Algorithmic Armor: Rethinking Section 230's Protection of Platform Design*

INTRODUCTION

Social media platforms no longer merely host content—they engineer attention.[1] Platforms like Facebook have evolved from virtual notice boards into sophisticated behavioral environments, using algorithmic tools to shape what users see, what they feel, and even how they act.[2] What began as a medium for interpersonal communication[3] has become a powerful system of influence that trades on emotional vulnerability and ideological division for profit.[4] This transformation raises novel legal questions, particularly when algorithmic systems lead to tangible harm.

1. *See* Kevin Driscoll, *A Prehistory of Social Media*, 38 ISSUES SCI. & TECH. 20, 20–23 (2022), https://issues.org/prehistory-social-media-modem-world-driscoll/ [https://perma.cc/R6M3-QUZ7] (recalling the early internet's user-initiated bulletin-board systems and strictly chronological feeds); Surabhi Parida, *Evolution of Social Media Algorithms: The Invisible Hand Guiding Our Online Experience*, RESO (Oct. 6, 2024), https://resoinsights.com/insight/evolution-of-social-media-algorithms/ [https://perma.cc/8LAK-SCHU] (describing how early social-media feeds shifted from reverse-chronological to relevance-based algorithmic ranking); Ian MacRae, Harris Eyre, Andy Keller & Sandi Chapman, *How Social Media Is Changing Our Brains*, DALL. MORNING NEWS, https://www.dallasnews.com/opinion/commentary/2021/12/05/how-social-media-is-changing-our-brains/ [https://perma.cc/89FT-ETQG (staff-uploaded, dark archive)] (last updated Dec. 5, 2021, at 06:30 CT) (noting that social media platforms now "deliberately use techniques from psychology and neuroscience to capture our attention, play with our emotions and keep us coming back as often as possible").

2. *See* Adam D.I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PROC. NAT'L ACAD. SCI. 8788, 8788 (2014) (explaining how "emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks"); Joseph Firth, John Torous, José Francisco López-Gil, Jake Linardon, Alyssa Milton, Jeffrey Lambert, Lee Smith, Ivan Jarić, Hannah Favian, Davy Vancampfort, Henry Onyeaka, Felipe B. Schuch & Josh A. Firth, *From "Online Brains" to "Online Lives": Understanding the Individualized Impacts of Internet Use Across Psychological, Cognitive and Social Dimensions*, 23 WORLD PSYCHIATRY 176, 176–80 (2024) (reviewing evidence that digital environments shape user behavior, including compulsive engagement driven by a "constant stream of entertainment" that disrupts sleep and work, addiction-like patterns such as tolerance and interference with offline responsibilities, and feedback loops in which users repeat behaviors that platform algorithms reward).

3. *See* Driscoll, *supra* note 1, at 21–23.

4. *See* Katherine J. Wu, *Radical Ideas Spread Through Social Media. Are the Algorithms To Blame?*, NOVA (Mar. 28, 2019), https://www.pbs.org/wgbh/nova/article/radical-ideas-social-media-algorithms/ [https://perma.cc/YA6D-9SNC (staff-uploaded archive)] (explaining that recommendation algorithms are built to maximize engagement and profit by showing users emotionally charged and attention-capturing content and that these systems "nudge[]" users toward radical political content or conspiracies when such content drives more engagement than neutral material).

Indeed, courts have had to grapple with these questions in the face of tragedy. In *M.P. v. Meta Platforms Inc.*,[5] the Fourth Circuit considered whether Meta[6] should be held liable for white supremacist content that Dylann Roof allegedly saw on Facebook before he committed the 2015 Charleston Massacre.[7] Facebook designed and deployed a recommendation algorithm that allegedly directed Roof towards that content, but the court held that Meta was immune from liability under Section 230 of the Communications Decency Act ("CDA").[8] The court found that Facebook's algorithmic sorting and promotion of content—however targeted or profit-driven—was equivalent to "traditional editorial functions" and therefore protected by § 230(c)(1) of the CDA.[9] In the court's view, the algorithm's role in Roof's radicalization could not be disentangled from Meta's status as a publisher of third-party speech.[10]

But this reasoning stretches the statute far beyond its original purpose. Section 230 was passed to encourage voluntary content moderation without exposing platforms to publisher liability.[11] It was not passed to shield the architectural design of the user experience itself.[12] Treating product design as editorial judgment collapses a critical doctrinal boundary.[13] The harm alleged in *M.P.* did not stem from what someone posted on Facebook; it arose from how Facebook's system delivered that content—automatically, repeatedly, and predictably calibrated to maximize engagement.[14] By casting algorithmic design

5. 127 F.4th 516 (4th Cir. 2025).

6. Meta Platforms, Inc., is the parent company that owns and operates Facebook. *M.P.*, 127 F.4th at 521. For clarity and consistency, this Recent Development uses "Meta" when referring to the corporate entity and "Facebook" when referring to the social media platform itself.

7. *Id.* at 520; Complaint at 2–3, M.P. v. Meta Platforms, Inc., 692 F. Supp. 3d 534 (D.S.C. 2023) (No. 2:22-cv-3830).

8. *M.P.*, 127 F.4th at 520–21; Communications Decency Act of 1996, Pub. L. No. 104-104, § 509, 110 Stat. 133, 137–39 (codified at 47 U.S.C. § 230).

9. *M.P.*, 127 F.4th at 525.

10. *Id.* at 525–26.

11. Publisher liability is the common-law rule that a party who disseminates another's defamatory or otherwise unlawful content may be held liable as if it authored the material, because publishing involves exercising traditional editorial functions such as selecting, editing, or disseminating content. *See* VALERIE C. BRANNON & ERIC N. HOLMES, CONG. RSCH. SERV., R46751, SECTION 230: AN OVERVIEW (2024) (explaining that under defamation law, "a newspaper 'who repeats or otherwise republishes a libel is subject to liability as if he had originally published it'" and that Section 230 was enacted against this backdrop).

12. *See* Danielle K. Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 406–08 (2017) ("The judiciary's long insistence that the CDA solely reflected 'Congress' desire to promote unfettered speech on the Internet' so ignores its text and history . . . .").

13. *See* Anderson v. TikTok, Inc., 116 F.4th 180, 188–89, 191 (3d Cir. 2024) ("Today, § 230 rides in to rescue corporations from virtually any claim loosely related to content posted by a third party, no matter the cause of action and whatever the provider's actions.").

14. *See M.P.*, 127 F.4th at 521–22.

as a publishing function,[15] the court barred the plaintiffs from discovery that may have revealed how Facebook's recommendation system operated.[16] Plaintiffs had no way of learning what data Facebook collected or whether it steered Roof toward extremist groups, leaving causation unprovable and remedy unattainable. The Fourth Circuit reached this conclusion by improperly treating Facebook's algorithmic outputs as a protected editorial decision, rather than distinguishing between content-management choices and system-level design.[17] This Recent Development explores the court's misapplication of Section 230 in *M.P.* and the broader dangers of continuing to extend immunity to the underlying architecture of online platforms.

