CSSF Women, Peace and Security Helpdesk

Understanding the Role of Gender-based Online Hate Speech in Exacerbating Conflict and Violence

Submitted: 08/08/22

Assignment Code: WPS003

WPS Women, Peace & Security Helpdesk

The Women Peace and Security Helpdesk, managed by Saferworld in partnership with Conciliation Resources, GAPS UK, University of Durham and Women International Peace Centre (WIPC), was established in December 2021 to increase capability across the UK Government on WPS policy and programming in order to make its work on conflict and instability more effective. If you work for the UK government and you would like to send a task request, please email us at wpshelpdesk@saferworld.org.uk. If you do not work for the UK government but have an enquiry about the helpdesk or this report, please email us at enquiries.wpshelpdesk@saferworld.org.uk

Experts: Ojaswi Shah

Direct Audience: FCDO Online VAWG Policy Team

Confidentiality Status: Open Source

Abstract

This report provides an evidence review summary of gender-based online hate speech in conflict contexts, exploring published evidence available to the issue in order to understand how prevalent gender-based online hate speech is in conflict contexts, the different impacts it has and evidence on how this is being addressed. The summary shows that while there is strong evidence on GBV and online hate speech and similarly for conflict and online hate speech, the intersection and linkages between these aren't strong in available literature. Evidence on women, peace and security (WPS) and digital harm through social media is critically missing from currently available literature, particularly looking at India, Nigeria, Myanmar and Sri Lanka.

Introduction:

This evidence summary report focuses on reviewing secondary literature from conflict contexts to inform FCDO on the nuances of gendered online hate speech in conflicts that have multi-ethnic identities. The review looks at the prevalence of using social media to fuel gender based hate speech, its impact and the strength of the evidence base.

As part of this, the analysis for this report is informed by secondary literature on Sri Lanka, Nigeria, Myanmar and India. These case studies were selected because of ongoing identity-based conflict and tensions that affect both women and men from diverse social, religious and ethnic backgrounds and provided an opportunity to explore whether and how online hate speech further entrench existing intersectional gender norms and conflict drivers. The review also looks at existing social media conflict frameworks to inform its methodology and analysis. Frameworks include Mercy Corps' Social Media and Conflict: Understanding Risks and Resilience – An Applied Framework for Analysis (2021); swisspeace's Social Media in Peace Mediation: A Practical Framework (2021), and the United Nations (UN)'s toolkit on Digital Technologies and Mediation in Armed Conflict (2019). These reports provide analytical frameworks for assessing social media risks in conflict context and highlight how conflict sensitivity might be applied when using digital approaches to peace and conflict prevention

Additionally, the task adopted a non-systematic evidence review attempting to summarise available literature review which includes practitioner methodologies for social media in conflict and crisis contexts and research reports, articles and blogs on Sri Lanka, Nigeria, Myanmar and India. Methodology also employs techniques such as systematic searches using key words and phrases to source evidence base. These include hate speech; online hate speech; cyberhate; gendered hate speech in conflict; social media; online disinformation; gender and cybersecurity in conflict; women, peace and security (WPS) and online hate speech; misogyny within online hate speech in conflict; patriarchal norms, masculinities and online hate speech; and online participation. These specific key words were further distilled to case study countries in order to explore literature at a more contextual level. Analysis presented in this report is derived from research, reports and publications from these country contexts along with research reports that provide conceptual clarity on either gender and online hate speech or conflict and online hate speech.

Limitations:

It is also important to note that there are two key limitations of this research which includes the inability to unpack granular data at the country level which would require fluency in local languages to comprehend and analyse social media vernacular of those contexts, and secondly, lack of consultations with experts working on this field and in those specific contexts.

The review relies solely on analysis that has been carried out in English and does not extend to literature that might have been available in other languages. Finally, the evidence summary also does not extensively cover news media that base their reportage on online hate speech and disinformation.

Key terminology used in the evidence summary:

Below are definitions adopted to undertake this analysis. It is important however to acknowledge that hate speech and online hate speech in particular is dynamic and continues to have multiple interpretations and that also stems from how fast online digital space is evolving with new forms of online syntax being developed and used constantly.

'**Cyberspace**' is an interactive domain made up of digital networks that is used to store, modify and communicate information. It includes the internet, but also the other information systems that support businesses, infrastructure and services.[1]

'**Social media**' is an umbrella term for a variety of interactive applications that allow users to create content (text, photos, videos) and share ideas with each other through an online community (UN, 2019).[2]

'**Hate speech**' is commonly defined as any communication that disparages a person or a group on the basis of some characteristics such as race, colour, ethnicity, gender, sexual orientation, nationality or religion.[3] The United Nations (UN) Strategy and Plan of Action on Hate Speech defines hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent gender or other identity factor".

'**Disinformation**' refers to incorrect information spread by people intentionally in order to deceive or manipulate others, and '**misinformation**' refers to incorrect information spread by people without the intent to deceive. [4]

'**Online hate speech**' refers to the use of offensive language, focused on a specific group of people who share a common property. The identification of the potential targets of hateful or antagonistic speech is key to distinguishing the online hate from arguments that represent political viewpoints.[5]

'**Cyberhate**' is conceptualised as the use of violent, aggressive or offensive language, focused on a specific group of people who share a common property, which can be religion, race, gender or sex or political affiliation, through the use of Internet and Social Networks, based on a power imbalance, which can be carried out repeatedly, systematically and uncontrollably, through digital media and often motivated by ideologies.[6]

---

[1] *The UK Cyber Security Strategy, November 2011*

[2] *United Nations Centre for Humanitarian Dialogue (2019), Toolkit on 'Digital Technologies and Mediation in Armed Conflict', UN. https://peacemaker.un.org/sites/peacemaker.un.org/files/DigitalToolkitReport.pdf*

[3] *Tontodimamma, A., Nissi, E., Sarra, A. et al (2021), 'Thirty years of research into hate speech: topics of interest and their evolution', Scientometrics 126, 157–179. https://doi.org/10.1007/s11192-020-03737-6*

[4] *Bertolini, G. (May 2016), 'Digital Hydra: Security Implications of False Information Online', NATO StratCom COE*

