

SAIFE session proposal IGF
Background Document
Spotlight on AI-Driven Content Governance in Times of Crisis

The free flow of information and access to accurate, diverse and public interest content is crucial at all times. It becomes vital at times of crisis – be it a pandemic, climate crisis, or conflict.

In today's information environment, content governance by online platforms plays a significant role in shaping and arbitrating political and public discourse. The way online information is curated and moderated directly individual human rights as our collective rights to global peace and security. It is therefore essential that policies – and their enforcement – are in line with international human rights standards.

Artificial intelligence and machine-learning technologies increasingly decide which content is removed, what content is prioritized or to whom it is disseminated. These decisions directly influence individuals' right to seek, impart and receive information – and ultimately, their opinions. But AI tools are regularly deployed with little or no transparency, accountability or public oversight.

While this already poses many human rights challenges, these become even more detrimental, or disastrous, in times of crisis. Automated content governance impacts the entire conflict cycle. The amplifying power of AI in content curation can contribute to rising tensions, radicalization and hatred in the heat-up of conflict. During conflict, many of the unintended consequences of automated content moderation become detrimental – by potentially silencing critical voices online, at a time when their voices and their safety are essential. The same technologies are then also used to surveil critical voices, and to promote violent agendas. All too often, it is leveraged for digital authoritarianism. Moreover, AI-based content governance can hamper reconciliation by magnifying disinformation and other polarizing content. This undermines and delays post-conflict recovery processes, and thwarts peace movements.

Furthermore, AI facilitates the constant observation and analyzing of data in order to personalize and target content and advertising. The subsequent personalized online experiences risk fragmenting online information spaces and limit individuals' exposure to a diversity of information, which infringes upon principles of media pluralism. A lack of information pluralism provides a perfect breeding ground for manipulation and deception – furthering inequalities, undermining democratic debates, and at times used for digital authoritarianism, fueling hatred, violence and propaganda for war.

The [OSCE Policy Manual on the Impact of Artificial Intelligence on Freedom of Expression](#) (SAIFE) serves as general guidance to states, on regulation of AI in order to protect freedom of expression and other human rights, as well as media pluralism. Over the last several months, the OSCE Representative on Freedom of the Media has explored its applicability in times of crises, to ensure there are no tradeoffs to human rights. This session will present and discuss key findings of this research and provide concrete recommendations on protecting freedom of expression and other human rights in the use of AI for content governance in times of crises.

Session contact:

Deniz Wagner, Adviser to the OSCE Representative on Freedom of the Media

Deniz.Wagner@osce.org +436648590956

and

Julia Haas, Project Officer, Office of the OSCE Representative on Freedom of the Media

Julia.haas@osce.org

Expected Outcomes

The aim of the session is to present OSCE key recommendations for the protection of freedom of expression and other human rights in the context of content governance during crisis situations.

Policy Questions

- To what extent are international standards and guidance on AI and freedom of expression applicable in times of crises?
- Which situations require a tailored approach to safeguarding free speech in the use of AI?
- How can and should international standards on AI and freedom of expression be tailored to crisis situations?

Format

Debate – Auditorium – 60 min

Hybrid Session

- How will you facilitate interaction between onsite and online speakers and attendees?
Two facilitating moderators, both on-site and online to ensure inclusive debates. Speakers/key contributors will be both on-site and online.
- How will you design the session to ensure the best possible experience for online and onsite participants?
Interactive engagement is at the core of this session. Key contributors will set the scene but the focus is on moderated discussions rather than a formal panel set-up.
- Please note any complementary online tools/platforms you plan to use to increase participation and interaction during the session.
Joint collaborative document (google doc) and/or MIRO board.

Approach to Diversity – experience, gender, geographic area, stakeholder group

The session will include speakers from different geographic areas and backgrounds with varying experiences, including high-level eminent experts. It is aimed at discussing diverse crisis situations to ensure an exchange of experiences and building on lessons learned from different regions, emergency situations and conflict areas to identify common patterns and necessary policy responses to ensure freedom of expression and media freedom at all times. It will pay particular attention to power imbalances and the uneven distribution of resources in addressing emerging crisis situations around the globe.

SDG link – 10, 16.10

Comprehensive security, lasting peace, and sustainable development necessitate that human rights including freedom of expression and media freedom are respected, protected and fulfilled at all times, including in emergency situations. While technologies provide ample opportunities for increasing access to information and freedom of expression in times of crisis, it is essential to address the challenges to human rights posed by use of artificial intelligence and machine-learning technologies for shaping and arbitrating information spaces.

Speakers:

1. **Teresa Ribeiro**, OSCE Representative on Freedom of the Media
2. **Tetiana Avdieieva**, Digital Security Lab Ukraine
3. **Arzu Geybulla**, Journalist Azerbaijan
4. **Courtney Radsch**, Media Freedom Expert
5. **Marwa Fatafta** or **Eliska Pirkova**, Access Now
6. **Matthias Kettemann**, Professor of Law Austria

Moderation: Deniz Wagner or Julia Haas, OSCE deniz.wagner@osce.org / julia.haas@osce.org