॥ न्यायस्तत्र प्रमाणं स्यात् ॥
NLU DELHI

CENTRE FOR COMMUNICATION GOVERNANCE

“BUT THAT'S NOT ME!”

Sexualised Deepfakes in Law, Tech and Society



Published by: Centre for Communication Governance at National Law University Delhi, Sector 14, Dwarka, New Delhi – 110078

Patrons: Professor (Dr.) G.S. Bajpai, Vice Chancellor, NLUD, and Professor (Dr.) Risham Garg, Registrar

NLUD Faculty Director, CCG: Dr. Daniel Mathew

Executive Director, CCG: Jhalak M. Kakkar

Authors: Sukriti, Rishiti Choudaha, Divyansha Sehgal
(author names are listed in reverse alphabetical order)

Supervision: Sukriti

Review: Jhalak M. Kakkar, Sukriti

Conceptualisation: Torsha Sarkar, Rishiti Choudaha, Sukriti

Illustrations by: Pritish Bali and Sameera Masoodi

Report Design and Visualisation (Infographics): Anamika Dubey

Design supervision and conceptualisation: Shivani Mago

Suggested Citation: Sukriti, Rishiti Choudaha, Divyansha Sehgal '"But that's not me!": Sexualised Deepfakes in Law, Tech, & Society' (Centre for Communication Governance, National Law University Delhi 2026)

(CC BY-NC-SA 4.0)

“BUT THAT'S NOT ME!”

Sexualised Deepfakes in Law, Tech and Society



Acknowledgements

We are continuously grateful for the generous support we received from the National Law University Delhi (NLUD). The Centre for Communication Governance (CCG) would therefore like to thank our patrons, the Vice Chancellor, Professor (Dr.) G.S. Bajpai, and the Registrar, Professor (Dr.) Risham Garg of NLUD, for their constant guidance. CCG would also like to thank our Faculty Director, Dr. Daniel Mathew, for his continuous direction and mentorship. We would also like to thank Jhalak M. Kakkar, Executive Director at CCG, for her unwavering support, encouragement, and mentorship, which have been instrumental in publishing this research. We would also like to thank all members of CCG for the many ways in which they supported the report. Special thanks to the ever-present and ever-patient CCG Administration Team for their unwavering support in all the work we do at CCG.

*We acknowledge the support and guidance provided to us by Ford Foundation.

Foreword

The rapid advancement of generative AI technology has transformed the manner in which audiovisual content is created, altered, and disseminated. The proliferation of accessible AI technology has exponentially increased the volume, speed, and reach of harms experienced by women globally. The creation of non-consensual sexualised imagery (NCSI) has been reduced to an act requiring just a few seconds. The sheer scale of the harms caused by NCSI have exploded over the years, and technology has been weaponised in ways that affect women from all walks of life, from celebrities and public figures to women in their everyday lives.

While debates have ignited worldwide around how best to mould the law to address these harms, the specific context of the Global South demands dedicated scholarship. Here, the harms of sexualised deepfakes are exacerbated owing to intersectional factors such as gender, caste, religion, and community, and cannot be viewed through a one-dimensional individualist approach. This marginalization must be studied specifically.

This report, *"But that's not me!": Sexualised Deepfakes in Law, Tech and Society*, from the Centre for Communication Governance at National Law University Delhi, presents one such Global South view, where the lived experiences of victims of technology-facilitated gender-based violence have been placed front and centre. Laudably, this project employs digital illustrations as a tool of articulating research in a visually engaging and impactful manner. Art and design can often express depth with sensitivity and bring out the human dimension of the issue which is so deeply grounded in the lived experiences of those affected. Part I of the report documents instances of sexualised deepfake abuse in India and elsewhere, and traces how intersecting identities and marginalities can shape the injury itself along with the avenues for redress. Part II expands the conceptualisation of harms caused by AI-generated NCSI from a single, unitary 'harm' to multivariate harms impacting women's dignity, economic and professional lives, public participation, bodily autonomy, privacy, and consent. Part III studies how social stigma and traditional norms of honour and shame in India complicate gender justice, and drive victims into silence. It then scrutinises the challenges to meaningful recourse, unravelling the structural and technological conditions that enable harm in the first instance,

including low-cost deepfake generation tools, recommendation algorithms, payment infrastructures, and inaccessible grievance redressal mechanisms which combine to place the burden for justice almost entirely upon victims. In Part IV, the report discusses approaches to guide the way forward. It assesses the promises and limitations of a range of socio-legal, ideological and socio-technical measures and presents a critically sound feminist, rights-based approach going forward.

As Vice Chancellor of the National Law University Delhi, I commend the Centre for Communication Governance for undertaking research on a subject as deep and pressing as sexualised deepfakes while placing the principle of consent and intersectionality at the heart of the project. As policymakers in India and the rest of the world brainstorm how best to respond to AI-generated NCSI, there is an opportunity to shape law to move beyond well-worn ideas of detection and takedown towards a victim-centred justice system that emphasises accountability and human dignity. I am certain that this report will serve as a useful reference for a wide range of stakeholders, and it is yet another feather in the cap for the Centre, which has produced excellent rights-forward policy-relevant research and advocacy over the years.



Prof. (Dr.) G.S. Bajpai
Vice Chancellor
National Law University Delhi

About Us

The Centre for Communication Governance at the **National Law University Delhi (CCG-NLUD)** was established in 2013 to ensure that Indian legal education establishments engage more meaningfully with information technology law and policy and contribute to improved governance and policy making. CCG is a leading academic research centre dedicated to undertaking rigorous research in India on technology law and policy. Through its academic and policy research, CCG engages meaningfully with policy making domestically and globally by participating in public consultations, contributing to parliamentary committees and other consultation groups, and holding seminars, courses and workshops for capacity building of different stakeholders in the technology law and policy domain.

The National Law University Delhi is one of the leading law universities in the capital city of India. Established in 2008 by an Act of the Delhi legislature (Act. No. 1 of 2009), the University is ranked second in the National Institutional Ranking Framework for the last five years. Dynamic in vision and robust in commitment, the University has shown terrific promise to become a world-class institution in a very short span of time. It follows a mandate to transform and redefine the process of legal education. The primary mission of the University is to create lawyers who will be professionally competent, technically sound and socially relevant, and will not only enter the Bar and the Bench but also be equipped to address the imperatives of the new millennium and uphold the constitutional values. The University aims to evolve and impart comprehensive and interdisciplinary legal education which will promote legal and ethical values, while fostering the rule of law.

Table of Contents

Glossary 01

Executive Summary 02

Introduction 09

Section 1: NCSI abuse through the lens of Intersectionality and Marginality 15

I. Religious Identity 16
II. Casteism 16
III. Queerness 17
IV. Personal Disagreements 17
V. Educational Institutions 18
VI. Scams and Blackmail 18

Section 2: Understanding the Harms 21

I. Harm to Dignity 22
II. Economic and Professional Harms 24
III. Harm to Public Participation and Free Expression 26
IV. Actual Bodily Harms 29
V. Privacy Harms 30
VI. Consent 35

Section 3: Challenges to Meaningful Recourse 39

I. Virtue Signalling, Socio-Cultural Norms and NCSI - "लोग क्या कहेंगे?" 39
II. Gender Justice and Meaningful Redressal 43
III. Structural and Design Conditions that Enable Harm 46

Section 4: Pathways to Solutions 53

I. Socio-Legal and Ideological 53
II. Socio-Technical 57

Conclusion 61

Endnotes 64

Glossary

NCSI: Non-Consensual Synthetic Imagery (also referred to as ‘sexualised deepfakes’)

NCII: Non-Consensual Intimate Imagery

TFGBV: Technology-Facilitated Gender-Based Fiolence

TFSV: Technology-Facilitated Sexual Violence

IBSA: Image-Based Sexual Abuse

CSAM: Child Sexual Abuse Material (previously was also referred to as ‘child pornography’)

LGBTQIA+: Lesbian, gay, bisexual, transgender, queer, intersex, asexual plus
Deepfakes: Synthetic audio, video and image media created using AI tools

LLMs: Large Language Models

Deepfakes: Synthetic audio, video and image media created using AI tools.

Disclaimer (Trigger Warning)
This report contains discussions of non-consensual sexual imagery, sexual abuse, caste and gender-based violence, and their psychological and bodily impact, including references to suicide. The content may be distressing to victim-survivors and readers. Reader discretion is advised.

Executive Summary

The problem of deepfakes has long persisted, but has accelerated in scale and sophistication with the expansion of Generative AI, consistently outpacing both public awareness and institutional responses. While sexualised deepfakes are often studied within a wider umbrella of Tech-facilitated Gender based Violence (TfGBV) for the harms and implications involved, it poses nuanced and specific challenges that relate to the social, cultural, technical, institutional, and political realities of a region. AI-generated sexual deepfakes, or non-consensual synthetic imagery (NCSI), as this report terms them, have come to be used against any woman, agnostic to her private or public status, and weaponised against women from all sections of society.

The report is written with the objective of contributing to the emerging body of research that distinctly centres Global South contexts in understanding NCSI abuse and its implications. Through this, it hopes to look beyond a predominantly Western lens to the issue. It offers insights into the uniqueness of the NCSI abuse as it occurs in the Global South, and the legal, policy and technical limitations that result from this, by demonstrating the role of socio-cultural factors in the weaponisation of deepfakes against women.

Impact of NCSI for Marginalised Groups, and the Need for an Intersectional Lens

Within the social and cultural realities of India and South Asia, women themselves exist as a marginalised and vulnerable group. However, within these contexts, the degrees of marginalisation are further influenced by an intersection of factors that serve to shape women's vulnerability towards and experience of abuse. In India, the factors determining this marginalisation exist on a spectrum of personal and religious identity (caste, class, religion, sexuality), or professional and economic realities (politics, journalism, content creators, actors). These are not causal factors, but undeniably influence who is impacted and in what ways, and the available recourse. Thus, it is pertinent to view NCSI from an intersectional perspective and observe how different types of discrimination faced by individuals compound to create differential impacts and unique experiences of abuse in the digital sphere.

In the Indian context, gendered digital violence occurs within deeply unequal social structures, where women's sexuality, mobility, public visibility and freedom are policed by invoking social norms of family honour, respectability, and communal reputation. These structures enable weaponisation of deepfakes against women through violation of their dignity and privacy, while threatening their social status, familial relationships, safety, livelihoods and capacity to participate in public life. In this way, the deliberate, and coordinated weaponisation of sexualised deepfakes against women attempts to control and keep women subjugated within the prevailing misogynistic societal structures. For instance, for women belonging to marginalised and oppressed communities in India, such as, Muslim women in India, deepfakes can be layered with Islamophobic tropes and communal harassment; for Dalit and Adivasi women, it can reproduce historical patterns of caste-based sexual violence, humiliation and narratives of 'pollution' that reaffirm caste hierarchy; for queer and gender-nonconforming people, sexualised deepfakes can compound stigma in a socio-political environment that remains hostile to their identities; and for rural women and those in tier-two cities, it contributes to intensified community surveillance and communal pressure to remain silent against the abuse.

While the report studies the links between marginalisation and NCSI abuse, it does not make a categorical distinction between marginalised and non-marginalised women in conducting its analysis. Given the pervasive and agnostic nature of NCSI abuse, a distinction between marginalised and non-marginalised women is not only impossible but also an incorrect point of interrogation into the question of marginalisation in this issue. A woman impacted need not herself be on a digital platform - it is enough that the perpetrator had the means to carry out the abuse against her. Similarly, an otherwise privileged woman may not be more secure or less vulnerable to this form of abuse given the pervasive nature and persistence of the abuse in time and space, and the harm resulting for her person, as discussed in this report. However, factors contributing to marginalisation impact the experience, available response, and her survival through this form of abuse.

Understanding the Harms from Deepfakes in an Indian Context

The report maps the layers of harm that surface from the case studies of NCSI and how they intersect within the socio-cultural and institutional realities of India and the Global South. It analyses how caste, religion and other community norms bring about these harms. The objective of such analysis is to map the patterns in injury, the underlying factors aiding the abuse, understand areas for response and redressal, and bring forth the regional and cultural complexities shaping the abuse, its implications, and what meaningful response would mean within these regional contexts. Based on this, the report provides the following matrix of harms:

a. Harm to Dignity: The harm to dignity from NCSI results from the non-consensual grafting of one's likeness for depiction in situations they werenever in. This severing of the link between private self and public representation, persists irrespective of whether the content is sexual in nature. In societies where women are communally policed through social norms of modesty, honour and respectability, a loss of control over the representation of one's own personal identity is experienced as an assault on one's personhood.

b. Economic and Professional Harms: These entail loss of jobs or difficulty in new recruitment due to the stigma and reputational harms resulting from NCSI. The vulnerability towards and severity of this harm is heavily influenced by pre-existing marginalisation of a woman. It can make it harder to recover from the economic and professional shock resulting from NCSI abuse, and entail greater struggle to find institutional support or access to legal assistance to remedy one's circumstances. Here, a single instance of NCSI abuse can result in a durable economic injury, reducing income, limiting advancement, and shrinking a survivor's long-term career choices.

c. Harm to Public Participation: Due to the shame and trauma associated with NCSI abuse, survivors withdraw from public and digital spaces, and self-censor. This harm is not just after-the-fact but also arises from a mere threat of manipulation of one's online images with AI, resulting in a chilling effect for speech, expression and free participation in both physical and online spaces. The report identifies public participation to not only be limited to a public presence, such as is the case with politicians and celebrities, but also as a woman's right to occupy public spaces that connects to her pursuit of friendships/personal social relationships, education, a career, free expression, and a sense of safety in public and online spaces. This right is severely constricted with NCSI abuse.

d. Actual Bodily Harms: Even though there is no physical infliction of the abuse in instances of TfGBV, this form of abuse has still been found to have tangible implications for both the mind and the body. These include credible threats of physical violence, stalking and doxxing that transport the harm from the screen into the street; and bodily/ somatic manifestation like stomach ulcers caused by psychological trauma from the incident. The report emphasises on the need for dismantling the presupposed dualism between the body and the mind within prevailing law and policy towards a view that considers these harms as 'embodied, real and tangible'.

e. Harm to Privacy: NCSI violates the privacy of a woman by using her likeness to construct and circulate a scenario she did not consent to be a part of. NCSI creates a distinctive privacy injury because it manufactures an intimate act that could still be socially attributed to the identifiable person, regardless of the 'artificial' nature of the content.

f. Consent-based Harms: Consent is not centered within both data protection and criminal laws for providing meaningful legal recourse for NCSI abuse. The existing understanding of consent within the law is situated in the assumption that it originates from a tangible source that involves a subject, a creator and the moment of capture of the media in question. This assumption is ruptured by AI-generated deepfakes because any image media can be modified. This renders traditional consent frameworks redundant for synthetically generated media, such as NCSI.

Challenges to Meaningful Recourse

This report illustrates how the socio-cultural pressures in turn undermine the effectiveness of legal and platform-based technical interventions in the following ways:

a. Virtue Signalling and Socio-Cultural Norms: These point to the jurisprudential gaps in legal and rights-based frameworks that fail to reconcile the provision of these rights to socio-political factors in place. These factors, such as, caste, gender, class, and religion shape an individual's ability to safely claim these rights in practice. Similarly, retrofitting the injuries resulting from NCSI into pre-existing legal categories results in the erasure of the uniqueness of the harm arising from NCSI abuse. This means that the formal availability of legal mechanisms, such as a criminal provision, platform reporting mandates, or a content takedown mechanism, do not themselves enable a meaningful recourse.

b. Structural and Design Conditions: The ability of a survivor to call out abuse, safely report it, secure prompt removal, obtain accountability and recover from the consequences is shaped by an interaction between institutional capacity, platform design, commercial incentives and entrenched socio-cultural hierarchies.

c. Gender Justice and Meaningful Redressal: The above raises the question: what does justice mean to a victim-survivor? Victim-survivors may seek consequences for the perpetrator and find themselves let down to know that there were none. This itself can further add to the trauma of what they experienced on finding their deepfakes on the internet. Similarly, traditional court-centered notion of justice might fail the victim-survivor for the delay in its delivery, and result for them in a reliving of the trauma.

Limits of Existing Remedies, and Future Pathways

This report chooses to identify pathways and does not offer policy or any specific form of recommendations as it believes that the research and the analysis required for identifying such recommendations requires further stakeholder coordination, and field insights that were not within the scope of this report. Below, we identify some future pathways, some of the limitations of existing pathways and potential approaches to overcome these limitations:

I. Socio-legal and Ideological Pathways:

i. From Consent to Structural Accountability: Legal responses must centre the evolving (and often inapplicable) nature of consent in cases of NCSI abuse. Potential responses include statutory recognition of ongoing, revocable agency over one's likeness, strict liability for platforms that design for virality without safeguards; and doctrinal shifts that treat the non-consensual creation of sexual imagery as a violation of bodily autonomy regardless of falsity or reputation impact.

ii. Intersectionality and the Doctrinal Method: There is a need to embed the feminist intersectional conceptualisations in doctrinal design: recognising NCSI as structural violence, rather than an individual offence; mandating investigation protocols sensitive to socio-political discriminatory factors, such as caste and religion, and creating independent oversight mechanisms to counter institutional bias.

iii. Cultural Transformation and Ideological Fracture: Honour-based, caste-based and other patriarchal and misogynistic norms that feed these ecosystems of abuse should be overcome by a shift in public discourse.' This should be supported by digital literacy initiatives through educational curricula and other public awareness efforts to critical curricula on consent, dignity and gendered politics of online presence that challenges honour-shame narratives that blame survivors, and frames viewing, sharing and commenting as participatory violence.

iv. Solidarity, Care and Collective Resistance: This entails accessible, trauma-informed support: legal aid, mental health services, safe housing and economic assistance to mitigate professional fallout. Grassroots organisations, feminist collectives and digital rights groups should be resourced to provide rapid response, documentation and advocacy.

II. Socio-technical pathways:

i. Hash-matching: Hash-matching is currently the gold standard for taking down known offensive content effectively. There are drawbacks, including, most importantly the burden falls almost entirely on the victim currently. It can be combined with other methods, like improved reporting mechanisms to be more effective in addressing the harms of NCSI and reducing the burden on the victim.

ii. Watermarking: Watermarking efforts can help users identify the synthetic nature of the content and discredit it as necessary. However, content designed to demean, embarrass, or discourage public participation would succeed in its intended manipulation regardless of how the image was manipulated.

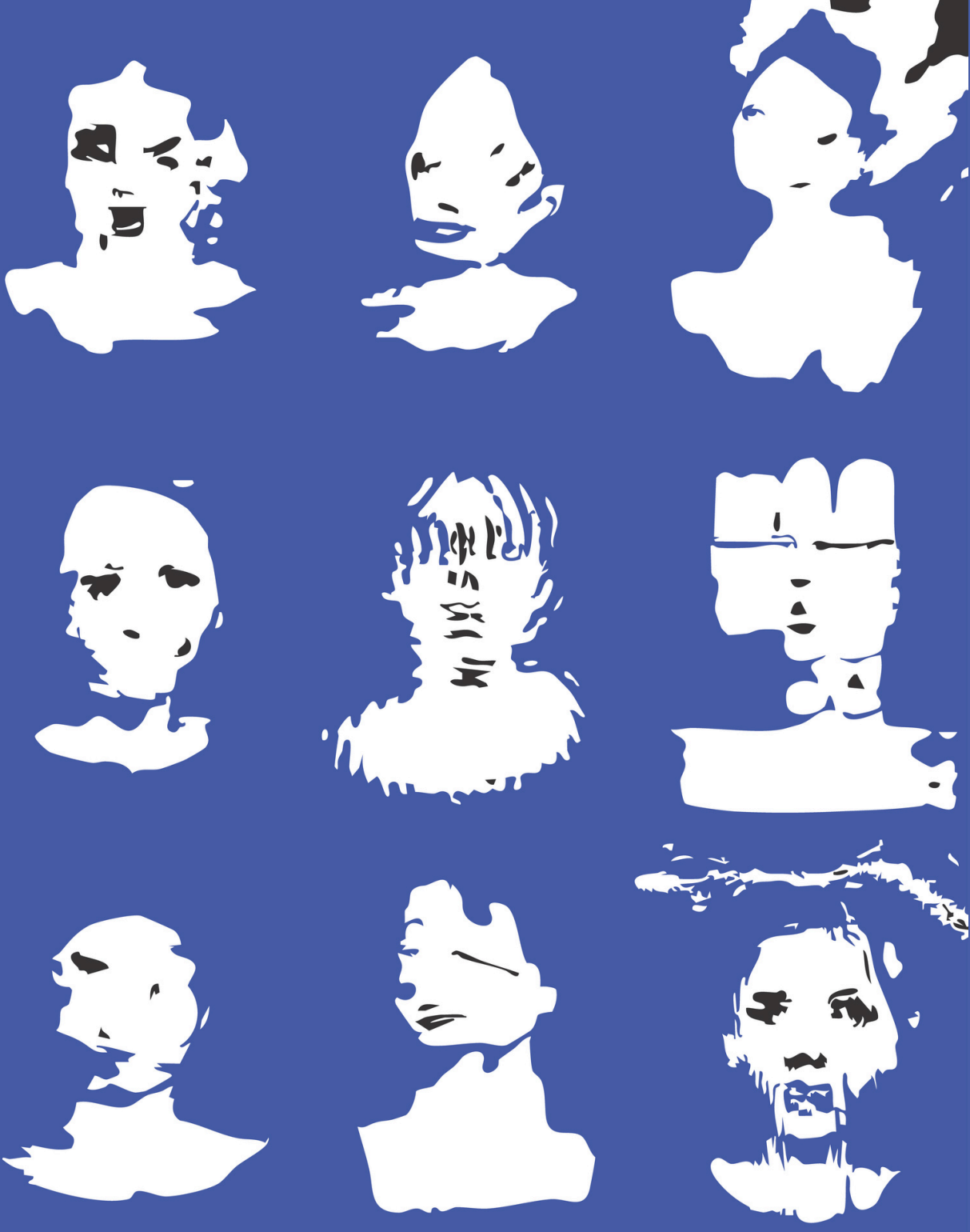
iii. App-Stores & Payment Infrastructures: Technical infrastructure like appstores, payment processors and search indexes should serve as gatekeepers against tools that enable creation and dissemination of NCSI. They can disincentivise or prohibit monetisation of abusive content,

and force platforms into better content moderation. There are, however, concerns of over-implementation of bans which results in over-removal of acceptable speech as automated filters may catch unrelated apps, payment or links on their respective infrastructure.

iv. Predictive Classifiers: Predictive classifiers developed for content moderation can provide a novel way to create automated moderation systems that are more aligned to company policies, and can thus help enforce them more effectively. It allows for identification of somewhat modified content as well, but may inadvertently catch material that is not NCSI infringing on acceptable speech on the platform.

v. LLMs for Content Moderation: LLMs have shown better performance on hate speech identification than predictive classifiers on some benchmarks, and newer LLM based approaches allow trust and safety teams to bring a policy written in plain English to the model. However, these approaches, while promising, are currently more suited to text based media than to image based media.

vi. Transparency and Reporting Mechanisms: Improved transparency mechanisms, such as allowing auditors and third party researchers to access data related to gendered harms of training images will be helpful in breaking the black box of moderation, so that researchers can study impacts of various tools and their efficacy. Improved reporting mechanisms that are intuitive and allow users to quickly point to offending material, and that can be actioned on by a team sensitized to deal with such material, can provide immediate remedy for victims of NCSI abuse.



Introduction

In an AI-driven online ecosystem, deepfakes are easy digital weapons exploiting an intersection of social vulnerabilities.[1] In countries such as India, deepfakes have to come be weaponised to marginalise and target individuals and groups along the lines of gender, sexuality, race, religion, caste and community. Non-consensual sexualised deepfakes, termed in this Report as non-consensual synthetic imagery ('NCSI') primarily target women, and are proliferating widely with easy creation, dissemination and consumption through Generative AI, social media, and networked online communication.[2] Studies have estimated that between 2023 and 2025, the number of deepfakes shared across social media platforms has increased from 500,000 to 8 million.[3] Although these figures constitute all types of deepfakes, it is estimated that most of these would be in the non-consensual category, primarily targeting women and girls.[4] A recent research studying the role of technology infrastructure in enabling proliferation of NCSI identified 88 publicly accessible websites over a duration of six weeks that hosted NCSI content, out of which 38 were entirely dedicated to such content.[5] Apps and platforms developed for the sole purpose of creation and consumption of NCSI utilise these ecosystems of abuse to generate profit from abusive content.[6]

In late 2025, Grok AI, the generative AI chatbot for the platform X (formerly Twitter) is estimated to have created millions of sexualised images of women and children.[7] These included not only famous women like popstars Taylor Swift and Sabrina Carpenter, politicians like Sweden's deputy Prime Minister Ebba Busch[8] but also private individuals from across the world. In 2018, years before ChatGPT was a household name, Hollywood actress Scarlett Johansson raised the issue of NCSI in popular media.[9] In 2024, millions of people accessed deepfaked sexualised imagery for musician Taylor Swift before platforms and regulators stepped in.[10] Other examples from the last year include Northern Irish politician Cara Hunter,[11] actress Collen Fernandes in Germany,[12] and Ukrainian MP Svitlana Zalishchuk.[13] A recent investigative analysis of websites shows that female politicians from 22 countries were targets of sexually explicit deepfakes.[14] In India, actress Shalini Pandey is the latest victim of the technology,[15] but actresses Priyanka Chopra Jonas, Alia Bhatt and Rashmika Mandanna among others are known to be prior targets.[16]

While sexualised deepfakes of famous women find most public visibility and discourse, the harms are not limited to them.[17] As this report attempts to demonstrate, deepfakes are weaponised against women from all

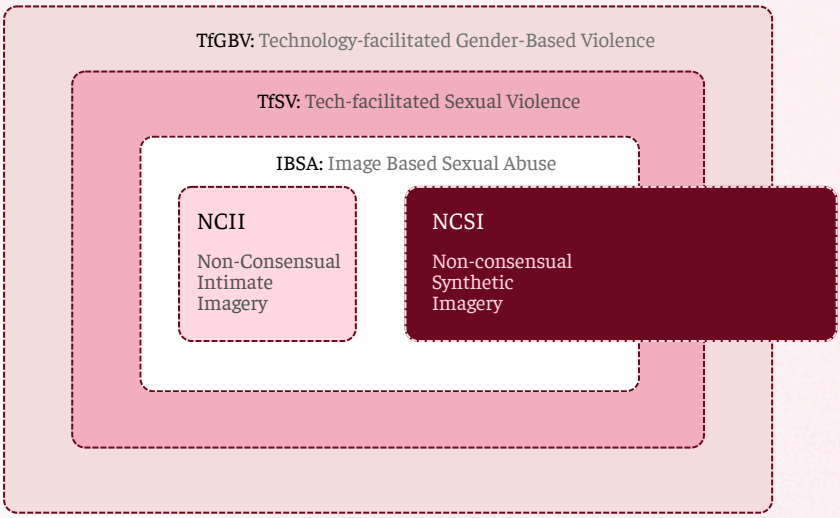
sections of society for merely existing outside of the traditional and patriarchal gender roles assigned to them. Women from marginalised and vulnerable communities will often find themselves at the center of abuse they are ill-equipped to cope with. Thus, it is pertinent to view NCSI from an intersectional perspective and observe how different types of discrimination faced by individuals compound to create unique experiences of harm in the digital sphere.

While deepfakes are often studied within a wider umbrella of TfGBV for the harms and implications involved, it poses nuanced and specific challenges that relate to the social, cultural, technical, institutional, and political realities of a region. In particular, the Global South, facing significant disadvantages in technological infrastructure, data sovereignty, judicial capacity, alongside a fragmented governance landscape, often possesses limited institutional understanding and tools to respond appropriately and swiftly to the problem[18]. As our analysis will note, this creates a dependence on platform-specific procedures (such as, a report to block or takedown content), coordination (such as, with the platform's grievance redressal cell, lawyers, or non-profits supporting victim-survivors of NCSI abuse), and multiple follow-ups for content removal. Resultantly, for victim-survivors of NCSI abuse, seeking accountability and redressal is not possible without access to substantial support mechanisms like mental health, financial support, and legal guidance.

