The Corporate Case for Platform Transparency: A Regulatory Proposal After *Moody v. NetChoice*

Hunter Thompson*

I. INTRODUCTION

Amazon fake review schemes.[1] Russian bots circulating "fake news."[2] The Christchurch Massacre livestream.[3] Instagram's toxic influence on teen girls' bodies.[4] The

* Hunter Thompson, J.D. Candidate, University of Iowa College of Law, 2026. Thanks to the *Journal of Corporation Law*, Volume 51 team, for their invaluable assistance and meticulous attention to detail. This work would not have been possible without my family's unwavering love and support.

1. Lisa Fickenscher & Thomas Barrabi, *Amazon Shoppers Are Being Duped By Manipulated Customer Reviews as China-Based Company Vevor Rakes in $500M in Sales: Bombshell Whistleblower Claims*, N.Y. POST (Nov. 18, 2024), https://nypost.com/2024/11/18/business/china-based-tool-giant-vevor-manipulates-customer-reviews-while-amazon-turns-a-blind-eye-sources/ [https://perma.cc/AL76-3Q3Z].

2. Adrian Chen, *The Agency*, N.Y. TIMES (June 2, 2015), https://www.nytimes.com/2015/06/07/magazine/the-agency.html?_r=0 (on file with the *Journal of Corporation Law*).

3. Ryan Mac, Kellen Browning & Sheera Frenkel, *The Enduring Afterlife of a Mass Shooting's Livestream Online*, N.Y. TIMES (May 19, 2022), https://www.nytimes.com/2022/05/19/technology/mass-shootings-livestream-online.html (on file with the *Journal of Corporation Law*).

4. Georgia Wells, Jeff Horwitz & Deepa Seetharaman, *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, WALL ST. J. (Sept. 14, 2021), https://www.wsj.com/articles/facebook-knows-

Hunter Biden laptop scandal.[5] Together, these examples illustrate the intensifying scrutiny digital platforms face over how they manage and explain content moderation.

The Hunter Biden laptop episode is especially revealing of how quickly content moderation decisions can spiral into public controversy, even when they have little actual effect. In that case, several platforms, in the final weeks of the 2020 election, temporarily restricted access to or deprioritized a *New York Post* story detailing Hunter Biden's Ukrainian business dealings and suggesting possible involvement by then-candidate Joe Biden.[6] In response, conservatives decried the incident as evidence of state-controlled media.[7] Yet experts, drawing on available data, argued that if suppression was the goal, the platforms failed: the story still reached a broad audience.[8] The incident did not reveal effective censorship, but a broader failure of communication. Fueled by a lack of transparency, the controversy deepened public distrust and intensified calls for regulatory oversight.[9]

This Note contends that a well-designed federal transparency law would be in the interest of most platforms, which have little to hide. Supreme Court Justice Louis Brandeis once famously noted, "[s]unlight is said to be the best of disinfectants."[10] Transparency, Brandeis explained, "will aid the investor in judging [] the safety of the investment" by leveling the information playing field of the market.[11] Today, instead of fighting "money trust[s]," well-crafted transparency laws can serve platform interests by publicizing the difficult trade-offs behind content moderation and algorithmic decisions.[12] Although some platforms have introduced self-regulating transparency initiatives, the resulting reports are often inadequate and inconsistent. Thus, a uniform standard would create equitable benchmarks for similarly situated platforms and counteract incendiary political rhetoric by demonstrating that most content moderation decisions are made in good faith rather than driven by corporate malfeasance.

Part II examines platform transparency through key terms: platform, trust and safety, automated detection, transparency reports, and Section 230. These terms highlight how

---

instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_lead_pos7&mod=article_inline (on file with the *Journal of Corporation Law*).

5. The Daily, *A Misinformation Test For Social Media*, N.Y. TIMES, at 6:24–8:45 (Oct. 21, 2020), https://www.nytimes.com/2020/10/21/podcasts/the-daily/hunter-biden-new-york-post-twitter-facebook.html (on file with the *Journal of Corporation Law*).

6. *See* Chris Mills Rodrigo, *Twitter, Facebook Clamp Down on New York Post Article About Hunter Biden*, THE HILL (Oct. 14, 2020), https://thehill.com/policy/technology/521072-twitter-facebook-clamp-down-on-new-york-post-article-about-hunter-biden/ [https://perma.cc/RF82-F82S] (reporting that Twitter cited its "hacked materials policy," while Facebook did not report what triggered its policy but did discuss a recent policy about misinformation).

7. The Daily, *supra* note 5, at 9:17–11:10.

8. *See* David Gilbert, *Facebook Failed Miserably in Its Attempt to Stop the Hunter Biden Story*, VICE (Oct. 23, 2020), https://www.vice.com/en/article/4ady3g/facebook-failed-miserably-in-its-attempt-to-stop-the-hunter-biden-story [https://perma.cc/L5CW-APMZ] (offering data to suggest that the story was still viewed over 54 million times on Facebook).

9. James Czerniawski & Jess Miers, *Twitter Hid the Hunter Biden Story. That's Not a Good Reason to Nuke Section 230*, WASH. EXAM'R (Oct. 24, 2020), https://www.washingtonexaminer.com/opinion/431775/twitter-hid-the-hunter-biden-story-thats-not-a-good-reason-to-nuke-section-230/ (on file with the *Journal of Corporation Law*).

10. Louis D. Brandeis, *What Publicity Can Do*, HARPER'S WKLY., Dec. 20, 1913, at 10.

11. *Id.* at 12.

12. *Id.* at 10.

federal policy debates about platforms have overlooked transparency as a regulatory tool. To illustrate how these concepts are applied in practice, this Note examines two significant platform moderation controversies, which reveal how moderation is often misunderstood by the public and misrepresented in political debates. Next, Part II reviews recent state-level transparency laws, with a particular focus on Florida and Texas. Florida and Texas are notable because they were passed in response to claims of conservative bias and go beyond general or limited transparency to mandate that platforms carry specific content and require explanations to individuals for content that has been removed. These laws have been extensively litigated, culminating in the Supreme Court's recent decision in *Moody v. NetChoice, LLC*.

Part III identifies legal and political trends suggesting that a federal transparency law may increasingly serve the interests of many platforms. This Note argues that *Moody* highlights a strategic dilemma: the more platforms claim that transparency rules violate their First Amendment rights, the greater the risk that courts will treat algorithmic outputs as expressive first-party speech, narrowing Section 230 protections. A single federal standard could mitigate this risk, preempt a patchwork of state laws, and clarify legal questions surrounding algorithmic discovery. It would also allow platforms to demonstrate that their moderation practices are consistent and grounded in policy, rather than driven by one-sided political partisanship. In a landscape shaped by high-profile but ultimately shallow gestures, such as the selective release of the "Twitter Files" or the symbolic authority of Facebook's Oversight Board, a credible legal framework would distinguish between performative efforts and the serious, often costly transparency work that some platforms are already undertaking in good faith. And for platforms like TikTok, already the subject of sustained scrutiny over content governance and former foreign ownership, transparency is likely to persist as a regulatory flashpoint and may become an ongoing condition for continued access to the U.S. market.

Part IV proposes changes to the Platform Accountability and Transparency Act (PATA). This federal bill aims to make platform data accessible to researchers to support oversight and accountability. Though promising in principle, the Act in its current form risks overburdening smaller platforms and keeping the public in the dark. This Section calls for raising the threshold for applicability, expanding public access to aggregated moderation data, and embedding transparency into platform design through tools that give users meaningful insight and control.

II. BACKGROUND

A. Five Keywords in Platform Transparency Debates

Understanding the debate around content moderation and platform transparency laws requires knowledge of five keywords, starting with the concept of a *platform*. Platforms are a new type of corporate organization—often associated with social media but encompassing services like email, search engines, and e-commerce marketplaces—whose business centers are built around digital infrastructures designed to "host, organize, and circulate users' shared content or social interactions . . . without having produced or commissioned (the bulk of) that content," while processing this data "for customer service,

advertising, and profit."[13] Microsoft researcher Tarleton Gillespie expands on this definition by emphasizing that "[m]oderation is not an ancillary aspect of what platforms do. It is essential, constitutional, definitional. Not only can platforms not survive without moderation, they are not platforms without it."[14] According to Gillespie, the tension surrounding platforms that present themselves as open, yet are subject to constant moderation, creates a core contradiction: platforms position themselves as neutral spaces, yet continuously decide what users see and share.[15]

If a platform is not neutral, it is essential to understand who decides what users see and share. This leads to the next key term: *trust and safety teams*. One important agent in making decisions about content is a platform's trust and safety team, which works "toward a shared goal of ensuring online safety."[16] Trust and safety teams set and enforce policies on various issues, including addressing spam and fraud, handling law enforcement requests, and safeguarding personal information.[17] Over the past decade, trust and safety teams have become increasingly professionalized, with senior members collaborating with corporate legal departments to set platform policies that help shape recommendation systems, influencing what content is prioritized, demoted, or removed.[18]

To enforce content policies, platforms rely on a second tier of workers tasked with screening disturbing material, including graphic violence and child sexual abuse.[19] This brings the discussion to the next key term: *automatic detection*. Faced with the emotional toll and overwhelming volume of harmful content, these workers increasingly rely on automated tools.[20] Unlike community flagging, which relies on users to report and review content, automatic detection utilizes PhotoDNA and other AI-driven tools, to identify and remove harmful material at scale, often without requiring human moderator approval and before users ever encounter it.[21] The origins of these systems date back to the enforcement of the Digital Millennium Copyright Act (DMCA), which led to the development of technologies designed to detect and remove infringing content, such as pirated movies and

13. TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 18 (2018).

14. *Id.* at 21.

15. *Id.*

16. Intelligence Desk, *The History of Trust & Safety*, ALICE (Sept. 14, 2023), https://www.activefence.com/blog/the-history-of-trust-and-safety/ [https://perma.cc/S73N-G5NG].

17. *See* Databite, *Origins of Trust and Safety*, DATA & SOC'Y, at 9:00–15:00 (July 8, 2020), https://datasociety.net/library/origins-of-trust-and-safety/ [https://perma.cc/KCY9-EK5H] (discussing the role of trust and safety teams).

18. *See* On with Kara Swisher, *Social Media's Original Gatekeepers on Moderation's Rise and Fall*, VOX MEDIA PODCASTS NETWORK, at 5:20–7:10 (Jan. 27, 2025), https://podcasts.voxmedia.com/show/on-with-kara-swisher [https://perma.cc/34E4-UHCG] (interviewing Dave Willner, former Head of Content Policy at Facebook, about how automatic detection tools targeting likely disinformation and hate speech are integrated into ranking algorithms and influence the speed and reach of certain types of information).

19. *See* SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA 104–05, 180–93 (2019) (describing an outsourced operation of low-wage workers, primarily in Southeast Asia, whom platforms treat as an unpleasant necessity with minimal investment).

20. *Id.* at 98–99, 107–08, 13031; *see also* GILLESPIE, *supra* note 13, at 97–110 (discussing various technologies used for automatic detection).

21. *See* GILLESPIE, *supra* note 13, at 97–110 (discussing the history of PhotoDNA).

music.[22] Over time, tools like PhotoDNA became essential in detecting and removing other types of content, such as child sexual abuse material.[23] Today, automatic detection technologies have emerged as the dominant method for moderating and enforcing platform policies.[24]

However, while automatic detection enables platforms to scale moderation efficiently, it often results in accidental removals. For example, after banning pro-Nazi content in 2019, then-Twitter mistakenly took down an image of Captain America punching a Nazi simply because it depicted a Nazi.[25] Similarly, at the start of the pandemic in 2020, platforms removed posts about volunteer efforts to distribute masks, misapplying filters designed to prevent price gouging by banning mask advertisements.[26] In another notable example, Facebook removed a post from a Black woman that discussed racial slurs directed at her children due to a moderation algorithm's inability to understand literary context.[27] There are also more humorous examples of accidental removals. For instance, posts about Role-Playing Games (RPGs) have been mistakenly flagged as references to a dietary supplement with the same name.[28] Similarly, a comment on a cat photo account referring to a cat as a "beautiful puss" has been misinterpreted as explicit content.[29] And this list of moderation blunders could go on indefinitely.

A variety of factors cause this infinite list. First, even before automation begins, the moderation process still involves subjective judgment, often leading to disagreement about how to apply policy. At a 2018 panel hosted by the Internet Society, trust and safety professionals participated in a group exercise in applying platform policies to eight different hypothetical fact sets.[30] Although they were given the same guidelines to apply to the hypothetical scenarios, this small group did not reach a unanimous consensus on whether any

22. *See* Databite, *supra* note 17, at 15:42–19:09 (discussing Google's early development of content moderation in relation to DMCA notices).

23. GILLESPIE, *supra* note 13, at 100.

24. *See* Nafia Chowdhury, *Automated Content Moderation: A Primer*, STAN. CYBER POL'Y CTR. (Mar. 19, 2022), https://cyber.fsi.stanford.edu/news/automated-content-moderation-primer [https://perma.cc/YV7P-NLTF?type=standard] (explaining that platforms like Facebook state that they rely on automatic detection—rather than user reports—to "identify 97% of content the platform removes for violating its hate speech policies. Regulation of platforms and online information-sharing must reflect this reality").

25. Blake Montgomery, *Twitter Suspends an Account for a Cartoon of Captain America Punching a Nazi*, DAILY BEAST (Sept. 11, 2019), https://www.thedailybeast.com/twitter-suspends-an-account-for-tweeting-a-cartoon-of-captain-america-punching-a-nazi [https://perma.cc/NEH3-A57U].

26. Mike Isaac, *Facebook Hampers Do-It-Yourself Mask Efforts*, N.Y. TIMES (Apr. 5, 2020), https://www.nytimes.com/2020/04/05/technology/coronavirus-facebook-masks.html (on file with the *Journal of Corporation Law*).

27. Tracy Jan & Elizabeth Dwoskin, *A White Man Called Her Kids the N-word. Facebook Stopped Her from Sharing It.*, WASH. POST (July 31, 2017), https://www.washingtonpost.com/business/economy/for-facebook-erasing-hate-speech-proves-a-daunting-challenge/2017/07/31/922d9bc6-6e3b-11e7-9c15-177740635e83_story.html (on file with the *Journal of Corporation Law*).

28. Mike Masnick, *Content Moderation at Scale Is Impossible: Recent Examples of Misunderstanding Context*, TECHDIRT (Feb. 26, 2021), https://www.techdirt.com/2021/02/26/content-moderation-scale-is-impossible-recent-examples-misunderstanding-context/ [https://perma.cc/MF2L-BCRU].

