA & COMM. 2, (2013), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2309065 [https://perma.cc/W29H-RK5C]; Esther Weltevrede, Anne Helmond, & Carolin Gerlitz, *The Politics of Real-Time: A Device Perspective on Social Media Platforms and Search Engines*, 31 THEORY, CULTURE & SOC'Y, 125, (2014).

[5] *See generally* Jack Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, U.C. DAVIS L. REV. (forthcoming 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3038939 [https://perma.cc/2GZ2-6RYK]; Laura DeNardis & Andrea Hackl, *Internet Governance by Social Media Platforms*, 39 TELECOMM. POL'Y 761, (2015); James Grimmelmann, *The Virtues of Moderation: Online Communities as Semicommons*, 17 YALE J. L. & TECH. 42 (2015), http://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1110&context=yjolt [https://perma.cc/QPF3-ZN24]; Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018), https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/ [https://perma.cc/YR3U-YRW7]; Alice Marwick, *Are There Limits to Online Free Speech?,* DATA & SOC'Y RES. INST. (Jan. 5, 2017), https://points.datasociety.net/are-there-limits-to-online-free-speech-14dbb7069aec [https://perma.cc/NX4D-BUXQ]; SARAH ROBERTS, *Commercial Content Moderation: Digital Laborers' Dirty Work*, *in* INTERSECTIONAL INTERNET: RACE, SEX, CLASS AND CULTURE ONLINE, 147, (Noble and Tynes, Eds., 2016); Rebecca Tushnet, *Power without Responsibility: Intermediaries and the First Amendment.*, 76 GEO. WASH. L. REV. 101 (2008).

~ ~ ~

Our thinking about platforms must change. It is not just that all platforms moderate, nor that they have to moderate, nor that they tend to disavow it while doing so. It is that moderation, far from being occasional or ancillary, is in fact an essential, constant, and definitional part of what platforms do. Moderation is the essence of platforms. It is the commodity they offer. It is their central value proposition.

First, moderation is a surprisingly large part of what platforms do in a practical, day-to-day sense and in terms of the time, resources, and employees devoted to it. Each social media platform has cobbled together a content moderation labor force consisting of company employees, temporary crowdworkers, outsourced review teams, legal and expert consultants, community managers, flaggers, administrators, moderators, superflaggers, nonprofits, activist organizations, and the entire user population.[6] Social media platforms have built complex apparatuses to manage all of this work. This creates both innovative workflows and problematic labor conditions, nearly all of which remain invisible to users. Given the scope and variety of platform moderation, we should be more skeptical when platforms venture to present themselves as open and frictionless flows of user participation.

Second, moderation shapes how platforms conceive of their users—not just the ones who break the rules or seek the platforms' help. By shifting some of the labor of moderation to users (i.e. through flagging), platforms deputize users as amateur editors and police.[7] From that moment, platform managers must think of, address, and manage users as such. This relationship changes how platforms must conceive of their

[6] *See, e.g.*, Adrian Chen, *The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed*, WIRED: BUSINESS (Oct. 23, 2014), https://www.wired.com/2014/10/content-moderation/ [https://perma.cc/3THX-XSVE]; Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, VERGE (Apr. 13, 2016), http://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech [https://perma.cc/3C35-RCRR]; Sarah Roberts, *Social Media's Silent Filter,* ATLANTIC (Mar. 8, 2017), https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/ [https://perma.cc/XB2U-AW4W].

[7] Kate Crawford & Tarleton Gillespie, *What Is a Flag For? Social Media Reporting Tools and the Vocabulary of Complaint*, 18 NEW MEDIA & SOC'Y 410, (2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2476464 [https://perma.cc/6UBQ-CJNN].

users, not just as customers and producers or as a data commodity, but as an essential labor force.[8]

More fundamentally, content moderation is the central service platforms offer. Anyone can make a website on which any user can post anything he or she pleases without rules or guidelines. Such an anarchical website would, in all likelihood, quickly become a cesspool of hate and porn and subsequently be abandoned. A website like that would not be difficult to build and would require little in the way of skill or financial backing. To produce and sustain an appealing platform requires moderation of some form: Platforms are defined not by what they permit but by what they disallow.[9] Content moderation is an elemental part of what makes social media platforms different and what distinguishes them from the open web. It is hiding inside every promise social media platforms make to their users, from the earliest invitations, to "Broadcast Yourself,"[10] to Mark Zuckerberg's promise to make Facebook "the social infrastructure to give people the power to build a global community that works for all of us."[11]

Content moderation is part of how platforms shape user participation into a deliverable experience. Platforms moderate (through removal, filtering, and suspension); they recommend (through news feeds, trending lists, and personalized suggestions); and they curate (through featured content and front-page offerings). Platforms use these three levers together to actively and dynamically tune the participation of users in order to generate the "right" feed for each user, the "right" social exchanges, and the "right" kind of community. "Right" in these contexts may mean ethical, legal, and healthy, but it also means whatever will promote engagement, increase ad revenue, and facilitate data collection.

