Essays

The First Amendment, Global Corporate Responsibility Standards, & the Quest for Online "Speech Nirvana"

EVELYN MARY ASWAD[†]

With large and powerful social media companies operating as worldwide speech regulators, it is unsurprising that governments have attempted to not only regulate how these companies address platform speech but also pressure them to deliver preferred speech outcomes. In 2024, the Supreme Court decided two cases addressing both themes in the U.S. context. In Moody v. NetChoice*, the Court explained how legislation regulating private platform curation of user-generated content runs afoul of First Amendment protections. And, in* Murthy v. Missouri*, the Court appeared to erect significant hurdles to challenging alleged governmental coercion of such platforms. These cases have left many wondering what can be done to achieve the legitimate societal aim of promoting broad and fair protections for user-generated speech on large social media platforms while addressing illicit governmental interference with such discourse. This Essay seeks to spur a dialogue about the potential for global corporate responsibility and free expression standards to provide a way forward. Under such standards, companies would apply a principled free expression framework in curating platform speech as well as proactively resist illicit governmental pressure. That said, as corporate adoption of these standards remains voluntary, this Essay urges a range of stakeholders to seize this norm-building moment to encourage platform adoption of these global standards or otherwise risk cementing unprincipled approaches to online speech.*

[†] Herman G. Kaiser Chair in International Law, Director of the Center for International Business & Human Rights, and Faculty Director of the International Business Law Graduate Program at the University of Oklahoma College of Law. The author thanks Nadine Strossen for her review and comments on a draft of this Essay. The author also thanks her research assistants, Madison Taylor and Jason Kersey, for their excellent assistance with this Essay. The views are solely those of the author.

TABLE OF CONTENTS

INTRODUCTION

"On the spectrum of dangers to free expression, there are few greater than allowing the government to change the speech of private actors in order to achieve its own conception of speech nirvana."[1]

Attempts at government-imposed online "speech nirvana" can occur in various ways, including through regulation of content on social media platforms and illicit pressure resulting in changes to corporate content moderation approaches. In 2024, the Supreme Court issued two important opinions regarding allegations that regulation of and pressure on platforms violated First Amendment protections.[2] One case focused on state governments' attempts to regulate the online expressive realm.[3] Specifically, Florida and Texas adopted laws in 2021 policing large online platforms[4] with the purpose of protecting their citizens from "Big Tech" censorship.[5] In adopting these laws, policymakers called social media platforms the "modern-day public square," urged that the future of democracy required "stand[ing] up to these technological oligarchs," and highlighted the "dangerous movement by social media companies to silence conservative viewpoints and ideas."[6] Though these laws emerged in response to

1. Moody v. NetChoice, LLC, 144 S. Ct. 2383, 2407 (2024).

2. *See id.*; *see also* Murthy v. Missouri, 144 S. Ct. 1972, 1997 (2024).

3. *Moody*, 144 S. Ct. at 2393.

4. Florida's law regulates any "social media platform" (which is defined broadly) with a gross annual revenue of over $100 million or more than 100 million active monthly users. FLA. STAT. ANN. § 501.2041(1)(g) (West 2025). The Texas law regulates platforms with more than fifty million monthly active users. TEX. BUS. & COM. CODE ANN. §§ 120.001(1), 120.002(b) (West 2025); *see also Moody*, 144 S. Ct. at 2395. This Essay focuses on the largest social media platforms.

5. *See, e.g.*, Press Release, Exec. Off. of the Governor, Governor Ron DeSantis Signs Bill to Stop the Censorship of Floridians by Big Tech (May 24, 2021), https://www.flgov.com/eog/news/press/2021/governor-ron-desantis-signs-bill-stop-censorship-floridians-big-tech; *see also* Press Release, Off. of the Texas Governor, Governor Abbott Signs Law Protecting Texans from Wrongful Social Media Censorship (Sept. 9, 2021) [hereinafter Texas Press Release], https://gov.texas.gov/news/post/governor-abbott-signs-law-protecting-texans-from-wrongful-social-media-censorship.

6. Texas Press Release, *supra* note 5. It should be noted that accusations of bias and retaliation have also been levied by those on the same political side of the aisle. *See, e.g.*, Donie O'Sullivan, *MAGA Opponents of Elon Musk Claim He Stripped Them of Their X Badges*, CNN BUS. (Dec. 27, 2024), https://www.cnn.com/2024/12/27/tech/maga-musk-x-visas/index.html. Various researchers have not found evidence of conservative bias in online content moderation. *See, e.g.*, Mohsen Mosleh, Qi Yang, Tauhid Zaman, Gordon Pennycook & David G. Rand, *Differences in Misinformation Sharing Can Lead to Politically Asymmetric Sanctions*, 634 NATURE 609, 609 (2024); PAUL M. BARRETT & J. GRANT SIMS, FALSE ACCUSATION: THE UNFOUNDED CLAIM THAT SOCIAL MEDIA COMPANIES CENSOR CONSERVATIVES 19 (2021); *Political Bias on Social Media Emerges from Users, Not Platform*, IND. UNIV. (Sept. 27, 2021), https://research.impact.iu.edu/key-areas/social-sciences/stories/social-media-platform-bias.html. Despite this research, polling shows conservatives believe there is bias in content moderation. *See* Monica Anderson, *Americans' Views of Technology Companies*, PEW RSCH. CTR. (Apr. 29, 2024), https://www.pewresearch.org/internet/2024/04/29/americans-views-of-technology-companies-2/ ("93% of Republicans say it's likely that social media sites intentionally censor political viewpoints that they find objectionable.").

alleged censorship of conservative speech, calls to regulate online speech and large platforms have come from both sides of the U.S. political aisle.[7]

The other Supreme Court case focused on "jawboning." Jawboning is an indirect form of censorship, involving the government imposing its will through the (often, but not always, secret) application of pressure on private companies rather than through the democratic, public law-making process.[8] In this case, several platform users, as well as Missouri and Louisiana, alleged that Biden Administration officials improperly pressured platforms to censor conservative views with respect to the COVID-19 pandemic and election-related speech.[9] Over the years, officials from both of the main U.S. political parties have been accused of applying improper governmental pressure on platforms.[10]

Although it's understandable for society to want the largest social media platforms with many millions (if not billions) of users to moderate content on their sites in a fair and non-discriminatory way, the Supreme Court made it clear in *Moody* that government-enforced viewpoint neutrality in corporate moderation is not consistent with the First Amendment.[11] Similarly, expecting governments to engage in any speech regulation through appropriate and legal means, rather than coercive means, is important for any well-functioning democracy. That said, the Supreme Court in *Murthy* seems to have erected significant hurdles for challenging potential jawboning, including with respect to standing.[12]

Given this constitutional backdrop, it is worth considering other ways to achieve fair and principled content moderation by the largest social media companies while maintaining platform resilience in the face of governmental jawboning attempts. One potential way forward involves adherence to global corporate responsibility standards. The UN Guiding Principles on Business and Human Rights ("UNGPs") calls on companies to "respect" human rights by avoiding infringements on rights in their operations and addressing any adverse

---

7. Eric Goldman, *"Speech Nirvanas" on the Internet: An Analysis of the U.S. Supreme Court's* Moody v. NetChoice *Decision*, 2024 CATO SUP. CT. REV. 125, 155 (observing that "[b]oth Democrats and Republicans favor censorial restrictions of the Internet; it's a rare topic that brings together legislators across the aisle. This leaves the Supreme Court as the last line of defense for Internet freedoms of speech and press.").

8. Jawboning occurs when "the government us[es] its power—or the threat of it—to indirectly bully individuals, institutions, or organizations into doing their bidding when it can't flat-out force them." *What is Jawboning? And Does It Violate the First Amendment?*, THE FOUND. FOR INDIVIDUAL RTS. & EXPRESSION (Nov. 8, 2024), https://www.thefire.org/research-learn/what-jawboning-and-does-it-violate-first-amendment.

9. Murthy v. Missouri, 144 S. Ct. 1972, 1981–84 (2024).

10. *See, e.g.*, Cat Zakrzewski & Cristiano Lima, *GOP Lawmakers Allege Big Tech Conspiracy, Even as Ex-Twitter Employees Rebut Them*, WASH. POST (Feb. 8, 2023), http://www.washingtonpost.com/technology/2023/02/08/house-republicans-twitter-files-collusion/ (noting allegations of jawboning of platforms by Republican government officials).

11. *See infra* notes 42–45 and accompanying text (discussing the Supreme Court's analysis of regulation of viewpoint neutrality).

12. *See infra* notes 68–89 (discussing the Supreme Court's standing analysis).

impacts.[13] Of the various ways the largest social media companies should respect human rights in their operations, two are directly related to freedom of expression.[14] First, those social media platforms should align their own speech rules to respect global protections for freedom of expression. These standards include avoiding discrimination on a variety of bases (including political ideology) in the regulation of speech.[15] Second, if governments require or pressure platforms to prohibit or burden speech, companies should tailor their reactions accordingly, in light of the same global free expression standards.[16] By voluntarily fulfilling both expectations embedded in global corporate responsibility standards, companies can better align with societal goals for fair and principled content moderation.

Part I of this Essay examines the two recent Supreme Court cases involving state regulation of social media and the alleged jawboning of such companies. Specifically, this Part focuses on how the Court's jurisprudence limits governmental regulatory attempts to "rejigger the expressive realm"[17] while setting a high bar for lodging challenges to jawboning. These two cases will likely make it difficult to achieve societal goals with respect to securing broad and fair protections for platform speech as well as resisting government jawboning. Part II examines global corporate responsibility standards as a potential alternative to realizing these goals.

