



HARVARD Kennedy School

**MOSSAVAR-RAHMANI CENTER**  
for Business and Government

# **Content Moderation in the United States and Europe: Similar Values, Different Approaches**

**Philip L. Verveer**  
**Harvard Kennedy School**

January 2023

## **M-RCBG Associate Working Paper Series | No. 197**

The views expressed in the M-RCBG Associate Working Paper Series are those of the author(s) and do not necessarily reflect those of the Mossavar-Rahmani Center for Business & Government or of Harvard University. The papers in this series have not undergone formal review and approval; they are presented to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

## Content Moderation in the United States and Europe: Similar Values, Different Approaches

*“The modern Internet in the United States is built on more than two decades of reliance on Section 230. To remove those twenty-six words from the United States Code would unravel the Internet that we know today.”<sup>1</sup>*

*The significance of Section 230 of the Communications Decency Act is undeniable. Without it, the open microphone-engendered “dynamic, communal, and vicious public square”<sup>2</sup> that is social media and other familiar Internet genre would be very different.*

*Gonzalez v. Google LLC, Petition for Certiorari granted, No. 21-1333 (October 3, 2022).*

*Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022—Digital Services Act*

One hundred years ago the eminent British economist Arthur Pigou identified the problem of externalities, of a business not absorbing all of the costs associated with the goods or services it produced and sold. Classic examples of negative externalities are environmental pollution and health effects from tobacco. Today, in addition to the carcinogenic effects of chemical runoffs and first and secondhand tobacco smoke, we have to contend with a new problem: the poisoning of democratic systems through foreign influence campaigns, intentional dissemination of misinformation, and incitements to violence inadvertently enabled by Facebook, YouTube and other major digital platform companies.

The principal response of the American policy community to the externality problem has focused on Section 230 of the Communications Decency Act.<sup>3</sup> The 1996 statute shelters the platforms’ editorial judgments from liability to an extent otherwise unknown in U.S. jurisprudence.<sup>4</sup> It effectively holds them harmless with respect to third party content that they

---

<sup>1</sup> Jeff Kosseff, *THE TWENTY-SIX WORDS THAT CREATED THE INTERNET* 8 (2019).

<sup>2</sup> *Id.*, at 1.

<sup>3</sup> Telecommunications Act of 1996, Pub. L. No. 104-104, § 230, 110 Stat. 56, 137–39, (codified as amended at 47 U.S.C. § 230 (2018)).

<sup>4</sup> “This was and remains an idiosyncratic and exceptional treatment under law,” Olivier Sylvain, “Intermediary Design Duties,” 50 *Connecticut L. Rev.* 203, 212 (2018).

host or decline to host. The principal response of the European Union is embodied in the Digital Services Act, adopted in October 2022, which also shields platforms from liability for third party content, but also assigns them significant responsibilities.<sup>5</sup>

Sheltering digital platforms from liability for content provided by unrelated parties raises, inter alia, three observations at the threshold:

First, the familiar open microphone business models deployed by Facebook, YouTube, Twitter, and others either literally rely on the liability protection or very nearly do. If and as the protection changes, so very likely will the business models.

Second, by virtue of the protection, the amount that platforms invest in content moderation is a function of their assessment of their customers’—both users and advertisers—preferences. The near absolute immunity means platforms can, if they choose, avoid or minimize expenditures associated with content moderation and with litigation, including payment of damages.

Third, any consideration of the existing level of platform immunity and of adjusting it inevitably foregrounds the value a society places on freedom of speech and conversely the value a society ascribes to suppressing speech likely to lead to harmful consequences. The sweeping immunity accorded platforms by the United States and the European Union more than two decades ago reflected a very strong free speech orientation. Whether implicit or explicit, proposals to reduce or condition the immunity necessarily revisit that orientation.

The level of concern about the negative externalities thrown off by social media company services has escalated steadily in recent years. The level of scholarly and other commentary describing and admiring the problem has increased commensurately. And the number of proposed legislative ameliorations also has increased. The production of proposals to adjust (or not) Section 230 has progressed from the artisanal to the industrial. The supply has flooded the market. What has not happened is the adoption in the United States of any public measure or measures aimed at diminishing the problem.<sup>6</sup>

---

<sup>5</sup> Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022—Digital Services Act

<sup>6</sup> Although this has changed in the European Union. See *infra* notes 83-129 and accompanying text.

But that could change in the next few months. For the first time since Section 230 was adopted, the Supreme Court will hear a case that has the distinct potential to affect the statute’s scope and, in consequence, the prevailing business models employed by many digital platforms.<sup>7</sup>

Concerns about insufficient and ineffective content moderation are exacerbated by a perceived lack of competition. High levels of concentration in the social media space, associated among other things with significant network effects, provide an additional predicate for government intervention. If consumer preferences for improved levels of content moderation are not being met by marketplace competition, there is a case for intervention to reduce market power in addition to the case for intervention designed to reduce negative externalities. The European Union has taken up both of those cases, adopting major initiatives in the last months of 2022.

The market power of the major digital platforms and the desire to reduce it, in addition to major enforcement actions against them,<sup>8</sup> has produced a parallel set of scholarly proposals. The most prominent among them converge on the recommendation that traditional antitrust remedies alone are insufficient. Rather, in addition to the application of competition law, new purpose built regulatory regimes will be required.<sup>9</sup> The recommendations contributed to changes in law

---

<sup>7</sup> *Gonzalez v. Google LLC*, cert granted No 21-1333. The Court also will hear a second case, *Twitter, Inc. v. Mehiar Taamneh*, cert granted, No. 21-1496 (October 3, 2022), that bears upon platform content moderation practices.

<sup>8</sup> *See, e.g.* *Federal Trade Commission v. Facebook, Inc.* (DDC 1:20-cv-03590-JEB); *New York v. Facebook, Inc.* (DC Cir. 21-7078); *United States v. Google LLC* (DDC 1:20-cv-03010-APM); *Colorado v. Google LLC* (DDC 1:20-cv-03715-APM); *Texas v. Google LLC* (SDNY 1:21-cv-06841-PKC).

<sup>9</sup> *United Kingdom Competition & Markets Authority, Online platforms and digital advertising, Market study final report*, 1 July 2020. (“CMA Report”);

Jacques Cremer, Yves-Alexandre De Montjoye, & Heike Schweitzer, *Euro. Comm’n Directorate General for Competition, Competition Policy for the Digital Era* (Apr. 4, 2019) (“Vestager Report”)

<https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf> ;

Jason Furman, et al., H.M. Treasury (UK), *Unlocking Digital Competition: Report of the Digital Competition Expert Panel* (March 13, 2019) (“Furman Report”),

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/785547/unlocking\\_digital\\_competition\\_furman\\_review\\_web.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf)

Stigler Committee on Digital Platforms: *Final Report*, September 16, 2019,

<https://research.chicagobooth.edu/stigler/media/news/committee-on-digital-platforms-final-report>.

Tom Wheeler, et al., *New Digital Realities; New Oversight Solutions in the U.S.*, Shorenstein Center, Harvard Kennedy School (August 2020) (“Wheeler Report”), [https://shorensteincenter.org/wp-content/uploads/2020/08/New-Digital-Realities\\_August-2020.pdf](https://shorensteincenter.org/wp-content/uploads/2020/08/New-Digital-Realities_August-2020.pdf)

*See also* G7 (United Kingdom 2021), surveying initiatives of eleven countries, at paras. 3.9 and 4.29:

[T]here is a growing consensus that additional mechanisms, powers, or safeguards are necessary and existing approaches should be modernised or strengthened to address the specific attributes of digital

in the European Union that address both competition and externality issues.<sup>10</sup>

One thesis is that it is both possible and desirable to establish a new government institution with the remit of assisting and thus improving the effectiveness of antitrust-based remedies and of requiring and supervising improved platform content moderation with due sensitivity to freedom of expression, privacy, and other values fundamental to the U.S. Constitution and legal system.<sup>11</sup>

At the time of enactment of Section 230, there was a recognition that widening of the possibilities for individual expression would produce benefits. There also was a recognition that the Internet and the business models taking advantage of its opportunities presented costs.<sup>12</sup>

Twenty-seven years later, we have a more concrete sense of the benefits and the costs. Both the benefits and the costs are incalculable except through the use of inevitably arbitrary assumptions, but it is indisputable that both are large. Accepting the proposition, just as Section 230 did, that we want to secure the benefits, there remains the difficult issue of finding and exercising mechanisms that would lower the costs without commensurately compromising the benefits.<sup>13</sup>

---

markets. While the reforms and reform proposals vary in content and scope, most facilitate easier or faster agency intervention or contemplate new regulatory regimes.

