osce

the representative on freedom of the media

THERE IS NO SECURITY WITHOUT MEDIA FREEDOM

SPOTLIGHT ON
ARTIFICIAL INTELLIGENCE AND
FREEDOM OF EXPRESSION

Impact of the Use of Artificial Intelligence on Freedom of Expression

- Artifical intelligence shaping and arbitrating information spaces

- Content governance
  - AI in content moderation
  - AI in content curation

AI in Content Moderation

AI in Content Curation

Key recommendations for OSCE participating States
Transparency, accountability & public oversight

1. Protect and promote freedom of expression and other human rights as the centre of AI-related strategies and policies
2. Preserve and foster the internet as a space for democratic participation and representation and for media pluralism
3. Develop evidence-based policies, built on inclusive processes, to respond to challenges to freedom of opinion, freedom of information and freedom of expression
4. Promote compliance with the UN Guiding Principles on Business and Human Rights, to prevent the prioritisation of profit maximisation at the expense of human rights and democratic values
5. Oblige online platforms to conduct human rights due diligence, including through human rights impact assessments (HRIAs) for their content governance policies and automated decision-making, as well as for their business practices, such as data harvesting, targeted advertising and interface design
6. Enforce clarity, explainability, and accessibility on the use of AI for content moderation, content curation and targeted advertising
7. Ensure that human rights protections are not fully outsourced or automated, and provide transparency about any public-private-partnerships
8. Enact strong transparency frameworks, including by mandating comprehensive transparency reports that contain detailed information on the use of AI
9. Make certain that robust remedy mechanisms against censorship and surveillance power are in place, including through human review and independent appeal mechanisms
10. Guarantee strong accountability, including through independent oversight and independent auditing, particularly of compliance with human rights and non-discrimination
11. Respect the right to privacy and data protection, including by identifying limits to surveillance-based advertising and by ensuring robust transparency and user agency in tracking- and profiling-based business practices
12. Promote media and digital literacy and strengthen users' empowerment, agency and control over content governance and the use of their data, including by providing the possibility to opt-out of all automated decision-making
13. Address unbalanced and monopolised market powers and promote plurality, and technological and media innovation
14. Engage on the multilateral level to ensure human rights s afeguards in the development and deployment of AI for online content curation and content moderation

What can NHRIs do?

Imagine you're writing an article about photographs that changed the history of warfare. Your article is shared widely on Facebook. One of the images you list in your article is a Pulitzer prize winning image, but Facebook's algorithm cannot distinguish between child pornography and famous war photographs and blocks your post due to the social media platform's community standards. How does this affect your right to free speech?

Imagine you posted satirical content on a social media platform, on a matter of public interest, and the company's machine learning algorithms have blocked your post. The platform does not offer any kind of remedy nor provide you with any insight into their algorithm-based decision-making. How does this affect your right to free speech online?

What can NHRIs do?

Imagine you're covering a conflict zone and are suddenly blocked from posting videos and images on a social media platform because its machine learning algorithms detect violent content. Your visual documentation related to human rights violations and other crimes committed by all sides of the conflict is gone, and you can no longer reply on uploading your content to social media as an evidence-based tool for reporting, advocacy and accountability purposes. What do you do?

What can NHRIs do?

Imagine you're an investigate journalist receiving threats like "I know where your kids go to school" but the content moderating algorithms don't delete the reported posts because they cannot detect the harm. Your kids will see them too, and the police will not launch an investigation because it's only on social media, and therefore not an imminent threat. Will you continue your work?

What can NHRIs do?

