Page: 868-905 *URL: www.lijdlr.com*

LAWFOYER
INTERNATIONAL
JOURNAL OF
DOCTRINAL LEGAL
RESEARCH

ISSN: 2583-7753

LawFoyer International Journal of Doctrinal Legal Research

[ISSN: 2583-7753]

Volume 3 | Issue 4

2025

DOI: https://doi.org/10.70183/lijdlr.2025.v03.158

DEEPFAKE AI AND CRIMINAL LAW: A NEW AGE THREAT TO WOMEN'S SAFETY

Srishti Sehgal[1]

I. ABSTRACT

Technological innovation in Artificial Intelligence (AI) has given rise to "deepfakes" — hyper-realistic synthetic images, videos, and audio generated through deep learning algorithms that can convincingly depict individuals in fabricated scenarios. While this technology has creative potential, its misuse has evolved into a disturbing digital threat, particularly against women. Non-consensual sexual deepfakes, cyberstalking, identity theft, defamation, and extortion have become modern forms of gender-based violence, undermining women's dignity, privacy, and mental health. This research critically examines the intersection of deepfake technology and criminal law, assessing whether existing legal provisions under the Indian Penal Code (IPC) and the Information Technology Act, 2000 are sufficient to address AI-driven sexual exploitation and image-based abuse. It adopts a doctrinal, comparative, and socio-legal methodology, integrating psychological studies and international legal developments, including the U.S. Take It Down Act (2025), the U.K. Online Safety Act (2023), and the EU AI Act. Through analysis of case law, policy gaps, and emerging judicial responses—such as the Bombay High Court's 2025 deepfake-takedown order—this paper argues that India's existing legal mechanisms remain fragmented and inadequate. It advocates for a dedicated deepfake legislation, mandatory takedown timelines, platform accountability, and institutional support systems for victims. The study concludes that safeguarding women in the age of artificial intelligence requires a proactive, rights-based legal framework that harmonizes technological innovation with human dignity.

II. KEYWORDS

Deepfake Technology, Non-Consensual Intimate Imagery, Criminal Law Reform, Women's Digital Safety, Artificial Intelligence Regulation.

[1] B.A. LL.B (Hons.), K.R. Mangalam University (India). Email: srishtisehgal8@gmail.com

III. INTRODUCTION

Artificial Intelligence (AI) has emerged as the defining technology of the twenty-first century, influencing domains as varied as healthcare, finance, education, and criminal justice. Among its most potent yet perilous manifestations are deepfakes—synthetic audio-visual artefacts generated through *deep-learning* algorithms that simulate human likeness with extraordinary fidelity. By mapping facial landmarks, synthesizing voice timbre, and emulating gestures, deepfakes collapse the evidentiary distinction between the real and the fabricated. What began as an experiment in creative mimicry has metamorphosed into a sophisticated tool of misinformation, political subversion, and gendered exploitation.

The rise of open-source generative models such as Stable Diffusion, DeepFaceLab, Roop, and Face Fusion has democratized this capability. Anyone with a standard GPU can fabricate a convincing video within minutes. This accessibility, while technologically remarkable, erodes social trust in imagery and destabilizes long-standing evidentiary assumptions in law and journalism. A century ago, a photograph was treated as near-conclusive proof; today, even the most vivid recording may be a data-driven illusion.

In the gendered context, the stakes are far graver. Over 95 percent of publicly available deepfakes depict sexualized representations of women without consent. These *non-consensual sexual deepfakes* (NCSDs) constitute a new species of image-based sexual abuse. They weaponize a woman's face and voice to generate false pornography, spreading humiliation that cannot be contained by geography or time. The result is what scholars now term "digital sexual violence."

For Indian women—already navigating an online environment marked by trolling, stalking, and moral policing—deepfakes represent an *amplified continuum* of patriarchal control. Victims experience severe reputational and psychological injury: depression, self-censorship, and social isolation. Yet offenders exploit legal lacunae and technological anonymity with near impunity.

India's principal criminal statutes, the Indian Penal Code (IPC) of 1860 and the Information Technology (IT) Act of 2000, were conceived for an analogy world. They envisage tangible acts—photography, recording, or publication—but not algorithmic simulation. The doctrine of *actus reus* tied to physical acts becomes tenuous when the "act" is a computational process generating synthetic data. This mismatch creates a juridical vacuum between harm and accountability.

Further complicating enforcement is the cross-border nature of digital dissemination. A deepfake produced in one jurisdiction may go viral through servers in another before the victim even discovers it. Mutual Legal Assistance Treaties (MLATs) remain sluggish; takedown requests to global platforms often languish without response. For victims, time is justice, and delay equates to denial.

From a constitutional standpoint, the proliferation of deepfakes tests the contours of Article 21, which guarantees the right to life and personal liberty—including dignity and privacy as recognized in *Justice K.S. Puttaswamy v. Union of India* (2017). When a woman's digital likeness is manipulated, her bodily and informational autonomy are simultaneously violated. Protecting *digital dignity* must therefore become as central to constitutionalism as protecting physical safety. This research proceeds from a normative conviction that the defence of truth and dignity in cyberspace is an extension of the right to life itself. Legal inertia in the face of technological disruption endangers not only women's safety but the credibility of justice.

A. RESEARCH PROBLEM

The study interrogates whether India's present criminal-law and cyber-law framework adequately shields women from deepfake-enabled offences—voyeurism, cyber-stalking, defamation, extortion, and synthetic pornography. Where protection proves inadequate, the inquiry identifies the legislative, judicial, and institutional reforms required to close this gap. The problem is therefore both doctrinal and policy-oriented: *how can Indian criminal jurisprudence evolve to confront harms born in the algorithmic age?*

B. OBJECTIVES

- Trace the technological evolution, mechanics, and gendered misuse of deepfakes.

- Analyse overlaps between deepfake pornography and existing offences under the IPC and IT Act.
- Assess the psychological, social, and reputational consequences for women victims.
- Compare Indian jurisprudence with global regulatory responses (U.S., U.K., E.U., Denmark, South Korea).
- Recommend reforms integrating legal, technological, and psychosocial safeguards consistent with constitutional morality and international norms.

C. HYPOTHESIS

India's current criminal-law framework is inadequate to address the sui generis harms created by deepfake technology. The absence of explicit statutory definitions, evidentiary standards, and victim-support mechanisms results in ineffective deterrence and limited psychological or compensatory redress for women subjected to AI-driven image-based sexual abuse.

D. SCOPE AND LIMITATIONS

The paper confines itself to gender-specific misuse: non-consensual sexual deepfakes, impersonation, defamation, and cyber-extortion. Political or satirical deepfakes are referenced only where they intersect with gendered targeting or reputational injury.

Given the rapid evolution of generative models and the paucity of empirical reporting (most victims remain silent due to stigma), this research adopts a doctrinal comparative rather than empirical methodology. It analyses statutes, judgments, and policy instruments to derive normative and procedural recommendations.

Temporal scope extends from 2017 — the advent of publicly available GAN architectures — to 2025, encompassing the global legislative surge on synthetic-media regulation.

E. METHODOLOGY

A hybrid approach combining doctrinal, comparative, and socio-legal analysis is adopted.

