Digital Policy Alert

Transparency in International AI Frameworks

Nils Deeg

November 2026

An initiative of the
St.Gallen Endowment
for Prosperity through Trade

Transparency in International AI Frameworks

Our text-level analysis of nine international AI frameworks shows that all nine cover transparency, across a total of 47 provisions with various degrees of focus and actionability.

- 23 provisions focus on transparency as the core concern while 24 provisions only touch upon transparency tangentially.
- 42 provisions provide actionable language for governments to pursue transparency. Only five provisions, in the G7 documents, the Joint Declaration, and the UN General Assembly resolution, address transparency in non-actionable language.
- The ASEAN Guide, G7 Code of Conduct, and G7 Principles have the most provisions on transparency (8 each). The ASEAN Guide also has the most actionable provisions (8), while the UN General Assembly resolution has the most core provisions (4).

We now outline the prevalence, focus, and actionability of relevant provisions in each international framework, before summarising their language.

OECD AI Principles

Two OECD AI Principles address transparency. Principle 1.3 (Transparency and Explainability) covers it as core concern and Principle 1.5 (Accountability) covers it tangentially. Both principles provide actionable measures for governments.

Principle 1.3 focuses on transparency and specifies four ways in which AI actors should provide information:

- general understanding of capabilities and limitations;
- awareness of interactions with AI systems, including in the workplace;
- plain-language information on the data, factors and logic behind an output; and
- information enabling adversely affected persons to challenge an output.

Principle 1.5 focuses on accountability but includes actionable language on transparency. Specifically, the Principle states that AI actors should ensure the traceability of datasets, processes and decisions across the lifecycle, which in turn enables the analysis of outputs and "responses to inquiry".

UN General Assembly resolution

Six provisions in the UN General Assembly resolution A/RES/78/265 address transparency. Four provisions are core and two are tangential. All except one are actionable. Out of the core provisions:

- 6(d) encourages raising public awareness of AI systems' core functions, capabilities, limitations and domains of appropriate civil use;
- 6(g) encourages the development of content authentication and provenance mechanisms;
- 6(k) encourages the provision of notice and explanation for systems that make or support decisions affecting end-users;
- 6(n) encourages information-sharing among entities across the AI lifecycle.

Out of the tangential provisions:

- 6(e) encourages the disclosure of risk monitoring and management mechanisms, privacy policies and impact assessments;
- Sub-paragraph 6(j) mentions transparency and reporting requirements only as a context in which privacy must be safeguarded, and is therefore non-actionable.

All measures are framed as encouragement to Member States and, where applicable, other stakeholders. Sub-paragraphs 6(d) and 6(g) closely mirror the wording of G7 Principles 3 and 7.

Council of Europe Framework Convention

Four articles in the Council of Europe Framework Convention are related to transparency. The concern is core to one of them and tangential to three of them. All are actionable. The Convention is the only instrument examined that contains "shall" obligations. These bind Parties rather than AI actors directly, and several are softened by qualifiers.

Article 8, on transparency and oversight, is the core provision requiring parties to adopt transparency measures. It leaves the substance of transparency requirements to Parties, requiring only that they be "tailored to the specific contexts and risks". Its one specified element is the mention of the identification of AI-generated content as a relevant activity in the AI lifecycle.

Three provisions refer to transparency tangentially:

- Article 14(2) requires documentation of information on systems that could significantly affect human rights, access for authorised bodies and communication to affected persons. The provided information must allow affected persons to contest the decision and, where relevant, the use of the system itself.
- Article 15(2) requires notification that a person is interacting with AI rather than a human. This is the Convention's weakest transparency obligation, since Parties need only "seek to ensure" it "as appropriate for the context".
- Articles 16(2)(f) and 16(3) prescribe the documentation of risks and adverse impacts, including on human rights.

G7 Hiroshima Process Guiding Principles

Eight Guiding Principles feature transparency. Transparency is a core concern for three and tangential to five others. Six Principles include an actionable element with regard to transparency, while two are non-actionable. The Principles are voluntary and addressed to organisations developing advanced AI systems.

Transparency is core to three principles:

- Principle 3, on the public reporting of system capabilities, suggests transparency reports for "all new significant releases", a clarity standard (enabling deployers and users to interpret output) and underlying documentation. While the G7 Principles do not specify contents for the report, the Code of Conduct does.

