DEEPFAKE-INDUCED VICARIOUS TRAUMA: RETHINKING PSYCHOLOGICAL HARM AND TORT DAMAGES IN INDIA

Piya Chowdhury, BBA LL.B. (Hons.), Amity Law School, Noida, Uttar Pradesh

ABSTRACT

The human mind is ingenious in its technological innovations. Yet, this proliferation of technology carries severe consequences for the very mind that created it. Issues like deep fakes, now trigger mental health challenges, ranging from body dysmorphia, digital addiction to severe stress and anxiety. This raises a pivotal questions, is the current legal framework equipped to protect the human mind from the repercussions of its own technological creations? Or What does the law owe a person whose dignity and identity have been weaponized, sexualized and spread across the internet by deepfakes, only to be deemed legally "resolved" as soon as the post is removed? This paradox represents the India's digital legal landscape. The IT (Intermediary Guidelines) Amendment Rules, 2026 promises to impose such rapid erasure of the unlawful content, yet conspicuously fails to take into account the profound psychological trauma inflicted upon the victims, rendering their mental harm irrelevant. More so Indian laws rely on a flawed assumption that the digital harm ends when the unlawful content is removed. This assumptions becomes more fatal in the context of deepfakes, which are known for their nature of speed, virality and content recidivism. This article argues that such silence on mental harms is rather doctrinal rather than an incidental oversight.

Keywords: Non-material Harm, Psychological Harm, Victim Compensation, Deepfakes, Content Recidivism, Mental Health law, AI Governance.

Introduction

The first and foremost immediate injury inflicted by deepfakes is not economical, nor is reputational damage, it is psychological. The assault is upon the mind, which manifests loss of autonomy, humiliation, anxiety and so far, as causes the slow erosion of one's self esteem. Yet, ironically the law is least equipped to recognise and provide remedy for this very mental injury.

In contrast, the statues of U.S, U.K and analogous GDPR have already moved ahead of such conceptual thresholds and have begun compensating the victims for their emotional injury/non-material harms. Whereas, Indian regulations by design operates *ex facto*, intervening when the harm has been materialised. This is further exacerbated when perpetrators are difficult to identify and are often obfuscated by their anonymity. As evidently observed by contemporary scholars Indian laws continue to rely on indirect and fragmented statues that fails to contain the enormity of deepfake harms.[1] The article highlights a central proposed framework that by calibrating a doctrinal shift towards victim centric damages and institutional reforms, the article seeks to reorient the India's regulations from reactive censorship of unlawful content to repositioning psychological harm as a legally cognizable and compensable wrong.

Regulating Deepfakes: A Shift Towards Stricter Regulation

At present India continues to tackle the issue of deepfakes through its existing dispersed set of legal statues rather than a consolidated one. The Information Technology Act, 2000 for an instance has particular provisions such as Section 66C and Section 66D are invoked to prosecute identity theft and cheating by impersonation, whereas Section 67 and Section 67A are primarily used when the deepfakes takes on the pornographic or sexually explicit nature.[2] Parallelly the Bharatiya Nyaya Sanhita prosecutes deepfakes that in nature offensive and fall under the ambit expressly defined offences such as defamation, impersonation, obscenity, or cheating.[3] Whereas, Regulatory governance of deepfakes are further shaped through the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 imposes due diligence obligations on intermediaries, wherein they have to take vigilantly and

[1] Chintan Pathak, *Impact of Deepfake Technology on Indian Society*, 10 Vidhi Bharati L.J. 1, 5 (2024), https://www.vtclaw.ac.in/upload/journal/1-%20IMPACT%20OF%20DEEPFAKE%20TECHNOLOGY%20ON%20INDIAN%20SOCIETY-%20Dr-%20Chintan%20Pathak.pdf.
[2] Information Technology Act, 2000, §§ 66C, 66D, 67, 67A
[3] Bharatiya Nyaya Sanhita, 2023, §§ 356, 319, 292, 318

expeditiously to remove unlawful content within stipulated timelines, if they fail to do so, they risk losing their legal immunity given to exempt themselves from liability for any third-party interferences. Rule 7 reinforce this by explicitly stating that if any intermediary fails to perform its due diligence prescribed under the rules, the safe harbour protection ensured under the Section 79(1) of the Information Technology Act, 2000 will no longer be applicable.[4] Thereby positioning intermediaries in the frontline of enforcements. The most apparent changes and progressive shift is seen through the recent amendments of IT Rules, 2026[5] that shifts towards a stricter and faster intervention to catch up with the velocity and exposure of deepfakes.

