Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

AI-Powered Content Moderation and Criminal Liability: Constitutional Safeguards vs. Platform Accountability for Objectionable Content on OTT

Saumya Tripathi*[*]& *Kshem Dixit**

Ph.D. Scholar, Department of Law,
Gurugram University

Submission date 12.04.2026 | Acceptance date: 25.04.2026 | Publication: 29.05.2026

ABSTRACT

The rise of AI-powered content moderation on Over-The-Top (OTT) platforms presents a novel and urgent constitutional dilemma at the intersection of criminal law and corporate accountability. This paper critically examines the tension between the fundamental right to freedom of speech and expression under Article 19(1)(a) of the Indian Constitution and the state's legitimate interest in preventing the dissemination of illegal content such as hate speech, obscenity, and incitement to violence through these pervasive digital mediums. Central to this conflict is the evolving standard of criminal liability for intermediaries under Section 79 of the Information Technology Act, 2000, and the IT Rules, 2021. The paper argues that the deployment of opaque, autonomous Artificial Intelligence (AI) systems for content curation and removal fundamentally disrupts traditional legal doctrines of mens rea (guilty mind) and knowledge. It probes whether a platform can be held liable for criminal abetment or negligence when its AI either fails to detect patently illegal content or, conversely, erroneously flags and removes protected speech. This analysis is framed within the constitutional guarantee of due process under Article 21, questioning whether automated takedowns, driven by algorithmic interpretation, satisfy principles of natural justice for affected users. The research navigates a tripartite challenge: the constitutional validity of imposing criminal sanctions for algorithmic failures, the state's positive obligation to protect citizens from online harms, and the corporate platform's duty of due diligence. It concludes by proposing a refined legal framework that balances these interests. This framework advocates for "procedural due process" safeguards in automated moderation, a redefinition of "due diligence" to include algorithmic transparency and auditability, and clear statutory guidelines to prevent the chilling of legitimate speech, thereby ensuring that the Constitution remains the paramount guide in the digital courtroom.

Keywords: Criminal Liability of Intermediaries, Constitutional Due Process (Article 21), OTT Platform Regulation, Algorithmic Accountability, Automated Takedowns, Censorship, Digital Fundamental Rights

[*]Ph.D. Research Scholar, Indian Law Institute (Deemed University), New Delhi.
**Senior Engineer at Harman International, Pune.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

I. Introduction: The Digital Conundrum

The explosive growth of Over-The-Top (OTT) platforms in India from Netflix and Amazon Prime Video to Disney+ Hotstar and homegrown services like Jio-Cinema and ALT-Balaji has fundamentally transformed the media landscape[1]. These platforms deliver audio-visual content directly to consumers via the internet, bypassing traditional broadcasters and cable networks. While this technological leap has democratized content creation and consumption, it has simultaneously created a regulatory quagmire concerning the moderation of objectionable content. Unlike traditional broadcast media governed by the Cable Television Networks (Regulation) Act, 1995, and the Programming Code, OTT platforms initially operated in a legal grey area[2]. This regulatory vacuum was partially filled by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (hereafter "IT Rules, 2021"), which brought digital news publishers and OTT platforms under a self-regulatory framework overseen by the Ministry of Information and Broadcasting[3].

To manage the sheer volume of user-generated content and uploaded media, OTT platforms increasingly rely on Artificial Intelligence (AI) and Machine Learning (ML) systems for content moderation. These automated tools scan, flag, and sometimes remove content deemed to violate platform policies or national laws[4]. This shift from human-led curation to algorithmic governance raises profound legal and constitutional questions that lie at the heart of this paper. The core dilemma is tripartite. First, there exists a fundamental tension between the constitutional guarantee of free speech under Article 19(1)(a) and the state's compelling interest, under Article 19(2), to impose reasonable restrictions in the interests of sovereignty, security, public order, decency, and morality[5]. Second, the legal standard for holding an intermediary like an OTT platform criminally liable for user-generated content is governed by the conditional "safe harbor" immunity under Section 79 of the Information Technology Act, 2000 (IT Act). This immunity is contingent upon the intermediary exercising "due diligence" and having no "actual knowledge" of unlawful content[6]. The deployment of opaque AI systems disrupts the very foundations of these legal concepts: what constitutes "knowledge" for an algorithm, and can an algorithmic failure satisfy the mens rea requirement for criminal liability? Third, automated content removal actions engage the right to life and personal liberty under Article 21, which has been interpreted by the Supreme Court to include the right to a fair procedure and the principles of natural justice[7]. Does an AI-

