THE GATEKEEPER'S GAMBIT: NAVIGATING FREE SPEECH AND REGULATION IN INDIA'S DIGITAL DOMAIN

Paransh M. Desai*

ABSTRACT

This article provides a comprehensive analysis of the evolution of intermediary liability law in India, charting its trajectory from a judicially-protected "safe harbour" framework to a prescriptive, executive-driven regulatory regime. Initially, Section 79 of the Information Technology Act, 2000, granted broad immunity to online platforms for third-party content. This principle was constitutionally fortified by the Supreme Court's landmark 2015 judgment in Shreya Singhal v. Union of India, which established that an intermediary's obligation to remove content was triggered only by a formal court order or government notification, thereby safeguarding against a "chilling effect" on free speech. This article argues that the subsequent notification of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, represents a fundamental paradigm shift. The 2021 Rules dismantle the previous safe harbour by imposing onerous due diligence obligations, mandating the traceability of the "first originator" of information, a measure that threatens end-to-end encryption and the right to privacy and establishing a government-controlled oversight mechanism that circumvents judicial review. By examining this transition, the paper contends that intermediaries have been effectively recast from neutral conduits into frontline enforcers of state policy, intensifying the conflict between legitimate regulation and state-sponsored censorship. The analysis further contextualises India's unique path by comparing it with international models, including the broad immunity of Section 230 in the United States and the co-regulatory approach of the European Union's Digital Services Act, ultimately interrogating the future of free expression in one of the world's largest and most critical digital ecosystems.

*BBA LLB, SECOND YEAR, AURO UNIVERSITY, SURAT.

Keywords: Intermediary Liability, Information Technology Act, 2000, Safe Harbour Doctrine, Freedom of Speech and Expression.

INTRODUCTION

The modern internet is a space of profound duality. It serves as an unprecedented engine for economic growth, a vibrant public square for democratic discourse, and an essential conduit for information and social connection. Yet, this same digital architecture acts as a vector for societal harms, including the proliferation of hate speech, the rapid spread of misinformation, and the facilitation of criminal activity. At the heart of this dynamic are “intermediaries”, the digital gatekeepers that structure our online experience.

Under India's Information Technology Act, 2000 (IT Act),[1] an intermediary is broadly defined as any entity that, on behalf of another person, "receives, stores or transmits" an electronic record or provides any service with respect to it. This expansive definition encompasses a vast ecosystem of entities, from internet service providers (ISPs) and search engines to e-commerce platforms and the social media giants like X (formerly Twitter), Facebook, and Instagram, that have become central to modern communication. This central role raises a critical legal and philosophical question: to what extent should these platforms be held responsible for the content created and disseminated by their millions of users?

This question of intermediary liability is the legal fulcrum on which the balance between a free, open internet and a safe, regulated digital environment rests. The Indian legal framework governing this liability has undergone a profound transformation. It has shifted from a judicially fortified "safe harbour" model, which prioritised free expression and insulated passive platforms from liability, to a prescriptive, executive-driven regulatory regime that mandates active content moderation. This evolution has effectively recast intermediaries from neutral conduits into frontline enforcers of state policy. This is not merely a technical legal development; it is a direct reflection of the changing relationship between the Indian state, global technology companies, and the citizenry, mirroring a broader global trend of states seeking to assert digital sovereignty in an increasingly interconnected world.

[1] Information Technology Act 2000

FORGING THE SHIELD: THE GENESIS OF 'SAFE HARBOUR' UNDER THE IT ACT, 2000

When the Information Technology Act (IT Act) was enacted in 2000, India's internet landscape was in its infancy. The primary goal of the legislature was to encourage the growth of e-commerce and the digital economy. To achieve this, it was deemed essential to protect service providers from the overwhelming burden of being held liable for the content created by their users. Without such protection, it was feared that nascent internet businesses would be crippled by endless litigation.

The true architect of the 'safe harbour' doctrine in Indian law is Section 79 of the Act, introduced through the 2008 amendments. It granted a broad, almost blanket, immunity to network service providers. This provision was designed to immunise intermediaries from liability for any third-party information, data, or communication link made available or hosted by them.

