

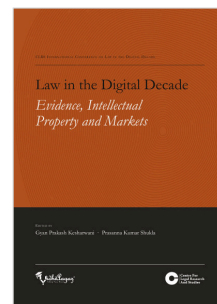
CHAPTER 7

Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations

Sharon Thomas and Soorya Narayanan

in *Law in the Digital Decade: Evidence, Intellectual Property and Markets*, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla (VidhiAagaz 2026), pp. 59–68

DOI <https://doi.org/10.63108/VAB.LDD.2.7>



CITATIONS

BLUEBOOK (21ST ED.)

Sharon Thomas & Soorya Narayanan, *Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations*, in *LAW IN THE DIGITAL DECADE: EVIDENCE, INTELLECTUAL PROPERTY AND MARKETS* 59 (Gyan Prakash Kesharwani & Prasanna Kumar Shukla eds., VidhiAagaz 2026).

OSCOLA (4TH ED.)

Sharon Thomas and Soorya Narayanan, ‘Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations’ in Gyan Prakash Kesharwani and Prasanna Kumar Shukla (eds), *Law in the Digital Decade: Evidence, Intellectual Property and Markets* (VidhiAagaz 2026) 59.

CHICAGO (17TH ED.)

Thomas, Sharon, and Soorya Narayanan. “Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations.” In *Law in the Digital Decade: Evidence, Intellectual Property and Markets*, edited by Gyan Prakash Kesharwani and Prasanna Kumar Shukla, 59–68. Maihar: VidhiAagaz, 2026. <https://doi.org/10.63108/VAB.LDD.2.7>.

APA (7TH ED.)

Thomas, S., & Narayanan, S. (2026). Artificial abuse: A socio-legal analysis on the effect of usage of generative artificial intelligence against vulnerable populations. In G. P. Kesharwani & P. K. Shukla (Eds.), *Law in the digital decade: Evidence, intellectual property and markets* (pp. 59–68). VidhiAagaz. <https://doi.org/10.63108/VAB.LDD.2.7>

SOURCE

<https://doi.org/10.63108/VAB.LDD.2.7>

LICENCE

Open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0): it may be copied, shared and adapted for non-commercial purposes, with credit to the authors and the source.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

CHAPTER 7

Artificial Abuse: A Socio-Legal Analysis on the Effect of Usage of Generative Artificial Intelligence Against Vulnerable Populations

SHARON THOMAS* AND SOORYA NARAYANAN**

The rapid increase in the reliance on Artificial Intelligence in everyday life has resulted in the creation of a new kind of gendered harm, in the form of exploitative materials created using Generative Artificial Intelligence (A.I.). This has led to an expansion of existing gaps between technological development and legal protection. This paper examines the intersection of images produced using A.I. tools and legal remedies available for the same, with a specific focus on how these tools are used as another medium to harass women. It analyses how A.I. exacerbates structural inequalities and highlights the critical need for reforms in the current Indian criminal justice system. Several forms of online harassment that are already faced by individuals such as stalking, image alteration through creation of deepfakes, sexual harassment, nonconsensual recording etc. are adding to the psychological and sexual abuse against women, making their commission easier and their prosecution harder. We also explore the technological gaps in the creation of these A.I. tools, especially the gender bias in training A.I., and its effect on the vulnerable populations of India where caste, class, religion, and disability intensify the vulnerability for women. The psychological and economic impact on the victims including trauma, digital withdrawal, job loss, reputational harm, and barriers to online participation are understood through case studies. Finally, the paper examines the legal framework addressing the digital abuse against women by comparing international legislations like the European Union A.I. Act and United Nations guidelines with the Indian legislations and identifies gaps in the Bharatiya Nyaya Sanhita, Information Technology Act, and Digital Personal Data Protection Act. The paper concludes by proposing recommendations to make AI governance gender-responsive and intersectional.

Keywords: Artificial Intelligence, Generative AI, Deepfakes, Gender-based Violence, Discrimination.

