Standards Body · Institutional-design proposal, public edition · Released July 17, 2026
Canonical record: https://standardsbody.ai/library/institutional-design/transparency-framework/

TRANSPARENCY_FRAMEWORK.md

Standards Body Transparency Framework

Project: Standards Body
Primary domain: standardsbody.ai
Core line: Foundations for Frontier AI
Document type: Canonical transparency, disclosure, confidentiality, public-record, correction, access, and accountability framework
Version: 1.0
Status: Approved foundational source
Document owner: Standards Body
Present institutional stage: Foundational research and institutional design
Applies to: Governance, research, standards development, evaluations, protocols, independent review, evaluator ecosystems, partnerships, funding, conflicts, incidents, registries, public communications, corrections, appeals, complaints, security, controlled evidence, institutional transitions, and future assurance activities
Related canonical sources: `PROJECT_IDENTITY.md`, `PROJECT_MANIFESTO.md`, `INSTITUTION_DESIGN.md`, `GOVERNANCE_FRAMEWORK.md`, `STANDARDS_DEVELOPMENT_PROCESS.md`, `FOUNDATIONS.md`, `FOUNDATIONS_APPENDIX.md`, `TERMINOLOGY.md`, `EVIDENCE_STANDARDS.md`, `RESEARCH_METHODOLOGY.md`, `TAXONOMY.md`, `EVALUATION_PHILOSOPHY.md`, and the eight foundation papers
Research basis reviewed through: July 16, 2026
Review cycle: Annual review, with event-triggered revision after a material change in legal status, institutional authority, information-security posture, public-reporting commitments, evaluation practices, incident experience, privacy obligations, standards processes, or international transparency expectations

---

Authority Note

This document defines the transparency practices Standards Body should apply to its own work and the disclosure architecture it proposes for a possible future institution.

It does not create:

- A legal right of access equivalent to a freedom-of-information statute
- A public-record obligation imposed by government law

- A regulatory disclosure mandate
- A legal privilege
- A confidentiality exemption from applicable law
- A right to receive protected model information
- A right to receive personal data
- A right to inspect every evaluation task
- A right to compel publication of dangerous information
- A certification of transparency
- An official transparency rating
- A universal reporting standard

Standards Body is not presently a government agency, regulator, accreditation body, certification body, statutory standards authority, or international organization.

Nothing in this framework authorizes Standards Body to disclose information it does not lawfully control.

Nothing in this framework overrides:

- Applicable law
- court orders
- contractual duties
- privacy rights
- security obligations
- intellectual-property rights
- responsible-disclosure duties
- legitimate national-security restrictions
- ethical duties to research participants

Transparency commitments should be interpreted together with proportional confidentiality, security, privacy, due process, and evidence integrity.

Document Purpose

This document establishes the complete transparency framework for Standards Body.

It defines:

- Why transparency is necessary
- What transparency can and cannot establish
- The presumption of disclosure
- The categories of information the institution should publish proactively
- The categories that may require controlled, confidential, restricted, or highly restricted treatment
- Public, controlled, and protected records
- Disclosure by audience and function
- Governance transparency

- funding and conflict transparency
- research transparency
- standards-process transparency
- evaluation transparency
- held-out and confidential evidence
- security and dangerous-information controls
- privacy and personal-data protection
- incident and failure transparency
- decision records
- appeals, complaints, and whistleblowing
- corrections, withdrawals, and supersession
- public registries
- source, version, and provenance records
- transparency requests
- redaction and aggregation
- publication timing
- accessibility, language, and machine readability
- partnership and external-influence disclosure
- automated and AI-assisted content disclosure
- institutional transparency reports
- transparency performance, audit, maturity, and failure recovery
- operational templates

The framework is designed for frontier AI, where transparency serves several different functions.

It can support:

- Accountability
- technical scrutiny
- reproducibility
- public understanding
- institutional legitimacy
- informed participation
- correction
- legal and policy oversight
- international interoperability
- incident learning

Transparency can also cause harm.

Improper disclosure can:

- Compromise held-out evaluations
- expose exploitable vulnerabilities
- reveal dangerous capability information
- disclose personal data
- violate research-participant rights
- create security risk

- breach contractual or legal obligations
- distort public understanding through partial evidence
- enable manipulation of standards or safeguards

The governing position is:

> **Standards Body should disclose enough information for others to understand, scrutinize, challenge, and correctly interpret its work, while protecting information whose disclosure would create material harm, violate rights, or undermine the integrity of the evidence.**

Executive Summary

Transparency is often treated as the opposite of secrecy.

That framing is too simple.

A credible frontier AI institution must distinguish among:

- Information that should always be public
- information that should be public after a delay
- information that should be available to qualified reviewers
- information that should be available only to defined participants
- information that should be aggregated or redacted
- information that should remain restricted because disclosure would create material harm

The central transparency question is not:

> Should this institution be transparent?

The better questions are:

- Transparent about what?
- To whom?
- At which level of detail?
- At which time?
- For which purpose?
- Under which safeguards?
- With which review and appeal rights?
- With which correction obligations?

Transparency Is Necessary but Not Sufficient

NIST describes accountability and transparency as related characteristics of trustworthy AI and notes that documentation can improve transparency, human review, and accountability.[^nist-rmf] [^nist-core]

Transparency does not itself establish:

- Accuracy
- safety
- fairness
- legality
- competence
- independence
- security
- validity
- public legitimacy

A weak process can be fully visible.

A harmful system can be well documented.

A false claim can be transparently repeated.

Transparency is valuable when it permits informed understanding, challenge, correction, and accountable action.

Presumption of Disclosure

Standards Body should adopt a presumption that institutional information is public unless a specific, documented reason justifies protection.

The public baseline should include:

- Mission and authority
- current institutional stage
- governing documents
- governing bodies
- appointment methods
- biographies and affiliations
- conflict disclosures
- funding sources and concentration
- work programs
- research methods
- standards procedures
- public-review drafts
- comment disposition
- decision summaries
- evaluation scope and conditions
- evidence levels
- limitations
- current status
- expiration
- corrections

- appeals
- incident summaries
- annual institutional performance

Transparency Classes

The framework uses five information classes.

Public

Information available without special authorization.

Controlled

Information available to defined participants, implementers, or reviewers under basic conditions.

Confidential

Information available only to authorized persons with a documented need and confidentiality duty.

Restricted

Sensitive information requiring enhanced access controls, logging, security review, and limited distribution.

Highly Restricted

Information whose unauthorized disclosure could create severe harm, compromise critical evaluation or security infrastructure, or violate exceptional legal duties.

Classification should not be permanent by default.

Each protected record should identify:

- Owner
- rationale
- access group
- handling rules
- review date
- release conditions
- appeal or challenge path

Transparency Layers

A frontier AI institution should provide several layers of transparency.

Institutional Transparency

Who the institution is, what authority it has, how it is governed, and who funds it.

Process Transparency

How research, standards, evaluations, reviews, and decisions are conducted.

Method Transparency

What methods, protocols, assumptions, and scoring rules are used.

Evidence Transparency

What evidence supports a claim, at which level, with which uncertainty and provenance.

Outcome Transparency

What was found, decided, corrected, withdrawn, or left unresolved.

Impact Transparency

How standards and evaluations affect organizations, people, markets, rights, security, and public institutions.

Status Transparency

Whether a document, result, certificate, protocol, or claim is draft, current, conditional, expired, superseded, suspended, withdrawn, or retired.

Functional Transparency

Different functions require different disclosure.

A standards process should disclose:

- Project need
- scope
- participants
- conflicts
- drafts
- comments
- objections
- consensus
- approval
- maintenance

A high-stakes evaluation should disclose:

- Evaluated object
- protocol
- evaluator
- conditions
- elicitation

- score or finding
- uncertainty
- evidence level
- review
- limitations
- status
- expiration

The exact held-out tasks may remain protected.

A governance decision should disclose:

- Decision owner
- authority
- evidence
- conflicts
- reasons
- dissent
- conditions
- appeal
- review date

A funding disclosure should identify:

- Source
- amount or range
- purpose
- restrictions
- control rights
- publication rights
- relationship to affected work

Transparency and Advanced AI Reporting

The OECD's Hiroshima AI Process Reporting Framework provides a voluntary international reporting structure for organizations across the advanced-AI value chain. Version 2.0 was launched in May 2026 and broadened participation, including attention to smaller organizations. [^haip-v2][^haip-overview]

Standards Body should learn from such reporting efforts while preserving several distinctions:

- Self-reporting is not independent verification.
- Disclosure of a policy is not evidence of effective implementation.
- A response may be complete in form and weak in evidence.
- Confidential review may be needed for sensitive claims.
- Version and status should remain visible.
- Public reporting should not become a marketing exercise.

Transparency and Legal Frameworks

The European Union AI Act contains context-specific transparency obligations and expressly distinguishes compliance with transparency obligations from broader lawfulness.[^eu-ai-act][^eu-summary]

The Council of Europe Framework Convention includes transparency and oversight among its lifecycle principles and connects AI governance to human rights, democracy, and the rule of law.[^coe-convention][^coe-principles]

These legal frameworks do not govern every Standards Body activity automatically.

They demonstrate that transparency should be:

- Contextual
- proportionate
- connected to oversight
- connected to affected persons
- distinct from full legal compliance

Transparency With Security

Transparency should not require publication of:

- Active cyber exploits
- detailed harmful biological procedures
- protected task banks
- vulnerabilities before remediation
- personal information
- trade secrets with no public-interest justification
- restricted model-access credentials
- operational security procedures whose disclosure increases risk

The correct response to sensitive evidence is not to eliminate accountability.

It is to provide:

- Public purpose
- public scope
- public governance
- public status
- public limitations
- qualified independent access
- protected review records
- release or declassification review

Corrections

Transparency includes the visibility of error.

A credible institution should publish:

- What changed
- why
- who authorized the change
- which conclusions were affected
- which records were updated
- whether the prior version remains usable
- whether a decision, standard, or result was withdrawn

Material changes should not occur silently.

Transparency Is a System

A transparency framework requires more than publishing documents.

It requires:

- Classification
- record ownership
- metadata
- versioning
- access controls
- request procedures
- redaction
- review
- audit
- correction
- retention
- public registries
- accessible communication

The final transparency principle is:

> **Make authority, process, evidence, uncertainty, status, and correction visible. Protect only what requires protection, explain the protection, and preserve qualified paths for review.**

---

1. Foundational Transparency Propositions

1.1 Transparency Serves Accountability

Disclosure should enable someone to understand, evaluate, challenge, or act.

1.2 Transparency Is Not Total Disclosure

Responsible transparency permits justified confidentiality.

1.3 Presumption of Public Access

Institutional governance and public-interest work should be public by default.

1.4 Protection Requires Reason

A protected classification should identify a specific harm, duty, or integrity need.

1.5 Transparency Is Audience-Specific

Different audiences may require different information and levels of access.

1.6 Transparency Is Function-Specific

Governance, research, standards, evaluation, assurance, incidents, and security require different disclosures.

1.7 Transparency Is Time-Specific

Information may move from restricted to public as risk changes.

1.8 Transparency Is Not Proof

Disclosure does not establish that a claim is true or a system is safe.

1.9 Status Is Essential

A record without current status can mislead even when its contents are accurate.

1.10 Limitations Must Travel With Claims

Public summaries should preserve material limitations and uncertainty.

1.11 Confidential Evidence Requires Independent Review

A public conclusion based on protected evidence should not rely only on the interested party's assertion.

1.12 Corrections Are Transparency

A correction system is part of disclosure, not an admission that transparency failed.

1.13 Funding Is Material Information

Financial dependence can affect judgment and should be disclosed.

1.14 Conflicts Are Material Information

Role-specific conflicts should be visible and governed.

1.15 Participation Should Be Visible

The institution should disclose who had influence and which perspectives were missing.

1.16 Public Records Should Be Usable

Publication without organization, accessibility, or current status is weak transparency.

1.17 Security Should Be Proportionate

Security should not become reputation management.

1.18 Privacy Should Be Protected

Public accountability does not require exposing personal data unnecessarily.

