THE STATE OF IMPLEMENTATION OF THE OECD AI PRINCIPLES FOUR YEARS ON

OECD ARTIFICIAL INTELLIGENCE PAPERS

October 2023   **No. 3**

This report was approved and declassified by written procedure by the Committee on Digital Economy Policy on 31 August 2023 and prepared for publication by the OECD Secretariat. For more information, please visit www.oecd.ai.

Note to Delegations: This document is also available on O.N.E. under the reference code:
DSTI/CDEP/AIGO(2023)5/FINAL

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

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by the Republic of Türkiye

The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Türkiye recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Türkiye shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union

The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Foreword

The state of implementation of the OECD AI Principles four years on was prepared under the aegis of the OECD Working Party on Artificial Intelligence Governance (AIGO) and the OECD Committee on Digital Economy Policy (CDEP). AIGO delegates and national contact points for the database of national AI policies of the OECD.AI Policy Observatory contributed significantly.

This report was a collective effort co-ordinated by the AI Unit in the OECD Digital Economy Policy division. Lucia Russo, Noah Oder (consultant to the OECD), and Karine Perset led the report development and drafting. Audrey Plonk (head of the OECD Digital Economy Policy Division) provided comments and feedback. Valéria Silva (consultant to the OECD) contributed to drafting the section on emerging AI-specific regulations. András Hlács and Sharon Ho (consultant to the OECD) contributed to the report with background research.

This report benefited from the contribution of the following experts who presented their AI policy work at the AIGO meetings in November 2022 and April 2023. They are: Paula Garnero and Vanina Martínez (Argentina), Ciro Eduardo Ferreira (Brazil), Roxane Sabourin (Canada), Juraj Bilic (Croatia), Kilian Gross, Tatjana Evas and Yordanka Ivanova (European Commission), Kristine Alanko (Finland), Jibu Elias (India), Sangwon Ko (Korea), Luis Ricardo Sánchez Hernández (Mexico), Ioan Istrate (Romania), Larissa Rim and Zee Kin Yeong (Singapore), Jana Novohradska (Slovakia), Miguel Valle Del Olmo (Spain), Isabelle Lois and Livia Walpen (Switzerland), Oleksandr Bornyakov (Ukraine), Abraham Baldry (United Kingdom), Lilly McFeeters, Lindsey Barrett, Elham Tabassi, Sam Schofield and Sorelle Friedler (United States).

In addition, the quality of the report benefited significantly from the engagement and written contributions of national delegations to the OECD Working Party on Artificial Intelligence Governance and the OECD Committee for Digital Economy Policy, namely: Tim McInerney (Australia's International, Trade & National Security Division), Ciro E Ferreira (Brazil's Ministry of Foreign Affairs), Allison O'Beirne and Roxane Sabourin (Canada's Ministry of Innovation, Science and Economic Development), Philipp Schulte (Germany's Federal Ministry for Digital and Transport), Kása Ferenc (Hungary's Neumann Technológiai Platform), Ziv Katzir (Israel Innovation Authority), Nobuhisa Nishigatan and Takayuki Honda (Japan's Ministry of Internal Affairs and Communications), Soohyeon Kang (Korea's Ministry of Science and ICT), Luis Ricardo Sánchez Hernández (Mexico's National Institute of Transparency, Access to Information and Protection of Personal Data), Christine Hafskjold (Norway's Ministry of Local Government and Regional Development), and David C. Turnbull (United States' Department of State).

The authors are grateful to John Tarver for editing this report and to Andreia Furtado for editorial and publishing support. The overall quality of this report benefited significantly from their engagement.

Abstract

The OECD AI Principles were adopted in 2019. They set out a framework containing ten principles – divided into five values-based principles and five recommendations to governments – for Members and adhering non-Members to promote and implement in their policies responsible stewardship of trustworthy AI. This report takes stock of the initiatives launched by countries worldwide to implement the OECD AI Principles and reported to the OECD.AI Policy Observatory as of May 2023. It provides an overview of national AI strategies, including their oversight and monitoring bodies, expert advisory groups, as well as their monitoring and evaluation frameworks. This report also discusses the different regulatory frameworks countries are implementing to ensure the trustworthiness of AI systems. These include national ethics frameworks and principles, emerging AI-specific regulation, and regulatory sandboxes. Finally, it presents illustrative examples of policies that implement each of the ten OECD AI Principles, to help policy makers learn from one another.

Abrégé

Les Principes de l'OCDE sur l'IA ont été adoptés en 2019. Ils fixent un cadre qui s'articule autour de dix principes – cinq principes fondés sur des valeurs et cinq recommandations à l'intention des pouvoirs publics – destinés à aider les Membres et les non-Membres ayant adhéré à l'instrument à promouvoir et mettre en œuvre, dans leurs politiques, une approche responsable à l'appui d'une IA digne de confiance. Le présent rapport propose un tour d'horizon des initiatives lancées par les pays du monde entier pour donner corps aux Principes de l'OCDE sur l'IA et consignées dans l'Observatoire OCDE des politiques relatives à l'IA (OECD.AI) jusqu'en mai 2023. Il donne un aperçu des stratégies nationales en matière d'IA, y compris des organes de suivi et de contrôle, des groupes consultatifs d'experts et des cadres de surveillance et d'évaluation. En outre, il examine les différents cadres réglementaires mis en place par les pays pour veiller à la fiabilité des systèmes d'IA, qu'il s'agisse de cadres et de principes d'éthique nationaux, de nouvelles réglementations propres à l'IA, ou de bacs à sable réglementaires. Enfin, il présente des exemples concrets de politiques qui donnent effet aux dix Principes de l'OCDE sur l'IA, afin d'aider les décideurs à tirer des enseignements de leurs expériences respectives.

Table of contents

FIGURES

TABLES

BOXES

Executive summary

In 2022, artificial intelligence (AI) took centre stage in policy discussions. The release of ChatGPT in November exposed the potential of generative AI to a mainstream audience. AI tools continue to demonstrate potential to bring transformative changes in many areas, promising productivity gains and new opportunities for individuals, businesses, and society. At the same time, AI's potential risks, such as the perpetuation of existing inequalities, the massive spread of manipulated content, and the threat to personal autonomy, have been brought to the fore, particularly by generative AI.

As the first intergovernmental standard on AI in 2019, the OECD AI Principles are now a global reference point for trustworthy AI. Countries are now working on policies based on the principles to tackle AI's risks and capitalise on opportunities.

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. Beyond OECD countries, Arab, African, and South American partner economies have also committed to actions that promote the AI principles. Overall, over 930 related policy initiatives across 70 jurisdictions had been reported to the OECD.AI policy hub by May 2023.

Countries have continued to develop policies to foster research and development (R&D) in AI and increase efforts to build and provide access to the required infrastructure to enable broader AI adoption. Some countries have strengthened their efforts to ensure different social groups develop AI skills, recognising the need to both prepare workers and citizens at large, and to monitor and accompany labour market transitions. Because AI transcends borders, many international co-operation initiatives have been launched or deepened. While some types of policy are still emerging, others are more advanced allowing countries to reflect on achievements to date and analyse further steps to advance national ambitions.

The most notable development pertains to policy actions that translate the OECD AI values-based principles into concrete, operational initiatives. This report takes stock of various types of policies designed to enhance inclusive growth, address bias and increase fairness and to make AI systems transparent, safe and accountable. Ethics framework, guidelines, codes of conduct, standards, and algorithmic impact assessment are being piloted at both national and international level, by both the public and private sectors.

Existing legislation, including on data protection and consumer protection, includes provisions relevant to AI. However, a major development in recent years has been the proposal of AI-specific regulatory frameworks that address AI high-risk systems or impacts, albeit with key differences in approach across jurisdictions. AI-specific regulation also raises new challenges in relation to international interoperability, which calls for international action to promote alignment of key definitions and their technical implementation where appropriate.

The report presents illustrative examples of the implementation of each AI Principle to help policy makers learn from one another.

Résumé

En 2022, l'intelligence artificielle (IA) s'est retrouvée au cœur des débats sur l'action des pouvoirs publics. Avec le lancement de ChatGPT, en novembre, le grand public a découvert le potentiel de l'IA générative. Les outils d'IA montrent chaque jour un peu plus leur capacité à impulser des changements profonds dans de nombreux domaines, avec à la clé des perspectives de gains de productivité et de nouvelles possibilités pour les individus, les entreprises et la société. Dans le même temps, les avancées de la technologie, en particulier de l'IA générative, ont mis en lumière un certain nombre de risques (persistance des inégalités existantes, propagation massive de contenus manipulés, ou encore risque de perte d'autonomie des personnes).

Première norme intergouvernementale dans le domaine de l'intelligence artificielle lors de leur adoption en 2019, les Principes de l'OCDE sur l'IA sont devenus une référence mondiale pour la mise en place d'une IA digne de confiance. Les pays s'en inspirent aujourd'hui pour élaborer des politiques afin de gérer les risques qu'elle induit, tout en capitalisant sur ses potentialités.

En 2017, seuls quelques pays disposaient d'une stratégie nationale en matière d'IA. Désormais, l'Observatoire OCDE des politiques relatives à l'IA (OECD.AI) recense plus de 50 stratégies nationales et initiatives gouvernementales traitant des moyens d'orienter le développement et le déploiement d'une IA digne de confiance. Au-delà des Membres de l'OCDE, des pays partenaires du monde arabe, d'Afrique et d'Amérique du Sud se sont également engagés à prendre des mesures favorisant l'application des Principes sur l'IA. Au total, plus de 930 programmes d'action mis en place dans 70 pays et territoires étaient recensés en mai 2023 sur la plateforme OECD.AI.

Les pays ont continué d'élaborer des politiques propres à stimuler la recherche et le développement (R-D) dans le domaine de l'IA et encourager les efforts pour bâtir les infrastructures indispensables à une adoption plus large de l'IA et y donner accès. Certains ont intensifié leurs efforts pour faire en sorte que différents groupes sociaux acquièrent des compétences en IA, conscients de la nécessité de préparer non seulement les travailleurs, mais aussi l'ensemble des citoyens, et de suivre et d'accompagner les mutations à l'œuvre sur les marchés du travail. Parce que l'IA transcende les frontières, de nombreux programmes de coopération internationale ont été lancés ou étendus. Si de nouveaux types de politiques continuent de voir le jour, d'autres sont plus avancés ; les pays peuvent alors en examiner les premiers résultats et analyser la marche à suivre pour donner corps aux ambitions nationales.

Les mesures transposant les principes de l'OCDE fondés sur des valeurs en initiatives opérationnelles concrètes constituent une évolution particulièrement notable. Ce rapport propose un tour d'horizon de différents types de politiques visant à favoriser la croissance inclusive, lutter contre les biais et accroître l'équité, et à faire en sorte que les systèmes d'IA soient synonymes de transparence, de sécurité et de responsabilité. Des cadres d'éthique, lignes directrices, codes de conduite, normes et évaluations d'impact des algorithmes sont mis à l'essai aux niveaux national et international, par les secteurs public comme privé.

Les législations existantes, notamment celles sur la protection des données et des consommateurs, intègrent certes des dispositions applicables à l'IA. Néanmoins, depuis quelques années, des propositions de cadres réglementaires propres à l'IA voient également le jour, qui traitent des systèmes ou des impacts à haut risque de l'IA, même si les approches varient sensiblement d'un pays à l'autre. Par ailleurs, la réglementation de l'IA pose des difficultés nouvelles en termes d'interopérabilité internationale, qui exigent une action coordonnée pour favoriser une harmonisation des principales définitions et de leur mise en œuvre technique, le cas échéant.

Ce rapport présente des illustrations concrètes de la mise en application de chacun des Principes de l'OCDE sur l'IA, afin d'aider les décideurs à tirer des enseignements de leurs expériences respectives.

Introduction

This report provides an update on policy initiatives launched since 2021, the year the OECD published the first report on the "State of Implementation of OECD AI Principles" (OECD, 2021[1]). It presents data from the OECD.AI Policy Observatory as of May 2023. While the first report focused on the implementation of the five recommendations to governments to foster trustworthy AI, this report also provides an overview of initiatives to put the OECD AI values-based Principles into practice, as well as a comparative analysis of selected emerging AI-specific regulations.

The report is organised as follows. Section 1 details national AI strategies, i.e., countries' strategic documents and bodies which structure government's efforts in the field of AI. Section 2 presents an overview of regulatory approaches to promote trustworthy AI and includes a comparative analysis of key AI-specific emerging regulations in selected jurisdictions. Section 3 reports on national initiatives operationalising the five OECD value-based principles. Section 4 discusses national initiatives implementing the five recommendations to governments.

1 National AI strategies are on the rise worldwide

In 2019, only a few countries had national AI strategies. **Canada**, **Finland**, and **Japan** were among the first to develop national AI strategies, setting targets and allocating budgets in 2017. **Australia**, **Denmark**, **France**, **Germany**, **Korea**, and the **United States** followed suit in 2018 and 2019. In 2020 and 2021, additional countries announced national AI strategies, including **Brazil**, **Chile**, **Spain**, the Republic of **Türkiye** (hereafter "Türkiye"), the **United Kingdom**, and **Ukraine**. In 2022, **Belgium, Israel, Italy** and **Thailand** launched their national AI strategies, while **Croatia, Greece, Iceland,** and **Romania** are establishing theirs. Some countries, such as **Canada**, **France** and **Germany**, have updated their national AI strategies, taking stock of achievements and keeping pace with technical, societal and economic developments. As of May 2023, 51 countries had reported a national AI Strategy to the database of national AI policies in OECD.AI (Figure 1.1), from all regions in the world.

Each country's national AI strategy has its specificities and tackles different aspects of AI policy. 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, investing in AI R&D and building human capacity.

Israel's National AI Programme (2022) aims to address all the key challenges at the national level and ensure Israel's leading position on AI for the years to come. Key pillars of the programme include, but are not limited to, supporting academic excellence and increasing the number of AI researchers in the Israeli

academia, providing the needed computation infrastructure for R&D in academia, the private as well as the public sector, and accelerating AI adoption within the public sector (Ben-Israel, Matania, E and Friedman, 2020[1]), (Artificial Intelligence and Data Science Committee, 2020[2]). In 2022, **Japan**'s Cabinet Office published the "AI Strategy 2022" (Japanese Government, 2022[3]), which sets five strategic objectives (human resources, industrial competitiveness, technology systems, international co-operation, and AI for imminent crises, e.g., natural disaster pandemic) to elaborate the three philosophies (dignity, diversity and inclusion, and sustainability) and mitigate Japan' social issues by accelerating social implementation of AI. The **United Kingdom**'s National AI Strategy (2021-2031) rests upon the three pillars of investment and planning for the long-term needs of the UK's AI ecosystem; support to the UK's transition to an AI-enabled economy; and national and international governance of AI to encourage innovation, investment, and protect the public and fundamental values (AIGO, 2022[4]). **Ukraine**'s Strategy of AI Development (2021-2030) seeks to harness the country's existing AI capacity and channel it towards a number of strategic national priorities (Atlantic Council, 2021[5]). According to the Ukrainian government, its strategy states that AI is an important basis for Ukraine to develop its digital economy, improve quality of life, increase public administration efficiency, and to deliver goods and services. It focuses on education, defence and security, the public sector, and Ukraine's key economic sectors. It also aims to attract investment in AI through the Ukrainian Start-up Fund – the largest angel investor – and the Blue and Yellow Heritage Fund, which supports AI start-ups. **India**'s approach to AI is based on the following structure: (1) AI for All, (2) Responsible AI for Social Empowerment; (3) Becoming the global garage for AI tools, especially in the developing world; (4) Focus on AI-powered language technology (Indian Government, 2023[6]). Its National Data Governance Framework Policy is a crucial development through which the government gathers datasets from all Ministries to make them machine readable and available on demand. India's government took the lead on AI and crafting the country's AI ecosystem, which will diffuse AI learning material in local languages (AIGO, 2022[4]).

In some countries, the national AI strategy is part of a broader digitalisation strategy. "Digital **Switzerland**" (2020-2022) is an umbrella strategy for all digital topics, including AI. Its principles are applicable to AI and other technologies.

For policy coherence and effective implementation of national AI policies, governments are using different models, including: (1) creating a new governmental or co-ordination body for AI; (2) assigning oversight of the development and implementation of a strategy to an existing ministry and establishing AI inter-ministerial and multi-stakeholder committees; (3) receiving input from oversight and expert advisory bodies or groups for AI and data ethics bodies. Countries are also establishing monitoring and evaluation frameworks for their national AI strategies.

Oversight and monitoring bodies manage national AI strategy implementation

In 2019, **Korea** established the Artificial Intelligence Policy Bureau as the lead centre for strategies related to AI, including data and cloud. The Bureau develops and implements government-wide AI strategies. The **United Kingdom**'s Government Office for AI, a unit within the Department for Science, Innovation and Technology, is responsible for overseeing the implementation of the National AI Strategy (OECD.AI, 2023[8]). Likewise, the **United States**' National Artificial Intelligence Initiative Office (NAIIO), established in 2021, is mandated by the National Artificial Intelligence Initiative Act (NAIIA) to coordinate and support the NAIIA and is located within the White House Office of Science and Technology Policy (National Artificial Intelligence Initiative Office, 2023[7]).

Figure 1.1. A timeline of national AI strategies

National AI strategies reported to the OECD.AI database of national AI policies, as of May 2023

LAUNCH OF
NATIONAL AI STRATEGIES

2018
European Union
(*updated in 2021*),
France (*updated in 2021*), Germany
(updated in 2020),
Hungary, India,
Mauritius, Mexico,
Sweden

2020
Bulgaria, Latvia,
Norway, Poland, Saudi
Arabia, Serbia, Spain,
Switzerland

2022
Belgium, Israel, Italy,
Thailand

2017
Canada (*updated in 2021*), People's
Republic of China,
Finland (*updated in 2021*)

2019
Argentina, Australia,
Colombia, Cyprus,
Czech Republic,
Denmark, Egypt,
Estonia, Japan
(*updated in 2022*),
Korea, Lithuania,
Luxembourg, Malta,
Netherlands, Portugal,
Singapore, United Arab
Emirates, United
States, Uruguay

2021
Austria, Brazil, Chile,
Ireland, Peru, Slovenia,
Türkiye, Ukraine,
United Kingdom,
Viet Nam

In Progress
Croatia, Greece,
Iceland, Romania

Other countries have established inter-ministerial and multi-stakeholder committees to monitor the implementation of their national AI strategy. The Governance Committee of the **Brazilian** AI Strategy is responsible for monitoring and evaluating the Brazilian AI Strategy. It is composed of private and public organisations, civil organisations, and specialists (OECD.AI, 2023[8]). In **Egypt**, the National Council for AI (NCAI) outlines, implements, and governs the AI strategy in close coordination with the relevant experts and entities. The NCAI is chaired by the Minister of Communications and Information Technology (MCIT) and includes representatives from government entities, private sector, and independent experts and heads of several bodies concerned (OECD.AI, 2023[8]). The **Serbian** government will form an AI Council composed of Ministers as well as AI industry and research leaders to follow the implementation of the AI strategy (OECD.AI, 2023[8]).

In the **People's Republic of China** (hereafter "China"), the National New Generation AI Promotion Office was established by the Ministry of Science and Technology and 14 other government agencies. It oversees the execution of the Next Generation AI Development Plan (OECD.AI, 2023[8]). The Saudi Data and Artificial Intelligence Authority (SDAIA) was established in 2019 as the owner of **Saudi Arabia**'s national and AI agenda, mandated with unlocking the value of data and AI to elevate the country as a pioneering nation among the elite league of data-driven economies. Addressing data, innovation and capability-building, SDAIA fosters the digital ecosystem while supporting the values-based G20 AI Principles (OECD.AI, 2023[8]).

Expert groups advise governments on advantages and challenges linked to AI systems

Countries have established AI expert advisory groups and multi-stakeholder groups of AI experts to provide recommendations while identifying and reporting on the current and future opportunities, risks, and challenges of the public use of AI.

In 2019, as part of its national AI strategy, **France** established a national consultative committee (FNCDE) on digital ethics and AI. The FNCDE has 27 members from different disciplines, and the working groups that prepare FNCDE's opinions often include external experts. The Prime Minister has already asked the committee for opinions on ethical issues related to digital applications that use machine learning: 1) conversational agents (chatbots); 2) autonomous cars; and 3) medical diagnosis and health AI (OECD.AI, 2023[8]). Since February 2022, **Korea** has been operating a multi-stakeholder forum (with participants from industry, academia, civil society, education and legal experts) to discuss ethical aspects of AI technologies and form a social consensus on how to build trust in AI. There are three expert committees (ethics, technology, education) within the forum to facilitate the consensus building. **Sweden**'s Committee for Technological Innovation and Ethics continuously delivers policy proposals to the government and, where relevant, also surveys the need to adapt regulatory frameworks (OECD.AI, 2023[8]). Similarly, **Singapore**'s Advisory Council on the Ethical Use of AI and Data was established in 2018 to advise the government on policy or regulatory intervention in the commercial deployment of AI.

In 2019, **Canada** established an Advisory Council on AI to advise its government on global leadership, building strengths, identifying opportunities for economic growth through AI that benefits all Canadians, and ensuring that AI advancements reflect national values. The Advisory Council is composed of experts from Canadian academia, industry, and civil society. Since 2019, it has convened two working groups: one on the commercialisation of AI, which explored ways to translate Canadian-owned AI into economic growth, and one on public awareness, which is exploring the public understanding of AI and its potential benefits and risks (Government of Canada, 2023[8]). In response to its Action Plan for the Digital Transformation of **Slovakia** for 2019-2022, the country's government formed a Standing Committee for Ethics and Regulation of AI (2020). The Committee is an independent expert advisory committee to the Ministry of Investments, Regional Development and Informatisation, which centrally co-ordinates Slovakia's AI agenda (AIGO, 2022[4]). The Swiss Federal Council established **Switzerland**'s Interdepartmental Working Group on AI within the framework of the "Digital Switzerland" strategy (OECD.AI, 2023[8]). **China**'s New Generation AI Expert Governance Committee (2019) was created by its Ministry of Science and Technology to research policy recommendations for AI governance and identify areas for international co-operation (Laskai and Webster, 2019[9]). It produced the "Governance Principles for a New Generation of Artificial Intelligence: Developing Responsible Artificial Intelligence" (Laskai and Webster, 2019[9]).

Countries establish monitoring and evaluation frameworks for national AI strategies

A few countries have launched policy intelligence activities and annual reports to evaluate the implementation of their national AI strategies. **Canada, Germany,** the **United Kingdom,** the **United States**, the **European Commission**, and **Singapore** published reports after monitoring and evaluating the implementation of their AI strategies (OECD, 2021[1]). Several national or regional institutions, such as **Chile,** the **Czech Republic, France, Germany**, and **Quebec** in Canada, have also established AI observatories to monitor the implementation of national AI strategies and policies. **Brazil** established the Brazilian Observatory of AI (OBIA) on the model of the OECD.AI Policy Observatory, serving as a

repository of data to monitor the development of AI as well as a hub for collaboration and knowledge exchange among stakeholders in the Brazilian AI ecosystem.

To date, only a few countries have conducted evaluations of their national AI strategies. The most recent example is the interim evaluation of the French National Strategy for AI (SNIA) conducted by **France**'s court of auditors (Court des Comptes, 2023[11]). The evaluation focuses on "Research" and "Higher Education", the main areas of funding of the French national AI strategy, endowed with EUR1.527 billion and EUR1.545 billion in the first and second phases, respectively. The analysis provided a total of seven key recommendations, which range from governance and monitoring (i.e., translate public policy on AI in a budget document to measure its impact; clarify the missions and funding of the centres of excellence; establish shared objectives and priority indicators for public policy on AI; create a scientific and steering committee to monitor the implementation of the strategy and to define future strategic directions), to specific actions to strengthen AI skills (i.e., produce a map of AI training courses to be promoted with a common label; develop a skills assessment needs for AI trainers and AI researchers and establish appropriate education plans), to measures related to environmental impact and the development of responsible AI (i.e., draw up a charter and catalogue of best practices to define and monitor the environmental impact of AI research; promote the development responsible AI).

These types of monitoring and evaluation frameworks and observatories are still too scarce, yet they are expected to expand across countries as national AI strategies move into later stages of implementation.

2 Countries use different regulatory frameworks to ensure the trustworthiness of AI systems

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 (Table 2.1), and iv) promoting controlled environments for regulatory experimentation (Figure 2.1).

Several countries have issued national ethics frameworks and principles

Several countries have launched national ethical frameworks and principles for AI development and deployment that largely overlap with the OECD AI Principles (up to May 2023, 17 such guidelines were reported in the OECD database). Some countries, such as **Japan**, **Korea**, and **India**, provide guidelines to developers and operators on how to implement the principles. Furthermore, **Colombia** has set up an online platform to monitor the framework's implementation.

Table 2.1. Examples of existing and emerging AI-specific regulatory approaches in select countries

Country	Legislation and regulation	Standards	Principles
Canada	• Directive on Automated Decision-Making (2019) • Proposed Bill C-27, Digital Charter Implementation Act, including AI and Data Act (AIDA) (2022l)	• Proposed CAN-ASC-6.2: Accessible and Equitable AI Systems (2023)	• Canada's Digital Charter (2019)
United Kingdom	• Proposed Online Safety Bill (2022) • Proposed Data Protection and Digital Information Bill (2023)	• Algorithmic Transparency Standard (Central Digital Data Office, 2021)	• A pro-innovation approach to AI regulation (2023)
United States	• Federal Trade Commission Act, for deceptive practices from deepfakes or chatbots (1914) • Proposed Algorithmic Accountability Act (US AAA) (2022) • Executive Order 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government (2020)	• National Institute of Standards and Technology (NIST) AI Risk Management Framework (2023)	• Blueprint for an AI Bill of Rights (2023)
European Union	• Proposed EU AI Act (2021) • Proposed updates to the EU Product Liability Directive (2022) • Proposed AI Liability Directive (2022) • EU's Digital Services Act (2022)	• CEN/CENELEC standards for AI and related data (forthcoming)	• Ethics guidelines on AI (2018)
Brazil	• Report and proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021 (2022) • Proposed Bill 705 on the compatibility of AI use in the public sector with ESG practices (2022)	• Incorporation of international standards National standards by the Brazilian Association of Technical Norms (ABNT)	• Proposed Art. 3 of the proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021 (2022)
China	• Chinese Internet Information Service Algorithmic Recommendation Provisions (2021) • Opinion on Strengthening the Ethics and Governance of Science and Technology (2022)	• National Standards for Autonomous Vehicle Testing (2018)	• New Generation AI Ethics Specifications (2019) • New Generation AI Code of Ethics (2021) • White Paper on Trustworthy AI (2021) • Internet Information Service Algorithmic Recommendation Management Provisions (2021)
Intergovernmental organisations	• Proposed Council of Europe Convention on AI, Human Rights, Democracy and the Rule of Law (2023)	• ISO 31000 Risk management (2009, 2018) • ISO/IEC 23053:2022 Framework for AI Systems Using Machine Learning (ML) (2022)	• OECD Recommendation of the Council on AI (2019) • UNESCO Recommendation on the Ethics of AI (2021)

Note: This table is a sample of emerging AI-specific initiatives from select jurisdictions at the time of writing in May 2023. Elements may have changed since. The table should thus be taken for illustrative purposes only.

