OECD Global Parliamentary Network

Paris, 11 October 2019

OECD AI PRINCIPLES – THE ROLE OF MPS IN LEVERAGING THE BENEFITS OF AI

Dirk Pilat
Deputy Director
Directorate for Science, Technology and Innovation
dirk.pilat@oecd.org

OECD
BETTER POLICIES FOR BETTER LIVES

1. UNDERSTANDING THE CHALLENGE

What is Artificial Intelligence?

- **OECD understanding** of an "**AI System**": A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.

- Today's Artificial Intelligence is **"narrow" or "applied"** and designed to accomplish a specific problem-solving or reasoning task. Even the most advanced AI systems available today, such as IBM's Watson or Google's AlphaGo, are still "narrow".

- Applied AI can be contrasted to a (hypothetical) **Artificial General Intelligence (AGI)**, in which autonomous machines would become capable of general intelligent action, like a human being, including generalising and abstracting learning across different cognitive functions.

AI has a wide range of applications…

- **Transport**: autonomous vehicles with potential cost, safety and environmental benefits
- **Science**: accelerate discovery, facilitate reproducibility and lower experimentation costs.
- **Health**: detect health conditions, deliver preventative services, discover treatments, facilitate clinical research, and optimise health systems.
- **Criminal justice**: predictive policing, assessing recidivism risk, and predicting court procedure outcomes.
- **Security**: detect anomalies and threats in digital security and surveillance.
- **Agriculture**: agricultural robots, crop and soil monitoring and impact prediction.
- **Financial services**: detect fraud, assess credit worthiness, automate trading, etc.
- **Marketing and advertising**: target and personalise content, advertising and products, demand forecasting, translation, etc
- And, of course, **economic analysis**.

Policy challenges and questions

- Driving AI **innovation** while ensuring AI systems are **trustworthy**: fair, safe, transparent, with accountability mechanisms in place
- Impact on **growth and productivity**
- Impacts on **jobs** and **skills**, including how to foster and attract **AI talent**
- Improving **access to data** while safeguarding **privacy, security, and intellectual property rights**

2. THE OECD RESPONSE

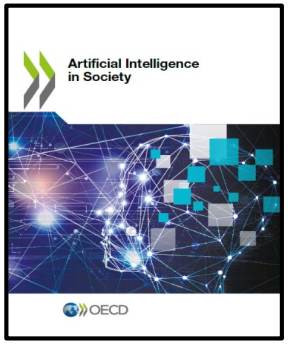
OECD work on Artificial Intelligence

- **G7 ICT Ministerial meeting in Japan**, Apr 2016

- **OECD events in 2016 and 2017,** incl.
  "AI: Intelligent Machines, Smart Policies", Oct 2017

- **AI expert group at the OECD (AIGO)**
  Sep 2018-Feb 2019
  - scoped principles to foster trust in AI
  - multi-stakeholder and multi-disciplinary

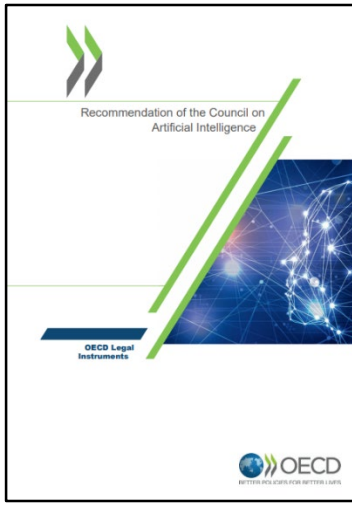
- **OECD Recommendation on AI**, MCM, May 2019
  - OECD-wide consultations
  - adopted by 42 countries (36 Members + 6 others)
  - basis for **G20 AI Principles**, June 2019



- **Measurement and analytical work:**
  Report on AI in Society, June 2019

OECD Principles on Trustworthy AI

10 Principles, covering two areas:



Principles for responsible stewardship of trustworthy AI

1.1. Inclusive growth, sustainable development and well-being
1.2. Human-centred values and fairness
1.3. Transparency and explainability
1.4. Robustness, security and safety
1.5. Accountability

