OECD.AI
Policy Observatory

1. OECD work to operationalise trustworthy AI

Andrew Wyckoff

High-Level Roundtable on Artificial Intelligence: Regulating Innovation and Innovating Regulation

22 March 2022

OECD

OECD AI Principles

OECD.AI Policy Observatory

10 Principles, covering two areas:

Principles for responsible stewardship of trustworthy AI

- Inclusive growth, sustainable development and well-being
- Human-centred values and fairness
- Transparency and explainability
- Robustness, security and safety
- Accountability

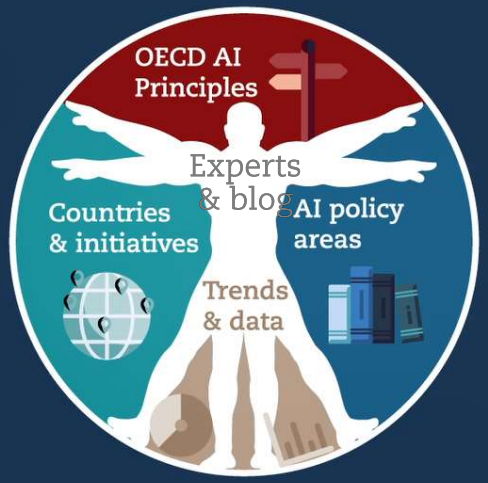
National policies and international cooperation for trustworthy AI

- Investing in AI research and development
- Fostering a digital ecosystem for AI
- Providing an enabling policy environment for AI
- Building human capacity and preparing for labour transition
- International cooperation

OECD AI Policy Observatory (OECD.AI)

A platform to share & shape public policies for responsible, trustworthy & beneficial AI

5 pillars:
- Network of experts and AI Wonk blog
- AI Principles & implementation
- AI trends & data
- AI policy areas
- Countries & initiatives



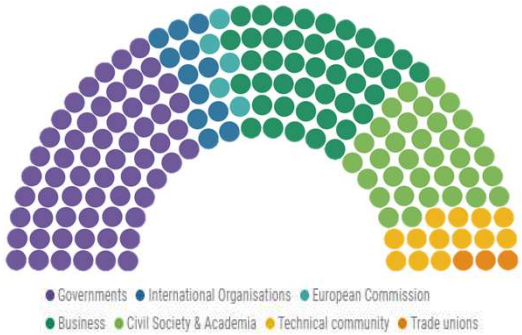
OECD Working Party on AI Governance and AI Experts Groups

Developing practical guidance to implement the AI Principles.

1 formal working party, 3 expert groups:
- OECD Working Party on AI Governance (AIGO)

Supported by:
- Expert Group on AI Classification & Risk
- Expert Group on AI Tools & Accountability
- Task force on AI compute

In addition to:
- The OECD Global Parliamentary Group on AI
- The Global Partnership on AI (GPAI)

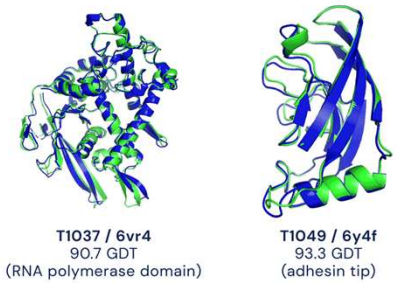


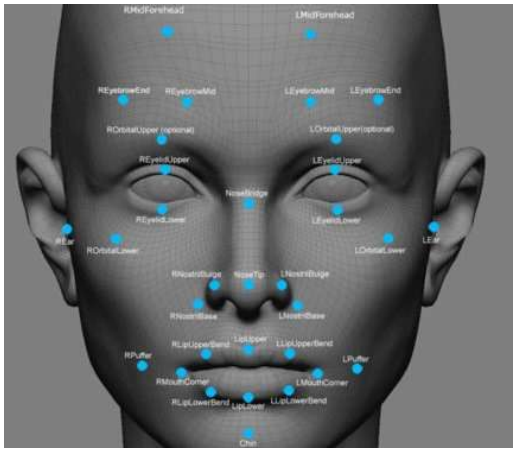
The OECD Framework for Classifying AI Systems

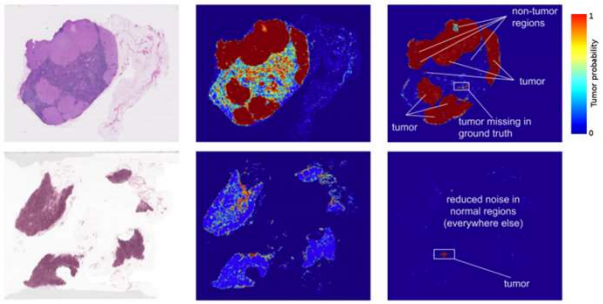
Why classify AI systems?

