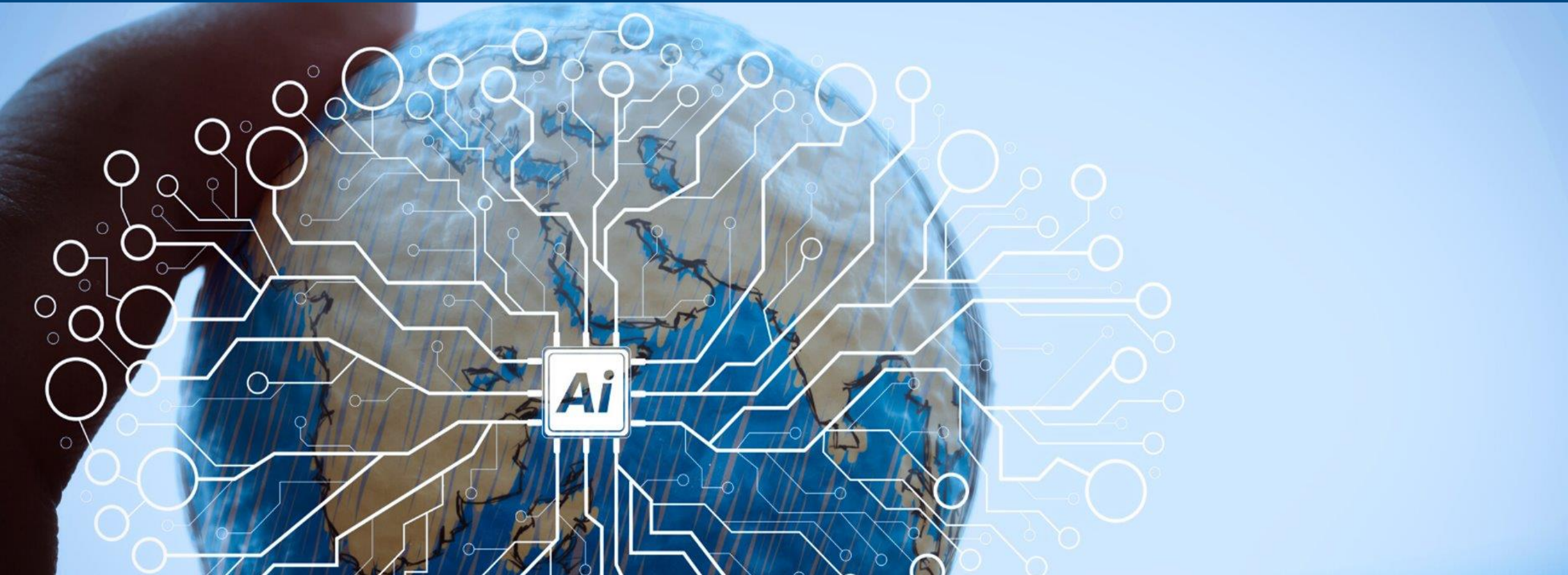




unesco

Recommendation on the Ethics of AI

4th Session of the WIPO Conversation



unesco

Dafna Feinholz / Chief of Section, Bioethics and Ethics of Science and Technology

22 September 2021

Towards a Global Ethical Framework of AI: the Recommendation

UNESCO is elaborating the first global standard-setting instrument on the ethics of artificial intelligence (adoption by the General Conference is anticipated in November 2021).

The Recommendation pays specific attention to the central domains of UNESCO: **Education, Sciences, Culture, and Communication and Information.**

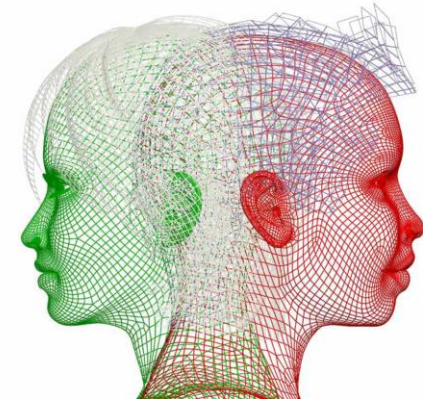
What makes the draft Recommendation unique?

- ✓ Global scope
- ✓ Strong language
- ✓ Emphasis on *how* (policy areas)
- ✓ Focus on the entire life cycle of AI systems
- ✓ Addressing issues of Data, Gender and Environment



Importance of Ethical Approach

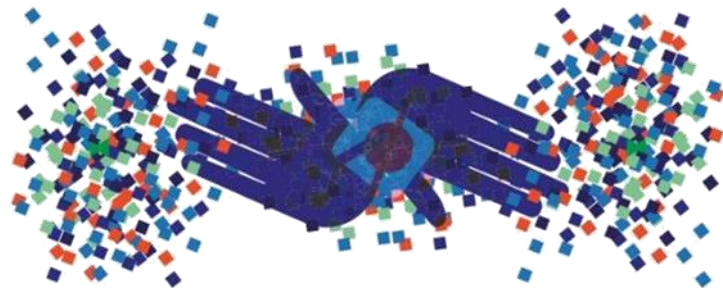
- Ethics is about making explicit the values that lie behind our policies, choices and actions, and applying these to human decision making. Human rights are the basic common ground.
- AI technologies are not value neutral, they should be firmly grounded in certain common foundational human values, in particular the international human rights framework.
- AI technologies are always evolving, and needs to be constantly accompanied by ethics as a dynamic process to ensure that human rights and fundamental freedoms are always protected.
- Ethical frameworks provide a response in cases where regulation does not or cannot exist.



