Centre
for Internet
& Society

TECHNOLOGY-FACILITATED GENDER-BASED VIOLENCE AND WOMEN'S POLITICAL PARTICIPATION IN INDIA

A Position Paper

December 2024

Technology-facilitated Gender-based Violence and Women's Political Participation in India

A Position Paper

Table of Contents

Abstract

Political participation of women is fundamental to democratic processes and the building of more equitable and just futures. Although women have traditionally struggled to access physical public spaces due to patriarchal norms and violence in the physical sphere, technology has created newer avenues for women to access the virtual public sphere. While technology has provided tools for political participation, information seeking, and mobilisation, it has also created unsafe online spaces for women, thus often limiting their ability to actively engage online.

This paper examines the emotional and technological underpinnings of gender-based violence faced by women in politics in India. We explore the ways in which gender-based violence is weaponised to diminish the political participation and influence of women in the public eye or political spaces. Through real-life examples of gendered disinformation and sexist hate speech targeting women in politics in India, we identify affective patterns in the strategies deployed to adversely impact public opinion and democratic processes. We highlight the emotional triggers that play a role in exacerbating these gendered online harms, particularly for women in public life. We also examine the critical role of technology and online platforms in this ecosystem – both in perpetuating and amplifying this violence, as well as in attempting to combat it.

We argue that it is critical to investigate and understand how existing affective structures and patriarchal hegemony operate to create unsafe access to public spheres, both online and offline, for women. We also advocate for understanding technology design and identifying tools that could actually aid in combating TFGBV. Lastly, we point to the continued need for greater accountability from platforms, to mainstream gender related harms and combat it through diversified approaches.

1. Introduction

On 22 September 2024, the United Nations General Assembly adopted the Global Digital Compact (GDC)[1] as a part of the Pact for the Future.[2] The GDC advocates the empowerment of all women and girls, encourages leadership by women, and promotes a gender perspective to "counter and eliminate all forms of violence, including sexual and gender-based violence that occurs through or is amplified by the use of technology."[3]

While widespread internet and technological access has created avenues for political participation, citizen engagement, and online information-sharing, it has also enabled various online harms, including attacks on freedom of expression, information disorder, and online violence. These challenges inevitably take on gendered dimensions, with women activists, journalists, and politicians facing a rise in targeted, emotionally charged campaigns intended to undermine their political participation and influence – a key focus of this essay.

The United Nations General Assembly Resolution 68/181 affirms that violence and abuse against women on social media platforms is a form of systematic gender discrimination against women, stating:

> *"Information-technology-related violations, abuses, discrimination and violence against women, including women human rights defenders … with a view to discrediting them and/or inciting other violations and abuses against them, are a growing concern and can be a manifestation of systemic gender-based discrimination."*[4]

Ever since, there have been attempts to arrive at a consensus-driven definition of technology-facilitated gender-based violence. During an expert group meeting in November 2022, convened primarily by UN Women and WHO, a definition was adopted by consensus:

> *"Technology-facilitated violence against women (TF VAW) is any act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political, or economic harm, or other infringements of rights and freedoms."*[5]

1. United Nations, "Global Digital Compact", accessed 7 November, 2024, https://www.un.org/global-digital-compact/en.

2. United Nations, "Pact of the Future", accessed 16 December, 2024, https://www.un.org/sites/un2.un.org/files/sotf-pact_for_the_future_adopted.pdf.

3. United Nations, "Global Digital Compact", September 2024, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf.

4. United Nations General Assembly, "Resolution Adopted by the General Assembly on 18 December 2013", 18 December 2013, https://documents.un.org/doc/undoc/gen/n13/450/31/pdf/n1345031.pdf?OpenElement.

5. UN Women, World Health Organization, "Technology-facilitated Violence against Women: Towards a common definition", November 2022, https://www.unwomen.org/sites/default/files/2023-03/Expert-Group-Meeting-report-Technology-facilitated-violence-against-women-en.pdf.

A 2020 Amnesty International report analysed over 114,000 tweets to 95 women politicians during and after the 2019 Indian elections. It found that they received nearly 1 million hateful mentions on Twitter (now X), with 1 in 5 of them being sexist or misogynistic.[6] The study also found that Indian women politicians experienced a substantially higher number of abusive attacks online (almost double) than their UK and US counterparts.[7] It is clear that while global conversations around digital rights have progressed, the issue of violence against women online continues to remain unresolved.

In this essay, we explore how gendered misinformation and disinformation, hate speech, and technology-facilitated gender-based violence (TFGBV) interact with and feed on each other; we look at manifestations of TFGBV in India against women in politics, explore the emotional roots that lead to TFGBV in certain contexts, and finally try to understand some of the technological underpinnings of TFGBV. Ahead of the 69th session of the Commission on the Status of Women (Beijing +30)[8], we focus on the conversation around TFGBV and its impact on political participation, particularly in the Global South. The last few years have been critical for discussions on internet governance processes, and 2025 will also be a critical year with the upcoming WSIS + 20 review. In these conversations, it is essential to continue mainstreaming dialogues around gender and not allow for it to be relegated to the background. We hope that some of the provocations put forward in this essay will further deliberations on the convergence of internet governance processes and gender-related processes globally, notably bringing in voices from the Global South, and lead to improved local responses to TFGBV in India.

A note on the public sphere, the private sphere and the virtual sphere

Women's access to public spaces has been a contentious issue for a long time, and we see similar manifestations in the virtual sphere. The reason we speak about this is because women in politics need to have access to the public sphere, both online and offline. We discuss and theorise the public sphere articulations in relation to women's access to show how the physical public sphere is often made inaccessible to women, which in turn is now also manifesting in the virtual public sphere.

The notion of the public sphere was first conceptualised by Habermas, who stated that the public sphere encompasses "coexistence with public authority", while the private sphere is, in the narrower sense, the realm of commodity exchange and social labor.[9] This emphasised two central themes: the physical location that creates social ties (public space) and the assemblage of attributes to the formation of public discourse[10] (Low, 2015).

6. Amnesty International India, "Troll Patrol India: Exposing Online Abuse Faced by Women Politicians in India", 2020, https://decoders.blob.core.windows.net/troll-patrol-india-findings/Amnesty_International_India_Troll_Patrol_India_Findings_2020.pdf.

7. Ibid.

8. UN Women, "CSW69 / Beijing+30 (2025)", https://www.unwomen.org/en/how-we-work/commission-on-the-status-of-women/csw69-2025.

9. J. Habermas, "The structural transformation of the public sphere" 1989, Cambridge, MA: MIT Press.

10. Low, S. (2015). "Public Space and the Public Sphere: The Legacy of Neil Smith." Antipode. Retrieved from http://dx.doi.org/10.1111/anti.12189.

In challenging the public/private divide, various feminist scholars note that the influence of the divide differs on the basis of race, class, (dis)ability, and sexual identity.[11] Susen, in his critique of Habermas' notion of the public sphere, lucidly brings up the problem of gender inequality and calls it 'gender-blind' due to its lack of mention of the "heterosexist and masculinist forms of normativity" in societies, which is a central element of the modern public sphere:

> *"The marginalization of women within the modern public sphere as well as their de facto exclusion from particular–for example, religiously defined–positions of power can be seen as an expression of the patriarchal character of society in general and as an extension of the patriarchal character of the bourgeois family in particular. To account for the normative significance of the gendered division of power in modern society requires acknowledging that modern public spheres cannot escape the ubiquity of materially and discursively constructed forms of patriarchy."[12]*

The rallying around "personal is political,"[13] signified the first attempt to break down the gendered division between the private sphere attributed to women and the public sphere traditionally reserved for men.

Papacharissi (2002) in considering the revival of the public sphere virtually, addressed how the internet could concurrently reduce or expand that potential. He further clarifies that while the internet and related technologies have created "a new public space for politically oriented conversation; whether this public space transcends to a public sphere is not up to the technology itself."[14]

Some media scholars speak about a networked public sphere as an extension of the already existing public sphere to an online platform, "a resuscitation of an ailing public sphere, and/or a first-time venture whose success has been made possible by the advent of digital network technologies."[15]

It is also worth noting that virtual sphere scholarship has produced critiques of the public sphere, which Benkler (2008) placed into the following categories:

> *"Information overload leading to political discourse polarization and fragmentation, the internet not being as open as it is deemed to be with a few sites monopolizing the web, and the digital divide which points out the question of access in the political discourse."[16]*

We consider this theoretical framing important to any conversation on women in politics and their access to political and public life across both online and offline contexts.

11. S. Boyd. "Challenging the Public/Private Divide: Feminism, Law, and Public Policy," Toronto, ON: University of Toronto Press. 1997.

12. Susen, S. (2011). "Critical Notes on Habermas's Theory of the Public Sphere." Sociological Analysis, 5, 37-62.

13. Theresa Man Ling Lee. "Rethinking the Personal and the Political: Feminist Activism and Civic Engagement." Hypatia 22, no. 4 (2007): 163–79. http://www.jstor.org/stable/4640110.

14. Papacharissi, Z. (2002). "The virtual sphere: The internet as a public sphere." New Media & Society, 4, 9-27.

15. Goldberg, G. (2011). Rethinking the public/virtual sphere: The problem with participation. New Media & Society, 13, 739-754.

16. Benkler, Y. (2006). The Wealth of Networks: how social production transforms markets and freedom. New Haven, CT: Yale University Press.

