BEYOND THE LABEL: BUILDING A FUTURE-READY INDIAN FRAMEWORK FOR DEEPFAKES, AI PORNOGRAPHY & SEXUAL IMAGE HARM

COMMENTS ON THE DRAFT AMENDMENTS TO INFORMATION TECHNOLOGY (INTERMEDIARY GUIDELINES AND DIGITAL MEDIA ETHICS CODE) RULES, 2021 [IN RELATION TO SYNTHETICALLY GENERATED INFORMATION]

Created by MAYA

Website: https://www.ourmaya.in | Email: anjana@ourmaya.in/ initiative.maya@gmail.com

TABLE OF CONTENTS

A. SUMMARY

'Beyond the Label: Building a Future-Ready Indian Framework for Deepfakes, AI Pornography & Sexual Image Harm' addresses India's approach to regulating synthetically generated information with a special focus on sexual crimes. While appreciating India's rights-based and preventive approach to synthetically generated information, we problematise not characterising the misuse of synthetically generated information to perpetrate sexual crimes. By undertaking a comparative approach with various jurisdictions across the world, we argue that while the Draft Amendment [hereinafter, "*Amendment*"] is a step in the right direction to prevent sexual crimes, more needs to be done. This includes defining non-consensual synthetic sexual imagery, constituting a liability architecture to make it more preventive in nature, having survivor-centric redressal mechanisms, introducing cross-border and inter-platform coordination, having digital rights safeguards that are sensitive to the gendered nature of sexual crimes, and having strict data and transparency mandates.

B. INTRODUCTION

[B.I.] Problem Statement

In brief: AI's rapid rise has enabled the easy creation of synthetic content, leading to severe misuse through **deepfakes** and **non-consensual sexual imagery that disproportionately harms women, children, and gender minorities**. Amid growing global concern, regulatory measures like the Draft Amendments to the IT Rules, 2021, are crucial to curb AI-facilitated gender-based violence and protect vulnerable groups.

AI or Artificial Intelligence has been taking over digital spaces like a storm. The Draft Amendments to IT Rules, 2021, rightfully address the urgent need to regulate synthetically generated information because, like every other technological tool created by humanity, AI comes with its own shortcomings, especially in its treatment of issues related to gender, gender-based violence, and sexual violence against gender minorities and children.

It is rather easy to generate synthetic data with the surge of AI; including images, texts, audio, and video clips. Data creation today is just a few taps away, with varied categories of image and multimedia generation easily accessible after a small investment. But with this ease comes misuse,[1] and this misuse impacts women

[1] Williams, Sophie, Jonas Schuett & Markus Anderljung, *On Regulating Downstream AI Developers*, 2 Eur. J. Risk Regul. (Aug. 4, 2025) https://link.springer.com/article/10.1007/s00146-024-02130-8.

and gender minorities[2] most disproportionately, with AI being used to generate sexually-suggestive images without consent, affecting women and children the most. Examples of such inappropriate material include deepfake images,[3] pornography,[4] and CSAM[5] (Child Sexual Abuse Material). AI has also been used as a weapon against gender minorities, being used for blackmail, extortion, and other such crimes. This is coupled with a rise of nudification apps like ClothOff, where people can 'undress girls for free' and 'undress anyone free of charge', based on a photograph.

As the AI wave takes the world by storm, there are growing global concerns and calls for regulatory action on the creation and circulation of deepfakes and other AI-generated content. A report by the Internet Watch Foundation has recognised the rapid rise of child sexual abuse material (CSAM) produced using AI as a growing and significant threat.[6] Data shows that **60%** of consumers have **encountered a deepfake video within the last year,**[7] while **human detection** of deepfake images averages only **62%** accuracy.[8] According to a 2024 study on technology-facilitated gender-based violence, **96%** of deepfakes are **nonconsensual**, and **99%** of deepfakes are of **women.**[9] Moreover, **100%** of the examined content on the **top five 'deepfake pornography websites'** was targeting **women.**[10] Globally, **57%** of women have experienced some sort of image-based abuse.[11] This means that not only is such content being **rapidly produced**, but it has also become harder to **detect**, leaving survivors vulnerable to harassment, exploitation, and abuse.

---

[2] Shrestha, Sunny & Sanchari Das, *Exploring Gender Biases in ML and AI Academic Research Through Systematic Literature Review*, 5 Front Artif Intell. 976838 (Oct. 11, 2022), https://pmc.ncbi.nlm.nih.gov/articles/PMC9593046/.

[3] NDTV, *AI Used To Create Celebrity Deepfakes In Multi-Million Dollar Instagram Scam*, NDTV (Oct. 4, 2025), https://www.ndtv.com/world-news/ai-used-to-create-celebrity-deepfakes-in-multi-million-dollar-instagram-scam-9393755.

[4] *IT Student Uses AI to Create Porn Pics of 36 Women Students, Suspended*, NDTV (Oct. 8, 2025), https://www.msn.com/en-in/news/India/it-student-uses-ai-to-create-porn-pics-of-36-women-students-suspended/ar-AA1O5oXQ.