- **Primary Sources:**
  - Indian Penal Code §§ 354C (Voyeurism), 354D (Stalking), 499–500 (Defamation), 509 (Insult to Modesty).
  - Information Technology Act §§ 66E (Violation of Privacy), 67 & 67A (Obscenity and Sexually Explicit Material).
  - Constitution of India Art. 21 (Right to Life and Dignity).
  - Leading precedents such as *Justice K.S. Puttaswamy v. Union of India* (2017) and *Shreya Singhal v. Union of India* (2015).
- **Secondary Sources:**
  - Academic and policy literature on AI ethics, gendered cyber-violence, and digital privacy.
  - Reports from MeitY, UN Women, Pew Research Center, and Regula Forensics.
  - Expert interviews and media analyses documenting victim experiences.
- **Comparative Lens:**
  - U.S. federal and state legislation such as the TAKE IT DOWN Act (2025) and ELVIS Act (2024).
  - U.K. Online Safety Act (2023) imposing platform duties of care.
  - E.U. Digital Services Act (2022) and forthcoming AI Act (2025) mandating watermarking.
  - Denmark's personality-rights amendments (2024).
  - South Korea's zero-tolerance model under its Sexual Violence Prevention Law.

Each jurisdiction is examined for conceptual clarity, enforcement practicality, and cultural transposability to the Indian context.

F. SIGNIFICANCE OF STUDY

The significance of this inquiry lies in bridging the gap between technological capability and normative protection. While digital India envisions AI as an engine of growth, the same infrastructure may amplify gendered harm if unregulated. By mapping legal insufficiencies and proposing a deepfake-specific statutory response, this study contributes to the evolving discourse on digital constitutionalism—the extension of constitutional values into cyberspace.

It also serves a pragmatic function for policymakers, suggesting model provisions for a proposed *Deepfake Regulation Act*, and for educators, framing digital literacy as an essential component of women's safety.

IV. DEEPFAKE CRIMES AGAINST WOMEN

A. MECHANICS AND ACCESSIBILITY

Deepfakes are created primarily through Generative Adversarial Networks (GANs) and diffusion models, both subsets of machine learning. In a GAN, two neural networks—the *generator* and the *discriminator*—compete in a feedback loop: the generator attempts to produce synthetic data (for example, a human face), while the discriminator evaluates its authenticity. Through repeated iterations, the generator learns to produce hyper-realistic output that deceives even human observers.

More recently, diffusion models such as *Stable Diffusion* and *DALL-E 3* have overtaken GANs due to their superior capacity to capture fine-grained features like lighting, texture, and micro-expressions. By 2025, repositories such as GitHub, Hugging Face, and CivitAI collectively host over 35,000 pre-trained models, many of which are explicitly optimized for pornographic generation.

Techniques like Low-Rank Adaptation (LoRA), DreamBooth, and ControlNet have further reduced computational costs and increased personalization. Users can now "train" models using as few as ten reference photographs to replicate an individual's face or body in minutes. Combined with cloud rendering and AI video synthesis platforms, deepfake production has been democratized—a process once confined to expert coders is now accessible through smartphone applications and web-based generators.

This democratization marks a shift from elite cybercrime to mass participation in digital abuse. The same algorithms that drive innovation in art, film, or marketing are weaponized to degrade, blackmail, or humiliate women. As the barriers to entry fall, so too does the threshold of accountability, creating an ecosystem where technological empowerment coexists with ethical erosion.

B. TYPOLOGIES OF HARM

Deepfake abuse manifests in multiple interrelated forms, each engaging distinct legal and moral considerations:

- **Non-Consensual Sexual Deepfakes (NCSD):** The most prevalent form, NCSD involves inserting a woman's likeness into sexually explicit imagery or video without consent. Globally recognized as Image-Based Sexual Abuse (IBSA), this category blurs the distinction between voyeurism and sexual assault. Victims experience trauma equivalent to being "digitally violated."
- **Defamation and Character Assassination:** Deepfakes have been deployed to manufacture scandals—showing women, often activists or politicians, in compromising acts to destroy public credibility. This misuse merges gendered harassment with political manipulation, weaponizing sexual morality as a tool of suppression.
- **Impersonation and Identity Theft:** Synthetic cloning of voice and face facilitates financial fraud, phishing scams, or "catfishing." For example, AI voice replication has been used to trick family members into transferring money or to fabricate audio of women "confessing" to moral transgressions.
- **Cyber-Extortion and Coercion:** Offenders threaten to publish manipulated sexual material unless victims pay money or comply with

sexual or political demands. The threat alone—regardless of authenticity—can devastate victims socially and psychologically.

- **"Nudification" Applications:** AI programs that algorithmically remove clothing from photographs, such as *Deep Nude* or *Undress AI*, effectively recreate the mechanics of voyeurism. Their widespread circulation on encrypted messaging platforms reflects how ordinary users, not professional criminals, perpetuate gendered abuse.

Each form of deepfake harm intersects multiple legal doctrines—defamation, obscenity, extortion, intimidation, and data theft. However, the synthetic nature of the act often makes prosecution uncertain: courts must decide whether falsified imagery can be treated as *real* harm when no physical act occurred.

C. ILLUSTRATIVE GLOBAL AND INDIAN CASES

Empirical evidence underscores the rapid escalation of deepfake offences worldwide:

- **Taylor Swift Deepfake Scandal (2024):** Explicit AI-generated images of the singer flooded X (formerly Twitter) and Discord, receiving millions of views before takedown. Despite community guidelines, the virality of the images exposed weaknesses in automated moderation and delayed takedown responses by major platforms.
- **Indian Case – Bombay High Court (2025):** The court ordered the immediate removal of a synthetic defamatory video portraying actor Akshay Kumar in a religiously sensitive scene. the case established an early judicial precedent recognizing "synthetic defamation" as a cognizable offence, even without a real recording.
- **South Korea (2023 Amendment):** South Korea's National Assembly introduced criminal penalties not only for creating or distributing but even for *viewing or possessing* explicit deepfakes. This stringent model treats participation itself as complicity, reflecting a policy of zero tolerance.

- **China (2024 Regulation on Deep Synthesis Services):** Requires AI content providers to embed digital watermarks disclosing synthetic origin. Failure to do so attracts fines and license suspension.
- **United Kingdom (2023 Online Safety Act):** Empowers the regulator Ofcom to compel immediate takedown of deepfake pornography. It also mandates risk assessments and duty-of-care obligations for tech companies.

These developments collectively signal a global recognition that deepfake misuse is not merely technological misconduct but a violation of human dignity.

D. SOCIO-TECHNOLOGICAL DRIVERS

The proliferation of deepfake crimes is not accidental; it emerges from a confluence of social, technical, and cultural factors:

- **Anonymity and Virality:** Encrypted platforms like Telegram, Discord, and Reddit allow offenders to distribute synthetic material anonymously. Viral reposting ensures that even deleted content resurfaces within hours.
- **Algorithmic Amplification:** Social-media algorithms prioritize engagement, meaning sensational or scandalous content—such as fabricated sexual videos—spreads faster than truth.
- **Forensic Deficiency:** Indian cyber-forensic labs rarely possess AI-authentication tools capable of detecting pixel-level inconsistencies. Manual verification can take weeks, by which time the harm is irreversible.
- **Cross-Border Hosting:** Platforms often store data on servers outside India, making jurisdictional reach difficult. Mutual Legal Assistance Treaties (MLATs) operate slowly, sometimes taking months for a single evidence request.
- **Cultural Patriarchy and Victim-Blaming:** In patriarchal societies, women’s reputations are closely tied to chastity and “modesty.” Victims of deepfake abuse face social ostracization, while perpetrators exploit this stigma to silence them.

- **Economic Incentivization:** Monetized adult-content sites sometimes host synthetic pornography because it attracts clicks, creating a profit motive for exploitation.

Together, these drivers create a perfect storm of impunity, where rapid technological innovation outpaces both law enforcement and social ethics.

