SUPREMO AMICUS

INDIA'S FIRST AI INTEGRATED LAW JOURNAL

Peer Reviewed, Refereed and Open access Journal

- Available in 331+ International Libraries
- Indexed at 32 Databases

ISSN NO. 2456-9704
Volume 10 Issue 2
www.supremoamicus.org

DISCLAIMER

The information presented in this article is intended for general informational and educational purposes only. While every effort has been made to ensure that the content is accurate, up-to-date, and reliable at the time of publication, the editorial board and publisher make no representations or warranties of any kind, express or implied, regarding the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

The views and opinions expressed in this article are those of the author and are based on personal research, experience, and interpretation. They do not necessarily reflect the official policy, position, or opinions of any affiliated organization, institution, or entity.

This article is not intended to serve as professional advice of any kind. The editorial board and publisher shall not be held liable for any errors or omissions in the content, nor for any losses, injuries, or damages arising from the use of or reliance on this information.

ABOUT THE JOURNAL

Supremo Amicus is an online, peer-reviewed international journal devoted to the interdisciplinary fields of law and science. In an era marked by rapid technological progress and evolving legal frameworks, the journal seeks to bridge the gap between these two dynamic domains by offering comprehensive and critical insights into their various aspects. The journal places a strong emphasis on contemporary advancements, emerging trends, and the complex challenges faced by both the legal community.

The primary objective of the journal is to encourage and promote original, high-quality research. It is committed to publishing well-researched, analytically sound, and thought-provoking articles that adhere to rigorous academic standards. Each submission undergoes a thorough peer-review process to ensure authenticity, relevance, and scholarly integrity. In doing so, the journal maintains its commitment to excellence and credibility.

In addition to fostering research, the journal aims to make complex ideas accessible and engaging for a diverse readership. It strives to present content that is not only intellectually enriching but also clearly written and reader friendly.

Furthermore, the journal is committed to promoting interdisciplinary collaboration and global engagement. It welcomes diverse perspectives from contributors across different regions and backgrounds, thereby enriching the quality and scope of discussions presented within its pages.

With this vision we proudly present Supremo Amicus to our readers.

-Editorial Team
Supremo Amicus

TOWARD A SAFER DIGITAL TOMORROW: REIMAGINING INDIA'S LEGAL RESPONSE TO EMERGING AI-DRIVEN HARMS

By *Helan Jesus Mary L.*
From *The Central Law College, Salem*

By *Nallamuniappan N.*
From *The Central Law College, Salem*

Abstract

India's democratic processes, human dignity, and informational integrity are at previously unheard-of peril due to the explosive growth of AI-generated synthetic media, especially deepfake audio, video, and image. In order to protect human rights in the era of artificial intelligence, this study contends that India needs a more comprehensive, victim-centred, and technologically sophisticated regulatory ecosystem. Recently, society is trying to distinguish which video, audio, or image is real or fraudulent, producing uncertainty and panic among users. The inability to identify legitimate content from manipulated media endangers victims and reduces faith in digital information. India's current legal framework lacks explicit protections, required disclosures, and platform accountability for the spread of deepfakes. Victims are left without sufficient remedies or quicker takedown and investigative procedures because of this gap.

The emergence of deepfakes has intensified a range of unresolved structural problems within India's digital ecosystem. The lack of a single national system for recognising, reporting, investigating, and removing deepfake content results in delayed justice. Furthermore, those who are not familiar with legal or technological procedures cannot access or use India's current cybercrime reporting platforms. Ultimately, the paper presents a forward-looking perspective for defending human rights and establishing a safer, more just digital future for all. A strong, future-ready legal framework is essential to secure the digital rights and human dignity of women and girls.

Keywords: Deepfakes, AI-generated content, cybercrime, digital safety, complaint portal.

1. Introduction

Artificial Intelligence (AI) has evolved from a theoretical concept into one of the most transformative technological forces of the 21st century. What began as rule-based computational models in the mid-20th century has now expanded into complex machine-learning systems capable of self-learning, predicting, generating, and mimicking human behaviour. The rapid advancement of deep learning, neural networks, and generative models — especially Generative Adversarial Networks (GANs) and large language models — has accelerated AI's integration into governance, healthcare, education, finance, and communication. AI is no longer a distant innovation; it shapes everyday human interaction, social behaviour, and the global information ecosystem.