[5] *Alarming Rise in AI-Produced Child Abuse Material, Warns Regulatory Body*, NDTV (Oct. 18, 2024), https://www.ndtv.com/world-news/alarming-rise-in-ai-produced-child-abuse-material-warns-regulatory-body-6817927. (ndtv.com)

[6] Internet Watch Foundation (IWF), *What Has Changed in the AI-CSAM Landscape? Report Update* (July 2024) (v.1.1), https://www.iwf.org.uk/media/opkpmx5q/iwf-ai-csam-report_update-public-jul24v11.pdf. (iwf.org.uk)

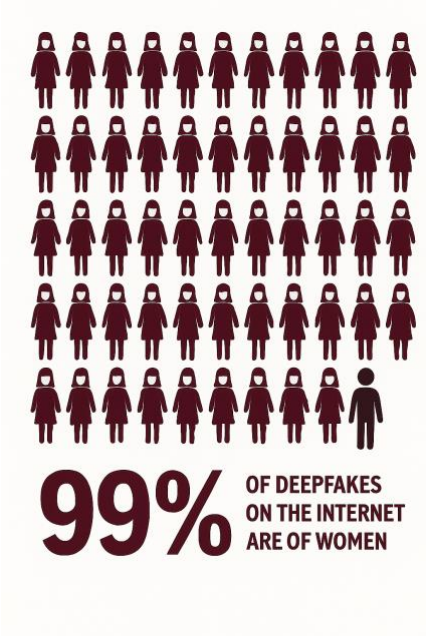
[7] Catherine Chipeta, *Deepfake Statistics (2025): 25 New Facts for CFOs*, Eftsure (May 29, 2025), https://www.eftsure.com/statistics/deepfake-statistics/.

[8] Chipeta, Catherine, *Deepfake Statistics (2025): 25 New Facts for CFOs*, Eftsure (May 29, 2025), https://www.eftsure.com/statistics/deepfake-statistics/. (eftsure)

[9] Kristine Baekgaard, *Technology-Facilitated Gender-Based Violence: An Emerging Issue in Women, Peace and Security* (June 2024), Georgetown Institute for Women, Peace and Security, https://giwps.georgetown.edu/wp-content/uploads/2024/06/Technology-Facilitated-Gender-Based-Violence.pdf. (giwps.georgetown.edu)

[10] Henry Ajder, Giorgio Patrini, Francesco Cavalli & Laurence Cullen, *The State of Deepfakes: Landscape, Threats, and Impact* (Sept. 2019), https://www.regmedia.co.uk/2019/10/08/deepfake_report.pdf.

[11] Cherie Blair Foundation for Women, *Online Gender-Based Violence Stifles Women's Digital Inclusion & Financial Freedom* (Jul. 2025), https://cherieblairfoundation.org/news-list/online-gbv/.



A recent press release by the Department of School Education & Literacy (DoSE&L) and the Ministry of Education revealed the government's plan to implement the use of AI in the curriculum of children from class 3 onwards.[12] This will lead to a boost in the use of AI among young children, which only increases the urgency for proper planning and execution of safety policies for children's protection against the children's victimisation by AI or its misuse. Childlight's report, the only global study of its type, reaffirms our claim. It highlights a 1,325% rise (2023-2024) in harmful AI-generated online abuse material, such as deep fakes placing real children's faces onto sexual images. The number rose from 4,700 reports logged by the National Center for Missing and Exploited Children in 2023 to over 67,000 in 2024.[13]

In light of these events, national governments have started to amend their laws to safeguard their citizens against any threats from AI-generated media. Generative AI has thus become a tool for sexual and gendered violence, and monitoring this is the need of the hour. Without equitable measures to ensure the safety of those most vulnerable, AI cannot be celebrated as a milestone.

[B.II.] Aim

Through these comments and suggestions, we wish to highlight the need for **specifically merging crimes related to synthetically generated information and sexual and gender-based minorities**. Since they are

[12] Press Information Bureau, "Curriculum on AI to be introduced in all schools from Class 3 onwards," Press Release No. 2184211, (Oct. 30, 2025) https://www.pib.gov.in/PressReleasePage.aspx?PRID=2184211.

[13] Childlight Global Child Safety Institute, *Study finds millions of children face sexual violence – and AI deepfakes surge is driving new harm* (Oct. 7, 2025) https://www.childlight.org/newsroom/study-finds-millions-of-children-face-sexual-violence-and-ai-deepfakes-surge-is-driving-new-harm.

more vulnerable to crimes related to synthetically generated information, we believe that the IT Rules (2021) can be better equipped to **prevent** and **prosecute** and **raise awareness** about them.

C. WHAT THE 2025 AMENDMENT DELIVERS AND ITS SIGNIFICANCE

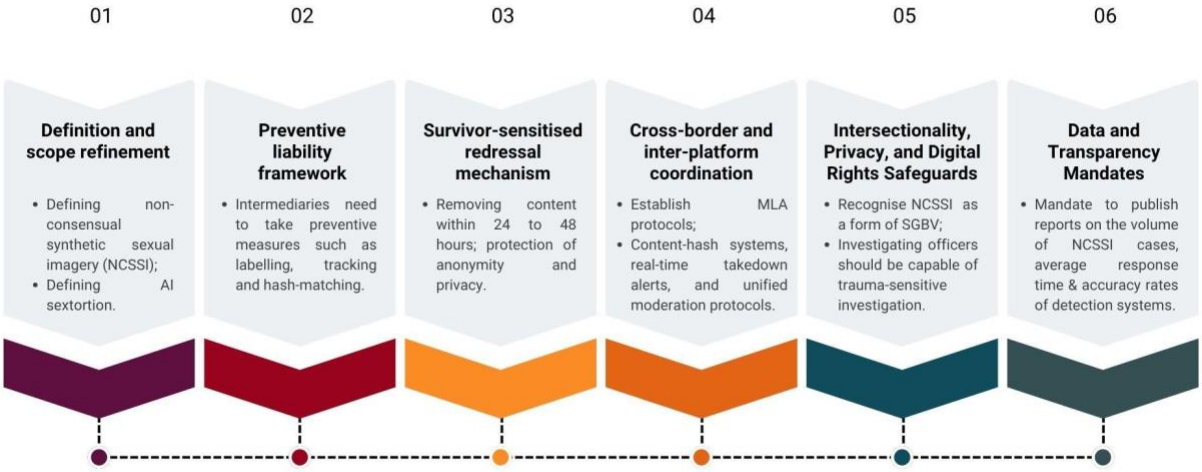
The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) follows up on the amendments of October 2022 and April 2023, reflecting the Government's continued effort to ensure an "open, safe, trusted and accountable internet" to its citizens. These proposed Amendments provide a legal basis for labelling, traceability, and accountability regarding synthetically generated information.

