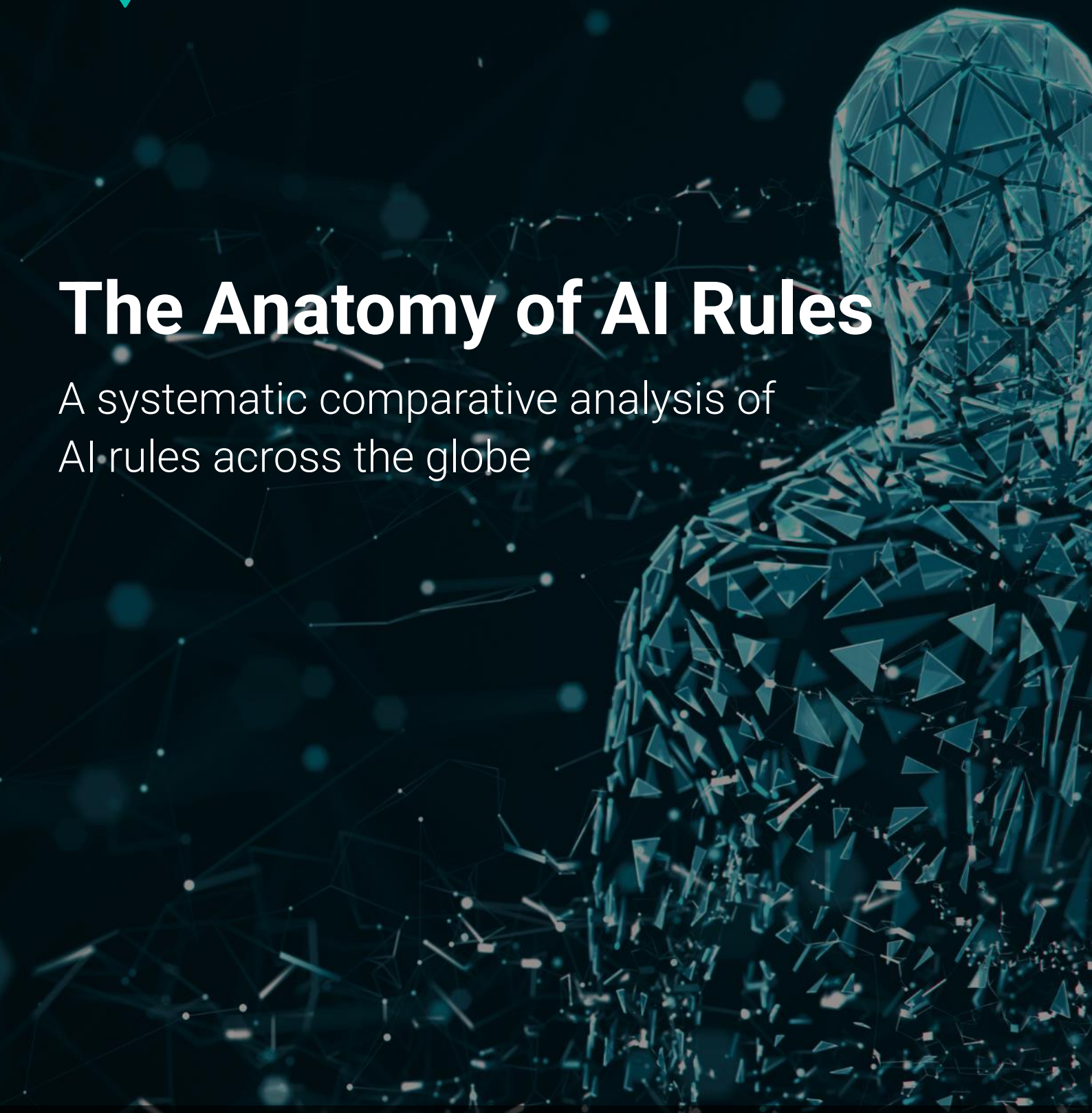
Digital Policy Alert

The Anatomy of AI Rules

A systematic comparative analysis of AI rules across the globe



In collaboration with the

Law and Economics Foundation St.Gallen

An initiative of the

St.Gallen Endowment for Prosperity through Trade

Executive Summary

The growing pace of **AI regulation demands international coordination**. The Digital Policy Alert has documented over 600 regulatory developments targeting AI providers since January 2023. Often, governments share regulatory objectives, such as safety and transparency, enabling international alignment under the OECD AI Principles. To translate international alignment into national AI rules, however, governments need a common factual base. To this end, this report provides a comprehensive comparative analysis of AI rules across the globe.

We find that **AI rules diverge on three layers**. First, governments prioritise the OECD AI Principles differently, focusing mainly on accountability and fairness. Second, when implementing the same OECD AI Principle, governments employ different regulatory requirements, creating a regulatory patchwork. Third, even when governments use the same regulatory requirement to implement the same OECD AI Principle, granular differences create hurdles to interoperability.

Divergence presents an **opportunity for governments**. Building on shared goals, governments can learn from each other and develop effective AI regulation. This report enables such learning with four value propositions:

1. A **common language** for AI rulemakers: We code the text of AI rulebooks across the globe using a single taxonomy, tackling terminological differences.
2. **Detail** for international alignment: We translate the five OECD AI Principles into 74 regulatory requirements and analyse interoperability with unique precision.
3. **Clarity** on current AI rules: We systematically compare the world's 11 most advanced AI rulebooks from seven jurisdictions.
4. **Transparency** for AI rulemakers: We make all our findings accessible through our CLaiRK suite of tools to navigate, compare, and interact with AI rules.

Lead authors

Tommaso Giardini
Johannes Fritz

Contributors

Nora Fischer (Digital Policy Scholar)
Philine Jenzer (Digital Policy Scholar)
Gian-Marc Perren (Digital Policy Scholar)
Nicolà Seeli (Digital Policy Scholar)
Anna Pagnacco

Acknowledgement

This report is the result of a collaboration with the Law and Economics Foundation St.Gallen. Our joint Digital Policy Scholarship develops young talents from the Law and Economics programme of the University of St.Gallen into experts on global digital policy developments.

Table of Contents

The three layers of divergence in AI rules

Divergence in AI rules emerges on three layers. Governments 1) prioritise the OECD AI Principles differently, 2) use different policy areas to implement the same OECD AI Principles, and 3) establish idiosyncratic regulatory requirements within the same policy areas.

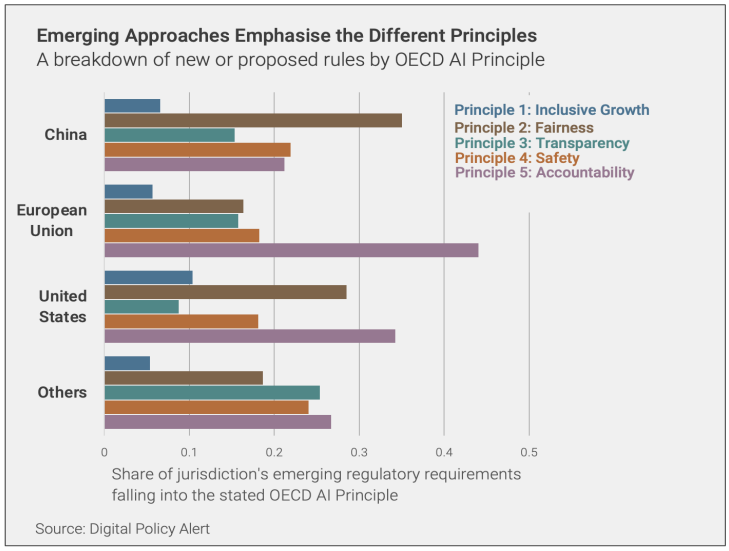
Governments prioritise different OECD AI Principles

The **global flurry of AI regulation** presents both an opportunity and a challenge. On the one hand, the diversity of regulatory approaches could spur governments to learn from each other in a new regulatory field, leading to more effective AI regulation. On the other hand, there is a considerable risk of creating a fragmented regulatory landscape, reminiscent of current data transfer rules. Fortunately, this dichotomy has catalysed a notable willingness among governments to coordinate on AI rules. The problem governments face, though, is what exactly to coordinate on.

Bridging international coordination and national regulation

International alignment on AI rules demands abstraction, as evidenced by the widely recognised **OECD AI Principles' lack of prescriptive detail**. The OECD AI Principles are high-level by design and advocate for AI technology that (1.1) promotes inclusive growth, (1.2) respects human rights and fairness, (1.3) ensures transparency and explainability, (1.4) maintains robustness and safety, and (1.5) enforces accountability. To effectively draw lessons from regulation abroad and promote interoperable AI regulation, governments need a high-resolution view of the regulatory landscape.

The Digital Policy Alert can now provide **clarity on emerging AI rules**, building on an unprecedented comparative analysis. Our team meticulously analysed 11 comprehensive AI rulebooks from Argentina, Brazil, Canada, China, the European Union, South Korea, and the United States. Paragraph by paragraph, we tagged every provision with our novel taxonomy of over 70 regulatory requirements. This rigorous, text-based analysis offers a comprehensive and detailed snapshot of the current state of emerging AI regulation, revealing both commonalities and disparities across borders. Moreover, we mapped each regulatory requirement into an OECD AI Principle to investigate governments' high-level priorities.



The high-level comparison reveals **how countries prioritise different OECD AI Principles**. Accountability is a universally shared priority, commanding a significant share of AI rules across all jurisdictions. The EU AI Act devotes over 40 percent of its requirements to this OECD AI Principle, while in the United States over 30 percent of the requirements pursue accountability. Fairness and safety are also global priorities, albeit less salient than accountability. China dedicates over 30 percent of its requirements to fairness, surpassing all other jurisdictions. Safety is a common priority, to which approximately 20 percent of requirements are devoted across jurisdictions. Transparency is emphasised most strongly outside the three big economic powers, covering over 25 percent of requirements. Finally, inclusive growth is currently the least salient OECD AI Principle, featured most prominently in the United States (slightly over 10 percent of requirements).

The challenge and opportunity of granular differences

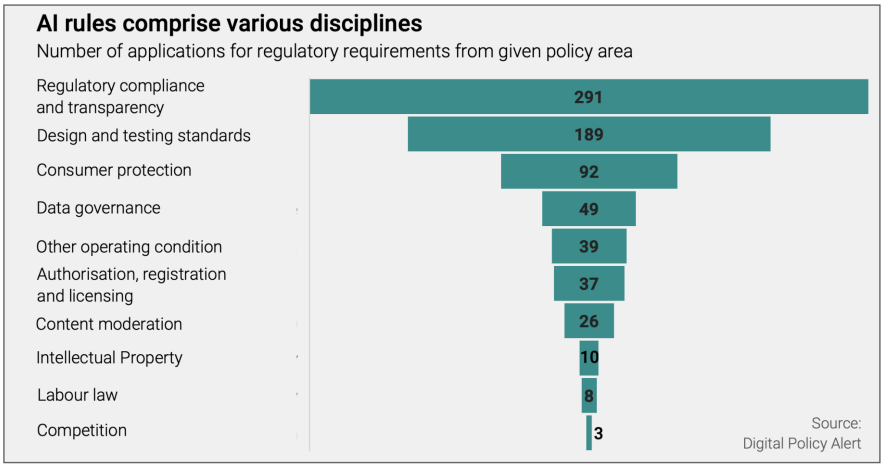
National differences in the prioritisation of the OECD AI Principles are only the **tip of the iceberg**. Even in the pursuit of the same OECD AI Principle, governments employ different regulatory requirements, from various policy areas. Since divergence – at all levels of granularity – is rising, it is imperative to learn from alternative approaches and pursue international coordination. To this end, we now move to the next layer, analysing which requirements governments use to implement each OECD AI Principle.

Governments use different policy areas for each OECD AI Principle

Governments draw from **ten different policy areas** to establish AI rules and impose different requirements to operationalise each OECD AI Principle. This presents a unique opportunity to learn from diverse regulatory approaches.

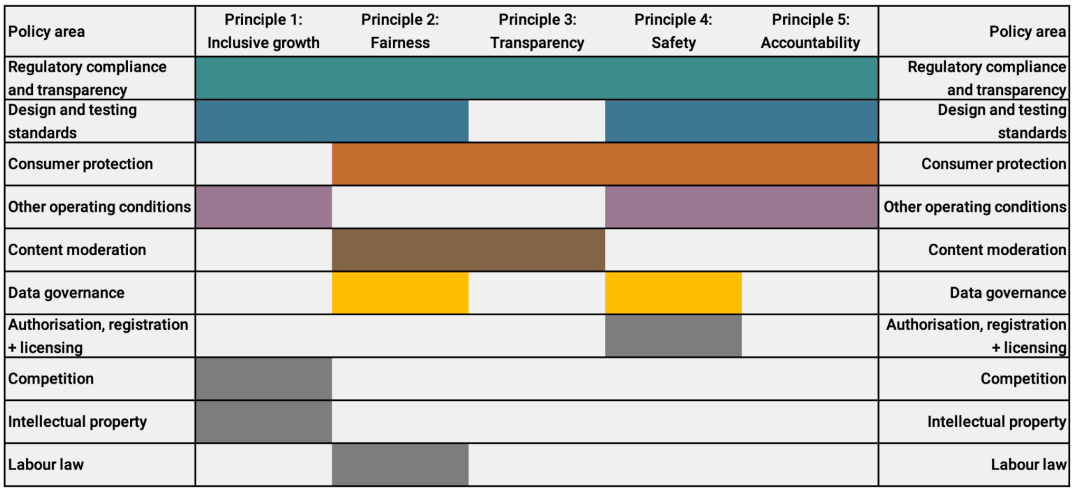
AI rules draw from ten different policy areas

Our analysis of 11 AI rulebooks reveals that **AI rules are not a single, delineated policy area**, but rather draw from almost a dozen existing ones. Over half of the total requirements (744) concern either regulatory compliance and transparency, or design and testing standards. Less frequently used policy areas include consumer protection, data governance, and content moderation. AI rules are diverse because AI is a multifaceted technology. Data governance rules regulate the data with which AI is trained and protect each AI user's privacy. Content moderation rules set guardrails for AI-generated output. Transparency rules address the opacity of AI systems. While these policy areas all pursue legitimate objectives, their interplay complicates international alignment.



Multiple policy areas intersect within each OECD AI Principle

When governments operationalise the OECD AI Principles, they **combine regulatory requirements from different policy areas**. To implement the principle of human rights and fairness (1.2) as well as safety (1.4), governments draw from six policy areas. The principle of transparency and explainability (1.3) is implemented through rules on regulatory compliance and transparency, consumer protection, and content moderation. The rules implementing the other principles span across at least four policy areas.



In turn, **several policy areas implement multiple OECD AI Principles**. For instance, regulatory compliance and transparency are relevant to all five principles. Design and testing standards as well as consumer protection are pertinent to the implementation of four principles. Content moderation and data governance are pertinent to the implementation of two principles. Other policy areas implement only one principle, namely competition, intellectual property, and labour law.

AI rules create a risk of multidimensional divergence

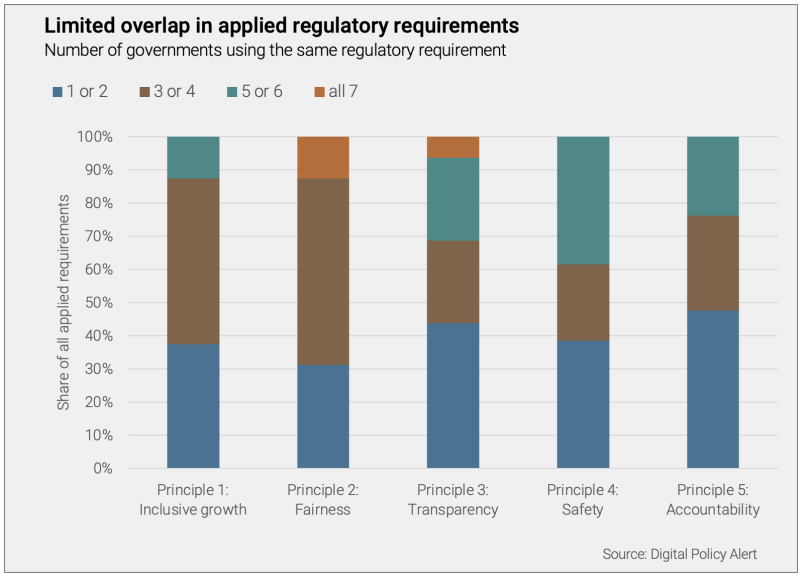
The diversity of AI rules creates **risk for divergence in the implementation of the OECD AI Principles** on three layers. For example, multidimensional divergence is visible in how governments implement the principle of respect for the rule of law, human rights and democratic values (1.2).

1. China and the United States emphasise this OECD AI Principle more than other governments.

2. Some governments establish rules regarding data governance, such as data protection requirements. Other governments demand consumer protection, for example through non-discrimination obligations.

3. Even within these policy areas, a patchwork of divergent requirements emerges. Within data protection, some governments establish data subject rights while others focus on data security requirements. Within non-discrimination, some governments establish rights to contest discriminatory AI output, while others impose prohibitions on discriminatory AI systems.

Multidimensional divergence, across the OECD AI Principles, is evidenced by **how rarely a single regulatory requirement is used across borders**. Our comparative analysis found 74 different regulatory requirements, applied a total of 744 times across the seven studied jurisdictions. Only three requirements – regarding data protection, non-discrimination, and the disclosure of technical documentation – are featured in all the jurisdictions we studied. In contrast, over a third of all regulatory requirements are foreseen in only one or two jurisdictions.



The opportunity to coordinate AI rules resembles multidimensional chess

Governments working towards international alignment on AI rules face a **unique opportunity**. The diversity of AI rules enables governments to learn from both previous experience and each other. Governments can draw from their experience in other policy areas, including the expertise accumulated by national regulators. In addition, governments are currently experimenting to find effective AI rules. Studying and comparing different approaches to operationalising the OECD AI Principles is an opportunity for rapid learning.

The **urgency for international alignment** on AI rules is underestimated: Multidimensional divergence on AI rules can amplify digital fragmentation risk. Currently, the global digital economy is struggling with different national rules regarding data transfers. Concerning AI, such differences multiply since they can occur within each pertinent policy area. International coordination is imperative to avoid fragmentation.

When governments pursue the **coordination of AI rules**, they should approach it like a game of multidimensional chess:

- Understand how the pieces move, by knowing the relevant policy areas in AI rules.
- Be aware of all the dimensions, by differentiating between the high-level OECD AI Principles and the granular requirements that implement them in national AI rules.
- Know their counterparts, by studying and learning from national regulatory approaches.

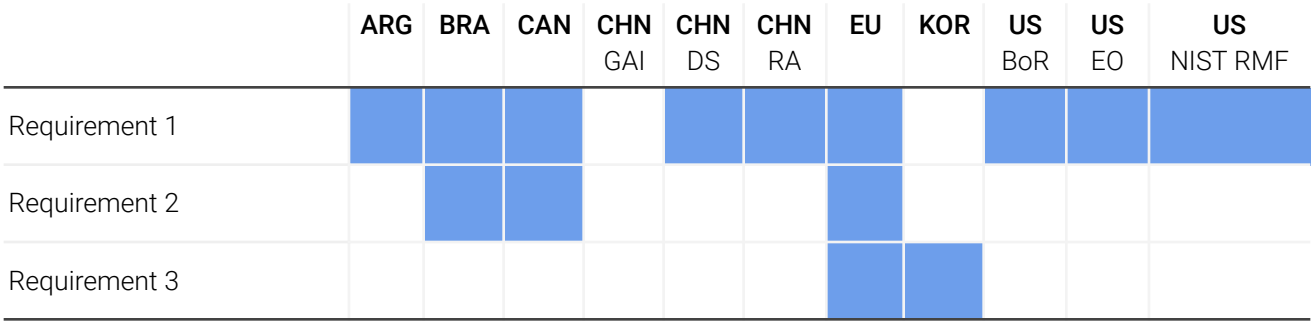
Understanding granular differences: How to read this report

Having established that countries prioritise different OECD AI Principles and use different regulatory requirements to implement the same OECD AI Principles, we move to the **granular layer**. The following sections dissect the implementation of four OECD AI Principles.

- First, we outline the regulatory requirements that implement the OECD AI Principle and note which governments employ which requirements.
- Then, we compare the details of every requirement, comparing the text of each AI rulebook that contains a given requirement.

