TikTok the Tortfeasor: A Framework to Discuss Social-Platform Externalities and Arguments Favoring Ex Ante Mitigations

Karan Lala†

In recent years, social media platforms have grown increasingly complex in how they invite, intersect with, and influence third-party speech. This complexity lies in stark contrast to the simplicity of the statute that governs those very platforms: Section 230 of the Communications Decency Act. Although Section 230 has cabined liability for platforms in the past, some have advocated for activist judges to deploy tools available to them to hold platforms accountable and mitigate harm to users as research documenting the negative impact of social products on user well-being has matured.

This Comment reviews Section 230 jurisprudence to develop a novel taxonomy that explicates the statute's boundaries and provides both an opening for ex post liability and a rough metric for its limits. It divides claims against platforms into three categories—content specific, content dependent, and content agnostic—based on the proximity of the alleged injury to user-generated content and the degree of the platform's participation. Noting the incentive for plaintiffs to frame claims as content agnostic to evade premature dismissal under Section 230, this Comment also formalizes a remedies test that courts can use to distinguish legitimate content-agnostic claims from those in name only. Armed with this vocabulary, this Comment turns its attention to a number of cases pending against social platforms. Applying the remedies test, it determines that a handful of pending allegations give rise to legitimate content-agnostic claims.

Noting that content-agnostic injuries are material but not yet fully understood, this Comment ultimately argues that an ex ante regulatory regime operationalized by an expert agency is better suited to address social-platform externalities than an ex post liability regime. It discusses several reasons to disfavor an ex post regime or favor an ex ante regime before outlining what an adequate ex ante regulatory regime could look like with respect to its mandate, powers, structure, and staffing.

† B.S. 2018, University of California, San Diego; J.D. Candidate 2025, The University of Chicago Law School. I would like to thank Professor Omri Ben-Shahar and the editors and staff of the *University of Chicago Law Review* for their thoughtful advice and support.

INTRODUCTION

Public perception of social platforms and their impact on society has radically shifted over the last two decades. Buoyant optimism about platforms like Facebook and Twitter serving as digital conduits to bring global communities together gave way to a vision of social platforms as untrustworthy monoliths.[1] Many now openly worry about the consequences of allowing "a billionaires boys' club" to exert "greater influence over the flow of information than has been possible in human history."[2]

[1] *See* Sean Kates, Jonathan Ladd & Joshua A. Tucker, *How Americans' Confidence in Technology Firms Has Dropped*, BROOKINGS INST. (June 14, 2023), https://perma.cc/4SL8-69KJ.

[2] Michael Scherer & Sarah Ellison, *How a Billionaires Boys' Club Came to Dominate the Public Square*, WASH. POST (May 1, 2022), https://perma.cc/G3V6-C2HQ (quoting political scientist Brendan Nyhan); *see also* Alex Krasodomski, *Democratic Infrastructure Must Not Be Owned by the Few*, CHATHAM HOUSE (Nov. 1, 2022), https://www.chathamhouse.org/2022/11/democratic-infrastructure-must-not-be-owned-few.

What changed? On one hand, quite a lot changed. Social platforms are more complex[3] and more integrated into day-to-day life than ever before.[4] On the other hand, much stayed the same. The same regulatory framework—Section 230 of the Communications Decency Act of 1996[5] (CDA)—that governed public message boards has also governed AOL Instant Messenger, Myspace, Facebook, Snapchat, and TikTok. By broadly immunizing "online service providers from legal liability stemming from content created by the users of their services,"[6] the twenty-six words in § 230(c)(1)[7] have "created the Internet that we know today."[8]

As social technologies have evolved and the body of research documenting the negative impact of social products on user health and well-being has grown,[9] critics have argued that this regulatory framework has proven inflexible and insufficient to meet the moment.[10] Courts, in response, have been forced to contort doctrine[11] to walk the line between promoting freedom of expression and protecting users from harm without inadvertently

[3] Susan Etlinger, *What's So Difficult About Social Media Platform Governance?*, *in* CTR. FOR INT'L GOVERNANCE INNOVATION, MODELS FOR PLATFORM GOVERNANCE 20, 22 (2019).

[4] Summer Allen, *Social Media's Growing Impact on Our Lives*, AM. PSYCH. ASS'N (Sept. 20, 2019), https://perma.cc/R8PJ-2RPA.

[5] 47 U.S.C. § 230.

[6] KATHLEEN A. RUANE, CONG. RSCH. SERV., LSB10082, HOW BROAD A SHIELD? A BRIEF OVERVIEW OF SECTION 230 OF THE COMMUNICATIONS DECENCY ACT 1 (2018).

[7] "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." 47 U.S.C. § 230(c)(1).

[8] JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 8 (2019).

[9] *See infra* notes 44–46, 96–97, and accompanying text.

[10] *See infra* note 121.

[11] Some courts have emphasized the impact the specter of litigation has on free speech when interpreting § 230. *See, e.g.*, Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 23 (1st Cir. 2016) (explaining that § 230 "reflect[s] a legislative judgment that it is best to avoid the potentially chilling effects that private civil actions might have on internet free speech"). In contrast, some scholars have argued that by "singl[ing] out the promotion of 'freedom of speech in the new and burgeoning Internet medium' as Congress's overriding purpose in enacting Section 230," courts have made a fundamental mistake because "a central if not primary goal of the bill was to encourage the removal of 'indecent' content online." Alan Z. Rozenshtein, *Interpreting the Ambiguities of Section 230*, BROOKINGS INST. (Oct. 26, 2023), https://perma.cc/P8CD-KCTJ (quoting Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997). These critics have noted that the text making up "Section 230 was not enacted as part of the 'Internet Freedom and Family Empowerment Act'" as originally proposed by the bill's authors—"[r]ather, it became law as part of the 'Communications Decency Act.'" *Id.* This title, sitting "alongside the 'freedom of speech' language" in the text of the statute itself, creates "a puzzling if not outright contradictory text for courts to grapple with." *Id.*

destabilizing the technology industry, creating "immense uncertainty[,] and flood[ing] lower courts with years of litigation."[12] These contortions, subject to fact-specific inquiry, when combined with increasingly complex platform operations, make it difficult to know where certain immunities granted by the CDA to social platforms end and where tort and product liability begins. Frustrated by congressional lethargy, some litigants and members of civil society have begun to advocate for more activist judges to deploy tools available to them to mitigate harm to users.[13]

Questions about the CDA and the role courts could play in policing social-platform conduct are relevant now more than ever. A new wave of litigation by users of social products and state attorneys general about alleged mental harms and physical injuries caused by the use of social products[14] provides a compelling backdrop to revisit precedent to explicate the CDA's outer edges and the line between user-generated content and platform conduct. It also raises the question: Is the judiciary the right institution to mitigate harms propagated by social platforms, and, if not, then what is?

This Comment seeks to address these issues in three parts. Part I introduces a novel taxonomy to categorize the types of claims plaintiffs generally bring against social platforms based on the specific source of their injuries. Using this new vocabulary, Part II unpacks the specific claims and allegations raised by plaintiffs in pending cases against social platforms. Part III then posits that an ex ante regulatory system operationalized by an expert agency is better suited than an ex post litigation regime to address plaintiffs' injuries and other negative externalities of social products and proposes a design for such a regime.

[12] *Id.*:

> If courts allow their interpretation [of the CDA] to be driven by their evaluation of the policy consequences, they will be substituting their own, potentially inaccurate, views over that of the democratic process. If they continue the status quo, they will lock in the benefits of the current system, but also its harms, and will potentially lower the chance of congressional involvement. And if they interpret the statute narrowly, . . . they risk destabilizing the internet itself.

[13] *See infra* Part II.A.

[14] *See infra* Part II.A.

I. A TAXONOMY OF CLAIMS AGAINST SOCIAL PLATFORMS

Although lawsuits against social platforms are not a new phenomenon, no comprehensive taxonomy exists that cleanly explicates the boundaries of § 230. I posit that claims against social platforms generally fall into one of three categories: content specific, content dependent, or content agnostic. What differentiates these categories is the proximity of the alleged injury to user-generated content and the degree of the platform's participation or passivity in causing the injury: Content-specific claims attempt to tie platform liability to *specific pieces of user-generated content hosted on the platform*. Content-dependent claims attempt to tie platform liability to *a combination of specific, user-generated content and underlying platform mechanics*. Content-agnostic claims attempt to tie platform liability to *platform mechanics and design independent of any specific piece of user-generated content*. This Part discusses each of these categories in turn and introduces a test that courts can use to distinguish between true content-agnostic claims and claims that are content agnostic in name only.

A. Content-Specific Claims

Content-specific claims attempt to tie platform liability to specific pieces of user-generated content hosted on the platform—for example, "Facebook injured me when I used it because it hosted specific pieces of content promoting terrorism."

In the United States, content-specific claims are mostly losing arguments. Section 230(c) of the CDA[15] immunizes "online service providers from legal liability stemming from content created by the users of their services, with some exceptions."[16] This liability shield is sweeping, applying broadly to "many civil actions or state criminal prosecutions brought against" social platforms,[17] save for the small set of exceptions outlined in § 230(e).[18] Section 230(c)(1) "shields conduct if the defendant (1) is a 'provider or user of an interactive computer service'; (2) the claim is based on 'information provided by another information content

[15] 47 U.S.C. § 230(c).

[16] RUANE, *supra* note 6, at 1.

[17] *Id.* at 2.

[18] *See* 47 U.S.C. § 230(e). These exceptions include violations of federal criminal law, intellectual property law, the Electronic Communications Privacy Act of 1986, and certain federal sex-trafficking laws. *Id.*

provider'; and (3) the claim would treat [the defendant] 'as the publisher or speaker' of that information."[19] Platforms (as providers of an "interactive computer service") cannot be held liable as publishers or speakers for claims that stem purely from content provided by other users (as "information content provider[s]").[20]

Although the CDA predates most modern social platforms, courts have paid great deference to Congress's broad drafting of § 230's liability shield when addressing new harms.[21] When a civil rights group attempted to hold Craigslist accountable for discriminatory housing listings posted by other Craigslist users, the Seventh Circuit held that § 230 immunity applied.[22] The court noted that although Craigslist played "a causal role in the sense that no one could post a discriminatory ad if Craigslist did not offer a forum," the plaintiff could not simply "sue the messenger just because the message reveals a third party's plan to engage in unlawful discrimination."[23] Due to the broad statutory construction of § 230(c), plaintiffs seeking to bring content-specific claims against social platforms, aside from the exceptions highlighted in § 230(e), find it to be a fruitless venture.

B. Content-Dependent Claims

In contrast to content-specific claims, content-dependent claims attempt to tie platform liability to a combination of specific user-generated content and underlying platform mechanics—for example, "YouTube injured me when I used it because its recommendations system pushed content promoting terrorism into my

[19] *Backpage.com*, 817 F.3d at 19 (quoting Universal Commc'n Sys. v. Lycos, Inc., 478 F.3d 413, 418 (1st Cir. 2007)).

[20] *Id.*

[21] *See, e.g.*, Marshall's Locksmith Serv. v. Google, LLC, 925 F.3d 1263, 1267 (D.C. Cir. 2019) ("Congress[ ] inten[ded] to confer broad immunity for the re-publication of third-party content."); *Backpage.com*, 817 F.3d at 18 ("There has been near-universal agreement that section 230 should not be construed grudgingly.").

[22] *See* Chi. Laws.' Comm. for C.R. Under L. v. Craigslist, Inc., 519 F.3d 666, 672 (7th Cir. 2008).

[23] *Id.* at 671–72.

feed." Content-dependent claims rest on the idea that the platform's underlying mechanics, such as feed-ranking systems,[24] recommendations systems,[25] or the general use of algorithms, change the effect harmful content has on the user. Although the content is injurious, the platform's conduct with respect to the content also invites culpability. Plaintiffs differentiate content-specific claims from content-dependent claims by arguing that a platform's systems materially affect the role the platform plays such that the platform is no longer a neutral publisher. They argue that the platform is instead an active and key participant in any injury perpetuated by the harmful content it hosts and is thereby liable for the role it played in the injury. Some scholars have opined that this is similar to a contributory liability regime where the platform (as a secondary party) shares the blame for harmful content with its original creator.[26]

Although content-dependent claims seem facially promising, courts have held that the use of tools like algorithms does not automatically imbue liability, preclude immunity, or turn a social platform into the speaker for the content it hosts. When victims of a Hamas attack in Israel sued Facebook, alleging that Facebook's algorithms directed content posted by Hamas encouraging violence and promoting specific terror tactics into the newsfeeds of the individuals that committed those acts, the Second Circuit extended § 230 immunity to the platform.[27] Noting that "tools such as algorithms that are designed to match [third-party] information with a consumer's interests" fall well within the range of publisher functions covered by § 230, the court held that "Facebook's use of algorithms [did not] rende[r] it a non-publisher."[28] More specifically, the court stated that a defendant

[24] Feed-ranking systems utilize machine learning to identify and order the specific pieces of content a given user is most likely to engage with from the general inventory of content available. *See* Akos Lada, Meihong Wang & Tak Yan, *How Machine Learning Powers Facebook's News Feed Ranking Algorithm*, ENG'G AT META (Jan. 26, 2021), https://perma.cc/5LAL-T4X9.