[1] FICCI and EY, FICCI-EY Media & Entertainment Report 2023: *The Era of Consumer Interactivity* (2023).
[2] Government of India, "Recommendations of Telecom Regulatory Authority of India on Regulatory Framework for Over-the-top (OTT) Communication Services" (Ministry of Information and Broadcasting, 2020).
[3] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, notified under s 87, The Information Technology Act, 2000 (Act of 2000).
[4] Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press, New Haven 2018) 5–7.
[5] Constitution of India, Art 19(2).
[6] Information Technology Act, 2000, (Act 21 of 2000), s 79.
[7] *Maneka Gandhi* v. *Union of India*, (1978) 1 SCC 248.

driven takedown, often without meaningful human review or appeal, satisfy this constitutional guarantee of due process?

This paper argues that the current legal framework is inadequately equipped to address the unique challenges posed by AI-powered content moderation on OTT platforms. The blurring of lines between platform as passive conduit and active curator, coupled with the "black box" nature of algorithmic decision-making, creates a regime of potential over-censorship and unaccountability. Through a doctrinal analysis of constitutional provisions, criminal law principles, and intermediary liability rules, this article seeks to construct a refined legal framework. This framework aims to reconcile platform accountability for genuine harms with robust safeguards for free speech and due process in the algorithmic age.

II. Constitutional Foundations: Article 19(1)(a) and the State's Power to Restrict

A. The Expansive Right to Freedom of Speech and Expression

Article 19(1)(a) of the Constitution guarantees to all citizens the right to freedom of speech and expression. The Supreme Court has consistently interpreted this right in the broadest possible terms. In *Romesh Thappar* v. *State of Madras* (1950), the Court held that freedom of speech "includes the freedom to propagate ideas which are received by others, and is ensured by the freedom of circulation[8]." This right is not confined to mere words but extends to all forms of expression, including artistic, cinematic, and digital expression. The Court in *K.A. Abbas* v. *Union of India* (1970) recognized film as a "powerful medium of communication" deserving of Article 19(1)(a) protection, a logic that applies a fortiori to content streamed on OTT platforms[9].

The internet and digital platforms have been recognized as critical enablers of this right. In *Anuradha Bhasin* v. *Union of India* (2020), the Supreme Court affirmed that "the freedom of speech and expression and the freedom to practice any profession or carry on any trade, business or occupation over the medium of internet enjoys constitutional protection under Article 19(1)(a)[10]." This landmark judgment cemented the status of digital speech within the protective ambit of the Constitution.

B. Permissible Restrictions under Article 19(2) and the Proportionality Test

The right under Article 19(1)(a) is not absolute. Article 19(2) empowers the state to impose "reasonable restrictions" by law on the exercise of this right in the interests of:

- the sovereignty and integrity of India
- the security of the state
- Friendly relations with foreign states

[8] *Romesh Thappar* v. *State of Madras*, AIR 1950 SC 124.
[9] *K.A. Abbas* v. *Union of India*, (1970) 2 SCC 780.
[10] *Anuradha Bhasin* v. *Union of India*, (2020) 3 SCC 637, para 86.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

- public order
- Decency or morality
- Contempt of court\
- Defamation
- Incitement to an offence

The Supreme Court has developed a rigorous jurisprudence around the "reasonableness" of such restrictions. The classic test from *Chintaman Rao* v. *State of Madhya Pradesh* (1950) requires that the restriction should not be "arbitrary or excessive," and the "means adopted" should have a "real and substantial relation" to the object sought to be achieved[11]. More recently, in *Modern Dental College* v. *State of Madhya Pradesh* (2016), the Court adopted the doctrine of proportionality as a key component of the reasonableness test[12]. A proportionality analysis typically involves a four-pronged test:

1. The measure must be designated for a proper purpose.

2. The measure must be rationally connected to that purpose.

3. The measure must be necessary, meaning there is no less restrictive alternative available.

4. There must be a fair balance between the benefits of the measure and its restrictions on rights (proportionality stricto sensu).

