

Standards Body · Foundational source, public edition · Released July 17, 2026

Canonical record: <https://standardsbody.ai/library/foundational-source/terminology/>

Standards Body is an independent research and institutional-design project. It is not currently a regulator, accreditation body, certification body, or governmental authority. This document is research; it is not an adopted standard.

TERMINOLOGY.md

Standards Body Terminology

Project: Standards Body

Primary domain: standardsbody.ai

Core line: Foundations for Frontier AI

Document type: Canonical controlled vocabulary and terminology standard

Version: 1.0

Status: Approved foundational source

Document owner: Standards Body

Applies to: All canonical files, website copy, research outputs, evaluation protocols, standards proposals, institutional documents, partnerships, public statements, diagrams, templates, registries, and contributor materials

Review cycle: Annual review, with event-triggered revision after material changes in project identity, evaluation science, assurance practice, standards, law, interoperability, or institutional structure

Document Purpose

This document establishes the canonical terminology of Standards Body.

It is the authoritative source for:

- Preferred terms
- Definitions
- distinctions among related concepts
- prohibited or misleading usage
- capitalization and naming conventions
- status labels
- public-claims language
- terminology used across the eight foundational papers
- terminology for future standards, protocols, registries, reports, and institutional frameworks
- rules for introducing, revising, deprecating, and retiring terms

This document exists because unclear terminology creates real institutional failure.

When terms such as *evaluation*, *audit*, *certification*, *accreditation*, *approval*, *safety*, *capability*, *risk*, *threshold*, and *standard* are used imprecisely, readers may misunderstand:

- What evidence exists
- What process was performed
- Who performed it
- What authority was involved
- What claim is justified
- Whether a result is current
- Whether a requirement is voluntary or binding
- Whether a system has been tested, reviewed, certified, or legally approved

Standards Body uses controlled terminology to make these distinctions explicit.

Where another Standards Body document conflicts with this file on terminology, this file governs unless a specialized approved standard expressly defines a narrower domain-specific term.

1. Terminology Principles

1.1 Precision Before Familiarity

Use the most accurate term, even when a more familiar term would sound simpler or more impressive.

1.2 Meaning Before Branding

Terms should describe the actual function performed.

They should not be selected to make a project, process, organization, or result appear more authoritative.

1.3 Scope Before Generalization

Every important term should be interpreted within a stated scope.

Examples:

- Capability under which conditions?
- Risk to whom?
- Evaluation of which system version?
- Certification against which scheme?
- Recognition by which institution?
- Compliance with which requirement?

1.4 Evidence Before Conclusion

Words implying certainty should match the quality and limits of the evidence.

1.5 Authority Before Institutional Label

Use *regulator*, *accreditation body*, *certification body*, *standards organization*, and related labels only when the organization actually performs that function under a legitimate mandate.

1.6 Version Before Reuse

Terms used in protocols, standards, and schemas should be versioned when their meaning changes materially.

1.7 Plain Language With Technical Integrity

Definitions should be understandable without removing distinctions necessary for institutional accuracy.

1.8 Stable Core, Extensible Detail

The vocabulary should maintain a stable common core while allowing domain-specific extensions.

1.9 No Hidden Synonymy

Two terms should not be treated as interchangeable merely because ordinary language uses them similarly.

1.10 No False Difference

Different institutions may use different labels for substantially similar concepts.

Mappings should identify when the difference is linguistic rather than substantive.

2. Term Status Labels

Every controlled term may receive one of the following statuses.

2.1 Preferred

The default Standards Body term.

2.2 Accepted

Permitted where context makes the meaning clear, but not preferred for canonical use.

2.3 Context-Specific

Permitted only in a defined technical, legal, standards, or institutional context.

2.4 Discouraged

Avoid because it is ambiguous, misleading, overly broad, or likely to imply unsupported authority.

2.5 Prohibited

Do not use in Standards Body materials except when quoting or analyzing another source.

2.6 Deprecated

Previously accepted but replaced by a clearer term.

2.7 Retired

No longer used for current work but retained in historical records.

3. Definition Format

Canonical entries may include:

- Preferred term
- Status
- Definition
- Use when
- Do not confuse with
- Usage note
- Related terms

Not every entry requires every field.

4. Core Project and Institutional Terms

4.1 Standards Body

Status: Preferred proper name

Definition: The independent research and institutional-design project developing foundations for frontier AI evaluation, assurance, standards, and governance.

Usage note: Always capitalize both words when referring to the project.

Do not use as: A generic substitute for every standards organization.

Authority note: The name does not imply present regulatory, accreditation, certification, or governmental authority.

4.2 Foundations for Frontier AI

Status: Preferred core line

Definition: The canonical line describing the project's focus on the intellectual and institutional foundations required for credible frontier AI evaluation and standards.

4.3 Project

Status: Preferred for present institutional stage

Definition: The current Standards Body research and institutional-design effort.

Use when: Describing current status.

Do not confuse with: A formally recognized standards organization, regulator, or certification authority.

4.4 Institution

Status: Context-specific

Definition: An enduring organization, system of rules, or arrangement that coordinates roles, authority, incentives, and behavior.

Usage note: Standards Body may study future institutional forms without claiming that all such functions presently exist.

4.5 Institutional Design

Status: Preferred

Definition: The deliberate design of roles, authority, governance, incentives, procedures, accountability, and relationships within or among institutions.

4.6 Institutional Infrastructure

Status: Preferred

Definition: The governance, standards processes, assurance systems, registries, contributor structures, transparency rules, and other durable arrangements that support repeatable institutional work.

4.7 Public-Interest Orientation

Status: Preferred

Definition: A commitment to improving evidence, accountability, beneficial development, public decision quality, and institutional legitimacy beyond the interests of any one company, government, or professional group.

4.8 Independent Project

Status: Preferred

Definition: A project whose conclusions and public positions are not controlled by a developer, evaluator, funder, political party, government, or other outside institution.

Usage note: Independence is multidimensional and should be supported by governance, funding disclosure, and publication rights.

4.9 Authority

Status: Preferred

Definition: Legitimate power to make, enforce, recognize, or decide within a defined scope.

Usage note: Authority may be legal, contractual, organizational, technical, or procedural. The source and scope should be stated.

4.10 Mandate

Status: Preferred

Definition: A documented assignment of purpose, scope, authority, responsibility, and limits.

4.11 Jurisdiction

Status: Preferred

Definition: The legal, geographic, organizational, or institutional domain within which an authority or requirement applies.

4.12 Public Claim

Status: Preferred

Definition: A statement intended for or reasonably available to the public concerning a system, organization, evaluation, capability, safeguard, standard, or institutional status.

4.13 Canonical Source

Status: Preferred

Definition: The authoritative Standards Body document governing a defined subject unless formally superseded.

4.14 Source of Truth

Status: Accepted

Definition: The designated authoritative record for a defined set of information.

Usage note: Prefer *canonical source* in formal governance language.

4.15 Working Paper

Status: Preferred status label

Definition: A substantive research document intended for review and revision that does not itself establish a standard or binding requirement.

4.16 Canonical Working White Paper

Status: Preferred status label

Definition: A developed, versioned Standards Body paper that serves as the current authoritative project position on a foundational subject while remaining revisable.

4.17 Approved Foundational Source

Status: Preferred status label

Definition: A canonical document approved to govern a foundational project area such as identity, terminology, evidence, or governance.

5. Artificial Intelligence System Terms

5.1 Artificial Intelligence

Status: Preferred

Definition: Machine-based systems capable of producing outputs such as predictions, recommendations, classifications, content, decisions, or actions in pursuit of explicit or implicit objectives.

Usage note: Avoid treating artificial intelligence as one homogeneous technology.

5.2 AI Model

Status: Preferred

Definition: A computational model whose learned parameters or structured logic produce outputs from inputs.

Do not confuse with: An AI system, which may include prompts, tools, retrieval, interfaces, policies, monitoring, and human processes.

5.3 AI System

Status: Preferred

Definition: The complete operational arrangement through which an AI model is configured, accessed, integrated, monitored, and used.

An AI system may include:

- One or more models
- system prompts
- tools
- retrieval
- memory
- policies
- interfaces
- infrastructure
- human operators
- safeguards
- monitoring
- deployment rules

5.4 Frontier AI

Status: Preferred but scope-sensitive

Definition: AI models or systems at or near the leading edge of broadly relevant capability, scale, autonomy, or strategic significance.

Usage note: The term does not imply that every frontier system is high risk.

5.5 Frontier Model

Status: Preferred but scope-sensitive

Definition: A model at or near the leading edge of relevant capability or scale at a stated time.

Usage note: A model may cease to be frontier as the field changes.

5.6 Frontier AI System

Status: Preferred

Definition: An AI system incorporating one or more frontier models or capabilities and whose behavior depends materially on system configuration and deployment context.

5.7 General-Purpose AI Model

Status: Context-specific

Definition: A model designed or capable of performing a wide range of distinct tasks across domains.

Usage note: Legal definitions may vary by jurisdiction and should be quoted precisely when relevant.

5.8 Advanced AI System

Status: Accepted

Definition: A comparatively capable or complex AI system.

Usage note: Less precise than *frontier AI system*. Use only when the broader term is intended.

5.9 Foundation Model

Status: Accepted

Definition: A model trained on broad data and adaptable to many downstream tasks.

Usage note: Do not treat *foundation model* and *frontier model* as synonyms.

5.10 Open-Weight Model

Status: Preferred

Definition: A model whose learned weights are made available under specified access and license conditions.

Do not confuse with: Fully open-source software or a model with fully disclosed training data and methods.

5.11 Open-Source AI

Status: Context-specific

Definition: AI software, models, data, or infrastructure made available under recognized open-source or open-use conditions.

Usage note: State which components are open.

5.12 Closed Model

Status: Accepted

Definition: A model whose weights or material internal components are not publicly available.

5.13 Model Family

Status: Preferred

Definition: A set of related models sharing lineage, architecture, training approach, or product identity.

5.14 Model Version

Status: Preferred

Definition: A distinct identified release, checkpoint, or deployed state of a model.

5.15 System Version

Status: Preferred

Definition: A distinct identified state of the full AI system, including material configuration and deployment components.

5.16 Checkpoint

Status: Preferred technical term

Definition: A saved state of model parameters at a particular point in training or development.

5.17 Fine-Tuning

Status: Preferred

Definition: Additional training that modifies a pretrained model for a task, behavior, domain, or preference.

5.18 Post-Training

Status: Preferred

Definition: Training or adaptation performed after base pretraining, including instruction tuning, preference optimization, safety tuning, or domain adaptation.

5.19 System Prompt

Status: Preferred

Definition: A high-priority instruction or context provided to a model as part of system configuration.

5.20 Scaffold

Status: Preferred

Definition: Software, prompts, tools, memory, planning loops, agents, or other structure added around a model to support task performance.

5.21 Tool Use

Status: Preferred

Definition: Model interaction with external software, APIs, environments, instruments, or information sources.

5.22 Retrieval-Augmented Generation

Status: Preferred

Definition: Generation supported by retrieving external information and providing it to the model.

5.23 Agent

Status: Context-specific

Definition: An AI system or configuration that selects and executes actions over multiple steps in pursuit of an objective.

Usage note: Do not assume autonomy, persistence, or independent agency beyond the specified design.

5.24 Agentic System

Status: Preferred

Definition: An AI system that plans, acts, observes outcomes, and adapts over multiple steps with some degree of operational autonomy.

5.25 Autonomy

Status: Preferred

Definition: The degree to which a system can select, sequence, and execute actions without direct human instruction at each step.

5.26 Deployment

Status: Preferred

Definition: Making a model or system available for operational use in a defined environment.

5.27 Release

Status: Preferred

Definition: Making a model, system, weights, code, or access mechanism available to a defined audience.

5.28 Open Release

Status: Context-specific

Definition: Release under access conditions permitting broad public use or redistribution.

5.29 Access Tier

Status: Preferred

Definition: A defined level of model or system access with specified permissions, restrictions, identity requirements, or safeguards.

5.30 Deployment Context

Status: Preferred

Definition: The users, environment, permissions, integrations, scale, purpose, and institutional conditions under which an AI system operates.

6. Capability Terms

6.1 Capability

Status: Preferred

Definition: The ability of a model or system to perform a defined task or class of tasks under specified conditions.

