Recommendation of the Council on Artificial Intelligence



OECD Legal Instruments

OECD

BETTER POLICIES FOR BETTER LIVES

This document is published under the responsibility of the Secretary-General of the OECD. It reproduces an OECD Legal Instrument and may contain additional material. The opinions expressed and arguments employed in the additional material do not necessarily reflect the official views of OECD Member countries.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

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

© OECD 2019

This document is provided free of charge. It may be reproduced and distributed free of charge without requiring any further permissions, as long as it is not altered in any way. It may not be sold.

This document is available in the two OECD official languages (English and French). It may be translated into other languages, as long as the translation is labelled "unofficial translation" and includes the following disclaimer: *"This translation has been prepared by [NAME OF TRANSLATION AUTHOR] for informational purpose only and its accuracy cannot be guaranteed by the OECD. The only official versions are the English and French texts available on the OECD website http://legalinstruments.oecd.org"*

Date(s)

Adopted on 22/05/2019

Background Information

The Recommendation on Artificial Intelligence (AI) – the first intergovernmental standard on AI – was adopted by the OECD Council at Ministerial level on 22 May 2019 on the proposal of 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. Complementing existing OECD standards in areas such as privacy, digital security risk management, and responsible business conduct, the Recommendation focuses on AI-specific issues and sets a standard that is implementable and sufficiently flexible to stand the test of time in this rapidly evolving field.

The Recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI and calls on AI actors to promote and implement them:

- inclusive growth, sustainable development and well-being;
- human-centred values and fairness;
- transparency and explainability;
- robustness, security and safety;
- and accountability.

In addition to and consistent with these value-based principles, the Recommendation also provides five recommendations to policy-makers pertaining to national policies and international co-operation for trustworthy AI, namely:

- investing in AI research and development;
- fostering a digital ecosystem for AI;
- shaping an enabling policy environment for AI;
- building human capacity and preparing for labour market transformation;
- and international co-operation for trustworthy AI.

The Recommendation also includes a provision for the development of metrics to measure AI research, development and deployment, and for building an evidence base to assess progress in its implementation.

The OECD's work on Artificial Intelligence and rationale for developing the OECD Recommendation on Artificial Intelligence

Artificial Intelligence (AI) is a general-purpose technology that 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. It is deployed in many sectors ranging from production, 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 over the past two years, starting with a Technology Foresight Forum on AI in 2016 and 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 Committee on Digital Economy Policy (CDEP) agreed to develop a draft Council Recommendation to promote a human-centric approach to trustworthy AI, that fosters research, preserves economic incentives to innovate, and applies to all stakeholders.

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" 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)* human-centred values and fairness; *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, this section provides five recommendations to Members and non-Members having adhered to the draft Recommendation (hereafter the "Adherents") to implement in their national policies and international co-operation: *i)* investing in AI research and development; *ii)* fostering a digital ecosystem for AI; *iii)* shaping an enabling policy environment for AI; *iv)* building human capacity and preparing for labour market transformation; and *v)* international co-operation for trustworthy AI.

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 CDEP agreed to form an expert group to scope principles to foster trust in and adoption of AI, with a view to developing a draft Council Recommendation in the course of 2019. The AI Group of experts at the OECD (AIGO) 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: in Paris, France, in September and November 2018, in Cambridge, MA, United States, at the Massachusetts Institute of Technology (MIT) in January 2019, back to back with the MIT AI Policy Congress, and finally in Dubai, United Arab Emirates, at the World Government Summit in February 2019. The work benefited from the diligence, engagement and substantive contributions of the experts participating in AIGO, as well as from their multi-stakeholder and multidisciplinary backgrounds.

Drawing on the final output document of the AIGO, a draft Recommendation was developed in the CDEP and with the consultation of other relevant OECD bodies. The CDEP approved a final draft Recommendation and agreed to transmit it to the OECD Council for adoption 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.

Follow-up, monitoring of implementation and dissemination tools

The OECD Recommendation on AI provides the first intergovernmental standard for AI policies and a foundation on which to conduct further analysis and develop tools to support governments in their implementation efforts. In this regard, it instructs the CDEP to monitor the implementation of the Recommendation and report to the Council on its implementation and continued relevance five years after its adoption and regularly thereafter. The CDEP 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 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 CDEP 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. This will be achieved largely through the OECD AI Policy Observatory, 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 will combine resources from across the OECD with those of partners from all stakeholder groups to provide multidisciplinary, evidence-based policy analysis on AI. The Observatory is planned to be launched late 2019 and will include 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 will also be continuously updated with AI metrics, measurements, policies and good practices that could lead to further updates in the practical guidance for implementation.