C. Application to OTT Content and AI Moderation

This constitutional framework directly applies to the regulation of content on OTT platforms. The state has a legitimate interest, under Article 19(2), in preventing the streaming of content that constitutes hate speech (affecting public order), obscenity (affecting decency or morality), or direct incitement to violence (an offence). The IT Act and Rules, along with provisions of the Bharatiya Nyaya Sanhita (BNS) like Sections 152 (promoting enmity), 294 (obscenity), and 353 (statements conducing to public mischief), provide the legal basis for these restrictions.

The critical question is whether the means employed specifically, mandating or incentivizing platforms to use automated AI systems for pre-emptive content moderation satisfies the proportionality test. Algorithmic systems are notoriously poor at understanding context, satire, artistic intent, or linguistic nuance. A blanket takedown of content flagged by an AI for potential violation may constitute an excessive and disproportionate restriction if a less restrictive alternative (e.g., human review post-flagging, age-gating, content descriptors) could achieve the

[11] *Chintaman Rao* v. *State of Madhya Pradesh*, AIR 1951 SC 118.
[12] *Modern Dental College & Research Centre* v. *State of Madhya Pradesh*, (2016) 7 SCC 353.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

same legitimate state aim. The risk of "collateral censorship", where platforms, to avoid liability, over-remove legal content, directly chills free speech and may fail the proportionality balance[13].

III. The Architecture of Intermediary Liability: Section 79 IT Act and the "Safe Harbour"

A. The Conditional Immunity of Section 79

The linchpin of intermediary liability in India is Section 79 of the IT Act. Modelled on the U.S. Digital Millennium Copyright Act's safe harbour provisions, it grants intermediaries immunity from liability for any thirdparty information, data, or communication link hosted or transmitted by them[14]. However, this immunity is not absolute. It is conditional upon the intermediary fulfilling two primary conditions, as outlined in Section 79(2):

1. The intermediary's function is limited to providing access to a communication system.

2. The intermediary does not: (a) initiate the transmission; (b) select the receiver of the transmission; or critically, (c) select or modify the information contained in the transmission.

3. The intermediary must observe "due diligence" while discharging its duties and also observe such other guidelines as the Central Government may prescribe.

The proviso to Section 79(3) further clarifies that this immunity ceases if the intermediary, upon receiving actual knowledge that any information is being used to commit an unlawful act, fails to expeditiously remove or disable access to that material[15].

B. The "Actual Knowledge" Standard and Judicial Interpretation

The phrase "actual knowledge" has been the subject of significant judicial scrutiny. In *Shreya Singhal* v. *Union of India* (2015), the Supreme Court struck down the draconian Section 66A of the IT Act and provided a crucial interpretation of intermediary liability[16]. The Court held that "actual knowledge" can only arise in two scenarios:

1. Upon receiving a court order directing the takedown of specific content.

2. Upon receiving a notification from the "appropriate government or its agency" under Section 69A of the IT Act (which deals with blocking of information for specific, stated reasons).

The Court explicitly rejected the notion that a private complaint or a "notice" from any individual could impose a takedown obligation on the intermediary, as this would lead to excessive

[13] David S Ardia, "Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act" 43 *Loyola of Los Angeles Law Review* 373 (2010).
[14] Information Technology Act, 2000, (Act 20 of 2000), s. 79(1).
[15] Information Technology Act, 2000, (Act 20 of 2000), s. 79(1).
[16] *Shreya Singhal* v. *Union of India*, (2015) 5 SCC 1.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

censorship. This interpretation was a significant victory for free speech, placing the onus of identifying illegal content on state authorities, not private platforms or individuals.