Within law and policy, solutions for tackling NCSI often emphasise detection and takedown of such content from the platform. While effective takedown mechanisms are key, scholars point to the need for moving beyond credibility and detection techniques for deepfakes toward understanding and dismantling the intersectional power dynamics underlying gender based violence, as discussed in this report in Chapter IV[19]. Research has noted that there is still little known about the harms from deepfakes as a distinct form of gender based violence beyond explorations of some aspects of victim experience, like feelings of helplessness and loss of control[20].

In research, NCSI are often subsumed within the broader categories of tech-facilitated gender based violence (TfGBV) or Tech-facilitated Sexual Violence (TFSV/TSV) and its subset of Image Based Sexual Abuse (IBSA). This report will employ any of these terminologies as relevant. Although NCSI is closely linked to the wider and pre-existing harms within these categories of online gender based violence;[21] we posit that, due to the synthetic and manipulative nature of NCSI created with ease using LLMs, this approach takes away its distinctiveness as an issue,[22] undermines the resulting complexity of harms, and fails to highlight the range of systemic and structural challenges underlying the problem. The infrastructure of an LLM enabling the generation of sexualised deepfakes raises unique challenges for questions around privacy and consent, survivor-centric redressal, ethical and legal implications, and the social and cultural responses, among others.[23]

Figure 1: **Relationship between TfGBV, TFSV, IBSA, and NCSI**



This report advances some of the insights from our preceding cross-jurisdictional regulatory analysis on NCSI through a deep-dive into the intersection of socio-cultural factors that influence weaponisation of deepfakes against women, characterises the harms experienced by victim-survivors, and defines the responses required for tackling the problem.[24] It moves beyond existing data and regulatory principles to offer a holistic conceptualisation of the harm and its implications for those most affected. Drawing on case studies, critical, evidence-based theoretical frameworks, including feminist legal theory, intersectionality and socio-technical analysis, the Report asks not only what the law could do to protect victims and hold perpetrators accountable, but what evolutions must occur in platform design, corporate accountability, institutional practice and cultural norms.

The objective of our report is three fold. First, to document the real harms caused by sexualised deepfake content, using news reports, academic studies and testimonial evidence from across the world, with a special focus on India and specific Global South contexts. Second, to demonstrate how socio-cultural factors such as caste, religion, and honour-based harms compound these harms for women from marginalised communities, shaping both the intensity of the injury, and the access to justice. Third, to map the full range of challenges that obstruct effective responses, from legal gaps and platform failures to structural and socio-cultural barriers.

The scope of this Report is intentionally broad. By centering the victim-survivor experience, it examines harms to dignity, bodily autonomy and self-representation, and the exclusionary harms that silence women. This highlights the uneven distribution of harm, showing how women from marginalised and vulnerable communities bear a disproportionate burden.

It centres consent as a normative and legal foundation, while interrogating how consent collapses under existing platform designs, legal inertia and infrastructural conditions shaped by algorithms, corporate policies and regulatory gaps.

While tech-facilitated gender based violence has been pervasive across the different regions of the world and has resulted in a diverse range of harms and survivor experience in different cultural contexts, prevailing platform governance and policy discourses have been dominated by Western perspectives. Our report hopes to advance ongoing research by scholars from different communities to decentre Western perspectives and bring forth an analysis of harms based on women's personal experience and accounts, and identify gaps in current policy measures to tackle the problem of NCSI. Given the global nature of digital platforms, the cross-jurisdictional spread of information and NCSI's consequent harms, an intersectional and culturally diverse understanding of harms is important to find survivor centric, robust measures to tackle NCSI.

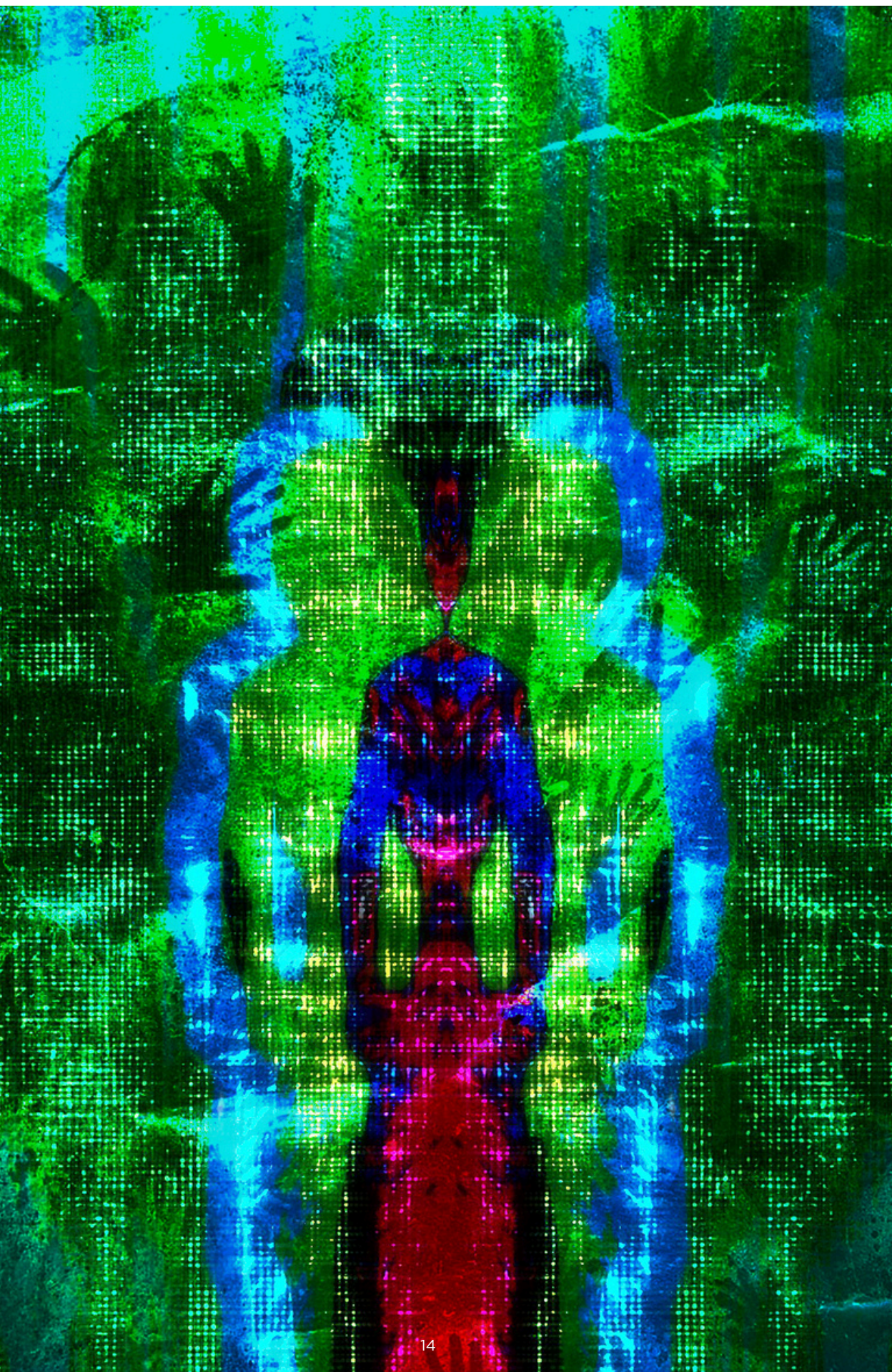
Women's experiences of online abuse can have overarching common and shared experiences, but can also diverge with regional differences that shape not only how women experience the harms but also the infliction and implications of deepfake abuse. These divergences have consequences for conversations on governance, regulation, institutional reforms, platform accountability, victim-survivor empowerment and support, and data collection for research and policy making. This demands a move away from the neutrality within institutions of power, towards a reckoning with the systemic nature of harm, and the power asymmetries underlying governance and regulatory endeavours when adapted for Global South contexts.[25] This forms a central basis for our approach to the analysis in this Report.

Marginalisation and intersectionality are central to understanding NCSI in India, because digital violence is not experienced as 'gendered violence' alone. It is shaped by the overlapping social positions of caste, class, religion, sexuality and geography. India's gendered digital abuse occurs within deeply unequal social structures, where women's sexuality, mobility, public visibility and freedom are policed through norms of family honour, respectability, and communal reputation. These norms enable sexualised deepfakes to act as a weapon that can violate privacy or dignity, but also threaten women's social status, familial relationships, safety, livelihoods and capacity to participate in public life. For religious minority women, for example Muslim women, NCSI can be layered with Islamophobic tropes and communal harassment, and for Dalit and Adivasi women, it can reproduce historical patterns of caste-based sexual violence, humiliation and narratives of 'pollution' that reaffirm caste hierarchy. For queer and gender-nonconforming people, sexualised deepfakes can compound stigma in a socio-political environment that remains hostile to their identities. Class and geography further determine vulnerability; rural women and those in tier-two cities can face intensified community

surveillance, be pressured to remain silent against abuse, and rely on informal dispute settlements which do not prioritize accountability. Intersecting identities shape access to justice because stigma, institutional bias, fear of retaliation, and the absence of disaggregated data on caste, sexuality and religion in digital crime reporting, effectively invisibilises harms. An intersectional lens is therefore necessary to study NCSI beyond a generic account of deepfake abuse, and to recognise NCSI as a structural and socio-technical form of gendered violence, and how the harms and barriers to redress are unevenly distributed across global, and Indian social hierarchies.

The scope of this research is restricted to use of NCSI against adults and does not include use of deepfakes, or online violence against children or Child Sexual Abuse Material (CSAM), as the considerations involved are different for minors and we did not wish to create any equivalences of experience, concerns, and challenges. While acknowledging that the appropriateness of terminologies for describing individuals with experiences of sexual abuse is a debated subject, this Report uses the terms survivor, victim and victim-survivor based on what may be deemed as the most appropriate usage in a given context.

The analysis in this Report has been divided across four sections. Section 1 provides an overview of the types of instances of NCSI abuse, focusing on the Indian context. Section 2 takes forward the insights and identifiable patterns from the preceding discussion to develop an understanding of the harms originating from the varied instances of NCSI abuse and their experience by the victim-survivors. It traces the various forms of harms originating from NCSI abuse by centering victim-survivor experiences within the socio-cultural norms and the misogynistic cultures in India. It maps these harms across six inter-related categories: dignity, economic and professional, public participation, bodily, privacy and consent. Based on this understanding of harms, Section 3 identifies the challenges in meaningful recourse for victim-survivors as well as the structural conditions, systemic and institutional gaps that aid the perpetuation of deepfakes against women. Section 4 uses the learnings from the preceding analysis to provide an overview of existing technical and legal measures against NCSI, their shortcomings and potential ways to address these gaps.



SECTION 1

NCSI abuse through the Lens of Intersectionality and Marginality

Instances of non-consensual synthetic imagery of famous women find most media scrutiny, but any woman can be a target of such abuse. Popularity is not a pre-condition for abuse, and neither is obscurity a guarantee of safety. Perpetuators of image based abuse can be strangers caught up in a voyeuristic frenzy, motivated by a networked storm of algorithmic virality like in case of the popular Grok incident on X. Or they can be colleagues or fellow students playing around with online tools meant to demean and embarrass. They can also be vengeful ex-boyfriends, or ex-colleagues after an interpersonal fall out in an effort to punish and discredit the victim. A marginalised victim identity will often add additional layers of harm, but it can happen to any woman who is interacting with the world in any way. This section provides examples and case studies from India and elsewhere for how NCSI exploits gendered social structures and other marginalities to actively harm women, discourage them from public participation, exert control and coercion over their actions, and ultimately advance abuse against them.

Further, in regional contexts such as those of South Asia, where notions of family honour are tied to preservation of women's subjugated statuses, the images need not be explicitly sexual to be dangerous. A recent study of image based abuse in Pakistan further showed that control and exercise of a woman's image through the hands of a bad actor can result in reputational and social harms that have long lasting repercussions for their public participation and personal freedom.[26] Families will often set restrictions on their movement and digital participation if there is even a whiff of perceived impropriety, regardless of the sexual nature of the images circulated.[27] In 2023, an 18-year-old in Pakistan was killed because a doctored image showed her appearing with a man.[28] In a similar loss of control over her own identity, a 28- year- old engineer in Nagpur took her own life after being a victim of AI impersonation online.[29]

Another form of NCSI is the use of AI to create unauthorised images that are not sexually explicit in nature. Instead it shows them being framed and posed in sexually provocative positions for their audience's voyeurism and satisfaction.[30] AI deepfakes – regardless of their sexual nature – wrest control of a person's ability to present themselves to the public and to their community, and robs them of the opportunity for self expression.

In such cases, it does not matter whether the content is real or fake – the mere illusion of social impropriety is enough to degrade, embarrass, and cause perceptible and, sometimes, irreparable harm.

I. Religious Identity

Religious minorities face compounded harassment due to their social identities. In India, the intersections of gender, islamophobia, and hate speech make NCSI an easy conduit for furthering abuse and marginalisation against Muslim women. Given the social and political vulnerabilities involved, it has serious implications for their physical safety, professional security, public participation, and social visibility. In a study analysing AI generated images and videos from May 2023 to May 2025, sexualized depictions of Muslim women received over 6.5 million interactions in the dataset studied.[31] Another Delhi-based Muslim model reported that her images had been manipulated with AI and shared on Instagram to make it seem like she was in the adult entertainment industry and her brother was her "pimp".[32] She has since been harassed online, taken a reputational hit resulting in lost revenue, and generally feels unsafe on the platform.

There is a long history of gendered Islamophobia online even before the advent of AI. In a widely publicised incident from 2021, Muslim women were harassed online by being shown to be "auctioned" in an Islamophobic app.[33] Women's rights activists classify this as a campaign of "sexual harassment, criminal intimidation, and cyber stalking" which "regulate their political speech and political participation."[34]

II. Casteism

In India, a person's caste is inextricable from their experience of the world, and this shows up online as well. A Dalit activist shared that in 2024, she and her husband had been the victim of AI-morphed images that threatened sexual and physical violence against her and their then 2-year-old child. She has reported this and other incidents of online abuse to the cyber cells to no avail.[35]

Further, interviews and qualitative research have shown that lower-caste women receive hate speech that is not only sexist and misogynistic in nature, but will often include casteist elements meant to undermine their credibility and demean without interacting with the substance of their arguments.[36] In addition to misogynistic hate, studies have shown that women on X (formerly Twitter) received casteist slurs in response to many of their posts.[37] A 2020 Amnesty study also analysed tweets sent to 95 women politicians during the 2019 election. Women from marginalized castes received 59% more caste-based abuse than general-caste women politicians.[38]

There is unfortunately, no formally reported data on caste-based NCSI harms in India at this time, since it is still early in the surfacing and documentation of this harm. As we will see in later sections, Dalit survivors

face a triple violence both online and off -- with discrimination based on caste, gender, and economic status put together. The absence of data should not be taken to mean the absence of harm to members of the community.

III. Queerness

In a conservative country like India which has no legal recognition of same sex marriage, and a social climate that is actively hostile to queerness despite decriminalisation of homosexuality,[39] LGBTQIA+ individuals navigate digital spaces both as a site of community building and targeted abuse.[40] The intersections of queerness and digital violence in the Global South are well documented:[41] in a highly publicised case in India, a 16 year old who shared make up tutorials on Instagram was a target of a barrage of homophobic abuse which continued with vehement vitriol after the teenager committed suicide.[42] The introduction of AI to this mix to create manipulated material will only further this trend of abuse for visible minorities. While there is no formal reported data on NCSI harms to queer people in India at this time, it is still early in the surfacing and documentation of this harm. The absence of data should not be taken to mean the absence of harm to members of this community.

We propose that this lack of reporting is due to the fact that National Crime Records Bureau (NCRB) does not disaggregate cyber crime data to include queerness or caste as a category, only gender.[43] Further, NCRB data is often out of date and unrepresentative of current trends due to the lag in reporting.[44]

It is also well reported that many people of intersecting marginalities do not prefer to self-identify on public fora because of the stigma and abuse that often comes with the disclosure, as discussed in the following sections.[45]

IV. Personal Disagreements

News reports show that personal disagreements are a common theme among the perpetrators' motives for creating this content. For example, a homemaker in Assam found that her likeness was stolen by an ex-boyfriend to create a sexualised AI influencer whose account was viewed millions of times before it was taken down. [46] The homemaker and her family did not initially know about the account, which had been online for years and had a large, dedicated following. The victim only filed a police report once the account broke containment of the original audience and went viral. While the account has now been taken down, images and videos created by the AI influencer still circulate on Instagram. In another incident, a Delhi man created AI morphed images of a 21-year-old woman in Ahmedabad, Gujarat, in a convoluted effort to manipulate her and gain her trust.[47] The woman reported online harassment and mental health harms for months before the perpetrator was caught.[48] In yet another case, an Assamese man created sexually explicit AI images of his former college classmate to tarnish her reputation after they had a falling out.[49] He harassed the victim through these images, tried to undermine her professionally by falsely

claiming that she works in the adult entertainment industry, and even monetised the AI-created images on subscription based platforms. Deepfakes provide an easy way for men to act out their frustrations and leave victims with social consequences that last for a long time.

V. Educational Institutions

Schools and universities are other sites where NCSI is proliferating. For teachers, who are both in positions of power and care over students, this content circulating within the community presents a threat to their authority and credibility as educators. For example, a police report shows two male students in Moradabad, Uttar Pradesh had created sexualised AI videos of a female teacher and circulated them on social media.[50] As a result of this type of harassment at the hands of their students, teachers in the United States too report mental health harms and that they reconsidered their relationship to teaching.[51] Not only is this disrespectful to teachers as disciplinarians in positions of authority, as an employment issue it is also a classic case of sexual harassment in the workplace.

Further, although NCSI that involves minors as targets would be considered CSAM, and is beyond the scope of this study, it is worth noting the proliferation of deepfakes as a systemic problem across institutions, with minors as perpetrators as well as victims. There are countless cases of female students being cyberbullied by their peers through the creation of AI generated intimate imagery. In a particularly harrowing case, a 10th class student in West Bengal took her own life after being a victim of NCSI that was created by a man living in her neighbourhood, and shared on social media.[52] A college student in Chhattisgarh created sexualised deepfakes of 36 of his female classmates creating an unsafe environment for his peers.[53] In Assam, three school boys were suspended for creating 64 sexually explicit images using AI of female peers and teachers.[54] A survey of 1200 young people aged 13-20 in the US reported that 1 in 10 people personally know someone who had NCSI created of them.[55] A similar survey conducted in South Korea of 1500 college students, over 14% of them said that they had created deepfake images or videos, and 1 in 5 say they created them for their own sexual gratification and for harassing the victim.[56] There are reports from the UK, Spain, Australia, and Brazil where students are creating and sharing AI-generated intimate images of their peers in an effort to bully and embarrass them.[57] No female student in any part of the world seems to be safe from their unconsented sexualisation and humiliation by their peers.

VI. Scams and Blackmail

In regional contexts such as those of South Asia – where notions of family honour are tied to preservation of women's statuses and "clear records" of fidelity and propriety – women are susceptible to blackmail and scams from bad actors who seek to profit off of their embarrassment. In an illustrative case, two men from Uttar Pradesh, created sexually explicit AI-generated images from pictures the victim had uploaded on her social media accounts, then blackmailed the woman with threats of circulating

the AI-morphed images online.[58] After the woman paid them, they blackmailed her new husband in an effort to embarrass him through his relationship to her. At the time of writing, the accused are still at large and police suspect they may have run this scheme on several other women.[59] This AI-facilitated blackmail is also increasingly being used as a tactic by predatory loan sharks who, in addition to verbally abusing and harassing the borrowers, threaten to release sexually-explicit AI-generated images to their communities via their phone's contact lists.[60] Investigative reporting has shown that unscrupulous loan apps offer almost instant money transfers to ease financial burden, but in the process create copies of the borrowers' contact lists, photo libraries etc. which are later used to coerce them to comply with unreasonable re-payment schedules.[61]

Table 1: **How NCSI harms women, case by case**

THE HARM	WHAT IT AFFECTS?	CASE STUDY/EXAMPLE
DIGNITY AND SELF-REPRESENTATION	Control over her own image, identity and sexual representation.	A homemaker discovered that an ex-boyfriend had used her likeness to create a sexualised AI influencer account. The account attracted millions of views and had existed for years before she became aware of it.
ECONOMIC AND PROFESSIONAL HARMS	Credibility, income, job chances and career growth, even when people know the content is fake.	Following her medal-winning performance as an Indian powerlifter, Keyaa Banerji reported sexual harassment, impersonation and thousands of AI-generated images using manipulated versions of her face and body. She eventually created a new public account to regain control of her online identity.
HARM TO PUBLIC PARTICIPATION	Place in public, professional and digital spaces, as the threat pushes her to step back.	Images and videos of female protesters were AI-manipulated and circulated by malicious social media accounts in an effort to discredit the protests. This was accompanied by misogynistic online commentary and targeted abuse of a woman who became prominent during the protests.
BODILY AND PSYCHOLOGICAL HARMS	Health and safety: anxiety, trauma, hypervigilance, poor sleep and fear of physical safety.	A 28-year-old engineer died by suicide after reportedly experiencing AI impersonation that she had no control over.
PRIVACY AND AUTONOMY	How the victims' images, identity and intimate representations are used and shared.	A Muslim model discovered that her images had been manipulated and circulated on Instagram to falsely portray her as working in the adult entertainment industry, with her brother depicted as her pimp. She experienced online harassment, reputational damage and lost revenue.
CONSENT AND COERCION	Sharing an image is not agreeing to have it altered, sexualised, spread or sold.	Two men allegedly used a woman's social media photographs to create sexually explicit AI images and blackmailed her with threats of circulation. After she paid them, they also attempted to blackmail her husband.



SECTION 2

Understanding the Harms

As discussed in the previous section, we have seen the layers of social and cultural factors that shape the experience of women who are victim survivors of NCSI in jurisdictions such as India and the nature in which such intersectionality of identity can complicate the adequate redressal of NCSI. Dominant AI governance notions have been argued to assume Western norms of individualism.[62] While these liberal norms have allowed for the constitutionalisation of individual rights and liberties across much of the post-colonial world, including India,[63] individual-centric norms are unable to fully account for and provide meaningful redress within the complex social and cultural diversities of the Global South.[64] Arguably, then, besides prevailing governance and policy measures that target individual rights and impacts, there is a need to contextualize policy and governance proposals to adequately address the societal and structural differences within the Global South.[65] This will ensure governance and policy measures that demonstrate technological sensitivity and center user needs and experiences from varied cultural contexts.

This section attempts to lay out a matrix of harms arising from NCSI. A harms matrix helps break down the research and understanding of NCSI by refusing to treat 'deepfake harms' as a single, uniform injury. In the previous section, we discussed case studies and how various intersectional identities come to bear, and in this section, we map the layers of harms that surface within these cases. This chapter identifies dignity harms, participation harms, economic and professional harms, bodily and psychological harms, and privacy and consent harms - showing how each of these operate, and how they intersect and overlap. This breakdown of the harms matters especially in Global South contexts, where caste, religion and community norms shape who is targeted, how severely they are targeted and whether they can access support or justice. By mapping these layers of injury, a matrix of harms makes it possible to identify patterns that would otherwise be hidden, and to design responses that address the full range of damage caused by sexualised deepfakes - not just the visible parts.

In this section, we discuss privacy, consent and dignity as individual harms, among others. Privacy, consent and dignity operate as distinct but overlapping normative foundations in responses to NCSI. Privacy centres on 'control' over one's body, image and intimate representation, and is a legally protected position.[66] As scholars have stressed, it includes the violation of the sexual self-determination of the subject.[67] Consent

foregrounds the absence of voluntary, informed authorisation at each stage of creation and distribution of NCSI. Harm to dignity captures the ontological injury of being unexpectedly sexually exposed, regardless of whether privacy was breached or consent was formally obtained.[68] Conceptually, these three are deeply interlinked and cannot be isolated when researching the harms. While our matrix traces harms within each of these categories separately, our mapping highlights the overlaps and intersections between all the identified categories within this section.

I. Harm to Dignity

Harm to dignity represents the primary experience of NCSI, because it severs the link between a woman and her own sexual representation, treating her body as raw material for public consumption and control. Critical and feminist scholarship increasingly frames sexualised deepfakes not primarily as misinformation or deception, but as a form of image-based sexual abuse that undermines agency, dignity and autonomy, while producing collective harm by creating a climate of threat that constrains women's participation in digital spaces.[69] The central injury is not whether viewers are fooled, but that a sexualised version of the person is created and circulated without their consent, wresting control of their likeness and grafting it onto situations they did not choose.[70] It is important to further note that the dignitary injury and loss of control inherent in NCSI are not contingent upon the sexual nature of the depicted conduct.[71] The degradation arises fundamentally from the unauthorised appropriation and recombination of a person's likeness to fabricate actions they did not perform.[72] The sexualisation is an aggravating layer of harm, but the core violation - the severing of the link between private self and public representation, persists irrespective of whether the content is sexual in nature.

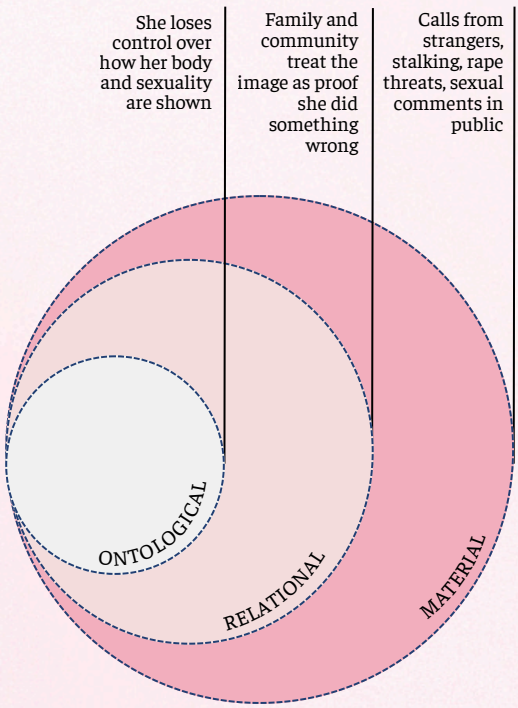
Deepfakes extend existing practices of visual domination (pornography, revenge porn, morphed images) by using AI to decouple face, voice and body from the person's will, enabling malicious actors to construct intimate scenarios that never occurred.[73] The non-consensual creation or sharing of intimate imagery is a form of sexual abuse that robs the individual of dignity and security in their own persona, violating what Indian scholarship describes as "informational sexual autonomy": the right to control how one's sexual identity is constructed, represented and disseminated in informational environments.[74] For women, whose bodies are already policed through norms of modesty, honour and respectability, this loss of control is experienced as a fundamental assault on personhood: they are made to inhabit a sexualised identity imposed from outside, over which they have no authority.[75]

In misogynistic societies, the creation and dissemination of fake images itself harms dignity due to the implications for a woman's day-to-day lived experience. Once NCSI is in circulation, it can trigger doxxing, lewd phone calls, stalking, rape threats and other forms of harassment that move the harm from the screen into everyday life.[76] Survivors report being

contacted by strangers who have seen the material, being subjected to sexual comments in public spaces, and are forced to navigate heightened surveillance from family and community who treat the existence of the image as evidence of moral failure.[77] The dignity harm thus operates on multiple levels: it is ontological (a violation of self-representation and personhood), relational (a disruption of how others see and treat the depicted individual), and material (a catalyst for concrete threats and intrusions into daily life).[78]

This is especially acute in contexts like India, where criminal law has historically regulated sexualised images through the lens of obscenity, defamation and "modesty," rather than through a framework of bodily autonomy and sexual dignity.[79] Within such a paradigm, the focus is often on whether the image is obscene or damaging to reputation, rather than on the violation of the woman's right to control her own sexual representation. [80] NCSI exposes the inadequacy of this approach: even when the image is known to be fake, the harm to dignity persists because the woman's body has been used without consent to produce a sexual narrative that she must then live with and against.[81] The upcoming section on consent will also delve into how NCSI reveals inadequacy of existing approaches. Recognising dignity harm as central to NCSI therefore requires shifting from a narrow focus on falsity or reputational damage to a broader understanding of sexual autonomy, personhood and the right to exist in public and private spaces without being sexually defined by others.[82]

Figure 2: **Nature of Injury within Harm to Dignity: ontological at the centre** (loss of control over self-representation)**, relational in the middle** (family and community surveillance, being treated as morally at fault)**, material on the outside** (strangers' calls, stalking, rape threats, comments in public spaces).



