SAIFE MASTERCLASS

SPOTLIGHT ON

Artificial Intelligence & Freedom of Expression

This OSCE masterclass focuses on addressing the human rights impact of the use of AI in content moderation and content curation by large online platforms. It is part of the OSCE RFoM project putting a *Spotlight on Artificial Intelligence and Freedom of Expression* (SAIFE). Following the launch of the SAIFE Policy Manual in early 2022, the masterclass aims to exemplify the impact of AI-based content governance and surveillance-based business models of online platforms on opinion shaping, freedom of expression, media pluralism, and democratic processes. It explores necessary safeguards for building a healthier online information ecosystem.

Through interactive discussions, presentations and group work, the masterclass seeks to build capacity of local stakeholders in view of addressing the adverse impact of AI-based content governance on human rights, peace and security.

The masterclass consists of three components. The first component addresses the general landscape of our online information ecosystem, focusing on the impact of AI and human rights policy aspects. Alongside the socio-technological landscape, the first component also lays down the international standards for freedom of expression and their interplay with the global socio-technological dynamic, before focusing on the role of state authorities and media governance regimes in mitigating and decreasing adverse human rights impact. This part aims at advancing participants' knowledge of the fundamental human rights framework on freedom of expression, specifically as it relates to the broad concepts of 'information gatekeeping' and 'content governance'. In order to learn more about how these standards interact with content governance policies and practices and particularly AI-driven processes applied by large social media platforms, the masterclass will exemplify current challenges and unfold key principles such as transparency, accountability, human rights due diligence, and public oversight.

The second component proceeds to highlight regional-specific dynamics and the impact of the global content governance challenges on the individual opinion-shaping process, public debate, and media pluralism. It particularly addresses the specific position and responsibility that different state and non-state actors should play in order to prevent and overcome the

shortcomings of algorithmic and AI-based content governance, as well as the challenges stemming from the gatekeeping role of a handful large online platforms that prioritize user engagement over public interest.

The third component contains interactive group work centering on case studies, in view of identifying safeguards for addressing freedom of expression challenges in today's digital ecosystem, and ultimately fostering an enabling environment for freedom of expression and media freedom in the age of algorithms and AI. This part should help participants to critically assess the roles of multiple actors, including whether and how to overcome current shortcomings concerning their respective – and differing – responsibilities as they relate to freedom of expression online. Following the group work, the masterclass will bring together outcomes to jointly develop recommendations and ways forward.

The training was developed jointly with the independent experts Prof. Krisztina Rozgonyi, Prof. Noah Giansiracusa, Bojana Kostic, and local experts for the regional-specific components.