DEEPFAKE TECHNOLOGY, GENERATIVE AI AND PERSONAL DATA MISUSE: EXAMINING LEGAL GAPS UNDER THE DIGITAL PERSONAL DATA PROTECTION ACT, 2023

Prashant Kumar, Uttar Pradesh State Institute of Forensic Science, Lucknow, India

ABSTRACT

The exponential rise of generative artificial intelligence has ushered in a technological epoch that challenges the very foundations of personal identity, informational privacy, and democratic integrity. Among the most insidious products of this revolution is deepfake technology synthetic media fabricated through machine learning architectures, capable of placing real human faces, voices, and identities into fabricated scenarios with unnerving authenticity. In India, the misuse of deepfakes has escalated from a peripheral concern to a pressing socio-legal crisis, yet the legislative response remains markedly inadequate.

This article undertakes a rigorous doctrinal and analytical examination of the Digital Personal Data Protection Act, 2023 (DPDPA) India's foundational data privacy legislation in the context of deepfake-related personal data misuse. The central argument advanced is that the DPDPA, despite representing a significant legislative milestone, contains critical structural lacunae that leave individuals exposed to harms arising from AI-generated impersonation, non-consensual biometric cloning, synthetic identity fraud, and mass scraping of personal data for AI training purposes.

Through a comparative analysis of international regulatory frameworks including the European Union's AI Act, China's Provisions on Deep Synthesis Internet Information Services, and emerging United States state-level legislation this article identifies six principal legal gaps in the DPDPA: the absence of a definition or regulatory treatment of biometric manipulation; no oversight mechanism for AI training datasets; the lack of deepfake-specific criminal liability; weak cross-border enforcement architecture; inadequate platform accountability; and the near-total absence of a victim compensation mechanism.

The article also critically evaluates whether ancillary legislation including the Bharatiya Nyaya Sanhita, 2023, the Information Technology Act, 2000,

and the Bharatiya Sakshya Adhiniyam, 2023 can fill these gaps, concluding that they cannot without significant amendment. The article closes with a comprehensive reform model proposing a standalone Deepfake Regulation Act, mandatory AI content watermarking, an AI licensing regime, enhanced platform liability, and victim-centric redressal mechanisms.

Keywords: Deepfake, Generative AI, DPDPA 2023, Biometric Data, Personal Data, Synthetic Media, Cyber Law, AI Regulation, Privacy, Data Protection.

1. INTRODUCTION

We are living through a peculiar paradox. The same technological ingenuity that has democratised creativity allowing a student in Jaipur to generate photorealistic art, or a filmmaker in Mumbai to produce visual effects without a studio has simultaneously handed malicious actors an unprecedented toolkit for fabricating reality. The weapon in question is deepfake technology, and its proliferation represents one of the gravest threats to personal dignity, political discourse, and legal order in the digital age.

The term 'deepfake' derives from 'deep learning' and 'fake.' What began in 2017 as a Reddit sub thread where a user superimposed celebrity faces onto adult content has evolved into a sophisticated industrial capability. Today, with tools like Stable Diffusion, Runway ML, Eleven Labs, and freely available open-source models, a technically literate person can generate a convincing video of any public figure saying anything in minutes, on a consumer laptop. The barriers of expertise and cost have collapsed. The barriers of law have not kept pace.

India's engagement with this crisis has been characterised by alarming delay. The country witnessed high-profile deepfake incidents targeting Bollywood celebrities including Rashmika Mandanna and Katrina Kaif in 2023, where digitally manipulated videos circulated on social media platforms gathering millions of views before any intervention was made. Political operatives have employed AI-generated voice clones of electoral candidates during state assembly elections. Financial fraudsters have used deepfake video calls to impersonate company executives and authorise fraudulent wire transfers a crime category now termed 'CEO fraud' that has cost Indian businesses crores of rupees.

The legislative response to this crisis has been fragmented and reactive. The Digital Personal Data Protection Act, 2023, passed by Parliament after years of deliberation, was

hailed as a cornerstone of India's data governance architecture. Yet the Act does not mention deepfakes, does not regulate the scraping of personal data for AI training, does not define biometric manipulation, and offers no specific remedy to individuals whose digital likeness has been cloned without consent. This is not a minor oversight it is a foundational gap in the country's digital rights framework.

1.1 Research Questions

This article is anchored in four principal research questions that together illuminate the depth of India's regulatory vacuum:

- Does the DPDPA, 2023 regulate or prohibit the creation and dissemination of deepfakes, whether directly or by implication?
- Can personal data including photographs, voice recordings, and biometric identifiers scraped from social media be legally used for AI model training under the DPDPA without the data principal's informed consent?
- Who bears civil and criminal liability when deepfake content causes harm the creator, the platform host, the developer of the AI tool, or all three?
- Are existing Indian laws, taken together, sufficient to address the harms arising from deepfake technology, or is targeted legislative reform imperative?

1.2 Research Objectives

- To conduct a granular doctrinal analysis of the DPDPA, 2023 to identify explicit and implicit regulatory gaps concerning deepfakes and generative AI misuse.
- To evaluate whether ancillary Indian legislation criminal, evidentiary, and sectoral provides meaningful recourse to deepfake victims.
- To undertake a comparative international analysis of deepfake and AI regulation across major jurisdictions.
- To propose a coherent, evidence-based reform model suited to India's socio-legal context.

1.3 Research Methodology

This research employs a doctrinal legal methodology, supplemented by comparative analysis and empirical case study review. Primary sources include the DPDPA, 2023, the EU AI Act (2024), China's Provisions on Deep Synthesis Internet Information Services (2022), the Bharatiya Nyaya Sanhita, 2023, and the Information Technology Act, 2000. Secondary sources include peer-reviewed legal scholarship, government committee reports, and documented incidents of deepfake misuse in India and globally.

