

OECD Principles on Artificial Intelligence

What is it?

The OECD Principles on Artificial Intelligence promote artificial intelligence (AI) that is innovative and trustworthy and that respects human rights and democratic values. They were adopted on 22 May 2019 by OECD member countries when they approved the OECD Council Recommendation on Artificial Intelligence. The OECD AI Principles are the first such principles signed up to by governments. Beyond OECD members, other countries including Argentina, Brazil, Colombia and Costa Rica have already adhered to the AI Principles, with further adherents welcomed.

The OECD AI Principles sets standards for AI that are practical and flexible enough to stand the test of time in a rapidly evolving field. They complement existing OECD standards in areas such as privacy, digital security risk management and responsible business conduct. The general scope of the Principles ensures they can be applied to AI developments around the world.

Who developed them?

The OECD set up a 50+ member expert group on AI to scope a set of principles. The group consisted of representatives of 20 governments as well as leaders from the business, labour, civil society, academic and science communities. The experts' proposals were taken on by the OECD and developed into the OECD AI Principles.

OECD AI Principles

The *Recommendation* identifies five complementary values-based principles for the responsible stewardship of trustworthy AI:

- AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.
- AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and implement appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
- There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
- AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

What can governments do?

Consistent with these value-based principles, the *OECD* also provides five recommendations to governments:

- Facilitate public and private investment in research & development to spur innovation in trustworthy AI.
- Foster accessible AI ecosystems with digital infrastructure and technologies and mechanisms to share data and knowledge.

- Ensure a policy environment that will open the way to deployment of trustworthy AI systems.
- Empower people with the skills for AI and support workers for a fair transition.
- Co-operate across borders and sectors to progress on responsible stewardship of trustworthy AI.

What is an OECD Recommendation?

OECD Recommendations are not legally binding but they are highly influential and have many times formed the basis of international standards and helped governments design national legislation. For example, the OECD Privacy Guidelines adopted in 1980 and stating that there should be limits to the collection of personal data underlie many privacy laws and frameworks in the United States, Europe and Asia.

What's next?

A particular focus of the *Recommendation* is the development of metrics to measure AI research, development and deployment, and to gather the evidence base to assess progress in its implementation. The OECD's AI Policy Observatory will facilitate this by providing evidence and guidance on AI metrics, policies and practices to help implement the Principles, and constitute a hub to facilitate dialogue and share best practices on AI policies.

Further reading

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

OECD (forthcoming), *Artificial Intelligence in Society*, OECD Publishing, Paris, <https://doi.org/10.1787/eedfee77-en>.