Recommendation of the Council on Artificial Intelligence

OECD Legal Instruments

OECD

BETTER POLICIES FOR BETTER LIVES

For access to the official and up-to-date texts of OECD Legal Instruments, as well as other related information, please consult the Compendium of OECD Legal Instruments at http://legalinstruments.oecd.org.

Please cite this document as:
OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449

Series: OECD Legal Instruments

Photo credit: © kras99/Shutterstock.com

Background Information

The Recommendation on Artificial Intelligence (AI) (hereafter the "Recommendation") – the first intergovernmental standard on AI – was adopted by the OECD Council meeting at Ministerial level on 22 May 2019 on the proposal of the Digital Policy Committee (DPC, formerly the Committee on Digital Economy Policy, CDEP). The Recommendation aims to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values. In June 2019, at the Osaka Summit, G20 Leaders welcomed the G20 AI Principles, drawn from the Recommendation.

The Recommendation was revised by the OECD Council on 8 November 2023 to update its definition of an "AI System", in order to ensure the Recommendation continues to be technically accurate and reflect technological developments, including with respect to generative AI. On the basis of the 2024 Report to Council on its implementation, dissemination and continued relevance, the Recommendation was revised by the OECD Council meeting at Ministerial level on 3 May 2024 to reflect technological and policy developments, including with respect to generative AI, and to further facilitate its implementation.

The OECD's work on Artificial Intelligence

Artificial Intelligence (AI) is a general-purpose technology that has the potential to: improve the welfare and well-being of people, contribute to positive sustainable global economic activity, increase innovation and productivity, and help respond to key global challenges. It is deployed in many sectors ranging from production, education, finance and transport to healthcare and security.

Alongside benefits, AI also raises challenges for our societies and economies, notably regarding economic shifts and inequalities, competition, transitions in the labour market, and implications for democracy and human rights.

The OECD has undertaken empirical and policy activities on AI in support of the policy debate since 2016, starting with a Technology Foresight Forum on AI that year, followed by an international conference on AI: Intelligent Machines, Smart Policies in 2017. The Organisation also conducted analytical and measurement work that provides an overview of the AI technical landscape, maps economic and social impacts of AI technologies and their applications, identifies major policy considerations, and describes AI initiatives from governments and other stakeholders at national and international levels.

This work has demonstrated the need to shape a stable policy environment at the international level to foster trust in and adoption of AI in society. Against this background, the OECD Council adopted, on the proposal of DPC, a Recommendation to promote a human-centred approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and applies to all stakeholders.

An inclusive and participatory process for developing the Recommendation

The development of the Recommendation was participatory in nature, incorporating input from a broad range of sources throughout the process. In May 2018, the DPC agreed to form an expert group to scope principles to foster trust in and adoption of AI, with a view to developing a draft Recommendation in the course of 2019. The informal AI Group of experts at the OECD was subsequently established, comprising over 50 experts from different disciplines and different sectors (government, industry, civil society, trade unions, the technical community and academia) - see http://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf for the full list. Between September 2018 and February 2019 the group held four meetings. The work benefited from the diligence, engagement and substantive contributions of the experts participating in the group, as well as from their multi-stakeholder and multidisciplinary backgrounds.

Drawing on the final output document of the informal group, a draft Recommendation was developed in the DPC and with the consultation of other relevant OECD bodies and approved in a special meeting on 14-15 March 2019. The OECD Council adopted the Recommendation at its meeting at Ministerial level on 22-23 May 2019.

Scope of the Recommendation

Complementing existing OECD standards already relevant to AI – such as those on privacy and data protection, digital security risk management, and responsible business conduct – the Recommendation focuses on policy issues that are specific to AI and strives to set a standard that is implementable and flexible enough to stand the test of time in a rapidly evolving field. The Recommendation contains five high-level values-based principles and five recommendations for national policies and international co-operation. It also proposes a common understanding of key terms, such as "AI system", "AI system lifecycle", and "AI actors", for the purposes of the Recommendation.

