RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Combating Online Violence Against Women: Evaluating AI-Enabled Legal Mechanisms under Indian Criminal Law

Ashish Dwivedi

Abstract

Online violence against women has become a compulsive growing menace in the rapidly digitising Indian society. With internet penetration never seen before, exceeding 850 millennials users, women are becoming more and more vulnerable to cyber harassment, stalking, image-based abuse, deepfakes, doxing, and coordinated online attacks. The conventional legal systems which target the physical areas find it difficult to deal with the rapid, scale, anonymity and trans-national character of digital violence. This research paper discusses the ways in which Artificial Intelligence (AI) technologies can be used to complement legal processes of resisting online violence against women in the system of Indian criminal law.

The article examines the existing legal framework, such as the Information Technology Act 2000, the Bhartiya Nyaya Sanhita 2023 and related laws and determines their sufficiency in dealing with modern manifestations of digital gender-based violence.

It discusses AI-enabled applications like automated content moderation, deepfake detection, predictive analytics to detect patterns of abuse, natural-language processing to analyse threatening communications, and evidence preservation based on blockchains. Both the potential transformational promise and the limitations of AI technologies are critically evaluated in the research, discussing the issues of algorithmic bias, violation of privacy, excessive dependence on automation, and the digital divide.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

The paper, through the case study, comparative international frameworks, and views of the stakeholders, leads to the argument that AI facilitated legal systems with proper safeguarding and transparency and human supervision can contribute substantially to the ability of India to deal with online violence against women.

Nonetheless, the technological solutions should be supported by wholesome legal changes, institutional capacity development, digital literacy, and cultural change to deal with the realities of gender-based violence.

Keywords: Online Violence, Cyber Harassment, Gender-Based Violence, Artificial Intelligence, Criminal Law, Digital Rights, Content Moderation, Deepfakes, Legal Technology, Women's Safety

1. Introduction

1.1 Background and Context

India has one of the largest and most rapidly expanding internet users and therefore, the challenge is long and sharp. The National Crime Records Bureau (NCRB) showed that cyber crimes against women had risen by 36 percent in 2019 to 2021, and the most reported types of crimes were sexual harassment and morphing of images [1]. Such statistics are, however, probably grossly underestimated because of underreporting, ignorance on legal remedies and the volatile nature on Internet evidence. Studies by other organizations like the Amnesty International and the United Nations Women have reported that online violence against Indian women is rampant and has been widespread especially among women journalists, activists, politicians and public figures in marginalized communities[2].

1.2 Role of Artificial Intelligence

The use of AI technologies can provide a possible way out of these issues. AI utilises computer systems that could execute functions that one would have performed with human intelligence, such

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

as learning through data, recognising patterns, making decisions, and natural language comprehension. Within the framework of fighting internet violence, AI may allow providing automated identification of offensive content at scale, language patterns analysis to detect threats and harassment, deepfakes and manipulated images, predicting the tendencies of abuse escalation, preserving and examining digital evidence, and recognizing coordinated abuse campaigns.[3].

Some technological applications already have AI-based content moderation systems to filter out harmful content. Nonetheless, these systems have some major weaknesses such as high error rates, cultural and linguistic biases, absence of transparency and inability to deal with context-specific content. Implementing AI into legal tools that are specifically crafted to address online violence against women has to consider the technological limitations and capabilities with great care, legal, ethical, and social consequences of the matter.

1.3 Research Objectives

This research paper aims to:

1. Analyze the current legal framework addressing online violence against women in India
2. Examine how AI technologies can enhance detection, investigation, and prosecution of online violence
3. Evaluate existing AI-enabled mechanisms and their effectiveness
4. Identify challenges and limitations in deploying AI for combating online violence
5. Propose recommendations for developing an integrated framework combining AI technologies with strengthened legal mechanisms

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

1.4 Methodology

The methodology used in this research is a doctrinal legal analysis and empirical analysis of technological capabilities. The methodology will involve the review of law and case law and the policy documents; discussion of the academic literature on the topic of online violence, AI technologies and the legal framework; case study and examples of implementation; comparison of approaches on the topic of online violence across different countries; and the interests of stakeholders such as the law enforcement, technology platforms, civil society organisations, and victims.

2. Understanding Online Violence Against Women

2.1 Forms and Manifestations

Violence against women online takes various forms that have different kinds of harms.