However, the same technological capability that powers innovation has also enabled unprecedented forms of harm. Generative AI tools can now produce hyper-realistic images, audio, and videos — commonly known as deepfakes — that are indistinguishable from authentic content. This has raised critical concerns around privacy, consent, misinformation, electoral manipulation, sexual violence, and digital harassment. The mass accessibility of AI tools, combined with their speed and accuracy, has radically increased the scale and impact of synthetic media misuse. Recently, India has experienced a surge in deepfake incidents. Viral fabricated videos have sparked public outrage, misinformation, and online harassment. Ordinary citizens now struggle to identify what is real or fake, resulting in erosion of trust in digital environments. Women and girls are disproportionately victimised, reflecting gendered vulnerabilities in online spaces. As India embraces digital transformation and seeks to become a global AI hub, the nation also faces an

urgent need to recalibrate its legal, ethical, and human rights frameworks. The challenge is not merely technological but fundamentally societal: how can India ensure that innovation does not compromise dignity, autonomy, and democratic values? The development of AI demands a new governance ecosystem — one capable of addressing emerging risks while enabling responsible technological progress.

Women and girls face a significantly higher vulnerability within this evolving AI landscape. Studies worldwide indicate that over 90% of deepfake content targets women, often without their knowledge or consent, leading to irreversible emotional, social, and professional harm. AI-enabled sexualised deepfakes, morphed images, and fabricated intimate videos have become tools of harassment, blackmail, revenge pornography, and character assassination, perpetuating gender-based violence in digital form. The anonymity of perpetrators, rapid virality of content, and the absence of robust legal remedies leave women with limited recourse and heightened psychological trauma. In patriarchal societies like India, such misuse often results in victim-blaming, social stigma, and long-term consequences for education, employment, and marriage prospects. Thus, the gendered impact of AI-generated harms underscores the urgent need for a rights-based legal framework that prioritises the safety, dignity, and agency of women and girls in the digital age.

This paper situates the deepfake crisis within the broader evolution of AI, arguing for stronger legal safeguards, human-rights-based protections, and technologically empowered regulatory mechanisms to defend individuals in the digital era. The study concludes that deepfake harm is not merely a technological challenge but a human rights imperative requiring integrated legal, technological, and policy solutions.

2. Statement of the Problem

The rapid advancement of artificial intelligence has enabled the creation of highly realistic synthetic audio, video, and images—commonly known as *deepfakes*—which pose unprecedented threats to personal dignity, public trust, and democratic discourse in India. As AI-generated media increasingly becomes indistinguishable from authentic content, individuals and authorities face growing difficulty in determining what is real and what is fabricated. This inability to verify authenticity has resulted in widespread confusion, panic, misinformation, and the manipulation of public perception.[1] Women and girls remain the most frequent victims of such misuse, as deepfake technology is commonly weaponised to create sexually explicit, defamatory, or coercive content, leading to severe emotional, social, and psychological harm.[2]

A recent national controversy involving a 19-minute viral Instagram video illustrates the scale and complexity of this emerging challenge. Conflicting claims over whether the clip was AI-generated or genuine left law-enforcement agencies, digital experts, and the public unable to conclusively determine its authenticity.[3] Despite police warnings that circulating such videos — real or AI-generated — constitutes an offence under provisions of the Information Technology Act, 2000, the content continued to spread rapidly. The incident caused

[1]Drishti IAS, "Deepfakes," available at https://www.drishtiias.com/daily-updates/daily-news-analysis/deepfakes-6.

[2]Information Technology Act, 2000; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021; Digital Personal Data Protection Act, 2023; Bharatiya Nyaya Sanhita, 2023 — none of these statutes contain an explicit definition of "deepfake." See Advocate Tanwar, "Deepfake Misuse in India: Need for Stronger Regulation" (legal commentary).

