

ARTICLE

AI-Generated Deepfakes as Structural Gender Discrimination Rethinking Constitutional Rights in India



Shelli Siddhu* & Prof.(Dr.)Vaibhav Goel Bhartiya**

Abstract

Artificial Intelligence (AI) has introduced new forms of harm that challenge existing legal frameworks, particularly in the context of gender-based digital violence. Among these, AI-generated deepfake pornography represents a significant threat to women's privacy, dignity, and equality. This article contributes to existing scholarship by developing an integrated constitutional and feminist analytical framework that reconceptualises deepfake harm as a form of structural gender discrimination rather than merely a privacy violation. The study adopts a comparative doctrinal methodology, analysing constitutional principles and statutory frameworks in India alongside the European Union's AI Act. Drawing on constitutional jurisprudence, including *Justice K.S. Puttaswamy v Union of India* and *Shreya Singhal v Union of India*, the paper examines how dignity, proportionality, and free speech doctrines apply to AI-enabled harms. The novelty of the paper lies in its shift from a state-centric to a platform-centric constitutional framework, highlighting the growing role of private digital platforms in shaping rights violations. The article identifies key regulatory gaps in Indian law, particularly in relation to synthetic media and platform accountability, and proposes a gender-sensitive governance model integrating algorithmic accountability, platform liability, and survivor-centred remedies.

Keywords:- Artificial Intelligence, Deepfake, Gender-based violence, Constitutional law, Privacy, India

I. INTRODUCTION

Artificial Intelligence is increasingly embedded in social, economic, and governance systems, reshaping how individuals interact, communicate, and exercise their rights. While AI offers transformative potential, it has also enabled new forms of harm that existing legal frameworks struggle to regulate.¹ Among these, AI-generated deepfake pornography has emerged as one of the most severe threats to women's privacy, dignity, and equality.²

Deepfake technologies allow the creation of hyper-realistic synthetic media, often without consent, disproportionately targeting women. Unlike traditional cyber harms, deepfake abuse combines elements of identity theft, sexual exploitation, and reputational harm, producing long-lasting

* Research Scholar, Faculty of Law, Swami Vivekanand Subharti University, Meerut, Uttar Pradesh, India.

** Dean of Research and Development cell, Faculty of Law, Swami Vivekanand Subharti University, Meerut, Uttar Pradesh, India.

¹ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*

² Citron DK (2014) *Hate crimes in cyberspace*. Harvard University Press, Cambridge.

psychological, social, and professional consequences.³ These harms are not incidental but structurally gendered, reflecting and amplifying existing social inequalities.⁴

Despite growing scholarship on AI harms, existing studies largely address privacy, cybercrime, or platform regulation in isolation. This paper departs from such approaches by offering an integrated constitutional analysis that connects gender, technology, and platform power. The originality of this study lies in demonstrating that AI-generated harms require a shift from traditional state-centric constitutionalism to a platform-centric model of rights protection. By situating deepfake harm within broader structures of algorithmic governance, the paper provides a novel framework for understanding how digital technologies reshape constitutional rights in contemporary democracies. This paper argues that AI-generated deepfake harm constitutes a form of structural gender discrimination that existing Indian constitutional doctrine is inadequate to address, particularly in the context of platform-mediated harms.

II. RESEARCH METHODOLOGY

This study adopts a qualitative, comparative doctrinal methodology to analyse the constitutional and regulatory implications of AI-generated deepfake harms. India, the European Union, and the United States are selected as comparative jurisdictions due to their distinct approaches to AI governance, constitutional rights, and platform regulation.

The analysis is based on primary legal sources, including constitutional provisions, statutory frameworks, and judicial decisions, as well as secondary sources such as academic literature, policy reports, and international guidelines.

The comparative framework is structured around three analytical dimensions:

- (1) privacy and dignity,
- (2) freedom of expression, and
- (3) algorithmic accountability.

These dimensions are used to evaluate the adequacy and limitations of existing legal frameworks. The study further employs a critical and feminist analytical approach to assess how AI harms disproportionately affect women and reinforce structural inequalities.