2. The interplay of gendered disinformation, sexist hate speech, and TFGBV

Claire Wardle and Hossein Derakhshan categorise information disorder as disinformation, misinformation, and malinformation.[17] Disinformation refers to the intentional creation and spread of falsehoods; misinformation involves the unintentional sharing of inaccuracies; and malinformation refers to the malicious use of factual information. Gendered disinformation specifically targets women and marginalised genders, aiming to cause harm through manipulated narratives. Demos refers to this as "information activities" that undermine individuals based on gender for political or social gain.[18]

Hate speech, though not universally defined, generally includes communications that attack individuals based on inherent traits, such as race or gender, to incite discrimination or violence. The United Nations defines it as any communication that disparages individuals based on their identity (e.g., race, gender, sexuality, or caste).[19] The Law Commission of India adds a dimension of intentionality to this definition – that is, the aim to instil fear or provoke violence.[20] Sexist or gendered hate speech is a manifestation of sexism and involves the expression of contempt based on gender and the promotion of harmful stereotypes.[21]

Cynthia Khoo introduced the term "technology-facilitated gender-based violence" to describe misogyny and abuse against women (including sexist hate speech and gendered disinformation) facilitated by technology.[22] TFGBV encompasses a spectrum of harmful behaviours, such as stalking and defamation, often amplified by technology.[23] The United Nations Population Fund (UNFPA) describes it as violence perpetrated through digital means against individuals based on gender.[24]

17. Claire Wardle and Hossein Derakhshan, "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making", Council of Europe, 27 September 2017, https://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf.

18. Ellen Judson, Asli Atay, Alex Krasodomski-Jones, Rose Lasko-Skinner, and Josh Smith, "Engendering Hate: The Contours of State-Aligned Gendered Disinformation Online", National Democratic Institute, October 2020, https://dera.ioe.ac.uk/id/eprint/36786/1/Engendering-Hate-Report-FINAL.pdf.

19. United Nations, "What is Hate Speech?", accessed November 7, 2024. https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech.

20. Law Commission of India, "Hate Speech", Report No. 267, 23 March 2017, https://cdnbbsr.s3waas.gov.in/s3ca0daec69b5adc880fb464895726dbdf/uploads/2022/08/2022081654-1.pdf.

21. Council of Europe, "Online Resources: Combating Sexist Hate Speech (2016)", accessed 7 November 2024, https://edoc.coe.int/en/gender-equality/6995-combating-sexist-hate-speech.html.

22. Cynthia Khoo, "Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence", Women's Legal Education and Action Fund, 2021, https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf.

23. United Nations Population Fund, "Preventing Technology-Facilitated Gender-Based Violence (TF GBV)", https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/GDC-Submission_UNFPA.pdf.

24. Ibid.

TFGBV, alongside its components – sexist hate speech and gendered disinformation – represents a multifaceted threat aimed at dehumanising individuals based on gender.

While sexist hate speech focuses on derogatory communication targeting a person's sex or gender, gendered disinformation deliberately seeks to spread false or harmful narratives about women. TFGBV includes all forms of online violence and harassment that leverage technology to inflict harm on women and marginalised genders.

In this essay, we will explore how some of these phenomena have played out in the Indian context and some of the emotional and technological tactics used to propagate gendered hate speech, TFGBV, and gendered disinformation.

3. The emotional triggers behind TFGBV in Indian politics

Affect and emotions play a crucial role in the effectiveness of TFGBV, influencing how we perceive, interpret and respond to information. While "affect" and "emotion" are somewhat distinct terms, they are often used interchangeably.[25] In modern psychological usage, "affect" refers to the *"mental counterpart of internal bodily representations associated with emotions, actions that involve some degree of motivation, intensity, and force, or even personality dispositions."* More generally, affect can been used to mean anything emotional.[26]

Emotionally charged content is more likely to be shared and remembered, making it a powerful tool for those seeking to influence public opinion.[27]

This emotional manipulation can be particularly potent in India's political environment, where politics is deeply rooted in cultural and religious sentiments.[28,29] This manifests in hierarchies of power that result in systemic discrimination against women and marginalised voices in political spaces.[30] In this section, we look at how affect, cultural attitudes and gender norms plays out in gendered contexts online, and how this further exacerbates harms for women in public life.

25. Murray Alpert and Anna Rosen, "A semantic analysis of the various ways that the terms "affect," "emotion," and "mood" are used", 23 Journal of Communication Disorders, 237-246 (1990), https://www.sciencedirect.com/science/article/pii/002199249090002G.

26. Lisa Feldman Barrett and Eliza Bliss-Moreau, "Affect as a Psychological Primitive", 41 Advances in Experimental Social Psychology, 167–218 (2009), 10.1016/S0065-2601(08)00404-8.

27. Marco Guerini and Jacopo Staiano, "Deep Feelings: A Massive Cross-Lingual Study on the Relation between Emotions and Virality", 16 March 2015, https://arxiv.org/abs/1503.04723.

28. Sharu Goel, Sandeep Kumar Pandey, and Hanumant Singh Shekhawat, "Analysis of Emotional Content in Indian Political Speeches", in International Conference on Intelligent Human Computer Interaction, eds. Jong-Hoon Kim, Madhusudan Singh, Javed Khan, Uma Shanker Tiwary, Marigankar Sur, and Dhananjay Singh (Springer, 2022), 177–185, https://link.springer.com/chapter/10.1007/978-3-030-68449-5_18.

29. Punyajoy Saha, Binny Mathew, Kiran Garimella, and Animesh Mukherjee, "'Short Is the Road That Leads from Fear to Hate': Fear Speech in Indian WhatsApp Groups", in WWW '21: Proceedings of the Web Conference 2021, eds. Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Association for Computing Machinery, 3 June 2021), 1110–1121, https://dl.acm.org/doi/10.1145/3442381.3450137.

30. Ranjana Kumari, "Patriarchal Politics: The Struggle for Genuine Democracy in Contemporary India", Heinrich Böll Foundation, 26 February 2014, https://www.boell.de/en/2014/02/26/patriarchal-politics-struggle-genuine-democracy-contemporary-india.

3.1 Gendered dimensions of affect

Women in politics frequently face targeted violence and harassment designed to undermine their authority and discourage their participation in public life. The United States Agency for International Development identifies violence against women in politics (or VAWPP) as a deliberate strategy aimed at inhibiting women's engagement in the public sphere that can ultimately lead to the silencing of women's voices.[31]

Cultural attitudes towards gender play a crucial role in shaping women's political experiences. Traditional gender norms often dictate that women should occupy subordinate roles and reinforce social hierarchies that marginalise women's contributions.[32,33]

TFGBV in Indian politics is characterised by its emotional intensity and cultural specificity.[34,35]

It exploits traditional gender norms and stereotypes and uses emotionally charged narratives to attack and discredit women in politics. This can range from false claims about a candidate's personal or professional life[36] to manipulated images or videos[37,38] designed to provoke visceral, emotional responses from voters.

31. DRG Learning Digest, "Combatting Technology-Facilitated Gender-Based Violence in Politics", 7 March 2023, https://content.govdelivery.com/accounts/USAIDHQ/bulletins/34c7e57.

32. Mona Lena Krook, "Violence Against Women in Politics", Journal of Democracy 28, no. 1 (2017): 74–88, https://muse.jhu.edu/article/645539.

33. Madeleine Arnot and Sharlene Swartz, "Gender Respect: Empirical Insights for (Moral) Educators about Women's Struggles for Respect in the Global South", Journal of Moral Education 47, no. 4 (2018): 481–497, https://www.tandfonline.com/doi/full/10.1080/03057240.2018.1433644.

34. Proma Ray Chaudhury, "The Political Asceticism of Mamata Banerjee: Female Populist Leadership in Contemporary India", Politics & Gender 18, no. 4 (2022): 942–977, https://www.cambridge.org/core/journals/politics-and-gender/article/political-asceticism-of-mamata-banerjee-female-populist-leadership-in-contemporary-india/618F669661A8387B2B8DFE39C43EF0F1.

35. Tarini Bedi, "Motherhood, and Its 'Lack': Personal Loss and Political Community among Shiv Sena Women in Mumbai", Women's Studies International Forum 35, no. 6 (2012): 478–486, https://www.sciencedirect.com/science/article/abs/pii/S0277539512001318?via%3Dihub.

36. Proma Ray Chaudhury, "The Political Asceticism of Mamata Banerjee: Female Populist Leadership in Contemporary India", Politics & Gender 18, no. 4 (2022): 942–977, https://www.cambridge.org/core/journals/politics-and-gender/article/political-asceticism-of-mamata-banerjee-female-populist-leadership-in-contemporary-india/618F669661A8387B2B8DFE39C43EF0F1.

37. Rakesh Vuppu, "Photoshopped Image of Sonia Gandhi and Maldives Ex-President Maumoon Abdul Gayoom Is Being Circulated with False Claims", Factly, 2 May 2019, https://factly.in/photoshopped-image-of-sonia-gandhi-and-maldives-ex-president-maumoon-abdul-gayoom-is-being-circulated-with-false-claims/.

38. Prafulla Das, "Biju Janata Dal's Chandrani Murmu, Youngest MP in India, Alleges Harassment by a Media House in Odisha and Seeks Justice", Frontline, 18 October 2020, https://frontline.thehindu.com/dispatches/biju-janata-dals-chandrani-murmu-youngest-mp-in-india-alleges-harassment-by-a-media-house-in-odisha-and-seeks-justice/article32885611.ece.

The impact of such abuse extends beyond the targeted individual and discourages women from entering political spaces altogether.[39] This creates a cyclical problem where the under-representation of women in Indian politics is entrenched by an increasingly hostile online environment.[40]

Gendered hate campaigns operate by exploiting specific emotional triggers. For instance, portraying a female politician as morally corrupt and a threat to established cultural norms incites fear and anger, especially in certain culturally conservative contexts. These strategies are not unique to India; women in politics in other countries have also faced similar online misogyny and abuse targeting their morality, bodies, sexuality, or credibility. These hate campaigns are not restricted to women politicians with particular political leanings and span the entirety of the political spectrum[41], leveraging cultural norms to portray women in politics as threats to societal values.[42]

AVERAGE LIKES   AVERAGE SAVES

39. Dhanaraj Thakur and DeVan L. Hankerson Madrigal, "An Unrepresentative Democracy: How Disinformation and Online Abuse Hinder Women of Color Political Candidates in the United States", Center for Democracy & Technology, 27 October 2022, https://osf.io/preprints/osf/bwta3.