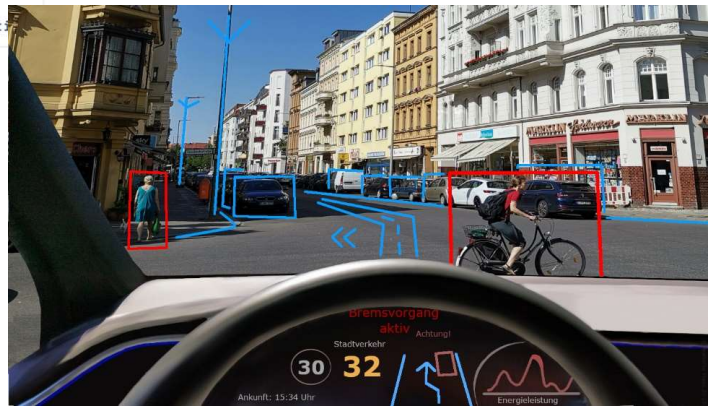
A variety of systems and policy implications

OECD.AI Policy Observatory

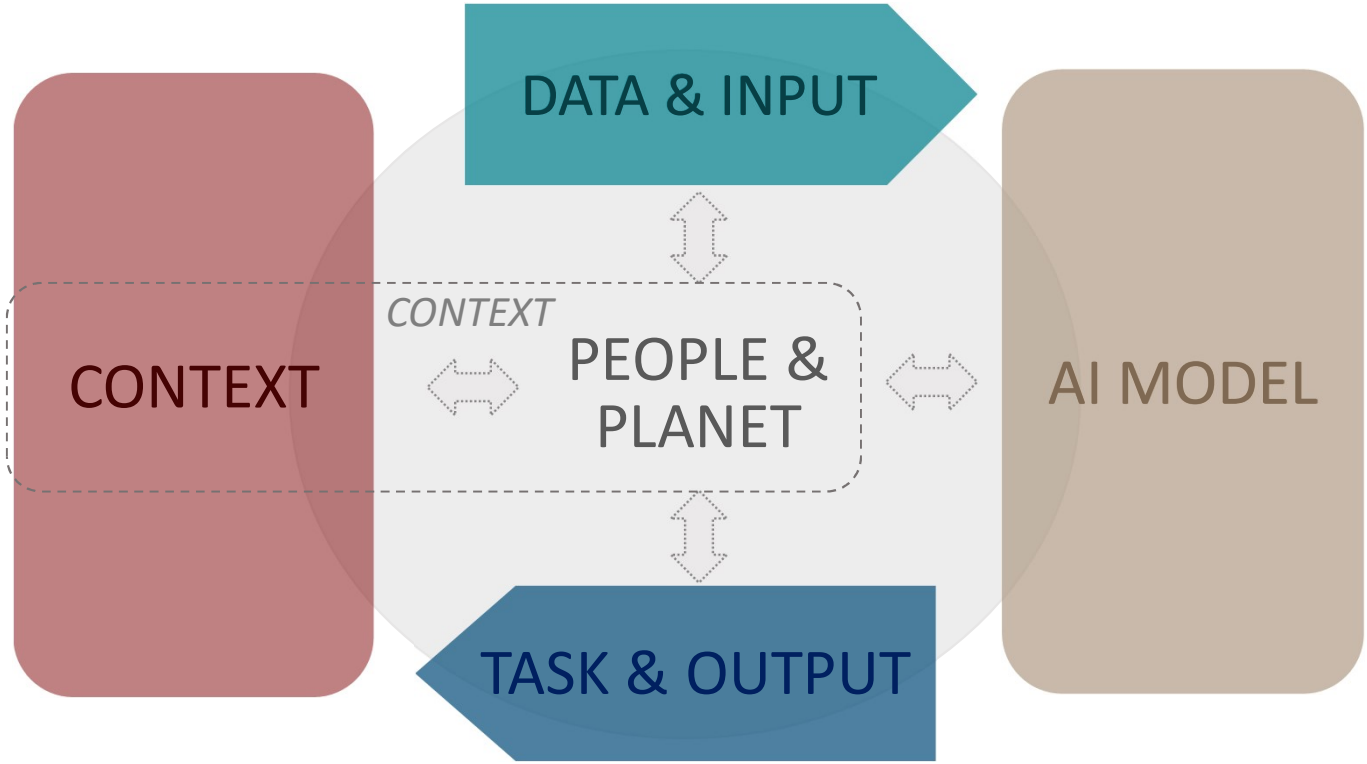