III. THEORETICAL FRAMEWORK

This article advances a central claim that AI-generated deepfake harm constitutes not merely a technological misuse or an extension of cybercrime, but a qualitatively distinct form of structural gender discrimination that exposes fundamental limitations within existing Indian constitutional and statutory frameworks.⁵ Unlike traditional forms of harm, deepfake technologies collapse the boundary between reality and fabrication, enabling the creation of hyper-realistic, non-consensual representations that directly undermine an individual's informational self-determination, bodily

³ McGlynn C, Rackley E, Houghton R (2017) Beyond “revenge porn”: the continuum of image-based sexual abuse. *Fem Leg Stud*.

⁴ Crenshaw K (1991) Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Rev*.

⁵ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*; Crenshaw K (1991) Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Rev*.

autonomy, and reputational integrity.⁶ What distinguishes such harm is not only its scale and speed but its embeddedness within platform architectures and algorithmic amplification systems that disproportionately target women, thereby reproducing and intensifying pre-existing gender hierarchies.⁷ In this sense, deepfake abuse operates as a form of digitally mediated structural violence rather than isolated individual wrongdoing.⁸ The article therefore reconceptualises deepfake harm as a constitutional issue, arguing that doctrines articulated in cases such as *Justice K.S. Puttaswamy v Union of India*⁹ must be extended beyond state surveillance to encompass private, platform-driven violations of privacy and dignity. At the same time, it critically interrogates the continued reliance on formal free speech frameworks, as reflected in *Shreya Singhal v Union of India*¹⁰, demonstrating that AI-generated sexualised content does not fit comfortably within existing categories of protected expression, thereby necessitating a re-evaluation of the harm principle in digital contexts.¹¹

Further, the article contends that prevailing equality jurisprudence remains inadequately equipped to address algorithmically mediated harms, as it often fails to capture the systemic and intersectional nature of digital discrimination.¹² Drawing on feminist constitutional theory, the analysis reveals how deepfake technologies are embedded within broader socio-technical systems that privilege visibility, virality, and commodification of women's bodies, thereby transforming individual acts of abuse into structural patterns of subordination.¹³ Consequently, the article argues for a shift from reactive, individualised legal remedies toward a proactive, structural regulatory approach that integrates constitutional principles with platform accountability, algorithmic transparency, and gender-sensitive safeguards. In doing so, it positions AI governance not merely as a question of technological regulation, but as a site for reimagining constitutional rights in the age of algorithmic power.

Why AI Harms Are Gendered?

Unlike general privacy violations, deepfake pornography disproportionately targets women and reinforces structural patterns of gender inequality. This brings such harms within the scope of Article 14 of the Indian Constitution, requiring a shift from formal equality to substantive equality in constitutional analysis. AI-enabled harms cannot be understood as gender-neutral technological externalities; rather, they constitute a form of structurally produced inequality that is deeply embedded in both social hierarchies and the design logics of digital systems.¹⁴ The disproportionate

⁶ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*.

⁷ Citron DK (2014) *Hate crimes in cyberspace*. Harvard University Press, Cambridge; Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge

⁸ McGlynn C, Rackley E, Houghton R (2017) Beyond "revenge porn": the continuum of image-based sexual abuse. *Fem Leg Stud*.

⁹ Justice K.S. Puttaswamy v Union of India (2017) 10 SCC 1.

¹⁰ Shreya Singhal v Union of India (2015) 5 SCC 1.

¹¹ MacKinnon CA (1993) *Only words*. Harvard University Press, Cambridge; Fiss O (1996) *The irony of free speech*. Harvard University Press, Cambridge.

¹² Crenshaw K (1991) Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Rev*.

¹³ Braidotti R (2013) *The posthuman*. Polity Press, Cambridge.