40. Drupa Dinnie Charles, Azhagu Meena, Simiran Lalvani, Syeda Zainab Akbar, Divya Siddharth, and Joyojeet Pal, "Performing Gender, Doing Politics: Social Media and Women Election Workers in Kerala and Tamil Nadu", in ICTD '20: Proceedings of the 2020 International Conference on Information and Communication Technologies and Development (Association for Computing Machinery, 17 June 2020), 1–11, https://dl.acm.org/doi/10.1145/3392561.3394648.

41. Lucina Di Meco and Saskia Brechenmacher, "Tackling Online Abuse and Disinformation Targeting Women in Politics", 30 November 2020, Carnegie Endowment for International Peace, https://carnegieendowment.org/research/2020/11/tackling-online-abuse-and-disinformation-targeting-women-in-politics?lang=en.

42. Ibid.

3.2 Moral disgust as an emotional trigger for TFGBV

Despite the notable increase in women's participation in public life and appointment to positions of power, research indicates that leadership is still widely considered to be a 'masculine' pursuit that women are innately unsuited for.[43]

Gender-based abuse is therefore fundamentally driven by deeply embedded moral emotions that emerge when traditional gender hierarchies are challenged. When women politicians demonstrate assertive leadership qualities – particularly in digital spaces – they trigger a complex web of negative moral emotions, including contempt, disdain, disgust, and moral revulsion[44] among specific segments of the population who view them as violating established gender norms.

Disdain could be expressed by treating women leaders as inherently less capable or less serious political actors, and it is often expressed through infantilising language (e.g., calling them a political puppet) or by reducing women to their familial roles (e.g., someone's wife or daughter)[45]. The intensity of this backlash appears to be amplified when women in politics demonstrate traditionally masculine-coded behaviours, such as aggressively questioning or directly confronting political opponents on public online or offline platforms.[46]

Research indicates that this backlash does not emerge from merely political disagreement but rather stems from a visceral response of moral disgust.[47] This response of disgust is expressed through specific forms of TFGBV that includE gendered slurs, sexualised threats, and coordinated harassment campaigns that specifically target the politician's gender identity rather than her political views. These findings align with broader research on moral evaluation processes and disgust sensitivity. Just as physical disgust can amplify moral judgments, perceived violations of gender norms by women politicians may trigger moral disgust, which fuels abuse.[48]

43. Anne M. Koenig, Alice H. Eagly, Abigail A. Mitchell, and Tiina Ristikari, "Are Leader Stereotypes Masculine? A Meta-analysis of Three Research Paradigms", Psychological Bulletin 137, no. 4, (2011): 616–642, https://psycnet.apa.org/buy/2011-11069-001.

44. Victoria L. Brescoll, Tyler G. Okimoto, and Andrea C. Vial, "You've Come a Long Way … Maybe: How Moral Emotions Trigger Backlash against Women Leaders", Social Issues 74, no. 1 (2018): 144–164, https://spssi.onlinelibrary.wiley.com/doi/10.1111/josi.12261https://spssi.onlinelibrary.wiley.com/doi/10.1111/josi.12261.

45. Kalpana Sharma, "The Paradox of India's Woman Politician", The Hindu, updated 1 October 2024, https://www.thehindu.com/society/india-woman-politician-balance-ambitions-identity-expectations-mamata-banerjee-atishi-vinesh-phogat/article68676929.ece.

46. Peter Beinart, "Fear of a Female President", The Atlantic (October 2016), https://www.theatlantic.com/magazine/archive/2016/10/fear-of-a-female-president/497564/.

47. Victoria L. Brescoll, Tyler G. Okimoto, and Andrea C. Vial, "You've Come a Long Way … Maybe: How Moral Emotions Trigger Backlash against Women Leaders", Social Issues 74, no. 1 (2018): 144–164, Brescoll et al., "You've Come a Long Way", 144–164..

48. Ibid.

4. TFGBV and women in politics in India

Note: This section contains specific examples of TFGBV that may be distressing to some. Please move ahead to the next chapter if you wish to avoid this.

Given the sensitive and often graphic nature of the material we identified, we have approached its documentation with care. The images and analysis that we have chosen to share are with the intent to communicate important insights while avoiding any needlessly gratuitous or sensationalistic content.

The emotional toll of TFGBV on women in politics is significant, potentially affecting their campaign strategies and public appearances and leading to harassment fatigue[49], all of which further continues to tilt the playing field against them.

Gendered narratives frequently focus on personal attributes, physical appearance, and the private lives of women while ignoring their qualifications and achievements. This is in stark contrast to the treatment of male politicians, who are more often criticised for their policies and intellect rather than their gender.[50,51]

Character assassination[52] is usually the first line of attack on women in the public eye, turning them into targets for a disproportionate amount of sexualised harassment. This now includes the use of deepfake technologies to morph their images to create pornographic content.[53,54]

49. Nitish Goyal, Leslie Park, and Lucy Vasserman, "'You Have to Prove the Threat Is Real': Understanding the Needs of Female Journalists and Activists to Document and Report Online Harassment", in CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–17, https://arxiv.org/pdf/2202.11168v1.

50. Daphne Joanna Van der Pas and Loes Aaldering, "Gender Differences in Political Media Coverage: A Meta-Analysis", Journal of Communication 70, no. 1 (2020), 114–143, https://academic.oup.com/joc/article/70/1/114/5761879.

51. Sharon Mavin, Patricia Bryans, and Rosie Cunningham, "Fed-up with Blair's Babes, Gordon's Gals, Cameron's Cuties, Nick's Nymphets: Challenging Gendered Media Representations of Women Political Leaders", Gender in Management 25, no. 7, https://www.emerald.com/insight/content/doi/10.1108/17542411011081365/full/html.

52. Yudhajit Shankar Das, "Politics and Sexism: What Have Parties Really Done on Women Empowerment?", India Today, updated 22 September 2023, https://www.indiatoday.in/india/story/womens-reservation-bill-smriti-irani-politics-character-assassination-2438478-2023-09-22.

53. Suzie Dunn, "Women, Not Politicians, Are Targeted Most Often by Deepfake Videos", Centre for International Governance Innovation, 3 March 2021, https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/.

54. Sophie Maddocks, "'A Deepfake Porn Plot Intended to Silence Me': Exploring Continuities between Pornographic and 'Political' Deep Fakes", Porn Studies 7, no. 4 (2020): 415–423, https://www.tandfonline.com/doi/full/10.1080/23268743.2020.1757499.

In this section, we look at a few instances of women in politics in India who have experienced various forms of TFGBV. Our methodology to identify these case studies began with a review of existing news coverage, which was then supported by primary research of user-generated content across major social media platforms, including Instagram, X, YouTube, and Reddit. Although we have attempted to make our case studies as diverse as possible, language constraints limited our investigation of regional contexts to content in English, Hindi and Marathi.

Gender-based attacks are not uniform, and vary significantly based on intersecting identities such as caste, religion, language, disability, and sexual orientation.

Women from marginalised backgrounds – such as those belonging to lower castes or religious minorities – experience heightened levels of abuse and discrimination. This was confirmed by an Amnesty International investigation which revealed that Muslim women politicians and those from marginalised castes experienced disproportionately higher levels of online abuse compared with their peers.[55]

Chandrani Murmu, a first-time politician from the Biju Janata Dal, is from a reserved tribal community in Odisha and the youngest member of Parliament.[56] When she was appointed as the BJD nominee for a district from which she subsequently won during the 2019 Lok Sabha election, a media channel owned by former BJD member reportedly circulated a morphed pornographic video of her across social media.[57]

Sonia Gandhi, the longest-serving president of the Indian National Congress, has faced decades of gendered disinformation and hate speech weaponising her Italian identity and claiming that her moral character was incongruent with Indian cultural norms. For example, her face was morphed into an image of James Bond actress Ursula Andress in a bikini[58] and a culturally iconic image of Marilyn Monroe in a white dress (with the caption claiming that she had been a "bar dancer" and sex worker before she married Rajiv Gandhi)[59]. A digitally altered picture of her sitting on the lap of the former Maldives president, Maumoon Abdul Gayoom, was also widely shared.[60]

55. Amnesty International India, "Troll Patrol India: Exposing Online Abuse Faced by Women Politicians in India", 2020, https://decoders.blob.core.windows.net/troll-patrol-india-findings/Amnesty_International_India_Troll_Patrol_India_Findings_2020.pdf.

56. Indo-Asian News Service, "Meet 25-year-old Chandrani Murmu, the youngest Member of Parliament", India Today, May 26, 2019, https://www.indiatoday.in/education-today/news/story/chandrani-murmu-youngest-member-of-parliament-engineering-graduate-1535040-2019-05-26.

57. Prafulla Das, "Biju Janata Dal's Chandrani Murmu, youngest MP in India, alleges harassment by a media house in Odisha and seeks justice", Frontline, October 18, 2020, https://frontline.thehindu.com/dispatches/biju-janata-dals-chandrani-murmu-youngest-mp-in-india-alleges-harassment-by-a-media-house-in-odisha-and-seeks-justice/article32885611.ece.

58. Pooja Chaudhuri, "Fact Check: Images of James Bond Actress Ursula Andress Being Passed off as Photos of Sonia Gandhi", Scroll, 25 March 2019, https://scroll.in/article/917822/fact-check-images-of-james-bond-actress-ursula-andress-being-passed-off-as-photos-of-sonia-gandhi.

59. Saurabh Jaiswal Hindu, Facebook, 23 December 2015, https://www.facebook.com/sankat.mochan.376/posts/554837954673427.

60. Rakesh Vuppu, "Photoshopped Image of Sonia Gandhi and Maldives Ex-President Maumoon Abdul Gayoom Is Being Circulated with False Claims", Factly, 2 May 2019, https://factly.in/photoshopped-image-of-sonia-gandhi-and-maldives-ex-president-maumoon-abdul-gayoom-is-being-circulated-with-false-claims/.

Mayawati is a Dalit leader who broke caste and gender barriers to rise through the ranks of the Bahujan Samaj Party and become chief minister of Uttar Pradesh, India's most populous state, four times.[61] When she joined Twitter in 2019, she was met with a torrent of casteist and misogynist abuse. These spanned remarks on her skin colour, her perceived lack of education (despite her law degree), and jokes about reservation[62] (see Image 1).

