Private-sector white paper | 2026

Hiroshima AI Process: Private-Sector White Paper 2026

From international cooperation to infrastructure for trust across businesses, industries, and society

Going beyond AI adoption
to share the conditions for its use and the evidence behind them

International cooperation. Sharing principles and practices

Business and industry. Bringing principles into procurement and operations

Trust in society. Checking the evidence that matters

International milestones with analysis and recommendations for implementation

Jointly authored and edited by

AI Assurance, Inc.
Hiroshima AI Assurance Council

Published by AI Assurance, Inc.
Research cutoff date: October 26, 2026

This white paper presents independent private-sector research and recommendations. It is not an official publication of any government or international organization.

Executive summary

Why businesses need to understand the Hiroshima AI Process

The conclusion of this white paper

It is not enough to recognize the Hiroshima AI Process (HAIP) as an outcome of international meetings. The next opportunity for industry lies in connecting its outcomes to arrangements that enable businesses to explain the conditions for AI use and their partners to check the supporting evidence.

Decisions about AI increasingly need to address which operations it can support and under what conditions, as well as whether to use it at all. This white paper reviews the international milestones of the Hiroshima AI Process and recommends ways to connect its outcomes to business procurement and contracts, operational decisions, and cooperation across industries.

Documented developments	Implications for business and government
International cooperation continues	The outcomes of 2023 have been carried forward by subsequent G7 presidencies. This is an ongoing international initiative, rather than a declaration confined to a single year. [03, 07]
The process has moved into reporting on practice	The OECD has launched a voluntary Reporting Framework. Version 2.0 makes participation more accessible to organizations deploying AI and to small and medium-sized enterprises. [05, 06]
Procurement is increasingly asking for supporting evidence	Japanese government procurement materials set out examples of supporting information for checking whether suppliers meet requirements. [14, 15]
Assessment and verification are becoming an industry in their own right	The UK government has studied the AI assurance market and treats the quality and professional expertise of third-party assessment as policy issues. [21, 22]

The next step builds on reporting

Common reporting provides a starting point for understanding an organization's practices. This white paper proposes linking that reporting to the subject being assessed, the conditions, the evidence, the verification method, and the scope of validity, so that important decisions can be checked. Disclosure, audits, testing, and formal verification should be combined according to their purpose.

Scope: The table above summarizes published materials. The further commercial and economic effects discussed here are this white paper's analysis and recommendations; they do not demonstrate that participation in the Hiroshima AI Process increases sales.

About this white paper

Purpose and research approach

This private-sector white paper on policy and industry prospects is jointly authored and edited by AI Assurance, Inc. and the Hiroshima AI Assurance Council (HAAC). It is not an official white paper of any government, the G7, the OECD, or any other international organization, and does not imply their approval or endorsement.

Connecting international outcomes to trust that works in practice

The primary readers are business leaders, public administrators and policymakers, industry associations, and research and assessment organizations. Here, infrastructure for trust means the combination of institutions, practices, and technologies that enables stakeholders to share the conditions for AI use and their supporting evidence, check them to the extent necessary, and revisit decisions as conditions change. It does not refer to a single product or a new central institution.

In this white paper, AI governance refers to the policies, roles, and management arrangements that support the appropriate development and use of AI. AI assurance refers to activities that support claims about AI trustworthiness with evidence, including assessments, verification, and records. [20, 21]

Chapter	Coverage	Pages
Chapter 1	Milestones from 2023 to 2026 and scope	4–5
Chapter 2	Business competitiveness and connections to procurement and contracts	6–7
Chapter 3	Connections between frameworks and the AI assurance industry	8–9
Chapter 4	Evidence and verification methods that support explanations	10–11
Chapter 5	Applications in industry, local government, and cybersecurity	12–14
Chapter 6	Approaches to adoption, policy recommendations, and the outlook toward 2030	15–17
References	Primary sources, materials published by the parties concerned, and public reference implementations	18–21

Research and editorial standards

The research cutoff date is October 26, 2026. The research draws primarily on materials published by governments, international organizations, and standards bodies. Companies' own accounts, articles by named authors, and public proofs of concept (PoCs) are identified as materials published by those parties. Facts about institutions and frameworks, analysis drawn from those facts, and this white paper's own recommendations are kept distinct. Numbers in square brackets in the footnotes correspond to the references at the end of the white paper.

This white paper is not a comprehensive legal commentary, a determination of legality in any individual case, or an independent product certification. AI Assurance develops related technologies, and its associated PoCs in Chapter 5 are not presented as the results of independent audits. When applying the material, each party should check for revisions to the original sources and any applicable conditions of use.

Chapter 1 | International milestones

From declarations to a framework for explaining practice

The Hiroshima AI Process began following the G7 Hiroshima Summit in May 2023. The International Guiding Principles and Code of Conduct adopted in October that year were followed in December by endorsement of a Comprehensive Policy Framework, including International Guiding Principles for All AI Actors. The process encompasses both expectations for organizations developing advanced AI and principles relevant to those using AI.[1]

Date	Documented milestones
2023	Launched following the G7 Hiroshima Summit. A policy framework was established, including international guiding principles and a code of conduct. [03]
May 2024	The Friends Group was established to support cooperation among countries and regions that endorse the process, extending collaboration beyond the G7. [03]
February 2025	The OECD launched a voluntary Reporting Framework for the International Code of Conduct, enabling organizations to explain their practices in a common format. [05]
March 2026	A Friends Group meeting adopted Action Plan 2026. At that time, the group comprised 66 countries and regions. [04]
May 2026	The OECD published Version 2.0 of the Reporting Framework. The G7 ministerial declaration welcomed the revision and addressed both AI safety and adoption. [06, 07]

Safety and adoption are compatible goals

The 2026 G7 ministerial declaration addresses safety and business adoption of AI within the same policy agenda, and also refers to government procurement of AI. Wider adoption depends on clarifying what deploying organizations need to check, as well as raising awareness of the frameworks available.[2]

Analysis in this white paper

The continuity of the Hiroshima AI Process is evident in the way subsequent G7 presidencies have carried forward the outcomes of Japan's 2023 presidency. For Japanese businesses, it may also provide a way to engage in international discussions and contribute lessons from implementation.

