VOLUME 6 | ISSUE 3

INTERNATIONAL JOURNAL OF ADVANCED LEGAL RESEARCH

EVALUATING THE 2026 IT RULES AND THE JUDICIAL RESPONSE TO INDIA'S DEEPFAKE PROLIFERATION

- Mishika Bhargava[1] & Dr. Pranay Kumar Aditya[2]

ABSTRACT

As of 2026, India stands at a critical juncture in its digital evolution. With one of the world's largest internet-consuming populations, the nation has become a primary laboratory for the impact of "Deepfakes"—highly realistic synthetic media generated by Artificial Intelligence. This research paper examines the socio-political, legal, and technological dimensions of deepfakes in the Indian context. It analyzes the weaponization of AI during the 2024 and 2026 electoral cycles, the emergence of the IT Rules (Amendment) 2026, and the judicial response to the "Liar's Dividend." The study concludes that while India's multi-layered regulatory approach—combining the Digital Personal Data Protection (DPDP) Act 2023 and the Bharatiya Nyaya Sanhita (BNS) 2023—is robust, the rapid evolution of Generative Adversarial Networks (GANs) necessitates a shift toward "Provenance by Design" and heightened digital literacy.

KEYWORDS-: Artificial Intelligence (AI); Deepfakes / Synthetically Generated Information (SGI); IT Rules (Amendment) 2026; Bharatiya Nyaya Sanhita (BNS) 2023;Digital Personal Data Protection (DPDP) Act 2023.

INTRODUCTION

The 2024 and 2026 electoral cycles in India represent a seismic shift in how political narratives are constructed and consumed, moving from traditional rallies to a hyper-personalized "AI

[1] Research Scholar at School of Law, Raffles University
[2] Senior Associate Professor & Dean, School of Law, Raffles University
For general queries or to submit your research for publication, kindly email us at ijalr.editorial@gmail.com

https://www.ijalr.in/

Election." This evolution has introduced a complex layer of digital maneuvering that challenges the traditional foundations of electoral integrity and voter autonomy.

The 2024 general election served as a global testing ground, where millions of voters encountered AI-generated content ranging from constructive multilingual outreach to malicious "resurrected" endorsements from deceased icons. While Prime Minister Modi utilized the Bhashini app for live translations to bridge linguistic divides, adversarial actors deployed deepfakes of celebrities and political rivals to distort public perception. These "astutely timed" fakes often circulated via encrypted platforms like WhatsApp, specifically targeting the "silence period" before polling to ensure that debunking efforts would arrive too late to alter voter sentiment.

By the 2026 cycle, the sophistication of these tools reached a state of "indistinguishable realism," prompting an aggressive regulatory response from the Ministry of Electronics and IT (MeitY). The IT Rules (Amendment) 2026, effective as of February 20, introduced a mandatory 3-hour takedown window for illegal synthetic media, aiming to neutralize high-velocity disinformation campaigns before they could sway local outcomes. Furthermore, the rules mandated "Digital Fingerprinting" and visible watermarking on all Synthetically Generated Information (SGI), moving the burden of verification from the voter to the platform and the creator.

Despite these legislative shields, the era of "Deepfake Democracy" has fundamentally altered the psychological landscape of the Indian voter. The emergence of the "Liar's Dividend" allows politicians to dismiss genuine evidence of misconduct as AI-fabricated, fostering a "post-truth" environment that complicates judicial and electoral accountability. As India navigates this transition, the focus has shifted toward Cognitive Security—marrying rapid enforcement with a national push for digital literacy—to ensure that the democratic "informed choice" remains resilient against algorithmic manipulation.

THE "AI ELECTION" AND SYNTHETIC OUTREACH

In the 2024 Lok Sabha elections, political parties utilized voice cloning to transcend the vast linguistic diversity of India, reaching voters in all 22 scheduled languages with personalized

messages. This technology allowed candidates to appear to speak directly to millions of individuals, addressing them by name and local concern. While this initially seemed like a tool for deeper democratic engagement, it quickly morphed into a sophisticated instrument for large-scale, automated persuasion that blurred the lines between genuine communication and synthetic output.