Australia's Department of Industry, Science, and Resources developed its AI Ethics Framework (2019) in an effort to guide businesses and governments on the responsible design, development, and implementation of AI. The AI Ethics Framework includes eight principles designed to ensure that AI is safe, secure, and reliable, aligned with the OECD AI Principles (OECD.AI, 2023[8]). Similarly, **Colombia**'s Ethical Framework for AI provides a set of principles to consider in the design, development, and implementation of AI systems. It illustrates a methodology for determining how these principles should be considered and implemented. It includes a toolbox that describes a series of strategies to develop these principles in public entities (OECD.AI, 2023[8]). The **Korean** government has developed a checklist for implementing AI ethical standards, which is based on the National Guidelines for AI Ethics (AIGO, 2022[4]). The checklist is guided by ten key AI ethics requirements, including human rights, protection of privacy, respect for diversity,

prevention of harm, public good, solidarity, data management, accountability, safety, and transparency (OECD.AI, 2023[8]). **Switzerland** has developed its Guidelines on AI for the Confederation (2020), which are meant as a general frame of reference for the use of AI within the Federal Administration. The Guidelines overlap with several of the OECD AI Principles (OECD.AI, 2023[8]). **Argentina**'s Ethics Principles for the Development of AI also reflect both the five value-based OECD AI Principles (Principles 1.1–1.5) as well as the five recommendations to national governments (Principles 2.1-2.5) (AIGO, 2022[4]).

India's approach document is divided into two parts. Part 1, *Principles for Responsible AI*, proposes principles for the responsible management of AI systems that stakeholders may leverage. Part 2, *Operationalizing Principles for responsible AI*, identifies a series of actions for the government, the private sector and for research institutions that must be adopted to drive responsible AI (OECD.AI, 2023[8]). **Serbia**'s Ethical Guidelines for Development, Implementation, and Use of Robust and Accountable AI (2022) set ethical standards all AI solutions should embrace (OECD.AI, 2023[8]). **Singapore**'s AI Ethics and Governance Body of Knowledge is based on the Infocomm Media Development Authority's (IMDA) Model AI Governance Framework. It is tailored for practical issues related to human safety, fairness, and the prevailing approaches to privacy, data governance, and general ethical values (OECD.AI, 2023[8]). **Thailand'**s Ministry of Digital Economy and Society (DES) drafted the country's first AI ethics guidelines (2021) for researchers, developers, and service providers engaging in tech development (OECD.AI, 2023[8]). The AI Principles and Ethics for the **Emirate of Dubai** (2019) aim to help AI developers, government and society develop AI in a safe, responsible, and ethical way (OECD.AI, 2023[8]).

AI-specific regulation is emerging in several jurisdictions

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. This section provides an overview and comparative analysis of selected legislative developments across world regions. Further information is provided in Annex A and Annex B.

Canada

Canada opted for separate regulations to address automated decision systems for the provision of public federal services to the population, vis-à-vis AI systems aimed at trade and commerce.

Since 2019, Canada has implemented specific federal policy requirements around the use of automated decision-making systems in the provision of services through the Directive on Automated Decision-Making. The Directive establishes measures for algorithmic impact assessment, transparency, quality assurance, and recourse automated systems used in administrative decision-making. The Treasury Board of Canada Secretariat is obliged to review the Directive on a regular basis to account for technological and regulatory change (the third review was completed in April 2023, when the amended Directive was published).

As for the use of AI systems by the private sector in the digital economy, Canada has put forward a comprehensive regulatory framework at the federal level, the Digital Charter Implementation Act (Canadian Parliament, 2022[12]). It includes a reform of the national data protection legislation and introduces overarching, mandatory rules on AI systems. The Digital Charter comprises three self-standing pieces of legislation: Part 1, the Consumer Privacy Protection Act, aimed at updating the country's personal data protection rules to align them better with the digital economy; Part 2, the Personal Information and Data Protection Tribunal Act, creates a specialized administrative tribunal to enforce the Consumer Privacy Protection Act provisions also through fines; and Part 3, the Artificial Intelligence and Data Act (AIDA), the central piece of national legislation aimed at regulating the use of AI systems in the digital market in the course of international and inter-provincial trade and commerce.

The proposed AIDA's approach is to ensure the safe and responsible design, development, and deployment of AI systems that respect Canadians' values, and to establish an Artificial Intelligence and Data Commissioner that administers and enforces the Act, while relying on sectoral regulators to enforce in their areas of responsibility. It establishes an impact-based approach that focuses on mitigating the risks of harm and bias of "high-impact" AI systems (Fasken, 2022[13]). Its main objectives are: (1) to establish common requirements throughout Canada with respect to the design, development, and use of AI in the private sector; and (2) to prohibit conduct in relation to AI systems that could result in serious harm to individuals or their interests (Landry et al., 2022[14]).

The Canadian Government recently published a companion document to the proposed AIDA (Canadian Government, 2023[15]) to clarify the legislation's goals, approach, scope, and enforcement choices. For example, AIDA does not specify the scope of high-impact systems, whereas its companion document explains that AIDA will cover systems that could significantly impact health, safety, and human rights. This aligns with the proposed approach in the EU AI Act. It stresses that the proposed AIDA does not intend to discourage innovation or target private actors acting in good faith but rather "to regulate the most powerful uses of this technology that pose the risk of harm". The AIDA companion document 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. It also recognizes the relevance of international coordination with these and other international partners such as the United Kingdom and the United States, to ensure regulatory interoperability in the global marketplace.

The Digital Charter Implementation Act is currently under debate in the House of Commons and will then proceed to the Senate. The new legal provisions are therefore projected to come into force as early as 2025 (Canadian Government, 2023[15]).

Israel

In November 2022, the Israeli Ministry of Innovation Science and Technology and the Ministry of Justice published a "Draft Policy White Paper for Regulation and Ethics in the Field of AI" for public consultation (Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal counsel and legislative affairs, 2022[16]). The white paper covers the main regulatory and legal challenges related to AI and makes regulatory policy recommendations.

Drawing on the OECD AI Principles, the white paper adopts non-binding AI ethical principles to be considered when developing, using, and regulating AI. It calls for sector-based regulatory efforts (rooted in risk assessment and management approaches), rather than overarching sector-crossing regulation, encouraging regulators to develop their respective frameworks in a manner that is consistent with those of leading countries in the field. The white paper also prioritises the use of "soft" regulation and advanced regulatory tools such as ethical principles, standards, recommendations for voluntary adoption or self-regulation. It further suggests adopting a gradual modular framework to develop regulation by using tools for controlled regulatory experimentation, such as sandboxes. Finally, the white paper calls for continued multi-stakeholder dialogue to develop AI regulation.

In July 2022, the Office of Legal Counsel and Legislative Affairs (Department of Economic Law) in the Israeli Ministry of Justice published a report on the use of AI in the financial sector, compiled by an interdisciplinary team of renowned researchers from Tel Aviv University. After reviewing the report, the financial regulators, the Ministry of Justice, the Ministry of Finance, and the Competition Authority decided to establish a joint taskforce to provide appropriate recommendations on these matters. The taskforce was established in December 2022, with a mandate to review the regulatory and legal implications of the use of AI in the financial sector, to recommend steps to promote innovation in this sector, and to suggest required amendments to current regulation (Israeli government, 2023[17]). The taskforce includes members from the Ministry of Finance, the Ministry of Justice, the Bank of Israel, the Capital Market, Insurance and

Savings Authority, Israel Securities Authority, and the Competition Authority. It is expected to present its recommendations by October 2023.

Japan

In Japan, the Cabinet Office has published the Social Principles of Human-Centric AI (Japanese Government, 2019[18]). The Social Principles resulted from multi-stakeholder expert group discussion, set forth three basic philosophies (dignity, diversity and inclusion, and sustainability), as well as seven AI principles (human-centric, education and literacy, privacy, fair competition, security, innovation and fair, accountable and transparent). These principles share some common elements with the OECD AI Principles since several members of the group have made contribution to the development of the OECD AI Principles. The Governance Guidelines for the Implementation of AI Principles (2022) summarise how to implement the Social Principles of Human-Centric AI decided by the Council for Integrated Innovation Strategy in 2019. The AI Governance Guidelines provide practical guidance for AI system operators and developers (Ministry of Economy, Trade and Industry, 2022[19]).

In April 2023, Japan hosted the G7 Digital and Tech Ministers' Meeting in Takasaki. The Ministers agreed on the Ministerial Declaration, which emphasises the importance of: i) international discussions on the interoperability between different AI governance frameworks, and ii) stock-taking of the opportunities and challenges brought by generative AI. The Ministers also adopted the "G7 Action Plan for promoting global interoperability between tools for trustworthy AI" (MIC, 2023[20]). The Ministerial discussion on AI was escalated to the Leaders' discussion at the G7 Summit meeting in May, hosted in Hiroshima. The Leaders agreed to task their Ministers to establish the "Hiroshima AI process", where G7 members continue the discussion on generative AI in an inclusive manner and in co-operation with the OECD and GPAI by the end of 2023 (Ministry of Foreign Affairs of Japan, 2023[21]). In addition to these efforts on the governmental level, the G7 Data Protection and Privacy Authorities adopted a "Statement on Generative AI" at their Tokyo roundtable meeting in June 2023 highlighting specific challenges related to data protection in generative AI.

Working group meetings to advance for the Hiroshima AI process, supported by the OECD, took place from June through September 2023, and will continue through the remainder of 2023. On 7 September, 2023, G7 Digital and Tech Ministers issued a statement (MIC, 2023[23]) endorsing: 1) a report by the OECD summarising a stocktaking of priority risks, challenges, and opportunities of generative AI based on priorities highlighted in the G7 Leaders' Statement (OECD, 2023[23]), (2) ongoing work towards international guiding principles applicable for all AI actors, (3) developing a code of conduct for organisations developing advanced AI systems, to be presented to the G7 Leaders, and (4) developing responsible AI tools and best practices.

Responding to the emergence of generative AI and related international discussion advancements, Prime Minister Kishida announced the launch of the "AI Strategic Council" in May 2023. The Council brings together the AI experts and responsible Ministers in the Prime Minister's office to identify the opportunities and challenges of generative AI, and to develop a whole-of-government policy programme to accelerate the R&D and use of AI in Japan, including to lead the international discussion on rulemaking. The Council published a preliminary summary of discussion and list of AI issues in May 2023 (Japanese government, 2023[22]).

United Kingdom

In a policy paper that delineates the country's regulatory choice on AI, the United Kingdom lays down a context-specific, sectoral approach to regulating AI. "Establishing a pro-innovation approach to Regulating AI", presented to the Parliament on July 18, 2022, is a vertical approach that characterises regulation with the following key principles: (1) context-specific; (2) pro-innovation and risk-based; (3) coherence; and (4)

proportionate and adaptable (UK Government, 2022[23]). On March 29, 2023, the United Kingdom published its policy paper "AI regulation: A pro-innovation approach" (UK Government, 2023[24]).

The government opted for what it calls a pro-innovation framework, based on the following four key elements: cross-sectoral principles, leveraging existing regulator expertise, context-specific approach, and central functions to drive coherence (AIGO, 2023[25]). The policy paper establishes five cross-sectoral principles for AI systems, based on the OECD AI Principles. It also recommends regulators to favour a soft approach instead of mandatory regulation, and leaves them to decide on the rules for each sector. This regulatory approach seeks to enable sectoral flexibility and adjustment as the technology evolves. Instead of defining AI, the paper opts to determine what it considers to be its core characteristics and capabilities. It proposes that regulators take them into account when developing their own definitions of AI according to their specific domains or sectors.

The United Kingdom's approach differs significantly from the proposed EU AI Act: the former is highly decentralised, focuses on actual risks and harms, relies on sectoral regulation, and favours voluntary measures and guidance over mandatory regulation (UK Government, 2022[26]).

On March 3, 2023, the United Kingdom introduced the Data Protection and Digital Information Bill in the House of Commons (UK Parliament, 2023[27]). This proposed bill brings an entire new section on automated decision-making. It adopts a more business-friendly approach in comparison to the existing UK General Data Protection Regulation (UK GDPR). Whereas the UK GDPR generally prohibits automated profiling decision-making - for example, to obtain a job or a bank loan -, the proposed bill allows it by default, placing the responsibility to challenge such decisions and request a human review on the individuals affected. In addition, only activities posing high risks will have to carry out data processing requirements (UK Parliament, 2023[27]), with the aim to reduce paperwork and compliance obligations for businesses.

United States

The United States has advanced federal AI policies on four fronts.

A horizontal framework, the Algorithmic Accountability Act (AAA) (US Congress, 2022[28]), was proposed in the Senate on February 3, 2022. The US AAA establishes a horizontal framework for companies to assess the impact of automated decision systems they sell and use through impact assessments and post-market monitoring. The bill also aims to increase the transparency and traceability of automated decisions. There is still uncertainty on whether the United States Senate and House will support the bill, which two members of Congress introduced. Other legislative packages are also being considered (e.g., proposals by Sen. Schumer (NY)), but their future path remains equally uncertain.

Given this context, Executive Branch policies have acquired major relevance in the United States. Two major documents of voluntary nature have shaped these policies: the Blueprint for an AI Bill of Rights, a white paper published by the White House Office of Science and Technology Policy in October 2022 (US Government, 2022[29]), and the AI Risk Management Framework (AI RMF), consisting of technical guidelines released by the National Institute of Standards and Technology (NIST) on January 26, 2023 . The Blueprint 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[29]). It establishes five principles: (1) safe and effective systems; (2) algorithmic discrimination protections; (3) data privacy; (4) notice and explanation; and (5) human alternatives, considerations, and fallback, to mitigate risks to civil rights and democratic values posed by the use of automated systems across sectors. The AI RMF issued by NIST incorporates similar rights-preserving principles into technical guidelines and standards, and builds on the OECD AI Principles and the OECD Framework for the Classification of AI Systems (OECD, 2022[34]) to propose a framework to map, measure, and manage AI risks.

Binding federal regulations in force are mostly sectoral or domain-specific, for existing regulations and emerging regulations expressly targeting AI and related technologies, both of which may have the recommendations of the Blueprint or the AI RMF incorporated into their application. An important piece of AI-specific policy is Executive Order 13960, which mandates all design, development, acquisition, and use of AI systems by the US federal government to adhere to eight principles (Executive Office of the President, 2020[31]). Many of these principles, including lawfulness, safety, security, and resilience, understandability, responsibility and traceability, transparency, and accountability, overlap significantly with the OECD AI Principles. In addition, Executive Order 14091 includes provisions to root out bias in the design and use of AI, and protect the public from algorithmic discrimination by instructing the Federal Government, when designing, developing, acquiring, and using artificial intelligence and automated systems, to do so, consistent with applicable law, in a manner that advances equity. Binding AI implementation guidance for federal agencies from the White House Office of Management and Budget is forthcoming (Executive Office of the President, 2020[31]).

Finally, on July 21, 2023, the White House secured voluntary commitments from seven leading AI companies to manage the risks posed by AI and to help move toward safe, secure, and transparent development of AI technology (The White House, 2023[32]). The White House also announced it would be developing another Executive Order and will pursue bipartisan legislation on AI and responsible innovation.

European Union

The EU is advancing the EU AI Act ("AI Act"), proposed by the European Commission in 2021 (European Commission, 2021a[33]). The proposed Act follows a risk-based approach and presents a uniform, horizontal legal framework for AI to ensure legal certainty (European Commission, 2021a[33]).

The proposed EU AI Act is part of the European Coordinated Plan on Artificial Intelligence (2018, revised in 2021 (European Commission, 2021b[34])), aimed at shaping Europe as a major world player in AI innovation, embedding it with human-centric, trustworthy, secure, and sustainable values, and addressing risks through harmonised regulation. This regulation emerged as one of the major outputs of the Plan. It seeks to establish an overarching regulation to avoid fragmented regulation across Member States. The proposed AI Act seeks to balance the EU's goals to accelerate innovation and to mitigate the risks of AI systems, including potential threats to European values.

The proposed EU AI Act introduces a classification of AI systems based on the levels of risk they represent. Risks subject to the AI Act are those for health and safety, adverse impact on fundamental rights (as well as on the environment, democracy and rule of law, in the compromise text adopted by the European Parliament), or those that can result from the areas of application of AI specified in Annex III of the AI Act (European Commission, 2021a[33]). Risks are classified as unacceptable (prohibited), high (subject to conformity assessment procedures before placing an AI system on the market, as well as to post-market monitoring); limited (subject to transparency obligations), and minimal or no risk (not covered by the Regulation). The proposed EU AI Act allows some level of flexibility in the specification of AI systems that pose a high risk. As such, rather than listing them in the text of the regulation, an annex (Annex III) is proposed for this purpose. This choice will enable the European Commission, when the conditions of Article 7 are satisfied, to amend Annex III and adequate it to new or unforeseen uses, as well as to emerging technologies that may pose significant risks.

Negotiations in the Council of the European Union led to a Compromise Version of the proposed EU AI Act on December 6, 2022 (European Council, 2022[35]). In June 2023, negotiations at the European Parliament led to a compromise text which notably adopts the OECD definition of AI systems (Bertuzzi, 2023[36]). The inter-institutional negotiations between these institutions and the European Commission ("trilogue") are expected to lead to a final proposal in the second half of 2023, with the AI Act set to come into force at the end of 2023 or in early 2024.

The original text of the proposed EU AI Act did not include general purpose AI (that is, AI systems that can be used for several purposes). Following negotiations at the Council of the European Union, general purpose AI systems were included in a compromise version of the AI Act. As per the text, the European Commission will carry out consultations and impact assessments, to determine when the requirements for high-risk AI systems would be equally applicable to general purpose AI systems.

The Parliament's compromise text (European Parliament, 2023[37]) introduces a new definition of "foundation model", an AI model that is trained on broad data at scale, designed for generality of output, and that can be adapted to a wide range of distinctive tasks. Generative AI is a subcategory of AI foundation models. Generative AI is also defined in a newly inserted Article 28a: "foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI")".

Council of Europe

From 2019 to 2021, the Council of Europe's Ad Hoc Committee on Artificial Intelligence (CAHAI) examined the feasibility and potential elements of a possible legal framework to ensure that AI is used to promote and protect CoE's standards. CAHAI was succeeded by the Committee on Artificial Intelligence (CAI) in 2022 that built on the work of its predecessor and began to work on a "Framework Convention Artificial Intelligence, Human Rights, Democracy and the Rule of Law" (Council of Europe, 2023[79]). The objective of the drafting process is to focus on common principles ensuring the seamless application and respect for human rights, democracy, and the rule of law in a context where AI systems assist or replace human decision-making. CAI is expected to finish drafting its framework by the end of 2023 and send it to the Committee of Ministers for approval. Once approved, the Convention will become an international legal instrument binding to its signatories.

Brazil

In Brazil, a committee of legal practitioners and experts was charged by the government to draft a proposal for an AI regulation. The findings and the report of the committee resulted in the proposal of a Bill (Bill nº 2338/2023), whose main aspects are: human rights-oriented approach; risk classification of AI systems; AI governance and risk-based-approach; the establishment of a supervisory authority; rules for civil liability; fostering of innovation by promoting regulatory sandboxes, among others (OECD.AI, 2023[8]).

The proposed Brazilian AI Bill differs from the proposed EU AI Act in which it addresses the rights of those affected by AI systems and puts them upfront (Chapter II). Those rights are grounded on principles such as human-centred AI; human rights and democratic values; personality, privacy and data protection; environment protection and sustainable development; equality, non-discrimination, plurality and labour rights; free enterprise, competition and consumer protection; informative self-determination; access to information and education; and the promotion of R&D to stimulate innovation (Art. 2).

Chapter IV addresses the governance of AI systems with the same approach as the proposed EU AI Act and the proposed Canadian AIDA. The proposed Brazilian Bill establishes duties for providers and deployers of AI systems, including transparency related to their use and internal governance. Conformity assessment procedures are required for high-risk systems.

Chapter V of the proposed Brazilian Bill addresses civil liability and is applicable when the violation of rights leads to damages, whether patrimonial, moral, individual or collective in nature. The rights of those affected by AI systems and the resulting civil liability are established regardless of the levels of risk: those not classified as unacceptable or high-risk lead to a legal presumption of guilt on the part of the agent who allegedly caused the damage, thus shifting the burden of proof in favour of the allegedly injured party.

The Bill was submitted by the President of the Brazilian Senate in May 2023, and it is currently under analysis before the National Congress (OECD.AI, 2023[8]).

China

China is advancing AI policies with both soft principles and hard rules, in addition to incentives for private actors and local governments to innovate (CAIDP, 2021[39]); (Roberts et al., 2021[40]). In 2019, the Chinese Ministry of Science and Technology issued ethical norms for the use of AI in the country, although several reports have expressed concern that AI systems' purpose and application in China, notably in facial recognition applications, negatively impact human rights, which may not align with the OECD and G20 AI Principles (CAIDP, 2021[39]); (Shaughnessy et al., 2023[41]). The Ministry of Industry and Information Technology published a white paper acknowledging the risks posed by AI systems and proposing an overarching "trustworthy AI framework" for the implementation of ethical principles and the creation of industry-wide standards of trust. The white paper calls for the acceleration of legislation and supervision of trustworthy AI by the central government and for regulatory experimentation through the use of sandboxes and other methods. Finally, in 2022, the Central Office of the Communist Party of China and the Office of the State Council issued an Opinion on Strengthening the Ethics and Governance of Science and Technology (Chinese Government, 2022[42]). This document in intended to have broad application and signals the creation of additional regulation to determine the scope of its implementation.

In China, there are also efforts on the provincial level to develop AI regulation. For example, in 2022, the Shanghai Municipal People's Congress introduced the Shanghai Regulations on Promoting the Development of the AI Industry (Holistic AI, 2023[43]). The Shanghai Regulations are the first provincial-level regulations in the field of AI. They set a graded management system and enforce sandbox supervision whereby companies have a designated space to test and explore AI technologies (Holistic AI, 2023[43]). Similarly, Shenzhen has its Regulation for the Promotion of the AI Industry. The Regulation sets a risk-based approach to encourage governmental organisations in Shenzhen to be at the forefront of AI adoption and development by increasing financial support for these endeavours (Holistic AI, 2023[43]).

A comparative analysis of similarities and differences in AI-specific regulations across selected jurisdictions

Some jurisdictions are taking a cross-sectoral "horizontal" approach to AI regulation, while others consider a more sectoral or "vertical" approach

Canada and the **European Union** have proposed to regulate AI systems across domains and applications, building an AI-specific regulatory framework applicable to all sectors. They have proposed a "horizontal" regulatory approach, so as to establish minimum standards of mandatory application across industries. Standards will clearly determine how businesses should manage the technology, enabling them to do it responsibly, while simultaneously enhancing consumer trust in their use and creating safeguards to protect individuals and groups against risks to health, safety, and human rights.

Other jurisdictions consider a more sectoral or "vertical" approach, developing regulations by sector or domain. This is the case in Israel, the United Kingdom, the United States, and in China.

The **United Kingdom** laid down a context-specific, sectoral approach on AI in the policy paper that delineates the country's regulatory approach (OECD.AI, 2023[44]). It establishes cross-sectoral, non-binding principles while leaving regulators the task of implementing, regulating, and enforcing them in their respective sectors and domains. It has also expressly called regulators to favour "lighter-touch options, such as guidance or voluntary measures" as a first choice (OECD.AI, 2023[8]). **Israel** follows a similar approach, refraining from enacting broad, horizontal legislation and rather preferring a sectoral based approach (Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal counsel and legislative affairs, 2022[16]).

The **United States** has some regulation that applies to federal government activities around AI, but for other actors it does not have AI-specific cross-regulatory frameworks at the federal level at this stage. It has so far relied on voluntary frameworks, such as the White House Blueprint for AI Bill of Rights (US Government, 2022[29]) and NIST's technical guidelines and voluntary AI Risk Management Framework (National Institute of Standards and Technologies - US Department of Commerce, 2023[33]). Sectoral regulation is still incipient. For example, the United States Department of Transportation announced it will develop regulation that advances innovation for automated vehicles while accounting for safety in the integration of new technologies (US Government, 2021[45]). In addition, individual state (e.g., Colorado and Illinois) and local governments (e.g., New York) have undertaken initiatives to regulate limited and specific uses of AI such as for recruitment, insurance, and auditing. The AI industry has even adopted its own initiatives, such as the Algorithmic Bias Safeguards for the Workforce, a questionnaire for employers to assess algorithmic discrimination prior to acquiring software for the evaluation of their workers (Data & Trust Alliance, 2022[46]).

China has advanced domain-specific regulation and standards in recent years, in areas such as data protection (Chinese Government, 2021[47]) and automated driving (OECD.AI, 2023[8]). Another recent regulatory initiative is the Internet Information Service Algorithmic Recommendation Management Provisions (Zhang, 2021[48]), which establish rules on the use of algorithms that recommend and disseminate online information. They prohibit the dissemination of misinformation and place legal liability for algorithmic-based recommendations on the digital information providers.

Both sectoral and cross-sector regulations usually exist within one country:

- **Canada** and the **European Union** have aimed for a higher level of harmonisation of definitions, the scope of application and centralisation of enforcement at the national level. This can be combined with targeted sector or domain-specific initiatives (laws, guidelines, or standards) to take into account specific characteristics and needs.

- **Israel**, the **United Kingdom** and the **United States** have opted for horizontal, overarching strategies and ethical principles to set a common ground for sectoral regulations, decentralising competence to gain agility when accounting for changes of scenarios. In these jurisdictions, courts will have a fundamental role in harmonizing interpretation across regulators when appropriate.

