

DIGITAL MEDIA AND DISCRIMINATION: IS ADVENT OF ARTIFICIAL INTELLIGENCE EXACERBATING GENDER DISPARITY?

Debasrita Choudhury *

ABSTRACT

The Constitution of India encompasses a wide range of negative liberties and enjoyment of rights free from any incursions. In Article 19(1)(a), citizens are provided the right of free speech and expression while Article 14 provides the equal protection of all. Besides Article 15(3) empowers the State to make any special provisions to facilitate women further. Apart from the fundamental rights, Part IV of the Constitution having Directive Principles of State Policy urges the State to uphold equality between men and women. However, there is a pervasive nature of disparity between the genders. Positive freedom is lacking when it comes to guaranteeing women's liberation since persistent, socially sanctioned discrimination against them exists even in the digital age and the period of Artificial Intelligence (referred to as AI). Rather, that disparity has only increased.

* Advocate, Calcutta High Court, West Bengal.

INTRODUCTION

Bonnie A Nardie, in her book titled, “My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft”⁹³ speaks on the different treatment meted out to male and female players in the massively popular multiplayer online game. She challenges the depiction of female characters in the game, as well as its hypermasculine culture. There are very few female players, and those that do tend to be subversive and tolerant of the seeming normality of the arena dominated by men. The author suggests that one of the reasons for it is that men, who are typically in their 20s and 30s, expressed themselves in the game in a way that would not be appropriate in any other setting.⁹⁴ Overtly sexist behaviour is considered the norm that attest to a player’s masculinity. The choice of female players, who are a glaring minority, is only restricted to tolerating this kind of discourse and engage minimally. According to many interviews with female players, they tend to overlook it since participation in the majority of the competitive groups in the game demands a "high abuse tolerance," which players must accept in order to join.⁹⁵

John Stuart Mill in his “On Liberty”⁹⁶ explains the harm principle that a person’s liberty can be curtailed only if it causes harm to others. This apparent garb of hypermasculinity is harmful for the society at large. Male players' right to free speech ought to end at the point when decency dictates otherwise. A lot of impressionable young men play these kinds of games. Despite being a fantasy game, it still sets a negative precedent for the permissible bounds of hypermasculine behaviour. Furthermore, it suggests, a false assumption, that men can only be who they really are in these settings when they act in such a way. This is a fine line to tread upon considering civility and the law, whether it's appealing to young minds or giving them an outlet to vent their frustrations of life.

The counter argument of creative liberty falls short when it causes long term widespread harm. One such example could be the Congress member Ranjeet Ranjan attacking Sandeep Reddy Vanga's film "Animal" in a sharply worded statement during Zero Hour on December

⁹³ Bonnie A. Nardie, “My Life as a Night Elf Priest Book Subtitle: An Anthropological Account of World of Warcraft”, *University of Michigan Press*, Digital culture books, available at: <https://www.jstor.org/stable/j.ctvndv9nw.12>

⁹⁴ Ibid

⁹⁵ Supra at 5

⁹⁶ Piers Norris Turner, “Harm and Mill’s Harm Principle”, 124, No. 2 *The University of Chicago Press, Ethics*, 299-326 (2014).

7, 2023, in the Rajya Sabha.⁹⁷ She said that the way sexism and violence are portrayed and justified is "shameful" and a "disease to society." She voiced her profound worry. Several well-known individuals have expressed concern over how women are portrayed in mass-media, generally young male-oriented commercial films. The adverse effect that such a representation in mainstream media may have, is shown by the fact that a Rajya Sabha MP had to denounce it in Parliament.⁹⁸

SOCIAL MEDIA AND STALKING

As more nations throughout the globe are embracing the digital age, changes are observed in the portrayal of women in digital media. Restrictive gender-based portrayals continue to exist in a variety of circumstances, notwithstanding advancements in civil rights. In the media, stereotypes, objectification, and sexualization of women are still prevalent. Exposure to representations that perpetuate stereotypes strengthens gender norms and encourages sexism and violence.⁹⁹ Representations that objectify and sexualize causes the internalization of beauty standards, adopting sexist viewpoints, or tolerating abuse. Such portrayals have the potential to have a detrimental effect on one's physical and mental health.¹⁰⁰

Social media discourses and narratives that are gendered are greatly influenced by female influencers. Their pictures influence how the general public views them and help shape social norms and expectations.¹⁰¹ Women are still underrepresented in conventional and digital journalism, despite the democratizing potential of digital media. Merely 26% of those included in media news tweets and online news pieces are female, and a mere 4% question gender norms.¹⁰²

⁹⁷ 'Animal', 2023 Hindi film revolves around a notorious son of an industrialist who vows to take revenge on those threatening his father's life. It depicts female characters in a very subservient manner, glorifying toxic masculinity. Its Box Office collection was Rs. 917.82 Crores.

⁹⁸ Shemin Joy, "Negative Review for 'Animal' in Rajya Sabha; Cong MP calls Film 'Disease' to Society" *Deccan Herald*, December 7, 2023, available at: <https://www.deccanherald.com/india/negative-review-for-animal-in-rajya-sabha-cong-mp-flays-film-for-promoting-violence-misogyny-2801308> (last accessed on 16 October 2024)

⁹⁹ Coyne, S. M., Linder, J. R., et al. "It's a bird! It's a plane! It's a gender stereotype!: Longitudinal associations between superhero viewing and gender stereotyped play." 70 *Sex Roles*, 416 (2014).