Part I of this analysis outlines the history and purpose of Section 230 and details the Fourth Circuit's decision and reasoning in *M.P.* Part II argues that algorithmic content delivery is better understood as architectural design, not a standard editorial act. Part III critiques the *M.P.* majority's doctrinal approach and presents more functional readings of Section 230 that distinguish between speech-hosting and system engineering. Part IV proposes a path forward, one that preserves the protections Section 230 was designed to offer while closing a growing accountability gap for platform-generated harm. Part V addresses counterarguments and alternative routes to establish liability on the part of platforms.

I. BACKGROUND AND THE *M.P. V. META* DECISION

Congress enacted Section 230 of the Communications Decency Act in 1996 to encourage platforms to moderate harmful content without fear of publisher liability.[18] The law was a response to early court decisions suggesting that platforms could be punished for exercising any editorial control.[19] Congress feared the results of these cases would discourage efforts to screen or remove

15. *Id.* at 526 ("[A] newspaper company does not cease to be a publisher simply because it prioritizes engagement in sorting its content. And the fact that Facebook uses an algorithm to achieve the same result of engagement does not change the underlying nature of the act that it is performing. Decisions about whether and how to display certain information provided by third parties are traditional editorial functions of publishers, notwithstanding the various methods they use in performing that task.").

16. *See id.*

17. *Id.*

18. *See* Citron & Wittes, *supra* note 12, at 404.

19. *See Section 230: Legislative History*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230/legislative-history [https://perma.cc/XND9-GJV8] (explaining that in *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995 WL 323710 (N.Y. App. Div. 1995), *superseded by statute*, Communications Decency Act of 1996, Pub. L. No. 104-104, 110 Stat. 133, *as recognized in*, *Free Speech Coalition v. Paxton*, 145 S. Ct. 2291 (2025), the court held that "just for attempting to moderate some posts, Prodigy took on liability for *all* posts," prompting Congress to act).

illicit material.[20] Section 230(c)(1) thus provides that no platform shall be treated as "the publisher or speaker of any information provided by another information content provider."[21] In other words, Congress meant to preserve space for platforms to remove harmful material without transforming those efforts into a basis for treating them as legally responsible for users' posts.

Over time, courts interpreted that provision expansively.[22] One year after the passage of Section 230, the Fourth Circuit held in *Zeran v. AOL*[23] that the statute barred claims based on "traditional editorial functions," including decisions to publish, remove, or alter third-party content.[24] That precedent became the foundation for broad immunity across jurisdictions, even though the platforms of today bear little resemblance to the passive speech hosts contemplated in 1997.[25] This rationale has even extended to provide Section 230 immunity against claims arising from algorithmic amplification and targeted delivery of terrorist content on platforms like Facebook, Twitter, and YouTube.[26] The result is a body of precedent that improperly treats even complex, platform-driven dissemination choices as insulated from liability.[27]

20. *See* 47 U.S.C. § 230(b)(4) ("It is the policy of the United States . . . to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material . . ."); Jeff Kosseff, *A User's Guide to Section 230, and a Legislator's Guide to Amending It (or Not)*, 37 BERKELEY TECH. L.J. 757, 768–71 (2022) ("Congress wanted to pass a law to overturn *Stratton Oakmont* and ensure that platforms did *not* have an incentive to be neutral.").

21. 47 U.S.C. § 230(c)(1).

22. *See, e.g.*, Zeran v. Am. Online, Inc., 129 F.3d 327, 331–33 (4th Cir. 1997) (emphasizing that holding platforms liable for their decisions about publishing or removing content would impose an obvious "chilling effect" on internet speech and discourage moderation efforts).

23. 129 F.3d 327 (4th Cir. 1997).

24. *Id.* at 330.

25. Richard MacManus, *What the Internet Was Like in 1997*, CYBERCULTURAL (June 11, 2025), https://cybercultural.com/p/internet-1997 [https://perma.cc/X496-QR7Z] (describing early online services such as ICQ and AOL Instant Messenger, which were simple downloadable programs allowing users to send messages through buddy lists and basic chat functions); *see* Kosseff, *supra* note 20, at 788.

26. *See* Force v. Facebook, Inc., 934 F.3d 53, 67, 84–87 (2d Cir. 2019) ("[S]o long as a third party willingly provides the essential published content, the interactive service provider receives full immunity regardless of the specific edit[orial] or selection process.").

27. The *Zeran* rationale is not without judicial critics. *See* Anderson v. TikTok, Inc., 116 F.4th 180, 191 (3d Cir. 2024) (Matey, J., concurring in the judgment in part and dissenting in part) ("Today, § 230 rides in to rescue corporations from virtually any claim loosely related to content posted by a third party, no matter the cause of action and whatever the provider's action . . . this conception of § 230 immunity departs from the best ordinary meaning of the text and ignores the context of congressional action."); *see also* David Lukmire, *Can the Courts Tame the Communications Decency Act?: The Reverberations of* Zeran v. America *Online*, 66 N.Y.U. ANN. SURV. AM. L. 371, 389 (2010) (noting that although the court accurately recounted part of Congress's rationale, it "failed to consider" that promoting speech was "subsidiary to" Section 230's core purpose of enabling providers to block offensive material; and that while the legislative history "emphasizes repeatedly" the goal of protecting children, the Fourth Circuit nevertheless elevated "protecting the then-nascent Internet" as "the most important purpose" of the statute).

The Fourth Circuit's recent decision in *M.P.* reflects the continued progression of this doctrinal trajectory.[28] M.P., the daughter of Reverend Clementa Pinckney—one of the nine victims of the 2015 Charleston church massacre—alleged that Facebook's engagement-driven algorithm contributed to Dylann Roof's radicalization by promoting white supremacist content and extremist groups.[29] Her claims were grounded not in the specific speech Roof consumed, but in the platform's design—its structural emphasis on material that maximizes user interaction regardless of subject matter.[30]

Her complaint described how Facebook's design and architecture were optimized to maximize user engagement and profit with internal research showing that divisive content generates the highest engagement.[31] According to M.P., Facebook's algorithm repeatedly exposed Roof to extremist content, recommended that he join white supremacist groups, and ultimately "nurtured, encouraged, and . . . solidif[ied]" his racist views.[32] As evidence, she pointed to Roof joining extremist groups on Facebook and changing his profile picture to display white supremacist symbols shortly before the attack.[33] The complaint also traced Roof's radicalization back to a 2012 Google search for "black on white crime," which directed him to a white nationalist website.[34] From there, M.P. alleged, Facebook's engagement-driven system compounded and accelerated his exposure to extremist material.[35]

Nevertheless, the majority dismissed her claims, holding that Facebook's recommendation algorithm fell squarely within the traditional publisher functions protected by Section 230.[36] Under the binding precedent of *Zeran*, the court reasoned that the algorithmic recommendation, sorting, and ranking of posts and groups are covered by the same immunity as human curation.[37] Even

28. *See* M.P. v. Meta Platforms Inc., 127 F.4th 516, 527 (4th Cir. 2025).

29. *Id.* at 521–22.