- The Amendment is the first of its kind to define synthetically generated information as "*information that is artificially or algorithmically created, generated, modified or altered by a computer source which makes the information seem authentic or true*".
- In order to increase the traceability of such information and to distinguish it from authentic data, it is proposed that such information be labelled or embedded with a unique metadata identifier.
- While uploading any content, the user has to declare any use of synthetically generated information via precise labelling. For audio content, such labelling should be present in the first 10% of the duration, whereas the label should cover at least 10% of the surface area for visual content. This could look similar to the declaration by content creators on their sponsored content.
- The role of social media intermediaries would be crucial in verifying the authenticity of data by flagging the synthetically generated ones. Such intermediaries are also restricted from removing, suppressing or modifying any such labels.

By acknowledging AI-generated media and deepfake content as material regulatory concerns and by bringing this into the ambit of the Rules, the government recognises the evolving nature of online risks and attempts a step in enhancing the privacy and safety of online users. By safeguarding the users' right to self-expression as well as keeping the intermediaries accountable for a secure digital environment, this Amendment aims to build public trust and enhance the security of a person's activity online.

D. FUTURE-ORIENTED GAPS & WHAT NEEDS TO BE BUILT

Multi-tier Approach to Address AI-Facilitated Sexual Violence



[D.I.] Definition & scope refinement

Proposed Amendment by the Ministry: The Ministry has defined synthetically-generated information under rule 2(1)(wa). It refers to "information which is artificially or algorithmically created, generated, modified or altered using a computer resource, in a manner that such information reasonably appears to be authentic or true".

Proposed recommendations:

- The Amendment must also clearly **define non-consensual synthetic sexual imagery (NCSSI)**. A proposed wording of the definition could read: "*Any synthetic, generated or digitally modified image, voice, video, or depiction that realistically portrays a person as nude, semi-nude, or engaged in sexual activity or sexually-suggestive conduct, created, shared, or stored without that person's consent*." Here, a "person" must extend beyond natural persons to bring fictional

characters under the ambit of the law. The Ministry must also criminalise the creation, sharing or storing of NCSSI.[14]

- The Amendment must also define **synthetic sexual extortion or AI sextortion**. This refers to the *creation or threat of creation or dissemination of synthetic sexual imagery to extort, blackmail, or coerce an individual into providing money, sexual acts, or personal information*. These definitions must be mentioned as a separate aggravated category to capture consent, harm, and intent.

Background/justification: The Draft Amendments to the Information Technology Rules, 2021, which introduce the term "synthetically generated information", mark an important step toward recognising the harms created by artificial intelligence. Yet, to address the structural, gendered, and transnational dimensions of deepfakes and AI-generated sexual image harm, the framework should move beyond definitional minimalism. **Defining technology-facilitated gender-based violence** and integrating it into regulatory frameworks is often viewed as one of the **primary steps toward ending it**.[15]

From the current wording of the definition, it becomes clear that the Amendment brings in non-consensual sexual imagery within the scope of synthetically generated information. This is also evident from the *Explanatory Note*, where the Government is encouraging accountability and traceability related to synthetically generated information. However, without a robust definition of non-consensual sexual synthetically generated information, it does not distinguish harm or intent. **This would lead to the law treating a "face-swap filter" and a "non-consensual sexual deepfake" as the same regulatory object.** For instance, the law will not be able to differentiate between a person using a filter on social media (for instance, Snapchat, where filters can make one look older/ younger) and a person using a filter to *nudify* a person. Intermediaries would risk **over-regulating benign uses (which stifles innovation)** while **under-regulating sexual violence**. Defining NCSSI would capture AI-generated "nudification apps," "fake porn videos," and "face-swap sexual deepfakes". These definitions could be interpreted in the broader context of intimate abuse via AI, AI impersonation, and consent fraud, such as using AI to falsify or manipulate indications of consent by generating content appearing to show a person consenting to a sexual act or uttering consent words that were never spoken.

Internationally, regulatory discourse now uses people-centred terminology such as "non-consensual synthetic sexual imagery," "revenge deepfake pornography," and "intimate image abuse via AI," which recognise the survivor and moral wrong rather than the technology itself. The United Kingdom's Online Safety Act (2023)[16] and subsequent Government announcement in January 2025 go further by explicitly

[14] The criminalisation of NCSSI has not been addressed in this Report, as the Amendment pertains solely to intermediary liability. Nevertheless, to effectively address and prevent NCSSI, it is essential to criminalise its creation, storage, and possession.

[15] Kristine Baekgaard, *Technology-Facilitated Gender-Based Violence: An Emerging Issue in Women, Peace and Security*, Georgetown Institute for Women, Peace and Security (2024) https://giwps.georgetown.edu/wp-content/uploads/2024/06/Technology-Facilitated-Gender-Based-Violence.pdf

[16] Online Safety Act 2023 (c. 50) UK
https://www.legislation.gov.uk/ukpga/2023/50

criminalising both the creation and sharing of sexually-explicit deepfake imagery.[17] Section 75 of Australia's Criminal Code Amendment (Deepfake Sexual Material) Act 2024, which allows survivors or the eSafety Commissioner to take legal action against non-consensual distribution of deepfake material and imposes a penalty of up to 7 years' imprisonment on those who create or share the media, is a significant step towards ensuring online safety.[18] The law also targets the source by considering the creator of the deepfake as a deterrent and imposing harsher penalties; a model that could be adopted in India to address the root cause of the harm. Furthermore, by enacting the Crimes Amendment (Intimate Images and Audio Material) Bill 2025, which imposes a penalty of up to 3 years' imprisonment, the New South Wales Government has made a significant stand against gender-based violence online.[19] This new law targets the creation, distribution, and even threat of sharing deepfake media, including audio.