¹⁰⁰ Fabrizio Santoniccolo, Tommaso Trombetta, et al. "Gender and Media Representations: A Review of the Literature on Gender Stereotypes, Objectification and Sexualization", 20 (10) *International journal of environmental research and public health* 5770 (2023) doi:10.3390/ijerph20105770

¹⁰¹ Roberti G, "Female influencers: Analyzing the Social Media Representation of Female Subjectivity in Italy", 7 *Front. Sociol.* 1024043 (2022) doi: 10.3389/fsoc.2022.1024043

¹⁰² "Visualizing the data: Women's representation in society", *UN Women*, published in February, 2020, available at: <https://www.unwomen.org/en/digital-library/multimedia/2020/2/infographic-visualizing-the-data-womens-representation> (last accessed on 15 October 2024)

The media's gender imbalance feeds negative preconceptions. To improve women's visibility and accurate depiction in the media, consistent efforts must be made.

In terms of gender equality, social media has both advantages and disadvantages. Stereotypes are reinforced and there are disparate depictions. Social media has the power to reinforce pre-existing gender norms.

One of the pre-dominant issues of the day is online harassment. Online abuse and harassment disproportionately target women. This may reduce agency of individuality by silencing women's voices and discouraging from engaging in public discourse.

Social media has the potential to bring about positive change in providing women a forum to talk about their experiences, dispel myths, and encourage societal change. It might be an instrument for elevating voices, if used as such.¹⁰³

However, the banes outweigh the boons. Harassment as well as stalking has increased manifold through social media with perpetrators having the garb of anonymity.¹⁰⁴

The word "cyberstalking" refers to a digital "attack" on a person who has been picked out for dominance, revenge, or fury. Cyberstalking may manifest itself in a variety of ways, including demeaning, harassing, and shaming the victim.

In the Covid 19 lockdown, there was a surge of such cases. Section 78 in Bharatiya Nyaya Sanhita, (hereinafter referred to as BNS) provides the primary definition for this word. It stipulates the penalties for stalking, which also include cyberstalking. Men are charged with this offence if they deliberately pursue a woman, contact her, or attempt to get in touch with her despite the woman showing disinterest.

The first known instance of cyberstalking in India occurred in 2000, when a lady lodged a complaint about being harassed via chat by a man who was harassing her on call and he had shared her contact information with others who perpetrated the same. Police located the defendant and filed charges against him under Section 79 of BNS for violating a woman's right to modesty. This Section, however, solely addresses statements or deeds meant to

¹⁰³ He, L., Firdaus, A., Gong, J. *et al.* "How the social media impact women's psychological well-being in the patriarchal structure? The moderating effect of social capital", 24 *BMC Public Health* 581 (2024). Available at: <https://doi.org/10.1186/s12889-024-18013-y> (last accessed on 16 October 2024)

¹⁰⁴ According to NCRB Data, in 2021, 10,730 cases were reported by women and these crimes mostly include cyber blackmail, threatening, cyber pornography, cyber-stalking, bullying.

offend modesty; activities carried out online are not under its scope. Therefore, more exacting laws had to be implemented. The Information Technology Act, 2000 in India contains additional measures to stop crimes of digital media and cyberstalking.

"Punishment for sending offensive messages through communication service, etc." is covered under Section 66A of the Act. According to this Section, "anyone who sends, by means of a computer resource or a communication device, (a) any information that is grossly offensive or has menacing character; or (b) any information which he knows to be false, but for the purpose of causing annoyance, inconvenience, danger, obstruction, insult, injury, criminal intimidation, enmity, hatred or ill will, persistently by making use of such computer resource or a communication device; (c) any electronic mail or electronic mail message with the intent to cause annoyance or inconvenience or to deceive or to mislead the addressee or recipient about the origin of such messages, shall be punished with a term which may extend to three years of imprisonment and a fine."

Section 66E of the Act addresses voyeurism on the internet, which is often the outcome of cyberstalking and imposes penalties for it. The dissemination of pornographic materials that are lascivious in character and cater to the prurient interest in electronic form is covered under Section 67. Penalties for the offence include a maximum 5-year jail sentence and a fine of up to Rs.10 lakh. Penalties for communicating explicit sexual material involving minors are outlined in Sections 67A and 67B. Section 72 of the IT Act, which addresses the disclosure of personal information or violation of confidentiality, is the provision most often utilised.

In the case of *Yogesh Prabhu v. State of Maharashtra*¹⁰⁵, two individuals were having an online discussion when the man proposed marriage to the woman which she declined broke off all communication, deleting him from her social media "friends list." The perpetrator continued to pursue her, sending explicit movies to her email address from unidentified IDs. The offender, was tried under Section 79 of BNS and 66E of the Information Technology Act, 2008 since Section 354D (Section 78 BNS) could not be used retroactively to a 2009 offence as it was enacted in 2013.¹⁰⁶

¹⁰⁵ *Yogesh Prabhu Vs State of Maharashtra*, 2006 (3) MhLJ 691

¹⁰⁶ Ibid

Social networking has made such crimes easier to commit. Via social media, these offences can be quickly carried out while disguising identities. Therefore, it is difficult to find the culprit. It has let in a torrential stream of cybercrimes, most of which target women and children as their most susceptible demographics.¹⁰⁷ AI has brought up a multitude of new kinds of crimes against women committed through digital platforms.

The Supreme Court issued guidelines in the 2012 case of *Inspector General of Police & Anr. Vs. S. Samuthiram*¹⁰⁸ to prevent eve teasing and stalking. A few of the protective measures suggested include placing CCTV cameras, setting up women's hotlines, and assigning female police officers in plain clothing to keep watch against eve teasing in public areas. At a more holistic level, legislators and society as a whole must make genuine efforts to prevent these crimes.