30. *Id.* at 522; *see also* Complaint, *supra* note 7, at 30 (describing the way Facebook's algorithm promotes inflammatory content).

31. *See* Complaint, *supra* note 7, at 35; Appellant's Brief at 24–25, M.P. v. Meta Platforms, Inc., 127 F.4th 516 (4th Cir. 2025) (No. 23-1880).

32. *M.P.*, 127 F.4th at 521–22.

33. *Id.* at 522.

34. *Id.* at 521.

35. *See* Complaint, *supra* note 7, at 6–7 ("Meta's platforms are underpinned by algorithmic systems that process huge volumes of data to infer and develop detailed profiles of individual users and then use this mined data and artificial intelligence to shape and individually tailor each user's online experience in a way best designed to modify that user's behavior."); *see also* Sean Craig, *How Does Facebook Know What I Searched on Google*, ELITE DIGITAL MKTG. (June 6, 2023, at 10:21 ET), https://elitedigitalmarketing.ca/seo/how-does-facebook-know-what-i-searched-on-google/ [https://perma.cc/LQK2-9HXD] (explaining that platforms like Facebook collect "vast amounts of data from various sources, including user profiles, clickstream data, purchase history, and search data," and use that data "to understand user preferences" and serve customized content and advertising).

36. *M.P.*, 127 F.4th at 523–26.

37. *Id.* at 525 ("The case before us is more like *Zeran* than like *Erie Ins. Co*. M.P.'s state tort claims are inextricably intertwined with Facebook's role as a publisher of third-party content.").

if the system's design promoted harmful content or did so to drive revenue, the court reasoned that the platform's role in delivering that content was still a function of publication.[38] It analogized Facebook's algorithm to a newspaper editor's decision about what to place "above the fold" to attract readers.[39] In a footnote, the court acknowledged allegations that Facebook had auto-generated a page for the Council of Conservative Citizens—the same group Roof discovered through his Google search—but dismissed this as irrelevant because M.P. did not allege that Roof ever saw or joined that page.[40] Although the majority framed this omission as a straightforward pleading defect, it sits against the reality that M.P. had no access to discovery regarding what Facebook showed Roof, how its systems influenced him, or what pages he interacted with. These constraints complicate the line between a true missing detail in the complaint, and a gap created by the statute's bar on fact gathering.

The majority also concluded that even absent Section 230 immunity, M.P.'s claims failed because she had not plausibly alleged proximate cause under South Carolina law.[41] The opinion emphasized gaps in the complaint, such as the absence of detail about how long Roof was in extremist groups, how much time he spent on Facebook, or how those factors tied to the attack.[42] By framing causation this way, however, the court underscored the catch-22 created by Section 230: the statute prevented M.P. from accessing discovery into precisely these issues—what data Facebook collected on Roof, how its algorithms used that data, what groups it suggested, and what timeline linked his online activity to his offline violence. Without access to discovery, M.P. had no realistic way to gather the facts necessary to plead a detailed, plausible causal chain.

The dissent by Judge Rushing drew a sharper line around the limits of Section 230.[43] She agreed that Meta could not be held liable for simply publishing harmful third-party content but emphasized that its group recommendation feature—prompts like "Groups You Should Join"—was Facebook's own speech, not that of users.[44] In her view, recommending that a

38. *Id.* at 526. ("[N]ewspaper editors choose what articles merit inclusion on their front page and what opinion pieces to place opposite the editorial page. These decisions, like Facebook's decision to recommend certain third-party content to specific users, have as a goal increasing consumer engagement.").

39. *Id.*

40. *Id.* at 526 n.7.

41. *Id.* at 527. South Carolina requires a plaintiff to plead both factual and legal causation, with cause-in-fact shown by establishing the injury would not have occurred "but for" the defendant's conduct, and legal cause turning on whether the harm was foreseeable—which is determined by whether the plaintiff's injury was "a natural and probable consequence" of defendant's actions. *Id.* (quoting Bramlette v. Charter-Med.-Columbia, 393 S.E.2d 914, 916 (S.C. 1990)).

42. *Id.* at 527–28.

43. *See id.* at 529–33 (Rushing, J., concurring in part and dissenting in part).

44. *Id.* at 529, 531.

person "join a group, connect with another user, or attend an event" is conduct attributable to Facebook itself, even under existing Fourth Circuit precedent.[45] Because of this feature, she would have reversed the dismissal of M.P.'s negligence claims at least as to Facebook's own conduct and remanded for further proceedings.[46] Rushing stressed that Section 230 "does not insulate a company from liability for all conduct that happens to be transmitted through the internet."[47] Accordingly, collapsing Meta's independent design choices into "publishing" erased the statutory boundary between user speech and platform-created activity.[48] The majority responded only on narrow factual grounds, and once again in a footnote. They explained that M.P. had not alleged that Roof actually saw a "Groups You Should Join" prompt or joined a specific hate group through a Facebook referral and thus declined to reach the legal question.[49] That response sidestepped the dissent's core concern by resolving the case on the complaint's factual gaps rather than addressing whether direct platform recommendations should be treated as the company's own speech.

The result in *M.P.* confirms the Fourth Circuit's current consensus: under *Zeran*, so long as the harm even tenuously connects to the dissemination of third-party material—even when finely tuned, machine-driven, personally targeting, and commercially motivated—regardless of the resulting harm, Section 230 shields the platform from liability.[50] This expansive interpretation treats algorithmic amplification not as a product design choice but as an editorial act indistinguishable from human moderation.[51] That understanding leaves little room for plaintiffs to challenge system architecture, even when that technology is the source of the damage.[52]

II. ALGORITHMS AS PLATFORM ARCHITECTURE

The *M.P.* decision mischaracterizes how Facebook operates, treating its engagement-optimized algorithm as if the platform were exercising editorial discretion over third-party speech.[53] That framing ignores the reality: Facebook has built a behavioral engine designed to manipulate attention for profit.[54] This is not editorial judgment—it is product architecture.

45. *Id.*

46. *Id.*

47. *Id.* at 530 (quoting Henderson v. Source for Pub. Data, L.P., 53 F.4th 110, 129 (4th Cir. 2022)).

48. *Id.* at 531–32.

49. *Id.* at 525 n.6 (majority opinion).

50. *Id.* at 526–27.

51. Force v. Facebook, Inc., 934 F.3d 53, 66 (2d Cir. 2019) ("[A]rranging and distributing third-party information inherently forms 'connections' and 'matches' among speakers, content, and viewers . . . . That is an essential result of publishing.").

52. *See* Kosseff, *supra* note 20, at 777–78.

53. *See* 127 F.4th at 526.