Comparative Analysis of the Takedown Timelines (IT Rules 2021 vs 2026)

Classification of the Digital Harm	**Tiggering Mechanism/ Authority**	**IT Rules (2021)**	**Amended IT Rules (2026)**
Non-consensual Sexually explicit Deepfakes	Complaint initiated by Victim (Rule 3 (2)(b))	24 hours	2 hours
Unlawful Deepfakes/ SGI	Government/Court Order (Rule 3(1)(d))	36 hours	3 hours
Individual Rights Violations	Grievances regarding Reputation or Identity	72 hours (Acknowledgment)	36 hours (Resolution)
General Synthetic Content	General user reporting or third-party	15 days	7 days

1. Taxonomy of 'Synthetically Generated Information' (SGI)

The IT Rule, 2026 Amendment reflects a jurisprudential leap in the terminology. Earlier deepfakes were addressed through archaic and narrowly defined terms such as 'morphed imagery', now the terminology has evolved to comprehensively defined term "Synthetically Generated Information" under the Rule 2(1)(wa).[6] The earlier rules were

[4] Information Technology Act, 2000 § 79(1).
[5] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026.
[6] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026,

ambiguous and were failing to capture the complexities AI-driven content.

2. Recalibrating Intermediary Liability

Through the insertion of Rule 4 (1A)(b)[7], wherein the Intermediaries are now given scope to employ automated tools or other technological advanced measures to remove and detect the unlawful synthetic content and thereby maintaining their legal immunity "safe harbour." This reflects evolving jurisprudence to encourage the deployment of technological safeguards without penalising the intermediaries. The framework encourages to use automated tools in the right places eventually to instil a better understanding how to adapt with technological advancements.

3. The Drastic rapid takedown

The most notable amendment is the reduction of the earlier takedown window of 24 hours to now 2 hours under the 2026 Amendments, which reflects a paradigm shift in targeting deepfakes and non-consensual intimate imagery. The transition was imperative as even though the earlier timelines were considered expeditious still, they were largely undermined and insufficient by the spread of these unlawful content because of their nature to be viral and their speed.

4. Sovereign oversight and the 36-hour Grievance standards

The 2026 Amendments has accelerated the governance oversight and has now introduced speedy compliance rules through the 3 hour takedown timelines. Under the Rule 3(1)(d) the intermediaries are mandated to remove the unlawful SGI within 3 hour upon receiving a Government Notice or Court's Order. Parallelly the Grievance mechanisms has also been transitioned from a sluggish channel where earlier grievances were only 'acknowledged' within 72 hours towards a more resolution based model by resolving the harms within 36 hours.

Inadequacies of Indian Laws: Absence of Victim-Centric Justice

A critical infirmity can be seen in the present remedies that how they are reactive, yet also are

Rule 2(1) (wa).

[7] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Rule 4 (1A)(b)

devoid of victim centred remedies. The absence is paramount in the current statues, regulations, guidelines and even in the IT Amendment Rules, 2026 as well as in the DPDP, 2023. The statutory provision imposes heavy fines or penalties, rather than directing those amounts towards the victims, such amounts are directed towards government exchequers.[8] There is no compensation fund for the victims that falls prey to these vicious crimes, nor is there any mandatory institutional reforms to provide psychological rehabilitation to the victims. They are left to navigate their trauma, reputational damages, fear and societal consequences in isolation.[9] Even the grievance redressals like the platform takedowns, they are still inadequate as empirical studies have showed that the nature of deepfake are pervasive and the removed content often resurfaces, phenomenon known as 'content recidivism' creating the endless cycle of abuse.[10] The harm therefore is not cured rather it is a continuing wrong. They are left with surface level 'justice' with no support, their mental harms are conveniently left to heal itself with time. The cost of therapies, psychiatrists, medical treatments all are to borne by the victims along with their trauma. It is in no doubt that the IT amendment rules 2026, is a progressive step towards better governance. The 2-3 hours takedown from earlier of 36-72 hours will create a stricter environment for intermediaries and it is imperative for the intermediaries to act responsibly as public trust is reposed on them. But still the problem remains the same, on the surface level it appears that the vicious unlawful content is dealt with a strict hand yet it is not victim centric, the victims are left with endless confusion, helplessness, fears, trauma etc.

