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

Bridging the Digital Divide: A Comparative Analysis of Deepfake Regulation in India and the United States

Himanshu Thakran[1]
Research Scholar, Department of Law,
Gurugram University

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

ABSTRACT

The rapid development of deepfake technology synthetic media created by artificial intelligence to convincingly imitate human appearance and behaviour poses a structural problem to the global legal order. Beyond being a mere advancement in disinformation, deepfakes represent a fundamental challenge to the epistemic trust upon which democratic, legal, and diplomatic processes rely. By enabling low-cost, high-impact deception, they exploit the grey zones of digital governance, complicating principles of cyber sovereignty and making it difficult to attribute wrongful acts in the digital sphere.

To address this crisis of truth, this paper employs a comparative analytical framework to evaluate divergent regulatory models, specifically focusing on India and the United States. First, it critically analyses India's legislative reforms, namely, the Information Technology (Intermediary Guidelines) Amendment Rules, 2026. It assesses India's transition from platform neutrality to active accountability through means of mandatory labelling of Synthetically Generated Information (SGI) and ultra-fast, 2-hour takedown requirements. Second, to fulfill its comparative mandate, the paper contrasts India's centralized, state led approach with the fragmented, First Amendment constrained regulatory landscape of the United States, evaluating US federal initiatives and state level legislative patchworks.

This contrast reveals the ongoing clash between protecting free speech and maintaining a stable society. Ultimately, the paper concludes that bridging the regulatory "digital divide" in deepfake governance requires moving beyond isolated domestic policies. Instead of outright technological bans, the paper proposes a path forward toward harmonized, universal standards of technical attribution and cryptographic provenance. Reconciling the US's decentralized, market driven ethos with India's stringent accountability mandates will be essential to safeguarding the integrity of reality and reclaiming digital sovereignty in a shared global information space.

Keywords: Deepfake Technology ,Artificial Intelligence (AI) , Digital Divide

I. Introduction

[1] [1] Research scholar; Department of Law, Gurugram University Gurugram

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

The proliferation of artificial intelligence has birthed a new era of digital interaction, characterized most notably by the advent of "deepfakes." Derived from the underlying technology of "deep learning" and the intent to create a "fake," deepfakes are highly realistic, synthetically generated audio, video, or image files. While early iterations of this technology required substantial computational power and technical expertise, the modern landscape is defined by the democratization of AI. Today, anyone with a smartphone and internet access can generate convincing synthetic media, precipitating a crisis of epistemic trust a breakdown in society's shared understanding of objective reality.[2]

The legal and societal implications of this technological leap are profound. Deepfakes have been deployed to non-consensually place individuals into pornographic material, orchestrate sophisticated financial frauds, and manufacture political disinformation designed to sway elections and incite public unrest.[3] This capacity for low cost, high impact deception challenges the very foundations of the global legal order. It blurs the lines of cyber sovereignty and complicates traditional legal mechanisms of attributing liability and establishing intent.

As nations grapple with this epistemic threat, a pronounced "digital divide" has emerged in regulatory philosophies. This divide is not merely technological, but deeply ideological. It represents the schism between nations that prioritize state-led security and rapid intervention, and those that prioritize free expression and market-driven self-regulation. This research paper critically examines this regulatory divide by analysing the contrasting approaches of the Republic of India and the United States of America.

India has recently adopted an aggressive, centralized approach, culminating in the Information Technology (Intermediary Guidelines) Amendment Rules, 2026, which mandate stringent platform accountability and rapid takedown mechanisms. In stark contrast, the United States relies on a fragmented, decentralized patchwork of state laws and federal advisories, heavily constrained by the constitutional protections of the First Amendment.

The objective of this paper is to conduct a detailed comparative analysis of these two distinct paradigms. Through this analysis, the paper aims to highlight the inadequacies of isolated domestic policies in a borderless digital ecosystem and propose a harmonized, technology-neutral legal framework. It argues that bridging this regulatory digital divide requires adopting universal standards of cryptographic provenance and technical attribution, balancing the Indian imperative for accountability with the American commitment to free expression.

II. The Anatomy of the Threat: Synthetic Media in the Digital Age

To effectively regulate deepfakes, one must first understand their underlying mechanics and the unique challenges they present to legal systems.