I. FIRST AMENDMENT PROTECTIONS & ASSOCIATED CHALLENGES TO "SPEECH NIRVANA"

This Part examines the Supreme Court's analysis of content moderation regulation in *Moody* and describes the Court's approach to jawboning complaints in *Murthy.* This Part also assesses how the Court's jurisprudence not only limits state regulation as a means of addressing societal concerns about key aspects of platform content moderation but also seems to establish significant barriers to addressing alleged jawboning.

13. Human Rights Council Res. 17/4, UN Doc. A/HRC/RES/17/4 (July 6, 2011); John Ruggie, *Rep. Of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises to the Human Rights Council*, UN Doc. A/HRC/17/31, Principle 11 (Mar. 21, 2011) [hereinafter UNGPs].

14. Beyond respecting freedom of expression, social media companies should also respect, for example, the right to hold opinions without interference and privacy in their business models. *See, e.g.*, Evelyn Aswad, *Losing the Freedom to Be Human*, 52 COLUM. HUM. RTS. L. REV. 306, 359-363 (2020) (assessing how collecting and monetizing personal information undermines, *inter alia*, the right to not disclose or be punished for one's opinions).

15. *See infra* Part.II.B.1.

16. *See infra* Part.II.B.2.

17. Moody v. NetChoice, LLC, 144 S. Ct. 2383, 2402 (2024).

A. *MOODY* & CONTENT MODERATION

1. *Background on Recent Regulation in Florida and Texas*

In the name of addressing claims of bias and other complaints relating to social media content moderation,[18] Florida and Texas each adopted statutes regulating large social media platforms.[19] Though the two laws varied, both limited platforms' discretion in showing user-generated content and mandated that platforms give users reasons for certain content moderation decisions.[20] For example, the Florida statute prohibited methods of "'censor[ing]' or otherwise disfavoring posts–including deleting, altering, labeling, or deprioritizing them–based on its content or source."[21] The Texas law, subject to certain exceptions, prohibited platforms from "'censor[ing]' a user or a user's expression based on viewpoint."[22] This statute defined "censoring" as actions that "block, ban, remove, deplatform, demonetize, de-boost, restrict, deny equal access or visibility to, or otherwise discriminate against expression."[23]

Unsurprisingly, two trade associations (NetChoice, LLC and the Computer & Communications Industry Association) brought First Amendment challenges against both laws.[24] Both associations secured preliminary injunctions in Florida and Texas district courts, enjoining enforcement on the basis that the laws unconstitutionally restricted the platforms' "editorial judgment."[25] The Eleventh Circuit ruled against Florida's appeal, generally upholding the lower court's preliminary injunction.[26] The Fifth Circuit, however, vacated the preliminary injunction, holding that content moderation does not implicate the First Amendment and, even if it did, Texas could legislate in this field to protect "a diversity of ideas" on platforms.[27] Given this circuit split, the Supreme Court agreed to hear the appeals in both cases.[28]

---

18. *See supra* notes 5–6 and accompanying text (describing the nature of the bias criticisms lodged against social media companies).

19. *Moody*, 144 S. Ct. at 2395; *see supra* note 4 (explaining how the laws defined large platforms).

20. In *Moody*, the Court's opinion did not focus on the individualized-explanation provisions. 144 S. Ct. at 2395. This Essay similarly focuses on governmental regulation of corporate discretion in deciding whether and how to display user-generated content on platforms rather than the topic of providing users with explanations for content decisions.

21. *Id.* (citing FLA. STAT. ANN. § 501.2041(1)(b) (West 2025)). The Supreme Court cited particular examples in the statute that restricted content moderation, including its ban on "deprioritizing posts by or about political candidates" and the requirement to engage in content moderation "in a consistent manner." *Id.* (quoting FLA. STAT. ANN. §§ 501.2041(2)(h), 501.2041(2)(b) (West 2025)).

22. *Id.* at 2395–96 (quoting TEX. CIV. PRAC. & REM. CODE ANN. §§ 143A.002(a), 143A.006 (West 2025)).

23. *Id.* at 2396 (quoting TEX. CIV. PRAC. & REM. CODE ANN § 143A.001(1) (West 2025)).

24. *Id.*

25. *Id.*

26. *Id.*; *see also* Goldman, *supra* note 7, at 133 (explaining which provisions of the lower court's injunction were upheld).

27. *Moody*, 144 S. Ct. at 2396; *see also* Goldman, *supra* note 7, at 134 (describing the procedural history regarding the lifting of the injunction).

28. *Moody*, 144 S. Ct. at 2397; *see also* Goldman, *supra* note 7, at 135.

2. *Supreme Court Application of First Amendment Jurisprudence in the Digital Age*

While the Supreme Court did not decide the merits of the free speech challenge in *Moody*, the opinion provides an explanation of how First Amendment jurisprudence applies to governmental regulation of platform content moderation.[29] First, the Court emphasized that the First Amendment challenges in both cases were *facial* challenges to the laws–that is, the trade associations were arguing the laws must be struck down as unconstitutional before their implementation rather than raising a challenge with respect to a particular enforcement action.[30] In *Moody*, the Court lamented that its traditional test for *facial* challenges in the First Amendment context had been largely ignored in lower court proceedings.[31] Relying on its past jurisprudence, the Court explained that, to win a facial challenge in this context, the laws' various applications must be considered in order to decide if "a substantial number of the law's applications are unconstitutional, judged in relation to the statute's plainly legitimate sweep."[32] In other words, to invalidate a law for violating the First Amendment based on a facial challenge, a law's potential unconstitutional applications must outweigh its constitutional ones.[33] The Court observed that a variety of applications of the law on large social media platforms beyond regulation of their "heartland applications"–or "best-known services" such as Facebook's Newsfeed and YouTube's homepage–may fall within the law's reach and were not considered in the lower courts.[34] The Court therefore vacated the Circuit court judgments and remanded the cases for proper assessment of the test relating to such facial challenges, which maintained the injunctions.[35]

The Court's repeated emphasis on the continuing utility and applicability of existing First Amendment jurisprudence in the digital age, including with respect to attempts to regulate corporate content moderation, is notable. For example, the Court underlined many times that corporate entities, including social media platforms, have free speech rights.[36] The Court viewed the "essence" of platform content moderation as similar to "[t]raditional publishers and editors [who] select and shape other parties' expression into their own

29. It should be noted that, although the decision to vacate and remand the Circuit court cases was unanimous, the "Justices' unanimity was only superficial" as there were a variety of concurring opinions, with three justices not supporting First Amendment protection for content moderation. Goldman, *supra* note 7, at 135.

30. *Moody*, 144 S. Ct. at 2397.

31. *Id.* at 2394, 2398.

32. *Id.*at 2397 (quoting Ams. for Prosperity Found. v. Bonta, 594 U.S. 595, 615 (2021)).

33. *Id.*

34. *Id.* The Court noted the lower courts "did not address the full range of activities the laws cover, and measure the constitutional against the unconstitutional applications. In short, they treated these cases more like as-applied claims than like facial ones." *Id.* at 2397-98. The Court highlighted direct messaging as an example of the application of the law that had not been considered in the lower courts. *Id.* at 2398. The Court further noted that lower courts had also not assessed the potential applicability of the laws to other platforms such as Uber and Venmo, which was relevant to assessing the facial challenge on remand. *Id*.

35. *Id.* at 2399.

36. *Id.* at 2393, 2399–2403, 2405–09.

curated speech products" and noted such editorial choices are subject to First Amendment protection.[37] In reviewing its precedents, the Court emphasized that "'[w]hatever the challenges of applying the Constitution to ever-advancing technology, the basic principles' of the First Amendment 'do not vary.'"[38]

In *Moody*, the Court flagged fundamental errors in the Fifth Circuit's analysis of First Amendment principles. For instance, the Fifth Circuit found that the content choices of large platforms were "not speech" and therefore not subject to First Amendment protections.[39] The Court observed that, "[d]espite the relative novelty of the technology . . . the main problem in this case . . . is not new."[40] Accordingly, the Court rebutted the Fifth Circuit's reasoning and explained why platform content moderation *does* trigger First Amendment protections. To start, the Court cited a litany of precedents to emphasize (1) regulation mandating private actors who are engaged in their own expressive activity "to provide a forum for someone else's views implicates the First Amendment" and (2) curating speech generated by others has consistently been found to qualify as an "expressive activity."[41]

Additionally, the Court affirmed that under the First Amendment, the government cannot justify its actions by claiming it is "improving, or better balancing, the marketplace of ideas," which is what Texas stated it was doing in banning viewpoint discrimination in corporate content moderation.[42] Though the Court seemed to acknowledge that states and society may desire "access to a wide range of views,"[43] the First Amendment prohibits the government from undermining "private actors' speech to advance its own vision of ideological balance."[44] In one of its most memorable statements of the case, the Court found that with respect to "dangers to free expression, there are few greater than allowing the government to change the speech of private actors in order to achieve its own conception of *speech nirvana*."[45]

---

37. *Id.* at 2393.

38. *Id.* at 2403 (quoting Brown v. Ent. Merch.'s Assn., 564 U.S. 786, 790 (2011)). The Court went on to discuss at length how social media is different from traditional forms of communication but emphasized the continuation of the application of First Amendment principles that had "served the Nation well over many years." *Id*.

39. *Id.* at 2396.

40. *Id.* at 2399.

41. *Id.* at 2399–2402 (citing Supreme Court case law from 1974–2006).

42. *Id.* at 2402; *see also id.* at 2407 (deciding that the Court need not determine if strict or intermediate scrutiny applies as Texas's stated purpose of promoting diversity in the marketplace of ideas is related to speech suppression and does not constitute a "substantial" governmental interest).

43. *Id.* at 2407 ("States (and their citizens) are of course right to want an expressive realm in which the public has access to a wide range of views. That is, indeed, a fundamental aim of the First Amendment. But the way the First Amendment achieves that goal is by preventing the government from 'tilting public debate in a preferred direction.'" (quoting Sorrell v. IMS Health Inc., 564 U.S. 552, 578–79 (2011) (emphasis omitted))).