II. Economic and Professional Harms

NCSI produces concrete employment consequences, even if viewers know, or suspect, that the material is fabricated.[83] Sexualised deepfakes draw on persisting gendered assumptions about women's sexuality, credibility, and reputation.[84] Employers, co-workers, clients and professional networks might associate the targeted person with the content, resulting in workplace stigma, bullying, loss of trust, fear of reputational harm by association, or even dismissal.[85] The recruitment and job application process is particularly vulnerable to this form of harm, since employers routinely conduct online background checks on prospective employees, and digital traces of NCSI connected to an individual can affect hiring decisions, even if the imagery is known to be manipulated or false.[86] Regulatory systems in the Global South focus on detecting the synthetic and manipulated nature of AI content and its imitation of reality, which "overlooks structural exclusions embedded in AI design, deployment, and governance,"[87] thus failing to acknowledge the systemic nature of these harms.

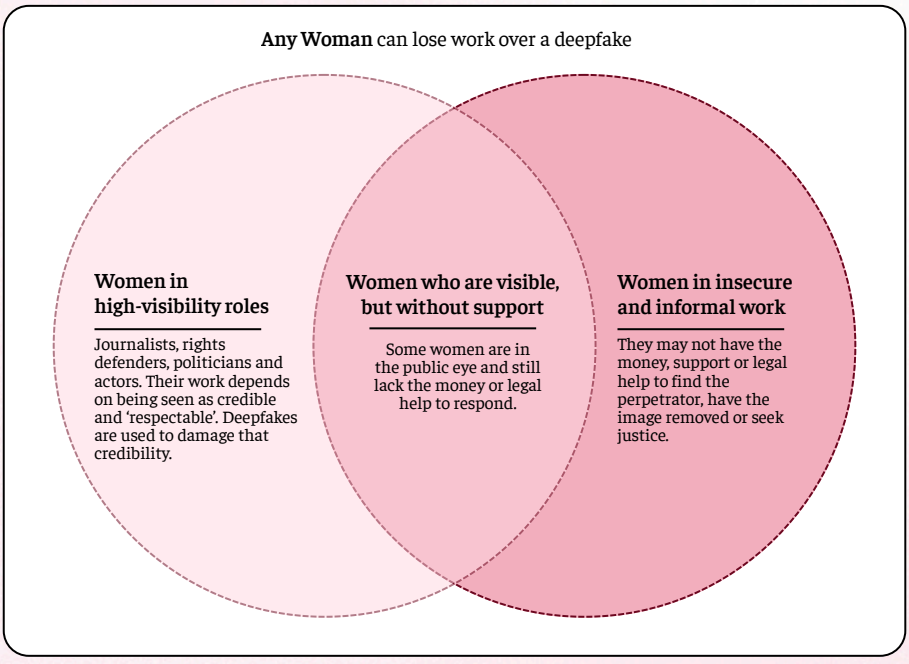
The professional damage is rarely limited to the initial circulation of NCSI. The possibility to download, reupload and redistribute NCSI across platforms makes them repeatedly discoverable through ordinary searches of a survivor's name and allow sexualised associations to persist long after the content has been taken down.[88] This can undermine professional credibility, weaken relationships with clients and colleagues, affect promotions and damage career mobility, despite the complete absence of any wrongdoing by the victim. This impact is acute in professions dependent on public trust and visibility, such as journalism, politics, academia and advocacy, where NCSI can be deliberately deployed to recast women as sexual objects rather than credible professionals.[89] Indian powerlifter Keyaa Banerji's experience reflects how NCSI can transform professional visibility into a source of sustained vulnerability.[90] After her medal-winning performance attracted public attention, Banerji reported sexual harassment, impersonation and thousands of AI-generated images using manipulated versions of her face and body. Disturbed by the AI-generated morphed images of herself, fake accounts impersonating her, and Instagram's suspension of her original private account, Banerjee decided to create a new public account to regain control over her online identity, and personal and professional narrative.[91] The onus of maintaining her autonomous identity, by having to create a new account also speaks to the constraints and added work load NCSI creates, where the victims are left to bear the practical costs of repairing a harm they did not cause.

These harms are also likely to be most severe for marginalised women in precarious, informal or public-facing work (such as, journalism, politics, social service, or government services), who may have limited financial resources, institutional support or access to legal assistance needed to identify perpetrators, obtain removal, and pursue redress.[92] Time spent documenting abuse, navigating report systems, managing safety concerns

and responding to reputational fallout can itself interrupt employment or education, while fear of further exposure may cause survivors to withdraw from professional networks and public opportunities.[93] In this way, what begins as a single act of NCSI becomes a durable economic injury, reducing income, limiting advancement, and reshaping a survivor's long-term career choices.

Women in sensitive or high-visibility roles such as investigative journalists, human-rights defenders, politicians or even actors playing feminist characters, are especially vulnerable since their professional authority depends on being perceived as credible and morally 'respectable' in patriarchal contexts.[94] NCSI is deployed strategically as a weapon designed to erode their legitimacy, generating harassment from audiences and employers, and deterring them from taking on controversial assignments.[95] Multiple recent cases illustrate these dynamics. Journalist Rana Ayyub has long been the target for online hate and abuse in response to her investigative[96] work, and in the last few years, this doxxing has also been extended to include sexually explicit deepfake videos. In September 2026, actress Shruti Hassan moved the Bombay High Court alleging circulation of 'AI-generated sexually explicit videos and images' featuring her, alongside fake endorsements and unauthorised merchandise, and actress Preity Zinta approached the court for similar content, including chatbot-style conversations using her likeness without consent.[97] In both instances, the Court directed social media intermediaries to remove the identified content, noting prima facie harm to their reputation, goodwill and commercial interests.

Figure 3: **Two kinds of exposure to NCSI: being seen and lacking the means to respond**



These instances also act as examples of how perpetrators can employ NCSI to profit off the identities, manipulating images to feed into a prurient and objectifying male gaze.[98] Indian courts, however, do not seem to view violations of personality rights as an umbrella - they litigate sexually explicit content but call for further proof of monetisation and commercial exploitation, even if said exploitation is rooted in the NCSI content.[99] In actresses Janhvi and Khushi Kapoor's suit, the court did not resolve the distinction on where the content attracts views that are later monetised through advertising, or where names and likeness are used indirectly to promote merchandise or services.[100] Interestingly, this verdict was a departure from previously awarded broad injunctions protecting personality rights of male-presenting public figures, including actors, politicians and entrepreneurs.[101] This suggests a broader doctrinal uncertainty in Indian jurisprudence; protection is being applied less consistently when sexualised misuse is involved, particularly where commercial exploitation is indirect, or difficult to trace.

The deliberate deployment of NCSI against women in public roles also reveals how control and misogyny continue to operate as core motivations. By weaponising sexual shame to undermine credibility and further deter public participation, perpetrators use NCSI not merely to cause individual distress but to reinforce patriarchal hierarchies that seek to keep women (especially those who challenge power) as subordinates, and to silence and marginalise them.[102] As a form of abuse, NCSI makes social existence costly, by attaching the threat of sexual humiliation to women's exercise of professional, political and expressive agency.

III. Harm to Public Participation and Free Expression

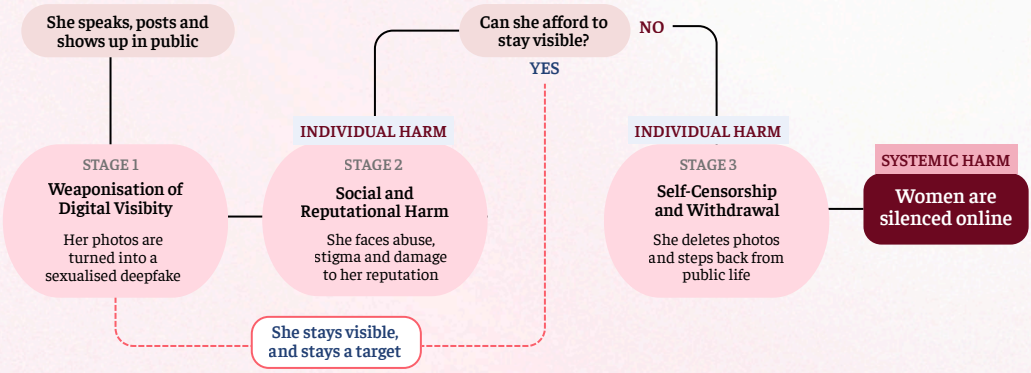
These instances also act as examples of how perpetrators can employ NCSI to profit off the identities, manipulating images to feed into a prurient and objectifying male gaze.[98] Indian courts, however, do not seem to view violations of personality rights as an umbrella - they litigate sexually explicit content but call for further proof of monetisation and commercial exploitation, even if said exploitation is rooted in the NCSI content.[99] In actresses Janhvi and Khushi Kapoor's suit, the court did not resolve the distinction on where the content attracts views that are later monetised through advertising, or where names and likeness are used indirectly to promote merchandise or services.[100] Interestingly, this verdict was a departure from previously awarded broad injunctions protecting personality rights of male-presenting public figures, including actors, politicians and entrepreneurs.[101] This suggests a broader doctrinal uncertainty in Indian jurisprudence; protection is being applied less consistently when sexualised misuse is involved, particularly where commercial exploitation is indirect, or difficult to trace.

The deliberate deployment of NCSI against women in public roles also reveals how control and misogyny continue to operate as core motivations. By weaponising sexual shame to undermine credibility and further deter public participation, perpetrators use NCSI not merely to cause individual

distress but to reinforce patriarchal hierarchies that seek to keep women (especially those who challenge power) as subordinates, and to silence and marginalise them.[102] As a form of abuse, NCSI makes social existence costly, by attaching the threat of sexual humiliation to women's exercise of professional, political and expressive agency.

Deepfakes are systematically and deliberately weaponised in a coordinated effort against women with a public presence, to prevent them from participating and expressing themselves in public spaces, including digital platforms.[103] For instance, there tends to be a notable rise in deepfakes against female politicians ahead of elections.[104] Before her election, Chandrani Murmu, a young Indian parliamentarian, was targeted by a deepfake in which her face was digitally superimposed onto an obscene video.[105] In another instance from Karnataka, India, studied in a recent report, an Instagram content creator found her reels AI-manipulated to depict her as nude.[106] The report pointed to the potentially systematic and targeted nature of the abuse that suggested a regional troll ecosystem against women from the region.[107] Existing research has sought to demonstrate how this kind of harassment and abuse discourages women politicians from expressing their views on the platform or even standing for an election; and the same can be extended to professionals in all fields.[108] Owing to the resulting trauma and emotional toll, TfGBV results in survivors withdrawing from digital spaces, self-censoring, and participating in the public.[109] Many times, survivors are forced to make this choice in order to prevent a loss of education, employment, housing, or family sanctions on their basic freedoms of movement and social engagement. The harm to public participation, especially when expressing online, is not just after-the-fact but also arises from a mere threat of manipulation of one's online images with AI.[110] This leads to a chilling effect for speech, expression and free participation in both physical and online spaces. Deepfakes weaponise visibility - they utilise the digital footprint left by a woman's online presence to turn it into material for her silencing, discrediting, and personal annihilation across the various spheres of her life.[111] In Delhi's recent protests by students, images and videos of female protestors on the site were AI-manipulated into lewd or objectionable material by malicious accounts on social media in a campaign to discredit the protests, and resulting in a deluge of misogynistic comments on social media.[112] This also included targeted deepfake and other forms of online abuse against an actor and model who went viral during these protests for blocking a police van to free the protesting students.[113]

Figure 4: **How NCSI forces women to choose between staying visible and staying safe**



Public participation requires a wider understanding than public presence, which is commonly associated with politicians, celebrities, or social media influencers. However, public participation may be said to be tied closely with women's right to occupy public spaces that connects to their pursuit of friendships/personal social relationships, education, a career, free expression, and a sense of safety in public and online spaces. Significantly, women's public participation is also shaped by a stripping away of safety at home. The vitriol resulting from NCSI threatens actual bodily safety, and the resulting suspicion at home leads to increased surveillance, censorship and restrictions on women's means of participation in South Asian societies. In contexts where family and community 'honour' is tightly bound to women's sexual purity, the bodily harms of NCSI are further compounded by the risk of punitive violence directed at the survivor rather than the perpetrator.[114] Honour-based reasoning can lead to communal restrictions on movement, forced confinement, withdrawal from education or work, and in extreme cases physical punishment or 'honour' attacks justified as restoring family reputation.[115] Survivors may face pressure to remain silent, to marry quickly, or to leave their homes, while community retaliation, ranging from social boycott to physical assault, can be triggeredby the mere existence of sexualised imagery, regardless of its authenticity.[116] All of these implications relate directly with a woman's pursuit of basic liberties and translate into an inability to participate socially, or in public, or safely occupy educational, professional or public spaces.

This dynamic means that the social ecosystem treats the victim's visible sexuality as a social stain requiring correction through control and violence. Women halt and retreat from their ambitions, silence themselves online, and avoid having social media accounts and photos.[117] Any form of visibility in the public, including otherwise mundane speaker engagements at professional or public events pose a threat of finding oneself in a potential non-consensual deepfake, for all that a creator of a deepfake

needs is a face.[118] For example, online abuse faced by women on dating apps has been studied to find that it pushes women to altogether withdraw from dating apps.[119] With the easy manufacture of deepfakes through AI and an available image, this harm from dating app use gets compounded further. Harm to public participation is therefore wide-ranging and multidimensional in its span, with serious implications for women's equality and democratic participation.

IV. Actual Bodily Harms

Harms associated with gender based violence are often viewed narrowly from the perspective of physical or bodily harms, due to which harms arising from TfGBV are often not viewed from the perspective of bodily harms. In the context of TfGBV and its various forms, this distinction between online and offline is a limited way of looking at this issue, and there is a need to view bodily harms to encompass psychological harm. When it comes to the varied nature of harms from technology facilitated violence, the question is "whether harm has to be enacted on the body of a subject to be thought of as harm."[120] Digital violence is frequently treated as less serious than physical sexual abuse, yet the bodily consequences of NCSI demonstrate that psychological violence can manifest in concrete somatic and behavioural harms.[121] Although gender-based violence has come to be understood as a public health concern, the same emphasis is yet to be found for TfGBV.[122] Researchers argue that harms from TFSV can be similarly tangible in both the bodily and psychical experience and central to our everyday lived experience in the physical realm. For instance, Henry and Powell contend, "The harms experienced by women in the 'sociospatial'[123] world may have at least as much impact on a person as traditional harms occurring against the physical body."[124] They emphasise that the first step to theorising TFSV is for harm to no longer be thought of in relation to the physical or bodily harm in our "techno-social lived experience." Accordingly, the need to understand the harms of TFSV is unique.[126] They dismantle the presupposed dualism between the body and the mind within law and policy that results in the obfuscation of "social and intraphysical harms", to argue that these harms are "embodied, real and tangible."[127]

Feminist and trauma scholarship increasingly rejects a sharp distinction between 'online' and 'offline' injury, showing that the body registers digital sexual abuse through chronic stress responses, altered self-perception and changes in everyday movement and social interaction.[128] NCSI, as a form of image-based sexual abuse, like rape and sexual assault, can produce complex trauma. Survivors report anxiety, depressive disorders, post-traumatic stress symptoms, intrusive thoughts, hyperarousal, avoidance behaviours and, in some cases, non-suicidal self-harm and suicidal ideation.[129] Each resurfacing of the material - through re-sharing, search results or new notifications, can reactivate trauma pathways, producing renewed panic, shame and hypervigilance that are experienced physically as sleep disturbance, headache and gastrointestinal problems, which are typical stress-related symptoms.[130]

Psychological violence in NCSI cases must be taken more seriously in developing legal recourse. Empirical studies show that mental health consequences for victims are comparable in nature and severity to those experienced by survivors of rape and sexual assault, including profound shame, humiliation, fear and social isolation.[131] Experimental research on sexualised deepfakes further finds that women anticipate intense disgust, anxiety and anger when imagining themselves as victims, with NCSI generating stronger negative emotions than non-sexual manipulated content.[132] These emotional responses are not merely feelings, but embodied reactions that shape how survivors inhabit their bodies, with reports of feeling alienated from their own image, distrusting mirrors and cameras, and experiencing their bodies as permanently contaminated or exposed.[133]

The non-consensual visibility resulting from NCSI also translates into direct bodily safety. Survivors often describe changing their routines, withdrawing from social spaces and limiting their physical mobility out of fear of being recognised, confronted or attacked by someone who has viewed the NCSI.[134] For women in the public domain like journalists, activists, politicians and performers, these can include credible threats of physical violence, stalking and doxxing that transport the harm from the screen into the street.[135] Legal and critical analysis on TfGBV reveals that such threats are not ancillary but central to the harm.[136] The knowledge that one's image is circulating in sexualised form can make the world feel pervasively unsafe, producing a state of constant bodily alertness that itself constitutes injury.[137]

NCSI demonstrates that sexual violence can be enacted through the unauthorised capture, fabrication and circulation of intimate representations, producing durable psychological and somatic injury, constraining bodily freedom and, in some cases, precipitating direct physical danger. Regulation must therefore treat the mental health impacts, behavioural changes and safety threats associated with NCSI as actual bodily harms, not secondary or symbolic effects, and design legal, medical and support responses accordingly.

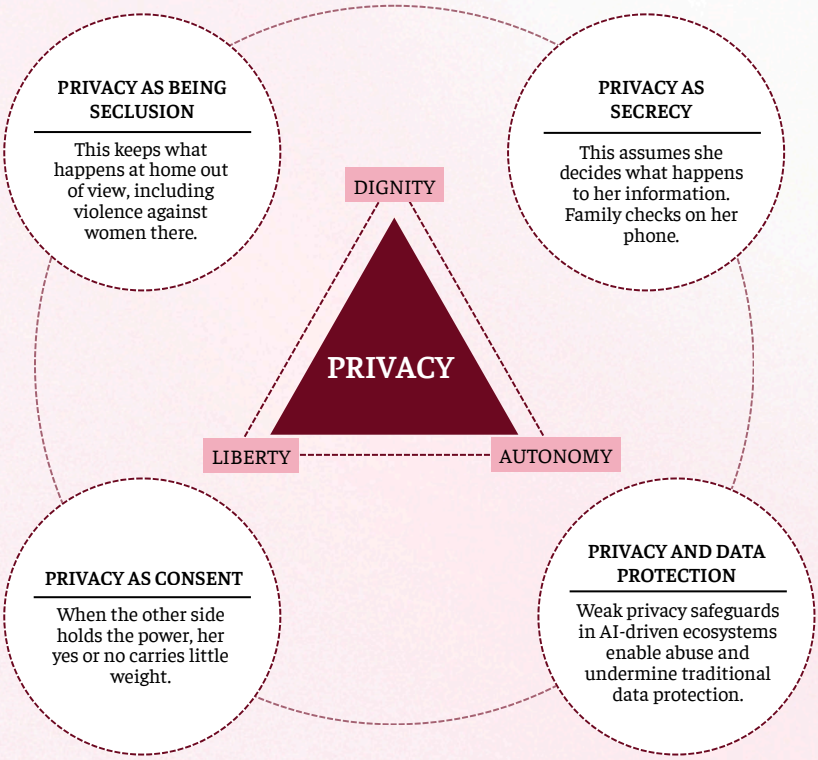
V. Privacy Harms

The Indian Supreme Court's verdict in Puttaswamy[138] laid important foundations as well as raised critical questions to the understanding of the right to privacy in Indian constitutional jurisprudence. Puttaswamy's dominant reading of the right to privacy, and as relevant for our purposes, has come to be understood as conceptualised within a 'dignity-liberty-autonomy triangle'.[139] This means that privacy's rationale may be said to be situated within the "notion of individual liberty operationalized through the ideas of autonomy and dignity."[140] The exact relationship of privacy with these three rights has been debated in the context of the operationalisation of the right to privacy.[141] However, that remains beyond the scope of this discussion. What remains relevant for our purposes is that

the conceptualisation of privacy in India goes beyond narrow notions of the private realm, consent, and secrecy, and cuts across notions of dignity and autonomy.[142] This can also include a capacity to exercise autonomous choices and the right to control dissemination of information about oneself, and bodily and sexual autonomy. [143]

Women's sexual privacy in India is shaped by patriarchal structures of honour, shame and respectability that position female bodies as markers of family and communal morality.[144] Feminist scholarship shows how women's sexuality is regulated through expectations of purity and modesty, with perceived transgressions triggering social exclusion, abuse or violence.[145] Privacy law, while recognising bodily autonomy, can obscure lived realities by treating privacy primarily as an individual right, rather than a site of gendered power and control, as will be discussed in this section.[146]

Figure 5: **Four common ideas of privacy**



NCSI operates in a contested terrain, undermining what limited privacy women are able to carve out by imposing a sexualised identity upon them without consent and broadcasting it to an expanding, potentially unlimited audience.[147] Even when the depicted act never occurred, the use of a woman's face, body or likeness to construct and circulate sexualised content intrudes upon her intimate autonomy.[148] Unlike defamation, which

concerns damage to reputation through false statements, the NCSI-privacy injury lies in the unauthorised creation and dissemination of an intimate representation regardless of its 'artificial' nature.[149] The 'artificial' element to deepfakes creates a legal conundrum - where injuries from NCSI are retrofitted into existing legal injuries like defamation, which effectively erases the unique harm that sexual deepfakes create.

In this way, NCSI creates a distinctive privacy injury. It does not disclose an existing intimate act, but it manufactures one that could still be socially attributed to the identifiable person.[150]

Privacy as Autonomy

The loss of privacy in instances of TfGBV may be wider in reach than in cases of physical abuse. In cases of physical abuse, laws are designed to ensure non-disclosure of survivor identity in judicial and journalistic reporting.[151] In the instances of TfGBV, a survivor loses complete control over information about herself. In privacy jurisprudence, this is also known as the right to informational self-determination,[152] and this right is completely lost in instances of TfGBV. While a survivor of physical abuse may be able to guard the experience with secrecy so as not to be identified as a victim, a similar nature of control is beyond a victim-survivor of TfGBV who finds herself in the midst of multiple mentions, replications and circulations of abusive content.[153] The persistent presence and continued visibility and consumption of sexually violative, non-consensual deepfakes in the public domain itself serves as a repeated violation of her dignity, autonomy and privacy In other words, a survivor may be able to find no escape from the violence caused to her through the creation and dissemination of an NCSI.

In this way, NCSI undermines the privacy claim at its core: by fabricating and disseminating sexualised images, it strips women of the power to decide who sees them, in what context and for what purpose, converting their likeness into raw material for public consumption and moral judgment. The harm begins at creation and is compounded by each act of circulation; the falsity of the material does not mitigate the intrusion but rather underscores how privacy is violated through the unauthorised sexualisation of identity itself.[154]

Personality rights (protecting name, image, voice and other identifying features) offer one legal vocabulary for this injury, recognising that these features are integral to identity, reputation and economic opportunity.[155] Yet, NCSI exceeds reputational harm: it is a privacy violation that expropriates the body, producing a persistent sense of exposure and loss of control over one's digital self. In the case of India, the privacy jurisprudence, shaped significantly by Supreme Court's landmark verdict in Puttaswamy,[156] already enables protection against such violation by protecting decisional autonomy over intimate matters, including how one's body is represented.[157]



Privacy is Gendered

Feminist legal theory emphasises that privacy and boundaries are lived through the body and shaped by structural inequalities of gender, caste and class.[158] Notions of privacy for women in the Global South tend to be shaped by ideas of shame, that tie closely to familial honour and used as a means of social control.[159] As discussed earlier, this enables social surveillance of women's public participation, free speech, and sexual self-determination, towards preservation of patriarchal and misogynistic social structures. This heavily shapes women's experiences of safety and risk.[160] When it comes to women's online existence, it finds itself up against the same patriarchal norms and surveillance that seeks to control women's autonomy and self-expression, such as through monitoring of digital activities by family members, or higher moral scrutiny and policing of women with online presence and incidents of hate speech against women.[161] For instance, a Kerala study on rural women's smartphone use illustrated how intersecting identities shape digital autonomy even in a high-literacy, formally progressive context, with patriarchal norms, caste-based labour roles and religious customs constraining women's control over their devices and images.[162]

In view of the continuum between the offline and the online, where the online can significantly influence one's everyday lived experience, and where online exposure for women has significant social costs attached, privacy needs to be viewed through a gendered lens. Community surveillance over women's everyday lives in an offline world, such as through "neighbourhood whispers, reputational sanctions or control of movement" continues in her online existence through platform-enabled scrutiny, executed via cyberstalking, doxxing, gendered and communal hate speech, and NCSI, with implications for her offline existence.[163]

Privacy and Personal Data

Scholars note the limitations of conceptualising privacy in any one particular way: privacy as seclusion risks masking violence at home; privacy as control over data risks ignoring the structural realities limiting meaningful exercise of this agency, and privacy as consent risks being diffused by power asymmetries.[164] Looked at this way, a complex picture of privacy emerges from the preceding discussion.

Given the layered nature of privacy, it is imperative to look at privacy beyond narrow notions of secrecy, and within the systems of power that define its limits. In the prevailing infrastructures of digital platforms, there is no true agency with people to define privacy for themselves. Agency loses its meaning with the way platforms are designed with default settings that can expose contact lists, auto-linking of phone numbers to social profiles, and no ways of accessing a service without a complete opt-in.[165] Traditional data protection principles such as purpose limitation and minimal data collection then lose meaning when access to any digital service provides no meaningful exercise of rights but only a wholesale opt-in to the service offered. This creates digital vulnerabilities easily exploited

by bad actors and platforms for targeting minority and vulnerable communities of individuals, who have little oversight or agency or understanding over the utilisation of their personal data. In an age of AI, this has accelerated to companies using data of individuals on their platforms to train their large language models through default user opt-ins.[166] Infrastructural designs such as these create an ecosystem that does not take privacy seriously, and effectively, furthers ecosystems of abuse[167] - where models trained on non-consensual sexual content allow users to turn any image into NCSI,[168] platforms allow such content to circulate, and victims are doxxed online, all the while desperately reporting NCSI for takedowns.[169]

In a digitised world taken over by AI, personal data has been commodified to the advantage of corporations, while perpetuating pre-existing systemic inequalities.[170] This commercialisation of personal data normalises surveillance, diminishes individual control over use of one's personal information, and disproportionately impacts women and people from marginalised groups, along the lines of caste, race, religion, or sexuality, by aiding reinforcement of pre-existing discriminatory and surveillance logics against them.[171] Legal and institutional responses must therefore move beyond narrow conceptions of privacy as secrecy, content moderation and takedown, or content labelling and watermarking, and recognise NCSI as a violation of intimate autonomy and personality rights that begins at creation, persists through circulation, and cannot be negated by the material's falsity.

VI. Consent

Consent, as traditionally construed in data protection and criminal law, is largely absent or illusory in the context of NCSI. India's DPDP Act, 2023, centres consent as "free, specific, informed, unconditional and unambiguous" for data processing, yet it operates as a civil regime with monetary penalties. It was not fully operative for key provisions until late 2026, offering little deterrence against the non-consensual extraction and sexual recombination of biometric data such as faces and voices.[172] Criminal law similarly fails: while some provisions penalise the sharing of intimate images "appearing to show" a person, there is no express offence targeting the non-consensual creation of AI-generated sexual imagery, leaving a gap where consent is neither sought nor obtainable because the victim may never have interacted with the perpetrator.[173] Consent in this context becomes a fiction: posting an ordinary photograph or sharing it privately does not authorise sexual alteration, AI generation, or public distribution, yet, by not criminalising creation explicitly, the law effectively treats such acts as 'implied permission'.