Too often, social media platforms discuss content moderation as a problem to be solved—and solved privately and reactively. In this customer service mindset, platform managers understand their responsibility primarily as protecting users from the offense or harm they are experiencing. But now, platforms find they must also answer to users

[8] Lilly Irani, *The Cultural Work of Microwork*, 17 NEW MEDIA & SOC'Y 720, (2015), https://www.researchgate.net/publication/259333001_The_Cultural_Work_of_Microwork [https://perma.cc/N63Y-3SL6].

[9] FINN BRUNTON, SPAM: A SHADOW HISTORY OF THE INTERNET (2012); TOM MALABY, MAKING VIRTUAL WORLDS: LINDEN LAB AND SECOND LIFE (2009).

[10] This represents YouTube's corporate tagline. *See, e.g.*, *Broadcast Yourself*, YOUTUBE (Mar. 18, 2010), https://youtube.googleblog.com/2010/03/broadcast-yourself.html [https://perma.cc/7NGC-D8ME].

[11] Mark Zuckerberg, *Building Global Community,* FACEBOOK (Feb. 16, 2017), https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/ [https://perma.cc/65BY-LCYC].

who find themselves implicated in and troubled by a system that facilitates the reprehensible, even if the users never see it. Even if I never saw, clicked on, or liked a fraudulent news post, I still worry others may have done so. I am troubled by the very fact of it and concerned for the sanctity of the political process as a result. Protecting users is no longer enough. The offense and harm in question is not just to individuals but also to the public itself and to the institutions on which it depends. This, according to John Dewey, is the very nature of a public: "The public consists of all those who are affected by the indirect consequences of transactions to such an extent that it is deemed necessary to have those consequences systematically cared for."[12] What makes something a concern to the public is the potential need for its inhibition.

Social media platforms now inhabit a new position of responsibility, not only to individual users but also to the public they powerfully affect. When an intermediary grows so large and so entwined with the institutions of public discourse, it has an implicit contract with the public that, whether platform management likes it or not, may be quite different from the contract it required users to click through. The impact these platforms have on essential aspects of public life now lies at their doorstep.

~ ~ ~

The early logic of content moderation, and particularly the robust safe harbor protections offered to intermediaries by U.S. law, makes sense in the context of the early ideals of the open web, fueled by naïve optimism, a pervasive faith in technology, and single-minded entrepreneurial zeal. Ironically, these protections were wrapped up in the first wave of public concern over what the web had to offer.

In 1996, the U.S. Congress crafted its first legislative response to online pornography: the Communications Decency Act (CDA). [13] The CDA made it a criminal act, punishable by fines and/or up to two years in prison, to display or distribute "obscene or indecent" material online to anyone under age eighteen.[14] It also imposed similar penalties for harassing or threatening someone online.[15] That law was deemed

[12] JOHN DEWEY, THE PUBLIC AND ITS PROBLEMS 15–16 (1927).

[13] *See* Telecomm. Act of 1996, 47 U.S.C.. *See also* PATRICIA AUFDERHEIDE, COMMUNICATIONS POLICY AND THE PUBLIC INTEREST (1999); Robert Cannon, *The Legislative History of Senator Exon's Communications Decency Act: Regulating Barbarians on the Information Superhighway*, 49 FED. COMM. L.J. 51 (1996).

[14] *See* Telecomm. Act of 1996, 47 U.S.C. § 223.

[15] *See id.*

unconstitutional by the Supreme Court less than a year later.[16] But one element survived. During the legislative process, the U.S. House of Representatives added a bipartisan amendment, largely a response to early lawsuits trying to hold Internet Service Providers (ISPs) and web-hosting services liable for defamation by their users.[17] It carved out a safe harbor for ISPs, search engines, and interactive computer service providers; so long as they only provided access to the Internet or conveyed information, they could not be held liable for the content of that speech.[18]

This safe harbor statute, 47 U.S.C. § 230 (known as "Section 230"), has two parts.[19] The first ensures that intermediaries that merely provide access to the Internet or other network services cannot be held liable for the speech of their users; these intermediaries will not be considered "publishers" of their users' content in the legal sense. The implication is that, like the telephone company, intermediaries do not need to police what their users say and do. The second, less familiar part, adds a twist. If an intermediary *does* police what its users say or do, it does not lose its safe harbor protection by doing so. In other words, choosing to delete some content does not suddenly turn the intermediary into a "publisher" nor does it require the service provider to meet any standard of effective policing. As Milton Mueller writes, Section 230

> was intended both to immunize OSPs who did nothing to restrict or censor their users' communications, and to immunize OSPs who took some effort to discourage or restrict online pornography and other forms of undesirable content. Intermediaries who did nothing were immunized in order to promote freedom of expression and diversity online; intermediaries who were more active in managing user-generated content were immunized in order to enhance their ability to delete or otherwise monitor "bad" content"[20]

This second part of the statute was crafted so that the safe harbor would not create legal jeopardy for intermediaries that chose to moderate

[16] Reno v. Am. Civil Liberties Union, 521 U.S. 844, 849 (1997).