Usage note: Capability claims should state the system, conditions, resources, tools, time, and success criteria.

6.2 Demonstrated Capability

Status: Preferred

Definition: A capability supported by observed performance under stated evaluation conditions.

6.3 Latent Capability

Status: Preferred but uncertain

Definition: A capability that may exist in the model or system but has not been reliably elicited or demonstrated under available conditions.

6.4 Effective Capability

Status: Preferred

Definition: Capability available in practical use after accounting for access, scaffolding, reliability, cost, time, safeguards, and deployment constraints.

6.5 Potential Capability

Status: Accepted

Definition: A capability that may emerge under plausible improvements in elicitation, scaffolding, fine-tuning, tools, or system configuration.

6.6 Capability Domain

Status: Preferred

Definition: A field or class of activity in which capability is evaluated.

Examples:

- Cyber operations
- biology
- software engineering
- persuasion
- autonomous research
- critical infrastructure

6.7 Capability Level

Status: Preferred

Definition: A defined degree of capability within a domain or task framework.

6.8 Capability Profile

Status: Preferred

Definition: A multidimensional description of a system's capabilities across tasks, conditions, reliability, cost, autonomy, and limits.

6.9 Capability Graph

Status: Preferred

Definition: A structured representation of related capabilities, prerequisites, task families, and progression within a domain.

6.10 Capability Threshold

Status: Preferred

Definition: A defined boundary in capability evidence used to trigger additional review, safeguards, governance, or other action.

Usage note: A threshold should not be treated as a natural law. It is an institutional decision rule supported by evidence.

6.11 Critical Capability

Status: Preferred

Definition: A capability whose credible possession or availability may create severe, strategic, systemic, or difficult-to-reverse consequences.

6.12 High-Stakes Capability

Status: Preferred

Definition: A capability relevant to decisions with substantial potential harm, public consequence, strategic significance, or institutional importance.

6.13 Dangerous Capability

Status: Context-specific

Definition: A capability that materially enables harmful activity under a stated threat model.

Usage note: Use with a defined domain and actor model.

6.14 Capability Uplift

Status: Preferred

Definition: The increase in task performance, speed, reach, reliability, or sophistication attributable to use of an AI system.

6.15 Human Uplift

Status: Preferred

Definition: Improvement in human performance resulting from AI assistance.

6.16 Superhuman Performance

Status: Context-specific

Definition: Performance exceeding a specified human reference group under comparable conditions.

Usage note: The human reference group and conditions must be stated.

6.17 Generalization

Status: Preferred

Definition: The ability to perform on relevant cases beyond the specific examples used in development or visible evaluation.

6.18 Transfer

Status: Preferred

Definition: Application of learned capability from one task, domain, or context to another.

6.19 Reliability

Status: Preferred

Definition: The consistency with which a system achieves the intended outcome under stated conditions.

6.20 Robustness

Status: Preferred

Definition: The ability to maintain acceptable performance or behavior under variation, disturbance, attack, distribution shift, or uncertainty.

6.21 Task Horizon

Status: Preferred

Definition: The duration, number of steps, complexity, or dependency depth of a task.

6.22 Autonomy Level

Status: Preferred

Definition: A defined degree of independent planning, action, persistence, or control within a system.

7. Risk, Harm, and Safety Terms

7.1 Risk

Status: Preferred

Definition: The combination of the likelihood and consequence of an adverse outcome under a defined context.

Usage note: Capability is not identical to risk.

7.2 Hazard

Status: Preferred

Definition: A source or condition with the potential to cause harm.

7.3 Harm

Status: Preferred

Definition: Adverse effect on people, institutions, rights, systems, property, security, the environment, or public welfare.

7.4 Impact

Status: Preferred

Definition: The realized or reasonably expected consequence of an event, capability, deployment, or decision.

7.5 Consequence

Status: Preferred

Definition: The outcome or effect that follows from an event, action, system behavior, or decision.

7.6 Severity

Status: Preferred

Definition: The magnitude of harm or consequence.

7.7 Likelihood

Status: Preferred

Definition: The estimated probability or plausibility of an event or outcome within a stated period and context.

7.8 Exposure

Status: Preferred

Definition: The degree to which people, systems, assets, or institutions are subject to a hazard.

7.9 Vulnerability

Status: Preferred

Definition: A weakness that can be exploited or activated to produce adverse effects.

7.10 Threat

Status: Preferred

Definition: A potential cause of an unwanted incident, including an actor, event, condition, or capability.

7.11 Threat Actor

Status: Preferred

Definition: A person, group, organization, or state capable of intentionally causing harm.

7.12 Threat Model

Status: Preferred

Definition: A structured description of relevant actors, goals, capabilities, access, constraints, attack paths, and protected assets.

7.13 Misuse

Status: Preferred

Definition: Use of an AI system for a harmful, prohibited, deceptive, or unintended purpose.

7.14 Accident

Status: Preferred

Definition: Harm arising without an actor's deliberate intention to cause that harm.

7.15 Systemic Risk

Status: Preferred but scope-sensitive

Definition: Risk capable of producing broad, interconnected, or cascading effects across systems, sectors, markets, institutions, or populations.

7.16 Catastrophic Risk

Status: Context-specific

Definition: Risk involving exceptionally severe harm at large scale.

Usage note: Use only with explicit scope, pathway, and evidence. Avoid rhetorical use.

7.17 Existential Risk

Status: Context-specific

Definition: Risk threatening the long-term survival or potential of humanity.

Usage note: Use only where directly relevant and carefully supported. It is not a synonym for high-stakes risk.

7.18 High-Stakes Risk

Status: Preferred

Definition: Risk relevant to decisions where error could produce substantial, severe, strategic, or difficult-to-reverse consequences.

7.19 Residual Risk

Status: Preferred

Definition: Risk remaining after controls or safeguards are applied.

7.20 Risk Tolerance

Status: Preferred

Definition: The degree of risk an institution is prepared to accept within a defined context.

7.21 Risk Appetite

Status: Context-specific

Definition: The broad amount and type of risk an organization is willing to pursue or retain.

7.22 Risk Threshold

Status: Preferred

Definition: A defined boundary at which a risk response, escalation, or decision is triggered.

7.23 Risk Indicator

Status: Preferred

Definition: A measurable signal associated with a change in risk.

7.24 Safety

Status: Preferred but highly bounded

Definition: Freedom from unacceptable risk within a defined system, context, period, and set of assumptions.

Usage note: Do not use *safe* as an absolute system label.

7.25 Safe AI

Status: Discouraged

Reason: Too broad and likely to imply universal assurance.

Preferred alternatives:

- Met the specified requirement
- demonstrated the stated safeguard performance
- remained within the defined risk tolerance under assessed conditions

7.26 AI Safety

Status: Accepted but broad

Definition: The field concerned with reducing unacceptable risks and harms associated with AI systems.

Usage note: State the relevant safety domain where possible.

7.27 Security

Status: Preferred

Definition: Protection against unauthorized access, compromise, misuse, interference, theft, disruption, or disclosure.

7.28 Reliability and Safety

Status: Accepted phrase

Usage note: Do not collapse reliability, security, robustness, and safety into one measure.

8. Safeguard and Control Terms

8.1 Safeguard

Status: Preferred

Definition: A technical, procedural, organizational, contractual, or institutional measure intended to reduce risk.

8.2 Control

Status: Preferred

Definition: A measure that modifies risk, supports conformity, or governs system behavior.

8.3 Mitigation

Status: Preferred

Definition: Action taken to reduce the likelihood or consequence of harm.

8.4 Preventive Control

Status: Preferred

Definition: A control intended to prevent an adverse event.

8.5 Detective Control

Status: Preferred

Definition: A control intended to identify an event, deviation, or failure.

8.6 Corrective Control

Status: Preferred

Definition: A control intended to contain, remediate, or recover from failure.

8.7 Compensating Control

Status: Preferred

Definition: An alternative control used when the primary control is unavailable or insufficient.

8.8 Defense in Depth

Status: Preferred

Definition: Use of multiple, independent or partially independent controls so that failure of one does not determine the entire outcome.

8.9 Refusal

Status: Preferred

Definition: A model or system response declining to perform a requested action.

Usage note: Refusal rate is not a complete measure of safeguard quality.

8.10 Access Control

Status: Preferred

Definition: A mechanism governing who or what can access a system, model, capability, tool, or information asset.

8.11 Rate Limit

Status: Preferred

Definition: A restriction on the frequency or quantity of permitted requests or actions.

8.12 Monitoring

Status: Preferred

Definition: Ongoing observation and analysis of system behavior, use, conditions, or controls.

8.13 Human Oversight

Status: Preferred

Definition: Human review, intervention, supervision, or decision authority over an AI system.

8.14 Human in the Loop

Status: Context-specific

Definition: A human participates directly in an operational decision or process step.

Usage note: Do not use as a vague assurance claim. State what the human can see and do.

8.15 Human on the Loop

Status: Context-specific

Definition: A human supervises an automated process and may intervene.

8.16 Human in Command

Status: Context-specific

Definition: Human authority remains responsible for the system's objectives, deployment, and ultimate decisions.

8.17 Rollback

Status: Preferred

Definition: Reversion to a prior system state, configuration, or release.

8.18 Kill Switch

Status: Discouraged

Reason: Often implies a simple universal shutdown mechanism where none exists.

Preferred alternatives:

- Emergency shutdown control
- access revocation
- deployment suspension
- system isolation

8.19 Containment

Status: Preferred

Definition: Limiting the spread, access, effect, or persistence of an adverse condition.

8.20 Safeguard Effectiveness

Status: Preferred

Definition: The degree to which a safeguard reduces risk under a stated threat model and set of conditions.

9. Evaluation and Testing Terms

9.1 Evaluation

Status: Preferred

Definition: A structured process for producing and interpreting evidence about a model, system, method, process, control, or claim.

Usage note: Evaluation includes more than executing tasks. It includes design, administration, scoring, analysis, interpretation, and limitations.

9.2 Test

Status: Preferred

Definition: A defined procedure used to observe or measure one or more characteristics.

9.3 Testing

Status: Preferred

Definition: Execution of one or more tests according to a specified procedure.

Do not confuse with: Evaluation, which includes interpretation and decision context.

9.4 Test Item

Status: Preferred

Definition: A single question, task, scenario, case, prompt, or challenge administered within an evaluation.

9.5 Task

Status: Preferred

Definition: A defined activity the model or system is asked to complete.

9.6 Task Family

Status: Preferred

Definition: A group of tasks measuring related aspects of a capability or construct.

9.7 Scenario

Status: Preferred

Definition: A structured situation describing context, actors, objectives, constraints, and events for evaluation or analysis.

9.8 Benchmark

Status: Preferred

Definition: A standardized set of tasks, procedures, and metrics used to compare performance.

9.9 Public Benchmark

Status: Preferred

Definition: A benchmark whose tasks, data, or procedures are publicly available.

9.10 Held-Out Evaluation

Status: Preferred

Definition: An evaluation using protected tasks, data, environments, scoring elements, or combinations not available to the evaluated party before administration.

9.11 Private Benchmark

Status: Accepted

Definition: A benchmark whose content is not publicly available.

Usage note: Prefer *held-out evaluation* when protection and evidentiary purpose are central.

9.12 Secret Test

Status: Discouraged

Reason: Informal and may imply arbitrary secrecy.

Preferred alternative: Held-out evaluation or protected evaluation material.

9.13 Dynamic Evaluation Protocol

Status: Preferred

Definition: A versioned evaluation system designed to change as models, threats, tasks, evidence, and measurement limits change.

9.14 Static Evaluation

Status: Preferred

Definition: An evaluation whose tasks, methods, and scoring remain fixed during a defined period.

9.15 Evaluation Protocol

Status: Preferred

Definition: The complete versioned specification governing evaluation purpose, scope, tasks, administration, configuration, scoring, analysis, security, reporting, and change control.

9.16 Test Procedure

Status: Preferred

Definition: The operational instructions for executing a particular test.

9.17 Evaluation Harness

Status: Preferred

Definition: Software and infrastructure used to administer tasks, connect models, record outputs, score results, and preserve metadata.

9.18 Evaluation Environment

Status: Preferred

Definition: The technical or simulated setting in which a model or system performs evaluation tasks.