From a critical legal and feminist perspective, practising consent as a Global South woman is unevenly distributed: those with economic security, caste privilege and institutional support can better absorb reputational loss and pursue redress, while marginalised women face compounded harms



where any perceived sexual transgression threatens family honour, community standing and physical safety.[174] This uneveness is also analogous to marital rape, where Indian law's presumption of implied consent within marriage erodes bodily autonomy and treats women's bodies as available regardless of their actual will.[175] Marriage in itself becomes a continuing, unending form of consent, which cannot be revoked verbally within the boundaries of the relationship, regardless of the woman withdrawing her consent. When reading this analogy for NCSI, the inaccurate presumption of consent is rooted in the mere act of having a digital presence, whether that is with social media accounts, or images existing online in the public domain. Viewing NCSI through this lens exposes how consent doctrines, whether in data protection, criminal law or social norms, often function to legitimise rather than prevent violation.[176]

Existing understanding of consent in the realm of digital media is situated in the assumption that they originate from an identifiable physical source or actors that involve a subject, a creator and the moment that was captured in the media. Synthetic media such as that in NCSI, does not require a subject or a particular moment of capture and the creator can utilise a generative AI model's data extraction and modification to obtain any form of media.[177] Researchers argue that "this creates a category mismatch where consent frameworks designed for conventional media become inadequate for AI-generated content which has no stable origin, author, or authentic referent."[178]

Moreover, consent has taken on a different relevance that goes beyond what is usually captured within ideas of privacy and representation, and now stands as an infrastructural concern, where consent may be thwarted by the design of a technology. Hence, in the case of AI-generated deepfakes, consent needs to be imagined within the technological lifecycle of synthetic media, which is an "ongoing pipeline of extractions (scraping faces and voices), transformations (generation/face swappings, nudification), circulations (uploads, reposts, encrypted groups), and amplifications (recommendation systems and group dynamics)."[179] This lifecycle may be said to constitute the relevant infrastructure for NCSI creation and dissemination, within which, consent gains relevance for processing aspects of personal identity or likeness, modification of one's likeness, scale and nature distribution of the resulting content, and algorithmic amplification. However, consent finds no meaningful presence across any stage of this production pipeline.[180]

Translating consent into the nature of production and distribution of the synthetically generated media therefore demands an "ongoing agency" over how one's identity or likeness are used.[181] This would imply a right to withdraw consent and demand removal of content. When it comes to deepfakes generation, these requisites remain unfulfilled. Generating a deepfake does not require a physical presence of a person, and upon creation, it owes no clear origin or author and can be reproduced globally without any costs. Whether consent was present at any stage in the production pipeline is immaterial, unless it is found present across all stages, and its absence at any stage renders moot the presence of consent at a prior stage.

Similarly, the traditional architecture of data protection laws[182] that requires consent before processing of data for a limited purpose also fails in the infrastructural realities of deepfake generation and dissemination. Platforms take advantage of such legal and governance vacuums and create systems that commodify attention and amplify harm by their design. In this way, consent is rendered "structurally impossible to enact or be enforceable."[183] Consent, thus, collapses under platform design and regulatory gaps, and remains no longer a matter of individual choice but an infrastructural concern shaped by algorithms, corporate policies and regulatory gaps. NCSI thus demands a reconception of consent as an ongoing, revocable claim over one's sexual representation, embedded in structures of power that determine who is believed, protected and afforded dignity.



SECTION 3

Challenges to Meaningful Recourse

Dominant governance frameworks combat deepfakes through detection-oriented measures. While important, such approaches do not solve for the complex nature of harms resulting from deepfakes. As also discussed earlier, the harm with deepfakes is not situated in the fact that they may appear realistic, rather regardless of their realism, and lies in the non-consensual nature of their creation itself. Research has found that the harm does not disappear or abate because the deepfake can be recognised as fake or appears less realistic.[184] Our preceding analysis has also indicated that meaningful recourse for NCSI cannot be reduced to the formal availability of a criminal provision, a platform reporting function or a content takedown mechanism. Whether a survivor can identify abuse, safely report it, secure prompt removal, obtain accountability and recover from its consequences is shaped by the interaction of institutional capacity, platform design, commercial incentives and entrenched socio-cultural hierarchies. This section examines how these conditions systematically shift the labour of documenting, reporting and repairing NCSI onto victim-survivors, while limiting the effectiveness and accessibility of the legal, technical and social systems ostensibly designed to protect all individuals.

I. Virtue Signalling, Socio-Cultural Norms and NCSI - *"लोग क्या कहेंगे?"/"what will people say?"*

The wider persistence of TfGBV, and the specific iterations of NCSI, persist not only because of legal gaps or platform failures, but also because the basis of such violence is embedded in the socio-cultural ecosystems which normalise the weaponisation of women's bodies and punish those who speak out against it.[185] As examined earlier, in India and other Global South contexts, gender norms that tie women's value to sexual purity and family 'honour' create a fertile ground for deepfake abuse. Thus, deepfakes serve to operate as tools for social control, defining what spaces women can legitimately occupy, and determining what is considered societally acceptable. In times when threats transcend the physical into the digital and the real into the synthetic, the boundaries of what is considered 'acceptable' continues to shrink.[186]

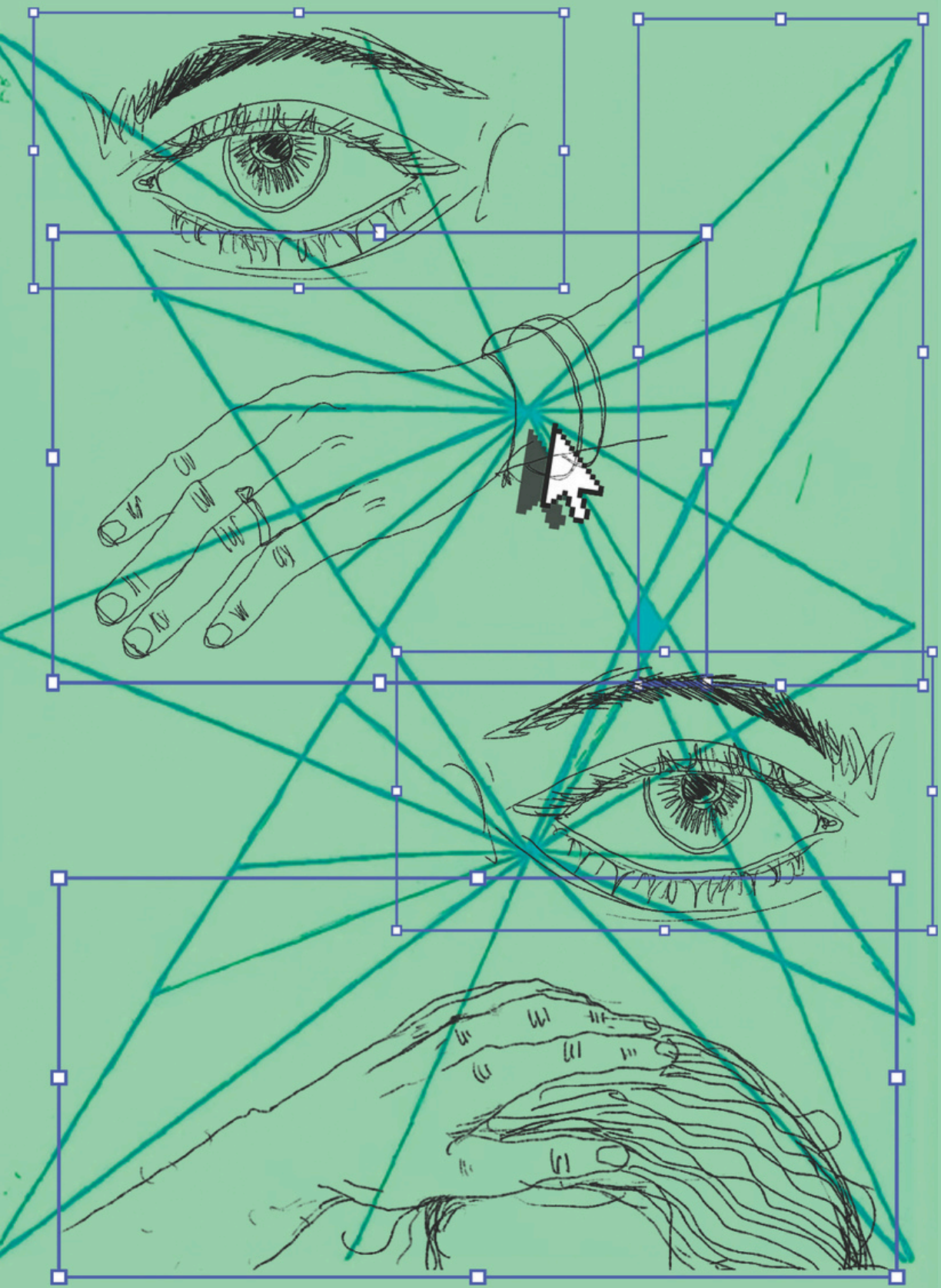
From a critical legal studies perspective, this is not merely a cultural problem but also a jurisprudential one. When the law presumes a neutral, rights-bearing subject, it obscures how caste, gender, class, and religion, or other socio-political factors, structure who can safely claim rights in practice.[187] This principle can be read in Indian jurisprudence, through

cases that unsettled moral majoritarianism, and furthered constitutional rights and liberties, but their transformative promise remains uneven when applied to NCSI harms.[188] As has been argued by scholars previously, rights formalism without attention to social power reproduces hierarchies even as it promises protection; and sexual abuse, even in the digital age, cannot be isolated from this thinking. [189]

Caste hierarchies intensify these harms for Dalit, Adivasi and similarly marginalised women. Historical patterns of caste-based sexual violence, where dominant-caste men employ sexual abuse to assert power and reinforce social boundaries, are reproduced in the digital sphere.[190] Digital sexual violence (which arguably now extends to NCSI) targeting Dalit women can be layered with casteist slurs, fetishisation and narratives of "pollution", making the abuse not only sexual but also a reaffirmation of caste supremacy.[191] Dalit survivors face a triple violence - with discrimination based on caste, gender, and economic status. This drives their lived experiences of sexual violence through patriarchal discourses on protection, honour and shame, making them less likely to seek redress due to anticipated institutional bias and community backlash.[192] The result is a regulatory blind spot: even where (and if) laws exist, caste-based stigma and impunity mean that deepfake cases involving Dalit women are under-reported, under-investigated and rarely prosecuted, allowing the cycle of abuse to continue.[193]

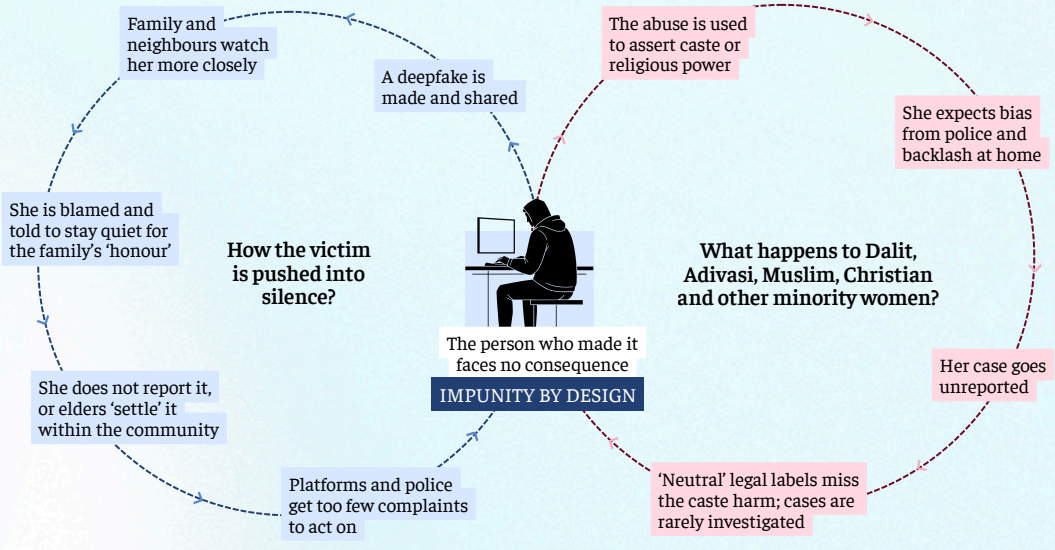
Doctrinally, this maps onto what critical caste research describes as "impunity by design".[194] Legal framing like 'voyeurism' and 'intimidation' may appear neutral but operationally undermine caste-inflected harms at the stages of complaint registration and investigations.[195] This is because caste-specific NCSI injury can lie not only in the image itself but also in casteist humiliation, sexualised degradation, coordinated harassment, reputational attacks, or threats that draw on the survivor's caste position and limited institutional protection. Discourse that sexual violence against Dalit women is a mechanism of caste domination has not translated into a parallel doctrinal shift for digital sexual harms, leaving NCSI cases to be treated as individual offences rather than structural violence.[196] Consequently, even robust statutory tools function as paper tigers when the survivor's caste position forecloses access to the very authoritative institutions tasked with enforcement.[197] The positions of authoritative institutions will be discussed further next, and deeper in the next section.

Community policing of women, especially in tier-two towns and rural areas, further entrenches the persistence of TfGBV.[198] Families, neighbourhoods and online mobs act as moral guardians, scrutinising women's clothing, friendships and online presence. When a deepfake is circulated, this policing can intensify, with survivors blamed for 'inviting' abuse and being pressured to stay silent and protect their family's 'honour'. Research in rural areas shows how community councils mobilise 'honour' to regulate women's movement, work and marital decisions and explains how community intervention can divert women away from formal legal



remedies.[199] Reporting on TfGBV in India documents how women are abused, shamed and silenced online, with many adolescent and young adult women withdrawing from education, work or public life to avoid further stigma.[200] This illustrates how socio-cultural pressures directly undermine the effectiveness of legal and platform-based technical interventions. Religious and communal stigma compound these effects for Muslim, Christian and other minority women, for whom NCSI is weaponised with communal tropes (labels such as 'characterless minority woman').[201] These harms are relational and reputational, and are more complex than the privacy and autonomy harms often discussed for NCSI. This in turn amplifies shame and safety fears, and discourages engagement with formal justice mechanisms due to a distrust in local police authorities or pressure from religious leaders to resolve matters 'within the community'.[202] This produces a jurisdictional gap, not of law, but of legitimacy: survivors are channelled into informal, community-managed resolutions that prioritise reputation over accountability, effectively immunising perpetrators from state sanction.[203] This suggests that, in India's legal regime of intermediary liability, when takedown and redress depend on user reporting, the same honour-shame calculus that silences survivors could work to starve the system of the complaints it needs to function.[204]

Figure 6: **Impunity by Design**



Together, these norms create a culture of imposed silence where survivors internalise shame, fear re-victimisation in courtrooms or in the media, and tolerate economic and social losses rather than risking further exposure.[205] This silence is not incidental, but structural: it is produced by patriarchal discourses that prioritize male honour and female modesty. It ensures that regulatory systems, no matter how well-designed, remain underutilised because they depend on reporting, evidence and witness testimony, requirements that are actively suppressed by socio-cultural pressures.[206]

The persistence of NCSI, specifically in the Indian context, cannot be resolved through technical countermeasures or the formal enactment of new offences alone. It calls for a cultural reckoning that interrogates how gendered respectability politics, caste-structured hierarchies, and community and religion-enforced moral boundaries normalise and enable such abuse.[207] The result, as discussed above, is the constrained reach and legitimacy of regulatory interventions.

II. Gender Justice and Meaningful Redressal

Studies reporting survivor narratives of experiences of TfGBV raise the question - what does justice mean to a victim-survivor? Understanding the meaning of justice by centering the victim-survivor experience is key to understanding the gaps in approaches to responding to the issue and supporting and empowering the victim-survivors. For instance, victim-survivors may seek consequences for the perpetrator and find themselves let down to know that there were none.[208] This itself can further add to the trauma of what they experienced on finding their deepfakes on the internet. Similarly, traditional court-centered notion of justice might fail the victim-survivor for the delay in its delivery, and result for them in a reliving of the trauma.[209]

Unlike harms caused in the physical world, such as through theft or accident, harms resulting from NCSI are not similarly quantifiable or visible. Hence, measures not accounting for women's experience of the harms are inadequate, ineffective and fail to provide a reliable recourse to the victims to repair the harms caused.

A study on women's perceptions of harm and favoured responses from 4000 participants across 14 regions[210] found that perceptions of harm differ by gender, where women perceive greater harm than men for the same scenario.[211] Designed across three types of harms: physical, sexual and psychological, the study found differences in preferred platform responses to the harm caused between men and women. Most preferred responses across participants were removal of content, banning the perpetrator account, and labelling content, while the least preferred across all scenarios was payment as compensation for the damage caused. Across all scenarios, amongst the different responses, women preferred banning and apology more than men. At the same time, the authors also noted that many survivors do not express confidence in the platform's judgement of what content deserves removal.

Another study from India on experiences of TfGBV and legal responses, found that women are systematically targeted, and their ability to exercise their rights against digital violence was significantly curtailed due to the social structures around them.[212] The study engaged with survivors who reported experiencing mental fatigue, public shaming, career disruption and shared a sense that justice systems were not designed to protect them. They found that some survivors opt not to take a legal recourse due to the

emotional toll, time constraints, financial burden and the inefficacy of the justice system in India. Further, for survivors who belong to marginalised communities, such as oppressed caste communities, the violence is entrenched in not only their gender but also social identity. This shrinks the available recourse and adds to pre-existing marginalisation and intensifies the harm experienced. The study also found gaps in the current effectiveness of current legal frameworks in India in responding to digital violence, "where experiences do not neatly fit into legal definitions or institutional categories."[213]

The study narrates first hand accounts from victim-survivor experiences on the effectiveness of available recourse. For instance, a well-known journalist was able to seek legal support only with the backing of her employer. Another survivor, a student belonging to a scheduled tribe and the LGBTQIA+ community, chose not to pursue a legal remedy for fear of stigmatisation and due lack of resources at their disposal. This demonstrates the structural and systemic constraints that define survivor capability and choices in pursuing available remedies. The study also found the current legal frameworks to be ill-equipped to respond to digital violence as the experiences of digital violence fail to be captured by the pre-existing legal definitions and the institutional categories.

As the preceding sections show, India's legal response to digital violence is fragments across offences such as stalking, obscenity, defamation, sexual harassment.[214] The system does not currently recognise digitally inflicted sexualised harm as a coherent category in its own right. This forces victim-survivors' experiences into pre-existing offence-based and institution specific categories, leaving forms of harm that are cumulative, gendered, anonymous, platform-mediated or sustained over time difficult to navigate, investigate and redress.

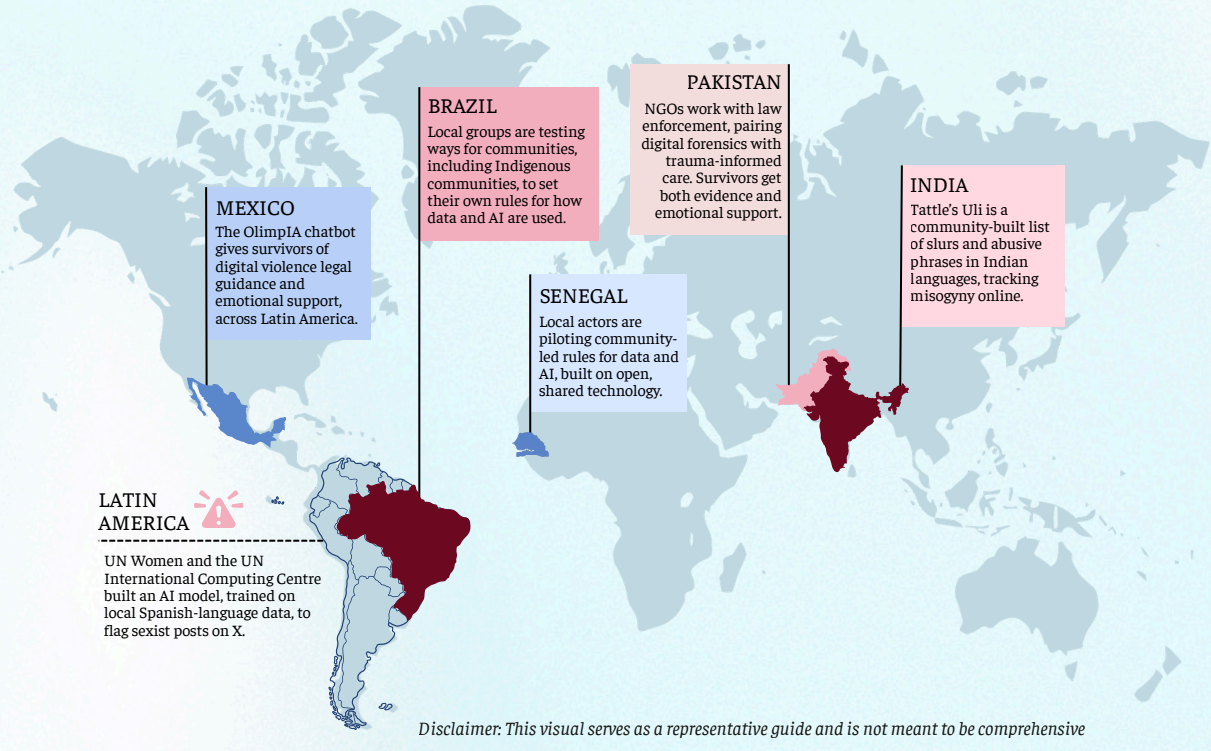
This also calls for research on the absence of survivor-centric frameworks, responses and recourses. Survivors find themselves let down by the legal, justice and platform systems and are unable to pursue any remedy for the harms they experience, without significant financial and social backing.

Scholars have explored and highlighted different models of justice, such as, shame-based, punitive, reparative, and restorative.[215] Others have pointed to user-based and platform-based responses that can help disincentivise and detect toxic and harmful conduct on platforms.[216] For survivors, justice may not mean approaching the judiciary or getting a legal outcome. Survivor-centric justice would mean understanding what survivors understand by justice in their circumstances. This implies widening our understanding and view of justice to move beyond traditional justice infrastructures which primarily involve institutions of police, law and the judiciary. Often, survivors feel let down or express lack of faith in these institutions and question their effectiveness for the magnitude of the problem that demands a sensitive and speedy response.[217] Survivor-centric justice can take many forms, and should start by understanding and

addressing the prevailing institutional gaps that fail to support the victim-survivors, introduce legal reforms, and demand accountability from tech platforms.[218] Governments should support civil society and grassroots actors to build; community-centric practices; and education and literacy programs for TfGBV; gather data on the instances and the nature of violence, the links between online and offline violence along lines of gender, caste, community, region, and religion.[219] Examples of ongoing efforts from other regions offer guidance, as discussed below.

Even as countries in the Global South struggle with judicial inefficiencies and technical capacity, alongside transnational regulatory gaps, some of them have managed to come up with innovative local practices to bridge the systemic governance gaps.[220] In countries like Brazil, Senegal, Pakistan, and Mexico local actors have collaborated to offer alternative paths to the prevailing institutions and to fill some of these gaps.[221] These include community-based and indigenous data governance models,[222] collaboration between federal agency and local NGOs that combine digital forensics and trauma-informed care to offer emotional support and evidentiary resources to survivors,[223] and the OlimpIA chatbot to provide support and guidance to survivors of TfGBV.[224] In India, Tattle's Uli database serves as an important initiative for creating safer online spaces.[225] Uli is a dataset of slurs and abusive phrases in Indian languages created through community contributions, and serves as a database on online misogyny in India.[226]

Figure 7: **Local actors offering alternative paths to bridge the systemic governance gaps**



Disclaimer: This visual serves as a representative guide and is not meant to be comprehensive

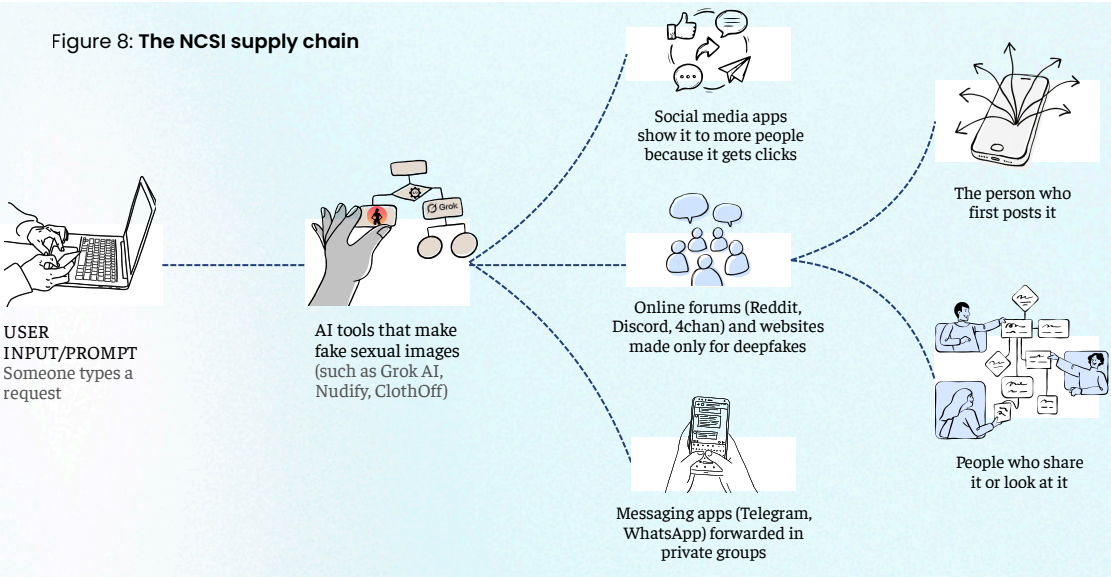
Databases like Uli support platform safety measures like content moderation, by helping filter out offensive words, phrases and slurs in different languages borrowed from offline into the online, and serve to archive problematic content. In another instance, UN Women and UN International Computing Centre developed a multilingual AI model trained on Spanish, culturally specific datasets to detect sexist content on X for addressing online sexism in Latin American countries.[227] Community led and design-justice oriented responses rooted in shared responsibility can be key in managing the problem of NCSI and responding to the needs of and empowering the victim-survivors.

III. Structural and Design Conditions that Enable Harm

The fact that the architecture of a physical space influences who can use it is well known. When the physical design of public spaces seeks to keep out vulnerable populations without explicitly banning their use through law, it is known as "hostile architecture".[228] Anti-homelessness interventions in public spaces like uncomfortable, slanted resting slabs in place of public benches, or metal bars that divide public benches that do exist are common examples of this type of design choices meant to explicitly keep people out.

This phenomenon of design influencing the experience of a space also translates into the digital world through decisions made by platforms and the infrastructure that surrounds them. There is a deep well of dark pattern literature that focuses on how interface decisions can manipulate users into making choices they may otherwise not have made were they presented with other options.[229] Common examples include flashing screens and banners that hide data protection warnings, and newsletters that don't allow users to unsubscribe. Digitally mediated work is another area where the design of platforms influences working conditions for workers in a variety of industries like ridesharing, delivery, domestic work, and other platform based gig work.[230]

Figure 8: **The NCSI supply chain**



We extend this lens of platform design and its affordances towards NCSI. We explore how the architecture of online spaces creates the conditions in which deepfake content is generated and rapidly spread. Prior work, including recent research by CCG, shows that there is an entire supply chain in the creation and dissemination of NCSI where multiple actors may profit, but the burden of response falls almost entirely on the victim. [231]

i. Low or No Cost Tools

As we've previously discussed, AI tools now make it incredibly easy to create this type of content with remarkably little pushback, if any. Tools like GrokAI, ClothOff, Nudify[232] and others provide an easy, no code way for perpetrators to generate non-consensual images of their targets and share them at scale.[233]

Further, training data for both open and closed weight AI models includes a lot of intimate imagery and CSAM, making it easier to generate non-consensual images and videos using existing content that may have been shared by social media users.[234] Researchers have repeatedly reported that Common Crawl[235] and ImageNet[236] – largest corpus of web and image data respectively which is used to train AI models – contains sexually explicit and hate content[237], which directly influences what content is generated from these tools.[238]

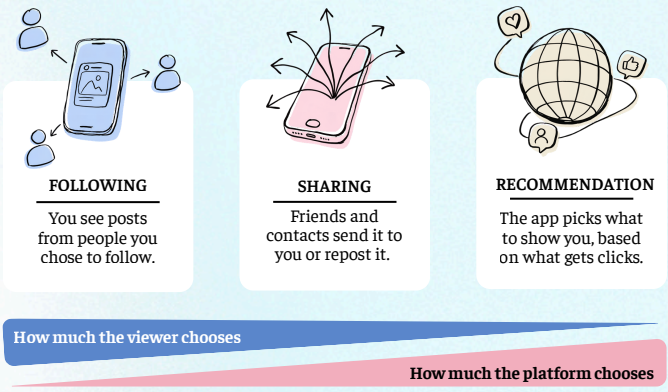


ii. Distribution Algorithms

At its simplest articulation, there are three main ways in which content reaches an audience on social media– direct subscription, sharing through a network, and recommendation algorithms.[239] In all of these, the content viewer has varying degrees of control over how an individual piece of content reaches them. In the direct subscription model, the user needs to actively be "following" or "subscribed to" the creator in question to see what they produce. In the network model, all the users "friends" and "connections" share content with them directly - either through direct messaging or just re-posting to their own feeds. While the user did not actively choose to see this content, they did sign up to get updates that are curated by the people they follow and thus have some control over the information they see. In the recommendation model, the platform almost exclusively defines what they see in their feeds, with the user having very little control over their experience, and is completely at the mercy of platform incentives that prioritise engagement and profit metrics.