C. The IT Rules, 2021: Diluting "Actual Knowledge" and Expanding "Due Diligence"

The IT Rules, 2021, introduced a more complex and arguably expansive regime for intermediaries, particularly "significant social media intermediaries" (SSMIs). While OTT platforms are categorized separately under Part III of the Rules, the overarching shift in philosophy is relevant.

For SSMIs, the Rules introduced a proactive monitoring obligation through the requirement to deploy technology-based measures, including automated tools, to identify certain categories of unlawful content[17]. This marked a departure from the *Shreya Singhal* "actual knowledge" standard, moving towards a "constructive knowledge" or "ought to have known" standard. Although OTT platforms are not explicitly subjected to the same automated tools mandate, the Rules require them to implement a "grievance redressal mechanism" and classify content based on age suitability (U, U/A 7+, U/A 3+, U/A 16+, and A)[18]. The practical pressure on platforms to preemptively identify and restrict content to avoid complaints and regulatory scrutiny is immense. This pressure incentivizes the use of AI moderation tools, effectively creating a de facto proactive monitoring regime.

The concept of "due diligence" under Section 79, further elaborated in the Rules, now implicitly encompasses the deployment of content moderation systems. Failure to effectively deploy such systems could be construed as a failure of due diligence, potentially stripping the platform of its safe harbour immunity. This creates a legal Catch: use imperfect AI and risk over removing legal speech, or don't use it and risk liability for missing illegal content.

IV. Criminal Liability in the Algorithmic Age: Mens Rea, Abetment, and Negligence

The integration of AI into content moderation directly challenges foundational principles of criminal law, particularly the requirement of mens rea (a guilty mind).

A. The Doctrine of Mens Rea and Its Inapplicability to AI

Indian criminal law, drawing from the common law tradition, generally requires the presence of *mens rea* alongside the *actus reus* (guilty act) for the imposition of criminal liability[19]. Crimes can be of specific intent (e.g., sedition under Section 152 BNS) or general intent (e.g., obscenity under Section 294 BNS). However, an AI system is not a legal person. It possesses no consciousness, intent, or moral agency. It operates on statistical correlations and pattern recognition within its training data. When an AI fails to flag a piece of hate speech, it is not exhibiting "negligence" in the human sense; it is producing an output based on its programming and data. Attributing *mens*

[17] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r 4(4).
[18] *Ibid*.
[19] Ratanlal & Dhirajlal, *The Indian Penal Code* (LexisNexis, 36th edn., 2023) Introduction.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

rea whether intention, knowledge, recklessness, or negligence to a corporate entity for the action or inaction of its AI is a profound legal fiction.

B. Potential Avenues for Criminal Liability of Platforms

Despite the *mens rea* conundrum, prosecutors may attempt to hold OTT platforms criminally liable under several theories:

1. Direct Liability for Hosting Illegal Content: If the safe harbour of Section 79 is lost (due to failure of due diligence or upon receipt of actual knowledge), the platform could be treated as a publisher of the illegal content. For example, it could be charged under Section 294 BNS for obscenity or Section 196 for promoting enmity. The legal battle would center on whether the platform's use of (faulty) AI constituted a failure of "due diligence."

2. Abetment under Section 45 BNS: Abetment involves instigating, conspiring, or intentionally aiding the commission of an offence[20]. Could providing a platform that an AI fails to adequately police be considered "intentional aiding"? This would require proving that the platform intended its infrastructure to be used for unlawful purposes, a high threshold unlikely to be met by mere algorithmic failure. However, if it can be shown that the platform deliberately designed its AI to be underinclusive to maximize engagement with provocative content, an abetment case becomes more plausible.

3. Criminal Negligence under Section 106 BNS: This section prescribes punishment for causing death by a negligent act. While not directly applicable to content, the principle highlights the potential for liability based on corporate negligence. A novel argument could be that a platform's failure to deploy a reasonably effective AI moderation system, leading to the spread of content that incites real-world violence (e.g., lynching), constitutes gross negligence. However, establishing the proximate causal link between the algorithmic failure and a specific violent act would be exceptionally difficult.