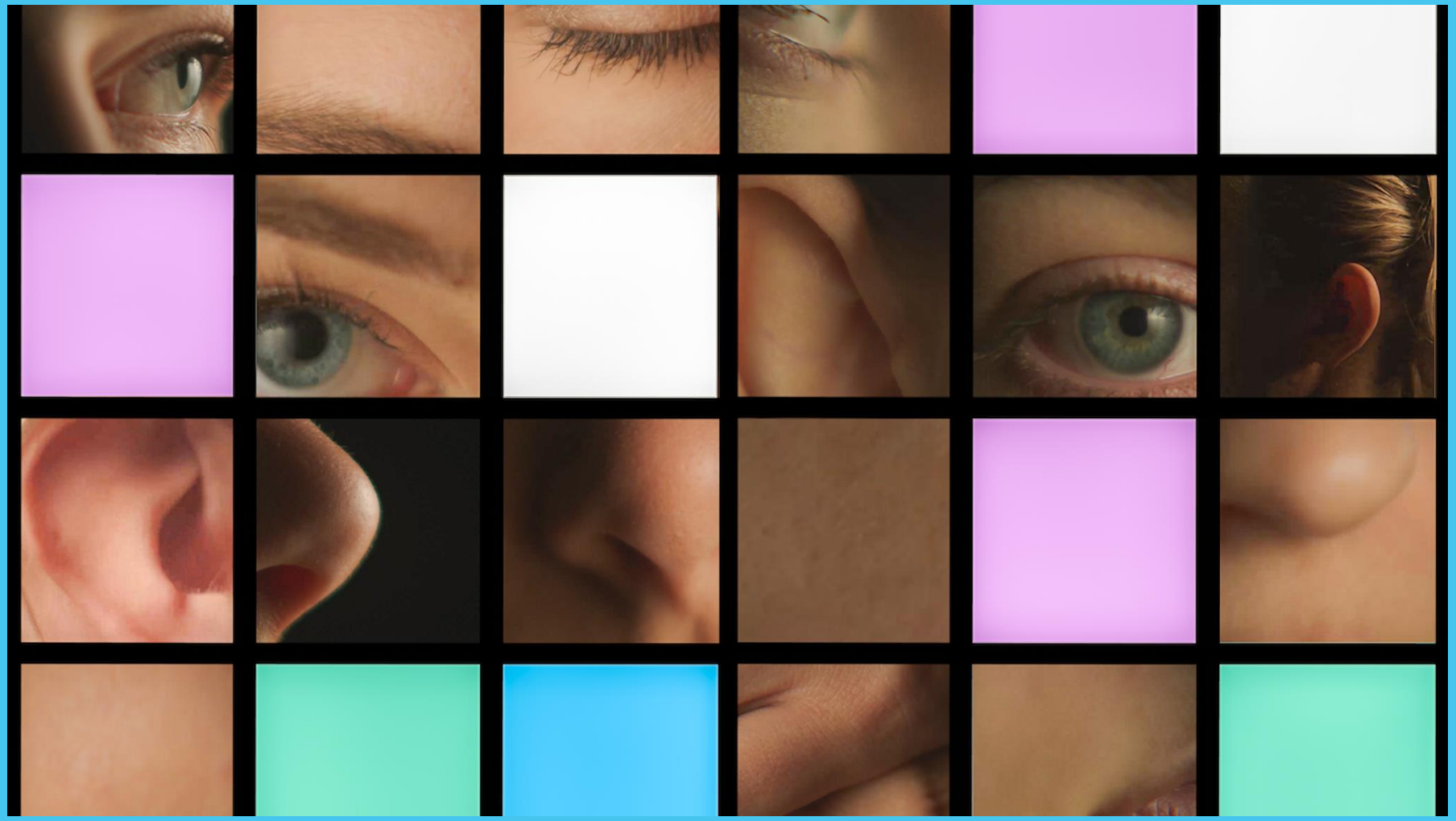
Imagine you're a journalist reporting on a protest and the police is using facial recognition software to record all protesters. You are added to a database of people considered a risk to public safety. Would you still report on the protest?

- What is the wider impact of your case?
- Do any of your NHRI's confront such cases (or similar cases)?
- And how can/should you deal with this case, if it is brought to your attention?

NHRIs‘ responses

- Monitoring/reporting on AI-based violations of freedom of opinion and expression and awareness-raising/human rights education
  - Case study: Error rates – false positives and false negatives; biased datasets
- Providing guidance and demanding a human rights-based approach to AI
  - Case study: Personalization and fragmentation
- Supporting human rights impact assessments
  - Case study: Discriminatory impact
  - Link to surveillance-based advertising and micro-targeting
- Ensuring policy coherence and human rights-compliant enforcement
- Multi-stakeholder engagement and partnerships
- Strategic litigation and access to justice

MADE TO MEASURE



Follow us & engage online!

www.osce.org/fom
www.osce.org/fom/ai-free-speech

Twitter @OSCE_RFoM

www.facebook.com/osce.rfom

www.youtube.com/osce.rfom

Telegram https://t.me/osce_rfom

Deniz Wagner:
deniz.wagner@osce.org
@wagner_deniz

Julia Haas:
julia.haas@osce.org
@TheJuliaHaas

Recommended Reading

- **SAIFE Policy Manual: https://www.osce.org/files/f/documents/8/f/510332_0.pdf**
- Additional reading
  - Artificial Intelligence and Disinformation as a Multilateral Policy Challenge: https://www.osce.org/files/f/documents/d/0/506702.pdf
  - Freedom of the Media and Artificial Intelligence: https://www.osce.org/files/f/documents/4/5/472488.pdf
  - SAIFE Strategy Session: https://www.osce.org/files/f/documents/9/f/456319_0.pdf
  - SAIFE Non-paper: https://www.osce.org/files/f/documents/b/a/447829.pdf