9.19 Reference Environment

Status: Preferred

Definition: A maintained environment used to support comparable evaluation across systems or institutions.

9.20 Evaluation Suite

Status: Preferred

Definition: A defined collection of related tests or task families.

9.21 Evaluation Run

Status: Preferred

Definition: One execution of an evaluation or part of an evaluation under identified conditions.

9.22 Trial

Status: Preferred

Definition: One attempt at a test item or task.

9.23 Replication

Status: Preferred

Definition: Independent or repeated execution intended to determine whether a result can be reproduced under stated conditions.

9.24 Reproduction

Status: Accepted

Definition: Recreating a result using the same or materially equivalent methods and artifacts.

Usage note: Distinguish replication from exact rerun where necessary.

9.25 Reperformance

Status: Preferred assurance term

Definition: Independent execution of a procedure to verify reported work.

9.26 Validation

Status: Preferred but context-sensitive

Definition: Confirmation that a method, model, process, or requirement is suitable for its intended use.

9.27 Verification

Status: Preferred but context-sensitive

Definition: Confirmation through objective evidence that specified requirements have been fulfilled.

9.28 Benchmark Saturation

Status: Preferred

Definition: A condition in which benchmark performance becomes too high or compressed to distinguish meaningful capability differences.

9.29 Contamination

Status: Preferred

Definition: Exposure of evaluation content, answers, or materially similar examples to model training, development, or preparation in a way that weakens the evidentiary value of the result.

9.30 Leakage

Status: Preferred

Definition: Unauthorized or unintended disclosure of protected evaluation material or information.

9.31 Overfitting

Status: Preferred

Definition: Excessive adaptation to a limited dataset, benchmark, or evaluation condition at the expense of generalization.

9.32 Benchmark Gaming

Status: Preferred

Definition: Optimization for benchmark performance in a way that does not produce corresponding improvement in the underlying construct.

9.33 Evaluation Gaming

Status: Preferred

Definition: Strategic behavior by an organization, evaluator, developer, or system that improves the reported result without achieving the intended evaluation objective.

9.34 Sandbagging

Status: Context-specific

Definition: Deliberate underperformance or concealment of capability.

Usage note: Do not infer sandbagging from low performance alone.

9.35 Evaluation Awareness

Status: Preferred

Definition: A system's ability to recognize or infer that it is being evaluated.

9.36 Evaluation Integrity

Status: Preferred

Definition: The degree to which evaluation design, administration, security, scoring, evidence, and reporting preserve the intended meaning of the result.

10. Measurement and Statistical Terms

10.1 Construct

Status: Preferred

Definition: The underlying concept or attribute an evaluation intends to measure.

10.2 Construct Validity

Status: Preferred

Definition: The degree to which an evaluation meaningfully measures the intended construct.

10.3 Content Validity

Status: Preferred

Definition: The degree to which evaluation content adequately represents the relevant domain.

10.4 Criterion Validity

Status: Preferred

Definition: The degree to which evaluation results correspond with an external outcome or reference.

10.5 Ecological Validity

Status: Preferred

Definition: The degree to which evaluation conditions reflect relevant real-world contexts.

10.6 Internal Validity

Status: Preferred

Definition: The degree to which observed results support the claimed causal or comparative interpretation within the study.

10.7 External Validity

Status: Preferred

Definition: The degree to which findings generalize beyond the assessed sample, setting, or conditions.

10.8 Reliability of Measurement

Status: Preferred

Definition: The consistency of a measurement process across repetitions, items, raters, or conditions.

10.9 Metric

Status: Preferred

Definition: A defined quantitative or categorical measure.

10.10 Score

Status: Preferred

Definition: A value assigned according to a defined scoring procedure.

10.11 Scoring Rule

Status: Preferred

Definition: The method used to convert outputs or observations into scores.

10.12 Rubric

Status: Preferred

Definition: Structured criteria used to judge or classify performance.

10.13 Threshold

Status: Preferred

Definition: A defined boundary used for classification, escalation, or decision.

Usage note: Always state what the threshold triggers.

10.14 Cut Score

Status: Context-specific

Definition: A score used to separate categories or decisions.

10.15 Baseline

Status: Preferred

Definition: A reference level of performance or condition used for comparison.

10.16 Human Baseline

Status: Preferred

Definition: Performance by a defined human reference group under stated conditions.

10.17 Reference Group

Status: Preferred

Definition: The population or category used as a comparison.

10.18 Confidence

Status: Preferred

Definition: The degree of justified belief in a conclusion based on available evidence.

Usage note: Distinguish qualitative confidence from statistical confidence.

10.19 Confidence Interval

Status: Preferred statistical term

Definition: An interval estimated by a defined procedure that expresses statistical uncertainty around a parameter.

10.20 Uncertainty

Status: Preferred

Definition: The degree and sources of incomplete knowledge concerning a measurement, result, interpretation, or future outcome.

10.21 Measurement Uncertainty

Status: Preferred

Definition: Uncertainty associated with a measured or estimated value.

10.22 Epistemic Uncertainty

Status: Preferred

Definition: Uncertainty arising from limited knowledge, evidence, models, or understanding.

10.23 Aleatoric Uncertainty

Status: Context-specific

Definition: Uncertainty arising from inherent variability or randomness.

10.24 Error

Status: Preferred

Definition: Difference between an observed, estimated, recorded, or reported value and the relevant correct or reference value.

10.25 False Positive

Status: Preferred

Definition: A positive finding when the relevant condition is absent.

10.26 False Negative

Status: Preferred

Definition: A negative finding when the relevant condition is present.

10.27 Sensitivity

Status: Preferred statistical term

Definition: The ability of a method to identify relevant positive cases or detect meaningful change.

10.28 Specificity

Status: Preferred statistical term

Definition: The ability of a method to exclude relevant negative cases.

10.29 Calibration

Status: Preferred

Definition: The degree to which stated probabilities or confidence levels correspond to observed frequencies or outcomes.

10.30 Discrimination

Status: Preferred

Definition: The ability of a metric or task to distinguish among relevant levels of performance.

10.31 Statistical Power

Status: Preferred

Definition: The probability that a study or test will detect an effect of a specified size when it exists.

10.32 Sample

Status: Preferred

Definition: A subset selected from a broader population or task universe.

10.33 Task Population

Status: Preferred

Definition: The broader set of relevant tasks that an evaluation sample is intended to represent.

10.34 Distribution Shift

Status: Preferred

Definition: A change between the data, tasks, users, or environment used in development or evaluation and those encountered later.

10.35 Aggregate Score

Status: Preferred

Definition: A combined score derived from multiple items, tasks, domains, or metrics.

10.36 Composite Score

Status: Preferred

Definition: A score combining different measures according to a defined method.

Usage note: Composite scores can conceal important tradeoffs and should be decomposable.

10.37 Ordinal Scale

Status: Preferred

Definition: A scale indicating order without guaranteeing equal distance between levels.

10.38 Ceiling Effect

Status: Preferred

Definition: Reduced ability to distinguish performance because results cluster near the maximum.

10.39 Floor Effect

Status: Preferred

Definition: Reduced ability to distinguish performance because results cluster near the minimum.

11. Elicitation Terms

11.1 Elicitation

Status: Preferred

Definition: The process of configuring prompts, tools, examples, scaffolds, resources, or procedures to reveal a model or system's available capability.

11.2 Elicitation Budget

Status: Preferred

Definition: The time, compute, attempts, human effort, tools, and resources permitted for capability elicitation.

11.3 Prompting

Status: Preferred

Definition: Providing instructions, examples, context, or questions to a model.

11.4 Few-Shot Prompting

Status: Preferred

Definition: Prompting that includes a small number of examples.

11.5 Chain-of-Thought

Status: Context-specific

Definition: Intermediate reasoning text generated or requested during problem solving.

Usage note: Do not assume that visible chain-of-thought faithfully represents internal model reasoning.

11.6 Tool-Augmented Elicitation

Status: Preferred

Definition: Elicitation using tools or external systems.

11.7 Fine-Tuning-Based Elicitation

Status: Preferred

Definition: Elicitation that uses additional training to reveal or improve performance.

11.8 Best-Effort Elicitation

Status: Preferred but should be specified

Definition: A documented attempt to obtain strong performance using reasonable available methods and resources.

Usage note: The phrase is incomplete without an elicitation budget and method description.

11.9 Capability Ceiling Estimate

Status: Preferred

Definition: An estimate of the highest performance plausibly available under defined elicitation conditions.

11.10 Elicitation Failure

Status: Preferred

Definition: Failure to reveal capability because the elicitation method, integration, tools, instructions, or resources were insufficient.

12. Evidence and Epistemic Terms

12.1 Evidence

Status: Preferred

Definition: Information, observation, artifact, record, result, or testimony relevant to supporting or challenging a claim.

12.2 Direct Evidence

Status: Preferred

Definition: Evidence derived directly from the object, event, system, or process at issue.

12.3 Indirect Evidence

Status: Preferred

Definition: Evidence supporting a claim through inference rather than direct observation.

12.4 Primary Source

Status: Preferred

Definition: An original source of data, law, standards text, evaluation results, institutional policy, or first-hand record.

12.5 Secondary Source

Status: Preferred

Definition: A source interpreting, summarizing, or analyzing primary material.

12.6 Expert Judgment

Status: Preferred

Definition: A reasoned assessment by a qualified person based on relevant knowledge and evidence.

12.7 Expert Opinion

Status: Accepted

Definition: A conclusion or interpretation offered by an expert.

Usage note: Prefer *expert judgment* where a structured evidentiary process exists.

12.8 Claim

Status: Preferred

Definition: A proposition asserted to be true, supported, or justified.

12.9 Finding

Status: Preferred

Definition: A supported conclusion arising from evaluation, review, inspection, audit, or analysis.

12.10 Conclusion

Status: Preferred

Definition: A reasoned determination based on evidence and analysis.

12.11 Inference

Status: Preferred

Definition: A conclusion derived from evidence rather than directly observed.

12.12 Assumption

Status: Preferred

Definition: A proposition accepted for purposes of analysis without full proof.

12.13 Hypothesis

Status: Preferred

Definition: A testable proposed explanation or prediction.

12.14 Open Question

Status: Preferred

Definition: A materially unresolved question requiring further evidence or analysis.

12.15 Knowledge Gap

Status: Preferred

Definition: A missing body of evidence, method, understanding, or institutional capability.

12.16 Evidence Gap

Status: Preferred

Definition: Missing evidence necessary to support or resolve a defined claim.

12.17 Evidence Standard

Status: Preferred

Definition: The level, type, quality, and sufficiency of evidence required for a defined claim or decision.

12.18 Burden of Proof

Status: Context-specific

Definition: Responsibility to provide sufficient evidence for a claim or decision.

12.19 Weight of Evidence

Status: Preferred

Definition: The combined strength, relevance, quality, consistency, and independence of available evidence.

12.20 Corroboration

Status: Preferred

Definition: Independent or distinct evidence supporting the same conclusion.

12.21 Triangulation

Status: Preferred

Definition: Use of multiple methods, sources, or perspectives to evaluate a claim.

12.22 Provenance

Status: Preferred

Definition: The origin, history, custody, modification, and ownership of evidence or an artifact.

12.23 Chain of Custody

Status: Preferred

Definition: A documented record of possession, transfer, access, and handling of evidence or protected material.

12.24 Traceability

Status: Preferred

Definition: The ability to connect a claim, result, decision, or artifact to its sources, methods, versions, and responsible parties.

12.25 Reproducibility

Status: Preferred

Definition: The degree to which a result can be obtained again using the same or adequately specified data, code, methods, and conditions.

12.26 Replicability

Status: Preferred

Definition: The degree to which a finding can be supported through an independent study or materially independent implementation.

12.27 Falsifiability

Status: Preferred

Definition: The degree to which a claim can be meaningfully tested and potentially shown false.

12.28 Negative Result

Status: Preferred

Definition: A result that does not support the tested hypothesis, expected effect, or claimed capability.

Usage note: A negative result is not automatically evidence of absence.

12.29 Absence of Evidence

Status: Preferred phrase

Definition: Lack of observed evidence supporting a claim.

12.30 Evidence of Absence

Status: Preferred phrase

Definition: Evidence that meaningfully supports the conclusion that a condition is absent under stated bounds.