29. *Id.*

30. *See generally* Internet Society North America Bureau, *You Make the Call: Audience Interactive*, YOUTUBE (May 15, 2018), https://www.youtube.com/watch?v=VIXGkoKfOS0 [https://perma.cc/RMX9-63T7] (demonstrating that a small group of trust and safety experts could not reach 100% consensus on any of eight hypothetical content moderation cases presented during a one-hour exercise).

of the scenarios violated policy.[31] Second, large-scale moderation tools often function as blunt instruments, typically relying on keywords or images without fully grasping the context as a human reviewer might, resulting in seemingly absurd or unfair enforcement.[32] Third, even an impressive 99.9% accuracy rate leaves room for visible mistakes given the scale of moderation. With Facebook processing over 350 million photos daily, a 99.9% accuracy rate still results in 350,000 errors, potentially creating the perception of widespread failure despite strong overall performance.[33] As an illustration of this sentiment against too many errors, consider Meta CEO Mark Zuckerberg's January 2025 announcement that the company was suspending its third-party fact-checking program and some of its misinformation-filtering algorithms.[34] In Zuckerberg's words, these systems excessively misclassify harmless content, leading to too many users being unfairly placed in "Facebook jail."[35]

The next key term, *transparency reports*, has become central to platforms' efforts to explain moderation decisions and address criticism. Emerging in the early 2010s, following Edward Snowden's revelations about government surveillance, transparency reports initially focused on data collection practices.[36] By 2018, concerns over election interference and disinformation shifted the scope of these transparency reports to cover content enforcement.[37] Platforms like Google, Facebook, and Twitter began publishing enforcement data.[38] Today, transparency reports track metrics such as flagged content, automated removals, and outcomes of appeals. For example, in the second quarter of 2025, Meta removed 21 million posts for nudity and sexual content, with 97% detected automatically.[39] On further review, 1.88 million posts were restored.[40] 879,000 of which required a direct appeal.[41] Although these transparency reports help illuminate content enforcement trends, they still leave many questions about moderation unanswered.

Finally, *Section 230*, a legal provision frequently invoked in platform moderation debates, is the last keyword to understand. Enacted as part of the Communications Decency Act of 1996, Section 230 provides safe harbor protections that shield platforms from

31. *Id.*

32. Masnick, *supra* note 28; *see also* Chowdhury, *supra* note 24.

33. Mike Masnick, *Masnick's Impossibility Theorem: Content Moderation at Scale Is Impossible to Do Well*, TECHDIRT (Nov. 20, 2019), https://www.techdirt.com/2019/11/20/masnicks-impossibility-theorem-content-moderation-scale-is-impossible-to-do-well/ [https://perma.cc/G48G-QUUL].

34. Joel Kaplan, *More Speech and Fewer Mistakes*, META (Jan. 7, 2025), https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/ [https://perma.cc/D5HU-PRP2].

35. *Id.*

36. Kevin Bankston, Ross Schulman & Liz Woolery, *Getting Internet Companies to Do the Right Thing—Case Study #3: Transparency Reporting*, NEW AM., https://www.newamerica.org/in-depth/getting-internet-companies-do-right-thing/case-study-3-transparency-reporting/ [https://perma.cc/S6AX-SKQ8].

37. Arcangelo Leone de Castris, *Types of Platform Transparency: An Analysis of Discourse Around Transparency and Global Digital Platforms*, 27 PUB. INTEGRITY 340, 342 (2025).

38. *Id.* at 342–43; *see also* Robert Gorwa & Timothy Garton Ash, *Democratic Transparency in the Platform Society*, *in* SOCIAL MEDIA AND DEMOCRACY 296–99 (Nathaniel Persily & Joshua A. Tucker eds., 2020) (discussing 2018 as a turning point in the expansion of transparency reports).

39. *Adult Nudity and Sexual Activity*, META, https://transparency.meta.com/policies/community-standards/adult-nudity-sexual-activity/ [https://perma.cc/9F9N-P66V].

40. *Id.*

41. *Id.*

liability for user-generated content.[42] Notably, Section 230 originates from a time when public concern about the burgeoning internet was focused on children's exposure to "cyberporn."[43] To address these concerns, Congress sought to regulate "obscenity" and "indecency" online.[44] However, a series of early judicial rulings about the commercial internet posed a significant obstacle to the passage of the Communications Decency Act. These landmark cases suggested that forums would not face the same liability as traditional publishers, such as *The New York Times*, if they refrained from moderating content.[45] However, if these forums did moderate content, they would be liable for the content posted by their users.[46] This created a dilemma: while Congress sought to mandate content moderation to address concerns about online indecency, this mandate would have exposed early websites to legal liability. In response, Section 230 was added to the Communications Decency Act to shield early internet forums engaging in "Good Samaritan" moderation by defining them as information distributors rather than publishers.[47] Although much of the Communications Decency Act was struck down by the United States Supreme Court a year after its passage, Section 230 and its crucial safe harbor provision endured, laying the groundwork for Section 230 to become a foundational law in modern internet governance.[48]

The distinction that Section 230 provides between a publisher and a distributor plays a critical role in litigation, often leading to the quick dismissal of tort suits and lowering legal costs for technology companies.[49] Without this protection, many digital platforms would likely adopt stricter moderation algorithms, leaving users with only the most basic,

---

42. Communications Decency Act of 1996, 47 U.S.C. § 230(c)(1).

43. Philip Elmer-Dewitt, *Online Erotica: On a Screen Near You*, TIME (July 3, 1995), http://content.time.com/time/subscriber/article/0,33009,983116,00.html [https://perma.cc/2CM7-CZVB]; Robert Cannon, *The Legislative History of Senator Exon's Communications Decency Act: Regulating Barbarians on the Information Superhighway*, 49 FED. COMMC'N L.J. 51, 64 (1996).

44. Cannon, *supra* note 43, at 77–78.

45. *Compare* Cubby, Inc. v. CompuServe Inc., 776 F. Supp. 135, 140 (S.D.N.Y. 1991) (holding that online form CompuServe that did not engage in content moderation akin to a distributor, and therefore, liability did not attach for defamation claim), *with* Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 031063/94, 1995 WL 323710, at *5 (N.Y. Sup. Ct. May 24, 1995) (holding that online forum Prodigy was a publisher subject to liability for defamation because its "conscious choice, to gain the benefits of editorial control, has opened it up to a greater liability than CompuServe and other computer networks that make no such choice").

46. *See Cubby Inc.*, 776 F. Supp. at 140.

47. Cannon, *supra* note 43, at 63–64; *see also* Christopher Cox, *The Origins and Original Intent of Section 230 of the Communications Decency Act*, UNIV. RICHMOND J.L. & TECH. (2020), https://jolt.richmond.edu/2020/08/27/the-origins-and-original-intent-of-section-230-of-the-communications-decency-act/ [https://perma.cc/WBE4-3YJG] (offering an account from Section 230 drafter on its origins and intent).

48. Reno v. ACLU, 521 U.S. 844, 882 (1997).

49. *See, e.g.*, David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Ac*t, 43 LOY. L.A. L. REV. 373 (2010) (analyzing 186 cases in which Section 230 emerged as an issue and finding that defendants won dismissal in 76% of the cases studied). Even though Section 230 was initially imagined to apply only to speech posted by third parties, it has been broadly applied to include non-publication torts and been applied to other areas of law. *See, e.g.*, Airbnb, Inc. v. San Francisco, 217 F. Supp. 3d 1066, 1072–73 (N.D. Cal. 2016) (applying Section 230 to a local ordinance requiring registration of short terms rentals); Hinton v. Amazon, LLC, 72 F. Supp. 3d 685, 687 (S.D. Miss. 2014) (applying Section 230 to a products liability suit involving a recalled product sold on Amazon).

uncontroversial content, like cat videos or cooking tutorials.[50] For this reason, the Electronic Frontier Foundation calls Section 230 "one of the most valuable tools for protecting freedom of expression and innovation on the Internet."[51] Others have praised it as the "Magna Carta of cyberspace,"[52] "the twenty-six words that created the internet,"[53] and even claimed it is "better than the First Amendment."[54] However, with such freedom comes controversy, and in recent years, Section 230 has become a symbol of the tech giants' corrosive influence.[55] Notably, in the 2020 United States presidential election, both major-party candidates pledged to reform or repeal the law.[56] Nevertheless, regardless of one's stance, Section 230 has helped create a landscape in which platforms have few mandated custodial obligations for policing content, and how they police is left mainly to their discretion.

The keywords in this Part highlight a recent evolution in platform moderation efforts. While platforms have increasingly adopted automated detection tools and expanded transparency reporting, federal discussions of platforms have heavily centered on Section 230. This gap reflects a growing disconnect: as platforms develop on a safe harbor model that insulates them from liability, policy debates often overlook transparency as a regulatory approach.

B. Public Misconception About Moderation

Now that keywords related to platform transparency have been explained, real-world cases help illustrate how moderation operates—and why it often sparks controversy. This Section examines two key examples—Facebook's Trending Topics during the 2016 election and Tumblr's 2018 ban on sexual content—demonstrating that misunderstandings about moderation arise across the political spectrum. While platforms are not neutral and inevitably shape public discourse, these cases highlight the complexity of moderation decisions and their unintended consequences. Moreover, they illustrate that when platforms

---

50. *See* Steve Randy Waldman, *The 1996 Law That Ruined the Internet*, THE ATL. (Jan. 3, 2021), https://www.theatlantic.com/ideas/archive/2021/01/trump-fighting-section-230-wrong-reason/617497 [https://perma.cc/MRZ7-LKZ9] (discussing how a shift in public policy would likely result in more homogenous or palatable content).

51. Cindy Cohn & Jamie Williams, *20 Years of Protecting Intermediaries: Legacy of "Zeran" Remains a Critical Protection for Freedom of Expression Online*, ELEC. FRONTIER FOUND. (Nov. 14, 2017), https://www.eff.org/deeplinks/2017/11/20-years-protecting-intermediaries-legacy-zeran-remains-critical-protection [https://perma.cc/4FTV-QU5K].

52. Noa Yachot, *The 'Magna Carta' of Cyberspace Turns 20: An Interview with the ACLU Lawyer Who Helped Save the Internet*, ACLU (June 23, 2017), https://www.aclu.org/blog/free-speech/internet-speech/magna-carta-cyberspace-turns-20-interview-aclu-lawyer-who-helped [https://perma.cc/Q3DS-XQY7].

53. JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 2 (2019).

54. Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 NOTRE DAME L. REV. 33, 33 (2019).

55. Adi Robertson, *How America Turned Against the First Amendment*, THE VERGE (Nov. 2, 2022) https://www.theverge.com/23435358/first-amendment-free-speech-midterm-elections-courts-hypocrisy [https://perma.cc/Y6K2-HRFW] (discussing Section 230 as a synecdoche for bipartisan panic about the power of Silicon Valley platforms).

56. *Id.*; *see also* Morgan Chalfant, *Trump Accuses Twitter of Unfair Targeting After Company Labels Tweet 'Glorifying Violence'*, THE HILL (May 29, 2020), https://thehill.com/homenews/administration/500078-trump-accuses-twitter-of-unfair-targeting-after-company-labels-tweet [https://perma.cc/6GQ2-SZBZ].

fail to communicate their decisions, speculation fills the gaps, often leading the public to have a distorted understanding of how moderation operates.

If there is a catalyst event that brought platform moderation into the public spotlight, it is the controversy surrounding the demise of Facebook's now-defunct Trending Topics feature.[57] Trending Topics was billed as showcasing the most popular and widely discussed news stories on Facebook.[58] That perception was upended in the summer of 2016, when *Gizmodo* reported several Facebook "news curators" told the outlet that the company "would routinely 'blacklist' stories that were actually organically trending on the social network but were generated by conservative news sources."[59] Facebook responded to these claims of bias in the Trending Topics feature by removing human editors and eliminating a policy that required stories to be verified before being featured on the Trending Topics page.[60]

A likely explanation for the controversy was not an intentional effort to suppress conservative news, but rather Facebook's news verification policy at the time. According to internal policy documents, human editors ensured highly discussed stories were authentic by requiring at least five mainstream sources to report on a story before it appeared in Trending Topics.[61] As a result, stories lacking coverage from mainstream sources would not be featured.[62] The removal of this policy had immediate consequences. Within just three days, Facebook promoted false stories, including claims that *Fox News* had fired then-Trump critic Megyn Kelly and that the Pope had endorsed Trump.[63] This event sparked

---

57. In 2018, Facebook retired its Trending Topics product in the United States in response to the difficulties of moderating news content. *See* Deepa Seetharaman, *Facebook to Drop Trending-Topics Feature*, WALL ST. J. (June 1, 2018), https://www.wsj.com/articles/facebook-to-drop-trending-topics-feature-1527877457 (on file with the *Journal of Corporation Law*) (speculating that the perception of the feature was biased, along with pressure not to circulate misinformation or disinformation, led to the retirement).

58. *Id.*

59. Michael Nunez, *Former Facebook Workers: We Routinely Suppressed Conservative News*, GIZMODO (May 9, 2016), https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006 [https://perma.cc/R92G-K6UW]; John Cook, *The Story Behind the Story That Created a Political Nightmare for Facebook*, HUFFPOST (Aug. 10, 2018), https://www.huffpost.com/entry/facebook-gizmodo-gawker-trending-conservatives_n_5b6c9b16e4b0530743c83f58 [https://perma.cc/JGY2-GKHX] (discussing an article by Michael Nunez, *Former Facebook Workers: We Routinely Suppressed Conservative News*).

60. Alex Johnson & Matthew DeLuca, *Facebook's Mark Zuckerberg Meets Conservatives Amid 'Trending' Furor*, NBC NEWS (May 18, 2016), https://www.nbcnews.com/tech/social-media/facebook-s-mark-zuckerberg-meets-conservatives-amid-trending-furor-n576366 [https://perma.cc/357Y-7DJE] (reporting that Facebook CEO Mark Zuckerberg met with members of the GOP following the Trending Topics controversy); Colin Stretch, *Response to Chairman John Thune's Letter on Trending Topics*, META (May 23, 2016), https://about.fb.com/news/2016/05/response-to-chairman-john-thunes-letter-on-trending-topics/ [https://perma.cc/357Y-7DJE]. Colin Stretch, Facebook's General Counsel, penned an open letter claiming that no employee wrongdoing was found relating the Trending Topics controversy but the company would be implementing changes to Trending Topics anyways. *Id.*

61. FACEBOOK, TRENDING REVIEW GUIDELINES 10, https://s3.documentcloud.org/documents/2830513/Facebook-Trending-Review-Guidelines.pdf [https://perma.cc/9RX3-W8UH].

62. The Guardian obtained this document as part of its investigative reporting. *See* Sam Thielman, *Facebook News Selection Is in the Hands of Editors Not Algorithms, Documents Show*, THE GUARDIAN (May 12, 2016), https://www.theguardian.com/technology/2016/may/12/facebook-trending-news-leaked-documents-editor-guidelines [https://perma.cc/GK6Y-EBZN] (providing links and context for Facebook's 21-page Trending Review Guidelines marked as "Internal Facebook Use Only").

