INTERNATIONAL JOURNAL OF HUMAN RIGHTS LAW REVIEW
An International Open Access Double Blind Peer Reviewed, Referred Journal

Volume 5 | Issue 6 | 2026 Art. 13

Conscripting the Gatekeepers: State Co-Regulation, Intermediary Liability, and the Politics of Digital Platform Governance

Deepika Sahu
Law Student,
3rd Year, B.A. LL.B. (Hons.),
Central Law College, Lucknow

Recommended Citation
Deepika Sahu, *Conscripting the Gatekeepers: State Co-Regulation, Intermediary Liability, and the Politics of Digital Platform Governance,* 5 IJHRLR 130-140 (2026).
Available at www.ijhrlr.in/current-issues/.

This Article is brought to you for free and open access by the International Journal of Human Rights Law Review by an authorized Lex Assisto & Co. administrator.
For more information,
please contact submission.ijhrlr@gmail.com

Conscripting the Gatekeepers: State Co-Regulation, Intermediary Liability, and the Politics of Digital Platform Governance

Book Title	**The Politics of Platform Regulation: How Governments Shape Online Content Moderation**
Author	Robert Gorwa
Publisher	Oxford University Press
Place of Publication	Oxford, United Kingdom
Year of Publication	2024
Edition & Extent	First Edition, xiv + 312 pp.
ISBN	978-0-19-887462-1 (Hardcover) / 978-0-19-887464-5 (Electronic)
Price & Access	£83.00 (Hardcover) / Open Access (Electronic / Oxford Scholarship Online)

1. INTRODUCTION AND CONTEXTUAL BACKGROUND

For nearly a quarter of a century, digital communications governance was arrested by the ideology of cyber-exceptionalism.[1] Early internet theorists posited cyberspace as an extraterritorial realm fundamentally resistant to state coercion, establishing an era of uninhibited 'governance by platforms.'[2] Under this private ordering regime, transnational tech giants exercised quasi-sovereign authority, policing public discourse through non-negotiable Terms of Service, private community standards, and opaque algorithmic sorting systems.[3] Shielded by statutory safe harbors—chiefly Section 230 of the United States Communications Decency Act of 1996 and Article 14 of the European Union's E-Commerce Directive of 2000—commercial gatekeepers claimed the status of neutral technological conduits while constructing expansive socio-technical bureaucracies to govern billions of users.[4]

Over the past decade, however, the global political consensus underwent a convulsive realignment. Catalyzed by electoral interference, algorithmic polarization, viral disinformation, and live-streamed terror, the myth of benign self-regulation collapsed into an international 'techlash.' In The Politics of Platform Regulation: How Governments Shape Online Content Moderation, Robert Gorwa provides an incisive institutional political economy analysis of this structural transformation: the shift from 'governance by platforms' to the state-led 'governance of platforms.'[5] Moving beyond technological determinism and formalistic legal commentary, Gorwa establishes that platform regulation represents

the normalization of digital communications into the contested arena of ordinary domestic politics, electoral calculation, and institutional bargaining.

2. CENTRAL THESIS AND INTELLECTUAL FRAMEWORK

Gorwa's central thesis posits that platform regulation is an institutional political economy problem wherein sovereign states shape the architecture, policies, and practices of digital intermediaries. At the foundational level, Gorwa anchors his analysis upon the Regulator–Intermediary–Target (R-I-T) governance model formulated by Kenneth Abbott, David Levi-Faur, and Duncan Snidal.[6] Confronting an acute enforcement deficit—lacking the technical infrastructure and administrative bandwidth to monitor billions of user interactions directly—the state (Regulator) conscripts commercial platforms (Intermediaries) as regulatory proxies to govern the behavior, speech, and transactions of end-users and creators (Targets).

Gorwa advances beyond the descriptive R-I-T model by developing an original Demand-and-Supply Framework. Regulatory intervention is determined by the intersection of political appetite and administrative capacity: (1) Political Will (Demand) is driven by exogenous salience shocks (e.g., refugee-related hate speech in Germany, Christchurch terror) and champion policy entrepreneurs, while depressed by tech lobbying, trade pressure, and free speech advocacy; and (2) Power to Intervene (Supply) is bounded by domestic market size (sovereign bargaining leverage), state technical resources, institutional veto points (judicial review, legislative gridlock), and normative legal traditions.