The immunity, however, was not absolute. Section 79(2)[2] laid down two key conditions for an intermediary to claim this protection. First, its function had to be limited to providing access to a communication system over which information made available by third parties is transmitted or temporarily stored. Second, the intermediary must not have initiated the transmission, selected the receiver of the transmission, or selected or modified the information contained in the transmission. This framework was built on the mere conduit principle, treating intermediaries as neutral pipelines, not as publishers or editors of content.

Crucially, Section 79(3)(b)[3] (before the Shreya Singhal[4] judgment) stated that the safe harbour would be withdrawn if the intermediary, upon having 'actual knowledge' or on being notified by the appropriate Government or its agency, failed to expeditiously remove or disable access to the unlawful material. This 'actual knowledge' clause became the epicentre of legal battles, as its ambiguous wording left platforms in a state of uncertainty. Did a simple user complaint constitute 'actual knowledge'? Did a takedown notice from a private entity suffice? The answers to these questions would shape the very nature of online speech in India.

[2] Information Technology (Amendment) Act 2008, s 79(2)
[3] ibid. s 79(3)(b)
[4] Shreya Singhal vs U.O.I AIR 2015 SUPREME COURT 1523

CRACKS IN THE SHIELD: EARLY JUDICIAL TESTS OF INTERMEDIARY IMMUNITY

Indian courts grappled with the scope and meaning of intermediary liability before the Supreme Court provided definitive clarity. Two cases are particularly instructive in understanding the initial judicial thinking.

The first was the infamous 'MMS clip' case, Avnish Bajaj v. State (NCT of Delhi) (2005),[5] which pre-dated the 2008 amendment. Avnish Bajaj, the CEO of Baazee.com (now eBay India), was arrested after a user listed an obscene video clip for sale on the platform. The Delhi High Court, while acknowledging the practical difficulties for platforms to pre-screen all content, distinguished passive and active involvement. It was observed that while Bajaj could not be held liable for the initial posting, the failure to remove the content after being notified of its existence could potentially attract liability. This case highlighted the judiciary's early attempts to apply traditional principles of criminal liability to the novel context of the internet, emphasising the importance of a reactive, notice-and-takedown mechanism.

A more direct engagement with Section 79 came in the copyright domain with MySpace Inc. v. Super Cassettes Industries Ltd., 2016.[6] Super Cassettes (T-Series) sued MySpace for hosting music and videos that infringed its copyright. The single-judge bench of the Delhi High Court initially held that MySpace had an obligation to proactively filter content, effectively placing the burden of policing on the intermediary. This was a significant blow to the safe harbour principle. However, the Division Bench, on appeal, overturned this decision. It held that the 'actual knowledge' required to revoke safe harbour had to be specific and not general. The court clarified that an intermediary would only lose its immunity if, after receiving specific knowledge of infringing content from the content owner, it failed to take it down. This judgment was a crucial victory for intermediaries, reinforcing that their role was reactive, triggered by specific, credible notices, not proactive and constant monitoring.

THE JUDICIAL ZENITH: SHREYA SINGHAL V. UNION OF INDIA

The landscape of online speech and intermediary liability was irrevocably altered by the Supreme Court's landmark 2015 judgment in Shreya Singhal v. Union of India.[7] While the case

[5] Avnish Bajaj v State (NCT of Delhi) (2005)3COMPLJ364(DEL)
[6] MySpace Inc v Super Cassettes Industries Ltd FAO (OS) 540/2011
[7] Shreya Singhal vs U.O.I AIR 2015 SUPREME COURT 1523

is most famous for striking down the draconian Section 66A of the IT Act[8] as unconstitutional for its vagueness and overbreadth, its interpretation of Section 79 was equally profound. The Court addressed the ambiguity surrounding the 'actual knowledge' clause head-on. It is feared that a broad interpretation would lead to a chilling effect on free speech, where intermediaries would engage in excessive self-censorship to avoid liability.