[3]Ministry of Electronics and Information Technology (MeitY), Draft Amendment to the Information Technology Rules, 2021, released 22 October 2025 (defining "synthetically generated information"); Aroon Deep, The Hindu, 22 October 2025.

reputational damage, emotional trauma, and widespread speculation, demonstrating how deepfake-like ambiguity can intensify harm even when the underlying video is real.[4] Additionally, misinformation related to alleged suicide attempts and unrelated viral clips highlighted how social media ecosystems can amplify psychological distress and public misunderstanding.[5]

Despite these growing threats, India lacks a comprehensive legal framework specifically addressing deepfakes. Existing laws such as the Information Technology Act, 2000, the IT Rules, 2021, the Digital Personal Data Protection Act, 2023, and relevant provisions of the Bharatiya Nyaya Sanhita, 2023 regulate only fragmented aspects of identity misuse, privacy invasion, obscene content, or misinformation, without explicitly recognising deepfakes as a distinct technological harm requiring specialised regulation.[6] The absence of mandatory authenticity disclosures, watermarking requirements, rapid takedown mechanisms, and consistent platform accountability leaves victims vulnerable and without timely remedies.[7]

Moreover, current reporting and redressal mechanisms remain difficult to access, especially for women, rural populations, and those unfamiliar with legal or technological procedures. Law-enforcement agencies often lack the technical capacity to detect, trace, and curb AI-generated content swiftly, resulting in delayed justice and prolonged victimisation. Consequently, personal autonomy, democratic integrity, and digital safety remain under increasing threat.

In this context, there is an urgent need to examine the limitations of existing laws, assess the human rights implications of deepfake misuse, and propose a coherent, victim-centred, technologically robust legal framework that can effectively safeguard citizens — especially women and girls — in an AI-driven digital environment.

3. Objective of the Study

The primary objective of this study is to critically examine the legal and regulatory implications of the IT Ministry's proposed amendment mandating the disclosure and labelling of AI-generated synthetic content on social media platforms. It seeks to analyse the effectiveness of compulsory self-declaration by users and proactive detection mechanisms by intermediaries in addressing the growing threat of deepfakes, misinformation, and impersonation. The study further aims to evaluate the technological feasibility and compliance preparedness of social media companies in embedding visible metadata or identifiers covering at least ten per cent of the content area, as required under the draft rules. In addition, it endeavours to assess the broader impact of mandatory AI-labelling on user autonomy, democratic decision-making, and the protection of privacy and reputation in the digital ecosystem. By comparing India's regulatory approach with emerging global standards, the study also aims to identify enforcement challenges and potential risks such as over-regulation, chilling effects on expression, or gaps in implementation. Ultimately, the study intends to offer informed recommendations for strengthening the governance of synthetic content while balancing innovation, user safety, and constitutional freedoms.

[4]Medianama, "Government Releases Draft Proposal Mandating AI Content Labelling" (2025) (analysing the absence of an explicit deepfake offence in Indian law).

[5]Shodh Sagar Law Journal, "Challenges in Regulating Deepfakes in India" (2024) (arguing that the absence of a formal statutory definition of deepfakes creates enforcement gaps and burdens victims).

[6]Cyber Peace Foundation, "Women's Vulnerability in Digital Spaces" (2023) (identifying structural gender biases in online abuse).

[7]Sensity AI, "The State of Deepfakes" (2021) (finding that 90–95% of deepfakes circulating online are non-consensual sexual content targeting women).

4. Understanding Deepfakes and Its Gendered Digital Harm

4.1 What Are Deepfakes?

Deepfakes are synthetic media — videos, images, or audio — that are digitally altered using Artificial Intelligence (AI) to make it appear that someone said or did something they never actually did. They blur the line between reality and manipulation.

Deepfakes are powered by deep learning, a subset of machine learning, which itself is a subset of AI. They are created using Generative Adversarial Networks (GANs), where two neural networks — a generator and a discriminator — work together to create and refine fake content. GANs use real data to recreate faces, voices, or movements: the generator produces fake content while the discriminator attempts to detect it, and the generator improves until it can fool the discriminator. Natural Language Processing (NLP) is used for cloning voices, and lip-syncing techniques align deepfake audio with video.

Deepfakes are synthetic media generated using machine-learning algorithms such as Generative Adversarial Networks (GANs). These systems superimpose or reconstruct faces, voices, or gestures to create media that appears authentic. The sophistication of AI has made deepfakes increasingly accessible, affordable, and difficult to detect. Common types include *face swaps* (replacing a person's face in a video with another's), *voice clones* (imitating someone's voice to say anything), and *source video manipulation* (making someone appear to do or say things they never did).