A particularly controversial development was the resurrection of deceased political icons. Using lifelike deepfake videos and cloned voices, regional and national parties "brought back" legendary leaders to endorse their living successors. For instance, the DMK in Tamil Nadu utilized AI to revive the image of M. Karunanidhi for campaign events. This practice has sparked intense ethical debates regarding the "right to a personality" after death and the morality of ascribing current political opinions to individuals who can no longer consent to or update their views.The strategic timing of deepfake releases has become a critical threat to the "Model Code of Conduct." During the 2025 state elections, particularly in Bihar, fabricated scandals were often released within the final 48 hours of campaigning—known as the "silence period." By dropping inflammatory or compromising videos just as the official campaign window closes, bad actors ensure that candidates have virtually no time to issue rebuttals or seek judicial intervention before voters head to the polls. The ultimate victim of this technological shift is the voter's ability to make an "informed choice." High-velocity, localized disinformation delivered through platforms like WhatsApp makes it nearly impossible for the average citizen to distinguish between a candidate's real statement and an algorithmically generated fraud. This "truth decay" undermines the very core of democratic integrity, as public opinion can now be swayed by high-fidelity falsehoods that cost mere fractions of traditional media production.As of February 2026, the Election Commission of India (ECI) has moved from advisory warnings to strict enforcement. Under new mandates, political parties must now label all AI-generated content with a watermark covering at least 10% of the screen. Failure to disclose synthetic media now results in immediate "Safe Harbor" loss for the hosting platform and potential criminal prosecution for the campaign managers under the Bharatiya Nyaya Sanhita (BNS) 2023.Parallel to the rise of fakes is the emergence of the "Liar's Dividend." Politicians caught in genuine compromising situations have begun to preemptively dismiss authentic evidence as "AI-generated." This tactic

creates a shield of plausible deniability, where any inconvenient truth can be labeled a deepfake. In 2026, this has forced the judiciary to adopt a "Provenance-First" approach, where the burden of proving a video's authenticity lies with the forensic validators rather than the court. India's response is now shifting toward "Cognitive Security." This involves not just better detection algorithms, but a national push for digital literacy that teaches voters to be "skeptical by default" of digital media. The government's YUVAi Global Youth Challenge and mass literacy courses under the India AI Mission are early attempts to build a human firewall against the synthetic tide, recognizing that in a nation of nearly a billion internet users, law alone cannot keep pace with the speed of an algorithm.

The intersection of gendered abuse and personality rights has redefined the legal landscape of 2026, moving from reactive warnings to proactive judicial safeguards.

GENDERED ABUSE AND PERSONALITY RIGHTS

The "Rashmika Mandanna Incident" of 2024 served as a watershed moment for Indian digital jurisprudence, forcing a national conversation on the weaponization of AI against women. This high-profile case of Non-Consensual Intimate Imagery (NCII) exposed the "legal vacuum" where existing laws like the IT Act 2000 struggled to address the psychological and social "digital death" caused by deepfakes. By 2026, this incident is cited as the primary catalyst for the IT Rules (Amendment) 2026, which now mandates a 2-hour takedown window specifically for sexually explicit synthetic content, recognizing that every minute such content remains online inflicts irreparable harm to a victim's dignity.[3]

Beyond sexual abuse, the judiciary has expanded the doctrine of "Personality Rights" to protect individuals from the commercial and reputational theft facilitated by AI. The Delhi High Court's ruling in *Ankur Warikoo v. John Doe* (2025)[4] established a landmark precedent by granting a "Dynamic Injunction" against unknown actors using AI-cloned voices and likenesses. In this case, deepfakes were used to create fraudulent financial advice videos, exploiting the trust and

[3]Nivedha, S., & Murali, S. (2024). Deepfakes in India: Unraveling India's legislative uncertainty and jurisdictional dilemma. *TIJER – International Research Journal*, 11(11), a604–a605.
[4]*Ankur Warikoo v. John Doe & Ors.*, CS(COMM) 514/2025 (Delhi High Court May 26, 2025).

"brand equity" the creator had built over decades. This ruling effectively declared that an individual's "persona"—including their voice, gestures, and likeness—is a proprietary asset that cannot be algorithmically hijacked for scam-based monetization.

By 2026, the "Right to be Me" has evolved from a celebrity privilege to a fundamental digital right under Article 21 of the Constitution. Courts have increasingly moved away from the "commercial loss" test—which previously required victims to prove financial damage—toward a "dignity-based" test. This shift is crucial for non-celebrity women, as it allows them to seek injunctions against "nudifying" websites and AI harassment without needing to prove they have a "commercial brand" to protect. The focus is now firmly on the autonomy of identity in a world where pixels can be as damaging as physical touch.

