Free markets. Real solutions.

R STREET POLICY STUDY NO. 244
November 2021

PRAGMATIC PRINCIPLES FOR CONTENT POLICY AND GOVERNANCE

By Chris Riley

CONTENTS

INTRODUCTION

Over the past two years the United States has seen a flurry of legislative proposals to modify Section 230 of the Communications Act of 1934, the infamous immunity provision designed to allow the use of moderation for user-generated content shared online without introducing liability as a consequence of the resulting implicit knowledge and control. Yet, none of the proposals have advanced substantially in either the House or Senate. Meanwhile, the European Union (EU) has assumed the driver's seat in global internet policy yet again by introducing its proposal, the Digital Services Act (DSA), in December 2020.

However, the need for intervention persists. For example, the case of *Herrick v. Grindr* attempts to hold an online platform accountable for the real-life harm caused by users of its platform.[1] Another larger-scale example circulates around public feelings of bias by major social media and technology companies, which persist despite a lack of supporting data and evidence.[2] Trust has broken down online, and change is needed. Whether that change requires legislation remains an open question. If so, the more challenging question remains of how to design an intervention to deliver meaningful benefits and minimize harmful externalities. However, the status quo seems unsustainable, so Congress is actively engaged in holding hearings and introducing legislation.

While the underlying rationales for reform vary widely and lead to equally varied regulatory approaches, a few ideas appear to have emerged that offer the potential for broad (though not universal) appeal.[3] A future American law aiming to establish greater responsibilities for online intermediaries of user generated content seems likely to include timely compliance with duly issued court orders, incentives or requirements to publish content policies and mechanisms to hold companies liable for some types of procedural insufficiencies. In principle, these ideas align with the spirit of "consumer protection" and balanced intervention, which were proposed in the EU's DSA. This alignment serves as a sign for future transatlantic cooperation in internet policy.

However, these concepts only tell a portion of the story. No legislative proposal is likely to gain traction without coming to grips with the harder questions that motivate many active draft bills, including in particular how to calibrate incentives for the proper management of lawful but contextually (or universally) harmful content, and whether criminal law is sufficient to govern modern online harmful behavior.

1. Carrie Goldberg, "Herrick v. Grindr: Why Section 230 of the Communications Decency Act Must be Fixed," Lawfare, Aug. 14, 2019. https://www.lawfareblog.com/herrick-v-grindr-why-section-230-communications-decency-act-must-be-fixed.

2. See, e.g., Paul M. Barrett and J. Grant Sims, "False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives," NYU Stern Center for Business and Human Rights, February 2021. https://bhr.stern.nyu.edu/bias-report-release-page.

3. David Morar and Chris Riley, "A guide for conceptualizing the debate over Section 230," Brookings TechStream, Apr. 9, 2021. https://www.brookings.edu/techstream/a-guide-for-conceptualizing-the-debate-over-section-230.

The harms of online content today are multifaceted and complex, and no single law or policy change can address them in full. Yet while Americans debate and introduce countless scattered proposals derived from a smorgasbord of conflicting rationales, other countries race ahead leaving the United States further behind in its policy leadership. This paper offers principles and analysis to contribute to future legislative proposals that seek to offer pragmatic implementations of the widely held ideas noted above along with answers to the more difficult questions. The specific proposals for consideration in this paper are offered with the intention of catalyzing further and deeper study of their consequences over the coming months.

Critically, none of the proposals included in this paper include any modifications to Section 230 itself. While the law has become a lightning rod for political engagement, and the problems associated with harm online are significant, the governmental interventions suitable for making positive progress towards those problems need not be centered on that law.

As a key note of context, this paper is intended to focus solely and specifically on content policy decisions, and in particular does not propose changes to copyright law's existing notice and takedown systems. These are distinguishable on several grounds in principle, and while there are certainly some meaningful analogies or comparisons to be made between copyright infringement and the space of online harms, this paper does not seek to make them.