The new offences also extend to those installing or maintaining equipment used to capture or generate intimate images without consent. Adopting such rights-based, preventive, and victim-centred approach and terminology in India would redirect attention to consent and control over one's digital embodiment rather than merely categorising content as synthetic.

[D.II.] Preventive Liability Framework

Proposed Amendment by the Ministry: The Amendment states that synthetically generated information must be prominently labelled or embedded with a permanent, unique metadata or identifier. This must cover at least ten per cent of the surface area of the visual display or, in the case of audio content, appear during the initial ten per cent of its duration. The intermediary shall not enable the modification, suppression or removal of such label, permanent unique metadata, or identifier. The intermediary shall also require the users to declare whether such information is synthetically generated, and shall deploy reasonable and appropriate technical measures to verify the accuracy of such declaration. The intermediary shall be deemed to have failed to exercise due diligence if the intermediary becomes aware, or is otherwise established, that the intermediary knowingly permitted, promoted, or failed to act upon such synthetically generated information in contravention of these rules.

Proposed recommendations:

- A forward-looking framework should impose varying responsibilities, requiring intermediaries to create **origin tools, watermarking systems**, and clear processes that prioritise the **survivor's consent and dignity**.

[17] "Government crackdown on explicit deepfakes", Ministry of Justice & Alex Davies-Jones, GOV.UK (7 Jan 2025) https://www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes

[18] "The Criminal Code Amendment (Deepfake Sexual Material) Act 2024: Policy Reform to Strengthen Online Safety in Australia," CODEA (2024) https://www.codea.com.au/publication/the-criminal-code-amendment-deepfake-sexual-material-act-2024-policy-reform-to-strengthen-online-safety-in-australia/

[19] "NSW Government strengthens protections against deepfakes and image-based abuse", New South Wales Department of Communities and Justice (19 Sept 2025) https://dcj.nsw.gov.au/news-and-media/media-releases/2025/nsw-government-strengthens-protections-against-deepfakes-and-ima.html

- Liability should not just depend on independent audits; **regular transparency reports** and accountability at the user level can enhance compliance.
- If anything seems synthetically generated but lacks the label by the user, the content must be labelled as '**suspected synthetic**'.

Case study: Hash-matching

A preventive measure could be a "hash-matching" system for intimate or sexually explicit imagery. Platforms like Meta and Reddit already use similar tools to combat child sexual abuse material. When a survivor reports a non-consensual sexual image or deepfake, the platform converts that image into a unique digital "hash," or a fingerprint of the file. This hash is then added to a shared industry database. If anyone attempts to upload the same or a visually similar image again, the platform automatically blocks it before it goes live. This prevents the re-circulation and re-victimisation that typically happens with deepfake sexual abuse. Such a system would make intermediary compliance preventive rather than reactive, while also protecting survivors' privacy since only the hash is stored, not the actual image.

Intermediaries should be required to use AI-supported hash-matching for non-consensual intimate imagery, establishing an early barrier before harm can occur.

Background/justification: Intermediaries must follow Section 79[20] and remove harmful content once they receive notice, as confirmed in Super Cassettes Industries Ltd vs. Myspace Inc. & Another (2011).[21] However, the rapid rise of AI-generated media, which can lead to harassment, defamation, and blackmail, has created enforcement gaps.

The liability rules under the Amendment are mainly reactive and procedural, whereas it must also focus on prevention and the needs of survivors. While platforms must take down flagged content and notify users, **they do not have to create systems that identify or trace harmful synthetic images before they spread**.

For example, China has updated its regulations from 2022 and 2023 to focus on labelling and tracking as a way to identify and disclose deepfakes.[22] This mandates all platforms in China to add a watermark or embedded metadata to any data created or altered by AI. According to the new Amendment, anything that cannot be traced back to a synthetically created will be labelled '**suspected synthetic**' for users. The Digital Services Act of the European Union also provides a relevant example that requires quick responses to illegal content and transparency in algorithms.[23]

[20] Information Technology Act, 2000, s. 79

[21] Super Cassetes Industries Ltd. v. Myspace Inc. & Another, CS(OS)2682/2008

[22] "Measures for Labeling of AI-Generated/Synthetic Content," China Law Translate (2025) https://www.chinalawtranslate.com/en/ai-labeling/

[23] Regulation (EU) 2022/2065 of the European Parliament and of the Council (Digital Services Act), OJ L 277

[D.III.] Survivor-sensitised Redressal Mechanisms

Proposed Amendment by the Ministry: While the Amendment seeks to strengthen due diligence by social media intermediaries and significant social media intermediaries (through takedowns and warnings), it does not mention any measures that can be adopted to ensure that the redressal process considers the survivors' trauma.

Proposed recommendations: Intermediaries must be required to take up the following steps to ensure that their grievance redressal mechanism is survivor-sensitised:

- Intermediaries should be legally required to create **dedicated support systems**;
- Intermediaries must remove content within **24 to 48 hours** and block reported material **across platforms** using shared hash databases;
- Intermediaries must protect the **anonymity and privacy of survivors**. They should allow reports **without requiring repeated verification** or causing further trauma. Survivors should have the option to report violations through **anonymous systems** that keep their identities safe from the perpetrator and minimise distress; and
- Once the investigation has been completed, the survivors' data must not be stored by the intermediary.

