# Delete and Repeat: The Problem of Protecting Social Media Users' Free Speech from the Moderation Machine

Jacob Kosakowski*

*"[W]hen men have realized that time has upset many fighting faiths, they may come to believe . . . that the ultimate good desired is better reached by free trade in ideas—that the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out. That at any rate is the theory of our Constitution."*[1]

## I. INTRODUCTION

As of January 2021, more than four billion people across the world actively use social media, averaging two hours and twenty-five minutes each day on social media platforms.[2] Facebook, created in 2004 by Mark Zuckerberg, maintains the largest social media platform with about 1.66 billion active users.[3] Twitter, founded by Jack Dorsey in 2006, cultivates roughly 152 million active users on its platform.[4] As the economic growth and societal influence of social media continues to flourish, many Americans increasingly rely on platforms like Facebook and Twitter for their daily news, social commentary, and individual expression.[5] With continued growth in user numbers and profits, commentators

* J.D. Candidate, Suffolk University Law School, 2022; University of Massachusetts, Amherst, B.A. 2018. My immense gratitude goes to Professor Rebecca Curtin, whose guidance and support was invaluable to developing the writing that became this Note. I also want to acknowledge Professor Alan Gaitenby for inspiring my interest, and eventual fascination, in this topic. Finally, I want to thank all Suffolk University Law Review editors and staffers for their helpful contributions to this Note.

1. Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

2. *See* Simon Kemp, *Digital 2021: The Latest Insights into the 'State of Digital*,*'* WE ARE SOC. (Jan. 27, 2021), https://wearesocial.com/us/blog/2021/01/digital-2021-the-latest-insights-into-the-state-of-digital/ [https://perma.cc/4J7A-8JQH] (highlighting increase in social media usage to 4.2 billion users across mobile devices and computers).

3. *See World Wide Web Timeline*, PEW RSCH. CTR.: INTERNET & TECH. (Mar. 11, 2014), https://www.pewresearch.org/internet/2014/03/11/world-wide-web-timeline/ [https://perma.cc/9Z5B-AA3C] (acknowledging Facebook's development and emergence); Katie Mellinger, Comment, *The Section 230 Standoff: Safe Harbor Rollbacks Would Not Solve Alleged "Anti-Conservative Bias" in Social Media Content Moderation*, 10 WAKE FOREST J.L. & POL'Y 389, 400 (2020) (stating Facebook's active daily users).

4. *See World Wide Web Timeline*, *supra* note 3 (acknowledging Twitter's founding and emergence); Mellinger, *supra* note 3, at 400 (stating Twitter's active daily users).

5. *See* Kemp, *supra* note 2 (emphasizing year-to-year percentage growth of social media's active users and mobile users); Barrie Sander, *Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation*, 43 FORDHAM INT'L L.J. 939, 941 (2020)

worry that an ever-increasing concentration of power and control in the hands of social media corporations may threaten free speech on the internet.[6]

On the surface, social media platforms like Facebook and Twitter appear to promote free speech by allowing individual users to share content on a wide range of personal, social, and political topics.[7] While an individual user's content is often benign in nature, some social media users will trigger the implicit and explicit mechanisms used for content moderation based on that particular platform's rules of private governance.[8] Within the past two years, however, public awareness of platforms' content-moderation practices has increased through notable censorship complaints such as Parler's removal from Amazon computing services, the congressional big-tech hearings, and Democratic Senator Elizabeth Warren's complaints against Facebook.[9] Looking critically at

(identifying social media's ability to share information, expand conversation, and create cultural participation). Major online social platforms exert enormous influence over the digital public sphere, acting as a global gatekeeper of information and expression. *See* Sander, *supra*, at 955.

6. *See* Gregory Day, *Monopolizing Free Speech*, 88 FORDHAM L. REV. 1315, 1317-18 (2020) (noting observers' concerns over powerful corporations outside of First Amendment's scope threatening free speech). Facebook, Twitter, and other social media firms can exploit their expanding market power and control of the social media sphere to impede upon the free trade of ideas. *Id.* at 1318; *see* Sander, *supra* note 5, at 943-44 (expressing concern over few platforms' almost complete power and control over internet's content layer); Annemarie Bridy, *Remediating Social Media: A Layer-Conscious Approach*, 24 B.U. J. SCI. & TECH. L. 193, 195 (2018) (explaining narrative surrounding overall expansion of big tech and online platforms). While the narrative that online platforms pose a real threat to democracy is becoming more widely accepted, the underlying concern about the power of platforms to influence public discourse is justified. *See* Bridy, *supra* (expressing concern about power social media platforms wield in public discourse).

7. *See* Mellinger, *supra* note 3, at 390 (noting modern social media communications reach hundreds to thousands of consumers). Modern social media platforms have increased the inundation of information by providing access to news, politics, and entertainment, as well as unsavory content. *See id.* (describing type of content shared on platforms); *see also* Day, *supra* note 6, at 1318 (explaining social media platforms allow users to share and discuss personal, social, and political topics); Giancarlo F. Frosio, *Why Keep a Dog and Bark Yourself? From Intermediary Liability to Responsibility*, 26 INT'L J.L. & INFO. TECH. 1, 2 (2018) (describing how social media platforms increasingly shape contemporary life). Social media platforms allow almost anyone to share ideas, assertions, or images, and to disseminate numerous types of content to the world. *See* Frosio, *supra* (explaining social media's role in freedom of expression and worldwide communication).

8. *See* Sander, *supra* note 5, at 946 (distinguishing between implicit and explicit rules of content moderation). Publicly available platform community standards and terms of service most commonly include restrictions on hate speech, harassment, copyright infringement, graphic sexual or violent content, and other illegal activity. *See id.* (describing range of content restrictions platforms establish); *see also* James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 47 (2015) (noting content moderation form of governance which structures participation by facilitating cooperation and preventing abuse). *But see* Sander, *supra* note 5, at 946 (discussing how rules of private governance determine platform content moderation). For social media platforms, content moderation constitutes an "indispensable and definitional part of what platforms do." Sander, *supra* note 5, at 945 (emphasizing online platforms' mechanism for navigating abundance of information available on internet).

9. *See* Jack Nicas & Davey Alba, *How Parler, a Chosen App of Trump Fans, Became a Test of Free Speech*, N.Y. TIMES (Feb. 15, 2021), https://www.nytimes.com/2021/01/10/technology/parler-app-trump-free-speech.html [https://perma.cc/8CA4-5AW7] (noting social media platform Parler's removal from various app stores and computing services); Tony Romm et al., *Facebook, Google, Twitter CEOs Clash with Congress in Pre-Election Showdown*, WASH. POST (Oct. 28, 2020, 5:42 PM), https://www.washingtonpost.com/technology/2020/10/28/twitter-facebook-google-senate-hearing-live-updates/ [https://perma.cc/NE96-F4RD] (summarizing Senate Commerce Committee's disagreements over internet police power and speech moderation); Mahita

social media as a mechanism for free speech, it becomes apparent that social media platforms are not implementing policies to protect individual speech, but rather to monetize users' attention.[10] Onerous speech regulation contradicts the idea that social media promotes free speech or a "marketplace of ideas," leaving the individual user vulnerable to the violation of their freedom of speech.[11]

---

Gajanan, *Facebook Removed Elizabeth Warren's Ads Calling for the Breakup of Facebook*, TIME (Mar. 11, 2019, 8:54 PM), https://time.com/5549467/elizabeth-warren-facebook-breakup-ads [https://perma.cc/ZG6F-BM4R] (summarizing Facebook's justification for removing Warren's presidential campaign ads). The social media platform Parler closely copies Twitter's platform structure. *See* Nicas & Alba, *supra.* In early January 2021, "Apple and Google had removed Parler from their app stores and Amazon said it would no longer host the site on its computing services," effectively forcing Parler to shut down. *See id.* (acknowledging companies removed Parler for not sufficiently policing "posts that incited violence and crime"). Prior to Parler's removal from various hosting sites, the Senate Commerce Committee and House Judiciary Committee met separately with the CEOs of the largest technology companies to discuss government concerns over the moderation and censorship power these internet platforms hold. *See* Romm et al., *supra* (summarizing Senate Commerce Committee's disagreements over internet police power and speech moderation); *see also* Bobby Allyn & Shannon Bond, *4 Key Takeaways from Washington's Big Tech Hearing On 'Monopoly Power,'* NPR (July 30, 2020, 1:45 AM), https://www.npr.org/2020/07/30/896952403/4-key-takeaways-from-washingtons-big-tech-hearing-on-monopoly-power [https://perma.cc/C7U7-4HJB] (describing accusations of conservative speech suppression during hearing on monopoly power). During Warren's 2020 presidential campaign, she notably raised concerns over the power large technology companies maintained, and Facebook removed her campaign ads that called for the breakup of Facebook and other large technology companies. *See* Gajanan, *supra* (noting large technology companies' power creates influence in economy, politics, and rulemaking); *see also* David Shepardson, *Facebook, Google Accused of Anti-Conservative Bias at U.S. Senate Hearing*, REUTERS (Apr. 10, 2019, 5:35 PM), https://www.reuters.com/article/us-usa-congress-so-cialmedia/facebook-google-accused-of-anti-conservative-bias-at-u-s-senate-hearing-idUSKCN1RM2SJ [https://perma.cc/9XLQ-NYAN] (acknowledging Warren's statements on removal of her presidential campaign ad). "I want a social media marketplace that isn't dominated by a single censor." Shepardson, *supra*.

10. *See* Bridy, *supra* note 6, at 217 (detailing how platforms only show users what will keep their attention). A user's timeline or newsfeed is "engineered to keep them logged in and interacting with content on the platform for as long as possible." *Id.* (noting social media platforms moderate content based on users' personal preference data); *see also* Niva Elkin-Koren & Maayan Perel, *Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law*, *in* THE OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY 669, 671 (Giancarlo Frosio ed., 2020) (asserting purpose behind social media platforms' content moderation policies). Social media platforms are private, for-profit entities whose financial interests are connected to the same content they are expected to moderate. *See* Elkin-Koren & Perel, *supra* (noting platforms act like commercial players and governors of other people's speech online); *see also* Sander, *supra* note 5, at 952 (noting largest online platforms' business model involves "sale of human attention"). Major social networks generate revenue by monetizing their users' attention. *See* Sander, *supra* note 5, at 953 (describing "attentional scarcity" within digital public sphere); *see also* Mary Anne Franks, *The Miseducation of Free Speech*, 105 VA. L. REV. ONLINE 218, 239 (Dec. 30, 2019), https://www.virginialawreview.org/articles/miseducation-free-speech/ [https://perma.cc/U5R2-7J55] (emphasizing social media prioritizes engagement, controversy, and attention over education, quality, and expertise).