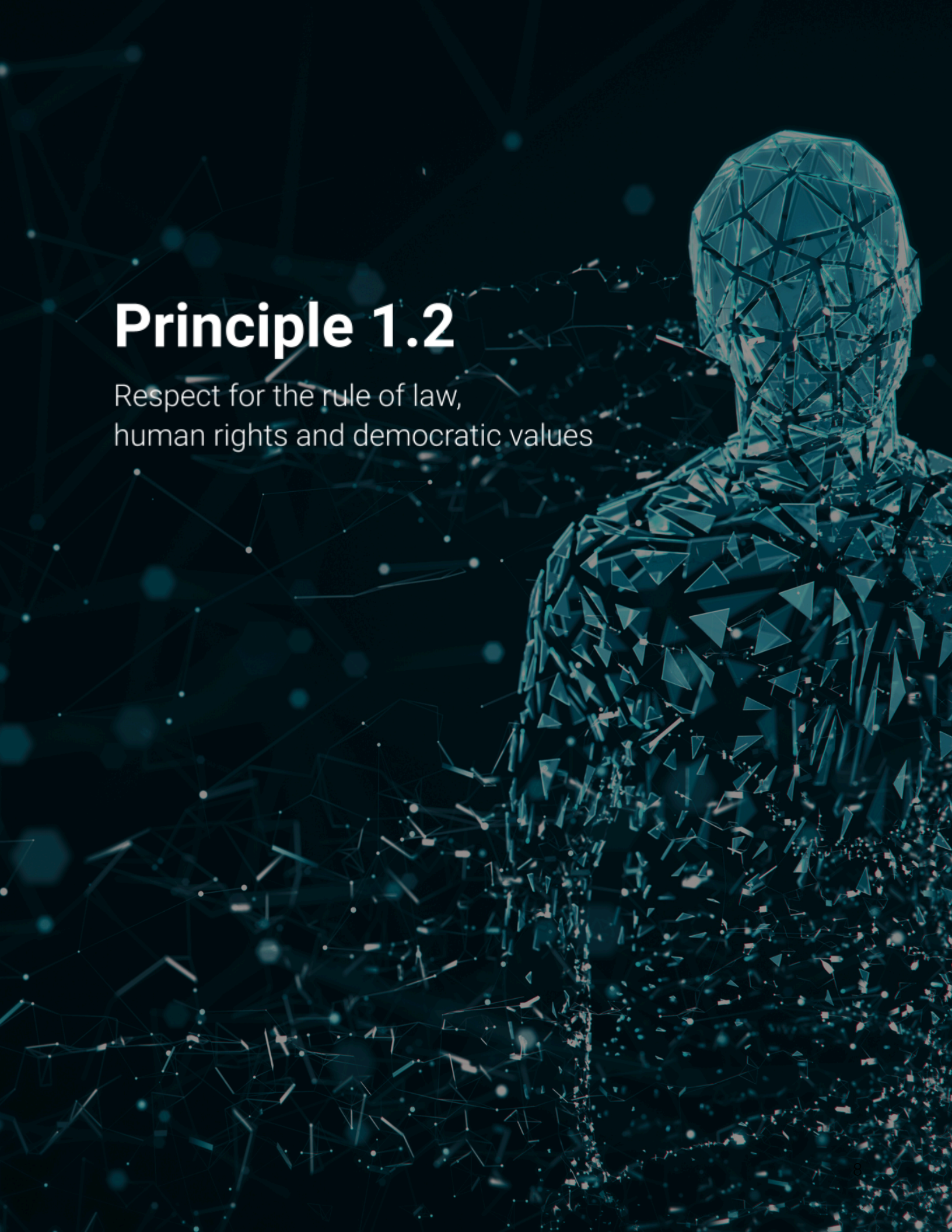
For every finding, we **link to the relevant text passage** of the AI rulebooks using our CLaiRK interface. This approach empowers readers to select their topics of interest, receive an overview of regulatory approaches, and dive into the details of the relevant legal text with unprecedented ease.

Heatmaps are the main visual tool of this report. All heatmaps share the same basic layout: The horizontal axis enumerates the AI rulebooks,[1] while the vertical axis lists different AI regulatory requirements. A highlighted field signals that the requirement exists in the given AI rulebook. We recommend reading each heatmap first vertically, to familiarise yourself with the requirements in focus, then horizontally, to compare requirements across jurisdictions.



This approach **empowers readers** to select their topics of interest, receive an overview of regulatory approaches, and dive into the details of the relevant legal text with unprecedented ease.

[1] The Annex lists the analysed rulebooks. Heatmaps feature China's regulations on generative AI ("GAI"), deep synthesis services ("DS") and recommendation algorithms ("RA"), from left to right. For the United States, we feature the Blueprint for an AI Bill of Rights ("BoR"), the Executive Order on AI ("EO"), and the NIST Risk Management Framework ("NIST RMF").



Principle 1.2

Respect for the rule of law,
human rights and democratic values

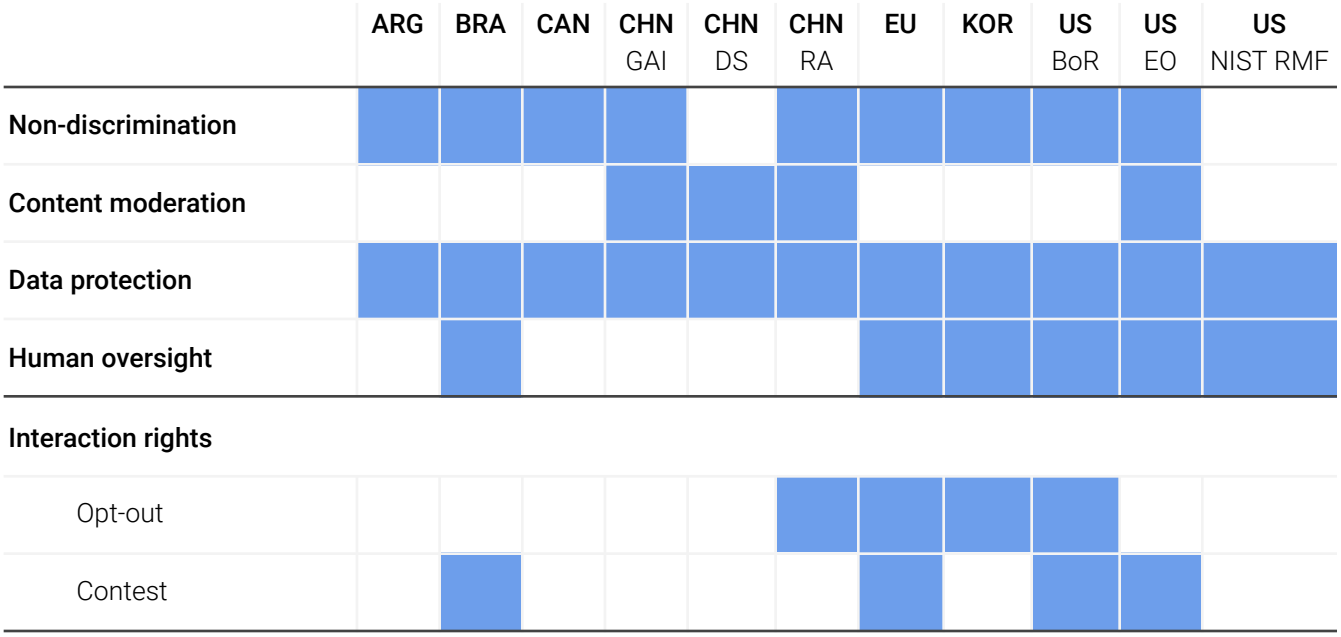
Principle 1.2: Respect for the rule of law, human rights and democratic values

As AI permeates into all areas of life, governments around the world worry about the **erosion of human values**. Ceding human agency to AI can both create new problems and exacerbate existing problems – from algorithmic discrimination, to AI privacy breaches, to AI-generated misinformation. This flurry of regulatory concerns has led different governments to similarly demand respect for the rule of law, human rights, and democratic values. Governments differ, however, in the regulatory requirements they choose to impose in pursuit of this shared goal.

A patchwork of regulatory requirements implements OECD AI Principle 1.2

The **OECD AI Principle 1.2** demands that AI actors should respect the rule of law, human rights, as well as democratic and human-centred values throughout the AI system lifecycle. The principle specifically lists non-discrimination, freedom, dignity, autonomy, privacy, diversity, fairness, social justice, and labour rights. In addition, actors should address AI's amplification of misinformation while respecting freedom of expression. To pursue this goal, AI actors should implement safeguards, such as human oversight, and also address risks arising from uses outside of intended purpose and un-/intentional misuse.

In national AI rules, a patchwork of regulatory requirements implements the OECD AI Principle 1.2. The heatmap visualises divergence within a selection of these requirements. Below, we explain each requirement in detail.



Non-discrimination is a common regulatory requirement

Non-discrimination requirements aim to address the concern of **algorithmic discrimination**. AI can create new forms of discrimination and perpetuate existing discriminatory practices. Hence,

non-discrimination requirements oblige AI providers not to discriminate between users of their AI systems and to avoid that their AI systems have discriminatory effects.

Non-discrimination requirements are widespread, since they are prevalent in each analysed jurisdiction (albeit not in all analysed rulebooks). Differences persist, however, regarding the definition of discrimination and the remedies against discrimination – from strict prohibitions to compliance obligations.

Content moderation is rarely required

Content moderation requirements address concerns that **AI-enabled content generation** facilitates the creation of illegal or harmful content. Discussions on free online speech have grown ever since user-generated content platforms enabled users to directly disseminate content. AI-enabled content generation merely raises the salience of this concern, prompting governments to impose requirements on what output generated by an AI system is permissible.

Content moderation requirements are rarely imposed, namely in China's regulations on generative AI, deep synthesis services, and recommendation algorithms, as well as the US Executive Order. Granular differences concern the type of content that is to be moderated and the procedure for moderation, including redressal mechanisms.

Data protection requirements are ubiquitous

Data protection requirements aim to prevent **privacy violations through AI systems**. Privacy concerns arise throughout AI systems' lifecycle: Personal data can be used to train AI systems, appear in the output of AI systems, and be inferred by AI systems from user interactions.

Data protection requirements are **prevalent in all the analysed AI rulebooks**. These requirements range from rules on AI providers' handling of data in the development of AI to user rights regarding the processing of their personal data. Differences persist, however, regarding the novelty of rules, ranging from references of existing frameworks, specifications thereof, and newly established rules. In addition, the specific obligations for AI providers vary significantly across borders.

Human oversight are frequently mandated

As many of the regulatory concerns regarding AI arise from **humans handing over agency to AI**, governments establish human oversight requirements. Such requirements demand that humans can oversee or interfere in AI decision making processes, aiming to improve capacity for human agency and oversight.

Human oversight requirements are regularly imposed, namely in six AI rulebooks, three of which stem from the US. Granular differences persist regarding the required extent of human oversight and the qualification criteria for the responsible humans.

Interaction rights are scarcely used

Interaction rights empower users to **interfere with the use of AI systems**. Governments primarily establish two kinds of interaction rights:

- The right to object to AI-assisted decisions enables users to refuse to be subjected to decisions made by AI systems.
- The right to contest AI-assisted decisions empowers users to challenge decisions made by AI.

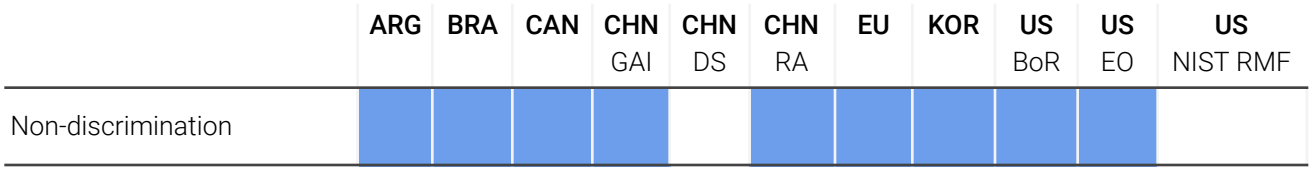
Interaction rights are rare, featuring in four AI rulebooks. Differences arise regarding the specific execution of user rights. For instance, the right to object can cover the general use of an AI system or specific components thereof (user labels).

Dive deeper into each requirement

The patchwork of regulatory requirements that implement OECD AI Principle 1.2 is only the **tip of the iceberg**. Granular differences emerge even within the jurisdictions that impose the same regulatory requirements. To showcase granular divergence, we now proceed with a detailed comparative analysis of the abovementioned requirements. Jump directly to the section that interests you:

Non-discrimination

The OECD AI Principle 1.2 (Respect for the rule of law, human rights and democratic values) demands that AI actors **emphasise non-discrimination** throughout the AI system's life cycle. Although governments share the objective of avoiding AI discrimination, their definitions and regulatory approaches differ, ranging from prohibitions to specific regulatory requirements. This article systematically analyses non-discrimination requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises the jurisdictions that foresee non-discrimination requirements.



Comparison

AI rulebooks differ regarding the **definition of discrimination**. Brazil, Canada, the EU, and South Korea, and the US Bill of Rights offer explicit definitions of discrimination. These definitions overlap in considering discrimination to occur when AI systems lead to unjustified differential treatment or adverse impacts based on factors such as ethnicity, religion, age, or social status. Argentina, China's regulations on generative AI and recommendation algorithms, and the US Executive Order regulate discrimination without explicitly defining it.

Regulatory approaches to tackle discrimination further differ in how they employ **prohibitions and regulatory requirements**. Argentina and South Korea explicitly prohibit discrimination. Canada, China (regulations on generative AI and recommendation algorithms), the EU, and the US (Bill of Rights and Executive Order) address discrimination through specific regulatory requirements, for instance data governance and system design measures. Brazil combines both approaches, explicitly prohibiting discriminatory AI systems and additionally imposing requirements to prevent discrimination.

Finally, non-discrimination rules differ regarding the **AI systems they address**. Argentina, South Korea, and the US Bill of Rights address all AI systems. Rules in Canada and the EU target only high risk AI. The Chinese rules concern specific technologies, namely generative AI and recommendation algorithms. The US Executive Order covers discrimination based on AI use cases, including criminal justice and the federal government, as well as the "broader economy." Brazil provides both general provisions and rules for high risk AI systems and biometric systems.

Country details

Argentina

Argentina prohibits the use of AI for discriminatory purposes, without providing an explicit definition thereof. Additionally, the Artificial Intelligence Supervision Authority is empowered to require AI providers to adopt specific measures to address discrimination and bias, including the suspension of the system in case of non-compliance. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil provides a detailed definition of discrimination and includes both a prohibition and specific regulatory requirements. Brazil considers discrimination to comprise any distinction, exclusion, restriction that limits freedom based on personal characteristics such as race, colour, gender, or religion. The definition also covers indirect discrimination, for instance when apparently neutral practices or criteria have the potential to disadvantage specific groups.

Brazil prohibits the implementation and use of AI systems that may result in direct, indirect, illegal or abusive discrimination. Brazil foresees exceptions, however, for differentiation that is based on a demonstrated objective and justification, as well as reasonable and legitimate in view of fundamental rights.

Brazil further imposes several regulatory requirements aimed to prevent discrimination, establishing non-discrimination as a principle and user right. Providers and operators of high risk AI systems must implement data management measures to mitigate and prevent discriminatory biases. These measures include evaluating data regarding human bias, avoiding bias generation from faulty data collection, and taking corrective action to prevent AI from perpetuating and amplifying structural biases. In addition, individuals impacted by AI systems have the right to comprehensive information regarding non-discrimination measures before using an AI system. [Check the specific provisions on CLaiRK↗]

Canada

Canada defines discrimination and imposes specific regulatory requirements to prevent it. Canada considers AI output and decisions to be biassed when they unjustly discriminate against individuals (directly or indirectly) based on characteristics prohibited by law. The use of such characteristics, however, is lawful if it prevents disadvantages stemming from them.

Canada imposes several regulatory requirements to prevent discrimination. Responsible individuals overseeing high impact AI systems must establish measures to identify, assess, and mitigate biassed output. Responsible individuals must further establish mechanisms to monitor compliance and maintain records. If the use of a high impact AI system could result in biassed output, the government can review records, mandate audits, and impose additional measures. [Check the specific provisions on CLaiRK↗]

China

China's regulations on generative AI and recommendation algorithms include requirements to prevent discrimination, without providing an explicit definition thereof.

China's regulation on generative AI requires providers to implement effective measures to prevent discrimination based on race, ethnicity, beliefs, nationality, origin, gender, age, occupation, or health, among others. Providers must implement these measures throughout the AI life cycle, including algorithm design, training data selection, model generation and optimisation, and the provision of services. In addition, the regulation on generative AI prohibits the generation of artificial content containing ethnic discrimination. [Check the specific provisions on CLaiRK↗]

China's regulation on recommendation algorithms generally mandates providers to respect the principles of justice and fairness. In addition, specific regulatory requirements for providers that sell goods and services to consumers aim to enable fair transactions. Namely, providers shall not use algorithms to impose unreasonable differential treatment in prices and other transaction conditions based on consumer preferences, transaction habits and other characteristics. [Check the specific provisions on CLaiRK↗]

European Union

The EU provides a definition of discrimination and imposes both a prohibition and several regulatory requirements to prevent discrimination. The EU outlines that non-discrimination means that AI systems are developed and used in a manner that avoids discriminatory impacts and unfair biases prohibited by EU or national law. Specifically, the EU considers the risk of such biassed results and discriminatory effects to be particularly relevant regarding age, ethnicity, race, sex or disabilities.

The EU prohibits social scoring systems that evaluate or classify natural persons based on their social behaviour or personality characteristics and lead to detrimental or unfavourable treatment of these persons. Specifically, this treatment shall not occur in social contexts that are unrelated to the social scoring context and should not be unjustified or disproportionate in view of the social behaviour and its gravity.

The EU's regulatory requirements to prevent discrimination apply to high risk AI systems. The training, validation, and testing datasets of high risk AI systems using "model training techniques with data" must undergo data governance and management practices. This includes an examination for potential biases that could result in discrimination. Additionally, high risk AI systems that continue learning after market deployment must be developed to eliminate or reduce the risk of biassed outputs influencing future input (feedback loops). In addition, providers' technical documentation must contain detailed information about the monitoring, functioning and control of the AI system, including risk of discrimination. Finally, oversight authorities enforcing the right to non-discrimination concerning high risk AI systems can request and access AI providers' documentation. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea states that discrimination through algorithmic distortion is a social concern that motivates its AI rulemaking, albeit without providing a detailed definition. The development and use of AI are not allowed to discriminate against individuals or groups based on gender, age, ethnicity, religion, social status, economic situation, or political views. Moreover, AI businesses must ensure fair protection of user rights during AI development or use, and take proactive measures to provide redress in case of user harm. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI emphasises non-discrimination as a core principle and instructs various government agencies to address AI discrimination in various contexts.

- Regarding the criminal justice system, the Attorney General must initiate discussions to prevent and address AI discrimination.

- Regarding the "broader economy," several agencies are to issue guidance and consider rulemaking to address discrimination in hiring, housing, and consumer financial markets.

- Regarding the health sector, the Department of Health and Human Services is to develop a strategic plan, including the use of representative datasets, the monitoring of algorithmic performance for discrimination and bias, and the identification and mitigation of discrimination and bias.

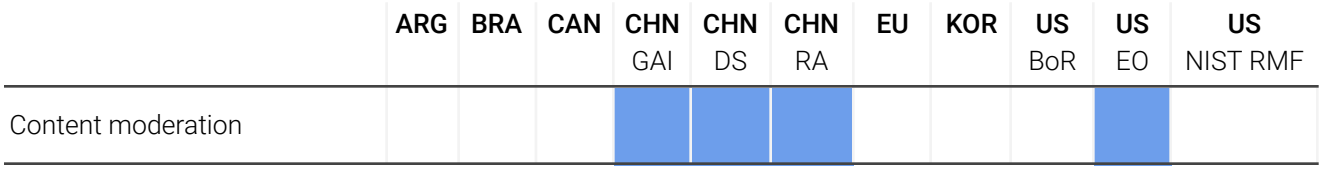
- Regarding the administration of public benefits via AI, several agencies are to issue guidance on AI to prevent discrimination. Regarding the government's use of AI, the Office of Management and Budget is to establish risk management practices including the mitigation of discrimination.

In this effort, the Executive Order references the Blueprint for an AI Bill of Rights and the US NIST Risk Management Framework, which requires the evaluation and documentation of fairness and bias. Finally, the Executive Order encourages agencies to use their authorities to safeguard consumers from discrimination. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights defines discrimination and provides voluntary requirements to prevent it. Discrimination is present when automated systems contribute to unjustified differential treatment or adversely affect individuals based on characteristics such as race, colour, ethnicity, or sex. The Bill of Rights calls for designers, developers, and implementers of automated systems to take proactive and continuous measures to protect individuals and communities from discrimination. Such measures include proactive equity assessments, the use of representative data, protection against proxies for demographic features, as well as disparity testing and mitigation. [Check the specific provisions on CLaiRK↗]

Content moderation

The OECD AI Principle 1.2 (Respect for the rule of law, human rights and democratic values) demands that AI actors address **misinformation and disinformation amplified by AI**. The spread of AI-generated content has exacerbated many regulatory concerns regarding online content that contravenes human values. To address this issue, governments require AI providers and platforms that host their content to moderate certain output. This article systematically analyses content moderation requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which AI rulebooks mandate content moderation.