12.31 Epistemic Status

Status: Preferred

Definition: A concise description of how strongly a claim is supported and what uncertainty remains.

12.32 Confidence Rating

Status: Preferred

Definition: A defined qualitative or quantitative expression of confidence in a conclusion.

13. Review Terms

13.1 Review

Status: Preferred

Definition: Structured examination of evidence, methods, reasoning, processes, or claims.

13.2 Peer Review

Status: Preferred

Definition: Review by people with relevant expertise who are peers in the relevant field or practice.

13.3 Internal Review

Status: Preferred

Definition: Review performed within the organization responsible for the work or claim.

13.4 External Review

Status: Preferred

Definition: Review performed by a person or organization outside the reviewed organization.

Usage note: External does not automatically mean independent.

13.5 Independent Expert Review

Status: Preferred

Definition: Review by qualified people or institutions sufficiently free from controlling conflicts to form and communicate a genuine judgment.

13.6 Technical Review

Status: Preferred

Definition: Review focused on technical methods, evidence, systems, or results.

13.7 Methodological Review

Status: Preferred

Definition: Review focused on design, validity, measurement, assumptions, and analysis.

13.8 Security Review

Status: Preferred

Definition: Review focused on threats, vulnerabilities, controls, access, and information protection.

13.9 Governance Review

Status: Preferred

Definition: Review focused on authority, roles, conflicts, decisions, accountability, and institutional process.

13.10 Public-Interest Review

Status: Preferred

Definition: Review focused on public consequences, affected parties, rights, access, distribution, and institutional legitimacy.

13.11 Review Panel

Status: Preferred

Definition: A group appointed to conduct or advise a review.

13.12 Review Chair

Status: Preferred

Definition: The person responsible for coordinating a review process and preserving procedural integrity.

13.13 Review Mandate

Status: Preferred

Definition: The documented question, scope, authority, access, methods, outputs, constraints, and decision relationship of a review.

13.14 Review Finding

Status: Preferred

Definition: A supported conclusion produced through review.

13.15 Dissent

Status: Preferred

Definition: A reasoned disagreement with the primary or majority conclusion.

13.16 Minority Report

Status: Preferred

Definition: A documented dissenting assessment issued alongside the principal review conclusion.

13.17 Right of Reply

Status: Preferred

Definition: Opportunity for the reviewed party to respond to factual or procedural findings without obtaining veto power over the conclusion.

13.18 Factual Correction Review

Status: Preferred

Definition: Review of a draft for factual error, misunderstood configuration, confidentiality, or security.

Usage note: It should not become conclusion negotiation.

13.19 Meta-Review

Status: Preferred

Definition: Review of the quality, consistency, or validity of another review.

13.20 Review Independence

Status: Preferred

Definition: The degree to which reviewers can select methods, access evidence, interpret results, communicate findings, and resist improper influence.

14. Audit, Inspection, and Assurance Terms

14.1 Audit

Status: Preferred but criteria-dependent

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

Usage note: Do not call every review an audit.

14.2 AI Audit

Status: Context-specific

Definition: An audit concerning an AI model, system, organization, process, control, or requirement.

Usage note: The criteria and scope must be stated.

14.3 Inspection

Status: Preferred

Definition: Examination of a product, process, service, system, installation, or design and determination of conformity with specified or professional requirements.

14.4 Assessment

Status: Accepted broad term

Definition: A structured process of determining characteristics, condition, performance, conformity, or risk.

Usage note: Prefer a more specific term where possible.

14.5 Assurance

Status: Preferred

Definition: Confidence supported by evidence, review, and institutional process that a claim or requirement is sufficiently reliable for a defined purpose.

14.6 Assurance Engagement

Status: Preferred

Definition: A structured engagement in which an independent practitioner evaluates evidence and provides a conclusion or defined findings.

14.7 Limited Assurance

Status: Context-specific

Definition: Assurance providing moderate confidence based on narrower evidence or procedures than reasonable assurance.

14.8 Reasonable Assurance

Status: Context-specific

Definition: A high but not absolute level of assurance supported by sufficiently extensive evidence and procedures.

14.9 Absolute Assurance

Status: Prohibited as a practical claim

Reason: No evaluation or audit can eliminate all uncertainty.

14.10 Continuous Assurance

Status: Preferred

Definition: Ongoing or repeated evidence collection, evaluation, monitoring, and review rather than a one-time assessment.

14.11 Attestation

Status: Preferred conformity term

Definition: Issue of a statement, based on a decision following review, that fulfillment of specified requirements has been demonstrated.

14.12 Conformity Assessment

Status: Preferred

Definition: Demonstration that specified requirements relating to a product, process, system, person, or body are fulfilled.

14.13 First-Party Assessment

Status: Preferred

Definition: Assessment performed by the organization responsible for the object or claim.

14.14 Second-Party Assessment

Status: Preferred

Definition: Assessment performed by a party with a user, purchaser, contractual, or direct stakeholder interest.

14.15 Third-Party Assessment

Status: Preferred

Definition: Assessment performed by a body sufficiently independent of provider and immediate user interests to support an impartial judgment.

14.16 Third-Party Evaluator

Status: Preferred

Definition: An organization or person performing an evaluation in a third-party role.

14.17 Auditor

Status: Context-specific

Definition: A person qualified and authorized to perform an audit against defined criteria.

Usage note: Do not use as a prestige label for general evaluators.

14.18 Evaluator

Status: Preferred

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

14.19 Evaluation Laboratory

Status: Preferred

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

14.20 Conformity-Assessment Body

Status: Preferred

Definition: A body performing conformity-assessment activities such as testing, inspection, certification, validation, or verification.

14.21 CAB

Status: Accepted abbreviation after first use

Definition: Conformity-assessment body.

14.22 Assurance Level

Status: Preferred

Definition: A defined degree of review depth, evidence, access, independence, continuity, and confidence.

14.23 Scope of Assurance

Status: Preferred

Definition: The exact claims, systems, versions, requirements, and period covered by an assurance conclusion.

15. Certification and Accreditation Terms

15.1 Certification

Status: Preferred but strictly controlled

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

Usage note: Certification does not mean universal safety or government approval.

15.2 Certification Body

Status: Preferred

Definition: A third-party body that conducts certification under a defined scheme.

15.3 Certificate

Status: Preferred

Definition: A formal record of certification identifying the subject, scheme, scope, issuer, version, date, and validity.

15.4 Certification Scheme

Status: Preferred

Definition: The rules, procedures, governance, criteria, assessment methods, surveillance, claims, and responsibilities governing certification.

15.5 Certification Mark

Status: Preferred

Definition: A controlled mark indicating certification under a defined scheme.

15.6 Certified Safe

Status: Prohibited unless quoting another source

Reason: Certification is always against specified requirements and cannot establish universal safety.

15.7 Accreditation

Status: Preferred but strictly controlled

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

15.8 Accreditation Body

Status: Preferred

Definition: A body that performs accreditation.

15.9 Scope of Accreditation

Status: Preferred

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

15.10 Accredited Evaluator

Status: Context-specific

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

Usage note: Never omit the scope.

15.11 Self-Accreditation

Status: Prohibited as a legitimacy claim

Reason: Accreditation requires independent recognition.

15.12 Accreditation-Like Review

Status: Discouraged

Preferred alternative: Evaluator qualification review or pilot recognition process.

15.13 Recognition

Status: Preferred

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

15.14 Mutual Recognition

Status: Preferred

Definition: An arrangement through which participating institutions accept specified results, qualifications, certificates, or decisions issued under another recognized system.

15.15 Unilateral Recognition

Status: Preferred

Definition: Acceptance by one institution of another institution's evidence or status without reciprocal obligation.

15.16 Recognition Scope

Status: Preferred

Definition: The defined activities, evidence, domains, methods, or decisions covered by recognition.

15.17 Approval

Status: Discouraged unless authority is clear

Definition: Formal acceptance by an authorized body.

Usage note: State who approved what and under which authority.

15.18 Official Approval

Status: Prohibited for Standards Body's present work

15.19 Endorsement

Status: Context-specific

Definition: Public expression of support.

Usage note: Partnership, citation, or participation does not automatically imply endorsement.

16. Standards and Requirements Terms

16.1 Standard

Status: Preferred

Definition: A document established through a recognized process that provides rules, requirements, guidelines, characteristics, or common practices for repeated use.

Usage note: A standard may be voluntary.

16.2 Proposed Standard

Status: Preferred

Definition: A draft or candidate standard not yet approved through its governing process.

16.3 Technical Standard

Status: Preferred

Definition: A standard specifying technical requirements, methods, measurements, interfaces, or performance characteristics.

16.4 Management-System Standard

Status: Preferred

Definition: A standard defining organizational requirements for establishing, operating, maintaining, and improving a management system.

16.5 Performance Standard

Status: Preferred

Definition: A standard defining required outcomes or performance without necessarily prescribing the exact method.

16.6 Process Standard

Status: Preferred

Definition: A standard defining required procedures, controls, records, or workflows.

16.7 Interface Standard

Status: Preferred

Definition: A standard defining how systems, data, protocols, or institutions interact.

16.8 Voluntary Standard

Status: Preferred

Definition: A standard whose adoption is not legally required by default.

16.9 International Standard

Status: Context-specific

Definition: A standard developed and approved through a recognized international standards process.

Usage note: Do not use merely because a document has global aspirations.

16.10 Consensus Standard

Status: Preferred

Definition: A standard developed through a process intended to balance relevant interests and resolve comments toward broad agreement.

Usage note: Consensus does not require unanimity.

16.11 Specification

Status: Preferred

Definition: A detailed description of technical, procedural, or functional requirements.

16.12 Technical Specification

Status: Preferred

Definition: A specification focused on technical requirements or methods.

16.13 Recommended Practice

Status: Preferred

Definition: Nonbinding guidance describing a preferred method based on available evidence and professional judgment.

16.14 Guidance

Status: Preferred

Definition: Advisory material explaining principles, interpretation, or implementation.

16.15 Framework

Status: Preferred

Definition: A structured set of concepts, functions, outcomes, or practices that supports consistent thinking and action while allowing adaptation.

16.16 Code of Conduct

Status: Preferred

Definition: A set of expected behaviors or commitments adopted voluntarily or recognized by an institution.

16.17 Code of Practice

Status: Preferred

Definition: Operational guidance describing accepted ways to implement or satisfy expectations.

16.18 Requirement

Status: Preferred

Definition: A condition that must be fulfilled within a defined context.

16.19 Voluntary Requirement

Status: Preferred

Definition: A requirement accepted through choice, contract, framework, membership, or certification scheme.

16.20 Mandatory Requirement

Status: Preferred

Definition: A requirement imposed through law, regulation, contract, or authorized institutional rule.

16.21 Binding Requirement

Status: Preferred

Definition: A requirement carrying enforceable legal, contractual, or organizational effect.

16.22 Best Practice

Status: Discouraged unless evidence is strong

Reason: Often used without proof that the practice is best.

Preferred alternatives:

- Recommended practice
- established practice
- current leading practice

16.23 Good Practice

Status: Accepted

Definition: A practice supported by relevant experience or evidence as useful or responsible.

16.24 Minimum Requirement

Status: Preferred

Definition: The least condition that must be met within a scheme or rule.

16.25 Baseline Requirement

Status: Preferred

Definition: A common foundational requirement applying across a defined class.

16.26 Progressive Requirement

Status: Preferred

Definition: A requirement whose rigor, coverage, assurance, or consequence increases through defined stages or triggers.

16.27 Tiered Requirement

Status: Preferred

Definition: A requirement organized into levels of increasing rigor or applicability.

16.28 Performance-Based Requirement

Status: Preferred

Definition: A requirement specifying an outcome rather than one required method.

16.29 Prescriptive Requirement

Status: Preferred

Definition: A requirement specifying a particular method, control, procedure, or design.

16.30 Risk-Based Requirement

Status: Preferred

Definition: A requirement whose applicability or rigor varies according to risk.

16.31 Capability-Based Requirement

Status: Preferred

Definition: A requirement triggered by demonstrated or reasonably anticipated capability.

16.32 Deployment-Based Requirement

Status: Preferred

Definition: A requirement triggered by deployment context, access, scale, autonomy, or use.

16.33 Incorporation by Reference

Status: Preferred legal term

Definition: Use of an external standard or document as part of a binding requirement.