Social media is a double edged sword, it can create a platform that can give voice to the marginalised, but on the other it can propagate more crimes. It is the collective responsibility of legislators, social media platforms and intermediaries, and society at large to combat the issues that women are facing and providing women the positive liberty to be part of more inclusive and representative spaces both online and offline. Social media businesses are to be held accountable to ensure that all users may feel safe and welcome on their services.¹⁰⁹

NATURAL LANGUAGE PROCESSING (NLP)

Natural language processing is a subfield of Artificial Intelligence that studies human-computer natural language interactions. The ultimate objective of NLP is to enable computers to perceive, interpret, and generate meaningful, contextually relevant human language. NLP is capable of handling a broad variety of tasks, from simple ones like sentiment analysis and language translation to more complex ones like understanding and generating human-like replies in a conversation. Tokenization, semantic analysis, part-of-speech tagging, Named

¹⁰⁷ "Cybercrime against women up 28% since 2019, Karnataka's share highest: NCRB", *Business Standard*, Aug 30, 2022, available at: https://www.business-standard.com/article/current-affairs/cybercrime-against-women-up-28-since-2019-national-crime-records-bureau-122083001139_1.html (last accessed on 13 October 2024)

¹⁰⁸ *Inspector General of Police & Anr. Vs S. Samuthiram*, AIR 2013 SC 14

¹⁰⁹ Sharad Mehrotra, "Social Media & Women's Safety in a Digital World", *Economic Times*, June 29, 2017 <https://telecom.economictimes.indiatimes.com/tele-talk/social-media-and-women-s-safety-in-a-digital-world/2456> (last accessed on 11 October 2024)

Entity Recognition (NER), and other methods are used to carry out these functions.¹¹⁰ NLP regularly trains models on massive datasets using machine learning techniques.

Natural language processing (NLP) systems have the potential to unintentionally perpetuate or mimic gender stereotypes present in the training set. This issue arises because linguistic data used to train NLP systems typically reflects prejudice from society, including gender bias. Gender stereotypes in NLP may be seen in a variety of settings.

- **Inaccurate Training Data**

Textual data in vast quantities is used to train Natural Language Processing (NLP) models. Should this data include gender prejudices or stereotypes, the models could inadvertently perpetuate such bias. The normalisation of gender stereotypes in our culture and day-to-day activities means that prejudice against women is ingrained in AI models by default, as it acts as a mirror to society.¹¹¹

- **Word Embeddings**

Word embeddings, which use training data to represent words as vectors in a multidimensional space, have the ability to both detect and propagate gender stereotypes. For instance, certain roles or jobs may be associated with specific genders. Medicine is one such example in this way. It is quite likely that one will find photographs of a male doctor and a female nurse when they search for images of doctors and nurses.¹¹² The establishment of this occupational prejudice in society may be attributed to its cultural transmission. Thus, the innate prejudice and bias inadvertently seep into NLP or AI models.

- **Bias in Occupational Practice**

Based on training data, NLP models could associate certain genders with particular jobs, which might not fairly represent diversity in the actual world. As was previously said, prejudice in the workplace is common; for example, a woman is often associated with the word 'parent', assigning secondary functions of child rearing to the father. In a recent

¹¹⁰ Josette Coughlin, "Perspectives on Natural Language Processing", 64(1) *The French Review*, 172 (1990).

¹¹¹ Jim Holdsworth, "What is NLP: Natural language processing", IBM, 2024, available at: <https://www.ibm.com/topics/natural-language-processing> (last accessed on 11 October 2024)

¹¹² Nikhil Garg, Londa Schiebinger, et al. "Word embeddings quantify 100 years of gender and ethnic stereotypes", 115(16) *Proceedings of the National Academy of Sciences of the United States of America*, *National Academy of Sciences* E3635-E3644, April 17, 2018.

interview with UN Women, Beyza Doğuç said, "Artificial intelligence mirrors the bias that are present in our society and that manifest in AI training data."¹¹³

- **Stereotypes and Gender Pronouns**

Prejudice in the use of gender pronouns and their links with certain attributes or behaviours may reinforce preconceived notions about what behaviour is typical for a given gender.

- **Sentiment Analysis Bias**

Sentiment analysis algorithms may display bias in their evaluation of sentiment based on the gender of the author or the topic. There is still more work to be done in this relatively unexplored field. With further attempts to increase the fairness and equality of these systems, there is room for improvement by addressing gender stereotypes. The influence of gender stereotypes may also be lessened by implementing strategies to identify and counteract bias throughout the training and testing stages of NLP models.¹¹⁴

Bias may be lessened in the training set by using representative and varied data. A fair representation of genders, races, and other demographic characteristics should be ensured via collective interdisciplinary efforts. Creating and implementing NLP systems requires the development and observance of ethical standards. This entails being aware of possible bias and making an effort to overcome them.

Enhancing the transparency and interpretability of NLP models may help detect and comprehend bias, enabling more successful mitigation techniques. To guarantee that NLP models stay impartial and fair over time, regular monitoring and updates are necessary. This entails dealing with newly discovered bias and adjusting to linguistic alterations.