Example clause of Survivor Confidentiality and Privacy

The intermediary shall ensure that in cases involving non-consensual intimate imagery, deepfake or synthetically generated sexually explicit content, the identity, confidentiality, and privacy of the person depicted shall be protected at all stages of complaint handling, takedown, and redressal. This shall also apply to all persons involved in handling the information, including law enforcement.

Explanation: For the purposes of this clause—
(a) The intermediary shall not disclose or publish any personally identifying information of the affected individual without their explicit consent;
(b) The intermediary shall ensure that any communication with the complainant is conducted through secure and confidential channels;
(c) The intermediary shall, upon verification of such content, take all reasonable steps to prevent its further circulation, including but not limited to, content hashing, blocking of derivative uploads, and coordination with other intermediaries or platforms for cross-platform removal;
(d) The intermediary shall maintain records of all such actions taken in a manner that does not compromise the complainant's privacy.

The Ministry must also conduct digital literacy campaigns as a preventive tool, as well as psychosocial support and local accessibility (for instance, digital literacy must be taught in different languages).

Background/justification: A common experience among survivors of deepfake pornography is that even after their material is removed from one platform, it reappears elsewhere. This leads to ongoing humiliation and distress. India's intermediary liability framework needs to go beyond simple content removal. The model from the Australian eSafety Commissioner also shows how confidential reporting, guided support, and victim-centred processes can help build trust and make support accessible.

The lack of sensitised mechanisms is one of the biggest barriers survivors face while wanting to report technology-facilitated gender-based violence.[24] **Survivors face slut-shaming and outright dismissal** when they approach authorities to report technology-facilitated gender-based violence.[25] In order to ensure that this Amendment benefits the maximum number of survivors, the Ministry must **train** and **sensitise** authorities (for instance, police and judiciary) on how to approach these cases from a **trauma-informed lens**.

[D.IV.] Cross-Border and Inter-Platform Coordination

Proposed Amendment by the Ministry: The scope of the Amendment remains confined to domestic intermediary compliance and user-level labelling mechanisms within India. The Amendment does not introduce any mechanism for enforcing Indian takedown orders on platforms hosted abroad, nor does it establish interoperability among intermediaries to prevent the same deepfake from being re-uploaded across multiple digital spaces.

Proposed recommendations: To strengthen the regulatory effectiveness of the amendment, it is recommended that specific clauses be added to address cross-border cooperation and inter-platform coordination in cases of AI-generated sexual harms.

- The Rules should mandate the creation of a **shared national content-hash database**, linked with both the National Cybercrime Reporting Portal (NCRP) and SAHYOG Portal, which will record and track all flagged instances of deepfake pornography or related NCSSI content. Such a database should be interoperable with verified intermediaries and designed to issue real-time takedown alerts across platforms.
- Once flagged and confirmed as non-consensual synthetic content, the same should be **automatically delisted from all participating digital platforms**. This will ensure that harmful material, once detected, cannot be redistributed or monetised elsewhere.
- Furthermore, the amendment should provide for **mutual legal assistance protocols** and bilateral or multilateral data-sharing memoranda with foreign regulators, enabling India's takedown orders to be recognised and implemented globally.

[24] Sippy, T., Hembram, U., Durani, A., and Chandiramani, S. (January 2023). "Institutional responses to digital harms: A case study of India's public and private institutional response mechanisms to tech-facilitated gender-based violence", Working Paper 02. Mumbai: Artha Global.

[25] Gurumurthy, A. (2022). On Cybercrime Against Women and its Larger Causes. Accessed via: https://itforchange.net/on-cybercrime-against-women-and-its-larger-causes

- **Collaboration with international AI safety and digital governance initiatives**, particularly under the Bletchley Declaration (2023),[26] should be institutionalised, promoting responsible AI governance that extends to regulating NCSSI and related transnational harms.

At the inter-platform level, the amendment should mandate participation in a cross-platform coordination network supported by **shared content-hash systems, AI-assisted traceability tools, and human moderation oversight**. This mechanism would ensure consistency in moderation standards and facilitate joint accountability. These technical and legal provisions must be accompanied by capacity-building measures, including the training of cybercrime units, digital forensic experts, and law enforcement personnel in handling synthetic media and NCSSI cases. This holistic approach would operationalise a trauma-informed, survivor-centric model of online safety, ensuring rapid redressal and effective deterrence across borders.

Background/ justification: Deepfake pornography and other AI-generated sexual harms are inherently **transnational** in nature.[27] These forms of abuse often originate, circulate, and are monetised across **multiple jurisdictions**, transcending the scope of any single nation's regulatory authority. The same synthetic image that violates Indian law may be hosted on servers abroad or shared through intermediaries headquartered outside the country, beyond the enforcement reach of domestic agencies. As a result, even when platforms remove such content in compliance with Indian takedown requests, the same material can quickly reappear on foreign-hosted platforms, creating a **cycle of re-victimisation and digital impunity**.

This becomes particularly evident in cases such as that of Noelle Martin, whose social media images were digitally altered into explicit pornographic material and circulated globally across platforms, including Reddit, X (formerly Twitter), and pornographic sites such as MrDeepFakes and Fan-Topia.[28] Despite repeated takedown attempts, the content continually resurfaced across interconnected platforms, revealing the futility of fragmented, country-specific interventions.[29] Similarly, in Assam's "Babydoll Archi" case, a woman's private photographs were used by her former partner to create an AI-generated pornographic persona that gained over a million followers on Instagram before the deception was uncovered. Despite the perpetrator's arrest, the deepfake content continued to circulate across multiple platforms, including those hosted abroad, underscoring the porous and transnational nature of AI-enabled sexual violations.[30] These incidents indicate how deepfake pornography operates as part of a fluid global network of digital abuse, resistant to unilateral legal or technological control.

