Standards Body · Foundation paper, public edition · Released July 17, 2026
Canonical record: https://standardsbody.ai/library/foundation-paper/global-interoperability/

FOUNDATION_08_GLOBAL_INTEROPERABILITY.md

Foundation 8: Global Interoperability

Series: Foundations for Frontier AI Evaluation Infrastructure
Version: 1.0
Status: Canonical working white paper
Project: Standards Body
Primary domain: standardsbody.ai
Core line: Foundations for Frontier AI
Research basis reviewed through: July 16, 2026
Document owner: Standards Body
Review cycle: Annual review, with event-triggered revision after material changes in international standards, evaluation networks, treaty systems, accreditation arrangements, model capabilities, or cross-border deployment practice

---

Document Purpose

This paper defines the Standards Body position on global interoperability for frontier artificial intelligence evaluation, assurance, standards, and governance.

It is intended to serve as:

- A first-principles explanation of why cross-border AI assurance requires interoperability rather than forced uniformity
- A framework for making evaluation evidence understandable, portable, comparable, and reviewable across institutions and jurisdictions
- A design guide for shared terminology, metadata, protocol mapping, evaluator recognition, incident exchange, secure information sharing, and standards crosswalks
- A bridge between national authority, international coordination, technical standards, conformity assessment, and plural governance systems
- A reference for future Standards Body partnerships, registries, working groups, standards-development processes, and institutional agreements
- A durable source document from which shorter articles, technical specifications, data schemas, and international proposals can be developed

This paper does not propose a single global regulator.

It does not claim that all jurisdictions should adopt identical laws, thresholds, risk tolerances, or institutional structures.

It does not authorize Standards Body to recognize governments, accredit evaluators, certify systems, or negotiate treaties.

It defines the infrastructure required for distinct institutions to exchange and interpret credible evidence without erasing legitimate differences.

Executive Summary

Frontier AI systems cross borders more easily than the institutions responsible for evaluating and governing them.

A model may be:

- Trained in one country
- Fine-tuned in another
- Hosted across several cloud regions
- Integrated by a multinational company
- Evaluated by laboratories in multiple jurisdictions
- Accessed by users worldwide
- Modified by an open-source community
- Connected to tools and data governed by different legal systems
- Deployed in sectors with distinct professional obligations
- Updated continuously after initial review

The evidence surrounding that system may also be distributed.

One institution may hold:

- Capability evaluations

Another may hold:

- Safeguard results

Another may possess:

- Security incidents

Another may assess:

- Organizational governance

Another may issue:

- A certification or procurement decision

Another may possess:

- National-security-relevant information that cannot be published

Without interoperability, these institutions can produce a large quantity of evidence that does not combine into shared understanding.

The same term may have different meanings.

The same score may be produced under different conditions.

The same model name may refer to different configurations.

The same certification language may imply different levels of assurance.

A threshold in one framework may not correspond to a threshold in another.

An evaluator accepted in one jurisdiction may be unrecognized in another.

An incident category may not transfer across legal systems.

A confidential finding may be too sensitive for public disclosure but too important to remain isolated.

A requirement may be technically equivalent to another requirement while appearing different because the documentation and institutional form are different.

These failures create practical consequences:

- Duplicated evaluations
- delayed deployment
- inconsistent safeguards
- regulatory arbitrage
- contradictory public claims
- unnecessary trade barriers
- weak incident coordination
- evaluator shopping
- exclusion of smaller countries and organizations
- dependence on a few dominant institutions
- pressure for lowest-common-denominator standards
- public confusion about what has actually been established

Global interoperability is the discipline of making evidence, standards, protocols, qualifications, and institutional decisions usable across boundaries.

It does not require one universal test.

It does not require one legal system.

It does not require one threshold.

It does not require one definition of acceptable risk.

It requires enough shared structure that different institutions can answer:

- Are we evaluating the same construct?
- Are we evaluating the same system?
- Were the test conditions materially comparable?
- What was held constant?
- What differed?
- How confident is the result?
- Is the evaluator competent for this scope?
- What does the assurance claim cover?
- Is the evidence current?
- Has the result been suspended, corrected, or superseded?
- Which local requirements remain unresolved?
- Can the result be recognized, partially recognized, or used only as supporting evidence?

Standards Body adopts the following core position:

> **Global frontier AI governance should pursue interoperable evidence, protocols, terminology, and assurance systems while preserving legitimate jurisdictional, cultural, institutional, and policy differences. The objective is shared understanding and portable trust, not forced global uniformity.**

A mature interoperability system should operate across at least ten layers.

1. Semantic interoperability

Institutions use shared definitions or explicit mappings among terms such as:

- Model
- system
- capability
- incident
- evaluator
- threshold
- safeguard
- audit
- certification
- accreditation
- critical capability
- deployment change

2. Identity interoperability

Evidence refers to a verifiable model, system, configuration, protocol, evaluator, and date.

3. Protocol interoperability

Evaluation procedures can be compared, mapped, reused, or reproduced across institutions.

4. Measurement interoperability

Scores, uncertainty, baselines, task populations, and thresholds can be interpreted without false equivalence.

5. Metadata interoperability

Reports include a common minimum set of machine-readable and human-readable information.

6. Assurance interoperability

Evaluator qualifications, scopes, review levels, accreditation, certification, and result status can be understood across borders.

7. Incident interoperability

Organizations can classify, share, escalate, and learn from incidents using compatible structures.

8. Legal and policy interoperability

Institutions can map technical evidence into different legal regimes without pretending the regimes are identical.

9. Security interoperability

Sensitive evidence can be exchanged through trusted, tiered, and accountable channels.

10. Institutional interoperability

Governments, laboratories, standards bodies, auditors, developers, researchers, and open communities can cooperate without surrendering their respective authority.

Interoperability should be built through modular agreements.

A country may recognize another country's evaluator competence while retaining its own deployment decision.

A purchaser may accept an evaluation report while requiring additional local testing.

A standards body may map two protocols without declaring them equivalent.

A regulator may recognize an international standard as evidence without making it the only compliance route.

A confidential incident exchange may operate among trusted institutions while a public summary remains available.

A multilingual protocol may preserve a shared construct while allowing locally valid tasks.

The central institutional distinction is between:

- **Recognition of evidence**
- **Recognition of competence**
- **Recognition of process**
- **Recognition of legal effect**

These are not the same.

A result can be scientifically informative without being legally determinative.

An evaluator can be competent without being authorized by every jurisdiction.

A certification can be valid under one scheme without satisfying every local requirement.

Interoperability should therefore support graded outcomes:

- Fully compatible
- Compatible with stated conditions
- Partially comparable
- Supporting evidence only
- Not comparable
- Recognition suspended
- Recognition withdrawn

The global landscape already contains important building blocks.

NIST has published a plan for global engagement on AI standards centered on cooperation, consensus standards, and information sharing.[^nist-global] The OECD's Hiroshima AI Process Reporting Framework provides a common voluntary reporting structure for advanced AI governance and risk-management practices.[^haip] The Council of Europe Framework Convention on Artificial Intelligence establishes a treaty-level framework focused on human rights, democracy, and the rule of law.[^coe-convention] The United Nations Global Digital Compact establishes a broad global framework for digital cooperation and AI governance.[^un-gdc] International accreditation arrangements such as the ILAC Mutual Recognition Arrangement demonstrate how testing and inspection results can gain cross-border acceptance through peer-evaluated accreditation systems.[^ilac-mra] The International Network for Advanced AI Measurement, Evaluation and Science has begun articulating areas of consensus and open questions for advanced AI evaluation across national institutes.[^aisi-network]

These initiatives are not interchangeable.

They operate at different levels:

- Technical
- organizational
- voluntary
- legal
- diplomatic
- scientific
- assurance
- capacity-building

Global interoperability should connect them without collapsing them.

The eighth foundation of Standards Body is therefore a shared evidentiary language for plural institutions.

---

1. Foundational Proposition

1.1 Core Thesis

> **Frontier AI evidence should be understandable and usable across institutional and jurisdictional boundaries without requiring every institution to adopt the same law, protocol, threshold, or policy judgment.**

1.2 Pluralism Thesis

> **Interoperability should preserve legitimate diversity while making differences explicit.**

1.3 Evidence Thesis

> **The first object of international alignment should be the structure and meaning of evidence, not immediate agreement on every policy outcome.**

1.4 Identity Thesis

> **No evaluation, certification, or incident record is globally useful unless the system, configuration, protocol, evaluator, and time are identifiable.**

1.5 Recognition Thesis

> **Recognition should be granular. Institutions should be able to recognize competence, process, evidence, or legal effect separately.**

1.6 Capacity Thesis

> **Global interoperability is incomplete when only a small number of wealthy countries, companies, or laboratories can produce or interpret the required evidence.**

1.7 Security Thesis

> **International evidence sharing should protect sensitive models, vulnerabilities, incidents, personal data, and dual-use information through proportionate access controls and accountable disclosure.**

1.8 Revisability Thesis

> **Mappings, equivalence decisions, recognition arrangements, and shared schemas should change as standards, systems, and institutional capabilities change.**

---

2. Scope and Boundaries

2.1 What This Foundation Covers

This paper covers interoperability concerning:

- Evaluation terminology
- Model and system identity
- Capability taxonomies
- Evaluation protocols
- Scoring and uncertainty
- Report metadata
- Evaluator competence
- Accreditation and mutual recognition
- Certification claims
- Incident reporting
- Secure evidence exchange
- Standards mapping
- Legal crosswalks
- multilingual evaluation
- open-source participation
- capacity building
- international networks
- dispute resolution
- registry design

2.2 What This Foundation Does Not Establish

This paper does not establish:

- A universal AI law
- A global licensing regime
- One international capability threshold
- One global task bank
- One mandatory evaluator
- One global certification mark
- Automatic recognition of foreign legal decisions
- Mandatory transfer of protected data
- A replacement for national or regional authority
- A treaty position on behalf of Standards Body

2.3 Interoperability Versus Harmonization

Interoperability enables systems or institutions to exchange and use information effectively.