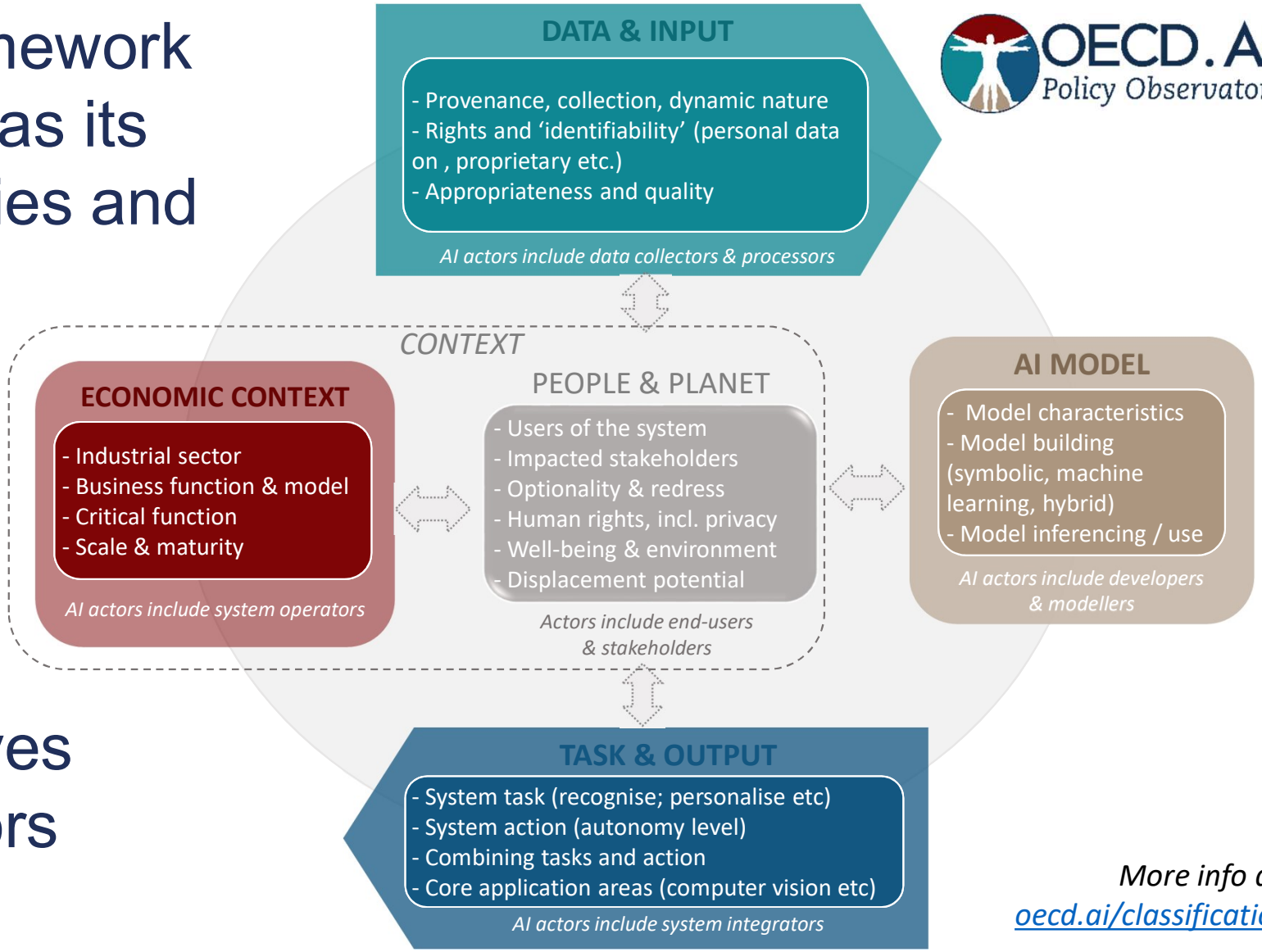




OECD Framework for Classifying AI systems: Key dimensions characterise AI systems' policy impact



