

DEEPPAKES, MISINFORMATION AND THE NEED FOR STRONGER INDIAN LAWS

By Sushanth Chikkatur

Christ (Deemed to be University), Bengaluru

I. INTRODUCTION

The advent of Artificial Intelligence (AI) has democratized creativity, but it has simultaneously weaponized deception. In January 2026, a video was uploaded on social media sites showcasing Sundararaman Ramamurthy, Chief Executive of Bombay Stock Exchange, giving advice to investors on which stock to buy. It was also called the **Bombay Stock Exchange (BSE) Incident**¹. However, it was a synthetic fabrication created using AI. "It was in the public domain where many people could see it, and get cheated into buying or selling stocks, as if I'd recommended them," explained Ramamurthy. The unquantifiable financial detriment caused to retail investors by such market manipulation remains a severe regulatory blind spot. And it was even more so due to Jamtara 2.0. It is an evolved version of phone-call phishing into high tech, industrial crime. It is based on the Jamtara, a district of Jharkhand, which is known for online-phishing and cybercrimes. The fraudsters use open-source AI models to create synthetic videos to manipulate online video KYC, voice cloning, identity theft etc., and these are what we call '**Deepfakes**'.

Misinformation on the other hand is false, inaccurate or mistaken information that is shared without harmful intent. The person sharing this misinformation genuinely believes it to be true and often send it to help, inform or warn others. In spring and summer of 2025, there was an increase in TikTok videos, YouTube and personal testimonies about the miracle '**Pink Salt Diet Drink**', also called the 'pink-salt trick' or 'sole water', which supposedly helped lose weight very quickly. This claim was merely misinformation which people believed to be true, until the doctors stated- "If you're looking to improve your health, sole water isn't the way to do it. "We say it all the time, but the best way to care for your health is to lower your salt intake and get your vitamins and minerals through whole, unprocessed foods." (Cleveland Clinic, June 2024)².

¹NSE, BSE Issue Warnings Against Deepfake Videos of Their CEOs Offering Stock Advice, The Hindu (May 25, 2024),

² What Is Sole Water and Is It Good for You?, Cleveland Clinic (June 12, 2024)

All these incidents expose a structural crisis at the intersection of technology, democracy, and individual dignity, and they reveal a legal system entirely unprepared to meet it. This article argues that India's existing legal framework is architecturally inadequate to address the menace of deepfakes and AI-generated misinformation. It maps the underlying technology and its typology, analyses the societal consequences of the deepfake phenomenon, identifies critical lacunae in current Indian law and proposes a concrete legislative reform agenda.

II. TECHNOLOGY FUNDAMENTALS AND TYPOLOGY

The term 'Deepfake' is a portmanteau of 'deep learning' and 'fakes'. It is a piece of digital media, such as a video, image or audio recording which uses AI to alter a person's face or voice to make them appear to do or say something which they never actually did. The most widely used underlying architecture for deepfake is the Generative Adversarial Network (GAN). It is a deep learning AI architecture comprising two competing neural networks: a 'generator' that fabricates synthetic content, and a 'discriminator' that attempts to identify it as artificial. After multiple iterations and successive training, this generator improves to a level where neither a discriminator nor a human can distinguish the fabrication from reality.

Modern deepfakes are ordinarily classified into 4 different categories. Face swapped videos, in which a target's face is grafted onto another's body; voice cloning, which replicates an individual's vocal patterns with limited vocal data; full body re-enactment, in which, the entire body movement, posture and action of a source person are transferred onto the target person in a video; and text based conversational style which stimulates writing style and reasoning. Each category creates different legal challenges and India's existing statutes are not adequate to address any of them.

A viral TikTok account called '@deptomcruise'³, amassed millions of followers using face-swap technology to generate convincing videos of Hollywood Actor Tom Cruise, enacted by an impersonator called Miles Fisher. It demonstrated entertainment and reputational risks through AI generated deepfakes. In November 2023, a morphed video of Rashmika Mandanna, a Bollywood actress, went viral across Indian social media platforms within hours of its upload. The video also used AI face-swap to swap her face onto another person's body with sufficient realism, to deceive millions of viewers, triggering national outrage and urgent

³ Drew Harwell, The viral Tom Cruise deepfakes are a glimpse into the future of deception, The Washington Post (Mar. 2, 2021)

governmental statements(Rashmika MandannaDeepfake Incident)⁴.Another instance in which an AI generated footage featured late former Tamil Nadu CM, C. N. Annadurai⁵,which showed him asking for support for a newly formed political party during Tamil Nadu state elections. These deepfakes have caused various social and legal challenges, and **MeitY**ⁱ advisories have drawn attention to the detection challenges that arise as these tools proliferate without adequate regulatory oversight.