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1036995/Compendium\\_final\\_format.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1036995/Compendium_final_format.pdf)

<sup>10</sup> In addition to the Digital Services Act, *supra* note 5, the European Union has adopted the Digital Markets Act, Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022.

<sup>11</sup> *See, e.g.*, S. 4201, 117<sup>th</sup> Congress, 2<sup>nd</sup> Session, Bill “To establish a new federal body to provide reasonable oversight and regulation of digital platforms” and H.R. 7858 (same), introduced by Sen. Bennet and Rep. Welch. *See also* Wheeler Report, *supra* note 9.

<sup>12</sup> *See infra* notes 15-17. *And see* *Packingham v. North Carolina*, 137 S.Ct. 1730, 1736 (2017) (Kennedy, J.):

For centuries now, inventions heralded as advances in human progress have been exploited by the criminal mind. New technologies, all too soon, can become instruments used to commit serious crimes. The railroad is one example, see M. Crichton, *The Great Train Robbery*, p. xv (1975), and the telephone another, see 18 U.S.C. § 1343. *So it will be with the Internet and social media.* (emphasis supplied)

<sup>13</sup> There is not universal agreement that adjustments to Section 230 are desirable. Professor Goldman is a leading proponent of the view that the existing arrangements need not and should not be changed in a material way. *See, e.g.*, Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 *Notre Dame L. Rev. Reflection* 33 (2019); Eric Goldman, “Speech Showdowns at the Virtual Corral,” 21 *Santa Clara Computer & High Tech. L. J.* 845, 852-53 (2005):

Section 230 grants online providers a near-blanket immunity from liability for their users' content. This immunity applies whether the online provider tries to control content it deems objectionable or not, meaning that online providers can figure out the best way to serve their communities. With this legal protection, a thousand online communities have bloomed, spanning the spectrum from tightly controlled to

The harmful effects from the operations of major Internet platform companies adversely affect the American population every day, resulting in offsets to the undeniable value that the companies' services also provide. These harms take the form of defamation of individuals, groups, and institutions, dissemination of hate speech, incitements to violence, spreading of misinformation, promotion of fraudulent schemes, and foreign interference in our democratic processes.<sup>14</sup> Significantly, the services that are provided by platforms convey both the valuable and the destructive, the benign and the malign, with equal efficiency. To reiterate, the importance to our society of the major platform companies is indisputable, but so is the fact that their products and services have been exploited by geopolitical adversaries, criminals, disaffected individuals, and others in ways that inflict material costs on the public. Those costs can be tangible, as when houses of worship are required to hire security guards because of platform-disseminated hate speech and incitement to violence, or intangible, as when state actors and state-sponsored actors seek to weaken our social bonds and influence our elections, or both,<sup>15</sup> or when misinformation about COVID-19 vaccines prolongs a pandemic and intensifies social divisions.<sup>16</sup>

---

virtually unregulated. This diversity has allowed individuals to find venues that serve their needs, giving customers the power to reward (or punish) providers for their choices. Section 230 played a non-trivial role in the Internet's ascension as a dominant media, a development from which we have all benefited.

(citations omitted)

<sup>14</sup> There is a wide range of views concerning the significance of this problem. *See*, for example, Matt Perault, "Section 230: A Reform Agenda for the Next Administration," 4 (October 26, 2020): "Research has found that the impact of problematic online content may be relatively small". Available at [https://9381c384-0c59-41d7-bbdf-62bbf54449a6.filesusr.com/ugd/14d834\\_16adf8519cc64ab5a6bc5e6a700126da.pdf](https://9381c384-0c59-41d7-bbdf-62bbf54449a6.filesusr.com/ugd/14d834_16adf8519cc64ab5a6bc5e6a700126da.pdf) By contrast, Julie. E. Cohen, "Tailoring Election Regulation: The Platform is the Frame," 4 *Geo. L. Tech. Rev.* 641, 647 (2020) argues that "In a networked media ecosystem designed for content targeting, optimization for engagement, and amplification of information flows, polarized and polarizing content spreads rapidly from one platform to another and between online and traditional media, gaining in volume as it travels." (citation omitted).

<sup>15</sup> *See, e.g.*, Robert S. Mueller, III, Report on the Investigation into Russian Interference in the 2026 Presidential Election, "Russian 'Active Measures' Social Media Campaign," (2019); *United States v. Internet Research Agency LLC* (DDC 1:18-cr-00032-DLF) (indictment secured by Special Counsel Mueller); US Department of State, Pillars of Russia's Disinformation and Propaganda Ecosystem GEC Special Report: August 2020. The concern, of course, is not limited to the United States. *See also*, United Kingdom Foreign, Commonwealth and Development Office, Press release, "UK exposes sick Russian troll factory plaguing social media with Kremlin propaganda" (May 1, 2022).

<sup>16</sup> The Report of the Joint Committee of the House of Lords and the House of Commons on the Online Safety Bill Draft 3-4 (December 14, 2021) raises similar concerns:

[S]ome service providers [value] the engagement of users at all costs, regardless of what holds their attention. This can result in amplifying the false over the true, the extreme over the considered, and the harmful over the benign. The human cost can be counted in mass murder in Myanmar, in intensive care beds full of unvaccinated Covid-19 patients, in insurrection at the US Capitol, and in teenagers sent down rabbit holes of content promoting self-harm, eating disorders and suicide.

As a result of these and other readily available examples, the “victims” include those who have no *a priori* ability to protect themselves from the resulting damage.<sup>17</sup> What is much worse, the damage occurs to the country’s underlying economic and social fabric, with losses both material and immaterial spreading in ways that threaten the social stability and cohesion on which democratic societies depend.

The legal domain in which these externalities arise is affected by the existence of the First Amendment, which significantly limits the potential responses of the United States government to them.<sup>18</sup> Although some negative externalities — those associated with child pornography and imminent incitement to violence, for example, -- plainly fall outside of First Amendment protection,<sup>19</sup> there obviously are First Amendment values that must be respected in dealing with others.

And so the public policy question, what can be done to maintain the benefits while reducing the costs?

Excluding the clearly illegal, the ability or inability to address most of the perceived problems in this area is a function of Section 230 that was intended, among other things, to promote the growth of the then-nascent Internet by effectively holding platforms harmless for third party content that they host or decline to host. It did not affect the potential liability of those who provide content through platform facilities, the ultimate sources of the information that is disseminated over the Internet.<sup>20</sup> Today, Section 230 exists in a technological and

---

<sup>17</sup> Tarleton Gillespie, “Platforms Are Not Intermediaries,” 2 *Geo. L. Tech. Rev.* 198, 203 (2018). The worst example of harm to innocent victims occurred outside of the United States, involving the targeting of the Rohingya people in Rakhine State, Myanmar. *See* Sheera Fankel and Cecilia Kang, AN UGLY TRUTH: INSIDE FACEBOOK’S BATTLE FOR DOMINATION 169-187 (2021)

<sup>18</sup> For a review of the juxtaposition of First Amendment jurisprudence and the challenges presented by social media, *see, e.g.*, Kyle Langvardt, “A New Deal for the Online Public Sphere,” 26 *Geo. Mason L. Rev.* 341 (2019).

<sup>19</sup> *See, e.g.*, Robert A. Sedler, “An Essay on Freedom of Speech: The United States versus the Rest of the World,” 2006 *Mich. St. L. Rev.* 377, 379 (2006) and Brett M. Frischmann, “Speech, Spillovers, and the First Amendment,” 2008 *U. Chi. Legal Forum* 301, 304, 317 observing that “...the First Amendment is not absolute; the government can and does regulate speech in some limited cases, often with the aim of internalizing negative externalities. ... [F]or certain types of speech, the costs and benefits of speech are distributed unevenly across groups so that speech that is beneficial to some is harmful to others.” (citations omitted).

<sup>20</sup> The Communications Decency Act’s legislative history is complicated. *See infra* notes 35 and 36 and accompanying text.

institutional milieu that is very different from the Internet environment of twenty-seven years ago, in part because of Section 230 itself.<sup>21</sup>

Section 230 of the Communications Decency Act contains three significant provisions. Subsection (c)(1) provides 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.”<sup>22</sup> That is, if a computer service merely provides a mechanism through which others disseminate information, it *cannot* be held liable for any adverse effects of the dissemination of that information on others *nor can users who simply repost content originated by others*. Subsection (c)(2) provides that “No provider or user of an interactive computer service shall be held liable on account of— (A) 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; or (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).”<sup>23</sup> That is, a service *can* restrict access to information that it finds “objectionable” without incurring liability for doing so. Importantly, subsection (e) carves out certain matters from the immunity protection: violations of federal criminal law, state laws, intellectual property law, certain communications privacy laws, and sex trafficking law.<sup>24</sup>

On the one hand, an interactive computer service platform cannot be held liable if it chooses to do nothing to limit access to the information that is available through its service and, on the other, it cannot be held liable if it chooses to limit such access if it finds the information to

---

<sup>21</sup> See Langvardt, *supra* note 18 at 343-346.