[17] *See* Telecomm. Act of 1996, 47 U.S.C. § 230.

[18] Christopher Zara, *The Most Important Law in Tech Has a Problem*, WIRED (Jan. 3, 2017), https://www.wired.com/2017/01/the-most-important-law-in-tech-has-a-problem/ [https://perma.cc/6NTG-RC9P]; DANIELLE CITRON, HATE CRIMES IN CYBERSPACE 168–75 (2014).

[19] Milton Mueller, *Hyper-Transparency and Social Control: Social Media as Magnets for Regulation*, 39 TELECOMM. POL'Y 804 (2015), http://iranarze.ir/wp-content/uploads/2017/08/E4563-IranArze.pdf [https://perma.cc/L8NP-C79F].

[20] *Id.* at 805.

in good faith by making them any more liable for moderating content than if they had simply turned a blind eye to it.21

These competing impulses—allowing intermediaries to stay out of the way and encouraging them to intervene—continue to shape the way we think about the role and responsibility of all Internet intermediaries, including how we regulate contemporary social media platforms. From a legal standpoint, broad and unconditional safe harbors are profoundly advantageous for Internet intermediaries. As Rebecca Tushnet put it, "[c]urrent law often allows Internet intermediaries to have their free speech and everyone else's too."22 Section 230 also provides ISPs and search engines with the framework upon which they have depended for the past two decades, allowing them to intervene on the terms they choose, while proclaiming their neutrality in order to avoid obligations they prefer not to meet. Offering ISPs and search engines the protection afforded to conduits, like telephone companies, granted them a powerful safe harbor with no matching obligation to serve the public in any specific way. Offering intermediaries indemnity, even if they do intervene, meant that they could pick and choose how and when to do so without being held accountable as "publishers" or for meeting any particular standards for how they do so.

In a phrase common to their terms of service agreements (and many advantageous legal contracts), social media platforms can claim "the right but not the responsibility"23 to remove users and delete content. This is classic legal language, designed to protect a provider from as much liability as possible while ensuring it the most discretionary power.24 But the phrase captures the enviable legal and cultural standing that platforms enjoy.

It is worth noting that Section 230 was not designed with social media platforms in mind, though platforms have managed to enjoy it anyway. Most of the policies that currently apply to social media platforms were intended for a broader category of online services and access providers. At the time Section 230 was crafted, few social media

[21] Citron, *supra* note 18.

[22] Tushnet, *supra* note 5, at 117.

[23] This particular phrase is common to Terms of Service and contractual documents. It appears in spirit in many of the Terms of Service documents for the major U.S. social media platforms. For example, from Instagram: "We may, but have no obligation to, remove, edit, block, and/or monitor Content or accounts containing Content that we determine in our sole discretion violates these Terms of Use." *Terms of Use*, INSTAGRAM https://help.instagram.com/478745558852511 [https://perma.cc/X8ND-JSL2].

[24] Nicolas Suzor, *The Role of the Rule of Law in Virtual Communities*, 25 BERKELEY TECH. L.J. 1844 (2010), https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=1862&context=btlj [https://perma.cc/3LP6-BJDT].

platforms existed. U.S. lawmakers were regulating a web largely populated by ISPs and web "publishers"—amateurs posting personal pages, companies designing stand-alone websites, and online communities having discussions. Besides the ISPs that provided access to the network, the intermediaries at the time were ISPs that doubled as content "portals," like AOL and Prodigy; the earliest search engines, like Altavista and Yahoo; and operators of BBS systems, chatrooms, and newsgroups. Blogging was in its infancy, well before the invention of large-scale blog hosting services like Blogspot and Wordpress. eBay, Craigslist, and Match.com were less than a year old. The ability to comment on a web page had not yet been modularized into a plug-in. The law predates not just Facebook but also MySpace, Friendster, and Livejournal; not just YouTube but also Veoh and Metacafe; not just Soundcloud but also Last.fm and Lala. It even predates Google.