Cyber Harassment and Trolling: a series of unsolicited messages and comments or communications that are meant to intimidate, threaten, or demean. Rape, death threats, sexualised abuse, body shaming and coordinated harassment campaigns where multiple accounts are directed to one person are considered under this category.[4]

Cyber Stalking: The continuing online harassment and stalking and may be resources to track the online actions of a person, constant unwanted attention by phone, email and various other platforms, spying using spyware or unauthorised entry to the devices and use of technology to determine the physical location of an individual.

Non-Consensual Intimate Image Sharing (NCII): Often known as revenge porn, involves the sharing or threat of sharing intimate photos or videos without their consent with devastating psychological consequences, reputation, and often social ostracism and professional consequences.

Deepfakes and Image Manipulation: Photographs that are created synthetically through means of AI are used to imprint the face of a person into a pornographic or embarrassing material or even to

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

modify existing images to create such a false impression. Women have become the targets of deep-fake pornography, which has become even more widespread and advanced.[5]

Doxing: Unauthorized disclosure of personal data Doxing is the release of personal information (residential address, telephone number, workplace, family, and so on) with the aim of causing offline harm or harassment.

Impersonation and Fake Profiles: The establishment of a fraudulent account in the name of a particular person in order to share harmful content, request undesirable interaction, or ruin a reputation.

Online Sexual Harassment: The unwanted sending or reception of sexual pictures or videos, sexual messages, demands or proposals of sexual acts or unsolicited sexual advances.

2.2 Impact on Victims

The consequences of online violence are much farther reaching than the online space.

Psychological Harm: This is visible through anxiety, depression, post-traumatic stress disorder (PTSD), sleeplessness, and in extremely extreme cases, depressive suicidal thoughts. The enduring quality and viral quality of online data can continue to create trauma.[6]

Social and Economic Consequences: The victims might retreat to the online space, thus reducing professional prospects, politics and socializing. Women reporters, feminists and women in general can either censor themselves or withdraw completely, a process known as silencing.

Physical Safety Risks: Threats of violence often turn into offline threats or real-life violence over online violence. Doxing may subject the victims to stalking, beating, or other offline malice.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Reputational Damage: Spread of fake news, doctored photos or other non-consented intimate information may cause permanent harm to personal and professional reputation, and employment, relationship, and social status.

2.3 Indian Context and Prevalence

There is a unique situation of online violence against women in India.

Digital Divide: Internet penetration is growing at a fast rate but there is unequal access between the genders, classes, caste and geography. The rural population and marginalised groups of women are both digitally illiterate and at risk, when they venture into the internet.

Patriarchal Social Structures: The violence of the Internet is a mirror of the existing gender inequalities and patriarchal attitudes, which are aimed at controlling the behaviour and mobility of women and their expression.

Intersectionality: Women who occupy a marginalised group such as Dalit women, Muslim women, tribal women, transgender women and others are extra abused not only based on gender but also on other affinities.[7]

Underreporting: Most of the online violence goes unreported because of shame, fear of revenge, ignorance of legal resolutions, distrust towards the police and believing that they are to blame or no one will believe them.

Studies including the work of organisations like the Point of View and the Internet Democracy Project have reported systematic violence against women in India on the Internet on sites like social media, messaging apps, online forums, and even dating sites. Women in digital space, content creators, journalists and activists are subjected to particularly serious and long-lasting abuse aimed at increasing their silence. [8]

3. Legal Framework in India

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

3.1 Constitutional Provisions

The Indian Constitution can provide the basic protections in case of online violence.

Article 21: Right to life and individual liberty has been construed by the Supreme Court as living with dignity, violence, and harassment free life and right to privacy, which is violated by the internet violence.

Article 14 and 15: Ensure equality before the law and no discrimination on the basis of sex, the targeting of women is disproportionate and this is gender-based discrimination.

Article 19(1)(a): The right to communicate freely and express is guaranteed and online violence is aimed at silencing this freedom especially when we are talking about the expression of women.

3.2 Information Technology Act, 2000 and Amendments

The first law regarding cybercrime in the Republic of India is the Information Technology Act, 2000 (IT Act), which has undergone amendments in 2008 and 2011. In particular, the online violence regulation is especially relevant to the following provisions:

Section 66A (Struck Down): this was initially a criminal offence which made the transmission of offensive messages through communication services a criminal offence, but it was declared invalid by the Supreme Court in Shreya Singhal v. Union of India (2015) based on its constitutional unconstitutionality and violation of right to free-speech [9].