<sup>22</sup> 47 USC 230(c)(1) (emphasis added) The statute defines an interactive computer service as “any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions” and an information content provider as “any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.”

“[P]ublication involves reviewing, editing, and deciding whether to publish or to withdraw from publication third-party content.” *Barnes v. Yahoo, Inc.*, 570 F.3d 1096, 1102 (9<sup>th</sup> Cir. 2009). “Publisher,” an undefined term, is a central issue in *Gonzalez*.

<sup>23</sup> 47 USC 230(c)(2) (emphasis added).

<sup>24</sup> 47 USC 230(c)(e).

be “objectionable”.<sup>25</sup> To the extent that there is liability for the dissemination of harmful information, it is to be imposed only on the source of the information, not, with limited exceptions, on the computer service through which it was disseminated.

It has been observed that the combination of this broad immunity and the business models, especially those of advertiser-supported internet services, has diminished the incentives of Internet platforms to engage in content moderation. For example, as the Stigler Committee on Digital Platforms has noted:

[T]he goal of all these [digital platforms] is to maximize engagement, often through extreme or divisive content, as recognized by Facebook itself. Unlike other media, however, [digital platforms] do not have any legal liability for promoting this content, thanks to Section 230 of the Communications Decency Act. This immunity, combined with the limited competition these platforms face, means that [digital platforms] have very weak incentives to promote quality content or to limit the spread of false or divisive information.<sup>26</sup>

The effects of the Section 230 incentive structure on interactive computer services are reflected in the outcomes of numerous lawsuits claiming harm. The most prominent of the early cases tended to involve garden variety, if appalling, tortious acts.<sup>27</sup> More recently, they have come to include harms emanating from geopolitical controversies. The holdings of the vast majority of cases prove that Section 230 is a very effective shield.<sup>28</sup>

In many cases, the platform services have been challenged on the grounds that their failure to control the information to which they provide access has injured third parties -- that their actions are the sources of negative externalities. In most cases, Section 230 has shielded them from liability for any adverse effects that the dissemination of information through their

---

<sup>25</sup> Kathleen Ann Ruane, “How Broad a Shield? A Brief Overview of Section 230 of the Communications Decency Act,” Congressional Research Service Legal Sidebar 1 (February 21, 2018) lists Facebook, Twitter, and Google as among the entities that “are permitted to publish others’ content without reviewing it for criminality or other potential legal issues.” It is important to note, however, that the list of interactive computer services, the entities whose behavior is protected under Section 230, is far longer than this and, in fact, Ms. Ruane notes that “Reviewing courts have interpreted [the] definition [of interactive computer services] to cover many entities operating online, including broadband Internet access service providers (e.g., Verizon FIOS and Comcast Xfinity), Internet hosting companies (e.g., DreamHost and GoDaddy), search engines (e.g., Google and Yahoo!) online message boards and many varieties of online platforms.” *Id.* at 2.

<sup>26</sup> Stigler Committee Final Report, *supra* note 9, at 10.

<sup>27</sup> Professor Kosseff, *supra* note 1, has supplemented the courts’ summary descriptions of the salient, troubling facts in many of the litigated cases with enlarged recitations of the record evidence.

<sup>28</sup> *See, e.g.*, Eric Goldman, “Why Section 230 Is Better Than the First Amendment,” 95 Notre Dame L. Rev. Reflection 33 (2019). (sweeping nature of 230 immunity).

services may have had.<sup>29</sup> This has led to proposals to modify Section 230, most often by adjusting the subsection (e) immunity carveouts to make interactive computer services liable if the information to which they provide access causes certain defined harms.<sup>30</sup>

The understanding of Section 230's immunizing provisions was established by the first appellate case considering the statute and has remained largely undisturbed ever since. In 1997 in *Zeran v. America Online* Chief Judge J. Harvie Wilkinson III of the 4<sup>th</sup> Circuit gave its immunizing provisions the broadest possible scope.<sup>31</sup>

The purpose of this statutory immunity is not difficult to discern. Congress recognized the threat that tort-based lawsuits pose to freedom of speech in the new and burgeoning Internet medium. The imposition of tort liability on service providers for the communications of others represented, for Congress, simply another form of intrusive government regulation of speech. Section 230 was enacted, in part, to maintain the robust nature of Internet communication and, accordingly, to keep government interference in the medium to a minimum. In specific statutory findings, Congress recognized the Internet and interactive computer services as offering "a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity." *Id.* § 230(a)(3). It also found that the Internet and interactive computer services "have flourished, to the benefit of all Americans, *with a minimum of government regulation.*" *Id.* § 230(a)(4) (emphasis added). Congress further stated

---

<sup>29</sup> See *infra* notes 31-34 and 37-45 and accompanying text. Professor Kosseff finds that "Section 230's first decade was marked by a rapid expansion of immunity for websites, [but] the second decade saw a gradual—but real—erosion of Section 230 immunity . . ." Kosseff, *supra* note 1, at 203. See also Jeff Kosseff, "The Gradual Erosion of the Law that Shaped the Internet: Section 230's Evolution Over Two Decades," 18 Colum. Sci. & Tech L. Rev. 1 (2016). He also notes that "[S]ome plaintiffs, knowing of courts' relatively broad interpretation of Section 230, may be discouraged from ever bringing a lawsuit against online intermediaries." *Id.*, at 36-37.

<sup>30</sup> The most controversial proposal emanated from the Trump Administration. See letter from Attorney General William P. Barr to The Honorable Michael R. Pence dated September 23, 2020 in which the Attorney General proposes that platforms would be excluded from immunity if they: "(1) purposefully promote, facilitate, or solicit third party content that would violate federal criminal law; (2) have actual knowledge that specific content it is hosting violates federal law; and (3) fail to remove unlawful content after receiving notice by way of a final court judgment." More controversially, the letter also proposes that the definition of information content provider be broadened to include "situations in which a platform 'solicits, comments upon, funds, or affirmatively and substantively contributes to, modifies, or alters the content of another person or entity.'" The "comments upon" provision appears designed to address a contemporaneous controversy involving Twitter's, Facebook's, and others' warnings concerning posts that contained misinformation. "The Justice Department Unveils Section 230 Legislation," (September 23, 2020). Available at <https://www.justice.gov/opa/pr/justice-department-unveils-proposed-section-230-legislation>. See Exec. Order No. 13,925, May 28, 2020, 85 Fed. Reg. 34,079 (June 2, 2020); *revoked*, Exec. Order No. 14, 029, May 14, 2021, 86 Fed. Reg. 27, 025 (May 19, 2021).

Other suggested reforms include requiring that services increase transparency in their handling of content and content-related disputes and conditioning immunity on the adoption of "reasonable" content moderation practices. See, e.g., Danielle Keates Citron and Benjamin Wittes, "The Problem Isn't Just Backpage: Revising Section 230 Immunity," 2 Geo. L. Tech. Rev. 453, 455-456 (2018). (condition immunity on a service provider taking reasonable steps to prevent or address unlawful third-party content that it knows about).

<sup>31</sup> *Zeran v. America Online Inc.*, 129 F.3d 327 (4<sup>th</sup> Cir. 1997).

that it is "the policy of the United States ... to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, *unfettered by Federal or State regulation.*" *Id.* § 230(b)(2) (emphasis in original ).<sup>32</sup>

The decision went on to note that "By its plain language, § 230 creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service."<sup>33</sup> Although it is indisputable that part of the motivation for the adoption of Section 230 was to enable freer exchange of information on the Internet, the court also recognized that "[a]nother important purpose of § 230 was to encourage service providers to self-regulate the dissemination of offensive material over their services."<sup>34</sup> This observation points to an important part of the Communications Decency Act's origin. When the Act was adopted, there were conflicting Congressional views about how best to discourage asserted harms, especially the availability to children of indecent and offensive material.<sup>35</sup> The important point is that there was some expectation of active content moderation despite the fact that subsection (c)(1) taken in isolation apparently affords service providers an opportunity to avoid liability through complete passivity.<sup>36</sup>

---

<sup>32</sup> *Id.*, at 330.