Harmonization reduces differences among requirements, methods, or standards.

Interoperability can exist without full harmonization.

2.4 Interoperability Versus Equivalence

Two approaches may interoperate without being equivalent.

Equivalence is a stronger claim that outcomes or requirements are sufficiently comparable for a stated purpose.

2.5 Interoperability Versus Mutual Recognition

Mutual recognition is an institutional agreement to accept specified results, qualifications, certificates, or processes.

It normally depends on interoperability but adds legal or organizational commitment.

2.6 Interoperability Versus Uniformity

Uniformity requires sameness.

Interoperability requires understandable and manageable difference.

2.7 Technical Versus Political Agreement

Technical comparability does not automatically create political agreement.

A shared cyber capability result may support different national responses.

---

3. Canonical Definitions

3.1 Interoperability

The ability of distinct systems, organizations, protocols, or jurisdictions to exchange, interpret, and use information or evidence effectively.

3.2 Semantic Interoperability

Shared or mapped meaning among terms, classifications, and data elements.

3.3 Syntactic Interoperability

Compatibility in data structure, format, encoding, and transmission.

3.4 Procedural Interoperability

Compatibility among processes, workflows, responsibilities, and decision steps.

3.5 Technical Interoperability

Ability of technical systems, tools, APIs, schemas, and environments to work together.

3.6 Measurement Interoperability

Ability to interpret and compare measurements produced by different methods, instruments, task sets, or institutions.

3.7 Institutional Interoperability

Ability of organizations with different mandates and governance structures to coordinate and rely on one another's work.

3.8 Legal Interoperability

Ability to map or coordinate requirements and evidence across legal systems while preserving each system's authority.

3.9 Mutual Recognition

An arrangement through which parties accept specified results, qualifications, certificates, or decisions issued under another recognized system.

3.10 Unilateral Recognition

Acceptance by one party of another party's evidence or status without reciprocal obligation.

3.11 Equivalence

A determination that different requirements, methods, or systems achieve sufficiently comparable outcomes for a defined purpose.

3.12 Comparability

The degree to which results can be meaningfully compared.

3.13 Compatibility

The degree to which systems or requirements can operate together without unacceptable conflict.

3.14 Crosswalk

A structured mapping between terms, requirements, controls, classifications, or standards.

3.15 Concordance

A mapping showing relationships among multiple vocabularies or classification systems.

3.16 Reference Architecture

A common conceptual structure that supports implementation by different organizations.

3.17 Common Minimum

A baseline set of shared requirements or metadata accepted across participating systems.

3.18 Localization

Adaptation of a protocol, standard, or system to local language, law, culture, infrastructure, or professional practice.

3.19 Translation Validity

The degree to which a translated evaluation preserves the intended construct and interpretation.

3.20 Recognition Scope

The specific activities, methods, domains, systems, and conditions covered by a recognition arrangement.

3.21 Trust Anchor

An institution, credential, cryptographic root, registry, or assurance mechanism relied upon to establish identity or legitimacy.

3.22 Registry

A maintained record of protocols, systems, evaluators, certificates, incidents, mappings, or recognition status.

3.23 Evidence Package

A structured set of artifacts supporting a claim or decision.

3.24 Interoperability Profile

A declaration of the standards, schemas, protocols, identifiers, and recognition conditions supported by an organization or system.

3.25 Recognition Decision

A documented determination to accept, conditionally accept, partially accept, suspend, or reject external evidence or status.

3.26 Localization Layer

The part of a shared system that permits jurisdictional, linguistic, cultural, or domain-specific adaptation.

3.27 Core Layer

The shared components that should remain stable across implementations.

3.28 Bridge Study

An analysis that connects results across protocols, versions, languages, or task forms.

3.29 Technical Barrier

An unnecessary or disproportionate difference in requirements or procedures that impedes cross-border use or trade.

3.30 Regulatory Arbitrage

Movement of activity toward jurisdictions or institutional arrangements with weaker or more favorable requirements.

3.31 Data Sovereignty

The principle that data is subject to the laws, governance, and control arrangements associated with relevant jurisdictions or communities.

3.32 Federated Evaluation

Evaluation in which data, tasks, models, or evidence remain distributed while coordinated methods produce shared results.

3.33 Confidential Exchange Network

A governed network through which authorized institutions share nonpublic information under defined security, use, and accountability rules.

3.34 Interoperability Debt

The accumulated cost created by incompatible terminology, formats, identifiers, protocols, and institutional arrangements.

3.35 Recognition Drift

Deterioration in the validity of a recognition arrangement after methods, standards, institutions, or systems change.

---

4. Why Global Interoperability Is Necessary

4.1 AI Systems Are Transnational

Model supply chains, hosting, users, data, capital, and applications cross jurisdictions.

4.2 Evidence Is Fragmented

No single institution sees the entire system.

4.3 Duplicate Evaluation Is Expensive

Repeated assessments can consume:

- Expert time
- developer access
- compute
- security resources
- held-out tasks
- deployment schedules

Some duplication is valuable for independent replication.

Unnecessary duplication is not.

4.4 Incompatible Language Creates False Disagreement

Institutions may use different words for similar concepts or the same word for different concepts.

4.5 Incompatible Scores Create False Comparison

A score of 70 under one protocol may not be comparable with 70 under another.

4.6 Legal Fragmentation Creates Operational Burden

Organizations may need separate evidence packages for each market.

4.7 Incidents Cross Borders

Misuse, security failures, and model behavior can affect users in multiple countries.

4.8 Evaluator Capacity Is Uneven

Some regions have advanced laboratories and institutes.

Others do not.

4.9 International Trust Is Limited

Governments and organizations may distrust foreign evaluators, proprietary evidence, or politically influenced institutions.

4.10 Open Models Are Globally Distributed

No single developer can control all downstream versions or deployments.

4.11 Standards Can Reduce Barriers

Shared standards and recognition can reduce unnecessary repeated testing.

4.12 Coordination Can Create New Risks

Interoperability can also:

- Spread weak standards
- centralize sensitive data
- enable surveillance
- favor dominant languages
- export powerful jurisdictions' assumptions
- create systemic dependence on one registry or evaluator network

It requires governance.

---

5. What Should Be Shared and What Should Remain Local

5.1 Strong Candidates for Shared Foundations

- Core terminology
- Protocol identifiers
- Model and system identity metadata
- Evaluator scope metadata
- Result status
- Versioning
- Uncertainty reporting
- Incident minimum fields
- Security classification labels
- Change records

- Recognition status
- Claims boundaries

5.2 Candidates for Shared Technical Methods

- Evaluation harness interfaces
- Task packaging
- Logging
- scoring metadata
- bridge-study methods
- reference systems
- proficiency testing
- result schemas

5.3 Candidates for Localization

- Language
- professional context
- legal duties
- cultural assumptions
- user populations
- local infrastructure
- risk tolerance
- enforcement
- remedies
- public disclosure rules

5.4 Candidates for National or Institutional Authority

- Deployment approval
- legal classification
- sanctions
- procurement decisions
- national-security action
- licensing
- constitutional rights balancing

5.5 Shared Evidence, Local Judgment

The preferred model is:

> Shared evidence structure, explicit local interpretation, accountable local or international decision authority.

5.6 No Hidden Localization

Local differences should be documented rather than silently embedded in test content or scoring.

---

6. Layers of Interoperability

Layer 1: Vocabulary

Can institutions understand each other's terms?

Layer 2: Identity

Can institutions verify the system, protocol, evaluator, and evidence source?

Layer 3: Data Format

Can reports and artifacts be exchanged?

Layer 4: Protocol

Can methods be reproduced or mapped?

Layer 5: Measurement

Can results be compared with appropriate uncertainty?

Layer 6: Assurance

Can evaluator competence and result status be recognized?

Layer 7: Security

Can sensitive information move safely?

Layer 8: Legal Mapping

Can evidence be used under different legal systems?

