OSCE
The Representative on
Freedom of the Media

#SAIFE

Public Consultation on the Impact of AI on Free Speech

#SAIFE questionnaire on the impact of AI on Free Speech

1. Country of origin

2. Organization (civil society, tech industry, government, academia other – please specify)

3. Position in organization

4. The RFoM Strategy Paper to Put a Spotlight on Artificial Intelligence and Freedom of Expression (#SAIFE) outlines various challenges to free speech when AI is deployed. What do you consider to be the **biggest risk for freedom of expression** when it comes to the use of AI? Please specify and, if possible, provide examples on what you, in your field of work, expertise or region, would consider to be the most important issue(s).

5. How can the **understanding** of the implications of AI on free speech be increased among all stakeholders, including States, internet intermediaries and the general public? Please provide examples of good practices.

6. The #SAIFE Paper introduces preliminary recommendations for OSCE participating States and internet intermediaries. For which other stakeholders, if any, should recommendations be developed?

7. In the #SAIFE Paper, several **underlying issues** are identified that need to be taken into account when considering the impact of the use of AI on free speech. Internet intermediaries, especially social media platforms, act as gatekeepers by engaging in the selection of information that is published, in the ranking and editorial control over it, as well as in the removal of content. For these interventions, AI-powered tools are often deployed. The business model of most internet intermediaries, which is advertisement-based, builds on the collection and processing of massive amounts of data about their users, feeding into these AI-driven tools. Another underlying issue is that a few dominant internet intermediaries act as particularly powerful information gatekeepers in the online ecosystem. Are there additional underlying issues that need to be taken into account? Please explain.

8. The #SAIFE Paper addresses the role of "surveillance capitalism" **business models** in creating AI systems that can threaten freedom of expression and privacy. What alternative business models, besides behaviourally targeted advertising and the monetization of users' data, can support hosting of user-generated content on a massive scale? Which alternative business model could better protect diversity of voices online, and how could such a model be designed for different kinds of content-hosting and other online services? Please provide examples of good practices.

9. Do you consider the **dominance** of a few internet intermediaries to be a challenge for freedom of expression online? If so, which specific concerns do you see?

10. Is there a need to create a policy and normative environment that is conducive to a diverse, pluralistic information environment in the AI domain, or to ensure competition to prevent the concentration of AI expertise? Should the "network effect"[1] and limited interoperability of online services be addressed? If so, how?

11. Is there a need for **different approaches** or different free speech safeguards on AI depending on the specific internet intermediary, their size, capability, extent of risks of human rights impact, and services offered?

12. The #SAIFE Paper addresses how the surveillance of individuals' activities through AI technologies, by States (often relying on data collected through, and shared by, private companies) and by the private sector resulting from their business model, can seriously impede freedom of expression. What are the main risks stemming from the use of AI for **surveillance** techniques?

13. Can you provide examples of AI-powered surveillance technologies that are used in accordance with human rights principles of lawfulness, legitimacy, necessity and proportionality, and of available legal redress mechanisms for victims of surveillance-related abuses?

14. How can **users' agency** and choice regarding the application of AI processes be enhanced to ensure better free speech protection? What role does the principle of "privacy by design" play, or opt-in and opt-out provisions in respect of AI systems?

15. The #SAIFE Paper outlines how the use of AI for content moderation can lead to the **removal** of legitimate expression, or failure to remove content that could have a negative impact on those who access it ("false positives" and "false negatives"). Do safeguards need to be introduced in AI-powered tools to address this? If so, what kind/which ones?

16. The #SAIFE Paper outlines how the assessment of the (il)legality of content is a complex task, and depends on local context, local languages, and other societal, political, historical and cultural nuances. AI-driven decisions for content removal can fail to understand nuances underpinning the pieces of content, resulting in the filtering and taking down of legitimate content. Is there a need for a "**human in the loop**" in AI applications? If so, what level of human review or genuine human involvement should be ensured?

17. Is there a need for **different levels of human review** depending on the context in which AI is used (e.g., for the curation and prioritization of media content) compared to the use of AI to identify and flag potentially illegal content?