When demand overcomes institutional constraints, governments deploy a tripartite menu of strategies: Contest (binding statutory coercion backed by punitive fines and executive liability, such as Germany's NetzDG and Australia's AVM Act); Collaborate (voluntary co-regulation and multi-stakeholder codes in the 'shadow of hierarchy,' such as the Christchurch Call); and Convince (opaque informal pressure and 'jawboning'). Crucially, Gorwa shows that modern states have converted statutory safe harbors from absolute liability shields into conditional bargaining chips, conditioning legal immunity on platforms enforcing intensive, state-mandated due diligence.

3. OVERVIEW OF THE BOOK'S THEMATIC ARCHITECTURE

Gorwa organizes his comparative inquiry across three thematic movements, avoiding descriptive chapter-by-chapter summaries in favor of an institutional diagnosis of regulatory evolution:

The first movement (Foundations of Governance) deconstructs the historical origins of platform self-regulation. Gorwa traces how early copyright disputes (such as the US Digital Millennium Copyright Act of 1998) established the architectural template for automated filtering and notice-and-takedown protocols. This technical apparatus laid the institutional groundwork for states to later conscript platforms into policing broader political speech. Gorwa then reveals that platform content moderation combines low-wage commercial moderation, automated hashing, and opaque corporate rules designed primarily for private risk mitigation.

The second movement (The Western Regulatory Laboratory) subjects the 'Three Cs' typology to qualitative process-tracing across three jurisdictions: Germany shifted from a voluntary 2015 Task Force (Collaborate) to the coercive 2017 Network Enforcement Act (NetzDG, Contest), imposing statutory fines up to €50 million and a 24-hour removal deadline for 'manifestly illegal' speech; Australasia exhibited divergent responses to the 2019 Christchurch massacre, where Australia enacted the punitive Sharing of Abhorrent Violent Material (AVM) Act 2019 (Contest) threatening executive imprisonment, while New Zealand pursued diplomatic co-regulation via the multi-stakeholder Christchurch Call (Collaborate); and the United States witnessed partisan polarization after January 6, prompting Texas (HB 20) and Florida (SB 7072) to enact 'common carrier' must-carry statutes restricting platforms' editorial discretion.

The third movement (The Majority World) scales the model to rising powers—China, Brazil, and India. Gorwa illustrates how these sovereigns assert digital sovereignty against transnational tech dominance, establishing that platform regulation outside the West is driven by post-colonial resistance, electoral sensitivities, and national security imperatives.

4. CRITICAL APPRAISAL AND SCHOLARLY ENGAGEMENT

Gorwa's core intellectual contribution lies in his deconstruction of regulatory intermediation and safe harbor conditionality. Historically, legal scholars treated intermediary safe harbors as static defensive shields designed to foster internet commerce. Gorwa exposes how modern states have weaponized safe harbors into instruments of sovereign coercion, converting immunity into a contingent privilege granted only upon strict adherence to state-mandated due diligence:

1. **Accelerated Compliance Windows**: Intermediaries face compressed removal deadlines—from Germany's 24-hour NetzDG window to India's 36-hour takedown rule under Rule 3(1)(d) of the IT Rules, 2021—forcing platforms to deploy

automated algorithmic filters to adjudicate speech at machine scale.[7]

2. **Personnel Liability**: To eliminate extraterritorial evasions, states mandate the domestic appointment of resident compliance officers, holding local employees personally and criminally liable for corporate non-compliance.

3. **Systemic Platform Responsibility**: As embodied by the EU Digital Services Act (DSA), regulation has transitioned from transactional notice-and-takedown liability toward structural obligations: mandatory systemic risk assessments, algorithmic audits, and verified researcher data access.

However, a rigorous appraisal exposes two unresolved structural pathologies in Gorwa's analysis: First, Perverse Market Concentration: Highly complex, punitive compliance regimes demand vast technical and legal resources. While incumbent tech monopolies (Meta, Alphabet) easily absorb these compliance overheads, emerging competitors and decentralized alternatives cannot. Consequently, state-driven platform regulation perversely entrenches market concentration, raising barriers to entry and consolidating the corporate dominance states seek to discipline. Second, The Reality of 'Collateral Censorship': When sovereign states impose rigid removal deadlines backed by severe corporate fines or executive imprisonment, private platforms face an asymmetric risk calculus. To insulate themselves from liability, platforms systematically over-moderate, pre-emptively suppressing legitimate, constitutionally protected expression. Co-regulation thus degenerates into delegated, privatized censorship, outsourcing state police powers to unaccountable corporate algorithms.