1 INTRODUCTION

The world is entering into the era of Artificial Intelligence (AI), where technology has been rapidly integrated into various aspects of everyday life and systems. The scale of development and training of these AI models, especially Generative AI models, has moved faster than regulators can address. This has led to the creation of statutory and regulatory vacuums which have allowed the harm caused by AI systems to vulnerable populations to slip through the regimes. These vulnerable populations include women, children, individuals belonging to the LGBTQIA+ community, individuals belonging to poor communities and more. Through this paper, we explore the disproportionate harm caused to them by AI systems, the systematic reinforcement of discrimination due to the way these models are created and trained, real-life examples of how these tools have actively harmed individuals, existing statutory frameworks and their

* Student at VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.

** Student at VIT School of Law, Vellore Institute of Technology, Chennai, Tamil Nadu, India.

scope and final recommendations to create a safer framework to utilize these tools in a way that reduces the harm to these populations.

2 UNDERSTANDING AI AS A TOOL FOR GENDERED VIOLENCE

Artificial Intelligence can be understood as the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. Generative AI technology is a subset of the same and refers to technology that is trained on large amounts of data that can generate content in the form of text, audio, video etc. in response to ‘prompts’ given to them. Rapid technological advances in the field of AI are creating unprecedented risks to facilitate violence against women. Technology-facilitated violence against women and girls (TF VAWG) is comparable to AI-assisted and AI-generated digital violence, but the generative AI models have greater access to data that make them more harmful and realistic. The humongous scale and unpredictability of AI models enable them to cause significantly more harm. The added layer of anonymity of the perpetrators makes access to justice even harder.

Common forms of online TF VAWG are mostly exacerbated by the usage of AI tools, which increases the harm that is caused by them. The volume of media that is created through Generative AI tools has obscured the distinction between genuine information and generated information, which is utilized to increase malicious targeted disinformation to a greater audience. The promulgation of misinformation is also easier as the AI models are trained on datasets that are inherently biased against women, reinforcing gender stereotypes. The non-consensual sharing of personal information is made much easier through AI tools, which can identify vulnerabilities and exploit them when interacting with female victims.

Gender inequality in society is reflected completely in these AI systems and is embedded in every part of their life cycle, from the people developing the AI to who accesses it and finally to who regulates it. A study conducted in 2021 analyzing 133 AI systems across various industries found that about 44.2% (59) of them showed gender bias, and 25.7% (34) exhibited both gender and racial bias¹. When the foundation of AI runs entirely on data and information that is given to it, the systems are built to perpetuate harm from their inception. The underrepresentation of women systematically in AI research and development influences the data used, priorities undertaken and evaluation of the systems. The structural challenges faced by women, which include lack of digital access, limited financial ability, unequal education, care responsibilities and lack of relevant training, restrict their ability to even use these AI systems which have quickly become part of everyday life. Women risk becoming passive users or completely excluded from AI opportunities, reinforcing the digital inequalities that facilitate Technology-facilitated Gender-based Violence (TFGBV). Without safeguards, these systems risk deepening inequality².

¹ G. Smith and I. Rustagi. 2021. When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. Stanford Social Innovation Review. 31 March

² Jinha Kim and Byeogjo Kong 2025 Bridging the Digital Divide: Harnessing Artificial Intelligence for Gender Equality in Asia and the Pacific ADB Brief October

3 FORMS OF AI ABUSE AGAINST WOMEN

3.1 Deepfakes and Non-Consensual Imagery

A deepfake is a digital image, video, voice, etc. that has been convincingly generated or altered to misrepresent someone as doing or saying something that was not actually done or said. Deepfakes are primarily made by GANs (Generative Adversarial Networks) which is an artificial intelligence system where two neural networks compete against one another to form hyper-realistic synthetic data³. A GAN system works through multiple competitive loops where the outputs which create synthetic images are repeated till it achieves the most accuracy that a common man would not be able to differentiate from a real image⁴. However, the increase in access to the use of AI has made it easy for non-tech users who do not have special expertise in computer or AI systems to be able to generate these hyper-realistic images within seconds, with a much larger number of people now using these tools for this purpose.

