
DEEFAKE TECHNOLOGY: EMERGING THREATS AND LEGAL SAFEGUARDS FOR WOMEN IN THE DIGITAL ERA

Harshita Malviya, LL.M. (Criminal Law), Barkatullah University, Bhopal.

I. INTRODUCTION

Deepfake technology represents one of the most transformative and troubling developments in the digital era. The term *deepfake* merges “deep learning” and “fake,” referring to synthetic media generated using artificial intelligence (AI) and machine learning algorithms that convincingly mimic real people’s faces, voices, and actions in videos, images, or audio recordings. Relying primarily on **Generative Adversarial Networks (GANs)**ⁱ, Deepfakes pit two neural systems, the generator, which produces artificial content, and the discriminator, which evaluates its realism against one another to create simulations indistinguishable from authentic media. This technology, originally developed for research and entertainment, has since become easily accessible through open-source software and mobile applications, contributing to its exponential proliferation across digital platforms.

While deepfakes have legitimate and creative applications in cinema, education, and virtual reality, their malicious use has become a major global concern. The ability to fabricate hyper-realistic audiovisual content undermines public trust, distorts reality, and poses a profound threat to personal dignity, identity, and privacy. This threat is particularly severe for women, who disproportionately become victims of non-consensual deepfake pornography and image-based sexual abuse. The gendered dimension of this harm highlights broader societal issues of digital misogyny, victim-blaming, and the silencing of women’s voices in online spaces.

In India and across the world, the emergence of deepfakes has exposed significant gaps in existing legal and regulatory frameworks. Traditional cybercrime and privacy laws struggle to address these complex, technologically mediated violationsⁱⁱ. Addressing such challenges requires a nuanced understanding that bridges technology, ethics, and law. Hence, this article aims to explore the contemporary threats posed by deepfake technology and analyse the legal safeguards, both existing and emerging, that seek to protect women in the digital era. By situating the discourse within global and Indian contexts, the discussion underscores the urgent need for a human-centric, victim-oriented approach to counter this evolving menace.

II. THE EVOLUTION OF DEEPPFAKE TECHNOLOGY

Deepfake technology has experienced a rapid evolution from its niche origins in academic research to becoming one of the most disruptive forces in digital media. Early experiments with artificial intelligence and neural networks set the foundation in the 2010s, enabling computers to recognise, generate, and manipulate images and sounds with increasing sophistication. The advent of Generative Adversarial Networks (GANs), where two neural networks compete to produce and verify synthetic outputs, marked a breakthrough, allowing for the creation of hyper-realistic deepfake videos, audio files, and images that are nearly impossible to distinguish from genuine contentⁱⁱⁱ.

Initially, deepfake technology was celebrated for its creative potential. In cinema and entertainment, it enabled seamless visual effects, de-ageing of actors, and digital resurrection of historical figures. Virtual reality and educational platforms leveraged deepfakes to create immersive experiences and dynamic learning environments. However, as open-source code and consumer applications democratized access, malicious actors soon exploited these capabilities for deception, defamation, and exploitation.

The global proliferation of deepfakes has accelerated dramatically in the past five years. With smartphones and personal computers capable of running powerful deepfake algorithms, millions of altered media pieces now flood digital networks annually. Women and public figures have become frequent targets, facing risks of non-consensual pornography, cyberbullying, identity theft, and character assassination.^{iv} Automated tools and viral sharing amplify the reach and impact of such content, making detection and control increasingly difficult.

Today, deepfake technology walks a fine line between innovation and abuse. The boundary between creative and criminal use is blurred as media realism outpaces conventional methods of authentication and verification. This unsettling trajectory compels policymakers and technologists to rethink digital safety, emphasising the urgent need for robust countermeasures and legal safeguards to prevent harm, especially for vulnerable groups in society.

