Organisation for Economic Co-operation and Development

C/MIN(2024)17

For Official Use **English - Or. English**

24 April 2024

COUNCIL

Meeting of the Council at Ministerial Level, 2-3 May 2024

REPORT ON THE IMPLEMENTATION OF THE OECD RECOMMENDATION ON ARTIFICIAL INTELLIGENCE

JT03542423

Table of contents

FIGURES

TABLES

EXECUTIVE SUMMARY

1. This Report provides an assessment of the implementation, dissemination, and continued relevance of the OECD Recommendation on Artificial Intelligence (AI) [OECD/LEGAL/0449] (the "Recommendation" or "AI Recommendation"), the first inter-governmental standard in AI. The OECD AI Principles set out in the Recommendation provided the basis for the G20 AI Principles endorsed by Leaders in June 2019. The Principles aim to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI and ensuring respect for human rights and democratic values. As of December 2023, in addition to the 38 OECD Members, eight non-Members (Argentina, Brazil, Egypt, Malta, Peru, Romania, Singapore and Ukraine), have adhered to the Recommendation (the "Adherents").

2. The technological, economic, policy and geopolitical landscape has evolved significantly since the adoption of the Recommendation in 2019, notably with the rise of AI models that can be used for many different purposes, including language and multimodal models that can generate novel content, transpose text-to-video and-image, and interact with people in natural language through chatbots. While existing regulation and legislation apply to AI, OECD countries are considering proposals for a diverse set of regulatory measures including AI-specific regulatory frameworks (both cross-cutting and sectoral). Against this backdrop, the review of the Recommendation's implementation is particularly timely, making it an opportunity to both take stock of developments and trends over the past five years, assess the continued relevance of the OECD AI Principles, and identify potential next steps.

3. To ensure the continued relevance of the Recommendation, the OECD definition of an "AI system" was already revised in November 2023 to ensure it continues to be technically accurate and reflect technological developments, including with respect to generative AI.

4. The Report's findings show that the Recommendation is being implemented by Adherents. The main tool to track progress in implementation at the national level is a survey of national AI policies updated twice a year, that feeds into the OECD.AI Policy Observatory. Since 2019, Adherents have advanced national and international level initiatives both to follow the five policy recommendations to governments and to translate the values-based OECD AI Principles into action.

5. In 2017, only a few countries had national AI strategies. Today, the OECD.AI Policy Observatory contains over 50 national strategic and government-wide initiatives on how to comprehensively steer trustworthy AI development and deployment, with 41 Adherents currently having a national AI Strategy in place, and three in the process of developing one. While each country's national AI strategy is unique, a mapping of national AI strategies to the OECD AI Principles shows significant commonalities with the OECD AI Principles. For example, most strategies focus on inclusive growth, sustainable development and well-being, human-centred values and fairness, investing in AI R&D and building human capacity.

6. The number of initiatives and countries included in the observatory has consistently increased: as of January 2024, the database included information on 1 020 initiatives in 70 jurisdictions, out of which 850 initiatives were reported by the 46 Adherents to the Recommendation and the European Union. The wealth of policy initiatives and their direct applicability to the OECD AI Principles demonstrate Adherents' commitment and continuing efforts in implementing the Principles as well as their continued relevance.

7. Several jurisdictions also explicitly refer to or draw from the OECD AI Principles and subsequent OECD analytical work in their national guidelines, legislative actions, or voluntary frameworks. This is the case for instance in Israel's and the United Kingdom's cross-sectoral principles, which reflect the OECD AI Principles. Japan referenced them in major government initiatives, including for formulating discussions at the AI Strategy Council. The AI Risk

Management Framework by the United States' National Institute of Standards and Technologies uses the OECD Framework of Classification of AI Systems, while the European Union's EU AI Act and the Council of Europe's Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law adopted the updated OECD definition of an "AI system" included in the Recommendation. Furthermore, several Adherents, including Korea, Italy, Lithuania and Türkiye leveraged the OECD AI Principles as foundational pillars in their national AI strategies and governance frameworks.

8. The analysis in this Report indicates that the OECD AI Principles provide a significant and useful international reference in national AI policymaking by Adherents. They are widely disseminated, and remain fully relevant, including as a solid framework to analyse recent evolutions such as those related to generative AI.

9. The Recommendation in its current form continues to be fit for purpose overall. However, there is an opportunity to update the Recommendation to support implementation by stakeholders and to reflect emerging issues and technological advancements, including with respect to generative AI. Specific updates: i) reflect the growing importance of addressing misinformation and disinformation, and safeguarding information integrity in the context of generative AI; ii) address uses outside of intended purpose, intentional misuse, or unintentional misuse; iii) clarify the information AI actors should provide regarding AI systems to ensure transparency and responsible disclosure; iv) outline mechanisms to address potential harm or undesired behavior throughout the AI lifecycle; and v) emphasise responsible business conduct throughout the AI lifecycle, involving co-operation with suppliers of AI knowledge and AI resources, AI system users, and other stakeholders. Furthermore, some of the Principles' headings and texts could be expanded, and the text on traceability and risk management further developed and moved to the "Accountability" Principle as the most appropriate Principle for these concepts.

1. Background

10. The OECD began undertaking empirical and policy activities on artificial intelligence (AI) in 2016 with its Technology Foresight Forum on AI. The OECD then organised conferences and conducted analytical and measurement work that provides an overview of the AI technical landscape, economic and social impacts, as well as policy considerations.

11. 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 Recommendation on Artificial Intelligence (AI) [OECD/LEGAL/0449] was developed through an inclusive and participatory process, integrating input from a broad range of stakeholders. In particular, the work was led by the AI Group of experts at the OECD, comprising over 50 experts from different disciplines and different sectors (government, industry, civil society, trade unions, the technical community and academia).

12. Drawing on the final output document of the AIGO, a draft Recommendation was developed in the Digital Policy Committee (DPC)[1] and with the consultation of other relevant OECD bodies and approved in a special meeting on 14-15 March 2019. On 22 May 2019, the Recommendation was adopted by the OECD Council at Ministerial level [C/MIN(2019)3/FINAL and C/M(2019)10, Item 102], becoming the first intergovernmental standard on AI (hereafter, the "Recommendation", or "AI Recommendation"). 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.

[1] Formerly the Committee on Digital Economy Policy (CDEP).

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 (OECD, 2019[1]).

13. The Recommendation identifies five complementary values-based Principles and five recommendations to policymakers. Together, these five value-based principles and five recommendations are referred to in this Report as the "OECD AI Principles". (Table 1.1). The Recommendation also calls 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 Recommendation helps standardise language between Adherents by providing definitions of key concepts such as a definition of an "AI system".

Table 1.1. The ten Principles of the OECD Recommendation on Artificial Intelligence

Principles for responsible stewardship of trustworthy AI	National policies and international co-operation 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 Shaping an enabling policy environment for AI
1.4 Robustness, security and safety	2.4 Building human capacity and preparing for labour market transformation
1.5 Accountability	2.5 International co-operation for trustworthy AI

Source: OECD/LEGAL/0449.

14. The Recommendation was revised by the Council on 8 November 2023 [C(2023)151 and C/M(2023)14, Item 218] to update the definition of an "AI System" in order to ensure that it continues to be technically accurate and reflect technological developments, including with respect to generative AI. In particular, the update of the definition aimed to: (i) clarify the objectives of an AI system (which may be explicit or implicit); (ii) underscore the role of input which may be provided by humans or machines; (iii) clarify that the Recommendation applies to generative AI systems, which produce "content"; (iv) substitute the word "real" with "physical" for clarity and alignment with other international processes; and (v) reflect the fact that some AI systems can continue to evolve after their design and deployment. In addition, DPC, on the proposal of AIGO, approved an Explanatory Memorandum to accompany the updated definition of an "AI System" on 15 December 2023 (OECD, 2024[2]).

15. The Recommendation is open to adherence by non-Members. As of December 2023, in addition to the 38 OECD Members, eight non-Members (Argentina, Brazil, Egypt, Malta, Peru, Romania, Singapore and Ukraine), have adhered to the Recommendation.

16. When adopting the Recommendation, the Council instructed the DPC "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".

17. The present Report addresses the implementation, dissemination, and continued relevance of the Recommendation five years after its adoption. During this period, there have been considerable advancements in AI, both in the technological and in the policy landscape. Against this backdrop, reporting to the OECD Council on implementation, dissemination and continued relevance of the Recommendation is particularly timely, making it an opportunity to both take stock of developments and trends in the past five years, and to identify their impact on the relevance of the Recommendation and potential next steps.

2. Methodology

18. The OECD began keeping track of implementation of AI policies in 2020 through a survey of national AI policies, available on a rolling basis on the OECD.AI Policy Observatory. National contact points in the reporting jurisdictions (currently 70) are invited twice per year to report national and regional AI policy developments, according to a structured survey, organised in policy initiatives, policy instruments and other specific fields, such as description, budget and target groups. The data feed into the OECD.AI database of national AI policies on the OECD.AI Policy Observatory (OECD, 2024[3]). The database is an online hub that allows users to explore policy initiatives by country, by policy instrument category and type, or by target group, among others.

19. Since its establishment, the OECD.AI Policy Observatory's comprehensive database has been instrumental in following policy developments and taking stock of countries and international organisations' efforts to implement the OECD AI Principles. The number of initiatives and reporting countries have been constantly increasing since 2020. As of January 2024, the database included information on 1 020 initiatives in 70 jurisdictions, out of which 850 initiatives were reported by the 46 Adherents to the Recommendation and the European Union.

20. Two reports on the State of Implementation of the OECD AI Principles were also approved by the DPC respectively in 2021 [DSTI/CDEP(2020)15/REV1] (the "2021 Report") and 2023 [DSTI/CDEP/AIGO(2023)5/REV2] (the "2023 Report"). The 2021 Report (OECD, 2021[4]) took stock of Adherents' endeavors in designing AI policies and in establishing governance mechanisms for national AI initiatives. It also showcased numerous policy examples from Adherents that implement the five recommendations to governments. The 2023 Report (OECD, 2023[5]) also provided updates on Adherents' national AI strategies and included an additional focus on select AI-specific regulatory frameworks being developed around the globe. Moreover, the 2023 Report illustrated different Adherents' efforts to implement the OECD AI Principles.

21. In order to complement this information and ensure a thorough and comprehensive reporting on implementation, dissemination and continued relevance, a questionnaire was circulated to AIGO delegates in December 2023 to inform some aspects of the reporting to Council, i.e., challenges in implementation, continued relevance and dissemination of the OECD AI Principles at national level. The questionnaire also included questions on main opportunities and risks from advanced AI models, drawn from the questionnaire circulated to G7 members to support the G7 Hiroshima Process on generative AI.

3. Implementation and continued relevance

3.1. Definitions

3.1.1. AI system

22. To ensure the continued relevance of the OECD's definition of an AI system, the Recommendation was revised by the Council on 8 November 2023 [C(2023)151 and C/M(2023)14, Item 218] to update this definition, in order to ensure that it continues to be technically accurate and reflects technological developments, including with respect to generative AI. In addition, the DPC approved and declassified an Explanatory Memorandum to accompany the updated definition of an "AI System" on 15 December 2023 (OECD, 2024[2]).

3.1.2. AI system lifecycle

23. The definition of an AI system lifecycle is still accurate and relevant. However, both the OECD – in its work on the OECD Framework for Classification of AI Systems (OECD, 2022[6]) and in work on risks and accountability (OECD, 2023[7]) – and members, such as the National Institute of Standards and Technology (NIST) in its AI Risk Management Framework (National Institute of Standards and Technology - US Department of Commerce, 2023[8]), have been using a simplified, one-level definition (Figure 3.2), instead of two levels (Figure 3.1). It is deemed substantively the same but clearer to describe the different stages of development and deployment of an AI system.

Figure 3.1. AI system lifecycle stages – current definition (two levels)

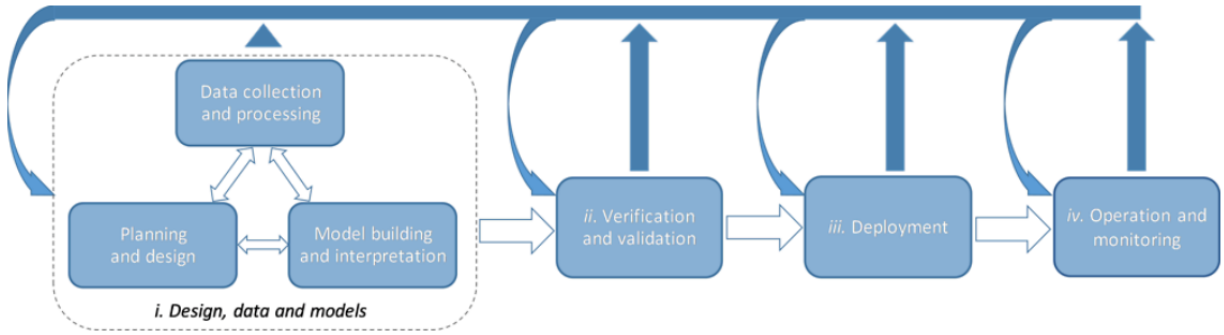
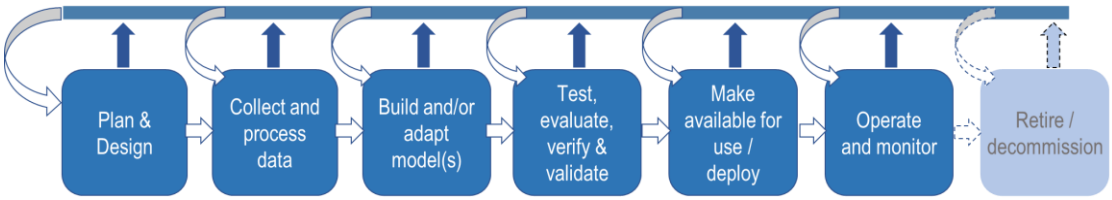


Figure 3.2. AI system lifecycle stages – proposed updated definition (one level)



24. **The definition of an AI system lifecycle would benefit from simplifications to ensure accuracy in the description of how AI systems are developed and operated.**

3.1.3. AI knowledge, AI actors and stakeholders

25. The definitions of AI actors and stakeholders are still fit for purpose and have provided/are providing the basis for other similar definitions at national and regional levels.