5. ENGAGEMENT WITH EXTERNAL SOURCES AND COMPARATIVE JURISPRUDENCE

Gorwa's framework enters into productive dialogue with foundational cyberlaw and constitutional theory. His application of the R-I-T model translates the administrative sociology of Abbott, Levi-Faur, and Snidal into the digital communications sphere.[8] Gorwa's structural critique intersects directly with Jack Balkin's 'Free Speech Triangle,' which demonstrated that digital speech is governed not by a dyadic relation between citizen and state, but by a triadic dynamic among the sovereign state, transnational private gatekeepers, and speaking citizens.[9]

Furthermore, Gorwa's institutional focus complements Julie E. Cohen's Between Truth and Power, which theorizes how informational capitalism reconfigures legal institutions to entrench platform power.[10]

While Cohen focuses on corporate legal capture, Gorwa demonstrates how sovereign states counter-mobilize to subordinate private platforms to domestic public law. Similarly, Gorwa's comparative analysis enriches Terry Flew's media-policy approach in Regulating Platforms, which grounds platform governance in public interest broadcasting traditions and co-regulatory design.[11]

In constitutional jurisprudence, Gorwa's analysis illuminates tensions within Indian legal doctrine. In the landmark case of Shreya Singhal v. Union of India, the Supreme Court of India struck down Section 66A of the Information Technology Act, 2000, and read down Section 79(3)(b) to protect intermediary safe harbor immunity.[12] The Court held that an intermediary is only obligated to remove content upon obtaining 'actual knowledge' strictly in the form of a court order or an authorized government directive.[13] The Supreme Court reasoned that compelling private intermediaries to adjudicate the legality of speech upon private complaints would undermine free expression and establish private censorship. Gorwa's analysis reveals how modern executive rulemaking has systematically bypassed this constitutional safeguard, re-establishing non-judicial takedown pressures through administrative due diligence mandates.

6. CONTEMPORARY RELEVANCE: THE INDIAN REGULATORY LANDSCAPE & FINANCIAL CO-REGULATION

The contemporary relevance of Gorwa's thesis is demonstrated in India, which serves as the premier battleground for platform regulation in the Global South. Over recent years, MeitY has operationalized Gorwa's coercive Contest strategy through the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (amended 2022–2023):[14] Rule 3(1)(d) mandates content removal within 36 hours of actual knowledge via court or government notification; Rule 3(2)(b) enforces a 24-hour turnaround for non-consensual sexual imagery; Rule 4(1)(a) holds resident Chief Compliance Officers personally and criminally liable; and Rule 3A introduces executive-appointed Grievance Appellate Committees (GACs) empowered to overturn platform moderation decisions. Under Rule 7, procedural non-observance strips statutory safe harbor under Section 79(1), exposing platforms to criminal prosecution.[15]

Crucially, this co-regulatory architecture extends into the commercial creator economy. In India, retail demat accounts surged from 4.09 crore in March 2020 to over 18.5 crore by early 2025. Unregistered financial influencers ('finfluencers') weaponized algorithmic reach to dispense unauthorized advice and speculative derivatives tips. A landmark SEBI study revealed that over 93% of retail traders in equity derivatives

(Futures & Options) incurred net financial losses between FY22 and FY24, surrendering aggregate losses exceeding ₹1.81 lakh crore (averaging ₹1.2 lakh per trader).

Recognizing that policing millions of decentralized video reels was impossible, SEBI deployed Gorwa's R-I-T model: (1) Conscripting Regulated Intermediaries: Under the SEBI (Intermediaries) (Amendment) Regulations, 2024, SEBI barred registered brokers and mutual funds from commercial ties with unregistered creators;[16] (2) Economic Choke-Point Enforcement: SEBI choked monetization pipelines—banning referral commissions, trailing fees, and sponsorships, while enforcing a three-month historical price lag on educational content;[17] and (3) Delegated Algorithmic Enforcement: SEBI introduced a safe harbor for 'Specified Digital Platforms' (SDPs) maintaining automated AI/ML guardrails against unauthorized advice. This operates alongside ASCI self-regulatory codes and CCPA guidelines mandating upfront commercial disclosures.[18]