E. INTERSECTION WITH TRADITIONAL CRIMES

Deepfake-related offences straddle multiple statutory categories but fail to fit neatly into any.

- **Voyeurism (IPC § 354C):** This provision penalizes capturing or disseminating images of a woman engaged in a private act without consent. It presupposes an *actual recording*; hence, synthetic content falls outside its ambit.
- **Stalking (§ 354D):** Criminalizes repeated electronic communication or surveillance. Deepfake harassment via social media may constitute "virtual stalking," but only if direct targeting is proven.
- **Defamation (§§ 499–500):** Applies when false representations harm reputation. However, proving falsity becomes complex when content appears realistic but is algorithmically generated.
- **Insult to Modesty (§ 509):** Could encompass sexualized deepfakes, but jurisprudence has yet to test this interpretation.
- **IT Act §§ 66E, 67, 67A:** Penalize violation of privacy and transmission of obscene material. Yet the requirement of "publication of real material" limits their reach against synthetic imagery.

Prosecutors thus struggle to establish *actus reus* when the "recording" never occurred, and *mens rea* when algorithms perform automated synthesis. The result is doctrinal friction—law built for tangible evidence confronting virtual harm.

F. SCALE OF THE THREAT

A 2025 Regula Forensics study found that 96% of all deepfakes contained sexualized depictions of women, and 85% targeted identifiable individuals. India ranked among the top five countries in global uploads, particularly through Telegram channels and Reddit communities like *r/AIgirls* or *FakeNudeHub*.

The economic undercurrent is equally concerning websites hosting explicit synthetic media report millions in monthly revenue through ads and subscriptions. The boundary between user-generated content and organized cybercrime is dissolving.

Furthermore, AI-image generators trained on scraped data often reproduce gender and racial biases, perpetuating stereotypical sexualization of women. These systemic biases render women's digital likenesses disproportionately vulnerable to exploitation.

G. PSYCHOLOGICAL, ECONOMIC, AND LEGAL RAMIFICATIONS

Deepfake exploitation inflicts multi-dimensional harm. Victims endure psychological trauma equivalent to sexual assault: loss of control over one's image triggers anxiety, insomnia, and post-traumatic stress. Economically, victims face loss of employment or professional credibility, particularly in public-facing professions such as teaching, law, or entertainment.

From a legal perspective, deepfakes challenge traditional evidentiary paradigms. The authenticity of digital evidence—long considered objective—is now suspect. Courts must navigate questions of admissibility, chain of custody, and expert verification under the Indian Evidence Act, 1872. Without standardized forensic protocols, even genuine evidence risks being dismissed as "fake."

H. PRELIMINARY CONCLUSION

Deepfakes extend age-old gendered offences into a post-truth era, where the body becomes data and violation occurs without touch. The resulting harms—psychological, social, and reputational—mirror those of sexual assault yet remain largely invisible in criminal law.

Unless Indian legislation evolves to define synthetic imagery, strengthen forensic capacity, and expedite takedown mechanisms, the digital environment will continue to normalize gendered violence under the guise of innovation. Technology without accountability, the chapter concludes, is a threat to autonomy itself.

V. PSYCHOLOGICAL AND SOCIAL CONSEQUENCES OF DEEPFAKE VICTIMIZATION

A. DIGITAL SEXUAL TRAUMA

The psychological injury arising from non-consensual deepfakes parallels the trauma of sexual assault even when no physical contact occurs. The violation is informational yet profoundly embodied: the victim's likeness becomes a site of public consumption without consent. Feminist scholars describe this as a *technological rape of the image*—a process through which digital code re-inscribes patriarchal power over women's bodies.

Neuroscientific research confirms that such incidents activate the same neural pathways associated with fear and shame responses in victims of in-person sexual assault. Cortisol and amygdala reactivity spike, producing lasting hyper-arousal. The sense of violation is intensified by the medium itself: because deepfakes appear visually "real," survivors encounter disbelief or mockery when asserting falsity. The trauma thus acquires a doubly hermeneutic character—first through the assault on dignity, and second through the social denial of victimhood.

Psychologists have coined the term "digital rape trauma syndrome" (DRTS) to describe symptoms including intrusive recollections, self-blame, and somatic anxiety triggered by any online interaction. Victims often report compulsive self-monitoring and withdrawal from digital communication, equating online visibility with exposure. The result is a shrinking of digital citizenship for women: fear becomes an invisible barrier to participation in online education, employment, and public debate.

B. POST-TRAUMATIC AND LONG-TERM IMPACTS

Where traditional sexual offences culminate in a discrete event, deepfake abuse is perpetual and renewable. Once uploaded, synthetic material is infinitely replicable;

deletion offers only symbolic closure. The persistence of duplicates on mirror sites and dark-web repositories sustains chronic anxiety—what scholars' term *permanent victimhood.

Clinical evidence from the *Journal of Cyber Psychology & Behaviour (2023)* shows that 72 percent of image-based-abuse victims develop long-term depression and 41 percent disengage from professional or academic life. The repetition of exposure—each time the content resurfaces or is re-shared—re-opens psychological wounds, preventing the natural recovery curve observed in time-bound trauma.

In patriarchal contexts such as India, the burden of proof of innocence falls disproportionately on women. Survivors expend emotional energy persuading employers, family, and community that the video is fabricated. This relentless labour of vindication compounds trauma through what sociologists call secondary victimization—the harm inflicted by disbelief, ridicule, or bureaucratic apathy.

Over time, the continual apprehension of rediscovery can manifest as complex PTSD: dissociation, self-isolation, and distrust of digital technologies. Therapists observe that victims often conflate the Internet itself with threat, leading to digital abstinence. In a society where connectivity equals opportunity, this retreat widens gender gaps in professional advancement and digital literacy.

C. REPUTATIONAL AND PROFESSIONAL FALLOUT

Reputation functions as a form of social capital, particularly in cultures where women's respectability is tied to notions of chastity and modesty. Deepfakes convert this moral economy into a mechanism of control. Employers, universities, and matrimonial prospects frequently react punitively—even when falsity is proven—because the stigma of "immorality" adheres more powerfully than truth.

The economic repercussions are severe. Women in public professions—journalism, politics, education, or entertainment—risk losing livelihoods as organizations fear association with controversy. In the gig-economy era, where reputational metrics determine employment, one viral deepfake can erase years of professional credibility.

Public figures face a paradoxical vulnerability: the same visibility that affords influence magnifies exposure. The media's appetite for scandal exacerbates this harm, algorithms reward outrage, not accuracy. Even when courts or fact-checkers debunk content, the liar's dividend—the residual doubt surrounding authenticity—persists.

From a sociological perspective, deepfakes thus reproduce the logic of the *male gaze* in digital form. They objectify women, re-inscribing patriarchal hierarchies under the guise of technological creativity. The contagion of reputational harm becomes collective: communities internalize narratives of female culpability, reinforcing misogyny and moral policing.

D. UNDER-REPORTING AND LEGAL SILENCE

Despite escalating prevalence, official data remain fragmented. Pew Research (2023) reported that more than 60 percent of Indian women experiencing online sexual abuse abstain from legal recourse due to **fear of stigma and institutional indifference. Cyber-crime portals capture only a small fraction of incidents because complaint categories fail to recognize "AI-generated" or "synthetic" content.

Law-enforcement officers often misclassify deepfakes as simple defamation or photo-editing, betraying limited technological literacy. Victims recount instances where police demanded physical-assault evidence or dismissed synthetic imagery as "not real harm." The absence of confidentiality provisions under current law intensifies reluctance—complainants fear media exposure and cross-examination that replicate the very humiliation they seek to redress.