Comparison

Content moderation requirements differ primarily regarding the **types of content** that must be moderated. The US Executive Order focuses primarily on child-sexual abuse material and non-consentual intimate images, but also contains a provision on discriminatory, misleading, inflammatory, unsafe, or deceptive AI output. The Chinese regulations all demand the moderation of content that is illegal or undermines the state. In addition, the regulations on deep synthesis services and on recommendation algorithms require the moderation of news content. Finally, the regulation on generative AI mandates the moderation of "false" content, while the regulation on recommendation algorithms requires the moderation of content for minors.

Moreover, content moderation requirements differ regarding the types of **AI systems they address**. The US Executive Order covers synthetic content in general and the use of generative AI in the federal government. China's regulations regulate generative AI, deep synthesis services, and recommendation algorithms, respectively. Notably, China's rules thus extend beyond moderating the generation of synthetic content, to the dissemination of such content via recommendation algorithms.

Finally, differences arise regarding the specific **methods prescribed to implement content moderation** requirements. The US Executive Order does not provide detail on how content is to be moderated, rather requiring government agencies to investigate techniques in a report and guidance. China's regulations specify precautions to be taken by providers, including the establishment of libraries of prohibited content, as well as manual and automatic content detection, and reporting. Furthermore, China's regulations outline steps to be taken upon the identification of content that is to be moderated, including content removal and reporting. Finally, the regulation on deep synthesis services further instructs providers on how to address users that contravene rules on illegal content, through warnings, restrictions, and account suspensions.

Country details

China

The Chinese regulation on generative AI mandates the moderation of content that undermines the state, as well as "harmful" and "false" content. Generated content shall adhere to core socialist values and shall not subvert state power, endanger national security, damage the national image, or undermine national unity. Regarding harmful content, generated content shall not promote terrorism, extremism, ethnic hatred, ethnic discrimination, violence, obscenity or pornography. Finally, the regulation prohibits the dissemination of false information. When generative AI providers find illegal content, they must promptly cease generation and transmission, eliminate the content, optimise the model training, and report to relevant authorities. [Check the specific provisions on CLaiRK↗]

The Chinese regulation on deep synthesis services demands the moderation of illegal content and news content. The regulation prohibits organisations and individuals from using deep synthesis services to produce, copy, publish, or disseminate information prohibited by law. Furthermore, the regulation prohibits deep synthesis service providers and users from producing, copying, publishing, or disseminating "fake news information." To moderate content, deep synthesis service providers and "technical supporters" must manually or automatically review users' input data and synthesis results, and establish a library of illegal and harmful content. If providers find illegal or harmful content, they must store records, report to authorities, and adopt measures against users, including warnings, restrictions, suspensions, and account closures. When providers find that their services are used to produce false information, they must further "dispel rumours." [Check the specific provisions on CLaiRK↗]

The Chinese regulation on recommendation algorithms requires content moderation regarding illegal information, news, and minor protection. Specifically, providers shall not disseminate information prohibited by law or generate "fake news" information. In addition, providers are prohibited from recommending information to minors that may cause them to imitate unsafe behaviour, violate social morality, or induce bad habits, including internet addiction. To this end, providers must strengthen content management, including manually and automatically identifying illegal content. When providers find illegal information, they must cease dissemination, prevent future dissemination, and report to authorities. In addition, the dissemination of artificially generated content is to be halted until the content is correctly watermarked. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI instructs several agencies to address the moderation of synthetic content, specifically child sexual abuse material and non-consensual intimate images. The Secretary of Commerce is instructed to submit a report on standards to prevent AI systems from generating child sexual abuse material or non-consensual intimate imagery of real individuals. Furthermore, the Office of Management and Budget is instructed to issue guidance on government use of AI. Regarding generative AI, the guidance is to include testing procedures and safeguards to prevent outputs that are discriminatory, misleading, inflammatory, unsafe, or deceptive, as well as child sexual abuse material and non-consensual intimate images. [Check the specific provisions on CLaiRK↗]

Data protection

The OECD AI Principle 1.2 (Respect for the rule of law, human rights and democratic values) demands that AI actors **protect privacy**. AI can affect privacy throughout its lifecycle: Personal data can be used to train AI systems, appear in the output of AI systems, and be inferred by AI systems from user interactions. This article systematically analyses data protection requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises the jurisdictions imposing data protection requirements.

	ARG	**BRA**	**CAN**	**CHN** GAI	**CHN** DS	**CHN** RA	**EU**	**KOR**	**US** BoR	**US** EO	**US** NIST RMF
Data protection											

Comparison

Data protection requirements differ primarily regarding the **AI systems they address**, ranging from requirements for all AI systems to risk- and technology-specific rules. Argentina, Brazil, Canada, South Korea, and the US Bill of Rights and NIST Risk Management Framework demand data protection from all AI providers. The EU's provisions apply to high risk AI systems and deployers of emotion recognition systems and biometric categorisation systems. China's rules apply to providers of generative AI, deep synthesis services, and recommendation algorithms, respectively. The US Executive Order contains provisions on privacy regarding AI use in the federal government and the health sector

Data protection requirements further differ in their **novelty**, including references of existing frameworks, specifications thereof, and novel rules. The Chinese regulation on recommendation algorithms and South Korea merely state that providers must uphold existing privacy rules. Argentina, Brazil, China's regulations on generative AI and deep synthesis services, and the EU reference existing data protection frameworks and specify their application to AI, for instance regarding training, validation and testing datasets, or consent. Canada imposes new measures regarding data anonymisation and establishes a new criminal offence regarding privacy violations in the context of AI. Finally, in the absence of federal privacy legislation, the US establishes novel provisions on data protection and AI, on a voluntary basis.

Country details

Argentina

Argentina requires all AI systems to respect and protect the privacy of users and process personal data in accordance with the applicable data protection framework. Argentina prohibits the unauthorised use of personal data collected by AI systems and obliges providers to obtain the informed consent of individuals for the use of their data. In addition, Argentina requires AI system developers to implement ethical design and development practices, considering privacy, as well as social responsibility, equity, security, and transparency. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil imposes data protection requirements as part of the governance measures that apply to all AI agents. Data processing must occur in accordance with the existing data protection framework, including measures to ensure privacy by design and privacy by default, as well as techniques to minimise the use of personal data. [Check the specific provisions on CLaiRK↗]

Canada

Canada establishes data protection measures for persons engaging in "regulated activities," including AI providers and individuals who process data or make it available for AI development and use. Specifically, Canada imposes measures regarding both how data is anonymised and how anonymised data is used, without further specifications. Furthermore, Canada establishes the possession and use of personal information that is knowingly obtained through criminal activities for the purpose of designing, developing, using or making available AI systems as a criminal offence. [Check the specific provisions on CLaiRK↗]

China

China's regulation on generative AI requires providers to obtain individual consent when using personal information and comply with the existing personal information protection framework. Providers shall not collect non-essential personal information and shall protect users' input information and usage records. Specifically, providers must refrain from retaining information that can identify users and shall not disclose user information to others. [Check the specific provisions on CLaiRK↗]

China's regulation on deep synthesis services requires providers to establish safe and controllable technical safeguards for personal information protection. Providers and "technical supporters" must further strengthen the management and safety of their systems' training data. If training data contains personal information, the regulation demands compliance with the existing personal information protection framework. [Check the specific provisions on CLaiRK↗]

China's regulation on recommendation algorithms requires systems to comply with the existing personal information protection framework and establishes personal information protection as a duty of algorithm providers. [Check the specific provisions on CLaiRK↗]

European Union

The EU imposes several data protection requirements for high risk AI systems which "make use of techniques involving the training of AI models with data." Such systems must be developed using training, validation and testing datasets that are subject to data governance and management practices. Specifically, these practices concern design choices, data collection processes and data origin, as well as the purpose of data collection. In addition, the practices include data preparation operations, such as data annotation and labelling, the assessment of the availability, quantity and suitability of the datasets, as well as bias examination.

In addition, deployers of emotion recognition systems and biometric categorisation systems must process personal data in accordance with the existing data protection framework, with the exception of systems permitted by law to detect, prevent and investigate criminal offences. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea states that personal information shall be governed by the Personal Information Protection Act, without providing further detail. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI dedicates a section to the protection of privacy, with provisions on the use of commercially available information (CAI) and privacy-enhancing technologies (PETs).

- Regarding CAI, the Executive Order instructs the Office of Management and Budget to guide agencies on the mitigation of privacy risks when using CAI that contains personally identifiable information. In addition, the Attorney General is to issue a request for information regarding potential revisions to the guidance for agencies to implement the privacy provisions of the E-Government Act of 2002, including through privacy impact assessments.

- Regarding PETs, the Executive Order instructs the Secretary of Commerce to draft guidelines that enable agencies to use PETs to safeguard privacy, including the evaluation of differential-privacy-guarantee protections for AI. Moreover, the Executive Order contains measures to advance PET research, development, and implementation, including funding for a Research Coordination Network dedicated to advancing privacy research.

In addition, the Executive Order contains provisions on data protection in the health sector, federal government use of AI, and international standards.

- The Department of Health and Human Services is to develop a strategic plan for AI in the health sector, including the incorporation of privacy standards in the software-development lifecycle.

- To protect federal government information in AI use, agencies are encouraged to implement measures to ensure compliance with privacy and data protection requirements. Additionally, foreign resellers of US IaaS products are required to limit third-party access to model data.

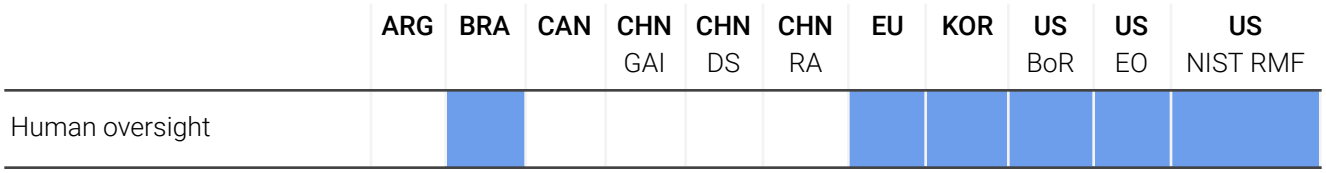
- To advance global technical standards for safe AI development and use, the Secretary of Commerce is to establish a plan for global engagement including best practices regarding data protection and privacy. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights dedicates a section to data privacy, with provisions on design choices, consent, and sensitive domains. AI systems' design should ensure privacy by default and only use data that is strictly necessary for the specific context. Designers, developers, and deployers should seek permission to collect, use, access, transfer, and delete data, via brief and understandable consent requests. In sensitive domains, including health, work, and education, data should only be used for necessary functions and continuous surveillance should not be used. In addition, the Bill of Rights calls for providers to design a user experience that does not obfuscate choice, provide reports to document data protection, and refrain from using inappropriate or irrelevant data in the development and deployment of AI systems. [Check the specific provisions on CLaiRK↗]

The US NIST Risk Management Framework calls for the examination and documentation of privacy risks in AI systems. [Check the specific provisions on CLaiRK↗]

Human oversight

The OECD AI Principle 1.2 (Respect for the rule of law, human rights and democratic values) demands the promotion of **human rights and human-centred values**. To this end, AI rulebooks establish human oversight mechanisms in the development and operation of AI systems. Human oversight can range from mere human supervision, to active human involvement ("in-the-loop"), to human fallback as an alternative to automated processes. This article systematically analyses human oversight requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises the jurisdictions that require human oversight.



Comparison

Human oversight requirements apply to **different AI systems**. The US Bill of Rights and the NIST Risk Management Framework mandate human oversight for all AI systems. The US Executive Order demands human oversight for the use of AI in the health sector and the federal government. The EU and South Korea limit human oversight to high risk AI systems, to which the EU adds specific oversight measures for biometric identification systems. Finally, Brazil only demands human oversight for AI systems with significant impact.

Human oversight requirements differ in the **qualification criteria** for responsible humans and the extent of human oversight. Brazil and the EU outline the necessary qualifications for responsible humans, while South Korea and the US do not provide such detail. Brazil, the EU, and the US include the option of halting the AI system and replacing it with human decision-making, while South Korea does not.

Country details

Brazil

Brazil establishes a detailed human oversight requirement for AI decisions, predictions, and recommendations that may pose risks to the physical integrity of individuals or have an impact that is difficult to reverse. Individuals responsible for human supervision must understand the capabilities and limitations of the AI system, properly control its technical operation and interpret results, and intervene or interrupt its operation when necessary. The aim of this human oversight is to prevent or minimise risks to human rights resulting from normal use or reasonably foreseeable conditions of misuse. [Check the specific provisions on CLaiRK↗]

European Union

The EU requires high risk AI systems to be designed and developed in a manner that enables effective human oversight. The goal of this human oversight is to prevent or minimise risks to health, safety, or fundamental rights resulting from the intended use or reasonably foreseeable misuse of the high risk AI

system. Oversight measures shall thus be commensurate with the risks, autonomy, and context of the AI use.

Specifically, the EU requires measures to be either built into the high risk AI system by the provider before market placement or identified by the provider before market placement and then implemented by the deployer. The responsible human must be competent and accordingly trained and supported. They should understand the AI system's capabilities and limitations, monitor its operation, interpret its output correctly, and, if necessary, stop its operation. In addition, for remote biometric identification systems, the EU prohibits deployers from taking action or decision on the basis of the AI's identification, unless that identification is separately verified and confirmed by at least two natural persons with the necessary competence, training and authority. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea mandates that high risk AI development business operators must ensure human management and supervision of high risk AI, without further specification. [Check the specific provisions on CLaiRK↗]

United States

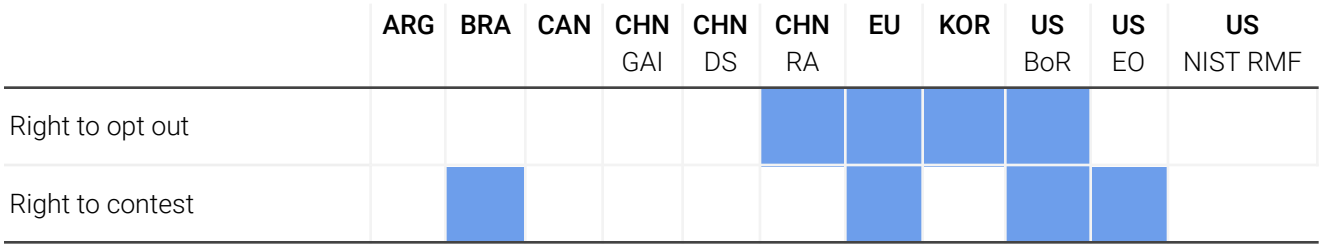
The Executive Order on AI mandates the Department of Department of Health and Human Services to develop a strategic plan on responsible AI deployment in the health and human services sector, considering human oversight of AI-generated output. Additionally, the Executive Order instructs the Office of Management and Budget to issue guidance regarding government use of AI, including minimum risk management practices such as human consideration. [Check the specific provisions on CLaiRK↗]

The NIST Risk Management Framework calls for the establishment of policies and procedures that define responsibilities for human-AI configurations and AI oversight. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights stipulates that users of automated systems should have access to a human who can quickly consider and remedy problems. This applies both when automated systems produce an error and when users contest the AI's decision. Human consideration should be timely, accessible, and effective. In addition, automated systems intended for use in sensitive domains, such as criminal justice, employment, education, and health, should incorporate human consideration for high risk decisions. [Check the specific provisions on CLaiRK↗]

Interaction rights

The OECD AI Principle 1.2 (Respect for the rule of law, human rights and democratic values) calls for the **promotion of human rights**. To this end, AI rulebooks provide users with rights throughout their interactions with AI systems. Interaction rights give users the option to interfere with the usage of an AI system, namely to opt out of AI use and to contest AI decisions. This article systematically analyses AI interaction rights across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises the jurisdictions that establish interaction rights.



Comparison

The **right to opt out** differs regarding the addressed AI systems and the granularity of options to opt out. Regarding addressed AI systems, the EU and South Korea extend this right to interactions with high risk AI systems. In China, this right applies to recommendation algorithms, regardless of risk. The US Bill of Rights extends this right to all AI systems. Regarding granularity, all four jurisdictions grant the right to opt out of AI use in its entirety. In addition, China enables users to opt out of specific components underlying the AI system's decision-making, namely user labels.

The **right to contest AI decisions** differs regarding the addressed AI systems and the subsequent level of human involvement. Brazil restricts this right to decisions with significant impacts. The EU restricts the right to the real-world testing of high risk AI. The US Bill of Rights grants this right regarding all AI systems, while the Executive Order focuses on AI use to administer public benefits and in the federal government. In terms of subsequent human involvement, Brazil establishes a right for human "participation," without further specification. The US Bill of Rights and Executive Order demand that a human considers and remedies problems upon contestation. The EU does not provide details on human involvement.

Country details

Brazil

Brazil grants the right to contest AI decisions that carry legal effects or significantly impact the interests of the affected person. This right extends to decisions based on discriminatory, unreasonable, or bad-faith inferences. Individuals are empowered to receive information to contest the decision and can request human "participation" in the decision-making. Notably, human participation is not required if the provider demonstrates that it is infeasible and instead provides alternative measures that ensure the reanalysis of the contested decision, taking into account the arguments raised by the affected person and repairing any damages. [Check the specific provisions on CLaiRK↗]

China

The Chinese regulation on recommendation algorithms demands that users receive convenient options to deactivate algorithmic recommendation. Additionally, users must have the option to select or delete specific labels used to tailor recommendations to their personal characteristics. [Check the specific provisions on CLaiRK↗]

European Union

The EU limits the right to opt out and the right to contest AI-assisted decisions to the specific context of high risk AI testing in real-world settings, outside designated AI regulatory sandboxes. For such testing to be permissible, individuals must be informed on the nature and objectives of the testing, provide previous informed consent, and retain the right to revoke consent at any point.

In addition, modalities must enable users to request the reversal or disregard of AI predictions, recommendations, or decisions. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea grants users the right to object to products or services using high risk AI. High risk AI business operators must inform users that a product or service is processed by high risk AI and about their right to object to the high risk AI use. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI contains provisions on the right to contest when AI is for the administration of public benefits and in the federal government.

