How to combat hate speech and gendered disinformation online?

UNESCO dialogue provides some ideas.

Beyond the urgency, focus on long term sustainable responses

> "Faced with this situation, we need firm responses, focused on the long term.
> This is UNESCO's way of acting – and it is our commitment."
> **Audrey Azoulay,** UNESCO Director-General

Several recommendations were made during the global dialogue Online Gendered Disinformation: what are the solutions? by experts from around the world to make digital spaces safe for women and girls, especially given the new and harmful narratives that are emerging at unprecedented speed, "forcing women and girls to self-censor, de-platform or reduce their interaction in online spaces".[1] They will inform the *Guidelines for Regulating Digital Platform: A Multi-stakeholder approach to safeguarding freedom of expression and access to information* that will be debated during its Global Conference Internet For Trust.

The compilation of recommendations made by the panellists of this global dialogue target both government regulators and digital platform companies to urgently address the global problem of online gendered disinformation and gender-based violence using international human rights law and the principles of equality, diversity, transparency, accountability as their compass. Civil society stakeholder groups can use the recommendations below in their advocacy campaigns and accountability actions.

Their recommendations reaffirm that as "women and girls represent half of the world's population and therefore also half of its potential"[2], there is a need to adopt stand alone, targeted measures in the global guidelines for regulating digital platforms to address the specific and daily threats that women and girls, in all their diversity, face online. The goal is to ensure their full participation and contribution to a digital world where they are free from fear of harassment, hate and toxic speech, as well as all other forms of technology facilitated gender-based violence[3] that has severe consequences for their mental well-being and, in some cases, physical security.

Contributing to global guidelines for regulating social platforms

Governments, social media and other digital platforms are called upon to work together in order to:

1. Enhance the accountability of digital platforms towards users by listening to the victims/targets of online harassment and set up mechanisms to ensure the voices of women and girls are heard and that they participate in creating solutions;

2. Provide due process rights for women and girls who are experiencing harassment and threats online, including the right to redress against decisions as well as conduct human rights due diligence with a view to policy, structural and content related change;

3. Introduce sanctions and hold accountable perpetrators of online gender-based violence and gendered disinformation, in line within international human rights law. Impunity must end, requiring specific training for law enforcement, prosecutors, and judges to prosecute such cases;

4. Provide meaningful transparency of processes and data, including algorithms, and come up with tailored solutions for trust and safety for women and girls that is stress-tested / red teamed in advance;

---

[1] 67th CSW Bureau (2023) *Innovation and technological change, and education in the digital age for achieving gender equality and the empowerment of all women and girls* Zero Draft, paragraph 19.

[2] https://www.un.org/sustainabledevelopment/wp-content/uploads/2016/08/5_Why-It-Matters-2020.pdf

[3] 67th CSW Bureau (2023) *Innovation and technological change, and education in the digital age for achieving gender equality and the empowerment of all women and girls* Zero Draft, paragraph 19.

5. Prevent the removal or blocking of legitimate content under international human rights law, including the rights of women and girls to freedom of expression, opinion and creation;

6. Monitor the impact of financial sanctions against digital platforms for enabling online gendered disinformation, hate speech and harassment and build on lessons learned to further hold digital platforms accountable;

7. Conduct annual human rights and gender impact assessments including algorithmic approaches to gender-specific risk assessment, with a view to identify the systemic risks to women and girls and to adjust regulations and practises to mitigate such risks more effectively;

8. Design content moderation rules and practises by investing in understanding local contexts and providing content moderation in local languages that can help address the particular forms of gendered biased speech that may be hidden in jokes, puns, cultural expressions of different communities that are not detectable by machines or human content moderators from abroad. In addition, investigate the working conditions, training and resources available for women who are content moderators and regularly monitor their mental health and well being;

9. Establish independent, inclusive oversight boards with the power and resources to influence rules, regulations and practices of individual digital platform companies and mandate them to review transparency reports and interrogate the results against external data-based evidence regarding gendered disinformation, online gender-based hate and toxic speech;

10. Use privacy enhancing technology to provide external and independent researchers access to internal data of platforms to help identify patterns of gendered disinformation, gender-based harassment, hate speech and toxic speech;

11. Create dedicated AI engineering teams that are made up of both men and women who are specifically trained to develop algorithmic solutions to different forms of gendered disinformation, including violent and other forms of toxic speech and harmful, stereotypical content;

12. Develop and launch inclusive structured community feedback mechanisms to eliminate gender bias in generative AI and generative algorithms producing content that perpetuates or creates gendered disinformation, harmful or stereotypical content;

13. Invest in digital, media and information literacy campaigns and education programmes to raise awareness of the impact of online harassment towards women and girls, to produce guidance on navigating online spaces safely and to build a culture of solidarity against online gendered disinformation and gender-based violence.

By implementing global guidelines for regulating digital platforms based on international human rights laws, the global community is taking one large step forward to ensure that the threats faced by women and girls around the world each day are not left behind in global governance of the digital ecosystem.

Useful links:

- How to combat hate speech and gendered disinformation online? UNESCO dialogue provides some ideas.
- Global Dialogue – Online Gendered Disinformation: what are the solutions?
- UNESCO's Internet for Trust conference – Register now
- UNESCO Guidelines for Regulating Digital Platforms – draft version 2.0
- UNESCO's work on the Safety of Women Journalists
- The Chilling: Global trends in online violence against women journalists
- UNESCO Recommendation on the Ethics of AI