[25] Social recommendations systems predict whether a specific user is likely to enjoy or engage with a given piece of content based on metadata about the content, how that content has performed across the social network, and whether other users similar to the user in question have enjoyed or engaged with that piece of content or others like it in the past. Ido Guy, *Social Recommender Systems*, in RECOMMENDER SYSTEMS HANDBOOK 511, 511–31 (Francesco Ricci, Lior Rokach & Bracha Shapira eds., 2015).

[26] *See, e.g.*, Madeline Byrd & Katherine J. Strandburg, *CDA 230 for a Smart Internet*, 88 FORDHAM L. REV. 405, 431–32, 434–35 (2019).

[27] Force v. Facebook, Inc., 934 F.3d 53, 64–71 (2d Cir. 2019).

[28] *Id.* at 66.

social platform "will not be considered to have developed third-party content unless the defendant directly and 'materially' contributed to what made the content itself 'unlawful.'"[29] Content-dependent liability turns on a material contribution to the unlawful content, and social platforms do not "develop [ ] information (or create new content)" by simply selecting, ordering, and surfacing user-generated content because "the underlying 'information [is] entirely provided by the third party, and the choice of presentation' [falls] within the interactive computer services' prerogative as publishers."[30] The use of algorithms simply automates what is already a publication function and the use of recommendations systems demonstrates nothing more than "Facebook vigorously fulfilling its role as a publisher."[31] In order to succeed, plaintiffs need to demonstrate something more: an indication that the social platform has in some way developed or materially contributed to user-generated content such that the platform becomes more than the mere publisher of that content.[32]

The Supreme Court indirectly tackled this distinction in *Twitter, Inc. v. Taamneh*.[33] Plaintiffs who were injured by an ISIS attack at a nightclub in Turkey sued Twitter, Facebook, and Google for aiding and abetting ISIS in carrying out the attack.[34] More specifically, the plaintiffs argued that the platforms aided and abetted terrorism[35] under the Justice Against Sponsors of Terrorism Act[36] when their algorithms enabled ISIS to spread its message and reach users it could not otherwise have reached. Although the Court did not specifically address § 230 immunity in its opinion in *Taamneh* (similarly sidestepping the issue in *Gonzalez v. Google, LLC*,[37] a sister case heard by the Court a day

[29] *Id.* at 68 (quoting FTC v. LeadClick Media, LLC, 838 F.3d 158, 174 (2d Cir. 2016)).

[30] *Id.* at 69 (emphasis omitted) (quoting *Marshall's Locksmith Serv.*, 925 F.3d at 1269).

[31] *Id.* at 70; *see also id.* at 67.

[32] *See, e.g.*, *LeadClick Media*, 838 F.3d at 158 (holding that a defendant platform developed and materially contributed to unlawful content generated by its affiliates when it provided its affiliates specific instructions on how to edit their websites); *Marshall's Locksmith Serv.*, 925 F.3d at 1263 (holding that platform-mapping services did not develop or contribute to false location data provided by locksmiths seeking to mislead consumers by translating the data into a visual map).

[33] 143 S. Ct. 1206 (2023).

[34] *Id.* at 1214–15.

[35] Brief for Respondents at 6–10, *Taamneh*, 143 S. Ct. 1206 (2023) (No. 21-1496).

[36] Pub. L. No. 114-222, 130 Stat. 852 (2016) (codified as amended in scattered sections of 18 and 28 U.S.C.).

[37] 143 S. Ct. 1191 (2023).

before *Taamneh*), the Court's opinion rejecting the plaintiffs' arguments appeared to hinge on the contribution that algorithms and targeted recommendations systems make to the user-generated content they surface.[38] Much of the Court's analysis rested on an assumption of algorithmic passivity toward the third-party content shown, drawing a line between harmful content and algorithms as neutral infrastructure.[39] Under this paradigm, litigants seeking to hold social platforms liable for harms arising in part or in whole from the user-generated content they host, by highlighting the role platforms play in elevating and circulating that content, have not succeeded because their injuries are inextricably linked to third-party speech.

C. Content-Agnostic Claims

Facing difficulty in pinning liability to platforms under the content-specific or content-dependent frameworks, some plaintiffs have turned to content-agnostic claims. Content-agnostic claims attempt to tie platform liability to platform mechanics and design independent of any specific piece of user-generated content—for example, "Snapchat injured me when I used it, and my injury was caused by some inherent aspect of the platform, not the content I saw on it."

Content-agnostic claims are not a new phenomenon. Recognizing the difficulties posed by § 230's sweeping shield, plaintiffs in cases as early as 2007 attempted to bypass user-generated content in their (sometimes "artful"[40]) pleadings.[41] Some academics have since recognized the need for plaintiffs to pivot to

[38] *Taamneh*, 143 S. Ct. at 1226 ("[P]laintiffs assert that defendants' 'recommendation' algorithms go beyond passive aid and constitute active, substantial assistance. We disagree.").

[39] *Id.* at 1227 ("[D]efendants' 'recommendation' algorithms are merely part of [ ] infrastructure. . . . As presented here, the algorithms appear agnostic as to the nature of the content, matching any content (including ISIS' content) with any user who is more likely to view that content."). This view on algorithmic passivity warrants further examination for factual accuracy. *See infra* Part III.B.4.

[40] Doe v. MySpace, Inc., 474 F. Supp. 2d 843, 849 (W.D. Tex. 2007).

[41] *See, e.g.*, *id.*:

> Plaintiffs argue this suit is based on MySpace's negligent failure to take reasonable safety measures to keep young children off of its site and not based on MySpace's editorial acts. The Court, however, finds this artful pleading to be disingenuous. It is quite obvious the underlying basis of Plaintiffs' claims is . . . postings on MySpace.

content-agnostic claims to succeed.[42] What has changed recently, however, is the context in which content-agnostic claims are being brought: a growing mental health crisis in the United States,[43] increased evidence suggesting a connection between psychological well-being and social media use,[44] scholarship outlining the social and policy benefits of extending liability to social platforms and the natural deficiencies of § 230,[45] and a frustration among some that § 230's "liability protections are overbroad or unwarranted."[46] As a result, some judges appear willing to sever a social platform's publication function from its recommendations function.[47] This effectively lays the conceptual foundation for content-

[42] *See, e.g.*, Allison Zakon, Comment, *Optimized for Addiction: Extending Product Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act*, 2020 WIS. L. REV. 1107, 1135 (noting that "[c]ases [against social platforms] will fail when litigants cannot separate the source of the harm from choices that the platform made about what content can appear on the site," but they will succeed "when they do not tie their claims to specific harmful content").

[43] *See, e.g.*, MADDY REINERT, THERESA NGUYEN & DANIELLE FRITZE, MENTAL HEALTH AM., THE STATE OF MENTAL HEALTH IN AMERICA 8 (2023); Monica Anderson, *A Majority of Teens Have Experienced Some Form of Cyberbullying*, PEW RSCH. CTR. (Sept. 27, 2018), https://perma.cc/75ED-DHZK; *AAP-AACAP-CHA Declaration of a National Emergency in Child and Adolescent Mental Health*, AM. ACAD. OF PEDIATRICS (Oct. 19, 2021), https://perma.cc/L4FC-YFJM.

[44] *See generally* Fazida Karim, Azeezat Oyewande, Lamis F. Abdalla, Reem C. Ehsanullah & Safeera Khan, *Social Media Use and Its Connection to Mental Health: A Systematic Review*, 12 CUREUS, no. 6, 2020.

[45] *See, e.g.*, Stanley M. Besen & Philip L. Verveer, *Section 230 and the Problem of Social Cost*, 30 J.L. & POL'Y 68, 68 (2021) (applying the Coase Theorem to argue that it is efficient to hold internet platforms accountable as the best-situated problem solvers for the negative externalities they perpetuate because platforms "will often be easier to identify [than the original source of the injurious content] and because they have greater ability to engage in content moderation"); Byrd & Strandburg, *supra* note 26, at 434 ("CDA 230 was simply not designed or intended to handle situations in which a service provider's *activities* as a publisher are actionable but the published *content* is not." (emphasis in original)).

[46] RUANE, *supra* note 6, at 1; *see also* Joe Biden, President of the United States, State of the Union Address (Mar. 1, 2022) ("[W]e must hold social media platforms accountable for the national experiment they're conducting on our children for profit.").

[47] *See, e.g.*, *Force*, 934 F.3d at 82 (Katzmann, J., concurring in part and dissenting in part):

> [C]laims based on [recommendations systems] algorithms do not inherently treat Facebook as the publisher of third-party content. First, Facebook uses the algorithms to create and communicate its own message: that it thinks you, the reader—you, specifically—will like this content. And second, Facebook's suggestions contribute to the creation of real-world social networks. The result of at least some suggestions is not just that the user consumes a third party's content. Sometimes, Facebook's suggestions allegedly lead the user to become part of a unique global community, the creation and maintenance of which goes far beyond and differs in kind from traditional editorial functions.

agnostic claims to circumvent § 230 immunity. So long as plaintiffs challenge the infrastructure serving third-party content, and not the third-party content itself, some courts may be willing to buck the dominant "CDA-driven, hands-off approach to social media."[48]

This strategy has already proven effective in holding some social platforms accountable for injuries arising from their use. For example, in *A.M. v. Omegle.com, LLC*,[49] a young woman sued Omegle—an online chat room that randomly connected strangers—for connecting her "with an adult man who sexually abused her online through [the platform]" while she was a minor.[50] The district court held that § 230 did not bar product liability claims arising from Omegle's failure to warn users or design a safe social product.[51] Days after the parties reached an undisclosed settlement, Omegle was shut down for good.[52]

D. Distinguishing Real Content-Agnostic Claims from Content-Specific and Content-Dependent Claims in Disguise

Given the near-absolute shield provided by § 230 and the potential opening presented by content-agnostic claims, plaintiffs bringing cases against social platforms have a strong incentive to reframe and present their injuries as content agnostic in nature. This muddies the water for judges, because it is not always clear how to appropriately attribute blame between the platform and the underlying content in any particular controversy. However, courts have demonstrated a willingness to look past creative labeling; injuries that are inseparable from third-party content may be losing content-dependent claims in disguise. For example, when a mother sued TikTok after her child died while partaking in the viral "blackout challenge,"[53] the court determined that the mother's claims were premised "on the 'defective' manner in

[48] *Id.* at 86.

[49] 614 F. Supp. 3d 814 (D. Or. 2022).

[50] *Id.* at 817.

[51] *Id.*

[52] Bill Chappell, *Video Chat Site Omegle Shuts Down After 14 Years—and an Abuse Victim's Lawsuit*, NPR (Nov. 9, 2023), https://perma.cc/E4Y4-XKMN.

[53] In the blackout challenge, users record themselves strangling themselves with household items and then encourage others to do the same. Olivia Carville, *TikTok's Viral Challenges Keep Luring Young Kids to Their Deaths*, BLOOMBERG (Nov. 29, 2022), https://perma.cc/2SMH-RFBQ.

which [TikTok] *published* a third party's dangerous content."[54] Looking past the label applied by the plaintiff,[55] the court noted that "[b]ecause [the plaintiff's] design defect and failure to warn claims [were] 'inextricably linked' to the manner in which [TikTok] cho[se] to publish third-party user content, Section 230 immunity applie[d]."[56]

So how should courts distinguish real content-agnostic claims from claims that are content agnostic in name only? One way to separate the wheat from the chaff is to focus on the remedy. For true content-agnostic claims, the specific injury alleged can be remedied (both for the current plaintiff and for a similarly situated hypothetical user in the future) *without referring to or implicating content generated by a third party in any way*.

This remedies test is a corollary of the "material contribution" test first outlined by the Ninth Circuit in *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*,[57] which states that a "website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct."[58] If a platform materially contributed to the content that caused injury, then it should be possible for a platform to remedy the harm by changing its conduct or operations without referring to or implicating third-party content. If a platform cannot remedy the harm without implicating third-party content, then the platform's contribution is not material and the underlying claim is not content agnostic.