44. *Id.*

45. *Id.* (emphasis added). The Court emphasized this point numerous times, including in its conclusion by stating starkly the government cannot promote some viewpoints by burdening the expression of other private actors, which would "enable it to control the expression of ideas, promoting those it favors and suppressing those it does not. And that is what the First Amendment protects us all from." *Id*. at 2409.

3. *Reflections*

The *Moody* opinion reflects an important application of First Amendment protections in the digital age. It significantly curtailed possibilities for state regulation of key aspects of corporate content moderation by, *inter alia,* reiterating a trifecta of principles. These principles flow from the Court's jurisprudential anchor in the general ongoing applicability of First Amendment principles to evolving technology.[46] First, the Court reiterated the free speech rights of companies, including social media.[47] As one scholar has observed, the Supreme Court's prior jurisprudence "reaffirmed the protection that American law provides for corporate power, including over quintessential democratic public discourse" and the development of the Internet.[48] Second, the Court found that heartland applications of corporate content curation constitute editorial choices and that expressive activity garners First Amendment speech protection.[49] Third, the Court explained that imposing a "balanced" speech environment on platforms is not a legitimate governmental objective.[50] The intersection of these principles makes governmental regulation of key aspects of platform content curation challenging.

A variety of commentators have also interpreted *Moody* as raising significant barriers to such regulation. Professor Noah Feldman called the decision a "blockbuster" and "the Brown v. Board of Education of the emerging field of social media law."[51] He noted "the U.S. Supreme Court has held for the first time that social media platforms, just like newspapers, have First Amendment rights that bar the government from forcing them to leave up or take down content."[52] Similarly, Professor Eric Goldman stated "Justice Elena Kagan's majority opinion was a rousing celebration of the First Amendment online. Critically, the majority said that the First Amendment protects social media platforms' content moderation efforts."[53] Human rights non-

46. *See supra* note 38 and accompanying text.

47. *See supra* note 36 (listing *Moody*'s references to corporate speech rights).

48. Evelyn Douek, *The Tug Between Private and Public Power Online*, 61 DUQ. L. REV. 209, 210 (2023) (referring to Citizens United v. FEC, 558 U.S. 310 (2010)).

49. *See supra* notes 40–41 and accompanying text.

50. *See supra* notes 42–45 and accompanying text.

51. Noah Feldman, *Social Media Ruling Is a Free-Speech Landmark: Noah Feldman*, BLOOMBERG LAW (July 1, 2024), https://news.bloomberglaw.com/us-law-week/social-media-ruling-is-a-free-speech-landmark-noah-feldman.

52. *Id.*

53. Goldman, *supra* note 7, at 126. Professor Goldman noted that "[a]fter reading the majority opinion [of both liberal and conservative justices], many legislators ought to rethink their censorial agendas toward Internet services. Otherwise, those laws will be invalidated." *Id.* at 142.

governmental organizations[54] and practitioners[55] appear to have drawn similar conclusions from the opinion. That said, *Moody* itself noted that it focused on how the laws related to social media platforms' "heartland applications" rather than other potential applications.[56] This will be considered by the lower courts and important to monitor.[57] In sum, though an important victory for First Amendment protection from governmental regulation of the expression of ideas, *Moody* leaves many wondering how society can encourage extraordinarily powerful platforms to provide broad protections for freedom of expression as well as other key assurances, including with respect to protection from discrimination.[58]

---

54. *See, e.g.*, *Supreme Court Ruling Underscores Importance of Free Speech Online*, AM. C.L. UNION (July 1, 2024), https://www.aclu.org/press-releases/supreme-court-ruling-underscores-importance-of-free-speech-online ("The [Supreme C]ourt recognized that government attempts to control the editorial decisions of social media companies violate the First Amendment."); David Greene, *In These Five Social Media Speech Cases, Supreme Court Set Foundational Rules for the Future*, ELEC. FRONTIER FOUND. (Aug. 14, 2024), https://www.eff.org/deeplinks/2024/08/through-line-suprme-courts-social-media-cases-same-first-amendment-rules-apply ("[T]he Supreme Court explained that social media platforms typically exercise their own protected First Amendment rights when they edit and curate which posts they show to their users, and the government may violate the First Amendment when it requires them to publish or amplify posts.").

55. *See, e.g.*, Aaron Schur, *More Wins for Free Speech on the Internet*, YELP BLOG (July 1, 2024), https://blog.yelp.com/news/more-wins-for-free-speech-on-the-internet/ ("First Amendment fully protects companies' editorial decisions about what to include and what to exclude on their platforms."); Adam S. Sieff, Ambika Kumar & David M. Gossett, *Moody Decision Confirms First Amendment Protects Online Platforms*, DAVIS WRIGHT TREMAINE (July 11, 2024), https://www.dwt.com/insights/2024/07/scotus-moody-ruling-a-win-for-online-platforms (concluding that platforms have a First Amendment right to "select and arrange content"); Seth P. Waxman, Ari Holtzblatt & Allison Schultz, *What's Next After Major First Amendment Win for Online Companies in Supreme Court's NetChoice Decision?*, WILMERHALE (July 3, 2024), https://www.wilmerhale.com/en/insights/client-alerts/20240702-whats-next-after-major-first-amendment-win-for-online-companies-in-supreme-courts-netchoice-decision ("[L]ower courts are likely to follow the entire majority opinion, including those portions setting out robust First Amendment principles and applying those principles to hold that the Texas law likely violates the First Amendment as applied to paradigmatic social-media functions.").

56. For example, the Court noted it was not clear "to what extent, if at all, [the Florida and Texas laws] affect social-media giants' other services, like direct messaging, or what they have to say about other platforms and functions." Moody v. NetChoice, LLC, 144 S. Ct. 2383, 2398 (2024). The opinion also noted that it did not focus on certain aspects of content moderation, including whether "feeds whose algorithms respond solely to how users act online–giving them the content they appear to want, without any regard to independent content standards" would be protected by the First Amendment. *Id*. at 2404 n.5.

57. Some scholars have noted that, by mandating a "close, fact-specific analysis, *Moody* . . . suggests a more nuanced First Amendment jurisprudence than many expected." Kyle Langvardt & Alan Z. Rozenshtein, *Beyond the Editorial Analogy: First Amendment Protections for Platform Content Moderation After* Moody v. NetChoice, 6 J. FREE SPEECH L. 1, 3 (2025). These scholars recognize the hurdles *Moody* imposes on regulation but argue there are potentially three "promising avenues" for content moderation regulation: "content-neutral design mandates [that] shape how platforms distribute and amplify content," procedural mandates that increase transparency, and "narrowly targeted access rights" in certain situations. *Id.* at 44–45.

58. Scholars have expressed concerns that the Supreme Court's approach focuses on the rights of corporations at the expense of users' free speech interests. *See* Jack M. Balkin, Moody v. NetChoice*: The Supreme Court Meets the Free Speech Triangle*, 2024 SUP. CT. REV. 127, 158 (2025) (observing how traditional First Amendment jurisprudence can mean that "[t]he speech interests of end users are left completely out of the picture"); Langvardt & Rozenshtein, *supra* note 57, at 35 ("This relationship between private moderation and constitutional values creates a seeming paradox: While the First Amendment restricts only government action, laws requiring platforms to host speech might actually advance First Amendment interests."). Professors

B. *MURTHY* & JAWBONING

1. *Case Background*

In *Murthy*, several social media users and two states, Missouri and Louisiana, sued a variety of Executive Branch agencies and officials, alleging they violated the First Amendment by improperly pressuring platforms (including Facebook, Twitter, and YouTube) to suppress their speech about COVID-19 and elections.[59] The Court described a variety of these governmental interactions with the platforms. For example, officials from the White House and Surgeon General's Office had raised concerns with the platforms about inadequate content moderation relating to COVID-19.[60] Moreover, the White House raised the specter of potential regulation as a response.[61] The Center for Disease Control (CDC) shared its expertise relating to the COVID-19 pandemic, often responding to fact-check requests about particular claims.[62] The Federal Bureau of Investigations (FBI) and the Cybersecurity and Infrastructure Security Agency (CISA) flagged concerns to the platforms about election misinformation and foreign interference.[63] The dissenting justices in *Murthy* seemed to view the majority's rendition of these interactions as sweeping under the carpet their

Langvardt and Rozenshtein have also noted that "conservatives are correct about one crucial point: If major platforms choose to, they could indeed systematically censor disfavored views and distort public discourse." *Id*. at 10.

59. *See* Murthy v. Missouri, 144 S. Ct. 1972, 1982–84 (2024). The five social media users who alleged demotion or removal of their speech included a health activist, a website owner, and three doctors. *Id*. at 1984. The two states asserted that the platforms had removed the speech of "state entities and officials, as well as their citizens' speech." *Id.* In its opinion, the Supreme Court seemed to understand platform "suppression" of speech as covering a range of content moderation tools, including labeling, demoting, and deleting content and suspending users for repeated violations of platform rules. *Id.* at 1982.

60. The White House officials who spoke to the social media platforms included the Director of Digital Strategy and COVID-19 response team members. *Id.* The interactions included sharing the "concern that Facebook in particular was 'one of the top drivers of vaccine hesitancy'" and complaints of prior violence when Facebook did not act on problematic content. *Id.* at 1982–83. One official shared with Facebook that the White House had been considering its options for dealing with platform content moderation that it disagreed with. *Id.* The Supreme Court noted that some White House "communications were more aggressive than others." *Id.* at 1983. The Surgeon General publicly encouraged platforms to change their content moderation policies and enforcement in various ways, including not amplifying health misinformation. *Id.*

61. In public statements, the White House urged platforms to increase their moderation of COVID-19 misinformation and publicly "raised the possibility of reforms aimed at the platforms, including changes to the antitrust laws" and Section 230 of the Communications Decency Act ("Section 230"). *Id.* Section 230 provides certain extensive protections from liability for platform content moderation of user-generated content. *See* Eric Goldman, *Why Section 230 is Better Than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33, 36–44 (2019) (explaining Section 230's broad substantive and procedural protections for platforms).

