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

EVALUATOR_ACCREDITATION_FRAMEWORK.md

Standards Body Evaluator Accreditation Framework

Project: Standards Body
Core line: Foundations for Frontier AI
Document type: Canonical evaluator-accreditation, competence-recognition, conformity-assessment, oversight, surveillance, international-recognition, and institutional-design framework
Version: 1.0
Status: Approved foundational source
Document owner: Standards Body
Present institutional stage: Independent foundational research and institutional design
Applies to: Organizations and organizational units performing frontier AI testing, evaluation, inspection, audit, validation, verification, certification assessment, technical review, protected evaluation, red teaming, safeguard assessment, incident analysis, and related conformity-assessment activities
Related canonical sources: `PROJECT_IDENTITY.md`, `INSTITUTION_DESIGN.md`, `GOVERNANCE_FRAMEWORK.md`, `STANDARDS_DEVELOPMENT_PROCESS.md`, `TRANSPARENCY_FRAMEWORK.md`, `CONTRIBUTOR_FRAMEWORK.md`, `WEBSITE_SOURCE_OF_TRUTH.md`, `SOURCES.md`, `VERSION_HISTORY.md`, `TERMINOLOGY.md`, `TAXONOMY.md`, `EVIDENCE_STANDARDS.md`, `RESEARCH_METHODOLOGY.md`, `EVALUATION_PHILOSOPHY.md`, `FOUNDATIONS.md`, `FOUNDATIONS_APPENDIX.md`, Foundation 2, Foundation 3, Foundation 4, Foundation 5, Foundation 6, and Foundation 8
Research basis reviewed through: July 17, 2026
Review cycle: Annual complete review, quarterly standards-and-recognition review, and event-triggered revision after a material change in conformity-assessment standards, frontier AI evaluation methods, legal requirements, international accreditation arrangements, evaluator market structure, security practice, or Standards Body authority

---

Authority Note

This document develops a proposed framework for the future accreditation of frontier AI evaluator organizations.

It does not establish that Standards Body is currently:

- An accreditation body

- A national accreditation body
- A conformity-assessment body
- A certification body
- A governmental authority
- A notified body
- A formal peer-evaluation body
- A signatory to an international mutual-recognition arrangement
- Authorized to issue legally recognized accreditation
- Authorized to designate evaluators for regulatory conformity assessment
- Authorized to grant a mark implying government, international, or statutory recognition

Standards Body may presently:

- Research evaluator competence
- define proposed accreditation criteria
- develop scope models
- publish assessment templates
- design pilot qualification reviews
- convene expert review
- compare existing standards
- develop interoperability mappings
- support future accreditation-program design
- recommend safeguards for evaluator markets

Any present-stage evaluator review conducted by Standards Body should be described as:

- Evaluator qualification review
- pilot competence assessment
- readiness assessment
- research assessment
- accreditation-preparation review

It should not be described as accreditation.

The term **accredited evaluator** may be used only where a legitimate accreditation body has independently accredited the evaluator for a defined scope.

The accreditation body and scope should always be identified.

Canonical Accreditation Position

> **Accreditation should provide independent, scope-specific, time-bounded, and continuously monitored recognition that an evaluator organization is competent, impartial, secure, and consistently capable of performing defined conformity-assessment activities. It should never be treated as a universal judgment that the evaluator is trustworthy for every AI system, method, risk, jurisdiction, or decision.**

The defining accreditation rule is:

> **Accredit the function, method, competence, and scope, not the reputation of the organization.**

Document Purpose

This framework establishes the proposed institutional architecture for evaluator accreditation in frontier AI.

It defines:

- The purpose of evaluator accreditation
- the distinction between accreditation, certification, recognition, qualification, designation, and licensing
- the relationship among standards bodies, scheme owners, evaluators, certification bodies, accreditation bodies, governments, and international recognition systems
- the requirements for an accreditation body
- the requirements for an evaluator organization seeking accreditation
- the source standards and sector applications that may apply
- scope design
- competence models
- independence and impartiality
- quality management
- method validation
- evaluation security
- protected and held-out evaluations
- measurement and uncertainty
- personnel authorization
- assessment teams
- witnessing
- technical review
- decision independence
- initial accreditation
- surveillance
- reassessment
- proficiency testing
- inter-evaluator comparison
- complaints
- appeals
- scope extension
- suspension
- withdrawal
- registry publication
- marks and public claims

- cross-border recognition
- regulatory and legal interfaces
- pilot implementation
- maturity
- metrics
- failure modes
- operational templates

The framework is designed for an environment in which frontier AI evaluators may perform very different activities.

Examples include:

- Laboratory-style technical testing
- secure held-out capability evaluation
- red-team exercises
- inspection of systems and controls
- audit of organizational management systems
- validation or verification of claims
- certification assessment
- evaluation of safeguards
- evaluation of deployment context
- incident analysis
- continuous monitoring
- independent expert review

These activities should not be placed under one generic "AI auditor" label.

Each function requires an appropriate conformity-assessment model, competence basis, scope, review method, and decision structure.

Executive Summary

Frontier AI evaluation is becoming too consequential to depend solely on informal reputation.

An evaluator may be technically skilled and still lack:

- A controlled quality system
- independence
- secure task custody
- reproducible methods
- qualified personnel
- change control
- complaint handling
- surveillance
- public scope boundaries
- international legibility

Accreditation is a possible institutional response.

Accreditation is not a badge for general excellence.

It is independent recognition that a conformity-assessment body is competent and impartial to perform specified activities within a defined scope.

The scope is the center of the accreditation.

A frontier AI evaluator might be competent to:

- Administer a defined cyber-capability evaluation protocol
- test model robustness under specified conditions
- inspect deployment controls
- verify a defined technical claim
- audit an AI management system
- perform a protected evaluation in a secure facility

That evaluator may not be competent to:

- Evaluate biological-risk capabilities
- interpret constitutional rights
- certify an AI product
- assess every model modality
- perform work in every jurisdiction
- evaluate systems outside the listed access mode
- issue a universal safety conclusion

Why Frontier AI Accreditation Is Distinct

Frontier AI evaluation creates unusual accreditation challenges.

Rapid Method Change

Methods may become obsolete within months.

Dynamic Objects

Model and system versions change while retaining the same public product name.

Protected Evidence

Held-out tasks, model access, security findings, and severe-risk results may require restricted handling.

Strategic Behavior

Systems or providers may optimize for known tests, limit access, or shape evaluation conditions.

High Consequence

Errors may affect deployment decisions, procurement, regulation, public safety, or national security.

Interdisciplinary Competence

A credible assessment may require:

- Machine learning
- cybersecurity
- statistics
- human factors
- domain science
- law
- rights
- organizational governance
- security operations

Market Pressure

Client-paid evaluation can create:

- Evaluator shopping
- low-bid pressure
- scope inflation
- favorable interpretation
- dependence on a small number of developers

Cross-Border Use

An evaluator may work in one jurisdiction while its results are relied upon elsewhere.

Accreditation Architecture

The preferred long-term architecture is distributed.

Standards Body or Other Standard Setter

Develops standards and technical requirements.

Scheme Owner

Defines the rules for a particular conformity-assessment program.

Evaluator or Conformity-Assessment Body

Performs testing, inspection, audit, validation, verification, or certification activity.

Accreditation Body

Assesses and recognizes the evaluator's competence and impartiality for a defined scope.

Peer-Evaluation and Recognition Network

Evaluates accreditation bodies and supports cross-border acceptance.

Government or Regulator

Determines legal recognition, designation, notification, procurement, or enforcement.

No one actor should perform all functions without strong separation and independent review.

Existing Conformity-Assessment Baseline

The proposed framework should build on the existing conformity-assessment architecture rather than inventing accreditation from first principles.

Relevant current standards include:

- ISO/IEC 17011:2017 for accreditation bodies
- ISO/IEC 17025:2017 for testing and calibration laboratories
- ISO/IEC 17020:2026 for inspection bodies
- ISO/IEC 17021-1:2015 for bodies auditing and certifying management systems
- ISO/IEC 42006:2025 for bodies auditing and certifying AI management systems
- ISO/IEC 17029:2019 for validation and verification bodies
- ISO/IEC 17065:2012 for bodies certifying products, processes, and services, noting that a replacement was under development as of the review date
- ISO/IEC 17024:2026 for bodies certifying persons
- ISO/IEC 17043:2023 for proficiency-testing providers
- ISO/IEC 17030:2021 for third-party marks of conformity
- ISO/IEC 17040:2005 for peer assessment
- ISO 19011:2018 for management-system auditing guidance

ISO/IEC 42006:2025 is especially important because it adds AI-specific requirements to the audit and certification of AI management systems under ISO/IEC 42001.[^iso42006]

Management-system certification is not the same as technical accreditation of frontier AI evaluation laboratories.

A certified AI management system does not establish that a specific AI system is safe.

A technically accredited evaluator does not establish that its client conforms to a management-system standard.

The accreditation framework should preserve those distinctions.

Proposed Accreditation Pathways

The framework supports separate pathways for:

1. Technical evaluation laboratories
2. inspection bodies
3. management-system certification bodies
4. validation and verification bodies
5. product, process, or service certification bodies
6. personnel certification bodies
7. proficiency-testing providers
8. hybrid organizations with separately controlled functions

An organization may hold more than one accreditation.

Each accreditation should have a separate scope and applicable requirements.

Core Accreditation Requirements

An evaluator should demonstrate:

- Legal and organizational identity
- defined scope
- impartiality
- independence
- competent personnel
- validated methods
- secure infrastructure
- controlled evidence
- measurement integrity
- reproducibility
- quality management
- documented decisions
- complaint handling
- corrective action
- proficiency
- change control
- surveillance readiness
- public claims discipline

Accreditation Decision

The people who assess the evaluator should not be the only people who make the accreditation decision.

The decision should be made by a competent, conflict-screened function that reviews the assessment record.

Surveillance

Accreditation should not be permanent.

It should be monitored through:

- Scheduled surveillance
- witness assessments
- document review
- proficiency testing
- complaint review
- incident reporting
- source and method updates
- personnel changes
- client-concentration review
- task-compromise review
- unannounced or short-notice review where justified

International Recognition

On January 1, 2026, Global Accreditation Cooperation Incorporated began operations as the single international accreditation organization formed from the work of the International Accreditation Forum and the International Laboratory Accreditation Cooperation, including a new multilateral recognition arrangement.[^globalaci]

A future frontier AI accreditation ecosystem should seek compatibility with legitimate national accreditation bodies and international peer-evaluation structures.

Standards Body should not create a competing global mark without:

- Legal basis
- institutional independence
- conformity with ISO/IEC 17011
- peer evaluation
- international recognition
- competent staffing
- public-interest governance
- operational maturity

European Union Interface

The European Union Artificial Intelligence Act provides a concrete legal example of accreditation and designation interacting.

Member States designate notifying authorities.

Those authorities may use national accreditation bodies to assess and monitor conformity-assessment bodies.

Applicants for notification may provide accreditation certificates showing conformity with notified-body requirements.

The Act requires notified bodies to demonstrate legal personality, quality management, resources, cybersecurity, independence, impartiality, competence, confidentiality, insurance, and sufficient technical, legal, and scientific personnel.[^eu-ai-act-notified]

This framework does not assume that every frontier AI evaluator will become an EU notified body.

It uses the example to demonstrate that:

- Accreditation and regulatory designation are distinct
- accreditation may support notification
- a government may retain designation authority
- scope and continuous monitoring matter
- certificates may be suspended or withdrawn
- evaluator capacity affects regulatory implementation

Preferred Institutional Strategy

Standards Body should not begin by issuing accreditation.

The preferred sequence is:

1. Publish the framework.
2. define evaluator scopes.
3. create competence matrices.
4. conduct readiness pilots.
5. partner with qualified existing accreditation bodies.
6. develop sector-specific application documents.
7. establish proficiency-testing programs.
8. test surveillance and appeals.
9. support international crosswalks.
10. consider a separate accreditation function only after independent governance and readiness review.

The most credible first outcome is not an accreditation mark.

It is a technically serious accreditation framework that an established accreditation body can inspect, challenge, adapt, and potentially use.

---

1. Foundational Accreditation Propositions

1.1 Competence Is Scope-Specific

An evaluator is not competent in the abstract.

Competence exists for defined activities, methods, systems, domains, conditions, and decisions.

1.2 Accreditation Is Independent Recognition

An evaluator cannot accredit itself.

A scheme owner should not treat its own internal approval as accreditation.

1.3 Accreditation Is Not Certification of the Evaluated System

Accreditation recognizes the evaluator.

Certification attests conformity of a different object under a certification scheme.

1.4 Accreditation Is Not Regulatory Designation

A regulator or government may recognize, designate, notify, license, or authorize an evaluator.

Accreditation may support that decision.

It does not replace it automatically.

1.5 Accreditation Should Be Time-Bounded

Competence, methods, personnel, systems, and risks change.

Accreditation should require surveillance and reassessment.

1.6 Independence and Competence Are Both Necessary

Technical skill does not cure conflict of interest.

Independence does not cure technical weakness.

1.7 Method Validity Is Central

An evaluator should not receive accreditation merely because it follows documented procedures.

The methods should be fit for their claimed purpose.

1.8 Protected Evaluation Requires Governed Security

Accreditation should include task custody, access control, incident response, and compromise management where protected evidence is used.

1.9 Accreditation Should Support Pluralism

Accreditation should allow multiple competent evaluators rather than granting monopoly status without justification.

1.10 Accreditation Should Resist Market Capture

Revenue dependence, developer access control, consulting, and client concentration require explicit controls.

1.11 Public Scope Should Be Precise

Every public accreditation claim should identify:

- Accreditation body
- evaluator
- scope
- applicable standard
- location
- version
- validity
- restrictions

1.12 International Recognition Requires Institutional Work

Cross-border acceptance should arise from compatible standards, peer evaluation, and recognition arrangements.

It should not arise from marketing language.

1.13 Accreditation Does Not Establish Universal Safety

An evaluator may be accredited to perform an assessment.

The assessment may still be limited.

1.14 Accreditation Should Be Correctable

Scope may be reduced, suspended, withdrawn, or corrected.

1.15 Public Interest Is Part of Accreditation Design

Evaluator competence affects people who may never become the evaluator's client.

1.16 Accreditation Should Encourage Method Improvement

The framework should not freeze obsolete methods.

1.17 Accreditation Should Preserve Dissent and Appeals

Applicants, clients, affected parties, and evaluators should have defined routes to challenge decisions.

1.18 Accreditation Claims Should Be Machine-Verifiable

Public registries should support verification of current scope and status.

---

2. Scope and Boundaries

2.1 Covered Organizations

- Independent evaluation laboratories
- commercial assurance providers
- nonprofit evaluators
- academic consortia
- government laboratories
- inspection bodies
- management-system certification bodies
- validation and verification bodies
- product or process certification bodies
- sector-specific technical evaluators
- secure-evaluation facilities
- proficiency-testing providers
- hybrid conformity-assessment organizations

2.2 Covered Activities

- Testing
- evaluation
- inspection
- audit
- validation
- verification
- certification assessment
- red teaming
- safeguard evaluation
- capability evaluation

- deployment assessment
- incident investigation
- evidence review
- continuous monitoring

2.3 Excluded Claims

This framework does not itself:

- Accredit an organization
- certify an AI system
- approve deployment
- create legal compliance
- replace national accreditation
- designate notified bodies
- create a global mutual-recognition arrangement
- determine regulatory penalties
- guarantee safety
- eliminate evaluator judgment

2.4 Personnel Versus Organization

This framework primarily concerns organizations or defined organizational units.

Individual competence may be:

- Authorized internally
- certified under a personnel-certification scheme
- registered professionally
- assessed through training and performance

Individual certification does not replace organizational accreditation.

2.5 Scheme-Specific Requirements

An accreditation body should apply:

- General accreditation-body requirements
- the applicable conformity-assessment-body standard
- AI-specific sector requirements
- scheme requirements
- jurisdictional requirements
- scope-specific technical criteria

2.6 Voluntary and Regulatory Use

The framework may support:

- Voluntary market recognition
- procurement
- insurance
- developer qualification
- regulatory designation
- legal conformity assessment
- research partnerships

Legal effect depends on the applicable authority.