China also applies a horizontal approach to bring its vertical regulations into coherence. However, it differs from the United States and the United Kingdom with regards to the level of compliance required. The latter countries favour voluntary horizontal principles and guidance, as well as inter-ministerial exchanges to foster coordination towards coherence (UK Government, 2022[23]), whereas China seems to prefer centralized, mandatory horizontal rules. For example, it has created ethical norms applied to AI and conferred the National New Generation Artificial Governance Specialist Committee exclusive competence for the publication, interpretation, and guidance on the implementation of the norms (Chinese Government, 2021[47]).

New AI-specific regulation also calls for new governance and enforcement bodies

In **Canada,** the Minister of Innovation, Science and Industry would be in charge of the governance and enforcement of the proposed AIDA. AIDA intends to create a new AI and Data Commissioner, which would effectively carry out administration and enforcement; track potential systemic effects of AI systems to inform decisions; and establish coordination across the government to ensure consistency in the implementation of the proposed AIDA (Canadian Government, 2023[15]). Enforcement for criminal violations would remain under the competence of the Public Prosecution Service of Canada.

In February 2023, **Israel** established a government centre for AI regulation, which will assist and coordinate the work of sectoral regulators.

The **United Kingdom** preconises policy coordination to ensure policy coherence across sectors. The already existing Digital Regulation Cooperation Forum, which fosters co-operation across national regulatory authorities on online matters, is likely to steer co-operation on AI policy without excluding other initiatives. The Alan Turing Institute has called attention to the current lack of capacity, understanding of AI and co-operation among regulators. It proposes the creation of a dedicated regulatory hub to address these issues (Aitken et al., 2023[49]).

Under the proposed EU AI Act, the **European Union** will create the European AI Board or Office, a new authority to ensure EU-wide oversight of the AI Act. As with the GDPR, national authorities will be charged with enforcement and establishing administrative sanctions according to the baseline established in the proposed EU AI Act. At this initial stage, some EU member states (among which the Netherlands, see below) have opted to create a unit in charge of algorithms inside their national Data Protection Authorities. In addition, each EU member state will determine whether, under what conditions, and to what extent civil and criminal sanctions will be applicable.

National courts of EU member states already have the competences to establish non-contractual civil liability and to sanction manufacturers of defective products for the harms they cause to consumers under the existing EU Product Liability Directive. Two proposals published by the European Commission in September 2022 – one to update the Product Liability Directive, another to create an AI Liability Directive – aim to equip national courts to establish liability throughout the lifecycle of an AI system. Courts will further be able to establish liability and sanction not only for manufacturers outside of the EU, but also operators and users of AI systems.

China established a new, centralised data regulator, the National Data Administration (2023), which is in charge of the creation and enforcement of rules on data, investigation of algorithmic manipulation and deepfakes, although the scope of its regulatory power remains unknown (Yang, 2023[50]). The National Data Authority will help advance smart cities and governmental digital services, facilitate data sharing across governmental agencies, and improve digital infrastructure (Yang, 2023[50]).

Technical standards will be crucial to implementing trustworthy AI

Standard setting bodies are stepping up to create technical standards for AI systems.

In the **United States,** NIST has established voluntary guidelines and technical standards for AI risk management (NIST, 2023[51]). These standards build on the *OECD Framework for the Classification of AI Systems* and have received broad support. Given their voluntary nature, their effectiveness will depend on the extent to which organisations adopt and implement them.

In 2022, **Türkiye** established the AI Mirror Committee to effectively oversee the standardization efforts within the field of AI, and to ensure that the country's perspectives and evaluations are duly incorporated. This committee brings together representatives from the public sector, private sector, academia, and from Non-Governmental Organisations (NGOs). By convening diverse expertise and perspectives, the Mirror Committee aims to represent the interests and viewpoints of Türkiye, contributing to the development of robust and inclusive standards in AI.

In the **European Union**, the European Committee for Electrotechnical Standardization (CEN-CENELEC) will develop technical standards to operationalise the EU AI Act. The International Organization for Standardization (ISO) has developed *ISO/IEC 23053* (2022), establishing a *Framework for Artificial Intelligence Systems Using Machine Learning*, in addition to *ISO 31000* (2009) for risk management that is applicable across sectors and activities (ISO, 2022[52]; ISO, 2009[53]).

Standard-setting organisations play a critical role in building consensus among AI actors. They can also help to promote interoperability between jurisdictions and offer market certainty for those using or developing AI systems in different parts of the world. Promoting wide participation from relevant parties in the establishment of such standards will be critical, to ensure different perspectives are considered for effective risk management. To increase participation in AI standards development, the United Kingdom established the AI Standards Hub (Box 2.1).

Box 2.1. The United Kingdom AI Standards Hub

As part of the **UK**'s National AI Strategy, the Hub's mission is to advance trustworthy and responsible AI, focusing on standards' role as governance tools and innovation mechanisms. The Alan Turing Institute leads the AI Standards Hub in partnership with the British Standards Institution (BSI) and the National Physical Laboratory (NPL). The initiative is supported by the UK Government through the DCMS Digital Standards team and the Office for AI.

The AI Standards Hub aims to help stakeholders navigate the rapidly growing range of AI-related activity by an increasing number of Standards Development Organisations (SDOs) around the world. It actively participates in international AI standardisation and informs the direction of these efforts. Dedicated to knowledge sharing, community and capacity building, and strategic research, the Hub brings together industry, government, regulators, consumers and civil society, and academia to shape debates about AI standardisation and promote sound, coherent, and effective standards. These standards inform and strengthen AI governance practices domestically and internationally, and increase multi-stakeholder involvement, while facilitating the assessment and use of relevant published standards.

Source: (AI Standards Hub, 2023[54]).

Governments and innovators are testing innovative AI solutions in controlled environments

An increasing number of countries use regulatory sandboxes (Figure 2.1), i.e., spaces in which authorities engage firms to test innovative products and services that challenge existing legal frameworks (OECD, 2023[59]). **Germany'**s Federal Ministry for Economic Affairs and Climate Action (BMWK) has launched the Regulatory Sandbox Strategy (2019) with the aim of fostering digital innovation and further developing the regulatory framework for AI and other digital technologies (German Government, 2022[56]). Promoting regulatory experimentation is also one of **Israel**'s national AI strategy's key tools to ensure safe and innovative AI deployment. **Spain** created an AI regulatory sandbox in 2022 as the first pilot programme to test the future proposed EU AI Act. The initiative is in collaboration with the EC and is seeking to onboard other EU member countries (OECD, 2023[59]). The key distinguishing feature of the Spanish AI regulatory sandbox is that it was established to test a regulation that has not yet entered into force. The objective of this pilot programme is to test the proposed regulatory framework with real AI applications to assess how both the regulation and applications respond, and to suggest modifications or explanatory guidelines.

The **United Kingdom** launched two regulatory sandboxes through the Financial Conduct Authority (FCA) and the Information Commissioner's Office (ICO). The FCA Sandbox (2016) focuses on FinTech while also admitting AI-related solutions applied in the financial sector (OECD, 2021[57]). Inspired by the ICO regulatory sandbox, the **Norwegian** Data Protection Authority (*Datatilsynet*) Regulatory Sandbox (2020) aims to promote ethical, privacy-friendly, and responsible innovation within AI. The Norwegian sandbox follows the principles of responsible AI as proposed by the EU High Level Group on Trustworthy AI, and

in 2020-2023, it selected twelve projects (Datatilsynet, 2020[61]), (Datatilsynet, 2021[61]). Another example of sandboxes in the financial sector is represented by the FinTech Regulatory Sandbox established by the Monetary Authority of **Singapore**'s (MAS), which has facilitated the live testing of AI applications such as the Kristal.AI case (Singapore Government, 2023[58]).

Figure 2.1. Regulatory sandboxes in AI: definition, risks, opportunities and policy considerations

What are regulatory sandboxes?

Sandboxes create spaces in which authorities engage firms to test innovative products or services that challenge existing legal frameworks. They are promising for areas with fast innovation cycles such as Artificial Intelligence (AI) and financial technology (fintech), since participating firms obtain a waiver from specific legal provisions or compliance processes to allow them to innovate. Approaches to regulatory sandboxes differ widely but share common characteristics:

- they are temporary, with a testing process usually limited to six months;
- they bring together regulators and firms;
- they waive existing legal provisions and provide tailored legal support for a specific project, often based on trial-and-error; and
- the technical and market information and data they collect can help regulatory authorities assess whether specific legal frameworks are fit-for-purpose or whether they need to be adapted.

Policy considerations

To maximize the benefits of regulatory sandboxes and address the associated challenges, countries should first enhance multi-disciplinary and multi stakeholder co-operation. AI products and services are complex and often affect several areas, so that several regulatory authorities must be involved in their testing. There must be institutional co-operation between firms, competition authorities, intellectual property offices, national standardisation bodies, and data protection authorities, among others. Second, regulatory authorities need to build AI technical expertise to allow them to make efficient decisions on access to sandboxes and to develop testing frameworks. Third, international co-operation on interoperable experimentation frameworks, including testing parameters for AI, could benefit innovation and decrease regulatory fragmentation. Fourth, considering the impacts on innovation and competition is key. Regulatory sandboxes could impact consumers, fundamental rights, innovation, and competition, which should assessed as early as possible.

Opportunities

Regulatory sandboxes bring about several benefits. By providing a flexible regulatory framework, sandboxes can facilitate a firm's financing and market entry, as they reduce administrative, compliance and transactional costs. Companies thus need to devote fewer resources to regulatory compliance, which is particularly beneficial for startups and small businesses. Sandboxes also enable regulators to understand new technologies and business models early on and identify risks. This can contribute to developing appropriate regulations and safeguards to minimize potential harm to consumers or the market as a whole. Indeed, a common outcome of sandboxes is that regulators issue regulatory amendments or guidance on how to interpret legal frameworks.

Risks

Implementing sandboxes also comes with certain challenges and risks. There is a lack of harmonization and standardisation of eligibility criteria and testing processes. Sandboxes require careful design and thorough testing, and inadequate specifications can harm competition, consumers, and public or personal data. Differing standards can furthermore lead to some more lenient sandboxes that promote arbitrage and forum shopping. If regulatory sandboxes are not adequately communicated or understood, there is also a risk that public trust in the regulatory process will be compromised.

Source: (OECD, 2023[59]), "Regulatory sandboxes in Artificial Intelligence", https://doi.org/10.1787/8f80a0e6-en.

3 Implementing the OECD AI values-based principles

The OECD Recommendation of the Council on Artificial Intelligence (OECD, 2019[65]) identifies five complementary values-based principles for the responsible stewardship of trustworthy AI: (1.1) inclusive growth, sustainable development and well-being, (1.2) human-centred values and fairness, (1.3) transparency and explainability, (1.4) robustness, security and safety, and (1.5) accountability.

These principles have been endorsed by 46 countries (OECD and partner economies) as the core values to lead the trustworthy deployment, development, and use of AI. When developing national AI strategies, most countries refer to these guiding principles. National and international ethics frameworks and principles also largely embed these principles. Finally, emerging AI-specific legislation require the implementation of the five OECD AI values-based principles.

The following sections describe practical initiatives for the implementation of the OECD values-based principles, based on an allocation framework of policy initiatives to each principle. This framework enhances the practical use of the values-based principles by defining them in a clearer and more practical way. Multiple principles can be attributed to each policy and is desirable in most cases, since this means that a policy considers several principles.

The following sections include policy examples for each principle. While the examples are drawn from the database of national AI policies, the rapid development of AI policies and regulations makes it challenging to be exhaustive. This collection is therefore illustrative.

Figure 3.1. Breakdown of issues covered by policy initiatives linked to OECD AI values-based principles
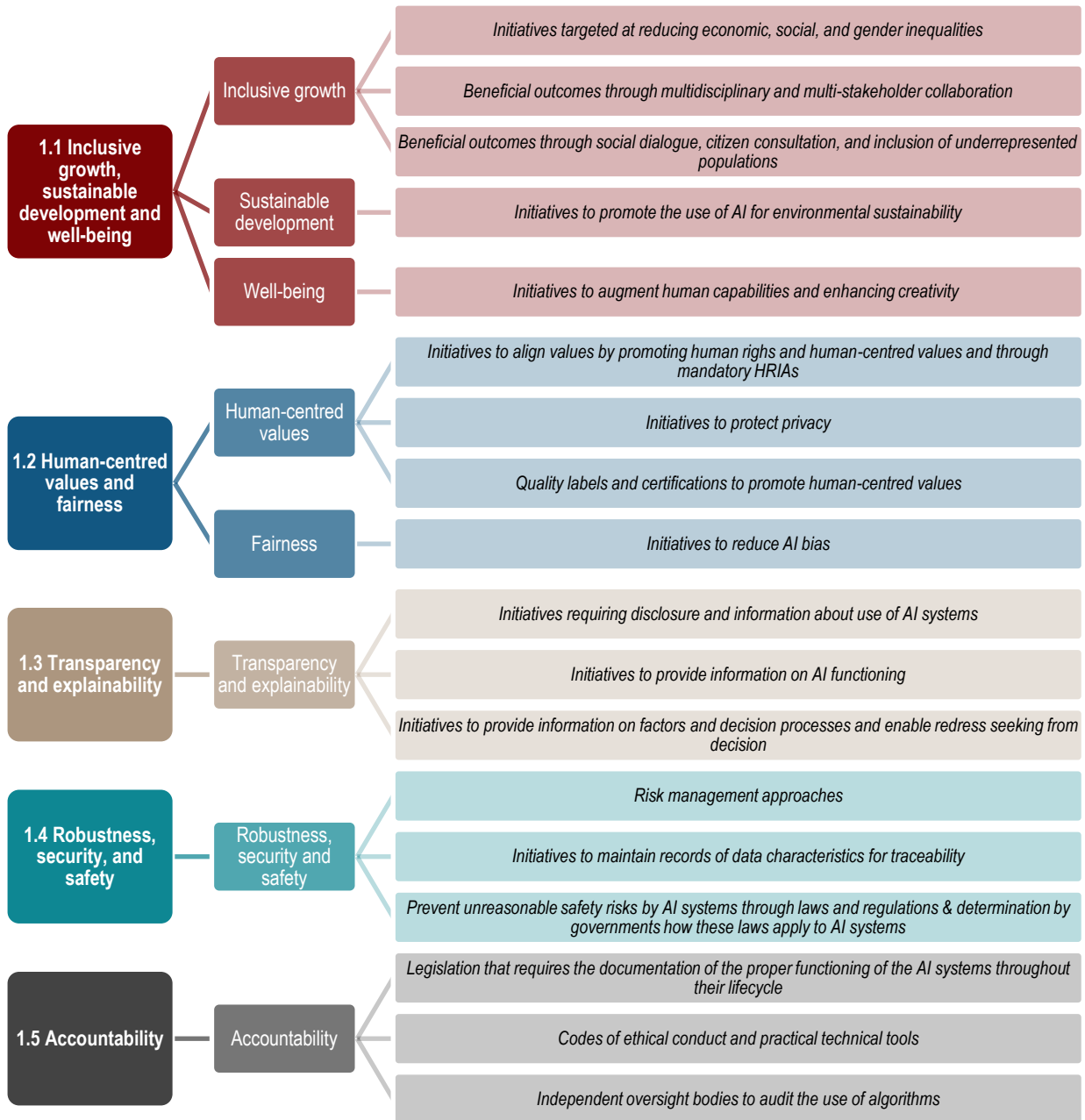
Inclusive growth, sustainable development, and well-being (Principle 1.1)

> *"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."*

AI has the potential to increase countries' productivity and lead to economic growth. Most countries have recognised this and are trying to boost AI research and development (R&D), infrastructure, capacities, and tools through diverse initiatives. However, AI systems could 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. This principle therefore calls for countries to steer AI development, deployment, and use in a way that empowers all members of society. If AI is not driven towards societal benefit at large and AI policies are not developed in an inclusive way, there is a risk that economic growth is unequal and endangers the environment (OECD, 2023[66]).

National approaches to implementing principle 1.1 Inclusive growth, sustainable development and well-being

To achieve inclusive growth, sustainable development and well-being, governments are pursuing different approaches. Most national AI strategies and AI ethics frameworks or guidelines for the implementation of AI refer to this principle. At the policy level, countries have launched 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. Governments are also funding projects that use AI to address environmental challenges (Figure 3.2).

Figure 3.2. Select national policies that implement OECD AI Principle 1.1 on inclusive growth, sustainable development and well-being

Reducing economic, social and gender inequalities	• France's "IA Booster" • United Kingdom's "Women in AI and Data Science"
Multidisciplinary and multi-stakeholder collaboration	• Canada's "Quebec AI Forum" • Colombia's "Coordination Bodies for AI Policy Implementation"
Social diaogue, citizen consultation and inclusion of underrepresented populations	• Chile's "Participation Process on AI" • United States' "AIM-AHEAD Program"
Promote the use of AI for environmental sustainability	• Germany's "AI Lighthouses for the Environment, Climate, Nature and Resources" • European Commission's "Destination Earth (DestinE)"
Augment human capabilities and enhancing creativity	• Germany's "AUTONOM - Performing Arts and AI" • Türkiye's "Breast Cancer Detection with AI"

Inclusive growth

Initiatives targeted at reducing economic, social, and gender inequalities

Australia is addressing inequalities across society through progressing practical initiatives that are underpinned by Australia's AI Ethical Principles (which are built upon OECD values). Programmes such as the National Artificial Intelligence Centre (NAIC) provide a country-level perspective on Australia's AI capability and capacity and support the uplift in AI skills across the nation. The NAIC will be supported by a network of Responsible AI Centres that will be located throughout the country, providing a flow of contemporary knowledge and information from industry leaders directly to business to grow ethical AI principals at a grassroots level. In building this pipeline of capacity, both programs aim to address how Responsible AI can be directly used to have positive impact on significant regional, social, economic, and gender inequalities.

France's IA Booster's mission is to reduce social and economic inequalities by supporting small- and medium-sized enterprises (SMEs) by accelerating the digitalisation of their activities through AI solutions. The programme is tailored to meeting the specific needs of each company and provides support throughout the entire transformation process, from the initial audit phase to selecting and implementing the appropriate solution. The programme also considers the evolution of workstations, job roles, and necessary skills to ensure a seamless transition (French Government, 2021[61]).

In the **United Kingdom**, the Alan Turing Institute's public policy programme's Women in Data Science and AI initiative provides a notable example of reducing gender inequalities. They collaborate with policy makers and industry stakeholders to provide practical insights and recommendations to address various ethical, economic, and governance-related issues arising from AI inequalities. Their approach involves three tiers: Firstly, they map the participation of women in data science and AI both in the UK and globally, with the goal of increasing the number of women in these fields; secondly, they examine diversity and inclusion in online and physical workplace cultures; lastly, they explore how the gender gap affects scientific knowledge and technological innovation while promoting responsible, gender-inclusive AI design (The Alan Turing Institute, 2023[62]).

Beneficial outcomes through multidisciplinary and multi-stakeholder collaboration

Canada's Quebec AI Forum uses AI as a lever for the economic and social development of Quebec. It follows a collaborative approach. In doing so, it rallies and mobilises a wide array of stakeholders around common projects and carries out monitoring and strategic thinking activities. The Forum works on several projects to increase the international competitiveness of Quebec's AI solutions providers and to promote the responsible adoption of AI by all organisations, especially SMEs. Moreover, it supports government stakeholders in their efforts to adopt AI in an ethical and socially responsible manner (Forum IA Quebec, 2023[63]).

Colombia's Coordination Bodies for AI Policy Implementation represent an illustrative example of coordinating AI policies among different stakeholders. These bodies consist of the "Technical Committee for Transformation and Digital Economy" and the "Presidential Advisory Office for Digital Transformation", both public entities dedicated to promoting, coordinating, and supporting the implementation of AI policies across national and local public entities in Colombia. Additionally, they facilitate decision-making related to digital transformation and the digital economy and provide guidance for developing the digital ecosystem among public entities, the private sector, academia, and the national government (OECD.AI, 2023[8]).

Germany's initiative "Civic Coding – Innovation Network AI for the Common Good" sets an example for inter-ministerial and multi-stakeholder collaboration in the field of common good-oriented AI. It was created as a joint effort by the Federal Ministry of Labour and Social Affairs (BMAS), the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ), and the Federal Ministry for the Environment,

Nature Conservation, Nuclear Safety and Consumer Protection (BMUV). The network strengthens the data and AI competencies of civil society, enables interdepartmental support programmes and measures, and thus promote the societal use of AI for the common good (Civic Coding, 2018[64]).

Since February 2022, **Korea** has been operating a multi-stakeholder AI Ethics Policy Forum (with participants from industry, academia, education circles, legal circles, and civil society) to discuss ethical issues induced by the development of AI technologies and to form a social consensus on how to build trust in AI. There are three expert committees (ethics, technology, education) within the forum to facilitate the consensus building. The forum also develops practical tools such as a self-assessment checklist for AI Ethics Guidelines and Trustworthy AI Development Guidelines.

Beneficial outcomes through social dialogue, citizen consultations and inclusion of underrepresented populations

The **Austrian** National AI Strategy was developed and implemented with the involvement of more than 160 experts from a variety of disciplines (technology, economics, natural sciences to law, social sciences, or educational sciences) and civil society organisations. These organisations include social partners such as the Chamber of Labour and the Chamber of Commerce, which have historically played an important role in Austrian policymaking. Mobilising civil society organisations in the design of AI policies can be seen as a positive continuation of traditional policy design patterns and as being inclusive of the interests of diverse social groups (Austrian Government, 2021[66]).

In 2020, **Canada**'s Advisory Council on AI launched the Public Awareness Working Group to explore public awareness of and trust in AI. The Working Group partnered with the Canadian Institute for Advanced Research and Algora Lab to conduct virtual workshops across Canada as part of the Open Dialogue: AI in Canada initiative. The workshops engaged the public in discussions about AI, its potential uses, and associated risks to achieve regional representation and inclusivity for marginalized populations and youth (Government of Canada, 2023[67]).

Chile's Participation Process on AI provides a strong example of how citizens' voices can be taken into account in AI policy design. The Science, Technology, Knowledge, and Innovation Ministry launched a process to collect the visions, perceptions, opinions, and concerns of people and organisations regarding the use and development of AI in Chile. This process fosters discussions about AI opportunities and challenges in Chile and diffuses AI-related knowledge (OECD.AI, 2023[8]).

Mexico's National Alliance of Artificial Intelligence (ANIA) was launched on April 21, 2023, by the Mexican Senate as an inclusive, open, plural, and objective space for the analysis of advances, opportunities, challenges, and risks of the use of AI in different sectors and by different individuals, organisations, and society in general. It seeks to recognize and strengthen the AI ecosystem in Mexico and maintain an open dialogue on AI and its impacts. It uses a comprehensive, pluralistic, and multidisciplinary perspective, including the participation of different stakeholders and society in general. Its main objective is to deploy AI for the benefit of humanity and as a transversal axis for the sustainable development of Mexico. Furthermore, ANIA operates on the following principles: democratising AI discussions, engaging multiple stakeholders, being pluralistic and non-partisan, encouraging collaboration and shared responsibility, adopting a multidisciplinary and transdisciplinary approach, and promoting a multisectoral vision for the present and future of AI.

Scotland's AI Alliance commissioned the Democratic Society to develop AI Co-Creation Public Engagement Workshops in Edinburgh, Inverness, and online. The programme developed design principles and a participatory decision tree on how people in Scotland should engage in future AI decision making. Participants did not need to have any knowledge about AI nor have AI skills (Democratic Society, 2022[65]).

In 2021, the **United Kingdom**'s Alan Turing Institute conducted the AI Ecosystem Online Survey in collaboration with the AI Council to gather the perspectives of individuals involved in the AI ecosystem,

including researchers, developers, and users of AI technologies. The survey aimed to inform the Office for AI's development of the National AI Strategy. With over 400 respondents, the survey stands as an excellent example of incorporating a broad range of diverse voices into AI policy design (OECD.AI, 2023[8]).

The **United States**' National Institutes of Health's AI/Machine Learning Consortium to Advance Health Equity and Researcher Diversity's (AIM-AHEAD) objective is to enhance diversity in the development of AI/ML models, and to improve their capacities to address health disparities and inequities. The AIM-AHEAD Coordinating Centre represents the heart of the AIM-HEAD consortium and consists of institutions and organisations dedicated to serving underrepresented and underserved groups affected by health disparities. The centre consists of four cores: (1) the Administration/Leadership Core, which is responsible for leading, recruiting, and coordinating the AIM-AHEAD Consortium; (2) the Data Science Training Core, which assesses, develops, and implements data science training curricula; (3) the Data and Research Core, which prioritizes and addresses research needs to create an inclusive basis for AI/ML; and (4) the Infrastructure Core, which evaluates data, computing, and software infrastructure to facilitate AI/ML and health disparities research (OECD.AI, 2023[8]).

Sustainable development

Initiatives to promote the use of AI for environmental sustainability

Since 2019, the AI Lighthouses for the Environment, Climate, Nature and Resources initiative, funded by the **German** Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMUV), has been supporting efforts to leverage AI to address environmental challenges and promote sustainable digitalisation. The initiative focuses on two main areas. First, the development of "AI innovations for climate protection" aims at reducing greenhouse gas emissions and adapting to rapid climate change. Second, supporting projects seek to reduce energy and resource consumption of AI systems and their infrastructure. The funded projects range from smart grid control for the energy transition to urban climate adaptation and AI-optimized rail transport. Other projects focus on mitigating nitrate pollution in groundwater or improving waste sorting (German Government, n.d.[68]).

Portugal has been using AI to promote environmental sustainability. For example, it deploys AI to fight against illegal fishing by identifying marine areas with the highest fish population, quantity and quality, and by measuring which fish species are most abundant in specific marine zones. Furthermore, Portugal has also been developing algorithms that enable the analysis and control of waste management. With these initiatives, Portugal contributes to the protection of biodiversity and the ecosystem.

The **European Commission**'s Destination Earth (DestinE) initiative represents another illustrative example of how AI can be leveraged to fight climate change. DestinE is a flagship initiative that develops a highly accurate digital model or "digital twin" of the Earth. By utilizing advanced observation and simulation capabilities powered by Europe's HPC computers and AI capacity, DestinE will enhance preparedness for natural disasters, climate change adaptation, and socioeconomic impact prediction. It consolidates access to valuable sources of data across Europe and allows non-scientific experts to access and interact with vast amounts of system and socio-economic data. By generating highly accurate simulations of the Earth, DestinE will support both EU policymaking and the practical implementation thereof. For example, it will contribute to achieving the objectives stated in the European Commission's Green Deal and its Digital Strategy, and therefore help to align the green and digital twin transition (OECD.AI, 2023[8]).