54. Complaint, *supra* note 7, at 26.

Facebook's algorithm is "optimized to maximize user engagement," not to facilitate communication.[55] It promotes content based not on truth or value, but on its ability to capture attention;[56] the algorithm aims not to express ideas but to keep users online.[57] The platform prioritizes content that drives clicks, shares, and emotional response.[58] That is not curation—It is the core of the product's design. And the effects are measurable: research shows that Facebook can "transfer emotional states" to users simply by changing what they see.[59] Rather than evaluating meaning, the system maximizes reaction.[60] Content is displayed not by editorial choice on a case-by-case basis, but by a preset algorithmic prediction lacking the human discretion Section 230 was meant to shield.[61]

What makes this architecture most distinct from traditional publishing is its opacity.[62] On its "Help Center" page, Facebook describes its recommendations in only the broadest terms.[63] The page points to factors like a user's geographic location, demographic traits such as age or language, groups joined by friends, activity in Pages and searches, items viewed in Marketplace or Reels, and trending topics across the platform.[64] Beyond this high-level list, however, the system's workings, such as the weight and interaction of these

55. *M.P.*, 127 F.4th at 521.

56. Complaint, *supra* note 7, at 30 ("Facebook . . . tends to reaffirm harmful content while not balancing that content with differing views due to the algorithm's preference for continuously reengaging the user with more divisive content.").

57. *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 73 (2015) (describing how YouTube's recommendation system can steer users from innocuous content, like Federal Reserve speeches, into extremist and anti-Semitic conspiracy videos through algorithmically suggested content).

58. Force v. Facebook, Inc., 934 F.3d 53, 87 (2d Cir. 2019) (Katzmann, C.J., concurring in part and dissenting in part) ("Facebook's algorithm 'tends to promote the most provocative content' on the site." (quoting Max Fisher, *Inside Facebook's Secret Rulebook for Global Political Speech*, N.Y. TIMES (Dec. 27, 2018), http://www.nytimes.com/2018/12/27/world/facebook-moderators.html [https://perma.cc/2Z4P-D3QC (staff-uploaded, dark archive)])).

59. Kramer et al., *supra* note 2, at 8788 (reporting experimental evidence that altering the emotional content of users' News Feeds changes their own expression through "emotional contagion," showing that such shifts occur "without their awareness").

60. *See* Complaint, *supra* note 7, ¶ 89.

61. *See id.*

62. *See* PASQUALE, *supra* note 57, at 15.

63. *See How Facebook Suggests Groups To Join*, FACEBOOK: HELP CTR., https://www.facebook.com/help/382485908586472 [https://perma.cc/T459-CJJP (staff-uploaded archive)].

64. *Id.* Pages refers to Facebook's label for a public profile created by individuals, businesses, organizations, or public figures to share information with audiences on the platform. *See Pages*, FACEBOOK: HELP CTR., https://www.facebook.com/help/282489752085908 [https://perma.cc/VCR5-YYHM (staff-uploaded archive)]. Marketplace is an integrated digital feature where a user can "buy and sell items with people in [their] community on Facebook." *See Marketplace*, FACEBOOK: HELP CTR., https://www.facebook.com/help/1713241952104830 [https://perma.cc/2QUX-CKDZ (staff-uploaded archive)]. Reels is Facebook's video content functionality, which includes features for posting, streaming, and viewing video content. *See Reels*, FACEBOOK: HELP CTR., https://www.facebook.com/help/753046815962474 [https://perma.cc/SPW4-QMNC (staff-uploaded archive)].

variables, are proprietary and closed to outside scrutiny.[65] A Pew study illustrates the scale of this concealed structure: Nearly nine in ten users had categories—a list of a user's purported interests crafted by the algorithm and displayed on their "Your Ad Preferences" page—automatically generated for them.[66] Heavy and long-term users were typically assigned more than twenty-one categories.[67] These classifications draw not only from on-platform activity but also from tracking user behavior across millions of websites through tools like the Meta Pixel, which even links offline purchases to online profiles.[68] What remains unknown is how all these diverse inputs are weighted, combined, and deployed to generate outputs. Since algorithms can operate as "black boxes," with layers of legal and technical obscurity shielding them from scrutiny,[69] what remains unknown is how all these diverse inputs are weighted, combined, and deployed to generate outputs.

Facebook's architecture extends beyond content ranking to design features that actively incentivize specific user behavior. Tools like group badges—labels such as "Conversation Starter," "Visual Storyteller," or "Rising Star"—reward activity and reinforce hierarchies inside communities.[70] These badges act as behavioral nudges, signaling who has influence and incentivizing regular activity in the group in order to maintain status and gain greater influence.[71]

Facebook also routinely auto-generates group pages whenever its systems detect repeated user interest in a topic, creating placeholder communities without any human judgment about whether a subject should have a dedicated

65. Haochen Sun, *The Right To Know Social Media Algorithms*, 18 HARV. L. & POL'Y REV. 3, 10–12 (2023) (noting that social media algorithms are protected as trade secrets, are "not generally known," and remain inaccessible to researchers because companies treat the details of how variables are weighted, combined, and operationalized as proprietary).

66. Paul Hitlin, Lee Rainie & Kenneth Olmstead, *Facebook Algorithms and Personal Data*, PEW RSCH. CTR. (Jan. 16, 2019), https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/ [https://perma.cc/S4ZY-EDML].

67. *Id.*

68. *Id.* The Meta Pixel is a piece of JavaScript code that a website owner embeds in their site's code to track activity (such as page views, conversions, and other user actions) and send that information back to Meta Platforms, Inc., for purposes of measuring, optimizing, and retargeting advertising across Meta's services. It enables advertisers to understand how people interact with their website after clicking on or viewing ads on Meta properties and to tailor ad delivery accordingly. *See About Meta Pixel*, META: BUS. HELP CTR., https://www.facebook.com/business/help/742478679120153 [https://perma.cc/D6EP-ZWQC].

69. PASQUALE, *supra* note 57, at 6.

70. *Discover New Badges To Recognize Admins and Outstanding Members*, FACEBOOK CMTY. (Nov. 1, 2018), https://www.facebook.com/community/whats-new/facebook-group-badges/ [https://perma.cc/P992-TDDE] (describing the different types of badges that Facebook assigns to users based on participation and engagement).

71. *Id.* (explaining that the "Conversation Starter" badge is awarded to members whose posts receive the most engagement in the past month, and the "Rising Star" badge is given to new members that receive the most comments and reactions on their posts and comments).

space.[72] This automated process applies across the board—including to militia movements and extremist ideologies; the same mechanism that creates pages for benign hobbies also produces official-looking spaces for groups associated with violence or radicalization, without any human discretion.[73] By presenting users with a preexisting community structure, the feature lowers the barrier to entry: it suggests that these topics have an established home on the platform and makes it far easier for like-minded individuals to gather and begin organizing than if they had to build such a group from scratch. And once these spaces exist, the badge system layers an additional set of incentives on top, rewarding high-engagement users and encouraging sustained activity within these radicalized environments. These mechanics are not instances of traditional editorial judgment; they are structural tools that craft and legitimize communities. As with the other operations of the algorithm, information explaining how this process takes place and what data influences it is inaccessible to the public.