1.19 Automation Should Be Disclosed Where Material

AI-assisted institutional outputs should remain attributable to accountable humans and processes.

1.20 Transparency Should Be Evaluated

The institution should measure whether disclosure improves understanding and accountability.

---

2. Scope and Non-Claims

2.1 Functions Covered

This framework applies to:

- Institutional identity
- governance
- funding
- conflicts
- partnerships
- research
- standards development
- evaluations
- reviews
- assurance relationships
- registries
- incidents
- complaints
- appeals
- corrections
- security
- public communications
- institutional transitions

2.2 Information Covered

- Documents
- data
- code
- logs
- model outputs
- task banks
- decisions
- minutes
- reports
- evidence
- funding records
- contracts
- conflicts
- complaints
- personal data
- security information
- registry records
- public claims

2.3 Persons Covered

- Founders
- directors
- officers
- employees
- contractors
- fellows
- advisors
- council members
- working-group participants
- reviewers
- evaluators
- contributors
- funders
- partners
- members

2.4 External Information

Standards Body should not publish third-party information beyond its lawful rights and responsibilities.

2.5 Legal Access Rights

This framework creates internal commitments.

It does not create statutory access rights unless incorporated into a contract, charter, law, or formal policy that does so.

2.6 No Universal Disclosure Rule

Different jurisdictions and domains may impose different duties.

2.7 No Transparency-Washing

Publishing selected favorable information while concealing material limitations does not satisfy this framework.

---

3. Canonical Transparency Definitions

Definitions in `TERMINOLOGY.md` govern.

3.1 Transparency

The degree to which relevant information concerning an institution, system, process, method, evidence base, decision, or outcome is available and understandable to the appropriate audience.

3.2 Disclosure

The act of making information available to a defined audience.

3.3 Proactive Disclosure

Publication without a specific request.

3.4 Responsive Disclosure

Disclosure in response to a request, complaint, appeal, review, legal duty, or incident.

3.5 Public Record

A record available publicly without special authorization.

3.6 Controlled Record

A record available to an eligible group under defined conditions.

3.7 Confidential Record

A record available only to authorized persons with a defined need and duty.

3.8 Restricted Record

A sensitive record subject to enhanced security and access controls.

3.9 Highly Restricted Record

A record whose unauthorized disclosure could create severe harm or exceptional legal or security consequences.

3.10 Redaction

Removal or masking of information from a disclosed record.

3.11 Aggregation

Combination of information to reduce identification or security risk while preserving useful meaning.

3.12 Declassification

A decision to reduce the protection level of information.

3.13 Transparency Report

A periodic report describing disclosures, requests, restrictions, incidents, corrections, and institutional practices.

3.14 Provenance

The origin, custody, transformation, and source history of information.

3.15 Status Transparency

Visible indication of whether a record is draft, current, conditional, expired, superseded, suspended, withdrawn, retired, or archived.

3.16 Public Minimum

The minimum information that should remain public even when detailed evidence is protected.

3.17 Explainability

The degree to which relevant persons can understand reasons, factors, logic, or behavior at a level appropriate to the context.

3.18 Auditability

The ability to inspect evidence and reconstruct a process or decision.

3.19 Traceability

The ability to follow the lineage of a system, record, decision, requirement, or evidence object.

3.20 Transparency Debt

Accumulated missing, stale, inaccessible, or misleading disclosure that impairs accountability.

---

4. Transparency Objectives

4.1 Accountability

Transparency should allow responsible actors to be identified and held answerable.

4.2 Correct Interpretation

Disclosure should reduce the risk that:

- Drafts are mistaken for standards
- research is mistaken for certification
- a model result is mistaken for a deployment result
- an expired result is treated as current
- membership is mistaken for endorsement
- a protected review is mistaken for public verification

4.3 Technical Scrutiny

Qualified persons should be able to evaluate:

- Methods
- evidence
- assumptions
- limitations
- conflicts
- uncertainty
- reproducibility
- interoperability

4.4 Public Understanding

The public should be able to understand:

- What the institution does
- what authority it lacks
- what a result means
- what a standard requires
- what remains uncertain
- how to challenge an error

4.5 Informed Participation

Participants should receive enough information to contribute meaningfully before decisions are finalized.

4.6 Correction

Others should be able to identify and report errors.

4.7 Institutional Legitimacy

Transparency should support justified trust rather than reputation-based trust.

4.8 International Interoperability

Comparable disclosures should help institutions understand and reuse evidence across jurisdictions.

4.9 Incident Learning

Safe incident disclosure should support prevention and standards improvement.

4.10 Market Integrity

Transparency should reduce misleading claims concerning:

- Safety
- certification
- accreditation
- evaluation
- compliance
- standards status
- institutional relationships

---

5. Transparency Architecture

The framework contains six layers.

5.1 Identity Layer

Discloses:

- Mission
- status
- authority
- nonauthority
- legal form
- institutional stage
- contact

- governance documents

5.2 Governance Layer

Discloses:

- Governing bodies
- appointments
- terms
- affiliations
- conflicts
- delegations
- decisions
- dissent
- appeals

5.3 Work Layer

Discloses:

- Research programs
- standards work
- evaluation programs
- pilots
- registries
- partnerships
- funding

5.4 Evidence Layer

Discloses:

- Sources
- methods
- evidence levels
- confidence
- provenance
- limitations
- protected-evidence status

5.5 Outcome Layer

Discloses:

- Findings
- standards

- decisions
- corrections
- incidents
- withdrawals
- current status

5.6 Impact Layer

Discloses:

- Adoption
- implementation
- burden
- access
- distribution
- failures
- institutional performance

5.7 Layer Rule

A public summary should not jump directly from identity to favorable outcome while omitting process, evidence, and limitations.

---

6. Information Classification System

6.1 Public

Definition

Information suitable for unrestricted public access.

Examples

- Mission
- governance framework
- public standards
- public-review drafts
- annual reports
- public funding disclosures
- public result summaries
- corrections
- public incident notices

Controls

- Integrity
- version
- accessibility
- archival preservation

6.2 Controlled

Definition

Information available to defined eligible persons under basic access or use conditions.

Examples

- Implementation materials
- member discussion records
- controlled research data
- evaluator guidance
- nonpublic meeting materials
- draft task-generation methods

Controls

- Eligibility
- terms of use
- redistribution rule
- access record where appropriate

6.3 Confidential

Definition

Information whose disclosure could create material but bounded harm, violate legitimate duties, or undermine a defined process.

Examples

- Unpublished research
- confidential partner information
- personnel matters
- nonpublic complaint records
- draft legal analysis
- detailed model-provider evidence

Controls

- Need-to-know access

- confidentiality duty
- secure storage
- review date
- approved disclosure process

6.4 Restricted

Definition

Information requiring enhanced protection because disclosure could materially compromise security, evaluation integrity, safety, privacy, or legal duties.

Examples

- Active held-out tasks
- detailed system vulnerabilities
- sensitive incident evidence
- protected model-access information
- security architecture
- high-risk evaluation records

Controls

- Explicit authorization
- least privilege
- access logging
- enhanced security
- no uncontrolled export
- periodic access review
- release review

6.5 Highly Restricted

Definition

Information whose unauthorized disclosure could create severe harm or exceptional institutional, legal, or security consequences.

Examples

- Certain severe-risk capability details
- active critical infrastructure vulnerabilities
- uniquely identifying whistleblower records
- highly sensitive model or system credentials
- exceptional national-security material lawfully held

Controls

- Exceptional authorization
- compartmentalization
- named custodians
- strong technical and physical controls
- access monitoring
- emergency response
- mandatory review
- limited retention where possible

6.6 Classification Criteria

Consider:

- Harm from disclosure
- harm from nondisclosure
- legal duty
- privacy
- evaluation integrity
- public-interest need
- reversibility
- time sensitivity
- availability elsewhere
- independent-review need

6.7 Highest Necessary Protection

Use the lowest classification sufficient to address the legitimate risk.

6.8 Mixed Records

A mixed record should be:

- Segmented
- redacted
- summarized
- released in part

rather than withheld entirely where possible.

6.9 Classification Metadata

Every protected record should include:

- Classification

- owner
- authority
- rationale
- date
- permitted roles
- handling
- review date
- release conditions
- destruction or archive rule

6.10 Classification Review

Review after:

- Time period
- project completion
- incident resolution
- vulnerability remediation
- publication
- appeal
- legal change
- evidence compromise

6.11 Classification Dispute

A qualified person may challenge classification through a secure process.

---

7. Public Minimum

Even when detailed information is protected, the following should ordinarily remain public.

7.1 Existence

That the work, process, or decision exists.

7.2 Purpose

Why the work is being conducted.

7.3 Responsible Body

Which institutional body owns it.

7.4 Authority

Under which mandate it operates.

7.5 General Scope

What category of system, risk, standard, or evidence is involved.

7.6 Process Type

Whether the activity is:

- Research
- evaluation
- review
- standards development
- incident investigation
- assurance research

7.7 Access Model

That information is protected and which qualified roles can review it.

7.8 Status

Draft, active, complete, suspended, corrected, withdrawn, or archived.

7.9 Limitations

What the public output cannot establish.

7.10 Review

Whether independent review occurred and at what level.

7.11 Protection Rationale

A safe explanation of why more detail is not public.

7.12 Review Date

When the restriction will be reconsidered.

7.13 Exception

The public minimum itself may be narrowed only when acknowledging existence would create exceptional harm or violate law.

Such exceptions should receive high-level independent review.

---

8. Audience-Based Transparency

8.1 General Public

Needs:

- Plain-language meaning
- authority
- status
- limitations
- risks
- correction
- contact

8.2 Researchers

Needs:

- Methods
- evidence
- data
- code
- assumptions
- uncertainty
- reproducibility
- access process

8.3 Evaluators and Auditors

Needs:

- Protocol
- implementation rules
- evidence criteria
- decision rules
- security
- competence
- status

8.4 Developers and Deployers

Needs:

- Requirements
- evaluation conditions
- factual correction
- transition
- appeal
- implementation guidance

8.5 Affected Persons

Needs:

- Notice
- relevance
- effect
- rights
- explanation
- complaint
- remedy
- contact

8.6 Governments and Regulators

Needs:

- Technical scope
- evidence
- legal nonclaims
- current version
- implementation
- limitations
- recognition
- secure access where lawful

8.7 Funders and Partners

Needs:

- Mission
- governance
- use of funds
- conflicts
- project status

- outcomes

They should not receive privileged control of conclusions.

8.8 Qualified Restricted Reviewers

Need:

- Sufficient evidence
- system access
- methods
- limitations
- ability to dissent
- publication rights or review record

8.9 Audience Conflict

Where audiences have competing needs, use layered disclosure.

---

9. Proactive Disclosure

9.1 Baseline

Standards Body should publish important institutional information before a request is made.

9.2 Proactive Disclosure Categories

- Identity
- governance
- funding
- conflicts
- work programs
- project registers
- standards drafts
- public reviews
- approved standards
- research publications
- evaluation summaries
- decision records
- appeals
- corrections
- incident summaries
- annual reports
- version history

9.3 Timeliness

Proactive disclosure should occur close enough to the event to support participation and accountability.

9.4 Currentness

Pages should display:

- Last updated
- current version
- review date
- owner
- status

9.5 Archive

Superseded information should remain available and clearly labeled.

9.6 Discoverability

Important disclosures should not be hidden in inaccessible repositories or unindexed pages.

9.7 Machine Readability

High-value records should be available in structured formats where useful.

---

10. Responsive Disclosure and Transparency Requests

10.1 Purpose

A transparency-request process allows persons to request records or clarification not already public.

10.2 No Statutory Claim

The process is voluntary unless a legal or contractual right applies.

10.3 Eligible Requests

Requests may concern:

- Governance

- funding
- standards process
- research methods
- evaluation status
- conflicts
- corrections
- registries
- institutional decisions

10.4 Request Information

Requesters should identify:

- Record or topic
- date range
- preferred format
- public-interest reason where relevant
- urgency
- accessibility need

10.5 Intake

The institution should:

1. acknowledge;
2. clarify if needed;
3. search;
4. classify;
5. consult record owners;
6. decide;
7. release, redact, summarize, or deny;
8. provide reasons;
9. offer review.