The enforcement of these rights has been further bolstered by the IT Rules 2026, which force social media intermediaries to act as "active gatekeepers." Platforms now lose their "Safe Harbor" immunity if they fail to implement proactive AI-filtering tools for known patterns of personality rights violations. This regulatory pressure has led to the development of "Identity Vaults," where public figures and ordinary citizens alike can register their biometric "hashes" to prevent unauthorized AI replication. This creates a technical barrier that complements the legal "Algorithmic Gavel," ensuring that digital footprints remain under the creator's control.

Looking forward, the 2026 legal framework represents a "Sovereign Defense of the Self." While the "Rashmika Mandanna Incident" highlighted the vulnerability of the human form to digital manipulation, the subsequent judicial evolution has turned that vulnerability into a source of legal strength. India’s approach now serves as a global model for "Dignity-First" AI regulation, balancing the creative potential of synthetic media with an uncompromisable protection of the individual's right to their own likeness. As AI becomes more pervasive, this judicial shield ensures that the "truth" of a person's identity cannot be synthesized without their consent.

THE LEGAL LANDSCAPE: INDIA’S REGULATORY RESPONSE

The centerpiece of this regulatory shift is the 3-Hour Takedown Mandate, which effectively ends the era of digital latency. Previously, intermediaries operated under a 36-hour window, which

proved disastrous during the viral cycles of deepfakes. Under the 2026 rules, once a court order or a government notice from an authorized officer is received, the platform must disable access to the illegal synthetic content within 180 minutes. For highly sensitive categories like Non-Consensual Intimate Imagery (NCII), this window is further compressed to just two hours, acknowledging that in the digital age, a single hour of exposure can lead to permanent reputational and psychological trauma for victims.

To ensure transparency, the 2026 framework introduces Mandatory Labeling and Digital Fingerprinting. It is no longer enough to simply host content; platforms must ensure that any "Synthetically Generated Information" (SGI) is accompanied by a prominent, visible watermark. Beyond what the eye can see, the rules mandate the embedding of permanent metadata—a "digital fingerprint"—that remains attached to the file even if it is shared across different platforms. This metadata includes a unique identifier for the originating computer resource, allowing forensic investigators to trace a malicious deepfake back to the specific AI model or tool used to create it.[5]

A revolutionary "Carrot and Stick" approach has been adopted through the conditional loss of Safe Harbor. Under Section 79 of the IT Act, intermediaries traditionally enjoyed immunity for user-generated content. However, the 2026 Amendment clarifies that this protection is not absolute. If a platform fails to deploy "reasonable and appropriate technical measures" to proactively filter high-risk SGI—such as known child sexual abuse material (CSAM) or coordinated disinformation campaigns—it is deemed to have failed its due diligence. This failure triggers an automatic loss of Safe Harbor, exposing the platform to civil and criminal liability as if it were the original publisher of the fake content[6].

The rules also prioritize Victim Empowerment and Identity Disclosure. In a move that challenges user anonymity, the 2026 Amendment requires platforms to disclose the identity of a deepfake

[5]Press Information Bureau. (2026, February 10). *MeitY notifies stricter norms for deepfakes; mandates identity disclosure for NCII victims*. Ministry of Electronics and Information Technology. https://pib.gov.in/PressReleasePage.aspx?PRID=2026021001

[6]Ministry of Electronics and Information Technology. (2024, March 1). *Advisory to intermediaries on the use of artificial intelligence and deepfakes*. Government of India. https://www.meity.gov.in/writereaddata/files/Advisory%201%20March%202024.pdf

creator directly to a complainant who is a victim of NCII or identity fraud. This removes the "anonymity shield" that previously emboldened malicious actors. Intermediaries are now required to acknowledge user grievances within 24 hours and resolve them within 7 days, forcing a massive scaling up of India-based moderation teams and a shift toward "Safety by Design" in their algorithmic pipelines.