Whatever happens behind the curtain, the algorithm performs as intended. Content that provokes strong engagement signals—whether sensational, polarizing, extreme, or otherwise attention-grabbing—is surfaced and amplified in a personalized format to keep a specific user online.[74] This is not accidental; it is the intended, foreseeable output of the design.[75] Yet courts continue to treat these platforms as if they were only mere forums for speech.[76] Legally, this distinction matters because the claims brought by M.P. against Facebook, and the claims that plaintiffs in similar suits bring against other platforms, do not challenge the ideas contained in any user's post, which would implicate traditional First Amendment concerns. Instead, they target the platform's own conduct: the proprietary architecture that packages and delivers speech—an architecture that operates according to engagement metrics, not editorial discretion. Because Section 230 blocks discovery at the threshold, plaintiffs cannot even probe which features were involved, how they were

72. *See* Caitlin Dickson, *Facebook Continues To Autogenerate Pages for Proud Boys, Other Extremist Groups*, YAHOO NEWS (Oct. 4, 2022), https://www.yahoo.com/news/facebook-continues-to-auto-generate-pages-for-the-proud-boys-other-extremist-groups-194527393.html [https://perma.cc/H9MD-RC6Q] (explaining that a long-standing Facebook feature "automatically generates a new page any time users list a job title, business, interest or location on their profile that doesn't already have an official page of its own").

73. *See* Tess Owens, *Facebook Is Auto-Generating Militia Group Pages as Extremists Continue To Organize in Plain Sight*, WIRED (Oct. 29, 2024, at 07:00 ET), https://www.wired.com/story/facebook-militia-organizing-election/ [https://perma.cc/QEX2-9QCV] (reporting that Facebook auto-generated pages for groups like the Arizona chapter of American Patriots Three Percent, despite the organization being banned as a "militarized social movement").

74. *See* Complaint, *supra* note 7, ¶ 90.

75. *Id.*

76. *See, e.g.*, M.P. v. Meta Platforms Inc., 127 F.4th 516, 526 (4th Cir. 2025); Force v. Facebook, Inc., 934 F.3d 53, 66 (2d Cir. 2019).

deployed, or whether the platform knew of their risks. When a system predictably amplifies material based on engagement metrics alone, the risk flows from the architecture of the product, not from the ideas it transmits. That mechanism is not editorial discretion. It is engineered exposure meant to keep the user online through any means necessary and regardless of side effects. It should be treated as defective product design.[77]

III. THE LIMITS OF THE PUBLISHER ANALOGY

The publisher analogy underlying Section 230's application is no longer sufficient. The technological landscape has shifted from rudimentary content-hosting to advanced, behavior-shaping algorithms, a reality far removed from the world in which the statute was enacted. There is certainly merit in protecting platforms from some level of liability, particularly for the actions of their users. It is essential to avoid deterring content moderation, which could be viewed as assuming responsibility for user speech, or burdening innovation with unpredictable legal exposure.[78] But the dangers posed by modern engagement-optimized systems—especially those designed to amplify outrage or extremism—are too great to justify clinging to a framework that fails to distinguish between speech and system.[79] Although courts have long treated content curation as a basic function of publication editing, that model fails to account for the structural and behavioral engineering that now shapes digital environments.[80]

A small number of recent decisions have begun to question the outer limits of this framework.[81] In *Fair Housing Council v. Roommates.com*,[82] the Ninth Circuit held that Section 230 does not apply when a platform "contributes materially to the alleged illegality of the conduct."[83] The platform's own design, which required users to disclose protected characteristics, made it more than a passive host.[84] The court explained that "a website helps to develop unlawful content" when it solicits or structures responses that lead to illegal conduct.[85]

77. For a more detailed discussion of products liability in the context of Section 230, see *infra* notes 119–21 and accompanying text.

78. *See* Citron & Wittes, *supra* note 12, at 412.

79. *See id.* at 413.

80. *M.P.*, 127 F.4th at 526; *see* Citron & Wittes, *supra* note 12, at 414.

81. *See* Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1167 (9th Cir. 2008); Henderson v. Source for Pub. Data, L.P., 53 F.4th 110, 129 (4th Cir. 2022); Anderson v. TikTok, Inc., 116 F.4th 180, 192 (3d Cir. 2024).

82. 521 F.3d 1157.

83. *Id.* at 1164–68.

84. *Id.* at 1166.

85. *Id.* at 1166–68. The same logic underlies Judge Rushing's dissent in *M.P.*: Facebook's "Groups You Should Join" could be viewed as Facebook prompting participation in illegal activity just like the questionnaires and sorting tools in *Roommates.com*. *See* M.P. v. Meta Platforms Inc., 127 F.4th 516, 529, 531 (4th Cir. 2025) (Rushing, J., concurring in part and dissenting in part). Both features were

Similarly, in *Henderson v. Source for Public Data*,[86] the Fourth Circuit denied immunity to a data broker accused of assembling and selling inaccurate public records.[87] The court emphasized that a but-for causal relationship between the act of publication and liability is insufficient for immunity.[88] These opinions, despite exemplifying judicial reasoning which more precisely identifies the bounds of publisher liability, remain exceptions to the prevailing trend.[89]

The most explicit doctrinal break came in the August 2024 case of *Anderson v. TikTok, Inc.*[90] There, the Third Circuit rejected TikTok's attempt to claim immunity for its algorithm's promotion of the "Blackout Challenge," a deadly online trend that encouraged users—often children—to asphyxiate themselves on camera.[91] The plaintiffs alleged that TikTok's recommendation system pushed the challenge to ten-year-old Nylah Anderson, who later died after attempting it.[92] The court found that this conduct went beyond merely publishing third-party content: TikTok's algorithm "[d]ecid[ed] on the third-party speech that will be included in or excluded from a compilation—and then organiz[ed] and present[ed] the included items."[93] Because the recommendation occurred without "any specific user input," the algorithm was not simply indexing existing material—it was promoting it.[94] The Third Circuit rejected TikTok's claim to immunity, holding that Section 230 applies only to information provided solely by users themselves, not when the platform itself engages in its own "expressive activity," in this case, the targeted, automated promotion of life-threatening behavior.[95]

The Third Circuit grounded its reasoning in the Supreme Court's decision in *Moody v. NetChoice, LLC*,[96] where the majority noted that a platform's algorithmic curation reflecting "editorial judgments" about "compiling the third-party speech it wants in the way it wants" constitutes an expressive

built by the platforms themselves as part of each respective product and then displayed for users to engage with.

86. 53 F.4th 110.

87. *Id.* at 117–18, 120.

88. *Id.* at 123.

89. Kosseff, *supra* note 20, at 782. ("[I]n more than a decade since the opinion, other courts have used its reasoning relatively sparingly to deny § 230 protections.").

90. Anderson v. TikTok, Inc., 116 F.4th 180, 192 (3d Cir. 2024).

91. *Id.* at 182.

92. *Id.*

93. *Id.* at 184 (citing Moody v. NetChoice, LLC, 144 S. Ct. 2383, 2402 (2024)).

94. *Id.* at 184–85 n.12.

95. *Id.* at 184. According to a compilation by Bloomberg of news reports, court records, and interviews, at least fifteen children at or below the age of twelve were killed by the Blackout Challenge in an eighteen-month period between 2021 and 2022. Olivia Carville, *TikTok's Viral Challenges Keep Luring Young Kids to Their Deaths*, BLOOMBERG BUSINESSWEEK (Nov. 30, 2022, at 00:01 ET), https://www.bloomberg.com/news/features/2022-11-30/is-tiktok-responsible-if-kids-die-doing-dangerous-viral-challenges [https://perma.cc/6FZH-BG4T (staff-uploaded, dark archive)].

