



# Jus Corpus Law Journal

Open Access Law Journal – Copyright © 2026 – ISSN 2582-7820

Editor-in-Chief – Prof. (Dr.) Rhishikesh Dave; Publisher – Ayush Pandey

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International (CC-BY-NC-SA 4.0) License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

---

## Examining Algorithmic Bias and Gender Harassment: A Case for Intermediary Accountability

Saba Urooj<sup>a</sup>

<sup>a</sup>Bangalore Institute of Legal Studies, Bengaluru, India

Received 13 February 2026; Accepted 13 March 2026; Published 17 March 2026

---

*Gender harassment refers to the acts of verbal, nonverbal, or physical aggression, intimidation or hostility based on sex or gender identity and that reinforce heteronormative gender roles.<sup>1</sup> Gender harassment is a historical societal ill that has undergone a profound transformation in the present times. The advent of social media platforms and artificial intelligence has given rise to sophisticated forms of online abuse. Compounding these harms, algorithmic systems amplify misogynistic content and reinforce deeply entrenched gender biases, thereby normalising discrimination and endangering women in digital spaces. The central issue addressed in the present research concerns the legal accountability of social media intermediaries for algorithmic bias that amplifies gender-based harassment. This paper examines the legal and institutional responses, rooted in anti-discrimination legislation and landmark judicial pronouncements, that remain largely reactive and ill-equipped to address the architecture-driven nature of online harms. Employing analytical methodology, this paper traces the evolution of gendered harassment from physical to digital spaces, analyses key legislative frameworks and judicial decisions in India, and critically evaluates the transition of online intermediaries from passive conduits to active gatekeepers. Drawing on global best practices, it proposes a statutory framework that requires full disclosure of content moderation policies and proactive design obligations to mitigate discriminatory outcomes. By prioritising oversight of the underlying digital architecture rather than criminalising content, India can move beyond a compliance-centric model and foster platforms that actively promote safety, equality, and*

---

<sup>1</sup> 'Gender-Based Harassment' (University of Illinois Chicago)

<<https://sexualmisconduct.uic.edu/policy/definitions/gender-based-harassment/>> accessed 10 February 2026

*justice. Such reforms are indispensable for creating a digital public sphere that does not exacerbate offline gender stereotypes but instead dismantles them, ensuring an equitable and dignified online existence for all.*

**Keywords:** *gender harassment, accountability, bias, algorithms, reforms.*

---

## INTRODUCTION

The pre-digital world was shaped by women saying, “I am scared to walk outside.” The post-digital moment is defined by women whispering, “I am scared to be online.” As Gaatha Sarvaiyya, an Indian law graduate, poignantly notes about posting images online: “The thought immediately pops in that maybe it’s not safe...”<sup>2</sup> Her fear echoes the paradox of millions of women navigating a digital environment where invisibility offers no protection and visibility makes one vulnerable.

Today, there are 5.04 billion social media users, over 60% of the world’s population,<sup>3</sup> all functioning within ecosystems governed by algorithms designed, owned, and controlled by online intermediaries. These platforms decide what is amplified, what is buried, and what is allowed to exist. This unregulated autonomy of these intermediaries has enabled the rapid proliferation of harmful, sexualised, and misogynistic content.

The consequences are devastating. The video of two women being paraded naked in Manipur circulated across Twitter, WhatsApp, and other platforms within hours, remaining online for days before meaningful action was taken.<sup>4</sup> In the Sulli Deals incident (2021),<sup>5</sup> photographs of women were uploaded without consent and ‘auctioned’ in a grotesque display of digital humiliation. Similar episodes have created a chilling effect for women who fear the very platforms meant to connect them.

---

<sup>2</sup> Aisha Down, “The chilling effect’’: how fear of ‘nudify’ apps and AI deepfakes is keeping Indian women off the internet’ *The Guardian* (05 November 2025) <<https://www.theguardian.com/global-development/2025/nov/05/india-women-ai-deepfakes-internet-social-media-artificial-intelligence-nudify-extortion-abuse>> accessed 21 November 2025

<sup>3</sup> ‘GLOBAL SOCIAL MEDIA STATISTICS’ (*Data Reportal*) <<https://datareportal.com/social-media-users>> accessed 21 November 2025

<sup>4</sup> Cherylann Mollan, ‘Manipur: India outrage after women paraded naked in violence-hit State’ *BBC* (20 July 2023) <<https://www.bbc.com/news/world-asia-india-66253389>> accessed 26 November 2025

<sup>5</sup> Alok Deshpande and Hemani Bhandari, ‘Sulli deals, Bulli Bai and the young and educated hatemongers’ *The Hindu* (22 January 2022) <<https://www.thehindu.com/news/national/sulli-deals-bulli-bai-and-the-young-and-educated-hatemongers/article38305009.ece>> accessed 26 November 2025

This underlines a disturbing reality. Intermediaries cannot remain passive conduits of information while their platforms become theatres of gendered violence. The time has come to shift from passive intermediaries to active digital actors who bear meaningful responsibility for the environments they create and profit from.