Well-being

Initiatives to augment human capabilities and enhance creativity

The AUTONOM - Performing Arts and AI programme, offered by the **German** Performing Arts Fund, exemplifies how the intersection of AI and art can enhance creativity. As one of the six federal cultural promotion funds, the German Performing Arts Fund focuses on promoting the liberal performing arts and the cultural landscape of the Federal Republic. It offers three general funding programs with four application periods each year to facilitate the targeted and professional development of artistic concepts. In addition, the fund offers special programmes that address topics such as diversity, digitization, art in rural areas, and the development of aesthetic-artistic formats. In doing so, it provides opportunities for qualification and further development. Participants are expected to explore various questions, including whether AI is a catalyst for innovation or a frightening portrayal of dystopian scenarios, whether it renders artists obsolete, who will determine what is presented and expressed in theatre spaces, and whether art and artists will continue to be autonomous in the future (Fond Darstellende Künste, 2021[69]).

Portugal's analysis and prediction of patterns in the utilisation of emergency and medical services is an example of how AI can enhance human capabilities. It proposes the use of microdata from medication prescriptions by healthcare professionals from healthcare institutions and medical services. Analysing the relationship between prescription patterns at this level of detail can serve as a proxy for predicting the utilisation of these services. In doing so, the initiative renders the management of healthcare entities in the National Health Service (SNS) more efficient.

Türkiye's Breast Cancer Detection with AI project represents another example of increasing human capabilities through AI. Radiologists use an AI labelling tool developed by the Presidency Digital Transformation Office to identify benign and malignant anomalies in mammography images. This process reduces the likelihood of errors, particularly in mammography screening, and enables radiologists to prioritize images detected as high-risk by AI. The project hence serves as a decision support system, allowing radiologists in Türkiye to work more efficiently and effectively while also freeing up time and enhancing their capabilities. Ultimately, the project has the potential to save lives by enabling early diagnosis of breast cancer (OECD.AI, 2023[8]).

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."

Some uses of AI systems have implications for human rights, including risks that (as defined in the Universal Declaration of Human Rights) human-centred values can be deliberately or accidentally infringed upon. To avoid this, countries are implementing human-centred values and fairness through policy initiatives (OECD, 2023[66]).

National approaches to implementing principle 1.2 Human-centred values and fairness

To implement principle 1.2, governments 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. By contrast, only very few Human Rights Impact Assessments (HRIA) and quality seals have been developed to date (Figure 3.3).

Figure 3.3. Select policies that implement OECD AI Principle 1.2 on human-centred values and fairness

Initiatives to align values by promoting human rights and human-centred values and through mandatory HRIAs	• France's "Addressing Human Rights Concerns Arising From Facial Recognition Technology" • The Netherland's "Fundamental Right and Algorithmic Impact Assessment"
Initiatives to protect privacy	• Korea's "AI Personal Information Protection Self-Checklist" • Mexico's "Recommendations for the Processing of Personal Data derived from the use of AI"
Quality labels and certifications to promote human-centred values	• Germany's "AI Seal of Quality" • Türkiye's "Trustworthy AI Stamp"
Initiatives to reduce AI bias	• United Kingdom's "Review into Bias in Algorithmic Decision-Making" • United States' "Artificial Intelligence and Algorithmic Fairness Initiative"

Human-centred values

Initiatives to align values by promoting human rights and human-centred values and through mandatory human rights impact assessments (HRIAs)

National AI policies often underline the significance of human-centred AI and reference one or several human rights. In this regard, the right to privacy is most frequently raised in national AI strategies, possibly because of its interconnectivity with other human rights, such as freedom of expression, association, and personal autonomy. This is followed by the right to equality/non-discrimination and the right to an effective remedy. Merely referencing human rights in a National AI Strategy does not guarantee, however, that a government's approach and actions will effectively ensure respect for human rights in practice (Stanford University, 2021[70]).

In addition, human-centred AI and the adherence to human rights is stated as an important guiding principle in many national and international AI ethical guidelines. However, in some cases, guidelines and tools for operationalizing human-centred values and human rights are also being developed for more specific AI use cases. For example, with "Addressing human rights concerns arising from facial recognition technology", **France** and the **World Economic Forum** are co-designing a policy framework to deal with human rights concerns arising from the use of facial recognition technology. This framework will establish a set of principles for responsible use of the technology. Furthermore, it will also comprise an assessment questionnaire that applies these principles to specific use cases, allowing organisations to evaluate their risk mitigation strategies. Furthermore, it will contain an audit framework to ensure compliance with the principles for action (OECD.AI, 2023[8]).

The **United States'** State Department "Guidance on products or services with surveillance capabilities" is a tool designed to offer practical and accessible human rights guidance to US businesses. Its goal is to help these businesses prevent their products or services with surveillance capabilities from misuse by government end-users to commit human rights violations. The tool is aligned with the UN Guiding Principles on Business and Human Rights, as well as with the OECD Guidelines for assessing the human rights impacts of relevant products or services. It is meant to be an easy-to-use roadmap for assessing the human rights impacts of relevant products or services and for evaluating a series of considerations before engaging in transactions with governments. It also recommends human rights safeguards, such as creating

a grievance mechanism and publicly reporting on sales practices, if a U.S. business decides to proceed with a transaction (OECD.AI, 2023[8]).

While OECD countries widely acknowledge that AI should be human-centred and respect human rights in their national AI strategies, countries have yet to implement this in a practical and legally binding way. National regulations do not seem to address significant risks that AI systems pose to fundamental human rights beyond discriminatory biases.

To date, only the **Netherlands** has introduced mandatory HRIA for the use of algorithms within Dutch public authorities. The Netherlands' Ministry of Interior and Kingdom Relations has created a Fundamental Rights and Algorithms Impact Assessment (FRAIA), which facilitates an interdisciplinary dialogue to help map the risks to human rights from the use of algorithms and determine measures to address these risks. This example is, however, a unicum across countries (Dutch Government, 2021[71]).

At the international level, the **Council of Europe**'s Recommendation CM/REC(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems suggests that "States should ensure that they, as well as any private actors engaged to work with them or on their behalf, regularly and consultatively conduct human rights impact assessments prior to public procurement, during development, at regular milestones, and throughout their context-specific deployment in order to identify the risks of rights-adverse outcomes" (Council of Europe, 2020[72]). The CoE's Committee on Artificial Intelligence (CAI) is currently working on a "Framework Convention Artificial Intelligence, Human Rights, Democracy and the Rule of Law", aimed at developing common principles to ensure respect for human rights, democracy, and the rule of law in a context where AI systems assist or replace human decision-making (Council of Europe, 2023[79]).

Initiatives to protect privacy

Countries are implementing various policies to protect privacy in AI, including through regulatory sandboxes to promote the development of privacy-friendly use of AI solutions (Figure 2.1). The **Korean** government issued guidance relative to development of AI in its "AI Personal Information Protection Self-Checklist" in May 2021. 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), an autonomous constitutional entity, developed "Recommendations for the Processing of Personal Data derived from the use of AI" (INAI, 2022[81]) as well as two instruments with the Ibero-American Data Network (Ibero-American Data Protection Network, 2019[82]); (Ibero-American Data Protection Network, 2019[83]). Several countries (e.g. Estonia, Turkey, the United Kingdom, United States) are also promoting Privacy Enhancing Technologies (PETs) to prevent privacy infringement in developing or operating AI technologies and services (OECD, 2023[82]).

Quality labels and certifications to promote human-centred values

So far, only a few countries have developed quality labels and certifications confirming that an AI tool is ethical and human-centred.

The **German** AI Association has created the AI seal of quality aimed at promoting the use of human-centred and human-serving AI. The seal of quality enforces a shared set of values and processes, thus ensuring that services and products developed using it are ethically compatible. Its key quality criteria comprise ethics, impartiality, transparency, security, and data protection. Each criterion has defined measures that must be met (KI Bundesverband, 2019[73]).

In 2019, **Malta**'s Digital Innovation Authority (MDIA) launched the AI Certification Programme. It consists of a set of guidance notes that assist Service Providers and AI Innovative Technology Arrangements (ITA) applicants when approaching the MDIA for registration and certification. The objectives are to expand the

ITA certification framework for AI-based solutions and to become the world's first national AI certification programme developed in an ethical, transparent, and socially responsible manner (OECD.AI, 2023[8]).

Similarly, **Türkiye** is in the process of developing a Trustworthy AI Trust Stamp that signals the use of trustworthy AI in products, thereby increasing customer trust. The Turkish Standards Institute is currently engaged in research to establish criteria and metrics for the stamp.

Fairness

Initiatives to reduce AI bias

France's HPC Support to NLP Bigscience Workshop is an international collaboration of over 800 researchers working on large multilingual language models and datasets for a year. France provides the HPC support for the training of the workshop's models, along with scientific contributions from the French National Institute for Research in Digital Science and Technology (INRIA) and the National Centre for Scientific Research (CNRS). The workshop has several objectives, including offering advanced large language models and datasets for French and other EU languages, and addressing legal and ethical issues related to ownership and storage of large datasets. Most importantly for this section, the project also analyses biases within and ethical problems regarding the language models and proposes metrics and tools for their evaluation and mitigation (OECD.AI, 2023[8]).

In 2021, the **Netherlands**' Ministry of the Interior and Kingdom Relations commissioned an investigation into the causes of discrimination in AI. A team of experts from Tilburg University, Eindhoven University of Technology, Vrije Universiteit Brussel, and The Netherlands Institute for Human Rights collaborated on creating the guideline "Non-discrimination by design". The guideline aims at explaining how organisations can prevent their systems from being discriminatory. It discusses the technical, legal, and organisational conditions that should be applied before, during and after creating an AI system (Dutch Ministry of Internal Affairs, 2021[74]).

In 2020, the **United Kingdom**'s Centre for Data Ethics and Innovation (CDEI) published its Review into Bias in Algorithmic Decision-Making. The review analyses the influence of the growing utilisation of algorithmic tools on decision-making bias, the measures necessary to mitigate risks, and the potential for improving fairness through better data utilisation. It focuses on significant decisions made by algorithms about individuals in four sectors: recruitment, financial services, policing, and local government. The review also presents overarching suggestions to create appropriate systems that enhance, rather than undermine, decision-making through algorithms (UK Government, 2020[75]).

The **United States**' Equal Employment Opportunity Commission has launched the agency-wide AI and Algorithmic Fairness Initiative to ensure that the use of AI complies with American anti-discrimination laws. The initiative has several objectives, i.e., offering technical guidance on algorithmic fairness and AI's use in employment decisions, identifying effective practices, hosting listening sessions with significant stakeholders about algorithmic tools and their employment implications, and collecting data on the adoption, development, and effect of hiring and other employment-based technologies (OECD.AI, 2023[8]). Furthermore, the Executive Order on Advancing Racial Equity and Support for Underserved Communities through the Federal Government of 2023 includes an AI clause, which places new equity obligations on federal agencies that deploy AI systems. It directs agencies to "prevent and remedy discrimination, including by protecting the public from algorithmic discrimination" (OECD.AI, 2023[8]) .

In 2021, the **United States**' Federal Trade Commission (FTC) designated eight key areas of focus for enforcement and regulatory action, one of which directly implicates investigations in unfair, deceptive, anticompetitive, collusive, coercive, predatory, exploitative, and exclusionary acts or practices relating to algorithms and biometrics. The FCT will "investigate whether any persons, partnerships, corporations, or others have engaged or are engaging in unfair, deceptive, anticompetitive, collusive, coercive, predatory,

exploitative, or exclusionary acts or practices relating to algorithms and biometrics, in or affecting commerce, including but not limited to bias in algorithms and biometrics, in violation of Section 5 of the Federal Trade Commission Act, 15 U.S.C. § 45, as amended or any statutes or rules enforced by the Commission" (OECD.AI, 2023[8]). It will then determine the appropriate action or remedy, including whether injunctive and monetary relief would be in the public interest.
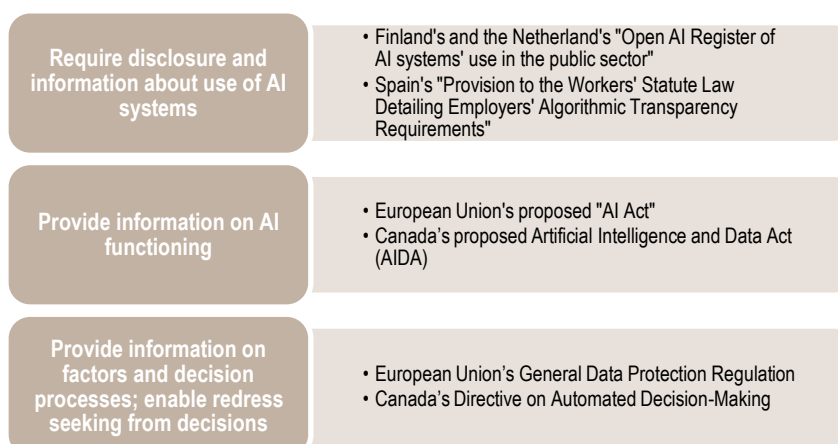
Transparency and explainability (Principle 1.3)

> *"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."*

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 need for transparent and explainable AI, operationalising these concepts is complex due to their multifaceted nuances. AI transparency entails: i) clearly communicating to users that they are dealing with an AI system, ii) the interpretability of decision-making processes, and iii) the explainability of decision-making logic.

National approaches to implementing principle 1.3 transparency and explainability

Governments are taking a variety of approaches to ensure AI transparency, ranging from guidelines for implementation of AI to the establishment of oversight bodies. Regulatory bodies have recognised the importance of AI transparency and explainability. 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, governments are enhancing transparency around the use of AI for public services, for example through AI registers (Figure 3.4).

Figure 3.4. Select policies that implement OECD AI Principle 1.3 on transparency and explainability

Require disclosure and information about use of AI systems	• Finland's and the Netherland's "Open AI Register of AI systems' use in the public sector" • Spain's "Provision to the Workers' Statute Law Detailing Employers' Algorithmic Transparency Requirements"
Provide information on AI functioning	• European Union's proposed "AI Act" • Canada's proposed Artificial Intelligence and Data Act (AIDA)
Provide information on factors and decision processes; enable redress seeking from decisions	• European Union's General Data Protection Regulation • Canada's Directive on Automated Decision-Making

Initiatives requiring disclosure and information about use of AI systems

Japan has introduced transparency requirements in the Digital Platform Transparency Act (DP Transparency Act) by requiring designated digital platform providers (online malls, app stores, and digital advertising businesses) to ensure transparency and fairness in transactions with business users (Habuka, 2023[76]). Other transparency requirements are included in soft law, such as the AI utilisation guidelines, the contract guidelines on utilising AI and data, and machine learning quality management guidelines.

The proposed **EU** AI Act includes transparency obligations for high-risk AI systems (Article 13), as well as those that meet at least one of the criteria below (Article 52):

- intended to interact with natural persons (e.g., chatbots);
- used for emotion recognition;
- used for biometric categorization; or
- used to generate or manipulate image, audio or video content.

The transparency obligations arising under Article 52 concern informing users that they are interacting with an AI system.

The **EU**'s Digital Services Act (DSA) (passed in July 2022) also includes requirements for enhanced transparency of algorithms. The DSA requires Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) operating in the EU to identify, analyse, and assess certain systemic risks stemming from the design and functioning of their service and related systems, including algorithmic systems. Moreover, they must commit to addressing identified risks, whether directly or indirectly related to the functioning of the algorithmic system in use (European Union, 2022[77]). The European Commission also set up the European Centre for Algorithmic Transparency (ECAT) to support its supervisory role with in-house and external multidisciplinary knowledge. The Centre, hosted by the Joint Research Centre (JRC) in close co-operation with the Directorate General Communications Networks, Content and Technology (DG CONNECT), will support the regulator in assessing whether the functioning of algorithmic systems is in line with the risk management obligations established by the DSA for VLOPs and VLOSEs (OECD.AI, 2023[8]).

As governments are increasingly integrating AI into the design and delivery of public policies and services, they are also increasing efforts to enhance transparency and explainability.

In **France**, the 2016 Digital Republic Law mandates transparency of government-used algorithms. Public agencies are required to publicly list any algorithmic tools they use and to publish their rules. Etalab, a department of the Inter-ministerial Digital Direction (DINUM), has provided guidance on the implementation of this commitment by publishing two guidance documents: one shows how to open public source codes, and the other explains the legal framework of accountability and transparency of public sector algorithms (OECD.AI, 2023[8]). Helsinki (**Finland**) and Amsterdam (**Netherlands**) have launched open AI registers that track how algorithms are being used in the municipalities. Following this example, nine European cities (Barcelona, Bologna, Brussels Capital Region, Eindhoven, Mannheim, Rotterdam, and Sofia) have collaborated in 2023 through the Eurocities Digital Forum Lab network to develop an AI algorithm registers standard (OECD.AI, 2023[8]). At the end of 2022, the **United Kingdom** government published the Algorithmic Transparency Recording Standard, which comprehensively organizes how the public sector, including government, should disclose information when using algorithmic tools. The United Kingdom, through The Alan Turing Institute, has also created an AI Standards Hub to advance trustworthy AI through standards. The Algorithmic Transparency Standard is one example of this (OECD.AI, 2023[8]).

Various approaches are being also employed to promote the transparent utilisation of AI in the workplace. These comprise the reliance on existing policies as well as the development of self- and co-regulation approaches and new polices (Salvi del Pero, Wyckoff and Vourc'h, 2022[78]).

Examples include the requirement to prior agreement with workers' representatives on the monitoring of workers using digital technologies (e.g., **France, Germany** and **Italy**), and regulations requiring employers to notify employees about electronic employee monitoring policies. For example, such regulations have been adopted by the **Canadian** Province of Ontario in 2022 (Ontario Working for Workers Act). Similarly, in the **United States**, a number of states have laws in place that require employers to notify employees of electronic monitoring. New York's Electronic Monitoring Bill (2022) imposes compliance obligations on New York City employers using AI tools to notify employees of their electronic monitoring practices. Illinois' AI Video Interview Act and Maryland Facial Recognition Law require employers to disclose the use of AI analysis in video interviews. In the **United Kingdom**, the Information Commissioner's Office issued guidance in October 2022 on employers' legal obligations when monitoring workers (Monitoring at Work Draft Guidance).

Initiatives to provide information on AI functioning

Canada's proposed Artificial Intelligence and Data Act (AIDA) introduces requirements to promote transparency on the use of AI. Transparency means providing the public with appropriate information about how high-impact AI systems are being used. The act stipulates in Article 11 that where the system is made available for use, the person responsible must publish on a publicly available website a plain-language description of the system that explains how the system is to be used, the types of content that it is intended to generate, and the types of decisions, recommendations, or predictions it is intended to make, along with the risk mitigation measures established. The information provided should be sufficient to allow the public to understand the capabilities, limitations, and potential impacts of the systems (Government of Canada, 2023[79]).

Spain's Royal Decree-Law 9/2021 (the "Rider Law") modifies the Workers' Statute Law detailing employers' algorithmic transparency requirements. The legislation renders transparency mandatory for AI systems that make decisions about or influence working conditions or employment status. It requires employers using algorithmic decision-making to disclose the key attributes of algorithms to employee representatives. The "attributes include the algorithm's parameters and general logic used to make decisions, a significant leap in regulation beyond the GDPR" (Chavez, Bahr and Vartanian, 2022[80]).

The **United States**' proposed Algorithmic Accountability Act (AAA) of 2022 includes transparency requirements for companies employing Automated Decision Systems (ADS) to make critical decisions, i.e., any decision that has significant legal or material effects on a consumer's life. This includes access to education, employment, essential utilities, healthcare, and financial services (Section 2.7). The AAA also requires organisations deploying new ADS to "describe the existing decision-making process [and] explain the intended benefits of augmenting [it]" (Section 4) (US Congress, 2022[28]).

The proposed **EU** AI Act includes transparency obligations for high-risk AI systems (Article 13). The Regulation prescribes information requirements which would allow users to interpret a system's output and use it appropriately (European Commission, 2021a[33]).

The **Chinese** regulation on algorithmic recommendation systems, which entered into force in March 2022, focuses on the use and impact of algorithmic recommendation systems. It creates transparency obligations which entail user notifications regarding the criteria for recommendation and clear indicators of algorithmically generated or synthetic information. It orders the implementation of mechanisms of manual intervention and autonomous user choice. The regulation also mentions the creation of a registry and categorisation system to manage algorithms placed on the market (OECD.AI, 2023[8]).

Initiatives to provide information on factors and decision processes and enable redress seeking from decisions

Canada's Consumer Privacy Protection Act includes transparency requirements regarding "the organization's use of any automated decision system to make predictions, recommendations or decisions about individuals that could have a significant impact on them" (Article 62). It also features a "right to explanation" in Article 63 (2): "If the organization has used an automated decision system to make a prediction, recommendation or decision about the individual that could have a significant impact on them, the organization must, on request by the individual, provide them with an explanation of the prediction, recommendation or decision" (Canadian Parliament, 2022[12]).

Canada's Directive on Automated Decision-Making sets a wide range of mandatory requirements to ensure the responsible use of AI by federal institutions. The Directive applies to systems used to make decisions affecting legal rights or carry out assessments about clients to inform these decisions. The requirements mandate completion and publication of an algorithmic impact assessment; adoption of several transparency measures such as notice and explanation to clients; application of quality assurance measures such as bias testing, peer review, and ongoing monitoring of outcomes; provision of recourse to subjects of automated administrative decisions; and public reporting on system effectiveness and efficiency (OECD.AI, 2023[8]).

Mexico's legal framework in matters of personal data protection, applicable to the public and private sector, also contains provisions related to automated decisions.

Norway's Public Administration Act states that public sector decision-making pertaining to specific individuals must provide explanations for the decisions taken in order to ensure accountability and support a complaints/appeals process. There is also a requirement for accountability/transparency and equal treatment. This law is the legal foundation for all types of casework systems in the public sector, including systems that use AI/machine learning.

In the **United Kingdom**, the proposed Data Protection and Digital Information Bill (No.2) introduced in the House of Commons on March 28, 2023, introduces a more business-friendly approach in comparison to the existing UK General Data Protection Regulation (UK GDPR). Whereas the UK GDPR generally prohibits automated profiling decision-making, the proposed bill allows it by default, placing on the affected individuals the responsibility to challenge such decisions and request a human review (UK Parliament, 2023[30]).

The **EU**'s General Data Protection Regulation (GDPR), which entered into force in 2018, implies a "right to explanation" in Article 22 by giving individuals the right not to be subjected to "a decision based solely on automated processing, including profiling, in cases where they produce legal or similarly significant effects affecting the data subject" (European Union, 2018[84]). A number of cases have been brought to EU Courts on the use of AI in the workplace, based on the legal rights accorded by the GDPR (Salvi del Pero, Wyckoff and Vourc'h, 2022[78]).

Robustness, security, and safety (Principle 1.4)

> *"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.*
>
> *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.*
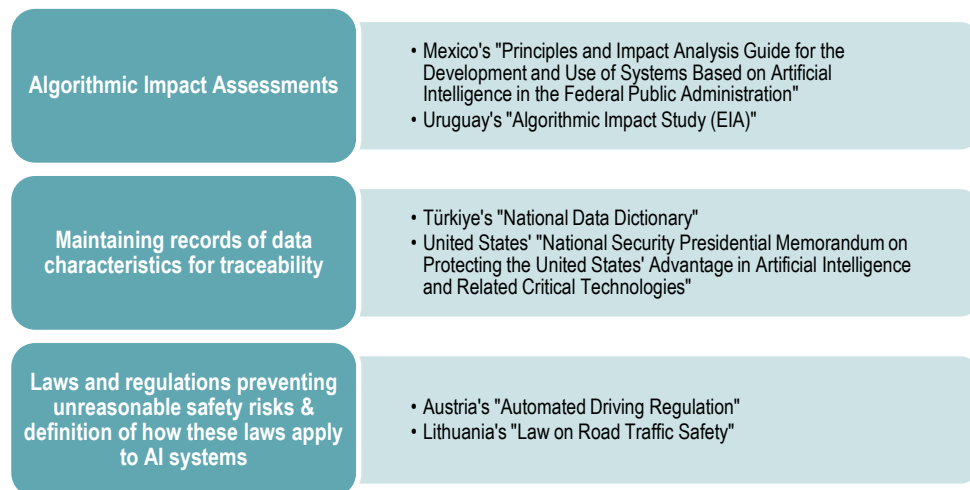
> *AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias."*

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 (OECD, 2023[66]). Therefore, they are analysed together. However, there are different, not mutually exclusive ways in which countries can operationalise them.

National approaches to implementing Principle 1.4 on robustness, security and safety

Countries are drawing on guidelines, ethics frameworks, impact assessments, new legislation, amendments to existing legislation, and other instruments to implement Principle 1.4 (Figure 3.5).

Figure 3.5. Select policies that implement OECD AI Principle 1.4 on robustness, security and safety

Algorithmic Impact Assessments	• Mexico's "Principles and Impact Analysis Guide for the Development and Use of Systems Based on Artificial Intelligence in the Federal Public Administration" • Uruguay's "Algorithmic Impact Study (EIA)"
Maintaining records of data characteristics for traceability	• Türkiye's "National Data Dictionary" • United States' "National Security Presidential Memorandum on Protecting the United States' Advantage in Artificial Intelligence and Related Critical Technologies"
Laws and regulations preventing unreasonable safety risks & definition of how these laws apply to AI systems	• Austria's "Automated Driving Regulation" • Lithuania's "Law on Road Traffic Safety"

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. The AIA helps evaluate and anticipate the potential social, economic, ethical, and legal effects of AI systems before they are implemented and throughout their lifecycle, by providing a framework for identifying and mitigating risks. The Canadian AIA was developed in consultation with academia, civil society, and other public institutions to help departments and agencies understand and manage the risks associated with automated decision systems. The AIA consists of over 80 questions in different formats aimed at evaluating six areas of risk, i.e., risks related to the project, system, algorithm, decision, impact, and data, and assessing mitigation measures in place to manage identified risks. The AIA evaluates the impact of automation projects based on the responses to the questions, which are assigned weights that are used to calculate raw impact and mitigation scores. A final score is produced based on these scores, which is used to classify the automated decision system into one of four possible impact levels (little to no impact; moderate impact; high impact; and very high impact). The level of impact determines applicable requirements under the Directive (Canadian Government, 2023[85]).

