



CEB
Chief Executives Board
for Coordination

High-Level Committee on Programmes (HLCP)

Inter-Agency Working Group on Artificial Intelligence

Principles for the Ethical Use of Artificial Intelligence in the
United Nations System

Introduction

1. Artificial Intelligence has profound and dynamic positive and negative impacts on societies, the environment, ecosystems and human lives, including the human mind, in part because of the ways in which its use influences human thinking, interaction and decision-making. It can be a positive force for tackling the world's most urgent problems and be used to accelerate progress on achieving the Sustainable Development Goals and the Secretary General's Common Agenda. However, along with multiple advantages, these technologies also generate downside risks and challenges, derived from malicious use of technology or deepening inequalities and divides.
2. To ensure that the United Nations system uses AI for the greater interest of the people it serves, to benefit humanity and the planet, principles for the ethical use of Artificial Intelligence in the United Nations system have been developed to guide the design, development, deployment and use of AI. Based on UNESCO's Ethics of AI recommendation,¹ the principles contained in this document provide a basis for a framework for the ethical use of artificial intelligence (AI) by United Nations system organizations throughout all stages of an AI system lifecycle.² They must be read together with relevant policies, principles or guidance developed for United Nations system entities, notably: United Nations Human Rights Due Diligence Policy,³ the United Nations Personal Data Protection and Privacy Principles,⁴ and IASC Operational Guidance on Data Responsibility in Humanitarian Action.⁵
3. AI systems are understood as systems which have the capacity to automatically process data and information in a way that resembles intelligent human behaviour, and typically include aspects of reasoning, learning, perception, prediction, planning or control.⁶
4. An "ethical approach" to the use of artificial intelligence ("AI ethics") is defined as one that a) is consistent with the Charter of the United Nations, respecting all applicable rules of international human rights law, including the right to privacy, as well as with the principles identified below, and b) that entails the assessment of such consistency at all stages of the AI lifecycle.
5. The principles developed by the HLCP Inter-Agency Working Group on AI aim to provide a basis for United Nations system organizations to make decisions about how to develop, design, deploy and use AI systems, including multiple interacting AI systems, in a way that is trustworthy and puts human

¹ UNESCO's Ethics of AI Recommendation, available at: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>.

² The stages of the AI system lifecycle range from the research, design and development to procurement, deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end of use, disassembly and termination (see para. 2(b) of UNESCO's Ethics of AI Recommendation, available at <https://unesdoc.unesco.org/ark:/48223/pf0000380455>).

³ UN Secretary-General (UNSG), UN Human Rights Due Diligence Policy on Support for Non-United Nations Security Forces ("HRDDP"), 5 March 2013, A/67/775-S/2013/110, which applies to all United Nations entities providing support to non-United Nations security forces.

⁴ Available at: <https://unsceb.org/personal-data-protection-and-privacy-principles>.

⁵ Available at: <https://interagencystandingcommittee.org/operational-response/iasc-operational-guidance-data-responsibility-humanitarian-action>. Partners across the system will implement these guidelines in accordance with their respective mandates and the decisions of their governing bodies.

⁶ For more on the approach to AI, see para. 2 of UNESCO's Ethics of AI Recommendation, available at <https://unesdoc.unesco.org/ark:/48223/pf0000380455>. It is acknowledged, however, that the definition of AI may need to change over time as a result of technological developments.

dignity, equality of all human beings, preservation of the environment, biodiversity and ecosystems, respect for cultural diversity, and data responsibility at the centre.⁷

6. These principles, which are complementary and mutually reinforcing, are to be implemented by each United Nations system entity in alignment with the values, purposes and principles of the Charter of the United Nations and their respective individual mandates, governing instruments, rules, regulations and procedures, and without prejudice to the privileges and immunities of the relevant United Nations system organizations.
7. These principles require guidance to support their implementation by all United Nations system organizations. This includes, but is not limited to, developing an ethical assessment framework with due regard for human and child rights due diligence; reviewing and developing policies and guidance; ensuring appropriate governance systems, including risk management; ensuring internal procedures and policies are aligned with the United Nations Personal Data Protection and Privacy Principles and with the IASC Operational Guidance on Data Responsibility, or such principles as developed by individual organizations; promoting ethical principles of AI with United Nations system partners; putting in place learning and development opportunities for the United Nations system workforce; and developing pilot projects to assess the operationalization of the principles. To support the implementation of the principles by the United Nations system, HLCP⁸ will continue to develop practical guidance on how to embed these principles throughout the process, ensure awareness-raising of these ethical principles, and promote capacity development to support governance and ethical mechanisms on AI.