11. *See* Day, *supra* note 6, at 1317-18 (noting social media platforms' market power and control may impede upon free trade of ideas); Bridy, *supra* note 6, at 217 (emphasizing platforms' monetization of attention where monetary interests correlate with users' time spent on platform). Even if done properly, speech regulation can create heavy costs on society by intentionally or accidentally condemning meritorious ideas. *See* Day, *supra* note 6, at 1326 (expressing concern over banning or restricting speech containing ideas in tension with status quo); *see also* Elkin-Koren & Perel, *supra* note 10, at 670-71 (discussing how online intermediaries challenge rule of law). Social media platforms blur the line between private interests and public responsibilities, potentially circumventing constitutional safeguards for speech. *See* Elkin-Koren & Perel, *supra* note 10, at 670-71 (highlighting conflict of interest between private interests and public duties). *But see* Franks, *supra* note 10, at 238-39 (relaying how principles of free speech become misguided and misconstrued online).

In the wake of these free speech concerns on social media, many legal writers have largely written off the First Amendment as a viable means of confronting social media censorship.[12] This notion extends from the idea that private entities are not subject to the constitutional constraints of the First Amendment.[13] Throughout its history, however, the Supreme Court of the United States has applied the concept of a "public forum" to public, semipublic, and private entities to accommodate the free speech rights of individuals based on the particular use of the property.[14] Likewise, the Supreme Court has constantly reimagined free speech rights in response to ever-changing media such as radio, television, and the internet.[15] While prior Supreme Court precedent oscillated on the First Amendment's application to private entities, the Court's decision in *Packingham v. North Carolina*[16] might redefine online speech platforms as public forums and revitalize the First Amendment's applicability to social media platforms.[17]

---

12. *See* Jason Wiener, Note, *Social Media and the Message: Facebook, Forums, and First Amendment Follies*, 55 WAKE FOREST L. REV. 217, 221 (2020) (contending First Amendment only protects against government restrictions on free speech); Day, *supra* note 6, at 1317-18 (noting how powerful social media corporations lie outside First Amendment's scope); Jonathan Peters, *The "Sovereigns of Cyberspace" and State Action: The First Amendment's Application—or Lack Thereof—to Third-Party Platforms*, 32 BERKELEY TECH. L.J. 989, 991 (2017) (reviewing scholars' comments on social media existing unconstrained by First Amendment). Corporate censors can prevent citizens from accessing social media platforms in various degrees because the platforms are not governmental entities, and thus the Constitution does not safeguard their abridgement of First Amendment principles. *See* Weiner, *supra*, at 242 (asserting social media platforms may constitutionally restrict users' speech in ways government entities cannot).

13. *See supra* note 12 (highlighting how First Amendment restricts only government actors, not social media providers). *But cf.* Peters, *supra* note 12, at 999-1000 (noting tenuous applicability of Supreme Court doctrine to social media content moderation). Platforms like Facebook and Twitter offer free public access and a place to engage in expressive activities, ultimately maintaining dual public and private characteristics. *See* Peters, *supra* note 12, at 1000 (acknowledging potential applicability of First Amendment to social media platforms via public forum doctrine); *see also* Kate Klonick, Article, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1611 (2018) (predicting future litigation arguing social media platforms perform quasi-municipal functions subject to First Amendment).

14. *See* Wiener, *supra* note 12, at 223-24 (describing types of forums and examples of each); *see also* Peters, *supra* note 12, at 1020 (arguing social media platforms have replaced traditional public forums). Where once public streets and parks were held in trust for the purposes of publicly assembling and communicating, platforms such as Facebook and Twitter—while privately owned—have been "freely and openly accessible to the public" for a variety of expressive purposes. *See* Peters, *supra* note 12, at 1020-21 (comparing traditional public forums to third-party platforms). *But see* Klonick, *supra* note 13, at 1610 (highlighting inconsistency in Supreme Court's reasoning on individual rights and forums).

15. *See* Red Lion Broad. Co. v. FCC, 395 U.S. 367, 400-01 (1969) (discussing radio and First Amendment); Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 656-57 (1994) (discussing television and First Amendment); Reno v. ACLU, 521 U.S. 844, 849 (1997) (holding decency restrictions relating to internet unconstitutional under First Amendment); *see also* Klonick, *supra* note 13, at 1611 (explaining how First Amendment doctrines have developed around media like radio and television).

16. 137 S. Ct. 1730 (2017).

17. *See infra* Part II.B (describing development of First Amendment jurisprudence); *Packingham*, 137 S. Ct. at 1736-37 (acknowledging *Packingham* decision first to address relationship between First Amendment and modern internet). Writing for the majority, Justice Kennedy concluded that users of social media platforms engage in a "wide array of protected First Amendment activity." *See Packingham*, 137 S. Ct. at 1735-36; *see also* Klonick, *supra* note 13, at 1611 (explaining implications of *Packingham* decision). Although the decision only applies to a total restriction of access to social media platforms, it leaves open "questions of how robust that

This Note identifies the various problems surrounding the protection of individual users' First Amendment rights on the internet.[18] In Section II.A, the Note provides a broad overview of the internet's infrastructure and the protections the Communications Decency Act of 1996 (CDA) provides to social media platforms.[19] Section II.B also examines the history of the Supreme Court doctrine surrounding what constitutes a public forum, as well as the relationship between the First Amendment and different types of media.[20] In Part III, the Note analyzes the various legal issues leaving internet users' free speech rights unprotected from an internet platform's moderation practices.[21] Finally, Part IV of the Note argues for the need to develop legal tools and policies that protect individuals from the unprecedented power of social media platforms.[22]

## II. HISTORY

### *A. The Internet and Content Moderation*

#### *1. Two Layers, Two Purposes*

Broadly, the internet's infrastructure consists of two layers: a bottom layer where internet service providers operate and a top layer where internet applications run.[23] On the bottom layer, internet service providers operate on an end-to-end principle, which carries and routes users' data from one endpoint to another endpoint.[24] Internet Protocol (IP), the "open-standard networking protocol," allows the configuration of local area networks from all over the world to connect with one another, enabling each internet endpoint to have a unique IP address to accurately route data between any source and destination.[25] IP is the backbone of the internet's physical infrastructure because it allows online

---

access must be or where in the internet pipeline a choke point must lie in order to abridge a First Amendment right." *See* Klonick, *supra* note 13, at 1611 (noting *Packingham* decision may further extend First Amendment protections to social media users); *see also* Wiener, *supra* note 12, at 239 (suggesting *Packingham* decision equates Facebook and Twitter to "modern public square"). "In this analogy, the platforms are the forums that are like public squares and user accounts are representations of speakers with the capacity to speak in many ways." Wiener, *supra* note 12, at 239 (comparing access to social media with access to modern public square).

18. *See infra* Part III (identifying various legal problems leaving individual users' free speech rights unprotected on internet platforms).

19. *See infra* Section II.A (detailing internet's infrastructure, monetary motivations, and CDA protections).

20. *See infra* Section II.B (comparing Supreme Court decisions on public forums, radio, and television).

21. *See infra* Part III (analyzing CDA, public forum doctrine, and media regulation shortcomings relating to social media platforms).

22. *See infra* Part IV (proposing solutions to address current shortcomings).

23. *See* Bridy, *supra* note 6, at 199 (highlighting two layers of internet's infrastructure).

24. *See id.* (recognizing basic principle of internet networking). The end-to-end principle provides that the core of the network "should be functionally limited to carrying and routing data." *See id.* (defining internet's basic end-to-end principle and sole purpose of internet's bottom layer).

25. *See id.* at 200 (explaining IP and how it works). IP makes the internet possible by allowing packets of data to be routed accurately from a unique IP address at each endpoint where it can receive and send data. *See id.* (emphasizing IP makes details of data irrelevant to carrying and routing data).

developers to design and develop new services and applications without having to rely on internet service providers to build any new physical or virtual infrastructure into the network.[26] On the top layer of the internet, online platforms, applications, and services run independently of the internet's underlying bottom layer.[27]

The concept of neutrality was the main characteristic associated with the internet as an open information system.[28] In December 2017, however, the Federal Communications Commission (FCC) expressly reversed internet neutrality rules that required internet service providers to provide equal access to websites across the bottom layer of the internet.[29] Rather than carrying and treating all data equally, irrespective of content or purpose, internet service providers can now block, throttle, or prioritize access to specific platforms on the bottom layer.[30] Meanwhile, applications and services on the internet's top layer encourage content awareness, or the conscious interaction with intelligible words and images on online platforms.[31] With data surfacing as content in the form of tangible words or images, top-layer applications' desire to protect themselves from legal consequences and to protect users from exposure to harmful material compels them to moderate content appearing on their platforms and services.[32] Thus, top-

26. *See id.* at 200-01 (describing how data transfers occur between endpoints regardless of underlying wires or transfer protocols). *But see* Grimmelmann, *supra* note 8, at 51 (noting shared internet infrastructure limits users' capacity).

27. *See* Bridy, *supra* note 6, at 201 (explaining how applications and services run "on top" of internet). Internet applications include web browsers and email. *Id.*

28. *See supra* note 7 and accompanying text (describing internet's ability to openly communicate various types of information, particularly through social media platforms); Bridy, *supra* note 6, at 199 (asserting concept of neutrality central internet characteristic).

29. *See* David Shepardson, *U.S. Appeals Court Will Not Reconsider Net Neutrality Repeal Ruling*, REUTERS (Feb. 6, 2020, 6:42 PM), https://www.reuters.com/article/us-usa-internet/u-s-appeals-court-will-not-reconsider-net-neutrality-repeal-ruling-idUSKBN20032K [https://perma.cc/3ZHX-2QMB] (highlighting Trump Administration's reversal of Obama-era rules encouraging level playing field for providers and consumers); *see also* Keith Collins, *Net Neutrality Has Officially Been Repealed. Here's How That Could Affect You.*, N.Y. TIMES (June 11, 2018), https://www.nytimes.com/2018/06/11/technology/net-neutrality-repeal.html [https://perma.cc/2F5W-4LHY] (acknowledging FCC's repeal of net neutrality rules).

30. *See* Bridy, *supra* note 6, at 202-03 (explaining indiscriminate routing of data at bottom layer of internet). An end-to-end model is not specialized to carry any specific type of data or to support any particular type of application. *See id.* at 199-200 (explaining end-to-end principle nondiscriminatorily routed data before 2017); *see also* Collins, *supra* note 29 (acknowledging prior rules prohibited internet service providers from giving preferential treatment to certain websites). Before the repeal of net neutrality rules, internet service providers could not discriminate against any lawful content by blocking websites, slowing the transmission of data, or creating prioritization for faster service to companies and consumers who paid premiums. *See* Collins, *supra* note 29 (noting internet providers now able to legally block, throttle, or prioritize access to specific platforms); *see also* Shepardson, *supra* note 29 (stating ability to offer paid "fast lanes" and block or throttle traffic creates unfair access).

31. *See* Bridy, *supra* note 6, at 205 (explaining difference between internet's top and bottom layer).