This paper adopts an analytical methodology to examine existing statutory frameworks, judicial precedents, and relevant international standards on intermediary accountability. This paper is divided into four parts. The first part traces the evolution of gender harassment to the contemporary digital harassment within its socio-legal context. The second part examines the nature of online harassment against women. The third part identifies and critiques algorithmic bias embedded in platforms, particularly where such bias reinforces gender stereotyping. The final part studies the existing international regulatory approaches on this subject and proposes a statutory framework for intermediary accountability in India.

## LITERATURE REVIEW

**Women in Rigveda (1933)<sup>6</sup> – Bhagwat Saran Upadhyaya:** *Women in Rigveda*, written by Bhagwat Saran Upadhyaya, is one of the earliest texts on the position of women during the Vedic times. The position of women during those times is highlighted as a ‘golden age’ where women enjoyed high status, dignity and social equality. This is relevant as it marks the evolution of women in the current society, which has weakened the position of women by reinforcing gender biases.

**The Position of Women in Hindu Civilisation (1938)<sup>7</sup> – AS Altekar:** This book, written by AS Altekar, provides a comprehensive examination of the evolution of the position of women from prehistoric times to the modern era. It critiques the decline in the position of women over a period of time, owing to foreign invasions.

**The Second Sex (1949)<sup>8</sup> – Simone de Beauvoir:** In *The Second Sex*, Simone de Beauvoir conceptualised gender as a socially constructed identity, arguing that women are considered weak by social conditioning and not biological determinism. It provides the theoretical

---

<sup>6</sup> ‘Full text of “Women In Rigveda”’ (*Internet Archive*)

<[https://archive.org/stream/WomenInRigveda/Women%20in%20Rigveda\\_djvu.txt](https://archive.org/stream/WomenInRigveda/Women%20in%20Rigveda_djvu.txt)> accessed 20 December 2025

<sup>7</sup> AS Altekar, *Position of Women in Hindu Civilization: From Prehistoric Times to the Present Day* (Motilal Banarsidass 2016)

<sup>8</sup> Simone de Beauvoir, *THE SECOND SEX* (1st edn, Jonathan Cape 1949)

foundation for gender harassment. However, it does not address technology-mediated gender harm.

**The Underside of History (1976)<sup>9</sup> - Elise M Boulding:** In this work, Elise M Boulding explored the changing roles of women and identified the division of labour as the earliest manifestation of gender differentiation, which subsequently entrenched the differences between men and women, laying the groundwork for gender harassment.

**The Creation of Patriarchy (1986)<sup>10</sup> - Gerda Lerner:** In the Creation of Patriarchy, Gerda Lerner draws focus on the creation of patriarchy and the origins of women's subordination. It highlights control over women's labour, sexuality, and mobility. Lerner traced the historical formation of patriarchy through legal, economic, and social institutions. It also propounds the idea that patriarchy was the root cause of gender harassment.

**Sexual Harassment in Work Organisations: A View from the 21<sup>st</sup> Century (2018)<sup>11</sup> - Louise F. Fitzgerald and Lilia M. Cortina:** This work by Louise F. Fitzgerald and Lilia M. Cortina provides a comprehensive examination of workplace harassment against women. This work has been cited to trace the changing forms of gender harassment over time.

**Accelerating Efforts to Tackle Online and Technology-facilitated Violence Against Women and Girls (VAWG)<sup>12</sup> - UN Women:** This report by UN Women, released in 2022, provides an overview of existing data on online violence against women and offers recommendations to address this challenge, including measures to enhance intermediary accountability. However, it does not offer policy recommendations to India.

Existing literature does not sufficiently connect gendered digital violence with algorithmic amplification and the legal responsibilities of intermediaries in India. This paper addresses that gap by examining how algorithmic systems shape online misogyny and proposes the enactment of a new rights-based statutory framework to ensure intermediary accountability.

---

<sup>9</sup> Elise Boulding, *The Underside of History: A View of Women Through Time* (1st edn, Westview Press 1978)

<sup>10</sup> Gerda Lerner, *The Creation of Patriarchy* (OUP USA 1988)

<sup>11</sup> Louise F Fitzgerald and Lilia M Cortina, 'Sexual harassment in work organizations: A view from the 21st century' in CB Travis et al., *APA Handbook of the Psychology of Women: Perspectives on Women's Private and Public Lives* (vol 2, American Psychological Association 2017)

<sup>12</sup> *Accelerating efforts to tackle online and technology-facilitated violence against women and girls* (UN Women 2022)

## METHODOLOGY

This paper adopts an analytical research methodology to examine the intersection of algorithmic bias, gendered harassment, and intermediary accountability. It is grounded in the study of Bharatiya Nyaya Samhita, 2023,<sup>13</sup> POCSO, 2012,<sup>14</sup> Information Technology Act, 2000<sup>15</sup> and international human-rights standards, and policy instruments shaping platform liability. Secondary materials include scholarly articles, global reports, regulatory frameworks, and empirical research on online gender-based violence, which provide interpretative depth. A comparative legal method assesses global models such as the EU Digital Services Act,<sup>16</sup> China's 2022 Algorithmic Regulation,<sup>17</sup> and the EU AI Act (2022),<sup>18</sup> enabling identification of gaps and lessons for India. The study also adopts a socio-legal lens to understand how cultural, historical, and technological factors contribute to digital misogyny. Together, these methods facilitate a rights-centred evaluation of intermediary responsibility in India.