---

3. Canonical Definitions

Definitions in `TERMINOLOGY.md` govern.

3.1 Accreditation

Independent recognition that a conformity-assessment body is competent and impartial to perform specified activities within a defined scope.

3.2 Accreditation Body

A body that performs accreditation.

3.3 Evaluator

A person or organization that designs, conducts, administers, interprets, or reviews an evaluation.

3.4 Accredited Evaluator

An evaluator accredited for a defined scope by a legitimate accreditation body.

The scope should never be omitted.

3.5 Conformity-Assessment Body

A body performing testing, inspection, certification, validation, verification, or another defined conformity-assessment activity.

3.6 Evaluation Laboratory

An organization or unit performing technical evaluation under controlled methods and quality systems.

3.7 Scope of Accreditation

The specific activities, methods, domains, systems, locations, and limits for which a body is recognized as competent.

3.8 Scheme

The rules, procedures, criteria, governance, claims, decisions, and responsibilities for a conformity-assessment program.

3.9 Scheme Owner

The organization responsible for a scheme.

3.10 Recognition

Acceptance of evidence, competence, process, status, or decision for a defined purpose.

3.11 Qualification

A determination that a person or organization meets defined prerequisites or competence criteria.

Qualification is not accreditation unless performed under a legitimate accreditation process.

3.12 Designation

Formal selection or authorization by a competent authority for a specified role.

3.13 Notification

A formal governmental communication recognizing a conformity-assessment body for activities under a legal framework.

3.14 Certification

Third-party attestation that specified requirements have been fulfilled within a defined certification scheme.

3.15 Testing

Determination of characteristics of an object according to a procedure.

3.16 Inspection

Examination of an object, process, installation, system, or service and determination of conformity with requirements or professional judgment.

3.17 Audit

A systematic, independent, and documented process for obtaining evidence and evaluating it against defined criteria.

3.18 Validation

Confirmation that requirements for a specific intended use or application have been fulfilled.

3.19 Verification

Confirmation that specified requirements have been fulfilled.

3.20 Impartiality

Presence of objectivity and management of interests that could compromise judgment.

3.21 Independence

The degree to which the evaluator's organization, governance, funding, method, information, operation, and publication are free from inappropriate control.

3.22 Competence

Demonstrated ability to apply knowledge and skills to achieve intended results.

3.23 Witness Assessment

Observation of the evaluator performing an activity within the proposed or accredited scope.

3.24 Surveillance

Continuing monitoring of an accredited body to confirm ongoing fulfillment of accreditation requirements.

3.25 Reassessment

A complete or substantial periodic review for renewal of accreditation.

3.26 Proficiency Testing

Evaluation of participant performance against pre-established criteria through interlaboratory or inter-evaluator comparison.

3.27 Flexible Scope

A scope allowing controlled introduction or modification of methods within defined boundaries without prior approval for every individual change.

3.28 Accreditation Decision

The independent decision to grant, maintain, extend, reduce, suspend, restore, or withdraw accreditation.

3.29 Accreditation Certificate

A formal record identifying the accredited body, accreditation body, scope, standard, locations, validity, and status.

3.30 Accreditation Mark

A controlled symbol indicating accredited status within defined rules.

3.31 Evaluator Shopping

Selection or repeated switching of evaluators to obtain a preferred result rather than a more valid assessment.

3.32 Accreditation Shopping

Seeking the least demanding accreditation pathway, jurisdiction, or assessor while presenting the result as equivalent to stronger recognition.

3.33 Scope Inflation

Public or operational expansion of accredited claims beyond the actual accredited scope.

3.34 Accreditation Capture

Improper influence over accreditation criteria, assessment, decisions, surveillance, or recognition by regulated, accredited, commercial, political, or institutional interests.

---

4. Institutional Role Architecture

4.1 Standard Setter

Develops normative or technical standards.

A standard setter may define requirements suitable for conformity assessment.

It should not automatically accredit organizations against its own standards.

4.2 Scheme Owner

Defines:

- Object of assessment
- eligibility
- requirements
- methods
- sampling
- decisions
- surveillance
- claims
- marks
- appeals
- maintenance

A scheme owner may be Standards Body, a government, a standards organization, an industry consortium, or another competent institution.

Scheme ownership does not create accreditation authority.

4.3 Evaluator Organization

Performs the activity within its competence and authorized scope.

4.4 Certification Body

Makes certification decisions under a defined scheme.

It may rely on testing, inspection, evaluation, or audit evidence.

4.5 Accreditation Body

Assesses whether the conformity-assessment body is competent and impartial for the defined scope.

4.6 Peer-Evaluation Body

Assesses accreditation bodies for compliance with recognized requirements and supports mutual recognition.

4.7 Government or Regulator

May:

- Recognize standards
- designate evaluators
- notify conformity-assessment bodies
- require accreditation
- adopt results
- enforce law
- restrict or withdraw authorization

4.8 Registry Operator

Maintains current status and scope records.

4.9 Proficiency-Testing Provider

Designs and administers inter-evaluator comparisons.

4.10 Reference-Asset Custodian

Maintains reference tasks, environments, datasets, scoring keys, incident cases, or other comparison assets.

4.11 Appeals Function

Reviews eligible accreditation decisions independently.

4.12 Public-Interest Function

Represents effects on:

- Affected communities
- rights

- labor
- accessibility
- safety
- competition
- public services
- environmental interests

4.13 Role-Separation Rule

The following functions should not be combined without explicit conflict controls:

- Consulting and accreditation assessment
- evaluation and accreditation decision
- scheme ownership and appeal
- client account management and technical decision
- standards drafting and exclusive commercial certification
- accreditation and evaluator operation
- proficiency-test design and participant result manipulation
- regulatory enforcement and paid consulting

4.14 Small-Ecosystem Exception

During early ecosystem development, one institution may perform several functions.

It should establish:

- Organizational separation
- separate personnel
- separate budgets
- decision recusal
- external review
- public disclosure
- transition plan

4.15 Preferred Standards Body Role

The preferred present and near-term role for Standards Body is:

- Research institution
- framework developer
- standards-process designer
- possible future scheme owner
- public registry and knowledge-infrastructure designer
- pilot coordinator

It should not become the evaluator accreditation body merely because it authored this framework.

5. Accreditation-Body Requirements

A body performing accreditation under this framework should meet the principles and requirements of ISO/IEC 17011:2017 and applicable law.[^iso17011]

5.1 Legal Identity

The accreditation body should possess:

- Legal personality
- documented authority
- clear ownership
- liability arrangements
- continuity plan

5.2 Accreditation Authority

The body should state:

- Source of authority
- jurisdiction
- voluntary or legal status
- functions
- limits
- recognition arrangements
- prohibited claims

5.3 Impartial Governance

The accreditation body should protect decisions from:

- Accredited clients
- prospective clients
- developers
- scheme owners
- funders
- governments outside lawful authority
- consultants
- commercial affiliates
- assessors
- founders

5.4 No Competitive Conformity Assessment

The accreditation body should not offer the same evaluator services it accredits on a competitive basis.

5.5 Financial Stability

The body should maintain sufficient resources to:

- Assess applicants
- conduct surveillance
- investigate complaints
- handle appeals
- maintain records
- secure protected information
- preserve continuity after withdrawal or failure

5.6 Competent Personnel

The body should have access to competence in:

- Accreditation
- conformity assessment
- frontier AI systems
- evaluation science
- statistics and measurement
- cybersecurity
- protected information
- domain risks
- law and fundamental rights
- organizational governance
- scheme operation

5.7 Decision Independence

Persons making accreditation decisions should not have performed the complete assessment alone.

5.8 Confidentiality

The body should protect:

- Applicant methods
- held-out tasks
- system access
- client information

- incident evidence
- trade secrets
- personal information
- assessor deliberation

5.9 Public Accountability

The body should publish:

- Accreditation process
- eligibility
- fees
- scope rules
- status definitions
- complaint and appeal routes
- current accredited bodies
- suspensions and withdrawals
- use-of-mark rules
- governance and conflicts
- international recognition status

5.10 Quality Management

The accreditation body should maintain:

- Document control
- record control
- personnel competence
- assessor authorization
- decision review
- internal audit
- management review
- corrective action
- risk management
- information security
- business continuity

5.11 Peer Evaluation

A body seeking international recognition should undergo qualified peer evaluation.

ISO/IEC 17040 provides general requirements for peer assessment of conformity-assessment bodies and accreditation bodies.[^iso17040]

5.12 Scheme Acceptance

Before accrediting against a scheme, the body should determine whether the scheme is:

- Clear
- assessable
- technically valid
- governed
- maintained
- legally usable
- nondeceptive
- sufficiently resourced

5.13 Public-Interest Governance

The body should consider impacts on parties beyond applicant organizations.

5.14 Accreditation-Body Failure

The body should have a plan for:

- Loss of recognition
- insolvency
- cyber incident
- corruption
- legal restriction
- merger
- dissolution
- transfer of accreditation records

---

6. Applicable Conformity-Assessment Pathways

A frontier AI evaluator should be accredited through the pathway matching its actual function.

6.1 Technical Testing and Evaluation Laboratory

Primary baseline:

- ISO/IEC 17025:2017
- AI-specific application criteria
- scheme-specific methods
- protected-evaluation controls where applicable

ISO describes ISO/IEC 17025 as the international standard for testing and calibration laboratories, with requirements for competence, impartiality, and consistent operation.[^iso17025]

Appropriate activities may include:

- Capability testing
- robustness testing
- benchmark administration
- quantitative safeguard testing
- system-performance measurement
- red-team task administration
- model-judge validation
- human-baseline measurement

6.2 Inspection Body

Primary baseline:

- ISO/IEC 17020:2026
- AI-specific inspection criteria

ISO/IEC 17020:2026 sets requirements for competence, impartiality, and consistent operation of inspection bodies.[^iso17020]

Appropriate activities may include:

- Inspection of deployment controls
- inspection of secure facilities
- inspection of incident-response implementation
- inspection of monitoring systems
- inspection of data-governance controls
- conformity determination using professional judgment

6.3 Management-System Certification Body

Primary baseline:

- ISO/IEC 17021-1:2015
- ISO/IEC 42006:2025
- ISO/IEC 42001:2023
- applicable accreditation-body application documents

ISO/IEC 42006 establishes AI-specific additional requirements for bodies auditing and certifying AI management systems under ISO/IEC 42001.[^iso42006]

This pathway concerns organizational management systems.

It should not be represented as direct certification of the safety or capability of a specific model.

6.4 Validation and Verification Body

Primary baseline:

- ISO/IEC 17029:2019
- ISO/IEC TS 17035:2024 where useful
- scheme-specific requirements

Appropriate claims may concern:

- Emissions or resource claims
- risk-management claims
- evaluation-process claims
- reported governance controls
- incident-remediation claims
- evidence-package completeness

ISO/IEC 17029 addresses competence, consistent operation, and impartiality for validation and verification bodies.[^iso17029]

6.5 Product, Process, or Service Certification Body

Primary baseline:

- ISO/IEC 17065:2012
- applicable successor when published and effective
- ISO/IEC 17067 for scheme design
- AI-specific certification criteria

As of July 17, 2026, ISO/IEC 17065:2012 remained current, while a final draft replacement was under development.[^iso17065]

This pathway may become relevant for certification of:

- AI-enabled services
- defined AI processes
- specific system configurations
- safeguard packages
- evaluation services

It requires a valid certification scheme.

6.6 Personnel Certification Body

Primary baseline:

- ISO/IEC 17024:2026
- defined personnel-certification scheme

Possible personnel schemes:

- Frontier AI evaluation lead
- protected-task custodian
- technical assessor
- automated-judge validation specialist
- severe-risk domain evaluator

Personnel certification should not substitute for organizational accreditation.

6.7 Proficiency-Testing Provider

Primary baseline:

- ISO/IEC 17043:2023
- AI-specific inter-evaluator comparison design

Possible programs:

- Common task packages
- scoring calibration
- incident classification
- secure administration
- replication exercises
- evaluator consistency studies

6.8 Hybrid Organization

A hybrid organization may perform several activities.

It should maintain:

- Separate scopes
- separate applicable standards
- function-specific competence
- conflict controls
- independent decisions
- transparent public claims

6.9 Pathway Determination

The accreditation body should classify the actual activity before selecting the standard.

The label preferred by the applicant should not control the pathway.

---

7. Normative and Reference Stack

7.1 Core Accreditation Standard

- ISO/IEC 17011:2017

7.2 Conformity-Assessment-Body Standards

- ISO/IEC 17025:2017
- ISO/IEC 17020:2026
- ISO/IEC 17021-1:2015
- ISO/IEC 17029:2019
- ISO/IEC 17065:2012 and applicable successor
- ISO/IEC 17024:2026
- ISO/IEC 17043:2023

7.3 AI-Specific Standards

- ISO/IEC 42001:2023
- ISO/IEC 42005:2025
- ISO/IEC 42006:2025
- ISO/IEC 23894:2023
- ISO/IEC 25059:2023
- ISO/IEC 24029 series
- ISO/IEC TS 42119-2:2025
- future ISO/IEC 42007 when published
- relevant data-quality, security, privacy, and governance standards

7.4 Supporting Conformity-Assessment Standards

- ISO/IEC 17000:2020
- ISO/IEC 17030:2021
- ISO/IEC 17040:2005
- ISO/IEC 17060:2022
- ISO 19011:2018
- ISO/IEC TS 17012:2024 for remote auditing where applicable

7.5 Standards Body Canonical Requirements

- `EVIDENCE_STANDARDS.md`
- `RESEARCH_METHODOLOGY.md`
- `EVALUATION_PHILOSOPHY.md`
- Foundation 1
- Foundation 2

- Foundation 3
- Foundation 4
- Foundation 5
- Foundation 8
- scheme-specific standards

7.6 Legal Requirements

Applicable law may add:

- Designation
- notification
- licensing
- insurance
- residency
- data localization
- professional secrecy
- regulatory reporting
- fundamental-rights competence
- cybersecurity
- public procurement
- market surveillance

7.7 Standards Currency

The accreditation body should monitor changes to applicable standards.

A revision may trigger:

- Gap analysis
- transition plan
- assessor training
- evaluator migration
- scope review
- public notice

7.8 No Checklist Substitution

Conformity to a general standard should not replace AI-specific method assessment.

---

8. Scope of Accreditation

8.1 Scope Principle

The scope should allow a competent user to understand exactly what the evaluator is recognized to do.

8.2 Required Scope Dimensions

Activity

- Testing
- evaluation
- inspection
- audit
- validation
- verification
- certification assessment
- proficiency testing

Object

- Model
- system
- agent
- deployment
- organization
- management system
- safeguard
- process
- service
- evidence package

System Class

- General-purpose model
- multimodal model
- open-weight model
- agentic system
- domain-specific system
- robotics system
- safety-critical system

Capability or Risk Domain

- Cyber
- biological

- chemical
- radiological
- autonomy
- persuasion
- fraud
- privacy
- discrimination
- critical infrastructure
- scientific research
- AI research acceleration

Method

- Named protocol
- method family
- standard
- laboratory-developed method
- inspection procedure
- audit scheme

Access Mode

- Public interface
- API
- privileged API
- weights
- internal logs
- secure enclave
- on-site access
- post-deployment monitoring

Lifecycle Stage

- Research
- predeployment
- deployment
- update
- incident
- retirement

Assurance Level

- Preliminary
- supported
- substantiated
- decision-grade
- scheme-specific assurance category

Location

- Physical site
- virtual facility
- mobile activity
- remote activity
- regional unit

Jurisdiction

- Voluntary international
- named jurisdiction
- regulatory scheme
- procurement scheme

Security Level

- Public evaluation
- controlled evaluation
- restricted evaluation
- highly restricted evaluation

Decision Type

- Measurement report
- inspection finding
- verification statement
- certification recommendation
- conformity decision

8.3 Example Scope Statement

> Technical evaluation of general-purpose and agentic AI systems through controlled API access, using listed capability and safeguard protocols in cyber operations, autonomous task completion, and software engineering, excluding biological-risk evaluation, legal compliance certification, and unrestricted product safety claims.