Notably, some courts have already endorsed the delineating principle outlined in the remedies test. The test mirrors the analysis conducted in *Airbnb, Inc. v. City and County of San Francisco*[59] and *HomeAway.com v. City of Santa Monica*.[60] In those cases, the respective courts allowed platforms hosting third-party vacation rentals to face liability for violating local ordinances against unlicensed rentals because the "platforms made

[54] Anderson v. TikTok, Inc., 637 F. Supp. 3d 276, 280 (E.D. Pa. 2022) (emphasis in original).

[55] *Id.* ("[W]hat matters is not the name of the cause of action—defamation versus negligence versus intentional infliction of emotional distress—what matters is whether the cause of action inherently requires the court to treat the defendant as the 'publisher or speaker' of content provided by another." (quoting Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1101 (9th Cir. 2009))).

[56] *Id.* at 281 (quoting Herrick v. Grindr, LLC, 765 F. App'x 586, 591 (2d Cir. 2019)).

[57] 521 F.3d 1157 (9th Cir. 2008).

[58] *Id.* at 1168.

[59] 217 F. Supp. 3d 1066 (N.D. Cal. 2016).

[60] 918 F.3d 676 (9th Cir. 2019).

decisions about where, how, and to whom to offer listings, in ways that violated local law."[61] "[I]f a platform could modify its own use of recommender algorithms—its own conduct—to comply with applicable law without reference to users' content, the illegality comes from the platform's choices, and Section 230 immunity does not apply."[62] The district court in *Omegle* also emphasized remedies in its analysis, noting that "Omegle would not have to alter the content posted by its users" in order to meet the obligations the plaintiff sought to impose on the platform; instead, "it would only have to change its design and warnings."[63]

This remedies test proves effective when applied to borderline cases or cases where it is facially difficult to isolate platform mechanics from third-party content. For example, fatalities in a car accident that occurs while the driver is using Snapchat's "speed filter" (which allows users to overlay their current speed over a previously captured video or photo) may give rise to content-agnostic claims. Although the filter is applied onto user-generated content and the filter's sole purpose is to be a content-authoring tool, the underlying claim is content agnostic because the injury in question could be prevented by disabling the speed filter when the platform detects movement at high speeds. The Ninth Circuit and the Court of Appeals of Georgia both agreed that the injuries in similar cases stemmed from Snapchat's own conduct[64] and that plaintiffs' claims do "not seek to hold Snap responsible as a publisher or speaker" in a manner that would invoke § 230 immunity.[65] In contrast, injuries resulting to users of Grindr and Facebook stemming from other users creating fake or impersonating profiles to harass or lure victims into sex trafficking are not content agnostic because the underlying injury cannot be mitigated without removing these profiles (third-party content) or restricting their ability to message other users (third-party speech). The Second Circuit and the Supreme Court of Texas both agreed[66] that claims arising from these fake

[61] Brief of the Integrity Institute and Algotransparency as Amici Curiae in Support of Neither Party at 24, *Gonzalez*, 143 S. Ct. 1191 (2023) (No. 21-1333) [hereinafter Brief of the Integrity Institute].

[62] *Id.*

[63] *Omegle*, 614 F. Supp. 3d at 820.

[64] Lemmon v. Snap, Inc., 995 F.3d 1085, 1093 (9th Cir. 2021); Maynard v. Snapchat, Inc., 816 S.E. 2d 77, 81 (Ga. Ct. App. 2018).

[65] *Lemmon*, 995 F.3d at 1093 (quoting *Maynard*, 816 S.E. 2d at 81).

[66] *See Herrick*, 765 F. App'x at 586; *In re* Facebook, Inc., 625 S.W. 3d 80, 93 (Tex. 2021).

profiles would be predicated upon "second-guessing of [the platform's] decisions relating to the monitoring, screening, and deletion of [third-party] content from its network," and thereby implicated § 230.[67]

Because the remedies test focuses attention on the platform's contribution to the injury and its ability to mitigate the injury, it serves as an effective sieve for claims that are content agnostic in name only. As noted in *Roommates.com*, if a website provides search functionality that allows users to filter third-party housing advertisements by protected characteristics under state and federal law, the website "forfeit[s] any immunity to which it was otherwise entitled under section 230."[68] These circumstances give rise to a content-agnostic claim because the website can remedy the injury without involving third-party content in any way—by removing the ability to filter content by protected characteristics. In contrast, if the same website allows users to create their own criteria for choosing roommates by providing "a blank text box," the website retains its § 230 immunity "so long as it does not require the use of discriminatory criteria."[69] These circumstances do not give rise to content-agnostic claims because the underlying injury (housing discrimination) cannot be addressed without editing or deplatforming third-party content.

E. The Clarity Aided by This Taxonomy

Applying the proposed taxonomy to organize claims raised by plaintiffs into content-specific, content-dependent, and content-agnostic buckets using the remedies test allows courts to focus their attention on injuries for which platforms themselves are more plausibly responsible. For product engineers and the nonlegal professionals who build and maintain social platforms every day, adopting this taxonomy enables an easier understanding of legal obligations and liabilities. The remedies test provides a practicable guiding principle to distinguish product mitigations that *should* be shipped[70] from those that *must* be. For advocates unhappy with the status quo, this taxonomy invites a clearer conversation on § 230 reform; rather than debating the imposition of

[67] *In re Facebook*, 625 S.W. at 93.

[68] *Roommates.com*, 521 F.3d at 1170.

[69] *Id.* at 1173, 1169.

[70] In software development, shipping is the act of publishing, deploying, or otherwise making an application or feature available to users. *See Software: Code Shipping Cycle*, DATA PANDA (Oct. 12, 2023), https://perma.cc/EVJ7-BNP5.

liability in the abstract, the proposed taxonomy provides a discrete spectrum—from pure third-party speech (content specific) to pure platform conduct (content agnostic)—with multiple intermediate points at which the line of liability may be drawn.

This taxonomy is also useful because it allows courts to weaken § 230's stronghold without undermining its purpose. The content-agnostic framework, supported by application of the remedies test, provides both an opening for ex post liability and a rough metric for its limits.

II. Current Plaintiffs and Their Content-Agnostic Claims

Several plaintiffs today are hoping to capitalize on the shifting sentiment toward § 230 and the role that increasingly complex social platforms play in the growing mental health crisis. Collectively, these lawsuits constitute one of the most significant legal challenges social platforms have faced in recent memory. Analyzing these plaintiffs' claims through the content-specific, content-dependent, and content-agnostic framework and the remedies test provides a valuable opportunity to assess both the efficacy of the taxonomy and the validity of the claims before courts today. Part II.A reviews plaintiffs' identities and claims, Part II.B postulates their objectives and litigation strategy, and Part II.C analyzes the authenticity of their alleged content-agnostic claims using the remedies test.

A. The Plaintiffs

Today's litigants fall into one of five groups. The first three groups comprise different categories of users injured by social products: users who claim injuries to themselves, parents who allege nonfatal injuries to their children, and parents suing for the wrongful deaths of their children resulting from social media use or platform (in)action. The fourth group consists of school districts alleging injury due to increased expenditures on mental health resources necessitated by platforms' harmful product-design choices. The final group of plaintiffs consist of state attorneys general suing on behalf of young users of social products. This Section discusses each of these plaintiffs and their injuries in more detail.

The first group includes social media users who claim injury to themselves, alleging that their use of social platforms has damaged their mental health or exacerbated issues such as eating disorders, anxiety, or depression. For example, a twenty-one-year-old plaintiff, who is a heavy user of Meta's social platforms, alleges that her lack of sleep, development of an eating disorder, depression, body dysmorphia, anxiety, suicidal ideation, and practice of self-harm stem from her addiction to social media.[71] She claims that she was unaware of the dangers of using social platforms and that Meta "misrepresented the safety, utility, and non-addictive properties of [its] products."[72]

The second group of plaintiffs are parents who allege similar injuries to their children. For example, parents of a thirteen-year-old boy, who is a heavy user of Meta's social platforms, allege "addictive and problematic" use of the social platform resulting in reduced sleep, depression, social media compulsion, and anxiety.[73]

A third group of plaintiffs are parents suing for the wrongful deaths of their children resulting from social media use or platform (in)action. In one case, a minor made social media accounts without her parents' knowledge and began incessant use of the platforms.[74] She used platforms like Snapchat to send sexually suggestive photographs to other users, which were then circulated or threatened to be circulated to others.[75] One night, her parents took away her phone as punishment for sneaking out of the house and discovered the inappropriate photographs; the following morning, she died by suicide.[76] In another case, a child created a Facebook account with his parents' permission and used it throughout the day and night.[77] His parents later learned that he was engaging in self-harm and was getting insufficient sleep.[78] His phone was confiscated by his parents after he had a fight with his brother; later that night, the boy's parents came home to find that he had died by suicide.[79] The boy's parents directly attribute his death to social-platform design.[80]

[71] Complaint at 21–23, Murden v. Meta Platforms, Inc., No. 3:22-CV-01511 (S.D. Ill. July 13, 2022) [hereinafter Murden Complaint].

[72] *Id.* at 22.

[73] Complaint at 22, Williams v. Meta Platforms, Inc., No. 1:22-CV-03470 (N.D. Ill. July 5, 2022).

[74] Complaint at 53–58, Gill v. Meta Platforms, Inc., No. 1:22-CV-02173 (W.D. La. July 20, 2022) [hereinafter Gill Complaint].

[75] *Id.* at 61.

[76] *Id.* at 64.

[77] Complaint at 50–59, Aranda v. Meta Platforms, Inc., No. 4:22-CV-04209 (N.D. Cal. July 20, 2022) [hereinafter Aranda Complaint].

[78] *Id.* at 54–56.

[79] *Id.* at 56–57.

[80] *Id.* at 58 (alleging that the child's anxiety, depression, and death would have been avoided "[b]ut for Meta's refusal to conduct . . . age verification or confirm parental consent," the platform's social features that enable "harmful social comparison," its use of

Given the similarities in the causes of action being pursued by these first three groups of plaintiffs, in 2022, the United States Judicial Panel on Multidistrict Litigation grouped more than eighty lawsuits by these three groups of plaintiffs into a single multidistrict litigation action centralized in the Northern District of California.[81]

The fourth group of litigants are school districts. Asserting that "[s]chool districts are on the frontlines of [the] unfair fight" between "students . . . being victimized and exploited" and "Social Media Companies . . . ruthlessly extracting every dollar possible with callous disregard for the harm to mental health,"[82] dozens of school districts have alleged injury due to increased expenditures on mental health resources.[83] More specifically, these school districts have attributed a number of additional costs they now bear to deliberate choices made by social platforms. These include costs associated with obtaining mental health resources, hiring more counselors and mental health professionals, training teachers to notice and address mental health issues in their students, and wasted time due to classroom disruption arising from students' use of social media in class.[84]

State attorneys general comprise the fifth group of litigants. In a heavily redacted federal complaint (representing a coalition of thirty-three state attorneys general)[85] and in a coordinated but separate set of complaints (filed by an additional eight states and the District of Columbia),[86] plaintiffs allege that Meta schemed to "exploit[ ] young users of its Social Media Platforms."[87] Plaintiffs allege that Meta injured young users by failing to "disclose . . .

"recommendation systems," its "endless feed and explore features," and its "content promotion and amplification, public profile, and direct messaging settings"—all of which coalesce to perpetuate the "harmful dependencies that [they] were designed to promote").

[81] Transfer Order, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., MDL No. 3047 (J.P.M.L. Oct. 6, 2022).

[82] Complaint at 2, Sch. Dist. of the Chathams v. Meta Platforms, Inc., No. 2:23-CV-00910 (D. N.J. Feb. 16, 2023) [hereinafter Chathams Complaint].

[83] *Id.* at 58–61; *see also* Complaint at 65–69, Tyrone Area Sch. Dist. v. Meta Platforms, Inc., No. 3:23-CV-155 (W.D. Pa. July 12, 2023).

[84] *See, e.g.*, Complaint at 47–53, Pittsburgh Pub. Schs. v. Meta Platforms, Inc., No. 4:23-CV-02085 (W.D. Pa. Apr. 6, 2023).

[85] *See generally* Complaint, Ariz. v. Meta Platforms, Inc., No. 4:23-CV-05448 (N.D. Cal. Oct. 24, 2023) [hereinafter Arizona Complaint].