This systemic silence mirrors broader gender-justice deficits in cyberspace. In contrast, jurisdictions like the United Kingdom and South Korea operate dedicated cyber-sexual-violence units staffed with digital-forensics experts and trauma counsellors. Their success rates demonstrate that reporting rises when victims trust both technological competence and procedural sensitivity.

For India, bridging this gap requires not merely statutory reform but cultural transformation within policing and prosecution—embedding empathy, gender training, and technical proficiency into investigative practice.

E. NEED FOR INTEGRATED SUPPORT SYSTEMS

Effective redress for deepfake-based violence demands a holistic triad of legal, psychological, and technological support.

At present, the *National Cyber Crime Reporting Portal* offers a "Women/Child" category but lacks AI-forensic verification pipelines. A re-engineered system should integrate automated metadata capture, secure evidence storage, and referral to mental-health services. Anonymous reporting options are essential to reduce barriers for marginalized or rural victims.

The creation of Digital Trauma Response Centres (DTRCs) at state level could operationalize this approach. Jointly administered by the *National Commission for Women (NCW)*, *CERT-In*, and accredited NGOs, each centre would provide,

- **Immediate crisis counselling** and psychiatric assessment.
- **Technical forensics** to authenticate synthetic content for evidentiary use.
- **Legal navigation assistance**, including drafting FIRs and coordinating with cyber-cells.
- **Data-erasure liaison units** engaging platforms for expedited takedown.

Parallelly, India could establish a Digital Violence Relief Fund, financed through technology-sector levies, to compensate victims for reputational and economic losses. Education plays a preventive role: embedding digital-ethics curricula in schools and universities can cultivate critical literacy around consent and image sharing. Public-awareness campaigns—akin to the U.K.'s *"Revenge Porn Helpline"* initiative—should communicate the message that sharing or viewing deepfakes constitutes participation in sexual violence.

Finally, collaboration with private industry is indispensable. Social-media companies and hosting providers must be bound by co-regulatory codes of practice requiring 24-hour takedown, transparency reports, and algorithmic-detection tools. Victim support, therefore, is not solely humanitarian; it is structural governance for a safe digital public sphere.

F. INTERSECTIONALITY AND SOCIAL JUSTICE

Deepfake victimization intersects with India's hierarchies of caste, class, region, and sexuality. Dalit, Adivasi, and minority women, already facing offline discrimination, suffer compounded marginalization online. The weaponization of deepfakes against outspoken Dalit women journalists or activists illustrates how digital patriarchy fuses with caste oppression.

LGBTQ+ persons encounter parallel risks. Non-binary and transgender individuals, whose images are often hyper-sexualized, report synthetic pornography used to mock or "out" them. Because Indian anti-discrimination law provides limited recourse for gender diversity, such victims remain invisible to both police and policy.

In rural India, infrastructural inequities—shared devices, limited privacy, male-dominated Internet cafés—magnify exposure. A single leaked deepfake can precipitate honour-based violence or forced migration. Addressing these realities requires embedding intersectional feminism and data-justice in AI regulation: laws must recognize that technology amplifies existing social asymmetries.

Global south perspectives emphasize community-based healing. Women's collectives in Kenya, Brazil, and the Philippines have pioneered peer-support circles that combine digital literacy with trauma counselling. Similar grassroots models could complement India's institutional response, ensuring accessibility beyond metropolitan centres.

G. THE BROADER SOCIAL IMPLICATIONS

Beyond individual harm, deepfakes corrode the epistemic foundations of society. When authenticity itself is suspect, women's testimonies lose credibility—a phenomenon termed *gendered epistemic injustice.* Deepfakes thus not only silence victims but distort collective truth, undermining journalism, activism, and democratic discourse.

Unchecked proliferation normalizes voyeurism as entertainment, desensitizing audiences to consent. The spectacle of female degradation becomes a recurring algorithmic pattern, reinforcing what media theorist Laura Mulvey termed the "pleasure of looking" as domination.

In the long run, the erosion of trust in images threatens judicial and evidentiary integrity: video evidence once considered irrefutable can now be challenged as synthetic, complicating prosecutions in every domain from domestic violence to corruption.

VI. LEGAL FRAMEWORK IN INDIA

A. EXISTING STATUTORY ARCHITECTURE

The Indian Penal Code (IPC) and the Information Technology Act (IT Act) constitute the backbone of India's criminal accountability regime in cyberspace. Yet both were conceived long before the rise of generative AI and synthetic media. Their application to deepfakes remains largely interpretive, forcing courts and enforcement agencies to retrofit nineteenth and early twenty-first century provisions to twenty-first century harms.

- **IPC § 354C (Voyeurism):** This provision prohibits the capture, publication, or transmission of images of a woman engaged in a private act without consent. Deepfakes, however, *replicate* rather than *record* such acts. Because the statute presumes the existence of a real image, its language fails to encompass algorithmic fabrication. The offence collapses conceptually when no actual private act took place.

- **IPC § 354D (Stalking):** Covers repeated contact or monitoring through electronic communication. The creation or use of deepfakes to intimidate or harass women could fall under this provision by analogy. Yet its evidentiary foundation—proof of "contact" or "monitoring"—is weak when the perpetrator hides behind pseudonymous online profiles or automated accounts.

- **IPC §§ 499–500 (Defamation):** Criminalize false statements intended to harm reputation. Courts must now determine whether an AI-generated video qualifies as a "representation" under § 499. The essence of defamation lies in the publication of falsehoods; synthetic videos satisfy this requirement, but judges must expand interpretation to include *non-verbal, visual misrepresentations* of fact.

- **IPC § 509:** Penalizes words, gestures, or acts intended to insult a woman's modesty. Deepfake dissemination clearly fits the mischief targeted by this section, though its moralistic language— "modesty"—reflects Victorian-era sensibilities. A gender-neutral, dignity-based reframing is overdue.
- **IT Act § 66E:** Addresses privacy violations through capturing, publishing, or transmitting images of private parts without consent. Like § 354C, it assumes the existence of real footage; deepfakes generate *illusory nudity* that nevertheless violates autonomy.
- **IT Act §§ 67 and 67A:** Prohibit publishing or transmitting obscene or sexually explicit material electronically. These provisions are technology-agnostic and thus applicable to synthetic content, but they fail to differentiate consensual adult expression from non-consensual manipulation.

Collectively, these provisions demonstrate partial coverage. The law punishes *what deepfakes imitate*—voyeurism, obscenity, defamation—but not *the act of fabrication itself.* The result is doctrinal slippage: a victim must prove an offence designed for tangible imagery to an intangible, algorithmic abuse. Moreover, mens rea (criminal intent) becomes elusive. When an offender merely "prompts" an AI model or shares pre-existing synthetic content, establishing intent and causation within the chain of creation, dissemination, and amplification requires new evidentiary frameworks.

B. CONSTITUTIONAL FOUNDATION

The Indian Constitution's moral core—Article 21—anchors the right to life and personal liberty, expansively interpreted by the judiciary to encompass privacy, dignity, and autonomy. In *Justice K.S. Puttaswamy v. Union of India* (2017), the Supreme Court declared that privacy includes control over personal information and identity.

Non-consensual deepfakes obliterate that control, amounting to a constitutional tort against informational privacy.

The dissemination of deepfakes implicates not only Article 21 but also Articles 14 and 19:

- **Article 14 (Equality before Law)** demands gender-sensitive enforcement. When deepfake abuse disproportionately targets women, state inaction perpetuates indirect discrimination.
- **Article 19(1)(a) (Freedom of Speech)** protects expression, but Article 19(2) allows restrictions for decency, morality, and defamation. Regulation of deepfake pornography thus falls squarely within permissible limits.