Recommendation algorithms are now common across social media and play a key role in how creators and content finds new audiences. They also do not usually discriminate between the type and form of content, only on its potential for engagement.[240] Thus, recommendation algorithms will often amplify NCSI before it is detected as such, for example, in the highly publicised GrokAI case the X algorithm distributed this content at scale.[241] NCSI cases cited in the previous sections got millions of views often due to their being recommended to users who may or may not have gone to actively search for it.

Figure 9: **As reach grows, the choice moves from the viewer to the platform, and the viewer's control shrinks at each step**



iii. Internet and Payments Infrastructure

The infrastructure of the internet does not often meaningfully differentiate between NCSI and other types of content – payment processors will still allow transactions for AI-NCSI, search engines will search and list this content and recommendation algorithms will recommend it.[242] Examples cited earlier show how perpetrators of abuse will easily create specific

images, videos and even entire influencer accounts based on the likeness of their victims, and earn money from their virality. None of this would be possible without app stores hosting these 'nudify' apps, and without payment processors approving transactions of NCSI content. There is an entire supply chain that feeds off of the creation and distribution of sexualised deepfakes – from creation till consumption.[243] An Indian creator previously discussed is reported to have made Rs.10 lakh before he was arrested by the police;[244] other investigations from around the world suggest that nudify apps may have a budget that is in the millions of dollars.[245] Thus as long as the monetary incentives for creating NCSI content exist, people are encouraged to create this type of content, regardless of the regulatory interventions that may be put in place.

iv. Inconsistent Content Moderation

Content moderation – the process of evaluating and actioning unacceptable user-generated content based on platform policies and local regulations – is inherently subjective and difficult to do at scale.[246] Content policies for platforms are often not comprehensive for all types of harms and harassment concerns, especially in a feminist context. While certain platforms may have policies for sexually explicit content overall, they rarely do so for NCSI specifically. Policies are also implemented inconsistently and haphazardly when they do exist. For example, a case referred to the Meta Oversight Board shows that when NCSI were generated for female public figures in India and the US, the automated takedown system caught the case of the US public figure, but the image of the Indian public figure remained online despite multiple appeals to the grievance redressal. system.[247] In this case, even when Meta recognised that the images were explicitly against their "Bullying and Harassment" policies, users had no way to remove the image before the case was selected for the Board's review.

Prior work from the CCG has noted that content takedown is flawed as a mechanism for preventing the spread of NCSI and its harms.[248] But even when takedown systems are in place, they do not catch 100% of the material the companies prohibit. And in that gap, real harms persist. Further, taking down a single piece of content is slow and ineffective in addressing the networked nature of harassment that the victim is subject to.

Content takedown is not a complete solution either. Prior work from CCG has shown that cross border enforcement is incredibly difficult.[249] Content can be copied and re-shared once it is shared on a platform. Due to the global nature of the internet, content takedown in one jurisdiction means that it can be reuploaded in another with less strict legal and social incentives for takedown, and no clear cross-jurisdiction enforcement mechanisms.

v. Inaccessible Grievance Redressal

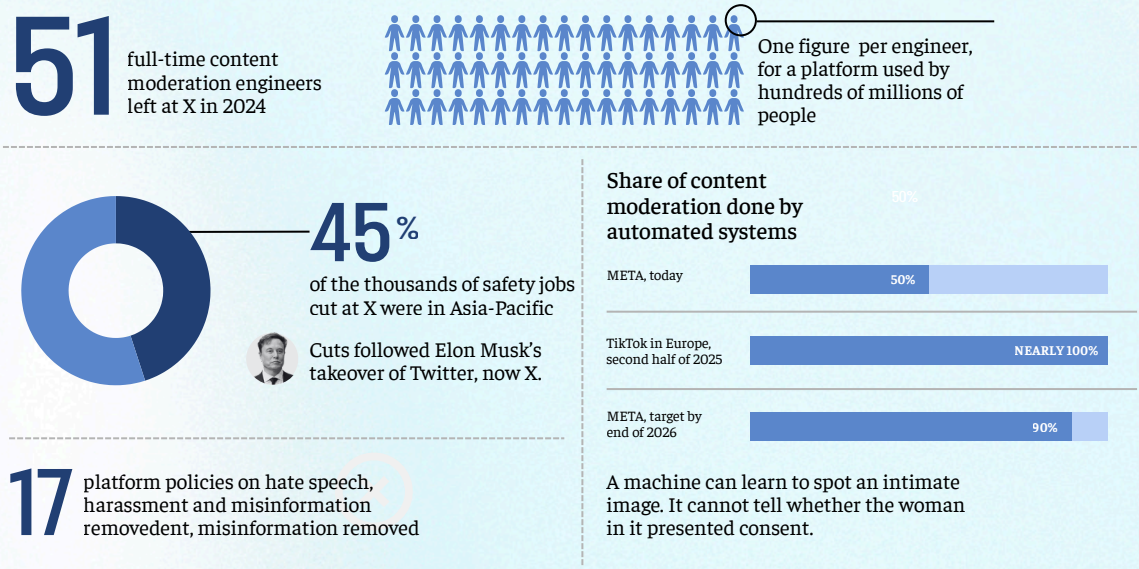
Grievance redressal systems for social media platforms tend to be hard to use. The reporting features are often hidden behind a series of incomprehensible user interfaces: screens that are hard to find and harder to understand for even sophisticated users of a platform.[250] Even when users are able to file complaints through the forms, these complaints may not get adequately actioned, like in a case referred to the Meta oversight board where a user reported a piece of NCSI content about an Indian public figure but it was automatically closed within 48 hours without review.[251] Further, grievance redressal mechanisms do not take into account the volume of hateful and abusive NCSI content a user can be facing in a single instant of networked, algorithmic harassment, where the only recourse seems to be to report each individual piece of content separately for review.

In India, grievance redressal gets further complicated by the varying rates of digital literacy across demographics. If an English speaking, digitally native user can have a hard time finding reporting workflows, it would not be a surprise to find that other users with different language skills and digital literacy will struggle with this process to a higher degree. Qualitative research on user experiences reflects the reality of inaccessible grievance redressal in the Global South: in a study of 299 users of Instagram, only 50% of the respondents said that they were updated about the status of their reports, while over 60% of respondents did not know that they could further raise their inadequately closed report to the Oversight Board for review.[252]

vi. Reduced Investment in Trust and Safety work

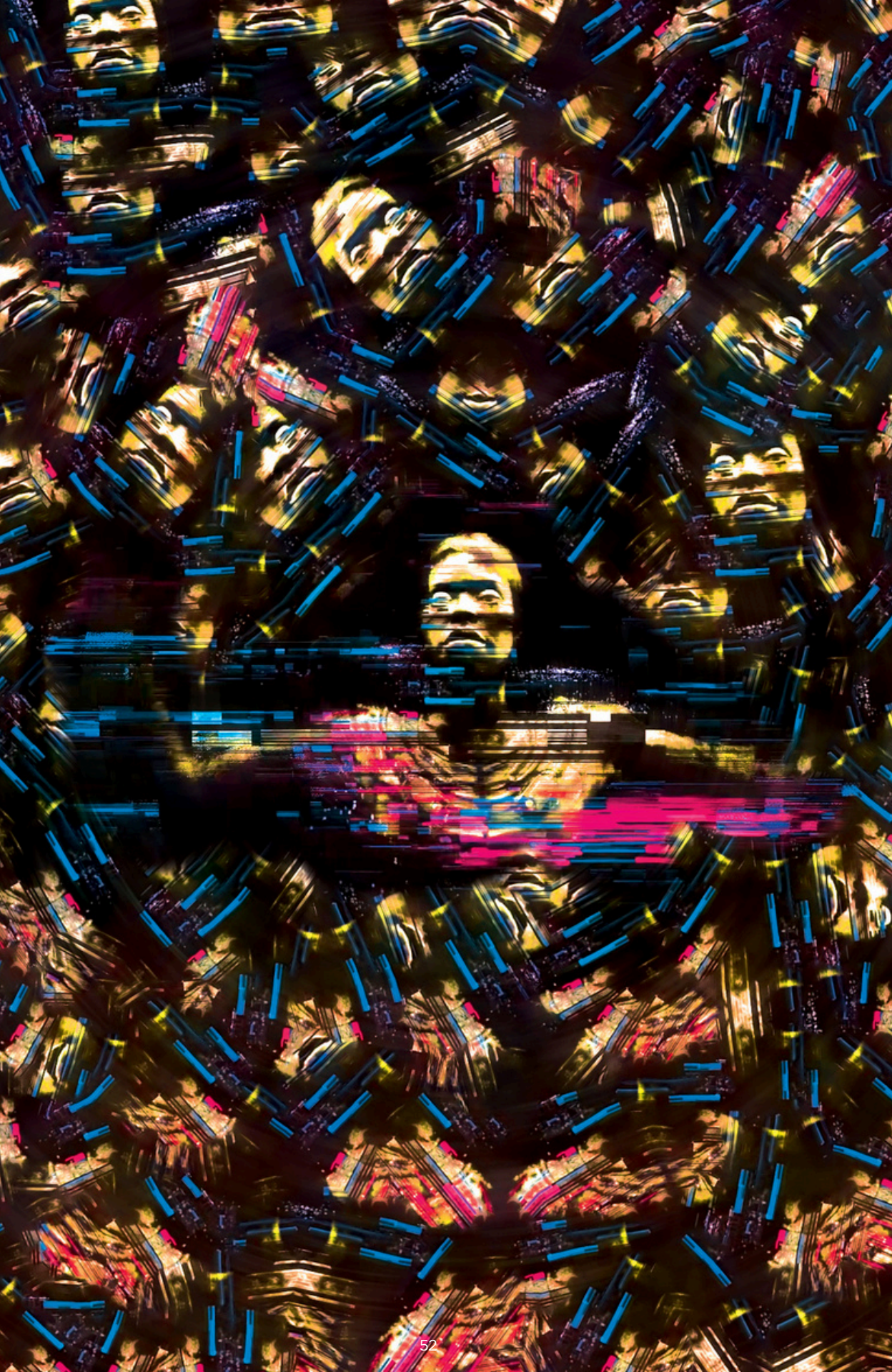
There is a trend of social media platforms cutting funding for policy, ethics or other trust and safety teams in recent years which affects how companies are able to proactively monitor content, take down harmful information and respond to user reports and requests for appeals. Elon Musk's takeover of Twitter, now X, resulted in thousands of cuts to their safety teams, with 45% of them coming from the Asia Pacific region.[253] In 2024, full time content moderation engineers fell to just 51 people for the entire platform that services hundreds of millions of users.[254] Meta, Amazon, TikTok and Alphabet have all laid off trust and safety teams around the world in favor of automated moderation.[255]

Figure 10: **Lack of resources for platform safety directly translates into how unsafe the platform gets for its users**



Disclaimer: Data obtained from EUobserver, Financial Times, Social Media today, CNBC, The guardian and the Harvard Kennedy School and analysed by the CCG team.

This lack of resources for platform safety directly translates into how unsafe the platform gets for its users. Researchers have tracked that about 17 different policies that focus on hate speech, harassment and misinformation across platforms were removed, which functions as a measure of realignment of company agendas on content moderation.[256] More recently, employees and users warned the company xAI about the fact that Grok could easily be used for generating sexually explicit material from real images, but their concerns went unaddressed,[257] which eventually resulted in the previously cited instance of users using Grok to undress women on X at scale.



SECTION 4

Pathways to Solutions

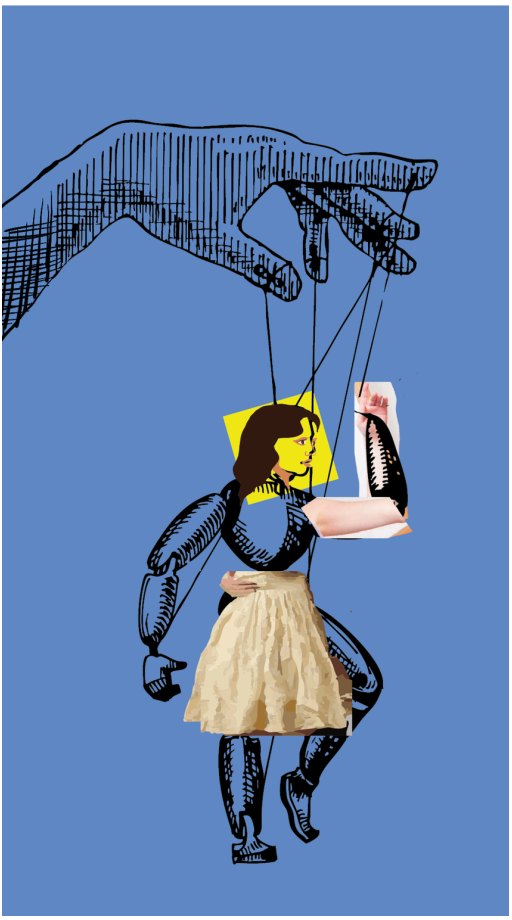
The research presented so far showcases how complicated the impact of sexualised deepfakes is in law and society. Accordingly, NCSI cannot be addressed through a single technical fix, legal prohibition or platform policy. Prevailing AI safety endeavours often lack evidence-based, gender-diverse data embedded in socio-cultural realities, and result in undermined evidence-based policy-making and a reinforcement of pre-existing systemic and structural inequalities.[258] For example, prevailing understandings of AI safety may frame harm as risk aversion rather than centering the understanding, prevention and redressal of harm within the structures of power in which they operate, and on gendered agency and justice. To conclude this report, the following sections sketch pathways to solutions that can help tackle the multifaceted obstacles presented by proliferation of NCSI.

This report is underlined by the understanding that NCSI creation and distribution are enabled by interlocking systems of technological design, commercial incentive, legal inadequacy and patriarchal social power.[259] Accordingly, an effective response must operate across each of these sites. Drawing on the above research that documents the layered harms of NCSI - from injury to dignity and psychological trauma to economic precarity and the erosion of public participation, this section outlines a combined socio-technical and socio-legal approach that contends with the scale and persistence of abuse while confronting the gendered, caste-based and communal structures that make sexualised digital violence both possible and effective.

I. Socio-Legal and Ideological

This report has demonstrated how the persistence of NCSI points to the need to interrogate the role of social structures in the South and the structural violence they enable on women's bodies. It underlines the need to build upon and then beyond the liberal legalist approaches and technocratic governance in the Global South to account for the context of Southern societies, and build pathways that account for the intersectionality, contextual complexity and social structures to ensure they meaningfully address the needs of victim-survivors. Although liberalism has furthered individual liberty and autonomy, which has been key to the feminist struggles against patriarchal structures, experience from the Global South shows that in pluralistic cultures, women's experiences and

freedoms are determined by social structures that constrain the exercise of basic individual freedoms in complex ways.[260] Hence, strengthening the rights of women and furthering empowerment requires emphasis on these social structures of the South. Detection tools, takedown regimes and consent-based data protection frameworks presume a rights-bearing subject who can meaningfully authorise, control, and withdraw permission over her digital self, but exclude from their consideration the various socio-cultural factors that constrain a meaningful exercise of rights. Yet, for women in the Global South, consent is structurally foreclosed, privacy is policed, and dignity is sacrificed to patriarchal moral orders. This underlines that responses to NCSI must reject the fiction of the neutral, autonomous individual and adopt a feminist socio-legal response that centres the structural conditions that produce NCSI as a technology of domination. This requires not only legal reform but ideological rupture: a confrontation with the power asymmetries that render women's bodies and sexuality legible as public property and a site of communal control. This can look like making available privacy-protected, anonymised and structured case studies and learnings on gender based violence for institutional actors such as civil society, lawyers, researchers and policy makers to further survivor centered advocacy, support, oversight and policy reform towards a gender just governance environment.[261]



From Consent, To Structural Accountability

Consent doctrines, whether in data protection, criminal law or platform's terms of services, operate as systems of legitimisation, rather than protection. They presume symmetry between data subject and data processor, between victim-survivor and perpetrator, between users and platform owners. Feminist legal theory has long shown that consent, under conditions of structural inequality, can serve to function as a tool to sanitise distasteful experiences of coercion.[262] The married woman is presumed to have given perpetual sexual consent; the Dalit woman's body is treated as available to dominant-case men; the public-facing woman is assumed to have waived her right to privacy.[263] NCSI reproduces these logics in digital form: posting a photograph is treated as

implied permission for reuse and alteration; visibility online is read as sexual availability; the absence of an explicit 'no' is conflated with an affirmative 'yes'. This is starkly exemplified by the fact that reporting copyright infringement results in faster action by platforms than content violations through NCSI.[264]

A critical legal response must therefore center consent as the organising principle and impose affirmative duties on platforms, intermediaries and state actors to prevent foreseeable harms. This includes statutory recognition of ongoing, revocable agency over one's likeness, strict liability for platforms that design for virality without safeguards; and doctrinal shifts that treat the non-consensual creation of sexual imagery as a violation of bodily autonomy regardless of falsity or reputation impact.[265] A doctrinal shift can be triggered with amendments to existing legislation and additional rules. Embedding specific protections into the law exemplifies tangible efforts to work towards structural accountability.

Intersectionality and the Doctrinal Method

Legal formalism obscures how caste, gender, class and religion structure who can safely claim rights in practice. Dalit, Adivasi and Muslim women face compounded harms: NCSI is weaponised with casteist slurs and communal tropes that reaffirm historical hierarchies of sexual violence.[266] Yet anti-discrimination law remains silent on digital sexual harms, and criminal procedures insufficiently account for caste-specific injuries at the stages of complaint registration, investigation and prosecution.[267] A feminist socio-legal conceptualisation can embed intersectionality into legal doctrinal design: recognising NCSI as structural violence, rather than just an individual offence; mandating investigation protocols sensitive to socio-political discriminatory factors, such as caste and religion, and creating independent oversight mechanisms to counter institutional limitations.[268] Embedding intersectionality into legal provisions designed to protect against online harms also requires expanding the vocabulary of harms beyond privacy and bodily autonomy, to include dignity-specific injury, collective trauma and the erosion of public participation rights.

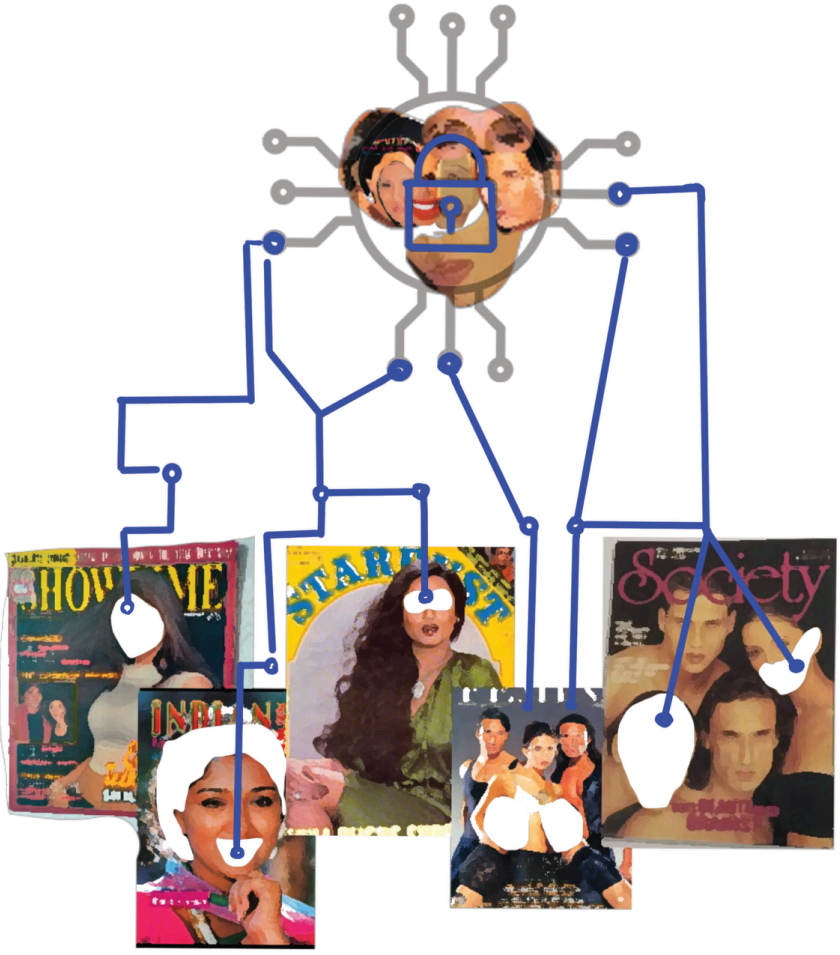
Cultural Transformation and Ideological Fracture

Legal and technical interventions will remain inadequate without confronting the socio-cultural ecosystems that normalise the weaponisation of women's bodies. Patriarchal honour economies, caste-based sexual domination and communal moral policing create a culture of imposed silence where survivors internalise shame, fear re-victimisation and tolerate economic and social losses, rather than risking further exposure.[269] Public discourse must shift from asking 'Why did she post that?' to 'Why do we treat women's bodies as public property?' Faith leaders, community elders, journalists and influencers can play pivotal roles in reframing honour as collective protection rather than individual control, and in modeling accountability for perpetrators rather than silence for survivors. Digital literacy initiatives must expand beyond technical skills to include critical curricula on consent, dignity and gendered politics of online

presence. This approach triggers public cognitive dissonance and challenges honour-shame narratives that blame survivors and framing viewing, sharing and commenting as participatory violence.

Solidarity, Care and Collective Resistance

Survivors need accessible, trauma-informed support: legal aid, mental health services, safe housing and economic assistance to mitigate professional fallout. Grassroots organisations, feminist collectives and digital rights groups can be resourced to provide rapid response, documentation and advocacy. Cross-sector coalitions can amplify survivor voices, pressure legal and policing institutions and build alternative narratives that centre dignity over same. The fight against NCSI is not only about regulating technology; it is about reimagining the social contract around gender, power and representation. A feminist socio-legal approach that centers consent and intersectionality, survivor agency and demands structural accountability can relabel NCSI from an inevitable byproduct of generative AI into a turning point for collective resistance and renewal.



II. Socio-Technical

Companies, governments and regulators rely heavily on technical methods of content moderation including to take down and limit the spread of NCSI if their policies prioritise it. Transparency documents from TikTok in Europe showed that the platform used automated methods of content moderation for nearly 100% of offending content in the second half of 2025.[270] For Meta platforms, 50% of moderation tasks are currently automated, with the company aiming to increase this to 90% by the end of 2026.[271] It is, thus, important to look at various ways in which automated moderation can be useful and what their drawbacks are.

i. Hash-Matching

Hash-matching is currently the gold standard for taking down known offensive content effectively. Non-profits and industry groups like StopNCII[272] and Tech Coalition[273] collect hashes or digital fingerprints of user and media reported instances of NCSI. Partner platforms then refer to these databases to prevent the reupload and spread of the reported NCSI content. As of 2026, Ofcom, the communications technology regulator in the UK, requires technology companies operating in the UK to implement hash-matching as a safety measure or pay fines up to 10% of their global revenue.[274] Once an image is identified its hash is added to the database, platforms use automated means to takedown instances of the image shared on their platform and prevent further reuploads.

While this is a great first step, it is insufficient to address all the harms caused by NCSI. By definition, hash-matching cannot identify new unrelated instances of NCSI, nor can it identify modified versions of the same image of the same person. Further, the burden of ensuring that the image is added to the databases that are hashed falls almost entirely on the victim. For StopNCII, the victim (or someone on her behalf) needs to create a digital fingerprint, called a hash, using the non-profit's website. If the image follows all the listed rules, a case is created which ensures that the hash generated is shared with participating platforms. These platforms then use the generated hash in their automated content moderation process to take down the offending images.[275]

This method of reporting and takedown is not trivially easy. Victims of NCSI harassment would need to ensure both that they catch all variants of the abuse against them, *and* that they are added to all the different databases maintained by many industry coalitions and NGOs. It is unreasonable to expect that the victim, in their moment of need, knows all the laws, relevant company policies, and the different collaborations that have come together to prevent the very harm that they are facing.

ii. Watermarking

Watermarking efforts like C2PA[276] and Google's SynthID[277] aim to mark the provenance of an image at generation. In this method, the AI model that creates the image adds a watermark which can then be used by a

social media platform to label the image as AI-generated when shared. This enables concerned users to identify the synthetic nature of the content they are seeing. However, not all harms of NCSI originate from its provenance.[278] A lot of them depend on their intent. As cheapfakes already show,[279] content used to demean, embarrass, discredit, or disencourage public participation would succeed in their intended manipulation regardless of how the image was manipulated. While watermarking is an important tool in the fight against NCSI, it is insufficient in itself to tackle all its socio-technical harms.

iii. App-Stores & Payment Infrastructures

Technical infrastructure like app-stores, payment processors, and search indexing, is another avenue where researchers, policymakers and activists have pressured companies to intervene to prevent the spread of NCSI.[280] While not a complete solution in itself, the infrastructure layer interventions address the supply chain inherent in the production and sharing of NCSI content and tackle the incentives associated with their creation and dissemination.[281]

App-store level bans of "nudification" apps can act as gatekeepers to the tools of the production of these images. Both Apple's and Google's app stores have hosted apps that allow users to non-consensually undress women, making this technology easily accessible from a perpetrator's phone.[282] Taking this away provides an easy level to remove access to such apps. The EU has done exactly this, by directing app-stores to ban all nudifier apps that create sexually explicit material of real people without their consent.[283] This measure goes into effect in December. De-listing search results that contain these images is a way for search engines to mitigate the harms related to continued access that a victim has to bear.

As previously noted, the monetisation infrastructure of this content incentivises its creation and spread. Case studies cited earlier in the report show that perpetrators can make real money out of the generation and sale of NCSI content, including via various subscription based services. Removing these incentives has been shown to be an effective way of reducing the volume and forcing platforms into better content moderation. For instance, CivitAI, an AI-art platform known for the proliferation of NCSI content of real people moved to using pre-paid gift cards and cryptocurrency for transactions after its credit card processor cut off access to its services over NCSI content.[284] Statistics are still pending but this measure effectively works to reduce the number of people who can pay for this content easily.