Although they were not included in or anticipated by the law, social media platforms have generally claimed its safe harbor.[25] Section 230, designed to apply to online services and access providers, included a third category awkwardly called "access software providers" to capture these early sites that hosted content provided by users.[26] Such sites are defined as "a provider of software (including client or server software), or enabling tools that do any one or more of the following: (a) filter, screen, allow, or disallow content; (b) pick, choose, analyze, or digest content; or (c) transmit, receive, display, forward, cache, search, subset, organize, reorganize, or translate content."[27] But contemporary social media platforms profoundly exceed that description. While this definition might capture YouTube's ability to host, sort, and queue up user-submitted videos, it is an ill fit for YouTube's ContentID techniques for identifying and monetizing copyrighted material. It may approximate some of Facebook's more basic features; it certainly did not anticipate the intricacy of the NewsFeed algorithm.

Social media platforms are eager to hold on to the safe harbor protections enshrined in Section 230 that shield them from liability for nearly anything that their users might say or do. But all of them also take advantage of the second half of its protection: nearly all platforms impose

[25] Orly Lobel, *The Law of the Platform*, 101 MINN. L. REV. 87 (2016), http://www.minnesotalawreview.org/wp-content/uploads/2016/11/Lobel.pdf [https://perma.cc/X36W-P6TJ].

[26] Section 230 expands the definition of "interactive computer service" to include "access software provider," and there have been cases where CDA was extended to include MySpace and others. 47 U.S.C § 230. *See also CDA 230: Key Legal Cases*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230/legal [https://perma.cc/RAL7-QH2L].

[27] *See* Telecomm. Act of 1996, 47 U.S.C. § 230.

their own rules and police their sites for offending content and behavior themselves. In fact, in most cases their ceaseless and systematic policing cuts much, much deeper than the law requires. In terms of impact on public discourse and the lived experience of users, the rules these platforms impose probably matter more than the legal restrictions under which they function.

~ ~ ~

A slow reconsideration of platform responsibility has been spurred by categories of content particularly abhorrent to users and governments. Public and policy concerns around illicit content, at first largely focused on sexually-explicit and graphically violent images, have expanded to include hate speech, self-harm, and extremism. Platforms have to deal with the enormous problem of user behavior targeting other users, including misogynistic, racist, and homophobic attacks, trolling, harassment, and threats of violence. And these hesitations are growing in every corner of the world.

The United States has largely held to the safe harbor protections first offered to online intermediaries. But growing concerns about terrorism and extremist content, harassment and cyber bullying, and the distribution of nonconsensual pornography (commonly known as "revenge porn") have tested this commitment. Many users, particularly women and racial minorities, are so fed up with the toxic culture of harassment and abuse that they believe platforms should be obligated to intervene.[28] Some critics suggest that "so much deference to the content policies of private technology platforms in fact causes a unique brand of reputational and psychological indignity."[29] A number of platforms have developed specific policies prohibiting revenge porn, modeled on the notice-and-takedown arrangements in copyright law; platforms are not obligated to proactively look for violations but will respond to requests to remove them.[30] This involves the kind of adjudicating platform moderators

[28] Dia Kayyali & Danny O'Brien, *Facing the Challenge of Online Harassment*, ELEC. FRONTIER FOUND. (Jan. 8, 2015), https://www.eff.org/deeplinks/2015/01/facing-challenge-online-harassment [https://perma.cc/5VVW-76LH]; Matias et al., *Reporting, Reviewing, and Responding to Harassment on Twitter*, WOMEN, ACTION & THE MEDIA, https://womenactionmedia.org/twitter-report/twitter-abuse-infographic/ [https://perma.cc/5MYC-5CY5].

[29] Nicolas Suzor et al., *Non-Consensual Porn and the Responsibilities of Online Intermediaries*, 40 MELB. U.L. REV. 1057 (2017).

[30] *See* Ben Medeiros, *Platform (Non-) Intervention and the 'Marketplace' Paradigm for Speech Regulation*, SOCIAL MEDIA + SOC'Y, Jan.–Mar. 2017, http://journals.sagepub.com/doi/pdf/10.1177/2056305117691997 [https://perma.cc/8QJL-X4KT].

typically prefer to avoid: determining whether a complainant (who may not even be a user of that platform) is in fact the subject of the video or photo, whether the material was posted with or without the subject's consent, and who owns the imagery and, thus, has the right to circulate it.

As terrorist organizations increasingly turn to social media to spread fear with shocking images of violence, to coordinate with supporters, and to radicalize the disaffected, Western governments have pressured social media companies to crack down on terrorist organizations. European legislators have slowly imposed something like a notice-and-takedown approach around hate speech and terrorist propaganda and have gradually decreased the required time within which platforms must respond.[31] In early 2016, the Obama administration urged U.S. tech companies to develop new strategies for identifying extremist content, either to remove it or to report it to national security authorities.[32] Also in 2016, European lawmakers persuaded the four largest tech companies to commit to a code of conduct on hate speech, promising to develop more rigorous review and to respond to takedown requests within twenty-four hours. And most recently, the European Commission delivered expanded, non-binding guidelines requiring platform companies to be prepared to remove terrorist and illegal content within one hour of notification.[33]

In the United States, the controversial "Allow States and Victims To Fight Online Sex Trafficking Act" (FOSTA), signed into law in April 2018, penalizes classified ad sites or other platforms if they allow advertising that facilitates sex trafficking operations cloaked as escort services. Strong advocates for Section 230 immunity worry that FOSTA would cause chilling effects on sites that engage in volunteer moderation.[34] Because it does not include a notice-and-takedown mechanism, social media platforms may be forced to proactively look for possible sex trafficking violations.