Image 1: A post mocking Mayawati's physical appearance and caste identity (from *X*)

Atishi, a member of the Aam Aadmi Party, has faced numerous attacks that target her sexuality, including accusations of promiscuity and claims that she engages in sexual favours.[63] During the 2019 general elections, Atishi broke down at a press conference, accusing her political rival, Gautam Gambhir, of circulating a pamphlet containing "obscene and derogatory" remarks about her. This incident illustrates the intimidation tactics used against women in politics and their emotional toll. The pamphlets, titled "Atishi Marlena – Know Your Candidate", were allegedly distributed by placing them inside newspapers throughout East Delhi, and were filled with allegations of sexual impropriety and attacks on her character.[64] This content was also widely circulated on social media, with counter-allegations that Atishi had exploited these personal attacks for her own political gain (see Image 2).

Image 2: When Atishi broke down in tears during a press conference citing the personal attacks on her in the pamphlet, she was accused of making false accusations to victimise her male opposing candidate (from X)

61. Anand Mishra, "Mayawati: An Icon in Retreat", Frontline, 16 November 2023, https://frontline.thehindu.com/politics/mayawati-profile-an-icon-in-retreat-dalit-politics-darling-of-marginalised-masses-now-struggling-to-remain-relevant/article67489133.ece.

62. Dilip Mandal, "Who Trolls Mayawati on Twitter and What It Says about Indians", The Print, 12 February 2019, https://theprint.in/opinion/who-trolls-mayawati-on-twitter-and-what-it-says-about-the-indian-society/191196/.

63. Ayush Tiwari, "On the Trail of the Mysterious 'Atishi Pamphlet': Here's What We Know So Far", Newslaundry, 10 May 2019, https://www.newslaundry.com/2019/05/10/atishi-aam-aadmi-party-pamphlet-gautam-gambhir-elections-2019.

64. Ibid.

Mahua Moitra, a member of parliament from the All India Trinamool Congress, is also heavily scrutinised for her fashion choices and her looks, and is the target of comments that often veer towards sexualisation[65] (see Image 3).

Image 3: Mahua Moitra is frequently the target of comments sexualising her or referring to her physical attributes (from *X*)

Smriti Irani, a former union minister and high-ranking member of the ruling BJP, has been scrutinised for her educational qualifications[66], political controversies[67], and policy decisions[68]. These attacks are amplified because of her previous career as an actress and model (Image 4). She is now subject to comments on her body size whenever she posts on social media (Images 5).
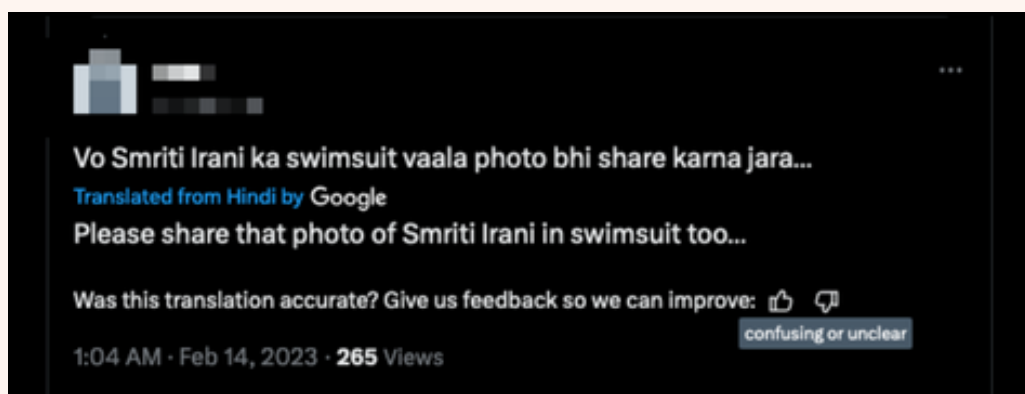

Image 4: This comment references a widely circulated picture of Smriti Irani wearing a swimsuit at a beauty pageant (from *X*)

65. Chanchal, "'This Is Also Louis Vuitton': Mahua Moitra's Dig on 'Fashion Sense' Tweet", Mint, updated 9 August 2022, https://www.livemint.com/news/this-is-also-louis-vuitton-mahua-moitra-dig-on-fashion-sense-tweet-handbag-11660041524072.html.

66. Gaurav Vivek Bhatnagar, "BA to No Degree, Arts to Commerce: The Shifting Story of Smriti Irani's Education", The Wire, 12 April 2019, https://thewire.in/politics/smriti-irani-affidavit-education-qualification.

67. Rohini Mohan, "Smriti Irani's Rise as a Case Study of Patriarchy and Sexism in the BJP", The Caravan, 30 April 2019, https://caravanmagazine.in/politics/smriti-iranis-rise-as-a-case-study-of-patriarchy.
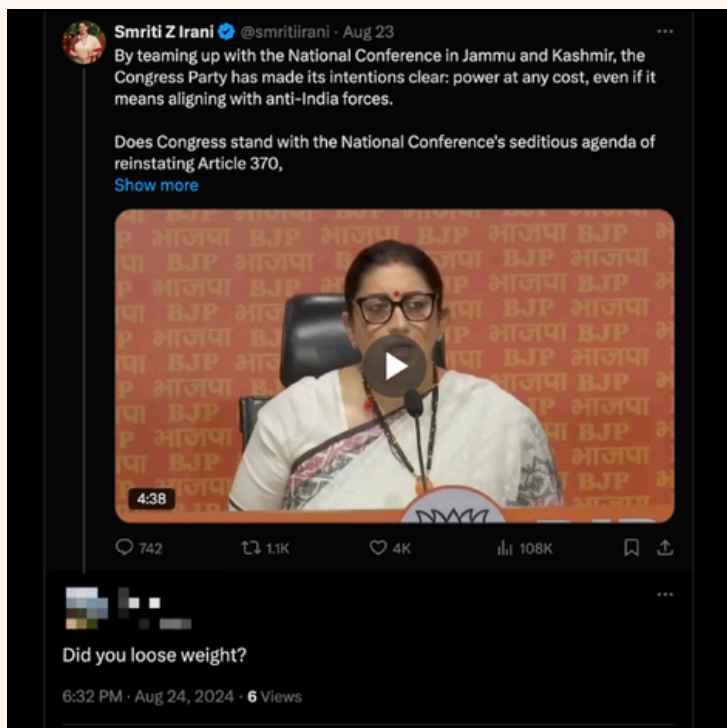
68. Ibid.

Image 5: Smriti Irani is subjected to comments about her body when she posts on official matters (from *X*)

Kangana Ranaut, a member of the BJP, often faces attacks targeting her personal life, sexuality, and past as an actress, which are used to undermine her capabilities as a political leader. She has frequently been labelled a "drama queen" and "attention seeker" in media coverage and by political opponents[69], which diminishes her credibility as a serious political leader by focusing on her personality rather than political achievements. Additionally, during her recent election campaign, memes referencing her alleged past relationship with actor Hrithik Roshan were circulated on social media whenever she was in the news.[70] She is routinely subject to derogatory sexual references due to her career as an actress, including by other politicians and opposition parties.[71]

69. Ketan Narottam Tanna, "When the Drama's So Good, You Just Have to Hit Replay!", The Free Press Journal, 26 September 2024, https://www.freepressjournal.in/analysis/when-the-dramas-so-good-you-just-have-to-hit-replay.

70. Isha Sharma, X, 20 May 2024, https://x.com/ishassharma123/status/1792486434106208679.

71. Abhishek De, "How Kangana Ranaut Row Is Tip of Sexism Iceberg", India Today, updated 27 March 2024, https://www.indiatoday.in/india/story/kangana-ranaut-mandi-row-sexist-remarks-women-politicians-misogyny-lok-sabha-election-season-2519807-2024-03-27.

Women who are related to political actors (as their wives or partners, mothers, sisters or daughters) are not spared from politically-driven, misogynistic scrutiny and abuse.

Amruta Fadnavis, the wife of 3-time Maharashtra Chief Minister Devendra Fadnavis, has had a decades-long career as a banker, in addition to an active music career.[72] Despite her professional achievements, she is a frequent target of gendered trolling and abuse[73], including for her fashion choices[74] and constant rumours about cosmetic surgery[75] – all of which are considered by many to be inappropriate given her husband's position. This is seemingly exacerbated by the fact that she continues to be a prominent public figure[76] who seemingly ignores such trolling and continues to actively participate in her husband's political campaigns on her own terms.[77]

Tabassum Rao, the wife of Karnataka minister Dinesh Gundu Rao, pointed out that she is routinely subjected to derogatory and communal comments because of her relationship to a politician and on the basis of her faith. "As someone not involved in politics, I find it reprehensible that my personal life and community are being dragged into the political arena. It is unacceptable to target family members of politicians simply because of their relationship."[78]

72. NDTV News Desk, "Meet Amruta Fadnavis - Banker, Singer, Wife Of Devendra Fadnavis", NDTV, 6 December 2024, https://www.ndtv.com/india-news/meet-amruta-fadnavis-banker-singer-wife-of-devendra-fadnavis-7180399.

73. Vijay Kumar Yadav, "Thane woman arrested for abusing Dy CM's wife Amruta Fadnavis on Facebook", Hindustan Times, September 14, 2022, https://www.hindustantimes.com/cities/mumbai-news/thane-woman-arrested-for-abusing-dy-cm-s-wife-amruta-fadnavis-on-facebook-101663096838763.html.

74. Pinkvilla, "Did Amruta Fadnavis change her style due to the scrutiny she faced on the internet about her dressing style?", Facebook, 15 February 2024, https://www.facebook.com/pinkvillamedia/videos/did-amruta-fadnavis-change-her-style-due-to-the-scrutiny-she-faced-on-the-intern/935107838133553/.

75. My Mahanagar Manini, "Amruta fadnavis before surgery | लग्नापूर्वी पार्लरलापण नाही गेले ,नंतर सर्जरी केली", Youtube, 29 November 2024, https://www.youtube.com/shorts/pK87TvzCFcw.