Another form of TFGBV is non-consensual Synthetic Intimate Imagery (NCSII), commonly known as revenge porn, which reveals a structural form of electronic sexual violence that does not comply with traditional sexual assault laws that exist making it difficult to handle such cases⁵. Historically, most non-consensual sharing of sexually explicit images happens within intimate relationships where the photos could've been taken with consent but shared against the will of the victim. With the rise of Generative AI and easier access to synthetic images, the earlier requirement of physically capturing the photo isn't necessary, and almost anyone can fall victim to this form of gendered violence. These deepfakes can be created with just an innocent headshot of a person, and any other images which can be easily found online.

The widespread access to deepfake technology affects public personalities and ordinary people alike. One of the earliest cases in India which made people aware of the dangers of AI technology was when the actress Rashmika Mandanna's face was digitally superimposed onto an explicit video of another woman⁶. The incident which occurred in November of 2023, soon after Gen AI was made accessible to the public, sparked a huge debate on the dangers of it, and led law enforcement to act against the perpetrator under the Information Technology Act (IT Act) and the Indian Penal Code (IPC). In January 2024, the 24-year-old accused was arrested for online sexual harassment. These cases do not happen only against high-profile public figures. Schools, universities, and workplaces often have cases where male workers use non-consensual synthetic imagery as a weapon to humiliate, silence, or blackmail their female peers⁷. These images spread quickly throughout the internet, ruining the reputation of the victims and affecting their everyday personal and professional lives, both online and offline.

³ Preeti, Kumar, M. and Sharma, H.K. (2023) 'A gan-based model of deepfake detection in social media', *Procedia Computer Science*, 218, pp. 2153–2162.

⁴ Ramanizan, B.N. (2025) 'Generative Adversarial Networks (gans): A comprehensive review of architectures, training challenges, and advancements', *Journal of Artificial Intelligence and Capsule Networks*, 7(4), pp. 304–319.

⁵ Preeti, Kumar, M. and Sharma, H.K. (2023) 'A gan-based model of deepfake detection in social media', *Procedia Computer Science*, 218, pp. 2153–2162.

⁶ Mankermi, S. (2023, November 8). *Etimes explainer: The truth behind Rashmika Mandanna's deepfake video, decoded with violation of privacy rules, legal implications and technological drawbacks*. Hindi Movie News. [etimes.in](https://www.etimes.in)

⁷ Karagianni, A. and Doh, M. (2024). A feminist legal analysis of non-consensual sexualized deepfakes: contextualizing its impact as AI-generated image-based violence under EU law. *Porn Studies*, pp.1–18.

3.2 AI-Enabled Stalking and Surveillance

Modern stalking tools are being developed by utilising predictive AI where a person's routines can be tracked through location data and analyzing past movements to forecast possible future outcomes. Perpetrators can use cheap tracking devices which they can attach to the victim's personal belongings or vehicles and connect them to AI models which analyse the data and provide real life updates on the location along with future predictions⁸. This allows stalkers to predict when their victims would be most vulnerable, to physically confront or harm them.

Modern AI has also easily taken advantage of facial recognition technology. For example, PimEyes is one of several facial-search services reported to index images scraped, without the consent of the people pictured, from public websites and social media where even one picture of the person can reveal their full identity along with personal details⁹. It isn't necessary for there to be a picture of the woman available on the internet as an individual can even capture an image of a woman in public without her knowledge and use that picture to track her down. Such searches can reveal employment histories, locations, family members or regularly visited public areas of the victim, threatening their safety and privacy.

3.3 AI in Intimate Partner Violence

Domestic violence advocates have begun to document regular cases where abusive partners use security cameras, smart door locks and other AI home technology to automatically lock doors trapping their partner with them or manipulate security alarms to control and isolate the victim¹⁰. These physical forms of interruption with AI tools can lead to coercive physical and emotional forms of control, and abuse, with very minimal force. This method of tracking a partner's domestic life leads to a severe lack of privacy leading to psychological distress.