62. *Murthy*, 144 S. Ct. at 1983.

63. *Id.* For example, the FBI "warned the platforms about the potential for a Russian hack-and-leak operation" and some platforms changed their rules to bar "posting hacked materials." *Id.* CISA had sent the platforms reports of election misinformation, with accompanying assurances that the agency "'w[ould] not take any action, favorable or unfavorable, toward social media companies" regardless of "how and whether" they would react to the information. *Id.*

severity and impact given social media companies' reliance on the federal government to protect key aspects of their underlying business models.[64]

The plaintiffs persuaded lower courts to issue an injunction preventing federal officials and agencies from pressuring platforms to censor and take other content moderation steps regarding protected speech.[65] The district court found that the officials had "coerced" or "significantly encouraged" the platforms, such that the companies' content moderation decisions were attributable to the federal government.[66] The Fifth Circuit ultimately maintained the injunction, albeit with some modification, after finding the federal officials and agencies had at least "significantly encouraged" the platforms, while some had also "coerced" the companies; this "rendered [the government] responsible for the private platforms' content moderation decisions."[67]

2. *Supreme Court Analysis*

The Supreme Court held that all plaintiffs lacked standing to bring the lawsuit, which meant there was no jurisdiction to hear the merits of the case.[68] In reaching this conclusion, the Court articulated the following test for standing in the context of an injunction: "the plaintiffs must show a substantial risk that, in the near future, at least one platform will restrict the speech of at least one plaintiff in response to the actions of at least one Government defendant."[69] The Court explained that the allegations of past censorship were "relevant only for their predictive value" as plaintiffs were seeking an injunction rather than compensatory relief.[70] The Court found the "record of past restrictions" lacked "*specific causation findings* with respect to *any discrete instance* of content moderation."[71] This lack of causation ultimately undermined the plaintiffs' ability to prove likely future injury attributable to the defendants, as well as the likelihood that such injury could be redressed through an injunction.[72]

64. The dissenting justices (Justice Alito, J., Thomas, J., and Gorsuch, J.) agreed with the district court that the governmental interactions with platforms were a "far-reaching and widespread censorship campaign." *Id.* at 1997) (Alito, J., dissenting) (quoting Missouri v. Biden, 680 F. Supp. 3d 630, 729 (WD La. 2023)). These justices noted that social media platforms are now "a leading source of news" but far more vulnerable to governmental pressure than legacy news outlets because social media entities rely, *inter alia*, on Section 230 protections, have deep concerns about antitrust actions, and need the assistance of the federal government to push back on foreign (especially European Union) regulation. *Id.* at 1999.

65. *Id.* at 1984.

66. *Id.*

67. *Id.* at 1981, 1985. The Fifth Circuit's modified injunction prohibited the defendants from coercing or significantly encouraging social media companies "to remove, delete, suppress, or reduce, including through altering their algorithms" user-generated content "containing protected free speech." *Id.* at 1984–85. The injunction was not limited to social media used by the plaintiffs. *Id.*

68. *Id.* ("We begin—and end—with standing.").

69. *Id.* at 1986.

70. *Id.* at 1987.

71. *Id.* (emphasis added).

72. *Id.* at 1992–93.

The Court rejected the lower court's "broad" findings and deductions from the record. The Court cited multiple reasons leading it to doubt the causal connection between the federal officials' communications and the reason for the platforms' various content moderation decisions. For example, it found the platforms had adopted significant speech restrictions before federal officials became involved, and "continued to exercise their independent judgment" including by correcting federal officials' allegations that content policies had been violated, and regularly consulting with non-government experts on content moderation issues.[73] The Court concluded that "the platforms had independent incentives to moderate content and often exercised their own judgment."[74] Additionally, the Court held that the lower courts erred in treating the plaintiffs and defendants in the aggregate and should have required each plaintiff to demonstrate standing for each claim against each defendant.[75] The Court found the plaintiffs failed to demonstrate how platform restrictions on their speech were tied to governmental communications.[76] This failure, coupled with a lack of demonstrated "continued pressure" from those officials, essentially sealed the majority's finding of a lack of standing.[77] The opinion also rejected the broad assertion that the "right to listen" to others' speech could be the basis for granting these plaintiffs standing.[78] The Court, therefore, reversed the Fifth Circuit's judgment and remanded the case for action consistent with its opinion.[79]

3. *Implications and Open Questions*

While *Murthy* did not address the substantive test for government coercion in the context of social media, the Court's opinion raises several questions about how high and complex the bar is for plaintiffs to assert standing in First Amendment challenges when alleging governmental jawboning of platforms. Professors Evelyn Douek and Genevieve Lakier have raised concerns that the "highly formalistic test of standing that *Murthy* set out is likely to undermine the [Court's] substantive rule against jawboning."[80] They explain that "*Murthy*'s insistence that plaintiffs in jawboning cases must establish that social media

73. *Id.* at 1987. The Court also noted that key factual findings of the district court were clearly erroneous. *Id*. at 1988 n.4.

74. *Id.* at 1987–88.

75. *Id.* at 1988.

76. *Id.* at 1989–93.

77. *Id.* at 1993–96. The three dissenting justices disagreed with the majority's standing analysis. *Id.* The dissent reviewed the standing of one of the plaintiffs and found she would be "likely to prevail on her claim that the White House coerced Facebook into censoring her speech." *Id.* at 2015 (Alito, J., dissenting).

78. *Id.* at 1996 (noting that the "right to listen" basis for standing requires a display of "the listener [having] a concrete, specific connection to the speaker").

79. *Id.* at 1997.

80. Evelyn Douek & Genevieve Lakier, *Lochner.com?*, 138 HARV. L. REV. 100, 147 (2024). Professor Goldman has also raised concerns about the implications of the Court's standing analysis in *Murthy*. *See* Goldman, *supra* note 7, at 149 ("Together, the *Moody* and *Murthy* cases are a one-two punch for challengers of government censorship. *Moody* drives challengers away from facial challenges and toward as-applied challenges, but *Murthy* highlights potential standing difficulties with as-applied challenges.").

companies did not suppress their speech because of their independent incentives to do so strongly suggests that unless they have *smoking gun* evidence, plaintiffs in platform jawboning cases will rarely be able to establish the kind of causation that is necessary to sue."[81] Indeed, with a myriad of factors affecting content moderation decisions, including technological capabilities and scalability of platform interventions,[82] advertiser preferences to exhibit their ads near uncontroversial content,[83] civil society demands,[84] foreign pressure and regulation,[85] and various (often contrasting views) by content moderation teams within a company,[86] the exact cause of removal or demotion of content may often be difficult to prove. This can undermine standing arguments and thus challenges to governmental jawboning.

In addition, the *Murthy* dissent raised concerns about the largest platforms' complex relationship with the federal government and how this may impact a jawboning analysis. The dissent took the position that social media companies are vulnerable to governmental pressure because their business model relies on continuing certain federal government policies, including with respect to Section 230 protections from liability and the enforcement of antitrust laws, and corporate dependence on the Executive Branch to repel problematic regulation in foreign jurisdictions.[87] Though not highlighted by the dissent, another vulnerable aspect of social media companies' business model is its foundation in harvesting and monetizing personal data to fuel engagement and targeted advertising.[88] Therefore, the ability to continue operating with no contemporary federal data privacy law may also impact how platforms calculate their responses to government pressure. These and evolving governmental pressure points should be assessed in context when considering where to draw the line on illicit jawboning of platforms in the digital age.

---

81. Douek & Lakier, *supra* note 80, at 148 (emphasis added). That said, the scholars did note that *Murthy* "may be more an indication of the Court's frustration with the Fifth Circuit's indulgence of the culture wars than a reliable indicator of the procedural hurdles that plaintiffs will face in platform jawboning cases going forward." *Id.* at 149–50.

82. *See, e.g.*, Rachel Griffin, *Algorithmic Content Moderation Brings New Opportunities and Risks*, CTR. FOR INT'L GOVERNANCE INNOVATION (Oct. 23, 2023), https://www.cigionline.org/articles/algorithmic-content-moderation-brings-new-opportunities-and-risks/.

83. *See, e.g.*, Suzanne Vranica & Patience Haggin, *Meta's Free-Speech Shift Made It Clear to Advertisers: 'Brand Safety' Is Out of Vogue*, WALL ST. J. (Jan. 25, 2025), https://www.wsj.com/business/media/meta-brand-safety-content-moderation-policy-changes-17308d9e.

84. *See, e.g.*, Press Release, The Leadership Conf. on Civ. & Human Rts., Civil Society Organizations Demand Big Tech Protect Voters Against Disinformation (June 13, 2024), https://civilrights.org/2024/06/13/civil-society-big-tech-voters-disinformation/.

85. *See* Dawn Carla Nunziato, *The Digital Service Act and the Brussels Effect on Platform Content Moderation*, 24 CHI. J. INT'L L. 115, 117 (2023); Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1041–50 (2018).

86. *See, e.g.*, Aimee Picchi, *Twitter Files: What They Are and Why They Matter*, CBS NEWS, (Dec. 14, 2022, 6:46 PM), https://www.cbsnews.com/news/twitter-files-matt-taibbi-bari-weiss-michael-shellenberger-elon-musk/ (discussing debates about Twitter's content moderation decisions).