## ANALYSIS

'One is not born, but becomes a woman,' coined by Simone de Beauvoir,<sup>19</sup> reflects the deep, entrenched meaning of gender, as distinct from sexual identity. Gender identity is not created by nature but by man himself, as Beauvoir observes, 'humanity is male, and man defines woman not in herself but as relative to him.'<sup>20</sup>

Gender refers to the characteristics of women, men, girls and boys that are socially constructed, which include the norms, behaviours and roles associated with it. As a social construct, gender varies from society to society and can change over time.<sup>21</sup> Acts of verbal,

<sup>13</sup> Bharatiya Nyaya Samhita 2023

<sup>14</sup> Protection of Children from Sexual Offences Act 2012

<sup>15</sup> Information Technology Act 2000

<sup>16</sup> EU Digital Services Act 2022

<sup>17</sup> 'Provisions on the Management of Algorithmic Recommendations in Internet Information Services' (*China Law Translate*, 04 January 2022)

<<https://www.chinalawtranslate.com/en/algorithms/#:~:text=Provisions%20on%20the%20Management%20of%20Algorithmic%20Recommendations,Services.%20By%20China%20Law%20Translate%20on%202022/01/04>> accessed 10 February 2026

<sup>18</sup> EU Artificial Intelligence Act 2024

<sup>19</sup> Beauvoir (n 8)

<sup>20</sup> Meriel Colenutt, 'One is not born, but becomes a woman' (*TORCH | The Oxford Research Centre in the Humanities*, 22 March 2021) <<https://www.torch.ox.ac.uk/article/one-is-not-born-but-becomes-a-woman>> accessed 22 November 2025

<sup>21</sup> 'Gender and health' (*World Health Organisation*) <[https://www.who.int/health-topics/gender#tab=tab\\_1](https://www.who.int/health-topics/gender#tab=tab_1)> accessed 22 November 2025

nonverbal, or physical aggression, intimidation or hostility based on sex or sex-stereotyping, sexual orientation or gender identity and that reinforce heteronormative gender roles. Gender-based harassment is not generally motivated by sexual interest or intent. It is more often based on hostility and is often an attempt to make the target feel unwelcome in their environment. This provides the theoretical foundation by conceptualising gender as a socially constructed identity and gender harassment as a form of structural hostility rather than sexual misconduct.

## EVOLUTION OF GENDER HARASSMENT

In the earliest forms, as noted by Elise M. Boulding in her work 'The Underside of History', she talks about how the findings from anthropological studies of hunting-gathering groups divide the roles of genders as 'Man: the hunter' versus 'woman: the child bearer and nurturer.'<sup>22</sup> In India, Raja Ram Mohan Roy's contributions are reflected in the works of *Women in Rigveda*<sup>23</sup> and *The Position of Women in Hindu Civilisation*, both of which characterise the Vedic period as a golden age for women. As documented by American historian Gerda Lerner, written records from that time show women gradually disappearing from the public world due to the practice of patrilocal marriage. Over time, marriage turned into a rigid legal institution, leading to the establishment of a patriarchal society. Patriarchy is defined as the 'system of social structures and practices in which men dominate, oppress and exploit women'.<sup>24</sup> Gender harassment is rooted in the concept of patriarchal societies.

During this time, early codes like the *Manusmriti*<sup>25</sup> institutionalised women as weak and dependent on men. Scholars such as Gerda Lerner, Sylvia Walby and Riane Eisler identify this period as the consolidation of patriarchal social structures. Harassment began to be expressed not only interpersonally but also through customs like early marriages, dowry and the sati system that constrained women's participation in public life, education, ownership, and governance. With British colonialism, the structures of gender harassment became even

---

<sup>22</sup> Angela Saini, 'How did patriarchy actually begin?' *BBC* (30 May 2023) <<https://www.bbc.com/future/article/20230525-how-did-patriarchy-actually-begin>> accessed 26 November 2025

<sup>23</sup> Bhagwat Saran Upadhyaya, *Women in Rigveda* (Khama Publishers 1994)

<sup>24</sup> Sylvia Walby, 'Theorising Patriarchy' (1989) 23(2) *Sociology* <<https://doi.org/10.1177/0038038589023002004>> accessed 26 November 2025

<sup>25</sup> Hirday N Patwari, 'Women as Depicted in the Manusmriti' (*The Modern Rationalist*, 23 December 2020) <<https://modernrationalist.com/women-as-depicted-in-the-manusmriti/>> accessed 19 December 2025

more entrenched, shaped by new social hierarchies and colonial norms. India's post-independence era marked significant progress in the status and role of women in society. With the adoption of the Constitution and various progressive legislations, women's public participation increased. However, gender harassment has shifted its battleground to workplace harassment.