[5] *Castano-Pulgarin, Suarez-Betancur, Vega and Lopez (2021), 'Internet, social media and online hate speech: systematic review', Aggression and Violent Behaviour: ScienceDirect Vol. 58 May – June. https://doi.org/10.1016/j.avb.2021.101608*

[6] *Ibid.*

'**Online disinformation**' includes all forms of false, inaccurate or misleading information designed, presented and promoted to intentionally cause public harm or for profit in the cyberspace.[7]

The *OSCE Office for Democratic Institutions and Human Rights (ODIHR)* describes '**gender-based hate crimes**' as criminal offences motivated by bias against a person's gender. These crimes often seek to intimidate and suppress ways of life or expressions of identity that are perceived as not complying with traditional gender norms. The victims of such crimes are often targeted due to their perceived deviation from gender norms, including non-conformance with traditional male-dominated relationships and patriarchal norms.[8]

[7] *European Union (2018), 'Multi-dimensional Approach to Disinformation: Report of the independent High Level Group on fake news and online disinformation', EU: Belgium.*

[8] *https://blogs.icrc.org/law-and-policy/2021/11/30/gender-based-hate-crime-conflict/*

Summary Findings:

This section briefly outlines how gender-based online hate speech and gendered disinformation tactics are used in social media messaging to further political divisions and/or aggravate conflict dynamics, particularly in conflict contexts that have multi-ethnic identities with a historical legacy of tensions.

Framing online hate speech, conflict and gender:

Reach of social media

The global growth of social media has created new platforms for hate speech amplifying hate crimes and rhetoric. To put things into perspective, according to one estimate,[9] in 2020, 490 million new users joined social media, a growth rate of 13.2%. By the start of 2021, there were some 4.2 billion active social media users, 53% of the global population – of whom women users accounted for 45.9%.[10] On average, users have accounts on eight different platforms and spend two-and-a-half hours per day on social media. By April 2022 this year, there were 4.65 billion social media users around the world, equating to 58.7% of the total global population and ten new users registering every second.[11] Collectively the world spends more than ten billion hours using social media platforms each day according to recent statistics,[12] which provides a compelling insight into how rapidly the syntax of communication is changing, globally. Social media has amplified the reach and spread of political, religious, social, cultural, environmental discussion and messaging. Over the last two decades, social media has been increasingly used to develop and disseminate political, social and cultural narratives influencing how people interact. Published and up-to-date evidence on social media penetration at a global scale and in many country contexts including India, Nigeria and Sri Lanka is easily accessible, which is highly useful to understand the reach and scope of digital platforms through social media – this includes understanding user behaviour online. However, the limitation within this is that much of the evidence feeding into this comes from language that needs to be in English or an accepted language that the software accepts. For instance, if there is a gender-based hate speech attack using local dialect on women from Rakhine community in Myanmar, this is often not picked up in the body of evidence analysing gender-based hate speech in conflict unless it is captured by local organisations working in that context.

[9] *Data based on Digital 2021: Global Overview Report.*

[11] *Global Social Media Statistics informed by Kepios report. https://datareportal.com/social-media-users*
[12] *Ibid.*

Contextualising online hate speech within conflict and gender

> *"Sexist hate speech is a form of "social shaming" that aims to degrade women, instil fear and insecurity thus contributing towards maintaining and reinforcing a gender hierarchy and patriarchy in public places*".
>
> Liri Kopaçi-Di Michele, Head of the Equality Division, Council of Europe (Peraro, 2016)

It is now acknowledged that 'hate speech' on the Internet is a global concern and with no kill-switch solution.[13] Hate speech often acts as an extension of conflict and tensions that already exist, and online hate speech is a further augmentation of that reality. Social media has created and amplified opportunities for accountability, collaboration and dialogue within different contexts, including in conflicts. For example social media platforms can connect people to social movements and provide the tools and opportunities for new social movements to emerge, empowering those who seek gender equality, whether they are posting as individuals or as part of organised activism.[14] Evidence also strongly suggests that online narratives fuel divisive political narratives and perpetuate misogynistic attitudes, including reinforcing toxic gender expectations. There is strong research and evidence – including at the UN level, and gathered in multiple contexts dealing with different conflicts and/or crises – showing clear causal linkages between online hate speech and divisions within communities. Similarly, strong evidence on online hate speech being used to target women and minority groups is also found and seems set to increase, with research projects currently underway looking into this. For example, a recent UN report states that, *"in many countries, three quarters or more of the victims of online hate speech are members of minority groups and that women belonging to these groups are disproportionately targeted… it can too easily prepare the ground for dehumanisation and scapegoating of minorities, and for normalising hate*."[15]

Analysis around the weaponisation of social media to fuel conflict can also be found, particularly with the Mercy Corps (2019,2021) which provides a strong conceptual framework for this. Since 2017/18 there has been a decrease in research on the links between social media and radicalisation and violent extremism, which suggests this is less of a priority area. The majority of this analysis tends to focus on political and identity divisions and rarely integrates gender norms in conflict into their analysis. For instance, social media is being weaponised to carry out targeted attacks on minority groups, including women and girls, and to attack political targets. This has been documented in various peer-reviewed research reports, but these reports don't explicitly mention why these groups are being attacked, other than providing the offline causes to conflict and gender-based violence. In India, for instance, Muslim communities, other tribal and ethnic groups, and over the last several years, groups that are not 'favourably considered' within a brahminical hindu structure, are experiencing more offline-online attacks. A recent ODI research report presents evidence that social media is shaping our engagement with gender norms, often along sexist, racist, and discriminatory lines.[16] From racial and ethnic targeting[17] linked to data monetisation by private companies

---

[13] *Samaratunge, S. and Hattotuwa, S. (2014), 'Liking Violence: A study of hate speech on Facebook in Sri Lanka', Centre for Policy Alternatives (CPA). https://www.cpalanka.org/wp-content/uploads/2014/09/Hate-Speech-Executive-Summary.pdf*

[14] *Washington, K. and Marcus, R. (2022), 'Hashtags, memes and selfies: can social media and online activism shift gender norms?'. ALIGN Report. London: ODI.*

[15] *Fernand de Varennes (2021), 'Recommendations made by the Forum on Minority Issues at its thirteenth session on the theme "Hate speech, social media and minorities" Report of the Special Rapporteur on minority issues', UN Human Rights Council: 46th Session: 22 February – 19 March 2021: Agenda item no. 5, UN General Assembly*

[16] *Diepeveen, S. (2022) 'Hidden in plain sight: how the infrastructure of social media shapes gender norms'. ALIGN Report. London: ODI. https://www.alignplatform.org/ resources/report-hidden-in-plain-sight*

[17] *Ibid.*

cashing in on social media users and algorithmic digital identifiers[18], social media is being used to either recreate or reinforce specific gender expectations and ethnic stereotypes from women, men, boys, girls and sexual and gender minorities.