¹⁴ Citron DK (2014) *Hate crimes in cyberspace*. Harvard University Press, Cambridge; Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

targeting of women through deepfake pornography, non-consensual image generation, and AI-amplified harassment reflects not merely patterns of misuse but the reproduction and intensification of pre-existing gendered power asymmetries within algorithmic environments.¹⁵ In this sense, AI does not create new hierarchies in isolation but reorganises and scales existing forms of gender-based subordination through automated, data-driven processes.¹⁶

The concept of intersectionality, developed by Kimberlé Crenshaw, is crucial in demonstrating that vulnerability to AI-mediated harm is unevenly distributed and structurally conditioned.¹⁷ Women located at the intersections of caste, class, religion, and sexuality experience compounded exposure to digital exploitation, while simultaneously facing greater barriers to legal redress. Algorithmic systems trained on historically biased and socially skewed datasets reproduce these inequalities by embedding discriminatory patterns into automated decision-making and content generation processes.¹⁸ As a result, AI systems do not merely reflect bias but institutionalise it, transforming historically contingent inequalities into persistent and scalable forms of digital harm.

From a normative perspective, the capabilities approach associated with Martha Nussbaum provides a critical framework for understanding the constitutional implications of such harms.¹⁹ Deepfake pornography undermines core human capabilities, particularly those relating to bodily integrity, autonomy, and dignity, by enabling the non-consensual appropriation and manipulation of an individual's identity.²⁰ These harms extend beyond informational privacy to constitute a deeper violation of personhood, as individuals lose meaningful control over their representation in digital space. This raises significant constitutional concerns, particularly in relation to dignity and autonomy as recognised under Indian fundamental rights jurisprudence.

Techno-feminist critiques, including those advanced by Rosi Braidotti, further illuminate how AI systems are embedded within broader socio-technical and economic structures that privilege dominant groups while marginalising others.²¹ The underrepresentation of women in AI development, combined with platform architectures that prioritise engagement, visibility, and virality, creates conditions in which gendered abuse is not only enabled but algorithmically amplified.²² This shifts the nature of harm from isolated acts of individual misconduct to systemic patterns of platform-mediated violence.

Accordingly, AI-generated abuse must be conceptualised as a form of gender-based structural harm that implicates constitutional guarantees of equality, dignity, and freedom. The failure to recognise

¹⁵ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*; McGlynn C, Rackley E, Houghton R (2017) Beyond "revenge porn": the continuum of image-based sexual abuse. *Fem Leg Stud*.

¹⁶ Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

¹⁷ Crenshaw K (1991) Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Rev*.

¹⁸ Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

¹⁹ Nussbaum M (2011) *Creating capabilities: the human development approach*. Harvard University Press, Cambridge.

²⁰ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*.

²¹ Braidotti R (2013) *The posthuman*. Polity Press, Cambridge.

²² Zuboff S (2019) *The age of surveillance capitalism*. Profile Books, London; Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

its gendered nature risks reducing such harms to neutral technological issues, thereby obscuring the deeper structural inequalities they reproduce. A meaningful legal response therefore requires a shift from formal neutrality to a substantive equality framework that acknowledges and addresses the gendered dimensions of algorithmic harm within contemporary constitutional analysis. This necessitates a shift from formal equality to substantive equality in addressing algorithmic harms.²³

IV. CONSTITUTIONAL ANALYSIS

Privacy, Dignity, and Informational Self-Determination

While *Justice K.S. Puttaswamy v Union of India*²⁴ recognises privacy as including dignity and informational self-determination, it does not adequately address harms arising from fabricated identities such as deepfakes. This reveals a doctrinal limitation, as existing privacy jurisprudence is premised on the protection of real information rather than synthetic misrepresentation. However, AI-generated harms expose limitations in this framework, particularly where violations originate from private platforms rather than the State.²⁵ The emergence of AI-generated deepfake technologies exposes a critical limitation in this framework, as the most pervasive violations of privacy and dignity now originate not from the State but from private, platform-mediated ecosystems.