96. 144 S. Ct. 2383 (2024).

product protected by the First Amendment.[97] Building on that logic, the *Anderson* court agreed with the plaintiff's contention that when a platform's algorithm amalgamates user videos into a curated stream that communicates to users that the compilation "will be interesting to them," the platform engages in first-party speech.[98] If that curation can trigger First Amendment protection, the court reasoned, it necessarily follows that such conduct cannot be immunized as third-party speech under Section 230.[99] Immunity applies only to information "provided by another information content provider," not to expressive choices the platform itself makes about what to select, arrange, and deliver.[100]

The *Anderson* decision thus marks a rare moment of doctrinal realism in Section 230 jurisprudence. By acknowledging that algorithmic systems do not merely host speech but actively construct and shape it, the Third Circuit applied Section 230 in a way that reflects how modern platforms functionally operate. Its reasoning recognizes that recommendation algorithms are not passive intermediaries but active design mechanisms capable of producing their own expressive and harmful effects. In grounding its analysis in both statutory text and technological reality, *Anderson* offered a path forward—one that preserves Section 230's core protections for user speech while allowing courts to scrutinize the architecture of systems that amplify it.

The *M.P.* decision makes clear that, at present, the dominant approach remains tethered to an outdated model that fails to reflect how platforms actually operate.[101] Rather than engaging with the emerging recognition—most prominently reflected in the Third Circuit's *Anderson* discussion—that algorithms can reflect a platform's own expressive or design-driven activity, the Fourth Circuit adhered to its existing precedent, reasoning that "acts of arranging and sorting content are integral to the function of publishing" and that Facebook's use of generated recommendation systems merely "automates its editorial decision-making."[102] Yet the court never explained how this reasoning avoids the "but-for" causal relationship *Henderson* rejected as insufficient for immunity. It simply stopped at the threshold finding of "publishing activity," treating that as dispositive and declining to engage with any of the plaintiff's factual or design-based arguments. In doing so, the Fourth

97. *Id.* at 2394.

98. 116 F.4th at 183.

99. *Id.* at 184.

100. *Id.* at 183 ("In other words, [interactive computer services] are immunized only if they are sued for someone else's expressive activity or content (i.e., third-party speech), but they are not immunized if they are sued for their own expressive activity or content (i.e., first-party speech).").

101. *See* M.P. v. Meta Platforms Inc., 127 F.4th 516, 527 (4th Cir. 2025) ("We are not free to disregard Section 230 or to limit its application based on our own assessment of the merits of its expansive reach.").

102. *Id.* at 526.

Circuit effectively demonstrated how Section 230 *is* a "but-for" shield—one that triggers dismissal whenever a platform's conduct can be tenuously connected to the act of publication. The result is a doctrinal kill switch that forecloses discovery and prevents plaintiffs from demonstrating that the harm truly arose from third-party speech or from the architecture that delivered it.

Compounding the concern, the Fourth Circuit failed to engage with *Anderson*'s reasoning. It did not ask whether Facebook's recommendation system—one that operates "without any specific user input" and autonomously promotes divisive material—could constitute the platform's own expressive act. By simply reviving the newspaper metaphor, the Fourth Circuit collapsed the modern complexity of algorithmic recommendation into a pre-digital image of editorial discretion. In doing so, it insulated Meta's system design from scrutiny and foreclosed discovery into how the algorithm actually functions—precisely the type of inquiry *Anderson* allowed to proceed.[103] The *M.P.* majority's refusal to see this parallel reflects a broader reluctance to confront how engagement-driven systems blur the line between content and conduct.

The result was a doctrinal step backward. *Anderson* suggested a path toward reconciling Section 230 with technological reality, recognizing that algorithms are capable of producing first-party expressive conduct and independent harm.[104] *M.P.*, by contrast, re-entrenched a model of near-total immunity that ignores how those systems work. It reaffirmed a view of Section 230 that shields not just speech but structure, transforming a statute meant to protect moderation into one that conceals the mechanisms of amplification.

IV. A FUNCTIONAL READING OF SECTION 230

The time has come for courts to turn the few decisions that properly apply Section 230 into a coherent judicial standard. A consistent, functional reading of Section 230 should begin where *Anderson* left off. That decision did not rewrite the statute; it read it as written and applied it properly to technology to which the publisher analogy is inapplicable. The *Anderson* majority recognized that algorithms capable of autonomously selecting, organizing, and promoting content are not mere conduits for user speech but active participants in its dissemination. In doing so, the Third Circuit applied Section 230 in a way that reflects how modern platforms actually operate.

103. *See* Appellant's Reply Brief at 14, Anderson v. TikTok, Inc., 116 F.4th 180 (3d. Cir. 2024) (No. 22-3061) (explaining that, unlike traditional publishers whose content is "the exact same for every consumer," TikTok's targeting "is equivalent to the New York Times spying on 10-year-old Nylah Anderson and determining that she likes farm animals and then clipping out an article on farm animals . . . and mailing that article, and only that article, directly to her doorstep").

104. *See* 116 F.4th at 184 (describing how TikTok's recommendation algorithm was TikTok's own 'expressive activity,'" which can form the basis of a plaintiff's tort claim).

Courts can follow this reasoning without disrupting the statute's basic structure. *Anderson* provides a workable template: when a platform's recommendation system operates without specific user input, determines what content a user will see, and thereby shapes the user's experience through its own expressive choices, the conduct at issue is the platform's—not the user's. That line—between hosting and engineering—is precisely what Section 230's text contemplates. As *Roommates.com* and *Henderson* suggested, the statute protects passive publication decisions but not acts of design, solicitation, or manipulation that materially contribute to the resulting harm. *Anderson* turned those hints into a principled rule: if a platform's algorithm engages in expressive activity by deciding what to include or exclude in a curated compilation, then the resulting injury flows from the platform's design, not from user speech.

Adopting this framework would do more than clarify doctrine—it would restore fairness to the litigation process itself. Under the current expansive interpretation of Section 230, plaintiffs like M.P. are denied access to even basic discovery about how a platform's algorithms operate, what data they collect, and how those systems determine what users see. A functional reading would permit claims to advance past the threshold stage when they plausibly allege harm caused by a platform's own architecture, allowing courts to evaluate liability based on an evidentiary record rather than on an assumption. This modest shift would not expose platforms to ruinous damage remedies; companies like Meta are among the most powerful and profitable in history and can easily sustain damages when harm stems from their own design choices.[105] For entities fully capable of bearing its costs, liability for harms produced by their creations is a necessary mechanism of accountability.

