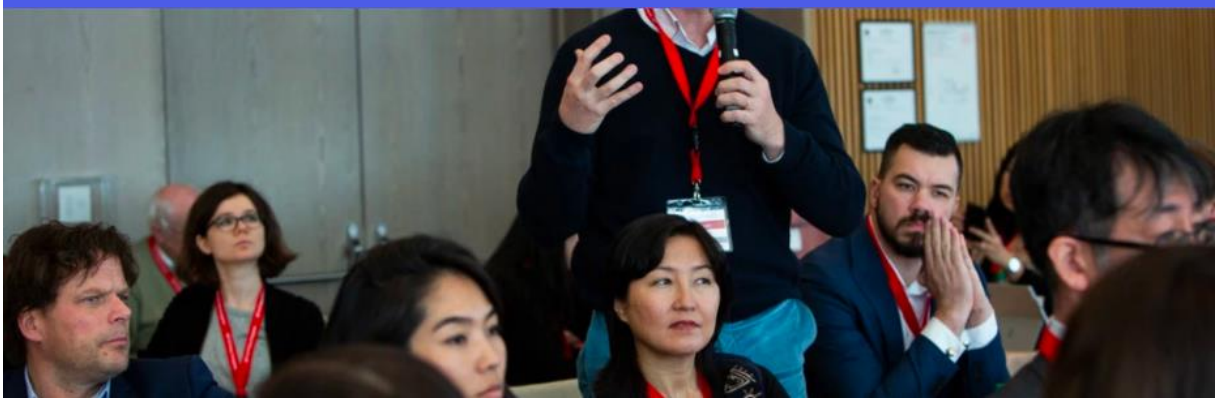
The Future Society, Law & Society Initiative, Principles for the Governance of AI

MAIN INSIGHT

Central Value: “Human beings are equal in rights, dignity, and the freedom to flourish”.



Policy Research The Law & Society Initiative
July 15, 2017
3 min read

Principles for the Governance of AI

Principle 1: AI shall not impair, and, where possible, shall advance the equality in rights, dignity, and freedom to flourish of all humans. Accordingly, the purpose of governing artificial intelligence is to develop policy frameworks, voluntary codes or practice, practical guidelines, national and international regulations, and ethical norms that safeguard and promote the equality in rights, dignity, and freedom to flourish of all humans.

Principle 2: AI shall be transparent. Transparency is the ability to trace cause and effect in the decision-making pathways of algorithms and, in hybrid intelligence systems, of their operators.

Principle 3: Manufacturers and operators of AI shall be accountable. Accountability means the ability to assign responsibility for the effects caused by AI or its operators.

Principle 4: AI’s effectiveness shall be measurable in the real-world applications for which it is intended. Measurability means the ability for both expert users *and the ordinary citizen* to gauge concretely whether AI or hybrid intelligence systems are meeting their objectives.

Principle 5: Operators of AI systems shall have appropriate competencies. When our health, our rights, our lives or our liberty depend on hybrid intelligence, such systems

should be designed, executed and measured by professionals with the requisite expertise.

Principle 6: The norms of the delegation of decisions to AI systems shall be codified through thoughtful, inclusive dialogue with civil society. In most instances, the codification of the acceptable uses of AI remains the domain of the technical elite with legislators, courts and governments struggling to catch up to realities on the ground, while ordinary citizens remain mostly excluded. Principle 6 is intended to ensure that standards and codes of practice result from more inclusive dialogue and are grounded in truly broad consensus.

Why a Framework, why now?

"To what extent should societies delegate to machines decisions that affect people?" Humanity's answer to this question will have profound consequences on how we experience every aspect of life, including our very conception of what it means to be human. The stakes are particularly meaningful in the extended legal domain, on which our safety, our rights and our duties, our dignity, and the prosperity of our societies so centrally relies. In order to reap AI's promise, while mitigating its risks, the adoption of a framework for its governance is indispensable, before realities on the ground render any attempts at such governance strategies futile.

Design Components & Constraints

In the interest of effectiveness, comprehensiveness, and durability, The Future Society's framework adheres to the following design constraints.

1. The framework seeks to be universal and free of cultural bias, but allow, where appropriate, for culture-specific implementation.
2. The framework seeks to be capacious enough to allow for experimentation and innovation, but also specific enough to enable meaningful mitigation of risk.
3. The framework seeks to be comprehensive in its coverage of AI, both forms currently in existence and forms yet to be invented.
4. All elements of the framework shall be transparent and accessible to all individuals.
5. The well-being of the individual human being shall be the standard against which the framework's effectiveness is measured.

The framework includes four components, sequenced from the general to the specifics:

1. Overarching value
2. General principles for the governance of AI
3. Recommendations for public policy
4. Implementation-level standards and codes of practice