



BEYOND THE PROMPT: DECONSTRUCTING INTERMEDIARY LIABILITY IN THE AGE OF GENERATIVE AI

Adithri P Sajeesh*

ABSTRACT

The “Safe Harbour” doctrine that has protected internet platforms for decades has been fundamentally disrupted by the emergence of generative AI (“genAI”). This article deconstructs the shift from passive intermediary liability to a “Safety by Design” mandate using the January 2026 Grok AI controversy as a focal point. This article has aimed to address the inadequacy of Section 79 of the IT Act, 2000, the revolutionary provisions of the AI Ethics and Accountability Bill, 2025, and the need to centre digital dignity and consent over traditional ideas of morality and decency.

Keywords: Section 79, IT Act 2000, Safe Harbour, Due Diligence Obligations, Action Taken Report (ATR), Safety-by-Design.

FACTS: THE GROK AI CONTROVERSY

xAI, the parent company of the artificial intelligence chatbot, Grok, and the social media platform, X (formerly known as Twitter), launched a new image generation and editing feature within Grok in December 2025. This development led to significant enhancements in the platform’s capabilities, which enabled users to create and modify images directly through the interface. In addition to being a standalone platform, Grok AI is also embedded within the social media platform X, which amplified the reach and usage of the tool.

Subsequently, the platform witnessed a seemingly unprecedented increase in the creation of Non-Consensual Intimate Imagery (NCII) within days of its release. Users of the platform

*BSC LLB (HONS.), FIRST YEAR, THE WEST BENGAL NATIONAL UNIVERSITY OF JURIDICAL SCIENCES.

engaged in a mass “digital undressing spree” that targeted both public and private figures, predominantly women and children.

This incident resulted in significant outrage from the public, who voiced their concerns regarding privacy violations and exploitation of individuals, particularly minors and the potential long-term harm caused by the dissemination of such imagery. These concerns were echoed and amplified by parliamentary stakeholders, who called for urgent action and accountability.

In response to these widespread protests, the Ministry of Electronics and Information Technology (MeitY) issued a notice to the platform, X, which stated that the corporation’s failure to comply with statutory due diligence obligations under the Information Technology Act, 2000 and the IT Rules, 2001. The Ministry has grave concerns regarding this misuse of technology within the platform and the display of obscene or indecent content, targeting women and minors. The government gave X Corp an ultimatum of 72 hours, within which they have to supply an ATR (Action Taken Report) that lists the measures adopted, compliance with mandatory reporting under the Bharatiya Nagarik Suraksha Sanhita (BNSS), 2023 and the duty of the chief compliance officer.

After a two-day extension requested by and granted to X, the corporation reportedly acknowledged its mistake and returned a detailed document. However, the ministry found the corporation’s reply to be largely “inadequate” and demanded more information on the AI’s preventive mechanisms and other relevant frameworks in place.

India is not alone in locking horns with the xAI. Countries such as Britain, France, Malaysia, Brazil and the European Union have officially condemned the platform for its lack of a regulatory framework. The global reaction was swift, with Indonesia and Malaysia blocking Grok within 48 hours and the UK’s Ofcom launching a formal investigation under the Online Safety Act (OSA).

THE IT ACT AND ITS INSUFFICIENCIES IN THE AI ERA

The Information Technology Act, 2000, with specific reference to Section 79, which predominantly governs the digital atmosphere of India, is designed according to a necessary legal fiction which states that a digital platform is merely a postman and is not responsible for the contents that are posted by its users. This principle, known as the “intermediary liability

protection” or “safe harbour protection”, was necessary for the digital economy to flourish. However, it is crucial to note that this principle is conditional and requires that,

1. The intermediary must not initiate the transmission.
2. The intermediary must not select the receiver.
3. The intermediary must not select or modify the information.

The performance of any of the above actions transforms a platform from passive to proactive – a distinction that was solidified in the landmark case, *Christian Louboutin SAS v. Nakul Bajaj* (2018). The idea of digital platforms being passive platforms is largely challenged by the advent of generative AI and user algorithm-based feed content. Generative AI models like Grok do not merely host explicit content but synthesise a response. When an AI generates a deepfake, it initiates the creation of a new digital object, thus transforming from a mere passive host of the user’s prompt to an active agent in the creation and dissemination of the explicit content. Under a strict and preferred reading of the IT Act, 2000, a platform that hosts generative AI loses its status as a “passive conduit”. The legislation simply does not have a category that hosts “Intermediary-Hybrid Creator”.

THE GLOBAL SHIFT: INDIA’S POSITION ON THE WORLD STAGE

Regulation of generative AI has come to the forefront of law and policy-making, both in the international and national spheres. The latest and most comprehensive regulatory framework is the EU AI Act. This legislation has established that the majority of obligations fall on the providers (developers) of high-risk AI systems. Chapter II, Article 5 of the EU AI Act outlines the factors that would prohibit an AI’s usage, which include exploitation of vulnerabilities, social scoring and distortion of behaviour by manipulation or deception techniques.

India has historically occupied a middle ground by relying on the IT Act, 2000 and IT Rules, 2021. However, the IT Rules 2021 are a tethered attempt to stretch the IT Act, 2000, to fit a new reality. By 2025, with the increased presence of generative AI, it has become sufficiently evident that it is imperative that India move away from the present Western “reactive” models towards a uniquely Indian “Safety-by-Design” framework.