87. Murthy v. Missouri, 144 S. Ct. 1972, 1999 (Alito, J., dissenting).

88. *See* Aswad, *supra* note 14, at 306.

As flagged by Professor Daphne Keller, another troubling (but related) scenario may be neglected in the majority's approach to causation: what if a platform is passive in the face of governmental requests?[89] In other words, what if a platform voluntarily implements governmental content moderation requests to make its life easier? For example, imagine a platform that goes along to get along with the government and not because it was coerced. Will that be considered acceptable governmental conduct and not jawboning under the First Amendment? Should a corporate posture of preemptive capitulation to future jawboning determine whether a violation of First Amendment free speech rights has occurred? This scenario and the overall hurdles to standing for plaintiffs alleging jawboning seem to chip away at the societal goals of eliminating illicit governmental pressure on platform-created digital town squares.

II. GLOBAL FREEDOM OF EXPRESSION AND CORPORATE RESPONSIBILITY STANDARDS

There are legitimate societal aims that call for the largest social media platforms to allow broad discussion online without discrimination and seek to prevent illicit government intervention that could distort such digital expressive realms. The recent Supreme Court cases, *Moody* and *Murthy*, raise questions about how such goals can be achieved. *Moody* seems to significantly limit governmental regulation of these platforms as a way of promoting "speech nirvana," while *Murthy* appears to make bringing complaints of governmental jawboning of platforms quite difficult. Part II seeks to spur a dialogue on an alternate route to achieve legitimate societal aims: reliance on global human rights standards. This Part begins by delving into the global standard for the human right to freedom of expression. It next examines the UNGPs' call on the largest social media platforms to respect this global free speech standard, not only in curating individuals' speech on their platforms, but also in their dealings with governments. Unlike the Supreme Court's approach, the UNGPs keep the focus on individuals' rights to free expression. Part II concludes by urging stakeholders to incentivize adoption of this voluntary framework.

A. THE GLOBAL FREEDOM OF EXPRESSION STANDARD

The International Covenant on Civil and Political Rights (ICCPR), a foundational treaty of the United Nation's (UN) human rights system, embodies the global standard for the protection of the right to freedom of expression.[90]

89. Stan. Cyber Pol'y Ctr., *The Supreme Court and Internet Platforms*, YOUTUBE, at 28:00–30:30 (Oct. 1, 2024), https://www.youtube.com/watch?v=60kpHwc5Nl0.

90. International Covenant on Civil and Political Rights art. 19, ¶ 2, *adopted by the UN G.A.* Dec. 16, 1966, 999 U.N.T.S 171 [hereinafter ICCPR]. This treaty along with the Universal Declaration of Human Rights (UDHR) and the International Covenant on Economic, Social, and Cultural Rights (ICESCR) are collectively known as the "International Bill of Human Rights" in the UN system. *International Bill of Human Rights*, UN OFF. OF THE HIGH COMM'R FOR HUM. RTS., https://www.ohchr.org/en/what-are-human-rights/international-bill-human-rights (last visited June 29, 2025).

There are 174 State Parties to this treaty, including the United States.[91] ICCPR Article 19(2) provides broad protection for free speech, including the right "to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."[92] In addition, the ICCPR requires that its rights be implemented without discrimination "of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status."[93]

ICCPR Article 19(3), however, provides governments with discretion to limit speech if they can prove that each prong of a three-part test is met. If one prong is not met, the restriction is not permitted.[94] The text provides that freedom of expression can be "subject to certain restrictions, but these shall only be such as are [1] provided by law and [2] are necessary: [3] (a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals."[95] This tripartite analysis for evaluating restrictions on speech is often referred to as the legality, necessity, and legitimacy test.[96] Each part of the tripartite test covers different aspects of governmental speech regulation, together forming a principled web of hurdles for governments to overcome when seeking to burden speech.

The "legality" test, which is derived from the phrase "provided by law" in Article 19(3), means that any governmental limitation on speech be, among

91. *International Covenant on Civil and Political Rights*, UNITED NATIONS TREATY COLLECTION, https://treaties.un.org/PAGES/ViewDetails.aspx?chapter=4&clang=_en&mtdsg_no=IV-4&src=TREATY [hereinafter *UN Treaty Collection: ICCPR*] (last visited Aug. 31, 2025). The United States became a party in 1992. *Id.* For relevant reservations lodged by the United States when joining this treaty, see *infra* note 130.

92. ICCPR, *supra* note 90, art. 19, ¶ 2.

93. *Id.* art. 2, ¶ 1.

94. *Id.* art. 19, ¶ 3. The UN Human Rights Committee (HRC), which is comprised of independent experts elected by ICCPR State Parties to monitor implementation of the treaty, has noted that the burden of demonstrating this three-part test is met falls on the government. UN Hum. Rts. Comm., General Comment No. 34, ¶¶ 27, 35, UN Doc. CCPR/C/GC/34 (Sept. 12, 2011) [hereinafter GC 34]. In other words, there is a presumption in favor of speech, i.e., the burden is not on speakers to prove a right to speak or on listeners to prove a right to listen. The HRC discharges its monitoring functions by, *inter alia*, assessing periodic reports of State Parties and issuing recommended interpretations of the treaty (known as "General Comments"). ICCPR, *supra* note 90, art. 40, ¶¶ 1, 4.

95. ICCPR, *supra* note 90, art. 19, ¶ 3.

96. *See, e.g.*, David Kaye (Special Rapporteur), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN GAOR, 38th Sess., Agenda Item 3, UN Doc. A/HRC/38/35, ¶ 8 (Apr. 6, 2018) [hereinafter SR 2018 Report], https://docs.un.org/en/A/HRC/38/35 (observing that burdens on speech must meet "the cumulative conditions of legality, necessity and legitimacy."). The UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression is an independent expert appointed by UN Human Rights Council member states to monitor implementation of UN standards on freedom of speech. *Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN OFF. OF THE HIGH COMM'R FOR HUM. RTS., https://www.ohchr.org/en/special-procedures/sr-freedom-of-opinion-and-expression (last visited July 9, 2025).

other things, based on a properly promulgated law[97] and not unduly vague.[98] The law must give proper notice both to the public and to the governmental authorities implementing the law.[99] UN experts have found laws on a variety of topics to be improperly vague, including those involving hate speech,[100] national security,[101] counter-terrorism,[102] and public order.[103]

The necessity condition, which derives from the word "necessary" in Article 19(3), means that the limitation or burden on speech must be (1) the "least intrusive means" to achieve a legitimate aim and (2) proportional to the expected benefit.[104] Limitations on speech will not survive the "least intrusive means" test if the government cannot demonstrate three things.[105] First, if the government can achieve its legitimate aim without burdening speech, then a speech restriction will not be the least intrusive means of achieving the aim (e.g., deploying media and digital literacy initiatives before resorting to speech

97. *See, e.g.*, David Kaye (Special Rapporteur), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN GAOR, 72d. Sess., Agenda Item 73(b), UN Doc. A/72/350, ¶ 37 (Aug. 18, 2017), https://docs.un.org/en/A/72/350 ("The requirement of legality . . . requires that regular procedures be followed in the adoption of restrictions . . . .").

98. *See* GC 34, *supra* note 94, at ¶ 25.

99. *Id.* ("[A] norm, to be characterized as a 'law', must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly and it must be made accessible to the public. A law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution.").

100. *See, e.g.*, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Communication to Egypt, UN Doc. OL EGY 13/2018, at 1–2 (Aug. 9, 2018), https://bit.ly/3jQK7N4 (finding Egypt's law that would suspend websites if they "spread[] hateful views" to be unduly vague).

101. *See, e.g.*, Frank La Rue (Special Rapporteur), *Addendum to the Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN GAOR, 17th Sess., Agenda Item 3, UN Doc. A/HRC/17/27/Add.1, at ¶ 584 (May 27, 2011), https://digitallibrary.un.org/record/706200 (critiquing China's ban on speech relating to the "subversion of state power").

102. *See, e.g.*, UN Working Group on Arbitrary Detention, Independent Expert on the Promotion of a Democratic and Equitable International Order, the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, the Special Rapporteur on the Situation of Human Rights Defenders, and the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism, Communication to France, UN Doc. OL FRA 1/2015, at 6 (Feb. 3, 2015) (translated text), https://bit.ly/2UASOk3 (criticizing France for using vague speech restrictions, including bans on "glorification" or "promotion" of terrorism).

103. *See, e.g.*, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Expression & Special Rapporteur on the Situation of Human Rights Defenders, Communication to Bangladesh, UN Doc. OL BGD 4/2018, at 4 (May 14, 2018), https://bit.ly/3hsW6P1 (criticizing a draft Bangladeshi bill that would prohibit speech which "ruins communal harmony or creates instability or disorder or disturbs or is about to disturb the law and order situation.").

104. GC 34, *supra* note 94, at ¶ 34 (restrictions on speech "must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected . . . .").

105. Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 47 (2018). UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression David Kaye endorsed this trilogy of questions as a means of assessing the necessity of a speech restriction. David Kaye (Special Rapporteur), *Rep. on Online Hate Speech*, UN GAOR, 74th Sess., Agenda Item 70(b), UN Doc. A/74/486, ¶ 52 (Oct. 9, 2019) [hereinafter SR 2019 Report], https://docs.un.org/en/A/74/486 (citing positively to these three inquiries in assessing if platforms have selected the least intrusive means of burdening speech in content moderation).

bans).[106] Second, if achieving the aim without burdening speech is not possible, the restriction will still fail this test if the government has not selected the option that least burdens speech (e.g., the government selects criminal penalties when civil sanctions are sufficient to achieve the aim).[107] Third, if the burden on speech is not effective in achieving the legitimate objective, it is not an appropriate burden on speech.[108]

With respect to the "least intrusive means" test, UN human rights experts have often determined that where a range of options exist to achieve a legitimate aim, silencing speech will rarely, if ever, be appropriate unless there is imminent and likely harm.[109] Where the harm is long-term or unlikely, there are governmental interventions short of burdening speech that are more appropriate.[110] With respect to proportionality, the the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression has advised this means ensuring not only that "restrictions 'target a specific objective and do not unduly intrude upon the other rights of targeted persons'" but also "interference with third part[y] rights . . . be limited and justified in light of the [public interest aim]."[111] Thus, the proportionality test serves as a second "check" on the government to ensure the speech ban is not only the least intrusive means to achieve the aim, but also proportional to the burdens placed on speakers and listeners.