Deepfake pornography fundamentally disrupts the notion of informational self-determination by enabling the non-consensual creation, manipulation, and circulation of intimate representations. Unlike conventional privacy breaches, these harms do not merely involve the unauthorized disclosure of existing information but the fabrication of synthetic identities that are indistinguishable from reality. This produces a qualitatively different form of harm, wherein the individual loses control not only over personal data but over the very constitution of their digital personhood²⁶. In this sense, deepfake abuse must be understood as an “identity harm” that implicates both dignity and autonomy, extending the scope of privacy beyond its traditional informational boundaries. Consequently, a constitutional reading of *Puttaswamy* must evolve to address these technologically mediated violations, recognising that privacy in the digital age includes protection against the creation of false yet credible representations of the self.

Proportionality and Platform Power

The doctrine of proportionality, central to Indian constitutional law, requires that any restriction on fundamental rights must pursue a legitimate aim, be suitable to achieve that aim, be necessary in the absence of less restrictive alternatives, and maintain a balance between competing interests.²⁷ While traditionally applied to state action, the rise of AI-driven platform governance raises the question of whether and how proportionality should extend to private actors whose technological infrastructures significantly shape the exercise of fundamental rights.

AI systems deployed by digital platforms operate through opaque algorithms, automated amplification mechanisms, and engagement-driven design choices that often prioritise virality over

²³ Nussbaum M (2011) *Creating capabilities: the human development approach*. Harvard University Press, Cambridge.

²⁴ *Justice K.S. Puttaswamy v Union of India* (2017) 10 SCC 1.

²⁵ Zuboff S (2019) *The age of surveillance capitalism*. Profile Books, London.

²⁶ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev.*

²⁷ Barak A (2012) *Proportionality: constitutional rights and their limitations*. Cambridge University Press, Cambridge.

harm prevention.²⁸ In the context of deepfake abuse, such systems enable the rapid and large-scale dissemination of harmful content, frequently outpacing existing moderation and redress mechanisms. This creates a structural asymmetry while the harm is immediate, scalable, and often irreversible, the regulatory response remains reactive and fragmented.

This raises a critical constitutional question: - *Can algorithmic systems that systematically facilitate the amplification of harmful content satisfy the requirements of proportionality?*

From a doctrinal perspective, the answer is increasingly in the negative. Where platform architectures are designed in ways that foreseeably enable harm accordingly, this raises the question of whether fundamental rights under the Indian Constitution should have horizontal application to private digital platforms such as through recommendation systems that amplify sensational or exploitative content they fail the necessity and balancing prongs of proportionality. In such cases, the absence of adequate safeguards, transparency, and accountability mechanisms suggests that these systems impose disproportionate burdens on fundamental rights, particularly the rights to dignity and equality. This necessitates a reconceptualisation of proportionality to account for platform power, where private technological actors exercise quasi-regulatory control over digital spaces.²⁹

Freedom of Speech and the Limits of Protection

The decision in *Shreya Singhal v Union of India*³⁰ protects free speech against vague restrictions; however, its framework becomes inadequate in the context of AI-generated deepfake pornography, where the issue is not legitimate expression but non-consensual, harmful fabrication. Treating such content as protected speech risks undermining dignity and equality. However, the application of this framework to AI-generated deepfake content reveals significant conceptual tensions. Deepfake pornography does not fit neatly within traditional categories of protected speech, as it involves the non-consensual fabrication of sexualised content that directly harms the subject's dignity, privacy, and reputation.³¹

Unlike legitimate expression, deepfake content lacks an underlying expressive intent that contributes to public discourse. Instead, it operates as a form of harmful, manipulative, and often exploitative representation. Treating such content as protected speech risks collapsing the distinction between expression and abuse, thereby undermining the very constitutional values that free speech doctrine seeks to preserve.³²

Accordingly, regulating deepfake pornography does not constitute an impermissible restriction on free speech but rather a necessary intervention to protect competing constitutional rights, particularly dignity and equality. This requires a doctrinal shift from a purely speech-centric analysis toward a harm-based framework, where the constitutional permissibility of regulation is assessed in light of the nature, impact, and structural context of the harm involved. Such an approach aligns with the broader constitutional commitment to substantive equality and ensures that freedom of speech is not weaponised to legitimise gendered forms of digital violence.

²⁸ Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

²⁹ Zuboff S (2019) *The age of surveillance capitalism*. Profile Books, London; Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

³⁰ *Shreya Singhal v Union of India* (2015) 5 SCC 1.