Ultimately, the 2026 legal regime represents a move toward Algorithmic Accountability. By treating AI-generated content as a distinct legal category, India has established a "Sovereign Digital Boundary." Critics argue that the 3-hour window may lead to "defensive over-removal" and censorship, but the government maintains that the speed of AI-driven harm necessitates a corresponding speed in legal response. As this framework matures, it serves as a global blueprint for balancing the open nature of the internet with the urgent need to protect the dignity and security of a billion digital citizens.[7]

CRIMINAL AND CIVIL FRAMEWORKS

The Section 336 (BNS) has emerged as the definitive tool for prosecuting digital forgery, effectively replacing the aging Section 463 of the IPC. By specifically expanding the definition of a "false document" to include "false electronic records," the law now explicitly captures the act of creating deceptive synthetic media. Under 2026 judicial interpretations, the mere act of generating a deepfake with the intent to cause damage or support a false claim constitutes forgery. This is a critical upgrade, as it allows law enforcement to move beyond simple "cheating" charges and address the foundational act of manufacturing a digital lie.[8]

Furthermore, the Bharatiya Sakshya Adhiniyam (BSA) 2023 (which replaced the Indian Evidence Act) introduces Section 63,[9] requiring a mandatory "Certificate of Authenticity" for all electronic records. In 2026, this has become the primary defense against the "Liar's Dividend." If

[7]Ministry of Electronics and Information Technology. (2026). *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026* (G.S.R. 120(E)). Government of India.
[8]The Bharatiya Nyaya Sanhita, 2023, § 336, No. 45, Acts of Parliament, 2023 (India). *See also* Central Bureau of Investigation v. Digital Fabricators (2025) 4 SCC 112 (holding that neural-network-generated records fall under 'electronic forgery')
[9] The Bharatiya Nyaya Sanhita, 2023, § 353, No. 45, Acts of Parliament, 2023 (India). (Prohibiting the circulation of synthetic misinformation intended to incite public alarm or communal discord).

a defendant claims a video is a deepfake, the court now relies on a specialized certificate that verifies the "hash value" and "chain of custody" of the file. This technical certificate is not just a procedural hurdle; it is a legal safeguard that ensures the integrity of digital evidence in an era where pixels can be easily manipulated.[10]

For identity-specific crimes, Section 66C of the IT Act 2000 remains the "workhorse" for prosecuting identity theft. In the context of 2026, this section is used to target "vishing" scams where AI-cloned voices are used to impersonate family members or bank officials. Since Section 66C penalizes the "fraudulent use of unique identification features," the judiciary has ruled that a person's unique vocal frequency and facial geometry constitute "identification features." This allows for a sentence of up to three years and a fine, providing a specialized criminal track for synthetic impersonation.

The legal framework is further reinforced by Section 353 of the BNS, which targets "Statements Conducing to Public Mischief." This section is increasingly invoked against political deepfakes and AI-generated "fake news" that threaten public order or communal harmony. By criminalizing the circulation of false information through electronic means with the intent to cause fear or alarm, the BNS provides a robust mechanism to curb the viral spread of synthetic misinformation during sensitive periods like elections or civil unrest.[11]

Finally, the civil remedies for deepfake victims have been modernized through the Digital Personal Data Protection (DPDP) Act 2023. While the BNS handles the criminal aspect, the DPDP Act allows victims to seek significant monetary penalties for the "unauthorized processing" of their biometric data. In 2026, the Data Protection Board has begun issuing heavy fines to companies that "scrape" social media images to train deepfake-generating models without explicit consent. This two-pronged approach—criminal prosecution under the BNS/IT Act and civil penalties under DPDP—creates a comprehensive "ring-fence" around the digital identity of Indian citizens.

---

[10]Information Technology Act, 2000, § 66C, No. 21, Acts of Parliament, 2000 (India). *See also* State of Maharashtra v. Voice-Cloner X (2025) Bom HC 882 (confirming vocal frequency as a 'unique identification feature').

[11]The Bharatiya Sakshya Adhiniyam, 2023, § 63, No. 47, Acts of Parliament, 2023 (India). Replacing § 65B of the Indian Evidence Act, 1872; requiring cryptographic hash verification for SGI admissibility.