- Regarding public benefits, the Executive Order instructs the Secretary of Health and Human Services to issue guidance regarding the use of automated or algorithmic systems to administer public benefits. The guidance is to identify instances when applicants can appeal benefit determinations and receive reconsideration by a human reviewer, as well as receive human customer support.
- Regarding federal government use of AI, the Executive Order instructs the Office of Management and Budget to issue guidance, including minimum risk management practices such as mechanisms for human consideration and remedies for adverse decisions made by AI systems. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights calls for the right to opt out and the right to contest AI-assisted decisions. Users should have the right to opt out from automated systems in favour of a human alternative, where appropriate based on reasonable expectations and the protection from harmful impacts. Users should further be able to appeal or contest the impacts of automated systems on them and have access to a person that quickly considers and remedies their problems. [Check the specific provisions on CLaiRK↗]

Principle 1.3

Transparency and explainability

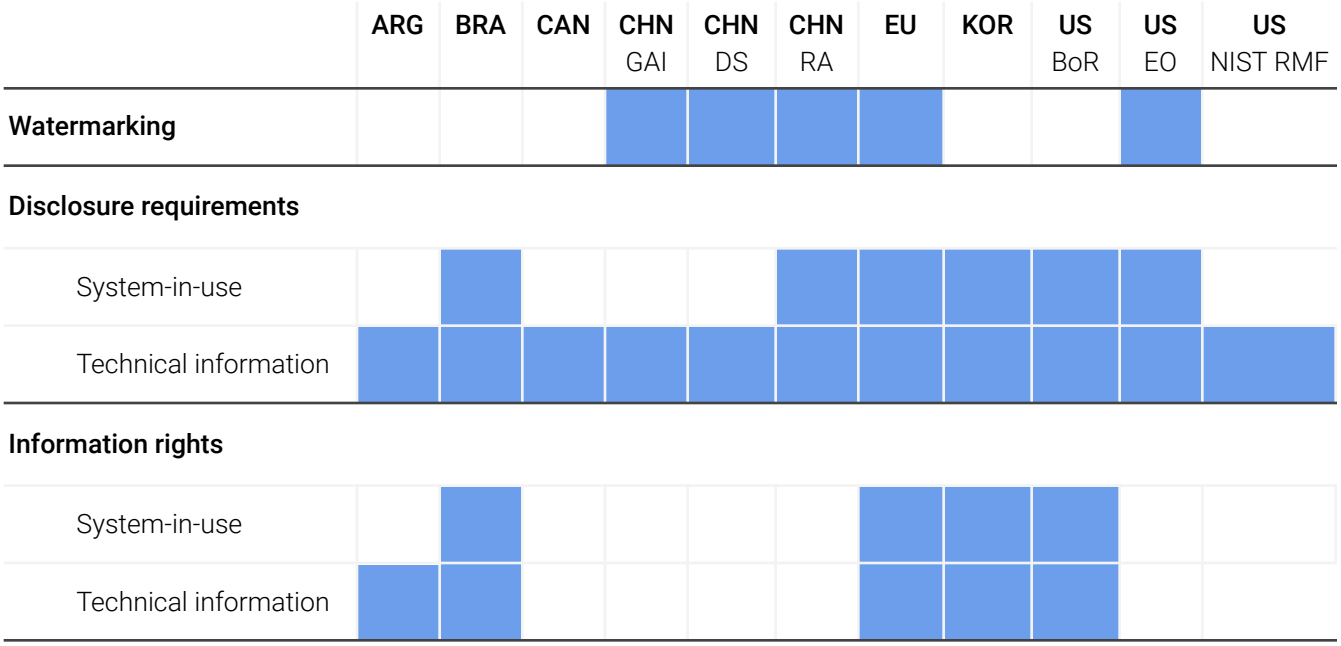
Principle 1.3: Transparency and explainability

AI raises several **transparency and explainability concerns**, two of which are top-of-mind for governments across the globe. First, the interaction with AI systems increasingly mimics human interaction. Second, AI systems are inherently opaque, leaving humans that interact with AI systems in the dark on the factors behind AI decisions. Despite sharing these regulatory concerns, governments choose different regulatory requirements to counter them.

A patchwork of regulatory requirements implements OECD AI Principle 1.3

The **OECD AI Principle 1.3** demands that AI actors commit to transparency and responsible disclosure regarding their AI systems. They should provide meaningful information to foster a general understanding of AI, make stakeholders aware of their interactions with AI, and provide information on the factors behind AI output.

In national AI rules, a **patchwork of regulatory requirements** implements the OECD AI Principle 1.3. The heatmap below visualises divergence within a selection of these requirements, grouped in three categories. Watermarking requirements directly attach to AI systems' output. Disclosure requirements demand that AI actors actively provide information. Information rights empower users to reactively request information. Below, we explain each category in detail.



Content watermarking is rarely required

Watermarking requirements require AI providers to add a **visible label or disclaimer** on AI-generated output. This technical approach aims to enhance transparency by default, enabling humans to see that content is created using AI, not solely by human effort. Watermarking is required rarely, namely in China (regulations on generative AI, deep synthesis services, and recommendation algorithms), the EU, and

the US (Executive Order). Adding to the patchwork, the three countries differ in the types of content that must be watermarked and the specific characteristics of watermarks.

Disclosure requirements are widespread

Public disclosure requirements oblige AI actors to **actively provide information**, either on their use of AI systems or on the AI systems' functioning.

- System-in-use disclosure requirements demand that the use of an AI system is disclosed. This addresses the concern that AI systems increasingly mimic human interaction.
- Technical disclosure requirements demand that information on the technical functioning of the AI system is disclosed, for example underlying datasets and heuristics. This addresses the concern that AI systems are inherently opaque.

Disclosure requirements are **widespread across borders**. System-in-use disclosure is required in seven rulebooks. Going further, each of the 11 analysed rulebooks demands some form of technical disclosure. For technical disclosure, however, granular differences persist regarding the information that must be disclosed and the format and timing of disclosure.

Information rights are scarcely used

Information rights empower users to **request information** on AI systems, which AI providers must reactively deliver.

- The basic right to be informed that an AI system is in use addresses the concern that AI systems increasingly mimic human interaction.
- The right to specific information about the AI systems' functioning can cover both the general functioning of an AI system and the processes behind a specific decision or output. It thus addresses the concern that AI systems are inherently opaque.

Information rights are **rarely established**. Four rulebooks establish the basic right to be informed that an AI system is in use, while five rulebooks establish the right to specific information about the AI systems' functioning. Adding to the patchwork, information rights differ regarding who is empowered (only users or anyone who is affected) and how the information is to be conveyed.

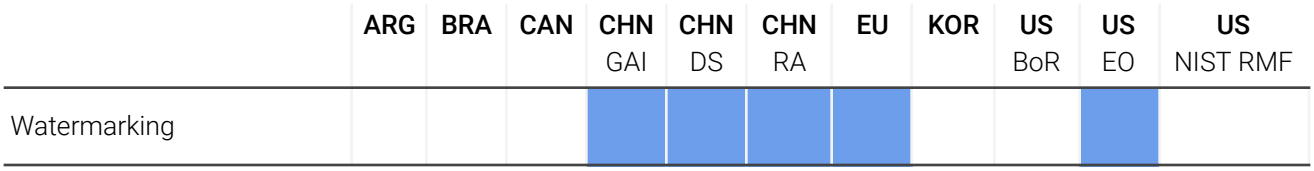
Dive deeper into each regulatory requirement

The patchwork of regulatory requirements that implement OECD AI Principle 1.4 is only the **tip of the iceberg**. Granular differences emerge even within the jurisdictions that impose the same regulatory requirements. To showcase granular divergence, we now proceed with a detailed comparative analysis of the following requirements. Jump directly to the section that interests you:

- Content watermarking
- System-in-use disclosure
- Technical disclosure
- Information rights

Content watermarking

The OECD AI Principle 1.3 (Transparency and explainability) demands that stakeholders are made aware of their **interactions with AI systems**. The spread of generative AI has raised the demand for transparency, since AI-generated content increasingly resembles human content. Governments have thus started to demand that artificially generated content is watermarked. This article systematically analyses watermarking requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish watermarking requirements.



Comparison

Watermarking requirements generally apply to AI systems that generate synthetic content, but **minor differences in scope** arise nevertheless. Only the EU outlines exceptions, including AI systems used in criminal investigations and human review or editorial control over content. China's deep synthesis regulation and the EU AI Act establish special rules for "deep fakes," while the EU establishes rules for text published to inform the public on matters of public interest. The US does not establish exceptions or specific obligations, although these may emerge upon the mandated investigation of watermarking.

AI rulebooks impose different **watermarking procedures**. China stipulates that watermarking should be implemented through technical measures that do not hinder users' ability to use the content. For deep synthesis services that may lead to confusion or misrecognition, significant watermarking must be applied in a reasonable area of the content. The EU mandates technical solutions for watermarking that are effective, interoperable, robust, and reliable. For content that is evidently artistic, creative, satirical, or fictional, the watermarking requirement should not hinder the display or enjoyment of the work. The US does not provide procedural details.

Country details

China

China's has issued a comprehensive regulation specifically dedicated to "deep synthesis" services. Deep synthesis is a technology that utilises generative synthesis algorithms, such as deep learning and virtual reality, to create various forms of content including text, images, audio, video, virtual scenes, and other network-based information. All providers of deep synthesis services are required to implement technical measures to incorporate watermarking without impeding users' ability to utilise the generated or edited content. Additionally, they are obligated to store log information in compliance with legal requirements. For deep synthesis service providers offering services that may lead to confusion or misrecognition by the public, significant watermarking must be applied in a reasonable location or area of the information content generated or edited. These services include simulations of natural persons for text generation

or editing, voice imitation, face generation, and immersive anthropomorphic scenes. [Check the specific provisions on CLaiRK↗]

China's regulations on recommendation algorithms and generative AI also contain provisions on content watermarking. When recommendation algorithm providers notice that synthetic content is not labelled as such, they must halt its dissemination until correctly labelling. Generative AI providers must watermark content in accordance with the deep synthesis regulation. [Check the specific provisions on CLaiRK: recommendation algorithms↗ I generative AI↗]

European Union

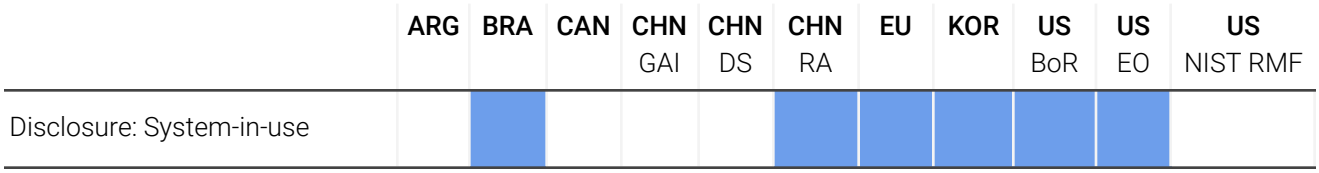
The EU AI Act requires providers of AI systems that generate synthetic audio, image, video or text content, to mark the output in a machine-readable format and make it detectable as artificially generated or manipulated. The technical solution must be effective, interoperable, robust and reliable. In addition, deployers of AI systems that generate either "deep fakes" or text published to inform the public on matters of public interest must disclose that the content was artificially generated or manipulated. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI mandates several government agencies to address watermarking. The Secretary of Commerce must submit a report that identifies standards, tools, methods and practices to authenticate content and detect synthetic content. The Office of Management and Budget must develop guidance regarding digital content authentication and synthetic content detection. Specifically, the guidance should provide recommendations to agencies regarding reasonable steps to watermark or otherwise label generative AI output. [Check the specific provisions on CLaiRK↗]

System-in-use disclosure

The OECD AI Principle 1.3 (Transparency and explainability) demands that stakeholders are made aware of their **interactions with AI systems**. AI rulebooks often require AI providers to publicly disclose whether an AI system is in use. This general disclosure requirement is complemented by technical disclosure obligations, requiring information on specific elements of an AI system such as training datasets. This article systematically analyses system-in-use disclosure requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish system-in-use disclosure requirements.



Comparison

System-in-use disclosure requirements are different in **scope**. South Korea applies the requirement only to high risk AI systems. In Brazil, providers must implement transparency measures for AI systems that interact with natural persons. The EU foresees requirements for both high risk AI systems and all AI systems that interact with natural persons. In China, the requirement is technology-specific, applying only to recommendation algorithms. The US Bill of Rights and the US Executive Order do not further specify the scope of their provisions.

Another factor of divergence is the specific **format of disclosure**. The Chinese regulation on recommendation algorithms simply stipulates “conspicuous” notification by providers. Similarly, a basic notice suffices to uphold system-in-use disclosure requirements in South Korea and the US. Brazil specifically regulates the design of human-machine interfaces, including requirements for accessibility. The EU enables various formats for system-in-use disclosure, ranging from simple notifications, to disclosure through design and registration in a public registry.

Country details

Brazil

Brazil requires transparency measures concerning the use of AI systems that interact with individuals, including human-machine interfaces that provide ”adequate” clarity and information. In addition, people exposed to emotion recognition or biometric categorisation systems must be informed regarding the environment in which the exposure occurs. [Check the specific provisions on CLaiRK↗]

China

The Chinese regulation on recommendation algorithms requires providers to notify users in a conspicuous manner about the use of algorithmic recommendation. Providers must appropriately

publicise the “basic principles” of their systems, their purpose, and their main operating mechanisms. [Check the specific provisions on CLaiRK↗]

European Union

The EU AI Act requires any AI system that directly interacts with natural persons to be designed and developed in a way that reveals the interaction with the AI system. Exceptions are foreseen if this interaction is obvious or the AI system is authorised by law to detect, prevent, investigate or prosecute crimes. In addition, deployers of high risk AI systems that make decisions related to natural persons and deployers of emotion recognition or biometric categorisation systems must inform natural persons of their exposure to said systems. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea requires business operators of high risk AI to inform users in advance that services using high risk AI are being provided and inform them about their right to request information. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI mandates the Department of Health and Human Services to publish a plan regarding the use of automated or algorithmic systems in the implementation of public benefits and services by states and localities, including to ensure that recipients are informed of the use of such systems. In addition, to improve transparency for government agencies’ use of AI, the Office of Management and Budget is to issue yearly instructions on the collection, reporting, and publication of agency AI use cases. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights calls for designers, developers, and deployers to provide a notice when automated systems are in use, along with a clear description of the overall system functioning, the role of automation, the responsible individual or organisation, and outcome explanations. [Check the specific provisions on CLaiRK↗]

Technical disclosure

The OECD AI Principle 1.3 (Transparency and explainability) demands **information on the functioning of AI systems**, including factors and decision processes. AI rulebooks often require AI providers to publicly disclose how an AI system functions. Such "technical disclosure requirements" can be general or relate to specific elements of the AI system, such as the training data or algorithm. This article systematically analyses obligations to disclose the technical elements of an AI system across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish public disclosure requirements.

	ARG	**BRA**	**CAN**	**CHN** GAI	**CHN** DS	**CHN** RA	**EU**	**KOR**	**US** BoR	**US** EO	**US** NIST RMF
Disclosure: Technical											

Comparison

Technical disclosure requirements differ in **scope**, ranging from general requirements to technology- and risk-specific requirements. Argentina, Brazil, and the US demand technical disclosure from all providers. In China, these requirements apply to providers of specific AI technologies. The EU and South Korea require providers of high risk AI systems to disclose technical information. Canada combines a general requirement with a specific requirement for providers of high impact AI systems.

Technical disclosure requirements diverge significantly in their **level of detail**. Brazil, China, South Korea, and the US Bill of Rights provide relatively little detail, mandating the disclosure of principles, purposes and governance measures. Conversely, Argentina, Canada and the EU impose precise disclosure obligations, listing the various types of technical information to be disclosed. This information spans from descriptions of data collection processes and methods, user instructions, technical characteristics and capabilities, limitations, computational and hardware resources, to risks and necessary precautions.

Finally, technical disclosure is to be conveyed in **different formats**. Argentina, China, and the EU, require registration in a public registry. The other jurisdictions don't specify the format, requiring AI providers to provide a suitable interface for the information to reach users.

Country details

Argentina

Argentina requires detailed documentation and disclosure of operations and algorithms used in AI systems to enable auditing and evaluation of their impact. Specifically, those responsible for AI systems must disclose information when publicly registering their systems, including technical characteristics, purposes and objectives, design and operation, as well as measures regarding transparency, accountability, and security. In addition, the personal data processed, along with its origin, nature, sources, and recipients, must be disclosed in the register. [Check the specific provisions on CLaiRK↗]

Brazil

In Brazil, AI providers must uphold transparency measures, including regarding the governance measures in the development and use of the AI system. In addition, providers of high risk AI systems must, in the context of impact assessment, provide a publicly available description of the intended purpose, context of use, and territorial and temporal scope of the AI system. Finally, Brazil requires the disclosure of results from testing and impact assessments. [Check the specific provisions on CLaiRK↗]

Canada

In Canada, those responsible for AI systems must provide clear and understandable information regarding the responsible usage of these systems. This information covers intended uses, limitations, risks, and necessary precautions, as well as descriptions of the content, decisions, recommendations, or predictions the AI system makes. Moreover, those responsible for high impact AI systems must publish a plain-language description of the system, including an explanation of mitigation measures, on a publicly accessible website. [Check the specific provisions on CLaiRK↗]

China

China's three AI regulations all demand technical disclosure, including to publicly display information submitted to the "algorithm filing" registration regime. This information includes identification, field of application, and algorithm type, among others.

The regulation on generative AI requires providers to clearly specify and disclose the intended group of users, circumstances, and purposes of their services, to guide users towards a rational understanding and lawful use of the technology. [Check the specific provisions on CLaiRK↗]

The regulation on deep synthesis services requires providers to formulate and publish management rules, platform conventions, and service agreements. In addition, China's regulation of deep synthesis services emphasises the prevention of false information and only allows for news information released by internet news information source units to be reproduced. [Check the specific provisions on CLaiRK↗]

The regulation on recommendation algorithms requires providers to publish service rules and operating mechanisms. [Check the specific provisions on CLaiRK↗]

European Union

The EU AI Act contains extensive rules on "technical documentation," which is to be publicly disclosed. High risk AI systems must be accompanied by instructions for use, with information on the characteristics, capabilities, limitations, underlying datasets, accuracy, and purpose of the AI system. Further technical disclosure requirements cover the AI system's performance, the needed computational and hardware resources, and the maintenance necessary for proper functioning.

Moreover, providers of general-purpose AI systems must publish detailed summaries of the content used for training. Further disclosure providers cover information on human oversight, prevalent risks, testing results, as well as the "CE" marking for high risk AI systems, affirming conformity with European health, safety, and environmental protection standards. The EU will establish a public database for high risk AI systems, on which providers must disclose contact details, the purpose and function of the

system, the data and inputs used by the system and the operation logic. [Check the specific provisions on CLaiRK↗]

South Korea

In South Korea, high risk AI developers (business operators) must notify users and stakeholders of the "operating principles," without disclosing trade secrets. Furthermore, South Korea requires the disclosure of prevalent risks when using AI systems. [Check the specific provisions on CLaiRK↗]

United States

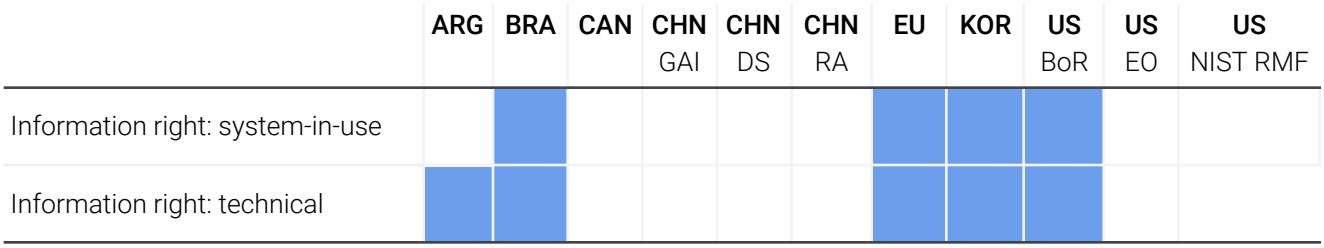
The Executive Order on AI underscores the importance of AI transparency and encourages independent regulatory agencies to consider rulemaking. In addition, the Executive Order mandates the Secretary of Commerce to submit a report which identifies the existing standards, tools, methods and practices as well as potential further standards and techniques to track content provenance. [Check the specific provisions on CLaiRK↗]

The NIST Risk Management Framework also calls for the elucidation and documentation of AI systems, as well as the interpretation of AI output within its context to inform responsible use and governance. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights calls for designers, developers, and deployers of automated systems to provide generally accessible, plain language documentation. The documentation encompasses clear descriptions of the system's functioning and the role of automation, as well as explanations of outcomes. The information must be regularly updated and individuals affected by the system must be informed of significant changes. In addition, the Bill of Rights calls for the disclosure of information on human oversight. [Check the specific provisions on CLaiRK↗]

Information rights

The OECD AI Principle 1.3 (Transparency and explainability) demands **information on the use and functioning of AI systems**. Several governments grant users of AI systems information rights. Information rights can cover basic information that an AI system is in use or specific information on how an AI system functions, for instance processes behind a specific decision. This article systematically analyses AI information rights across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish information rights.



Comparison

Although the **basic right to be informed** about the use of an AI system is similar across jurisdictions, differences persist regarding the addressed AI systems. Brazil and the US Bill of Rights grant all users the right to be informed, irrespective of its risk or application area. The EU grants this right only in specific application areas and only when users are unlikely to be aware of interacting with an AI system. South Korea grants this right only to users of high risk AI.

Similarly, the **right to specific information** about the AI system differs across jurisdictions regarding addressed AI systems and level of detail. Argentina, Brazil and the US Bill of Rights grant this right for all AI systems. South Korea and the EU only afford this right to users of high risk AI systems. Regarding the level of detail, only Brazil provides a detailed elaboration on the specific elements that must be disclosed to the user. South Korea, on the other hand, provides the most detail regarding the information request procedure. Notably, Argentina, Brazil, the EU, and South Korea include an explicit right to request an explanation of AI decisions, which in turn provides information about the AI system.

Country details

Argentina

Argentina demands a level of transparency that allows users of AI systems to understand the decision-making process and output of AI systems. In addition, the rulebook explicitly establishes a right for affected persons to request explanations of AI decisions. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil establishes a basic right for people affected by AI systems to be informed on the automated character of the interaction before use.