Researchers and practitioners may help create more impartial and equitable natural language processing systems by actively striving to reduce bias in training data and enhance the fairness of NLP models. It's a continuous task that calls for cooperation across computer science, technology, languages, ethics and law.¹¹⁵

¹¹³ Beyza Doğuç is the one who found in a generative AI the doctor-male, nurse-female bias

¹¹⁴ Chen Y., Mahoney C., et al. "Gender Bias and Under-Representation in Natural Language Processing Across Human Languages". *New York, NY: Association for Computing Machinery*. (2021)

¹¹⁵ Aylin Caliskan, "Detecting and Mitigating Bias in Natural Language Processing", Report from The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative is part of "AI and Bias," (2021)

Discussing NLP and Netiquette (online etiquette) requires consideration of gender disparity in order to foster fair and courteous interactions in digital environments. When developing NLP models, it is imperative to use vocabulary that is gender-neutral. Developers need to try to create models that treat each gender equally in order to avoid reinforcing stereotypes. Encouraging inclusive language and steering clear of gender-based stereotypes are two crucial netiquette rules to abide by whether using social media, online forums, or discussions. It is important for individuals to be conscious of the language and expressions they use in order to create an environment where everyone is treated equally, regardless of gender.¹¹⁶

Netiquette principles should be followed while encouraging participation in online conversations where opinions pour in from different individuals having multitudes of diverse backgrounds. In netiquette rules, harassment and discrimination based on gender should be explicitly addressed and denounced. Online platforms should have explicit rules and processes in place for reporting and managing these kinds of incidents. Including instructional components in netiquette guidelines may help expose gender inequity and prejudice.¹¹⁷

Moderation in online communities should be proactive in addressing harassment based on gender and enforcing rules about netiquette. All participants benefit from a more secure and welcoming atmosphere as a result.

Applications for natural language processing (NLP) should be created with inclusiveness and accessibility in mind, making sure that users of all genders can easily navigate the interfaces and interactions.

DEEP FAKES

Deep Fakes pose a serious threat to our political system, society, and industry because of a number of reasons. News media hangs in jeopardy as deep fakes make it harder for journalists to distinguish between real and fake news. There could be a question of compromise of national security by spreading misinformation and tampering in elections. It erodes the confidence citizens have in government. There could be major financial scams

¹¹⁶ Nancy Flynn, "Encourage Online Civility with Netiquette Rules", *Training Industry*, 2022, available at: <https://trainingindustry.com/articles/compliance/encourage-online-civility-with-netiquette-rules/> (last accessed 12 October 2024)

¹¹⁷ Supra at 28

perpetrated with the use of deep fake technology. Most importantly, there are obvious cybersecurity risks for individuals and institutions.

Deep fakes have a serious and diverse effect on women's rights. Deep fake technology, which uses AI to create fake videos and pictures that are very lifelike, presents serious risks to people's privacy, autonomy, and safety. This problem is faced by both men and women, however, its adverse impact is felt much more by women. Deep fakes are very lifelike audio and video content modifications that have the potential to violate women's rights, encourage harassment, and reinforce negative stereotypes.¹¹⁸ The production of non-consensual pornography, also known as "revenge porn," is one common use of deep fakes directed at women. These fake films may result in intimidation, extortion, and damage to one's reputation.¹¹⁹

- **Non-Consensual Sexual Imagery**

It is possible to create sexual photos of individuals without their consent by using deepfake technology. Women are majorly the targets, which may have devastating consequences on their personal and professional life, such as trauma, harassment, and reputational injury.¹²⁰

- **Objectification and Misrepresentation**

The Delhi High Court maintained the protection of an actor's identity, personal traits, and acting career against exploitation in *Anil Kapoor vs. Simply Life India and Ors.*¹²¹, especially with regard to the use of AI technology to create deep fakes. The ex parte order issued by the Court essentially barred sixteen firms from utilizing the actor's name, likeness, or image for financial gain or commercial objectives. The injunction also forbade the use of technological tools such as AI.

In a similar vein, in the case of *Amitabh Bachchan vs. Rajat Negi and Ors.*¹²², well-known actor Mr. Bachchan was given an ad interim in rem injunction against the unlawful

¹¹⁸ Edvinas Meskys & Ors., "Regulating Deep Fakes: Legal and Ethical Considerations", 15(1) *Journal of Intellectual Property Law & Practice*, 24 (2020).

¹¹⁹ Jordan Fairbairn, "Rape Threats and Revenge Porn: Defining Sexual Violence in the Digital Age", *University of Ottawa Press* (2015)

¹²⁰ Suzie Dunn, "Forms of TFGVB", from the Report "Technology Facilitated Gender Based Violence: An Overview", Centre for International Governance Innovation, (2020)

¹²¹ *Anil Kapoor Vs Simply Life India & Ors.*, 2023 SCC Online Del 6914

¹²² *Amitabh Bachchan Vs Rajat Negi & Ors.*, 2022 SCC Online (Del) 4110

exploitation of his personality rights and personal traits, including voice, image, and resemblance, for commercial purpose.

Here, the notable distinctions between male and female victims can be observed. The greatest loss a man may suffer in this regard is his reputation, and many would sympathize with him. They would not question how much access he may or may not have provided for the crime to be perpetrated. But when a woman's reputation is damaged, society as a whole begins to scrutinize and demonize her. Although the law may treat men and women equally in this sense, society has a fundamental role in the creation of laws and their interpretation when prosecuting these offences. Deepfakes may promote negative stereotypes and objectify women by superimposing their faces over explicit or degrading content. This reinforces harmful societal stereotypes that see women as nothing more than objects of sex and lessens the value of women.

A number of actresses expressed worry about their facial characteristics being improperly placed on other people and used to falsely represent them.