Evidence talks about either conflict and hate speech or gender inequality, GBV (gender-based violence) and hate speech. On the one hand, offline conflict is already seeping into the online syntax benefitting mostly the political elite and the status quo gatekeepers, and on the other, companies that profit from increased social media users and interaction are being questioned on accountability to comply with and develop new legal requirements to promote cybersecurity by putting up measures to combat online hate speech, cyberhate and gender-based online violence. Evidence on cyberhate and gender-based online violence is very strong and at present a hotly discussed issue during this evidence review, which includes how social media companies benefit from an increasing user base and their online data and behaviours by having control and ownership of that data for further manipulation, either by selling that data to other companies or targeting individuals based on their interaction and user information.

Deepening gendered conflict and tensions within communities

Evidence and analysis on social media deepening gender-based conflicts that target minority groups, including LGBTQI, from different social, cultural, political and economic identities within these contexts is available and documented. In some instances, they have overlapped with socio-economic disparities that emerged during the Covid-19 pandemic, particularly in India (example provided below). Social media platforms, search engines and online news organisations, for example, play an increasingly significant role in elections integrity, civic discourse and group identity formation, with offline impacts on peace and social cohesion.[19] The spread of malicious or inaccurate information has long been a driver of conflict through in-person communication and traditional media, and digital platforms and behaviours — specifically on social media — uniquely contribute to conflict. Conflict actors are increasingly using social media and cyberspace to carry out targeted attacks both in conflict and post-conflict contexts. Repressive authorities, armed groups and violent extremists are making innovative use of digitally enabled tools and methods to distort facts on the ground and spread incendiary rhetoric with the goal of obfuscating accountability, undermining community acceptance, eroding social cohesion or inciting panic and/or violence.[20] There is growing evidence around conflict actors employing tactics on social media to further exacerbate tensions either for political gain or to control the narrative against their opposition or other ethnic/ identity groups. This evidence includes the Mercy Corps, ALIGN research reports, UN research and various research articles, published blogs, and op-eds from the four country contexts.

In the context of Myanmar, the root causes and drivers of hate speech are multiple and originate in both the historical and current context providing space for racist, sexist and religio-nationalist sentiments to remain widespread in popular society.[21] Recent research on Myanmar identifies how online hate speech is informed by mutually reinforcing constructed narratives aimed at advancing Buddhist-Burman dominance at the expense of ethnic and religious

[18] *Ibid.*

[19] *Guay, J. and Gray, S. (2019), 'The Weaponization of Social Media: How social media can spark violence and what can be done about it', Mercy Corps.*

[20] *Mercy Corps (2021), 'Social Media and Conflict: Understanding Risks and Resilience – An Applied Framework for Analysis. https://www.mercycorps.org/sites/default/files/2021-08/Assessing-Digital-Conflict-Risks-Resilience-073021.pdf*

[21] *Minority Rights Group (2020), 'Online hate speech in Myanmar: an evolving threat'. https://minorityrights.org/2020/12/20/hate-speech-myanmar/*

minorities,[22] and often disproportionately affects women, girls, young men and boys even within the minority groups. The report also catalogues[23] a number of key drivers of hate speech, including the role of ultranationalist groups, the political and business interests of the elite, and socio-economic factors such as poverty, education, and historical divisions; these key drivers help explain the root causes of hate speech as well as how systemic and entrenched narratives of hate and discriminatory structures remain. It should be noted that the root causes of online hate speech mirror those of conflict, playing out offline to online hate speech. Granular information available on gender, conflict and use of online hate speech is currently inadequate and proved difficult to obtain through secondary online research. Much of the focus was either on sexist hate speech as a gender-based violence issue or on political disinformation to manage the status quo, control political dissent and even abet ethnic cleansing. The need for better gendered analysis that addresses the problem through a women, peace and security lens in conflict-affected and fragile contexts is critically missing from the body of evidence that was researched.

Similarly, in Sri Lanka, the resurgence of Sinhala Buddhist ultranationalist groups has grown in parallel to targeted disinformation and the spread of hate speech on social media since 2018. Much of the online hate speech is targeted towards Muslim minorities, Tamils, and against women and young girls. Attacks on women activists, politicians, journalists are commonplace, along with targeting Muslim minorities using both abusive and misogynistic messages against them. From denigrating their physical appearance and sexist remarks to attempted character assassination, women and minorities in Sri Lanka face a great deal of gender-based hate speech on social media, as evidenced in the research carried out. Reportedly, Sri Lanka's culture of impunity and the breakdown in the rule of law is what affords the space for extremist groups with politico-religious identities to say what they do and get away with it.[24] In India, studies have shown how digital rumours have spurred mob lynchings of minority Muslim women and men by Hindu nationalist groups (Mirchandani 2018).[25] Muslim women and men have been subject to repeated harassment over the years, with women often being targeted more often through a combination of a religious and patriarchal lens.