In Fitzgerald & Cortina's (2018)<sup>26</sup> comprehensive review of research on sexual harassment, four subcategories of gender harassment have been recognised. These categories include Sexist hostility, sexual hostility, gender policing and work/family policing. Although these categories were originally developed in the context of workplace harassment, the underlying patterns of hostile gender expression appear across all social and digital environments.

This underlines that gender harassment has always existed, but the forms have changed with social norms, power structures and technology. Taken together, this historical evolution is of significance as it implies that gender harassment is not just an individual misconduct, but is deeply embedded within institutional and societal setups. Today, digital platforms constitute the latest site for harassment that works through their opaque and unaccountable algorithms.

## **DIGITAL MEDIA AND GENDER HARASSMENT**

Digital media's journey began with the rise of the internet in the 1990s, a transformative period that enabled the creation of new online platforms and services. Initially, the internet was used primarily for email and basic websites. However, as technology evolved, so did the capabilities of digital media. The 2000s witnessed the emergence of social media platforms like Facebook, Twitter, YouTube and LinkedIn. A 'Social Media Intermediary' means an intermediary which primarily or solely enables online interaction between two or more users and allows them to create, upload, share, disseminate, modify or access information using its services.<sup>27</sup> The emergence and growth of Artificial Intelligence has ushered in an era with an unprecedented series of extraordinary advances in chatbots, virtual assistants, natural language processing and machine learning. The shift to digital spaces has amplified the scale,

---

<sup>26</sup> Amir Karami et al., 'Hidden in Plain Sight For Too Long: Using Text Mining Techniques to Shine a Light on Workplace Sexism and Sexual Harassment' (2024) 14(1) *Psychology of Violence* <<https://doi.org/10.1037/vio0000239>> accessed 20 November 2025

<sup>27</sup> Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021

speed and severity of gender harassment, transforming it from a localised social practice into a pervasive, networked phenomenon.

In 2018, the Special Rapporteur on violence against women and girls, its causes and consequences defined online violence against women to encompass any act of gender-based violence against women that is committed, assisted or aggravated in part or fully by the use of information and communications technology (ICT), such as mobile phones and smartphones, the Internet, social media platforms or email, against a woman because she is a woman, or affects women disproportionately. Cyber harassment also include trolling women for their opinions, gender-based hate speeches, image abuse, blackmailing, threats, fraud, deep fake images, revenge pornography, insults, obscene messages, bullying, rape threats, cyberstalking, sextortion, cybersex trafficking, publishing sexually explicit images, publicly humiliating women, negatively portrayals, reiterating conventional stereotypes to devalue, defamation, moral policing, breach of the right to privacy, identity theft and economic violence and other forms of violence.<sup>28</sup> Available data demonstrate the prevalence and severity of the problem. A UN report suggests that 16-58% of women have experienced online violence,<sup>29</sup> and another report suggests that 58% of girls reported that they have personally experienced some form of online harassment on social media platforms.<sup>30</sup>

Platforms that host content often employ automated systems that take in user-supplied data, data gleaned from the user's use of the platform, and, through a set of parameters, determine what content should be flagged for review or removal or could be shown to users. People often refer to the personalised nature of these technologies as their algorithm, and this supports a colloquial conception of a 'social media algorithm'.<sup>31</sup> These algorithms are embedded with bias, leading to the amplification of online gender harassment.

---

<sup>28</sup> Adv. Dr. Shalu Nigam, 'Ending Online Violence against Women in India: Calling for an Inclusive, Comprehensive, and Gender-Sensitive Law and Policy Framework' (*IMPRI Impact and Policy Research Institute*, 22 April 2024) <<https://www.impriindia.com/insights/ending-online-violence-against-women/>> accessed 18 November 2025

<sup>29</sup> 'FAQs: Digital abuse, trolling, stalking, and other forms of technology-facilitated violence against women and girls' (*UN Women*, 13 November 2025) <<https://www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women/>> accessed 24 November 2025

<sup>30</sup> *THE STATE OF THE WORLD'S GIRLS REPORTS* (Plan International 2023)

<sup>31</sup> Garrett Crites, 'Demystifying the "Social Media Algorithm": The Legacy of Surveillance Advertising and Platformization' (2024) Honors Projects <<https://digitalcommons.spu.edu/cgi/viewcontent.cgi?article=1220&context=honorsprojects>> accessed 27 November 2025

Research shows how algorithms used by digital platforms perpetuate racism and misogynistic stereotypes.<sup>32</sup> Misogynistic, sensational, or inflammatory content often produces higher levels of user interaction, which in turn feeds into advertising revenue metrics. As a result, recommendation engines designed to optimise attention may inadvertently prioritise harmful gendered content, not because of malice, but because of misaligned commercial incentives.