8.4 Scope Exclusions

Exclusions should be explicit.

8.5 Fixed Scope

A fixed scope lists each approved method or activity.

Appropriate when:

- Methods are immature

- consequences are high
- security is sensitive
- evaluator experience is limited

8.6 Flexible Scope

A flexible scope may permit controlled method introduction within defined boundaries.

It requires:

- Mature method validation
- competent authorization
- change control
- internal approval
- retrospective reporting
- surveillance
- no expansion into a new risk domain without approval

8.7 Scope Extension

Requires assessment proportionate to the new:

- Domain
- method
- system class
- location
- access mode
- security level
- decision authority

8.8 Scope Reduction

May be voluntary or imposed.

8.9 Scope Suspension

Suspension may apply to:

- One method
- one domain
- one site
- one personnel group
- one security level
- the entire accreditation

8.10 Public Scope Registry

The registry should show current and historical scope.

---

9. Organizational Competence

9.1 Competence Model

Organizational competence combines:

- Personnel
- methods
- infrastructure
- data
- security
- governance
- quality system
- decision process
- experience
- learning

9.2 Legal and Organizational Identity

The evaluator should disclose:

- Legal entity
- ownership
- controlling interests
- affiliates
- locations
- subcontractors
- governance
- insurance
- insolvency and continuity arrangements

9.3 Technical Leadership

The evaluator should designate responsible technical leadership for each accredited domain.

9.4 Method Ownership

The evaluator should know:

- Who owns each method

- who approves changes
- who may administer it
- who interprets results
- how validity is monitored

9.5 Operational Capacity

The evaluator should demonstrate capacity to complete work at required quality without excessive overload.

9.6 Experience

Experience may be demonstrated through:

- Completed evaluations
- pilot assessments
- witnessed work
- method-development history
- replication
- incident handling
- proficiency performance

9.7 Knowledge Management

The evaluator should maintain current knowledge of:

- Model and system developments
- evaluation science
- security threats
- standards
- law
- incidents
- benchmark compromise
- domain risks

9.8 Competence Gaps

The organization should identify and control gaps.

9.9 Continuing Competence

Use:

- Training
- supervised practice

- witness review
- periodic authorization
- proficiency testing
- technical discussion
- external review
- incident learning

9.10 Loss of Competence

The evaluator should report material loss of:

- Key personnel
- secure facility
- method validity
- system access
- insurance
- legal authority
- infrastructure
- independence

---

10. Personnel Competence

10.1 Role-Based Competence

Competence should be defined for:

- Evaluation designer
- task author
- administrator
- operator
- scorer
- automated-judge specialist
- statistician
- technical reviewer
- domain expert
- security custodian
- decision maker
- audit lead
- inspector
- report approver

10.2 Competence Components

- Education

- training
- knowledge
- practical skill
- experience
- judgment
- conduct
- security eligibility
- communication
- continuing performance

10.3 Competence Matrix

Map each role to:

- Required competencies
- evidence
- evaluator
- authorization
- limits
- renewal
- supervision

10.4 Domain Competence

High-risk domains require genuine subject-matter competence.

General machine-learning expertise does not establish competence in:

- Biology
- chemistry
- cybersecurity operations
- nuclear systems
- medicine
- finance
- critical infrastructure
- constitutional rights

10.5 Evaluation Science Competence

Personnel should understand:

- Constructs
- validity
- reliability
- uncertainty
- elicitation

- contamination
- baselines
- scoring
- generalization
- limitations

10.6 System Competence

Personnel should understand:

- Model architecture
- system configuration
- tools
- scaffolds
- memory
- safeguards
- deployment
- version identity
- access effects

10.7 Security Competence

Protected work requires competence in:

- Access control
- secret handling
- threat modeling
- incident response
- secure environments
- chain of custody
- responsible disclosure

10.8 Legal and Rights Competence

Where scope includes legal or rights claims, teams need competent legal or rights expertise.

10.9 Authorization

Qualified personnel should receive documented authorization for defined activities.

10.10 Supervised Personnel

Trainees may participate under documented supervision.

10.11 Personnel Rotation

Rotation may reduce capture and single-person dependence.

10.12 External Experts

External experts should meet the same relevant competence, conflict, confidentiality, and security requirements.

---

11. Impartiality and Independence

11.1 Impartiality Risk Analysis

The evaluator should maintain a current analysis of threats from:

- Ownership
- governance
- funding
- client revenue
- consulting
- employment
- shared personnel
- intellectual commitments
- future work
- political pressure
- system access dependence
- publication control
- litigation
- competition

11.2 Client Concentration

High dependence on one developer or client should trigger enhanced controls.

11.3 Result-Dependent Compensation

Compensation should not depend on:

- Passing result
- favorable rating
- certification
- absence of findings
- continued deployment
- avoided regulatory action

11.4 Consulting

The evaluator should not assess work it designed without strong separation and disclosed safeguards.

For certification functions, applicable standards may impose stricter limits.

11.5 Access Dependence

A developer should not be able to threaten unrelated access to influence a finding.

11.6 Publication Independence

The evaluator should have defined rights to report findings under the engagement and scheme.

Protected information may limit public detail.

It should not permit falsification.

11.7 Independent Decision Function

Certification or conformity decisions should be made by persons not responsible for the complete assessment activity.

11.8 Impartiality Committee

A mature evaluator may use an external or balanced impartiality function.

11.9 Conflict Controls

- Disclosure
- monitoring
- reviewer addition
- recusal
- team replacement
- scope limitation
- independent replication
- client reallocation
- refusal

11.10 Public Conflict Summary

Material conflicts should be disclosed at a safe level.

11.11 Independence Profile

The accreditation record may summarize:

- Organizational independence
- financial independence
- methodological independence
- operational independence
- publication independence
- security independence

11.12 Structural Failure

Where threats cannot be mitigated, the evaluator should not perform the activity.

---

12. Quality Management System

12.1 Quality Policy

The evaluator should define its commitment to:

- Competence
- impartiality
- valid methods
- accurate reporting
- security
- correction
- public-interest responsibility
- continual improvement

12.2 Document Control

Control:

- Protocols
- methods
- rubrics
- system manifests
- templates
- standards
- policies
- software
- task specifications
- scoring keys
- release notes

12.3 Record Control

Preserve:

- Engagement records
- raw results
- system identity
- prompts and configurations
- access conditions
- personnel
- scoring
- uncertainty
- reviews
- decisions
- incidents
- corrections
- client communications

12.4 Retention

Retention should reflect:

- Scheme requirements
- legal requirements
- certificate or result lifetime
- appeal period
- incident risk
- privacy
- security
- research value

12.5 Personnel Control

Maintain:

- Competence records
- authorizations
- training
- performance
- conflicts
- confidentiality
- access
- discipline
- departure

12.6 Method Control

Every accredited method should have:

- Identifier
- version
- owner
- purpose
- validation
- limitations
- approved users
- change history
- retirement rules

12.7 Equipment and Infrastructure

Control:

- Compute
- networks
- secure environments
- test harnesses
- software
- time sources
- storage
- logging
- physical facilities
- remote systems

12.8 External Providers

Assess:

- Cloud providers
- software vendors
- model providers
- task suppliers
- contractors
- translators
- security services
- data sources

12.9 Nonconforming Work

The evaluator should identify work that:

- Used the wrong system version
- used a compromised task
- departed from protocol
- lacked authorized personnel
- lost evidence
- used invalid scoring
- experienced security compromise
- contained material reporting error

12.10 Control of Nonconforming Work

Actions may include:

- Pause
- segregation
- client notice
- review
- rerun
- correction
- result invalidation
- regulator notice
- accreditation-body notice

12.11 Corrective Action

The organization should:

1. Contain the problem.
2. assess cause.
3. identify affected work.
4. correct.
5. prevent recurrence.
6. verify effectiveness.
7. notify where required.

12.12 Internal Audit

Internal audit should test both documented conformity and actual technical practice.

12.13 Management Review

Review:

- Performance
- complaints
- appeals
- proficiency
- incidents
- client concentration
- competence
- methods
- resources
- security
- improvement
- accreditation status

12.14 Quality Culture

The evaluator should reward:

- Error reporting
- correction
- dissent
- careful uncertainty
- refusal of invalid work

It should not reward only throughput or client satisfaction.

---

13. Method Validation and Control

13.1 Method Categories

- Standard method
- scheme method
- modified method
- laboratory-developed method
- client-specified method
- research method
- emergency method
- retired method

13.2 Method Validation Questions

- What construct is measured?

- What object is evaluated?
- What decision will use the result?
- What evidence supports validity?
- What reliability is expected?
- What uncertainty remains?
- What failure modes exist?
- What access assumptions apply?
- What system versions are covered?
- What task exposure is possible?
- What scoring method is used?
- What safety controls apply?

13.3 Standard Methods

A published standard method still requires:

- Correct implementation
- competent administration
- local verification
- version control
- equipment and software validation
- scope fit

13.4 Modified Methods

The evaluator should validate the effect of modifications.

13.5 Laboratory-Developed Methods

Require stronger evidence and review.

13.6 Elicitation Validation

Capability evaluation should state whether the method reasonably approaches the capability ceiling relevant to the decision.

13.7 Automated Judge Validation

Record:

- Judge model
- version
- rubric
- prompt
- calibration

- human comparison
- disagreement
- bias
- failure conditions
- replacement trigger

13.8 Human Scoring

Control:

- Scorer competence
- blinding
- rubric
- calibration
- inter-rater reliability
- adjudication
- fatigue
- conflict

13.9 Software Validation

Validate:

- Evaluation harness
- data pipelines
- scoring code
- logging
- environment
- reproducibility
- change control

13.10 Benchmark Integrity

Assess:

- Public exposure
- contamination
- solution availability
- repeated optimization
- saturation
- strategic behavior
- task retirement

13.11 Method Verification Before Use

Before each engagement, verify:

- Current method version
- authorized personnel
- correct system access
- task integrity
- expected environment
- safety readiness

13.12 Method Change

Classify as:

- Editorial
- nonbreaking
- substantive
- breaking
- emergency

13.13 Comparability

A change should not be described as comparable without evidence.

13.14 Method Registry

The accreditation body should maintain the methods included within each scope.

13.15 Method Sunset

Retire methods that are:

- Invalid
- compromised
- obsolete
- unsafe
- unsupported
- no longer decision-relevant

---

14. Measurement, Evidence, and Uncertainty

14.1 Evidence Standard

Accredited work should follow `EVIDENCE_STANDARDS.md`.

14.2 Measurement Model

The evaluator should identify:

- Construct
- indicator
- metric
- scale
- baseline
- threshold
- uncertainty
- interpretation

14.3 Traceability

Classical metrological traceability may not be available for all AI evaluations.

The evaluator should instead preserve appropriate traceability to:

- Protocol
- task source
- reference dataset
- system version
- environment
- scorer
- software
- decision criteria

14.4 Uncertainty Sources

- Task sampling
- system stochasticity
- prompt variation
- access conditions
- scorer disagreement
- judge model
- environment
- distribution shift
- human baseline

- task contamination
- incomplete elicitation

14.5 Uncertainty Reporting

State:

- Estimated uncertainty
- method
- assumptions
- unquantified uncertainty
- sensitivity
- limitations

14.6 Thresholds

A threshold should identify:

- Decision purpose
- evidence
- false-positive cost
- false-negative cost
- uncertainty treatment
- review
- version

14.7 Repeated Trials

Use enough trials to support the decision.

14.8 Missing Data

Document and assess missingness.

14.9 Outliers

Use defined rules.

14.10 Aggregation

Do not hide severe domain failure through broad averages.

14.11 Evidence Package

A complete package may include:

- System manifest
- protocol
- raw records
- task identifiers
- result data
- scoring
- uncertainty
- deviations
- reviewer findings
- security status
- final report

14.12 Evidence Access

The accreditation body should receive enough evidence to assess competence while respecting protected information.

---

15. Security and Protected Evaluation

15.1 Security Scope

Accreditation should assess security whenever the evaluator handles:

- Held-out tasks
- model weights
- privileged access
- dangerous capabilities
- vulnerabilities
- client secrets
- personal data
- critical-infrastructure information
- incident evidence

15.2 Security Management System

The evaluator should maintain controls for:

- Governance
- risk assessment
- access
- identity

- devices
- networks
- storage
- logging
- incident response
- recovery
- suppliers
- personnel
- physical security

15.3 Task Custody

Protected task systems should record:

- Creation
- validation
- storage
- access
- transfer
- administration
- scoring
- retirement
- compromise

15.4 Least Privilege

Personnel should receive only the access required.

15.5 Separation

Consider separation among:

- Task author
- task custodian
- administrator
- scorer
- client contact
- result reviewer

15.6 Secure Facilities

The accreditation scope should identify approved:

- Physical facilities
- remote environments
- secure enclaves

- cloud environments
- jurisdictions

15.7 Evaluation-Induced Risk

The evaluator should assess whether testing itself could:

- Create dangerous knowledge
- reveal vulnerabilities
- enable misuse
- train the evaluated system
- expose tasks
- harm human participants

15.8 Security Incident

Report material incidents to:

- Internal authority
- accreditation body
- client
- scheme owner
- regulator

as required.

15.9 Compromise Assessment

Determine:

- Asset affected
- time
- access
- exposure
- affected evaluations
- result validity
- replacement
- public notice

15.10 Accreditation Effect

A severe security failure may lead to:

- Corrective action
- special surveillance
- scope suspension

- full suspension
- withdrawal

15.11 Public Minimum

The registry should indicate whether the evaluator is accredited for protected work without exposing sensitive details.

15.12 Security Competence Witnessing

Assessment teams should include security competence for protected scopes.

---

16. Data, Software, and Infrastructure

16.1 Data Governance

Control:

- Source
- rights
- privacy
- quality
- version
- transformation
- retention
- access
- deletion
- correction

16.2 Evaluation Data

Preserve enough to reproduce or review results.

16.3 Personal Data

Apply lawful and ethical safeguards.

16.4 Synthetic Data

Document generation and validation.

16.5 Software Inventory

Maintain:

- Evaluation code
- dependencies
- versions
- licenses
- vulnerabilities
- owners
- validation
- release status

16.6 Repositories

Use controlled repositories and protected branches.

16.7 Compute

Assess:

- Availability
- isolation
- performance
- reproducibility
- regional constraints
- logging
- failure

16.8 Cloud Services

Review:

- Contract
- security
- data location
- access
- continuity
- vendor dependency
- incident notification

16.9 System Access

Record:

- Interface

- credentials
- rate limits
- model version
- system prompt
- tool access
- safeguards
- changes

16.10 Time and Logging

Use reliable timestamps and immutable or protected logs where consequence warrants.

16.11 Infrastructure Change

Material changes may require validation and notification.

16.12 Business Continuity

Plan for:

- Cloud outage
- model-access loss
- cyberattack
- facility loss
- key-person departure
- vendor failure
- task compromise

---

17. Reports and Public Claims

17.1 Report Minimum

An accredited report should identify:

- Evaluator
- accreditation body
- accredited scope
- method
- system
- version
- conditions
- result
- uncertainty
- limitations

- deviations
- date
- status
- authorized signatory

17.2 Accreditation Reference

The evaluator may refer to accreditation only for work within scope.

17.3 Mixed-Scope Reports

Clearly distinguish accredited and nonaccredited activities.

17.4 No Universal Claim

Avoid:

- Accredited for AI
- officially approved evaluator
- certified safe
- universally trusted
- government approved

unless exact basis exists.

17.5 Accreditation Mark

Use should follow controlled rules consistent with ISO/IEC 17030:2021.[^iso17030]

17.6 Digital Verification

Reports should support registry verification.

17.7 Opinion and Result

Separate:

- Measured result
- professional judgment
- legal interpretation
- policy recommendation
- client statement

17.8 Corrected Report

Preserve original and correction.

17.9 Report Withdrawal

Withdraw when the basis is materially invalid.

17.10 Marketing Review

The evaluator should review public marketing for scope inflation.