In addition, Brazil grants people affected by AI systems the right to receive clear and adequate information before use. This includes a comprehensive description of the AI system, including the role of AI and humans in decision-making, the underlying data, the output, as well as measures to ensure non-discrimination and reliability. In addition, affected persons can request an explanation of the decision, recommendation or prediction made by an AI system, including information about the criteria, procedures, and factors that underlie the decision. This includes the rationality and logic of the system, the meaning and predicted consequences of decisions, the processed data and its source, and the criteria for decision-making and their weighting. This information must be provided for free, in understandable language, within fifteen days of the request. [Check the specific provisions on CLaiRK↗]

European Union

The EU establishes a basic right to be informed about the use of an AI system, providing details mainly regarding several exception mechanisms. Notably, this right does not apply to AI systems authorised by law to detect, prevent, investigate, and prosecute criminal offences. In addition, this right does not apply in circumstances where it is evident that the user is interacting with AI. This exception, in turn, does not apply to emotion recognition and biometric categorisation systems. In addition, regarding the use of AI in the workplace, both affected workers and representatives must be informed.

In addition, the EU grants individuals affected by high risk AI systems to request clear and meaningful explanations regarding the AI system's role in decision-making processes and the key elements of the decisions made. Furthermore, high risk AI systems must be accompanied by instructions for use, with information on the technical capabilities and characteristics of the AI system that are relevant to explain its output. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea establishes the basic right for users to be informed when the service or product they use involves high risk AI processing.

In addition, users have the right to request relevant materials from high risk AI business operators to verify whether they have been adversely affected by such systems. Specifically, users can request the Information and Communication Strategy Committee to compel high risk AI business operators to furnish this information if they refuse upon direct user request. In addition, users of high risk AI must be informed on the AI system's "operating principles" and the possibility of serious risk to life or physical safety through its use. [Check the specific provisions on CLaiRK↗]

United States

The Blueprint for an AI Bill of Rights declares that users should have the right to know that an automated system is being used. In addition, the Bill of Rights calls for a right to be informed about how

and why an AI system influences outcomes that affect the user. In addition, the Bill calls for users to have the right to understand the AI decision-making process, stipulating that explanations should be technically valid, meaningful, and useful. This information should be calibrated based on the level of risk and context. [Check the specific provisions on CLaiRK↗]

Principle 1.4

Robustness, security, and safety

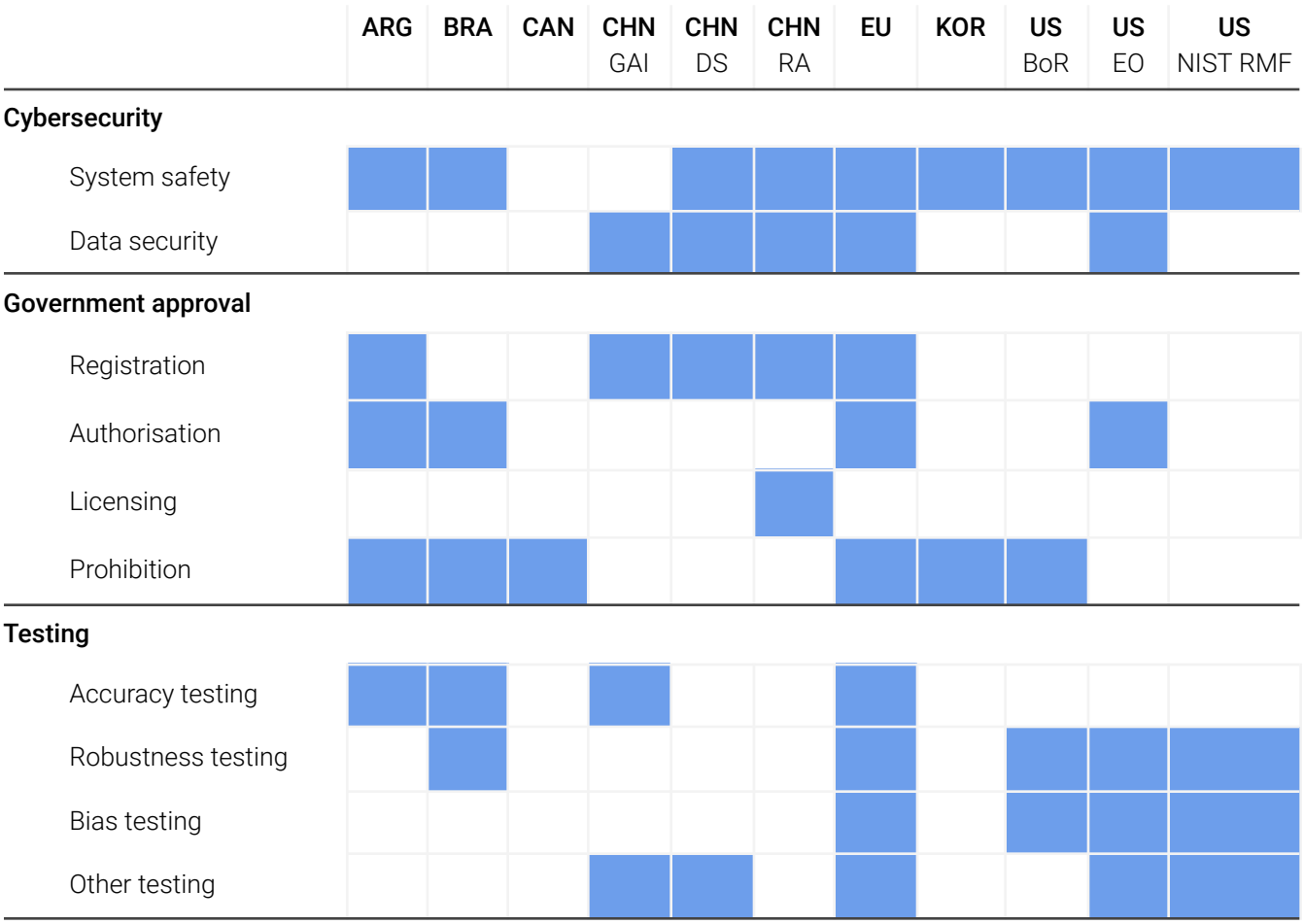
Principle 1.4: Robustness, security, and safety

The **safety risks** brought by AI systems are a salient and shared regulatory concern. In a rare display of international alignment, governments including China, the EU, and the US, jointly discussed AI safety in November 2023 and issued the Bletchley Declaration. At the following AI Seoul Summit, certain countries signed a declaration to address severe AI risks, a declaration for safe, innovative, and inclusive AI, and a statement of intent toward international cooperation on AI safety science. On the national level, however, regulatory approaches to address AI safety diverge.

A patchwork of regulatory requirements implements OECD AI Principle 1.4

The **OECD AI Principle 1.4** demands that AI systems should be robust, secure and safe throughout their entire lifecycle, in order to function appropriately and not pose unreasonable safety risks. AI actors should establish mechanisms to ensure that AI systems that risk causing undue harm or exhibit undesired behaviour can be overridden, repaired, or decommissioned. This extends to conditions of normal use, foreseeable misuse, and other adverse conditions.

In national AI rules, a **patchwork of regulatory requirements** implements the OECD AI Principle 1.4. The heatmap visualises divergence within a selection of these requirements, grouped in three categories. Below, we explain each category in detail.



Cybersecurity requirements often address systems and rarely data

Cybersecurity requirements demand that AI providers **address risks through technical measures**. Governments require two types of cybersecurity measures:

- System safety requirements demand safeguards for the AI system itself, including against unintended use or external manipulation.
- Data security requirements demand safeguards for the data in AI systems, including against data leaks.

Cybersecurity requirements showcase that jurisdictions can **converge and diverge on similar issues**. System safety requirements are established by all rulebooks, except Canada. Data security requirements are only established in China, the EU, and the US. Adding to the patchwork, governments' elaborations of these same requirements differ, for instance regarding the specific measures to be taken to safeguard AI systems or data.

Government approval requirements are scattered

Government approval requirements aim to grant the government **oversight over AI systems**, to prevent risks. Governments require four types of government approval:

- Registration requirements oblige providers to enter a register to launch AI systems.
- Authorisation requirements oblige providers to obtain government pre-approval to launch AI systems.
- Licensing requirements oblige providers to obtain an operating licence before launching AI systems.
- Prohibitions forbid certain AI systems, through heuristics or specific prohibitions.[2]

Every jurisdiction requires government approval in a certain form. All jurisdictions except China use prohibitions. Authorisation and registration requirements are more rare, featuring in four and three jurisdictions, respectively. Notably, China establishes a licensing regime, hinting at the novelty of AI as a regulatory object, for which licensing regimes are yet to be established. Adding to the patchwork, governments' elaborations of these same requirements differ, for instance regarding the specific AI systems that are prohibited, as well as authorisation and registration procedures.

Testing requirements are neither frequent nor rare

Testing requirements demand that AI providers **test and evaluate their AI systems.** Governments impose four types of testing requirements:

- Accuracy testing requirements cover the accuracy, reliability and effectiveness of AI systems.

[2] Heuristics are abstract rules that determine which AI systems are prohibited. Specific prohibitions enumerate forbidden AI systems based on their technical capabilities or their use context.

- Robustness testing requirements cover the robustness and security of AI systems.
- Bias testing requirements involve detecting, preventing, and mitigating potential biases and discrimination in AI systems.
- Other testing requirements encompass all AI testing measures that do not fall under the other three categories.

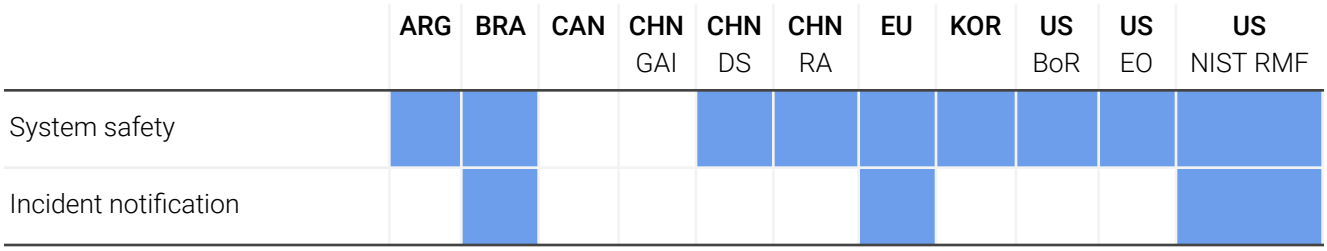
Testing requirements are all used in **similar frequency**, appearing in four or five of the 11 analysed AI rulebooks. The EU is the only rulebook that employs all testing requirements. Robustness and other testing requirements feature in four different rulebooks. Accuracy and bias testing are included in three other rulebooks. Adding to the patchwork, governments' elaborations of these testing requirements differ, for instance regarding the timing and scope of tests.

Dive deeper into each requirement

The patchwork of regulatory requirements that implement OECD AI Principle 1.4 is only the **tip of the iceberg**. Granular differences emerge even within the jurisdictions that impose the same regulatory requirements. To showcase granular divergence, we now proceed with a detailed comparative analysis of the following requirements. Jump directly to the section that interests you:

System safety

The OECD AI Principle 1.4 (Robustness, security, and safety) calls for the **prevention of unreasonable AI safety risks**. AI risks can be addressed through technical system safety requirements, which demand safeguards for the AI system itself, for example against unintended use or manipulation through intruders. Moreover, incident notification requirements. require the communication of AI safety incidents. This article systematically analyses system safety requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish system safety and incident notification requirements.



Comparison

System safety requirements differ regarding the **AI systems they address** and their imposition of technical, ethical, or organisational measures. The EU and South Korea establish risk-based system safety requirements. China and the EU impose technology-specific system safety rules. Argentina, Brazil, and the US apply system safety requirements to all types of AI systems. Regarding the types of safety measures, all AI rulebooks establish technical safety measures. Argentina and China also highlight the need for ethical measures. Only Argentina emphasises organisational measures for system safety.

Incident notification requirements differ in terms of the **types of incidents** that trigger notification, the notification timeline, and the addressee. Regarding the type of incidents, the EU provides the most detail, differentiating between serious incidents and widespread infringements. Brazil mandates notification for serious incidents and provides criteria, for example human rights infringements. The US NIST Risk Management Framework calls for the notification of incidents, regardless of their seriosity. Regarding the notification timeline, the EU establishes clear timelines that are calibrated to the incident type. Brazil mentions that the competent authority will determine timelines, while the NIST Risk Management Framework does not specify any timeline. Regarding the addressee, Brazil and the EU only demand notification to government authorities, while the US NIST Risk Management Framework extends notification to affected communities.

Country details

Argentina

Argentina states that developers, providers, and users are responsible for their AI systems and must ensure that they are safe, reliable, and in compliance with quality standards (Art. 5.1). AI systems must be designed and developed with adequate security measures, both technical and organisational, to

guarantee robustness and prevent unauthorised access, manipulation, or malicious interference. Argentina's verification and certification process for AI systems assesses safety based on technical, legal, and ethical criteria, including equity, transparency, privacy, and security. Finally, risks identified through risk and impact assessments must be mitigated through adequate security measures to minimise their impact on fundamental rights and individual security. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil mandates all AI agents to establish governance structures and internal processes to ensure the security of systems and uphold the rights of affected people. Specifically, security measures must be adopted throughout the AI system's lifecycle, from the design to the operation.

Regarding incident notification, Brazil mandates all AI agents to communicate serious security incidents to the competent authority in a "reasonable time" to be defined by the competent authority. Serious incidents include risks to the life and physical integrity of people, violations of fundamental rights, and the interruption of critical infrastructure operations. The competent authority then verifies the seriosity of the incident and can order measures to reverse or mitigate the effects of the incident. [Check the specific provisions on CLaiRK↗]

China

China's regulations on recommendation algorithms and deep synthesis services both contain system safety requirements. Providers must establish robust management systems and technical safeguards to ensure information security, data protection, and compliance with laws and regulations. In addition, providers must implement measures regarding user registration, algorithm audit, ethics assessment, data security, personal information protection, and anti-fraud measures. [Check the specific provisions on CLaiRK↗]

The deep synthesis regulations further note that application stores must review the security assessment and, in case of deficiencies, issue warnings, suspensions, or removals. [Check the specific provisions on CLaiRK↗]

European Union

The EU mandates high risk AI systems to be designed and developed with the aim of consistent accuracy, robustness, and security. Specifically, high risk AI systems must be resilient regarding attempts by unauthorised third parties to exploit system vulnerabilities to alter the use, output, or performance of AI. Appropriate technical solutions must prevent, detect, respond to, resolve, and control for attacks trying to manipulate the AI training dataset (data poisoning), components of the training model (model poisoning), inputs designed to cause the model to make a mistake (adversarial examples/model evasion), confidentiality attacks, and model flaws. In addition, users must be enabled to safely interrupt the functioning of high risk AI systems with a "stop" button or similar procedure. To verify system safety, the EU establishes a presumption of compliance for systems certified under the EU Cybersecurity Act.

Regarding incident notification, the EU mandates providers of high risk AI systems to report serious incidents to authorities. The report is due immediately once the provider establishes a causal link

between the AI system and the serious incident, or at the latest 15 days after becoming aware of the incident. For "widespread infringements," such as disruption of critical infrastructure, the timeline is reduced to two days. In addition, providers of general-purpose AI models with systemic risk must document and report relevant information about serious incidents and possible measures to address them to the AI Office and national authorities. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea mandates the strengthening of cybersecurity in the development of high risk AI systems, without further specification. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI, NIST Risk Management Framework, and Blueprint for an AI Bill of Rights all establish system safety requirements.

The Executive Order mandates agencies to draft guidelines for safe, secure, and trustworthy AI systems.

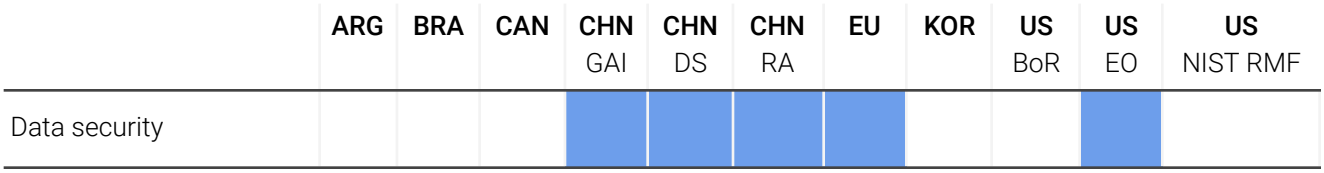
- To address malicious cyber activities, the Executive Order instructs the Secretary of Commerce to require companies developing dual-use AI foundation models to provide the federal government with ongoing information and records. This extends to the cybersecurity measures taken to protect dual-use foundation models, their model weights, and the results of red-team testing.
- In addition, the Executive Order mandates agencies to assess and mitigate AI risks in critical infrastructure sectors, establish cybersecurity best practices for financial institutions, and incorporate AI risk management in infrastructure security guidelines.
- Finally, the Executive Order mandates agencies to pilot AI capabilities for identifying and remediating vulnerabilities in government systems, and calls for measures for safe AI use across sectors including healthcare and education. [Check the specific provisions on CLaiRK↗]

The NIST Risk Management Framework (RMF) establishes system safety requirements for both normal and extraordinary circumstances. During normal operations, the NIST RMF calls for regular evaluation of safety risk, monitoring of reliability and robustness, and demonstration of failsafe mechanisms. To address extraordinary circumstances, the NIST RMF demands mechanisms to override or deactivate AI systems exhibiting unintended performance or outcomes, to safely decommission AI systems without increasing risks or undermining trustworthiness, and to handle failures or incidents involving high risk third-party AI systems or data. Regarding incident notification, the NIST RMF states that incidents and errors are to be communicated to relevant actors, including affected communities, and demands the tracking, response, and recovery from incidents. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights mandates AI systems to be designed prioritising user safety, avoiding foreseeable risks, and proactively protecting users from unintended harmful impacts. If an AI system fails or produces errors, users should have access to timely human intervention and redress mechanisms. [Check the specific provisions on CLaiRK↗]

Data security

The OECD AI Principle 1.4 (Robustness, security, and safety) calls for the prevention of **unreasonable safety risks by AI systems**. AI risks can be addressed through data security requirements, meaning safeguards for the data underlying an AI system, including the prevention of data leaks and manipulation. This article systematically analyses data security requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish data security requirements.



Comparison

Data security requirements have a **different scope** across jurisdictions. The EU restricts data security requirements to high risk AI. China imposes requirements only regarding specific technologies, namely recommendation algorithms, deep synthesis services, and generative AI. The US Executive Order covers all types of AI systems, but only in specific contexts, such as health and critical infrastructure.

The **level of detail** with which jurisdictions outline data security requirements differs greatly. The EU's single provision on data security specifically lists the data security risks that must be addressed by technical solutions, including data and model poisoning. The Chinese regulations assign clear responsibilities to AI providers but do not mention specific measures – instead, they reference the existing (and evolving) data security regime in China. Finally, the US Executive Order does not provide detail since it instructs government agencies to scrutinise data security in their respective sectors.

Country details

China

The Chinese regulations on recommendation algorithms, deep synthesis services, and generative AI all mandate providers to take responsibility for algorithmic security and ensure compliance with data security requirements enshrined in law. [Check the specific provisions on CLaiRK: recommendation algorithms↗ I generative AI↗]

The regulation on deep synthesis specifically stipulates that providers must strengthen the management of training data, implement necessary measures to guarantee its safety, and uphold data security and protection. [Check the specific provisions on CLaiRK↗]

European Union

The EU mandates high risk AI systems to be resilient regarding attempts by unauthorised third parties to exploit system vulnerabilities to alter the use, output, or performance of AI. The EU thus mandates

appropriate technical solutions to ensure the cybersecurity of high risk AI systems. Specifically, the EU lists measures to prevent, detect, respond to, resolve, and control for attacks trying to manipulate an AI training dataset (data poisoning) or components of the training model (model poisoning). Furthermore, such protections must be raised against inputs designed to cause the model to make a mistake (adversarial examples / model evasion), as well as "confidentiality attacks" and "model flaws." [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI does not directly impose data security requirements, but instructs several agencies to address the issue. Relevant regulatory agencies must thus:

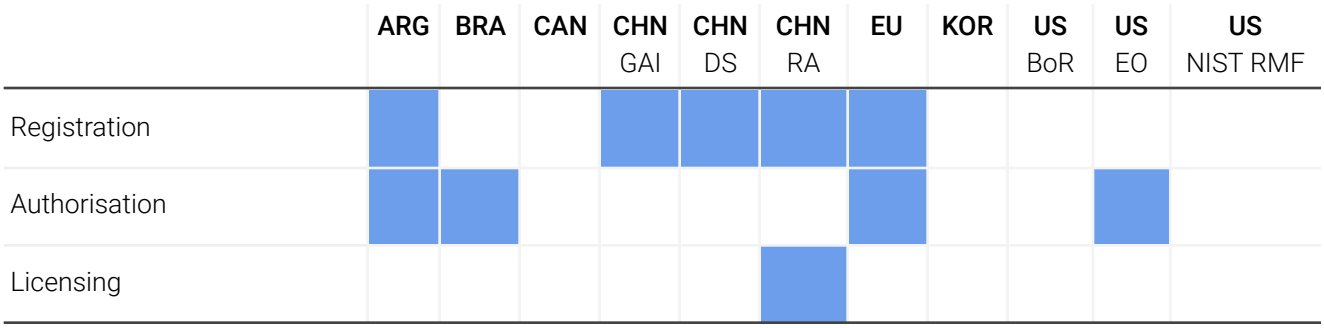
- develop initial guidelines for security reviews, with focus on risks of releasing federal data connected to chemical, biological, radiological, and nuclear weapons, as well as autonomous offensive cyber capabilities
- evaluate potential risks of AI use in critical infrastructure, and
- develop a strategic plan on responsible AI deployment in the health and human services sector, including the incorporation of safety, privacy, and security standards and measures to address AI-enhanced cybersecurity threats. [Check the specific provisions on CLaiRK↗]

Registration, authorisation, and licensing

The OECD AI Principle 1.4 (Robustness, security, and safety) demands the prevention of **unreasonable safety risks by AI systems**. Governments establish registration, authorisation, and licensing regimes to establish and maintain oversight of AI systems.