To prevent this, the Court performed a reading down of the provision. It ruled that the 'actual knowledge' mentioned in Section 79(3)(b) could not be derived from a mere user complaint or a private takedown request. Instead, it had to be knowledge acquired through a formal court order or a notification from a government agency. Justice R.F. Nariman, writing for the bench, delivered a powerful articulation of this principle:

"It is...clear that Section 79(3)(b) has to be read down to mean that the intermediary upon receiving actual knowledge that a court order has been passed asking it to expeditiously remove or disable access to certain material must then fail to expeditiously remove or disable access to that material... Otherwise, it would be very difficult for intermediaries to function."

This interpretation was a constitutional masterstroke. It ring-fenced the safe harbour, protecting intermediaries from arbitrary takedown demands and the pressure to adjudicate the legality of content themselves. The judgment solidified a judicial-led, rights-respecting framework for content moderation. It ensured that any restriction on online speech would have to pass the muster of judicial scrutiny, thereby placing a high premium on due process and the freedom of expression enshrined in Article 19(1)(a)[9] of the Constitution. For a time, Shreya Singhal established a clear, predictable, and constitutionally sound regime for intermediary liability in India.

THE PARADIGM SHIFT: THE IT RULES, 2021

This judicially settled equilibrium was upended in 2021 with the notification of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules (IT Rules, 2021).[10] Framed under Section 87 of the IT Act, these rules introduced a comprehensive and

[8] Information Technology (Amendment) Act 2008, s 66A
[9] The Constitution of India, art 19(1)(a)
[10] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021

far-reaching regulatory regime that significantly diluted, if not entirely dismantled, the protections established by the Shreya Singhal case.[11]

The government justified the rules as necessary to combat the rising tide of fake news, hate speech, and online harm, and to make large technology companies more accountable to Indian law. However, critics argue that the rules go far beyond the Parent Act and grant the executive branch sweeping powers to control online discourse without adequate oversight.

The key and most contentious provisions of the 2021 Rules include-

Dilution of Safe Harbour: The rules state that an intermediary loses its safe harbour protection if it fails to comply with any of the provisions. This makes the safe harbour conditional on adherence to a complex and onerous set of regulations.

Due Diligence and Content Categories: The rules expand the categories of content that intermediaries must proactively block or remove. The list includes content that "threatens the unity, integrity, defence, security or sovereignty of India," a phrase so broad that it could be used to suppress political dissent or critical journalism.

Traceability of the First Originator: For "Significant Social Media Intermediaries" (SSMIs) platforms with over 5 million users, Rule 4(2) mandates the identification of the "first originator" of information within India. This is seen as a direct attack on end-to-end encryption, a technology crucial for user privacy and security, used by platforms like WhatsApp and Signal. This provision directly conflicts with the fundamental right to privacy established in Justice K.S. Puttaswamy (Retd.) v. Union of India.

Proactive Monitoring: The rules require SSMIs to deploy technology-based measures, including automated tools, to proactively identify and remove content depicting sexual violence or abuse. While the objective is laudable, this amounts to pre-censorship and relies on fallible algorithms to make complex contextual decisions about speech.

Grievance Redressal and Executive Oversight: The rules mandate a three-tier grievance redressal mechanism, culminating in an "Oversight Mechanism" controlled by an inter-departmental committee of government officials. This effectively creates a system where the

[11] Shreya Singhal vs U.O.I AIR 2015 SUPREME COURT 1523

executive branch acts as the final arbiter of what content is permissible online, circumventing the judicial oversight mandated by the Shreya Singhal judgment.

THE TIGHTROPE WALK: REGULATION VS. CENSORSHIP

The IT Rules, 2021, have intensified the long-standing debate in India over the difference between content regulation and censorship. Proponents of the rules, including the government, argue that they are a necessary measure to make intermediaries more accountable to Indian laws and users. They contend that global tech giants cannot operate in a legal vacuum and must be subject to domestic regulations to curb the spread of fake news, hate speech, and other societal harms. The stated aim is to empower users with a timely grievance redressal mechanism and assist law enforcement in preventing, detecting, and prosecuting serious crimes. From this perspective, the rules are a legitimate exercise of the state's regulatory power to protect its citizens and maintain public order.