- Principle 4, on responsible information sharing and incident reporting, suggests sharing evaluation reports, safety and security risks, dangerous capabilities and attempts to circumvent safeguards.
- Principle 7 suggests the development of content authentication and provenance mechanisms.

Transparency is tangentially related to five further principles:

- Principle 2, on vulnerability identification and mitigation, encourages the documentation of reported incidents.
- Principle 5, on governance and risk management policies, encourages the disclosure of such policies, including privacy policies.
- Principle 10, on the development of technical standards, explicitly mentions the need to develop standards for watermarking.
- Principle 1, on appropriate measures to address risks, is non-actionable, simply mentioning that developers should aim to enable traceability of datasets, processes, and decisions.
- Principle 11, on data input and protection measures, is also non-actionable, stating that appropriate transparency of training datasets should be supported.

G7 Hiroshima Process Code of Conduct

The Code of Conduct builds on the G7 Principles, adding additional detail in several cases. Transparency appears in the same eight provisions, as the core concern of three provisions and tangential to five. Seven of eight provisions contain an actionable element on transparency, with Provision 1 receiving an actionable element that was not present in the Principles. Apart from this, the added value of the Code of Conduct lies in elaboration:

- Provision 3 specifies what transparency reports should contain: evaluations of safety, security, societal and human rights risks; capabilities and significant limitations; effects on safety and society, including bias, discrimination and privacy; and red-teaming results. It adds instructions for use and technical documentation, all to be kept up to date.
- Provision 4 extends information sharing to the public and to public authorities. It adds lifecycle documentation for systems posing significant risks and mechanisms for shared standards. It also provides that reporting "should safeguard intellectual property rights".
- Provision 2 adds bounty systems, contests and prizes as incentives for responsible disclosure.
- Provision 10 extends standards work to content authentication, provenance and public reporting, and to interoperable standards for distinguishing AI-generated content.
- Provision 1, on risk identification and mitigation, adds an actionable transparency element by requiring regularly updated technical documentation to support risk testing.
- Provisions 5, 7, and 11 substantially reproduce the transparency elements of the Principles. Provision 7 adds a commitment to research on provenance.

ASEAN Guide on AI Governance and Ethics

Eight provisions across the ASEAN Guide focus on transparency. Three provisions have transparency as a core concern, and five provisions are tangentially related to transparency. All eight

provisions contain an actionable transparency element. The Guide is addressed to deployers and developers. Its distinguishing feature is practical specificity, with worked examples, named tools and fallback options where full transparency is not feasible.

Three provisions take transparency as their core concern:

- Principle 1 focuses on Transparency and Explainability. It sets out several actionable measures. These include disclosure of AI use, of its role in decisions, of the data used and of its purpose, and explanations ranging from simple text to heatmaps. Where models cannot be explained, it offers alternatives: outcome-based explanations, documentation of repeatability, audit trails, a "black box recorder" capturing input data, records of data provenance, and AI Model Cards. The black box recorder is the only explicit logging-type measure in the dataset.
- Section 4 centres on Stakeholder interaction and communication. It suggests chatbot disclosure requirements and disclaimers that user inputs may be used for training.
- The section on 'Modelling' guidance has a strong focus on explainability. Its actionable measures include:
    - documentation of training, model selection and automated machine-learning parameters;
    - expected-behaviour descriptions in product documentation;
    - contractual transparency obligations on developers and joint audits;
    - Use of explainability tools such as AI Verify, SHAP and LIME;
    - implicit and counterfactual explanations for lay users.

The section also flags fraud detection and proprietary trading algorithms as cases where disclosure may be inappropriate.

In the five tangential provisions, transparency serves other principles:

- With regard to security and safety, risks, limitations and safeguards should be made known to users.
- Internal documentation of errors can ensure accountability and integrity.
- Robustness and reliability can be improved through documentation of data sources and lineage.
- Determining the level of human involvement in AI-augmented decisions is facilitated by recording system and user behaviour for auditing purposes.
- Proper outcome analysis can be conducted if an organisation has documented the intended purpose of an AI system and established communication channels between technical teams and stakeholders.

