United Nations System CEB/2022/2/Add.1

Chief Executives Board for Coordination

Distr.: General
27 October 2022

Original: English

Second regular session of 2022
Manhasset, New York, 27 and 28 October 2022

Summary of deliberations

Addendum

Principles for the ethical use of artificial intelligence in the United Nations system

I. 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 addressing the world's most urgent problems and can be used to accelerate progress in achieving the Sustainable Development Goals and the goals of Our Common Agenda. However, along with multiple advantages, artificial intelligence technologies have a downside; they entail risks and challenges, such as those resulting from the malicious use of technology or from deepening inequalities and divides.

2. Principles for the ethical use of artificial intelligence in the United Nations system have been developed to ensure that the United Nations uses artificial intelligence for the best interest of the people whom it serves, in order to benefit humanity and the planet. The principles, which are contained in the present report, guide the design, development, deployment and use of artificial intelligence. They are based on the Recommendation on the ethics of artificial intelligence of the United Nations Educational, Scientific and Cultural Organization[1] and provide a basis for a framework for the ethical use of artificial intelligence by United Nations system organizations throughout all stages of an artificial intelligence system's life cycle.[2] They must be read together with relevant policies, principles and guidance developed for United Nations system entities, including the Human rights due diligence policy

[1] United Nations Educational, Scientific and Cultural Organization (UNESCO), Recommendation on the ethics of artificial intelligence, available at: https://unesdoc.unesco.org/ark:/48223/pf0000380455.

[2] The stages of the artificial intelligence system life cycle range from research, design and development to procurement, deployment and use, and include maintenance, operation, trade, financing, monitoring and evaluation, validation, end of use, disassembly and termination (see UNESCO, Recommendation on the ethics of artificial intelligence, sect. I, para. 2 (b)).

on United Nations support to non-United Nations security forces (A/67/775-S/2013/110, annex), the United Nations Personal data protection and privacy principles,[3] and the Operational guidance on data responsibility in humanitarian action of the Inter-Agency Standing Committee.[4]

3. Artificial intelligence systems are understood to be systems that have the capacity to automatically process data and information in a way that resembles intelligent human behaviour, and they typically include aspects of reasoning, learning, perception, prediction, planning or control.[5]

4. An "ethical approach" to the use of artificial intelligence is defined as one that 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 that entails the assessment of such consistency at all stages of the artificial intelligence life cycle.

5. The principles developed by the Inter-Agency Working Group on Artificial Intelligence of the High-level Committee on Programmes are intended to provide a basis for United Nations system organizations to make decisions on how to develop, design, deploy and use artificial intelligence systems, including multiple interacting systems, in a way that is trustworthy and centred on human dignity, the equality of all human beings, preservation of the environment, biodiversity and ecosystems, respect for cultural diversity, and data responsibility.[6]

6. The principles, which are complementary and mutually reinforcing, are to be implemented by all the United Nations system entities in alignment with the values, purposes and principles of the Charter of the United Nations and their relevant mandates, governing instruments, rules, regulations and procedures, and without prejudice to the privileges and immunities of the relevant United Nations system organizations.

7. Guidance is needed to support adherence to the principles by all United Nations system organizations. This includes, but is not limited to, the development of an ethical assessment framework, with due regard for human and child rights due diligence; the revision and development of policies and guidance; the establishment of appropriate governance systems, including risk management systems; the alignment of internal procedures and policies with the Personal data protection and privacy principles, the Operational guidance on data responsibility in humanitarian action of the Inter-Agency Standing Committee or such other principles as are developed by individual organizations; the promotion of ethical principles of artificial intelligence use with United Nations system partners; the creation of learning and development opportunities for the United Nations system workforce; and the development of pilot projects to assess the operationalization of the principles. To support adherence to the principles by the United Nations system, the High-level Committee on Programmes will continue to develop practical guidance on how to embed the principles throughout the process of implementation, raise awareness of these ethical principles and promote capacity development to support artificial intelligence governance and related ethical mechanisms.

[3] Available at: https://unsceb.org/personal-data-protection-and-privacy-principles.

[4] Available at: https://interagencystandingcommittee.org/operational-response/iasc-operational-guidance-dataresponsibility-humanitarian-action. Partners across the system will implement these guidelines in accordance with their mandates and the decisions of their governing bodies.

[5] For more on the approach to artificial intelligence, see UNESCO, Recommendation on the ethics of artificial intelligence, para. 2. It is acknowledged, however, that the definition of artificial intelligence may need to change over time as a result of technological developments.