Nonetheless, we must take the current limitations of AI into considerations. Deepfakes often exhibit detectable artefacts such as unnatural blinking, lighting inconsistencies, audio-video misalignmentsetc...Shallow fakes (or cheap fakes) where media is altered using basic non-AI editing techniques, such as slowing down audio or video to make someone sound intoxicated, selective cropping or recaptioning old photos. Fabricated content which are entirely false;imposter contents where genuine sources are impersonated. Creating deepfakes requires substantial amount of time, softwareand hardware. Amateurs currently cannot create highly realistic deepfake contents. However, even these shallow fakes, risk reputational harm. An example would be the 2019 slowed-down video of United States House Speaker, Nancy Pelosi⁶, which distorted her speech patterns and was amplified by high-profile political actors, demonstrated that even low-cost manipulations produce outsized reputational harm.

III. DETECTION AND FORENSIC RESPONSE

Deepfake detection uses advanced computer vision, machine learning, and multi-artifact forensic techniques to spot various anomalies introduced by generative models like GANs and diffusion systems. These include, inspection of pixel-level discrepancies; colourdistribution; lighting mismatches; evaluating frame-to-frame coherence; tracking inconsistent motions; analysing transformed image representations to find sleekness or high-frequency grid artifacts, left behind by up sampling decoders; discrepancies occurring at the boundary where a generated or swapped face meets the original background or native skin texture.

⁴ Delhi Police files FIR in Rashmika Mandanna deepfake case, The Hindu (Nov. 11, 2023)

⁵ India's state elections see AI moving from the margins to the mainstream, Channel News Asia (Apr. 21, 2026)

⁶ Drew Harwell, Faked Pelosi videos, slowed to make her appear drunk, spread across social media, The Washington Post (May 23, 2019)

Content provenance and authentication include, provenance tracking in AI deepfakes, which is the process of recording the complete life cycle and origin of a digital file such as photos, audio clips, or videos. Instead of guessing if a file is fake after it is made, it builds a secure, unalterable chain of custody from the moment of creation. The core of provenance and authentication is the C2PAⁱⁱ and the CAI. C2PA is an open technical specification created by the Coalition for Content Provenance and Authenticity. It embeds tamper-evident, cryptographically signed metadata known as content credentials. They are embedded directly into digital images, videos, and audio and documents to track their origin, creator, and editing history. CAI is a global cross-industry group working to fight digital misinformation and fake news. This was founded in 2019 by Adobe, The New York Times, and Twitter (now X) and it supports media transparency, curbing disinformation, creator attribution, and AI training control.

IV. MISINFORMATION: DEFINITIONS, TYPES AND ECOSYSTEM

Misinformation refers to false, inaccurate, or mistaken information which is shared without the harmful intent. The person sharing this information genuinely believes that it is true. Disinformation refers to false and deliberately engineered or spread to deceive, manipulate or cause harm. Misleading content involves distortion of context where the information presented contains a core element of truth but is framed, cropped, or presented in a way that leads the audience to a false conclusion. And Malinformation is sharing of truthful, private information to specifically cause harm. During COVID-19, many viral messages circulated through social media platforms claimed that drinking hot water, consuming cow urine, eating garlic, or using specific herbal drinks would cure coronavirus. Even more so during election period, many exit-polls show misinformation and misleading content to influence people's thoughts and emotions. Disinformation such as WWII Black Propaganda⁷ in which planes dropped fake ration books, official army letters, and newspapers with altered facts. These fake items encouraged soldiers to desert, fake illnesses, or waste materials. Most recently, phishing phone calls have evolved into high-tech cybercrimes in which fraudsters use phone calls to get vocal and biometric patterns of the target individuals to misuse them.