³¹ MacKinnon CA (1993) *Only words*. Harvard University Press, Cambridge.

³² Fiss O (1996) *The irony of free speech*. Harvard University Press, Cambridge.

V. STATUTORY GAPS IN INDIA

Despite the constitutional recognition of privacy, dignity, and equality, India's statutory framework remains structurally inadequate to address the distinct harms posed by AI-generated deepfakes. Existing legal regimes were designed for earlier forms of cyber misconduct and consequently fail to capture the technological specificity, scale, and gendered nature of synthetic media abuse. This results in a regulatory mismatch, where rapidly evolving forms of harm are addressed through fragmented and conceptually outdated legal provisions.³³

The Digital Personal Data Protection Act 2023³⁴ exemplifies this limitation. While the Act establishes a consent-based framework for personal data processing, it is premised on the assumption that harm arises from the unauthorized collection, storage, or transfer of existing personal data. Deepfake technologies, however, frequently operate through the generation of synthetic content that may not directly rely on identifiable personal data in a conventional sense, thereby falling outside the Act's core regulatory logic. As a result, individuals subjected to non-consensual deepfake pornography may find limited recourse under data protection law, as the injury lies not only in data misuse but in the fabrication and circulation of manipulated identities.

Similarly, the criminal law framework under the Bharatiya Nyaya Sanhita 2023³⁵ does not explicitly recognise AI-generated sexual abuse or synthetic impersonation as distinct offences. While certain provisions relating to obscenity, defamation, or identity misuse may be invoked, these are ill-suited to address the unique characteristics of deepfake harm, including its scalability, anonymity, and cross-border dissemination.³⁶ The absence of offence-specific definitions results in doctrinal ambiguity, inconsistent enforcement, and evidentiary challenges, particularly where victims must prove intent, authorship, or direct harm in technologically complex environments.

These limitations point to three interrelated structural gaps. First, there is an absence of clear and technologically informed legal definitions of deepfake harm, which prevents the law from adequately recognising and categorising the injury involved. Without such definitional clarity, enforcement agencies are compelled to rely on analogical reasoning, often leading to under-criminalisation or misclassification of offences. Second, platform accountability remains weak and largely reactive. Current intermediary liability frameworks do not impose sufficiently robust obligations on platforms to detect, prevent, or rapidly remove AI-generated harmful content, nor do they mandate transparency in algorithmic amplification processes.³⁷ This is particularly problematic given that platforms are not passive intermediaries but active participants in shaping the visibility and circulation of content. Third, existing legal regimes provide limited survivor-centric remedies. Procedural delays, lack of specialised digital evidence mechanisms, and the absence of fast-track takedown protocols exacerbate harm, especially in cases where reputational damage and psychological injury are immediate and irreversible.³⁸

³³ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*

³⁴ Digital Personal Data Protection Act (2023) India.

³⁵ Bharatiya Nyaya Sanhita (2023) India.

³⁶ McGlynn C, Rackley E, Houghton R (2017) Beyond "revenge porn": the continuum of image-based sexual abuse. *Fem Leg Stud*.

³⁷ Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

³⁸ Citron DK (2014) *Hate crimes in cyberspace*. Harvard University Press, Cambridge; McGlynn C,

Taken together, these gaps reveal that India's statutory framework is not merely incomplete but conceptually misaligned with the realities of AI-mediated harm. Addressing deepfake abuse therefore requires more than incremental legal reform it demands a shift toward a technology-sensitive, gender-responsive regulatory approach that integrates clear legal definitions, proactive platform obligations, and effective, victim-oriented remedies.³⁹

VI. COMPARATIVE ANALYSIS: THE EU AI ACT AND PREVENTIVE REGULATION

Applying the above comparative framework, the following analysis evaluates the regulatory approaches of India, the European Union, and the United States in addressing AI-generated harms. Comparative analysis of the European Union's regulatory framework provides critical insights into how legal systems can move beyond reactive approaches toward anticipatory governance of AI-related harms. The European Union's Artificial Intelligence Act adopts a risk-based regulatory model that classifies AI systems according to the level of threat they pose to fundamental rights, safety, and democratic values.⁴⁰ This approach marks a significant departure from traditional *ex post* legal frameworks by embedding preventive safeguards within the design, development, and deployment stages of AI systems, thereby shifting regulatory focus toward systemic risk mitigation.⁴¹