16.34 Presumption of Conformity

Status: Preferred legal or scheme term

Definition: A presumption that compliance with a recognized standard provides evidence of compliance with a specified requirement.

16.35 Safe Harbor

Status: Preferred legal term

Definition: A provision offering defined protection or reduced exposure when specified practices are followed.

16.36 Regulatory Sandbox

Status: Preferred

Definition: A controlled environment for testing innovations under oversight and defined conditions.

16.37 Phase-In

Status: Preferred

Definition: Staged introduction of a requirement across time, systems, actors, or risk tiers.

16.38 Sunset Clause

Status: Preferred

Definition: A provision causing a requirement or program to expire unless renewed.

16.39 Review Clause

Status: Preferred

Definition: A requirement that a rule, standard, or program be reassessed after a period or trigger event.

17. Compliance Terms

17.1 Compliance

Status: Preferred but source-dependent

Definition: Fulfillment of applicable legal, regulatory, contractual, or organizational obligations.

Usage note: State the governing requirement.

17.2 Conformity

Status: Preferred

Definition: Fulfillment of specified requirements.

Distinction: Conformity is broader and may apply to voluntary standards. Compliance often implies binding obligations.

17.3 Nonconformity

Status: Preferred

Definition: Failure to fulfill a specified requirement.

17.4 Noncompliance

Status: Preferred

Definition: Failure to fulfill an applicable binding obligation.

17.5 Compliant AI

Status: Discouraged

Reason: Too broad. Compliance always depends on a defined requirement and jurisdiction.

17.6 Compliance Certification

Status: Context-specific

Usage note: Use only when a valid certification scheme assesses defined compliance requirements.

17.7 Compliance Review

Status: Preferred

Definition: Structured review of adherence to a defined legal, contractual, policy, or framework requirement.

17.8 Compliance Theater

Status: Preferred critical term

Definition: Formal or documentary conformity without meaningful achievement of the underlying objective.

17.9 Checklist Compliance

Status: Preferred critical term

Definition: Narrow completion of listed items without sufficient attention to actual effectiveness or context.

17.10 Legal Compliance Opinion

Status: Context-specific

Definition: A legal determination or opinion provided by a qualified legal authority or professional.

Usage note: Standards Body does not issue legal compliance opinions unless a formally qualified and authorized legal process exists.

18. Governance Terms

18.1 Governance

Status: Preferred

Definition: The system of authority, roles, decisions, accountability, oversight, and control through which an organization or activity is directed.

18.2 Governance Framework

Status: Preferred

Definition: A documented structure defining governance principles, roles, procedures, decisions, and accountability.

18.3 Governing Body

Status: Preferred

Definition: The body with formal authority to direct or oversee an organization or scheme.

18.4 Board

Status: Context-specific

Definition: A formally constituted governing or oversight body.

18.5 Secretariat

Status: Preferred

Definition: The administrative and coordination function supporting governance, committees, records, and process.

18.6 Committee

Status: Preferred

Definition: A formally appointed group responsible for a defined area of work or decision.

18.7 Working Group

Status: Preferred

Definition: A group formed to develop, analyze, or recommend work on a defined subject.

18.8 Technical Committee

Status: Preferred

Definition: A committee responsible for technical standards, methods, or decisions.

18.9 Advisory Group

Status: Preferred

Definition: A group providing advice without final decision authority unless expressly granted.

18.10 Decision Right

Status: Preferred

Definition: Authority to make a defined decision.

18.11 Accountability

Status: Preferred

Definition: Obligation to explain, justify, and accept responsibility for decisions and outcomes.

18.12 Responsibility

Status: Preferred

Definition: Assigned duty to perform or ensure an activity.

18.13 Oversight

Status: Preferred

Definition: Independent or supervisory observation and review of an activity, institution, or decision.

18.14 Delegation

Status: Preferred

Definition: Assignment of authority or responsibility from one body to another.

18.15 Escalation

Status: Preferred

Definition: Transfer of an issue or decision to a higher or different authority because of consequence, conflict, uncertainty, or defined trigger.

18.16 Appeal

Status: Preferred

Definition: A formal request for review of a decision.

18.17 Complaint

Status: Preferred

Definition: An expression of dissatisfaction or allegation concerning process, conduct, evidence, or outcome that requires documented handling.

18.18 Recusal

Status: Preferred

Definition: Withdrawal from participation in a decision or review because of a conflict or other disqualifying condition.

18.19 Quorum

Status: Preferred

Definition: The minimum participation required for a body to conduct official business.

18.20 Consensus

Status: Preferred

Definition: Broad agreement characterized by resolution of substantial objections, without requiring unanimity.

18.21 Unanimity

Status: Preferred

Definition: Agreement of every eligible decision participant.

18.22 Majority Decision

Status: Preferred

Definition: A decision supported by more than half of eligible votes under defined rules.

18.23 Minority Position

Status: Preferred

Definition: A documented position not adopted by the majority.

18.24 Emergency Decision

Status: Preferred

Definition: A time-sensitive decision made under a defined emergency process with limited scope and subsequent review.

18.25 Institutional Legitimacy

Status: Preferred

Definition: Justified acceptance of an institution's role and decisions based on competence, process, participation, accountability, authority, and public purpose.

18.26 Institutional Capture

Status: Preferred

Definition: Distortion of an institution's decisions or agenda by a stakeholder, funder, industry, government, ideology, or professional group.

19. Independence, Impartiality, and Conflict Terms

19.1 Independence

Status: Preferred

Definition: Sufficient freedom from controlling relationships or interests to form and communicate a genuine judgment.

19.2 Impartiality

Status: Preferred

Definition: Presence of objectivity and management of conflicts so that judgment is not improperly influenced.

19.3 Objectivity

Status: Preferred

Definition: Reliance on evidence and defined criteria rather than improper preference or interest.

19.4 Neutrality

Status: Discouraged when used as a universal institutional claim

Reason: Institutions and decisions always operate within values, mandates, and assumptions.

Preferred alternatives:

- Independent
- impartial
- evidence-based
- nonpartisan

19.5 Conflict of Interest

Status: Preferred

Definition: A relationship, incentive, obligation, or commitment that could impair, or reasonably appear to impair, impartial judgment.

19.6 Material Conflict

Status: Preferred

Definition: A conflict significant enough to require disclosure, mitigation, role limitation, recusal, or exclusion.

19.7 Financial Conflict

Status: Preferred

Definition: A conflict arising from payment, investment, equity, grants, employment, or economic dependence.

19.8 Organizational Conflict

Status: Preferred

Definition: A conflict arising from governance, ownership, institutional role, or reporting relationships.

19.9 Intellectual Conflict

Status: Preferred

Definition: A conflict arising from strong prior commitments, proprietary methods, public positions, or reputational stakes.

19.10 Political Conflict

Status: Preferred

Definition: A conflict arising from party, government, national, or political interests.

19.11 Reviewer Capture

Status: Preferred

Definition: Material influence over a reviewer by the reviewed organization, funder, professional community, government, or ideology.

19.12 Client Capture

Status: Preferred

Definition: Dependence on a client relationship that weakens willingness or ability to issue unfavorable findings.

19.13 Regulatory Capture

Status: Preferred

Definition: Distortion of regulatory action toward the interests of regulated entities or other powerful stakeholders.

19.14 Disclosure

Status: Preferred

Definition: Communication of relevant information to defined recipients.

19.15 Conflict Disclosure

Status: Preferred

Definition: Documentation of relationships or interests relevant to impartiality.

19.16 Mitigation of Conflict

Status: Preferred

Definition: Action to reduce the effect or appearance of a conflict.

19.17 Independence Profile

Status: Preferred

Definition: A structured account of organizational, financial, methodological, informational, operational, publication, intellectual, political, and security independence.

20. Transparency and Information-Access Terms

20.1 Transparency

Status: Preferred

Definition: Availability of sufficient information to understand relevant decisions, methods, interests, evidence, and limitations.

20.2 Opacity

Status: Preferred

Definition: Lack of visibility into relevant methods, evidence, decisions, or system behavior.

20.3 Confidentiality

Status: Preferred

Definition: Protection of information from unauthorized disclosure.

20.4 Secrecy

Status: Discouraged unless describing intentional concealment

Preferred alternatives:

- Confidential
- restricted
- protected
- nonpublic

20.5 Public Information

Status: Preferred

Definition: Information available without access restriction.

20.6 Controlled Information

Status: Preferred

Definition: Information available only under defined access and use conditions.

20.7 Confidential Information

Status: Preferred

Definition: Nonpublic information protected by legal, contractual, security, privacy, or institutional duties.

20.8 Restricted Information

Status: Preferred

Definition: Sensitive information subject to stronger access limitations than ordinary confidential material.

20.9 Redaction

Status: Preferred

Definition: Removal or obscuring of information before disclosure.

20.10 Public Summary

Status: Preferred

Definition: A nontechnical or reduced-detail account of a finding or process suitable for public release.

20.11 Disclosure Level

Status: Preferred

Definition: A defined category governing who may access information and under what conditions.

20.12 Need to Know

Status: Preferred

Definition: Access principle limiting information to people whose role requires it.

20.13 Least Privilege

Status: Preferred

Definition: Access principle granting only the minimum permissions necessary for a role or task.

20.14 Information Hazard

Status: Preferred

Definition: Information whose creation or disclosure may materially increase risk.

20.15 Responsible Disclosure

Status: Preferred

Definition: Coordinated reporting and communication of vulnerabilities, incidents, or sensitive findings in a manner intended to support remediation and reduce harm.

20.16 Publication Right

Status: Preferred

Definition: The right to communicate findings subject to defined legal, security, and confidentiality constraints.

20.17 Publication Veto

Status: Preferred critical term

Definition: Power to prevent publication of a finding.

Usage note: A reviewed party should not possess unrestricted publication veto over independent conclusions.

21. Security Terms

21.1 Information Security

Status: Preferred

Definition: Protection of information confidentiality, integrity, and availability.

21.2 Cybersecurity

Status: Preferred

Definition: Protection of digital systems, networks, software, data, and services against unauthorized activity or disruption.

21.3 Model Security

Status: Preferred

Definition: Protection of model weights, access, behavior, interfaces, and associated assets against theft, compromise, or misuse.

21.4 Weight Security

Status: Preferred

Definition: Protection of model weights from unauthorized access, theft, copying, or modification.

21.5 Secure Evaluation

Status: Preferred

Definition: Evaluation performed under controls appropriate to the sensitivity of the model, tasks, evidence, and outputs.

21.6 Secure Enclave

Status: Context-specific

Definition: A controlled technical environment designed to isolate sensitive computation or data.

21.7 Sandbox

Status: Preferred technical term

Definition: An isolated environment limiting the effect of system actions.

Do not confuse with: Regulatory sandbox.

21.8 Air Gap

Status: Context-specific

Definition: Physical or logical separation from external networks.

21.9 Threat Surface

Status: Preferred

Definition: The set of points through which a system may be attacked or compromised.

21.10 Attack Surface

Status: Preferred

Definition: The exposed components, interfaces, permissions, and pathways available to an attacker.

21.11 Security Incident

Status: Preferred

Definition: An event compromising or threatening confidentiality, integrity, availability, access control, or system security.

21.12 Security Classification

Status: Preferred

Definition: A category governing protection, access, handling, and disclosure of information or assets.

21.13 Dual-Use Information

Status: Preferred

Definition: Information with legitimate beneficial uses and plausible harmful uses.

21.14 Sensitive Capability Information

Status: Preferred

Definition: Information about model capabilities, methods, or weaknesses whose disclosure may increase misuse or security risk.

22. Incident Terms

22.1 Incident

Status: Preferred

Definition: An event or condition that caused, could have caused, or revealed material harm, failure, compromise, misuse, or loss of control.

22.2 AI Incident

Status: Preferred

Definition: An incident in which an AI model or system materially contributed to the event, condition, or consequence.

22.3 Safety Incident

Status: Preferred

Definition: An incident involving unacceptable risk or harm related to system safety.

22.4 Evaluation Incident

Status: Preferred

Definition: An event compromising evaluation validity, integrity, security, evidence, or interpretation.

22.5 Misuse Incident

Status: Preferred

Definition: Harmful or prohibited use of an AI system.