2. UNDERSTANDING DEEPFAKE TECHNOLOGY AND GENERATIVE AI

2.1 The Meaning of Deepfake

A deepfake is a piece of synthetic media video, audio, image, or text in which a real person's likeness, voice, or identity has been computationally fabricated or manipulated using artificial intelligence, to such a degree that the result is indistinguishable from authentic content to an ordinary human observer. The word carries within it both its technical genealogy ('deep learning') and its essential danger ('fake'). Unlike traditional photomanipulation or voice dubbing, which leave artefacts detectable by trained eyes, deepfakes generated by modern neural architectures can produce output that defeats visual scrutiny even from forensic experts without specialised tools.

It bears emphasis that 'deepfake' is not a neutral or inherently criminal term. The same underlying technology enables legitimate applications dubbing films into regional languages, restoring damaged historical footage, assisting individuals with speech impairments, and creating digital avatars for virtual assistants. The harm lies not in the technology per se, but in its non-consensual, deceptive, or malicious application using another person's identity without authorisation.

2.2 How Deepfakes Work: The Technical Architecture

To understand the legal problem, one must first understand the technical mechanism. Deepfakes operate at the intersection of three machine learning paradigms: Generative Adversarial Networks (GANs), diffusion models, and neural codec language models for voice synthesis.

2.2.1 Generative Adversarial Networks

Introduced by Ian Goodfellow in 2014, GANs consist of two competing neural networks a generator and a discriminator. The generator attempts to produce synthetic content indistinguishable from real content; the discriminator attempts to detect the fake. Through iterative adversarial training, the generator progressively improves until its output fools the discriminator with high frequency. Face-swap deepfakes the archetypal form primarily use GAN-based architectures such as Deep Face Lab and Face Swap.

2.2.2 Diffusion Models

More recent systems such as Stable Diffusion, DALL-E, and Midjourney employ diffusion models, which generate images by learning to reverse a process of progressive noise addition. These models, trained on billions of internet-scraped images many of which contain identifiable persons can produce photorealistic images of real individuals in fabricated contexts with alarming ease.

2.2.3 Voice Cloning and Neural Codec Language Models

Voice deepfakes have advanced to the point where a model trained on as little as three seconds of a target's voice can generate unlimited synthetic audio in that voice. Tools such as ElevenLabs, Resemble AI, and open-source platforms like Tortoise-TTS can clone vocal characteristics tone, cadence, accent, emotional inflection with fidelity that defeats telephone-based voice authentication systems widely used by Indian banks and government services.

2.2.4 Facial Mapping and Reenactment

A parallel class of deepfakes does not swap faces but rather reenacts them controlling the facial expressions and head movements of the target through driving video. First Order Motion Model and similar architectures can animate a single photograph of a person using the head movements of an entirely different actor, effectively making the person appear to speak words they never uttered.

2.3 Taxonomy of Deepfakes

Understanding the diverse forms deepfakes take is essential for legal categorisation and

targeted regulation:

- Video Deepfakes: Face-swapped or reenacted videos, typically targeting sexual exploitation, political misinformation, reputational destruction, or financial fraud.
- Audio Deepfakes: Cloned voice recordings used to impersonate individuals in telephone fraud, ransom calls, boardroom impersonation, or to fabricate incriminating admissions.
- Image Deepfakes: AI-generated photographs placing real persons in fabricated contexts compromising, criminal, or politically charged.
- Text Impersonation: Large language models prompted to mimic a specific individual's writing style, used to generate fake statements, social media posts, or legal documents attributed to real persons.
- Document Forgery: AI-assisted generation of fake identity documents, certificates, and official records that incorporate authentic-looking personal data.

2.4 The Role of Personal Data in Deepfake Creation

Every deepfake begins with data specifically, personal data belonging to the target. The creation of a convincing face-swap deepfake requires a dataset of photographs or videos of the target from multiple angles and lighting conditions. Voice cloning requires audio recordings. Together, these constitute sensitive personal data under any reasonable reading of privacy law.

The sources from which this data is drawn are both varied and largely unregulated. Social media platforms are the primary reservoir: a person with an active Instagram profile of several years may have, without knowing it, provided thousands of high-resolution photographs from which a deepfake could be trained. Public speeches, interviews, podcasts, and news appearances provide voice training data. Biometric identifiers embedded in these media facial geometry, voice prints, gait patterns are automatically extracted by AI systems without any active act of disclosure by the data subject.

This is the foundational legal problem: the personal data being misused for deepfake

creation was almost never provided by the subject for that purpose. It was shared for social communication, political participation, professional visibility, or journalistic coverage. Its weaponisation into a deepfake training dataset represents a fundamental betrayal of the purpose limitation principle that underpins modern data protection philosophy.

3. PERSONAL DATA MISUSE IN THE GENERATIVE AI ECOSYSTEM

3.1 The Data Hunger of AI Systems

Generative AI systems are voracious consumers of data. The Large Language Models that power ChatGPT, Google Gemini, and Meta's Llama series were trained on hundreds of billions of words scraped from the open internet, books, academic papers, and social media. The diffusion models that generate photorealistic images were trained on datasets such as LAION-5B a collection of approximately 5.85 billion image-text pairs harvested from the web the vast majority sourced without the knowledge or consent of the individuals pictured or the copyright holders of the content.

This data collection paradigm creates a profound consent vacuum. When a person posts a family photograph on social media, they consent typically under a platform's terms of service to the platform displaying that image. They do not consent to that image being scraped, stored in a training corpus, used to train a commercial AI model, and eventually enabling the generation of synthetic images of their face. The transactional chain from original upload to AI training is several steps removed from anything the user contemplated.

3.2 Web Scraping and Unauthorised Dataset Creation

Systematic web scraping for AI training purposes has become one of the defining privacy controversies of the current decade. The Clearview AI case exemplifies the scale and severity of this problem. Clearview AI, a facial recognition company, scraped over thirty billion photographs from social media platforms, news websites, and public databases without any consent from the individuals pictured to build a facial recognition system marketed to law enforcement agencies. Regulatory authorities in Canada, Australia, France, Italy, and the United Kingdom have imposed substantial fines on Clearview AI for violating their respective data protection laws. India's DPDPA, notably, has no comparable enforcement mechanism in its current form.