Specifically, Article 5 (prohibited practices), Article 6 (high-risk systems), and Article 27 (fundamental rights impact assessments) of the EU AI Act provide a structured framework for identifying and mitigating risks to fundamental rights.⁴² Article 5 prohibits certain forms of manipulative and deceptive AI systems that undermine individual autonomy, while Article 6 designates specific applications as "high-risk," thereby subjecting them to enhanced regulatory obligations, including risk management, data governance, and human oversight requirements. Article 27 further institutionalises a preventive, rights-based approach by mandating fundamental rights impact assessments prior to deployment, ensuring that potential harms are evaluated and mitigated at the design stage rather than addressed solely through post hoc remedies.

In the context of deepfake technologies, these provisions are particularly significant, as they recognise that harms arising from synthetic media are not merely incidental but structurally embedded within the functioning of AI systems.⁴³ By requiring *ex ante* safeguards, transparency, and accountability, the EU framework shifts the regulatory focus from individual liability to systemic risk management. This reflects a broader normative commitment to integrating fundamental rights considerations into technological governance.⁴⁴

In contrast, India's current regulatory approach remains largely reactive and fragmented. A comparative evaluation reveals important structural differences between jurisdictions. The European Union adopts a proactive, risk-based regulatory model, embedding safeguards such as fundamental rights impact assessments and *ex ante* compliance obligations. In contrast, the United States follows

Rackley E, Houghton R (2017) Beyond "revenge porn": the continuum of image-based sexual abuse. *Fem Leg Stud*.

³⁹ European Union (2024) Artificial Intelligence Act. <https://eur-lex.europa.eu>.

⁴⁰ European Union (2024) Artificial Intelligence Act. <https://eur-lex.europa.eu>.

⁴¹ OECD (2024) OECD AI principles (updated). <https://doi.org/10.1787/ai-principles-en>.

⁴² European Union (2024) Artificial Intelligence Act. <https://eur-lex.europa.eu>.

⁴³ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev*.

⁴⁴ OECD (2024) OECD AI principles (updated). <https://doi.org/10.1787/ai-principles-en>.

a market-oriented and speech-protective model, where regulatory intervention is limited by strong First Amendment protections. India reflects a hybrid but underdeveloped framework, combining constitutional protections with fragmented statutory regulation, yet lacking a comprehensive AI governance strategy.

This comparison demonstrates that India's current approach remains predominantly reactive, addressing harms only after their occurrence, whereas the EU model emphasises prevention and systemic risk mitigation. This distinction highlights the need for India to transition toward a more anticipatory and rights-oriented regulatory framework.

The EU model therefore demonstrates the importance of transitioning toward a preventive, rights-oriented regulatory paradigm that integrates constitutional values into the lifecycle of AI systems. It suggests that effective AI governance must move beyond reactive enforcement toward proactive design-based regulation, where the protection of dignity, autonomy, and equality is embedded within technological architectures themselves.⁴⁵

VII. DISCUSSION

This paper argues that AI governance must be understood as a site of constitutional transformation rather than mere regulatory adaptation. The analysis demonstrates that existing constitutional doctrines, designed to regulate state action, are inadequate in addressing harms originating from private, platform-driven ecosystems. The findings highlight a shift from state-centric to platform-centric constitutionalism, where private digital actors exercise significant influence over the realisation of fundamental rights. This requires a reconceptualisation of privacy, equality, and free speech in the context of algorithmic governance. However, AI-driven harms increasingly originate within private, transnational platform ecosystems that exercise significant control over the creation, circulation, and amplification of digital content.⁴⁶ This transformation challenges the adequacy of existing doctrines, as the most consequential threats to fundamental rights now arise outside the conventional state-centric paradigm.

