6

Making Social Media Pay for Its Sins

Repealing or Amending Section 230

Section 230 of the Communications Decency Act[1] has been described as "the twenty-six words that created the internet."[2] Though initially widely supported, in recent years Section 230 has become a lightning rod for attacks on social media platforms and concerns about the social impact of the internet. Section 230, as we shall discuss in more detail, provides platforms with immunity from legal liability both for third-party content that they host *and* for the platforms' good faith efforts to moderate harmful content. In essence, critics argue that Section 230 has operated as a "get out of jail free" card for social media platform operators, permitting them to ignore harmful content when it suits them, but at the same time to block content that they do not like, both free of legal restraints.

As a result of this continuing criticism, there have been innumerable calls for Section 230 reform coming from across the political spectrum. Remarkably, during the 2020 presidential campaign *both* President Joe Biden and President Donald Trump called for Section 230 to be "revoked" outright. Biden, for example, said in an interview with the *New York Times* in December 2019 that "Section 230 should be revoked, immediately should be revoked."[3] The reason he gave was Facebook's failure to block harmful speech (with a particular focus on falsehoods), though around the same time House Speaker Nancy Pelosi cited harassment and abuse as another reason to eliminate Section 230 immunity.[4]

[1] 47 U.S.C. § 230.

[2] Jeff Kosseff, The Twenty-Six Words that Created the Internet (2019).

[3] Editorial Board, *Joe Biden*, N.Y. Times (Jan. 17, 2020), www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html.

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

In May 2020 President Trump joined the bandwagon, tweeting "REVOKE 230!" in response to a dispute with Twitter/X,[5] a position he reiterated in December 2020 in the course of vetoing a major defense appropriation bill.[6] Importantly, Trump's calls to revoke Section 230 were triggered not by awful content but rather by social media firms' alleged anti-conservative bias; but the two presidents' agreement on the needed *remedy* for the different problems they identify with social media is noteworthy. Nor are Trump and Biden alone in calling for repeal of Section 230.[7]

6.1 SECTION 230: WHAT IT DOES AND WHY IT DOES IT

Section 230 was enacted by Congress in 1996, in the early days of the internet long before the rise of social media platforms and ubiquitous user-generated content. It was a part of the Communications Decency Act, which in turn was part of the Telecommunications Act of 1996, which is why in popular parlance the provision is often described as Section 230 of the Communications Decency Act, or CDA.[8] Section 230 was a direct congressional response to two early defamation cases brought against internet service providers who hosted bulletin boards and discussion forums containing third-party, user-generated content.[9] Read in combination, the two decisions appeared to establish the principle that internet platforms hosting user- or third party-generated content could be held liable as publishers of defamatory content if, but only if, they made efforts to control and suppress harmful or offensive content on their platforms. If the platform owners did *not* exercise such control, they could be held liable only as distributors, meaning that they were liable only for defamatory content of which they were aware, or should have been aware, but nonetheless took no action to stop distributing.

It was immediately apparent that the legal regime created by these decisions would have been an utter disaster for the development of the internet, and especially for those corners of the internet that specialized in hosting third-party content – including what eventually become social media platforms.

[5] *ibid.*

[6] *Presidential Veto Message to the House of Representatives for H.R. 6395*, White House (Dec. 23, 2020), www.whitehouse.gov/briefings-statements/presidential-veto-message-house-representatives-h-r-6395/.

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

[8] www.govinfo.gov/link/uscode/47/230.

[9] The cases were Cubby, Inc. v. CompuServe, Inc., 776 F. Supp. 135 (S.D.N.Y. 1991), and Stratton Oakmont, Inc. v. Prodigy Service Co., 1995 WL 323710 (N.Y. Sup. Ct. 1995).

For one, the cases in combination strongly disincentivized any form of content moderation, for fear of opening the door to publisher liability. But it was apparent to observers – and to Congress – that some forms of content moderation are unquestionably desirable. At the same time, the possibility of distributor liability incentivized providers to pull down information at the first hint that it might be defamatory or otherwise illegal, inevitably resulting in the suppression of large amounts of marginal but legal speech. Furthermore, such platform actions opened the door to publisher liability, putting platforms between a rock and a hard place. The internal contradiction here – and the resulting chaos – should have been and were apparent to all.

Section 230 was Congress's response. The first, key operative provision of that law, Section 230(c)(1), states that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[10] The clear intent and effect of this language is to flatly preclude imposing publisher liability on platforms for user-generated or other third-party content that they host. As such, this provision ensures (contrary to the early *Stratton Oakmont* decision described earlier) that engaging in some content moderation does not open platforms to unlimited liability. This protection was in itself essential to permit providers of hosting services (including, eventually, social media platforms) to provide controlled and curated experiences, as opposed to the anything-goes chaos of the public square – though, of course, if a particular platform *wants* to provide fully open, unmoderated spaces they are welcome to do so, as the Telegram and increasingly Twitter/X platforms do.

Important as this immunity was, courts quickly expanded the immunity provided by Section 230(c)(1) well beyond the obvious scope of its language. The key case in this regard was the very important 1997 decision in *Zeran v. American Online*,[11] issued by the United States Court of Appeals for the Fourth Circuit (the regional federal court of appeals covering the states of Maryland, Virginia, West Virginia, North Carolina, and South Carolina). *Zeran* held that Section 230(c)(1) shielded platforms not only from publisher liability but also from *distributor* liability – meaning that even if a platform *knew* or had reason to know that it was hosting illegal or harmful content, it still could not be held legally liable for it. This holding, which has been widely followed by other courts, eliminated any *legal* obligation or incentive (business incentives, as we have seen, are another thing) for platforms to moderate harmful or illegal third-party content.

[10] 47 U.S.C. § 230(c)(1).

[11] 129 F.3d 327 (4th Cir. 1997).

In recent years, the *Zeran* decision and its progeny have come under sharp attack.[12] In particular, in an important separate opinion US Supreme Court Justice Clarence Thomas expressed serious doubts about *Zeran* and urged the Supreme Court (which has never ruled on the issue) to reconsider the scope of Section 230(c)(1) and whether it should be limited to publisher immunity.[13] The Court has not yet taken up Justice Thomas's invitation, but with the growing unpopularity of social media platforms, especially on the political right, it would be unsurprising if it does so soon. It should be noted, however, that at the time *Zeran* was decided its holding was broadly supported. The judge who authored the decision was then-Chief Judge J. Harvey Wilkinson, one of the most respected judges in the federal judiciary (and, it should be noted, an appointee of President Ronald Reagan). And, as already noted, other courts quickly adopted Chief Judge Wilkinson's reasoning. It is only when the implications of the *Zeran* rule for *social media* platforms (which did not exist in their modern form in 1997) became clear that the decision became controversial.

The impact and importance of Section 230 does not end, however, with providing immunity from publisher or distributor liability. Its second key provision, subsection (c)(2)(A), also grants platforms immunity for their affirmative actions in moderating harmful user-generated and other third-party content. It reads as follows:

> No provider or user of an interactive computer service shall be held liable on account of any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.[14]

Several things should jump out from this language. First, unlike the immunity provided by subsection (c)(1), this immunity is not absolute, it is limited to *good faith* content moderation. But given the extreme difficulty of proving lack of good faith, this limitation has not turned out to be terribly important in practice.

Second, as the last phrase of this provision suggests, Section 230(c)(2)(A) immunity is *not* limited to platform actions imposing restrictions on illegal content or content beyond the scope of the First Amendment. It is true that

[12] For a careful discussion of the difficulties raised by the *Zeran* decision, *see* Alan Z. Rozenshtein, *Interpreting the Ambiguities of Section 230*, 41 Yale J. Reg. Bull. 60, 68–71 (2024).

[13] Malwarebytes, Inc. v. Enigma Software Group USA, LLC, 141 S. Ct. 13, 15–16 (2020) (Statement of Justice Thomas respecting the denial of certiorari).