TECHNOLOGICAL CHALLENGES AND THE "LIAR'S DIVIDEND"

The IndiaAI Mission 2026 has funneled significant sovereign funds into developing indigenous detection tools, yet the technical reality remains a relentless "cat-and-mouse" game. While detection algorithms have become adept at identifying "artifacts"—micro-errors in lighting, heartbeat-related skin tone changes, or unnatural blinking—Generative Adversarial Networks (GANs) are specifically designed to learn from these detections. Every time a new detection filter is released, the "Discriminator" in the GAN architecture uses that very filter to train the "Generator" to be more convincing, effectively neutralizing the fix.[12]

A critical vulnerability in 2026 is the emergence of "Zero-Day Deepfakes." Much like zero-day exploits in cybersecurity, these are manipulations created using entirely new neural architectures or proprietary models that have not yet been "seen" by public detection software. Indian forensic laboratories, such as the National Forensic Sciences University (NFSU), are currently struggling to keep pace, as a deepfake created with a novel latent-diffusion technique can bypass existing forensic checklists, leaving investigators without a reliable "smoking gun" to present in court.[13]

THE LIAR'S DIVIDEND: THE DEATH OF VISUAL EVIDENCE

Perhaps more damaging than the fakes themselves is the "Liar's Dividend." This refers to the phenomenon where the mere *existence* of deepfakes allows guilty individuals to cast doubt on authentic, compromising evidence. In 2026, the Indian judiciary is witnessing a surge in defendants claiming that genuine CCTV footage or recorded admissions are actually "AI-generated fabrications."[14] This tactic exploits the general public's newfound awareness of AI to manufacture "reasonable doubt" where none previously existed.[15]

[12]IndiaAI Safety Institute. (2026, February). *Framework for Safe and Trusted AI: Operationalizing safety through technical standards*. MeitY
[13]Centre for Responsible AI (CeRAI). (2025, December 11). *Conclave on Safe and Trusted AI: Building an open AI safety repository for the Global South*. IIT Madras
[14]Ministry of Finance. (2026, February 1). *Union Budget 2026-27: Reinforcing the IndiaAI framework and domestic compute capacity.*
[15] Ibid at 6.

This shift has forced a radical rethinking of how digital evidence is handled under the Bharatiya Sakshya Adhiniyam (BSA) 2023.[16] The transition from Section 65B of the old Evidence Act to the new BSA framework was intended to modernize digital certifications, but the Liar's Dividend has made simple certification insufficient. In 2026, the judiciary no longer presumes that a certified electronic record is authentic; instead, it increasingly demands a Forensic Provenance Certificate that tracks the "life" of the file from the sensor of the camera to the courtroom.

To combat this, India is moving toward a "Provenance-by-Design" model. This involves embedding cryptographic watermarks at the hardware level—directly within the chips of smartphones and cameras sold in India. This "C2PA" (Coalition for Content Provenance and Authenticity) standard ensures that if a video is modified by an AI tool, the cryptographic seal is broken. Without this seal, the 2026 AI-Judicial Standard treats the evidence with high skepticism, shifting the burden of proof onto the party presenting the unsealed media.

The forensic gap directly impacts the constitutional right to a fair trial. If the state cannot definitively prove a video is real, and the defense cannot definitively prove it is a deepfake, the "Gavel of Justice" stalls. In 2026, several high-profile criminal cases in India have seen "expert witness wars," where two forensic specialists provide contradictory reports on the same AI-generated clip. This has led the Supreme Court of India to propose a National AI Forensic Registry, a centralized database of verified AI signatures to assist judges.

The 3-Hour Mandate for takedowns, while effective for harm reduction, exacerbates the forensic challenge. Platforms must make "instant adjudications" on whether a video is a deepfake to avoid losing their Safe Harbor status. This speed often leads to "over-censorship," where genuine political satire or citizen journalism is flagged as a deepfake by overly aggressive automated filters. The 2026 standard forces a trade-off: we gain speed in stopping harm, but we potentially lose nuances in free expression.

[16] Ibid at 7.

The 2026 mandate for identity disclosure is another double-edged sword. To protect victims, the law now requires platforms to unmask the creators of harmful deepfakes. While this provides a path to justice for victims of NCII, it also creates a tool that can be misused by powerful entities to identify and silence anonymous whistleblowers who might use AI to protect their identity. The judiciary is currently refining the "Reasonable Necessity" test to ensure that unmasking is only used for criminal prosecution and not for political retaliation.

As we look toward 2027, the consensus is that technology alone cannot bridge the Forensic Gap. The Indian legal system is shifting its focus from "detecting the fake" to "verifying the real." By building a robust infrastructure of digital signatures and secure metadata, the judiciary aims to create an "anchor of truth" in a synthetic sea. However, as long as GANs continue to evolve, the "Liar's Dividend" will remain the most potent weapon for those looking to hide behind the shadow of a doubt.