The fundamental challenge is that criminal law is anthropocentric. It is designed to judge human moral culpability. Holding a corporation liable for the stochastic errors of a complex statistical model stretches these doctrines to their breaking point and raises serious concerns of fair notice and constitutional vagueness.

V. Due Process under Article 21 and the "Automated Takedown"

Article 21 of the Constitution states: "No person shall be deprived of his life or personal liberty except according to procedure established by law." The Supreme Court has infused this Article with substantive content, holding that the "procedure established by law" must be "fair, just and reasonable," not merely a procedure prescribed by statute[21]. This encompasses the principles of

[20] Bharatiya Nyaya Sanhita (BNS), 2023, s 45.
[21] *Maneka Gandhi* v. *Union of India*, (1978) 1 SCC 248.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

natural justice: *audi alteram partem* (hear the other side) and *nemo judex in causa sua* (no one should be a judge in their own cause).

A. The Right to a Meaningful Hearing

When a human moderator removes a user's video or post, there is at least the possibility, however limited, of understanding the rationale and contesting it. An AI-driven takedown operates differently. The decision is instantaneous, often based on opaque criteria. The user typically receives a generic notification: "Your content violates our Community Guidelines." The specific reason (e.g., "hate speech," "graphic violence") may be given, but the algorithmic basis for that classification is almost never disclosed.

This process fails the basic tenets of natural justice. The content creator is not given a meaningful opportunity to be heard before the deprivation (the takedown). The appeal process, if it exists, is often another automated or outsourced system. The "judge "the AI is not only in the platform's cause but is also inscrutable. As the Supreme Court held in Maneka Gandhi v. Union of India (1978), the right to be heard must be meaningful and effective[22]. An opaque, automated decision-making process with a feeble appeals mechanism is neither.

B. The Right to Know the Reasons

A corollary of the right to a hearing is the right to know the reasons for a decision that adversely affects one's rights[23]. When an AI system flags a documentary on caste discrimination as "hate speech" or a satirical sketch as "inciting violence," the user has a right to understand why. This is not merely a platform policy issue; when the takedown is undertaken to comply with or avoid potential liability under Indian law (e.g., BNS sections), it becomes a state action adjacent deprivation engaging Article 21. The current lack of algorithmic transparency violates this fundamental due process right.

C. The Chilling Effect as a Due Process Violation

The uncertainty and arbitrariness of AI moderation have a profound chilling effect on free speech[24]. When creators cannot predict what will be flagged, they engage in self-censorship, avoiding topics related to sexuality, religion, politics, or social critique. This chilling effect is itself a deprivation of the right to free expression, a right which the Supreme Court in *K.S. Puttaswamy* v. *Union of India* (2017) held is an integral part of the right to privacy and liberty under Article 21[25]. Thus, an unfair, opaque moderation system violates Article 21 both in its direct procedural unfairness and in its secondary chilling effect on Article 19(1)(a).

[22] *Ibid*.
[23] *S.N. Mukherjee* v. *Union of India,* (1990) 4 SCC 594.
[24] *Umesh Kumar* v. *State of Andhra Pradesh*, (2013) 10 SCC 591.
[25] *Justice K.S. Puttaswamy (Retd.)* v. *Union of India*, (2017) 10 SCC 1.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

VI. Towards a Constitutional and Accountable Framework: Proposed Reforms

The current trajectory risks creating a regime where constitutional rights are silently eroded by private algorithmic systems operating under the shadow of vague liability threats. To prevent this, a refined legal and regulatory framework is urgently needed. This framework must be anchored in constitutional principles and aim for a proportionate balance.

A. Legislate a "Procedural Due Process" Safeguard for Automated Actions

Parliament should amend the IT Act or enact a new Digital Rights Act to incorporate explicit due process requirements for any automated content moderation action that restricts access to lawful speech. This should include:

- Pre-Takedown Notice: For nonemergency content, the user must be given a clear, detailed notice of the alleged violation and the specific content in question, with an opportunity to submit a written response before takedown.
- Post-Takedown Appeal: A robust, timely, and human moderated appeals process must be mandatory. The appeal must be reviewed by a person with relevant expertise and cultural/linguistic context.
- Statement of Reasons: Every final takedown decision must be accompanied by a clear, reasoned order that explains the legal or policy basis, referencing specific portions of the content.
- Judicial Oversight: A fast-track mechanism should be established in designated courts or tribunals to hear appeals from platform decisions, especially where the content involves political, artistic, or journalistic speech.