PROPOSAL

Four actionable principles outlined below present a substantial and meaningful legislative intervention that could incentivize private sector investment in responsible safeguards against online harm without imposing unwarranted cost—including on those parties least well positioned to engage in effective mitigation. The recommendations that these principles collectively lead to offer a granular framework readily suitable for translation into statutory language. The principles are:

1. Uphold, but do not privatize, the law
2. Protect consumers
3. Empower critical community
4. Target specific concerns with specific solutions

PRINCIPLE I: UPHOLD, BUT DO NOT PRIVATIZE, THE LAW

One of the major U.S. legislative proposals on content issues is the bipartisan Platform Accountability and Consumer Transparency Act (PACT) cosponsored by Sens. Schatz (D-HI) and Thune (R-SD).[4] Key provisions of the PACT Act mandate compliance with court orders to remove content and activity determined to be illegal, while making clear that platforms themselves are not obligated to make determinations regarding the legality of any content.[5] The EU's DSA proposal takes a similar approach toward illegal content: an intermediary's duty to take down illegal content is triggered through receipt of a standardized order from a Digital Services Coordinator (a newly defined government authority set up for DSA enforcement).[6]

In its approach to upholding the law, the PACT Act as introduced in 2021 does four things that seem worth including in future legislation on this topic:

1. **Court order standard:** The PACT Act adopts a court order standard, requiring a clear and well-structured determination of illegality by a court to trigger takedown obligations.[7] As Daphne Keller notes on the previous bill, the limitation of state law applicability to only defamation law may deserve further thought; however, further consideration of federal criminal law as proposed below in this paper may help close some gaps.[8]
2. **Time and flexibility for takedown:** The original PACT Act required takedowns within 24 hours of sufficient notice, which is an untenable threshold; the 2021 text expands this to four days, but further analysis is needed to determine the reality of such a timetable. The bill also permits "reasonable exceptions, including concerns about the legitimacy of the notice" to the four-day threshold. New language should provide clarity and specificity to service providers on means and rationales for claiming such an exception.[9]
3. **No monitoring obligation:** Like Europe's DSA proposal, the PACT Act includes clear language speci-

4. S.797, Platform Accountability and Consumer Transparency Act, 117th Congress (Hereafter: PACT Act).

5. Ibid., sec. 6(a).

6. COM/2020/825 final, Digital Services, Act, European Commission, article 8 (Hereafter: DSA).

7. PACT Act, sec. 6(a).

8. Daphne Keller, "CDA 230 reform grows up: the PACT Act has problems, but it's talking about the right things," Center for Internet and Society of Stanford Law School, July 16, 2020. https://cyberlaw.stanford.edu/blog/2020/07/cda-230-reform-grows-pact-act-has-problems-it%E2%80%99s-talking-about-right-things.

9. PACT Act, sec. 5(c)(1)(A)(i).

fying that intermediaries are not required by law to take affirmative steps to identify potentially illegal content. Rather, the obligations of intermediaries are limited to reactive responses that are triggered by the provision of external identification of illegality.[10]

4. **Infrastructure exemption:** The PACT Act includes clear language exempting non-user-facing providers of infrastructure from the bill's transparency and process obligations, including a broad range of services from web-hosting to cloud services.[11]

Study the State of Federal Criminal Law

Federal criminal law is broadly exempted from Section 230's immunity, but whether federal criminal law itself is up to the task for modern online crime is an open and legitimate question. There may be gaps that limit law enforcement's ability to identify and penalize online harm in ways that deserve remedy, and while this is a cross-cutting issue that goes beyond the core of Section 230, some understandably combine it with Section 230 proposals. For example, the Perault proposal, suggests amending criminal law in the context of Section 230 reform by creating new prohibitions related to voting disinformation to target specific vectors of online harm.[12]

Rather than modifying federal criminal law directly, legislation should include guidance and resources for the Department of Justice to study the scope of protection in federal criminal law in the context of online harm, along the lines of the limitations raised by the Perault proposal, but through a holistic view. A constructive approach to such inquiry would focus on the principals engaging in activity rather than the platform's contributory activities, and in particular whether their actions ought to be considered criminal. Where federal criminal behavior takes place, platform responsibility becomes legally relevant through the existing carve-out within Section 230 for federal crimes. Further research and analysis, along with separation of the legislative vehicles for addressing platform practices and the underlying legality of content and user behavior, seems warranted and suitable for addressing such complex issues. In practice, there remains a substantial difference between the possibility of justice through effectively designed law and the reality of equitable access to justice for victims of harm; however, this challenge lies well beyond the scope of this paper.