[26] The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023 (Nov. 1, 2023), available at the-bletchley-declaration-ai-safety-summit PDF (www.astrid-online.it).
[27] Felipe Romero Moreno, Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content, 38 Int'l Rev. L. Comput. & Tech. 297, 309 (2024), https://doi.org/10.1080/13600869.2024.2324540.
[28] Ruby Harris, How It Feels to Find Your Face Photoshopped Onto Internet Porn, VICE (Apr. 17, 2019), Accessed via: How it Feels to Find Your Face Photoshopped Onto Internet Porn.
[29] Kat Tenbarge, Found Through Google, Bought with Visa and Mastercard: Inside the Deepfake Porn Economy, NBC News (Mar. 27, 2023), Accessed via:The deepfake AI porn industry is operating in plain sight.
[30] Geeta Pandey, Deepfake Deception: Indian Woman's Identity Stolen for Erotic AI Content, BBC News (July 23, 2025), Accessed via: https://www.bbc.com/news/articles/cn0znk47x9eo.

An example of a reactive law is the 'TAKE IT DOWN' Act (Tools to Address Known Exploitation by Immobilising Technological Deepfakes on Websites and Networks Act),[31] which was passed in the US in May 2025. This law considers using non-consensual intimate images a crime, with violation leading to monetary fines and up to three years of imprisonment. Although the law requires platforms to remove flagged content within 48 hours, it doesn't prevent the content from being uploaded to foreign-hosted platforms. This renders laws meant to address deepfakes as ineffective.

[D.V.] Intersectionality, Privacy, and Digital Rights Safeguards

Proposed Amendment by the Ministry: The Amendment remains limited in its attention to the intersectional and gendered nature of technology-facilitated violence. It lacks an explicit recognition that AI-generated sexual abuse reproduces and amplifies existing hierarchies of gender, caste, class, and sexuality. Furthermore, there is insufficient articulation of privacy safeguards and survivor protections, despite the growing dependence on AI-based detection systems that often entail surveillance risks. The Ministry's draft does not yet include a non-discrimination clause ensuring equal protection or mandate that investigation and moderation processes be sensitive to gender and intersectional vulnerabilities. Nor does it provide guidance for embedding privacy-by-design standards or data protection principles within technological interventions.

Proposed recommendations:

- The Amendment should include an explicit non-discrimination clause guaranteeing that no person is denied protection or redress based on gender, caste, class, sexuality, disability, or other identity markers.
- It should further specify that all enforcement, investigation, and content moderation processes be conducted in a manner sensitive to gender, privacy, and intersectional vulnerabilities.
- Recognising that AI-generated sexual abuse is a form of technology-facilitated gender-based violence (TFGBV), the Amendment can include language such as "*with due regard to the gendered nature of harm arising from technology-facilitated abuse*" or "*ensuring trauma-informed, survivor-centric, and consent-based investigation and redressal*." The Amendment should also incorporate privacy and digital rights safeguards to prevent overreach in the name of regulation. Provisions may specify that all technological interventions comply with data minimisation, purpose limitation, and informed consent principles, and that content moderation and detection systems follow privacy-by-design standards. This alignment with the Digital Personal Data Protection Act (2023)[32] and the

[31] "TAKE IT DOWN Act: Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act" (119th Cong., 2025–2026) https://www.congress.gov/bill/119th-congress/senate-bill/146

[32] The Digital Personal Data Protection Act, No. 22 of 2023, INDIA GAZETTE(Aug. 11, 2023). 2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf.

subsequent DPDP Rules (2025)[33] would ensure that technological responses do not compromise privacy while addressing AI-driven abuse.

- Further, the Amendment should define NCSSI as a serious digital offence and mandate the establishment of Digital Gender Units within cybercrime cells staffed by trained officers capable of trauma-sensitive investigation.
- The Amendment should also require platforms to conduct Gender and Digital Impact Assessments before deploying AI-based moderation or recommendation systems, ensuring these technologies neither reinforce bias nor ignore intersectional harm.
- Drawing from existing consent-management frameworks such as TRAI's Digital Consent Management framework (2025)[34] for telecom and banking sectors, and opt-out mechanisms in generative AI platforms, the Digital Consent Registry could also be explored as a future policy innovation. It would link individuals, platforms, and regulators by embedding consent at the creation stage rather than relying only on post-reporting interventions. However, its design would require careful deliberation on issues such as interoperability across platforms, privacy protection, and cross-border enforceability. As a forward action item, the government could initiate multi-stakeholder consultations with technology experts, social media intermediaries, and civil society organisations to assess its legal and technical feasibility. It is therefore recommended that the Amendment or subsequent policy instruments include a pilot framework for a Digital Consent Registry, accompanied by privacy-by-design standards and grievance redressal mechanisms, to operationalise informed and revocable consent within digital ecosystems.
- To remain inclusive, reporting systems must be available in multiple regional languages, supported by local legal aid, and designed with the participation of marginalised groups to reflect the full diversity of digital experiences.

Background/ justification: The UNPFA (2024)[35] defines technology-facilitated gender-based violence (TFGBV) as digital forms of abuse that reproduce existing gender hierarchies and power imbalances. AI-generated sexual abuse exemplifies this, as synthetic imagery and deepfakes disproportionately target women, and gender-diverse persons, causing psychological, social, and reputational harms that reinforces gendered vulnerability. UN Women (2024)[36] notes that between 16 and 58 per cent of women and girls globally have experienced online violence, with the most extreme and frequent forms reported among young women, LGBTQIA+ persons, women with disabilities, and racialised, caste-oppressed, and migrant groups. Such intersectional risks demonstrate that digital harms are magnified for those already facing

[33] Ministry of Electronics & Information Technology, Draft Digital Personal Data Protection Rules, 2025, G.S.R. 02(E) (Jan. 3, 2025), INDIA GAZETTE

[34] Press Information Bureau, Ministry of Communications, TRAI Launches Pilot Project for Digital Consent Management in Partnership with RBI and Banks (June 16, 2025), available at https://www.pib.gov.in/PressReleasePage.aspx?PRID=2136728.