22.6 Near Miss

Status: Preferred

Definition: An event that could have caused material harm but did not, because of chance, detection, intervention, or incomplete progression.

22.7 Adverse Event

Status: Accepted

Definition: An event producing harm or unacceptable effect.

22.8 Failure

Status: Preferred

Definition: Inability of a system, process, control, or institution to fulfill an intended function or requirement.

22.9 Failure Mode

Status: Preferred

Definition: A specific way in which a system, process, control, or institution can fail.

22.10 Root Cause

Status: Preferred

Definition: An underlying cause or set of causes whose correction would materially reduce recurrence.

22.11 Contributing Factor

Status: Preferred

Definition: A condition that increased the likelihood or severity of an incident without being the sole cause.

22.12 Corrective Action

Status: Preferred

Definition: Action taken to eliminate or reduce the cause of a detected failure or nonconformity.

22.13 Preventive Action

Status: Preferred

Definition: Action taken to reduce the likelihood of a potential failure or incident.

22.14 Remediation

Status: Preferred

Definition: Action taken to correct, contain, or repair a problem.

22.15 Incident Response

Status: Preferred

Definition: The organized process of detecting, assessing, containing, communicating, remediating, and learning from an incident.

22.16 Incident Report

Status: Preferred

Definition: A structured record of an incident, evidence, impact, response, and lessons.

22.17 Incident Taxonomy

Status: Preferred

Definition: A classification system for incident types, severity, causes, and effects.

23. Interoperability Terms

23.1 Interoperability

Status: Preferred

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

23.2 Semantic Interoperability

Status: Preferred

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

23.3 Syntactic Interoperability

Status: Preferred

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

23.4 Technical Interoperability

Status: Preferred

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

23.5 Measurement Interoperability

Status: Preferred

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

23.6 Institutional Interoperability

Status: Preferred

Definition: Ability of institutions with different mandates and structures to coordinate and rely on one another's work.

23.7 Legal Interoperability

Status: Preferred

Definition: Ability to map or coordinate evidence and requirements across legal systems while preserving jurisdictional authority.

23.8 Harmonization

Status: Preferred

Definition: Reduction of differences among requirements, methods, or standards.

Distinction: Interoperability can exist without full harmonization.

23.9 Compatibility

Status: Preferred

Definition: The degree to which systems, methods, or requirements can operate together without unacceptable conflict.

23.10 Comparability

Status: Preferred

Definition: The degree to which results can be meaningfully compared.

23.11 Equivalence

Status: Preferred

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

23.12 Crosswalk

Status: Preferred

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

23.13 Concordance

Status: Preferred

Definition: A mapping showing relationships among multiple vocabularies or classification systems.

23.14 Bridge Study

Status: Preferred

Definition: An analysis connecting results across protocols, versions, languages, forms, or systems.

23.15 Common Core

Status: Preferred

Definition: The stable shared elements used across multiple implementations or jurisdictions.

23.16 Local Extension

Status: Preferred

Definition: A documented addition or adaptation specific to a domain, language, jurisdiction, or institution.

23.17 Localization

Status: Preferred

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

23.18 Translation Validity

Status: Preferred

Definition: The degree to which translation preserves the intended construct and interpretation.

23.19 Federated Evaluation

Status: Preferred

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

23.20 Evidence Portability

Status: Preferred

Definition: The ability to reuse and interpret evidence across organizations, schemes, or jurisdictions.

23.21 Recognition Drift

Status: Preferred

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

23.22 Interoperability Debt

Status: Preferred

Definition: Accumulated cost created by incompatible terminology, formats, identifiers, protocols, and institutional arrangements.

24. Registry and Identity Terms

24.1 Registry

Status: Preferred

Definition: A maintained authoritative or verified record of systems, protocols, evaluators, certificates, incidents, mappings, or recognition status.

24.2 Register

Status: Accepted

Definition: A maintained list or record.

Usage note: Use *registry* for structured institutional systems.

24.3 Record Identifier

Status: Preferred

Definition: A unique value used to identify a record.

24.4 Persistent Identifier

Status: Preferred

Definition: An identifier intended to remain stable over time.

24.5 Model Identifier

Status: Preferred

Definition: A unique identifier for a model artifact or release.

24.6 System Identifier

Status: Preferred

Definition: A unique identifier for a defined AI system configuration.

24.7 Protocol Identifier

Status: Preferred

Definition: A unique identifier for an evaluation protocol and version.

24.8 Evaluator Identifier

Status: Preferred

Definition: A unique identifier for an evaluator organization or qualified unit.

24.9 Evidence Package

Status: Preferred

Definition: A structured set of artifacts supporting a claim, review, certification, recognition, or decision.

24.10 Manifest

Status: Preferred

Definition: A structured record identifying components, versions, configuration, provenance, and integrity information.

24.11 System Manifest

Status: Preferred

Definition: A manifest identifying the material components and configuration of an AI system.

24.12 Signed Record

Status: Preferred

Definition: A record protected by a digital signature or equivalent integrity mechanism.

24.13 Status Record

Status: Preferred

Definition: A record indicating whether an item is current, expired, suspended, withdrawn, corrected, or superseded.

25. Versioning and Document-Control Terms

25.1 Version

Status: Preferred

Definition: An identified state of a document, protocol, system, model, standard, or schema.

25.2 Major Version

Status: Preferred

Definition: A version containing material changes that may break compatibility, alter meaning, change decisions, or require re-evaluation.

25.3 Minor Version

Status: Preferred

Definition: A version containing substantive but backward-compatible improvement.

25.4 Patch Version

Status: Preferred

Definition: A version containing corrections or clarifications without intended substantive change.

25.5 Revision

Status: Preferred

Definition: A documented change to a document, protocol, standard, or system.

25.6 Amendment

Status: Preferred

Definition: A formally approved change to an existing document or rule.

25.7 Change Log

Status: Preferred

Definition: A record of changes across versions.

25.8 Revision Record

Status: Preferred

Definition: A formal summary of document version, date, change type, and rationale.

25.9 Current

Status: Preferred status

Definition: The active and applicable version or record.

25.10 Draft

Status: Preferred status

Definition: Work not yet approved for canonical use.

25.11 Proposed

Status: Preferred status

Definition: Submitted for review or adoption but not yet approved.

25.12 Approved

Status: Preferred status

Definition: Formally accepted through the applicable process.

25.13 Active

Status: Preferred status

Definition: In current operational use.

25.14 Expired

Status: Preferred status

Definition: No longer valid because its defined validity period ended.

25.15 Suspended

Status: Preferred status

Definition: Temporarily inactive or invalid pending review, correction, or resolution.

25.16 Withdrawn

Status: Preferred status

Definition: Formally removed from current use or recognition.

25.17 Superseded

Status: Preferred status

Definition: Replaced by a newer approved version.

25.18 Deprecated

Status: Preferred status

Definition: Still available for transition or historical use but no longer preferred.

25.19 Retired

Status: Preferred status

Definition: Removed from active use after planned discontinuation.

25.20 Archived

Status: Preferred status

Definition: Preserved for historical or evidentiary purposes but not active.

25.21 Corrected

Status: Preferred status

Definition: Modified to address error while preserving a visible correction record.

25.22 Reissued

Status: Preferred status

Definition: Issued again after correction or formal update.

25.23 Effective Date

Status: Preferred

Definition: The date on which a version, decision, or requirement becomes applicable.

25.24 Expiration Date

Status: Preferred

Definition: The date after which a version, result, certificate, or recognition is no longer valid unless renewed.

25.25 Review Date

Status: Preferred

Definition: The scheduled date for reassessment.

26. Status of Evaluation Results

26.1 Valid Result

Status: Context-specific

Definition: A result that meets the defined validity requirements for its stated purpose.

Usage note: Valid does not mean universally true.

26.2 Current Result

Status: Preferred

Definition: A result still applicable to the identified system, protocol, configuration, and period.

26.3 Expired Result

Status: Preferred

Definition: A result whose validity period has ended.

26.4 Superseded Result

Status: Preferred

Definition: A result replaced by a newer evaluation or interpretation.

26.5 Corrected Result

Status: Preferred

Definition: A result modified after error identification, with the original and correction history preserved.

26.6 Withdrawn Result

Status: Preferred

Definition: A result formally removed because of invalidity, serious error, conflict, compromise, or other material reason.

26.7 Inconclusive

Status: Preferred

Definition: Evidence is insufficient to support a defined positive or negative conclusion.

26.8 Indeterminate

Status: Preferred

Definition: A determination cannot be made within the available evidence, scope, or method.

26.9 Not Demonstrated

Status: Preferred

Definition: The evaluation did not produce sufficient evidence that the claimed capability or condition was present under the tested conditions.

Distinction: Not demonstrated is not the same as absent.

26.10 Demonstrated With Qualifications

Status: Preferred

Definition: Evidence supports the claim subject to stated limitations or conditions.

26.11 Partially Substantiated

Status: Preferred

Definition: Evidence supports only part of the claim.

26.12 Substantiated

Status: Preferred

Definition: Available evidence sufficiently supports the claim within the stated scope.

26.13 Not Substantiated

Status: Preferred

Definition: Available evidence does not sufficiently support the claim.

27. Incentive and Recognition Terms

27.1 Incentive

Status: Preferred

Definition: A condition that changes the expected benefit, cost, status, opportunity, or consequence associated with behavior.

27.2 Intrinsic Motivation

Status: Preferred

Definition: Motivation arising from interest, purpose, mastery, identity, or satisfaction inherent in the activity.

27.3 Extrinsic Motivation

Status: Preferred

Definition: Motivation arising from external reward, sanction, status, access, or requirement.

27.4 Prestige

Status: Preferred

Definition: Durable esteem or status granted by a relevant community or institution.

27.5 Recognition

Status: Preferred, see Section 15

Additional meaning: Formal or informal acknowledgment of contribution, competence, or achievement.

27.6 Award

Status: Preferred

Definition: Formal recognition granted under stated criteria.

27.7 Prize

Status: Preferred

Definition: A reward offered for achieving a defined objective.

27.8 Bounty

Status: Preferred

Definition: A reward offered for identifying a specified vulnerability, failure, flaw, or contribution.

27.9 Grant

Status: Preferred

Definition: Funding provided to support research, infrastructure, training, public-interest work, or capacity.

27.10 Contribution Credit

Status: Preferred

Definition: Attribution of value to people or organizations responsible for a contribution.

27.11 Corrective Credit

Status: Preferred

Definition: Recognition for timely disclosure, correction, remediation, or withdrawal of invalid work.

27.12 Prestige Capture

Status: Preferred

Definition: Control of recognition systems by actors able to convert existing status into further authority or reward.

27.13 Recognition Inflation

Status: Preferred

Definition: Decline in the meaning of a designation as awards, badges, certificates, or titles proliferate.

27.14 Goodhart Effect

Status: Preferred

Definition: Degradation of a measure's value when it becomes a target for optimization.

27.15 Reward Hacking

Status: Preferred

Definition: Maximizing the measured reward while avoiding or undermining the intended goal.

27.16 Crowding Out

Status: Preferred

Definition: Reduction of intrinsic or prosocial motivation after external rewards or controls are introduced.

27.17 Public Good

Status: Preferred

Definition: A good whose benefits are broadly shared and difficult to restrict to paying contributors.

27.18 Free Rider

Status: Preferred

Definition: An actor that benefits from shared infrastructure or risk reduction without contributing proportionately.