Closer to home, Indian citizens' photographs from platforms such as Sharechat, Josh, and Moj where linguistic and regional minorities are disproportionately represented have reportedly been included in international training datasets without disclosure. The data flows across jurisdictions in ways that are effectively invisible to individuals and presently ungoverned by Indian law.

3.3 The Consent Problem in AI Training

The DPDPA, 2023 requires consent for the processing of personal data, and that consent must be free, specific, informed, unconditional, and unambiguous. None of these criteria are typically satisfied in the context of AI training data collection. The 'consent' embedded in social media terms of service is neither specific (it covers a broad range of undefined future uses) nor informed (users cannot reasonably anticipate that their photographs will train facial recognition or generative AI systems). India's regulatory framework does not yet address whether scraping publicly available personal data for AI training constitutes unlawful processing under the Act.

3.4 Cross-Border Data Flows and Foreign AI Platforms

The jurisdictional dimension of this problem is particularly acute for India. The dominant generative AI platforms OpenAI, Google DeepMind, Meta AI, Stability AI are headquartered outside India, subject to different regulatory regimes, and process data from Indian users on servers located abroad. When an Indian citizen's voice recording is used to fine-tune an American speech synthesis model, the data protection violation occurs across multiple jurisdictions and the victim has no effective legal remedy under current Indian law.

3.5 Dark Web Markets for Personal Data

A particularly alarming vector for deepfake enablement is the trade in personal data on dark web marketplaces. Stolen databases containing photographs, identity documents, phone numbers, and financial records from Indian data breaches are routinely offered for sale on platforms accessible through the Tor network. Deepfake-as-a-service operations purchase these datasets to create targeted impersonation attacks against specific individuals executives, politicians, celebrities, and ordinary citizens. The availability of this data on criminal markets means that even individuals who exercise reasonable caution about their online presence are not fully insulated from deepfake targeting.

4. THE DIGITAL PERSONAL DATA PROTECTION ACT, 2023: FRAMEWORK AND PROVISIONS

4.1 Legislative History and Context

The Digital Personal Data Protection Act, 2023 represents the culmination of over seven years of legislative deliberation in India about data privacy. The process began with the Supreme Court's landmark judgment in Justice K.S. Puttaswamy v. Union of India (2017), which unanimously declared the right to privacy a fundamental right under the Constitution of India. This judgment provided the normative foundation for a comprehensive data protection statute.

Three successive draft Bills the Personal Data Protection Bill, 2018; the Personal Data Protection Bill, 2019; and the Data Protection Bill, 2021 preceded the current Act. Each was withdrawn or substantially revised amid controversy over surveillance carve-outs, government data processing exemptions, and the scope of individual rights. The 2023 Act was finally enacted as a leaner, more industry-friendly legislation, but critics argue that the dilutions have come at the cost of meaningful individual protection.

4.2 Key Provisions of the DPDPA

4.2.1 Definition of Personal Data

Section 2(t) defines 'personal data' as any data about an individual who is identifiable by or in relation to such data. This is a broad, technology-neutral definition that, on its face, should encompass photographs, voice recordings, and biometric identifiers all of which are used in deepfake creation. However, the Act does not specifically enumerate biometric data as a category deserving heightened protection, a critical omission distinguishing it from the GDPR and India's own earlier draft Bills.

4.2.2 Consent Requirements

Part II of the Act establishes consent as the primary basis for personal data processing. Section 6 requires consent to be free, specific, informed, unconditional, and unambiguous, with a clear affirmative action. This standard, if rigorously applied, would prohibit scraping personal data from social media for AI training without explicit consent. However, the Act's

enforcement mechanisms are significantly weaker than this aspirational standard, and its exemptions particularly in Section 7 covering 'legitimate uses' provide potential workarounds for data fiduciaries.

4.2.3 Obligations of Data Fiduciaries

Entities that determine the purpose and means of processing personal data termed 'Data Fiduciaries' under the Act bear obligations of accuracy, security, and data minimisation. They are required to implement appropriate technical and organisational measures and to erase personal data when the purpose of processing is completed. These obligations, while sound in principle, are framed at a level of generality that offers little specific guidance on AI training practices or deepfake-related harms.

4.2.4 Rights of Data Principals

The Act grants individuals termed 'Data Principals' rights including the right to information about processing, the right to correction and erasure, and the right to grievance redressal. Notably absent are rights explicitly addressing AI-specific harms: there is no right to object to inclusion in AI training datasets, no right to demand deepfake takedown, and no right to algorithmic explanation or transparency regarding decisions made using AI systems trained on one's personal data.

4.3 The Central Silence: What the DPDPA Does Not Address

The most significant characteristic of the DPDPA in the context of deepfake regulation is its comprehensive silence. The Act does not:

- Define, regulate, or prohibit the creation of deepfakes or synthetic media incorporating personal data.
- Establish any category of 'biometric sensitive data' requiring heightened consent or protection.
- Create obligations specific to AI systems that process personal data for training purposes.
- Address synthetic identity theft or AI-generated impersonation as distinct legal harms.

- Provide for injunctive relief, platform takedown obligations, or victim compensation in the context of AI-generated content misuse.

This silence is not incidental. It reflects the political economy of AI regulation an unwillingness to impose obligations on the rapidly expanding AI industry but its consequences for individual rights are severe.

5. CRITICAL LEGAL GAPS IN THE DPDPA, 2023

This chapter constitutes the analytical core of this article. Six discrete and compounding legal gaps are identified, each of which creates specific vulnerabilities for individuals at risk of deepfake harm.

Gap 1: Absence of Definition or Regulation of Biometric Manipulation

The DPDPA's definition of personal data is broad but unarticulated with respect to biometric information. Earlier drafts of the legislation, drawing from frameworks like India's own Aadhaar Act, 2016, and the EU GDPR, explicitly listed biometric data as a 'sensitive personal data' category subject to stricter processing conditions. This categorisation was abandoned in the 2023 Act.