[14] 47 U.S.C. § 230(c)(2)(A).

legally "obscene" content does fall outside First Amendment protection,[15] as perhaps does "harassing" content; but the other forms of content listed, including notably "excessively violent" content, quite clearly are protected by the First Amendment under well-established Supreme Court precedent.[16]

The final important question about the scope of subsection (c)(2)(A) immunity concerns the meaning of the final category of content listed, that which is "otherwise objectionable." Many courts have interpreted this language in an open-ended way, granting platforms immunity for decisions to block or restrict *any* content that a platform, in good faith, finds "objectionable" for any reason.[17] There is an argument to be made, however (and indeed has been made by Professors Adam Candeub and Eugene Volokh), that the phrase "otherwise objectionable" should be interpreted in light of the words preceding it, and so limited to content that is "objectionable" for similar reasons to the other listed categories – i.e., objectionable because it was highly sexual, vulgar, or violent.[18] If such a reading was adopted by, say, the Supreme Court (the question remains unresolved as of this writing), that would severely curtail the power of social media platforms to moderate content because it is false (for example, mis- or disinformation about, say, vaccines), hateful (e.g., racist, sexist, homophobic, and other hate speech), or ideologically troubling (e.g., pro-ISIS/Islamic State or pro-KKK propaganda that is not itself "excessively violent").

One further point about the relationship between the two subsections of Section 230 discussed earlier. On its face, the language of the two provisions would appear to protect distinct things. Subsection (c)(1) immunizes internet service providers regarding third-party content that remains on their platforms, thereby eliminating (on the *Zeran* reading) any obligation on the part of platforms to remove objectionable material. Subsection (c)(2)(A), on the other hand, immunizes platforms when they *do* choose, in good faith, to remove objectionable content (leaving aside for now the meaning of the word "objectionable"). But in fact, some courts have read the first subsection more broadly, to immunize *any* platform decision made in their role as a "publisher," including "reviewing, editing, and deciding whether to publish *or to withdraw from publication* third-party content."[19]

[15] Roth v. United States, 354 U.S. 476 (1957).

[16] *See* Brown v. Entertainment Merchants Association, 564 U.S. 786 (2011) (violent speech); Cohen v. California, 403 U.S. 15 (1971) (curse words); United States v. Playboy Entertainment Group, 529 U.S. 803 (2000) (non-obscene pornography).

[17] Adam Candeub and Eugene Volokh, *Interpreting 47 U.S.C. § 230(c)(2)*, 1 J. FREE SPEECH L. 175, 177 n.4 (2021).

[18] *Ibid*. at 179–83.

[19] Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1102 (9th Cir. 2009) (*citing* Fair Housing Council of San Fernando Valley v. Roommates.Com, LLC, 521 F.3d 1157, 1170–71 (9th Cir. 2008)) (emphasis added).

This reading creates serious overlap between the two separate provisions of Section 230, seemingly making subsection (c)(2)(A) redundant and eliminating that provision's limitation of immunity to content moderation decisions made in "good faith." Courts have justified this by arguing that (2)(A) is not redundant because it also immunizes a decision to moderate non-third-party (i.e., at least in part platform-generated) content.[20] But this expansion of subsection (c)(1) immunity is in deep tension with the statutory language, and indeed in the separate opinion mentioned earlier, Justice Clarence Thomas also raised and criticized the judicial interpretation of subsection (c)(1) to reach decisions to remove content.[21]

The reason why all of this matters is simple. It demonstrates that there are several ongoing disputes about the precise scope of Section 230 immunity for platforms. Because the Supreme Court has never spoken to *any* of these issues to date, it remains possible that moving forward Section 230 immunity will become less expansive than what courts have granted in the past. But regardless of how these various disagreements are ultimately resolved, few people seriously doubt that the existence of *some* statutory immunity has been essential for platforms to develop and expand as they have since the turn of the Twenty-First century.[22] And so, any proposals to repeal or reform Section 230 must be evaluated in light of that broad consensus.

6.2 THE WAR ON SECTION 230

As noted at the beginning of this chapter, Section 230 is currently not a very popular law. To the contrary, it is fair to say that Section 230 is loathed across the political spectrum, with few or any politicians or journalists coming to its defense (academics are another matter[23]). And the basic shape of the attack on Section 230 is also familiar and predictable. It is, as noted earlier, that Section

[20] *Ibid.* at 1105.

[21] Malwarebytes, Inc. v. Enigma Software Group USA, LLC, 141 S. Ct. 13, 16–17 (2020) (Statement of Justice Thomas respecting the denial of certiorari).

[22] The primary argument *against* the need for some form of Section 230 immunity, aside from Luddites who wish to simply extinguish platforms for third-party content as a technology, rests on the idea that even absent statutory protection, the First Amendment to the Constitution would give platforms the space they need to operate. Professor Eric Goldman has strongly and articulately refuted that view, however. Eric Goldman, *Why Section 230 Is Better than the First Amendment*, 95 Notre Dame L. Rev. Reflection 33 (2019).

[23] While there are many members of the academy who defend Section 230 in its current form, including this author, probably the most vociferous defender is Professor Eric Goldman of Santa Clara University. One can get a good sense of Professor Goldman's views on his Technology and Marketing Law Blog: https://blog.ericgoldman.org/.

230 operates as a "Get Out of Jail Free" card, permitting social media (and other) platforms to do whatever they want, safe in the knowledge that the two parts of Section 230 permit them to leave harmful speech on their platforms and to "censor" whatever content they oppose, in both cases with confidence that they will be free of liability. As a result, platforms can cause great harm both to individuals and to society as a whole in the pursuit of unprecedented profits, without taking any responsibility for their actions.

Despite this seeming unanimity about the ills of Section 230, there exists a basic conundrum: While critics of Section 230 across the political spectrum agree that it enables misbehavior by platforms, they disagree fundamentally regarding the *kinds* of allegedly bad conduct Section 230 enables – which should come as no surprise given the discussion in Chapters 1 and 2 about the nature of conservative and progressive critiques of social media more generally. As a result, the two sides take polar opposite positions regarding *how* Section 230 should be reformed. This is not true, admittedly, about the most extreme calls, such as those from Presidents Biden and Trump described earlier, for the complete repeal of Section 230 – a call Senator Linsey Graham of South Carolina echoed in early 2024, when grandstanding about online child safety issues.[24] But it is not clear what repeal advocates envision taking the place of Section 230, given the unworkability of the legal regime that preceded it. And one suspects that conservatives and progressives would not agree on that replacement.

In any event, few people take calls to totally repeal Section 230 seriously because such a complete repeal would, realistically, make it impossible for platforms to host any user or other third-party content. This is because given the sheer scale at which social media platforms operate, perfect policing of defamatory and other illegal content is effectively impossible (especially defamation of private persons, which is exceedingly hard to factcheck). Repealing Section 230, in short, would completely eliminate the role of platforms as *social* media, and instead put them in the business of traditional media, which is to say serving up content created or carefully selected and vetted by the platform owners.

The destruction of social media may well be the unstated goal of some politicians and journalists, for whom social media has been deeply disempowering – most significantly, as discussed in Chapter 5, by eliminating the gatekeeper power that traditional media enjoyed and that politicians sometimes exploited. Furthermore, in the case of journalists the growth of social media

[24] Editorial Board, *Congress's Social-Media Spectacle*, Wall Street Journal (Feb. 1, 2024), www.wsj.com/articles/big-tech-hearing-congress-meta-social-media-mark-zuckerberg-1aeb2044.

platforms has been financially ruinous. But there are billions of people worldwide who use social media, presumably because they find engaging with their friends and acquaintances as well as the broader public on platforms a net positive experience. For them, the elimination of social media would hardly be a positive outcome or one they would support.

Indeed, moving beyond purely individual desires, eliminating social media as a technology would be a net negative for society and – yes – democracy. Such a move would reverse the profoundly democratizing impact of the internet, which has for first the first time in human history permitted millions of ordinary individuals to reach and engage with large audiences and communities. Perhaps from the point of view of some elites (notably media elites) such a reversal would not be such a bad thing; but from the perspective of most people and of society as a whole, eliminating platforms for user-generated content because they cause some social harm would truly be throwing out the baby with the bathwater.