26. **The definition of AI knowledge would benefit from an addition to include knowledge on how to manage risks within the set of skills and resources needed to understand and participate in the AI system lifecycle.**

27. The sections below address in turn implementation of each Principle and discuss the continued relevance of each Principle, including whether updates are needed.

3.2. Principles

3.2.1. AI national strategies and governance models

28. In 2017, only a few countries had national AI strategies. By January 2024, the OECD.AI Policy Observatory contained over 850 AI policy initiatives reported by Adherents, including

41 national AI strategies, i.e. a government-wide initiatives aimed at steering trustworthy AI development and deployment in a comprehensive manner.

29. Each country's national AI strategy has its specificities and tackles different aspects of AI policy. Yet, mapping national AI strategies to the recommendations to governments included in the OECD AI Principles shows commonalities, as most strategies focus on inclusive growth, sustainable development and well-being, human-centred values, and fairness. Most strategies also include pillars mapping to all the five policy recommendations, with investing in AI research and development, and building human capacity among the key priorities of policymakers. Several Adherents, including Italy, Lithuania, Korea and Türkiye, reported that their national AI strategies integrate or draw from the OECD AI Principles.

30. Adherents to the Recommendation use different types of governance models to implement national AI policies. Some have created governmental coordination bodies for AI. The United Kingdom founded the Government Office for AI, a unit within the Department for Science, Innovation and Technology, and the United States established the National Artificial Intelligence Initiative Office (NAIIO), located within the White House Office of Science and Technology Policy. Other Adherents leverage existing ministries and have established AI inter-ministerial and multi-stakeholder committees to oversee the development and implementation of AI strategies. Examples of this approach include the creation of the Governance Committee of the Brazilian AI Strategy and Egypt's National Council for AI (NCAI).

31. Adherents are also setting up multi-stakeholder groups of AI experts to advise and report on current and future opportunities, risks and challenges of the use of AI. France's National Consultative Committee on Digital Ethics and AI (FNCDE) and Canada's Advisory Council on AI are pertinent examples.

32. Adherents like Canada have established dedicated AI observatories to monitor and evaluate their AI strategy rollouts. While still rare, monitoring and evaluation mechanisms can be expected to expand across other Adherents as national AI strategies move into later stages of implementation.

3.2.2. Section 1: Principles for responsible stewardship of trustworthy AI

Inclusive growth, sustainable development, and well-being (Principle 1.1)

> *"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."*

33. This Principle calls on stakeholders to steer AI development, deployment and use to achieve societal benefits at large, empowering all members of society, mitigating inequalities and minimising harms to the environment.

34. AI has the potential to increase countries' productivity, lead to economic growth and contribute to sustainable development, including environmental sustainability. Most countries have recognised this and are promoting the development of AI research and development (R&D), infrastructure, capacities, and tools through diverse initiatives. However, AI systems can also perpetuate existing inequalities and have disparate impact on vulnerable and underrepresented populations, such as ethnic minorities, women, children, the elderly and the less educated or low-skilled.

35. Furthermore, the computational resources required for the training and use of AI systems can have significant environmental footprints from energy and water use, greenhouse gas (GHG) emissions and end-of-life considerations (OECD, 2022[9]). The Principle underscores the importance of harmonising AI advancements with environmental sustainability objectives. Addressing environmental degradation and climate change necessitates the integration of sustainability within AI development and deployment. High energy consumption, increased emissions, resources consumption and degradation, as well as unchecked rapid growth highlight the critical need for environmentally responsible AI initiatives. Consequently, the Principle should be understood as endorsing the pursuit of energy-efficient AI architectures and encourage leveraging AI for environmental protection, monitoring, and facilitating the shift towards renewable energy sources. Prioritising the creation and application of AI technologies that bolster and hasten the environmental transition is essential for ensuring AI's positive contribution to global ecological health and sustainable development goals.

36. As of January 2024, the database of national AI policies included over than 300 initiatives related to Principle 1.1 on inclusive growth, sustainable development, and well-being. Most overarching national AI strategies and AI ethics frameworks or guidelines for the implementation of AI refer to these themes. Adherents have notably launched policy initiatives to ensure vulnerable groups in the population are involved in, and benefit from, the development of AI systems, either through targeted initiatives or in policy design. Adherents also fund and promote projects that use AI to address environmental challenges. Other initiatives, for instance in the creative and healthcare sectors, illustrate how AI can enhance people's well-being.

- **Inclusive growth:** initiatives include France's "*IA Booster*" to support small- and medium-sized enterprises' (SMEs) digitalisation through AI solutions, and programmes such as "Women in Data Science" by the Alan Turing Institute in the United Kingdom, which focuses on reducing gender inequalities by analysing participation, examining workplace cultures, and promoting gender-inclusive AI design.

- **Citizen consultations:** Several Adherents proactively involve citizens in AI policy design. For instance, Austria involved over 160 experts and civil society organisations in the development of its National AI Strategy. In Canada, a Public Awareness Working Group engages the public in discussions about AI through virtual workshops. Scotland (United Kingdom) has held AI Co-Creation Public Engagement Workshops and developed design principles for how people should engage in future AI decision-making. Similarly, Chile's Participation Process on AI collects opinions and concerns of citizens and organisations regarding AI use and development. In Mexico, the Artificial Intelligence National Alliance fosters inclusive, open dialogue on AI and its impacts, by engaging multiple stakeholders to democratise AI discussions. The United Kingdom's AI Ecosystem Online Survey of June 2021 carried out by the Alan Turing institute gathered over 400 responses reflecting the perspectives of actors involved in the AI ecosystem.

- **Beneficial outcomes fostered through multi-stakeholder collaboration:** Canada's "Quebec AI Forum" collaborates with various stakeholders to use AI for economic and social development. In Colombia, the "Coordination Bodies for AI Policy Implementation" align AI policies across national and local public entities and provide guidance to public entities, the private sector, academia, and the national government. Germany's "Civic Coding" initiative is a collaborative effort between ministries to strengthen AI competencies and promote the use of AI for the common good. Since February 2022, Korea has established a multi-stakeholder forum to discuss ethical concerns arising from the advancement of AI technologies and to form consensus on

how to build trust in AI. Three expert committees (on ethics, technology, and education) within the forum facilitate consensus building.

- **Sustainable Development:** Several Adherents have launched initiatives to promote the use of AI for environmental sustainability. Germany's "AI Lighthouses for the Environment", for example, supports AI projects to address environmental challenges; Portugal uses AI to combat illegal fishing and improve waste management for environmental protection, and the European Union's "Destination Earth" project uses AI to create a digital twin of the planet Earth to conduct simulations and help prepare for natural disasters and adapt to climate change.

37. **Evidence indicates progress by Adherents in implementing Principle 1.1 on inclusive growth, sustainable development, and well-being. The Principle remains relevant but could benefit from a minor addition to specifically refer to environmental sustainability, of which the importance has grown considerably over the past five years.**

Human-centred values and fairness (Principle 1.2)

> *"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. 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."*

38. Some uses of AI systems have implications for human rights, including risks that human-centred values can be deliberately or accidentally infringed upon. To address these risks, Adherents have issued primarily non-binding guidelines or initiatives targeted at reducing AI biases as well as at values-alignment by promoting human rights and human-centred values. A few Human Rights Impact Assessments (HRIA) and quality seals have also been developed.

- Initiatives to align values by promoting human rights and human-centred values and through mandatory Human Rights Impact Assessments (HRIAs): France, in collaboration with the World Economic Forum, has established a policy framework to address human rights concerns stemming from the use of AI facial recognition. Korea established the National AI Ethics Guidelines in December 2020, laying out comprehensive guidelines to realise human-centered AI for all members of society in all stages of AI development and use. In the United States, the State Department provides practical human rights guidance to U.S. businesses dealing with products or services involving surveillance capabilities. A "Fundamental Rights and Algorithms Impact Assessment" (FRAIA) is now mandatory for the use of algorithms by Dutch public authorities.

- **Protecting privacy**: AI and particularly generative AI raise considerable threats to privacy. Adherents are implementing various policies to protect privacy in AI, including through existing privacy and data protection legislation, through guidance on data protection in AI applications (Korea and Mexico), through regulatory sandboxes to promote the development of privacy-friendly use of AI solutions (OECD, 2023[10]), and by promoting Privacy Enhancing Technologies (PETs) to prevent privacy infringement in developing or operating AI technologies and services (e.g. Estonia, Türkiye, the United Kingdom, United States) (OECD, 2023[11]).

- **Quality labels and certifications to promote human-centred values:** a few Adherents have developed or are developing quality labels and certifications confirming

that an AI tool is ethical and human-centred. Examples to date include Türkiye's Trustworthy AI Trust Stamp.

- **Initiatives to reduce AI bias:** Several Adherents have launched initiatives to identify and rectify biases and ethical concerns in AI systems. These include the "Bigscience Workshop" in France, which analyses biases and ethical problems in language models and proposes metrics and tools for their evaluation and mitigation; an investigation into the causes of discrimination in AI by the Dutch Ministry of the Interior and Kingdom Relations; and the "Review into Bias in Algorithmic Decision-Making" published by the United Kingdom's Centre for Data Ethics and Innovation (CDEI) in 2020. In the United States, the Equal Employment Opportunity Commission has launched an agency-wide "AI and Algorithmic Fairness Initiative" to ensure that the use of AI complies with American anti-discrimination laws. Furthermore, the US Federal Trade Commission (FTC) has designated AI as a key area of focus for enforcement and regulatory action in investigations of unfair, deceptive, anticompetitive, collusive, coercive, predatory, exploitative, and exclusionary acts or practices relating to algorithms and biometrics. Countries are also introducing AI auditing requirements for hiring tools (OECD, 2023[12]).
- **Democratic values**: Adherents and international organisations are proposing legislation to ensure AI systems respect democratic values, e.g. the proposed Canada's Artificial Intelligence and Data Act (AIDA), the European Union's AI Act and the Council of Europe's Framework Convention on AI, Human Rights, Democracy and the Rule of Law include provisions on transparency and accountability (see Principle 2.3).

39. This Principle remains relevant but could benefit from some updates to reflect policy concerns arisen with recent developments of generative AI.

40. The heading of this Principle could better reflect the text, by calling out the need to respect the rule of law, human rights, and democratic values. Concerns over privacy infringement were among the main risks prioritised by the OECD in the context of Generative AI (Lorenz, Perset and Berryhill, 2023[13]). While privacy is already considered in the text of the Principle, its heading could usefully call out privacy explicitly, alongside fairness, to align with policy priorities expressed by OECD and G7 members and the EU in the proposed EU AI Act, as well as priority risks addressed by practical guidance documents (including, for example the United States in the NIST AI Risk Management Framework).

41. The recent development of generative AI is accompanied by significant concern over the potential for misuse in the creation and propagation of synthetically generated content which is increasingly realistic and convincing. Consequences could extend to the spreading of misinformation and disinformation, perpetuation of discrimination, distortion of public discourse and markets, the incitement of violence, and associated threats to democratic processes and human rights. Given the fact that governments recognise the transformative impact of generative AI and are actively working to address these challenges, the Principles could explicitly call out the need address misinformation and disinformation, in line with the policy priorities expressed by Adherents (Figure 3.4) and the G7 for Generative AI (OECD, 2023[14]).

42. As AI technologies become increasingly sophisticated, potential risks may emerge from their use beyond intended purpose, or from intentional or unintentional misuse. Due to their adaptability and learning capabilities, AI systems can be repurposed for tasks or applications beyond their original design, which may lead to unforeseen consequences. Intentional misuse of AI also presents a significant concern, particularly in scenarios where malicious actors seek to exploit AI systems for purposes such as cyber-attacks or mis and disinformation. Furthermore, unintentional misuse of AI can occur due to human error, technical failures, or

unforeseen circumstances. In light of these risks, the Principle could call on AI actors to implement robust mechanisms and safeguards, including the capacity for human agency and oversight throughout the AI lifecycle.