III. EMERGING THREATS: GENDERED IMPLICATIONS OF DEEPPFAKE TECHNOLOGY

Deepfake technology poses distinct and severe threats to women, making it a pressing gendered

issue within the broader context of digital harm. Several key dimensions define how women are impacted:

A. Non-consensual pornography and image-based sexual abuse:

Research consistently finds that the overwhelming majority of deepfake videos circulating online, up to 96%, are non-consensual pornographic or explicit content, with 99% targeting women. Public figures such as actresses, musicians, journalists (e.g., Scarlett Johansson, Gal Gadot, Rana Ayyub), and increasingly “everyday” women are targeted with fabricated explicit content. These digital forgeries are not only tools of humiliation, but also acts of revenge and blackmail, exposing victims to extortion and threat.

B. Psychological and social harm:

The fallout from deepfake abuse can be devastating, leading to anxiety, depression, social ostracism, reputational ruin, and, in some cases, threats of physical violence or stalking. Shame and stigma attached to being a victim of deepfake abuse can cause lasting trauma and impact personal relationships as well as professional opportunities for women^v.

C. Reinforcement of gendered stereotypes and misogyny:

Deepfakes perpetuate objectification, unrealistic beauty standards, and male gaze dynamics, which reinforce societal patterns of control, discrimination, and virtual domination over women. This often silences female voices and discourages women from participating in online life and public discourse.

D. Cyber harassment and blackmail:

Deepfakes have amplified the reach and impact of online harassment, including stalking, doxing, and sextortion. Because perpetrators can remain anonymous behind digital platforms, victims frequently lack the means to identify or pursue justice against their abusers.

E. Global patterns: India, US, Europe:

While celebrities have historically been the most visible victims, cases from India (e.g.,

journalists, activists), Australia, the UK, South Korea, and the Americas show that deepfake abuse affects women across cultures. In Latin America, for instance, teenage girls have increasingly reported being targeted with deepfake “revenge porn” following social disagreements or breakups.

Despite some progress through content bans and legal changes in select countries, most women victims have few avenues for swift redress or protection. The rapid viral spread and realistic nature of deepfakes make removal and reputational repair nearly impossible in many cases, underscoring the urgent need for stronger legal, technological, and community safeguards.

IV. LEGAL FRAMEWORK: GLOBAL OVERVIEW

The alarming rise in deepfake-related threats has spurred governments and international bodies to examine and update their legal frameworks, with a special focus on protecting women and vulnerable populations from digitally mediated harm^{vi}. Approaches vary widely across regions, with a mix of criminal legislation, content regulation, and civil remedies.

- A. Europe:** The law is broad enough to address both sexual exploitation and reputation damaging videos. Other EU member states reference the Digital Services Act (DSA), which imposes strict platform accountability and user rights, demanding the timely takedown of illegal content and transparency about automated detection tools.
- B. United Kingdom:** The UK’s Online Safety Act (2023) represents a pioneering regulatory move, criminalising malicious deepfakes and placing direct obligations on social media platforms to detect, remove, and report such content. The Act also creates new criminal offences for fake media designed to harm or harass individuals, especially in gender-based abuse contexts, and gives regulators the authority to levy significant fines against non-compliant tech firms.
- C. United States:** The US approach remains fragmented, with state-level laws providing varying degrees of protection. California’s AB-602, for example, prohibits non-consensual deepfake pornography and enables victims to sue for damages. Virginia’s Code § 18.2-386.2 similarly criminalises the distribution of deepfake sexual content. However, federal law lags; there is no uniform nationwide statute addressing deepfakes directly, resulting in jurisdictional gaps and inconsistent enforcement.

D. International Regulation: The United Nations and European Union have introduced mandates for watermarking or labelling AI-generated content, and the upcoming EU AI Act includes requirements for platforms to clearly declare synthetic media and provide means for users to contest or report it. Cross-border digital rights treaties aim to harmonise content standards and facilitate swift takedowns of harmful deepfakes.

Overall, while some countries have moved swiftly to criminalise deepfake abuse targeting women, global regulation remains uneven and fragmented. The sheer speed and adaptability of the technology continue to challenge lawmakers, highlighting the need for greater international cooperation and harmonisation of enforcement practices.