Mexico's Principles and Impact Analysis Guide for the Development and Use of Systems Based on AI in the Federal Public Administration are based on Canada's AIA and are designed to assess the societal and ethical implications of AI systems developed by the Federal Public Administration. It is an online

questionnaire that determines the impact level of an automated decision-system, outlining appropriate safeguards that must be put in place according to the potential impacts of these systems. Each level requires the system to fulfil a different set of requirements before, during, and after its implementation (OECD, 2022[93]).

Similarly, **Uruguay**'s Agency for Electronic Government and the Information and Knowledge Society (AGESIC) developed the Algorithmic Impact Study (EIA) to analyse machine learning-based automated decision support systems. It poses different questions that assess various aspects of AI systems. These include the predicted social impact, an impact evaluation of the automated decision system, the origin of the data used, stakeholders involved, actions to reduce and mitigate the risks of the automated decision system, and procedural fairness. It is primarily aimed at project managers or teams involved in AI projects, and designed to identify crucial aspects of the systems requiring additional attention or treatment. Users of the tool can then share, analyse, and evaluate the results obtained from the questions (OECD, 2022[93]).

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 (Turkish Government, 2019[87]). The **United States**' National Security Presidential Memorandum (NSPM) Protecting the United States Advantage in AI and Related Critical Technologies safeguards the country's advantage in critical technologies, including AI, against foreign adversaries and strategic competitors. One of the primary objectives of the NSPM is to improve access to high-quality and completely traceable Federal data, models, and computing resources (US Presidential Office, 2019[88]).

Laws and regulations preventing unreasonable safety risks of AI systems: autonomous driving

The main field where countries are increasingly implementing laws and regulations pertains to autonomous driving. For example, **Austria**'s Automated Driving Laws and Regulations (32nd Amendment to the Austrian Motor Vehicles Act; Automated Driving Regulation) grants permission for certain automotive functions or tasks to be done by automated systems. It regulates the basic requirements for the testing of in-vehicle drive assistance systems, automated, or networked driving systems and defines the first applications of autonomous vehicles (e.g., autonomous minibuses in urban areas) (OECD.AI, 2023[8]). **Germany**'s Automated Vehicles (AV) Bill in the Road Traffic Act as well as its Act Amending the Road Traffic Act and the Compulsory Insurance Act ("Autonomous Driving Act") represent further examples for the increase in legislation in this sector. They legalise automated vehicles by modifying the current Road Traffic Act and define the requirements for highly and fully automated vehicles to use public roads (OECD.AI, 2023[8]). Similarly, **Denmark**'s Road Directorate, **Japan**'s Legal Regulation of Autonomous Driving Technology, **Lithuania**'s Law on Road Traffic Safety as well as the **United Kingdom**'s Automated and Electric Vehicles Bill all represent new legislations that define the use of self-driving cars on the countries' national roads (OECD.AI, 2023[8]).

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."*

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 (OECD, 2019[65]).
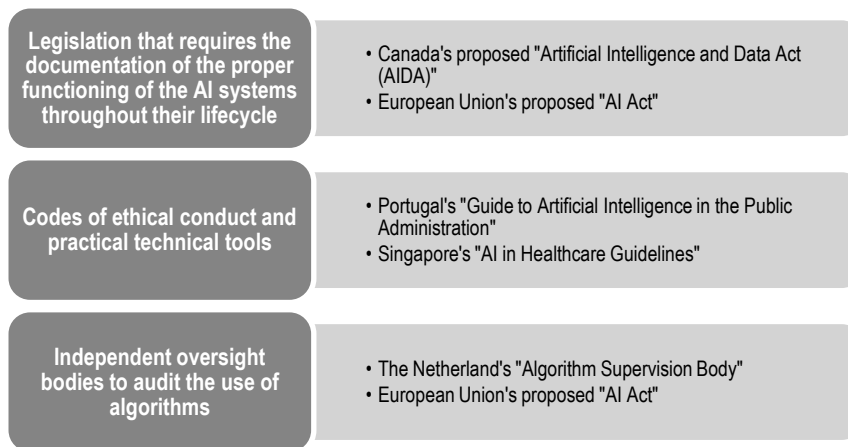
Upcoming specific-AI legislations establish the scope for accountability (both in terms of who should be held responsible and in which circumstances) in relation to the development, deployment, and use of AI systems, which may vary according to the jurisdiction (Annex A).

Demand for tools and processes to document AI system decisions and to ensure accountability is on the rise in both the public and private sectors. This field encompasses major AI standardization initiatives led by organizations 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 proposed AI Act, the UK's AI Standards Hub, the European AI Alliance, the Council of Europe's Committee on Artificial Intelligence (CAI), and the EU-US Trade and Technology Council. Certification schemes are also a part of this landscape (OECD, 2023[66]).

National approaches to implementing Principle 1.5 on accountability

Countries have developed codes of ethical conduct for the use or implementation of AI in several sectors (public administration, health care, autonomous driving). Proposed AI-specific regulation requires the documentation of the proper functioning of the AI systems throughout their lifecycle. Lastly, countries have established independent oversight bodies that audit the use of algorithms (Figure 3.6).

Figure 3.6. Select policies that implement OECD AI Principle 1.5 on accountability

Legislation that requires the documentation of the proper functioning of the AI systems throughout their lifecycle	• Canada's proposed "Artificial Intelligence and Data Act (AIDA)" • European Union's proposed "AI Act"
Codes of ethical conduct and practical technical tools	• Portugal's "Guide to Artificial Intelligence in the Public Administration" • Singapore's "AI in Healthcare Guidelines"
Independent oversight bodies to audit the use of algorithms	• The Netherland's "Algorithm Supervision Body" • European Union's proposed "AI Act"

Legislation that requires documenting the proper functioning of the AI systems throughout their lifecycle

Canada's proposed Artificial Intelligence and Data Act (AIDA) defines accountability as follows: "Accountability means that organisations must put in place governance mechanisms needed to ensure compliance with all legal obligations of high-impact AI systems in the context in which they will be used" (Government of Canada, 2023[79]). To ensure this in practice, policies, processes, and measures implemented as well as ways to meet requirements for design and development, must be proactively documented. Moreover, organisations are required to provide appropriate documentation to users concerning datasets, limitations, and appropriate uses.

The proposed **United States** Algorithmic Accountability Act (AAA) includes requirements on documentation for Automated Decision Systems. Article 4(7) requires operators to "maintain and keep updated documentation of any data or other input information used to develop, test, maintain, or update the automated decision system or augmented critical decision process" (US Congress, 2022[28]).

The **EU**'s proposed AI Act also makes technical documentation obligatory. Article 11 states that the technical documentation of a high-risk AI system must be prepared in a way that proves the system's conformity with the requirements specified in the AI Act. It includes all pertinent information for national competent authorities and notified bodies to evaluate the system's compliance. This documentation must be completed prior to the system's introduction to the market or its utilisation and must be kept up-to-date (European Commission, 2021a[33]).

Codes of ethical conduct and practical technical tools

Codes of ethical conduct and practical technical tools are the most commonly employed instruments by countries to ensure accountability, with particular prominence in the public and healthcare sectors.

France's Etalab can serve as the first example of such codes of ethical conduct in the public sector. Etalab has issued guidance on Accountability for Public Algorithms, which sets out how public organisations should report on their use to promote transparency and accountability. The guidance proposes six principles for the accountability of algorithms in the public sector (OECD.AI, 2023[8]).

Colombia's Dashboard for monitoring the Ethical Framework for AI provides an overview of AI projects by public entities and of how these entities are implementing ethical principles and tools in their projects. By having them report on several key issues, it is an example of holding public entities that use AI accountable. The issues include the ethical principles employed, non-discrimination measures taken, and potential ethical risks, along with mechanisms installed to mitigate them (OECD.AI, 2023[8]).

Portugal has also published a Guide to AI in the Public Administration. The guide outlines the definition AI, highlights its presence in society, and offers insights into the potential effects of its use. In recognition of the critical role data plays in developing and sustaining these systems, the guide also touches upon the data ecosystem within the Public Administration, including the principles that must govern it. These comprise human rights as well as inclusion, equality, sustainable development, and well-being (OECD.AI, 2023[8]).

Norway's Guidance on the development and use of AI in the public sector aims at operationalising the ethics principles from its national AI strategy. It promotes the use and development of AI in the public sector while ensuring this is done following an ethical framework. Another code of ethical conduct can also be found in the **United Kingdom**. Its Guide to Using AI in the Public Sector, published by the Government Digital Service (GDS) and the Office for Artificial Intelligence (OAI), is based on the UK's Data Ethics Framework, which sets out clear principles for how data should be used in the public sector. The "Guide to Using AI in the Public Sector" applies these principles to the context of public procurement (OECD.AI, 2023[8]).

Codes of ethical conduct in the healthcare sector have been established in **France**, among other countries. The French "Good Practice Recommendations to Integrate Ethics in the Development of AI Solutions in Healthcare" provides a comprehensive list of ethical guidelines necessary for AI solutions in healthcare, deploying a two-tier "ethics by design" approach. This firstly entails incorporating ethical values, to the greatest extent feasible, into the hardware and software architecture of the systems during the design phase. Secondly, it involves validating proposed solutions and preventing the risks arising from today's predominantly inductive nature of AI at the end of the development process (OECD.AI, 2023[8]). Likewise, **Singapore** has also implemented AI in Healthcare Guidelines. Co-developed by the Ministry of Health (MOH), the Health Sciences Authority (HSA), and the Integrated Health Information Systems (IHiS), they are to strengthen patient safety and trust of AI in healthcare. Further, the recommendations aim at

encouraging the safe development and implementation of both AI medical devices ("AI-MDs") and any other AI system implemented in healthcare settings (OECD.AI, 2023[8]).

Codes of ethical conduct are also implemented in various other sectors. **Austria**'s Codes of Practice for Testing of Automated Driving on Public Roads can be mentioned as an example for the transport sector. The Codes comprise a set of measures for companies seeking to test automated vehicles (AV), which should minimise risk and maximise safety when AV are deployed on roads (OECD.AI, 2023[8]).

Korea also established National AI Ethics Guidelines in December 2020, which lay out comprehensive basic guidelines that should be followed by all members of society in all stages of AI development and use to realize human-centred AI. For practical implementation of the National AI Ethics Guidelines, Korea announced the Trustworthy AI Implementation Strategy in May 2021 and developed a self-assessment checklist in February 2022. The country also provided detailed guidelines, which consider sector-specific and/or case-specific in April 2023, to help the private sector to voluntarily follow the guidelines and verify their development methods in various fields (e.g., general purpose, health, autonomous driving, public/social services).

The **OECD.AI Expert Group on Risk and Accountability** developed a catalogue of tools and metrics for trustworthy AI (OECD, 2023[89]) to provide an interactive collection of resources for the development and implementation of AI systems that respect human rights and are fair, transparent, explainable, robust, secure, and safe. These tools, mapped to the OECD AI Principles and the phases of the AI system lifecycle, are expected to facilitate accountability in AI, from documenting and monitoring risks to certification and assurance. The **United Kingdom** Centre for Data Ethics and Innovation (CDEI) is also developing a portfolio of use cases and an online searchable repository of AI assurance tools. AI Assurance tools are market-based means of managing AI risks and a complement to regulation that will empower industry to ensure that AI systems meet their regulatory obligations (OECD.AI, 2023[8]). **Singapore**'s A.I. Verify, an AI governance testing framework and toolkit for companies, is another tool that helps companies with transparency. It is a framework and software tool to conduct objective, verifiable tests and to record process checks. Key features include covering crucial international governance frameworks and guidelines, validating companies' claims about AI systems' performance, a single and integrated toolkit for self-testing, and customised testing reports to be available for different groups of stakeholders (OECD.AI, 2023[8]).

Independent oversight bodies to audit the use of algorithms

While voluntary codes of ethical conduct and practical tools to ensure accountability are already widespread, only very few countries have implemented legally-binding, independent oversight bodies to audit the use of algorithms.

In December 2022, **Spain** announced the establishment of the Spanish Agency for the Supervision of AI (*Agencia Española de Supervisión de la Inteligencia Artificial, AESIA*). One of its objectives is to promote responsible, sustainable, and trustworthy AI, while also fostering collaboration and coordination with other national and supranational authorities responsible for AI oversight. (Proteccion Data, 2022[90]). Another independent oversight body is the **Dutch** algorithms supervision unit, located within the Dutch Data Protection Authority (DPA). The unit's objective is to enhance the supervision of algorithms that process personal data. Its responsibilities include monitoring algorithms to promote transparency, prevent discrimination and bias, and identifying and analysing potential risks of these systems (OECD.AI, 2023[8]). In parliament on June 16, 2023, the **Norwegian** Liberal Party proposed the establishment of an Algorithmic Supervision Authority. They also urged the government to assess how relevant aspects of the Norwegian legal framework should be interpreted and applied to AI use.

Through the implementation of new structures for algorithm oversight, Spain and the Netherlands also anticipate a critical element of the **EU**'s proposed AI Act. In compliance with Article 59, member states are required to appoint national supervisory authorities that typically act as Market Surveillance Authorities

(MSAs). These public bodies are endowed with regulatory powers, such as accessing training, validation, and testing datasets used by the provider and the AI source code. They also have the authority to withdraw products and require intermediaries to cooperate in removing products from the market. Given this legal requirement under the proposed AI Act, many independent oversight bodies are expected to emerge in EU countries once the proposed AI Act is officially adopted (European Commission, 2021a[33]).

4 Implementing the five recommendations to governments

Investing in AI Research & Development (Principle 2.1)

> *"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."*
>
> *"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."*

Many countries have recognised the importance of policies that support AI R&D and are responding with initiatives to ramp up efforts in this area. Most national AI strategies focus on AI R&D as one of the key areas for action. Countries have dedicated AI R&D funding and support it through different instruments. Main trends include launching AI R&D-focused policies, plans, and programmes, supporting the creation of national AI research institutes and centres, and consolidating AI research networks and collaborative platforms (Figure 4.1).

One way to measure progress in AI R&D is to examine the quantity of AI research published by countries. Over the past decades, there has been consistent growth in the number of AI research publications in both the **United States** and the **European Union**, including journal articles, books, conference proceedings, patents, and academic repositories. AI publications have increased dramatically in **China** and **India** in recent years (Figure 4.2). Since 2019, China has published more AI research than the United States or the European Union. India has also recently made significant advances, with its number of AI research publications more than doubling since 2015. If these trends continue, India can be expected to catch up to the leading three players in AI research (OECD.AI, 2023[91]).

Figure 4.1. Select policies that implement OECD AI Principle 2.1 on investing in AI R&D

AI R&D policies, plans, and programmes	• France's "National AI Research Programme" • United Kingdom's "AI Sector Deal"
AI research institutes and centres of excellence	• Chile's "National AI Research Centre" • Canada's "National AI Institutes (Amii, Vector Institute, Mila)
AI research networks and portals	• Japan's "AI R&D Network" • United States' "AI Researchers Portal"
Funds to support AI R&D diffusion	• European Union's "Horizon Europe 2021-2027" • Australia's "AI Solutions to Build a Stronger Australia"

Figure 4.2. AI research publications, 2000-2021, top publishing countries
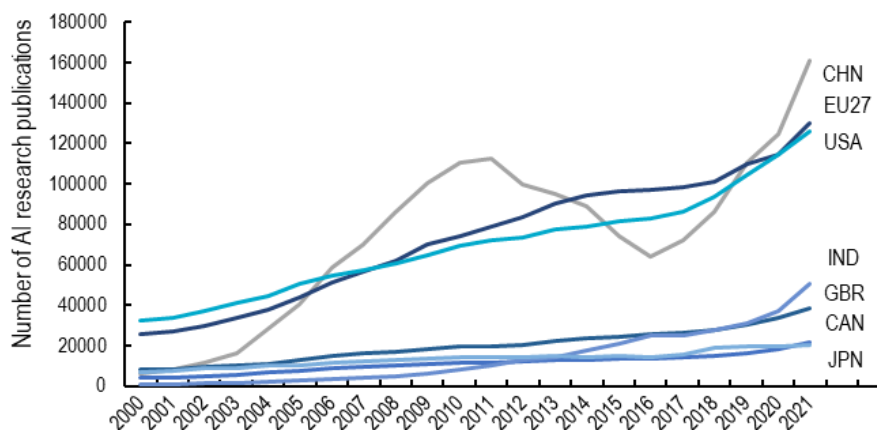
Note: This figure shows the quantity of AI research publications for a sample of top countries for 2000-2021. OpenAlex publications are scholarly documents such as journal articles, books, and dissertations.
Source: (OECD.AI, 2023[91]), visualisations powered by JSI using data from OpenAlex.

Merely assessing the quantity of AI publications does not offer a comprehensive view of publication quality and impact. Analysing the frequency of citations a publication receives can serve as proxy indicator for quality, as a higher number of citations tends to indicate a "higher impact". As of 2019, **Chinese** publications have been cited more frequently, with more AI research papers in the top 1% of the most-cited papers compared to those from both the **United States** and the **European Union** (Figure 4.2). While this trend indicates that China is currently ahead in terms of AI research publications, a different picture emerges when considering the proportion of high-impact publications relative to the total number of AI publications. Since 2000, the **United Kingdom** has led in the proportion of high-quality AI publications, followed by **Canada, China**, and the **United States** (OECD.AI, 2023[91]).
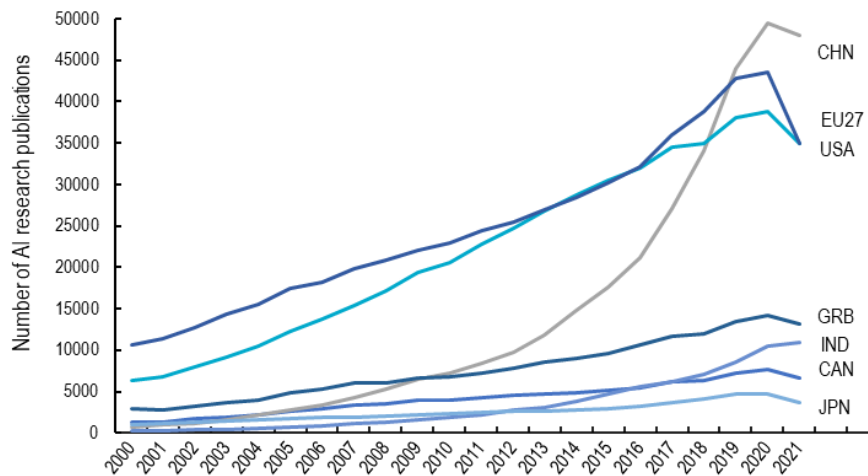
Figure 4.3. High-impact AI research publications, 2000-2021, top publishing countries

Note: This figure shows the quantity and impact of AI research publications for a sample of top countries for 2000-2021. OpenAlex publications are scholarly documents such as journal articles, books, and dissertations. Publication quality ranking is calculated by dividing the number of citations by the average citations in the subdiscipline, discounted by the number of years since the publication was published, plus one (as the number of years impacts the number of citations). Publications are classified as "high-impact" if their score falls in the highest quartile.
Source: (OECD.AI, 2023[91]), visualisations powered by JSI using data from OpenAlex.

Public funding to AI R&D

The allocation of public budgets to AI R&D varies in scale across countries. In many instances, the distribution of budgets per year to AI R&D and other aspects of the strategy is not explicitly mentioned. The OECD has assessed government spending on AI-related R&D up to 2019 through key terms matching of national (and EU) databases (Yamashita and et al., 2021[92]). However, there is no comprehensive method for tracking and comparing AI R&D funding across countries and agencies. While there are no official or comparable estimates of public investment in non-defence AI R&D, several budgetary elements are provided below.

In the **United States**, the funding requested for non-defence AI R&D by the Networking and Information Technology Research and Development (NITRD) Program and the National AI Initiative Office (NAAIO) was USD 1.8 billion in 2023 (National Council of Science and Technology, 2022[93]), making it the highest spending since 2019 (NITRD, 2023[99]). The U.S. National AI Initiative Act of 2020 provides a co-ordinated programme across the Federal government to accelerate AI research and application for the country's economic prosperity and national security (OECD.AI, 2023[8]). Most recently, the NSF and the Office of Science and Technology Policy (OSTP) proposed to establish a National AI Research Resource (NAIRR) with an estimated funding of USD 2.6 billion over the next six years (OECD.AI, 2023[8]). To connect AI researchers to Federal resources that can support their research, the Networking and Information Technology R&D Coordination Office, in partnership with Federal departments and agencies, created the AI Researchers Portal in 2022. The Portal serves as an online AI research platform that includes information about navigating federal research funding processes, data and computing resources, an AI research programme repository, and an AI R&D testbed inventory (NAIIO, 2022[95]).

The **EU** allocated EUR 1 billion per year for AI, including for R&D, within the Horizon Europe and Digital Europe programmes (European Commission, 2023[97]). This follows a Coordinated Plan on AI released in 2018, in which all European Union Member States emphasise the importance of coordinating AI R&D action to maximise the impact of investments at both EU and national levels. The plan, which was reviewed in 2021, includes the development of "shared agendas for industry-academia collaborative AI R&D and

innovation" (OECD.AI, 2023[8]). Horizon Europe is the EU's main research and innovation funding programme. Within this programme, several calls support research and innovation focusing on the development and deployment of AI, the increase of private and public investment, and the promotion of trustworthy AI (European Commission, 2020[98]).

China's 2018 public AI R&D spending, including basic research through the National Natural Science Foundation of China and applied research through the National Key R&D Programmes, has been estimated to range from USD 1.7 billion USD to USD 5.7 billion (Center for Security and Emerging Technology, 2019[99]). Although estimates for more recent years are lacking, it is likely that the budget for AI R&D in China is closer to the upper bound estimate and to have increased in recent years, in line with the country's ambitions set in official strategic documents. The 2017 New Generation AI Plan sets the objective of positioning China as the world's primary innovation centre by 2030, while China's 14th Five-Year-Plan lists AI among the top priorities for scientific research.

AI R&D-focused policies, plans, and programmes

France launched its National AI Research Programme in 2018 as part of its National AI Strategy, the coordination of which has been entrusted to the French National Institute for Research in Digital Science and Technology (INRIA) (Inria, 2023[100]). From 2018 to 2022, EUR 445 million, or nearly 30% of the funding allocated to the strategy, were dedicated to research, compared to EUR 134 million, or 8.7%, in the second phase. **Spain**'s AI Strategy in RDI outlines specific priorities within the new Spanish Strategy for Science, Technology, and Innovation for the period 2021 to 2028. These priorities will be implemented through defined initiatives and activities funded via the Science, Technology, and Innovation Strategic Plans. The goal is to mobilize synergies among various levels of public administration and through cooperation between the public and private sectors. (OECD.AI, 2023[8]). **Portugal**'s R&D programme on data science and AI in Public Administration, which was undertaken between 2018 and 2020, aimed at deepening the processing of public data and stimulating the production of new knowledge relevant to citizens based on the use of advanced techniques of AI and data science. Its main objectives were to promote deriving scientific knowledge from large amounts of data available in public administration, and to assist decision making processes and the definition of public policies in areas such as health, employment, education, sustainable development, and road prevention.

National AI research institutes and centres

Some national AI strategies call for the establishment of AI research institutes and centres of excellence supporting AI R&D efforts. Countries have created national AI research institutes and centres to strengthen their AI research capabilities and to create interdisciplinary research communities.

As part of the Australian AI Action Plan, the Government of **Australia** has allocated AUD 124 million to establish its National AI Centre to further develop Australia's AI and digital ecosystem (CSIRO, 2022[101]). **Canada**'s Pan-Canadian AI Strategy, launched in 2017 and renewed in 2021, has a strong emphasis on research, talent, and commercialization. Through the strategy, Canada is investing in three National AI Institutes: Amii (Edmonton), the Vector Institute (Toronto), and Mila (Montréal). Funding by the strategy includes CAD 60 million over five years (2021 to 2026) to help the institutes drive commercialization and adoption of AI, and CAD 208 million over ten years (2021 to 2031) for the Canadian Institute for Advanced Research (CIFAR) and the institutes to attract, retain and develop academic research talent in AI, and advance AI research (OECD.AI, 2023[8]).

In February 2023, **Korea** opened the Research Data Centre of AI Innovation Hub, the country's top-tier AI research network for AI research and workforce development. It consists of a computing lab with 35 PetaFlops, i.e., machines that allow up to 100 researchers to work on large-scale AI projects simultaneously.

Brazil's Ministry of Science, Technology and Innovation (MCTI), the Sao Paolo Research Foundation and CGI.br are creating up to eight Applied Research Centres in AI (AIGO, 2022[4]). They will collectively invest BRL 1 million per year in each of the new Applied Research Centres for a period of up to ten years. Partner firms will invest a matching amount, taking the total to BRL 20 million per Applied Research Centre. The focus areas of the first centres will be Health (CIIA-*Saúde* and CEREIA), Agriculture (BIOS), Industry (IPT/SP, CIMATEC), and Smart Cities (IARA).

Egypt established an Applied Innovation Center (AIC) within the Ministry of Communications and Information Technology to adopt AI in various fields and to identify innovative solutions to the challenges faced by the Egyptian society. The AIC has already developed an application that uses AI in diagnosing diabetic retinopathy and is currently in the process of developing an initiative aimed at the early detection of diabetic retinopathy for one million citizens. The AIC has also developed an Automatic Speech Recognition system for Arabic, which is being used in tribunals to transcribe judges' sentences (OECD.AI, 2023[8]). **Peru** has established a National Centre for Innovation and AI aimed at accelerating the development and adoption of AI in the country. The centre has a mandate to carry out R&D in AI and to coordinate its activities with national and international academia as well as with the private and public sector.

The **United States**' National AI Institutes represent a cornerstone Federal Government commitment to fostering long-term, fundamental research in AI while also delivering significantly on each of the other eight objectives in that strategy. The programme represents a multisector effort led by the National Science Foundation (NSF), in partnership with the Simons Foundation (SF), the National Institute of Standards and Technology (NIST), Department of Defense (DOD) Office of the Under Secretary of Defense for Research and Engineering (OUSD (R&E)), Capital One Financial Corporation (Capital One), and Intel Corporation (Intel) (National Science Foundation, 2023[102]).

Reinforcing collaborative networks of experts and researchers

Many countries also seek to promote collaboration between experts and researchers in AI and ML through the establishment of national and/or international networks.