Section 66C The provision relates to identity theft and it punishes the fraudulent or dishonest use of electronic signatures or passwords. It is applicable to impersonation and fake profile.

Section 66D: This section is a punishment against impersonation through the use of computer resources.

Section 66E: This section pertains to the breach of privacy by capturing, publishing, or transmitting images of the personal areas of a person without their permission thus it can be applied to revenge

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

pornography and image-based abuse cases.

Section 67: The clause punishes the publication or transmission of obscene content in electronic format; the issue of whether or not specific intimate images that are consensually created and then shared non-consensually should be punished is still controversial.

Section 67A: This provision deals with sexually explicit content and the penalties that are given are tougher with the first and recurring offences.

Section 67B: This particular provision is criminalising content that shows children engaging in sexually explicit activities, and thus can offer effective protection against child sexual abuse content on the internet.

3.3 Bhartiya Nyaya Sanhita, 2023

The Bhartiya Nyaya Sanhita (BNS), which came to replace the Indian Penal Code in 2023, contained important provisions on cybercrimes against women:

Section 77: this section directly focuses on the subject of non-consent to the sharing of intimate images, which makes sharing of intimate images without consent an offence. The provision addresses directly the issue of revenge pornography and associated image-based abuse.[10].

Section 78: This section deals with sexual harassment over the digital platform, such as sending sexually explicit content, soliciting sexual actions, and making sexually suggestive comments on electronic platforms.

Section 74: The BNS has integrated cyberstalking in its interpretation of stalking because stalking can also happen on the internet like social-media surveillance, frequent text messaging, internet monitoring.

Section 79: This section deals with cyber bullying and cyber abuse via the Internet.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

These provisions are a major step in the consideration of digital aspects of gender-based violence and in the provision of special legal provisions to be used in prosecution.

3.4 Other Relevant Legislation

Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013: Although this Act targets physical workplaces, courts have understood this Act to extend to harassment through digital means as long as they are relating to the workplace situations.

Protection of Children from Sexual Offences (POCSO) Act, 2012: The Act is detailed in providing protection against sexual offences against children even those committed via digital means.

3.5 Gaps and Limitations

Although there are such legal provisions, there are still many gaps: **Definitional Gray areas:** The words, which include harassment, stalking, and offensive, are not clearly defined and this makes it difficult to interpret them.

Complexity of Procedures: Victims are faced with complicated reporting systems; most of the police stations have the knowledge and desire to take complaints seriously.

Jurisdictional Problems: Jurisdictional problems arise when both the perpetrator and the victim are in separate states or countries making investigation and prosecution tricky.

Platform Accountability: The legal practices in place are insufficient to encompass the platform liability in the hosting or the amplification of harmful content.

Speed and Scale: The current system of investigations and prosecutions is unable to match the pace of the online propagation of violence or the sheer volume of abusive campaigns, [11].

4. AI Technologies for Combating Online Violence

4.1 Automated Content Moderation

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

With the amount of content being posted on the internet, billions of posts, images, videos, etc., human moderators cannot scroll through all the content. The AI-based systems are able to analyze large amounts of data in a short period of time and identify possible infractions due to textual, visual, or audio information.

Text-Based Moderation: Natural Language Processing (NLP) algorithms perform the analysis of the text data to identify threatening language, sexual harassment, hate speech, and other harmful communications. These systems will allow detecting explicit threats, sexual solicitation, abusive linguistic patterns, and organized harassment campaigns.[12]

Image-Based Moderation: These systems will recognize the banned content without the human examination of every picture.

Video Moderation: Video material is more complex, and it requires visual and sound analysis. AI systems are capable of analyzing video data to detect banned content but this is also currently technologically difficult because of the complexity of video signals.

Contextual Analysis: Advanced AI systems attempt to make inference about contexts, understanding that the same language can be acceptable in one context (e.g. educational discourse about sexuality) but harassment in another (e.g. unsolicited sexual messages).

4.2 Deepfake Detection

The most dangerous types of AI-created synthetic media include deepfakes, which is a manipulation of faces or voice, especially deepfake pornography. Machine learning is applied by Deepfake detection technologies to detect synthetic media by:

Facial and Biological Markers: Study of discrepancies in movements of the face, the pattern of blinking, the texture of the skin and other biological indicators that deep fake algorithms cannot imitate without flaws.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Digital Forensics: Digital fingerprints, compression artefacts, and metadata analysis to expose manipulation.