Evidence of the impact on the broader community is weak, although emerging research published over the last two–three years discusses how women, men and LGBTQI groups from diverse backgrounds use social media and are open for manipulation by handing over their online data to companies and being open for targeted profiling, messaging and even attacks. For instance, a Mercy Corps (2021) report talks about how ordinary citizens become embroiled in these processes, whether intentionally or unconsciously, and contribute to networks of online and offline actions that undermine healthy societies or foment violence.[26] Identity and context frame how social media narratives are received, and which stories gain traction and which do not – the character of the information ecosystem is key to assessing how and where social media weaponisation is likely to succeed (Mercy Corps, 2021). The fluidity of offline-online spaces has allowed gender hierarchies and power dynamics to permeate the digital sphere.[27] Outcomes for gender equality today hinge on understanding the manner in which social norms and rules are recast in the criss-crossing experiences of physical-virtual life.[28]

---

[22] *(2020), 'Hate Speech Ignited: Understanding Hate Speech in Myanmar', Burma Monitor, Progressive Voice and International Human Rights Clinic: Harvard Law School. http://hrp.law.harvard.edu/wp-content/uploads/2020/10/20201007-PV-Hate-Speech-Book-V-1.4-Web-ready1.pdf*

[23] *Ibid.*

[24] *S. Samaratunge and S. Hattotuwa (2014), op. cit.*

[25] *Udupa, S., Gagliardone, I., Deem, A. and Csuka, L. (2020), 'Hate Speech, Information Disorder, and Conflict', SSRC.*

[26] *Mercy Corps (2021), op. cit.*

[27] *IT for Change (2021). Participatory Action Research on Gender-Based Hate Speech Online with a Karnataka-Based Youth Group. Recognize, Resist, Remedy: Addressing Gender-Based Hate Speech in the Online Public Sphere. https://itforchange.net/sites/default/files/1738/PAR-on-gender-based-hate-speech-online-with-a-Karnataka-based-youth-group.pdf*

[28] *Ibid.*

Conflict, masculinities and gender-based hate speech

Evidence on masculinities and conflict in itself is not as strong compared to gender inequality in conflict or GBV in conflict. One of the focus of this review was to undertake a search on whether and how harmful masculinities in conflicts interact with gender-based online hate speech. This isn't strongly present in the available literature unless reviewers were to take literature on misogyny within online hate speech and masculinities in conflict and use the analysis to infer findings. Recent literature such as the ALIGN reports however attempt to unpack this within a global context and not necessarily discuss them within conflict contexts. Below are some findings inferred from available literature as well as gaps identified in the available evidence.

Online-offline patriarchy, masculinities, and conflict

Gender-based hate and sexual harassment are intended to restore both women and men to "their place" and reinforce the difference between the genders.[29] Gender norms and expectations for both women and men are recreated on social media, often playing out the same tropes of being traditional man and traditional woman. From expected gendered hobbies to pressures on social performance, economic performance, political and civic performance, both women and men who don't conform to the rules face higher risks of cyberhate and bullying. This extends to the political domain, particularly where national and local politics are governed by identity-, religion- or ethnicity-based ideology. There is growing evidence, mostly from the last decade, that provides an insight into how the need to perform on social media puts pressure on women, men and LGBTQI groups. This evidence shows that much of this performance is dictated by patriarchal norms, unequally affecting women, girls and LGBTQI people from ethnic and religious minorities.

For instance, analysis of masculinities, digital capitalism and social media exists, but is largely insufficient. Besides India and (to a lesser extent) Sri Lanka, thorough evidence on this was largely missing from the contexts studied. Where analysis does exist, it is very clear. For example:

"*Bounded within the norms of surveillance capitalism, digital sociality spaces make way for desire and agency, but are also seamless extensions and powerful determinants of hegemonic and violent masculinity.*[30] *They span innumerable, inscrutable worlds of toxic maleness — multiplying misogynistic hashtags and proliferating homosocial (male-only) communities — that are subterranean, but always ready to strike, from bro-clubs to incel groups and women-hater gangs, digital space is constantly growing the patriarchal space."*[31]

Within contexts of sociopolitical and economic fragility, 'class' as an identity underpins gender and power hierarchies, particularly in the conflict contexts researched for this report. Understanding this within the context of patriarchy and masculinities is important, as different men and women coming from different backgrounds determine their class, based on caste, education, language, region, income and so forth. Exploring these important intersectional linkages within online hate speech is critically missing from the majority of literature reviewed and presents a critical gap in research, particularly where social conflicts are complex and protracted. Typically, it is men and women from

---

[29] *Marjan Nadim, Audun Fladmoe (2019), 'Silencing Women? Gender and Online Harassment', Institute for Social Research. https://creativecommons.org/licenses/by-nc/4.0/*

[30] *Gurumurthy, A., Jha, B. (2020), 'Public participation is a woman's first order claim to being recognised as a human being, the pandemic can't be allowed to undermine that', 28-04-2020: Firstpost. https://www.firstpost.com/tech/news-analysis/public-participation-is-a-womans-first-order-claim-to-being-recognised-as-a-human-being-the-pandemic-cant-be-allowed-to-undermine-that-8419571.html*

[31] *Ibid.*

marginalised and minority communities who are at the receiving end of systemic violence in the off-line reality, and it is important to understand and acknowledge how they criss-cross between the offline and online. The majority of published evidence on online hate speech looks at it from a political aspect exploring targeted disinformation tactics within India, Sri Lanka, Myanmar and Nigeria, and most of the online gendered hate speech content links back to GBV, violence against women and girls and finally notions of traditional patriarchy linked to male honour and ultra-nationalism. Through the evidence, it can be inferred that within contexts where identity-based tensions and structural violence occurs, it adds a layer of harm to online abuse and violence wherein gender intersects with identity and is used as a tool to diminish the 'others / enemy / threat / opposition' through the weaponisation of gender norms. Direct causal linkages aren't abundantly clear within online hate speech analysis and there is scope to look into this more closely in the future.

For instance, during the Covid-19 pandemic, the narrative around Muslim men using Covid-19 as a biological bomb to attack Hindu communities took over Indian national and social media, gaining momentum and fanning communal tensions and oppression of an already vulnerable group during the lockdown, not just in India but also spilling over to neighbouring Nepal. A right-wing Hindu political party in India tried to gather support using social media manipulation tactics, floating the idea that Muslim men are waging "*jihad*" in India through love marriages – claiming that Muslim men use love, seduction and trickery to convert Hindu (and Christian) women to Islam and therefore the women should keep themselves "pure" in order that they might be fit mothers of the nation.[32] Patriarchy veiled as ultra-nationalism has been rife in Indian social media, reinforcing patriarchal norms women and girls have been subjugated to in the offline, i.e. the need to protect the honour of their families and the country by keeping their 'purity' intact. Social media only provides another avenue for patriarchy to play out and once again it is women, girls and marginalised minorities who are acutely targeted, as is evident in the literature on this issue. Cyberspace amplifies sexist, racist and caste-ist disinformation, in multiple folds affecting vulnerable minority groups with women, girls and LGBTQI groups being disproportionately affected.