Each AI framework dimension has its own properties and attributes…



…and involves specific actors

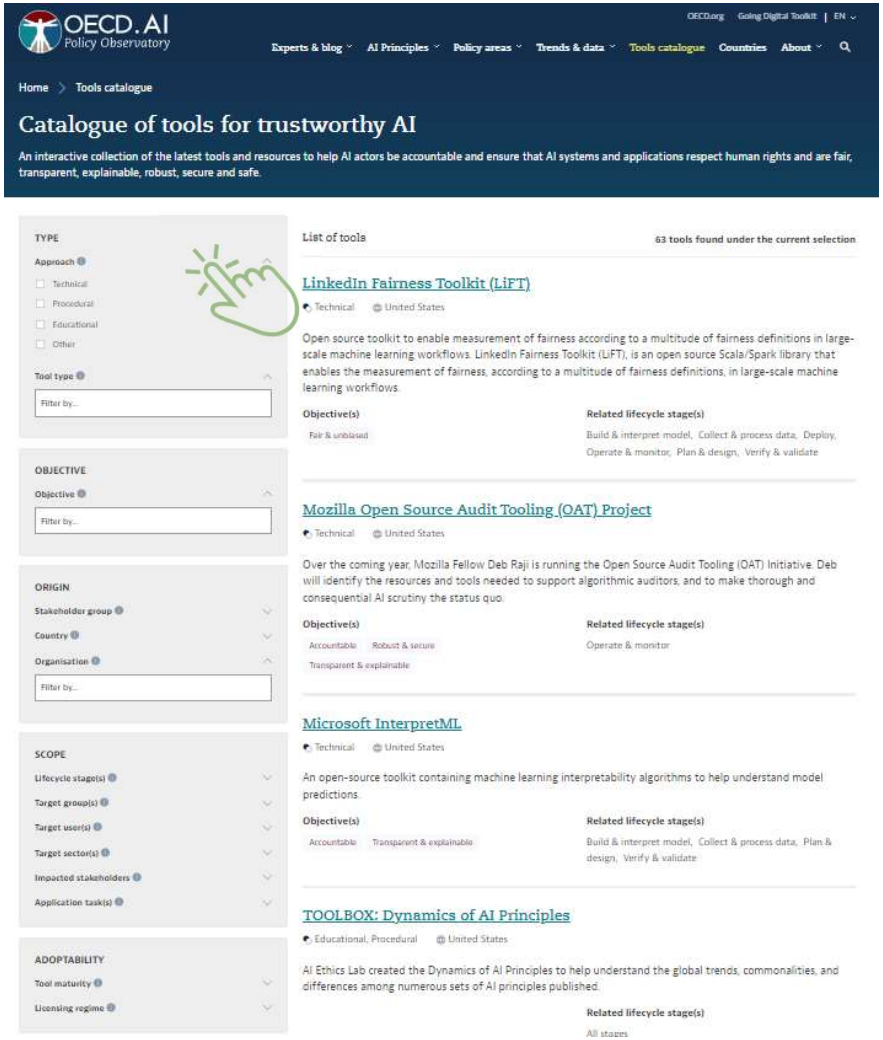
More info at:
oecd.ai/classification

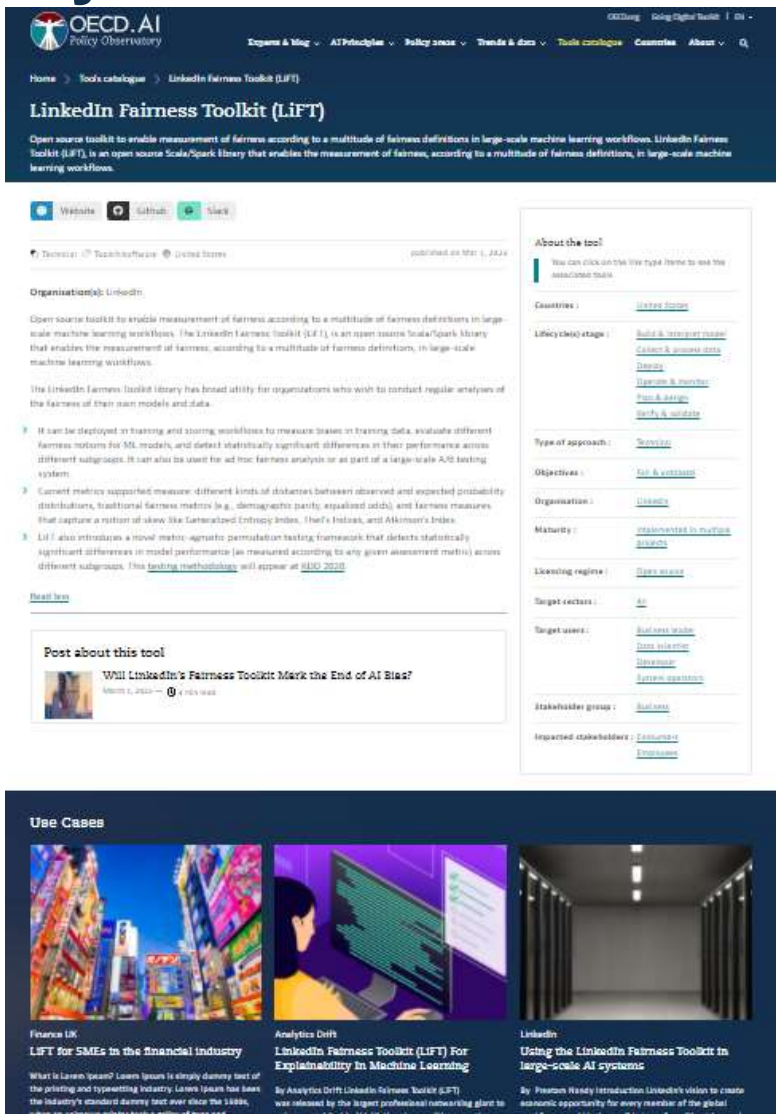
Next steps at the OECD:

- **Refine classification criteria**
  - Add more real-world AI systems and identify possible indicators
- **Develop a risk assessment framework to facilitate global interoperability**
  - Leveraging work in partner organisations, including EU, US, ISO
  - Leveraging risk assessment work in other parts of the OECD
- **Develop a common framework for reporting about AI incidents**
- **Support risk management**: Inform related work on mitigation, compliance and enforcement along the AI system lifecycle, and responsible business-impact assessment.

Catalogue of tools for trustworthy AI

Catalogue of tools for trustworthy AI





Analytical work on auditing AI systems

