**CI/FEJ contributions to the call for inputs on the Special Rapporteur’s**

**Report on freedom of expression and the gender dimensions of disinformation**

1. Conceptual issues

* What do you consider to be ‘gendered disinformation’?
* How is ‘gendered disinformation’ similar to or different from online gender-based violence?

In her [2021 report on Gender Justice and the Right to Freedom of Opinion and Expression](https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/212/16/PDF/N2121216.pdf?OpenElement), the Special Rapporteur referred to disinformation as a factor limiting women’s freedom of expression: “In the digital age, the spate of online violence, hate speech and disinformation often compel women to self-censor, limit what they post or leave platforms.” It also revealed that gendered disinformation is on the rise, making reference to the Special Rapporteur’s [2021 report on Disinformation and freedom of opinion and expression](https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement).

The report further distinguishes gendered disinformation from gender-based violence: “While it is a subset of gender-based violence, it has some distinct characteristics, using ‘false or misleading gender and sex-based narratives against women, often with some degree of coordination, aimed at deterring women from participating in the public sphere. It combines three defining characteristics of online disinformation: falsity, malign intent, and coordination.’ It is often intersectional in nature, promoting both sex and race-based narratives, and used as part of issue-based campaigns to undermine public trust. Research data indicates that female politicians, especially those who speak out on feminist issues or come from racial, ethnic, religious or minority groups are targeted with much higher levels of disinformation than their male counterparts.”

This definition is common to that elaborated in[*The Chilling: A Global Study On Online Violence Against Women Journalists*](https://www.icfj.org/sites/default/files/2023-02/ICFJ%20Unesco_TheChilling_OnlineViolence.pdf)(2022), a report by the International Center for Journalists that was supported by UNESCO. The authors of *The Chilling* emphasize the coordination element in defining disinformation, based on responses of over 900 surveyed journalists based on their experience of gender-based online violence. According to the study, disinformation may come in the form of false accusations of professional misconduct, the spread of smears about female journalists’ character designed to damage their personal reputations and professional standing, and malicious misrepresentation (e.g., synthetic media such as deepfake porn videos, abusive memes, manipulated images).

Finally, *The Chilling* describes a threefold function of disinformation in gendered online violence, rendering it not only a danger in itself, but also as a factor provoking gendered online violence:

1. Disinformation tactics are routinely deployed in targeted multiplatform online attacks against women journalists;

2. Reporting on disinformation and intertwined issues, such as digital conspiracy networks, conflicts, and far right extremism, is a trigger for heightened attacks on women journalists;

3. Disinformation purveyors operationalise misogynistic abuse, harassment and threats against women journalists to undercut public trust in critical journalism and facts in general.

“Attacking women journalists is a favoured tactic of disinformation agents; **disinformation methods are co-opted in misogynistic attacks**; and reporting on disinformation and associated themes (e.g., digital conspiracy communities; far-right networks) is a lightning rod for attacks (Posetti, 2018a; Posetti, 2020b; Bontcheva and Posetti, 2020; Jankowicz et al., 2021). In the words of UK columnist and investigative reporter Carole Cadwalladr who writes for the Observer: **“If you report on disinformation, you become a target of disinformation"**. At the global level, 16% (n=114) of the women journalists surveyed identified disinformation as a beat likely to lead to an increase in online violence.”

As a result, 30% of the women journalists surveyed by UNESCO reported self-censoring on social media and 20% withdrew from all online interaction due to harassment and threats.

1. Responses

* What measures have States, digital companies or international organizations taken to combat ‘gendered disinformation’?
* To what extent do these responses comply with international human rights law standards, in particular freedom of expression?
* How effective have these measures been in addressing ‘gendered disinformation’?
* Please provide references / links to legal or policy frameworks developed to address the gender dimensions of disinformation.
* Please provide references / links to relevant case law concerning this topic.

*Countering gendered disinformation through regulation.*

Disinformation is difficult to tackle as it cannot be restricted under international human rights law simply for being false, and laws set to restrict disinformation quite often are not alligned with international freedom of expression standards. It is therefore important that public policy that aims to counter gendered disinformation has a comprehensive multistakeholder approach involving educational, cultural and rights-based regulatory solutions aligned with the international human rights standards. Regulation of gendered disinformation as a solo strategy could increase harms and reduce participation of women as well as of other vulnerable groups in the digital civil space.

Yet, when talking about regulatory solutions to counter harmful content online such as gendered disinformation, UNESCO’s [Guidelines for Regulating Digital Platforms](https://unesdoc.unesco.org/ark:/48223/pf0000384031.locale=en) approach is one that aims to safeguard freedom of expression and access to information and focuses on the structures and processes to curate and moderate content rather than in specific pieces of content. UNESCO believes that any regulatory process that attempts to deal with harmful content online should have five key characteristics: 1) Has a legitimate purpose, 2) Has a rights-based approach. Ensure that any restriction of content is made on the basis of the conditions established under international human rights law. 3) Is produced in open, transparent, and evidence-based manner. 4) Allows multistakeholder meaningful participation. 5) Focuses on the systems and processes used by platforms to moderate and curate content.