7. STRENGTHS AND LIMITATIONS (WHAT REMAINS 'UNSAID')

Robert Gorwa's volume possesses significant scholarly strengths: methodological rigor, extensive interview archives (77 elite interviews across industry and government), process-tracing precision, and the sustained inclusion of the Majority World in Chapter 8. However, a critical reading reveals three profound thematic and epistemic boundaries:

- ***Critique 1: Epistemic Neglect of the Creator Economy & Algorithmic Monetization***

  Gorwa conceptualizes content moderation as an expressive speech-policing problem rather than an advertising business model problem. By focusing on political harms (hate speech, terror, disinformation), he overlooks commercial deceptive practices, creator monetization, and algorithmic financial fraud. Platforms do not merely host text; their algorithms actively optimize for engagement to maximize ad revenues. Creators peddle speculative advice because affiliate brokerage links, referral payouts, and ad-revenue splits yield massive private returns. Gorwa's reliance on the traditional 'passive conduit' doctrine fails to account for established jurisprudence, such as Christian Louboutin SAS v. Nakul Bajaj, where the Delhi High Court affirmed that an intermediary actively promoting, organizing, or monetizing transactions forfeits Section 79(2)(a) safe harbor and becomes an active participant.[19] Regulating

content solely through deletion clocks treats symptoms; effective co-regulation must target monetization choke-points.

- ***Critique 2: The Flaw of the 'Horizontal Regulator' Assumption***

  Gorwa's R-I-T model presumes a unitary state authority enforcing a single horizontal statute against Big Tech. This creates a 'horizontal illusion' that overlooks inter-agency friction in modern governance. In India, horizontal platform rules framed by MeitY inevitably collide with vertical market protection mandates issued by SEBI and consumer protection bodies (CCPA/ASCI).[20] Under Section 79(3)(b) as interpreted in Shreya Singhal, platforms routinely dismiss informal regulatory notices from market regulators as non-binding, allowing financial scams to proliferate until formal judicial orders are secured. Governing the algorithmic economy demands inter-agency regulatory convergence—modeled after the UK Digital Regulation Cooperation Forum (DRCF, synchronizing the FCA, Ofcom, CMA, and ICO), joint SEC-FTC initiatives in the US, and the pairing of the EU Market Abuse Regulation (MAR) with Articles 34 and 35 of the Digital Services Act (DSA).[21] For India, this requires a formalized SEBI–MeitY Co-Regulatory Protocol.

- ***Critique 3: Democratic Deficits and Outsourced Private Adjudication in the Global South***

  Gorwa portrays co-regulation as a functional administrative compromise between state power and corporate technical capacity. In doing so, he understates the severe constitutional perils and democratic deficits created by coercive co-regulation in developing democracies. Compressed compliance deadlines (such as India's 36-hour takedown rule under Rule 3(1)(d)) backed by safe harbor revocation induce systemic collateral censorship, forcing risk-averse platforms to over-moderate speech to avoid legal penalties. Furthermore, where independent judicial oversight is bypassed—exemplified by executive-staffed Grievance Appellate Committees under Rule 3A of the IT Rules—co-regulation becomes an instrument of executive coercion. It privatizes the judicial function, subverting the constitutional safeguards articulated in Shreya Singhal and subordinating digital human rights to administrative expediency.

8. OVERALL SCHOLARLY CONTRIBUTION

The Politics of Platform Regulation represents a major intellectual milestone in internet governance studies, regulatory theory, and administrative law. Robert Gorwa performs an indispensable service by

dismantling the lingering assumptions of cyber-libertarianism and demonstrating that platform regulation follows recognizable patterns of political mobilization, salience shocks, and bureaucratic bargaining. His conceptual taxonomy—operationalized through the Demand-and-Supply model and the Contest, Collaborate, Convince spectrum—provides legal scholars and policymakers with an enduring analytical grammar. Gorwa convincingly proves that while digital platforms govern through code and algorithms, they remain fundamentally vulnerable to the sovereign power of the administrative state.