Separately, the Protecting Americans from Dangerous Algorithms Act (PADAA) of Reps. Malinowski (D-NJ) and Eshoo (D-CA) would waive Section 230 immunity where a large platform uses algorithmic amplification and is accused in a civil action of liability under three specific sections of law related to foreign terrorism, domestic extremism and threats to civil rights.[13] As these particular harms have a clear nexus to potential criminal law, rather than creating greater civil liability, the aforementioned DOJ study should consider these sections of law as well, and evaluate whether criminal liability is sufficient in the online context or whether there are meaningful gaps in current law leading to skewed incentives not to take greater action to limit the algorithmic amplification of harmful content.

Do Not Privatize Legal Determinations Through Rigid Processes

Perceived deficiencies in the court order standard as a centerpiece for private sector compliance come not just in its scope, but also in its speed (e.g. the four-day compliance window of the PACT Act). Thus, many proposals identify non-judicial, non-legislative process shortcuts to develop minimum standards of responsible conduct above and beyond compliance with court orders. Such proposals result in the effective privatization of governmental legal functions in highly problematic ways.

Facebook's proposal, as laid out in Congressional testimony, includes a third party evaluation of the adequacy of their platform systems for identifying and removing unlawful conduct.[14] The EARN IT Act by Sens. Graham (R-SC), Blumenthal (D-CT), and many others addresses the same high-level problem by designating a commission of mixed government and non-government stakeholders to set reasonability standards.[15] The chosen commission must recommend best practices by a certain deadline to Congress, which then enacts said practices into law through accelerated procedures. While the two proposals differ in many respects, both will put onus on a company to make proactive determinations about the legality of content and behavior online as a way of appearing more effective at identifying and removing unlawful content. Furthermore, both assume the ability to reach agreement on very difficult procedural balancing questions, a fragile assumption that is likely to fail in practice.

It is not feasible to make behavioral obligations fit moving targets when it comes to questions of sufficiency of process in speech regulation. Such regulation need not be forced, as the proposed DSA in Europe demonstrates by including

10. Ibid., sec. 6(a).

11. Ibid., sec 5(f).

12. Matt Perault, "Section 230: A Reform Agenda for the Next Administration," Day One Project, Oct. 26, 2020, p. 5. https://www.dayoneproject.org/post/section-230-reform.

13. H.R. 8636, Protecting Americans from Dangerous Algorithms Act, 116th Congress.

14. Testimony of Mark Zuckerberg, House Subcommittees on Consumer Protection & Commerce and Communications & Technology, Committee on Energy and Commerce, "Testimony of Mark Zuckerberg, Facebook, Inc.," 117th Congress, March 25, 2021. https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-U1.pdf.

15. S.3398, EARN IT Act of 2020, 116th Congress.

heightened obligations for "very large online platforms," under which platforms must conduct risk assessments, take reasonable and effective measures to mitigate risks through process improvements, and submit themselves to external and independent audits.[16] This combination of meaningful obligations preserves a focus on responsible processes as developed and evaluated by non-governmental entities, while drawing a clear line short of setting specific obligations through government action. Rather than privatize governmental determinations of legality, the better approach is to uphold duly protected legal processes and set higher bars for good behavior through consumer protection and community empowerment models, as the next two principles will articulate in more detail.

Proposals for Consideration:

- Codify the court order standard as the necessary and sufficient private sector compliance for content takedowns.
- Codify compliance with court orders through four-day windows to take down content, with clarified exceptions language.
- Incorporate in statute that no general monitoring for potential illegality is or can be required of platforms, following the example of Europe's E-Commerce Directive.
- Exempt infrastructure services from the scope of content transparency and process obligations (as proposed in this principle and in the next).
- Authorize and resource a Department of Justice study on the scope and effectiveness of criminal law for modern day online harm, seeking to identify gaps to close through future legislation.
- State the intention of legislation to avoid the creation of incentives for private determination of legality of online activity, a fundamentally governmental authority.

PRINCIPLE 2: PROTECT CONSUMERS

Section 230-related reform aims far beyond setting processes and standards regarding the handling of content or behavior that is illegal in any particular legal jurisdiction. For such "lawful but harmful" content, the policies of service providers serve as the primary rule. Where moderators interpreting such policies implement blocking or other differential treatment, the platform's consumers are affected and their perception of the fairness of such action depends largely on the policies and processes offered by the service provider. Many Section 230 reform proposals seek to change that calculus by mandating the adoption and disclosure of some level of content policies and attendant processes to ensure platform consumers are protected.