---

106. Aswad, *supra* note 105, at 47.

107. *Id.*

108. *Id.*

109. For example, the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression has stated:

> For a restriction to be necessary, it must . . . not be more restrictive than is required for the achievement of the desired purpose or protected right . . . . [T]he authorities must demonstrate, in specific and individualized fashion, the precise nature of the *imminent* threat, as well as the necessity for and the proportionality of the specific action taken. A direct and immediate connection between the expression (or the information to be disclosed) and the alleged threat must be established.

Frank La Rue (Special Rapporteur), *Rep. on the Right to Access Information and the Right to Truth*, UN GAOR, 68th Sess., Agenda Item 69(b), UN Doc. A/68/362, ¶¶ 52–53 (Sept. 4, 2013) (emphasis added), https://docs.un.org/en/A/68/362; s*ee also* Frank La Rue (Special Rapporteur), *Rep. on the Internet and Freedom of Expression*, UN Human Rights Council, 17th Sess., Agenda Item 3, UN Doc. A/HRC/17/27, ¶ 36 (May 16, 2011), https://docs.un.org/en/A/HRC/17/27 (observing that limitations on speech for counter-terrorism or national security purposes should only occur when "(a) the expression is intended to incite *imminent* violence; (b) it is *likely* to incite such violence; and (c) there is a *direct and immediate connection* between the expression and the likelihood or occurrence of such violence." (emphasis added)).

110. *See, e.g.*, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression & Special Rapporteur on the Situation of Human Rights Defenders, Communication to Qatar, UN Doc. OL QAT 1/2020, at 3 (Apr. 14, 2020), https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=25158 (urging Qatar to fight disinformation through fact checking, medial literacy, and education rather than speech bans).

111. *See, e.g.*, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression & Special Rapporteur on Freedom of Religion or Belief, Communication to the United States, UN Doc. OL USA 6/2017, at 3 (May 9, 2017), https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/Legislation/OL-USA-6-2017.pdf.

The third prong of the tripartite test (the "legitimacy" condition) requires the government to pursue one of the enumerated legitimate public interest aims in limiting expression.[112] The listed legitimate aims are respect for the rights and reputations of others and the protection of national security, public order, health, and morals.[113] UN independent experts made clear that governments may not seek to justify speech bans based on any additional aims,[114] and the listed aims may not be invoked as pretexts for illicit governmental objectives.[115]

While ICCPR Article 19(3) permits governments to limit speech if they can demonstrate compliance with the tripartite test, the treaty also contains a mandatory speech ban. Article 20 requires speech prohibitions in two instances: (a) "[a]ny propaganda for war"[116] and (b) "[a]ny advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence."[117] While the ban on war propaganda has not been the subject of significant focus in the UN system,[118] the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression has recommended interpreting the second prohibition on advocacy of incitement to require consideration of both imminent and likely harm.[119] In addition to this high bar, the UN Human Rights Committee has stressed the ICCPR Article 19(3)'s tripartite test must be met even when a speech restriction is imposed pursuant to Article 20.[120] These two interpretations raise significant and principled hurdles for a government to surmount when imposing a speech prohibition under Article 20.

It should be noted that the UN Convention on the Elimination of All Forms of Racial Discrimination (CERD), which has 182 State Parties, also contains a

112. ICCPR, *supra* note 90, art. 19(3).

113. *Id.*

114. *See* GC 34, *supra* note 94, ¶ 22 ("Restrictions are not allowed on grounds not specified in [ICCPR Article 19(3)], even if such grounds would justify restrictions to other rights protected in the Covenant.").

115. *See id.* ¶ 30 ("It is not compatible with [ICCPR Article 3], for instance, to invoke [national security or sedition] laws to suppress or withhold from the public information of legitimate public interest that does not harm national security or to prosecute journalists, researchers, environmental activists, human rights defenders, or others, for having disseminated such information.").

116. ICCPR, *supra* note 90, art. 20, ¶ 1. For an analysis of the scope and interpretation of the ban on war propaganda, see generally Evelyn Mary Aswad, *Propaganda for War & International Human Rights Standards*, 24 CHI. J. INT'L L. 1 (2023).

117. ICCPR, *supra* note 90, art. 20, ¶ 2.

118. *See* Aswad, *supra* note 116, at 6–7, 21–29 (observing that the prohibition on war propaganda has been generally overlooked by the international community and developing a framework for a principled limitation of the scope of this provision).

119. Frank La Rue (Special Rapporteur), *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN GAOR, 67th Sess., Agenda Item 70(b), UN Doc. A/67/357, ¶¶ 44(c), 45(e), (Sept. 7, 2012), https://docs.un.org/en/A/67/357.

120. GC 34, *supra* note 94, ¶¶ 50–52. In addition, the Special Rapporteur has clarified that this provision does not entail a heckler's veto. S R 2019 Report, *supra* note 105, ¶ 10 ("There is no 'heckler's veto' in international human rights law."). In other words, if someone reacts to offensive expression with violence or other illicit acts, that is not a justification to suppress the offensive speech. Article 20(2) covers situations in which the speaker is calling for harm against a targeted group and the risk of that harm is imminent and likely. *See* Evelyn Aswad, *To Ban or Not to Ban Blasphemous Videos*, 44 GEO. J. INT'L L. 1313, 1322 (2013).

mandatory prohibition on racist hate speech.[121] Article 4 requires the criminalization of "all dissemination of ideas based on racial superiority or hatred,[and] incitement to racial discrimination."[122] The CERD's monitoring body of experts endorsed limiting the potential breadth of this ban by emphasizing that ICCPR Article 19(3)'s legality and necessity tests be met for any bans imposed under this provision.[123] This means any such ban must not be unduly vague; the ban must also be the least intrusive means to achieve legitimate objectives. In addition, UN experts have noted that imposing this ban means assessing the intention of the speaker to incite harm as well as the imminence and likelihood of such harm.[124]

For U.S. social media companies, a natural question may be how the First Amendment and global free speech standards compare, although platform speech rules have not been aligned with First Amendment standards for some time.[125] A prior scholarly examination of the differences between global and U.S. First Amendment free speech protections revealed four key similarities. First, both bodies of law start with a presumption in favor of speech with the burden on the government to show restrictions are justified.[126] Second, neither permit unduly vague bans on speech and this requirement has been interpreted vigorously.[127] Third, both require showing an important public interest reason to burden speech.[128] Fourth, both require narrowly tailoring limitations on speech, with First Amendment jurisprudence using a "least restrictive means test" and the global standard deploying a "least intrusive means" test.[129] Although the

121. International Convention on the Elimination of All Forms of Racial Discrimination art. 4, *opened for signature* Mar. 7, 1966, 660 U.N.T.S. 195, 218 (entered into force Jan. 4, 1969) [hereinafter CERD]; *International Convention on the Elimination of Racial Discrimination*, UNITED NATIONS TREATY COLLECTION [hereinafter *UN Treaty Collection: CERD*], https://bit.ly/3dQPVCh (last visited Sept. 5, 2025); *see also* Convention on the Prevention and Punishment of the Crime of Genocide, art. 3(c), *opened for signature* Dec. 9, 1948, 78 U.N.T.S. 277, 280 (entered into force Jan. 12, 1951) (banning incitement to genocide).

122. *CERD*, *supra* note 121, art. 4. For relevant reservations lodged by the United States when joining this treaty, see *infra* note 130.

123. Comm. on the Elimination of Racial Discrimination, General Recommendation No. 35, ¶ 12, UN Doc. CERD/C/GC/35 (Sept. 26, 2013) [hereinafter GR 35] ("The application of criminal sanctions should be governed by principles of legality, proportionality and necessity.").

124. *Id.* at 5 (observing that "the intention of the speaker, and the imminent risk or likelihood that the conduct desired or intended by the speaker will result from the speech in question" should be considered when determining if speech that disseminates racism or incites harms rises to the level of an offense under CERD Article 4). Patrick Thornberry, a former CERD Committee Member, has noted that this General Recommendation "decisively rejects any suggestion of a 'strict liability' approach to dissemination and incitement . . . [by linking] them with principles of criminal law on mental elements in crime." Patrick Thornberry, *International Convention on the Elimination of All Forms of Racial Discrimination: The Prohibition of 'Racist Hate Speech,' in* THE UNITED NATIONS AND FREEDOM OF EXPRESSION AND INFORMATION: CRITICAL PERSPECTIVES 121, 131 (Cambridge Univ. Press 2015). He states "the scope of potential hate speech prosecutions" is also narrowed by the inclusion of an imminence requirement. *Id.* at 132.

125. *See* Aswad, *supra* note 105, at 60.

126. Evelyn Mary Aswad, *Taking Exception to Assessments of American Exceptionalism: Why the United States Isn't Such an Outlier on Free Speech*, 126 DICK. L. REV. 69, 126 (2021).

127. *Id.* at 127.

128. *Id.*

129. *Id.* at 128.