43. **Evidence indicates progress by Adherents in implementing Principle 1.2 on human-centred values and fairness. While the Principle remains relevant, clarifications and additions would be appropriate to reflect important challenges that have arisen over the past five years. These include calling out privacy explicitly in the heading, add in the text the need to address misinformation and disinformation. Finally, the text could call on AI actors to implement mechanisms to address risks arising from use outside of intended purpose, intentional or unintentional misuse of AI systems**.

Transparency and explainability (Principle 1.3)

> *"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: to foster a general understanding of AI systems, to make stakeholders aware of their interactions with AI systems, including in the workplace, to enable those affected by an AI system to understand the outcome, and, 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."*

44. Most national AI strategies, ethical frameworks, and general principles for the implementation of AI list transparency and explainability among the key properties of a trustworthy AI system. Transparency and explainability also figure prominently in several non-binding guidelines for ethical AI implementation. However, despite broad agreement on the importance of transparent and explainable AI, operationalising these concepts is complex, due to their multifaceted nuances. AI transparency entails clearly communicating to users that they are dealing with an AI system, enabling users to interpret system outputs, and in some cases explaining the decision-making logic.

45. Adherents are taking a variety of approaches to ensure AI transparency, ranging from guidelines for implementation of AI, to the establishment of oversight bodies. Transparency provisions are laid down in existing legislation (e.g., data protection and privacy legislation, consumer protection legislation), and are also being included in proposed AI-specific regulations, with several specific provisions pertaining to the workplace. In the public sector, adherents are enhancing transparency around the use of AI for public services through AI registers. Examples of initiatives aimed at promoting transparency and explainability include:

- **Initiatives requiring disclosure and information about use of AI systems:** Japan has introduced transparency requirements in the Digital Platform Transparency Act by requiring designated digital platform providers to ensure transparency and equal treatment in transactions with business users. The EU regulation on Artificial Intelligence (hereafter "EU AI Act") (in the latest stages of the legislative process at the time of writing) includes transparency obligations for high-risk AI systems, as well as for those intended to interact with natural persons; used for emotion recognition; used for biometric categorisation; or used to generate or manipulate image, audio or video content, and for general-purpose AI systems.
    - **In the public sector,** the 2016 French Digital Republic Law mandates transparency of government-used algorithms. Finland and Netherlands have launched open AI registers in their capitals (Helsinki and Amsterdam) that track how algorithms are being used in the municipalities. The UK Algorithmic Transparency Recording Standard (2022) outlines comprehensive guidelines

for the public sector, including government, on disclosing information when using algorithmic tools.

- **In the workplace:** Among others, France, Germany, and Italy have laws that include the requirement to secure the prior agreement of worker representatives on using digital technologies including AI to monitor workers. The Canadian Province of Ontario's Working for Workers Act requires employers to notify employees about electronic monitoring policing. Similarly, several federal states in the US have laws in place requiring employers to notify employees of electronic monitoring.

- **Initiatives to provide information on AI functioning:** The Canadian proposed Artificial Intelligence and Data Act (AIDA) introduces requirements to promote transparency on the use of AI. Spain's Royal Decree-Law 9/2021 (the "Rider Law") makes transparency mandatory for AI systems that make decisions about or influence either working conditions or employment status. The EU AI Act prescribes information requirements which would allow users to interpret system's output and use it appropriately.

- **Initiatives to provide information on factors and decision processes and enable redress seeking from decisions:** Canada's Consumer Privacy Protection Act includes transparency requirements. Moreover, the country's directive on Automated Decision-Making sets a wide range of mandatory requirements to ensure the responsible use of AI by federal institutions. Mexico's legal framework in matters of personal data protection also contains provisions related to automated decisions. The Norwegian Public Administration Act states that public sector decision-making that pertains to specific individuals must provide explanations for the decisions taken in order to ensure accountability and support a complaints/appeals process. The EU's General Data Protection Regulation establishes a "right to explanation" in its Article 22.

46. **Evidence indicates progress by Adherents in implementing Principle 1.3 on transparency and explainability; the Principle remains relevant but could benefit from some clarifications on the information that AI actors should provide in relation to AI systems in general, as well as to those affected by an AI system and to these negatively affected by the outputs of an AI system**.

Robustness, security, and safety (Principle 1.4)

> *"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."*

47. Issues of robustness, security and safety of AI are interlinked. For example, digital security can affect the safety of connected products such as automobiles and home appliances when risks are not appropriately managed. Therefore, they are analysed together. However, there are different, not mutually exclusive ways in which Adherents can operationalise them.

48. Adherents are drawing on guidelines, ethics frameworks, impact assessments, new legislation, amendments to existing legislation and other instruments to implement Principle 1.4. Examples include:

- **Algorithmic Impact Assessments:** Canada's Directive on Automated Decision-Making requires federal institutions planning to use an automated system, including those that rely on AI, to make or support administrative decisions to complete and publish an Algorithmic Impact Assessment (AIA) before the launch of the system. Mexico's Principles and Impact Analysis Guide for the Development and Use of Systems Based on AI in the Federal Public Administration are designed to assess the societal and ethical implications of AI systems developed by the Federal Public Administration.
- **Initiatives to maintain records of data characteristics for traceability:** Efforts are currently underway in Türkiye to introduce a National Data Dictionary, which aims at compiling a national data inventory and establishing management and monitoring processes through national data integration architecture. The US National Security Presidential Memorandum (NSPM) Protecting the United States Advantage in AI and Related Critical Technologies aims at improving access to high-quality and completely traceable Federal data, models, and computing resources.
- **Laws and regulations preventing unreasonable safety risks of AI systems - autonomous driving:** In Austria, Germany, Denmark, Japan, Lithuania, and United Kingdom, new legislation has been passed (or existing legislation has been amended) to define the use of self-driving cars on their respective national roads.

49. Work on accountability has highlighted the need for traceability and for a systematic risk management approach to each phase of the AI lifecycle on a continuous basis. While originally envisaged as part of the Principle on "Robustness, security and safety", it is now clear that traceability and risk management apply across all value-based Principles and all phases of the lifecycle, and therefore should be part of the Principle on "Accountability." These are currently points (b) and (c) under Principle 1.4. (Robustness, security and safety), and would be more appropriate under Principle 1.5 on accountability.

50. As highlighted in the discussion in Principle 1.2, the rise of generative AI is accompanied by concerns over the potential for creation and misuse of synthetically created content, and the propagation of misinformation and disinformation, which has the potential to harm democratic processes. This highlights the need to support efforts to protect information integrity, which is essential to ensuring AI systems do not adversely affect individuals' ability to obtain accurate, evidence-based, and plural information sources. Upholding information integrity is crucial to safeguarding freedom of expression, including the freedom to seek, receive, and impart information and ideas (OECD, 2024[15]).

51. Furthermore, in light of discussions on potential risk of humans losing control over AI systems – a key discussion topic at the AI Safety Summit convened by the United Kingdom in November 2023 (AI Safety Summit, 2023[16]) and at forthcoming Summits in Korea and France –, the Principle could include a reference to the need to have mechanisms in place to ensure that AI systems that risk causing undue harm or exhibit undesired behaviour can be overridden, repaired, and/or decommissioned as needed.

52. **Evidence indicates progress by Adherents in implementing Principle 1.4 on robustness, security, and safety. While the Principle remains relevant, there could be merit in explicitly calling out the need to have mechanisms on place to bolster information integrity, and to retire AI systems that risk causing undue harm or exhibit undesired behaviour. Furthermore, the text currently under Principle 1.4. (Robustness, security and**

safety) on traceability and on systematic risk management approach would be more appropriately placed under Principle 1.5 (Accountability).

Accountability (Principle 1.5)

> *"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."*

53. Accountability refers to the expectation that organisations or individuals will ensure and be held responsible for the proper functioning, throughout their lifecycle, of the AI systems that they design, develop, operate or deploy, in accordance with their roles and applicable regulatory frameworks, and for demonstrating this through their actions and decision-making processes. In the case of a negative outcome, it also implies taking action to ensure a better outcome in the future.

54. Demand for tools and processes to document AI system decisions and ensure accountability is on the rise in both the public and private sectors. This field encompasses major AI standardization initiatives led by organisations like the International Organization for Standardization (ISO), Institute of Electrical and Electronics Engineers (IEEE), International Telecommunication Union (ITU), NIST, European Telecommunications Standards Institute (ETSI), Internet Engineering Task Force (IETF), and European Committee for Electrotechnical Standardization (CEN-CENELEC). These initiatives focus on various aspects including AI design (like trustworthiness by design), impact assessments, conformity evaluations, and risk management frameworks for AI. Additionally, there are governmental and intergovernmental efforts such as the EU's AI Act, the UK's AI Standards Hub, the European AI Alliance, the Council of Europe's Committee on Artificial Intelligence, and the EU-US Trade and Technology Council. Certification schemes are also a part of this landscape (OECD, 2023[7]).

55. Adherents have developed guidelines for the use or implementation of AI in several sectors (public administration, health care, autonomous driving), i.e., initiatives that can be regarded as codes of ethical conduct. Proposed AI-specific regulation requires the documentation of the proper functioning of the AI systems throughout their lifecycle. Lastly, Adherents and the EU have established independent oversight bodies for AI and algorithms. Examples include:

- **Legislation that requires documenting the proper functioning of the AI systems throughout their lifecycle:** In Canada, the proposed AIDA ensures accountability through the proactive documentation of policies, processes, and measures implemented as well as ways to meet requirements for design and development. In the EU, the AI Act makes technical documentation obligatory.
- **Codes of ethical conduct and practical technical tools:** France, Portugal, Norway, and United Kingdom have, among others, issued guidance, including transparency requirements, on the development and use of AI in the public sector. France and Singapore have also established codes of ethical conduct in the healthcare sector. In February 2022, Korea developed a "Self-assessment Checklist" as a practical way to implement the National AI Ethics Guidelines, with more detailed guidelines (sector and case specific examples) provided in April 2023. In Mexico the National Institute for Transparency, Access to Information, and Personal Data Protection (*Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales*, INAI) developed jointly with the Iberoamerican Data Protection Network (RIPD), the "General Recommendations for the Processing of Personal Data in Artificial Intelligence" (Ibero-American Data Protection Network, 2019[17]) and the "Specific Guidelines for Compliance with the Principles and Rights that Governs the Protection

of Personal Data in Artificial Intelligence Projects" (Ibero-American Data Protection Network, 2019[18]), providing guidance to developers and manufacturers of AI on the regulatory requirements on treatment of personal data. The UK Centre for Data Ethics and Innovation (CDEI) developed a portfolio of use cases and an online searchable repository of AI assurance tools (CDEI, 2024[19]). Singapore has launched A.I. Verify, an AI governance testing framework that helps companies with transparency.

- **Independent oversight bodies for the supervision of AI**: Supervisory agencies and oversight bodies have an increasing role in ensuring the responsible use of AI. In December 2022, Spain announced the establishment of the Spanish Agency for the Supervision of AI which should promote responsible, sustainable, and trustworthy AI, as well as collaboration and coordination with other national and supranational authorities for AI oversight. The Agency's statutes were approved in August 2023 and its governing body was designated in December 2023. In the Netherlands, the algorithm supervision unit, situated within the Dutch Data Protection Authority (DPA), aims to enhance the oversight of algorithms across all sectors and to coordinate cooperation and risk detection.

56. As AI permeates sectors across of the economy and society, and in line with ongoing reflections of the OECD Expert Group on AI Risk and Accountability, it would be important to reference in this Principle the need for AI actors to adopt responsible business conduct to address risks related to AI systems. Looking forward, they will need to increasingly co-operate with other AI actors but also with upstream and downstream actors, i.e. suppliers of AI knowledge and resources, AI system users, and other stakeholders. These may include, for example, hardware providers, investors, rights holders, and research institutions.

57. Further revisions could specify that risks of bias that need to be addressed refer specifically to "harmful" bias rather than all bias, e.g. bias that leads to unfairness, discrimination, or reinforces stereotypes and inequalities. The text could explicitly mention risks of infringing intellectual property rights, a policy concern that has increased dramatically with the rise of generative AI, as well as labour rights.

58. **Evidence indicates progress by Adherents in implementing Principle 1.5 on accountability. While the Principle remains relevant, text currently under Principle 1.4. (Robustness, security and safety) on traceability and systematic risk management approach would be more appropriate under Principle 1.5 (Accountability). Furthermore, it would be important to reference the need for AI actors to ensure responsible business conduct to address risks related to AI systems. Further revisions could specify that risks of bias that need to be addressed refer specifically to "harmful" bias rather than all bias, and explicitly mention risks of infringing labour and intellectual property rights.**

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

Investing in AI research and development (Principle 2.1)

> *"a) Governments should consider long-term public investment, and encourage private investment in research and development, including inter-disciplinary 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."*

59. Many Adherents have recognised the importance of policies that support AI R&D and are responding with initiatives to ramp up efforts in this area. To date, Adherents have reported about 290 policy initiatives related to this Principle. Most national AI strategies focus on AI R&D as one of the key areas for action. Adherents have dedicated AI R&D funding programmes and are using different instruments for their implementation. A key trend is the creation of national AI research institutes and centres. Other actions include establishing AI R&D-focused policies, plans, programmes, and funds to support AI diffusion in enterprises and in the public sector and consolidating AI research networks.