Abusers utilize smart devices along with AI technology to manipulate their victim's perception of reality, easily manipulating and dominating them¹¹. They can alter the victim's digital records of plans in their calendar, delete home security camera footage, and use AI voice modification to manipulate conversations which creates an extremely stressful and confusing environment for the victim to navigate. These forms of manipulation are easy to commit but hard to prove, reducing the already slim chances for the victim to get justice. This form of AI tracking and surveillance further increases the chances for physical and sexual violence since the abuser has full knowledge of their daily routines and can discard the evidence. Constant surveillance also reduces the likelihood for the victims to plan an escape route as all their movements are tracked. Victims often fear that their abusers could have non-consensually recorded or generated sexually explicit videos or images that they could publicize to shame and control them even further.

⁸ Nameeta Rana (2024) 'Artificial Intelligence, cyber stalking, and the future of Digital Privacy: A Legal Perspective', *International Journal For Multidisciplinary Research*, 6(4).

⁹ Wakefield, J. (2020). PimEyes facial recognition website "could be used by stalkers." [online] BBC News, 11 June. Available at: <https://www.bbc.com/news/technology-53007510>.

¹⁰ Brown, A., Harkin, D. and Tanczer, L.M. (2024) 'Safeguarding the "internet of things" for victim-survivors of domestic and family violence: Anticipating exploitative use and encouraging safety-by-design', *Violence Against Women*, 31(5), pp. 1039–1062.

¹¹ Peter Novitzky. (2023). A systematic review of ethical challenges and opportunities of addressing domestic violence with ai-technologies and online tools. *Heliyon*, 9(3–4).

3.4 Non-Consensual Recording through AI Devices

The creation and distribution of AI hardware products such as camera-equipped smart glasses, including Meta's AI glasses, whose small cameras and microphones may go unnoticed by others, can enable people to record others in public without their consent. The increasing use of Meta AI glasses has raised concerns regarding violation of privacy and bodily autonomy¹². Because such devices can look like ordinary accessories, they may be misused to violate someone's privacy. This is a violation of Article 21 of the Indian Constitution¹³, and section 77 of the Bharatiya Nyaya Sanhita (BNS)¹⁴ which handles voyeurism, that is, the non-consensual recording or sharing of private videos or pictures taken without consent. Despite this, due to the elusive nature of the AI devices, it is highly improbable that the perpetrators are caught. The concern for safety escalates with devices that are attached to real-time facial recognition software and LLMs. Researchers have found that these devices allow the perpetrators to access all the private information about a person with the help of AI¹⁵.

4 SOCIAL AND INTERSECTIONAL ISSUES WITH AI ABUSE

4.1 Gender Bias and Structural Discrimination in AI Systems

AI is trained on available data on the internet, including historical statistics and structures. When LLMs consume vast amounts of data from the internet, they also collect and learn the biased discriminatory systems and practices followed historically, which leads to algorithmic bias. AI models¹⁶ thus end up learning sexism, racism, classism, transphobia, and other forms of discriminatory information. For example, when an AI tool is told to generate a person in a leadership position, it would generate images of men. In contrast, when it is told to generate a person in a caregiving job, it would generate a woman¹⁷. This bias affects job opportunities for women since AI data would show preference for male workers over female workers in most jobs, aggravating the existing issues of gender inequality in workplaces.

4.2 Intersectionality

To understand intersectional bias within AI, it is important to first understand the intersectional framework where overlapping systems of discrimination cause a layered form of oppression. Kimberlé Crenshaw, a legal scholar and an intersectional feminist, showed how structures of marginalization overlap, making individual experiences different. She highlighted¹⁸ how a white woman and a black woman would have different forms of discrimination that they must overcome, and how Black women can face discrimination on the combined grounds of race and sex that neither a gender-only nor a race-only analysis

¹² Suraksha P. (2026, June 20). Meta AI glasses pilot faces privacy scrutiny. *The Economic Times*.

¹³ Constitution of India, art. 21.

¹⁴ Bharatiya Nyaya Sanhita, 77 (2023).

¹⁵ Kaira Software (n.d.). AI privacy in crisis: 5 Scandals that expose the dark side of artificial intelligence. [online] Kaira Software. Available at: <https://kairasoftware.com/blog/ai-privacy-in-crisis-5-scandals-that-expose-the-dark-side-of-artificial-intelligence> [Accessed 22 Sept. 2026].