Specifically, a multistakeholder approach within regulation means that whichever regulatory arrangement is chosen in a specific jurisdiction, all stakeholders share responsibility for sustaining an enabling environment for freedom of expression and the right to information, while ensuring there is an open, safe and secure environment for users and non-users and specific consideration should be given to vulnerable groups, including vulnerable groups of women and girls. Therefore, participation of the different stakeholder groups should be considered in the complete regulatory cycle.

Therefore, on the Guidelines, UNESCO outlines a set of governmental responsibilities to safeguard freedom of expression and information online including among others, promoting and implemented different Media and Information Literacy Policies or by discouraging , including through measures such as professional codes of conduct-, public officials from disseminating disinformation, and adopt measures aimed to avoid expressions amounting to advocacy of hatred that constitute incitement to discrimination, hostility or violence, as prohibited under international.

The Guidelines also call for the digital platforms to fulfill 5 principles:

1. Platforms conduct human rights due diligence, evaluating the risks and impact on human rights of their policies and practices as well as defining the mitigation measures.

2. Platforms should adhere to the International Human Rights standards, including in platform design, content moderation and curation.

3. Platforms are transparent, being open about how they operate, with understandable and auditable policies as well as multistakeholder-agree metrics for evaluating performance.

4. Platforms make available information and tools for users to understand and make informed decisions about the digital services they use.

5. Platforms are accountable to relevant stakeholders

Finally, as outlined in the Guidelines UNESCO strongly believes that any regulatory arrangement for digital platforms must ensure independence of the regulatory entity. For instance, in the cases of statutory regulation, independence must be reflected in the constitution and in the powers of the regulator, when self-regulation is implemented, there should also be independent oversight mechanisms and accountability from those.

*Targeting gendered disinformation through emerging technologies.*

New technology, in the form of AI and machine learning models, is providing some solutions to recognize hate and toxic speech characteristic of online gendered disinformation campaigns. There are, however, shortcomings inherent in pure AI-based models. Human review alongside the use of automated tools to detect violent content targeting women and girls in all their diversity will be necessary to reverse gender biases and stereotypes and take into account cultural and linguistic contexts and nuances.

“There has to be a significant degree of human input into these kinds of detection and reporting mechanisms and enforcement overall.”

Dhanaraj Thakur, Research Director, Center for Democracy and Technology, USA & Member of the Advisory board of #Shepersisted[[1]](#footnote-2)

Monitoring the impact of financial sanctions against digital platforms for enabling online gendered disinformation and build on lessons learned to further hold digital platforms accountable could also reinforce the effectiveness of mechanisms for users to report abuse and prevent platforms from deliberately allowing gendered disinformation in order to meet ‘bottom lines’ and investor demands.

*There is a massive problem related to the business model of social media platforms because online disinformation is being “monetised” by many of these companies and by many other smaller players who are selling “followers lists” to attack women and others. This is a huge market. It is hard to ask the platforms to self-regulate because it is not in their interest to go against very serious, yet profitable, disinformation campaigns.*

*Octavio Kulesz, digital publisher and philosopher, Argentina*

Furthermore, privacy enhancing technology could be used to provide external and independent researchers access to internal data of platforms to help identify patterns of gendered disinformation.

Last, the development of inclusive structured community feedback mechanisms could contribute to eliminate gender bias in generative AI and generative algorithms producing content that perpetuates or creates gendered disinformation.

*Promoting diversity in media and social media teams.*

Social media reflects hetero-cis normative culture in both their technological and human decisions. Decisions made by social media platforms (even in the design phase) strengthen patterns to define the way content has been managed, enforcing gender narratives and leading to gendered disinformation. Thus, there should be greater diversity among those building emerging technologies. Currently, only 12% of machine learning researchers and 22% of all AI professional developers are women.[[2]](#footnote-3) Solutions must be found to these barriers of entry to obtain a gender-balanced perspective to existing tools and to content moderation more specifically, in addition to gathering community feedback in an inclusive manner.

In media outlets, this problem remains and impacts the way women journalist do their work. According to a survey underpinning [The Chilling](https://www.icfj.org/sites/default/files/2023-02/ICFJ%20Unesco_TheChilling_OnlineViolence.pdf), newsrooms and media companies employing women journalists have been falling short of preventing and protecting journalists from the harms of gender-based disinformation. While in recent years progress has been made by many employers, respondents maintain that reporting mechanisms were insufficiently installed and responses to such reports were weak or lacking. Only 25% (n=179) of the women survey respondents said they had reported online violence incidents to their employers, and among these, some received no response at all why others were given discouraging responses.

Gaps therefore remain in news industry responses, with many media organizations still lacking formal response protocols and the integration of holistic strategies that blend digital security, physical safety, psychological support, upward referral mechanisms and gender-sensitive editorial policies. Even in those newsrooms where a formal protocol for responding to online violence was in place, the study revealed a lack of understanding of the psychological harm caused by gendered and intersectional online violence, and a corresponding failure to address the mental health consequences for those targeted.