[31] Giancarlo Frosio, *Reforming Intermediary Liability in the Platform Economy: A European Digital Single Market Strategy,* 112 NW. U.L. REV. 19 (2017).

[32] Eric Geller, *White House and Tech Companies Brainstorm How to Slow ISIS Propaganda*, DAILY DOT (Jan. 8, 2016), https://www.dailydot.com/layer8/white-house-tech-companies-online-extremism-meeting/ [https://perma.cc/QR3X-DZ2U].

[33] Thuy Ong, *The European Commission Wants Facebook, Google To Remove Terrorist Content Within An Hour After Being Flagged*, VERGE (Mar. 1, 2018), https://www.theverge.com/2018/3/1/17066362/european-commission-eu-tech-illegal-terrorist-content-google-youtube-facebook [https://perma.cc/3H7K-LZDX].

[34] *See, e.g.*, Elliot Harmon, *How Congress Censored the Internet*, ELEC. FRONTIER FOUND. (Mar. 21, 2018), https://www.eff.org/deeplinks/2018/03/how-congress-censored-internet [https://perma.cc/L6UX-8G2L].

~ ~ ~

Even in the face of compelling concerns like harassment and terrorism, the logic underlying Section 230 persists. The promise of openness, neutrality, meritocracy, and community was powerful and seductive, resonating deeply with the ideals of network culture and much older dreams of a truly democratic information society.[35] But as social media platforms multiply in form and purpose, become more central to how and where users encounter one another online, and extend themselves into the circulation of goods, money, services, and labor, the safe harbor afforded to Internet providers seems increasingly problematic.

Social media platforms are intermediaries, of course, in the sense that they mediate between users who speak and users who might want to hear them. But this makes them similar not only to search engines and ISPs but also to all forms of traditional media and telecommunications.[36] Media industries of all kinds face some kind of regulatory framework designed to oversee how they mediate between producers and audiences, speakers and listeners, and the individual and the collective.

But social media do violate the century-old distinction deeply embedded in how we think about media and communication. On the one hand, we have trusted interpersonal information conduits like the telephone companies and the post office. Users give them information aimed at others and trust it will be delivered. We expect them not to curate or even monitor that content. In fact, we make it illegal to do so. We expect that our communication will be delivered for a fee, and we understand that the service is the commodity, not the information it conveys. On the other hand, we have media content producers, such as radio, film, magazines, newspapers, television, and video games. The entertainment they deliver feels like the commodity we pay for (sometimes with money, sometimes with our attention to ads) and is designed to speak to us as an audience. We understand that the public obligation of these providers is to produce information and entertainment for public consumption. And we task them in that role with moderating away the kinds of communication harms we worry about most: sexual and

[35] *See generally* FRED TURNER, FROM COUNTERCULTURE TO CYBERCULTURE: STEWART BRAND, THE WHOLE EARTH NETWORK, AND THE RISE OF DIGITAL UTOPIANISM (U. Chi. Press 2006).

[36] Manuel Puppis, *Media Governance: A New Concept for the Analysis of Media Policy and Regulation*, *in* COMMUNICATION, CULTURE, & CRITIQUE 3 134 (2010); Stefan Verhulst, *The Regulation of Digital Content*, *in* THE HANDBOOK OF NEW MEDIA: SOCIAL SHAPING AND CONSEQUENCES OF ICTS 329 (Leah Lievrouw & Sonia Livingstone eds., 2006).

graphic content, violence and cruelty, dangerous kinds of information and knowledge. At times, we debate the values of public media as a way to debate our values as a people.

We are now dealing with a third category: a hybrid between mere information conduits and media content providers, perhaps, or something new emerging from their convergence. Social media platforms promise to connect users person-to-person, entrusted with messages to be delivered to a select audience (sometimes one person, sometimes a friend list, sometimes all users who might want to find it). But as a part of their service, these platforms not only host that content, they organize it, make it searchable, and in some cases even algorithmically select some subset of it to deliver as front-page offerings, news feeds, subscribed channels, or personalized recommendations. In a way, those *choices* are the central commodity platforms sell, meant to draw users in and keep them on the platform, in exchange for advertising and personal data. Users entrust platforms with their interpersonal "tele-" communication, but those contributions then serve as the raw material for the platforms to produce an emotionally engaging flow, more like a "broadcast."