The Report recognises that the initial harm may already have taken place before measures like bans on 'nudify' apps or search engine de-listing go into effect, but infrastructure level interventions are a promising avenue to stop secondary harms of repeated exposure of the NCSI content online, and keep app-developers and creators of NCSI on notice. The main drawback here is the risk of over-implementation of bans which results in

over-removal of acceptable speech as automated filters may catch unrelated apps, payment or links on their respective infrastructure.

iv. Predictive Classifiers

Predictive classifiers can be developed by technology companies for content moderation.[285] For companies with the significant resources needed to train a predictive classifier – label offensive data, and train a machine learning model to identify it – this approach can provide a good way to create automated moderation systems that are aligned to company policies, that can thus help enforce them more effectively, and reduce the problem of inconsistent application of policies. It solves for the problem of exact matches that hash-matching is known for, and instead allows for identification of somewhat modified content as well. Known issues with predictive models include over-and-under moderation due to the inadvertent takedown of content that does not violate any company policies or leaving up of content that does.[286] A major drawback for their use in NCSI prevention however, is that while it may be possible to train a classifier to identify intimate imagery, it is almost impossible to train to identify consent. There is no way for a model to differentiate between a sexual image created by the user's permission and one without.

v. LLMs for Content Moderation

Large Language Models (LLMs) are the newest way in which content moderation is being implemented on tech platforms.[287] LLMs have shown better performance on hate speech identification than predictive classifiers on some benchmarks,[288] and newer LLM-based approaches allow trust and safety (T&S) teams to bring a policy written in plain English to the model rather than having dedicated data scientists assigned to re-create models for moderation purposes every time a policy changes.[289] In other words, a member of the T&S team can prompt the moderation LLM in natural language to remove content based on keywords like "explicit" or "suggestive" images, and the LLM can detect the content, take into account its context, and make decisions about whether a policy is violated.

This also improves the T&S team's ability to adapt existing policies to cover new types of content manipulation. When new forms of abuse get noticed on platforms, T&S teams can test if the specificity of their written policy adequately responds to it or whether they need a new policy. For example, if a platform sees a new instance of bullying starting to emerge from a viral trend, the team can update their policies to include the new related keywords. These new keywords can then be quickly implemented through the LLM-based system to improve takedown timelines.

However, these approaches, while promising, currently show better results on text-based media than on image-based media, especially when focused on deepfake detection.[290]

vi. Transparency Mechanisms

Improved transparency mechanisms from both social media platforms and

from AI model developers will help improve trust in the systems of generation and dissemination.[291] Allowing auditors and third party researchers to access data related to use of AI systems, and to the gendered harms of images used for model training more specifically will be helpful in breaking the black box of algorithmic training. More data on moderation tools will allow researchers to study impacts of various tools and their efficacy in NCSI detection and moderation.[292]

vii. Reporting Mechanisms

As previously noted, reporting and grievance redressal mechanisms are inaccessible for large swaths of the population. Improved reporting mechanisms that are intuitive and allow users to quickly point to offending material that needs to be removed from the platform is an easy win that can reduce the lasting harms of NCSI abuse. This reported material should then be actioned on by a team sensitized to deal with such material.[293] Grievance redressal mechanisms should also be able to recognise and protect against networked harassment and coordinated campaigns against the victim.[294] This can be done by allowing users to report keywords and platform trends instead of only systematically reporting single pieces of content.

Table 2: **Available socio-technical interventions**

INTERVENTION	WHAT IT GETS RIGHT	WHERE IT FALLS SHORT	WHO DOES THE WORK
WATERMARKING	It marks an image as AI-made when it is created	It does not tackle nature of abuse, only provenance	AI model developers
INTERNET INFRASTRUCTURE	It cuts off the money that keeps nudify apps running	It acts only after the first harm, and may block unrelated apps	App stores and payment processors
HASH-MATCHING	It stops known images from being uploaded again	It misses new images and edited copies	Victim must report every copy to every database
PREDICTIVE CLASSIFIERS	It can catch some edited copies and may be faster	It may take down unrelated images	Large tech platforms that can afford it
AI LANGUAGE MODELS	Safety teams can update rules quickly, in plain English	It works better on text than on images	Platform safety teams
REPORTING TOOLS	It allows users to report items automated systems have missed	It is only relevant once the victim or concerned parties find the content	The victim or concerned parties must find and report it herself.
TRANSPARENCY	It lets researchers see how moderation really works	It does not remove any content by itself	Platforms and AI developers

Conclusion

NCSI is not merely a novel misuse of generative AI systems, it is a digitally accelerated form of gendered violence that exploits existing inequalities of gender, caste, religion, class and sexuality. The report studied the weaponisation of NCSI within social norms and notions that are designed to keep women in subjugation. It pointed to how such weaponisation serves to marginalise women generally, but interacts in complex ways with their social, economic, or political identity to create differential implications for women belonging to marginalised or minority communities in countries like India.

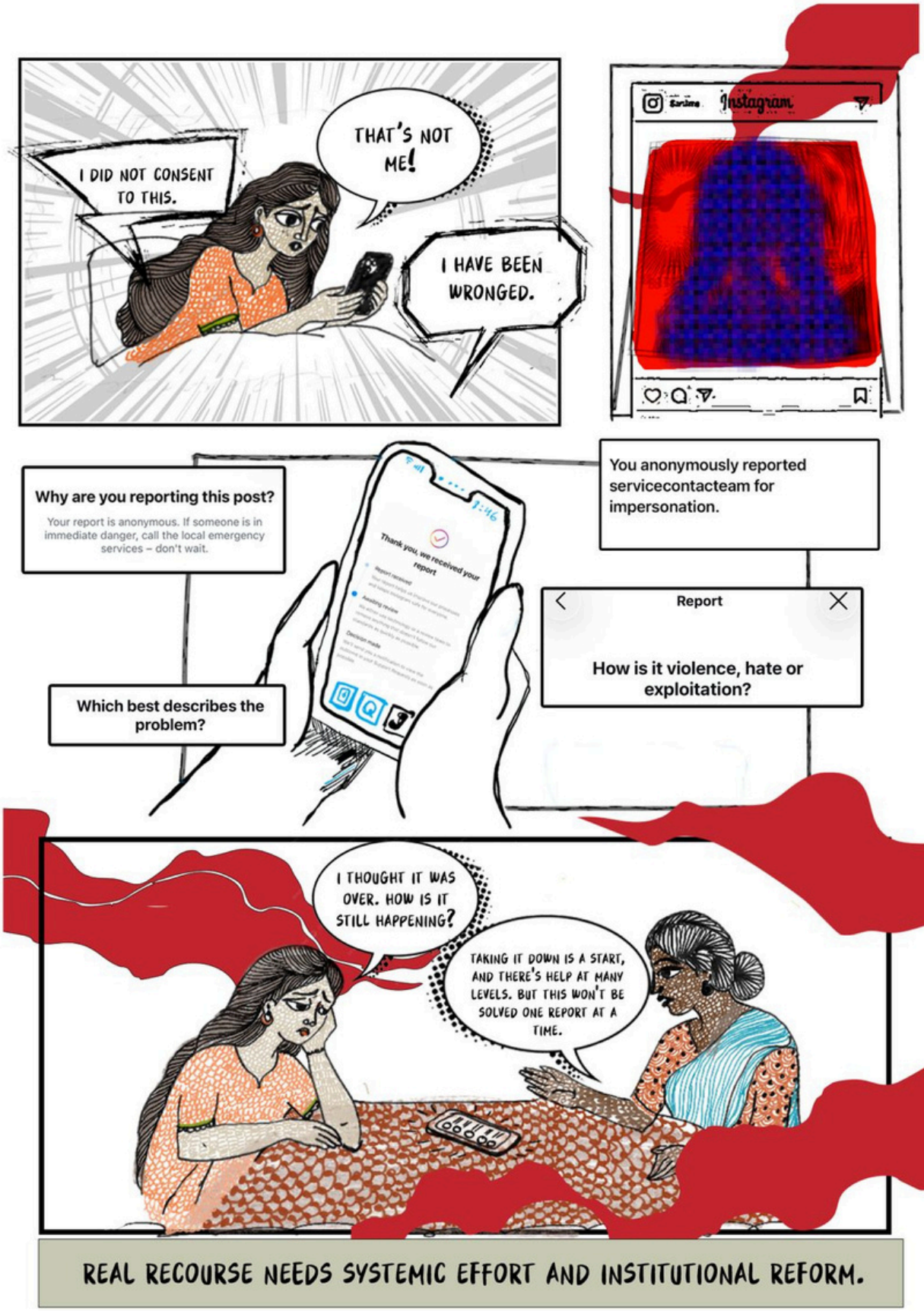
Mapping the layers of harm that surface from the case studies included in this report, it identified the different types of harms, and how they operate, intersect and overlap within the socio-cultural and institutional realities of the Indian and the Global South context. The harms analysis conducted by this report provides how women in India are abused, shamed and silenced online through NCSI abuse, and has direct implications for women's pursuit of education, career, free expression and an exercise of autonomy on account of the attached stigma. This report illustrates how the socio-cultural pressures in turn undermine the effectiveness of legal and platform-based technical interventions. For instance, Dalit women faced with NCSI are likely to face a violence that is not just gender-based but also rooted in caste-based subjugation. Combined with a potential economic constraint and a pre-existing hesitation to seek redress due to anticipated institutional bias and community backlash, the result is that deepfake cases involving Dalit women are likely to go under-reported, under-investigated and rarely prosecuted, allowing the cycle of abuse to continue.

This Report further shows that the harms extend beyond reputational damage or privacy violations. NCSI attacks the dignity, bodily autonomy, financial security, public participation and the ability of women to fearlessly exist in digital and public spaces. The Report demonstrated how socio-cultural factors combine to censor women's consensual and free expression not just online but also in their day-to-day lived experience.The effects of NCSI are especially acute for women who belong to marginalised communities, for whom sexualised synthetic imagery is entangled with honour-based control, and communal and casteist harassment. The report mapped the range of socio-legal, socio-cultural, and socio-technical challenges that obstruct meaningful redressal of harms, and provided insights for identifying pathways towards meaningful reform, recourse, and redressal.

The Report shows that existing responses are inadequate because they treat NCSI as a content-moderation or individual criminal liability problem. Detection, hash-matching, watermarking and takedown mechanisms are necessary but they are reactive, and place an unreasonable burden on victim-survivors to identify, report and pursue removal of abuse across platforms. Technical measures, such as automated filters and predictive classifiers, also risk removal of acceptable speech, or inadvertently catch material that could qualify as consensual expression. Legal consent-based frameworks are also insufficient because they tend to assume an autonomous individual who can freely provide, monitor and withdraw permission, while ignoring the social conditions that make consent, privacy and redress inaccessible to many women.

Meaningful responses must therefore move from consent alone to structural accountability. Platforms, AI developers, app stores, payment aggregators and state institutions should bear explicit affirmative duties to prevent foreseeable harm, enable rapid deletion, protect survivors from coordinated harassment, and ensure transparent, survivor-centric grievance processes. Legal reform must recognise NCSI as a violation of bodily autonomy and dignity, embed intersectional enforcement, and address the compounded vulnerabilities. Legal and technical responses should further ensure a balanced approach to protecting against NCSI while safeguarding privacy and freedom of expression, through targeted and proportionate measures that guard against a risk of general monitoring and censorship.

The report explored how the architecture of online spaces create the conditions in which deepfake content is generated and rapidly spread. Noting the enablement of NCSI by interlocking systems of technological design, commercial incentive, legal inadequacy and patriarchal social power, the report concludes that effective response must operate across each of these sites. Ultimately, addressing NCSI requires more than better technology or stricter law. It demands cultural transformation, rejecting narratives that blame women for their visibility, treating the creation and distribution of NCSI as participatory violence, and rebuilding digital governance around dignity, equality and survivor agency. By combining legal reform, platform accountability, technical safeguards and feminist socio-cultural change, NCSI can be challenged as an inevitable consequence of bad actors accessing generative AI, to instead contend with it as a preventable form of systemic gendered violence.


I DID NOT CONSENT TO THIS.
THAT'S NOT ME!
I HAVE BEEN WRONGED.
Instagram
Why are you reporting this post?
Your report is anonymous. If someone is in immediate danger, call the local emergency services – don't wait.
Thank you, we received your report
You anonymously reported servicecontacteam for impersonation.
Report
How is it violence, hate or exploitation?
Which best describes the problem?
I THOUGHT IT WAS OVER. HOW IS IT STILL HAPPENING?
TAKING IT DOWN IS A START, AND THERE'S HELP AT MANY LEVELS. BUT THIS WON'T BE SOLVED ONE REPORT AT A TIME.
REAL RECOURSE NEEDS SYSTEMIC EFFORT AND INSTITUTIONAL REFORM.

Endnotes

[1] Anubhav Mishra and Bibhudatta Dash, “Deepfakes and the crisis of digital consent: Artificial intelligence and synthetic media violations against women in Asia” (2026) <https://doi.org/10.1177/01634437261465231> accessed 31 October 2026.

[2] Sarah Morgan, Hany Farid, and Sophie J. Nightingale, “The Backbone of Abuse: How Infrastructure Providers Enable the Proliferation of AI-Generated Non-Consensual Intimate Imagery” (2026) 3(4) Journal of Online Trust and Safety <https://tsjournal.org/index.php/jots/article/view/348> accessed 31 October 2026; Rishiti Choudaha, Torsha Sarkar, Sukriti, ‘Pixelated Images, Real Harms: Insights from the Regulation of Non-Consensual Synthetic Imagery across UK, India and Brazil’ (Centre for Communication Governance, National Law University Delhi 2026) (forthcoming).

[3] Morgan, Farid and Nightingale, (n 2).

[4] Ibid.

[5] Ibid.

[6] Ibid.; Adrija Bose, Tarunima Prabhakar and Tora Agarwala, “Desi Faces, Foreign Servers: Inside the AI Economy Undressing Women” (Decode, 3 June 2026) <https://www.decodeinternet.in/life/investigation/desi-faces-foreign-servers-inside-the-ai-economy-undressing-women-31691> accessed 23 September 2026; Michelle L Ding, Harini Suresh, Suresh Venkatasubramanian, “How to stop playing whack-a-mole: Mapping the Ecosystem of Technologies Facilitating AI-Generated Non-Consensual Intimate Images” (arxiv, 2026) <https://arxiv.org/abs/2602.04759> accessed 23 September 2026.

[7] Center for Countering Digital Hate, 'Grok Floods X with Sexualized Images of Women and Children' (CCDH, 22 January 2026) <https://counterhate.com/research/grok-floods-x-with-sexualized-images/> accessed 10 September 2026.

[8] Reuters Staff, ‘EU orders X to keep Grok documents for longer amid sexualised AI photos furore’, (Reuters, 8 January 2026) <https://www.reuters.com/world/eu-commission-has-ordered-x-retain-all-grok-documents-until-end-2026-2026-01-08/> accessed 25 September 2026.

[9] Drew Harwell, 'Scarlett Johansson on Fake AI-Generated Sex Videos: 'Nothing Can Stop Someone from Cutting and Pasting My Image'' (The Washington Post, 31 December 2018) <https://www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/> accessed 10 September 2026.

[10] Emine Saner, 'Inside the Taylor Swift Deepfake Scandal: 'It's Men Telling a Powerful Woman to Get Back in Her Box'' (The Guardian, 31 January 2024) <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box> accessed 10 September 2026.

[11] David Wilson, 'NI Politician Targeted by Deepfake Quits X over Grok AI Concerns' (BBC News, 12 January 2026) <https://www.bbc.com/news/articles/cgezg5rlz7wo> accessed 10 September 2026.

[12] Susan Dominus, 'Someone Was Impersonating Her in Online Affairs. She Says It Was Her Husband.' (The New York Times, 23 August 2026) <https://www.nytimes.com/2026/08/23/magazine/collien-fernandes-impersonators-germany-metoo.html> accessed 10 September 2026.

[13] Nina Jankowicz, 'How Disinformation Became a New Threat to Women' (Coda Story, 11 December 2017) <https://www.codastory.com/polarization/how-disinformation-became-a-new-threat-to-women/> accessed 10 September 2026.

[14] Matt Burgess, 'Sexually Explicit Deepfake Sites Target 100-Plus Politicians in Europe' (Wired, 14 September 2026) <https://www.wired.com/story/sexually-explicit-deepfake-sites-target-100-plus-politicians-in-europe/> accessed 25 September 2026.

[15] Rishabh Suri, 'Shalini Pandey Reacts to Viral Video of Her on Social Media, Calls It "AI Generated": 'I Condemn This Kind of Abuse'' (Hindustan Times, 31 August 2026) <https://www.hindustantimes.com/htcity/cinema/shalini-pandey-reacts-to-viral-video-of-her-on-social-media-calls-it-ai-generated-i-condemn-this-arjun-reddy-101788169763444.html> accessed 10 September 2026.

[16] Noor Nanji and Shruti Menon, 'Why Are There So Many Deepfakes of Bollywood Actresses?' (BBC News, 22 December 2023) <https://www.bbc.com/news/entertainment-arts-67726019> accessed 10 September 2026.

[17] See generally, Yongjie Gao, 'Sexualized Deepfake Abuse as Image Based Sexual Abuse: A Scoping Review of Platform-Mediated Circulation and Cumulative Harm' Trauma, Violence and Abuse (2026) <https://doi.org/10.1177/15248380261480163>.

[18] Weijie Huang, Payal Arora, Marta Zarzycka, "Geographies of Hope: Rethinking Deepfake Harms and Gender AI Safety in the Global South" (2026) 14 Media and Communication Article 10969.

[19] Lisa Lazard et al, "Deepfake Technology and Gender- Based Violence: A Scoping Review" (2025) 0 Trauma Violence & Abuse 1.

[20] Ibid.

[21] Other examples of online gender based violence include doxing, cyberstalking, sextortion and grooming. See generally, UN Women, 'FAQs: Digital abuse, trolling, stalking, and other forms of technology-facilitated violence against women and girls' (UN Women, 13 November 2025) <https://www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women>.

[22] Lazard et al (n 19).

[23] On NCSI as a supply chain, read more here: Choudaha, Sarkar, Sukriti (n 2).

[24] Ibid.

[25] See Jane Im et al, "Women's Perspectives on Harm and Justice after Online Harassment" (2022) 6:CSCW2 Proceedings of the ACM on Human-Computer Interaction 1.

[26] Abigail Bentley, 'Not Just Nudes: Defining Image-Based Abuse in Pakistan and the Diaspora' (Chayn, 2026) <https://reports.chayn.co/image-based-abuse-in-pakistan> accessed 28 October 2026.

[27] Ibid.

[28] Frances Mao, Kelly Ng and Muhammad Zubair, 'Pakistan: Woman Killed After Being Seen with Man in Viral Photo' (BBC News, 28 November 2023) <https://www.bbc.com/news/world-asia-67551554> accessed 10 September 2026.

[29] Press Trust of India, '"Fabricated Video, Lookalike": Woman Techie's Note Before Suicide' (NDTV, 3 April 2025) <https://www.ndtv.com/india-news/woman-techie-kills-self-at-nagpur-home-leaves-behind-note-alleging-conspiracy-threats-to-life-8081003> accessed 10 September 2026.

[30] Emanuel Maiberg, 'Subtlefakes: Slightly Altered Nonconsensual AI Images Are Taking Over X' (404 Media, 20 August 2026) <https://www.404media.co/subtlefakes-slightly-altered-nonconsensual-ai-images-are-taking-over-x/> accessed 10 September 2026.

[31] Nabiya Khan, Aishik Saha and Zenith Khan, 'AI-Generated Imagery and the New Frontier of Islamophobia in India' (Center for the Study of Organized Hate, 29 September 2025) <https://www.csohate.org/2025/09/29/ai-generated-hate-in-india/> accessed 10 September 2026.

[32] Jyoti Thakur, '"Looked So Real": How AI Is Being Weaponised Against India's Muslim Women' (Al Jazeera, 15 June 2026) <https://www.aljazeera.com/features/2026/6/15/looked-so-real-how-ai-is-being-weaponised-against-indias-muslim-women> accessed 10 September 2026.

[33] Geeta Pandey, 'Sulli Deals: The Indian Muslim Women "Up for Sale" on an App' (BBC News, 9 July 2021) <https://www.bbc.com/news/world-asia-india-57764271> accessed 10 September 2026.

[34] The Wire Staff, '"Act of Intimidation and Harm": Rights Activists on "Sulli Deals" App Targeting Muslim Women' (The Wire, 13 July 2021) <https://thewire.in/women/sulli-deals-muslim-women-cyber-harassment-statement> accessed 10 September 2026.

[35] Anoushka Dalmia and Preeksha Malhotra, 'AI Tools & Social Norms are Making the Internet Unsafer For Women' (Project Multatuli, 6 April 2026) <https://projectmultatuli.org/en/ai-tools-social-norms-making-internet-unsafer-for-women/> accessed 28 October 2026.

[36] Kiruba Munusamy, 'Intersection of Identities: Online Gender and Caste Based Violence' (GenderIT.org, 7 June 2018) <https://www.genderit.org/articles/intersection-identities-online-gender-and-caste-based-violence> accessed 24 September 2026; Anita Gurumurthy, Amrita Vasudevan and Nandini Chami, Born Digital, Born Free? A Socio-Legal Study on Young Women's Experiences of Online Violence in South India (IT for Change 2019) <https://itforchange.net/sites/default/files/1618/Born-Digital_Born-Free_SynthesisReport.pdf> accessed 24 September 2026; Anita Gurumurthy and Amshuman Dasarathy, Profitable Provocations: A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-Political Life (IT for Change, July 2022) <https://itforchange.net/sites/default/files/2132/ITfC-Twitter-Report-Profitable-Provocations.pdf> accessed 24 September 2026; Manjusha Madhu and Richa Singh, Experiencing Technology-Facilitated Gender-Based Violence in India: Survivor Narratives and Legal Responses (Equality Now and Breakthrough, November 2025) <https://equalitynow.org/wp-content/uploads/2025/10/Experiencing-technology-facilitated-gender-based-violence-in-India-English-1.pdf> accessed 24 September 2026.

[37] Ibid.

[38] Scroll Staff, '95 Indian Women Politicians Received One Million Abusive Tweets in Three Months: Amnesty Study' (Scroll.in, 23 January 2020) <https://scroll.in/latest/950836/nearly-14-tweets-sent-to-women-politicians-during-2019-polls-were-hurtful-or-abusive-amnesty-study> accessed 28 October 2026.

[39] Paola Bacchetta, "When the (Hindu) Nation Exiles Its Queers" (1999) 61 Social Text 141; Tanvi Kanchan, '"Instagram Is like a Karela": Transnational Digital Queer Politics and Online Censorship and Surveillance in India' (2024) 17(3) Communication, Culture and Critique 162.

[40] Gurumurthy, Vasudevan and Chami (n 36); Ashmita Biswas, 'Exploring Queered Digital Spaces in India: Prospects and Precarities of Looking for Love Online' (2023) 7(4) Erothanatos 1 <https://www.academia.edu/108615976/Exploring_Queered_Digital_Spaces_in_India_Prospects_and_Precarities_of_Looking_for_Love_Online> accessed 24 September 2026; Choudaha, Sarkar, Sukriti (n 2).

[41] Adnan Bhat and Kaisar Andrabi, 'How Big Tech Platforms Like Instagram Are Failing LGBT Teens in India' (The Daily Beast, 1 April 2024) <https://www.thedailybeast.com/how-big-tech-platforms-like-instagram-are-failing-lgbt-teens-in-india/> accessed 24 September 2026; Madhu and Singh (n 36).

[42] Navya Kharbanda, '16-Year-Old Queer Child Pranshu Dies by Suicide Due to Bullying; Did We Fail as a Society? Mental Health Expert Opines' (Hindustan Times, 28 November 2023) <https://www.hindustantimes.com/htcity/cinema/16yearold-queer-child-pranshu-dies-by-suicide-due-to-bullying-did-we-fail-as-a-society-mental-health-expert-opines-101701172202794.html> accessed 24 September 2026; Bhat and Andrabi (n 41).

[43] Bose, Prabhakar and Agarwala (n 6); Sneha Thomas, 'What India's 2024 Crime Statistics Reveal (and Miss) About Violence Against Women' (Centre for Economic Data and Analysis, Ashoka University, 22 June 2026) <https://ceda.ashoka.edu.in/what-indias-2024-crime-statistics-reveal-and-miss-about-violence-against-women/> accessed 24 September 2026; Ministry of Home Affairs, 'National Cyber Crime Data' (Press Information Bureau, Government of India, 21 July 2026) <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2287039&reg=48&lang=2> accessed 24 September 2026.

[44] Ritika Jain, 'NCRB Data on Cyber Crimes Dated, Not Indicative of True Picture: Experts' (BOOM, 6 October 2025) <https://www.boomlive.in/news/ncrb-report-notes-sharp-rise-in-cyber-crimes-in-india-29662> accessed 28 October 2026.

[45] For more, see, Section 2.

[46] Geeta Pandey, 'Deepfake Deception: Indian Woman's Identity Stolen for Erotic AI Content' (BBC News, 23 July 2025) <https://www.bbc.com/news/articles/cn0znk47x9eo> accessed 10 September 2026; Bose, Prabhakar and Agarwala (n 6).

[47] Samridhi Tewari, 'Gujarat Woman's "Spiritual" Friend Turns Stalker, Held for Posting Her Morphed Photos Online for Months' (The Print, 29 May 2026) <https://theprint.in/india/gujarat-womans-spiritual-friend-turns-stalker-held-for-posting-her-morphed-photos-online-for-months/2944720/> accessed 10 September 2026.

[48] Ibid.

[49] Rajib Dutta, 'Tinsukia Techie Held for Creating, Circulating Morphed Pics Using AI' (The Times of India, 14 July 2025) <https://timesofindia.indiatimes.com/city/guwahati/tinsukia-techie-held-for-creating-circulating-morphed-pics-using-ai/articleshow/122422410.cms> accessed 10 September 2026.

[50] Press Trust of India, '2 School Students Booked for Posting AI Generated Obscene Image of Teacher in Uttar Pradesh' (Deccan Herald, 28 September 2024) <https://www.deccanherald.com/india/uttar-pradesh/2-school-students-booked-for-posting-ai-generated-obscene-image-of-teacher-in-uttar-pradesh-3211295> accessed 10 September 2026.

[51] Caroline Haskins, 'They Dedicated Their Lives to Teaching. Then the Deepfakes Started' (Wired, 24 August 2026) <https://www.wired.com/story/teachers-deepfake-ai-students-content/> accessed 10 September 2026.

[52] Press Trust of India, 'Bengal: Class 10 student found hanging in her room after AI-generated nude images circulated online' (The Print, 28 November 2025) <https://theprint.in/india/bengal-class-10-student-found-hanging-in-her-room-after-ai-generated-nude-images-circulated-online/2794447/> accessed 24 September 2025.

[53] Press Trust of India, 'IIIT Student Held in Raipur for Using AI Tools to Create Obscene Images of 36 Female Students' (Deccan Herald, 10 October 2025) <https://www.deccanherald.com/india/chhattisgarh/iiit-student-held-in-raipur-for-using-ai-tools-to-create-obscene-images-of-36-female-students-3759231> accessed 10 September 2026.

[54] Tathagata Bhattacharjee, '3 Assam Students Suspended for Creating Deepfake Images of 64 Female Teachers, Peers; Probe On' (The Times of India, 12 June 2026) <https://timesofindia.indiatimes.com/city/guwahati/3-assam-students-suspended-for-creating-deepfake-images-of-64-female-teachers-peers-probe-on/articleshow/131689385.cms> accessed 10 September 2026.

[55] Thorn, Deepfake Nudes & Young People: Navigating a New Frontier in Technology-Facilitated Nonconsensual Sexual Abuse and Exploitation (Thorn, March 2025) <https://info.thorn.org/hubfs/Research/Thorn_DeepfakeNudes&YoungPeople_Mar2025.pdf> accessed 10 September 2026.

[56] Kim Nam-young, 'Survey Finds 1 in 5 Male Students Who Create Deepfakes Use Them for Sexual Gratification, Harassment' (Korea JoongAng Daily, 18 January 2026) <https://www.koreajoongangdaily.com/korea/survey-finds-1-in-5-male-students-who-create-deepfakes-use-them-for-sexual-gratification-harassment/12134132> accessed 28 October 2026.