76. "Amruta Fadnavis", accessed 17 December 2024, https://www.amrutafadnavis.in/.

77. News18 Lokmat, "Amruta Fadnavis Dance In Rally: अमृता फडणवीसांचा भर रॅलीत डान्स", Youtube, 15 December 2024, "Amruta Fadnavis", accessed 17 December 2024, https://www.amrutafadnavis.in/. .

78. The Hindu Bureau, "Tabassum Rao, wife of Dinesh Gundu Rao, complains to women's commission against BJP for 'continuous derogatory and communal references' against her", The Hindu, 6 October 2024, https://www.thehindu.com/news/national/karnataka/tabassum-rao-wife-of-dinesh-gundu-rao-complains-to-womens-commission-against-bjp-for-continuous-derogatory-and-communal-references-against-her/article68719249.ece.

5. Technological underpinnings of TFGBV

Expanded access to the internet, coupled with the anonymity afforded by online spaces, has coincided with an increase in information disorder, hate speech, and violence online. Popular mediums such as social media and messaging platforms are commonly used to perpetrate TFGBV. Advanced technologies, including artificial intelligence (AI), can help multiply these abuses by enabling perpetrators to remain anonymous. Technology plays a dual role in the context of TFGBV, both amplifying and mitigating its effects. In the following subsections, we look at what are some of the ways in which technologies are being spread to perpetrate gender-based violence (5.1) and also look at some ways in which technology is being used to find potential solutions, mitigation tactics for such attacks (5.2). While this essay focuses on women in politics, these levers and responses, in most instances, apply to women and gender and sexual minorities as well.

While this essay focuses on women in politics, these levers and responses, in most instances, apply to women and gender and sexual minorities as well.

5.1 Technological levers for TFGBV

Technology has intensified and fundamentally restructured gendered harassment and abuse into distinctly digital manifestations with their own recognisable and replicable patterns (such as coordinated harassment campaigns, algorithmically-driven pile-on effects, the rapid spread of targeted disinformation, and other anonymised, scalable forms of intimidation).[79] Social media platforms and messaging applications have become sites of violence against communities that are marginalised for their gender, sexuality, religion, caste, and culture. The exponential growth of AI has further exacerbated this issue, making the creation and amplification of false or harmful content more effortless than ever, with women becoming the primary victims of these technologically-facilitated harms.[80,81]

79. Gen Pol, "When Technology Meets Misogyny: Multi-level, Intersectional Solutions to Digital Gender-Based Violence", 2019, https://gen-pol.org/wp-content/uploads/2019/11/When-Technology-Meets-Misogyny-GenPol-Policy-Paper-2.pdf.

80. Deepak P. Navya Sahadevan, "Gender Bias in Fake News: An Analysis" (paper presented at the Integrity in Social Networks and Media 2023 workshop, 24 September 2022), https://arxiv.org/abs/2209.11984.

81. Laura Hinson, Jennifer Mueller, Lila O'Brien-Milne, and Naome Wandera, "Technology-Facilitated Gender-Based Violence: What Is It, and How Do We Measure It?", International Center for Research on Women, 2018, https://www.svri.org/sites/default/files/attachments/2018-07-24/ICRW_TFGBVMarketing_Brief_v8-Web.pdf.

Engagement-based ranking: A critical factor in the spread of TFGBV is the fundamental architecture of online platforms and the growing shift in their design toward monetising their user bases by trading on user data and engagement.[82]

Engagement-based ranking algorithms, a common feature across platforms, reward virality and hashtags with higher engagement, even though popularity is no indication of the truth or quality of content.

Engaging with a certain kind of content results in recommendations for similar content – which often includes even more misleading or abusive content that may sometimes violate the platform's own terms of service). Personalised search results also recommend similar content, which may be harmful.[83] As a result, emotionally charged, sensational, and inflammatory content targeting women in politics (rather than positive content or attempts to fact-check disinformation) often receive disproportionately high visibility, amplifying their reach and impact.

Engagement-based ranking strategies vary across digital platforms. As a social media platform, Instagram prioritises user engagement and curated content and deploys distinct algorithms for different sections of their app – Feed, Stories, Reels.[84] Particularly, Feed balances content from both followed and recommended accounts by predicting user engagement through factors like interaction history, post popularity, and individual preferences.[85] While the role of engagement-based ranking in their search function is less clear, Jeff Allen surmises that it similarly prioritises engagement by identifying posts containing the search term and estimating the likelihood of user interaction, such as likes or comments, while assessing the probability that the content violates community guidelines. Its final ranking is then determined by weighing its engagement potential against its likelihood of policy violations, often leading to the promotion of content that may be harmful or of low quality due to its high engagement rates.

82. "3 Social Media Stocks Monetizing Their Massive User Bases", Look First/Then Leap, 29 August 2024, https://www.tradingview.com/news/stocknews:1866f629f094b:0-3-social-media-stocks-monetizing-their-massive-user-bases/.

83. Cécile Simmons and Zoé Fourel, "Hate in Plain Sight: Abuse Targeting Women ahead of the 2022 Midterm Elections on TikTok and Instagram", Institute for Strategic Dialogue, 2022, https://www.isdglobal.org/wp-content/uploads/2022/12/Hate-in-Plain-Sight.pdf.

84. Adam Mosseri, "Instagram Ranking Explained", 31 May 2023, https://about.instagram.com/blog/announcements/instagram-ranking-explained/.

85. Ibid.

In contrast, Google Search claims to include both quality and relevance in its ranking system via established quality signals such as PageRank, which measures the authority of web pages, to then prioritise high-quality, relevant content that aligns with user queries.[86] However, its rollout of "AI Overviews" (an AI-powered search feature that generates short summaries and links at the top of the search result page) for all searches on its platform earlier this year seemingly negated the existence of any such safeguards by generating results that were often completely untrue, due to its inability to verify the credibility of its sources.[87]

A more recent example of this platform-specific content distribution is on X, which has confirmed that it implements a deliberate strategy of deprioritising external link sharing. By promoting native content and discouraging what CEO Elon Musk terms "lazy linking," the platform prioritises internal user engagement.[88] However, by making it more challenging to access external content and diverse information sources, this approach could potentially increase user reliance on platform-generated content that enjoys higher engagement regardless of whether it is biased or unverified.

Content moderation gaps: Content moderation requires a strategy extending beyond simply boolean logic, necessitating a multi-layered approach that accounts for linguistic and cultural complexities. Even when limited to English-language content, detecting toxic content requires sophisticated analysis combining text and media classifiers with nuanced account signals to build comprehensive "bad actor" profiles.[89]

Gaps in content moderation persist, particularly for non-English languages where content in regional dialects is poorly moderated, and cultural nuances are missed.

In the Indian context, the inability to detect coded language, gender-, religious- or caste-based hate speech, and the use of more than 19,500 mother tongues within India's 121 regional languages[90] adds to this complexity, as many moderators lack training in recognising coded threats and the dynamics of intersectional harassment.[91]

86. Jeff Allen, "Why Is Instagram Search More Harmful Than Google Search? On Instagram's Decisions to Disable Search for Sensitive Topics and the Instagram We Can't Have", Integrity Institute, February 13, 2024. https://integrityinstitute.org/blog/why-is-instagram-search-more-harmful-than-google-search.

87. Rhiannon Williams, "Google's new AI search feature is a mess. So why is it telling us to eat rocks and gluey pizza, and can it be fixed?", MIT Technology Review", May 31, 2024. https://www.technologyreview.com/2024/05/31/1093019/why-are-googles-ai-overviews-results-so-bad/.

88. Io Dodds, "Elon Musk confirms that he's limiting people's ability to share outside news on X", The Independent, November 26, 2024. https://www.independent.co.uk/tech/elon-musk-x-news-links-b2653614.html.

89. Amit Sheth, Valerie L. Shalin and Ugur Kursuncu, "Defining and detecting toxicity on social media: context and knowledge are key", 490 Neurocomputing, 312-318 (2022), https://doi.org/10.1016/j.neucom.2021.11.095.

90. Language (Paper 1 of 2018)-Census of India 2011 (June 25, 2018), https://censusindia.gov.in/nada/index.php/catalog/42458.

91. Bertie Vidgen, Alex Harris, Dong Nguyen, Rebekah Tromble, Scott Hale, and Helen Margetts, "Challenges and Frontiers in Abusive Content Detection", in Proceedings of the Third Workshop on Abusive Language Online, eds. Sarah T. Roberts, Joel Tetreault, Vinodkumar Prabhakaran, and Zeerak Waseem (Association for Computational Linguistics, 2019), 80–93, https://aclanthology.org/W19-3509.pdf.

Major platforms such as Instagram continue to depend on current content moderation approaches that are inherently reactive, with classifiers trained exclusively on historical harms that have been identified, labeled, and systematically documented. This retrospective approach creates a perpetual game of digital cat-and-mouse, where harmful content creators are continuously incentivised to develop more sophisticated evasion tactics, rather than being motivated to create genuinely constructive and positive online interactions that contribute to healthier digital environments.[92]

Generative AI and deepfakes: The proliferation of increasingly sophisticated deepfakes is an example of targeted weaponisation of technology. Deepfakes leverage machine learning, Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs), a type of deep neural network particularly adept at analysing visual imagery and trained on datasets of real images or videos. This results in hyper-realistic synthetic content that is able to accurately replicate human features and movements.[93]

The rapid evolution of deepfake technologies and the increasing accessibility of generative AI tools have made it easier to create convincing false images and videos with minimal input.