⁷ For more on UN system-wide strategy on AI capacity development, also see [UN System-wide Strategic approach and Road map for supporting capacity development on AI](https://unsceb.org/united-nations-system-wide-strategic-approach-and-road-map-supporting-capacity-development) (endorsed by the CEB in May 2019). Available at <https://unsceb.org/united-nations-system-wide-strategic-approach-and-road-map-supporting-capacity-development>

⁸ HLCP, the High-level Committee on Programmes, <https://unsceb.org/high-level-committee-programmes-hlcp>

Principles

<p>Do no harm</p>	<p>AI systems should not be used in ways that cause or exacerbate harm, whether individual or collective, and including harm to social, cultural, economic, natural, and political environments. All stages of an AI system lifecycle should operate in accordance with the purposes, principles and commitments of the Charter of the United Nations. All stages of an AI system lifecycle should be designed, developed, deployed and operated in ways that respect, protect and promote human rights and fundamental freedoms. The intended and unintended impact of AI systems, at any stage in their lifecycle, should be monitored in order to avoid causing or contributing to harm, including violations of human rights and fundamental freedoms.</p>
<p>Defined purpose, necessity and proportionality</p>	<p>The use of AI systems, including the specific AI method(s) employed, should be justified, appropriate in the context and not exceed what is necessary and proportionate to achieve legitimate aims that are in accordance with each United Nations system organization’s mandates and their respective governing instruments, rules, regulations and procedures.</p>
<p>Safety and security</p>	<p>Safety and security risks should be identified, addressed and mitigated throughout the AI system lifecycle to prevent where possible, and/or limit, any potential or actual harm to humans, the environment and ecosystems. Safe and secure AI systems should be enabled through robust frameworks.⁹</p>
<p>Fairness and non-discrimination</p>	<p>United Nations system organizations should aim to promote fairness to ensure the equal and just distribution of the benefits, risks and costs, and to prevent bias, discrimination and stigmatization of any kind, in compliance with international law. AI systems should not lead to individuals being deceived or unjustifiably impaired in their human rights and fundamental freedoms.</p>
<p>Sustainability</p>	<p>Any use of AI should aim to promote environmental, economic and social sustainability. To this end, the human, social, cultural, political, economic and environmental</p>

⁹ These may include, for example, the development or enhancement of a) sustainable, privacy-protected data access frameworks, b) appropriate safeguards against function creep, and c) fair and inclusive training, validation, and maintenance of AI models utilizing quality data.

	<p>impacts of AI technologies should continuously be assessed and appropriate mitigation and/or prevention measures should be taken to address adverse impacts, including on future generations.</p>
<p>Right to privacy, data protection and data governance</p>	<p>Privacy of individuals and their rights as data subjects must be respected, protected and promoted throughout the lifecycle of AI systems. When considering the use of AI systems, adequate data protection frameworks and data governance mechanisms should be established or enhanced in line with the United Nations Personal Data Protection and Privacy Principles¹⁰ also to ensure the integrity of the data used.</p>
<p>Human autonomy and oversight</p>	<p>United Nations system organizations should ensure that AI systems do not overrule freedom and autonomy of human beings and should guarantee human oversight. All stages of the AI system lifecycle should follow and incorporate human-centric design practices and leave meaningful opportunity for human decision-making. Human oversight must ensure human capability to oversee the overall activity of the AI system and the ability to decide when and how to use the system in any particular situation, including whether to use an AI system and the ability to override a decision made by a system. As a rule, life and death decisions or other decisions affecting fundamental human rights of individuals must not be ceded to AI systems, as these decisions require human intervention.</p>
<p>Transparency and explainability</p>	<p>United Nations system organizations should ensure transparency and explainability of AI systems that they use at all stages of their lifecycle and of decision-making processes involving AI systems. Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Individuals should be meaningfully informed when a decision which may or will impact their rights, fundamental freedoms, entitlements, services or benefits, is informed by or made based on AI algorithms and have access to the reasons for a decision and the logic involved. The information and reasons for a decision should be presented in a manner that is understandable to them.</p>

¹⁰ <https://unscebor.org/principles-personal-data-protection-and-privacy-listing>

<p>Responsibility and accountability</p>	<p>United Nations system organizations should have in place appropriate oversight, impact assessment, audit and due diligence mechanisms, including whistle-blowers' protection, to ensure accountability for the impacts of the use of AI systems throughout their lifecycle. Appropriate governance structures should be established or enhanced which attribute the ethical and legal responsibility and accountability for AI-based decisions to humans or legal entities, at any stage of the AI system's lifecycle. Harms caused by and/or through AI systems should be investigated and appropriate action taken in response. Accountability mechanisms should be communicated broadly throughout the United Nations system in order to build shared knowledge resources and capacities.</p>
<p>Inclusion and participation</p>	<p>When designing, deploying and using AI systems, United Nations system organizations should take an inclusive, interdisciplinary and participatory approach, which promotes gender equality. They should conduct meaningful consultations with all relevant stakeholders and affected communities, in the process of defining the purpose of the AI system, identifying the underlying assumptions, determining the benefits, risks, harms and adverse impacts, and adopting prevention and mitigation measures.</p>