First Amendment does not contain any mandatory prohibitions on harmful speech like the UN human rights framework,[130] this distinction appears to have been narrowed significantly during the last fifteen years in UN expert interpretations that recommend imposing such bans only where there is likely and imminent harm.[131] While this study did not identify a complete convergence between the two bodies of law, it found some important similarities and highlighted that the global speech standard provides a thorough and principled check on speech regulators.[132]

B. UNGPs

The UNGPs set forth the global corporate responsibility framework with regard to business and human rights matters.[133] This voluntary framework, which companies can pledge to implement, has been endorsed by governments, including the U.S. government.[134] Principle 11 of the UNGPs calls on companies to "respect" human rights, which is defined as seeking to avoid infringing on rights as well as addressing harms.[135] This corporate responsibility exists even when companies are operating in states without particular human rights obligations or states that are hostile to human rights.[136] When faced with "conflicting requirements" between local law and global standards, the UNGPs do not require or authorize companies to violate local law, but they *do* call on companies to strive to respect human rights standards to the greatest extent possible.[137]

The UNGPs delineate a variety of steps that companies should proactively take to operationalize their responsibility and avoid infringing on human

130. Because the First Amendment does not contain mandatory speech bans, the United States lodged reservations to ICCPR Article 20 and CERD Article 4 to the extent such obligations require action by the U.S. government that is inconsistent with the U.S. Constitution. *See UN Treaty Collection: ICCPR*, *supra* note 91 ("[A]rticle 20 does not authorize or require legislation or other action by the United States that would restrict the right of free speech and association protected by the Constitution and laws of the United States"); *UN Treaty Collection: CERD*, *supra* note 121 ("The Constitution of the United States contains provisions for the protection of individual rights, such as the right of free speech, and nothing in the Convention shall be deemed to require or to authorize legislation or other action by the United States of America incompatible with the provisions of the Constitution of the United States of America").

131. Aswad, *supra* note 126, at 129.

132. *Id.* It should be noted that a former American Civil Liberties Union (ACLU) President has advocated for global platforms to align their speech codes with the UN standard on freedom of expression, given UN experts' recent speech friendly interpretations thereof. Nadine Strossen, *United Nations Free Speech Standards as the Global Benchmark for Online Platforms' Hate Speech Policies*, 29 MICH. ST. INT'L L. REV. 307, 361 (2021) ("A promising approach, consistent with the UN Guiding Principles on Business and Human Rights, is for the Platforms to honor and advance the key speech-protective aspects of UN free speech law.").

133. UNGPs, *supra* note 13.

134. Indeed, the U.S. government has called on companies to treat the UNGPs as a "floor" rather than a "ceiling" in their operations. *See, e.g.*, U.S. DEP'T OF STATE, UNITED STATES GOVERNMENT NATIONAL ACTION PLAN ON RESPONSIBLE BUSINESS CONDUCT 7 (2024).

135. UNGPs, *supra* note 13, at Principle 11.

136. *Id.* at Commentary to Principle 11.

137. *Id.* at Principle 23(b).

rights.[138] For example, they should have a meaningful human rights policy[139] as well as appropriately staff, resource, and mainstream human rights issues throughout business operations.[140] Moreover, the UNGPs call on companies to engage in "human rights due diligence," including thoughtfully assessing their most salient human rights risks and developing ways to minimize those risks.[141] The UNGPs call on companies to be transparent to the extent possible about their efforts to address human rights risks.[142]

Of note, the UNGPs specifically define the concept of human rights according to the UN's global standards. Principle 12 explains that "internationally recognized human rights" are to be understood to include the International Bill of Human Rights and the International Labour Organization's Declaration on Fundamental Principles and Rights at Work.[143] The commentary on Principle 12 explains that companies may need to consult additional UN instruments on particular human rights topics in conducting their human rights due diligence.[144] Given that regional human rights systems have at times provided different or fewer protections than UN human rights standards,[145] it is not surprising the UN chose to peg the UNGPs' substantive human rights standards to its own global instruments and human rights machinery to maximize coherence and consistency in operationalizing the framework.[146]

When considering how the largest social media platforms would implement the UNGPs in curating content on their platforms, potential infringements on freedom of expression emerge as a salient human rights risk.

---

138. For an in-depth list of steps companies should take, see *id.* at Principles 11–24.

139. *Id.* at Principle 16.

140. *Id.* at Principle 19.

141. *Id.* at Principles 17–18.

142. *Id.* at Principle 21.

143. *Id.* at Principle 12. *See* ICCPR, *supra* note 90 (defining the UN instruments that comprise the International Bill of Human Rights).

144. UNGPs, *supra* note 13, at Principle 12. For example, the Commentary notes it may be relevant to consult the UN Declaration on the Rights of Indigenous Peoples and the UN Convention on the Elimination of Discrimination Against Women. *Id.* at Commentary to Principle 12.

145. For example, the high bar set by UN experts in interpreting ICCPR Article 19(3)'s tripartite test for speech restrictions stands in contrast to the European Court of Human Rights' (ECtHR) method of interpreting similar conditions in Europe's regional treaty. Evelyn Aswad & David Kaye, *Convergence & Conflict: Reflections on Global and Regional Human Rights Standards on Hate Speech*, 20 NW. J. HUM. RTS. 165, 206–09 (2022). The ECtHR has not been rigorous in applying the vagueness test, allows for additional governmental aims to qualify as legitimate, and only applies a "proportionality" and not a "least intrusive means" test. *Id.* In addition, the ECtHR frequently refuses to hear appeals for speech it considers particularly offensive whereas the UN system applies the tripartite test to assess all speech bans. *Id.* at 206. The ECtHR also applies a "margin of appreciation" doctrine, which entails significant deference to governments on speech restrictions, whereas the UN system rejects this doctrine. *Id.* at 206–07.

146. It should be noted that as a matter of basic treaty law, "[r]egional human rights norms cannot, in any event, be invoked to justify departure from international human rights protections." SR 2019 Report, *supra* note 105, ¶ 26. In other words, regional instruments that provide fewer rights protections than UN instruments do not justify states not implementing their obligations under UN treaties and can be a form of improperly invoking cultural relativism to try to "legitimize the failure to implement global minimum standards." Aswad & Kaye, *supra* note 145, at 171.

This risk presents itself in at least two ways. First, the largest social media companies have their own rules for permissible platform speech.[147] When those rules are vague or overly restrictive, they can infringe on individuals' enjoyment of the right to freedom of expression.[148] Second, governments often pressure these platforms to interpret or change their own rules in a particular way, or implement laws that do not adhere to global free speech standards.[149] Corporate complicity in such situations can also undermine individuals' enjoyment of freedom of expression. The remainder of Part II(B) describes how the largest social media companies could implement the UNGPs in their own content moderation and when dealing with governments as well as how this would help to realize some of the societal aims animating the *Moody* and *Murthy* cases.[150]

1. *Content Moderation*

In his 2018 and 2019 reports, the UN Special Rapporteur on Freedom of Opinion and Expression proposed that social media companies should fulfill their UNGPs responsibility in content moderation by aligning their platform speech codes with ICCPR speech standards.[151] In practice, this would mean the largest platforms need to demonstrate their speech rules meet ICCPR Article 19's tripartite test of legality, necessity, and legitimacy.[152] With respect to the legality test, the Special Rapporteur noted that a variety of platform rules were

147. *See, e.g.*, *Community Standards*, META, https://transparency.meta.com/policies/community-standards (last visited May 31, 2025) (listing Meta's Community Standards outlining "what is and isn't allowed on Facebook, Instagram, Messenger and Threads" within larger categories); *Our Policies*, YOUTUBE, https://www.youtube.com/howyoutubeworks/our-policies/ (last visited May 31, 2025) (describing YouTube's speech rules); *Community Guidelines Overview*, TIKTOK, https://www.tiktok.com/community-guidelines/en (last visited May 31, 2025) ("The guidelines apply to everyone and everything on our platform. They include rules for what is allowed on TikTok.").

148. *See, e.g.*, SR 2018 Report, *supra* note 96, ¶¶ 26–28, 44, 77 (observing that company speech codes should not be vague or limit speech in ways that would undermine the necessity and proportionality tests in ICCPR Article 19(3)).

149. *See, e.g.*, *Turkiye: Big Tech Should Protect Free Speech and Resist State Censorship*, FREEDOM HOUSE (Mar. 4, 2024), https://freedomhouse.org/article/turkiye-big-tech-should-protect-free-speech-and-resist-state-censorship (calling on companies to resist the Turkish government's "formal and informal pressure targeting expression protected under international human rights law").

150. It should be noted that the UNGPs also mandate the creation of accountability mechanisms to address infringements on human rights. UNGPs, *supra* note 13, at Principles 22, 28–31. While such mechanisms are important to ensuring companies are acting consistently with their pledges to implement the UNGPs, the creation and functioning of such mechanisms are beyond the scope of this Essay.

151. SR 2018 Report, *supra* note 96, ¶¶ 41–43 (calling on companies to use human rights law free speech principles not only to push back on illicit governmental pressure but also in their content moderation); SR 2019 Report, *supra* note 105, ¶ 41 ("Companies do not have the obligations of Governments, but their impact is of a sort that requires them to assess the same kind of questions about protecting their users' right to freedom of expression.").