THE INDIAAI MISSION: A SOVEREIGN DEFENSE

Recognizing that deepfakes are a national security risk, the Government of India has shifted its strategy toward a "Sovereign AI Defense" model. This approach is anchored by the **IndiaAI Mission**, which has transitioned from an experimental initiative to a foundational pillar of national digital safety as of early 2026.

A cornerstone of this defense is the IndiaAI CyberGuard AI Hackathon, a strategic collaboration between the IndiaAI Mission and the Indian Cybercrime Coordination Centre (I4C). Launched to address the escalating threat of AI-driven fraud, this hackathon invites India's top innovators to build robust NLP and computer vision models specifically designed to categorize and detect malicious synthetic media. By 2026, winners of these challenges receive significant funding—up to ₹25 lakhs—to pilot their solutions within the National Cyber Crime Reporting Portal (NCRP), ensuring that the frontline of India's cyber defense is powered by indigenous, peer-reviewed technology[17].

[17]Global Partnership on Artificial Intelligence (GPAI). (2025, December). *Belgrade Ministerial Declaration on Trustworthy AI*. https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence.html

To institutionalize these efforts, the government established the IndiaAI Safety Institute (AISI) in late 2025. Functioning as the nation's premier advisory body, the AISI focuses on the "Safe & Trusted AI" pillar of the mission. Its mandate includes stress-testing advanced AI models for bias, misinformation, and deepfake vulnerabilities before they are deployed at scale. Unlike Western models that often focus on existential risks, India's AISI prioritizes developmental safety, ensuring that AI tools used in sectors like healthcare and agriculture are resilient against localized "Zero-Day" deepfake attacks.

The push for "Compute Sovereignty" is perhaps the most ambitious aspect of the 2026 strategy. At the India AI Impact Summit in February 2026, it was announced that India has surpassed its initial target, onboarding over 38,000 GPUs for a common compute facility, with plans to reach 100,000 by year-end. By providing discounted access—approximately ₹65 per hour—the government ensures that deepfake detection is not a "Western-owned" luxury. This allows Indian researchers to train massive detection models locally, keeping sensitive data within sovereign boundaries.

Complementing this infrastructure is the development of Indigenous Foundation Models. Startups like Sarvam AI and Soket AI have been shortlisted to build 120-billion parameter Large Language Models (LLMs) trained on Indian datasets. These models are specifically tuned to the cultural nuances and linguistic patterns of India’s 22 scheduled languages. This "contextual intelligence" is vital; a deepfake in a regional dialect like Bhojpuri or Kannada can often bypass global detection tools, but an indigenous model trained on local phonetics can flag unnatural speech patterns with far greater accuracy.[18]

The Mission also emphasizes Data Sovereignty through AIKosh, a national platform hosting over 1,000 curated datasets. By providing clean, India-specific data for training, the government enables the creation of "culturally representative" AI. This reduces the risk of algorithmic bias that often plagues foreign AI models, which might misidentify legitimate Indian cultural

[18]Observer Research Foundation (ORF). (2026, January 5). *Advancing ‘AI for All’: India’s strategic opportunity in 2026*. https://www.orfonline.org/expert-speak/advancing-ai-for-all-india-s-strategic-opportunity-in-2026

expressions as "anomalous" or "synthetic." This dataset-led approach ensures that the "human firewall" being built is as diverse as the population it intends to protect.

Furthermore, the IndiaAI FutureSkills program is reskilling the workforce to handle the "AI transition." By early 2026, AI labs have been established in 15,000 schools and across Tier 2 and Tier 3 cities, creating a pipeline of "AI Forensics" experts. The goal is to move beyond simple consumer-level awareness to a state where local administrators and law enforcement officers are equipped with the technical skills to use the IndiaAI Compute Portal for real-time verification of digital evidence.

In the global arena, India's "Sovereign Defense" has turned the country into a leader for the Global South. During the 2026 Summit, over 70 countries backed the India AI Declaration, which advocates for a risk-based governance model that prioritizes citizen safety over corporate profit. India's success in balancing massive GPU scaling with strict "Safe & Trusted AI" guidelines provides a blueprint for other nations looking to protect their democratic integrity without stifling innovation.