Jurisprudentially, the right to be forgotten, emerging from *Puttaswamy* and subsequent Delhi High Court rulings, strengthens victims' entitlement to digital erasure. Deepfakes make this right urgent content that never existed must still be forgotten in law.

A constitutional deepfake regime must therefore reconcile two imperatives: safeguard women's dignity without overbroad censorship of satire or artistic expression. This balance mirrors the *Shreya Singhal* doctrine—laws must be precise, proportionate, and narrowly tailored to avoid chilling speech.

C. JUDICIAL RESPONSES AND DOCTRINAL DEVELOPMENTS

Although sparse, judicial recognition of synthetic-media harm is growing. The Bombay High Court (2025) ordered takedown of an AI-generated defamatory video depicting actor Akshay Kumar in a distorted religious role, observing that "synthetic realities can inflict genuine reputational injury." This acknowledgment—though in a celebrity context—marks India's entry into AI jurisprudence.

Earlier, in *Shreya Singhal v. Union of India* (2015), the Supreme Court struck down § 66A IT Act for vagueness, emphasizing that speech offences must be narrowly defined. The same principle applies to deepfake regulation: definitions of "synthetic media" must be technologically precise to survive constitutional scrutiny.

In *Faheema Shirin v. State of Kerala* (2019), the Kerala High Court recognized Internet access as integral to the right to education and expression. This case indirectly underlines the need for safe digital spaces; rights to online participation lose meaning when women are deterred by fear of violation.

Judicial trends indicate a nascent but progressive approach—courts willing to extend privacy and dignity doctrines to AI contexts. The challenge lies not in normative recognition but in procedural realization: without clear evidentiary rules or forensic competence, these ideals remain unenforced.

D. PRACTICAL ENFORCEMENT CHALLENGES

- **Absence of Definition:** Neither IPC nor IT Act defines "deepfake," leaving ambiguity over scope. Police reports often misclassify incidents under obscenity or defamation, missing the synthetic element entirely.
- **Evidentiary Complexity:** Authenticating manipulation requires AI-forensic experts and metadata analysis—resources unavailable in most district courts. The Indian Evidence Act, 1872, does not contemplate algorithmic falsification, creating admissibility confusion.
- **Platform Liability:** The 2021 Intermediary Guidelines mandate takedown on notice but lack statutory deadlines or tiered penalties. Major platforms remove content reactively, often after virality inflicts irreversible harm.
- **Cross-Border Data Barriers:** Servers hosting deepfakes frequently lie outside India. Mutual Legal Assistance Treaties (MLATs) take months, rendering digital evidence obsolete.
- **Victim Retraumatization:** The slow judicial process forces victims to repeatedly testify about intimate content, compounding trauma. The absence of in-camera proceedings or anonymity provisions deters complaints.
- **Technical Illiteracy in Law Enforcement:** Police manuals seldom address AI-generated media. Investigators lack the capacity to distinguish real from synthetic, leading to dismissals or improper charge-sheeting.

Collectively, these obstacles reveal not a lack of law but a lack of literacy and infrastructure. Without institutional investment in training and technology, legislative reform alone will remain cosmetic.

E. GOVERNMENT AND TECHNOLOGICAL INITIATIVES

Recognizing AI's dual-use potential, the Ministry of Electronics, and Information Technology (MeitY) issued *AI Ethics Guidelines* (2024), recommending algorithmic transparency, watermarking, and consent protocols. These remain advisory but signal policy intent. The Zero Defend Security Initiative (ZDSI)—a public–private partnership launched in 2025—developed *Vastav AI*, a detection tool that analyses pixel anomalies to verify image authenticity.

While promising, its legal utility depends on certification under the Indian Evidence Act. Courts currently accept such reports only as *expert opinion*, not conclusive proof. The Intermediary Rules (2021) introduced traceability obligations for messaging platforms but sparked privacy concerns. A nuanced amendment could reconcile traceability with encryption by allowing metadata verification rather than content decryption.

Further, the Digital India Cybersecurity Agency (DICA) is piloting a *National AI Forensics Grid* linking police cyber-cells, CERT-In, and judicial repositories. If institutionalized, this network could standardize evidence collection and prevent tampering—a crucial step toward credible prosecution. However, these efforts remain fragmented. India lacks an integrated *Deepfake Response Framework* akin to South Korea's Cyber Sexual Violence Bureau or the U.K.'s Ofcom-led regime.

F. NEED FOR LEGISLATIVE MODERNIZATION

India's legal infrastructure must undergo a structural overhaul. The IPC's colonial origins (1860) and the IT Act's millennial vintage (2000) leave them unfit for the generative-AI epoch. Incremental patchwork amendments cannot bridge the doctrinal gap; a comprehensive statute dedicated to synthetic-media offences is imperative.

Key legislative reforms should include:

- **Definition and Classification:**

  - Define "synthetic media," "deepfake," and "non-consensual intimate synthetic content."
  - Distinguish between harmful (non-consensual or deceptive) and benign (satirical, educational) uses to preserve freedom of expression.

- **Criminal Provisions:**
  - Penalize creation, distribution, or threat of non-consensual deepfakes with graded sentences—up to five years for creation/dissemination, enhanced penalties for extortion or recidivism.
  - Include offence of synthetic identity impersonation covering fraudulent AI voice or likeness use.
- **Civil and Administrative Remedies:**
  - Introduce victim-friendly injunctions for immediate takedown and restraining orders.
  - Establish a statutory "Digital Rights Tribunal" with powers equivalent to consumer forums for expeditious adjudication.
- **Evidentiary Reform:**
  - Amend the Indian Evidence Act to recognize AI-forensic certification and metadata logs as primary evidence.
  - Mandate preservation of digital trails by intermediaries for 180 days post-takedown to aid prosecution.
- **Platform Accountability:**
  - Replace the notice-based model with a time-bound obligation—verified harmful deepfakes must be removed within 24–48 hours.
  - Introduce financial penalties proportional to user base, incentivizing proactive moderation.

- **Victim Protection:**
  - Guarantee anonymity in proceedings, akin to rape survivors under § 327 CrPC.
  - Provide government-funded legal aid and trauma counselling.
- **Periodic Review:**
  - Mandate legislative reassessment every three years to adapt to evolving AI capabilities—a sunset clause ensuring dynamism.

Such modernization would align India with international best practices—U.S. federal acts, the U.K. Online Safety Act (2023), and the forthcoming EU AI Act (2025). Importantly, it would shift India's posture from reactive to preventive governance, embedding accountability across the AI ecosystem.

VII. COMPARATIVE LEGAL FRAMEWORK

A. UNITED STATES — FEDERAL AND STATE RESPONSES

The United States represents a multi-tiered governance model in its handling of synthetic media, balancing the constitutional sanctity of free speech under the First Amendment with the need to curb digital impersonation and non-consensual sexual deepfakes. The TAKE IT DOWN Act (2025) is the country's first federal law explicitly criminalizing the knowing creation, sharing, or threat to share AI-generated intimate imagery without consent. The statute recognizes "digitally fabricated likeness" as an independent legal category, acknowledging that harm arises from the *representation* of identity rather than the veracity of the image. The Act's practical mechanisms—mandatory 48-hour takedown, survivor hotline, and FTC-enforced penalties—represent a strong procedural framework for rapid intervention. It also establishes an online portal where victims can submit takedown requests directly to platforms, reflecting a bureaucratic model of victim support and automation.