- **Public funding to support AI R&D**: Information on publicly funded AI R&D remains imprecise and difficult to compare, although information is available for some countries. For instance, France's National AI Research Programme allocated EUR 445 million to AI research between 2018 and 2022. In the United States, funding of USD 1.8 billion was requested for non-defence AI R&D in 2023 (National Council of Science and Technology, 2022[20]). Most recently, the National Science Foundation (NSF) and the Office of Science and Technology Policy (OSTP) proposed to establish a National AI Research Resource (NAIRR) with estimated funding of USD 2.6 billion over the years 2023-2029. The European Union has allocated EUR 1 billion per year for AI including funding in the Horizon Europe and Digital Europe programmes (European Commission, 2023[21]).
- **National AI research institutes, centres and networks**: Several governments have established national AI research institutes and centres, by tasking specialised institutions or organisations to promote research, development, and innovation in AI. Such centres typically have a mandate to advance AI technologies, foster collaboration between academia, industry, and government, and contribute to the broader AI ecosystem in the country. Examples of such centres are the three Canadian national AI institutes (Amii, the Vector Institute, and Mila); in France, the National Institute for Research in Digital Science and Technology (INRIA), the AI for Science and Science for AI (AISSAI) within the National Centre for Scientific Research (CNRS), LaborIA, and a network of four interdisciplinary institutes for AI (3AI) established by the French national AI strategy; the six German AI excellence centres (including the German Research Center for Artificial Intelligence "DFKI") and the four German AI service centres, the Australian national AI centre, and the Korean Research Data Centre of AI Innovation Hub. Italy established in 2022 "FAIR" (Future Artificial Intelligence Research), a national research network including research institutions, universities, and companies aimed at advancing research and innovation in AI. In the United States, the NSF funds various AI research institutes across the country, supporting interdisciplinary projects that tackle fundamental AI research questions. In addition, the Applied Research Centres in AI, the Applied Innovation Centre and the National Centre for Innovation and AI were established respectively in Brazil, Egypt and Peru.
- **Investments in open datasets:** Evidence from the 2023 OECD Open, Useful, and Re-usable data (OURdata) Index show that "OECD countries have improved the quality of open government data, an important capability considering recent advancements in AI" (OECD, 2023[22]). Several Adherents have initiatives to support open datasets for research and development, while ensuring privacy and data protection. Examples include the Canadian government's Open Government Portal, which provides access to a wide range of datasets from federal departments and agencies, the Etalab platform in France and the Dutch Open Data Portal in the Netherlands. The United Kingdom's Open Data Institute (ODI) promotes the use of open data and provides guidance on privacy and data protection, and Data.gov in the United States provides access to a vast array of datasets across different sectors.

60. **Evidence indicates progress by Adherents in implementing Principle 2.1 on investing in AI research and development; the Principle remains relevant. It but could benefit from some additions to reflect the importance of investing in open science and open source tools; both of which have been critical to AI's development to date. Furthermore, the text could be revised to be consistent with the clarification in Principle 1.5 regarding "harmful" bias**.

Fostering a digital ecosystem for AI (Principle 2.2)

> *"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."*

61. Embracing AI-enabled transformation depends on the availability of data, infrastructure, and software to train and use AI models at scale. Fostering a digital ecosystem for AI hence represents a crucial component of countries' efforts to advance in their AI adoption. Adherents have reported to date 400 policy initiatives related to Principle 2.2 on fostering a digital ecosystem for AI. These include several types of initiatives, ranging from high-level strategic documents, such as national AI strategies, data strategies, technologies roadmaps, to specific actions, such as AI measures in response to COVID-19.

62. Key overall trends are efforts to increase computing capacity and access to infrastructure, open data policies and policies for data sharing, and investments in Natural Language Processing technologies.

- **Initiatives to increase computing capacity and access to infrastructure:** Access to AI technologies and computing capacity is crucial for researchers and industries to develop innovative solutions. Half of the Adherents and the European Union have reported to date policy instruments addressing AI computing and research infrastructure. These include for instance, policies to support high-performance computing (HPC) capacity (e.g. Czech Republic, Finland, France, Portugal, Japan, Slovenia, Türkiye, the United Kingdom), including to improve access for AI researchers and startups (Canada, Israel, United States and the European Union), or to pool resources for computing capacity (European Union), investments to enhance cloud competitiveness (Korea and European Union), and to strengthen domestic AI chips supply (Korea and the United States). Few Adherents to date have conducted mapping of their national compute capacities and needs. Examples are the 2020 Canadian Digital Research Infrastructure Needs Assessment (Digital Research Alliance of Canada, 2020[23]) and the 2023 review of digital research infrastructure needs carried out in the United Kingdom (UK DSIT, 2023[24]).
- **Data policies:** AI requires a large amount of data to recognize patterns, learn, and make accurate predictions or decisions. Linking data policies to AI policies is important because it helps keep data ethical, protected, and safe from privacy issues, while at the same time ensuring that AI can do its job effectively. Whereas evidence from the 2023 OECD Digital Government Index shows that "only 59% of OECD countries have a data strategy or similar instrument in place for the public sector" (OECD, 2024[25]), several Adherents have begun to link their data access and sharing policies with AI policies. Examples of these initiatives include strategies to enhance data access and sharing (Korea, Sweden) providing access to scientific information and data (Czech Republic), including in specific sectors (e.g. health in Norway), creating platforms (Belgium, France, Hungary, Norway) or data infrastructure and platforms (Peru) to

centralise public sector data and share open government data (OECD, 2023[26]), providing accessible data infrastructure to support the development of AI systems (Colombia). Other Adherents such as the United Kingdom, have created data trusts, to enable individuals or organisations to collectively manage, share, and govern data for a specific purpose. Several Adherents and the EU are also establishing data spaces to store, manage, process and make data accessible (e.g. European Union), including with sector-specific focuses (e.g. health in France, and health and mobility in Germany).

- **Investments in NLP technologies:** Several Adherents have launched policy initiatives to promote language models in their national languages. Examples of these initiatives include – but are not limited to – Denmark's "Danish Gigaword Project", Estonia's "Estonian Language Technology 2018-2027", France's "Pour des IA Francophones", Israel's plans to create datasets, models and tools for Hebrew and Arabic NLP, Japan's "Global Communication Plan 2025", Korea's "National Initiative for Language Technologies", Norway's "Norwegian Language Bank", and Spain's "National Plan for the Advancement of Language Technologies".

63. Given the increasing importance of computational power for AI development and adoption, this Principle is increasingly relevant, and Adherents should continue working on ensuring availability of advanced computing infrastructure, and to make compute capacity available to researchers and start-ups. These efforts should be informed by assessments of national computing capacity and needs, focusing on capacity, effectiveness, and resilience (OECD, 2023[27]). On this basis, Adherents could develop national AI compute plans, or integrate plans for domestic AI compute capacity into their national AI strategies.

64. Policies aimed at fostering compute capacity should also take in consideration the environmental and sustainability impacts of compute infrastructure. It is important to note that, in parallel, DPC is reviewing the 2010 OECD Recommendation on Information and Communications Technologies (ICTs) and the Environment [OECD/LEGAL/0380], in order to clarify the broader policy landscape in this regard.

65. **Evidence indicates progress by Adherents in implementing Principle 2.2 on fostering a digital ecosystem for AI. While the Principle remains relevant, the Principle could better reflect in its heading the enabling nature of the digital ecosystem and the need to foster an inclusive digital ecosystem, i.e. ensure access to AI resources for a diverse set of users. The text could be made clearer with explicit references to the elements of the digital ecosystem that should be addressed under this heading, including *inter alia* data, AI technologies, computational and connectivity infrastructure, and knowledge-sharing. Third, the text should also refer to the characteristics that an AI ecosystem should embody, namely inclusivity, dynamism, sustainability, and interoperability. This indicates that governments should foster an ecosystem that promotes access to and contribution by diverse stakeholders, quickly responds to rapidly evolving technology, prioritises long-term societal, environmental, and economic well-being – including by minimising its carbon footprint and resource consumption – and facilitates the seamless integration of AI systems across domains.**

Shaping an enabling policy environment for AI (Principle 2.3)

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.

66. Countries are exploring approaches to ensure trustworthy AI and mitigate risks associated with the development and deployment of AI systems. In addition to exploring the application and need to adapt current legislation for AI, emerging regulatory actions for trustworthy AI include: i) establishing ethical frameworks and principles, ii) considering hard law approaches, iii) supporting international standardisation efforts and international law efforts.

- **Guidelines for trustworthy AI:** Many Adherents have introduced guidelines for trustworthy AI that are largely aligned with the OECD AI Principles and that provide standards for the ethical use of AI and its governance. Depending on the case, they are addressed to policy makers, businesses, research institutions and other AI actors. Examples include Australia's AI Ethics Framework, Belgium's online self-assessment tool to foster trustworthy AI specifically tailored to the public sector, Colombia's Ethical Framework for AI, the Korean Trustworthy AI Development Guidelines, Hungary's AI Ethical Guidelines, Japan's AI R&D Guidelines and Human-centric AI Principles, Switzerland's Guidelines on AI for the Confederation, Argentina's Ethics Principles for the Development of AI, and Egypt's Charter on Responsible AI.

- **Controlled environments for regulatory experimentation**: An increasing number of Adherents use regulatory sandboxes for AI, i.e., spaces in which authorities engage firms to test innovative products and services that challenge existing legal frameworks (OECD, 2023[10]). Examples include Germany's Regulatory Sandbox Strategy (2019) (German Government, 2022[28]), and the Norwegian Data Protection Authority (*Datatilsynet*) Regulatory Sandbox (2020). Promoting regulatory experimentation is also one of Israel's national AI strategy's key tools to ensure safe and innovative AI deployment. Spain, in collaboration with the European Commission, created an AI regulatory sandbox in 2022 as the first pilot programme to test the future proposed EU AI Act with real AI applications, to assess how both the regulation and applications respond, and to suggest modifications or explanatory guidelines (OECD, 2023[10]).

- **Emerging AI-specific regulation:** Existing provisions in different fields of legislations already regulate AI systems. But in recent years, countries have started codifying OECD AI Principles into binding, AI-specific legislative and regulatory frameworks that address AI high-risk systems or impacts, albeit with key differences in approach across countries and jurisdictions.

- Some Adherents and international organisations are taking a cross-sectoral approach to AI regulation building an AI-specific regulatory framework applicable to all sectors. Canada, Brazil, the European Union and the Council of Europe have proposed to regulate AI systems across domains and applications, building an AI-specific regulatory framework applicable to all sectors:

    - Canada has put forward a comprehensive regulatory framework at the federal level, the Digital Charter Implementation Act, which includes the Artificial Intelligence and Data Act (Canadian Parliament, 2022[29]). The proposed AIDA's approach is to ensure the safe and responsible design, development, and deployment of AI systems that respect Canadians' values. It establishes an impact-based approach that focuses on mitigating the risks of harm and bias of "high-impact" AI systems (Fasken, 2022[30]). The companion document to the proposed AIDA (Canadian Government, 2023[31]) acknowledges AIDA's alignment with the OECD AI Principles, the proposed EU AI Act, and the

United States National Institute of Standards and Technology (NIST) Risk Management Framework.

- Brazil proposed Bill nº 2338/2023 (the "Brazilian AI Act"), of which key themes are: a human rights-oriented approach; a risk-based-approach and risk classification of AI systems; the establishment of a supervisory authority; rules for civil liability; the fostering of innovation by promoting regulatory sandboxes, among others (OECD.AI, 2023[32]).
- The European Union is in the last stages of the legislative process to adopt the EU AI Act. The AI Act follows a risk-based approach and presents a uniform, horizontal legal framework for AI. The EU AI Act leverages the OECD definition of an AI system as updated by the OECD Council in November 2023. It introduces a classification of AI systems based on the levels of risk they represent for health, safety, and fundamental rights, including democracy, rule of law and environmental protection. "High risk" AI systems are subject to conformity assessment procedures before being placed on the market, and to post-market monitoring provisions. AI systems posing "unacceptable" risks are banned.
- The Council of Europe's Committee on Artificial Intelligence (CAI) finalised a "Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law" (Council of Europe, 2024[33]) in mid-March 2024. The Convention will be sent to the Committee of Ministers for adoption and opened for signature at a later stage. The Convention is expected to become an international legal instrument binding on its Parties once adopted. The Convention uses the OECD revised definition of an AI system as updated by the OECD Council in November 2023 and the proposed revised definition of an AI system lifecycle.