- **Eroding Credibility**

Deepfakes may be used to create erroneous and deceptive audio or video recordings of actresses, political activists or any other women showcasing them behaving improperly or spreading incorrect information. This may harm their standing in society, influence public opinion, and make it more difficult for them to promote social justice and women's rights. A politician named Manoj Tiwari, who was vying for the Indian Legislative Assembly in 2020, had his speech modified so as to be in the dialect of Haryanvi when it was in reality in English. This is sometimes intentionally done to sway voters before elections.¹²³

- **Cyberbullying and Harassment**

Although "trolling" happens to both men and women, women are more likely to suffer from it because of the patriarchal nature of our culture. Online abuse and cyber bullying of women are already commonplace. Deep fakes heighten the risks connected with them since they provide offenders with powerful tools to create and disseminate damaging and inaccurate information intended to frighten, discredit, or silence women. Women lose control over their

¹²³Tom Wicker, "Deep fake Technology presents Imminent Challenge to Democracy", *International Bar Association*, Feb. 25, 2020, available at: <https://www.ibanet.org/article/432E7100-35BE-44B4-B6AB-AD55758411F4>(last accessed 10 October 2024)

public appearance when they are the subject of deep fakes. As deep fakes are so easy to create, a woman may find her face and body used in ways she never meant, which may be quite horrifying.¹²⁴ The fact that India is ranked third in the world for cyber bullying, after China and Singapore, is not shocking.¹²⁵

Deep fakes may be used to spread false information about women, which may have an impact on their work, personal life, and safety. It is becoming more difficult to trust visual media in general due to the difficulty to differentiate between authentic and fake information. Deep fakes are now regarded by governments and law enforcement organizations throughout the globe as a major criminal and national security risk. This was proven during a hearing on "Artificial Intelligence and Human Rights" held by the United States Senate Committee on the Judiciary in June 2023.¹²⁶ In her testimony, Jennifer DeStefano detailed how she became a victim of a deep fake abduction scam and extortion. During a phone conversation that seemed to be from Ms. DeStefano's fifteen-year-old daughter, she said that she had been kidnapped and was being held for a million dollars in ransom. In June 2023, there was also talk of deep fake "sextortion" schemes, in which victims, particularly young ones, are either misled into disclosing personal information or having sexual imagery morphed onto them. The perpetrators then threaten to release the material to the public unless they are paid the ransom.¹²⁷

Deep fake threats are detrimental to elections, finances, and the dignity of those involved. A recent appeal against the use of deep fake technology in political campaigns was filed in the Delhi High Court for the most recent Lok Sabha election. Additionally, the petitioner demanded social media platforms like Google, Meta, and X to delete and prohibit deep fake

¹²⁴ Samer Hussain Al-Khazreji & Ors., "Impact of Deep fake Technology on Social Media: Detection, Misinformation and Societal Implications", 23 *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)* 429 (2023).

¹²⁵ According to Microsoft's "Global Youth Online Behavior Survey" in 2012, covering more than 7,600 children across 25 countries aged between 8 and 17, India ranked third in the number of cyber bullying cases (53%).

¹²⁶ "Artificial Intelligence and Human Rights." United States Senate Committee on the Judiciary, June 13, 2023. Available at: <https://www.judiciary.senate.gov/committee-activity/hearings/artificial-intelligence-and-human-rights> (last accessed on 12 October 2024)

¹²⁷ Jared Gans, "FBI Warns of 'deep fakes' in Sextortion Schemes," *The Hill*, June 7, 2023. Audrey de Rancourt-Raymond and Nadia Smali, "The unethical use of deep fakes," 30(4) *Journal of Financial Crime* 1066 (2023).

material on political candidates, representatives, leaders, and/or public figures until the results were made public on June 4, 2024.¹²⁸

In December 2023, the Union published a proposal requesting that Social Media Intermediaries (SMI) identify deep fake news and take necessary action against it, among other things.¹²⁹ The Union advised SMIs as follows:

- Information that violates the terms of user agreements, rules and regulations, or both is specifically identified using reasonable diligence and efforts; in accordance with the IT Rules, 2021, such cases are handled promptly and access is disabled well within applicable timelines.
- Users of SMI are urged not to host any content, including deep fake content, and to remove any such content within 36 hours of being reported. SMIs may face legal prosecution under the criminal laws in force and Rule 7 of the Information Technology Rules (Intermediary Guidelines and Digital Media Ethics) Code, 2021, should they neglect to take appropriate action in this respect.
- The advisory cautioned that failure to comply with the requirements of the IT Act, 2000 and the IT Rules, 2021 might result in SMIs losing their intended exemption from liability for illegal user-generated content under Section 79 of the IT Act, 2000.

The Union then said that in order to combat the creation and spread of deep fake information, it will be creating new laws and guidelines. The cornerstone of these actions, according to the Union, would be the detection, prevention, reporting, and public awareness of deep fake technology.

An investigation by the cybersecurity company Deeptrace, situated in Amsterdam, claims that women are the primary victims of deep fake pornographic movies. 90% of victims of crimes such as non-consensual porn, revenge porn, and other types of harassment are women, and deep fake is just another on the list.¹³⁰

Technological Remedies

¹²⁸ Malavika Peasad, “‘Not Proper for Court to Intervene’: Delhi HC on plea over Deep fake Videos”, *Indian Express*, May 3, 2024, available at: <https://indianexpress.com/article/cities/delhi/elections-delhi-hc-election-commission-pil-9303559/> (last accessed on 10 October 2024)

¹²⁹ Press Release on Dec. 26, 2023. Available at: <https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1990542> (last accessed 20 August 2024)

¹³⁰ Deep Trace, The State of Deep fakes, Landscape, Threats, and Impact. Available at: https://regmedia.co.uk/2019/10/08/deep_fake_report.pdf (last accessed on 31 July 2024)

Researchers are developing tools and techniques to recognize and authenticate digital information. These developments have made it possible for users to verify the legitimacy of material before sharing it. One possible use of machine learning algorithms with blockchain-based verification is the ability to distinguish between authentic and falsified content. Collaboration between tech firms, scholars, and lawmakers is necessary to stay ahead of deep fakes' constantly developing capabilities.¹³¹

Legal Defence

Legislation is necessary to eradicate deep fake damages and hold criminals accountable. It should be against the law to produce and distribute harmful deep fake material, including non-consensual pornography. These laws need to have severe enough punishments to deter offenders and ensure victims' rights. To satisfy the needs of jurisdiction, anonymity, and cross-border enforcement, legal systems must adapt. Together, lawmakers, tech companies, law enforcement, and civil society groups can mitigate the negative impact of deep fakes on women's rights.¹³²

Deepfakes restrict the freedom of speech and expression for women. The introduction of deep fake and algorithmic prejudice has enhanced the decades of discrimination against women. Strong detection and verification techniques must be developed; media and digital literacy must be promoted; victims must be given access to legal and support resources, and an atmosphere that upholds women's autonomy and dignity both online and offline must be fostered.