10.6 Response Targets

Suggested internal targets:

- Acknowledgment within five business days
- ordinary decision within twenty business days
- complex request status update within twenty business days
- expedited review where serious safety or rights concerns exist

These are service targets, not statutory deadlines.

10.7 Fees

Routine digital requests should ordinarily be free.

Exceptional costs may justify a disclosed fee or narrowed request.

10.8 Denial Grounds

Possible grounds:

- No record
- information not controlled
- personal privacy
- security
- evaluation integrity
- legal restriction
- active investigation
- contractual duty
- excessive burden
- bad-faith abuse

10.9 Partial Release

Use redaction or summary before full denial.

10.10 Review

A requester may seek internal review of a denial.

10.11 Request Log

Publish aggregate request statistics and significant decisions.

---

11. Transparency Decision Test

Before publishing or withholding, ask:

11.1 Relevance

Would disclosure support accountability, understanding, participation, correction, or safety?

11.2 Harm

Could disclosure create a material and plausible harm?

11.3 Rights

Would disclosure violate privacy or other rights?

11.4 Integrity

Would disclosure undermine evaluation or research validity?

11.5 Alternatives

Can the information be:

- Redacted
- delayed
- summarized
- aggregated
- reviewed confidentially
- released to a qualified group

11.6 Authority

Who may decide?

11.7 Duration

How long should protection last?

11.8 Review

Can the decision be challenged?

11.9 Record

Is the rationale documented?

11.10 Public Minimum

What should remain visible?

---

12. Governance Transparency

12.1 Governing Documents

Publish:

- Charter or articles
- bylaws
- governance framework
- committee charters
- delegations policy
- conflict policy
- appeals policy
- transparency policy

12.2 Governing Bodies

Publish:

- Body name
- mandate
- authority
- limits
- membership
- chair
- terms
- appointment
- quorum
- reporting
- appeal

12.3 Members

Publish for directors and major council members:

- Biography
- affiliation
- role
- term
- competence
- independence status
- material conflict summary

12.4 Appointment

Publish:

- Selection criteria
- nomination route
- appointing authority
- term
- constituency classification
- decision

12.5 Meetings

Publish:

- Calendar
- agendas where appropriate
- safe summaries
- decisions
- recusals
- dissent
- action items

12.6 Closed Sessions

State:

- That a closed session occurred
- general category
- authority
- public outcome where possible

12.7 Delegations

Publish material delegations of authority.

12.8 Board Votes

Publish vote totals for:

- Mission changes
- institutional-stage changes
- mergers
- dissolution
- creation of certification or accreditation functions
- major governance changes

12.9 Dissent

Material board and council dissent should be preserved where safe.

---

13. Funding Transparency

13.1 Why Funding Matters

Funding can shape:

- Agenda
- access
- priorities
- staffing
- publication
- partnerships
- institutional survival

13.2 Public Funding Record

Publish:

- Funder name or lawful category
- amount or range
- period
- purpose
- restrictions
- control rights
- publication rights
- related projects
- related-party status

13.3 Amount Disclosure

Preferred:

- Exact amount for institutional grants and major gifts

Where exact disclosure is not possible:

- Use a meaningful range
- explain the limitation

13.4 In-Kind Support

Disclose material:

- Compute
- model access
- data
- personnel
- travel
- legal support
- facilities
- security services

13.5 Funding Concentration

Publish:

- Largest funder share
- top five share
- restricted versus unrestricted revenue
- commercial service share
- membership share

13.6 Funding Conditions

Publish whether the funder may:

- Review drafts
- receive early access
- delay publication
- appoint participants
- approve expenses
- use branding

A funder should not control findings.

13.7 Anonymous Funding

Disclose:

- Amount
- purpose
- due-diligence process
- board approval
- reason identity is not public

Large anonymous funding should be exceptional.

13.8 Donor Privacy

Personal donor privacy may justify limited disclosure for small gifts.

13.9 Annual Financial Transparency

Publish:

- Audited statements when available
- revenue categories
- expenses
- reserves
- compensation principles
- related-party transactions
- material liabilities

---

14. Conflict Transparency

14.1 Public Conflict Register

Publish role-relevant conflict summaries for:

- Directors
- officers
- council chairs
- standards chairs
- lead reviewers
- lead evaluators
- decision makers

14.2 Scope

The register should identify:

- Employment
- board roles
- ownership
- consulting
- clients
- funding
- intellectual property
- recent affiliations
- matter-specific relationships

14.3 Privacy Balance

Do not publish unnecessary personal financial details.

14.4 Recusal Record

Publish material recusals with:

- Matter
- person
- reason category
- role excluded

14.5 Conflict Decision

State whether the response was:

- Disclosure
- role limit
- recusal
- independent review
- replacement
- rejection of relationship

14.6 Intellectual Conflict

Disclose when a reviewer:

- Authored the method
- owns relevant intellectual property
- publicly committed to the conclusion
- depends professionally on adoption

14.7 Funding Conflict

Reports should identify project-specific funding and publication control.

---

15. Partnership Transparency

15.1 Material Partnership Summary

Publish:

- Partner
- purpose

- duration
- funding
- governance
- data
- intellectual property
- publication
- branding
- termination
- authority limits

15.2 No Implied Endorsement

The summary should state that partnership does not imply approval of all partner activities.

15.3 Government Relationship

Identify whether the relationship is:

- Informal
- advisory
- research
- contractual
- recognized
- delegated
- regulatory

15.4 Model-Provider Relationship

Disclose:

- Access provided
- restrictions
- evaluation independence
- publication rights
- factual review
- funding
- conflicts

15.5 Joint Work

Identify which institution:

- Owns the output
- approved it
- maintains it
- hears appeals

- controls publication

15.6 Termination

Publish termination when material to public reliance.

---

16. Research Transparency

16.1 Research Register

Every material research project should have a register entry containing:

- Project identifier
- title
- purpose
- question
- responsible lead
- sponsor
- method
- consequence level
- sensitivity
- review level
- status
- start date
- expected output
- correction channel

16.2 Research Protocol

Publish the protocol or a safe summary before results when feasible.

16.3 Registration

Confirmatory and high-consequence work should use prospective registration where appropriate.

16.4 Methods

Disclose:

- Design
- data source
- sampling
- measures
- analysis

- exclusions
- deviations
- uncertainty
- limitations

16.5 Data

Data access should be classified as:

- Open
- controlled
- confidential
- restricted
- unavailable

State why.

16.6 Code and Artifacts

Publish code, prompts, schemas, and analysis artifacts when safe and legally permitted.

16.7 Research Participants

Protect:

- Consent
- privacy
- safety
- confidentiality
- vulnerable populations

16.8 Negative and Null Findings

Do not suppress findings because they are unfavorable, inconclusive, or inconsistent with institutional expectations.

16.9 Sponsor Role

Publish:

- Research funding
- question influence
- method influence
- access
- review rights
- publication rights

16.10 Research Deviations

Disclose material departures from the planned method.

16.11 Peer and Independent Review

State:

- Review type
- reviewer selection
- access
- conflicts
- findings
- unresolved dissent

16.12 Retraction and Withdrawal

Research status should be updated visibly.

16.13 Responsible Publication

Sensitive research may require:

- Redaction
- staged release
- qualified access
- safe abstraction
- delayed publication
- nonpublication of dangerous details

The decision should be reviewed and documented.

---

17. Standards-Process Transparency

17.1 Work Program

Publish:

- Proposed work
- active projects
- paused projects
- discontinued projects
- maintenance projects
- responsible groups
- timelines

- participation opportunities

17.2 New Work

Publish the proposal, need, scope, evidence maturity, and decision.

17.3 Charter

Publish:

- Purpose
- scope
- non-scope
- deliverables
- membership categories
- decision rules
- security
- milestones
- appeal

17.4 Participants

Publish:

- Member names or authorized representatives
- affiliations
- interest categories
- leadership
- editor
- liaisons
- conflict summaries

17.5 Balance

Publish a balance assessment and identified gaps.

17.6 Drafts

Public drafts should display:

- Version
- status
- date
- changes
- open issues

- nonapproval notice

17.7 Public Review

Publish:

- Notice
- review period
- draft
- supporting evidence
- comment route
- accessibility
- translation
- next stage

17.8 Comments

Publish comments and responses where lawful and safe.

17.9 Protected Comments

State the number and categories of protected submissions.

17.10 Objections

Publish substantial objections, responses, and unresolved dissent.

17.11 Consensus

Publish the consensus report.

17.12 Voting

Publish:

- Eligible-voter rules
- quorum
- totals
- affiliation controls
- invalid or recused votes

17.13 Approval

Publish the decision, authority, conditions, appeal, effective date, and maintenance owner.

17.14 Maintenance

Publish:

- Interpretations
- errata
- amendments
- reviews
- supersession
- withdrawal

17.15 Standards Access

Core Standards Body standards should be freely readable.

---

18. Evaluation Transparency

18.1 Evaluation Register

Each material evaluation should have a record containing:

- Evaluation identifier
- evaluated object
- protocol
- evaluator
- sponsor
- date
- lifecycle stage
- consequence level
- integrity status
- result status
- expiration
- public and protected components

18.2 Object Identity

Disclose enough to identify:

- Model
- system
- version
- access date
- tools
- scaffolds
- safeguards

- deployment context

18.3 Purpose

State:

- Claim assessed
- decision supported
- invalid interpretations

18.4 Protocol

Publish:

- Protocol identifier and version
- construct
- task domains
- administration
- elicitation
- scoring
- uncertainty
- change control

Exact tasks may remain protected.

18.5 Elicitation

Disclose:

- Prompting regime
- tools
- retries
- examples
- fine-tuning
- human effort
- time
- compute
- developer support

at the level necessary for interpretation.

18.6 Results

Disclose:

- Scores or findings
- distribution

- failure modes
- uncertainty
- invalid runs
- limitations
- evidence level
- confidence

18.7 Negative Results

Use bounded language.

Preferred:

> The capability was not demonstrated under the assessed conditions.

18.8 Reviewer Information

Disclose:

- Review type
- reviewer competence
- independence
- access
- dissent

18.9 Sponsor and Provider Influence

Disclose:

- Funding
- system access
- method control
- factual review
- publication rights
- restrictions

18.10 Integrity Status

Use:

- Unexposed
- exposure unknown
- partially exposed
- contaminated
- compromised
- retired

18.11 Expiration

Every consequential result should show:

- Effective date
- valid-through date
- re-evaluation triggers
- current status

18.12 Protected Evidence

Where tasks or evidence are protected, publish the public minimum and independent-review architecture.

18.13 Public Claims

The public summary should not exceed the evidence.

18.14 Evaluation Incident

Disclose material:

- Scoring error
- system mismatch
- task compromise
- excluded runs
- access limitation
- withdrawal

18.15 Comparative Results

State whether protocols and conditions are comparable.

---

19. Transparency for Held-Out Evaluations

19.1 Protected Content

Possible protected elements:

- Exact tasks
- solutions
- task-selection rules
- scoring details
- environment configuration

- attack methods
- compromise indicators

19.2 Public Protocol Layer

Publish:

- Construct
- task universe
- administration class
- scoring class
- evaluator competence
- integrity controls
- uncertainty
- status

19.3 Independent Access

Qualified reviewers should receive enough protected access to evaluate:

- Task quality
- construct coverage
- scoring
- administration
- integrity
- conclusion

19.4 Chain of Custody

Maintain controlled records of:

- Task creation
- access
- administration
- transfer
- scoring
- storage
- retirement

19.5 Exposure Disclosure

Publicly state material exposure or compromise promptly.

19.6 Rotation

Publish task-rotation policy without revealing current content.

19.7 Post-Retirement Release

Consider releasing retired tasks when:

- Security risk is acceptable
- privacy and rights permit
- future task-generation methods are not compromised
- research value remains

19.8 No Black-Box Authority

A conclusion based entirely on secret evidence with no qualified independent review should not support a high-consequence public claim.