32. *See id.* (describing how data presents itself on top layer and legally actionable injuries relating to content). Content on internet applications can "cause legally actionable injuries, including defamation, harassment, infliction of emotional distress, invasion of privacy, and intellectual property infringement." *Id.*; *see* Mariarosaria Taddeo & Luciano Floridi, *The Debate on the Moral Responsibilities of Online Service Providers*, 22 SCI. & ENG'G ETHICS 1575, 1584 (2016) (describing moral dilemma internet platforms face in disseminating

layer applications inherently become gatekeepers of information through the moderation of content.[33] While the bottom layer of the internet generally does not interact with individualized content, the top layer is more susceptible to content moderation due to potential public blowback for internet applications hosting disagreeable content.[34]

### *2. Liability Shield or Responsibility Shield?*

On top-layer applications, such as social media platforms, there is a basic need for some degree of content moderation.[35] Moderation consists of structural mechanisms that govern an online community in order to facilitate collaborative conversation, enhance user cooperation, and prevent abuse.[36] Because social media platforms are privately owned businesses, owners of these platforms get to decide who can moderate content, what content is subject to moderation, and how moderation will occur.[37] Broadly speaking, a social media platform's moderator can support or hide posts, encourage or denounce content, and promote or

information). While internet service providers are not liable for user-generated content, internet platforms might be responsible for the consequences of certain material hosted on their platforms. *See* Taddeo & Floridi, *supra* (describing potential responsibility of certain internet actors).

33. *See* Taddeo & Floridi, *supra* note 32, at 1583 (defining information gatekeepers). A platform is an information gatekeeper if that platform "controls access to information, and acts as an inhibitor by limiting access to . . . the scope of information," and acts as a facilitator of information. *See id*. (characterizing platforms' central role in management of information).

34. *See* Marcelo Thompson, *Beyond Gatekeeping: The Normative Responsibility of Internet Intermediaries*, 18 VAND. J. ENT. & TECH. L. 783, 805, 819 (2016) (noting differences between internet layers and proposing internet intermediaries moderate content according to public norms). Unlike platforms or services on the top layer of the internet, an internet service provider on the bottom layer only interacts with the data it routes for about a second; only a deeper-level inspection of IP data would reveal its content. *See id.* at 805 (highlighting bottom layer's inherent neutrality towards content); *see also* Bridy, *supra* note 6, at 205 (emphasizing how end-to-end principle assigns responsibility for content to internet users and platforms). On the top layer, internet users and platforms control the content of communications and shared information because users and platforms interact with the content in the form of actual words and images. *See* Bridy, *supra* note 6, at 205 (describing inherent differences of layers).

35. *See* Grimmelmann, *supra* note 8, at 45 (discussing differences between moderated and unmoderated online communities). A key factor in generating healthy online communities is moderation. *See id.* (comparing moderated communities to flourishing, healthy environments and comparing unmoderated communities to anarchy).

36. *See id.* at 47 (defining moderation, its purpose, and its basic importance in creating healthy online communities for communication); *see also* Sander, *supra* note 5, at 945 (describing critical and definitional role of content moderation in what online platforms do).

37. *See* Grimmelmann, *supra* note 8, at 49 (noting platform owners' unchallenged control). Owners of platforms are in a privileged position because their control over the community's rules allow them to dictate the flow of content from authors to readers through content moderation. *See id*. (emphasizing impact of owners' control of content moderation on social media platforms).

ban users.[38] A moderator's decision will determine what users see, influence what the online community values, and reinforce the platform's content norms.[39]

In the United States, the CDA protects social media platforms' decisions as to what content to moderate.[40] Initially, Congress added section 230 to the CDA to protect children from viewing inappropriate online material.[41] In doing so, section 230 protects internet applications from being treated as speakers or publishers of any user-generated content hosted on their platforms and immunizes these applications from liability for such content.[42] Likewise, section 230's "Good Samaritan" provision further protects internet applications from liability for the good faith moderation of "otherwise objectionable" content hosted on their services.[43] While section 230 immunizes platforms from claims alleging they are publishers and claims arising from their good faith efforts, Congress's

38. *See id.* at 45 (describing moderator's role and purpose on online platforms); *see also* Sander, *supra* note 5, at 944 (emphasizing extent to which social media platforms control content). Major online social platforms exert enormous influence over the digital public sphere, "acting as a gateway for information and expression around the world." *See* Sander, *supra* note 5, at 955.

39. *See* Grimmelmann, *supra* note 8, at 45 (describing moderator's influence on online communities' speech and values); *see also* Klonick, *supra* note 13, at 1615 (asserting online platforms' self-regulation reflects free speech norms). While platforms publicly cite a "sense of corporate social responsibility" as the justification for generating rules and systems to moderate speech, the reality is "their economic viability depends on meeting users' speech and community norms." *See* Klonick, *supra* note 13, at 1625 (arguing online platforms' rules and systems reinforce perceived community norms); *see also* Florian Wettstein, *Silence as Complicity: Elements of a Corporate Duty to Speak Out Against the Violation of Human Rights*, 22 BUS. ETHICS Q. 37, 41 (2012) (asserting corporations maintain positive and negative duties). A positive duty is a duty to actively help or assist persons in distress, while a negative duty is a duty to do no harm. *See* Wettstein, *supra* (contending positive duties improve state of affairs, while negative duties avoid making situation worse).

40. *See* Communications Decency Act § 509, 47 U.S.C. § 230 (articulating protected moderation practices for internet applications); *see also* Taddeo & Floridi, *supra* note 32, at 1584 (explaining internet platforms not legally liable for user-generated content). While not legally liable, the CDA encourages internet platforms to monitor and filter online content to the extent they can. *See id.* (emphasizing role internet platforms play in moderation of online content).

41. *See* Communications Decency Act of 1996, Pub. L. No. 104-104, sec. 509, § 230, 110 Stat. 56, 137-39 (codified as amended at 47 U.S.C. § 230) (setting forth CDA); 47 U.S.C. § 230(b)(4) (explaining policy consideration behind section 230); *see also* 47 U.S.C. § 230(b)(5) (emphasizing desire to "deter and punish trafficking in obscenity, stalking, and harassment" on internet).

42. *See* 47 U.S.C. § 230(c)(1)–(2) (shielding applications from liability and distinguishing publisher or speaker of information from users); *see also* Grimmelmann, *supra* note 8, at 103 (describing how section 230 shields platforms from publisher or speaker liability).

43. *See* 47 U.S.C. § 230(c)(2) (describing protection for Good Samaritan blocking and screening of offensive material). Section 230 says in part:

> No provider or user of an interactive computer service shall be held liable on account of any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected . . . .

*Id.* § 230(c)(2)(A).

elimination of the risk of liability for "mismoderation" encourages internet applications to moderate content.[44]

The CDA provides top-layer applications with a liability shield, which in turn allows applications wide discretion to moderate online content.[45] As internet applications become increasingly prevalent fixtures of daily life, these companies will continue to engender a wider set of responsibilities towards society, ranging from economic to moral responsibilities.[46] In regard to content moderation, however, the CDA shields internet platforms from responsibility no matter the inconsistency or ineffectiveness of their moderation practices.[47] Thus, while the societal responsibility of internet platforms increases, laws such as the CDA will continue to emphasize liability protection for moderation practices and

44. *See* Daphne Keller, *Who Do You Sue? State and Platform Hybrid Power over Online Speech* 12 (HOOVER INST., Aegis Series Paper No. 1902, 2019), https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf [https://perma.cc/LQ5Y-2AY8] (noting two specific aspects of section 230's liability shield); Grimmelmann, *supra* note 8, at 103 (discussing moderators' blanket immunity via section 230 protections). So long as moderation practices and actions are done in "good faith," no moderation decision can lead to liability for the platforms performing the moderation. *See* Grimmelmann, *supra* note 8, at 103 (emphasizing policy underlying enactment of section 230); *see also* Bridy, *supra* note 6, at 209 (observing Good Samaritan provision allows good faith discrimination).

45. *See* Bridy, *supra* note 6, at 209 (characterizing moderation policies under existing law to maintain extreme discretion). Existing laws provide internet platforms with broad leeway to engage in content-based discrimination of what constitutes acceptable content. *See id*. at 209-10 (acknowledging platform's discretion comes just short of "blank check" moderation practices with broad immunity); *see also* Daphne Keller, *Internet Platforms: Observations on Speech, Danger, and Money* 10 (HOOVER INST., Aegis Series Paper No. 1807, 2018), https://www.hoover.org/sites/default/files/research/docs/keller_webreadypdf_final.pdf [https://perma.cc/B2RW-L8C5] (recognizing Congress's pragmatic decision to rely on immunities and incentives). By relying on immunities and incentives rather than legal mandates, Congress hoped internet companies would try harder to moderate harmful content once those practices could not expose them to legal liability. *See* Keller, *supra* (emphasizing CDA prioritizes economic development and free expression on internet at cost of imperfect enforcement).

46. *See* Stephen Chen, *Corporate Responsibilities in Internet-Enabled Social Networks*, 90 J. BUS. ETHICS 523, 524 (2009) (describing modern view of business responsibilities). While economic responsibilities remain the foundation of most businesses, the most common view adds required legal responsibilities, expected ethical responsibilities, and desirable philanthropic responsibilities for businesses flourishing today. *Id.* (concluding these responsibilities culminate in overarching moral responsibilities). Social networks generate moral responsibilities through the content their users publish, the socially expected behavior they assume, and the responsible decision making they engage in regarding their platform and moderation practices. *See id.* (establishing moral responsibility based on three criteria: causality, rule-following, and decision making).

47. *See supra* note 45 (recognizing platform's wide latitude of discretion and immunity in their moderation practices); Bridy, *supra* note 6, at 224 (concluding inherently flexible and subjective standard approach to moderation leads to uneven, unpredictable decision making). While social media platforms may adopt detailed content removal policies, it is significantly more difficult to enforce these policies consistently given the scale at which these platforms now operate. *See* Bridy, *supra* note 6, at 222 (arguing for openness and accountability in moderation to counter claims of arbitrariness and political bias); *see also* Giancarlo Frosio & Martin Husovec, *Accountability and Responsibility of Online Intermediaries*, *in* THE OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY 613, 628 (Giancarlo Frosio ed., 2020) (highlighting difficulty to measure or evaluate current moderation practices due to their incredible vagueness).

actually shield platforms from responsibility for inconsistent or ineffective content moderation.[48]

### *3. The Invisible Police Force*

Content moderation is a definitional and indispensable aspect of what internet platforms do.[49] With ever-increasing activity and content online, internet platforms have expanded the use of private algorithmic tools as a means of moderating content.[50] In the context of online content moderation, an algorithmic tool is a computing code that automatically operates underneath internet platforms to implicitly influence the types of possible interactions on a platform, which in turn determines how the platform organizes, promotes, and presents content to users.[51] Algorithmic content moderation depends on "dynamic machine learning" that allows the system, either through human input or its own computing power, to automatically identify patterns and relationships from seemingly unrelated sets of data.[52] Upon identifying and categorizing online content, algorithmic tools use several organizational techniques to strategically shape online users' behavior through the content they see and how the platform presents the content to them.[53] When viewing content online, it is almost impossible to determine whether algorithms or humans moderate the content.[54]

---

48. *See* Bridy, *supra* note 6, at 209-10 (describing platform's discretion to moderate with broad immunity); Sander, *supra* note 5, at 950 (arguing regulatory and liability laws affect how platforms moderate their sites). *But see* Mellinger, *supra* note 3, at 400 (asserting absence of third-party liability has allowed social media industry to flourish). "Imposing platform liability for third-party content and content moderation practices would yield destructive results, including diminished online communities and onerous litigation." Mellinger, *supra* note 3, at 400 (arguing imposing liability would fail to adequately remedy complaints of unfavorable content moderation).