[86] *See generally, e.g.*, Complaint, Att'y Gen., Fla. v. Meta Platforms, Inc., No. 8:23-CV-02412 (M.D. Fla. Oct. 24, 2023) [hereinafter Florida Complaint]; Complaint, Utah Div. of Consumer Prot. v. Meta Platforms, Inc., No. 230908060 (Utah Dist. Ct. Oct. 24, 2023) [hereinafter Utah Complaint].

[87] Arizona Complaint, *supra* note 85, at 11.

that it is weaponizing young users' data to capture and keep their attention" and by misrepresenting and omitting information regarding the addictive and harmful nature of its social products in public statements.[88] Authorized by their "respective states' Unfair and Deceptive Acts and Practices statutes . . . to enforce such statutes," these state attorneys general seek monetary damages and injunctive relief.[89]

B. Plaintiffs' Arguments for Liability

Having established who these plaintiffs are, it is important to understand what they are alleging and why. This Section reviews the general arguments raised by these five groups of plaintiffs, discusses their pleading and litigation strategy, and utilizes the content-agnostic framework and remedies test to unpack the validity of their claims.

1. Structure of arguments.

These five sets of plaintiffs are attempting to pin liability to social platforms via a myriad of theories and causes of action, such as public nuisance, negligence and products liability, fraud and fraudulent concealment, negligent infliction of emotional distress, and more.[90] These claims generally proceed in three parts. First, plaintiffs allege that platform algorithms designed to maximize the time users spend on the platform and feed-ranking systems designed to encourage endless scrolling and use of social media are inherently harmful. Plaintiffs specifically point their fingers at features such as "[p]ublicly visible social metrics" that turn popularity into a competition;[91] algorithms that are optimized for maximized engagement, even if that engagement happens in the middle of the night or the school day; endless feeds of harmful content being served to users; direct messaging; and other manifestations of "social reciprocity" on the platform, such as read receipts, which make users feel compelled to return to the platform and respond to messages received.[92] They allege that

[88] *Id.* at 32; *see id.* at 67–70. Plaintiffs also allege violations of the Children's Online Privacy Protection Act; however, these claims are outside the scope of this Comment. *Id.* at 105.

[89] *Id.* at 6; *see also* Arizona Complaint, *supra* note 85, at 198–217.

[90] *See, e.g.*, Murden Complaint, *supra* note 71, at 23–92.

[91] Chathams Complaint, *supra* note 82, at 12–13.

[92] Murden Complaint, *supra* note 71, at 11–15.

these features "take advantage of biological systems, human behavior, and psychology to addict and condition users to engage in repetitive content-consuming actions such as scrolling, liking, and sharing content in search of repeated dopamine releases."[93] As evidence, some plaintiffs cite remarks made by Facebook's first director of monetization, Tim Kendall. In his congressional testimony in 2020, Kendall testified that Facebook "took a page from Big Tobacco's playbook, working to make [its] offering addictive at the outset" and analogized enabling misinformation and conspiracy theories via algorithms to advancements in cigarette design that allowed nicotine to be more effectively delivered to a smoker's brain and lungs.[94] Plaintiffs underscore their content-agnostic claims by highlighting academic research[95] and comments by public authorities[96] referencing the negative externalities of social media use on individual health.

Second, plaintiffs have argued that social media products exploit the underdeveloped "decision-making capacity, impulse control, emotional maturity, and psychological resiliency" of teen minds, and that the use of social platform products led to them or their family members being ridiculed at school, being hospitalized for psychiatric care, experiencing a multitude of mental health issues, or dying by suicide.[97] They have claimed that many features available on modern social platforms operate as intermittent variable rewards and prey on the "chemical reward system

[93] *Id.* at 28.

[94] Chathams Complaint, *supra* note 82, at 11–12.

[95] *See, e.g.*, Kira E. Riehms, Kenneth A. Feder, Kayla N. Tormohlen, Rosa M. Crum, Andrea S. Young, Kerry M. Green, Lauren R. Pacek, Lareina N. La Flair & Ramin Mojtabai, *Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth*, 76 JAMA PSYCHIATRY 1266, 1271–72 (2019); Meg Pillion, Michael Gradisar, Kate Bartel, Hannah Whittall & Michal Kahn, *What's "App"-ning to Adolescent Sleep? Links Between Device, App Use, and Sleep Outcomes*, 100 SLEEP MED. 174, 179 (2022) (demonstrating that YouTube was the "only app consistently and negatively related to sleep outcomes").

[96] *See, e.g.*, Alvaro M. Bedoya, Chairman, FTC, Prepared Remarks at the National Academies of Sciences, Engineering & Medicine Meeting of the Committee on the Impact of Social Media on the Health and Wellbeing of Children & Adolescents (Feb. 7, 2023) ("[W]e live in an attention economy . . . . [C]ompanies very literally compete for our thoughts, our time, our minds. No one should be surprised if that economy affects our mental health."); U.S. DEPT. OF HEALTH AND HUM. SERVS., SURGEON GEN.'S ADVISORY ON SOC. MEDIA AND YOUTH MENTAL HEALTH 4 (2023) ("[T]here are ample indicators that social media can also have a profound risk of harm to the mental health and well-being of children and adolescents.").

[97] Complaint at 31, Mt. Lebanon Sch. Dist. v. Meta Platforms, Inc., No. 2:23-CV-00651 (W.D. Pa. Apr. 20, 2023).

of users' brains (especially young users),"[98] as likes and other measures of social approval trigger "enhanced dopamine responses to stimuli on [ ] social media platforms," hormonal responses to which adolescents are inherently more sensitive.[99] Plaintiffs have alleged that social platforms knew that their products would have such a psychological effect on minors[100] but that they wantonly continue to "grow the use of their platforms by minors through designs, algorithms, and policies that promote addiction, compulsive use, and other severe mental harm."[101] They have also claimed that platforms intentionally "thwart[ ] the ability of parents to keep their children safe and healthy by supervising and limiting social media use."[102] Citing whistleblower Frances Haugen's testimony demonstrating that Meta knew about the effects its platforms were having on teen mental health and subsequent leaked internal documents from various platforms,[103] they have concluded that social platforms should be liable for operating addictive and unreasonably dangerous products and for failing to warn users of the dangers of social media use.[104]

Lastly, plaintiffs have posited that social platforms have a duty of care toward their users and that they violated that duty of care. Noting that platforms knew[105] or should have foreseen that their design choices would inflict mental health harms on children,[106] plaintiffs have contended that platforms violated their

[98] Complaint at 24, Estevanott v. Meta Platforms, Inc., No. 6:22-CV-03149 (S.D. Mo. June 7, 2022) [hereinafter Estevanott Complaint].

[99] Second Amended Complaint at 21–22, Rodriguez v. Meta Platforms, Inc., No. 3:22-CV-00401 (N.D. Cal. May 6, 2022).

[100] *Id.* at 4 ("[Social platforms] also know . . . that minor users of their social media products are much more likely to sustain serious physical and psychological harm through their social media use than adult users.").

[101] Chathams Complaint, *supra* note 82, at 2.

[102] *Id.*

[103] *See, e.g.*, *Primary Source Documents*, THE FACEBOOK PAPERS (Nov. 24, 2021), https://perma.cc/AJX6-WZQ8; Ben Smith, *How TikTok Reads Your Mind*, N.Y. TIMES (Dec. 5, 2021), https://perma.cc/YE8B-JP67.

[104] Aranda Complaint, *supra* note 77, at 13–15.

[105] Some plaintiffs reference a chart published by Facebook founder Mark Zuckerberg—which shows that natural engagement with a piece of content increases exponentially as the content gets closer to violating platform policies—as proof that platforms were aware that their algorithms steered users toward the most negative content. *See* Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (May 5, 2021), https://www.facebook.com/notes/751449002072082/.

[106] *See, e.g.*, Murden Complaint, *supra* note 71, at 19 ("Meta is aware that teens often lack the ability to self-regulate."); Complaint at 12, Youngers v. Tiktok, Inc., No. 4:22-CV-06456 (N.D. Cal. Oct. 24, 2022) ("Children are particularly susceptible to TikTok's manipulative algorithm and have diminished capacity to avoid self-destructive behaviors.").

duty by not providing accurate information and warnings concerning the risks and potential adverse effects of using social products.[107] This argument is plausible because, arguably, social platforms' duty of care has changed over time: whereas platforms may not have known about the negative impacts of social products on mental health a decade ago, sufficient research exists today linking product-design choices made by social platforms to plaintiffs' injuries.[108] Plaintiffs have concluded by highlighting cost-effective, reasonably feasible design alternatives available to platforms to mitigate the alleged injuries, such as developing less aggressive recommendations systems, limiting the number or length of user sessions, and using more effective age-gating and verification technologies.[109]

2. Cementing content-agnostic claims and litigation strategy.

Although it is valuable to note what these plaintiffs *are* doing, it is more important to note what they are expressly *not* doing. Due to the difficulties plaguing content-specific and content-dependent claims, and to avoid premature dismissal, these plaintiffs have expressly disavowed any claims seeking to hold platforms liable as the publishers or speakers of third-party speech.[110] Each plaintiff has been careful to state that their claims arise from the platform product itself and that their injuries can be mitigated "without altering, deleting, or modifying the content of a single third-party post or communication."[111]

This disclaimer is of utmost importance to litigants on both sides of the matter. For those seeking to hold social platforms accountable, there is value in these plaintiffs successfully pleading past the motion to dismiss stage and reaching discovery, even if they eventually lose at trial. Given how little is publicly known

[107] *See, e.g.*, Complaint at 4, 18–19, Cerone v. Meta Platforms, Inc., No. 4:22-CV-06417 (S.D. Ga. Sept. 14, 2022).

[108] *See generally, e.g.*, Karim et al., *supra* note 44.

[109] *See, e.g.*, Complaint at 30–32, Harrison v. Meta Platforms, Inc., No. 2:22-CV-12038 (E.D. Mich. Aug. 30, 2022).

[110] *See, e.g.*, Murden Complaint, *supra* note 71, at 21 ("Plaintiff's claims arise from Defendants' status as the designer and marketer of dangerously defective social media products, not as the speaker or publisher of third-party content. . . . None of Plaintiff's claims for relief set forth herein require treating Defendants as a speaker or publisher of content posted by third parties."); Utah Complaint, *supra* note 86, at 58 ("Meta's unconscionable design choices include deploying features . . . that unfairly harm children independently of any actions taken by third-party users of Meta's Platforms.").

[111] Murden Complaint, *supra* note 71, at 21.

about how specific social platforms operate[112] (beyond what was made public through the Facebook Papers[113] and other leaks), any discovery conducted is likely to benefit future litigants, academics, and researchers. Furthermore, given the complexity and size of social platforms, any discovery conducted could be akin to throwing a dart in the dark, with a nontrivial chance of hitting something significant. On the other hand, social platforms, unwilling to expose themselves to these risks and perhaps unable to comply with the financial cost and technical complexity of complying with discovery, may be willing to settle at any cost if plaintiffs survive the motion to dismiss stage. As a result, a lot rests on whether plaintiffs can frame their content-agnostic claims as plausible.

C. Analyzing Plaintiffs' Alleged Content-Agnostic Claims

Do plaintiffs' injuries really raise content-agnostic claims? Upon closer examination, most of the plaintiffs' alleged injuries are inseparable from third-party content, but a handful of allegations do give rise to legitimate content-agnostic claims.

Although plaintiffs have claimed that they are not seeking to hold platforms accountable for third-party speech and are adamant to highlight that their mental health injuries (such as body dysmorphia, depression, and anxiety) stem purely from the social platforms' conduct, applying this Comment's proposed remedies test demonstrates the inextricable link to third-party content. Under the test, plaintiffs' injuries give rise to true content-agnostic claims only if their specific injuries can be remedied without referring to or implicating third-party content in any way. But remedies to these mental health injuries inherently require social platforms to make different choices about how they select, display, organize, or promote third-party content, and plaintiffs recognize this in their briefs.

[112] *See* Renée DiResta, Laura Edelson, Brendan Nyhan & Ethan Zuckerman, *It's Time to Open the Black Box of Social Media*, SCI. AM. (Apr. 28, 2022), https://perma.cc/6AZR-R5M2 (emphasis in original):

> [S]ocial media companies are stingy about releasing data and publishing research, especially when the findings might be unwelcome . . . . [W]e need access to data on the *structures* of social media, such as platform features and algorithms, so we can better analyze how they shape the spread of information and affect user behavior.

[113] *Primary Source Documents*, *supra* note 103.