The first-ever AI safety conference saw the signing of the Bletchley Declaration, 2023¹³³, which recognized both the positive transformational effects of AI and the grave risks it may pose to privacy, individuality, and safety. One of the major signatories to it is India. The Declaration delineates a trajectory for global collaboration about the present and prospective obstacles brought about by AI. It also has an agenda item that seeks to identify risks associated with Artificial Intelligence and to put appropriate risk-based laws into place in different countries, and ensure transparency by encouraging private industries to develop

¹³¹ Nicholas Mesa-Cucalon, "Deepfakes: Effective Solutions for Rapidly Emerging Issues", May, 27, 2021. Available at: <https://medium.com/analytics-vidhya/deepfakes-effective-solutions-for-rapidly-emerging-issues-8b1685feef56>(last accessed on 31 July 2024)

¹³² Caroline Criado Perez, "Invisible Women: Data Bias in a World Designed for Men", *Abrams Press, New York*, (2019).

¹³³ Bletchley Declaration on AI Safety, November 2023

cutting-edge AI capabilities. If not, all of the previous century's effort to emancipate women would face a major setback in the digital age.

PORNOGRAPHY

Indian courts have evolved distinct standards of scrutiny throughout the last decades to determine the extent to which the state may intervene in cases involving pornography or other potentially "obscene" content.

Although not directly related to pornography, the British case of *R v. Hicklin*¹³⁴ established a test that set a precedent for extremely low bar of review by allowing a conviction for "obscenity" if it was shown that it served the most vulnerable member of society's prurient interests. Later, the criteria was replaced by one that was somewhat more reasonable. In the American case of *Roth v. United States*¹³⁵ a defendant was found guilty of marketing and selling erotica that was deemed to be "obscene, lewd, lascivious, and filthy." According to the Roth test, the requirements of obscenity should be applicable to the regular individual and be comparable to the standards of the community. In the instances of *Memoirs v. Massachusetts*¹³⁶ and *Miller v. California*¹³⁷, this standard was further altered. In the latter instance, the proprietor of a mail order company that specialized in pornographic material mass distributed a number of pamphlets with explicit sexual content. The Miller test, a three-prong standard, was created by the court to determine obscenity in this particular case. This uses a community norm that is offensive in nature, using offence as a yardstick for obscenity.

But pornography ought to be seen via John Stuart Mill's Harm Principle, discussed earlier, which holds that consequentiality of pornographic material is directly or indirectly harmful to women. Despite advances in law, women's status and treatment remain largely unchanged due to the pervasiveness of sexual exploitation and violence against them. Pornography's negative impacts on society perpetuate discrimination, injustice, and the subordination of women to men.¹³⁸ Speaking out against these injustices puts women in danger of social exclusion, threats, and derision, among other serious challenges. In the case of *Paris Adult*

¹³⁴ *R Vs Hicklin, England and Wales High Court of Justice, Queen's Bench Division*, 11 Cox C.C. 19 (1868)

¹³⁵ *Roth Vs United States*, 354 US 476 (1957)

¹³⁶ *Memoirs Vs Massachusetts*, 383 US 413 (1966)

¹³⁷ *Miller Vs California*, 413 US 15 (1973)

¹³⁸ Andrea Dworkin, "Against the Male Flood: Censorship, Pornography and Equality", 8 HARV. WOMEN'S L.J. 1 (1985).

*Theatre v. Slaton*¹³⁹, displaying pornographic material at a theatre was deemed against community standards and did not pass the Miller test, even though it was consensual and only adults over 21 were permitted to purchase tickets.

It is illegal to forcefully record and distribute sexually explicit material, include minors in it, or publicize it. The accused in the Canadian case of *R v. Butler*¹⁴⁰ operated a store where he rented and sold pornographic publications, 'hard core' videotapes, and other sexual paraphernalia. He faced several charges of selling pornographic content, having pornographic material on hand with the intent to distribute or sell it, and displaying pornographic and lewd material for public consumption. In this decision, the Court mostly relied on feminist arguments about the negative effects of pornography. As a result, it is both offensive and cannot take the defence of protected speech.

There are two schools of thought on pornography and whether or not the government should get involved. While advocating for freedom of speech and expression and speaking out against censorship are important aspects of individual liberty, there is the case of public morality, the defence of women's rights, and the necessity for government engagement which are additional factors to take into consideration. Although some feminist writers advocate for the intentional use of pornographic material as a method of emancipation, this anti-censorial viewpoint is not supported in today's world when there are several ways to procure and abuse seemingly "wilful" content. This profusion of sexual information causes more damage than good in a culture where sexual literacy is low to nil. While the author does not fully disagree with feminist writers and jurists who support libertarian ideas, state involvement is imperative in the digital age to stop the harm that pornography does to women, particularly with the introduction of AI and its many potential misuse. The author believes that governmental action would secure the maintaining of women's freedom of speech and expression, rather than restricting it.

Normally, the fundamental rights outlined in Part III of our Constitution enforce negative liberties against the state; however, in this instance, proactive state action is required to ensure both positive and negative liberty in order to lessen the evils committed against women. Only then can women be treated as equals, not as second-class citizens.