Misogynistic narratives and tactics to malign individuals and communities

Gendered online harassment is seen as offline misogyny moved to a new arena.[33] Evidence on how misogynistic narratives are developed, shared and accessed with the purpose of maligning, intimidating, threatening and silencing women, girls and LGBTQI groups is available and strongly present in published literature. However, within conflict contexts, this is mostly discussed within the identity-based lens described above and the research is moderate as it focuses mostly on political disinformation tactics and examples. Gauging the impact on affected and broader communities requires more work and this might require longitudinal research, rather a short piece of work. Another factor is to understand user behaviour, and while strong research analysis is starting to emerge that provides insights into the technical aspects of user behaviour, it is mostly generic and does not focus explicitly on gender and conflict. This review does not take into account research and data in local languages which might provide more insight into what narratives are being pushed and how.

Essentially, once content is produced for the web and originally for a single platform, given user interactions and responses, it often replicates and mutates into other content over dozens of other websites and platforms, making it impossible to completely erase a record from existence even if the original was taken down, deleted or redacted.[34]

---

[32] *https://thewire.in/law/love-jihad-ordinance-communal-rhetoric-divisive-justice-ap-shah*

[33] *Marjan Nadim, Audun Fladmoe (2019), 'Silencing Women? Gender and Online Harassment', Institute for Social Research. https://creativecommons.org/licenses/by-nc/4.0/*

[34] *S. Samaratunge and S. Hattotuwa (2014), 'Liking Violence: A study of hate speech on Facebook in Sri Lanka', Centre for Policy Alternatives (CPA). https://www.cpalanka.org/wp-content/uploads/2014/09/Hate-Speech-Executive-Summary.pdf*

According to Bishop (2018), the net result is a gendered splitting of content on social media, in terms of what is most popular and most visible to men and to women, in ways that reflect hegemonic views of beauty, femininity and masculinity.[35] User characteristics, such as age, gender and geography, shape opportunities and vulnerabilities to platform manipulation.[36] Within conflict contexts, it is often women activists, political actors, journalists and women in public spaces including women in the workforce, who are targeted, often with the intention to intimidate and/or silence them. For instance, a recent UNESCO report[37] found that disinformation tactics are routinely deployed in targeted multiplatform online attacks against women journalists. This included attacks on women journalists reporting on far-right extremism or digital conspiracy networks, who were more prone to being targeted. Tactics to malign included misogynistic abuse, harassment and threats against women journalists to undercut public trust in critical journalism and facts in general.[38]

EXAMPLES OF TACTICS EMPLOYED BY ACTORS IN CONFLICT CONTEXTS INCLUDES:[39]

Research shows that conflict actors use various tactics and methods to create and spread online disinformation, some of which includes:

A. **targeting** through disinformation operations that carry out intelligence collection on their target audiences via open-source channels on the web and analysis gathered by digital advertising agencies;
B. operatives **create and curate** emotionally resonant or otherwise inflammatory **content** (audio/visual, text-based information) for weaponisation, including propaganda, misinformation and disinformation employing elite strategists and creatives to use marketing techniques to weaponise popular vernaculars and maximise the reach of social media posts;
C. **leveraging digital influencers**;
D. **amplifying disinformation** through community-level fake operators;
E. **engaging grassroots intermediaries** to provide consistent messaging, including the employment of locally informed creative writers;
F. **employing rumour as a tactic** using unverified information transmitted from one person to others. Rumours can be true, false or a mixture. At their core, mis- and disinformation are rumours; and,
G. use of **hate speech and dangerous speech** to amplify and/or catalyse violence against a person or members of group with shared characteristics such as race, gender, religion, sexual orientation or disability.

Source: Mercy Corps, Weaponisation of Social Media, 2019

Research shows that when these tactics are interwoven with statistics on mobile phone ownership, access to internet and use of social media, it is mostly men and the elite groups within conflict contexts, often individuals or groups

[35] *Diepeveen, S. (2022) 'Hidden in plain sight: how the infrastructure of social media shapes gender norms'. ALIGN Report. London: ODI*

[36] *Mercy Corps (2021), op. cit.*

[37] *Posetti J., Shabbir N., Maynard D. et al. (2021), 'The Chilling: Global trends in online violence against women journalists', UNESCO. https://en.unesco.org/publications/thechilling*

[38] *Ibid.*

[39] *Guay, J. and Gray, S. (2019), op. cit.*

with access and ability to manipulate political and economic decision-making, who would have the time, resources and interest to mobilise these tactics for their personal gain.

In Myanmar, political parties, candidates and their supporters espousing Buddhist extremism deployed hate speech targeting religious minorities on social media in advance of the 2020 elections.[40] During the last uprising majority of protestors out in the frontlines were women in Myanmar. A long history of suppressing women in political and public spaces in Myanmar has meant that women face discrimination in general across Burmese and other ethnic groups, the later having to face dual discrimination as a result of their gender and identity. An example from 2015 shows how women and men are impacted differently: 17 out of 18 candidates from a Rohingya political party were barred from contesting local elections despite having successfully registered in the previous election. The only one Rohingya candidate that was allowed was a woman, solely because her mother's side of the family were Buddhists. Fuelled by increased availability of high-speed internet, radical Buddhist nationalists have leveraged social media to further spread their message, exacerbating Myanmar's ethnic and religious divisions, as fake news, nationalist slogans and viral misinformation have combined in unpredictable and combustible ways to steer political debate that demonises Rohingya men and women.[41] During the campaign period, Muslim Rohingya men and women candidates faced attacks and threats online, while political parties fielding Muslim candidates were targeted as "*not protecting race and religion*."[42] Tactics employed included disinformation, falsely acclaiming that a Muslim candidate wanted Arabic to be taught in schools (Leibowitz, *et al*., 2021). As mentioned earlier, within identity-based conflicts, religion, ethnicity and caste intersect with gender and often do more harm in the real world and transcend to the online. Online hate speech tactics often tend to favour the political elite, mostly older men from Burmese communities, and their interests.