Temporal Consistency Analysis: Temporal Consistency Analysis: Detection of inconsistencies between video frames, which suggest synthesis generation.

Audio Analysis: Synthetic audio detection: Voice patterns and phonetic consistency analysis and audio artefact detection.[13]

Challenges:

- **Arms Race Dynamic:** With the advancement of detection technology, the technologies of deepfakes are also improved, which leads to the fact that the competition of technologies is constantly taking place.
- **Authentic Content Misidentified:** System can label authentic content as a fake one.
- **Resource Intensive:** Deepfake detection is resource-intensive.

4.3 Predictive Analytics and Pattern Recognition

AAI is able to identify trends that signal the increase of the abuse or organised campaigns:

Behavioral Pattern Analysis: This is done to identify accounts that have stalking behaviour, e.g. the constant viewing of the profile of another, constant messages, monitoring of online activity patterns.

Coordinated Behavior Detection: Detection of co-ordinated harassment campaigns in which several accounts attack a person at the same time indicating that they are not isolated occurrences; meaning the abuse is organised.

Escalation Prediction: This predictive step involves monitoring the behavior of conversations in order to detect threats that are escalating, which might represent a higher risk of offline violence, and therefore can be acted upon.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Network Analysis: Visualising the links between accounts that are involved in harassment to determine the abuse networks and those who organise them.[14]

4.4 Natural Language Processing for Threat Assessment

NLP technologies will aid in the systematic study of text-based communications to assess the threat level:

Threat Severity Classification: This is the classification of threats on the basis of specificity, imminence, and severity to guide the law-enforcement prioritization.

Sentiment Analysis: Evaluation of emotional tone and intensity that are used to determine the communications that are associated with a true intent to cause harm versus those that are associated with hyperbole or venting.

Linguistic Profiling: Study of writing style and vocabulary and stylistic features to possibly define the perpetrators or connect many accounts with one.

Multilingual Analysis: The processing of content in more than one language and dialect relevant to Indian contexts such as Hindi, Tamil, Bengali and so on and code switching phenomena[15].

4.5 Evidence Collection and Preservation

The evidence provided in digital format is volatile, capable of being deleted, changed, or lost. Evidence preservation can thus be improved using AI technologies:

Automated Evidence Capture: The automated systems are the ones that capture and date digital media before it may be deleted.

Blockchain-Based Authentication: Use of blockchain technology to establish tamper-resistant records of digital evidence, and to guarantee authenticity and chain-of-custody.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Metadata Preservation: Metadata (timestamps, IP addresses, device information, etc.) taken and preserved that can help identify perpetrators.

Cross-Platform Tracking: It is a method of monitoring the spread of content on various platforms to record viral transmission and the amplifiers of harmful content.

5. Existing AI-Enabled Mechanisms

5.1 Platform-Based Content Moderation

Major technology platforms including Facebook (Meta), Twitter (X), Instagram, YouTube, and WhatsApp deploy AI-based content moderation systems. These platforms use machine learning to:

- Recognize and filter out automatically content that breaches the standards in the community.
- Take proactive action in detecting accounts which are involved in harassment.
- Minimize exposure of potentially risky material.
- Offer reporting abuse tools to users.

Indian Context:

Under the pressure of the government and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code), 2021, the platforms used in India must have grievance-redressal frameworks, designate compliance officers, and delete the content that is labeled as illegal within set timeframes. Most platforms have established India-specific content-moderation groups and implemented AI systems that have been tailored to the Indian languages and cultures.[16].

Limitations:

- Social networks are driven by commercial interests, and this can be against user safety.
- ontent-moderation has a lack of transparency and responsibility.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

- The process of appeals is poor.
- Regional languages are poorly handled by the systems.
- Indian generated volume of content is under-resourced in its operations.

5.2 Law Enforcement Integration

he use of AI technologies has been introduced in some Indian law-enforcement agencies:

Cyber Crime Cells: Cyber-crime cells have been created in some states that have digital forensics and artificial intelligence-based investigation tools.

Automated Complaint Systems: Certain jurisdictions have online portals on which cyber-crime victims may report their complaints, and AI systems are used to classify and rank complaints.

Facial Recognition Systems: The facial-recognition technology has been used on rare occasions to target the offenders of image based violence, though most of the time these systems are used in the security sector.

