

OECD Work on Implementing Trustworthy AI

Like the steam engine, electrification and computing, artificial intelligence (AI) is a general-purpose that has been transforming and revolutionizing everything we do. On a societal level, AI can improve the welfare and well-being of people, contribute to sustainable global economic activity, increase innovation and productivity, and help respond to key global challenges such as climate change. AI has as many potential benefits as it does challenges: economic shifts and inequalities, competition, transitions in the labour market, and implications for democracy and human rights.

As policy makers look for guidance and ways to work together and to get the most out of AI, the OECD is becoming the natural centre of gravity for AI policy by doing what it does best: facilitating international dialogue and collaboration while providing evidence-based policy guidance. In 2016, the Japanese Presidency to the G7 initiated the OECD's mandate to undertake empirical and policy analysis on AI in support of the policy debate, starting with a Technology Foresight Forum on AI in 2016.





In 2019, OECD member countries adopted the first intergovernmental standard on AI, the OECD [Recommendation on Artificial Intelligence](#) based upon the rigorous policy analysis that now makes up the publication [AI in Society](#). Now, the OECD focuses on developing evidence-based policy guidance to support countries' efforts to promote innovative and trustworthy AI. At the centre of this work is [OECD.AI](#) Policy Observatory.

Principles for Trustworthy AI

The OECD AI Principles provide guidance on how governments and other actors can shape a human-centric approach to trustworthy AI. As an OECD legal instrument, or soft law, the principles represent a high-level aspiration for the forty-six countries who have adhered to it. Adherents include OECD members and beyond, and the Principles have served as the basis for the G20 AI Principles, adopted in June 2019.

To develop the AI Principles, the OECD set up a multi-stakeholder, multi-disciplinary process built around a wide variety of expertise perspectives.

The five principles are value-based, to ensure that AI systems are trustworthy and human-centric. They are accompanied by five key actions that policy makers need to execute to foster thriving AI ecosystems that follow the principles and benefit societies.

Principles for responsible stewardship of trustworthy AI	National policies and international cooperation for trustworthy AI
 1.1. Inclusive growth, sustainable development and well-being	 2.1. Investing in AI research and development
 1.2. Human-centred values and fairness	 2.2. Fostering a digital ecosystem for AI
 1.3. Transparency and explainability	 2.3. Providing an enabling policy environment for AI
 1.4. Robustness, security and safety	 2.4. Building human capacity and preparing for labour transition
 1.5. Accountability	 2.5. International cooperation

Moving from principles to practice

Two key initiatives are underway at the OECD to implement the AI Principles. The first concentrates on building and sharing the evidence and data on AI policies, while the second focuses on developing tools and practical guidance to facilitate implementation.

OECD.AI: A platform to share and shape AI policies

Online since February 2020, the [OECD.AI](#) Policy Observatory is an online platform that fosters AI policy dialogue and provides access to:

- **Real-time trends and data** on AI development, research, jobs and skills.
- **National AI policies**, covering more than **600 policies** in over **60 countries** (in partnership with the European Commission).
- **Information about how AI impacts different policy domains**, from agriculture to healthcare or finance.
- **Analysis on the role of AI during the COVID-19 pandemic** ([OECD.AI/Covid-19](#)).
- **The AI Wonk: a blog** on cutting-edge AI research and policies ([OECD.AI/Wonk](#)).



