Improving Online Safety for Women and Children in India:

An assessment of current Indian legislative framework, global regulatory approaches, and potential pathways

Prepared by The Quantum Hub (TQH), India

Funding Support by Meta and YouTube

Table of Contents

Acknowledgements

*This research and its accompanying analysis would not have been possible without the generous funding support from **Meta** and **YouTube**. We extend our sincere gratitude to both organizations for their commitment to fostering critical discussions in this evolving landscape.*

We are also deeply thankful to all the experts and participants who contributed their invaluable insights and perspectives during the various roundtable discussions. Your engagement significantly enriched this work.

Authors & Contributors

Authors:

- Rhydhi Gupta
- Senu Nizar
- Akanksha Ghosh
- Manan Katyal
- Tithi Neogi
- Devika Oberai
- Nikhil Iyer

Design Inputs and Project Coordination:

- Kashvi Verma

Research Directed and Edited By:

- Aparajita Bharti
- Sidharth Deb

Disclaimer

At The Quantum Hub (TQH), we are committed to transparency in all our work. We consistently disclose the sponsors or funders of any commissioned research. When a project is supported by a funder, we clearly identify them and acknowledge their role in shaping the research scope, providing methodological inputs, and sharing feedback. However, TQH retains full editorial independence and takes sole responsibility for the final research output, including its analysis and recommendations.

All insights and analysis presented in this publication are solely attributable to The Quantum Hub (TQH) and the named authors. Any errors or omissions are exclusively our own. This report was prepared between December 2024 and August 2025. Given that laws, regulations, and policies are subject to continuous updates, some of the views expressed here may reflect the understanding at the time of writing and could evolve as new developments emerge. In particular, where we discuss recent or nascent policies, the analysis should be read as preliminary, with the recognition that their full implications will only become clear over time.

Suggested Citation

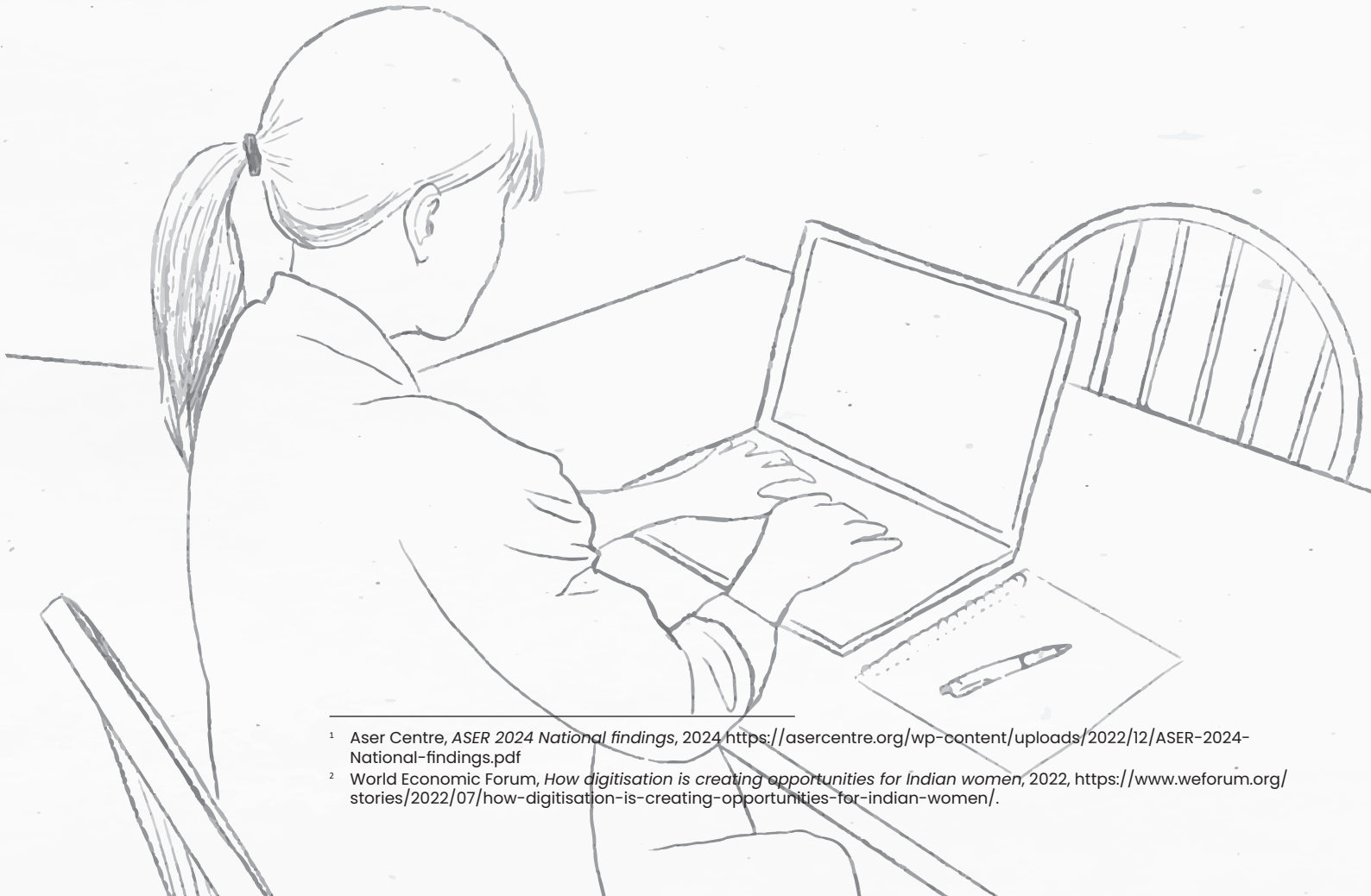
TQH Consulting. **(2025, October)**. *Improving online safety for women and children in India: An assessment of current Indian legislative framework, global regulatory approaches, and potential pathways.*

List of Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
BNS	Bharatiya Nyaya Sanhita, 2023
CCPWC	Cyber Crime Prevention against Women and Children
CSEA	Child Sexual Exploitation and Abuse
CSEAM	Child Sexual Exploitation and Abuse Material
CSAM	Child Sexual Abuse Material
CSO	Civil Society Organisation
DPDPA	Digital Personal Data Protection Act, 2023
FIR	First Information Report
GBV	Gender-Based Violence
I4C	Indian Cyber Crime Coordination Centre
ICRW	International Center for Research on Women
IO	Investigating Officer
IPC	Indian Penal Code, 1860
IT Act	Information Technology Act, 2000
ITPA	Immoral Traffic (Prevention) Act, 1956
JCCT	Joint Cyber Crime Coordination Team
JJB	Juvenile Justice Board
LEA	Law Enforcement Agency
LGBTQI+	Lesbian, Gay, Bisexual, Transgender, Queer, Intersex and related identities
MLAT	Mutual Legal Assistance Treaty
MMS	Multimedia Messaging Service
MoU	Memorandum of Understanding
NCII	Non-Consensual Intimate Imagery
NCRB	National Crime Records Bureau
NCPCR	National Commission for Protection of Child Rights
NCW	National Commission for Women
NCMEC	National Center for Missing and Exploited Children
NFHS	National Family Health Survey
NHRC	National Human Rights Commission
Ofcom	Office of Communications
Ofsted	Office for Standards in Education, Children's Services and Skills
OSA	Online Safety Act
OSAEC	Online Sexual Abuse and Exploitation of Children
PICACC	Philippine Internet Crimes Against Children Center
POCSO	Protection of Children from Sexual Offences Act, 2012
POSH	Prevention of Sexual Harassment Act, 2013
RPH	Revenge Porn Helpline
SMIs	Social Media Intermediaries
SSMIs	Significant Social Media Intermediaries
SOP	Standard Operating Procedure
TFGBV	Technology-Facilitated Gender-Based Violence
UK	United Kingdom
UN	United Nations
UNICEF	United Nations Children's Fund
US	United States
VAW	Violence Against Women
VAWC	Violence Against Women and Children
VLOPs	Very Large Online Platforms
VLOSEs	Very Large Online Search Engines
VPN	Virtual Private Network

Executive Summary

This report examines the complex and evolving online harms faced by women and children in India, situating them within both the country's rapid digital growth story and its entrenched socio-cultural realities. Digital spaces have ushered in new possibilities for children and women, offering access to opportunities that were once unfeasible. For young people, nation-wide surveys have disclosed that the internet can provide opportunities for learning, self-expression, skill development, and community building[1]. For women, the World Economic Forum highlights that it can expand access to professional opportunities, enable entrepreneurial ventures, strengthen social networks, and offer platforms for civic participation and advocacy[2]. The need to design an online ecosystem that preserves and enhances the benefits of digital engagement while systematically reducing the risks is therefore critical. The issue of online trust and safety has been a major topic of policy discussions in India - previously with discussions around a proposed 'Digital India Act', as well as more recent discussions around cybercrime that are stemming from different Parliamentary committees.



[1] Aser Centre, *ASER 2024 National findings*, 2024 https://asercentre.org/wp-content/uploads/2022/12/ASER-2024-National-findings.pdf

[2] World Economic Forum, *How digitisation is creating opportunities for Indian women*, 2022, https://www.weforum.org/stories/2022/07/how-digitisation-is-creating-opportunities-for-indian-women/.

While India's internet user base exceeds 965 million, its regulatory, legal and enforcement ecosystem for online safety remains reactive, fragmented, and often triggered in response to high-profile incidents. Women and children experience distinct and intersecting forms of risk, ranging from clearly illegal acts to legally ambiguous harms. This report examines the nature, scope, and systemic drivers of online harms faced by women and children and proposes actionable ecosystem-wide reforms. It is the culmination of a multi-stakeholder consultation process including two roundtables with civil society organisations (CSOs) working with survivors of online harms, focus group discussions and listening conferences with children and teenagers, and extensive secondary research.

The analysis begins by mapping the evolving risk landscape for both groups, highlighting how anonymity, permanence of content, and cross-platform abuse compound the impact of technology-facilitated gender-based violence, while children face additional threats from grooming, peer-to-peer abuse, harmful online communities, and challenges associated with excessive screen time. The report also offers a critical examination of India's capacity to measure such harms. It finds that while official crime statistics capture some cybercrimes, the absolute numbers of cybercrimes and online risks that are reported in India remain disproportionately low. The absence of robust taxonomies, disaggregated and intersectional data, and standardised reporting frameworks significantly limits the ability of policymakers and law enforcement agencies to assess the magnitude of harms and design targeted interventions.

The report also assesses the adequacy of India's laws, policies, and enforcement systems, reviewing the Information Technology (IT) Act, potentially applicable provisions under criminal laws like the Bharatiya Nyaya Sanhita, child protection statutes, general laws that apply to women's safety, etc. The report similarly analyses platform obligations that emanate from India's legal frameworks. The report finds that legislative design is unable to adequately distinguish between cyber-enabled and cyber-dependent offences leading to a difficulty in capturing the scope and scale of cyber threats faced by women and children in India. The report also finds that IT legislation perpetuates unintentional incentives where first responder cops are required to pursue cyber investigations through general criminal laws. Additionally, enforcement practices are fragmented across different states, and victims' representatives / survivors encounter lengthy, unpredictable legal processes with insufficient institutional support. Online platforms, while subject to takedown obligations, have to contend with significant informal pressures during live flashpoint incidents, and are not yet structurally incentivised to allocating institutional resources towards adopting preventive safety-by-design measures that align with the lived realities of India's internet users. Moreover, India's current approach to platform obligation struggles to balance online safety with constitutional imperatives like people's rights to free speech and privacy. Civil society organisations, while being critical actors in survivor support and awareness, suffer from a lack of institutional recognition and support despite their reach and trust within communities.

Drawing on comparative analysis of international approaches to online safety for women and children, the report distils lessons that are transferable to the Indian context. The recommendations present a systemic roadmap that promotes an all-of-ecosystem approach to the online safety of women and children, since interventions should by-design be a shared responsibility where stakeholders can all work towards a common objective.

The recommendations prioritise proactive risk assessment and prevention over reactive enforcement. These include modernised policy definitions for both cybercrimes and legal but harmful cyber risks affecting women and children. Some other key recommendations also highlight: (a) the importance of disaggregated data collection by policy and law enforcement institutions; (b) strengthening platform accountability through systemic risk assessment and disclosure requirements; and (c) introducing gender-sensitive and child-specific legal provisions. The report further calls for standardised law enforcement protocols, investment in training and capacity building, and the formal integration of CSOs into reporting, support, and policy processes. Additionally, the report calls for concerted efforts by industry and civil society to work together to build platform safety codes of practices that create a shared understanding of best practices around safe and inclusive platform design, and parallelly enable stakeholder coordination to prevent cross-platform abuse.

The study concludes that safeguarding women and children online requires a coordinated approach where legislative clarity, institutional capacity, platform responsibility, and survivor-centred support work in concert. Achieving this demands not only reforms in law and enforcement but also cultural change, cross-sector collaboration, and sustained investment in prevention and rehabilitation. By embedding these principles into policy and practice, India can shift from a reactive posture to one that systematically mitigates harm, ensuring safer and more inclusive digital spaces for its most vulnerable users.

Chapter 1
Background and Objectives

1.1. Background

India, with a staggering 969.60 million internet users as of 2024,[3] is witnessing rapid digital transformation. Rising internet adoption creates many opportunities but also risks. Online safety risks are heterogeneous and their extent varies based on digital and media literacy, gender, age, socio-economic background, religious/caste/regional identity, and location. Indian laws largely address online risks under the Information Technology Act, 2000 ("IT Act") as well as general criminal laws.

Despite incremental amendments since the enactment of the IT Act in the year 2000[4], questions have organically emerged about its suitability to address newer online safety risks, especially with the growing availability and pervasive use of many digital services. In recent years, India has had various discussions on replacing the IT Act with a new **Digital India Act**, given the former's limited provisions on user rights, trust and safety, and its inability to address emerging and complex forms of cybercrime such as doxxing, cyberstalking, and online harassment.[5] While the momentum for a new Digital India Act has since receded, different Parliamentary Committees have increasingly taken up issues of cybersecurity, cybercrimes and online safety. For example in **August 2025, the Parliamentary Committee on Empowerment of Women** recently engaged with social media companies on the issue of women's cyber safety, amid growing concerns over online harassment, stalking, trolling, and the misuse of digital platforms to target women.[6] The same month saw the **Parliamentary Standing Committee on Home Affairs publish a report on cybercrime** which called for wider reform of existing IT laws, amending the legal obligations of online intermediaries, and called for improvements across investigation and enforcement.[7]

Against this backdrop, when stakeholders revisit the suitability of India's existing online safety frameworks, specific attention is warranted towards the safety of women and children. These groups contend with specialised profiles of online risks, such as sexual harassment, grooming by online predators, reputational risks etc., and must navigate unique socio-cultural norms to reap the benefits of digital technologies. Unfortunately, India

[3] The Indian Telecom Services Performance Indicator - April to June 2024 (Available at - https://www.trai.gov.in/sites/default/files/QPIR_09102024.pdf)

[4] The IT Act was enacted in 2000 to legally recognise and facilitate cross border e-commerce transactions that became increasingly relevant during the 1990s

[5] PIB Press Release, *MoS Rajeev Chandrasekhar to hold a Digital India Dialogues' session tomorrow in Mumbai on principles of Digital India Act*, 2022, npci.org.in/PDF/npci/knowledge-center/partner-whitepapers/UPI-For-Her-Enabling-Digital-Payments-for-Women-in-India.pdf

[6] PTI News, *Parliament panel to hear social media giants on cyber safety of women*, 2025, https://www.ptinews.com/story/national/parliament-panel-to-hear-social-media-giants-on-cyber-safety-of-women/2831246

[7] Parliament of India Rajya Sabha, *254th Report on Cybercrime - Ramifications, Protection and Prevention*, 2025, https://www.medianama.com/wp-content/uploads/2025/08/rsnew_Committee_site_Committee_File_ReportFile_15_197_254_2025_8_12-1.pdf

lacks a systematic approach to online safety for women and children. There have only been *ad-hoc* amendments to the IT Act and attendant rules to address women's and children's online safety. Coincidentally, incidents involving women's safety have shaped India's intermediary liability and safe harbour protection laws that govern most digital services. For example, the 2008 reform of the IT Act brought online digital services ("intermediaries") within the scope of safe harbour liability exemptions against unlawful third party content/behaviour. The amendment was a direct result of the ***DPS MMS scandal (Bazee.com) case*** that involved the online circulation/sale of explicit content featuring a teenage schoolgirl and the subsequent arrest of the Managing Director of *Bazee.com*.[8]

The 2008 IT Act amendments also introduced cybercrime offences that deal with threats which disproportionately affect women and children. These include provisions that punish the violation of intimate privacy of individuals[9], online obscenity[10], distribution of sexually explicit materials[11], and distribution of materials depicting children performing sexually explicit acts.[12] **Yet expert commentary observes that the cybercrimes addressed in the 2008 IT Act amendments seem to be largely reactive, driven by immediate concerns rather than a comprehensive evaluation of the digital landscape's needs.[13]** Similarly, recent amendments to the IT Act's intermediary liability rules[14] also contain provisions on women and children's safety, but were a reaction to the Indian Supreme Court's directions in the *Prajwala case*[15] wherein a committee was constituted to examine technological solutions that would pre-emptively filter, block and prevent the circulation of videos depicting rape, gangrape and child pornography.

1.2. Objectives and Approach of the Study

India's reactive approach to online safety regulation often leads to ad-hoc interventions that result in fragmentation and responses to cases that become publicly prominent.[16] A reactive approach is not conducive for safety frameworks to adequately address challenges associated with the "*pacing problem*"[17] in technology policymaking. **This report aims to propose an alternative approach where safety interventions help make online spaces saf er and more accessible by-default for women and children.**

Chapter 2 traces the **type of safety risks and challenges that women and children face online**, India's ability to measure these risks (especially **cybercrimes**) and some of the **newer risks** affecting these groups.

Chapter 3 studies the suitability and identifies gaps within **India's existing law, policy and enforcement ecosystems**. We examine "***enforcement***" at the levels of **regulators, law enforcement agencies (LEAs) and online platforms**. The analysis focuses on **preventing harms, and when harm arises offering survivors and victims accessible pathways to remedy and/or access to justice.** The gap analysis targets both

---

[8] *Avnish Bajaj vs State*, 116 (2005) DLT 427

[9] Information Technology Act 2000, Section 66E

[10] Information Technology Act 2000, Section 67.

[11] Information Technology Act 2000, Section 67A.

[12] Information Technology Act 2000, Section 67B.

[13] Nappinai, N.S., *Report on the Information Technology Act, 2000 (as amended) & the need for further amendments to the Act,* 2019, https://www.cybersaathi.org/report-on-the-information-technology-act-2000-as-amended-it-act-the-need-for-further-amendments-to-the-act-july-30-2019-authored-by-ms-n-s-nappinai/.

[14] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021

[15] *Re: Prajwala letter dated 18.02.2015 Videos of Sexual Violence and Recommendations* (2018) 15 SCC 551

[16] Soumyarendra Barik, *Centre issues advisory to social media platforms over deepfakes after viral 'Rashmika Mandanna' video*, 2023, https://indianexpress.com/article/business/centre-deepfake-advisory-to-social-media-platforms-9017283/

[17] The pacing problem describes how technological innovation outpaces the ability of laws and regulations to keep up, resulting in significant ramifications for the governance of these technologies. Adam Thierer, *The Pacing Problem and the Future of Technology Regulation*, 2018, https://www.mercatus.org/economic-insights/expert-commentary/pacing-problem-and-future-technology-regulation.

law/policy design as well as current enforcement practices. Chapters 4 and 5 study **international trends on online safety for children and women, respectively**. Chapter 6 concludes the report with **final observations and recommendations**. The report's recommendations are a reflection of our insights from our **desk research, private stakeholder interactions and the insights derived from expert roundtable discussions[18] involving representatives from civil society, academia, industry and others**.

We have attempted to align the final recommendations with India's socio-cultural and on-ground realities. This includes considerations like India's (a) digital divide[19], (b) low digital media literacy[20], (c) linguistic, vernacular and regional diversity, (d) caste, tribal, sexual, religious and other forms of identity, (e) disability and accessibility, and (f) unique patterns of internet use. **Our recommendations also factor patriarchal social norms that constrain the agency of women and girls in India[21]** that add an additional layer of vulnerability to various online risks, including those to their reputation within their communities and beyond. Each of these factors contribute immensely to on-ground complexities. **Overall, it is unrealistic to assure online safety through only strong laws and regulations. Instead, it becomes important to identify systemic gaps and subsequently offer a series of corresponding safety interventions.**

In that context, a snapshot of our recommendations is provided in the table below:

Theme	Recommendation
Modernizing Policy & Regulatory Design	Emphasizes better definitions and categorization of offences, improved measurement of online safety risks, and systems for proactive risk evaluations and mitigations.
Enforcement & Capacity Building	Improve enforcement practices of LEAs and government agencies by enhancing procedural predictability, improving on-ground incentives, better training and technical capabilities, improving cyber safety capabilities of institutions dedicated to the safety of women and children, and streamlining LEA coordination with digital services.
Platform Design & Responsibility	Creating industry and civil society led codes of practices for online safety of women and children that focus on incentivizing safer product and platform design.
Cross-Ecosystem Collaboration	Promote all-of-ecosystem collaboration through partnerships across industry and civil society, especially for issues like cross-platform abuse.
Victim Support & Rehabilitation	This includes policy, platform and institutional interventions that help provide resources and support for those who have been affected by online safety risks.

These recommendations aim to equitably distribute the **roles and responsibilities of each stakeholder**, where all actors collectively work towards the common objective of keeping India's women and children safe online.

[18] LinkedIN, *TQH Roundtable on Online Safety*, 2025, https://www.linkedin.com/posts/thequantumhub_onlinesafety-digitalsafety-womensafety-activity-7320341066346348546-sZ-E?utm_source=share&utm_medium=member_desktop&rcm=ACoAADc1JXIBgdbmjixRuxCROzYrHjgpD-HRkc8

[19] People's Archive of Rural India, *Digital Divide: India Inequality Report 2022*, 2022, https://ruralindiaonline.org/en/library/resource/digital-divide-india-inequality-report-2022/.

[20] Gupta, Jana, Maiti & Y., *Gender-gap in Internet Literacy in India: a state-level analysis*, 2023, https://www.researchgate.net/publication/374951642_Gender-Gap_in_Internet_Literacy_in_India_A_State-Level_Analysis.

[21] NORC at the University of Chicago and the International Center for Research on Women, *Case Study: Technology-facilitated Gender Based Violence in India*, 2022,https://www.icrw.org/wp-content/uploads/2021/09/USAID-TFGBV-India.pdf.

Chapter 2

Profiling the Online Risk Environment for Women and Children

2.1. Preliminary Considerations

A comprehensive understanding of online risks affecting women and children is essential for appropriate evidence-based online safety interventions. This requires both a **qualitative profile** of existing and emerging types of risk. Additionally, safety measures should be informed via adequate **quantitative measures** of the extent of risks under each category.

Different enforcement actors – namely regulators, LEAs and online platforms – benefit from a commonly understood taxonomy of risks as it helps them with targeted enforcement and resource allocation. These measures also assist enforcement actors implement accessible survivor/victim centric

redressal mechanisms – ultimately facilitating safer online experiences.

When evaluating the tapestry of online risks faced by women and children it is important to consider their unique online and offline realities. The complex nature of these risks can broadly be categorised into risks arising due to **clearly illegal and criminal activities** – which may or may not be viewed as ***"cybercrimes"***, along with more **legally ambiguous activities/ experiences which contribute to harmful outcomes** for individuals.

Risk identification must also contend with the **interplay between online and offline environments**. Harms or threats might arise in the online space (e.g. grooming) but might lead to a physical manifestation in offline settings. Thus, we need to look at both environments as interconnected and mutually influential. For instance, if girls receive unsolicited messages or images from boys at their school, these events unfold both on the social media platform and within the school and classroom settings where they interact.[22] For women, tech-facilitated violence[23] (for instance, image-based sexual abuse) is often used as a tool for blackmail, coercing survivors into remaining in a toxic relationship, providing additional images, or engaging in sexual acts - indicating that something that starts with technology ultimately translates into harmful consequences in the offline world.[24] Regulators have recognised the influence of both online and offline environments in defining cybercrime, with global frameworks creating nuanced distinctions between cyber-enabled and cyber-dependent crimes to reflect the dynamic interplay of different environments in shaping cybercrime.[25]

2.1.1. Aligning Safety Interventions with Social Realities

The internet often embodies and amplifies the values and power dynamics that perpetuate across societies and institutions.[26] As mentioned in Chapter 1, safety interventions must consider India's unique socio-economic tapestry.

Online safety frameworks must also align with evolving usage patterns. For example, while there is considerable commentary on harms by adults to children online, young internet users are also often subjected to harm as a result of peer-to-peer abuse via other people under the age of 18 years.[27] These can manifest into harmful online experiences like online bullying, online harassment, online sexual abuse and exploitation.[28] Similarly, discussions on children's online safety must also contend with the fact that children at-risk are often subjected to abuse/ harm by family members, relatives and other people in their social proximity.[29]

2.2. The Expanding Scope of Online Risk Landscape(s)

Online risk environments are evolving due to the **transformation of media and information consumption** – defined by content curation, and sequencing practices, as well as interactive and immersive P2P environments. They have combined to

---

[22] European Parliament, *The impact of the use of social media on women and girls*, 2023, https://www.europarl.europa.eu/RegData/etudes/STUD/2023/743341/IPOL_STU(2023)743341_EN.pdf.

[23] UN Women, *FAQs: Digital abuse, trolling, stalking, and other forms of technology-facilitated violence against women*, February 2025 https://www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women

[24] Megan O'Brien, *Online violence: real life impacts on women and girls in humanitarian settings*, 2024, https://blogs.icrc.org/law-and-policy/2024/01/04/online-violence-real-life-impacts-women-girls-humanitarian-settings/#:~:text=Women%20and%20adolescent%20girls%20also,in%20some%20cases%20honor%20killing.

[25] United Kingdom Home Office, *Cyber crime: A review of the evidence*, 2013, https://www.gov.uk/government/publications/cyber-crime-a-review-of-the-evidence.

[26] Brandee Easter, *'Feminist_brevity_in_light_of_masculine_long-windedness:' Code, Space, and Online Misogyny*, 2018, https://www.tandfonline.com/doi/full/10.1080/14680777.2018.1447335.

[27] Gill, Monk & Day, *Qualitative research project to investigate the impact of online harms on children*, 2022, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1167838/Online_Harms_Study_Final_report_updated_51222_updated_290623.pdf.

[28] United Nations Office on Drugs and Crime, *Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children*, 2015, https://www.unodc.org/documents/Cybercrime/Study_on_the_Effects.pdf.

[29] Katz & Asam in partnership with Internet Matters, *Vulnerable Children in a Digital World*, 2019, https://www.internetmatters.org/wp-content/uploads/2019/04/Internet-Matters-Report-Vulnerable-Children-in-a-Digital-World.pdf

create wholly unexpected outcomes with regard to human outlook and behaviour.[30] For women and children, specific risks are coming up across a broad range of issues. According to global literature this includes:

- Major privacy breaches and abuse in the form of non-consensual intimate imagery (NCII), doxxing and other means[31],
- Targeted trolling, harassment, bullying and hate speech which can impact people from marginalised communities[32]; public figures like journalists[33], actors[34], politicians[25] and athletes[36]; and risks withdrawal from various online spaces[37].
- Reinforcing of sexist stereotypes[38],
- Cognitive distortions by being fed a constant stream of specific topics[39],
- The perpetuation of eating disorders[40] and negative body image perceptions (largely in girls and young women)[41].

The discussions on risky online behaviour by young people escalated in India (*circa 2017*) over the *Blue Whale Challenge*.[42] The trend reportedly saw many young people engaging in self-harm and even dying by suicide. The issue was so prominent that Indian authorities issued formal warnings and regularly engaged with internet companies on the same.[43] Additionally, the **social and economic gains associated with visibility and virality on online platforms** has led to risky user behaviour among children / adolescents[44], and even parents (see: *sharenting*[45]). Possibly recognising risks with the latter, India's National Commission for Protection of Child Rights (NCPCR) released guidelines enumerating safeguards during the production of online content that involves the participation of children (under 14s) and adolescents (between the age of 14 and 18 years).[46]

Young people may be vulnerable to the **distribution and consumption of**

---

[30] Valkenburg & Piotrowski, *Plugged In: How Media Attract and Affect Youth*, 2017, https://drupal.yalebooks.yale.edu/sites/default/files/files/Media/9780300228090_UPDF.pdf; Amnesty International, *Driven into Darkness: How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation*, 2023, https://www.amnesty.org/en/documents/POL40/7350/2023/en/.

[31] United Nations General Assembly Human Rights Council at its 38th Session, Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective, 2018,https://www.ohchr.org/en/documents/thematic-reports/ahrc3847-report-special-rapporteur-violence-against-women-its-causes-and.

[32] Chowdhury & Lakshmi, *"Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI*, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000387483.

[33] Posetti, Shabbir, Maynard, Bontcheva & Aboulez, *The Chilling: global trends in online violence against women journalists; research discussion paper*, 2021, https://unesdoc.unesco.org/ark:/48223/pf0000377223.

[34] Sharma, Sultana, Alam & Banu, *Trolling as a Disruptive Tool for Human Rights Violations: An Exploration of the Challenges Faced by Performance Artists,* 2024, https://www.sciedupress.com/journal/index.php/wjel/article/view/25690.

[35] Center for Countering Digital Hate, *Abusing Women in Politics: How Instagram is failing women and public officials*, 2024, https://counterhate.com/research/abusing-women-in-politics/?utm_source=substack&utm_medium=email.; European Parliamentary Research Service, *Violence against women active in politics in the EU: A serious obstacle to political participation*, 2024, https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/759600/EPRS_BRI(2024)759600_EN.pdf.

[36] World Athletics, *Key Findings: Online Abuse in Athletics*, 2024, https://worldathletics.org/news/press-releases/four-year-analysis-online-abuse-athletics

[37] Centre for Strategy and Evaluation Services, *Rapid Evidence Assessment: The Prevalence and Impact of Online Trolling*, 2019, https://assets.publishing.service.gov.uk/media/60607f64d3bf7f717df98839/DCMS_REA_Online_trolling__V2.pdf.

[38] United Nations Regional Information Centre for Western Europe, *How technology-facilitated gender-based violence impacts women and girls*, 2023, https://unric.org/en/how-technology-facilitated-gender-based-violence-impacts-women-and-girls/.

[39] Kathy Katella, *How Social Media Affects Your Teen's Mental Health: A Parent's Guide*, 2024, https://www.yalemedicine.org/news/social-media-teen-mental-health-a-parents-guide.

[40] Ofcom, *Online Content Qualitative Research: Experiences of children encountering online content relating to eating disorders, self-harm and suicide*, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/keeping-children-safe-online/experiences-of-children/experiences-of-children-encountering-online-content-relating-to-eating-disorders-self-harm-and-suicide.pdf?v=368019.

[41] Amelia Hill, *Social media triggers children to dislike their own bodies*, says study, 2023, https://www.theguardian.com/society/2023/jan/01/social-media-triggers-children-to-dislike-their-own-bodies-says-study.

[42] Ant Adeane, *Blue Whale: What is the truth behind an online 'suicide challenge'?,* 2019, https://www.bbc.com/news/blogs-trending-46505722.

[43] Economic Times, *IT Ministry asks Google, Facebook, WhatsApp and Instagram to remove Blue Whale game links*, 2017, https://economictimes.indiatimes.com/magazines/panache/it-ministry-asks-google-facebook-whatsapp-and-instagram-to-remove-blue-whale-game-links/articleshow/60070772.cms?from=mdr.

[44] Al Jazeera, *Top US health official warns of social media's risk to children,* 2023, https://www.aljazeera.com/news/2023/5/23/top-us-health-official-warns-of-social-medias-risk-to-children.

[45] Keskin, Kaytez, Damar, Elibol & Aral, *Sharenting Syndrome: An Appropriate Use of Social Media?,* 2023, https://pmc.ncbi.nlm.nih.gov/articles/PMC10218097/

[46] National Commission for Protection of Child Rights, *Guidelines for Child and Adolescent Participation in the Entertainment Industry and any Commercial Entertainment Activity*, 2023, https://ncpcr.gov.in/public/uploads/16844053596465fc6f115d1_guidelines-for-child-and-adolescent-participation.pdf.

child sexual exploitative and abuse material over online platforms.[47] Additionally, the transformation of individual interactions using online platforms have also led to **online child sexual exploitation activities** which includes child luring, inappropriate contact, or grooming via deceptive activities like gifting that ostensibly serves as a precursor to abuse.[48] The challenges to track and prevent such activities that threaten children are complicated by the fact that **end-to-end encrypted messaging platforms** can be used by perpetrators with reduced detection.[49]

Children also are navigating large amounts of **age inappropriate content** and information across several different digital platforms. This can manifest in exposure to terrorist/extremist content[50], information streams that encourage unhealthy eating habits[51], and other content that is **misaligned with a young person's age or cognitive development stage.**

2.3. Profiling Types of Online Risks

The contemporary trends described in the previous section serves as a backdrop to our forthcoming analysis. This section offers a **theoretical and quantitative overview** of the profile of online risks faced by women and children in India. The analysis presents a snapshot of India's ability to measure online risks to women and children via formal institutions like the **National Crime Records Bureau (NCRB)**. It also divides the analysis on the basis of **(a) illegality (e.g. cybercrimes), and (b) legally ambiguous risks** that nevertheless cause negative or harmful outcomes for women and children.

2.3.1. Online Risks Faced by Women ("Technology-Facilitated Gender Based Violence")

When evaluating challenges like technology-facilitated gender based violence (TFGBV), it is essential to situate discussions within offline realities. **This helps ascertain the balance required between online and offline interventions**.

According to India's 5th National Family Health Survey (NFHS– 2019-21) about 32% of women between 18 – 49 years experience some form of spousal or physical violence during pregnancy.[52] One study of the NFHS data suggests that these **risks tend to differ across various parameters and situations**. This includes age, education level, marital (or divorce) status, caste, family wealth status, employment status, state/regionality, alcohol consumption, and so on.[53] More generally, women are at higher risk of murder, stalking, sexual violence, harassment at public/social events and work places, and are often vulnerable to gendered violence and rape in conflict environments.[54] The same vulnerabilities are felt by women online as gender based violence remains persistent in all spheres of society.

TFGBV has been defined as "...*any action carried out using the internet and/or mobile technology that harms others based on their sexual or gender identity or by enforcing harmful*

---

[47] National Center for Missing and Exploited Children, *2024 Cybertipline Report*, 2024, https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata.
Internet Watch Foundation, *IWF Annual Report 2023*, 2023, https://www.iwf.org.uk/annual-report-2023/.

[48] Özçalık & Atakoğlu, *Online child sexual abuse: Prevalence and characteristics of the victims and offenders*, 2021, https://jag.journalagent.com/phd/pdfs/PHD-30643-REVIEW-KARA_OZCALIK%5BA%5D.pdf.

[49] Teunissen & Napier, *Child sexual abuse material and end-to-end encryption on social media platforms: An overview*, 2022, https://www.aic.gov.au/sites/default/files/2022-07/ti653_csam_and_end-to-end_encryption_on_social_media_platforms-v2.pdf.

[50] UNICEF Innocenti, *Children's exposure to hate messages and violent images online,* 2023, https://www.unicef.org/innocenti/documents/childrens-exposure-hate-messages-and-violent-images-online.

[51] Carroll, Edmund, Griffin, Bertone-Johnson, VanKim & Sturgeon, *Children's Perception of Food Marketing Across Digital Media Platforms*, 2024, https://www.sciencedirect.com/science/article/pii/S2773065424000245.

[52] Government of India - Ministry of Health & Family Welfare, *Compendium of Fact Sheets: Phase II NFHS 5*, 2021, https://mohfw.gov.in/sites/default/files/NFHS-5_Phase-II_0.pdf.

[53] Population Research Centre at the Gokhale Institute of Politics & Economics, *Gender-based Violence: A Shred of Evidence from NFHS 5*, 2023, https://gipe.ac.in/wp-content/uploads/2024/03/Gender-Based-Violence-A-Shred-of-Evidence-from-NFHS-5.pdf.

[54] World Health Organization, *Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence*, 2013, https://www.who.int/publications/i/item/9789241564625.; Constance Backhouse, *Sexual Harassment: A Feminist Phrase that Transformed the Workplace*, 2012, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2268095.; Vera-Gray & Kelly, *Contested gendered space: public sexual harassment and women's safety work*, 2020, https://www.tandfonline.com/doi/full/10.1080/01924036.2020.1732435.

gender norms."[55] Experts describe TFGBV as modernised GBV where digital technologies are used to cause harm.[56] International institutions observe that as COVID-19 triggered the need for people to shift online, there has also been a surge in TFGBV.[57] The shift to digital has allowed perpetrators to broaden the scope of violence that they enact (or perpetrate) upon their victims, and can **range across intimate partner violence, gender-based harassment, misinformation campaigns or even hate campaigns**.[58] Digital technologies have also made traditional abusive behaviours like child luring[59] or stalking[60] more accessible. New innovations like artificial intelligence are enabling newer forms of abuse such as the non-consensual creation of sexual imagery, e.g. using deepfakes or virtual reality technologies.[61] Its rising prominence has even led to a recognition by major online platforms who are working on special initiatives and interventions to give women an opportunity to defend themselves against deepfake and AI-based online image abuse.[62] These initiatives are discussed in greater detail in later chapters of the report as well as the final recommendations.

Existing social norms also exacerbate TFGBV. **Two recent examples from India that highlight these trends in India are the *Boys Locker Room*[63] and *Sulli Deals*[64] online group incidents, which show how gender and religious identity can make people vulnerable to certain kinds of online abuse.** Moreover, investigative reports demonstrate that aside from abusive views, Indian women have been the victims of Non Consensual Intimate Imagery (NCII) distribution and/or online trafficking facilitated via online messaging platforms like Telegram.[65] Non-binary, transgender and other gender non-conforming persons further face exacerbated risks.

Experts believe that some factors amplify risks for women in online spaces. **These include the scope for abusers to remain anonymous, the potential for distanced/cross-**

---

[55] Hinson L, Mueller J, O'Brien-Milne L & Wandera N, *Technology-Facilitated Gender-based Violence: What is it, and how do we measure it?* 2018, https://www.icrw.org/publications/technology-facilitated-gender-based-violence-what-is-it-and-how-do-we-measure-it/. Also See: USAID, *Technology-Facilitated Gender-based Violence in Asia: India*, 2021, https://www.icrw.org/wp-content/uploads/2021/09/USAID-TFGBV-India.pdf.

[56] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf, Also see: Delanie Woodlock, *ReCharge: Women's Technology Safety, Legal Resources, Research and Training*, 2015, (https://www.researchgate.net/publication/310015457_ReCharge_Women%27s_Technology_Safety_Legal_Resources_Research_and_Training.

[57] UN Women, *COVID-19 and Ending Violence Against Women and Girls: Addressing the Shadow Pandemic,* 2020, https://www.unwomen.org/en/digital-library/publications/2020/06/policy-brief-covid-19-and-violence-against-women-and-girls-addressing-the-shadow-pandemic; Dehingia N, McAuley J, McDougal L, Reed E, Silverman JG, Urada L, Raj A, *Violence against women on Twitter in India: Testing a taxonomy for online misogyny and measuring its prevalence during COVID-19*, 2023, https://doi.org/10.1371/journal.pone.0292121.

[58] European Institute for Gender Equality, *Cyber violence against women and girls*, 2017, https://eige.europa.eu/publications-resources/publications/cyber-violence-against-women-and-girls?language_content_entity=en.; Also see: Freed, Palmer, Minchala, Levy, Ristenpart & Dell, *Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders*, 2017, https://dl.acm.org/doi/10.1145/3134681.

[59] Adriane Van Der Wilk, *Cyber violence and hate speech online against women*, 2018, https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU(2018)604979_EN.pdf.

[60] Khoo, Robertson & Deibert, *Installing Fear: A Canadian Legal and Policy Analysis of Using, Developing, and Selling Smartphone Spyware and Stalkerware Applications*, 2019, https://citizenlab.ca/docs/stalkerware-legal.pdf.

[61] Suzie Dunn, *Identity Manipulation: Responding to Advances in Artificial Intelligence and Robotics*, 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3772057.

[62] Meta, *Taking Action Against 'Nudify' Apps, 2025*, https://about.fb.com/news/2025/06/taking-action-against-nudify-apps/; StopNCII, *How does StopNCII Work?*, https://stopncii.org/.

[63] The Print, *Delhi Police to question 26 boys in 'Bois Locker Room' case, suspect more groups are active*, 2020, https://theprint.in/india/delhi-police-to-question-26-boys-in-bois-locker-room-case-suspect-more-instagram-groups-are-active/414872/; In May 2020, minor and non-minor boys from prominent schools in Delhi were found to operate an instagram account by the name "bois locker room", where they shared explicit and obscene images of girls, gave rape threats and discussed plans to gangrape targeted female classmates.

[64] The Print, Muslim women 'on sale' — website targets journalists, activists, taken down after outrage, 2021, https://theprint.in/india/muslim-women-on-sale-website-targets-journalists-activists-taken-down-after-outrage/690318/; In July 2021, an X (then Twitter) user shared the link of a website floating on the Internet, in which she, among 90 Muslim women from India and other nations, had been put up for "auction". The website was called "Sulli Deals", based on a slur used against Muslim women. The website described itself as a "community driven open source project" and was hosted by GitHub. The website creators had targeted and profiled Muslim women working as journalists, activists, researchers etc in order to harass them, by putting them up "for sale".

[65] Monica Jha, *The dark hand of tech that stokes sex trafficking in India*, 2019, https://archive.factordaily.com/tech-phone-calls-whatsapp-facebook-sex-trafficking-india/; Benson Rajan, *Harassment and abuse of Indian women on dating apps: a narrative review of literature on technology-facilitated violence against women and dating app use*, 2025, https://www.nature.com/articles/s41599-024-04286-6.

jurisdictional abuse, the ease with which content (including photos and videos) can be copied and distributed, and the enduring nature of digital connectivity. Additionally, TFGBV can amplify harm to victims/survivors as a result of the breadth of audiences who observe the abuse, and the opportunities for abusers to join together[66] on digital platforms and collectively harm targeted individuals or groups.[67]

Another complexity that exacerbates online harms for victims of TFGBV is the permanence of online content especially in the context of activities like image based abuse. Specifically, even if the original piece of content has been removed, the internet allows for the easy download, copying and redistribution of the abusive content across multiple platforms, and leaves victims/survivors at perpetual risk of grave abuse.[68] **One example that helps illustrate how women are especially vulnerable to long term online harassment is the *Gamergate* campaign.** These attacks were carried out by large groups of sexist gamers targeting women media critics/video game developers, who were offering gendered critiques of the ecosystem. There were coordinated attacks that accused these critics of using their sexuality to move forward in the online gaming world and that their presence was not welcome. The campaign led to death threats and ultimately caused fear, anxiety and even withdrawal from these online spaces.[69]

Since TFGBV is an evolving concept we lack an exhaustive list of what types of activities fall within it. However, the UN Declaration on the Elimination of Violence Against Women's definition[70] of GBV is a good starting baseline. Some of the type of threats[71] that emerge as a result of TFGBV include threats/hate speech that incite gender based violence, harassing digital communication, dissemination of harmful lies, defamation/misrepresentation, impersonation, trafficking, disclosure (or a threat thereof) of private information, doxxing, sexploitation/ sextortion, trolling, unauthorised access to devices or private information, stalking, networked (mob) harassment, manipulated images, synthetic media, and other forms of image based abuse.

As discussed above, not all women are equally exposed to the same kinds of risks and vulnerabilities. These differ[72] on the basis of intersectional identity, scenarios that involve intimate partner violence, women in leadership/public roles e.g. politics, journalism, sports, etc.

In 2022, International Center for Research on Women (ICRW) and the University of Chicago undertook a case study of TFGBV in India.[73] This study observes that TFGBV in India is a continuation of violence and inequities from the offline sphere. It specifically says, "*... inequalities that make women and girls vulnerable to*

---

[66] Michael Salter, From geek masculinity to Gamergate: the technological rationality of online abuse, 2017, (https://www.researchgate.net/publication/312155283_From_geek_masculinity_to_Gamergate_The_technological_rationality_of_online_abuse.

[67] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[68] Carrie Goldberg, *Nobody's Victim: Fighting Psychos, Stalkers, Pervs, and Trolls*, 2019, https://archive.org/details/carrie-goldberg-nobodys-victim-fighting-psychos-stalkers-pervs-and-trolls-pengui.

[69] The New York Times, *Feminist Critics of Video Games Facing Threats in 'GamerGate' Campaign*, 2014, https://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html.

[70] Article 1 defines GBV as any act "***that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or private life***."

[71] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[72] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[73] USAID, NORC at the University of Chicago and the International Center for Research on Women (ICRW), *Technology-Facilitated Gender-based Violence in Asia: India*, 2021, https://www.icrw.org/wp-content/uploads/2021/09/USAID-TFGBV-India.pdf.

offline violence penetrate the online space." **It categorically states that TFGBV must be understood through an intersectional lens of class, caste, gender, sexuality, religion, education and access to technology. The study observes that in rural India women and girls usually only access the internet in heavily surveilled environments since the relevant device is often controlled by the patriarchal head (husbands and fathers) of the household.** This environment of control leads to under-reporting by survivors. They end up fearing that reports of violence can lead to greater restrictions of device ownership and opportunities for usage. The study also observed that male dominated cultural norms also make the internet inaccessible/ inhospitable for women and girls. These social structures have been observed to create unique challenges in terms of self-censorship and online participation for Indian women in public positions/spaces e.g. female journalists, politicians and women's rights activists.[74]

Overall, the study observed that in India women and girls must navigate two distinct kinds of harassment related risks.

- **The first is violence that is committed by an individual in either a public or private setting.** These types of harms are committed by individuals, usually men, who can be either strangers or ex/current intimate partners. Harms often range across harassing phone calls, online sexual harassment, or non-consensual distribution of intimate images.
- **The second kind of risk that the study identifies is violence perpetrated by groups, often in a public manner.** It says that such attacks are usually administered by groups known as "*cyber trolls*" or *"troll armies"* led often by men that repeatedly harass or threaten targeted individuals.[75] Such group based attacks are amplified through the anonymity gained via fake profiles. These tactics disproportionately impact women, members of the LGBTQI+ community, and then gendered minorities from various intersectional identities.[76]
- **Lastly, the study observes that the onset of the COVID-19 pandemic has exacerbated TFGBV several fold.** It has even led to newer forms of TFGBV e.g. the phenomenon of *Zoom Bombing* or *Zoom Flashing* in online classroom or work meetings.

As discussed earlier, there is no exhaustive list of the types of activities covered as TFGBV. Table 1 offers insights on certain unique risks that women recurrently face when using the internet in India. The analysis attempts to distinguish for the reader cybercrimes that are measured by India's National Crime Records Bureau (NCRB)[77], and other forms of TFGBV that may not be strictly unlawful but in totality cause harm to women internet users. It also illustrates India's current ability to comprehensively track TFGBV.

---

[74] USAID, NORC at the University of Chicago and the International Center for Research on Women (ICRW), *Technology-Facilitated Gender-based Violence in Asia: India*, 2021, https://www.icrw.org/wp-content/uploads/2021/09/USAID-TFGBV-India.pdf.

[75] Former Cabinet Minister Sushma Swaraj faced coordinated trolling by extremist right-wing cyber trolls for helping an interfaith couple get their passports, Sushma Swaraj's Reaction to Bitter Trolling Was Graceful but Inadequate – The Wire; Journalist Rana Ayyub was targeted by cyber trolls on X with sexist and Islamophobic comments, including rape threats and pictures of her face morphed onto pornographic images, Delhi Journalists Body Condemns Relentless Trolling of Rana Ayyub – The Wire.

[76] USAID, NORC at the University of Chicago and the International Center for Research on Women (ICRW), *Technology-Facilitated Gender-based Violence in Asia: India*, 2021, https://www.icrw.org/wp-content/uploads/2021/09/USAID-TFGBV-India.pdf.; Also see: Kiruba Munusamy, *Intersection of identities: online gender and caste based violence*, 2018, https://www.genderit.org/articles/intersection-identities-online-gender-and-caste-based-violence

[77] National Crime Records Bureau, *Crime in India 2022: Statistics Volume II*, 2022, https://www.ncrb.gov.in/uploads/nationalcrimerecordsbureau/custom/1701608364CrimeinIndia2022Book2.pdf.

Table 2.1: Types of Harms Most Commonly Experienced by Women

Type of Harm	Description	Indian Data	Covered in NCRB or other government databases?
Cybercrimes			
Image Based Sexual Abuse[78]	While there are multiple forms, mainstream policy conversations usually revolve around the non-consensual distribution of intimate images (NCII) by ex/current partners.[79] However, this category of online sexual abuse is actually broader and also has wider swathe of perpetrators/abusers.[80] It involves the creation and distribution of sexual imagery without consent of the person featured in the concerned media.	NCRB tracks such offences using a relatively broad category of '***publishing obscene sexual materials without consent***'. As per their reporting, this number has gone up from **1158 in 2019 to 2251 in 2022 - signalling a 94% increase in reported incidents.**	Partially Captured
	This form of online sexual abuse also involves any threat to create and distribute such images.[81] Under this, we observed certain types of activities namely: (a) distribution of NCII; (b) voyeurism/ creepshots; (c) sexploitation and sextortion; (d) documentation and online broadcasting of sexual violence; (e) non-consensual creation of synthetic/ deepfake sexual media.[82]	Overall these low baseline figures demonstrate that Indian institutions are **unable to track image based sexual offences in a sufficiently disaggregated manner**. Further, the NCRB shows disproportionately low reporting numbers. **Thus, we see a systemic gap of imperfect taxonomies and poor reporting practices**.	

[78] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[79] Valente, Giorgetti, Neris, Ruiz and Bulgarelli, *The Body is Code: Legal Strategies to Combat Revenge Porn in Brazil, 2018, https://internetlab.org.br/en/news/internetlab-releases-the-book-the-body-is-the-code/*.

[80] McGlynn and Rackley, *Image-Based Sexual Abuse*, 2017, https://academic.oup.com/ojls/article-abstract/37/3/534/2965256; Also see: McGlynn, Rackley and Houghton, *Beyond 'revenge porn': The continuum of image-based sexual abuse*, 2017, https://link.springer.com/content/pdf/10.1007/s10691-017-9343-2.pdf.