---

18. Application and Eligibility

18.1 Application Eligibility

The applicant should have:

- Legal identity
- defined activities
- operational experience or pilot evidence
- quality system
- competent personnel
- methods
- infrastructure
- impartiality controls
- financial stability
- security readiness
- agreement to surveillance

18.2 Application Information

- Organization
- ownership
- affiliates
- locations
- requested scope
- standards
- schemes
- personnel
- methods
- equipment
- clients
- conflicts

- insurance
- complaints
- incidents
- prior recognition
- sanctions
- subcontractors

18.3 Prior Accreditation

Disclose:

- Existing accreditation
- suspension
- withdrawal
- denial
- pending application
- cross-border activity

18.4 Application Review

Determine:

- Completeness
- pathway
- assessor competence
- security
- legal eligibility
- assessability
- resource needs
- conflict

18.5 Refusal to Accept

The accreditation body may refuse when:

- Scope is not assessable
- competence is unavailable
- scheme is invalid
- security is inadequate
- applicant is deceptive
- authority is absent
- conflict is unmanageable

18.6 Preassessment

Optional preassessment may identify readiness gaps.

It should not become consulting that compromises later assessment.

18.7 Readiness Review

Standards Body may conduct a readiness review at the present stage.

The output should state clearly that it is not accreditation.

---

19. Accreditation Assessment Team

19.1 Team Composition

A team may include:

- Lead assessor
- technical assessor
- domain expert
- security assessor
- legal or rights expert
- quality-system assessor
- witness assessor
- observer
- trainee

19.2 Team Competence

Collectively cover:

- Applicable conformity-assessment standard
- scheme
- requested scope
- AI systems
- methods
- security
- measurement
- domain risks
- jurisdiction

19.3 Assessor Independence

Screen:

- Employment
- consulting
- research collaboration

- funding
- competition
- prior assessment
- personal relationship
- future opportunity
- ideological commitment

19.4 Applicant Objection

The applicant may raise reasoned objections to team members.

The accreditation body retains appointment authority.

19.5 Assessor Authorization

Assessors should be authorized for defined roles and scopes.

19.6 Confidentiality and Security

Assessors should receive access only as required.

19.7 Assessor Performance

Monitor:

- Technical quality
- consistency
- conduct
- timeliness
- report quality
- complaint history
- decision reversal

19.8 Rotation

Rotate assessors where capture or familiarity risk becomes material.

19.9 Observers

Government, peer-evaluation, or scheme observers may participate under controlled roles.

19.10 Team Record

Publish or preserve:

- Names or protected identities
- roles
- competence
- conflicts
- dates
- scope

---

20. Initial Accreditation Assessment

20.1 Assessment Plan

Define:

- Standards
- scheme
- scope
- locations
- methods
- witnessing
- personnel
- security
- records
- schedule
- language
- remote and on-site activity

20.2 Document Review

Review:

- Governance
- quality manual
- methods
- competence records
- impartiality
- security
- complaints
- reports
- internal audit
- management review
- incidents

- proficiency
- subcontractors

20.3 On-Site Assessment

Inspect:

- Facilities
- systems
- access
- records
- staff practice
- security
- equipment
- task custody
- continuity

20.4 Remote Assessment

May support, but should not replace, necessary direct observation.

ISO/IEC TS 17012:2024 provides guidance concerning remote auditing methods for management-system audits.[^iso17012]

20.5 Witness Assessment

Observe the evaluator conducting representative activities.

20.6 Sample Selection

Sample across:

- Methods
- personnel
- domains
- clients
- reports
- locations
- security levels
- decisions

20.7 Interviews

Interview:

- Leadership

- technical staff
- scorers
- security staff
- reviewers
- decision makers
- complaints personnel
- trainees

20.8 Vertical Trace

Follow one evaluation from:

- Contract
- scope
- system identity
- method
- administration
- result
- review
- report
- correction

20.9 Horizontal Trace

Compare one control across several engagements.

20.10 Findings

Classify:

- Conformity
- observation
- improvement
- minor nonconformity
- major nonconformity
- critical nonconformity

---

21. Nonconformity and Corrective Action

21.1 Minor Nonconformity

A bounded failure unlikely to invalidate competence or results alone.

21.2 Major Nonconformity

A systemic or consequential failure affecting competence, impartiality, validity, or reliable operation.

21.3 Critical Nonconformity

An immediate threat to:

- Result validity
- protected information
- public safety
- legal integrity
- independence
- accreditation credibility

21.4 Examples of Critical Nonconformity

- Fabricated evaluation data
- deliberate result manipulation
- concealed task compromise
- false system identity
- unauthorized dangerous disclosure
- result-dependent payment
- unqualified personnel making severe-risk conclusions
- accreditation-mark fraud
- retaliation against a whistleblower
- undisclosed ownership by a primary client

21.5 Corrective-Action Plan

Include:

- Containment
- cause
- correction
- affected work
- systemic action
- owner
- deadline
- evidence
- verification

21.6 Closure

The accreditation body should verify effectiveness.

21.7 Open Major Finding

Accreditation should not be granted while an unresolved major finding undermines the requested scope.

21.8 Dispute

The applicant may challenge a finding through defined process.

---

22. Technical Review and Accreditation Decision

22.1 Assessment Report

The report should contain:

- Applicant
- scope
- criteria
- team
- activity
- evidence
- findings
- corrective action
- unresolved issues
- recommendation

22.2 Technical Review

A competent reviewer not responsible for all assessment activity should test:

- Evidence sufficiency
- scope wording
- consistency
- finding classification
- corrective-action closure
- legal and scheme requirements
- impartiality
- security

22.3 Decision Function

The decision function may:

- Grant
- grant conditionally
- narrow
- defer
- deny
- require further evidence

22.4 Decision Independence

Decision makers should be conflict screened and separate from sales and account management.

22.5 Accreditation Conditions

Possible conditions:

- Restricted method
- enhanced surveillance
- personnel limitation
- no protected work
- limited location
- proficiency requirement
- corrective-action milestone

22.6 Decision Record

Record:

- Decision
- authority
- date
- scope
- conditions
- reasons
- dissent
- appeal
- effective date
- validity

22.7 Denial

Provide reasons and appeal route.

22.8 Public Notice

Publish granted scope only after decision and acceptance of terms.

---

23. Accreditation Certificate and Public Registry

23.1 Certificate Content

The accreditation certificate should identify:

- Accreditation body
- accredited legal entity
- organizational unit
- identifier
- applicable standard
- scheme
- scope
- locations
- effective date
- expiration or reassessment date
- restrictions
- status
- registry link
- authorized signatory

23.2 Scope Annex

A detailed scope annex should be part of the accreditation record.

23.3 Digital Registry

The registry should permit verification by:

- Evaluator name
- identifier
- scope
- method
- domain
- location
- standard
- status
- date

23.4 Historical Registry

Preserve:

- Prior scopes
- reductions
- suspensions
- withdrawals
- restorations
- expired status

23.5 Public Minimum

Where protected work is involved, publish enough to verify competence without exposing sensitive assets.

23.6 No Certificate as Marketing Substitute

The complete registry controls.

A copied certificate may be outdated.

23.7 Machine-Readable Status

Provide structured fields for:

- Active
- conditional
- suspended
- reduced
- withdrawn
- expired
- under appeal

23.8 Registry Correction

Correct errors visibly.

---

24. Surveillance

24.1 Purpose

Surveillance confirms continuing competence and impartiality between complete reassessments.

24.2 Surveillance Plan

Risk-based planning should consider:

- Scope
- consequence
- novelty
- complaint history
- proficiency
- client concentration
- method change
- personnel change
- security
- regulatory role
- prior findings

24.3 Surveillance Methods

- Document review
- on-site assessment
- remote assessment
- witness assessment
- report sampling
- proficiency testing
- unannounced or short-notice review
- interview
- registry check
- complaint review
- incident review
- client feedback
- public-claim review

24.4 Frequency

High-consequence, immature, or protected scopes may require more frequent surveillance.

24.5 Continuous Reporting

The evaluator should report material changes promptly.

24.6 Trigger Events

- Key personnel departure
- new ownership
- major client concentration

- method revision
- security incident
- task compromise
- significant complaint
- adverse legal finding
- scope expansion
- new location
- outsourcing
- insolvency risk
- repeated proficiency failure

24.7 Unannounced Review

May be justified for:

- Suspected manipulation
- security concern
- false public claims
- repeated nonconformity
- risk to affected parties

24.8 Surveillance Findings

May lead to:

- Continued accreditation
- condition
- special surveillance
- reduction
- suspension
- withdrawal

24.9 Proportionality

Surveillance should be rigorous without creating unnecessary barriers to smaller evaluators.

24.10 Public Status

Material restrictions should be updated promptly.

---

25. Reassessment and Renewal

25.1 Reassessment Purpose

Reassessment determines whether accreditation should continue for a new cycle.

25.2 Complete Review

Review:

- Quality system
- competence
- methods
- results
- security
- independence
- proficiency
- complaints
- corrections
- changes
- public claims
- continuing need for scope

25.3 Renewal Is Not Automatic

Good prior history supports but does not guarantee renewal.

25.4 Method Currency

Obsolete methods should be removed or updated.

25.5 Personnel Continuity

Confirm current authorized personnel.

25.6 Scope Relevance

Confirm the scope remains clear and decision-relevant.

25.7 Accreditation Cycle

The cycle should be specified by the accreditation body and scheme.

A shorter cycle may be appropriate during ecosystem formation.

25.8 Expiration

If renewal is not completed, status should become expired unless an authorized extension exists.

25.9 Temporary Extension

Use only with:

- Reason
- risk review
- defined duration
- public status
- no unresolved critical issue

---

26. Scope Extension, Reduction, Suspension, and Withdrawal

26.1 Extension

Requires evidence that competence and controls apply to the added activity.

26.2 Reduction

May result from:

- Voluntary request
- loss of competence
- discontinued method
- low activity
- failed proficiency
- security limits
- legal change

26.3 Suspension

Appropriate where:

- Requirements are not currently met
- risk is material but remediable
- investigation is pending
- evaluator cannot perform reliably
- accreditation claims are misused

26.4 Partial Suspension

Prefer targeted action where the failure is bounded.

26.5 Full Suspension

Use where institutional competence or integrity is broadly affected.

26.6 Withdrawal

Appropriate for:

- Fraud
- persistent nonconformity
- severe impartiality failure
- deliberate concealment
- repeated task compromise
- insolvency
- refusal of surveillance
- unlawful activity
- unrecoverable competence loss

26.7 Voluntary Withdrawal

The evaluator may withdraw.

The accreditation body should assess effects on current reports, certificates, and client reliance.

26.8 Restoration

Requires verified correction and decision.

26.9 Effect on Prior Work

The accreditation body should assess whether prior results:

- Remain valid
- need qualification
- require review
- should be withdrawn

26.10 Public Notice

Status changes should be timely and specific.

---

27. Complaints

27.1 Complaint Sources

- Applicant
- accredited evaluator
- evaluator client
- affected party
- employee
- competitor
- regulator
- scheme owner
- assessor
- public-interest organization

27.2 Complaint Subjects

- Evaluator conduct
- invalid methods
- false scope claim
- discrimination
- conflict
- security
- report error
- accreditation-body conduct
- assessor conduct
- delay
- confidentiality
- retaliation

27.3 Intake

Record:

- Complainant
- matter
- evidence
- risk
- confidentiality
- retaliation concern
- requested outcome

27.4 Conflict Screening

The person complained about should not control the response.

27.5 Investigation

Use proportionate:

- Record review
- interviews
- witness assessment
- technical review
- security review
- legal review

27.6 Outcome

- No finding
- guidance
- correction
- additional surveillance
- finding
- scope action
- disciplinary action
- systemic change
- referral

27.7 Complainant Communication

Provide a safe summary of outcome where lawful.

27.8 Public Reporting

Publish aggregate complaint data and material systemic findings.

27.9 Nonretaliation

Protect good-faith complaints.

---

28. Appeals

28.1 Appealable Decisions

- Application refusal
- finding
- scope limitation
- denial

- condition
- suspension
- withdrawal
- complaint outcome where eligible
- public registry statement

28.2 Appeal Independence

The appeal body should not consist solely of people who made the original decision.

28.3 Appeal Grounds

- Procedural error
- factual error
- technical error
- conflict
- disproportionate action
- new evidence
- inconsistent application

28.4 Appeal Record

- Decision
- grounds
- evidence
- response
- reviewer
- conflicts
- outcome
- reasons
- effective date

28.5 Interim Status

An appeal should not automatically stay a safety-critical suspension.

28.6 Outcome

- Affirm
- modify
- remand
- restore
- reverse
- dismiss

- require reassessment

28.7 External Appeal

Law or recognition arrangements may provide external routes.

---

29. Proficiency Testing and Inter-Evaluator Comparison

29.1 Purpose

Demonstrate that evaluators can achieve sufficiently consistent and valid results.

29.2 Program Types

Common Reference System

Evaluators assess the same controlled model or system.

Common Task Package

Evaluators administer equivalent protected tasks.

Common Evidence Package

Evaluators review the same records and reach findings.

Blind Incident Case

Evaluators classify and analyze an incident.

Scoring Comparison

Evaluators score common outputs.

Replication Exercise

One evaluator reproduces another's result.

Security Drill

Evaluators demonstrate protected-task handling.

29.3 Proficiency-Testing Provider

The provider should be competent and independent.

ISO/IEC 17043:2023 establishes general competence requirements for proficiency-testing providers.[^iso17043]

29.4 Program Design

Control:

- Objective
- reference value
- object stability
- task security
- scoring
- uncertainty
- confidentiality
- participant comparability
- result use

29.5 Performance Evaluation

Avoid simplistic rankings.

Assess:

- Accuracy
- consistency
- uncertainty
- process
- explanation
- security
- reproducibility

29.6 Unsatisfactory Performance

Require:

- Cause analysis
- corrective action
- follow-up
- scope review
- surveillance

29.7 Repeated Failure

May lead to scope suspension or withdrawal.

29.8 Public Disclosure

Publish program-level findings where useful.

Individual results may be protected unless status action is required.

29.9 Scarce Domains

Where formal proficiency testing is unavailable, use:

- Witnessing
- cross-evaluator replication
- case review
- reference exercises
- peer comparison

---

30. Subcontracting and External Resources

30.1 Responsibility

The accredited evaluator remains responsible for subcontracted work.

30.2 Approval

The evaluator should assess and approve subcontractors before use.

30.3 Competence

Subcontractors should meet relevant competence, impartiality, security, and quality requirements.

30.4 Client Consent

Obtain where required.

30.5 Public Disclosure

Material subcontracting should be disclosed in the report.

30.6 Accreditation Reference

The evaluator should not imply that subcontracted work is accredited unless covered appropriately.

30.7 Protected Work

Apply enhanced controls.

30.8 External Models and Judges

Treat third-party models used for scoring or evaluation as critical external resources.

30.9 Subcontractor Register

Maintain:

- Identity
- role
- competence
- contract
- conflicts
- security
- performance
- status

30.10 Affiliate Risk

Affiliates and subsidiaries require the same conflict review.

---

31. Multi-Site and Remote Operations

31.1 Central Control

A multi-site evaluator should demonstrate central control of:

- Quality
- methods
- competence
- records
- security
- reporting
- correction

31.2 Site Scope

List each accredited site and activity.

31.3 Sampling

The accreditation body may sample sites only where the system supports valid inference.

31.4 Virtual Sites

Remote teams and cloud systems should be treated as operational locations where relevant.

31.5 Cross-Border Personnel

Assess:

- Jurisdiction
- access
- employment
- data transfer
- security
- legal eligibility
- language

31.6 Remote Evaluation

The scope should identify remote activities.

31.7 Remote Limitations

Some activities may require direct or on-site access.

31.8 Site Closure

Report and update the registry.

---

32. Financial Model and Market Integrity

32.1 Fee Principle

Fees should support competent assessment without creating result dependence.

32.2 Published Fees

Publish or explain:

- Application

- assessment
- surveillance
- travel
- scope extension
- appeal
- registry
- mark use

32.3 No Pass Fee

The applicant should pay for the process, not a favorable result.

32.4 Small Evaluator Access

Possible support:

- Scaled fees
- grants
- public funding
- pooled infrastructure
- shared proficiency programs
- staged scope

Support should not weaken criteria.

32.5 Revenue Concentration

The accreditation body should monitor dependence on large evaluator groups, developers, or schemes.