This shift gives rise to a significant governance gap, wherein neither constitutional law nor statutory regulation effectively addresses platform-mediated harms. While constitutional jurisprudence provides robust protections for privacy, dignity, and free speech, its application to private technological actors remains underdeveloped.⁴⁷ At the same time, statutory frameworks lack the conceptual and technical specificity required to regulate AI-generated harms, particularly those that operate through algorithmic amplification and cross-border dissemination.⁴⁸

Moreover, the gendered nature of deepfake abuse exposes deeper limitations within existing equality jurisprudence. Formal equality frameworks, which focus on non-discrimination in abstract terms, are ill-equipped to address the structural and systemic dimensions of algorithmic harm.⁴⁹ Deepfake technologies do not simply produce isolated instances of discrimination, they reinforce and scale

⁴⁵ European Union (2024) Artificial Intelligence Act. <https://eur-lex.europa.eu>

⁴⁶ Zuboff S (2019) *The age of surveillance capitalism*. Profile Books, London; Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

⁴⁷ Justice K.S. Puttaswamy v Union of India (2017) 10 SCC 1; Shreya Singhal v Union of India (2015) 5 SCC 1.

⁴⁸ Chesney R, Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *Calif Law Rev.*

⁴⁹ Crenshaw K (1991) Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Rev.*

gendered patterns of exploitation by embedding them within platform architectures and data-driven systems. This necessitates a shift toward a substantive equality approach that recognises how digital infrastructures can reproduce and intensify existing social hierarchies.⁵⁰

The analysis also underscores the need to reconceptualise constitutional rights in light of algorithmic governance.⁵¹ First, fundamental rights must extend to platform-mediated harms, recognising that private actors exercising significant control over digital environments can effectively shape the conditions under which rights are realised. Second, the principle of equality must be reinterpreted to account for algorithmic bias, data asymmetries, and the gendered distribution of harm. Third, legal frameworks must transition from reactive, remedy-based models toward preventive and structural regulation, incorporating mechanisms such as algorithmic audits, design-based safeguards, and proactive content moderation obligations.⁵²

Taken together, these insights suggest that the challenge posed by AI-generated harms is not merely regulatory but constitutional in nature. Addressing this challenge requires a reorientation of legal thought, moving beyond traditional categories toward a framework that integrates technological realities, structural inequality, and constitutional values.

VIII. CONCLUSION

AI-generated deepfake pornography presents a profound and evolving challenge to the protection of constitutional rights in India. As this article has demonstrated, such harms cannot be adequately understood within existing legal categories of cybercrime, privacy violation, or obscenity. Rather, they constitute a distinct form of structurally embedded, gender-based harm that undermines core constitutional values of dignity, autonomy, and equality. The current legal framework both constitutional and statutory remains insufficiently equipped to address the scale, complexity, and gendered nature of these harms.

An effective response requires a multi-dimensional and forward-looking approach that integrates constitutional principles with targeted statutory reform and robust platform accountability. This includes, first, the explicit legal recognition and definition of deepfake-related harms within both civil and criminal law to ensure doctrinal clarity and consistent enforcement. Second, intermediary liability frameworks must be strengthened to impose proactive obligations on digital platforms, including timely content removal, transparency in algorithmic processes, and responsibility for the amplification of harmful content. Third, legal and institutional mechanisms must prioritise survivor-centred remedies, including rapid takedown procedures, accessible reporting systems, and specialised support structures that address the unique harms caused by digital abuse.

In addition, regulatory frameworks must incorporate preventive safeguards such as mandatory algorithmic audits, bias mitigation requirements, and fundamental rights impact assessments to ensure that harms are addressed at the level of system design rather than solely through post hoc remedies. Ultimately, the governance of AI must be grounded in the principles of constitutional morality, ensuring that technological innovation is aligned with the protection of human dignity,

⁵⁰ Nussbaum M (2011) *Creating capabilities: the human development approach*. Harvard University Press, Cambridge.

⁵¹ Pasquale F (2015) *The black box society: the secret algorithms that control money and information*. Harvard University Press, Cambridge.