---

20. Independent Review Transparency

20.1 Review Register

Publish:

- Review identifier
- mandate
- subject
- reviewers
- competence
- independence profile
- conflicts
- access level
- date
- status

20.2 Review Scope

State what the review did and did not examine.

20.3 Access Sufficiency

Disclose whether reviewers considered access sufficient.

20.4 Reviewer Selection

Disclose who selected reviewers and under which criteria.

20.5 Funding

Disclose who paid reviewers.

20.6 Right of Reply

State whether the reviewed party had factual-response rights.

20.7 Publication Control

State who controlled publication.

20.8 Findings

Publish findings, limitations, and recommendations where safe.

20.9 Dissent

Publish minority findings or state that protected dissent exists.

20.10 Review Status

Distinguish:

- Advisory review
- technical review
- methodological review
- assurance review
- audit
- certification
- accreditation

20.11 No Review-Washing

Do not describe a limited factual check as independent expert review.

---

21. Evaluator and Assurance Transparency

21.1 Evaluator Profile

A public evaluator record should identify:

- Legal identity
- ownership
- governance
- services
- competence scopes
- methods
- quality system
- conflicts
- security
- accreditation or recognition status
- complaints and status

21.2 Scope

State exact:

- Domain
- protocol
- system type
- activity
- security level
- jurisdiction
- status

21.3 Client Dependence

Disclose material client concentration in aggregate where confidentiality prevents exact detail.

21.4 Service Bundling

Disclose whether the evaluator also provides:

- Consulting
- implementation
- model development
- certification
- training
- standards participation

21.5 Proficiency

Publish proficiency status and limitations.

21.6 Certification

A certificate record should identify:

- Scheme
- certification body
- certified object
- scope
- effective date
- expiration
- surveillance
- status
- exclusions

21.7 Accreditation

An accreditation record should identify:

- Accreditation body
- criteria
- scope
- locations
- status
- expiration
- suspension

21.8 Recognition

Standards Body recognition should identify:

- Purpose
- evidence
- scope
- conditions
- nonaccreditation status where applicable
- expiration
- appeal

21.9 Complaints

Publish aggregate complaint and disciplinary information without compromising fairness or privacy.

21.10 No Badge Ambiguity

Every mark should link to a registry explaining its meaning.

---

22. Decision Transparency

22.1 Material Decision Record

A material decision should identify:

- Decision identifier
- question
- owner
- authority
- evidence standard
- evidence
- participants
- conflicts
- options
- decision
- reasons
- conditions
- dissent
- appeal
- review date
- status

22.2 Reasons

Reasons should be sufficiently specific to permit understanding and challenge.

22.3 Protected Reasons

Where detail is protected, publish a safe summary and provide qualified review.

22.4 Decision Alternatives

State material alternatives considered.

22.5 Uncertainty

State the uncertainty that remained.

22.6 Dissent

Preserve material dissent.

22.7 Implementation

State who is responsible for action.

22.8 Review

State whether the decision is:

- Permanent
- time-limited
- provisional
- subject to monitoring
- automatically reviewable

22.9 Decision Registry

Maintain a searchable registry of material public decisions.

---

23. Protocol and Standard Status Transparency

23.1 Status Labels

Use controlled statuses:

- Draft
- proposed
- public review
- provisional
- approved
- active
- conditional
- suspended
- corrected
- superseded
- withdrawn
- retired
- archived

23.2 Status Placement

Display status:

- On the document
- on the website page
- in the registry
- in machine-readable metadata
- in public citations where feasible

23.3 Effective and Expiration Dates

Display:

- Publication
- effective date
- expiration
- review
- supersession

23.4 Superseded Access

Keep superseded versions available with a warning.

23.5 Status Propagation

A status change should update dependent:

- Website pages
- registries
- claims
- certificates
- evaluation results
- partner materials

23.6 No Orphaned Documents

A file without a current owner or status should not remain presented as canonical.

---

24. Incident Transparency

24.1 Incident Categories

- AI safety

- AI security
- privacy
- evaluation
- governance
- research
- standards
- assurance
- data
- institutional operations

24.2 Immediate Priorities

Immediate response should prioritize:

- Safety
- containment
- evidence preservation
- legal duties
- affected persons
- responsible disclosure

24.3 Public Notice Trigger

Public notice may be required when:

- Public reliance is affected
- a result or standard is invalid
- personal data is affected
- an active safety risk exists
- registry status changes
- institutional integrity is materially impaired

24.4 Initial Notice

State:

- What is known
- what is not known
- affected scope
- current action
- public implications
- next update

24.5 Investigation Update

Avoid premature certainty.

24.6 Final Report

Where feasible, publish:

- Timeline
- system
- failure mode
- root cause
- contributing factors
- harm
- response
- corrective action
- standards and protocol effects
- remaining uncertainty

24.7 Protected Details

Redact:

- Exploitable details
- personal data
- confidential evidence
- active investigation material

24.8 Near Miss

Publish safe learning from near misses.

24.9 Incident Status

Use:

- Reported
- triaged
- investigating
- confirmed
- contained
- remediated
- closed
- reopened

24.10 Incident Correction

Update prior statements visibly when facts change.

---

25. Complaint and Appeal Transparency

25.1 Process Publication

Publish:

- Eligibility
- filing method
- grounds
- timelines
- confidentiality
- decision authority
- appeal
- nonretaliation

25.2 Case Transparency

For material cases, publish:

- Category
- process
- outcome
- reasoning
- corrective action
- appeal

subject to privacy and fairness.

25.3 Aggregate Reporting

Publish:

- Number
- type
- response time
- substantiation
- outcomes
- appeals
- recurrence

25.4 Anonymous and Confidential Cases

Protect identity while reporting systemic lessons.

25.5 No Retaliation

Report retaliation findings and corrective action where lawful.

25.6 Appeal Decisions

Publish reasoned appeal decisions or safe summaries.

25.7 Accessibility

Complaint and appeal procedures should be understandable and usable.

---

26. Correction and Withdrawal Transparency

26.1 Correction Principle

The institution should make the history of material change visible.

26.2 Correction Notice

State:

- Record
- prior version
- corrected version
- error
- change
- date
- effect
- responsible body
- dependent records

26.3 Correction Types

- Editorial
- factual
- methodological
- statistical
- status
- security
- evidence level
- confidence
- governance

- withdrawal

26.4 Material Correction

A material correction should be linked prominently from the original.

26.5 Withdrawal

State:

- Reason
- effective date
- replacement
- effect on prior claims
- archive
- appeal

26.6 No Silent Deletion

Do not silently remove a consequential public record.

26.7 Search and Indexing

Corrections should be discoverable in search and registries.

26.8 Downstream Notification

Notify known affected:

- Partners
- adopters
- evaluators
- authorities
- media
- citation users

where feasible.

---

27. Source and Evidence Transparency

27.1 Source Register

Maintain a source record containing:

- Identifier
- title
- author or institution
- date
- version
- type
- access
- relevance
- status
- correction
- link

27.2 Primary Sources

Prefer primary sources for:

- Law
- regulation
- standards
- official policy
- technical claims
- organizational status

27.3 Secondary Sources

Use secondary sources for context, synthesis, and critique.

27.4 Personal Communications

Disclose:

- Role
- date
- purpose
- confidentiality

without exposing identity unnecessarily.

27.5 Source Limitations

State when a source is:

- Self-reported
- unaudited
- incomplete
- outdated
- anonymous
- inaccessible
- translated
- disputed

27.6 Evidence Level

Public claims should display or link to the applicable evidence level.

27.7 Contrary Evidence

Disclose material contrary evidence.

27.8 Source Updates

Update records when sources are corrected, withdrawn, or superseded.

27.9 Citation Integrity

Citations should support the precise claim made.

---

28. Version and Provenance Transparency

28.1 Version Metadata

Every canonical file should display:

- Version
- status
- owner
- publication date
- current-through date
- review date
- superseded version

28.2 Revision Record

State:

- What changed
- why
- approval
- date
- effect

28.3 Authorship

Disclose:

- Responsible institutional author
- contributors
- reviewers
- AI assistance where material
- approval authority

28.4 Provenance Chain

Preserve:

- Source
- creation
- edits
- review
- approval
- publication
- correction
- archive

28.5 Repository

Use controlled repositories and immutable release records where feasible.

28.6 Machine-Readable Metadata

Publish metadata for:

- Identifier
- version
- status
- relationships
- sources

- license
- access
- corrections

28.7 Forks and Derivatives

A derivative should identify its source and changes.

28.8 Canonical Location

Each public record should identify the canonical location.

---

29. Privacy and Personal Data

29.1 Privacy Principle

Transparency should not expose personal information beyond what accountability requires.

29.2 Personal Data Categories

- Identity
- contact
- employment
- financial
- health
- biometric
- location
- communications
- complaints
- research participation
- security logs

29.3 Data Minimization

Collect and publish only what is necessary.

29.4 Public Professional Information

Public roles may justify disclosure of:

- Name
- institutional affiliation
- role

- term
- professional biography
- material conflict summary

29.5 Private Information

Do not publish:

- Home address
- personal telephone number
- personal financial account
- unnecessary family information
- protected health information
- security credentials
- confidential complaint details

29.6 Consent

Consent may support disclosure.

It should not be the sole basis where power imbalance makes consent weak.

29.7 Research Participants

Use:

- De-identification
- aggregation
- controlled access
- ethics review
- retention limits

29.8 Complaint and Whistleblower Records

Protect identities unless:

- The person consents
- disclosure is legally required
- due process requires limited disclosure
- public interest clearly outweighs harm under qualified review

29.9 Re-Identification Risk

Assess whether combined public records can identify protected persons.

29.10 Data Subject Requests

Where applicable, provide pathways for:

- Access
- correction
- deletion
- restriction
- objection

subject to law and evidence-preservation duties.

29.11 Privacy Incident

A privacy incident should trigger:

- Containment
- legal review
- affected-person response
- public notice where required
- corrective action

---

30. Security and Dangerous-Information Transparency

30.1 Security Principle

Transparency should not materially increase the ability to cause harm.

30.2 Dangerous-Information Categories

- Exploit details
- bypass methods
- severe-risk capability procedures
- critical infrastructure weaknesses
- active safeguard vulnerabilities
- sensitive operational locations
- access credentials
- protected task-generation methods

30.3 Security Review

Before publication, assess:

- Harm capability
- actor access

- novelty
- reproducibility
- target exposure
- mitigation
- public-interest value
- existing availability

30.4 Publication Options

- Full publication
- redacted publication
- high-level summary
- delayed publication
- qualified-access release
- responsible disclosure to affected parties
- nonpublication

30.5 Responsible Disclosure

Coordinate with affected organizations where disclosure concerns remediable vulnerabilities.

30.6 No Indefinite Vendor Veto

An affected organization should not control publication indefinitely.

30.7 Embargo

An embargo should have:

- Purpose
- duration
- remediation milestones
- review
- expiration
- escalation

30.8 Independent Security Review

High-consequence restrictions should receive independent review.

30.9 Public Security Statement

State:

- That details are withheld

- general reason
- review process
- status
- next review

30.10 Security Through Obscurity

Secrecy should not substitute for sound security design.

30.11 Release After Remediation

Reassess publication when the vulnerability is mitigated.

---

31. Redaction

31.1 Redaction Principle

Redaction should remove only what requires protection.

31.2 Redaction Categories

- Personal data
- security
- trade secret
- legal privilege
- active investigation
- confidential source
- held-out content
- protected model information

31.3 Redaction Marking

A redacted document should indicate:

- That material was removed
- category
- approximate location
- authority
- date

unless doing so would itself create harm.

31.4 Redaction Review

A second qualified person should review material redactions.

31.5 Technical Redaction

Redaction should remove underlying data, metadata, comments, layers, and hidden text.

31.6 No Misleading Redaction

Remaining text should not create a false interpretation.

31.7 Segregability

Release nonprotected portions.

31.8 Redaction Appeal

Material redactions may be challenged through the transparency-request process.

---

32. Aggregation and Statistical Disclosure

32.1 Aggregation Purpose

Aggregation can disclose patterns while reducing privacy, security, or confidentiality risk.