[1] [01, 03] Ministry of Foreign Affairs of Japan. The October documents for organizations developing advanced AI are distinguished from the December principles for all AI actors.

[2] [07] G7 Digital and Technology Ministerial Declaration, sections on Promoting Secure AI and Boosting AI Adoption for Economic Growth.

Chapter 1 | Scope and participation

Relevance extends beyond major AI developers

Version 2.0 of the Reporting Framework distinguishes between the roles of model developers, application developers, and deployers, making it easier for each to answer the questions relevant to them. This broadens the opportunities for companies to explain their use of AI and their management arrangements, including those that do not develop their own foundation models.[3]

Initiative	Participants and roles	What it means
International Guiding Principles and Code of Conduct	International cooperation originating in the G7	Shared principles and directions for action. These are not laws that apply directly and uniformly to every company.
Friends Group	Countries and regions	Endorsement of the spirit of the Hiroshima AI Process and voluntary cooperation. [03, 04]
Partners Community	Companies, associations, and other organizations	A framework for private-sector dialogue and cooperation. Participation is distinct from submitting a report to the OECD. [08, 12]
OECD Reporting Framework	Eligible organizations	Reporting and publication of an organization's own practices. A listing is not a certification of safety. [08]

Publication supports transparency but does not constitute verification

The OECD Secretariat checks matters such as eligibility, completeness of responses, and availability of links, but does not assess or verify the substantive content of reports. A HAIP Brand listing likewise does not constitute certification of compliance with the Code of Conduct or an endorsement. When using reports for procurement or operational decisions, the evidence supporting the relevant claims therefore needs to be checked separately.[4]

The scope extends beyond technical safety to human rights, privacy, intellectual property, information integrity, security, and contributions to addressing societal challenges. Individual verification methods support these broader purposes; they do not replace value judgments or societal agreement.[5]

Companies considering participation should check the activities and workload involved in each initiative. Under the OECD Reporting Framework, organizations are expected to update or resubmit their reports at least every 12 months to remain listed as participants. Decisions on whether to participate should take account of the value gained and the work required for updates. [08]

[3] [06, 08] The OECD's announcement of Version 2.0 and its FAQ. Eligibility and submission requirements are set out in the official FAQ.

[4] [08] FAQ entries How will submissions be processed? and What should organisations know about the HAIP Brand listing?

[5] [02, 32] International Code of Conduct and International Guiding Principles.

Chapter 2 | Implications for business

Explaining trustworthiness becomes a condition of doing business

This research has not established a causal link between participation in the Hiroshima AI Process and increased orders or profits. Even so, explaining the conditions a partner needs to understand and providing the necessary supporting evidence have practical value in advancing AI adoption and business transactions.

Source of the statement	What the statement supports	What it does not establish
KDDI	Links participation in the Reporting Framework to external transparency and trust, and to taking stock of and reviewing internal practices. [11]	It does not provide measurements showing that participation increased profits.
Salesforce executive	Argues that a common reporting language can help organizations navigate differences between frameworks and contribute to customer trust. [10]	It is not an official OECD determination of economic benefits.
Autodesk	Discusses participation in the Partners Community alongside its own AI governance practices. [12]	Participation does not itself certify the safety of individual products.

Three possible pathways to economic benefits

The first is to organize the checks required by deploying organizations and reduce repeated exchanges of explanations. The second is to make existing assessment materials usable in other projects where the relevant conditions match. The third is to clarify the scope of permissible use by distinguishing the conditions under which AI proposals can and cannot be accepted. This white paper presents these as hypotheses about how benefits could arise.

For example, the assessment of the same system changes when its purpose, data, or allocation of responsibilities changes. What is needed is an explanation of what has already been checked and what must be checked again after a change, rather than a claim that one assessment establishes safety for every use.

Questions for management

Assess the value of AI governance by whether the evidence needed for an adoption decision is available, whether checks on conditions with business partners have progressed, and whether decisions can be revisited after changes or failures. The number of documents produced is not a useful measure on its own.