49. *See* Sander, *supra* note 5, at 945 (noting moderation curates user experience).

50. *See* Kemp, *supra* note 2 (emphasizing yearly percentage growth of social media's active users); Frosio & Husovec, *supra* note 47, at 615 (discussing worldwide move towards private, online law enforcement).

51. *See* Sander, *supra* note 5, at 946 (explaining rules of private governance determine platform moderation). While publicly available platform community standards and service agreements document explicit rules, algorithms implicitly enforce these rules by operating out of public view. *See id.* (distinguishing between explicit rules and implicit rules of governance online); *see also* Grimmelmann, *supra* note 8, at 62-63 (explaining how moderators influence norms, shape behavior, and reinforce community expectations).

52. *See* Elkin-Koren & Perel, *supra* note 10, at 674-75 (asserting disclosure of moderation algorithm's coding will not reveal much about its performance); *see also* Frosio, *supra* note 7, at 13 (asserting public lacks technical knowledge or resources to address global online behavior and moderation needs). Even if moderation was put into the hands of governments or international bodies, these institutions would delegate content moderation to algorithmic tools, just as private internet platforms already do. *See* Frosio, *supra* note 7, at 13 (expressing concerns regarding limited accountability with algorithmic tools).

53. *See* Grimmelmann, *supra* note 8, at 58-59 (introducing informational capabilities and organization techniques internet platforms use to influence content). Algorithmic tools use organizational techniques based on the following operations: deletion, editing, annotation, synthesis, filtering, and formatting. *Id.* at 58-59.

54. *See* Chen, *supra* note 46, at 525 (expressing difficulty in separating members of network from structure of network); *see also* Sander, *supra* note 5, at 947 (describing how automatic algorithms and manual human moderators work together in content moderation).

When adopting algorithms as the default moderation system rather than utilizing humans, an internet platform sacrifices quality in favor of cost.[55] An algorithm can cheaply moderate thousands or millions of cases online, whereas a human cannot.[56] While algorithmic moderation is programmed to be rule bound and more inflexible, human moderation generally allows for better results in more difficult moderation decisions due to human attention, responsiveness, and deliberation.[57] In a worst-case scenario, algorithmic moderation could generate thousands of "false positives," negatively impacting users' freedom of expression across an internet platform and creating a chilling effect.[58] Human moderation is not infallible either; humans are susceptible to imprecision, inconsistency, and discriminatory enforcement.[59] In either situation, unreasonable moderation practices threaten to create the appearance of moderation biases, which, regardless of their validity, may trigger "feelings of alienation, frustration, and moral outrage" for those whose content the moderation system erroneously censors.[60]

In assessing algorithms that internet platforms employ, both commentators and consumers have an inherent desire to evaluate the technology to ensure it is acting responsibly.[61] Disclosing the software code that underlines these

55. *See* Grimmelmann, *supra* note 8, at 65 (discussing how automation allows more moderation at lower cost).

56. *See id.* at 64-65 (comparing manageable caseloads and costs of human and automated moderation systems).

57. *See id*. at 65 (maintaining human moderation generally provides better moderation decisions); *see also* Frosio, *supra* note 7, at 24 (arguing inability to program human determinations of fairness into algorithms substantially impacts individual rights).

58. *See* Frosio, *supra* note 7, at 23 (describing present level of technological sophistication and potential for false positives); *see also* Keller, *supra* note 45, at 6 (noting moderation accuracy rates of algorithmic text-matching software). One research study showed that sophisticated algorithms relying on language filters to moderate content showed 70%–80% accuracy rates, meaning the algorithm removed the wrong content roughly once every four or five times. *See* Keller, *supra* note 45, at 6 (depicting difficulties in using algorithms to identify objectionable text).

59. *See* Sander, *supra* note 5, at 956-57 (highlighting human vulnerabilities in moderating online content). Further, there is emerging evidence that human moderators suffer serious psychological harm as a result of their work. *See* Andrew Arsht & Daniel Etcovitch, Commentary, *The Human Cost of Online Content Moderation*, JOLT DIG. (Mar. 2, 2018), https://jolt.law.harvard.edu/digest/the-human-cost-of-online-content-moderation [https://perma.cc/26QU-TL77] (noting impact of moderation on human moderators). In attempting to meet arbitrary numerical quotas, many human moderators filter through graphic, sexual, and violent content. *See id.* (acknowledging PTSD and other mental health issues arising among human moderators); *see also* Casey Newton, *The Trauma Floor*, VERGE (Feb. 25, 2019, 8:00 AM), https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona [https://perma.cc/V278-E7PT] (finding moderators develop PTSD, secondary traumatic stress, and anxiety disorders after exposure to online content). The psychological trauma human moderators experience due to the content they filter causes significant changes in behavior, as well as anxiety, sleep loss, and dissociation. *See* Newton, *supra* (acknowledging traumatic workplace conditions and inadequate mental health safety for human moderators).

60. *See* Sander, *supra* note 5, at 958 (explaining unintended risks and perceptions of discriminatory moderation on certain individuals and groups); *see also* Chen, *supra* note 46, at 531 (explaining expounding effect of harmful practices online).

61. *See* Thompson, *supra* note 34, at 818 (suggesting algorithms' creators try to construct reasonable functioning algorithms). Internet platforms must be able to justify the actions their algorithms are programmed to

algorithms, however, does not reveal much about the actual performance of these systems of content moderation.[62] Rather, transparency would only reveal the underlying values embedded in the code at the time of its creation because an algorithm's dynamic machine learning causes the algorithm's moderation practices to evolve over time.[63] Likewise, internet platforms are not subject to disclosure obligations for their algorithms and are thus free to determine which data to share in accordance with their business interests as well as to prevent interested parties from manipulating their content moderation systems.[64] In juxtaposition, transparency by the platforms themselves would explicitly detail what content the system moderated and why, instructing the public on moderation policies and how they apply in specific instances.[65]

*4. Attention over Protection*

Social media platforms' policies surrounding algorithmic content moderation are not meant to protect users, but rather to monetize users' attention by presenting users with content that interests them.[66] In this custom-tailored model, algorithms organize and present information for specific users to create personalized informational ecosystems, or echo chambers, that isolate individuals from conflicting viewpoints.[67] Through algorithms, social media platforms exercise editorial power by curating, moderating, and prioritizing certain content, while

perform or restricted from enabling. *See id.* at 819 (explaining "justificatory dimension" to algorithm responsibility); *see also* Elkin-Koren & Perel, *supra* note 10, at 674-75 (arguing increased transparency will not necessarily result in accountability of algorithms); Sander, *supra* note 5, at 994-95 (expressing concern over transparency of algorithms and impact of unpredictable moderation).

62. *See* Elkin-Koren & Perel, *supra* note 10, at 675 (discussing transparency in relation to dynamic machine learning and what it would reveal).

63. *See id.* at 674-75 (noting how algorithms' coded values change with actual performance and learning); *see also* Sander, *supra* note 5, at 995 (arguing frequent algorithm alteration complicates meaningful transparency). For example, "Google . . . updates its algorithms hundreds of times per year—and often rely on machine learning techniques, which can lead to a divergence between what programmers believe an algorithm does and how it actually behaves." Sander, *supra* note 5, at 995.

64. *See* Elkin-Koren & Perel, *supra* note 10, at 675-76 (contending business interests and security concerns compel restricting algorithm transparency); *see also* Sander, *supra* note 5, at 995-96 (arguing making algorithms transparent will expose platforms to "manipulation and gaming").

65. *See* Grimmelmann, *supra* note 8, at 65 (emphasizing secret moderation practices hide significant details); *see also* Sander, *supra* note 5, at 995 (describing steps needed to improve algorithmic transparency). Enhancing algorithmic transparency would allow users to understand the critical factors, frequency, and ramifications underlying the use of algorithmic tools. *See* Sander, *supra* note 5, at 995.

66. *See* Sander, *supra* note 5, at 954 (noting platforms incentivized to moderate content to maximize advertising revenue); *see also* Taddeo & Floridi, *supra* note 32, at 1579 (explaining "rich get richer" dynamic of filtered content). Like publishers of other media, search engines and social media platforms filter information according to users' tastes and preferences. *See* Taddeo & Floridi, *supra* note 32, at 1579.

67. *See* Taddeo & Floridi, *supra* note 32, at 1582 (arguing custom-tailored information leads to individuals isolated in "informational bubbles"). Custom-tailored information for specific users undermines the possibility of exposing users to different experiences and viewpoints. *See id*. at 1581-82 (criticizing approaches relying on users' preferences and market dynamics to shape information access and communication).

diverting attention away from other kinds of content.[68] Platforms such as Facebook and Twitter purposely design timelines and newsfeeds to keep users logged in and engaged with content for as long as possible.[69] In this context, content moderation culminates in one primary objective: monetizing the individual user's attention.[70] Economic pressure incentivizes internet platforms to moderate content to maximize user engagement, data surveillance, and targeted advertising in order to generate revenue and growth.[71]

As private, profit-maximizing entities, social media is based on a broad notion of freedom of expression, and a social media platform's interest derives from the content they are expected to moderate.[72] These notions create an idealized perception of social media as a marketplace of ideas, where everyone has the opportunity to compete to convince the general public to accept their ideas.[73] This idealized marketplace, however, cannot exist within the framework of current governance, algorithmic involvement, and individualized platform practices surrounding online content moderation.[74] In today's digital public sphere, platforms shape what communication is possible, what content is permissible, and what content is visible to the individual user.[75] With a large amount of moderation power in the hands of a small number of "private mega platforms," internet

68. *See* Elkin-Koren & Perel, *supra* note 10, at 671 (contending moderation immunity intended to promote free flow of information rather than create online editors); *see also* Sander, *supra* note 5, at 954 (contending many internet platforms offer algorithmic personalization to keep users glued to their sites); Bridy, *supra* note 6, at 217 (describing how social media platforms avoid exposing users to content they might find uninteresting).

69. *See* Bridy, *supra* note 6, at 217 (explaining platforms inundate users with content based on aggregated data about their personal interests).

70. *See* Sander, *supra* note 5, at 952 (noting largest online platforms' business model involves "sale of human attention"). Major social networks generate revenue by monetizing the attention of their users. *See id.* at 953 (describing "attentional scarcity" within digital public sphere).

71. *See id.* at 954 (asserting demands of advertising industry drives content moderation policy); *see also* Bridy, *supra* note 6, at 217 (emphasizing effects of algorithmic personalization). Social media platforms fill each user's timeline or newsfeed with content based on a user's personal preferences that the platform surreptitiously extracts and aggregates. *See* Bridy, *supra* note 6, at 217 (highlighting how this benefits advertisers by allowing them to reach "more particularized" cohorts of consumers).