However, a chorus of voices from civil society, academia, and the tech industry has warned that these rules cross the line from regulation into a new regime of state-sponsored censorship. The sheer compliance burden and the threat of personal liability for employees, they argue, incentivise platforms to err on the side of caution and take down any content that is even remotely controversial, a phenomenon known as the chilling effect. By creating mechanisms for rapid takedowns and constant law enforcement coordination, the rules make it easier for the state to silence dissent and critical voices. The traceability mandate is seen not as a tool for crime prevention, but as an instrument that undermines the fundamental right to privacy, a right affirmed as intrinsic to the right to life and personal liberty by the Supreme Court in K.S. Puttaswamy v. Union of India (2017).[12] The core of the critique is that the IT Rules, 2021, shift the power to regulate speech from the judiciary, as envisioned in the Shreya Singhal case,[13] to the executive, creating a framework susceptible to political misuse and arbitrary action.

A TALE OF THREE REGIMES: INDIA'S PATH IN GLOBAL INTERMEDIARY LIABILITY

India's regulatory journey does not exist in isolation. Globally, nations are struggling to craft legal frameworks for the digital age. The United States has long been defined by Section 230

[12] Justice KS Puttaswamy (Retd) v Union of India (2017) 10 SCC 1
[13] Shreya Singhal vs U.O.I AIR 2015 SUPREME COURT 1523

of the Communications Decency Act,[14] which provides broad immunity to platforms, stating that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." This has been credited with fostering innovation and protecting free speech online, though it is now facing bipartisan criticism for allegedly shielding platforms from accountability for harmful content.

In contrast, the European Union has moved towards a more co-regulatory model with its recently enacted Digital Services Act (DSA).[15] The DSA maintains the core principle of conditional immunity for intermediaries but imposes stringent, risk-based due diligence obligations, particularly on Very Large Online Platforms (VLOPs). It mandates greater transparency in content moderation, algorithmic accountability, and provides robust mechanisms for user redress, all enforced by a powerful regulatory apparatus. The DSA's approach is often described as setting a gold standard for tech regulation that aims to protect fundamental rights while making platforms more responsible.

India's IT Rules, 2021, appear to cherry-pick elements from different models but ultimately carve a path that prioritises state control. While it borrows the idea of tiered obligations like the EU's DSA, its emphasis on traceability, short takedown timelines, and the personal liability of employees creates a far more coercive environment than what is seen in either the US or the EU. India's framework appears less concerned with user rights and due process and more focused on ensuring the government's ability to control the flow of information and enforce its writ in the digital sphere.

CONCLUSION

The journey of intermediary liability in India is a story of a constant tug-of-war between the judiciary, the executive, and the intermediaries themselves, with the fundamental rights of citizens hanging in the balance. The initial promise of a broad, judicially-guarded safe harbour, which fostered an environment of open expression, has given way to a prescriptive and demanding regulatory regime that many fear will stifle speech and compromise privacy. The constitutional validity of the IT Rules, 2021, is currently under challenge before various High

[14] Communications Decency Act 1996, s 230
[15] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

Courts and is likely to reach the Supreme Court, which will once again be called upon to adjudicate this critical issue.

The path forward requires a delicate balancing act. An unregulated internet is not a viable option; platforms must have robust and transparent systems to deal with genuinely harmful and illegal content. However, a regulatory framework that prioritises state control over individual rights and due process is equally dangerous. The future of free expression in India's digital domain will depend on whether the nation can find a middle path, one that holds powerful intermediaries accountable without turning them into extensions of the state, and that protects citizens from harm without sacrificing the freedoms that are the very essence of a democracy. The gatekeeper's gambit has been played; the outcome for India's billion-plus internet users is yet to be decided.