[81] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[82] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

Type of Harm	Description	Indian Data	Covered in NCRB or other government databases?
Technology Facilitated Grooming as a Precursor to Trafficking	The UN Special Rapporteur's 2018 report on online violence against women and girls identifies **online trafficking as a prevalent form of TFGBV**. *This often involves recruitment or exploitation through deceptive or coercive online communications. Sometimes accompanied with false school/work opportunities, marriage proposals, romantic relationships, or promises of travel.* There are usually signs of individuals being lured into situations of exploitation through suspicious offers or messages. Online sextortion[83] and impersonation activities can also trick vulnerable women into dangerous situations such as scenarios involving human trafficking.[84]	The annual crime statistics of India's NCRB reported more than 10,000 cases of trafficking between 2018 and 2022. **However, the statistics have not been disaggregated to disclose the number of cases where trafficking has been facilitated by the internet and related ICTs.[85]** The current reporting mechanisms do not lend itself to a consolidated understanding of how sextortion, online impersonation and other related methods enable tech facilitated trafficking. **The absence of this data is particularly concerning** since press reports suggest that traffickers in India are using **new digital avenues like instant loan apps[86], gaming sites, and deep fake technology to traffic young women.[87]**	Inadequate Data and Measurement Practices

[83] Occurs when an individual has, or claims to have, a sexual image of another person and uses it to coerce a person into doing something they do not want to do. See: Wittes, Poplin, Jurecic & Spera, *Sextortion: Cybersecurity, teenagers and remote sexual assault*, 2016, https://www.brookings.edu/wp-content/uploads/2016/05/sextortion1-1.pdf

[84] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.; Also see: Association for Progressive Communications, *Voices from digital spaces: technology-related violence against women*, 2012, https://www.apc.org/en/pubs/voices-digital-spaces-technology-related-violence-against-women.

[85] Information and Communication Technologies.

[86] BBC World Service Eye Investigations, *Inside the deadly instant loan app scam that blackmails with nudes*, 2023, https://www.bbc.com/news/world-asia-india-66964510.
Several Loan apps in India have harassed clients, resulting in loss of life.

[87] Shiv Sahay Singh, *Over 10,000 cases of trafficking but only 1,031 convictions between 2018-2022*, 2024, https://www.thehindu.com/news/national/other-states/over-10000-cases-of-trafficking-but-only-1031-convictions-between-2018-2022/article67713302.ece; Sreeparna Chakrabarty, *Traffickers move online to search for victims*, 2022, https://www.thehindu.com/news/national/traffickers-move-online-to-search-for-victims/article66223005.ece.

Type of Harm	Description	Indian Data	Covered in NCRB or other government databases?
Stalking and Monitoring	Repeated, unwanted contact or surveillance through digital means, including social media, messaging apps, or emails that can cause a targeted person to feel fear.[88] Indicators include obsessive messaging, constant monitoring of online activity, threats, or impersonation to harass the victim. It can also involve the use of ICTs to track the target's location or the installation of commercial *stalkerware* to granularly track online activities including their digital interactions.[89]	In India, the extent of cyber-stalking is measured using data from NCRB, which suggests that a total of 1457 cases of cyber-stalking were registered, where Maharashtra topped the list with over 500 cases.[90]	Yes, but does not seem to be extensively captured.
Legally Amorphous Forms of TFGBV			
Doxxing	The public disclosure of private information can have harmful ramifications for women and girls.[91] Often such activities are meant to harass, embarrass or harm a target's reputation.[92] Doxing is one of the more dangerous forms of such activities. **It involves the non-consensual public release of personal information e.g. legal name, driver's license, workplace, home address, phone number, email or other personal details. The term originates from a hacking term "*dropping dox*"[93] where documents are published online.** Doxxing is usually intended to make victims/survivors experience either widespread online harassment or to instil fear of in-person harm/ harassment.	India lacks consolidated data to measure the nature, scope and extent of doxxing activities in India.	No

[88] Danielle Keats Citron, *Hate Crimes in Cyberspace*, 2016, https://www.hup.harvard.edu/books/9780674659902.

[89] Lenhart, Ybarra, Zickuhr & Price-Feeney, *Online Harassment, Digital Abuse and Cyberstalking in America*, 2016, Online Harassment, Digital Abuse, and Cyberstalking in America.; Khoo, Robertson & Deibert, *Installing Fear: A Canadian Legal and Policy Analysis of Using, Developing, and Selling Smartphone Spyware and Stalkerware Applications*, 2019, stalkerware-legal.pdf; National Network to End Domestic Violence, *Glimpse From the Field: How Abusers Are Misusing Technology*, 2014, Glimpse From the Field: How Abusers Are Misusing Technology | Office of Justice Programs.

[90] National Crime Records Bureau, *Crime in India 2022: Statistics Volume II*, 2022, https://www.ncrb.gov.in/uploads/nationalcrimerecordsbureau/custom/1701608364CrimeinIndia2022Book2.pdf

[91] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[92] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.

[93] Sarah Jeong, *The Internet of Garbage*, 2018, https://cdn.vox-cdn.com/uploads/chorus_asset/file/12599893/The_Internet_of_Garbage.0.pdf.

Type of Harm	Description	Indian Data	Covered in NCRB or other government databases?
Trolling/ Bullying/ Harassment (Including Sexual Harassment)	Persistent and targeted negative behaviour, including offensive comments, threats, or derogatory remarks, is typically intended to provoke, intimidate, or distress the individual. Indicators may include repeated personal attacks, the spreading of malicious rumours, coordinated efforts to shame or embarrass, and an escalating pattern of abusive language or imagery directed at an individual.	In India, a 2017 study found that sexual harassment is a significant concern for individuals under 40, with 40% of this age group having experienced it.[94] The highest reports of sexual harassment in India came from victims in Delhi and Mumbai (43%), followed by Kolkata (37%) and Bangalore (36%) .[95]	Absence of adequate details with NCRB
Misinformation / Defamation (Networked Harassment[96])	The spread of false or misleading information aimed at damaging an individual's reputation. This involves fake news, doctored images, or false accusations that misrepresent facts, particularly when it targets individuals with the intent to harm their reputation. There is usually repeated sharing of incorrect information across different online platforms.	Defamation cases are tracked within NCRB data, which observes that 385 defamation cases were reported in 2022.	India lacks appropriate data to track networked harassment.

2.3.2. Profiling Online Risks Faced by Children

One-third of internet users worldwide are children.[97] 15% of active internet users in India are estimated to be between the ages of 5 and 11 years.[98] Children and adolescents use the internet and digital services in unique ways and thus are vulnerable to novel kinds of risks. The first main category of risk that they are exposed to relates to **child sexual exploitation and abuse i.e. CSEA**, and within that category one type of harm constitutes being victimised through the creation and distribution of **child sexual exploitative and abuse materials (CSEAM)**.

The WeProtect Global Alliance surveyed 18-20 year olds about their online experiences as adolescents. The results estimated that **globally an alarming 54% of individuals who regularly use the internet as children have been victims of at least one form of online sexual harm**.[99] According to the Childlight Global Child Safety Institute

[94] DQChannels, *Norton study reveals widespread online harassment in India*, 2017, https://www.dqchannels.com/norton-study-reveals-widespread-online-harassment-in-india/.

[95] Deccan Chronicle, *Study reveals widespread online harassment in India*, 2017, https://www.deccanchronicle.com/technology/in-other-news/091017/study-reveals-widespread-online-harassment-in-india.html.

[96] Suzie Dunn, *Technology-Facilitated Gender-based Violence: An Overview*, 2020, https://www.cigionline.org/static/documents/SaferInternet_Paper_no_1_coverupdate.pdf.; Marwick & Caplan, *Drinking male tears: language, the manosphere, and networked harassment*, 2018, https://www.tandfonline.com/doi/full/10.1080/14680777.2018.145056.

[97] World Health Organization, *Online violence against children*, https://www.who.int/teams/social-determinants-of-health/violence-prevention/online-violence-against-children.

[98] The Hindu Businessline, *66 mn children aged 5-11 years are active Internet users in India*, 2021, https://www.thehindubusinessline.com/info-tech/66-mn-internet-users-in-india-aged-between-5-and-11-years/article29518418.ece.

[99] WeProtect, *Estimates of childhood exposure to online sexual harms and their risk factors*, 2024, https://www.weprotect.org/economist-impact-global-survey/?utm_source=chatgpt.com#report

at the University of Edinburgh, **over 300 million children fall victim to online sexual exploitation and abuse annually**, representing a clear and present danger to the world's children.[100]

Risks are exacerbated on digital services due to the unprecedented ease with which child sex offenders can contact potential victims, share illicit images, and incite others to commit offences. Children may be victimized through the creation, distribution, and consumption of sexual abuse material, or they may be **groomed for exploitation**, with abusers attempting to meet them in person or coercing them into providing explicit content.[101] **A key challenge to children's online safety stems from the persistent stigma surrounding child sexual exploitation and abuse. Therefore, international organisations opine that the official reported figures likely represent a mere fraction of the true extent of the issue's real prevalence.[102]**

Children are also at risk of **off-platforming or a type of cross-platform risk** wherein sexual abusers form relationships and manipulate children into migrating from public social media platforms to private messaging apps where sexually abusive exchange of messages and images cannot be detected. **A study[103] found that 65% of the children they sampled had been asked to migrate their communication by an online contact, and 52% percent of this sample complied with their online contact by migrating. Research indicates that applications that allow ephemeral nature of messages were more likely to be used to share sexual images, even though the intention of the design is to grant more privacy/security.[104]**

Additionally, children are **vulnerable to a host of new age harms such as cyberbullying, impersonation, trolling, harassment, exposure to hate speech, exposure to misogyny and misogynistic echo chambers[105], hostile peer activity, gambling, encouragement of self-harm, and identity theft.[106]**

While some of these risks are covered under criminal laws, other risks operate in **legally ambiguous terrain** that nevertheless lead to harmful experiences for the victim. Other ***"legal but harmful"*** risks that negatively impact young people, regardless of gender, are extreme body image content, depressive content, cyberbullying, eating disorders, exposure to misogyny.[107]

---

[100] The University of Edinburgh, *Scale of online harm to children revealed in global study*, 2024, https://www.ed.ac.uk/news/2024/scale-of-online-harm-to-children-revealed-in-globa.

[101] UNICEF, *Keeping children safe online*, https://www.unicef.org/protection/keeping-children-safe-online#:~:text=When%20browsing%20the%20internet%2C%20children,collect%20data%20for%20marketing%20purposes.

[102] World Economic Forum, *Online dangers for children are rife. We must both pre-empt them and treat the consequences,* 2022, https://www.weforum.org/stories/2022/06/child-safety-protection-internet/

[103] Thorn, *Online Grooming: Examining risky encounters amid everyday digital socialization: Findings from 2021 qualitative and quantitative research among 9-17-year-olds*, 2022, https://info.thorn.org/hubfs/Research/2022_Online_Grooming_Report.pdf.

[104] Revealing Reality, *Not just flirting: The unequal experiences and consequences of nude image-sharing by young people*, 2022, https://revealingreality.co.uk/wp-content/uploads/2022/06/Revealing-Reality_Not-Just-Flirting.pdf.

[105] Media and Society, *Incel Culture: The rise of online toxic norms and misogyny*, 2025, https://mediaandsociety.org/2025/04/07/toxic-online-norms-incel-culture/. Manosphere is an umbrella term for online communities that promote harmful ideas of toxic masculinity and the use of wealth, physical appearance, dominance and control to subjugate women. Manosphere is increasingly making its way to young people's social media feeds, exposing youngsters to harmful content created by masculinity influencers posing as men's rights activists, thus creating an echo chamber of misogynistic values.

[106] Livingstone, Haddon, Görzig and Ólafsson, *Risks and safety on the internet: the perspective of European children: full findings and policy implications from the EU Kids Online survey of 9-16 year olds and their parents in 25 countries*, 2011, https://eprints.lse.ac.uk/33731/1/Risks%20and%20safety%20on%20the%20internet%28lsero%29.pdf.

[107] Children & young people's Commissioner Scotland, *Protecting children from harms online*, 2024, https://www.cypcs.org.uk/resources/ofcom-july24/#:~:text=The%20Online%20Safety%20Act%20makes%20provision%20for%20additional,inclusion%20of%20body%20image%20content%20and%20depressive%20content; World Economic Forum, *Online dangers for children are rife. We must both pre-empt them and treat the consequences,* 2022, https://www.weforum.org/stories/2022/06/child-safety-protection-internet/.

Another increasingly discussed risk is imbalance in a child's life due to excessive screen times and reported risks of online addiction.[108] This extended screen time can disrupt sleep patterns, reduce face-to-face interactions, and limit participation in physical activities, all of which are crucial for healthy development.[109] Here is a list of commonly recognised harms to children:[110]

Table 2.2: Types of Harms Most Commonly Experienced by Children

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
Child Sexual Exploitative and Abuse Material (CSEAM)	The presence of explicit images or videos depicting children, shared or accessed online, is a serious concern.[111] While certain platforms, such as encrypted messaging apps, serve legitimate privacy needs, they can sometimes be misused for harmful activities.	In the European Union, Europol reported a "considerable increase" in the sharing of child sexual exploitative and abuse materials (CSEAM) during 2020-21.[113] In the United States (US), cases of online child sexual abuse and exploitation (OCSAE) more than doubled in the first half of 2020 compared to the same period in 2019.[114] UK-based Internet Watch Foundation noted that 2021 was the worst year on record for online child sexual abuse.[115] Similarly, Australia experienced a 90 per cent increase in online illegal content between 2019 and 2020, with the majority being CSEAM.[116]	**According to the National Human Rights Commission (NHRC), out of 32 million reports of child sexual exploitative and abuse material, 5.6 million were uploaded by perpetrators based in India.[118]** CSEAM data in India is further recorded by NCRB, and is broadly categorised as **'publishing obscene sexual materials online'**.	Yes

[108] Huijuan Yu, Chan Xu, Jiamin Lu, Qishan Li, Qian Li, Kefan Zhou, Jiawen Zhong, Yingyu Liang, Wenhan Yang, *Associations between screen time and emotional and behavioral problems among children and adolescents in US, National Health Interview Survey (NHIS), 2022*, 2025, https://www.sciencedirect.com/science/article/abs/pii/S0165032725003684; Newsweek, *Psychologists Tracked 292,000 Kids' Screen Time—What They Found Is Alarming*, 2025, https://www.newsweek.com/kids-screen-time-vicious-circle-psychologists-warning-2082727.

[109] UNICEF, *Keeping children safe online*, https://www.unicef.org/protection/keeping-children-safe-online#:~:text=When%20browsing%20the%20internet%2C%20children,collect%20data%20for%20marketing%20purposes.

[110] National Crime Records Bureau, *Crime in India Yearwise*, 2022, https://www.ncrb.gov.in/crime-in-india-year-wise.html?year=2022&keyword=.; World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms*, 2023, https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf.

[111] Protect Children, *Online Child Sexual Abuse and Exploitation: Current and Emerging Threats*, 2024, https://www.ohchr.org/sites/default/files/documents/issues/children/sr/cfis/existing-emerging/subm-existing-emerging-sexually-cso-suojellaan-lapsia-ry.pdf

[113] Europol, *Internet Organised Crime Threat Assessment*, 2021,https://www.europol.europa.eu/cms/sites/default/files/documents/internet_organised_crime_threat_assessment_iocta_2021.pdf

[114] Dustin Racioppi, *'People don't want to talk about it,' but reports of kids being exploited online have spiked amid coronavirus pandemic*, 2020,https://www.usatoday.com/story/news/nation/2020/10/22/coronavirus-child-abuse-nj-online-child-exploitation-reports-increase/6004205002/

[115] Euronews, *COVID lockdowns saw a record rise in online child sexual abuse reports, says watchdog*, 2022, https://www.euronews.com/next/2022/01/14/covid-lockdowns-saw-a-record-rise-in-online-child-sexual-abuse-reports-says-watchdog.

[116] Helena Burke, *Covid-19 lockdowns cause disturbing spike in online child exploitation activity in Australia,* 2021, https://www.news.com.au/national/crime/covid19-lockdowns-cause-disturbing-spike-in-online-child-exploitation-activity-in-australia/news-story/a8a2e904ae7f4704bc4484e81d470890.

[118] The Hindu, *How safe is the online space for children in India?* 2023, https://www.thehindu.com/podcast/how-safe-is-the-online-space-for-children-in-india/article67543810.ece.

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
	Indicators of misuse may include suspicious file-sharing or attempts to hide illicit activity using coded language.[112]	In 2021, the National Center for Missing & Exploited Children (NCMEC) escalated over 49,000 urgent reports to law enforcement involving children in imminent danger. Similarly, in its 2023 report, WeProtect Global Alliance (which includes governments, companies, and charities collaborating for digital safety) reported an 87% increase in such cases since 2019.[117]	The number of reported cases has surged from 102 in 2019 to 1,171 in 2022, marking a staggering 1,048% increase. **However, despite this dramatic rise, these figures underrepresent the true scale of the problem and fail to capture the full extent of the issue.** The global data as well as the NHRC data cited in this table suggests that the actual prevalence is much higher.[119]	
Grooming	An adult engages in personal or private conversations with a child online, **often via social media or gaming platforms.**[120] Signs include attempts to isolate the child, requests for inappropriate images, or offering gifts.	**Europol** reported a "steep increase" in online "grooming" activities in 2020-21.[121] **UK police** have recorded nearly 34,000 online grooming crimes against children. Alarmingly, one in four of these crimes over the past five years targeted primary school children.[122] **The '*Global Threats Assessment*' Report (2023) from We Protect Alliance** suggests that the number of online grooming and coercion cases is on the rise.[123] In 2022 alone, NCMEC received over 10,000 reports of such cases - a staggering increase from just 139 reports in 2021.[124]	While no concrete data quantifies the prevalence of grooming in the Indian context, on-ground reportage suggests that an increase in internet use might be associated with an increased risk for children.[127]	No

[112] Protect Children, *Online Child Sexual Abuse and Exploitation: Current and Emerging Threats*, 2024, https://www.ohchr.org/sites/default/files/documents/issues/children/sr/cfis/existing-emerging/subm-existing-emerging-sexually-cso-suojellaan-lapsia-ry.pdf

[117] WeProtect Global Alliance, *Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response*, 2023, https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2023-English.pdf.

[119] National Crime Records Bureau, *Crime in India Yearwise*, 2022, https://www.ncrb.gov.in/crime-in-india-year-wise.html?year=2022&keyword=.

[120] Thorn, *Online Grooming: what it is, how it happens and how to defend children*, 2024, https://www.thorn.org/blog/online-grooming-what-it-is-how-it-happens-and-how-to-defend-children/.

[121] Europol, *Internet Organised Crime Threat Assessment*, 2021, https://www.europol.europa.eu/cms/sites/default/files/documents/internet_organised_crime_threat_assessment_iocta_2021.pdf.

[122] NSPCC, *82% rise in online grooming crimes against children in the last 5 years*, 2023, https://www.nspcc.org.uk/about-us/news-opinion/2023/2023-08-14-82-rise-in-online-grooming-crimes-against-children-in-the-last-5-years/.

[123] WeProtect Global Alliance, *Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response*, 2023, https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2023-English.pdf.

[124] WeProtect Global Alliance, *Alarming escalation in child sexual abuse online revealed by Global Threat Assessment 2023,* 2023, https://www.weprotect.org/wp-content/uploads/Global-Threat-Asssesment-2023-Press-Release.pdf.

[127] The New Indian Express, *India at number two in internet usage, leads to more child abuse*, 2023, https://www.newindianexpress.com/cities/bengaluru/2023/Aug/15/india-at-number-twoin-internet-usage-leads-to-more-child-abuse-2605358.html.

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
		Similarly, according to data from the National Society for the Prevention of Cruelty to Children, reports of online grooming in the UK have increased by 89% from 2017 to 2023.[125] One prevalent tactic used by groomers is ***"off-platforming"***, which involves shifting conversations to private messaging or end-to-end encrypted platforms where the risk of detection is significantly lower. **According to institutions like UNICEF,** groomers use this technique not only to target children but also to network with others and share abusive material.[126]		
Legally Amorphous Forms of Online Harms faced by Children				
Bullying and Harassment	Repeated and targeted negative behaviour online, including offensive comments, threats, or derogatory remarks aimed at an individual.[128] **Indicators include persistent personal attacks, spreading of harmful rumours, or coordinated efforts to shame or embarrass**.	**According to UNICEF, more than a third of young people across 30 countries have reported experiencing cyberbullying, with 1 in 5 students skipping school as a result.[131]** **Surveys from the US** indicate that cyberbullying has persisted and increased between 2011 and 2019 — where one in ten high school boys and one in five high school girls experienced cyberbullying each year, indicating that the move online made bullying and harassment a larger part of everyday lives for children.[132]	A ten-country survey conducted by McAfee placed India at the top of global cyberbullying rankings, with 85% of Indian children report having been cyberbullied and 1 in 3 children facing cyber racism, sexual harassment and threats of physical harm as early as at the age of 10.[135]	Yes

[125] National Society for the Protection of Children to Cruelty, *Online grooming crimes against children increase by 89% in six years*, 2024, https://www.nspcc.org.uk/about-us/news-opinion/2024/online-grooming-crimes-increase/.

[126] WeProtect Global Alliance, *Analysis of the sexual threats children face online: Global threat assessment 2023*, 2023, https://www.weprotect.org/global-threat-assessment-23/analysis-sexual-threats-children-face-online/.

[128] Parents, *The real-life effects of cyberbullying on children and teens,* 2024, https://www.parents.com/what-are-the-effects-of-cyberbullying-460558.

[131] UNICEF, *Keeping children safe online,* https://www.unicef.org/protection/keeping-children-safe-online#:~:text=When%20browsing%20the%20internet%2C%20children,collect%20data%20for%20marketing%20purposes.

[132] Jonathan Haidt, *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness*, 2024, https://www.penguinrandomhouse.com/books/729231/the-anxious-generation-by-jonathan-haidt/.

[135] McAfee, *Cyberbullying in Plain Sight,* 2022, https://media.mcafeeassets.com/content/dam/npcld/ecommerce/en-us/docs/reports/rp-cyberbullying-in-plain-sight-2022-global.pdf; Sukriti Vats, *85% Indian kids have experienced cyberbullying, highest in the world, finds new survey,* 2022, https://theprint.in/india/85-indian-kids-have-experienced-cyberbullying-highest-in-the-world-finds-new-survey/1074175/.

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
	This could also include **peer on peer abuse** among children, extending to sexual abuse. Peer on Peer, or child on child abuse means abuse of one child by another child, regardless of age, stage of development or the age differential between them.[129] Peer on peer abuse can include bullying, hate incidents and hate crimes, physical abuse, racism, teenage relationship abuse, hazing and harmful sexual behaviour (HSB).[130]	**In Europe**, approximately 12% of adolescents, or 1 in 8, report engaging in cyberbullying. Boys are more likely to engage in cyberbullying, with 14% reporting such behaviour compared to 9% of girls.[133] **In the UK**, an Ofsted study revealed that around 90% of girls and 50% of boys in schools reported peer on peer sexual abuse as "commonplace".[134]	These harms may not be restricted to online platforms. **For instance, a UNICEF study found that at least 36% of Indian students experience harassment and bullying on school campuses.[136]** Online bullying has severe consequences.[137] For instance, young adolescents who experience cyberbullying are at an increased risk of suicidal ideation.[138]	
Excessive Screen Time and Internet Use Disorder	Excessive and compulsive use of online platforms or digital devices can impact daily responsibilities and social interactions.[139]	Unhealthy use of the internet for prolonged periods of time has been discussed as a significant mental health issue in many countries. A meta-analysis of 31 countries estimated that the global prevalence of **Internet Use Disorder is 6.0% among individuals aged 12 to 41.**[140]	In India, an independent survey with a sample size of 1,000 parents reveals that almost 60% of children aged 5 to 16 display behaviours that suggest a potential for digital addiction and 70-80% are surpassing recommended screen time limits daily.[143]	

[129] Farrer & Co, *Addressing child-on-child abuse: a resource for schools and colleges*, 2024, https://www.farrer.co.uk/globalassets/clients-and-sectors/safeguarding/addressing-child-on-child-abuse.pdf.

[130] Farrer & Co, *Addressing child-on-child abuse: a resource for schools and colleges*, 2024, https://www.farrer.co.uk/globalassets/clients-and-sectors/safeguarding/addressing-child-on-child-abuse.pdf.

[133] World Health Organization, *One in six school-aged children experiences cyberbullying, finds new WHO/Europe study*, 2024, https://www.who.int/europe/news/item/27-03-2024-one-in-six-school-aged-children-experiences-cyberbullying--finds-new-who-europe-study.

[134] Office for Standards in Education, Children's Services and Skills (Ofsted), Government of the UK, *Ofsted: A Review on Sexual Harassment in Schools and Colleges,* 2021, https://www.governmentevents.co.uk/ofsted-a-review-on-sexual-harassment-in-schools/.

[136] UNICEF, *Strategy for Ending Violence Against Children, 2020, https://www.unicef.org/india/sites/unicef.org.india/files/2020-07/UNICEF%20India%20EVAC%20Programme%20Strategy_web%20version.pdf.*

[137] Power of Zero, *Why India needs to take bullying seriously*, 2024, https://powerof0.org/bullying-india/#:~:text=85%25%20of%20Indian%20children%20say,and%20bullying%20in%20school%20campuses.

[138] National Institutes of Health, *Cyberbullying linked with suicidal thoughts and attempts in young adolescents*, 2022, https://www.nih.gov/news-events/nih-research-matters/cyberbullying-linked-suicidal-thoughts-attempts-young-adolescents.

[139] Children and screens, *Digital Addictions: a family guide to prevention, signs and treatment*, 2024, https://www.childrenandscreens.org/learn-explore/research/digital-addictions-a-family-guide-to-prevention-signs-and-treatment/.

[140] Ding & Li, *Digital Addiction Intervention for Children and Adolescents: A Scoping Review*, 2023, https://pmc.ncbi.nlm.nih.gov/articles/PMC10049137/#:~:text=1.1.&text=A%20meta%2Danalysis%20of%2031,to%2019%20years%20%5B5%5D.

[143] The Hindu, *60% of children at risk of digital addiction: Survey*, 2024, https://www.thehindu.com/sci-tech/technology/60-children-at-risk-of-digital-addiction-survey/article67989890.ece.

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
	Indicators include prolonged screen time, withdrawal symptoms when not online, and a noticeable decline in physical or mental health.	Another meta-analysis, spanning three decades, found that 4.6% of adolescents aged 10 to 19 suffer from **Internet Gaming Disorder (IGD)**.[141] Additionally, multiple studies suggest that such increased use is more prevalent among younger age groups.[142]	Similarly, due to the affordability and widespread access to touchscreen mobiles, tablets, and WiFi, some studies estimate that at least 24.6% of adolescents in India are exhibiting problematic internet use or Internet Addiction Disorder.[144] While disaggregated data is scarce, these trends **likely highlight patterns in urban India more than in rural India.** In a school-based cross sectional study conducted in the rural and urban field practice area of a medical college hospital in Mangaluru, the prevalence of internet addiction among urban school students was at 83.3%, while for rural school students it was at 78%.[145]	No
General Mental Health Harms	There have been concerns around changes in a child's behaviour due to online interactions, such as increased anxiety, depression, or body image issues.[146]	Increased social media use has been linked to the potential for self-harm in adolescent girls and boys. Research published by the **American Psychological Association** found that teens and young adults who cut their social media use by 50% for three weeks experienced notable improvements in their body image and overall	In India, a comprehensive study[149] conducted in Coimbatore surveyed **1,200 college girls** and found that 77.6% were dissatisfied with their bodies. Key factors contributing to this dissatisfaction included higher	No

[141] Fam JY, *Prevalence of internet gaming disorder in adolescents: A meta-analysis across three decades*, 2018, https://pubmed.ncbi.nlm.nih.gov/30004118/.

[142] Fam JY, *Prevalence of internet gaming disorder in adolescents: A meta-analysis across three decades*, 2018, https://pubmed.ncbi.nlm.nih.gov/30004118/.

[144] Maheshwari & Sharma, *Internet Addiction: a growing concern in India*, 2018, https://journals.lww.com/iopn/fulltext/2018/15010/internet_addiction__a_growing_concern_in_india.15.aspx#:~:text=According%20to%20a%20study%2C%20the,addiction%20disorder%20(IAD)21.

[145] Sowndarya T. A.and Mounesh Pattar, *Pattern of internet addiction among urban and rural school students, Mangaluru, India: a comparative cross-sectional study*, 2018, https://www.ijpediatrics.com/index.php/ijcp/article/download/1849/1308/6876.

[146] The BMJ Opinion, *Screen time and social media: Interventions to protect our children's health*, 2019, https://blogs.bmj.com/bmj/2019/02/07/screen-time-and-social-media-interventions-to-protect-our-childrens-health/

[149] Docvita, *Is social media affecting how you see yourself?* 2024, https://docvita.com/blog/the-impact-of-social-media-on-body-image/#:~:text=A%20larger%20study%20in%20Coimbatore,of%20the%20%E2%80%9Cideal%E2%80%9D%20body.

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
	Indicators include withdrawal from social activities, frequent exposure to harmful content, and distressing posts or messages.	appearance compared to their peers who continued to use social media at consistent levels.[147] Similarly, studies on Amercian teens confirm that exposure to videos and photos on social media platforms can lead to body dissatisfaction and eating disorders among teenage and adolescent girls, potentially resulting in mental health issues, including suicidal behaviour.[148]	BMI, societal pressure to be thin, and depression. Influenced by social media's depiction of the "ideal" body, many of these girls resorted to skipping meals or eating very little in an effort to control their weight.	
Exposure to hate speech and ideological manipulation	Children are particularly vulnerable to extremist material online[150] due to the nascent and evolving nature of their identities and political ideologies.[151] Globally, exposure to hate speech has had real world consequences for young people, be it the normalisation of misogynistic views by masculinity influencers in the West[152], or the targeting of women from minority communities in India as in the Sulli Deals case[153].	Recommender algorithms used by social media platforms can amplify male - supremacist content. **A study conducted on 29 hours of short format online user generated videos demonstrated that all male identified accounts were shown anti-feminist and extremist male supremacist content, regardless of individual preferences,** and the number of such videos increased on their feeds once they registered interest by watching these videos.[158]	In a user survey[159] conducted to find the exposure of hate speech among **Instagram users in India,** users from relevant age groups of 13-17 years, 18-24 years and 25-34 years participated in sharing their insights. The participants comprised of female, male and non-binary users, with 65 participant responses being taken as valid. 43.8% of the participants reported exposure to hate speech via the direct messaging feature on Instagram.	No

[147] American Psychological Association, *Reducing social media use significantly improves body image in teens, young adults*, 2023, https://www.apa.org/news/press/releases/2023/02/social-media-body-image

[148] Harvard T.H. Chan School of Public Health, *Exploring the effect of social media on teen girls' mental health*, 2023, https://hsph.harvard.edu/news/exploring-the-effect-of-social-media-on-teen-girls-mental-health/.

[150] UN Office on Drugs and Crime, Handbook on Children Recruited and Exploited by Terrorist and Violent Extremist Groups: The Role of the Justice System, 2017 https://www.unodc.org/documents/justice-and-prison-reform/Child-Victims/Handbook_on_Children_Recruited_and_Exploited_by_Terrorist_and_Violent_Extremist_Groups_the_Role_of_the_Justice_System.E.pdf

[151] Jennifer H Pfeifer, Elliot T Berkman, *The Development of Self and Identity in Adolescence: Neural Evidence and Implications for a Value-Based Choice Perspective on Motivated Behavior,* 2019 https://pmc.ncbi.nlm.nih.gov/articles/PMC6667174/

[152] The Guardian, *Inside the violent, misogynistic world of Tiktok's new star, Andrew Tate*, 2022, https://www.theguardian.com/technology/2022/aug/06/andrew-tate-violent-misogynistic-world-of-tiktok-new-star.

[153] The Economic Times, *'Sulli Deals', form of hate speech in India, must be condemned: UN official*, 2022, https://economictimes.indiatimes.com/news/india/sulli-deals-form-of-hate-speech-in-india-must-be-condemned-un-official/articleshow/88848351.cms.

[158] Baker, Ging & Andreasen, *Recommending Toxicity: The role of algorithmic recommender functions on YouTube Shorts and TikTok in promoting male supremacist influencers*, 2024, https://antibullyingcentre.ie/wp-content/uploads/2024/04/DCU-Toxicity-Full-Report.pdf.

[159] Bhutkar, Raghvani & Juikar, *User survey about exposure of hate speech among instagram users in India*, 2021, https://www.ijcaonline.org/archives/volume183/number19/bhutkar-2021-ijca-921536.pdf.

Type of Harm	Description	Global Insights	Indian Data	Covered in NCRB/other databases?
	Extremist online content challenges democratic values, reinforces stereotypes and incites violence.[154] Such content usually targets individuals or groups with protected characteristics such as gender, race or disability[155], and incites, promotes and justifies violence, hatred, discrimination or even denigrating those with protected characteristics, whether real or attributed.[156] There are recurrent patterns of repeated and targeted attacks on kids and adolescents representing minority or vulnerable groups based on race, nationality, gender, or sexuality, supported by beliefs that champion negative stereotypes around these groups.[157]		Direct messaging was reported to be a source of explicit content and sexist slurs by these participants. 78.4% of the participants reported psychological effects such as anxiety, depression and even suicidal thoughts from exposure to hate speech, while 61.5% experienced social effects such as isolation and deactivating Instagram accounts.	

[154] Nienierza, Reinemann, Fawzi, Riesmeyer& Neumann, *Too dark to see? Explaining adolescents' contact with online extremism and their ability to recognize it,* 2019,. https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1697339.

[155] Social and Emotional Learning for Mutual Awareness (SELMA), *Hacking Hate Toolkit*, 2019, https://hackinghate.eu/

[156] Council of Europe, *Recommendation of the Committee of Ministers to member States on combating hate speech,* 2022, https://search.coe.int/cm#{%22CoEIdentifier%22:[%220900001680a67955%22],%22sort%22:[%22CoEValidationDate%20Descending%22]}.

[157] Adrienne Katz, Dr Aiman El Asam, Vulnerable Children in a Digital World, 2019, https://www.internetmatters.org/wp-content/uploads/2019/04/Internet-Matters-Report-Vulnerable-Children-in-a-Digital-World.pdf

2.4. Data Quality Issues Exacerbate Challenges of Measuring Online Safety Risks for Women and Children

Data on technology-facilitated harms is crucial for understanding the scope of the problem and informing evidence-based policy responses. However, countries vary in how they track and report these trends. **In India, the National Crime Records Bureau (NCRB) is the primary source for assessing online violence, specifically through its reporting of 'cyber crimes' against women and children. Yet, the limitations of the NCRB's data highlight gaps in capturing the full extent of the issue.** The NCRB reports online violence in India **under the sections that capture 'cyber-crimes' against women and children**. Data is **disaggregated by category for both women and children, as highlighted in *Figures 2.1 and 2.2* below.** As the name suggests, this data can capture– to an extent– 'illegal' acts by perpetrators. **Yet there are limitations to these aggregate statistics, since many online harms get missed due to India's *principal offence rule*[160] that we discuss further later in this report.[161] Due to this rule, Indian institutions are unable to track how cyber-enabled technologies exacerbate or play into the commission of traditional criminal offences.**

Beyond this as we have demonstrated, certain online risks for women and children go beyond illegal activities, and **occupy legally ambiguous territory**. Concerningly, India lacks meaningful consolidated/transparently available data that helps us scope out, classify and quantify such issues.

2.4.1. Insights from NCRB Data on Cybercrime Risks for Women and Children in India

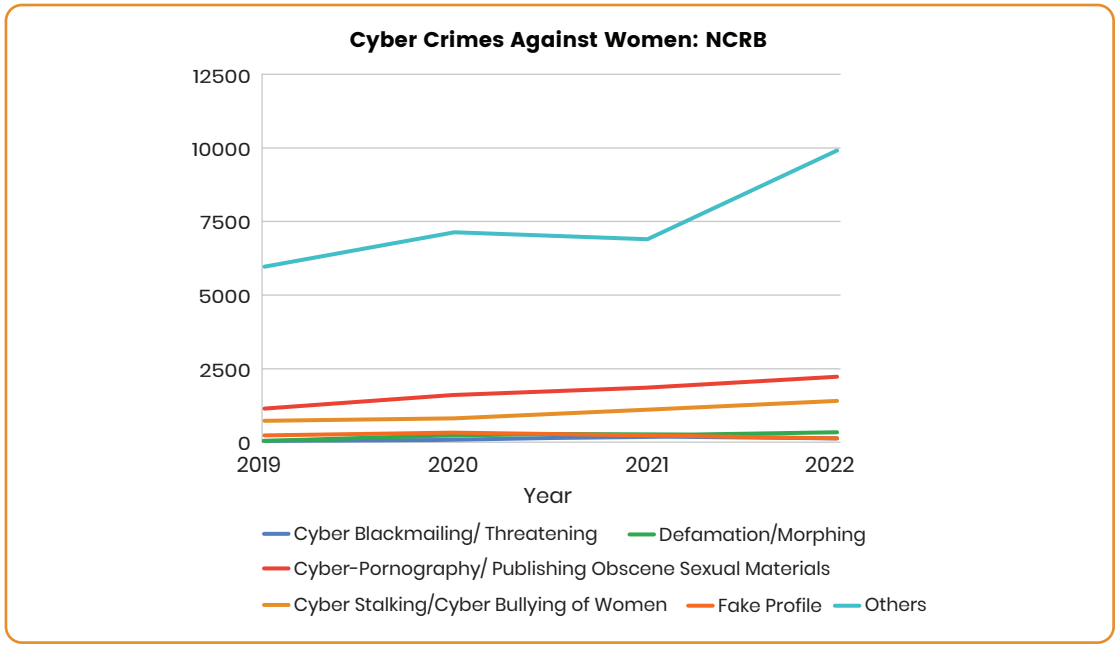


Figure 2.1: NCRB Data Highlighting Crimes Against Women Over The Years

The NCRB annual reports from 2019-2022 indicate that the **total number of reported cybercrimes against women** increased from 8,379 in 2019 to 14,409 in 2022, marking a 71% rise in reports of violence facilitated by technology over four years.[162] The data further disaggregates these crimes into categories such as cyber-stalking, cyber-bullying, defamation, fake profiles, and cyber-pornography.

---

[160] People's Archive of Rural India, *Crime in India 2022: Volume 1,* 2023, https://ruralindiaonline.org/en/library/resource/crime-in-india-2022-volume-i/: "The Principal offence rule means that where multiple offences committed by the same perpetrator are registered in a single First Information Report (FIR), the most heinous offence (with the maximum punishment) is considered and recorded with the NCRB's database. The principal offence rule is followed by most developed nations to avoid an exaggerated perception of crime. However, it can also lead to under-reported figures, especially for crimes against women".

[161] Abhishek Dey, *How the National Crime Records Bureau under-reports crime against women,* 2017, https://scroll.in/article/847053/how-the-national-crime-records-bureau-under-reports-crime-against-women;

Despite these classifications, a **significant proportion of the reported crimes against women fall under the ambiguous category of 'others',** indicating that many reported incidents do not align with the pre-existing categories or data heads. As seen in the analysis above, the act of doxxing—sharing private information about others on the internet is becoming prevalent. However, Indian systems are unable to adequately capture these activities. Similarly, gender-based hate speech against women, though prevalent, is not reflected in the data as the current legal provision on hate speech, Section 196 of the Bharatiya Nyaya Sanhita, 2023 (BNS) does not mention gender expressly as one of its protected categories.[163]

The reliance on the 'others' category suggests that many forms of online harm against women remain under-recognized in official statistics, complicating efforts for policymakers to understand the scope of the full problem. This incomplete categorisation hinders our ability to accurately assess the ground realities, underscoring the need for a more comprehensive and nuanced classification system in reporting. Improved reporting and documentation would not only provide a clearer picture of the types and prevalence of online harms but also enhance accountability and support the development of more effective risk/harm prevention systems.

According to NCRB data, crimes reported against children increased from 305 in 2019 to 1,823 in 2022, indicating a 497% increase across 4 years. According to Lok Sabha data, child pornography cases surged from 44 in 2018 to 738 in 2020, representing a 1,600% increase in reporting during the pandemic.[164] The pandemic played a pivotal role in this surge, as children of all ages shifted online for education and social interaction, amplifying their vulnerability to harmful online behaviour.[165] **However, the absolute numbers of cybercrimes and online risks that are reported in India remain disproportionately low to the size of the population.**

The graph below further highlights trends of heightened cybercrime risks for children in India since the onset of the pandemic:

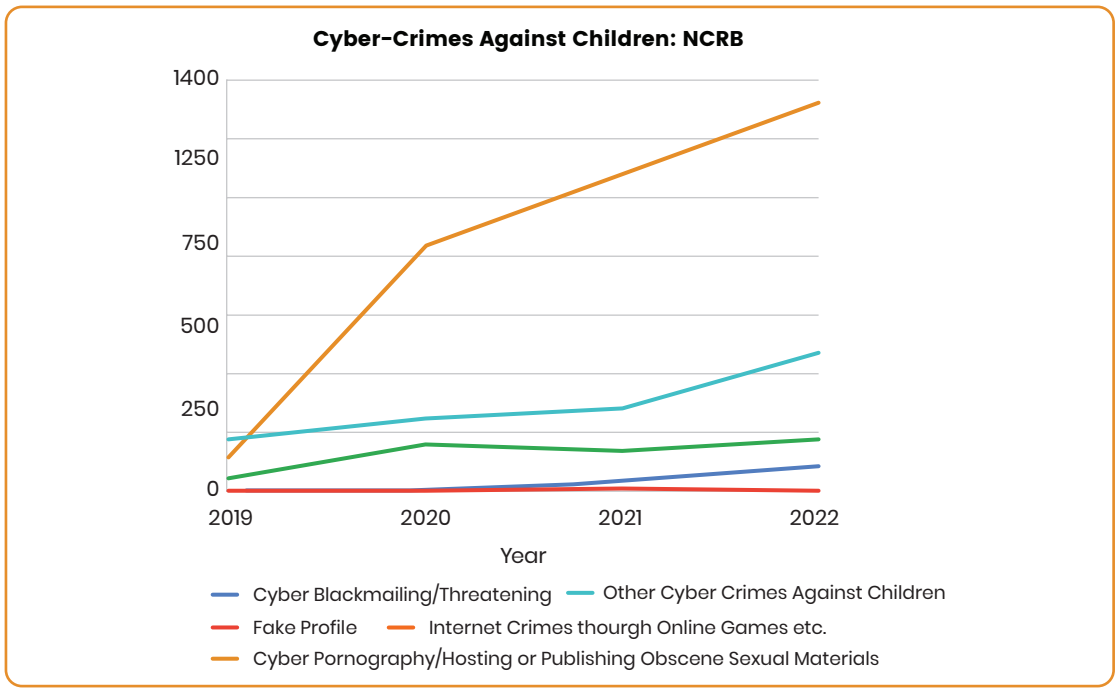


Figure 2.2: NCRB Data Highlighting Crimes Against Children Over The Years

[162] National Crime Records Bureau, *Crime in India Yearwise*, 2022, https://www.ncrb.gov.in/crime-in-india-year-wise.html?year=2022&keyword=.

[163] IT for Change, *Recognize, Resist, Remedy: Addressing Gender-based Hate Speech in the Online Public Sphere*, 2020, https://itforchange.net/online-gender-based-hate-speech-women-girls-recognise-resist-remedy.

[164] Orin Basu, *India witnessed 17-fold rise in child pornography cases, UP & Kerala on top*, 2022, https://zeenews.india.com/india/india-witnessed-17-fold-rise-in-child-pornography-cases-up-kerala-on-top-2512885.html

[165] UNICEF Maldives, *Children at increased risk of harm online during global COVID-19 pandemic - UNICEF*, 2020, https://www.unicef.org/maldives/press-releases/children-increased-risk-harm-online-during-global-covid-19-pandemic-unicef.

2.4.2. Comparing NCRB Data Against Select Platform Transparency Disclosures under India's Intermediary Rules, 2021

Transparency Reports from certain significant social media platforms under India's IT Rules, 2021 provide valuable insights into the scale of online safety risks impacting internet users**.**

For example, in May-June 2024, reports sampled across select platforms[166] collectively indicate that 3,500,051 posts were flagged for sexual content, 1,731,500 posts for harassment and bullying, and 683,052 posts for child sexual exploitation.[167]

Despite the availability of platform-level data and NCRB data, it is still challenging to measure the extent and nature of the risks women and children face on the internet in India.

2.5. Systemic Observations of Indian Data on Online Safety Risks for Women and Children

Some key systemic observations when studying publicly available data are as follows:

1. **Platform-reported incidents within a single month far exceed the figures reported by government institutions. The table below highlights these disparities.** While the NCRB reports 1,171 cases of CSEAM and publishing obscene material, platform transparency reports that have been made available under India's IT Rules, 2021 indicate that over 600,000 posts are removed due to sexually abusive content involving children – **indicating a significant disparity**. This disparity could be a reflection of reporting practices pertaining to abusive content. While many incidents of abuse go unreported at the LEAs, platforms proactively monitor and report activities around online abuse. The graph below compares reporting across three major categories:[168]

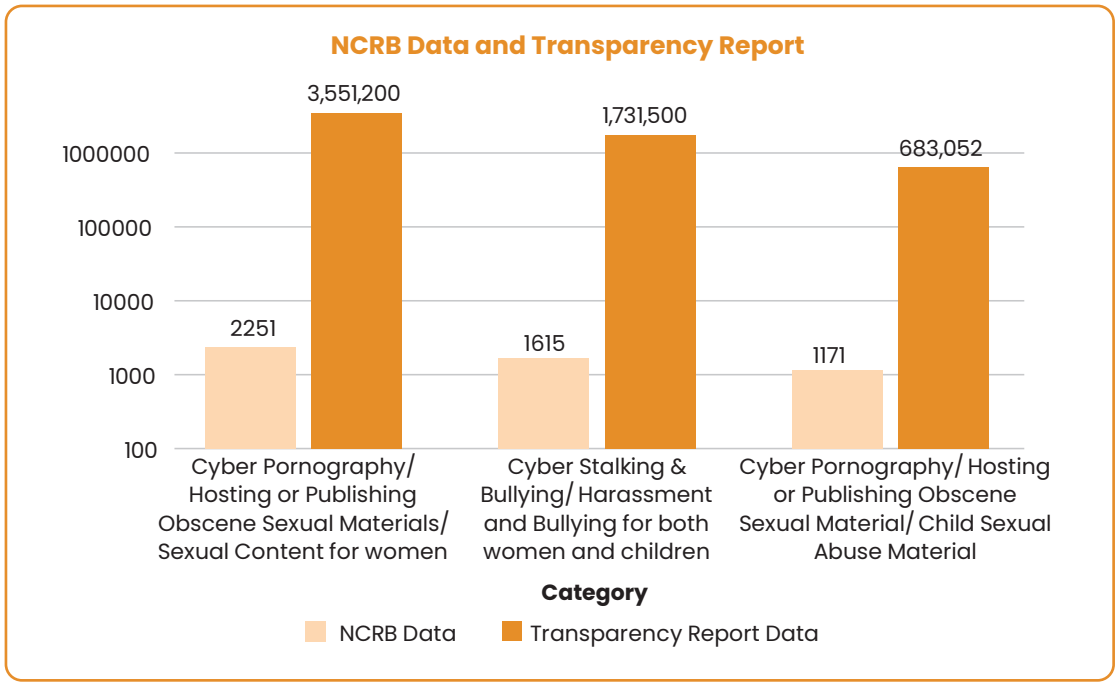


Figure 2.3: NCRB Data And Transparency Reports Data

[166] We referred to monthly transparency reports issued by Meta under the IT Rules 2021, containing data on Facebook and Instagram for the months of May and June 2024: india-monthly-report-June28-2024, India Monthly Report - 31 July 2024 (FB/IG) - Google Docs and the monthly transparency reports issued by Snapchat for May and June 2024: India May 2024 Transparency & Data | Snapchat Transparency, India June 2024 Transparency & Data | Snapchat Transparency.

[167] Numbers aggregated across Instagram, Facebook, Snapchat

[168] Data sources from Transparency reports by Instagram, Facebook, and Snapchat compared with the NCRB report of 2022.

2. **Principal Offence Rule**: **Crimes involving online elements, such as sexual harassment, trafficking or rape facilitated through MMS or videography, are not classified as cybercrimes** due to the '***principal offence rule***'.[169] Instead, these incidents are recorded solely as sexual harassment cases and excluded from cybercrime data, resulting in an incomplete portrayal of the true extent to which the internet facilitates already existing risks for women.[170] India's inability to capture such activities perhaps is attributable to its legal system's inability to categorise and measure "cyber-enabled" crimes.[171]

3. **Absence of Intersectional Data:** Gender-based vulnerability to violence and discrimination is often exacerbated by intersecting forms of marginalisation, including ethnicity, class, sexual orientation, and religion. However, the consolidated data collected is not disaggregated enough to understand specific vulnerabilities. **Research indicates, for instance, that sexual violence against Dalits and other lower-caste girls is more prevalent[172] compared to that against upper-caste women[173]. Similarly, specific religious groups face heightened susceptibility to online violence, as illustrated by the 'Sulli Deals' case, where images of Muslim women were circulated for auction.[174]** Without detailed and disaggregated data, decision makers struggle to identify trends, allocate resources appropriately, and implement targeted interventions that protect the interests of marginalised / vulnerable groups.

Regarding children, the available NCRB data is not disaggregated by sex. Despite existing evidence suggesting that girls may be disproportionately affected, the lack of detailed, sex/gender-disaggregated data prevents a nuanced understanding of these dynamics. **Further, the lack of large-scale or national-level surveys disaggregated by age, gender, caste, or geography prevents a comprehensive picture of online harms against women and children.**

2.6. Absence of Shared Understanding(s) and Taxonomies Impede Online Safety

The discourse on technology-facilitated violence is often overly broad, failing to address its multifaceted nature. **While conversations frequently focus on 'violence,' they tend to overlook other types of 'legal but harmful' practices that have proliferated with the rise of digital media**. A lack of taxonomy further contributes to the data quality issues discussed above. This deficiency in detailed classification results in **ill-fitting one-size-fits-all policy and enforcement responses** that ultimately fail to address the specificities of each type of harm.

To address this problem, the World Economic Forum (WEF) has developed a typology of online harms.[176] Developed by a working group of the Global Coalition for Digital Safety (which includes representatives from industry, governments, civil society, and academia), this typology serves as a foundation for facilitating multistakeholder discussions to establish a common terminology and shared understanding of online safety.

---

[169] If multiple offences are registered under a single first information report (FIR) case, only the most heinous crime is counted as a unit

[170] Gurumurthy & Vasudevan, *Hidden figures- A look at technology-mediated violence against women in India*, 2018, https://itforchange.net/index.php/hidden-figures-a-look-at-technology-mediated-violence-against-women-india.

[171] United Nations, *Basic Facts about the Global Cybercrime Treaty* https://www.un.org/en/peace-and-security/basic-facts-about-global-cybercrime-treaty

[172] Coalition of Feminists for Social Change (COFEM), *Sexual and gender-based violence against Dalit women and girls in India*, 2022, https://cofemsocialchange.org/sexual-gbv-dalit-women-girls-india/#:~:text=Violence%20Against%20Dalit%20Women%20in%20India&text=National%20Crime%20Record%20Bureau%20reported,high%20rate%20of%20almost%20159%25.

[173] Ajay Kumar, *Sexual Violence against Dalit Women: An Analytical Study of Sexual Violence against Dalit Women: An Analytical Study of Intersectionality of Gender, Caste, and Class in India Intersectionality of Gender, Caste, and Class in India*, 2021, https://vc.bridgew.edu/cgi/viewcontent.cgi?article=2680&context=jiws.