32.6 Cross-Subsidy

Disclose material cross-subsidies.

32.7 Funder Influence

Funders should not control technical decisions.

32.8 Assessor Compensation

Compensation should not depend on findings or approval.

32.9 Accreditation Shopping

The framework should resist:

- Jurisdiction arbitrage
- repeated applications after undisclosed denial
- weak scheme selection
- inflated equivalence claims

32.10 Market Competition

Accreditation should support plural competent evaluators without lowering rigor.

---

33. Evaluator Shopping and Client Selection

33.1 Client Choice

In many voluntary markets, the client chooses the evaluator.

This creates a risk of shopping.

33.2 Controls

- Public scope
- common methods
- registry
- report transparency
- result portability
- scheme allocation
- rotating panels
- regulator selection
- insurer selection
- multi-evaluator review

33.3 Prior Evaluator Disclosure

Schemes may require disclosure of:

- Prior assessment
- withdrawal
- unresolved findings
- repeated applications
- evaluator changes

33.4 Transfer Rules

A new evaluator should obtain relevant prior records where lawful.

33.5 Refusal to Cooperate

May be a scheme-level warning.

33.6 Lowest-Bid Risk

Procurement should evaluate competence, independence, access, and quality, not price alone.

33.7 Developer Access Control

Schemes should reduce the ability of developers to reward favorable evaluators with privileged access.

---

34. Public Claims, Marks, and Misuse

34.1 Permitted Claim

> Accredited by [accreditation body] for the scope listed at [registry link].

34.2 Required Scope

Do not omit:

- Activity
- standard
- domain
- status
- location where material

34.3 Prohibited Claims

- Accredited by Standards Body, when no accreditation exists
- Accredited for all AI
- Official global evaluator
- Government approved, without designation
- Certified safe evaluator
- Accreditation guarantees correct results

34.4 Mark Use

Rules should cover:

- Placement
- digital use
- reports
- websites
- advertising
- mixed-scope work
- suspension
- withdrawal
- affiliates

34.5 Mark Misuse

Actions:

- Correction
- takedown
- notice
- suspension
- withdrawal
- legal enforcement where authorized

34.6 Client Use

Clients should not use evaluator accreditation to imply certification of their AI system.

34.7 Directory Integrity

Directories should show current registry data.

---

35. International Recognition

35.1 Recognition Objective

Support acceptance of competent evaluator results across borders without assuming universal equivalence.

35.2 Global Accreditation Cooperation

Global Accreditation Cooperation Incorporated began full operations in January 2026, unifying the international accreditation functions previously associated with IAF and ILAC and launching a multilateral recognition arrangement.[^globalaci]

A frontier AI accreditation program should engage with that evolving system where applicable.

35.3 National Accreditation Bodies

Legal and market recognition often depends on national or regional accreditation bodies.

35.4 Peer Evaluation

International recognition should rely on:

- Common standards
- peer evaluation
- competent assessors
- transparent scopes
- complaint cooperation
- status exchange
- surveillance
- sanctions

35.5 Recognition Scope

A recognition arrangement should identify:

- Accreditation standards
- conformity-assessment activities
- schemes
- jurisdictions
- exclusions
- conditions
- dispute process
- withdrawal

35.6 Equivalence

Do not assume that two accreditations are equivalent because both concern AI.

35.7 Crosswalk

Map:

- Standards
- scope fields
- competence
- surveillance
- reports
- status
- legal effect
- security

35.8 Cross-Border Accreditation

Prefer local accreditation where competent and available.

Cross-border accreditation may be appropriate when:

- No local competence exists
- scheme is international
- regulator permits
- regional arrangements apply
- capacity building is planned

35.9 Capacity Building

Support:

- Assessor training
- regional laboratories
- shared proficiency programs
- translation
- secure infrastructure
- mentorship
- fee support

35.10 No Global Monopoly

International interoperability does not require one accreditation body.

---

36. Government and Regulatory Interfaces

36.1 Accreditation and Law

Law may:

- Require accreditation
- define national accreditation
- designate conformity-assessment bodies
- establish notified bodies
- recognize foreign results
- impose liability
- control marks
- require reporting

36.2 European Union AI Act

Under the EU AI Act, Member States designate notifying authorities for assessment, designation, notification, and monitoring of conformity-assessment bodies.

Member States may have national accreditation bodies conduct the assessment and monitoring.

An applicant for notification may use an accreditation certificate to demonstrate compliance with notified-body requirements.

Notification remains a governmental process.[^eu-ai-act-notified]

36.3 Notified-Body Requirements

The AI Act requires relevant notified bodies to address:

- Legal personality
- organizational confidence
- quality management
- resources
- cybersecurity
- independence
- impartiality
- confidentiality
- liability insurance
- professional integrity
- technical, legal, administrative, and scientific competence
- coordination and standards awareness

36.4 Conformity Assessment Under the AI Act

For certain high-risk systems, the Act provides pathways involving internal control or notified-body assessment, depending on system category and application of harmonized standards or common specifications.[^eu-ai-act-conformity]

This framework should not be represented as an implementation of EU notification unless formally adopted.

36.5 Regulatory Designation

A regulator may add requirements beyond accreditation.

36.6 Procurement

A purchaser may require:

- Accreditation
- scheme qualification
- domain experience
- result format
- security
- jurisdiction

36.7 Insurance

Insurers may use accreditation as one input.

It should not replace underwriting judgment.

36.8 Legal Liability

Accreditation does not eliminate evaluator liability.

36.9 Government Evaluators

A government evaluator may be accredited, designated, recognized, or operate under direct statutory authority.

The status should be explicit.

---

37. Frontier AI Scope Modules

37.1 General-Purpose Model Evaluation

Competence in:

- Broad capability taxonomy
- model identity
- elicitation
- system variation
- benchmark integrity
- uncertainty
- cross-domain interpretation

37.2 Agentic Systems

Competence in:

- Long-horizon tasks
- tools
- memory
- planning
- environment control
- intervention
- trajectory analysis
- multi-agent effects

37.3 Cyber Capability

Competence in:

- Defensive and offensive security
- controlled environments
- vulnerability handling
- authorization
- misuse safeguards
- responsible disclosure

37.4 Biological and Chemical Risk

Competence in:

- Relevant science
- threat assessment
- information hazards
- controlled task design

- institutional safeguards
- legal restrictions

37.5 Persuasion and Social Manipulation

Competence in:

- Behavioral science
- research ethics
- sampling
- human subjects
- cultural context
- harm
- statistical inference

37.6 AI Research Acceleration

Competence in:

- Machine-learning research workflows
- experiment design
- coding
- model improvement
- automation
- attribution
- counterfactual baseline

37.7 Critical Infrastructure

Competence in:

- Sector systems
- operational technology
- reliability
- safety
- cascading effects
- incident response
- legal requirements

37.8 Fundamental Rights and Public Sector

Competence in:

- Rights
- administrative systems
- discrimination

- due process
- public accountability
- accessibility
- affected-party engagement

37.9 Open-Weight Systems

Competence in:

- Weight access
- fine-tuning
- local deployment
- reproducibility
- safeguard removal
- distribution risk
- version control

37.10 Multimodal Systems

Competence across:

- Text
- image
- audio
- video
- sensor
- action
- cross-modal risk

---

38. Accreditation Stages and Levels

These stages describe program maturity, not current Standards Body authority.

Stage A0: Research Framework

- Criteria published
- no accreditation
- open critique
- standards mapping

Stage A1: Readiness Review

- Voluntary gap assessment
- no accredited status

- private improvement record
- public nonclaim

Stage A2: Pilot Qualification

- Limited pilot
- witnessed activity
- independent review
- no accreditation mark
- time-bounded qualification label

Stage A3: Provisional Accreditation

Possible only through a legitimate accreditation body.

- Narrow scope
- enhanced surveillance
- limited duration
- proficiency requirement
- public conditions

Stage A4: Full Scope Accreditation

- Mature quality system
- validated methods
- demonstrated competence
- surveillance
- public registry
- appeal

Stage A5: Internationally Recognized Accreditation

- Accreditation body peer evaluated
- relevant recognition arrangement
- cross-border scope
- status exchange
- continuing surveillance

38.1 No Automatic Progression

An organization advances only through evidence and decision.

38.2 Provisional Status

Should not be marketed as full accreditation.

38.3 Standards Body Present Position

Standards Body remains at the framework and pilot-design stage.

---

39. Accreditation Program Governance

39.1 Program Charter

Every evaluator-accreditation program should have a public charter defining:

- Purpose
- legal and institutional basis
- accreditation pathways
- eligible applicants
- applicable standards
- scope rules
- decision authority
- surveillance
- complaints
- appeals
- funding
- public registry
- international-recognition status
- correction and withdrawal

39.2 Governing Body

The governing body should protect:

- Public interest
- technical quality
- impartiality
- international compatibility
- due process
- financial sustainability
- security
- institutional continuity

39.3 Technical Accreditation Council

A technical council may advise on:

- Sector criteria
- assessor competence
- scope design

- method validity
- proficiency testing
- flexible scopes
- technical interpretations
- transition periods

The council should not make individual accreditation decisions where conflicts exist.

39.4 Impartiality Oversight

An independent or balanced function should review threats created by:

- Applicant influence
- developer funding
- scheme ownership
- government pressure
- consulting
- market concentration
- assessor relationships
- accreditation-body commercial incentives

39.5 Public-Interest Participation

Governance should include meaningful channels for:

- Civil society
- affected communities
- labor
- accessibility experts
- consumer interests
- public-sector users
- independent researchers

39.6 Scheme Governance

Each accepted scheme should have:

- Owner
- version
- maintenance process
- interpretation process
- conflict rules
- public review
- transition policy
- complaint and appeal coordination
- termination plan

39.7 Decision Governance

Accreditation decisions should remain insulated from:

- Sales targets
- political preference
- client importance
- assessor convenience
- media pressure
- funder expectation

39.8 Records

Preserve:

- Charter
- governance membership
- conflicts
- meetings
- decisions
- interpretations
- appeals
- funding
- peer-evaluation findings
- public corrections

39.9 Governance Review

Review the program after:

- Major incident
- repeated assessor inconsistency
- evaluator-market concentration
- recognition failure
- legal change
- scheme controversy
- public legitimacy challenge

39.10 Constitutional Change

Changes to accreditation authority, institutional form, decision rights, or legal recognition should follow the enhanced change-control requirements in `VERSION_HISTORY.md`.

---

40. Scheme-Owner Interface

40.1 Scheme Acceptance Review

Before offering accreditation for a scheme, the accreditation body should assess:

- Scheme purpose
- object of conformity assessment
- normative requirements
- evidence model
- decision rule
- surveillance
- marks
- public claims
- legal use
- maintenance
- governance
- appeals
- implementation feasibility

40.2 Scheme Validity

The accreditation body should not accept a scheme that:

- Makes unverifiable claims
- treats evaluation as proof of universal safety
- omits system identity
- lacks clear requirements
- permits uncontrolled conflicts
- has no correction process
- creates misleading marks
- cannot support consistent assessment
- has no competent owner
- is materially obsolete

40.3 Scheme Interpretations

Interpretations should be:

- Versioned
- approved
- public where possible
- consistently applied
- subject to correction
- linked to the relevant scheme version

40.4 Scheme Changes

A scheme change should trigger:

- Materiality review
- accreditation-body gap analysis
- assessor training
- evaluator transition
- registry update
- compatibility statement
- surveillance adjustment

40.5 Scheme Owner and Accreditation Body

The scheme owner may provide technical expertise.

It should not control individual accreditation decisions.

40.6 Scheme Owner and Evaluator

The scheme should prevent exclusive access or favoritism that advantages selected evaluators without public-interest justification.

40.7 Scheme Termination

The scheme owner should provide:

- Notice
- current-record treatment
- transition
- archive
- mark withdrawal
- certificate and report implications
- successor information

40.8 Standards Body as Possible Future Scheme Owner

If Standards Body later owns an evaluator scheme, it should maintain separation among:

- Standards development
- scheme management
- accreditation assessment
- accreditation decision
- evaluation services
- complaints
- appeals

41. Capture, Corruption, and Concentration Controls

41.1 Capture Threats

Accreditation may be captured by:

- Large developers
- dominant evaluators
- accreditation clients
- governments
- scheme owners
- insurers
- funders
- professional networks
- technical monocultures
- founders

41.2 Market Concentration

Track:

- Share of accredited work
- client concentration
- assessor concentration
- method ownership
- infrastructure dependency
- geographic concentration
- domain bottlenecks

41.3 Revolving Door

Review employment and consulting relationships among:

- Accreditation staff
- assessors
- evaluators
- developers
- regulators
- scheme owners

41.4 Gifts and Benefits

Prohibit or govern:

- Gifts

- travel
- hospitality
- investment
- referral payments
- future-employment arrangements
- privileged data access
- research sponsorship

41.5 Corruption Reporting

Provide protected channels for:

- Bribery
- result manipulation
- concealed conflicts
- accreditation shopping
- mark misuse
- assessor coercion
- retaliation

41.6 Case Allocation

Use controlled allocation rather than permitting applicants to select favorable assessors.

41.7 Technical Monoculture

Recruit assessors with diverse:

- Institutions
- regions
- methods
- disciplines
- professional backgrounds
- public-interest perspectives

41.8 Monopoly Risk

An accreditation program should not create an exclusive evaluator monopoly unless law and evidence justify it.

41.9 Transparency

Publish aggregate information on:

- Market concentration

- complaints
- suspensions
- assessor pools
- recognition
- major conflicts
- corrective action

41.10 External Audit

High-consequence accreditation programs should receive independent governance and integrity review.

---

42. Evaluator Types and Proportionate Requirements

42.1 Commercial Evaluators

Commercial status is compatible with accreditation.

Controls should address:

- Client dependence
- marketing
- consulting
- ownership
- result incentives
- access dependence

42.2 Nonprofit Evaluators

Nonprofit status does not establish independence automatically.

Review:

- Donors
- board
- advocacy
- grants
- developer access
- mission pressure
- financial continuity

42.3 Academic Evaluators

Academic evaluators may offer strong expertise and weak operational systems.

Accreditation should assess:

- Institutional authority
- laboratory continuity
- quality management
- personnel turnover
- intellectual property
- student supervision
- research and service separation
- security

42.4 Government Laboratories

Government status may provide authority and resources.

Review:

- Political independence
- legal mandate
- confidentiality
- procurement
- public reporting
- cross-border acceptance
- appeal

42.5 Open-Source Evaluator Communities

An open-source community may contribute:

- Methods
- code
- tasks
- review
- replication
- incident discovery

Formal accreditation ordinarily requires a legally accountable organizational unit with:

- Quality management
- authorized personnel
- secure operations
- liability
- records
- decision control

42.6 Small Evaluators

Proportionate support may include:

- Narrow scope
- shared secure infrastructure
- pooled proficiency testing
- staged assessment
- fee support
- technical assistance separated from decision staff

Criteria should not be weakened.

42.7 International Consortia

Define:

- Legal responsibility
- central control
- member roles
- scope
- locations
- data transfer
- decision authority
- liability
- withdrawal

42.8 In-House Developer Evaluation Units

An internal unit may be competent and quality controlled.

It should not be described as an independent third-party evaluator.

It may receive internal qualification or a scope-specific accreditation where applicable, but independence claims must remain precise.

42.9 Hybrid Research and Commercial Units

Separate:

- Research experimentation
- accredited methods
- client work
- publication
- intellectual property
- staff authorization

---

43. Multi-Domain and Sector-Specific Accreditation

43.1 Multi-Domain Risk

An organization may appear broad while relying on shallow expertise.

43.2 Domain Scope

Each high-consequence domain should identify:

- Technical lead
- qualified personnel
- method
- safety controls
- legal constraints
- reference assets
- proficiency
- continuing education

43.3 Cross-Domain Evaluation

Some risks arise from interaction among:

- Cyber and biological capability
- autonomy and critical infrastructure
- persuasion and public services
- AI research acceleration and safeguards
- finance and agentic action

Cross-domain scopes should demonstrate integrated competence.