[57] Jean Mackenzie and Leehyun Choi, 'Inside the Deepfake Porn Crisis Engulfing Korean Schools' (BBC News, 3 September 2024) <https://www.bbc.com/news/articles/cpdlpj9zn9go> accessed 24 September 2026; SaferNet Brasil, 'Mapeamento da SaferNet identifica deepfakes sexuais em escolas em 10 dos 27 estados brasileiros' (SaferNet Brasil, 6 October 2025) <https://new.safernet.org.br/content/mapeamento-da-safernet-identifica-deepfakes-sexuais-em-escolas-em-10-dos-27-estados> accessed 24 September 2026; Sally Weale, 'The Rise of Deepfake Pornography in Schools: "One Girl Was So Horrified She Vomited"' (The Guardian, 2 December 2025) <https://www.theguardian.com/society/ng-interactive/2025/dec/02/the-rise-of-deepfake-pornography-in-schools> accessed 24 September 2026.

[58] TOI City Desk, 'Pay or Go Viral: UP Woman Blackmailed Using AI-Generated Obscene Photos; Cyber Cafe Operator, Brother Booked' (The Times of India, 16 May 2026) <https://timesofindia.indiatimes.com/city/lucknow/pay-or-go-viral-up-woman-blackmailed-using-ai-generated-obscene-photos-cyber-cafe-operator-brother-booked/articleshow/131132657.cms> accessed 10 September 2026.

[59] Ibid.

[60] Snigdha Poonam, 'The Dark World of Deepfaked Debt' (The Dial, 30 May 2023) <https://www.thedial.world/articles/news/issue-5/india-loan-app-blackmail-deepfake-images> accessed 24 September 2026; Poonam Agarwal, Nupur Sonar and Stephanie Hegarty, 'Inside the Deadly Instant Loan App Scam That Blackmails with Nudes' (BBC News, 11 October 2023) <https://www.bbc.com/news/world-asia-india-66964510> accessed 24 September 2026.

[61] Ibid.

[62] Huang, Arora and Zarzycka (n 18); Chayn, "Taxonomy of images and content women consider private, intimate, or harmful if shared without consent" <https://reports.chayn.co/image-based-abuse-in-pakistan/taxonomy?from=home> accessed 28 September 2026.

[63] Albeit the constitutionalisation of individual rights and liberties across much of the post-colonial world has not been without contestations and adjustments to liberal constitutionalism. Martha C Nussbaum, "The Feminist Critique of Liberalism" (The Lindley Lecture, University of Kansas, 1997) <https://kuscholarworks.ku.edu/server/api/core/bitstreams/a6b6a3cb-83a9-4841-bae8-7270c5879312/content> accessed 31 October 2026; Tim Connolly, "A cure for individualism" (aeon, 27 May 2025) <https://aeon.co/essays/what-is-the-cure-for-the-wests-individualist-worldview> accessed 31 October 2026; Theunis Roux, "Liberal Constitutionalism Beyond the West?" (Verfassungsblog, 30 October 2026) <https://verfassungsblog.de/liberal-constitutionalism-beyond-the-west/> accessed 31 October 2026; Arun K. Thiruvengadam, "Fortifying Liberal Constitutional Democracy in our Times: Reflections on 'Abusive Constitutional Borrowings" (2021) 12(2) The Journal of Indian Law and Society 116 <https://www.nls.ac.in/wp-content/uploads/2021/09/Thiruvengadam-IJLS-Final-version.pdf> accessed 31 October 2026; Sudhir Krishnaswamy, "Is the Indian Constitution Liberal?" (Academic Foundation, 03 Jul 2019) <https://clpr.org.in/wp-content/uploads/2019/06/Is-The-Indian-Constitution-Liberal.pdf> accessed 31 October 2026; Theunis Roux, "Liberal Constitutionalism in the Post-Colony" (Verfassungsblog, 12 February 2026) <https://verfassungsblog.de/liberal-constitutionalism-in-the-post-colony/> accessed 31 October 2026; Daniel Bonilla Maldonado (ed), Constitutionalism of the Global South (Cambridge University Press, 2013); John Christman, "Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves" (2004) 117 Philosophical Studies 143.

[64] Huang, Arora and Zarzycka (n 18); Chayn (n 62).

[65] Ibid.

[66] In India, the right to privacy was bolstered and reaffirmed within Article 21 of the Constitution of India by the Supreme Court judgement of Justice K.S. Puttaswamy (Retd.) v Union of India & Ors. (2017) 10 SCC 1 (Puttaswamy).

[67] Robert Chesney and Danielle K Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 California Law Review 1753.

[68] See Section 2.1, See generally Huang, Arora and Zarzycka (n 18); Chayn (n 62).

[69] Clare McGlynn et al, '"It's Torture for the Soul": The Harms of Image-Based Sexual Abuse' (2021) 30(4) Social & Legal Studies 541 <https://doi.org/10.1177/0964663920947791>; Dana Mahr, 'Sexualized Deepfakes as a Socio-Technical Continuation of Gendered Power' (2026) 41 AI & Society 7407.

[70] Asher Flynn et al, 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non-Consensually Created and Shared Sexualized Deepfake Imagery' (2025) Journal of Interpersonal Violence <https://doi.org/10.1177/08862605251368834>.

[71] Chayn (n 62).

[72] See more, Vandita Morarka, Karishma Shafi and Arshiya Kochar, ''Has Your Family Seen This?' Using Non-Sexual Images as Sexual Control in India' (Heinrich-Böll-Stiftung, 28 July 2026) <https://www.boell.de/en/2026/07/28/has-your-family-seen-using-non-sexual-images-sexual-control-india>.

[73] Mahr (n 69); Flynn et al (n 70).

[74] Danielle K Citron, 'Sexual Privacy' (2019) 128 Yale Law Journal 1870 <https://scholarship.law.bu.edu/faculty_scholarship/620/>; SK Vaishnavi, 'Synthetic Intimacy and Informational Sexual Autonomy: Regulating Deepfake Pornography in India' (2026) 2(1) International Journal of Law, Culture & Society <https://www.ijlcs.in/Journal/index.php/ijlcs/article/view/28>.

[75] Ibid.

[76] Clare McGlynn, Erika Rackley and Ruth Houghton, 'Beyond "Revenge Porn": The Continuum of Image-Based Sexual Abuse' (2017) 37(3) Oxford Journal of Legal Studies 534.

[77] McGlynn (n 69); Antoinette Huber, '"A Shadow of Me Old Self": The Impact of Image-Based Sexual Abuse in a Digital Society' (2023) 29(2) International Review of Victimology 199.

[78] Ibid.

[79] Kalika Mehta and Avantika Tiwari, 'Between Sexual Violence and Autonomy: Rethinking the Engagement of the Indian Women's Movement with Criminal Law' (2021) 22(5) German Law Journal 860; Sahana Sarkar and Rahul Sinha-Roy, 'Cybersecurity Through an Intersectional Lens: Survivors' Responses to Technology-Facilitated Sexual Violence in India' (2025) Feminist Criminology <https://doi.org/10.1177/1557085125140>.

[80] Clare McGlynn and Erika Rackley, 'Image-Based Sexual Abuse' (2017) 37 Oxford Journal of Legal Studies 534 <https://doi.org/10.1093/ojls/gqw033>.

[81] Mahr (n 69).

[82] Citron (n 74); Clare McGlynn and Rüya Tuna Toparlak, 'The "new voyeurism": criminalizing the creation of "deepfake porn"' (2025) 52(2) Journal of Law and Society 204.

[83] Mahr (n 69).

[84] McGlynn and Toparlak (n 82).

[85] Mahr (n 69); Kathleen Ross, 'Pornographic Deepfakes and Ugly Social Facts: The Costs of a Normative Approach to Defamation' (March 2026) Michigan Law Review <https://michiganlawreview.org/journal/pornographic-deepfakes-and-ugly-social-facts-the-costs-of-a-normative-approach-to-defamation/>.

[86] Ibid.

[87] Huang, Arora and Zarzycka (n 18); Chayn (n 62).

[88] McGlynn and Toparlak (n 82).

[89] Ibid; Mohamed Chawki, Subhajit Basu and Kyung-Shick Choi 'Deepfake Sextortion in England, Wales and Northern Ireland: A Doctrinal and Regulatory Analysis' (2026) 15 Laws 11 <https://doi.org/10.3390/laws15010011>. Chawki and authors discuss Cara Hunter's case, where a Northern Irish serving member of the Legislative Assembly was the subject of NCSI during an election campaign.

[90] Outlook Sports Desk, '"Thousands Of AI Images": Keyaa Banerji's Viral 140kg Lift Turns Into Online Abuse Horror' (Outlook India, 4 September 2026) <https://www.outlookindia.com/sports/others/keyaa-banerji-140kg-lift-ipf-junior-world-championships-online-abuse-ai-generated-images-morphed-body-face>.

[91] Ibid; Hijam Raju Singh, "Keya Banerji lifted 140 kg and went viral. Then came thousands of AI images, fake accounts, and online abuse – why she created a new Instagram account" (Times of India, 4 September 2026) <https://timesofindia.indiatimes.com/sports/more-sports/others/keyaa-banerji-lifted-140kg-and-went-viral-then-came-thousands-of-ai-images-fake-accounts-and-online-abuse-why-she-created-a-new-instagram-account/articleshow/133766556.cms> accessed 31 October 2026.

[92] Li Qiwei et al, 'A Law of One's Own: The Inefficacy of the DMCA for Non-Consensual Intimate Media' (2025) <https://arxiv.org/abs/2409.13575>.

[93] Ibid.

[94] McGlynn & Rackley (n 80); McGlynn and Toparlak (n 82).

[95] Ibid.

[96] Maktoob, 'Right-wing handles continue harassing journalist Rana Ayyub, this time with deepfake video' (Maktoob Media, 18 November 2024) <https://maktoobmedia.com/post?id=95063&slug=right-wing-handles-continue-harassing-journalist-rana-ayyub-this-time-with-deepfake-video>.

[97] Neha Joshi, 'Bombay High Court grants actress Shruti Haasan interim relief on plea against AI deepfakes' (Bar & Bench, 3 September 2026) <https://www.barandbench.com/news/bombay-high-court-grants-actress-shruti-haasan-interim-relief-on-plea-against-ai-deepfakes>; Preity G Zinta v Google LLC & Ors, Interim Application (L) No 20703 of 2026 in Commercial IP Suit No 245 of 2026 (Bombay HC, 8 July 2026) <https://indiankanoon.org/doc/19159052/>.

[98] Ibid.

[99] Aakriti Bansal, 'Personality rights cases: What issues are Indian courts really considering?' (Medianama, 22 September 2026) <https://www.medianama.com/2026/09/223-personality-rights-indian-courts/>

[100] Aakriti Bansal, 'Delhi HC refuses blanket takedown of fan pages in Janhvi, Khushi Kapoor personality rights suits' (Medianama, 13 August 2026) <https://www.medianama.com/2026/08/223-delhi-hc-personality-rights-janhvi-khushi-kapoor/>.

[101] Bansal (n 100).

[102] Mahr (n 69).

[103] Siddharth Pillai, Tarunima Prabhakar and Kaustubha Kalidindi, "Make It Real: Mapping AI-Facilitated Gendered Harm" (Tattle and Rati Foundation, 2025) <https://tattle.co.in/blog/make-it-real/> accessed 23 September 2026; "Sexual deepfakes used to silence voices of women in politics and media" (RFI, 28 May 2025) <https://www.rfi.fr/en/international/20260528-sexual-deepfakes-fuel-efforts-to-silence-women-in-politics-and-media> accessed 23 September 2026; Posetti, J, Williams, K, Hellmueller, L, Renaud, P., Shabbir, N., Aboulez, N. (2026) TIPPING POINT: Online Violence Impacts, Manifestations and Redress in the AI Age (UN Women: New York, April 2026); See Anita Gurumurthy and Anshuman Dasarathy, "Profitable Provocations A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-political Life" (2022) <https://itforchange.net/sites/default/files/2132/ITfC-Twitter-Report-Profitable-Provocations.pdf> accessed 23 September 2026.

[104] Vandinika Shukla, "Deepfakes and Elections: The Risk to Women's Political Participation" (Tech Policy Press, 29 February 2024) <https://www.techpolicy.press/deepfakes-and-elections-the-risk-to-womens-political-participation/> accessed 23 September 2026; Risper Arose, "Digital Shadows: Deepfakes Used As Violence Against Women in Journalism and Politics During African Elections" (CIPESA, 5 February 2025) <https://cipesa.org/2025/02/digital-shadows-deepfakes-used-as-violence-against-women-in-journalism-and-politics-during-african-elections/> accessed 23 September 2026; Rina Chandran, "Flood of AI and deepfake images underline threat to women, sexual minorities in South Asia" (The Scroll, 16 December 2023) <https://scroll.in/article/1060585/flood-of-ai-and-deepfake-images-underline-threat-to-women-sexual-minorities-in-south-asia> accessed 23 September 2026; Sophie Maddocks, "A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes" (2020) Porn Studies <https://doi.org/10.1080/23268743.2020.1757499> accessed 23 September 2026.

[105] Eliza Mackintosh and Swati Gupta, 'Troll armies, 'deepfake' porn videos and violent threats. How Twitter became so toxic for India's women politicians' (CNN, 22 January 2020) <https://edition.cnn.com/2020/01/22/india/india-women-politicians-trolling-amnesty-asequals-intl/index.html>.

[106] Pillai, Prabhakar, Kalidindi (n 103).

[107] Ibid.

[108] Mackintosh and Gupta (n 105).

[109] Seung Gyeong Ji, "#MeToo in an AI-generated deepfake sexual violence era in South Korea" (2025) 112 Women's Studies International Forum.

[110] Aisha Down, "The chilling effect': how fear of 'nudify' apps and AI deepfakes is keeping Indian women off the internet | Global development" (The Guardian, 5 November 2025) <https://www.theguardian.com/global-development/2025/nov/05/india-women-ai-deepfakes-internet-social-media-artificial-intelligence-nudify-extortion-abuse> accessed 23 September 2026; Ji-yeon Lee, "Silenced by non-consensual deepfakes? Perceived threat and online self-censorship among female users: Mediating roles of perceived control, belief in a just world, and passive SNS use" (2026) 21 Telematics and Informatics Reports 100279.

[111] Mishra and Dash (n 1); Yummna Hina Khan, "Digital Violence In Pakistan: From Online Abuse To Real-World Danger" (Friday Times, 11 March 2026) <https://www.thefridaytimes.com/11-Mar-2026/digital-violence-pakistan-onlineno> accessed 23 September 2026; Van Nguyen, "Patterns of Silence" (Dawn, 25 November 2025) <https://www.dawn.com/news/1957241> accessed 23 September 2026.

[112] Pawan Kumar, "Alt News Impact: Meta removes doctored, AI generated visuals of women CJP protesters" (AltNews, 11 September 2026) <https://www.altnews.in/alt-news-impact-meta-removes-ai-manipulated-videos-of-women-from-jantar-mantar-protest/> accessed 23 September 2026.

[113] "Woman who stopped Mumbai police van files complaint against online abuse" (The Hindu, 28 July 2026) <https://www.thehindu.com/news/national/maharashtra/woman-who-blocked-police-van-during-mumbai-protest-alleges-trolling/article71272430.ece> accessed 23 September 2026; "Online Hate, Deepfake Images: Rhiya Ahir Opens Up After Mumbai Gen Z Protest | AI with Sanket" (The Federal, 30 July 2026) <https://www.youtube.com/watch?v=LpsLxEVn3EE> accessed 23 September 2026.

[114] Jessica Ringrose and Katy Regehr, 'Central Comprehensive: Religion, honour, Digital Sexual Double Standards and Victim Shaming and Blaming' in Teens, Social Media, and Image Based Abuse (Palgrave Macmillan, 2025) <https://doi.org/10.1007/978-3-031-92322-7_5>; Law Commission, 'Taking, making and sharing intimate images without consent' (7 July 2022) <https://lawcom.gov.uk/project/taking-making-and-sharing-intimate-images-without-consent/>.

[115] Hasret Çetinkaya, 'The Coloniality of Contemporary Human Rights Discourses on "Honour" in and Around the United Nations' (2023) 31 Feminist Legal Studies 343 <https://doi.org/10.1007/s10691-023-09517-w>.

[116] Sarkar and Roy (n 79).

[117] Bose, Prabhakar, Agarwala (n 6); Down (n 110); Ingrid Beck, Florencia Alcaraz and Paula Rodríguez, "Online gender-based violence against women with a public voice: Impact on freedom of expression" (UN Women, 2022) <https://lac.unwomen.org/sites/default/files/2023-12/report_onlineviolence_21dec23.pdf> accessed 23 September 2026.

[118] Id.

[119] 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) 12:55 Humanities and Social Sciences Communications 1; Tavishi and Angelina Dash, "Platform Accountability in Online Dating: A Critical Analysis of Privacy, Discrimination, and Safety Harms in India" in The Platform Question: Power, Accountability and Global South (2026) <https://ccgdelhi.org/research-academic/platform-accountability-online-dating> accessed 23 September 2026.

[120] Nicola Henry and Anastasia Powell, "Embodied Harms: Gender, Shame, and Technology-Facilitated Sexual Violence" (2015) 21(6) Violence Against Women 758; Fabio Patrone and Marco Viola, "Non consensual sexual deepfakes as direct personal harm" (2026) 39:94 Philosophy and Technology 93; Garry Young, "Some concerns over the attribution of blame to non consensual sexual deepfakes: A response to Patrone and Viola" (2026) 39 Philosophy and Technology 120.

[121] Per Moum Hellevik, Linn-Eirin Aronsen Haugen and Carolina Överlien, 'Outcomes of image-based sexual abuse among young people: a systematic review' (2025) 16 Frontiers in Psychology 1599087 <https://doi.org/10.3389/fpsyg.2025.1599087>.

[122] "Gender based violence is a public health issue: using a health systems approach" (WHO, 25 November 2021) <https://www.who.int/news/item/25-11-2021-gender-based-violence-is-a-public-health-issue-using-a-health-systems-approach> accessed 23 September 2026.

[123] Here, the authors refer to Gillian Young's reference of socio-spacial, "technologically mediated spaces we can access, work and play in, build relationships and communities in, pursue learning, etc."

[124] Henry and Powell (n 120).

[125] Ibid.

[126] Ibid.

[127] Henry and Powell (n 120).

[128] Ajay Jose and Sonia Mathew, 'Addressing the psychiatric implications of AI-enabled non-consensual sexual imagery' (2026) 228 The British Journal of Psychiatry 73 <https://doi.org/10.1192/bjp.2025.10489>.

[129] Ibid.

[130] Ibid; Stine Nygård, Ingela Lundin Kvalem and Bente Træen, '"It Spread Like Wildfire, as These Things Do": Exploring Mechanisms of Harm in Young Norwegians' Experiences of Image-Based Sexual Abuse' (2025) 62 The Journal of Sex Research 582 <https://doi.org/10.1080/00224499.2024.2341129>.

[131] Hellevik, Haugen and Överlien (n 121).

[132] Lee Shepherd et al, 'Fake Content, Harmful Emotions: The Influence of Being a Victim-Survivor of Sexualised Deepfakes on Women's Aversive Emotions and Desire for Legal Action' (2026) 92 Sex Roles 46 <https://link.springer.com/article/10.1007/s11199-026-01677-8>.

[133] Ibid.

[134] Hellevik, Haugen and Överlien (n 121), Carlotta Rigotti and Clare McGlynn, 'Towards an EU criminal law on violence against women: The ambitions and limitations of the Commission's proposal to criminalise image-based sexual abuse' (2022) 13 New Journal of European Criminal Law 539 <https://doi.org/10.1177/20322844221140713>.

[135] McGlynn and Toparlak (n 82); See Section 2.4.

[136] McGlynn & Rackley (n 80); Luise Koch, Maria Paula Russo Riva and Janina Isabel Steinert, 'Technology-Facilitated Gender-Based Violence Against Politically Active Women: A Systematic Review of Psychological and Political Consequences and Women's Coping Behaviors' (2025) Trauma, Violence, & Abuse <https://doi.org/10.1177/15248380251343185>; X Corp v Union of India (Bombay High Court, 24 September 2025) <https://indiankanoon.org/doc/7614885/>.

[137] Ibid.

[138] Justice K.S. Puttaswamy (Retd.) v Union of India & Ors. (2017) 10 SCC 1.

[139] John Sebastian & Aparajito Sen, "Unravelling the Role of Autonomy and Consent in Privacy" (2020) 9 Indian J Const L 1; "Case brief: Justice K.S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors." (Centre for Communication Governance) <https://privacylibrary.ccgnlud.org/case/justice-ks-puttaswamy-ors-vs-union-of-india-ors> accessed 28 September 2026.

[140] M. Kamil, "Puttaswamy: Jury still out on some privacy concerns?", (2017) 1(2) Indian Law Review 190, 202-03.

[141] Sebastian and Sen (n 139).

[142] Gautam Bhatia, "The Supreme Court's Right to Privacy Judgment – II: Privacy, the Individual, and the Public/Private Divide" (Constitutional Law and Philosophy, 28 August 2017) <https://indconlawphil.wordpress.com/2017/08/28/the-supreme-courts-right-to-privacy-judgment-ii-privacy-the-individual-and-the-publicprivate-divide/> accessed 28 September 2026; Independent Thought vs. Union of India and Anr. (2017) 10 SCC 800 in Privacy and the Indian Supreme Court: A Compendium of Case Briefs (Centre for Communication Governance National Law University Delhi, 2023) 254 <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/final-ccg-pll-book-513.pdf> accessed 31 October 2026; X vs. Principal Secretary, Health and Family Welfare Department, Govt. of NCT of Delhi & Anr. 2022 SCC OnLine SC 1321 in Privacy and the Indian Supreme Court: A Compendium of Case Briefs (Centre for Communication Governance National Law University Delhi, 2023) 378 <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/final-ccg-pll-book-513.pdf> accessed 31 October 2026; Jacob Puliyel vs. Union of India & Ors. 2022 SCC OnLine SC 533 in Privacy and the Indian Supreme Court: A Compendium of Case Briefs (Centre for Communication Governance National Law University Delhi, 2023) 372 <https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/final-ccg-pll-book-513.pdf> accessed 31 October 2026; Surabhi Singh, "The Puttaswamy Effect: Exploring the Right to Abortion in India" (2021) Centre for Communication Governance at NLU Delhi.

[143] Sebastian and Sen (n 139); Gautam Bhatia, "The Supreme Court's Right to Privacy Judgment – IV: Privacy, Informational Self-Determination, and the Idea of Consent" (Constitutional Law and Philosophy, 30 August 2017) <https://indconlawphil.wordpress.com/2017/08/30/the-supreme-courts-right-to-privacy-judgment-iv-privacy-informational-self-determination-and-the-idea-of-consent/> accessed 28 September 2026; Torsha Sarkar, "Privacy through the Ages: India's Privacy Jurisprudence in Gender and Sexuality Rights" (2021) 12 J Indian L & Soc'y 53; Severyna Magill, "The Right to Privacy and Access to Abortion in a Post-Puttaswamy World" (2020) 2020 U Oxford Hum Rts Hub J 160.

[144] Usha Tandon and Sidharth Luthra, 'Rape: Violation of the Chastity or Dignity of Woman? A Feminist Critique of Indian Law' (2016) FICHL Policy Brief Series No 51; Sarkar and Roy (n 79); Chayn, "Explicit Harms of Non-Explicit Images Defining Image-Based Abuse in Pakistan and Diaspora" <https://reports.chayn.co/> accessed 28 September 2026.

[145] Ibid.

[146] Ishani Mookherjee, 'Individual's Right to Privacy: Panchayati Eclecticism and Feminist Constitutionalism' (2017) 8 Journal of Indian Law and Society 43; Kalpana Kannabiran and Swethaa S. Ballakrishnen, Gender Regimes and the Politics of Privacy: A Feminist Re-Reading of Puttaswamy vs. Union of India (Zubaan Academic 2021) <https://jilsblognujs.wordpress.com/wp-content/uploads/2025/01/ishani-mookherjee_8.pdf>.

[147] McGlynn and Rackley (n 80); Mahr (n 69).

[148] Ibid.

[149] Tsachi Keren-Paz, 'Pornographic Deep Fakes: Liability for Breach of Privacy in Cases of Parody?' (2025) 18(2) Journal of Tort Law 553 <https://doi.org/10.1515/jtl-2025-0030>.

[150] Ibid.

[151] Krishnadas Rajagopal, "Supreme Court flags repeated breaches of law on sexual assault survivors' identity" (The Hindu, 28 March 2026) <https://www.thehindu.com/news/national/sc-flags-repeated-breaches-of-law-on-disclosure-of-sexual-assault-survivors-identity/article70792238.ece> accessed 23 September 2026.

[152] Nidhi Singh, "Defining the Right to Privacy in India in light of Justice KS Puttaswamy & Anr. v. Union of India (2017)" (Oxford Human Rights Hub, 5 October 2017) <https://ohrh.law.ox.ac.uk/defining-the-right-to-privacy-in-india-in-light-of-justice-ks-puttaswamy-anr-v-union-of-india-2017/> accessed 23 September 2026.

[153] Madhu and Singh (n 36).

[154] McGlynn and Rackley (n 80); Keren-Paz (n 149).

[155] Elizabeth F Judge and Amir M Korhani, 'Deepfakes, Counterfeits, and Personality' (2021) 59 Alberta Law Review, Ottawa Faculty of Law Working Paper No 2021-21 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3893890>; Claire Benn, 'Deepfakes, Pornography and Consent' (2025) 25 Philosophers' Imprint 1 <https://doi.org/10.3998/phimp.2653>.

[156] Justice K S Puttaswamy (Retd) v Union of India (2017) 10 SCC 1.

[157] Mookherjee (n 146); Satyavati Rajesh Khaladkar, 'Right to Privacy for Women: Legal Framework and Social Imperatives' (2025) 5(3) Technoarete Transactions on Advances in Social Sciences and Humanities <https://technoaretepublication.org/socialsciences-humanities/article/privacy-for-women.pdf>; Puttaswamy (n 156).

[158] Tandon and Luthra (n 144); Sarkar and Roy (n 79).

[159] Mahera Imam and N. Manimekalai, "Gender Dynamics and the Misuse of Personal Data: Understanding Vulnerabilities in the Digital Age" (2026) 11(1) Journal of Development Policy and Practice 22; Chayn (n 144).

[160] Ibid.

[161] Ibid.

[162] Renza Kalarikkandy Maliyekkal, Payal Arora and Manisha Shelat, 'Negotiating Digital Autonomy in Rural Kerala: Gendered Constraints and Everyday Resistance' (2026) Women's Studies in Communication (advance online publication) <https://www.tandfonline.com/doi/full/10.1080/07491409.2026.2690644#abstract>.

[163] Imam and Manimekalai (n 159).

[164] Ibid.

[165] Ibid.

[166] Melissa Heikkilä, "How to opt out of Meta's AI training" (MIT Technology Review, 14 June 2024) <https://www.technologyreview.com/2024/06/14/1093789/how-to-opt-out-of-meta-ai-training/> accessed 31 October 2026; Paige Freshwater and Véronique Hawksworth, "AI warning for Facebook and Instagram users who have until November 20 to opt out of changes" (MyLondon, 28 October 2026) <https://www.mylondon.news/lifestyle/ai-warning-facebook-instagram-users-34829895> accessed 31 October 2026; "Meta to use public posts, AI interactions to train models in EU", (Economic Times, 15 April 2025) <https://economictimes.indiatimes.com/tech/artificial-intelligence/meta-to-use-public-posts-ai-interactions-to-train-models-in-eu/articleshow/120295964.cms> accessed 31 October 2026; Jess Weatherbed, "Meta fed its AI on almost everything you've posted publicly since 2007" (The Verge, 12 September 2024) <https://www.theverge.com/2024/9/12/24242789/meta-training-ai-models-facebook-instagram-photo-post-data> accessed 31 October 2026; Terrence O'Brien, "Meta removes language that may have allowed it to scrape news sites" (The Verge, 17 October 2026); Kurt Wagner and Jordan Robertson, "Child Predators turn to Open source AI to make unlimited Illegal Images" (Bloomberg, 7 October 2026) <https://www.bloomberg.com/features/2026-open-source-ai-models-fuel-csam/> accessed 31 October 2026; "Musk's Grok AI chatbot is still making sexual deepfakes, despite X's promise to stop it" (NBC News, 14 April 2026) <https://www.nbcnews.com/tech/tech-news/musks-ai-chatbot-grok-xai-making-sexual-deepfakes-imagine-rcna265855> accessed 31 October 2026.