That then leaves more modest reform proposals. But here, we again face the conundrum of fundamental inconsistency between conservatives and progressives. Conservative critics are focused mainly on weakening the provisions of Section 230 that shield platform moderation of third-party content from liability, on the grounds that they enable platform owners to effectuate their alleged anti-conservative bias. Conservative reform proposals thus consistently seek to restrict the scope of platform authority to moderate content. This would certainly be the result of Justice Thomas's argument, described earlier, that subsection (c)(1) of Section 230 should *not* have been read to immunize platform decisions regarding content moderation. If such an interpretation prevailed, this would mean that platforms could only evoke subsection (c)(2)(A) when defending decisions to remove or deemphasize content, which in turn would permit an attack based on claims of bad faith on the part of platforms – which, arguably, political discrimination would qualify as.

Similarly, the proposal – advanced among others by Professors Candeub and Volokh – to read the "otherwise objectionable" language of subsection (c)(2)(A) narrowly would also severely limit platform immunity for content moderation decisions (especially in combination with Justice Thomas's narrow reading). In fact, as noted earlier, such as reading would probably completely eliminate platform power to suppress speech, including hate speech and some terrorist propaganda, based on the viewpoint expressed in such content (note the parallel to Texas's HB 20, discussed in Chapters 1 and 4).

But conservative attempts to limit Section 230 immunity for content moderation decisions go beyond interpretive arguments, to legislative proposals. Most notably, in September 2020 the Trump Administration Justice

Department sent legislation to Congress proposing substantial revisions to Section 230.[25] Among other things (some discussed later), this proposal would have completely eliminated platforms immunity under subsection (c)(2)(A) for decisions to remove content that it believes to be "otherwise objectionable," thereby mooting the interpretive debate. Under such a regime, in practice platform owners' discretion to remove harmful content would be limited to sexual, violent, or unlawful materials.

Other examples of conservative efforts to amend Section 230 in response to allegations of political bias include proposals by Republican Senators Josh Hawley of Missouri and Marco Rubio of Florida that would condition Section 230 immunity for social media platforms on their making politically neutral and/or viewpoint-neutral content moderation decisions.[26] Another, narrower example of such an legislative initiative, also proposed by Senator Hawley, would condition Section 230 immunity on tech platforms passing an independent audit which confirmed that the platforms were not politically biased.[27] All these initiatives, like the Texas HB 20 and Florida S.B. 7072 laws discussed in Chapter 1, are at bottom efforts to attack platform content moderation power head-on (the only difference being that as states, Texas and Florida have no power to amend a federal statute such as Section 230).

In contrast to conservative proposals, unsurprisingly, progressive reform efforts take exactly the opposite tack, seeking to incentivize *more* rather than less content moderation by platforms by limiting immunity under subsection (c)(1) of Section 230. For example, Senator Amy Klobuchar, a Democrat from Minnesota (and former presidential candidate) proposed legislation in July of 2021 that would have created an exception to Section 230 for platforms whose algorithms promoted health misinformation (July of 2021 was of course in the middle of the COVID-19 pandemic).[28] Senator Elizabeth Warren of Massachusetts, another prominent Democrat (and former presidential

[25] *The Justice Department Unveils Proposed Section 230 Legislation*, U.S. Dep't of Just. (Sept. 23, 2020), www.justice.gov/opa/pr/justice-department-unveils-proposed-section-230-legislation.

[26] *See Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies*, Josh Hawley, U.S. Senator for Missouri (June 19, 2019), https://perma.cc/HFM2-93VA; *see also* Jane Coaston, *A Republican Senator Wants the Government to Police Twitter for Political Bias*, Vox (June 26, 2019, 3:30 PM), www.vox.com/2019/6/26/18691528/section-230-josh-hawley-conservatism-twitter-facebook.; *Rubio Introduces Sec 230 Legislation to Crack Down on Big Tech Algorithms and Protect Free Speech*, Marco Rubio, U.S. Senator for Florida (June 24, 2021), https://perma.cc/43R6-HRWV.

[27] Allyn, *supra* n. 4.

[28] S.2448 – Health Misinformation Act of 2021, www.congress.gov/bill/117th-congress/senate-bill/2448/text; Shannon Bond, *Democrats Want to Hold Social Media Companies Responsible for Health Misinformation*, NPR (July 22, 2021), www.npr.org/2021/07/22/1019346177/democrats-want-to-hold-social-media-companies-responsible-for-health-misinformat.

candidate) has similarly criticized social media platforms, including especially Facebook, sharply for failing to block misinformation in political advertising – in one famous instance by taking out an ad on Facebook containing false information about Facebook and its CEO Mark Zuckerberg.[29]

Nor, of course, is left-wing criticism limited to politicians. Left-leaning journalistic outlets, including well-respected ones such as the New York Times and the Washington Post, have repeatedly published stories and editorials criticizing social media platforms' failure to block mis- and disinformation, as well as their spreading of polarizing content. And several of those editorials have strongly suggested that Section 230 reform was the only way to incentivize better behavior[30] – though as with politicians, the journalists are typically notably short on details on what should take Section 230's place.

Needless to say, not all Section 230 reform proposals are as relentlessly aggressive (or politically motivated) as this. Law professors Danielle Keats Citron and Mary Anne Franks have proposed a set of more thoughtful, and more limited, reforms of Section 230. One would limit Section 230 immunity to *speech*, thereby clarifying that online commercial transactions and the like would not fall within the provision.[31] They would also deny immunity to truly bad actors, meaning websites that knowingly keep up illegal content, encourage illegality, principally host illegal content, or solicit illegal content.[32] Such a revision would presumably have little impact on major platforms such as Facebook and YouTube, but would permit action against the seediest parts of the internet.

Finally, Citron and Franks propose language that would condition Section 230 immunity on platforms taking "reasonable steps to address unlawful uses of its service that create serious harm to others."[33] Such a provision would, of course, require courts to determine what constitutes "reasonable steps" in a world in which, all acknowledge, content moderation will necessarily be imperfect. Citron and Franks argue, however, that courts have proven capable

[29] Cecelia Kang and Thomas Kaplan, *Warren Dares Facebook with Intentionally False Political Ad*, Washington Post (Oct. 12, 2019), www.nytimes.com/2019/10/12/technology/elizabeth-warren-facebook-ad.html.

[30] *See*, e.g., Jennifer Rubin, *It's Time to Stand Up to Facebook*, Washington Post (Oct. 4, 2021), www.washingtonpost.com/opinions/2021/10/04/its-time-stand-up-facebook/; Joe Scarborough, *Zuckerberg Says He's "Disgusted" by Trump's Rhetoric. It's Just Crocodile Tears*, Washington Post (June 18, 2020), www.washingtonpost.com/opinions/why-are-facebook-and-its-founder-not-held-responsible-for-the-damage-they-deal/2020/06/18/85c4017e-b0cb-11ea-856d-5054296735e5_story.html.

[31] Danielle Keats Citron and Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 45 U. Chi. Legal F. 45, 69–74 (2020).

[32] *Ibid.* at 70–71.

[33] *Ibid.* at 71.

of making such judgments in the past, and that over time best practices will emerge.[34]

Interestingly, the Trump Administration's 2020 proposal to reform Section 230, discussed earlier, picked up on some of Citron and Frank's proposals. In particular, the Trump Administration would have amended Section 230 to deny immunity to "Bad Samaritans," meaning platforms that knowingly facilitate criminal behavior, or knowingly failed to remove material that violated criminal law.[35] The proposed amendments would also have eliminated Section 230 immunity for actions brought under a wide swath of laws, including laws regulating terrorism, child sex abuse, cyber-stalking, as well as antitrust laws. Some of these proposals, notably the Bad Samaritan exception, respond to widely shared concerns. However, it seems fairly clear, given President Trump's disputes with social media firms, that a major objective of some of these proposals, especially the removal of the "otherwise objectionable" language discussed earlier, was to restrict social media platforms' ability to block politically charged posts by conservative politicians.