⁵² European Union (2024) Artificial Intelligence Act. <https://eur-lex.europa.eu>

substantive equality, and democratic values. Only through such an integrated and rights-based approach can the law effectively respond to the challenges posed by AI and ensure that digital transformation does not come at the cost of justice and human rights.

In doing so, this study advances both theoretical and practical contributions by bridging constitutional law, feminist theory, and AI governance, offering a novel framework for addressing emerging forms of digital harm in contemporary legal systems.

IX. RECOMMENDATIONS: TOWARDS A GENDER-RESPONSIVE AI GOVERNANCE FRAMEWORK

The preceding analysis demonstrates that existing constitutional and statutory frameworks are insufficient to address the structural and gendered nature of AI-generated deepfake harms. In light of these limitations, a multi-layered and forward-looking regulatory approach is necessary. This section proposes targeted recommendations aimed at bridging doctrinal, statutory, and institutional gaps within the Indian legal framework.

Legal Recognition of Deepfake Harm

A foundational step is the explicit legal recognition of deepfake-related harms within Indian law. Existing provisions under the Bharatiya Nyaya Sanhita 2023 and the Digital Personal Data Protection Act 2023 do not adequately capture the nature of synthetic media abuse. Legislators should introduce clear and technology-specific definitions of deepfake harm, particularly in relation to non-consensual sexual content, identity manipulation, and synthetic impersonation. Such recognition is essential to ensure doctrinal clarity, consistent enforcement, and effective victim protection.

Strengthening Platform Accountability

Given the central role of digital platforms in the creation and dissemination of deepfake content, regulatory frameworks must impose stronger intermediary obligations. These should include proactive detection of AI-generated harmful content, time-bound takedown requirements for sexualised deepfakes, and greater transparency in algorithmic amplification systems. Additionally, platforms should incur liability for failing to act on verified complaints. Such measures would shift the regulatory approach from passive intermediary immunity to active responsibility, recognising platforms as key actors in addressing AI-mediated harm.

Survivor-Centred Remedies and Evidence Preservation

A critical gap in the current framework is the absence of effective, survivor-oriented mechanisms to address deepfake harm. Legal reforms must prioritise fast-track takedown procedures for harmful content, alongside robust digital evidence preservation protocols to prevent deletion or manipulation. In addition, reporting systems should be confidential, accessible, and responsive to the needs of women and marginalised groups. Institutional support, including legal aid and psychological assistance, is equally essential. Such measures are necessary to respond to the immediacy, scale, and often irreversible nature of harm caused by deepfake technologies.

Preventive Regulation through Algorithmic Governance

India must transition from a reactive enforcement model to a preventive regulatory approach in governing AI systems. Drawing on comparative frameworks such as the European Union's AI Act, this requires the introduction of mandatory algorithmic audits for high-risk systems, alongside bias

detection and mitigation obligations. Regulatory mechanisms should also mandate fundamental rights impact assessments prior to deployment and incorporate design-based safeguards to prevent misuse of generative AI technologies. Such an approach embeds constitutional values within system architecture, reducing reliance on post hoc remedies.

Integrating a Feminist Constitutional Perspective

Legal and policy responses to AI harms must explicitly incorporate feminist constitutional principles to address their gendered dimensions. This entails recognising intersectional vulnerabilities across caste, class, and gender, and adopting a substantive equality framework within AI governance. It also requires the meaningful inclusion of women and marginalised groups in AI design, policy-making, and regulatory processes. A gender-responsive approach ensures that AI regulation moves beyond formal neutrality to actively address structural inequalities embedded within digital systems.

Constitutional Extension to Platform-Mediated Harms

Finally, there is a need to reconsider the traditional state-centric orientation of constitutional law. Given the significant influence of private platforms over digital spaces, constitutional principles particularly those relating to dignity, equality, and privacy must extend to platform-mediated harms. This may require doctrinal development toward the horizontal application of fundamental rights, ensuring that private technological actors are subject to constitutional scrutiny where they exercise substantial control over rights.