4.2 Is There a Law in India That Defines "Deepfake"?

India currently does **not** have any enacted law that explicitly defines the term *"deepfake."* Existing legislation — including the Information Technology Act, 2000, the IT Rules, 2021, the Digital Personal Data Protection Act, 2023, and the Bharatiya Nyaya Sanhita, 2023 — regulates only fragmented aspects of digital harm such as identity misuse, obscenity, defamation, and privacy violations, without recognising deepfakes as a distinct legal category. However, in October 2025, the Ministry of Electronics and Information Technology (MeitY) introduced a draft amendment to the IT Rules, 2021 proposing a formal definition for *"synthetically generated information,"* described as information that is artificially or algorithmically created, generated, modified, or altered using a computer resource in a manner that such information reasonably appears to be authentic or true. Although this definition indirectly encompasses deepfake images, audio, text, and videos, it does not specifically use the term *"deepfake."* Thus, unless this draft is formally notified and implemented, India continues to operate within a legal vacuum, addressing deepfake-related harms only through broad and outdated statutory provisions not designed for modern AI-generated content. Scholars have consistently stressed that the absence of an explicit statutory definition complicates enforcement, accountability, evidence collection, and victim protection in deepfake cases.

4.3 Why Women Are Disproportionately Targeted

Women and girls are disproportionately affected by deepfake misuse due to entrenched gender inequalities, patriarchal social norms, and a digital environment that weaponises female identity for exploitation. Studies show that over 90% of deepfake videos circulating online are sexually explicit, and the vast majority of these target women, reflecting a gendered pattern of online violence. In India, this vulnerability is amplified by cultural stigma associated with female sexuality, where even fabricated intimate content can destroy reputations, lead to social isolation, or trigger harassment, blackmail, and threats. Women in public-facing professions — including influencers, journalists, actors, and activists — are especially targeted, as perpetrators use manipulated images or videos to

silence, shame, or discredit them.[8] The rapid spread of such content across social media, combined with the absence of strong legal protections and slow platform response systems, intensifies the emotional and psychological trauma faced by victims.[9] The fear of being disbelieved or blamed further deters many women from reporting these incidents, resulting in significant under-reporting and continued impunity for offenders.[10]

These acts amount to sexual harassment, privacy invasion, and psychological harm, directly infringing on the dignity guaranteed under Article 21 of the Constitution of India.

The widespread circulation of AI-generated synthetic media has created unprecedented levels of public confusion, undermining society's ability to distinguish between truth and fabrication. Deepfakes blur the line between authentic and manipulated content, generating an "information disorder" that destabilises public trust in digital communication. This confusion became starkly visible in the recent 19-minute viral Instagram incident, where conflicting claims about whether the video was AI-generated or genuine left the public, media outlets, and even law-enforcement agencies uncertain.[11] Such ambiguity fuels mass speculation, panic, and emotional distress, allowing misinformation to spread rapidly and unchecked. Deepfake-related confusion also erodes institutional trust, as false videos can be used to manipulate democratic discourse, target political leaders, or incite communal tensions.[12] When citizens are unable to rely on visual or audio evidence, the credibility of digital information collapses, enabling malicious actors to create or deny evidence at will — a phenomenon scholars describe as the "liar's dividend." This erosion of informational integrity results in broader societal harms, including reputational damage, breakdown of social cohesion, increased cyber-vigilantism, and psychological trauma among affected individuals. In the absence of accessible verification mechanisms, strong regulatory protections, and rapid takedown systems, the Indian public remains vulnerable to manipulation, deception, and large-scale misinformation.[13]

5. Constitutional Framework and Human Rights Principles

India's constitutional framework provides strong normative grounds for safeguarding individuals from the harms caused by deepfakes and other AI-generated synthetic media. At its core, Article 21 guarantees the right to life and personal liberty, which the Supreme Court has expansively interpreted to include the right to privacy, dignity, and protection of one's reputation.[14] Deepfake misuse — especially involving non-consensual intimate content — directly violates these constitutional protections by stripping individuals of control over their identity, bodily autonomy, and digital presence. The Supreme Court's landmark judgment in *K.S. Puttaswamy v. Union of India* affirmed privacy as a fundamental right and categorically recognised informational autonomy as

[8]Internet Freedom Foundation, "Gendered Disinformation in India" (2023) (documenting the misuse of AI and manipulated media to silence women journalists and activists).