The consequence is stark: when a person's facial geometry is extracted from photographs and fed into a GAN to generate deepfake imagery, there is no provision in the Act that specifically characterises this as processing of sensitive biometric data requiring explicit consent. The facial geometry is arguably 'personal data' under the broad definition, but the absence of heightened protection means Data Fiduciaries can potentially rely on the general consent framework or legitimacy exceptions to justify such processing.

Furthermore, the concept of 'biometric cloning' the extraction and replication of unique biometric identifiers for the purpose of impersonation has no legal analogue anywhere in the Act. This is an extraordinary gap given India's increasing reliance on biometric authentication for everything from banking to welfare distribution under the Aadhaar ecosystem.

Gap 2: No Regulatory Oversight of AI Training Datasets

The DPDPA does not address the data lifecycle specific to AI model training, which is

fundamentally different from conventional data processing. In ordinary data processing, data is collected for a defined purpose (processing a transaction, delivering a service), used for that purpose, and then retained or erased. AI training involves a different paradigm: data is used to modify the internal parameters of a model, and once training is complete, the data's influence persists within the model's weights a form of data retention that is neither transparent nor easily reversible.

The Act's 'purpose limitation' principle requires that data collected for one purpose cannot be used for a different purpose without fresh consent. Applying this principle strictly, photographs shared on a social media platform cannot be scraped and used to train an AI image generation model without separate consent. However, the Act provides no enforcement mechanism or guidance for this scenario, and the Data Protection Board of India the regulatory body established under the Act has not issued any advisory or rule addressing AI training data practices.

This gap is particularly consequential because the most capable AI systems require training data at a scale that effectively precludes individually obtained consent for each data point. The regulatory challenge is genuine but the absence of any framework at all leaves individuals entirely unprotected.

Gap 3: No Deepfake-Specific Criminal Liability

The DPDPA is primarily a civil regulatory instrument. Its penal provisions concentrated in Chapter VII impose monetary penalties on Data Fiduciaries and Data Processors for specified violations, to be adjudicated by the Data Protection Board. The maximum penalty under Section 33 is Rs. 250 crore for certain categories of violation. However, these penalties are institutional they attach to organisations, not individuals, and they address data processing violations, not the independent harm of deepfake creation and dissemination.

There is no provision in the Act that criminalises the non-consensual use of personal data to generate deepfake content, the distribution of such content, or the use of deepfakes for fraud, impersonation, or sexual exploitation. A person who scrapes photographs of a private individual from social media and uses them to generate and distribute non-consensual intimate deepfake imagery commits, in the terms of the DPDPA, a 'data processing violation' attracting

a financial penalty. The profound dignitary and psychological harm inflicted on the victim finds no articulation in the Act.

Gap 4: Cross-Border Enforcement Deficits

The dominant players in the generative AI ecosystem OpenAI, Google, Meta, Stability AI, Midjourney are domiciled outside India and subject to the laws of the United States, the European Union, or other jurisdictions. When their AI systems process Indian citizens' personal data, or when deepfakes targeting Indian citizens are generated using their tools, the question of jurisdictional reach arises acutely.

Section 3 of the DPDPA provides that it applies to processing of personal data collected within India, as well as processing by entities outside India offering goods or services to individuals within India. This extra-territorial provision follows the GDPR's approach and is well-intentioned. However, the Act provides no mechanism for enforcing Data Protection Board orders against foreign entities that do not voluntarily comply no mutual legal assistance treaty framework for data protection enforcement, no blacklisting mechanism, and no provision for cross-border regulatory cooperation analogous to GDPR's cooperation mechanism among EU data protection authorities.

The practical consequence is that a victim of a deepfake generated by a foreign AI platform using illegally scraped Indian data may obtain a favourable order from the Data Protection Board, but that order will be effectively unenforceable against the foreign entity. This jurisdictional impotence fundamentally undermines the Act's protective purpose.

Gap 5: Inadequate Platform Accountability

Social media platforms and AI-hosting services occupy a unique structural position in the deepfake ecosystem: they are neither the creators of deepfakes nor their ultimate victims, but they are the essential distribution infrastructure through which deepfake harm is scaled from individual incidents to societal crises. The Rashmika Mandanna deepfake video reached millions of Indian viewers within hours not because of any sophisticated distribution effort by the creator, but because the platform's algorithmic amplification engine promoted engaging content regardless of its authenticity or harm.

The DPDPA does not address platform accountability for AI-generated content

specifically. The existing regime of intermediary liability under the Information Technology Act and the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 provides some framework platforms must take down notified illegal content within specified timeframes but these rules were not designed with AI-generated content in mind and do not require platforms to proactively detect or label deepfakes.

The absence of obligations for AI watermarking, provenance disclosure, or automated deepfake detection in either the DPDPA or the IT Rules means that platforms have no legal obligation to invest in the technical infrastructure that would enable victims to identify, trace, and seek redressal for deepfake harm.

Gap 6: Absence of Victim Compensation Mechanism

Perhaps the most significant practical gap in the DPDPA from a victim's perspective is the near-total absence of a compensation mechanism. Section 33 of the Act creates penalties payable to the Consolidated Fund of India not to victims. There is no provision for a Data Protection Board order to award monetary compensation to the individual whose data was misused, analogous to civil damages awards in personal injury litigation.

A person who's intimate deepfake imagery was created and distributed without consent may sustain profound psychological trauma, reputational devastation, professional loss, and relationship harm. Under the DPDPA, her only remedy within the data protection framework is a complaint to the Data Protection Board, which may result in a financial penalty being paid to the government. She receives nothing. She must pursue separate civil litigation for tortious harm an expensive, technically complex, and practically difficult avenue that most victims will be unable to access.

6. DEEPFAKES AND EXISTING INDIAN LAWS: A CRITICAL ASSESSMENT

Given the DPDPA's structural inadequacy, an important question arises: can other limbs of Indian law provide effective recourse to deepfake victims? The answer, on examination, is partial and unsatisfactory.

6.1 Bharatiya Nyaya Sanhita, 2023

The BNS replaced the Indian Penal Code, 1860, and came into force in July 2024.