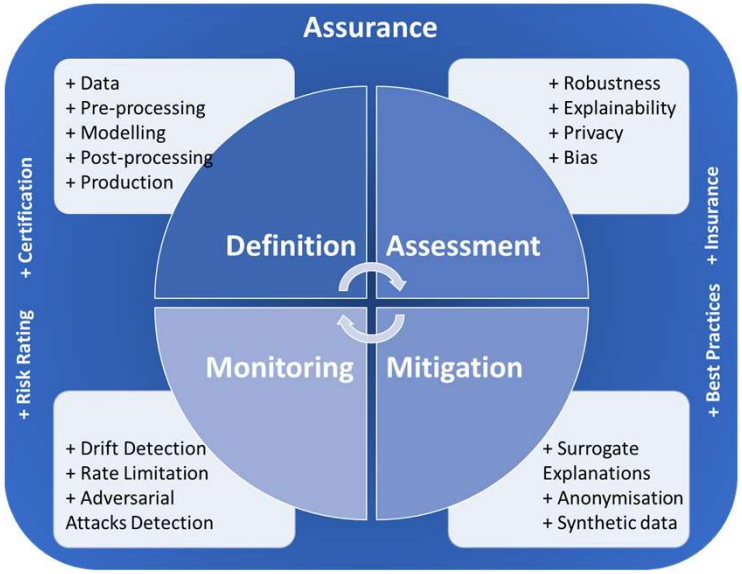
Next step: encouraging international interoperability in risk assessment

OECD.AI Policy Observatory

Lifecycle stage	*Use or are impact by*	Plan & design	Operate & monitor	Collect & process data	Build & use	Build & validate	Deploy

At each stage of the lifecycle conduct, conduct a risk assessment "DAMMA":

1. **DEFINE**: relevant principles for that stage, and relevant stakeholders and actors
2. **ASSESS**: risks to principles at individual and aggregate/societal levels (i.e., many small risks can amount to a big risk)
3. **MITIGATE**: in a way that is appropriate and commensurate to risk, considering likelihood and impact of risk
4. **MONITOR**: measure, evaluate and feedback results of the implementation
5. **ASSURE**: verify (audit, certify etc.) and communicate.



OECD.AI
Policy Observatory

For more information visit
www.oecd.ai

email: ai@oecd.org

OECD