[9]National Commission for Women, Annual Report (2024) (noting delays in takedown and inadequate cybercrime response mechanisms for women victims).

[10]Amnesty International, "Toxic Twitter Report" (2020) (explaining under-reporting due to fear of stigma and mistrust in reporting systems).

[11]One India News, "Is the Instagram 19-Minute Viral Video Real or AI-Made? Police Warn Against Sharing" (December 2025).

[12]Brookings Institution, "The Threat of Deepfakes to Democracy" (2020) (discussing the manipulation of elections, leaders, and public opinion).

[13]NITI Aayog, "Responsible AI for All" (2023) (highlighting India's vulnerability to misinformation and low digital literacy).

[14]Maneka Gandhi v. Union of India, (1978) 1 SCC 248 (expanding the interpretation of Article 21 of the Constitution of India).

an essential component of human dignity.[15] Deepfakes, by manipulating a person's face, voice, or identity, intrude upon this autonomy and therefore engage constitutional scrutiny.

Further, Article 14 ensures equality before the law and protection against arbitrary state action.[16] When deepfake harms disproportionately affect women and marginalised groups, the State has a constitutional duty to design gender-sensitive and accessible mechanisms for redressal. Article 19(1)(a) protects freedom of speech, but this right does not extend to spreading defamatory, obscene, or deceptive AI-generated content that harms others.[17] At the same time, excessive censorship of AI tools may raise constitutional concerns, making it essential to balance free expression with the rights to privacy, dignity, and safety.

Collectively, the constitutional scheme demands a rights-based regulatory approach to AI governance — one that prioritises harm prevention, ensures accountability of platforms, and upholds the digital dignity of individuals. In the context of a rapidly evolving AI ecosystem, the State's obligation under Articles 21, 14, and 19 requires proactive legislative and policy responses to address deepfake-related risks, protect vulnerable communities, and preserve the integrity of India's democratic and digital spaces.[18]

5.1 International Human Rights Standards and Deepfakes

The rise of AI-generated deepfakes raises serious concerns under international human rights law, particularly in relation to the rights to dignity, privacy, equality, freedom of expression, and protection from gender-based violence. Key global human rights instruments — including the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR), and the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) — establish a clear normative framework obligating states to safeguard individuals against technologically amplified harms.[19] Deepfakes, especially sexually explicit or defamatory forms, threaten these rights by enabling non-consensual imagery, eroding informational authenticity, undermining reputations, and facilitating social exclusion. These harms disproportionately affect women and girls, aligning deepfake abuse with digital gender-based violence recognised under international women's rights standards.

The UDHR, under Articles 1, 3, and 12, protects inherent human dignity, security, and privacy rights directly compromised when deepfakes are used to manipulate a person's image, voice, or likeness without consent.[20] Similarly, ICCPR Article 17 obligates states to prevent arbitrary interference with privacy, reputation, and honour, requiring legal safeguards against digitally altered images that can destroy a person's social standing within minutes.[21] Deepfake-driven misinformation also threatens ICCPR Article 19, which guarantees the right to receive and impart information. The proliferation of synthetic media erodes public trust, blurs the line between truth and falsehood, and undermines democratic discourse — problems increasingly

[15]K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1 (recognising privacy as a fundamental right).
[16]Constitution of India, art. 14 (equality before law and equal protection of laws).
[17]Constitution of India, art. 19(2) (reasonable restrictions on speech, including defamation, public order, and decency).
[18]NITI Aayog, "Responsible AI for All: 2023 Strategy Paper" (emphasising a human-rights-based approach to AI regulation).
[19]Universal Declaration of Human Rights (1948); International Covenant on Civil and Political Rights (1966); Convention on the Elimination of All Forms of Discrimination Against Women (1979).
[20]Universal Declaration of Human Rights, arts. 1, 3, 12.
[21]International Covenant on Civil and Political Rights, art. 17.

recognised by UN Special Rapporteurs on freedom of expression.