72. *See* Elkin-Koren & Perel, *supra* note 10, at 671 (arguing content-sharing platforms thrive on democratic notions of freedom and openness). These notions of freedom and openness allow diverse users to potentially share their ideas on a global scale. *See id.*

73. *See* Bridy, *supra* note 6, at 216 (contending marketplace of ideas works differently in age of social media compared to traditional media); *see also* Frosio, *supra* note 7, at 2 (asserting most creative expression takes place on internet today).

74. *See* Frosio, *supra* note 7, at 32-33 (emphasizing current interested parties trying to coerce online platforms to implement more effective self-regulation). Governments and interested third parties increasingly push internet platforms to adopt more effective and transparent moderation policies, which only further increases the power private actors have on the internet. *See id.* (arguing more power in private actors jeopardizes fundamental individual rights). "[T]ransferring regulation and adjudication of Internet rights to private actors does jeopardize fundamental rights . . . by limiting access to information, causing chilling effects, or curbing due process." *Id.* at 33.

75. *See* Sander, *supra* note 5, at 944 (arguing social media platforms effectively dismantle marketplace of ideas online). By establishing private content moderation policies and practices, a small number of internet platforms control how people exchange ideas and information online. *See id.* (comparing private online platforms to "governors of online speech").

platforms essentially remain unaccountable for their moderation practices, resulting in these private platforms maintaining the ability to shape public discourse and silence lawful speech with impunity.[76]

### *B. The First Amendment and the Idea of the "Public Square"*

#### *1. The Public Forum*

A traditional "public forum" refers to a place where a speaker accesses public property or private property dedicated to public use for the purpose of speaking.[77] For First Amendment purposes, the government may not, through its exercise of viewpoint discrimination via censorship or exclusion, infringe upon the speech of a speaker in a space deemed a public forum.[78] As a result of the ever-growing privatization of many areas of public life, the traditional notion of what constitutes a public forum has become increasingly difficult to determine.[79] The rapid development of popular communication technologies, particularly social media platforms, has forced lawmakers and legal scholars to reexamine the traditional public forum doctrine.[80]

In 1946, the United States Supreme Court in *Marsh v. Alabama*[81] addressed the applicability of First Amendment protections for public speakers on private property.[82] In *Marsh*, a public speaker was arrested for distributing religious literature in a privately owned town against the wishes of the town's owner.[83] Gulf Shipbuilding Corporation owned the town, but otherwise the town maintained all the characteristics of any other municipality, including residential areas, public streets, and a shopping center that the general public accessed and

76. *See* Elkin-Koren & Perel, *supra* note 10, at 670 (commenting on blurred distinction between private interests and public responsibilities). Private entities have no duty to respect free speech or other fundamental rights, and that leads to social media platforms diminishing their sites as a true marketplace for ideas. *See id.* at 673 (highlighting internet platforms' potential invasion into individual civil rights).

77. *See* Wiener, *supra* note 12, at 223 (defining traditional public forum for purposes of First Amendment). A public forum traditionally includes places associated with "open speech, debate, and political speech." *Id.* at 224 (acknowledging public afforded maximum protection from speech restrictions and exclusions in public forums).

78. *See id.* at 223-24 (emphasizing importance of public forum in relation to free speech and expression).

79. *See* Peters, *supra* note 12, at 994 (noting increased privatization blurs distinction between public and private under public forum doctrine). For example, prisons, hospitals, and schools are areas of public life increasingly subject to privatization. *See id.* (highlighting complexity in applying public forum doctrine today where traditional notions evolved).

80. *See* Wiener, *supra* note 12, at 225 (addressing need for change or adaption of public forum paradigm). For example, as social media platforms consume more of the communications infrastructure, it is becoming increasingly more difficult to differentiate between government speech and private speech. *See id.* (noticing pre-internet analogies for forums and media untenable for current disputes over social media); *see also* Peters, *supra* note 12, at 994-95, 999-1000 (noting challenge of applying traditional doctrines to social media platforms). "Platforms like YouTube, Facebook, and Twitter defy easy classification in this area." Peters, *supra* note 12, at 1000 (recognizing classification of virtual public forums difficult because of their private ownership).

81. 326 U.S. 501 (1946).

82. *See id.* at 502 (summarizing relevant constitutional issue).

83. *See id.* at 502-03.

used freely.[84] In reversing the public speaker's conviction, the Court held that the more an owner opens up his property for general public use, the more the constitutional rights of others constrain an owner's rights.[85] Thus, the state may regulate a property built and operated primarily for the benefit of the public—one whose operation essentially performs a public function.[86] The Court recognized that the general public has an interest in the property functioning so that "channels of communication remain free," regardless of whether a public or private entity owns the property.[87]

In 1968, the Supreme Court in *Amalgamated Food Employees Union Local 590 v. Logan Valley Plaza, Inc.*[88] reinforced its holding in *Marsh* by continuing to expand an individual's First Amendment rights on private property dedicated to public use.[89] In *Logan Valley*, the Court held that a privately owned shopping mall was the functional equivalent of the town in *Marsh*.[90] Similar to the town in *Marsh*, the shopping mall "serve[d] as the community business block 'and [wa]s freely accessible and open to the people in the area and those passing through.'"[91] The Court held that the First Amendment protected peaceful picketing in a location generally open to the public.[92]

Eight years later, in *Hudgens v. NLRB*,[93] however, the Court expressly overruled its holding in *Logan Valley*.[94] In reexamining a shopping mall as a public forum—extensively relying on *Lloyd Corp. v. Tanner*—the Court determined that private properties do not inherently perform a public function by merely remaining open for public use.[95] Rather, owners of private property must assume governmental functions or exercise a form of municipal power—like the

84. *See id.* (describing characteristics of Chickasaw, Alabama).

85. *See Marsh*, 326 U.S. at 506 (rejecting coextension of corporation's rights with homeowner's rights where property dedicated to public use). For example, "owners of privately held bridges, ferries, turnpikes and railroads may not operate them as freely as a farmer does his farm." *Id.*

86. *See id.* (recognizing ownership does not equal "absolute dominion").

87. *See* Marsh v. Alabama, 326 U.S. 501, 507 (1946) (explaining private or public ownership of publicly used property does not diminish citizens' constitutional protections).

88. 391 U.S. 308 (1968), *overruled by* Hudgens v. NLRB, 424 U.S. 507 (1976).

89. *See id.* at 318-19 (comparing *Logan Valley* to *Marsh*).

90. *See id.* at 318 (emphasizing similarities between situations in *Marsh* and *Logan Valley*). "The shopping center premises are open to the public to the same extent as the commercial center of a normal town." *Id.* at 319 (highlighting mall's public function resulting from public access to property).

91. *Id.* (quoting *Marsh*, 326 U.S. at 508). When private property performs a public function, public speakers exercising their First Amendment rights should be protected on said property. *See id.* at 319-20 (emphasizing purpose and use of property critical in distinguishing between public and private spaces). *But see id.* at 320 (holding property owners can reasonably regulate exercise of First Amendment on their land).

92. *See Logan Valley*, 391 U.S. at 313, 319-20 (noting free expression under First Amendment includes speech and conduct).

93. 424 U.S. 507 (1976).

94. *See id.* at 518 (noting conflicting rationales of *Logan Valley* and *Lloyd Corp. v. Tanner*); *see also* Lloyd Corp. v. Tanner, 407 U.S. 551, 568-69 (1972) (disagreeing with rationale of *Logan Valley* without explicitly overruling it).

95. *See* Hudgens v. NLRB, 424 U.S. at 519 (quoting *Tanner*, 407 U.S. at 568-69) (emphasizing significance between constitutional restraints on government action versus private action on property).

privately owned town in *Marsh*.[96] Thus, without the performance of a public function, private property open to the general public need not accommodate individuals' First Amendment free speech rights.[97]

*2. Media Forums—Radio, Television, and the Early Internet*

Although the right to expressive conduct in private online forums remains unsettled, the Court has upheld First Amendment protection of speech on platforms such as radio and television.[98] In *Red Lion Broadcasting Co. v. FCC*, the Supreme Court of the United States upheld the regulation of radio broadcasting, reasoning that the scarcity of broadcast frequencies required the government to allocate those frequencies to ensure individual access.[99] Under the First Amendment, the right of free speech for a broadcast entity or any other individual "does not embrace a right to snuff out the free speech of others."[100] Specifically, the Court observed that the FCC already required broadcasters to give adequate coverage to public issues by accurately and fairly presenting coverage of opposing views.[101] Similarly, in *Turner Broadcasting System, Inc. v. FCC*, the Court upheld the regulation of cable television by emphasizing the power of broadcast companies as "gatekeepers" over all television programming, particularly in their ability to prevent subscribers from obtaining access to programs and speakers they choose to exclude.[102] The Court reaffirmed that the First Amendment does

96. *See id.* (quoting *Tanner*, 407 U.S. at 569) (comparing town in *Marsh* to extension of state exercising "semi-official municipal functions"). The owner of Chickasaw was a proxy for the government and exercised broad municipal powers. *Id.* (quoting *Tanner*, 407 U.S. at 569) (distinguishing private property for public use from private property performing public function).

97. *See id.* at 520-21 (holding no right to picket in privately owned mall). The Court reiterated that the constitutional guarantee of free speech only applies to an abridgment by the government. *Id.* at 513.

98. *See* Klonick, *supra* note 13, at 1611-12 (summarizing Court's holdings in cases involving radio and television regulation).

99. *See* 395 U.S. 367, 400-01 (1969) (upholding governmental regulation of radio broadcasting for First Amendment purposes). In regulating radio broadcasting, the Court previously recognized that the FCC neither oversteps its authority nor infringes upon the First Amendment in "interesting itself in [the] general program format and the kinds of programs broadcast by licensees." *See id.* at 395 (citing NBC v. United States, 319 U.S. 190, 226-27 (1943)) (emphasizing regulation of broadcast content via licensing scheme did not violate First Amendment).

100. *See id.* at 387 (citing Associated Press v. United States, 326 U.S. 1, 20 (1945)) (underscoring need for government regulation).

101. *See id.* at 377 (articulating emergence of fairness doctrine where each side of public issue must receive fair coverage). Without the fairness doctrine, a licensee could devote their station entirely to the support of one ideology or group of candidates to the exclusion of all others. *See id.* at 382-83 (arguing fairness doctrine operates in public interest by preventing biased broadcasting on important issues).