B. Redefine "Due Diligence" to Encompass Algorithmic Transparency and Auditability

The government should issue clear guidelines under Section 79 of the IT Act redefining "due diligence" for platforms using AI moderation[26]. "Due diligence" should not mean the mere deployment of AI, but its responsible deployment, measured by:

- Transparency Reports: Mandatory periodic publication of transparency reports detailing the number of takedowns, categories, automation rates, and appeal success rates.
- Algorithmic Audits: Platforms should be required to commission and publish independent, third-party audits of their AI moderation systems for bias (linguistic, cultural, political), accuracy (precision and recall rates), and error rates. The audit methodology and key findings should be public.
- Risk-Based Proportionality: The intensity of AI filtering should be proportionate to the assessed risk of the content and the context. A live news stream should not be subjected to the same pre-emptive filtering as a permanently hosted film.

[26] *Ibid*.

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

C. Clarify and Fortify the "Actual Knowledge" Standard

The Shreya Singhal standard must be reaffirmed and statutorily entrenched. The law should explicitly state that "actual knowledge" for the purpose of intermediary liability under Section 79 arises only from:

1. A binding order from a competent court or tribunal.

2. A notification from the designated government authority under Section 69A, following the strict procedures outlined therein (including a hearing to the originator, where possible). This would insulate platforms from the pressure to over-censor based on unverified complaints and place the primary responsibility for identifying unlawful speech on state authorities, as the Constitution envisages.

D. Establish a Co-Regulatory Body for OTT Content

Drawing from models like the Broadcasting Content Complaints Council (BCCC), a dedicated, independent coregulatory body for OTT content should be established by statute. This body, comprising retired judges, legal experts, technologists, and civil society representatives, would have multiple functions:

- Develop nuanced, context-sensitive classification guidelines beyond mere age-rating.
- Serve as an independent appellate authority for user grievances against platform decisions.
- Commission and review algorithmic audits.
- Advise the government on policy, ensuring it remains within constitutional bounds.

VII. Conclusion

The integration of AI into the content moderation systems of OTT platforms is an irreversible technological reality. However, the law must not blindly capitulate to technological determinism. The Indian Constitution, with its robust framework of fundamental rights and its living tree doctrine of interpretation, provides the necessary tools to govern this new frontier. The central finding of this analysis is that the current regulatory approach, embodied in the IT Rules, 2021, creates a dangerous imbalance. It incentivizes platforms to deploy opaque, error-prone AI systems, leading to the privatized censorship of lawful speech without the procedural safeguards demanded by Articles 19(1)(a) and 21. Simultaneously, the threat of criminal liability f-or algorithmic failure is based on legal concepts (*mens rea*, knowledge) that are fundamentally mismatched with the reality of automated systems, creating a regime of unpredictable and potentially unfair corporate prosecution. The path forward is not Luddisman's outright ban on AI moderation nor is its laissez-faire abdication. The solution lies in constitutionally aligned governance. The proposed framework of procedural due process, transparent and auditable due diligence, a fortified actual knowledge standard, and independent coregulation seek to achieve this alignment. It acknowledges the state's legitimate interest in preventing online harm and the platform's need for operational efficiency, while placing inviolable constitutional rights free speech, due process, and liberty at the very center

Cadernos de Pós-Graduação em Direito Político e Econômico
Published by: Centro de Estudos Acadêmicos Press
ISSN: 1678-2127
Volume 26 *Issue S1*, 2026
Gurugram University Conference : 09-10 April 2026
Website: https://ceapress.org

of the digital public square. In the final analysis, the "digital courtroom" where algorithmic judgments are rendered must have the Constitution as its presiding judge. Only then can we ensure that the immense power of AI serves to enhance, rather than erode, the democratic discourse that is the lifeblood of the Indian republic.