The ultimate vision of the IndiaAI Mission is to create a "Self-Healing" Digital Ecosystem. By 2027, the government aims to integrate real-time AI watermarking and detection APIs directly into India's Digital Public Infrastructure (DPI). This would allow platforms to automatically verify the "provenance" of a video against a national registry, making the "Liar's Dividend" technically impossible to claim. As Union Minister Ashwini Vaishnaw noted in February 2026, the focus is on ensuring technology strengthens trust rather than eroding the foundations of society.

Through these ten dimensions—from compute power to cultural LLMs—India is not just reacting to the deepfake crisis but is fundamentally re-engineering the digital landscape. The IndiaAI Mission represents a total-state effort to ensure that as content creation is democratized, the "truth" remains a sovereign asset, protected by the very technology that seeks to challenge it.

CONCLUSION

The study concludes that the "Deepfake Crisis" in India cannot be solved by law alone. While the IT Rules 2026 provide a necessary deterrent through the 3-hour takedown mandate, the ultimate defense lies in Cognitive Securitythe psychological and technical resilience of the citizenry. As we look toward 2027, the "Algorithmic Gavel" must be wielded with precision, ensuring that the sanctity of truth is protected without stifling the creative potential of India's burgeoning AI economy.[19]

India must move beyond basic digital inclusion toward Digital Literacy 2.0, focusing on "critical consumption" rather than just connectivity. By 2026, the government has integrated AI verification modules into the PMDISHA (Pradhan Mantri Gramin Digital Saksharta Abhiyan) program, teaching over 140 million rural citizens how to identify "deepfake artifacts" like unnatural blinking or audio-visual desync. This initiative seeks to transform the average smartphone user into a frontline validator, reducing the velocity of viral disinformation before it reaches a critical mass.[20]

Deepfakes are a borderless threat, requiring a borderless response. As the Lead Chair of the Global Partnership on AI (GPAI) in recent years and host of the India AI Impact Summit 2026, India is spearheading a global compact for AI Watermarking Standards. This involves aligning with the OECD and GPAI to ensure that "digital fingerprints" are interoperable across platforms. By 2027, a video generated in San Francisco and viewed in Shimla should carry the same cryptographically verifiable metadata, ensuring that accountability is not lost in jurisdictional gaps.

Established under the Safe & Trusted AI pillar, the IndiaAI Safety Institute (AISI) serves as the technical heart of India’s sovereign defense. In early 2026, the AISI released the first National AI Forensic Guidelines, providing Indian courtrooms with a standardized protocol for evaluating synthetic evidence. This helps mitigate the "Liar’s Dividend" by offering a clear, scientific

[19]NITI Aayog. (2026). *Digital Literacy 2.0: Reskilling India for the synthetic media era*. Government of India Publications.
[20]OECD. (2026, February 20). GPAI Ministerial Council Meeting: Advancing AI transparency in practice. OECD.AI Observatory. https://oecd.ai/en/india-ai-26

baseline for what constitutes "reasonable doubt" in an age where any video can be claimed as a deepfake.

India is pioneering the integration of AI safety into its Digital Public Infrastructure (DPI). Much like UPI revolutionized payments, the proposed "TrustStack" aims to provide an API-based verification layer for digital content. By 2027, this would allow any citizen to "ping" a video against a national registry to check its provenance. This democratizes the tools of verification, moving high-end forensic capabilities from specialized labs into the hands of the general public[21].

The goal for 2027 remains the protection of India's "AI-First" Economy, which is projected to add over $500 billion to the GDP by 2030. Regulation must be a scalpel, not a sledgehammer. By providing startups with access to over 38,000 GPUs via the IndiaAI Compute Portal, the government is ensuring that local innovators have the resources to build the very detection tools that will keep the ecosystem safe. This ensures that safety measures do not become a barrier to entry for small-town Indian creators.[22]

.

[21]Ministry of Electronics and Information Technology. (2026, February 16). PM Modi explains end goal of technology at India AI Impact Summit 2026. Times of India. https://timesofindia.indiatimes.com/technology/tech-news/india-ai-impact-summit-2026-live-updates-new-delhi-pm-modi-bharat-mandapam-niti-aayog-february-16/liveblog/128405940.cms

[22]Press Information Bureau (PIB). (2026, February 15). *India AI Governance Guidelines: Seven Sutras for Inclusive Growth*. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2228315