Because of this, they are neither distinctly conduit nor content, nor only network or media, but a hybrid that has not been anticipated by information law or public debates. It is not surprising that users mistakenly expect platforms to be one or the other and are taken aback when they find they are something altogether different.[37] And social media platforms have been complicit in this confusion, as they often present themselves as trusted information conduits and have been oblique about the way they shape our contributions into their commodities. It takes years, or even decades, for a culture to adjust itself to the subtle workings of a new information system and to stop expecting from it what traditional systems provided. This shift, not just in the size and prominence of platforms but in their purpose and practice when it comes to mediating our content, may warrant a full reconsideration of the workings of content moderation.

The promise that platforms are impartial is a powerful one, and it is supported by the way Section 230 offers platforms double protection from accountability. But it is a distraction. Even the early web tools, used only to help design a page or run a blog, shaped how people communicated. Even ISPs that served as mere conduits influenced what we did over them. But the moment that social media platforms introduced profiles, the moment they added comment threads, the moment they added

[37] Tarleton Gillespie, *Facebook's Algorithm—Why Our Assumptions Are Wrong, and Our Concerns Are Right*, CULTURE DIGITALLY (July 4, 2014), http://culturedigitally.org/2014/07/facebooks-algorithm-why-our-assumptions-are-wrong-and-our-concerns-are-right/ [https://perma.cc/3XGF-RS5F].

ways to tag or sort or search or categorize what users posted, the moment they indicated what was trending or popular or featured, the moment they did anything other than list users' contributions in reverse chronological order, they moved from delivering content for the person posting it to constituting it for the person accessing it.

As Frank Pasquale has noted, "policymakers could refuse to allow intermediaries to have it both ways, forcing them to assume the rights and responsibilities of content or conduit. Such a development would be fairer than current trends, which allow many intermediaries to enjoy the rights of each and responsibilities of neither."[38] But if social media platforms are neither conduit nor content, then legal arrangements premised on those categories may be insufficient.

One possibility is to recommit to Section 230, even double down on it, but with a sober and unflinching eye for which platforms, or aspects of platforms, warrant it and which exceed it. If a platform offers to connect you to friends or followers and deliver what they say to you and what you say to them, then it is a conduit. This would enjoy Section 230 safe harbor, including the good faith moderation that safe harbor anticipated. But the moment a platform begins to select some content over others, based not on a judgment of relevance to a search query but in the spirit of enhancing the value of the experience and keeping users on the site, it becomes a hybrid. As soon as Facebook changed from delivering a reverse chronological list of materials that users posted on their walls to curating an algorithmically selected subset of those posts in order to generate a News Feed, it moved from delivering information to producing a media commodity out of it. If this is a profitable move for Facebook, its administrators can do so, but it would make them more liable for the content they assemble, even though it is entirely composed out of the content from users.[39] This new category of hybrids would certainly

[38] Frank Pasquale, *Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power*, 17 THEORETICAL INQUIRIES IN L. 487 (2016).

[39] An intriguing exception to how Section 230 has generally been applied in U.S. courts is the decision in *Roommates.com.* Roommates.com was charged with facilitating discrimination on the basis of race and sexual preference, in part because the roommate preference survey required that these questions be answered and used that information to algorithmically show some profiles to some users. Roommates.com argued that Section 230 offered it protection from any liability for what its users did. But the court ruled that because the platform had asked for these personal details and had algorithmically selected what to show to whom, it was so involved in the creation and distribution of that information that it could not enjoy the immunity offered by Section 230. This case offers an intriguing alternative for how to think about the responsibilities of platforms now that they algorithmically sort and display user contributions. *See* Fair Housing Council of San Fernando Valley, et al. v. Roommates.com LLC, 489 F.3d 921 (9th Cir. 2007), *aff'd en banc*, 2008 WL 879293 (9th Cir. 2008).

include the marketplace services that present themselves as social media platforms, like Airbnb, Etsy, and Uber. Even though, as part of their services, they host and distribute users' speech (profiles, comments, reviews, and so on), they are also new kinds of employers and brokers, and they should not get to use Section 230's protection to avoid laws prohibiting discrimination in employment, housing, or pricing.[40]

~ ~ ~

What I just proposed is heresy to some. There are many who, even now, strongly defend Section 230. The "permissionless innovation"[41] it provides arguably made the development of the Internet and contemporary Silicon Valley possible, and some see it as essential for that to continue.[42] As David Post remarked, "No other sentence in the U.S. Code . . . has been responsible for the creation of more value than that one."[43] But among defenders of Section 230, there is a tendency to paint even the smallest reconsideration as if it would lead to the demise of the Internet, the end of digital culture, and the collapse of the sharing economy. Without Section 230 in place, some say, the risk of liability will drive platforms either to remove everything that seems the slightest bit risky or to turn a blind eye.[44] Entrepreneurs will shy away from investing in new platform services because the legal risk would appear too costly.