Understanding scale of digital expansion and gender-based online hate speech

Research carried out by CARE India found that 40% of the male and female participants surveyed said that over the last year they had been called offensive names as a result of being a Muslim, and 60% of the respondents reported coming across content on digital platforms stating Muslim immigrants will take over India.[43] Growing polarisation between the Hindu and Muslim communities in India, instigated by Hindu right-wing politics, has witnessed overt online hate speech targeting religious minorities, Muslim men and women as well as anyone who are seen as anti-establishment. For example, a Bengali Muslim migrant worker was brutally murdered on December 2017 in northern India by a man who filmed the entire attack and uploaded it on YouTube. The video went viral along with his speech justifying his act in the name of "*love jihad*",[44] a concept touted by ultra-right Hindu nationalists who claim that Hindu women and girls are being seduced and tricked into marriages and partnerships by Muslim men to convert them. The video including his message met with both negative backlash but also found him supporters, particularly from radicalised Hindu groups who saw him as a hero who acted to stop "*love jihad*" – a divisive term denoting the marriage of a Muslim man with a Hindu woman, popularised on social media and picked up without critique by many

[40] *J. Leibowitz, G. Macdonald, V. Shivaram and S. Vignaraja (2021), 'The Digitalization of Hate Speech in South and Southeast Asia: Conflict-mitigation Approaches', Journal of International Affairs: Georgetown University.*

[41] *Hunter Marston (2020), 'Stirring hatreds ahead of Myanmar elections', The Interpreter: Lowy Institute.*

[42] *Ibid.*

[43] *Roy, T. (2022), 'Hindutva's Circulation of Anti-Muslim Hate Aided by Digital Platforms, Finds Report', The Wire.*

https://thewire.in/communalism/india-anti-muslim-hate-twitter-facebook-whatsapp-hindutva-modi-bjp

[44] *The Wire (2017), ' As Filmed Murder of Muslim Man in Rajasthan Goes Viral, Official Culture of Impunity Towards Hate Crimes Blamed'.*

https://thewire.in/communalism/rajasthan-muslim-man-hacked-to-death-video

in the mainstream.[45] As a result, there were some social media calls to attack Muslim women and girls in order to retaliate, while threats and intimidation against women increased. Crucially, while there were several incendiary tweets across India about so-called "*love jihad"* in the months before the attack, none were geotagged within a hundred kilometres of where the murder took place, implying that the online material that radicalised the perpetrator was unlikely to have been created within his community (Mirchandani, 2018). We see, therefore, the importance of reviewing social media penetration and reach alongside tactics and strategies to fully understand its impact on broader communities.

The Nigerian Stability and Reconciliation Programme (NSRP)[46] reported that 76% of hate speech messages in Nigeria are transmitted through Facebook, either as a post on a private page or in a group. The most prevalent messages call for discrimination (45%), for war (38%) and advocate the killing of others (10%). This online speech tends to be actively recirculated by audiences with over 75% of messages receiving moderate to significant responses and observation (NSRP, 2017). Research showed that women activists at the local and national levels were often most targeted along with political actors that oppose the status quo. Similar to other contexts, misogynistic online abuse, threats and intimidation against women are rife within gender-based online hate speech.

The falling price of SIM cards and the expansion of telecommunications coverage in Myanmar rapidly magnified the risks of hate speech and disinformation (Mercy Corps, 2021). Improved opportunities for digital and network expansion also meant that more users joined cyberspace. This led to heightened risks for women, minorities and LGBTQI community, who saw an increase in social media attacks. Similar to the case above of India, even when internet penetration is low or confined to urban centres, the impact of social media can reach far beyond the user base as mentioned in the research carried out by Mercy Corps where they document cases in which online narratives appeared to "spill over", reaching populations with limited or no internet connection.[47] Understanding gender- and conflict-sensitive patterns of interactions between social media insertion in conflict contexts and how they interact with broader communities is missing from the evidence reviewed here.

The root causes and drivers of hate speech in Myanmar are multiple and originate in both the historical and current context providing space for racist, sexist and religio-nationalist sentiments to remain widespread in popular society.[48] Published literature and evidence shows that hate speech messages often target religious and ethnic minorities as well as women and people from the LGBTQI community.[49] For example, an already vulnerable group in Myanmar, Rohingya men and women minorities face state persecution and systemic violence against their ethnic community.

On the other side, literature coming out from Bangladesh, which continues to host Rohingya refugees, shows that examining the impact of digital disinformation on intercommunity conflicts in Bangladesh, illustrates that digital media is impeding the peaceful coexistence of religious communities, playing a role in inciting aggressive behaviour by Muslims (the dominant religious group) against Hindus and Buddhists (religious minorities), and successfully staging communal violence along religious fault lines.[50] Tactics include coordinated mob violence by the Muslim

---

[45] *Maya Mirchandani (2018), 'Digital hatred, real violence: Majortiarian radicalisation and social media in India', ORF Occasional Paper: 167, Observer Research Foundation. https://www.orfonline.org/research/43665-digital-hatred-real-violence-majoritarian-radicalisation-and-social-media-in-india/#_edn9*

[46] *Nigeria Stability and Reconciliation Programme (2017). 'How-to-guide. Mitigating dangerous speech. Monitoring and countering dangerous speech to reduce violence. http://www.nsrp-nigeria.org/wp-content/uploads/2017/12/NSRP-How-to-Guide-Mitigating-Hate-and-Dangerous-Speech.pdf*

[47] *Mercy Corps (2021), op. cit.*

[48] *Minority Rights Group (2020), 'Online hate speech in Myanmar: an evolving threat'. https://minorityrights.org/2020/12/20/hate-speech-myanmar/*

[49] *Madzima-Bosha, T. (2021), 'Harnessing the power of social media for conflict prevention'. https://www.peaceinsight.org/en/articles/social-media-conflict-prevention/?location=&theme=conflict-prevention-early-warning*

[50] *Udupa, S. et al. (2020), op. cit.*

majority population spurred by a (fake) Facebook post allegedly created by a Hindu fisherman "defaming" Islam. Similarly, a fake Facebook account linked to a young Buddhist man was used to spread a post portraying the desecration of the Quran and subsequently led to the framing of a religious minority that resulted in mass mob violence, even though the post merely tagged the alleged Buddhist perpetrator (i.e. did not even picture him) and featured a pair of white, apparently female feet with painted nails stepping on a Quran.[51] On the one hand, the evidence shows that disinformation is being developed and targeted against religious minorities, inciting physical harm against what appears to be mostly men while using the same online space to reinforce gender stereotypes causing long-term harm to women, girls and LGBTQI people on how they are viewed societally. However, evidence on tactics employed at the local level is weak from conflict contexts and that merits more partnership with local research organisations who have strong gender and conflict expertise to look into this issue.