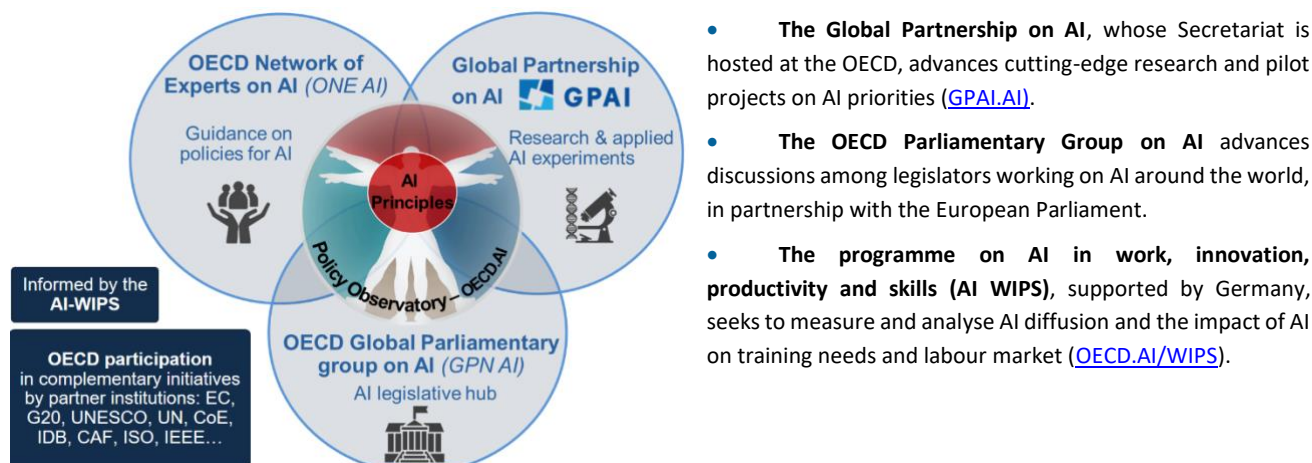
Tools and guidance for implementing the OECD AI Principles

Building on the expertise of a multi-stakeholder network of over 200 AI experts (ONE AI), the OECD allows policy makers and other actors from all disciplines and sectors to compare policy responses, learn from each other's experiences, monitor collective progress and to develop good practices. Some of the current tools that ONE AI's is focusing on are:

- **The OECD framework to classify AI systems for policymakers** will assess the policy implications and risks of different types of AI systems. A major component of this tool will be a risk assessment framework. The OECD AI system classification framework will be published by the end of 2021.
- **A database of tools to implement trustworthy AI** (a recent report is available at [OECD.AI/Tools](#)) and an analysis of the AI accountability ecosystems of public and private governance mechanisms to ensure accountable, trustworthy AI.
- **Practical guidance on national AI policies that promote AI deployment**, including R&D policies and policies to prepare societies and workers for the changes that AI will bring to the workplace. See the most recent **report on national policy implementation** at [OECD.AI/Policies](#).
- **Developing the means for measuring countries' AI compute infrastructure**, as one of the three key enablers of AI, alongside data and algorithms.

Fostering global dialogue and collaboration

The OECD's convening power and robust multi-stakeholder approach facilitates co-operation between all relevant actors: the private sector, the technical community, civil society, academia, governments, the regulatory community and other international organisations. Complementary initiatives on AI supported by the OECD include:



- **The Global Partnership on AI**, whose Secretariat is hosted at the OECD, advances cutting-edge research and pilot projects on AI priorities ([GPAI.AI](#)).
- **The OECD Parliamentary Group on AI** advances discussions among legislators working on AI around the world, in partnership with the European Parliament.
- **The programme on AI in work, innovation, productivity and skills (AI WIPS)**, supported by Germany, seeks to measure and analyse AI diffusion and the impact of AI on training needs and labour market ([OECD.AI/WIPS](#)).

More information

OECD (2021), "Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems", OECD Digital Economy Papers, No. 312, OECD Publishing, Paris, <https://doi.org/10.1787/008232ec-en>, available at: [OECD.AI/Tools](https://www.oecd.ai/Tools)

OECD (2021), "State of implementation of the OECD AI Principles: Insights from national AI policies", OECD Digital Economy Papers, No. 311, OECD Publishing, Paris, <https://doi.org/10.1787/1cd40c44-en>

Not all intelligence is artificial. Keep yours real with the AI Wonk blog [OECD.AI/Wonk](https://www.oecd.ai/Wonk)

<http://www.oecd.ai> and ai@oecd.org