These issues are further exacerbated when the statues are silent on what really constitutes as harm, for instance the DPDP Act, 2023 has removed the broad definition of 'harm' that was present in the prior DPDP iterations where in the ambit of definition of harm it recognised mental injury, reputational loss, discriminatory treatment and so on.[11] Simultaneously, the laws and regulations assumes that the digital harms can be addressed or confined to a single unlawful act and more so can be dealt in a straightforward manner where the use of deepfakes can only be curbed in a limited sense, either someone is defamed, cheated, deceived or has been exposed

[8] *Supra note 1*
[9] Equality Now, *Research Exposes How Women in India Are Being Abused, Shamed and Silenced Online* (2025), https://equalitynow.org/news/press-releases/research-exposes-how-women-in-india-are-being-abused-shamed-and-silenced-online/
[10] Aisha Down, *'The Chilling Effect': How Fear of 'Nudify' Apps and AI Deepfakes is Keeping Indian Women Off the Internet*, The Guardian (Nov. 5, 2025), https://www.theguardian.com/global-development/2025/nov/05/india-women-ai-deepfakes-internet-social-media-artificial-intelligence-nudify-extortion-abuse.
[11] S&R Associates, *How Much and How Bad? – "Significant Others" in India's New Data Regime*, S&R Assocs. (Nov. 2, 2023), https://www.snrlaw.in/how-much-and-how-bad-significant-others-in-indias-new-data-regime/.

to sexual or obscene content. But it fails to take into account that deepfakes are notorious in nature as they may involve elements of various offences at once, such as reputational harms, sexual exploitations, impersonations, identity thefts, and privacy violations, yet the crime is restricted to one of the offences and never really fitting entirely into these categories. This happens because the laws were constructed around conventional scenarios that existed in the pre AI evolution era. Therefore they are bound to fragment the harms caused by deepfakes. The fragmentation results in inevitable consequences, firstly reducing the complex harms/injuries into narrow and confined labels. For an instance Image-Based Sexual Abuse IBSA deepfakes are currently dealt under 'obscene content' under the BNS and other statues but it silently shifts the focus to public morality rather than the victim's state of being and towards their loss of autonomy and identity. The offences such as sextortion for an instance overlooks the fear, trauma, PTSD and ongoing psychological distress of the victims. Secondly, the frameworks are designed only to curb and acknowledge creation, publication or dissemination of the unlawful content, that once the content is removed its repercussions vanishes. It is imperative to shift our regulations towards better perspective, the injuries or harm does not end with the act, they persist in the minds of victims, through endless circulation and constant possibility of the content resurfacing across multiple platforms. Due to this misalignments, courts are compelled to stretch the existing statues to address situations for which they were never meant to address. For example prosecuting perpetrators of deepfakes on obscenity or identity theft only solves one part of the problem but conveniently fails to capture the victim's experience, who may be subjected trauma, loss of autonomy, constant anxiety over their digital persona. Lastly, the current regulatory measures are mainly platform driven and ensures compliance through notice and takedown methods, it is not meant to handle with post effects of the digital exposure. There is a clear and evident absence in addressing non material harms. The courts are left with limited doctrinal measure to compensate for harms suffered.

Indian courts for an instance has been dealing with vicious deepfakes since few years now and acknowledges the disruptive potential of such AI-generated content, albeit within limited statutory confines. For an instance in the recent case of *Arijit Singh v. Codible Ventures LLP*, the Bombay High Court recognised that the singer's personality rights were infringed by AI-enabled voice cloning, synthetic recording and by generating fake videos to gain unauthorized millions of views. The Court noted that the singer was subjected to reputational injury and

commercial harms.[12] Similarly seen in the same sense in various cases of celebrities from Asha Bhosle, Anil Kapoor and the Rashmika Mandana cases. Though all these cases majorly focused on the parameters of economical and reputational harms. Now from perspective of recognising mental agony, there has been limited jurisprudence but an evolving one such as in the case of *Sukdeb Saha v. State of Andhra Pradesh*[13] has set a broader constitutional foundation where the Hon'ble Supreme Court had affirmed that under the 'right to life' Article 21 mental health is not only intrinsic but should be treated at par with other rights. The judgement gives rise to normative affirmation and constitutional validity to recognising mental health yet this thinking is under enforced or absent in the context of dealing with cybercrimes.