I am sympathetic to the desire to defend some of the safe harbor offered by Section 230. But the typical defense of Section 230 in the face of these compelling concerns tends to adopt an all-or-nothing rhetoric. Some claim that any conditional liability opens the door to proactive

[40] Christopher Zara, *The Most Important Law in Tech Has a Problem*, WIRED (Jan. 3, 2017), https://www.wired.com/2017/01/the-most-important-law-in-tech-has-a-problem/ [https://perma.cc/K29X-GGKT].

[41] ADAM THIERER, PERMISSIONLESS INNOVATION: THE CONTINUING CASE FOR COMPREHENSIVE TECHNOLOGICAL FREEDOM (Mercatus Center at George Mason Univ. 2014).

[42] *Manila Principles on Intermediary Liability*, ELEC. FRONTIER FOUND. (Mar. 24, 2015), https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf [https://perma.cc/Q4S2-VFGU]; Adam Thierer, *Celebrating 20 Years of Internet Free Speech & Free Exchange*, MEDIUM (June 21, 2017), https://readplaintext.com/celebrating-20-years-of-internet-free-speech-free-exchange-8a10f236d0bd [https://perma.cc/G4DP-QSNW].

[43] David Post, *A Bit of Internet History, or How Two Members of Congress Helped Create a Trillion or so Dollars of Value*, WASH. POST (Aug. 27, 2015), https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/08/27/a-bit-of-internet-history-or-how-two-members-of-congress-helped-create-a-trillion-or-so-dollars-of-value/ [https://perma.cc/9QY8-JVTG].

[44] Giancarlo Frosio, *Why Keep a Dog and Bark Yourself? From Intermediary Liability to Responsibility*, 25 OXFORD INT'L J.L. & INFO. TECH, 1 (2017).

moderation and could prove a slippery slope to full liability; some imply that platforms do not moderate already. This rigorous defense of "expressive immunity," as Julie Cohen calls it, requires a "carefully tended hysteria about censorship and injured protestations of First Amendment virtue."[45] And, it is worth noting, this perspective lines up well with the way platforms themselves defend Section 230's protections. Realistically, there is a great deal of room between complete legal immunity offered by a robust Section 230 without exceptions and total liability for platforms as Section 230 crumbles away.[46]

However as Section 230 grows to meet today's challenges, we must redress the opportunity that was missed when Section 230 was first drafted. Safe harbor, including the right to moderate in good faith and the permission not to moderate at all, was an enormous gift to the young Internet industry. Over the history of U.S. regulation of the media and telecommunication industries, gifts of this enormity were always fitted with a matching obligation to serve the public: a monopoly granted to a telephone company comes with the obligation to serve all users; a broadcasting license comes with obligations about providing news or weather alerts or educational programming.

The gift of safe harbor could finally be paired with public obligations—not external standards for what to remove, but parameters for how moderation should be conducted. Such matching obligations might include:

- **Transparency Obligations** – Platforms could be required to report data on the process of moderation to the public or to a regulatory agency. Several of the major platforms already voluntarily report takedown requests, but these have typically focused on government requests. Until recently none systematically reported data on flagging, policy changes, or removals made on their own accord. Facebook and YouTube began to do so in 2018 and should be encouraged to continue.[47]

- **Minimum Standards for Moderation** – Without requiring that moderation be handled in a particular way, minimum standards for the worst content, minimum response times, or obligatory mechanisms for

[45] Julie Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133 (2017).

[46] In this, I am in agreement with Danielle Citron and Ben Wittes in their contributions to this special issue. *See* Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage: Revising Section 230 Immunity*, 2 GEO. L. TECH. REV. 453 (2018).

[47] Tarleton Gillespie, *Facebook and YouTube Just Got More Transparent. What Do We See?*, NEIMAN LAB (May 3, 2018), http://www.niemanlab.org/2018/05/facebook-and-youtube-just-got-more-transparent-what-do-we-see/ [https://perma.cc/3J98-8AW7].

redress or appeal could help establish a base level of responsibility and parity across platforms.

- **Shared Best Practices** – A regulatory agency could provide a means for platforms to share best practices in content moderation without raising antitrust concerns. Outside experts could be enlisted to develop best practices in consultation with industry representatives.

- **Public Ombudsman** – Most major platforms address the public only through their corporate blogs, when announcing major changes in policy or responding to public controversies. But this is on their own initiative and offers little room for public response. Platforms could each be required to have a public ombudsman who responds to public concerns and translates those concerns to policy managers internally, or a single "social media council"[48] could field public complaints and demand accountability from platforms.