[35] United Nations Population Fund (UNFPA), An Infographic Guide to Technology-Facilitated Gender-Based Violence (2024), available at https://www.unfpa.org/sites/default/files/pub-pdf/An%20Infographic%20Guide%20to%20An%20Infographic%20Guide%20to%20TFGBV.pdf.

[36] UN Women, *Creating Safe Digital Spaces Free of Trolls, Doxing, and Hate Speech* (June 28, 2024), https://www.unwomen.org/en/articles/explainer/creating-safe-digital-spaces-free-of-trolls-doxing-and-hate-speech.

structural exclusion. Feminist scholars such as Sharmila Rege, Nivedita Menon, and Gopal Guru[37] have shown that sexual violence enforces power hierarchies rather than moral codes. Deepfakes extend these hierarchies by weaponising visibility and manufacturing humiliation. They are not merely obscene but disciplinary; designed to silence, humiliate, and control. Any legal response should, therefore, recognise non-consensual synthetic sexual imagery as a form of gender-based and sexual abuse and treat it as an aggravated digital offence.

[D.VI.] Data and Transparency Mandates

Proposed Amendment by the Ministry: The Amendment stops short of mandating detailed disclosures on the nature, frequency, and resolution of synthetic sexual or gender-based imagery cases. The Ministry's draft does not yet include standardised requirements for reporting detection accuracy, takedown response time, or algorithmic bias audits. Moreover, it lacks a data-driven and gender-sensitive approach that integrates existing national and state-level initiatives. The Amendment's framework also remains limited to procedural compliance, without incorporating independent social or third-party audits that evaluate the broader social consequences of moderation failures.

Proposed recommendations:

- Platforms should be mandated to publish quarterly transparency reports detailing the volume of synthetic sexual or gender-based imagery cases detected and removed, average response times for takedown, and accuracy rates of detection systems. These disclosures, consistent with the recommendations of the CERT-In Advisory (2024),[38] would promote algorithmic transparency and fairness.
- Independent audits should be instituted to assess whether moderation systems disproportionately target or neglect specific communities, ensuring that content governance practices uphold equality and digital rights.
- Building on international ethical frameworks such as UNESCO's Ethical Impact Assessment (EIA) and the Readiness Assessment Methodology (RAM),[39] the Amendment could require gender-sensitive AI impact assessments and bias audits as part of the intermediary liability framework.

[37] Sunaina Arya, Dalit or Brahmanical Patriarchy? Rethinking Indian Feminism, 1 CASTE: A Global Journal on Social Exclusion 7 (2020), https://doi.org/10.26812/caste.v1i1.54.

[38] Indian Computer Emergency Response Team (CERT-In), Advisory CIAD-2024-0060: Deepfakes - Threats and Countermeasures (Nov. 27, 2024), https://www.cert-in.org.in/s2cMainServlet?pageid=PUBVLNOTES02&VLCODE=CIAD-2024-0060.

[39] UNESCO, Women for Ethical AI: Outlook Study on Artificial Intelligence and Gender (preprint for discussion at the W4EAI Conference, Oct. 30, 2024, UNESCO Headquarters, Paris, France) (SHS/REI/EAI/W4EAI/2024/Outlook), https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000391719&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_306db80f-6a44-4db4-a0f7-35fde8eebb52%3F_%3D391719eng.pdf.

- Platforms should also undergo periodic third-party and social audits to evaluate the broader accountability of their systems, recognising that the harm from NCSSI extends beyond legal violation to encompass deep social and psychological consequences.
- The Amendment should additionally link the National Cybercrime Reporting Portal (NCRP), SAHYOG Portal, and state-level cybercrime units to establish a unified, real-time database that can track complaints and ensure faster inter-agency coordination. Drawing inspiration from state initiatives like Kerala's CyberDome, the Counter Child Sexual Exploitation Cell, and Telangana's Women Safety Wing, the Ministry can replicate these as scalable public-private models of prompt detection and coordinated response.
- Finally, transparency measures must be accompanied by education and cultural change. The Amendment should mandate digital literacy programmes, gender-sensitivity training for law enforcement, and public awareness campaigns promoting respect for digital consent and autonomy.

Background/justification: Effective transparency mechanisms are indispensable for building public trust and ensuring accountable governance in the digital sphere. The National Cybercrime Reporting Portal (NCRP) documented a **118.4% increase in online crimes against women between 2020 and 2024**, under categories including **sexual abuse content, obscene material, and online exploitation**.[40] In the same document, reported cases under "*Online and Social Media related Crimes*" on NCRP have risen sharply from **21,589** in 2020 to **1,56,938** in 2024 (**626%** increase), with significant increases in impersonation, identity theft, and cyberstalking, reflecting the expanding scale of digital gender-based abuse.[41]

The recent NCRB 2023 data further feeds into this trend.[42] Given below is a table comparing data from NCRB 2022 and NCRB 2023. The crimes mentioned are likely to increase with the advent of AI, which calls for data-driven approaches to intermediary liability.