[167] Matt Burgess, "Deepfake 'Nudify' Technology Is Getting Darker—and More Dangerous" (Wired, 26 January 2026) <https://www.wired.com/story/deepfake-nudify-technology-is-getting-darker-and-more-dangerous/> accessed 31 October 2026.

168] Reece Rogers, "Google's and OpenAI's Chatbots Can Strip Women in Photos Down to Bikinis" (Wired, 23 December 2025) <https://www.wired.com/story/google-and-openais-chatbots-can-strip-women-in-photos-down-to-bikinis/> accessed 31 October 2026; Harriette Boucher, "ChatGPT 'can be made to generate sexualised and violent images'" (Independent, 18 June 2026) <https://www.independent.co.uk/news/uk/home-news/chatgpt-openai-images-sexual-violence-b2998272.html> accessed 31 October 2026; Jessica Klein, "The shock of seeing your body used in deepfake porn" (MIT Technology Review, 14 May 2026) <https://www.technologyreview.com/2026/05/14/1137161/ai-porn-nonconsensual-deepfakes-takedown-piracy-copyright> accessed 31 October 2026; Jim the AI Whisperer, "When ChatGPT's guardrails fail, it doesn't just undress women. It mutilates them." (Medium, 25 July 2026) <https://medium.com/the-generator/chatgpt-generates-gruesome-explicit-images-of-women-when-guardrails-fail-my-research-shows-c0edfac4f129> accessed 31 October 2026.

[169] Mariella Moon, "Hugging Face reportedly plagued with AI models generating adult deepfakes" (engadget, 28 July 2026) <https://www.engadget.com/2224899/hugging-face-reportedly-plagued-with-deepfakes/> accessed 31 October 2026; Julie Inman Grant, "'Deepfakes, doxxing, death threats': The cost women pay to be online" (women'sagenda, 24 November 2025), <https://womensagenda.com.au/latest/deepfakes-doxxing-death-threats-the-cost-women-pay-to-be-online/> accessed 31 October 2026; "Law Professor Talks Cornell 7, Doxing of Jane Doe & 'The Burnerverse'" (Forbes, 30 September 2026) <https://www.wired.com/video/watch/big-interview-dr-mary-anne-franks> accessed 31 October 2026; Jean Wang, Michelle Cera, Blake Spendley, "Dox for Me, O Muse: Meta's New AI Agent Built Lists of People in Vulnerable Groups on Request" (Hunterbrook, 28 September 2026) <https://hntrbrk.com/breaking-news/muse-doxxing> accessed 31 October 2026; Richie Koch, "Meta's smart glasses open door to new levels of facial surveillance" (Proton, 22 October 2024) <https://proton.me/blog/smart-glasses-real-time-doxxing> accessed 31 October 2026. Imran Rahman-Jones, "Meta AI searches made public - but do all its users realise?" (BBC, 13 June 2025) <https://www.bbc.com/news/articles/c0573lj172jo> accessed 31 October 2025.

[170] Imam and Manimekalai (n 159).

[171] Ibid.

[172] Digital Personal Data Protection Act 2023; Annpurna Kumar, 'Protecting Privacy from Deepfake Technology in India: A Critical Analysis of the Indian Legal Framework with Lessons from the European Union' (2026) 9(IV) International Journal of Law Management & Humanities 1950 <https://ijlmh.com/article/view/protecting-privacy-from-deepfake-technology-in-india-a-critical-analysis-of-the-indian-legal-framework-with-lessons-from-the-european-union>.

[173] Ibid; Bharatiya Nyaya Sanhita 2023, s 77 <https://devgan.in/bns/section/77/>; Information Technology Act 2000, ss 66E, 67, 67A.

[174] Jyoti Diwakar, 'The Court of Bigotry: Contested Sexuality of Dalit and Non-Dalit Women' (2022) 14(2) Contemporary Voice of Dalit 199 <https://doi.org/10.1177/2455328X221106537>; "Artificial intelligence: What are deepfakes?" (United Nations, 15 May 2026) <https://unric.org/en/artificial-intelligence-what-are-deepfakes/> accessed 23 September 2026.

[175] Ajita Sharma, 'Marital Rape As Domestic Violence: A Case for Criminalizing Marital Rape' in Indira Jaising and Pinki Mathur Anurag (eds), Conflict in the Shared Household: Domestic Violence and the Law in India (OUP 2019) 128; Bharatiya Nyaya Sanhita 2023, s 63, Exception 2.

[176] Nishasri S, 'Sexual Autonomy of Women in India: Challenges and Way Forward' (2025) 8(2) International Journal of Law Management & Humanities 4741 <https://ijlmh.com/paper/sexual-autonomy-of-women-in-india-challenges-and-way-forward/>.

[177] United Nations (n 174).

[178] Mishra and Dash (n 1).

[179] Ibid.

[180] See Choudaha, Sarkar, Sukriti (n 2).

[181] Mishra and Dash (n 1).

[182] General Data Protection Regulation (GDPR) 2018; Digital Personal Data Protection Act 2023.

[183] Mishra and Dash (n 1); See Choudaha, Sarkar, Sukriti (n 2).

[184] Ding et al. (n 6); Michelle L. Ding and Harini Suresh, "The Malicious Technical Ecosystem: Exposing Limitations in Technical Governance of AI-Generated Non-Consensual Intimate Images of Adults" (arxiv, 2025) <https://arxiv.org/html/2504.17663> accessed 23 September 2026.

[185] McGlynn and Toparlak (n 82).

[186] Carlotta Rigotti, 'International Law, Technology, and Gender-Based Violence' (2024) Oxford Research Encyclopedia of International Studies.

[187] See generally, Ratna Kapur, 'The Tragedy of Victimization Rhetoric: Resurrecting the "Native" Subject in International/Post-Colonial Feminist Legal Politics' (2002) 15 Harvard Human Rights Journal 1

<https://www.researchgate.net/publication/228143416_The_Tragedy_of_Victimization_Rhetoric_Resurrecting_the_Native_Subject_in_InternationalPost-Colonial_Feminist_Legal_Politics>; Aanchal Kabra and Rohit Gupta, 'Carving an Indian Mosaic for Image-Based Sexual Abuse' (2022) 34(1) National Law School of India Review 207 <https://repository.nls.ac.in/cgi/viewcontent.cgi?article=1992&context=nlsir>.

[188] See generally, landmark cases from 2018 like Navtej Singh Johar v Union of India (2018) 10 SCC 1 and Joseph Shine v Union of India (2018) 3 SCC 39. These are Supreme Court of India judgments that struck down colonial-era penal provisions, affirming individual dignity, privacy, autonomy and constitutional morality over prevailing notions of orthodox socio-cultural norms.

[189] Kapur (n 187), Mahr (n 69); Mihaela Mihailova, 'Discriminators for Gender Equality: Deepfakes in Feminist Filmmaking' (2026) 67(2) Screen 111, Rachel Killean, Anne-Marie McAlinden and Eithne Dowds, 'Sexual Violence in the Digital Age: Replicating and Augmenting Harm, Victimhood and Blame' (2022) 31 Social & Legal Studies 625 <https://journals.sagepub.com/doi/10.1177/09646639221086592>.

[190] Rupali Bansode, 'Moments of Erasure of the Testimonies of Sexual Violence Against Dalit Women' in Nicky Falkof, Shilpa Phadke and Srila Roy (eds), Intimacy and Injury: In the Wake of #MeToo in India and South Africa (Manchester University Press 2022); Sarkar and Roy (n 79).

[191] Sarkar and Roy (n 79); Sudeep Mangat and Aisha K Gill, 'Negotiating between Speech and Silence as a Form of Agency: Understanding Dalit Women's Experiences of Sexual Violence in India' (2024) 105 Women's Studies International Forum 102916.

[192] See generally, Diwakar (n 174); Ibid.

[193] Ibid.

[194] Dalit Human Rights Defenders Network, 'Caste Based Sexual Violence and State Immunity' March, 2022 <https://feministlawarchives.pldindia.org/wp-content/uploads/caste-based-sexual-violence-and-state-impunity-english.pdf?>.

[195] Ibid; Jayshree Bajoria, 'Everyone Blames Me': Barriers to Justice and Support Services for Sexual Assault Survivors in India (Human Rights Watch, 2017) <https://www.hrw.org/report/2017/11/08/everyone-blames-me/barriers-justice-and-support-services-sexual-assault-survivors>.

[196] Ibid, Bansode (n 190).

[197] Information Technology Act 2000, Section 66E.

[198] Shazia Choudhry, 'Patriarchal Forestalling - the Anticipatory Structure of Legal Failure on Violence against Women' (2026) Oxford Journal of Legal Studies <https://doi.org/10.1093/ojls/gqag014>.

[199] Ibid.

[200] Anita Gurumurthy, Amrita Vasudevan and Nandini Chami, Born Digital, Born Free? A Socio-Legal Study on Young Women's Experiences of Online Violence in South India (IT for Change 2019) <https://itforchange.net/sites/default/files/1618/Born-Digital_Born-Free_SynthesisReport.pdf>.

[201] Jessica Ringrose and Kaitlyn Regehr, 'Central Comprehensive: Religion, Honour, Digital Sexual Double Standards and Victim Shaming and Blaming' in Teens, Social Media, and Image Based Abuse (Palgrave Macmillan 2025) 123 <https://doi.org/10.1007/978-3-031-92322-7_5>.

[202] Choudhry (n 198); Bajoria (n 195).

[203] Diwakar (n 174).

[204] See generally, Vasudev Devadasan, "Report on Intermediary Liability in India" (2022) Centre for Communication Governance, National Law University Delhi <https://ccgdelhi.org/research-reports/report-on-intermediary-liability-in-india-ccg> accessed 31 October 2026; Choudaha, Sarkar, Sukriti (n 2).

[205] Ibid.

[206] Mangat and Gill (n 191).

[207] See generally, Mahr (n 69), Killean (n 189).

[208] Madhu and Singh (n 36).

[209] Ibid.

[210] Austria, Cameroon, China, Colombia, India, South Korea, Malaysia, Mexico, Mongolia, Pakistan, Russia, Saudi Arabia, the United States, and nine countries within the Caribbean: St. Kitts and Nevis, Barbados, Dominica, St. Lucia, St. Vincent and the Grenadines, Jamaica, Grenada, Montserrat, and Antigua.

[211] Im et al (n 25).

[212] Madhu and Singh (n 36).

[213] Ibid.

[214] See generally, Section 2.

[215] Im et al (n 25).

[216] Ibid.

[217] Madhu and Singh (n 36); Pillai, Prabhakar, Kalidindi (n 103).

[218] UNFPA and UN WOMEN, 2025. "Research Study on Technology-Facilitated Violence Experienced by Women, Girls, LGBTQIA+ persons, and Ethnic Minorities in Sri Lanka." Colombo: UNFPA and UN WOMEN <https://asiapacific.unwomen.org/en/digital-library/publications/2025/09/research-study-on-technology-facilitated-violence-experienced-by-women> accessed 23 September 2026.

[219] "Technology-facilitated gender-based violence: Developing a shared research agenda" (UN Women, 2024) <https://www.unwomen.org/en/digital-library/publications/2024/09/technology-facilitated-gender-based-violence-developing-a-shared-research-agenda> accessed 23 September 2026; "UN Women strategy: Preventing and eliminating technology-facilitated violence against women and girls" (UN Women, 2025) <https://www.unwomen.org/en/digital-library/publications/2025/12/un-women-strategy-preventing-and-eliminating-technology-facilitated-violence-against-women-and-girls> accessed 23 September 2026.

[220] Huang, Arora and Zarzycka (n 18).

[221] Ibid.

[222] Ibid.

[223] Amna Batool, Mustafa Naseem and Kentaro Toyama, “Expanding Concepts of Non-Consensual Image-Disclosure Abuse: A Study of NCIDA in Pakistan” (2024) CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Article No. 398) 1.

[224] “AI Chatbot Helps Victims of Digital Sexual Violence In Latin America” (The Latin Times, 12 March 2025) <https://www.latintimes.com/ai-chatbot-helps-victims-digital-sexual-violence-latin-america-578164> accessed 23 September 2026.

[225] Uli <https://uli.tattle.co.in/> accessed 23 September 2026.

[226] Uli Dataset <https://uli.tattle.co.in/dataset/> accessed 23 September 2026.

[227] “Unveiling sexist narratives: AI approach to flag content on social media” (AI for Good, 7 March 2024) <https://www.youtube.com/watch?v=altcufbFyiw> accessed 23 September 2026; Sebastián Felipe Zambrano, et al. “Detecting Sexist Language in Spanish Social Media Using NLP Models” (UNICC) <https://docs.ie.edu/school-science-technology/Posters%20MBDS/Capstone_Poster_UNICC_MBDS.pdf> accessed 23 September 2026.

[228] Winnie Hu, ''Hostile Architecture': How Public Spaces Keep the Public Out' (The New York Times, 8 November 2019) <https://www.nytimes.com/2019/11/08/nyregion/hostile-architecture-nyc.html> accessed 24 September 2026.

[229] Harry Brignull, 'Dark Patterns' (2018) <https://darkpatterns.org/> accessed 24 September 2026; Arunesh Mathur et al, 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites' (2019) 3(CSCW) Proceedings of the ACM on Human-Computer Interaction, art 81 <https://doi.org/10.1145/3359183> accessed 24 September 2026.

[230] Alex Rosenblat and Luke Stark, 'Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers' (2016) 10 International Journal of Communication <https://ijoc.org/index.php/ijoc/article/view/4892> accessed 24 September 2026; Mohammad Hossein Jarrahi and Will Sutherland, 'Algorithmic Management and Algorithmic Competencies: Understanding and Appropriating Algorithms in Gig Work' in Information in Contemporary Society: 14th International Conference, iConference 2019 (Springer 2019) <https://doi.org/10.1007/978-3-030-15742-5_55> accessed 24 September 2026; Jamie Woodcock, 'The Algorithmic Panopticon at Deliveroo: Measurement, Precarity, and the Illusion of Control' (2020) 20(3) ephemera 67; Divyansha Sehgal and Yatharth, 'Designing Domestic Work Platforms: A Case Study of Urban Company' (Centre for Internet and Society 2022) <https://firn.genderit.org/research/designing-domestic-work-platforms-case-study-urban-company> accessed 24 September 2026.

[231] Choudaha, Sarkar, Sukriti (n 2); Madhu and Singh (n 36).

[232] Aliza Chasan, 'AI "Nudify" Sites Lack Transparency, Researcher Says' (CBS News, 15 December 2024) <https://www.cbsnews.com/news/ai-nudify-sites-payment-redirects-60-minutes/> accessed 10 September 2026; Liv McMahon, 'Deepfake "Nudify" Site Fined £55,000 Over Lack of Age Checks' (BBC News, 20 November 2025) <https://www.bbc.com/news/articles/c0qpq5q9712o> accessed 10 September 2026.

[233] Will Hawkins, Brent Mittelstadt and Chris Russell, 'Deepfakes on Demand: The Rise of Accessible Non-Consensual Deepfake Image Generators' in Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25) (Association for Computing Machinery 2025) 1602 <https://doi.org/10.1145/3715275.3732107> accessed 10 September 2026.

[234] Max Kamachee et al, 'Video Deepfake Abuse: How Company Choices Predictably Shape Misuse Patterns' (arXiv, 2025) arXiv:2512.11815 <https://arxiv.org/abs/2512.11815> accessed 24 September 2026.

[235] 'Common Crawl' (Common Crawl Foundation) <https://commoncrawl.org/> accessed 24 September 2026.

[236] ImageNet, 'ImageNet' (Stanford Vision Lab, Stanford University and Princeton University) <https://www.image-net.org/> accessed 24 September 2026.

[237] Vinay Uday Prabhu and Abeba Birhane, 'Large Image Datasets: A Pyrrhic Win for Computer Vision?' (arXiv, 2020) arXiv:2006.16923 <https://arxiv.org/abs/2006.16923> accessed 24 September 2026; Alexandra Luccioni and Joseph Viviano, 'What's in the Box? An Analysis of Undesirable Content in the Common Crawl Corpus' in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (Association for Computational Linguistics, 2021) <https://aclanthology.org/2021.acl-short.24/> accessed 24 September 2026; Kamachee et al (n 234); Matt Burgess, 'Hugging Face Has a Deepfake Nudes Problem' (Wired, 28 July 2026) <https://www.wired.com/story/hugging-face-has-a-nonconsensual-deepfakes-problem/> accessed 24 September 2026.

[238] Ibid.

[239] Arvind Narayanan, 'Understanding Social Media Recommendation Algorithms' (Knight First Amendment Institute at Columbia University, 9 March 2023) <https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms> accessed 24 September 2026.

[240] Ibid.

[241] Elissa Welle, 'Grok Is Undressing Anyone, Including Minors' (The Verge, 3 January 2026) <https://www.theverge.com/news/853191/grok-explicit-bikini-pictures-minors> accessed 10 September 2026.

[242] Bose et al (n 6).

[243] Choudaha, Sarkar, Sukriti (n 2).

[244] Bose et al (n 6).

[245] Max Hoppenstedt Heinrichs and Marvin Milatz, 'Using AI to Humiliate Women: The Men Behind Deepfake Pornography' (Der Spiegel, 1 July 2025) <https://www.spiegel.de/international/zeitgeist/using-ai-to-humiliate-women-the-men-behind-deepfake-pornography-a-0de338f9-9cec-4ae8-a5ef-9a356b0a5bd4> accessed 24 September 2026.

[246] Tarleton Gillespie, Patricia Aufderheide and Elinor Carmi, 'Expanding the Debate About Content Moderation: Scholarly Research Agendas for the Coming Policy Debates' (2020) 9(4) Internet Policy Review <https://doi.org/10.14763/2020.4.1512> accessed 10 September 2026; Nilay Patel, 'Welcome to Hell, Elon' (The Verge, 28 October 2022) <https://www.theverge.com/2022/10/28/23428132/elon-musk-twitter-acquisition-problems-speech-moderation> accessed 10 September 2026.

[247] Oversight Board, 'Multiple Case Decision: Explicit AI Images of Female Public Figures' (Cases IG-JPEE85LL and FB-4JTP0JQN, Oversight Board, 25 July 2024) <https://www.oversightboard.com/decision/bun-7e941o1n/> accessed 24 September 2026.

[248] Choudaha, Sarkar, Sukriti (n 2).

[249] Ibid.

[250] Divyansha Sehgal and Lakshmi T Nambiar, Online Gender Based Violence on Short Form Video Platforms (Centre for Internet and Society, 2024) <https://cis-india.org/raw/online-gender-based-violence-on-short-form-video-platforms> accessed 10 September 2026.

[251] Oversight Board (n 247).

[252] Tanmay Durani et al, CASCA Report: User Awareness and Experience of Content Moderation on Meta Platforms in India (Centre for Advanced Studies in Cyber Law and Artificial Intelligence 2025) <https://globalfreedomofexpression.columbia.edu/publications/casca-report-user-awareness-and-experience-of-content-moderation-on-meta-platforms-in-india/> accessed 24 September 2026.

[253] Rod McGuirk, 'X Corp Has Slashed 30% of Trust and Safety Staff, an Australian Online Safety Watchdog Says' (AP News, 10 January 2024) <https://apnews.com/article/x-corp-musk-australia-staff-safety-bc4772369cab1fe8dd975132fd8d61ed> accessed 10 September 2026.

[254] Ibid.

[255] Hayden Field and Jonathan Vanian, 'Tech Layoffs Ravage the Teams That Fight Online Misinformation and Hate Speech' (CNBC, 26 May 2023) <https://www.cnbc.com/2023/05/26/tech-companies-are-laying-off-their-ethics-and-safety-teams-.html> accessed 10 September 2026; Lauren Almeida, 'Hundreds of TikTok UK Moderator Jobs at Risk Despite New Online Safety Rules' (The Guardian, 22 August 2025) <https://www.theguardian.com/technology/2025/aug/22/tiktok-uk-moderator-jobs-at-risk-despite-new-online-safety-rules> accessed 10 September 2026.

[256] Shorenstein Center on Media, Politics and Public Policy, 'Trust and Safety Cutbacks' (Election Essentials, Harvard Kennedy School) <https://electionessentials.shorensteincenter.org/tip-2> accessed 10 September 2026.

[257] Faiz Siddiqui, Nitasha Tiku and Elizabeth Dwoskin, 'Inside Musk's Bet to Hook Users That Turned Grok into a Porn Generator' (The Washington Post, 2 February 2026) <https://www.washingtonpost.com/technology/2026/02/02/elon-musk-grok-porn-generator/> accessed 10 September 2026.

[258] Huang, Arora and Zarzycka (n 18).

[259] For more, see Choudaha, Sarkar, Sukriti (n 2).

[260] Nussbaum (n 63); Roux (n 63); Christman (n 63).

[261] Ibid.

[262] Sharma (n 175); Nishasri (n 176).

[263] See Section 3.1; See generally Sharma (n 175); Bajoria (n 195); Bansode (n 190).

[264] Li Qiwei et al, Reporting Non-Consensual Intimate Media: An Audit Study of Deepfakes.

[265] See Section 2.5; Choudaha, Sarkar, Sukriti (n 2).

[266] See generally, Diwakar (n 174); Bansode (n 190).

[267] At the time of development of protections against discrimination; digital sexual harms were not a pervasive concern. The lack of legal evolution is being problematised here. Diwakar (n 174); also see generally, Li Qiwei et al (n 264); See Section 3.1.

[268] See Section 3.2.

[269] Tandon and Luthra (n 144); Sarkar and Roy (n 79).

[270] Owen Carpenter-Zehe, 'TikTok Used Automation in Nearly 100% of Violating-Content Moderation in Europe' (EUobserver, 2 March 2026) <https://euobserver.com/205083/tiktok-used-ai-in-nearly-100-of-violating-content-moderation-in-europe/> accessed 28 September 2026.

[271] 'Meta Looks to AI to Review Harmful Content in Cost-Cutting Drive' (Financial Times, 24 June 2026) <https://www.ft.com/content/39251a31-4a9d-4870-b86c-dc6353d67fdd> accessed 28 September 2026; Andrew Hutchinson, 'Meta Plans to Replace 90% of Content Review Staff with AI' (Social Media Today, 25 June 2026) <https://finance.yahoo.com/technology/ai/articles/meta-plans-replace-90-content-175748836.html> accessed 28 September 2026.

[272] StopNCII.org, 'StopNCII.org' (SWGfL) <https://stopncii.org/> accessed 28 September 2026.

[273] Tech Coalition, 'Tech Coalition' <https://technologycoalition.org/> accessed 28 September 2026.

[274] Ofcom, 'Ofcom to Crack Down on Spread of Illegal Intimate Images and Deepfakes, Boosting Protections for Women and Girls Online' (Ofcom, 9 September 2026) <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-to-crack-down-on-spread-of-illegal-intimate-images-and-deepfakes-boosting-protections-for-women-and-girls-online> accessed 28 September 2026.

[275] StopNCII.org, 'Create Your Case' (StopNCII.org) <https://stopncii.org/create-your-case/> accessed 30 October 2026.

[276] Coalition for Content Provenance and Authenticity, 'C2PA: Verifying Media Content Sources' (C2PA) <https://c2pa.org/> accessed 31 October 2026.

[277] Google DeepMind, 'Google SynthID' (Google DeepMind) <https://deepmind.google/models/synthid/> accessed 31 October 2026.

[278] Li Qiwei et al, 'Position: AI/ML Deepfake Research Is Misaligned with AI-Generated Non-Consensual Intimate Imagery (AIG-NCII)' (arXiv, 2026) <https://arxiv.org/abs/2607.18263> accessed 28 September 2026.

[279] 'Cheapfakes and the Manipulative Editing of Media' (Fake News & Digital Media Literacy, Central Methodist University, 24 July 2026) <https://centralmethodist.libguides.com/fake_news/cheapfakes> accessed 28 September 2026.

[280] National Association of Attorneys General, 'State and Territory Attorneys General Urge Tech and Payment Platforms to Address Deepfake Exploitation' (NAAG, 26 August 2025) <https://www.naag.org/press-releases/state-and-territory-attorneys-general-urge-tech-and-payment-platforms-to-address-deepfake-exploitation/> accessed 28 September 2026; Tech Transparency Project, 'Apple and Google Are Steering Users to Nudify Apps' (Tech Transparency Project, 15 April 2026)

<https://www.techtransparencyproject.org/articles/apple-and-google-are-steering-users-to-nudify-apps> accessed 28 September 2026.

[281] Choudaha, Sarkar, Sukriti (n 2).

[282] Tech Transparency Project, 'Nudify Apps Widely Available in Apple and Google App Stores' (Tech Transparency Project, 27 January 2026) <https://www.techtransparencyproject.org/articles/nudify-apps-widely-available-in-apple-and-google-app-stores> accessed 30 October 2026.

[283] European Parliament, 'AI Act: EP approves simplification measures and "nudifier" app' (Press Release, 16 June 2026) <https://www.europarl.europa.eu/news/en/press-room/20260611IPR45207/ai-act-ep-approves-simplification-measures-and-nudifier-app-ban> accessed 30 October 2026.

[284] Emanuel Maiberg, 'Civitai, Site Used to Generate AI Porn, Cut Off by Credit Card Processor' (404 Media, 20 May 2025) <https://www.404media.co/civitai-site-used-to-generate-ai-porn-cut-off-by-credit-card-processor/> accessed 28 September 2026; Emanuel Maiberg, 'Civitai Ban of Real People Content Deals Major Blow to the Nonconsensual AI Porn Ecosystem' (404 Media, 27 May 2025) <https://www.404media.co/civitai-ban-of-real-people-content-deals-major-blow-to-the-nonconsensual-ai-porn-ecosystem/> accessed 28 September 2026.

[285] Nafia Chowdhury, 'Automated Content Moderation: A Primer' (The Program on Platform Regulation, Stanford Program in Law, Science & Technology, 1 May 2022) <https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/automated_content_moderation_a_primer.pdf> accessed 28 September 2026.

[286] Ibid.

[287] Penn MEDIATED, 'LLMs are the Third Wave of Automated Content Moderation' (Penn MEDIATED, September 2026) <https://mediated.upenn.edu/blog/#llms-third-wave-moderation> accessed 30 October 2026.

[288] Aish Albladi et al, 'Hate Speech Detection Using Large Language Models: A Comprehensive Review' (2025) 13 IEEE Access <https://doi.org/10.1109/ACCESS.2025.3532397> accessed 28 September 2026.

[289] ROOST, 'A New Milestone for Open Source Safety Infrastructure and Transparency: gpt-oss-safeguard' (ROOST, 29 October 2025) <https://roost.tools/blog/a-new-milestone-for-open-source-safety-infrastructure-and-transparency> accessed 28 September 2026; Cambridge Consultants, Use of AI in Online Content Moderation (Ofcom, 18 July 2019) <https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/other/cambridge-consultants-ai-content-moderation.pdf?v=324081> accessed 30 October 2026.

[290] Shahroz Tariq et al, 'LLMs Are Not Yet Ready for Deepfake Image Detection' (arXiv, 2025) <https://arxiv.org/abs/2506.10474> accessed 28 September 2026.

[291] Rumman Chowdhury and Dhanya Lakshmi, '"Your Opinion Doesn't Matter, Anyway": Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI' (UNESCO 2023) <https://unesdoc.unesco.org/ark:/48223/pf0000387483> accessed 28 September 2026.

[292] Gabriel Nicholas, 'Grounding AI Policy: Towards Researcher Access to AI Usage Data' (Center for Democracy and Technology, 13 August 2024)

<https://cdt.org/insights/grounding-ai-policy-towards-researcher-access-to-ai-usage-data/> accessed 28 September 2026; Oversight Board, 'Content Moderation in a New Era for AI and Automation' (Oversight Board, 2024) <https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/> accessed 28 September 2026; Kamachee and others (n 234).

[293] Chowdhury and Lakshmi (n 291).

[294] Ibid.



Centre for Communication Governance
National Law University Delhi
Sector 14, Dwarka
New Delhi – 110078
011-28031265

ccgdelhi.org | privacylibrary.ccgnlud.org |
ccg@nludelhi.ac.in | x.com/CCGNLUD | Instagram/ccgnlud