For example, as part of their negligence and design-defect claims, plaintiffs have proposed "cost effective, reasonably feasible alternative designs" that platforms should have utilized to "minimize [their] harms."[114] These include "[a]ge-based content filtering"; "[g]eneral content filtering"; "[a]lgorithmic . . . reductions or elimination in a user's feed of potentially harmful content," such as "in the genres of lifestyle, influencer, beauty, fitness, success flaunting, and/or heavily edited images and videos," "inappropriate or salacious content," "controversial, political, or emotionally weighted content," and "content encouraging or promoting eating disorders, depressive thinking, self-harm, or suicide"; and "[c]hronological presentation of content rather than algorithmic."[115] Each of these proposals requires platforms to make different editorial choices about the third-party content they host, and as a result, the underlying claim cannot be construed as content agnostic under the remedies test. Because these claims are, in fact, content-dependent claims in disguise, they should fail due to § 230.

These plaintiffs' briefs do raise other legitimate content-agnostic claims, however. For example, claims arising from misrepresentations by platforms in public statements regarding the safety of social products are genuine content-agnostic claims because the requisite remedy would simply be to issue more accurate public statements. Similarly, a failure to warn users of the known harms of social-product use also gives rise to real content-agnostic claims in these cases because the necessary remedy implicates the business entity's speech about what it knows about its social products and does not implicate third-party speech.[116] Some product-feature design choices, such as enabling

[114] Murden Complaint, *supra* note 71, at 25.

[115] *Id.* at 25–26, 40–41.

[116] Note that the Ninth Circuit conducted a similar analysis and reached a similar conclusion in Doe v. Internet Brands, Inc., 824 F.3d 846 (9th Cir. 2016). In that case, the court concluded that the plaintiff's failure to warn claim, arising from an incident where two individuals used the defendant platform to lure and rape the plaintiff, was not barred by the CDA because "[a]ny alleged obligation to warn could have been satisfied without changes to the content posted by the website's users." *Id.* at 851. The warning was made necessary not because of content posted by the two rapists on the platform, but because the two rapists used information posted by the plaintiff on the platform to target her. This can be contrasted with the Eastern District of Pennsylvania's holding in *Anderson*, where the warning was made necessary due to harmful content about the blackout challenge posted by other users on TikTok. 637 F. Supp. 3d at 276. In *Anderson*, the warning remedy necessarily implicated third-party content because the need to warn arose from third-party content and the warning's content was specific to the harmful third-party content

read receipts or defaulting users to less protective privacy settings, also give rise to content-agnostic claims under the remedies test, as mitigation does not require platforms to treat third-party content differently.[117] As legitimate content-agnostic claims, these claims evade § 230's broad shield and pose a material challenge to the respective platform defendants.

Courts appear amenable to bucketing plaintiffs' claims in this fashion. In November 2023, the United States District Court for the Northern District of California granted in part and denied in part social platforms' motion to dismiss in the multidistrict litigation that consolidated lawsuits filed by users and their parents.[118] The court's analysis mirrored the remedies test, as the court distinguished allegations barred by § 230 from those that could proceed to trial by determining whether "such allegations . . . can be fixed by [defendant platforms] without altering the publishing of third-party content."[119] However, the court was inconsistent in applying its reasoning to the facts, dismissing some claims that would be classified as content agnostic under the proposed taxonomy while allowing some claims to proceed that would be classified as content agnostic in name only.[120]

III. THE CASE FOR AN EX ANTE REGULATORY REGIME TO ADDRESS CONTENT-AGNOSTIC INJURIES

Although the content-agnostic framework provides a way for courts to weaken § 230's stronghold while preserving its fundamental purpose, it is not the most effective way to address the externalities of social products. Courts, as ex post institutions, are

on the platform; in contrast, in *Internet Brands*, the warning remedy neither arose from third-party content nor addressed on-platform third-party conduct.

[117] Although privacy settings affect visibility of third-party content, the mitigation does not implicate third-party content because the injury to a user stems from how the platform treats *that* user's content, not content produced by someone else. The primary harm of privacy violations is not that user *A* sees content from a user *B* he does not like, but rather that user *A*'s content is shown to user *B* without user *A*'s consent.

[118] *See generally* Order Granting in Part and Denying in Part Defendants' Motions to Dismiss, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047 (N.D. Cal. Nov. 14, 2023).

[119] *Id.* at 14.

[120] *Compare id.* at 16–19 (dismissing claims associated with the timing and clustering of notifications of third-party content, even though the remedy plausibly rests more in platforms' decision about *when* to send a notification, rather than on the third-party *content* for which the notification is being sent), *with id.* at 14–15 (allowing claims associated with platforms' failure to label filtered content to proceed, even though the remedy would inherently require platforms to treat certain types of third-party content differently during publication).

limited in the sorts of tools they can deploy to remedy content-agnostic injuries. The sorts of cases in which they can issue those remedies is further restrained by § 230's statutory language. Because there is "no regulatory apparatus" governing social platforms "to displace the ordinary tort law duty of care,"[121] some, including plaintiffs before courts today, have argued that courts should pay deference to plaintiffs bringing content-agnostic claims and dispense tort liability more generously.[122] However, judicial intervention is not without its drawbacks, and there is little reason to believe that simply tweaking the model of liability subject to safe harbors adopted by the United States to regulate social technologies will lead to preferred outcomes.

Rather, developing an ex ante regulatory regime via a new expert agency with statutory authority[123] over social platforms is more likely to address the root cause of plaintiffs' injuries and maximize societal welfare in the long run. An agency with the ability to impose transparency obligations onto platforms, coordinate research and third-party oversight, and affect platform incentives or their decisions to ship harmful features can be effective in ways that a court cannot. An appropriately empowered, structured, and staffed agency can provide a floor of protection for users as our collective understanding of social technologies evolves.

Part III.A discusses several reasons to disfavor an ex post liability regime or prefer an ex ante regulatory regime. Part III.B

[121] Kyle Langvardt, *Regulating Habit-Forming Technology*, 88 FORDHAM L. REV. 129, 167 (2019).

[122] *See generally* Matthew P. Bergman, *Assaulting the Citadel of Section 230 Immunity: Products Liability, Social Media, and the Youth Mental Health Crisis*, 26 LEWIS & CLARK L. REV. 1159 (2023); Nancy S. Kim, *Beyond Section 230 Liability for Facebook*, 96 ST. JOHN'S L. REV. 353 (2022); Kaidyn McClure, Comment, *A Case for Protecting Youth from the Harmful Mental Effects of Social Media*, 26 CHAP. L. REV. 325 (2022); Tyler Lisea, Comment, Lemmon *Leads the Way to Algorithmic Liability: Navigating the Internet Immunity Labyrinth*, 50 PEPP. L. REV. 785 (2023).

[123] Establishing statutory authority for the expert agency to act is critical to the success of any ex ante regulatory regime. *See, e.g.*, Rob Frieden, *Ex Ante Versus Ex Post Approaches to Net Neutrality: A Comparative Assessment*, 30 BERKELEY TECH. L.J. 1561, 1574 (2015). Clear statutory authority is also necessary to sidestep allegations of the state using "official speech to inappropriately compel" action by private-speech intermediaries and constitutional evasion. Will Duffield, *Jawboning Against Speech*, CATO INST. (Sept. 12, 2022), https://www.cato.org/policy-analysis/jawboning-against-speech; *see also* Genevieve Lakier, *Jawboning as a Problem of Constitutional Evasion*, KNIGHT FIRST AMEND. INST. AT COLUMBIA UNIV. (Oct. 13, 2023), https://perma.cc/XBW3-GJFH. This Comment recognizes, but does not address, the political and constitutional difficulties of establishing such a statutory authority.

outlines what an adequate ex ante regulatory regime could look like with respect to its mandate, powers, structure, and staffing.

A. Reasons to Disfavor an Ex Post Liability Regime or Prefer an Ex Ante Regulatory Regime

Given the breadth and simplicity of the liability-insulating provisions of the CDA, it is likely that social products (which integrate complex modeling, feed-ranking systems, algorithmic recommendations, and now technologies like generative artificial intelligence) have become too complex for the current all-or-nothing CDA framework. It is also evident that the evolution of the CDA into its current framework has imposed negative externalities on users beyond what was understood at the time of the CDA's drafting.[124] Although the content-agnostic framework provides a way for courts to impose liability on social platforms under the CDA within a bounded scope, there are several reasons to generally favor the use of an ex ante regulatory regime over an ex post liability regime to address the externalities presented by social products.

1. Incentivizing focus on the right part of the product-development funnel.

At many social platforms, product decisions are governed by metrics. Deciding which piece of content a user sees first or which change to a recommendations system is deployed is a balancing act involving a suite of metrics—how much time the user spends on the platform, how engaged the user is during that session, how the user feels when she is not using the platform, and more.[125] The goal metrics that platforms use when deciding whether to ship particular changes, what weight is given to those goal metrics,

[124] *See* Hassell v. Bird, 420 P.3d 776, 824 (Cal. 2018) (Cuéllar, J., dissenting) ("To the extent the Communications Decency Act merits its name, it is because it was not meant to be—and it is not—a reckless declaration of the independence of cyberspace.").

[125] *See, e.g.*, Rachel Mentz, *Likes, Anger Emojis and RSVPs: The Math Behind Facebook's News Feed—and How It Backfired*, CNN (Oct. 27, 2021), https://perma.cc/5GYT-Z85A (noting Facebook's use of the "meaningful social interactions" metric to rank content and the user-sentiment surveys and experiments that undergird the metric's mathematical definition); *Using P(Anger) to Reduce the Impact Angry Reactions Have on Engagement Ranking Levers*, DOCUMENTCLOUD (Sept. 4, 2020), https://www.documentcloud.org/documents/21601629-tier0_rank_ro_0920 (highlighting a change to Facebook's ranking models that was shipped based on movement in goal and countermetrics).

and whether adequate and appropriate countermetrics are utilized all shape how the social product functions.[126] Accordingly, seeking to reform platform conduct without understanding the incentive structure that led to the problematic product decisions is short-sighted.

For example, suppose that a platform can measure the number of peers a given user follows and that changing the color of the "Follow User" button makes users more likely to click the button. If the number-of-users-followed metric is a platform's sole goal metric, then the platform is likely to ship the change to the button's color because the change maximizes the platform's goal metric. If the platform utilizes five different goal metrics but gives the number-of-users-followed metric substantially more weight than the other four goal metrics in its analysis, then the platform is likely to ship the change to the button's color regardless of the change's negative impact on the other four goal metrics. If following more peers makes users more likely to post harmful content but the platform does not measure the prevalence of harmful content or include it in its suite of goal metrics, then the platform is likely to ship the change to the button's color without noticing that doing so increases the amount of harmful content circulating within the platform's ecosystem. This is the data-driven process that guides product development at most social platforms. If designing safer social products is the war, building the right suite of goal metrics and countermetrics is half the battle. Good features cannot survive bad metrics, and bad features can be elevated by bad metrics.

The primary drawback of ex post liability regimes governing social technologies is that ex post judicial remedies focus efforts on product outcomes rather than inputs. Courts can tell platforms *what* to do, but they cannot reform *how* platforms operate the way an expert agency can. An ex post regime does not necessarily incentivize building safer products so long as platform profit incentives and decision structures remain unchanged. Affecting the

[126] *See, e.g.*, *One-Go Summary Post for Recent Goaling and Goal Metric Changes for News Feed: Groups, US+CA DAP, Public Figures and Integrity Metrics*, DOCUMENTCLOUD (Mar. 9, 2021), https://www.documentcloud.org/documents/21748448-tier0_news_ro_0321 (highlighting some of the goal metrics used in Facebook's News Feed ranking and referencing integrity guardrail metrics); *A Proposal for Bold Experiments to Learn About Users and Craft Proxy Metrics for Integrity*, DOCUMENTCLOUD (2019), https://www.documentcloud.org/documents/23605722-tier0_civ_pr_1119 (discussing the difficulty a team at Facebook faced in launching product interventions to improve user experience without having the requisite metrics to justify their launch).

process by which social platforms make decisions or affecting the inputs into the decision process through the proposed ex ante regulatory regime is more likely to address the root of content-agnostic injuries.