Both the Digital Services Act and the PACT Act embrace the frame of consumer protection as central to their respective regulatory approaches.[17] Under a consumer protection approach, the role of government is to ensure transparent content policies and processes are followed, not to second-guess their substance or interpretation. In particular, under this philosophy, courts and government functions are not meant to make independent judgments regarding the interpretation of content policy.

Four pieces are critical for effective consumer protection in the context of moderating lawful content online:

1. **Transparent content policies:** Intermediaries must have content policies providing guidance on acceptable content and behavior, and must make these content policies readily available. Content policies should provide sufficient information to be useful for the intermediary's users, however there is today no clear, shared standard for what constitutes sufficiency. Such policies should include some disclosure of triggers that generate deprioritizing or other forms of non-organic treatment of content or accounts, in addition to blocking. At the same time, content policies may be described with some generality to ensure forward-looking agility and resist gamification and abuse.
2. **Notification for affected users:** When content or accounts are affected by targeted, reactive actions to enforce platform content policies, the affected users must be notified, consistent with the guidelines of the Santa Clara Principles.[18] In particular, such notice should include enough specificity of detail to allow a user to take corrective measures and learn from the action to improve the quality of future engagement.
3. **Meaningful appeals mechanisms:** Users whose content or accounts are affected by content policy driven mitigation must be provided with some mechanism for appealing to the platform service provider, and should be informed of such mechanism in the

16. See, e.g., DSA, article 25; Ibid., article 26; Ibid., article 27, article 28.

17. Ibid.; Office of Sen. Thune, "Thune, Schatz Introduce Legislation to Update Section 230, Strengthen Rules, Transparency on Online Content Moderation, Hold Internet Companies Accountable for Moderation Practices," Press Release, June 24, 2020. https://www.thune.senate.gov/public/index.cfm/2020/6/thune-schatz-introduce-legislation-to-update-section-230-strengthen-rules-transparency-on-online-content-moderation-hold-internet-companies-accountable-for-moderation-practices.

18. "The Santa Clara Principles on Transparency and Accountability in Content Moderation," Santa Clara Principles, last accessed June 23, 2021. https://santaclaraprinciples.org.

notification of action, consistent with the guidelines of the Santa Clara Principles.

4. **Legal consequences for procedural, not substantive, deficiencies:** Content-based judgments should be managed entirely by the platform itself, and not a court or other government actor or agency. When a company does not comply with its own procedural obligations, some form of government intervention and redress may be appropriate, particularly if noncompliance is systemic. However, deference to the company should be given regarding the substantive interpretation of a company's content policies.

The distinction between permissible legal review of alleged procedural deficiencies and deference on alleged substantive deficiencies parallels the "don't privatize the law" disclaimer of the first principle. The proposed framework here essentially seeks to keep companies and governments in their own lanes of expertise with the government authorized to determine legality, and companies authorized to interpret and apply their own policies and services. In the context of this principle, where courts, legislatures, or regulators substitute their own judgment for that of a company in the interpretation of the company's own content policies, the governance framework has gone beyond the boundaries of "consumer protection" and entered directly the territory of speech regulation.

The Federal Trade Commission (FTC) can serve as the governmental backstop for consumer protection issues arising from content policy and practice disputes given its active engagement with company privacy policies. The FTC is fully capable of evaluating structural and procedural questions, such as whether a company's promises as articulated in its terms of service are implemented through correct procedures. However, walking the fine line between consumer protection and speech regulation requires clear guardrails on FTC review to prevent the agency from inserting its own judgment in place of a company's in the interpretation of the company's own policy.

One consequence of this separation of powers is that government actors would lack the power to set any particular standards with regards to lawful but harmful content. The status quo is insufficient in this regard, and better standards are needed. A more effective and less harmful mechanism for continuous ratcheting up of reasonability lies not in government action or other forms of hard law, but rather in the close engagement of an expert critical community and other levers for pressure, which will be discussed in the next principle of this framework.[19]

19. Chris Riley, "The need for a robust critical community in content policy," *Medium*, Sep. 25, 2020. https://mchrisriley.medium.com/the-need-for-a-robust-critical-community-in-content-policy-7572679d008c.