[2] Robert Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security" 107 California Law Review 1753 (2019).
[3] Shruti Agarwal, et.al., "Protecting World Leaders Against Deep Fakes" Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 38 (2019)

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

A. The Technological Mechanics of Deception

Deepfakes are primarily generated using Generative Adversarial Networks (GANs) and advanced diffusion models. A GAN consists of two distinct neural networks: the "generator" and the "discriminator." The generator attempts to create a synthetic image or video, while the discriminator analyses it against a dataset of real media to determine if it is fake.[4] This adversarial process continues iteratively; the generator constantly improves its output to fool the discriminator, resulting in synthetic media that is virtually indistinguishable from reality to the human eye.

Furthermore, voice cloning technologies require only mere seconds of a person's recorded audio to synthesize their voice flawlessly, capturing cadence, tone, and emotional inflection. The resulting Synthetically Generated Information (SGI) is not merely a manipulated photograph, akin to traditional image editing; it is an entirely fabricated digital reality.

B. The Epistemic Crisis and the "Liar's Dividend"

The most pernicious threat posed by deepfakes is not the fake media itself, but the resulting erosion of trust in genuine media. Legal scholars identify this phenomenon as the "Liar's Dividend."[5] In an information ecosystem where anything can be convincingly faked, wrongdoers can easily dismiss genuine evidence of their misdeeds—such as an authentic video of corruption or a human rights violation—by simply claiming it is a deepfake.

This strikes at the heart of the judicial process. Courts have historically relied on audio-visual evidence as an objective record of fact. The ubiquitous presence of deepfakes threatens the probative value of such evidence, forcing legal systems to expend massive resources on digital forensics merely to establish the baseline authenticity of an exhibit. In democratic processes, this epistemic rot manifests as a vulnerability to foreign and domestic influence operations. Synthetic media can be deployed hours before an election, leaving victims and regulators with insufficient time to debunk the falsehood before the electoral damage is done.

C. The Regulatory Challenge

Regulating deepfakes presents a unique jurisprudential puzzle. Unlike malware or traditional cyberattacks, deepfakes do not damage physical infrastructure or breach computer systems. They target human cognition.[6] Therefore, they often fall into the "grey zone" of international and domestic law. Banning the underlying technology is legally untenable and practically impossible, as AI possesses legitimate dual-use applications in education, entertainment, and

[4] Ian J. Goodfellow, et.al., "Generative Adversarial Nets" 27 Advances in Neural Information Processing Systems 2672 (2014).
[5] Danielle Keats Citron and Robert Chesney, "Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics" 98 Foreign Affairs 147 (2019).
[6] Chris Tenove, "Deepfakes and the Epistemic Crisis of Democracy" 33 Philosophy & Technology 581 (2020).

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

accessibility services. The challenge, therefore, lies in regulating the *malicious application* of the technology without stifling innovation or infringing upon fundamental rights.

III. The Indian Approach: State-Led Accountability and Rapid Remediation

India's approach to deepfake regulation is rooted in its broader philosophy of digital sovereignty and internet governance. As the world's most populous democracy with the largest user base for global social media platforms, India views digital disinformation as a direct threat to public order and national security. Consequently, the Indian regulatory model emphasizes proactive state intervention, strict intermediary liability, and rapid remediation.

A. The Existing Statutory Framework

Prior to specific deepfake regulations, India utilized a combination of the Information Technology Act, 2000 (IT Act) and the Bharatiya Nyaya Sanhita, 2023 (BNS)—which recently replaced the Indian Penal Code—to prosecute malicious synthetic media.[7]

Under the IT Act, Section 66D addresses the offence of cheating by personation by using a computer resource, which can be applied to deepfake enabled financial frauds. Section 66E protects privacy, penalizing the capturing or publishing of images of a person's private areas without consent, a provision frequently invoked in cases of deepfake non-consensual pornography.[8] Furthermore, provisions under the BNS concerning forgery, defamation, and the creation of enmity between different groups have been utilized to prosecute those who create deepfakes intended to incite communal violence or damage reputations.

However, the rapid dissemination capabilities of social media rendered post facto criminal prosecution insufficient. The damage inflicted by a viral deepfake is often irreversible by the time law enforcement identifies the perpetrator.