Complementing this, the ELVIS Act (2024) (Ensuring Likeness, Image, and Voice Security) extends intellectual-property-style protection to a person's identity, voice, and likeness, addressing both entertainment and non-consensual domains. For instance, the unauthorized use of a deceased singer's voice for a commercial deepfake song would now attract civil penalties. State-level enactments such as California Penal Code §653.35, Texas Penal Code §21.16, and Virginia's HB2678 preceded federal harmonization, criminalizing the creation or dissemination of sexual deepfakes when identifiable persons are depicted.

However, federalism also introduces inconsistency. Jurisdictional fragmentation across 50 states often results in uneven enforcement. A video generated in Texas but hosted on a foreign website may escape immediate prosecution due to transnational complexities. Moreover, *Section 230* of the Communications Decency Act (CDA) still shields intermediaries from liability for user-generated content, creating friction between platform immunity and victim redress. Nonetheless, the U.S. framework excels in procedural clarity, integrating civil, criminal, and regulatory dimensions. The focus on time-bound takedown, mental-health support, and civil damages offers a rights-based yet pragmatic model India can adapt within its own federal structure.

B. UNITED KINGDOM — ONLINE SAFETY ACT (2023)

The U.K. Online Safety Act (2023) represents a holistic attempt to redefine platform accountability in the age of algorithmic harm. It imposes a statutory "duty of care" on service providers to identify, prevent, and promptly remove illegal content—including AI-manipulated sexual or defamatory material. Enforcement lies with Ofcom, which wields the authority to impose fines amounting to up to 10% of a company's global turnover for non-compliance. The Act complements amendments to the Sexual Offences Act (2003) that criminalize both the creation and the threat to share non-consensual sexual deepfakes.

Beyond punishment, the Act emphasizes design-level safety. The principle of *"safety-by-design"* mandates that platforms incorporate risk assessments, verification mechanisms, and automated filters capable of detecting synthetic nudity before upload. This preventive model shifts the onus from the victim—who traditionally bore the burden of reporting—to the platform, thereby operationalizing restorative justice. Another hallmark is its "transparency reporting" obligation: companies must periodically disclose how their algorithms mitigate the spread of manipulated media.

Critics argue that these obligations may burden start-ups and curtail creative freedom. However, the U.K. model offers a powerful lesson in proactive governance—that harm prevention is as crucial as post-facto redress. For India, embedding a similar duty of care into the IT Rules or a prospective Deepfake Regulation Act could recalibrate the balance between innovation and accountability.

C. EUROPEAN UNION — DIGITAL SERVICES ACT AND AI ACT

The European Union adopts a rights-driven, ex ante regulatory philosophy anchored in transparency and human dignity. The Digital Services Act (DSA) (2022/2065) establishes layered obligations for digital intermediaries, mandating traceability of content origin and algorithmic accountability. It classifies platforms into tiers— "very large online platforms" (VLOPs), "hosting services," and "intermediary services"—assigning each specific compliance burdens. VLOPs must maintain a content-moderation infrastructure and appoint an independent compliance officer answerable to the European Commission.

Complementing the DSA is the AI Act (2025), which marks the world's first horizontal AI-specific legislation. Article 52 obliges deepfake creators to disclose synthetic origin through visible watermarking or metadata embedding, effectively codifying transparency as a technological standard. Generative AI models producing realistic human likenesses are designated "high-risk AI systems," triggering strict pre-market testing, data governance audits, and post-market surveillance.

The EU's method is not punitive but preventive seeking to engineer trust into AI ecosystems. For India, which often struggles with post-incident justice, adopting similar compliance obligations could shift the focus toward risk mitigation. India's MeitY or CERT-In could emulate the EU's risk-tier model, requiring generative AI platforms to register, label outputs, and provide APIs for authenticity verification. This approach would harmonize India's constitutional emphasis on privacy with international digital due process norms.

D. DENMARK — PERSONALITY RIGHTS AS PROPERTY

Denmark's 2024 Copyright and Personality Rights Amendment Bill offers a novel perspective by framing digital likeness as intellectual property. Unlike moral-based criminalization, it treats unauthorized AI-generated usage of one's face, voice, or mannerisms as a form of identity theft in property law. This conceptual shift enables victims to seek injunctions and damages under civil law rather than criminal prosecution, expediting redress. For instance, a Danish influencer could sue an AI developer for monetizing her likeness without consent even if no sexual element is involved.

This property-rights lens reclaims personal agency by granting individuals control over their digital selves. For India, where Article 21 enshrines the right to dignity and privacy, such reasoning could ground a jurisprudence of "image rights" akin to the *right of publicity* in U.S. law. Recognizing image rights as property would allow Indian victims to claim civil compensation alongside criminal remedies under the IPC and IT Act.

However, critics note the risk of commodifying identity—transforming personhood into a transactional asset. A balanced approach, therefore, would treat personality rights as both moral and proprietary, aligning with India's mixed constitutional ethos that values dignity while enabling enforcement through civil claims.

E. SOUTH KOREA — ZERO-TOLERANCE CRIMINALIZATION

South Korea presents perhaps the most stringent model globally. The 2023 amendment to the Sexual Violence Punishment Act criminalizes not only the creation or dissemination but even possession or viewing of sexually explicit deepfakes. Penalties include up to three years' imprisonment and substantial fines. Complementing this, the Cyber-Sexual Violence Centre (CSVC) operates as a centralized government agency coordinating forensic authentication, legal aid, and psychological counselling for victims. The CSVC's Rapid Deletion Protocol (RDP) facilitates real-time takedowns through direct collaboration with domestic ISPs and global tech firms.

This administrative proactivity illustrates that effective deterrence depends as much on institutional infrastructure as on legislative precision. The South Korean approach reframes deepfake crimes as a public morality issue, akin to sexual assault, rather than a mere privacy violation. For India, where enforcement delays often nullify legislative intent, replicating South Korea's centralized forensic and victim-aid ecosystem could bridge the gap between law and implementation. However, India must tailor such a model to its federal and linguistic diversity, ensuring localized support through state-level cyber cells linked to a national coordination hub.

F. COMPARATIVE INSIGHTS FOR INDIA

From this comparative panorama, five key takeaways emerge for India's reform agenda:

- **Definition and Scope:** Every advanced jurisdiction—from the U.S. to Denmark—has introduced statutory definitions of "synthetic media," "deepfake," or "digitally fabricated likeness." India must legislate similar clarity to avoid interpretative confusion and evidentiary delay. A clear definition should encompass *creation, dissemination, and threat of dissemination* to pre-empt extortion.
- **Platform Accountability and Timelines:** The 24–48-hour takedown deadlines observed in the U.S. and U.K. balance urgency with due process. India's IT Rules (2021) could be amended to mandate *"trusted flagger"* systems, empowering verified victims and NGOs to expedite removals without lengthy bureaucratic approval.
- **Victim-Centric Remedies:** Psychological counselling, anonymity during trial, and compensation funds—standard features in Western models—should be embedded within India's cyber-crime infrastructure. Legal aid clinics attached to law schools or NCW partnerships could handle representation and counselling.
- **Forensic Infrastructure:** Establishing Digital Forensic Verification Units in every state, integrated with CERT-In's national grid, would ensure authenticity verification for AI-generated media. Such units could be

accredited to testify under the Indian Evidence Act, bridging the current proof gap.

