ENABLE PEACEBUILDING THROUGH SAFEGUARDING THE INTEGRITY OF THE INFORMATION SPACE IN TIMES OF CONFLICT

10 POLICY RECOMMENDATIONS TO PLATFORMS AND STATES
NOVEMBER 2023

Forum on Information & Democracy
10 POLICY RECOMMENDATIONS
TO PLATFORMS AND STATES

These priority policy recommendations align with core peacebuilding principles, including conflict prevention and transformation, inclusivity, accountability, transparency, human rights and reconciliation. These recommendations contribute to efforts to prevent conflicts, mitigate polarisation, and promote a balanced and comprehensive understanding of conflicts, which are central to peacebuilding.

As events around the world today illustrate, violent conflicts of a seemingly local, national or cross-border nature play out globally on smart screens. This has the twin effect of raising awareness of the conflict and drawing in more parties to the conflict outside immediate geographical areas. The velocity at which we create, disseminate and consume information dramatically impacts conflict dynamics, challenges notions of who influences on-the-ground actions, global reactions, and positions staked out by allies and foes. The global spread, consumption and abuse of information at velocity now contributes to expanding a conflict or conflict dynamics to seemingly unrelated areas and places beyond the local.

During times of conflict and especially during acute crises, information spreads rapidly, often without filter, analysis and verification, with real-world consequences on civilians and communities. Fear speech¹ exploits grievances and can mobilise conflict parties, while more easily escaping content moderation policies. Misinformation can distort realities, create, reinforce or exacerbate grievances and trigger latent conflict. Disinformation is weaponized and used intentionally to influence the outcome of the conflict, cause harm, incite increased levels of violence, create invented facts on the ground, and shift regional and global public or political opinion. It is often used to fan the flames of hate that encourage retaliation or acts of revenge against communities outside immediate zones of conflict, sometimes geographically very far away.

Risks are high that military decisions, political responses and global engagements quickly spiral based on faulty, biased or invented information. Further, trust can be quickly eroded among conflict parties, national and international humanitarian/development workers, and observers seeking to explain and understand the violence as they are tasked to identify the structures and stakeholders from which peacebuilding can begin. The search for sustainable peace becomes increasingly elusive as people, driven by fear, mis/disinformation or extreme bias, respond with rage and hate.

Unfortunately, research shows that divisive, hateful content and disinformation spread faster than content aimed at seeking to build peace. When emotions are inflamed further by bad actors running bots or by deep fakes generated by AI, the work of peacebuilding becomes more complex and challenging. This creates added financial, human and social costs with people suffering longer as fear, mis/dis and biased information has to be disentangled. As much as access to reliable information is vital in crisis response, to enable people to seek shelter and safety or to know the wellbeing of family, friends and community, reliable and bridging information² is also the basis for peacebuilding processes, consolidating pro-peace

---

¹ “Fear speech is speech that attempts to incite fear about a target community and that has almost zero toxic content, making it look plausible” as in “On the rise of fear speech in online social media”, Punyajoy Saha et al, 2023

² “Bridging information is content that meets approval (or generates positive engagement) across diverse groups of people” as in “The Algorithmic Management of Polarization and Violence on Social Media” by Jonathan Stray, Ravi Iyer, and Helena Puig Larrauri, 2023
policy consensus and ensuring policies are driven by commonly shared facts rather than emotionally charged reactions to faulty information.

These recommendations draw on policy frameworks already developed by the Forum on Information and Democracy\textsuperscript{3} to ensure technology serves citizens, our democratic institutions and integrity of information. They encourage platforms to implement conflict sensitive design principles and practices and states to create an information environment conducive to peacebuilding.

Here we present the Top 10 most salient policy recommendations for States and private companies to ensure the integrity of our information space and thus enable sustainable peacebuilding in the digital age.

The recommendations fall under four categories:

- ensuring reliability and plurality of information about all conflict parties and promoting peacebuilding messages;
- diminishing the reach and effects of dis- and misinformation;
- enhancing accountability of highly influential actors; and
- enabling independent scrutiny of efforts.

\textsuperscript{3} These include including: How to End Infodemics (2020), A New Deal for Journalism (2021), Accountability Regimes for Social Networks and their Users (2022) and Pluralism of Information in Curation and Indexation of Algorithms (2023).
RECOMMENDATIONS

RECOMMENDATIONS TO PLATFORMS

To ensure reliability and plurality of information about all conflict parties and promote peacebuilding messages

1. Promote peacebuilding and public interest journalistic content and access to bridging content and a diversity of viewpoints on the conflict and on capacities for peace.