1. Solutions

* What recommendations do you think the Special Rapporteur should make and to whom on combating gendered disinformation?
* What issues or areas of gendered disinformation require further research in your opinion?

Please provide references or links to relevant research or reports.

Drawing on the draft UNESCO [*Guidelines for Digital Platform Regulation*](https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000384031&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_c084184b-eca2-416c-b937-f4b6254db7f6%3F_%3D384031eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000384031/PDF/384031eng.pdf#%5B%7B%22num%22%3A37%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C69%2C254%2C0%5D), created through various phases of multistakeholder consultations and including the work done on gendered disinformation in the framework of the 2023 International day of women in multilateralism, the Special Rapporteur could recommend that digital platforms conduct human rights due diligence for the human rights impacts or their policies and practices on all vulnerable and marginalized groups, including due to online gender-based violence.

Digital platforms may thus be recommended to :

* 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 policies and practices to mitigate such risks more effectively.
* Use privacy-protecting technology to provide external researchers access to internal data of platforms to help identify algorithmic amplification of gender-based violence;
* Create dedicated engineering teams that are made up of both men and women who are specifically trained to develop algorithmic solutions to content moderation.
* Develop and launch inclusive structured community feedback mechanisms to eliminate gender bias in generative AI and generative algorithms producing content that perpetuates or creates gender-based violence.

These recommendations were inspired by a global dialogue on gendered disinformation that was organized by UNESCO in January 2023[[3]](#footnote-4).

The [recommendations that emerged out of *The Chilling*](https://unesdoc.unesco.org/ark:/48223/pf0000383788), tailored to a variety of stakeholders including news organizations, big tech, governments, NGOs and INGOs, lawmakers, also elaborate possible steps to finding solutions to online violence, including gendered disinformation. Some salient points addressing gendered disinformation in particular include:

* The need for internet companies to develop markers for abuse perpetrator accounts and systems to identify disinformation purveyors
* The development of newsroom protocols designed to identify, monitor, prevent and respond to online violence. They should be sensitive to intersectional threats, and it should also be regularly reviewed so that it is responsive to the changing nature of gendered online violence and the technology used to perpetuate it. These protocols should take account of contexts of weaponized social media platforms, viral disinformation, far right extremism and conspiracy networks.

UNESCO’s work on the safety of women journalists will continue to include gendered disinformation. In 2022, on the occasion of the 10th Anniversary of the UN Plan of Action on the Safety of Journalists and the Issue of Impunity, a roadmap was agreed upon for the next ten years and beyond, to build upon successes and address the emerging risks. It was built upon a high-level ministerial conference which took place in Vienna, Austria, as well as a series of [multistakeholder consultations](https://unesdoc.unesco.org/ark:/48223/pf0000383337) including one on the safety of women journalists. This consultation reaffirmed the need for solutions to online gender-based violence, including disinformation, and among its recommendations cited training women journalists on online safety, training law enforcement on how to investigate and process complaints, and improving transparency around the tech solutions and tools that women may use to protect themselves online.

Finally, it is important to recognize the role that improved media and information literacy plays in equipping citizens with the knowledge and tools to recognize disinformation, including gendered disinformation, and prevent its harmful effects. To this end, governments must take steps to develop national policy and strategy on media and information literacy, which should be included in school curricula to allow future generations of internet users to make informed choices about how they participate in and understand freedom of expression, dialogue, and access to information. In particular, governments should invest in digital, media and information literacy campaigns and education programmes to raise awareness of the impact of online gendered disinformation, to produce guidance on navigating online spaces safely, to build emotional resilience against harassment online and a culture of solidarity among women and between men and women to call out online gendered disinformation.

While UNESCO is working with its stakeholders to provide solutions to gendered disinformation, there is a need for further data on the violence and discrimination which vulnerable and marginalized groups face online, including gendered disinformation, and the role of States, companies and civil society organizations in countering such violence. One crucial area requiring further research is the role of generative AI in producing or countering gendered disinformation.

1. On the occasion of the UNESCO [Global Dialogue](https://www.unesco.org/en/articles/global-dialogue-online-gendered-disinformation) bringing together public, private and civil society experts to reflect on innovative solutions, [International Day of Women in Multilateralism](https://www.unesco.org/en/days/women-multilateralism), 25th January 2023. [↑](#footnote-ref-2)
2. World Economic Forum, 2018, Global Gender Gap Report. https://www.weforum.org/agenda/2018/12/artificial-intelligence-ai-gender-gap-workplace/ [↑](#footnote-ref-3)
3. Outcomes of the Global Dialogue Online Gendered Disinformation: what are the solutions? https://articles.unesco.org/sites/default/files/medias/fichiers/2023/02/OnlineGenderedDisInformation-BeyondTheUrgency.pdf [↑](#footnote-ref-4)