



Digital Policy Alert

# AI Transparency and Explainability

Operationalising the  
OECD AI Principle 1.3

Written by Tommaso Giardini and Nora Fischer

Edited by Johannes Fritz

In collaboration with the



Law and  
Economics  
Foundation  
St. Gallen

An initiative of the

**St. Gallen  
Endowment**  
for Prosperity through Trade

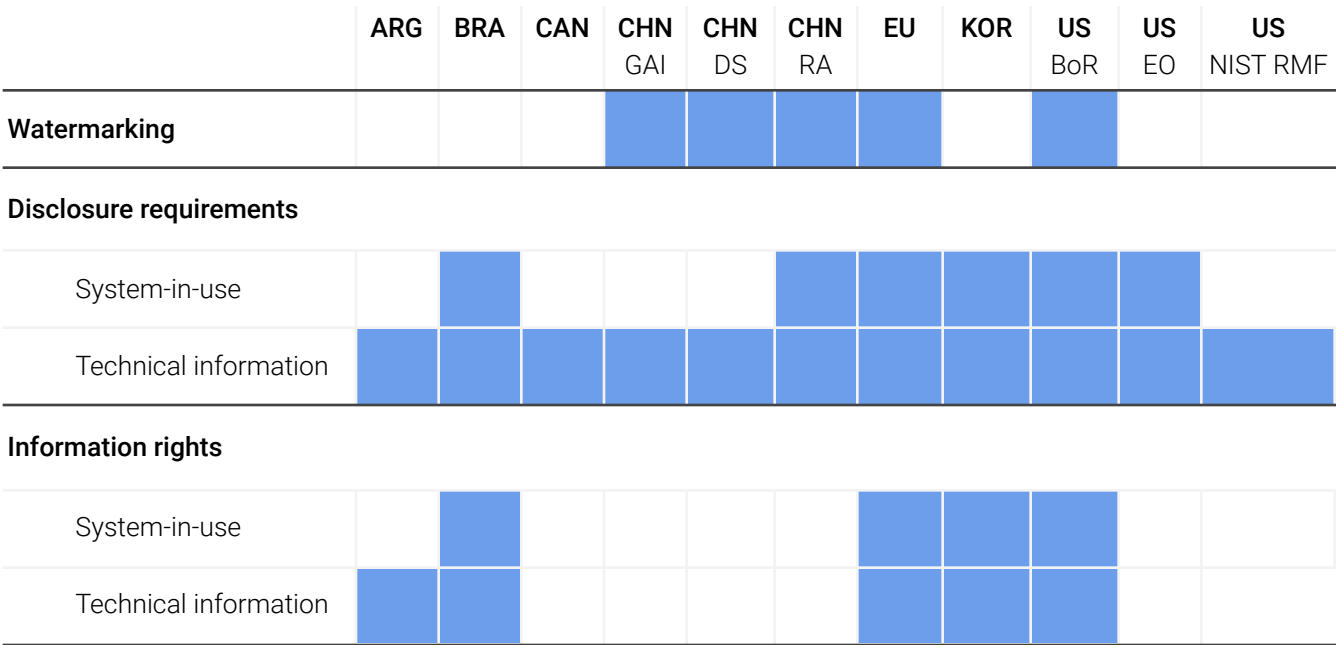
# Principle 1.3: Transparency and explainability

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.<sup>1</sup> 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.



<sup>1</sup> To provide a common language for international AI rules, we analyse rulebooks that differ in their legal nature and current lifecycle stage. For China, we analyse the regulations on generative AI (“GAI”), deep synthesis services (“DS”) and recommendation algorithms (“RA”). 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”).

## 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 six jurisdictions. 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 1.3 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.

	ARG	BRA	CAN	CHN GAI	CHN DS	CHN RA	EU	KOR	US BoR	US EO	US NIST RMF
Watermarking											

## 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↗](#) | [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.

	ARG	BRA	CAN	CHN GAI	CHN DS	CHN RA	EU	KOR	US BoR	US EO	US NIST RMF
Disclosure: System-in-use											

## 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.

	ARG	BRA	CAN	CHN GAI	CHN DS	CHN RA	EU	KOR	US BoR	US EO	US NIST RMF
Information right: system-in-use											
Information right: technical											

## 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↗](#)]