The Recommendation on the Ethics of AI: Values and Principles

VALUES

- Respect, protection and promotion of human rights and fundamental freedoms and human dignity
- Environment and ecosystem flourishing
- Ensuring diversity and inclusiveness
- Living in peaceful, just and interconnected societies



PRINCIPLES

- Proportionality and Do No Harm
- Safety and security
- Fairness and non-discrimination
- Sustainability
- **Right to Privacy, and Data Protection**
- Human oversight and determination
- Transparency and explainability
- Responsibility and accountability
- Awareness and literacy
- Multi-stakeholder and adaptive governance and collaboration

The Recommendation on the Ethics of AI: Privacy & Data Protection

- **Privacy**, a right essential to the protection of human dignity, human autonomy and human agency, must be respected, protected and promoted throughout the life cycle of AI systems.
- Adequate **data protection frameworks and governance mechanisms** should be established in a multi-stakeholder approach, protected by judicial systems, and ensured throughout the life cycle of AI systems.
- Algorithmic systems require adequate **privacy impact assessments**, which also include societal and ethical considerations of their use and an innovative use of the **privacy by design** approach.
- AI actors need to ensure that they are **accountable** for the design and implementation of AI systems in such a way as to ensure that personal information is protected throughout the life cycle of the AI system.



The Recommendation on the Ethics of AI: Policy Action Areas

Concrete, impact-oriented, and innovative policy actions are proposed in the following areas:

1. Ethical Impact Assessment
2. Ethical governance and stewardship
3. **Data policy**
4. Development and international cooperation
5. Environment and ecosystems
6. Gender
7. Culture
8. Education and research
9. Communication and information
10. Economy and labour
11. Health and social well-being



The Recommendation on the Ethics of AI: Data Policy Area

Data is the lifeblood of AI systems. The Recommendation provides evidence-based guidance on:

- Mechanisms to require AI actors to disclose and combat stereotyping in the outcomes of AI systems and data
- Strategies that strike a balance between the collection of metadata and users' privacy
- Safeguards to respect, protect and promote privacy throughout the lifecycle of AI systems
- Increased security for personal data and sensitive data



The Recommendation on the Ethics of AI: Data Policy Area

- Develop data governance strategies that ensure:
 - continual evaluation of the quality of training data for AI systems
 - proper data security and protection measures
 - feedback mechanisms
- Encourage all AI actors, including business enterprises, to carry out privacy impact assessments and to apply privacy by design in their systems
- Devise effective and meaningful accountability schemes and mechanisms
- Promote open data
- Promote public and private efforts to create collaborative platforms to share quality data in trusted and secured data spaces



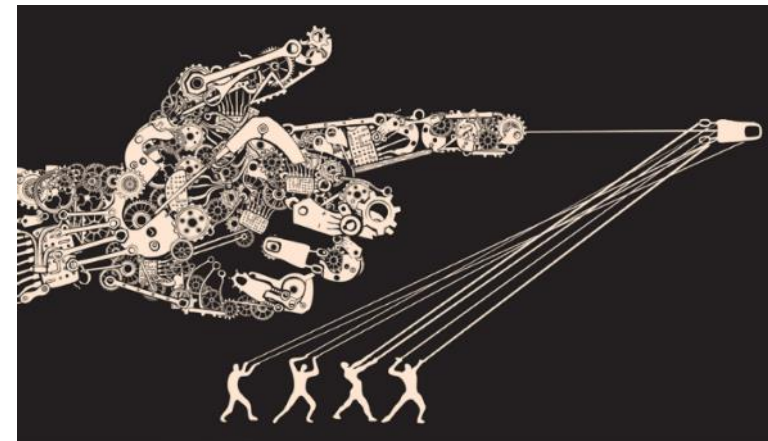
The Recommendation on the Ethics of AI: Innovative tools

Readiness benchmarking methodology

- assist Member States in identifying their readiness status related to the Recommendation

Ethical impact assessment

- help Member States and other stakeholders identify and assess benefits, concerns and risks of AI systems, as well as risk prevention, mitigation and monitoring measures



Thank you!

More on the UNESCO Recommendation on the Ethics of AI:

<https://en.unesco.org/artificial-intelligence/ethics>