Overview

Addressing Transparency & Explainability When Using AI Under Global Standards

Arsen Kourinian, Mayer Brown

Bloomberg
Law

Addressing Transparency & Explainability
When Using AI Under Global Standards

Contributed by **Arsen Kourinian**, *Mayer Brown*

Editor's Note: For additional guidance on practice-specific areas of risk associated with the use of generative and other forms of AI, see our AI Legal Issues Toolkit. For additional information on laws, regulations, guidance, and other legal developments related to AI, visit In Focus: Artificial Intelligence (AI).

Global artificial intelligence (AI) principles, laws, and guidelines often require the transparent use of AI and an explanation on how it works. Commonly referred to as "transparency" and "explainability," these principles are important components organizations should consider as part of their AI governance plan.

To address these principles, organizations developing or using AI should consider drafting a public-facing AI notice, using plain and easy-to-understand language that incorporates common requirements observed under global guidelines and laws.

What Are Transparency & Explainability?

Global AI standards often group transparency and explainability together because they are interconnected terms. Transparency answers the question "what happened" in the AI system, while explainability addresses "how" a decision was made using AI. See the National Institute of Standards and Technology's (NIST) **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**.

Organizations are transparent when they provide meaningful information so that individuals are aware that they are interacting with AI (e.g., chatbots), content was AI-generated (e.g., output of generative AI), or a decision was made about the individual using AI (e.g., decision to invite a candidate for an interview). See the Organisation for Economic Co-operation and Development's (OECD) **Recommendation of the Council on Artificial Intelligence**.

Explainability means that organizations should provide individuals with a plain-language explanation of the AI system's logic and decision-making process so that individuals know how the AI generated the output or decision. See the White House's **Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People**.

Why Are Transparency & Explainability Important?

Organizations should provide individuals impacted by AI systems with a transparency and explainability notice for several reasons.

Initially, an AI notice is important to gain the public's trust and confidence in their AI systems by empowering individuals to better understand the AI systems and providing them a mechanism to exercise rights, challenge AI decisions, and take appropriate recourse. See the UK Information Commissioner's Office's **Explaining Decisions Made With AI**. It is also important to have an open relationship with the public through an AI notice because it provides an opportunity for the organization to solicit external feedback regarding how its AI systems function in the real environment, detect model drift, and establish bug bounty programs to help improve the AI systems.

Next, transparency and explainability are widely recognized AI principles necessary for trustworthy AI. For example, the OECD includes transparency and explainability as one of its five values-based principles. See **OECD AI Principles Overview**. With countries on six continents committed to the OECD's AI principles, including the G7 (US, UK, Japan, Canada, France, Germany, and Italy), transparency and explainability will be important components of future AI regulations in major jurisdictions. See **Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System**.

Lastly, it is important for organizations to provide a transparency and explainability notice because it is required in different contexts under domestic and international laws. A non-exclusive list of some of these laws are provided in the chart below.

Country	Law
US (federal)	Federal Trade Commission (FTC) guidance and **Section 5 of the FTC Act**. See the FTC's **Using Artificial Intelligence and Algorithms**.
US (California)	California's Chatbot Law, **Cal. Bus. & Prof. Code § 17941** (making it unlawful to mislead online users that they are interacting with a bot and requiring a disclosure that a person is interacting with a bot to avoid liability). California Consumer Privacy Act, as amended by the California Privacy Rights Act (current **working draft** of regulations incorporates transparency and explainability requirements in the AI pre-use notice).
US (Colorado)	Colorado Privacy Act Rules, **4 CCR 904-3-9.03** (requiring a transparency and explainability notice when a consumer's personal data is used for profiling in furtherance of decisions that produce legal or other similarly significant effects).
US (Illinois)	Illinois Artificial Intelligence Video Interview Act, **820 ILCS 42/5** (requiring transparency and explainability notice before using AI to analyze a job applicant's video interview).
US (New York City)	**New York City Local Law 144 of 2021** (prohibiting employers and employment agencies from using automated employment decision tools (AEDT) to screen job candidates or employees for promotion unless, among other things, they provide prior notice regarding their use of AEDT).
EU	EU AI Act, **Artificial Intelligence – Questions and Answers**, European Commission (stating that for certain AI systems, such as chatbots, the EU AI Act will impose transparency requirements so that users are aware that they are interacting with a machine). General Data Protection Regulation, **Article 5(1)(a)** (stating that personal data shall be "processed lawfully, fairly and in a transparent manner in relation to the data subject"); **Article 12(1)** ("The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child."); **Article 13**(2)(f) (requiring a controller to describe in a privacy notice "the existence of automated decision-making, including profiling . . . and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject"); and **Article 14**(2)(g) (same).
China	**Personal Information Protection Law**, Article 24 (requiring transparency for automated decisionmaking).
Nigeria	**Implementation Framework of the Nigeria Data Protection Regulation** (requiring transparency and consent before making a decision based solely on automated processing that produces legal effects concerning or significantly affecting data subjects).
Brazil	Brazilian Data Protection Law (LGPD), **Article 20** (requiring a controller to provide clear and adequate information regarding the criteria and procedures used for an automated decision).
Canada (Quebec)	Quebec's Act Respecting the Protection of Personal Information in the Private Sector **§ 12.1** ("Any person carrying on an enterprise who uses personal information to render a decision based exclusively on an automated processing of such information must inform the person concerned accordingly not later than at the time it informs the person of the decision.").