More specifically, the Recommendation includes two substantive sections:

1. **Principles for responsible stewardship of trustworthy AI**: the first section sets out five complementary principles relevant to all stakeholders: i) inclusive growth, sustainable development and well-being; ii) respect for the rule of law, human rights and democratic values, including fairness and privacy; iii) transparency and explainability; iv) robustness, security and safety; and v) accountability. This section further calls on AI actors to promote and implement these principles according to their roles.
2. **National policies and international co-operation for trustworthy AI**: consistent with the five aforementioned principles, the second section provides five recommendations to Members and non-Members having adhered to the Recommendation (hereafter the "Adherents") to implement in their national policies and international co-operation: i) investing in AI research and development; ii) fostering an inclusive AI-enabling ecosystem; iii) shaping an enabling interoperable governance and policy environment for AI; iv) building human capacity and preparing for labour market transformation; and v) international co-operation for trustworthy AI.

2023 and 2024 Revisions of the Recommendation

In 2023, a window of opportunity was identified to maintain the relevance of the Recommendation by updating its definition of an "AI System", and the DPC approved a draft revised definition in a joint session of the Committee and its Working Party on AI Governance (AIGO) on 16 October 2023. The OECD Council adopted the revised definition of "AI System" at its meeting on 8 November 2023. The update of the definition included edits aimed at:

- clarifying the objectives of an AI system (which may be explicit or implicit);
- underscoring the role of input which may be provided by humans or machines;
- clarifying that the Recommendation applies to generative AI systems, which produce "content";
- substituting the word "real" with "physical" for clarity and alignment with other international processes;
- reflecting the fact that some AI systems can continue to evolve after their design and deployment.

In line with the conclusions of the 2024 Report to Council, the Recommendation was further revised at the 2024 Meeting of the Council at Ministerial level to maintain its continued relevance and facilitate its implementation five years after its adoption. Specific updates aimed at:

- reflecting the growing importance of addressing misinformation and disinformation, and safeguarding information integrity in the context of generative AI;
- addressing uses outside of intended purpose, intentional misuse, or unintentional misuse;
- clarifying the information AI actors should provide regarding AI systems to ensure transparency and responsible disclosure;
- addressing safety concerns, so that if AI systems risk causing undue harm or exhibit undesired behaviour, they can be overridden, repaired, and/or decommissioned safely by human interaction;
- emphasising responsible business conduct throughout the AI system lifecycle, involving co-operation with suppliers of AI knowledge and AI resources, AI system users, and other stakeholders,
- underscoring the need for jurisdictions to work together to promote interoperable governance and policy environments for AI, against the increase in AI policy initiatives worldwide, and
- introducing an explicit reference to environmental sustainability, of which the importance has

grown considerably since the adoption of the Recommendation in 2019.

Furthermore, some of the headings of the principles and recommendations were expanded for clarity, and the text on traceability and risk management was further elaborated and moved to the "Accountability" principle as the most appropriate principle for these concepts.

For further information please consult: oecd.ai.
Contact information: ai@oecd.org.

Implementation

The Recommendation instructs the DPC to report to the Council on its implementation, dissemination and continued relevance five years after its adoption and regularly thereafter.

2024 Report to Council

The DPC, through AIGO, developed a report to the Council on the implementation, dissemination and continued relevance of the Recommendation five years after its implementation, and proposed draft revisions drawing from its conclusions.

The 2024 Report concluded that the Recommendation provides a significant and useful international reference in national AI policymaking. The Recommendation is being implemented by its Adherents, is widely disseminated, and remains fully relevant, including as a solid framework to analyse technology evolutions such as those related to generative AI.

However, the 2024 Report found that updates were needed to clarify the substance of some of the Recommendation's provisions, facilitate implementation, increase relevance, and ensure the Recommendation reflects important technological developments, including with respect to generative AI.

Further work to support the implementation of the Recommendation

In addition to reporting to the Council on the implementation of the Recommendation, the DPC is also instructed to continue its work on AI, building on this Recommendation, and taking into account work in other international fora, such as UNESCO, the European Union, the Council of Europe and the initiative to build an International Panel on AI (see https://pm.gc.ca/eng/news/2018/12/06/mandate-international-panel-artificial- intelligence and https://www.gouvernement.fr/en/france-and-canada-create-new-expert-international-panel-on- artificial-intelligence).