28. Standards Body Public-Claims Vocabulary

28.1 Preferred Phrases

Use:

- Independent research and institutional-design project
- Developing foundations for frontier AI
- Canonical working white paper
- Proposed framework

- Draft protocol
- Pilot specification
- Independent review
- Qualified evaluator
- Evidence under specified conditions
- Met the defined requirement
- Not demonstrated under the tested conditions
- Current project position
- Subject to revision
- Does not constitute certification or regulatory approval

28.2 Discouraged Phrases

Avoid unless precisely justified:

- Official
- approved AI
- safe model
- global authority
- universal standard
- industry consensus
- experts agree
- certified secure
- government-grade
- validated safe
- trusted AI
- fully compliant
- guaranteed
- proven harmless
- risk-free
- failsafe
- unbiased

28.3 Prohibited Phrases for Present Standards Body Status

Do not use:

- Official AI Standards Body
- Frontier AI regulator
- AI accreditation authority
- AI certification authority
- Government-approved framework
- Standards Body approved model
- Standards Body certified safe
- Legally compliant according to Standards Body
- Internationally recognized standard, unless formal recognition exists

- Licensed AI evaluator, unless a lawful licensing regime applies

29. Critical Distinction Table

Terms	Canonical distinction
Model vs system	A model is a learned computational component. A system includes configuration, tools, interfaces, safeguards, infrastructure, and operational context.
Capability vs risk	Capability is ability. Risk combines likelihood, consequence, exposure, actor, access, and safeguards.
Safety vs compliance	Safety concerns acceptable risk. Compliance concerns fulfillment of applicable obligations.
Test vs evaluation	A test is a procedure. Evaluation includes design, administration, scoring, interpretation, and limitations.
Benchmark vs protocol	A benchmark is a standardized comparison instrument. A protocol is the complete specification governing an evaluation.
Public benchmark vs held-out evaluation	A public benchmark is visible. A held-out evaluation protects material to preserve evidentiary value.
Review vs audit	Review examines evidence and reasoning. Audit is systematic, independent, documented, and criteria-based.
External vs independent	External means outside the organization. Independent means sufficiently free from controlling influence.
Assurance vs certification	Assurance is evidence-supported confidence. Certification is formal third-party attestation under a defined scheme.
Certification vs accreditation	Certification concerns conformity of a product, process, service, person, or management system. Accreditation recognizes the competence of the body performing conformity assessment.
Conformity vs compliance	Conformity is fulfillment of specified requirements. Compliance usually concerns binding legal, regulatory, or contractual obligations.
Standard vs requirement	A standard is a document or agreed framework. A requirement is a condition that must be fulfilled.
Voluntary vs nonbinding	Voluntary concerns adoption. A voluntarily adopted contract or scheme can create binding obligations.
Recognition vs approval	Recognition accepts evidence, competence, process, or status for a defined purpose. Approval is formal acceptance by an authorized body.
Validation vs verification	Validation asks whether something is suitable for intended use. Verification asks whether specified requirements were fulfilled.
Harm vs hazard	Harm is adverse effect. A hazard is a source or condition with potential to cause harm.
Safeguard vs guarantee	A safeguard reduces risk. It does not guarantee absence of harm.
Not demonstrated vs absent	Not demonstrated means evidence was insufficient under tested conditions. Absent is a stronger claim requiring evidence of absence.

Terms	Canonical distinction
Confidence vs certainty	Confidence expresses justified belief. Certainty implies no meaningful uncertainty.
Consensus vs unanimity	Consensus is broad agreement after addressing substantial objections. Unanimity requires agreement by all.
Transparency vs full disclosure	Transparency provides sufficient visibility. Full disclosure may be unsafe, unlawful, or unnecessary.
Interoperability vs uniformity	Interoperability supports exchange and use across differences. Uniformity requires sameness.
Equivalence vs comparability	Comparability allows meaningful comparison. Equivalence supports acceptance as sufficiently similar for a stated purpose.

30. Capitalization Rules

30.1 Proper Names

Capitalize:

- Standards Body
- Foundations for Frontier AI
- Names of formal organizations
- Formal document titles
- Formal program names
- Defined committee names

30.2 Generic Terms

Use lowercase for generic references:

- standard
- evaluation
- certification body
- accreditation body
- regulator
- protocol
- working group

30.3 Foundation Titles

Use title capitalization in headings:

- Foundation 1: Dynamic Evaluation Protocols

Use lowercase in ordinary generic references:

- the dynamic evaluation protocols foundation

30.4 Artificial Intelligence

Use *AI* after first use where clear.

30.5 Frontier AI

Capitalize only at the start of a sentence or within a proper title.

30.6 Standards Body Files

Use exact uppercase filenames in code formatting:

- PROJECT_IDENTITY.md
 - TERMINOLOGY.md
 - FOUNDATION_01_DYNAMIC_EVALUATION_PROTOCOLS.md
-

31. Abbreviation Rules

31.1 First Use

Write the full term followed by the abbreviation in parentheses.

Example:

conformity-assessment body (CAB)

31.2 Avoid Unnecessary Abbreviations

Do not abbreviate a term used only a few times.

31.3 Approved Common Abbreviations

- AI, artificial intelligence
- API, application programming interface
- CAB, conformity-assessment body
- COI, conflict of interest, only where unambiguous
- RAG, retrieval-augmented generation
- TEVV, test, evaluation, validation, and verification

31.4 Ambiguous Abbreviations

Avoid abbreviations that can refer to multiple concepts without explicit definition.

31.5 File and Protocol Identifiers

Identifiers may use standardized abbreviations if defined in the applicable registry or schema.

32. Hyphenation and Style Rules

32.1 Preferred Forms

Use:

- third-party
- first-party
- second-party
- high-stakes
- open-weight
- public-interest
- risk-based
- capability-based
- performance-based
- decision-relevant
- model-level
- system-level
- cross-border
- cross-institutional
- machine-readable
- human-readable
- version-specific
- scope-specific
- evidence-based

32.2 Artificial Intelligence Terms

Use:

- AI system
- AI model
- AI evaluation
- AI governance

Do not hyphenate *AI* as a prefix unless part of a longer compound requiring clarity.

32.3 Em Dashes

Standards Body canonical documents should not use em dashes.

Use:

- commas
- semicolons
- parentheses
- colons
- separate sentences

32.4 Slashes

Use slashes sparingly.

Prefer explicit conjunctions when meaning matters.

32.5 Quotation Marks

Use quotation marks for:

- Direct quotations
- terms under analysis
- exact public claims

Do not use quotation marks to signal vague skepticism.

33. Introducing New Terms

A new canonical term should be introduced only when:

- Existing terms are inadequate
- The concept is materially distinct
- The term improves precision
- A definition can be written
- Related terms can be mapped
- The term can be translated or localized where relevant
- The term will be used beyond one isolated sentence

33.1 Required Proposal Fields

- Proposed term
- definition
- purpose
- scope

- examples
- related terms
- distinctions
- source or rationale
- status
- owner
- version

33.2 Review

New terms should be reviewed by relevant technical and institutional experts.

33.3 Temporary Terms

Exploratory papers may use temporary terms if marked as provisional.

33.4 Avoid Coining for Prestige

Do not create new terminology merely to brand ordinary concepts as proprietary.

34. Revising Terms

A term should be revised when:

- Evidence changes
- usage becomes misleading
- legal meaning changes
- international mapping reveals ambiguity
- scope expands or narrows
- implementation shows conflict
- a superior term becomes established

34.1 Material Revision

A material change to meaning requires:

- New version
- change record
- affected-document review
- compatibility note
- transition guidance

34.2 Nonmaterial Revision

Spelling, grammar, or clarification without meaning change may be treated as a patch.

34.3 Deprecation

Deprecated terms should identify:

- Replacement term
- reason
- transition period
- historical mapping

34.4 Retirement

Retired terms remain discoverable for historical interpretation.

35. Translation and Localization Rules

35.1 Concept Before Word

Translate the concept, not only the literal word.

35.2 Preserve Institutional Distinctions

Terms such as:

- audit
- certification
- accreditation
- regulator
- standard
- compliance

may have jurisdiction-specific legal meanings.

35.3 Translation Record

A translated controlled vocabulary should record:

- Source language
- target language
- translator
- domain reviewer
- date

- version
- non-equivalent terms
- local notes

35.4 No Forced Equivalence

Where no exact equivalent exists, preserve the source term and explain the distinction.

35.5 Legal Review

Legal terms require qualified jurisdiction-specific review.

36. Terminology Governance

36.1 Owner

The canonical owner is Standards Body.

36.2 Stewardship Function

A future terminology steward or committee should:

- Maintain definitions
- review proposals
- map external vocabularies
- resolve conflicts
- manage translations
- publish revisions
- track deprecated terms

36.3 Contributor Input

Contributors may propose changes through a documented process.

36.4 Decision Standard

Terminology decisions should consider:

- Precision
- usability
- technical validity
- institutional meaning
- legal implications

- international mapping
- public interpretation
- existing practice

36.5 Dissent

Material unresolved disagreement should be recorded.

36.6 Emergency Correction

A misleading or harmful definition may be corrected through an expedited process followed by full review.

37. Terminology Quality Tests

Before approving a term, ask:

37.1 Definition Test

Can the term be defined without circular language?

37.2 Distinction Test

Is it materially different from related terms?

37.3 Evidence Test

Does the term imply more evidence than is usually available?

37.4 Authority Test

Does it imply legal or institutional authority?

37.5 Scope Test

Can the term be used with explicit scope?

37.6 Public Interpretation Test

Would a reasonable reader misunderstand it?

37.7 International Test

Can it be mapped across jurisdictions and languages?

37.8 Operational Test

Can the term support a decision, record, protocol, or standard?

37.9 Stability Test

Is the term likely to remain useful as the field changes?

37.10 Anti-Branding Test

Is the term needed for precision rather than prestige?

38. Deprecated and Discouraged Vocabulary Register

Term	Status	Preferred replacement or instruction
Safe AI	Discouraged	State the assessed requirement, risk level, or safeguard performance
Certified safe	Prohibited	Certified against named requirements within scope
AI approval	Discouraged	State the approving authority and exact decision
Official framework	Discouraged	Approved by named body, or published framework
Secret test	Discouraged	Held-out evaluation
AI watchdog	Discouraged for Standards Body	Independent research or oversight organization, as applicable
AI audit, without criteria	Discouraged	Name the audit criteria and scope
Accredited model	Prohibited	Models may be certified or assessed; bodies are accredited
Self-accredited	Prohibited	Internally qualified or independently reviewed
Validated safe	Prohibited	Validated for a defined intended use, or met defined safeguard criteria
Fully compliant	Discouraged	Compliant with named requirements as of a defined date
Unbiased model	Prohibited as an absolute claim	Evaluated for defined bias or disparity measures
Failsafe	Prohibited as an absolute claim	Fault-tolerant or protected by defined controls
Risk-free	Prohibited	Residual risk assessed as low under stated conditions
Proven harmless	Prohibited	No material harm observed under assessed conditions

Term	Status	Preferred replacement or instruction
Global consensus	Discouraged	State participating institutions and degree of agreement
Experts agree	Discouraged	Identify expert population, evidence, and dissent
Best practice	Discouraged	Recommended practice or current leading practice
Kill switch	Discouraged	Emergency shutdown, access revocation, or deployment suspension
Black box, without definition	Discouraged	State which information or access is unavailable
Hallucination, for every factual error	Discouraged	Unsupported output, fabrication, factual error, or confabulation, as applicable
Alignment, without object	Discouraged	State alignment with whose objective, values, policy, or intent
Responsible AI, without criteria	Discouraged	State the practices, framework, or responsibility at issue
Trustworthy AI, without evidence	Discouraged	State the assessed reliability, safety, security, transparency, or governance properties

39. High-Risk Ambiguity Register

The following terms require special care because they are frequently used to imply more than they establish.

39.1 Alignment

Always state:

- Aligned with what objective?
- Whose intent or values?
- Under which conditions?
- How measured?

39.2 Safety

Always state:

- Safe from which harm?
- For which user and context?
- Under which assumptions?
- For what period?

39.3 Trustworthy

Always state the properties supporting trust.

39.4 Robust

Always state the perturbation, threat, shift, or condition.

39.5 Secure

Always state the threat model and scope.

39.6 Independent

Always state relevant relationships, access, funding, and authority.

39.7 Validated

Always state:

- What was validated?
- For which intended use?
- By whom?
- Under which method?

39.8 Audited

Always state the criteria, scope, period, and auditor.

39.9 Certified

Always state the scheme, requirements, issuer, subject, scope, and validity.

39.10 Accredited

Always state the accreditation body and scope.

39.11 Compliant

Always state the applicable requirement and jurisdiction.

39.12 Frontier

Always state the relevant capability, time, or comparison basis where material.

40. Canonical Usage Examples

40.1 Capability Result

Preferred:

The system demonstrated the defined cyber capability under Protocol CYBER-AGENT 1.2 with tool access, five attempts per task, and a 90-minute task limit.

Avoid:

The model is cyber-capable.

40.2 Negative Result

Preferred:

The evaluation did not demonstrate the capability under the tested conditions.