63. Sam Thielman, *Facebook Fires Trending Team, and Algorithm Without Humans Goes Crazy*, THE GUARDIAN (Aug. 29, 2016), https://www.theguardian.com/technology/2016/aug/29/facebook-fires-trending-

debate over the definition of "fake news," as various groups clashed over its meaning, making it a heated public topic during and after the election.[64]

These events are widely credited as kicking off a "tech-lash."[65] Public focus on "fake news" expanded to include Russian influence operations on platforms during the 2016 election.[66] Attention also turned to Cambridge Analytica, a data firm that sought to use Meta's data to build psychological profiles of millions of voters for micro-targeting campaigns in both the 2016 presidential election and the Brexit referendum.[67] Despite scrutiny, the platforms' actual impact remained unclear, fueling speculation that moderation failures had tipped the scales in the 2016 election.[68] Some experts cautioned that fears about disinformation and voter manipulation were likely exaggerated.[69] But in the absence of meaningful transparency, the public was left to fill in the gaps.[70]

Turning to a different example, Tumblr's 2018 purge of sexual content illustrates how policy changes can reshape platform moderation choices in ways the public often fails to appreciate. In late 2018, Tumblr, a year after Verizon had acquired the platform, announced

---

topics-team-algorithm [https://perma.cc/PK6B-7ZQ7]; Craig Silverman & Jeremy Singer-Vine, *The True Story Behind the Biggest Fake News Hit of the Election*, BUZZFEED NEWS (Dec. 16, 2016), https://www.buzzfeednews.com/article/craigsilverman/the-strangest-fake-news-empire [https://perma.cc/262E-3548].

64. *See* Johan Farkas & Jannick Schou, *Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood*, 25 J. EUR. INST. COMMC'N & CULTURE, 298, 303–07 (2018) (discussing shifting uses of the term fake news during the 2016 election).

65. Rachel Botsman, *Dawn of the Techlash*, THE GUARDIAN (Feb. 10, 2018), https://www.theguardian.com/commentisfree/2018/feb/11/dawn-of-the-techlash [https://perma.cc/UTG5-PZ2Q]; Jane Mayer, *How Russia Helped Swing the Election for Trump*, THE NEW YORKER (Sept. 24, 2018), https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump (on file with the *Journal of Corporation Law*) (discussing various analyses of how platforms were used in the election and ongoing public speculation).

66. *Id.*

67. Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, THE GUARDIAN (Mar. 17, 2018), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [https://perma.cc/FS2F-2RYQ].

68. Mayer, *supra* note 65.

69. *See* Andrew M. Guess, Brendan Nyhan & Jason Reifler, *Exposure to Untrustworthy Websites in the 2016 U.S. Election*, 4 NAT. HUM. BEHAV. 472, 472 (2020) (finding that "fake news" made up only a small percentage of people's information diets and was consumed by those who already had strong ideological preferences); *see also* Brendan Nyhan, *Fake News and Bots May Be Worrisome, but Their Political Power Is Overblown*, N.Y. TIMES (Feb. 13, 2018), https://www.nytimes.com/2018/02/13/upshot/fake-news-and-bots-may-be-worrisome-but-their-political-power-is-overblown.html (on file with the *Journal of Corporation Law*) ("Twitter, for instance, reported that Russian bots tweeted 2.1 million times before the election—certainly a worrisome number. But these represented only 1 percent of all election-related tweets and 0.5 percent of views of election-related tweets."); *see also* Elizabeth Nolan Brown, *Cambridge Analytica Was Doing Marketing, Not Black Magic*, REASON MAG. (Mar. 19 2018), https://web.archive.org/web/20180322072147/http://reason.com/blog/2018/03/19/cambridge-analytica (on file with the *Journal of Corporation Law*) (discussing inaccurate press coverage of the Cambridge Analytica scandal and quoting political scientist Eitan Hersh that "every claim about psychographics etc. made by or about [Cambridge Analytica] is BS").

70. The Mueller Report became a central focus of the national debate over digital disinformation and foreign interference, with many hoping it would offer definitive answers about the role social media played in the 2016 election. Its findings, however, left key questions unresolved and did little to settle broader concerns about platform influence. *See* Ross Douthat, *The Mueller Exposé*, N.Y. TIMES (Apr. 20, 2019), https://www.nytimes.com/2019/04/20/opinion/sunday/mueller-report-trump.html (on file with the *Journal of Corporation Law*).

a ban on "real-life human genitals or female-presenting nipples," as well as any content depicting sex acts.[71] This represented a significant shift for Tumblr, which had long allowed explicit content and served as a space for LGBTQ, sex-positive, and body-positive communities.[72]

The ban was likely driven by pressure from Apple, which removed Tumblr from the IOS App Store.[73] This removal occurred as many platforms started to implement new policies in response to FOSTA-SESTA, a law enacted in 2018 that modified Section 230 to make it so platforms could be held liable for facilitating or supporting sex trafficking.[74] Tumblr stated that the ban was prompted by the discovery of at least one instance of child abuse material on its platform.[75] However, experts observed that the ban could be attributed to the enactment of FOSTA-SESTA, which heightened scrutiny from business partners, such as payment processors, concerned about potential criminal and civil liability.[76] Faced with these pressures, Tumblr implemented new moderation tools.

These new moderation algorithms, designed to target explicit content, were overly broad, leading to the removal of entirely safe-for-work material.[77] Mistakenly flagged content included educational blogs, fan art, sketches depicting people hugging or kissing, and photos documenting gender transitions because they referenced sexuality or gender identity.[78] Ironically, even the platform's ban announcement was flagged and removed when users reposted it.[79] These removals underscored the algorithm's inability to differentiate between prohibited and acceptable material.[80] And despite a federal policy shift and algorithmic flaws driving this collateral censorship, many saw the sanitation of the platform as a calculated move to suppress LGBTQ and sexual content for advertisers, with users

71. Paris Martineau, *Tumblr's Porn Ban Reveals Who Controls What We See Online*, WIRED (Dec. 4, 2018), https://www.wired.com/story/tumblrs-porn-ban-reveals-controls-we-see-online/ (on file with the *Journal of Corporation Law*); Steven Thrasher, *What Tumblr's Porn Ban Really Means*, THE ATL. (Dec. 7, 2018), www.theatlantic.com/technology/archive/2018/12/tumblr-adult-content-porn/577471/ (on file with the *Journal of Corporation Law*).

72. Martineau, *supra* note 71.

73. *Id.*

74. *Id.*

75. Chance Miller, *Tumblr Was Removed from the App Store After It Was Found to Be Serving Child Pornography*, 9TO5MAC (Nov. 19, 2018), https://9to5mac.com/2018/11/19/why-tumblr-was-removed-from-app-store/ [https://perma.cc/58FK-9B6N].

76. Martineau, *supra* note 71.

77. Hudson Hongo, *Tumblr's Porn Filter Flags Its Own Examples of 'Permitted' Nudity*, GIZMODO (Dec. 17, 2018), https://gizmodo.com/tumblrs-porn-filter-flags-its-own-examples-of-permitted-1831151178 [https://perma.cc/GGZ6-UHNY].

78. Fandom Is My Fandom, *List of Banned Searches/Tags on Tumblr Purge*, TUMBLR (Nov. 25, 2018), https://web.archive.org/web/20181126062130/https://meeedeee.tumblr.com/post/180508518516/list-of-banned-searchestags-on-tumblr (on file with the *Journal of Corporation Law*); Michael Kan, *Tumblr's Child Porn Crackdown Ensnares Legit Blogs in Purge*, PCMAG (Nov. 19, 2018), https://www.pcmag.com/news/tumblrs-child-porn-crackdown-ensnares-legit-blogs-in-purge [https://perma.cc/2WKJ-C6ZC].

79. Hongo, *supra* note 77.

80. Paige Leskin, *A Year After Tumblr's Porn Ban, Some Users Are Still Struggling to Rebuild Their Communities and Sense of Belonging*, BUS. INSIDER (Dec. 20, 2019), www.businessinsider.com/tumblr-porn-ban-nsfw-flagged-reactions-fandom-art-erotica-communities-2019-8 [https://perma.cc/89NH-VLA4]; Kathryn Macapagal, *How Tumblr's 'Adult Content' Ban Could Hurt LGBTQ Teens*, REWIRE NEWS GRP. (Dec. 12, 2018), www.rewire.news/article/2018/12/12/how-tumblrs-adult-content-ban-could-hurt-lgbtq-teens/ [https://perma.cc/AGX8-PDE5].

framing it as an attempt to satisfy "Verizon's corporate greed."[81] This reaction reflects a broader misunderstanding of moderation, in which algorithmic errors and the collateral effects of policy shifts are often perceived as intentional acts of platform bias rather than as unintended consequences of automated enforcement.

The examples presented here from Facebook and Tumblr are not intended to deny that platforms have, at times, enforced policies aligned with geopolitical agendas. Evidence, albeit contested, suggests that TikTok suppresses negative content about China, with topics like Tiananmen Square and Tibetan independence conspicuously absent from the platform.[82] Similarly, Elon Musk's acquisition of Twitter (now X) and his rollback of

81. *See* Carolyn Bronstein, *Pornography, Trans Visibility, and the Demise of Tumblr*, 7 TSQ: TRANSGENDER STUD. Q. 240, 241–45 (2020) (describing public understanding of Tumblr's demise in the LGBTQ community).

82. The question of whether TikTok suppresses or amplifies content in alignment with Chinese Communist Party (CCP) interests remains highly contested, with early reporting based on leaked documents and later studies offering divergent empirical findings. *Compare, e.g.*, Alex Hern, *Revealed: How TikTok Censors Videos That Do Not Please Beijing*, THE GUARDIAN (Sept. 25, 2019), https://www.theguardian.com/technology/2019/sep/25/revealed-how-tiktok-censors-videos-that-do-not-please-beijing [https://perma.cc/R26R-JMQT] (reporting, based on leaked internal policy documents, that TikTok's moderation algorithms were designed to avoid amplifying content on politically sensitive topics, including Tiananmen Square, Tibetan independence, Falun Gong, and references to political leaders such as Xi Jinping and Vladimir Putin), *with TikTok vs. Douyin: A Security and Privacy Analysis*, CITIZEN LAB (Mar. 22, 2021), https://citizenlab.ca/2021/03/tiktok-vs-douyin-security-privacy-analysis/ [https://perma.cc/4D8B-QZ66] (finding no clear empirical evidence of systematic censorship of political terms in TikTok's search bar, but noting that algorithmic amplification may still reflect CCP-aligned narratives). More recent claims of pro-CCP bias in TikTok's content delivery systems have largely originated from researchers at Rutgers University's Network Contagion Research Institute (NCRI), which has published the most sustained empirical work on the subject to date. In a 2023 report, NCRI and affiliated researchers found that TikTok systematically promotes content favorable to the CCP. NETWORK CONTAGION RESEARCH INSTITUTE, A TIK-TOK-ING TIMEBOMB: HOW TIKTOK'S GLOBAL PLATFORM ANOMALIES ALIGN WITH THE CHINESE COMMUNIST PARTY'S GEOSTRATEGIC OBJECTIVES 1 (2023), https://networkcontagion.us/reports/12-21-23-a-tik-tok-in-timebomb-how-tiktoks-global-platform-anomalies-align-with-the-chinese-communist-partys-geostrategic-objectives [https://perma.cc/2S2C-KC28]. The study compared hashtag frequencies on TikTok and Instagram and reported that politically sensitive content, such as references to Uyghurs, Tiananmen Square, and Tibet, appeared far less frequently on TikTok, while narratives aligned with CCP foreign policy goals, such as pro-Kashmir independence content, were disproportionately amplified. *Id.* Following publication of the NCRI 2023 report, TikTok restricted public access to its Creative Center search tool, stating that "some individuals and organizations have misused the Center[] . . . to draw inaccurate conclusions." Sapna Maheshwari, *TikTok Quietly Curtails Data Tool Used by Critics*, N.Y. TIMES (Jan. 8, 2024), https://web.archive.org/web/20240109205120/https://www.nytimes.com/2024/01/08/business/media/tiktok-data-tool-israel-hamas-war.html (on file with the *Journal of Corporation Law*). However, the NCRI study has drawn criticism for methodological limitations, including its failure to account for differing user demographics across the two platforms, making comparisons between the two platforms' content trends limited in probative value. *See, e.g.*, Paul Matzko, *Lies, Damned Lies, and Statistics: A Misleading Study Compares TikTok and Instagram*, CATO INST. (Jan. 2, 2024), https://www.cato.org/blog/lies-damned-lies-statistics-misleading-study-compares-tiktok-instagram [https://perma.cc/ED9F-WMWE] (offering a list of criticisms of the NCRI report). In response to these criticisms, NCRI researchers investigated TikTok's search function using a similar methodology, finding that its search results overwhelmingly suppressed anti-CCP content compared to those on Instagram and YouTube, with few negative videos appearing for terms such as "Tiananmen Square," "Tibet," and "Uyghurs." *See e.g.*, JOEL FINKELSTEIN ET AL., THE CCP'S DIGITAL CHARM OFFENSIVE: HOW TIKTOK'S SEARCH ALGORITHM AND PRO-CHINA INFLUENCE NETWORKS INDOCTRINATE GENZ USERS IN THE UNITED STATES (2024), https://networkcontagion.us/reports/the-ccps-digital-charm-offensive/ [https://perma.cc/26TV-8FC5]. A key takeaway from this growing body of literature is that while there is substantial anecdotal evidence of content suppression and some empirical support for algorithmic bias that favors pro-CCP narratives, the extent of this influence remains

moderation policies appear tied to his growing alignment with right-wing political movements.[83] But TikTok and Twitter are high-profile cases that represent only one facet of a broader, more complex ecosystem. What distinguishes them is their ownership and governance structures.[84]

Taken together, the cases discussed in this Part illustrate a central dilemma facing all platforms today: without transparency, the public is left to speculate. Has corporate culture driven a decision? Financial incentives? Algorithmic design flaws? Automated detection tools? Legal and regulatory pressure? Such uncertainty breeds distrust, making platform moderation a persistent source of controversy.

C. Recent and Proposed Platform Transparency Laws

Several states, including California, Florida, Maryland, New York, Texas, and Washington, have enacted platform transparency laws, signaling a growing legislative trend in response to moderation controversies. Three notable trends emerge from these efforts. The first, exemplified by Washington and Maryland, specifically targets election law and political advertising. The second, exemplified by Florida and Texas, takes a broad approach to transparency while also attempting to limit the ability to moderate or remove certain forms of speech. The third, exemplified by New York and California, focuses on limited disclosure requirements for statutorily enumerated moderation categories, such as how platforms define and handle hate speech and online harassment. At the federal level, several proposals have echoed these efforts, reflecting a patchwork of state-level ideas scaled up for national debate. This Part surveys these developments to provide a snapshot of the evolving legal landscape around platform regulation in the United States.

The first major category of state-level transparency laws focuses on political advertising and election-related disclosures. One of the earliest examples came in 2018, when

---

contested, with room for disagreement about which specific platform functions, such as search and recommendation, are implicated.