<sup>33</sup> *Id.*

<sup>34</sup> *Id.*, at 331.

<sup>35</sup> The way in which, with the aid of the Supreme Court, Section 230's approach to content moderation, won out over an alternative employing Constitutionally suspect direct prohibition is described in Chief Judge Katzmann's opinion in *Force v. Facebook, Inc.*, 934 F.3d 53, 77-80 (2<sup>nd</sup> Cir. 2019) (Katzmann, C.J., concurring in part and dissenting in part). See also Kosseff, *supra* note 1 (offering an excellent history of Section 230 and the cases leading to its passage) and Zeran v. America Online Ebook, Eric Goldman and Jeff Kosseff (eds.), (2020) Available at <https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=3286&context=historical> (including essays describing additional history of Congress' deliberations over Section 230 and the history of the Zeran litigation). See also Danielle Keates Citron and Benjamin Wittes, "The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity," 86 *Fordham L. Rev.* 401, 404-406 (2017) and Sylvain, *supra* note 4, at 235-38.

<sup>36</sup> As former Representative Christopher Cox and Senator Ron Wyden, the authors of Section 230, recalled, "[W]e began with these two propositions: let's make sure that every internet user has the opportunity to exercise their First Amendment rights; and let's deal with the slime and horrible material on the internet by giving both websites and their users the tools and the legal protection necessary to take it down." Reply Comments of Co-authors of Section 230 of the Communications Act of 1934, 8 FCC Docket RM-11862 (September 17, 2020). For additional perspectives on the significance of content moderation, see Danielle K. Citron & Benjamin Wittes, *supra* note 30 and Stigler Committee, *supra* note 9, at 192, (News Industry Subcommittee Report) which observes that:

Although sometimes viewed as a sweeping libertarian intervention, Section 230 actually began life as a smut-busting provision: an amendment for the "Protection for Private Blocking and Screening of Offensive Material." Its purpose was to allow and encourage Internet service providers to create safe spaces, free of pornography, for children. (citations omitted)

The European Union's e-Commerce Directive, which was adopted four years after the Communications Decency Act, is similar to Section 230(c)(1). The Directive's Articles 12, 13, and 14 addressing conduit, caching, and hosting services afford providers immunity and do not require them "to monitor the information which they transmit or

Following *Zeran*, interactive computer service defendants most often have been found *not* to be liable for the adverse effects of the content that appeared on their services, even where those effects were indisputably noxious. For example, in *Force v. Facebook, Inc.*, the plaintiffs, who included “U.S. citizen victims, and relatives and representatives of the estates of those victims, of certain terrorist attacks committed by Hamas in Israel,” sued Facebook alleging that it had provided Hamas “...with a communications platform that enabled those attacks.” The Court of Appeals for the Second Circuit upheld a district court’s dismissal of the complaint based on Section 230 immunity.<sup>37</sup>

In *Doe v. Backpage.com, LLC*, each of the three plaintiffs alleged that she was the victim of sex trafficking at age 15 and each alleged that she was subject to rape--“over 1000 times,” “over 900 times,” and “on numerous occasions,” respectively. Although the First Circuit panel, including retired Supreme Court Justice Souter, concluded that the plaintiffs had made a persuasive case “that Backpage has tailored its website to make sex trafficking easier,” Section 230 “requires that we ... deny relief to plaintiffs whose circumstances evoke outrage.”<sup>38</sup>

In *Barnes v. Yahoo!*, Section 230 was found to preclude recovery on a state law negligence claim notwithstanding “a dangerous, cruel, and highly indecent use of the internet for the apparent purpose of revenge.”<sup>39</sup> In that case, the court found that:

[The plaintiff] Barnes did not authorize her now former boyfriend to post the profiles, which is hardly surprising considering their content. The profiles contained nude photographs of Barnes and her boyfriend, taken without her knowledge, and some kind of open solicitation, whether express or implied is unclear, to engage in sexual intercourse. The ex-boyfriend then conducted discussions in Yahoo's online "chat rooms," posing as Barnes and directing male correspondents to the fraudulent profiles he had created. The profiles also included the addresses, real and electronic, and telephone number at Barnes' place of employment. Before long, men whom Barnes did not know were peppering her office with emails, phone calls, and personal visits, all in the expectation of sex.<sup>40</sup>

---

store, nor ... to seek facts or circumstances indicating illegal activity.” Article 15. The Digital Services Act retains this policy, which has been referred to as the “passivity paradox.” See Joris V.J. van Hoboken, European Intermediary Liability in Historical Perspective, March 21<sup>st</sup> 2017, Available at <https://www.ceps.eu/wp-content/uploads/2017/03/CEPS%20-%20Limited%20liability%20for%20the%20Net%20-%20Joris%20van%20Hoboken.pdf>.

<sup>37</sup> 934 F.3d 53 (2<sup>nd</sup> Cir. 2019).

<sup>38</sup> 817 F.3d 12, 17, 29 (1<sup>st</sup> Cir. 2016).

<sup>39</sup> 570 F.3d 1096, 1098 (9<sup>th</sup> Cir. 2009).

<sup>40</sup> *Id.*

Similarly, in *Herrick v. Grindr LLC*, the plaintiff, who was “the victim of a campaign of harassment by his ex-boyfriend, who created Grindr profiles to impersonate Herrick and communicate with other users in his name, directing the other users to Herrick’s home and workplace,” brought suit against Grindr, a “hook-up” application, arguing that it “was defectively designed and manufactured because it lacks safety features to prevent impersonating profiles and other dangerous conduct, and that Grindr wrongfully failed to remove the impersonating profiles created by his ex-boyfriend.” A district court ruled that Grindr was an interactive computer service, so that Herrick’s claims were barred by Section 230, among other factors, and the ruling was upheld by the Court of Appeals for the Second Circuit.<sup>41</sup>

More recently in *Dyroff v. Ultimate Software, Inc.*,<sup>42</sup> the 9<sup>th</sup> Circuit considered a case in which a platform connected an individual seeking to acquire heroin with a drug dealer. The individual bought heroin laced with fentanyl, which led to his death. The court rejected the plaintiff’s contention that the platform’s recommendation and notification functions—which connected the decedent and the dealer—was not entitled to Section 230 immunity.<sup>43</sup>

The *Dyroff* decision, and its rejection of *Fairhousing Council of San Fernando Valley v. Roommates.com, LLC*<sup>44</sup> as a precedent, led the 9<sup>th</sup> Circuit to affirm Section 230 immunity in *Gonzalez v. Google LLC*, the case now before the Supreme Court.<sup>45</sup>

The broad construction of Section 230 immunity has been questioned repeatedly by appellate judges, including by Justice Thomas in a statement on the denial of certiorari in *Malwarebytes, Inc. v. Enigma Software Grp USA, LLC*.<sup>46</sup> “Extending § 230 immunity beyond the natural reading of the text can have serious consequences. Before giving companies immunity from civil claims for ‘knowingly host[ing] illegal child pornography,’ or for race discrimination, we should be certain that is what the law demands.”<sup>47</sup>

---

<sup>41</sup> 765 Fed. Appx. 586 (2d Cir. Mar. 27, 2019) (unpublished summary order).

<sup>42</sup> 934 F.3d 1093 (9<sup>th</sup> Cir. 2019).

<sup>43</sup> *Id.*, at 1096-1098.

<sup>44</sup> See *infra* notes 53-54 and accompanying text.

<sup>45</sup> See *infra* notes 67-69 and accompanying text.

<sup>46</sup> 141 S.Ct. 13 (2020).

<sup>47</sup> *Id.*, at 18 (internal citations omitted).

Similarly, circuit judges have made noteworthy attempts to locate more flexibility in Section 230 than *Zeran* and its progeny recognized. Judge Gould of the 9<sup>th</sup> Circuit,<sup>48</sup> Judge Tymkovich of the 10<sup>th</sup> Circuit,<sup>49</sup> and Chief Judge Katzmann of the 2<sup>nd</sup> Circuit,<sup>50</sup> among others, have urged a narrower reading, but unsuccessfully.