- Other Adherents consider more sectoral approaches, developing regulations by sector or domain.
  - Drawing on the OECD AI Principles, Israel's "Draft Policy White Paper for Regulation and Ethics in the Field of AI" adopts non-binding AI ethical principles to be considered when developing, using, and regulating AI, and calls for sector-based regulatory efforts rooted in risk assessment and management approaches, rather than overarching sector-crossing regulation (Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal counsel and legislative affairs, 2022[34]).
  - The United Kingdom's "AI regulation: A pro-innovation approach" (UK Government, 2023[35]) establishes cross-sectoral, non-binding principles – based on the OECD AI Principles -, leaving regulators the task of implementing, regulating and enforcing them in their respective sectors and domains. Israel follows a similar approach.
  - The United States' Blueprint for an AI Bill of Rights (US Government, 2022[36]) aims to support the development of policies and practices that protect civil rights and promote democratic values in the development, deployment, and governance of AI systems (US Government, 2022[36]). It establishes five non-binding principles to mitigate risks to civil rights and democratic values posed by the use of automated systems across sectors. In 2023, the White House released an Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, directing establishment of standards for AI safety and security, including various requirements for AI

developers, government entities, and other relevant actors, to protect privacy, consumer, workers, and civil rights, among others (The White House, 2023[37]; The White House, 2023[38]). The AI Risk Management Framework (AI RMF), issued by the National Institute of Standards and Technology (NIST) (National Institute of Standards and Technology - US Department of Commerce, 2023[8]), takes a rights-preserving approach to technical guidelines and standards for trustworthiness characteristics and responsible practice and use, and builds on the OECD AI Principles and the OECD Framework for the Classification of AI Systems (OECD, 2022[6]) to propose a framework to map, measure, and manage AI risks.

67. The Recommendation specifically underlines 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". This need for ensuring the continued relevance of regulatory and governance approaches and adjusting them as appropriate is also a key tenet of the Recommendation for Agile Regulatory Governance to Harness Innovation (OECD, 2021[39]) and the Best Practice Principles on reviewing the Stock of Regulation (OECD, 2020[40]).

68. Intellectual property rights (IPR) issues have emerged especially with generative AI, particularly concerning unlicensed content in training data, potential copyright, patent, and trademark infringement by AI creations, and IPR ownership of AI-generated works. Adherents are undergoing assessments and developing approaches regarding IPR and AI. In August 2023, for instance, the United States Copyright Office issued a notice of inquiry in the Federal register on copyright and AI (US Copyright Office, 2023[41]), seeking factual information and views on copyright issues raised by advances in generative AI. The United States' Executive Order on AI (The White House, 2023[38]) acknowledges the significant issues posed by AI to intellectual property protection. It instructs the U.S. Patent and Trademark Offices (USPTO) to issue guidance on inventorship and other key issues at the intersection of AI and intellectual property and to review both the U.S. Copyright Office's forthcoming AI report and recommendations by USPTO on copyright and AI for executive action. The EU AI Act requires providers of General-Purpose AI Models to draw up and make publicly available a sufficiently detailed summary of the content used for training the model.

69. **Evidence indicates progress by Adherents in implementing Principle 2.3 on shaping an enabling policy environment for AI; the Principle remains relevant. Notably, emerging AI-specific regulation is aligned or explicitly refers to the OECD AI Principles, or to OECD work based on the Principles (the OECD Framework for Classification of AI system,** (OECD, 2022[6])**). The Principle could possibly benefit from clarification in its heading as "an enabling policy environment" is very broad and as such could refer to many non-AI areas of policy. Furthermore, the heading could explicitly reference the governance of AI in the heading, alongside the importance of ensuring that different governance frameworks are "interoperable", i.e. compatible and able to address common policy and practice challenges. This interoperability of governance frameworks differs from technological interoperability and is at the core the OECD's mandate. As such it can provide helpful clarification. Furthermore, while regulatory experimentation is one of the options for agile regulation to enable innovation, further approaches have also been identified in OECD work** (OECD, 2021[39])**, and could be referenced in the text**.

Building human capacity and preparing for labour market transformation (Principle 2.4)

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.

70. AI is already changing the nature of many aspects of life as it diffuses across sectors, particularly in the context of labour, employment, and the workplace. While to date AI has mainly impacted the quality of jobs rather than their quantity (OECD, 2023[12]), there are signals that labour markets could soon face a significant shakeup with both positive and negative effects. AI can benefit jobs by creating demand for new tasks and complementary skills, resulting in the creation of new jobs for which human labour has a comparative advantage. Recent research shows that generative AI can improve the performance of less skilled workers (OECD, 2023[12]). At the same time, advances in generative AI have heightened focus on the potential impact of AI on labour markets.

71. Recent developments in generative AI, including GPT-4, have broadened the scope of tasks AI systems can perform, thereby increasing their potential influence on the job market. While lower-skilled occupations have until now been most exposed to automation (Lassébie, J. and Quintini, 2022[42]), recent advances in generative AI have led to questioning traditional assumptions about the types of occupations that might be impacted by automation. Research on language-based generative AI finds that 32.8 percent of jobs in the International Standard Classification of Occupations (ISCO) could be impacted on a full scale, 36.5 percent could be partially impacted, and that only 30.7 percent would not be affected by generative AI models (Zarifhonarvar, 2023[43]). This puts pressure on organisations to adapt to generative AI and support their workforces, and on policymakers to steer labour market developments and transitions. However, research on the labour-market effects of generative AI is relatively recent and further peer-reviewed research is needed for more definitive conclusions to be drawn.

72. Countries recognise that both managing a fair transition of the labour market and leading in research, development, and adoption of AI requires policies for AI skills development in tandem with talent attraction. Most national AI strategies include a pillar on AI education and skills development, and to date, Adherents have reported over 200 policy initiatives related to Principle 2.4. Adherents have mainly put in place initiatives to prepare the work force with the skills required for AI through formal education programmes and lifelong learning initiatives. Many of these programmes focus on developing talent among those who will develop AI systems, whether fundamental AI systems or those aimed at use in particular sectors/domains. They also launched initiatives to attract and retain AI talent.

73. While most OECD countries have launched training initiatives for digital skills, countries will need to put in place training programmes that specifically prepare workers to work with AI and that adequately target low skilled workers. Ensuring that workers in many industries and

sectors have the skills to apply AI in their specific sectors appears to be critical, as is developing talent among those who will develop AI systems.

74. Initiatives to monitor the impact of AI in the labour market and to accompany transitions in the labour market appear limited to date. As examined in the programme on *AI in Work, Innovation, Productivity and Skills* (AI-WIPS) - an OECD programme supported by Germany which analyses the impact of AI on the labour market, skills and social policy (OECD.AI, 2024[44]) -, there is an important need to continue to develop initiatives to monitor the impact of AI on labour markets and to identify evidence-based policy responses.

75. Many OECD countries have legislation that needs to be respected when AI systems are used in the workplace, including anti-discrimination legislation, occupational safety and health regulation, worker privacy regulation, and freedom of association (OECD, 2023[45]). Emerging AI-specific legislation, notably the EU AI Act, has important implications for AI in the workplace. The EU AI Act classifies certain AI systems used for recruitment and decisions in work-related contractual relationships as "high risk", making such systems subject to legal requirements relating to risk management, data quality and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy, and security. In February 2024, Spain and the United States signed a joint statement on "Algorithmic Bias in the World of Work" calling on the international community to manage appropriately the harmful risks to workers' social and labour rights posed by certain automated systems (Department of Labour, 2024[46]).

- **Formal education programmes for STEM, AI, and AI-related fields:** Several Adherents and the EU are supporting AI talent development through targeted funding for students in higher education programmes. Examples include Australia's Next Generation AI Graduates Programme, an industry-co-funded PhD scholarship programme, Israel's government funded scholarships to support Master's, PhD and Post-doc students in AI, Italy's National PhD Programme in AI, and the United Kingdom's 2 500 Master's conversion courses for applicants from near- and non-STEM backgrounds, which also encourage greater diversity in AI careers. The European Union's Digital Europe Programme also funds actions to boost advanced digital skills in Europe, including in AI. Korea developed three types of AI ethics textbooks for students in elementary and secondary schools and three types of teacher manuals for teaching AI ethics.
- **Training and lifelong learning AI and related programmes:** Examples of initiatives aimed at providing training to professionals aligned to labour market needs include Chile's Digital Talent programme, Korea's Comprehensive Strategy for Digital Workforce Development, Japan's Practical Guidebook on Data Provision for Fostering Human Resources of Experts in AI and Data Science, Sweden's AI Competence for Sweden, a collaboration among ten Swedish universities to develop courses for professionals who can contribute to Sweden's development in the area of AI, and Singapore's Chartered AI Engineer designation, a professional qualification programme by the AI Professionals Association to recognise and award credentials to working professionals in AI-related engineering roles.
- **Initiatives to retain and attract AI talent:** Canada was one of the early adopters of a skills focused strategy, supporting the attraction and retention of leading academic talent in its Pan-Canadian AI Strategy. Likewise, attracting and training new professorships in AI is a key initiative of the German National AI Strategy.
- **Initiatives to foster collaboration with stakeholders to ensure a fair transition for workers.** Countries are developing measures to foster or strengthen collaboration with stakeholders, including social partners through social dialogue. Spain's Charter of

Digital Rights specifies that "workers' representatives shall be informed of the technological changes taking place in the company and shall participate in decision-making regarding the digital transformation and the consequences it may have for work." The United States' Executive Order (The White House, 2023[38]) specifies that "as AI creates new jobs and industries, all workers need a seat at the table, including through collective bargaining, to ensure that they benefit from these opportunities" and that the "next steps in AI's development should be built on the views of workers, labour unions, educators, and employers to support responsible uses of AI that improve workers' lives, positively augment human work, and help all people safely enjoy the gains and opportunities from technological innovation."

- **Monitoring the impact of AI on the labour market:** The United States' Office of the White House established the American Workforce Policy Advisory Board as part of a national initiative to help bridge the skills gap that is widening due in part to the rise of automation and the increasing need for high-tech skills. United States' Executive Order (The White House, 2023[38]) calls for action to protect the rights and safety of workers, emphasising the need to adapt job training and education to support a diverse workforce and help provide access to opportunities that AI creates. Singapore's Guide to Job Redesign in the Age of AI is a document that helps organisations and employees understand how existing job roles can be redesigned to harness the potential of AI and increase the value of their work.

76. **Evidence indicates progress by Adherents in implementing Principle 2.4 on building human capacity and preparing for labour market transformation. However, more needs to be done in monitoring the impact of AI on the labour market, looking at the quantity, quality, and inclusiveness of jobs, and in implementing labour market specific regulation to promote trustworthy use of AI in the workplace. The Principle remains relevant and could benefit from two additions. These refer to social protection as one of the means to support those affected by displacement, and to the potential for AI to improve the quality of public services.**

International co-operation for trustworthy AI (Principle 2.5)

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.

77. Adherents are increasingly engaged in international co-operation to promote the beneficial use of AI and address its challenges. This is happening through several types of initiatives, which include:

- **International AI research collaboration:** The Global Partnership on AI (GPAI) is an international and multi-stakeholder initiative jointly founded by Canada and France.

Launched in June 2020, it undertakes cutting-edge research and pilot projects on AI priorities to advance the responsible development and use of AI. Examples of cross-border research collaboration at regional level include the Declaration on AI in the Nordic-Baltic Region, the Japan-Singapore Economic Partnership Agreement (JSEPA) framework, and the Quadrilateral Security Dialogue (QUAD), and informal strategic forum involving the United States, India, Australia, and Japan.

- **International and multi-stakeholder co-operation on AI:** In May 2023, under Japan's G7 Presidency, G7 leaders established the G7 "Hiroshima Process on Generative AI" to examine opportunities and challenges related to generative AI (OECD, 2023[47]). In December 2023, G7 leaders endorsed the "Hiroshima AI Process Comprehensive Policy Framework", that includes guiding principles (MIC, 2023[48]) and a code of conduct (MIC, 2023[49]) aimed at promoting the safe, secure and trustworthy development of advanced AI systems. In November 2023, the United Kingdom hosted the AI Safety Summit culminating in several outcomes including the establishment of the United Kingdom AI Safety Institute to conduct advanced AI safety research, possibly in collaboration with similar institutes emerging in other countries, and the commissioning of an AI "State of the Science" report (AI Safety Summit, 2023[16]). The United Nations (UN) Technology Envoy created an "AI Advisory Body" that delivered its interim report in December 2023 (UN Advisory Board on AI, 2023[50]), and is tasked with delivering a final report with its recommendations at the "Summit of the Future", a major event planned for September 2024. In Korea, the AI Ethics Policy Forum involves a multi-stakeholder approach to disseminate the AI Ethics System. The United States and the European Commission have together created a Trade and Technology Council to support stronger transatlantic relations, including on AI. The private sector is also coming together with other stakeholders. The Partnership on AI is a collaboration between major tech companies like Amazon, Google, and Microsoft, as well as civil society and non-profit organisations including the American Civil Liberties Union (ACLU) and the Electronic Frontier Foundation (EFF), academic institutions, and media organisations.
- **Trade agreements that include AI:** In 2020, Chile, New Zealand and Singapore signed the Digital Economy Partnership Agreement (DEPA), whose aims include promoting the safe and responsible use of AI technologies. In 2023, Korea agreed to join DEPA from 2030. Australia and Singapore, building on their pre-existing trade agreement, signed the Singapore-Australia Digital Economy Agreement (SADEA) also in 2023, where parties agreed to advance co-operation on AI. Furthermore, the Korea-Singapore Digital Partnership Agreement (KSDPA), which includes reference to AI and a Memorandum of Understanding (MoU) to enhance practical collaboration and promote the responsible development and use of AI, entered into force in January 2023.
- **Co-operation for AI capacity building in developing countries:** International co-operation is needed to avoid the emergence of an "AI divide" and ensure that the benefits of AI are distributed more evenly across different regions and communities. Germany's "AI for All – FAIR Forward" (2019-2023) is a policy initiative launched by the Federal Ministry for Economic Co-operation and Development focused on the open and sustainable development and application of AI. The Asian and Pacific Training Centre for Information and Communication Technology for Development (APCICT) provides developing countries in the Asia-Pacific region with ICT training, knowledge sharing and policy consultations. The United Nations Development Programme (UNDP) has developed iVerify, an open-source, automated fact-checking tool that is used to help identify false information and prevent and mitigate its spread in Zambia, Kenya, and Honduras, with plans for deployment in Liberia. The "Harnessing AI for

Development" initiative, prepared with the support of the Digital Development Partnership, is an ongoing work within the World Bank's Digital Development Global Practice.