This is particularly relevant for causes of action where, in response, companies may retain easier alternatives than reforming their profitable products. Take, for instance, a plaintiff's failure to warn claim. History is replete with instances where companies, when held liable for failing to warn consumers, have responded to liability by adding new warnings to their products without changing the underlying products.[127] The same may be true here. Platforms governed by ex post liability are more likely to build the product first and then determine whether warnings or disclosure are required, rather than focusing efforts on the inputs into product decisions that led them to build harmful products in the first place. If plaintiffs succeed in their failure to warn claims today, platforms' first objective tomorrow will be to update their Terms and Conditions or ship new website pop-ups warning users of the dangers of social products and requiring them to click an "I Agree" button to access their respective services. These changes are unlikely to affect consumer behavior—because virtually no one reads them[128]—and are unlikely to make social platforms safer, even though they protect defendant platforms from future liability. These changes prove similarly ineffective when applied

[127] A quintessential example of this phenomenon is *Liebeck v. McDonald's Restaurants, P.T.S., Inc.*, 1995 WL 360309 (N.M. Dist. Aug. 18, 1994), *vacated sub nom.* Liebeck v. Restaurants, 1994 WL 16777704 (N.M. Dist. Nov. 28, 1994). In the aftermath of the lawsuit, "[m]any McDonald's drive-thrus now have a sign warning, 'Coffee, tea and hot chocolate are VERY HOT!'" and "the lids of McDonald's hot beverage cups are now embossed with the words 'HOT! HOT! HOT!'" Kevin G. Cain, *And Now the Rest of the Story . . . the McDonald's Coffee Lawsuit*, 11 J. CONSUMER & COM. L. 14, 17 (2007). Some sources indicate that McDonald's did not change its behavior in response to liability; the chain continued to serve coffee at the same temperature as before the *Liebeck* case. *See Burger Chain Sued After Boy's Ordeal*, CAMBRIDGE NEWS ONLINE (June 22, 2007), https://web.archive.org/web/20090515122340/http://www.cambridge-news.co.uk/cn_news_huntingdon/displayarticle.asp?id=180135.

[128] In a study, 98% of users signing up for a fictitious social network consented to terms and conditions that contained an agreement to provide their first-born child as payment for access to the website. *See* Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, 23 INFO., COMMC'N & SOC'Y 1, 25 (2018). Users generally take the path of least resistance when facing obstructions such as interstitials and can be easily led to "making decisions they are likely to regret or misunderstand," such as affirming receipt of a product warning, by "prompting impulsive System 1 decision-making and discouraging deliberative System 2 decision-making." Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEGAL ANALYSIS 43, 52 (2021).

to the fact patterns presented by plaintiffs before courts today. Many plaintiffs were children when they started using social media and were therefore unlikely to understand or heed product warnings.[129] Other plaintiffs created social media accounts either without their parents' knowledge or against their parents' explicit wishes, making it unlikely that product warnings would be seen by adults who could make and enforce an informed decision.[130] Although additional warnings and revised platform representations may be helpful on the margins, they are unlikely to address the root cause of content-agnostic injuries because these interventions do not affect how social products are built in the first place.

2. Difficulty in drawing intelligible and defensible lines.

Given the complexity of product features and the myriad of ways by which third-party content and platform features can interact to cause injury to users, overreliance on an ex post liability regime to govern social platforms may lead courts to entangle themselves in an attempt to avoid opening the floodgates of litigation. This difficulty in drawing intelligible and defensible lines generally materializes in three ways. First, courts may struggle to draw lines around the scope of injury-causing social products. Imagine a digital product with many of the same attributes as social platforms—highly engaging, optimized "to addict and condition users to engage in repetitive content-consuming actions,"[131] and designed to prey on the "chemical reward system of users' brains (especially young users)."[132] However, this digital product contains no third-party content; rather, it shows users a series of colorful shapes and allows users to interact with them. If plaintiffs prevail in the cases before courts today, should this hypothetical digital product also be liable for inflicting addiction-related injuries onto users? This presents a problem, as this hypothetical product seems functionally indistinguishable from Candy Crush, one of the most popular mobile games available today.[133] If social

[129] Aranda Complaint, *supra* note 77, at 51.

[130] *See* Gill Complaint, *supra* note 74, at 54–55.

[131] Murden Complaint, *supra* note 71, at 28.

[132] Estevanott Complaint, *supra* note 98, at 24.

[133] There are numerous reported stories of user harm arising from Candy Crush's addictive design, including users leaving "their children stranded at school, abandon[ing] housework[,] and even injur[ing] themselves." Eliana Dockterman, *Candy Crush Saga: The Science Behind Our Addiction*, TIME (Nov. 15, 2013), https://perma.cc/G7BG-ST9N. Candy Crush is not unique in this regard. *See, e.g.*, Michelle Boudin, *Medical Professionals: Video Games Like Fortnite Can Be As Addictive As Heroin*, WGRZ

products can be held liable for being designed to addict, then many other digital products should face the same legal risks. While a discussion of digital addiction is long overdue, introducing legal liability in this fashion with no reasonable limiting principle can be destabilizing given the commercial significance and entrenchment of the modern "attention economy."[134]

Second, courts may have difficulty drawing lines around the scope of users' injuries. For example, plaintiffs attribute their injuries in part to push notifications.[135] However, push notifications can be turned off by users on both the application and device levels. In adjudicating this claim, courts will likely need to delineate where the platforms' duties end and where those of product users begin.[136] Courts will similarly need to distinguish users who could control or limit their use of social products but chose not to from users who could not. Holding that platforms are per se unreasonably addictive also becomes difficult when one considers the ubiquity of social media; billions of individuals who use social media platforms daily do not become compulsive users. Ex post adjudication faces challenges in determining what level of loss shifting is adequate, appropriate, or welfare maximizing.

Third, courts will be challenged to draw lines that do not undermine or contradict other lines governing social platforms. For example, to rule in favor of the Florida Attorney General in her aforementioned case against Meta,[137] the court must find that Meta is not sufficiently aggressive in policing its platforms. But in a different case, the Florida Attorney General is concurrently defending[138] a Florida statute that requires social platforms not

(Sept. 15, 2018), https://perma.cc/4W86-84SZ. *See also* Wilson v. Midway Games, Inc., 198 F. Supp. 2d 167, 170 (D. Conn. 2002) ("[Plaintiff] [ ] alleges that Midway designed Mortal Kombat to addict players to the exhilaration of violence, and specifically targeted a young audience, intending to addict them to the game.").

[134] Lexie Kane, *The Attention Economy*, NIELSEN NORMAN GRP. (June 30, 2019), https://perma.cc/FEH5-6AWY.

[135] *See, e.g.*, Chathams Complaint, *supra* note 82, at 15; Arizona Complaint, *supra* note 85, at 51; Aranda Complaint, *supra* note 77, at 56; Gill Complaint, *supra* note 74, at 40.

[136] *See* Douglas H. Cook, *Personal Responsibility and the Law of Torts*, 45 AM. U. L. REV. 1245, 1253 (1996) ("[I]f a plaintiff could reasonably take action to eliminate the damages, rather than merely minimizing them, the law would require that he or she do so." (emphasis omitted)).

[137] *See generally* Florida Complaint, *supra* note 86.

[138] *See generally* NetChoice, LLC v. Att'y Gen., 34 F.4th 1196 (11th Cir. 2022), *cert. granted in part sub nom.* Moody v. NetChoice, LLC, 144 S. Ct. 478 (2023), and *cert. denied sub nom.* NetChoice, LLC v. Moody, 144 S. Ct. 69 (2023), and *vacated and remanded sub nom.* Moody v. NetChoice, LLC, 144 S. Ct. 2383 (2024).

to moderate content created by certain categories of users.[139] To rule in favor of the Florida Attorney General in this latter case, the court must adopt the view that Meta should do less to interfere with what users see on its platforms. The same litigant is arguing in one case that Meta should not be allowed to deplatform certain explicit and harmful content on its platforms while arguing in a different case that Meta allows too much explicit and harmful content to propagate and injure users. Should courts fail to grasp the evolving regulatory landscape and move in lockstep with other courts around the country on factual findings and burdens placed on social platforms, social platforms may find themselves in paradoxes they cannot unravel without further litigation. The proposed ex ante regime centralizes rulemaking, which may help avoid these line-drawing difficulties.

3. Overcorrection risks.

It is generally presumed that liability incentivizes firms to invest in product safety and quality.[140] However, liability, especially when it turns on what a firm knew about its products and when it knew it, can disincentivize disclosure. Opening the doors to ex post liability may lead platforms to overcorrect and more tightly guard internal data about their operations. This is dangerous because it blunts the efficacy of regulators, impairs the studies of researchers seeking to understand social products and their externalities, and diminishes the ability of users to take informed precautions.

Social platforms' ability to lock down their internal data is tremendous,[141] and several attempts by researchers to work around existing restrictions on data access have been unsuccessful.[142] As cases proceed, the risk of further data lockdown is both

[139] FLA. STAT. §§ 106.072, 501.2041 (2023). Among placing other restrictions on social platforms, the statute limits "platforms' ability to engage in deplatforming, censorship, shadow-banning, or post prioritization" and "prohibit[s] platforms from deplatforming or restricting the content of political candidates or 'journalistic enterprises.'" VALERIE C. BRANNON, CONG. RSCH. SERV., LSB10748, FREE SPEECH CHALLENGES TO FLORIDA AND TEXAS SOCIAL MEDIA LAWS 2 (2022).

[140] *See, e.g.*, A. Mitchell Polinsky & William P. Rogerson, *Products Liability, Consumer Misperceptions, and Market Power*, 14 BELL J. ECON. 581, 584 (1983).

[141] *See* DiResta et al., *supra* note 112.

[142] *See, e.g.*, Heidi Ledford, *Researchers Scramble as Twitter Plans to End Free Data Access*, NATURE (Feb. 14, 2023), https://www.nature.com/articles/d41586-023-00460-z; Alex Engler, *Platform Data Access Is a Lynchpin of the EU's Digital Services Act*, BROOKINGS INST. (Jan. 15, 2021), https://perma.cc/5M59-5CP9 ("Despite the widespread impression of far right-wing news dominating Facebook, it's actually impossible to know

present[143] and significant[144] for a few reasons: First, psychological and sociological research of social technologies is still a burgeoning field. Second, the risks that social products present to individuals and communities are unknown, complex, and difficult to measure from the outside looking in. And third, a handful of industry titans have exclusive access to data about the effects of their technologies and a lot to lose.

Although "[e]x ante rules . . . have the potential to trigger false positives (i.e., a determination that a rule violation has occurred, despite the absence of harm to consumers and competitors),"[145] these risks can be mitigated if rules are made by an informed regulator with statutorily mandated access to platform data in partnership with industry. In contrast, risks associated with platform data lockdowns are present and pose a substantially higher risk to public safety because there are few alternatives to platform data transparency. Without this data, efforts to understand social technologies would be stymied, making it difficult to predict and proactively mitigate the next long-tailed harm that may arise from the use of social products. Without this data, researchers and regulators operate purely in a reactive mode, addressing harms only after they have reared their ugly heads.

4. Institutional expertise.

Ex ante regulatory regimes are preferrable when a potential regulator may have superior access to information and subject-matter expertise relative to the average plaintiff in an ex post regime. Social technologies represent such an instance.

Because generalist judges rely on litigants' framing of social technologies in word-limited briefs, they can fall victim to two sorts of risks. The first risk is that courts may oversimplify how

if that's the case with currently available data. . . . The voluntary measures taken by the internet platforms to enable researcher access are simply not working.").

[143] For example, in response to a series of leaks of internal documents, Meta has vigorously locked down internal access to research conducted by its Integrity teams. Alex Heath, *Meta Goes into Lockdown*, VERGE (Nov. 16, 2021), https://perma.cc/29DK-D58M.

[144] *See, e.g.*, SARA BUNDTZEN & CHRISTIAN SCHWIETER, INST. FOR STRATEGIC DIALOGUE, ACCESS TO SOCIAL MEDIA DATA FOR PUBLIC INTEREST RESEARCH: LESSONS LEARNT AND RECOMMENDATIONS FOR STRENGTHENING INITIATIVES IN THE EU AND BEYOND 6 (2023) ("[A]ccess to social media data has become a prerequisite to investigating and understanding most contemporary problems 'in the real world'—whether in the context of election cycles, foreign interference, public health, or societal attitudes towards climate change, migration or LGBT+ rights.").