6.3 THE LIMITS OF SECTION 230

Finally, the idea that Section 230 completely shields platforms from liability for their actions – i.e., Section 230 is an unlimited "Get Out of Jail Free" card – is an exaggeration. Certainly Section 230 has been read by courts (perhaps mistakenly, as noted earlier) to immunize platforms against essentially all claims arising from third-party content, or their efforts to moderate such content; but it does *not* immunize the platforms' own actions aside from content moderation, including the platforms' own speech. Thus if I post content defaming someone else on Facebook, Meta is immune. But if Meta itself posts defamatory content, Section 230 provides no defense. Furthermore, courts have consistently held that decisions platforms make about how to design their platforms, and what features to offer on platforms, are not protected by Section 230. Two examples can help illustrate what this means.

In May of 2017, three Wisconsin teenagers were speeding in a car at well over 100 MPH when they swerved off the road and hit a tree, killing all three of them. In 2019, the parents of two of the boys sued Snap, Inc., the owner of the Snapchat platform, in a federal district court in California (Snap's headquarters are in Santa Monica) claiming that the crash was a result of a Snapchat design feature. In particular, the parents' complaint claimed that at the time

[34] *Ibid.* at 71–73.

[35] *The Justice Department Unveils Proposed Section 230 Legislation, supra* n. 25.

of the crash, Snapchat had a feature called a "Speed Filter" which permitted users to record their actual current speed, and overlay that information on content they were uploading to Snapchat. The parents claimed a belief had emerged among Snapchat users that the app rewarded uploads taken at 100 MPH or higher, and that shortly before the accident, one of the boys (not the driver) opened Snapchat and used the Speed Filter.[36] The basis of their claim was that Snapchat was negligent in designing the Speed Filter feature, knowing as they did or should have that it incentivized dangerous conduct among their (mainly young) user base.

After the parents filed their lawsuit against Snap, Snap sought to have the case dismissed on the grounds of Section 230 immunity. Snap succeeded at the trial court level, but on appeal the United States Court of Appeals for The Ninth Circuit (covering the western states) reversed. In doing so the court drew a crucial distinction between claims which seek to treat a defendant platform as a "publisher or speaker" of third-party content, and claims based on negligent design of the platform itself. In this case, the court said, the parents were not seeking to impose liability on Snap based on the content uploaded just before the accident, but rather on the design of the Speed Filter feature itself. Of course, the accident was ultimately related to the fact that the boys were uploading content; but that was not the basis of the claim against Snap. As a result, the court permitted to lawsuit to proceed.[37]

The importance of the Ninth Circuit's *Lemmon* decision (the title of the case was *Lemmon v. Snap*) should be obvious. It clarifies that Section 230 does not generally shield social media platforms from responsibility for all misconduct, but only for their actions with respect to third-party content. The decision has, unsurprisingly, been highly influential, and has opened an (admittedly narrow) door to holding platforms accountable for harm for which they, themselves, are directly responsible. And it is widely accepted and supported, so much so that even the Electronic Frontier Foundation, one of the preeminent cyber-libertarian organizations in the world and a prominent defender of Section 230, expressed support for it.[38]

Important though it is, however, it should be acknowledged that the exact scope of the distinction drawn in *Lemmon* is unclear. Perhaps the greatest unresolved issue in this regard is whether Section 230 permits holding platforms liable for the results of computer algorithms they use to recommend

[36] Lemmon v. Snap, Inc., 995 F.3d 1085, 1088–89 (9th Cir. 2021).

[37] *Ibid*. at 1091–94.

[38] Sophia Cope, *Lawsuit against Snapchat Rightfully Goes Forward Based on a "Speech Filter," Not User Speech*, Electronic Frontier Foundation (May 18, 2021), www.eff.org/deeplinks/2021/05/lawsuit-against-snapchat-rightfully-goes-forward-based-speed-filter-not-user.

specific content to specific users. The Supreme Court was faced with this issue in its 2022–2023 Term in a pair of cases alleging that the major social media platforms (Facebook, Twitter/X, and YouTube) had facilitated terrorist attacks abroad by permitting the Foreign Terrorist Organization ISIS (also known as Islamic State) to use their platforms as recruiting tools. The lower court (as it happens, the Ninth Circuit again) had dismissed the plaintiffs' claims under the authority of Section 230, and crucially, had reaffirmed its conclusion in an earlier case that Section 230(c)(1) immunity barred legal claims based on platform recommendation algorithms, so long as the algorithms were not designed to favor illegal content.[39] On appeal, however, the Supreme Court ducked the Section 230 issue and instead held that the relevant federal anti-terrorism statutes did not reach the platforms' conduct.[40]

Of course, the Court's decision to avoid determining the scope of Section 230 did not either resolve the issue or make it go away, and the question of whether Section 230 *should* be read to immunize algorithmic recommendations has important implications for ongoing and future litigation.[41] Most notably, in the wake of the *Lemmon* decision (and others) literally hundreds of lawsuits were filed by local school districts, as well as a bipartisan group of state attorneys general, against the major platforms based on the allegedly "addictive" nature of platforms, which the claimants say has contributed to a mental health crisis among adolescents[42] (as noted in Chapter 2, the empirical basis for such claims remain highly disputed, but that is a separate question from the scope of Section 230 immunity). Though they vary slightly, the essential element of all of these lawsuits are claims that the platforms have intentionally created a series of features, including recommendation algorithms and endless loops, designed to addict children in ways that cause child users significant mental harm.

These lawsuits have all been consolidated before a single federal judge, Judge Yvonne Gonzales Rogers, who sits in Oakland, California. In the most recent significant development in this litigation as of this writing, on November 14, 2023, Judge Rogers issued a long and careful opinion resolving a motion by the platforms to dismiss the litigation. She concluded that based on Ninth Circuit precedent which she was bound to follow (including *Lemmon* and the terrorism cases discussed earlier), claims against the platforms based

[39] Gonzales v. Google LLC, 2 F.4th 871, 894–97 (9th Cir. 2021).

[40] Twitter, Inc. v. Taamneh, 598 U.S. 471 (2023); Gonzales v. Google LLC, 598 U.S. 617 (2023) (per curiam).

[41] The arguments are summarized in Rozenshtein, *supra* n. 12, at 71–73.

[42] Isaiah Poritz, *Social Media Addiction Suits Take Aim at Big Tech's Legal Shield*, Bloomberg Law (Oct. 25, 2023), https://news.bloomberglaw.com/tech-and-telecom-law/hundreds-of-social-media-addiction-suits-face-first-legal-hurdle.

on the features built into the platforms, rather than on decisions directly tied to specific third-party content, fell outside Section 230 as interpreted in the *Lemmon* decision.[43] However, Judge Rogers also confirmed that some of the claims against platforms, based on such things as failing to limit the length of time users could spend on platforms, creating endless streams of content, and using algorithms "to promote addictive engagement," *were* barred by Section 230 because in practice, such claims, if accepted, could be cured only by publishing less third-party content.[44] And notably, in reaching these conclusions Judge Rogers fully embraced the view that Section 230 immunity extends to recommendation algorithms.[45]

In short, despite the seemingly broad consensus among social media critics that Section 230 is fatally flawed, those critics cannot come close to agreeing *why* it is flawed. And furthermore, there remain important, open questions about the actual scope of Section 230 immunity vis-à-vis various kinds of legal claims, including (outside the Ninth Circuit) claims based on the results of platform recommendation algorithms.

6.4 IN DEFENSE OF SECTION 230: IMMUNITY FOR THIRD-PARTY CONTENT

One reason why the basic protections that Section 230 provides to social media platforms are socially beneficial is that to some significant degree, Section 230 protections overlap with First Amendment protections already available to social media (though this does *not* mean that the First Amendment makes Section 230 unnecessary[46]). Significant parts of modern First Amendment law are driven by an insight about "chilling effects": the idea that imposing broad or ill-defined liability on speakers and other First Amendment actors will cause those actors to "voluntarily" silence themselves or others out of fear of inadvertently crossing a legal line. This idea was the driving force behind one of the most influential and significant First Amendment decisions of all time, *New York Times v. Sullivan*.[47] In that case, which arose during the Civil Rights era, an Alabama city official sued the New York Times as well as four individuals active in the civil rights movement for libel, based on factual errors in a fund-raising advertisement that the individuals had placed in the Times. The US

[43] In re Social Media Adolescent Addition/Personal Injury Products Liability Litigation, 2023 W.L. 7524912 at *11–13 (N.D. Cal. 2023).