Consumer protection solves two distinct problems with content moderation. First, clarifying expectations and processes around how content policies are interpreted helps protect consumers directly. And second, transparency helps to enable and catalyze the effective functioning of the critical community, which then develops ever-evolving standards of responsibility and holds platforms to account where they fall short. Without sufficient transparency in content policies and processes, the critical community cannot know the full extent of what platforms are doing in practice, and without the assurance of procedural implementation the critical community cannot rely on information that is voluntarily provided.

Both the PACT Act and the Digital Services Act adopt their consumer protection rules as standalone obligations for platforms, rather than implementing them as requirements for the receipt of immunity under Section 230. For any mandates related to consumer protection, such an approach makes the most sense, as there is no inherent relationship between the benefits of these obligations and the specific context of civil immunity.

The question remains whether and to what extent a mandate for disclosure is in fact necessary, and how such a mandate could be designed when the nature of content policies and practices must change (sometimes rapidly) and include some amount of generality. In practice, services of sufficient scale already offer content policies and appeals mechanisms, although there is always room for improvement and not all stakeholders believe current levels of disclosure are sufficient to provide clarity for users.[20] While smaller services may not meet the same standards, the increasing professionalization of the trust and safety space will help build better practices.[21] Furthermore, given the vast diversity of services and content policies, and the utility and benefits of continuing to encourage diversity rather than to constrain it, the design of a legislative mandate to apply to both current and future intermediaries of user-generated content would be difficult.

However, state law continues to move forward, and the passage of one or more state bills may change both the political calculus and the policy impact of adopting a federal mandate. In particular, California's AB-587 bill includes specific requirements for changes to the terms of service and content moderation practices to be included within a company's public terms of service as well as obligatory quarterly

20. See, e.g., Kara Frederick, "Steven Crowder Is Suing YouTube Over Vague Rules, but It's Not Just About Him," The Heritage Foundation, May 21, 2021. https://www.heritage.org/technology/commentary/steven-crowder-suing-youtube-over-vague-rules-its-not-just-about-him.

21. See, e.g., "About us," The Trust & Safety Professional Association, last accessed, June 23, 2021. https://www.tspa.info.

reports.[22] AB-587 notably limits its scope to companies with over $100 million in annual gross revenue, following the lead of the DSA which defines "Very Large Online Platforms" and subjects them to its highest regulatory obstacles.[23] Should AB-587 or a similar law pass, there may be added value in a federal disclosure mandate to create consistency across states through policy action and preemption.

In the interim, as the Federal Trade Commission separately considers the scale of its current rulemaking authority, a component of such inquiry should be whether it would be both feasible and desirable to undertake a rulemaking process on the current state of content policy disclosure practices. These considerations would need to include the specificity and clarity of *ex ante* content policies as well as the procedural completeness of associated processes including notifications and appeal mechanisms.

Proposals for Consideration:

- Provide resources to the FTC sufficient for the agency to use its existing authority regarding unfair and deceptive practices to receive complaints of procedural deficiencies with respect to its publicly disclosed practices; pair such resources with guardrails limiting FTC action solely to procedural deficiencies, and barring the FTC from engaging in the substantive interpretation of content policy.
- Encourage the FTC to conduct an inquiry into the current state of content policy disclosure practices and determine if further action such as a rulemaking proceeding is desirable to calibrate disclosure obligations.

PRINCIPLE 3: EMPOWER CRITICAL COMMUNITY

There are a number of mechanisms that can hold platforms accountable for taking meaningful steps to prevent harm that do not center around a government setting standards for responsible behavior. Soft law forces—including voluntary best practices and codes of conduct, as well as normative pressure brought about through public campaigns—provide effective levers in many contexts through a robust critical community of independent civil society advocates, researchers and other stakeholders. In the nuanced and fluid context of online content control, critical community as an exercise of soft law is not inherently less effective than hard law. Rather, it builds flexible, adaptive, and evolving pressure towards responsibility and accountability that can set and enforce higher and more effective bars for responsible behavior than any terminal legislation, particularly in a country with as many limitations on governmental authority as the United States.