83. *See* Valerie Richardson, *Babylon Bee CEO: Twitter's Bee Lockout May Have Been 'Last Straw' for Elon Musk*, WASH. TIMES (Apr. 4, 2022), https://web.archive.org/web/20220404185748/https://www.washingtontimes.com/news/2022/apr/4/babylon-bee-ceo-twitters-bee-lockout-may-have-been/ (on file with the *Journal of Corporation Law*) (discussing how Twitter's suspension of the satirical conservative news site *The Babylon Bee* contributed to Elon Musk's decision to purchase the platform); *see also The Musk Bump: Quantifying the Rise in Hate Speech Under Elon Musk*, CTR. FOR COUNTERING DIGIT. HATE (Dec. 6, 2022), https://counterhate.com/blog/the-musk-bump-quantifying-the-rise-in-hate-speech-under-elon-musk/ [https://perma.cc/CCA5-LA2V] (analyzing the increase in hate speech on Twitter following Musk's acquisition and policy changes); Charlie Warzel, *Elon Musk Is a Far-Right Activist*, THE ATL. (Dec. 11, 2022), https://www.theatlantic.com/technology/archive/2022/12/elon-musk-twitter-far-right-activist/672436/ (on file with the *Journal of Corporation Law*) (arguing that Musk's actions as Twitter's owner, including reinstating far-right figures and dismantling moderation policies, reflect his political alignment with right-wing movements).

84. Both platforms have distinct ownership structures that complicate direct comparisons to others. TikTok, owned by the China-based company ByteDance, faces ongoing scrutiny due to the Chinese government's broad legal authority over domestic firms, including laws that may compel companies to cooperate with state intelligence efforts. Similarly, Twitter is privately owned by Elon Musk, making it unusually susceptible to the views and priorities of a single individual. Unlike publicly traded or IPO-driven platforms such as Meta and Snap, which are subject to obligations to shareholders, TikTok and Twitter operate without that layer of oversight. However, this accountability is attenuated by dual-class share structures that confer outsized voting power to founders and early executives. While structurally distinct from private ownership, these arrangements raise parallel concerns regarding transparency, accountability, and the concentration of decision-making authority in the hands of a few.

Maryland passed the Online Electioneering Transparency and Accountability Act in response to mounting concerns about foreign interference, particularly following the 2016 Russian disinformation campaign. The law required online platforms to disclose the identities of political ad purchasers and how much they spent.[85] But the law's broad definition of "platform" encompassed several members of the press, prompting a legal challenge from *The Washington Post*.[86] The U.S. Court of Appeals for the Fourth Circuit struck down the law as applied to the press, warning that it risked giving the government excessive influence over newsroom operations; however, the Court left unresolved whether the law could still apply to tech platforms.[87] Washington State later passed a similar political advertising statute, which survived a legal challenge from Meta in December 2024, following the company's appeal of a $35 million judgment for noncompliance.[88]

The second major category of transparency laws requires platforms not only to explain takedown decisions and offer appeals but also limit their discretion over what content remains up or is taken down.[89] In doing so, these laws blur the line between a private company and a public utility, treating platforms more like common carriers or town squares that function as limited public forums subject to constraints on viewpoint discrimination. This legislative trend gained momentum in 2021, when Florida and Texas enacted similar laws in response to the controversy surrounding President Trump's removal from social media following the January 6th Insurrection.[90] Although the laws of Texas and Florida share many similarities, they differ in several key ways. Notably, Texas imposes broader obligations than Florida because its law applies to smaller platforms and all users, not just journalistic enterprises and politicians. And by some accounts, Texas's individual notice explanations surpass the transparency standards of the European Union's Digital Services Act.[91]

---

85. MD. CODE ANN., ELEC. LAW § 13–405.

86. Wash. Post v. McManus, 944 F.3d 506, 511–12 (4th Cir. 2019).

87. *See id.* at 518–20; *id.* at 513 ("[T]he ultimate issue before us is a narrower one, i.e., whether the Maryland Act as applied to these particular plaintiffs is unconstitutional. To that end, we do not expound upon the wide world of social media and all the issues that may be pertinent thereto.").

88. *See* State v. Meta Platforms, Inc., 560 P.3d 217, 224, 243, 248 (Wash. Ct. App. 2024) (rejecting Meta's arguments that the First Amendment protections established in *McManus* applied and that the law was preempted by Section 230).

89. TEX. H.B. 20 §§ (3)–(4), 120.052, 120.101–104; FLA. STAT. §§ 501.2041 (1), (2)(a)–(d), (2)(e)–(h). Both Texas and Florida's laws share common transparency elements requiring platforms generally to (1) notify individual users affected by content moderation decisions, (2) provide explanations for removing posts or accounts, (3)offer an appeals process to challenge these decisions, and (4) publish general transparency reports. For each statute's common carrier requirement, see FLA. STAT. § 501.2041(1)(j); TEX. H.B. 20 §§ 120.102(b), 143A.002.

90. Pooja Salhotra, *Does the First Amendment Apply to Social Media Moderation? The U.S. Supreme Court Will Decide.*, TEX. TRIB. (Feb. 26, 2024), https://www.texastribune.org/2024/02/26/texas-social-media-law-supreme-court/ [https://perma.cc/PQ3V-XDMN]; Bob Unruh, *Conservative Leaders Call for Breakup of Big Tech After It Declared 'War on Free Speech'*, WORLD NET DAILY (Jan. 12, 2021), https://www.wnd.com/2021/01/conservatives-call-breakup-big-tech-declared-war-free-speech/ [https://perma.cc/6R3P-U9ZB].

91. Daphne Keller, *Platform Transparency and the First Amendment*, 4 J. FREE SPEECH L. 1, 14 (2023). Former Associate General Counsel for Google, Daphne Keller, argues that detailed reporting and explanation provisions in the Texas statute apply to individual user notice requirements for removed content. *Id.* at 14, 34–36. The provisions that concern Keller, (TEX. H.B. 20 §§ 120.051(a) & 120.053(7)), mandate that platforms provide "accurate information" about their "content management, data management, and business practices," including a "description of each tool, practice, action, or technique used in enforcing the policy." TEX. H.B. 20

Texas v. Florida Platform Transparency Laws

Category	Florida	Texas
Must Carry Focus Applied To	Only journalistic enterprises & political candidates.[92]	All users must be carried without "viewpoint discrimination."[93]
Platform Rule Change Notification	30-day notice for affected groups.[94]	Not specified.
Minimum User Requirement for Legal Applicability	100 million monthly users.[95]	50 million monthly users.[96]
Individual Notice	Standard notice that content has been removed or account restricted.[97]	Stricter, detailed explanations are required.[98]
Remedies	Actual & Punitive Damages. Statutory Damages: $100,000 per claim. Injunctive Relief. Attorney's Fees.[99]	Injunctive Relief. Attorney fees.[100]

Due to the potentially burdensome nature of these laws, a trade group, NetChoice, representing major platforms, challenged Texas and Florida in court.[101] The case reached the Supreme Court in *Moody v. NetChoice, LLC*,[102] the details of which are covered in Part II.D. A central argument in that litigation was that, even without the constitutionally

§§ 120.051(a) & 120.053(7). Per the statute, this "specific information" on how the platform "curates and targets content to users" ensures users have the details needed to make an "informed choice" about using or purchasing platform services. *Id.* However, given the vast number of tools, practices, actions, and techniques involved in content moderation—from automated filters and machine learning models to human review protocols and internal escalation procedures—this requirement could result in massive, unwieldy disclosures detailing every decision-making process in a way that is impractical for both platforms and users. According to Keller, this amount of disclosure would also far surpass the European Union's Digital Services Act. Keller, *supra* note 91, at 14–36.

92. FLA. STAT. § 501.2041(2)(a).

93. TEX. H.B. 20 §§ 120.102(b), 143A.002.

94. FLA. STAT. § 501.2041(1)(g)–(j).

95. *Id.* at § 501.2041(1)(g)(b).

96. TEX. H.B. 20 §§ 120.102(b), 143A.002.

97. FLA. STAT. §§ 501.2041(2)(g)–(i), (3)(a)–(d).

98. TEX. H.B. 20 §§ 120.053, 120.102(b), 143A.002.

99. FLA. STAT. § 501.2041 (5)–(6) (allowing the State or private individuals to bring a cause of action to receive statutory damages of $100,000, as well as actual and punitive damages).

100. TEX. H.B. 20 § 143A.007–8 (allowing the State or private individuals to bring a cause of action to receive declaratory or injunctive relief and recover attorney fees).

101. Krista Chavez, *NetChoice Wins at Supreme Court Over Texas and Florida's Unconstitutional Speech Control Schemes*, NETCHOICE (July 1, 2024), https://netchoice.org/netchoice-wins-at-supreme-court-over-texas-and-floridas-unconstitutional-speech-control-schemes [https://perma.cc/5JHH-4535].

102. *Id.*

suspect must-carry provisions, these laws would incentivize platforms to adjust their moderation practices to appease partisan enforcers rather than incur compliance costs: leaving content up to avoid enforcement actions and thereby reducing editorial autonomy, in violation of the First Amendment.[103]

The third major trend in platform transparency laws focuses on disclosure requirements for specific content categories, such as hate speech, disinformation, harassment, and extremism. That model has influenced recent laws in states such as California and New York. In 2022, both states adopted legislation requiring platforms to disclose moderation policies and enforcement outcomes for select categories.[104] In early 2023, the U.S. District Court for the Northern District of New York struck down New York's law, holding that it forced platforms "to weigh in on the debate about the contours of hate speech when they may otherwise choose not to speak."[105] In response, New York revised its law to require disclosures only if a platform already defines prohibited hate speech in its terms of service.[106] A similar conflict unfolded in California.[107] Although the state's law initially withstood challenge due to lack of standing, the U.S. Court of Appeals for the Ninth Circuit sided with Elon Musk's X Corp in February 2025, striking down the law's requirement to define and report on six specific categories of content as unconstitutional compelled speech.[108] Following that decision, California agreed to narrow the law. Platforms must now post their terms of service and biannual enforcement summaries but are no longer required to disclose how they define or moderate content related to hate speech, extremism, or misinformation.[109]

Unlike state-level efforts, recent federal proposals focus less on regulating content and more on enabling independent research into platform operations. A leading example is the

103. Keller, *supra* note 91, at 21 (arguing that "platforms will find that simply doing what state enforcers or litigants demand—changing individual editorial decisions or overall rules for online speech—is the cheapest and safest choice").

104. *California's Newsom Signs Bill Requiring Social Media Firms' Transparency*, REUTERS (Sept. 14, 2022), https://www.reuters.com/technology/californias-newsom-signs-bill-requiring-social-media-firms-transparency-2022-09-14/ (on file with the *Journal of Corporation Law*); Sara Grace Kennedy & Gillian Vernick, *New York Wades into Social Media Regulation Waters with 'Hateful Conduct' Law*, REPS. COMM. FOR FREEDOM PRESS (June 27, 2022), https://www.rcfp.org/new-york-hateful-conduct-law/ [https://perma.cc/2MXS-Y9B6]. For a more recent legislative effort from New York in response to legal challenges, see Press Release, N.Y. St. Assembly, Assemblymember Grace Lee & Senator Hoylman-Sigal Introduce 'Stop Hiding Hate' Bill to Hold Social Media Companies Accountable for Eliminating Harmful Content on Their Platforms (May 19, 2023), https://assembly.state.ny.us/mem/Grace-Lee/story/107273 [https://perma.cc/N37N-NSZL] (offering new legislation); Press Release, Off. Governor N.Y., Governor Hochul Signs Online Safety Legislation to Strengthen Protections for the Personal Data of Consumers (Dec. 24, 2024), https://www.governor.ny.gov/news/governor-hochul-signs-online-safety-legislation-strengthen-protections-personal-data-consumers [https://perma.cc/9P6X-3M3U] (enacting new terms of service hate speech disclosure law).

105. Volokh v. James, 656 F. Supp. 3d 431, 442 (S.D.N.Y 2023).

106. Press Release, Off. Governor N.Y., *supra* note 104; N.Y. GEN. BUS. L. § 1102.

107. A.B. 1027, 2023 Gen. Assembly, Reg. Sess. (Cal. 2023).

108. *See* Minds, Inc. v. Bonta, No. 2:23-cv-02705, 2023 WL 6194312, at *1–2 (C.D. Cal. Aug. 18, 2023) (rejecting challenge due to a lack of standing); Tyler Katzenberger, *California Agrees to Drop Parts of Social Media Law Challenged by Elon Musk's X*, POLITICO (Feb. 24, 2025), https://www.politico.com/news/2025/02/24/california-drop-parts-social-media-law-challenged-elon-musk-x-00205890 [https://perma.cc/5K9S-SBGS].

109. Katzenberger, *supra* note 108 (explaining the bounds of the California law governing social media platforms' terms of service).

Platform Accountability and Transparency Act (PATA), introduced by Senator Chris Coons in 2023.[110] The bill tasks the Federal Trade Commission (FTC) and the National Science Foundation (NSF) with establishing standards that would allow researchers to study internal platform data.[111] Platforms with over 50 million monthly users would be required to report on widely shared content, algorithmic rankings, and moderation practices.[112] Researchers approved under the program would gain access to these datasets, with judicial review available if platforms deny access due to technical concerns.[113] To encourage participation, the bill includes a safe harbor shielding both researchers and platforms from liability when they comply with the FTC's rules.[114] A related proposal, the Algorithmic Justice and Online Platform Transparency Act, goes further by empowering the FTC to investigate algorithmic discrimination.[115] Yet despite growing interest in regulation, neither bill has advanced beyond the proposal stage.

As these trends indicate, platform transparency regulation in the United States is rapidly evolving. With Congress unable to reach a consensus on a federal approach, states are stepping up to take the lead. While some state laws have failed judicial review (New York), others have been fully upheld (Washington) or upheld in part (California). As the next Section discusses, Florida and Texas's laws have not been struck down entirely and were recently reviewed by the United States Supreme Court, which provided limited guidance on their constitutionality.

D. Outlook of NetChoice Litigation

NetChoice challenged the laws of Texas and Florida in federal district courts on several grounds, with the appropriate standard of review under the First Amendment emerging as the central issue.[116] Texas and Florida advocated for the *Zauderer* standard, derived from a case on lawyer advertising, which permits the State to implement disclosure requirements for "factual and uncontroversial information" in consumer protection contexts.[117] NetChoice, however, argued that *Zauderer* was inapplicable, contending that

110. *See* Platform Accountability and Transparency Act, S. 1876, 118th Cong. (2023).

111. *Id.* § 3(a).

112. Platform Accountability and Transparency Act, S. 1876, 118th Cong. §§ 2(5)(b), 9(b)–(d). These disclosures are intended to help the public understand the "prevalence and size of the problem of hate speech, disinformation, incitement, child endangerment, and the like." Justin Hendrix, *Transcript: Senate Hearing on Platform Transparency*, TECH POL'Y PRESS, (May 5, 2022), https://www.techpolicy.press/transcript-senate-hearing-on-platform-transparency/ [https://perma.cc/WPH8-UDW3] (statement of Professor Nathaniel Persily).