A growing concern is the increasing use of opaque inferential analytics that exacerbate existing biases in society. One of the recurrent biases is gender stereotyping. Gender stereotyping refers to the practice of ascribing to an individual ‘woman’ or ‘man’ specific attributes, characteristics, or roles by reason only of their membership in the social group of ‘women or men’.<sup>33</sup> This gender stereotyping often replicates, reinforces and sometimes even amplifies the gender norms of the offline world. In an international study involving 500 Facebook profile photos, Tifferet and Vilnai-Yavetz found that males’ profile photos accentuated status and risk-taking, whereas females’ photos highlighted familial and emotional expression.<sup>34</sup> Women are more likely to experience negative body image, as well as face an increased risk of gender-based abuse on social media. Women politicians and journalists also face higher rates of online gender-based harassment, which can negatively impact their professional participation. Even digital advertising continues to reinforce gender stereotypes, repackaging old biases in modern formats. The Advertising Standards Authority in the UK produced a report that identifies six main stereotypes in online advertising: Roles, Characteristics, Mocking, Sexualisation, Objectification and Body image. In India, recent incidents, such as the ‘Bois Locker Room’ case,<sup>35</sup> the circulation of deepfakes targeting women public figures,<sup>36</sup> and large-scale trolling campaigns against women

---

<sup>32</sup> Soumya Awasthi, ‘From clicks to chaos: How social media algorithms amplify extremism’ (*ORF*, 13 February 2025) <<https://www.orfonline.org/expert-speak/from-clicks-to-chaos-how-social-media-algorithms-amplify-extremism>> accessed 19 December 2025

<sup>33</sup> Eduard Fosch-Villaronga, ‘Gendering algorithms in social media’ (2021) 23(1) ACM SIGKDD Explorations Newsletter <[https://kdd.org/exploration\\_files/4\\_specialsectionBIAS\\_5.pdf](https://kdd.org/exploration_files/4_specialsectionBIAS_5.pdf)> accessed 27 November 2025

<sup>34</sup> Kirsty Park et al., *The impact of the use of social media on women and girls* (European Union 2023)

<sup>35</sup> Andrew Clarence and Ayeshea Perera, ‘Bois Locker Room: Indian teens’ lewd Instagram group causes outrage’ *BBC* (05 May 2020) <<https://www.bbc.com/news/world-asia-india-52541298>> accessed 27 November 2025

<sup>36</sup> Suzie Dunn, ‘Women, Not Politicians, Are Targeted Most Often By Deepfake Videos’ (*Centre for International Governance Innovation*, 03 March 2021) <<https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/>> accessed 27 November 2025

journalists, illustrate how digital architectures enable harassment to scale rapidly and evade traditional legal remedies.

Violence against women in digital contexts also impedes women's participation in public life through humiliation, shame, fear and silencing. Women's voices are often silenced and censored by online violence. This is the chilling effect, whereby women are discouraged from actively participating in public life. Research shows that digital violence can result in women and girls having to restrict their online activity, which in turn inhibits their access to the Internet and increases the digital gender divide.<sup>37</sup> In the British newspaper *The Guardian*, Becky Gardiner found that out of the ten most targeted journalists, eight were women. This finding is especially significant, considering that only 28% of *Guardian* writers are women.<sup>38</sup> There is also the intergenerational impact of digital violence, as it serves as an impediment to young women entering professions such as politics and journalism. Koc-Michalska has found that the experience of 'mansplaining' (men explaining topics to women from an assumed position of greater knowledge) on Twitter is especially prevalent (54% of female participants said they had experienced it). The researchers suggest that this difference in men's behaviour on Twitter acts as a disincentive for women to discuss politics on the platform.<sup>39</sup>

Current laws addressing digital violence lack clear definitions and have not kept pace with different forms of violence in digital contexts. Under the *Bharatiya Nyaya Sanhita 2023*,<sup>40</sup> cyber harassment against women is covered under the provisions of Sexual Harassment,<sup>41</sup> Stalking,<sup>42</sup> Threats in the form of Criminal Intimidation,<sup>43</sup> Defamation,<sup>44</sup> Online Fraud<sup>45</sup> and

---

<sup>37</sup> S Oates et al., 'Running While Female: Using AI to Track How Twitter Commentary Disadvantages Women in the 2020 U.S. Primaries' (2019) SSRN <<https://www.semanticscholar.org/paper/Running-While-Female%3A-Using-AI-to-Track-how-Twitter-Oates-Gurevich/25e2f4c0362e0d1ed4c048c3920cf536bdf16994>> accessed 27 November 2025