[44] *Ibid.* at *13–16.

[45] *Ibid.* at *14.

[46] The reasons why are ably explained by Professor Eric Goldman. Goldman, *supra* n. 22.

[47] 376 U.S. 254 (1964).

Supreme Court held that when public officials sought to recover damages for defamatory speech about their official conduct, they were required to prove that the falsehood was made with "actual malice," meaning that speaker acted with "knowledge that it was false or with reckless disregard for the truth." Later cases in this line extend First Amendment immunity to defamation claims by all public figures, and to liability claims other than defamation.[48]

The crucial aspect of the decision, however, was that the Court did not base its new rule, which protected even negligent falsehoods from liability, on the supposed value of false speech. Rather, all these decisions were driven by the insight that imposing liability on the media even for seemingly low value or unprotected speech can lead to the "voluntary" suppression of valuable, protected speech. It was this "self-censorship" that, in the Court's view, would seriously interfere with the public discourse at the center of our democracy.

There is an obvious analogy between the chilling effects the *Sullivan* Court identified and the position of social media platforms today – and indeed, it was concerns about such chilling effects that drove Congress to adopt Section 230 in the first place. To understand why that is so, one must first absorb the sheer scale of content moderation that social media platforms engage in. Consider the largest of the platforms, Facebook. Facebook is available in almost every country in the world, in over 100 languages (Reuters reported in 2019 that Facebook was available in 111 languages supported by Facebook, and another thirty-one without support[49]). That translates to over 2 *billion* daily active users on Facebook, a substantial percentage of the human race. And with that scale comes an enormous amount of content moderation. The Supreme Court's recent *NetChoice* decision (discussed in detail in Chapter 4) reports that in one quarter of 2021, Facebook blocked 25 million pieces of content just under its rule against hate speech, and another 9 million under its bullying and harassment policy (YouTube similarly reported blocking 6 million videos in one quarter alone).[50] Indeed, it is fair to say that for massive platforms such as Facebook, Instagram, YouTube, and TikTok, content moderation practices are not only a major part of their businesses; they are industrial in scale.[51]

[48] Curtis Publ'g Co. v. Butts, 388 U.S. 130, 155 (1967) (extending *Sullivan* holding to public figures); Hustler Magazine v. Falwell, 485 U.S. 46, 56 (1988) (extending *Sullivan* holding to claims by public figures for intentional infliction of emotional distress).

[49] Maggie Fick and Paresh Dave, *Facebook's Flood of Languages Leave It Struggling to Monitor Content*, Reuters (Apr. 23, 2019), www.reuters.com/article/us-facebook-languages-insight/facebooks-flood-of-languages-leave-it-struggling-to-monitor-content-idUSKCN1RZ0DW.

[50] Moody v. NetChoice, LLC, 144 S. Ct. 2383, 2406 (2024).

[51] The classic, albeit at this point slightly dated, account of how platform content moderation operates is Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598 (2018).

But with that scale comes an inevitable byproduct: mistakes. A 2020 report issued by the NYU Stern School of Business suggests that content moderation mistakes are ubiquitous. Indeed, Mark Zuckerberg, the CEO of Meta (which owns Facebook and Instagram), himself conceded that one out of ten content moderation decisions are mistaken – which translates to 300,000 mistakes *a day* for Facebook alone.[52] And while recent advances in Artificial Intelligence (AI) *might* help mitigate this problem in the future, it could also make the problem more intractable given the scale at which generative AI can produce and post content, including problematic content. And note that the inevitability of content moderation mistakes has implications for both sides of the Section 230 puzzle: It means that platforms will inevitably sometimes block content that does not violate their rules (or the law), and sometimes fail to block content that does.

These well-known facts lead to a simple conclusion: Without some basic form of Section 230 immunity, social media as we know it could not exist. Indeed, without Section 230, defamation liability alone would shut down social media as we know it. But this principle itself has implications for Section 230 "reform." In particular, for the same reasons that, as discussed in Chapter 4, the Supreme Court extended First Amendment editorial rights to social media platforms in the *NetChoice* cases, it seems very likely that the principles underlying *Sullivan* and later cases will also apply to social media. Thus, it may well be that the First Amendment requires any reductions in Section 230 immunity to be tempered with some sort of scienter requirement for platforms, such as the *Sullivan* "actual malice" standard. Or to put it differently, it may well be that the Constitution itself forbids imposing *publisher,* meaning strict or negligence, liability on platforms regarding the types of speech that fall within the protection of the *Sullivan* line of cases, even if distributor liability, based on actual knowledge, was permissible.

I will discuss later why reading (or amending) Section 230 to permit distributor, but not publisher, liability creates its own set of serious problems. For now, however, we should note that the *Sullivan* case and its progeny do not in any way shield media from liability to entirely private individuals when the speech at issue does not involve matters of public concern.[53] As such, at least under current law private defamation, whether published in traditional media or posted to platforms, does not trigger First Amendment protections. For newspapers, this is not a major problem because most content in such

[52] John Koetsier, *Report: Facebook Makes 300,000 Content Moderation Mistakes Every Day*, Forbes (June 9, 2020), www.forbes.com/sites/johnkoetsier/2020/06/09/300000-facebook-content-moderation-mistakes-daily-report-says/.

[53] Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc., 472 U.S. 749 (1985).

publications is presumed to be of public concern. Unlike with *news* media, however, with *social* media there can be no such assumption; to the contrary, much content on social media is purely private and often personal (I for one tend to use Facebook to report on nice hikes).

This legal regime, which favors political and social commentary over private speech about private persons, certainly reflects the current state of First Amendment law; and it might well have made sense with respect to traditional media whose focus was public affairs. But its consequences for social media are quite troubling – and in this regard the Constitution is almost beside the point. The reality is that online defamatory statements about private figures are actually much harder to police than potential defamation of public figures, because information about private individuals is not readily accessible from available and trustworthy records. Nor, as just noted, would such statements receive First Amendment protections. So, without Section 230 immunity from at least publisher liability, platforms would have to simply ban potentially defamatory content, which is to say any critical statements about private persons. That, in practice, is the end of the "social" aspect of social media (aside from happy thoughts and cat videos, since cats can't sue for defamation).

Nor is the problem limited to defamatory statements. Consider threatening speech. It is well-established law that "true threats" are a form of unprotected speech under the First Amendment to the US Constitution.[54] However, in a recent case involving systematic, online harassment of a women by a stranger, the Supreme Court held that threats and harassment may be punished under the First Amendment only if the speaker was "reckless," meaning that he or she was consciously aware that others could see the speech at issue as threatening but decided to proceed anyway.[55] This is obviously an extraordinarily difficult line for courts and victims to draw, but for platforms it is an impossible one given their lack of access to the state of minds of individual users. Thus if platforms did face publisher liability for what are called "true threats" posted by users, their only recourse would be to block *all* sharp criticisms that could possibly be read to implicate violence, including sharp criticisms of public figures. Such chilling effects raise serious First Amendment concerns and have very troubling implications for democratic discourse.

The way to think about the consequences of eliminating or significantly limiting publisher liability is that doing so would leave platforms with effectively two choices: either to stop carrying unvetted third-party content (and so effectively end the social element of social media) or to massively ramp up

[54] Virginia v. Black, 538 U.S. 343, 359 (2012) (plurality opinion).

[55] Counterman v. Colorado, 600 U.S. 66, 79–80 (2023).

content moderation to catch all (or almost all) content that creates the risk of liability. The former "solution" would destroy a multi-billion-dollar industry that provides services that billions of people around the world evidently value, an outcome only a Luddite could support. And as for the second potential solution, massively tightening content moderation, that poses its own problems. First of all, for all of the reasons described earlier such almost-perfect content moderation is impossible, even (or especially) in the age of AI. Furthermore, even if a behemoth such as Facebook could afford to undertake the enormous amounts of content moderation that repeal of Section 230 would require, several of its smaller competitors have already expressed concerns about their ability to do so,[56] which means that such Section 230 "reform" would further concentrate an already overly concentrated industry.