Several of its provisions have potential, though imperfect, application to deepfake harms.

Section 318 (Cheating) may be invoked where deepfakes are used to deceive victims in financial fraud. Section 356 (Criminal Intimidation, which includes defamation) may apply to deepfakes designed to harm reputation. Section 79 (Word, gesture, or act intended to insult the modesty of a woman) and Section 74 (Assault or criminal force to woman with intent to outrage her modesty) have uncertain application to digital deepfake content. Section 85 read with Section 86 addressing obscene acts and songs may partially apply to sexual deepfakes.

The most significant provision is the newly introduced offence of publication of obscene material in electronic form under Section 294 and the provisions relating to identity theft under Section 319. However, none of these provisions are specifically calibrated to deepfakes. They do not account for the AI-mediated nature of the harm, they require proof of specific intent that may be difficult to establish, and they do not address the platform distribution dimension.

6.2 The Information Technology Act, 2000

The IT Act contains provisions of direct relevance: Section 66C criminalises identity theft, Section 66D criminalises cheating by impersonation using computer resources, Section 67 criminalises publication of obscene material in electronic form, and Section 67A specifically addresses publication of material containing sexually explicit acts.

These provisions were invoked in the aftermath of the Rashmika Mandanna deepfake incident, and the Delhi Police registered a case under Section 67A. However, the investigation revealed the structural limitations of these provisions they were enacted in 2000 and 2008, before deepfake technology existed. They do not address the supply chain problem: the AI tool developers, the dataset curators, the platform amplifiers. They focus on the terminal act of publication, not the upstream processes that enable it.

Section 79 of the IT Act, which grants intermediary platforms conditional immunity from liability for third-party content, creates a further obstacle. Platforms can claim safe harbour protection even when they host and algorithmically amplify deepfake content, as long as they act expeditiously upon receiving notification of its existence. This creates a perverse incentive structure: platforms have no reason to proactively detect deepfakes, and their legal

obligation only arises after harm has already been distributed at scale.

6.3 Bharatiya Sakshya Adhiniyam, 2023: Evidentiary Challenges

The admissibility of electronic evidence in deepfake cases raises profound challenges under the Bharatiya Sakshya Adhiniyam, 2023. Section 57 of the BSA provides that electronic records produced from electronic devices are admissible if accompanied by a certificate from a responsible official verifying the device and the integrity of the data.

Deepfake evidence presents a dual challenge: on one hand, prosecutors must produce video or audio evidence of the deepfake's creation or distribution, which must be authenticated under the BSA. On the other hand, the deepfake itself may be presented as evidence by a defendant claiming fabrication of evidence against them. The legal system currently lacks clear standards for determining the authenticity of AI-generated content, and there is no formally recognised protocol for deepfake forensic authentication.

The expert witness provision under Section 39 of the BSA may be used to admit testimony from AI forensic specialists, but Indian courts have not yet developed consistent jurisprudence on the admissibility standards for such evidence. A comprehensive evidentiary framework for AI-generated content is urgently needed.

6.4 Electoral Deepfakes and the Election Commission

The use of AI-generated content in Indian elections represents a particularly dangerous application of deepfake technology. The 2024 General Elections saw documented instances of deepfake audio clips attributed to political leaders circulating on WhatsApp, AI-generated videos of candidates making statements they never made, and synthetic social media accounts amplifying disinformation at scale.

The Election Commission of India issued advisories cautioning political parties against the use of deepfakes but lacks statutory authority to effectively regulate or sanction such conduct. The Model Code of Conduct does not contain deepfake-specific provisions, and the Representation of the People Act, 1951's provisions on corrupt practices and electoral offences do not contemplate AI-generated political disinformation. This is a democratic emergency requiring urgent legislative attention.

7. COMPARATIVE INTERNATIONAL ANALYSIS

7.1 The European Union: A Comprehensive Risk-Based Framework

The European Union's Artificial Intelligence Act, provisionally agreed in December 2023 and formally adopted in 2024, represents the world's most comprehensive regulatory framework for AI governance. The EU AI Act employs a risk-tiered architecture that classifies AI systems by the level of risk they pose and imposes proportionate obligations accordingly.

Of particular relevance to deepfakes, the EU AI Act requires that AI systems generating synthetic content including AI-generated images, audio, and video be equipped with technical measures ensuring that outputs are labelled as artificially generated or manipulated. This mandatory 'AI watermarking' obligation applies to all providers placing such systems on the EU market, regardless of where the system is developed. The Act further requires that AI systems used for real-time remote biometric identification in publicly accessible spaces be subject to strict prior authorisation, directly addressing the biometric data processing dimension.

The GDPR, which complements the AI Act, classifies biometric data as a 'special category' subject to explicit consent requirements and prohibits processing for the purpose of uniquely identifying a person unless one of a narrow list of conditions is met. Taken together, the EU framework treats deepfake-enabling data practices as presumptively unlawful, a position diametrically opposed to the permissive silence of India's DPDPA.

7.2 The People's Republic of China: Proactive Deepfake Regulation

China has adopted what is arguably the most direct legislative response to deepfakes globally. The Provisions on Deep Synthesis Internet Information Services, issued by the Cyberspace Administration of China and effective from January 2023, impose mandatory obligations on providers of 'deep synthesis' technology. These include: mandatory user consent before creating deepfakes using a person's likeness; prohibition on using deep synthesis technology to produce, reproduce, or disseminate false news information; mandatory labelling of all deep synthesis content with conspicuous notice of its synthetic nature; and preservation of logs for sixty days enabling traceability.

Service providers are required to prevent users from using the technology to infringe

personality rights, and must report and remove illegal content. The framework is backed by administrative penalties and, in serious cases, criminal liability. China's approach is not without its own concerns regarding state surveillance and the potential weaponisation of deepfake regulation against political dissent, but as a technical regulatory model it represents a sophistication that India's framework entirely lacks.