National Cyber Crime Reporting Portal: The government has a national portal of reporting cyber crimes; its AI capabilities are not extensive yet [17].

5.3 Civil Society and NGO Initiatives

AI-based tools have been created by the organizations that are concerned with the rights of women and their online safety:

Online Harassment Helplines: Organisations like the Internet Democracy Project or the Point of View offer help to the victims and use AI to help analyse the patterns of abuse and record evidence.

Automated Documentation Tools: The tools that help victims to document online harassment, screenshots, timestamps, and metadata are all automated.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Safety Applications: Mobile applications providing safety material, law, and evidence-gathering functions.

5.4 Judicial Recognition

The importance of digital evidence and investigations enabled by artificial intelligence has become recognized by the Indian courts more and more:

Electronic Evidence Admissibility: Courts have also defined rules to be used in admitting electronic evidence such as social media posts, messages, and digital images.

Speedy Disposal: The courts have developed expedited systems to handle cases related to cyber-crime since it is evident that digital evidence is very urgent.

Progressive Interpretation: The development of new legal approaches to curb new types of online violence that were not clearly envisioned in the laws has been through progressive interpretation of the current statutes.

6. Challenges and Limitations

6.1 Algorithmic Bias and Discrimination

The biases in AI systems always exist in the training data, as well as in the design:

Gender Bias: Gendered aspects of abuse can not be identified by content moderation systems, and gender-based violence has become the same as generic trolling.

Caste and Community Bias: Systems which are trained mostly on general datasets might not be able to detect caste-based abuse or communal violence against a particular community.

Language Bias: Sub optimal performance with non-english content means users of other languages, who speak Indian languages are at a disadvantage.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Class Bias: AI can give additional protection to people who belong to privileged backgrounds if their speech patterns and context are more overrepresented in training corpora.[18]

6.2 Privacy and Surveillance Concerns

The implementation of AI-based surveillance poses the question of salient privacy concerns:

Mass Surveillance: Mechanized attention to internet communications makes possible an unprecedented surveillance, which may have a chilling role in free expression.

Data Collection: AI systems need a large amount of user data, which raises the issue of data protection and the possible abuse of such data.

Function Creep: Technologies which were implemented as a safety measure can later be turned into a general surveillance or control tool.

Differential Privacy: Women in the marginalised groups have a higher risk of being over-surveilled and lack proper protective measures.

6.3 Over-Reliance on Technological Solutions

The sixth issue is that of over-reliance on technological solutions. Problems based on social inequities cannot be corrected through the use of technological interventions:

Patriarchal Attitudes: AI do not suit the context of dealing with the latent patriarchal attitudes that set the stage of the occurrence of gender-based violence.

Structural Inequalities: Technology cannot address caste, class, and religious inequalities which cross with gender-based violence.

Human Judgment: The human judgment, the empathies and the understanding of the contexts that require a human touch are complex cases that cannot be handled by AI systems.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Victim Agency: Over-automation can destroy the victim agency and reduce the voice of victims in the resolution of abuse.

6.4 Digital Divide

Solutions based on AI have a risk of causing an increase in existing inequalities:

Access Barriers: Women who are not digitally literate or have no access to technology will be deprived of the benefits of AI-enabled protections.

Resource Requirements: Advanced AI-based systems require computational resources that many jurisdictions have no available resources to support law-enforcement agencies that are resource-constrained.

Language Barriers: inadequate performance with state languages is unfavorable to non-English users.

Urban-Rural Divide: AI-based systems are mainly present in cities and leave rural women without proper protection [19]

6.5 Transparency and Accountability

Transparency and accountability with regard to the financial performance of the organization are enshrined in the transparent financial reporting procedures through which the firm presents financial performance to the stakeholders. AI systems are often black box systems:

Algorithmic Opacity: AI systems have their inner workings, which are frequently proprietary and opaque, and no one can scrutinize them.

Decision-Making Accountability: Accountability mechanisms are ineffective when AI systems make consequential decisions (e.g. the removal of content, flagging of accounts, etc.)

Appeal Mechanisms: There are few ways that users can challenge AI-based decisions.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Platform Power: The influence of the online discourse by private technology platforms is enormous with minimal public responsibility.

6.6 Legal and Regulatory Gaps

There is a lack of AI technologies in current legal practices: Algorithms: No complete regulation system is applied to AI systems used in content moderation or law enforcement.