Moreover, a shift from the widely adopted, improper reading of Section 230 would allow plaintiffs to address factual and causation issues to which the *M.P.* majority attributed dismissal. The Fourth Circuit noted that M.P. had not alleged whether Dylann Roof ever saw a "Groups You Should Join" prompt, whether Facebook specifically recommended hate groups to him, or whether he joined such groups because of a referral. But those gaps were the product of misreading Section 230: by barring the suit at the outset, the court prevented M.P. from obtaining discovery that could have answered those very questions. A refined doctrine would allow plaintiffs to investigate how long users were exposed to harmful material, how recommendation systems

105. Meta Platforms, Inc., reported total revenue of approximately $164.5 billion for the fiscal year ending December 31, 2024. *See Meta Platforms Revenue 2012–2025|META*, MACROTRENDS, https://www.macrotrends.net/stocks/charts/META/meta-platforms/revenue [https://perma.cc/9VHA-B3BY]. ByteDance Ltd. (who was the majority owner of TikTok in 2024), reported total revenue of approximately $156 billion for the same period. *See Key Financial Figures of ByteDance in 2024*, STATISTA, https://www.statista.com/statistics/1342778/bytedance-key-annual-financial-figures [https://perma.cc/3MDQ-EW8G].

interacted with off-platform behavior, and what data informed those connections. Such information lies exclusively within the company's control. Opening discovery in cases like M.P.'s would not predetermine liability—it would simply ensure that courts decide causation on facts rather than speculation.

Courts—not Congress—are best positioned to make this course correction.[106] The judiciary created the overly broad interpretation of Section 230,[107] and it has both the tools and the duty to refine it.[108] Legislatures face structural limits in responding to technology that evolves faster than statutes can be drafted or amended.[109] Even well-intentioned reforms often take years to negotiate, and each new amendment must navigate complex political dynamics and stakeholder pressures before passage.[110] The judicial process, by contrast, can adapt incrementally and responsively through case-by-case adjudication. Courts routinely refine ambiguous statutory language, delineate boundaries, and reconcile emerging technologies with enduring legal principles.[111] The boundary between speech and system is already visible in

106. Some commentators similarly in favor of curtailing the broad power of Section 230 believe the path to reform lies outside of the judicial system. *See* Ramya Krishnan & Alex Abdo, *How Do You Solve a Problem Like Facebook?*, KNIGHT FIRST AMEND. INST. (Oct. 14, 2021), https://knightcolumbia.org/content/how-do-you-solve-a-problem-like-facebook [https://perma.cc/N3DS-S4JS] (arguing that Congress must enact reforms—such as mandated ad transparency, safe harbors for independent research, and regulated access to platform data—because meaningful oversight of platform behavior requires statutory tools courts alone cannot supply).

107. *See* Gregory M. Dickinson, *Section 230: A Juridical History*, 28 STAN. TECH. L. REV. 1, 1, 7 (2025) (explaining that courts "constructed an immunity doctrine insufficiently resilient against technological change," identifying the "interpretive steps and missteps" by which courts broadened Section 230, and noting that at "each stage of evolution, courts made . . . questionable[] choices to broaden the scope of Section 230 immunity").

108. *See* Malwarebytes, Inc. v. Enigma Software Grp. USA, LLC, 141 S. Ct. 13, 13–14 (2020) (statement of Thomas, J., respecting denial of certiorari) (noting that although the Court denied review in that case, "in an appropriate case, we should consider whether the text of this increasingly important statute aligns with the current state of immunity enjoyed by Internet platforms," and emphasizing that the Court has "never interpreted" Section 230 despite lower courts having construed it broadly).

109. *See* Maya Kornberg, Marci Harris & Aubrey Wilson, *Congress Must Keep Pace with AI*, BRENNAN CTR. FOR JUST. (Feb. 8, 2024), https://www.brennancenter.org/our-work/research-reports/congress-must-keep-pace-ai [https://perma.cc/N7GY-4XJ5] (explaining that Congress faces an "ever-expanding gap between technological advancement (which is often exponential) and the ability of governing institutions to keep up with these changes (at their default linear pace)," and noting that the "accelerating rate of technology's evolution," combined with "political inertia," exacerbates this structural mismatch).

110. *See* Alan Z. Rozenshtein, *Interpreting the Ambiguities of Section 230*, BROOKINGS INST. (Oct. 26, 2023), https://www.brookings.edu/articles/interpreting-the-ambiguities-of-section-230/ [https://perma.cc/7WXS-SN7W] (observing that Congress is often hampered by "well-known pathologies and partisan gridlock" and noting that Section 230 reform must contend with the influence of "powerful interest groups," particularly large platforms, which makes legislative change slow and politically difficult).

111. *See, e.g.*, Bostock v. Clayton County, 140 S. Ct. 1731 (2020) (interpreting Title VII's prohibition on discrimination "because of sex" to include sexual orientation and gender identity); Riley

precedent—the task is now to define it with consistency. As *Anderson* illustrates, courts can enforce Section 230's text while recognizing the expressive and architectural roles modern platforms play. The pace of technological change only underscores why courts must lead this interpretive evolution: statutes may remain static, but judicial reasoning can evolve alongside the systems it governs—ensuring that immunity does not become impunity.

The misstep in *M.P. v. Meta* was not in applying Section 230,[112] but in applying it without precise distinction. The court treated Facebook's recommendation system—one intentionally engineered to target individual users and maximize engagement—as indistinguishable from an editorial decision about what to publish.[113] This reading ignores both how the platform functions and what the statute says. Courts must do better going forward. They can—and should—adopt a functional reading of Section 230 that draws a principled line between content and system. Where the claim is about what a user said, the platform remains immune from liability. Where the claim is about how an algorithm was built to exploit that speech for emotional manipulation or monetary gain, however, the platform should be held liable. Drawing that line is not policymaking; it is a long-overdue evolution of an archaic standard in the face of algorithmic architecture that is increasingly capable and dangerous.

V. COUNTERARGUMENTS AND ALTERNATIVES

The argument for narrowing Section 230 is incomplete without addressing the countervailing concerns—primarily that doing so could chill online expression, overwhelm courts with lawsuits, or threaten the "vibrant and competitive free market" Congress sought to protect.[114] These concerns are legitimate, but they lose force when applied to cases like *M.P. v. Meta*. A functional reading would not center around specific types of speech from the user's side. It would not seek to punish Facebook and other platforms for third-party content but for the design of a system that expectedly creates and magnifies harm through that platform's own expressive activity. Holding platforms accountable for engineering choices that cause foreseeable injury

v. California, 573 U.S. 373 (2014) (holding that police generally must obtain a warrant before searching digital information on a cell phone seized during an arrest); Carpenter v. United States, 585 U.S. 296 (2018) (requiring a warrant for government acquisition of historical cell-site location information); Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015) (holding that large-scale book digitization and searchable indexing constituted fair use); hiQ Labs, Inc. v. LinkedIn Corp., 31 F.4th 1180 (9th Cir. 2022) (interpreting the Computer Fraud and Abuse Act in the context of scraping publicly available website data).