ASEAN Expanded Guide on AI Governance and Ethics – Generative AI

The Expanded Guide on Generative AI builds on the 2024 Guide, adapting its underlying principles to the generative AI context. Four provisions feature transparency: it is core to three provisions and tangential to one further provision. All transparency elements are actionable.

Three provisions take transparency as their core concern:

- Section 3.4 contains the most detailed treatment of incident reporting in the dataset. It covers internal channels, external reporting to regulators, users, customers and value chain participants, vulnerability reporting, severity tiering, and interaction with existing breach notification regimes.
- Section 3.7 covers labelling, watermarking and cryptographic provenance.
- Section 1.3 restates all seven principles. Its transparency component is the Transparency and Explainability principle plus one element under privacy and data governance. The definitions reproduce the 2024 Guide. The new actionable elements are a mention of user awareness of interacting with generative AI and of how data is used, and transparency about training data sources and pre-training processing. Otherwise the section is largely diagnostic: it openly acknowledges that explanation and traceability have inherent limits for generative AI.

One provision is tangential:

- Section 3.3 combines safety and evaluation guidance with a transparency component, namely the disclosure of safety practices to downstream deployers and end users, likened to "food or ingredient labels", and proposed common disclosure elements for developers.

UNESCO Recommendation on the Ethics of AI

Three provisions in the UNESCO Recommendation feature transparency. All three are core and actionable.

Transparency is the core concern of the principle 'Transparency and explainability'. It is largely conceptual and offers the most developed rights-based rationale in the dataset, linking transparency to liability regimes, fair trial and effective remedy. It nevertheless contains concrete measures:

- informing people when decisions are AI-based and when products or services use AI;
- a right to request explanatory information and to access the reasons for decisions;
- submissions to a designated staff member able to review and correct decisions;
- meaningful explanations for decisions whose impact is not temporary, easily reversible or low risk.

Its most far-reaching suggested measure, not mentioned in any other framework, is the sharing of code or datasets where there are serious threats of adverse human rights impacts.

Transparency is also core to Policy Area 9, on communication and information. The Policy Area's transparency component refers to the transparency of online communication and information operators and prompt notification of the reasons for content removal or other treatment. It sits alongside material on access to information, literacy and media policy.

Policy Area 2 is a broad governance section. Its transparency component includes clear transparency and explainability requirements (paragraph 70), mandatory disclosure of stereotyping in AI outcomes and data (paragraph 66), and international standards setting "measurable, testable levels" of transparency (paragraph 64). It also covers auditability and traceability for redress (paragraph 55) and transparent self-assessment of AI systems by public authorities (paragraph 57).

Joint Declaration on AI, Freedom of Expression and Media Freedom

Four provisions of the Joint Declaration feature transparency: it appears in two as a core concern and in two as a tangential concern. Three provisions include actionable transparency elements. The Declaration addresses states, companies, as well as financial institutions and investors.

Transparency is a core concern of Principle 6, on transparency, accountability, and effective remedies. Its actionable component is its call for "meaningful information and access to data".

Further, transparency is also core to the section addressed to the private sector, which calls on AI companies to publish due diligence assessments ((a)); implement industry standards for the attribution and labelling of synthetic content ((e)); and provide meaningful information on system functioning, traceability, explainability, capabilities and limitations, and affected persons' rights ((f)), as well as, specifically, ensuring the explainability of AI in content moderation, curation and targeted advertising, paired with an appeal avenue ((h)).

Transparency is tangential to the section on state regulation of companies, which calls for transparency requirements aimed at bias correction, harm identification, liability and remedies ((a)(iii)); public availability of oversight and testing findings as "epistemic counterpower" ((b)(i)); and data protection laws enabling transparency in data collection and use ((c)).

Finally, transparency is also tangential to Principle 2, on a healthy, pluralistic, and diverse information environment. It only names transparency as a goal of regulation, without specifying any actionable measure.

Digital Policy Alert

Contact
Written by Nils Deeg
Please send questions and suggestions to nils.deeg@sgept.org

The independent Digital Policy Alert is a pillar of the Swiss-based
St. Gallen Endowment for Prosperity Through Trade

St.Gallen Endowment
for Prosperity through Trade