9. CONCLUSION

Robert Gorwa's volume succeeds as a definitive institutional post-mortem of platform self-regulation and a rigorous diagnostic guide to the rise of state-driven platform governance. Its insights into regulatory intermediation, safe harbor conditionality, and the political economy of content moderation are essential reading for legal scholars, human rights practitioners, and regulatory authorities navigating the intersections of technology and public power.

Yet, as digital platforms evolve from expressive forums into expansive financial marketplaces and creator economies, the frontiers of platform governance have outpaced Gorwa's political speech paradigm. To advance this field, contemporary legal scholarship must confront two vital research questions on the future of platform co-regulation:

1. **Jurisprudential & Regulatory Convergence**: How can administrative law reconcile jurisdictional friction between horizontal digital intermediary rules (e.g., MeitY's IT Rules) and vertical sectoral market regulators (e.g., SEBI) to effectively govern algorithmic monetization without fragmenting statutory safe harbor protections?

2. **Digital Constitutionalism & Private Adjudication**: Under what constitutional standards can democratic legal systems mandate automated algorithmic due diligence and executive appellate oversight on private intermediaries without eroding independent judicial review and inducing systemic collateral censorship?

REFERENCES

1. John Perry Barlow, A Declaration of the Independence of Cyberspace (Electronic Frontier Foundation, San Francisco, 1996).

2. Robert Gorwa, "What is Platform Governance?" 23(6) Information, Communication & Society 854 (2019).

3. Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media 24–45 (Yale University Press, New Haven, 2018).

4. Communications Decency Act of 1996, 47 U.S.C. § 230; Directive 2000/31/EC (E-Commerce Directive), art. 14.

5. Robert Gorwa, The Politics of Platform Regulation: How Governments Shape Online Content Moderation 1–35 (Oxford University Press, Oxford, 2024).

6. Kenneth W. Abbott, David Levi-Faur and Duncan Snidal, "The Theory and Practice of Regulatory Intermediaries" 670 The ANNALS of the American Academy of Political and Social Science 14–35 (2017).

7. Netzwerkdurchsetzungsgesetz [NetzDG], Sept. 1, 2017, BGBl. I at 3352, s. 3; Information Technology Rules, 2021, r. 3(1)(d).

8. Christopher T. Marsden, Internet Co-Regulation: European Law, Hybrid Governance and Towards a More Holistic Approach 45–68 (Cambridge University Press, Cambridge, 2011).

9. Jack M. Balkin, "Free Speech is a Triangle" 118 Columbia Law Review 2011 (2018).

10. Julie E. Cohen, Between Truth and Power: The Legal Constructions of Informational Capitalism 75–112 (Oxford University Press, New York, 2019).

11. Terry Flew, Regulating Platforms 32–58 (Polity Press, Cambridge, 2021).

12. The Information Technology Act, 2000 (Act 21 of 2000), ss. 66A, 79.

13. *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.

14. Ministry of Electronics and Information Technology, Notification G.S.R. 139(E), Feb. 25, 2021, as amended by Notification G.S.R. 794(E), Oct. 28, 2022.

15. The Information Technology Act, 2000 (Act 21 of 2000), s. 79; Information Technology Rules, 2021, r. 7.

16. Securities and Exchange Board of India (Intermediaries) (Amendment) Regulations, 2024, Notification No. SEBI/LAD-NRO/GN/2024/193, Aug. 27, 2024.

17. Securities and Exchange Board of India, Consultation Paper on Association of SEBI Registered Intermediaries with Unregistered Entities (SEBI, Mumbai, Aug. 25, 2023); SEBI, Study on Analysis of Profit and Loss of Individual Traders in Equity Derivatives (SEBI, Mumbai, Sept. 2024).

18. ASCI, The Code for Self-Regulation of Advertising Content in India (ASCI, Mumbai, updated 2023); Central Consumer Protection Authority, Guidelines for Prevention of Misleading Advertisements, Notification F. No. J-25/15/2020-CCPA, June 9, 2022.

19. *Christian Louboutin SAS v. Nakul Bajaj*, 2018 SCC OnLine Del 12215.

20. The Securities and Exchange Board of India Act, 1992 (Act 15 of 1992), ss. 11, 11B, 12A; The Consumer Protection Act, 2019 (Act 35 of 2019).

21. Digital Regulation Cooperation Forum, Workplan 2023–2024 (DRCF, London, 2023); Regulation (EU) 2022/2065 (Digital Services Act), arts. 34, 35.