

Editorial

AI and Gender-Based Violence A Double-Edged Sword in the Fight for Justice and Equality

Nazish Imran,¹ Khaula Tariq²

¹Professor, Child & Family Psychiatry Department/ Editor, Annals of KEMU, King Edward Medical University, Lahore;

²Senior Registrar, Department of Psychiatry and Behavioural Sciences, King Edward Medical University / Mayo Hospital Lahore Pakistan

Correspondence: nazishimrandr@gmail.com

The fight against gender-based violence (GBV) is a crucial one, as it affects millions of people around the world particularly women and girls. This issue has no borders since it touch every culture and society. Shockingly, an estimated 736 million women world-wide have experienced intimate partner violence or non-partner sexual violence at some point in their lives.¹

Artificial intelligence (AI) is widespread, it is playing an ever-expanding role in shaping how we communicate but when it comes to addressing GBV, AI is a mixed bag. On one hand, it can help to bring about innovative solutions to help identify and prevent GBV. On the other hand, it has the inherent risk of reinforcing the harmful biases already existing in the society.²

This editorial sheds light on how AI can act as a double-edged sword in the battle against GBV. Finding effective ways to tackle AI for the greater good, and as an aid in the pursuit of justice and equality, while being mindful and mitigating challenges, is the goal and way forward.

The Dual Role of AI in Gender-Based Violence

AI & Amplifying Risks of GBV:

The integration of Artificial intelligence (AI) into digital platforms has intensified gender-based violence (GBV) in complex ways.

Disinformation and Deep Fakes:

There are two major methods of disinformation with AI use: (1) AI-generated fabricated content and (2) AI-enabled dissemination. Women and girls, along with religious and ethnic minorities are at increased risk, where violence can be incited using these powerful tools. AI technologies can lead to the creation of more realistic fake media (e.g., Deep Fake Technology), which can be misused to create content or generate images of women (to which they did not consent), as well as videos and audio recordings of women in situations or where they are represented doing or saying something they didn't do or say. The deep fake phenomenon is getting more attention in low- and middle-income countries to disseminate misinformation and target female public figures or celebrities. Pornographic images or videos of girls or women created using deepfake directly constitute GBV and are being shared and used for blackmail.³ AI tools are also being used to build "synthetic histories"-realistic false narratives. These narratives can be used to defame the individual.



Production and Hosting by KEMU

<https://doi.org/10.21649/akemu.v31i1.6079>
2079-7192/© 2025 The Author(s). Published by Annals of KEMU on behalf of King Edward Medical University Lahore, Pakistan.
This is an open access article under the CC BY4.0 license
<http://creativecommons.org/licenses/by/4.0/>

Cyber Harassment:

As technology evolves, it can be manipulated for harmful purposes—particularly against women and marginalized groups—exacerbating incidents of technology-facilitated GBV (TF GBV).⁴ The use of digital platforms enables aggressors to stalk, harass, and exploit victims with increasing impunity. Cyber harassment, hate speech, misinformation, impersonation, and defamation are among the topmost vectors of technology-facilitated GBV. A concerning estimate of 58% of young women and girls experience online harassment on social media platforms, with the majority becoming victims by 14-16 years of age.⁵ 60% of the respondents experienced at least one of 13 harms (including being physically threatened online, spied on, having unwanted images taken or being sent to them, and networked harassment) in a 2023 survey of more than 18,000 people.⁶ Automated harassment campaigns and cyber harassment templates generated by AI models directly lead to a much wider proliferation of online harassment and the misuse of data.

Reinforcement of existing structural inequalities across society by AI systems:

AI-automated exploitation online is compounded by the structural inequalities that persist across society, which AI systems often mirror and reinforce, thus perpetuating cycles of violence. In this context, it is crucial to acknowledge that the continuum of online abuse does not exist in isolation; rather, it is deeply interconnected with offline violence, warranting an urgent need for dynamic legal responses that reflect these evolving realities.⁷ The discourse surrounding technology-facilitated gender-based violence (TFGBV) must also consider AI's role in shaping harmful social dynamics, as evidenced by Uber's insufficient recognition of how its platform may be used to facilitate such violence.⁸

Data Biases:

The data trained to use AI models can potentially introduce biases or hallucinations causing unintended harm. Most AI datasets currently available are trained on “WEIRD” datasets (i.e., originating from Wealthy, Educated, Industrialized, Rich, and Democratic countries), and thus the interpretation of data can lead

to bias and inaccurate profiling of minorities and ethnic groups.⁹ Social prejudices are frequently reflected in algorithms, which can intensify discriminatory actions against women and other marginalized genders. Moreover, the lack of diversity in AI development teams limits the identification of these biases, leading to solutions that often overlook the experiences of those most affected. Addressing these issues necessitates a concerted effort to implement inclusive practices in AI development, ensuring that the technologies designed to protect and empower individuals are equitable and just.

Privacy Concerns:

Surveillance technologies like facial and biometric recognition can expose vulnerable populations to GBV, especially in humanitarian settings as anyone (including aid workers and governments) can access detailed information like age, gender, location, etc., which can be misused. Data breaches could also endanger survivors of GBV. It is crucial that AI-generated biometric data is dealt with in line with standard ethical and safety protocols.¹⁰

The Role of AI in Identifying and Preventing Gender-Based Violence

The potential for AI to serve as a transformative tool in combating gender-based violence is significant, provided that ethical frameworks are established.