¹⁶ Nivedhaa N, (2024) 'A COMPREHENSIVE REVIEW OF AI'S DEPENDENCE ON DATA', *International Journal of Artificial Intelligence and Data Science (IJADS)*, 1(1), pp. 1–11.

¹⁷ Bhardwaj, A. (2026, June 23). AI is already rewriting reality, but getting women wrong, says UN Women. *Business Standard*. https://www.business-standard.com/technology/artificial-intelligence/ai-is-already-rewriting-reality-but-getting-women-wrong-says-un-women-126062300887_1.html

¹⁸ K. Crenshaw, (1991) 'Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color', *Stanford Law Review*, 43(6), pp. 1241–1299.

captures. Similarly, a woman's sexuality, class, caste, etc. can intensify the form of violence, harassment, or discrimination she faces¹⁹. This is applicable to Artificial Intelligence, as it feeds off online data and learns the same biases, leading to marginalized women such as those from minority religions, women with disabilities, Dalit and Adivasi women, and women from the LGBTQIA+ community often facing targeted algorithmic attacks and abuse online.

4.3 Psychological and Economic Impacts

Survivors of non-consensual deepfakes and AI-surveillance based stalking face severe psychological trauma, similar to the clinical mental impact caused by physical sexual assault and stalking²⁰. The realistic quality of modern AI and its generative images create severe emotional distress and anxiety among survivors of such attacks. Victims have reported that this form of AI crime has taken away their bodily autonomy and often get clinically depressed and develop Post Traumatic Stress Disorder²¹. To cope with this online abuse, victims often have no choice but to log off from all social media accounts and withdraw from public and professional areas to avoid hate, judgement, and scrutiny²². This withdrawal in turn leads to the removal of women from public media, and their voices being silenced and suppressed. Additionally, there are several economic impacts on women who face AI-harassment. Once their deepfakes become easily publicly accessible to anyone, employers may feel hesitant in letting the woman work in the company due to fear of public and moral perceptions, and may terminate her employment. Victims may end up losing their jobs over something they did not partake in, aggravating their trauma²³. Further, in abusive relationships where the partner uses AI tools to monitor the woman's financial and banking tools, it becomes difficult for the victim to have economic independence. These structural issues risk widening the gender pay gap and make it more difficult for women in leadership positions to succeed.

5 LEGAL AND REGULATORY FRAMEWORKS

5.1 International Legislative Responses

The European Union's Artificial Intelligence Act or the EU AI Act (2024)²⁴ establishes a thorough and risk-based regulatory system that sorts AI usage into different and specific risk tiers based on the potential harm it would cause to society. It puts an explicit ban on any AI use that poses an unacceptable risk, such as extracting and scraping of facial images from CCTV footage or content from the internet to develop facial recognition databases, and biometric classification systems that deduce or infer a person's race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. The EU AI Act enforces strict obligations on transparency for AI generated content to tackle synthetic manipulation.

The United Kingdom's Online Safety Act (2023)²⁵ acknowledges technology-based violence and abuse and includes criminal penalties for offenders. The Act makes the act of sharing non-consensual

¹⁹ Prince Kumar. (2026). Intersections of caste and gender in social stratification. Sociology. Institute.

²⁰ Goldberg, C. (2017). Nobody's victim: Fighting harassment online and off.

²¹ Muhammad Umar. (n.d.). Digital trauma: Deepfake victimisation and AI-generated violence. The Lancet Psychiatry, 13.

²² Ali, M. (2025). Deepfakes and victimology: Exploring the impact of digital manipulation on victims. Substantive Justice: International Journal of Law, 8.

²³ Bradley, D. (2026). Deepfakes, job losses, opaque models: Exploring the dark side of AI. Tech Explore.

²⁴ European Union's Artificial Intelligence Act or the EU AI Act (2024)

²⁵ United Kingdom's Online Safety Act (2023)

intimate media a punishable crime. It shuts down several legal loopholes regarding online harassment and abuse that allowed perpetrators to walk free without facing any sanctions. The Act also imposed a statutory duty of care for all digital media. Digital platforms are legally mandated to notice and remove non-consensual sexual content. The penalty for non-compliance is heavy, ensuring strong enforcement of the provisions.