[174] The Wire, *'Act of Intimidation and Harm': Rights Activists on 'Sulli Deals' App Targeting Muslim Women*, 2021, 'https://thewire.in/women/sulli-deals-muslim-women-cyber-harassment-statement.

[175] University of Bristol, *Seen but not heard: addressing the silent epidemic of child maltreatment in India*, 2021, Shttps://www.bristol.ac.uk/policybristol/policy-briefings/child-maltreatment-india/

This typology categorises online harms into sub-groups. **For instance, it includes threats to personal and community safety, such as child sexual exploitation material, pro-terror material, and extremist content. It also identifies harms to health and well-being caused by content promoting suicide or disordered eating. The typology recognizes the importance of dignity and privacy by listing examples like bullying, harassment, doxxing, and image-based abuse as violations of these principles.**[177]

Keeping this deficit in mind, the next chapter studies how India's legal system addresses online safety issues that impact women and children in India. The analysis assesses how these issues are covered under Indian IT laws, and parallelly under general and specific criminal laws. It also offers an overview of enforcement practices in India.

[176] World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations*: *Typology of Online Harms*, 2023, https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf.

[177] World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations*: *Typology of Online Harms*, 2023, https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf.

Chapter 3

Analysing India's Relevant Policy and Enforcement Landscape

3.1. Overview of Relevant Legal Frameworks

India's IT Act that regulates the internet was a pioneering legislation that facilitated India's IT/ITeS boom. Subsequent legislative and regulatory amendments have attempted to regulate other aspects of modern digital life. However, the rapid evolution of cyber threats leaves lingering gaps in the legislation. Traditional laws, which were primarily designed to address offline crimes, are now being interpreted to apply to the digital realm.[178] Given India has this patch work system that keeps adapting to changes in digital society, there is a need for a closer examination of its legal system and the potential need for new frameworks to address the complexities of online safety risks that affect women and children.

[178] The laws can often fail to account for the unique characteristics of online interactions and technology leading to ineffective enforcement and lack of resource for victims of cyber-crimes

This chapter analyses the following laws in order to map how online safety risks (especially criminal risks) against women and children are being addressed:

1. **Information Technology (IT) Act, 2000 and accompanying rules/ amendments**: The framework regulates online content and addresses cybercrime, including provisions to combat obscene content online, cheating, violation of privacy, cyber terrorism, misrepresentation, transmitting sexually explicit content (including those involving children). Additionally, the IT Act and attendant rules provide detailed due diligence obligations for online intermediaries like user-generated content platforms, social media intermediaries, P2P messaging services, etc. These obligations set expectations for platforms when they undertake content moderation, for their trust and safety practices, supporting LEAs with investigations, and responding to content takedown requests from government and judicial authorities.
2. **Bhartiya Nyaya Samhita (BNS), 2023**: This legislation overhauled and attempted to modernise India's criminal law framework under the Indian Penal Code ("IPC"), 1860. The BNS has introduced specific provisions addressing digital crimes like cyberstalking and transmitting of obscene content, while the general criminal law provisions continue to apply in online context as well.
3. **Protection of Children from Sexual Offences (POCSO) Act, 2012:** This framework focuses on the protection of children from sexual assault, sexual harassment and pornography, including cyberpornography and online sexual harassment.
4. **Immoral Traffic (Prevention) Act, 1956**: This law addresses trafficking for commercial sexual exploitation, which can intersect with online grooming and the exploitation of women and children. This law is also relevant for cases of technology-facilitated human trafficking[179] and exploitation.
5. **Indecent Representation of Women (Prohibition) Act (IRWA), 1986**: The IRWA prohibits the depiction of women in an indecent manner[180] in advertisements, publications, writings, paintings, figures or in any other manner. The Ministry of Women and Child Development had earlier proposed amending the definition of advertisement in the IRWA to include digital form or electronic form, hoardings, or through SMS, MMS, etc.[181] However, the IRWA remains unamended. While there are no explicit references to the internet and digital media within the framework, it is a special criminal law which has the potential to be repurposed for online activities as well.
6. **Juvenile Justice Act (JJA), 2015**: The JJA establishes protective mechanisms for children in need of care and protection and children in conflict with law. Additionally, the JJA offers a legal system through which criminal offences (including those committed digitally) are addressed via **proportionate punitive and rehabilitation mechanisms designed for offenders below the age of majority**.
7. **Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013 (POSH Act)**: This provides safeguards for women facing sexual harassment in digital workspaces and online professional networks.

Most online safety related interventions in India largely operate within the criminal legal system that we have outlined above. In principle, these frameworks allow victims, survivors and their families to register formal cases against perpetrators at

[179] Council of Europe, *Online and technology-facilitated trafficking in human beings,* https://www.coe.int/en/web/anti-human-trafficking/online-and-technology-facilitated-trafficking-in-human-beings

[180] "Indecent representation of women" means the depiction in any manner of the figure of a woman, her form or body or any part thereof in such a way as to have the effect of being indecent, or derogatory to, or denigrating, women, or is likely to deprave, corrupt or injure the public morality or morals; Section 2(c) of the Act.

[181] Press Information Bureau, *WCD proposes amendments to widen the scope of Indecent Representation of Women (Prohibition) Act (IRWA), 1986*, 2018, https://www.pib.gov.in/PressReleasePage.aspx?PRID=1534316

police establishments. Additionally, Indian institutions are creating presenceless and digital avenues for reporting such as the National Cybercrime Reporting Portal (**NCRP**) and the accompanying national helpline number (**1930**) that is **managed by the Ministry of Home Affair's Indian Cyber Crime Coordination Centre (I4C)**.[182]

Additionally, victims have avenues to pursue other forms of remedy (vis-a-vis content removal and harm reduction) through online platforms that qualify as intermediaries[183] and significant social media intermediaries ("SSMIs")[184], and host third-party user generated content. In these circumstances victims have the opportunity to file complaints with these platforms and these platforms have to:

- In general scenarios, acknowledge complaints within 24 hours and resolve the same within 15 days[185]; or
- In scenarios where harmful content relates to the **victim being depicted (even artificially)** performing sexual activities or exposing their private areas, intermediaries are required to take all reasonable and practicable measures to remove or disable access to such content within 24 hours.[186]

Despite the availability of such legal frameworks, subsequent analysis within the chapter demonstrates that the implementation of these regimes are not aligning with the needs of victims and survivors. This is due to a lack of citizen awareness, lengthy and burdensome processes with respect to LEAs, and challenges associated with the digital divide and other accessibility barriers. **Therefore, India's overall landscape presents a systemic enforcement deficit where children and women are unable to utilise these remedies effectively.**

Tables 3.1 and 3.2 represent how different digital crimes against women and children are addressed across different legislation. Each of these tables' classifications of these crimes draws on classifications that we observe on the National Cyber Crime Reporting Portal (NCRP) website.[187]

Table 3.1: Legal Provisions Governing Online Harms against Women in India

Offence	IT Act	BNS	IRWA	Other Laws
Pornography	S. 66E, 67, 67A	S. 75, 77, 294	S. 3, 4	-
Cyber Bullying	S. 67	S. 79, 75, 351	-	S. 3 POSH
Cyber Stalking	S. 67, 67A	S. 75, 78	-	S. 3 POSH
Online Sextortion	S. 67, 67A	S. 75, 77, 79, 111, 294, 351	-	-
Cyber Grooming	S. 66D	S. 69, 75	-	-
Identity Theft	S. 66C, 66D	S. 315, 335, 336	-	S. 15(b) DPDPA
Data Breach	S. 66B, 72, 72A	S. 303, 316	-	S. 8(5) DPDPA
Trafficking	-	S. 143	-	S. 5 ITPA Act
Deepfakes/Morphing (involving nudity)	S. 66D, 66E, 67	S. 294, 335, 336, 356	S. 3, 4	S. 51, Copyright Act

[182] Indian Cybercrime Coordination Centre (I4C), *National Cybercrime Reporting Portal*, https://i4c.mha.gov.in/ncrp.aspx
[183] Information Technology Act, Section 2(1)(w).
[184] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 Rule 2(1)(v)
[185] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rule 3(2)(a)(i)
[186] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 Rule 3(2)(b)
[187] Indian Cyber Crime Coordination Center (I4C), *National Cyber Crime Reporting Portal,* https://cybercrime.gov.in/.

Table 3.2: Legal Provisions Governing Online Harms Against Children in India

Offence	IT Act	BNS	POCSO	Other Laws
Pornography/CSAM	S. 66E, 67, 67A, 67B	S. 75, 77, 294, 295	S. 13, 15	-
Cyber Bullying	S. 67	S. 79, 75, 351	S. 11	-
Cyber Stalking	S. 67, 67A, 67B	S. 75, 78	S. 11	-
Online Sextortion	S. 67, 67A, 67B	S. 75, 77, 79, 111, 294, 351	S. 11	-
Cyber Grooming	S. 66D	S. 69, 75	S. 11	-
Trafficking	-	S. 143	S. 13, 14	S. 5, 6 ITPA
Deepfakes/Morphing (involving nudity)	S. 66D, 66E, 67, 67B	S. 294, 295, 335, 336, 356	S. 11, 13, 15	-

Even though the provisions mentioned in the table may not explicitly recognise the cyber crimes themselves, there are provisions corresponding to the elements that make up the crime in question. **For instance, there are no explicit provisions to cover cyber stalking, however law enforcement officials might utilise provisions covering harassment / stalking in general to book the perpetrators.**

Notably, NCRP classification fails to cover several major risks that children and women commonly face online, such as misinformation, deepfakes, or unsolicited explicit content. Additionally, the list overlooks the critical overlap between online and offline risks, where harm initiated in digital spaces—such as cyber grooming or stalking—can escalate rapidly into physical threats, exacerbating the impact on victims.

Therefore, India's legal design (and thus its enforcement culture) suffers from one systemic deficit i.e. its inability to distinguish between cyber-enabled and cyber-dependent digital offences.

Box 3.1: What are Cyber-Enabled Offences and Cyber-Dependent Offences?

***Cyber-enabled Offences*:**

- Traditional crimes that exist offline and are amplified, facilitated, exacerbated or scaled through digital technologies are cyber-enabled offences.[188]
- This includes criminal activity that conceptually exists beyond cyberspace and can be committed without digital technologies.
- In such instances, the digital service acts as a tool or medium to allow the criminal activity to be committed.
- These crimes do not depend on computer networks, but have been transformed in scale or form by the use of ICT.[189]
- Examples include drug and weapons trafficking, identity theft, fraud and incitement of violence.

***Cyber-dependent Offences*:**

- Crimes that can only be committed through the use of computer networks or ICT are cyber-dependent crimes.[190]
- In such offences, the ICT devices or networks are the medium or tool for committing the crime and the target of the crime.
- Without the internet, these crimes could not be committed.
- They broadly fall into two categories:[191]
  - *Illicit intrusions into computer networks (such as hacking), and*
  - *Disruption or downgrading of computer functionality and network space (such as malware or Denial of Service (DOS).*

[188] European Parliament, *Understanding cybercrime,* 2024, https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/760356/EPRS_BRI(2024)760356_EN.pdf.

[189] UNU - Centre for Policy Research, *Understanding the UN's new international treaty to fight cybercrime,* 2024, https://unu.edu/cpr/blog-post/understanding-uns-new-international-treaty-fight-cybercrime.

[190] UNU - Centre for Policy Research, *Understanding the UN's new international treaty to fight cybercrime,* 2024, https://unu.edu/cpr/blog-post/understanding-uns-new-international-treaty-fight-cybercrime.

[191] Crown Prosecution Service, Government of the UK, *Cybercrime - prosecution guidance,* 2024, https://www.cps.gov.uk/legal-guidance/cybercrime-prosecution-guidance.

Therefore, under the status quo Indian police officials rely on broad criminal provisions to prosecute cyber criminals. **At a structural level, the table highlights a landscape where cybercrime cases are predominantly managed under general criminal laws, with Indian IT laws addressing only a limited set of technology-faciliated offences that are particularly relevant to women and children**.

3.2. Role of Law Enforcement Agencies (LEAs)

In this section, we examine how LEAs navigate this landscape of broad criminal laws and how successfully they are able to protect women and children from online safety risks.

3.2.1. Coordination Efforts in Cybercrime Investigation in India

Under the Seventh Schedule of the Constitution of India, 'Police' and 'Public Order' are State subjects. This means that the States and Union Territories (UTs) are primarily responsible for the prevention, detection, investigation, and prosecution of crimes, through their respective state-level Law Enforcement Agencies (LEAs).

Moreover, cyber crimes transcend state and national boundaries, complicating jurisdictional issues that necessitate a coordinated response. **The absence of standardised response playbooks and coordinated mechanisms across state jurisdictions often leads to fragmented and delayed responses to cyber crimes.** Currently, individual states issue their own cyber crime investigation manuals, reporting checklists, and protocols - for instance, those adopted by Jharkhand, Puducherry, and Odisha[192] - which, while well-intentioned, differ in scope, structure, and procedural emphasis. **This fragmented approach results in inconsistent practices for identifying, recording, escalating, and investigating cyber offences. As a consequence, inter-state investigations suffer from delays in information sharing, misaligning standards for evidence, and confusion over jurisdictional authority - particularly when offences originate in one state but affect victims in another.[193] Online platforms also face challenges in assisting these different LEAs since they are navigating varying protocols across states, leading to delayed or incomplete responses to data requests.[194]**

The Indian Government has established the I4C under the Ministry of Home Affairs (MHA) for centralising the reporting and investigation of cyber crimes across the country through initiatives like the **National Cyber Crime Reporting Portal**.[195] Reports submitted through this portal are forwarded to state-level law enforcement agencies, which are responsible for converting them into First Information Reports (FIRs) and initiating further legal action, if necessary. I4C has also constituted seven Joint Cyber Crime Coordination Teams (JCCTs) to enhance inter-state coordination among law enforcement agencies across States and Union Territories. In 2023, seven JCCT workshops were held to train LEA in different states.[196] I4C is also responsible for facilitating joint identification, prioritisation, and initiation of multi-jurisdictional actions against cybercrimes. **However, there are currently no specialised JCCTs focused specifically on issues affecting women and children.**

---

[192] CID Crime Branch, Odisha Police, *Offences and Relevant Penal Sections in Cyber Crime*, 2021 https://odishapolicecidcb.gov.in/sites/default/files/Relevant%20Penal%20sections%20Cyber%20Crime.pdf; Cyber Crime Cell, Puducherry Police, *Cyber Crimes Investigation Guidelines*, 2024, https://police.py.gov.in/Cyber%20Crime%20-%201%20Investigation%20check%20list%20by%20Dr%20Bascarane%20SP%20Cyber%20dt%2014.02.24.pdf; Data Security Council of India (DSCI), *Cyber Crime Investigation Manual,* 2011, https://jhpolice.gov.in/sites/default/files/documents-reports/jhpolice_cyber_crime_investigation_manual.pdf.

[193] Kuldeep Bairwa, *Cyber Crime and Laws in India,* 2020, http://oldopac.nls.ac.in:8081/xmlui/bitstream/handle/123456789/407/887.pdf; Narayan Namboodiri, *Interstate Cybercrime Challenges Mumbai Cops,* 2023, https://timesofindia.indiatimes.com/city/mumbai/interstate-cybercrime-challenges-mum-cops/articleshow/97076507.cms.

[194] Smriti Parsheera and Prateek Jha, *Cross-Border Data Access for Law Enforcement: What Are India's Strategic Options?,* 2020, https://carnegieendowment.org/research/2020/11/cross-border-data-access-for-law-enforcement-what-are-indias-strategic-options?lang=en.

[195] I4C, *National Cybercrime Reporting Portal (NCRP),* https://i4c.mha.gov.in/ncrp.aspx.

[196] PIB Press Release, 2024, https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1992949.

The MHA also oversees the Cyber Crime Prevention against Women and Children (CCPWC) scheme,[197] implemented under the Nirbhaya Fund[198]. The initiative aims to address and prevent cybercrimes targeting women and children. However, an analysis of the scheme reveals that it largely consolidates pre-existing initiatives for general cybercrimes, such as the NCRP portal and a toll-free helpline (1930), which are **not tailored or sensitive to the needs of women and children in distress**. India's National Commission for Women (NCW) has recommended improving the scheme through online women-specific crime reporting units for seamless complaint forwarding and acknowledgment, as well as establishing a monitoring unit for monthly reporting on cybercrime complaints.[199] As of June 2025, the MHA has yet to take the recommendation forward.

Notably, partnerships such as India's MoU with the U.S.-based National Center for Missing and Exploited Children (NCMEC) for sharing reports on online child exploitation, have been subsumed into the aforementioned scheme. However, the Supreme Court has recently deemed this arrangement inadequate, asking platforms to report such cases directly to local law enforcement as well.[200] Responding to such gaps, there have been press reports that suggest that the MHA is considering a proposal to establish a specialised central police unit to address issues related to child pornography. Additionally, similar dedicated cells may be set up in each state.[201]

In parallel, the MHA has also launched the **Sahyog platform[202]** under the I4C to enable seamless real-time coordination between law enforcement agencies, social media platforms, telecom service providers and other stakeholders involved in cyber crime response. **The platform was created in response to persistent concerns from law enforcement about delays in content takedown, lack of cooperation from intermediaries, and the absence of a standardised communication channel.** The Sahyog portal is designed to facilitate quicker reporting and removal of unlawful content by **automating the process of sending notices from authorised government agencies to intermediaries, in alignment with the takedown process prescribed under Section 79(3)(b) of the IT Act.** This creates a mechanism that streamlines communication, **case tracking and real-time information sharing amongst state LEAs and offers a secure, centralised system for managing takedown requests, evidence preservation, and status updates.** As of April 2025, the portal has onboarded 65 intermediaries, notified nodal officers in 28 states and 5 Union Territories[203] and issued 130 takedown orders[204]. In a more recent push, I4C has directed all social media platforms to formally join the Sahyog portal and appoint dedicated grievance and compliance officers specifically for reporting and acting on CSAM (in line with the Supreme Court order referred above), signalling the government's effort to build a technologically integrated and accountable response to cybercrimes.[205] While the Sahyog platform has strong potential, it

---

[197] Ministry of Home Affairs, Government of India, *Details about CCPWC (Cybercrime Prevention against Women and Children) Scheme,* https://www.mha.gov.in/en/division_of_mha/cyber-and-information-security-cis-division/Details-about-CCPWC-CybercrimePrevention-against-Women-and-Children-Scheme.

[198] Post-2012, Nirbhaya Gang rape case, a dedicated fund was set up in 2013 with the focus on implementing the initiatives aimed at improving the security and safety of women in India.

[199] National Commission for Women, *Cyber Crime Prevention Against Women and Children,* http://ncw.nic.in/node/1762

[200] Satya Prakash, *Downloading, watching sexually explicit material involving children crime: SC*, 2024, https://www.tribuneindia.com/news/india/storage-of-child-pornography-crime-under-pocso-it-acts-rules-supreme-court/.

[201] Rahul Chabbra, *Central police unit likely to deal with child pornography cases,* 2023, https://sundayguardianlive.com/news/central-police-unit-likely-to-deal-with-child-pornography-cases.

[202] I4C, Ministry of Home Affairs, *Sahyog Portal*, https://sahyog.mha.gov.in/.

[203] Soumyendra Barik, *130 censorship orders issued via Home's Sahyog portal in 5 months,* 2025 https://indianexpress.com/article/express-exclusive/130-censorship-orders-issued-via-homes-sahyog-portal-in-5-months-9957698/

[204] Soumyendra Barik, *130 censorship orders issued via Home's Sahyog portal in 5 months,* 2025 https://indianexpress.com/article/express-exclusive/130-censorship-orders-issued-via-homes-sahyog-portal-in-5-months-9957698

[205] Aihik Sur, *I4C asks social media platforms to join Sahyog portal, appoint officers to report child sexual abuse material*,2025,https://www.moneycontrol.com/technology/i4c-asks-social-media-platforms-to-join-sahyog-portal-appoint-officers-to-report-child-sexual-abuse-material-article-13026765.html#google_vignette.

is currently facing constitutional challenges raised by X calling it a 'censorship portal' in the Karnataka High Court, so its future remains uncertain.[206]

Similarly, other institutions and states have made efforts to provide specialised and dedicated complaint mechanisms for children affected by cybercrimes. The National Commission for Protection of Child Rights (NCPCR) has established the **POCSO e-Box**, an online complaint management system to ease direct reporting by children of offences under the POCSO Act, including cyberbullying, cyberstalking and child pornography.[207]

Aside from the above examples, the launch of the "Pratibimb" platform, and weekly peer learning sessions for cyber cell officers from across the country are also key initiatives aimed at improving collaborative capabilities. These sessions have facilitated knowledge-sharing and real-time solutions to challenges in cybercrime investigations, which has led to the live tracking and arrest of 6,046 accused persons and assisted in 36,296 law enforcement investigation requests.[208]

To build on the Pratibimb software, the I4C launched the Samanvay Platform (formerly known as the Joint Cybercrime Investigation Facilitation System) which goes beyond facilitating coordination between online platforms and LEAs, extending its reach to a wider range of digital platforms including financial intermediaries, payment aggregators, banks, telecom providers and other intermediaries. **This platform brings heterogenous intermediaries, whose data conventionally exists in silos, together on a single digital interface and eases multi-agency collaboration by creating common protocols and an operational dashboard.** This allows investigating officers to simultaneously request and receive information on bank transactions, SIM card ownership, call records, IMEI numbers of devices and platform activity - dramatically reducing the time spent chasing data across entities.[209] **Certification from the Massive Open Online Courses (MOOC) platform 'CyTrain' developed by I4C for police officials, is also available to the police, law enforcement and judiciary to allow for a better informed response to emerging cyber challenges.[210]**

In a more localised context, state initiatives vary greatly. For example, the **Government of Kerala has launched the Kunjapp platform[211], in collaboration with the state's Child Welfare Committee (CWC) and Juvenile Justice Board (JJB)**, to facilitate the reporting of cybercrimes and improve access to rehabilitation services for child victims. While the CWCs and JJBs have traditionally focused on providing care, protection and rehabilitation to children facing physical neglect or abuse, the increasing prevalence of online harms urges the thought on how their protective and rehabilitative role could be extended to cases involving online risks, abuse and exploitation. **Leveraging the capacity and wide-reaching network of existing institutions and frameworks that are already well-positioned to respond to online harms against children would result in more effective policy outcomes**.

---

[206] Medianama, *Explained: What is the Sahyog Portal that X has called out for censorship?*, 2025, https://www.medianama.com/2025/04/223-explained-what-is-the-sahyog-portal-that-x-called-out-for-censorship/?utm

[207] Press Information Bureau, Cyber crimes against children can now be reported at the POCSO e-Box, 2017, https://www.pib.gov.in/newsite/PrintRelease.aspx?relid=166857.

[208] Press Trust of India, *6,046 cyber criminals arrested with help of 'Pratibimb' module: Govt*, 2025, https://www.business-standard.com/india-news/6-046-cyber-criminals-arrested-with-help-of-pratibimb-module-govt-125031100770_1.html

[209] Press Trust of India, *'Samanvaya' portal for data exchange among police forces on cyber criminals' activities: MHA*, 2024, https://theprint.in/india/samanvaya-portal-for-data-exchange-among-police-forces-on-cyber-criminals-activities-mha/2273246/

[210] National Crime Records Bureau, *National Cybercrime Training Centre (CyTrain),* https://cytrain.ncrb.gov.in/.

[211] Press Trust of India, *Kerala govt launches mobile app to prevent cyber crimes against children,* 2022, https://www.business-standard.com/article/current-affairs/kerala-govt-launches-mobile-app-to-prevent-cyber-crimes-against-children-122102200936_1.html

[212] Jharkhand Police, Online Service for Cyber Crime Related Investigation Cooperation Requests, https://jhpolice.gov.in/node/32549.

Separately, Jharkhand has been particularly proactive by launching an online platform[212] that allows cooperation requests from other states where victims have lodged FIRs related to cyber crimes with possible links to Jharkhand. However, most other states lack similar systems. The cost of inter-state investigations often acts as a disincentive, especially given the limited resources of police forces.[213] This challenge is compounded by current police procedures that do not encourage allocating resources to cases originating outside their own state, compounded by a shortage of personnel.[214]

3.2.2. Primary Challenges in Cyber Crime Investigation

While the Central Government has highlighted[215] that the NCRP Portal has been accessed over 140 million times and that 3.1 million cybercrime complaints have been filed, to demonstrate its credibility and effectiveness, significant challenges remain.

1. **Variability in Standard Operating Procedures (SOPs):** Different state agencies across India implement varying SOPs for cybercrime investigations, resulting in inconsistent approaches to similar cases.[216] This inconsistency complicates the process, making it difficult for law enforcement officers to follow uniform guidelines, particularly in cases spanning multiple jurisdictions. Inconsistent SOPs and the lack of technical expertise often result in poorly framed or incomplete data requests (for the purposes of investigations) to intermediaries. This leads to delays in receiving crucial evidence, such as user activity logs, IP addresses, or account details.
2. **Limited Availability of Internal Technical and Forensic Expertise Within Law Enforcement:** Cybercrime cases often require specialised technical knowledge, which many law enforcement agencies lack.[217] While LEAs sometimes rely on private firms and cyber intelligence agencies to provide technical/forensic assistance, such as tracking IP addresses or analysing responses from social media companies, there are no clear policies to determine when external experts can formally assist in investigations.[218] This creates gaps in technical support during investigations.
3. **Restrictive Legal Provisions under the IT Act:** Sections 78 and 80 of the IT Act **restrict the investigation of cybercrimes to police officers of the rank of inspector and above**.[219] However, this creates a procedural bottleneck, as there is a shortage of such officers with the necessary ranks/designations.[220] **Legal restrictions prevent sub-inspectors and lower officials, who are often on the front-lines, from taking lead roles in investigations under the IT Act, exacerbating delays. Further, these lower ranked officials thus are incentivised to find work arounds and investigate or file cases**

---

[213] Regina Mihindukulasuriya, *What can lawmakers, regulators and judiciary do to reduce cybercrime,* 2024 https://www.deccanherald.com/amp/story/business/what-can-lawmakers-regulators-and-judiciary-do-to-reduce-cybercrime-3019451

[214] Aditi Dehal and Mrigank Patel, *Inter-state Police Cooperation: Challenges, Policy Framework, and the Way Forward,* 2022, https://sprf.in/wp-content/uploads/2022/10/ISPC-IB.pdf

[215] Press Information Bureau, Press Release, 2024, https://pib.gov.in/PressReleasePage.aspx?PRID=1992949.

[216] CID Crime Branch, Odisha Police, *Offences and Relevant Penal Sections in Cyber Crime*, 2021 https://odishapolicecidcb.gov.in/sites/default/files/Relevant%20Penal%20sections%20Cyber%20Crime.pdf; Data Security Council of India (DSCI), *Cyber Crime Investigation Manual,* 2011, https://jhpolice.gov.in/sites/default/files/documents-reports/jhpolice_cyber_crime_investigation_manual.pdf.

[217] Vivek Sharma, Dr. Hemant Kumar Harti, *Need for Imparting Training to Officials to Investigate Cyber Crimes,* 2021, https://iajesm.in/admin/papers/648dbf9b81369.pdf.

[218] Anandi Chandrashekhar and Shashwat Mohanty, *Police in states across India are relying on firms and consultants to solve cyercrimes,* 2019, https://economictimes.indiatimes.com/news/politics-and-nation/police-in-states-across-india-are-relying-on-private-firms-and-consultants-to-solve-cybercrime-cases/articleshow/72499885.cms?from=mdr

[219] Vaishnavi Thakur and Anurag Gupta, *An Analysis Of Cybercrime Investigation And Surveillance*, 2022, https://www.ijllr.com/post/an-analysis-of-cybercrime-investigation-and-surveillance

[220] Umesh Kumar Ray, *Lessons From Bihar: Why It's Difficult For Cops To Track Down Cybercriminals*, 2024, https://www.boomlive.in/decode/why-its-difficult-for-cops-to-track-down-cybercriminals-26526

under general criminal laws. This often leads to sub-optimal enforcement, as traditional statutes were not designed for digital contexts - applying provisions like Section 318 of the BNS (cheating) to cybercrimes introduces ambiguities in jurisdiction, evidentiary standards, and jurisdictional reach, making convictions harder to secure.[221]

4. **Mismatched Legal Provisions:** When law enforcement applies legal provisions inconsistently in FIRs for cyber crimes, it often results in investigations falling short. Officers tend to use different sections of the BNS (formerly IPC] and the IT Act for what are essentially the same offences.[222] **A lack of a standardized approach can delay investigations, as officials struggle to navigate overlapping or mismatched legal frameworks.[223] It also makes it harder to build a body of legal precedents, which is crucial for developing the expertise needed to tackle increasingly complex cybercrimes.[224]** When the wrong charges are applied, individuals can find themselves facing undue legal difficulties, with the victim experiencing feelings of guilt, fear and helplessness, sometimes worse than their original victimization.[225] On the other hand, using insufficient or inappropriate provisions can give offenders an easy way out, eroding trust in the justice system. **For instance, police officials often tend to impose charges for more serious offences in cybercrime cases, which do not pass muster in a court of law.[226] This leads to a lose-lose situation where the accused may be wrongly implicated, genuine offenders might escape punishment, and valuable judicial resources are wasted in the process.**

5. **Reluctance in filing FIR -** A recent circular[227] by the Puducherry Police about the backlog of pending investigations of complaints filed in the NCRP Portal for their jurisdiction, particularly in sensitive cases involving women and children, is symptomatic of the state of cyber crime investigation in the country. **Key reasons behind such backlogs include delays in data collection from social media, telecom, and internet service providers, difficulties in organising case files, lack of clarity around the required investigation check-list, insufficient action on social media IDs involved in the crime, and confusion over the roles of platforms versus law enforcement authorities in initiating action.[228]** The situation is further complicated by the varying SOPs implemented by different state agencies, leading to increased complexity in the investigation process.

These challenges collectively highlight the urgent need for policymakers to work towards raising the capacity of Indian LEAs with:

- Standardisation of protocols for investigation, evidence collection and reporting across different states[229];

---

[221] Vinay K, *Challenges faced by law enforcement agencies in investigating and prosecuting cyber crimes in India,* 2024,https://www.researchgate.net/publication/383785171_CHALLENGES_FACED_BY_LAW_ENFORCEMENT_AGENCIES_IN_INVESTIGATING_AND_PROSECUTING_CYBER_CRIMES_IN_INDIA.

[222] R.M.Kamble & C.Vishwapriya, *Cyber Crimes and Information Technology*, 2008, https://nalsar.ac.in/images/Nalsar%20Law%20Review-Vol.%204.pdf

[223] *Legal gaps and concerns abound as cybercrime rises unabated in India*, 2024, https://ciso.economictimes.indiatimes.com/news/cybercrime-fraud/legal-gaps-and-concerns-abound-as-cybercrime-rises-unabated-in-india/106434980

[224] Information Security Forum, *Review and Gap Analysis of Cybersecurity Legislation and Cybercriminality Policies in Eight Countries,* https://www.securityforum.org/solutions-and-insights/review-and-gap-analysis-of-cybersecurity-legislation-and-cybercriminality-policies-in-eight-countries/

[225] Irazola, et al., *Addressing the Impact of Wrongful Convictions on Crime Violations*, 2014, https://www.ojp.gov/pdffiles1/nij/247881.pdf

[226] Joanna Curtis and Gavin Oxburgh, *Understanding cybercrime in 'real world' policing and law enforcement*, 2022, https://journals.sagepub.com/doi/10.1177/0032258X221107584

[227] Cyber Crime Cell, Puducherry Police, Cyber Crimes Investigation Guidelines, 2024, https://police.py.gov.in/Cyber%20Crime%20-%201%20Investigation%20check%20list%20by%20Dr%20Bascarane%20SP%20Cyber%20dt%2014.02.24.pdf.

[228] Cyber Crime Cell, *Puducherry Police, Cyber Crimes Investigation Guidelines*, 2024, https://police.py.gov.in/Cyber%20Crime%20-%201%20Investigation%20check%20list%20by%20Dr%20Bascarane%20SP%20Cyber%20dt%2014.02.24.pdf.

[229] Jain & Kanwar, *Legal Framework of Rising Threat of Cybercrime in India Challenges and Emerging Issues*, 2024, https://ijrpr.com/uploads/V5ISSUE11/IJRPR34835.pdf.

- Improved cross-jurisdictional coordination[230];
- Capacity building, specialised training and providing access to technical expertise and infrastructure[231];
- Real-time information sharing pathways amongst LEAs, intermediaries and other stakeholders that are proportionate and consistent with citizens' various constitutional interests like the right to privacy[232]; and
- Preventive measures, such as early threat detection and warning systems, real-time risk flagging and digital literacy campaigns[233].

3.2.3. Relevant Legal Framework for Coordination with Online Platforms i.e. Internet Intermediaries

The legal framework concerning intermediary obligations and law enforcement's access to information in India comprises disparate provisions under the IT Act, its associated rules, and the Bharatiya Nagarik Suraksha Sanhita, 2023. **These provisions have to balance national security, public order, and investigative needs, and have to be proportionate and consistent with citizens' various constitutional interests like the right to privacy.** However, ambiguities in definitions, broad discretionary powers, and procedural gaps often lead to challenges in implementation and ensuring consistency with the rule of law[234].

[230] Parsheera & Jha, *Cross-Border Data Access for Law Enforcement: What are India's Strategic Options?*, 2020, https://carnegie-production-assets.s3.amazonaws.com/static/files/ParsheeraJha_DataAccess.pdf.

[231] https://vidhilegalpolicy.in/wp-content/uploads/2020/06/VidhiBriefingBook_LawinNumbers.pdf

[232] World Economic Forum, *Cyber Information Sharing: Building Collective Security*, 2020, https://www3.weforum.org/docs/WEF_Cyber_Information_Sharing_2020.pdf.

[233] The Daily Guardian, *Delhi Police launches 'Cyber Suraksha' campaign to tackle cybercrime*, 2023, https://thedailyguardian.com/india/delhi-police-launches-cyber-suraksha-campaign-to-tackle-cybercrime/

[234] United Nations, *What is the Rule of Law,* https://www.un.org/ruleoflaw/what-is-the-rule-of-law/; World Justice Economy, *What is the Rule of Law?,* https://worldjusticeproject.org/about-us/overview/what-rule-law.

Table 3.3: Overview of Key Statutory Provisions related to investigation powers

Provision	Key Requirements	Concerns
Rule 3(1)(j) & 4(2), IT Rules 2021	• Digital services that host third party content i.e. intermediaries must provide information or assistance to authorised LEAs within 72 hours of a lawful information request (24 hours in case of an online gaming intermediary). Information requests can involve identity verification, or can be in relation to investigation, prevention, or cybersecurity purposes. • Order can be issued by any government agency that is notified under this provision as per the IT Act.[235] • Orders must be issued in writing and state the purpose.	• Terms like "information under its control or possessions"[236] and "assistance" are broad and undefined.[237] • Scope of information not clearly limited, leading to concerns of excessive data requests which can conflict with citizen privacy imperatives. • Local police units or LEAs with low expertise may lack the legal or technical drafting capacity to frame requests that meet platform standards in storing data. **This creates challenges for police and LEAs to lawfully request data from intermediaries in a manner that is consistent with Constitutional imperatives**. • In time-sensitive situations particularly involving high risks to women and children, there is a need for clear and predictable standards to guide platforms in prioritising information sharing with law enforcement along with timely content takedown. **It becomes important to consider real-time coordination solutions, to help victims effectively respond within the first few hours of an incident, in order to mitigate the far-reaching effects of online abuse when the harm is circulating online at a rapid rate**.

[235] The Economic Times, *10 central agencies can now snoop on "any" computer they want*, 2018, https://economictimes.indiatimes.com/news/politics-and-nation/10-central-agencies-can-now-snoop-on-any-computer-they-want/articleshow/67188875.cms?from=mdr

[236] Narrow interpretation of the phrase "information under its control or possession" could result in the platforms' refusal to share relevant metadata, logs, archived/cached content, etc. that could aid investigations.

[237] Sarkar, et al., On the legality and constitutionality of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, 2021, https://www.medianama.com/2021/06/223-legality-constitutionality-of-it-rules/.

Provision	Key Requirements	Concerns
	• Significant social media intermediaries (SSMI) providing messaging services are required to identify the first originator of any information that they have stored, if required by a court order or a government order under S.69 of the IT Act that deals with the interception and decryption of information for national security and public order purposes. **To be clear this provision on *traceability* has been currently challenged on grounds of constitutionality within Indian courts (see 'Concerns' column on the right hand side)**. • As per the IT Rules a traceability order can be issued to SSMI for offences related to CSAM, sexually explicit content among others.	• Rule 4(2) stipulates that the traceability order (to identify the first originator) can be issued only in the absence of less intrusive means to investigate or prevent a crime. However, the traceability mandate has been heavily contested as being violative of the right to privacy and free speech.[238] This provision is currently under challenge in the Delhi High Court by WhatsApp, which has argued that breaking encryption will sound the death knell for the application since users use it for that very purpose.[239] Petitioners, including some SSMIs, have raised concerns that a traceability mandate under Rule 4(2) may pose privacy and security risks and break end-to -end encryption.[240] The Government of India has defended the provision by claiming that this measure would only be adopted if less invasive and restrictive alternatives have been proven to be ineffective, that there exists international precedent for such an obligation, and that it is permitted under India's legal system.[241] Even so, scholars have pointed out that breaking encryption may be an excessively onerous obligation which will affect a platform's usability and users' privacy and security, and that such an obligation is arguably impermissible under India's constitution and the parent legislation, i.e. the IT Act, 2000.[242]
Section 69 & 69B, IT Act	• Governments can intercept, monitor, or decrypt information for national security or public order investigations.	• Civil society has expressed concerns about the rules that operationalise these provisions arguing they have inadequate procedural safeguards and transparency requirements.[245]

[238] Grover, Rajwade & Katira, *The Ministry and the Trace: Subverting End-to-End Encryption*, 2021, https://nujslawreview.org/wp-content/uploads/2021/07/14-2R-The-Ministry-and-the-Trace-Subverting-End-to-End-Encryption.pdf.

[239] Indian Express, *WhatsApp used for its encryption, will have to break it to implement IT rules, parent company tells HC*,2024,https://indianexpress.com/article/cities/delhi/people-use-whatsapp-because-of-encryption-will-have-to-break-it-to-implement-it-rules-messaging-app-counsel-tells-delhi-high-court-9291237/.

[240] Joyeeta Roy, *"If We are Told to Break Encryption, Then WhatsApp Goes" | WhatsApp Challenges Indian Govt's Rules in Delhi HC*, 2024, https://lawchakra.in/high-court/whatsapp-challenges-indian-govts-rules/.

[241] PIB Press Release, *The Government Respects the Right of Privacy and Has No Intention to Violate it When WhatsApp is Required to Disclose the Origin of a Particular Message,* 2021, https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=1721915.

[242] Grover, et al., *The Ministry and the Trace: Subverting End-to-end Encryption,* 2021, https://nujslawreview.org/wp-content/uploads/2021/07/14.2-Grover-Rajwade-Katira-2.pdf.

[245] SFLC.in, *S. 69 of the Information Technology Act and the Decryption Rules : Absence of adequate procedural safeguards*, 2021, https://sflc.in/s-69-information-technology-act-and-decryption-rules-absence-adequate-procedural-safeguards/

Provision	Key Requirements	Concerns
	• Section 69B enables governments (with the support of intermediaries) to monitor and collect "traffic data"[243] which is more commonly understood as metadata[244]. • Intermediaries must provide technical assistance. • Non-compliance by intermediaries may lead to 7 years of imprisonment and a fine.	• Limited transparency obligations means insufficient public accountability through recorded reasons and justifications. • The government lacks appropriate technical and infrastructural capacity to decrypt and analyze the intercepted data, especially at the state and district levels. • Any attempt by the Government to force intermediaries to comply with decryption orders might require breaking or bypassing an end-to-end encrypted platform's encryption layer, which is not technically feasible without compromising the encryption protocol of the entire platform. **Doing so could weaken security for all users, raising serious privacy concerns.**[246]
Section 67C, IT Act	• Intermediaries must preserve specified information for the duration and manner prescribed by the Central Government. • The Supreme Court mandates secure retention until the conclusion of the trial.[247]	• Ambiguity in procedural guidelines for duration and manner of storage and retention.
Section 94, BNSS[248]	• This is the erstwhile Section 91 of the CrPC, that is frequently used by law enforcement agencies and courts to seek information from internet services for investigation purposes.	• Provision often enables excessive information requests by the police. The vague wording of the provision, specifically enabling information requests *"necessary or desirable for the purposes of any investigation."*

[243] Section 69B, Explanation ii of the Information Technology Act, 2000 defines traffic data as "any data identifying or purporting to identify any person, computer system or computer network or location to or from which the communication is or may be transmitted and includes communications origin, destination, route, time, data, size, duration or type of underlying service and any other information."

[244] Jenn Riley, *Understanding Metadata: What is Metadata, and what is it for?*, 2017, https://guides.lib.utexas.edu/ld.php?content_id=29091241

[246] India Today, *WhatsApp says it will exit India if asked to break encryption,* 2024, https://www.indiatoday.in/technology/news/story/whatsapp-says-it-will-exit-india-if-asked-to-break-encryption-story-in-5-points-2532027-2024-04-26.

[247] Murali Krishnan, *Supreme Court directs telcos to maintain data seized in criminal cases*, 2020, https://www.hindustantimes.com/india-news/sc-directs-telcos-to-maintain-data-seized-in-criminal-cases/story-DwpVH4rzscsxgAfAhuNN6J.html

[248] Bharatiya Nagarik Suraksha Sanhita (BNSS)

Provision	Key Requirements	Concerns
	• The Court or Police can request information from intermediaries or users during investigations. • Allows issuing of orders to produce documents, information or communication devices.	creates a vague standard of issuing information requests.[249] In the absence of narrowly construed standards of what may be desirable or necessary for investigation, there have been instances where police have accessed excessive information leading to privacy breaches[250], or the police have accessed information without any ongoing investigation to justify such a broad information request.[251]

A 2022 report by Meta disclosed that India had made the second highest number of disclosure requests in the world, a position it has consistently held in recent years.[252] Similarly, Google's transparency report reveals that it addressed 49,313 disclosure requests in India between January and June 2024, making it the second-highest volume globally.[253] Such volumes of requests are mired with several areas of concern. Companies have pointed out that LEA requests often miss essential details such as the case background, the relevance of the information sought, and the roles of the individuals involved.[254] This lack of clarity makes it difficult for intermediaries to respond effectively. Intermediaries have also revealed that while they receive informal preservation requests from law enforcement, these are not acted upon.[255] **The absence of key information or the propensity of LEAs making informal data requests significantly hampers the intermediaries' ability to assess legality, urgency, and relevance of providing such data.** This not only delays legitimate investigations but also risks the rejection of otherwise valid requests due to procedural insufficiencies or non-compliance with due process. Moreover, informal or undocumented preservation requests fall outside formal accountability frameworks, and can raise concerns about people's right to privacy, data retention without cause, and reduce accountability of law enforcement officials.

The confusion in such interactions underscores a broader issue: the **absence of a standardised framework that governs how law enforcement agencies interact with social media platforms**. **In addition, the type of data that can be requested—such as**

[249] Krishnesh Bapat, *#Privacyofthepeople: 91 problems but this ain't one*, 2022, https://internetfreedom.in/privacyofthepeople-91-problems-but-this-aint-one/.

[250] In 2022, as part of an ongoing investigation of the Alt News, Delhi Police requested the payments platform, Razorpay to furnish information related to funding of Alt News. This information gave away details of all contributors who had donated for journalism, Abhinav Sekhri, *On Section 91 Notices and the Razorpay Furore*, 2022, https://theproofofguilt.blogspot.com/2022/07/on-section-91-notices-and-razorpay.html?ref=static.internetfreedom.in.

[251] Krishnesh Bapat, *#Privacyofthepeople: 91 problems but this ain't one*, 2022, https://internetfreedom.in/privacyofthepeople-91-problems-but-this-aint-one/.

[252] Scroll, *India made second-highest requests for user data from Meta, shows transparency report,* 2022, https://scroll.in/latest/1038147/india-made-second-highest-requests-for-user-data-from-meta-shows-transparency-report.

[253] Google, *Global requests for user information*, 2024, https://transparencyreport.google.com/user-data/overview?hl=en&user_requests_report_period=series:requests,accounts,compliance;authority:IN;time:2024H1;get_range:true&lu=user_requests_report_period.

[254] Bismah Malik, *India sought details of 62,754 user accounts from Facebook in second half of 2020*, 2021, https://www.newindianexpress.com/business/2021/May/20/india-sought-details-of-62754-user-accountsfrom-facebook-insecond-half-of-2020-2305243.html.

[255] Sarkar, et al., *Through the Looking Glass: Analysing Transparency Reports,* 2019, https://cis-india.org/internet-governance/files/A%20collation%20and%20analysis%20of%20government%20requests%20for%20user%20data%20%20and%20content%20removal%20from%20non-Indian%20intermediaries%20.pdf; Press Trust of India, *Received 40,300 govt requests for user data from India: Facebook report*, 2021, https://www.hindustantimes.com/india-news/received-40-300-govt-requests-for-user-data-from-india-facebook-report-101621505028130.html; Ananya Bhardwaj, *Apple refuses ED's 'informal' request to unlock Khejriwal's iPhone,* 2024, https://theprint.in/india/apple-refuses-eds-informal-request-to-unlock-kejriwals-iphone/2025098/

accounts interacted with, login-logout information, conversations, friends, and other related data—is not clearly provided for in any existing policy framework. This absence causes inefficiencies, delays, and potential barriers to effective investigation, highlighting the need for a more structured approach to handling digital evidence and requests. These challenges are even more pronounced in cases involving women and children, where delays in obtaining digital evidence can significantly impact access to justice.[256] Indian LEAs have conversely expressed concern about non-cooperation of intermediaries who impose their own standards for data disclosures[257], which delays investigation processes.

Remedying these challenges requires rules that **establish clear and predictable Standard Operating Procedures (SOPs) for law enforcement to seek information from intermediaries**. This was observed by the Delhi High Court where an information request made by Delhi Police to Meta (Instagram) to obtain details on a missing child allegedly led to delayed responses.[258] **This prompted the Court to examine the broader systemic issues in how LEAs obtain critical data relating to investigations from social media intermediaries (SMIs).** In response, a Status Report was filed by the Delhi Police and Ministry of Home Affairs (MHA) which outlines the following roadblocks LEAs face in getting information from platforms in a timely manner:

Table 3.4: Problems and Challenges Spelt out by MHA in its Status Report[259]

Issue	Description
Requirement of FIR for Data Requests	Intermediaries insist on an FIR before providing data, which hampers preventive actions requiring immediate access.
Challenges with VPNs and Proxy Servers	Intermediaries require data requests involving VPNs or proxies to be processed through MLAT channels, delaying LEA actions[260].
Delayed Response from IT Intermediaries	Responses to data requests are delayed by 15 days to a month, including emergency requests.
Complex Portal Procedures	Distinct portals for each intermediary create complexity, requiring separate logins and lengthy procedures for LEAs.
Lack of Designated Nodal Officers	In certain cases the absence of local resident grievance officers and non-responsive foreign-based officers complicate communication.
Incomplete IP Log Disclosure	Intermediaries provide only the last logged-in IP instead of full IP logs, obstructing thorough investigations.

As a solution, the Government has submitted with the Delhi Court that it is developing the Sahyog portal[261] (referred above), spearheaded by the I4C under the MHA, to centralize and streamline information requests between LEAs and intermediaries. Although some platforms have raised concerns about the procedural feasibility of the portal,[262] it might become an important communication channel for intermediaries in the near future in cases involving tech-faciliated violence against women and children.

---

[256] IT For Change, *Access to Justice Issues in Cases of Online Gender-Based Violence,* https://projects.itforchange.net/online-violence-gender-and-law-guide/module-3-access-to-justice/.

[257] Yashashvi Yadav, *Anachronistic cyber legislation related to cyber crimes needs upgradation,* 2023, https://timesofindia.indiatimes.com/blogs/voices/anachronistic-cyber-legislation-related-to-cyber-crimes-needs-upgradation/

[258] Shabana v Govt of NCT and Others, WP. CRL 1563/2024; Sakshi Sadashiv K, *Delhi High Court Tasks Police with Creating Handbook for Emergency Data Requests from Social Media Platforms,* 2024, https://www.medianama.com/2024/11/223-delhi-hc-directs-delhi-police-create-emergency-data-requests-handbook/.

[259] Shabana v Govt of NCT and Others, WP. CRL 1563/2024, Order dated December 11, 2024, https://www.legitquest.com/case/shabana-v-govt-of-nct-of-delhi-and-ors/7B3A62

[260] Bedavyasa Mohanty and Madhulika Srikumar, *Hitting Refresh*: *Making India-US data sharing work,* 2017, https://www.orfonline.org/english/research/hitting-refresh-india-us-data-sharing-mlat. Rethinking Data, Geography and Jurisdiction, *The Challenge of Extraterritorial Data,* 2018, https://www.jstor.org/stable/resrep17624.5?seq=1.

[261] I4C, Ministry of Home Affairs, *Sahyog Portal*, https://sahyog.mha.gov.in/.

[262] The Hindu, *Back door censor: On the government's Sahyog portal*, 2025, https://www.thehindu.com/opinion/editorial/back-door-censor-on-the-governments-sahyog-portal/article69382221.ece.

In parallel, in November 2024, the Court directed Delhi Police to prepare a detailed handbook for investigation officers (IOs), recognising that many of them were unaware of the proper processes for approaching each platform. The handbook, developed with inputs from intermediaries and approved by the Commissioner of Police, was completed in January 2025 and submitted to the Court[263] as a tool to standardize and expedite platform coordination.

Box 3.2: Delhi Police - Handbook for Investigating Officers

Procedure to Handle Data Disclosure Requests to Social Media Intermediaries

In pursuance of the directive passed under *Shabana v. Govt. of NCT Delhi and Ors.*, Delhi Police prepared a handbook delineating the procedure on coordinating with Social Media Intermediaries (SMIs).[264] The objective of this handbook is to serve as a comprehensive guide for **Investigating Officers (IOs)** while obtaining data disclosures from SMIs, as per the procedure laid under the IT Act, IT Rules and the BNSS.

The handbook details platform-specific procedures for data disclosure, model templates, escalation matrices, and timelines. It provides how IOs can engage with platforms like Meta, X (Twitter), WhatsApp, Google, Reddit, LinkedIn, and Telegram, **offering step-by-step instructions for both preservation and records requests.**

The Handbook offers structured templates, legal references and escalation protocols, ensuring that IOs raise data requests methodically and systematically. This is particularly critical for addressing TFGBV and CSEAM, which are time-sensitive and often reliant on intermediary-held data for investigation.