152. SR 2018 Report, *supra* note 96, ¶ 45 ("Companies should incorporate directly into their terms of service and 'community standards' relevant principles of human rights law that ensure content-related actions will be guided by the same standards of legality, necessity and legitimacy that bind State regulation of expression.").

inappropriately vague.[153] The Special Rapporteur called on companies to seek guidance from Article 19(3)'s legitimacy test in assessing public interest aims for restricting speech.[154] The Special Rapporteur also urged social media companies to apply the ICCPR Article 19 necessity test's least intrusive means and proportionality analysis in their content moderation,[155] and emphasized human rights law principles require nondiscrimination in the application of speech rules.[156] In addition, the Special Rapporteur called for platforms to publicly justify any exceptional departures from the norm of applying human rights principles–adding transparency and a measure of flexibility to content moderation.[157]

With regard to the necessity test, the Special Rapporteur noted that social media companies have a wide variety of digital tools to deploy and should publicly justify how they select the one that least burdens speech.[158] The Special Rapporteur endorsed a three-pronged test for assessing if platforms use the least intrusive digital tool to achieve a legitimate aim (e.g., protection of the rights of others).[159] First, platforms should assess if they have the tools to address a particular aim without burdening the speech of users (e.g., promoting digital literacy).[160] Second, if such means are unavailable or insufficient, then they should select the least intrusive digital tool to address the legitimate aim.[161] Third, platforms should monitor if the selected tool is effective in achieving the specified aim. If it is not effective, the burden on speech is not justified.[162]

The application of global human rights law principles to the largest social media platforms' content moderation would bring several benefits. While scholars have noted that *Moody*'s application of traditional First Amendment

153. *Id.* ¶¶ 26–28, 46 (criticizing platform rules for vagueness issues with respect to their hate speech and harassment rules); SR 2019 Report, *supra* note 105, at ¶¶ 46, 49 (criticizing the vagueness of company standards on various grounds, including not defining incitement).

154. SR 2019 Report, *supra* note 105, ¶ 47(b). The Special Rapporteur did note that certain aims (such as national security) fall more within the expertise of a government than a social media company. *Id.*

155. SR 2018 Report, *supra* note 96, ¶¶ 28, 44, 47; SR 2019 Report, *supra* note 105, ¶¶ 51–52.

156. SR 2018 Report, *supra* note 96, ¶ 48.

157. SR 2019 Report, *supra* note 105, ¶ 48.

158. *Id.* ¶ 51. The Special Rapporteur noted that social media companies can:

> [D]elete content, restrict its virality, label its origin, suspend the relevant user, suspend the organization sponsoring the content, develop ratings to highlight a person's use of prohibited content, temporarily restrict content while a team is conducting a review, preclude users from monetizing their content, create friction in the sharing of content, affix warnings and labels to content, provide individuals with greater capacity to block other users, minimize the amplification of the content, interfere with bots and coordinated online mob behaviour, adopt geolocated restrictions and even promote counter-messaging. Not all of these tools are appropriate in every circumstance, and they may require limitations themselves, but they show the range of options short of deletion that may be available to companies in given situations. In other words, just as States should evaluate whether a limitation on speech is the least restrictive approach, so too should companies carry out this kind of evaluation. *Id.*

159. *Id.* ¶ 51.

160. *Id.* ¶ 52.

161. *Id.*

162. *Id.*

jurisprudence in the social media context resulted in championing corporate rather than individuals' free speech rights,[163] the UNGPs approach to content moderation would restore the focus on individuals and their rights.

Pursuant to these principles, platforms would need to improve the clarity of their speech rules to give better notice to users and constrain the discretion of the rules' implementers; this would help curb arbitrary and discriminatory decisions and promote consistent application of the rules.[164] The necessity test would bring a strong measure of discipline and transparency in requiring platforms to justify their decisions to deploy the least intrusive digital tools to address legitimate aims.[165] Implementation of the nondiscrimination principle in speech regulation would be an important aspect of running a fair digital town square.[166] The ICCPR's call to address likely and imminent harmful acts (subject to meeting Article 19's tripartite test) would also mean social media companies need to proactively address serious harms.[167] The application of such human rights principles would not require or result in a harmonization of speech rules across the largest social media platforms (which is a good thing from a free expression viewpoint). Rather, it would require that platforms apply the ICCPR's tripartite test of legality, legitimacy, and necessity to their own rules and enforcement, and mainstream the nondiscrimination principle and the commitment to addressing likely and imminent harmful acts.

2. *Laws and Governmental Pressure*

In addition to aligning their own speech codes with global standards, the largest social media companies that adopt the UNGPs would need to avoid infringing on individuals' free speech by resisting, to the extent possible, implementing laws and governmental jawboning requests that limit expression in a manner inconsistent with the standards discussed in Part II.A. Rather than placing the onus on individuals to show jawboning resulted in particular content moderation decisions (as in *Murthy*), under the UNGPs approach, the burden would be on the platforms to proactively resist such illicit pressure or regulation. An international multistakeholder initiative, the Global Network Initiative (GNI), provides a way for companies to operationalize this pledge.[168] The initiative is comprised of information and communications companies, civil society organizations, academics, and investors.[169] GNI companies pledge to

163. *See* Balkin, *supra* note 58.

164. *See supra* notes 98–103 (describing the ICCPR's norm against vague speech prohibitions).

165. *See supra* notes 104–111 (describing how the ICCPR's least intrusive means test works).

166. *See* ICCPR, *supra* note 93 (noting the ICCPR's prohibition on discrimination).

167. *See supra* notes 116–124 (examining the UN human rights machinery's interpretations narrowing the scope of mandatory speech bans in foundational UN human rights treaties).

168. GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/ (last visited May 31, 2025). This initiative also covers privacy issues. *Id.*

169. *Members*, GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/members/ (last visited May 31, 2025). Participating companies include Google, Microsoft, Verizon, and Meta. *Id.* Civil society

identify risks of infringing on freedom of expression and resist, to the extent possible without violating local law, implementing speech bans or succumbing to illicit pressure to ban or burden speech in ways that contradict ICCPR Article 19 standards.[170] The means of seeking to avoid and mitigate human rights risks include initiating lawsuits in local courts to compel compliance with local laws and/or international human rights obligations, narrowly construing governmental requests, and seeking the aid of other governments as well as the UN's human rights machinery.[171] Participating companies submit to periodic assessments to provide a measure of oversight on implementation.[172] While a global social media company pledging to implement the UNGPs would not necessarily need to join the GNI, this multistakeholder approach is helpful to identify and resist both government jawboning and laws that do not comport with global free speech standards.

C. NORM-BUILDING

With respect to matters of digital governance, Professor Jonathan Zittrain observed, "we don't know what we want [and] we don't trust anybody to give it to us [and] we want it now."[173] It can certainly feel like online speech governance is trapped in a vicious cycle of competing societal critiques about content moderation followed by illicit governmental attempts to "fix" social media, which only triggers further criticisms. Recent Supreme Court jurisprudence has focused on corporate free speech rights in the social media context rather than individuals' free speech interests. Society is left relying on the ever–evolving speech preferences of corporate actors–such as Silicon Valley elites and advertisers–who are in turn subject to various ongoing commercial, governmental, and societal pressures. This is an unstable and unsatisfying situation for the protection of individuals' free expression interests on global platforms.

Rather than missing out on a norm-building opportunity at this critical juncture, the largest social media companies should adopt and implement the UNGPs. Platform adoption of the UNGPs would provide a means for addressing the concerns about users' speech interests that animated the regulation in *Moody* and the litigation in *Murthy*. First, adhering to the UNGPs means platforms would use a principled free expression standard that places the burden on

members include prominent NGOs such as Human Rights Watch, Freedom House, and the Center for Democracy & Technology. *Id.*

170. *The GNI Principles*, GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/gni-principles/ (last visited May 31, 2025).

171. *Implementation Guidelines*, GLOBAL NETWORK INITIATIVE, https://globalnet workinitiative.org/implementation-guidelines/ (last visited May 31, 2025).

172. *Company Assessments*, GLOBAL NETWORK INITIATIVE, https://globalnetwork initiative.org/company-assessments/ (last visited May 31, 2025).

173. Samantha Laine Perfas, *How to Make Social Media, Online Life Less of a 'Dumpster Fire'*, HARV. GAZETTE (Sept. 17, 2024), https://news.harvard.edu/gazette/story/2024/09/how-to-make-social-media-online-life-less-of-dumpster-fire/.

companies to demonstrate they have met rigorous tests in respecting users' free speech rights in platform speech codes.[174] Second, the pledge to implement the UNGPs would require platforms to be proactive and push back on laws and illicit pressure that are inconsistent with global free speech norms.[175] In essence, this approach would return the focus of content moderation to the individuals' speech interests.

But the road to platform adoption of the UNGPs contains roadblocks. Chief among them is the voluntary nature of accepting this global standard. While voluntary adoption helps to avoid a First Amendment problem, it also means platforms may lack the incentive to commit to a free expression framework in their content moderation operations. Societal demand from a wide range of stakeholders–including consumers, civil society, platform employees, advertisers–is needed to encourage platforms to implement these corporate responsibility principles. While adoption of the UNGPs will not necessarily result in universal "speech nirvana" as there will always be disputes about exactly where to draw the line on speech, it provides a principled, user-focused framework for addressing the issues that *Moody* and *Murthy* have shown can be difficult to resolve otherwise.

CONCLUSION

Together, *Moody* and *Murthy* raise questions about how society can promote principled platform curation of speech as well as resilience in the face of jawboning. *Moody* blocks governmental attempts to legislate "speech nirvana," including prohibiting viewpoint discrimination, in key aspects of how the largest social media companies curate content. *Murthy* appears to erect substantial barriers for jawboning claims. Rather than adopting approaches that contravene First Amendment protections against governmental regulation of speech, a potential way forward in seeking to promote respect for individuals' freedom of expression on such platforms is for companies to adopt global corporate responsibility standards. This approach would require platforms to use a principled framework in not only curating user-generated speech but also resisting illicit governmental laws and pressure. Platform adoption of global standards is likely to be achieved only if companies are encouraged to do so by a range of stakeholders. Until then, we remain stuck in a cycle of criticizing platform content moderation followed by illicit governmental attempts to "fix it" rather than developing the foundation for important norm-building to protect individuals' free expression on platforms in the digital age.

174. *See* Part.II.B.1.
175. *See* Part.II.B.2.

***