

*Note to facilitators: This document is a summary of the original text of the UNESCO draft recommendation on AI ethics, for which all this deliberative work is designed. It comes with an index allowing to find each of the points below with the ethical issues identified*

---

The UNESCO is involved in the matter of AI Ethics, as by its Constitution, it “seeks to construct the defences of peace in the minds of human beings and aims to promote cooperation among the nations”

This Recommendation **(2)** pays specific attention to the central domains of UNESCO: **(2a)** education, **(2b)** science, **(2c)** culture, **(2d)** communication and information. It is **(3)** addressed to States, public and private organizations and individuals that **Research, Design, Develop, Deploy, or Use (RDDDU)** AI.

The recommendation aims for the **(4)** formulation of ethical values, principles, policy recommendations for the RDDDU of AI, for the good of humanity. This document **(6)** is addressed primarily to policy-makers but also to serve as a framework stimulating a multi-stakeholder approach, for non-state actors.

## VALUES

**-Human dignity:** The RDDDU of AI should **(9)** respect and preserve human dignity and this value **(10)** should be respected by all actors involved

**-Human rights and fundamental freedoms:** RDDDU of AI should be **(11)** consistent and compliant with international human rights law, principles and standards

**-Leaving no one behind:** RDDDU of AI should be **(12)** in a way that respects all groupings of humanity and fosters creativity **(13)**. It must be compatible with empowering all humans

**-Living in harmony:** The RDDDU of AI should **(14)** recognise the interconnectedness of all humans (belonging to a whole), **(15)** should avoid conflict, violence, should not segregate, objectify or undermine the safety of humans, or divide and turn groups against each other

**-Trustworthiness:** AI should be **(16)** trustworthy and the RDDDU should inspire, instead of infringing on, trust among people and in AI. Trust has **(17)** to be earned in each use context

**-Protection of the Environment:** RDDDU of AI should **(18)** recognise the promotion of environmental well-being, follow environmental laws to minimise climate change risk factors, prevent exploitation and depletion of natural resources, and be used **(19)** to provide solutions to protect the environment.

## PRINCIPLES

**-For human and flourishing:** RDDDU should **(23)** let humans and the environment flourish, enhance their quality of life. RDDDU of AI may be used **(24)** in interactions with vulnerable people and should never objectify or undermine dignity, or violate or abuse human rights

---

<sup>1</sup> abbreviated as such in this document

- Proportionality:** RDDDU of AI **(25)** may not exceed what is necessary to achieve legitimate aims, **(26)** the method chosen should be **(26a)** desirable, proportional to the aim **(26b)** not have an excessive infringement on the values of this document **(26c)** be appropriate to context
- Human oversight and determination:** **(27)** Ethical and legal responsibility for the RDDDU of AI always attributable to a physical person or to a legal entity. **(28)** Shared control with AI is for reasons of efficacy; this decision in limited contexts is that of humans. RDDDU of AI is to assist humans in decision-making and acting, never to replace human responsibility
- Sustainability:** RDDDU of AI should **(29)** promote the achievement of sustainability related to globally accepted frameworks such as the sustainable development goals.
- Diversity and inclusiveness:** RDDDU of AI should **(30)** respect and foster diversity and inclusiveness at a minimum consistent with international human rights law
- Privacy:** AI should **(31)** respect, protect and promote privacy. Adequate data governance mechanisms should be ensured throughout the lifecycle of AI
- Awareness and literacy:** **(32)** Public awareness and understanding of AI and the value of data should be promoted so that citizens can make informed decisions about their use of AI
- Multi-stakeholder and adaptive governance:** Governance of AI should be **(33)** responsive to shifts in technology and associated business models **(34)**. It should consider a range of responses from soft governance through self-regulation and certification processes to hard governance with national laws and, where possible and necessary, international instruments.
- Fairness:** AI actors should **(35)** respect fairness, equity to minimise and avoid reinforcing or perpetuating socio-technical biases including racial, ethnic, gender, age, and cultural biases
- Transparency and explainability:** **(36)** Level of transparency and explainability should always be appropriate to the use context **(37)**. Transparency = allowing people to understand the RDDDU of AI appropriately to the use context and sensitivity of the AI system. **(38)** Explainability = making intelligible and providing insight into the outcome of AI
- Safety and security:** RDDDU of AI should **(39)** avoid unintended harms (safety risks) and vulnerabilities to attacks (security tasks). Governments should **(40)** play a leading role in ensuring safety and security. Research on safety and security risks, should be continual
- Responsibility and accountability:** AI actors should **(41)** assume moral, legal responsibility and **(42)** mechanisms should be developed to ensure accountability. Technical and institutional designs should be considered to ensure auditability and traceability of AI