78. **Evidence indicates progress by Adherents in implementing Principle 2.5 on international co-operation for trustworthy AI; this Principle is growing in importance, as AI policy initiatives worldwide multiply. Many challenges and opportunities that AI presents are global in nature, as demonstrated by the numerous initiatives launched in 2023 to foster dialogue on AI at the international level. Geopolitical considerations play an increasingly notable role, alongside the growing societal role of AI, and further highlight the importance of this Principle. The Principle could benefit from a minor revision in the text to substitute metrics with indicators as a clearer term.**

4. Results from the questionnaire to Adherents

79. Most Adherents reported positive outcomes from adhering to the Recommendation, with a few neutral responses. Adherents reported that the AI Principles have provided a solid foundation for shaping national AI policies, provided guidance for policy development, promoted international alignment in AI governance, and facilitated stakeholder collaboration.

- **International Alignment:** Several Adherents emphasised the advantage of working towards international alignment in AI governance through the Principles. They believed they serve as a universally recognised foundation for guiding AI policies, fostering interoperability, and promoting global cooperation in the AI domain.
- **Policy Guidance**: Adherents noted that the Principles provide essential guidance for policy development. They help identify high-level policy considerations, allocate resources efficiently, and steer the creation of new initiatives. Several Adherents reported that they have served as solid foundation for shaping their national AI strategies.
- **Cross-Sector Collaboration**: Adherents noted that the Principles facilitate cross-sector collaboration at the national level as they emphasise the need to engage academia, industry, and civil society. They noted that this multi-stakeholder approach enriches AI strategies, at is considers diverse perspectives.

80. However, some Adherents reported facing challenges in implementing the OECD AI Principles, with the main challenge being the difficulty in translating the values-based Principles into practical, actionable measures, given their high-level and wide-ranging nature. Limited knowledge and technical capabilities in understanding AI, as well as resource limitations affecting policy development and investment were also perceived as obstacles for national implementation of the Principles.

81. Adherents suggested the following additional activities that the OECD could undertake to assist Adherents in implementing the Principles:

- **Guidelines and practical guidance:** Some Adherents suggested providing detailed guidelines and practical explanation to facilitate the implementation of high-level Principles, such as the OECD's work on classifying AI systems or developing a Catalogue of Tool and Metrics for Trustworthy AI. Additionally, several Adherents emphasised the importance of translating the Principles into technical standards and certification processes.
- **Knowledge sharing and capacity building:** Several Adherents underscored the importance of knowledge sharing and capacity building to support policymakers and

stakeholders in adhering countries. Suggestions included creating training programmes and practical workshops for policymakers. Adherents recognised the importance of the OECD AI Policy Observatory as a platform for information and knowledge sharing, and encouraged further showcasing best practices through case studies, benchmarking, and examples of successful OECD AI Principles implementation.

- **Monitoring and reporting mechanisms**: Some Adherents recommended establishing mechanisms for monitoring and reporting on the progress of the OECD AI Principles implementation, including benchmarking and case studies.
- **Technical assistance and advisory services:** A further suggestion was to provide technical assistance and advisory services to countries, such as assistance in drafting AI policies, reviewing existing policies, and offering expert advice.
- **Collaboration and partnerships**: Adherents emphasised the importance of fostering international collaboration, public-private partnerships, and collaborations between governments, academia, and civil society to pool resources and expertise. They also noted the need for mechanisms to recognise assurance efforts and compliance and promote industry engagement.

82. Adherents agreed that the OECD AI Principles remain relevant in guiding AI development, fostering innovation, and building public trust. Adherents considered the OECD AI Principles as foundational for promoting human-centric values, transparency, accountability, and fairness, preventing biases, promoting inclusivity, and ensuring that AI respects human rights and democratic values. Additionally, they were reported to provide an ethical framework that supports inclusive growth, international collaboration, and the establishment of common standards in the global AI landscape.

83. Some Adherents highlighted the need for complementary national or regional legislation and collaboration between international initiatives to address potential AI risks effectively. Adherents also suggested tailoring the OECD AI Principles to different types of AI and creating specific guidelines for various stakeholders, such as users, developers, and sectors. Other Adherents provided recommendations to enhance the relevance of the OECD AI Principles, such as assessing the OECD AI Principles' applicability to (advanced) frontier models and foundation models, and the importance of regular reviews to maintain the adaptability of the Principles in the rapidly evolving AI landscape.

84. Adherents were also asked about key opportunities and risks associated with advanced AI systems, such as foundational models and generative AI, and on the role of OECD AI Principles in leveraging benefits and mitigating risks.

85. Adherents considered productivity gains as the greatest opportunity brought about by foundation models and generative AI, followed closely by improving healthcare and promoting innovation and entrepreneurship (Figure 4.1). They also pointed to the opportunity for such models to contribute to fighting cybersecurity and crime. Most Adherents agreed that the AI Recommendation can help leverage the opportunities presented by advanced AI systems, with some delegations taking a neutral stance. However, they shared a common view that supplementing the OECD AI Principles with more specific and actionable recommendations, especially in emerging areas like generative AI, could further enhance their effectiveness in leveraging opportunities presented by AI.

86. Threats to cybersecurity and disinformation were viewed by most Adherents as the main risks posed by advanced AI systems (Figure 4.2). Most Adherents also considered exacerbating bias and discrimination and risks to safety as major risks. Three Adherents indicated additional risks in their responses: threats to security, disruptions to social cohesion and polarisation, as well as threats to human rights and democratic processes.

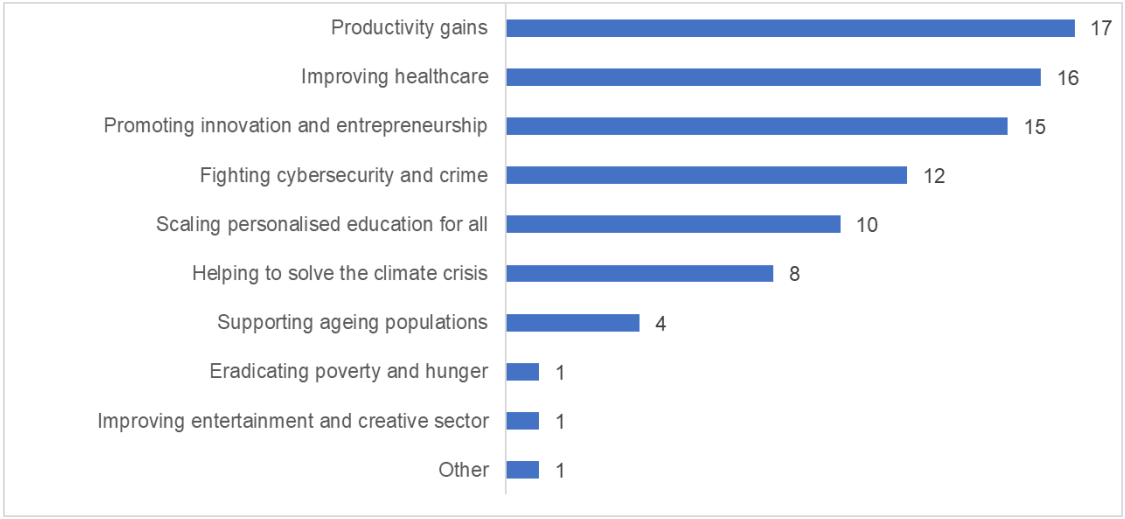
87. While there is a generally positive perception of the Recommendation's ability to address AI-related risks, Adherents' responses varied in this regard. Several Adherents emphasised that while the OECD AI Principles are valuable for addressing risks associated with AI, they may not be sufficient on their own, and that additional measures, such as regulation, are needed to effectively mitigate these risks. One Adherent highlighted that while the OECD AI Principles are helpful in outlining objectives, more work may be needed to assess their effectiveness, particularly in tackling risks posed by frontier AI systems, and stressed the need for additional assurance techniques and standards to ensure the successful implementation of these Principles. Another Adherent also highlighted the importance of detailed guidelines for addressing risks related to privacy, security, fairness, diversity, and non-discrimination. Finally, the importance of addressing risks throughout the lifecycle of AI systems and the need for a comprehensive approach to ensure that risks and impacts are addressed adequately were underscored.

88. The 'governance' of advanced AI systems was widely viewed as the most "important'" and "urgent" priority for policy. In addition to 'governance', the 'responsible use' of generative AI technologies was also viewed as the most "important" priority for policy, followed by 'safety of people', 'alignment of AI systems objectives and human values' and 'data governance'. Promoting accountability and privacy were also seen as urgent issues to address.

89. Overall, Adherents' responses suggest that the Recommendation is seen as a valuable tool to address important advanced AI systems-related priorities, but there were variations in how different Adherents perceived its effectiveness in this regard. While they stressed the importance of the OECD Recommendation as guidance for addressing critical AI-related challenges, Adherents also highlighted the need for practical governance frameworks and Principles tailored to emerging issues, such as disinformation. They emphasised the high-level nature of the OECD AI Principles and the need for more detailed guidance in crafting regulatory frameworks and non-regulatory options tailored to specific cases.

Figure 4.1. Adherents' views on top five opportunities offered by advanced AI systems

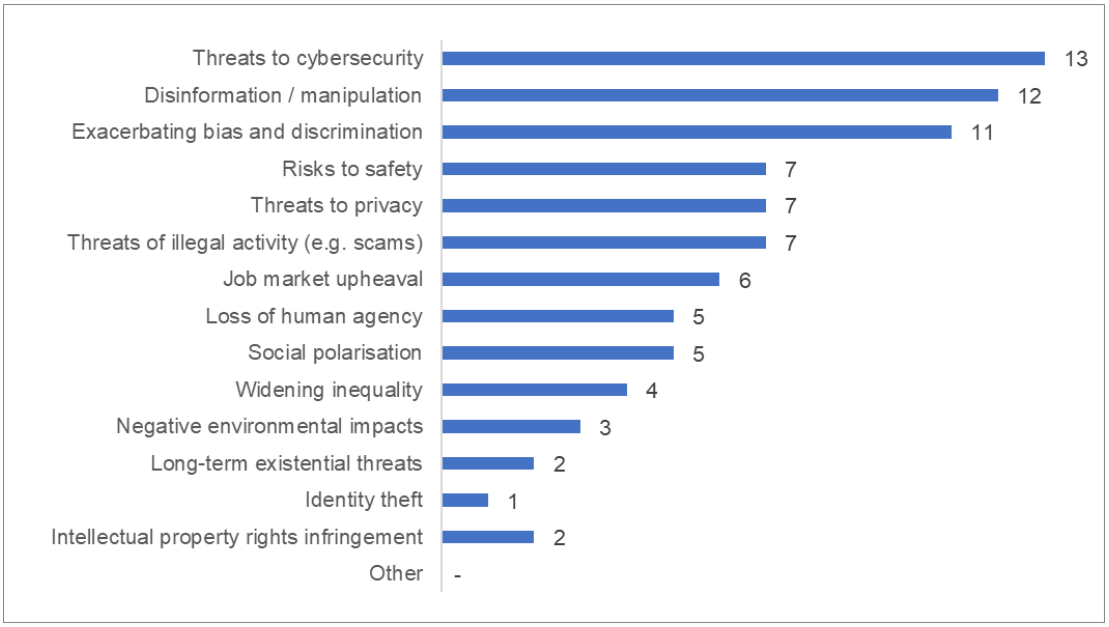
Number of Adherents that selected (five) specific opportunities from a pre-populated drop-down list



Note: The figure aggregates responses from seventeen respondents to the question: "*From your country or region's perspective, what are the top five opportunities generative AI presents to help achieve national and regional goals*? (Please select five options)".