43.4 Sector Modules

Possible sector modules include:

- Health
- finance
- employment
- education
- public administration
- defense
- energy
- communications
- transport
- scientific research

43.5 Local Legal Context

Sector competence may require jurisdiction-specific law and institutions.

43.6 Rights and Accessibility

Scopes affecting people should include competence in:

- Accessibility
- discrimination
- due process
- privacy
- remedy
- human oversight
- affected-party evidence

43.7 Domain Advisory Panels

Panels may support criteria and assessor competence.

They should not become substitutes for accountable decisions.

43.8 Scope Expansion Across Domains

A cyber scope does not establish biological, legal, or public-sector competence.

---

44. Flexible Scope and Rapid Method Evolution

44.1 Need

Frontier AI methods may change faster than traditional fixed accreditation cycles.

44.2 Flexible-Scope Eligibility

Require:

- Mature quality system
- strong method validation
- experienced technical leadership
- controlled authorization
- change records
- internal audits
- successful surveillance
- proficiency

- no unresolved major findings

44.3 Flexible-Scope Boundaries

Define:

- Method family
- construct
- system class
- risk domain
- access mode
- decision type
- security level
- excluded changes

44.4 Internal Method Approval

Before use, the evaluator should document:

- Need
- method
- validation
- competence
- security
- comparability
- scope fit
- approver
- version

44.5 Mandatory Prior Approval

Require accreditation-body approval before:

- Entering a new severe-risk domain
- adding a new conformity-assessment activity
- increasing security classification
- issuing a new certification decision type
- adding a jurisdiction with material legal differences
- making a noncomparable protocol change

44.6 Reporting

The evaluator should report changes within a defined period.

44.7 Surveillance Sampling

The accreditation body should sample newly introduced methods.

44.8 Flexible-Scope Failure

Reduce or suspend flexibility after:

- Invalid method
- poor change control
- unreported expansion
- repeated inconsistency
- security incident
- misleading public claim

44.9 Innovation Sandbox

A separate research sandbox may test new methods outside accredited claims.

44.10 No Accreditation Freezing

Accreditation should not force continued use of a known obsolete method.

---

45. Incident Reporting and Whistleblower Protection

45.1 Reportable Incidents

- Fabricated data
- task compromise
- unauthorized access
- wrong system version
- severe scoring error
- invalid report
- conflict concealment
- client coercion
- mark misuse
- data breach
- dangerous disclosure
- regulator investigation
- repeated method failure
- loss of key competence

45.2 Reporting Timeline

Define immediate, prompt, and periodic reporting categories.

45.3 Internal Escalation

Personnel should be able to bypass normal management where management is implicated.

45.4 Accreditation-Body Notification

The evaluator should notify the accreditation body when an incident may affect:

- Competence
- impartiality
- security
- results
- public claims
- accredited scope

45.5 Whistleblower Protection

Protect good-faith reporting from:

- Dismissal
- demotion
- access removal
- credit removal
- legal intimidation
- blacklisting
- public disparagement

within the institution's authority.

45.6 Anonymous Reporting

Permit protected reporting where justified.

45.7 Evidence Preservation

Preserve:

- Logs
- reports
- communications
- system identity
- task state

- decision records
- access history

45.8 Accreditation Response

Possible response:

- Information request
- special assessment
- temporary restriction
- witness assessment
- scope suspension
- result review
- public notice
- withdrawal

45.9 Systemic Learning

Publish safe aggregate lessons.

45.10 No Confidentiality Shield

Confidentiality should not prevent lawful reporting of serious wrongdoing or danger.

---

46. Evaluator Failure, Transfer, and Continuity

46.1 Failure Scenarios

- Insolvency
- acquisition
- closure
- loss of key personnel
- accreditation withdrawal
- cyber compromise
- legal prohibition
- loss of model access
- infrastructure failure
- reputation collapse

46.2 Continuity Plan

The evaluator should maintain:

- Record custody

- client notification
- report verification
- protected-data transfer
- archive
- credential revocation
- subcontractor closure
- correction channel
- successor arrangements

46.3 Transfer of Work

A successor evaluator should receive only lawful and necessary records.

46.4 Prior Reports

Assess whether prior reports:

- Remain verifiable
- remain valid
- require qualification
- require reassessment
- should be withdrawn

46.5 Certificates

Where the evaluator supports a certification function, applicable scheme and certification-body rules govern certificate transfer.

Accreditation alone does not transfer certificates.

46.6 Registry Continuity

The accreditation body should preserve historical status after evaluator closure.

46.7 Acquisition

A change in control should trigger impartiality and competence review.

Accreditation should not transfer automatically to a materially different legal entity.

46.8 Key-Person Loss

Narrow or suspend affected scope until competence is restored.

46.9 Accreditation-Body Failure

A recognized transition arrangement may protect valid accreditations.

The record should not imply continued recognition without an authorized basis.

46.10 Public Communication

Communicate status without unnecessary reputational speculation.

---

47. Accreditation Metrics

47.1 Competence Metrics

- Authorized personnel
- method validation
- witness performance
- technical-review findings
- continuing competence
- proficiency results

47.2 Impartiality Metrics

- Client concentration
- consulting overlap
- conflict disclosures
- recusals
- result-dependent incentives
- complaints

47.3 Quality Metrics

- Nonconforming work
- corrective-action effectiveness
- internal-audit findings
- report corrections
- review consistency
- management-review completion

47.4 Security Metrics

- Access reviews
- task compromise

- incidents
- recovery
- supplier findings
- protected-scope surveillance

47.5 Accreditation-Process Metrics

- Application time
- assessment time
- finding closure
- decision time
- surveillance completion
- appeal time
- assessor consistency

47.6 Market Metrics

- Accredited evaluator count
- geographic distribution
- domain distribution
- market concentration
- small-evaluator access
- cross-border use
- evaluator shopping indicators

47.7 Public-Integrity Metrics

- Scope-misuse findings
- mark misuse
- registry corrections
- withdrawn reports
- public complaints
- transparency timeliness

47.8 Outcome Metrics

- Replication success
- inter-evaluator consistency
- decision usefulness
- detected failures
- correction speed
- implementation feedback

47.9 Anti-Metric Rule

Do not equate:

- More accredited bodies with better assurance
- fewer findings with higher quality
- faster assessments with stronger competence
- higher client satisfaction with greater independence
- international recognition with universal equivalence

---

48. Accreditation Maturity Model

Level 0: Informal Evaluator Reputation

Characteristics:

- Expert reputation
- no accreditation
- inconsistent methods
- limited public scope
- weak surveillance

Level 1: Evaluator Readiness

Characteristics:

- Quality-system baseline
- competence matrix
- method registry
- impartiality analysis
- readiness review
- no accreditation claim

Level 2: Pilot Qualification

Characteristics:

- Witnessed work
- limited scope
- proficiency exercise
- independent review
- public nonclaim
- time-bounded pilot status

Level 3: Accredited Evaluator Ecosystem

Characteristics:

- Legitimate accreditation body
- scope-specific accreditation
- surveillance
- public registry
- complaints and appeals
- correction and suspension

Level 4: Multi-Jurisdiction Recognition

Characteristics:

- Peer-evaluated accreditation bodies
- recognition arrangements
- crosswalks
- multilingual scopes
- regional capacity
- cross-border complaint cooperation

Level 5: Adaptive Global Assurance Ecosystem

Characteristics:

- Dynamic scope control
- continuous proficiency
- interoperable evidence
- rapid method transition
- secure international exchange
- distributed oversight
- independently audited recognition system

48.1 Maturity Rule

A prestigious evaluator market does not establish an accreditation ecosystem.

---

49. Consolidated Accreditation Failure Modes

49.1 Accreditation by Reputation

Failure:

A well-known evaluator is recognized without scope-specific competence assessment.

Controls:

- Defined scope
- witness assessment
- method review
- personnel authorization
- surveillance

49.2 Self-Accreditation

Failure:

An evaluator or scheme owner labels its own approval as accreditation.

Controls:

- Independent accreditation body
- terminology control
- public correction
- mark enforcement

49.3 Universal Scope Claim

Failure:

An evaluator accredited for one method or domain markets itself as accredited for all AI.

Controls:

- Scope annex
- registry
- report review
- sanctions

49.4 Certification Confusion

Failure:

Evaluator accreditation is represented as certification of the evaluated AI system.

Controls:

- Public-claims rules
- client guidance
- mark separation
- correction

49.5 Regulatory Designation Confusion

Failure:

Accreditation is represented as governmental authorization or notification.

Controls:

- Status field
- authority note
- legal review
- separate designation record

49.6 Method Checklist Accreditation

Failure:

Assessment confirms documentation without testing whether methods are valid.

Controls:

- Technical assessment
- witnessing
- validity evidence
- proficiency

49.7 Obsolete Method Lock-In

Failure:

Accreditation preserves outdated benchmarks because changing them is administratively difficult.

Controls:

- Method monitoring
- flexible scope
- sunset rules
- scheme maintenance

49.8 Flexible-Scope Abuse

Failure:

An evaluator introduces materially new domains or methods without approval.

Controls:

- Boundaries
- change logs

- surveillance
- prior-approval triggers
- scope reduction

49.9 Client Capture

Failure:

A major client influences methods, findings, personnel, publication, or accreditation.

Controls:

- Concentration monitoring
- contractual independence
- replication
- external review
- refusal

49.10 Accreditation-Body Capture

Failure:

Accredited organizations control the institution that assesses them.

Controls:

- Balanced governance
- independent decision function
- funding diversity
- peer evaluation
- public reporting

49.11 Scheme-Owner Capture

Failure:

The scheme owner pressures accreditation decisions to increase adoption.

Controls:

- Role separation
- decision independence
- appeals
- public criteria

49.12 Consulting Conflict

Failure:

The same team designs the applicant's system and assesses conformity.

Controls:

- Separation
- cooling-off period
- conflict review
- independent assessors
- prohibition where required

49.13 Result-Dependent Payment

Failure:

Evaluator or assessor compensation depends on a favorable result.

Controls:

- Contract prohibition
- audit
- disciplinary action
- withdrawal

49.14 Evaluator Shopping

Failure:

Clients repeat assessments until receiving a preferred result.

Controls:

- Prior-assessment disclosure
- transfer rules
- common registry
- scheme allocation
- regulator oversight

49.15 Accreditation Shopping

Failure:

An organization seeks the weakest accreditation jurisdiction while claiming broad equivalence.

Controls:

- Recognition status
- peer evaluation
- scope crosswalk
- public legal effect

49.16 Assessor Familiarity Capture

Failure:

Repeated assessor-applicant relationships reduce challenge.

Controls:

- Rotation
- performance review
- independent technical review
- complaint monitoring

49.17 Competence by Credentials Alone

Failure:

Academic qualifications substitute for observed performance.

Controls:

- Witnessing
- practical authorization
- proficiency
- continuing competence

49.18 Generalist Overreach

Failure:

Machine-learning expertise is treated as sufficient for severe domain conclusions.

Controls:

- Domain modules
- multidisciplinary teams
- scope exclusions
- external experts

49.19 Security Theater

Failure:

Policies exist without secure task custody or operational controls.

Controls:

- Direct observation
- access logs

- incident drills
- technical testing
- witness assessment

49.20 Task Compromise Concealment

Failure:

An evaluator hides exposure to preserve results or accreditation.

Controls:

- Mandatory reporting
- whistleblower channel
- task lineage
- sanctions
- result review

49.21 Wrong-System Evaluation

Failure:

The report identifies a product family rather than the tested system version.

Controls:

- System manifest
- access records
- version verification
- report requirements

49.22 Automated-Judge Dependence

Failure:

An unvalidated model judge determines consequential results.

Controls:

- Judge validation
- human calibration
- disagreement analysis
- version control

49.23 Uncertainty Suppression

Failure:

Reports present precise scores without material uncertainty.

Controls:

- Evidence requirements
- technical review
- report template
- correction

49.24 Average-Score Masking

Failure:

Aggregate performance conceals catastrophic failure in a critical domain.

Controls:

- Domain reporting
- threshold governance
- no compensatory averaging
- decision-specific interpretation

49.25 Proficiency-Test Gaming

Failure:

Evaluators optimize for known exercises without improving normal work.

Controls:

- Protected exercises
- varied formats
- witness assessments
- routine-work sampling

49.26 Proficiency Monopoly

Failure:

One provider controls reference tasks and evaluator market access.

Controls:

- Provider accreditation
- multiple programs
- governance
- appeals
- portability

49.27 Small-Evaluator Exclusion

Failure:

Fees and infrastructure requirements entrench large incumbents unnecessarily.

Controls:

- Proportionate scope
- shared infrastructure
- grants
- staged assessment
- transparent fees

49.28 Weak-Criteria Inclusion

Failure:

Access goals are pursued by reducing competence requirements.

Controls:

- Support without lowered criteria
- narrow scopes
- mentoring separated from decisions

49.29 International Equivalence Inflation

Failure:

Accreditations are treated as equivalent despite different scopes or recognition.

Controls:

- Crosswalk
- peer evaluation
- explicit exclusions
- legal-effect disclosure

49.30 Registry Lag

Failure:

Suspended or reduced accreditation remains displayed as active.

Controls:

- Status automation
- time targets

- audit
- notification

49.31 Certificate Reliance Without Registry Check

Failure:

A copied certificate remains in use after status changes.

Controls:

- Digital verification
- registry precedence
- expiration
- machine-readable status

49.32 Mark Misuse by Client

Failure:

A developer uses evaluator accreditation to imply its system is certified.

Controls:

- Client terms
- monitoring
- correction
- mark enforcement

49.33 Complaint Capture

Failure:

The accreditation sales or account team controls complaints.

Controls:

- Independent intake
- conflict screening
- appeal
- governance reporting

49.34 Appeal Theater

Failure:

Appeals exist but are reviewed by the original decision makers.

Controls:

- Independent panel
- conflict rules
- reasoned decision
- external route where available

49.35 Suspension Without Result Review

Failure:

An evaluator is suspended while potentially affected prior reports remain unexamined.

Controls:

- Retrospective impact review
- client notice
- result qualification
- withdrawal where necessary

49.36 Acquisition Transfer

Failure:

Accreditation is assumed to transfer automatically after a change of control.

Controls:

- Legal-entity review
- impartiality review
- competence review
- transition decision

49.37 Accreditation-Body Failure

Failure:

The accreditation body closes and evaluator claims continue without a valid basis.

Controls:

- Recognition transition
- archive
- status notice
- transfer procedure
- no automatic continuation

49.38 Regulatory Outsourcing

Failure:

A government treats accreditation as a complete substitute for its own legal judgment.

Controls:

- Separate designation
- legal criteria
- regulator oversight
- enforcement authority

49.39 Accreditation as Safety Guarantee

Failure:

Users infer that accredited evaluation proves an AI system safe.

Controls:

- Nonauthority language
- bounded reports
- scheme education
- public testing

49.40 Standards Body Authority Inflation

Failure:

Standards Body's framework is marketed as if it already operates an internationally recognized accreditation system.

Controls:

- Present-stage statement
- no mark
- readiness-review terminology
- website claims audit
- governance transition gate

---

50. Serious Objections and Responses

Objection 1: Frontier AI methods change too quickly for accreditation

Rapid change makes static accreditation inadequate, not necessarily accreditation impossible.

The framework supports:

- Flexible scopes
- method registries
- rapid surveillance
- transition rules
- proficiency exercises
- breaking-change review

Accreditation should assess the evaluator's capacity to manage change as well as its present methods.

Objection 2: Existing accreditation standards were not designed for frontier AI

Existing standards provide institutional foundations for:

- Competence
- impartiality
- records
- decisions
- surveillance
- complaints
- peer evaluation

They require AI-specific application criteria rather than wholesale replacement.

Objection 3: Accreditation will freeze weak benchmarks

It will if schemes define competence as faithful use of fixed tests.

This framework requires validity monitoring, retirement, method change, and dynamic protocol governance.

Objection 4: Evaluators will become dependent on the developers they assess

The risk is real.

Controls include:

- Client-concentration monitoring
- independent funding
- access governance
- disclosure
- replication
- scheme allocation

- regulator or purchaser involvement

Some dependencies may remain unmanageable and require refusal.

Objection 5: Accreditation will favor large consulting firms

Traditional operational requirements can advantage incumbents.