Layer 9: Institutional Workflow

Can organizations coordinate review, escalation, and appeals?

Layer 10: Capacity

Can all participating regions meaningfully implement and use the system?

Failure at one layer can undermine the others.

---

7. Design Principles

7.1 Interoperability Before Uniformity

Begin with compatible evidence and explicit mappings.

7.2 Meaning Before Format

A common JSON field is not useful if institutions interpret it differently.

7.3 Identity Before Comparison

Confirm what was evaluated before comparing results.

7.4 Purpose-Bounded Equivalence

Equivalence should always state the purpose for which it is accepted.

7.5 Minimum Common Core

Share the smallest stable core necessary for cooperation.

7.6 Documented Localization

Allow local adaptation with a clear record.

7.7 No False Comparability

State when results cannot be compared.

7.8 Plural Trust Anchors

Avoid dependence on one institution, registry, cloud, country, or accreditation body.

7.9 Proportional Security

Share enough for coordination without unnecessary disclosure.

7.10 Capacity Inclusion

Interoperability should include funding, training, infrastructure, and translation.

7.11 Open Interfaces

Prefer documented interfaces and portable evidence.

7.12 Standards Neutrality with Mapping

Support multiple standards through crosswalks where appropriate.

7.13 Versioned Recognition

Recognition applies to identified versions and scopes.

7.14 Contestability

Mappings and recognition decisions should be appealable and reviewable.

7.15 Human Legibility

Machine-readable systems should have clear human explanations.

7.16 Dynamic Maintenance

Crosswalks, schemas, and recognition should update as the field changes.

---

8. Semantic Interoperability

8.1 The Vocabulary Problem

Terms such as "frontier model," "systemic risk," "critical capability," and "independent evaluator" vary across organizations.

8.2 Canonical Vocabulary

A shared vocabulary should include:

- Preferred term
- definition
- scope
- exclusions

- synonyms
- related terms
- source
- version
- jurisdictional notes

8.3 Term Mapping

Mappings may indicate:

- Exact match
- broader term
- narrower term
- partial overlap
- related but distinct
- no equivalent

8.4 Definition Governance

Definitions should be:

- Evidence-based
- versioned
- publicly reviewable
- translatable
- linked to use cases

8.5 Legal Definitions

Legal definitions may need precise jurisdictional meaning.

Do not overwrite them with technical vocabulary.

8.6 Translation

Translate concepts, not only words.

8.7 Ambiguity Register

Maintain disputed terms and unresolved interpretations.

8.8 Standards Body Terminology Role

`TERMINOLOGY.md` should become the canonical project vocabulary, with Foundation 8 defining mapping requirements.

9. Model and System Identity

9.1 Why Identity Matters

A model name is often insufficient.

Systems can differ through:

- Checkpoint
- fine-tuning
- system prompt
- tools
- retrieval
- safety layers
- inference settings
- region
- access tier
- update date

9.2 Minimum Identity Record

- Developer
- model family
- model identifier
- version
- release or deployment date
- endpoint or artifact
- fine-tuning
- system prompt status
- tools
- retrieval
- memory
- safety configuration
- inference settings
- evaluated date
- evaluator

9.3 Cryptographic Identity

Where feasible, use:

- Hashes
- signed manifests
- attestation

- artifact signatures
- reproducible build identifiers

9.4 Closed Systems

Closed APIs may require developer attestation and evaluator verification.

9.5 Open-Weight Systems

Record:

- Weight hash
- repository
- license
- quantization
- fine-tune
- merge
- tokenizer
- inference stack

9.6 Composite Systems

Identity should include all material components.

9.7 Identity Change Trigger

A material identity change should trigger re-evaluation or explicit inheritance analysis.

9.8 Global Identifier

A future system could assign persistent, non-proprietary identifiers for evaluated AI artifacts and configurations.

9.9 Privacy and Security

Identity systems should not disclose sensitive information unnecessarily.

---

10. Protocol Interoperability

10.1 Protocol Identity

Each protocol should have:

- Unique identifier
- owner
- version
- status
- scope
- construct
- administration
- scoring
- security
- expiration

10.2 Protocol Profile

A profile should describe:

- Required components
- optional components
- localization
- dependencies
- reference implementations

10.3 Common Execution Interface

Shared evaluation tools can support portability.

The UK AI Security Institute's Inspect framework demonstrates a modular approach to evaluation tasks, agents, tools, scorers, and model interfaces.[^inspect]

10.4 Reference Implementation

A reference implementation can clarify the protocol.

It should not become the only permitted implementation unless required.

10.5 Protocol Mapping

Compare:

- Construct
- task population

- model configuration
- tools
- retries
- scoring
- uncertainty
- human baseline
- security
- reporting

10.6 Compatibility Levels

Level A: Directly Reproducible

Same protocol and materially equivalent execution.

Level B: Form-Equivalent

Different task forms with validated comparability.

Level C: Construct-Compatible

Different methods measuring substantially the same construct.

Level D: Supporting Evidence

Related but not directly comparable.

Level E: Incompatible

No defensible comparison.

10.7 Bridge Studies

Use shared systems, anchor tasks, human baselines, and statistical analysis.

10.8 Protocol Forks

Allow forks with explicit lineage and compatibility statements.

10.9 Protocol Retirement

Retired protocols should remain discoverable with status.

---

11. Measurement Interoperability

11.1 Score Meaning

A score requires context.

11.2 Minimum Measurement Metadata

- Metric
- scale
- task population
- sample size
- uncertainty
- retries
- aggregation
- invalid runs
- baseline
- date
- configuration
- protocol version

11.3 Common Units

Use common units where valid.

Examples:

- Success probability
- time
- cost
- task horizon
- failure rate
- calibration error

11.4 Local Metrics

Some domains require specialized measures.

Map rather than force conversion.

11.5 Threshold Mapping

Thresholds may be:

- Technically comparable

- institutionally different
- legally different

11.6 Human Baselines

Human comparisons require common definitions of:

- Expertise
- tools
- time
- assistance
- selection
- compensation

11.7 Uncertainty

Interoperability should carry uncertainty, not only point estimates.

11.8 Statistical Equating

Use only when assumptions are defensible.

11.9 No Universal Score

Standards Body opposes a single global frontier safety score.

11.10 Measurement Registry

A future registry can document metrics, protocols, mappings, and limitations.

---

12. Common Metadata Architecture

12.1 Purpose

Metadata makes evidence discoverable and interpretable.

12.2 Core Record Categories

System

What was evaluated?

Protocol

How was it evaluated?

Evaluator

Who performed the work?

Evidence

What artifacts support the result?

Result

What was found?

Assurance

What review or recognition applies?

Status

Is the evidence current?

Jurisdiction

Where and how can it be used?

12.3 Required Fields

- Record identifier
- record type
- issuing organization
- date
- version
- system identifier
- protocol identifier
- evaluator identifier
- scope
- result
- confidence
- limitations
- security classification
- status
- expiry
- correction history
- recognition

12.4 Machine-Readable Format

Use open, documented schemas.

12.5 Human-Readable Summary

Every machine record should have a plain-language explanation.

12.6 Provenance

Record:

- Source
- modifications
- signatures
- chain of custody
- related records

12.7 Privacy

Minimize personal data.

12.8 Extensibility

Allow domain and jurisdiction extensions.

12.9 Schema Governance

Define:

- Versioning
- compatibility
- deprecation
- validation
- public comment
- security

---

13. Evaluator and Assurance Interoperability

13.1 Competence Recognition

An evaluator's competence should be tied to:

- Activity

- domain
- method
- system type
- assurance level
- security level
- jurisdiction

13.2 Accreditation

Accreditation can support recognition when accreditation bodies operate under shared requirements and peer evaluation.

13.3 ILAC Model

The ILAC Mutual Recognition Arrangement supports cross-border acceptance of results from accredited testing, calibration, inspection, proficiency-testing, and reference-material activities. [^ilac-mra]

Its general principle, accredited once and accepted across participating systems, is instructive.

Frontier AI will require narrower and more dynamic scopes.

13.4 Recognition Types

- Evaluator competence
- test result
- inspection report
- certification
- validation statement
- proficiency result
- security qualification

13.5 Conditional Recognition

Recognition may require:

- Additional local module
- supplemental test
- translation review
- local legal assessment
- security review
- updated system identity

13.6 Recognition Registry

Publish:

- Recognized body
- recognition authority
- scope
- status
- conditions
- dates
- suspensions
- withdrawals

13.7 Peer Evaluation

Institutions recognizing evaluators should themselves be reviewed.

13.8 Government Institutes

National institutes may recognize each other's technical evidence while retaining independent conclusions.

13.9 No Recognition by Prestige Alone

Institutional reputation is not a substitute for scope and evidence.

---

14. Certification and Conformity Claims Across Borders

14.1 Certification Portability

A certificate may travel only if:

- Scheme is understood
- evaluator is recognized
- system and version are identified
- local requirements are mapped
- certificate remains current

14.2 Management-System Certification

ISO/IEC 42001 provides requirements for AI management systems, and ISO/IEC 42006 provides requirements for bodies auditing and certifying those systems.[^iso-42001][^iso-42006]

These standards can support international consistency in organizational assurance.