In 2021, the **German** government created a Network of National Centres of Excellence for AI Research in 2021 (ML2R, 2021[103]). The Network includes the following six Institutes: (1) Berlin Institute for the Foundations of Learning and Data; (2) German Research Centre for AI; (3) Munich Centre for Machine Learning; (4) Competence Centre Machine Learning Rhine-Ruhr; (5) Tübingen AI Centre (TUE.AI Centre); and (6) Competence Centre for Scalable Data Services and Solutions (ScaDS.AI) Dresden/Leipzig. The Network's objectives are to increase synergies on AI Research through the exchange of competencies and research results, to implement joint activities, and to increase the national and international visibility of German AI research. The European Lighthouse on Secure and Safe AI (2022) – a joint initiative of the **EU** and the **UK** governments funded by Horizon Europe – is a virtual centre of excellence that brings together leading European experts in AI and machine learning (OECD.AI, 2023[8]). **Brazil**'s Ministry of Science, Technology and Innovation (MCTI) and the Brazilian Company for Industrial Research and Innovations (EMBRAPII) have established the MCTI / EMBRAPII Network of Innovation in AI to encourage AI use in the production process of the national industry. EMBRAPII Units support industry projects with infrastructure and qualified professionals. Thus far, the Network has provided support for 286 projects involving 257 companies and successfully concluded 197 of these projects (AIGO, 2022[4]).

Funds to support AI diffusion to businesses and the public sector

Countries have also established specific funds to support AI R&D and the translation of theory to practice. Some of these funds are specific to AI, while others encompass AI among other new technologies such as blockchain and the Internet of Things. Furthermore, some of the funds focus on AI R&D to promote greater adoption in the public sector, while others encourage greater adoption in business and industry.
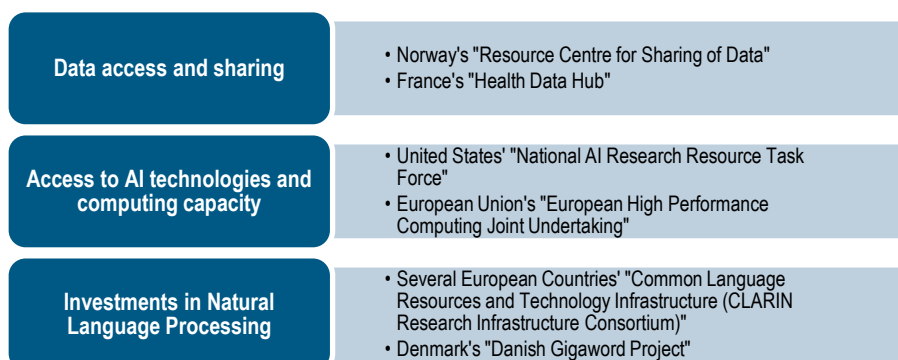
As part of **Denmark**'s national AI strategy, the Danish government launched the Danish National Uptake Fund for New Technologies (2019) to boost the use of AI and dissemination of digital welfare solutions (Agency for Digital Government, 2023[104]). From 2020 to 2022, the National Uptake Fund has funded a total of 40 projects, which provide experience with the use of AI in the public sector and knowledge of challenges when using the technology. The aim of these projects is to provide concrete experience with the use of AI in the areas of welfare services, climate impact, and public administration. **Türkiye**'s AI Ecosystem Call, launched in 2023 by the Ministry of Industry and the Technology & Scientific and Technological Research Council, acts as a support model for facilitating collaboration between companies seeking AI solutions and relevant stakeholders across sectors. The model encourages companies to form a consortium comprising a technology provider, a university research laboratory/centre or a public research centre/institute with AI expertise, and the Scientific and Technological Research Council of Türkiye's AI Institute. Companies in need of AI solutions for sector-specific problems can actively seek out partners to join their consortium. To expedite and optimize the process of addressing these needs, the AI Institute will provide support by pairing experienced technology providers with university research laboratories and public research centres or institutes. The **United Kingdom**'s AI Sector Deal (2018-2027) is a GBP 1 billion support package from the UK government and industry to realise the potential of AI in the UK. It sets out actions to promote the adoption and the use of AI in the UK and delivers on the recommendations of the independent AI review "Growing the AI industry in the UK" (UK Government, 2019[105]). The AI Sector Deal aims to attract and retain both domestic and global AI talent, deliver major upgrades to digital and data infrastructure, ensure that the UK has an enabling business environment, and contribute to communities' prosperity by spreading the benefits of AI across the country.

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."*

Embracing AI-enabled transformation depends on the availability of data, infrastructure, and software to train and use AI models at scale. It is thus critical for countries to ensure they have sufficient AI compute capacity to meet their needs to capture AI's full economic potential. Fostering a digital ecosystem for AI hence represents a crucial component of countries' efforts to advance in their AI adoption. Countries have implemented several policies to address this principle, including linking data access and sharing policies with AI policies, strengthening efforts to increase computing capacity and access to infrastructure, and investing in NLP technologies (Figure 4.4).

Figure 4.4. Select policies that implement OECD AI Principle 2.2 on fostering a digital ecosystem for AI

Data access and sharing	• Norway's "Resource Centre for Sharing of Data" • France's "Health Data Hub"
Access to AI technologies and computing capacity	• United States' "National AI Research Resource Task Force" • European Union's "European High Performance Computing Joint Undertaking"
Investments in Natural Language Processing	• Several European Countries' "Common Language Resources and Technology Infrastructure (CLARIN Research Infrastructure Consortium)" • Denmark's "Danish Gigaword Project"

Linking policies on data access and sharing with AI 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. Several countries have therefore begun to link their data access and sharing policies with AI policies.

The Government of **Colombia**, in partnership with the World Economic Forum, created a data marketplace and infrastructure called the Moonshot project (OECD, 2022[93]). The project was developed as part of the National Policy for Digital Transformation and AI – and specifically, its principle of easily accessible data infrastructure allowing for the design and implementation of AI systems. As part of the Forum's Data for Common Purpose Initiative, the Colombia Moonshot project seeks to implement a data marketplace in the country based on the principles of auditability, equity, ethics, inclusion, transparency, and social and environmental responsibility. The **Czech Republic**'s National Strategy of Open Access to Scientific Information and Data (2017-2020) includes open access to publications and data requirements for publicly funded research projects. Its objective is to define and implement a clear strategy for open access to scientific information for projects financed by public funds (OECD.AI, 2023[8]).

The **French** government created a data sharing commons (data.gouv.fr, 2023[106]), i.e., an open community platform that aims to centralise and structure open data in France. The portal includes a selection of public and private data with an appropriate license to allow their use for research and innovation purposes. Together with the Ibero-American Data Protection Network, in 2019, **Mexico** launched specific guidelines for compliance with the principles and rights that govern the protection of personal data in AI projects. As mentioned above (see p. 42), Mexico's Federal Institute for Access to Public Information and Data Protection (INAI) issued the "Recommendations for the Processing of Personal Data derived from the use of AI" in 2022. The recommendation aims at safeguarding the fulfilment of the data protection principles and duties when applying AI in different sectors such as education, public security, or health.

Sweden's Strategy for Enhanced Data Accessibility (2021) was formulated with the understanding that it is essential to bolster Sweden's international competitiveness by enhancing data sharing capabilities, accelerating data market development, and maintaining a flexible policy framework that can adapt to rapid technological advancements (OECD.AI, 2023[8]). **Peru**'s National Network of Open Access Digital Repositories of Science, Technology, and Innovation (2013) is a national network for open access repositories connecting 49 institutions and providing access to approximately 50 000 publications and datasets (OECD, 2022[93]). The Network's objective is to promote collaboration among its members, thereby fostering open access, use and preservation of information and knowledge in science, technology and innovation. **India**'s Biological Data Storage, Access, and Sharing Policy (2019) establishes a framework and principles for data sharing that safeguard the rights of individuals and populations, ensuring no harm is caused to them. Its objective is to define guidelines for sharing data generated by scientists in India using modern biotechnological tools and methods (OECD.AI, 2023[8]).

Access to AI technologies and computing capacity

Both researchers and industry need access to AI technologies and computing capabilities to develop innovative solutions, solve complex problems, and gain competitive advantages in the global economy. Many governments have therefore implemented policies providing these groups with access to their government-owned AI technologies and computing capabilities.

Canada's Pan-Canadian AI Strategy, launched in 2017 and renewed in 2021, leverages Canada's National AI Institutes—Amii (Edmonton), the Vector Institute (Toronto), and Mila (Montréal)—and supports the acquisition of HPC capacity dedicated for AI researchers. Canada's Digital Research Infrastructure Strategy, launched in 2019, funds the Digital Research Alliance of Canada, a new national not-for-profit organisation to advance and invest in national digital research infrastructure (DRI) activities. Computing

facilities at University of Victoria, Simon Fraser University, University of Waterloo, University of Toronto, and McGill University were expanded as part of the Advanced Research Computing Expansion Program in 2019. (Digital Research Alliance of Canada, 2020[107]).

Providing the needed computation infrastructure for R&D in the academia, private and public sectors is one of the key actions of **Israel**'s National AI Program (2022). Following a request for information for the establishment of an HPC/compute infrastructure in 2021 (Israel Innovation Authority, 2021[108]), the country has launched an advanced HPC R&D technology lab for the use of start-ups and academia in August 2023. Israel's Innovation Authority plans to invest up to NIS 30 million (USD 7.9 million) in its establishment (Israel Defense, 2023[109]). In December 2022, **Korea** announced the K-Cloud Project Implementation Strategy, investing a total of approximately USD 650 million from 2023 to 2030 in the development of ultra-fast, low-power domestic AI chips. The goal is to apply them to data centres, thereby raising domestic cloud competitiveness while contributing to carbon neutrality and providing citizens with advanced AI services. In 2022, **Portugal,** in collaboration with Google, launched "Calls for High Performance Computing R&D Projects: AI on the cloud". The collaboration agreement resulted in the availability of USD 2 million in Google Cloud platform credits over a period of two years for AI R&D projects, with 80% of resources dedicated to the ethics in AI and NLP.
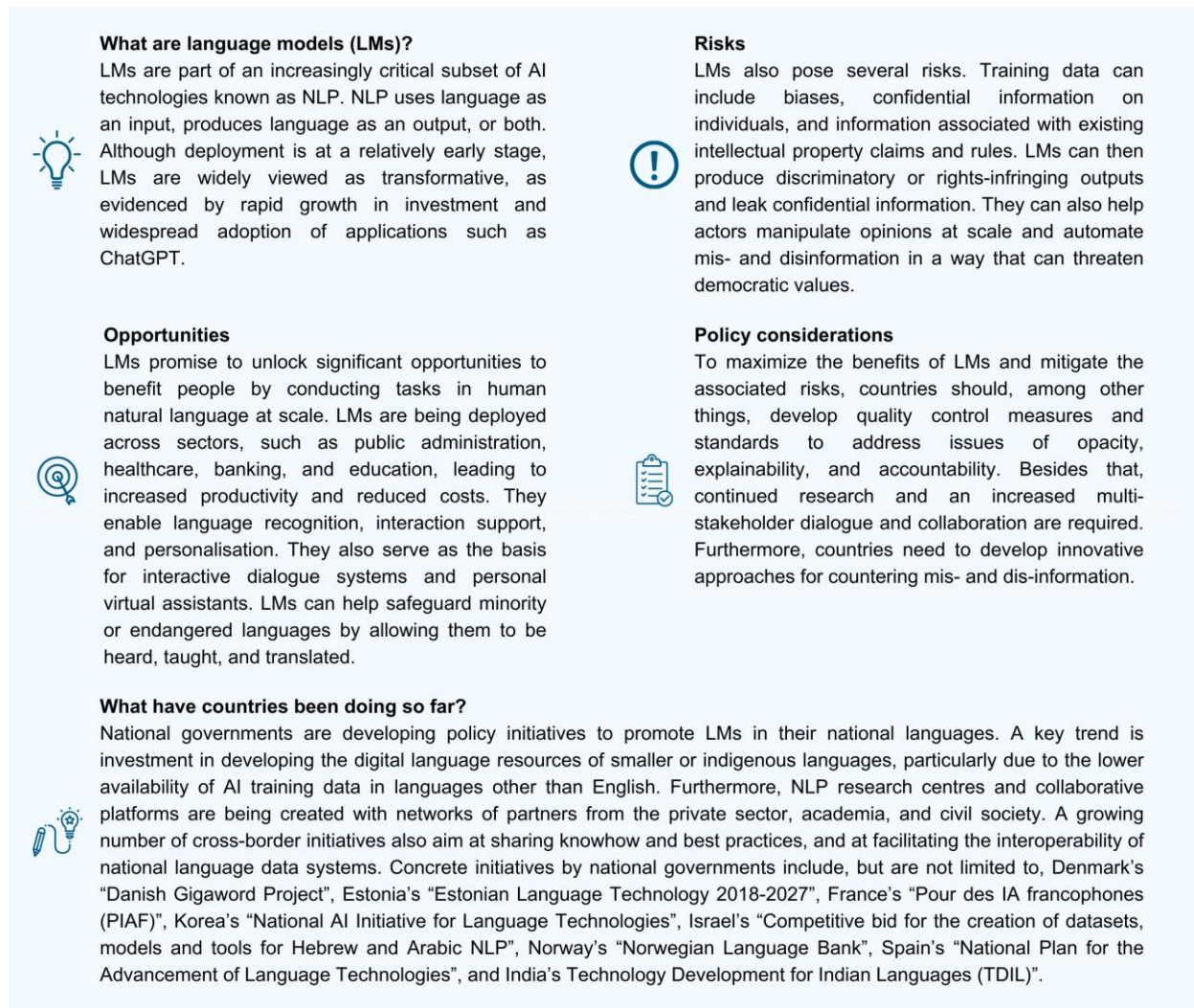
In 2022, the **United Kingdom** launched a review of its digital research infrastructure needs to support the development and use of AI, examining the provision of compute, data access, and talent, which will inform its ongoing national AI strategy (The Alan Turing Institute, 2022[110]). The review report, which was released in March 2023, called for a strategic vision and integrated compute ecosystem, significant investment in public AI compute infrastructure, and the empowerment of the compute community through skills programmes and the attraction of leading AI talent (UK Government, 2023[111]). In 2022, the **United States** Department of Energy launched the Frontier supercomputer as one of the world's most powerful HPCs for AI applications (US Department of Energy, 2019[112]). The National Science Foundation (NSF) invests significantly in next-generation AI R&D supercomputers, such as Frontera, deployed in June 2019 (National Science Foundation, 2019[113]), and provides programs for access to AI compute through the National AI Research Institutes (National Science Foundation, 2022[114]). The United States National AI Initiative Act of 2020 plans to make world-class computing resources and datasets available to researchers across the country through the forthcoming United States National AI Research Resource (NAIRR). In January 2023, an implementation plan was presented, proposing the democratisation of AI R&D through funding a widely accessible AI compute infrastructure with a budget of USD 2.6 billion over an initial six-year period (US Government, 2023[115]).

The Government of **India** established its MeitY Quantum Computing Applications Lab (2021) in collaboration with Amazon Web Services (MEIT, 2023[116]) to provide scientific, academic, and developer communities with access to a quantum computing development environment in the cloud.

Investments in language technologies

Natural Language Processing (NLP) refers to computer programs and tools that automate natural language functions by analysing, producing, modifying, or responding to human texts and speech. NLP is a subset of AI that uses language as an input, produces language as an output, or both. Chatbots, machine translation systems, and virtual assistants that recognise speech are all examples of NLP applications (OECD, 2023[123]). Increasingly powerful NLP such as Large Language Models (LLMs) also raises significant policy challenges related to trustworthiness, including privacy, digital security, misinformation and disinformation, inclusion, as well as financial and environmental cost (Figure 4.5). Limited access to digitally readable text for most languages, which is essential for training models, could limit the benefits of such technology extending to various groups, including those using minority languages. To address this issue, several countries have launched initiatives to promote NLP in their national languages (Figure 4.5).

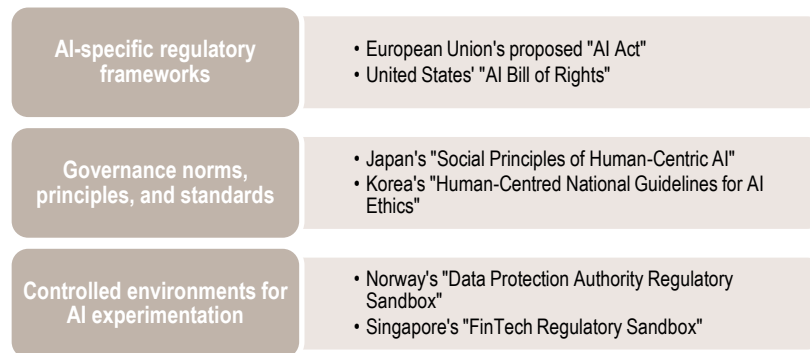## Figure 4.5. Language Models: definition, risks, opportunities, policy considerations and national initiatives

What are language models (LMs)?

LMs are part of an increasingly critical subset of AI technologies known as NLP. NLP uses language as an input, produces language as an output, or both. Although deployment is at a relatively early stage, LMs are widely viewed as transformative, as evidenced by rapid growth in investment and widespread adoption of applications such as ChatGPT.

Risks

LMs also pose several risks. Training data can include biases, confidential information on individuals, and information associated with existing intellectual property claims and rules. LMs can then produce discriminatory or rights-infringing outputs and leak confidential information. They can also help actors manipulate opinions at scale and automate mis- and disinformation in a way that can threaten democratic values.

Opportunities

LMs promise to unlock significant opportunities to benefit people by conducting tasks in human natural language at scale. LMs are being deployed across sectors, such as public administration, healthcare, banking, and education, leading to increased productivity and reduced costs. They enable language recognition, interaction support, and personalisation. They also serve as the basis for interactive dialogue systems and personal virtual assistants. LMs can help safeguard minority or endangered languages by allowing them to be heard, taught, and translated.

Policy considerations

To maximize the benefits of LMs and mitigate the associated risks, countries should, among other things, develop quality control measures and standards to address issues of opacity, explainability, and accountability. Besides that, continued research and an increased multi-stakeholder dialogue and collaboration are required. Furthermore, countries need to develop innovative approaches for countering mis- and dis-information.

What have countries been doing so far?

National governments are developing policy initiatives to promote LMs in their national languages. A key trend is investment in developing the digital language resources of smaller or indigenous languages, particularly due to the lower availability of AI training data in languages other than English. Furthermore, NLP research centres and collaborative platforms are being created with networks of partners from the private sector, academia, and civil society. A growing number of cross-border initiatives also aim at sharing knowhow and best practices, and at facilitating the interoperability of national language data systems. Concrete initiatives by national governments include, but are not limited to, Denmark's "Danish Gigaword Project", Estonia's "Estonian Language Technology 2018-2027", France's "Pour des IA francophones (PIAF)", Korea's "National AI Initiative for Language Technologies", Israel's "Competitive bid for the creation of datasets, models and tools for Hebrew and Arabic NLP", Norway's "Norwegian Language Bank", Spain's "National Plan for the Advancement of Language Technologies", and India's Technology Development for Indian Languages (TDIL)".

Source: (OECD, 2023[123]), "AI language models: Technological, socio-economic and policy considerations", https://doi.org/10.1787/13d38f92-en.

Fostering 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.

Figure 4.6. Select policies implementing OECD AI Principle 2.3 on Fostering an enabling policy environment for AI

AI-specific regulatory frameworks	• European Union's proposed "AI Act" • United States' "AI Bill of Rights"
Governance norms, principles, and standards	• Japan's "Social Principles of Human-Centric AI" • Korea's "Human-Centred National Guidelines for AI Ethics"
Controlled environments for AI experimentation	• Norway's "Data Protection Authority Regulatory Sandbox" • Singapore's "FinTech Regulatory Sandbox"

As discussed in Chapter 2, countries are encouraging trustworthy AI use and innovation by reviewing and adapting existing policies and legislations, as well as by adopting AI-specific regulatory frameworks. They are also supporting an agile transition from R&D to commercialisation or deployment of AI by providing controlled environments for experimentation and testing of AI systems
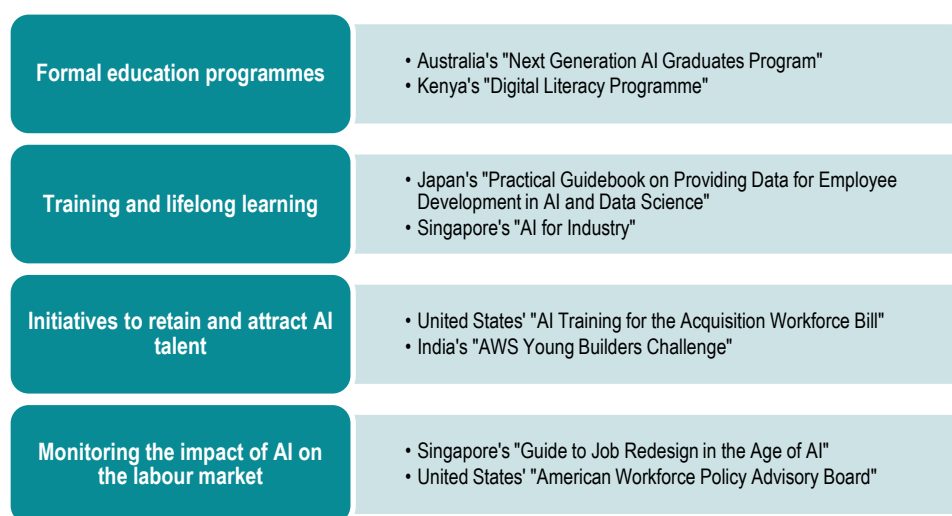
Figure 4.6).

AI skills, jobs and 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.

AI is already changing the nature of many aspects of life as it diffuses across sectors, particularly within the context of labour, employment, and the workplace. 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. Governments have mainly put in place initiatives to prepare the workforce with the skills required for AI through formal education programmes, training and lifelong learning initiatives. They also launched initiatives to attract and retain AI talent. Initiatives to monitor the impact of AI in the labour market and to accompany transitions in the labour market appear limited to date (Figure 4.7).

Figure 4.7. Select policies that implement OECD AI Principle 2.4 on AI skills, jobs and labour market transformation

Formal education programmes	• Australia's "Next Generation AI Graduates Program" • Kenya's "Digital Literacy Programme"
Training and lifelong learning	• Japan's "Practical Guidebook on Providing Data for Employee Development in AI and Data Science" • Singapore's "AI for Industry"
Initiatives to retain and attract AI talent	• United States' "AI Training for the Acquisition Workforce Bill" • India's "AWS Young Builders Challenge"
Monitoring the impact of AI on the labour market	• Singapore's "Guide to Job Redesign in the Age of AI" • United States' "American Workforce Policy Advisory Board"

Formal education programmes for STEM, AI and AI-related fields

Many countries have invested in expanding formal education programmes. They have, for example, designed action plans for stronger digital, AI-focused curricula, developed new AI-centric degree programmes, and launched AI-specific scholarships.

Korea, for example, announced the Comprehensive Strategy for Digital Workforce Development in August 2022, aimed at supporting the expansion of digital education opportunities and digital capacity building of all citizens and nurturing a total 1 million digital talents by 2026. Moreover, in February 2023, Korea developed three types of AI ethics textbooks for students in elementary and secondary schools and three types of teacher manual for teaching AI ethics.

Some countries launched AI PhD programmes to support emerging researchers in AI and machine learning. **Australia** has dedicated AUD 1.4 million to an industry-co-funded PhD scholarship programme – the Next Generation AI Graduates Programme (2021-2027) – to attract and train the next generation of specialist workers in AI by collaborating with industry on research projects and internships. **Italy**'s National PhD Programme in AI (PhD-AI.it) (2021) consists of five federated PhD courses that bring together 61 universities and research institutions (OECD.AI, 2023[8]). The five courses share a common basis in the foundations and developments of AI, while each course specialises in a different sector of AI application (Health and life sciences, Agrifood and environment, Security and cybersecurity, Industry 4.0, and Society) (OECD.AI, 2023[8]). Likewise, **Israel** has been supporting Master's, PhD and Post-doc students in AI with scholarships in 2022 and 2023. The **United Kingdom** has been promoting 2 500 Master's conversion courses at 28 Higher Education Institutes across the country for applicants from near- and non-STEM backgrounds, providing 1 000 scholarships for students from underrepresented groups, including women, black, disabled and lower socio-economic backgrounds, to encourage greater diversity in AI careers. Industry-funded AI Master's programmes (2019) have also been established to broaden access to AI-related education and training and help galvanise future sources of talent (OECD.AI, 2023[8]). The Institute of Coding matches partners in industry interested in sponsoring places on Industrial Masters in AI (IMAI) with higher education institutes able to provide additional places on Master's programmes which meet industry needs (OECD.AI, 2023[8]).

Under its Digital Europe Programme, the **European Union** funds actions to boost advanced digital skills in Europe, including in AI. With a total budget of EUR 580 million for digital skills over seven years, the

programme aims at enhancing co-operation between EU Member States and stakeholders in digital skills and jobs through specialised education programmes in key digital areas such as AI, blockchain, robotics, quantum and HPC. It also funds short-term training courses tailored to the needs of businesses, with an emphasis on SMEs in Europe, as well as jobseekers and citizens looking to reskill. Since 2021, the programme also supports the establishment of four new Master programmes on AI that focus on human centric AI, AI ethics, AI for public sector and AI in healthcare through the Connecting Europe Facility (CEF) Telecom programme (OECD.AI, 2023[8]).

Croatia's European Social Fund Plus (ESF+) Programme 2021-2027 provides support for the application of digital technologies in education. The Programme includes curriculum development and digital transformation of educational processes to strengthen workforce skills and to reinforce quality and accessible education related to the labour market (OECD.AI, 2023[8]). In **Egypt**, developing human capacity for AI is a key pillar of the national AI strategy (National Council for Artificial Intelligence, 2019[118]). The Information Technology Institute developed and is implementing the AI Capacity Building Framework, targeting different population groups, including young people, citizens, AI engineers and workers. In **India**, Amazon established the AWS Young Builders Challenge (2021) in partnership with the Ministry of Education, the Atal Innovation Mission (AIM), Niti Aayog, and the Central Board of Secondary Education (OECD.AI, 2023[8]). The Challenge seeks to enable school students in India to develop an early understanding and adoption of cloud computing and AI, inspire design and computational thinking, and help the rising generation develop a scientific calibre at a young age (AWS India, 2021[119]).

Training and lifelong learning AI and related programmes

Alongside the increased focus on AI in formal education programs, governments have launched initiatives to raise the level of AI skills in the population through vocational training and lifelong learning.