102. *See* 512 U.S. 622, 656 (1994) (explaining how cable television companies own essential pathway for cable speech). "A cable operator, unlike speakers in other media, can thus silence the voice of competing speakers with a mere flick of the switch." *Id.*

not limit the government from ensuring that private interests do not restrict "the free flow of information and ideas."[103]

Prior to the CDA, federal and state courts addressed an internet platform's liability for user-generated content.[104] In *Cubby, Inc. v. CompuServe, Inc.*,[105] CompuServe was an electronic library site that allowed subscribers access to over 150 special interest forums, most of which independent communications companies moderated.[106] The Southern District of New York found CompuServe not liable for libel on one of its forums because it did not actively review or control the forum's content.[107] The court reasoned that CompuServe was a distributor rather than a publisher, and a distributor could not be held liable for libel "if it neither knew nor had reason to know" of the allegedly defamatory statements of its user-generated content.[108] In contrast, in *Stratton Oakmont, Inc. v. Prodigy Services Co.*,[109] the New York Supreme Court found Prodigy to be a publisher because it publicly stated it controlled the content on its services, as well as utilized technology and manpower to edit content.[110] Prodigy was an online financial bulletin board where users could post and promote finance-related matters.[111] The court reasoned Prodigy was a publisher because it actively deleted forum posts, moderated posts based on its guidelines, and used automatic software to scan and remove posts, all of which showed a conscious choice to gain editorial control of its user-generated content.[112]

As different types of media forums grew in popularity, however, the Supreme Court of the United States determined that "[e]ach medium of expression, of course, must be assessed for First Amendment purposes by standards suited to it, for each may present its own problems."[113] Following the CDA's passage, the Supreme Court in *Reno v. ACLU* addressed whether certain CDA provisions

---

103. *See id.* at 657 (citing *Associated Press*, 326 U.S. at 20) (equating cable companies' television programs to First Amendment-protected speakers). If necessary, the government may take steps to physically control "critical pathway[s] of communication" to protect free speech. *See id.*

104. *See* Mellinger, *supra* note 3, at 392 (asserting several online moderation cases established precedent for online moderation practices).

105. 776 F. Supp. 135 (S.D.N.Y. 1991).

106. *See id.* at 137 (describing CompuServe's business structure).

107. *See id.* at 140-41 (equating CompuServe's lack of editorial control to public libraries, bookstores, or newsstands). "First Amendment guarantees have long been recognized as protecting distributors of publications . . . . [T]he national distributor of hundreds of periodicals has no duty to monitor each issue of every periodical it distributes. Such a rule would be an impermissible burden on the First Amendment." *Id.* at 140 (quoting Lerman v. Flynt Distrib. Co., 745 F.2d 123, 139 (2d Cir. 1984)).

108. *See id.* at 141 (reasoning mere distributors of user-generated content not liable for substance of said content).

109. No. 031063/94, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995).

110. *See id.* at *4 (distinguishing online publishers from online distributors).

111. *See id.* at *1 (describing Prodigy's online service).

112. *See id.* at *4-5 (explaining how Prodigy's editorial control may have chilling effect on freedom of communication). "Prodigy's 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." *Id.* at *5.

113. *See* Se. Promotions, Ltd. v. Conrad, 420 U.S. 546, 557 (1975) (citing Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 503 (1952)).

violated the First Amendment for being overly vague in their categorical prohibitions of internet communications.[114] While the Court saved various portions of the CDA, it held that the CDA's indecency and offensiveness provisions violated the First Amendment because the content-based restrictions of free speech were not narrowly tailored.[115] The Court stressed that regardless of important policy goals, the vagueness of content-based speech regulations raise special First Amendment concerns because of an "obvious chilling effect on free speech" as well as a "risk of discriminatory enforcement."[116]

### *3. Laying the Foundation*—Packingham *and the Public Square*

The Supreme Court revisited First Amendment free speech protections on the internet in the 2017 decision of *Packingham v. North Carolina*.[117] In 2008, a North Carolina statute made it a felony for registered sex offenders to gain access to certain social networking websites, including social media platforms like Facebook and Twitter.[118] Writing for the majority, Justice Kennedy observed that the statute created an "unprecedented prohibition in the scope of First Amendment speech it burden[ed]."[119] In declaring the statute unconstitutional, the Court held that the exclusion of a subset of internet users from completely accessing social media prohibited the legitimate exercise of First Amendment rights, as social media is a principal source for discovering current events, finding employment, and exploring ideas.[120]

The *Packingham* Court reiterated that allowing all persons access to places where they can speak and listen is a fundamental principle of the First Amendment.[121] In his opinion, Justice Kennedy equated access to and the usage of social media platforms to "speaking and listening in the modern public square," emphasizing the ease with which individual speakers on these platforms can reach an audience beyond their physical community.[122] While prior decades provided no clear consensus regarding the most important places for the exchange of ideas, the Court unequivocally identified the internet and social media as those places today.[123] Justice Kennedy cautioned, however, that the power and

114. *See* 521 U.S. 844, 870 (1997) (explaining CDA's categorical prohibitions too broad).

115. *See id.* at 849, 879 (holding CDA's scope unprecedented and poorly defined).

116. *See id.* at 871-72 (emphasizing special concerns regarding CDA's ambiguity).

117. *See* 137 S. Ct. 1730, 1733 (2017) (addressing whether North Carolina statute violates Free Speech Clause of First Amendment).

118. *See id.* (describing North Carolina statute). The statute prevented registered sex offenders from accessing websites that permit children to become members, or create or maintain webpages with their personal information. *See* N.C. GEN. STAT. § 14–202.5 (2021), *invalidated by Packingham*, 137 S. Ct. 1730.

119. *See Packingham*, 137 S. Ct. at 1737 (describing impact of North Carolina statute).

120. *See id.* at 1737 (noting social media allows users to gain access to information and communicate freely).

121. *See id.* at 1735 (discussing First Amendment's free speech protections and right to speak).

122. *See id.* at 1737 (illustrating ideas and viewpoints of users on social media platforms resonate farther than physical platforms).

123. *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017) (emphasizing importance of cyberspace for free expression). "[S]ocial media users employ these websites to engage in a wide array of protected

trajectory of the internet is "so new . . . and so far reaching that courts must be conscious that what they say today might be obsolete tomorrow."[124]

In his concurrence, Justice Alito voiced concerns about the majority's relaxed comparison of social media platforms to traditional public forums.[125] While agreeing that the Court should exercise restraint in applying free speech precedents to the internet, Justice Alito criticized the Court for not heeding its own warning of caution, believing the majority opinion's dicta proceeded in an undisciplined fashion.[126] Nevertheless, the United States Court of Appeals for the Ninth Circuit did not share Justice Alito's trepidation.[127] In *Prager University v. Google LLC*, the Ninth Circuit held that YouTube, an online service that hosts user-generated videos, remained a private forum and not a public forum subject to judicial scrutiny under the First Amendment.[128] Despite YouTube's ubiquity and self-representation as a virtual public square, the court reasoned that a platform merely hosting speech and inviting public discourse does not transform that platform into a state actor performing a public function.[129] Nonetheless, the Ninth Circuit recognized that arguments for First Amendment protections of online speech are significant to future policy discussions regarding the internet.[130]

## III. ANALYSIS

### *A. The CDA Problem*

In the age of the modern internet, social media platforms maintain an unprecedented ability to regulate individual expression unlike any other communication medium in human history.[131] In the spirit of facilitating collaboration, enhancing conversation, and preventing abuse, content moderation is important for creating

First Amendment activity on topics 'as diverse as human thought.'" *Id.* at 1735-36 (quoting Reno v. ACLU, 521 U.S. 844, 870 (1997)).

124. *See id.* at 1736 (equating Cyber Age to historic revolution of unknown dimensions and vast potential).

125. *See id.* at 1743 (Alito, J., concurring) (chastising Court for "loose rhetoric"). In criticizing the majority for declining to explain what the comparison meant regarding free speech law, Justice Alito found the Court's readiness to equate the internet with traditional public spaces particularly troubling. *Id.*

126. *See id.* at 1738, 1744 (expressing unease for implications of Court's unnecessary rhetoric).

127. *Prager Univ. v. Google LLC*, 951 F.3d 991, 995 (9th Cir. 2020).

128. *See id.* (noting online platform merely hosting others' speech on public issues does not change analysis).

129. *See id.* at 997 (concluding any other outcome would significantly weaken distinction between governmental and private entities). "YouTube does not perform a public function by inviting public discourse on its property." *Id.* at 998.

130. *See id.* at 999 (discussing arguments of unregulated big-tech companies versus government regulation of internet platforms).

131. *See* Keller, *supra* note 44, at 1 (characterizing large internet companies' "unprecedented technological capacity to regulate individual expression"). Social media companies can monitor every word or symbol a user shares and censor out anything they do not like. *See id.* (recognizing unprecedented power large technology companies wield).

healthy online communities and facilitating communication.[132] Without any content moderation, the value of communication in online communities would diminish, and communication channels would be subject to harmful, harassing, or even graphic content.[133] While this simple summation portrays an optimistic need for content moderation, it is critical to remember social media platforms do not moderate content to promote safe freedom of expression amongst users, but rather to monetize users' attention for profit maximization.[134] Under their current business model, social media platforms use algorithms to intentionally capture users' attention by curating, moderating, and prioritizing certain content from others, which create personalized echo chambers that isolate individuals.[135]

In the United States, the CDA protects platforms from liability for user-generated content as well as for the "good faith" moderation of such content.[136] While the CDA encourages platforms to maintain healthy online communities by eliminating liability for moderation, the CDA does not account for the economic reality of why platforms moderate content.[137] Under the current system, platforms use the CDA to maintain wide discretion in the moderation of online content without the risk of liability for potential mismoderation.[138] The CDA's

132. *See* Grimmelmann, *supra* note 8, at 47 (defining moderation and its purpose); *see also* Sander, *supra* note 5, at 945 (describing necessity of content moderation for online platforms); Keller, *supra* note 45, at 9 (asserting laws holding platforms liable for damages would doom many online businesses and hinder innovation).

133. *See* Grimmelmann, *supra* note 8, at 45 (asserting moderation key factor in creating successful, healthy online communities); Sander, *supra* note 5, at 945-46 (explaining how platform practices create better experience by curating, organizing, and moderating content). Through community terms of service, platforms expressly restrict content such as "hate speech, graphic or violent content, sexual content, harassment, copyright, and illegal activity." Sander, *supra* note 5, at 946 (describing need for some degree of moderation).

134. *See supra* notes 68-70 and accompanying text (noting largest online platforms' profit maximization involves sale of human attention).

135. *See* Sander, *supra* note 5, at 953-54 (contending demands of advertising industry and platforms' business models drive content moderation policy); Taddeo & Floridi, *supra* note 32, at 1582 (arguing custom-tailored information isolates individuals into "informational bubbles"). A by-product of algorithmic personalization is the "promotion of emotionally charged, extreme, and inflammatory content." *See* Sander, *supra* note 5, at 954 (describing how curation, moderation, and prioritization of information culminates over extended period of time). Custom-tailored information for specific users undermines the possibility of exposure to sources, opinions, and information that may support or convey differing viewpoints. *See* Taddeo & Floridi, *supra* note 32, at 1581-82.

136. *See supra* notes 42-43 and accompanying text (describing internet platforms' two-pronged liability shield); *see also* Keller, *supra* note 45, at 9 (asserting moderation's simple purpose to take bad internet content down and keep good content up).