[145] Frieden, *supra* note 123, at 1584.

social platforms function. For example, the Supreme Court left the door open for content-agnostic claims in *Taamneh*, calling recommendations algorithms "agnostic as to the nature of the content" and implying that they rely only on noncontent signals.[146] The Court also noted that "recommendation algorithms . . . are infrastructure," thereby drawing a line between recommendations algorithms as content-neutral pipes and the content flowing through those pipes.[147] It appears that the Court is not fully clear on this delineation, however, as at other points, the Court notes that recommendations algorithms do consider "information about the . . . content being viewed."[148]

Such a content-infrastructure boundary does not exist in practice. It is difficult to separate a complex recommendations model from the underlying third-party content because the features that serve as inputs into the model can depend on the third-party content,[149] the model's output consists entirely of third-party content, and the metrics evaluating the performance of the model depend on the third-party content.[150] Training a recommendations model involves selecting goal metrics (e.g., the amount of time a user spends on the social platform) and countermetrics (e.g., the prevalence of hate speech, borderline misinformation, or other low-quality content) to optimize, and these metrics are entirely reliant on measurements derived from third-party content. Recommendations algorithms are less like a neutral pipe through which content flows (as the Court describes), and more like a dynamic funnel that gets wider or narrower depending on what content flows through it.

The second risk is that courts may oversimplify social ecosystems generally. It is difficult to separate third-party content from a platform's engagement model when one views platforms as a social marketplace with users as both active producers and consumers of social media content, rather than just as passive

[146] *Taamneh*, 143 S. Ct. at 1227.

[147] *Id.*

[148] *Id.* at 1216.

[149] "Data about [ ] post content" and "about the media, like photo or video, contained in the post" factor into Facebook's Feed ranking system. *Our Approach to Facebook Feed Ranking*, FACEBOOK TRANSPARENCY CTR. (June 29, 2023), https://perma.cc/5PQG-6D4R.

[150] Daphne Keller, *What the Supreme Court Says Platforms Do*, LAWFARE (Sept. 14, 2023), https://perma.cc/B6ME-YNMY ("Algorithms' success, as judged by platforms' human evaluators using frameworks like Google's Search Quality Evaluator Guidelines, explicitly depends on what content they surface.").

consumers.[151] Platforms may draw users to sensationalist or harmful content, but in a social marketplace, users are also incentivized to create sensationalist or harmful content.[152] An ex post regime focused entirely on the demand side of the social marketplace misses half of the problem. These nuances can be difficult to litigate given their deeply technical nature and complex sociological causative relationships.[153]

For example, consider mental health injuries caused by exposure to third-party content promoting eating disorders on a social platform. This problem exists in a two-sided marketplace: On one side, some users produce harmful content for many reasons. On the other side, some users consume this harmful content, whether by choice or inadvertently. How should the platform address this issue? The platform could address the demand side by making it more difficult for consumers to discover the harmful content in question. (For example, Instagram redirects users who search for disordered-eating content to a community support line.[154]) However, doing so without addressing the supply side pushes producers to find ways around these restrictions by using alternative hashtags or intentional substitutions of letters with similar looking characters in banned phrases.

[151] Brendan Nyhan et al., *Like-Minded Sources on Facebook Are Prevalent but Not Polarizing*, 620 NATURE 137, 143 (2023) (demonstrating that reducing Facebook users' algorithmic exposure to content from like-minded sources had little effect on characteristics like affective polarization, ideological extremity, or susceptibility to misinformation because doing so "cannot fully counteract users' proclivity to seek out and engage with congenial information"). The study noted that although "popular narratives blam[e] social media echo chambers for the problems of contemporary American democracy," these sorts of algorithmic changes "do not seem to offer a simple solution for those problems." *Id.*

[152] Brief of the Integrity Institute, *supra* note 61, at 14 ("Put another way, optimizing for engagement means that harmful content will rise to the top of recommendation feeds. This happens in part because in their capacity as creators—rather than consumers—of content, users have strong incentives to post content that garners more engagement from other users.").

[153] Some courts have recognized their technical limitations in analyzing the evidence presented to them and identifying lasting solutions when addressing social technologies. *See, e.g.*, Force v. Facebook, Inc., 934 F.3d 53, 88 (2d Cir. 2019) (Katzmann, J., concurring in part and dissenting in part) ("Whether, and to what extent, Congress should allow liability for tech companies . . . is a question for legislators, not judges."); Transcript of Oral Argument at 45–46, *Gonzalez*, 143 S. Ct. 1191 (2023) (No. 21-1333) ("[W]e're a court. We really don't know about these things. . . . [T]hese are not like the nine greatest experts on the internet.").

[154] *See* Jacob Shamsiam, *Instagram Is Cracking Down on Its Pro-Anorexia Community*, BUS. INSIDER (Dec. 12, 2018), https://perma.cc/GF4D-DZKD.

Alternatively, the platform could address the supply side by more rigorously screening content by producers before it is published. However, without addressing the demand side, a heavy-handed platform may inadvertently push both producers and consumers to migrate to alternative, less scrupulous internet services where consumers may be exposed to even more harmful content. To successfully address the issue, the platform must engage in a delicate balancing act on both sides of the market. Monetization and other factors further complicate these producer-platform and consumer-platform relationships. As with all social technologies, when one changes one part of the social network, complex sociological forces can lead to outsized unintended consequences in other parts of the network. Courts should be hesitant to make changes when they cannot foresee or control those consequences.

In contrast, a regulatory regime that unites academics, researchers, industry representatives, and other experts on social technologies is more likely to understand the Rube Goldberg machine of social products and consider how changes mandated by regulations in one part of social-platform design are likely to affect other parts of the social ecosystem.[155] This is important as users are not monolithic (a change made to a recommendations system is likely to affect users with different digital literacy rates or social media use patterns differently), and an ex post liability regime provides redress for only the complaining user, not others.[156] Furthermore, an ex post regime is ideal for circumstances where the types of injuries and paths to injury are varied because it is hard to provide ex ante guidance to govern nonuniform circumstances.[157] This is not true of social platforms, which operate at massive scale with immense complexity, but in uniform

[155] *See, e.g.*, Steven Shavell, *Liability for Harm Versus Regulation of Safety*, 13 J. LEGAL STUD. 357, 369 (1984) [hereinafter Shavell, *Liability Versus Regulation*] ("[R]egulatory authority may not suffer an informational disadvantage, but instead may enjoy a positive advantage relative to private parties. Notably, . . . a regulatory agency may have better access to, or a superior ability to evaluate, relevant . . . knowledge.").

[156] Litigants are incentivized to request redress for their specific circumstances but have less incentive to evaluate whether their requested judicial remedies are welfare maximizing more broadly. *See, e.g.*, *id.* ("In certain contexts[,] information about risk will not be an obvious by-product . . . but rather will require effort to develop or special expertise to evaluate. In these contexts[,] a regulator might obtain information by committing social resources to the task, while private parties would have an insufficient incentive."). Therefore, ex post adjudication may inadvertently optimize for a local, not global, maxima of care.

[157] *See* Steven Shavell, *A Model of the Optimal Use of Liability and Safety Regulation*, 15 RAND J. ECON. 271, 274 (1984) [hereinafter Shavell, *Liability and Safety Regulations*].

ways.[158] The complexity and relative uniformity of social-product design indicates that an ex ante regime is better suited to collect data and address the externalities of social products.

5. Evading liability due to distributed injuries.

If plaintiffs are correct in asserting that social technologies as designed are inherently harmful and injure all users to some degree, then individually initiated litigation is an unattractive solution to the broader problem. Given the costs of litigation, aggrieved parties have little incentive to initiate legal action until their injuries in aggregate exceed some threshold.[159] This allows tortfeasor platforms to evade liability for every other user whose injuries do not exceed that threshold. Take, for instance, a user who has developed anxious tendencies due to his social media use. Unlike the plaintiffs before courts today, this user is not presenting symptoms severe enough to require medication or counseling. This user will likely never sue social platforms for his injuries even though he was harmed by their products. "[L]iability does not [always] create sufficient incentives to take appropriate care because of the possibility that parties . . . would not be sued for [harm done]."[160] Furthermore, when potential injuries to users include death and lasting impacts on mental health, injuries that are hardly made whole by monetary damages, an ex ante regime may be preferrable for the simple reason that an ounce of prevention is worth a pound of cure.

B. Designing an Adequate Ex Ante Regulatory Regime

There are two different ways that regulators can go about fulfilling their objectives. One approach is coregulation, defined by the European Union as "the mechanism whereby a [ ] legislative act entrusts the attainment of the objectives defined by the legislative authority to parties which are recognized in the field (such

[158] Arielle Pardes, *All the Social Media Giants Are Becoming the Same*, WIRED (Nov. 30, 2020), https://perma.cc/AM83-3BNN.

[159] *See* Shavell, *Liability Versus Regulation*, *supra* note 155, at 363. Note that class action lawsuits can offset this risk to a degree, though not in its entirety. *Id.*; *see also* John H. Beisner, Jordan M. Schwartz & Paden Gallagher, *Unfair, Inefficient, Unpredictable: Class Action Flaws and The Road to Reform*, U.S. CHAMBER OF COM. INST. FOR LEGAL REFORM (Aug. 2022), https://perma.cc/2ZDB-NES7 (raising valid structural and efficiency flaws in class actions and their ability to deter harmful conduct, despite the institute's obvious bias).

[160] Shavell, *Liability and Safety Regulations*, *supra* note 157, at 271.

as economic operators, the social partners, non-governmental organisations, or associations)."[161] The other approach is command-and-control regulation, in which the state issues firm directives and guidance "which is often assumed to take a particular form, that is the use of legal rules backed by [ ] sanctions."[162] Blending elements from a coregulatory regime and elements from a command-and-control approach may be the most promising avenue for mitigating plaintiffs' injuries and addressing the harmful effects of social products. Specifically, the regime best suited to address the root cause of content-agnostic injuries as society's understanding of social products evolves would have the power to (1) "determine the nature and extent of the negative externality caused by" social technologies; (2) define broad regulatory objectives for platforms; (3) allow platforms to experiment with product-specific solutions; and (4) mandate the "precise regulatory response[s] that [are] most efficient."[163] Additionally, given that Democrats and Republicans broadly agree that technology companies should be regulated but disagree vehemently on what that regulation should look like,[164] delegation to an independent and technically competent agency may be the only feasible and politically palatable solution on the table.

1. Agency mandate and powers.

In designing an adequate regulatory response to the harms of social products, one should first ask: What is the regulatory regime's mandate, and what regulatory powers does the regime need to fulfill that mandate? In effect, what does the regulator need to be able to do? Balancing the specific challenges posed by social products and the dangers of an overly powerful institution governing modern manifestations of speech, a regulator should be able to study social products and establish guardrails that indirectly affect platform incentives or their decisions to ship harmful features, but not be able to directly proscribe content-level decisions. In practice, this looks like an expert agency with the power to impose transparency obligations onto social platforms like

[161] Interinstitutional Agreement on Better Law-Making, 2003 O.J. (C 321/1) 3.

[162] Julia Black, *Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World*, 54 CURRENT LEGAL PROBS. 103, 105 (2001).

[163] Kyle D. Logue, *In Praise of (Some) Ex Post Regulation: A Response to Professor Galle*, 69 VAND. L. REV. EN BANC 97, 103 (2016).

[164] Justin Sherman, *There Is No Bipartisan Consensus on Big Tech*, WIRED (Oct. 13, 2021), https://www.wired.com/story/there-is-no-bipartisan-consensus-on-big-tech/.

those imposed under the European Union's Digital Services Act[165] (DSA) and use its learnings to affect the process by which social platforms make decisions. Among other obligations, the DSA requires platforms to conduct annual assessments; issue network transparency reports; conduct yearly assessments on the impacts of their design, algorithms, advertising, and terms of service on a range of societal issues; make certain platform data available to independent auditors and vetted researchers; and propose and implement concrete remedial measures under the scrutiny of independent auditors, vetted researchers, and an expert agency.[166] Granting the expert agency power to impose many of the same research and reporting obligations onto social platforms could go a long way toward building an informed and effective regulator.

A yet-unaddressed source of content-agnostic injuries is social platforms' incentives and product-launch criteria.[167] When making decisions about which product features will ship, social platforms often overindex on metrics that capture profitability and underrepresent countermetrics that capture user safety or network health.[168] An expert agency, in partnership with researchers and civil society, with the power to define specific countermetrics, require platforms to consider those countermetrics in product decisions, and provide guidance on how to adequately weigh those countermetrics against goal metrics[169] could meaningfully affect product outcomes and make social products safer for all users.[170]

The expert agency could also reasonably compel platforms to internalize at least some costs associated with their harmful features by publishing standards and best practices for feature

[165] Regulation 2022/2065 of the European Parliament and of the Council of 19 Oct. 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277).

[166] *See Questions and Answers: Digital Services Act*, EUR. COMM'N (Apr. 25, 2023), https://perma.cc/9BD2-NLKH; David Morar, *The Digital Services Act's Lesson for U.S. Policymakers: Co-regulatory Mechanisms*, BROOKINGS INST. (Aug. 23, 2022), https://perma.cc/YSL9-FEHW.