Courts similarly have ignored the nature of harms caused by nefarious deepfakes, they are endless and moreover they are qualitatively distinct from conventional cybercrimes. It is profoundly psychological and cannot be confined to be just reputational or financial. A growing number of interdisciplinary research shows that deepfakes has direct measurable emotional, cognitive and behavioural impacts. A comprehensive scoping review of 28 studies observes harms such as distress, reduced self-worth, anxiety, trust erosion in media and in some may cause behavioural changes such as social withdrawal. In particular victims of revenge pornography or deepfake pornography exhibits symptoms of severe emotional distress, analogous to PTSD, insomnia and intrusive memories. If observed more granularly studies of victims, studies highlights that they experience feelings of dehumanisation, violations, powerlessness and shame, which are often amplified by psychosomatic signs of high blood pressure, loss of appetite, etc. Severe cases victims also reports suicidal thoughts and their complete withdrawal from social and professional life. The feeling of helplessness due to inability to control such contents and living with constant fear that it might resurface further intensifies their trauma and suffering, in some instances, making them comparable and analogous to forms of sexual violence. [14]

Another dimension and in a devastatingly alarming way, deepfakes are heavily biased by gender, the domain of harm is very deliberated and discriminatory. Empirical data highlights that 96% of deepfake content are disproportionately target women and are by nature almost

[12] Arijit Singh v. Codible Ventures LLP, 2024 SCC OnLine Bom 2445
[13] Sukdeb Saha v. State of Andhra Pradesh, 2025 SCC Online Sc 1515
[14] Asher Flynn et al., *Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non-Consensually Created and Shared Sexualized Deepfake Imagery*, J. Interpersonal Violence (Sept. 9, 2025), https://doi.org/10.1177/08862605251368834.

always pornographic. The exploitation is staggering and massive in scale if looked that data that leading websites attract over 134 million views.[15] This is not a merely a question of misuse happening but it showcase the very dark and very real pattern of systematic digital sexual violence. Multiple studies links image-based sexual abuse (IBSA) to anxiety, depression, substance abuse and emotional distress.[16] In context of India, deepfakes has been weaponised to silence, intimidate and discredit the masses as seen in high profile cases public figures, artists such as Rashmika Mandana, journalists such as Rana Yubb, politicians have been subjected to sexual violence and targeted harassment campaigns.[17] These instances produces chilling effects, such as deterring participation in digital spaces. Cultural stigmas, discrediting professional credibility. Victims often face moral scrutiny, are questioned and subjected to victim-blaming thereby worsening the mental harm, which shifts the limelight from perpetrators to victims. Therefore the most apparent and gapping limitation of the current existing framework is entirely central to the act of offence and the offender and not on the harms suffered by the victims or for their rehabilitation.

Global discernible shift towards recognising psychological harm

Every country today face the challenges to keep in pace with rapid technological advances, but what set few jurisdictions apart is how they adapt, protect and tackle these ever evolving changes. Jurisdictions across the world for instance the United States, United Kingdom and the European Union proactively have begun to explicitly recognise the mental harms as a legally cognisable injury in the context of manipulated or synthesized content. The U.S in particular criminalises non-consensual intimate imagery under the TAKE IT DOWN Act, where the intent to cause mental distress has been mandated to be removed. Similarly, the U.K's Online Safety Act, 2023 goes further steps ahead by criminalising both the act of creation and dissemination of intimate or explicit deepfakes, the act expressly state that the intent to cause "alarm, distress or humiliation," thereby placing mental harms at the heart of addressing the offence. The European Union adopts the world's first comprehensive and structural approach to regulate AI through their Artificial Intelligence Act, 2024. The Article 5 of the Act