Components of a Transparency & Explainability Notice

To address the transparency and explainability requirements, organizations may prepare a public-facing notice regarding their use of AI systems, akin to a privacy policy. The AI notice should be written in plain-language and easy to understand, which organizations can validate using readability tools, such as the Fry readability graph, the Gunning Fog Index, and the Flesch-Kincaid readability test. Organizations may also consider using visualization tools, graphical representations, and/or summary tables in their AI notice to enhance readability. See the Singapore Personal Data Protection Commission's **Model Artificial Intelligence Governance Framework**.

To help address global laws and guidelines for an AI notice, organizations may consider including the elements below for transparency and explainability.

(A) Provide the name of the organization accountable for the AI systems and its outcomes. See the UK Department for Science, Innovation & Technology's **A Pro-Innovation Approach to AI Regulation**.

(B) A statement that individuals are interacting with AI, the nature and purpose of the AI, and what decisions are made using an AI system. See the European Commission's **Artificial Intelligence – Questions and Answers**; **4 CCR 904-3-9.03**.

(C) The type of data (including personal and sensitive data) that were or will be processed as part of the AI decision, along with the data used to train the AI. See **4 CCR 904-3-9.03**; the UK Department for Science, Innovation & Technology's **A Pro-Innovation Approach to AI Regulation**. Organizations should also maintain technical documentation and a data provenance record to document the lineage of the data and their right to use IP-protected data. See **ISO/IEC 42001:2023**.

(D) An explanation of the logic used in the AI decision, including the key parameters that affect the AI system's output. See the UK Department for Science, Innovation & Technology's **A Pro-Innovation Approach to AI Regulation**.

(E) The intended output of the AI system (e.g., numerical score). See the California Privacy Protection Agency's (CPPA) **Draft Automated Decisionmaking Technology Regulations**, at § 7017(b)(4)(D)(i)(2) (Dec. 2023).

(F) An explanation of how the AI is used in the decision-making process, including the role of human involvement. See **4 CCR 904-3-9.03**; the CPPA's **Draft Automated Decisionmaking Technology Regulations**, at § 7017(b)(4)(D)(i)(3) (Dec. 2023).

(G) Whether the AI system has been evaluated for accuracy, validity, reliability, fairness, or bias, and the outcome of any such evaluation. See **4 CCR 904-3-9.03**; the CPPA's **Draft Automated Decisionmaking Technology Regulations**, at § 7017(b)(4)(D)(i)(4) (Dec. 2023).

(H) The benefits and potential consequences of the decision made using AI. See **4 CCR 904-3-9.03**.

(I) Information about how the individual may exercise rights in connection with the AI, such as access to further information about the AI system and opting-out of or contesting AI decisions. See **4 CCR 904-3-9.03**; the CPPA's **Draft Automated Decisionmaking Technology Regulations**, at § 7017(b)(4)(B)&(C) (Dec. 2023).

(J) Organizations may consider providing a hyperlink to a risk assessment conducted for the AI system. See the CPPA's **Draft Automated Decisionmaking Technology Regulations**, at § 7017(b)(4)(D)(ii) (Dec. 2023).

(K) Contact methods if the public has any questions or feedback regarding the AI system. See the Singapore Personal Data Protection Commission's **Model Artificial Intelligence Governance Framework**.