In order to support implementation of the Recommendation, the Council instructed the DPC to develop practical guidance for implementation, to provide a forum for exchanging information on AI policy and activities, and to foster multi-stakeholder and interdisciplinary dialogue.

To provide an inclusive forum for exchanging information on AI policy and activities, and to foster multi-stakeholder and interdisciplinary dialogue, the OECD launched i) the AI Policy Observatory (OECD.AI) as well as ii) the informal OECD Network of Experts on AI (ONE AI) in February 2020.

OECD.AI is an inclusive hub for public policy on AI that aims to help countries encourage, nurture and monitor the responsible development of trustworthy artificial intelligence systems for the benefit of society. It combines resources from across the OECD with those of partners from all stakeholder groups to provide multidisciplinary, evidence-based policy analysis on AI. The Observatory includes a live database of AI strategies, policies and initiatives that countries and other stakeholders can share and update, enabling the comparison of their key elements in an interactive manner. It is continuously updated with AI metrics, measurements, policies and good practices that lead to further updates in the practical guidance for implementation.

The OECD.AI Network of Experts (ONE AI) is an informal group of AI experts from government, business, academia and civil society that provides AI-specific policy expertise and advice to the OECD. The network provides a space for the international AI community to have in-depth discussions about shared AI policy opportunities and challenges.

In 2022, in recognition of the growing importance and scope of its work on AI, the DPC created a dedicated Working Party to the topic – the Working Party on AI Governance (AIGO). The Working Party supports the implementation of the OECD Recommendation on Artificial Intelligence, develops practical guidance for implementation, provides a forum and online hub for exchanging information on AI policy and activities through the OECD.AI Policy Observatory, and foster multi-stakeholder and interdisciplinary dialogue through the Expert Groups of the OECD.AI Network of Experts.

THE COUNCIL,

HAVING REGARD to Article 5 b) of the Convention on the Organisation for Economic Co-operation and Development of 14 December 1960;

HAVING REGARD to standards developed by the OECD in the areas of privacy, digital security, consumer protection and responsible business conduct;

HAVING REGARD to the Sustainable Development Goals set out in the 2030 Agenda for Sustainable Development adopted by the United Nations General Assembly (A/RES/70/1) as well as the 1948 Universal Declaration of Human Rights;

HAVING REGARD to the important work being carried out on artificial intelligence (hereafter, "AI") in other international governmental and non-governmental fora;

RECOGNISING that AI has pervasive, far-reaching and global implications that are transforming societies, economic sectors and the world of work, and are likely to increasingly do so in the future;

RECOGNISING that AI has the potential to improve the welfare and well-being of people, to contribute to positive sustainable global economic activity, to increase innovation and productivity, and to help respond to key global challenges;

RECOGNISING that, at the same time, these transformations may have disparate effects within, and between societies and economies, notably regarding economic shifts, competition, transitions in the labour market, inequalities, and implications for democracy and human rights, privacy and data protection, and digital security;

RECOGNISING that trust is a key enabler of digital transformation; that, although the nature of future AI applications and their implications may be hard to foresee, the trustworthiness of AI systems is a key factor for the diffusion and adoption of AI; and that a well-informed whole-of-society public debate is necessary for capturing the beneficial potential of the technology, while limiting the risks associated with it;

UNDERLINING that certain existing national and international legal, regulatory and policy frameworks already have relevance to AI, including those related to human rights, consumer and personal data protection, intellectual property rights, responsible business conduct, and competition, while noting that the appropriateness of some frameworks may need to be assessed and new approaches developed;

RECOGNISING that given the rapid development and implementation of AI, there is a need for a stable policy environment that promotes a human-centric approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and that applies to all stakeholders according to their role and the context;

CONSIDERING that embracing the opportunities offered, and addressing the challenges raised, by AI applications, and empowering stakeholders to engage is essential to fostering adoption of trustworthy AI in society, and to turning AI trustworthiness into a competitive parameter in the global marketplace.