Corporate publications [10–12] are used as statements of the parties' own views. The OECD's analysis of the initial 20 organizations also examines voluntary reports; it is neither a representative sample of all companies nor an independent safety assessment. [09]

Chapter 2 | Procurement and contracts

Making explanations verifiable during supplier selection

Japan's Digital Agency revised its guidelines in June 2026 to address the procurement and safe use of generative AI together. The procurement checklist in Annex 3 identifies matters to check and gives examples of supporting information. This official practical guidance translates trustworthiness into specific checks for supplier selection.[6]

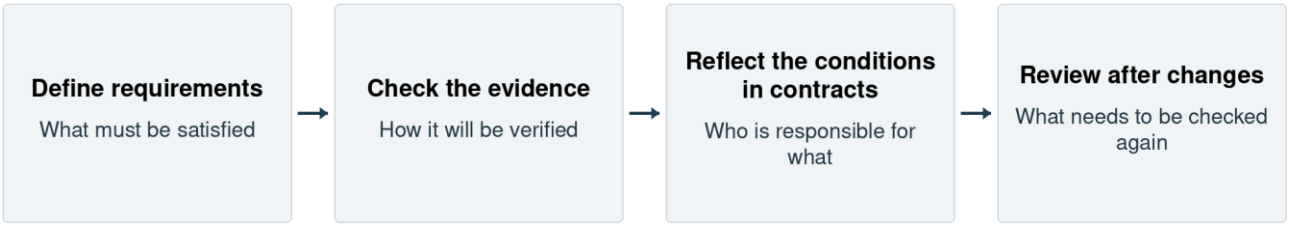


Figure 1. This white paper's proposed model for connecting evidence to procurement and contracts. It does not depict the procedures of an existing framework.

Questions for procurement and contracts	Examples of checks proposed in this white paper
Purpose and subject	Distinguish between drafting guidance and deciding to execute actions such as issuing benefits, placing orders, or transferring funds.
Information from the provider	Ask the provider to identify the assessed version, relevant data, assessment conditions, and limitations.
Responsibilities and scope of review	Clarify which decisions are to be checked by the provider, the deployer, and operational personnel.
Updates, failures, and termination	Agree how reassessment, notification, and suspension will be handled when specifications change, incidents occur, or use of data ends.

The contract checklist in Annex 4 and the Ministry of Economy, Trade and Industry's AI contract checklist also provide references for organizing conditions between parties. However, the examples in those materials cannot simply be treated as obligations in every contract. They must be adapted to the applicable rules, operational risks, and roles of the contracting parties.[7]

Connection to the Hiroshima AI Process

Use international reporting as a starting point for understanding business partners, then check the evidence relevant to your own use case. This is distinct from making participation in the Hiroshima AI Process a uniform condition of doing business.

[6] [14, 15] Version 2.0 of the guidelines and the structure and instructions for Annex 3. The scope of application to government is distinguished from voluntary reference by private-sector organizations.

[7] [16, 17] The contract checklist and AI contract checklist. The specific examples in the table are this white paper's proposals informed by these materials.

Chapter 3 | Connections between frameworks

Respecting the distinct roles of different frameworks

International principles, laws, management standards, and assessment tools serve different roles. Interoperability means reusing information that they require in common while continuing to address the requirements specific to each framework. It does not mean merging them into a single system.

Framework	Primary role	Limits on substituting for other assessments
Hiroshima AI Process	International principles, actions, and cooperation for safe and trustworthy AI. [01–03]	Does not replace obligations under national laws or conformity assessments for individual AI systems.
OECD Reporting Framework	Reporting and publication of organizational practices in a common format. [05, 08]	A listing does not constitute substantive verification of content or safety certification.
Japan's AI Guidelines for Business	Domestic guidance supporting the practices of developers, providers, and users. [13]	Following the guidelines alone does not establish compliance with every applicable law.
EU AI Act	Legal obligations according to scope, role, and risk category. [18]	Participation in voluntary reporting or certification under another framework does not by itself establish compliance.
ISO/IEC 42001	A standard for organizational AI management systems. [19]	Does not prove the correctness of individual outputs or decisions.
NIST AI RMF	A voluntary framework for identifying, measuring, and managing risk. [20]	Does not confer legal immunity or automatically authorize an individual use case.

Verification needs to go beyond mapping documents

Even when two documents both call for transparency, they may differ in their intended audience, level of detail, method of disclosure, and timing of updates. Beyond matching headings, it is necessary to check which materials support which requirements and to what extent.

This comparison outlines the roles of the frameworks; it is not a full interpretation of the laws or standards. The EU AI Act has different application dates for different provisions and is subject to amendments and transitional arrangements. Consult the latest official sources when applying it to a particular case. The discussion of ISO draws on its public overview.

Chapter 3 | Interoperability and the market

Reusing existing evidence where it remains applicable

Connections between frameworks are already being developed. The AI Verify Foundation has published an assessment framework comprising principles, outcomes, processes, and evidence, together with mappings to the Hiroshima Process International Code of Conduct, NIST AI RMF, ISO/IEC 42001, and other frameworks.[8]

An article by Ségerie and Salter published on OECD.AI also discusses overlap between the information needed for HAIP reporting and for compliance with the EU AI Act. This is the authors' analysis; it does not establish equivalence between the frameworks or certify compliance with EU law.[9]

Checks before reuse	How to address differences
Match of subject	Avoid transferring results to a different model, use case, or user group.
Match of conditions and period	Reassess the relevant aspects if the settings or assumptions used in the assessment have changed.
Match of assessment scope	Do not treat requirements that were not examined under one framework as already verified.
Permissible scope of sharing	Do not disclose personal information or trade secrets to parties that do not need them for verification.

The assessment and verification industry is also a subject of public policy

A 2024 study by the UK's Department for Science, Innovation and Technology (DSIT) estimated that 524 firms were involved in AI assurance, generating approximately £1.01 billion in gross value added (GVA). Its 2025 roadmap addresses the quality of third-party assurance, professional expertise, and market development. Activities that assess trustworthiness are themselves being recognized as an industry.[10]

Analysis in this white paper

Japan could develop an assessment and verification industry that understands use cases, selects appropriate evidence, and reduces duplicate checks, rather than a market for producing more paperwork. The UK estimates support the existence of such an industry; they do not measure a market created by the Hiroshima AI Process.