Knowing exactly what steps to follow - what legal provisions to invoke, which online portals to use, what forms to submit, and what data to request - reduces procedural errors that previously led to delays, rejections, or incomplete responses from platforms. Such handbooks can familiarise IOs with the nuances of each platform's data storage and retention policies (including their technical capabilities on storing metadata, IP logs, account information, etc.) and hence data access requests can be tailored accordingly. Moreover, predictable and standardized requests raised through this handbook can help standardise all platforms' internal data governance practices for the Indian market through dedicated data retrieval workflows and teams.

However, this Handbook presents several limitations - each chapter outlines a unique data disclosure workflow tailored to the internal processes of each individual platform. **This lack of a unified, standardized data disclosure framework imposes a disproportionate burden on IOs, who are now expected to be familiar with a wide array of portal interfaces, registration requirements, forms, legal citations, and case categorization systems, which demands extensive training and technical proficiency.**

Importantly, by aligning its procedures with pre-existing internal practices and workflows of SMIs, the Handbook reflects a broader challenge: the absence of a standardised framework for data sharing in criminal investigations. **When platforms follow differing procedures for law enforcement access, it becomes difficult for law enforcement to aggregate data, identify systemic patterns of harm, or evaluate compliance.** For example, Telegram only provides an email address through which IOs can send data requests, without providing a dedicated portal, structured intake system, publicly disclosed turnaround time or pathways for tracking, monitoring or escalating requests. Furthermore, Telegram stipulates that requests must be sent specifically from the official email ID of the designated SHO/ACP of the concerned police unit, which introduces additional procedural requirements not uniformly applicable across platforms.[265] **This highlights the need for clearer, centralised guidance to streamline law enforcement interactions with platforms.**

Despite the Handbook, its deference to platform-specific procedures should be replaced with a uniform, legally binding standard for all platforms operating in India determined by the State police department/MHA.

---

[263] Shabana v Govt of NCT and Others, WP. CRL 1563/2024

[264] Delhi Police, *Handbook For Investigating Officers*, 2025, https://delhipolice.gov.in/CircularsFiles/20717_SOP%20for%20SMI.pdf.

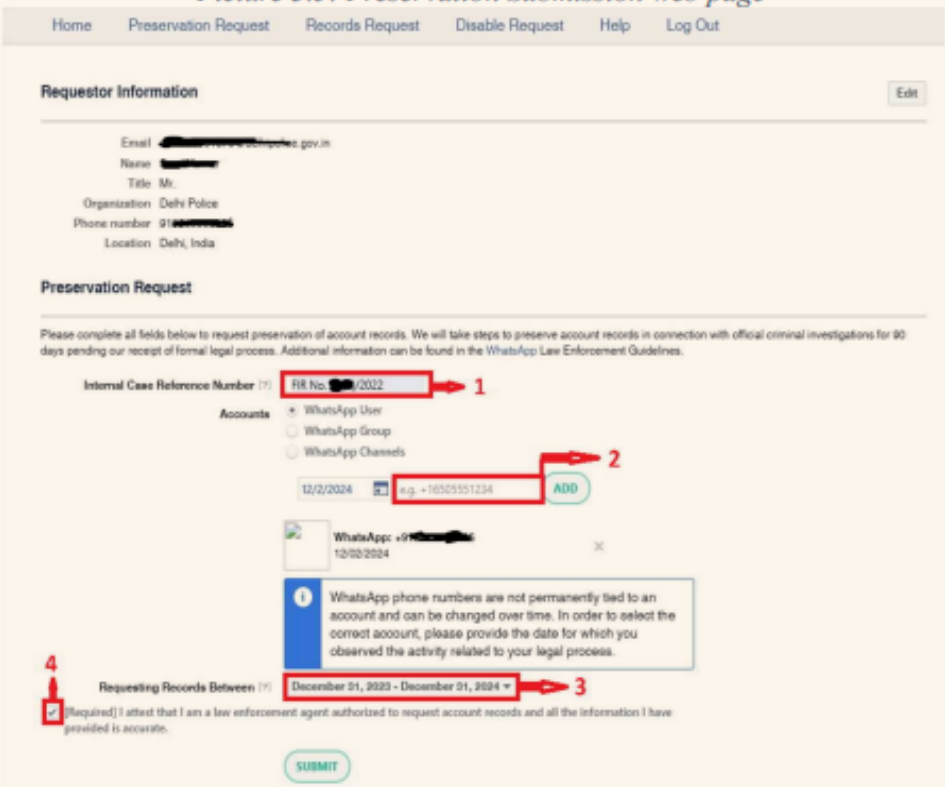
[265] Delhi Police, Handbook For Investigating Officers, 2025, https://delhipolice.gov.in/CircularsFiles/20717_SOP%20for%20SMI.pdf.

Illustration of Box 3.2

To Make a Preservation Request

When IO Clicks on "Make a Preservation Request" on the request dashboard on the Requests Web Page (shown in Picture 3.4 above), a new Web Page will open, as shown in Picture 3.5 below

Picture 3.5: Preservation Submission web page



The IO needs to submit the following details to submit a preservation request:

Field 1: **Internal Case Reference No.** refers to Complaint No. / FIR No. or any other reference number.

Field 2: **WhatsApp number** by prefixing the country code, i.e. +91, for India mobile number and the date on which this WhatsApp number was indulged in criminal activity. The IO may submit requests for multiple

While access to metadata is governed by the legal framework outlined in Table 3.2, recent analyses have highlighted concerns regarding proportionality and consistency with the right to privacy. Several judicial pronouncements have underscored that procedural safeguards, including limiting permissions to what is necessary for achieving the stated objective, serve as critical oversight mechanisms. The principle of proportionality must therefore guide the collection and use of metadata.[266] This indicates that although metadata is commonly used in cyber crime investigations by LEAs, it also has the potential for misuse by the police without clear guidelines and oversight. Without appropriate intervention this can undermine trust in investigation processes.

The lack of standardized procedures for law enforcement information request orders often results in overbroad or vague demands, creating friction between platforms and law enforcement agencies. Such orders may conflict with platforms' obligations under the UN Guiding Principles on Business and Human Rights (UNGPs)[267], particularly their international responsibility to respect user privacy and freedom of expression.[268] Addressing these deficits through capacity-building workshops and targeted training for law enforcement personnel could help establish clear guidelines, improve mutual understanding, and ensure that requests align with both legal and human rights standards. Explicitly incorporating such initiatives into the framework could reduce conflicts and enhance compliance.

3.3. Intermediary Obligations to Remove Harmful Content and Prevent Online Harms Under Indian Laws

India's digital governance framework obligates online intermediaries (such as ISPs, websites hosting third-party content, search engines and social media platforms) to take certain measures to protect women and children. These are intended to curb the prevalence of such crimes and harmful behaviours, and complement efforts by law enforcement authorities in this direction.[269]

These intermediary obligations are legitimised through the **Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021** (IT Rules), which were notified under the IT Act, 2000 by the Government of India.[270] The IT Rules 2021 were ostensibly shaped by a number of factors including a 2018 order from the Supreme Court of India directing the Government to act against the prevalence of child pornography, and rape and gangrape related content on online sites, among other reasons.[271]

The IT Rules 2021 detail due diligence obligations for internet intermediaries that are pre-conditions to them benefiting from safe harbour liability exemptions when third-parties upload unlawful user generated content.[272] Failure to comply with these due diligence obligations would expose the intermediary to penalties, including being asked to shut down.[273]

---

[266] Sarasvati NT, *How India's Police Is Using Metadata*, 2023, https://www.medianama.com/2023/11/223-india-police-metadata-use-tracking-2/.

[267] UN Human Rights Office of the High Commissioner, *Guiding Principles on Business and Human Rights,* 2011, https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

[268] UN Secretary General, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression,* 2017, https://digitallibrary.un.org/record/1304394?ln=en&v=pdf.

[269] Centre for Communication Governance (CCG), *CGG Working Paper: Tackling the Dissemination and Redistribution of NCII*, 2022, https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ncii-working-paper-december-2022-505.pdf.

[270] The IT Rules, 2021 have been challenged in at least 17 petitions in various High Courts in India. In March 2024, the Supreme Court directed that all petitions be clubbed and heard before the Delhi High Court. See here and here.

[271] The Supreme Court passed the order in the Prajjwala case. See here.

[272] Safe harbour refers to a statutory provision which offers immunity from legal liability to intermediaries such as social media platforms, e-commerce platforms, etc. for actions done by users on their platforms. This immunity is typically contingent on adherence to conditions specified in the law, such as taking down certain content or blocking certain accounts on receiving notice from a government agency or based on a court order. In India, safe harbour protection is governed by Sec 69A and Sec 79 of the IT Act, 2000, which provide parallel sets of obligations for intermediaries. Different sets of government agencies are empowered to issue notices to intermediaries under the two provisions, creating an overlap and a legal grey zone. For more information, see https://ccgnludelhi.wordpress.com/2023/01/18/report-on-intermediary-liability-in-india.

[273] See Rule 7, IT Rules 2021, read with Sec. 79, IT Act 2000.

The penalties could even lead to imprisonment for executives employed by the intermediary.[274]

Obligations range from notifying users of platform policies, setting up grievance redressal mechanisms, taking down content within specific timelines, and proactive monitoring (on a best effort basis) to detect rape and child sexual exploitative and abuse material. Most obligations pertain to *all* intermediaries, irrespective of size, sector or function.[275] Prominent among these are obligations relating to *due diligence*, under Rule 3 of the IT Rules 2021. Intermediaries must inform users that sharing prohibited content – a list of which is provided in the IT Rules 2021 – will be liable for penalties under various Indian laws such as the IT Act and other criminal laws in case of violations.[276]

Intermediaries which enable online interactions between two or more users, with more than 5 million registered users in India are known as '**significant social media intermediaries' (SSMIs)** under the Rules,[277] and have additional obligations. SSMIs are expected to upload monthly reports with details of complaints received from users and action taken, with information on content that was taken down for violating the Rules. Typically, SSMIs in India provide information on the nature of objectionable content, such as if it relates to child sexual abuse, hate speech, bullying, etc. in their monthly report.[278] ***It is notable that a lack of standardisation in how SSMIs present these monthly reports inhibits meaningful comparison of this information across intermediaries***.[279] As per a recent report, it was found that the SSMIs adopt different formats and levels of granularity, use inconsistent terminologies, and vary in how they structure and disclose data.[280] Some disclose detailed information on grievance types, automated tool usage and language-specific takedowns, while others only provide high-level counts of complaints and removals.[281] These divergences limit the utility of the reports for regulators and prevent any meaningful aggregation or benchmarking across SSMIs, limiting comparability.

The *due diligence* obligations also require intermediaries to '*make reasonable efforts* to cause the user not to host, display, upload, modify, publish, transmit, store, update, or share' any prohibited content. The prohibited content list includes content which is obscene, invasive of bodily privacy, harassing on basis of gender, or harmful to a child.

The tension between this obligation, which expects intermediaries to interfere in user behaviour, and the idea that by definition intermediaries merely facilitate content as third-parties, has been noted by commentators.[282] Since the notification of the IT Rules, 2021, the Government

---

[274] Aditi Agarwal, *What does it mean to lose safe harbour?*, 2021, https://www.forbesindia.com/article/take-one-big-story-of-the-day/what-does-it-mean-to-lose-safe-harbour/68573/1.

[275] Specific obligations on larger intermediaries are discussed in the next sub-section.

[276] See Rule 3(1)(b) of IT Rules, 2021. The list of prohibited content includes: (i) belongs to another person and to which the user does not have any right; (ii) is obscene, pornographic, paedophilic, invasive of another's privacy including bodily privacy, insulting or harassing on the basis of gender, racially or ethnically objectionable, relating or encouraging money laundering or gambling, or promoting enmity between different groups on the grounds of religion or caste with the intent to incite violence; (iii) is harmful to child; (iv) infringes any patent, trademark, copyright or other proprietary rights; (v) deceives or misleads the addressee about the origin of the message or knowingly and intentionally communicates any misinformation or information which is patently false and untrue or misleading in nature; (vi) impersonates another person; (vii) threatens the unity, integrity, defence, security or sovereignty of India, friendly relations with foreign States, or public order, or causes incitement to the commission of any cognisable offence, or prevents investigation of any offence, or is insulting other nation; (viii) contains software virus or any other computer code, file or program designed to interrupt, destroy or limit the functionality of any computer resource; (ix) violates any law for the time being in force.

[277] Gazette of India, *Notification S.O. 942(E),* 2021, https://www.meity.gov.in/static/uploads/2024/05/Gazette-Significant-social-media-threshold.pdf

[278] For instance, see here and here.

[279] Maheshwari, et al., *Social Media Transparency Reporting: A Performance Review*, 2024, https://igap.in/wp-content/uploads/2024/10/IGAP-Social-Media-Transparency-Reporting-A-Performance-Review.pdf.

[280] Maheshwari, et al., *Social Media Transparency Reporting: A Performance Review*, 2024, https://igap.in/wp-content/uploads/2024/10/IGAP-Social-Media-Transparency-Reporting-A-Performance-Review.pdf.

[281] Maheshwari, et al., *Social Media Transparency Reporting: A Performance Review*, 2024, https://igap.in/wp-content/uploads/2024/10/IGAP-Social-Media-Transparency-Reporting-A-Performance-Review.pdf.

[282] National Law University Delhi, *Submission of Comments on the proposed draft for amendment in Part-I and Part-II of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021,* 2022, https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccgnlud-comments-draftamendments-itrules2021-6jul22-301.pdf.

of India has shared advisories to intermediaries exhorting them to perform their ***reasonable efforts-based due diligence* obligations** on more than one occasion. A recent example of the same came in the weeks after a deepfake video of actor Rashmika Mandana was circulated online in November 2023.[283]

Box 3.3: India's Due Diligence Framework Promotes Reactionary Instead of Proactive Interventions

At this stage Indian law has a paradigm where the reasonable efforts requirements for an intermediary's due diligence expectations are interpreted by the Indian Government to push for informal directives for intermediaries to address certain online safety incidents over others. **This is indicative of a broader ecosystem outcome where platforms are not offered a certain pathway towards lawful online trust and safety practices. Moreover, the framework lends itself to a scenario where policy enforcement practices veer towards reactionary responses to publicly prominent incidents. This is representative of a wider gap in platform governance where problems are not identified and addressed proactively.**

Under the IT Rules, 2021, intermediaries have also set up **grievance redressal mechanisms** to enable users to raise issues. The Rules specify that when users flag content which is of sexually explicit nature, which shows private parts of individuals, partial or full nudity, or individuals engaged in sexual acts, such content must be taken down within 24 hours of receiving a complaint from an affected individual.[284] This includes content that is in the nature of impersonation, such as through morphed images or videos,[285] and extends to sexually explicit deepfake videos.[286] Complaints regarding other content which is in the list of prohibited content and pertaining to women and children safety must be similarly resolved within 72 hours.[287] **The Rules also require intermediaries to take down content notified by the Government order or by a court order as illegal within 36 hours of such notification.**

They mandate that SSMIs ought to '*endeavour to deploy technology-based measures, including automated tools or other mechanism to proactively identify*' content depicting rape or child sexual abuse. The Rules go on to classify this obligation – that they must be *proportionate* to interests of free speech and privacy of the users, and that they should include human oversight for periodic review of these tools. **The obligation on use of automated content filtering tools has led to concerns from civil society and from platforms on their accuracy and concerns that automated content moderation requirements might lead to excessive impingement upon legitimate speech.[288]** Commentators argue that such tools reduce transparency on how content moderation decisions are being taken, preventing individual users from meaningfully exercising their freedom of speech and right to privacy in a legally permissible manner.[289]

Specifically, automated systems have been noted to inaccurately flag and take down content, and are usually unable to identify irony, satire, and critical analysis needed to differentiate between legal and illegal content.[290] Even in the context of illegal content, the safeguards in the IT Rules, around proportionality and human oversight, lack appropriate detailing. For instance, it is unclear when

[283] The Indian Express, *Govt directs social media platforms to comply with IT rules amid concerns over deepfakes,* 2023, https://indianexpress.com/article/india/govt-social-media-intermediaries-it-rules-concerns-deepfakes-9083707/

[284] See Rule 3(2)(b) of IT Rules, 2021.

[285] PIB Press Release, *Government notifies Information Technology (IntermediaryGuidelines and Digital Media Ethics Code) Rules 2021,* 2021, https://www.pib.gov.in/PressReleasePage.aspx?PRID=1700749.

[286] The Indian Express, *Govt directs social media platforms to comply with IT rules amid concerns over deepfakes,* 2023, https://indianexpress.com/article/india/govt-social-media-intermediaries-it-rules-concerns-deepfakes-9083707/.

[287] See Rule 3(2)(a)(i) of IT Rules, 2021.

[288] Krishnesh Bapat, *Deep dive: How the intermediaries' rules are anti-democratic and unconstitutional*, 2021, https://internetfreedom.in/intermediaries-rules-2021.

[289] Aarathi Ganesan, *How Will the Proposed Amendments to the IT Rules Affect Free Speech and Intermediaries?*, 2022, https://www.medianama.com/2022/06/223-it-rules-amendments-india-free-speech-big-tech/

[290] Vasudev Devadasan, *Intermediary Guidelines and the Digital Public Sphere: Automated Filtering*, 2021, https://indconlawphil.wordpress.com/2021/04/05/intermediary-guidelines-and-the-digital-public-sphere-automated-filtering/

exactly the automated tools are expected to kick in, how they are expected to take down only illegal content in a targeted manner and not take down other content as collateral damage, what is a tolerable error rate for such automated review, how frequently should human oversight be engaged, and what form of data processing is permissible for this purpose considering impacts on user privacy.[291] **The obligation thus presents a complex problem involving trade-offs between individual rights and online safety**.

Table 3.5: Government-Led Content Blocking Regime for Unlawful Content | Story of Parallel Pathways

Section 69A and Access Blocking Rules, 2009	Section 79(3)(b): Exemption from liability of intermediary in certain cases
• The **Central Government** (or its authorized officers) can issue written orders to online intermediaries (such as social media platforms or internet service providers) to **block public access to unlawful content/ information** typically on grounds relating to national security, public order or to prevent commission/incitement of offences. • Normally, blocking requests originate from **Nodal Officers** within different government agencies, who then forward them to a **Designated Officer within the MeitY**. These requests are usually reviewed by a **Committee**. • Under this framework, the concerned intermediary has an **opportunity to be heard** by this Committee. The Committee's recommendation is then submitted to the **Secretary of MeitY** for final approval. • In **emergency situations**, the MeitY Secretary can issue an interim blocking order without prior committee review, provided it is a reasoned decision. This emergency order must subsequently undergo the standard review process for confirmation. • There have been concerns about the framework's opacity and the system's lack of independence. For example, blocking orders are not publicly disclosed and the review committee is composed entirely of **executive members**, without judicial representation. • It is important to note that **non-compliance** with a blocking order can lead to penal consequences for the intermediary.	• **Section 79(3)(b) of the IT Act** outlines circumstances where online intermediaries can be held responsible for unlawful content. According to the framework, an intermediary loses its safe harbour liability exemption if it fails to promptly remove unlawful content after it receives a lawful content takedown order from a judicial or government authority. • Various agencies, including the Indian Cybercrime Coordination Centre (I4C)[292], the NCRB[293], the Delhi Police[294], Deputy Secretary (Tobacco Control)[295] among others, have been designated as appropriate legal authorities under this framework. • The legality of this parallel framework, particularly its use as a separate online content blocking mechanism without the specific procedural safeguards found in **Section 69A of the IT Act**, is currently under legal challenge.[296] • **Non-compliance** with a government notification under Section 79(3) can result in the intermediary losing its safe harbor protection, making it legally liable for the content.

[291] Vasudev Devadasan, *Intermediary Guidelines and the Digital Public Sphere: Automated Filtering*, 2021, https://indconlawphil.wordpress.com/2021/04/05/intermediary-guidelines-and-the-digital-public-sphere-automated-filtering/

[292] Deccan Herald, *Centre designates I4C as agency of MHA to notify unlawful activities in cyber world*, 2024, https://www.deccanherald.com/india/centre-designates-i4c-as-agency-of-mha-to-notify-unlawful-activities-in-cyber-world-2936976

[293] Government of India, *Parliamentary question in the Lok Sabha to the Minister of Home Affairs*, 2021, https://sansad.in/getFile/loksabhaquestions/annex/176/AU3462.pdf?source=pqals

[294] Rahul Sundaram, *Delhi Police designated as Nodal Agency under Section 79(3)(b) of the Information Technology Act, 2000*, 2025, https://www.indialaw.in/blog/civil/delhi-police-nodal-agency-793b-it-act-2000/

[295] Ministry of Electronics & Information Technology, *Direction to Indian Council of Medical Research regarding Nomination of Nodal Officer for handling unlawful content/information/activities in Cyberspace as per the provisions of the IT Act, 2000*, 2022, https://www.icmr.gov.in/icmrobject/custom_data/1703666738_office_memorandum21022022.pdf

[296] Ajoy Sinha Karpuram, *IT Act and content blocking: Why X has challenged govt's use of Section 79*, 2025, https://indianexpress.com/article/explained/explained-law/it-act-content-blocking-why-x-has-challenged-govts-use-of-section-79-9899231/

Limited Instances where India's Content Blocking Framework Used for Incidents Relating to Women and Children:

Sections 69A and 79(3)(b) have been used in select instances to address content impacting women and children. For example, directions under Section 69A were issued to restrict public access to the Blue Whale online gaming challenge, which reportedly targeted vulnerable children and encouraged self-harm.[297] In July 2023, the government directed platforms including Twitter to take down a video showing two Manipuri women being paraded naked.[298] Even in a Union Government order from 2015, the Department of Telecommunications expressly cited Section 79(3)(b) when directing ISPs to disable access to 857 websites allegedly hosting pornographic content as they "relate to morality, decency as given in Article 19(2) of the Constitution of India".[299]

These examples illustrate that while Sections 69A and 79(3)(b) have occasionally been used to address harms against children and women, such interventions have occurred under broader rationales like public order or morality. **While these actions address serious and often highly publicised incidents, they reflect a pattern of reactive enforcement. There is little evidence that these provisions are being used to systematically or proactively address everyday harms faced by women and children online.**

To implement the mandates imposed by the IT Rules 2021, other intermediaries have evolved automated detection tools, content standards and community guidelines to proactively remove material that harms or comprises the safety of women and children. These include Artificial Intelligence (AI) and Machine Learning (ML) tools that rely on fingerprinting and hash-matching methods.[300] For instance, Google deploys such tools to prevent the dissemination of 'harmful content such as child sexual exploitative and abuse material and violent extremist content'.[301] Upon detection, the intermediary considers whether to remove the content for violating intermediary policies, restrict it according to age criteria, or leave it live when there is no violation. In just June 2024, Google reportedly took 7.72 lakh removal actions using automated detection.[302] Similarly, Instagram deployed a ML tool in 2018 to detect instances of bullying and abusive language in pictures and captions – flagging such content to be reviewed by an Instagram employee, and taking down the post if necessary.[303] At the same time, it is important to contend with the technical limitations associated with such automated content identification and filtering solutions.

---

[297] *Sneha Kalita v. Union of India, (2018) 12 SCC 674.*

[298] Business Standard, *Centre asks Twitter to take down Manipur video under Section 69A*, 2025, https://www.business-standard.com/india-news/centre-asks-twitter-to-take-down-manipur-video-under-section-69a-123072100415_1.html?utm_

[299] Software Freedom Law Centre, *DOT orders blockage of porn websites*, 2015, https://sflc.in/dot-orders-blockage-porn-websites/.

[300] Center for Internet and Society, *On the legality and constitutionality of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021,* 2021, https://www.medianama.com/2021/06/223-legality-constitutionality-of-it-rules/.

[301] Susan Jasper, *How we detect, remove and report child sexual abuse material*, 2022, https://blog.google/technology/safety-security/how-we-detect-remove-and-report-child-sexual-abuse-material/.

[302] Google, *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021: Monthly Transparency Report,* 2024, https://storage.googleapis.com/transparencyreport/report-downloads/india-intermediary-guidelines_2024-6-1_2024-6-30_en_v1.pdf.

[303] Josh Constine, *Instagram now uses machine learning to detect bullying within photos,* 2018, https://techcrunch.com/2018/10/09/instagram-bullying-photos/

Box 3.4: Case Study | Apple's technical intervention to detect CSEAM

Apple's attempt to address automated content monitoring through technical intervention illustrates the pitfalls of such approaches. In 2021, Apple announced several ML-powered features aimed at detecting CSEAM which would scan photos before they were uploaded to iCloud and compare them against known CSEAM hashes provided by NCMEC.[304] **The initiative drew swift backlash from civil liberties groups, cryptographers, and privacy experts, who argued that introducing such scanning mechanisms on user devices risked setting a dangerous precedent for invasive surveillance and could be repurposed by governments for broader content monitoring.[305]** Following this, Apple recalled these features. This incident brought out that specific technical tools that automate content moderation can be difficult to implement responsibly, especially if prescribed through regulation.

Separately, it is important to better understand the complexities of platform obligations in high-profile, time sensitive cases. Box 3.5 shows how a lack of defined response protocols makes compliance complicated as a result of ad-hoc requests from government agencies.

Box 3.5: Case Study| RG Kar Medical College Incident and Intermediary Obligations

The RG Kar Medical College incident highlights significant challenges in managing the dissemination of sensitive content involving a rape and murder victim (including but not limited to their identity) on social media. Despite directives from the Supreme Court to the MeitY[306] to remove references to the victim's name, images, and videos, **such content often continues to circulate online.[307]** This case underscores critical gaps in intermediary obligations under India's IT laws, particularly in high-profile cases.

Social media platforms were unable to curb the proliferation of such content, partly because the Supreme Court's requirement and the subsequent MeitY directive were not rooted with the realities of the internet. The information about the victim was shared across thousands of accounts, often reappearing with slight alterations to evade detection. The platforms' reliance on existing automated tools, machine learning classifiers and manual review processes struggled to deal with the sheer volume and virality of the posts. It was another instance where a sensitive flashpoint led to direct pressures on platforms to pursue absolute perfection when proactively removing illegal content.

This incident underscores the need to adopt layered multi-faceted approaches to detect, moderate and regulate the rapid virality of harmful content. It also highlights a regulatory gap: the absence of standardised protocols for intermediaries to follow in handling sensitive cases involving gender-based violence. Addressing this gap through a defined standard operating procedure, would offer clearer guidance to platforms, improve policy predictability and ensure greater accountability, and consistency in enforcement. Legal, policy and judicial institutions must also propose solutions that are in line with technical realities and best practices, and thus there is a need for open channels of consultation and communication regarding the possibilities of tech-enabled incident detection as well as alternative interventions that work around the limitations of such systems.

---

[304] Apple, *CSAM Detection Technical Summary,* 2021, https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf

[305] Zack Whittakar, *Apple's CSAM detection tech is under fire - again,* 2021, https://techcrunch.com/2021/08/18/apples-csam-detection-tech-is-under-fire-again/; Lily Hay Newman, *Apple's Decision to Kill Its CSAM Photo-Scanning Tool Sparks Fresh Controversy,* 2023, https://www.wired.com/story/apple-csam-scanning-heat-initiative-letter/.

[306] PIB Press Release, *Social Media Platforms to comply with Supreme Court order on removal of deceased's identity in RG Kar Medical college incident*, 2024, https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2047347

[307] NDTV, *Supreme Court Upset Over RG Kar Victim's Photos On Social Media,* 2024, https://www.ndtv.com/video/supreme-court-upset-over-rg-kar-victim-s-photos-on-social-media-843915.

3.4. User Challenges in Seeking Remedy and Accessing Justice

As cybercrime continues to rise, the systems designed to protect users and support victims/survivors fall short, leaving individuals vulnerable and disillusioned. These gaps exist across two interconnected dimensions:

- The challenges users face on digital platforms, and
- The systemic shortcomings in law enforcement and judicial responses.

On platforms, inadequate reporting mechanisms, lack of feedback, and accessibility barriers hinder users from addressing harmful content effectively. Meanwhile, victims navigating the legal justice system encounter biases, the burden of evidence collection, and prolonged legal processes, compounding the harm they endure. This section examines these critical gaps, shedding light on how platform-specific limitations and victim-specific policy failures intersect to create a fragmented and inequitable approach to combating TFGBV and other technology-facilitated safety risks that impact women and children.

3.4.1. Gaps in Reporting

1. **Limitations of Content Reporting Options:** One significant limitation lies in the predefined categories for reporting content. Platforms typically provide generic options such as "harassment," "hate speech," or "spam", but these categories often fail to capture the complexity of user experiences.[308] Many incidents do not fit neatly into these categories, forcing users to rely on the vague "other" option.[309] **This lack of specificity hampers platforms' ability to assess and address complaints effectively, leaving many users feeling unheard. Moreover, reporting mechanisms rarely allow users to include additional context, such as whether an incident forms part of a broader pattern of abuse.[310]** There is a need for more flexible reporting systems that enable users to describe their experiences in detail, with dynamic templates to capture recurring issues.

2. **Lack of Feedback:** A second challenge stems from the absence of feedback on complaint outcomes. **When users report harmful content, they often receive generic acknowledgments such as "your report has been reviewed", without any clear explanation of the decision made or actions taken. This lack of transparency discourages users from engaging with reporting mechanisms and diminishes trust in the platform's user safety protocols.[311]** Providing users with clear updates on the progress of their complaints, the rationale behind decisions, and specific measures taken, such as content removal or warnings go a long way toward restoring confidence in these systems. As a solution, some jurisdictions within the European Union (e.g. Ireland) let users file complaints regarding content moderation on platforms to a newly established independent body.[312] **In India, users can approach the Grievance Appellate Committee under the IT Rules, 2021.[313] However, the GAC has received only 2,322 appeals in the period between March 2023 and January 2025[314] – revealing that these appeals are being processed at relatively low volumes.**

---

[308] European Union Agency for Fundamental Rights, *Online Content Moderation Current Challenges in Detecting Hate Speech,* 2023, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2023-online-content-moderation_en.pdf

[309] Banko, et al., *A Unified Taxonomy of Harmful Content,* 2020, https://www.researchgate.net/publication/347232912_A_Unified_Taxonomy_of_Harmful_Content.

[310] Ángel Díaz and Laura Hecht-Felella, *Double Standards in Social Media Content Moderation,* 2021, https://www.brennancenter.org/sites/default/files/2021-08/Double_Standards_Content_Moderation.pdf.

[311] Ku, et al., *Social learning effects of complaint handling on social media: Self-construal as a moderator*, 2021, https://www.sciencedirect.com/science/article/abs/pii/S0969698920313515.

[312] Supantha Mukherjee, *New body to handle disputes between EU users and Facebook, TikTok, YouTube*, 2024, https://www.reuters.com/technology/new-body-handle-disputes-between-eu-users-facebook-tiktok-youtube-2024-10-08/?utm_source=chatgpt.com.

[313] Digital India, *Grievance Appellate Committee,* https://gac.gov.in/.

[314] Press Information Bureau, *Grievance Appellate Committee workshop organised by MeitY to enhance grievance redressal framework for a safer internet*, 2025, https://www.pib.gov.in/PressReleasePage.aspx?PRID=2091074

Thus, there might be a need for regulation to focus on platforms providing clearer feedback and taking more proactive measures in the first instance.

3. **Context Gaps:** Another shortcoming of current reporting mechanisms is their inability to capture the context of incidents. These tools focus predominantly on isolated events, overlooking patterns of abuse or recurring violations. **For example, users may experience repeated harassment from the same individual or coordinated attacks by multiple accounts, yet they lack a mechanism to report these patterns effectively.**[315] Platforms also fail to incorporate contextual data, such as prior interactions or related complaints, into their moderation processes. To address this gap, platforms should develop tools that allow users to report patterns of behavior and link their complaints to existing records, providing moderators with a more comprehensive understanding of the situation. In addition to these limitations, platforms often struggle with incidents of online abuse that are rooted in linguistic peculiarities that emerge within certain communities.[316] Global platforms typically apply a narrow, universalist lens to defining harms and reporting metrics and struggle to keep pace with on-ground cultural dynamics.[317] **While platforms have invested in trusted flagger networks to surmount these challenges, several studies from the European context have noted that without thoughtful design and implementation there can be limitations around scale and accountability.**[318]

4. **Accessibility Issues:** Finally, accessibility issues present a significant barrier for many users. **Reporting tools are often available only in a limited set of languages, making them inaccessible to non-English speakers or those unfamiliar with technical terminology.**[319] Automated systems dominate the complaint-handling process, leaving users with limited recourse when these systems fail to address their concerns. Outside of linguistic inclusion, users can also stand to benefit from human-moderated support channels, and enhanced compatibility with assistive technologies to ensure that all users can report concerns effectively. Moreover, linguistic accessibility is crucial for detecting harms experienced by vulnerable groups such as LGBTQI+ youth, women or caste-marginalised users, where abuse often takes the form of culturally-specific slurs or even emoji-based[320] and text-coded harassment that detection tools frequently overlook.

3.4.2. Gaps in Law Enforcement

As the prevalence of cybercrime grows, so do the systemic gaps in addressing the needs and challenges faced by survivors and victims' families. **Current policies and frameworks often fail to account for the complexities of navigating legal and judicial systems, particularly from the perspective of these individuals.** This section explores critical gaps, ranging from reluctance of law enforcement to take cases forward, other systemic biases, excessive burden of proof, digital divide, and the lengthy processes that victims endure.

---

[315] UNESCO, *The Chilling: assessing big tech's response to online violence against women journalists*, 2022, https://unesdoc.unesco.org/ark:/48223/pf0000383044.

[316] Prithvi Iyer, *Lost in Translation: How Content Moderation Fails Tamil Speakers Online,* 2025, https://www.techpolicy.press/lost-in-translation-how-content-moderation-fails-tamil-speakers-online/.

[317] Social & Media Matters, Is content moderation working in India?, 2024, https://drive.google.com/file/d/13L4eQgR0eqHseKh9SuSVAcKeQNwySvIo/view.

[318] "On trusted flaggers", Yale Journal of Law and Technology, Special White Paper Series https://yjolt.org/sites/default/files/0_-_appelman_leerssen_-_on_trusted_flaggers.pdf

[319] Global Witness, *How Big Tech platforms are neglecting their non-English language users,* 2023, https://www.globalwitness.org/en/campaigns/digital-threats/how-big-tech-platforms-are-neglecting-their-non-english-language-users/

[320] Zhou, et al., *The Hidden Language of Harm: Examining the Role of Emojis in Harmful Online Communication and Content Moderation,* 2025, https://arxiv.org/html/2506.00583v1.

1. **Reluctance of Law Enforcement to Prosecute Cybercrime -** Indian criminal law and LEA officials tend to operate primarily within the physical realm, often searching for tangible signs of injury. This focus persists even though the consequences of cyber violence can be as debilitating as those of offline crimes.[321] **Qualitative research from West Bengal highlighted that in investigations, police in the state frequently advised women who reported online harassment to simply block or ignore the perpetrator. Moreover, instead of filing FIRs, the police often recorded these complaints as general diary entries, which do not require follow-up investigations.[322]** This apathy creates a "*chilling effect*"[323], discouraging victims from reporting incidents and perpetuating a culture of impunity among perpetrators. **Improving law enforcement training, resources, and accountability mechanisms are essential to ensuring cybercrime complaints are treated with the seriousness they deserve. As a result of LEAs perceiving online harms as 'not serious enough', official reporting of online violence suffers. Data from The Economist highlights that despite its prevalence, only 14% of women reported online violence to an offline protective agency.[324]**

   The challenges are further compounded by the limitations of the 1930 national cyber crime helpline. While the helpline was set up to provide swift first-line support to victims of cyber crimes, it lacks personnel trained in responding to gender-based online harms or cyber crimes against children. Our interactions with relevant stakeholders pointed to an urgent need to sensitise helpline responders and embed specialised protocols for handling cases involving women and children, ensuring that the first point of contact does not become a barrier to justice.

2. **Judiciary and Law Enforcement Biases -** Judiciary and law enforcement biases further compound the challenges for victims. Stereotyping, gender biases, and a lack of sensitivity toward victims often undermine the pursuit of justice. For example, women reporting online harassment or stalking may encounter dismissive attitudes or victim-blaming narratives from police or judicial authorities.[325] Such biases are not limited to gender; marginalized groups, including those from certain castes, religions, or socio-economic backgrounds, also face discriminatory treatment.

3. **Burden of Proof on Victims** - Another pressing issue is the burden of proof, which disproportionately falls on victims. **Unlike in traditional crimes, where evidence collection is primarily the responsibility of law enforcement, cybercrime victims are often expected to gather extensive documentation themselves.[326] This includes screenshots, chat logs, email headers, and other digital evidence, which may be challenging for those without technical expertise.** Additionally, perpetrators often exploit digital anonymity/app features to erase traces of their activity, further complicating evidence collection.[327] This burden

[321] TISS and IT for Change, *Report of the National Dialogue on Gender-based Cyber Violence,* 2018, https://projects.itforchange.net/e-vaw/wp-content/uploads/2018/03/Event-Report-of-National-Dialogue-on-Gender-Based-Cyber-Violence.pdf

[322] Anita Gurumurthy and Amrita Vasudevan, *Hidden figures- A look at technology-mediated violence against women in India*, 2018, https://itforchange.net/index.php/hidden-figures-a-look-at-technology-mediated-violence-against-women-india

[323] Leslie Kendrick, *Speech, Intent and the Chilling Effect,* 2012, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2094443.

[324] The Economist, *Measuring the prevalence of online violence against women,* 2021, https://onlineviolencewomen.eiu.com/

[325] College of Policing, *Victim experience of the police response to stalking*, 2024, https://assets.college.police.uk/s3fs-public/2024-09/Victim-experience-of-the-policing-response-to-stalking-REA.pdf.

[326] Babasaheb Bhimrao Ambedkar University, Cyber Crime and its Classification, https://www.bbau.ac.in/dept/Law/TM/1.pdf.

[327] United Nations Office on Drugs and Crime, *Handling of digital evidence*, https://www.unodc.org/e4j/zh/cybercrime/module-6/key-issues/handling-of-digital-evidence.html

is even more acute in the context of children, where caregivers or parents often act as first responders but may lack the knowledge to file complaints. **Parents' and caregivers' unfamiliarity with digital platforms, formal reporting procedures, or the legal thresholds for what constitutes an offence leads to delays or incomplete reports.[328] Establishing victim support units equipped with forensic experts and simplifying evidentiary procedures can alleviate this burden.**

4. **Digital Divide and Accessibility Barriers** - The digital divide exacerbates accessibility barriers for rural and economically disadvantaged victims. Many cybercrime reporting mechanisms and support services are digital-first, which marginalizes those with limited internet access or technological literacy.[329] Rural populations, in particular, face additional hurdles, including the absence of nearby cybercrime cells or legal aid centers. Economic disadvantage also limits victims' ability to seek private legal counsel or afford the technology necessary to document and report incidents.[330]

5. **Lengthy Investigations and Judicial Delays -** Finally, victims of cybercrime endure lengthy investigations and court proceedings that are both emotionally and financially taxing. The complexities of collecting digital evidence, combined with overburdened judicial systems, leads to protracted delays. For victims, this extended timeline compounds the psychological trauma of the initial crime, often leaving them feeling re-victimized by the system itself. The financial strain of legal fees, travel costs, and lost income further discourages many from pursuing justice. **Streamlining cybercrime case handling through specialized fast-track courts, procedural reforms, and victim compensation programs might help reduce these burdens.**

The current policy framework for addressing cybercrime falls short in adequately supporting victims, particularly those from marginalized or vulnerable groups. Policymakers must consider victim/survivor-centric approaches, to ensure that people, regardless of their background or circumstances, can navigate the system with dignity and confidence.

3.5. Observations On Current Framework And Gaps Therein

An analysis of the current legal governance and enforcement frameworks that tackle online harms against women and children, as explained in this chapter, leads to the following observations:

1. **Criminalisation Through General Frameworks:** India's current approach to cybercrime through general (e.g. BNS) and specific laws (e.g. IT Act) primarily seek to empower the State to prosecute perpetrators of online harms. These laws give victims a legal remedy whereby they can approach the police/cyber-crime wing to file a criminal case, to eventually get the courts to impose imprisonment and/or monetary penalties on the perpetrator. To achieve these remedies, the prosecution must present evidence to the satisfaction of the court, as required in criminal cases. **Scholars have observed how this criminal trial process may be insufficient in adequately dealing with online harms, since it exposes the victim to the vagaries of the trial process.[331]** This includes

[328] Manoj, et al., *Behind the screens: Understanding the gaps in India's fight against online child sexual abuse and exploitation*, 2025, https://www.sciencedirect.com/science/article/pii/S2950193824000883#sec3

[329] Yoganandham Govindharaj, *Bridging the Digital Divide: Information Technology's Role in the Rural Economy and Women's Empowerment in Tamil Nadu, 2024,* https://www.researchgate.net/publication/382972745_BRIDGING_THE_DIGITAL_DIVIDE_INFORMATION_TECHNOLOGY'S_ROLE_IN_THE_RURAL_ECONOMY_AND_WOMEN'S_EMPOWERMENT_IN_TAMIL_NADU

[330] United Nations Office on Drugs and Crime, *Global Study on Legal Aid: Global Report*, 2016, https://www.unodc.org/documents/justice-and-prison-reform/LegalAid/Global_Study_on_Legal_Aid_-_FINAL.pdf

[331] Dr Shalu Nigam, *Ending Online Violence Against Women in India: Calling for an Inclusive, Comprehensive, and Gender-Sensitive Law and Policy Framework,* 2024, https://www.impriindia.com/insights/ending-online-violence-against-women/; Cyber Violence- Unpacking Joseph, et al., *Case Histories from Counselling Centres and Cyber Crime Cells*, 2016, https://projects.itforchange.net/e-vaw/wp-content/uploads/2018/01/MeghaBijiJoseph_AnuSwarajV-S_0ANargeesBasheer.pdf.

long trials which delay the delivery of justice, inadequately trained police officials who struggle to deal with technical cyber laws, and gender stereotypes prevalent in courts and legal jurisprudence. Another major issue is the gap in legal terminology which does not address emerging behaviour and harms as crimes (such as networked harassment, doxxing, and gendered hate speech or trolling).[332] This is potentially because Indian courts and police rely extensively on general criminal laws for cyber offences, which presents a wide scope for subjective interpretation. Similarly, there could be issues in gathering and presenting evidence in cyber-crime cases (due to data protection laws, lack of forensic capacity within investigatory institutions, etc.)[333] and also overlapping offences (such as voyeurism and obscenity) where the police may file a chargesheet with all such provisions aiming for the highest penalty, but which make the prosecution's case more complex.[334] **The outcome of these pitfalls is detrimental to the victim, who lacks immediate redressal of the online harm they have suffered**.

2. **Ad-hoc Nature of Redressal:** Various cyber-crimes which the National Cyber Crime Reporting Portal seeks to address do not have corresponding provisions in the Indian law. Yet, the police regularly rely on a combination of other provisions to proceed with the investigation and file a prosecution case. **This approach subjects the victims of cyber-crimes, and online intermediaries, to considerable uncertainty, as the police action may differ from state to state in India (as policing is a state subject under the Constitution).** Likewise, the police themselves may lack directions on providing effective redressal in cases of emerging harms, since the law may not explicitly recognize them.

3. **Ambiguity, Uncertainty and Implementation Complexities in Intermediary Obligations:** The IT Rules, 2021 which imposes wide-ranging intermediary obligations, incentivizes intermediary compliance with threats of losing their safe-harbour protection for failure to do so. Even so, intermediaries must wade through considerable uncertainty regarding the exact scope of their obligations. For instance, while the Rules initially required intermediaries to merely inform users not to upload or share prohibited content, an amendment in 2022 changed the obligation to '*make reasonable efforts to cause the user*' not to upload or share such material. **Commentators have noted ambiguity when it comes to what constitutes 'reasonable efforts'.[335] This phrasing is also at odds with the underlying principle of safe-harbour protection, which is that intermediaries cannot be held liable for actions done by users in which the intermediary had no role to play except as an intermediary.** Through this phrasing, intermediaries are arguably compelled to interfere and influence the actions of their users, and thereby lose out on the protection earlier afforded to them under the IT Act.[336] Similar ambiguity exists in the language around automated content filtering mechanisms and traceability provisions. Such provisions could lead to overreach by intermediaries wishing to remain on the right side of the law, but at

[332] Malavika Rajkumar and Shreeja Sen, *The Judiciary's Tryst with Online Gender Based Violence,* 2023, https://itforchange.net/sites/default/files/2190/The%20Judiciary's%20Tryst%20with%20OGBV_0.pdf.

[333] Pruthvi Ramkanta Hegde, *All About Digital Evidence,* 2024, https://blog.ipleaders.in/all-about-digital-evidence/#Challenges_in_handling_digital_evidence_in_India.

[334] Malavika Rajkumar and Shreeja Sen, *The Judiciary's Tryst with Online Gender Based Violence,* 2023, https://itforchange.net/sites/default/files/2190/The%20Judiciary's%20Tryst%20with%20OGBV_0.pdf.

[335] Centre for Communication Governance, *CCG-NLUD comments to the MeitY's proposed amendments to the 2021 IT Rules*, 2021, https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccgnlud-comments-draftamendments-itrules2021-6jul22-303.pdf.

[336] Centre for Communication Governance, *CCG-NLUD comments to the MeitY's proposed amendments to the 2021 IT Rules*, 2021, https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccgnlud-comments-draftamendments-itrules2021-6jul22-303.pdf.

the cost of individual free speech and privacy.[337] The ambiguity in these provisions create sufficient ground for political pressures to creep in, as the Government can invoke these provisions to argue that intermediaries are not doing enough, as has happened multiple times in the past[338] in the months after the Rashmika Mandana deepfake case in 2023.[339] Given the various implications on individual rights due to the language and framing of these obligations, it is **necessary to work towards governance systems that enable better platform design focussed on user risk assessments, safety-by-design and sufficient transparency of content moderation practices**.[340]

4. **Lack of mandate around legally ambiguous but harmful content:** While activities designated as 'crimes' have a straightforward redressal process (through the police and courts), harms emanating out of otherwise legal activities need further consideration. For instance, while harms arising out of systematic doxxing (publishing another person's personal information) may be actionable under provisions of the BNS or IT Act, enforcement remains uneven and they do not sufficiently address the variety of consequences that can follow a case of doxxing.[341] Other examples include online cyberbullying and online trolling.[342] **Users may rely on intermediaries' internal policies and content standards to access remedies, but these are subject to the vagaries of the intermediary's own stance on these issues, and the freedom of speech of the users in question**. Similarly, even if an affected user manages to obtain a court order under the IT Act and relevant criminal laws against such content, the legal process to do so may come at substantial cost of time and money, and may expose the affected user to offline harms as well.[343] **These issues are a symptom of a policy and enforcement framework that focuses excessively on individual content moderation related interventions, as opposed to investing in frameworks which promote responsive and responsible platform design which addresses systemic risks with proportionate risk mitigation measures.[344]**

---

[337] Sarkar, et al., *Legal Advocacy Manual: A primer on the jurisprudence of digital rights in India,* 2024, https://cis-india.org/internet-governance/legal-advocacy-manual/at_download/file; Aarathi Ganesan, *How Will the Proposed Amendments to the IT Rules Affect Free Speech and Intermediaries?*, 2022, https://www.medianama.com/2022/06/223-it-rules-amendments-india-free-speech-big-tech/

[338] Al Jazeera, *New Delhi gives itself power over social media content moderation*, 2022, https://www.aljazeera.com/economy/2022/10/28/in-india-govt-now-has-power-over-social-media-content-moderation.

[339] Soumyendra Barik, *Centre issues advisory to social media platforms over deepfakes after viral 'Rashmika Mandanna' video*, 2023, https://indianexpress.com/article/business/centre-deepfake-advisory-to-social-media-platforms-9017283/.

[340] Janjira Sombatpoonsiri and Sangeeta Mahapatra, *Regulation or Repression? Government Influence on Political Content Moderation in India and Thailand*, 2024, https://carnegieindia.org/research/2024/07/india-thailand-social-media-moderation?lang=en; Sangram Salgar, *A Comprehensive Approach to Content Moderation in Social Media Platforms,2024,*https://indiaai.gov.in/article/a-comprehensive-approach-to-content-moderation-in-social-media-platforms.

[341] Internet Freedom Foundation, *Why Doxxing Remains a Legal Grey Area: Navigating the Legal Uncertainty of Online Exposure*, 2025, https://internetfreedom.in/why-doxxing-remains-a-legal-grey-area-navigating-the-legal-uncertainty-of-online-exposure/.

[342] The Hindu, *Why laws fall short in combating the surge in cyber-bullying cases*, 2025, https://www.thehindu.com/news/national/why-laws-fall-short-in-combating-the-surge-in-cyber-bullying-cases/article69574500.ece.

[343] It is vital for the current legal framework to acknowledge the online-offline continuum, which explains that online actions can have real world consequences and even lead to physical harm. See Malavika Rajkumar and Shreeja Sen, *The Judiciary's Tryst with Online Gender Based Violence,* 2023, https://itforchange.net/sites/default/files/2190/The%20Judiciary's%20Tryst%20with%20OGBV_0.pdf.

[344] While the 'due diligence' obligations ask intermediaries to make 'reasonable efforts' to prevent harms to women and children, there is scope for greater deliberation and clarity on the design of these efforts. The United Kingdom's Online Safety Act 2023 attempted to address this issue recently. Though obligations around 'legal but harmful content' were dropped from the draft law earlier, there are hints these may be brought back. See Lucy Fisher, *UK considers forcing tech firms to remove 'legal but harmful' content after riots,* 2023, https://www.ft.com/content/d026a8d1-26d1-494d-83dc-5ff0204388e8; Chris Vallance, *Online Safety Bill: Crackdown on harmful social media content agreed,* 2024, https://www.bbc.com/news/technology-66854618.

5. **Persistent Problem of Respawning Sexual Content:** Both government and private stakeholders consider crimes involving CSEAM and NCII as high priority issues. Yet, it is noticeable that the redressal offered to affected users currently relies heavily on bona fide action by intermediaries on receiving a complaint, and a proactive judicial system that offers timely justice. Prior research indicates that intermediaries may be unresponsive for "*... reasons such as (i) lack of grievance officers, (ii) lack of content moderation capacity, (iii) belief that the risk of prosecution is remote, (iv) lack of concern over losing safe harbour, and (v) financial incentives such as advertising revenue derived from hosting NCII content*".[345] This is especially problematic when the harmful content in question may be taken down on one platform/ website, but may have already resurfaced elsewhere. Users will be compelled to repeat the complaint and takedown mechanism currently in place under the IT Act. Any future mechanism ought to ease this burden on users. **These researchers point to the Internet Watch Foundation (IWF), an independent forum in the EU for platforms, civil society and government stakeholders to work together to tackle CSEAM. The IWF uses a common hash database, assigning unique numeric values to the visuals, and enables intermediaries to take down content by providing access to this database.[346]**

6. **The Role of AI and Human Moderators in Addressing Abuse -** Moderation systems need to evolve to incorporate regional dialects, relevant subcultures, on-ground context, and intersectional perspectives, particularly in marginalized communities, to better detect and address issues like caste-based abuse, gendered violence, and other forms of discrimination.[347] This may require incentives for third party AI models that can understand diverse linguistic variations (e.g., Hinglish,[348] caste-specific terms) and collaborate with human moderators and local civil society experts to ensure nuanced, on-the-ground understanding.[349]

7. **Lack of legal mandate on prevention and remediation:** Even though various laws exist to deal with online harms against women and children, no specific organization is mandated to prevent the occurrence of such harms through digital literacy and sensitization activities. Although the government has shown increased responsiveness - evident in measures like enhancements to the National Cyber-Crime Portal and the formation of the I4C - these efforts remain largely reactive. **There is a pressing need to adopt a multistakeholder approach that actively involves intermediaries, civil society organizations, and users to ensure a more comprehensive and preventive response to online harms**.