Finally, however, excessive content moderation is objectively troubling for policy reasons. For one thing, some commentators have pointed out that eliminating or severely restricting Section 230 immunity will inevitably lead social media firms to over-filter borderline speech on topics such as sexuality, which could work to the detriment of marginal groups such as LGBTQ youth.[57] In addition, such a change would turbocharge conservative complaints (discussed on Chapter 1) about platform "bias" against conservative voices. And more generally, tightening up content moderation will inevitably result in the silencing of huge amounts of legitimate and protected speech, especially political speech, throughout the world. There is evidence that this is precisely what happened at Facebook when Germany adopted its so-called NetzDG law that imposed massive fines on platforms that, after being notified, permitted hate speech to remain available.[58] Eliminating Section 230 immunity would simply spread that political chill to the United States, on a much more massive scale.

Moreover, essentially all of the bad results I have argued will result from amending Section 230 to permit *publisher* liability would also follow from amending (or as Justice Thomas proposes interpreting) Section 230 to permit imposing *distributor* liability on platforms. Recall that while publisher liability

[56] Todd Shields and Ben Brody, *Facebook Worries Smaller Rivals with Openness on Liability*, Yahoo! Fin. (Dec. 23, 2020), https://finance.yahoo.com/news/facebook-support-liability-reform-little-070000635.html.

[57] *See* Bill Easley, *Revising the Law that Lets Platforms Moderate Content Will Silence Marginalized Voices*, Slate (Oct. 29, 2020), https://slate.com/technology/2020/10/section-230-marignalized-groups-speech.html.

[58] Rebecca Zipursky, Note, *Nuts about NETZ: The Network Enforcement Act and Freedom of Expression*, 42 Fordham Int'l L.J. 1325, 1359–60 (2019); Linda Kinstler, *Germany's Attempt to Fix Facebook Is Backfiring*, The Atlantic (May 18, 2018), www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/.

would hold platforms liable, either strictly or upon a showing of negligence, for harmful third-party content, distributor liability kicks in only if a platform *knew* or had reason to know that it was hosting illegal or harmful content. Justice Thomas and others would (contrary to the *Zeran* case) read Section 230 to permit distributor liability, and others have proposed amending Section 230 to do so. The problem is that distributor liability effectively forces platforms to implement a take-down regime, under which once *anyone* tells a platform that some specific content it is hosting is illegal, that platform must immediately make a judgment about whether to risk keeping the flagged content up or take it down. But of course, the incentives faced by platforms in this situation are asymmetric. Ideally, a platform would like to make the right call every time and only take down actually illegal, flagged content. But, given blurry lines between protected and illegal content, that is impossible. So, in case of doubt a platform has two choices – take down the content and irritate a single user, or leave up the content and potentially face massive fines. That is precisely the choice that the German NetzDG law gave to platforms, and their reaction was utterly predictable: in case of *any* doubt, take it down. The result is a significant burden on the speech of countless users who have posted sharply worded but legal content.

Even worse, a take-down regime not only creates perverse incentives for platforms; it also creates an opening for bad actors among the public. Platforms host lots of content that, while perfectly legal and even legitimate, angers others. That includes honest but negative online reviews, sharp criticisms of individuals, political opinions outside the mainstream (in either direction), and so forth. Take-down regimes effectively encourage those who dislike specific posts to flag them to the platform. Doing so is almost effortless (since take-down regimes inevitably require platforms to create easy means to flag content) and has little downside from the point of view of the flagger, with the potential "upside" of (at least temporarily) getting rid of the content. As a consequence, it is utterly predictable that take-down regimes will encourage huge numbers of dubious or even clearly false notices, with the consequent negative effects on free speech. Indeed, the take-down regime created by the US Digital Millennium Copyright Act regarding intellectual property has been so thoroughly abused that it has led the Electronic Frontier Foundation to create a "Takedown Hall of Shame."[59] Laws like NetzDG, or interpreting Section 230 to permit distributor (even if not publisher) liability, would spread this burden beyond speech implicating intellectual property to *all* speech, including especially political speech, on platforms.

[59] Electronic Frontier Foundation, *Takedown Hall of Shame*, www.eff.org/takedowns.

In fact, the kind of take-down regime I describe is precisely what the European Union (EU) put into place in late 2022 when it adopted its "Digital Services Act" (DSA).[60] Article 6 of the DSA states that platforms shall not be liable for third-party content that they host so long as the platform "does not have actual knowledge of illegal activity or illegal content" *and* "upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the illegal content"[61] – in other words, the DSA imposes distributor liability. And Article 16 implements this approach by mandating that platforms create "notice and action" (i.e., notice and take-down) mechanisms that permit individuals to flag content that an individual "considers to be illegal content."[62] Finally, another part of the DSA requires platforms to give priority to notices submitted by "trusted flaggers," which are entities that a member EU government has designated as trusted.[63]

While it is early days, the full impact of these provisions of the DSA is potentially appalling. For one thing, the DSA applies to any content that is illegal under the law of any member state. As Professor Dawn Nunziato points out, that extends to hate speech, Holocaust denial, and glorification of Nazi ideology in Germany; but also to criticism and parody of the President of France under French law, to pro-LGBTQ+ content accessible to minors under Hungarian law, and to *blasphemy* under Austrian and Finnish law.[64] Furthermore, the creation of the "trusted flagger" program is a clear invitation to illiberal governments (yes, they exist in the EU) to drown platforms in complaints – to which platforms are required to give priority – by designating "trusted flaggers" who will repeatedly objecting to content the government dislikes. This would surely put pressure on platforms to voluntarily block speech to which the authorities object, even if that speech is probably legal, simply to avoid the burden of processing complaints (and as a useful side effect, to get into the relevant government's good graces).

But even assuming official good faith, the DSA regime clearly creates a system ripe for abuse by individual flaggers, for reasons already explained. But the only response to this problem is Article 23(2) of the DSA, which states that platforms must stop accepting, "for a reasonable period of time," notices from individuals or entities "that frequently submit notices or complaints that are manifestly unfounded."[65] Presumably in the EU's view false flagging resulting

[60] https://eur-lex.europa.eu/eli/reg/2022/2065/oj ("DSA").

[61] *Ibid.*, Art. 6.

[62] *Ibid.*, Art. 16.

[63] *Ibid.*, Art. 22.

[64] Dawn Carla Nunziato, *The Digital Services Act and the Brussels Effect on Platform Content Moderation*, 24 Chi. J. Int'l Law 115, 119–20 (2023).

[65] DSA, Art. 23(2).

in wrongful take-downs is fine so long as it is not “frequent” and “manifestly unfounded.” It is too early to know (as of this writing) what the long term consequences of the DSA for free speech in Europe will be; but the prognosis is not good.

Finally, while I have focused on the EU’s DSA, the problem I describe is not unique to Europe. In particular India, the largest democracy in the world and (not coincidentally) home to the largest number of Facebook users,[66] has a similar problem. In 2021, India released what are commonly known as the Information Technology Rules (or “2021 IT Rules”),[67] which implemented provisions of the Information Technology Act of 2000.[68] Sections 14 and 15 of the Rules, acting on the authority of Section 69A of the 2000 Act, create a mechanism permitting executive branch officials to order the blocking of online content deemed to be illegal or otherwise unprotected. In addition, like the DSA, Rules 3(2) and 4(c) of India’s 2021 IT Rules also provide the public with a mechanism to file complaints with platforms, to which platforms must respond. But given gaps in enforcement, the primary impact of the Indian IT Rules has been to greatly enhance the power of the central government, led by Prime Minister Narendra Modi and his Bharatiya Janata Party, to pressure platforms to block content – especially because Indian law permits criminal prosecution of tech executives for violations of these rules. Press reports strongly suggest that the consequence has been a dramatic increase in government control over, and censorship of, online content the government considers objectionable.[69]

6.5 IN DEFENSE OF SECTION 230: IMMUNITY FOR CONTENT MODERATION

The discussion to this point demonstrates why any serious curtailment of platform immunity for third-party content that they carry would have (and in some countries has had) very bad consequences for the liberty of social media

[66] Tom Fish, *These Countries Have the Most People on Facebook*, Newsweek (Sept. 2, 2021), www.newsweek.com/counties-most-people-facebook-1624911. China has more total users, but famously blocks the major US platforms.