The United States' legal responses to AI technology have been developed through federal and state level measures. The DEFIANCE Act – Disrupt Explicit Forged Images and Non-Consensual Edits Act²⁶, would establish a civil cause of action for victims. This act would grant the victims of explicit deepfakes the right and opportunity to sue the creator of the image, the people who distributed it, and anyone else who facilitates the crime for monetary damages. Several individual states like California, Texas, and New York have made criminal and civil statutes that punish non-consensual image creators and distributors.

International organizations have established certain ethical standards over the years for national guidance when drafting legislation on Artificial Intelligence. The UNESCO Recommendation on Ethics of AI (2021)²⁷ requests its member states to create special provisions and resources to reduce the gender biases in AI training databases and to make clear standards regarding accountability for gender-based violence or harassment through AI. The OECD Recommendation on Artificial Intelligence (2019)²⁸, furthermore, stresses human-centric values and transparency in the way AI is consumed and used. Although these international frameworks provide a good base for the way AI law is legally handled, their non-binding nature limits their usage across countries.

5.2 Indian Legal Framework

The basis of the laws addressing gender-bias in India is the Constitution. Article 14²⁹ of the Constitution guarantees equality in the eyes of law and provides the foundation to challenge and question systemic inequalities including gender-based discrimination or violence. Article 19(1)(a)³⁰ protects the freedom of speech and expression of individuals, which may be affected for victims of AI-based violence or non-consensual image creation as they are the ones who bear the consequences of such acts. The Supreme Court case of Justice K.S. Puttaswamy Vs. Union of India (2017)³¹ declared the importance of personal dignity and rights over bodily autonomy as fundamentally protected under Article 21 of the Constitution³². Non-consensually morphing or generating an explicit image or video of someone directly infringes the ruling set by this court.

The Information Technology Act (2000), which is the primary Indian law for cybercrime, contains provisions for digital harassment. These provisions came into existence before the widespread usage of Generative AI and are therefore outdated in some scenarios. However, some provisions can still be relied upon such as Section 66E³³ of the act, which criminalizes intentional capture, publication, or transfer of

²⁶ Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2025, S. 1837, 119th Congress (passed by the US Senate on 13 January 2026; not yet passed by the House of Representatives).

²⁷ Recommendation on the Ethics of Artificial Intelligence, 41 C (2021).

²⁸ Recommendation of the council on artificial intelligence (May 22, 2019)

²⁹ Constitution of India, art. 14.

³⁰ Constitution of India, art. 19(1)(a).

³¹ Justice K.S. Puttaswamy v. Union of India [2017].

³² Constitution of India, art. 21.

³³ The Information Technology Act, 2000, s. 66E.

images taken of private body parts without consent. This provision can be used for non-consensual recording with Meta AI glasses but will not apply to generated images. Sections 67 and 67A punish publication and transfer of obscene or sexually explicit media³⁴. These sections do not mention the capturing of images and therefore can be applied to AI generated images as well.

The Bharatiya Nyaya Sanhita (2023) updated several gender-based provisions in criminal law. Section 78 of the act penalizes physical and online stalking of a person³⁵. Monitoring a woman's movements or electronic communication without her consent is punishable under this section. Section 79 punishes anyone who tries to insult the modesty of a woman through words, actions, and gestures³⁶. This would include creating and sharing non-consensual explicit images. Section 356 of the act handles defamation, i.e., false actions or words that will cause damage to someone's reputation³⁷. That section can be applied to cases where a woman's morphed or AI-generated images are created and shared without her consent, holding the offenders criminally liable.

The Digital Personal Data Protection Act (2023) requires owners of online websites to process the personal data of an individual only after receiving explicit consent. Yet, the act also contains exceptions that instantly weaken the effectiveness of the provision for non-consensual deepfake content. Under section 3(c)(ii) of the act, it exempts protection for personal data that is made publicly available by the individual themselves³⁸. This means that women who post public images for professional or social purposes can have their images used by Gen-AI users with bad intent and still be left without protection.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 makes it mandatory³⁹ for websites to take down non-consensual sexually explicit content within 24 hours of receiving a complaint. However, this is not followed in the strictest sense due to several restrictions, like the rapid increase in such content, low staff to handle such issues, and difficulty keeping track of them. Moreover, non-consensual digital content spreads rapidly among several different networks, websites, apps, etc. which may not always fall under India's legal jurisdiction.