## AREAS OF POLICY ACTION FOR MEMBER STATES

### *ACTION GOAL I: ETHICAL STEWARDSHIP*

- Policy Action 1: Promoting Diversity & Inclusiveness:** **(44)** Member states should actively participate in discussions on AI, **(45)** require AI actors to combat stereotyping in AI, whether by design or by negligence , **(46)** ensure that AI actors demonstrate awareness and respect for cultural and social diversity, **(47)** address diversity gaps in AI, and **(48)** mainstream AI ethics discussions

### *ACTION GOAL II: IMPACT ASSESSMENT*

- Policy Action 2: Addressing Labour Market Changes:** Member States should **(50)** assess and address AI impact on labour markets and for education requirements; 'learning how to learn' should be taught alongside technical skills. They should **(51)** ensure a fair transition for at-risk employees, **(52)** analyze AI impact on the local labour market, **(53)** develop labour force policies to support women and underrepresented populations
- Policy Action 3: Addressing social and economic impact of AI:** **(54)** Preventing the monopolisation of AI and resulting inequalities whether these are data, research, technology, market or others, **(55)** providing adequate AI literacy to the public, **(56)** establishing monitoring and evaluation mechanisms for initiatives and policies related to AI ethics, **(57)** considering certification mechanisms for AI similar to the ones used for medical devices, **(58)** encouraging private sector to involve different stakeholders in AI governance and

consider adding AI Ethics Officer to oversee impact assessment, (59) developing data governance strategies balancing metadata and users' privacy

**-Policy Action 4: Impact on Culture and on the Environment: (60)** Incorporating AI in cultural heritage where appropriate, (61) examining AI impact in human language, (62) researching the effects of interaction of people with AI, (63) promoting AI education for creatives, for (64) cultural industries and startups, (65) assessing and reducing the environmental impact of AI.

#### *ACTION GOAL III: CAPACITY BUILDING FOR AI ETHICS*

**-Policy Action 5: Promoting AI Ethics Education & Awareness:** Encouraging (67) the embedding of AI ethics and the collaboration between technical skills and humanities, (68) promoting awareness of AI its opportunities and challenges, (69) introducing flexibility into higher education given the innovations in AI, (70) promoting awareness programs of AI (71) encouraging research initiatives on use of AI in teaching, (72) supporting academia-industry collaboration, (73) promoting the participation of women and all sorts of minorities.

**-Policy Action 6: Promoting AI Ethics Research: (74)** Promoting AI research, ensuring that (75) AI researchers are trained in ethics, (76) facilitating access to data for research for scientific community, not at the expense of citizens' privacy, (77) promoting gender diversity in AI research, academia and industry, fighting gender stereotyping and harassment, (78) promoting interdisciplinary and critical AI research including disciplines other than STEMs.

#### *ACTION GOAL IV: DEVELOPMENT AND INTERNATIONAL COOPERATION*

**-Policy Action 7: Promoting Ethical Use of AI in Development: (80)** encouraging the ethical use of AI in areas of development, (81) striving to provide platforms for international cooperation on AI for development, (82) working to promote international collaborations on AI research

**-Policy Action 8: Promoting International Cooperation on AI Ethics: (83)** Working through international organizations and research institutions to conduct AI ethics research, ensuring that algorithms and data used in AI areas are applied equally and fairly, and (84) encouraging international cooperation in AI development to bridge geo-technological lines

#### *ACTION GOAL V: GOVERNANCE FOR AI ETHICS*

**-Policy Action 9: Establishing Governance Mechanisms for AI Ethics: (86)** Ensuring AI governance is (86a) Inclusive, (86b) Transparent, (86c) Multidisciplinary, (86d) Multilateral.(87) Fostering the development of, and access to a digital ecosystem for ethical AI, (88) encouraging the development and use of AI guidelines, (89) considering the development of international legal framework for international cooperation

**-Policy Action 10: Ensuring Trustworthiness of AI: (90)** Implementing proper measures to monitor all phases of an AI lifecycle, (91) setting requirements for AI transparency and explainability based on (91a) application domain, (91b) target audience and (91c) feasibility, (92) encouraging research into transparency and explainability, (93) developing international standards to measure levels of transparency for objective assessment of compliance levels

**-Policy Action 11: Ensuring Responsibility, Accountability and Privacy: (94)** adapting, legal frameworks to achieve accountability and responsibility for AI content and outcome to a natural or legal person; no responsibility or legal personality should be granted to AI, (95) identify and assess benefits and risks of AI, and oversight mechanisms, including auditability, traceability and explainability, and include external, multidisciplinary review of AI, (96) involving all actors of the AI ecosystem. (97) Insuring that harms caused to users through AI can be investigated, punished, and redressed, (98) applying appropriate safeguards of individuals' fundamental right to privacy, (99) ensuring that individuals can oversee the use of their private information/data, (100) ensuring increased security for personally identifiable data, (101) adopting a Commons approach to data to promote interoperability of datasets and exercising vigilance in overseeing their collection and use

*Credits: Designed by the research team of Algora Lab and of Mila Quebec Artificial Intelligence Institute (July 2020)*