<sup>38</sup> Becky Gardiner, "'It's a terrible way to go to work:' what 70 million readers' comments on the *Guardian* revealed about hostility to women and minorities online' (2018) 18(4) *Feminist Media Studies* <<https://doi.org/10.1080/14680777.2018.1447334>> accessed 19 December 2025

<sup>39</sup> Karolina Koc-Michalska et al., 'From Online Political Posting to Mansplaining: The Gender Gap and Social Media in Political Discussion' (2021) 39(2) *Social Science Computer Review* <[https://www.researchgate.net/publication/335588906\\_From\\_Online\\_Political\\_Posting\\_to\\_Mansplaining\\_The\\_Gender\\_Gap\\_and\\_Social\\_Media\\_in\\_Political\\_Discussion](https://www.researchgate.net/publication/335588906_From_Online_Political_Posting_to_Mansplaining_The_Gender_Gap_and_Social_Media_in_Political_Discussion)> accessed 27 November 2025

<sup>40</sup> *Bharatiya Nyaya Sanhita 2023*

<sup>41</sup> *Ibid* s 75

<sup>42</sup> *Ibid* s 78

<sup>43</sup> *Ibid* s 351

<sup>44</sup> *Ibid* s 356

<sup>45</sup> *Ibid* ss 318, 319

Extortion.<sup>46</sup> Though these provisions address online harm, they punish individual offenders and discrete acts of harm. The Information Technology Act, 2000,<sup>47</sup> addresses online content harm through Section 66 that penalises offensive messages, identity theft, cheating by impersonation and violation of privacy and Section 67<sup>48</sup> criminalises transmission and publishing of obscene and offensive content. In addition, Sections 69<sup>49</sup> and 79<sup>50</sup> empower the Government to block access to harmful content and provide conditional safe-harbour protection subject to takedown notices. These provisions operate post-harm and fail to impose obligations on intermediaries to prevent amplification or recurrence of such content. In cases involving child victims, the Protection of Children from Sexual Offences Act, 2012, provides protection.<sup>51</sup> Sections 11–15<sup>52</sup> criminalise online sexual harassment, the use of children for pornographic purposes, and the storage or dissemination of child sexual abuse material. This is limited to minors and provides for individual liability. The Digital Personal Data Protection Act (DPDP Act), 2023,<sup>53</sup> is a recent legislation that introduces a rights-based framework for regulating personal data in digital spaces. By mandating consent-based processing of data<sup>54</sup> and imposing data security obligations under Section 8, this Act implicitly prevents the harm of non-consensual sharing of images, identity theft and violation of privacy. However, this Act conceptualises harm as data and fails to address the broad scope of online gender violence through algorithmic amplification and does not impose obligations on intermediaries to prevent such harm. Additionally, the algorithmic due diligence norms are overly broad and undefined.<sup>55</sup> Thus, legal responses remain predominantly *individualised*, focusing on prosecuting specific offenders. Violence against women in digital contexts is particularly challenging in cases with multiple offenders and platforms across different jurisdictions.

In examining how digital platforms mediate free speech, it becomes essential to analyse the State's own regulatory tools, particularly the blocking and takedown architecture under

---

<sup>46</sup> *Ibid* s 308

<sup>47</sup> Information Technology Act 2000

<sup>48</sup> *Ibid* s 67

<sup>49</sup> *Ibid* s 69

<sup>50</sup> *Ibid* s 79

<sup>51</sup> Protection of Children from Sexual Offences Act 2012

<sup>52</sup> *Ibid* ss 11-15

<sup>53</sup> Digital Personal Data Protection Act 2023

<sup>54</sup> *Ibid* s 6

<sup>55</sup> *Ibid* s 13(3)

Section 69A<sup>56</sup> and Section 79<sup>57</sup> of the Information Technology Act, 2000. A landmark decision of the Supreme Court in *Shreya Singhal*<sup>58</sup> upheld Section 69A<sup>59</sup> for the following reasons: First, blocking can only be resorted to where the Central Government is satisfied that it is necessary so to do. Second, such a necessity is in line with the subjects set out in Article 19(2).<sup>60</sup> Third, reasons have to be recorded in writing in such a blocking order so that they may be assailed in a writ petition under Article 226 of the Constitution.<sup>61</sup> It also protected the rights of intermediaries by upholding the 'safe harbour provision' and is not liable for third-party content hosted on their platforms under Section 79.<sup>62</sup> Section 79 is an exemption provision that exempts intermediaries from third-party information hosted on their platforms. Thus, the court read down Section 79 of the IT Act, along with Section 69A.<sup>63</sup> This decision of the apex court was premised on the ground of procedural safeguards rather than transparency, leaving the broader issue of algorithmic amplification unresolved.