National policies and international cooperation for trustworthy AI

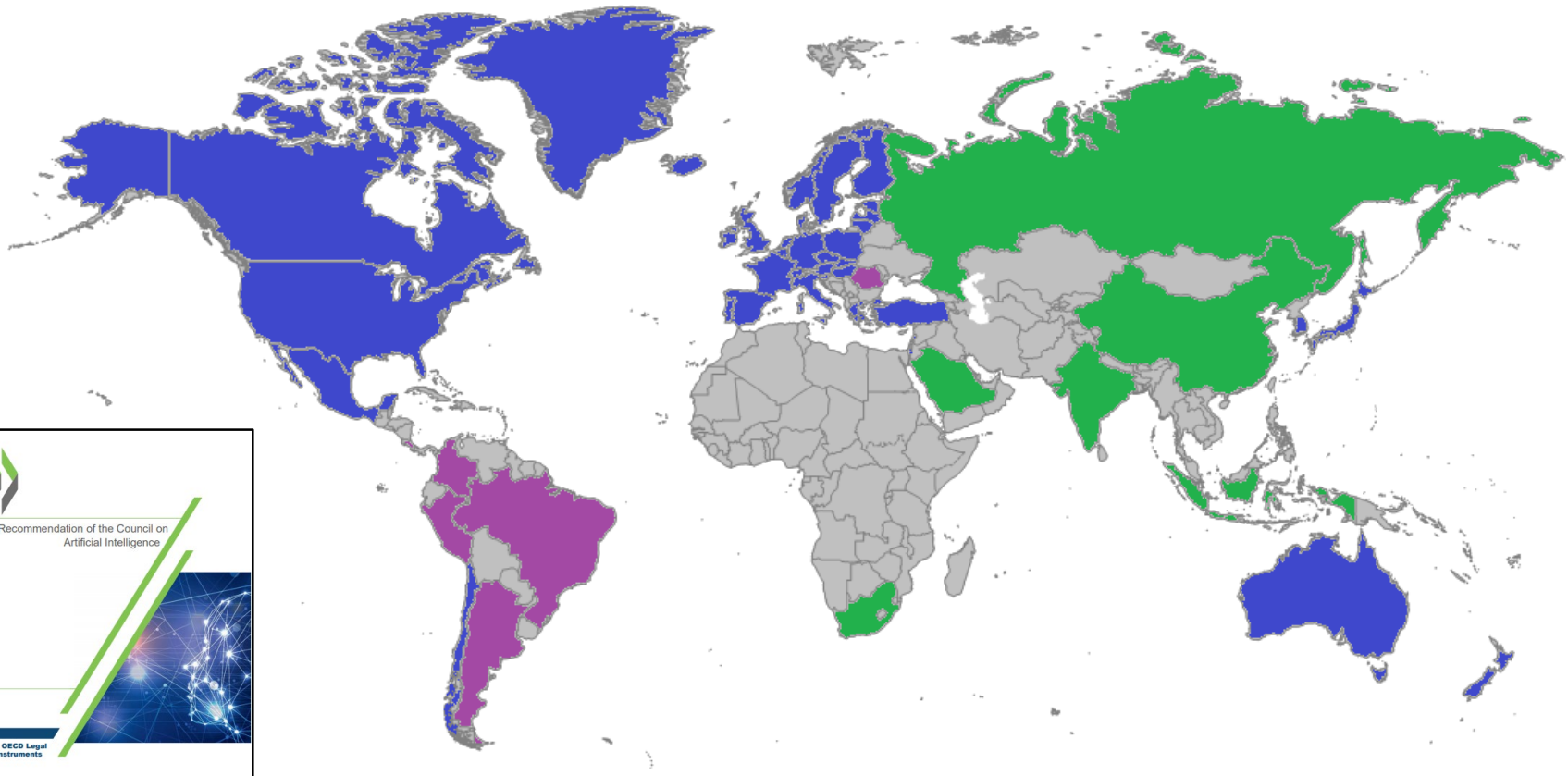
2.1. Investing in AI research and development
2.2. Fostering a digital ecosystem for AI
2.3. Providing an enabling policy environment for AI
2.4. Building human capacity and preparing for labour transition
2.5. International cooperation

Governments that have adhered to the OECD or OECD-based G20 AI principles

OECD members

Adherents

G20 principles, based on OECD



From principles to practice: OECD.AI

- Establishment of the **AI Policy Observatory**, to be launched **early 2020**
- Developing **practical guidance** on the implementation of the OECD AI Principles
  ⇔ Pillar 1 of **OECD.AI**, the AI Policy Observatory
  - Implementation guidance
  - rationale and illustrative actions,
  - implementation examples by countries (over time, best practices),
  - indicators
  - ***Continuously updated on OECD.AI***

3. EXAMPLES FROM COUNTRIES (WORK IN PROGRESS)

Principle 1: Inclusive growth, sustainable development and well-being

Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as (…) advancing inclusion of underrepresented populations, reducing (…) inequalities (…) thus invigorating inclusive growth, sustainable development and well-being.

- Canada's has recently recommended to establish *AI Centres of Excellence* in low- and-middle income countries.

Principle 2: Human-centred values and fairness

a) *AI actors should respect the rule of law, human rights and democratic values (…).*
b) *To this end, AI actors should implement mechanisms and safeguards (…) that are appropriate to the context and consistent with the state of art.*

- The *Ethics Guidelines for Trustworthy AI* developed by the independent high-level expert group on AI set up by the EU consider fundamental rights as a basis for trustworthy AI.
- Korea has established *Ethics Guidelines* to promote a human-oriented intelligent information society.

Principle 3: Transparency and explainability

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should (…):

i. *foster a general understanding of AI systems,*
ii. *make stakeholders aware of their interactions with AI systems, including in the workplace,*
iii. *enable those affected (…) understand the outcome, and*
iv. *enable those adversely affected by an AI system, to challenge its outcome (…).*

- The United Kingdom's *Project ExplAIn* is an initiative to create practical guidance explaining AI decisions.
- Japan's AI *Utilization Guidelines* provide methods to enhance explainability on the outcomes of AI systems.
- Denmark is collaborating with industry bodies to develop a *data ethics seal*.

Principle 4: Robustness, security and safety

a) *AI systems should be robust, secure and safe throughout their entire lifecycle so that (…) they function appropriately and do not pose unreasonable safety risk.*
b) *To this end, AI actors should ensure traceability (…) to enable analysis of the AI system's outcomes (…).*
c) *AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle (…).*

- Canada's Treasury Board *Directive on Automated Decision-Making* seeks to a consistent approach to risk management in AI across the public sector. Canada further developed an *Algorithmic Impact Assessment tool* to assess the potential impact of algorithms on citizens.

Principle 5: Accountability

AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.

- Singapore's *Model AI Governance Framework* helps organisations demonstrate accountability through practical measures in the areas of internal governance, risk management, operations and communications.

Thank you

Contact: dirk.pilat@oecd.org

OECD Going Digital website: http://oe.cd/goingdigital

OECD work on AI: http://www.oecd.ai