113. S. 1876, 118th Cong. § 4(e). Notably, the statute includes language that presumably would allow platforms to contest requests that "materially burden the technical operation of a platform." *Id.* § 8(a)(5).

114. *Id.* §§ 8(a), 4(d).

115. Press Release, Sen. Edward J. Markey, Sen. Markey, Rep. Matsui: Let's Ban Big Tech's Black-Box Algorithms That Perpetuate Discrimination, Inequality, and Racism in Society (July 13, 2023), https://www.markey.senate.gov/news/press-releases/sen-markey-rep-matsui-lets-ban-big-techs-black-box-algorithms-that-perpetuate-discrimination-inequality-and-racism-in-society [https://perma.cc/T3HR-A396].

116. Issues initially argued at the district level included the Dormant Commerce Clause, Section 230 federal pre-emption, and the platforms' First Amendment rights to set editorial policy without state interference. Chavez, *supra* note 101.

117. *See* NetChoice, LLC v. Paxton, 49 F.4th 439, 485 (5th Cir. 2022) (agreeing with the State of Texas that *Zauderer* controls); NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1209–10 (5th Cir. 2022) (agreeing with the State of Florida that *Zauderer* controls but only protects the general disclosure requirements); Zauderer v. Off. Disciplinary Couns., 471 U.S. 626, 651 (1985).

platforms function more like newspapers than legal advertisements.[118] Citing *Miami Herald v. Tornillo* and its progeny, NetChoice maintained that platforms, like private newspapers, cannot be forced to publish content, such as a political candidate's editorial.[119] Alternatively, even if *Zauderer* applied, NetChoice argued that the disclosure mandates in the Texas and Florida laws—requiring platforms to disclose their editorial processes for public-interest reasons—would still trigger heightened First Amendment scrutiny.[120]

These analogies were deeply flawed, yet each party strategically invoked them to support its preferred level of judicial review. NetChoice relied on *Tornillo* to argue for strict scrutiny to protect its members' business models from regulation.[121] In contrast, the states turned to *Zauderer* to advocate for rational basis review, seeking broad authority to regulate content moderation.[122]

The Eleventh and Fifth Circuits diverged sharply when applying these First Amendment precedents. NetChoice persuaded the Eleventh Circuit to enjoin the majority of Florida's law on First Amendment grounds.[123] The Eleventh Circuit held that the "must carry requirements" were unlikely to withstand strict scrutiny and that "the individual notice requirement" was "unduly burdensome and likely to chill platforms' protected speech."[124] However, the Court upheld Florida's general disclosure provisions as they did not impose burdens on speech or editorial discretion.[125] In contrast, the Fifth Circuit reached a fundamentally different conclusion regarding Texas's law.[126] Unlike the Eleventh Circuit, the Fifth Circuit determined that content moderation activities and the algorithms that drive them are not "speech" under *Zauderer* and do not trigger First Amendment protections.[127] And even if these moderation activities were considered speech, the Texas law was justified by a compelling government interest in promoting viewpoint diversity.[128] Furthermore, the Fifth Circuit found the transparency and individual notice requirements manageable and not overly burdensome on speech, as the affected platforms already had moderation systems in place that just needed to be "scale[d] up."[129]

118. *Paxton*, 49 F.4th at 451, 455–56; *Att'y Gen., Fla.*, 34 F.4th at 1208, 1210. For the leading case on requiring actors to host third-party speech, see *Miami Herald v. Tornillo*, 418 U.S. 241 (1974).

119. *Paxton*, 49 F.4th at 451, 455–56; *Att'y Gen., Fla.*, 34 F.4th at 1208, 1210.

120. *See Paxton*, 49 F.4th at 487–88 (responding to NetChoice invoking *Herbert v. Lando*, 441 U.S. 153, 173 (1979)). In *Herbert v. Lando*, the Supreme Court allowed a defamation plaintiff to obtain judicially guided discovery into a newspaper's editorial process, rejecting the newspaper's argument that such discovery would create a chilling effect and violate First Amendment protections afforded to the press. However, dictum from *Herbert* suggested that such discovery, if conducted for the general welfare, would likely be subject to strict scrutiny. *See* Herbert v. Lando, 441 U.S. 153, 174 (1979) ("There is no law that subjects the editorial process to private or official examination merely to satisfy curiosity or to serve some general end such as the public interest; and if there were, it would not survive constitutional scrutiny as the First Amendment is presently construed.").

121. *Att'y Gen., Fla.*, 34 F.4th at 1208.

122. *Id.* at 1227, 1230.

123. *Id.* at 1209, 1216, 1230.

124. *Id.*

125. *Id.* at 1231.

126. NetChoice, LLC v. Paxton, 49 F.4th 439 (5th Cir. 2022).

127. *Id.* at 448, 459 n.8.

128. *Id.* at 482.

129. *Id.* at 487.

With a circuit split, the United States Supreme Court granted certiorari to review the law's moderation restrictions and individual notice requirements.[130] A unanimous judgment, but not a unanimous opinion, was issued on July 1, 2024.[131] Justice Kagan, writing for the majority, explained that because the challenge to Texas and Florida's laws was facial rather than as-applied, NetChoice faced a high bar: it needed to show that "a substantial number of [the law's] applications [were] unconstitutional."[132] Neither circuit had correctly applied this standard, so the cases were remanded to their respective district courts for further factfinding.[133]

However, the majority opinion offered further guidance to the lower courts on remand: reaffirming the constitutional principle that the government cannot compel private parties to host expressive content.[134] The opinion also reaffirmed that public accommodation laws focused on non-expressive activities are constitutional.[135] While Justice Kagan was not required to apply these principles to the specific laws at issue, given the factual barriers that required a remand, she nonetheless chose to do so, concluding that Texas and Florida's laws requiring platforms to carry specific content are likely unconstitutional when applied to inherently expressive features, such as Facebook's newsfeed and YouTube's homepage.[136] However, Kagan and the other justices left unresolved whether other platform functions, such as private message services, might be constitutionally subject to must-carry requirements.[137] Nor did the Justices provide clear guidance on what standard to apply to the general transparency provisions of the laws. Instead, the Court signaled that the constitutionality of these laws must be evaluated feature by feature, while making clear that the First Amendment's prohibition on viewpoint discrimination constrains state

130. NetChoice, LLC v. Paxton *Consolidated with:* Moody v. NetChoice, LLC, SCOTUS BLOG, https://www.scotusblog.com/cases/case-files/netchoice-llc-v-paxton/ [https://perma.cc/6HUN-UPF8].

131. Moody v. NetChoice, LLC, 603 U.S. 707 (2024).

132. *Id.* at 723.

133. *Id.* at 726.

134. *Id.* at 743–45. For cases cited invoking the expressive content principle, see Pac. Gas & Elec. Co. v. Pub. Util. Comm'n, 475 U.S. 1, 12 (1986) (holding unconstitutional a state law that required a utility company to include in its mail to customers energy perspectives from a consumer advocacy group); *see also* Turner Broadcasting Sys., Inc. v. FCC, 512 U.S. 622, 643–44 (1994) (holding that cable operators enjoy free speech rights in determining what channels to carry, but declining to answer if must-carry rules at issue violated this right); Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp., Inc., 515 U.S. 557, 572–73 (1995) (holding that a municipal government's refusal to issue a permit for a St. Patrick's Day parade unless the organizers included a gay and lesbian group pursuant to a local anti-discrimination law violated the First Amendment because a parade is inherently expressive and the forced inclusion would alter the organizers' message).

135. *Moody*, 603 U.S. at 730–31; For public accommodation cases cited, see PruneYard Shopping Ctr. v. Robins, 447 U.S. 74, 100 (1980) (upholding California law requiring shopping malls to allow members of the public to distribute pamphlets on mall property); *see also* Rumsfeld v. F. for Acad. & Institutional Rts., Inc., 547 U.S. 47, 64 (2006) (upholding a federal statute requiring law schools to allow the military to participate in on-campus recruiting because law school recruiting services lack expressive quality).

136. *Moody*, 603 U.S. at 739–40.

137. *See id.* at 744 (emphasizing that the lower court "must then decide which of the law's applications are constitutionally permissible and which are not, and finally weigh the one against the other"); s*ee also id.* at 745 (Barrett, J., concurring) (opining that "dealing with a broad swath of varied platforms and functions in a facial challenge . . . [is] a daunting, if not impossible, task"); *id.* at 751, 788 (Alito, J., concurring) (opining that the common carrier doctrine likely applies to platform features like Gmail that carry messages rather than curate them).

action—not the editorial decisions of private platforms—because the government may not "tilt[] public debate in a preferred direction."[138]

Oddly, despite being a unanimous judgment, the ruling produced four concurring opinions, including one from Justice Alito that read like a dissent.[139] Alito placed the term "content moderation" in scare quotes, remarking sarcastically that this is a "gentle-sounding term used by internet platforms to denote actions they take purportedly to ensure that user-provided content complies with their terms of service and 'community standards.'"[140] Both Alito and Thomas departed from the majority by squarely addressing the common-carrier doctrine, noting that factual gaps in the case hindered its application before the Court, leaving it open for lower courts to apply.[141] Platform algorithm-driven moderation, Alito opined, bears no resemblance to an editor deciding what appears on a newspaper's front page.[142] Moreover, Alito emphasized that platforms' "network effects" on public discourse set them apart from traditional publishers, requiring greater caution when applying First Amendment precedents.[143] Building on this technological divergence, Justice Barrett similarly cautioned that artificial intelligence (AI), as a non-human actor, may not automatically receive the same First Amendment protection as human speakers.[144] In the end, with four concurrences accompanying a unanimous judgment, the Court left several legal theories open for lower courts to test and refine.

In the wake of the Supreme Court's decision to vacate and remand the cases, Chris Marchese of NetChoice described the ruling as "a victory for First Amendment rights online."[145] Meanwhile, the attorney general of Florida welcomed the decision, expressing confidence that the laws would ultimately be upheld.[146] However, given the remand, it is unclear how the litigation might progress and whether a discovery fight is brewing. In November 2024, the Fifth Circuit issued additional guidance to the district court, instructing NetChoice's members to disclose how their algorithms moderate content, as it "might change the constitutional analysis."[147] The Fifth Circuit further noted that its instruction

138. *Id.* at 741.

139. *Moody*, 603 U.S. at 745 (Barrett, J., concurring); *id.* at 748 (Jackson, J., concurring); *id.* at 749 (Thomas, J., concurring); *id.* at 766 (Alito, J., concurring).

140. *Id.* at 769 (Alito, J., concurring).

141. *See id.* at 752 (Thomas, J. concurring) (explaining that "the same factual barriers that preclude the Court from assessing the trade associations' claims under our First Amendment precedents also prevent us from applying the common-carrier doctrine in this posture. At a minimum, we would need to pinpoint the regulated parties and specific conduct being regulated. On remand, however, both lower courts should continue to consider the common-carrier doctrine.").

142. *See id.* at 767 (Alito, J., concurring) (criticizing the majority for "unreflectively" assuming "the truth of NetChoice's unsupported assertion that social-media platforms—which use secret algorithms to review and moderate an almost unimaginable quantity of data today—are just as expressive as the newspaper editors who marked up typescripts in blue pencil 50 years ago").

143. *Id.* at 795.

144. *Moody*, 603 U.S. at 746 (Barrett, J., concurring) ("If the AI relies on large language models to determine what is 'hateful' and should be removed, has a human being with First Amendment rights made an inherently expressive 'choice . . . not to propound a particular point of view?'").

145. Prithvi Iyer, *Reactions to the Supreme Court's* NetChoice *Cases*, TECH POL'Y PRESS (July 2, 2024), https://www.techpolicy.press/reactions-to-the-supreme-courts-netchoice-cases-/ [https://perma.cc/882E-G4DT].

146. *Id.*

147. NetChoice, LLC v. Paxton, 121 F.4th 494, 499 (5th Cir. 2024) (instructing individual members of the NetChoice coalition to disclose additional information about their algorithms to the district court).

applied to all NetChoice members as "the same covered actor might use a different algorithm (or use the same algorithm differently) on different covered services."[148] As the case progresses, attempts to probe the inner workings of various platform algorithms will likely spark intense legal battles over discovery. NetChoice might have won a Supreme Court battle, but the litigation is likely far from over.

III. ANALYSIS

A. What Is the Effect of Moody*?*

The exact implications of *Moody* for platforms remain uncertain, with commentators offering a range of interpretations in its aftermath. Jameel Jaffer, Executive Director of the Knight First Amendment Institute, who filed a brief in the case, argued that *Moody* struck a balance by recognizing that platforms hold some degree of editorial First Amendment protection while rejecting "the argument that regulation in this sphere is categorically unconstitutional."[149] Others referred to *Moody* as "squash[ing] the common carrier theory . . . once and for all."[150] Meanwhile, litigators from Davis Wright Tremaine, who also submitted a brief in the case, contended that *Moody* extended robust First Amendment protections to algorithmic curation, potentially dooming California's Stop Addictive Feeds Exploitation (SAFE) for Kids Act and similar laws targeting addictive product design.[151]

These divergent interpretations stem mainly from the nature of the Court's opinion. Instead of crafting platform-specific rules, the Court articulated three broad principles derived from established First Amendment jurisprudence to guide the lower courts: (1) platforms that curate or organize speech into expressive content are shielded from being compelled to include third party messages they would otherwise exclude; (2) platforms are not required to include specific content simply because they allow most other forms of content; and (3) the government cannot justify content-inclusion mandates by merely citing an interest in balancing the marketplace of ideas.[152] These generalized rules, while foundational, offer only a framework, leaving lower courts with the challenging task of applying them to specific platform features on remand.

---

148. *Id.*

149. Press Release, Knight First Amend. Inst. Colum. Univ., Knight Institute Comments on Supreme Court Ruling in Cases Involving Florida and Texas Social Media Laws (July 1, 2024), https://knightcolumbia.org/content/knight-institute-comments-on-supreme-court-ruling-in-cases-involving-florida-and-texas-social-media-laws [https://perma.cc/67RD-TWVQ].

150. Iyer, *supra* note 145.

151. *See* Adam S. Sieff, Ambika Kumar & David M. Gossett, Moody *Decision Confirms First Amendment Protects Online Platforms*, DAVIS WRIGHT TREMAINE (July 11, 2024), https://www.dwt.com/insights/2024/07/scotus-moody-ruling-a-win-for-online-platforms [https://perma.cc/2MUQ-ETHW] (stating "the [C]ourt held that the First Amendment protects 'the use of algorithms' to 'personalize[]' and target particular content to particular users through a 'continually updating stream' like a news feed, whether 'based on a user's expressed interests and past activities' or on the platforms' own prioritization decisions"). This language is drawn from an analysis section of the opinion and is used by the majority to describe how platforms curate content. The opinion does not explicitly establish a rule that algorithms are inherently expressive. Instead, the majority opinion narrowly addressed expressiveness, focusing only on the front pages of platforms like YouTube and Facebook as examples of expressive conduct.