[67] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.

[68] Information Technology Act, 2000.

[69] Karishma Mehrotra and Joseph Menn, *How India Tames Twitter and Set a Global Standard for Online Censorship*, Washington Post (Nov. 8, 2023), www.washingtonpost.com/world/2023/11/08/india-twitter-online-censorship/; Varsha Bansal, *India’s Government Wants Total Control of the Internet*, Wired (Feb. 13, 2023), www.wired.com/story/indias-government-wants-total-control-of-the-internet/

users, and more broadly for free public discourse. But in fact, significantly altering the *other* aspect of Section 230 immunity – freedom to moderate content in good faith without fear of liability – would be even worse.

To understand why, it is important to recall that the internet is full of content that is "lawful but awful," meaning content that most people do not want to have to confront but that constitutes fully protected speech under the First Amendment to the US Constitution (though not, in many cases, in other countries). Examples of lawful-but-awful content under US law include almost all hate speech,[70] non-obscene pornography,[71] calls for political violence that do not fall within the extremely narrow definition of "incitement" adopted by the US Supreme Court[72] (including expressing praise and support for terrorist organizations such as ISIS/Islamic State and Hamas), and even deliberate lies that do not cause tangible and provable harm.[73] At first cut, eliminating platform immunity for good faith content moderation would open platforms up to lawsuits for blocking any of this content. Indeed, presumably the whole purpose of Section 230 reform is to prevent platforms from blocking political content they do not like. But hate speech and ISIS propaganda are no less "political viewpoints" than speech supporting Democrats or Republicans. As such, proposals such as those advanced by Senators Rubio and Hawley that condition Section 230 immunity on viewpoint-neutral content moderation[74] would, strikingly, give terrorists and hate groups a free pass on social media.

Another possibility is to amend Section 230 to distinguish between legal and illegal (i.e., constitutionally unprotected) content and to limit Section 230(c)(2)(A) immunity to content moderation of illegal content – the effect of which would be to make it financially risky for platforms to moderate legal content. But this will not work because, as Chapter 4 discusses in more detail, it is extraordinarily difficult for platforms to easily distinguish between legal and illegal, for two separate reasons. The first is that the law is often far from clear in defining unprotected content. And second, even when the law is clear (as with child pornography), it can be very difficult to determine which side of the line any particular piece of content falls on. This, indeed, is precisely why the drafters of Section 230 did not attempt to draw such lines, instead saying

[70] Matal v. Tam, 582 U.S. 218, 243–44 (2017) (plurality opinion); *ibid*. at 248–50 (Kennedy, J., concurring in part and concurring in the judgment); R.A.V. v. City of St. Paul, Minn., 505 U.S. 377, 391–92 (1992).

[71] United States v. Playboy Entertainment Group, Inc., 529 U.S. 803, 811 (2000).

[72] Under that definition, speech calling for violence falls outside the First Amendment only if it is "directed to inciting of producing imminent lawless action and is likely to produce such action." Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).

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

[74] *See supra* n. 26–27.

explicitly that Section 230(c)(2)(A) immunity extended to content moderation of constitutionally protected content.

Moreover, an approach distinguishing legal and illegal content would do nothing to solve the "lawful but awful" problem. To do *that*, Section 230 would have to somehow draw a legally enforceable distinction between "good" and "bad" *legal* content, immunizing only content moderation of "bad" content. But again, it turns out that this is also a nonstarter.

For one thing, such a move may well violate the First Amendment. As explained in Chapter 4, the Supreme Court clearly held in the *NetChoice* cases that the First Amendment protects platforms' editorial rights to choose what content to carry and what content not to carry. This means that a direct government command to favor certain content – as Texas and Florida sought to impose – violates the First Amendment. But surely what the government cannot do directly, it also cannot do indirectly. And so, while there is no Supreme Court case that directly addresses this question, it seems likely that the Court would strike down a law, such as a gerrymandered version of Section 230(c)(2)(A), that legally incentivizes privately owned platforms to carry some government-favored viewpoints ("ISIS is terrible") while blocking government-disfavored viewpoints ("ISIS is great").

In addition, such a distinction is often impossible to make. Consider the problem of pornography. Presumably Section 230 reformers would want to permit platforms to block non-obscene pornography without fear of liability. But at the same time, surely not all nudity falls within this category, as illustrated by an episode in which Facebook first blocked, then unblocked, the iconic "Napalm Girl" photograph from the Vietnam War.[75] Drawing the line between pornography and valuable but explicit materials in a meaningful way has confounded courts for decades; which is why they do not try to draw such lines, and why Section 230 reformers will be unable to do so as well.

Or consider praise for political violence. Presumably reformers, especially conservative ones, would want to continue to permit platforms to block ISIS propaganda or praise for ISIS attacks, foreign and domestic. At the same time, presumably many conservatives would *not* want to immunize efforts to block praise for the January 6, 2021, incursion on the US Capitol, or calls for similar actions in the future. But imagine trying to draw *that* line in any coherent way. The reality is that distinguishing between content based on its social value in an objective, enforceable, and constitutional manner is well nigh impossible.

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

6.6 IN DEFENSE OF SECTION 230

Finally, it should be evident that as bad as repealing either specific provision of Section 230 would be, it would be worse if, as Presidents Trump and Biden proposed, we repealed Section 230 altogether. Under such a regime, platforms would be extremely reluctant to moderate any content because drawing lines between legal and illegal content is so difficult, and making the wrong decision (as they inevitably would do) creates serious financial risks. But on the other hand, failing to block content also creates risks. In a publisher-liability regime, the risks are so severe that platforms would need to block any content that could conceivably cross the line into illegality (meaning a potential source of liability). And even in a distributor-liability regime, the moment a platform was informed that particular content *might* be illegal, the same incentives apply.

So, without Section 230 in place, platforms both *cannot* block any content that might be legal and *must* block all content that could conceivably be illegal, to be safe. Obviously, doing both things at once is not possible, placing all platforms carrying third-party content between a rock and a hard place. Perhaps the very largest and wealthiest platforms (meaning Meta and Google/YouTube) could find a way to navigate these opposing risks, but for smaller platforms this would spell doom. And in truth, it is not clear that even the larger platforms could find a way through. Which would leave them with two ways forward: either block *all* third-party content and eliminate the "social" aspect of social media; or permit utterly anodyne third-party content that raises no risk of liability either way (lots of cat videos) but nothing else.

Either way, social media as the home of vibrant social and political dialogue would come to an end. And that is an outcome we should all be very, very wary of, involving as it does an inconceivable reversal in the democratization of public discourse that the internet has enabled. For that reason Section 230, mainly in its current form, must stay despite all of its weaknesses; the alternatives are much worse.

6.7 CODA: A FEW THOUGHTS ON ALGORITHMS

Moving forward, some of the most contentious and legally troubling issues surrounding social media will focus on the use of computer algorithms to recommend or amplify certain content (which means by implication not amplifying other content). As noted in Chapter 4, in the *NetChoice* litigation the Supreme Court strongly suggested that the use of algorithms to recommend and amplify favored content, as well as to deamplify or block other content, is itself an exercise of editorial rights protected by the First Amendment.[76] But the exact scope of that First Amendment protection remains quite unclear – in her concurring opinion, Justice Barrett (who provided the crucial fifth vote for the relevant parts of the majority opinion) suggested that such protection *might* not apply "if a platform's algorithm just presents automatically to each user whatever the algorithm thinks the user will like" (i.e., if the algorithm is not enforcing substantive community standards chosen by the platform).[77] She also suggested that protection might be denied if platforms turn over enforcement of their content standards (whatever they are) to AI.[78] Finally, she also raised questions about protection for content moderation decisions made under the influence of foreign owners of platforms – an obvious allusion to ongoing controversies over TikTok's Chinese ownership.[79] So it remains important whether Section 230, which unlike the First Amendment does not limit the scope of the immunity it grants, applies to algorithmic (and eventually AI-driven) recommendations and amplification.