6 WAY FORWARD

6.1 Ethics in Design and Technical Watermarking

To reduce the AI-assisted and enabled abuse that is committed against vulnerable communities, software developers must construct safety standards for the foundational structure of the AI systems and ensure that they are followed. These standards would mandate digital watermarks to accentuate AI-generated content⁴⁰, and the integration of cryptographic systems like the C2PA protocol into all generative platforms⁴¹. Cryptographic data lets web browsers, forensic tools, and content platforms verify whether media was artificially generated and automatically filters the content to ensure only intended recipients will

³⁴ The Information Technology Act, 2000, ss. 67, 67A.

³⁵ Bharatiya Nyaya Sanhita, 2023, s. 78.

³⁶ Bharatiya Nyaya Sanhita, 2023, s. 79.

³⁷ Bharatiya Nyaya Sanhita, 2023, s. 356.

³⁸ Digital Personal Data Protection Act, 2023, s. 3(c)(ii).

³⁹ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021.

⁴⁰ Tiwari, S. (2026). Mandatory watermarking and labelling of ai-generated content

⁴¹ Lisa Braswick (2026). C2PA explained - content credentials guide. AFIP.

access it⁴². AI developers must also implement strict boundaries for the training models which makes it possible for them to reject prompts regarding non-consensual nude image generations or face-swapping without verification⁴³. Incorporating such safety measures into the software models themselves would lower the possibility of digital harassment.

6.2 Liability of AI Platforms and Statutory Duty of Care

Relying only on individual reports to address harms is insufficient due to the high spread of non-consensual synthetic content on digital platforms. Social media regulations must establish a statutory duty of care on all social media platforms and hold them accountable for negligence if they fail to take down the non-consensual deepfakes or morphed images. Digital platforms must also implement automatic algorithms that will be able to identify, block, and report non-consensual imagery before they are uploaded. Statutory guidelines can also reduce mandatory takedown time frames from 24 hours to 2 hours for non-consensual sexually explicit content as they can spread quickly and cause irredeemable damage. Platforms that fail to follow these regulations should face legal disputes and must pay fines or penalties for damages.

6.3 Legislative Reform for India and Victim-Centric Approach

India must enact an act for AI safety and governance as it still does not have a specific framework for these tools. The act must focus specifically on regulating automated systems and protecting individuals from AI-enabled abuse. This act must define, criminalize, and punish non-consensual creation, possession, or distribution of sexually explicit content. The courts must ensure that a victim-centric approach is followed throughout the justice system to reduce their trauma and stress. The state should make the reporting process easier and more accessible for victims, where they can make one online report which must immediately trigger the takedown of the content across all online platforms and track down the perpetrator. Governments must also provide special services for the victims to help with their trauma by offering them psychological care free of cost, support groups, and safety services for anyone whose well-being was threatened after reporting. Special cyber-crime units with staff trained at understanding gender sensitivity and privacy rights and AI abuse are necessary for helping the victims recover faster.

7 CONCLUSION

The development of technology provides opportunities for the growth of society at large; however, this development must not be at the expense of those who are most vulnerable to the harm caused by these same technological developments. Understanding how Artificial Intelligence can be used as a tool to facilitate violence against marginalized communities ensures that further technological development is followed by appropriate regulations. By understanding the real-life harm caused by these virtual tools, Artificial Intelligence development can be done more holistically, by including all relevant stakeholders who are likely to be affected by its large-scale deployment across sectors. Despite the existence of multiple statutes that can be used to govern Artificial Intelligence in India, there are still significant statutory gaps

⁴² Latha, S. (2024). AI-Driven identification of cryptographic algorithms from encrypted data. *International Journal of Research in Technology and Innovation (IJRTI)*.

⁴³ Trent, A. (2026). *Ethical AI image generation: Best practices for developers*.

that necessitate the creation of a framework to regulate, supervise and govern Artificial Intelligence in the country.

* * *