⁷ Stanley Newcourt-Nowodworski, Black Propaganda in the Second World War (2005) (detailing the mass aerial distribution of forged ration cards, altered newspapers, and fake military orders to subvert Axis morale)

V. DEEPFAKES AS ACCELERANT OF MISINFORMATION

Deepfakes do not operate in isolation. They are the most potent accelerant in this volatile misinformation ecosystem which combine visual authenticity and emotional resonance. The phenomenon of agnotology, which is the deliberate, culturally induced ignorance or doubt found in deepfakes, is the most sophisticated contemporary instrument. Misinformation can exist without deepfakes, but deepfakes exponentially amplify misinformation.

The most viewed media - 'The Social Media', is optimized for engagement, not accuracy. It constitutes the primary information source for majority of the population. Misinformation and fakes are popularity and emotional driven. Consequently, the average user rarely verifies factual accuracy. A 2018 MIT study published in 'Science'⁸ confirmed that false news reaches people about six times faster than true stories, and falsehoods are 70% more likely to be retweeted. Fuzzy Trace Theory (FTT), drawn from cognitive psychology, provides a structural explanation for this vulnerability. It says that users process data on two levels, gist (emotional) and verbatim (detailed) and social media consumption operates almost at the gist level. And the **believability** of fakes and misinformation is therefore often greater for the crowd than the truth.

At individual level, misinformation and deepfakes cause reputational harm or psychological trauma. But at a societal level, they produce a phenomenon known as '**Liar's Dividend**'. This significantly reduces believability of the true news and in media alike as the people aren't able to verify facts properly. The BSE 2026, Tamil Nadu State Elections 2026 and the WWII's Black-propaganda incidents illustrated this starkly. The deepfakes represent the twenty-first century democratisation of this tactic, available now to state actors, criminal networks, and individuals alike.

Your One Stop Legal Destination

VI. INDIAN LEGAL FRAMEWORK: EXISTING PROVISIONS AND LACUNAE

The regulatory architecture governing synthetic media in India underwent a major overhaul with the notification of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026. Rather than relying exclusively on indirect

⁸ Soroush Vosoughi, Deb Roy & Sinan Aral, The Spread of True and False News Online, 359 Science 1146, 1146-1151 (2018)

provisions, the legal regime now explicitly defines, regulates, and penalizes unlawful synthetic media. With the combination of delegated legislation, criminal law under the Bharatiya Nyaya Sanhita (BNS), 2023, the DPDP frameworks,

1. Under the **Information Technology Act, 2000**⁹, **Section 66C** penalises identity theft through computer resources. **Section 66D** addresses cheating by impersonation through a computer resource. **Section 66E** prohibits the unauthorised capture, publication, or transmission of images violating personal privacy. **Section 69A** empowers the government to direct the blocking of online content threatening sovereignty, security, or public order.

The February 2026 Amendment strictly targets the spread of AI-generated deepfakes, online misinformation, and synthetic media. It was notified by MeitY and it took effect from 20th February 2026. This amendment officially defines Synthetically Generated Information (SGI)ⁱⁱⁱ mandated that platforms must ensure that all permitted SGI is prominently labelled with permanent metadata to trace its origin. **Significant Social Media Intermediaries (SSMIs)**^{iv} are obligated to take protective technical measures to prevent the creation or spread of unlawful synthetic content, such as **non-consensual intimate imagery (NCII)**^v and deepfake pornography; disabling access to unlawful content within 3 hours of receiving a government or court order and removing high-risk SGI within 2 hours of a complaint. Failure to adhere to these rules and timelines exposes the platform to direct civil and criminal liabilities.

2. **Under The Bharatiya Nyaya Sanhita (BNS), 2023**¹⁰, substantive criminal sanctions are imposed when SGI crosses into criminal fraud or social disruption. **Digital Forgery** penalizes the fabrication of false electronic records intended to harm reputation, directly covering defamatory deepfakes. **Cheating** penalises financial deception, particularly synthetic voice scams deployed by cybercrime rings. And **Public Mischief** criminalizes circulating statements or synthetic media intended to cause public panic, communal discord, or electoral distortion.
3. The **Digital Personal Data Protection (DPDP) Act, 2023**¹¹ regulates harvesting of personal data. Scraping an individual's facial geometry or voice recordings to train generative AI models without explicit, informed consent violates the statutory obligations of "**Data Fiduciaries**"^{vi}, attracting substantial financial penalties.