To date, as noted earlier, this issue has not been addressed by the US Supreme Court, but the lower courts have tended to treat algorithmic recommendations as an aspect of the "publication" process protected by Section 230(c)(1).[80] However, in the important Ninth Circuit *Gonzalez* decision regarding platform liability for terrorist acts (which later made its way to the Supreme Court), two of the three judges on the panel that decided the case raised doubts about applying Section 230 to recommendation algorithms (albeit one of the two voted for immunity on the grounds that she was bound by precedent);[81] and in a similar case in the Second Circuit, one of the three judges disagreed with the majority on the immunity point.[82] The issue thus remains very much a point of contention among judges, and also among commentators.

[76] Moody v. NetChoice, LLC, 144 S. Ct. 2383, 2403–06 (2024).; *see also ibid*. at 2410 (Barrett, J., concurring).

[77] *Ibid*. at 2410 (Barrett, J., concurring).

[78] *Ibid*.

[79] *Ibid*.

[80] Dryoff v. Ultimate Software Grp., Inc. 934 F.3d 1063, 1098 (9th Cir. 2019); Force v. Facebook, Inc., 934 F.3d 53, 64–72 (2nd Cir. 2019); Gonzalez v. Google LLC, 2 F.4th 871, 894–97 (9th Cir. 2021).

[81] *Gonzalez*, 2 F.4th at 912–18 (Berzon, J., concurring); *ibid*. at 920–21 (Gould, J., concurring in part and dissenting in part).

[82] *Force*, 934 F.3d at 80–84 (Katzmann, C.J., concurring in part and dissenting in part). In a recent decision, the Third Circuit held that in light of the Supreme Court's reasoning in the *NetChoice* cases the output of recommendation algorithms should be considered a platform's own speech, and so outside of Section 230 immunity. Anderson v. TikTok, Inc., 116 F.4th 180, 183-84 (3rd Cir. 2024).

The truth is that the strictly correct "legal" answer is to this question is almost impossible to resolve because when Section 230 was adopted by Congress in 1996 nothing like modern social media platforms, much less recommendation algorithms, existed. And while a fairly straightforward analogy can be drawn (based on plain language) between modern platforms and the message boards and discussion forums of 1996, the same is simply not true of algorithms. So, to ask what Congress "intended" in 1996 regarding recommendation algorithms is nonsensical. Furthermore, a purely linguistic analysis of Section 230 is also indeterminate. On the one hand, deciding what content should be emphasized on a platform (like what story goes on the front page of a newspaper) is an intrinsic part of the process of hosting third-party content protected by Section 230(c)(1); but on the other hand, arguably making recommendations goes beyond merely hosting "any information provided by another information content provider," the thing that Section 230 was designed to protect. As such, the idea that this disagreement can be clearly resolved using narrow legal tools of statutory interpretation is wishful thinking.

So, I would turn the issue around and ask, *should* Section 230 be read to protect algorithmic recommendations – and concomitantly, if courts conclude it does not provide such protection, should Congress amend Section 230 to add such protections? I think that the answer is clear – it should, and Congress should (what Congress actually *would* do, in its current state of dysfunction, is of course a separate question).

The reason for this is simple: In today's world, given the sheer scale of social media platforms, recommendations are an absolutely essential element of running platforms for third-party content, if they are to function at all. The alternative, after all, is to serve up content in either random or chronological order, which would make most feeds useless, especially on open-ended platforms like YouTube and Twitter/X, but also on platforms like Facebook that serve up a combination of posts by friends and other, more generalized content. Indeed, it is precisely because of this that recommendation and amplification choices are central to "publishing" third-party content on a social media platform, just as choosing what stories to run on the front page is an essential element of publishing a newspaper. Or as Tarleton Gillespie put it, content moderation "is, in many ways, *the* commodity that platforms offer."[83]

Now consider what Congress intended to accomplish in creating Section 230 immunity. It intended to permit platforms to serve up third-party content without fear of liability *and* to make sure that platforms did not lose that immunity by making choices about what content to carry (i.e., to engage in content

[83] Tarleton Gillespie, Custodians of the Internet 13 (2018).

moderation). Remember in this regard that Section 230 was a response to judicial decisions imposing publisher liability on platforms who engaged in content moderation. In those early days of the internet, the relevant content moderation consisted only of blocking harmful content, since, given low volumes of users and content (remember, we are talking about message boards and discussion forums), that was all that was required. Today, however, recommendation and amplification decisions are as essential an element of running a platform for third-party content as is blocking harmful content, given exponentially higher volumes of users and content. Without incorporating such a function, platforms would serve up mainly useless and random garbage, from the point of view of users. Publication of third-party content, in other words, *requires* making choices about what content to amplify and to which users to amplify it.

Furthermore, precisely because of the volume of content and users, algorithms are the only feasible way to implement those decisions. It is simply impossible to imagine human beings making the millions of decisions that social media algorithms constantly make, twenty-four hours a day, across the world, regarding what content to recommend to each individual user. So again, if we wish to accomplish Congress's goal of permitting effective platforms for third-party content to thrive, we must protect not only their right to make recommendations but also their ability to use algorithms to make and implement those recommendation and amplification decisions.

Finally, for practical purposes Section 230 immunity is essential if platforms are going to be able to make and implement those algorithmic choices. Just as it is practically impossible for platforms to ensure that all third-party content they carry is legal and does not cause harm, so too there is no way to ensure that their algorithms will never end up directing users to harmful content (whether harmful to themselves, such as content encouraging self-harm, or in the terrorism cases harmful to innocent third parties). Again, the volume of recommendations is simply too large to be perfectly policed. As such, without immunity, platforms would be paralyzed in their ability to run recommendation algorithms if they faced potential liability for every choice – precisely the result that Section 230 seeks to avoid.

None of which is to say that Section 230 should be read to automatically immunize *all* algorithmic choices. Certainly if a platform *deliberately* designed their algorithms to serve up illegal or harmful content, that would pose a very different situation – and there is no doubt that there exist some niche platforms who are conscious bad actors and who should not be able to shield their deliberate choices through Section 230. In fact, that sort of design choice would seem to fall within the *Lemmon* exception to Section 230 immunity.

But to eliminate any doubt, it is probably a good idea, as Professors Danielle Keats Citron and Mary Anne Frank have proposed, to amend Section 230 to make this point clear.[84] But for most mainstream platforms, so long as they can demonstrate that their algorithms were designed based on neutral criteria and were not intended to favor illegal or harmful content (even if, on occasion, they end up doing so), Section 230 should protect their choices.

Finally, what about AI? Recall that in *NetChoice*, Justice Barrett suggested that turning over content moderation to AI might argue against First Amendment protection, since no human being is any longer making choices about what content to permit/block and favor/disfavor (unlike with traditional algorithms, whose creation involves such human choices). I frankly am doubtful if Justice Barrett's First Amendment doubts are justified, since the First Amendment protects *speech*, not *speakers*.[85] But regardless of First Amendment questions, I think it would be utterly strange to suggest that Section 230 immunities rely on the technology being used to run a platform. Nothing in the text of the statute suggests such a limitation. Furthermore, if Section 230 immunity had been tied to specific technology, it would have quickly become obsolete as the message boards and discussion forums of the 1990s evolved into the modern internet. Tying such protections (or for that matter regulations – more on this in Chapter 8) to specific technologies is a terrible policy choice, given how quickly internet technology changes. Section 230 was originally, and wisely, written in much broader and more flexible terms, and we should continue to read it in that light to protect *all* new technologies that facilitate platforms for third-party content.

[84] Citron and Franks, *supra* n. 31, at 70–71.

[85] I would note that this view is supported by the text of the First Amendment, which simply says "Congress shall make no law … abridging the freedom of speech," without regard to the source of that speech. My position in this regard is also rooted in a broader view that the Bill of Rights as a whole was designed not to protect the rights of individuals but rather to provide structural protections against government overreach. *See* Ashutosh Bhagwat, The Myth of Rights (2010).