Joint Statement on Responsible Government Practices for AI Technologies

Freedom Online Coalition - September 2024

We, the member countries of the Freedom Online Coalition (FOC), recognize that safe, secure, and trustworthy artificial intelligence (AI) systems, when designed, developed, procured, deployed, used, and decommissioned responsibly, offer immense opportunities for governments to improve public service delivery, increase efficiency, and foster sustainable, inclusive development for all and advance the achievement of the 2030 Agenda for Sustainable Development.

At the same time, the design, development, procurement, deployment, use, and decommissioning of AI systems without adequate safeguards or in a manner inconsistent with international law pose risks that can undermine the protection of, promotion of, and ability to exercise and enjoy human rights and fundamental freedoms.

AI tools can create, contribute to, or exacerbate risks related to safety and privacy; the full exercise or enjoyment of human rights and fundamental freedoms for all; labor rights; democracy, good governance, and the rule of law; and equitable access to opportunities and critical resources and services. These risks can arise both as intended and unintended consequences of AI actors' actions and can include algorithmic bias and discrimination; the exposure of sensitive personal data; the creation and distribution of abusive and non-consensual deepfakes; the amplification of identity-based online harassment and abuse, including technology-facilitated gender-based violence; and the misuse of AI tools to facilitate repression or arbitrary or unlawful surveillance. The risk of harms facilitated or amplified by AI systems is especially pronounced for populations who experience systemic forms of violence, discrimination, and oppression related to gender and gender identity and expression, race, ethnicity, sexual orientation, disability status, or religion.

Risks and impacts on individuals can be particularly acute in public sector uses of AI—for example, when it is used to conduct surveillance, inform decisions in the judicial system, law enforcement, or government service delivery, or shape other governmental functions that can impact rights or safety. Biased AI tools deployed without appropriate safeguards, particularly by governments, can exacerbate existing inequalities and create new forms of marginalization and vulnerability.

Through the FOC and in our own national efforts, we recognize that policies, processes, and regulations regarding AI should protect and promote human rights. As articulated in the FOC 2020 Joint Statement on Artificial Intelligence and Human Rights, the FOC reaffirms that states must abide by their obligations under

international human rights law to ensure that human rights are fully respected and protected.

As also noted in the UN Guiding Principles on Business and Human Rights (UNGPs), "States must protect against human rights abuse within their territory and/or jurisdiction by third parties, including business enterprises." We recognize that government practices, including a government's procurement requirements, can also positively shape private sector practices to fulfill their responsibility to respect human rights in line with the UNGPs.

With the unprecedented exploration and adoption of AI tools, the time for governments to establish responsible and human rights-respecting AI practices and policies is now. As governments dedicated to upholding respect for universal human rights, democracy, and the rule of law, we are committed to integrating responsible AI practices into our policies and procedures for developing, deploying, procuring, and using AI that impact human rights or safety.

The measures below represent common practices and principles for which the implementation can vary across nations depending on legal frameworks and systems, with some frameworks providing even more robust safeguards such as the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Similar non-legally binding instruments that articulate the responsible use of surveillance tools and data include the work of the Freedom Online Coalition; the Organisation for Economic Co-operation and Development (OECD) AI Principles, Privacy Guidelines, and Declaration on Government Access to Personal Data Held by Private Sector Entities; the Global Privacy Assembly resolution on Government Access to Data, Privacy and the Rule of Law; the UNESCO Recommendation on the Ethics of Artificial Intelligence; relevant UN Human Rights Council and General Assembly resolutions including the resolution on Seizing the opportunities of safe, secure, and trustworthy artificial intelligence systems for sustainable development and the UN Guidance on Human Rights Due Diligence for Digital Technology Use. These policies and practices may complement efforts to address other impacts of AI, such as those on the environment. Some areas, such as government design, development, procurement, deployment, use, and decommissioning of AI for national security purposes are best addressed through separate processes, and the scope of this pledge is not intended to apply to those contexts.

We call on governments to:

1. **Respect International Obligations and Commitments:** Design, develop, procure, deploy, use, and decommission AI systems in a manner consistent with the Universal Declaration of Human Rights and international law, including the UN Charter and international human rights law.

2. **Assess Impacts of AI Systems in High-Risk Contexts[1] Prior to Deployment and Use:** Identify, assess, manage, and address potential risks to human rights, equity, fairness, and safety before deployment and use of an AI system on a case-by-case basis, and refrain from deploying AI systems where risks are incompatible with the protection of international human rights. The assessment may include:
    - Assessing the intended purpose and reasonably foreseeable uses of AI systems as well as their expected benefits to help ensure that AI is well-suited to accomplish the relevant task.
    - Assessing the potential risks of using AI in a given context, including by assessing the possible failure modes of the AI system and of the broader system, both in isolation and as a result of human users and other likely variables outside the scope of the system itself, documenting which stakeholders will be most impacted by the AI system, and enabling the meaningful participation of impacted stakeholders throughout the value chain of the AI system.
    - Evaluating the quality and representativeness of the data used in AI systems' design, development, training, testing, and operation and its fitness to the AI system's intended purpose, insofar as possible, including evaluation of:
        - The data collection, preparation, storage, and retention process, as well as the provenance of any data used to train, fine-tune, or operate the AI system;
        - The quality and representativeness of the data for its intended purpose;
        - How the data is relevant to the task being automated and may reasonably be expected to be useful for the AI system's development, testing, and operation; and
        - Whether the data contains sufficient breadth to address the range of real-world inputs the AI system might encounter and how data gaps, data inaccuracies, and other shortcomings can be addressed.