[15] Oscar Williams-Grut, *Deepfake Hype: Should We Really Be Worried?*, BBC News (Oct. 7, 2019), https://www.bbc.com/news/technology-49961089.
[16] Carmela Mento et al., *Psychological Violence in Image-Based Sexual Abuse (IBSA): The Role of Psychological Traits and Social Communications—A Narrative Review*, 12 Behav. Scis. 1, 4 (2025), https://pmc.ncbi.nlm.nih.gov/articles/PMC12428175/.
[17] Gaurav Yadav et al., *Psychological Trauma and Legal Challenges of Deepfake Technology*, 37 Scis. Conserv. & Archaeology, (2025)

specifically bans AI practices that are likely to cause psychological harms, which includes subliminal manipulation, distorting human behaviour and exploitation of vulnerabilities.[18] The AI Act works alongside and is linked with the EU Product Liability Directive which have mechanisms for compensation or damages for medically recognised harm to psychological health, inclusive of injuries not accompanied by physical harm as compensable damages. These are directly awarded to who have been victims of such malicious practices.[19]

CONCLUSION

We are in that day and age where the very foundation of trust built by media whether its print, digital or electronic is rapidly eroding. Now anybody with a basic subscription of generative AIs like ChatGPT or Claude and with minimal skills can manipulate, twist and fabricate the media, alter narratives and within a matter of minutes can destroy a person's livelihood or reputation, often shielded behind the layers of anonymity. In this context it becomes more concerning as at present the issue of deepfakes do not have any definitive solutions only stop-gaps. There is no reliable scientific methods or legal remedies to prevent the creation of malicious deepfakes, we are only dealing with its aftermaths. Whilst not all deepfakes are inherently malicious but more often than not, is synonymous with manipulative fabrication of content since the incumbent of AI, the pace at which AI has evolved it has outstripped the existing legal systems around the world to respond and deal with it meaningfully, although there has been gradual shift towards dealing with the harms caused by these content. Jurisdictions as highlighted above of the U.S, U.K and EU have shifted towards targeting the source of harm, rather than dealing with its consequences, their regulations demonstrates a clear and inevitable doctrinal evolution by giving recognition to non-material harms or injury and comprehensively dealing with digital harms.

India by contrast is yet to see such jurisprudential change, currently laws neither recognise mental harm as an injury nor does it provide effective remedial measures dedicated towards victims. There is growing need to evolve our legal responses that adapts to the unique and complex nature of such technological driven harms. If laws continues to adhere to rigid and traditional frameworks, it becomes a double edged sword, where victims will be forced to claim their remedies within ill-suited categories of laws, that neither captures the full extent of the

[18] Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), art. 5, 2024 O.J. (L).
[19] Council Directive 2024/2853, art. 4(6), 2024 O.J. (L) 2853 (EU).

injury nor the remedy provided is meaningful. Current frameworks continues to prioritise harms that are economical or reputational. There is no institutional pathway for victims of deepfake abuse to be mandatorily directed towards psychological rehabilitation. At best the current mental health services are voluntary and is often limited to one time interventions, despite the evidence showing how psychological issues require sustained interventions over months if not years. Moreover due to lack of compensation to victims they are burdened with the cost of therapies, medial treatment or counselling costs. Exacerbated by the existing penalty framework where the fines or penalties are absorbed by the government exchequer rather than using to build a support-based system for those affected. Furthermore limited institutional coordination from the authorities, particularly between the cybercrime authorities and the mental health services. The gap highlights a siloed approach between administrative departments and can be seen in existing mental health initiatives for an instance the tele-counselling service by Tele-MANAS is not directly connected to the current cybercrime redressals. Both the response systems work independently leaving victims without meaningful recovery

Therefore even though in theory our minds are constitutionally protected but still its harm remains undercompensated and unrecognised. A multi-pronged approach must be adopted for a meaningful response, the reforms must shift beyond mere removal of content and to victim-centric, recognising psychological harm as compensable injury, integrating better interdisciplinary redressals and establishing better inclusive awareness.

The Article 21 promises the right to dignity, but dignity cannot be realised without mental integrity. Deepfakes strikes at the core of our integrity, makes victims lose their autonomy, strips their confidence away, erodes the very trust reposed in technology and affect their societal relationships, not to forget the prolonged legal processes where patience is tested every day offer little solace. Thus, the scars left are even though invisible but are no less real and if our laws continues to ignore this, it risks offering incomplete justice.