They do not establish universal system safety.

14.3 Product and Capability Claims

More technical evidence may be needed for:

- Model capability
- safeguards
- deployment behavior
- high-stakes thresholds

14.4 Recognition Statement

A recognition statement should identify:

- What is accepted
- for which purpose
- under which conditions
- what remains local
- duration
- appeal

14.5 Marks

Global marks risk oversimplification.

Prefer verifiable registry records.

14.6 Suspension Propagation

Material suspension should be communicated across recognition networks.

14.7 Certificate Translation

Translation should preserve legal and technical meaning.

---

15. Incident Interoperability

15.1 Why Incident Exchange Matters

Incidents can reveal:

- New capabilities
- misuse

- safeguard failures
- security weaknesses
- evaluation gaps
- systemic interactions

15.2 Incident Categories

- Safety
- security
- misuse
- privacy
- human rights
- reliability
- evaluation integrity
- model control
- critical infrastructure
- near miss

15.3 Minimum Incident Record

- Identifier
- date
- system
- version
- location
- category
- severity
- impact
- detection
- cause
- safeguards
- response
- status
- confidence
- disclosure level

15.4 Severity Mapping

Different institutions may use different scales.

Provide crosswalks.

15.5 Public and Restricted Layers

Public

High-level incident and lesson.

Trusted Network

Detailed technical and operational evidence.

Restricted

Highly sensitive vulnerabilities, personal data, or national-security information.

15.6 Notification

Define:

- Who must be notified
- timing
- format
- confidentiality
- further dissemination

15.7 Near Misses

Include near misses to improve prevention.

15.8 Duplicate Incidents

Use common identifiers and linkage.

15.9 Incident Feedback

Incidents should update:

- Protocols
- standards
- safeguards
- evaluator guidance
- research agendas
- recognition status

15.10 International Confidential Network

A future network could support trusted exchange among qualified national institutes and evaluators.

16. Secure Evidence Exchange

16.1 Information Classes

- Public
- controlled
- confidential
- restricted security
- highly restricted

16.2 Sharing Agreement

Define:

- Purpose
- authorized users
- onward disclosure
- storage
- retention
- incident response
- legal process
- audit
- termination

16.3 Technical Controls

- Encryption
- identity
- access control
- logging
- signed artifacts
- secure compute
- data loss prevention
- compartmentalization

16.4 Federated Analysis

Keep evidence local while sharing:

- Aggregates
- queries
- model outputs
- signed findings
- privacy-preserving statistics

16.5 Confidential Computing

May support joint evaluation without full asset disclosure.

Limitations remain.

16.6 Legal Constraints

Account for:

- Privacy
- export control
- trade secrets
- national security
- discovery
- data localization

16.7 Trust Framework

A secure exchange network needs:

- Membership criteria
- identity assurance
- sanctions
- audits
- dispute process
- key management
- incident response

16.8 No Automatic Sharing

Interoperability should not compel unsafe or unlawful disclosure.

---

17. Standards Crosswalks

17.1 Purpose

Crosswalks reduce duplication and clarify overlap.

17.2 Crosswalk Objects

- NIST AI RMF
- ISO/IEC 42001
- ISO/IEC 23894

- organizational frameworks
- sector rules
- procurement requirements
- legal obligations
- evaluator standards

17.3 Mapping Types

- Equivalent
- substantially aligned
- partially aligned
- complementary
- conflicting
- absent

17.4 Evidence

Every mapping should include:

- Text
- interpretation
- rationale
- reviewer
- confidence
- version
- limitations

17.5 No Automated Equivalence by Keyword

Similar language does not prove equivalent requirements.

17.6 Control Inheritance

Evidence under one standard may support another requirement.

It should not be reused beyond its valid scope.

17.7 Crosswalk Governance

Use:

- Multi-institution review
- public comment
- conflict disclosure
- versioning

- appeal

17.8 Crosswalk Expiration

Update after either source changes.

17.9 Legal Crosswalks

Require jurisdiction-specific legal expertise.

---

18. Legal and Regulatory Interoperability

18.1 Different Legal Objectives

Legal systems may prioritize:

- Human rights
- product safety
- national security
- innovation
- competition
- consumer protection
- privacy
- administrative accountability

18.2 Evidence Portability

Technical evidence can be portable even when legal consequences differ.

18.3 EU AI Act

The European Union AI Act creates a risk-based legal structure with differentiated obligations for AI systems and general-purpose AI models.[^eu-ai-act]

Technical evidence may support compliance, but local legal interpretation remains necessary.

18.4 Council of Europe Convention

The Council of Europe Framework Convention provides a treaty-level framework focused on human rights, democracy, and the rule of law and was opened for signature in September 2024. [^coe-convention]

It illustrates a rights-centered international layer distinct from technical evaluation standards.

18.5 United Nations Global Digital Compact

The Global Digital Compact provides a global framework for digital cooperation and AI governance, including capacity building and international cooperation.[^un-gdc]

18.6 Presumption and Recognition

Law may:

- Recognize standards
- accept accredited reports
- require local assessment
- allow equivalent evidence
- impose additional duties

18.7 Regulatory Cooperation

Authorities can cooperate through:

- Memoranda
- joint guidance
- information exchange
- coordinated enforcement
- common reporting
- shared technical studies

18.8 Sovereignty

Interoperability should not obscure who has legal authority.

18.9 Conflict of Laws

A system may face incompatible requirements.

Use:

- Conflict register
- technical alternatives
- jurisdictional review
- escalation
- transparent limitation

18.10 No Regulatory Laundering

Private standards should not acquire legal effect without accountable recognition.

---

19. Multilingual and Cultural Interoperability

19.1 Language Is Part of the Construct

Translation can change:

- Difficulty
- ambiguity
- cultural reference
- legal meaning
- social norms
- system performance

19.2 Translation Process

Use:

- Forward translation
- independent review
- back translation
- domain expert review
- pilot testing
- measurement analysis

19.3 Language Coverage

Report:

- Language
- dialect
- script
- region
- task origin
- translator
- validation

19.4 Cultural Adaptation

Some tasks should be localized rather than literally translated.

19.5 Cross-Language Comparability

Require bridge studies.

19.6 Local Experts

Local professional and cultural expertise is essential.

19.7 Dominant-Language Bias

English-centered protocols can misrepresent global capability and risk.

19.8 Legal Translation

Use qualified legal translators and jurisdictional review.

19.9 Multilingual Incident Exchange

Maintain canonical codes with local-language descriptions.

19.10 Recognition

Translation work should receive professional credit and funding.

---

20. Open-Source and Open-Weight Interoperability

20.1 Global Distribution

Open-weight models are copied, modified, quantized, merged, and deployed across borders.

20.2 Identity Challenge

Record:

- Base model
- hash
- modification
- fine-tune
- quantization
- inference
- repository
- license

20.3 Evaluation Portability

Open models can support independent reproduction.

20.4 Decentralized Responsibility

No single organization may control downstream deployment.

20.5 Community Registries

Communities can maintain:

- Model lineage
- evaluation results
- security notices
- known modifications
- incident reports

20.6 Recognition

Community evaluation should be recognized when methods and evidence are credible.

20.7 Security

Open access changes safeguard and threat assumptions.

20.8 Small-Actor Burden

Interoperability schemas should be implementable without large compliance teams.

20.9 Licensing

License terms affect use but should not be confused with evaluation evidence.

20.10 Open Interfaces

Open tooling and schemas reduce institutional dependence.

---

21. Capacity Building and Global Inclusion

21.1 Capacity Is Part of Interoperability

A standard that only a few countries can implement is not globally interoperable in practice.

21.2 Capacity Dimensions

- Technical expertise

- evaluator workforce
- compute
- secure infrastructure
- legal expertise
- standards participation
- translation
- incident response
- accreditation
- research funding

21.3 Developing Countries

International systems should support meaningful participation, not only consultation.

The United Nations has emphasized capacity building, equitable access, and participation by developing countries in global AI governance and safe, secure, and trustworthy AI systems.[^un-resolution][^un-gdc]

21.4 Capacity Mechanisms

- Training
- fellowships
- shared facilities
- regional laboratories
- grants
- open tools
- standards access
- translation
- evaluator partnerships
- twinning programs

21.5 Regional Hubs

Regional centers can adapt methods and build local trust.

21.6 Avoiding Dependency

Capacity building should support autonomous local competence rather than permanent reliance.

21.7 Funding Governance

Local institutions should participate in priority setting.

21.8 Brain Drain

Programs should support careers and institutions in participants' regions.