As Professor Koseff has noted, despite the influence of *Zeran*, there have been decisions that have avoided granting immunity to interactive computer services. These have stemmed largely from two theories: that the sites have contributed to illegal content and that the sites were engaged in activities other than publishing and speaking.<sup>51</sup>

For example, in *MCW, Inc. v. badbusinessbureau.com*, the court held that the defendant that created such category headings as “Con Artists” and “Corrupt Companies” for consumer complaints was not immune. The Court found that

The CDA does not distinguish between acts of creating or developing the content of reports, on the one hand, and acts of creating or developing the titles or headings of those reports, in the other. The titles and headings are clearly part of the web page content. Accordingly, the defendants are information content providers with respect to the website postings and thus they are not immune from MCW’s claim.... the defendants are also information content providers because they are ‘responsible, in whole or in part, for the creating or development’ of third party derogatory messages.<sup>52</sup>

*Fair Housing Council of San Fernando Valley v. Roommates.com, LLC* involved a website that required disclosure of personal information—gender, family status, and sexual orientation—as a condition of use and employed profiles derived from the information as an integral part of its real estate service. “By requiring subscribers to provide the information as a condition of accessing its service, and by providing a limited set of pre-populated answers, Roommate becomes much more than a passive transmitter of information provided by others; it

---

<sup>48</sup> See *Batzel v. Smith*, 333 F.3d 1018 (2003) (Gould, J. dissenting); *Gonzalez v. Google, Inc.*, 2 F.4<sup>th</sup> 871, 918 (Gould, J. concurring and dissenting).

<sup>49</sup> *FTC v. Accusearch Inc.*, 570 F.3d 1187, 1204 (10<sup>th</sup> Cir. 2009) (Tymkovich, J. concurring).

<sup>50</sup> *Force v. Facebook*, 934 F.3d 53, 77-89 (2d Cir. 2019) (Katzmann, C.J., concurring and dissenting).

<sup>51</sup> Koseff, *supra* note 1, at 203.

<sup>52</sup> *MCW, Inc. v. badbusinessbureau.com*, 2004 WL 833595, 2004 U.S. Dist. Lexis 6678 (N.D. Tex. Apr. 19, 2004).

becomes the developer, at least in part, of that information.”<sup>53</sup> Based on this finding, the Ninth Circuit concluded that Roommates could not claim immunity under the terms of Section 230.<sup>54</sup>

Although these decisions illustrate that an entity risks losing the immunity provided under Section 230 if it is perceived as more than a mere conduit for information provided by others, the preponderance of cases reinforces a conclusion that an interactive computer service will be able to avoid being classified as an information provider, and thus subject to liability for harm caused by the content that it disseminates, if it is careful *not* to take actions that modify that content.

Section 230’s immunity stretches beyond platforms to include “users” who “republish” content created by others. As Professor Perry notes,

[A]ctions like “sharing” another person’s post on Facebook, “liking” a post to the extent that it brings the content to others’ attention, “retweeting” on Twitter, and “reblogging” on Tumblr, are mere dissemination of third-party content, and cannot give rise to liability. Forwarding an email with offensive content similarly enjoys § 230 immunity, as does hyperlinking to unlawful content, even if accompanied by explicit and enthusiastic endorsement of that content.<sup>55</sup>

This is another important way in which Section 230 privileges online activity over more traditional methods of information dissemination, with obvious if unintended reductions in individuals’ ability to secure redress for defamation, cyber bullying, intentional infliction of emotional harm, and other forms of denigration that have been amplified by repetition. Thus “user” immunity is another way in which Section 230 departs not only from offline jurisprudence, but also from the legal approaches of other nations.

American law makes a clear-cut distinction between offline republication, which gives rise to a new cause of action against the republisher (subject to a few limited exceptions), and online republication, which enjoys an almost absolute immunity. Other Western jurisdictions employ more generous republisher liability regimes, which usually require endorsement, a knowing expansion of exposure, or repetition.<sup>56</sup>

---

<sup>53</sup> 521 F. 3d 1157, 1166 (9<sup>th</sup> Cir. 2007) (en banc).

<sup>54</sup> *See also* FTC v. Accusearch, 570 F.3d 1187 (10<sup>th</sup> Cir. 2009) and Doe v. Internet Brands, 824 F.3d 840 (9<sup>th</sup> Cir. 2016).

<sup>55</sup> Ronen Perry, “The Law and Economics of Online Republication,” 106 Iowa L. Rev. 721, 740 (2020).

<sup>56</sup> *Id.*, at 773.

Two cases have a prominent place in the user immunity jurisprudence.

In *Batzel v. Smith*, the 9<sup>th</sup> Circuit determined that a Dutch art security expert who had reposted an email that falsely accused a U.S. citizen of possessing art stolen by Nazis, reinforced by the also false indication that she was a descendent of Heinrich Himmler,<sup>57</sup> would be immune from a defamation suit if found to be a “user.”<sup>58</sup> The court’s reasoning follows *Zeran* in concluding that Congressional intent required a very broad construction of Section 230.<sup>59</sup>

As noted,<sup>60</sup> the decision drew a strong dissent from Judge Gould:

The majority rule licenses professional rumor-mongers and gossip-hounds to spread false and hurtful information with impunity. So long as the defamatory information was written by a person who wanted the information to be spread on the Internet (in other words, a person with an axe to grind), the rumormonger's injurious conduct is beyond legal redress. ... Nothing in the text, legislative history, or human experience would lead me to accept the notion that Congress in § 230 intended to immunize users or providers of interactive computer services who, by their discretionary decisions to spread particular communications, cause trickles of defamation to swell into rivers of harm.<sup>61</sup>

Three years later, the California Supreme Court rejected Judge Gould’s dissent in favor of the broad immunity found by *Zeran* and by the *Batzel* majority in a case involving the reposting of an accusation of stalking. It found user immunity while simultaneously expressing concern about the consequences. “We share the concerns of those who have expressed reservations about the *Zeran* court's broad interpretation of section 230 immunity. The prospect of blanket immunity for those who intentionally redistribute defamatory statements on the Internet has disturbing implications.”<sup>62</sup> Nevertheless, the majority concluded that Congressional intent required the result:

[T]he congressional purpose of fostering free speech on the Internet supports the extension of section 230 immunity to active individual "users." It is they who provide much of the "diversity of political discourse," the pursuit of "opportunities for cultural development," and the exploration of "myriad avenues for intellectual activity" that the statute was meant to protect. (§ 230(a)(3).) The approach taken by the *Batzel* dissent would tend to chill the free exercise of

---

<sup>57</sup> *Batzel v. Smith*, 333 F.3d 1018, 1021-1022 (9<sup>th</sup> Cir. 2003).

<sup>58</sup> *Id.*, at 1035.

<sup>59</sup> *Id.*, at 1027-1028.

<sup>60</sup> *See supra* note 48.

<sup>61</sup> 555 F.3d, at 1038. (Gould, J. concurring and dissenting).

<sup>62</sup> *Barret v. Rosenthal*, 51 Cal.Rptr.3d 55, 77 (2006).

Internet expression, and could frustrate the goal of providing an incentive for self-regulation.<sup>63</sup>

Concerns about the effects of platform and user immunity have been accompanied by a consideration that wasn't prominent at the time *Zeran* and many of the other decisions broadly construing Section 230 issued. It is that latent problems in the fundamental social media business model intentionally have been actualized by Facebook<sup>64</sup> and others through algorithmic amplification.<sup>65</sup> The specific concern follows from an elemental business strategy: Facebook wants to maximize the attractiveness of its service and so has designed the algorithms underlying its News Feed to prolong each individual user's online sessions. This enables the acquisition of more individual data and the sale of more and better targeted advertising. The broader problem is that the inducement to prolong a session involves an appeal to emotion, and among the more powerful emotions are those involving confirmation of individual bias and anger. The sophisticated exploitation of these facets of human nature involves in too many cases, for example, the provision of misinformation to individuals particularly susceptible to it or increasingly extreme political perspectives to individuals particularly susceptible to radicalization.<sup>66</sup>

This, then, is the understanding of Section 230 as the Supreme Court addresses *Gonzalez v. Google*.

---

<sup>63</sup> *Id.*

<sup>64</sup> This is a particularly troubling narrative from individuals with direct experience with corporate Facebook. *See, e.g.,* Roger McNamee, *ZUCKED: WAKING UP TO THE FACEBOOK CATASTROPHE*, (2019); Dipayan Ghosh, *TERMS OF DISSERVICE: HOW SILICON VALLEY IS DESTRUCTIVE BY DESIGN* (2020); Chris Hughes, "It's Time to Break Up Facebook," *New York Times* (May 9, 2019); Dipayan Ghosh and Nick Couldry, "The lawsuits against Facebook don't go far enough," *Washington Post* (December 11, 2020).