[1] For the purpose of this Joint Statement, High-Risk AI refers to AI systems that impact human rights and/or safety, which may be more likely in sectors such as healthcare, law enforcement and justice, and provision of public benefits.

- Testing the AI system for performance in realistic conditions or contexts to ensure the AI, as well as components that rely on it, will perform its intended purpose in real-world contexts, and considering leveraging pilots and limited releases with strong monitoring, evaluation, and safeguards in place to carry out the testing before wider releases.
- Testing or performing internal audits for accuracy and discriminatory bias, particularly pertaining to race, ethnicity, disability, gender, sexuality, and gender identity and expression.

3. **Conduct ongoing monitoring of AI systems in high-impact contexts throughout their use:** Identify, assess, and mitigate AI systems that may impact human rights, equity, fairness, or safety during use by conducting ongoing monitoring to identify, for example, degradation of AI systems' functionality and to detect changes in the AI system's impact on equity, fairness, human rights, and safety throughout the entire AI value chain, and ceasing use of AI systems as soon as is practicable where an AI systems' risks to human rights or safety exceed an acceptable level and where mitigation strategies do not sufficiently reduce risk. Incorporate feedback mechanisms, including from affected stakeholders and/or by participating in external audits or participating in third-party evaluations, to allow evidence-based discovery and reporting by end-users and third parties of technical vulnerabilities and misuses of the AI system, and take action to correct and address them.

4. **Ensure adequate human training and assessment:** Build capacity of relevant personnel to sufficiently understand the capabilities and limitations of AI systems, including their potential human rights impacts, through adequate training to enable appropriate degrees of assessment, and oversight of the AI system to allow those personnel to interpret and act on AI systems' output, to address human-machine teaming issues (such as automation bias), and to ensure the human-based components of the system effectively manage risks from the use of AI.

5. **Communicate and Respond to the Public:** Publicize available policies regarding how governments will protect human rights in the context of their AI activities. Establish processes for public disclosure of high-risk uses of AI systems and seek out and incorporate feedback from stakeholders or the public on uses of AI systems that impact equity, fairness, human rights, and safety, including by providing and maintaining options to opt out of AI-enabled decisions when appropriate.

6. **Provide Effective Access to Remedy:** Provide and facilitate access to timely human consideration and effective remedy, if appropriate, when individuals

have been negatively impacted throughout the value chain of AI systems, including by developing clear, detailed protocols for timely redress when an individuals' rights or safety has been violated that include guidelines on process, responsible entities, and timelines; notifying individuals when an AI system results in an adverse decision or action impacting their human rights or safety; and providing relevant information needed for individuals to seek remedy.

7. **Procure Safe, Secure, and Trustworthy AI:** Integrate the above practices into procurement processes as necessary preconditions, as appropriate, so outside vendors who design, develop, deploy, use, or decommission AI systems on our governments' behalf adhere to high standards for safety, security, and trustworthiness, such as human and labor rights standards, which may include by:
    - Obtaining adequate documentation to assess the AI system's capabilities, such as through the use of model, data, and system cards, as well as documentation of known limitations of the AI system and any guidelines on how the system is intended to be used;
    - Obtaining adequate information about the provenance of the data and other inputs to the AI system throughout its value chain used to train, fine-tune, or operate the AI system, in order to evaluate the quality of the data used in the system;
    - Assessing if there are adequate measures in place to prevent, address, and mitigate human and labor rights risks of the AI system throughout the vendor's value chain in line with the ILO Fundamental Principles and Rights at Work, the UN Guiding Principles on Business and Human Rights, the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct, and other international standards; including risks of infringement on human rights that can arise in connection with the purchase and integration of data or AI systems from third-party vendors, such as risks to privacy and gaps in transparency;
    - Regularly evaluating claims made by vendors concerning both the effectiveness of their AI offerings as well as the risk management measures put in place, including by testing the AI system in the particular environment where the government expects to deploy the capability;
    - Including contracting provisions that incentivize the continuous improvement of procured AI system; and
    - Requiring sufficient post-award monitoring of the AI system, where appropriate in the context of the product or service acquired.

AI can help address society's greatest challenges and further progress towards the 2030 Agenda and Sustainable Development Goals. Together, we pledge to develop and use AI responsibly, and call upon all governments to join us.