The response should be shared infrastructure, narrow scopes, grants, staged entry, and transparent fees.

Competence and security requirements should not be abandoned.

Objection 6: A government laboratory does not need accreditation

A government laboratory may operate through statutory authority.

Accreditation may still provide:

- International legibility
- competence review
- method consistency
- procurement confidence
- cross-border acceptance

The legal regime should determine necessity.

Objection 7: Open-source evaluation communities cannot be accredited

A loose community may not have sufficient legal accountability.

A defined organizational unit may incorporate open-source methods and contributors while maintaining controlled quality and security.

Objection 8: Accreditation creates false confidence

Poorly communicated accreditation can.

Scope, status, limitations, valid-through dates, and registry verification are therefore central.

Objection 9: Proficiency testing is infeasible for unique frontier systems

Common-system comparison is only one model.

Alternatives include:

- Blind case review
- common evidence packages

- scoring calibration
- replication
- security drills
- witness assessment

Objection 10: Protected evaluations cannot be assessed transparently

Exact tasks may remain protected.

The accreditation body can still assess:

- Custody
- validation
- administration
- scoring
- compromise response
- reviewer access

A public minimum should describe the control model.

Objection 11: Accreditation bodies lack frontier AI expertise

That is a present capacity constraint.

The framework requires specialist assessors, domain experts, training, external panels, pilot work, and staged scope.

A body should decline work it cannot assess competently.

Objection 12: Accreditation will duplicate certification

Accreditation recognizes evaluator competence.

Certification attests conformity of another object under a scheme.

The functions complement rather than duplicate each other.

Objection 13: International recognition will take too long

Recognition does require institutional work.

Premature global claims would be less credible.

Early interoperability can begin with common scope fields, standards mappings, assessor exchanges, and pilot peer review.

Objection 14: Regulators should directly approve evaluators

Direct approval may be appropriate.

Accreditation can provide technical evidence to support designation while the regulator retains legal responsibility.

Objection 15: Evaluator independence is impossible when clients pay

Payment creates a threat, not automatic invalidity.

Fee structure, result independence, diversification, governance, and external oversight can mitigate some threats.

Objection 16: Accreditation cannot guarantee correct results

Correct.

Accreditation provides evidence of competence and controlled operation.

It does not guarantee every result.

Objection 17: This framework is too demanding for the present ecosystem

The complete framework is a target architecture.

Readiness reviews and narrow pilot scopes permit staged development without using accreditation claims prematurely.

Objection 18: Standards Body should issue accreditation quickly to establish leadership

Premature accreditation would create authority inflation, legal risk, weak recognition, and public confusion.

Leadership should arise first from a credible framework, useful pilots, qualified partnerships, and demonstrable institutional restraint.

---

51. Implementation Pathway

Phase 1: Framework and Standards Mapping

- Publish this framework
- map applicable conformity-assessment standards

- define terminology
- identify legal interfaces
- consult accreditation bodies
- publish present-stage nonauthority

Phase 2: Scope Taxonomy

- Define activity types
- define domain modules
- define access and security levels
- define scope templates
- define exclusions
- create machine-readable fields

Phase 3: Competence and Method Infrastructure

- Create personnel matrices
- create assessor competence requirements
- create method-validation criteria
- create witness-assessment protocols
- create quality-system guidance
- create protected-evaluation controls

Phase 4: Readiness Reviews

- Recruit voluntary evaluator organizations
- conduct gap assessments
- test evidence access
- test public nonclaims
- publish aggregate lessons
- revise criteria

Phase 5: Proficiency and Comparison Pilots

- Develop common exercises
- accredit or qualify providers where appropriate
- run blind comparisons
- study inter-evaluator consistency
- test security and correction

Phase 6: Partnership With Existing Accreditation Bodies

- Select qualified bodies
- establish role boundaries

- develop application documents
- train assessors
- conduct joint pilots
- preserve independent decisions

Phase 7: Limited Accreditation Scheme

Through a legitimate accreditation body:

- Use narrow scopes
- enhanced surveillance
- public registry
- defined validity
- complaints
- appeals
- no universal claims

Phase 8: Regulatory and Procurement Interface

- Map legal recognition
- test designation processes
- develop procurement language
- define report reliance
- assess liability
- monitor evaluator shopping

Phase 9: International Recognition

- Participate in peer evaluation
- support scope crosswalks
- engage national accreditation bodies
- test cross-border recognition
- invest in regional capacity

Phase 10: Institutional Review

Before Standards Body considers operating an accreditation function:

- Conduct legal review
- establish separate governance
- demonstrate competence
- establish financial independence
- satisfy ISO/IEC 17011
- undergo external assessment
- define international-recognition strategy

- amend project identity
- publish transition
- obtain governing approval

---

52. First Accreditation-Readiness Pilot

52.1 Pilot Title

Frontier AI Technical Evaluator Readiness Pilot

52.2 Purpose

Test the framework without issuing accreditation.

52.3 Participants

Recruit a diverse set of:

- Independent nonprofit evaluator
- commercial evaluator
- academic laboratory
- government laboratory where feasible
- open-source-affiliated organizational unit

52.4 Pilot Scope

Suggested initial scope:

> Technical evaluation of general-purpose and agentic AI systems through controlled interfaces, using listed software-engineering, cyber-safety, autonomous-task, and safeguard protocols, excluding biological-risk assessment, certification decisions, and regulatory designation.

52.5 Pilot Assessment Basis

- ISO/IEC 17025 principles
- Foundation 1
- Foundation 2
- Foundation 3
- Foundation 4
- Foundation 5
- this framework
- pilot scheme criteria

52.6 Activities

- Application
- scope design
- document review
- competence matrix
- impartiality review
- method review
- security review
- witness assessment
- common proficiency exercise
- report review
- corrective action
- independent decision-style review

52.7 Output

Each participant receives:

- Readiness report
- strengths
- nonconformity-style gaps
- corrective-action priorities
- proposed scope
- public nonclaim
- no accreditation certificate
- no accreditation mark

52.8 Public Aggregate Report

Publish:

- Common gaps
- assessor lessons
- method challenges
- security challenges
- market barriers
- framework revisions

Do not publish confidential participant findings without authorization.

52.9 Pilot Success Criteria

- Scopes are understandable
- assessor teams can evaluate competence
- witnessed work reveals information not visible in documents

- proficiency exercises distinguish performance
- public nonclaims prevent authority confusion
- security review is feasible
- small evaluators can participate with support
- corrective actions are measurable
- the framework produces consistent decisions
- established accreditation bodies find the output usable

52.10 Pilot Termination Criteria

Pause if:

- Participants market pilot status as accreditation
- protected information cannot be secured
- assessors lack competence
- conflicts cannot be managed
- methods are not assessable
- the pilot creates material public confusion

---

53. Evaluator Accreditation Scorecard

Dimension	Core question
Authority	Is the accreditation body legitimate and properly constituted?
Nonauthority	Are voluntary accreditation, legal designation, and certification distinguished?
Pathway	Does the conformity-assessment standard match the evaluator's actual function?
Scheme	Is the scheme valid, governed, assessable, and maintained?
Scope	Is recognized competence precise, bounded, and public?
Legal identity	Is the accredited organization or unit identifiable and accountable?
Governance	Are decisions protected from applicant, funder, and scheme-owner control?
Impartiality	Are organizational, financial, methodological, and publication conflicts managed?
Competence	Are personnel and organizational competence demonstrated?
Domain expertise	Does the team possess real high-risk-domain competence?
Methods	Are evaluation methods valid, current, versioned, and controlled?
Elicitation	Is capability elicitation fit for the decision?
Measurement	Are metrics, baselines, thresholds, and uncertainty defensible?
Software	Are harnesses, judges, dependencies, and environments validated?
System identity	Is the evaluated model or system version exact?
Security	Are held-out tasks, privileged access, and dangerous information protected?
Data	Are provenance, rights, privacy, quality, and retention governed?

Dimension	Core question
Quality system	Are records, nonconforming work, correction, audit, and management review effective?
Witnessing	Has actual evaluator performance been observed?
Proficiency	Has competence been tested through comparison, replication, or equivalent evidence?
Assessment team	Is the accreditation team collectively competent and independent?
Decision	Is the accreditation decision separate, reasoned, and conflict screened?
Registry	Can current scope and status be verified publicly?
Claims	Are accreditation references precise and nondeceptive?
Surveillance	Is continuing competence monitored proportionately?
Change control	Are methods, personnel, ownership, and scope changes governed?
Complaints	Can applicants, clients, workers, and affected parties report problems?
Appeals	Is material decision review independent?
Incidents	Are serious failures reported, investigated, and propagated?
Sanctions	Can scope be reduced, suspended, restored, or withdrawn?
Continuity	Are closure, acquisition, failure, and record transfer planned?
Market integrity	Are evaluator shopping, concentration, and small-evaluator barriers addressed?
International	Are recognition scope, peer evaluation, and legal effect clear?
Public interest	Are affected parties and societal consequences represented?
Audit	Can the complete accreditation decision and history be reconstructed?

53.1 Critical Failures

The following normally prevent accreditation or require immediate restriction:

- No legitimate accreditation authority
- false scope
- fabricated evidence
- deliberate result manipulation
- unmanageable conflict
- result-dependent compensation
- concealed task compromise
- invalid or unreviewable methods
- absent system identity
- unqualified severe-risk personnel
- uncontrolled protected information
- refusal of surveillance
- mark misuse after notice
- retaliation against good-faith reporters
- accreditation decision controlled by sales or the applicant
- no complaint or appeal process

- inability to preserve or verify records
- scheme makes inherently deceptive claims

53.2 No Composite Accreditation Score

Do not average the scorecard into one overall number.

A critical competence, impartiality, authority, security, or integrity failure cannot be offset by strengths elsewhere.

---

54. Accreditation-Body Readiness Template

Body:
Legal entity:
Jurisdiction:
Proposed authority:
Assessment date:

Legal and Institutional Basis

ISO/IEC 17011 Alignment

Governance

Impartiality

Financial Sustainability

Personnel and Assessor Competence

Scheme Acceptance

Decision Independence

Confidentiality and Security

Complaints and Appeals

Public Registry

Peer Evaluation

International Recognition

Continuity

Readiness Decision

- Research stage
- pilot ready
- limited-operation ready
- not ready

Required Actions

---

55. Evaluator Application Template

Applicant:
Legal entity:
Ownership:
Locations:
Requested pathway:
Requested scope:

Activities

Objects and System Classes

Domains

Methods and Protocols

Access Modes

Security Levels

Decision Types

Applicable Standards and Schemes

Personnel

Quality System

Impartiality and Conflicts

Infrastructure

Subcontractors

Prior Accreditation and Adverse History

Complaints and Incidents

Insurance and Continuity

Applicant Declaration

---

56. Scope of Accreditation Template

Accreditation identifier:
Evaluator:
Applicable standard:
Scheme:
Status:
Effective date:
Review or expiration:

Activity

Object

System Class

Capability or Risk Domain

Method or Protocol

Access Mode

Lifecycle Stage

Assurance Level

Location

Jurisdiction

Security Level

Decision Type

Exclusions

Conditions

Flexible-Scope Boundaries

Registry Link

---

57. Personnel Competence Matrix Template

Role	Scope	Required knowledge	Required skill	Experience	Security	Evidence	Authorization	Renewal

Domain Competence

Evaluation-Science Competence

System Competence

Legal and Rights Competence

Witnessed Performance

Proficiency Results

Limitations

Supervisor

---

58. Impartiality Risk Record Template

Evaluator or accreditation body:
Review date:
Owner:

Threat	Relationship	Severity	Likelihood	Existing control	Residual risk	Decision

Ownership

Funding and Client Concentration

Consulting

Shared Personnel

Access Dependence

Publication Rights

Future Employment

Intellectual Commitments

Required Recusals

Public Summary

---

59. Accreditation Assessment Plan Template

Applicant:
Requested scope:
Lead assessor:
Assessment dates:

Criteria

Team and Competence

Conflict Screening

Documents

Locations

Methods Sampled

Personnel Sampled

Engagements Sampled

Witness Activities

Proficiency Evidence

Security Arrangements

Remote Activities

Interviews

Vertical and Horizontal Traces

Reporting and Decision Timeline

---

60. Witness Assessment Template

Evaluator:
Activity:
Protocol:
System:

Witness assessor:
Date:

Scope Fit

Personnel

Preparation

System Identity

Method Execution

Elicitation

Task Integrity

Data and Logging

Scoring

Uncertainty

Security

Deviations

Review

Reporting

Findings

Competence Conclusion

---

61. Nonconformity Template

Finding ID:
Applicant or accredited body:
Criterion:
Classification:

- Minor
- major
- critical

Objective Evidence

Requirement

Nonconformity

Immediate Containment

Affected Work

Root Cause

Correction

Corrective Action

Deadline

Verification

Scope or Status Effect

---

62. Accreditation Decision Template

Decision ID:
Applicant:
Assessment report:

Decision authority:
Date:

Requested Scope

Assessment Findings

Corrective-Action Status

Technical Review

Impartiality Review

Security Review

Decision

- Grant
- conditional grant
- narrow
- defer
- deny
- maintain
- extend
- reduce
- suspend
- restore
- withdraw

Approved Scope

Conditions

Effective and Review Dates

Reasons

Dissent

Appeal

Registry Action

---

63. Surveillance Plan Template

Accredited body:
Accreditation cycle:
Risk level:
Owner:

Surveillance Frequency

Methods

Witnessing

Report Sampling

Proficiency

Complaints and Incidents

Personnel and Ownership Changes

Method Changes

Security

Public Claims

Trigger Events

Special Surveillance

Status Decision

---

64. Proficiency Exercise Template

Program:
Provider:
Participants:
Date:

Objective

Reference Asset

Participant Comparability

Security

Administration

Scoring

Uncertainty

Performance Criteria

Results

Outliers

Corrective Action

Follow-Up

Accreditation Effect

Public Summary

---

65. Complaint and Appeal Template

Record ID:
Type: Complaint or appeal
Complainant or appellant:
Subject:
Date:
Confidentiality:

Matter

Evidence

Immediate Risk

Conflict Screening

Interim Action

Investigation or Review

Response

Decision

Reasons

Corrective Action

Further Review

Nonretaliation Monitoring

---

66. Suspension or Withdrawal Notice Template

Accredited body:
Accreditation identifier:
Affected scope:
Action:
Effective date:
Authority:

Reason

Immediate Effect

Prior Reports and Decisions

Required Client or Regulator Notice

Mark and Claim Restrictions

Corrective Action

Restoration Conditions

Appeal

Registry Status

---

67. Public Accreditation Registry Template

Accreditation body:
Accredited evaluator:
Legal entity:
Identifier:
Applicable standard:
Scheme:
Status:

Scope

Locations

Conditions

Effective Date

Review or Expiration Date

Suspension or Reduction History

Current Mark Rights

Complaints Contact

Registry Verification

Historical Versions

---

68. Cross-Border Recognition Record Template

Recognition ID:
Parties:
Jurisdictions:
Effective date:

Accreditation Standards

Activities Covered

Schemes Covered

Peer-Evaluation Basis

Scope Mapping

Legal Effect

Exclusions

Status Exchange

Complaint Cooperation

Suspension and Withdrawal

Review

Public Claims

---

69. Readiness-Review Nonclaim Template

> This review is a voluntary evaluator-readiness assessment conducted against proposed technical and institutional criteria. It is not accreditation, certification, regulatory approval, governmental designation, or international recognition. The findings concern only the organization, scope, methods, evidence, personnel, locations, and dates identified in the report.

---

70. Canonical Standards Body Accreditation Positions

Standards Body adopts the following working positions.