21.9 Measurement

Track who can:

- Evaluate
- interpret
- contribute
- govern
- challenge
- recognize evidence

---

22. International Networks and Institutions

22.1 National AI Measurement and Security Institutes

National institutes can:

- Evaluate models
- coordinate research
- support government decisions
- share methods
- build public infrastructure

22.2 International Network for Advanced AI Measurement, Evaluation and Science

The network has articulated shared areas of consensus and open questions concerning advanced AI evaluation, including the need for common understanding across borders.[^aisi-network]

22.3 Bilateral Partnerships

Bilateral agreements can support:

- Joint evaluations
- staff exchange
- shared infrastructure
- incident cooperation
- protocol development

22.4 OECD

The OECD supports:

- AI policy principles
- reporting frameworks
- comparative analysis
- policy observatories
- international coordination

22.5 United Nations

The UN provides broad global participation and development-oriented coordination.

22.6 Council of Europe

The Council of Europe provides a human-rights, democracy, and rule-of-law treaty framework.

22.7 ISO and IEC

ISO and IEC provide international consensus standards and conformity-assessment infrastructure.

22.8 Accreditation Networks

ILAC and related international accreditation systems demonstrate peer-evaluated recognition mechanisms.

22.9 Industry Forums

Industry groups can support technical coordination but require conflict controls.

22.10 Civil Society and Academia

Independent research and public-interest scrutiny are necessary counterweights.

22.11 Polycentric Governance

No single institution is likely to govern all layers effectively.

A polycentric system can distribute authority while sharing evidence.

---

23. Recognition Architecture

23.1 Four Recognition Objects

Evidence Recognition

Accepts an external result as informative.

Competence Recognition

Accepts that an evaluator is qualified within scope.

Process Recognition

Accepts that a protocol, audit, or certification process meets shared requirements.

Legal Recognition

Gives formal effect under a legal system.

23.2 Recognition Levels

R0: No Recognition

Evidence may be reviewed but receives no presumed weight.

R1: Informational Use

Evidence may support analysis.

R2: Conditional Technical Recognition

Evidence is accepted subject to conditions or supplemental work.

R3: Full Technical Recognition

Evidence is accepted for the defined technical purpose.

R4: Institutional Recognition

Evaluator, process, or certificate is recognized within scope.

R5: Legal Recognition

Evidence or status receives formal legal effect.

23.3 Recognition Criteria

- Identity
- scope

- competence
- method
- security
- uncertainty
- version
- governance
- complaints
- status
- reciprocity where required

23.4 Recognition Is Not Permanent

Use:

- Expiry
- surveillance
- event triggers
- suspension
- withdrawal

23.5 Recognition Decision Record

Document:

- Applicant
- object
- evidence
- purpose
- conditions
- local gaps
- reviewers
- conflicts
- decision
- duration

---

24. Dispute Resolution

24.1 Dispute Types

- Definition
- protocol mapping
- score comparability
- evaluator competence
- recognition scope

- security
- legal conflict
- incident classification
- translation
- suspension

24.2 Technical Resolution

Use:

- Expert panel
- bridge study
- additional evaluation
- shared reference system
- independent review

24.3 Institutional Resolution

Use:

- Negotiation
- peer review
- appeal body
- mediation
- recognition committee

24.4 Legal Dispute

Handled by competent legal institutions.

24.5 Dissent

Unresolved disagreement should remain visible.

24.6 Temporary Status

Use:

- Under review
- conditionally recognized
- suspended
- disputed

24.7 No Forced Equivalence

Failure to agree should not produce false compatibility.

---

25. Governance

25.1 Governance Functions

- Vocabulary stewardship
- schema maintenance
- protocol registry
- recognition
- security
- dispute resolution
- capacity building
- public reporting
- change control

25.2 Distributed Governance

Different organizations can own different functions.

25.3 Standards Body Role

At its present stage, Standards Body should:

- Develop conceptual architecture
- publish schemas and crosswalk methods
- support pilots
- convene technical contributors
- document open questions
- avoid claiming formal recognition authority

25.4 Balanced Participation

Include:

- National institutes
- standards bodies
- evaluators
- developers
- open-source experts
- public-interest organizations
- affected sectors

- developing countries
- domain specialists

25.5 Conflict Controls

Disclose:

- Funding
- national interests
- commercial interests
- standards ownership
- evaluator relationships
- political mandates

25.6 Decision Separation

Separate technical mapping from legal recognition where possible.

25.7 Public Records

Publish:

- Versions
- decisions
- dissent
- recognition status
- corrections
- governance

25.8 Emergency Change

Use narrow, temporary processes for:

- Security compromise
- invalid mapping
- incident
- protocol failure

25.9 Appeals

Allow challenge of recognition and mapping decisions.

25.10 Sunset

Governance bodies and arrangements should be reviewable.

26. Maturity Model

Level 0: Fragmented Evidence

Characteristics:

- Inconsistent terms
- unverified identity
- incomparable scores
- bilateral improvisation

Level 1: Shared Vocabulary and Metadata

Characteristics:

- Core definitions
- identifiers
- report minimums
- versioning
- status

Level 2: Protocol and Evaluator Compatibility

Characteristics:

- Crosswalks
- bridge studies
- evaluator scopes
- common execution interfaces
- incident taxonomy

Level 3: Conditional Recognition Network

Characteristics:

- Recognition decisions
- trusted exchange
- registries
- peer review
- supplemental local modules
- dispute process

Level 4: Internationally Interoperable Assurance Regime

Characteristics:

- Mature mutual recognition
- secure incident exchange
- multilingual validation
- capacity inclusion
- continuous surveillance
- legal and procurement reliance
- plural trust anchors

---

27. Implementation Pathway

Phase 1: Vocabulary Inventory

Collect definitions across major frameworks and institutions.

Phase 2: Core Terminology

Publish a stable minimum and explicit mappings.

Phase 3: Metadata Schema

Create a machine-readable evaluation record.

Phase 4: Model Identity Pilot

Test identity records for closed and open systems.

Phase 5: Protocol Registry

Register protocol versions, owners, scope, and status.

Phase 6: Crosswalk Pilot

Map two evaluation protocols and two governance frameworks.

Phase 7: Bridge Study

Run common systems across protocols.

Phase 8: Evaluator Profile

Publish competence and recognition metadata.

Phase 9: Incident Taxonomy

Pilot public and restricted incident records.

Phase 10: Secure Exchange

Create a limited trusted network among qualified partners.

Phase 11: Recognition Pilot

Conduct a conditional technical-recognition decision.

Phase 12: Capacity Program

Support participation from underrepresented regions and smaller evaluators.

Phase 13: International Review

Invite external institutions to challenge the architecture.

Phase 14: Revise and Scale

Expand only after evidence.

---

28. Proposed Standards Body Pilot

28.1 Pilot Name

Global Frontier Evaluation Interoperability Profile

28.2 Purpose

Demonstrate that two independent institutions can exchange and interpret frontier AI evaluation evidence without using identical protocols or surrendering local decision authority.

28.3 Pilot Domain

Autonomous cyber capability.

This domain connects Foundations 1 through 7 and provides:

- Defined task families
- agent environments
- held-out components
- high-stakes relevance
- multiple evaluator types
- measurable outputs

28.4 Participants

Target participation from:

- One government or public research institute
- one academic or nonprofit evaluator
- one commercial evaluator
- one open-source evaluation group
- one standards or assurance expert
- one institution from an underrepresented region

28.5 Shared Core

- Capability vocabulary
- model identity
- protocol metadata
- result schema
- uncertainty
- evaluator profile
- security labels
- incident fields
- recognition status

28.6 Local Modules

Participants retain:

- Task variants
- local legal analysis
- deployment decision
- language
- supplemental security requirements
- local threshold response

28.7 Evaluation

Each institution evaluates:

- A shared reference model
- its own approved task form
- common anchor tasks
- equivalent system configuration where possible

28.8 Bridge Study

Compare:

- Construct coverage
- scores
- uncertainty
- failure patterns
- human baselines
- threshold interpretation
- cost
- time

28.9 Recognition Exercise

Each institution decides whether the other's evidence is:

- Fully recognized
- conditionally recognized
- supporting evidence
- not comparable

28.10 Incident Exchange Exercise

Simulate a cross-border safeguard failure and test:

- Classification
- secure exchange
- public summary
- protocol update
- recognition status

28.11 Outputs

- Interoperability profile
- metadata schema
- protocol crosswalk

- bridge-study report
- evaluator-recognition template
- incident taxonomy
- secure-exchange agreement
- localization report
- capacity analysis
- public lessons report

28.12 Success Criteria

The pilot succeeds if it:

- Preserves construct meaning across protocols
- identifies the evaluated system precisely
- supports conditional recognition
- makes noncomparability visible
- protects sensitive evidence
- includes multilingual or regional participation
- reduces unnecessary duplication
- preserves local decision authority
- generates reusable technical artifacts

---

29. Metrics for Evaluating Interoperability

29.1 Semantic Quality

- Definition agreement
- unresolved ambiguity
- mapping confidence
- translation validity

29.2 Identity Quality

- System traceability
- configuration completeness
- artifact verification
- update detection

29.3 Protocol Quality

- Reproducibility
- mapping completeness
- bridge-study success

- task-form equivalence

29.4 Measurement Quality

- Comparability
- uncertainty preservation
- baseline consistency
- threshold interpretation

29.5 Data Quality

- Schema validity
- completeness
- machine readability
- provenance
- correction

29.6 Recognition Quality

- Decision consistency
- scope clarity
- conditions
- suspension propagation
- appeal

29.7 Security

- Unauthorized disclosure
- access integrity
- incident response
- secure-exchange performance

29.8 Capacity

- Countries represented
- evaluator participation
- translation
- shared infrastructure
- local competence

29.9 Efficiency

- Reduced duplicate work
- evaluation time

- integration cost
- report reuse
- recognition delay

29.10 Governance

- Participation balance
- conflict handling
- dissent
- change speed
- public transparency

29.11 Decision Utility

- Procurement use
- policy use
- safeguard coordination
- incident learning
- evaluator reliance
- reduction in false claims

---

30. Failure Modes and Safeguards

30.1 Uniformity Disguised as Interoperability

Failure: One institution's system becomes the global default without meaningful consent.