<sup>65</sup> *See, e.g.,* Mary Anne Franks, "Justice Beyond Dispute," 131 *Harv. L. Rev.* 1374, 1381-1382 (2018):  
Unlike the commercial or institutional actors in pro-social disputes, social media platforms in anti-social disputes often have no incentive to resolve or prevent the conflicts at issue. They may in fact have incentives to ignore or even to aggravate them. This is due in large part to the business model of many social media companies. They do not make money by selling products; they make money by selling ads. Increased engagement with their platforms, whether for pro-social or anti-social purposes, translates into increased profits: "[A]busive posts still bring in considerable ad revenue and the more content that is posted, good or bad, the more ad money goes into their coffers." This can create incentives for platforms to be indifferent to or even encouraging of inequalities of power among users. For some of these platforms, online abuse may be, as the saying goes, "not a bug but a feature." (citations omitted)

<sup>66</sup> *See, e.g.,* Sheera Finkel and Cecilia Kang, *supra* note 17, at 182-83 (2021); Langvardt, *supra* note 18, at 357-363. "An algorithm designed to drive engagement at all costs is unlikely to push users toward contemplation, intellectual challenge, or doubt. Instead, it will tend to forward users a selection of bias-affirming materials that by turns soothe and provoke the user into more Facebooking." *Id.* at 358. *Accord*, Danielle Keats Citron, "Cyber Mobs, Disinformation, and Death Videos: The Internet as It Is (and as It Should Be)," 118 *Mich. L. Rev.* 1073, 1085-1086 (2020).

*Gonzalez* presents facts similar to those at issue in *Force*. Nohemi Gonzalez, a U.S. citizen studying in Paris, was killed in a 2015 attack by the Islamic State of Iraq and Syria (ISIS). Her family sued Google pursuant to the Antiterrorism Act.<sup>67</sup> It alleged that Google’s YouTube subsidiary’s algorithmic recommendations facilitated ISIS proselytization and recruiting. In other words, it argued that YouTube’s recommendation system was not immunized by Section 230; that recommendation systems producing the amplification and magnification of content should not “be treated as the publisher or speaker of any information provided by another information content provider.” Rather, the recommendation system was an addition supplied by YouTube the uses and consequences of which the company should be held legally responsible. The panel rejected the claim in a split decision<sup>68</sup> and the 9<sup>th</sup> Circuit declined an *en banc* request.<sup>69</sup>

The Supreme Court granted the Gonzalez family’s Petition for Certiorari at the beginning of the present term to consider the question

Does section 230(c)(1) immunize interactive computer services when they make targeted recommendations of information provided by another information content provider, or only limit the liability of interactive computer services when they engage in traditional editorial functions (such as deciding whether to display or withdraw) with regard to such information?

The matter to be decided has a certain metaphysical quality: is a targeted recommender algorithm an intrinsic part of an open microphone platform business, in which case immunized, or an external, non-intrinsic function added by the platform operator, in which case not immunized?

The Gonzalez family’s advocates have based their case on the text of Section 230 without providing an explicit perspective on any larger consequences of a decision favorable to them.<sup>70</sup> Although the case could well be decided on textual analysis, dictionary definitions, and the other conventional paraphernalia of appellate law, Google argues the “real world” consequences

---

<sup>67</sup> Antiterrorism Act of 1990, 18 U.S.C. 2331 *et seq.*

<sup>68</sup> *Gonzalez v. Google LLC*, 2 F.4<sup>th</sup> 871 (9<sup>th</sup> Cir. 2021).

<sup>69</sup> 21 F.4<sup>th</sup> 665 (9<sup>th</sup> Cir. 2022) (*en banc*).

<sup>70</sup> The Gonzalez brief does advert to the larger context in which to view platforms’ recommender algorithms: “Internet companies are constantly adjusting their recommendation systems to improve their effectiveness in inducing viewers to spend more time on the site looking at materials there, what YouTube refers to as “watch time.” These recommendation systems have been highly effective at increasing usage, and thus the profitability, of the sites.” Gonzalez brief at 17.

should not be ignored, that an adverse decision would “undercut a central building block of the modern internet”<sup>71</sup> and lead to a digital dystopia:

Without algorithmic sorting, Google Search would display an unordered, spam-filled list of every website. Gmail would not be able to deprioritize spam. YouTube would play every video ever posted in one infinite sequence—the world’s worst TV channel. Westlaw would display every judicial decision chronologically without headnotes. Amazon would intermingle jackets with knives and handbags with toothbrushes.<sup>72</sup>

The case has drawn a very large number of amicus filings, including most prominently the Solicitor General’s on behalf of the United States. This brief, among others, argues that YouTube’s targeted recommendations for additional videos do not qualify for Section 230 immunity. “Properly construed, Section 230(c)(1) protects YouTube from ... liability for hosting or failing to remove ISIS-related content, but not for claims based on YouTube’s own conduct in designing and implementing its targeted-recommendation algorithms.”<sup>73</sup> “[B]ecause YouTube created the algorithms which determine which videos will be recommended to which users, the recommendations are bound up with YouTube’s platform-design choices.”<sup>74</sup>

The case is scheduled for oral argument February 21. In the normal course, a merits decision would be handed down by the end of June.<sup>75</sup>

Beyond litigation construing Section 230, there have been numerous proposals in the United States and Europe to ameliorate the externality problem in more systematic ways, including increasing liability exposure and mandating platform internal processes.

A wide variety of proposals for amending Section 230 have emerged in the United States. Some commentators have proposed that an interactive computer service should be able to avoid liability if it can demonstrate that it has taken certain precautions to avoid the dissemination of harmful content. For example, Professor Citron and Mr. Wittes have argued that “If providers or users engage in good-faith efforts to address or restrict abusive material, they should be immune

---

<sup>71</sup> Google brief at 5.

<sup>72</sup> *Id.* at 32.

<sup>73</sup> Brief for the United States as Amicus Curiae in Support of Vacatur, at 12. The brief, however, implies that the plaintiffs would have a difficult time successfully asserting an Antiterrorism Act claim. *Id.* at 32 & note 5.

<sup>74</sup> *Id.* at 28.

<sup>75</sup> The Supreme Court also will consider *Twitter v. Taamneh*, *supra* note 7, with similar underlying facts and brought under the Antiterrorism Act. Google notes that a decision in *Taamneh* could obviate any need to address the merits of *Gonzalez*. Google brief at 54-55.

from liability even if they were negligent or reckless in doing so. By contrast, the immunity should not apply to platforms designed to host illegality or sites that deliberately choose to host illegal content.”<sup>76</sup>

Others propose more forceful approaches. Professor Perault has argued, for example, that “legislators should pass new federal criminal laws that prohibit some of the most harmful forms of online speech, such as voter suppression and incitement to riot.”<sup>77</sup>

Proposals made by members of Congress tend to retain the framework of Section 230, to rely on individual effort to suppress negative externalities where possible and to accept government intervention only where it is believed individual initiative would be ineffective. The principal proposals offered in the last Congress fall into three main categories.

First, some proposals would expand the categories carved out from blanket immunity, thereby permitting victims greater scope to bring actions against the platforms conveying harmful content. A leading example is the SAFE TECH bill<sup>78</sup>, which carves out additional categories of harmful material from platform immunity, including civil rights, antitrust, cyberstalking, international human rights, and wrongful death. It eliminates platform immunity in connection with material the platform is paid to distribute or in which it otherwise has an economic interest. It also makes it easier for a plaintiff to secure a take down injunction in the case of material likely to create irreparable harm.

Second, some proposals would require increased transparency and commitment to abide by user-friendly procedures in the event of controversy. For example, the PACT Act<sup>79</sup> emphasizes platforms’ transparency in their content moderation practices and requires take down within four days of material found by a court to be illegal. It also confirms the authority of the Justice Department, FTC, and State Attorneys General to engage in civil enforcement of an amended Section 230.

---

<sup>76</sup> Citron and Wittes, *supra* note 35, at 417 (citation omitted).

<sup>77</sup> Perault, *supra* note 14, at 5.

<sup>78</sup> S. 299, 117<sup>th</sup> Congress, 1<sup>st</sup> Session, Bill “To amend Section 230 of the Communications Act of 1934 to reaffirm civil rights, victim rights, and consumer protections.”