⁹ Information Technology Act, 2000, § 66C, No. 21, Acts of Parliament, 2000 (India)

¹⁰ Bharatiya Nyaya Sanhita, 2023, § 318, No. 45, Acts of Parliament, 2023 (India)

¹¹ Digital Personal Data Protection Act, 2023, No. 22, Acts of Parliament, 2023 (India)

4. Grounded in the constitutional **Right to Privacy under Article 21 (K.S. Puttaswamy v. Union of India)**¹², Indian courts protect an individual's likeness. Landmark rulings in **Titan Industries Ltd. v. Ramkumar Jewellers**¹³ and **Anil Kapoor v. Simply Life India (2023)**¹⁴ established that unauthorized commercial exploitation and synthetic cloning of an individual's persona violate both intellectual property and fundamental human dignity.
5. Until IT Amendment Rules 2026, IPR was the most effective legal tools available in India to combat the unauthorized replication of a person's identity. Personality rights (or the right of publicity) protect an individual's right to control the commercial exploitation of their name, image, voice, signature, and distinct persona. While India does not have a standalone "Right of Publicity"^{vii} statute, courts have recognized it through common law and as an extension of the Right to Privacy under Article 21 of the Constitution. And in Copyright Act, 1957, even though not directly, provides critical defence mechanisms for artists. **Sections 38A, 38B, and 57** of this act ensures that performers control how their live acts or recordings are used, and they guarantee that original authors retain lifelong moral rights over their work even after assigning the commercial copyright. In **Arijit Singh v. Codible Ventures LLP (2024)**¹⁵, The Bombay High Court ruled that mimicking a performer's vocal attributes using AI platforms without authorization infringed upon personality, publicity, and moral rights under the Copyright Act.
6. Functioning under the Ministry of Home Affairs, the **Indian Cyber Crime Coordination Centre (I4C)**^{viii} actively tracks the use of synthetic media in cybercrime and issues critical advisories. For example, it recently warned financial institutions and the public about cybercriminals utilizing AI deepfakes, voice cloning, and synthetic identities to bypass Video-KYC and facial authentication systems in the banking sector. It operates the National Cyber Crime Reporting Portal (NCRP)^{ix} and the 1930 helpline, which helps citizens to report scams through misinformation and deepfakes.
7. CERT-In^x handles the technical defence, threat assessment, and organizational preparedness against deepfakes. It advises companies to implement digital watermarking, adopt AI-driven detection tools, utilize media provenance verification, and build robust crisis management plans for deepfake incidents. It also advocates for enhanced digital forensics training to

¹² Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.

¹³ Titan Indus. Ltd. v. Ramkumar Jewellers, 2012 SCC OnLine Del 2382

¹⁴ Anil Kapoor v. Simply Life India & Ors., 2023 SCC OnLine Del 6914

¹⁵ Arijit Singh v. Codible Ventures LLP, Interim Application (L) No. 23560 of 2024 in Com IPR Suit (L) No. 23443 of 2024 (Bom. July 26, 2024)

swiftly identify AI artefacts in altered media, ensuring India's technical response matches the rapid sophistication of GANs.

Until February 2026, India did not have a dedicated statute or law penalising them. Legal action was generally available only after a case was filed and after the harm was already caused. Prevention of harm was limited. Violation of rights arises with the creation of the fakes itself, but it wasn't accounted for. However, with the introduction of the new IT rules from February 2026, we are leap-forward for platform accountability of deepfakes. But there still are very critical limitations and the statutory omissions in Indian statutes. First, a 'Fatal Flaw' in the IT Amendment Rules, 2026 must be addressed. It surely did penalise 'Creation', but it did not penalise **individual creation**. Rather it created **platform liability** by bringing in timelines and various other provision. The 2026 Rules, force platforms to use automated tools to block the creation of unlawful Synthetically Generated Information (SGI) on their servers. Therefore, these rules penalize the platform's failure to moderate, not the individual's act of creation or circulation. This repeats the original flaw with Indian laws, which only penalised publications. These new rules are merely an extension of the same publication-based violation, with superficial legislative drafting to ostensibly target the act of 'Creation'. Second, there are no appropriate rules which regulate the creation and publication of deepfake or misinformation, which could essentially cause harm to an Indian national, by foreigners and foreign media. **Section 75 of Information Technology Act, 2000** explicitly extends Indian jurisdiction over any offense or contravention committed outside India by any person, regardless of their nationality, so long as the act harms an Indian national and involves a computer, computer system, or network located in India. Theoretically this works paired with other Indian Laws, but as catching a fraudster abroad is extremely difficult, realistically, this law is still not a viable option.