Platform Liability: The existing rules are insufficient when it comes to platform liability on harms supported by their AI systems.

Data Protection: The current data protection system in India is unsophisticated, and the data on users are not properly secured.

International Coordination: Online violence is trans-national as it requires international cooperation but the existing systems fail to support such collaboration satisfactorily[20]

7. Recommendations

7.1 Legal and Regulatory Reforms

Comprehensive Cyber Violence Legislation: Establish new laws specifically on in relation to cyber violence against women, which offer well-defined definitions of crime, simplified reporting, victim compensation measures, and well-defined accountability policies against online sites.

AI Regulation Framework: Develop a detailed regulatory system on artificial intelligence systems used in content moderation and law-enforcement capabilities, which would require transparency, accuracy and fairness standards, human review of consequential decisions, and well-developed accountability methodology.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Platform Accountability: Establish clear legal requirements of platforms operating in India, which include a duty of care to the users, the need of safety-by-design principles, full disclosure of the content-moderation policies, and substantive appeals policies.

Data Protection: Enact stringent data-protection laws that would protect the privacy of users and still allow legitimate access of law enforcement.

7.2 Institutional Capacity Building

Specialized Cyber Crime Units: The units are the specialized cyber-crime units in every district that are provided with digital-forensics capacity, AI-based investigation procedures, and staff members trained in gender-sensitive investigative practices.

Training Programs: Have police officers, prosecutors and judges undergo extensive professional training regarding cyber crimes against women, collection of digital evidence and trauma-proof methods of investigation.

Technology Infrastructure: Invest into creation of a high-quality technological infrastructure that facilitates the workflows of AI-assisted investigation and prosecution.

Inter-State Coordination: This is to create inter-state mechanisms that work together to deal with interstate cyber crimes.

7.3 AI System Design and Deployment

Inclusive Design: Provide AI with an inclusive design process with the integration of contributions of different stakeholders, such as women with different socioeconomic statuses and cultures, and multilingual and cultural awareness.

Bias Mitigation: Perform strong testing and audit to detect and remove algorithmic bias.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Human Oversight: Maintain human review of consequential decisions, where trained human operators review AI-tagged content.

Transparency and Explainability: Have the full documentation of the AI system design and functionality so that it can be viewed by the public leading to accountability and scrutiny.

Victim-Centric Approach: Build systems in which victims are empowered through giving them agency, ensuring privacy, and respecting autonomy.

7.4 Multi-Stakeholder Collaboration

Public-Private Partnerships: Encourage the cooperation of government organisations, technology platforms, civil-society organisations and academic institutions.

Platform Cooperation: Build ways that allow global platforms to exchange information on coordinated campaigns on abuse, patterns of perpetrator behaviour, and useful interventions.

Civil Society Engagement: Promote civil-society groups which offer the frontline services to victims, and record the cases of web-based violence.

Academic Research: Invest in the online violence against women, AI technologies, and the effectiveness of intervention strategies research.

7.5 Victim Support and Empowerment

Legal Aid: Provide affordable legal services to victims who are going through the complicated process of cyber-crime.

Counseling Services: This is the provision of counselling and mental-health services to online violence victims.

Digital Literacy: Launch an extensive digital -literacy campaign involving training women to protect themselves on-line, record abuse, and seeking legal solutions.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

Community Support: Provide social support networks to the victims at the community level.

7.6 Prevention and Cultural Change

Public Awareness: Introduce awareness programs which clarify the nature, harms and the remedies available to online violence.

Education Programs: Incorporate digital citizenship and good online conduct in the formal education programs.

Bystander Intervention: Promote bystander intervention, so that the users can assist victims and report abuse.

Cultural Transformation: Resolve patriarchism which perpetuates gender-based violence by implementing long-term social-change efforts.

8. Conclusion

Online violence against women is not only an unrelenting problem but a growing serious problem that has posed a threat to the safety, dignity, and equality of women in India. Traditional legal systems are just one of the factors but are not as quick and complex as digital abuses. Artificial intelligence turns out as a possible tool; with a properly developed tool, AI will be able to spot dangerous content, detect deep-fakes, extract patterns, and preserve digital evidence, which has never been possible before.