A 2023 study of the deepfake ecosystem found that 98% of all online deepfake videos had pornographic content, and 99% of the targeted individuals were women.[94]

It is now possible to create a 60-second deepfake pornographic video for free in under 25 minutes with only a single clear image of the face of the intended target.[95] In India, during the recent 2024 elections, online platforms were the site of deepfake content created by political parties and their followers to influence voters, much of which was designed to stoke controversy and provoke strong emotional responses.[96]

Stalkerware: In addition to being widely accessible, stalkerware incorporates stealth features that enable malicious software to operate undetected on a victim's device, often disguised as legitimate programs. This deception and invisibility is critical for stalkers as it prevents victims from easily identifying the software until significant harm has occurred.[97] Existing security mechanisms on app stores frequently fail to adequately screen for stalkerware disguised as legitimate applications.[98]

92. Jeff Allen, "Why Is Instagram Search More Harmful Than Google Search? On Instagram's Decisions to Disable Search for Sensitive Topics and the Instagram We Can't Have", Integrity Institute, 13 February 2024. https://integrityinstitute.org/blog/why-is-instagram-search-more-harmful-than-google-search.

93. Lancashire Cyber Foundry, "An Introduction to Deepfakes", accessed 16 December 2024. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/cyber-foundry/lcf-articles/LCFArticle-Josh-Deepfakes_WEB.pdf.

94. Security Hero, "State of Deepfakes: Realities, Threats, and Impacts", accessed 8 November 2024, https://www.securityhero.io/state-of-deepfakes/#key-findings.

95. Ibid.

96. Nilesh Christopher, "Before India Election, Instagram Boosts Modi AI Images That Violate Rules", Al Jazeera, 12 April 2024, https://www.aljazeera.com/economy/2024/4/12/before-india-election-instagram-boosts-modi-ai-images-that-violate-rules.

97. Kaspersky, "The State of Stalkerware in 2023–2024", Securelist, 13 March 2024, https://securelist.com/state-of-stalkerware-2023/112135/.

98. Divyank Katira, "India's Parental Control Directive and the Need to Improve Stalkerware Detection", The Centre for Internet & Society, 4 April 2024, https://cis-india.org/internet-governance/blog/india2019s-parental-control-directive-and-the-need-to-improve-stalkerware-detection.

5.2 Technological responses to TFGBV

In this section, we highlight how certain technological tools, both AI-based and otherwise, are being piloted and deployed to prevent, detect and respond to TFGBV. These tools are not just specific to TFGBV or to women in politics specifically, but aim to capture a variety of attacks, including hate speech, harmful and toxic content, among others.

Prevention and early detection of TFGBV are crucial to curb harmful impacts. Developing robust systems for detecting abusive content is essential for effective online moderation. Some progress has been made in implementing technological solutions to detect and combat gendered harms online. Advances in machine learning and natural language processing (NLP) have improved the performance of these systems; however, various challenges remain.

For instance, there is a well know diversity problem in AI/ML teams that develop automated content moderation systems at social media companies, often on insufficient training data sets, and also employ non-native annotators, all of which negatively impact the accuracy of the automated content moderation process.[99]

Jurgens et al. argue that the NLP community should prioritise developing technologies that can assist authors, moderators, and platform owners in preventing abuse before it occurs.[100] A comprehensive repository for classifying online gender-based violence (oGBV) can aid in automating identification efforts.[101]

99. Mona Elswah, "Moderating Maghrebi Arabic Content on Social Media", Centre for Democracy and Technology, September 2024, https://cdt.org/wp-content/uploads/2024/09/2024-09-26-CDT-Research-Global-South-Moderating-Report-English-Arabic-final.pdf

100. David Jurgens, Libby Hemphill, and Eshwar Chandrasekharan, "A Just and Comprehensive Strategy for Using NLP to Address Online Abuse", in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, eds. Anna Korhonen, David Traum, and Lluís Màrquez (Association for Computational Linguistics, 2019), 3658–3666, https://aclanthology.org/P19-1357.pdf.

101. Gavin Abercrombie, Aiqi Jiang, Poppy Gerrard-Abbott, Ioannis Konstas, and Verena Rieser, "Resources for Automated Identification of Online Gender-Based Violence: A Systematic Review", in 7th Workshop on Online Abuse and Harms (WOAH), eds. Yi-Ling Chung, Aida Mostafazadeh Davani, Debora Nozza, Paul Rottger, and Zeerak Talat (Association for Computational Linguistics, 13 July 2023), 170–186, https://aclanthology.org/2023.woah-1.17.pdf.

Emerging tools like Perspective API[102] and OpenMeasures[103] are advancing online content moderation by developing sophisticated tools to detect toxic language, harassment, and sexually explicit content across digital platforms. These open-source solutions, complemented by more unique approaches such as StopNCII.org's digital fingerprinting[104] to tackle complaints of non-consensual intimate imagery, aim to systematically identify and remove harmful content. However, current technological approaches still lack specialised categorisation for TFGBV, highlighting the need for more nuanced detection and labeling mechanisms in digital safety frameworks.

Meta's CrowdTangle, a social media monitoring and sentiment analysis tool, used to be a critical resource for researchers and journalists to identify emerging disinformation narratives early. However, Meta shut it down earlier this year despite widespread protest[105] and replaced it with the Meta Content Library[106], which is less transparent and accessible.[107]

Despite their critical role, Meta Content Library and similar platforms such as Hootsuite[108], Sprinklr[109] and Tweetdeck (now X Pro)[110] suffer significant limitations that include restrictive rate limits, data retention constraints, and paywalled access. While large tech platforms develop sophisticated in-house tools with robust data collection capabilities, external researchers face increasingly complex barriers to comprehensive social media analysis, resulting in a growing tension between platform control and research transparency.[111]

102. Perspective API, "Using Machine Learning to Reduce Toxicity Online", accessed 2 December 2024, https://perspectiveapi.com/.

103. Open Measures, "About Us", accessed 2 December 2024, https://openmeasures.io/about-us/.

104. "StopNCII.org - Stop Non-Consensual Intimate Image Abuse", accessed 17 December, 2024, https://stopncii.org/.

105. Meta, "CrowdTangle", accessed 8 November 2024, https://transparency.meta.com/en-gb/researchtools/other-datasets/crowdtangle/.

106. Meta, "Meta Content Library and API", accessed 8 November 2024, https://transparency.meta.com/en-gb/researchtools/meta-content-library/.

107. Sarah Grevy Gotfredsen and Kaitlyn Dowling, "Meta Is Getting Rid of CrowdTangle – and Its Replacement Isn't as Transparent or Accessible", Columbia Journalism Review, 9 July 2024, https://www.cjr.org/tow_center/meta-is-getting-rid-of-crowdtangle.php.

108. "Hootsuite", accessed 17 December 2024, https://www.hootsuite.com/.

109. "Sprinklr", accessed 17 December 2024, https://www.sprinklr.com/.

110. X, "X Pro", accessed 17 December 2024, https://pro.twitter.com/.

111. Gabriel Nicholas and Dhanaraj Thakur, "Learning to Share: Lessons on Data-Sharing from Beyond Social Media", OSF Preprints, September 2022, https://osf.io/preprints/osf/2qnhv.

Blockchain-based content verification initiatives, like the News Provenance Project[112], aim to verify official campaign materials to establish a baseline of authentic content.

Educational tools such as the Bad News Game[113] and Join This Group[114] (which is tailored to the Indian context and helps users identify misinformation on direct messaging apps rather than social media and is also available in Hindi[115]) teach users about tactics used in disinformation campaigns by simulating the spread of fake news, helping them recognise and counter such tactics in real life.

Bot detection tools such as Bot Sentinel[116] track troll bots and untrustworthy accounts on Twitter, while Botometer[117] analyses Twitter accounts to assess their likelihood of being automated. However, these tools have faced resistance from the platform itself: in the midst of the company's legal battle with Elon Musk, Twitter threatened to revoke Bot Sentinel's access to its application programming interface (API).[118] It also dismissed Botometer as an unreliable web tool when it classified Musk's own account as a bot.[119]

Most of these tools are global in their application, but very few are specifically tailored to the Indian context.

112. "News Provenance Project", accessed 8 November 2024, https://www.newsprovenanceproject.com/.

113. "Bad News", accessed 8 November 2024, https://www.getbadnews.com/en.

114. "Join This Group", accessed 8 November 2024, https://whatsapp.aboutbadnews.com/#/intro.

115. Trisha Harjani, Melisa-Sinem Basol, Jon Roozenbeek, and Sander van der Linden, "Gamified Inoculation against Misinformation in India: A Randomized Control Trial", Journal of Trial and Error 3, no. 1 (28 February 2023), https://journal.trialanderror.org/pub/gamified-inoculation/release/2.

116. Bot Sentinel, "Auto Blocker – Get Started", accessed 8 November 2024, https://botsentinel.com/free-tools/autoblocker.

117. Botometer X, "Botometer X: An OSoMe project (bot•o•meter)", accessed 8 November 2024, https://botometer.osome.iu.edu/#!/faq.

118. Kali Hays, "In the Midst of Its Battle with Elon Musk, Twitter Threatens to Revoke Bot Sentinel's Data Access after Founder Suggests Twitter has More Than 5% 'Bots'", Business Insider India, 22 August 2022, https://www.businessinsider.com/twitter-threatens-to-remove-tracking-tool-bot-sentinel-from-its-api-2022-8.

119. Jon Brodkin, "Twitter Says Musk's Spam Analysis Used Tool That Called His Own Account a Bot", Ars Technica, 5 August 2022, https://arstechnica.com/tech-policy/2022/08/twitter-says-musks-spam-analysis-used-tool-that-called-his-own-account-a-bot/#gsc.tab=0.

Rapid responses and fact-checking can be assisted by AI-powered fact-checking tools such as Logically AI[120] that help streamline the verification process for claims, as well as crowdsourced verification platforms such as Emergent.Info[121] that allow users to track and verify rumours collaboratively. The Digital Polarisation Initiative[122] (which engaged university students to crowdsource verification efforts) and Dirt Protocol[123] (which enabled communities to collaboratively moderate data and incentivised accurate information-sharing through token rewards) were similar community-driven platforms, but are no longer active. The Political Ad Collector[124] (a browser extension) is a tool that promotes transparency by creating a publicly available database of political advertisements that it automatically copies from its users' Facebook feeds. NewsGuard[125] employs expert analysts to conduct detailed, transparent evaluations of online news sources using apolitical journalistic criteria, providing comprehensive reliability ratings for over 35,000 online sources that account for 95%+ of digital news engagement, while also tracking and analysing thousands of misinformation narratives to help clients identify and mitigate unreliable information.