1. Accreditation recognizes a conformity-assessment body's competence and impartiality for a defined scope.

2. Accreditation should never be described without scope.
3. Accreditation and certification are distinct.
4. Accreditation and regulatory designation are distinct.
5. Accreditation and licensing are distinct.
6. Qualification and readiness review are not accreditation.
7. An evaluator cannot accredit itself.
8. A scheme owner should not present its internal approval as accreditation.
9. Standards Body does not currently accredit evaluators.
10. Present-stage work should use readiness, qualification, pilot, or research terminology.
11. Accreditation should attach to an accountable organization or organizational unit.
12. Individual competence does not replace organizational accreditation.
13. Personnel certification does not replace organizational accreditation.
14. The applicable conformity-assessment pathway should match the actual evaluator function.
15. Testing, inspection, audit, validation, verification, certification, and accreditation should remain distinct.
16. Technical evaluation laboratories should ordinarily align with laboratory competence requirements.
17. Inspection work should use an inspection-body framework.
18. AI management-system certification should not be presented as direct technical certification of a specific model.
19. Product or service certification requires a valid certification scheme.
20. Hybrid organizations should maintain separate scopes and conflict controls.
21. The accreditation body should conform to recognized accreditation-body requirements.
22. The accreditation body should not compete in the activities it accredits.
23. Accreditation decisions should be separate from sales and complete assessment activity.
24. Scheme owners may provide technical input but should not control individual accreditation decisions.
25. Governments retain legal authority for designation, notification, licensing, and enforcement.
26. International recognition should rely on common standards, peer evaluation, and formal arrangements.
27. A self-created global mark does not establish international recognition.
28. Scope is the center of accreditation.
29. Scope should identify activity, object, system class, domain, method, access, lifecycle, location, security, and decision type.
30. Scope exclusions should be explicit.
31. Flexible scope should require mature validation and change control.
32. Flexible scope should not permit unapproved entry into new severe-risk domains.
33. Accreditation should be time-bounded and monitored.
34. Renewal should not be automatic.
35. Surveillance should be risk based.
36. High-consequence and protected scopes should receive enhanced surveillance.
37. Witness assessment is essential because documentation alone cannot prove operational competence.
38. Proficiency testing, replication, or equivalent comparison should support continuing competence.

39. Evaluator competence should include personnel, methods, infrastructure, security, governance, and quality systems.
40. Credentials alone do not establish practical competence.
41. High-risk domains require actual domain expertise.
42. General machine-learning expertise does not establish competence in every scientific, legal, or sector domain.
43. Protected evaluations require security competence.
44. Evaluators should maintain role-specific personnel authorizations.
45. Trainees may participate under supervision and should not independently make high-consequence decisions.
46. External experts should meet relevant competence and conflict requirements.
47. Technical skill does not cure conflict of interest.
48. Independence does not cure invalid methods.
49. Result-dependent compensation is incompatible with credible evaluation.
50. Client concentration should be monitored.
51. Consulting and evaluation of the same work require separation or refusal.
52. Developer access dependence should be treated as an impartiality threat.
53. Evaluators should preserve rights to report accurate findings within legitimate confidentiality constraints.
54. Quality systems should govern real technical work, not only documents.
55. Nonconforming evaluation work should be identified, contained, corrected, and propagated.
56. Error reporting and correction should be rewarded.
57. Methods should be validated for their intended claims and decisions.
58. Standard methods still require correct local implementation.
59. Modified and laboratory-developed methods require validation.
60. Evaluation software, model judges, scoring systems, and environments require control.
61. Automated judges should not determine high-consequence outcomes without validation and human oversight.
62. Benchmark contamination should affect method and result status.
63. Obsolete or compromised methods should be retired.
64. A method change should not be described as comparable without evidence.
65. Measurement uncertainty should remain visible.
66. Aggregate scores should not hide critical domain failure.
67. Evaluation evidence should preserve traceability to system, protocol, tasks, environment, software, scorers, and decisions.
68. Protected evidence should remain independently reviewable.
69. Security incidents should be reported when they may affect accreditation or results.
70. Concealed task compromise is a critical integrity failure.
71. Accreditation claims should identify the accreditation body, standard, scope, status, and registry.
72. Mixed accredited and nonaccredited work should be distinguished in reports.
73. Evaluator accreditation should not be used by clients to imply system certification.
74. Accreditation marks should be controlled and withdrawn after suspension or withdrawal.
75. The public registry should control over copied certificates.
76. Historical scope and status should remain visible.

77. Applicants should disclose prior denial, suspension, withdrawal, and relevant adverse history.
78. Accreditation assessment teams should be collectively competent.
79. Applicants may raise reasoned assessor objections but should not select favorable assessors.
80. Assessors should be rotated when familiarity threatens challenge.
81. Major and critical findings should prevent unconditional grant.
82. Accreditation decisions should be reasoned and appealable.
83. Complaints should be available to affected parties, workers, clients, regulators, and the public.
84. The subject of a complaint should not control its resolution.
85. Appeals should be independent from the original decision.
86. An appeal should not automatically stay a safety-critical suspension.
87. Scope may be reduced, suspended, restored, or withdrawn.
88. Partial action is preferable when failure is genuinely bounded.
89. Withdrawal should trigger review of potentially affected prior reports.
90. An acquisition should trigger review and should not transfer accreditation automatically.
91. Evaluator closure should preserve records, status, and correction pathways.
92. Small evaluators should receive proportionate support without weaker criteria.
93. Open-source methods may support accredited work, but accountable organizational control remains necessary.
94. Nonprofit, academic, government, and commercial evaluators all require conflict and competence review.
95. Accreditation should resist evaluator shopping and accreditation shopping.
96. Procurement should not select evaluators by price alone.
97. International equivalence should not be inferred from broad AI labels.
98. Cross-border recognition should identify scope, standards, peer-evaluation basis, and legal effect.
99. Local accreditation capacity should be strengthened where possible.
100. Accreditation should support evaluator pluralism rather than unnecessary monopoly.
101. Market concentration should be measured.
102. Public-interest and affected-party perspectives should inform scheme governance.
103. Accreditation governance should be protected from developers, evaluators, funders, and political capture.
104. Corruption and retaliation reporting should be protected.
105. Confidentiality should not prevent lawful reporting of serious wrongdoing or danger.
106. Accreditation metrics should not reward speed, volume, or client satisfaction at the expense of rigor.
107. Fewer findings do not necessarily indicate better evaluator quality.
108. More accredited evaluators do not necessarily produce stronger assurance.
109. A pilot readiness result should never be marketed as accreditation.
110. Standards Body should partner with competent existing accreditation bodies before considering its own accreditation function.
111. Any future Standards Body accreditation function should be organizationally separated from standards development and evaluation services.

112. Any such function should meet ISO/IEC 17011 and applicable law.
113. Any claim of international recognition should require actual peer evaluation and recognition arrangements.
114. Standards Body's identity should be formally amended before it claims accreditation authority.
115. The ultimate accreditation test is whether a qualified user can understand exactly what the evaluator is competent to do, under which methods and conditions, with which independent oversight, and with what current limitations.

---

71. Relationship to Other Canonical Files

`PROJECT_IDENTITY.md`

Controls Standards Body's present nonauthority and any future identity transition.

`INSTITUTION_DESIGN.md`

Defines the ecosystem relationship among standards, evaluation, assurance, accreditation, registries, government, and public-interest functions.

`GOVERNANCE_FRAMEWORK.md`

Defines authority, governing bodies, conflicts, decisions, complaints, appeals, funding, and institutional transitions.

`STANDARDS_DEVELOPMENT_PROCESS.md`

Defines how evaluator criteria, scheme requirements, technical specifications, and conformity-assessment-ready standards should be developed and maintained.

`TRANSPARENCY_FRAMEWORK.md`

Defines public registries, protected evidence, confidentiality, incidents, corrections, funding, conflicts, and public minimums.

`CONTRIBUTOR_FRAMEWORK.md`

Defines assessor, reviewer, domain-expert, decision-maker, scheme-maintainer, and public-interest contributor roles.

WEBSITE_SOURCE_OF_TRUTH.md

Defines public accreditation claims, registry fields, status, corrections, marks, authority notes, and machine-readable records.

SOURCES.md

Defines the sources and currentness of accreditation standards, legal recognition, evaluator evidence, methods, and decisions.

VERSION_HISTORY.md

Defines versioning for schemes, criteria, methods, scopes, decisions, certificates, suspensions, withdrawals, and recognition arrangements.

TERMINOLOGY.md

Controls accreditation, certification, conformity assessment, evaluator, auditor, scope, recognition, designation, and related language.

TAXONOMY.md

Classifies evaluator types, conformity-assessment activities, scope dimensions, statuses, risks, methods, and institutional relationships.

EVIDENCE_STANDARDS.md

Defines the evidence required to support competence, findings, accreditation decisions, surveillance, and correction.

RESEARCH_METHODOLOGY.md

Defines research and pilot methods used to validate accreditation criteria and evaluator performance.

EVALUATION_PHILOSOPHY.md

Defines validity, object boundaries, elicitation, scoring, uncertainty, generalization, and limits of evaluation.

Foundation 1

Provides the protocol versioning, change-control, bridge-study, and method-lifecycle basis.

Foundation 2

Provides the held-out task, access, custody, compromise, reproducibility, and confidentiality basis.

Foundation 3

Provides consequence-scaled competence, safeguards, security, and decision requirements.

Foundation 4

Provides the independent expert review, conflicts, access, dissent, and appeals basis.

Foundation 5

Provides the direct architecture for plural testing, inspection, audit, certification, accreditation, proficiency, surveillance, and recognition.

Foundation 6

Provides the path from voluntary frameworks to standards, procurement, legal recognition, mandate, and enforcement.

Foundation 7

Requires incentives and prestige associated with accreditation to resist gaming and false legitimacy.

Foundation 8

Provides international identifiers, scope crosswalks, peer recognition, multilingual records, and capacity-building principles.

---

72. Final Accreditation Position

Frontier AI evaluation should not depend only on reputation, access, or institutional prestige.

It should be possible to ask:

- Is the evaluator competent?
- For which activity?
- Under which method?
- In which domain?
- With which system access?
- At which security level?

- With which personnel?
- Under which quality system?
- With which independence?
- Who assessed that competence?
- Who made the recognition decision?
- How is continuing competence monitored?
- What happens after failure?
- Can the public verify the current scope?
- Can the evaluator appeal?
- Can affected parties complain?
- Is the recognition accepted across borders?
- What legal effect does it actually have?

Accreditation can help answer those questions.

It cannot answer every question about an evaluated AI system.

Accreditation should not become:

- A universal trust badge
- a substitute for valid methods
- a substitute for regulator judgment
- a substitute for public accountability
- a barrier protecting incumbent evaluators
- a mechanism for laundering developer influence
- a claim that every accredited result is correct
- a claim that every tested system is safe

The framework should create disciplined confidence rather than ceremonial confidence.

Disciplined confidence requires:

- Scope
- competence
- impartiality
- valid methods
- protected evidence
- quality systems
- witnessing
- proficiency
- surveillance
- correction
- sanctions
- public status
- appeals
- international clarity

Standards Body should demonstrate restraint in its own institutional development.

It should complete the foundations.

It should test the criteria.

It should learn from established accreditation bodies.

It should build evaluator capacity.

It should preserve clear public nonclaims.

It should not issue accreditation merely because the ecosystem needs it.

A future accreditation function should begin only when its competence, independence, governance, legal basis, security, and recognition are at least as reviewable as the evaluator organizations it intends to judge.

The defining evaluator-accreditation rule is:

> **Recognize only demonstrated competence, within a precise scope, through an independent decision, under continuing surveillance, with public status and correctable authority.**

References and Research Basis

[^iso17011]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17011:2017, Conformity assessment, Requirements for accreditation bodies accrediting conformity assessment bodies**. https://www.iso.org/standard/67198.html

[^iso17025]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17025:2017, General requirements for the competence of testing and calibration laboratories**. https://www.iso.org/standard/66912.html

[^iso17020]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17020:2026, Conformity assessment, Requirements for bodies performing inspection**. https://www.iso.org/standard/17020

[^iso17021]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17021-1:2015, Conformity assessment, Requirements for bodies providing audit and certification of management systems, Part 1: Requirements**. https://www.iso.org/standard/61651.html

[^iso17029]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17029:2019, Conformity assessment, General principles and requirements for validation and verification bodies**. https://www.iso.org/standard/29352.html

[^iso17065]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17065:2012, Conformity assessment, Requirements for bodies certifying products, processes and services**, with a replacement project under development as of July 17, 2026. https://www.iso.org/standard/46568.html

[^iso17024]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17024:2026, Conformity assessment, General requirements for bodies operating certification of persons**. See the ISO Committee on Conformity Assessment standards catalogue. https://www.iso.org/committee/54998/x/catalogue/

[^iso17043]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17043:2023, Conformity assessment, General requirements for the competence of proficiency testing providers**. https://www.iso.org/standard/80864.html

[^iso17030]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17030:2021, Conformity assessment, General requirements for third-party marks of conformity**. https://www.iso.org/standard/78283.html

[^iso17040]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17040:2005, Conformity assessment, General requirements for peer assessment of conformity assessment bodies and accreditation bodies**. https://www.iso.org/standard/31808.html

[^iso17012]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC TS 17012:2024, Conformity assessment, Guidelines for the use of remote auditing methods in auditing management systems**. https://www.iso.org/standard/82269.html

[^iso17000]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17000:2020, Conformity assessment, Vocabulary and general principles**. https://www.iso.org/standard/73029.html

[^iso17060]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 17060:2022, Conformity assessment, Code of good practice**. https://www.iso.org/standard/78919.html

[^iso19011]: International Organization for Standardization, **ISO 19011:2018, Guidelines for auditing management systems**. https://www.iso.org/standard/70017.html

[^iso42001]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system**. https://www.iso.org/standard/81230.html

[^iso42005]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 42005:2025, Information technology, Artificial intelligence, AI system impact assessment**. https://www.iso.org/standard/44545.html

[^iso42006]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC 42006:2025, Information technology, Artificial intelligence, Requirements for bodies providing audit and certification of artificial intelligence management systems**. https://www.iso.org/standard/42006

[^iso42007]: International Organization for Standardization and International Electrotechnical Commission, **ISO/IEC DIS 42007, Information technology, Artificial intelligence, Guidance for**

accredited testing of AI systems, under development as of July 17, 2026. See the ISO/IEC JTC 1/SC 42 catalogue. https://www.iso.org/committee/6794475/x/catalogue/

[^globalaci]: Global Accreditation Cooperation Incorporated, **Global Accreditation Cooperation**, operational from January 1, 2026, including the Global Accreditation Cooperation Multilateral Recognition Arrangement. https://global-aci.org/en/home/ ; International Laboratory Accreditation Cooperation, **Launch of the Global Accreditation Cooperation Incorporated**. https://ilac.org/latest_ilac_news/launch-of-the-global-accreditation-cooperation-incorporated/

[^eu-ai-act-notified]: European Union, **Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence**, especially Articles 28 through 37 concerning notifying authorities, applications for notification, notified bodies, monitoring, challenge, suspension, and withdrawal. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[^eu-ai-act-conformity]: European Union, **Regulation (EU) 2024/1689**, especially Article 43 concerning conformity-assessment procedures for high-risk AI systems. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[^eu765]: European Parliament and Council, **Regulation (EC) No 765/2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products**. https://eur-lex.europa.eu/eli/reg/2008/765/oj

[^nist-ai-rmf]: National Institute of Standards and Technology, **Artificial Intelligence Risk Management Framework and current revision program**. https://www.nist.gov/itl/ai-risk-management-framework

---

Revision Record

Version 1.0

Date: July 17, 2026

Change type: Complete foundational edition

Summary: Establishes the canonical Standards Body evaluator-accreditation framework. Defines present authority limits, accreditation architecture, conformity-assessment pathways, accreditation-body requirements, standards stack, scope design, organizational and personnel competence, impartiality, quality management, method validation, evidence and uncertainty, security, data, software, reports, application, assessment teams, witness assessment, nonconformity, technical review, decisions, certificates, registries, surveillance, reassessment, scope action, complaints, appeals, proficiency testing, subcontracting, multi-site operations, financial integrity, evaluator shopping, public claims, marks, international recognition, government and regulatory interfaces, frontier AI scope modules, accreditation stages, program governance, scheme-owner interfaces, capture controls, evaluator types, multi-domain work, flexible scope, incidents, whistleblowers, continuity, metrics, maturity, failure modes, objections, implementation,

pilot design, scorecard, operational templates, canonical positions, relationships to other sources, primary references, and complete revision record.

Status: Approved foundational source.