[8] [23] AI Verify Testing Framework. The mapping to the Hiroshima Process International Code of Conduct is dated May 29, 2025.

[9] [24] Analysis by the named authors. The article expressly states that it does not represent the official views of the OECD, GPAI, or their member countries.

[10] [21, 22] UK government research and policy documents. GVA is not revenue, and the estimate relates to the UK market.

Chapter 4 | Explanations and evidence

Infrastructure for trust that makes clear what has been verified

An organization's statement that it has implemented safety measures is broader than an operational decision that a particular action may proceed under specified conditions. Moving from the former to the latter requires checking whether the reported measures apply to the system, task, or action at hand.

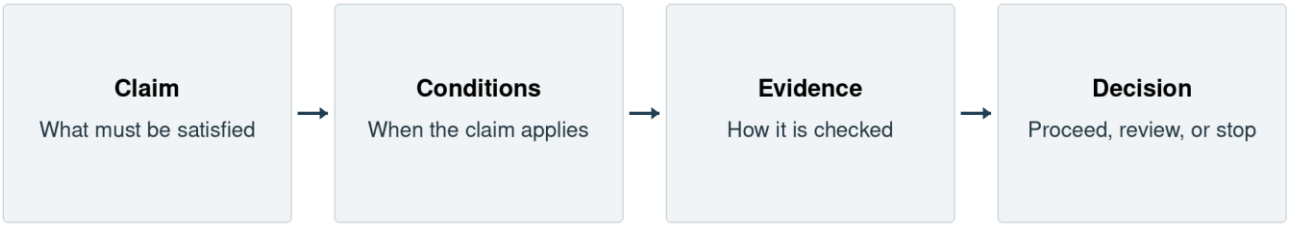


Figure 2. The connection proposed in this white paper. It allows the relevant parties to check what they need to know, within the necessary scope; it does not require every individual decision to be made public.

Information needed to communicate the basis for a decision	Details to record
Subject and intended use	The system, version, business process, users, and action to be performed.
Conditions and criteria	The assumptions underlying the decision, applicable policies and contracts, time limits, and required approvals.
Evidence and methods	Documents and test results used for the assessment, their sources, who performed the checks, and how.
Results and unresolved matters	Conditions satisfied, inconsistencies, missing information, and areas not evaluated.
Scope of validity and reassessment	Where the results may be used, changes that would invalidate them, and who is responsible for the next step.

These information fields are a proposal by this white paper; they have not been adopted as a new international standard. Existing operational records and assessment materials can be used to organize only the information needed for the intended use. Handling the evidence that supports trust does not require creating another large reporting exercise.

Transparency and independent verification complement each other

The Hiroshima Process International Code of Conduct has addressed evaluation, including independent external testing, from the outset. The next task is therefore to connect reported practices with evidence relevant to specific uses, rather than to introduce independent verification as a new concept.[11]

[11] [02] Action 1 addresses internal and independent external testing, traceability, and technical documentation. The table and figure are this white paper's own synthesis informed by these provisions.

Chapter 4 | Choosing verification methods

Combining testing and auditing with formal verification

Being able to check something independently does not mean proving everything mathematically. Human evaluation, operational audits, statistical testing, and formal verification each address different questions. NIST also addresses the importance of measurement, testing, evaluation during operation, and independent perspectives.[12]

Method	Questions it is well suited to answer	Further checks needed
Testing and red teaming	What errors occur with the inputs, attacks, and operational scenarios tested?	Untested scenarios, how representative the data are, and changes in the environment.
Audits and document reviews	Are responsibilities and procedures established, and are they consistent with actual operational records?	Limits of scope and sampling, and the independence of the reviewer.
Formal verification	Does a defined property logically hold under explicitly stated assumptions?	Whether the assumptions are valid, the real-world evidence is genuine and accurate, and the implementation corresponds to the model.

Example of preventing execution without valid approval

If a mechanism is defined to permit execution only when all required approvals are valid, it is possible to verify, within that model, the property that operations without approval are never authorized. Separate evidence is needed to establish whether approvers are who they claim to be, whether the governing rules are appropriate, and whether the actual system behaves as the model specifies. This is a general example explaining the role of formal verification, not a claim about the performance of a particular product.

Distinguishing what can be checked jointly from what requires agreement

Even when countries or organizations use different safety criteria, they can check an agreed proposition using a common method if they can share the proposition and the model and procedures used to verify it. Cooperation can begin with what is verifiable without waiting for complete alignment of criteria. This does not, however, automatically resolve value judgments about what counts as safe or whose rights should be protected and how.

The position of this white paper

Formal verification is neither a mandatory requirement nor a universal solution. What matters is choosing a method suited to the claim and enabling third parties to distinguish what is known from what remains unknown.

[12] [20] The MEASURE function of the AI RMF Core. For technical background on formal specification and analysis, see NASA's guide [26].

Chapter 5 | Applications across industry and society

Conditions for accepting AI proposals apply across sectors

The following are applications proposed in this white paper. They presuppose the rules, professional judgment, and allocation of responsibilities in each field; they are not presented as outcomes of technologies already deployed. Their common feature is to assess the quality of an AI proposal separately from the conditions for accepting and executing it in operations.