From a gender-justice perspective, CEDAW is particularly relevant, as it obligates states to eliminate all forms of discrimination and violence — including technology-facilitated gender-based harms.[22] UN Women and the UN General Assembly have emphasised that digital violence against women, such as non-consensual synthetic pornography and AI-enabled impersonation, constitutes a violation of women's rights to dignity, equality, and bodily autonomy.[23] Deepfake sexual imagery, which overwhelmingly targets women, reinforces harmful stereotypes, causes economic and emotional harm, and pushes victims out of online spaces — contrary to CEDAW's mandate for equal participation in public life.

Additionally, global cybersecurity and digital rights frameworks, including the Budapest Convention on Cybercrime, stress the need for states to strengthen investigative capacity, enhance cross-border cooperation, and develop modern legal tools to respond to evolving cyber harms.[24] Deepfakes, which can be produced anonymously and disseminated globally within minutes, require such coordinated responses. Several democracies — including the EU, the UK, South Korea, and the United States — have begun implementing transparency mandates, authenticity labelling, and explicit deepfake criminalisation, reflecting an emerging global consensus that synthetic-media regulation is essential for protecting human rights.

Taken together, these international norms impose a clear responsibility on India to adopt stronger domestic laws, platform accountability mechanisms, and victim-centred remedies. As deepfakes intensify threats to privacy, dignity, democratic participation, and gender equality, aligning national regulation with global human rights standards becomes essential for safeguarding individuals — especially women and girls — in the digital age.

6. Gaps in the Existing Legal Framework in India

Despite the increasing misuse of AI-generated deepfake technology, India does not yet have a dedicated, comprehensive statute that defines or regulates deepfakes as a specific category of digital harm. Existing provisions under the Information Technology Act, 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, the Digital Personal Data Protection Act, 2023, and the Bharatiya Nyaya Sanhita, 2023 provide only fragmented safeguards that fail to address the unique challenges posed by synthetic media. These laws focus largely on issues such as obscene content, identity misuse, privacy violations, and data protection, but none of them recognise the creation, alteration, or dissemination of deepfakes as an independent offence requiring specialised investigation, reporting, and takedown mechanisms.[25] This gap results in significant ambiguity for users, platforms, and law-enforcement authorities, leading to inconsistent enforcement and delays in victim relief.

A major limitation in India's regulatory architecture is the absence of a clear statutory definition of deepfakes. Without a legal definition, neither courts nor enforcement agencies possess a uniform basis for determining liability or establishing evidentiary standards in cases involving AI-generated manipulation. While the Ministry of Electronics and

[22]Convention on the Elimination of All Forms of Discrimination Against Women, arts. 2, 3, 5.
[23]UN Women, Technology-Facilitated Gender-Based Violence Reports; UN General Assembly Resolutions on Violence Against Women in Digital Contexts.
[24]Council of Europe, Budapest Convention on Cybercrime (2001).
[25]Information Technology Act, 2000; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021; Digital Personal Data Protection Act, 2023; Bharatiya Nyaya Sanhita, 2023.

Information Technology (MeitY), in October 2025, proposed amendments to the IT Rules mandating the labelling of AI-generated synthetic content and requiring platforms to embed unique metadata identifiers, these provisions remain at the consultation stage and have not yet been enacted into enforceable law.[26] This regulatory uncertainty enables perpetrators to exploit loopholes and allows platforms to avoid strict accountability obligations.

Furthermore, the laws currently in force do not impose mandatory timelines for platform takedown of deepfake content, nor do they require intermediaries to use proactive detection tools for harmful synthetic media. Although Rule 3(1)(b) of the IT Rules, 2021 obligates platforms to remove unlawful content upon notice, its application to deepfakes is unclear because the rule was not drafted with AI-generated media in mind.[27] As a result, victims are left to navigate slow complaint systems, unclear procedural pathways, and inconsistent platform practices. This situation disproportionately harms women and girls, who bear the social and psychological burden of synthetic sexual imagery circulating on social media.

The lack of dedicated investigative infrastructure also remains a major challenge. Law-enforcement agencies often lack the technical expertise and forensic tools needed to trace, authenticate, or confirm the origin of AI-generated videos. Existing cyber cells are overloaded, and police officers receive no mandatory training on AI-driven harms.[28] The burden therefore shifts to victims, who must prove authenticity, establish manipulation, and contend with jurisdictional delays across multiple digital platforms.