[6] For more on the United Nations system-wide strategy on artificial intelligence capacity development, see also CEB/2019/1/Add.3, A United Nations system-wide strategic approach and road map for supporting capacity development on artificial intelligence (endorsed by CEB in May 2019).

II. Principles

Do no harm	Artificial intelligence systems should not be used in ways that cause or exacerbate harm, whether individual or collective, including harm to social, cultural, economic, natural or political environments. All stages of an artificial intelligence system's life cycle should operate in accordance with the purposes, principles and commitments of the Charter of the United Nations. All stages of an artificial intelligence system's life cycle should be designed, developed, deployed and operated in ways that respect, protect and promote human rights and fundamental freedoms. The intended and unintended impacts of artificial intelligence systems, at any stage in their life cycle, should be monitored in order to avoid causing or contributing to harm, including violations of human rights and fundamental freedoms.
Defined purpose, necessity and proportionality	The use of artificial intelligence systems, including the specific artificial intelligence 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 mandate and governing instruments, rules, regulations and procedures.
Safety and security	Safety and security risks should be identified, addressed and mitigated throughout the artificial intelligence system's life cycle to prevent or, at least, limit any potential or actual harm to humans, the environment or ecosystems. Safe and secure artificial intelligence systems should be enabled through robust frameworks.[7]
Fairness and non-discrimination	United Nations system organizations should aim to ensure the equal and just distribution of the benefits, risks and costs associated with artificial intelligence systems and to prevent bias, discrimination and stigmatization of any kind, in compliance with international law. The use of artificial intelligence systems should not lead to individuals being deceived or to unjustifiable restrictions on their human rights and fundamental freedoms.
Sustainability	Artificial intelligence should be aimed at promoting environmental, economic and social sustainability. To this end, the human, social, cultural, political, economic and environmental impacts of such technologies should be continuously assessed and appropriate mitigation and prevention measures should be taken to address adverse impacts, including on future generations.
Right to privacy, data protection and data governance	Individuals' privacy and rights as data subjects must be respected, protected and promoted throughout the life cycle of artificial intelligence systems. When the use of artificial intelligence systems is considered, adequate data protection frameworks and data governance mechanisms should be established or enhanced, in line with the Personal data protection and privacy principles, also to ensure the integrity of the data used.

[7] 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 artificial intelligence models utilizing quality data.

Human autonomy and oversight	United Nations system organizations should ensure that artificial intelligence systems do not impinge on human beings' freedom and autonomy and should guarantee human oversight. All stages of an artificial intelligence system's life cycle should follow and incorporate human-centric design practices and leave meaningful opportunity for human decision-making. Human oversight includes ensuring that humans have the capability to manage the overall activity of an artificial intelligence system and the ability to decide when and how to use it in specific situations, including whether to use such a system, and the ability to override a decision made by such a system. As a rule, life or death decisions or other decisions affecting fundamental human rights require human intervention and must not be ceded to artificial intelligence systems.
Transparency and explainability	United Nations system organizations should ensure the transparency and explainability of artificial intelligence systems that they use, at all stages of their life cycles, and of decision-making processes involving such systems. Technical explainability requires that the decisions made by an artificial intelligence system can be understood and traced by human beings. Individuals should be fully informed when a decision that may or will affect their rights, fundamental freedoms, entitlements, services or benefits is informed by or made based on artificial intelligence algorithms, and should have access to the reasons and logic behind such decisions. The information and reasons for a decision should be presented in a manner that they can understand.
Responsibility and accountability	United Nations system organizations should have appropriate oversight, impact assessment, audit and due diligence mechanisms, including protection for whistle-blowers, to ensure accountability for the impacts of the use of artificial intelligence systems throughout their life cycles. Appropriate governance structures should be established or enhanced to ensure that humans or legal entities are made ethically and legally responsible and accountable for artificial intelligence-based decisions made at any stage of an artificial intelligence system's life cycle. Harm caused by or as a result of the use of artificial intelligence systems should be investigated, and appropriate action taken in response. Information on accountability mechanisms should be communicated widely throughout the United Nations system in order to build shared knowledge, resources and capacities.
Inclusion and participation	When designing, deploying and using artificial intelligence systems, United Nations system organizations should take an inclusive, interdisciplinary and participatory approach, and promote gender equality. They should conduct meaningful consultations with all relevant stakeholders and affected communities as part of the processes of defining the purpose of an artificial intelligence system, identifying the assumptions underpinning its use, identifying the associated benefits, risks, harm and adverse impacts, and adopting prevention and mitigation measures.