Within the framework that this paper proposes, soft law structures are natural and necessary complements to the hard law associated with court order compliance and consumer protection obligations. The most agile respondents to malign private sector behavior are not government agencies or courts, which are (rightly) slowed with procedural safeguards and other limitations. Instead, the critical community surrounding industry remains vigilant with every change in content policy and every visible outcome of a private sector decision, armed in an ideal state with sufficient resources and expertise to offer real-time feedback on specific corporate practices. Regardless of one's threat model assumptions with regard to the private sector, such a critical community serves immense value. For hostile actors or actions, a critical community serves as a watchdog function to corral and wield public opprobrium; for well-meaning but mistaken contexts, a critical community helps companies see around their own blind spots and craft more inclusive and effective policies and procedures.

Investment in a robust critical community offers a different vision of external oversight compared to one of the few experimental structures developed and tested at scale: the Oversight Board designed and funded by Facebook. The Oversight Board was built to be both independent and final, in that Facebook committed both to supporting the Board financially without retaining influence, and to abiding by decisions made by the Board. The inclusion of global and diverse board members helps bring new perspectives to complex decisions, similar to a critical community. However, the Board's sheen of formal authority comes at the expense of substantial time and political cost to build and execute it compared to a more scalable, fully independent critical community.

As described here, a critical community is fundamentally non-governmental in nature. Yet its success at oversight is highly dependent on sufficient insight into private sector practices, which can be improved through properly designed transparency and accountability mandates adopted under the auspices of consumer protection. Influence over private sector practices comes from a combination of insight and messaging, bringing public opprobrium that can harm reputation with consequences for business relationships and customer retention as well as the possibility of future direct legal and legislative consequences.[24]

One further dimension of added value in empowering critical community for normative development and influence is the ability to bring in perspectives from underrepresented com-

22. AB-587, Social media companies: terms of service, California Legislature 2021-2022 regular session.

23. DSA, article 25.

24. See, e.g., "Stop Hate for Profit," Stop Hate for Profit Campaign, last accessed June 23, 2021. https://www.stophateforprofit.org.

munities and new voices who may face barriers in the normal course of engagement with technology companies and even governmental processes. Diverse contributions enrich policy conversations immensely and help well-meaning actors see around their own blind spots, with the potential for immediate benefit and better collective long-term outcomes.

Two affirmative steps can be taken within a content regulatory framework to facilitate the emergence and effectiveness of critical community:

1. **Kickstart community through multi-stakeholder processes**: Congress can authorize and fund the execution of a multi-stakeholder process to better understand the problems inherent in online content facilitation. This will act as an accelerant for community awareness and productivity, and help develop roadmaps and frameworks for improvement.[25] A key discussion within that process should be the appropriate disclosure of factors used in recommendation engines as well as learnings from changes made to such systems to mitigate harm.
2. **Fund research and beneficial community activity**: The United States government, through the National Science Foundation among other agencies, provides substantial funding for basic science and research that leads to the development of powerful technologies. To compel further work robust support is needed for appropriate research into the real-world effects of technology systems.[26] Such research may face challenges with at scale access to relevant data and systems, and potentially with negotiations for greater access as a consequence of modern privacy law. These challenges may be overcome with time, or some form of intervention may be needed to reach the full benefits of funding.[27]

Hard law interventions designed to calibrate private sector practices for online content come in two broad forms: specifically mandated practices, and substantive limits to the use of Section 230. Specific behavioral mandates are difficult if not impossible in practice because of the diversity of online content contexts (making such remedies limited and/or unscalable) and the challenges of constitutionality (and politics) in legislating limits on lawful activity including speech.

Limiting the use of Section 230 in certain circumstances defaults the review of law to a case-by-case court review of potential contributions by online platforms to liability where injury has occurred. The EARN IT Act takes a slightly different approach through its establishment of a "Commission" consisting of high-ranking political appointees.[28] In practice, neither courts nor commissions are designed to evolve over time or to be inclusive in the same way as a critical community; thus, hard law runs the risk of having a continually incomplete lens for review.

It is understandably tempting to assume that courts will be reasonable or that Congress will legislate to ensure a standard of reasonability applies to their review of private sector behavior.[29] Courts then become empowered to determine whether a platform's behavior has sufficient centrality to an end user's harmful action, and whether a platform's normal content and service precautions against possible future harm were sufficient. Mistakes will be made in such a balancing exercise, resulting both in actual harm and in private sector chilling effects resulting in conservativism rather than the desired responsiveness.[30] This risk is greater than usual when reaching the right balance requires a detailed understanding of the underlying technology and its functionality, and it is greater still in contexts where the technologies and practices of online content management change frequently and rapidly. Unfortunately, these two amplifying conditions are almost always applicable in practice. Simply put, an expert critical community can adapt at pace with tech, but courts and caselaw cannot.