Figure 4.2. Adherents' views on top risks associated to advanced AI systems

Number of Adherents that selected (five) specific risks from a pre-populated drop-down list



Note: The figure aggregates responses from seven respondents to the question: "*From your country or region's perspective, what are the top five risks generative AI presents to achieving national and regional goals?* (Please select five options)".

5. Dissemination

5.1. Adherent activities

90. Several Adherents reported concrete steps undertaken to disseminate the OECD AI Principles at national level and to integrate them into the formulation and implementation of AI strategies and governance frameworks. These initiatives span from the dissemination among academic, technical and political communities to their explicit incorporation into national AI strategies and ethical standards. However, some Adherents did not explicitly mention examples of dissemination, suggesting that more efforts may be needed to promote awareness and understanding of the OECD AI Principles at various levels.

91. Several Adherents leveraged the OECD AI Principles as foundational pillars in their national AI strategies and governance frameworks. These include Italy, whose national AI strategy, and Strategic Programme for AI (2022-2024) implement the OECD AI Principles; and Korea, which incorporated the OECD AI Principles into its national AI strategy, AI ethical standards, and implementation strategy for trustworthy AI. Similarly, Türkiye integrated the OECD AI Principles into its national AI strategy and engaged in discussions at technical and high-level meetings. Similarly, Lithuania included some of the OECD AI Principles in its AI strategy, and Japan referenced them in major government initiatives, including to support discussions at the “Council for Social Principles of Human-Centric AI” and at the "AI Strategy Council, and for developing a "Tentative summary of AI issues" and "(Draft) AI Guidelines for Business." Denmark and Ireland incorporated the OECD AI Principles within their public sector guidelines, emphasising their importance for responsible AI adoption in government. Mexico used the OECD AI Principles as reference for issuing Recommendations for processing of personal data derived from AI use. Finally, the United Kingdom’s White Paper on AI Governance aligns directly with the OECD AI Principles.

5.2. Secretariat activities to disseminate and support the implementation of the Recommendation

92. The OECD Secretariat collaborates with other intergovernmental and regional organisations through the globalpolicy.AI coalition - established by the OECD and seven other intergovernmental organisations - that includes an online platform, and regularly convenes and participates in international discussions on AI policy. Through the collaboration, partner IGOs help policy makers navigate through the different international initiatives, keep each other up to date on their respective AI policy activities and try to ensure interoperability between their work, and work together to advance trustworthy AI where their mandates intersect.

93. The OECD supports implementation of the Recommendation in a number of ways, including by building a robust evidence base, engaging world-class experts, and advancing policy approaches, frameworks and tools. This work is implemented by the OECD Secretariat as part of the programme of work of the Working Party on AI Governance (WPAIGO) under the DPC.

- **The OECD AI Policy Observatory**: To support the practical implementation of the OECD AI Principles, provide an inclusive forum for exchanging information on AI policy and activities, and to foster multi-stakeholder and interdisciplinary dialogue, the OECD launched the AI Policy Observatory (OECD.AI) in February 2020. It brings together resources from across the OECD, its partners and stakeholder groups, in areas such as competition, innovation, trade, health and skills, with dedicated pages for related analytical work, news and data visualisations. Demonstrating both the growing interest in AI and the OECD’s global leadership, traffic to OECD.AI has increased by

350% in the past year, with more than 50 thousand monthly users from 100 countries mid-2023.

- **Novel approaches to measure AI development and adoption**: The OECD is also improving approaches for measuring AI development and adoption to provide governments with up-to-date evidence needed for AI policymaking. The OECD continually produces a range of AI metrics as references and research tools alongside live data showing the latest news in AI and main trends in areas such as: demographics and divides, research, investment, jobs and skills, patents, software development, education and more. Other ongoing work based on different sources of microdata is further exploring the links between AI use by firms and productivity and considering the role of complementary assets such as human capital.
- **The OECD AI Index:** Based on the metrics already available on OECD.AI (trends and data, and database of national AI policies), the OECD is developing a comprehensive measurement framework for AI.
- **The AI Incidents Monitor**: To understand actual incidents that result from the use of AI, the OECD is beginning to monitor AI incidents in real time through a newly developed global AI Incidents Monitor (AIM).

94. The OECD has in the past years provided practical guidance on the implementation of the OECD AI Principles, through its analytical work in several fields, and in collaboration with other parts of the OECD. Figure 5.1 summarises OECD support to implement the value-based OECD AI Principles, namely to contribute to managing AI risks effectively. Table 5.1 provides an overview of activities aimed at helping Adherents implementing the five policy recommendations.

Figure 5.1. OECD work to support implementation of trustworthy, values-based AI

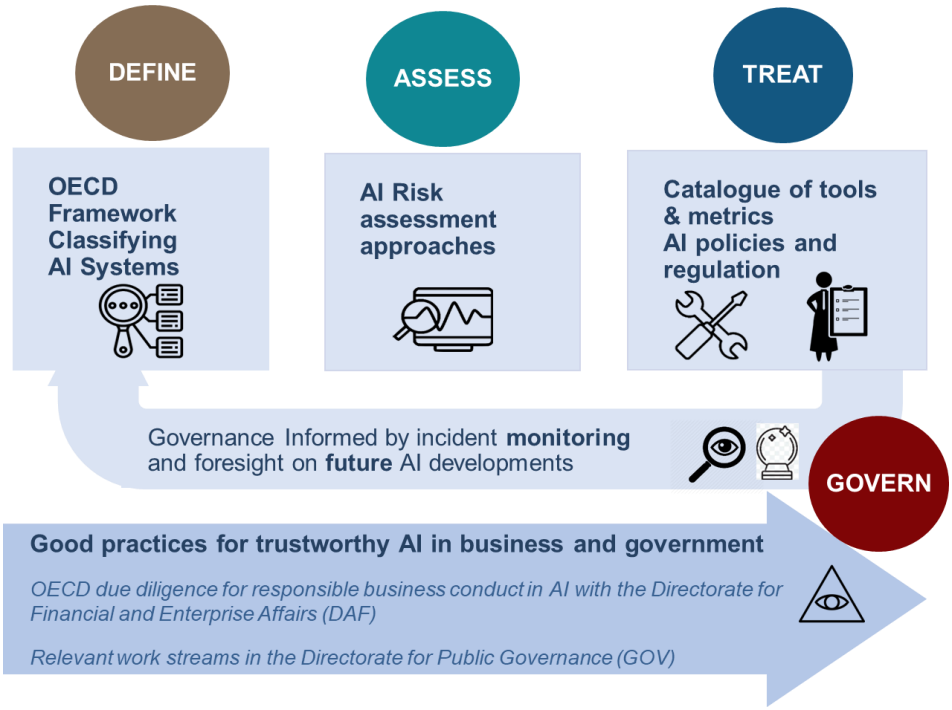


Table 5.1. OECD guidance to support Adherents in implementing the five recommendations to governments

Five recommendations to governments	Select projects
Investing in AI research and development	Joint work on AI futures (with the Committee on Scientific and Technological Policy (CSTP) and the foresight unit of the Secretary General) Events and analysis on smart energy systems (with the Working Party on Communications Infrastructure and Services (CISP) and the International Energy Agency (IEA))
Fostering a digital ecosystem for AI	Measuring the environmental impacts of artificial intelligence compute and applications (OECD, 2022[51]) A blueprint for building national compute capacity for artificial intelligence report (OECD, 2023[27]) AI language models: Technological, socio-economic and policy considerations (OECD, 2023[52]) Initial policy considerations for generative artificial intelligence (Lorenz, Perset and Berryhill, 2023[13]).
Shaping an enabling policy environment for AI	AI regulatory sandboxes report (OECD, 2023[5]) State of implementation reports (OECD, 2021[4]), (OECD, 2023[5]) National AI country reviews (Germany and Egypt, forthcoming)
Building human capacity and preparing for labour market transformation	Programme on AI & work, innovation, productivity and skills – OECD.AI/WIPS (with the OECD Directorate for Employment, Labour and Social Affairs (ELS) and the OECD Education Directorate (EDU)
International co-operation for trustworthy AI	Globalpolicy.ai (with 8 IGOs incl. EC, CoE, UNESCO, IDB)

5.2.1. The OECD Network of Experts on AI

95. The OECD Network of Experts on AI (ONE AI) is an informal multi-stakeholder and multi-disciplinary expert group that supports the AIGO. It comprises over 400 experts from all stakeholder groups organised into six informal expert working groups, which meet monthly or bimonthly to focus on providing input and advice AIGO on key policy priorities including:

- **Risk & Accountability:** The OECD is engaging with partner organisations, policy makers and experts, to identify common guideposts to assess AI risk and impact for Trustworthy AI. The goal is to help implement effective and accountable trustworthy AI systems by promoting global consistency.
- **AI Incidents**: While AI provides tremendous benefits, it also poses risks. Some of these risks are already materialising into harms to people and society like bias and discrimination, the polarisation of opinions, privacy infringements, and security and safety issues. These harms are broadly referred to under the developing term of an "AI incident". Monitoring AI incidents requires global consistency and interoperability in incident reporting, so that AI system operators and policy makers can learn from the risks and incidents of other actors internationally. The OECD expert group is working on a reporting framework for AI incidents including definitions. In parallel, the OECD has developed a global AI Incidents Monitor (AIM), tracking actual AI incidents in real time and providing a "reality check" to make sure that the reporting framework and definition function in practice (OECD.AI/incidents). AI incidents reported in international media are being used as a starting point, since many other incidents are not disclosed publicly.
- **Compute and Climate:** Alongside data and algorithms, AI computing capacity ("AI compute") is a key enabler for AI and related economic growth and competitiveness. Understanding domestic AI compute capacity is critical for policy makers who want to formulate effective AI policies and make intelligent national AI investment choices. The OECD.AI Expert Group on AI Compute and Climate is helping the OECD Secretariat

to create a basic framework for understanding, measuring and benchmarking domestic AI computing capacity by country and region.

- **AI Futures:** The OECD continues to closely monitor and analyse fast-paced advances in AI and their implications for policy. As AI evolves into generative AI and beyond, policymakers must be able to anticipate and harness these technological advances, equip individuals with the necessary skills and mitigate associated risks. In April 2023, the OECD formed an expert group on AI Futures under ONE AI to study the diffusion, impact and implications of generative AI systems, including aspects such as innovation, productivity, inclusion, employment and self-fulfillment, and education, to equip governments with the knowledge and tools necessary to accelerate the adaptation of policies accordingly.
- **OECD AI Index**: Drawing from the wealth of indicators on AI trends and policies in the OECD.AI Policy Observatory, the newly established expert group (September 2023) is working on developing a comprehensive and synthetic measurement framework on Trustworthy Artificial Intelligence. The work is carried out in collaboration with the Working Party on Digital Economics, Measurement and Analysis (DEMA) and University of Oxford Saïd Business School.
- **AI, Data and Privacy:** Established in early 2024, the Expert Group on AI, Data, and Privacy assists the OECD in creating linkages between the AI and Privacy/Data Protection communities, exploring policy opportunities and challenges, and considering further possible synergies between key existing privacy and AI frameworks drawing from the OECD's AI and Privacy Principles. The work is carried out in collaboration with the Working Party on Data Governance and Privacy (DGP).

5.2.2. OECD analytical work

96. The OECD Secretariat has carried out a wide range of analytical work based on the OECD AI Principles set out in the Recommendation and that help further disseminate and support their implementation:

- **The OECD Framework for the Classification of AI Systems** (OECD, 2022[6]) establishes a common understanding of how an AI system works in order to facilitate risk assessment nuanced to context (e.g., the difference between an AI system used for translation and one used for detecting diseases). The now widely-used Framework was developed through ONE AI with global input through public consultation. For example, it features in the United States National Institute of Standards and Technology's draft Risk Management Framework, and in the work of the European Parliament, Council and Commission towards the EU AI Act.
- **Advancing Accountability in AI report** (OECD, 2023[7]): Governing and managing risks through the lifecycle for trustworthy AI, provides a high-level interoperability framework and maps existing and emerging AI standards, frameworks and guidelines to core characteristics. The OECD plans to conduct a gap analysis and develop an interactive online tool to help organisations and stakeholders compare frameworks and navigate existing approaches for identifying, assessing, treating and governing AI risks.
- **The OECD catalogue of tools and metrics for trustworthy AI** (OECD.AI/tools): The catalogue provides a living database to which organisations around the globe contribute and update regularly. With currently over 700 tools (34 of which specifically addressing generative AI), it is a one-stop-shop for finding and sharing tools and methods – technical, normative, and educational – for making AI trustworthy. Tools are classified according to the objectives they serve, which reflect the OECD AI value-based Principles. The catalogue is cited for example in the EU-US Trade and Technology

Council's Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management.

5.2.3. Analysis of AI technological developments: generative AI

97. The technological and legal landscape of AI has rapidly advanced since the adoption of the OECD AI Principles in 2019.

98. The most notable development has been the rise of "foundation" or "general-purpose" AI models, including large language models (LLMs) that can generate novel content, translate text-to-video and-image, and offer advanced chatbots at people's fingertips. Late into 2022 and 2023, generative AI took centre stage in public, academic, and political discussions, following the launch of the chatbot ChatGPT and the rapid uptake of generative AI tools that produce new text, images, audio or video content.