Your One Stop Legal Destination

This creates a continuing reason not to abandon the notion of 'Stronger Indian Laws'. At the same time, we must acknowledge the fact that we are moving forward, and the current laws do tackle most of the issues raised by deepfakes and misinformation. The fundamental issue we now lies with the execution of these laws. If there is no proper enforcement of the laws, the statutes are technically useless. However, we must keep in mind that India, needs a standalone statutory criminal offense targeting the unauthorized creation, circulation, and work towards evolving Indian Laws tackling them.

VII. THE WAY FORWARD AND CONCLUSION

While recent intermediary rules aggressively regulate platforms, India's primary challenge lies in bridging substantive statutory gaps and resolving execution bottlenecks. First, the Parliament must enact a standalone criminal statute penalizing unauthorized synthetic replication at its inception, moving beyond mere platform compliance. Second, executive enforcement must match technological velocity. Institutions such as the I4C, CERT-In, and Grievance Appellate Committees (GACs) must be equipped with certified forensic infrastructure and automated detection tools to realistically execute the two-to-three-hour takedown mandates without resorting to algorithmic over-censorship. Third, statutory adoption of C2PA and CAI provenance standards should be mandated to cryptographically embed authenticity at the point of media generation.

Ultimately, synthetic deception threatens the constitutional guarantee of human dignity under Article 21 of the Indian Constitution. Regulating deepfakes is not merely about containing misinformation. It is about preserving individual autonomy and epistemic trust in democratic institutions. Substantive laws remain ineffective without an agile, technically sophisticated execution mechanism. India must harmonize *ex-ante* criminal deterrence with robust institutional enforcement, ensuring the legal framework functions as a proactive shield rather than a post-mortem remedy in the synthetic era.

Your One Stop Legal Destination

ⁱ MeitY stands for the Ministry of Electronics and Information Technology. It is a government branch in India that handles computer systems, the internet, and digital growth

ⁱⁱ C2PA stands for the **Coalition for Content Provenance and Authenticity**. It is an open technical standard founded in February 2021 by Adobe, Microsoft, Intel, BBC, and other tech giants

ⁱⁱⁱ Synthetically Generated Information (SGI) is audio, visual, or audio-visual media created or changed by algorithms or artificial intelligence to look or sound completely real

^{iv} Significant Social Media Intermediaries (SSMIs) are large online platforms or social media services that cross a specific user threshold of more than five million registered users in India

^v Non-consensual intimate imagery (NCII) is the sharing, distribution, or creation of nude or sexually explicit photos and videos of a person without their clear and ongoing permission. It is also known as image-based sexual abuse

^{vi} A data fiduciary is any person, business, or organization that decides why and how personal data is processed

^{vii} The right of publicity is a fundamental right under Right to Privacy, Article 21 of The Indian Constitution. It stops others from using your name, face, voice, or image for money without your permission. It lets people protect and profit from their own identity

^{viii} The Indian Cybercrime Coordination Centre (I4C) is a nodal agency established by the Ministry of Home Affairs, Government of India, in New Delhi. It provides a national framework and ecosystem for law enforcement agencies, banks, and stakeholders to prevent, detect, investigate, and tackle cybercrime comprehensively

^{ix} The National Cyber Crime Reporting Portal (NCRP) is an online platform launched by the Government of India under the Indian Cyber Crime Coordination Centre (I4C) initiative. It allows citizens to report all types of cyber crimes digitally, with a special focus on cases targeting women and children

^x CERT-In stands for the Indian Computer Emergency Response Team. It is the national nodal agency under India's Ministry of Electronics and Information Technology (MeitY) that handles cyber security threats, handles data breaches, and issues critical safety alerts.



Your One Stop Legal Destination