However, subsequent judgments have re-examined *Shreya Singhal* in light of the growing menace of deepfakes, algorithmic recommendation systems, generative models, and automated content moderation. In a recent case of *X v Union of India*,<sup>64</sup> X Corp challenged the state's power to issue take-down notices under Section 79(3)(b).<sup>65</sup> The Karnataka High Court observed that algorithms are often presented as neutral, mathematical systems that generate outcomes free from human bias. This is the myth; in reality, algorithms are conceived, designed and trained by human beings. Every line of the code is a part of the creation of the creator, while noting the role played by social media intermediaries. The Court in this decision upheld Section 79(3)(b) of the IT Act, in conjunction with Rule 3(1)(d) of the IT Rules, 2021.<sup>66</sup> This decision fundamentally re-contextualises the intermediary safe harbour, making its enjoyment strictly conditional: the exemption from civil and criminal liability now hinges on the intermediary's compliance with government takedown notices. Crucially, this is operationalised by a 36-hour window for notified content. This timeline

---

<sup>56</sup> Information Technology Act 2000, s 69A

<sup>57</sup> *Ibid* s 79

<sup>58</sup> *Shreya Singhal v Union of India* AIR 2015 SC 1523

<sup>59</sup> Information Technology Act 2000, s 69A

<sup>60</sup> Constitution of India 1950, art 19

<sup>61</sup> *Ibid* art 226

<sup>62</sup> Information Technology Act 2000, s 79

<sup>63</sup> *Ibid* s 69A

<sup>64</sup> *X Corp v Union of India and Ors* (2025) SCC OnLine Kar 19584

<sup>65</sup> Information Technology Act 2000, s 79(3)(b)

<sup>66</sup> Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, r 3(1)(d)

reflects the Court’s acknowledgement of the velocity and scale of modern digital harm – say, a deepfake video that can destroy a person’s life, dignity and privacy. The Court, by upholding such a timeline, has ensured minimal harm is done by such content. To facilitate this rapid compliance while ensuring transparency, the Court also upheld the creation of the Sahyog portal as a means of swift communication and coordination between state agencies and intermediaries. This ruling is an evolution of the Indian jurisprudence in the changing times. However, this ruling doesn’t look into the scope of intermediary accountability but notes the harm of algorithms.

In *Ranveer Allahbadia (2025)*,<sup>67</sup> the Supreme Court—citing *Apoorva Arora v UOI*<sup>68</sup>—emphasised balancing free speech with public morality, noting that vulgar content must be regulated to protect impressionable children. This video was taken after 2-4 days of the controversy, after amassing millions of views. Similarly, in *Samay Raina (2025)*,<sup>69</sup> involving offensive jokes about blind persons and a child with SMA, the Court held that while social media commercialises speech, it cannot do so at the cost of others’ dignity. However, a significant gap in these recent decisions of the apex court that evaluate harmful online content at the level of individual speakers, but have not engaged with the opaque algorithmic systems of intermediaries that determine visibility, dissemination, and harm at scale.

Taken together, the existing statutory framework and judicial decisions reveal a consistent pattern that cyber harassment is viewed through the lens of individual harm and culpability, rather than a systemic product of platform architecture and algorithmic governance. Platforms are understood as passive hosts of information, while the potential of the speed and scale of these platforms is overlooked. The existing framework is fragmented, reactive and ill-suited to the present times. This calls for a shift towards passive to active intermediaries. Such a shift is possible only through a recalibration of legislative intent – one

---

<sup>67</sup> Manasi Shah, ‘Supreme Court slams youtuber Ranveer Allahbadia for ‘obscene’ remarks, grants interim protection’ (*Supreme Court Observer*, 18 February 2025) <<https://www.scobserver.in/journal/supreme-court-slams-youtuber-ranveer-allahbadia-for-obscene-remarks-grants-interim-protection/>> accessed 01 December 2025

<sup>68</sup> *Apoorva Arora and Anr Etc v State (Govt of Nct of Delhi) and Anr* (2024) INSC 223

<sup>69</sup> Advay Vora, ‘Samay Raina, Sonali Thakkar and other comedians must publish unconditional apology online, Supreme Court Orders’ (*Supreme Court Observer*, 25 August 2025) <<https://www.scobserver.in/journal/samay-raina-sonali-thakkar-and-other-comedians-must-publish-unconditional-apology-online-supreme-court-orders>> accessed 03 December 2025

that moves beyond mere liability and towards meaningful accountability for the design and operation of intermediary systems.

## **A CASE FOR INTERMEDIARY ACCOUNTABILITY**

The foregoing analysis makes evident that gendered harassment online is inseparable from the technological architectures that structure digital interaction. This calls for a shift towards recognising and enforcing intermediary responsibility.

The Digital Services Act 2022<sup>70</sup> of the European Union is considered one of the first laws to regulate social media platforms and an attempt to hold them accountable for their intermediaries. Under Article 15, all providers are required to make publicly available annual reports on the content moderation that they are engaged in. These reports must include information about the moderation initiative, including information relating to illegal content, use of automated tools, training measures, and complaints received under complaints-handling systems. Also, Article 45 mandates the platforms to share information on any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review, as well as the rules of procedure of their internal complaint-handling system. Such transparency reports and disclosure mechanisms help in understanding how gendered harm circulates online, which helps in effective policy responses to address this challenge. However, transparency alone does not reduce the spread of harmful content, and the implementation falls short of imposing obligations on intermediaries to prevent the spread of such content.