Safeguard: modular core, localization, plural governance, alternative mappings.

30.2 Lowest Common Denominator

Failure: Shared standards become too weak to matter.

Safeguard: common baseline plus higher-assurance profiles.

30.3 False Equivalence

Failure: Different protocols are treated as interchangeable.

Safeguard: purpose-bounded recognition and bridge studies.

30.4 Semantic Drift

Failure: Shared terms acquire different meanings.

Safeguard: versioned vocabulary and definition governance.

30.5 Model Identity Failure

Failure: Evidence is applied to a different model or configuration.

Safeguard: signed manifests, configuration metadata, re-evaluation triggers.

30.6 Score Portability Error

Failure: Numeric results travel without context.

Safeguard: required protocol, uncertainty, and task metadata.

30.7 Recognition Capture

Failure: Powerful jurisdictions or organizations control accepted evidence.

Safeguard: distributed recognition, peer review, appeals, capacity support.

30.8 Weak Standard Propagation

Failure: Mutual recognition spreads poor assurance.

Safeguard: surveillance, proficiency testing, suspension, limited scope.

30.9 Registry Monopoly

Failure: One registry controls legitimacy.

Safeguard: interoperable registries, open formats, mirrored records.

30.10 Sensitive Data Centralization

Failure: Global coordination creates a high-value breach target.

Safeguard: federation, compartmentalization, minimal disclosure.

30.11 Legal Overreach

Failure: Technical mapping is treated as legal equivalence.

Safeguard: separate technical and legal decisions.

30.12 Translation Error

Failure: Local-language evaluation changes construct meaning.

Safeguard: translation validation and local experts.

30.13 English Dominance

Failure: English performance and institutions define global capability.

Safeguard: multilingual task development, funding, and bridge studies.

30.14 Capacity Exclusion

Failure: Countries are expected to adopt standards they cannot implement.

Safeguard: capacity building, regional hubs, shared infrastructure.

30.15 Regulatory Arbitrage

Failure: Actors use interoperability to select the weakest regime.

Safeguard: recognition conditions and local minimums.

30.16 Duplicate Burden Persists

Failure: Institutions collect the same evidence in different formats.

Safeguard: common evidence package and crosswalks.

30.17 Incident Suppression

Failure: Legal or reputational concerns block sharing.

Safeguard: protected exchange, de-identification, corrective incentives.

30.18 Recognition Staleness

Failure: Institutions continue accepting obsolete scopes or methods.

Safeguard: expiry, surveillance, automatic status updates.

30.19 Geopolitical Fragmentation

Failure: Technical networks divide into incompatible blocs.

Safeguard: neutral interfaces, multilateral participation, scientific cooperation.

30.20 Authority Confusion

Failure: Standards Body or another technical institution is mistaken for a regulator.

Safeguard: precise public language and mandate boundaries.

---

31. Serious Objections

Objection 1: Global Interoperability Is Unrealistic During Geopolitical Competition

Competition creates distrust and strategic secrecy.

Response:

Begin with narrow technical artifacts:

- Metadata
- protocol identifiers
- incident fields
- evaluator competence
- public reporting

Residual concern:

High-sensitivity domains may remain fragmented.

Objection 2: Different Values Make Shared Standards Impossible

Values differ.

Shared evidence does not require shared values.

Institutions can agree on what was measured while disagreeing on acceptable risk.

Objection 3: Interoperability Helps Regulatory Arbitrage

It can.

Response:

Recognition should preserve local minimum requirements and conditions.

Objection 4: Global Standards Favor Large Companies

They can.

Response:

Include small-actor pathways, open tools, capacity support, and competition review.

Objection 5: Mutual Recognition Spreads Weak Evaluations

Correct.

Response:

Use peer evaluation, narrow scopes, proficiency testing, surveillance, and suspension.

Objection 6: Common Metadata Creates Bureaucracy

It adds reporting burden.

Response:

Use a minimal core, reusable machine-readable records, and evidence portability.

Objection 7: Confidential Information Cannot Be Shared Internationally

Some information cannot.

Interoperability can still support:

- Public summaries
- signed findings
- federated analysis
- restricted bilateral exchange
- local review

Objection 8: Legal Systems Cannot Be Harmonized

Full harmonization is not required.

Technical and evidentiary crosswalks can coexist with legal difference.

Objection 9: Translation Makes Evaluation Incomparable

Translation creates real measurement challenges.

Response:

Validate translation, localize where necessary, and report comparability limits.

Objection 10: One Global Registry Would Be Efficient

It would also create:

- Concentration
- political control
- security risk
- single-point failure

Use interoperable registries.

Objection 11: National Institutes Should Trust Only Their Own Evaluations

Independent national testing is valuable.

Exclusive reliance creates duplication and limits learning.

Conditional recognition preserves sovereignty.

Objection 12: Open-Weight Systems Cannot Be Governed Through Institutional Recognition

Their decentralized nature changes responsibility.

Interoperable identity, evaluation, incident, and community-governance systems remain useful.

---

32. Evidence Gaps

32.1 Construct Mapping

How reliably can different frontier evaluation frameworks be mapped?

32.2 Score Comparability

Which bridge methods work across dynamic, agentic, and held-out protocols?

32.3 Model Identity

What identity mechanisms work for closed, continuously updated systems?

32.4 Recognition

Which recognition models preserve rigor without excessive duplication?

32.5 Incident Exchange

What information can be shared safely and usefully across borders?

32.6 Translation

Which methods preserve construct validity across languages and cultures?

32.7 Capacity

Which capacity-building models produce durable local institutions?

32.8 Legal Crosswalks

How should technical evidence be mapped into distinct legal systems?

32.9 Accreditation

Can existing international accreditation arrangements adapt to dynamic frontier AI evaluation?

32.10 Federated Evaluation

When can models and data remain local while evidence remains comparable?

32.11 Registry Governance

How can registries remain accurate, distributed, and politically legitimate?

32.12 Recognition Drift

How quickly do recognition arrangements become obsolete?

32.13 Geopolitical Resilience

Which technical layers can remain interoperable during political conflict?

32.14 Decision Impact

Does interoperability improve safety and efficiency enough to justify institutional cost?

---

33. Research Agenda

Priority 1: Canonical Vocabulary

Build and validate a multilingual frontier evaluation ontology.

Priority 2: System Identity

Develop portable identifiers and configuration manifests.

Priority 3: Evaluation Metadata

Create an open minimum schema for protocols and results.

Priority 4: Protocol Crosswalks

Develop methods for mapping constructs, tasks, scoring, and thresholds.

Priority 5: Bridge Studies

Test statistical and qualitative comparability across institutions.

Priority 6: Evaluator Recognition

Pilot scope-specific competence recognition.

Priority 7: Incident Taxonomy

Develop public and restricted international incident records.

Priority 8: Secure Exchange

Test federated, confidential, and tiered evidence sharing.

Priority 9: Translation Validity

Develop multilingual evaluation methodology.

Priority 10: Capacity Building

Compare regional hubs, fellowships, shared facilities, and twinning.

Priority 11: Legal Interoperability

Develop technical-to-legal crosswalk methods.

Priority 12: Registry Architecture

Design distributed, signed, machine-readable registries.

Priority 13: Mutual Recognition

Pilot conditional recognition in bounded domains.

Priority 14: Recognition Surveillance

Develop event-triggered updates and suspension propagation.

Priority 15: Interoperability Impact

Measure duplication, cost, safety, participation, and trust.

---

34. Near-Term Experiments

Experiment 1: Vocabulary Crosswalk

Map ten core terms across five major frameworks.

Experiment 2: Model Manifest

Create identity records for one closed and one open-weight system.

Experiment 3: Shared Metadata

Have three evaluators publish the same minimum result schema.

Experiment 4: Protocol Bridge

Run two protocols on common reference systems.

Experiment 5: Conditional Recognition

Have two institutions assess whether to recognize each other's evidence.