The central spirit of Section 230 is the shifting of responsibility for rapid reaction to online harm away from court systems and processes to the private sector, because companies can respond more quickly to moderate their systems than courts can evolve standards for responsibility over time through the development of caselaw. While legislative processes and commissions may mitigate some of the delay and limitations of common law, that small improvement comes at the expense of greater risk of politicization and procedural breakdowns.

Online harm is very real, and the challenge for policymakers is how to align incentives to encourage the agile and responsive use of all possible forms of mitigation in transparent and responsive ways. But introducing substantive second-guessing of careful speech balances through the heavy influence of courts or commissions is not the answer.

25. Emily Birnbaum, "Commerce Department nominee advocates for Section 230 reform," *Protocol*, Jan. 26, 2021. https://www.protocol.com/bulletins/gina-raimondo-section-230-reform.

26. See, e.g., "Designing Accountable Software Systems (DASS) Program Solicitation" National Science Foundation, April 19, 2021. https://www.nsf.gov/pubs/2021/nsf21554/nsf21554.htm.

27. See, e.g., Nathaniel Persily and Joshua A. Tucker, eds., *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge University Press, 2020), pp. 313-14.

28. S.3398, EARN IT Act, 116th Congress.

29. Danielle Keats Citron and Benjamin Wittes, "The Internet Will Not Break: Denying Bad Samaritans §230 Immunity," *Fordham Law Review* 86:401 (2017). https://ir.lawnet.fordham.edu/flr/vol86/iss2/3.

30. Eric Goldman, "Content Moderation Remedies," *SSRN*, March 21, 2021. https://ssrn.com/abstract=3810580.

Proposals for Consideration:

- Authorize and resource the National Telecommunications and Information Administration to convene multi-stakeholder discussions regarding the state of play of content moderation and management online, to include private sector, public interest, academic and government actors.
- Invest in research through the National Science Foundation to better understand the real-world effects of internet technologies that intermediate human social and economic activity online.
- State the intention of legislation to preserve the role of soft law in setting evolving standards of responsibility for online content management, and to avoid the use of common law or political processes to set rigid standards of sufficiency.

PRINCIPLE 4: TARGET SPECIFIC CONCERNS WITH SPECIFIC SOLUTIONS

Many legislative proposals specify certain crimes to create carve-outs from the immunity provisions of Section 230.[31] By and large, these proposals are designed to capture organic, unpaid speech, increasing the potential surface area for liability to virtually all online activity. These are politically advantageous proposals as they focus on the breadth and/or severity of the crimes involved, such as cyber-stalking or civil rights violations, and the proposals build on existing carve-outs for federal crimes like sex trafficking and intellectual property violations. However, many calls for carve-outs include broadly defined categories of harm, including incitement to violence, hate speech, and disinformation. Given that the entirety of Section 230 applies only in cases where credible injury or harm can be alleged, it's unclear exactly what kinds of injury or harm are considered sufficiently innocuous as not to require a carve-out.

Individual crime-specific carve-outs to Section 230 make the underlying issue a question of speech itself, which is not well-calibrated within the context of the delicate balance of that law. Furthermore, modifications to Section 230 are often not necessary nor the best means of achieving the objectives of the carve-outs. Limiting the immunity of Section 230 operates by shifting the burden of developing a standard of reasonable behavior to common law courts systems, and thus strategically helps address the inherent inability of legislation to determine *ex ante* what reasonable behavior should be. But, given the unpredictability, inefficiency, and delays involved with deferring such development of standard to common law, limiting Section 230 immunity becomes an extraordinarily ineffective and poorly tailored mechanism whenever the goal of Congress is to mandate private sector behavioral outcomes where the nature of the better behavior sought is already reasonably determinable.