In India, independent media operate many fact-checking initiatives such as Webqoof[126] and Boom[127] to fact-check false or misleading content as soon as it goes viral. They also provide Whatsapp-based tiplines that members of the public can use to submit content for verification or flagging.

Content moderation and removal are also vital components of this strategy. Given the high volume of user-generated content online, organisations are increasingly adopting automated solutions for efficient content moderation. Deepfake detection technologies, such as Microsoft's Video Authenticator[128], are important tools for identifying non-consensual deepfakes that disproportionately target women. Additionally, emotion recognition algorithms can help in identifying and mitigating harmful narratives targeting women in politics; emotion-aware models are more effective at recognising abusive language than traditional models that do not consider emotional contexts.[129,130]

120. "Logically", accessed 8 November 2024, https://www.logically.ai/.

121. "Emergent: A Real-Time Rumor Tracker", accessed 8 November 2024, http://www.emergent.info/.

122. RAND, "Fighting Disinformation: Digital Polarization Initiative", accessed 8 November 2024, https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search/items/digital-polarization-initiative.html.

123. RAND, "Fighting Disinformation: Dirt Protocol", accessed 8 November 2024, https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search/items/dirt-protocol.html.

124. ProPublica, "Political Ad Collector", accessed 8 November 2024, https://propublica.github.io/political-ad-collector/.

125. Newsguard, "NewsGuard: Global Leader in Information Reliability", accessed on 17 December 2024, https://www.newsguardtech.com/.

126. The Quint, "WebQoof", accessed 8 November 2024, https://www.thequint.com/news/webqoof.

127. "Boom", accessed 8 November 2024, https://www.boomlive.in/.

128. Tom Burt and Eric Horvitz, "New Steps to Combat Disinformation", Microsoft, 1 September 2020, https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/.

129. Santhosh Rajamanickam, Pushkar Mishra, Helen Yannakoudakis, and Ekaterina Shutova, "Joint Modelling of Emotion and Abusive Language Detection", in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, eds. Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Association for Computational Linguistics, 2020), 4270–4279, https://aclanthology.org/2020.acl-main.394.pdf.

130. Segun Taofeek Aroyehun and Alexander Gelbukh, "Aggression Detection in Social Media: Using Deep Neural Networks, Data Augmentation, and Pseudo Labeling", in Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), eds. Ritesh Kumar, Atul Kr. Ojha, Marcos Zampieri, and Shervin Malmasi (Association for Computational Linguistics, August 2018), https://aclanthology.org/W18-4411.pdf.

In a move to account for India's linguistic diversity, Shhor AI was locally developed as an AI-powered bot specifically to protect marginalised communities in India by recognising hate speech in vernacular and code-mixed languages (such as Hinglish or Tamlish).[131]

Finally, **post-attack support** is essential for those affected by TFGBV. Platforms such as HeartMob offer real-time support resources for individuals experiencing online harassment.[132] Users can share their experiences and receive assistance from vetted bystanders who provide supportive messages or help document abuse. Studies show that HeartMob effectively reduces feelings of isolation among users and aids in healing from trauma.[133] Uli, a browser extension, enables users to automatically redact abusive content in their feeds on Twitter/X using its crowdsourced database (this includes slurs in local Indian languages such as Hindi and Tamil). Importantly, it also allows users to instantly archive offensive tweets via screenshots, which they can store locally or email themselves.[134]

131. "ShhorAI", accessed 7 January 2025, https://www.shhorai.com/.

132. "About Heartmob", Right To Be, accessed 9 November 2024, https://hateandhope.righttobe.org/pages/about-page.

133. Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe, "Classification and Its Consequences for Online Harassment: Design Insights from HeartMob", Proceedings of the ACM on Human–Computer Interaction 1 (2017), 1–19, https://www.researchgate.net/publication/321636042_Classification_and_Its_Consequences_for_Online_Harassment_Design_Insights_from_HeartMob.

134. Uli, "Reclaim Your Online Space", accessed 9 November 2024, https://uli.tattle.co.in/.

6. Challenges and impact

Focus on English-only responses: Many TFGBV detection tools primarily focus on widely spoken languages, such as English. In India, where there are numerous regional languages and dialects, these tools may not be as effective in analysing and addressing misinformation in various languages or regional dialects.

A recent analysis of oGBV on four short-form video platforms popular in India revealed that none of these platforms consider the social and cultural contexts that affect the nature and consequences of oGBV in India, resulting in reporting mechanisms that lack space for context and do not require evaluation of local socio-cultural factors.[135]

> **TFGBV often exploits local cultural contexts, social norms, and political dynamics. Tools that are not customised to specific regional contexts may fail to identify or address gendered abuse that resonates with local audiences.**

Inadequate platform safety mechanisms: Despite platforms recognising the urgent need to address TFGBV, current safety mechanisms remain insufficient. Participants in a 2020 global survey of women journalists[136] reported that while platforms were aware of the risk of harassment, the journalists were provided with little to no support when actually faced with such a crisis, primarily due to a lack of resources and procedures to protect them. The burden of preparing for and handling online harassment continues to fall disproportionately on at-risk targets and victims rather than the platforms that facilitate it.[137]

User trust deficit: The effectiveness of fact-checking or abuse detection tools often depends on user engagement and trust. However, in India's politically polarised landscape, where online abuse and disinformation can spread rapidly, users may be sceptical of efforts to identify and combat these if they perceive such tools as biased or politically motivated.[138] Additionally, the sheer volume of content generated can overwhelm even the most sophisticated tools, leading to gaps in monitoring and response. While platforms provide transparency reports, they are not widely read and may sometimes not dissect the right categories that users, researchers and other stakeholders may want to see.

135. Divyansha Sehgal and Lakshmi T. Nambiar, "Online Gender Based Violence on Short Form Video Platforms: An Inquiry into Platform Policies and Safeguards", Centre for Internet and Society, 10 April 2024, https://cis-india.org/raw/online-gender-based-violence-pdf-10-april.

136. ICFJ-International Center for Journalists, "The Chilling: A global study of online violence against women journalists", https://www.icfj.org/sites/default/files/2023-02/ICFJ%20Unesco_TheChilling_OnlineViolence.pdf

137. Nitish Goyal, Leslie Park, and Lucy Vasserman, "'You Have to Prove the Threat Is Real': Understanding the Needs of Female Journalists and Activists to Document and Report Online Harassment", in CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–17, https://arxiv.org/pdf/2202.11168v1.

138. Clara Pretus, Camila Servin-Barthet, Elizabeth A. Harris, William J. Brady, Oscar Vilarroya, and Jay J. Van Bavel, "The Role of Political Devotion in Sharing Partisan Misinformation and Resistance to Fact-Checking," Journal of Experimental Psychology: General 152, no. 11 (2023), 3116–3134, https://psycnet.apa.org/fulltext/2023-83147-001.html.

Documentation and reporting mechanisms for TFGBV present significant structural, technical, and cultural barriers. Current reporting infrastructure across social media platforms have significant gaps in terms of accessibility: reporting interfaces are predominantly designed in English (and more recently Hindi), creating linguistic barriers for multilingual users in India, while non-compliance with Web Content Accessibility Guidelines often renders these mechanisms inaccessible to users with disabilities.[139] Current measures to address this issue put the onus on women to lock their accounts and therefore limit their reach, which is counter-productive to women in politics who depend on public platforms and engagement to share their work and views.[140]

Evidentiary requirements impose a further burden on victims, who must navigate complex and cumbersome reporting and escalation processes while experiencing ongoing harassment, which further compounds the impact of the abuse.[141]

The ephemeral nature of digital content (e.g., stories on Instagram that automatically disappear after 24 hours, disappearing messages on Whatsapp and other messaging apps, or simply posts that are quickly deleted) creates significant challenges in preserving evidence. This means that victims are tasked with collecting extensive documentary evidence while dealing with the trauma of ongoing harassment and threats or their immediate aftermath. This often requires proving a pattern of sustained harassment, which is particularly challenging given the diversity of languages in which such content is published and the distributed nature of online violence across various platforms and different user accounts.[142]

"There's this one user on one platform and he's on another now, and other platforms too… It takes weeks to get his social media accounts taken down. This person targets multiple people and has doxxed multiple people, but it's really hard to document and report it as an organised effort."[143]

The need to prove the existence of a coordinated campaign presents additional complications as victims typically lack access to data that would demonstrate orchestrated harassment networks or that the same actors were operating multiple accounts.[144]

139. Pen America, "Online Harassment Field Manual", accessed 9 November 2024, https://onlineharassmentfieldmanual.pen.org/.

140. Manish Singh, "Facebook rolls out feature to help women in India easily lock their accounts", Tech Crunch, 21 May, 2020, https://techcrunch.com/2020/05/21/facebooks-new-safety-feature-for-women-in-india-easily-lock-the-account-from-strangers/#:~:text=Manish%20Singh,of%20someone%20misusing%20their%20information.

141. ICFJ-International Center for Journalists, "The Chilling: A global study of online violence against women journalists", https://www.icfj.org/sites/default/files/2023-02/ICFJ%20Unesco_TheChilling_OnlineViolence.pdf

142. Anita Gurumurthy, Amrita Vasudevan, and Nandita 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/1662/Born-Digital_Born-Free_SynthesisReport.pdf.

143. Nitish Goyal, Leslie Park, and Lucy Vasserman, "'You Have to Prove the Threat Is Real': Understanding the Needs of Female Journalists and Activists to Document and Report Online Harassment", in CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–17, https://arxiv.org/pdf/2202.11168v1.

144. Amnesty International India, "Troll Patrol India: Exposing Online Abuse Faced by Women Politicians in India", 2020, https://decoders.blob.core.windows.net/troll-patrol-india-findings/Amnesty_International_India_Troll_Patrol_India_Findings_2020.pdf.

The systemic impact extends beyond immediate documentation challenges. Victims report harassment fatigue arising from the substantial time and resource costs of managing evidence collection and organisation while simultaneously navigating potential safety risks from increased exposure during the documentation process.