Category	NCRB 2022	NCRB 2023
Cyber Crimes/ Information Technology Act	2940	3678 (25% increase)
Publishing or Transmitting of Sexually Explicit Material	2251	2767
Other Women Centric Cyber Crimes (for instance, blackmailing/ defamation/ morphing/ creating a fake profile	689	911

Alarmingly, **1,143** cases under the IT Act were verified as true but could not be pursued due to insufficient evidence, untraceable offenders, or lack of actionable leads. These figures unravel that digital gender-based violence has grown in both frequency and complexity, necessitating data-driven coordination across

[40]Government of India, Ministry of Home Affairs, Lok Sabha Unstarred Question No. 2944 to be answered on 18th March 2025/Phalguna 27, 1946 (Saka): Cybercrime Against Women, at 1–7 (Mar. 18, 2025), 2944.pdf.
[41] Government of India, Ministry of Home Affairs, Lok Sabha Unstarred Question No. 2944 to be answered on 18th March 2025/Phalguna 27, 1946 (Saka): Cybercrime Against Women, at 1–7 (Mar. 18, 2025), 2944.pdf.
[42] Nat'l Crime Recs. Bureau, Crime in India 2023, Part I (Statistics), Vol. I (Ministry of Home Affairs, Gov't of India), https://www.ncrb.gov.in/uploads/files/1CrimeinIndia2023PartI1.pdf.

institutional levels. Introducing data privacy obligations would move beyond mere procedural compliance and embed fairness, non-discrimination, and inclusivity in the technological architecture of online governance.

E. CONCLUSION

The 2025 Amendment marks an important milestone in India's journey toward responsible AI governance, offering a crucial regulatory foundation to address possible harms, especially those linked to deepfakes and generative AI's sexual imagery. Yet, this legal step has a long way to go. The real transformation lies in building an ecosystem that moves beyond reactive takedowns to proactive, systemic preventions. Tackling AI-enabled abuse demands an integrated framework encompassing prevention, detection, liability, survivor protection, and robust rights safeguards.

Given below is a concise summary of our recommendations:

1. **Definition and scope refinement** - The Amendment must define non-consensual synthetic sexual imagery and AI sextortion to explicitly bring AI-facilitated sexual violence into its ambit;
2. **Preventive liability framework** - The Amendment must state that Intermediaries must take more preventive measures, such as labelling, tracking and hash-matching;
3. **Victim-centric redressal mechanisms** - Intermediaries must take down any sexually-violative content in 24-48 hours while protecting the anonymity and privacy of the complainant;
4. **Cross-border and inter-platform coordination** - The Ministry must establish Mutual Legal Assistance protocols; and introduce content-hash systems, real-time takedown alerts and have unified moderation procedures;
5. **Intersectionality, privacy, and digital rights safeguard** - The IT Act must recognise non-consensual synthetic sexual imagery as a form of sexual or gender-based violence; and investigators must be capable of conducting trauma-sensitive investigations; and
6. **Data and transparency mandates** - The Amendment must mandate Intermediaries to publish reports on the volume of NCSSI cases, average response time and accuracy rates of their systems which detect synthetically generated information.

For children, women, and marginalised groups, who face disproportionate exposure to digital harms, the challenge is not only technological but structural. Regulations must therefore embed principles of equity and protection into every layer of AI deployment. This includes mandating safety-by-design standards, enforcing platform accountability, and ensuring that survivors have accessible and timely redressal mechanisms. At the same time, AI systems that process or generate sensitive data must adhere to privacy-by-default norms and transparent governance structures to prevent overreach. At the same time, oversight mechanisms should be designed to protect free expression, ensuring that regulatory zeal does not devolve into censorship or surveillance.

Though the proposed solution will be better in taking precautions, India, with its unique social and digital ecosystem, has to account for the groups who are more vulnerable to technology-based violence. Hence, it should move from a harm-based approach that focuses on prevention to a user-based approach where the interests of the most affected groups are protected. A gender-neutral law might be dangerous to India's current socio-political milieu. As we can see, China focuses on accountability from the platforms,

Australian law targets the liability on the individual, and American law focuses on criminal liability for platforms. A mix of the instructional model along with precautionary measures might work the best for India.

Ultimately, the test of India's AI regulatory evolution will lie in how well it balances innovation with safety, and freedom with accountability. The next two to three years will be decisive: the frameworks we build today will determine whether the digital public sphere becomes safer, more equitable, and rights-respecting, or deepens existing inequalities. By centering children, women, and marginalized communities in policy design, India can set a global benchmark for inclusive AI regulation - one that ensures technology serves as a tool of empowerment, not exploitation. The Amendment is a beginning; the vision ahead must be a humane, just, and transparent AI ecosystem for all.

ABOUT MAYA

MAYA is a student-run organisation that aims to raise awareness about sexual violence. By breaking barriers and raising awareness about sexual violence, we work to empower every child to recognise and report it. We have collaborated with over **2000** students across **30+** schools, colleges, orphanages and residences and educated them about sexual harassment, online abuse and how to tackle it. To make mental health more accessible, we have also spearheaded a free therapy program, providing responsible and accurate mental health resources. We have released two easy-access guides on understanding child sexual violence and how to respond to a disclosure of it. In the past, we have collaborated with organisations such as Rotaract and IDIA. You may know more about us through our website.

ACKNOWLEDGEMENTS

Drafted by:

- Anjana Palamand, Founder, MAYA
- Joyotri Nandy, Founder's Associate, MAYA
- Arya Antherjanam V, Research Associate, MAYA
- Urbi Bhandary, Research Associate, MAYA
- Gopika Jayakumar, Research Intern, MAYA

Reviewed by:

- Preethi Kavilikatta, LL.M. (UCLA, NALSAR), LL.B., B.M.M (Journalism)
- Prof. (Dr.) Nachiketa Mittal, Registrar & Professor of Law, National Law University, Tripura
- Paakhhi Garg, Director, World Cyber Security Forum

CONTACT

Email: anjana@ourmaya.in/ initiatve.maya@gmail.com
Phone: +91 9008865558
Website: https://www.ourmaya.in