32.2 Aggregate Uses

- Complaints
- funding
- client concentration
- incidents
- evaluator performance
- participation
- access requests
- security events
- demographic data

32.3 Small Cell Risk

Suppress or combine small categories where identification is likely.

32.4 Meaning Preservation

Aggregation should not conceal:

- Severe outlier harm
- concentrated funding
- systematic discrimination
- repeat misconduct
- one dominant client

32.5 Ranges

Use ranges where exact values create unnecessary risk.

32.6 Method

Disclose aggregation method sufficiently for interpretation.

32.7 Differential Privacy and Advanced Methods

Advanced privacy-preserving methods may be used where valid and proportionate.

Their effect on accuracy should be disclosed.

32.8 Aggregate Correction

Correct aggregate reports when underlying data changes materially.

---

33. Timing and Publication Sequence

33.1 Timeliness

Information should be disclosed while it remains useful.

33.2 Premature Disclosure

Premature disclosure can:

- Distort incomplete findings
- compromise investigation
- expose vulnerabilities
- prejudice due process
- invalidate evaluation

33.3 Delayed Disclosure

Delay should identify:

- Reason
- authority
- expected release
- review date

33.4 Pre-Decision Transparency

Publish enough before a decision to permit meaningful participation.

33.5 Decision Transparency

Publish the outcome promptly after approval.

33.6 Post-Implementation Transparency

Publish implementation and impact evidence.

33.7 Incident Updates

Use staged updates as facts develop.

33.8 Quiet Corrections

Do not delay correction merely to coordinate favorable communications.

33.9 Market-Sensitive Information

Handle lawfully and avoid selective disclosure.

---

34. Accessibility and Understandability

34.1 Accessibility Principle

Information is not meaningfully transparent if intended audiences cannot access or understand it.

34.2 Formats

Provide:

- Accessible HTML
- structured text
- downloadable files
- machine-readable data where useful
- captions and transcripts
- readable tables

34.3 Disability Access

Follow applicable accessibility standards and test with users.

34.4 Plain Language

Provide plain-language summaries for consequential public outputs.

34.5 Technical Detail

Plain language should supplement, not replace, technical evidence.

34.6 Layered Documents

Use:

- Summary
- detailed report
- methods
- data and code
- protected annex

34.7 Jargon

Define necessary technical and institutional terms.

34.8 Visual Communication

Charts should include:

- Source
- scale
- uncertainty
- units

- accessible descriptions
- current status

34.9 Mobile and Low-Bandwidth Access

Important public information should remain usable on common devices and limited connections.

34.10 Accessibility Feedback

Provide a channel to report barriers.

---

35. Language, Translation, and Cultural Context

35.1 Language Transparency

State the original and controlling language.

35.2 Translation Status

Use:

- Official
- approved
- informative
- community
- machine-assisted
- draft

35.3 Technical Review

Normative and technical translations should receive domain review.

35.4 Translation Differences

Publish known non-equivalence.

35.5 Regional Context

Explain where:

- Law
- institutions
- examples

- risk assumptions
- public expectations

differ by region.

35.6 Multilingual Participation

International projects should provide meaningful multilingual pathways as resources permit.

35.7 Machine Translation

Disclose material machine translation and human review.

35.8 No English-Only Universal Claim

A process should not claim broad international legitimacy while practical access remains limited to one language.

---

36. Website and Digital Transparency

36.1 Source of Truth

`WEBSITE_SOURCE_OF_TRUTH.md` should govern public website facts and claims.

36.2 Required Pages

The website should include:

- About
- mission and status
- authority note
- governance
- funding
- conflicts
- work program
- standards
- research
- evaluations
- registries
- corrections
- complaints and appeals
- contact
- privacy and security

36.3 Page Metadata

Display:

- Owner
- last updated
- version
- status
- review date

36.4 Broken and Stale Links

Monitor and repair.

36.5 Search

Provide searchable access to standards, decisions, corrections, and registries.

36.6 Archive

Maintain access to superseded public records.

36.7 Analytics

Disclose material use of tracking and analytics.

36.8 Cookies and Privacy

Use proportionate data collection.

36.9 Website Claims

Review claims concerning:

- Authority
- independence
- international status
- partnerships
- standards
- certification
- accreditation
- safety

36.10 Contact Identity

State whether communications come from the project, a staff member, a working group, or an independent panel.

---

37. AI-Assisted and Automated Content Transparency

37.1 Accountability Principle

Human and institutional accountability remains even when AI assists.

37.2 Disclosure Trigger

Disclose AI assistance when material to:

- Authorship
- analysis
- translation
- evidence extraction
- scoring
- public communication
- decision support
- code generation

37.3 Materiality

Routine spelling or formatting assistance may not require item-level disclosure.

Material analytical or drafting assistance should be recorded internally and disclosed appropriately.

37.4 Human Review

AI-generated content should receive qualified human review before institutional publication.

37.5 Source Verification

AI-generated citations, summaries, and factual claims require source verification.

37.6 Decision Use

An AI system should not be presented as the accountable decision maker.

37.7 Automated Scoring

Disclose:

- Judge system
- version
- validation
- bias
- human review
- appeal

37.8 Translation

Identify machine-assisted translation where it may affect normative meaning.

37.9 Provenance

Maintain records of:

- Tool
- version
- purpose
- human reviewer
- material edits

37.10 Confidentiality

Do not submit protected information to unauthorized AI services.

---

38. Public Communications and Media

38.1 Communications Standard

Public communications should preserve:

- Authority
- evidence
- uncertainty
- status
- limitations
- current version

38.2 Press Release

A press release should not broaden a technical finding.

38.3 Social Media

Short-form posts should link to the full record.

38.4 Quotes

Institutional representatives should distinguish:

- Evidence
- interpretation
- opinion
- future proposal

38.5 Embargoed Media

Embargoes should not create privileged influence over findings.

38.6 Corrections

Correct media and social posts visibly.

38.7 Partner Announcements

Coordinate facts without allowing partners to control institutional conclusions.

38.8 Headline Risk

Avoid headlines that state:

- Safe
- approved
- certified
- official
- global consensus

unless those claims are precisely true.

38.9 Communications Archive

Preserve material public statements and corrections.

---

39. Transparency in High-Stakes Domains

39.1 Domain-Specific Balance

High-stakes domains may require greater public accountability and stronger information protection simultaneously.

39.2 Cybersecurity

Publicly disclose:

- Risk category
- methodology
- findings at safe level
- mitigation status
- residual risk

Protect actionable exploit details until responsible release.

39.3 Biological and Chemical Risk

Disclose governance, construct, high-level results, review, and safeguards.

Protect procedural detail that could materially increase harmful capability.

39.4 Critical Infrastructure

Coordinate with operators and authorized authorities.

Avoid disclosing vulnerabilities that create operational risk.

39.5 Persuasion and Information Integrity

Disclose:

- Population
- context
- outcome
- uncertainty
- ethical review
- privacy
- distributional effect

39.6 Human Subjects

Protect participants and provide ethics transparency.

39.7 National Security

Comply with lawful restrictions while maintaining public minimum and independent authorized review where possible.

39.8 No Security Exceptionalism

A high-stakes label should not eliminate all accountability automatically.

---

40. Legal and Regulatory Transparency

40.1 Legal Status

State the jurisdiction and status of cited law.

40.2 Legal Advice

Institutional publications should not imply individualized legal advice.

40.3 Regulatory Relationship

Disclose whether Standards Body is:

- Advising
- contracting
- partnering
- recognized
- delegated
- regulated

40.4 Legal Requests

Requests from authorities should be handled under:

- Law
- due process
- privacy
- security
- records
- transparency where permitted

40.5 Government Data Requests

Publish aggregate reporting where lawful.

40.6 Litigation and Investigations

Protect legitimate legal process while disclosing material institutional effect.

40.7 Legal Incorporation of Standards

State when an external authority adopts or references a Standards Body document.

40.8 No Legal Effect by Implication

A standard's technical importance should not be described as legal force.

40.9 Legal Change

Update affected public materials promptly.

---

41. International Reporting and Interoperability

41.1 Common Disclosure Core

A shared transparency profile should include:

- Organization
- system or standard
- governance
- risk process
- evaluation
- incidents
- status
- correction

41.2 Local Extensions

Allow jurisdictional and sectoral additions.

41.3 Reporting Frameworks

Standards Body should map its disclosures to relevant external frameworks rather than require duplicative reporting without justification.

41.4 HAIP Mapping

A future crosswalk may connect Standards Body records to the OECD Hiroshima AI Process Reporting Framework.

41.5 Self-Reporting Status

Self-reported information should be labeled as such.

41.6 Verification Status

Use:

- Unverified
- internally reviewed
- independently reviewed
- audited
- certified

only when accurate.

41.7 Recognition

Recognition of an external transparency report should state:

- Purpose
- scope
- evidence
- limitations
- expiration

41.8 International Accessibility

Support:

- Translation
- machine-readable formats
- stable identifiers
- common status terms
- regional context

---

42. Annual Transparency Report

42.1 Purpose

The annual report should make the institution's own transparency performance visible.

42.2 Institutional Information

- Legal and institutional status
- governance changes
- authority changes
- board and council changes

42.3 Funding

- Revenue categories
- concentration
- major grants
- in-kind support
- restricted funding
- related-party transactions

42.4 Work

- Research
- standards
- evaluations
- reviews
- registries
- partnerships

42.5 Disclosure

- Proactive publications
- transparency requests
- release
- partial release
- denial
- response time
- appeals

42.6 Classification

- Records by class

- new restrictions
- declassifications
- overdue reviews
- classification appeals

42.7 Corrections

- Editorial
- factual
- methodological
- withdrawals
- average correction time

42.8 Complaints and Appeals

- Volume
- category
- outcomes
- recurrence
- retaliation

42.9 Incidents

- Governance
- research
- security
- evaluation
- privacy
- standards

42.10 Accessibility

- Formats
- translations
- reported barriers
- remediation

42.11 Performance

- Metrics
- failures
- independent review
- next-year improvement plan

43. Transparency Metrics

43.1 Purpose

Metrics should test whether transparency supports understanding and accountability.

43.2 Publication Metrics

- Percentage of required records published
- publication timeliness
- stale-record rate
- broken-link rate
- metadata completeness
- archive completeness

43.3 Request Metrics

- Number
- response time
- full release
- partial release
- denial
- review
- requester satisfaction

43.4 Classification Metrics

- Records by class
- overdue review
- declassification
- overclassification findings
- unauthorized disclosure

43.5 Correction Metrics

- Correction time
- downstream propagation
- repeated errors
- withdrawal rate
- correction discoverability

43.6 Governance Metrics

- Conflict disclosure completion

- published recusals
- decision-record completeness
- dissent publication
- meeting-summary timeliness

43.7 Standards Metrics

- Draft availability
- comment publication
- disposition completeness
- participation transparency
- status accuracy

43.8 Evaluation Metrics

- System-identity completeness
- protocol disclosure
- uncertainty reporting
- expiration compliance
- protected-review coverage

43.9 Comprehension Metrics

- User testing
- misunderstanding reports
- public survey
- correction of common misinterpretations

43.10 Anti-Metric Rule

High document volume does not equal meaningful transparency.

---

44. Transparency Audit

44.1 Audit Purpose

Determine whether transparency commitments are implemented.

44.2 Audit Scope

- Governance
- funding
- conflicts

- standards
- research
- evaluation
- protected evidence
- incidents
- corrections
- requests
- accessibility
- website
- records

44.3 Audit Sample

Test:

- Current records
- archived records
- withheld records
- correction histories
- denied requests
- classification reviews
- public claims

44.4 Audit Questions

- Is required information public?
- Is it current?
- Is status visible?
- Are limitations preserved?
- Are restrictions justified?
- Are records accessible?
- Can decisions be reconstructed?
- Are corrections discoverable?
- Are confidential conclusions independently reviewed?
- Do public claims match authority?