Avoid:

The model cannot perform the task.

40.3 Safeguard Result

Preferred:

The safeguard reduced attack success under the stated threat model, but adaptive attacks remained outside scope.

Avoid:

The safeguard makes the model safe.

40.4 Independent Review

Preferred:

The report received independent external review under a documented mandate with access to raw evaluation records.

Avoid:

Experts verified the report.

40.5 Audit Claim

Preferred:

The organization underwent an independent audit against the specified control framework for the period stated.

Avoid:

The company was audited and approved.

40.6 Certification Claim

Preferred:

The management system was certified against the named standard by the identified certification body within the published scope.

Avoid:

The AI was certified safe.

40.7 Accreditation Claim

Preferred:

The evaluator is accredited for the listed testing scope by the named accreditation body.

Avoid:

The evaluator is officially accredited for AI.

40.8 Standards Body Identity

Preferred:

Standards Body is an independent research and institutional-design project.

Avoid:

Standards Body is the official authority for frontier AI.

40.9 International Recognition

Preferred:

The result is recognized for the defined technical purpose under the named arrangement.

Avoid:

The result is globally approved.

40.10 Compliance

Preferred:

The organization reported conformity with the named voluntary standard and compliance with the stated legal requirement as of the review date.

Avoid:

The organization is fully compliant.

41. Alphabetical Quick Reference

This quick reference lists preferred terms and the sections containing full definitions.

A

- Accreditation, Section 15
- Accreditation body, Section 15
- Accountability, Section 18
- Agent, Section 5
- Agentic system, Section 5
- AI model, Section 5
- AI system, Section 5
- Appeal, Section 18
- Approval, Section 15
- Assessment, Section 14
- Assurance, Section 14
- Attestation, Section 14
- Audit, Section 14
- Authority, Section 4
- Autonomy, Section 5

B

- Baseline, Section 10
- Benchmark, Section 9
- Benchmark gaming, Section 9
- Benchmark saturation, Section 9
- Binding requirement, Section 16
- Bounty, Section 27

- Bridge study, Section 23
- Burden of proof, Section 12

C

- CAB, Section 14
- Calibration, Section 10
- Canonical source, Section 4
- Capability, Section 6
- Capability domain, Section 6
- Capability profile, Section 6
- Capability threshold, Section 6
- Certificate, Section 15
- Certification, Section 15
- Certification body, Section 15
- Certification scheme, Section 15
- Chain of custody, Section 12
- Claim, Section 12
- Code of conduct, Section 16
- Code of practice, Section 16
- Committee, Section 18
- Comparability, Section 23
- Compatibility, Section 23
- Compliance, Section 17
- Confidence, Section 10
- Conformity, Section 17
- Conformity assessment, Section 14
- Consensus, Section 18
- Construct, Section 10
- Construct validity, Section 10
- Contamination, Section 9
- Control, Section 8
- Corrective action, Section 22
- Corrective credit, Section 27
- Critical capability, Section 6
- Crosswalk, Section 23
- Current, Section 25

D

- Dangerous capability, Section 6
- Decision right, Section 18
- Defense in depth, Section 8
- Demonstrated capability, Section 6
- Deployment, Section 5

- Deployment context, Section 5
- Deprecated, Section 25
- Direct evidence, Section 12
- Disclosure, Section 19
- Dissent, Section 13
- Distribution shift, Section 10
- Dynamic evaluation protocol, Section 9

E

- Effective capability, Section 6
- Elicitation, Section 11
- Elicitation budget, Section 11
- Epistemic status, Section 12
- Equivalence, Section 23
- Evaluation, Section 9
- Evaluation awareness, Section 9
- Evaluation harness, Section 9
- Evaluation incident, Section 22
- Evaluation integrity, Section 9
- Evaluation protocol, Section 9
- Evaluator, Section 14
- Evidence, Section 12
- Evidence package, Section 24
- Evidence standard, Section 12
- External review, Section 13

F

- Failure, Section 22
- Failure mode, Section 22
- False negative, Section 10
- False positive, Section 10
- Fine-tuning, Section 5
- First-party assessment, Section 14
- Foundation model, Section 5
- Framework, Section 16
- Frontier AI, Section 5
- Frontier AI system, Section 5
- Frontier model, Section 5

G

- General-purpose AI model, Section 5
- Good practice, Section 16

- Goodhart effect, Section 27
- Governance, Section 18
- Governance framework, Section 18
- Grant, Section 27
- Guidance, Section 16

H

- Harm, Section 7
- Harmonization, Section 23
- Hazard, Section 7
- Held-out evaluation, Section 9
- High-stakes capability, Section 6
- High-stakes risk, Section 7
- Human baseline, Section 10
- Human oversight, Section 8

I

- Impact, Section 7
- Impartiality, Section 19
- Incident, Section 22
- Incident response, Section 22
- Independence, Section 19
- Independent expert review, Section 13
- Inference, Section 12
- Information hazard, Section 20
- Inspection, Section 14
- Institutional capture, Section 18
- Institutional design, Section 4
- Institutional legitimacy, Section 18
- Interoperability, Section 23

J

- Jurisdiction, Section 4

K

- Knowledge gap, Section 12

L

- Latent capability, Section 6

- Legal interoperability, Section 23
- Likelihood, Section 7
- Limited assurance, Section 14
- Localization, Section 23

M

- Management-system standard, Section 16
- Mandate, Section 4
- Manifest, Section 24
- Material conflict, Section 19
- Measurement uncertainty, Section 10
- Metric, Section 10
- Minority report, Section 13
- Misuse, Section 7
- Model family, Section 5
- Model identifier, Section 24
- Model security, Section 21
- Model version, Section 5
- Monitoring, Section 8
- Mutual recognition, Section 15

N

- Near miss, Section 22
- Negative result, Section 12
- Noncompliance, Section 17
- Nonconformity, Section 17
- Not demonstrated, Section 26

O

- Objectivity, Section 19
- Open question, Section 12
- Open-source AI, Section 5
- Open-weight model, Section 5
- Oversight, Section 18

P

- Peer review, Section 13
- Performance standard, Section 16
- Phase-in, Section 16
- Post-training, Section 5
- Prestige, Section 27

- Primary source, Section 12
- Progressive requirement, Section 16
- Project, Section 4
- Proposed standard, Section 16
- Protocol identifier, Section 24
- Public claim, Section 4
- Public-interest review, Section 13

Q

- Quorum, Section 18

R

- Reasonable assurance, Section 14
- Recognition, Section 15
- Recommended practice, Section 16
- Registry, Section 24
- Reliability, Section 6 and Section 10
- Remediation, Section 22
- Replication, Section 9
- Reproducibility, Section 12
- Requirement, Section 16
- Residual risk, Section 7
- Responsible disclosure, Section 20
- Review, Section 13
- Review mandate, Section 13
- Risk, Section 7
- Risk threshold, Section 7
- Robustness, Section 6

S

- Safe harbor, Section 16
- Safeguard, Section 8
- Safeguard effectiveness, Section 8
- Safety, Section 7
- Sandbagging, Section 9
- Sandbox, Section 21
- Score, Section 10
- Scoring rule, Section 10
- Second-party assessment, Section 14
- Security, Section 7
- Security incident, Section 21
- Severity, Section 7

- Standard, Section 16
- Standards Body, Section 4
- Substantiated, Section 26
- Sunset clause, Section 16
- System identifier, Section 24
- System manifest, Section 24
- System prompt, Section 5
- System version, Section 5
- Systemic risk, Section 7

T

- Task, Section 9
- Task family, Section 9
- Task horizon, Section 6
- Technical review, Section 13
- Technical specification, Section 16
- Test, Section 9
- Testing, Section 9
- Third-party assessment, Section 14
- Third-party evaluator, Section 14
- Threat, Section 7
- Threat model, Section 7
- Threshold, Section 10
- Traceability, Section 12
- Transparency, Section 20
- Triangulation, Section 12

U

- Uncertainty, Section 10
- Unilateral recognition, Section 15
- Unanimity, Section 18

V

- Validation, Section 9
- Verification, Section 9
- Version, Section 25
- Vulnerability, Section 7

W

- Weight of evidence, Section 12
- Weight security, Section 21

- Working group, Section 18
 - Working paper, Section 4
-

42. Terminology Interfaces With Other Canonical Files

42.1 PROJECT_IDENTITY.md

Governs project identity, public role, authority, and approved descriptions.

42.2 EVIDENCE_STANDARDS.md

Will govern evidentiary sufficiency, source quality, confidence, citation, and claims.

42.3 RESEARCH_METHODODOLOGY.md

Will govern how research questions, evidence, expert input, and correction are handled.

42.4 TAXONOMY.md

Will organize terms into formal categories and hierarchical relationships.

42.5 EVALUATION_PHILOSOPHY.md

Will govern Standards Body's conceptual approach to measurement, capability, risk, thresholds, and uncertainty.

42.6 GOVERNANCE_FRAMEWORK.md

Will govern formal roles, decision rights, committees, conflicts, appeals, and accountability.

42.7 STANDARDS_DEVELOPMENT_PROCESS.md

Will govern the use of *draft*, *proposed*, *approved*, *standard*, *revision*, and *retired* within standards work.

42.8 TRANSPARENCY_FRAMEWORK.md

Will govern disclosure levels, confidentiality, redaction, public summaries, and restricted evidence.

42.9 EVALUATOR_ACCREDITATION_FRAMEWORK.md

Will govern accreditation terminology within a formal evaluator-recognition system.

42.10 WEBSITE_SOURCE_OF_TRUTH.md

Will apply this terminology to exact public website language.

43. Canonical Terminology Commitments

Standards Body commits to:

1. Use evaluation, review, audit, certification, and accreditation as distinct terms.
 2. Use capability and risk as distinct concepts.
 3. Use model and system as distinct concepts.
 4. Avoid absolute safety claims.
 5. State the scope of compliance claims.
 6. State the scheme and scope of certification claims.
 7. State the accreditation body and scope of accreditation claims.
 8. Avoid calling external review independent without supporting conditions.
 9. Distinguish not demonstrated from absent.
 10. Distinguish evidence recognition from legal approval.
 11. Carry uncertainty with evaluation results.
 12. Identify system and protocol versions.
 13. Use current status labels.
 14. Preserve deprecated and retired terms for historical interpretation.
 15. Avoid terminology created mainly for institutional prestige.
 16. Correct misleading terms when discovered.
 17. Document material changes in meaning.
 18. Support multilingual and jurisdiction-specific mappings.
 19. Keep public language consistent with present authority.
 20. Treat terminology as operational infrastructure rather than editorial decoration.
-

44. Final Terminology Position

Frontier AI standards cannot be credible when the language supporting them is unstable.

A project cannot claim to build trustworthy evaluation while using:

- audit as a synonym for review
- certification as a synonym for approval
- accreditation as a synonym for reputation
- safety as a synonym for passing a test
- capability as a synonym for risk
- external as a synonym for independent
- compliance as a synonym for good practice
- international as a synonym for globally recognized
- not demonstrated as a synonym for impossible

These distinctions are not semantic formalities.

They determine what institutions believe, what the public understands, what purchasers rely upon, what regulators enforce, and what developers can claim.

Standards Body therefore treats terminology as foundational infrastructure.

The vocabulary should make every important claim answerable.

What was evaluated?

Which system?

Under which protocol?

By whom?

With what access?

Against which requirement?

Under which authority?

With what uncertainty?

For how long?

Within which jurisdiction?

The purpose of this file is not to make the project sound more technical.

It is to make the project harder to misunderstand.

The canonical rule is:

Use the term that matches the function, the evidence, the scope, and the authority actually present.

Revision Record

Version 1.0

Date: July 16, 2026

Change type: Complete foundational edition

Summary: Establishes the canonical Standards Body controlled vocabulary. Defines project, AI system, capability, risk, safeguard, evaluation, measurement, elicitation, evidence, review, audit, assurance, certification, accreditation, standards, requirements, compliance, governance, independence, transparency, security, incidents, interoperability, registries, versioning, result

status, incentives, public claims, usage distinctions, capitalization, abbreviations, hyphenation, term introduction, revision, translation, governance, quality tests, discouraged vocabulary, ambiguity controls, examples, alphabetical reference, file interfaces, and terminology commitments.

Status: Approved foundational source.