- **Transparency and Innovation Safeguards:** Mandatory watermarking and cryptographic provenance (like Content Authenticity Initiative protocols) would deter misuse while protecting artistic freedom. India must also provide safe-harbour exemptions for parody, education, and journalistic use to preserve free expression.
- **Institutional Synergy:** A cross-ministerial task force combining MeitY, the Ministry of Women and Child Development, and the Ministry of Law could harmonize legislative, technological, and welfare dimensions of deepfake governance. Lessons from Ofcom and FTC show that *multi-agency coordination* ensures both enforcement and education.
- **International Cooperation:** India should negotiate bilateral data-sharing MOUs with jurisdictions hosting major social media servers to accelerate evidence retrieval. Participation in multilateral AI ethics coalitions, such as the Global Partnership on AI (GPAI), could also strengthen India's regulatory credibility.

VIII. FINDINGS AND RECOMMENDATIONS

A decade into the AI revolution, deepfakes mark one of the most disruptive intersections between technology, gender, and legal governance. The findings of this research converge on several interrelated themes that illuminate both the depth of the crisis and the opportunity for reform.

- **Gendered Pattern of Abuse:** Empirical data consistently confirm that over 96 percent of detected deepfakes depict women in sexualized contexts. Algorithms trained on biased datasets replicate patriarchal stereotypes, turning female bodies into digital commodities. The "male gaze," once confined to cinema and advertising, is now embedded in machine learning itself. Such algorithmic misogyny perpetuates historical gender hierarchies in a new, technological form of subjugation.

- **Psychological Harm Comparable to Physical Assault:** Clinical psychology increasingly recognizes that victims of image-based sexual abuse experience trauma mirroring that of physical assault survivors. The digital permanence of deepfakes exacerbates anxiety, shame, and isolation, leading to depression and suicidal ideation. These harms transcend the traditional distinction between tangible and intangible injury, demanding that criminal law expand its conception of harm beyond corporeality.

- **Fragmented Legal Coverage:** Existing Indian laws—IPC §§ 354C, 354D, 499–500, 509, and IT Act §§ 66E, 67, 67A—address partial aspects such as voyeurism, defamation, or obscenity but fail to capture the essence of synthetic fabrication. The absence of a statutory definition of "deepfake" or "synthetic media" creates interpretive uncertainty, forcing courts and police to stretch analogy provisions unsuited for digital realities.

- **Institutional Deficits:** Cybercrime units remain under-equipped, lacking forensic tools and AI expertise. Coordination among agencies like MeitY, CERT-In, and state police is inconsistent, leading to redundant or delayed investigations. Without dedicated AI-forensic laboratories, authenticity verification remains dependent on third-party vendors, compromising evidentiary integrity.

- **Judicial Awareness Emergent but Limited:** While the Bombay High Court's 2025 decision recognizing "synthetic defamation" represents progress, India's judiciary lacks uniform protocols for adjudicating AI-generated evidence. The need for standardized evidentiary principles—similar to the Daubert test in the U.S.—is acute.

- **Global Momentum Outpaces India:** The U.S., U.K., E.U., and South Korea have enacted targeted statutes and institutional frameworks addressing AI-enabled sexual exploitation. India's reactive case-by-case approach contrasts sharply with these proactive, exosystemic responses.

- **Ethical and Corporate Vacuum:** Despite the proliferation of corporate "AI ethics charters," compliance remains voluntary. Social media platforms continue to profit from engagement-driven algorithms that amplify harmful content. Without transparency and audit requirements, the commercial architecture of virality remains fundamentally misaligned with women's safety.
- **Socio-Cultural Blind Spots:** Patriarchal stigma and victim-blaming persist. Public discourse often questions victims' morality rather than the perpetrator's intent. This cultural inertia undermines deterrence even when legal mechanisms exist.

A. DOCTRINAL INSIGHTS

Doctrinally, deepfake abuse exposes the elasticity and the limits of Indian criminal law. The IPC's reliance on physical "acts" and "representations" presumes a material reality. Deepfakes, however, are algorithmic simulations—representations without referents. This epistemic shift demands a re-imagination of *actus reus* and *mens rea.*

The *Puttaswamy* judgment (2017) provides a constitutional anchor by linking privacy with dignity and informational autonomy. By extending this reasoning, synthetic violations of identity qualify as breaches of bodily integrity in a digital sense. The constitutional framework, therefore permits, and arguably mandates, legislative innovation to fill this void.

Further, criminal law must integrate the principle of technological neutrality—that the medium of harm should not dictate the validity of protection. Whether abuse occurs via camera or code, the injury to dignity is equivalent.

B. POLICY RATIONALE FOR REFORM

- **Normative Justice:** Deepfake exploitation violates equality (Articles 14 & 15) and dignity (Article 21), necessitating statutory reinforcement of gender justice.
- **Preventive Deterrence:** Explicit criminalization and time-bound procedures deter potential offenders and close enforcement gaps.

- **Restorative Justice:** A victim-centred approach ensures that legal recognition translates into emotional and social rehabilitation, aligning criminal law with restorative principles.
- **International Commitments:** India, as a signatory to CEDAW (1979) and participant in the Budapest Convention on Cybercrime, bears an obligation to curb technology-mediated sexual violence.
- **Technological Sovereignty:** Developing domestic capacity for AI forensics and content moderation strengthens India's autonomy in digital governance.

C. RECOMMENDED LEGAL REFORMS

- **Enactment of a Deepfake Regulation Act**
    - Define *deepfake*, *synthetic media*, and *non-consensual intimate imagery*.
    - Criminalize creation, possession, dissemination, or threat thereof without consent.
    - Establish graded punishments: up to five years for a first offence, ten for aggravated or commercial exploitation.
    - Include civil remedies—injunctions, damages, and public apologies.
    - Recognize *psychological injury* as compensable harm.
- **Time-Bound Takedown and Right to Erasure:**
    - Mandate 24–48-hour takedown for verified deepfake complaints.
    - Introduce a Right to Digital Oblivion, ensuring permanent removal from search engines and archives.
    - Establish a judicially supervised Content Removal Tribunal for expedited orders.
- **Platform Accountability and Transparency:**

  - Amend §79 IT Act: conditional immunity only for platforms that implement AI-content audits and watermark verification.
  - Mandate public AI Transparency Reports detailing takedowns, algorithmic moderation, and bias assessments.
  - Empower MeitY to issue compliance directions with enforceable penalties.

- **Forensic and Institutional Capacity Building:**
  - Create Digital Forensics Units (DFUs) in every state, equipped with certified AI detection tools.
  - Integrate these DFUs with CERT-In for real-time data exchange.
  - Partner with IITs and law universities to establish AI Forensics Centres of Excellence.

- **Victim-Centric Mechanisms:**
  - Implement anonymous complaint filing and in-camera proceedings.
  - Introduce the Digital Violence Relief Fund for rehabilitation and counselling.
  - Ensure free legal aid and access to trauma-informed therapy.

- **Judicial and Administrative Training:**
  - Include AI Evidence Interpretation modules at the National Judicial Academy.
  - Develop a judicial Benchbook on Synthetic Media.
  - Sensitize police officers through periodic workshops led by forensic experts.

- **Cross-Border Cooperation:**
  - Update MLATs to include rapid AI-content exchange protocols.

    - Collaborate with Interpol's Cybercrime Directorate and the EUROPOL Innovation Hub.
    - Create a shared Global Hash Database for illegal synthetic imagery.
- **Periodic Review and Sunset Clauses:**
    - The Deepfake Act should mandate five-year reviews and public consultations to adapt to evolving technology.