Sector and scenario	Checks before acceptance or execution	Potential value
Manufacturing **Process changes and shipment release decisions**	Whether quality criteria, equipment conditions, inspection results, and change approvals apply to the product and process concerned.	Make it easier to communicate what has already been checked between business partners and production processes.
Logistics **Adding participants to joint delivery**	Assess existing shippers' conditions, approvals, information-sharing boundaries, and profitability separately, alongside capacity.	Explore opportunities for additional participation and revenue while preserving existing conditions.
Local government **Resident guidance and document preparation**	The relevant program or rules, their period of applicability, the circumstances of the inquiry, personal information that may be used, and the scope of human review.	Support appropriate guidance and identify missing information or inconsistencies before an answer is accepted.
Finance and insurance **Support for assessment and contract decisions**	Supporting documents, explanations, authority, conflicts of interest, and whether assessment results apply to the intended use.	Make the reasons for decisions easier to trace. This does not guarantee lower insurance premiums or acceptance for coverage.
Employment and education **Support for assessment and selection**	The purpose of evaluation, conditions for comparison, data suitability, fairness considerations, and avenues for appeal.	Ensure that those being assessed have opportunities to receive explanations and seek reconsideration.

A common record format still requires criteria specific to each field

Manufacturing quality criteria cannot simply be transferred to fairness in recruitment. The aim is to develop a common way of communicating the subject, conditions, supporting evidence, and areas not yet checked. Decision criteria must be determined through each field's laws, contracts, professional knowledge, and dialogue with the parties concerned.

The applications in this chapter are proposals by the joint authors; they do not indicate regulatory approval or third-party certification in any field. Cybersecurity and data management are addressed in this chapter's section on AI agents and safeguards for systems and data as conditions that apply across all sectors.

Chapter 5 | Published technical examples

Limited scenarios can demonstrate how the connections work

AI Assurance, one of the joint authors, provides public proofs of concept (PoCs) involving local government and logistics. The summary below is based on their public descriptions as of October 26, 2026. No production-environment audit or independent rerun of the programs was performed for this white paper.

Public PoC	Design described in the public materials	Limits of the demonstration
Local government AI Answer acceptance [28]	Checks a proposed answer against the relevant program or rules, document versions, and circumstances of the inquiry, then distinguishes acceptance, hold, and rejection.	Fictional rules and synthetic data. Accepting a proposed answer is distinct from making an actual administrative decision.
Local government AI Privacy [29]	Separately checks the purpose, authority, and data fields before transmission and the conditions for using the returned output.	Simulated transmission and synthetic data. This does not prove the safety of handling actual residents' information or making real external transmissions.
Local government AI Managing the use of derived information [30]	Carries conditions forward to generated summaries and other derived information, then checks subsequent use and cessation of use within a defined scope.	Cessation of use along managed paths. This is distinct from physical erasure or proof that no copies exist outside the managed scope.
Joint delivery Expanding participation [31]	Assesses additional shipper participation, handoffs, and incremental profitability separately.	Synthetic conditions on a set route. This does not demonstrate actual operations, realized revenue, or measured reductions in labor.

The three local government examples are public reference implementations in which Crowdsien Inc. is responsible for business and deployment requirements, and AI Assurance for technical design, implementation, and verification. The joint delivery PoC is published by AI Assurance.[13]

What the public examples show and what remains to be assessed

These examples present designs that handle operational conditions and the reasons for decisions, going beyond statements of principle. Real-world data quality, integration with existing systems, exception handling, costs, and acceptance by professionals and users still require separate evaluation. A successful public PoC should not be treated as proof of commercial viability.

The role of publication

Publication communicates to external parties what can be checked and what has not been checked; it does not require disclosing confidential information. Public reference implementations matter because they provide a starting point for independent examination.

[13] [28–31] The README files of the respective repositories. The descriptions and scope of checks are based on the public documentation; they are not independent certification of performance.

Chapter 5 | AI agents and safeguards for systems and data

When AI takes action the conditions at execution must also be checked

The Hiroshima Process International Code of Conduct addresses cybersecurity and the provenance of generated content. In July 2026, Japan's AI Safety Institute (AISI) also added "Observation and Control" to its evaluation perspectives in response to risks involving AI agents and related technologies. When AI acts on an external environment, output quality alone is insufficient to determine whether an operation should proceed.[14]

Scenario	Operational conditions proposed in this white paper	Records to maintain
AI transmits information externally	Check the purpose of use, destination, data fields to be sent, and required authority at the time of transmission.	The data concerned, authorization conditions, whether transmission succeeded, and unresolved matters.
A critical operation is executed	Check the intended target, approvals, and permitted scope of action. Do not proceed automatically if conditions are unclear.	What was approved, time limits, execution results, and the person or entity to whom the matter was referred for review.
Recovery from a failure or attack	Check the current state before resuming, rather than relying only on old approvals or past assessments that found no problems.	The scope of recovery, remaining risks, conditions for resuming, and who performed the checks.
Generated information is reused or its use ends	Check the conditions governing use of the source information, distinguishing derived information, copies, and material outside the managed scope.	Provenance, applicable conditions, the scope of cessation of use, and exceptions.

Connecting security controls with operational decisions

This white paper identifies one direction for cyber assurance as examining the grounds for proceeding with critical operations, including when an attack or failure is suspected. It does not replace authentication, access control, monitoring, vulnerability management, or recovery plans. It checks whether the results of those measures provide a basis for authorizing the particular operation.