7.3 The United States: A Patchwork of State Legislation

The United States has not enacted federal deepfake legislation, but a growing number of states have passed targeted statutes. California's AB 602 and AB 730 (2019) address digital manipulation of political candidates and non-consensual intimate deepfakes, respectively. Virginia has criminalised the distribution of deepfake pornography. Texas and Georgia have enacted deepfake-specific electoral offence provisions.

The FTC has taken enforcement action against companies engaged in deceptive AI practices, and the proposed 'No FAKES Act' at the federal level would create a private right of action against persons who produce AI-generated content using an individual's voice or likeness without consent a model with significant potential applicability for India.

7.4 UNESCO's AI Ethics Framework

The UNESCO Recommendation on the Ethics of Artificial Intelligence (2021), adopted by 193 member states including India, provides a normative framework emphasising human dignity, right to privacy, non-discrimination, and accountability in AI systems. While non-binding, the Recommendation provides important interpretive guidance for domestic legislation and has been invoked in regulatory reform discussions globally. India's government, as a signatory, bears a soft law obligation to align domestic AI regulation with these principles.

8. LANDMARK CASES AND REAL-WORLD INCIDENTS

8.1 India: The Rashmika Mandanna Case (2023)

In November 2023, a deepfake video depicting popular Indian actress Rashmika Mandanna went viral across Indian social media platforms. The video superimposed Mandanna's face onto the body of a British Indian influencer in a morphed video of explicitly suggestive nature. The video accumulated millions of views before it was flagged, and it spread

on Instagram, X (formerly Twitter), and WhatsApp before platforms began taking it down.

The incident galvanised public discourse in India. The Ministry of Electronics and Information Technology issued an emergency advisory to social media platforms invoking Rule 3(1)(b) of the IT (Intermediary Guidelines) Rules requiring removal of content violating specified categories. The Delhi Police registered a case under Sections 66C and 67A of the IT Act. However, no individual was charged, no platform was penalised, and the victim had no legal mechanism to seek personal compensation. The case exposed, with clinical precision, the legal vacuum this article seeks to address.

8.2 Political Deepfakes: The 2024 General Elections

During the 2024 Indian General Elections, fact-checking organisations documented numerous instances of AI-generated political content. Deepfake audio clips purportedly featuring political leaders making inflammatory or damaging statements circulated widely on encrypted messaging platforms, particularly WhatsApp. The rapid viral spread of such content in the hours before the twenty-four-hour pre-election silence period created situations where corrections could not reach the same audience that had seen the original fabrication.

8.3 CEO Fraud via Voice Deepfakes

Multiple Indian corporations have reported incidents of voice deepfake fraud in which callers impersonating company executives their voices cloned using AI instructed finance personnel to authorise urgent wire transfers to fraudulent accounts. In one documented case, a mid-sized manufacturing company lost over Rs. 2 crore in a single transaction. Voice biometrics deployed by the company's bank were defeated by the AI-generated voice clone, illustrating the inadequacy of current authentication infrastructure.

8.4 Non-Consensual Intimate Deepfakes

India's National Commission for Women and various state women's commissions have received complaints involving deepfake intimate imagery. These cases disproportionately target women, often in domestic contexts following relationship breakdown, or as tools of harassment targeting women in public life activists, journalists, politicians, and educators. The psychological harm documented in these cases is severe and includes clinical depression, post-traumatic stress disorder, withdrawal from professional life, and in some reported cases,

suicidal ideation.

8.5 International Cases with Instructive Relevance

The 2023 case of AI-generated intimate imagery of Taylor Swift circulating on the X platform, reaching approximately twenty-seven million views before removal, forced a major international platform to publicly confront its deepfake content policies and led directly to proposed U.S. federal legislation. In South Korea, the discovery of encrypted Telegram channels dedicated to creating and sharing deepfake intimate imagery of schoolgirls and female university students triggered nationwide protests and emergency legislation. These cases illustrate the speed at which AI-enabled harm can scale, and the inadequacy of ad hoc responses.

9. CHALLENGES IN INVESTIGATION, DIGITAL FORENSICS, AND EVIDENCE

9.1 The Detection Problem

The forensic detection of deepfakes is a technically demanding and constantly evolving challenge. As generative AI models improve, their output becomes progressively harder to distinguish from authentic media using both human inspection and automated detection tools. This creates a technological arms race: forensic detection systems trained on the artefacts of current-generation deepfakes become obsolete as next-generation models learn to avoid those artefacts.

Current deepfake detection approaches include CNN-based classifiers trained to identify compression artefacts, temporal inconsistencies in video (unnatural blinking patterns, gaze anomalies), frequency domain analysis revealing GAN fingerprints, and physiological signal analysis detecting implausible variations in facial blood flow. State-of-the-art detection systems achieve high accuracy rates in controlled laboratory conditions, but their performance degrades significantly when deepfakes are further processed through social media compression, noise addition, or format conversion precisely the processing that occurs during online distribution.

9.2 Source Attribution and Blockchain Tracing

A critical investigative challenge is the attribution of deepfakes to their creators. The

tools used to generate deepfakes are globally distributed, often open-source, and can be run on consumer hardware without any account registration or usage logging. Tracing a deepfake to its creator requires identifying metadata creation timestamps, hardware fingerprints, software signatures that may be deliberately stripped, and following network traces that typically traverse multiple jurisdictions and anonymising layers.

Blockchain-based content provenance solutions such as the Coalition for Content Provenance and Authenticity (C2PA) standard offer a promising technical approach. Under C2PA, content creation tools embed cryptographically signed provenance metadata at the point of creation, creating an auditable chain of custody. Some camera manufacturers and AI platforms have begun adopting this standard voluntarily. However, adoption remains uneven and these standards do not apply retroactively to existing deepfakes or to tools whose developers choose not to participate.

9.3 Metadata Analysis and Digital Chain of Custody

EXIF metadata embedded in image files, and equivalent metadata in video files, can provide valuable forensic information including creation timestamp, device model, software used, and GPS coordinates. However, deepfake creation tools frequently strip or falsify this metadata, and images shared through social media platforms are typically stripped of all EXIF data as part of the upload processing. The forensic investigator is therefore often working with a decontextualised file whose authentic provenance cannot be established through metadata analysis alone.