44.5 Findings

- Critical
- material
- minor
- improvement

44.6 Critical Findings

Examples:

- False authority claim
- hidden controlling funding
- undisclosed material conflict
- silent withdrawal
- expired result shown as current
- restricted evidence used without independent review
- personal data exposure
- security redaction failure

44.7 Corrective Action

Assign:

- Owner
- deadline
- verification
- public notice
- systemic review

44.8 Independent Review

A mature institution should receive periodic external transparency review.

---

45. Transparency Maturity Model

Level 0: Selective Publication

Characteristics:

- Promotional disclosures
- no classification
- no status
- no correction register
- no request process

Level 1: Basic Institutional Disclosure

Characteristics:

- Mission
- authority

- governance
- funding categories
- publications
- contact

Level 2: Process Transparency

Characteristics:

- Work registers
- methods
- conflicts
- decisions
- public review
- corrections
- archives

Level 3: Evidence and Status Transparency

Characteristics:

- Evidence levels
- uncertainty
- provenance
- system identity
- expiration
- protected-review records
- machine-readable status

Level 4: Accountable Controlled Transparency

Characteristics:

- Formal classification
- request and appeal
- redaction
- declassification
- security review
- privacy controls
- independent audit

Level 5: Adaptive Transparency Infrastructure

Characteristics:

- Automated status propagation

- federated registries
- international crosswalks
- measured comprehension
- continuous audit
- transparent institutional learning

45.1 Maturity Rule

Transparency maturity depends on reliability and use, not the number of pages published.

---

46. Consolidated Transparency Failure Modes

46.1 Transparency Theater

Failure:

The institution publishes polished summaries while withholding the information needed to evaluate authority, evidence, conflicts, or limitations.

Controls:

- Required disclosure schema
- audit
- public minimum
- correction
- independent review

46.2 Data Dumping

Failure:

Large volumes of unorganized documents create the appearance of openness without usability.

Controls:

- Search
- metadata
- summaries
- status
- indexes
- machine readability

46.3 Authority Ambiguity

Failure:

The institution's role is described in language that implies regulatory, certification, accreditation, or international authority it does not possess.

Controls:

- Authority note
- controlled public descriptions
- website review
- correction registry

46.4 Status Omission

Failure:

Draft, expired, withdrawn, or superseded material appears current.

Controls:

- Mandatory status
- automated propagation
- registry links
- archive warnings

46.5 Limitation Removal

Failure:

A public summary omits the limitations contained in the technical report.

Controls:

- Claim review
- linked limitations
- summary checklist
- accountable owner

46.6 Funding Opacity

Failure:

Major financial dependence is hidden through intermediaries, categories, or in-kind support.

Controls:

- Beneficial-source review
- concentration reporting
- related-party disclosure
- audit

46.7 Conflict Minimalism

Failure:

A participant states "no conflict" without disclosing relevant relationships.

Controls:

- Structured disclosure
- matter-specific review
- public summaries
- recusal record

46.8 Participation Opacity

Failure:

A process appears open while the institution does not disclose who actually participated or influenced the result.

Controls:

- Participant list
- affiliation
- attendance
- comment influence
- balance report

46.9 Selective Evidence

Failure:

Only evidence supporting the preferred conclusion is published.

Controls:

- Contrary-evidence register
- independent review
- evidence standards
- dissent

46.10 Self-Reporting Inflation

Failure:

Organizational claims are published without clear self-reported or unverified status.

Controls:

- Verification labels
- source classification
- evidence level
- registry metadata

46.11 Confidentiality Capture

Failure:

Commercial, political, or reputational interests use confidentiality to prevent scrutiny.

Controls:

- Specific rationale
- independent review
- public minimum
- expiration
- appeal

46.12 Security Overreach

Failure:

Security classifications remain broad, indefinite, and unreviewed.

Controls:

- Lowest necessary class
- review dates
- declassification
- overclassification appeal
- aggregate reporting

46.13 Security Underreach

Failure:

The institution releases dangerous details, protected tasks, or vulnerabilities.

Controls:

- security review
- responsible disclosure
- redaction
- staged release
- publication options

46.14 Privacy Exposure

Failure:

Transparency exposes personal data, whistleblowers, research participants, or complainants.

Controls:

- minimization
- de-identification
- aggregation
- privacy review
- controlled access

46.15 Redaction Failure

Failure:

The redaction can be reversed or hidden metadata reveals the content.

Controls:

- technical validation
- second review
- secure export
- incident response

46.16 Misleading Redaction

Failure:

The remaining text creates a false impression.

Controls:

- contextual summary
- redaction marking
- reviewer sign-off

46.17 Independent-Review Ambiguity

Failure:

A factual check or sponsor-selected review is described as independent expert review.

Controls:

- review taxonomy
- mandate

- independence profile
- access disclosure

46.18 Evaluation Black Box

Failure:

A high-consequence public claim relies on hidden tests with no reviewable protocol or independent access.

Controls:

- public protocol layer
- qualified review
- integrity status
- limitations
- expiration

46.19 Benchmark Exposure

Failure:

Full disclosure compromises evaluation validity.

Controls:

- protected content
- rotation
- chain of custody
- retired-task release

46.20 Incident Silence

Failure:

A material incident remains undisclosed after it affects public reliance.

Controls:

- notice triggers
- incident status
- board escalation
- external audit

46.21 Premature Incident Certainty

Failure:

Early communications state a cause or scope not yet established.

Controls:

- known and unknown fields
- staged updates
- correction

46.22 Silent Correction

Failure:

A file changes without notice or history.

Controls:

- immutable releases
- correction notice
- versioning
- downstream notification

46.23 Link Rot

Failure:

Sources and records become inaccessible.

Controls:

- archival copies where lawful
- link monitoring
- persistent identifiers
- source register

46.24 Translation Drift

Failure:

A translation changes normative meaning.

Controls:

- controlling language
- domain review
- version alignment
- correction

46.25 Accessibility Failure

Failure:

Information is published in formats inaccessible to affected audiences.

Controls:

- accessible HTML
- captions
- alt text
- user testing
- alternative formats

46.26 Communications Overreach

Failure:

A social post or headline broadens the evidence.

Controls:

- public-claim review
- links
- correction
- spokesperson accountability

46.27 Partner Claim Drift

Failure:

A partner describes the relationship as endorsement, approval, or certification.

Controls:

- agreement language
- monitoring
- correction
- termination

46.28 AI-Generated Hallucination

Failure:

AI-assisted drafting introduces false sources, claims, or interpretations.

Controls:

- human review

- source verification
- provenance
- restricted-data controls

46.29 Request Obstruction

Failure:

The institution delays or narrows transparency requests without reason.

Controls:

- service targets
- request log
- review
- aggregate reporting

46.30 Permanent Embargo

Failure:

A temporary publication delay becomes indefinite.

Controls:

- expiration
- milestones
- independent review
- escalation

46.31 Transparency as Safety Proof

Failure:

Detailed disclosure is used to imply that a system or institution is safe.

Controls:

- explicit nonclaims
- evidence levels
- outcome review

46.32 Transparency as Compliance Proof

Failure:

A report is treated as proof of full legal compliance.

Controls:

- legal-status note
- jurisdiction
- verification status
- scope

46.33 Public-Record Fragmentation

Failure:

Relevant information is divided across inconsistent systems.

Controls:

- canonical identifiers
- integrated registry
- status propagation
- source-of-truth rules

46.34 Stale Transparency

Failure:

Accurate historical information is presented without an updated date or review.

Controls:

- current-through date
- owner
- stale warnings
- review schedule

46.35 Transparency Burden

Failure:

Reporting requirements consume resources without improving accountability.

Controls:

- proportionality
- reusable schemas
- data minimization
- interoperability
- outcome testing

---

47. Serious Objections and Responses

Objection 1: Full transparency is the only credible model

Full public disclosure can be incompatible with:

- Security
- privacy
- evaluation integrity
- responsible research
- lawful confidentiality

Credibility requires governed access and independent scrutiny, not indiscriminate release.

Objection 2: Confidential evidence can never support public decisions

Confidential evidence creates a legitimacy challenge.

It can support a bounded public decision when:

- The public minimum is available
- qualified independent reviewers have access
- reviewer competence and independence are disclosed
- limitations are visible
- appeal and later review exist

Objection 3: Transparency makes standards and evaluations easier to game

Some disclosure increases gaming.

The response is layered transparency:

- Public constructs and governance
- protected task details
- dynamic methods
- exposure monitoring
- versioning

Objection 4: Transparency requirements will slow urgent work

They can add time.

Urgent work may use provisional and staged disclosure.

Urgency should not remove:

- Authority
- owner
- status
- reasons
- expiration
- review

Objection 5: Funders will not support exact disclosure

Some legitimate donor privacy and contractual constraints exist.

Material institutional dependence should still be visible through:

- Names where possible
- ranges
- concentration
- purpose
- control rights
- independent review

Objection 6: Public governance records expose individuals to harassment

Personal safety may justify limited disclosure.

The institution should publish professional roles and decisions while minimizing unnecessary personal information.

Objection 7: Public comment and open records invite bad-faith attacks

Bad-faith behavior exists.

A credible process can use:

- Conduct rules
- structured submissions
- moderation
- evidence requirements
- anti-harassment protections

without eliminating legitimate scrutiny.

Objection 8: Transparency about incidents will damage trust

Concealed incidents can damage trust more severely.

Good incident transparency shows:

- What happened
- what remains uncertain
- what changed
- how recurrence is addressed

Objection 9: Self-reporting is sufficient for transparency

Self-reporting is useful and scalable.

It should be labeled accurately and supplemented with independent review where consequence is high.

Objection 10: Independent review cannot be transparent when the evidence is secret

The institution can disclose:

- Review mandate
- reviewers
- competence
- independence
- access level
- findings
- limitations

while protecting exact evidence.

Objection 11: Publishing all dissent will confuse the public

Unresolved material dissent is part of the evidence.

It can be presented clearly without giving every minor disagreement equal weight.

Objection 12: Machine-readable transparency is too technical

Structured records support registries, versioning, and interoperability.

They should supplement accessible human summaries.

Objection 13: Transparency creates legal risk

Disclosure can create legal risk.

So can concealment or misleading public claims.

Qualified legal review should support accurate, bounded publication.

Objection 14: A private institution has no duty to answer requests

The framework does not create statutory rights.

A voluntary request process supports accountability and institutional legitimacy.

Objection 15: Transparency standards will become paperwork exercises

They can.

The framework requires evaluation of comprehension, correction, decision quality, and outcome usefulness.

---

48. Transparency Implementation Pathway

Phase 1: Canonical Disclosure

Publish:

- `PROJECT_IDENTITY.md`
- institutional status
- authority note
- canonical file register
- version history
- correction channel

Phase 2: Governance Disclosure

Publish:

- Governance framework
- decision owners
- advisory roles
- conflicts
- material funding
- project register

Phase 3: Classification and Records

Adopt:

- Information classes
- record owners

- retention
- classification review
- redaction
- provenance
- status metadata

Phase 4: Work Registries

Create:

- Research register
- standards work program
- evaluation register
- review register
- correction register
- partnership register

Phase 5: Request and Review

Create:

- Transparency request process
- denial reasons
- partial-release process
- internal review
- request log

Phase 6: Protected Evidence

Establish:

- Secure reviewer access
- held-out evidence rules
- public minimum
- security committee review
- declassification

Phase 7: Reporting

Publish:

- Annual transparency report
- funding report
- governance report
- incident report
- accessibility report

Phase 8: Audit

Conduct:

- Internal transparency audit
- external review
- public correction of findings
- maturity assessment

Phase 9: Machine-Readable Infrastructure

Publish:

- Stable identifiers
- structured registries
- status APIs or exports
- correction links
- provenance metadata

Phase 10: International Crosswalk

Map disclosures to:

- Relevant international reporting frameworks
- standards bodies
- public authorities
- evaluator registries
- incident systems

---

49. First Transparency Pilot

49.1 Pilot Title

Standards Body Public Record and Status Registry

49.2 Purpose

Create a unified register for:

- Canonical documents
- standards projects
- research
- evaluations
- reviews
- decisions

- corrections
- partnerships
- funding disclosures

49.3 Minimum Fields

- Identifier
- title
- type
- owner
- version
- status
- publication date
- current-through date
- review date
- classification
- evidence level where relevant
- related records
- correction
- canonical link

49.4 Public Interface

Provide:

- Search
- filters
- status history
- current version
- archive
- correction alerts
- machine-readable export

49.5 Protected Interface

Qualified reviewers may access controlled metadata and records according to role.