99. In April 2023, the OECD was one of the first organisations to publish analysis of developments in a key area of generative AI, large language models, including benefits and risks of their use as well as policy considerations through the lens of the OECD Recommendation (OECD, 2023[52]). The OECD has also been supporting the G7 "Hiroshima Process on generative AI", namely by conducting a stocktaking exercise to examine main risks, opportunities and approaches to generative AI across G7 countries and the EU (OECD, 2023[47]). In September 2023, the OECD launched a series of Working Papers tracking generative AI developments, with a first publication focusing on technological and socio-economic policy considerations of AI language models, including on labour, bias, and misinformation and intellectual property (OECD, 2023[53]).

100. The three reports referenced above analyse generative AI through the lens of the OECD AI Principles, which proved to be a relevant and comprehensive framework to understand opportunities and risks of technological developments and their policy implications, and to identify potential mitigation measures.

6. Summary and conclusions

6.1. Implementation

101. The Report finds that the AI Recommendation is being implemented by Adherents. Since 2019, Adherents have advanced national and international level initiatives both to follow the five policy recommendations to governments and to translate the values-based Principles into action. The ten OECD AI Principles set out in the Recommendation have been used to guide national initiatives, including national AI strategies, governance and regulatory frameworks within Adherents, to frame the analytical work conducted by the OECD and Adherents, and to analyse AI technological developments.

102. Adherents reported that the OECD AI Principles provide a solid foundation for shaping national AI policies, provide guidance for policy development, promote international alignment in AI governance, and facilitate stakeholder collaboration. However, some Adherents found it difficult to implement the values-based Principles; the main challenge being the difficulty in translating them into practical, actionable measures, given their high-level and wide-ranging nature. Limited knowledge and technical capabilities in understanding AI, as well as resource limitations affecting policy development and investment were also perceived as obstacles for national implementation of the OECD AI Principles.

103. Adherents called for the OECD to continue providing guidelines and practical guidance on implementation. These include sharing best practices such as due diligence guidance on responsible business conduct in AI, international benchmarking and capacity building, and technical assistance to countries. Adherents also emphasised the need for more detailed implementation guidance for the OECD AI Principles across sectors, and including support for crafting regulatory frameworks and non-regulatory options tailored to specific cases and specific sectors like healthcare or finance, or to help address environmental challenges.

6.2. Dissemination

104. Several Adherents reported undertaking concrete initiatives to disseminate the OECD AI Principles at national level and to integrate them into the formulation and implementation of AI strategies and governance frameworks. These initiatives range from the dissemination among academic, technical, and political communities to their explicit incorporation into national AI strategies and ethical standards.

6.3. Continued relevance

105. Adherents consider that the OECD AI Principles remain relevant in guiding AI development, fostering innovation, and building public trust. The OECD AI Principles were also widely regarded as valuable guidelines for promoting responsible AI development that is in line with human rights and democratic values. The OECD AI Principles were viewed as foundational for promoting human-centric values, transparency, accountability, and fairness, preventing biases, promoting inclusivity, and ensuring that AI respects human rights and democratic values. Additionally, they were viewed as providing an ethical framework that supports inclusive growth, international collaboration, and the establishment of common standards in the global AI landscape. Some Adherents highlighted the need for complementary national or regional legislation and collaboration with other international initiatives to address potential AI risks effectively.

106. With regards to advanced AI systems, Adherents see the OECD AI Principles as a valuable tool to address important AI related policy priorities, but with differences on its

perceived effectiveness. While they stressed the importance of the Recommendation as guidance for addressing crucial AI-related challenges, Adherents also highlighted the need for practical governance frameworks and Principles tailored to address emerging concerns, such as disinformation. Adherents considered that supplementing the OECD AI Principles with more specific and actionable recommendations, especially in emerging areas like generative AI, could further enhance their effectiveness in leveraging opportunities presented by AI.

107. The OECD AI Principles serve as a significant and useful international reference in domestic AI policymaking by Adherents. They are widely disseminated, and remain fully relevant, including as a solid framework to analyse recent evolutions such as those related to generative AI.

108. The Recommendation in its current form continues to be fit for purpose overall. However, there is an opportunity to update the Recommendation to support implementation by stakeholders and to reflect emerging issues and technological advancements, including with respect to generative AI. Specific updates: i) reflect the growing importance of addressing misinformation and disinformation, and safeguarding information integrity in the context of generative AI; ii) address uses outside of intended purpose, intentional misuse, or unintentional misuse; iii) clarify the information AI actors should provide regarding AI systems to ensure transparency and responsible disclosure; iv) outline mechanisms to address potential harm or undesired behavior throughout the AI lifecycle; and v) emphasise responsible business conduct throughout the AI lifecycle, involving co-operation with suppliers of AI knowledge and AI resources, AI system users, and other stakeholders. Furthermore, some of the OECD AI Principles' headings and texts could be expanded, and the text on traceability and risk management further developed and moved to the "Accountability" Principle as the most appropriate Principle for these concepts.

References

AI Safety Summit (2023), *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023. [16]

Canadian Government (2023), *The Artificial Intelligence and Data Act (AIDA) – Companion document*, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document (accessed on 29 March 2023). [31]

Canadian Parliament (2022), *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, https://www.parl.ca/legisinfo/en/bill/44-1/c-27 (accessed on 28 March 2023). [29]

CDEI (2024), *Portfolio of AI Assurance Technique*, https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-technique. [19]

Council of Europe (2024), *Committee on Artificial Intelligence*, https://www.coe.int/en/web/artificial-intelligence/cai. [33]

Department of Labour (2024), *United States/Spain Joint Statement Algorithmic Bias in the World of Work*, https://www.dol.gov/sites/dolgov/files/ILAB/USA-SPAIN_JointStatementOnAlgorithmicBias.pdf. [46]

Digital Research Alliance of Canada (2020), *Canadian Digital Research Infrastructure Needs Assessment*, https://alliancecan.ca/en/initiatives/canadian-digital-research-infrastructure-needs-assessment. [23]

European Commission (2023), *A European approach to artificial intelligence*, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence. [21]

Fasken (2022), *The Regulation of Artificial Intelligence in Canada and Abroad: Comparing the Proposed AIDA and EU AI Act*, https://www.fasken.com/en/knowledge/2022/10/18-the-regulation-of-artificial-intelligence-in-canada-and-abroad (accessed on 29 March 2023). [30]

German Government (2022), *Regulatory Sandboxes – Testing Environments for Innovation and Regulation*, https://www.bmwk.de/Redaktion/EN/Dossier/regulatory-sandboxes.html (accessed on 29 March 2023). [28]

Ibero-American Data Protection Network (2019), *General Recommendations for the Processing of Personal Data in Artificial Intelligence*, https://www.redipd.org/sites/default/files/2020-02/guide-general-recommendations-processing-personal-data-ai.pdf. [17]

Ibero-American Data Protection Network (2019), *Specific Guidelines for Compliance with the Principles and Rights that Govern the Protection of Personal Data in Artificial Intelligence Projects*, https://www.redipd.org/sites/default/files/2020-02/guide-specific-guidelines-ai-projects.pdf. [18]

INAI (2022), *Recomendaciones para et tratamiento de datos pesonales derivado del uso de la Inteligencia Artificial*, https://home.inai.org.mx/wp-content/documentos/DocumentosSectorPublico/RecomendacionesPDP-IA.pdf. [54]

Lassébie, J. and G. Quintini (2022), "What skills and abilities can automation technologies replicate and what does it mean for workers?: New evidence", *OECD Social, Employment and Migration Working Papers, No. 282*, https://doi.org/10.1787/646aad77-en. [42]

Lorenz, P., K. Perset and J. Berryhill (2023), *Initial policy considerations for generative artificial intelligence*, OECD Publishing, Paris, https://doi.org/10.1787/fae2d1e6-en. [13]

MIC (2023), *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*, https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05_en.pdf. [49]

MIC (2023), *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System*, https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document04_en.pdf. [48]

Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal counsel and legislative affairs (2022), *Principles for Policy, Regulation and Ethics in AI: technical paper (annex to policy paper)*, https://www.gov.il/BlobFolder/rfp/061122/he/professional-letter.pdf. [34]

National Council of Science and Technology (2022), *The Networking & Information Technology R&D Program and the National Artificial Intelligence Initiative Office Supplement to the President's FY 2023 Budget*, https://www.nitrd.gov/pubs/FY2023-NITRD-NAIIO-Supplement.pdf. [20]

National Institute of Standards and Technology - US Department of Commerce (2023), *AI Risk Management Framework*, https://www.nist.gov/itl/ai-risk-management-framework (accessed on 28 March 2023). [8]

OECD (2024), *2023 OECD Digital Government Index: Results and key findings*, OECD Publishing, Paris. [25]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers* No.8, https://doi.org/10.1787/623da898-en. [2]

OECD (2024), *Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity*, OECD Publishing, https://doi.org/10.1787/d909ff7a-en. [15]

OECD (2024), *OECD database of national AI policies & strategies*, https://oecd.ai/en/dashboards/overview. [3]

OECD (2023), *2023 OECD Open, Useful and Re-usable data (OURdata) Index: Results and key findings*, OECD Publishing, Paris. [22]

OECD (2023), *A blueprint for building national compute capacity for artificial intelligence*, OECD Publishing, https://doi.org/10.1787/876367e3-en. [27]

OECD (2023), *Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI*, OECD Publishing, https://doi.org/10.1787/2448f04b-en. [7]

OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *OECD Digital Economy Papers* No. 352, https://doi.org/10.1787/13d38f92-en. [52]

OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *OECD Digital Economy Papers* No. 352, https://doi.org/10.1787/13d38f92-en. [53]

OECD (2023), *Digital Government Review of Latin America and the Caribbean: Building Inclusive and Responsive Public Services*, https://www.oecd.org/publications/digital-government-review-of-latin-america-and-the-caribbean-29f32e64-en.htm (accessed on 5 December 2023). [26]

OECD (2023), *Emerging privacy-enhancing technologies: Current regulatory and policy approaches*, OECD Publishing, https://doi.org/10.1787/bf121be4-en. [11]

OECD (2023), *G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI*, OECD Publishing, https://doi.org/10.1787/bf3c0c60-en. [14]

OECD (2023), *G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI*, OECD Publishing, Paris, https://doi.org/10.1787/bf3c0c60-en. [47]

OECD (2023), *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*, OECD Publishing, Paris, https://doi.org/10.1787/08785bba-en. [12]

OECD (2023), *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*, OECD Publishing, https://doi.org/10.1787/08785bba-en. [45]

OECD (2023), "Regulatory sandboxes in Artificial Intelligence", Vol. No. 356, https://doi.org/10.1787/8f80a0e6-en. [10]

OECD (2023), *State of implementation of the OECD AI Principles four years on*. [5]

OECD (2022), *Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint*, OECD Publishing, https://doi.org/10.1787/7babf571-en. [51]

OECD (2022), "Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint", *OECD Digital Economy Papers, No. 341*, https://doi.org/10.1787/7babf571-en. [9]

OECD (2022), *OECD Framework for the Classification of AI systems*, OECD Publishing, https://doi.org/10.1787/cb6d9eca-en. [6]

OECD (2021), "Recommendation of the Council for Agile Regulatory Governance to Harness Innovation", *OECD/LEGAL/0464*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0464. [39]

OECD (2021), *State of Implementation of the OECD AI Principles: Insights from National AI Policies*, https://doi.org/10.1787/1cd40c44-en. [4]

OECD (2020), *Reviewing the Stock of Regulation*, OECD Best Practice Principles for Regulatory Policy, OECD Publishing, Paris, https://doi.org/10.1787/1a8f33bc-en. [40]

OECD (2019), *Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449*, http://file:///C:/Users/russo_l/Downloads/OECD-LEGAL-0449-en%20(10).pdf. [1]

OECD.AI (2024), *OECD programme on AI in Work, Innovation, Productivity and Skills*, https://oecd.ai/en/work-innovation-productivity-skills. [44]

OECD.AI (2023), *Database of national AI policies & strategies*, https://oecd.ai/en/dashboards/overview (accessed on 2023). [32]

The White House (2023), *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. [38]

The White House (2023), *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/. [37]

UK DSIT (2023), *Independent Review of The Future of Compute: Final report and recommendations*, Department for Science, Innovation & Technology. [24]

UK Government (2023), *AI regulation: a pro-innovation approach*, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach (accessed on 29 March 2023). [35]

UN Advisory Board on AI (2023), *Governing AI for Humanity - Interim Report*, https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf. [50]

US Copyright Office (2023), *Copyright Office Issues Notice of Inquiry on Copyright and Artificial Intelligence*, https://www.copyright.gov/newsnet/2023/1017.html?utm_campaign=subscriptioncenter&utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term=. [41]

US Government (2022), *Blueprint for an AI Bill of Rights*, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed on 29 March 2023). [36]

Zarifhonarvar, A. (2023), "Economics of ChatGPT: A Labour Market View on the Occupational Impact of Artificial Intelligence", https://ssrn.com/abstract=4350925. [43]