137. *See* Keller, *supra* note 44, at 12 (emphasizing two significant aspects of section 230's liability shield for platforms moderating content); Sander, *supra* note 5, at 954 (asserting many internet platforms use algorithmic personalization to maintain users' attention); Bridy, *supra* note 6, at 217 (emphasizing how platforms' policies extract and aggregate individuals' information to capture their attention).

138. *See* Bridy, *supra* note 6, at 208-09 (noting existing CDA protections allow wide discretion in platforms' moderation policies). Section 230's existing protection from liability for good faith efforts provides internet platforms with wide-ranging immunity to engage in content-based discrimination predicated on their choice of what constitutes acceptable content. *See id.* at 209-10 (acknowledging broad discretion does not equate to absolute discretion); *see also* Grimmelmann, *supra* note 8, at 103 (discussing moderators' blanket immunity via section 230 protections). Under the current system, no improper or questionable moderation decision leaves a platform liable for its actions. *See* Grimmelmann, *supra* note 8, at 103 (asserting policy "encourages moderation by removing liability for mismoderation").

purpose of protecting online communities' ability to flourish without obscene material clashes with the tendency of platforms to use the CDA as a liability shield for their overall moderation practices.[139] While necessary to maintain healthy online communities, the CDA has problematically morphed into a mechanism for large internet platforms to restrict speech under the guise of protecting their users from material the platforms deem harmful.[140]

### *B. The Public Forum Problem*

The First Amendment only protects individuals or private entities from government restrictions on free speech.[141] In the United States, arguments equating social media platforms to public forums for the purpose of compelling platforms to carry users' speech have been unsuccessful.[142] In *Marsh*, the Supreme Court articulated the public forum doctrine, which declared a property built and operated primarily for the benefit of the public was subject to constitutional safeguards.[143] In subsequent cases, however, the Court quickly clarified that this doctrine only includes private enterprises performing ordinary governmental functions or exercising normal municipal powers.[144] Notwithstanding this clear

139. *See* Communications Decency Act § 509, 47 U.S.C. § 230(b)(5) (emphasizing desire to deter and punish obscenity); *id.* § 230(c)(2)(A) (limiting internet platforms' liability for moderation of obscene or "otherwise objectionable" content); *see also* Bridy, *supra* note 6, at 208 (describing discrepancy between intent of statute and its ultimate effect). The term "otherwise objectionable" content within section 230 is incredibly broad and should not give platforms absolute immunity in their moderation practices. *See* Bridy, *supra* note 6, at 210 (noting how section 230 should operate); *see also* Grimmelmann, *supra* note 8, at 103 (equating section 230 protections to blanket immunity for moderation by platforms).

140. *See* Grimmelmann, *supra* note 8, at 47 (explaining healthy online communities generate cooperation and minimize abuse amongst their members); *see also* Day, *supra* note 6, at 1318 (noting social media platforms' market power and control may impede free trade of ideas). Despite the intent and perceived benefits behind speech regulation, a platform's moderation policies may unintentionally constrain meritorious and innocent content. *See* Day, *supra* note 6, at 1326 (expressing concern over banning or restricting speech containing ideas in tension with mainstream thought); *see also* Sander, *supra* note 5, at 958 (describing risk and lack of transparency for erroneous or discriminatory moderation of content).

141. *See* Wiener, *supra* note 12, at 221 (noting constitutional distinction between private action and state action).

142. *See* Keller, *supra* note 44, at 11 (summarizing arguments equating platforms to public forums have failed since 1990s). In the United States, there is a widely held notion that large internet platforms must carry all lawful speech, but no claim asserting such an argument has succeeded. *See id.* (describing how intermediate liability statutes and First Amendment jurisprudence makes such claims fruitless).

143. *See* Marsh v. Alabama, 326 U.S. 501, 506 (1946) (explaining how properties dedicated to public use embody "public function"). The Court asserted that it does not matter whether a public or private entity owns the property, because the public has an identical interest in constitutional safeguards protecting community discourse. *See id.* at 507 (asserting type of ownership insignificant because property maintains public function).

144. *See* Lloyd Corp. v. Tanner, 407 U.S. 551, 569 (1972) (distinguishing private property for public use from private property performing public function); Hudgens v. NLRB, 424 U.S. 507, 519 (1976) (quoting *Lloyd*, 407 U.S. at 569) (comparing private property dedicated to public use and company-owned town in *Marsh*).

limitation, the increasing privatization of public life blurs the distinction between private property and public function.[145]

Social media platforms, as new virtual public forums, embody this blurred distinction by encompassing both public and private characteristics.[146] Despite this ambiguity, these new forums for expression ultimately do not embody a public function that the public forum doctrine protects.[147] In *Packingham*, Justice Kennedy's broad rhetoric made access to private online platforms a First Amendment right using the analogy of "the modern public square" to question how far "access" extends.[148] While *Packingham* raised questions about the future of social media, the Court never explicitly addressed the public forum doctrine and its relation to cyberspace.[149] Thus, without the performance of a public function, platforms simply remain private property open for public use.[150]

### *C. The Medium Regulation Problem*

In the past, the Court has held regulation of speech on media such as radio and television is consistent with the First Amendment.[151] In *Red Lion* and *Turner Broadcasting*, the Supreme Court used the First Amendment—paired with the inherent broadcasting limitation each medium possessed—to construct a public right to retain suitable access to various information and ideas on radio and television.[152] In dealing with online speech prior to the passage of the CDA, federal and state courts used the First Amendment to distinguish internet distributors and

145. *See* Peters, *supra* note 12, at 994 (noting complexity when private and public characteristics overlap); *see also Marsh*, 326 U.S. at 506 (explaining how owners of private property dedicated to public use concede certain rights to public).

146. *See* Peters, *supra* note 12, at 1000 (analogizing social media platforms to public forums).

147. *See id.* at 999-1000 (highlighting social media platforms nonexistent connection to state actors). As of now, the public forum doctrine does not permit the First Amendment's application to social media platforms. *See id.* at 1018.

148. *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017) (noting total foreclosure from access unconstitutional while leaving open questions about constitutionality of limited restrictions); *see also* Klonick, *supra* note 13, at 1611 (describing how sweeping language in *Packingham* leaves many questions about social media platforms unanswered). Regarding social media platforms, the *Packingham* decision does not answer how robust access must be, when an internet platform abridges a First Amendment right, or whether platforms perform "quasi-municipal" functions. *See* Klonick, *supra* note 13, at 1611.

149. *See Packingham*, 137 S. Ct. at 1743 (Alito, J., concurring) (criticizing Court's potentially unclear rhetoric regarding new meanings of forums and cyberspace within First Amendment).

150. *See* Prager Univ. v. Google LLC, 951 F.3d 991, 995 (9th Cir. 2020) (holding platforms remain private forums and not public forums when merely hosting others' speech).

151. *See* Klonick, *supra* note 13, at 1611 (summarizing speech regulation cases involving radio and television).

152. *See* Red Lion Broad. Co. v. FCC, 395 U.S. 367, 400-01 (1969) (asserting government regulation of radio broadcasting and its content does not violate First Amendment); Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 657 (1994) (citing Associated Press v. United States, 326 U.S. 1, 20 (1945)) (emphasizing government may regulate cable television without violating First Amendment); Klonick, *supra* note 13, at 1611-12 (analyzing holdings of *Red Lion* and *Turner Broadcasting*). In justifying the regulation of radio and television, the Court cited elements such as the media's invasive nature, prior history of regulation, and scarcity of frequencies. *See* Klonick, *supra* note 13 at 1612.

internet publishers to hold platforms accountable for their editorial control of user-generated speech.[153] Following the passage of the CDA, the Court in *Reno* held that certain provisions of the CDA regulating the internet violated the First Amendment for being unconstitutional in scope.[154] The Court determined that elements justifying the regulation of radio and television were not present on the early internet and explicitly restricted the regulation of the internet.[155]

With the growth of the internet in size and usage, any significant regulation would be monumental compared to the regulation of radio and television, especially since it would regulate the speech of individual users.[156] Internet platforms have contested regulation by representing themselves as First Amendment speakers, while simultaneously claiming regulatory immunity by labeling themselves as mere distributors of third-party speech.[157] In order for a regulation to overcome First Amendment strict scrutiny, the restriction must be "narrowly tailored and advance a compelling government interest."[158] In *Packingham*, the Supreme Court reexamined the internet and considered general access to social media, as opposed to specific access to parts of an internet platform.[159] Although Justice Kennedy's opinion initiated the discussion as to the essential nature of the internet, a more explicit declaration is necessary to overcome internet platforms' virtually exclusive control over First Amendment rights in the digital sphere.[160] Imprecise comparisons relating internet platforms to the modern public square found within judicial opinions are not sufficient to protect users and consumers on internet platforms.[161]

153. *See* Cubby, Inc. v. CompuServe, Inc., 776 F. Supp. 135, 140-41 (S.D.N.Y. 1991) (using First Amendment to protect speech of mere distributors of user-generated content); Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 031063/94, 1995 WL 323710, at *5 (N.Y. Sup. Ct. May 24, 1995) (acknowledging conscious choice to make decisions regarding content constitutes editorial control).

154. *See* Reno v. ACLU, 521 U.S. 844, 877, 879 (1997) (noting CDA's coverage wholly unprecedented).

155. *See id.* at 868-69 (emphasizing differences between radio, television, and internet); *see also* Klonick, *supra* note 13, at 1612 (asserting *Reno* exempted internet from doctrine established in *Red Lion* and *Turner Broadcasting*).

156. *See* Keller, *supra* note 44, at 14 (emphasizing difference in number of actors affected by regulation of traditional media and internet platforms).

157. *See* Bridy, *supra* note 6, at 197 (framing internet platforms' desire to remain liability free).

158. *See* Day, *supra* note 6, at 1325 (explaining First Amendment's strict scrutiny standard).

159. *See* Wiener, *supra* note 12, at 239 (analogizing Facebook and Twitter to modern public square and user accounts to representations of speakers). Access to the entirety of a social media platform is extremely different from access to an individual's personal page. *See id.* (discussing different forms of access exist within one social media platform).

160. *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017) (discussing First Amendment's free speech protections and internet platforms); Klonick, *supra* note 13, at 1613 (arguing Court's rhetoric equating internet platforms to forums threatens private companies claiming First Amendment rights).

161. *See Packingham*, 137 S. Ct. at 1743-44 (Alito, J., concurring) (criticizing Court for loose rhetoric). "The Court is unable to resist musings that seem to equate the entirety of the internet with public streets and parks." *Id.* at 1738.

### *D. What Can Be Done?*

The Supreme Court stressed that "[e]ach medium of expression . . . must be assessed for First Amendment purposes by standards suited to it."[162] With the growing pervasiveness of online speech, society increasingly attributes the role of information gatekeeper to internet platforms.[163] In particular, social media platforms' control of the digital sphere allows them to act as gatekeepers of public discourse, generating the expectation that these platforms will act with awareness and respect towards social standards.[164] With these realities and expectations in mind, online communities should not want private companies acting as their own moral arbiters or without the collaboration of other institutions creating online speech standards.[165] As private companies, social media platforms presently enjoy similar authority and power to that of a high court or administrative agency, acting with nearly undisputed finality in determining what content to host and moderate.[166] While First Amendment limitations significantly constrain actual judges and government agencies, internet platforms use First Amendment jurisprudence, the CDA, and the lack of American judicial precedent to remain free from liability and oversight for their speech moderation.[167]

---

162. Se. Promotions, Ltd. v. Conrad, 420 U.S. 546, 557 (1975) (citing Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 503 (1952)).