Chile's Digital Talent programme (2019-2022) was established to connect companies, training institutions, and the government to develop new capacities in people, aligned with the demands of the digital economy, and generate more opportunities to access quality jobs (OECD.AI, 2023[8]). **Japan**'s Practical Guidebook on Data Provision for Fostering Human Resources of Experts in AI and Data Science (METI, 2021[120]) focuses on employee development in AI and data science. It summarises issues in four categories based on increasing benefit and risk reduction for businesses (OECD.AI, 2023[8]). AI Competence for **Sweden** (2018) is a national initiative on AI education and competence that involves the collaboration between ten Swedish universities (OECD.AI, 2023[8]) to develop courses for professionals who can contribute to Sweden's development in the area of AI (OECD.AI, 2023[8]). Furthermore, Sweden's Digital Excellence (2019) represents a government assignment to the Swedish Agency for Economic and Regional Growth and the Swedish Higher Education Authority with the goal to enhance competencies of digital excellence in society and, more specifically, in the labour market. **Singapore** has three programmes to enhance skills of professionals. First, AI for Industry (AI4I) (2018) is a training programme designed to equip participants with basic AI and data competency skills. It brings together the country's AI research institutions, start-ups, and companies developing AI products. Second, the AI Apprenticeship Programme (AIAP) consists of a full-time 9-month programme to train local Singaporean AI talent in technology and in the skillsets needed to work with AI. Third, the Chartered AI Engineer (CAIE) designation (2020) is a professional qualification programme by the AI Professionals Association (AIP) to recognise and award credentials to working professionals in AI-related engineering roles (OECD.AI, 2023[8]).

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 first launched in 2017, and renewed in 2021, progress reports indicated that CIFAR and Canada's three National AI institutes had named 120 leading AI researchers, trained over 1,200 graduate students and post-doctoral fellows, and that Canada had been

ranked 6th in the Global AI Talent Report (OECD.AI, 2023[121]). **Germany**, whose technical universities already generate high numbers of homegrown AI talent, has made human capacity a key part of its National AI Strategy. The strategy aims to train 100 new professorships in AI as part of a federal STEM action plan conducted through nationally funded STEM clusters. It will open new centres of excellence and establish a programme for international Master's and PhD students through the German Academic Exchange Service (OECD.AI, 2023[121]). In the **United States**, the AI Training for the Acquisition Workforce Bill (Bill S-2551) (2021) requires the Director of the Office of Management and Budget to provide an AI training programme for the acquisition workforce and other purposes (OECD.AI, 2023[8]). **China** has introduced its AI Innovation Action Plan for higher education to "significantly enhance China's cadre of AI talent and its university AI curricula by 2030" by optimising college infrastructure, delivering first-class AI research, and operating leading AI innovation centres in universities (OECD.AI, 2023[121]). Attracting and developing AI talent is also a key objective of China's National New Generation AI Plan, which prioritises talent development, education, and skills acquisition (OECD.AI, 2023[121]).

Monitoring the impact of AI on the labour market

In order to better prepare for changes, a few countries have launched efforts to measure the impact of AI on the labour market.

The Office of the White House established the **American** Workforce Policy Advisory Board (2019) as part of a national initiative to help bridge the widening skills, driven in part by increased automation and the growing demand for high-tech skills. (OECD.AI, 2023[8]). The Board issues advice and recommendations to the National Council for the American Worker to encourage the private sector and educational institutions to combat the skills crisis through demand-driven education, training, and re-training. **Singapore**'s Guide to Job Redesign in the Age of AI (2020) 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 (OECD.AI, 2023[8]).

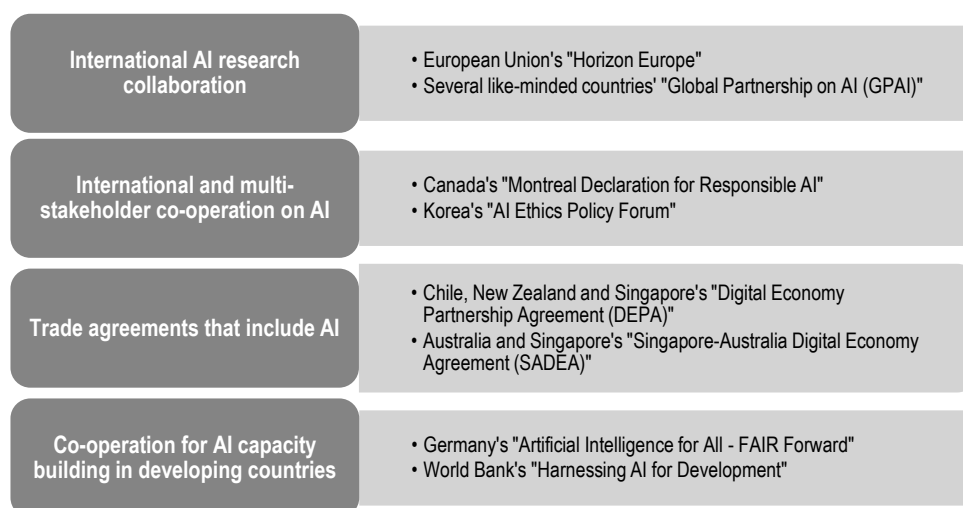## International and multi-stakeholder co-operation on AI (Principle 2.5)

> *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.*

Countries are increasingly engaged in international co-operation to promote the beneficial use of AI while addressing its challenges. This is happening through several types of initiatives, including: i) international research collaborations on AI, ii) trade agreements including language on AI, and iii) co-operation for AI capacity building in developing countries (Figure 4.8).

Intergovernmental organisations (IGOs) play a key role in helping policy makers develop common visions and solutions. Many of these organisations, with complementary mandates and memberships, are actively involved in AI initiatives and projects. The OECD and seven other treaty-based IGOs launched the coalition **GlobalPolicy.AI** (

Figure 4.9), which includes an online platform to align their work and share information on the AI initiatives and projects undertaken by the respective organisations (Globalpolicy.AI, 2023[122]). Through this collaboration, partner IGOs help policy makers navigate various international initiatives, stay informed about each other's AI policy activities, strive to ensure interoperability between their work, and work together to advance trustworthy AI in areas where their mandates intersect.

Figure 4.8. Select policies implementing the OECD AI Principle 2.5 on international co-operation

International AI research collaboration	• European Union's "Horizon Europe" • Several like-minded countries' "Global Partnership on AI (GPAI)"
International and multi-stakeholder co-operation on AI	• Canada's "Montreal Declaration for Responsible AI" • Korea's "AI Ethics Policy Forum"
Trade agreements that include AI	• Chile, New Zealand and Singapore's "Digital Economy Partnership Agreement (DEPA)" • Australia and Singapore's "Singapore-Australia Digital Economy Agreement (SADEA)"
Co-operation for AI capacity building in developing countries	• Germany's "Artificial Intelligence for All - FAIR Forward" • World Bank's "Harnessing AI for Development"

International AI research collaboration

The **Global Partnership on AI** (GPAI) is an international and multi-stakeholder initiative jointly founded by Canada and France during their respective 2018 and 2019 G7 presidencies. GPAI was launched in June 2020 to undertake cutting-edge research and pilot projects on AI priorities to advance the responsible development and use of AI. As of 2023, it counts 29 members (GPAI, 2020[123]). GPAI is led by a ministerial-level Council and a Steering Committee and is supported by a Secretariat hosted by the OECD, as well as by two Centres of Expertise: the International Centre of Expertise in Montreal for the Advancement of Artificial Intelligence (ICEMAI) and the French National Institute for Research in Digital Science and Technology (INRIA). GPAI brings together leading AI experts from industry, government, civil society, and academia to collaborate across four current working groups on the themes of: i) responsible AI (including a subgroup on AI and pandemic response), ii) data governance, iii) the future of work, and iv) innovation and commercialisation. The Montreal Centre of Expertise supports the first two working groups, while the Paris Centre of Expertise supports the latter two working groups.

Countries are promoting cross-border research collaboration on AI at the regional level. In Europe, Ministers responsible for digital development from Denmark, Estonia, Finland, the Faroe Islands, Iceland, Latvia, Lithuania, Norway, Sweden, and the Åland Islands released a **Declaration on AI in the Nordic-Baltic Region** (OECD.AI, 2023[8]). The **Working Group on AI of the African Union (AU)** aims to unify views, develop a single African AI strategy, and to create a joint capacity-building framework across the continent (OECD.AI, 2023[8]). In the **Japan-Singapore Economic Partnership Agreement (JSEPA)** framework, signed in 2002, Japan and Singapore plan to build upon the two Memoranda of Cooperation covering areas like AI, cybersecurity, and digital government transformation (OECD.AI, 2023[8]). The **Quadrilateral Security Dialogue** (QUAD) is an informal strategic forum of the United States of America, India, Australia and Japan. In September 2021, it released the QUAD Principles on Technology Design, Development, Governance, and Use (White House, 2021[124]). In the document, the signatories declare their commitment to facilitating the exchange of researchers and movement of highly skilled personnel, to

enhance science and technology collaboration, and to develop shared research and development agendas, joint projects and joint capacity building.

International and multi-stakeholder co-operation on AI

International research collaboration in AI has the potential to accelerate scientific progress by combining knowledge, resources, and expertise from different countries and cultures. It can help address and resolve ethical, social, and political challenges associated with the global implementation of AI technologies on a broader, international scale. Some countries have therefore launched initiatives to foster international and multi-stakeholder co-operation on AI. In **Canada**, the Montreal Declaration for Responsible AI is an initiative led by the Montreal AI Ethics Institute that seeks to establish a global consensus on the ethical and trustworthy use of AI. The declaration has been signed by about 2 400 individuals and organisations from around the world (OECD.AI, 2023[8]). In **Korea**, the AI Ethics Policy Forum involves a multi-stakeholder approach to spread the AI Ethics System. It establishes a foundation for AI reliability technology and consists of three committees: (1) Proliferation of AI ethics system; (2) Laying the foundation for trustworthy AI; and (3) Strengthen AI literary and ethics education (OECD.AI, 2023[8]).

The private sector is also coming together with other stakeholders. **Partnership on AI** is a collaboration between major tech companies like Amazon, Google, and Microsoft, as well as the civil society and non-profit organisations such as the American Civil Liberties Union (ACLU) and the Electronic Frontier Foundation (EFF), the academic community, and media organisations. The collaboration includes several members from Asia, including Baidu Research and the University of Tokyo. The partnership aims to promote AI technologies that are beneficial to society, ethical, transparent, and trustworthy (PAI, 2023[125]).

Trade agreements that include AI

AI has an undisputable potential to drive innovation, help firms create new value from data, and reduce trade costs. Trade can also be an important mechanism through which countries and firms access the inputs needed to build AI systems (Ferencz, López González and Oliván García, 2022[126]).

The growing interest in AI systems increases the focus on the intersection of trade policy and AI, resulting in the enhanced presence of AI in the current trade policy deliberations. Some countries are leveraging trade agreements to enhance co-operation on AI. In 2020, Chile, New Zealand and Singapore signed the **Digital Economy Partnership Agreement (DEPA)**, which aims to promote the safe and responsible use of AI technologies (DEPA, 2020[127]). In 2023, Korea has agreed to join DEPA as of 2030. Australia and Singapore, building on their pre-existing trade agreement, signed the **Singapore-Australia Digital Economy Agreement (SADEA)** in the same year, where parties agreed to advance their co-operation on AI (SADEA, 2020[128]). Furthermore, the **Korea-Singapore Digital Partnership Agreement (KSDPA)** entered into force in January 2023. In Article 14.28 of this digital economy agreement, both sides recognised the importance of developing ethical governance frameworks for the trusted, safe and responsible use of AI technologies, and agreed to cooperate through promoting dialogue and sharing experiences on regulations, policies, and initiatives relating to the use and adoption of AI technologies.

Co-operation for AI capacity building in developing countries

To ensure that the potential of AI is not only harnessed in the Global North, but also in countries of the Global South, some countries, international organisations, and companies have co-operated around AI capacity building in developing countries.

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. It aims to strengthen local technical know-how on AI in Africa and Asia, improve access to training data and AI technologies for local innovation, and to develop policy frameworks for ethical AI,

data protection, and privacy. To date, FAIR Forward is active in Ghana, Rwanda, Kenya, South Africa, Uganda, Indonesia, and India. The partnering countries are pursuing objectives in co-operation that include: (1) building, expanding, and transferring knowledge to AI; (2) improving access to training data and AI technology; and (3) developing political frameworks for ethical AI and improved data protection (OECD.AI, 2023[8]).

The Asian and Pacific Training Centre for Information and Communication Technology for Development (APCICT) was inaugurated in **Korea** in 2006 as a regional institute of the Economic and Social Commission for Asia and the Pacific (ESCAP). Its main goal is to build ICT capacity and narrow the digital divide in the region. The Centre provides developing countries in the Asia-Pacific region with ICT training, knowledge sharing and policy consultations. Among other things, the Centre also provides an "Academy of ICT Essentials for Government Leaders (ACADEMY)" course. Under ACADEMY, the Frontier ICT for Sustainable Development for Digital Leaders programme, which focuses on AI, big data and blockchain, was developed and launched in September 2022. As of July 2023, a total of three sessions have been held with 142 people finishing the programme. In June 2023, the Ethics of AI programme was developed, with the goal of being unveiled in Indonesia in August 2023. Furthermore, APCICT courses are requirements for central government official training programmes in 16 countries, including Kazakhstan, the Philippines, Cambodia, Bhutan, and Indonesia. The courses have been translated into 17 languages and distributed to 40 countries.

The **United Nations Development Programme (UNDP)** has developed iVerify, an open-source, automated fact-checking tool that can be used to help identify false information and prevent and mitigate its spread. iVerify is now listed as a Digital Public Good and has been deployed in Zambia, Kenya, and Honduras, with plans for deployment in Liberia, to tackle misinformation during elections. The UNDP is also developing AI Readiness Assessment to help governments understand the current state of AI adoption through multi-stakeholder engagement, including with marginalized groups. Furthermore, UNDP supports countries such as Moldova, Senegal, Mauritania, and Kenya in developing Data Governance frameworks that foster responsible and inclusive use of data in policymaking (OECD.AI, 2023[8]).

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. It aims to understand the role of governments in fostering AI development and adoption in developing country contexts. The work highlights how governments are designing policy and regulatory frameworks around AI to support their unique development needs and make progress towards reaching the 17 SDGs (World Bank, 2023[129]).

Figure 4.9. Globalpolicy.AI: a platform for intergovernmental co-operation on AI

Organisation for Economic Co-operation and Development (OECD)

The OECD has launched several AI-specific initiatives. First, the OECD AI Policy Observatory (OECD.AI) is an online platform that provides a valuable reference for international dialogue and collaboration on AI public policy issues. Second, the OECD's Committee on Digital Economy Policy created the Working Party on Artificial Intelligence Governance (AIGO) to oversee its work on artificial intelligence (AI) policy. Third, The OECD.AI Network of Experts works with the working party as an informal group of AI experts from government, business, academia, and civil society. The network provides AI-specific policy advice for the OECD's work on AI policy and contributes to the OECD Policy Observatory on AI.

Inter-American Development Bank (IDB)

The IDB leads the 'fAIr LAC' initiative, an alliance in Latin American and the Caribbean (LAC) region between public and private sectors, civil society and academia to promote the responsible and ethical use of AI. Its objective is to improve public services such as education, health, and social protection by developing knowledge products, informing the conversation around AI, implementing experiments and pilot projects of AI, and creating practical tools to apply the OECD Principles and mitigate potential risks. The initiative includes a regional observatory to map and track AI projects and use cases.

European Commission (EC)

The EC's vision on AI is based on the twin objective of excellence and trust. People's safety and fundamental rights are at the centre of all efforts. At the same time, the EC works together with its Member States and other partners to foster socio-economic and environmental wellbeing through a flourishing science, business, and innovation domain. In 2018, the EC, together with its Member States as well as with Norway and Switzerland, have agreed to work together on maximising investments to AI and on preparing a framework within which these actions can operate. This "Coordinated Plan on AI" was updated in 2021 to translate this joint commitment into action with a vision to accelerate, act and align. Furthermore, together with the United States, the EC has created the Trade and Technology Council (TTC). The TTC provides a mechanism to support stronger transatlantic relations and to deliver concrete outcomes through ten working groups. These groups assist the sustainable, inclusive economic growth and development, promote a human-centric approach to the digital transformation, and ensure that international norms are respected.

World Bank Group (WBG)

The WBG's initiative "Harnessing AI for Development" aims to understand the role of governments in fostering AI development and adoption in developing countries. The WBG's Technology & Innovation Lab works with internal operational teams across various sectors by applying emerging technology to solve business challenges. The WBG's Disruptive Technology for Development (DT4D) Trust Fund is a real-testing lab of AI approaches to tackle operational development challenges in WBG projects where traditional solutions have not worked. Between 2019-2022, the WBG conducted a project on NLP to evaluate the presence of gender biases in the written decisions of judges in Kenya by measuring gender attitudes in the language of decisions. It also measured gender biases in the US based on the usage of gender-stereotyped language in judges' authored opinions.

United Nations (UN)

In 2021, the UN published the "Resource Guide on AI Strategies" to lay out existing resources on AI ethics, policies, and strategies on the national, regional, and international level. In September 2021, the Secretary General released his report "Our Common Agenda", proposing a Global Digital Compact to "outline shared principles for an open, free and secure digital future for all", to be agreed at the Summit of the Future in September 2024**.**

European Agency of Fundamental Rights (FRA)

On December 8, 2022, FRA published a report on bias in algorithms, AI, and discrimination, looking at the use of AI in predictive policing and offensive speech detection. Furthermore, in line with the proposed EU AI Act and Council of Europe's international (framework) Convention on AI, FRA has started to work on a project to provide empirical analysis and guidance on how to assess high-risk AI in relation to fundamental rights.

Council of Europe (CoE)

From 2019 to 2021, the CoE's Ad Hoc Committee on AI (CAHAI) examined the feasibility and potential elements of a possible legal framework to ensure that AI is used to promote and protects CoE's standards (see Chapter 2).

United Nation Educational, Scientific and Cultural Organisation (UNESCO)

In November 2021, UNESCO adopted the Recommendation on the Ethics of Artificial Intelligence, which aims to make AI systems work for the good for humanity, individuals, societies, and the environment while preventing harm. During 2022-2025, UNESCO plans to elaborate specific tools for the implementation of the Recommendation, such as ethical impact assessment (EIA) and readiness assessment methodology; establish a Global Observatory of AI Ethics; assist Member States in building strong national institutions for promoting AI ethics; and fund experts' groups, including a Women for Ethical AI Network (W4ethicalAI). In 2021, UNESCO and the United Nations Institute for Training and Research (UNITAR) jointly launched a 3-year new microlearning course on AI and Human Rights for youths aged 16 to 24. Also in 2021, UNESCO launched a global Massive Open Online Course (MOOC) on AI and the Rule of Law to strengthen capacities of judicial operators in the use of AI in the administration of justice. In 2022, UNESCO and the International Research Centre in Artificial Intelligence (IRCAI) launched their second call for Top 100 innovative solutions that leverage AI to address human rights and the Sustainable Development Goals (SDGs) and to showcase promising AI initiatives.

References

Agency for Digital Government (2023), *Signature Projects Under the Danish National Uptake Fund for New Technologies*, https://en.digst.dk/policy/new-technologies/national-uptake-fund-for-new-technologies/. [110]

AI Standards Hub (2023), *About the AI Standards Hub*, https://aistandardshub.org/the-ai-standards-hub/ (accessed on 13 March 2023). [58]

AIGO (2023), *Presentation at the third meeting of the OECD Working Party on Artificial Intelligence Governance (AIGO)*. [28]

AIGO (2022), *Presentation at second meeting of the OECD Working Party on Artificial Intelligence Governance (AIGO)*. [5]

Aitken et al. (2023), *Common Regulatory Capacity for AI*, https://doi.org/10.5281/zenodo.6838946 (accessed on 29 March 2023). [53]

Artificial Intelligence and Data Science Committee (2020), *[Final report of the committee on AI and data science]*, https://innovationisrael.org.il/sites/default/files/%D7%93%D7%95%D7%97%20%D7%A1%D7%95%D7%A4%D7%99%20%D7%A1%D7%99%D7%9B%D7%95%D7%9D%20%D7%95%D7%95%D7%A2%D7%93%D7%AA%20%D7%AA%D7%9C%D7%9D%20%D7%9C%D7%AA%D7%9B%D7%A0%D7%99%D7%AA%20%D7%9E%D7%95%D7%A4%20%D7%9C%D. [3]

Atlantic Council (2021), *Ukraine's roadmap to an artificial intelligence future*, https://www.atlanticcouncil.org/blogs/ukrainealert/ukraines-roadmap-to-an-artificial-intelligence-future/. [6]

Austrian Government (2021), *Strategie der Bundesregierung für Künstliche Intelligenz*, https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/strategie-bundesregierung.html (accessed on 28 March 2023). [71]

AWS India (2021), *Students present innovative ideas at AWS Young Builders Challenge 2021*, https://www.aboutamazon.in/news/aws/students-present-innovative-ideas-at-aws-young-builders-challenge-2021 (accessed on 29 March 2023). [125]

Ben-Israel, I., E. Matania, E and L. Friedman (2020), *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Special Report to the Prime Minister*, https://icrc.tau.ac.il/sites/cyberstudies-english.tau.ac.il/files/media_server/cyber%20center/The%20National%20Initiative_eng%2 [2]

02021_digital.pdf.

Bertuzzi, L. (2023), *EU lawmakers set to settle on OECD definition for Artificial Intelligence*, https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/. [40]

CAIDP (2021), *Artificial Intelligence and Democratic Values Index*, https://www.caidp.org/reports/aidv-2021/ (accessed on 29 March 2023). [43]

Canadian Government (2023), *Algorithmic Impact Assessment tool*, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html (accessed on 28 March 2023). [92]

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). [16]

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). [13]

Center for Security and Emerging Technology (2021), *Ethical Norms for New Generation Artificial Intelligence Released*, https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/. [136]

Center for Security and Emerging Technology (2019), *Chinese Public AI R&D Spending: Provisional Findings*, https://cset.georgetown.edu/wp-content/uploads/Chinese-Public-AI-RD-Spending-Provisional-Findings-1.pdf. [105]

Chavez, K., J. Bahr and T. Vartanian (2022), *AI has made its way to the workplace. So how have laws kept pace?*, https://oecd.ai/en/wonk/workplace-regulation-2022. [90]

Chinese Government (2022), *Opinion on Strengthening the Ethics and Governance in Science and Technology*, http://www.gov.cn/zhengce/2022-03/20/content_5680105.htm (accessed on 29 March 2023). [46]

Chinese Government (2021), *Data Security Law*, http://www.npc.gov.cn/englishnpc/c23934/202112/1abd8829788946ecab270e469b13c39c.shtml (accessed on 29 March 2023). [51]

Civic Coding (2018), *Über uns*, https://www.civic-coding.de/ueber-uns (accessed on 28 March 2023). [70]

Council of Europe (2023), *Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66. [42]

Council of Europe (2020), *Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154 (accessed on 28 March 2023). [78]

Court des Comptes (2023), *La stratégie nationale de recherche en intelligence artificielle*, Court des comptes, https://www.ccomptes.fr/fr/publications/la-strategie-nationale-de-recherche-en-intelligence-artificielle. [12]

CSIRO (2022), *National Artificial Intelligence Centre*, https://www.csiro.au/en/work-with-us/industries/technology/national-ai-centre. [107]

Data & Trust Alliance (2022), *Data & Trust Alliance*, https://dataandtrustalliance.org/. [50]

data.gouv.fr (2023), *Plateforme ouverte des données publiques françaises*, https://www.data.gouv.fr/fr/. [112]

Datatilsynet (2021), *Sandbox for responsible artificial intelligence*, https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/. [63]

Datatilsynet (2020), *Sandbox for responsible artificial intelligence*, https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/#:~:text=The%20main%20mission%20for%20the%20Data%20Protection%20Authority%27s,full%20openness%20about%20the%20assessments%20that%20are%20made. (accessed on 29 March 2023). [62]

Democratic Society (2022), *Scottish Artificial Intelligence (AI) Co-Creation Public Engagement Workshops*, https://www.demsoc.org/projects/scottish-artificial-intelligence-ai-co-creation-public-engagement-workshops (accessed on 28 March 2023). [73]

DEPA (2020), *Digital Economy Partnership Agreement*, https://www.mti.gov.sg/-/media/MTI/Microsites/DEAs/Digital-Economy-Partnership-Agreement/Text-of-the-DEPA.pdf. [133]

Digital Research Alliance of Canada (2020), *Canadian Digital Research Infrastructure Needs Assessment*, https://alliancecan.ca/en/initiatives/canadian-digital-research-infrastructure-needs-assessment. [113]

Dutch Government (2021), *Fundamental Rights and Algorithms Impact Assessment (FRAIA)*, https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms (accessed on 28 March 2023). [77]

Dutch Ministry of Internal Affairs (2021), *Non-discrimination by design*, https://www.tilburguniversity.edu/sites/default/files/download/04%20handbook%20non-discrimination%20by%20design%28ENG%29.pdf. [84]

European Commission (2023), *A European approach to artificial intelligence*, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence. [103]

European Commission (2022), *AI Liability Directive*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0496 (accessed on 29 March 2023). [138]

European Commission (2022), *Proposal Product Liability Directive*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A495%3AFIN (accessed on 29 March 2023). [137]

European Commission (2020), *Horizon Europe*, European Commision, https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en. [104]

European Commission (2021b), *Coordinated Plan on Artificial Intelligence 2021 Review*, https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review. [38]

European Commission (2021a), *Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence and amending certain Union Legislative Acts (Artificial Intelligence Act)*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (accessed on 28 March 2023). [37]

European Council (2022), *Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights*, https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/ (accessed on 29 March 2023). [39]

European Parliament (2023), *DRAFT Compromise Amendments on the Draft Report - Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf. [41]

European Union (2022), *Digital Services Act (DSA) Package*, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package (accessed on 28 March 2023). [87]

European Union (2018), *General Data Protection Regulation (GDPR)*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679 (accessed on 28 March 2023). [91]

Executive Office of the President (2020), *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*, https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government. [35]

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). [14]

Ferencz, J., J. López González and I. Oliván García (2022), "Artificial Intelligence and international trade: Some preliminary implications", *OECD Trade Policy Papers*, Vol. No. 260, https://doi.org/10.1787/13212d3e-en. [132]

Fond Darstellende Künste (2021), *AUTONOM - Symposium zu Künstlicher Intelligenz in den Darstellenden Künsten*, https://www.fonds-daku.de/events-und-diskurs/archiv/autonom-symposium/ (accessed on 28 March 2023). [75]

Forum IA Quebec (2023), *Artificial intelligence will change the world*, https://forumia.quebec/en/ (accessed on 28 March 2023). [69]

French Government (2021), *Aide pour la maîtrise et la diffusion numérique dans le cadre de « IA Booster »*, https://www.economie.gouv.fr/plan-de-relance/mesures/aide-maitrise-diffusion-numerique-ia-booster (accessed on 28 March 2023). [67]

German Government (2022), *Regulatory Sandboxes – Testing Environments for Innovation* [60]

and Regulation, https://www.bmwk.de/Redaktion/EN/Dossier/regulatory-sandboxes.html (accessed on 29 March 2023).