China's Regulations on the Administration of Internet Information Service Recommendation Algorithms, 2022, also known as the Algorithmic Regulation Act, provides that algorithms must be regularly assessed for efficacy, fairness and security, as well as curtailing content-generating' algorithms' (including automated translation, news article-generation, and 'deepfakes'). China's new Regulations on algorithms are striking in terms of their scope, strength and ambition.<sup>71</sup> This legislation is of significance as it directly addresses the challenge of opaque algorithms and mandates transparency and audit obligations to reduce

---

<sup>70</sup> EU Digital Services Act 2022

<sup>71</sup> Steven Rolf, *CHINA'S REGULATIONS ON ALGORITHMS* (Friedrich-Ebert-Stiftung 2023)

gendered harm. Though this legislation is comprehensive, it is criticised as state censorship of content as it is tied to state values.

The European Union Artificial Intelligence Act 2022<sup>72</sup> prohibits biometric categorisation systems from inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation), preventing the use of AI in algorithmic generation.<sup>73</sup> This legislation addresses a critical regulatory gap by rejecting the narrative that algorithms are merely automated outcomes, beyond platform control. This Act is being implemented in phases, and the proper regulatory framework is still emerging.

These global legislation models can be adapted in India with context-specific modifications, considering the different socio-cultural conditions. India's regulatory model must align with India's constitutional values that balance freedom of speech and reasonable restrictions, as provided under Article 19.<sup>74</sup> Regulatory enforcement in India would require strengthening the current institutional frameworks by increasing human resources, training them for risk assessment and audits and establishing a dedicated digital regulator to serve this purpose. India's diverse linguistic and cultural diversity requires content governance to avoid bias against any language or rural areas, and to strengthen the unity of the country.

## FINDINGS

In light of these global legislations, India must move towards enacting a comprehensive law that reflects the standards articulated by the Forum on Information and Democracy.<sup>75</sup> Such legislation should require intermediaries to comply with domestic and international human-rights frameworks, ensure that content moderation remains legitimate, necessary and proportionate under Article 19(3) and Article 19(2), and adhere to the '5 Cs' of moderation - customisation, compatibility, community, care, and centralisation-free. It must legally mandate evaluation of the sender's influence, authorial intent, contextual factors, likelihood of harm, dissemination scale and contribution to public debate before any moderation action.

---

<sup>72</sup> EU Artificial Intelligence Act 2024

<sup>73</sup> 'The EU Artificial Intelligence Act: Up-to-date developments and analyses of the EU AI Act' (*EU Artificial Intelligence Act*) <<https://artificialintelligenceact.eu/>> accessed 20 November 2025

<sup>74</sup> Constitution of India 1950, art 19

<sup>75</sup> 'Accountability regimes for social networks and users: Shifting from immunity regimes to accountability regimes online' (*Forum on Information & Democracy*, 22 September 2022) <<https://informationdemocracy.org/publications/accountability-regimes-for-social-networks-and-users/>> accessed 20 November 2025

Further, the law should compel full algorithmic transparency, disclosure of logic, criteria, training data, error rates, biases, corrective procedures and the number of moderation decisions to an independent statutory body, and create a simplified, standardised notice-and-action system with real-time updates to complainants. Crucially, platforms whose algorithms amplify harmful content must be treated as content producers and held liable for resulting harm, supported by gender-inclusive civil-society oversight, periodic impact assessments, external audits and a requirement of adequate country-specific human moderation. The statute should also require the detection of artificial virality and impose proportionate sanctions, including fines of up to 6% of the company's global turnover, in line with the European Union's DSA, 2022. Such an accountability-focused law is essential for India to meaningfully address online gendered harm and regulate intermediary power.

## **CONCLUSION**

This paper analyses the emergence and growth of online gender harassment and its profound impact on the lives of women. It highlights how the algorithmic architecture of online intermediaries can inadvertently amplify misogyny, perpetuate bias, and create systemic barriers to women's participation. The findings demonstrate the urgent need for intermediary accountability. Implementing such measures would not only mitigate the scale and intensity of online harassment but also foster safer, more inclusive digital environments.

While existing laws in India address individual acts of harassment, they fail to regulate the technological structures that facilitate and magnify such harm. This gap underscores the necessity for comprehensive legislation that combines rights-based approaches with technological accountability. Implementing algorithmic accountability in India could expand the scope of the fundamental right to life and liberty to include the 'right to feel safe online'. By centering both human and algorithmic responsibility, India can take meaningful steps towards addressing online gendered harm and protecting the dignity and safety of women in digital spaces.