- Registration requirements demand that providers register with the government to launch certain AI systems.
- Authorisation refers to the obligation to obtain government pre-approval to launch certain AI systems.
- Licensing requirements formalise authorisation and oblige providers to obtain an operating licence before launching certain AI systems.

This article systematically analyses registration, authorisation, and licensing requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish registration, authorisation, and licensing requirements.



Comparison

Registration requirements converge in their demand for the submission of information on a public registry, but address different types of AI. Argentina demands those responsible for any AI system to register. The EU demands registration only for high risk AI systems. China demands a unique form of registration from providers of recommendation algorithms, deep synthesis services, and generative AI with "public opinion attributes" or "social mobilisation potential."

Authorisation requirements also address different types of AI. In Argentina, authorisation applies to all AI systems. Brazil and the EU only demand authorisation for certain biometric systems in criminal investigations. The US Executive Order envisages an authorisation regime for generative AI by the federal government.

Licensing requirements are only established in China, specifically for recommendation algorithms in a news context.

Country details

Argentina

Argentina requires those responsible for AI systems to register their systems in a public registry managed by the AI supervisory authority. The registry includes technical details such as purpose, design, algorithms, and processed personal data and its sources. The supervisory authority will define procedures and criteria for registering AI systems, including updates upon significant changes to the systems' technical characteristics or data processing. In addition, developers of AI systems must register their AI system according to the procedures of the Scientific and Technological Cabinet, for verification and certification.

In addition, those responsible for AI systems are subject to a verification and certification process to ensure the AI's quality and safety, as well as compliance with technical, legal, and ethical requirements. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil does not impose a general authorisation requirement but demands federal legislation and judicial authorisation for the use of remote biometric identification systems in public spaces for targeted criminal investigations. [Check the specific provisions on CLaiRK↗]

China

China's regulations on recommendation algorithms, deep synthesis services, and generative AI demand "algorithm filing," a unique form of registration. The regulation on recommendation algorithms originally established this requirement for providers of services with public opinion attributes or social mobilisation potential. The filing must be completed within ten days of market deployment, be updated within ten working days of significant changes, and be cancelled within twenty days of service termination. In addition, the regulation on recommendation algorithms requires providers who deliver Internet news and information services to obtain an "Internet News Information Services License." [Check the specific provisions on CLaiRK↗]

The regulations on deep synthesis services mention that such providers must follow the algorithm filing procedure. Specifically, providers must publicly submit information, including identification, field of application, algorithm type, and the security self-assessment to receive the filing. [Check the specific provisions on CLaiRK↗]

The regulations on generative AI also mention that such providers must follow the algorithm filing procedure. Specifically, providers must publicly submit information, including identification, field of application, algorithm type, and the security self-assessment to receive the filing. In addition, the Chinese regulations on generative AI mention that an administrative licence for generative AI services could be mandated by other laws, but does not demand a licence itself. [Check the specific provisions on CLaiRK↗]

European Union

The EU demands pre-deployment registration for providers of high risk AI systems, or their authorised representatives on a public database of the European Commission. The information required for

registration, specified in the AI Act's Annex, includes contact details and a description of the purpose and information used by the AI system, among others. Specific registration rules apply to the real-world testing of AI systems and to AI systems employed in law enforcement and border control.

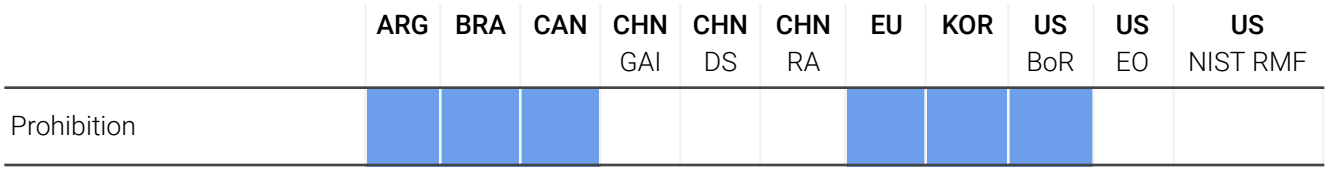
Regarding authorisation, the EU only requires prior judicial or administrative authorisation for post-remote biometric identification in targeted criminal investigations. If authorisation is denied, the use of the system must cease immediately, and any related personal data must be deleted. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI instructs the Administrator of General Services, in coordination with other agencies, to develop a framework for the authorisation of AI systems. This framework will initially focus on generative AI systems with primary purposes such as providing large language model-based chat interfaces, code-generation and debugging tools, and associated application programming interfaces, as well as prompt-based image generators. [Check the specific provisions on CLaiRK↗]

Prohibition

The OECD AI Principle 1.4 (Robustness, security, and safety) demands the **prevention of unreasonable AI safety risks**. One regulatory mechanism to prevent such risks is to prohibit certain AI systems or practices. In particular, we refer to proactive prohibitions, rather than sanctions that prohibit AI systems as a consequence of non-compliance with rules. This article systematically analyses active prohibition across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish prohibitions.



Comparison

The prohibition of AI systems occurs through **heuristics or specific prohibitions**, with considerable overlap. Heuristics are abstract rules that determine which AI systems are prohibited, such as AI systems that infringe human rights or lead to discrimination. Specific prohibitions enumerate forbidden AI systems based on their technical capabilities, such as biometric recognition, or their use context, such as work or education.

Country details

Argentina

Argentina generally prohibits AI systems that violate human rights, lead to unfair discrimination, serious harm or security risks, that are used for undue manipulation or influence, lack transparency and accountability, or perpetuate inequality. Argentina further prohibits AI systems with severe impacts on fundamental rights or people's security, with exceptions in the presence of adequate mitigation measures. The rulebook does not enumerate specific prohibited AI systems. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil generally prohibits AI systems that result in discrimination. The rulebook further prohibits AI systems that carry unacceptable risk, including AI systems that employ subliminal techniques, exploit vulnerabilities, and provide social scoring, as well as specified biometric identification systems. In addition, Brazil specifically states that AI systems whose risks cannot be sufficiently prevented or reduced are to be discontinued. [Check the specific provisions on CLaiRK↗]

Canada

Canada generally prohibits the use and deployment of AI systems that can cause serious harm to individuals and their interests. In addition, the government can issue orders to halt the use of a specific high impact AI system if the system poses a serious risk of imminent harm. Furthermore, Canada

establishes that two AI practices constitute a criminal offence: First, making available an AI system likely to cause serious physical or psychological harm to an individual or substantial damage to property. Second, making available an AI system with the intent to defraud the public and cause substantial economic loss to an individual. [Check the specific provisions on CLaiRK↗]

European Union

The EU enumerates prohibited AI practices, including AI systems that exploit vulnerabilities and employ subliminal techniques. The list also comprises specific AI technologies, including certain biometric identification systems, face recognition algorithms that rely on untargeted data scraping, personal classification algorithms that lead to detrimental or disproportionate social effects, and criminal risk assessment algorithms. In addition, the list prohibits AI systems in specific contexts, including AI systems that infer emotions in a workplace or educational institution. The Commission assesses the need to amend the list of prohibitions every year. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea does not explicitly establish a prohibition of specific AI systems but calls for the government to restrict the development of AI systems that may harm human life, body, and property. [Check the specific provisions on CLaiRK↗]

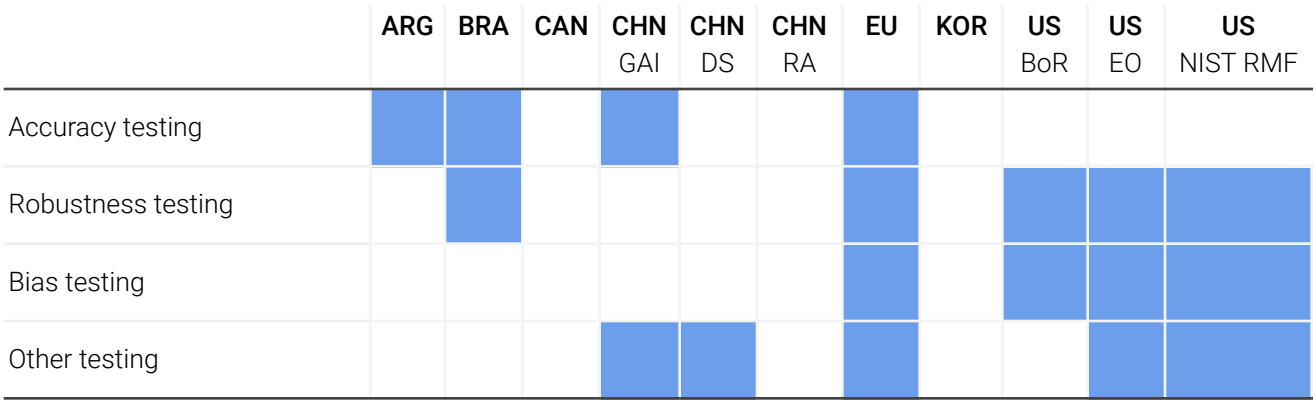
United States

The Blueprint for an AI Bill of Rights calls for the prohibition of continuous monitoring of individuals in the context of work, education, and housing, if the monitoring impacts individual rights and opportunities. [Check the specific provisions on CLaiRK↗]

Testing

The OECD AI Principle 1.4 (Robustness, security, and safety) demands the **prevention of unreasonable safety risks by AI systems**. Testing requirements are a common regulatory tool for providers to check their AI systems, both regarding general performance and specific issues such as robustness or bias. This article systematically analyses testing requirements across 11 AI rulebooks in seven jurisdictions.

We differentiate **four types of testing requirements.** Accuracy testing requirements cover the accuracy, reliability and effectiveness of AI systems. Robustness testing requirements cover the robustness and security of AI systems. Bias testing requirements involve detecting, preventing, and mitigating potential biases and discrimination in AI systems. Finally, other testing requirements encompass all AI testing measures that do not fall under the other three categories. The heatmap below visualises the testing requirements addressed by each jurisdiction.



Comparison

Testing requirements differ in the **types of AI systems they address**. In Argentina and the US (Bill of Rights and US NIST Risk Management Framework) the testing requirements apply to all AI systems. The US Executive Order instructs testing requirements for all AI systems as well as for underwriting AI models and dual-use foundation models. In Brazil and the EU, the testing requirements apply to high risk AI systems, as well as general purpose AI models in the EU. China demands testing from providers of deep synthesis and generative AI services.

The divergence in testing requirements also extends to the **timing of the tests**. Pre-deployment testing is explicitly mentioned in China (regulations on generative AI), the EU, and the US (Bill of Rights, Executive Order, and US NIST Risk Management Framework). China (regulations on generative AI and deep synthesis services) also explicitly mentions testing during the use of generative AI services and periodic reviews of deep synthesis services. In the EU, bias testing is to be conducted before, during, and after deployment. Furthermore, regarding discrimination, the US Bill of Rights also calls for ongoing testing in addition to pre-deployment testing.

Country details

Argentina

Argentina establishes that developers, providers and users of AI systems are responsible for their decisions and actions regarding AI, and must ensure its safety and reliability, as well as compliance with quality standards. Specifically, developers and providers must implement testing measures to minimise usage errors and ensure the quality and reliability of their systems. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil demands accuracy and robustness testing for operators of high risk AI systems. Regarding accuracy testing, Brazil requires an assessment of the high risk AI system reliability, including tests for accuracy, precision, and coverage. Furthermore, Brazil demands robustness testing to evaluate reliability of high risk AI systems, according to the sector and type of application. [Check the specific provisions on CLaiRK↗]

China

The Chinese regulations on generative AI impose accuracy testing and other testing requirements. Regarding accuracy testing, the regulations require those providing or utilising a generative AI service to take measures to improve the accuracy and reliability of the generated content. Furthermore, If data annotation is used in the research and development of generative AI services, providers must assess quality and verify accuracy. [Check the specific provisions on CLaiRK↗]

The Chinese regulations on deep synthesis services require providers and technical supporters to conduct periodic reviews, evaluations, and verifications of algorithms used for content generation. [Check the specific provisions on CLaiRK↗]

European Union

The EU demands accuracy, robustness, bias and other testing requirements for high risk AI systems.

- In terms of accuracy testing, high risk AI systems must meet a stringent level of precision. To ensure this, providers are mandated to implement a quality management system that includes verification procedures.
- Regarding robustness testing, the EU pursues resilience against errors and faults, especially in interactions with humans or other systems. To ensure this, high risk AI systems shall be designed and developed in a way that they achieve an appropriate level of robustness and cybersecurity.
- In terms of bias testing, the EU mandates that high risk AI systems that use data-trained models must be developed using training, validation, and testing. These datasets must comply with data governance practices, evaluating availability, quantity, and suitability to identify potential biases that could affect health, safety, rights, or lead to discrimination. The quality management system must also outline testing and validation procedures before, throughout and after development to ensure compliance.

Finally, high risk AI systems are subject to other testing requirements to identify appropriate risk management measures and ensure consistent performance for their intended purpose. These tests, which may simulate real-world conditions, are required at various stages of development and prior to market placement or service deployment. Testing is based on predefined metrics and thresholds relevant to each system's purpose. Providers of general purpose AI models must maintain detailed documentation, respect intellectual property rights, and ensure compliance with standardised protocols, including adversarial testing for systems with systemic risks. During development, any identified data gaps or deficiencies must be addressed. Additionally, providers based in the EU are permitted to test high risk AI systems under real-world conditions before market launch or service implementation. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI includes provisions regarding robustness testing, bias testing, and other testing.

- Regarding robustness testing, the Executive Order directs the creation of testing environments and tools to ensure the safety, security, and trustworthiness of AI systems, requiring collaboration between the Secretary of Energy and the Director of the National Science Foundation (NSF). The Secretary of Energy is also responsible for developing AI model evaluation tools and testbeds to address security risks, while the Secretaries of Defense and Homeland Security are tasked with piloting cyber defence AI projects. The Executive Order further mandates agencies to establish guidelines for AI red-team testing, particularly for dual-use foundation models, to ensure safe and reliable systems.

- Regarding bias testing, the Executive Order instructs AI underwriting models to be evaluated for biases affecting protected groups, with processes in place to minimise such bias. Additional guidelines may address discrimination in housing and real estate transactions by automated or algorithmic tools.

- Regarding other testing, the Executive Order tasks the Secretary of Commerce to report on current standards in testing risks resulting from synthetic content generated by AI. [Check the specific provisions on CLaiRK↗]

The NIST Risk Management Framework calls for robustness, bias, and other testing of AI systems.

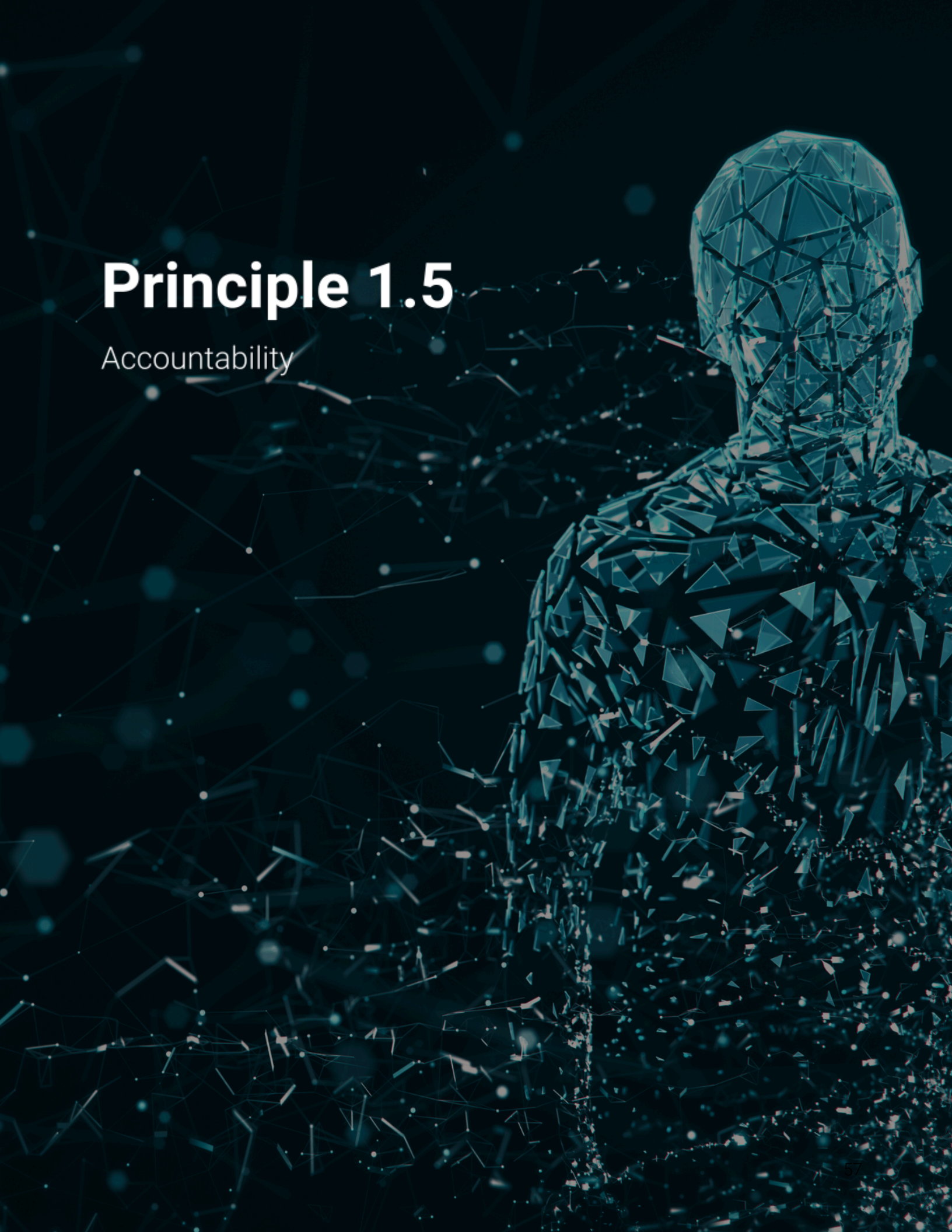
- Robustness testing should evaluate and document AI systems' security and resilience, guided by the MAP function, which sets the context for AI systems and promotes risk prevention and trustworthy development using diverse perspectives.

- Regarding bias testing, the NIST Risk Management Framework calls for the assessment and documentation of fairness, using MAP function criteria. Additionally, regarding other testing, organisations must implement practices to support other AI system tests.