163. *See* Taddeo & Floridi, *supra* note 32, at 1582 (acknowledging responsibility academic writers, politicians, and society at large willing to give to online platforms). A platform is an information gatekeeper if that platform "controls access to information, and acts as an inhibitor by limiting access to or restricting the scope of information," ultimately acting as a facilitator of information. *See id.* at 1583 (stating platforms have central role in management of information).

164. *See* Keller, *supra* note 44, at 16 (comparing internet platforms with high market share to de facto gatekeepers of public discourse); Taddeo & Floridi, *supra* note 32, at 1582 (attributing certain expectations and idealistic principles to internet platforms performance in hosting content).

165. *See* Wettstein, *supra* note 39, at 54 (arguing for institutional safeguards or outside input on decisions by private entities effecting global community).

166. *See* Elkin-Koren & Perel, *supra* note 10, at 674 (asserting social media platforms operate under private governance allowing them to determine legitimacy of content). Social media platforms' policies regarding moderation or censorship thus remain unconstrained by constitutional limitations with respect to free speech rights. *See id.* at 673-74 (describing functions of private internet platforms).

167. *See* Wiener, *supra* note 12, at 221, 242 (asserting First Amendment only protects against government restrictions on free speech); Elkin-Koren & Perel, *supra* note 10, at 674 (acknowledging private companies not subject to constitutional limitations or duty to protect fundamental rights); Peters, *supra* note 12, at 992 (arguing First Amendment does not constrain nongovernmental entities, including internet platforms); Communications Decency Act § 509, 47 U.S.C. § 230(c)(1) (shielding internet platforms from liability for any information or content they host for their users); Grimmelmann, *supra* note 8, at 103 (describing how section 230 shields platforms from publisher or speaker liability); 47 U.S.C. § 230(c)(2) (describing protection for good faith blocking and screening of offensive material or otherwise objectionable content); Reno v. ACLU, 521 U.S. 844, 868-69 (1997) (emphasizing differences between radio, television, and internet); Klonick, *supra* note 13, at 1612 (contending *Reno* explicitly exempted internet from doctrine established in *Red Lion* and *Turner Broadcasting*). Corporate censors can prevent citizens from accessing their platforms in various degrees because social media platforms are not a governmental entity, and the Constitution does not safeguard their abridgement of First Amendment principles. *See* Wiener, *supra* note 12, at 242 (contending First Amendment does not reach social media platforms). The Court held that the elements justifying radio and cable television regulation were not present with the internet and precluded the medium from being regulated under the First Amendment. *See*

In order to protect the fundamental free speech rights of individual internet users, the judicial or legislative branch should determine policy regarding individual free speech on platforms.[168] Regarding the judicial branch, courts should reassess First Amendment jurisprudence in order to be more responsive towards internet platforms because analogies for forums and media that predate the internet are inapt for current disputes over free speech and social media.[169] For example, the Supreme Court could expand the public forum doctrine in *Marsh* to make ownership just one factor in a case-by-case analysis of whether a private space is functionally public.[170] To find a rule more suitable to the digital world, reexamining whether a private space possesses public characteristics would allow courts to maintain limited constitutional safeguards and to reconsider the balancing of rights between individuals and pervasive, private internet platforms.[171] A better approach for the digital world would recognize that private property often takes on public characteristics, which would allow courts to maintain some statutory and constitutional rights for public individuals using such property.[172]

Regarding the legislative branch, legislatures should lessen corporate free speech rights in favor of human speakers.[173] For example, Congress could regulate platforms as digital utilities with specified must-carry obligations; major internet platforms are arguably like utilities in that they provide "essential, unavoidable, and monopolistic services" for their users.[174] By imposing must-carry obligations on internet platforms, Congress and administrative agencies like the

---

Klonick, *supra* note 13, at 1612 (noting elements like "invasive nature, history of extensive regulation, and the scarcity of frequencies").

168. *See* Thompson, *supra* note 34, at 785 (noting online platforms and other intermediaries should remain "neutral implementers of these decisions").

169. *See* Wiener, *supra* note 12, at 225 (emphasizing need for change or adaption of public forum paradigm with evolution in communication technologies); *see also* Peters, *supra* note 12, at 994 (noting challenge of applying traditional doctrines to social media platforms). Social media platforms such as YouTube, Facebook, and Twitter are not easily classified as public or private. Peters, *supra* note 12, at 1000 (recognizing difficulty classifying virtual public forums in wake of their private ownership).

170. *See* Peters, *supra* note 12, at 1023 (diminishing importance of ownership in analyzing forums). The easiest way to have social media platforms fall within the public forum doctrine would be to expand *Marsh* to assess whether a private space is functionally public, not whether a private space is comparable to a municipality. *See id.* at 1023-24 (noting this doctrine expansion would broaden *Marsh's* scope beyond its current application); *see also* Marsh v. Alabama, 326 U.S. 501, 506 (1946) (stating ownership rights become statutorily and constitutionally circumscribed when owner allows general public use).

171. *See* Peters, *supra* note 12, at 1023 (noting private property open to public use lessens owner's rights compared to public's constitutional rights).

172. *See id.* (emphasizing need for case-by-case analysis to balance rights of public individuals and private entities).

173. *See* Bridy, *supra* note 6, at 197-98 (advocating corporate free speech rights deemphasized to effectively regulate internet platforms).

174. *See* Keller, *supra* note 44, at 16 (asserting Congress could pass legislation regulating major internet platforms like utilities). "The Supreme Court has acknowledged similar rights for other entities opposing must-carry obligations, varying from cable companies to parade organizers." *Id.* at 17-18 (noting Court required entities to yield to protect speech and information rights of larger public).

FCC could diminish the perceived unfair moderation of political speech by creating an obligation for platforms to carry such speech, while also maintaining platforms' ability and discretion to moderate harmful content.[175] Without a judicial or legislative response, internet platforms will create their own policies and reconcile private motives and public interests in a manner unlikely to ease concerns over individual speech rights on the internet.[176]

Regardless of the proposed solution, internet platforms' ability to regulate and adjudicate online content via their own moderation practices jeopardizes the freedom of expression of internet users without recourse.[177] Without a judicial or legislative response, internet platforms will continue to benefit from the CDA's section 230 immunity arising from their good faith moderation efforts or from claims treating them as publishers of user-generated content.[178] In *Packingham*, Justice Kennedy equated the development and integration of the internet in the daily lives of Americans to a "revolution of historic proportions," one which has the ability to alter the way individuals think and express themselves.[179] Absent significant judicial or legislative regulation, internet platforms will continue to become more influential in society, without any oversight from the very institutions citizens rely on to protect their individual rights.[180] While individuals

175. *See id.* at 11, 25 (explaining must-carry obligations protect individual users' speech rights by forcing platforms to keep speech online). While making platforms carry all legal expression seems extreme, proponents of must-carry obligations argue internet platforms need to be held to some sort of standard. *See id.* at 14 (noting must-carry obligations on internet would dwarf similar obligations upheld in radio and television). "[A]ny law requiring Internet platforms to carry specific voices or perspectives would face serious First Amendment obstacles." *Id.*

176. *See* Keller, *supra* note 45, at 2 (arguing democratic processes and court rulings create unclear guidance on how to protect online speech). Private platforms cannot reasonably be expected to change in the absence of clear legislative or judicial action. *See id.* (acknowledging unclear guidance allows private platforms to interpret law themselves and set own ethical rules). Most well-known platforms enforce their own community guidelines or terms of service in accordance with their core business goals or advertisers' demands. *See id.* at 4 (recognizing without such moderation practices, internet platforms would lose advertisers). Facing potential legal risks, a platform's safest option is to err on the side of removing content. *See id.* at 5 (acknowledging speedy compliance eliminates legal and financial costs associated with unresponsive moderation).

177. *See* Frosio, *supra* note 7, at 33 (assessing trade off between private and public governance of moderation practices on internet platforms). Private governance of content moderation provides algorithmic efficiency at the cost of users' fundamental rights, potentially jeopardizing freedom of expression "by limiting access to information, causing chilling effects, or curbing due process." *See id.* (emphasizing increased private ordering and decreased public involvement crushes due process and fundamental guarantees).

178. *See* Keller, *supra* note 44, at 12 (noting two specific aspects of section 230's liability shield); *see also* Grimmelmann, *supra* note 8, at 103 (discussing how moderators maintain blanket immunity via section 230 protections).

179. *See* Packingham v. North Carolina, 137 S. Ct. 1730, 1736 (2017) (recognizing Cyber Age's impact and potential on individual expression will remain unknown).

180. *See id.* (noting courts must recognize what they say today will become obsolete tomorrow); *see also* Keller, *supra* note 44, at 27 (acknowledging private internet companies concentrated communication power in moderating and regulating individual speech). At the intersection of state and private platform power, concerns over what legal arguments and outlets to use to regulate internet platforms are just beginning. *See* Keller, *supra* note 44, at 27 (recognizing widely disjointed approaches in grappling with our communications shifting to one common online infrastructure).

should anticipate an advancement in the law relating to internet platforms, the legal tools and public policy needed to challenge these platforms' power remain weak and underdeveloped.[181]

## IV. CONCLUSION

As internet platforms continue to become increasingly present, prevalent, and pervasive in the daily lives of individuals, the need to protect individual users' online speech will continue to clash with the private interests of internet platforms. Without a meaningful judicial or legislative response, private internet platforms will continue to benefit from the current inability of judicial doctrine and legislation to hold these platforms responsible for any questionable moderation practices. Left to create their own online policies and practices, internet platforms like social media websites will continue to promote their profit-maximization interests under the guise of safe freedom of online expression.

Internet platforms maintain the ability to self-moderate online speech based on the CDA's liability shield, lack of judicial precedent, and First Amendment jurisprudence. In this respect, the development of legal tools and public policy surrounding online free speech is critical for generating consequential change to protect the rights of individual internet users. To overcome the shortcomings of judicial precedent and legislative inadequacy, citizens need to push governmental institutions to protect their individual rights. While public discourse slowly recognizes the unprecedented power social media platforms yield, identifying the legal problems regarding internet platforms is a critical first step in effectuating a change in policy towards the immunities of private platforms and the rights of individual internet users.

181. *See* Keller, *supra* note 44, at 27 (noting power of virtual "public square" where people share their creativity, opinions, and other speech). For the first time, individual communication and expression share a common infrastructure, and today's communication ecosystem and legal mechanisms have barely begun to consider the unprecedented power of internet platforms in relation to citizens' constitutional rights. *See id.* (recognizing current law provides few or no checks to private platforms' power over online speech).