152. Moody v. NetChoice, LLC, 603 U.S. 707, 731–33 (2024).

Lower courts have begun applying *Moody* in various ways. Its facial application rule—mandating that each platform feature affected by a statute be analyzed separately before determining a law's constitutionality—has proven particularly influential. It has, for instance, helped sustain transparency mandates and child-safety regulations that might otherwise have been invalidated under the First Amendment.[153] For a short period, *Moody* was also repeatedly invoked to justify striking down online age verification laws as unconstitutional.[154] *Moody* may also expose platform algorithms to legal discovery.[155] Perhaps most intriguingly, a recent Third Circuit case concerning the wrongful death of a youth who attempted TikTok's Blackout Challenge determined that *Moody's* characterization of algorithmic products as "expressive activity" rendered them first-party speech, stripping TikTok of Section 230 immunity.[156] Additionally, the case is being implicated in cases only tangentially related to content moderation. For example, *Moody* has been invoked in wrongful-death suits targeting AI chatbots alleged to have encouraged users to commit suicide, to resist early First Amendment defenses, reasoning that AI-generated speech may not warrant the same heightened protection afforded to human expression.[157] Meanwhile, a growing body of consumer-fraud and personal-injury litigation challenging addictive product design has distinguished *Moody*, rejecting it as a categorical shield where claims focus on non-expressive platform architecture rather than content curation.[158] As these examples highlight, *Moody's* influence on judicial review of platform regulations is expanding, yet its role in shaping digital speech jurisprudence remains unsettled.

Despite the uncertainty around *Moody,* the case does offer a basis for predicting which platform regulations are permissible or impermissible under the First Amendment. General transparency requirements aimed at consumer protection are likely permissible, provided they do not impose undue burdens on speech. Individual notice requirements tied to consumer protection might also survive constitutional review if they are not overly burdensome. By contrast, laws requiring platforms to carry political speech or give detailed

153. *See* X Corp. v. Bonta,116 F.4th 888, 899, 904 (9th Cir. 2024) (upholding parts of California's AB 587, which mandated content moderation disclosures by allowing provisions in the statute that compelled speech to be analyzed separately); NetChoice, LLC v. Bonta, 761 F. Supp. 3d 1202, 1226–28, 1230–31 (N.D. Cal. 2024) (denying a preliminary injunction against California's SAFE for Kids Act, based on the reasoning that most of the law's applications were constitutional and only likely First Amendment violations were its restrictions on push notifications and a disclosure requirement focused on the number of children using a platform).

154. *See generally* NetChoice, LLC v. Fitch, 738 F. Supp.3d 753, 769, 775 (S.D. Miss. 2024) (striking down age verification); NetChoice, LLC v. Yost, 778 F. Supp. 3d 923, 948 (S.D. Ohio 2025) (holding Ohio's parental-consent social media law was a content-based restriction and therefore unconstitutional under strict scrutiny). These early lower court decisions struck down or enjoined state age-verification and age-based design mandates. To the extent those opinions applied strict scrutiny or treated such regulations as presumptively unconstitutional, they are inconsistent with *Free Speech Coalition, Inc. v. Paxton*, 606 U.S. 461, 478–79 (2025), which held that age-based regulations aimed at protecting minors from historically unprotected categories of speech are subject to intermediate scrutiny.

155. *See* NetChoice, LLC v. Paxton, 121 F.4th 494, 497 (5th Cir. 2024) (holding that fact-intensive platform questions should be answered after discovery).

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

157. *See e.g.*, Garcia v. Character Techs., Inc., 785 F. Supp. 3d 1157, 1778–79 (M.D. Fla. 2025) (citing Justice Barrett's *Moody* concurrence holding that AI systems are not necessarily First Amendment speakers).

158. *See, e.g.*, TikTok, Inc. v. Eighth Jud. Dist. Ct., 578 P.3d 640, 650–52 (Nev. 2025) (holding that Section 230 did not immunize TikTok from claims alleging misrepresentations about youth safety and addictive design features, and that the First Amendment did not bar the claims at the pleading stage).

takedown explanations are likely unconstitutional and face serious First Amendment challenges.

Yet even as these bright lines begin to take shape, *Moody* leaves unresolved one of the thorniest questions in platform law: how to distinguish between expressive content and functional design. This distinction may well become a major fault line. For instance, Gmail may appear to be a common carrier akin to the U.S. Postal Service, but the expressive content it handles pushes it into murky First Amendment territory. Similarly, algorithms that rank or sort content by date, relevance, or popularity raise difficult questions about whether they reflect expressive editorial judgment or merely operate as technical infrastructure. As courts begin to grapple with these challenges, the future of platform regulations will likely hinge on how courts draw the line between the expressive and the infrastructural.

This unresolved boundary—what might be called the Expression-Infrastructure Threshold—will likely shape the next generation of platform litigation. How should courts evaluate claims that algorithms constitute protected speech? What distinguishes an expressive algorithm from a non-expressive one? If algorithms are considered expressive first-party speech, what are the implications for Section 230 protections? And how should the First Amendment apply to the growing wave of addictive-design claims targeting platform architecture rather than content moderation? As platform regulation continues to evolve, these are pressing legal questions courts will need to resolve moving forward.

B. Inadequacies with Current Transparency Efforts

Part II.D illustrated that the public frequently misunderstands how platforms operate and moderate content. These misconceptions are not solely the public's fault. Rather, they stem from the genuine complexity and opacity that surround these systems. Scholars of science and technology studies often describe platforms as "black boxes"—systems whose internal operations, including moderation algorithms and decision-making processes, remain largely hidden from public scrutiny and regulatory oversight.[159] Although voluntary transparency reports offer some insight, they provide only a limited and often inadequate view into the mechanisms that govern online content moderation.

Transparency reports typically present aggregated metrics, such as the total volume of flagged content across categories or the percentage of content restored after appeals. While such data is informative, it provides only a distant view of how moderation works. Without additional context on how policies are applied, the public is left to guess why certain content was removed or reinstated. Individual users are also often left confused, particularly when content is removed without explanation or flagged without notice, a practice known as shadow banning.[160] And in practice, transparency reports only reach a limited audience and do little to address public misunderstanding or rebuild trust.

---

159. *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015) (expanding on the "black box" metaphor to describe how algorithms in finance, search engines, and social media operate without sufficient transparency).

160. *See generally* Kelley Cotter, *'Shadowbanning Is Not a Thing': Black Box Gaslighting and the Power to Independently Know and Credibly Critique Algorithms*, 26 INFO., COMMC'N & SOC'Y 1226 (2023) (arguing that platforms use strategic denials and opacity to undermine users' ability to critique algorithmic content moderation, a dynamic the author terms "black box gaslighting"); Laura Savolainen, *The Shadow Banning Controversy: Perceived Governance and Algorithmic Folklore*, 44 MEDIA, CULTURE & SOC'Y 1091 (2022) (investigating how

In recent years, platform transparency to third-party researchers has also begun to erode. A combination of factors—including data privacy regulations, monetization strategies, and growing concerns about protecting proprietary information from large-scale scraping—has led many companies to restrict public access to platform data, particularly through Application Programming Interfaces (APIs).[161] This retrenchment has sparked what some researchers refer to as the "API apocalypse," as scholars and watchdogs had long relied on these infrastructures to generate independent accounts of platform behavior.[162] As a result, the current transparency model does little to address public misunderstandings or illuminate the complex trade-offs platforms face in moderation decisions.

In the absence of robust transparency, much of what is known about platform architecture comes from leaks, whistleblowers, and investigative journalism. This limited visibility has fueled a broad range of critiques regarding the societal effects of platforms. Some of these critiques are grounded in substantial evidence, while others rely more heavily on speculation. For example, early concerns about "echo chambers" and "filter bubbles," which emphasized the fragmentation of public discourse and the harm caused to journalism's business model by ad tech, are well-supported.[163] Emerging studies and recent lawsuits suggest that algorithmic systems may produce discriminatory outcomes. However, the supporting evidence remains relatively limited, and courts have often deemed such claims non-actionable under existing law.[164]

---

users believe platforms secretly suppress their posts through algorithms, and how these beliefs—shaped by online discussions—contribute to mistrust in content moderation).

161. Contested narratives have accompanied the decline in data access for researchers: while platforms cite user privacy, data security, and the protection of proprietary information, critics emphasize the role of monetization strategies, reputational control, and competitive self-interest. *See e.g.*, Axel Bruns, *After the 'APIcalypse': Social Media Platforms and Their Fight Against Critical Scholarly Research*, 22 INFO. COMMC'N & SOC'Y 1544 (2019) (arguing that API shutdowns were strategically motivated to suppress independent academic criticism and insulate platforms from public-interest research); Chris Stokel-Walker, *Twitter's API Crackdown Will Hit More Than Just Bots*, WIRED (Feb. 13, 2023), https://www.wired.com/story/twitters-api-crackdown-will-hit-more-than-just-bots/ (on file with the *Journal of Corporation Law*) (noting that Twitter framed its API changes as anti-bot and pro-privacy, while critics viewed them as a move to monetize data and eliminate unmonetized access); Bobby Allyn & Tilda Wilson, *Thousands of Reddit Communities 'Go Dark' in Protest of New Developer Fees*, NPR (June 12, 2023), https://www.npr.org/2023/06/12/1181376050/reddit-communities-go-dark-protest-new-api-developer-fees [https://perma.cc/V2PQ-R6WM] (reporting that Reddit's imposition of high API fees was justified as cost recovery but broadly interpreted as an effort to eliminate third-party tools that compete with the platform's monetization model); Josh Constine, *Instagram Suddenly Chokes Off Developers as Facebook Chases Privacy*, TECHCRUNCH (Apr. 2, 2018), https://techcrunch.com/2018/04/02/instagram-api-limit/ [https://perma.cc/V6D7-NL3Q] (describing how Instagram restricted API access following the implementation of the General Data Protection Regulation (GDPR), a major EU privacy law, citing compliance concerns).

162. *See* Bruns, *supra* note 161, at 1554.

163. *See, e.g.*, ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU (2011); *see also* David Bauder, *Decline in Local News Outlets Is Accelerating Despite Efforts to Help*, AP NEWS (Nov. 16, 2023), https://apnews.com/article/local-newspapers-closing-jobs-3ad83659a6ee070ae3f39144dd840c1b [https://perma.cc/7HBV-CYMM] (noting that over 3,000 newspapers have closed due to lost revenues to ad tech with local journalism hit particularly hard).

164. *See* SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM, 3–5 (2018) (offering a monograph-length study on Google search engines results that are more likely to turn up highly sexualized and pornographic content for the term "Black girls"); *see also* Olivier Sylvain, *Platform Realism, Informational Inequality, and Section 230 Reform*, 131 YALE L.J. F. 475, 508–10 (2021) (discussing ongoing litigation over Facebook product features that permit targeting based on categories like race and gender in violation of statutes such as the Fair Housing Act); *see also* Jess Rauchberg, *Articulating Algorithmic Ableism: The*

Other, more sensational critiques—for instance, theories likening platform design to psychological operations developed by intelligence agencies—tend to lack empirical grounding.[165] While platforms undeniably shape user behavior through interface design, metrics, and incentives, there is little credible evidence that they exert the kind of manipulative control over cognition some theorists suggest.[166] Most platforms sustain their business models by monetizing user attention through targeted advertising, often summed up by the phrase "you are the product."[167] Yet competition for engagement is not the same as control over thought or behavior. Platform design influences behavior, but it does so alongside—and not above—law, culture, and market incentives.

In parallel with these debates about platform power and responsibility, platforms have increasingly automated core decisions about what content is seen, promoted, or suppressed, shifting responsibility from human moderators to algorithmic systems. Processes that were once solely dependent on human judgment, such as recommendations, are now managed by what can be described as "court[s] of algorithmic appeal."[168] In such systems, decisions about what ideas, identities, or practices are surfaced are made with minimal human oversight. The full consequences of this shift remain unclear, but they raise fundamental

---

Suppression and Surveillance of Disabled TikTok Creators, 34 J. GENDER STUD. 1138, 1138 (2025) (arguing that ableism is embedded into TikTok's algorithmic infrastructure and documenting how the platform's systems—by tracking user behavior, facial features, and video content—systematically flagged disabled users as outside platform norms, leading to automated downranking and visibility suppression under the guise of user protection). Many forms of alleged platform discrimination have struggled to gain traction in court. For example, in August 2019, a group of LGBTQ+ content creators filed suit against YouTube, alleging that the platform systematically demonetized or restricted their videos based on identity-related keywords. Nico Lang, *This Lawsuit Alleging YouTube Discriminates Against LGBTQ+ Users Was Just Tossed Out*, THEM (Jan. 8, 2021), https://www.them.us/story/lawsuit-alleging-youtube-discriminates-against-lgbtq-users-tossed-out [https://perma.cc/EF7W-TDZH]. Although the plaintiffs presented empirical evidence that terms such as "gay," "lesbian," and "transgender" triggered demonetization, the court dismissed the case in January 2021, holding that the claims were non-actionable under existing law. *Id.*

165. *Facebook Emotion Experiment Sparks Criticism*, BBC NEWS (June 30, 2014), https://www.bbc.com/news/technology-28051930 [https://perma.cc/CM5E-EWM5] (discussing a secretive emotional contagion experiment conducted to manipulate users' emotional states). Shoshana Zuboff has emerged as the leading proponent of a gamification theory, arguing that platforms are not merely promoting engagement strategies but are creating behavioral modification tools designed to shape user actions without their conscious awareness. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 296–317 (2019). For example, Zuboff theorizes scenarios in which platforms might seek to induce negative emotions, such as sadness or insecurity, to make users more susceptible to certain advertisements (e.g., promoting weight-loss products or self-help content). *Id.* However, while platforms have designed themselves in ways analogous to slot machines (e.g., infinite scroll), the extent to which platforms can actively manipulate a user's affective state remains a contested topic. *See* CORY DOCTOROW, HOW TO DESTROY SURVEILLANCE CAPITALISM 11 (2020) (arguing that platforms lack any "mind-control rays" capable of brainwashing users into voting for a presidential candidate or joining an extremist group and that many platform tools are more akin to snake oil).

166. DOCTOROW, *supra* note 165, at 10–30 (noting the lack of empirical evidence for many of Zuboff's claims about persuasion and offering an alternative account of how surveillance capitalism or AdTech persuades).

167. Olivia Solon, *You Are Facebook's Product, Not Customer*, WIRED (Sept. 21, 2011), https://www.wired.com/story/doug-rushkoff-hello-etsy/ (on file with the *Journal of Corporation Law*).