On the proposal of the Digital Policy Committee:

I. AGREES that for the purpose of this Recommendation the following terms should be understood as follows:

- *AI system*: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

- *AI system lifecycle*: An AI system lifecycle typically involves several phases that include to: plan and design; collect and process data; build model(s) and/or adapt existing model(s) to specific tasks; test, evaluate, verify and validate; make available for use/deploy; operate and monitor; and retire/decommission. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.
- *AI actors*: AI actors are those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI.
- *AI knowledge*: AI knowledge refers to the skills and resources, such as data, code, algorithms, models, research, know-how, training programmes, governance, processes, and best practices required to understand and participate in the AI system lifecycle, including managing risks.
- *Stakeholders*: Stakeholders encompass all organisations and individuals involved in, or affected by, AI systems, directly or indirectly. AI actors are a subset of stakeholders.

Section 1: Principles for responsible stewardship of trustworthy AI

II. RECOMMENDS that Members and non-Members adhering to this Recommendation (hereafter the "Adherents") promote and implement the following principles for responsible stewardship of trustworthy AI, which are relevant to all stakeholders.

III. CALLS ON all AI actors to promote and implement, according to their respective roles, the following principles for responsible stewardship of trustworthy AI.

IV. UNDERLINES that the following principles are complementary and should be considered as a whole.

1.1. 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 augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, well-being, sustainable development and environmental sustainability.

1.2. Respect for the rule of law, human rights and democratic values, including fairness and privacy

a) AI actors should respect the rule of law, human rights, democratic and human-centred values throughout the AI system lifecycle. These include non-discrimination and equality, freedom, dignity, autonomy of individuals, privacy and data protection, diversity, fairness, social justice, and internationally recognised labour rights. This also includes addressing misinformation and disinformation amplified by AI, while respecting freedom of expression and other rights and freedoms protected by applicable international law.

b) To this end, AI actors should implement mechanisms and safeguards, such as capacity for human agency and oversight, including to address risks arising from uses outside of intended purpose, intentional misuse, or unintentional misuse in a manner appropriate to the context and consistent with the state of the art.

1.3. Transparency and explainability

AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

i. to foster a general understanding of AI systems, including their capabilities and limitations,

ii. to make stakeholders aware of their interactions with AI systems, including in the workplace,

iii. where feasible and useful, to provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output, and,

iv. to provide information that enable those adversely affected by an AI system to challenge its output.

1.4. Robustness, security and safety

a) AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety and/or security risks.

b) Mechanisms should be in place, as appropriate, to ensure that if AI systems risk causing undue harm or exhibit undesired behaviour, they can be overridden, repaired, and/or decommissioned safely as needed.

c) Mechanisms should also, where technically feasible, be in place to bolster information integrity while ensuring respect for freedom of expression.

1.5. Accountability

a) 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 the art.

b) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outputs and responses to inquiry, appropriate to the context and consistent with the state of the art.

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 on an ongoing basis and adopt responsible business conduct to address risks related to AI systems, including, as appropriate, via co-operation between different AI actors, suppliers of AI knowledge and AI resources, AI system users, and other stakeholders. Risks include those related to harmful bias, human rights including safety, security, and privacy, as well as labour and intellectual property rights.

Section 2: National policies and international co-operation for trustworthy AI

V. RECOMMENDS that Adherents implement the following recommendations, consistent with the principles in section 1, in their national policies and international co-operation, with special attention to small and medium-sized enterprises (SMEs).

2.1. Investing in AI research and development

a) Governments should consider long-term public investment, and encourage private investment, in research and development and open science, including interdisciplinary efforts, to spur innovation in trustworthy AI that focus on challenging technical issues and on AI-related social, legal and ethical implications and policy issues.

b) Governments should also consider public investment and encourage private investment in open-source tools and open datasets that are representative and respect privacy and data protection to support an environment for AI research and development that is free of harmful bias and to improve interoperability and use of standards.