V. LEGAL LANDSCAPE IN INDIA

India faces significant legal and regulatory challenges in combating deepfake-related harms, but a combination of statutory provisions, court rulings, and government advisories provides important legal safeguards for women in the digital era^{vii}.

1. Existing Statutes

A. Information Technology Act, 2000 (IT Act):

- **Section 66C:** Addresses identity theft.
- **Section 66D:** Covers impersonation and cheating using electronic means.
- **Section 66E:** Protects privacy against unauthorised publication of images or data.
- **Section 67, 67A, 67B:** Criminalise publishing or transmitting obscene, sexually explicit, or exploitative content, including child sexual abuse materials via electronic media; these sections grant prosecution rights for deepfake pornography.
- **Section 69A:** Empowers authorities to block access to unlawful online content, including deepfakes.
- **Section 79:** Mandates intermediaries (social media platforms) to exercise due diligence and remove unlawful content promptly, or risk losing “safe harbour”

protections.

B. Digital Personal Data Protection Act, 2023 (DPDP Act):

- Penalises use of personal data for deepfakes without user consent.
- Provides for robust security and privacy safeguards. C.

Bharatiya Nyaya Sanhita, 2023 (BNS):

- **Section 353:** Penalises the creation/distribution of misinformation and false statements, including those facilitated by deepfakes.
- **Section 111:** Covers organised cybercrimes involving synthetic media. D.

Indian Penal Code:

□ **Sections 499, 500** cover criminal defamation applicable to deepfake videos designed to damage reputation; **IPC Section 465:** Punishment for forgery. **IPC Section 469:** Forgery for the purpose of harming a person's reputation. **E. IT Rules 2021 (amended 2023):**

- These rules require platforms to proactively detect, block, and expeditiously remove deepfake content.
- Ministry advisories (2023-2024) reinforce intermediary obligations and platform accountability for “**synthetic media**” and **deepfakes**.

F. Indecent Representation of Women (Prohibition) Act, 1986 – Covers sexual deepfakes involving women. (Not specifically mentioned the term “deepfake”).

2. Judicial Precedents:

Landmark cases, such as **Ankur Warikoo v. John Doe and Ors** (CS(COMM) 514/2025) and **Sadhguru Jagadish Vasudev & Anr. v. Igor Isakov & Ors** CS(COMM) 578/2025. Also, the rulings involving Akshay Kumar and Rashmika Mandanna, resulted in courts granting swift injunctions for personality, image, and likeness rights against deepfake misuse.

The Delhi High Court and Bombay High Court recognise personality rights as a proprietary

asset, aligning legal protections with intellectual property law in the context of digital identity misuse.

While no definitive Supreme Court precedent exists yet, cases before High Courts provide insight into the legal remedies being pursued by women against AI deepfakes:

In September 2025, the Delhi High Court granted an injunction against the misuse of Aishwarya Rai Bachchan's identity through AI-generated deepfake videos.

Female influencer: In July 2025, the Delhi High Court ordered social media platforms Meta and X to remove AI-generated obscene content targeting a 15-year-old female influencer.^{viii}

3. Emerging Reforms and Challenges:

- A.** Recent government advisories urge platforms to enhance detection and deletion systems for malicious deepfakes.
- B.** Despite these efforts, India lacks dedicated legislation specifically targeting deepfakes; enforcement is slow, and redress for victims remains complex and often inaccessible, especially in cases of sexual content.

India's legal framework provides several avenues for both criminal and civil remedies, but the mosaic approach suffers from fragmentation and enforcement delays. The absence of a comprehensive deepfake-specific statute limits rapid response, victim support, and structured deterrence for digital harms, making urgent reform a top priority for lawmakers and civil society.

VI. ETHICAL, PSYCHOLOGICAL, AND SOCIETAL DIMENSIONS

The impact of deepfake technology on women transcends legal, technical, and privacy concerns, penetrating the ethical, psychological, and societal realms and reinforcing cycles of harm and stigma.