Experiment 6: Incident Simulation

Run a confidential cross-border incident exchange drill.

Experiment 7: Translation Study

Translate and validate an evaluation in three languages.

Experiment 8: Registry Federation

Synchronize signed records across two independent registries.

Experiment 9: Capacity Pilot

Support a regional evaluator to implement the common profile.

Experiment 10: Standards Crosswalk

Map NIST AI RMF, ISO/IEC 42001, and one frontier safety framework.

Experiment 11: Legal Evidence Package

Test one technical report against two jurisdictional requirements.

Experiment 12: Recognition Suspension

Simulate method invalidation and status propagation.

---

35. Implications for Future Standards

A future global interoperability standard could require:

35.1 Vocabulary

Canonical terms, mappings, and versioning.

35.2 Identity

Persistent identifiers for systems, protocols, evaluators, and reports.

35.3 Metadata

Minimum human-readable and machine-readable fields.

35.4 Protocol Profile

Construct, tasks, administration, scoring, security, and expiration.

35.5 Measurement

Units, uncertainty, baselines, and comparability statement.

35.6 Evaluator Profile

Competence, scope, independence, security, and recognition.

35.7 Result Status

Current, expired, suspended, corrected, withdrawn, or superseded.

35.8 Recognition

Purpose, scope, conditions, duration, and appeal.

35.9 Incident Exchange

Classification, minimum data, sensitivity, and notification.

35.10 Security

Access control, provenance, retention, onward disclosure, and incident response.

35.11 Localization

Language, cultural, legal, and professional adaptation.

35.12 Crosswalks

Methods for mapping standards and requirements.

35.13 Capacity

Minimum support and participation expectations.

35.14 Governance

Change control, conflicts, dissent, and public records.

35.15 Retirement

Deprecation, transition, suspension, and archive.

Such a standard should be developed through `STANDARDS_DEVELOPMENT_PROCESS.md` with international participation.

---

36. Relationship to the Other Foundations

Foundation 1: Dynamic Evaluation Protocols

Interoperability must preserve protocol versioning, bridge studies, and explicit discontinuity.

Foundation 2: Held-Out Evaluations

Cross-border use of protected evidence requires secure custody, access, and compromise response.

Foundation 3: High-Stakes Capability Evaluation

Capability evidence should be comparable enough to support coordinated preparation without forcing one global threshold.

Foundation 4: Independent Expert Review

Reviewer access, independence, competence, and findings need portable profiles.

Foundation 5: Third-Party Auditor Ecosystem

Accreditation, proficiency, registries, and mutual recognition are central interoperability mechanisms.

Foundation 6: Progressive Standards and Requirements

Voluntary frameworks, standards, procurement, and law require crosswalks and recognition.

Foundation 7: Incentives and Prestige

International recognition can reward interoperable evidence, while prestige competition can fragment it.

---

37. Canonical Standards Body Positions

Standards Body adopts the following working positions.

1. Global interoperability is necessary because frontier AI systems, evidence, and deployment cross borders.
2. Interoperability should not be confused with global uniformity.
3. Shared evidence structure should generally precede attempts to impose shared policy outcomes.
4. Technical comparability does not require identical legal consequence.
5. Evidence recognition, competence recognition, process recognition, and legal recognition should remain distinct.
6. Every interoperable evaluation record should identify the model or system, configuration, protocol, evaluator, date, uncertainty, limitations, and status.
7. A model name alone is not adequate system identity.
8. Numeric scores should never travel without protocol and measurement context.
9. Protocol equivalence should be demonstrated for a stated purpose, not assumed.
10. Noncomparability is a legitimate and important result.
11. Shared terminology should include explicit mappings and disputed-term records.
12. Global metadata should use a stable common core with local extensions.
13. Evaluator recognition should be scope-specific and versioned.
14. Accreditation and mutual-recognition systems should be adapted carefully rather than copied mechanically.
15. Certification issued under one scheme should not automatically satisfy every jurisdiction.
16. Incident reporting should support public, trusted, and restricted disclosure layers.
17. International evidence sharing should be secure, purpose-limited, and auditable.
18. Interoperability should not compel unlawful or unsafe disclosure.
19. Legal crosswalks require jurisdiction-specific expertise.
20. Multilingual evaluation requires construct validation, not literal translation alone.
21. English-language performance should not define global AI capability by default.
22. Open-weight model lineage and configuration should be recorded through portable identity methods.

23. Capacity building is part of interoperability, not a separate charitable addition.
24. Developing countries and underrepresented regions should participate in technical design and governance, not only implementation.
25. Recognition networks should use plural trust anchors and avoid one global monopoly.
26. Interoperable registries should be portable, signed, and independently governable.
27. Recognition should expire or suspend after material changes or loss of confidence.
28. International cooperation should preserve national and institutional decision authority.
29. Shared minimums should not prevent stronger local safeguards.
30. Global interoperability should be evaluated by whether it improves evidence, reduces unnecessary duplication, broadens participation, and strengthens real decisions.

---

38. Decision Rules

A protocol should be considered interoperable when:

- Its construct is defined
- its identity and version are clear
- its administration and scoring are documented
- required metadata is available
- differences can be mapped
- uncertainty is preserved
- local extensions are declared
- result status is current

Two results should be treated as directly comparable only when:

- The evaluated system is materially equivalent
- constructs align
- task populations are comparable
- administration conditions align
- scoring and uncertainty align
- bridge evidence supports comparison

A result may receive conditional recognition when:

- Core evidence is credible
- local gaps are identifiable
- supplemental assessment can address those gaps
- recognition purpose is narrow
- status and conditions are public or reviewable

Recognition should be suspended when:

- Evaluator scope changes
- protocol validity is challenged
- system identity is uncertain
- security is compromised
- accreditation is suspended
- evidence expires
- serious complaints remain unresolved

A global common requirement should not be proposed when:

- Construct meaning is unresolved
- implementation capacity is highly unequal
- local rights or legal systems require materially different treatment
- the shared requirement would create a weak lowest common denominator
- evidence is too immature
- dominant actors control the process

A cross-border incident should be shared through a trusted channel when:

- It has material relevance beyond one jurisdiction
- disclosure is lawful
- the recipient has a legitimate role
- security controls are adequate
- onward use is governed
- public learning cannot be achieved safely through public disclosure alone

---

39. Interoperability Profile Template

A. Organization

- Name
- jurisdiction
- role
- contact
- governance

B. Supported Standards

- Standard
- version
- scope
- status

C. Supported Protocols

- Identifier
- version
- domain
- execution environment

D. Vocabulary

- Canonical terminology
- mappings
- local terms
- unresolved differences

E. Identity

- Model identifiers
- system manifests
- signatures
- attestation

F. Metadata

- Schema
- version
- required fields
- extensions

G. Evaluation

- Task formats
- scoring
- uncertainty
- bridge methods

H. Evaluator Recognition

- Accreditation
- scopes
- recognition partners
- conditions

I. Incident Exchange

- Taxonomy
- channels
- sensitivity
- notification

J. Security

- Access
- encryption
- retention
- onward disclosure
- incident response

K. Localization

- Languages
- legal modules
- cultural validation
- regional requirements

L. Recognition

- Evidence accepted
- conditions
- exclusions
- appeal
- expiry

---

40. Protocol Crosswalk Template

Protocol A:
Protocol B:
Purpose of mapping:
Date:
Reviewers:

Construct

Scope

Evaluated Object

Task Population

Sampling

Tools and Scaffolds

Administration

Scoring

Uncertainty

Baselines

Security

Reporting

Versioning

Thresholds

Local Modules

Compatibility Assessment

- Directly reproducible
- form-equivalent
- construct-compatible
- supporting evidence
- incompatible

Conditions

Evidence Needed

Expiration

---

41. Recognition Decision Template

Recognizing body:
External body or evidence:
Recognition object:
Purpose:
Date:

Identity

Scope

Competence

Method

Accreditation or Review

Security

Version

Local Requirements

Gaps

Supplemental Work

Conflicts

Decision

- No recognition
- informational use
- conditional technical recognition
- full technical recognition
- institutional recognition
- legal recognition

Conditions

Duration

Surveillance

Appeal

Suspension Triggers

---

42. International Incident Exchange Template

Incident identifier:
Reporting organization:
System:
Date:
Jurisdictions affected:

Category

Severity

Impact

Detection

Technical Summary

Capability Implication

Safeguard Implication

Evaluation Implication

Root Cause

Response

Current Status

Information Classification

- Public
- controlled
- confidential
- restricted
- highly restricted

Authorized Recipients

Onward Disclosure

Legal Constraints

Requested Action

Public Summary

Follow-Up

---

43. Translation and Localization Validation Template

Source protocol:
Target language or region:
Version:
Date:

Construct

Source Material

Translation Team

Domain Review

Cultural Adaptation

Legal Adaptation

Back Translation

Pilot Systems

Human Baselines

Difficulty Analysis

Bias Analysis

Comparability

Non-Equivalent Items

Local Additions

Limitations

Approval

Review Date

---

44. Registry Record Template

A. Record Identity

- Identifier
- type
- issuer
- date
- version
- signature

B. Subject

- System
- protocol
- evaluator
- certificate
- incident
- recognition

C. Scope

D. Status

- Current
- expired
- suspended
- corrected
- withdrawn
- superseded

E. Evidence

F. Recognition

G. Security Classification

H. Jurisdiction

I. Related Records

J. Correction History

K. Expiration

L. Public Summary

---

45. Global Interoperability Scorecard

Dimension	Core Question
Purpose	Is the intended cross-border use defined?
Vocabulary	Are terms shared or explicitly mapped?
Identity	Is the model, system, protocol, evaluator, and date identifiable?
Metadata	Is a common minimum record available?
Protocol	Can methods be reproduced, mapped, or bridged?
Measurement	Are scores, uncertainty, and baselines interpretable?
Comparability	Is equivalence demonstrated rather than assumed?
Noncomparability	Can the system clearly state when comparison fails?
Evaluator competence	Is scope-specific qualification legible?
Recognition	Are purpose, conditions, duration, and status explicit?
Accreditation	Are recognizing bodies competent and peer reviewed?
Certification	Are claims limited to scheme and scope?
Incidents	Can material incidents be classified and exchanged?
Security	Can sensitive evidence move under accountable controls?
Provenance	Can evidence history and modification be traced?
Localization	Are language, culture, law, and professional context addressed?
Legal mapping	Are technical and legal equivalence kept distinct?

Dimension	Core Question
Open-source fit	Can distributed model lineage and evidence be represented?
Capacity	Can smaller and underrepresented institutions participate?
Registry	Are records portable, current, and independently verifiable?
Governance	Are changes, conflicts, appeals, and dissent managed?
Resilience	Does the system avoid one trust anchor or registry monopoly?
Efficiency	Does interoperability reduce unnecessary duplication?
International utility	Can evidence support real cross-border decisions?
Adaptation	Can mappings and recognition evolve as AI changes?

---

46. Final Perspective

Global AI governance is often discussed as a choice between two extremes.

One extreme is fragmentation.

Every country, company, evaluator, and standards body creates its own:

- Vocabulary
- test
- threshold
- report
- certificate
- incident process
- registry
- legal interpretation

Evidence becomes difficult to combine.

Organizations repeat the same work.

Smaller countries and institutions become dependent on dominant actors.

Incidents fail to travel.

The other extreme is forced uniformity.

One framework, one registry, one evaluator network, one risk model, or one political bloc becomes the global default.

Local law, culture, language, professional context, and institutional legitimacy are treated as obstacles rather than sources of knowledge.

Neither extreme is sufficient.

Frontier AI requires shared evidence and plural authority.

The shared layer should make it possible to know:

- What system was evaluated
- how it was evaluated
- who performed the work
- what the result means
- how uncertain it is
- whether it is current
- what another institution has recognized
- which differences remain

The plural layer should preserve the right of institutions and communities to decide:

- What risk is acceptable
- which rights receive priority
- what safeguards are required
- how law applies
- who may deploy
- what remedies exist
- what information can be shared

Interoperability is the bridge.

It does not eliminate disagreement.

It makes disagreement more precise.

It does not eliminate duplication.

It distinguishes independent replication from administrative repetition.

It does not eliminate national authority.

It allows national authority to use evidence produced elsewhere without surrendering judgment.

It does not eliminate confidential information.

It creates controlled pathways for information that should not remain isolated.

It does not guarantee trust.

It provides the infrastructure through which trust can be earned, limited, reviewed, and withdrawn.

The eighth foundation of Standards Body is therefore portable evidence across plural institutions.

The future international system should not require every institution to speak with one voice.

It should make it possible for different voices to understand the same evidence.

References and Research Basis

[^nist-global]: National Institute of Standards and Technology, **A Plan for Global Engagement on AI Standards**, NIST AI 100-5, released 2024 and updated in 2025. https://www.nist.gov/publications/plan-global-engagement-ai-standards

[^nist-ai-standards]: National Institute of Standards and Technology, **AI Standards**. https://www.nist.gov/artificial-intelligence/ai-standards

[^nist-rmf]: National Institute of Standards and Technology, **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**, 2023. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

[^haip]: OECD, **Hiroshima AI Process Reporting Framework**. https://oecd.ai/en/hiroshima

[^haip-overview]: OECD, **HAIP Reporting Framework Overview**. https://oecd.ai/en/transparency/overview

[^haip-insights]: OECD, **How Are AI Developers Managing Risks? Insights from Responses to the Reporting Framework of the Hiroshima AI Process Code of Conduct**, 2025. https://oecd.ai/en/ai-publications/how-are-ai-developers-managing-risks-insights-from-responses-to-the-reporting-framework-of-the-hiroshima-ai-process-code-of-conduct

[^coe-convention]: Council of Europe, **Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law**. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

[^coe-treaty-status]: Council of Europe Treaty Office, **Chart of Signatures and Ratifications of Treaty 225**. https://www.coe.int/en/web/conventions/full-list/?module=signatures-by-treaty&treatynum=225

[^un-gdc]: United Nations, **Global Digital Compact**, 2024. https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf

[^un-gdc-site]: United Nations Office for Digital and Emerging Technologies, **Global Digital Compact**. https://www.un.org/digital-emerging-technologies/global-digital-compact

[^un-resolution]: United Nations General Assembly, **A/RES/78/265, Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development**, 2024. https://digitallibrary.un.org/record/4043244/files/A_RES_78_265-EN.pdf

[^un-ai-report]: United Nations Secretary-General's High-level Advisory Body on Artificial Intelligence, **Governing AI for Humanity: Final Report**, 2024. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

[^aisi-network]: UK AI Security Institute, **International Consensus and Open Questions in AI Evaluations**, February 12, 2026. https://www.aisi.gov.uk/blog/international-ai-network-consensus-and-open-questions

[^aisi]: UK AI Security Institute. https://www.aisi.gov.uk/

[^inspect]: UK AI Security Institute, **Inspect AI**. https://inspect.aisi.org.uk/

[^iso-ai]: International Organization for Standardization, **Artificial Intelligence Standards**. https://www.iso.org/sectors/it-technologies/ai

[^iso-sc42]: International Organization for Standardization, **ISO/IEC JTC 1/SC 42 Artificial Intelligence Catalogue**. https://www.iso.org/committee/6794475/x/catalogue/

[^iso-42001]: International Organization for Standardization, **ISO/IEC 42001:2023, Artificial Intelligence Management Systems**. https://www.iso.org/standard/42001

[^iso-42006]: International Organization for Standardization, **ISO/IEC 42006:2025, Requirements for Bodies Providing Audit and Certification of Artificial Intelligence Management Systems**. https://www.iso.org/standard/42006

[^iso-23894]: International Organization for Standardization, **ISO/IEC 23894:2023, Artificial Intelligence, Guidance on Risk Management**. https://www.iso.org/standard/77304.html

[^iso-42005]: International Organization for Standardization, **ISO/IEC 42005:2025, AI System Impact Assessment**. https://www.iso.org/sectors/it-technologies/ai

[^ilac-mra]: International Laboratory Accreditation Cooperation, **ILAC Mutual Recognition Arrangement and Signatories**. https://ilac.org/ilac-mra-and-signatories/

[^ilac-about]: International Laboratory Accreditation Cooperation, **About ILAC**. https://ilac.org/about-ilac/

[^ilac]: International Laboratory Accreditation Cooperation. https://ilac.org/

[^eu-ai-act]: European Union, **Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence**. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

[^eu-summary]: EUR-Lex, **Rules for Trustworthy Artificial Intelligence in the European Union**. https://eur-lex.europa.eu/EN/legal-content/summary/rules-for-trustworthy-artificial-intelligence-in-the-eu.html

[^oecd-principles]: OECD, **OECD AI Principles**. https://oecd.ai/en/ai-principles

[^oecd-observatory]: OECD.AI Policy Observatory. https://oecd.ai/

[^iso-casco]: International Organization for Standardization, **CASCO Conformity Assessment Toolbox**. https://casco.iso.org/

[^iso-recognition]: International Organization for Standardization, **Recognition of Conformity Assessment Bodies**. https://casco.iso.org/recognition-of-cabs.html

---

Revision Record

Version 1.0

Date: July 16, 2026

Change type: Complete foundational edition

Summary: Establishes the fully developed canonical working white paper for Foundation 8. Defines the global interoperability problem, shared and local layers, terminology, system identity, protocol and measurement interoperability, metadata, evaluator recognition, certification portability, incident exchange, secure evidence sharing, standards and legal crosswalks, multilingual evaluation, open-weight systems, capacity building, international institutions, recognition architecture, dispute resolution, governance, maturity, implementation, a Standards Body pilot, metrics, failure analysis, objections, evidence gaps, research agenda, standards implications, operational templates, scorecard, and current primary-source research basis.

Status: Ready for internal review and future expert critique.