Adding human review alone may not be sufficient. The design must ensure that reviewers have the information and authority they need and that alternative procedures are available in an emergency. Rather than stopping indefinitely regardless of circumstances, the permitted scope should be made clear so that operations can proceed, be put on hold, or stop as appropriate.

The table presents design proposals, not demonstrated resistance to all cyberattacks. Protection of records against tampering, identity and authorization checks, and management of timestamps and versions also require separate evaluation.

[14] [02] Actions 6 and 7. [25] The explanation of revisions to the Guide to Evaluation Perspectives on AI Safety, version 1.20.

Chapter 6 | Where businesses can begin

Start with one important decision

There is no need to evaluate every AI system across an organization to the same depth from the outset. This white paper proposes choosing one decision where failure would have significant consequences, multiple organizations are involved, or deployment has stalled because checks are incomplete, and determining what can be established from existing materials.

Initial step	Minimum matters to establish
1 Identify the decision	For example, referring a proposed answer for staff review, accepting an additional shipper, or permitting an external transmission.
2 Establish acceptance conditions	Applicable policies and contracts, required approvals, data to be checked, and how to handle missing information.
3 Map the existing evidence	Which operational records, test results, contracts, and other materials support each condition, and where evidence is missing.
4 Check exceptions and changes	Whether the process can appropriately hold and reassess cases involving unmet conditions, inconsistencies, expiry, or a change in the subject being assessed.
5 Assess the value of deployment	Compare expected benefits with the costs of implementation, maintenance, and reassessment for the same subject and under the same conditions.

Assess both the benefits and the burden of evaluation

Benefits to compare include reductions in existing checking work, a wider range of usable applications, and less need to repeat explanations. Costs should include initial setup, data preparation, system integration, staff review, and reassessment after updates. Passing a check and achieving economic viability are separate judgments.

A framework for assessing deployment value

Benefits − Initial implementation costs − Ongoing operating costs − Reassessment costs after changes

This calculation alone must not determine whether to proceed when human rights or safety requirements cannot be expressed in monetary terms. Unmeasured benefits must not be treated as established profits.

Choose indicators according to the purpose. Examples include the number of exchanges needed to complete checks, staff workload, operations that proceeded before the necessary checks were completed, and operations that were incorrectly stopped. The aim is not to add uniform periodic measurement or reporting requirements, but to assess these indicators on a comparable basis when needed to inform a deployment decision.

Prioritize existing operational records when checking indicators and avoid unnecessary collection and reporting. Treat this separately from the submission and update requirements that apply when participating in international reporting frameworks.

Chapter 6 | Joint recommendations for policy and industry

Connecting a shared international language with usable evidence

The following are recommendations by the joint authors, not plans adopted by governments or international organizations. They prioritize connections between existing reporting, procurement, evaluation, and operational records over the creation of a new uniform certification scheme.

Recommendation	Main parties to consider it	Practical action
1 Develop a reference format linking claims and evidence	Reporting framework operators and participating companies and organizations	Develop an optional reference format that can be appended to existing reports and materials, specifying the subject, conditions, evaluation scope, and unresolved matters. Do not require blanket disclosure of confidential information.
2 Reduce duplicated checks	Governments, standards bodies, assessment organizations, and industry associations	Prepare mappings between frameworks, distinguishing reusable evidence from matters requiring additional checks. Keep this separate from claims of legal equivalence.
3 Connect demonstrations of specific operations with procurement and contracts	Organizations deploying AI, local governments, and organizations supporting demonstrations	Define the operation and acceptance conditions first, then examine cases involving unmet or changed conditions as well as normal operation. Communicate the scope of the demonstration to procuring organizations.
4 Improve the quality and range of third-party verification	Assessment organizations, universities, and professional bodies	Explain assessors' competence, conflicts of interest, and verification scope so that results from different methods can be understood. Avoid dependence on a single product.
5 Enable participation by small and medium-sized enterprises and users	Governments, industry associations, and regional support organizations	Prioritize simple procedures that use existing materials, sector-specific examples, and support with language and staffing. Select checks proportionate to risk.

Why communicate these proposals from Hiroshima

Connecting the outcomes of the Hiroshima AI Process with concrete challenges in manufacturing, logistics, and local government can generate lessons that inform both international cooperation and practical implementation. HAAC presents this possibility through its role in private-sector cooperation, and AI Assurance through its role as a technology developer.[15]

The parties responsible for implementation, funding, and timing require separate agreement. Listing these recommendations does not guarantee the participation of any specific institution, a budget allocation, formal adoption, or establishment of an institutional framework.

[15] [27] An announcement by a participating organization describing HAAC's structure and activities. It does not indicate delegated authority from a government or international organization, or official certification.

Chapter 6 | Outlook toward 2030

Making trust a foundation for using AI

The significance of the Hiroshima AI Process lies in establishing a shared understanding of risks and expanding opportunities for organizations to explain their practices. This white paper envisages a society that builds on those outcomes so that businesses and public authorities can share the conditions under which AI may be used for specific purposes.

Stage of development	Desired state	Conditions for progress
Share explanations	Stakeholders can understand practices, responsibilities, intended uses, and limitations.	A common language and clear reporting tailored to each role.
Examine the evidence	Third parties can check the subject and evaluation scope of important claims.	Appropriate evidence, evaluation methods, information protection, and independence.
Use the evidence in decisions	Verified evidence informs procurement, contracts, and decisions on whether to proceed with an operation, with reassessment when conditions change.	Correspondence with actual systems, allocation of responsibilities, exception handling, and reasonable costs.