The Recommendation is open to non-OECD Member adherence, underscoring the global relevance of OECD AI policy work as well as the Recommendation's call for international co-operation.

Unofficial translation(s): German.

For further information please consult: oecd.ai.

Contact information: ai@oecd.org.

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 the OECD Guidelines for Multinational Enterprises [OECD/LEGAL/0144]; Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data [OECD/LEGAL/0188]; Recommendation of the Council concerning Guidelines for Cryptography Policy [OECD/LEGAL/0289]; Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information [OECD/LEGAL/0362]; Recommendation of the Council on Digital Security Risk Management for Economic and Social Prosperity [OECD/LEGAL/0415]; Recommendation of the Council on Consumer Protection in E-commerce [OECD/LEGAL/0422]; Declaration on the Digital Economy: Innovation, Growth and Social Prosperity (Cancún Declaration) [OECD/LEGAL/0426]; Declaration on Strengthening SMEs and Entrepreneurship for Productivity and Inclusive Growth [OECD/LEGAL/0439]; as well as the 2016 Ministerial Statement on Building more Resilient and Inclusive Labour Markets, adopted at the OECD Labour and Employment Ministerial Meeting;

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 Committee on Digital Economy Policy:

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 can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
- *AI system lifecycle*: AI system lifecycle phases involve: *i)* 'design, data and models'; which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; *ii)* 'verification and validation'; *iii)* 'deployment'; and *iv)* 'operation and monitoring'. 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 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.
- *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.
- *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, sustainable development and well-being.

1.2. Human-centred values and fairness

a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.

b) To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of 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,

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

iii. to enable those affected by an AI system to understand the outcome, and,

iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.

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 risk.

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 outcomes and responses to inquiry, appropriate to the context and consistent with the state of 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 a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.

1.5. 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.

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, 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 datasets that are representative and respect privacy and data protection to support an environment for AI research and development that is free of inappropriate bias and to improve interoperability and use of standards.

2.2. Fostering a digital ecosystem for AI

Governments should foster the development of, and access to, a digital ecosystem for trustworthy AI. Such an ecosystem includes in particular digital technologies and 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 policy environment for AI

a) Governments should promote a policy environment that supports an agile transition 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.

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, 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 and the quality of jobs, 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 metrics 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 Committee on Digital Economy Policy:

a) to 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) to develop and iterate further practical guidance on the implementation of this Recommendation, and to report to the Council on progress made no later than end December 2019;

c) to 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) to monitor, in consultation with other relevant Committees, the implementation of this Recommendation and report thereon to the Council no later than five years following its adoption and regularly thereafter.

Adherents*

OECD Members		Non-Members	Other
Australia	United States	Argentina	
Austria		Brazil	
Belgium		Colombia	
Canada		Costa Rica	
Chile		Peru	
Czech Republic		Romania	
Denmark			
Estonia			
Finland			
France			
Germany			
Greece			
Hungary			
Iceland			
Ireland			
Israel			
Italy			
Japan			
Korea			
Latvia			
Lithuania			
Luxembourg			
Mexico			
Netherlands			
New Zealand			
Norway			
Poland			
Portugal			
Slovak Republic			
Slovenia			
Spain			
Sweden			
Switzerland			
Turkey			
United Kingdom			

*Additional information and statements are available in the Compendium of OECD Legal Instruments: http://legalinstruments.oecd.org

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, 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, Turkey, 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 450 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**: OECD legal instruments which are legally binding on all Members except those which abstain at the time of adoption. While they are not international treaties, they entail the same kind of legal obligations. Adherents are obliged to implement Decisions and must take the measures necessary for such implementation.

- **Recommendations**: OECD legal instruments which are not legally binding but practice accords them great moral force as representing the political will of Adherents. There is an expectation that Adherents will do their utmost to fully implement a Recommendation. Thus, Members which do not intend to do so usually abstain when a Recommendation is adopted, although this is not required in legal terms.

- **Declarations**: OECD legal instruments which are prepared within the Organisation, generally within a subsidiary body. They usually set general principles or long-term goals, have a solemn character and are usually adopted at Ministerial meetings of the Council or of committees of the Organisation.

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

- **Arrangement, Understanding and Others**: several ad hoc 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.