B. The Shift to Intermediary Accountability: The 2021 Rules to the 2026 Amendments

To address the inadequacy of traditional penal laws, the Indian government shifted its regulatory focus from the creators of deepfakes to the platforms that host them. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, established a framework requiring Significant Social Media Intermediaries (SSMIs) to exercise due diligence. Rule 3(1)(b) specifically mandated that platforms must not host content that impersonates another person.[9]

As deepfake technology advanced, the Indian government recognized that the 2021 rules were insufficiently robust. Following a series of high-profile deepfakes targeting Indian actors and politicians in late 2023 and 2024, the Ministry of Electronics and Information Technology

[7] The Information Technology Act, 2000 (Act 21 of 2000), s. 66D.
[8] Id., s. 66E.
[9] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, rule 3(1)(b).

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

(MeitY) issued stringent advisories, warning platforms that failure to remove deepfakes within 24 hours would result in the loss of their "safe harbour" immunity under Section 79 of the IT Act.[10]

This aggressive administrative posture culminated in the seminal Information Technology (Intermediary Guidelines) Amendment Rules, 2026.[11]

C. The 2026 Amendment Rules: A Paradigm of Stringency

The 2026 Amendments represent one of the world's most stringent legal regimes for synthetic media. The rules depart from traditional platform neutrality, transforming intermediaries from passive hosts into active gatekeepers.

Key provisions of the 2026 Rules include:

1. **Mandatory Labelling of SGI:** Platforms are legally obligated to detect and prominently label all Synthetically Generated Information. Failure to apply watermarks or labels to AI-generated content triggers immediate regulatory penalties.

2. **Ultra-Fast Takedown Mandate:** The traditional 24-hour grievance redressal window was dramatically compressed. For deepfakes concerning non-consensual sexual content or high-profile political disinformation likely to cause public disorder, platforms are mandated to effectuate a takedown within two hours of receiving a complaint or a government directive.[12]

3. **Proactive Monitoring Obligations:** The rules implicitly require platforms to deploy automated AI-driven detection tools to identify deepfakes proactively, a significant departure from the traditional "notice-and-takedown" model.

D. Critique of the Indian Model

India's model is highly effective in prioritizing victim remediation and mitigating the rapid viral spread of disinformation. By threatening the loss of safe harbour, the state successfully forces multinational tech conglomerates to comply with domestic laws.

However, this approach is not without substantial legal criticism. Scholars argue that the 2026 Rules induce a "chilling effect" on free speech.[13] Faced with the threat of losing safe harbour and facing criminal liability, intermediaries are economically incentivized to over censor. Automated detection tools are notoriously imperfect and often flag legitimate satire, parody, or political commentary as malicious deepfakes. By privatizing censorship and granting platforms

[10] Ministry of Electronics and Information Technology, "Advisory to Intermediaries on Deepfakes and Misinformation", *available at:* https://www.meity.gov.in (last visited on May 13, 2026).
[11] The Information Technology (Intermediary Guidelines) Amendment Rules, 2026.
[12] Ibid.
[13] Nikhil Pahwa, "The Chilling Effect of India's New IT Rules on Free Speech" 12 Internet Freedom Foundation Journal 45 (2026)

the power to determine the boundaries of acceptable speech within a two hour window, the Indian model risks undermining democratic discourse in the name of security.

IV. The United States Approach: The First Amendment and Fragmented Governance

The regulatory posture of the United States stands in stark contrast to India's centralized stringency. Deepfake regulation in the US is deeply constrained by the First Amendment of the

Constitution, which provides unparalleled protection for freedom of speech. Consequently, the US approach is decentralized, heavily reliant on a patchwork of state-level laws, federal agency advisories, and the voluntary self-regulation of tech monopolies.

A. The Constitutional Constraint: The First Amendment

In American jurisprudence, the government cannot ban speech simply because it is false. The Supreme Court of the United States, in cases such as *United States v. Alvarez*, has held that false statements of fact do not, by themselves, fall outside the protection of the First Amendment unless they cause a legally cognizable harm, such as fraud or defamation.[14]

Therefore, any federal legislation attempting to outright ban deepfakes would face "strict scrutiny" from the judiciary. The law would need to serve a compelling state interest and be narrowly tailored to achieve that interest without unnecessarily infringing on protected speech. This high constitutional bar has severely paralyzed comprehensive federal legislative efforts, as lawmakers struggle to draft definitions of deepfakes that capture malicious intent while exempting constitutionally protected parody, art, and satire.