D. ETHICAL AND SOCIETAL RECOMMENDATIONS

- **Digital Literacy and Education:**
    - Incorporate AI ethics, privacy, and consent modules into school and university curricula.
    - Encourage peer-to-peer awareness initiatives, especially for adolescents and rural internet users.
- **Corporate Accountability:**
    - Require tech companies to adopt Responsible AI Charters verified by independent audits.
    - Impose penalties for algorithmic negligence or failure to prevent foreseeable harm.
- **Public Awareness and Media Campaigns:**
    - National outreach under the slogan "Recognize, Report, Remove."
    - Collaborate with influencers and NGOs to destigmatize reporting.
- **Research and Innovation Funding:**
    - Establish a National AI Integrity Fund supporting open-source detection technologies.
    - Incentivize Indian start-ups to develop watermarking and provenance-tracking tools.
- **Media Ethics and Regulation:**

  - Amend Press Council norms to ban broadcast of synthetic sexual content.
  - Encourage responsible reporting—focus on systemic issues, not voyeuristic details.

- **Community Resilience:**
  - Promote survivor support networks and mentorship circles.
  - Encourage digital-rights NGOs to provide rapid legal counselling and media literacy training.
- **Gender-Sensitive Data Policy:**
  - Mandate gender-impact assessments for all AI datasets and applications.
  - Protect biometric and facial data under a strengthened Data Protection Act.

IX. CONCLUSION

Deepfake AI embodies the paradox of innovation—it enables creativity while simultaneously enabling harm. The same algorithms that entertain and inform can be weaponized to distort identity, violate consent, and erode trust. For women, deepfakes represent a new form of digital sexual violence: a violation of dignity through the manipulation of likeness and autonomy. By converting a person's image into an object of public consumption, deepfakes redefine violence in informational terms.

India's criminal-law framework, rooted in the Indian Penal Code (1860) and IT Act (2000), was never designed for crimes where reality itself can be fabricated. Existing provisions on obscenity, defamation, and modesty rely on physical acts and tangible evidence. Deepfakes, however, produce harm algorithmically, without cameras or witnesses. This doctrinal gap leaves victims without a timely remedy and blurs the line between truth and fabrication.

A. DEEPFAKES AS GENDERED VIOLENCE

Non-consensual synthetic imagery constitutes a form of gender-based violence. It exploits anonymity and virality, inflicting lasting psychological trauma. Victims often experience identity fragmentation and social exclusion. Such violations strike at the heart of constitutional rights to dignity, privacy, and equality under Articles 14, 15, and 21. By silencing women through humiliation, deepfakes also curtail freedom of expression under Article 19(1)(a).

B. RETHINKING CRIMINAL LAW

Criminal jurisprudence must evolve from act-based to harm-based accountability, recognizing that injury can be informational as well as physical. Liability must extend across the ecosystem—creators, distributors, platforms, and data brokers—rather than focusing solely on individual offenders. Governance must also become preventive: watermarking, provenance tracking, and algorithmic audits should be mandatory safeguards built into AI architecture.

C. CONSTITUTIONAL AND ETHICAL DIMENSIONS

The Supreme Court's decision in *K.S. Puttaswamy v. Union of India* (2017) affirmed privacy as integral to dignity. Deepfakes violate both informational autonomy and bodily integrity. Regulation must therefore be grounded not in censorship but in constitutional protection. Ethically, AI developers and platforms must adopt a "duty of non-maleficence"—to prevent foreseeable harm—through transparency, impact assessments, and ethical-by-design innovation.

D. SOCIETAL AND PSYCHOLOGICAL CONTEXT

Legal reform alone cannot restore the agency of victims. Survivors need counselling, anonymity protections, and mechanisms to reclaim digital identity. Public awareness must frame sharing deepfakes as complicity in violence, not entertainment. Educational curricula should promote digital ethics and empathy, while media outlets must adopt responsible reporting standards to avoid re-victimization.

E. GLOBAL LESSONS FOR INDIA

Global examples reveal a dual approach: criminalization and systemic accountability. The U.S. TAKE IT DOWN Act (2025) mandates 48-hour takedowns; the U.K.'s Online Safety Act (2023) enforces platform "duty of care"; South Korea's zero-tolerance policy combines criminal penalties with counselling and rapid deletion.

India can adapt these lessons by enacting a Deepfake Prevention and Accountability Act, establishing:

- National Deepfake Response Cell within CERT-In for rapid verification and takedown,
- AI forensic protocols admissible under the Evidence Act, and
- a victim-compensation framework similar to the Nirbhaya Fund.

F. THE MORAL IMPERATIVE

At its core, regulating deepfakes is a fight for epistemic integrity—the ability to trust evidence in a digital society. If images can lie perfectly, public trust in journalism, justice, and memory erodes. Thus, protecting women from synthetic exploitation is inseparable from protecting truth itself.

India, aspiring to lead in AI innovation, must pair technological progress with moral responsibility. Digital sovereignty must mean not only innovation but integrity.

G. FINAL REFLECTIONS

The weaponization of AI against women reflects deep-rooted patriarchy expressed through code. True reform, therefore, requires both legislative precision and cultural transformation.

To ensure women's autonomy in the digital age, India must:

- Enact explicit deepfake legislation.
- Educate citizens in digital ethics and consent.
- Embed authenticity safeguards into AI systems.
- Provide trauma-informed victim support.

The fight against deepfakes is, ultimately, a fight for human dignity. It challenges lawmakers to modernize justice, technologists to humanize design, and citizens to recognize that empathy—not voyeurism—is the measure of progress. Protecting truth and autonomy in the age of AI is not merely a legal necessity but a civilizational duty.

X. REFERENCES

- Justice K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1 (India).
- Shreya Singhal v. Union of India, (2015) 5 SCC 1 (India).
- Indian Penal Code, No. 45 of 1860 (India).
- Information Technology Act, No. 21 of 2000 (India).
- Ministry of Electronics & Information Technology (India), *AI Ethics and Watermarking Guidelines* (2024).
- Bombay High Court, Order in *Akshay Kumar v. Unknown*, (2025).
- Regula Forensics, *Deepfake Laws: Global Regulations in 2025* (2025).
- Reuters, "South Korea to Criminalise Watching or Possessing Sexually Explicit Deepfakes" (2024).
- U.S. Congress, *TAKE IT DOWN Act*, S. 146, 119th Cong. (2025).
- Tennessee General Assembly, *Ensuring Likeness Image and Voice Security (ELVIS) Act* (2024).
- Online Safety Act 2023 (U.K.).
- European Union, *Digital Services Act*, Regulation (EU) 2022/2065.
- European Commission, *Artificial Intelligence Act* (Proposal 2025).
- Denmark, *Copyright and Personality Rights Amendment Bill* (2024).
- Cyberpsychology & Behaviour Journal, "Perpetual Victimization and Online Trauma" (2023).
- Pew Research Center, *Online Harassment and Gender Inequality Report* (2023).

- Globsec, *Regulating Deepfakes: Comparative Global Approaches* (2024).
- Zero Defend Security Initiative, *Vastav AI Detection Tool White Paper* (2025).
- Will Crozier, "Deepfake Detection Technologies: A Policy Gap Analysis," *Journal of Cyber Law* (2024).
- Will Hawkins, Chris Russell & Brent Mittelstadt, "Deepfakes on Demand: The Rise of Accessible Non-Consensual Generators," (2025).
- European Commission, *Ethical Guidelines for Trustworthy AI* (2023).
- National Crime Records Bureau (India), *Cybercrime Statistics 2024* (2025).
- Rouse, "AI-Generated Deepfakes: What Does the Law Say?" (2024).
- UN Women, *Technology-Facilitated Gender Based Violence Report* (2024).
- Center for Internet and Society (India), *Policy Paper on AI Governance and Privacy* (2025).