- **Financial Contributions to Support Organizations and Digital Literacy Programs** – Major platforms like Twitter have leaned on non-profit organizations to advise and even handle some moderation, as well as to mitigate the socio-emotional costs of the harms some users encounter.[49] Digital literacy programs could expand to better address online harassment, hate speech, and misinformation. Enjoying safe harbor protections of Section 230 might entail platforms helping to fund these non-profit efforts.

- **An Expert Advisory Panel** – Without assuming regulatory oversight of a government body, a blue-ribbon panel of regulators, experts, academics, and activists could be given access to platforms and their data to oversee content moderation without revealing platforms' inner workings to the public.

- **Advisory Oversight from Regulators** – A government regulatory agency could consult on and review the content moderation procedures at major platforms. By focusing on reviewing procedures, such oversight could avoid the appearance of imposing a political

[48] *Regulating Social Media: We Need a New Model That Protects Free Expression*, ARTICLE19 (Apr. 25, 2018) https://www.article19.org/resources/regulating-social-media-need-new-model-protects-free-expression/ [https://perma.cc/7ABD-8428].

[49] Matias et al., *Reporting, Reviewing, and Responding to Harassment on Twitter*, WOMEN, ACTION & THE MEDIA, https://womenactionmedia.org/twitter-report/twitter-abuse-infographic/ [https://perma.cc/5MYC-5CY5].

viewpoint. Such review could instead be sensitized to the more systemic problems of content moderation.

- **Labor Protections for Moderators** – Content moderation at large platforms depends on crowdworkers, either internal to the company or contracted through third party temporary services. Guidelines could ensure these workers have basic labor protections like health insurance, assurances against employer exploitation, and greater care for the potential psychological harm that can be involved.

- **Obligation to Share Moderation Data with Qualified Researchers** – The right to safe harbor could come with an obligation to set up reasonable mechanisms for qualified academics to access platform moderation data, in order to allow for the investigation of questions that platforms might not think to, or want to, answer. The research partnership between Facebook and the Social Science Research Council[50] announced in 2018 has yet to work out the details, but some version of this model could be extended to all platforms.

- **Data Portability** – Social media platforms have been resistant to making users' profiles and preferences interoperable across platforms. But moderation data, like blocked users and flagged content, could be made portable so that users are not required to duplicate their efforts across the platforms they frequent.

- **Build Systems for Regular Audits** – Without requiring complete transparency in the moderation process, platforms could build in mechanisms for researchers, journalists, and even users to conduct their own audits of the moderation process in order to better understand how the rules play out in practice.

- **Regular Legislative Reconsideration of Section 230** – The Digital Millennium Copyright Act[51] stipulated that the Library of Congress revisit the list of exceptions every three years to account for changing technologies and emergent needs. Section 230, and whatever matching obligations that might be paired with it, could similarly come up for reexamination to account for the rapidly changing workings of social

[50] *Statement from SSRC President Alondra Nelson on the Social Data Initiative*, SOCIAL SCIENCE RESEARCH COUNCIL (SSRC), https://www.ssrc.org/programs/view/social-data-initiative/ [https://perma.cc/WZ54-YJUL].

[51] 17 U.S.C. § 512.

media platforms and the even more rapidly changing nature of harassment, hate, misinformation, and other harms.

~ ~ ~

We desperately need a thorough public discussion about the social responsibility of platforms. This conversation has begun, but too often it gets hamstrung between the defenders of Section 230 and those concerned by the harms it may protect. And until intermediary liability law is rethought, social media platforms will continue to enjoy the two sides of safe harbor: the right, but not the responsibility, to police their sites as they see fit. Adjustments can be made to Section 230 to balance some light, shared public obligations to go with the generous immunity it has offered to platforms. But there is a real risk that incremental improvements, as welcome as they would be, might in fact hold the existing logic of content moderation in place. What the law may need is a new way of thinking about platforms and their responsibility that recognizes that they constantly tune public discourse through their moderation, recommendation, and curation.[52]

Our conceptions of what these information providers "do" to the information they circulate is trapped inside of metaphors of passage: who gets through the gate and who does not. Such metaphors are wildly insufficient if we hope to attend to the complex ways in which platforms intervene in the flow of information and give shape to public discourse. New policy and new scholarship will need new ways of thinking about platforms that (a) attend to the complex feedback loops generated by the interaction between the ongoing contributions of users and the ongoing interventions of platforms; (b) offer frameworks of obligation based on the way platforms tune people's contributions through policy and design; and (c) extend responsibility to second order consequences from the proper working of these systems, not just their misuse.

[52] *Id.*; Mariarosaria Taddeo & Luciano Floridi, *New Civic Responsibilities for Online Service Providers*, *in* THE RESPONSIBILITIES OF ONLINE SERVICE PROVIDERS, 1 (Mariarosaria Taddeo & Luciano Floridi eds., 2017).