2. Discourage polarisation and extremist views by prohibiting sponsored campaigns to personalise and target people based on categories inter alia ethnicity, religion, language, sexual orientation, gender identity, political affiliation, and nationality. Similarly, recommender systems must not personalise recommendations based on these categories linked to the conflict.

3. Avoid engagement-driven metrics in recommender systems for conflict-related content and develop alternative engagement possibilities enabling users to flag informative and respectable content.

4. Empower people to make informed choices by showing and explaining to users, in an easy and intelligible manner, why they are targeted with specific information and how they can choose to see different information.

To diminish the reach and effects of dis- and misinformation

5. When content reaches a certain threshold, trigger an internal viral circuit breaker that would temporarily prevent the content from algorithmic amplification. When a piece of dis- or misinformation has been exposed to a large number of users, show a contextual analysis with diverse viewpoints conducted by independent fact checkers to every user exposed to that piece of information.

6. Expand the number of fact-checking initiatives and moderators in all local languages concerned by the conflict and have content policy and moderation teams work with peacebuilders. Develop cross-platform and multilingual solutions to scaling fact-checking efforts that reduce the burden on media and civil society and ensure transparency and auditability.

7. Make it possible for users and other stakeholders to easily report problematic content, including in flagging its potentially harmful impact on the conflict and ensure timely review and action, including reducing virality of so-called edge content that does not quite violate platform policies but nonetheless contributes to reducing capacities for peace.

To enhance accountability of highly influential actors

8. Put in place special accountability regimes for highly influential actors on all sides of the conflict and news media, journalists and influencers, with appropriate sanctions according to published and impartial rules or terms of service and fast tracked appeal mechanisms.
To enable independent scrutiny of efforts

9. Publish terms of service and policies in plain language and provide regular reporting on how algorithmic and AI recommender operations work, content moderation systems and modalities of cooperation with trusted flaggers. This includes transparency about Standard Operating Procedures and policies during times of conflict as well as data on moderation requests (from governments and civil society) and their decisions, the prevalence, treatment and reach of disinformation, fake accounts and bots and the activities of trusted flaggers and machine readable data to allow scrutiny by researchers and regulators.

10. Make transparent, accountable and inclusive the process of drafting, amending, and applying the terms of service agreements, community standards, and content-restriction policies that ensure non-discrimination and transparency in working with civil society and human rights organisations especially during times of crisis.

RECOMMENDATIONS TO STATES

To ensure reliability and plurality of information about all conflict parties

1. Honor all domestic and international commitments to fundamental human rights and close the implementation gap between treaties and the situation on the ground when it comes to freedom of opinion and expression, media freedom, and the safety of journalists.

2. Enhance the plurality and sustainability of funding for media organisations particularly to local journalism and journalism serving underserved, underprivileged and marginalised communities and by putting in place mechanisms to fairly compensate news provided on platforms, to enable media to play its social function.

3. Ensure that public service media are fully independent and funded through mechanisms that protect editorial independence.

4. Encourage the voluntary adoption of open standards for content authenticity and provenance, their implementation on platforms and begin using these technologies when developing and sharing government content online.

To diminish the reach and effects of dis- and misinformation

5. Hold platforms accountable and ensure that sanctions and moderation decisions on platforms are made in accordance with international human rights norms and standards.

6. Impose on platforms a transparency obligation and notification requirement to account holders when implementing sanctions on users.

7. Ensure platforms implement an appeal mechanism for account holders to expeditiously challenge decisions from platforms, by reference to judicial authorities, and in an expedited manner during times of conflict.

8. Hold platforms responsible for complying with decisions from independent regulatory and judicial authorities, and apply sanctions or reinstate accounts and content according to the decisions.
To enable independent scrutiny of efforts

9. Make available publicly and in a regular manner the number and nature of content restrictions, as well as the categories of personal data requested from platforms, and the reasons and legal frameworks that justified such requests, a clearly defined legal basis for request, responses from the companies and exact steps that were taken as a result of requests as well as agreements made with platforms.

10. Establish a new or strengthen an existing independent administrative authority with a democratic mandate to support users of digital services, oversee platforms, serve as a research hub and advise on public policies and regulations, including in conflict times.
Forum on Information & Democracy