¹³⁹ *Paris Adult Theatre Vs Slaton*, 413 US 49 (1973)

¹⁴⁰ *R Vs Butler*, [1992] 1 SCR 452

Prominent scholars have often discussed a particular facet of the radical school of feminist jurisprudence known as "bodily autonomy."¹⁴¹ Pornography undermines women's feeling of dignity and undermines this fundamental idea. Radical feminist philosophy holds that women become objects of sex, prostitution, social sexism, and patriarchy. They contend that pornographic media denigrates women while depicting uneven power dynamics between men and women. Since pornography falls outside of some categories of obscenity that may be protected speech, state involvement is required in this case.

Women have a fundamental right to be protected against such a deterioration of their rights, as stated in the Constitution's values of equality (Article 14), freedom of choice derived from the right to live with dignity (Article 21), and freedom of expression (Article 19).

The act of creating, sharing, and distributing pornographic material online is known as internet pornography. Furthermore, Section 354A of the Indian Penal Code (BNS Section 75) addresses sexual harassment. It states that a man may face legal action if he intentionally sends a woman pornographic content by email, or another platform without her consent. Section 67A of The Information Technology Act, which covers products that process or transmit sexually explicit content in any electronic format, forbids such activities even further. A first offence involves a maximum of five years in prison and a fine that might reach 10 lakhs, as per the Information Technology Act. A maximum penalty of seven years in jail and a fine exceeding 10 lakhs are imposed for a second conviction. However, this law was formerly governed by IPC Section 292 (294 BNS) which dealt with the obscenity violation and included anything voyeuristic, lewd, or intended to dehumanize and corrupt individuals. As a result, this section now forbids the sale, distribution, communication, exhibition, or profit-making of such a company in public and charges penalties for doing so.

Suhas Katti v. State of Tamil Nadu¹⁴² started as a complaint filed by a woman who claimed she was being harassed and defamed by a man in a Yahoo message group. The accused established a fake email account in the woman's name and started forwarding correspondence to it. More calls came in from individuals claiming she was a sex worker. Accused was consequently sentenced to two years of hard labour and a fine of Rs. 4000 for an offence under Section 67 of the Information Technology Act (punishment for publishing or communicating obscene material in electronic form), one year of hard labour and a fine of

¹⁴¹ Rosemarie Tong, "Women, Pornography and the Law", 73(5) *Academe* 14 (1987).

¹⁴² *Suhas Katti Vs State of Tamil Nadu*, CC No. 4680 of 2004

Rs. 500 for an offence under Section 79BNS (words, gestures, or conduct intended to insult a woman's modesty) and two years of rigorous imprisonment and a fine of Rs. 500 under Section 336(4) BNS(forgery in cases involving harm to reputation)

In the case of *Kalandi Charan Lenka v. State of Odisha*¹⁴³, the accused broke into the victim's social media accounts and sent her pornographic messages, harassing her continually. Everywhere she lived and studied, there were printed pamphlets about her with sexual comments scribbled on the walls. Eventually, to insult her reputation and offend her modesty, a fake Facebook account was made in her name and morphed nude photos were uploaded. The offender was charged under BNS Section 79, Section 67A of the IT Act, and BNS Section 77(Voyeurism) and the matter was described as cyber defamation.

In a recent judgment¹⁴⁴, the Court decided that the intermediaries were required to remove any offending content from their platform in relation to Non-Consensual Intimate Images (NCII). The Court highlighted the potential damage that might arise from the posting of NCII and the agony that victims may experience when they have to search the internet for new uploads of graphic images or videos in order to ask for their removal. The Court held that intermediaries must take a "reasonable effort" to stop its users from publishing anything that is offensive or does not belong to them, and that they must employ technology to ensure that objectionable photos are deleted from reposts.

In *Mrs. X v. Union of India*¹⁴⁵, the offender became an online acquaintance of the victim, a married woman. The victim was abused by the accused one day when he went to her home and took obscene pictures of her from her own phone and onto his. The woman was then subjected to blackmail by the accused, who even uploaded the photos to pornographic websites. He further posted her videos and pictures with her face altered. Mrs. X filed a request with the Delhi High Court to prohibit the websites that featured her personal images. The Court was confused about what orders it can give to internet intermediaries (like Google and Microsoft) in situations where those intermediaries remove images but then re-upload them in order to protect both the rights of individuals and the intermediaries' obligations, since despite efforts to eliminate any references to the pictures, they were "repeatedly being re-uploaded and re-produced." Even after the accused was apprehended and the case was

¹⁴³ *Kalandi Charan Lenka Vs State of Odisha*, 2017 SCC Online Ori 52

¹⁴⁴ *Infra* at 57

¹⁴⁵ *Mrs. X Vs Union of India*, Neutral Citation Number: 2023:DHC:2806

rendered moot, the court continued to pursue it in order to provide a lasting remedy for future victims.

After the judgment had been reserved, the government alerted the Court to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2022.

The main issue under discussion was, who is responsible for removing NCII¹⁴⁶ off the internet. While the Ministry of Electronics and Information Technology (MEITY) argued that "proactive monitoring and removal of content will adversely affect the freedom of speech and expression of other individuals having the same or similar name as the Petitioner," the intermediaries are required to remove offensive content within 24 hours of becoming aware of it. Delhi Police announced that it was establishing District Cyber Police Stations and a website feature that assists women and children in filing complaints in order to monitor and prosecute crimes against women and children on the internet.

The Information Technology Act, 2000 and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 mandated the intermediaries to remove any objectionable information from their platform. This duty went beyond the precise URLs that people submitted. Intermediaries would only be protected if all legal conditions were met.