<sup>79</sup> S.272, 117<sup>th</sup> Congress, 1<sup>st</sup> Session, Bill “To require transparency, accountability, and protections for consumers online”

Third, some proposals would condition immunity specifically on algorithmic reforms, reflecting concern about the amplification of material produced by the algorithms used by a digital service. One example that aligns with the claims advanced in *Force* and *Gonzalez* is the Protecting Americans from Dangerous Algorithms Act.<sup>80</sup> It would eliminate immunity for civil rights and international terrorism claims if a platform employs algorithms that amplify or recommend content related to the claims. The Filter Bubble Transparency Act, introduced in both the Senate and the House of Representatives, approaches the issue differently. It effectively requires covered digital platforms—essentially large ones—to make available “a service that provides a feed that is not amplified by user-specific data.”<sup>81</sup>

The European Union and the United States were in roughly similar positions with respect to open mic platform services at the turn of the 21<sup>st</sup> century. Section 230 protected platforms from most liability in the United States and the Electronic Commerce Directive of 2000<sup>82</sup> protected platforms from most liability in the European Union.

More than two decades later, with the accumulation of experience about the consequences of relieving platform operators from exposure to tort and other claims, things are different. The European Union has legislated significant changes in its approach and the United States has not.

The differences are not a reflection of differences in fundamental values. At a deep level, the two jurisdictions continue to share a commitment to freedom of expression and fundamental rights.<sup>83</sup> The divergence, rather, reflects differences in legal and government culture.

These differences are fundamental — differences in civil law and common law, differences in governing philosophies with respect to communal and individual responsibility,

---

<sup>80</sup> H.R. 2154, 117<sup>th</sup> Congress, 1<sup>st</sup> Session.

<sup>81</sup> S. 2024, 117<sup>th</sup> Congress, 1<sup>st</sup> Session and H.R. 5921, 117<sup>th</sup> Congress, 1<sup>st</sup> Session.

<sup>82</sup> Directive 2000/31/EC. Like Section 230, the eCommerce Directive did not require any content moderation effort by platforms. It provides for exemption from liability for the carriage of potentially harmful content so long as the platform is completely passive vis-à-vis third-party content. *Id.*, at Art. 12.

<sup>83</sup> The Digital Services Act, *supra* note 5 at Recital 52 reflects fundamental rights [that] include but are not limited to: for the recipients of the service, the right to freedom of expression and of information, the right to respect for private and family life, the right to protection of personal data, the right to non-discrimination and the right to an effective remedy; for the service providers, the freedom to conduct a business, including the freedom of contract; for parties affected by illegal content, the right to human dignity, the rights of the child, the right to protection of property, including intellectual property, and the right to non-discrimination.

differences in concomitant propensities to intervention or not--and also instantiated in the First Amendment.<sup>84</sup> The European approach utilizes increased state direction while the United States' approach relies more on individual initiative to suppress negative externalities. One dramatic manifestation of the differences is that Section 230 occupies approximately two pages in the U.S. Code and the Digital Services Act occupies 100 pages in the European Union's Official Journal.<sup>85</sup>

The Digital Services Act seeks to protect a strong commitment to freedom of expression<sup>86</sup> while simultaneously imposing an extensive series of responsibilities on the internal workings of the platforms. The responsibilities that are cumulative and increase with their size. Fundamental responsibilities apply to all intermediaries<sup>87</sup> and then scale up with respect to

---

<sup>84</sup> The Stigler Committee Report, *supra* note 9, at 192, makes a similar point:

Jurisdictions outside the US have adopted versions of Section 230, but none provides as much protection. In Europe, platforms have borne more liability and responsibility for removing illegal content. Under the European E-Commerce Directive, for example, intermediaries are exempt from liability for content they host so long as they “play a neutral, merely technical and passive role towards the hosted content.” Once they become aware that any hosted content is illegal, the intermediaries “need to remove it or disable access to it expeditiously.” Germany enacted the NetzDG law in 2018, enabling courts to fine social media companies with more than 2 million euros up to €50 million if they do not delete posts contravening German hate speech law within 24 hours of receiving a complaint or seven days in more ambiguous cases. There are a number of EU and member state proposals to hold platforms responsible not only for illegal content but also for harmful content and to impose a “duty of care” for managing content in the public’s interest. (citations omitted)

*See also* Alexandre de Stree, et al., “Liability of Online Hosting Platforms: Should exceptionalism end?,” CTR. ON REGUL. IN EUROPE (Sept. 2018), at 21-31. (“CERRE Report”). The CERRE Report was commissioned as part of the process leading to the Digital Services Act. It contains a useful summary of actions taken by the EU since the adoption of the e-commerce Directive. These actions include: (i) adopting rules that make it illegal for platforms to disseminate certain types of materials, including those involving child sexual abuse, terrorism, and hate speech and violence; (ii) providing guidance to platforms for the detection, removal, and prevention of illegal content; and (iii) promoting co- and self-regulation for the removal of some illegal material.

As Professor Gillespie notes, *supra* note 17, at 203, “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.”

<sup>85</sup> Official Journal of the European Union, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) 3-104 (October 27, 2022), consisting of 40 pages of Recitals and 60 pages covering 93 Articles. The Digital Services Act addresses more than content moderation—for example, obligations attaching to trading on eCommerce platforms. Art. 30. This description omits these other features of the DSA.

<sup>86</sup> The Act seeks to assure respect for a range of “fundamental rights guaranteed in the Charter of Fundamental Rights of the European Union (the ‘Charter’), in particular the freedom of expression and of information, the freedom to conduct a business, the right to non-discrimination and the attainment of a high level of consumer protection.” Recitation 3. Thus, the rights of both platform providers and platform users are addressed in the DSA. *See, e.g.*, Recitation 52. Due process considerations receive significant attention throughout the Act.

<sup>87</sup> Ch II and III, Section 1.

“hosts,” essentially online platforms disseminating information to the public,<sup>88</sup> and then again to the largest online platforms and search engines.<sup>89</sup>

The Act imposes extensive disclosure and transparency requirements. They begin with the terms and conditions, including disclosures relating to both algorithmic and human content moderation.<sup>90</sup> The Act also requires all intermediaries to file an annual public report describing any content moderation activities in which they engaged.<sup>91</sup> The Act requires the largest platforms and search engines to make these filings at six-month intervals.<sup>92</sup>

The DSA relies on transparency in other ways as well. There are substantial disclosure requirements concerning the source, sponsor, and related data about online advertising.<sup>93</sup> In the case of targeted advertising, an explanation of the algorithmic criteria determining the servicing of the ad must be included<sup>94</sup> and individuals must be given the opportunity to influence those criteria.<sup>95</sup>

An additional form of disclosure is applicable to very large platforms and search engines. They must make data they control available to “vetted researchers” for the limited but important purpose of contributing to the platforms’ due diligence responsibilities involving systemic risk assessment and associated mitigations.<sup>96</sup>

Recommender algorithms are subject to specific disclosure requirements. Providers are required to describe “the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters.”<sup>97</sup> They also are required to describe the criteria that caused the material to be recommended to a particular recipient and to enable the recipient to select alternative criteria.<sup>98</sup> The largest

---

<sup>88</sup> Art. 3 and Ch. III.

<sup>89</sup> Ch. III, Section 5.

<sup>90</sup> Art. 14.

<sup>91</sup> Arts. 15 & 24.

<sup>92</sup> Art. 42.

<sup>93</sup> Art. 39.

<sup>94</sup> *Id.*

<sup>95</sup> *Id.*

<sup>96</sup> Art. 40 and *infra* notes 103-111 and accompanying text.

<sup>97</sup> Art. 27.

<sup>98</sup> *Id.*

platforms have an additional obligation to make available a recommendation system that is not based on profiling of the recipient.<sup>99</sup>

The responsibilities include an elaborate notice-and-action requirement. The obligation facilitates individuals' reporting to the platform of assertedly illegal content and the platform's concomitant responsibility to assess the complaint and promptly take any action it finds to be warranted.<sup>100</sup> The platform is required to disclose its decision and, where it undertakes remediation of any sort—take down, demotion, etc.—to advise the content provider not only of the decision but also of its basis.<sup>101</sup> The action-and-notice regime relies on users' initiatives, but is supplemented by “trusted flaggers,” individuals or organizations with expertise appointed to identify suspect content and to report it to the relevant platform<sup>102</sup>

The notice-and-action provisions involve ex post content review. The Digital Services Act imposes additional requirements of an ex ante nature applicable to very large online platforms such as Google and Facebook in light of their importance “in facilitating public debate, economic transactions and the dissemination of information, opinions and ideas and in influencing how recipients obtain and communicate information online.”<sup>103</sup> The reach of these platforms presents “societal risk, different in scope and impact” from smaller platforms and thus could produce “disproportionately negative impact[s] in the Union.”<sup>104</sup> For that reason, very large online platforms and very large search engines should “assess the systemic risks stemming from the design, functioning and use of their services, as well as by potential misuses by the recipients of the service”<sup>105</sup> and “deploy the necessary means to diligently mitigate the systemic risks identified in the risk assessment.”<sup>106</sup> While the “systemic risks” that the largest platforms would be required to mitigate extend to harmful but not illegal activities — for example, “actual or foreseeable negative effects on civic discourse” and “serious negative consequences to the person’s physical and mental well-being “<sup>107</sup> — the DSA leaves the extent of harmful activity

---

<sup>99</sup> Art. 38. The DSA adopts the definition of “profiling” found in the General Data Protection Regulation. Art. 4 (4).