Addressing the obstacles to implementation

Challenges remain: reports that satisfy formal requirements without meaningful substance, rising evaluation costs, dependence on particular vendors, information leaks through evidence sharing, and the exclusion of affected parties from the development of evaluation criteria. Responses should combine checks proportionate to risk, verification that protects confidentiality, disclosure of assessors' conflicts of interest, and opportunities for users to appeal or seek reconsideration.

The year 2030 is a horizon for considering direction in this white paper, not a guaranteed deadline for delivery. Even without complete alignment of international rules, it is possible to expand the conditions on which parties can agree and the scope of what can be checked. The lessons from that work can, in turn, inform international discussions.

Conclusion

The reason to understand the Hiroshima AI Process is not that participation itself creates a competitive advantage.

It provides a shared international foundation for considering what businesses and society need to explain and verify, and who should decide, so that AI can be used with confidence.

Connecting its outcomes with conditions for specific uses and verifiable evidence offers a path toward new infrastructure for trust that supports AI deployment, transactions, and cooperation across industries.

This white paper does not presuppose adoption of any particular technology or product. Publication does not grant joint ownership of any existing patents, know-how, or other rights, transfer those rights, or grant a license to use them. Any necessary rights arrangements must be made separately at the time of deployment.

References

References 01–08

Reference date: October 26, 2026. Numbers correspond to the citations in square brackets in the text. The views of companies and individual authors are distinguished from the official views of governments and international organizations. Some document titles have been shortened for ease of identification.

[01] Ministry of Foreign Affairs of Japan | Primary government source

G7 Leaders' Statement on the Hiroshima AI Process (October 30, 2023)

Original source: www.mofa.go.jp

Endorsement and publication of the International Guiding Principles and International Code of Conduct.

[02] G7, published by the Ministry of Foreign Affairs of Japan | Primary government source

Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems (provisional Japanese translation, October 30, 2023)

Original source: www.mofa.go.jp

Actions 1 (testing), 3 (transparency), 6 (security), 7 (provenance) and 10 (standards).

[03] Ministry of Foreign Affairs of Japan | Primary government source

AI Diplomacy (in Japanese; updated July 1, 2026)

Original source: www.mofa.go.jp

The December 2023 Comprehensive Policy Framework, the guiding principles for all AI actors, and international developments.

[04] Ministry of Foreign Affairs of Japan | Primary government source

Second In-Person Meeting of the Hiroshima AI Process Friends Group (Results) (March 16, 2026)

Original source: www.mofa.go.jp

The number of participating countries and regions is as of the meeting date. Adoption of Action Plan 2026 and voluntary initiatives.

[05] OECD | Primary international organization source

OECD launches global framework to monitor application of G7 Hiroshima AI Code of Conduct (February 7, 2025)

Original source: www.oecd.org

Launch of a common voluntary reporting framework.

[06] OECD | Primary international organization source

OECD launches streamlined Hiroshima AI Process Reporting Framework to help small- and medium-sized enterprises participate (May 28, 2026)

Original source: www.oecd.org

Revision to Version 2.0, expanding participation by organizations that develop, provide or deploy AI, including small and medium-sized enterprises.

[07] G7, published by the UK Government | Primary government source

G7 Digital and Technology Ministerial Declaration (adopted May 29, 2026; published June 1, 2026)

Original source: www.gov.uk

Continuation of the Hiroshima AI Process, the Reporting Framework, AI procurement and business adoption of AI.

[08] OECD.AI | Primary international organization source

Hiroshima AI Reporting Framework: Frequently Asked Questions (accessed October 26, 2026)

Original source: oecd.ai

Eligibility, questions tailored to organizational roles, the scope of submission screening, the distinction from certification, updates, and the differences between participation mechanisms.

References

References 09–16

Numbers correspond to citations in the text and footnotes. The URLs link to the original sources. Note the subject, version and scope of each source.

[09] OECD | International organization research

How are AI developers managing risks? (September 25, 2025)

Original source: doi.org

Analysis of self-reports from the first 20 organizations. OECD Artificial Intelligence Papers, No. 45.

[10] Paula Goldman, published on OECD.AI | Commentary by a corporate executive

HAIP is transforming transparency from a compliance burden to a competitive advantage (July 16, 2026)

Original source: oecd.ai

Views from a Salesforce perspective. These are not an official OECD position or measured evidence of economic effects.

[11] KDDI | Company account

International Discussions on Ensuring Trustworthy AI: The Hiroshima AI Process International Code of Conduct and Reporting Framework (in Japanese; September 4, 2026)

Original source: tech-note.kddi.com

The company's account of external trust and transparency, and of taking stock of and reviewing its internal initiatives.

[12] Autodesk | Company account

Autodesk Advances Its Trusted AI Initiatives Through the Hiroshima AI Process (HAIP) (in Japanese; July 22, 2026)

Original source: adsknews.autodesk.com

An account of participation in the Partners' Community and the company's AI governance. This is distinct from submission to the Reporting Framework.

[13] Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry of Japan | Primary government source

AI Guidelines for Business, Version 1.2 (in Japanese; March 31, 2026)

Original source: www.meti.go.jp

Guidelines for businesses, including AI developers, providers and users.

[14] Digital Agency of Japan | Primary government source

Guidelines for the Procurement and Use of Generative AI for the Advancement and Innovation of Public Administration, Version 2.0 (in Japanese; June 12, 2026)

Original source: www.digital.go.jp

Official publication page for the guidelines, annexes and original files.