112. M.P. v. Meta Platforms Inc., 127 F.4th 516, 526 (4th Cir. 2025).

113. *Id.*

114. Communications Decency Act of 1996, Pub. L. No. 104-104, § 509, 110 Stat. 133, 137–38 (codified at 47 U.S.C. § 230).

would not silence anyone's speech—it would merely require billion-dollar companies to internalize the costs of their own conduct. Section 230 was never meant to confer absolute immunity for defective design and narrowing the statute's protection to restore that balance would not stifle online discourse. If anything, it would protect that discourse by ensuring that the same platforms profiting so immensely from directing their users toward creating and enduring harm are also tasked with cleaning up their own messes.

Defenders of broad immunity also argue that even if Section 230 were read more narrowly, the First Amendment would still bar claims like M.P.'s.[115] But this misreads both the statute and constitutional doctrine.[116] The First Amendment protects expressive activity; it does not shield a company's decision to build a product that causes harm through mechanical amplification and systematic targeting of individual users.[117] *Anderson* makes this distinction explicit: Algorithmic curation can constitute a platform's own expressive act when it organizes and presents content without user input—but that recognition of expressiveness does not mean platforms are exempt from accountability.[118] To the contrary, it clarifies where their speech ends and their conduct begins. Courts can respect the expressive dimension of platform design while still holding companies liable for the non-expressive consequences of that design—just as a newspaper's right to publish does not immunize it from liability for a printing press that injures its workers or a delivery truck that crashes into a pedestrian. The First Amendment limits the government's ability to regulate ideas, not a plaintiff's ability to seek redress for negligent architecture. Narrowing Section 230 would thus facilitate, not undermine, proper First Amendment analysis by ensuring that courts have access to the facts necessary to determine whether the challenged conduct is expressive or purely mechanical.

115. *See generally* Note, *Section 230 as First Amendment Rule*, 131 HARV. L. REV. 2027, 2027–28 (2019) (arguing that the First Amendment requires Section 230 to bar lawsuits based on defamation torts).

116. Eric Goldman, *Why Section 230 Is Better than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33, 34 (2019) (arguing that Section 230 provides "irreplaceable substantive and procedural benefits beyond the First Amendment's free speech protections" and that "the First Amendment does not backfill these benefits," because constitutional defenses are narrower, fact-intensive, and often unavailable for claims not targeting expressive content).

117. The *Brandenburg* test reflects a narrow exception under which plaintiffs might still ground claims in the content of speech itself. In *Brandenburg v. Ohio*, 395 U.S. 444 (1969) (per curiam), the Supreme Court held that "the constitutional guarantees of free speech and free press do not permit a State to forbid or proscribe advocacy of the use of force or of law violation except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action." *Id.* at 447. A plaintiff whose injury stems directly from user content meeting that standard—such as targeted incitement or coordinated calls for immediate violence—might overcome First Amendment protections.

118. Anderson v. TikTok, Inc., 116 F.4th 180, 184 (3d Cir. 2024).

Finally, even if courts remain reluctant to reframe Section 230 immunity, they could recognize that the harms at issue may support a successful products liability claim. When a platform's algorithm functions as a defective product—one that operates exactly as designed but unreasonably endangers its users, liability should be found, as it is across so many other industries. Traditional tort law already provides a framework for evaluating such risks.

The *Lemmon v. Snap, Inc.*[119] decision exemplifies this approach: the Ninth Circuit held that plaintiffs could pursue a negligent-design claim based on Snapchat's "Speed Filter," which encouraged and rewarded dangerous driving, because the alleged harm arose from the app's design, not from user speech.[120] Extending that reasoning to platforms like Facebook would not erode Section 230's purpose; it would simply place algorithmic design within the same duty framework that governs all consumer products.[121] Platforms would remain free to host speech, but they could no longer profit from designs that predictably endanger users while hiding behind statutory immunity. Although M.P. did not frame her claims as products liability, the underlying theory—harm caused by an engagement-driven design feature—is functionally analogous. If Facebook's algorithms had been treated as a defective product rather than as a publishing decision, the court could have reached the merits rather than ending the inquiry at Section 230.[122]

For now, the central task is not to resolve every doctrinal or constitutional question, but to ensure that plaintiffs can reach the point where those questions can even be asked. The overextension of Section 230 has foreclosed meaningful inquiry before facts are ever developed, transforming what should be evidentiary debates about causation, fault, and design into threshold dismissals. Correcting that misreading is therefore the necessary first step. Once plaintiffs are allowed to proceed past the immunity barrier, courts will have the tools to separate protected expression from negligent engineering, to evaluate First

119. 995 F.3d 1085 (9th Cir. 2021).

120. *Id.* at 1088–89, 1091–92 (explaining that Snapchat "rewards [users] with 'trophies, streaks, and social recognitions' based on the snaps they send" and that many users "suspect, if not actually 'believe,'" they will earn such rewards for recording high-speed snaps; concluding that this gamified reward system, combined with the Speed Filter, created a foreseeable design defect that encouraged teenagers to accelerate to dangerous speeds, and that Snap had notice of prior accidents tied to this feature).

121. *See generally* Marc H. Pfeiffer, *First, Do No Harm: Algorithms, AI, and Digital Product Liability*, CTR. FOR URB. POL'Y RSCH. (Sep. 2023) (proposing a modern liability framework for algorithmic systems that would impose proactive risk-management duties, require developers to assess and mitigate foreseeable harms before deployment, strengthen ongoing monitoring obligations, and align digital products with the safety expectations that govern traditional consumer goods).

122. Product liability theory could also resolve the asymmetry between individual plaintiffs and massive technology companies. Just as manufacturers of vehicles or pharmaceuticals are expected to bear the costs of foreseeable harm, so too should digital platforms whose design choices produce measurable real-world consequences.

Amendment defenses on a full record, and to determine whether product liability principles apply. Getting Section 230 right does not dictate outcomes—it simply restores the ordinary process of adjudication, where discovery, evidence, and reasoned analysis decide responsibility rather than immunity doctrine alone.

CONCLUSION

Section 230 was never meant to immunize harm designed into a platform's architecture. As online spaces have evolved from static forums into sophisticated systems of behavioral manipulation, the law's application must evolve with them. Courts must stop treating all algorithmic conduct as mere editorial discretion and start treating it for what it is: product design, deliberately engineered to keep users online and profit by any means. Immunity for the speech of others should remain. Immunity for the systems that deliver that speech—especially when those systems generate divisiveness and outrage—should not. The law already contains the tools for this distinction. What remains is the judicial will to use them. *M.P. v. Meta* illustrates the dangers of ignoring that reality and shows how firmly courts remain entrenched in an outdated model even in the face of mass violence. Until they adopt a more functional reading, they are not just overlooking the harm—they are safeguarding the production and dissemination of it, shielding algorithmic systems under the armor of Section 230.

AMELIA CHRISTIAN WALKER**

** J.D. Candidate, Class of 2027. I would like to think Katrina Smith, Robby Fensom, and everyone at the *North Carolina Law Review* for their thoughtful contributions to this piece. It is an immense honor to be published by such a kind and hardworking group of people. I am incredibly grateful to all of my family and close friends who have never wavered in their support of me and my academic pursuits. Most of all, I want to thank my mom, Christian Walker, who has been the ultimate champion of my education and happiness.