In sum, India's regulatory landscape remains significantly underprepared for the scale and sophistication of deepfake-related offences. Without a comprehensive, victim-centred statutory regime that explicitly defines deepfakes, mandates platform accountability, and provides fast-track remedies and forensic support, the digital rights and human dignity of users — especially women — remain inadequately protected.

7. Recommendations for a Future-Ready Legal Framework

To protect citizens — particularly women and girls — from AI-driven harms, India requires a robust, transparent, and technologically equipped legal ecosystem. First, the law must explicitly define deepfakes as a standalone offence within the IT Act or the Bharatiya Nyaya Sanhita (BNS), distinguishing between benign, creative uses and harmful, non-consensual, or defamatory applications. Clear categorisation is crucial for ensuring consistent enforcement and preventing misuse. A statutory mandate requiring AI provenance tools, such as cryptographic watermarking or metadata, should also be introduced to ensure traceability and support the admissibility of evidence in courts.

Second, India should implement mandatory labelling of AI-generated content, in line with the 2025 proposal by the Ministry of Electronics and Information Technology, which requires platforms to embed visible disclosures covering at least 10% of the screen area. Platforms must also adopt proactive detection technologies rather than relying solely on user complaints, especially for sexually explicit synthetic media disproportionately targeting women. Establishing platform liability norms — similar to the EU Digital Services Act — would ensure accountability for delayed removals, non-compliance

[26]Ministry of Electronics and Information Technology (MeitY), Draft Amendment to the Information Technology Rules, 2021, proposing mandatory labelling of AI-generated content (22 October 2025).

[27]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r. 3(1)(b).

[28]National Crime Records Bureau & Indian Computer Emergency Response Team (CERT-In), reports repeatedly highlighting capacity limitations in cyber-forensic training across police departments.

with takedown orders, and failure to protect vulnerable users.

Third, a National Deepfake Monitoring and Response Cell should be established to provide a centralised mechanism for reporting, verifying, and investigating deepfake incidents. This body should include cyber-forensics experts, AI engineers, legal officers, and mental health professionals to ensure holistic victim assistance. Fast-track procedures for takedowns — such as mandatory platform response within six hours for sensitive content — are essential to preventing further circulation and psychological harm.

Finally, digital literacy programmes must be integrated into school, college, and community-level curricula to enable citizens to identify manipulated content, understand legal protections, and access complaint portals. Special provisions should be developed for women, minors, and rural users who face the highest vulnerability. Together, these measures can create a future-ready legal and technological architecture that upholds human rights, safeguards dignity, and promotes trust in India's digital ecosystem.

8. Conclusion and Suggestions

The rapid advancement of artificial intelligence — especially deepfake technology — has created a complex intersection of legal, ethical, and human rights challenges. While AI offers transformative possibilities, its misuse threatens personal dignity, privacy, gender equality, electoral integrity, and public trust. Deepfakes, in particular, erode societal confidence by blurring the boundary between truth and fabrication. Women and girls endure the worst consequences, facing image-based sexual abuse, harassment, and lasting psychological harm. The absence of clear statutory definitions and specialised regulations in India further aggravates these vulnerabilities, creating loopholes that perpetrators can easily exploit.

A comprehensive legal, technological, and institutional response is urgently needed. Strengthening current laws, introducing deepfake-specific definitions, and ensuring platform accountability must become national priorities. Judicial recognition of informational privacy, dignity, and autonomy provides a constitutional foundation, but operationalising these rights requires proactive legislative action. Equally important is the creation of accessible reporting mechanisms, gender-sensitive investigative protocols, and fast-track remedies for victims of AI-generated harms. Public awareness and digital literacy must be expanded so that individuals can recognise manipulated media and protect themselves online.

Technological interventions — such as mandatory provenance metadata, watermarking, and the deployment of advanced AI-detection tools — should be integrated into a broader governance framework. Collaboration between government, technology companies, researchers, civil society, and international bodies is essential for developing global standards. As AI rapidly evolves, India must adopt a forward-looking approach that balances innovation with strong safeguards.

In conclusion, addressing deepfake harms requires a multi-layered strategy grounded in constitutional rights, human dignity, and gender justice. The State must adopt a future-ready legal framework that ensures accountability, protects vulnerable groups, and strengthens societal trust in the digital age. Meaningful reform will empower individuals, preserve democratic values, and ensure that technological progress serves humanity rather than undermining it.

*****