Although the term "revenge porn" is frequently used, the Court noted that this is only one instance of NCII and defined it as "sexual content that is distributed under the larger umbrella of cyber-harassment, largely meaning that its dissemination is non-consensual and comes under the said larger umbrella of being depicted in said content... [it] may or may not be taken with the consent of the individual involved."¹⁴⁷

The Court noted that the general perception of these acts as less serious than sexual harassment and molestation notwithstanding, the victims' mental health is negatively impacted by the material being shared about them causing further "disruptions to life." Given the increase in NCII cases and internet accessibility, the Court ruled that it was critical to understand how the IT Act and its rules effect NCII usage as well as the role and obligations

¹⁴⁶ Rule 3(2)(b) "defines NCII as any content which prima facie exposes any individual's private area, shows such individual in full or partial nudity, shows or depicts such individual in any sexual act or conduct, or is in the nature of impersonation in an electronic form, including artificially morphed images"

¹⁴⁷ *Mrs. X Vs Union of India*, Neutral Citation Number: 2023:DHC:2806

intermediaries have in the dissemination of NCII and the prevention of its abuse. The Court provided guidelines and instructions.¹⁴⁸

But none of these requirements seem to be nearly sufficient to address the plethora of crimes that seem to be emerging in tandem with the accessibility and ease of use of technology, particularly AI.

Courts acknowledged, the "abysmal absence of a collaborative effort that should ideally be undertaken by the intermediaries and the State" which further encourages offenders. The victim's sentiments, reputation, and access to legal choices are all negatively impacted by this, according to the Court, especially in India where "NCII abuse does indeed lead to harrowing consequences and everlasting stigma for the victim."¹⁴⁹

The Press Information Bureau (PIB) has clarified through its June 20, 2021 release that the Indian government has been directed by the Supreme Court to create regulations with the goal of eliminating child pornography, images of rapes and gang rapes, and related content from websites and applications.¹⁵⁰ In accordance with this decision, the Indian government drafted the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.

A start towards the formation of intermediary guidelines was made in the recommendation letter dated February 8, 2015, in the context of *Re: Prajwala Letters*¹⁵¹. A non-governmental organization (NGO) called Prajwala, situated in Hyderabad works to eradicate sex trafficking and forced prostitution. On February 18, 2015, the Hon'ble Supreme Court received a letter from NGO Prajwala about videos of sexual assault, rape, and child pornography that were making rounds on social media platforms. In response, the court mandated the creation of a committee to assist and counsel the court in order to guarantee that recordings containing the same are not distributed and are completely inaccessible. It is evident that the committee's main objective was to protect the victims' names and reputations.

¹⁴⁸ Ibid

¹⁴⁹ Supra at 60

¹⁵⁰ "Permanent Mission of India responds to the concerns raised by Special Branch of Human Rights Council about India's IT Rules, 2021", Report of Press Information Bureau, Ministry of Electronics & IT, (June 2021). Available at: <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=172873> (last accessed on 10 August 2024)

¹⁵¹ *Re: Prajwala Letters case*, (2018) 15 SCC 551

The committee made a number of recommendations, including that the Indian government work with civil society organizations and the companies it represents to suggest expanding the list of keywords that appear when Child Pornography warning and Public Service Announcements are displayed during searches. The Committee agrees that it is essential to develop a Central Reporting Mechanism India's hotline webpage, in line with other countries like the United States with NCMEC¹⁵². Law enforcement in this area also has to be reformed consistently to keep up. Internet enterprises should support the development of law enforcement, non-governmental organizations, and relevant authorities in India by providing a variety of trainings on online criminal investigations and the usage of relevant internet technology.

On December 11, 2018, Meity filed an affidavit. The order authorized the government two weeks to create the necessary standards and implement them so that child pornography, rape, gang rape photos, videos, and websites would no longer be present on content hosting platforms and other applications.¹⁵³

Stringent legislation and more progressive court rulings such as this one are needed to truly make the laws and their execution and implementation efficient.

CONCLUSION

The European Union passed General Data Protection Regulations (GDPR) in 2018 and the Artificial Intelligence Act (AI) Act in 2024. The United States deals with AI the same way as it deals with a regular tort. Besides, it has state specific legislations handling the same (for example, California Consumer Privacy Act, 2018). The world is gradually understanding the evils of unregulated use of AI and progressing towards safeguards accordingly.

The Information Technology Act, 2000, IT Rules, 2021 and IT amended Rules 2022 are the main Acts in force to combat the issues that are currently cropping up with AI. The Indian government is going towards a progressive direction as well with precautions and policies underway. The Digital Personal Data Protection Act, enacted recently in 2023 is a safeguard against rampant misuse of private data. The public policy think tank in India, NITI Aayog, was given the duty of creating guidelines and regulations for the advancement of AI in the

¹⁵² National Center for Missing and Exploited Children

¹⁵³ "Supreme Court asks govt to set up cell to fight child porn", *Indian Express*, Oct. 27, 2017, available at: <https://indianexpress.com/article/india/supreme-court-asks-govt-to-set-up-cell-to-fight-child-porn-4908458/> (last accessed on 1 August 2024)

country as part of the country's national AI program. The National Strategy for Artificial Intelligence #AIforAll ("NSAI 2018"), published by NITI Aayog in 2018, emphasized using AI to promote inclusive and social progress in accordance with the Government of India's planned AI roadmap.¹⁵⁴

Using AI ethically, legislating to mandate accountability of digital platforms, enacting an AI specific legislation like the European Union would be the next steps. Implementation and execution of the policies and guidelines effectively could bring about the much needed protection of women's rights in this regard.

¹⁵⁴ "National Strategy for Artificial Intelligence", 2018, *available at*: <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf> (last accessed on 10 October 2024)