49.6 Success Criteria

- No canonical file lacks status
- corrections link in both directions
- superseded records remain accessible
- public pages match the register
- user testing confirms understanding
- protected access is logged

- registry can export structured data

49.7 Nonclaims

The registry does not itself certify truth, safety, compliance, or authority.

---

50. Transparency Scorecard

Dimension	Core question
Identity	Is the institution's mission, stage, and authority clear?
Governance	Are bodies, appointments, powers, and limits public?
Funding	Are material sources, restrictions, and concentration visible?
Conflicts	Are relevant interests and recusals disclosed?
Work program	Can the public see current and planned work?
Process	Are methods and decision procedures understandable?
Participation	Is it clear who participated and who was missing?
Evidence	Are sources, levels, limitations, and contrary evidence visible?
Status	Are draft, current, expired, superseded, and withdrawn states clear?
Evaluation	Are object, protocol, conditions, uncertainty, and expiration disclosed?
Protected evidence	Is the public minimum available and independent access governed?
Standards	Are drafts, comments, objections, consensus, and maintenance visible?
Review	Are reviewer competence, independence, access, and findings disclosed?
Assurance	Are evaluation, audit, certification, accreditation, and recognition distinct?
Decisions	Are authority, reasons, conflicts, dissent, and appeal recorded?
Incidents	Are material incidents and corrective actions reported safely?
Corrections	Are changes, withdrawals, and downstream effects visible?
Provenance	Can the origin and history of a record be traced?
Privacy	Is personal information minimized and protected?
Security	Are dangerous details protected proportionately?
Redaction	Are redactions technically sound and nonmisleading?
Requests	Is there a usable disclosure-request and review process?
Accessibility	Can intended audiences access and understand information?
Language	Are controlling language and translation status clear?
Automation	Is material AI assistance attributable and reviewed?
Partnerships	Are purpose, funding, roles, and claim limits visible?
Legal status	Are jurisdiction and legal effect described accurately?

Dimension	Core question
International	Can disclosures map across institutions and jurisdictions?
Timeliness	Is information published while useful?
Audit	Is transparency independently evaluated?
Impact	Does disclosure improve understanding, challenge, and correction?

50.1 Critical Failures

The following normally invalidate a claim that Standards Body is operating transparently:

- False or ambiguous authority claim
- hidden controlling funder
- undisclosed material conflict
- expired or withdrawn record shown as current
- material correction without notice
- high-consequence secret conclusion without qualified independent review
- public release of protected personal data
- public release of dangerous information through negligence
- inaccessible complaint or appeal process
- public summary that omits decisive limitations
- certification or accreditation implication without authority
- inability to reconstruct a material decision
- no record owner or status
- refusal to disclose the existence of a material process without justified exceptional reason

50.2 No Composite Transparency Score

Do not average all dimensions into one number.

A critical failure cannot be offset by a large volume of lower-value publication.

---

51. Transparency Request Template

Request ID:
Requester:
Date:
Contact:
Preferred format:

Record or Topic Requested

Date Range

Public-Interest or Urgency Context

Accessibility Need

Clarification

Search Conducted

Records Located

Classification

Decision

- Full release
- partial release
- summary
- referral
- no record
- denial

Redactions or Reasons

Fees

Response Date

Review or Appeal

---

52. Information Classification Record Template

Record ID:
Title:
Owner:

Date:
Classification:

- Public
- controlled
- confidential
- restricted
- highly restricted

Information Description

Harm From Disclosure

Harm From Nondisclosure

Legal, Privacy, Security, or Integrity Basis

Authorized Roles

Handling

Redaction or Summary Options

Public Minimum

Review Date

Release Conditions

Retention or Destruction

Decision Authority

Appeal

---

53. Public Minimum Template

Protected activity or record:
Responsible body:
Status:

Existence

Purpose

General Scope

Authority

Process Type

Review Type

Current Outcome or Status

Limitations

Protection Rationale

Qualified Access Path

Next Review Date

Contact

---

54. Funding Disclosure Template

Funder:
Funding type:
Amount or range:
Period:
Purpose:

Restrictions

Control Rights

Draft Review

Publication Rights

Early Access

Appointment or Participation Rights

Related Standards, Evaluations, or Partners

In-Kind Support

Related-Party Status

Concentration Effect

Independence Safeguards

Public Notes

---

55. Conflict Disclosure Template

Person:
Role:
Period:

Employment and Offices

Ownership and Investments

Consulting and Clients

Grants and Funding

Board and Advisory Roles

Intellectual Property

Research or Standards Authorship

Political or Government Roles

Personal or Family Relationships

Access Dependencies

Matter-Specific Interests

Conflict Decision

Recusal or Controls

Review Date

---

56. Evaluation Transparency Record Template

Evaluation ID:
System ID and version:
Protocol ID and version:
Evaluator:
Sponsor:
Date:
Status:

Purpose and Claim

Evaluated Object

Conditions

Elicitation

Tools and Scaffolds

Task Domains

Integrity Status

Scoring

Result

Uncertainty

Evidence Level

Confidence

Review

Protected Components

Public Minimum

Limitations

Valid Through

Re-Evaluation Triggers

Corrections

57. Independent Review Disclosure Template

Review ID:
Subject:
Mandate:
Date:

Reviewers

Competence

Selection

Independence Profile

Conflicts

Funding

Access

Method

Findings

Dissent

Reviewed-Party Response

Publication Control

Protected Information

Limitations

Status

58. Incident Transparency Notice Template

Incident ID:
Date detected:
Status:
Responsible body:

What Is Known

What Is Not Known

Affected Systems, Records, or Persons

Actual and Potential Impact

Immediate Action

Public-Reliance Effect

Protected Details

Reporting to Authorities or Affected Parties

Standards, Protocol, or Registry Effect

Next Update

Correction History

---

59. Correction Notice Template

Correction ID:
Affected record:
Prior version:
Corrected version:
Date:

Error

Correction

Cause

Effect on Findings or Decisions

Dependent Records

Responsible Body

Review

Notification

Archive

---

60. Redaction Review Template

Record:
Reviewer:
Date:

Proposed Redactions

Protection Categories

Authority

Segregable Information

Public-Interest Need

Misleading-Context Risk

Technical Redaction Validation

Second Review

Release Decision

Review Date

---

61. Annual Transparency Report Template

Institutional Status

Governance and Appointments

Funding and Concentration

Conflicts and Recusals

Research and Standards Work

Evaluations and Reviews

Proactive Disclosures

Transparency Requests

Classification and Declassification

Protected Evidence

Incidents

Complaints and Appeals

Corrections and Withdrawals

Privacy and Security

Accessibility and Translation

AI-Assisted Content

Metrics

Audit Findings

Corrective Actions

Next-Year Commitments

---

62. Canonical Standards Body Transparency Positions

Standards Body adopts the following working positions.

1. Transparency should support accountability, understanding, participation, correction, and informed action.
2. Transparency is not identical to total public disclosure.
3. Responsible transparency includes proportionate confidentiality, privacy, security, and evaluation-integrity controls.
4. The institution should presume that governance and public-interest information is public unless a documented reason justifies protection.
5. Protection should use the lowest classification sufficient for the legitimate risk.
6. Every protected record should have an owner, rationale, access rule, review date, and release condition.
7. A public minimum should remain available for protected work whenever safe and lawful.
8. The existence of a material institutional process should not be secret without exceptional justification.
9. Transparency does not establish accuracy, safety, fairness, legality, competence, independence, or legitimacy by itself.
10. A disclosure should identify the audience and purpose it is intended to serve.
11. Public summaries should preserve material limitations and uncertainty.
12. Status is part of meaning.
13. Draft, current, conditional, expired, superseded, suspended, withdrawn, and retired records should be distinguishable.
14. A superseded record should remain discoverable with a visible warning.
15. Material decisions should identify owner, authority, evidence, conflicts, reasons, dissent, appeal, and review date.

16. The institution's current legal and institutional stage should be visible on major public materials.
17. Standards Body should not imply regulatory, certification, accreditation, governmental, or international authority it does not possess.
18. Membership should not be presented as endorsement.
19. Partnership should not be presented as approval.
20. Participation in drafting should not be presented as conformity.
21. A public registry listing should not be presented as certification unless the registry actually records a valid certificate.
22. Funding is material governance information.
23. Material funding should disclose source, amount or range, purpose, restrictions, control rights, and related work.
24. In-kind support can create influence and should be disclosed when material.
25. Funding concentration should be reported.
26. A funder should not control findings, standards language, reviewers, or publication.
27. Role-relevant conflicts should be disclosed and governed.
28. Disclosure alone may be insufficient to manage a conflict.
29. Material recusals should be visible.
30. Intellectual, reputational, client, political, and access-dependence conflicts can matter as much as direct financial conflicts.
31. Research projects should have visible purpose, owner, method, status, sponsor, and correction route.
32. Confirmatory and high-consequence research should use prospective registration where appropriate.
33. Negative, null, and inconclusive findings should not be suppressed for reputational reasons.
34. Material deviations from a research plan should be disclosed.
35. Research data, code, prompts, and artifacts should be published when safe, lawful, and useful.
36. Ethical and privacy duties may justify controlled access.
37. Standards work programs should be public.

38. New standards work should disclose need, scope, evidence maturity, participants, and current status.
39. Standards drafts should display clearly that they are not approved.
40. Public comments and dispositions should ordinarily be visible.
41. The number and categories of protected comments should be disclosed.
42. Substantial objections and unresolved dissent should remain visible.
43. Consensus reports should explain how consensus was determined.
44. Standards maintenance, interpretations, corrections, amendments, and withdrawals should be public.
45. Core public-interest standards should be freely readable.
46. Evaluation records should identify the exact model, system, version, protocol, evaluator, conditions, date, status, and expiration.
47. Tools, scaffolds, prompts, retries, fine-tuning, and human assistance should be disclosed at the level necessary for interpretation.
48. Evaluation uncertainty should not be removed from public summaries.
49. Failure to demonstrate capability should not be reported as proof of incapability.
50. A high-consequence evaluation result should identify evidence level and review status.
51. Public evaluation transparency does not always require publication of exact tasks.
52. Held-out tasks may remain protected when exposure would weaken validity.
53. Protection of task content should not eliminate transparency about construct, governance, method class, evaluator, uncertainty, status, and limitations.
54. A public conclusion based on protected evidence requires sufficient qualified independent review.
55. A secret test is not valid merely because it is secret.
56. Task compromise and contamination should be disclosed through result status.
57. Retired tasks should be considered for release when safe and useful.
58. Independent review should disclose mandate, reviewer competence, selection, funding, independence, access, findings, and limitations.
59. A factual review should not be described as independent expert review.
60. External status does not automatically establish independence.

61. Evaluator profiles should disclose competence scope, ownership, services, conflicts, security, quality, and recognition status.
62. Accreditation should be reported with its exact scope.
63. Certification should be reported with its exact object, scheme, version, status, and expiration.
64. Standards Body should not describe pilot recognition as accreditation.
65. Any future mark should link to a public record explaining its exact meaning.
66. Incident transparency should prioritize safety, containment, affected persons, evidence preservation, and accurate staged reporting.
67. Initial incident notices should distinguish known facts from unknowns.
68. Incident updates should be corrected visibly as facts change.
69. Near-miss learning should be shared at the safest useful level.
70. Security should not be used to conceal invalid evidence, conflicts, misconduct, or material institutional failure.
71. Transparency should not expose active vulnerabilities or dangerous procedures irresponsibly.
72. Responsible disclosure should use documented timelines, review, and escalation.
73. Embargoes should not become indefinite organizational vetoes.
74. Personal data should be minimized.
75. Whistleblower, complainant, and research-participant identities should be protected proportionately.
76. Transparency should not require publication of unnecessary private financial or family information.
77. Aggregate reporting should preserve material patterns and severe outliers.
78. Redactions should be technically irreversible in the released artifact.
79. Redacted records should indicate the fact and basis of removal where safe.
80. Redaction should not create a misleading interpretation.
81. Material corrections should not be silent.
82. The original public record should remain linked to its correction or withdrawal.
83. Corrections should propagate to dependent pages, registries, reports, and claims.
84. Source provenance should be preserved.