- Furthermore, the NIST Risk Management Framework calls for evaluations involving human subjects to comply with protection standards and representatively reflect the targeted

population. It also calls for both qualitative and quantitative assessments of AI system performance under conditions similar to actual deployment, with thorough documentation of all procedures. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights calls for both robustness and bias testing requirements for AI systems. The Bill of Rights calls for pre-deployment robustness testing to ensure safety and effectiveness based on intended use, alongside independent evaluation. Regarding bias testing, the Bill of Rights calls for proactive equity assessments and both pre-deployment and continuous testing and mitigation measures to address algorithmic discrimination. [Check the specific provisions on CLaiRK↗]



Principle 1.5

Accountability

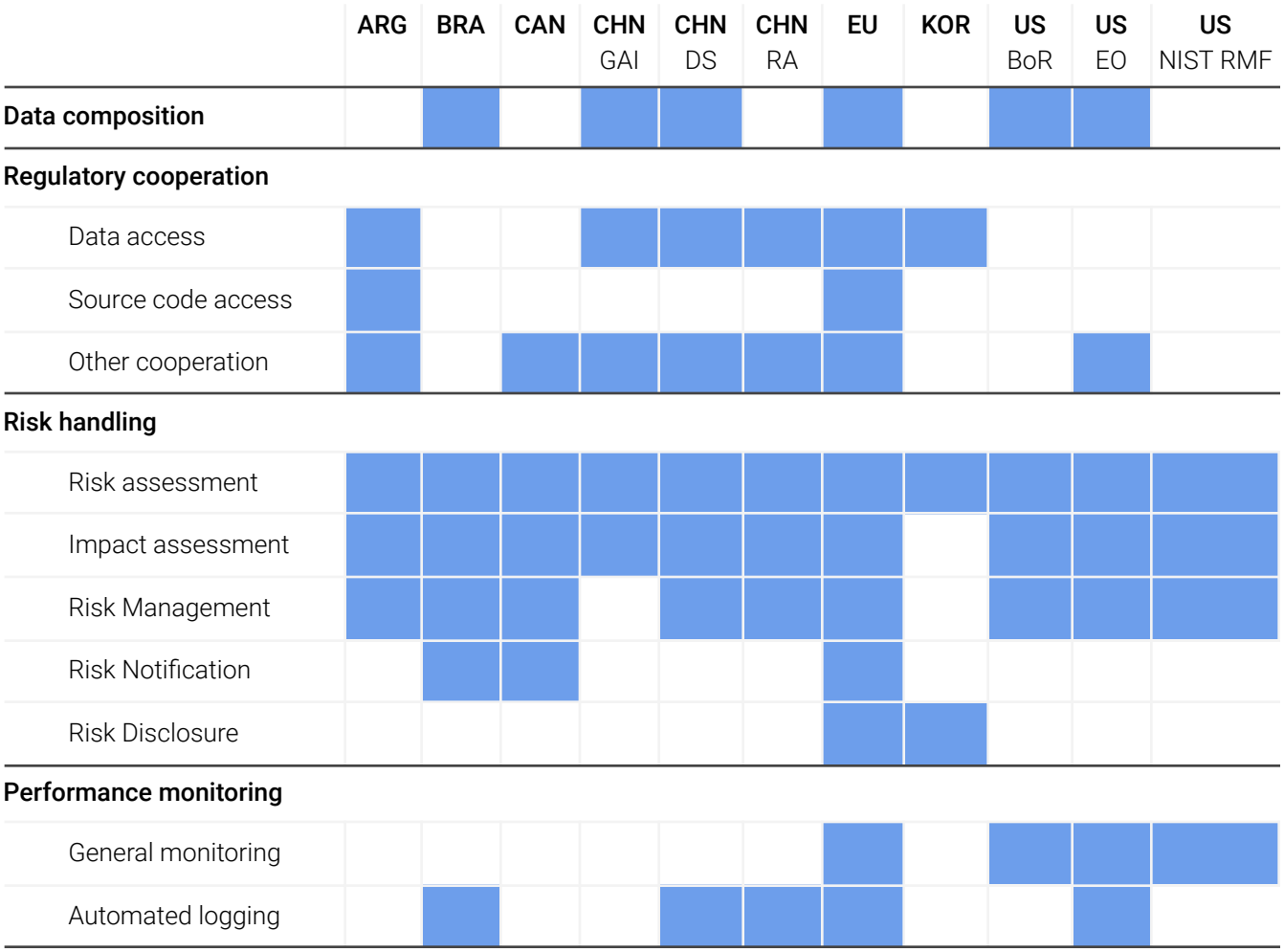
Principle 1.5: Accountability

Holding AI actors accountable for the impact of their AI systems is a goal that brings together governments across the globe. Regulators concerned with the risks of AI systems and the unforeseen consequences of AI's permeation into all economic sectors pursue accountability. The regulatory requirements imposed on AI actors, however, vary significantly.

A patchwork of regulatory requirements implements OECD AI Principle 1.5

The **OECD AI Principle 1.5** demands that AI actors should be accountable for the proper functioning of AI systems and for the respect of the OECD AI Principles. In the 2024 update of the principles, the OECD specified that AI actors should 1) ensure traceability to enable analysis of the AI system's outputs and 2) apply systematic risk management throughout the AI system lifecycle.

In national AI rules, a **patchwork of regulatory requirements** implements the OECD AI Principle 1.5. The heatmap visualises divergence within a selection of these requirements, grouped in four categories. Below, we explain each category in detail.



Data composition requirements are frequently imposed

Data composition requirements set rules regarding the **data used to train AI systems**. Since the training data influences an AI system's output significantly, faulty datasets can cause unintended consequences, for instance perpetuating biases. Data composition requirements aim to tackle this problem at the source.

Data composition requirements are **frequently imposed**, namely in six of the analysed AI rulebooks. These requirements differ, however, regarding the granular obligations for AI providers with regard to the data – from using legitimate data sources to mitigating discriminatory bias.

Regulatory cooperation are common, except regarding source code

Regulatory cooperation requirements oblige AI providers to **cooperate with government authorities**. Governments impose three kinds of cooperation requirements

- Data access requirements oblige AI providers to grant authorities access to internal data, including training datasets.
- Source code access requirements oblige AI providers to grant government authorities access to the source code of an AI application.
- Other regulatory cooperation requirements demand that AI providers cooperate with authorities, both on request and by default.

Regulatory cooperation requirements are **either frequent or rare**. Data access requirements and other regulatory cooperation requirements are prevalent in six and seven rulebooks, respectively. Source code access requirements, on the other hand, feature in two rulebooks. Granular differences persist regarding the procedures of cooperation, including defence mechanisms for firms to counter requests for data and source code access.

Risk handling requirements showcase both convergence and divergence

Risk handling requirements demand that AI providers **assess and manage risks** created by their systems. Governments impose five specific requirements:

- Risk assessment requirements demand an assessment of the risks brought by the specific AI application.
- Impact assessment requirements oblige providers to conduct an assessment of the impacts of the specific AI application.
- Risk management measures demand the mitigation and control of risks, once they are identified and assessed.
- Risk notification requirements require providers to notify the general risk level of an application or a specific risk (before an incident).
- Risk disclosure requirements require providers to publicly disclose the general risk level of an application or a specific risk (before an incident).

Risk handling requirements are **either widespread or rare**. Risk assessment is required in all jurisdictions. Impact assessment and risk management are also common requirements, established in all jurisdictions with the exception of Korea for impact assessment. Risk notification and disclosure, on the other hand, are rarely required. Adding to the patchwork, governments' elaborations of these same requirements differ, for instance regarding the specific measures to be taken to manage risks.

Performance monitoring requirements are scattered

Performance monitoring requirements demand **continuous observation of working AI systems**. Governments impose two kinds of performance monitoring requirements:

- General performance monitoring requirements refer to general obligations to monitor an AI system without specifying the attribute to be monitored.
- Automated logging requirements demand the automatic documentation of an AI system's workings, including outputs.

Performance monitoring requirements are **scattered**. General performance monitoring requirements are established in four AI rulebooks, while five AI rulebooks demand automated logging. Adding to the patchwork, governments differ in the specific information that must be monitored and logged, as well as the retention period of the monitoring data.

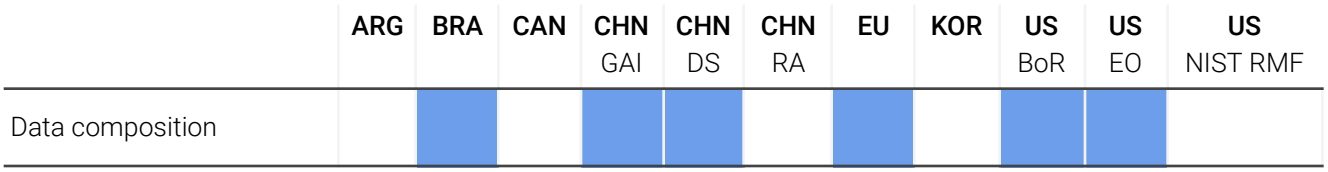
Dive deeper into each requirement

The patchwork of regulatory requirements that implement OECD AI Principle 1.5 are only the **tip of the iceberg**. Granular differences emerge even within the jurisdictions that impose the same regulatory requirements. To showcase granular divergence, we now proceed with a detailed comparative analysis of the following requirements:

- Data composition
- Regulatory cooperation
- Risk and impact assessment
- Risk management
- Performance monitoring

Data composition

The OECD AI Principle 1.5 (Accountability) demands that AI actors ensure **traceability**, including in relation to datasets, processes and decisions made during the AI system lifecycle. Data composition requirements set rules on the input data used to train AI applications. This article systematically analyses data composition requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises which jurisdictions establish data composition requirements.



Comparison

Data composition requirements differ regarding their **scope**. Brazil's requirement applies to all "AI agents." The EU regulates data composition regarding high risk AI systems. China imposes rules regarding specific technologies, namely generative AI and deep synthesis services. The US Bill of Rights contains a general, voluntary provision on data composition, while the US Executive Order instructs government agencies to include data composition in their rulemaking on AI in the health and public sectors.

Country details

Brazil

Brazil requires AI agents to implement adequate data management measures to mitigate and prevent discriminatory biases. This includes separating and organising data for training, testing, and validation, evaluating data for human cognitive biases, avoiding biases due to classification problems or lack of information about affected groups, and implementing corrective measures to prevent amplification of structural social biases. Additionally, public entities must ensure the use of accurate, relevant, updated, and representative data from secure sources when contracting, developing, or using high risk AI systems. [Check the specific provisions on CLaiRK↗]

China

China's regulations on generative AI and deep synthesis services both contain provisions on data composition. Providers of generative AI services must use data and base models with legitimate sources and take effective measures to improve the quality of training data, enhancing its authenticity, accuracy, objectivity and diversity. [Check the specific provisions on CLaiRK↗]

Providers of deep synthesis services must adopt technical methods to audit input data and synthesis results, and take measures to ensure the safety of training data. In particular, training data containing personal information must be treated in compliance with legislation on personal information protection. [Check the specific provisions on CLaiRK↗]

European Union

The EU imposes quality criteria for the datasets used to train, validate and test high risk AI systems which "make use of techniques involving the training of AI models with data." Relative to the intended purpose, these datasets must be relevant, sufficiently representative, and, to the extent possible, free of errors and complete. The datasets must further take into account the specific geographical, contextual, behavioural or functional setting in which the AI system is deployed. In addition, the EU demands data governance and management processes covering data collection and origin, data preparation (including annotation and cleaning) and data assessment regarding availability, quantity and suitability. The instructions for use of high risk AI systems must include information on the training, validation, and testing datasets, while the documentation of quality management systems must encompass data management procedures. [Check the specific provisions on CLaiRK↗]

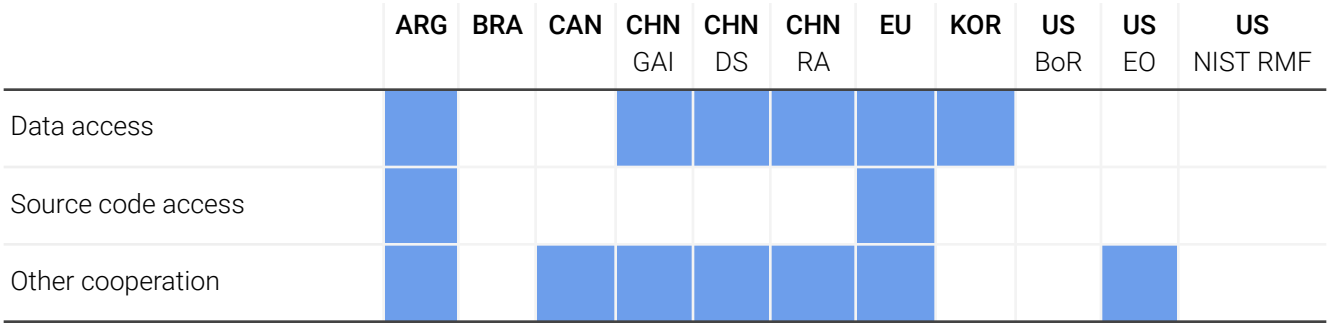
United States

The Executive Order on AI instructs several government agencies to include data composition requirements in their AI rulemaking. The Department of Health and Human Services's guidance on AI technology for the health and human services sector is to include equity principles, particularly as it regards using disaggregated data on affected populations and using representative population data sets when developing new models. The Office of Management and Budget's guidance on the use of AI in the federal government must demand a data quality assessment as part of the risk management requirements and establish an inventory of commercially available information (CAI) procured by agencies. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights and Executive Order on AI both contain provisions on data composition. The Bill of Rights calls for the use of representative data and protection against proxies for demographic features. [Check the specific provisions on CLaiRK↗]

Regulatory cooperation

The OECD AI Principle 1.5 (Accountability) demands that AI providers enable the **analysis of their systems' output**. To pursue accountability, governments demand regulatory cooperation, which includes granting access to the data or source code underlying AI systems, as well as other forms of cooperation. This article systematically analyses regulatory cooperation requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises the jurisdictions imposing regulatory cooperation requirements.



Comparison

Data access requirements differ in scope and procedure. Regarding scope, Argentina demands data access from all AI providers. The EU and South Korea demand data access from providers of high risk AI. China's requirements address providers of generative AI, deep synthesis services, and recommendation algorithms, respectively. Regarding the procedure, the EU and South Korea outline the conditions for governmental data access requests, while Argentina and China don't provide specific information.

Source code access requirements differ in level of detail and scope. Argentina establishes an obligation for all AI providers to grant the government access to source code, without providing further detail. The EU contains detailed source code access requirements for providers of high risk AI and general purpose AI. The requirements define the purposes for which access can be requested, the conditions for the validity requests, and the information to be included in the request, such as the legal basis and purpose.

Other regulatory cooperation requirements differ in scope and the nature of cooperation. Regarding scope, Argentina and Canada address all AI providers. China's regulations target providers of specific technologies (generative AI, deep synthesis services, and recommendation algorithms) with public opinion attributes or social mobilisation. The US Executive Order covers providers of dual-use foundation models, large-scale computing clusters and infrastructure as a service. The EU covers providers of high risk AI and general purpose AI. Regarding the nature of the cooperation, Argentina encourages dialogue between various stakeholders. Canada demands reporting to be shared with the government and auditors. China requires the public filing of basic information. The EU demands cooperation to mitigate AI risks and demonstrate conformity. Finally, the US Executive Order requires the reporting of specific transactions, especially with foreign counterparts.

Country details

Argentina

Argentina imposes data access, source code access, and other regulatory cooperation requirements for developers and deployers of AI systems. Specifically, Argentina establishes an obligation to cooperate with the proposed "Artificial Intelligence Oversight Authority." This cooperation includes providing access to systems, source code, data and relevant documentation for the purpose of monitoring and control. To this end, Argentina establishes the obligation to document and disclose the operation and algorithms used in AI systems. Regarding other regulatory cooperation, Argentina encourages dialogue between developers, researchers, users, and society for inclusive AI development and collaboration between educational institutions, research centres, and industry for AI training. Moreover, Argentina encourages the national and international exchange of knowledge, data, and best practices and seeks international cooperation to promote common standards and policies. [Check the specific provisions on CLaiRK↗]

Canada

Canada requires AI providers to cooperate with the government and independent auditors by providing access to internal records. To this end, providers are required to keep records on various other regulatory requirements. All AI providers must document how data is anonymised and used, as well as their AI impact assessment. In addition, high impact AI providers must keep records on measures to identify, assess and mitigate the risks of harm or biassed output, as well as compliance monitoring. The government can request access to records by order. [Check the specific provisions on CLaiRK↗]

China

The Chinese regulations on recommendation algorithms, deep synthesis services, and generative AI require data access to enable supervision and other regulatory cooperation to enter the "algorithm filing" system.

To enable supervision, including security assessment, each regulation imposes data access requirements. Recommendation algorithm providers must grant access to network logs. Deep synthesis services providers must grant access to "technical data." Generative AI providers must provide the source of training data and labelling rules, among others.

To enter China's "algorithm filing" system, providers must cooperate by publicly submitting information, including identification, field of application, algorithm type, and the security self-assessment. The regulation on recommendation algorithms originally established this requirement for providers of services with public opinion attributes or social mobilisation potential. The regulations on deep synthesis services and generative AI both mention that such providers must follow the algorithm filing procedure.

Check the specific provisions on CLaiRK: generative AI↗ I deep synthesis services↗ I recommendation algorithms↗

European Union

Concerning data access requirements, the EU requires high risk AI providers to provide access to market surveillance authorities. Specifically, authorities can issue a "reasoned request" for access to logs as well as training, validation, and testing datasets used during the development of high risk AI systems.

The EU's source code access requirements apply to providers of high risk AI and general purpose AI. High risk AI providers must grant source code access to market surveillance authorities with two conditions: The access is necessary to assess conformity with the law and other auditing procedures and verifications are exhausted. Regarding general purpose AI, the newly established AI Office can request access to source code when conducting evaluations to assess compliance and investigate systemic risk. Specifically, the request can demand access to the general purpose AI model through APIs or further technical means and tools, including source code. The request must state its legal basis, purpose, and reasons.

The EU further demands other regulatory cooperation requirements regarding high risk AI and general purpose AI. For instance, the EU demands cooperation with national authorities that take action to reduce and mitigate AI risks. In addition, national authorities can issue "reasoned requests" to receive information necessary to demonstrate conformity with requirements. [Check the specific provisions on CLaiRK↗]

South Korea

High risk AI business operators must provide information to users if it's requested and necessary to check whether the users were disadvantaged by the AI. However, the high risk AI business operators are not obliged to do so if a special provision in other laws or a "legitimate reason" exists. Further, South Korea mandates all AI systems to provide access to data for the dispute resolution committee if the data is necessary for the resolution of a dispute. Although, access to the data must not be provided if a "legitimate reason" exists. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI contains regulatory cooperation requirements regarding dual-use foundation models, large-scale computing clusters, and infrastructure as a service (IaaS), under the Defense Production Act.

Regarding dual-use foundation models, the Executive Order instructs the Secretary of Commerce to require companies developing such models to provide the federal government with ongoing information and records. This extends to activities related to training, developing, or producing dual-use foundation models, the ownership and possession of the model weights, the results of red-team testing, and cybersecurity measures to protect the models.

Regarding large-scale computing clusters, the Executive Order instructs the Secretary of Commerce to require cooperation from companies and individuals that acquire, develop, or possess such clusters. Specifically, the Executive Order demands information on the acquisition, development, or possession of clusters, including their location and computing power. The requirement applies to models trained using computing power greater than 10^{26} integer or floating-point operations (10^{23} for models using

primarily biological sequence data) and to computing clusters that have machines physically co-located in a single data centre, transitively connected by data centre networking of over 100 Gbit/s, and with a theoretical maximum computing capacity of 10^{20} integer or floating-point operations per second for AI training.

Regarding Iaas, the Executive Order instructs the Secretary of Commerce to propose regulations that require US IaaS providers to report each transaction with a foreign party that could train a large AI model with potential capabilities for "malicious cyber-enabled activity." The thresholds for such activity correspond to those for "large-scale computing clusters" (see above). The report must include the identity of the foreign person and the training run, among others. Furthermore, the Executive Order prohibits foreign resellers of US IaaS products from reselling such products, unless they report the same information to the IaaS provider and the Secretary of Commerce. [Check the specific provisions on CLaiRK↗]

Risk and impact assessment

The OECD AI Principle 1.5 (Accountability) demands the management of **unreasonable AI risks**, beginning with the identification, assessment, and evaluation of risks and their impact. This article systematically analyses risk and impact assessment requirements across 11 AI rulebooks in seven jurisdictions.

We distinguish between **risk and impact assessment**, which overlap in some AI rulebooks. The consistent application of our taxonomy as a common language for AI rules is a core value proposition. In our taxonomy, risk assessments estimate the likelihood of an undesirable outcome, while impact assessments quantify the consequences. Simply put, there is a risk of falling off the bed, and the impact could be to break an arm. Notably, both assessments differ from an audit, carried out by an external entity rather than the AI provider, and from security assessments, a Chinese idiosyncrasy outlined below. The heatmap below visualises the jurisdictions that foresee risk and impact assessment.