1. Psychological harm and trauma:

Victims of deepfake abuse report profound psychological distress, including anxiety, depression, shame, helplessness, and trauma akin to post-traumatic stress disorder. Women

targeted by non-consensual deepfake pornography and harassment often experience a loss of control over their identities and bodies, leading to a severely impaired sense of self. The fear of recurrence, public scrutiny, and the often-viral nature of deepfakes exacerbate feelings of violation and isolation.

2. Social stigmatisation and silencing:

The fallout of being falsely associated with explicit or compromising content leads to social ostracisation, damaged reputations, and even professional disadvantages regardless of the content's veracity. Victims may withdraw from online and public life, suffering the "silencing effect" described by human rights advocates. The persistence of deepfakes online further solidifies stigma and denies women the right to reclaim their image and dignity.

3. Ethical dilemmas- consent and autonomy:

Deepfake creation without explicit consent violates core ethical principles, infringing upon autonomy, privacy, and control over one's likeness. The weaponisation of a woman's image and identity turns digital harm into a form of gendered violence, reinforcing misogynistic narratives. The line between free expression and protection from abuse is frequently debated, calling for a nuanced approach that balances rights.

4. Reinforcement of stereotypes and objectification:

Deepfakes perpetuate objectification, dehumanisation, and unrealistic beauty standards, embedding harmful gender stereotypes in digital culture. By commodifying women's bodies, they reinforce specifically patriarchal ideologies and reduce women to objects of manipulation.

5. Challenges in memory and personal identity:

Research suggests deepfakes can even disrupt personal memory and identity, causing victims to doubt their own reality. Fabricated images and videos may be so convincing that they alter how others and even victims themselves recall past events.

6. Prevention through advocacy and literacy:

Raising public awareness, promoting digital literacy, and advocating for legal reform are crucial in countering the psychological harms and ethical transgressions produced by

deepfakes. Community support and survivor-led movements are crucial to the healing and empowerment of women facing these threats.

Deepfake technology not only injures at the level of individual victims but also undermines collective trust, social participation, and the prospects for gender equality in digital spaces.

Addressing these dimensions is essential for developing comprehensive safeguards and fostering more resilient, respectful online communities.

VII. COMBATING THE THREAT: POLICY AND TECHNOLOGICAL SOLUTIONS

The fight against deepfake-enabled abuse requires a concerted blend of regulatory reform, technological innovation, and international cooperation. Recent proposals and evolving policies, especially in India and worldwide, highlight significant directions for safeguarding women in the digital era.

1. Policy Recommendations:

A. Targeted Legislation:

India is introducing stricter IT rules that mandate platforms to clearly label AI-generated and deepfake content with visible markers covering at least 10% of a video or audio's display area. Social media giants must require users to declare if uploaded content is synthetic, take reasonable steps to check claims, and maintain transparent labelling systems.

B. Platform Accountability & Victim-Centric Reform:

Platforms are obliged to block, trace, and remove synthetic media promptly, and legal protection is offered to those acting in good faith against deepfake uploads. Feedback mechanisms, grievance redressal, and fast-track digital tribunals are under consideration to improve victims' ability to seek justice.

C. International Measures:

EU's Digital Services Act and US "Take It Down Act" require rapid takedowns of

harmful deepfakes, users' right to request removal, and substantial fines for non-compliance. Denmark is pioneering digital identity protection with compensation clauses for victims. These efforts aim to harmonise digital rights and law enforcement across borders.

2. Technological Interventions:

A. AI-Driven Detection and Watermarking:

Platforms like Sensity AI and others offer multimodal, real-time deepfake detection with accuracy rates above 95%. Tools analyse video, audio, and text for manipulations and integrate liveness and face-matching checks to prevent identity theft and fraud^{ix}.

B. Blockchain-Based Verification:

Blockchain solutions provide immutable records of original media, tracking its provenance and exposing alterations, thus supporting content authenticity and rapid identification of manipulated materials.

C. CERT-In Advisories:

The Indian Computer Emergency Response Team (CERT-In) has issued advisories on recognising, preventing, and reporting deepfake harm. Preventive measures include limiting high-resolution media sharing, cross-verifying sources, enabling multi-factor authentication, and strengthening privacy controls^x.