For example, Senator Manchin (D-WV) proposed the "See Something Say Something Online Act of 2020" with the objective of requiring internet services to submit reports of potential major crimes to the Department of Justice.[32] Failure to submit such reports triggers a waiver of immunity related to specific content that should have been reported.[33] The bill's restraint is commendable, in that it associates the loss of liability specifically to the suspicious content itself and not to the platform's full operations. Regardless, the incorporation of Section 230 is unnecessary here, although the political connection between the two issues is pertinent.

As a policy matter, the precise scope of "See Something Say Something" deserves further reflection. The private sector has ample incentive to support the identification of major criminal activity given that Section 230 provides no protection from federal criminal liability. In similar circumstances of largely shared incentives, Congress adopted the Cybersecurity Information Sharing Act of 2015 after substantial dialogue between government agencies and the private sector to better calibrate the mechanisms and safeguards associated with sharing information on active cybersecurity threats.[34] While that bill was criticized by privacy advocates and is not cited as an endorsement, its model of beginning from a position of assumed trust and cooperation would help in this instance.[35] It would be worthwhile to legislatively direct the Department of Justice and other federal agencies to undertake dialogue with stakeholders in industry and civil society to better explore mechanisms for trusted reporting of suspicious content online, above and beyond the now-established landscape of cybersecurity threat reporting.

In some circumstances, the potential harm motivating proposed changes to Section 230 may not be best addressed through actions by the private sector at all. For example, child safety issues remain a key source of anti-tech advocacy. Yet an alternative and compelling approach to increasing child protection online is to identify the resources and authority needed by law enforcement. Senator Wyden (D-OR) has made just such a proposal with the Child Safety Act, which would quadruple the number of prosecutors in the applicable division of the Department of Justice and substantially increase resources for other task forces and centers such as

31. See, e.g., S.299, SAFE TECH Act, 117th Congress.

32. S.27, See Something, Say Something Online Act, 117th Congress.

33. Ibid., sec. 5.

34. S.754, To improve cybersecurity in the United States through enhanced sharing of information about cybersecurity threats, and for other purposes, 114th Congress. (adopted through inclusion in its entirety within a budget bill in December 2015).

35. See, e.g., Lee Tien, "EFF Strongly Opposes CISA Cyber Surveillance Bill and CFAA Amendment," Electronic Frontier Foundation, Oct. 22, 2015. https://www.eff.org/deeplinks/2015/10/eff-strongly-oppose-cisa-cyber-surveillance-bill-and-cfaa-amendment.

the National Center for Missing and Exploited Children which works in partnership with technology companies to evaluate online harm.[36]

There are plenty of other proposals for legislative change that relate to the diverse perceptions of unchecked harm online.[37] One notable example is that federal privacy legislation remains unadopted at the time of this writing despite years of debate. That single, overdue step would help balance the power scales substantially and give regulators powerful new tools to limit harm online.

Proposals for Consideration:

- Authorize and resource the Department of Justice to convene stakeholders on mechanisms for the trusted reporting of suspicious online content and activity.
- Pass the Child Safety Act to provide proper support to law enforcement.
- Pass federal privacy legislation.

CONCLUSION

This paper suggests an affirmative framework for government action to make progress on the challenging problems of online harm, with minimal undue harm from regulatory overreach. Its four principles offer a roadmap for potential legislative development: uphold, but do not privatize, the law; protect consumers; empower critical community; and target specific concerns with specific solutions. The proposals for consideration translate these principles into actions, while distinguishing at length alternative approaches which for various reasons reflect a poorer balance of outcomes.

Regulating speech and speech-related activity online is complex and not to be taken lightly. Each of the proposals presented in this paper merits deeper research and broader perspectives on potential effect from a range of stakeholders. Nevertheless, Congressional action is indicated in due course, for both endogenous reasons that reflect the scale and perception of continued harm, as well as exogenous reasons that underscore the importance of re-establishing U.S. normative leadership in internet governance. This roadmap and analysis will help advance more constructive and collaborative discussions toward that end.

ABOUT THE AUTHOR

Chris Riley is a senior fellow of Internet Governance at R Street Institute. Prior to joining R Street, Chris led global public policy work for the Mozilla Corporation, and worked in the U.S. Department of State to help manage their Internet Freedom grants portfolio.

36. S.3629, Invest in Child Safety Act, 116th Congress.

37. Morar and Riley. https://www.brookings.edu/techstream/a-guide-for-conceptualizing-the-debate-over-section-230.