<sup>100</sup> Art. 16.

<sup>101</sup> *Id.*

<sup>102</sup> Art. 22.

<sup>103</sup> Art. 75.

<sup>104</sup> Art. 76.

<sup>105</sup> Recital 79. See also Art. 34.

<sup>106</sup> Recital 86.

<sup>107</sup> Art. 34.

largely undefined, something that would be problematic in United States jurisprudence delimited by the First Amendment. The largest platforms must conduct these assessments on an annual basis<sup>108</sup> and, where systemic risks have been identified, to engage in “reasonable, proportionate and effective” mitigation.<sup>109</sup> The risk assessments are reinforced by requirements for independent audits<sup>110</sup> and regular reports to the authorities on the mitigation efforts and audits results.<sup>111</sup>

Very large online platforms and search engines also must establish an internal compliance function to assure that risks “are identified and properly reported on and that reasonable, proportionate and effective risk mitigation measures are taken.”<sup>112</sup>

The use of “dark patterns” is prohibited.<sup>113</sup>

The Act encourages the Commission to engage standards setting bodies for the purpose of making use of and compliance with the various notice and disclosure provisions more efficient.<sup>114</sup> It also requires the Commission to encourage platforms to create, with the assistance of relevant stakeholders, voluntary codes of conduct “to contribute to the proper application of this Regulation” with specific reference to the “challenges of tackling different types of illegal content and systemic risks.”<sup>115</sup> The Act proposes that codes be available by 2025<sup>116</sup> and recommends the Code of conduct on countering illegal hate speech online of 2016 as a useful example.<sup>117</sup>

The DSA establishes non-governmental adjudication procedures, something that could be understood to remit disputes to more specialized and presumably faster, but non-binding review

---

<sup>108</sup> *Id.*

<sup>109</sup> Art. 35.

<sup>110</sup> Art. 37.

<sup>111</sup> Art. 42.

<sup>112</sup> Art. 41.

<sup>113</sup> Art. 25.

<sup>114</sup> Art 44.

<sup>115</sup> Art. 45.

<sup>116</sup> Recitations 103-107.

<sup>117</sup> Recitations 62, 87, 106. The 2016 Code is a three-page document agreed upon by some of the largest platform companies. It establishes expectations for internal processes and timetables for dealing with hate speech. The signatories’ performances pursuant to the Code are subject to annual review by the European Commission.

than the general court systems could afford.<sup>118</sup> Aggrieved individuals, however, also can seek redress in the national and EU courts for violations of the DSA.<sup>119</sup>

Oversight and enforcement of the DSA in the first instance is lodged in each national jurisdiction with a Digital Service Coordinator<sup>120</sup> endowed with very substantial powers for enforcement.<sup>121</sup> However, in the case of very large platforms and search engines, the European Commission is responsible for enforcement.<sup>122</sup> The Commission is given extensive investigatory authority, including the right to conduct dawn raids<sup>123</sup> and the right to impose interim measures (essentially injunctions *pendente lite*).<sup>124</sup> As with its other adjudicatory functions, it can accept and approve parties' commitments for purposes of settlement.<sup>125</sup> The penalty for noncompliance with the Act's requirements ultimately can amount to six percent of a firm's worldwide annual turnover.<sup>126</sup> Parties have the right to appeal adverse actions to the Court of Justice of the European Union.<sup>127</sup>

With the passage of the Digital Services Act the European Union and United States approaches to content moderation have diverged very substantially.

The principal manifestation of the differences is the European Digital Services Act's imposition of responsibilities on platforms, generally conditioning immunity on the adoption and adherence to detailed processes, both internal and external to the firm.

The Digital Services Act also increases government powers. And it continues a tradition of government intermediation between platform and citizen. European citizens who are adversely affected by the dissemination of harmful information have the opportunity to seek the

---

<sup>118</sup> Art. 21.

<sup>119</sup> *Id.* and Art. 54.

<sup>120</sup> Art. 49

<sup>121</sup> Art. 51.

<sup>122</sup> This contrasts with the experience of enforcement of the General Data Protection Regulation, which assigned enforcement responsibilities to national privacy authorities, the effectiveness of which has been severely criticized. *See, e.g.*, Brian Daigle and Mahnaz Khan, "The EU General Data Protection Regulation: An Analysis of Enforcement Trends by EU Data Protection Authorities." *Journal of International Commerce and Economics*. 20-21 (May 2020). (discussing Irish Data Protection Commission).

<sup>123</sup> Art. 69.

<sup>124</sup> Art. 70.

<sup>125</sup> Art. 71,

<sup>126</sup> Arts.52 and 74.

<sup>127</sup> Art.81.

assistance of government agencies with regulatory power and to seek redress for platforms' failures to meet DSA-imposed responsibilities.

The DSA does not define illegal content, but rather leaves the definition to existing, and of course changeable European Union and national laws.<sup>128</sup> It thus reflects the European Commission's and individual European nations' superior ability, compared to that of the United States, to declare harmful expression illegal and thus subject to take down as well as to encourage intermediaries to curtail harmful content without regard to its intrinsic illegality. This ability of the European Union and of the EU Member States to add to the categories of illegal content is obviously important in this context. To take a likely example, the transparency, self-assessment of risk, auditing, reporting and mitigation requirements will produce a great deal of salient information that in appropriate cases will translate into *malum prohibitum* content in some Member States if not the entire European Union.

While the eCommerce Directive's passivity rule granting liability immunity has been nominally preserved,<sup>129</sup> the DSA's addition of enforceable responsibilities and obligations has created a vastly different juridical environment. It has created something like a regulatory program, although admittedly an unusual one given the paucity of substantive rules.

The DSA's effects will be felt from early 2024, but the process of establishing a reasonably stable common understanding of its 109 Article requirements will take a very long time. The risk assessment and mitigation obligations constitute an obvious example of why this is so.

Relative to the European approaches, potential United States' efforts to suppress Internet-transmitted harmful content are significantly constrained by the combination of the First Amendment, the country's prevalent non-interventionist political/judicial culture, and the related inability to date to amend Section 230. Rather than direct or indirect prohibitions, the proposed responses tend to continue to rely on individual citizen initiative based on tort law. Inevitably that exposes potential revisions to platforms' content moderation cost-benefit calculations to the unavoidable exigencies of litigation: cost, delay, and uncertainty of outcome.

---

<sup>128</sup> Recital 17 and Arts. 3(h) and 9.

<sup>129</sup> Arts. 4-8.

That the most prominent content moderation bills introduced in the last Congress proposed incremental changes, it is likely that any near-term Section 230 amendments will be incremental. Eliminating the liability shield from interactive computer services for additional categories of harm would result in more actions being brought and at least some services adapting their content moderation practices in order to limit their carriage of harmful content. However, the severe limitations inherent in relying on individual citizens' legal actions to suppress the wide range of Internet-delivered harmful content will remain. For example, even if interactive computer services are liable, in many instances the number of persons or entities that are adversely affected by the carriage of specific harmful information is likely to be so large and the harmful effect on each is likely to be so small that no individual victim is likely to have the incentive to bring suit against the service.<sup>130</sup> These provide an impetus for seriously considering government enforcement, as *parens patriae* and otherwise, against the dissemination of certain types of harmful information that are not protected by the First Amendment.<sup>131</sup>

An alternative possibility finds an analogy in the experience with the GDPR. Just as that European Union approach to privacy protection, for good or ill, has fundamentally influenced similar initiatives around the world, including in the United States,<sup>132</sup> the Digital Services Act's emphasis on enhanced platform responsibility could prove to be a template for a revised legislative approach in the U.S.

---

<sup>130</sup> In at least some circumstances this complication could be ameliorated through the use of class actions. Rule 23, Fed. R. Civ. Pro.

<sup>131</sup> Cf. *California v. Amazon.Com, Inc*, Docket No. CGC22601826 (Cal. Super. Ct., Sep 15, 2022). (Parens patriae suit by California Attorney General alleging Cartwright Act and Unfair Competition Law violations.)

<sup>132</sup> See, e.g., California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.100-.199 (2018).