[15] Digital Agency of Japan | Government practical guidance

Annex 3 to the Guidelines: Procurement Checklist (in Japanese; June 12, 2026)

Original source: www.digital.go.jp

The introductory instructions and checklist structure are referenced. The document sets out requirements, verification methods and examples of supporting information.

[16] Digital Agency of Japan | Government practical guidance

Annex 4 to the Guidelines: Contract Checklist (in Japanese; June 12, 2026)

Original source: www.digital.go.jp

Contractual issues to consider and sample clauses. Review and adjustments are required for each case.

References

References 17–24

Numbers correspond to citations in the text and footnotes. The URLs link to the original sources. Note the subject, version and scope of each source.

[17] Ministry of Economy, Trade and Industry of Japan | Primary government source

Publication of a Contract Checklist for the Use and Development of AI (in Japanese; February 18, 2025)

Original source: www.meti.go.jp

Published guidance on issues to consider in contracts for AI use and development.

[18] European Commission | Official government guidance

Navigating the AI Act (updated August 7, 2026; accessed October 26, 2026)

Original source: digital-strategy.ec.europa.eu

Official FAQ on the scope and phased application of the EU AI Act. The legislative text must be consulted to determine its application to an individual case.

[19] ISO | Standards body

ISO/IEC 42001:2023 — Artificial intelligence management system (2023 edition; accessed October 26, 2026)

Original source: www.iso.org

The public overview is referenced. The standard covers organizational AI management systems.

[20] NIST | Government technical guidance

AI RMF Core (AI RMF 1.0; accessed October 26, 2026)

Original source: airc.nist.gov

GOVERN, MAP, MEASURE and MANAGE. Measurement, evaluation during operation and independent review.

[21] UK Department for Science, Innovation and Technology (DSIT) | Government market research

Assuring a Responsible Future for AI (November 6, 2024)

Original source: www.gov.uk

Estimates for the UK market: 524 companies and approximately £1.01 billion in gross value added (GVA).

[22] UK DSIT | Government policy document

Trusted third-party AI assurance roadmap (September 3, 2025)

Original source: www.gov.uk

Policy on the quality, expertise and market development of third-party AI assurance.

[23] AI Verify Foundation | Primary source from the implementing organization

AI Verify Testing Framework (accessed October 26, 2026)

Original source: aiverifyfoundation.sg

Connects principles, outcomes, processes and evidence. Publishes mappings to the Hiroshima Process International Code of Conduct and other frameworks.

[24] Charbel-Raphael Ségerie and Jonathan Salter, published on OECD.AI | Analysis by the authors

Bridging frameworks: How HAIP 2.0 supports interoperability with the EU AI Act (September 30, 2026)

Original source: oecd.ai

Discusses overlaps between frameworks and the reuse of evidence. This is not an official OECD position or a determination of compliance.

References

References 25–32

Numbers correspond to citations in the text and footnotes. The URLs link to the original sources. Note the subject, version and scope of each source.

[25] Japan AI Safety Institute (AISI) | Public-sector technical guidance

Release of the Guide to Evaluation Perspectives on AI Safety, Version 1.20 (in Japanese; July 7, 2026)

Original source: aisi.go.jp

Explains the revision adding “observation and control” as an evaluation perspective in response to AI agents and related developments.

[26] NASA | Government technical guidance

Formal Methods Specification and Analysis Guidebook for the Verification of Software and Computer Systems: A Practitioner’s Companion, Volume 2 (NASA-GB-001-97/VOL2)

Original source: ntrs.nasa.gov

Technical background on formal specification and analysis. This does not provide an unconditional guarantee of compliance with all laws or of real-world safety.

[27] ONZALINX | First-party announcement

Establishment of the Hiroshima AI Assurance Council: Bringing Technologies Supporting the Hiroshima AI Process from Hiroshima to the World (in Japanese; October 8, 2026)

Original source: prtimes.jp

Describes HAAC’s organizational structure and participation. This is distinct from official recognition or certification by a government or international organization.

[28] Crowdsien and AI Assurance | Public reference implementation

Civic AI Answer Adoption Assurance: README (accessed October 26, 2026)

Original source: github.com

A PoC that checks the conditions and evidence for proposed answers using synthetic data. This white paper refers to the publicly available description.

[29] Crowdsien and AI Assurance | Public reference implementation

Civic AI Privacy Assurance: README (accessed October 26, 2026)

Original source: github.com

Verification using synthetic data and simulated transmission. This does not demonstrate safety for actual resident data or actual external transmission.

[30] Crowdsien and AI Assurance | Public reference implementation

Civic AI Derived Information Lifecycle Assurance: README (accessed October 26, 2026)

Original source: github.com

Propagation of conditions and termination of use within the managed scope. Physical erasure and the absence of copies outside that scope are not covered.

[31] AI Assurance | Public reference implementation

Joint Delivery Participation Expansion Assurance: README (accessed October 26, 2026)

Original source: github.com

Separate assessments of participation, handoff and profitability using predefined routes and synthetic conditions. These do not represent actual delivery operations or realized profits.

[32] G7, published by the Ministry of Foreign Affairs of Japan | Primary government source

Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems (provisional Japanese translation, October 30, 2023)

Original source: www.mofa.go.jp

Principles covering safety, human rights, privacy and security, among other issues. Note the difference in scope from the guiding principles for all AI actors.

Published by AI Assurance, Inc.
Jointly authored and edited by AI Assurance, Inc. and the Hiroshima AI Assurance Council