85. Primary sources should be used for legal, regulatory, standards, and official-status claims where possible.
86. Self-reported information should be labeled as self-reported.
87. Verification status should be stated accurately.
88. Evidence contrary to a material institutional conclusion should not be hidden.
89. Every canonical document should display version, owner, status, current-through date, and revision history.
90. Material AI assistance in institutional analysis, drafting, translation, or scoring should be recorded and disclosed appropriately.
91. Human accountability remains when AI assists.
92. AI-generated sources and factual claims should be verified before publication.
93. Protected information should not be submitted to unauthorized AI systems.
94. The website should operate from a controlled source of truth.
95. Major pages should display update and status information.
96. Superseded public records should remain accessible through an archive.
97. Public communications should not broaden the underlying technical evidence.
98. Social-media summaries should link to the full record.
99. Accessibility is part of transparency.
100. Plain-language summaries should supplement technical detail rather than replace it.
101. International transparency should support translation, common metadata, local context, and machine-readable records.
102. A process should not claim broad international legitimacy when practical participation and disclosure are limited to one region or language.
103. External reporting frameworks should be mapped where useful to reduce duplicative burden.
104. Self-reporting and independent verification should remain distinct.
105. A transparency report is not proof of legal compliance.
106. Transparency requests should receive timely acknowledgment, reasoned response, and internal review.
107. Partial release, redaction, summary, or qualified access should be considered before complete denial.
108. Request procedures should not be used to obstruct legitimate scrutiny.

109. Transparency metrics should measure currentness, usability, comprehension, correction, and accountability, not page count alone.
110. Transparency systems should receive internal and periodic external audit.
111. Critical transparency failures may require correction, decision review, standard suspension, result withdrawal, governance change, or public notice.
112. Standards Body should publish its own transparency failures and improvement plans.
113. Transparency obligations should remain proportionate to institutional stage and consequence.
114. Early-stage limitations should be disclosed rather than hidden behind institutional presentation.
115. Mature authority should require mature transparency.
116. An expansion into standards, assurance, certification, accreditation, or public authority should trigger stronger disclosure and independent review.
117. Transparency should preserve the ability of others to challenge the institution.
118. The institution should remain willing to disclose that evidence is incomplete, access is insufficient, or no conclusion is justified.
119. The strongest transparency statement is sometimes an honest statement of uncertainty.
120. The ultimate test of transparency is whether others can understand what happened, who was responsible, what evidence existed, what remained hidden and why, and how error can be corrected.

---

63. Relationship to Other Canonical Files

`PROJECT_IDENTITY.md`

Defines Standards Body's present identity, mission, institutional stage, authority boundaries, audiences, and approved public descriptions.

This transparency framework operationalizes those public commitments.

`PROJECT_MANIFESTO.md`

Defines the deeper public-interest purpose that transparency should serve.

INSTITUTION_DESIGN.md

Defines the institutional architecture and separation among research, standards, evaluation, assurance, accreditation, certification, and public authority.

This framework defines what each layer should disclose.

GOVERNANCE_FRAMEWORK.md

Defines governing bodies, decision rights, conflicts, funding governance, appeals, records, security, and institutional transitions.

This framework makes those systems visible and reviewable.

STANDARDS_DEVELOPMENT_PROCESS.md

Defines standards work programs, drafting, participation, public review, comments, consensus, voting, approval, maintenance, and appeal.

This framework defines the disclosure requirements across that lifecycle.

FOUNDATIONS.md

Defines the overview of the eight foundations.

FOUNDATIONS_APPENDIX.md

Defines the integrated relationship among protocols, protected evidence, high-stakes evaluation, independent review, assurance, standards, incentives, and interoperability.

This framework supplies the shared public and protected disclosure architecture.

TERMINOLOGY.md

Defines transparency, disclosure, status, review, independence, certification, accreditation, and related terms.

EVIDENCE_STANDARDS.md

Defines source quality, evidence levels, confidence, claims, contrary evidence, and correction.

This framework defines how those properties should be disclosed.

RESEARCH_METHODOLOGY.md

Defines project registration, methods, review, ethics, security, publication, and correction.

This framework establishes their public and protected records.

`TAXONOMY.md`

Classifies records, statuses, security levels, evidence, actors, decisions, incidents, and relationships.

This framework operationalizes those classifications.

`EVALUATION_PHILOSOPHY.md`

Defines what evaluation results mean and where their authority ends.

This framework prevents public reporting from overstating those results.

Foundation 1

Requires transparent protocol purpose, construct, version, changes, comparability, and retirement.

Foundation 2

Requires public governance and qualified review around protected evaluation content.

Foundation 3

Requires disclosure proportional to the consequence of high-stakes evaluation.

Foundation 4

Requires review mandate, competence, independence, access, conflicts, findings, and dissent to be visible.

Foundation 5

Requires evaluator, certification, accreditation, and assurance status to be scope-specific and verifiable.

Foundation 6

Requires standards stages, evidence maturity, obligations, recognition, and legal effect to remain distinguishable.

Foundation 7

Requires incentives, funding, awards, rankings, and recognition systems to disclose their rules and gaming risks.

Foundation 8

Requires interoperable metadata, mappings, recognition conditions, local extensions, and noncomparability to remain visible.

`CONTRIBUTOR_FRAMEWORK.md`

Will define public contributor roles, credit, affiliations, conduct, confidentiality, and removal.

`PARTNERSHIP_PRINCIPLES.md`

Will define partnership disclosures, funding, authority, branding, data, intellectual property, and exit.

`LONG_TERM_ROADMAP.md`

Will define the timing at which more mature transparency infrastructure becomes required.

`WEBSITE_SOURCE_OF_TRUTH.md`

Will define the canonical public facts, page ownership, claims, update cadence, and correction workflow for standardsbody.ai.

`SOURCES.md`

Will maintain the source registry and source-status information.

`VERSION_HISTORY.md`

Will preserve canonical changes, supersession, correction, and institutional history.

---

64. Final Transparency Position

A frontier AI institution can appear transparent while remaining difficult to scrutinize.

It can publish thousands of pages without revealing:

- Who decided
- who paid

- which evidence was missing
- which system was actually tested
- which limitations applied
- which objections remained
- which records were protected
- why they were protected
- whether the result is still current
- how an error can be corrected

It can also create harm by publishing everything.

It can expose:

- Active vulnerabilities
- protected evaluation tasks
- research participants
- whistleblowers
- personal data
- dangerous technical procedures
- security systems

The solution is not maximum disclosure or maximum secrecy.

The solution is governed transparency.

Governed transparency begins with a presumption of public accountability.

It then asks whether a specific piece of information requires protection.

When protection is justified, the institution should narrow it.

It should identify:

- The record
- the harm
- the authority
- the access group
- the review date
- the release condition

It should preserve a public minimum.

It should provide qualified independent access when a public conclusion depends on protected evidence.

It should record challenges.

It should release or declassify when the reason for protection ends.

The institution should make several things especially difficult to hide:

- Authority
- funding
- conflicts
- process
- evidence status
- uncertainty
- dissent
- correction
- expiration
- institutional failure

These are the facts most likely to determine whether public trust is justified.

Transparency should not become a marketing layer placed over institutional work after decisions are complete.

It should be embedded in:

- Identity
- governance
- research design
- standards development
- evaluation protocols
- incident response
- records
- funding
- partnerships
- corrections
- succession

The mature institution should be able to answer:

- What information is public?
- What is protected?
- Why?
- Who has access?
- Who reviewed it?
- What conclusion does it support?
- What conclusion does it not support?
- When will the restriction be reviewed?
- How can a person challenge the decision?
- How will the public know if the underlying evidence changes?

The defining transparency rule of Standards Body is:

> **Make the institution understandable enough to challenge, the evidence traceable enough to assess, the status current enough to rely on, and the protected information governed enough to review without creating greater harm.**

References and Research Basis

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

[^nist-core]: National Institute of Standards and Technology, **AI RMF Core** and **Accountable and Transparent Trustworthiness Characteristic**, NIST AI Resource Center. https://airc.nist.gov/airmf-resources/airmf/5-sec-core/ and https://airc.nist.gov/airmf-resources/airmf/3-sec-characteristics/

[^nist-playbook]: National Institute of Standards and Technology, **AI RMF Playbook**, including Govern, Map, Measure, and Manage suggested actions. https://airc.nist.gov/airmf-resources/playbook/

[^nist-privacy]: National Institute of Standards and Technology, **NIST Privacy Framework**. https://www.nist.gov/privacy-framework

[^oecd-ai]: OECD, **Recommendation of the Council on Artificial Intelligence**, adopted 2019 and updated 2024. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[^haip-overview]: OECD.AI, **Hiroshima AI Process Reporting Framework**, reporting and transparency resources. https://oecd.ai/en/transparency/overview

[^haip-v2]: OECD.AI, **OECD Launches Hiroshima AI Process Reporting Framework 2.0**, May 29, 2026. https://oecd.ai/en/haip-2-launch

[^haip-instructions]: OECD.AI, **How to Complete the HAIP Reporting Framework**, including eligibility across the advanced-AI value chain. https://oecd.ai/en/transparency/instructions

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

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

[^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-principles]: Council of Europe, **Framework Convention Lifecycle Principles**, including transparency and oversight, accountability and responsibility, equality, privacy, reliability, and safe innovation. https://www.coe.int/en/web/cddh-handbook-on-ai-and-hr/the-framework-convention-on-artificial-intelligence-and-human-rights-democracy-and-the-rule-of-law

[^iso-37000]: International Organization for Standardization, **ISO 37000:2021, Governance of Organizations, Guidance**. https://www.iso.org/standard/65036.html

[^iso-42001]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 42001:2023, Artificial Intelligence Management System**. https://www.iso.org/standard/81230.html

[^iso-27001]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 27001, Information Security Management Systems**. https://www.iso.org/isoiec-27001-information-security.html

[^iso-27701]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 27701, Privacy Information Management Systems**. https://www.iso.org/standard/85819.html

[^iso-directives]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC Directives, Part 1, Procedures for the Technical Work**. https://www.iso.org/sites/directives/current/consolidated/index.html

[^wto-principles]: World Trade Organization, **Principles for the Development of International Standards, Guides and Recommendations**. https://www.wto.org/english/tratop_e/tbt_e/principles_standards_tbt_e.htm

[^w3c-process]: World Wide Web Consortium, **W3C Process Document**. https://www.w3.org/policies/process/

[^ietf-process]: Internet Engineering Task Force, **Guide to the IETF Standards Process**. https://www.ietf.org/process/process/

[^un-guiding]: Office of the United Nations High Commissioner for Human Rights, **Guiding Principles on Business and Human Rights**, 2011. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

[^accessibility]: World Wide Web Consortium, **Web Content Accessibility Guidelines**. https://www.w3.org/WAI/standards-guidelines/wcag/

Revision Record

Version 1.0

Date: July 16, 2026

Change type: Complete foundational edition

Summary: Establishes the canonical Standards Body transparency, confidentiality, disclosure, access, correction, and public-accountability framework. Defines authority limits, transparency objectives, disclosure layers, five information classes, public-minimum rules, audience-based and proactive disclosure, transparency requests, governance, funding, conflict and partnership

disclosures, research, standards, evaluation, held-out evidence, independent review, evaluator and assurance records, decision and status transparency, incidents, complaints, appeals, corrections, sources, provenance, privacy, security, redaction, aggregation, timing, accessibility, translations, website governance, AI-assisted content, public communication, high-stakes domains, legal and international reporting, annual reports, metrics, audits, maturity, failure modes, objections, implementation, pilot design, scorecard, operational templates, canonical positions, cross-file interfaces, and primary research basis.

Status: Approved foundational source.