

Opening remarks for the Expert Meeting on Deepfake News – Artificial Intelligence and Disinformation as a Multilateral Policy Challenge

7 December 2021



The Representative on
Freedom of the Media

OSCE Representative on Freedom of the Media
Ms. Teresa Ribeiro

Dear colleagues, distinguished delegates, ladies and gentlemen,

It is my great pleasure to welcome you to our expert meeting on “**Deepfake News – Artificial Intelligence and Disinformation as a Multilateral Policy Challenge**”, in our fourth roundtable dedicated to addressing the issue of disinformation, which remains a **timely and important** subject – including in the online context.

In today’s digital age, information and opinions are increasingly accessed, shared and received online, a trend which has been further accelerated by the global pandemic. Internet intermediaries have become **powerful actors, gatekeepers** to our information, and they deploy automation and **artificial intelligence** to curate and moderate online information spaces. Consequently, their **architecture and policies** shaping the information landscape are immensely relevant to our freedom of opinion and expression, and our common and comprehensive security.

My Office works on the impact of artificial intelligence on freedom of expression, by putting a spotlight on it in our **SAIFE** project. For the protection of our human rights, and for regional security, it matters enormously *how* and *why* certain information is more available than others. It matters when illegal or harmful content such as disinformation spreads online, when it discriminates against marginalized groups, or when it creates divides in our societies. It matters whether and how fact-checked, accurate and edited news and information of public interest is made available. All in all, it matters that we are all able to seek, receive, and impart the **same kind of information**, and to **access accurate, fact-checked and reliable information**, no matter who we are or where in the world we might be. This is the very foundation of how we as individuals form our opinions; it is the foundation of public debate and how we develop as democratic societies.

Today, however, algorithmic content recommendation and filter systems determine *who* sees *what* and *when*. These **AI-based tools** are not designed for diversity, accuracy or the public good. On the contrary, they are designed to maximize **user engagement** in order to increase advertising and thus profit. This means that **data harvesting and advertising-based business models** impact content curation. Various studies have shown how this leads to problematic, polarizing and radicalizing **filter bubbles** and echo chambers as well as to the **amplification of disinformation** and deceptive content.

What worries me further is if news and media content is subject to the same logic of maximizing profit, irrespective of fact-checking, or **professional journalistic standards**, as this leads to the prioritization of “clickworthy” over newsworthy content. And we all know that independent, diverse quality journalism is the best antidote to disinformation, including online.

Alongside this, online disinformation has severe **silencing effects**. It is important to underline the particular targeting of women and marginalized voices in this regard, not least because we are in the midst of the 16 days campaign to eliminate violence against women, but also because, unfortunately, disinformation is often **gendered, intersecting with various forms of discrimination** that can hinder women from exercising their right to freedom of expression. The aim of disinformation campaigns, in particular those targeting women journalists, is to deny them a public voice, and to undermine their credibility, and the credibility in facts. Recent **Nobel Peace Prize laureate Maria Ressa** continuously demonstrated this link between platforms’ business models, automation, disinformation and attacks against independent voices. Let me use her words: without facts, we cannot have truth; without truth, we cannot have trust or a shared reality; and if we don’t have a shared reality, we cannot have functioning democracies or solve any of the complex problems we are facing, be it climate change or the COVID-19 pandemic.

The **pandemic** has clearly highlighted the **impact** that online falsities and disinformation can have on our individual and public health, as well as on the health of our democracies.

The online “**infodemic**” accompanying the pandemic has caused internet intermediaries to take unprecedented steps to address false and misleading information, and inauthentic behavior, including through automated tools. But while we all agree on the need to address online disinformation, we also know it is a **complex endeavor**. Tackling disinformation requires different responses, depending on whether it incites violence, discriminates or offends others. The **use of AI to detect** these different categories of online falsehood brings with it an **additional** set of **challenges** to our **human rights**.

This holds particularly true as AI is **context-blind** and cannot analyze or understand the nuances of human speech. Disinformation, however, is highly context-dependent. The use of AI to detect disinformation thus leads to **false positives** and **false negatives**, which limits legitimate speech or fails to address the harms of disinformation that chill the ability of targeted individuals or communities to engage online. In short, **AI is no silver bullet**. We need to ensure that any use of automation impacting our information builds on **human rights due diligence, transparency and accountability**.

Here I see the clear **role of States** in ensuring **procedural safeguards**. States have a positive obligation to protect and promote human rights, including online. But while we need to find **entry points for States**, unfortunately, State regulation also entails its own risks, especially as we have seen how fragile our democracies and systems of rule of law can be. At the same time, States relying on intermediaries to “solve” online disinformation and **handing over human rights protection** to private actors that act with little or no accountability and use error-prone

automation falls short of the State obligation to protect freedom of expression. There are numerous **layers of regulatory responses needed**: One aspect is addressing the **failures of business practices** of certain platforms, another is the need for enforcing **transparency, human rights impact assessments and procedural and remedial safeguards**, and a third is to **moderate illegal and harmful content online**, within international human rights frameworks.

Less than two weeks ago, in my report to the **OSCE Permanent Council**, I identified the need to address harmful business practices that rely on the use of AI to shape and arbitrate content online. We need to address the lack of transparency, independent oversight and accountability. This is key in order to effectively address online disinformation. In addition, we also have to look at the **digital ecosystem** in which falsities and lies spread, or are even promoted.

Today, we want to discuss how the **AI-based digital ecosystem fuels disinformation** and what is needed to address this on a structural level, even if this means redefining the data-harvesting advertising models.

We also want to speak about whether and how **AI can be used to address disinformation**, in content moderation or as tool to detect manufactured content, to evidence human rights abuses or to analyze, identify and address online disinformation in a human rights-friendly way.

I am thrilled to have such renowned and excellent **speakers** with us today, and I am sure that today's presentations will provide the participating States with useful information that will help to ensure that the use of AI is conducive to our digital rights, and does not harm online information spaces, to the detriment of human rights and comprehensive security.

There is a clear need to continue talking about disinformation, its challenges, and the nexus to technologies such as AI and the overall digital ecosystem – and about the impact these phenomena have on our right to form and hold opinions, and to seek, receive and impart information. I am confident that today's event will help us connect all these dots together and to identify much needed responses to disinformation in the age of algorithms and AI.

I thank you for your attention and look forward to today's discussions.