	ARG	**BRA**	**CAN**	**CHN** GAI	**CHN** DS	**CHN** RA	**EU**	**KOR**	**US** BoR	**US** EO	**US** NIST RMF
Risk assessment	■	■	■	■	■	■	■	■	■	■	■
Impact assessment	■	■	■	■	■	■	■		■	■	■

Comparison

AI rulebooks differ in whether they foresee an **interaction between risk and impact assessment**. In some rulebooks, the first assessment triggers the second assessment. Brazil requires all AI providers to conduct a risk assessment to determine whether their systems are "high risk", which triggers compliance obligations including impact assessment. In contrast, Canada requires all AI providers to conduct an impact assessment to determine whether their systems are "high impact," which triggers compliance obligations including risk assessment. Argentina follows a similar structure, although it does not clearly distinguish the risk from the impact assessment. The second kind of rulebook does not foresee an interaction between the two assessments. The EU demands high risk AI providers to conduct both risk and impact assessment. China's security assessment similarly covers both. In the US, the two assessments don't interact. Finally, South Korea only mandates risk assessment, rendering any interaction impossible.

The **timing** of assessments differs regarding whether they are due before deployment of the AI system and whether they must be continuously updated throughout the AI's lifecycle. Risk assessment requirements apply before deployment in Argentina, Brazil, China, the EU, and the US (Bill of Rights), and throughout the AI lifecycle in Argentina, the EU, and the US (Bill of Rights and NIST Risk Management Framework). Impact assessments apply before deployment in Argentina, China, and the EU. Argentina and Brazil demand continuous impact assessment. The EU only requires updates when the factors underlying the assessment change, while the US Bill of Rights calls for the evaluation to be performed whenever possible to protect from the impacts. Notably, only the EU includes a reactive risk

assessment requirement, triggered by the occurrence of an incident. Canada and South Korea don't specify the timing of assessment requirements.

Finally, assessments differ regarding the **types of risk and impact** they address. Risk assessments most often addressed security and safety risks, namely in Argentina, South Korea, and the US NIST Risk Management Framework (RMF). Rarer risk types include fundamental rights (Argentina), third-party risks including intellectual property (US NIST RMF) and critical infrastructure (US EO). Notably, jurisdictions differ in their assessment of unforeseen risks: The US NIST RMF calls for an assessment of existing, unanticipated, and emergent risks. South Korea limits the assessment to risks associated with the intended use. The EU extends the requirement to reasonably foreseeable misuse. Regarding impact assessments, fundamental rights are covered in Argentina, Brazil, and the EU. In addition, Brazil and the US Bill of Rights cover discriminatory impacts, while Argentina and Brazil cover safety and security. Canada does not specify the scope of covered impacts.

Country details

Argentina

Argentina's risk and impact assessment requirements overlap considerably. All AI providers must conduct an impact assessment before deployment, to identify risks. Then, providers must implement measures to mitigate the risks identified during the "risk assessment" – a term used only once. This assessment must be updated periodically and serves as the basis for the risk classification, which in turn triggers a range of requirements. All providers must implement risk management measures to minimise the impact of the identified risks on fundamental rights and security. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil requires all suppliers of AI systems to carry out a preliminary assessment before market deployment, to identify risks and evaluate whether the AI risk level is excessive, high, or low. Providers of high risk AI systems must then uphold a range of obligations, including impact assessment. The impact assessment covers societal impacts, including individual rights and discrimination and must be conducted throughout the AI system's lifecycle by a functionally independent company unit with the necessary technical, scientific and legal knowledge. [Check the specific provisions on CLaiRK↗]

Canada

Canada demands that all persons responsible for an AI system conduct an impact assessment. The assessment determines whether the AI system is classified as "high-impact," a classification that triggers a flurry of other obligations, including risk assessment. Namely, high-impact AI systems must identify and assess the risks of harm or biassed output that could result from the use of their system. The criteria for classification as high-impact are yet to be developed by the government. [Check the specific provisions on CLaiRK↗]

China

China's regulations on recommendation algorithms, deep synthesis services, and generative AI all require providers with public opinion attributes or social mobilisation capabilities to conduct a security assessment.[3] The assessment evaluates the AI system's vulnerabilities, threats, and compliance with security standards. Providers are responsible for the security assessment, which is submitted to local government authorities, in order to enter the mandatory "algorithm filing" regime. The regulations on deep synthesis services mention that authorities oversee security assessment and emphasise providers' cooperation duties. Check the specific provisions on CLaiRK: recommendation algorithms↗ I deep synthesis ↗ I generative AI↗]

European Union

The EU requires providers of high risk AI systems[4] to conduct risk assessments. The prior classification as high risk is determined by a list of specific AI technologies and use cases. High risk AI providers must continuously identify and analyse known and foreseeable risks associated with the AI system, including to health, safety, and fundamental rights, in the context of risk management. In addition, providers must re-assess risks following serious incidents. During the AI Act negotiations, the European Parliament introduced a new provision requiring impact assessment. Art. 27 demands that deployers of high risk AI systems (with certain exceptions) perform a fundamental rights impact assessment prior to deployment. The assessment must cover the processes in which the AI system is used, the time period and frequency of use, human oversight, affected persons, risks of harm likely to have an impact on affected persons, and measures to be taken if those risks materialise. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea requires operators of high risk AI systems to determine if there are significant risks to life or safety related to their AI system, but does not mandate impact assessment. The prior classification as high risk is determined by a list of specific AI technologies and use cases deemed to have a significant impact on the protection of people's lives, safety, and rights. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI instructs the Secretary of Commerce, acting through NIST, to draft guidelines for the development and deployment of safe, secure, and trustworthy AI systems, including a companion resource to the NIST Risk Management Framework for generative AI and guidance for the evaluation of AI capabilities and harms. The Executive Order further mandates a risk assessment

[3] In addition, deep synthesis service providers and their technical supporters must conduct security assessments if their tools generate or modify biometric information, such as human faces and voices, or certain non-biometric information, such as content implicating national security or social welfare.

[4] The EU also establishes risk identification and assessment obligations for providers of general purpose AI models with systemic risk. The prior classification as "systemic risk" is determined based on the system's impact capabilities or by a decision from the European Commission.

regarding critical infrastructure and impact assessment regarding the impact of government use on AI on people's rights and safety. [Check the specific provisions on CLaiRK↗]

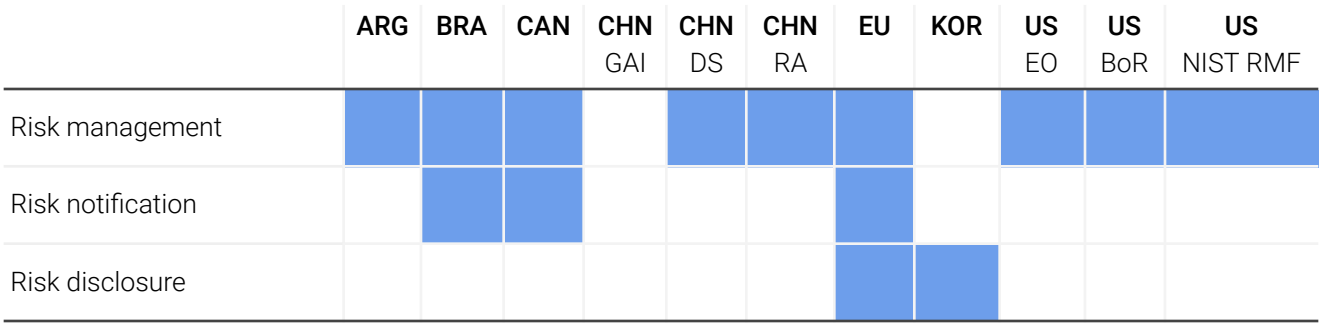
The NIST Risk Management Framework calls for the mapping of legal risks including third-party rights, the regular evaluation of safety risks, and the continuous identification of existing, unanticipated, and emergent AI risks. Beyond risk assessment, the NIST Risk Management Framework demands risk tracking and third-party risk monitoring, as well as the assessment of impacts on the environment and sustainability. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights demands that pre-deployment risk identification should ensure the safety of AI systems and that automated systems should be developed in consultation with stakeholders, to identify risks. The Bill of Rights also calls for a voluntary independent evaluation of algorithmic impacts, without specifying whether this is an internal or external assessment. [Check the specific provisions on CLaiRK↗]

Risk management

The OECD AI Principle 1.5 (Accountability) requires **risk management** to prevent unreasonable safety risks by AI systems. Risk management requirements call for the mitigation or control of AI risks. Risk management is complemented by requirements to assess risks, and mandates to notify or publicly disclose AI risk. This article systematically compares different risk management as well as risk notification and disclosure requirements across 11 rulebooks in seven jurisdictions.

AI rulebooks differ in the specific **regulatory requirements** they employ to address AI risks. Risk management requirements broadly oblige providers to mitigate and control AI risks. This requirement often follows mandates to assess AI risks. In addition, providers may be required to notify or publicly disclose AI risk. The heatmap below visualises the jurisdictions that require risk management, notification, and disclosure.



Comparison

Risk management requirements differ regarding the AI systems they target and their timing. Regarding the target, Brazil, Canada, and the EU only require risk management for high risk (or impact) AI systems, while the other rulebooks extend this requirement to all AI providers. Regarding timing, risk management requirements are continuous, but differ in whether they apply before market deployment. Only Argentina and the US Bill of Rights demand pre-deployment risk.

Risk notification requirements differ regarding the reason and addressee for the justification. Brazil and the EU require notification when a system meets the criteria for a certain risk class. Canada and Brazil (second requirement) demand notification when new risks arise that might cause harm. Regarding the addressee, Brazil, the EU, and Canada all demand notification to authorities. Brazil also demands the notification of affected people regarding new risks.

Risk disclosure requirements are rare and very similar. The EU and South Korea both require providers of high risk AI systems to disclose risks regarding the health and safety of users.

Country details

Argentina

Argentina requires the identification and assessment of potential risks before any AI system is deployed. The entity responsible for an AI system is obliged to implement adequate measures to mitigate previously identified risks. [Check the specific provisions on CLaiRK↗]

Brazil

Brazil requires suppliers and operators of high risk AI systems to implement risk mitigation measures throughout the lifecycle of the AI system. Brazil also requires two kinds of risk notification. First, in the context of the impact assessment, suppliers and operators must notify authorities of high risk AI systems. Second, if suppliers or operators take notice of an unexpected risk, they are obliged to notify both authorities and affected persons. [Check the specific provisions on CLaiRK↗]

Canada

Canada requires persons responsible for high impact AI systems to mitigate risks, specifically harm or biassed output resulting from the use of their systems. In addition, Canada requires persons responsible for high impact AI systems to notify the Minister of Industry when the use of the system causes or is likely to cause harm. Notification is due as soon as feasible. [Check the specific provisions on CLaiRK↗]

China

The Chinese regulations on recommendation algorithms and deep synthesis services require providers to implement management systems for algorithm and data security, among others. In addition, deep synthesis service providers and technical supporters must adopt measures to rectify and eliminate threats if government authorities find a significant information security risk in the security assessment. The regulations on generative AI don't require risk management, although they mention risk prevention as an area of potential collaboration between industry organisations, firms, and research institutions. [Check the specific provisions on CLaiRK: recommendation algorithms↗ I deep synthesis↗]

European Union

The EU establishes a risk management requirement for high risk AI systems. The risk management system must address risks that can be reasonably mitigated or eliminated, throughout the AI system's life cycle. Providers must continuously monitor and evaluate risks that may emerge from the intended use or reasonably foreseeable misuse of the AI system. For providers of high risk AI systems that continue to learn post-deployment, the risk of biassed outputs influencing future operations, including potential feedback loops, must adopt suitable mitigation strategies.

In addition, the EU requires risk notification. Providers of general-purpose AI systems must notify the European Commission if the AI system meets the prerequisites to be classified as "systemic risk" within two weeks.

Furthermore, the EU requires risk disclosure, since the instructions for the use of high risk AI systems must state circumstances that can lead to risks to health, safety, or fundamental rights. [Check the specific provisions on CLaiRK↗]

South Korea

South Korea does not establish a risk management requirement, but demands risk disclosure. Operators of high risk AI systems must explain possible serious risks to life or physical safety of users in an understandable manner. [Check the specific provisions on CLaiRK↗]

United States

All three US AI rulebooks demand risk management.

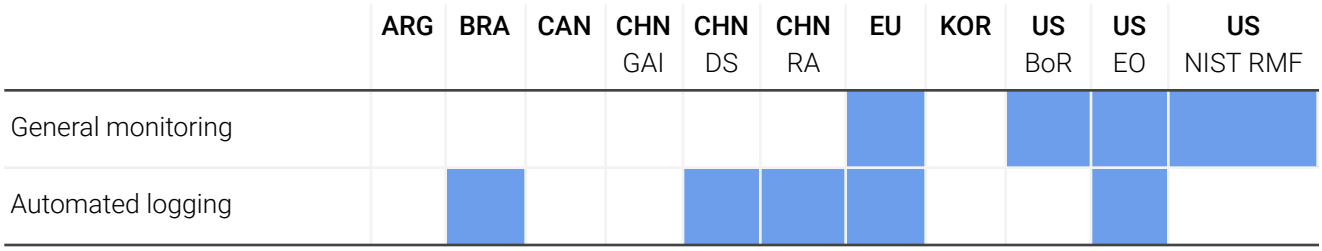
The Executive Order on AI instructs the Office of Management and Budget to issue guidance regarding government use of AI, including minimum risk management practices such as human consideration. In addition, the Executive Order references the NIST Risk Management Framework in two mandates to draft guidelines. First, the Secretary of Commerce, acting through the NIST, must draft guidelines for safe, secure, and trustworthy AI, and develop a companion resource to the NIST Risk Management Framework for generative AI. Second, the Secretary of Homeland Security must draft guidelines for the use of AI in critical infrastructure, incorporating the NIST Risk Management Framework. [Check the specific provisions on CLaiRK↗]

The NIST Risk Management Framework calls for the establishment of a risk management process to address AI risks. Risk management spans from the identification and tracking of risks, including previously unknown and third-party risks, to their mitigation, transfer, avoidance, or acceptance, aiming to reduce the magnitude or likelihood of potential impacts. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights calls for automated systems to undergo pre-deployment risk identification and mitigation. [Check the specific provisions on CLaiRK↗]

Performance monitoring

The OECD AI Principle 1.5 (Accountability) states that AI actors should be **accountable for the proper functioning of AI systems**. Performance monitoring requirements demand that AI actors continuously observe the workings of their AI systems. We distinguish between general performance monitoring requirements and logging requirements, which require automatic documentation of AI systems' workings. This article systematically analyses performance monitoring requirements across 11 AI rulebooks in seven jurisdictions. The heatmap below visualises the jurisdictions that demand performance monitoring.



Comparison

General performance monitoring requirements differ in scope and bindingness. The EU imposes a binding post-market monitoring regime for high risk AI systems, demanding continuous analysis of the performance and compliance of AI systems throughout their life cycle. The US Bill of Rights, NIST Risk Management Framework, and Executive Order all include provisions on performance monitoring that apply to all AI systems, regardless of risk, on a voluntary basis.

Automated logging requirements differ regarding the information that must be logged and the retention period. The EU provides the most detail on the logging information, including the time frame of each use of a high risk AI system, as well as the input and reference data. The US Executive Order demands logging regarding foreign transactions of US Infrastructure as a Service products, to train AI models that could be used for malicious purposes. China demands the storage of "network logs," while Brazil requires logging to assess robustness and discrimination without specifying the information to be logged. Regarding the retention period, only the EU states a specific timeframe, namely a minimum of six months.

Country details

Brazil

Brazil demands automated logging for AI agents who supply or operate high risk AI systems. Such providers must use "automatic system operation recording tools" to enable the assessment of the AI system's accuracy and robustness, the investigation of potential discriminatory effects, and the documentation of risk mitigation measures. [Check the specific provisions on CLaiRK↗]

China

The Chinese regulations on deep synthesis services and recommendation algorithms impose automated logging requirements. Deep synthesis service providers must store network logs related to unlawful and harmful content and also store "log information" when required by other legal frameworks. [Check the specific provisions on CLaiRK↗]

Recommendation algorithm providers must store network logs in the context of cybersecurity, to enable collaboration with government bodies. [Check the specific provisions on CLaiRK↗]

European Union

The EU requires automated logging and general performance monitoring for high risk AI systems. Automated logging is required throughout the life cycle of high risk AI systems. Providers must record events that are relevant to identify increased risk and to facilitate post-market monitoring. Specifically, the logging obligation covers the period of each use of the system, the reference database against which input data was checked, the input data for which the search has led to a match, and the identification of natural persons involved in the verification of the results. Providers must keep automatically generated logs for a minimum of six months and grant authorities access to logs.

High risk AI providers must further implement a post-market monitoring system that is proportional to the nature and risk of the AI system. The monitoring system must systematically collect, document and analyse relevant data regarding the performance and compliance of AI systems throughout their life cycle. The data can be provided by deployers or collected from other sources and includes an analysis of the interaction with other AI systems, where relevant. The monitoring system must be based on a post-market monitoring plan that is part of the technical documentation. In addition, providers must monitor risks as part of their risk management system and deployers must monitor systems based on their instructions for use. [Check the specific provisions on CLaiRK↗]

United States

The Executive Order on AI instructs government agencies to draft guidance regarding general performance monitoring and rules regarding automated logging.

- Regarding general performance monitoring, the Executive Order instructs the Office of Management and Budget to draft guidance regarding federal government use of AI, including continuous monitoring and evaluation of deployed AI. Moreover, the Executive Order instructs the Department of Health and Human Services to draft a strategic plan on responsible deployment and use of AI in the health and human services sector, which comprises the real-world performance monitoring of AI and the monitoring of algorithmic performance regarding discrimination and bias.

- Regarding automated logging, the Executive Order instructs the Secretary of Commerce to propose regulations that require US Infrastructure as a Service (IaaS) providers to ensure that foreign resellers verify the identity of foreign persons that obtains an IaaS account. Specifically, foreign resellers must report information including the identity of such foreign persons, the means of payment, the e-mail and telephone contact, the Internet Protocol addresses, and the date and time of each access. [Check the specific provisions on CLaiRK↗]

The NIST Risk Management Framework calls for monitoring throughout AI systems' lifecycle. During production, the functionality and behaviour of the AI system and its components should be monitored. Before deployment, developers should determine whether the system achieves its intended purposes and stated objectives, and thus whether deployment should proceed. Finally, post-deployment monitoring should include mechanisms to capture and evaluate user input, as well as mechanisms to appeal and override, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use. [Check the specific provisions on CLaiRK↗]

The Blueprint for an AI Bill of Rights calls for ongoing performance monitoring to demonstrate that AI systems are safe and effective for their intended use. Monitoring should be complemented by pre-deployment testing, risk identification and mitigation, and adherence to domain-specific standards. [Check the specific provisions on CLaiRK↗]

Annex: Analysed AI rulebooks

Our analysis focuses on the text of the following 11 AI rulebooks from seven jurisdictions. Notably, these rulebooks are diverse regarding their:

- legal nature, including laws and executive regulations, as well as non-binding and government-facing frameworks;
- rationale, including product safety and fundamental rights protection; and
- current lifecycle stage, including early-stage proposals and fully implemented rules.

Country	**Rulebook**	**Lifecycle stage**
Argentina	Proposed legal framework to regulate the development and use of AI *[Proyecto de ley 2505-D-2024, Marco legal para la regulación del desarrollo y uso de la Inteligencia Artificial]*	Under deliberation (introduced on 8 June 2023)
Brazil	Proposed Bill on the use of AI *[Projeto de Lei 2338/2023, Dispõe sobre o uso da Inteligência Artificial]*	Under deliberation (introduced on 3 May 2023)
Canada	AI and Data Act (Part 3 of Bill C-27, the Digital Charter Implementation Act 2022)	Under deliberation (introduced on 16 June 2022)
China	Regulations on the Management of Algorithm Recommendation for Internet Information Services *[互联网信息服务算法推荐管理规定]*	In force (published on 4 January 2022)
China	Regulations on the Management of Deep Synthesis Internet Information Services *[互联网信息服务深度合成管理规定]*	In force (published on 11 December 2022)
China	Interim Measures for the Management of Generative AI Services *[生成式人工智能服务管理暂行办法]*	In force (published on 13 July 2023)
European	AI Act	Adopted (last passage on

Union	(as adopted by the Council of the European Union)	25 May 2024)
Republic of Korea	Bill on AI Liability *[인공지능책임법안 (2120353)]*	Under deliberation (introduced on 28 February 2023)
United States	Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI	Adopted (published on 30 October 2023)
United States	National Institute of Standards and Technology (NIST) AI Risk Management Framework	In force - voluntary (adopted on 30 March 2023)
United States	Blueprint for an AI Bill of Rights	In force - voluntary (adopted in October 2022)