German Government (n.d.), *Unsere Förderinitiative "KI-Leuchttürme für Umwelt, Klima, Natur und Ressourcen"*, https://www.bmuv.de/themen/nachhaltigkeit-digitalisierung/digitalisierung/unsere-foerderinitiative-ki-leuchttuerme (accessed on 28 March 2023). [74]

Globalpolicy.AI (2023), *https://globalpolicy.ai/en/*, https://globalpolicy.ai/en/. [128]

Government of Canada (2023), *Advisory Council on Artificial Intelligence*, https://ised-isde.canada.ca/site/advisory-council-artificial-intelligence/en. [10]

Government of Canada (2023), *Artificial Intelligence and Data Act - Companion Document*, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document. [89]

Government of Canada (2023), *Public Awareness Working Group*, https://ised-isde.canada.ca/site/advisory-council-artificial-intelligence/en/public-awareness-working-group (accessed on 28 March 2023). [72]

GPAI (2020), *Global Partnersip on Artificial Intelligence*, https://www.gpai.ai/. [129]

Habuka, H. (2023), *Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency*, Center for Strategic and International Studies (CSIS), https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency. [86]

Holistic AI (2023), *Making Sense of China's AI Regulations*, https://www.holisticai.com/blog/china-ai-regulation. [47]

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. [80]

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. [81]

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. [79]

Indian Government (2023), *IndiaAI*, https://indiaai.gov.in/ (accessed on 28 March 2023). [7]

Inria (2023), *Programme national de recherche en intelligence artificielle*, https://www.inria.fr/fr/programme-national-recherche-intelligence-artificielle. [106]

ISO (2022), *ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*, https://www.iso.org/standard/74438.html. [56]

ISO (2009), *ISO 31000 Risk management*, https://www.iso.org/iso-31000-risk- [57]

management.html.

Israel Defense (2023), *Next Silicon will establish a national R&D laboratory for acceleration development of technologies for supercomputing*, https://www.israeldefense.co.il/node/59291 (accessed on 21 September 2023). [115]

Israel Innovation Authority (2021), *[National Infrastructure for supercomputing (HPC) - Request for Information]*, https://www.law.co.il/media/computer-law/innovation_authority_doc1.pdf. [114]

Israeli government (2023), *[New interagency team looking at the use of AI in the financial sector]*, https://www.gov.il/he/departments/news/ai-team. [18]

Japanese Government (2022), *AI Strategy 2022*, https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf. [4]

Japanese Government (2019), *Social Principles of Human-Centric AI*, https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf (accessed on 29 March 2023). [19]

Japanese government (2023), *Summary of AI issues*, https://www8.cao.go.jp/cstp/ai/ronten_youshi_yaku.pdf. [25]

KI Bundesverband (2019), *KI Gütesiegel*, https://ki-verband.de/wp-content/uploads/2019/02/KIBV_Guetesiegel.pdf (accessed on 28 March 2023). [83]

Landry et al. (2022), *Bill C-27 – Canada's proposed Artificial Intelligence and Data Act*, https://www.stewartmckelvey.com/thought-leadership/bill-c-27-canadas-proposed-artificial-intelligence-and-data-act/ (accessed on 29 March 2023). [15]

Laskai, L. and G. Webster (2019), *Translation: Chinese Expert Group Offers 'Governance Principles' for 'Responsible AI'*, https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/. [11]

MEIT (2023), *MeitY Quantum Computing Applications Lab*, https://quantumcomputing.negd.in/. [122]

METI (2021), *Practical Guidebook on Data Provision for Fostering Human Resources of Experts in AI and Data Science*, https://www.meti.go.jp/english/press/2021/0301_003.html. [126]

MIC (2023), *G7 Hiroshima AI Process, G7 Digital & Tech Ministers' Statement (September 7, 2023)*, https://www.soumu.go.jp/main_content/000900470.pdf (accessed on 7 September 2023). [23]

MIC (2023), *Results of the G7 Digital and Tech Ministers' Meeting in Takasaki, Gunma.*, https://www.soumu.go.jp/joho_kokusai/g7digital-tech-2023/en/topics/topics_20230430.htm. [21]

Ministry of Economy, Trade and Industry (2022), *Governance Guidelines for Implementation of AI Principles (Ver. 1.1)*, https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf. [20]

Ministry of Foreign Affairs of Japan (2023), *G7 Hiroshima Leaders' Communiqué*, https://www.mofa.go.jp/ms/g7hs_s/page1e_000690.html (accessed on 6 July 2023). [22]

Ministry of Innovation, Science and Technology and Ministry of Justice, the Office of legal [17]

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.

ML2R (2021), *Network of German AI Centres strengthens Scientific Exchange*, https://www.ml2r.de/en/network-of-german-ai-centres-strengthens-scientific-exchange/. [109]

NAIIO (2022), *AI Researchers Portal*, https://www.ai.gov/ai-researchers-portal/. [102]

National Artificial Intelligence Initiative Office (2023), *ABOUT*, https://www.ai.gov/about/#NAIIO_-_National_Artificial_Intelligence_Initiative_Office. [9]

National Council for Artificial Intelligence (2019), *Egypt National Artificial Intelligence Strategy*, https://mcit.gov.eg/Upcont/Documents/Publications_672021000_Egypt-National-AI-Strategy-English.pdf. [124]

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. [100]

National Institute of Standards and Technologies - US Department of Commerce (2023), *AI Risk Management Framework*, https://www.nist.gov/itl/ai-risk-management-framework (accessed on 28 March 2023). [33]

National Science Foundation (2023), *National Artificial Intelligence Research Institutes*, https://new.nsf.gov/funding/opportunities/national-artificial-intelligence-research. [108]

National Science Foundation (2022), *National Artificial Intelligence Research Institutes*, https://beta.nsf.gov/funding/opportunities/national-artificial-intelligence-research-institutes. [120]

National Science Foundation (2019), *NSF-funded leadership-class computing center boosts U.S. science with largest academic supercomputer in the world*, https://www.nsf.gov/news/news_summ.jsp?cntn_id=299134. [119]

NIST (2023), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. [55]

NITRD (2023), *ARTIFICIAL INTELLIGENCE R&D INVESTMENTS, FISCAL YEAR 2018 – FISCAL YEAR 2023*, https://www.nitrd.gov/apps/itdashboard/ai-rd-investments/#Chart-2-Federal-Budget-for-Nondefense-AI-RD-FYs-2018-2023 (accessed on 2023). [101]

OECD (2023), "Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI", *OECD Digital Economy Papers,* No. 349, https://doi.org/10.1787/2448f04b-en. [66]

OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *Digital Economy Papers,* No. 352, https://doi.org/10.1787/13d38f92-en. [123]

OECD (2023), *Catalogue of AI Tools & Metrics*, https://oecd.ai/en/catalogue/overview. [96]

OECD (2023), "Emerging privacy-enhancing technologies: Current regulatory and policy approaches", *OECD Digital Economy Papers,* No. 351, https://doi.org/10.1787/bf121be4-en. [82]

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.   [24]

OECD (2023), "Regulatory sandboxes in Artificial Intelligence", *OECD Digital Economy Papers,* No. 356, https://doi.org/10.1787/8f80a0e6-en.   [59]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers,* No. 323, https://doi.org/10.1787/cb6d9eca-en.   [34]

OECD (2022), *The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean*, OECD Publishing, Paris, https://doi.org/10.1787/22190414.   [93]

OECD (2021), *OECD Business and Finance Outlook 2021 : AI in Business and Finance*, https://www.oecd.org/finance/oecd-business-and-finance-outlook-26172577.htm.   [61]

OECD (2021), "State of Implementation of the OECD AI Principles: Insights from National AI Policies", *OECD Digital Economy Papers,* No. 311, https://doi.org/10.1787/1cd40c44-en.   [1]

OECD (2019), *OECD Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.   [65]

OECD.AI (2023), *Database of national AI policies & strategies*, https://oecd.ai/en/dashboards/overview (accessed on 5 October 2023).   [8]

OECD.AI (2023), *Live data: AI jobs and skills*, https://oecd.ai/en/data?selectedArea=ai-jobs-and-skills&selectedVisualization=relative-international-ai-skill-demand (accessed on 13 March 2023).   [127]

OECD.AI (2023), *Live data: AI research*, https://oecd.ai/en/data?selectedArea=ai-research&selectedVisualization=ai-publications-by-country-over-time and https://oecd.ai/en/data?selectedArea=ai-research&selectedVisualization=ai-publication-time-series-by-institution (accessed on 13 March 2023).   [98]

OECD.AI (2023), *Live data: investments in AI and data*, https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data (accessed on 13 March 2023).   [48]

PAI (2023), *Partnership on Artificial Intelligence*, https://partnershiponai.org/.   [131]

Proteccion Data (2022), *AESIA*, https://protecciondata.es/agencia-espanola-supervision-inteligencia-artificial-aesia/.   [97]

Roberts et al. (2021), "Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US", *Science and Engineering Ethics*, https://link.springer.com/article/10.1007/s11948-021-00340-7.   [44]

SADEA (2020), *Memorandum of Understanding between the Government of the Republic of Singapore and the Government of Australia on Cooperation on Artificial Intelligence*, https://www.mti.gov.sg/-/media/MTI/Microsites/DEAs/Singapore-Australia-Digital-Economy-Agreement/MOUs/MOU-on-Cooperation-onArtificial-Intelligence.pdf.   [134]

Salvi del Pero, A., P. Wyckoff and A. Vourc'h (2022), *Using Artificial Intelligence in the workplace: What are the main ethical risks?*, OECD Publishing,   [88]

https://doi.org/10.1787/840a2d9f-en.

Shaughnessy et al. (2023), *Lessons From the World's Two Experiments in AI Governance*, https://carnegieendowment.org/2023/02/14/lessons-from-world-s-two-experiments-in-ai-governance-pub-89035 (accessed on 9 March 2023). [45]

Singapore Government (2023), *Overview of Regulatory Sandbox*, https://www.mas.gov.sg/development/fintech/regulatory-sandbox (accessed on 29 March 2023). [64]

Stanford University (2021), *The AI Index Report*, https://aiindex.stanford.edu/ai-index-report-2021/ (accessed on 28 March 2023). [76]

The Alan Turing Institute (2023), *Women in Data Science and AI*, https://www.turing.ac.uk/research/research-projects/women-data-science-and-ai-new (accessed on 28 March 2023). [68]

The Alan Turing Institute (2022), *UK AI Research Infrastructure Requirements Review*, https://www.turing.ac.uk/work-turing/uk-ai-research-infrastructure-requirements-review. [116]

The White House (2023), *FACT SHEET: Biden-⎕Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/. [36]

Turkish Government (2019), *National Data Dictionary*, https://cbddo.gov.tr/en/project-description/4605/national-data-dictionary- (accessed on 28 March 2023). [94]

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). [27]

UK Government (2023), *Independent Review of The Future of Compute: Final report and recommendations*, https://www.gov.uk/government/publications/future-of-compute-review/the-future-of-compute-report-of-the-review-of-independent-panel-of-experts. [117]

UK Government (2022), *Establishing a pro-innovation approach to regulating AI*, https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement (accessed on 2023 March 2023). [26]

UK Government (2022), *National AI Strategy*, https://www.gov.uk/government/publications/national-ai-strategy (accessed on 29 March 2023). [29]

UK Government (2020), *Review into bias in algorithmic decision-making*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf (accessed on 28 March 2023). [85]

UK Government (2019), *Policy paper: AI Sector Deal*, https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal#further-information. [111]

UK Parliament (2023), *The Data Protection and Digital Information (No. 2) Bill 2022-2023*, https://commonslibrary.parliament.uk/research-briefings/cbp-9746/ (accessed on 28 March 2023).

[30]

US Congress (2022), *Algorithmic Accountability Act (AAA)*, https://www.congress.gov/bill/117th-congress/house-bill/6580 (accessed on 28 March 2023).

[31]

US Department of Energy (2019), *U.S. Department of Energy and Cray to Deliver Record-Setting Frontier Supercomputer at ORNL*, https://www.energy.gov/articles/us-department-energy-and-cray-deliver-record-setting-frontier-supercomputer-ornl.

[118]

US Government (2023), *United States National AI Research Resource (NAIRR)*, https://www.ai.gov/nairrtf/ (accessed on 29 March 2023).

[121]

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

[32]

US Government (2021), *Automated Vehicles Comprehensive Plan*, https://www.transportation.gov/av/avcp (accessed on 29 March 2023).

[49]

US Presidential Office (2019), *Maintaining American Leadership in Artificial Intelligence*, https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence (accessed on 28 March 2023).

[95]

White House (2021), *Quad Principles on Technology Design, Development, Governance, and Use*, https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/24/quad-principles-on-technology-design-development-governance-and-use/#:~:text=Build%20trust%2C%20integrity%2C%20and%20resilience,integrity%2C%20and%20resilience%20foster%20innovation.

[130

World Bank (2023), *Digital Development*, https://www.worldbank.org/en/topic/digitaldevelopment/overview (accessed on 29 March 2023).

[135]

Yamashita, I. and et al. (2021), "Measuring the AI content of government-funded R&D projects: A proof of concept for the OECD Fundstat initiative", *OECD Science, Technology and Industry Working Papers*, Vol. No. 2021/09, https://doi.org/10.

[99]

Yang (2023), *China just set up a new bureau to mine data for economic growth*, https://www.technologyreview.com/2023/03/15/1069814/china-new-bureau-data-economiy/ (accessed on 29 March 2023).

[54]

Zhang (2021), *Provisions on the Administration of Algorithm Recommendation of Internet Information Services in China*, https://www.lexology.com/library/detail.aspx?g=08867c3e-7ded-43d2-af2b-8f09878ef7a6#:~:text=Provisions%20on%20the%20Administration%20of%20Algorithm%20Recommendation%20of,information%20services%20...%203%20V.%20Legal%20liability%20 (accessed on 29 March 2023).

[52]

Annex A.

Table A.1. AI definitions, risk classifications, scope and oversight in select AI-specific regulations

Country/ Institution	Legislation and regulation	AI Definition	Risk classification	Scope & Accountability	Oversight
Canada	Proposed AI and Data Act (AIDA)	"a technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions"	• Classification according to high-impact system. The criteria applicable to the AIDA definition is to be established in regulations. • High-impact AI systems could result in two types of adverse impacts: (1) individual and collective harms; (2) biased outputs.	• Persons – including trusts, joint ventures, partnerships and other legal entities, as well as governmental institutions (except for those defined section 3 of the Privacy Act, as well as specific federal security institutions and provincial ones as prescribed by regulation) – under the following circumstances: "a person is responsible for an artificial intelligence system, including a high-impact system, if, in the course of international or interprovincial trade and commerce, they design, develop or make available for use the artificial intelligence system or manage its operation" – Art. 5(2). • When dealing with high-impact AI systems, obligations of such persons include: to keep general records throughout the lifecycle of these systems; monitor compliance with mitigation measures and their effectiveness; notify the competent authority; and make a description of the system publicly available. • Albeit not expressly	• Minister ("the member of the Queen's Privy Council for Canada designated under section 31 or, if no member is so designated, the Minister of Industry" – Art. 5). The Minister may order the person responsible for high-impact systems to provide records, as well as to conduct an audit and provide the resulting report (Arts. 13 to 15). It may also require, among others, the implementation of measures as well as the discontinuation of the use or availability of a high-impact AI system. • "In addition, the AIDA would create a new statutory role for an AI and Data Commissioner, who would support the Minister in carrying out" its responsibilities.

				prohibiting the use of AI systems that may lead to unacceptable risk, AIDA lists a number of actions as per se criminal offenses, e.g., the possession or use of personal information unlawfully obtained.	
Council of Europe	Proposed Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law	"an algorithmic system or a combination of such systems that, as defined herein and in the domestic law of each Party, uses computational methods derived from statistics or other mathematical techniques to carry out functions that are commonly associated with, or would otherwise require, human intelligence and that either assists or replaces the judgment of human decision-makers in carrying out those functions. Such functions include, but are not limited to, prediction, planning, classification, pattern recognition, organization, perception, speech/sound/image recognition, text/sound/image generation, language translation, communication, learning, representation, and problem-solving"	No classification of risks based on severity levels, as it leaves for each country to determine its priorities in terms of classification system. Instead, it addresses overall "risks and adverse impact resulting from the application of an artificial intelligence system in relation to the enjoyment of human rights, the functioning of democracy and the observance of rule of law"	• Public and private actors indistinctively, which design, develop and apply AI systems, throughout the lifecycle of the latter, where it "involves issues related to human rights, the functioning of democracy and the observance of rule of law". • The Convention is not intended to be applied to the design, development, and application of AI systems "used for purposes related to national defense".	• Each Party shall ensure that adequate oversight mechanisms are set in place within its jurisdiction and in accordance with its domestic law

European Union	Proposed AI Act	Proposed definitions by: • European Commission: "artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" • European Council: "artificial intelligence system' (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts" • European Parliament: "artificial intelligence system' (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments"	Four levels of risk: (1) Unacceptable risk (banned); (2) High risk (strictly regulated); (3) Limited risk (transparency obligations); (4) Minimal or no risk (no intervention).	• Applicable to natural persons, legal persons and public authorities. Specifically, to providers and users/ deployers of AI systems. • A specific set of obligations is established for providers of high-risk AI systems (including manufacturers, importers, distributors), which includes undertaking relevant conformity assessment procedures and taking the necessary corrective actions, among others. • Users/deployers of high-risk AI systems are also subject to obligations, such as to make due use, monitor and maintain documentation available. • AI systems developed or used exclusively for military purposes are outside the scope of the Regulation.	European Artificial Intelligence Board (or Office) to be established to: foster co-operation across national as well as the European Commission; coordinate and provide guidance to authorities; and ensuring consistency in the application of the EU AI Act.
United Kingdom	White Paper "A pro-innovation approach to AI Regulation" (March 2023)	Products and services that are 'adaptable' and 'autonomous'	• No predefined risk categories or risk classification, but rather illustrations of AI risks, such as risks to human rights (e.g., deepfake pornographic video damaging reputation,	• Needs to be further clarified by central government or a central body. • Areas outside of the scope of the regulatory framework include some of "the wider societal and global challenges that may relate to the	• No oversight nor cross-sectoral body will be created at a first moment. Government (not specified which part of it) will monitor the effectiveness of the application of values-based principles by

			relationships and dignity), safety (e.g., physical and/or mental harm), fairness (e.g., biases), privacy and agency, societal well-being (e.g., disinformation that could undermine access to reliable information and trust in democratic institutions and processes) and security (e.g., cyber attacks). • The regulatory framework is context-specific, focusing on the outcomes AI is likely to generate in particular applications rather than assigning rules or risk levels to entire sectors or technologies.	development or use of AI", such as "issues relating to access to data, compute capability, and sustainability, as well as the balancing o the rights of content producers and AI developers".	regulators (which are based on the OECD.AI respective principles). Horizontal collaboration across existing regulators will prevail for the application of an AI regulatory framework on a voluntary basis at a first moment (non-statutory), according to their own discretion. • In a second moment, the government will assess the need to introduce a statutory duty mandating regulators to implement further measures to support enforcement.
United States	Blueprint for an AI Bill of Rights	"An 'automated system' is any system, software, or process that uses computation as whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities. Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing or artificial intelligence techniques, and exclude passive computing infrastructure. "Passive computing infrastructure" is any intermediary technology that does not influence or determine the outcome of decision, make or aid in decisions, inform policy implementation, or collect data or observations, including web hosting, domain registration, networking, caching, data storage, or cybersecurity. Throughout this framework, automated systems that are considered in scope are only those that have the potential to meaningfully impact individuals' or communities' rights, opportunities, or access"	No predefined risk categories; applicability is to be judged based on potential for impact on rights, opportunities, and access to critical needs.	Non-binding proposals; it is left to each authority (whether federal or local) or AI actor who is applying the principles to establish accountability in relation to specific matters they seek to provide guidance or regulate.	The establishment of a horizontal authority for oversight, coordination and monitoring has not been foreseen.
	NIST Risk Management Framework	An AI system is "an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions,	No predefined risk threshold categories. The NIST RMF refers to risks to people (individuals, groups, organisations,	The NIST RMF makes reference to AI designers, developers, deployers and users.	

	recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy".	communities, society, the environment, and the planet.), organisations and systems/ecosystems (e.g., environment), but it does not seek to establish threshold categories nor metrics related to those risks. These are rather left for each organisation that voluntarily choose to apply the NIST RMF to determine using the guidance, suggested actions. and information references provided in the AI RMF Playbook.		

Note: This table is a sample of emerging AI-specific initiatives from select jurisdictions at the time of writing in May 2023. Elements may have changed since and thus it should be taken for illustrative purposes only.

Annex B.

Further aspects of selected AI-specific regulations

Territorial scope

1.   Some regulations expressly place obligations irrespective of the physical location or presence of the persons they intend to reach. The proposed Canadian AIDA is explicit about its extraterritorial reach. It creates responsibilities to persons that design, develop or make available AI systems in the course of international or interprovincial trade. The proposed **EU** AI Act will be applicable to users, manufacturers and providers inside or outside of the Union. Distributors, importers and representatives placing AI products or putting into service AI systems in the Single Market will have to abide to it (European Commission, 2021a[33]).

2.   **China** has mainly focused its regulatory efforts on implementation within its territory. The country's Internet Information Service Algorithmic Recommendation Provisions (Zhang, 2021[48]) specifies on Article 2 that it is applicable within mainland China. The Ethical Norms for New Generation Artificial Intelligence (Center for Security and Emerging Technology, 2021[130]) observe the same territorial scope.

Prohibitions

3.   **Canada**'s proposed AIDA sets from the outset, as one of its purposes, the prohibition of certain conducts in relation to AI systems that may result in serious harm. However, the text does not clarify what the prohibitions consist of. The Companion to AIDA completes this task, clarifying what is prohibited under AIDA and creating three new prohibited conducts under criminal law. In a nutshell, the prohibitions include (Canadian Government, 2023[15]):

* human rights discrimination (that is, biased output of an AI system leading to unjustifiable and adverse impact) as prohibited in the Canadian Human Rights Act;
* reckless and malicious uses of AI that cause serious harms (to Canadians and their interests).

4.   To address these uses, AIDA creates three new criminal law provisions:

* "Knowingly possessing or using unlawfully obtained personal information to design, develop, use or make available for use an AI system";
* "Making an AI system available for use, knowing, or being reckless as to whether, it is likely to cause serious harm or substantial damage to property, where its use actually causes such harm or damage"; or
* "Making an AI system available for use with the intent to defraud the public and or cause substantial economic loss to an individual, where its use actually causes that loss".
* other applicable provisions of the Canadian Criminal Code.

5.   Art. 5 of the proposed **EU** AI Act establishes bans to the use of certain AI systems (European Commission, 2021a[33]). The exact scope is still subject to negotiations before the European Union institutions and has been recently expanded in the Parliament. Currently, prohibitions include AI systems that:

- deploy subliminal techniques beyond a person's consciousness for certain objectives or effects;
- exploits vulnerabilities of a specific group of persons, due to age, disability or specific social or economic situation;
- evaluates or classifies natural persons (social scoring) leading to certain negative effects; or
- carry out 'real-time' remote biometric identification in public spaces, subject to certain exceptions set forth in Art. 5(d).

Non-contractual civil liability

6.   The **EU** opted to address non-contractual civil liability arising from harm caused by AI systems in legislation separate from the AI Act. Accordingly, in September 2022 it published two Directive proposals.

7.   The proposed revision of the Product Liability Directive (European Commission, 2022[131]) intends to update the existing 1985 Product Liability Act, modernizing the rules for the digital economy. It does it by a) clarifying liability for defective products in situations when there is substantial modifications, taking into account the circular economy; b) including provisions on liability for technological products and their updates, such as software and digital services; c) including manufactures outside of the EU in the scope of the Directive, as well as importers of defective products; and d) requiring manufactures to disclose evidence and alleviates the burden of proof to those harmed in certain situations, such as complex cases involving AI systems (European Commission, 2022[131]).

8.   The second Directive proposal under analysis is the newly proposed AI Liability Directive (European Commission, 2022[132]). It provides for damages caused by safety violations or breaches of privacy, as well as for compensation for discrimination in recruitment processes involving AI systems. In addition, it eases the burden of proof for those affected by AI systems to establish causality between the fault and the resulting harm in given circumstances, as well as makes it easier for injured parties to obtain evidence in cases involving high-risk AI (European Commission, 2022[132]).

Sanctions

9.   In **Canada**, the proposed AIDA establishes two types of penalties for non-compliance with the regulation: administrative monetary penalties and prosecution of regulatory offences. In additional, criminal offenses are also punishable under AIDA (Canadian Parliament, 2022[12]).

- The regulator would apply administrative monetary penalties to any type of violations, for the purpose of encouraging compliance with the obligations under AIDA. The government would have to carry out consultations and create new regulations to establish these penalties.
- Regulatory offenses would be applicable to serious cases of non-compliance with obligations. Here, guilt must be demonstrated. The authority responsible for examining whether a regulatory offense is in the public interest is the Public Prosecution Service

of Canada, which retains full autonomy in deciding whether to proceed with a prosecution.

- The Public Prosecution Service of Canada would also be in charge of assessing the public interest and prosecuting criminal offenses based on reckless and malicious uses of AI that cause serious harms.

10.   The proposed **EU** AI Act establishes fines up to 6% of annual worldwide annual turnover for placing a prohibited AI system on the market; up to 4% of worldwide annual turnover for violating requirements for high-risk AI systems; and up to 2% of the worldwide annual turnover for submitting false or incomplete information to authorities (European Commission, 2021a[33]).