9.4 Admissibility Under the Bharatiya Sakshya Adhiniyam, 2023

The admissibility of deepfake forensic evidence in Indian courts presents novel challenges under the BSA. Section 57 requires a certificate from a person in charge of the electronic device establishing its authenticity. For deepfake evidence, the court must grapple with a more complex question: not just whether the file is an accurate record of what was stored on a device, but whether the content itself is authentic or fabricated.

There is currently no standardised procedure in Indian courts for the appointment of deepfake forensic experts, no accreditation standard for AI forensic laboratories, and no judicial guideline on the weight to be accorded to AI-generated forensic reports. These gaps

create significant risks of miscarriage of justice both through the wrongful conviction of persons based on fabricated deepfake evidence presented as authentic, and through the wrongful acquittal of actual deepfake creators whose content cannot be forensically authenticated under existing standards.

9.5 Jurisdictional Complexity in Investigation

A deepfake created using an open-source model downloaded from a server in the United States, using training data scraped from Indian social media, processed on cloud computing infrastructure in Singapore, and distributed through an Irish-registered social media platform to an Indian victim presents an investigation spanning at least four jurisdictions. India currently has Mutual Legal Assistance Treaties (MLATs) with certain countries, but the process of obtaining evidence through MLATs is typically measured in months or years entirely inadequate for the speed at which digital evidence degrades and perpetrators disappear.

10. ETHICAL DIMENSIONS OF DEEPFAKE TECHNOLOGY

10.1 Dignity and Bodily Autonomy in Digital Space

The creation of a deepfake of a real person without their consent constitutes, at the most fundamental level, a violation of human dignity. It asserts a claim over that person's bodily identity their face, their voice, their physical presence that belongs exclusively to them. The fact that the violation occurs in digital space, using computational rather than physical means, does not diminish its severity. The individual's loss of control over their own identity representation is total.

This has particular resonance for women, who constitute the overwhelming majority of non-consensual intimate deepfake victims. The creation of sexualised deepfakes reproduces, in digital form, the logic of sexual objectification and the denial of bodily autonomy that feminist jurisprudence has long identified as central to gendered structural inequality. India's legal system, through its evolving jurisprudence on sexual harassment and dignity, has the normative resources to characterise such harm but the statutory instruments to enforce this characterisation against deepfake creators do not yet exist.

10.2 Epistemic Harm and the Collapse of Trust

Beyond individual harm, deepfakes inflict what philosophers of technology term

'epistemic harm' damage to the shared social infrastructure of truth on which democracy, journalism, and legal proceedings depend. When a sufficiently convincing deepfake of a political figure can be generated in minutes, the credibility of video evidence historically considered among the most powerful forms of proof is fundamentally undermined. The legal concept of 'beyond reasonable doubt' in criminal proceedings rests on the assumption that evidence presented to a court is capable of being authenticated. A world saturated with convincing deepfakes challenges this assumption.

10.3 Democratic Manipulation

The deployment of deepfakes in electoral contexts fabricated statements by candidates, synthetic rallies, AI-generated voter suppression messaging represents a direct attack on democratic self-determination. The harm here is not merely to individual candidates but to the collective capacity of citizens to make informed electoral choices. India, as the world's largest democracy with a voter base of nearly one billion, faces particular exposure to AI-enabled electoral manipulation. The absence of targeted legal provisions addressing deepfake electioneering is a democratic governance failure of the first order.

11. RECOMMENDATIONS: A COMPREHENSIVE REFORM MODEL

The legal gaps identified in this article are not amenable to incremental patch solutions. They require a systematic, multi-layered reform model that addresses the harms of deepfake technology at every stage of the production and distribution chain. The following reforms are proposed.

11.1 Enact a Standalone Deepfake Regulation Act

India requires a dedicated Deepfake and Synthetic Media Regulation Act that specifically criminalises: the non-consensual creation of deepfakes using another person's likeness; the creation of sexual deepfakes under any circumstances without the explicit consent of the depicted individual; the use of deepfakes for electoral influence operations; and the distribution or hosting of deepfake content with knowledge of its synthetic nature and harmful purpose.

The Act should adopt a clear definition of 'synthetic media' calibrated to technological reality not restricted to video but encompassing audio, image, and text deepfakes. It should

establish tiered offences based on harm severity, with penalties ranging from substantial fines to custodial sentences for aggravated offences involving sexual content, electoral manipulation, or financial fraud.

11.2 Mandatory AI Content Watermarking

All AI systems capable of generating synthetic media of real persons should be required, by subordinate legislation under the IT Act or a new AI regulation, to embed cryptographically secure provenance metadata C2PA-compliant digital watermarks in all generated content at the point of creation. This watermarking should survive common processing operations including social media compression and format conversion. The failure to implement mandatory watermarking should constitute a strict liability offence for the AI system provider.

11.3 AI Licensing Mechanism

Generative AI systems capable of creating synthetic media of identifiable persons should require registration and licensing from a designated regulatory authority potentially the Ministry of Electronics and Information Technology or a newly constituted AI Regulatory Authority of India. Licensing conditions should include: mandatory deepfake detection system integration; content watermarking compliance; user identity verification for generation of photorealistic human content; and retention of generation logs for a specified period to enable forensic tracing.

11.4 Amend the DPDPA to Address AI-Specific Harms

The following specific amendments to the DPDPA are proposed: insertion of a definition of 'biometric data' as a special category of sensitive personal data requiring explicit, purpose-specific consent; insertion of a new chapter addressing AI training data, requiring data fiduciaries that train AI models on personal data to obtain separate explicit consent for that specific purpose; creation of a private right of action enabling data principals to claim compensation directly from the Data Protection Board rather than merely seeking government-retained penalties; and provision of a data principal's right to demand exclusion from AI training datasets and to request deletion of their personal data from trained AI models where technically feasible.