[167] *See infra* Part III.A.1.

[168] *See, e.g.*, Andrew Mauboussin, *Moving Beyond Engagement: Optimizing Facebook's Algorithms for Human Values*, SURGE AI (Feb. 10, 2022), https://perma.cc/NM8B-WU75.

[169] Due to variability in platform features, content types, and user characteristics, the question of goal-metric and countermetric trade-offs invites greater complexity. An expert agency is best suited to make these determinations on a case-by-case basis.

[170] Academics have identified at least sixty-four additional product interventions that could be deployed by an expert agency on a case-by-case basis to make social platforms safer. *See, e.g.*, *Focus on Features*, INTEGRITY INST., https://features.integrityinstitute.org.

design, facilitating and coordinating independent research on social technologies, auditing platform operations, and imposing sanctions for noncompliance. By first defining regulatory objectives and standards, this expert agency would grant platforms a degree of freedom by allowing them to tailor their compliance to product-specific conditions. For example, suppose that the expert agency mandates that platforms maintain the prevalence of sexual content in minors' feeds below 0.1%. In this situation, Reddit may choose to comply by hiring more content moderators and more aggressively filtering such content, while Instagram may comply by algorithmically increasing the reach of trusted, high-quality content in minors' feeds. In this manner, Reddit reaches the agency's goal by reducing the numerator while Instagram does so by increasing the denominator. Such an agency establishes a floor of protection for users while incentivizing platforms to experiment and develop new solutions to long-standing product-safety issues. Once particularly effective solutions are identified, the agency could then issue standards to mandate that all social platforms integrate those solutions into their products.

One important limit must exist on this agency's power: the agency must lack the authority to proscribe specific content-level decisions. The inability of the proposed agency to order platforms to host or delete any specific piece of content is likely necessitated by the First Amendment. Codifying this limit on agency power is also made necessary by the fact that "it will often be far more rational for private companies to comply with even relatively soft government pressure to suppress or keep up speech than it will be for them to contest it."[171] Because platforms profit "from serving a great deal of speakers," "[t]he cost they face . . . of removing particular speakers or speech acts from their platforms will often be extremely minimal."[172] In contrast, because the agency would "possess a great deal of discretionary power [it could] wield to the benefit or the detriment of" the platforms it regulates, platforms would have a strong incentive to placate their regulator.[173] Giving the agency the power to set direction and standards, but not dictate whether specific pieces of content are hosted or deleted, ensures a baseline level of platform safety without turning social

[171] Lakier, *supra* note 123.

[172] *Id.*

[173] *Id.*

platforms into state-run media. Existing First Amendment jurisprudence likely provides a sufficient guardrail against this risk,[174] and in practice, may strongly incentivize the agency to focus its efforts on content-neutral product interventions and remedies to content-agnostic injuries, neither of which implicates the First Amendment in the first place.

Striking the right balance between a coregulatory and command-and-control regime is difficult because doing so requires "regulators to have an enormous amount of information that they typically do not have at their disposal."[175] Furthermore, adopting provisions of the DSA as is will likely engender tension with existing U.S. law.[176] But "[j]ust because command-and-control regulation is difficult to implement effectively and requires a great deal of information on the part of the regulator, [ ] does not mean that it is never the best regulatory instrument."[177] Such a regime "works especially well to provide a given 'floor' of protection from certain negative externalities."[178] It is also preferrable to an ex post regime because an expert agency is able to provide ongoing guidance and regular nudges to social platforms, rather than allowing long periods of laissez-faire operation interrupted by bursts of intense judicial intervention. Establishing a floor of protection is valuable because that floor changes as our collective understanding of social technologies evolves.

2. Agency structure and staffing.

Appropriately structuring and staffing the proposed agency is critical to its success because these choices impact the agency's

[174] *See* VALERIE C. BRANNON & ERIC N. HOLMES, CONG. RSCH. SERV., R48751, SECTION 230: AN OVERVIEW 43–48 (2024) (addressing First Amendment issues with proposals to reform § 230); *see also* VICTORIA L. KILLION, CONG. RSCH. SERV., IF12308, FREE SPEECH: WHEN AND WHY CONTENT-BASED LAWS ARE PRESUMPTIVELY UNCONSTITUTIONAL 1 (2023) (providing an overview of permissible content-based regulation); VICTORIA L. KILLION, CONG. RSCH. SERV., IF11072, THE FIRST AMENDMENT: CATEGORIES OF SPEECH 2 (2019) (discussing the historically "unprotected" categories of speech, including obscenity, that may be fair grounds for agency regulation).

[175] Jon D. Hanson & Kyle D. Logue, *The Costs of Cigarettes: The Economic Case for Ex Post Incentive-Based Regulation*, 107 YALE L.J. 1163, 1265 (1998). Note that the proposed regulatory regime requires that the regulator be able to impose transparency and disclosure obligations onto social platforms to function.

[176] *See* Ioanna Tourkochoriti, *The Digital Services Act and the EU as the Global Regulator of the Internet*, 24 CHI. J. INT'L L. 129, 144–146 (2023) (discussing, for instance, potential conflicts between some obligations imposed on platforms under the DSA and U.S. free-expression standards).

[177] Logue, *supra* note 163, at 105.

[178] *Id.* at 106.

behavior and performance. There are a few considerations to bear in mind. First, the proposed agency will likely conduct a wide breadth of activities under its mandate to achieve its objectives. Second, it must manage a web of complex and interdependent relationships between actors that will both influence the agency and be governed by it. For example, social platforms are subject to the agency's rules but are also key participants in developing those rules. Academics and civil society serve both as consumers and potential auditors of data made available by social platforms under transparency obligations imposed by the agency. Given these considerations, the proposed agency must have two key qualities: independence and technical competence.

Several provisions in the agency's enabling legislation could help foster regulatory independence. These include "restricting politicians from removing agency heads except for 'good cause'; [ ] creating multi-member boards to head the agency; [ ] establishing set tenures for board members; [ ] ensuring partisan board balance; . . . making the agency self-funded, where the agency collects fees from the regulated industry for the duties it performs"; and choosing "longer-term civil servants" as agency heads.[179] Staffing the agency with career professionals and technologists and placing civil society, independent researchers, industry, and the public at the heart of any rulemaking authority can help relieve the political pressure that such an agency will undoubtedly face.

Independence, in turn, establishes the foundation for the agency to build technical competence. By limiting the ability for political actors to direct the agency toward partisan ends or influence agency decisions, "personnel in the [agency] may become more willing to invest effort in the analysis leading to that decision" and thereby gain more expertise.[180] Note that this creates a virtuous cycle: "independence leads to the development of expertise, and expertise can become a source of independence . . . , as the more expertise [the agency] has relative to others, including lawmakers, the more authority it has."[181] Independence and technical expertise together create "a stable environment for the regulated industry and more durable policy decisions,"[182] a

[179] Christopher Carrigan & Lindsey Poole, *Structuring Regulators: The Effects of Organizational Design on Regulatory Behavior and Performance*, PENN PROGRAM ON REGUL. at ii, 3 (June 2015), https://perma.cc/UNL5-PFNB.

[180] *Id.* at 9.

[181] *Id.*

[182] *Id.* at iii.

preferrable outcome given the scope, complexity, and significance of social technologies in modern life.

One concern with such an agency is the potential for regulatory capture.[183] Social platforms are large companies with large lobbying budgets, and many technologists with relevant experience who may staff the proposed agency will likely work for these platforms at some point in their careers. The risk of a revolving door may undermine agency efficacy because "revolving doors raise concerns that: (i) prior experience in industry makes [agency] personnel unduly sympathetic to industry's interests; or (ii) [agency] personnel go easier on violations to curry favor with future employers."[184] However, structural and procedural options exist to mitigate this risk. For example, in Europe, compliance with the DSA is audited by independent commercial and professional auditors with privileged access to internal data and systems.[185] This reduces the risk of regulatory capture because it is harder for a platform to capture both the agency and the auditor. Additionally, in the United Kingdom, the Digital Regulation Cooperation Forum (a collective comprised of the Competition and Markets Authority, Information Commissioner's Office, Office of Communications, and the Financial Conduct Authority)[186] has proposed a multitier framework to auditing social products involving governance, technical, and empirical auditing.[187] By asking different parties to conduct different kinds of audits of social products, the agency leaves noncompliant platforms little room to hide.

For example, suppose that YouTube is noncompliant with the proposed agency's standards on terrorist content, which mandate that platforms take sufficient steps to reduce the prevalence of terrorist content to below 0.1%. Also, suppose that under the agency's authority, the agency conducts a governance audit, a trusted nongovernmental organization (NGO) conducts a technical audit, and YouTube is required to publish certain prevalence metrics in quarterly transparency reports. There are

[183] *Id.* at 3.

[184] Ed deHaan, Simi Kedia, Kevin Koh & Shivaram Rajgopal, *The Revolving Door and the SEC's Enforcement Outcomes: Initial Evidence from Civil Litigation*, 60 J. ACCT. & ECON. 65, 66 (2015).

[185] Digital Services Act, 2022 O.J. (L 277) art. 37.

[186] *The Digital Regulation Cooperation Forum*, GOV.UK (Mar. 10, 2021), https://perma.cc/EKY6-P438.

[187] *Auditing Algorithms: The Existing Landscape, Role of Regulators and Future Outlook*, GOV.UK (Sept. 23, 2022), https://perma.cc/H6TD-TJHZ.

multiple avenues by which YouTube's noncompliance is detected: In conducting the governance audit, the regulator may notice that YouTube has not established a team tasked with addressing terrorist content on its platform. In conducting the technical audit, the NGO may notice that the code supporting YouTube's recommendations system does not check for terrorist content when suggesting new videos to users. YouTube's transparency report will show that the prevalence of terrorist content on the platform is 1%, higher than the agency's 0.1% cap. This distribution of responsibility across multiple parties blunts the potential for and the effectiveness of regulatory capture in this context.

Note that regulatory bodies with such a structure and mandate are not novel. Several regulators already exist in the United States with a similar suite of factfinding, rulemaking, and enforcement powers as the proposed agency. Many of these regulators are also tasked with regulating the conduct of a narrow set of industries, like the scope of the proposed agency. Agencies like the Nuclear Regulatory Commission, Consumer Financial Protection Bureau, Food and Drug Administration, and Securities and Exchange Commission could serve as different models for how the proposed agency could be structured, staffed, and operated.[188]

CONCLUSION

The ways in which social platforms invite, intersect with, and influence third-party speech are complex. Given § 230's broad statutory language, plaintiffs seeking to hold social platforms accountable for the externalities of their products have an incentive to plead artfully to circumvent premature dismissal. This Comment's novel content-specific, content-dependent, and content-agnostic taxonomy provides a useful tool to categorize claims raised by plaintiffs. Additionally, this Comment formalizes

[188] The proposed agency could even be nested under the umbrella of existing agencies like the Federal Communications Commission (FCC) or Federal Trade Commission (FTC), though it will likely need to exist as a separate entity from those specific agencies. This is due to the specialized focus, specific staffing needs, and expertise needed by the agency to fulfill its mandate, as well as the specific statutory authority needed to compile the agency powers identified earlier into a single regulatory body (no exact parallel to which is found in the FCC's or FTC's enabling statutes). *See* Communications Act of 1934, Pub. L. No. 73-416, 48 Stat. 1064 (codified as amended at 47 U.S.C. § 151 et seq.); Federal Trade Commission Act of 1914, Pub. L. No. 63-203, 38 Stat. 717 (codified as amended at 15 U.S.C. § 41 et seq.).

a remedies test to help courts separate legitimate content-agnostic claims from those in name only.

Recognizing that content-agnostic injuries propagated by social products are material but not yet fully understood, this Comment also advocates for an ex ante regime that blends a coregulatory and command-and-control approach as a preferred alternative to ex post liability and the right vehicle to address social-product externalities. An expert agency tasked with refining product norms, defining broad regulatory objectives for private platforms, and establishing a floor of precaution is better suited to focus intervention on product inputs and more likely to reform the root design decisions that engender harmful products. An ex ante regime is also more likely to sidestep line-drawing difficulties and regulatory paradoxes, avoid platform overcorrection, have superior access to information and subject-matter expertise relative to the average plaintiff, and account for those injured who may be unable or unwilling to sue.

Social technologies simultaneously represent human connection at its best and humanity's vices at their worst. Our discussion of social technologies and our approach to governing them ought to be no less technical or nuanced than the technologies themselves are.