Online hate speech can also be platform-specific, as identified in the Mercy Corps report. In Nigeria the popularity of WhatsApp has confounded efforts to regulate or counter political manipulation and digital hate speech. Closed WhatsApp discussion groups, in which members are vetted by moderators, reduce opportunities for external monitoring and help ensure the 'echo chamber' effect that contributes to polarisation. Social media influencers who sow online disinformation on behalf of political leaders preferred to use WhatsApp given the potential for creating multiple online identities: given that WhatsApp accounts are tied to mobile phone numbers, an influencer can have as many proxy accounts as they have SIM cards (Mercy Corps, 2021).

From a gendered perspective, in Nigeria's Middle Belt, a focus group of Salafi Muslim women emphasized that social media was a "lifeline" providing social connection for wives confined to the home by conservative husbands, but this dependence also increased susceptibility to online disinformation and misinformation.[52] On the one hand, women are finding ways to navigate their patriarchal culture through social media, especially in conservative contexts; while the same pathway also comes with drawbacks. Managing the two realities while retaining the sense of social connectivity will be challenging but also provides an important perspective on how nuanced social media's presence is in this context.

Implications of online hate speech in conflict contexts

"*In the age of digital capitalism, patriarchy does not even pretend to strike gender bargains. To belong in the privatised quasi-public online spaces, women have no choice but to fall in line and the costs of rebelling against the online norm are brutally punishing, extending beyond the online into the offline*."[53]

Digital spaces have amplified women's voices, and are being employed by women to spearhead their voice and initiative, and increase the spotlight on the work that they do. However, as noted above and substantially recorded in the evidence review, these platforms are also notoriously and increasingly infiltrated by hate speech, including misogynist comments.[54] A recent study[55] found that the spikes in online abuse correspond with events related to feminist movements or gender rights across South Asia (India, Sri Lanka, Bangladesh, Pakistan and Nepal). Although

[51] *Ibid.*

[52] *Ibid.*

[53] *Gurumurthy, A., Jha, B. (2020), op. cit.*

[54] *Dehingia, N., Lundgren, R., Dey, A.K., Raj, A. (2021), 'Trends in online misogyny before and during the COVID-19 pandemic: Analysis of Twitter data from five South-Asian countries', 'Big data and gender in the age of Covid-19: A Brief Series', University of California, San Diego.*

[55] *Ibid.*

seemingly obvious, this backlash is alarming given that such content discredits reports of violence and everyday discrimination.[56] While including this report finding here, it would also be useful to highlight that more research will be required to understand the longitudinal aspect of change and violence caused by online gender-based hate speech – research looking into this was not accessible through secondary data review.

Gender norms on mobility and visibility on the Internet are seamlessly tied to cultures of female responsibility and self-discipline.[57] For instance, in the Indian context, women are told to build walled gardens through technological aids, so that they can remain online and benefit from the wonders of technology without encountering any of the harm.[58] While research shows that young men are more politically active online, fewer young women than men post opinions on social and political issues or take part in online voting.[59] Young women are more likely to censor themselves, as they take into consideration possible backlash for their online political participation.[60] This reveals the magnitude of online violence when considered alongside the fact that girls, women and LGBTQI individuals are disproportionately over-represented as victims of antisocial online behaviour.

As reported on the reviewed evidence, one of the first incidents of inter-group violence in India, where hate speech was circulated on SMSs, MMSs and posts on Facebook, was in 2012.[61] It targeted men and women from northeast India who had moved to cities in the south and west I due to better employment and educational opportunities, experiencing systemic marginalisation for many decades. This online-offline hate speech, which includes stereotyping women and girls from the northeast who face multiple SGBV threats on a regular basis, subsequently led to a mass exodus, with more than 30,000 men and women from northeast ethnic communities leaving Bengaluru within days of online disinformation, leading the government to ban bulk texting (Narrain, 2018).

Women and young girls often have to deal with threats of sexual violence, including rape threats, extortion, publicly sharing of private messages / photos / videos without consent, and cyberstalking through online misogyny. The stigma has even driven young girls to suicide in Sri Lanka.[62] Conflict and sexual violence in Sri Lanka has had a long history: spreading disinformation and targeting women with hate speech, particularly women activists and journalists, has been a repeat pattern over several decades, often along with misogynistic abuse, intimidation and threats of sexual violence. Research on this from a sexual violence in conflict perspective was readily available but the body of evidence is weak on linkages with conflict-related women, peace and security themes.

Unsurprisingly, reports show that gender-based online hate speech often doesn't get taken down even when reported. For instance, attempts to report violations to Facebook are sometimes met with the response that they do not violate community standards. Evidence shows that this is more frequent when the language used is local and not easily picked up by online moderators. For instance, it was reportedly an issue in Sri Lanka, when the content is in Sinhala or Tamil (Perera and Wickrematunge, 2019).

Strong evidence exists on how social media reproduces unrealistic standards of how women and men should appear and behave, and how it manipulates their choices and agency. Evidence also suggests that these reproductions are

---

[56] *Ibid.*

[57] *Gurumurthy, A., Vasudevan, A. (2019), 'Masculinity, Femininity, Equality: Gender Scripts in the Lives of the Born Digital', 'Righting Gender Wrongs' series, IT for Change. https://itforchange.net/masculinity-femininity-equality*

[58] *Ibid.*

[59] *European Institute for Gender Equality (2018), 'Gender Equality and Youth: the opportunities and risks of digitalisation', EIGE.*

[60] *Ibid.*

[61] *Narrain, S. (2018), 'Social Media, Violence and the Law: "Objectionable Material" and the Changing Contours of Hate Speech Regulation in India', Culture Unbound: Journal of Current Cultural Research Volume 10, issue 3, 2018: 388–404. Published by Linköping University Electronic Press. http://www.cultureunbound.ep.liu.se*