---

[346] Internet Watch Foundation, https://www.iwf.org.uk/.

[347] Khandelwal, et al., *Casteist but Not Racist? Quantifying Disparities in Large Language Model Bias between India and the West,* 2023, https://openreview.net/forum?id=ED7lj2ThbB.

[348] Gabriel Nicholas and Aliya Bhatia, *The Dire Defect of 'Multilingual' AI Content Moderation*, 2023, https://www.wired.com/story/content-moderation-language-artificial-intelligence/.

[349] Oversight Board, *Content Moderation in a New Era for AI and Automation,* https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/?utm_source=chatgpt.com.

Chapter 4
International Trends on Children's Safety Interventions

4.1. Introduction

As we have mapped extensively in Chapter 2, kids and adolescents are having to navigate various online risks, including cyberbullying, exposure to harmful content, and online grooming.[350] This chapter examines different international experiences of crafting regulatory and policy interventions to improve online safety of children.

In April 2022, approximately[351] 70 countries signed the Declaration for the Future of the Internet, a set of voluntary principles that included safeguarding young people in a more digitized society. Moreover, G7 member states agreed upon internet safety principles in April 2021 that urged companies to address both illegal and harmful content toward children and supported further civil society and academic engagement on the issue. Shortly after, the OECD amended its Recommendation on Children in the Digital Environment, which called for a balance between mitigating online harms to children and maintaining the rights to free speech and online access for all.

---

[350] Oversight Board, *Content Moderation in a New Era for AI and Automation,* https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/?utm_source=chatgpt.com.

[351] Center for Strategic & International Studies, *A New Chapter in Content Moderation: Unpacking the UK Online Safety Bill*, 2023, https://www.csis.org/analysis/new-chapter-content-moderation-unpacking-uk-online-safety-bill.

More recently, the UN's 2024 Global Digital Compact[352] offers a framework for addressing children's online safety. It recommends that digital technology and social media companies undertake interventions that prioritise transparency and accountability across their terms of service, content moderation algorithms and better civil society access to platform data, among other priority areas. Further, the UN GDC encourages platforms to be more transparent about the details and performance of their online safety interventions.

A central principle guiding international regulatory discussions is safety by design.[353] Regulatory discourse is advocating that platforms proactively incorporate safe by default measures into the development and operation of online platforms.[354] Such proactive design related interventions are observed to be promising alternatives to more heavy handed strategies centered around age-gating children from accessing certain corners of the internet.

Definitional specificity of offenses and risk is also an essential element for effective online safety legislation. Clear and precise definitions of online harms help to ensure that laws are applied consistently and that individuals and organizations are held accountable.[355] Clear legal language can enhance the preciseness and effectiveness of legislative measures to address risks that are connected with cybercrime as well as risks that are not necessarily criminal, but yet may have harmful effects on children. Definitional specificity can help facilitate more tailored policy interventions that are suited to specific profiles of online safety risks.

Finally, discussions around co-regulation and self-regulation have increased when it comes to children's online safety.[356] While governments have a responsibility to protect children online, industry and civil society also play a crucial role in mitigating risks. By fostering collaboration between government, industry, civil society and other field experts, policymakers are increasingly realising the opportunity of leveraging the collective expertise of all actors to create safer online environments. Critically, such efforts must be inclusive and prevent risks of regulatory capture by large tech companies.[357] To appreciate the need for nuanced online children's safety regulatory interventions, let us first trace the growing political discourse on issues like age-gating.

4.2 Tracing the Intensification of Global Discourse on Children's Online Safety

The debate over children / adolescents' access to digital services has intensified in recent years. First we look at the dialogue emerging from **Australia[358], the EU and the US[359]** that has been at the forefront of discussions. In the United States, the Senate held a high-profile hearing with executives from major social media companies in 2024.[360] Lawmakers questioned the CEOs about their platforms' impact on young users, **particularly in relation to mental health, screen time, and the spread of misinformation**.

A more prominent measure came from the Australian parliament when it passed the **Online Safety Amendment (Social Media Minimum Age) Act in November 2024**. This measure

---

[352] United Nations, *Global Digital Compact*, 2024, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf.

[353] Organisation for Economic Co-operation and Development (OECD), *Towards digital safety by design for children*, 2024, https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/towards-digital-safety-by-design-for-children_f1c86498/c167b650-en.pdf.

[354] John Perrino, *Using 'safety by design' to address online harms*, 2022, https://www.brookings.edu/articles/using-safety-by-design-to-address-online-harms/

[355] World Economic Forum, *How can we prevent online harm without a common language for it? These 6 definitions will help make the internet safer*, 2023, https://www.weforum.org/stories/2023/09/definitions-online-harm-internet-safer/.

[356] Kevin & Rokša- Zubčević, *Towards coregulation of harmful content online in Bosnia & Herzegovina: A study of European standards and co-regulatory practices for combating harmful content online*, 2022, https://rm.coe.int/co-regulation-of-harmful-content-online-study-eng/1680adeef7.

[357] Organization for Security and Co-operation in Europe, *Self-regulation, Co-regulation, State Regulation* (2003), https://www.osce.org/files/f/documents/2/a/13844.pdf

[358] Hannah Ritchie, *Australia approves social media ban on under-16s*, 2024, https://www.bbc.com/news/articles/c89vjj0lxx9o.

[359] Times of India, *California becomes the 15th US state to limit smartphones in schools*, 2024, https://timesofindia.indiatimes.com/technology/tech-news/california-becomes-15th-us-state-to-limit-smartphones-in-schools/articleshow/113630879.cms.

[360] The Economic Times, *US Senate hearing: Meta, TikTok, X CEOs grilled by senators about child sexual exploitation*, 2024, US Senate hearing: Meta, TikTok, X CEOs grilled by senators about child sexual exploitation - The Economic Times.

mandates technology companies to restrict users under the age of 16 years from accessing social media platforms. The Act seeks to restrict minors under 16 from using video-sharing platforms TikTok and Snapchat, Instagram and Facebook and X by the end of 2025. Initially, the law had an exemption for certain UGC platforms like YouTube citing it as a valuable educational tool and not a core social media application, however now YouTube has also been included in the ban.[361] This proposal was heavily debated in Australia. **Julie Inman Grant, Australia's National eSafety Commissioner**, expressed concerns during a parliamentary inquiry on social media use in June 2024, stating that the ban "*... may limit young people's access to critical support.*"[362] **Daniel Argus, director of the digital media research center at Queensland University of Technology**, argued that there were two significant flaws with banning young people's access to social media. They argue that it "*... threatens to create serious harm by excluding young people from meaningful, healthy participation in the digital world, potentially driving them to lower quality online spaces, and removing an important means of social connection.*".[363]

France has seen a governmental **push to set 15 years as the minimum age** for being allowed access to social media[364] and recently the **French Digital Affairs and AI Minister called for unified efforts across Europe** to mobilise and push for 15 years as the minimum age for access to social media.[365] Recent reports have revealed an **EU-wide proposal for the age of digital adulthood[366]**, below which minors would require parental consent to be able to access social media.

In the context of these proposals, it becomes important to understand whether bans are feasible and are they a strong way to protect the interests of children. **Box 4.1 showcases an example from South Korea which elucidates the practical realities of access restrictions for children and how they are likely to respond to such realities**.

Box 4.1: South Korea's Gaming Law An Example of Why Bans Have Limitations

South Korea's **Youth Protection Revision Act[367]**, commonly known as the *Shutdown Law or Cinderella Law*, was an act of the country's National Assembly which forbade children under the age of 16 years from playing online video games between the hours of **00:00 hrs and 06:00 hrs**. The legislature passed the law in May 2011 and it went into effect in November 2011. The law was proposed by the government following an outcry prompted by several tragic events, most drastically including a teen suicide and a couple whose infant died of neglect, which the government blamed on addiction to video games.[368] The most popular games were played in Internet cafés called "PC bangs" and enable microtransactions, adding financial costs to the loss of time for studying and other pursuits and the deleterious effect on general well-being cited by authorities.

Impact: The law led those under sixteen to commit identity theft—underage South Koreans stole **national 13-digit resident registration numbers that detail people's name, gender, DoB, etc.** in an effort to elude the law.

Ultimately, the law was abolished in August 2021.[369] The Korean National Assembly repealed the provisions of Cinderella law from the Youth Protection Act (effective from January 1, 2022) on the grounds that it **hindered autonomy of the players and infringed upon the rights of game users and developers. Additionally, regulatory impact assessments had revealed that the Cinderella law failed to achieve its desired objective of reducing gaming usage for the target adolescents.**

---

[361] Reuters, *Despite Australia's strict social media ban for minors, a YouTube exemption poses risks,* February 2025, https://www.reuters.com/technology/despite-australias-strict-social-media-ban-minors-youtube-exemption-poses-risks-2025-02-03/

[362] The Guardian, *Social media age restrictions may push children online in secret, Australian eSafety commissioner says*, 2024, https://www.theguardian.com/australia-news/article/2024/jun/23/social-media-age-restrictions-may-push-children-online-in-secret-australia-regulator-says.

[363] Dwayne Oxford, *Australia's social media ban for minors: Has this worked elsewhere?*, 2024, https://www.aljazeera.com/news/2024/9/19/australias-social-media-ban-for-minors-has-this-worked-elsewhere.

[364] Politico, *France doubles down on age minimum of 15 for social media use*, 2024, https://www.politico.eu/article/france-doubles-down-on-social-media-age-limit-at-15/.

[365] Euronews, *France's AI minister calls for a Europe-wide ban on social media for children under 15*, 2025, https://www.euronews.com/next/2025/05/12/frances-ai-minister-calls-for-a-europe-wide-ban-on-social-media-for-children-under-15.

[366] Politico, *Europe's efforts to block kids from social media gather pace*, 2025, https://www.politico.eu/article/eu-children-social-media-regulation-platforms-big-tech/

[367] Korea Legislation Research Institute, *Youth Protection Act*, 2024, https://elaw.klri.re.kr/eng_service/lawView.do?hseq=67171&lang=ENG.

[368] Digital Media Wire, *Korea slaps curfew on gamers*, 2011, https://digitalmediawire.com/2011/11/28/korea-slaps-curfew-on-gamers/

[369] The Korea Herald, *Korea moves to abolish controversial anti-game rule, but tasks remain unresolved*, 2021, https://www.koreaherald.com/article/2725997.

4.2.1 Regulatory Outlook and International Proposals on Age Verification and their Feasibility

As discussed above, we observe a rising trend where different countries are putting forth age verification proposals. They seek to ensure that young users below certain thresholds do not access either (a) specific types of online content or (b) certain kinds of websites/apps. For example, the UK[370], EU[371], and Australia[372] have introduced guidelines and/or laws requiring social media companies enforcing age limits.

The UK regulator Ofcom's Draft Guidance on Children's Access Assessments[373] provides technical criteria that service providers should fulfil to ensure their age assurance is highly effective. **These include technical accuracy, robustness, reliability and fairness.** Service providers are also required to conduct annual assessment of children's access to their platforms, including an assessment of the effectiveness of their age assurance mechanisms.

The European Data Protection Board (EDPB) adopted the Statement 1/2025 on Age Assurance[374], which provides guidance on a risk-based and proportionate implementation of age assurance, centering the best interests of the child. Service providers are required to conduct age assurance in compliance with data protection principles such as purpose limitation and data minimisation. The Statement provides the criteria of **accessibility, reliability and robustness** to evaluate the effectiveness of age assurance measures.

In the US, Meta has proposed a federal legislative framework that shifts the responsibility for key online safety measures from individual apps to app stores.[375] This proposal suggests that app stores should handle tasks like getting parental approval for verifying a user's age. The rationale is to create a more consistent and streamlined process for parents, who would no longer need to provide sensitive information to hundreds of different apps. It is important to note that this is one of several proposals being considered, and its acceptance and implementation are still being discussed.

The Australian government's Age Assurance Technology trial report that seeks to implement the country's social media age gating legislation has found that while selfie-based age estimation software could be broadly accurate, it also revealed significant inconsistencies.[376] As media researcher Justine Humphry highlighted, these systems showed "a lot of variations in accuracy," particularly for individuals near the age cut-off. **The trial found a grey zone of high uncertainty for users up to three years on either side of the 16-year-old minimum, noting that teenage girls and non-caucasians experienced a higher rate of inaccuracies.**

Overall what we observe globally is that there is ongoing debate[377] about the effectiveness of age assurance. Some of the challenges include:

- **Accuracy**: The reliability of age estimation methods can vary, especially when dealing with factors like lighting conditions, image quality, and individual variations.

---

[370] Information Commissioner's Office, *Age Assurance for the Children's Code*, 2021, https://ico.org.uk/about-the-ico/what-we-do/information-commissioners-opinions/age-assurance-for-the-children-s-code/.

[371] European Commission, *Digital Services Act: Task Force on Age Verification*, 2024, https://digital-strategy.ec.europa.eu/en/news/digital-services-act-task-force-age-verification-0.

[372] Australian Government Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, *Tender awarded for Australian Government's age assurance trial*, 2024, https://www.infrastructure.gov.au/department/media/news/tender-awarded-australian-governments-age-assurance-trial#:~:text=The%20Age%20Assurance%20Technology%20Trial,harmful%20and%20inappropriate%20content%20online.

[373] Ofcom, Children's Access Assessments: Draft Guidance for Consultation, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/284469-consultation-protecting-children-from-harms-online/associated-documents/a5-draft-childrens-access-assessments-guidance.pdf?v=336055.

[374] European Data Protection Board, *Statement 01/2025 on Age Assurance*, 2025, https://www.edpb.europa.eu/system/files/2025-04/edpb_statement_20250211ageassurance_v1-2_en.pdf.

[375] Antigone Davis, *A framework for legislation to support parents and protect teens online*, 2024, https://medium.com/@AntigoneDavis/a-framework-for-legislation-to-support-parents-and-protect-teens-online-6565148b26b1.

[376] Australian Government Department of Infrastructure, Transport, Regional Development, Communications, Sports and the Arts, *Age Assurance Technology Trial Part A: Main Report*, 2025, https://www.infrastructure.gov.au/sites/default/files/documents/aatt_part_a_digital.pdf.

[377] CNIL, *Online age verification: balancing privacy and the protection of minors,* 2022, https://www.cnil.fr/en/online-age-verification-balancing-privacy-and-protection-minors.

- **Privacy Concerns[378]:** Some methods, such as facial recognition/estimation, raise privacy concerns, as they involve collecting, processing and storing biometric data.
- **Circumvention**: Users may find ways to bypass age verification checks, including using fake identities or obtaining adult accounts from others.
- **Scalability**: Age assurance methods might not translate in poorly networked areas and for populations facing challenges of digital access.

To address these challenges, regulators are increasingly collaborating with industry to develop more effective age verification and age estimation solutions.[379] Besides age verification, regulators are increasingly turning towards interventions that encourage safer platform design for interventions. The next section focuses on one such intervention i.e. platform risk assessments.

4.3 Risk Assessment as an Element of Safety By Design

Flowing from the idea of safety-by-design[380] regulators are increasingly steering the industry towards identifying risks for users on digital services or within specific platform features. In this context, international regulators are increasingly considering risk assessment mandates. For instance, as per draft Ofcom guidelines (2025), under the UK's Online Safety Act, 2023, regulated services must conduct an assessment of online safety risks.[381] The assessment is expected to enable platforms to delve into how harm might manifest while taking into account user demographics, platform features, and other relevant factors. Concurrently, platforms are expected to develop appropriate safety measures, particularly for safeguarding minors.[382]

Under the UK law, services have to complete an **illegal content risk assessment, and also a children's risk assessment if they are likely to be accessed by children**. Service providers must follow a four-step risk assessment process as part of the illegal content risk assessment. These steps are 1) understanding the kinds of illegal content that need to be assessed, 2) assessing the risk of harm by separately evaluating the likelihood and impact of all illegal content, 3) deciding the risk mitigation measures, recording the outcome of the assessment, and 4) reporting, reviewing and updating risk assessments.[383] Separately, ***services likely to be accessed by children*** must conduct a children's risk assessment, comprising a similar four-step process centred around harms/risks for children.[384] **The children's risk assessment is a separate duty that is distinct from the illegal content risk assessment.[385]** Ofcom's proposals present a structured approach consisting of four steps that are applicable across various services[386]:

1. **Context Establishment**: Identify and understand the risks of harm, referencing Ofcom's risk profiles[387] (comprising user-to-user risk profile and search-risk profile) and addressing any knowledge gaps.

[378] IAPP, *Online age verification could have broad privacy implications*, 2023, https://iapp.org/news/b/online-age-verification-could-have-broad-privacy-implications.

[379] Digital Trust and Safety Partnership, *Age Assurance: Guiding Principles and Best Practices*, 2023, https://dtspartnership.org/wp-content/uploads/2023/09/DTSP_Age-Assurance-Best-Practices.pdf.

[380] John Perrino, Using 'safety by design' to address online harms, 2022, https://cyber.fsi.stanford.edu/news/using-safety-design-address-online-harms.

[381] Ofcom, *Ofcom's approach to implementing the Online Safety Act*, 2023, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation.

[382] Ofcom, *Quick guide to children's risk assessments: protecting children online*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-risk-assessments.

[383] Ofcom, *Quick guide to illegal content risk assessments*, 2023,https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-online-safety-risk-assessments.

[384] Ofcom, *Quick guide to children's risk assessments: protecting children online*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-risk-assessments.

[385] Ofcom, *Quick guide to children's risk assessments: protecting children online*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-risk-assessments.

[386] Ofcom, *Quick guide to children's risk assessments: protecting children online*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-risk-assessments.

[387] The Ofcom guidance on children's risk assessment lays down some general and specific risk factors that inform children's user to user risk profile and search risk profile, Ofcom, *Children's risk assessment guidance and children's risk profiles*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/childrens-risk-assessment-guidance-and-childrens-risk-profiles.pdf?v=396653.

2. **Risk Assessment**: Evaluate the probability and consequences of potential harm, factoring in user demographics, platform features, and other pertinent variables, such as business model or commercial profile of the platform.
3. **Mitigation Identification**: Identify and assess potential measures to mitigate identified risks effectively.
4. **Review and Update**: Regularly review and update the risk assessment, especially in response to significant changes in the service.

Based on the determination of the risk assessment, the Protection of Children Codes of Practice recommend[388] different measures for different types of services depending on the size, capacity, and risk level (low-risk, specific-risk, or multi-risk) of a service. These are broadly divided into seven categories:

1. Recommender systems,
2. Terms of service and publicly available statements,
3. User support such as tools to restrict harmful online interaction and educational resources[389],
4. Search moderation,
5. Governance and accountability,
6. Content moderation, and
7. Age assurance.

Similarly, the EU's Digital Service Act ("DSA") calls for a broad-based assessment[390] of systemic risks including any actual or foreseeable negative effects in relation to the protection of minors. Articles 34 and 35 of the DSA require online platforms and search services, designated by the European Commission as very large (active monthly EU users above 45 million),to annually assess negative effects of their services for the protection of minors, the rights of the child, and serious negative consequences for their physical and mental well-being and mitigate any identified systemic risk. For each category of systemic risk set out in Article 34 of the DSA, a **meta-analysis across risk assessments should aim to synthesize[391]:**

- The taxonomy of harms related to the categories of systemic risks,
- The harms arising from these systemic risks ,
- The regulatory perspectives on addressing these systemic risks, and
- Data gathering and reporting on the categories of systemic risks.

What we observe is that the European approach to risk assessments and risk mitigations differs from the UK. The UK adopts a more pointed approach towards specific risk assessments for children's online safety and the EU DSA takes a broader approach when harm to minors is one type of online safety challenge that must be attended to by large platform operators. In the next section we take a look at how regulators are contending with the issue of content moderation in the context of children's online safety.

4.4 Proportionate Content Moderation Standards

In this section we juxtapose the UK and Australia's respective Online Safety Act(s) to better understand different ways that other countries are attempting to classify and define harms. The following tables (4.1 and 4.2) provide an overview of how online harms/risks are defined under the UK and Australian OSAs respectively. What we observe is a repeated emphasis on specific definitions with concrete detail, accompanied with nuanced categorisation of different types of risks

[388] Ofcom, *Quick guide to Protection of Children Codes*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-safety-codes;ReedSmith, *UK Online Safety Act: Children Protection Consultation*, 2024, https://www.reedsmith.com/en/perspectives/2024/06/uk-online-safety-act.

[389] Ofcom has proposed certain user support measures to give children more control over their online experience and safety. These could include tools for controlling risk in interaction, such as an option to decline group invites, block or mute user accounts and disable comments on their posts, or some user support materials that can augment their understanding of harmful online interactions, in Ofcom, *Protecting children from harms online: a summary of our consultation*, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/284469-consultation-protecting-children-from-harms-online/associated-documents/summary-of-consultation.pdf?v=336045.

[390] European Union, *Article 34 of the Digital Services Act*, 2024, https://www.eu-digital-services-act.com/Digital_Services_Act_Article_34.html.

[391] Centre on Regulation in Europe, *Cross-cutting issues for DSA systemic risk management: an agenda for cooperation*,2024,https://cerre.eu/wp-content/uploads/2024/07/240709_CERRE_DSA-Systemic-Risk_Cross-Cutting-issues-for-DSA-risk-management_FINAL.pdf.

that can help platforms distinguish between risks that are clearly illegal, as well as risks that can be treated as legal but harmful. The latter becomes essential for proportionate interventions and can be viewed as consistent with legal principles like the need for *different and graduated responses* to harmful online speech[392].

Table 4.1: Definition of Online Harms in UK

UK's Online Safety Act, 2023[393]

The UK's OSA adopts a tiered approach to identifying harmful content for children. Each category is treated with varying degrees of legal severity and regulatory expectation, reflecting the graduated risk they pose to children. This tiered approach allows Ofcom, to mandate stricter duties and enforcement actions for the most egregious forms of content, while still providing a framework for managing other harmful material. Specifically, the distinctions primarily lie in:

Legal Duties on Platforms: Services face stronger, often absolute, duties to prevent access to and swiftly remove "**Primary Priority Content" for children**. Platforms also have duties for another category designated as **"Priority Content"**. While robust these may involve more nuanced risk mitigation strategies. For "**Non-designated content**" that is harmful to children, platforms are expected to conduct risk assessments and implement proportionate measures based on identified risks. Given below are the definitions of these content categories:

Primary Priority Content: This category represents the **most severe and highly regulated types of harmful content**. Platforms have the strongest duties to prevent children from encountering this material. It includes:

- Pornography
- Content that encourages, promotes, or provides instructions for self-harm, eating disorders, or suicide.

Priority Content: This category encompasses other forms of content that are **significantly harmful to children**, and platforms have clear duties to implement measures to address them. It includes:

- Bullying, abusive, or hateful content
- Content which depicts or encourages serious violence or injury
- Content which encourages dangerous stunts and challenges
- Content which encourages the ingestion, inhalation, or exposure to harmful substances.

Non-designated content that is harmful to children: This is a broader, more flexible category designed to capture other **evolving or context-specific harms not explicitly listed above.** While not subject to the same strict specific duties as the "Priority" categories, **platforms are still legally obligated to assess and mitigate risks arising from such content.** It includes any content not within the "Primary Priority" or "Priority Content" categories but which presents a **material risk of serious harm to an appreciable number of children in the UK**. Some examples include:

- **Content promoting unrealistic body image or cosmetic procedures:** Outside of content promoting eating disorders (classified as Priority Content), this category may include (among other things) features that promote filtered or edited images, promote extreme cosmetic surgery trends, or set unrealistic beauty standards for children. This is a way to address issues that are significantly harmful to teens' self-esteem and mental health.

[392] Faris, Robert, Amar Ashar, Urs Gasser, and Daisy Joo, *Understanding Harmful Speech Online*, 2016, https://cdn.prod.website-files.com/646feeef1697362ad70b19d9/64af4fb5e332fcdec6cca838_Understanding%20Harmful%20Speech%20Online%20.pdf.

[393] UK Public General Acts, *Online Safety Act*, 2023, https://www.legislation.gov.uk/ukpga/2023/50.

UK's Online Safety Act, 2023[393]

- **Misinformation/Disinformation specifically harmful to children:** This could include false health claims (e.g., about vaccinations or fake cures) that, if believed by children, could lead to physical harm. It might also encompass conspiracy theories designed to incite fear or anxiety in children, or content that distorts historical events in a way that is profoundly upsetting or radicalizing for young audiences.
- **Content encouraging excessive or addictive behaviors:** This could include content that glorifies or normalizes excessive screen time, online gambling, or addictive gaming mechanics in a way that significantly harms a child's development, sleep, or social life.
- **Harmful trends or "challenges" that don't involve serious physical injury:** While "dangerous stunts and challenges" are Priority Content, there might be online trends that encourage less overtly physically dangerous but still **psychologically damaging behaviors (e.g., public shaming challenges, or trends that lead to social exclusion or harassment)**.

Enforcement Actions: Ofcom has greater powers to take enforcement action, including substantial fines, for failures related to Primary Priority Content compared to other categories.

Table 4.2: Definition of Cyberbullying in Australia

Australia's Online Safety Act, 2021

In Australia, the Online Safety Act (OSA) outlines a clear regulatory framework to address harmful online content, including **cyberbullying targeting children**. The legal threshold for cyberbullying under Australia's OSA centers on whether the material is "*likely to have the effect of... **seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child.***" If the content meets this threshold, the **eSafety Commissioner has the authority to issue content removal notices** at online service providers and in some cases the individual concerned. Failure to comply can result in significant penalties.

For greater specificity, the Australian legislation also provides detailed explanation of each of the key qualifying terms to allow regulators and companies to interpret if a particular activity constitutes as cyberbullying. **The Australian OSA typically excludes content that is merely offensive or insulting.** An explanation of those terms is presented below:

- **Seriously Threatening:** This involves content where an individual explicitly posts or comments about intending to harm a child, or encourages others to do so. For example, direct statem-ents like *"We're going to wait outside your house and hurt you when you come out,"* or *"If you stop talking to me, I will publicize your address and phone number across the internet."*
- **Seriously Intimidating:** This refers to content designed to instill significant fear in a child, even if it doesn't contain an overt, direct threat. An illustration could be **sharing an image depicting a person's head in a guillotine with the caption "Informers face consequences" and tagging the child**.
- **Seriously Harassing:** This category applies when a person consistently sends messages to a child or repeatedly reposts material about them. While individual messages or posts might not, on their own, meet the definition of cyberbullying, the **cumulative effect of persistent, unwanted communication elevates the impact and constitutes serious harassment**.
- **Seriously Humiliating:** This describes instances where an individual posts a comment or distributes an image that **causes a child profound embarrassment**. An example might be a **video showing a child with a physical disability falling in mud at school** and then crying about having to wear soiled clothes, accompanied by captions and comments ridiculing the child's disability and perceived inability to move properly.

We also observe some other relevant factors that regulations assess when determining online harms. For example, the UK's OSA and EU's DSA adopt two differing approaches.[394] The UK considers harm in a narrow sense of physical and psychological harms that **impact individuals resulting from specific illegal activities**. Conversely the **EU considers harms more broadly, and assesses the impact on both individuals and society**. Thus, the EU's approach to harm identification is more systemic rather than exclusively examining individual incidents. Overall, the following constitutes key considerations that other countries have factored in when defining and categorising online harms:

- **On Specificity of Definitions:** Both the UK and Australia have taken care to define prohibited content in a specific and precise manner. This helps to ensure that platforms and regulators have a clear understanding of what constitutes harmful content and can take appropriate action without disproportionately affecting people's right to free speech online.
- **Legal but Harmful Acts:** The **UK's Online Safety Act (OSA)** sets out categories of harmful content which are not illegal. **It calls for complete prohibition of children's access to Primary Priority Content, and only age-appropriate access to Priority Content**.
- **Contextual Understanding & Balancing Free Speech:** Australia's OSA emphasizes the **importance of context in determining whether material constitutes cyberbullying**. This helps to prevent the overreach of content moderation and ensures that only material that is truly harmful to a specific child is targeted. Both the UK and Australia have sought to balance the need to protect users from harmful content with the importance of preserving free speech. This is reflected in the **careful consideration of the context in which content** is shared and the **avoidance of overly broad or vague definitions**.

Having looked at content or speech related aspects of online safety, another vector of online risks emerge from peer to peer interactions. The next section looks into global interventions on this issue.

4.5 Peer on Peer Abuse/ Harmful Sexual Behaviour

Inappropriate behaviour among children and adolescents that is abusive in nature, including behaviours within intimate personal relationships, are increasingly manifesting into online spaces.[395] Often referred to as child on child or peer on peer abuse, such behaviours pose an additional layer of complexity to online safety strategies. The UK Government's Office for Standards in Education, Children's Services and Skills (Ofsted) defines peer on peer abuse to include physical and sexual abuse, sexual harassment and violence, emotional harm, online and offline bullying, teenage relationship abuse etc.[396]

In its statutory guidance on abuse within educational institutions, the UK's Department of Education[397] defines sexual harassment as "unwanted conduct of a sexual nature" that can occur both online and offline. It specifically defines online sexual harassment as inclusive of both consensual and non-consensual sharing of nude or semi nude images of those under 18, sharing of unwanted explicit content, sexualised online bullying, unwanted sexual comments and messages,

[394] Benjamin Farrand, *How do we understand online harms? The impact of conceptual divides on regulatory divergence between the Online Safety Act and Digital Services Act*, 2024, https://www.tandfonline.com/doi/full/10.1080/17577632.2024.2357463#d1e540.

[395] Safeguarding Network, *Child on child abuse*, 2023, https://safeguarding.network/content/safeguarding-resources/peer-peer-abuse/.

[396] Ofsted, *What is peer on peer abuse?*, 2019, https://educationinspection.blog.gov.uk/2019/10/04/what-is-peer-on-peer-abuse/.

[397] UK Department of Education, *Keeping children safe in education 2024: statutory guidance for schools and colleges*, 2024, https://assets.publishing.service.gov.uk/media/66d7301b9084b18b95709f75/Keeping_children_safe_in_education_2024.pdf.

including on social media, and coercing others into performing acts or sharing images online, which they may be uncomfortable with.

The UK guidance uses the term Harmful Sexual Behaviour (HSB) to refer to problematic sexual behaviours by children. HSB can take place either online or face to face or even simultaneously in both contexts. The guidance specifically lists some support tools for schools to defer to in cases of online sexual harassment.[398] The report of the UK Children's Commissioner on findings on online peer on peer abuse[399] which preceded the Online Safety Act, proposed that commercial pornography sites should have dedicated child-facing complaints route for children, and directed companies to use accurate software to scan for child abuse and grooming in private messaging services.

In Australia, the eSafety Commission recognises online sexual behaviours that are inappropriate at an early developmental stage and online sexual behaviours that are exploitative or abusive.[400] Inappropriate online sexual behaviours include sending nude photos or videos, participating in sexual chat or explicitly describing sexual acts and participating in sexual acts by webcam. Exploitative or abusive online sexual behaviours include posting sexually harassing comments or content about someone else, pressuring others to send nudes, sharing intimate images or videos without the consent of the person shown and sextortion. Australia's eSafety Commissioner adopts a restorative, instead of punitive approach to online HSB by minors. **It recommends using a child-centric and trauma-informed response to online HSB, besides using whole-of-community awareness and education programs to educate youth on respectful relationships. Its enforcement strategies to tackle online HSB include a cyberbullying team, image-based abuse team and cyber report team**.

In South Korea social interventions (by students) and civil society interventions have leant greater credibility and success to legislative interventions. For example, **South Korea's 'Sunfull Movement'** has achieved success in combating cyberbullying. The Sunfull or the good reply movement was conceptualised for school children. It uses a strategy of flooding online forums with positive messages to discourage bullies from making negative or vicious comments. Another example of a social intervention is the Nuri Cop, which refers to civil society volunteers who clean up and patrol the Internet by deleting child pornographic images.

The next section discusses how regulators are trying to address issues relating to online platform design. This includes how regulations are attempting to address challenges associated with (a) content recommender systems; and (b) content classification.

4.6 Improving Online Platform Design

4.6.1 Content Recommender Systems and User Remedy

Internationally organisations like the Integrity Institute[401] have mapped how proactive content recommendation systems present challenges for the well-being and development of children. UK's OFCOM agency maintains that recommender systems (algorithms which provide personalised recommendations to users) can harm children online.[402] The UK regulator

[398] Support tools for protecting children from online sexual harassment include an online safety helpline by the UK Safer Internet Centre, Image and videography removal support by the Internet Watch Foundation and Childline, UKCIS guidance on response to NCII and educational material for parents by the CEOP Education Program.

[399] UK Children's Commissioner, *Interim findings on Government's Commission on online peer on peer abuse*, 2021, https://assets.childrenscommissioner.gov.uk/wpuploads/2021/09/occ-interim-findings-on-governments-commission-on-online-peer-on-peer-abuse.pdf.

[400] Australia eSafety Commissioner, *Online harmful sexual behaviours in children and young people under 18 – position statement*, 2020, https://www.esafety.gov.au/industry/tech-trends-and-challenges/harmful-sexual-behaviours-under-18.

[401] Integrity Institute, *Child Safety Online*, 2024, https://integrityinstitute.org/blog/child-safety-online.

[402] Ofcom, *Protecting children from harms online: a summary of our consultation*, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/284469-consultation-protecting-children-from-harms-online/associated-documents/summary-of-consultation.pdf?v=336045.

argues that these systems may serve up solicited, dangerous content to children in their personalised news feeds. **As a consequence, Ofcom proposes that any service which operates a recommender system and is at higher risk of harmful content must use highly-effective age assurance to identify who their child users are. They must then configure their algorithms to filter out the most harmful content from these children's feeds and reduce the visibility and prominence of other harmful content.** Children must also be able to provide negative feedback directly to the recommender feed, so it can better learn what content they don't want to see. **This shows a distinct preference for regulators to nudge platforms to deliver age appropriate experiences for children**.

Ofcom's Protection of Children Codes prescribe proportionate measures to be implemented by **user-to-user services**[403] and **search services**[404]. Under the draft Codes[405], all user-to-user services must have content moderation systems and processes that ensure swift action is taken against content harmful to children. **Services must aim to prevent children from accessing Primary Priority Content. Services must also mitigate risks and harms from Priority Content and Non-Designated harmful content and demonstrate such endeavours through their *terms of service***.

Reporting and complaints must be enabled for users of service whenever they come across content that is harmful for children. User-to-user services must also provide age-appropriate user support materials for children. Such materials should contain information on how users or affected persons can report content that they consider harmful to children. For child accessible parts of a user-to-user service, the service provider must explain how to block or mute other user accounts, how to restrict other users from commenting on one's post, and how to control whether to become part of a group chat. User-to-user services must present such information in a clear, comprehensible and easy manner, with audio-visual and interactive elements wherever directed at children, and a separate explanatory section for the parent or guardians of children.

Search engines are expected to take similar action. **Where a user is believed to be a child, large search services must implement a safe search setting which cannot be turned off and must filter out the most harmful content.** Search engines are also expected to conduct children's risk assessment, and mitigate risks and impact of harms for the specific age groups that they identify in such assessments. **Search engines must enable users to report predictive search suggestions which direct users (children) towards Primary Priority Content or Priority Content.** Subsequently, if a search service provider finds on risk assessment that a predictive search suggestion presents a clear and material risk of a user encountering *Primary Priority Content or Priority Content*, then it must take appropriate risk mitigation steps for child accessible parts of the service. **For transparency, search engines are required to explain how they fulfil their safety duties in a publicly available statement, as well as enable reporting of harmful content.**

These measures are founded in an **Ofcom guidance (RS2)**[406] from May 2024 **which requires platforms to significantly reduce the visibility of potentially harmful content (Priority Content) in children's recommendation feeds, and extend such measures to NDC as well.** This includes content flagged for review

[403] Ofcom, *Protection of children code of practice for user to user services*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/protection-of-children-code-of-practice-for-user-to-user-services-.pdf?v=395966.

[404] Ofcom, *Protection of children code of practice for search services*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/protection-of-children-code-of-practice-for-search-services.pdf?v=395675.

[405] Ofcom, *Protecting children from harms online: Volume 3: the causes and impacts of online harms to children*, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/284469-consultation-protecting-children-from-harms-online/associated-documents/vol3-causes-impacts-of-harms-to-children.pdf?v=336052.

[406] Ofcom, *Protecting children from harms online: a summary of our consultation*, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/284469-consultation-protecting-children-from-harms-online/associated-documents/summary-of-consultation.pdf?v=336045.

but not yet removed. While specific metrics for downranking are not provided, platforms' measures should prioritize reducing content visibility over user engagement factors. However, the effectiveness of this measure will depend on factors like the extent of downranking, platform design, and the initial amount of harmful content.

Meanwhile, **Article 27 of Europe's Digital Services Act[407] requires platforms that use recommendation systems to set out in their T&Cs the main parameters they use for such systems, including any available options for recipients to modify or influence them.** It also mandates risk assessments in this context from the perspective of potential negative consequences on minors.

In the USA, New York's Empire State's SAFE Kids Act[408] is the first law to focus on the algorithms used by platforms to serve customized content to users. It specifically bans the use of machine-learning-designed content feeds for kids under age 18, unless their parents give consent. The law, which has been in effect since June 2025, may alter the way social media platforms operate within the state by requiring them to come up with an alternate feed for minors that would not involve an algorithm based on a child's individual user history. This alternate feed also cannot withhold, degrade, lower the quality, or increase the price of any product. **The law has prompted Meta (the parent company of Facebook and Instagram) to implement a non-algorithmic feed as the default for users under 18 in New York, and restrict late-night notifications for these users.[409]**

4.6.2 Content Classification Labels for Age Appropriate Experiences

Before the UK's OSA came into force in July 2025, Video Sharing Platforms (VSPs) like Twitch were required to take appropriate measures to prevent under-18 users from accessing pornography and other harmful content, under erstwhile regulations. These content classification requirements under the older VSP regulatory regime now overlap with regulatory obligations under the UK's OSA.[410] Per our analysis, UK's Ofcom is using regulatory levers to encourage video sharing platforms to adopt best-in-class content classification standards in order to ensure children are not exposed to age inappropriate content/experiences. Usually these platforms use content classification standards to label content as mature-rated, depicting intoxication, containing violence, depicting nudity/obscenity, etc.

These incentives/nudges from the UK regulator are observable through examples like Ofcom's September 2024 evaluation of content classification measures by the popular interactive live streaming service– Twitch. Ofcom's assessment observes that Twitch's measures required content creators to apply content classification labels to tell viewers if a stream contains certain mature themes. **Ofcom found that, following changes to Twitch's content classification guidelines, the accuracy of content labeling increased substantially.[411]** The regulator observed that Twitch adopted certain policy changes[412] to its content classification guidelines that allowed creators to categorise their content into seven adult-themed content classification categories, namely:

- mature-rated game;
- sexual themes;
- drugs, intoxication or excessive tobacco use;
- violent and graphic depictions;
- significant profanity or vulgarity,

[407] Article 27, Europe's Digital Services Act, 2022.

[408] The New York State Senate, *Senate Bill S7694A establishing the Stop Addictive Feeds Exploitation (SAFE) for Kids act prohibiting the provision of addictive feeds to minors*, 2023, https://www.nysenate.gov/legislation/bills/2023/S7694/amendment/A.

[409] Brooklyn Daily Eagle, New York's SAFE for Kids Act drives Instagram's changes to protect minors, September 2024, https://brooklyneagle.com/articles/2024/09/20/new-yorks-safe-for-kids-act-drives-instagram-changes/#:~:text=The%20changes%2C%20they%20said%2C%20come,and%20protecting%20children's%20privacy%20online.

[410] Twitch, *Content Classification Guidelines*, https://safety.twitch.tv/s/article/Content-Classification-Guidelines?language=en_US.

[411] Ofcom, *How accurate are Twitch's new content classification labels?*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/how-accurate-are-twitchs-new-content-classification-labels.

[412] Twitch, *Content classification labels*, https://help.twitch.tv/s/article/content-classification-labels?language=en_US#Top.

- Gambling; and
- Politics and sensitive social issues

The UK OSA's **Children's Safety Codes of Practice**[313], published by **Ofcom** in May 2025, now include a **specific duty** for **user-to-user** and **search services** to **use content labelling, classification, or age rating systems to support age-appropriate experiences for children.** This includes: (i) Labelling or tagging content according to **age-appropriateness;** (ii) Using classification systems like **BBFC** or in-house automated tools and (iii) Displaying **age ratings** on **videos, games, and potentially harmful posts**. Thus, what we observe is that in creator-led ecosystems age classification policies are being encouraged by international regulators to facilitate age appropriate online experiences.

Having discussed how regulators are encouraging healthier platform design, the next section discusses how co-regulatory and self-regulatory efforts are being explored abroad to standardise children's online safety efforts.

4.7 Co-Regulation and Self Regulation

The robustness of online safety measures can be better tested against common benchmarks. Such measures can promote uniform safety standards for children across platforms and use cases. To this end we look at the role played by self-regulation and co-regulatory efforts.

4.7.1 UK: Age Appropriate Design Code (AADC)

The AADC is a statutory code of practice meant to reinforce the General Data Protection Regulation (GDPR) for providers of information services most likely to be accessed by children. The AADC presents a set of 15 interlinked standards for service providers to comply with depending on a proportionate risk-based assessment. **The AADC mandates service providers to implement privacy by default and design their platforms keeping in mind the best interests of the child.**

A Children and Screens report[314] that mapped the impact of the AADC on digital platforms identified 91 changes made across major social media and digital platforms towards a safer and age-appropriate Internet for children. These changes were implemented by digital platforms across four key areas -

- **Youth safety and well being**: Platforms have implemented targeted features to protect minors and manage their online interactions. Examples include YouTube Kids disabling the autoplay feature by default to limit continuous consumption. Similarly parents on the platform have tools that can tailor the experience, with different content settings available for preschoolers, younger kids, or older kids. Other platform initiatives include Instagram introducing the "take a break" feature and Snapchat introducing methods to allow users to pause Snap Streaks.
- **Privacy/security and data management**: One example cited is TikTok updating its reporting mechanisms and explaining the same in guidelines. Additionally, TikTok, Instagram and Youtube **enabled users to filter offensive comments** and TikTok enabled **opt-in feature for personalised advertisements for under-18 users**.
- **Encouraging Age Appropriate Consumption Habits**: Examples include Instagram disabling users from tagging minors who do not follow them, Instagram notifying minors when they interact with an adult flagged for suspicious

[413] Ofcom, *Statement: Protecting Children from Harms Online*, 2025, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-protecting-children-from-harms-online?utm_medium=email&utm_campaign=New%20rules%20for%20a%20safer%20generation%20of%20children%20online&utm_content=New%20rules%20for%20a%20safer%20generation%20of%20children%20online+CID_3ea132bd5b0f1ba3656a2368b6c35bbd&utm_source=updates&utm_term=under%20transformational%20new%20protections%20finalised%20by%20Ofcom%20today.

[414] Children and Screens, Institute of Digital Media & Child Development, *UK Age Appropriate Design Code Impact Assessment*, 2024, https://www.childrenandscreens.org/wp-content/uploads/2024/03/Children-and-Screens-UK-AADC-Impact-Assessment.pdf.

behaviour, and Google enabling users under 18 and their parents to remove their images from Google search.

- **Time management**: Examples include TikTok enabling a time bound curfew on push notifications for children, Youtube Kids removing overly commercial content and Google automatically enabling a SafeSearch feature that filters inappropriate content for users under 18. Similarly, platforms like YouTube have created product nudges that encourage users aged between 13-17 years to take breaks and incorporate bedtime reminders.

According to a report[415] by the Digital Futures Commission[416] Meta, TikTok, Google and Snapchat made 42 changes focussed on children's privacy in 2021, after the AADC came into effect the same year. **Some of the most important changes recorded, linked to legislation and regulation, included social media accounts defaulting to privacy-enhanced settings, changes to recommender systems and restrictions on targeted advertising to children. Other changes were made to product tools, information and support mechanisms**. However, further research is needed to assess the full extent of the benefits. According to Professor Sonia Livingstone, LSE, Child Internet Expert, "*The AADC marks a cultural shift — from blaming children for online harms to holding tech companies responsible for designing with children's best interests at heart.*"[417]

4.7.2 UK: Ofcom's Codes of Practice

In December 2024, UK's Ofcom published[418] its **final codes of practice under the OSA.**[419] The code of practice forms the framework for how children will be protected from illegal activity and content, and other harms such as bullying. These are different from the aforementioned AADC.

The Ofcom Codes of Practice were developed through a rigorous and multi-stage process led by the UK's online safety regulator, and is mandated by the Act itself. This process **emphasized wide public consultation, evidence-gathering, and an iterative approach to ensure the codes are effective and proportionate.** Volume 4, specifically concerning the ***Protection of Children Codes***, outlines the practical measures and regulatory framework that applies to digital services in furtherance of children's online safety.[420] Given below are the key features and areas addressed in Volume 4 of Ofcom's Codes of Practice[421]:

- **Risk Mitigation Strategies:** Under this umbrella the code of practice includes measures related to:
  - **Content Moderation:** How services should manage and remove content harmful to children, including illegal content (like child sexual abuse material) and legal but harmful content (e.g., content promoting self-harm, eating disorders, or bullying).

---

[415] Steve Wood, Impact of regulation on children's digital lives, 2024, https://www.digital-futures-for-children.net/our-work/regulation-impact.

[416] The Digital Futures Commission preceded the Digital Futures for Children Centre (DFC), was a research programme led by Prof. Sonia Livingstone, London School of Economics. Guided by Commissioners with expertise on the intersection of children with digital technologies, the Commission facilitated collaboration between innovators, policymakers, regulators, academics and civil society. The aim was to center children's rights and needs in policymaking goals for digital technologies.

[417] Sonia Livingstone, *Child online safety - next steps for regulation, policy and practice*, 2025, https://blogs.lse.ac.uk/medialse/2025/01/22/child-online-safety-next-steps-for-regulation-policy-and-practice/.

[418] Liv McMahon, *Social media given 'last chance' to tackle illegal posts*, 2024, https://www.bbc.com/news/articles/cwy83jdpgw5o.

[419] Ofcom, *Statement: Protecting children from harms online*, 2025, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-protecting-children-from-harms-online?utm_medium=email&utm_campaign=New%20rules%20for%20a%20safer%20generation%20of%20children%20online&utm_content=New%20rules%20for%20a%20safer%20generation%20of%20children%20online+CID_3ea132bd5b0f1ba3656a2368b6c35bbd&utm_source=updates&utm_term=under%20transformational%20new%20protections%20finalised%20by%20Ofcom%20today

[420] Ofcom, *Protecting children from harms online: Codes at a glance*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/codes-at-a-glance.pdf?v=395791#:~:text=This%20document%20summarises%20the%20measures%20in%20our%20Protection,in%20detail%20in%20Volume%204%20of%20this%20statement.

[421] Ofcom, *Protecting children from harms online: Codes at a glance*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/statement-protecting-children-from-harms-online/main-document/codes-at-a-glance.pdf?v=395791#:~:text=This%20document%20summarises%20the%20measures%20in%20our%20Protection,in%20detail%20in%20Volume%204%20of%20this%20statement.

- **Recommender Systems:** Updates and measures to ensure that algorithmic recommender systems do not amplify or expose children to harmful content.
- **Age Assurance:** Emphasizing the use of highly effective age assurance to identify child users and deliver age-appropriate experiences.
- **Safe Search Settings:** For large general search services, measures include applying "safe search" settings to filter out primary priority content (like pornography) from children's search results.
- **User Controls:** Measures to provide children and their caregivers with tools and support, such as blocking and muting other users, disabling comments, and easily accessible reporting and complaints functions.
- **Design Choices:** Encouraging safety by design in platform operations and processes to prevent harm to children from the outset.

- **Scope and Applicability:** Volume 4 clarifies which services and content categories these measures apply to. While focused on services likely to be accessed by children, it **distinguishes between user-to-user services and search services**, and often applies measures proportionally based on the service's size and the level of risk it poses to children.
- **Governance and Accountability:** It sets out expectations for robust internal governance structures within service providers, ensuring senior oversight and accountability for children's safety, as well as internal monitoring and assurance functions to ensure safety measures are effective.
- **Baseline Protection:** It aims to establish a strong baseline level of protection for children of all ages, while also encouraging providers to implement actions that are appropriate for different age groups.

4.7.3 New Zealand: Code of Practice for Online Safety and Harms

The **Aotearoa New Zealand Code of Practice for Online Safety and Harms**[422] is a **voluntary industry code** that provides a self-regulatory framework aimed at improving users' online safety and minimizing harmful content online with a focus on organizations providing online services to people in New Zealand.

The code was developed between April 2021 and March 2022 by Netsafe – an independent, non-profit online safety organization, that provides online safety support, expertise and education to people in Aotearoa New Zealand – in collaboration with industry and consultation with Māori advisers, government, civil society and the public. **The code was initially drafted with the involvement of major digital platforms, including Meta (Facebook and Instagram), Google (YouTube), TikTok, Twitch and Twitter, who are current signatories**.

Signatories have committed to implementing policies, processes, products and/or programmes that seek to promote safety and mitigate risks that may arise from the propagation of harmful content online while respecting freedom of expression, privacy and other fundamental human rights. **This can include measures to prevent known child sexual exploitation and abuse material from being made available on their platforms, to protect children against predatory behaviours like online grooming, and to reduce or mitigate the risk to individuals (minors and adults) or groups from being the target of online bullying or harassment.**

Notably, the New Zealand code takes[423] a systems and outcomes based approach towards online safety and content moderation. It facilitates accountability through transparency

[422] The Code, *Aotearoa New Zealand Code of Practice for Online Safety and Harms*, 2022, https://thecode.org.nz/wp-content/uploads/sites/38/2023/06/THE-CODE-DOCUMENT-FINAL.pdf.

[423] World Economic Forum, *Digital Safety Risk Assessment in Action: A Framework and Bank of Case Studies*, 2023, https://www3.weforum.org/docs/WEF_Global_Coalition_Digital_Safety_Risk_Assessments_2023.pdf.

of policies, processes, systems and outcomes. The Code advocates that instead of implementing interventions that may become outdated in the rapidly changing digital ecosystem, platforms should focus on adaptable measures.