2.2. Fostering an inclusive AI-enabling ecosystem

Governments should foster the development of, and access to, an inclusive, dynamic, sustainable, and interoperable digital ecosystem for trustworthy AI. Such an ecosystem includes inter alia, data, AI technologies, computational and connectivity infrastructure, and mechanisms for sharing AI knowledge, as appropriate. In this regard, governments should consider promoting mechanisms, such as data trusts, to support the safe, fair, legal and ethical sharing of data.

2.3. Shaping an enabling interoperable governance and policy environment for AI

a) Governments should promote an agile policy environment that supports transitioning from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate. They should also adopt outcome-based approaches that provide flexibility in achieving governance objectives and co-operate within and across jurisdictions to promote interoperable governance and policy environments, as appropriate.

b) Governments should review and adapt, as appropriate, their policy and regulatory frameworks and assessment mechanisms as they apply to AI systems to encourage innovation and competition for trustworthy AI.

2.4. Building human capacity and preparing for labour market transformation

a) Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills.

b) Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programmes along the working life, support for those affected by displacement, including through social protection, and access to new opportunities in the labour market.

c) Governments should also work closely with stakeholders to promote the responsible use of AI at work, to enhance the safety of workers, the quality of jobs and of public services, to foster entrepreneurship and productivity, and aim to ensure that the benefits from AI are broadly and fairly shared.

2.5. International co-operation for trustworthy AI

a) Governments, including developing countries and with stakeholders, should actively co-operate to advance these principles and to progress on responsible stewardship of trustworthy AI.

b) Governments should work together in the OECD and other global and regional fora to foster the sharing of AI knowledge, as appropriate. They should encourage international, cross-sectoral and open multi-stakeholder initiatives to garner long-term expertise on AI.

c) Governments should promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI.

d) Governments should also encourage the development, and their own use, of internationally comparable indicators to measure AI research, development and deployment, and gather the evidence base to assess progress in the implementation of these principles.

VI. **INVITES** the Secretary-General and Adherents to disseminate this Recommendation.

VII. **INVITES** non-Adherents to take due account of, and adhere to, this Recommendation.

VIII. **INSTRUCTS** the Digital Policy Committee, through its Working Party on AI Governance, to:

a) continue its important work on artificial intelligence building on this Recommendation and taking into account work in other international fora, and to further develop the measurement framework for evidence-based AI policies;

b) develop and iterate further practical guidance on the implementation of this Recommendation to meet evolving developments and new policy priorities;

c) provide a forum for exchanging information on AI policy and activities including experience with the implementation of this Recommendation, and to foster multi-stakeholder and interdisciplinary dialogue to promote trust in and adoption of AI; and

d) report to Council, in consultation with other relevant committees, on the implementation, dissemination and continued relevance of this Recommendation no later than five years following its revision and at least every ten years thereafter.

About the OECD

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD Member countries are: Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Legal Instruments

Since the creation of the OECD in 1961, around 460 substantive legal instruments have been developed within its framework. These include OECD Acts (i.e. the Decisions and Recommendations adopted by the OECD Council in accordance with the OECD Convention) and other legal instruments developed within the OECD framework (e.g. Declarations, international agreements).

All substantive OECD legal instruments, whether in force or abrogated, are listed in the online Compendium of OECD Legal Instruments. They are presented in five categories:

- **Decisions** are adopted by Council and are legally binding on all Members except those which abstain at the time of adoption. They set out specific rights and obligations and may contain monitoring mechanisms.

- **Recommendations** are adopted by Council and are not legally binding. They represent a political commitment to the principles they contain and entail an expectation that Adherents will do their best to implement them.

- **Substantive Outcome Documents** are adopted by the individual listed Adherents rather than by an OECD body, as the outcome of a ministerial, high-level or other meeting within the framework of the Organisation. They usually set general principles or long-term goals and have a solemn character.

- **International Agreements** are negotiated and concluded within the framework of the Organisation. They are legally binding on the Parties.

- **Arrangement, Understanding and Others**: several other types of substantive legal instruments have been developed within the OECD framework over time, such as the Arrangement on Officially Supported Export Credits, the International Understanding on Maritime Transport Principles and the Development Assistance Committee (DAC) Recommendations.