168. *See* Blake Hallinan & Ted Striphas, *Recommended for You: The Netflix Prize and the Production of Algorithmic Culture*, 18 NEW MEDIA & SOC'Y 117, 129 (2014) (discussing the implications of algorithmic sorting as it removes humans from decision-making processes, blurring the line between culture and computation).

concerns about legitimacy, accountability, and the role of platforms in shaping public discourse.

Even with greater visibility into platform operations, understanding their broader social impact would remain difficult. Cause-and-effect in systems like these is rarely straightforward. For example, while social media is often blamed for rising mental health issues among young people, many overlapping factors—social, economic, and cultural—are likely at play.[169] Platforms likely contribute to mental health struggles, but they cannot be isolated as the sole cause. No single platform action tells the whole story. The effects we observe are built from many moving parts acting together, often in unpredictable ways.

C. A Federal Solution? The Growing Case for National Platform Transparency Legislation

In light of First Amendment protections and ongoing uncertainty surrounding platform regulation, transparency laws are poised to play a central role in shaping future U.S. policy. Notably, several converging trends suggest that federal legislation may be in the interest of many platform companies. The proliferation of state-level regulations may make federal preemption an increasingly attractive solution, especially if it becomes more difficult for platforms to maintain Section 230 immunity in cases where their product design is deemed expressive under the First Amendment. In this environment, federal legislation

169. The rise of smartphones and the platform ecosystem has coincided with a significant increase in adolescent mental health problems, particularly since 2012. Jonathan Haidt, a leading advocate for adolescent mental health, observes that since smartphones became widely available, Generation Z has reported increased levels of anxiety and depression. S*ee* JONATHAN HAIDT, THE ANXIOUS GENERATION: HOW THE GREAT REWIRING OF CHILDHOOD IS CAUSING AN EPIDEMIC OF MENTAL ILLNESS 24–31, 40–44 (2024) (analyzing statistical trends in adolescent mental health across the United States and comparable nations). Beyond self-reported distress, Haidt also notes that emergency room visits for self-harm and suicide rates have also risen substantially amongst Generation Z. *Id.* at 40–44. Some research suggests the longer an adolescent spends on social media, the more likely they are to suffer from depression, anxiety and other disorders. *See, e.g.*, Jean M. Twenge, Brian H. Spitzberg & W. Keith Campbell, *Less In-Person Social Interaction with Peers Among U.S. Adolescents in the 21st Century and Links to Loneliness*, 36 J. SOC. & PERS. RELATIONSHIPS 1892 (2019). Haidt relies on studies like these to argue that smartphones fundamentally undermine well-being by fostering social deprivation, sleep deprivation, attention fragmentation and addiction. HAIDT, *supra* note 169, at 113–41. Others have taken a more skeptical view of smartphones, arguing that critics like Haidt overuse the term causation when discussing correlational studies. *See* Candice L. Odgers, *The Great Rewiring, Unplugged: Is Social Media Really Behind an Epidemic of Teenage Mental Illness?*, 628 NATURE 29, 29–30 (2024). Other leading clinical psychologists contend that while smartphones may be a contributing factor, a more balanced approach would recognize that academic pressure, overscheduling and economic uncertainty are equally significant drivers of adolescent mental health issues. *See, e.g.*, LISA DAMOUR, THE EMOTIONAL LIVES OF TEENAGERS: RAISING CONNECTED, CAPABLE, AND COMPASSIONATE ADOLESCENTS (2023) (offering a skeptical account of phones being the catalyst for the teen mental health crisis); *see also* Amy Orben & Andrew K. Przybylski, *The Association Between Adolescent Well-Being and Digital Technology Use*, 3 NATURE HUM. BEHAV. 173, 173 (2019) (finding only a small correlation between time spent using digital technology and self-reported adverse mental health). Rather than viewing smartphone critics' focus as purely diagnostic, their arguments can also be interpreted as a pragmatic response to a complex issue, as regulating smartphone use and age-gating platforms offers a bright-line intervention. *See* Jean M. Twenge, *Smartphones Are Damaging Our Kids*, NAT'L REV. (Mar. 28, 2024), https://www.nationalreview.com/magazine/2024/05/smartphones-are-damaging-our-kids/ (on file with the *Journal of Corporation Law*) (arguing that government regulation of smartphone usage offers a more feasible policy solution compared to alternative approaches, as "putting more regulation on children's and teens' use of social media is straightforward and would cost very little").

may no longer represent a threat to platform autonomy, but rather a stabilizing solution. This Section examines these intersecting trends.

1. Federal Preemption to Curb Litigation

In 2023, Senator Chris Coons introduced the Platform Accountability and Transparency Act (PATA), a legislative proposal designed to give qualified researchers access to internal platform data.[170] Under the bill, the Federal Trade Commission (FTC and the National Science Foundation (NSF) would set standards for evaluating research proposals and overseeing access to data.[171] PATA targets platforms with over 50 million monthly users and includes several provisions aimed at increasing transparency, particularly around algorithmic.[172]

The legislation has three major prongs. First, it mandates disclosures about highly disseminated content, algorithmic ranking systems, and moderation practices.[173] Notably, platforms would be required to explain the inputs to their recommender systems, how user data is used, and the extent to which specific content is amplified.[174] Second, it requires platforms to provide these data categories to approved public interest researchers, with judicial review available in cases where platforms contest a request on the grounds that compliance would pose undue technical burdens or safety risks.[175] Third, it creates a safe harbor shielding both researchers and platforms from liability, provided they comply with FTC standards—an effort to address common legal and privacy concerns that have hindered prior transparency efforts.[176]

PATA has sparked a range of reactions. Advocates view it as a long-overdue corrective that could help researchers expose platform harms such as disinformation, hate speech, and algorithmic bias.[177] Critics, however, warn of unintended consequences. Some caution that broad researcher access could become a vector for government surveillance.[178] Others argue that compliance could impose technical burdens on platforms, potentially degrading user experience.[179] Still, PATA reflects a growing consensus that meaningful oversight of

---

170. John Perrino, *Platform Accountability and Transparency Act Reintroduced in Senate*, TECH POLICY PRESS (June 8, 2023), https://www.techpolicy.press/platform-accountability-and-transparency-act-reintroduced-in-senate/ [https://perma.cc/24XU-6PM4].

171. *Id.*

172. Platform Accountability and Transparency Act, S. 1876, 118th Cong. §§ 2(5)(b), 9(b)–(d).

173. *Id.*

174. *Id.* § 9(d)(2)(A)–(D).

175. *See id.* § 8(a)(5). This provision concerns a requirement to protect against liability but may also permit platforms to object to requests that impose a material burden on their operations. Presumably, platforms could use this language to object to certain data requests, claiming that Congress only intended for projects to be performed that would not materially burden platform operations.

176. *Id.* § 8(a); *id.* § 4(d).

177. Hendrix, *supra* note 113 (statement of Professor Nathaniel Persily).

178. *Hearing on Platform Transparency: Understanding the Impact of Social Media Before the Subcomm. on Priv., Tech. & L. of S. Comm. on the Judiciary*, 117th Cong. 6 (2022) (statement of Daphne Keller, Stanford University Cyber Policy Center), https://www.judiciary.senate.gov/imo/media/doc/Keller%20Testimony1.pdf [https://perma.cc/GH8P-S99A].

179. *Hearing on Platform Transparency: Understanding the Impact of Social Media Before the Subcomm. on Priv., Tech. & L. of S. Comm. on the Judiciary*, 117th Cong. 13–14 (statement of Jim Harper, Nonresident Senior Fellow, Am. Enter. Inst.), https://www.judiciary.senate.gov/imo/media/doc/Harper%20Testimony.pdf [https://perma.cc/WAZ5-675J].

digital platforms requires access to the very systems that have, until now, operated largely in the dark.

For platforms, a federal law like PATA, despite its demands, may offer a preferable alternative to navigating an increasingly complex and fragmented regulatory environment. As Part II.B showed, states have adopted divergent approaches to platform transparency. Some—such as Washington and New York—focus narrowly on political advertising disclosures or select categories of moderation, while others—like Texas and Florida—combine transparency mandates with speech restrictions that approach compelled hosting obligations. This patchwork creates significant legal uncertainty and operational complexity, especially for national or global companies subject to overlapping and sometimes contradictory rules. Moreover, recent litigation has underscored the potential for burdensome discovery processes under state laws. In the Fifth Circuit, following the Supreme Court's remand in *Moody v. NetChoice*, platforms have faced potentially extensive discovery demands that could compel them to disclose sensitive internal documents and communications, raising the prospect of significant legal costs and operational disruptions.[180] In the light of litigation such as this, even a federal measure like PATA may ultimately be the easier option for platforms, as it would streamline their obligations and help them avoid more extreme or legally risky state laws.

2. Holding Platforms to the Same Standard

A federal transparency law would hold similarly situated platforms to a consistent standard, helping to resolve collective action problems. Without uniform requirements, platforms face little incentive to pursue meaningful transparency, especially when competitors can gain public favor through selective or superficial disclosures. Twitter under Elon Musk illustrates this dynamic. After acquiring the platform, Musk released the so-called "Twitter Files" to a select group of journalists, presenting it as a bold act of transparency.[181] At the same time, Musk dismantled much of the platform's content moderation infrastructure and withdrew from international regulatory commitments.[182] Critics noted that the disclosures were limited in scope, revealed little new information, and largely served to reinforce a preferred narrative.[183]

---

180. NetChoice, LLC v. Paxton, 121 F.4th 494, 499 (5th Cir. 2024).

181. Katherine Alejandra Cross, *The Transparency Theater of the Twitter Files*, WIRED (Dec. 12, 2022), https://www.wired.com/story/twitter-files-elon-musk-shadowbanning-censorship/ (on file with the *Journal of Corporation Law*).

182. *See Musk's Twitter Has Dissolved Its Trust and Safety Council*, NPR (Dec. 12, 2022), https://www.npr.org/2022/12/12/1142399312/twitter-trust-and-safety-council-elon-musk [https://perma.cc/DE86-TTAH] (reporting that Twitter disbanded its Trust and Safety Council, a body that advised on moderation and policy issues); Natasha Lomas, *Twitter Layoffs Trigger Oversight Risk Warning from Brussels*, TECHCRUNCH (Nov. 24, 2022), https://techcrunch.com/2022/11/24/elon-musk-twitter-layoffs-eu-dsa-vlop-warning/ [https://perma.cc/ML4P-D2ZE] (noting EU concerns about Twitter's ability to comply with the Digital Services Act following cuts to moderation and compliance teams).

183. *See* Shirin Ghaffary, *What the Twitter Files Don't Tell Us*, VOX (Dec. 9, 2022), https://www.vox.com/recode/2022/12/9/23502237/twitter-files-elon-musk-conservatives-right-wing-bari-weiss-matt-taibbi-blacklist-shadowban (on file with the *Journal of Corporation Law*) (arguing that the disclosures offered limited new insights and were selectively framed to support a political narrative); Shannon Bond, *Elon Musk Is Using the Twitter Files to Discredit Foes and Push Conspiracy Theories*, NPR (Dec. 14, 2022), https://www.npr.org/2022/12/14/1142666067/elon-musk-is-using-the-twitter-files-to-discredit-foes-and-push-

Similar concerns apply to Meta and its Oversight Board. When Mark Zuckerberg introduced the Oversight Board in 2019, he likened it to a "Supreme Court" for content moderation, with the intent that the Board would make final decisions on online speech. In reality, however, the Board has often been seen as largely symbolic, serving more to signal transparency than as a check on Meta's power.[184] In practice, the board lacks the authority to investigate Meta's broader policies and cannot compel the company to follow its rulings.[185] This performative posture has become even more apparent in the wake of Meta's recent decision to remove its misinformation filters and shut down its third-party fact-checking program. This move was broadly interpreted as a political concession to the incoming Trump administration.[186] Although presented as a major policy shift, the fact-checking system had limited reach and produced only a few reviews each day.[187] It rarely served as a tool for large-scale enforcement.[188] And because Meta offers little visibility into how it applies its rules, it remains unclear whether this rollback changes anything at all, or codifies a reality in which meaningful oversight never truly existed.

These examples reveal what happens when self-regulation turns into theater. Platforms stage moments of transparency—the "Twitter Files," an oversight board branded as

---

conspiracy-theor [https://perma.cc/SEN5-U83C] (noting that the Twitter Files contained no major revelations but revealed internal debate); Ariel Levin-Waldman, *Analysis: 'Twitter Files' Not a Smoking Gun, but They Are Significant*, I24NEWS (Dec. 6, 2022), https://www.i24news.tv/en/news/international/culture/1670351304-analysis-twitter-files-not-a-smoking-gun-but-they-are-significant [https://perma.cc/57GS-NZCL] (stating that while the Twitter Files provided some insights, they did not amount to a "smoking gun" and largely confirmed existing understandings of Twitter's content moderation practices).

184. *See* Gorwa & Ash, *supra* note 38, at 304–05 (discussing concern that oversight board could become a "transparency proxy").

185. *See generally* Happiness N. Okereke, Note, *Examining the Autonomy of Social Media Content Moderation Oversight Boards: A Case Study of Facebook's Oversight Board*, 50 J. CORP. L. 1423 (2025) (providing an overview of the corporate structure governing the Oversight Board and critiquing its limited institutional authority); Josh Cowls et al., *Constitutional Metaphors: Facebook's 'Supreme Court' and the Legitimation of Platform Governance*, 26 NEW MEDIA & SOC'Y 2448 (2024) (critiquing the metaphor of Facebook's Oversight Board as a "Supreme Court" and analyzing its role in legitimizing platform governance). Observers have also raised concerns that Meta has, at times, restricted the Oversight Board's access to information, thereby limiting its capacity to adjudicate key factual issues fully. Kara Swisher, *Inside the Decision on Trump's Facebook Fate: Interview with Alan Rusbridger of the Facebook Oversight Board*, N.Y. TIMES, at 19:15–20:14 (May 7, 2021), https://www.nytimes.com/2021/05/07/opinion/sway-kara-swisher-alan-rusbridger.html (on file with the *Journal of Corporation Law*) (interviewing Facebook oversight board member who described asking Meta 46 questions when investigating the suspension of Donald Trump's account but that there "were about two questions they didn't answer properly, and seven where they didn't give us an answer. So again, we're chipping away, but we're not getting as far as we want to.").

186. Kelvin Chan, Barbara Ortutay & Nicholas Riccardi, *Meta Eliminates Fact-Checking in Latest Bow to Trump*, AP NEWS (Jan. 7, 2025), https://apnews.com/article/meta-facts-trump-musk-community-notes-413b8495939a058ff2d25fd23f2e0f43 [https://perma.cc/Z2Z7-4PFD].