[62] *The Sunday Times (2014), 'Facebook – the good, the bad and the ugly'. http://www.sundaytimes.lk/140302/sunday-times-2/facebook-the-good-the-bad-and-the-ugly-87267.html.*

harmful as they stem from patriarchal and unequal gender norms, but what is critically lacking is whether they affect – and to what degree – the political and cultural gains women and minority groups have made over the last decades. This is important as it seeps into political online rhetoric used to target women and ethnic and religious minority groups in public space. For example, in Sri Lanka during the 2018 elections, a comment recorded on the Facebook page of a woman MP comments on her physicality : '*Iresha [name changed] you've put on weight*'.[63] The researchers recorded a high volume of such harassment on the Facebook page of first-time candidate Senani Samaranayake, standing for the Colombo Municipal Council under the Sri Lanka Podujana Peramuna (SLPP).[64]

Another example from Sri Lanka, a hotbed for attacking women and religious minorities, shows how women are forced to experience misogyny when they report against or try to avail justice. In 2018, a human rights activist who pursued a contempt of court case against a Buddhist monk of the Bodu Bala Sena (BBS) was viciously targeted online. One such attack was a text superimposed onto her photo and posted on her public Facebook page: *"For sending this monk to jail, you Tiger prostitute, we curse you"*. (Perera and Wickrematunga, 2019) While there are many examples of tactics and narratives used to target women and minorities, research that looks into the roll-back of women's rights or identity rights and even human rights is so far very limited. This would be an important future research focus from a conflict sensitivity point of view.

Integrating gender perspectives towards solutions and pathways for peace and development

The majority of evidence is divided into two categories: one that looks at solutions to resolve GBV and gender-based online violence and the other that looks at addressing communal and political divisions and conflicts. Evidence that explicitly combines the two to develop recommendations or solutions wasn't available unless inferred through available data. Similar to some of the evidence review findings above, it is possible to collate and combine literature to make inferences based on multiple publications, particularly if research or initiatives have been carried out in the conflict contexts discussed here.

Conflict actors use social media platforms tactically to advance their goals.[65] They may use one platform for disseminating text, another to share video material and a third for internal communication, or they may use one platform to communicate with international audiences and another to communicate with local constituencies, using different languages in each case.[66] A policy paper produced by International Alert[67] discusses how social media can also play a bridging function between local and national spheres to exert political influence: from an activist core to the public, from user-generated content to mainstream mass media, and from local struggles to international

---

[63] *Perera, S. and Wickrematunge, R. (2019), 'Opinions, B*tch: Technology-based Violence Against women in Sri Lanka', Centre for Policy Alternatives: Counterpart International.*

[64] *Ibid.*

[65] *Lanz, D., Eleiba, R. Formica, E., Kavanagh, C. (2021), 'Social Media in Peace Mediation: A Practical Framework', DPPA Mediation Support Unit and Swisspeace.*

[66] *Ibid.*

[67] *International Alert (2020), 'Realising the potential of social media as a tool for building peace: Policy Paper'. https://www.international-alert.org/publications/realising-potential-social-media-tool-building-peace/#:~:text=It%20can%20create%20new%20spaces,activities%20and%20strengthen%20peace%20processes*

attention.[68] Nigerian citizens are utilising social media to document and share their lived realities of conflict, thus shaping the narrative around the actions and narratives of Boko Haram and the Nigerian government.[69]

The degree to which hate speech exists on social media is often undetected by platform, domain or app owners (e.g. Facebook's own hate speech monitoring mechanisms) due to the expression being predominantly in Sinhala (the language used to annotate photos, illustrate videos or draw memes is predominantly if not exclusively Sinhala). This is why content in English that runs completely counter to Facebook's policies around hate speech finds free expression in Sinhala, only subject to scrutiny and compliance when reported by conscientious users.[70] Tracking user behaviours online assumes that they mirror offline behaviours and that they exist in a traceable format. In reality, social media users may go from public social media to private spheres, making it difficult to track for impact and hate speech even when public data is available (R. Brown and L. Livingston, 2019).[71] It is not possible to assume that the mere output – some number of tweets – equals impact.[72]

Other societies have developed unique mechanisms to identify and counter hate speech, which may variously combine customary law and formal law. In Somalia for example, where poetry constitutes a popular vehicle for the dissemination of information and ideas, poets who are seen as repeatedly composing poems which community elders consider to be derogatory of individuals or groups can be banned from composing new work (Stremlau 2012).[73] What could seem traditional and more straightforward has probably not been as simple, with social media companies and their policy on monitoring hate speech and their continually changing parameters of what constitutes hate speech. Integrated gendered pathways could include leveraging social media for gender-sensitive analysis of a conflict-affected context; using gender-responsive social media for communication purposes which includes understanding and responding to conflict parties' use of social media; and finally understanding and addressing social media as a source of mis- and disinformation, all within the context of mediation processes that take into account the role of women and men in peace processes (swisspeace, 2021).

From a gender-aware conflict sensitivity perspective, analysis of local actors in conflict contexts is often complex and comes with potential pitfalls and risks as well as opportunities. Working with local actors is vital particularly due to vernacular challenges of online communication and interaction, which, as mentioned in the report, take place in different languages that are not picked up by social media companies. Equally important is the need to support women and girls in conflict contexts, particularly where women are under-represented in public spaces. Equipping them with digital technology skills, creating an environment for them to employ social media and digital tools to enhance development and peace can lead to sustainable and resilient outcomes.

---

[68] *Aday, I.S. et al (2012), 'Blogs and Bullets II: New media and conflict after the Arab Spring', United States Institute of Peace. https://www.usip.org/publications/2012/07/blogs-and-bullets-ii-new-media-and-conflict-after-arab-spring*

[69] *Ibid.*

[70] *Samaratunge, S. and Hattotuwa, S. (2014), 'Liking Violence: A study of hate speech on Facebook in Sri Lanka', Centre for Policy Alternatives (CPA). https://www.cpalanka.org/wp-content/uploads/2014/09/Hate-Speech-Executive-Summary.pdf*

[71] *International Alert (2020), op. cit.*

[72] *Ibid.*

[73] *I. Gagliardone, et al (2015), 'Countering Online Hate Speech', United Nations Educational, Scientific and Cultural Organization.*