Provisions of Information about GBV services:

AI-powered chatbots are being used in different countries to provide information and resources to GBV survivors. Information includes where to report as well as how to save forensic evidence, provide legal support, and make safety plans anonymously.¹¹ Innovations such as the “Fear Not” intervention demonstrate the potential for AI to promote resilience among vulnerable populations, suggesting that virtual learning can effectively impart coping mechanisms to children, thereby reducing overall victimization rates.¹² AI for Reporting and Intervention: Amid the growing prevalence of gender-based violence, the integration of artificial intelligence technologies offers promising avenues for monitoring and reporting incidents effectively.

AI Role in identifying GBV online:

AI tools can help not only in identifying GBV online but, based on people's social media activity, abusive language can also flag people at risk of perpetrating GBV offline. Online detection algorithms have been classified to identify sexism, online misogyny, and child sexual abuse among others.¹³

Facilitation of Risk Assessment of GBV:

Predictive models using machine learning algorithms have been used to identify potential threats, victims, as well as flag cases of domestic violence, intimate partner violence, and child abuse with good accuracy resulting in timely intervention. It can also help in detecting and preventing domestic violence.¹⁴ However, it is critical to consider ethical considerations; biased data can aggravate existing discrimination in vulnerable communities and further endanger marginalized populations, if not used with caution.

Implications for Policy and Future Research:

Strategic frameworks need to be developed that not only reduce the risks of AI-facilitated violence but also leverage its abilities in violence prevention. Some potential solutions are:

- Improving Literacy to comprehend the potential harms of AI and strong advocacy to address TFGBV.
- Accountability principles for data processing must be upheld, including Data protection and privacy.
- Adequate reporting policies are necessary and should be in place following proactive identification of falsified content related to GBV.
- It is imperative that individuals and networks disseminating hate speech and misinformation perpetuating GBV, must be held accountable followed by swift action.
- In order to ensure transparency, and liability of all important stakeholders a more human rights centric approach should be followed.
- Algorithms should be reviewed on a regular basis in order to remove biases related to age, gender, ethnicity, etc.
- It is important to examine and analyze different social media platform policies on an ongoing basis in order to address the current and future risks related to AI & GBV.

- Complete transparency in the use of AI systems to prevent GBV, must be in place to establish trust and improve its effectiveness in combating these modern threats.

To wrap things up, it's evident that the rise of artificial intelligence (AI) technologies is closely linked with the growing problem of gender-based violence (GBV). Social media and online platforms often serve as tools for toxic behavior due to anonymity and the spread of falsified information. Sadly enough, these harmful actions particularly affect marginalized gender groups, making the digital space feel unsafe and unfair.

It's high time that the authorities start looking at AI regulations through a gender-sensitive lens. By doing so, society can work towards creating an online environment that is not only safer but also more just and welcoming for everyone.

References

1. Gayatri, Pradhan. A sociological review of gender-based violence across the globe. *Int J Multidiscip Res.* 2024;6(5):27970. doi:10.36948/-ijfmr.2024.v06i05.27970.
2. Chowdhury R, Lakshmi D. "Your opinion doesn't matter, anyway": Exposing technology-facilitated gender-based violence in an era of generative AI. UNESCO Publishing; 2023. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000387483>.
3. Pauwels E. Artificial intelligence and data capture technologies in violence and conflict prevention. Global Centre on Cooperative Security. 2020 [accessed 2024 Dec 30]. Available from: <https://www.jstor.org/stable/resrep27551>.
4. Daniel CP. Socio-economic determinants of gender-based violence [GBV]: SDG analytics on the global GBV scenario with special reference to technology-facilitated gender-based violence and adolescent birth rates. 2023. NYMC Student Theses and Dissertations. 63. Available from: https://touro scholar.touro.edu/nymc_students_theses/63.
5. Women UN. Accelerating Efforts to Tackle Online and Technology Facilitated Violence Against Women and Girls (VAWG). Policy brief. 2022. www.unwomen.org/en/digital-library/publications/2022/10/accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls.

6. Centre for International Governance Innovation. Supporting safer digital spaces. 2023 Jun 8. Available from:<https://www.cigionline.org/publications/supporting-saferdigital-spaces/>.
7. De Silva de Alwis R. A rapidly shifting landscape: why digitized violence is the newest category of gender-based violence. *La Revue des Juristes de Sciences Po*. 2024;29(25):62. Available at SSRN: <https://ssrn.com/abstract=4648409>.
8. Turnbull, A. Onlife Harms: Uber and Sexual Violence. *CJLT*. 2022; 19:277-306.
9. Edit. AI has a WEIRD problem. [date unknown]. Available from:<https://edit.co.uk/blog/ai-has-a-weird-problem/>.
10. Gender-Based Violence Area of Responsibility Helpdesk. Gender-based violence and artificial intelligence (AI): Opportunities and risks for women and girls in humanitarian settings. 2023. Available from:<https://gbvaor.net/node/1903>.
11. UNDP. Sara: The new artificial intelligence tool to tackle gender violence in Central America. 2023. Available from:<https://www.undp.org/latin-america/press-releases/sara-new-artificial-intelligence-tool-tackle-gender-violence-central-america>.
12. Sapouna M, Wolke D, Vannini N, Watson S, Woods S, Schneider W ,et al. Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. *J Child Psychol Psychiatry*. 2010;51(1):104-12. doi: 10.1111/j.1469-7610.2009.02137.x.
13. Rodríguez DA, Díaz-Ramírez A, Miranda-Vega JE, Trujillo L, Mejia-Alvarez P. A systematic review of computer science solutions for addressing violence against women and children. *IEEE Access*. 2021;9(9):114622-39. <https://doi.org/10.1109/ACCESS.2021.3103459>.
14. Janssen J, Kokkeler B, Novitzky P. A systematic review of ethical challenges and opportunities of addressing domestic violence with AI technologies and online tools. *Heliyon*. 2023;9(6):17140. <https://doi.org/10.1016/j.heliyon.2023.e17140>