CRITICAL ANALYSIS

The Grok controversy of 2026 is not a failure of moderation but rather a failure of legal imagination that is necessary for an innovative solution to the problem of a lack of regulatory frameworks. To solve this crisis of “synthetic violence”, we must integrate the relevant analysis of technology with a critical assessment of constitutional rights and corporate ethics.

The Consent Paradigm: Shifting the Legal Lens: The most profound failure with India’s current response is the obsession with archaic ideas of “lewdness” and “obscenity” and a lesser focus on concepts of “consent” of the victims. As an example, Section 67 of the IT Act, 2000, focuses on aspects such as “lasciviousness” and whether it “appeals to prurient interests”, which is a colonial principle that does not meet the standards of today’s changing, wider society.

The inherent harm in the Grok-orchestrated controversy is not the nature of the image, and it is supposedly not fitting present moral standards, but is rather the radical absence of consent. Indian jurisprudence must pivot from a “morality-based” approach to a “consent-based” system. In the present age of generative AI, the sole trigger of liability should be the lack of consent, regardless of whether the generated output meets a threshold of “obscenity”.

The Chilling Effect: A Constitutional Crisis: When genAI tools such as Grok, in collaboration with social media platforms like X, “pornify” the likeness of unsuspecting women and minors, it creates a systematic “chilling effect” on the affected parties, suppressing this community from participating and interacting in public forums. This is a direct assault on the right of Freedom of Speech and Expression, enshrined in Article 19 (1) (A) of the Constitution of India. While it's integral to preserve the “free speech” of the platform, this right should not supersede the fundamental right of a citizen to exist in a digital space without being subjected to non-consensual sexualization.

Strict Liability: Managing "Digital Pollution": The Indian judicial system must move towards a “strict liability” model for high-risk AI. Just as a factory is liable for a chemical leak regardless of the organisation’s intent, a platform should be strictly liable for the generation and dissemination of NCII imagery.

The burden to prove the negligence must not rest with the victim and must instead lie with the platform to prove that it implemented “Safety-by-Design- a technical standard where safety is the primary requirement and not an afterthought.

CONCLUSION

The legal fiction of “passive postman” has officially retired with the advent of the January 2026 Grok controversy. For decades, digital platforms operated under the “safe harbour” shield, protected by principles and rules designed for a world of digital world of static hosting. However, as algorithms traverse beyond delivery to synthesis, this shield becomes a weapon of immunity for the creation of societal harm. The “digital undressing” spree on X exposes a fundamental truth that a platform cannot provide the ink, hand and voice and then claim to be a mere bystander.

As the idea of intermediary liability is deconstructed, Indian jurisprudence must shed its colonial obsession with archaic ideas of “obscenity” and progress towards a modern, robust framework of digital consent. This systematic “chilling effect” on women and children is not to be treated as a mere by-product of innovation and is instead a constitutional crisis that threatens Article 19. If a tool found to be capable of “digital pollution” is designed by a platform, it must be held to the same threshold of “strict liability” as any other hazardous industry.

REFERENCES

1. Information Technology Act 2000, s 79.
2. Shreya Singhal v Union of India AIR 2015 SC 1523.
3. Christian Louboutin SAS v Nakul Bajaj (2018) 253 DLT 728.
4. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).
5. Supreme Court Observer, ‘Controversy Over Grok’s Images: Uncertainty Surrounds Intermediary Liability’ (Supreme Court Observer, 21 August 2024) <https://www.scobserver.in/journal/controversy-over-groks-images-uncertainty-surrounds-intermediary-liability/> accessed 24 January 2026.

6. 'X: Elon Musk's Grok chatbot to be investigated over deepfakes' (BBC News, 19 August 2024) <https://www.bbc.com/news/articles/ce8gz8g2qnlo> accessed 24 January 2026.
7. IBM, 'What is the EU AI Act?' (IBM Think, 2024) <https://www.ibm.com/think/topics/eu-ai-act> accessed 24 January 2026.
8. 'EU AI Act: High-level summary' (Artificial Intelligence Act, 2024) <https://artificialintelligenceact.eu/high-level-summary/> accessed 24 January 2026.
9. ICAEW, 'Legal considerations when using generative AI' (ICAEW, 2024) <https://www.icaew.com/technical/technology/artificial-intelligence/generative-ai-guide/legal-considerations> accessed 24 January 2026.
10. The SoftCopy Legal Digest, 'Shreya Singhal v. Union of India: A Critical Analysis' (TSCLD, 2021) <https://www.tsld.com/shreya-singhal-v-union-of-india-a-critical-analysis> accessed 24 January 2026.
11. Fei Sun, Damir Iovic and Gordana Dodig-Crnkovic, 'Axiology and the Evolution of Ethics in the Age of AI: Integrating Ethical Theories via Multiple-Criteria Decision Analysis' (2025) 126(1) Proceedings 17 <https://doi.org/10.3390/proceedings2025126017> accessed 23 January 2026.
12. 'Notice to X for Misuse of Grok to Create and Share Obscene, Indecent, Explicit Contents' (Revoi, 2025) <https://english.revoi.in/notice-to-x-for-misuse-of-grok-to-create-and-share-obscene-indecent-explicit-contents/> accessed 23 January 2026.