B. The Absence of Comprehensive Federal Legislation

Due to these constitutional hurdles, the US lacks a unified federal statute akin to India's IT Rules. Federal efforts have largely been piecemeal. The Deepfake Task Force Act and various iterations of the DEEPFAKES Accountability Act have been introduced in Congress, aiming to mandate watermarks and alter Section 230 of the Communications Decency Act (the US equivalent of safe harbour).[15] However, these bills have frequently stalled in the legislative process due to free speech concerns and heavy lobbying by the tech industry.

Instead, federal action has been largely executive. Recent Executive Orders on Artificial Intelligence have directed federal agencies, such as the Department of Commerce and the National Institute of Standards and Technology (NIST), to develop standards for watermarking and content provenance.[16] However, these standards are largely voluntary guidelines for the industry, lacking the punitive enforcement mechanisms seen in India.

C. The State-Level Patchwork

[14] United States v. Alvarez, 567 U.S. 709 (2012).
[15] Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability (DEEPFAKES) Act, 2019 (H.R. 3230, 116th Cong.).
[16] Executive Order No. 14110, 88 Fed. Reg. 75191 (2023).

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

In the absence of federal preemption, individual US states have become the laboratories for deepfake legislation, resulting in a highly fragmented legal landscape. State laws primarily target two specific harms: non consensual synthetic pornography and election interference.

- **Non-Consensual Pornography:** States like Virginia and California have amended their existing "revenge porn" statutes to explicitly include digitally created or altered images, making it a criminal offence to distribute synthetic sexual imagery without consent.[17]
- **Election Interference:** Texas pioneered electoral deepfake regulation with a law criminalizing the creation and distribution of a deepfake video intended to injure a political candidate or influence an election within 30 days of voting.[18] California followed suit with similar laws, focusing on the deceptive intent of the creator.

D. Critique of the US Model

The American model prioritizes the protection of civil liberties and fosters technological innovation by avoiding heavy-handed state intervention. However, its fragmented nature is a severe liability in the borderless digital ecosystem.

A deepfake created in Eastern Europe and hosted on a server in California can cause severe reputational damage to an individual in Texas. The patchwork of state laws makes jurisdictional enforcement incredibly complex for victims seeking civil remedies. Furthermore, relying on the voluntary self-regulation of technology platforms allowing companies like Meta, Google, and X to set their own terms of service regarding synthetic media essentially outsources the governance of truth to private, profit-driven entities.[19] This market-driven ethos often prioritizes user engagement (which is heavily driven by sensationalized synthetic media) over the rapid removal of disinformation.

V. Comparative Analysis: Bridging the Regulatory Digital Divide

The comparative analysis of India and the United States reveals a profound digital divide in deepfake governance. This divide is characterized by a fundamental trade-off: India sacrifices a degree of free expression to guarantee rapid accountability and societal stability, whereas the United States tolerates a higher degree of societal risk and disinformation to preserve absolute freedom of speech and market autonomy.

A. The Intermediary Liability Divide

[17] California Penal Code (West 2024), s. 647(j)(4).
[18] Texas Election Code Annotated (West 2024), s. 255.004.
[19] Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech" 131 Harvard Law Review 1598 (2018).

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

The starkest contrast lies in the treatment of digital intermediaries. India utilizes the threat of revoking safe harbour as a blunt instrument to ensure compliance. The 2026 Rules force platforms to act as quasi-judicial bodies, assessing the legality of synthetic media within hours.

Conversely, the United States continues to shield platforms through Section 230 of the Communications Decency Act, which generally prevents platforms from being held liable for the content created by their users. While there is growing bipartisan support in the US to reform Section 230, it currently remains a formidable barrier to the type of platform accountability mandated in India.

B. The Problem of Forum Shopping and Jurisdictional Arbitrage

This regulatory divergence creates a massive vulnerability in global cyberspace. Malicious actors engage in "jurisdictional arbitrage" or forum shopping.[20] An organized disinformation campaign targeting Indian elections can deliberately utilize servers and platforms based in the United States. While the Indian government may issue a 2-hour takedown notice, United States based platforms may resist compliance, citing the First Amendment and the absence of a reciprocal legal obligation under US law.