The code applies broadly and provides flexibility to all signatories to **respond and comply in a way that best matches their risk profiles. The code facilitates accountability through transparency reporting that helps certify if a signatory exceeds, meets or falls short of code commitments.** The launch of the New Zealand Code of Practice led to the inclusion in transparency reports of online service providers of a New Zealand-specific focus, providing visibility to the community on policy enforcement, data requests handling and intellectual property protection.[424]

4.7.4 Combined Takeaways from Platform Design Codes of Practices in the UK and New Zealand

The case studies in this section highlight the potential of regulatory and self-regulatory codes to assist with standardising industry responses towards children's online safety. The UK AADC demonstrates how statutory codes, linked to overarching data protection laws, can drive significant, tangible changes in platform design that embed privacy and child welfare by default. Ofcom's Codes of Practice under the UK's OSA aim to incentivise proactive safety platform design, ensure a baseline protection for child users, grant more control to users and their caregivers, prioritise governance and accountability, and grant platforms flexibility to develop risk mitigation strategies that are curated to their service features.

New Zealand's Voluntary Code of Practice showcases the potential of inclusive and context driven, industry-led initiatives to foster a culture of accountable and adaptable safety measures. Its system- and outcomes-based approach, coupled with transparency reporting, offers flexibility while still encouraging commitment to mitigating online harms.

The evidence suggests a hybrid model, where clear legislative frameworks set foundational standards and drive compliance (as seen with the AADC), while flexible, industry-led codes enable rapid adaptation to technological changes and foster innovation in safety solutions. **Future efforts to advance children's online safety will likely benefit most from a dynamic interplay between these approaches, fostering both accountability and agility in an ever-changing digital landscape.** In the next section we spotlight key international strategies that deal specifically with child sexual exploitation and abuse (CSEA)

4.8 CSEAM

Online child sexual exploitation and abuse is a persistent threat with devastating consequences for those affected. **Child sexual exploitation and abuse (CSEA)** encompasses a range of different behaviours, including sharing of **child sexual exploitative and abuse material (CSEAM)**, online grooming of children - which can lead to coercing a child to send sexual images of themselves, sexual extortion, or arranging in-person child sexual abuse of child victims.

For content explicitly falling under CSEAM, regulators endorse a proactive monitoring approach, facilitating bridges between law enforcement and platforms to effectively deliver grievance redressal as well as track the perpetrators. **The UK's Online Safety Act has set out duties so that online services must carry out risk assessments to understand the likelihood and impact of child sexual exploitation and abuse (CSEA) appearing on their service.**[425] They must also take steps to mitigate the risks identified in their risk assessment and to identify and remove illegal content where it appears. The higher the risk on a service, the more measures and safeguards they will need to take

[424] Dr. Philippa Smith, *Independent Review: Aotearoa New Zealand Code of Practice for Online Safety and Harms Transparency Report,* 2024, https://thecode.org.nz/wp-content/uploads/sites/38/2024/01/The-Code-Independent-Review-Report_31_January-2024.pdf.

[425] Ofcom, *Tackling child sexual abuse under the online safety regime*, 2024, https://www.ofcom.org.uk/online-safety/protecting-children/tackling-child-sexual-abuse-under-the-online-safety-regime.

to keep their users safe from harm, and prevent their services being used as a platform to groom and exploit children.

Notably, in 2023 OfCom carried out an Illegal Harms Consultation[426] towards suggesting codes of practice that services can adopt in order to protect children from CSEA. The strategies discussed included the use of:

- **Hash-matching technologies**, which automatically detect known CSEAM images shared by users in their public content. For the codes of practice, this was not intended to apply to private or end-to-end encrypted communications. An example of hash-matching technology being used to remove nude images of minors is the service "**Take it Down**", a free tool developed by the National Center for Missing and Exploited Children (**NCMEC**). Minors can use this tool anonymously to remove nude/semi nude images, or prevent them from being posted at all.[427]

- **URL detection technologies** which scans public posts to remove illegal URLs that lead to material depicting the abuse of children.

- **Applying warning messages on search services** when users search for content that explicitly relates to CSEAM.

- **Measures to tackle the online grooming of children**, including safer default settings that make it **harder for strangers to find and interact with children online**.

- **Supportive prompts and messages for child users** during their online journey, to empower them to make safe choices online, such as when they turn off default settings or receive a message from a user for the first time.

The **European Union's strategy 2020-2025[428]** outlines a holistic approach to address the increasing risks of child sexual abuse both offline and online. To that end, the European Commission proposed in May 2022 a draft Regulation laying down rules to prevent and combat child sexual exploitation and abuse (CSEAM Regulation).[429] The CSEAM Regulation includes[430] a significant number of elements such as:

- The creation of an **EU Centre to coordinate the fight against CSEAM in Europe**,

- An obligation for hosting service providers and those providing inter-personal communications services to **carry out risk assessments,**

- Subsequently take **risk mitigation measures** to tackle the potential use of their services for the exchange of CSEAM, and

- The **recognition of orders from judicial authorities** to fight against CSEAM, such as **detection, removal, blocking and delisting orders**.

To ensure progress in reporting requirements under the framework, the Commission developed a **uniform standard** for the region on data to be included in reports by companies and organizations combatting online child sexual abuse. **The form would require detailed reporting on nine primary categories, including the type and volumes of data processed, retention policies and safeguards, the specific grounds relied on for processing, and transfers under the GDPR, and the number of cases of CSEAM identified**.

Data points for these categories require, for instance, the number of bytes of text processed to detect online grooming in relation to non-EU users, the average time needed to make the decision to restore or keep the suspension of user accounts in the EU, error rates in automatic flagging of user accounts, etc.

---

[426] Ofcom, Consultation: *Protecting people from illegal harms online*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/protecting-people-from-illegal-content-online.

[427] Justin W. Patchin, Take it Down: A new tool to combat the unauthorized sharing of explicit images of minors, 2023, https://cyberbullying.org/take-it-down.

[428] European Union, *Strategy 2020 - 2025*, 2020, https://pro.europeana.eu/files/Europeana_Professional/Publications/EU2020StrategyDigital_May2020.pdf.

[429] European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse*, 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0209#:~:text=It%20seeks%20to%20provide%20legal,general%20principles%20of%20EU%20law.

[430] Inhope, *the European Union CSEAM Regulation*, 2023, https://inhope.org/EN/articles/the-european-union-csam-regulation.

Similar to India, European authorities mainly rely on the US to receive reports of OCSEA. This is because the US National Center for Missing and Exploited Children (NCMEC) and applicable US legislation requires service providers to report such content to NCMEC, which then forwards EU-related reports to relevant LEAs. **At the moment, there is no central entity in the EU that service providers can send their reports to. As a result, reports of abuse in the EU are sent to the US and then back to EU law enforcement agencies. The proposed EU centre aims to simplify this process.** The EU centre will receive the reports of child sexual abuse online, make sure that they are not false positives and distribute them to national law enforcement agencies in Member States. The initiative aims to reduce the amount of time that law enforcement spends on filtering reports that are not actionable. It also aims to mitigate the incidence of false positives and erroneous reporting to LEAs, thereby reducing the administrative and financial burden on said providers.

Parallelly, Eurojust (European Union Agency for Criminal Justice Cooperation)[431] supports judicial cooperation between Member States to facilitate the prosecution of child sexual abuse perpetrators in crossborder cases. The European Commission funds and supports several initiatives and networks, including the **Better Internet for Kids portal**[432], raising awareness of the potential risks children face online, and **INHOPE**[433], a network of hotlines combating online child sexual exploitation and abuse material by analysing illegal content.

The Home Office in the UK launched Outcome 21 in 2016 for policing self-generated sexual imagery among teens, colloquially known as "sexting". Outcome 21 is a code that provides an option for the police to record that no formal criminal justice action would be taken after police investigation concerning a case of sexting, as such action would not be in public interest. **Outcome 21 enables the police to conclude their investigation without issuing punitive action against children, as often children indulge in sexting consensually, without being aware of the potential dangerous ramifications of self generated sexual imagery. Outcome 21 was intended to allow police forces to issue a proportionate response to sexting without criminalising children. It is important to note that Outcome 21 is to be considered an appropriate solution only in those cases of youth generated sexual imagery where no evidence of abuse, exploitation, inappropriate sharing or aggravating factors is found after investigation.**[434] Legitimate cases of peer on peer abuse, or harmful sexual behaviour fall outside the purview of Outcome 21.

In the Philippines, formal reporting mechanisms for OSCEA cases include hotlines, websites, calls and text-based channels. eProtectKids is Philippines' internet based hotline against CSAM. It is a global hotline that enables takedown of CSAM content hosted both nationally and internationally, and facilitates further investigation. Bantay Bata 163 is a toll free 24x7 helpline service for children. **It receives CSAM reports and forwards them to Philippines' law enforcement groups dedicated to cybercrime for report verification and victim identification.** Victims are then offered counselling, legal support and referral to follow up on their case.[435] **In South Africa, formal reporting mechanisms for CSAM include the Childline South Africa Helpline (member of Child Helpline International), which provides phone counselling, online counselling and information services.** The online counselling services provided

[431] European Parliamentary Research Service, *Combating child sexual abuse*, 2024, https://www.europarl.europa.eu/RegData/etudes/BRIE/2024/757611/EPRS_BRI(2024)757611_EN.pdf.

[432] European Commission, *A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+)*, 2022, https://better-internet-for-kids.europa.eu/en/bik.

[433] Inhope, *European Commission Funded Initiative*, 2020, https://inhope.org/EN/articles/inhope-a-european-commission-funded-initiative-all-about-collaboration.

[434] United Kingdom GD8, *Youth produced sexual imagery: Guidance for disclosure*, 2016, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/578979/GD8_-_Sexting_Guidance.pdf.

[435] ECPAT, INTERPOL & UNICEF, *Disrupting harms in the Philippines: Evidence on Online child sexual exploitation and abuse*, 2022, https://www.unicef.org/philippines/media/7396/file/Disrupting%20Harm%20in%20the%20Philippines.pdf.

by Childline in South Africa also cater to children with speech and hearing disabilities. Most police stations are equipped with victim empowerment centres with child friendly facilities.[436]

Overall what we observe is that internationally governments are investing in facilitating defined standards and protocols for engagement with industry, while simultaneously investing in LEA and judicial capacity, better coordination across institutions, sensitised enforcement practices and more accessible complaints and redressal mechanisms. In the next section we look at how countries are also investing in rehabilitation for child victims of online abuse.

4.9 Rehabilitation and Efforts to Limit Revictimisation

Traditionally, online risk mitigation strategies largely focus on punishing perpetrators. However, regulators now recognise the need for support and rehabilitation among child-victims to move past abusive experiences and regain confidence to reintegrate into online activities.

Australian initiatives emphasize a survivor-centered, trauma-informed approach. The eSafety Commissioner notes that its assistance helps victims *"regain control over their situation"* and feel safer,[437] showcasing crucial elements of trauma-informed care. Victims are empowered to make decisions, such as whether to involve law enforcement. Safety features on the eSafety website, like the "quick exit" button bolster confidentiality in a trauma-informed approach. The coordinated model has shown measurable improvements: harmful content is often removed within hours of a report, minimizing trauma duration, and eSafety boasts a ~90% success rate in takedowns[438]. This rapid response is associated with real relief for victims – *"making a real difference to their distress,"* according to the Commissioner[439]. **Australia's cumulative approach of enforcing a strong legal mandate and facilitating a one-stop national help agency is widely seen as an effective model in improving victim safety and access to justice.**

In the **UK**, *Childline* provides counseling for cyberbullying or grooming,[440] and law enforcement works with the *Internet Watch Foundation* to rapidly remove child sexual abuse material, thus limiting revictimization.[441] **Ireland** boasts a unique multi-stakeholder approach to non-state interventions. The Irish Safer Internet Centre[442] is a partnership between four organisations that collectively provides awareness materials and programmes for schools, a public hotline to report suspected illegal content or behaviour and programmes to enhance youth participation in developing online safety solutions.

Victims of online harms in the Philippines have access to comprehensive services often integrated with existing anti-trafficking and Violence Against Women (VAW) programs. Under RA 11930, child victims of online sexual abuse are entitled to *"immediate rescue and rehabilitation,"* including provision of **shelter, counseling, legal and medical assistance**[443]. The Department of Social Welfare and Development (DSWD) operates temporary shelters

---

[437] Australia eSafety Commissioner, *Insights from eSafety's image-based abuse reporting and removal scheme*, 2025, https://www.esafety.gov.au/research/insights-from-esafetys-image-based-abuse-reporting-and-removal-scheme#:~:text=under%20the%20scheme%20and%20eSafety%E2%80%99s,responses%20to%20reports.

[438] Monash University, *Reports of "revenge porn" skyrocketed during lockdown. We must stop blaming victims for it*, 2020,https://lens.monash.edu/@politics-society/2020/06/04/1380604/reports-of-revenge-porn-skyrocketed-during-lockdown-we-must-stop-blaming-victims-for-it#:~:text=If%20you%E2%80%99re%20a%20victim%20of,helplines%20such%20as%201800%20RESPECT.

[439] The Straits Times, *Australia's online safety regulator responds fast to remove harmful content*, 2023, https://www.straitstimes.com/tech/australia-s-online-safety-regulator-responds-fast-to-remove-harmful-content.

[440] NSPCC Learning, *Childline*, https://learning.nspcc.org.uk/services/childline.

[441] Internet Watch Foundation, Data That Drives Change: IWF 2024 Annual Data & Insights Report, 2024, https://www.iwf.org.uk/.

[442] *Irish Safer Internet Centre*, https://better-internet-for-kids.europa.eu/en/sic/ireland.

[443] Respicio & Co, *The Anti-Online Sexual Abuse or Exploitation of Children and Anti-Child Sexual Abuse or Exploitation Materials (CSAEM) Act [R.A. No. 11930, July 30, 2022] | Special Penal Laws*, 2025, https://www.respicio.ph/bar/2025/criminal-law/special-penal-laws/the-anti-online-sexual-abuse-or-exploitation-of-children-osaec-and-anti-child-sexual-abuse-or-exploitation-materials-csaem-act-ra-no-11930-july-30-2022#:~:text=The%20law%20mandates%3A.

and recovery programs for children rescued from online exploitation, where they receive therapy and education. There are dedicated **hotlines** such as *1343* (for reporting human trafficking and OSAEC) provide initial counseling and can refer victims to shelters or Women and Children Protection police units. The *Philippines National Police* and *National Bureau of Investigation* have specialized cybercrime divisions. For example, the PNP's Women and Children Protection Center launched the **Philippine Internet Crimes Against Children Center (PICACC)** in 2019[444]. This center improves cross-border investigation of online child abuse and coordinates victim identification and safeguarding – an innovative international collaboration (with Australia and the UK) that has led to more children being rescued and less reliance on victims having to testify, thereby reducing trauma. **On the civil society side, organizations like *International Justice Mission (IJM)* and *Child Rights Network*** assist law enforcement with victim aftercare, ensuring survivors get long-term counseling, livelihood training for older teens, and reintegration support.

Overall, a general trend that emerges across jurisdictions on rehabilitation policies for children involves collaboration between civil society and law enforcement. While the efforts of law enforcement are usually geared more towards identification, removal and penalisation of abusive content and offenders, CSOs offer survivor centric support mechanisms such as victim support helplines, counselling, legal aid and therapy.

Next, our research spotlights international examples of institutional involvement of teens and adolescents in tech policy and regulation.

4.10 Inclusive Institutional Design and Youth Engagement

Children and youth use social media to engage in unique ways, and are often exposed to content specifically targeted towards shaping their online and offline experiences. In this final section of this chapter, we spotlight examples where international regulators and institutions are formally including children's voices in policy discussions. **First, in April 2022 Australia's eSafety Commissioner set up an online safety youth advisory council[445]** to provide young people a voice about online safety policy. Members of the eSafety Youth Council are **aged 13 to 24 years** and are from a diverse range of experiences, genders, cultural and linguistic backgrounds, and locations. The eSafety Youth Council provides advice to the Government about issues that young people experience online and explores ways of supporting them to have positive online experiences. **Their insights also inform the eSafety Commission's youth policies and programs.** The operation of the Council has **been informed by recommendations included in a Youth engagement report, commissioned by the Australian regulator.** The report explored the online experiences of young Australians and their concerns and ideas for the future. **Second, the UK's youth parliament engages young people** to shape national policies and has made significant contributions to shaping laws on online safety.[446] The Scottish Youth Parliament is conducting a youth-led project over a two year period to amplify young people's views on improving online safety for children in Scotland. **Third, UNICEF's Innocenti Global Youth Network** serves as a global platform for youth perspectives through participatory research and foresight analysis, besides equipping young people to conduct research through youth fellowships.[447] The research conducted by young people informs insights of UNICEF reports, and helps co-create UNICEF's Leading Minds Conference Series.[448]

Through inclusive regulation, children are able to contribute to an Internet where they can exist with autonomy, agency and safety.

---

[444] Beh Lih Yi, UK, *Australian police help Philippines fight child cybersex trafficking*, 2019, https://www.reuters.com/article/world/uk-australian-police-help-philippines-fight-child-cybersex-trafficking-idUSKCN1QG1PK/#:~:text=At%20the%20Philippine%20Internet%20Crimes,border%20abuse%20and%20protect%20children.

[445] Australia eSafety Commissioner, *eSafety Youth Council*, https://www.esafety.gov.au/young-people/esafety-youth-council.

[446] Hayley Clarke, *Social media ban not practical or effective, teens say*, 2025, https://www.bbc.com/news/articles/c8x40qplk15o

[447] UNICEF, *Youth engagement: a world where young people can create their futures*, https://www.unicef.org/innocenti/innocenti/approach/youth-engagement.

[448] UNICEF, *Leading Minds: Young people changing the world,* https://www.unicef.org/innocenti/innocenti/leading-minds.

4.11 Snapshot Summary of International Trends on Children's Online Safety

This chapter demonstrates how countries have dealt with online safety for children through a holistic set of interventions including safe platform design, age appropriate experiences, definitions for targeted harm identification, investments in enforcement and rehabilitation, participative dialogue and codes of practices. The following table is meant to represent a snapshot of these trends for quick reference.

Table 4.3 Global Trends in regulating online harms against children

Themes	Country	Legislation / Tool	Takeaways
Age Gating / Age Assurance	**UK**	**Online Safety Act**	Age assurance for children's access to platforms to be determined on a **best effort basis**
	US	**California's AADC**	Requires age estimation with a **reasonable level of certainty** appropriate to the risks that arise from the data management practices of the business **Challenges**: Accuracy, Privacy, and Circumvention
Risk Assessments	**EU**	**Digital Services Act**	Calls for a **broad-based assessment of systemic risks** including any actual or foreseeable negative effects in relation to the protection of minors
	UK	**Online Safety Act**	Requires platforms to ensure that child users have **age-appropriate experiences** and are shielded from harmful content
Specificity of Definitions	**UK**	**Online Safety Act**	Defines specific categories of illegal content like **Primary Priority Content** (e.g., child abuse) and **Priority Content** (e.g., bullying), ensuring platforms know exactly what to target.
	Philippines	**Anti-Online Sexual Abuse or Exploitation of Children**	Criminalizes specific acts like **online grooming** and **live-streaming sexual abuse**, making the legal framework clear and improving enforcement outcomes
Legal but Harmful Acts	**UK**	**Online Safety Act**	Addresses content that is **harmful to children but not illegal**, such as extreme diet promotion
Product Design	**UK**	**Children's Code**	**42 changes (eg. privacy by default etc) across 4 platforms in 2021** after the enactment
	New Zealand	**Code of Practice for Online Safety and Harms**	Led to **context-specific transparency reporting**
Participative Policymaking	**UK**	**Youth Parliament**	Young people shape national policy
		Digital Futures Commission	Youth actively involved in research and policy
	Australia	**eSafety Youth Council**	Youth directly advise on online safety

Chapter 5
International Trends on Online Safety Interventions Directed Towards Women

5.1. Introduction

Women constitute one of the most vulnerable groups facing online harms. While digital platforms have opened new opportunities for women's empowerment, TFGBV is a systemic problem and undermines women's ability to take advantage of these spaces fully. According to the Institute of Development Studies, around 16% to 58% of women have faced TFGBV, underscoring its pervasive nature.[449] This is further supported by the Economist Intelligence Unit's study which revealed that 38% of women have personally experienced online abuse and 85% of women have witnessed such violence against other women.[450] As we have profiled in detail in Chapter 2 of this report, the study also highlighted that common forms of TFGBV including misinformation and defamation (67%), cyber harassment (66%), hate speech (65%), impersonation (63%), hacking and stalking (63%), and video or image-based abuse (57%). Given the disproportionate risks faced by women online, and

[449] Jacqueline Hicks, *Global evidence on the prevalence and impact of online gender-based violence*, 2021, https://www.ids.ac.uk/publications/global-evidence-on-the-prevalence-and-impact-of-online-gender-based-violence-ogbv/

[450] The Economist, *Measuring the prevalence of online violence against women,* 2021, https://onlineviolencewomen.eiu.com/

in order to ensure that the internet remains a positive avenue for upward social mobility, online safety becomes very important. This chapter examines how even globally online safety is a gendered issue and then offers an overview of how other jurisdictions are attempting to address TFGBV.

5.1.1. The gendered nature of harms

Online harms and violence against women are rooted in structural sexism: they exploit social norms that police women's speech, appearance, and autonomy, and they are intensified by platform designs that reward virality and anonymity (and the internet's networked realities) without adequate safeguards. Women are further victimised based on negative gender stereotypes, through various forms of violence like pornography, sexist games and breaches of privacy.[451] Online violence is a complex multifaceted problem, in as much as it is interrelated to other forms of gender violence. This complexity must be acknowledged by those responsible for addressing TFGBV. **For instance, as we have alluded to in Chapter 2 as well, the 'Gamergate' online harassment campaign that targeted women in the video game industry during 2014–2015, revealed how female gamers experienced anxiety and loneliness due to lack of social support, and this mirrored their experiences with social support outside of gaming - demonstrating the interconnectedness and pervasive nature of TFGBV.[452] Women also bear the additional psychological burden associated with safety concerns. Addressing TFGBV therefore requires a multidisciplinary approach and the participation of various stakeholders.[453]**

5.1.2. Intersectionality in online harms

As discussed previously in Chapter 2, women with multiple identities are targeted based on an interplay of their characteristics, such as caste, race, gender identity, ethnicity, age, and abilities, among others. **The discrimination and violence they experience are therefore intersectional, resulting in more severe consequences.[454] Women belonging to ethnic minorities, sexual minorities, gender minorities, women with disabilities, and women from other marginalized groups are particularly targeted.[455] Even women in public roles, like human rights defenders, journalists, and politicians are at heightened risk of online harms and are subjected to online threats, which are misogynistic and sexualized.** A 2018 study titled '*Sexism, Harassment and Violence against Women in Parliaments in Europe*' revealed that 58.2% of women parliamentarians had been targets of sexist attacks online. Further, **Amnesty International's research corroborates this figure by showing that women politicians are 27 times more likely to face abuse online than their male counterparts.[456]** This is further worsened for women from marginalized communities, with black women being 84% more likely to be referred to in abusive posts than their white counterparts.[457]

---

[451] United Nations Office of the High Commissioner for Human Rights, *Report of the Working Group on the issue of discrimination against women in law and in practice - A/HRC/23/50, 2013*, http://daccess-ods.un.org/access.nsf/Get?Open&DS=A/HRC/23/50&Lang=E.

[452] Lavinia McLean and Mark D. Griffiths, *Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study,* 2018, https://link.springer.com/article/10.1007/s11469-018-9962-0.

[453] UNFPA and eSafety Commissioner (Australia), *A Framework for TFGBV Programming*, 2024, https://www.unfpa.org/sites/default/files/pub-pdf/A%20Framework%20for%20TFGBV%20Programming.pdf?utm

[454] United Nations Office of the High Commissioner for Human Rights, *A/HRC/38/47: Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective,* 2018, https://www.ohchr.org/en/documents/thematic-reports/ahrc3847-report-special-rapporteur-violence-against-women-its-causes-and.

[455] United Nations Office of the High Commissioner for Human Rights, *Promotion, protection, and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective - Report of the United Nations High Commissioner for Human Rights* - A/HRC/35/9, https://documents.un.org/doc/undoc/gen/g17/111/81/pdf/g1711181.pdf.

[456] Commonwealth Parliamentary Association UK, *Online violence against women parliamentarians hinders democracy, and all parliamentarians are responsible for addressing it,* 2021, https://www.uk-cpa.org/news-and-views/online-violence-against-women-parliamentarians-hinders-democracy-and-all-parliamentarians-are-responsible-for-addressing-it.

[457] Commonwealth Parliamentary Association UK, *Online violence against women parliamentarians hinders democracy, and all parliamentarians are responsible for addressing it,* 2021, https://www.uk-cpa.org/news-and-views/online-violence-against-women-parliamentarians-hinders-democracy-and-all-parliamentarians-are-responsible-for-addressing-it.

This creates a 'chilling effect' on women who hold such positions, forcing them to self-censor, which adversely impacts their professional lives. According to UNESCO's Report[458] on online violence against women journalists, the safety and security concerns associated with such threats overlap with the curbs on freedom of expression and freedom of press and are often inseparable.[459] This includes online harassment and abuse, and can further include threats of physical or sexual violence, privacy breaches exposing identifying information, coordinated disinformation campaigns, and hate speech. Notably, the most common consequence reported by respondents was the significant mental health impact stemming from such abuse. **These experiences severely undermine their ability to perform their jobs to their full potential.[460]** Online abuse and violence therefore represent a threat to not only women's safety online but also to women's civic participation.[461] **This can also result in far-reaching consequences in women's civic participation with online abuse being recognized as a future barrier to women's representation in public service roles in the United Kingdom[462], Sweden[463] and Finland[464], including having the potential of driving women out of politics[465].** This intersectional nature of TFGBV faced by women in public roles has garnered attention from international and regional regulatory bodies, including the United Nations[466] and the Council of Europe[467].

5.2. International Law and Principles governing TFGBV

Certain core international law instruments govern women's rights and can be extended to protecting their rights around online safety as well. These include the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)[468], the Declaration on the Elimination of Violence against Women[469], and the Beijing Declaration and Platform for Action[470] - **all of which predate the development of ICT and the emerging forms of online violence against women**. However, the emerging risks women face online have been acknowledged and **progressively analysed by the**

---

[458] UNESCO, *Online violence against women journalists: a global snapshot of incidence and impacts*, 2020, https://unesdoc.unesco.org/ark:/48223/pf0000375136.

[459] UNESCO, *Safety of Women Journalists*, https://www.unesco.org/en/safety-journalists/safety-women-journalists.

[460] Inter-Parliamentary Union, *Sexism, harassment and violence against women in parliaments in the Asia-Pacific region,* 2025, https://www.ipu.org/resources/publications/issue-briefs/2025-03/sexism-harassment-and-violence-against-women-in-parliaments-in-asia-pacific-region.

[461] United Nations Office of the High Commissioner for Human Rights, *A/HRC/38/47: Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective,* 2018, https://www.ohchr.org/en/documents/thematic-reports/ahrc3847-report-special-rapporteur-violence-against-women-its-causes-and.

[462] Harmer, et al., *Digital microaggressions and everyday othering: an analysis of tweets sent to women members of Parliament in the UK*, 2021, https://www.tandfonline.com/doi/full/10.1080/1369118X.2021.1962941.

[463] Erikson, et al., *Three dimensions of gendered online abuse: Analyzing Swedish MPs' experiences of social media*, 2023, https://www.cambridge.org/core/journals/perspectives-on-politics/article/three-dimensions-of-gendered-online-abuse-analyzing-swedish-mps-experiences-of-social-media/F52E7389E355C1C78335B44B9E66811E.

[464] M Mannevuo, *Uneasy self-promotion and tactics of patience: Finnish MPs' ambivalent feelings about personalised politics on social media*, 2023, https://journals.sagepub.com/doi/10.1177/13678779221120028.

[465] Tom Felle, *Online abuse could drive women out of politics*, 2023, https://www.ips-journal.eu/topics/democracy-and-society/online-abuse-could-drive-women-out-of-politics-7036/.

[466] UN Women, *Preventing Violence Against Women in Politics*, 2021, https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2021/Guidance-note-Preventing-violence-against-women-in-politics-en.pdf.

[467] Council of Europe, , *No space for violence against women and girls in the digital world,* 2021, https://www.coe.int/en/web/commissioner/-/no-space-for-violence-against-women-and-girls-in-the-digital-world.

[468] The Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW) is an international treaty adopted in 1979 by the UN General Assembly. Described as an *international bill of rights for women*, it requires UN member states to undertake legal obligations to protect and fulfill women's rights and promote gender equality. This treaty has currently been ratified by 189 member states. See: https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women.

[469] The Declaration on the Elimination of Violence Against Women is a resolution adopted by the UN General Assembly in 1993, aimed at eradicating violence against women and girls worldwide. It laid the authority on defining violence against women as *"any act of gender-based violence that results in physical, sexual, or psychological harm"*. See: https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-elimination-violence-against-women.

[470] The Beijing Document was adopted by the UN at the Fourth World Conference on Women in 1995. It is a global agenda for achieving gender equality, considered the most comprehensive and transformative plan for these goals. The document outlines 12 critical areas of concern, including poverty, education, violence, and political participation, and provides a framework for governments to implement policies and programs. See: https://www.unwomen.org/en/digital-library/publications/2015/01/beijing-declaration.

UN Committee on the Elimination of Discrimination against Women in several general recommendations. While recognising the role of online spaces in empowering women[471], the Committee has clarified that the CEDAW applied to digital spaces as well, since those have become sites for violence against women and girls.[472] In addition, the Committee has specifically highlighted how cyberbullying impacts girls in their right to access education.[473] **The Committee recommended that States encourage private businesses and transnational corporations to take appropriate measures to eliminate all forms of violence against women.[474]**

More recently, the UN General Assembly released a report in 2022, titled ***Intensification of Efforts to Eliminate All Forms of Violence Against Women and Girls***[475], identifying that forms and patterns of online facilitated violence against women will continue to evolve with the evolution in technology, e.g. *zoombombing*[476], metaverse harassment[477], etc., especially with the increasing adoption of AI. It draws a correlation between such emerging harms and specific features of digital platforms that create a particularly conducive environment for violence against women, such as the scale, speed and ease of online communication combined with anonymity and pseudonymity. **To ensure harm mitigation keeps pace with its evolving forms, the report identifies four key action points:**

- ***Updating legal frameworks*** *to address online violence against women and* ***equating it with physical violence****;*
- *Bringing ecosystem-level changes by* ***urging intermediaries to design products*** *and services in a* ***safe, accessible and representative manner****;*
- *Establishing a* ***common methodology to guide data collection and promoting platform transparency*** *to inform early detection and warning systems; and*
- *Promoting* ***greater collaboration*** *between technology/ communication companies, civil society, governments and experts.*

Even at the regional level, the Council of Europe's ***Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention)*** urges its Member States to encourage private companies to prevent violence against women and educate users on tackling harmful online content.[478] The Istanbul Convention is one of the most comprehensive human rights instruments recognising violence against women (including technology and ICT-enabled violence) as violence occurring because of their gender and establishes the state's obligation to address it.[479]

---

[471] UN Committee on the Elimination of Discrimination against Women, *General Recommendation No. 33 on Women's Access to Justice,* 2015, https://digitallibrary.un.org/record/807253?ln=en&v=pdf.

[472] UN Committee on the Elimination of Discrimination against Women, *General recommendation No. 35 on gender-based violence against women*, 2017, https://digitallibrary.un.org/record/1305057?ln=en&v=pdf.

[473] UN Committee on the Elimination of Discrimination against Women, *General recommendation No. 36 on the right of girls and women to education*, 2017, https://digitallibrary.un.org/record/3843534?v=pdf.

[474] UN Committee on the Elimination of Discrimination against Women, *General recommendation No. 35 on gender-based violence against women*, 2017, https://digitallibrary.un.org/record/1305057?ln=en&v=pdf.

[475] UN Secretary General, *Intensification of Efforts to Eliminate all Forms of Violence Against Women and Girls - A/77/302*, 2022, https://digitallibrary.un.org/record/3988297?v=pdf.

[476] Zoombombing is a term for when unauthorized users disrupt video conferencing meetings, often on platforms like Zoom, by sharing inappropriate content, broadcasting disturbing or explicit imagery or disrupting the meeting flow. These intruders can be individuals or bots, and they often exploit vulnerabilities like publicly shared meeting IDs or weak security settings. See: University of Georgia, *Preventing, Managing & Recovering from Zoombombing*, https://ctl.uga.edu/teaching-resources/teaching-amid-disruption/preventing-managing-and-recovering-from-zoombombing/

[477] Metaverse harassment is the sexual violence and abuse that female-presenting avatars in the metaverse are subjected to. Experiencing these acts in the metaverse can be deeply distressing for victims, with emotional and psychological responses resembling reactions to incidents that happen in the physical world, with studies showing that metaverse harassment can have real-world impact as well. See: Carlotta Rigotti and Gianclaudio Malgieri, *Sexual violence and harassment in the Metaverse: A new manifestation of gender-based harms*, 2024, https://equalitynow.storage.googleapis.com/wp-content/uploads/2024/04/26125953/EN-AUDRi-Sexual-violence-and-harassment-in-the-metaverse-03.pdf; Equality Now, *Sexual violence in the metaverse has a real-world impact on victims,* 2024, https://equalitynow.org/press_release/sexual-violence-in-the-metaverse-has-a-real-world-impact-on-victims/

[478] Council of Europe, *The relevance of the Istanbul Convention and the Budapest Convention on Cybercrime in addressing online and technology-facilitated violence against women,* 2021, https://rm.coe.int/the-relevance-of-the-ic-and-the-budapest-convention-on-cybercrime-in-a/1680a5eba3.

[479] Centre for Communication Governance, *CCG-NLUD comments to the MeitY's proposed amendments to the 2021 IT Rules*, 2021, https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccgnlud-comments-draftamendments-itrules2021-6jul22-303.pdf.

Building on the framework, the **Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO),** an independent expert body responsible for monitoring the implementation of the Istanbul Convention, issued its ***General Recommendation No. 1 on the Digital Dimension of Violence Against Women*** in 2021.[480] This recommendation significantly advances the Convention's scope by defining how its obligations apply in the digital context and calls on states for **specific legislative reform to criminalise new forms of TFGBV, capacity-building for law enforcement to secure digital evidence, platform accountability through reporting tools and moderation obligations, and integration of online harms into victim support systems**. Together, the Istanbul Convention and GREVIO's recommendation demonstrate how internationally States are developing **gender-sensitive policies** that reflect specific types of online violence against women and their impact on the victim/survivor.

Predating this, in 2018, the **UN Special Rapporteur on Violence Against Women and Girls (VAWG)** released a report analysing the causes and consequences of online violence against women and girls from a human rights perspective.[481] The Report identified a **few key areas governing state's obligations in preventing online harm:**

- ***Prevention***: This includes the obligation of states to provide adequate information and awareness on online violence against women and girls, including the availability of legal entitlements and services, and to take steps to prevent human rights violations committed by internet intermediaries.
- ***Protection***: This includes the state's obligation to establish procedures to ensure immediate removal of harmful content, timely judicial intervention, prompt action by internet intermediaries, and providing accessible legal services for survivors.
- ***Sensitive and Accessible Prosecution*:** LEAs that engage in insensitive victim-blaming attitudes contribute to underreporting by survivors and a culture of silence. Further, technical and financial barriers also hinder the survivor's right to access justice effectively. **States need to ensure access to justice and remedies for survivors both at the time of investigation and institution of proceedings against perpetrators.** This entails the state providing efficient, sensitive, and well-trained first responders, including internet intermediaries, police, helplines, judiciary and regulators.
- ***Punishment*:** The state needs to ensure sanctions are put in place as punishment for online violence against women, to convey the message that such violence will not be tolerated. This is especially important for survivors who are scared to speak up or the impunity enjoyed by their perpetrators.
- ***Redress, reparation, and remedies*:** Remedies must therefore include a variety of material, individual, collective, and symbolic measures tailored to the needs and circumstances of the survivor. This includes the removal of the alleged content and injunction to prevent the publication of harmful content, coupled with other forms of **restitution, rehabilitation and guarantees of non-repetition**.

The UN Special Rapporteur Report also outlined some key recommendations to the State parties, including:

- to recognize online violence against women as a human rights violation and a form of GBV;
- to enact laws to address new emerging forms of online GBV and to adapt appropriate criminal and civil causes of action;
- to implement measures to prevent publication of harmful material and ensure their prompt removal;

[480] Council of Europe, *GREVIO General Recommendation No. 1 on the digital dimension of violence against women,* 2021, https://rm.coe.int/grevio-rec-no-on-digital-violence-against-women/1680a49147.

[481] United Nations Office of the High Commissioner for Human Rights, *A/HRC/38/47*: *Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective,* 2018, https://www.ohchr.org/en/documents/thematic-reports/ahrc3847-report-special-rapporteur-violence-against-women-its-causes-and.

- to prohibit and criminalise online violence against women;
- to apply a gender perspective to all online forms of violence;
- to provide training for magistrates, lawyers, police, and all other law enforcement officials and frontline workers;
- to develop **clear and transparent internal and external protocols and codes of conduct for law enforcement officials addressing TFGBV**;
- to provide protective measures and services for victims of TFGBV, including helplines, shelters and protection orders;
- to provide reparation measures not limited to compensation but also include forms of restitution, rehabilitation, satisfaction and guarantees of non-repetition, depending on the circumstances and the preferences of the victim.

The Report also made key recommendations to intermediaries, including adopting transparent and accessible complaint mechanisms, policies and procedures for reporting and requesting removal of harmful content; publishing clear content moderation policies and human rights safeguards against arbitrary censorship and appeal processes; and ensuring that terms of service and reporting tools are accessible, user-friendly and provided in local languages.

In addition to this, the **Global Partnership for Action on Gender-Based Online Harassment and Abuse (the Global Partnership)** is a coordinated and interdisciplinary action to address TFGBV which was launched in 2022 by the US and Danish governments. The Global Partnership is now a 12-country coalition and is working alongside a multi-sectoral advisory group that brings evidence-based and coordinated solutions, principles and policies to address the issue of TFGBV.[482] The Global Partnership calls on countries and partners to advance three goals: create shared principles that hold perpetrators and platforms accountable and recognize online abuse as a human rights issue; and expand programs, resources, and training to prevent and respond to such abuse. It also seeks to strengthen data collection, research, and platform transparency to better understand the prevalence and impacts of gender-based online harassment and abuse.[483]

5.3. Different regulatory approaches to the regulation of TFGBV

5.3.1. Individual Harms v. Systemic Harms: A comparison between the UK OSA and the EU DSA

Users face several difficulties while addressing TFGBV and are forced to fit their experiences to predetermined categories while reporting online content. This fails to account for the multifaceted nature of online abuse. Moreover, content moderators are not provided with gender-sensitive training and as a result complaints are often incorrectly adjudged to not violate community standards, leaving harmful content unaddressed.[484]

To address these gaps, the ***UK's Online Safety Act* (OSA)** and ***EU's Digital Services Act* (DSA)** have adopted two different regulatory approaches to co-regulating platforms. As we have discussed in Chapter 4 as well, the way *'harm'* has been defined in both legislations varies significantly, both on a theoretical and practical basis. **While the OSA defines harm in a narrow sense constituting physical or psychological harms arising out of specific actions, the DSA takes a more broad approach to harms, covering harms to individuals and society at a systemic level rather than individual actions**.

[482] United Nations Population Fund, *2022 Global Symposium on Technology-facilitated Gender-based Violence Results: Building a Common Pathway*, 2023, https://www.unfpa.org/sites/default/files/pub-pdf/2022-GlobalSymposium-TFGBV_EN.pdf?utm_source=chatgpt.com.

[483] White House Gender Policy Council, *Launching the Global Partnership for Action on Gender-based Online Harassment and Abuse*, 2022, https://bidenwhitehouse.archives.gov/gpc/briefing-room/2022/03/18/launching-the-global-partnership-for-action-on-gender-based-online-harassment-and-abuse/?utm_source=chatgpt.com.

[484] World Wide Web Foundation, *The impact of online gender-based violence on women in public life,* 2025, https://webfoundation.org/2020/11/the-impact-of-online-gender-based-violence-on-women-in-public-life/.

The UK OSA stirred much controversy as the draft bill did not have any mention of women or girls.[485] **It was only pursuant to sustained campaign efforts led by the End Violence Against Women Coalition and other civil society groups that the extant provision under the OSA dealing with harm against women and girls came in**. However, the focus of harm remains limited to individualised harms which are defined as '*physical or psychological harm*' under Section 234 of the OSA. **Critics say that the definition neglects the broader systemic harms impacting society, especially combating online violence against women and girls.[486]** As a result, while the UK's OSA obliges platforms to remove material that promotes hate crimes, **misogynistic content on its own is treated as "legal but harmful"** and therefore falls outside the Act's scope.[487]

However, in February 2025, **Ofcom issued a Call for Evidence and released draft guidance on *A Safer Life Online for Women and Girls*[488],** which **proposes that platforms conduct gender-informed risk assessments** to investigate how features like algorithmic amplification, anonymity, or recommendation systems may facilitate abuse against women. It also calls for targeted safety measures such as **improved content moderation of misogynistic and abusive content, stronger protections for public figures (e.g. female politicians, journalists),** and **better tools for reporting and redress** while taking into account **intersectional identities**. However, the proposed guidance on women and girls will not be as strongly enforceable as a Code of Practice under the UK OSA framework.[489]

On the other hand, the **European DSA requires very large online platforms (VLOPs) and very large online search engines (VLOSEs)** to '*diligently identify, analyse and assess **any systemic risks** in the Union stemming **from the design or functioning of their service** and its related systems, **including algorithmic systems**, or from the use made of their services… This risk assessment shall be specific to their services and proportionate to the systemic risks, taking into consideration their severity and probability, and shall **include the following systemic risks**:… any actual or foreseeable negative effects in relation to **gender-based violence***'.[490]

The DSA risk assessment includes assessing how the various system designs - recommender systems, content moderation algorithms, terms and conditions, advertising selection systems, or data-related practices - impact system risks. **VLOPs and VLOSEs are required to assess the role of their services in the threat and take specific, proportionate measures to mitigate it, on an annual basis or prior to deploying functionalities that are likely to have a critical impact on the risk of gender-based violence.[491]** This broad mechanism appears to suggest a holistic understanding of the complex and multidirectional nature of online harms, providing more comprehensive and effective risk management than OSA's narrow framework. Adopting a systemic risk-based approach identifies that TFGBV is not a consequence of sporadic or individual instances of abuse, but one that has several causal factors, is widespread and recurrent in nature and is facilitated by the design of online platforms.

---

[485] End Violence Against Women, *Women and girls failed by government's Online Safety Bill*, 2022, https://www.endviolenceagainstwomen.org.uk/women-girls-failed-governments-online-safety-bill/

[486] Benjamin Farram, *How do we understand online harms? The impact of conceptual divides on regulatory divergence between the Online Safety Act and Digital Services Act,* https://www.tandfonline.com/doi/full/10.1080/17577632.2024.2357463.

[487] Somerset and Avon Rape and Sexual Abuse Support, *Online Safety Act 2023: What protections are there for women and girls? (Part 2)*, 2024, https://www.sarsas.org.uk/online-safety-act-2023-what-protections-are-there-for-women-and-girls-part-2/.

[488] Ofcom, *A Safer Life Online for Women and Girls, Practical Guidance for Tech Companies*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/consultation-on-draft-guidance-a-safer-life-online-for-women-and-girls/main-docs/annex-a-draft-guidance.pdf?v=391669.

[489] Jess Eagelton, *The Online Safety Bill - What does it mean for women and girls?*, 2023, https://refuge.org.uk/news/the-online-safety-bill-what-does-it-mean-for-women-and-girls/.

[490] Article 34 - Risk Assessment, Digital Services Act.

[491] European Commission, Q&A on risk assessment reports, audit reports and audit implementation reports under DSA, https://digital-strategy.ec.europa.eu/en/faqs/qa-risk-assessment-reports-audit-reports-and-audit-implementation-reports-under-dsa.

5.3.2. Criminalisation of TFGBV

The inclusion of TFGBV as a systemic risk within the DSA aligns with the EU's aim to align with the ***Directive to Combat Violence against Women and Domestic Violence***, published in May 2024.[492] The Directive seeks to establish minimum criminal standards for the perpetration of cyber stalking, non-consensual sharing of intimate or manipulated material, and cyber-incitement to violence or hatred. For each of these forms of TFGBV, the Directive sets out harmonised standards in order to ascertain if the conduct reaches the level of criminality. This includes the prerequisite that any content in question was publicly disseminated and for the conduct to have caused serious harm (including psychological) to the victim, or caused them to fear for their own safety or that of their dependents in order to meet the standards for the criminalisation.

With this, EU lawmakers have tried to strike a balance in recognising the severe harm that can be caused by TFGBV whilst ensuring safeguards are in place to prevent cases being inappropriately criminalised – such as when images are shared consensually. From the initial publication of the Directive, the provision on cyberharassment caused intense debate[493] due to the concerning proposal to criminalise "initiating an attack with third parties, directed at another person, by making threatening or insulting material accessible...," **without any supplementary definitions or specified safeguards to prevent the criminalisation of legitimate speech. EU negotiators conveyed that they made some improvements to the text in the final agreement and additionally criminalised the unsolicited sending of content containing the genitals of a person (i.e., cyberflashing), a prevalent issue which significantly impacts young women and girls.** EU lawmakers also clarified that threatening conduct is only criminalised when it is repeated/continuous and when that conduct involves threats to commit criminal offences.

Similarly, on a national level, **Canada has adopted a criminal-law based approach to addressing TFGBV treating many forms of online abuse as criminal offences under the Criminal Code. Rather than relying on a separate, standalone cybercrime law, Canada embeds relevant offences across multiple provisions of the Code**, such as those prohibiting criminal harassment[494], publication and distribution of intimate image without consent[495], voyeurism[496], identity fraud[497], hate propaganda[498], and criminal threats[499].

These offences must be proven beyond a reasonable doubt, typically requiring evidence of intent, harm, and lack of consent - placing a high evidentiary burden on the complainant.[500] Operationally, TFGBV cases are investigated and prosecuted through the standard criminal justice system, with survivors often engaging police to file complaints, followed by formal charges and trial proceedings. While many of these provisions pre-date the rise of online abuse, their application has been extended in the digital contexts as well[501], especially where the behaviour is persistent, targeted, and harmful. **Complementing these laws, Canada has introduced civil remedies and protective orders in some provinces, such as Nova Scotia's Intimate Images and Cyber-Protection Act[502]** that create

[492] European Commission, *Ending gender-based violence,* https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence/ending-gender-based-violence_en.

[493] Asha Allen and Dhanaraj Thakur, *CDT Europe Reacts to EU Directive on Gender-Based Violence (GBV) – New Rules to Tackle Online GBV Create Free Expression Concerns*, 2024, https://cdt.org/insights/cdt-europe-reacts-to-eu-directive-on-gender-based-violence-gbv-new-rules-to-tackle-online-gbv-create-free-expression-concerns/

[494] Section 264, Criminal Code.

[495] Section 162.1, Criminal Code.

[496] Section 162, Criminal Code.

[497] Section 403, Criminal Code.

[498] Section 318, Criminal Code.

[499] Section 264, Criminal Code.

[500] Tech Safety Canada, *Legal Protections for TFGBV: What Laws Apply to You?*, https://techsafety.ca/resources/toolkits/legal-protections-for-tfgbv-what-laws-apply-to-you

[501] Through several legislative amendments to the Criminal Code, the Parliament of Canada has extended the application of the Code in digital contexts. For example, C-15A was passed to allow Section 162, relating to NCII, to be applicable to digital content/media. Similarly, C-13 was passed to extend the Act's application to instances of cyber harassment.

civil remedies to deter, prevent and respond to the harms of non-consensual sharing of intimate images and cyber-bullying. However, challenges remain in terms of accessibility as the process is complex, costly and inaccessible for vulnerable communities without any legal support.[503] **Further, critics also point out that while the Act offers protective orders and takedown requests, it does not guarantee timely or enforceable outcomes, particularly when content is hosted on non-Canadian platforms, which has been corroborated by studies that highlight a gap between the law's intent and its real-world impact.[504]**

This approach of addressing TFGBV through existing or expanded criminal law frameworks is increasingly common across jurisdictions. Many countries are choosing to integrate TFGBV into their broader criminal codes rather than developing new, standalone statutes. **For instance, the UK's OSA creates new criminal offences under the Sexual Offences Act, 2003 specifically targeting online harms disproportionately affecting women, such as cyberflashing and NCII sharing.[505]** Offenders who share non-consensual intimate deepfakes can face up to two years of imprisonment and additional monetary penalties.[506] This reflects a global trend where there is a growing legal recognition of integrating specialised forms of TFGBV within criminal law frameworks.

5.3.3. Gender-neutral laws addressing TFGBV

Australia's Online Safety Act, 2021 (that was previously discussed in Chapter 4) is a gender-neutral legislation which regulates **adult cyber-abuse and the non-consensual sharing of intimate images**. The Act defines adult cyber abuse as material targeting a particular Australian adult that is both[507]:

1. *intended to cause serious harm, and*
2. *menacing, harassing or offensive in all the circumstances.*

If the material only meets one of the two criteria above (for example, if the post is offensive but is found to not be intended to cause serious harm), it will **not be considered adult cyber abuse under the Act**. The Act defines **'serious harm'** to mean serious physical harm or serious harm to a person's mental health, whether temporary or permanent. This includes serious psychological harm and serious distress that goes beyond '**mere ordinary emotional reactions such as those of only distress, grief, fear or anger**'. On its own, purely financial harm, defamatory material that causes purely reputational harm, or incidental harm experienced as part of social or community interaction is not enough to be considered 'serious harm'. Generally the eSafety Commissioner will consider the occurrence and prominence of the following factors while determining *serious harm* in the context of adult cyber abuse[508]:

- *Revealing personal information to deliberately make someone feel unsafe, which is known as doxxing;*
- *Urging or encouraging violence against a person including actively inciting self-harm;*
- *Threats of violence;*
- *Posts designed to generate volumetric and 'pile-on' attacks from others;*
- *Relevant history between the target and the end-user;*
- *Behaviour which is clearly targeting a known vulnerability of the person targeted that exacerbates that vulnerability. This might occur, for example, where there is evidence that the person posting, sharing or sending*

[502] Intimate Images and Cyber-protection Act, Province of Nova Scotia, Canada.

[503] Global News, *Nova Scotia's proposed anti-cyberbullying bill creates access problems: experts*, 2017, https://globalnews.ca/video/3820511/nova-scotias-proposed-anti-cyberbullying-bill-creates-access-problems-experts/

[504] Anam Khan, *Nova Scotians aren't getting the help they need to remove online intimate images, expert says,* 2023, https://www.cbc.ca/news/canada/nova-scotia/intimate-images-increasing-not-getting-help-1.6777240.

[505] Jankowicz, et al., *It's Everyone's Problem: Mainstreaming Responses to Technology-Facilitated Gender-Based Violence*, 2024, https://igp.sipa.columbia.edu/sites/igp/files/2024-09/IGP_TFGBV_Its_Everyones_Problem_090524.pdf

[506] Section 188, Online Safety Act

[507] Section 7, Online Safety Act.