However, it is undoubted that AI cannot solve the issue on its own. Isolated technology is not sufficient. Algorithms come with other challenges that include bias, violation of privacy, digital divide and most importantly, they do not solve the problems that cause violence. Relying on technologic solutions too much, and ignorance of underlying social injustices and patrarchy, mean very little. Furthermore, poorly constructed AIs may make the lives of disadvantaged groups worse and the process of securing the most vulnerable less efficient.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

In this way, the interventions require a complex strategy. This requires the combination of AI and solid laws, highly qualified organizations, real cooperation among stakeholders, full support of victims, and the campaigns to change culture. AI must be used as an addition to the human decision, but not to replace it, and strict protective measures must be established in order to guarantee transparency, responsibility, and fairness.

The Bhartiya Nyaya Sanhita 2023 is an important step forward, considering the fact that gender-based violence is not confined in the physical space only but spreads to the virtual space as well. However, additional steps are required. India needs to implement specific rules on AI, bring platforms to task, and protect the data integrity. The police departments need better equipment and specialized education to address cybercrime. It will require a concerted effort of the government agencies, technology companies, civil society, and even academia to reach this goal; no individual player can attain this goal alone.

Finally, the control of the digital spaces through women would be beyond technology or legislation. It is in line with the wider fight against gender inequality and social injustices and the deinstitutionalization of patriarchy. With digitalization in the air, ensuring that women feel safe and empowered on the internet should jump to a national agenda and it will require a concerted effort by all stakeholders including technology, law, institutions and the rest of the society.

References

i. National Crime Records Bureau. *Crime in India 2021: Statistics Report*. Ministry of Home Affairs, Government of India, 2022.

ii. Amnesty International. *Toxic Twitter: A Toxic Place for Women*. Amnesty International, 2018. https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/

iii. Gorwa, Robert, et al. "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance." *Big Data & Society*, vol. 7, no. 1, 2020, pp. 1-15.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

iv. Citron, Danielle Keats and Mary Anne Franks. "Criminalizing Revenge Porn." *Wake Forest Law Review*, vol. 49, no. 2, 2014, pp. 345-391.

v. Chesney, Robert and Danielle Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review*, vol. 107, 2019, pp. 1753-1820.

vi. Pew Research Center. *Online Harassment 2021*. Pew Research Center, 2021. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

vii. Tharoor, Ishaan. "India's Digital Divide Keeps Millions of Women Offline." *The Washington Post*, 15 March 2020.

viii. Point of View. *Technology-Enabled Violence Against Women in India*. Point of View, 2020. https://pointofview.org/technology-violence-against-women-india/

ix. Supreme Court of India. *Shreya Singhal v. Union of India*, (2015) 5 SCC 1. This landmark judgment struck down Section 66A of the IT Act for violating free speech rights.

x. Bhartiya Nyaya Sanhita, 2023, Section 77. This section specifically addresses non-consensual sharing of intimate images, representing significant advancement in legal protections.

xi. Kulkarni, Anushka. "Challenges in Prosecuting Cyber Crimes Against Women in India." *Indian Journal of Criminology*, vol. 48, no. 2, 2020, pp. 234-256.

xii. Schmidt, Anna and Michael Wiegand. "A Survey on Hate Speech Detection Using Natural Language Processing." *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 2017, pp. 1-10.

xiii. Tolosana, Ruben, et al. "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." *Information Fusion*, vol. 64, 2020, pp. 131-148.

xiv. Kumar, Srijan, et al. "An Army of Me: Sockpuppets in Online Discussion Communities." *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 857-866.

xv. Waseem, Zeerak and Dirk Hovy. "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter." *Proceedings of the NAACL Student Research Workshop*, 2016, pp. 88-93.

RMSG (Rashtrakavi Maithili Sharan Gupt) PUBLICATION
ISSN:3049-4109
INTERNATIONAL DOUBLE-PEER-REVIEWED MULTIDISCIPLINARY E-RESEARCH JOURNAL

xvi. Ministry of Electronics and Information Technology. *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*. Government of India, 2021.

xvii. National Cyber Crime Reporting Portal. *Annual Report 2022*. Ministry of Home Affairs, Government of India, 2022. https://cybercrime.gov.in/

xviii. Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018. This seminal work examines how algorithmic systems perpetuate and amplify existing biases and inequalities.

xix. Irani, Lilly and M. Six Silberman. "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 611-620.

xx. European Union. *Regulation (EU) 2022/2065 on a Single Market For Digital Services (Digital Services Act)*. Official Journal of the European Union, 2022. This comprehensive regulation establishes platform accountability framework applicable across EU member states.