This asymmetry renders domestic laws inherently inadequate. Deepfakes do not respect national borders; therefore, a fragmented approach where one nation strictly regulates while another permits unregulated dissemination ensures that the epistemic threat remains active globally.

VI. The Path Forward: Towards a Harmonized Global Framework

Bridging the digital divide between the Indian accountability paradigm and the American free expression paradigm is essential to safeguarding the integrity of reality. Total regulatory homogenization is impossible due to immutable constitutional differences. However, the

solution lies not in harmonizing the *laws regarding speech*, but in harmonizing the *technical standards of reality*.

A. Shifting from Content Censorship to Content Provenance

Both nations must pivot from reactive, content-based censorship (which angers US free-speech advocates and overwhelms Indian regulators) toward proactive content provenance. Provenance focuses on establishing the origin and history of a digital file rather than judging its truthfulness.

The international legal framework should universally mandate adherence to technical standards such as the Coalition for Content Provenance and Authenticity (C2PA).[21] C2PA utilizes secure,

[20] Andrew K. Woods, "Litigating Data Sovereignty" 128 Yale Law Journal 328 (2018).
[21] Coalition for Content Provenance and Authenticity, "C2PA Technical Specification Version 1.3", *available at:* https://c2pa.org/specifications/ (last visited on May 13, 2026).

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

cryptographically bound metadata to track the origin of an image or video from the moment it is captured by a camera or generated by an AI.

If this standard is adopted globally:

1. Genuine media will possess an unbroken, cryptographic "chain of custody."
2. Synthetic media will inherently carry a cryptographic watermark detailing the AI model used to generate it.

B. A Balanced Global Treaty on Synthetic Media

A new international treaty or an addendum to the Budapest Convention on Cybercrime is necessary to bridge the US-India divide. This framework should establish:

1. **Universal Definitions:** A shared legal definition of malicious synthetic media, distinguishing it from legitimate artistic expression.
2. **Mutual Legal Assistance Treaties (MLAT) Reform:** Fast-tracking MLAT requests specifically for the attribution and identification of deepfake creators, bypassing the current bureaucratic delays that allow disinformation to go unpunished.
3. **Tiered Intermediary Obligations:** A compromise between the US and Indian models. Platforms should not be strictly liable for all user content, but they must be held liable if they intentionally strip cryptographic provenance metadata or fail to implement baseline AI detection tools.

By mandating technical transparency rather than engaging in ideological battles over censorship, legal systems can empower citizens to make informed decisions about the media they consume. In this proposed paradigm, the "Liar's Dividend" is neutralized; a politician cannot claim a genuine video is a deepfake if the video possesses an unbroken, mathematically verifiable cryptographic signature proving it was recorded on a specific smartphone at a specific time.

VII. Conclusion

The proliferation of deepfake technology represents a watershed moment in the evolution of human communication and international law. The ability to seamlessly fabricate reality threatens the evidentiary foundations of justice, the integrity of democratic elections, and the basic epistemic trust required for a functioning society.

This paper has demonstrated, the current global response is deeply fractured. The Republic of India has courageously prioritized accountability, utilizing the Information Technology Amendment Rules, 2026, to force digital platforms to swiftly remediate the harms of synthetic media. However, this state-led approach risks encroaching upon free expression through overly

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

broad censorship. Conversely, the United States has allowed constitutional constraints and market reliance to paralyze federal action, resulting in a fragmented state-level patchwork that provides inadequate protection against a borderless digital threat.

Bridging this digital divide is an urgent legal imperative. Isolated domestic legislation is futile against an internet architecture that inherently routes around localized restrictions. To reclaim digital sovereignty and protect the integrity of reality, the international community must synthesize the strengths of both paradigms. This requires embracing the American ethos of technical innovation by investing in cryptographic provenance and verifiable metadata, while simultaneously adopting the Indian ethos of holding powerful technological intermediaries accountable for the architectures they build. Only through a harmonized, transparency first legal framework can the global legal order survive the truth crisis of the synthetic age.