11.5 Strengthen Platform Accountability

Amendments to the IT (Intermediary Guidelines) Rules should impose proactive obligations on Significant Social Media Intermediaries regarding deepfake content: mandatory integration of deepfake detection tools meeting specified minimum accuracy standards; automatic labelling of content detected as AI-generated; expedited takedown timelines for deepfake content within six hours of reporting, reduced from the current twenty-four-to-thirty-six-hour regime for other illegal content; and removal of the safe harbour protection under Section 79 of the IT Act where a platform has been notified of deepfake content and fails to act within the expedited timeframe.

11.6 Establish a Victim Compensation Framework

A dedicated Deepfake Victim Compensation Fund, seeded by penalties collected from violators and supplemented by mandatory contributions from AI platform operators, should be established to provide rapid relief to victims without requiring them to navigate expensive civil litigation. The fund should be administered by the Data Protection Board in coordination with the proposed Deepfake Regulation Authority, and should provide compensation for documented harms including psychological treatment costs, lost income, and the costs of content removal.

11.7 Strengthen Cross-Border Enforcement

India should negotiate bilateral data protection enforcement agreements with major AI-producing countries the United States, the United Kingdom, the European Union, and Canada modelled on existing MLAT frameworks but specifically calibrated to data protection and AI regulation. These agreements should provide for rapid information sharing between regulatory authorities, mutual recognition of enforcement orders, and joint investigation mechanisms for cross-border deepfake incidents.

11.8 Invest in Forensic Capacity Building

The Central Forensic Science Laboratory and state forensic science laboratories should be mandated and resourced to develop specialised AI forensics units capable of deepfake authentication, source attribution analysis, and expert witness testimony in criminal proceedings. Judicial training programmes should include deepfake forensics modules, and the

Supreme Court's E-Committee should develop model guidelines for the admissibility of AI forensic evidence under the Bharatiya Sakshya Adhiniyam, 2023.

12. CONCLUSION

The relationship between artificial intelligence and personal identity sits at the frontier of the most important legal debates of our time. In this rapidly shifting landscape, India has enacted a data protection statute that, while a genuine legislative achievement in principle, fails in substance to address the specific harms that make AI-mediated personal data misuse so distinctive and so destructive.

The Digital Personal Data Protection Act, 2023 was the product of years of deliberation and represented hard-won political compromise. It should be recognised for what it is: a necessary foundation. But foundations are not walls, and walls are not roofs. The harms that Indian citizens face from deepfake technology require the full architectural edifice of legal protection specific definitions, targeted liability, platform obligations, victim remedies, and cross-border enforcement. The DPDPA provides none of these in the deepfake context.

The six legal gaps identified in this article concerning biometric manipulation, AI training datasets, deepfake-specific liability, cross-border enforcement, platform accountability, and victim compensation are not peripheral technicalities. They are structural failures that leave millions of Indians, and disproportionately Indian women, without meaningful legal protection against a category of harm that existing evidence shows to be spreading rapidly.

The international comparisons undertaken in this article are instructive: the European Union's AI Act, China's Deep Synthesis Provisions, and even the patchwork of American state legislation all demonstrate that targeted deepfake regulation is achievable and that its absence is a policy choice, not a technical necessity. India's legislative capacity is not in question what is in question is the political will to prioritise the dignity and data rights of ordinary citizens over the short-term convenience of an unregulated AI industry.

Artificial intelligence will continue its exponential advancement regardless of what any legislature does. The question is whether the law will run alongside it, providing guardrails and remedies, or whether it will trail behind, arriving at each crisis scene after the harm is already

done. This article has argued, with evidence and analysis, that the time for trailing behind has passed. The reform recommendations proposed here are neither radical nor impractical they draw on established international models, are technically feasible, and are constitutionally grounded. What they require is urgency, and that urgency is not a matter of academic preference. It is a matter of democratic necessity.

REFERENCES

Primary Sources

Legislation

- The Digital Personal Data Protection Act, 2023 (India)
- The Information Technology Act, 2000 (India)
- The Bharatiya Nyaya Sanhita, 2023 (India)
- The Bharatiya Sakshya Adhiniyam, 2023 (India)
- The Bharatiya Nagarik Suraksha Sanhita, 2023 (India)
- The Representation of the People Act, 1951 (India)
- The Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016 (India)
- Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU Artificial Intelligence Act)
- Regulation (EU) 2016/679 (General Data Protection Regulation)
- Provisions on Deep Synthesis Internet Information Services (People's Republic of China, 2022)
- California AB 602 and AB 730 (United States)

Judicial Decisions

- Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1 (Supreme Court of India)
- Shreya Singhal v. Union of India, (2015) 5 SCC 1 (Supreme Court of India)
- People v. Harris, No. 23CR (California, 2023)

Secondary Sources

Books and Monographs

- Srinivasan S., 'Data Sovereignty and the Right to Privacy in India: Constitutional and

Statutory Dimensions' (LexisNexis India, 2022)

- Citron D.K., 'The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age' (W.W. Norton, 2022)

- Chesney R. and Citron D.K., 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (107 California Law Review 1753, 2019)

- Lynskey O., 'The Foundations of EU Data Protection Law' (Oxford University Press, 2015)

Articles and Reports

- Mittal S. and Singh J., 'Deepfakes and Indian Law: An Analysis of Legislative Lacunae' (2023) 35 National Law School of India Review 112

- Westerlund M., 'The Emergence of Deepfake Technology: A Review' (2019) 15 Technology Innovation Management Review 40

- Diakopoulos N. and Johnson D., 'Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections' (2021) 23 New Media & Society 2072

- Ministry of Electronics and Information Technology, 'Advisory on Deepfakes and AI-Generated Content' (Government of India, November 2023)

- UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (UNESCO, 2021)

- European Parliament, 'Artificial Intelligence Act: Agreed Text' (European Parliament, March 2024)

- Basu M., 'Regulating Artificial Intelligence in India: Towards a Rights-Based Framework' (2024) 6 Indian Journal of Law and Technology 45

- Goyal R., 'The DPDPA 2023 and Its Limitations: A Critical Assessment' (2024) 12 Journal of Indian Law and Society 78