[508] eSafety Commissioner, Australia, *Adult Cyber Abuse Scheme, Regulatory Guidance - eSC RG 3*, 2025, https://www.esafety.gov.au/sites/default/files/2025-01/Adult-Cyber-Abuse-Scheme-Regulatory-Guidance-January2025.pdf?v=1747897746222.

the material is aware of the targeted person's mental health history and the material is intended to worsen the targeted person's wellbeing;

- *Age of the end user is one of the mitigating factors which will not definitively rule out seeking removal action, however it is a factor to be taken into account in determining appropriate responses; and*
- *Online incitement of any of the above activities.*

Serious harm in the context of adult cyber abuse is to be considered objectively. It is not enough that a person felt seriously harmed by the material but rather whether an ordinary reasonable person would likely conclude that the post was intended to cause serious harm. Such a high threshold for intervention means that instances of women being distressed by harms such as online trolling might not meet the criteria for takedown orders.

While the law in itself is gender neutral, the eSafety Commissioner's guidance provides that while determining whether any content or behaviour amounts to *menacing, harassing or offensive*, it will consider whether a person has been targeted because of their gender, sexual orientation, disability, mental health condition or family or domestic violence situation.[509] The eSafety Commissioner also has dedicated resources[510] for women to understand online risks and available remedies.

It is essential to note that Australia's eSafety Commissioner works on a complaint-based system and can only act on individual complaints received from users or organisations. **While it currently does not have formal powers to compel the removal of adult cyber abuse materials, in extreme cases, it is able to leverage its strong relationship with platforms to request for the voluntary removal of these materials.[511]** If cyber abuse appears to reach the criminal threshold, eSafety Commissioner can recommend that the complainant approach relevant law enforcement agencies for appropriate action. However, since they cannot exercise *suo motu* powers to proactively monitor or censor content, their **role as a regulator remains constrained.**

5.3.4. Gendered Laws addressing TFGBV

Mexico's Frente Nacional para la Sororidad or National Front for Sorority (FNS), was founded in 2013 by Olimpia Coral Melo Cruz to defend the digital rights of women and girls and to demand online spaces free of gender-based violence. **FNS developed and promoted a legal reform known as the Ley Olimpia (Olimpia Law) to address the legal vacuum for online harms against women in Mexico**.

Between 2018 and 2022, the Olimpia Law[512] was approved by all the country's federal institutions and in 2021 it was incorporated into national and federal laws. **Till 2022, this resulted in 35 legal reforms across 28 local legislatures to include different forms of online violence and abuse in their principal laws on ending violence against women and girls.[513]** This includes **criminalization of sextortion, threats, cyber harassment, sexual harassment, and non-consensual image sharing**. The Olimpia Law transformed the way sexual crimes are defined in Mexico's different criminal codes, which until then had only recognised actions involving sexual abuse, mistreatment, or transgression of the physical integrity of individuals.

---

[509] eSafety Commissioner, Australia, *Adult Cyber Abuse Scheme, Regulatory Guidance - eSC RG 3*, 2025, https://www.esafety.gov.au/sites/default/files/2025-01/Adult-Cyber-Abuse-Scheme-Regulatory-Guidance-January2025.pdf?v=1747897746222.

[510] eSafety Commissioner, Australia, *eSafety - Women,* https://www.esafety.gov.au/women.

[511] UN Office of the High Commission of Human Rights, *UN Special Rapporteur on violence against women, its causes and consequences, Questionnaire on gender-based violence against women journalists - Response of the Government of Australia,* https://www.ohchr.org/sites/default/files/Documents/Issues/Women/SR/VAWJournalists/Government/australia.pdf

[512] Olimpia Coral Melo Cruz, a women's-rights activist from the Mexican city of Puebla, is a survivor of revenge porn—sexual content that is shared without the consent of those featured within it. She turned her experience into action, and in April 2021, Mexico passed Olimpia's Law, federally prohibiting the sharing of such content without the subject's permission.

[513] UN Women, *Accelerating Efforts to Tackle Online and Technology Facilitated Violence Against Women and Girls (VAWG)*, 2022, https://www.unwomen.org/sites/default/files/2022-10/Accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls-en_0.pdf

As a result, it has been a ground-breaking piece of gender legislation in Latin America, with a parallel now also adopted in Argentina.[514]

After the enactment of the Act in 2021, government institutions began to[515]:

- Launch campaigns such as the Violentómetro Digital ('Digital Violence Meter') to publicise the Olimpia Law;
- Establish certain institutional channels for reporting of crimes against sexual intimacy;
- Train state officials, especially those in prosecutor's offices; and
- Reform prosecutor's offices to enable them to pursue the new crime. **For example, the Specialised Agency for Crimes Against Sexual Privacy was created in the Mexico City Prosecutor's Office**.

Both male and female officials highlighted[516] the significant impact of training on their perceptions and understanding of online violence. The categorisation of digital sexual violence (DSV) under the Olimpia Law helped them identify the acts that constitute such violence. Additionally, their direct engagement with victims/survivors has fostered greater empathy and a deeper awareness of the consequences of DSV.

However there **remains local resistance** to laws that include a gender perspective, as well as a reluctance to recognise online spaces as real. In addition, women are still subjected to stigmatising ideas about female sexuality, which suggest women are responsible for the violence inflicted on them. **Although some officials consider it inappropriate, a culture of machismo and male complicity still exists among male officials.** This appears to limit the application of the Olimpia Law.[517] According to a digital survey[518], 72.2% of victim-survivors of DSV did not make a complaint, and those that did mostly reported to social media platforms (11.3%). **The reasons given for not reporting cases of DSV were that it was easier to let it slip, fear of complaining, and distrust of authorities**.

Commentators have noted that the experience of the Olimpia Law highlights the meaningful impact of adopting frameworks that embrace gender-specific definitions as well as gender-sensitive enforcement approaches to addressing online harms. Regarding the latter, the Olympia law filled a critical legislative vacuum and also reshaped institutional practices, such as establishing specialised prosecution units, training law enforcement officials in victim-centred approaches, developing gender-specific reporting channels, etc.[519]

5.4. Emerging Risks to Women's Safety Online and Responses by Regulators

5.4.1. Non-consensual intimate image abuse

Countries around the globe are taking proactive efforts to address abuse enabled by the sharing of non-consensual intimate images (NCII). This includes incorporating specific provisions in their digital safety legislations with **notable examples include Australia, the UK, and the EU.**

Section 16 of the Australia Online Safety Act defines NCII as follows: *if (a) an intimate image of a person is*

[514] Recognised as such by UN Women in a message published on X (Twitter at the time) on 30 April 2021.

[515] ALiGN - Advancing Learning and Innovation on Gender Norms, *Digital sexual violence against women in Mexico*, 2024, https://www.alignplatform.org/sites/default/files/2024-10/align-mexico-digitalsexualviolence-execsummary-eng-digital.pdf

[516] ALiGN - Advancing Learning and Innovation on Gender Norms, *Digital sexual violence against women in Mexico*, 2024, https://www.alignplatform.org/sites/default/files/2024-10/align-mexico-digitalsexualviolence-execsummary-eng-digital.pdf

[517] ALiGN - Advancing Learning and Innovation on Gender Norms, *Digital sexual violence against women in Mexico*, 2024, https://www.alignplatform.org/sites/default/files/2024-10/align-mexico-digitalsexualviolence-execsummary-eng-digital.pdf

[518] ALiGN - Advancing Learning and Innovation on Gender Norms, *Digital sexual violence against women in Mexico*, 2024, https://www.alignplatform.org/sites/default/files/2024-10/align-mexico-digitalsexualviolence-execsummary-eng-digital.pdf

[519] Suzie Dunn, *Addressing Gaps and Limitations in Legal Frameworks and Law Enforcement on Technology-Facilitated Gender-based Violence*, 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4852083

provided on: (i) a social media service; or (ii) a relevant electronic service; or (iii) a designated internet service; and (b) the person did not consent to the provision of the intimate image on the service; and (c) the provision of the intimate image on the service is not an exempt provision of the intimate image, the intimate image is a non-consensual intimate image of the person. The framework has a **complaints and objections system for NCII** which is administered by the e-Safety Commissioner.

Beyond this, the provider of a social media service or internet service, the end-user who posts the image, or a hosting service provider that hosts an intimate image, **may be given a removal notice requiring the intimate image to be removed from the service. The Act also covers an extended range of NCII, including deepfakes (which is explored in detail below), by providing that it is not relevant if the material has been digitally altered in any way**.[520] However, there are two restricting conditions: the image must have a link to Australia (i.e. the person depicted in the image or the user who posted the image must be a resident of Australia); and in most cases, the complaints must be made by the individual depicted in the image itself (except in case where the complaint is made by an authorised person, parent or guardian where the individual is a child or is mentally or physically incapacitated). This means that a member of the general public cannot report such content even if the harm is apparent and it is reasonable to expect that the person depicted would not have consented to it. **As a result, these images could spread and cause significant damage, even before the targeted person becomes aware of it, which can be particularly problematic for incidents like those involving deepfake intimate images**.

The **UK OSA** makes it a serious offense to share NCII. **Section 188 of the UK OSA amends the Sexual Offences Act, 2003 by introducing a new Section 66B, which creates four criminal offences related to intimate image abuse.**[521]

- **Section 66B(1)** establishes a base offence of sharing intimate images without the consent of the person depicted.
- **Section 66B(2)** introduces a more serious offence where such sharing is done with the intent to cause alarm, distress, or humiliation.
- **Section 66B(3)** further criminalises the sharing of intimate images without consent—or without a reasonable belief in consent—when done for the purpose of obtaining sexual gratification.
- **Section 66B(4)** makes it an offence to threaten to share intimate images if the intention is to cause fear that the threat will be carried out, or if the perpetrator is reckless as to whether such fear will arise in the victim or someone who knows them.

Under the UK OSA framework, victims/survivors do not have the burden to prove 'intent to distress' for it to be considered a base offense. The law also provides for strong enforcement measures which include:

- Imprisonment of 6 months for sharing an intimate image without consent or reasonable belief of consent;
- Imprisonment of 2 years to anyone who shares the image with intent to cause distress or humiliation;
- Imprisonment of 2 years and be subject to the sex offenders register for anyone who shares an intimate image to obtain sexual gratification;
- A maximum of 2 years imprisonment to anyone who threatens to share an intimate image.

In March 2025, a UK House of Commons committee report recommended that the government expand the definition of NCII to include imagery that is non-sexual, however *culturally intimate* for the victim, such as a Muslim woman

[520] Australian Government, *Statutory Review of the Online Safety Act 2021,* 2024, https://www.infrastructure.gov.au/sites/default/files/documents/online-safety-act-2021-review-issues-paper-26-april-2024.pdf.

[521] House of Commons, UK Parliament, *Tackling non-consensual intimate image abuse*, 2025 https://publications.parliament.uk/pa/cm5901/cmselect/cmwomeq/336/report.html.

being pictured without her hijab.[522] This report further suggested that NCII should be brought in line with CSAM in law, so that mere possession of NCII can be treated as a criminal offence.

In **Italy**, under its criminal code, the **illegal dissemination of sexually explicit images or videos**[523] has been added as a new offence. This provision punishes anyone who, after having taken or stolen sexually explicit images or videos that were intended to remain private, distributes them without the express consent of the persons concerned. Those who further distribute the images disseminated by the offender are also punished. Under the law, the punishment increases if a spouse commits the act, whether divorced or divorced, or by a person who is or has had an intimate relationship with the person concerned or if the act is committed against a physically or mentally weak person or to the detriment of a pregnant woman. **With the new law, the quality of criminal repression of revenge pornography in Italy has drastically improved.** The nature of the sentence (imprisonment with up to six years that can be extended) shows that Italy is regarding revenge pornography as a serious sexual offence that protects the sexual integrity of an individual, and not only as a minor privacy violation offence. **With the new legislation Italy has become the county in continental Europe with the most severe and serious approach to revenge pornography.**[524] Moreover, under **Italy's Privacy Code**, an urgency procedure was introduced in January 2022 which allows victims of revenge porn to submit a complaint to the **Data Protector Authority 'Garante'** and upload the pornographic material in encrypted form from the DPA's own website. The DPA will then verify the complaint and take urgent measures within 48 hours. Specifically, the DPA will order social media or content sharing platforms to stop broadcasting the pornographic material. **Platforms are also required to store the pornographic material for up to 12 months for evidentiary use, and to take measures to prevent the identification of the data subject while the material is stored.**

In **South Korea**, the *Act on Special Cases Concerning the Punishment of Sexual Crimes* imposes severe penalties for illegal filming and distribution, with up to seven years of imprisonment for the distribution of sexual images or videos recorded without consent.[525] Article 14(1) of the Sex Crimes Act makes it a crime when a person "*takes photographs or videos of a person's body that may cause sexual stimulus or humiliation without the victim's consent, by means of using a camera or other devices with similar functions*". Experts have raised concerns about the language of this provision which excludes abuses, such as someone filming another without consent in her home when there is no nude or sexual content and phrases like 'sexual stimulus' or 'shame' have a subjective standard.[526] The ***Act on Promotion of Information and Communications Network Utilization and Information Protection*** regulates various illegal activities conducted via online networks and also prohibits the distribution of illegal sexual images or videos and allows for the punishment of those who disseminate such content.

Further, as mentioned above, the EU's Directive to Combat Violence against Women establishes minimum criminal standards for the perpetration of NCII or digitally manipulated material like deepfake.

---

[522] House of Commons, UK Parliament, *Tackling non-consensual intimate image abuse*, 2025 https://publications.parliament.uk/pa/cm5901/cmselect/cmwomeq/336/report.html.

[523] Italy, *Article 612 of the Codice Penale*, (approved by Royal Decree No. 1398 of October 19, 1930, as amended up to Legislative Decree No. 63 of May 11, 2018), https://www.wipo.int/wipolex/en/legislation/details/18132.

[524] Miha Šepec, *Revenge Pornography or Non-Consensual Dissemination of Sexually Explicit Material as a Sexual Offence or as a Privacy Violation Offence*, 2019, https://www.cybercrimejournal.com/pdf/MihaSepecVol13Issue2IJCC2019.pdf.

[525] Jang & Suh, *Cyber Sex Crimes Targeting Children and Adolescents in South Korea: Incidents and Legal Challenges*, 2024, https://www.mdpi.com/2076-0760/13/11/596#:~:text=The%20Act%20on%20Promotion%20of,and%20psychological%20treatment%20to%20victims.

[526] Human Rights Watch, *"My life is not your porn": Digital Sex Crimes in South Korea*, 2021, https://www.hrw.org/sites/default/files/media_2021/06/southkorea0621_web_1_0.pdf.

5.4.2. Deepfake-based Image Abuse

Deepfake image-based abuse constitutes an emerging form of technology-enabled violence against women that uses AI (and software or other digital tools) to morph and create deceptive and non-consensual content.[527] There has been an exponential increase in deepfake content and such content is gendered, disproportionally targeting women and girls. Studies show that there has been a 550% increase in deepfake videos from 2019 to 2023 and that 98% of deepfakes flagged in this period consisted of pornographic content featuring women.[528]

Under the EU AI Act, deepfake is defined as an '*AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful*'. While the AI Act does not impose a blanket ban on deepfake content, it adopts a nuanced approach and imposes obligations for systems generating or manipulating such content. Some of the main provisions regarding deepfakes are:

- **Transparency Obligations**: To prevent misinformation, persons who create or disseminate deepfake content are required to disclose and label the content as AI-generated.
- **Accountability and Traceability**: In order to ensure that the origins of deepfake content is traceable for authorities, records of the processes and data used to generate deepfakes are to be maintained.

Despite these disclosure obligations, it is questionable whether the AI Act would be effective in tackling deepfakes. One concern that remains is the risk classification of deepfakes. The EU AI Act adopts a four-level risk-based approach in regulating AI systems: unacceptable, high, limited, and minimal (or no) risk. **However, the AI Act classifies deepfake AI as limited risk subject to lighter transparency obligations.**[529] Moreover, while monetary fines act as an *ex-post* deterrent to address violations, it does not prevent the creation and spread of deepfakes.[530]

As mentioned above, the UK OSA makes sharing of non-consensual deepfakes an offence.[531] However, it is critical to note that the creation of deepfake content is still not criminalised. Notably, apart from criminalizing the sharing of deepfake content, the law does not place any responsibilities on social media companies or other digital services. **OSA only provides that its guidelines 'contain advice and examples for best practice for assessing risks of harm to women and girls'.**[532]

In South Korea, the Sexual Violence Prevention and Victims Protection Act (discussed in the previous section), states that producing sexually explicit deepfakes with the intent to distribute is punishable with up to five years in prison or a fine of 50 million KRW ($37,900).[533] However, in September 2024, South Korea passed a new bill to tackle the surging dissemination of deepfake content. This was in response to media reports that revealed increased instances of networks of Telegram chatrooms across schools and universities, which generated deepfake porn featuring female staff and students. The new law proposed to increase the maximum sentence to seven years, regardless of

---

[527] Equality Now and AUDRi, *Deepfake image-based sexual abuse, tech-facilitated sexual exploitation and the law*, 2024, https://equalitynow.storage.googleapis.com/wp-content/uploads/2024/01/17084238/EN-AUDRi-Briefing-paper-deepfake-06.pdf

[528] Security Hero, *2023 State of Deepfakes: Realities, Threats, and Impact, 2023, https://www.securityhero.io/state-of-deepfakes/.*

[529] EU AI Act,*High-level summary of the AI Act,* 2024, https://artificialintelligenceact.eu/high-level-summary/.

[530] Mauro Fragale and Valentina Grilli, *Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation, 2024,* https://cjel.law.columbia.edu/preliminary-reference/2024/deepfake-deep-trouble-the-european-ai-act-and-the-fight-against-ai-generated-misinformation/.

[531] Section 66B, Sexual Offences Act 2003 and Section 28A, Online Safety Act 2023.

[532] Manasa Narayana, *The UK's Online Safety Act Is Not Enough To Address Non-Consensual Deepfake Pornography*, 2024, https://www.techpolicy.press/the-uks-online-safety-act-is-not-enough-to-address-nonconsensual-deepfake-pornography/.

[533] Nandini Singh, *South Korea criminalises explicit deepfake possession amid public outcry,* 2024, https://www.business-standard.com/world-news/south-korea-criminalises-explicit-deepfake-possession-amid-public-outcry-124092700334_1.html.

intent. **Under these new regulations, possessing and viewing sexually explicit deepfake material also attracts serious penalties, such as imprisonment for up to three years or a fine. The new law also requires platform companies to remove the content within 24 hours of a request being made. Moreover, platforms are also encouraged to adopt proactive measures, including blocking suspicious content preemptively**. Furthermore, an automated AI system to detect deepfake content and send real-time removal requests to platforms was also proposed.

In May 2025, the US enacted the first federal law that introduces strict penalties for the distribution of deepfake pornography, AI-generated images that depict real people, and authentic media that is non-consensual.[534] The act makes it illegal to "knowingly publish" NCII, including visual depictions "*created through the use of software, machine learning, artificial intelligence, or any other computer-generated or technological means*." The act establishes a "reasonable person" test for determining NCII, which when "viewed as a whole by a reasonable person, is indistinguishable from an authentic visual depiction of the individual". Penalties include up to three years of prison time. **The law aims to enforce 48-hour take downs, marking the US federal government's first attempt on imposing horizontal safety obligations on online platforms.** However, the law has faced criticism within the US for posing unacceptable risks to user expression that does not involve NCII as well as to users' privacy and online security. By heavily incentivizing platforms to rely on automated content detection and filtering systems, the law raises concerns about overreach, as these technologies often have significant limitations, potentially resulting in the erroneous removal or suppression of lawful and legitimate speech.[535]

Box 5.1: Deepfake-Based Image Abuse: Technical Interventions for Content Provenance and Authenticity

Deepfake-based image abuse - especially that involving non-consensually generated sexual or intimate images - is a rapidly escalating threat to women's safety online. In response, technical interventions aimed at validating content provenance and authenticity have emerged as critical tools.

Digital Watermarking & Invisible Markers
Platforms have implemented imperceptible watermarks in AI-generated media. Meta's 'Stable Signature' system[536] and Google DeepMind's SynthID[537] embed hidden identifiers directly into generated images to help provenance detection and flag AI-originated content. The Stable Signature system is a watermarking technique, invisible to the naked eye but detectable by algorithms, that can distinguish when an image is created by an open source generative AI model. SynthID uses a similar technique to distinguish AI-generated content by embedding technology-detectable digital watermarks.

Proactive Watermarking Models
Efforts like FaceSigns[538] and SepMark have proposed embedding semi-fragile watermarks into original media. These watermarks survive benign transformations but degrade or disappear when content is manipulated by deepfake tools, including tampering.

[534] Section 146, Take It Down Act 2025.

[535] Electronic Frontier Foundation, Congress Passes TAKE IT DOWN Act Despite Major Flaws, 2025, https://www.eff.org/deeplinks/2025/04/congress-passes-take-it-down-act-despite-major-flaws?utm..

[536] Meta, *Stable Signature: A new method for watermarking images created by open source generative AI*, 2023, https://ai.meta.com/blog/stable-signature-watermarking-generative-ai/.

[537] Google DeepMind, *SynthID*, https://deepmind.google/science/synthid/.

[538] Neekhara, et al., *FaceSigns: Semi-Fragile Neural Watermarks for Media Authentication and Countering Deepfakes*, 2022, https://arxiv.org/abs/2204.01960.

C2PA & Content Credentials
The Coalition for Content Provenance and Authenticity (C2PA) - backed by Adobe, Microsoft, Intel, OpenAI, and others - publishes open-source standards for embedding cryptographically signed metadata in images and videos.[539] These verifiable metadata, established by C2PA's technical standards, are called content credentials. Content credentials provide details about who produced a particular content, when they produced it, and the tools and editing processes utilised in producing the content. Content credentials are free, open-standard and capture creation context, editing history, and platform information to serve as a "digital nutrition label" for media.[540]

Detection-Focused AI Tools
Deepfake detection algorithms use convolutional neural networks[541] (e.g., DeepRhythm) and forensic analysis to flag manipulated content with high accuracy. Yet, as deepfake techniques evolve, this remains an ongoing, adversarial challenge. **Thus, these technologies remain experimental and must keep pace with the trajectory of misuse**.

Content Authenticity Initiative (CAI)
CAI, led by Adobe, promotes the large-scale adoption of provenance tools, including content credentials.[542] In adherence to the C2PA standards, the CAI creates a protected end to end system for digital content provenance, using open source tools. Simultaneously, the CAI promotes these tools to a widespread cross-industry community of adopters.[543]

Limitations of Technical Interventions

- **Data Poisoning:** A significant emerging threat is data poisoning, where malicious actors intentionally inject corrupted data into the training datasets of AI models.[544] For deepfake detection, this could involve subtly altering images or videos before they are used to train detection algorithms. This poisoned data can then cause the AI model to misclassify future content, effectively undermining the detector's accuracy or even leading it to identify real content as fake. This poses a severe challenge as it directly targets the integrity of the AI's learning process.
- **Computational Complexity:** Many of the most accurate detection methods require substantial computing power, limiting their real-time application, especially on user devices or for large-scale content processing.[545]

[539] Coalition for Content Provenance and Authenticity (C2PA), *Overview - C2PA*, https://c2pa.org/.

[540] Andy Parsons, *Authenticity in the age of AI: Growing Content Credentials momentum across social media platforms, AI companies and rising consumer awareness,* 2024, https://blog.adobe.com/en/publish/2024/09/18/authenticity-age-ai-growing-content-credentials-momentum-across-social-media-platforms-ai-companies-rising-consumer-awareness.

[541] Convolutional Neural Networks (CNNs) are a specialized type of neural network, primarily used for analyzing visual imagery. They excel at tasks like image recognition, object detection, and image classification due to their ability to automatically learn features from raw pixel data.

[542] Andy Parsons, *Authenticity in the age of AI: Growing Content Credentials momentum across social media platforms, AI companies and rising consumer awareness,* 2024, https://blog.adobe.com/en/publish/2024/09/18/authenticity-age-ai-growing-content-credentials-momentum-across-social-media-platforms-ai-companies-rising-consumer-awareness.

[543] Content Authenticity Initiative, *How a voice cloning marketplace is using Content Credentials to fight misuse*, 2024, https://contentauthenticity.org/blog/community-story-respeecher.

[544] Opera, Singhal & Vassilev, *Poisoning Attacks Against Machine Learning: Can Machine Learning Be Trustworthy?*, 2022, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934932.

[545] Balafrej & Dahmane, *Enhancing practicality and efficiency of deepfake detection,* 2024, https://www.nature.com/articles/s41598-024-82223-y.

5.4.3. Platform Safety Codes for Online Dating

Given that NCII and image based abuse is closely linked with intimate partner violence, it is useful to consider how platform safety codes have emerged for activities like online dating. The **Australian Online Safety Code[546] for Dating Services** is a voluntary industry code of practice which establishes safeguards to reduce the risk of online harm to users of dating services operating in Australia. **The Code was led by an Industry Working Group including representatives from Match Group, Bumble, The Meet Group, eharmony, Spark, Grindr and RSVP.** The working group engaged in extensive consultation with the Australian Government and other stakeholders throughout the drafting process, including external advocacy groups. The Code includes provisions designed to protect users and foster a safer online dating environment. **Dating apps are encouraged to implement profile verification processes, in-app reporting mechanisms for abuse or harassment and options to block or mute other users. Platforms are urged to educate users about safe online dating. This includes providing resources on how to recognise and report suspicious behaviour, as well as promoting awareness about the risks.** The code encourages dating app companies to cooperate with law enforcement in cases of serious misconduct or criminal activity. This involves sharing relevant data and providing support to investigations. By involving industry stakeholders, there is a greater likelihood of buy-in and compliance.

The Australian Government has no statutory powers to enforce compliance with the Code or seek the imposition of civil penalties on non-compliant dating services. However, the dating services which have signed up to the Code have agreed to comply with the obligations set out within it. These obligations include annual reporting to regulators, and standards of accessibility and safety across platforms. **The independent Code Compliance Committee can subject non-compliant online dating platforms to enforcement measures such as formal warnings, directive to create plans for meeting compliance or suspension or removal from the Code.[547]**

5.4.4. Tech-enabled Trafficking

During both stages of trafficking - recruitment and exploitation - technology can play an exacerbating role. While technology assists malicious actors in identifying and contacting potential victims during the recruitment stage, technology can facilitate sale of sexual services during the exploitation stage and allow traffickers to monitor and control the survivors as well. Importantly, for LEAs the intermediation of technology is concerning as it allows coordinating the trafficking from a place that is separate from the place where the sexual activity is performed.[548]

In the UK, the OSA provides a list of "priority illegal content" which imposes additional specific duties for platforms. Human trafficking is included within the list of "priority illegal content". In March 2025, Ofcom implemented its illegal harms code of practice, which requires platforms to conduct illegal harms risk assessment and implement appropriate measures to remove illegal material promptly, including content or behaviour relating to human trafficking.[549] **This effort sets out over 40 safety measures for platforms to tackle human trafficking, such as easier reporting mechanisms, strict performance targets for content moderation**

---

[546] eSafety Commissioner (Australia), *Code of Practice - Online Safety Code for dating services,* 2024, https://www.australianonlinedatingcode.com.au/Australian%20Voluntary%20Code%20for%20Online%20Dating%20Services%20Code%20of%20Practice.pdf

[547] The Hon Michelle Rowland MP, *Online dating platforms now subject to enforcement*, 2025, https://minister.infrastructure.gov.au/rowland/media-release/online-dating-platforms-now-subject-enforcement.

[548] Council of Europe, *Online and technology-facilitated trafficking in human beings,* 2022, https://rm.coe.int/online-and-technology-facilitated-trafficking-in-human-beings-summary-/1680a5e10c&sa=D&source=docs&ust=1734500535701316&usg=AOvVaw3PdeF3VO3LFlUlo9wPx_t0.

[549] Ofcom, *Enforcing the Online Safety Act: Platforms must start tackling illegal material from today*, 2025, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/enforcing-the-online-safety-act-platforms-must-start-tackling-illegal-material-from-today.

systems, closer monitoring of platform governance structures, etc. and make them safer by design.[550]

A separate study by the European Council titled '***Online and technology-facilitated trafficking in human beings***' provides the following recommendations[551]:

- Need to invest in capacity building for LEAs in internet monitoring, cyber-patrols, undercover online investigations, automatic searching tools to analyze evidence, and tools to assist investigators in handling and processing large-volume data.
- Stricter regulations and control over job advertisement websites by leveraging technological support.
- States and private parties must work together to provide confidential and anonymous reporting mechanisms for survivors of trafficking, along with developing tools to detect such instances online.
- Training to be given to law enforcement officers in both ICT and trafficking, ensuring they have the expertise to handle e-evidence.
- Regular study and analysis of emerging patterns on offenders' *modus operandi* to be carried out.
- Increased cross-border cooperation between countries is essential as they can benefit from the sharing of technological support and best practices, including a more streamlined and simplified process for Mutual Legal Assistance Requests.
- Countries should develop cooperation protocols and data-sharing procedures with private companies.

5.4.5. Other Emerging Harms

1. **Doxxing:** There is no legal definition of "doxxing" in international human rights law and no international legal framework that directly regulates or tackles "doxxing". Also, across the different jurisdictions, no law specifically references the term 'doxxing', however, some laws make it a criminal offence to share another's personal information. **In France, Article 223-1 of the Criminal Code prohibits revealing personal information when such an act enables that person to be identified or located with the intention of exposing them or their family to a direct risk of harm.** Other countries have a variety of laws that could, in theory, be applied to tackle doxxing. In Australia, doxxing could be regulated under the Criminal Code Act and the Privacy Act[552], while it also finds mention in the Adult Cyber Abuse Regulatory Guidance[553]. In the UK, several laws might provide some redress, such as the Protection from Harassment Act and the Malicious Communications Act, and victims of doxxing might be able to get posts containing their personal information taken down under the UK's OSA[554].

   In the EU, the DSA could also be relevant to prohibiting doxxing with respect to the removal of 'illegal content' by 'intermediary' services such as social media platforms. Doxxing could be considered as 'illegal content' within the meaning of the DSA in Member States where it is expressly prohibited under national law[555]. Parallelly, under the EU's GDPR, processing personal data on any non-consensual, unlawful grounds is prohibited[556], which can

[550] Ofcom, *Time for tech firms to act: UK online safety regulation comes into force*, 2024, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/time-for-tech-firms-to-act-uk-online-safety-regulation-comes-into-force.

[551] Council of Europe, *Online and technology-facilitated trafficking in human beings,* 2022, https://rm.coe.int/online-and-technology-facilitated-trafficking-in-human-beings-summary-/1680a5e10c.

[552] Privacy and Other Legislation Amendment Act 2024, No. 128, 2024.; Library of Congress, *Australia: New Privacy Legislation Criminalizes Doxxing*, 2024, https://www.loc.gov/item/global-legal-monitor/2024-12-16/australia-new-privacy-legislation-criminalizes-doxxing/.

[553] eSafety Commissioner, *Adult Cyber Abuse Scheme Regulatory Guidance eSC RG 3,* 2023, https://www.esafety.gov.au/sites/default/files/2023-12/Adult-Cyber-Abuse-Scheme-Regulatory-Guidance-Updated-December2023.pdf?v=1734600780633.

[554] AUDRi and Equality Now, *Doxing, digital abuse and the law,* 2024 https://audri.org/wp-content/uploads/2024/02/EN-AUDRi-Briefing-paper-doxing-04.pdf.

[555] AUDRi and Equality Now, *Doxing, digital abuse and the law,* 2024 https://audri.org/wp-content/uploads/2024/02/EN-AUDRi-Briefing-paper-doxing-04.pdf.

[556] Article 5, General Data Protection Regulation.

misconstrue and undermine doxxing as a privacy violation, rather than a larger online safety violation requiring proactive risk assessment and mitigation. **Additionally, while the EU Directive on Combating Violence Against Women and Domestic Violence does not explicitly mention doxxing, it can nonetheless be used to regulate such behaviour. The Directive (albeit while dealing with cyber harassment) criminalises the act of making any material containing the personal data of a person accessible to the public without that person's consent, for the purpose of inciting others to cause physical or serious psychological harm[557].**

Evidently, we don't observe a consistent approach across different legal jurisdictions for a victim of doxxing to seek help by having their personal details removed from online platforms or to seek redress from the harm caused. In some countries, a patchwork of several laws might apply, including criminal and civil remedies. This makes it likely that action to seek remedies would neither be fast nor easy[558].

2. **Networked Harassment**: Generally, coordinated trolling and mass attacks of women are covered under individual cases of **hate speech in most jurisdictions**. **However, Australia's Adult Cyber Abuse Policy[559] includes posts designed to generate volumetric and 'pile-on' attacks** from others as a metric for evaluating serious harms for the purpose of determining cyber abuse. Ofcom's Code of Practice[560] published in December 2024 includes harassment, stalking, threats and abuse as priority offences under the Online Safety Act, 2023. However, it does not mention pile-on attacks.

3. **Cyberflashing**: The UK's Online Safety Act, 2023 added a new offence of "cyber-flashing" under section 66A of the Sexual Offences Act 2003 (which was added by section 187 of the OSA). The offence requires a relevant image to be sent or given to another person with the intention that the recipient will be caused alarm, distress or humiliation, or for the purpose of obtaining sexual gratification and being reckless as to whether the recipient will be caused alarm, distress or humiliation. The offence is based on proof of the perpetrator's motive to cause the survivor harm or for sexual gratification - but intent and motivation can be hard to prove, and this may create a worrying loophole in the law.[561] In the Crown Court, the maximum penalty is a fine and/or a term of imprisonment of up to two years.

5.5. Programmatic Interventions for addressing TFGBV

To effectively prevent and respond to TFGBV, it is essential to examine programmatic interventions that go beyond legislation and policy.[562] This section analyses **notable state-led and non-state initiatives** that have developed practical models to address the complex and evolving nature of online harms against women.

[557] Article 7(d), Directive on Combating Violence Against Women and Domestic Violence.

[558] AUDRi and Equality Now, *Doxing, digital abuse and the law,* 2024 https://audri.org/wp-content/uploads/2024/02/EN-AUDRi-Briefing-paper-doxing-04.pdf.

[559] eSafety Commissioner, *Adult Cyber Abuse Scheme Regulatory Guidance eSC RG 3,* 2023, https://www.esafety.gov.au/sites/default/files/2023-12/Adult-Cyber-Abuse-Scheme-Regulatory-Guidance-Updated-December2023.pdf?v=1734600780633.

[560] Ofcom, *Overview of Illegal Harms,* 2023, https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/overview-of-illegal-harms.pdf?v=387538.

[561] Jess Eagelton, *The Online Safety Bill - What does it mean for women and girls?*, 2023, https://refuge.org.uk/news/the-online-safety-bill-what-does-it-mean-for-women-and-girls/

[562] Georgetown Institute for Women, *Peace and Security, Beyond engaging men: Masculinities, (Non) Violence and Peacebuilding*, 2023, https://giwps.georgetown.edu/wp-content/uploads/2023/10/Beyond_Engaging_Men.pdf.

Table 5.1: State-led Interventions and non-state initiatives

State-led Interventions
United Kingdom
The UK Revenge Porn Helpline (RPH)[563] was established in 2015 and helps prevent individuals from becoming victims of non-consensual intimate image abuse. The RPH has a 90% removal rate, successfully removing over 200,000 individual non-consensual intimate images from the internet till 2022.
Non-state Interventions
Germany
HateAid[564] offers a dedicated legal consultation service for victims of digital violence specifically in the context of the DSA, by helping individuals understand their rights under the regulation - particularly around platform obligation and risk mitigation duties of VLOPs and VLOSEs. This service includes guidance on how to file user-led complaints under the DSA and creating a direct channel of platform accountability.
African Union
The Africa Online Safety (AOS) Program[565] supports local organisations in developing innovative, community-based solutions to online harms. They create a network of local organisations and volunteers that can offer a range of victim-focused services, including counselling, digital literacy training, public awareness campaigns, training programs for youth on identifying online safety risks and assistance with reporting online harms. By tailoring interventions to local contexts and needs, the Program helps build sustainable online safety ecosystems across Africa.

5.6. Conclusion : Key Takeaways of International Trends to Address TFGBV

Table 5.2 Key Trends in regulation of TFGBV

Country	Legal Instrument	Approach	Takeaways
United Kingdom	*Online Safety Act*	**Individualised harms** which are defined as 'physical or psychological harm'	• Narrow definition **excludes broad social impact of online harms,** overlooking its complex and non-linear nature. • Eg. individual impact of self-harm v. algorithm-driven content recommendations normalising self-harm online
European Union	*Digital Services Act*	**Systemic risks,** including GBV	• Requires VLOPs to assess proactively: **recommender systems, content moderation algorithms, terms and conditions, advertising selection systems, or data-related practice** • More holistic understanding of the complex nature of online harms
Australia	*Online Safety Act*	**Gender-neutral** legislation regulating adult cyber-abuse	• **High bar** for "serious harm", i.e. serious psychological harm or physical harm • Many GBV cases, like abusive content targeting women of **colour** or public figures would not be covered
Mexico	*Olimpia Law*	**Gendered law** on digital violence	• Express recognition of digital GBV led to **more awareness and sensitivity** towards survivors of online harms • **Training** for law enforcement **improved perception** of online sexual violence

[563] *Revenge Porn Helpline,* https://revengepornhelpline.org.uk/.

[564] *Hate Aid,* https://hateaid.org/en/.

[565] *The Africa Online Safety Platform,* https://www.africaonlinesafety.com/.

[566] European Parliament, *Answer given by Mr Breton on behalf of the European Commission*, 2022, https://www.europarl.europa.eu/doceo/document/E-9-2022-003143-ASW_EN.pdf.

5.6.1. Legal Policies and Design

Gender neutral v. Gender sensitive laws: Some legislations, such as Australia's Online Safety Act, while being tailored for digital spaces, continue to remain gender-neutral while addressing cyber abuse. On the contrary, Mexico's Olimpia Law specifically recognizes the gendered nature of online harms and provides a framework for prevention, response, and elimination, embedding gender sensitivity within legislative and enforcement design. **It was reported that the inclusion of gender-sensitive provisions in law led to better enforcement outcomes in detecting, reporting and handing of TFGBV.**

Systemic Harms v. Individual Harms: One area of divergence is in the nature of harms that are sought to be addressed by different legislations. While the UK's OSA provides a narrow approach limited to physical or psychological harms directed at individuals, the EU's DSA seeks to address systemic risks, while specifically providing for risk assessments to be carried out for gender-based violence.

Mandatory v. Voluntary Compliance by Platforms: In some jurisdictions, platforms are legally required to follow specific safety rules, while in others, they voluntarily adopt such measures without a legal obligation. The EU Digital Services Act requires VLOPs to carry out risk assessments (including systemic risks like gender-based violence) and this is a legally binding requirement with remedial action plans, fines and interim measures for non-compliance.[566] While Australia's OSA provides penalties for non-compliance with removal notices or industry codes/ standards, the penalty is quite low and makes no distinction between various types of offences (**e.g. maximum penalty for failure to remove child sexual exploitation material is the same as for failure to take down harmful but not unlawful material**). In some instances, like the Australian Online Safety Code for Dating Services, platforms voluntarily implement collectively agreed baseline measures.

5.6.2. Enforcement Practices

Approaches to regulation: Countries use different approaches to regulate online platforms, which can be seen on a spectrum from self-regulation to co-regulation. **The UK's OSA follows a co-regulation model:** platforms must assess and manage risks, while the regulator, Ofcom, sets out codes and guidance and monitors compliance. **Similarly, Australia's OSA combines self-regulation**—where digital services develop their own codes—**with oversight from the eSafety Commissioner**, who approves the codes and ensures compliance. By encouraging cooperation between government, industry and civil society through models like co-regulation or self-regulation, **policymakers (with the right safeguards that prevent regulatory capture) can leverage the strengths of all actors to build a safer online space**.

Regulator's enforcement powers: Entities like Australia's eSafety Commissioner do not have formal investigative authority but leverage relationships with platforms for voluntary removals. While the UK's Ofcom has limited enforcement power and prefers to encourage voluntary compliance with the OSA, it will launch investigatory and enforcement actions (including imposing fine, requiring companies to utilise certain accredited technologies or seeking a court order imposing business disruption measures in case of serious violations) when deemed necessary.[567]

Role of Law Enforcement Agencies: The insensitive attitudes of LEAs often lead to underreporting by survivors. The involvement of law enforcement agencies is vastly different across jurisdictions. For instance, in Australia, it is only when the abuse reaches a

[563] *Revenge Porn Helpline,* https://revengepornhelpline.org.uk/.

[564] *Hate Aid,* https://hateaid.org/en/.

[565] *The Africa Online Safety Platform,* https://www.africaonlinesafety.com/.

[566] European Parliament, *Answer given by Mr Breton on behalf of the European Commission*, 2022, https://www.europarl.europa.eu/doceo/document/E-9-2022-003143-ASW_EN.pdf.

[567] Ofcom, *Online safety enforcement guidance*, 2024, https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/online-safety-enforcement-guidance.pdf?v=391925.

certain threshold that the eSafety Commissioner recommends the survivor to the relevant agency. In other jurisdictions, the criminalisation of certain online harms against women leads to law enforcement agencies being more involved. **A critical factor in improving law enforcement is training and awareness. In Mexico, for example, it was reported that training helped officers change their perceptions on online harms, enabling a more sensitive approach.**

5.6.3. Access to Justice for Survivors

Support Mechanisms: South Korea's Digital Sex Crime Victim Support Center offers comprehensive support for women and survivors of online abuse, focusing on training, counselling, and legal and medical aid support. The UK's Revenge Porn Helpline also helps to provide preventive support and boasts a 90% removal success rate.

Lacking Survivor-Centric Approach: Legislation like Australia's Online Safety Act has a very high threshold for a material to qualify as harmful (e.g., serious harm must be demonstrated), excluding several distressing but less severe forms of abuse. Some legislations only allow victims to file complaints and not any member of the general public - causing significant delay and amplifying damage. Moreover, law enforcement, content moderators and judiciary personnel across jurisdictions often lack adequate training to handle cases involving gender-based online violence sensitively.

Intersectional Challenges: Existing legal frameworks and risk assessment do not account for the intersectional nature of discrimination or abuse faced by women from marginalized communities, women in political roles, journalists and human rights defenders.

Remedies Beyond Law: There's a growing call to adopt more holistic measures. The UN emphasises a mix of restitution, rehabilitation, and symbolic measures to address survivors' needs comprehensively. Beyond formal laws, the involvement of civil society organisations becomes essential. They not only provide crucial support services for survivors, but also drive public awareness and promote online safety by challenging harmful gender norms and stereotypes - which are the underlying drivers of TFGBV.

Chapter 6
Recommendations

India needs to revisit its reactive approach to online safety of women and children. This includes a careful consideration of policy design and enforcement practices. A systemic overhaul must suitably address the *pacing problem* of technology regulation and keep adapting to the evolving nature of digital risks. To that end, the recommendations in this chapter outline a comprehensive whole-of-ecosystem approach to overall online safety for women and children in India. **The recommendations promote a proactive and systemic framework** grounded in:

- **Strengthening data collection and modernising cybercrime and online harm classification frameworks** to better profile online risks faced by women and children;
- **Evaluating the need for international best practices** such as risk assessments, risk mitigation, harm reduction and platform accountability;
- Inserting **gender-sensitive provisions and frameworks** into India's criminal enforcement frameworks to account for the disproportionate impact of online abuse on women and girls;
- Emphasising the **criticality of clear definitions of the risks that kids and adolescents face online** across both clearly illegal and legally amorphous safety challenges;
- Investing in **more enabling, technically capable, streamlined and sensitive enforcement practices** across LEAs and the judiciary;
- **Empowering local civil society organisations (CSOs)** to contribute more effectively towards the online safety of women and children;
- **Conducting multi-stakeholder consultation for the creation of codes of practices to promote an ecosystem-wide culture of safety by design;** premised on collaboration, shared responsibility, safer and inclusive platform design, transparency, user empowerment, and local relevance; and

- **Investing in survivor-centred mechanisms for rehabilitation and reintegration** into digital economies.

The recommendations also call for more inclusive and participatory policymaking that meaningfully involve on-ground experts, as well as affected demographics.

6.1. Effective Measurement and Modernising Cybercrime and Online Risk Classification Frameworks

Appropriate measurement of online risks is essential for sound policy interventions. It helps decision makers understand the scope of challenges, **efficiently allocate ecosystem resources in addressing the identified risks, and pursue targeted interventions** that advance online safety.[568] Such data and attendant insights can be useful for policymakers, platforms, researchers, and civil society in **identifying evidence-based solutions**. Such data can help decision-makers determine when interventions should prioritise legislative or regulatory reform; institutional overhaul; or stakeholder partnerships.

6.1.1 Need for a robust taxonomy for consistency

Data quality issues stem from limited legal and policy clarity, and consistency in how online harms including cybercrimes are defined and recorded within government systems. **India needs to develop a robust taxonomy for online harms targeted at specific groups like women and children.** Without such a taxonomy, responses to cybercrimes and lower level risks will remain inadequate, inconsistent, and reactive, even potentially leading to improper understandings of prevailing risks.[569] Taxonomies on online risks should be able to distinguish between risks that are clearly illegal in nature, as compared to harmful content or conduct that is legally amorphous. **They can be informed by large-scale survey-based studies by government and civil society**. Additionally, India's taxonomy of online risks should be informed by transparency disclosures by online platforms that are made accessible to the public; and periodic stakeholder consultations. With respect to platform transparency disclosures, efforts should be made to standardise these practices to the best extent possible with respect to online safety risks for women and children. A combination of these activities can help Indian policies **determine the prevalence and impact of niche issues like (i) promotion of eating disorder materials and ensuing body image challenges; (ii) issues relating to screen time; and (iii) challenges faced by public figures from gender minorities (e.g. women) due to mass online trolling**.

6.1.2 Need for disaggregated data

Data published by government agencies and other civil society research organizations should facilitate disaggregated analysis across gender, caste, religion, sexuality, region, age, and other identity markers. Disaggregating data helps uncover hidden patterns of abuse, highlights vulnerable populations, and makes these groups more visible to policymakers.[570] **There is a need for the government to strengthen data collection through the National Family Health Survey (NFHS) and other similar mechanisms.** Equally important is the need to study harmful online content that reflects the lived experiences of marginalised groups, particularly Dalit, Bahujan, Adivasi, and tribal women. Traditional surveys often fail to capture the full extent of online risks faced by these communities and therefore, participatory research methods should be prioritised, involving grassroots organisations.

---

[568] World Economic Forum (WEF), Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms, 2023, https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf.

[569] Akhilesh Chandra and Melissa J. Snowe, *A taxonomy of cybercrime: Theory and design,* 2020, https://www.sciencedirect.com/science/article/abs/pii/S1467089520300348.

[570] National Collaborative Centre for Aboriginal Health, *The Importance of Disaggregated Data,* https://www.nccih.ca/docs/context/FS-ImportanceDisaggregatedData-EN.pdf

6.1.3 Revisiting the 'principal offence rule' for online crimes

Additionally, India's 'principal offence rule' should be relaxed in cases involving online crimes. Such interventions will improve India's understanding of how digital tools exacerbate traditional crimes. Indian legal systems should consider efforts at multilateral institutions like the United Nations where the international community has been deliberating a global cybercrime treaty. **There is merit in Indian frameworks embracing the UN Cybercrime Convention's efforts at specifically identifying "cyber-enabled offences" as a specific category of cybercrime**. Cyber-enabled offences are described as traditional criminal offences that are enabled by digital tools.[571] With the **recognition of "cyber-enabled offences" within existing Indian laws**, there could be better opportunities to capture data on the volume and scope of the challenges faced by particular groups like women and children, without institutions being limited by legacy concepts like the ***principal offence rule***.

6.2 Preventive Measures Based on International Best Practices

6.2.1 Benefits of enabling proactive systemic risk assessments, disclosures and risk mitigations

India's recent attempts at addressing online safety through platform regulation merits revisiting. Online safety at the platform level is primarily pursued through the intermediary liability ("IL") and safe harbor protection as prescribed under **Section 79 of the Information Technology Act and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("IT Rules")**, that are periodically updated.

The 2021 IT Rules themselves are an overhaul of the previous IL and safe harbour regime that was enforced through the Information Technology (Intermediaries guidelines) Rules, 2011. Under the 2011 framework– that was **refined by the Supreme Court's landmark decision in the *Shreya Singhal case***– intermediaries were essentially supposed to communicate to third-party users that under their platforms' terms and conditions users were not allowed to host various categories of harmful and unlawful content; and that non-compliance could lead to service termination. **This was combined with an obligation for intermediaries to remove unlawful content expeditiously once they received an appropriate directive from a government agency or judicial authority**.

However, in the government's recent quest towards promoting online safety over digital services, **the IT Rules 2021 has gradually shifted towards a proactive monitoring and content regime which is at odds with the Supreme Court's observations in *Shreya Singhal*.** Notably, the Court in *Shreya Singhal* had held that legal requirements for proactive monitoring of individual posts is inconsistent with the fundamental **right to free speech and expression that is guaranteed under Article 19(1)(a)** of the Constitution. **Thus, intermediaries' content takedown requirements of individual posts should only stem from valid government or court order(s), designating a piece of content as unlawful.**

Nevertheless the IT Rules, 2021 has pushed forward an agenda to nudge online intermediaries (including significant social media intermediaries) to proactively monitor and remove unlawful content. The 2021 Rules leverage creative drafting around the idea of positive obligations to adopt *"reasonable efforts"* and best efforts "endeavors". However, the actual instances where the enforcements of these provisions have arisen, have been in *ex-post* settings after a major flash point or incidents that gained mainstream publicity.[572] Prominent examples include the use of **informal pressures via government advisories[573]** to push for proactive

[571] UN Peace and Security, *Basic facts about the global cybercrime treaty,* 2024, https://www.un.org/en/peace-and-security/basic-facts-about-global-cybercrime-treaty.

[572] Vasudev Devadasan, *Conceptualising India's Safe Harbour in the Era of Platform Governance*, 2024, https://repository.nls.ac.in/cgi/viewcontent.cgi?article=1420&context=ijlt.

cooperation from platforms in the *Rashmika Mandana deepfake incident*[574], and another was when malicious actors were violating the anonymity of the victim in the *RG Kar* case.[575]

Content takedown frameworks often target issues around law and order, public order, public safety and national security. **Thus, what we see is that India's regulatory framework is largely a reactive framework that may not effectively protect the interests of vulnerable groups and individuals. Hence, tackling problems at a wider systemic scale, and allocating resources towards positive preventive measures could be a beneficial solution.** The broader systemic factors that enable harm against vulnerable groups like women and children are therefore largely unaddressed.

This merits **prioritizing systemic risk assessments, particularly for harms related to children's online safety and technology-facilitated gender-based violence (TFGBV). This should be coupled with robust risk mitigation measures, including clearer obligations on platforms for effective content moderation and takedown procedures (especially in time sensitive cases), survivor-centric reporting mechanisms, and structured information-sharing protocols with civil society organizations and other ecosystem partners. The risk assessments should serve the following key purposes[576]**:

- **Create accountability** for the negative impacts platforms can have on people and societies and establish best practices to evaluate and address these impacts;
- **Incentivise best practices** in platform and algorithmic design, encouraging companies to proactively minimize harm through thoughtful and responsible design choices;
- **Inform the public** about the risks and harms associated with platform operations, enabling civil society to take informed steps to protect users and demand better standards.