187. *See* Priyanjana Bengani & Ian Karbal, *Five Days of Facebook Fact-Checking*, COLUM. JOURNALISM REV. (Oct. 30, 2020), https://www.cjr.org/analysis/five-days-of-facebook-fact-checking.php [https://perma.cc/8JNP-W2AQ] (analyzing the limited reach and inconsistent application of Facebook's third-party fact-checking system, which during a five day study produced only seventy fact checks total, an average of 14 fact checks per day across the platform); *see also* Judd Legum, *The Facts About Facebook's Fact-Checking Program*, POPULAR INFO. (Feb. 13, 2020), https://popular.info/p/the-facts-about-facebooks-fact-checking [https://perma.cc/D855-9927] (reporting that Facebook's U.S. fact-checking partners conducted only 302 fact checks in January 2020, highlighting the program's limited scope and minimal impact on the platform's vast content).

188. Bengani & Karbal, *supra* note 187.

a Supreme Court—while avoiding the deeper accountability such gestures pretend to signal.[189] But beneath the spectacle lies a real business problem: in a fragmented regulatory landscape, meaningful transparency can become a competitive liability. Why invest in costly moderation or safety systems when rivals win headlines with selective disclosures and little follow-through?

A federal transparency standard could shift that logic. By setting uniform rules, it would eliminate the commercial penalty for doing the right thing. For platforms like Snapchat, long praised for limiting virality, avoiding algorithmic amplification, and relying on in-house moderation, such a standard would validate an approach already oriented toward responsibility.[190] And it would ease one of the central dilemmas in content governance: the fear that a stricter policy, like age verification, will drive users elsewhere.[191] If the major platforms knew they would be held to the same standard—and that enforcement practices would be visible to the public—then transparency would no longer be a risk. It would be a competitive asset. In that light, regulation is not a constraint on corporate power. It is a way to reward companies that are doing a great job.

A federal transparency regime could also help ensure that the law is applied consistently rather than selectively. While a federal standard promotes consistency, critics correctly warn that it could also be used to extract political patronage, favoring some platforms while targeting others.[192] And there is some basis for this concern. The Trump administration has been credibly accused of leveraging allies like FTC Commissioner Brendan Carr to shape enforcement of television broadcasting outcomes, and similar anxieties about "jawboning" surfaced when the Biden administration flagged accounts to platforms for potential terms-of-service violations.[193] A transparency law would not eliminate these

189. Cross, *supra* note 181.

190. *See* Sara Fischer, *Snapchat Emphasizes Human Content Moderation in App Redesign*, AXIOS (Nov. 29, 2017), https://www.axios.com/2017/12/15/snapchat-emphasizes-human-content-moderation-in-app-redesign-1513307227 (on file with the *Journal of Corporation Law*) (reporting that Snapchat relied on editorial teams for content review and deliberately limited algorithmic amplification to avoid viral misinformation); *Snapchat Moderation*, *Enforcement, and Appeals*, SNAP INC. (Dec. 2025), https://values.snap.com/privacy/transparency/community-guidelines/moderation [https://perma.cc/9VQ5-TN6M] (describing Snapchat's use of human moderation and its design choices aimed at curbing amplification, such as avoiding public reposting and algorithmic ranking).

191. Alex Heath, *Facebook's Lost Generation*, THE VERGE (Oct. 25, 2021), https://www.theverge.com/22743744/facebook-teen-usage-decline-frances-haugen-leaks (on file with the *Journal of Corporation Law*) (reporting on internal Facebook documents showing significant teen user decline and the company's concerns about losing market share to competitors like TikTok, which influenced hesitance toward stricter age verification policies).

192. *See, e.g.*, Daphne Keller & Max Levy, *Getting Transparency Right*, LAWFARE (July 11, 2022), https://www.lawfaremedia.org/article/getting-transparency-right [https://perma.cc/ANM2-6HYQ] (critiquing proposed platform transparency regimes, including PATA, and warning of potential risks in their design). More recently, commentators have raised growing concerns that legal enforcement is perceived as tied to patronage, with some describing companies as effectively paying tribute to the Trump administration. These critics are also likely to argue that a transparency regime itself could be weaponized to help exert such patronage. As an example, Amazon's Melania documentary has drawn criticism, with observers pointing to its reported $40 million production cost relative to limited returns, as well as the attendance of prominent technology executives, including Apple CEO Tim Cook, at its premiere, as evidence of a patronage system. Steven Levy, *After Minneapolis, Tech CEOs Are Struggling to Stay Silent*, WIRED (Jan. 30, 2026), https://www.wired.com/story/after-minneapolis-tech-ceos-are-struggling-to-stay-silent/ [https://perma.cc/Y3GC-CZ9E].

193. George F. Will, *Trump, Kimmel and the Upside of Ignoring Big-Government Coercion*, WASH. POST (Sept. 24, 2025), https://www.washingtonpost.com/opinions/2025/09/24/jimmy-kimmel-donald-trump-abc/

risks, but by requiring requests for data or information to follow clear procedures, it would give platforms something to point to in a legal challenge that informal government pressure has crossed a line.

3. Transparency as an Alternative to Platform Bans

A federal transparency law would also serve as a constructive alternative to recurring calls for platform bans, particularly in the case of TikTok. In 2025, the app became the center of sustained public confusion after the Supreme Court allowed a congressionally enacted ban to take effect, prompting a brief shutdown, followed by the Trump administration extending the divestiture deadline due to diplomatic, economic, and technical complications.[194] On January 22, 2026, TikTok's U.S. operations were transferred to a newly formed U.S.-controlled joint venture majority-owned by American investors, including Oracle founder Larry Ellison, thereby averting a nationwide ban by separating operational control from ByteDance.[195] Yet controversy persisted, particularly over whether ByteDance retained influence over the platform's recommendation algorithm.[196] The transition was further marred by temporary blackouts reportedly stemming from a record-breaking winter storm disrupting operations at an Oracle data center, which coincided with the January 24, 2026, killing of Alex Perretti by Immigration and Customs Enforcement (ICE) officers.[197] Because users were unable to post for a brief time during that period, the platform faced public accusations of pro-Trump administration censorship.[198] The events underscore the same pattern of speculative attribution and public confusion that Part II.B argued has become endemic to platform governance debates.

In the future, if ownership of the TikTok algorithm or another foreign-owned platform were to return to the stage as a national issue, a federal transparency standard could offer a viable regulatory middle ground. Rather than relying solely on high-stakes divestment mandates or outright bans, Congress could impose disclosure obligations that address both national security and democratic accountability. For TikTok or platforms like it, supporting

---

[https://perma.cc/EW54-2ZB4]. *See also*, Murthy v. Missouri, 603 U.S. 43, 63–69 (2024) (holding that social media anti-vaccine influencers lacked Article III standing to challenge alleged Biden Administration pressure on platforms).

194. Exec. Order No. 14,350, Further Extending the TikTok Enforcement Delay, 90 Fed. Reg. 45,903 (Sept. 16, 2025) (extending the enforcement delay of the Protecting Americans from Foreign Adversary Controlled Applications Act as applied to TikTok until Dec. 16, 2025).

195. David McCabe & Emmett Lindner, *TikTok Strikes Deal for New U.S. Entity, Ending Long Legal Saga*, N.Y. TIMES (Jan. 22, 2026), https://www.nytimes.com/2026/01/22/technology/tiktok-deal-oracle-bytedance-china-us.html (on file with the *Journal of Corporation Law*) (reporting that ByteDance reached a deal to form a majority-American joint venture to run TikTok's U.S. operations and avoid a nationwide ban).

196. Alex Turvy & Rebecca Scharlach, *U.S. Power Play Over TikTok Did Nothing to Protect Americans*, TECH POL'Y PRESS (Jan. 30, 2026), https://www.techpolicy.press/us-power-play-over-tiktok-did-nothing-to-protect-americans/ [https://perma.cc/P7PV-NBWU].

197. Kanishka Singh, *Oracle Says Data Center Outage Causing Issues Faced by U.S. TikTok Users*, REUTERS (Jan. 27, 2026), https://www.reuters.com/business/energy/oracle-says-outage-data-center-causes-issues-faced-by-us-tiktok-users-2026-01-28/ (on file with the *Journal of Corporation Law*).

198. Blake Montgomery, *Why TikTok's First Week of American Ownership Was a Disaster*, THE GUARDIAN (Feb. 1, 2026), https://www.theguardian.com/technology/2026/feb/01/tiktok-first-week [https://perma.cc/R3J3-CVTK].

a federal transparency law may offer a path toward regulatory stability and continued access to the U.S. market, particularly as an alternative to more severe remedies.

IV. RECOMMENDATION

This Note proposes targeted revisions to the Platform Accountability and Transparency Act (PATA) to better reflect the realities of platform governance, constitutional limits, and the shifting demands of digital regulation. As explored in Part III, content-neutral transparency requirements are likely to survive constitutional scrutiny after *Moody*. More importantly, transparency requirements address a fundamental challenge: the public and lawmakers still lack access to the data held by platforms that is critical to understanding the scope and mechanics of misinformation, algorithmic bias, and political polarization. In the absence of that information, policy debates proceed in the dark, increasing the likelihood of badly written laws driven by speculation rather than evidence.

A federal transparency law would not only help close this information gap but also preempt a growing patchwork of state regulations, offering platforms a more straightforward path through the courts' emerging jurisprudence on algorithms and access to data in discovery. While transparency may invite greater scrutiny, many major platforms already invest in content governance and could benefit from a legal framework that distinguishes meaningful oversight from feigned transparency. The reforms that follow aim to refocus PATA without placing undue burdens on smaller or emerging competitors.

A. Raise the Threshold of PATA

First, raising the threshold for applicability from the proposed 50 million monthly active users to 100 million would help avoid regulatory capture, ease the burden on smaller platforms, and create more space for innovation. Reddit illustrates why this threshold matters. Despite being a household name, Reddit only recently crossed the 100 million monthly active user mark in the United States—a milestone it reached around the time of its March 2024 initial public offering.[199] Reddit's IPO helps illustrate why 100 million is a sensible threshold: the company reached that scale around the time it became mature enough, both organizationally and financially, to absorb the demands of transparency compliance. A threshold set at this level would more accurately reflect which platforms have the institutional capacity to meet those demands, while sparing smaller players from regulations they may not yet be equipped to follow.

B. Alternative Models of Transparency

Second, instead of restricting data sharing to academic researchers, a more effective approach might be to emphasize public access to aggregated platform data. Exclusive access for researchers raises concerns about fairness and necessitates extensive judicial oversight to prevent potential misuse. In contrast, public availability of key data would enhance transparency while reducing the need for case-by-case legal review. Publicly accessible

199. Blake Montgomery & Agencies, *Reddit Shares Soar After Company Turns First Profit as a Public Company*, THE GUARDIAN (Oct. 30, 2024), https://www.theguardian.com/technology/2024/oct/30/reddit-stock [https://perma.cc/N2JK-QDX2] (reporting that Reddit had nearly 100 million monthly users as of late 2024, coinciding with its growth into a publicly traded company).

data could include large-scale content moderation statistics, detailing the type of content flagged, the detection method used, the restriction applied, and whether enforcement actions—such as removal or suspension—were based on Community Guidelines, legal requirements, or government requests. While public reports may offer less insight than making data available to researchers, this approach would establish a baseline for transparency without the complexities of judicial oversight or questions about appointment powers with an independent commission.[200] Of course, such an approach would need to confine itself to the disclosure of purely factual and uncontroversial information, in line with *Zauderer*, since mandating commentary on or justification of platform policies risks rendering the law content-based and triggering heightened First Amendment scrutiny.

The PATA could still provide a structured mechanism for academic access to internal data on platform decision-making, including how platforms develop and test new features or adjust ranking algorithms. However, such access should be subject to a multimember agency commission structure designed to prevent potential abuse. One option would be to vest oversight authority in an independent commission whose members serve staggered, fixed terms, thereby reducing the likelihood that the body reflects the priorities of any single administration at a given moment. This structure would promote institutional continuity and help preserve balanced representation across political perspectives. The commission would be required to evaluate applications under clearly defined statutory criteria, including institutional affiliation, methodological rigor, research purpose, the substantial or compelling state interest served by the research, and demonstrated data-security protections. Based on these factors, the commission would issue reasoned, written determinations. Limited judicial review would ensure procedural regularity, accountability, and likely as-applied First Amendment challenges.

Additionally, to help combat misuse, PATA might create a presumption that government data or information requests made outside its procedures are unlawful unless proven otherwise. This would not eliminate informal pressure or jawboning concerns that increasingly define the current era, but it would give companies statutory language that could help provide a partial check in a legal challenge.

C. Build Transparency into Products

Finally, beyond legislating transparency research, legislatures and companies should begin to consider how transparency could be directly embedded into platform design, whether through legislation or private initiatives. One effective measure might be offering users more control over algorithmic recommendations, allowing them to filter or modify content suggestions to align with their preferences. Properly structured, such a "drop-box" or preference tool would fit squarely within *Zauderer's* consumer-protection framework (thereby avoiding strict scrutiny), as it would provide users with purely factual, noncontroversial information about how content one receives is categorized and recommended. By centering transparency at the user level, this approach would also minimize the need for sustained governmental interaction with platforms and reduce concerns about regulatory

200. Although a multi-member commission with staggered terms might better promote ideological balance and institutional independence, the continuing vitality of such structures has been called into question by the Court's recent separation-of-powers decisions. *See* Seila Law LLC v. Consumer Fin. Prot. Bureau, 591 U.S. 197 (2020).

jawboning or informal pressure. As a practical example, platforms could allow users to opt out of or limit AI-generated content in their recommendation feeds—a meaningful check on the deluge of low-quality synthetic material—thereby giving users greater control over their digital environment while illuminating the mechanics of algorithmic curation.

To be sure, customization comes with trade-offs. Greater transparency and user control may narrow informational exposure and intensify preference-confirming content streams, reinforcing polarization rather than mitigating it. Users who can fine-tune their feeds may filter out disagreeable viewpoints. They may construct more homogeneous information environments. That risk is real.

But it must be weighed against the structural shift that has already occurred. Early conceptions of the internet emphasized its "pull" architecture, in which users actively sought out information, but algorithmic feeds have increasingly transformed it into a "push" medium that delivers content regardless of user intent.[201] Tools that restore meaningful user control over recommendations would partially rebalance that shift, reintroducing elements of pull media into an environment now dominated by push dynamics. In this way, product design itself can function as a regulatory lever, fostering a more competitive, user-centric digital ecosystem without inviting heightened First Amendment scrutiny or expanding avenues for governmental abuse of regulatory power.

V. CONCLUSION

This Note advocated for well-designed transparency laws to clarify the complex, often misunderstood processes behind content moderation and algorithmic decision-making. The Supreme Court's recent analysis in *Moody v. NetChoice, LLC*, underscores the constitutional viability of these measures when they are carefully designed. A reformed PATA offers a balanced solution, enabling platforms to embrace transparency as both a regulatory safeguard and a public relations asset.

201. Chris Dixon, *Two Eras of the Internet: Pull and Push*, CDIXON (Dec. 21, 2014), https://cdixon.org/2014/12/21/two-eras-of-the-internet-pull-and-push/ [https://perma.cc/83Q5-AL2D].