"At any stage of the process, we might totally disengage, get really burned out, and [feel] like 'I don't care anymore. Everybody knows everything about me. What do I have to hide?' I don't know how many times I've heard that."[145]

In-depth interviews conducted by the International Centre for Journalists with 113 women journalists worldwide[146] revealed that most victims of such abuse believed this reporting process to be futile. This belief arises from their personal experience of long waiting periods, often extending to months, for material to be removed, accounts to be suspended, any corrective action to be taken, or for the incident to even be acknowledged. Even after this waiting period, victims may face the disappointment of their reports being rejected for not meeting the platform's stringent standards for determining offensive or harmful content despite the extensive evidence presented.

Platforms' responses frequently demonstrate inconsistent enforcement timelines and limited transparency regarding report status, while automated responses fail to capture nuanced cultural contexts.

The reasons provided for rejections are often vague, and the platform usually does not provide any meaningful channel to appeal the decision. All communications are usually routed via a complaint portal, with no option to directly contact the platform to act in case of imminent danger or sustained harassment.[147]

These barriers disproportionately affect women from marginalised or minority communities. Women belonging to linguistic minorities, having disabilities, or limited literacy skills face compounded challenges in accessing and utilising reporting mechanisms. The technical and procedural requirements often reflect and reinforce existing social hierarchies, creating additional barriers for those with limited digital literacy or access to resources.[148]

145. Nitish Goyal, Leslie Park, and Lucy Vasserman, "'You Have to Prove the Threat Is Real': Understanding the Needs of Female Journalists and Activists to Document and Report Online Harassment", in CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–17, Goyal et al., "'You Have to Prove the Threat Is Real'", 1–17.

146. ICFJ-International Center for Journalists, "The Chilling: A global study of online violence against women journalists", https://www.icfj.org/sites/default/files/2023-02/ICFJ%20Unesco_TheChilling_OnlineViolence.pdf.

147. Ibid.

148. Plan International, "Free to Be Online? Girls' and Young Women's Experiences of Online Harassment", 2020, https://www.plan.org.au/wp-content/uploads/2020/10/SOTWG-Free-to-Be-Online-2020.pdf.

Under-reporting of TFGBV: There is significant under-reporting of TFGBV, which presents barriers to accessing justice. The Global Partnership for Action on Gender-Based Online Harassment and Abuse reports that the reasons for this include "shame, stigma and discrimination; fear and distrust of law enforcement; the focus on the survivors/victims and users to report abuse they experience; inadequate response from social media platforms, including language barriers in responding to abuse in non-English speaking countries."[149]

The internet was initially viewed as a safe space for self-expression and freedom, particularly for women or individuals from marginalised communities, who sought to find like-minded others. Kovacs et al. visualise the internet as a large public domain with many streets and spaces:

> *"This allows us to interact with and meet others whom they may otherwise not come into contact with offline – to form friendships, plan campaigns, and inhabit or create spaces for themselves to be heard. However, to be a woman online – just like to be a woman walking the streets of an Indian city, town or village – is to transgress an unwritten law of patriarchy; to cross over into a space that isn't meant to be yours."[150]*

However, evolving internet infrastructures have resulted in an online environment that embodies (and amplifies) historical, real-world systems of misogyny, oppression, and violence, resulting in the rapid shrinking of safe spaces for women.

149. Wilton Park, "Events: Building a Shared Agenda on the Evidence Base for Gender-Based Online Harassment and Abuse", accessed 9 November 2024, https://www.wiltonpark.org.uk/event/building-a-shared-agenda-on-the-evidence-base-for-gender-based-online-harassment-and-abuse/.

150. Anja Kovacs, Richa Kaul Padte, and Shobha SV, "'Don't Let It Stand!' An Exploratory Study of Women and Verbal Online Abuse in India", Internet Democracy Project, April 2013, https://cdn.internetdemocracy.in/idp/assets/downloads/reports/women-and-verbal-online-abuse-in-india/Internet-Democracy-Project-Women-and-Online-Abuse.pdf.

7. **Concluding remarks**

The internet, much like society, has had a gendered upbringing duplicating and amplifying gendered harms and violence. Women's mobility in public spaces has historically been a social issue as a result of existing patriarchal systems and codes of conduct, and we have seen these patterns being replicated online.

Through this essay, we have demonstrated how women in politics in India are subjected to gendered disinformation, gendered hate speech, and oGBV. We have also presented some of the emotional attributes at play that contribute to this weaponisation of gender, leading to targeted hate against women in politics. Lastly, we have focused on how technology plays a role in increasing the dissemination of such gendered hate while providing some avenues for mitigation.

Much of current research on disinformation focuses on fact-checking as a response to dis/misinformation. This approach draws from a binary, 'true/false' understanding of information and posits that misinformation can be addressed by countering it with verified information.

However, studies have found that false news spreads faster than true news[151], and the reach of the fact-checked piece of content can be often much lower than that of the viral fake content. This exposes the limitations of fact-checking as the primary means to tackle misinformation; for instance, it tends to ignore the underlying issues that can rouse 'irrational' emotions in individuals and in-groups. Similarly, efforts to counter hate speech campaigns against structurally and historically marginalised groups cannot solely rely on presenting facts. The challenge with information disorder lies not only in its dissemination of false information but also in its capacity to mobilise specific affects and passions towards harmful ends.

An additional practical limitation is that fact-checking requires robust infrastructures, which are often lacking for under-resourced languages in the Global South. It is imperative to address these gaps to create better remedial measures and improve platform responses for combating TFGBV.

151. Peter Dizikes, "Study: On Twitter, false news travels faster than true stories" MIT News. 18 March 2018, https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.

We also highlighted several content moderation gaps, and how they are particularly exacerbated in regional languages. While we have discussed automated content moderation and the industry wide push for it, it is important to also reflect on human review for content moderation. This is both a highly resource intensive as well as challenging activity and has had significant impacts on moderators' mental and physical health.[152] Unfortunately, human oversight continues to remain an important part of the moderation process given that NLP detection systems still fail to capture and detect cultural and linguistic nuances, and that they also need to be constantly updated to adapt to the changing trends in language and culture. It would be beneficial for the entire ecosystem, if platforms can create more diverse NLP teams as well as invest in regional and local annotation teams that develop and annotate these detection systems such that these tools perform better and are able to detect harmful content more effectively.[153]

We have highlighted several technological solutions to address TFGBV, but it is essential to recognise that technology alone cannot resolve the problem, and that the tech solutions for detection of abuse that exist today are not necessarily at scale. Furthermore, the scale and impact of these solutions are not known.

Comprehensive strategies must also include policy reforms, educational initiatives, and community engagement to create a safer environment for women in politics.

All stakeholders – platforms, policymakers, law enforcement – need to do more to provide a safer experience for women online. Suggested responses to counter TFGBV tend to advocate protectionist measures and advocate for either adding controls that limit users' reach or often compel users to get off the platform but these further diminish women's participation online. Instead, it is necessary to focus on understanding the perpetrators' tactics and what choices platforms optimise for (including their business models). Researchers also need to be given greater access to study cross-platform patterns of TFGBV and hateful content. There are certain tech working groups and coalitions already in existence that work on sharing cross platform signals in particular areas (child safety for example).[154] Greater participation and engagement of industry and non-industry actors (civil society, academia, policymakers) could help in creating greater awareness around cross-platform effects and also reduce the ease for bad actors.

152. Ruth Spence, Antonia Bifulco, Paula Bradbury, Elena Martellozzo, Jeffrey DeMarco "Content Moderator Mental Health, Secondary Trauma, and Well-being: A Cross-Sectional Study." Cyberpsychol Behav Soc Netw. 2024 Feb;27(2):149-155. https://pubmed.ncbi.nlm.nih.gov/38153846/.

153. Mona Elswah, "Moderating Maghrebi Arabic Content on Social Media", Centre for Democracy and Technology, September 2024, https://cdt.org/wp-content/uploads/2024/09/2024-09-26-CDT-Research-Global-South-Moderating-Report-English-Arabic-final.pdf.

154. "Technology Coalition", accessed 17 December 2024, https://www.technologycoalition.org/about.

Attention must be paid to content moderation in non-English languages. Lastly, researchers have studied the potential of using non-engagement signals (such as more focused review of quality metrics as opposed to engagement) to improve outcomes for both platforms and society.[155] Whether or not this works out for platforms, it is essential that platforms not prioritise hateful viral content with higher engagement but seriously consider the societal level of harm such engagement is causing, particularly for marginalised identities. Lastly, while platforms share transparency reports, having more categorical and customised reports which allow one to look into specific harms (TFGBV in India by age for example), could help in understanding how and what are the ways in which harmful content is being shared online, and allow for more tailored mitigation strategies.

Feminist praxis has advocated for a feminist internet, with the emergence of feminist infrastructures in academia and policy.[156] Perhaps an alternative imagination which prioritises experiences of women and gender and sexual minorities, through alternative technological and non-technological layers (hardware, software, inclusive design among others) could be a starting point to rethink platforms and their use. Non-localised and context agnostic responses from platforms and limited policy interventions have facilitated an environment in which TFGBV continues to exist and even thrive. As we adopt various internet governance processes globally and continue to think about local responses and regulations, it is essential that we keep our focus on the gendered aspects of technologically mediated harm and continue to emphasise the need to improve women's online experiences.

155. Tom Cunningham, Sana Pandey, Leif Sigerson, Jonathan Stray, Jeff Allen, Bonnie Barrilleaux, Ravi Iyer, Smitha Milli, Mohit Kothari, and Behnam Rezaei, "What We Know About Using Non-engagement Signals in Content Ranking", 9 February 2024, https://arxiv.org/pdf/2402.06831.

156. Puthiya Purayil Sneha and Saumyaa Naidu, "Understanding Feminist Infrastructures: An Exploratory Study of Online Feminist Content Creation Spaces in India," Centre for Internet and Society, March 25, 2024. https://cis-india.org/Feminist_Infrastructures_Report.

Centre
for Internet
& Society