Given the limited regulatory and enforcement capacity to monitor and audit information disclosed by platforms in India, public participation should also be encouraged.[577] Platforms could also consider publishing periodic risk assessments using standardised and comparable metrics that are accessible to external researchers, experts, and civil society groups to allow for independent scrutiny and informed public discourse.

As discussed in Chapters 4 and 5 of the report, an approach revolving around risk assessments and mitigations aligns well with global practices already adopted under the **EU Digital Services Act (DSA) and the UK Online Safety Act (OSA)**. India could consider adapting elements from the **EU DSA**, which mandates VLOPs to provide access to "vetted researchers" for conducting research that contributes to the "detection, identification and understanding of systemic risks in the Union"(scope of which is outlined in Article 34(1) and noted above).

Finally, platforms' content moderation performance, trust and safety interventions, and design choices should be **benchmarked against their own risk assessments**, ensuring that mitigation efforts are proportionate, evidence-based, and continuously

---

[573] Genevieve Lakier, *Informal Government Coercion and the Problem of "Jawboning"*, 2021, https://www.lawfaremedia.org/article/informal-government-coercion-and-problem-jawboning.

[574] Mashable India, *Amid Rashmika Mandanna Deepfake Row Government Issues Advisory For Social Media: 1 Lakh Fine And 3 Years Jail*, 2023, https://in.mashable.com/tech/63436/amid-rashmika-mandanna-deepfake-row-government-issues-advisory-for-social-media-1-lakh-fine-and-3-ye.

[575] PIB Press Release, *Social Media Platforms to comply with Supreme Court order on removal of deceased's identity in RG Kar Medical college incident*, 2024, https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2047347.

[576] Integrity Institute, *Overview of Online Social Platform Transparency,* 2023, https://integrityinstitute.org/news/institute-news/integrity-institute-releases-overview-of-online-social-platform-transparency.

[577] Centre for Communication Governance (CCG), *Platform Transparency under the EU's Digital Services Act: Opportunities and Challenges for the Global South, 2025,* https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/ccg-nlud-platform-transparency-eu-dsa-gschallenges-738.pdf

improved. Embedding such an approach would foster a safety-by-design ethos, encouraging platforms to anticipate and prevent harm rather than react to it. This would also enhance **legal certainty for platforms**, while improving **transparency and accountability** for civil society actors and users—especially those most vulnerable to online harms.

6.2.2. Gender-Sensitive Provisions in Cybercrime Legal Frameworks

Ignoring the gendered nature of online abuse can weaken efforts toward gender equality and digital safety. **Incorporating gender-sensitive provisions into the law can lead to more effective enforcement, better identification and reporting of these harms, enhanced victim support and more targeted resource allocation.**[578] As seen in Mexico's Olimpia Law[579], the formal recognition of gendered offences in the statute helped authorities identify and respond to such acts, leading to more consistent enforcement and support for victims. **Therefore, we recommend two key interventions:**

1. **Work towards developing a comprehensive definition of TFGBV**: There is no single, universally accepted definition of TFGBV, instead a range of institutions and jurisdictions have articulated it differently, reflecting varying priorities and contexts. The United Nations Population Fund defines TFGBV as "*an act of violence perpetrated by one or more individuals that is committed, assisted, aggravated and amplified in part or fully by the use of information and communication technologies or digital media against a person on the basis of gender*".[580] The UN Women in November 2022 convened a diverse set of global experts to develop a more expansive definition of TFGBV which is "*any act, that is committed, assisted, aggravated or amplified by the use of ICTs or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political or economic harm, or other infringements of rights and freedoms*".[581] At the national level, the Mexican law offers a definition for digital violence as "*any action carried out through the use of printed materials, e-mail, text messages SMS, social media, internet platforms, or any technological device through which images, audios or videos of intimate sexual content of a person are obtained, exposed, distributed, disseminated, exhibited, reproduced, transmitted, commercialized, offered, exchanged and shared without their consent; that violates the integrity, dignity, privacy, freedom, and private life of women or causes psychological, economic, or sexual harm in the private or public sphere, as well as moral harm, to them and their families*".[582]

   These definitions illustrate **different legal and policy approaches**: (a) **scope of harm**: Some definitions (like UN Women's) are broad and include all forms of harm (psychological, social, economic, political), while others (like Mexico's) focus more narrowly on specific acts such as non-consensual image sharing; (b) **target group**: While Mexico's definition centers only on women, others refer to "gender" allowing inclusion of gender-diverse and gender-nonconforming individuals — a crucial distinction in contexts like India, where trans and

[578] UNDP, *Analysis of the legislation related to Technology Facilitated Gender Based Violence*, 2024, https://www.undp.org/sites/g/files/zskgke326/files/2024-12/final-analysis-tf-gbv.pdf.

[579] Chapter V, Olimpia Law.

[580] https://www.unfpa.org/TFGBV

[581] (UN Women, 2023, Technology Facilitated Violence against Women-Report of the Foundational Meeting of the Expert Group).

[582] Juan Carlos Tornel, *Government of Mexico publishes Digital Violence Act*, 2020, https://www.globalcompliancenews.com/2020/03/16/government-of-mexico-publishes-digital-violence-act/#:~:text=On%20January%2022nd%2C%202020%2C%20the%20Government%20of%20Mexico,Access%20to%20a%20Violence-Free%20Life%20in%20Mexico%20City.

queer persons face systemic online abuse; (c) **focus on consent**: The Mexican law emphasizes consent and privacy, while the UN definitions focus on technology as an amplifier of harm. In designing a definition of TFGBV for the Indian context, it will be important to: recognize both cyber-dependent and cyber-enabled harms, including actions that may appear benign—such as circulating photos of a girl and boy together—can have serious consequences for the girl and her family, especially in conservative or patriarchal settings; acknowledge that gender-based violence is not limited to cisgender women, and explicitly include transgender and gender-nonconforming individuals; and ensure the definition is technology-neutral, allowing it to evolve with emerging digital tools.

This must be accompanied by comprehensive capacity-building efforts, including: training of law enforcement, judiciary, and regulatory bodies to recognize and respond to TFGBV, sensitization of platform moderators and grievance officers, and awareness campaigns aimed at the general public to promote responsible digital behavior and encourage reporting. Without institutional alignment and shared understanding across actors involved in identification, prevention, investigation, and enforcement, even the most well-intentioned law may fail to offer real protection. **Embedding the definition within a broader programmatic framework ensures not just clarity in law, but consistency in practice**.

2. **Incorporate Gender-Sensitive Provisions**: To better capture the full spectrum of online harms, especially emerging threats like deepfakes, digital impersonation, and coordinated trolling (i.e. networked harassment), the IT Act should be expanded to include **gender-sensitive offences**. This would explicitly criminalize (or regulate) acts that disproportionately target women, trans, and gender-nonconforming individuals, and provide stronger legal recourse. Reports on the basis of calls received from Meri Trustline highlight that cyber-bullying, non-consensual intimate image-sharing (for sextortion and blackmail), impersonation, and deepfake are already significant and gendered online harms.[583] Incorporating these provisions would recognize and address new forms of TFGBV, signal a strong legal stance against orchestrated online harms targeting specific genders, and support effective investigation and prosecution by LEAs.

6.2.3. Definitions for Online Harms involving Children

Regarding children's online safety we need clear, precise and constitutionally consistent definitions of online risks/harms. Among other things the definitions cannot be overbroad and inadvertently cause a, constitutionally incompatible, chilling effect on people's legitimate online speech. Additionally, it is important to distinguish between harms associated with **illegal criminal offences** (e.g. child sexual exploitation and abuse material or incitement to violence) or **legal but harmful risks** that are not necessarily illegal but have a detrimental impact on children and adolescents. An adequate distinction between these concepts and their sub-categories allows for more tailored policy and online safety interventions that can eventually lead to more proportionate outcomes. In this regard, two prominent international models present instructive paradigms for consideration:

1. **The United Kingdom's Tiered and Categorical Approach (Online Safety Act 2023):** The UK's OSA employs a sophisticated, tiered system for content categorization. This includes specific categories within the framework for "Illegal Content" (material already unlawful), "Primary Priority Content Harmful to Children," "Priority Content Harmful to Children" and a broader, residual category, "Non-designated content that is harmful to children".

[583] Rati Foundation, *Meri Trustline: Annual Report Volume 2*, 2025, https://ratifoundation.org/wp-content/uploads/2025/05/Meri-Trustline-Year-2-Report-Final-Version_compressed.pdf.

Trade-offs for India: This approach offers significant benefits in its granular and prescriptive framework, furnishing platforms with clear directives regarding their duties concerning specific types of content. Such a structured methodology can facilitate highly targeted safety interventions and transparent regulatory expectations. However, the comprehensive definition of "legal but harmful" content may invite concerns regarding potential overreach or governmental influence over online speech, thereby necessitating precise drafting to withstand scrutiny under Article 19(1)(a) of the Indian Constitution. Furthermore, its effective implementation would demand a specialised enforcement institution that is endowed with substantial expertise and suitable independence from the executive.

2. **Australia's Incident-Based Approach (Online Safety Act 2021):** In contrast, Australia's Online Safety Act primarily adopts an "incident-based" approach, particularly concerning "class 1 and class 2 material" (seriously harmful content). While it defines categories such as "cyberbullying material" and other forms of "seriously harmful content" (e.g., non-consensual intimate images, content promoting terrorist acts), its eSafety Commissioner typically intervenes in response to specific instances of harm through takedown notices following a report. This methodology prioritizes addressing the *effect* of the material on the individual.

   Trade-offs for India: This approach offers greater flexibility and may be perceived as less susceptible to free speech concerns, as it is predicated upon demonstrated harm to an individual. This might align more closely with Indian constitutional principles that emphasize reasonable restrictions on expression. **Conversely, its efficacy can be more reactive than proactive, relying upon victim reporting before remedial action is initiated.** It also necessitates substantial resources for individual case investigation and enforcement, and may prove less effective in addressing systemic issues of widespread harmful content exposure.

Irrespective of the chosen regulatory model, the precise definition of "legal but harmful" content will be indispensable for an effective digital safety framework. Its explicit categorization is vital for several reasons:

- **Enabling Targeted Interventions:** By delineating specific types of harm (e.g., self-harm promotion versus bullying), platforms, parents, and educators can implement contextually appropriate and tailored safety measures. A comprehensive tiered system of access control, for instance, allows for outright prohibition for the most severe harms (e.g., Primary Priority Content), while enabling age-appropriate access controls, content warnings, or educational interventions for less severe "Priority Content."
- **Fostering Shared Understanding:** Clearly articulated definitions provide a common vocabulary and analytical framework for all stakeholders – individuals, families, educational institutions, digital platforms, and law enforcement agencies. This clarity empowers informed decision-making regarding escalation pathways: whether an issue can be effectively resolved at the platform level through content moderation, within the family and school environment via digital literacy programs, or if it necessitates intervention from law enforcement agencies.
- **Driving Platform Accountability:** Specific and unambiguous definitions translate directly into measurable duties and responsibilities for digital platforms, thereby enabling regulators to evaluate platforms' design choices, their design

choices, content moderation practices, and algorithmic amplification mechanisms.

- **Shaping Societal Norms:** Legal frameworks play a profound role in shaping societal norms and expectations. Articulating these distinctions contributes to a broader public understanding of online child safety, fostering responsible digital citizenship and cultivating a pervasive culture of online safety.

6.2.4. Allow lower-ranking (first responder) officers to lead cybercrime investigation

Restrictions under Sections 78 and 80 of the IT Act limit the investigation of cybercrimes to police officers of the rank of inspector and above.[584] This creates bottlenecks, as there are fewer officers at these ranks, while lower-ranking officers (who are often the first responders) are not authorised to lead investigations. **As a result, cases are often delayed and officers are incentivised to register cases under general criminal provisions instead of the relevant cybercrime sections**.

To address this, IT Act 2000 should permit specially trained junior-ranking officers to investigate cybercrimes. This would help improve response time, reduce delays, reduce evidentiary burdens to take case forward and ensure cases are registered under the appropriate legal provisions.

6.3. Reforming India's Law Enforcement Practices

Cybercrimes often involve complex technological systems and networks, making it necessary for law enforcement agencies to have specialized technical knowledge to investigate.[585] Jurisdictional issues also complicate investigations, as cybercrimes often span state and national borders. Inter-state coordination is limited, with only a few states, like Jharkhand, having mechanisms in place for cooperation. The absence of a uniform standard operating procedure (SOP) across states leads to inconsistent handling of cases and delays. Additionally, online crimes are often not taken as seriously as offline ones. As noted earlier, qualitative findings from West Bengal reveal that police often dismiss women's complaints of online harassment, advising them to block the perpetrator rather than initiating formal investigations—frequently recording such cases as general diary entries instead of FIRs.[586] These issues are worsened by gender bias, stereotyping, and a lack of sensitivity within law enforcement and the judiciary. Victims are also burdened with collecting their own evidence, which is particularly difficult in cybercrime cases. Together, these factors result in long, emotionally draining, and costly processes that often re-victimise complainants rather than providing them with justice. It is therefore important to implement SOPs for cybercrime cases involving women and children.

6.3.1. Implement Uniform Standard Operating Procedure (SOP)

Currently, there exists a patchwork of resources such as the Cyber Crime Investigation Manual (published by the Data Security Council of India)[587], periodic advisories issued by the Ministry of Home Affairs (MHA)[588], and SOPs issued by separate states (like the Handbook prepared by Delhi police outlining the process of obtaining data disclosure from social media intermediaries) or specific to certain offences.

While these documents provide valuable guidance, the absence of a **uniform model SOP for cybercrime cases involving women and children**

[584] Sections 78 and 80, Information Technology Act, 2000.

[585] Vivek Sharma, Dr. Hemant Kumar Harti, *Need for Imparting Training to Officials to Investigate Cyber Crimes,* 2021, https://iajesm.in/admin/papers/648dbf9b81369.pdf.

[586] Anita Gurumurthy and Amrita Vasudevan, *Hidden figures- A look at technology-mediated violence against women in India*, 2018, https://itforchange.net/index.php/hidden-figures-a-look-at-technology-mediated-violence-against-women-india.

[587] Data Security Council of India (DSCI), *Cyber Crime Investigation Manual,* 2011, https://jhpolice.gov.in/sites/default/files/documents-reports/jhpolice_cyber_crime_investigation_manual.pdf.

[588] Ministry of Home Affairs, Government of India, *Advisory on Preventing & Combating Cyber Crime against Children,* 2012, https://www.mha.gov.in/sites/default/files/CS-Adv-160112.pdf.

results in inconsistent and ineffective responses across various states. To address this, we recommend that **India's Ministry of Home Affairs and the Indian Cyber Crime Coordination Centre (I4C) collaborate with industry and other stakeholders** to develop a model uniform SOP on cybercrime investigation that may be applicable across all states and union territories. The SOP should provide clear and detailed guidelines for LEAs on the investigation of technology-facilitated gender-based violence and cyber harms against children. It should be gender-sensitive and developed in **close consultation with civil society organizations working with children and women**.

Additionally, the SOP should outline a **framework for inter-state cooperation**, enabling LEAs in different states and union territories to coordinate seamlessly during investigations, including clear protocols for communication and mutual assistance.[589] Furthermore, the SOP should specify the conditions and processes under which LEAs may engage certified external technical experts to support forensic and electronic evidence collection for cybercrime investigations.

It should also establish standardised procedures for data requests made to digital platforms, detailing the types of information LEAs can seek, such as account interactions, login/ logout records, conversations, and social connections. The SOP must set clear response timelines and escalation mechanisms to address any delays or refusals by platforms, drawing from existing models like the I4C's SOP that govern information sharing between financial institutions and law enforcement agencies in economic offences cases[590]. **This will help streamline cooperation, accelerate evidence gathering, and ensure timely justice for victims.**

6.3.2. Mandatory Training and Gender-Sensitisation for Key Stakeholders

The government has already put many important systems in place to protect children and women from violence and cybercrimes. Under the Juvenile Justice (Care and Protection of Children) Act, 2009, every police station is required to have at least one designated **Child Welfare Police Officer (CWPO)** who receives appropriate training and orientation to exclusively handle cases involving children, whether as victims or perpetrators, in coordination with non-governmental organizations.[591] Additionally, State Governments must constitute **Special Juvenile Police Units (SJPUs) in each district and city, comprising all CWPOs and two social workers**.[592] These personnel require specialized sensitisation and training tailored to address the nuances of cybercrime cases involving children. Similarly, **social workers and support personnel managing helplines and Police Control Rooms** should receive focused training to enhance their ability to assist victims effectively.

Separately, the government has **already established several women-centric initiatives[593] aimed at supporting women affected by violence, such as One Stop Centres**, which provide integrated services including medical aid, legal assistance, temporary shelter, police support, and psycho-social counseling. To make police stations more approachable and women-friendly, **Women Help Desks have also been set up,** which serve as the primary point of contact for women in distress. Additionally, the Emergency Response Support System facilitates rapid, computer-aided deployment of police resources, complemented by a dedicated **Women Helpline (WHL-181)**. It is essential to leverage these existing frameworks by ensuring that all personnel involved are sensitised and trained specifically to address cybercrimes against women with appropriate awareness and empathy.

---

[589] As noted above,while I4C has constituted seven Joint Cyber Crime Coordination Teams (JCCTs) to enhance inter-state coordination, there are no specialised JCCTs focused specifically on issues affecting women and children.

[590] Cyber Crime Cell, Puducherry Police, *Cyber Crimes Investigation Guidelines*, 2024, https://police.py.gov.in/Cyber%20Crime%20-%201%20Investigation%20check%20list%20by%20Dr%20Bascarane%20SP%20Cyber%20dt%2014.02.24.pdf

[591] Section 107(1), the Juvenile Justice (Care and Protection of Children) Act, 2009.

[592] Section 107(1), the Juvenile Justice (Care and Protection of Children) Act, 2009.

[593] PIB Press Release, 2025, https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2112763.

Police officers, in particular, must be **sensitised to take online harms as seriously as offline offenses** and be directed to register First Information Reports (FIRs), where it is legally required to do so. Additionally, **regular training on forensic and technological capacities should be incorporated to strengthen the overall investigative and support mechanisms for cybercrime cases** involving women and children.

6.4. Leveraging CSOs in tackling tech-facilitated violence against women and children

Civil Society Organizations (CSOs) play an important role in promoting digital literacy and online safety, undertaking initiatives to train LEAs and working closely with communities. Many CSOs also run helplines addressing technology-facilitated violence and, in doing so, gather dynamic intelligence on the experiences of victims and survivors across different parts of India. **They hold crucial on-ground, contextual expertise on the specific threats and challenges faced by marginalized groups.** Organizations such as Point of View (PoV), Rati Foundation, Icall, Space2Groww and many others serve as key examples of this vital work. PoV runs a helpline TechSakhi where responders answer queries on digital safety and online abuse or violence, while also running dedicated programs for LGBTIQA+ and other marginalised communities (like sex workers, domestic workers) equipping them with tools such as digital security, art healing, peer support training, digital strategising, among others to address online and offline threats[594]. The Rati Foundation runs "Meri Trustline" a helpline for children, women, and people from marginalized gender identities at risk of online harm, in addition to providing legal aid, rehabilitation and psycho-social support for victims.[595] Space2Groww focuses on digital rights for children by working with companies, policymakers, parents, educators, and caregivers, running digital literacy programs, digital safety initiatives and conducting social audits that address online threats.[596]

Even though they do valuable work, such as training LEAs, supporting survivors, and advising on policies, they **face a lack of formal recognition and institutional support.** Civil society organizations also face serious limits in their ability to raise resources, often relying heavily on funding from platforms themselves or other donors.[597] **Many CSOs step in when formal systems fail, however, their efforts remain largely informal and under-resourced, limiting their potential impact. To strengthen their functioning, CSOs need institutional support to bring insights from the grassroots to the top decision makers and key stakeholders.**

6.4.1. Institutionalising CSOs in the digital safety ecosystem

CSOs should be formally recognised as key partners in grievance redressal, education, and survivor support. MeitY and platforms should engage with them through liaison roles and advisory panels to integrate their expertise into policy and platform design. Policymakers and platforms should allocate dedicated budgets and resources to support CSOs in addressing online harms against women and children. This could include funding for capacity building and research. Moreover, there is a need to identify strategies which can pool government, philanthropic and industry resources for CSO-led online safety initiatives targeted at specific women and children communities.

6.4.2. Awareness and Digital Literacy Promotion

There is a need for locally relevant content in regional languages to raise awareness and build digital literacy.[598]

[594] Point of View, *Programmes | Gender and Technology*, https://pointofview.org/gender-technology/.

[595] Rati Foundation, *Meri Trustline,* https://ratifoundation.org/meri-trustline/.

[596] Space 2 Groww, *Digital Safety of Children: Creating Safe Online Spaces*, 2023, https://www.dqindia.com/news/space2grow-releases-report-on-digital-safety-of-children-8629275.

[597] Government of UK, *Department of Science, Innovation and Technology, Platform Design and the Risk of Online Violence against Women and Girls (Online VAWG)*, 2024, https://assets.publishing.service.gov.uk/media/67a39e2cad556423b636cadd/Platform_design_risk_of_online_violence_against_women_girls_A.pdf?utm_source=chatgpt.com.

[598] The United Indian, *Empowering the Nation: Strategies for Enhancing Digital Literacy in India,* 2024, https://theunitedindian.com/news/blog?Digital-literacy-in-India&b=193&c=3.

Children often fail to identify online harms due to limited awareness.[599] Using video explainers, story formats, and mobile-friendly designs can enhance outreach. **Community-based approaches, such as sports programs, local theatre, and legal aid clinics in schools, colleges, and communities, can raise awareness and offer legal aid, psychological support and rehabilitation to victims of online harm.** Awareness campaigns must also focus on informing girls of their digital rights and available support systems, with a specific focus on peer-led support systems. Since parents and teachers are often not the primary trusted contacts, senior peers and mentors should lead safety dialogues. Additionally, CSOs should disseminate clear guidance and assistance to users and survivors on gathering and preserving evidence of online harms.

6.4.3. Participatory and consultative decision-making process

Encourage consultative and participatory research that involves children and women from different communities in identifying online harm risks and creating solutions, in collaboration with CSOs. **CSOs can leverage their networks to facilitate intergenerational dialogues[600], involving teenagers, children, caregivers, parents, child safety organisations and educators, among others, to understand different views on online safety and shape better policies. Globally, examples like the UK Youth Parliament[601], UNICEF Innocenti's Global Youth Network[602], the Digital Futures Commission[603], and Australia's online safety youth advisory council[604] highlight the positive impact of directly involving youth in research, policy design, and digital safety.** Additionally, multistakeholder councils or coalitions involving NCW, NCPCR, state police nodal officers, and relevant officials from different government ministries and civil society organisations must be organised to build capacity, raise awareness, and ensure consistency across stakeholders.

6.4.4. Role of CSOs in Reporting Complaints and Offering Survivor Support

Platforms often invest in **trusted flagger programs that they tailor** to specifically collaborate with CSOs and researchers working on TFGBV and child safety. **Such partnerships can fast track the detection of niche, contextually relevant harmful content and online abuse. Such CSO networks** stand to benefit from formal recognition and eventual certification as **independent grievance intermediaries**. Certified CSOs can subsequently serve as formal partners of ecosystem coordination efforts where they can contribute by assessing harmful content, offering survivor-centric reporting options, and sharing key trends/information with platforms and LEAs in a timely manner.

Furthermore, **platforms and LEAs should partner with CSOs to establish dedicated victim support units staffed with forensic experts to assist survivors in collecting digital evidence.** In many rural areas, **CSOs also provide critical survivor support outside formal legal systems, often acting as first responders through existing networks such as ASHA workers[605] (community-based health workers).**

---

[599] Muvija M, *UK children exposed to violent content online, see it as 'inevitable', report finds*, 2024, https://www.reuters.com/world/uk/uk-children-exposed-violent-content-online-see-it-inevitable-report-finds-2024-03-15/?utm_source=chatgpt.com.

[600] Intergenerational dialogues are interactive participatory forums that bring together older and younger generations and are intended to create shared knowledge and meaning and a collective experience. See: https://tciurbanhealth.org/courses/global-community-group-engagement/lessons/intergenerational-dialogue-2/.

[601] National Youth Agency, *Youth Parliament Manifesto,* 2025, https://nya.org.uk/wp-content/uploads/2025/02/UK-Youth-Parliament-Manifesto-2024-2026-FINAL-DIGITAL.pdf

[602] UNICEF, *Youth engagement,* https://www.unicef.org/innocenti/approach/youth-engagement.

[603] LSE, *Youth engagement,* https://www.digital-futures-for-children.net/about/youth-engagement.

[604] eSafety Commissioner, *eSafety Youth Council*, https://www.esafety.gov.au/young-people/esafety-youth-council.

[605] Mahima Jain, *How India's Public Health System Can Reach Rural Women Suffering Domestic Abuse,* 2023, https://pulitzercenter.org/stories/how-indias-public-health-system-can-reach-rural-women-suffering-domestic-abuse?utm_source=chatgpt.com.

6.5 Conduct a Multi-Stakeholder Consultation For Establishing Codes of Practices

To effectively foster online safety, a phased regulatory approach is recommended for which we require inputs from all stakeholders- the government, industry and civil society organisations. This approach recognizes the value of all stakeholders easing regulatory enforcement, drawing upon the experiences of Australia, the EU, and the UK. Initially, the focus should be on establishing voluntary codes of conduct. **These codes should be developed through a collaborative process involving online platforms, industry associations, civil society organizations, and eventually regulatory agencies.**

At the first stage, India's emphasis should be on encouraging platforms to adopt safety-by-design principles, implement best practices, and promote a culture of online safety. **As seen in Australia's Safety by Design initiative and the EU's Code of Conduct on Countering Illegal Hate Speech Online, these mechanisms can raise awareness, promote the adoption of safety features, and encourage a "race to the top" in safety innovation.** Regulators should actively guide this process, providing expertise, facilitating dialogue, and monitoring progress. This initial phase allows for a more agile and flexible approach, enabling platforms to innovate and adapt quickly to emerging online safety challenges. Furthermore, the voluntary model fosters collaboration and quick wins, such as faster takedowns and more transparency.

After some experience with voluntary codes, and inputs gathered from relevant stakeholders, and an evaluation of their effectiveness, the regulatory framework can transition to statutorily backed co-regulatory codes. This phased approach could offer several advantages. Firstly, it encourages early action, as voluntary codes can be implemented relatively quickly, prompting platforms to take swift measures to address online safety concerns. Secondly, it fosters collaboration, since the initial emphasis on industry participation promotes dialogue and collaboration among stakeholders, building trust and facilitating the development of effective solutions. Thirdly, it ensures long-term accountability, as the eventual transition to statutorily backed codes provides a clear framework for long-term accountability, ensuring that online safety standards are consistently upheld. **Finally, it provides adaptability, allowing the regulatory framework to evolve over time, adapting to changes in technology and the online landscape through regular multistakeholder consultation.**

Through stakeholder consultations we can envision a proper construction of a code of practice that can balance safety with competing imperatives of innovation and free speech. The development of these codes could address the need for innovative technical solutions to complex issues. For instance, instead of mandating specific, potentially flawed technical approaches like widespread automated content filtering tools that could lead to unintended consequences, or enforcing measures that might compromise fundamental security features like breaking end-to-end encryption, **the emphasis should be on outcomes.**

This approach acknowledges that platforms are better positioned to develop effective technical responses to emerging threats. This balance – enabling platforms to innovate while preventing issues like excessive content removal – can only be achieved through **test-and-learn environments that are centered on key focus areas and principles**. These environments allow for the experimentation and refinement of safety mechanisms. Initiatives like Project Lantern[606] and Project ROOST[607] exemplify the kind of collaborative, innovation-driven solutions that platform design codes should encourage, focusing on shared tools and best practices rather than prescriptive mandates.

[606] Tech Coalition, Announcing Lantern: *The First Child Safety Cross-Platform Signal Sharing Program*, 2023, https://www.technologycoalition.org/newsroom/announcing-lantern.

[607] Boden Moraski, *ROOST: A collaborative effort for AI-driven online safety*, 2025, https://medium.com/nexstudent-network/roost-a-collaborative-effort-for-ai-driven-online-safety-42001150d45b

6.5.1. Safety By Design

Safety by Design is an approach to technology development that puts user safety and rights at the core of product design, rather than treating safety as an afterthought[608]. Instead of **retrofitting** safeguards after harm occurs, it urges companies to **anticipate, detect, and eliminate online harms upfront** – building protections **"from the get-go"** into platforms and apps[609]. The goal is to embed safety into corporate governance, institutional design and product architecture, fostering more positive and secure online experiences for all users, especially **those most at risk** (such as women and children). This includes:

- **Service Provider Responsibility:** Platforms must continue to actively **design against abuse**. This means anticipating risks, engineering out misuse during the design phase, and **maintaining dedicated safety teams with robust content moderation, reporting mechanisms, and proactive detection/removal processes.** The goal is to prevent harm before it occurs, ensuring users (especially children) aren't left to self-protect. Ofcom's practical guidance for service providers endorses nine high-level actions for taking responsibility, preventing harm and supporting women and children in tackling online abuse.[610]
- **User Empowerment and Autonomy:** Users, and their guardians (where applicable), should have **control over their online experience**. This includes intuitive **safety settings, privacy controls, filters or labels for harmful content including trigger warnings and easy ways to block or report harmful content.** Australia's eSafety Commissioner recommends using technical features to nudge users towards safer interactions (For instance, Tinder uses AI to detect potentially offensive messages, sending an in-app prompt to the user with a gateway to report the incident), and building support features with constructive feedback loops for users reporting TFGBV.[611]
- **Transparency and Accountability:** Platforms should continue to be **open about their safety efforts and accountable for results**. This involves clear guidelines, regular reporting about performance of safety features, and regular engagement with experts and users for continuous improvement. **Publishing data, such as harm prevalence and response times, fosters public scrutiny and incentivizes a "race to the top" in safety innovation.**

Some key safety-focused measures are spotlighted below:

1. **Reforming Engagement-Based Ranking Systems:** Platforms should critically evaluate and modify content ranking and recommendation systems that are primarily driven by engagement metrics (likes, shares, comments), as these can inadvertently amplify divisive, sensational, or harmful content. Shifting towards trust-based or quality-oriented ranking systems which prioritizes metrics like source credibility, information accuracy, or positive interaction indicators (e.g., YouTube focuses on a combination of clicks, total watch time, survey responses, sharing, likes, and dislikes to infer what users find satisfying and promote higher-quality videos) — can mitigate these harms while maintaining user engagement and prioritizing quality and safety.[612]

---

[608] eSafety Commissioner, *Safey by Design,* https://www.esafety.gov.au/industry/safety-by-design.

[609] eSafety Commissioner, *Safey by Design,* https://www.esafety.gov.au/industry/safety-by-design.

[610] Ofcom, *A safer life online for women and girls*: *Practical guidance for tech companies*, 2025, https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/consultation-on-draft-guidance-a-safer-life-online-for-women-and-girls/main-docs/consultation-document-a-safer-life-online-for-women-and-girls.pdf?v=391803

[611] Australian Government eSafety Commissioner, *Technology, gendered violence and safety by design*, 2024, https://www.esafety.gov.au/sites/default/files/2024-09/SafetyByDesign-technology-facilitated-gender-based-violence-industry-guide.pdf?v=1726531200021

[612] Lena Slachmuijlder & Sofia Bonilla, *Prevention by Design*: *A Roadmap for Tackling TFGBV at the Source*, 2025, https://techandsocialcohesion.org/wp-content/uploads/2025/03/Prevention-by-Design-A-Roadmap-for-Tackling-TFGBV-at-the-Source.pdf?utm_source=substack&utm_medium=email.

2. **Implementing Default Privacy Settings to Minimize User Vulnerability:** Platforms should implement privacy-first default settings, particularly **for new users and identified vulnerable groups**. This includes measures such as limited profile visibility, restricted discoverability by search engines or non-friends, and control over who can comment or contact them. The success of initiatives like Facebook's "Locked Profiles" in India[613], which significantly reduced harassment for women users by restricting non-friend access to photos and posts, underscores the impact of such proactive defaults. Similarly, YouTube implements age-appropriate protections for children where content upload and commenting functionalities are disabled, and for teenagers (ages 13-17), the default settings for uploading and live-streaming are set to the most private available.[614]

3. **Nudging Users Towards Respectful Conduct:** Platforms should implement "nudges"—subtle, real-time prompts designed to encourage users to pause and reconsider potentially harmful language or imagery before posting. **Machine learning models can detect harmful language or NSFW media and trigger tailored nudges, offering an opportunity for self-correction.**

4. **Empowering Users Through Content Filters:** Providing users with robust filtering tools is essential for enabling them to manage their online experience. **Filters allowing users to block specific keywords, topics, or accounts can significantly reduce exposure to unwanted material (e.g., Instagram's "Hidden Words" feature).**

5. **Implementing Community Noting for Contextual Information:** Platforms should consider developing and scaling "community noting" features. These allow vetted contributors or a broader community to add context to content that might be misleading or disputed, empowering users to better assess information.

6. **Designing for Shared-Device Usage and Diverse Literacy Levels:** Platforms must design safety tools and information dissemination with an understanding of diverse user contexts in India, including shared-device usage and varying literacy levels.[615] This includes options like profile-level locks operable within shared accounts, features that can make app usage less conspicuous if needed, robust local language settings for all safety information, and simple audio-visual safety guides or gamified reporting processes to help new users with limited literacy stay safe online.

6.5.2. Establishing a Clear Framework to Address Contextual Online Risks for Women and Children

Effectively addressing online harms necessitates a comprehensive understanding of their nature and prevalence. This begins with inclusive definitions, tailored guidelines, detailed data collection, and an intersectional approach to policy.

1. **Expanding and Refining Categories for Reporting Online Safety Risks:** While online platforms have made strides in content moderation, they could consider adding more **categories for reportable harms**, particularly concerning abusive behaviors that do not neatly fit into categories like

[613] Lena Slachmuijlder & Sofia Bonilla, *Prevention by Design: A Roadmap for Tackling TFGBV at the Source*, 2025, https://techandsocialcohesion.org/wp-content/uploads/2025/03/Prevention-by-Design-A-Roadmap-for-Tackling-TFGBV-at-the-Source.pdf?utm_source=substack&utm_medium=email.

[614] Youtube, *Child Safety Policy* https://support.google.com/youtube/answer/2801999?hl=en

[615] Annual Status of Education Report (ASER), *ASER 2023 Evidence Brief: Digital Readiness of India's Youth*, 2024, https://asercentre.org/wp-content/uploads/2022/12/EB_Digital-Readiness-of-Indias-Youth_11.03.2024.pdf; The Voices, *Digital Literacy on the Rise, Online Safety Need of the Hour*, 2025, https://medialit.in/thevoices/digital-literacy-on-the-rise-online-safety-need-of-the-hour/.

illegal content. This limitation often results in a failure to adequately address nuanced, yet deeply damaging, forms of online abuse.

This issue is particularly pronounced when considering the diverse socio-cultural contexts in which these harms manifest. For example, the experience of image-based abuse in India often differs significantly from that in countries like the US, highlighting the need for platforms to adopt contextually sensitive reporting categories.

As articulated by the Rati Foundation[616], the impact of image-based abuse in India is often compounded by deeply entrenched cultural norms around modesty, shame, and victim-blaming. When intimate images are shared without consent, victims, especially women and girls, face severe stigmatization and ostracization within their communities, often leading to profound psychological and social havoc. This cultural intensification of harm can make it incredibly difficult for victims to come forward, report the abuse, or seek help, underscoring the inadequacy of a one-size-fits-all reporting framework designed primarily for Western contexts. The research indicates how these violations are frequently intertwined with broader issues of control, coercion, and family honor, intensifying the victim's vulnerability beyond what might be immediately apparent to an automated system or a moderator unfamiliar with the on-ground context.

Therefore, platforms must **expand their taxonomy of reportable harms** to include these and other abusive behaviors, ensuring their reporting mechanisms and content policies are sensitive to the varied impacts of such harms across different cultural and geographical contexts.

2. **Adopting an Intersectional Approach to Online Safety:** Platforms and policymakers must ensure that online safety policies explicitly protect marginalized communities, including trans and queer persons, and individuals facing caste-based discrimination. **This includes taking strong action against deadnaming, sexualized abuse, and caste-based hate speech or harassment, even when such content does not achieve high virality. To ensure these harms are recognized and adequately addressed, content moderation practices should strive for direct input from Dalit, Adivasi, Bahujan, queer, and trans communities.** For instance, Facebook added caste as a distinct category of online hate speech within its global community guidelines, recognizing it as a specific basis for exclusion and harm.[617]

6.5.3 Developing Survivor-Centred Reporting, Support, and Evidence Mechanisms

When harm does occur, reporting and support mechanisms must be designed with the survivor's experience at the forefront, ensuring ease of access, comprehensive support, transparent processes, objective moderation, and robust evidence collection to assist with law enforcement (where required/possible).

1. **Accessible and Supportive Reporting Channels:** Survivors must be able to report abuse through simple, easily navigable interfaces available in multiple languages. Continuous in-app support, such as comprehensive FAQs and chat-based assistance, should be readily available, alongside mobile-friendly and "Lite" versions of reporting portals.

2. **Streamlined and Transparent Complaint Resolution:** Platforms should work towards empowering users by enabling one-click reporting where feasible, allowing users to track complaint progress,

[616] Rati Foundation, *Meri Trustline: Annual Report Volume 2*, 2025, https://ratifoundation.org/wp-content/uploads/2025/05/Meri-Trustline-Year-2-Report-Final-Version_compressed.pdf.

[617] Meta, *Hateful Conduct*, 2025, https://transparency.meta.com/en-gb/policies/community-standards/hateful-conduct/

and providing clear, justified information regarding decisions or actions taken, all within a consolidated platform interface.

3. **Automatic Evidence Capture for Legal Recourse:** Under current law, when any information is removed or access to it is disabled, intermediaries are required to preserve such information and associated records for 180 days, or longer if directed by a court or a lawfully authorised government agency.[618] Building on this, and with due technical inputs from industry, content that is flagged by victims (and their support systems) that are relevant to a complaint involving women and children could also be preserved in encrypted formats, including metadata and timestamps, in line with statutory timelines. This could be later made available for use in legal proceedings or other institutional actions, subject to due process.[619]

6.5.4. Fostering Cross-Platform Accountability, Collaboration, and Transparency

The dynamic and interconnected nature of the internet means that harmful incidents, particularly those involving Technology-Facilitated Gender-Based Violence (TFBGV) and risks to children, rarely remain confined to a single platform. Perpetrators often exploit the siloed nature of content moderation, moving between sites to evade detection and continue their abusive behaviors. To effectively combat this, a robust framework for **cross-platform coordination** is imperative[620], especially as India continues its rapid digital expansion.

1. **Implementing Open-Source and Global Cross-Platform Coordination Efforts to Prevent Multi-Platform Abuse:**

We recommend that Indian policymakers encourage the industry and wider ecosystem to **implement global coordination efforts within India's digital landscape**. In this regard, there are opportunities to mainstream initiatives like the **ROOST Initiative**[621] and **Project Lantern**[622] within the Indian landscape. It will be important for domestically relevant stakeholders to actively participate in these initiatives and help shape the broader conversation, ensuring that global frameworks reflect the safety needs of women and children in India.

Mainstreaming the ROOST Initiative in India: The **Robust Open Online Safety Tools (ROOST) Initiative** offers an open-source tooling hub to democratize access to advanced trust and safety infrastructure for platforms of all sizes. By providing **ready-to-deploy software components** – such as rules engines, case management systems, and content safeguards powered by foundation models – ROOST alleviates the need for every platform to reinvent the wheel. For India, mainstreaming ROOST would mean:

- **Capacity Building:** Empowering smaller Indian platforms and startups to implement robust safety measures, particularly for **child safety**, by providing free, accessible, and high-quality tools that were previously out of reach due to cost or technical complexity.
- **Accelerated Innovation:** Fostering a collaborative environment where Indian tech companies can contribute to and benefit from shared safety advancements, allowing for faster detection and mitigation

[618] Rule 3(1)(g), IT Intermediary Guidelines, 2021.

[619] Lena Slachmuijlder & Sofia Bonilla, *Prevention by Design: A Roadmap for Tackling TFGBV at the Source*, 2025, https://techandsocialcohesion.org/wp-content/uploads/2025/03/Prevention-by-Design-A-Roadmap-for-Tackling-TFGBV-at-the-Source.pdf?utm_source=substack&utm_medium=email.

[620] Institute for Strategic Dialogue, *Misogynistic Pathways to Radicalisation: Recommended measures for platforms to assess and mitigate online gender-based violence*, 2023, https://www.isdglobal.org/wp-content/uploads/2023/09/Misogynistic-Pathways-to-Radicalisation-Recommended-Measures-for-Platforms-to-Assess-and-Mitigate-Online-Gender-Based-Violence.pdf.

[621] Boden Moraski, *ROOST: A collaborative effort for AI-driven online safety*, 2025, https://medium.com/nexstudent-network/roost-a-collaborative-effort-for-ai-driven-online-safety-42001150d45b

[622] Tech Coalition, Announcing Lantern: *The First Child Safety Cross-Platform Signal Sharing Program*, 2023, https://www.technologycoalition.org/newsroom/announcing-lantern.

of emerging harms relevant to the Indian context.

- **Consistency in Safety Standards:** Promoting a baseline of effective safety practices across the diverse Indian digital ecosystem, which is crucial given the rapid growth of new online services and applications.
- **Contextualization:** Facilitating the development of AI safety tools within ROOST that are specifically trained on Indian linguistic, cultural, and behavioral nuances, making them more effective at identifying TFBGV and children's risks unique to the region.

Mainstreaming Project Lantern in India: Project Lantern, a collaboration between the Tech Coalition and major tech companies like Meta Snap, Google, Twitch, and Discord reflects a critical step towards **cross-platform signal sharing for child safety**. It is also being repurposed as an initiative to provide women internet users with opportunities to fight against risks associated with cross-platform deepfake based NCII abuse.[623] Project Lantern enables participating companies to share information about accounts and behaviors that violate their child safety policies, allowing each platform to conduct investigations and take action on their own services. This tackles the problem of predators simply moving between platforms. Mainstreaming Project Lantern in India would involve:

- **Enhanced Intelligence Sharing:** Establishing formal mechanisms for Indian platforms, especially those with significant child user bases, to participate in this global signal-sharing program. This would allow for the proactive identification of individuals attempting to perpetrate child abuse across multiple services, significantly reducing their ability to operate undetected.
- **Proactive Risk Mitigation:** Enabling platforms to act on intelligence received from other services, even if the initial harmful act did not occur on their platform. For instance, if a "nudify" app (as highlighted in Meta's recent actions)[624] is identified on one platform, signals could be shared through Lantern to help other platforms proactively detect and remove similar content or associated accounts.
- **Strengthening Enforcement:** Providing local law enforcement and child protection agencies with better intelligence, facilitated by cross-platform sharing, to support investigations into online child abuse originating or impacting individuals in India.
- **Promoting Accountability:** Emphasizing that online safety is a shared responsibility, and mainstreaming Project Lantern would cement the expectation that platforms actively collaborate to protect children, rather than operating in isolation. At the same time, it could also serve as an important opportunity to strengthen protections for victims of AI deepfake-based apps, who often face severe harm and have limited recourse under existing frameworks.

2. **Unified Cross-Platform Reporting Mechanisms:** Efforts should be made to develop tools or collaborative frameworks that allow users to report abuse spanning multiple platforms through a single point of contact, potentially in partnership with CSOs.[625]

---

[623] Meta, *Taking Action against 'Nudify' Apps*, 2025, https://about.fb.com/news/2025/06/taking-action-against-nudify-apps/?utm_source=chatgpt.com.

[624] Tom Gerken, *Meta urged to go further in crackdown on 'nudify' apps*, 2025, https://www.bbc.com/news/articles/cgr58dlnne5o.

[625] Slackmuijlder & Bonilla, *Prevention by Design: a Roadmap for tackling TFGBV at the source*, 2025, https://techandsocialcohesion.org/wp-content/uploads/2025/03/TFGBV-Report-2025-Final-March-2025.pdf.

3. **Interoperable Evidence Documentation:** Platforms should explore adopting interoperable documentation tools, empowering users to collect and share evidence of abuse (screenshots, metadata) seamlessly across different platforms, reducing retraumatization.[626]

4. **Leveraging Civil Society Expertise:** Platforms should actively collaborate with or integrate existing reporting mechanisms developed by CSOs specializing in tracking TFGBV and other online harms (e.g., StopNCII.org).

5. **Facilitating Research Through API Access to Transparency Reports:** SSMIs should publish transparency reports with API access to enable research by academic institutions and CSOs.

6.5.5. Specialized Measures for High-Virality and High-Sensitivity Harms

Certain types of online harm, particularly those involving TFGBV, child sexual abuse material (CSAM), or content that can incite imminent violence, require specialized and rapid intervention capabilities.

1. **Investing in Advanced AI for Real-Time Detection:** Platforms must continuously invest in and deploy advanced AI-powered content moderation systems capable of detecting known and emerging harmful content in real-time, including re-uploaded or modified versions of previously removed material, especially for high-sensitivity harms.

2. **Developing Early Warning Systems:** Platforms should develop robust early warning systems that can identify content rapidly gaining virality that has characteristics associated with harmful campaigns or sensitive topics, allowing for proactive assessment and intervention.

3. **Creating Dedicated Rapid-Response Task Forces:** Platforms should establish and maintain dedicated rapid-response task forces. These teams should consist of cross-functional experts (legal, policy, specialized content moderators, communications). They must be specifically trained to handle high-sensitivity content and have clear protocols for swift, decisive action, including containment and escalation.

6.6. Enhancing Victim Rehabilitation in the Digital Age

To effectively support victims of online harm, a comprehensive, multi-faceted approach to rehabilitation is crucial. This approach should be anchored in the following key principles and practices, **drawing from successful models implemented internationally**:

6.6.1. Strong Legal Foundations with Victim Provisions

Legislation should explicitly recognize online abuses and mandate support for victims. As seen in Australia and the UK, laws that create obligations to assist victims (e.g., rapid takedowns) are essential. **The Philippines' OSAEC law, which requires trauma-informed referral protocols, provides a strong example of such a framework.** These legal foundations not only punish perpetrators but also formally acknowledge the victim's right to protection and care, thereby compelling government agencies to take action.

6.6.2. Dedicated Agencies or Helplines

The establishment of specialized bodies or helplines, such as Australia's eSafety Commissioner and the UK's Revenge Porn Helpline, significantly improves outcomes for victims. These entities streamline processes, coordinate with platforms, and provide crucial guidance to survivors. In the absence of standalone agencies, well-funded and government-partnered

[626] Slackmuijlder & Bonilla, *Prevention by Design: a Roadmap for tackling TFGBV at the source*, 2025, https://techandsocialcohesion.org/wp-content/uploads/2025/03/TFGBV-Report-2025-Final-March-2025.pdf.

NGOs can also effectively enhance victims' access to relief.

India's national cybercrime helpline system currently suffers from shortages in trained personnel, particularly those equipped in responding to gender-based online harms and cybercrimes against children. This is compounded by an overall lack of staffing and reduced budget allocations. We have observed instances of childline funding cuts[627], and women's helplines operating largely as call-forwarding services that transfer calls to One Stop Centres or police, often with little (to no) follow-up[628]. India's national helplines require urgent reform. They require dedicated investments in capacity-building and staffing, with a focus on recruiting and training personnel equipped to respond to gender-based violence and cybercrimes against children. **Specialized training should include sensitivity to trauma, knowledge of digital safety tools, and familiarity with relevant legal provisions.** To ensure adequate coverage and timely response, **budgetary allocations must be enhanced**. In addition, helpline protocols should be updated to enable seamless coordination with law enforcement, mental health services, and CSOs that provide survivor support, ensuring that users are not left navigating fragmented systems on their own. **Further, there is a pressing need to develop separate SOPs and maintain dedicated teams for women and children.**

6.6.3. Survivor-Centered, Trauma-Informed Approach

Rehabilitation efforts must prioritize the victim's choices, privacy, and psychological needs. Adopting trauma-informed care, as seen in the work of organizations like *Chayn* and the protocols in the Philippines can be beneficial. This approach involves providing counseling, preventing re-victimization, and empowering the victim to regain control. **Training law enforcement, social workers, and hotline staff in trauma-informed practices is essential for fostering trust and encouraging victims to seek and adhere to support services.**

6.6.4. Measured Outcomes

The effectiveness of victim rehabilitation practices should be evaluated through measurable outcomes, including:

- **Content Takedown and Reduced Exposure:** Swift removal of abusive content, as demonstrated by the high success rates in Australia and the UK, is crucial in minimizing ongoing harm and anxiety for victims.
- **Empowerment and Well-being:** Efforts should focus on empowering victims and enhancing their sense of safety and control. Feedback from programs like *Australia's eSafety* and *Brazil's Maria d'Ajuda*, which indicate increased digital confidence among survivors, should be considered.
- **Access to Justice:** Increased reporting of online abuse and successful legal actions, as seen in the UK, South Africa, and the Philippines, demonstrate improved access to justice for victims.
- **Safety and Recidivism:** The reduction in repeat victimization, as evidenced by the success of protection orders and rehabilitation programs in South Africa and the Philippines, indicates the effectiveness of these interventions.

[627] The Times of India, *1098: where silence speaks volumes*, 2024, https://timesofindia.indiatimes.com/city/chennai/1098-child-helpline-revolutionizing-support-for-silent-calls-in-india/articleshow/115575758.cms.

[628] The Reporters' Collective, *Tuned out: Launched after Nirbhaya, helpline fails women*, 2024, https://www.reporters-collective.in/womens-safety/tuned-out-launched-after-nirbhaya-helpline-fails-women.

About The Quantum Hub

Founded in 2017, **The Quantum Hub (TQH)** is a multi-sectoral public policy research and consulting firm working with various stakeholders including industry, academia, civil society, philanthropies, regulators and governments. Over the years, it has consistently engaged on issues related to the digital economy and governance, and has closely tracked debates on data protection and online safety.

This report builds on TQH's previous work on children and women's participation on the internet. Our report, ***Navigating Children's Privacy and Parental Consent Under the DPDP Act 2023***, examined pathways for implementing the 'verifiable parental consent' requirement under the Digital Personal Data Protection Act, 2023. TQH's work on children's data protection and online safety, is also complemented by the work of its sister organisation the **Young Leaders for Active Citizenship (YLAC)** which was founded in 2016. **YLAC's Digital Champions programme** equips students to learn about various facets of online safety, become conscious consumers of information, and foster a healthier relationship with technology. YLAC also undertakes large scale surveys to explore how young people access the internet, what they use it for, the risks they encounter, and the support they require.

Within its work on gender inclusion, TQH engages on issues relating to women's labour force participation, women in leadership and women's overall health and wellbeing. In this context, TQH has spearheaded policy dialogues on women's online safety, unique challenges of women in public life, gender responsive policymaking in tech, women's privacy and agency online, and recent legislative discussions on regulating obscene content.

For more information, visit: **thequantumhub.com**

The Quantum Hub Consulting

New Delhi