[1] The network effect is a phenomenon whereby increased numbers of people or participants improve the value of a good or service. A social media platform might therefore grow in popularity because it has achieved a critical mass of users and new users will be deterred from using another platform. For more information, see the #SAIFE Strategy Paper.

18. What measures should be taken (and by whom) to ensure that **societal inequalities** in the production of, and access to, information, which impact freedom of expression, are not reinforced in the development and deployment of AI technologies?

19. The #SAIFE Paper outlines how the use of AI can impede the free flow of information and democratic discourse. AI-powered tools are often used to categorize users to determine their particular political, commercial and other preferences in order to target them with specifically curtailed content. What measures could be initiated, by State and non-State actors, to promote the use of AI to **foster diversity** and to create an enabling environment for **media pluralism online**?

20. Should **human rights impact assessments** for AI-powered tools be mandatory, and if so, how, on which level (design, development, use of training datasets, deployment of AI), and according to which timeframes? Should specific AI applications require specific evaluation prior to their deployment? Who should conduct such assessments, and what should be the response to any foreseeable risks?

21. What measures, if any, need to be implemented to ensure **effective remedies** for AI-powered tools? How can it be ensured that those impacted by partially or fully automated decisions enjoy protection against erroneous or discriminatory outcomes? What internal complaint and redress mechanisms need to be installed for users in relation to AI?

22. The #SAIFE Paper calls for stronger transparency of AI-powered tools in content moderation and curation. What measures are needed to increase **transparency** while ensuring a strong data protection framework? Please provide examples of good practices.

23. What **minimum standards of transparency** should be introduced for the use of AI? What elements should these standards contain?

24. Should a **multi-tiered approach to transparency** be introduced? If so, what should be the main considerations?

25. How often should **transparency reports** be made publicly available? Should there be any other criteria for publication?

26. How can transparency norms and expectations be **harmonized** to ensure that disclosures are comparable, accurate, and useful to a broad range of stakeholders?

27. Are there any **free speech impediments to mandatory transparency regimes** that need to be addressed, and how could they be mitigated?

28. Are **different levels of transparency and accountability** required for the use of AI in different stages of content moderation and content curation (from uploads, to making certain content more visible for users, to the removal of content), in search results, or for tackling inauthentic behaviour?

29. Based on increased transparency, where and how do you see possibilities and benefits of **multi-stakeholder** contributions?

30. The #SAIFE Paper emphasizes the need for more **accountability** of AI-driven tools in content moderation and curation. What measures are necessary to ensure that AI systems employ a variety of controls, to verify that they work in accordance with their intentions, and to ensure that the operator can identify and rectify harmful outcomes or reproducing inequalities? What governance arrangements would lead to an effective system for supervising and ensuring that free speech is protected when AI is used?

31. What good governance and accountability processes can serve as a **model** for algorithmic and AI accountability, for all actors involved and at all stages in the process (design, use of training datasets, human rights impact assessments etc.)? Please provide examples of good practices.

32. What could and should be the role of **independent oversight** and of an **auditing** mechanism to ensure meaningful accountability of AI systems?

33. What role could and should **self-regulatory initiatives** play? How can it be ensured that discussions around "**ethical principles**" are based on, and compatible with, human rights?

34. The #SAIFE Paper outlines how the use of AI-powered tools can create a chilling effect for the media, and can lead to self-censorship, particularly of marginalized voices, and to altered behaviour in both online and offline spaces due to surveillance. How can AI-driven tools support the **protection of journalists**, and how can AI be beneficial for journalistic work and the media?

35. Do you have any other comments on the #SAIFE Paper and its preliminary recommendations, or would you like to raise **additionally important aspects**, areas of concern or sets of recommendations? Are there any other points you would like to raise?

36. Do you want to add any specific observations in the context of the **COVID-19 pandemic**, and the tendency, as observed in the #SAIFE Paper, to revert to technocratic solutions, including AI-powered tools, which may lack adequate societal debate or democratic scrutiny?

37. If you wish to provide a written statement, or send a position paper, article, report, or any other relevant information for the attention of the OSCE RFoM, please do so by sending an e-mail to AIFreeSpeech@osce.org.

38. Do you want to engage further in the #SAIFE discussions of the RFoM Office? If so, please indicate your name and contact details below or contact us via AIFreeSpeech@osce.org.