D. Community and Advocacy:

Broad educational campaigns and digital literacy programs are essential for empowering users to detect deepfakes, understand the risks, and report abuse. Collaborative efforts between civil society, law enforcement, and technology firms are key to creating safer online environments for women.

Combining policy, technology, and advocacy yields a more resilient landscape for detecting, deterring, and punishing the misuse of deepfake technology, creating robust safeguards for women and all digital citizens.

VIII. CONCLUSION

Deepfake technology stands as a double-edged sword in the digital era, merging creative innovation with grave threats to personal integrity, privacy, and public trust. While its potential for positive change in sectors like entertainment and education is undeniable, its capacity for harm, particularly targeting women through non-consensual pornography, harassment, and reputation damage, puts human rights, safety, and equality at risk^{xi}.

Current legal and regulatory frameworks, both in India and globally, reveal a landscape in transition. Traditional laws struggle to address the speed and complexity of synthetic media abuse, and although new policies and technological interventions are emerging, enforcement and victim support remain inadequate for the scale of the threat^{xii}.

Ultimately, the solution lies in building synergies among law, technology, and community action. Effective labelling, rapid detection, platform accountability, and international cooperation must be paired with advocacy, education, and survivor support. Global solidarity, informed policymaking, and technological vigilance are essential to making digital spaces safer and more equitable, giving women the protection and dignity they deserve in the digital age.

ENDNOTES:

- ⁱ Ministry of Electronics & Information Technology (Government of India), 'Perspective: Combating Deepfakes' (Press Release, 3 January 2024) <https://www.drishtiias.com/loksabha-rajyasabhadiscussions/perspective-combating-deepfakes> accessed 2 January 2026
- ⁱⁱ eSafety Commissioner (Australian Government), 'Deepfake trends and challenges' (Position Statement, January 2022) https://www.esafety.gov.au/sites/default/files/2022-01/Deepfake-positionstatement%20_v2.pdf accessed 2 January 2026
- ⁱⁱⁱ Press Information Bureau (Government of India), 'India well-equipped to tackle evolving online harms and cyber crimes' (Press Release, 25 December 2023), <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2154268> accessed 2 January 2026
- ^{iv} CIGI, 'Women, Not Politicians, Are Targeted Most Often by Deepfake Videos' (2 March 2021) <https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/> accessed 2 January 2026
- ^v Andrea Birrer, 'What we know and don't know about deepfakes' (2024) 24 New Media & Society 741 <https://journals.sagepub.com/doi/10.1177/14614448241253138>
- ^{vi} Asian Institute of Research, 'Deepfake Technology in India and World: Foreboding and Forbidding' <https://www.asianinstituteofresearch.org/lhqrarchives/deepfake-technology-in-india-and-world:-forebodingand-forbidding> accessed 2 January 2026
- ^{vii} Press Information Bureau (Government of India), 'Government of India Taking Measures To Tackle Deepfakes' (Press Release, 3 April 2025) <https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2119050> accessed 2 January 2026
- ^{viii} Kamya Buch v. JIX5A & Ors. (CS(OS) 465/2025)
- ^{ix} India proposes strict rules to label AI content citing growing misuse Reuters (22 October 2025) <https://www.reuters.com/business/media-telecom/india-proposes-strict-it-rules-labelling-deepfakes-amid-aimisuse-2025-10-22/> accessed 2 January 2026
- ^x Blackbird.AI, 'Deepfake Detection Solutions: Innovations and Best Practices' (20 May 2025) <https://blackbird.ai/blog/deepfake-detection-solution/> accessed 2 January 2026
- ^{xi} Deepfake Pornography: Examining the Impact on Women's Rights' (2024) <https://pdfs.semanticscholar.org/283e/72f9eb3522426708f7cdc06e599fea3edfb9.pdf> accessed 2 January 2026
- ^{xii} India proposes strict rules to label AI content citing growing misuse' Reuters (22 October 2025) <https://www.reuters.com/business/media-telecom/india-proposes-strict-it-rules-labelling-deepfakes-amid-aimisuse-2025-10-22/> accessed 2 January 2026