

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

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

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.

# RESEARCH\_METHODODOLOGY.md

## Standards Body Research Methodology

**Project:** Standards Body

**Primary domain:** standardsbody.ai

**Core line:** Foundations for Frontier AI

**Document type:** Canonical research planning, execution, review, publication, and correction methodology

**Version:** 1.0

**Status:** Approved foundational source

**Document owner:** Standards Body

**Applies to:** All Standards Body research programs, foundation papers, technical evaluations, literature reviews, case studies, institutional analyses, standards proposals, legal and policy research, forecasts, public reports, contributor projects, pilots, and future assurance research

**Related canonical sources:** PROJECT\_IDENTITY.md, TERMINOLOGY.md, FOUNDATIONS\_APPENDIX.md, EVIDENCE\_STANDARDS.md, TAXONOMY.md, EVALUATION\_PHILOSOPHY.md, GOVERNANCE\_FRAMEWORK.md, TRANSPARENCY\_FRAMEWORK.md, VERSION\_HISTORY.md

**Research basis reviewed through:** July 16, 2026

**Review cycle:** Annual review, with event-triggered revision following material changes in frontier AI evaluation science, research integrity, reproducibility practice, human-subjects regulation, open-science practice, research security, or Standards Body institutional status

---

## Authority Note

This document defines the internal research methodology of Standards Body.

It does not:

- Replace institutional review board review
- determine whether a project is legally regulated human-subjects research
- provide legal advice
- authorize access to restricted systems or information
- establish a government research policy
- guarantee scientific validity
- certify that a study is ethical

- replace domain-specific laboratory, security, clinical, or professional requirements
- create accreditation, certification, or regulatory authority

Where applicable law, regulation, contract, ethics review, security classification, or domain-specific standard imposes a stronger requirement, the stronger requirement governs.

---

## Document Purpose

This document establishes how Standards Body conducts research.

It is the authoritative source for:

- Research-question formation
- project classification
- protocol development
- prospective planning
- literature and evidence review
- technical evaluation research
- institutional and governance research
- case-study methodology
- legal, standards, and regulatory research
- comparative and international research
- forecasting and expert judgment
- data and artifact management
- reproducibility and replication
- human-participant protection
- research security and dual-use review
- conflict management
- contributor roles and authorship
- use of artificial intelligence tools in research
- peer, independent, and adversarial review
- uncertainty, confidence, and evidence synthesis
- publication, transparency, confidentiality, correction, and retirement
- quality assurance and methodological audit

The methodology is designed for an environment in which:

- AI systems change quickly
- evaluation methods are immature
- public benchmarks become contaminated
- access to frontier systems is uneven
- evidence may be confidential
- technical results can have policy consequences
- severe claims can be difficult to test safely
- institutional roles are still developing
- research may affect standards, procurement, public perception, or regulation

- uncertainty is substantial
- premature certainty creates harm

The purpose is not to impose one method on every project.

The purpose is to require that each project select, justify, document, execute, review, and revise a method appropriate to its question.

---

## Executive Summary

Standards Body is a research and institutional-design project.

Its credibility depends on the quality of its research process, not only the quality of its conclusions.

The central methodological proposition is:

**Research should begin with a decision-relevant question, proceed through a documented and proportionate method, preserve the distinction between planned and exploratory work, expose material uncertainty and disagreement, protect sensitive information, and produce outputs that can be reviewed, reproduced where appropriate, corrected, and retired.**

Frontier AI research creates unusual methodological challenges.

A study may evaluate a model that changes before publication.

A capability result may depend more on scaffolding, tool access, test-time compute, or evaluator expertise than on the base model name.

A public task set may no longer measure general capability because the model was trained on similar material.

A controlled experiment may be impossible because the relevant event is rare, dangerous, proprietary, or international.

A qualitative interview may provide better evidence about institutional practice than a public policy document.

A confidential system log may be highly direct evidence but impossible for the public to inspect.

A technical result may inform a policy decision without determining the correct policy.

A research institution may be pressured to publish before the evidence is mature or to withhold findings that are commercially or politically inconvenient.

Standards Body addresses these challenges through a structured research lifecycle.

# The Research Lifecycle

Every substantial project should ordinarily pass through:

1. **Initiation** - Define the problem, decision, claim, audience, and consequence of error.
2. **Classification** - Identify the project type, risk level, evidence needs, security needs, and review requirements.
3. **Protocol** - Specify the research question, methods, sources, data, analysis, uncertainty, ethics, security, and change rules.
4. **Registration** - Preserve a time-stamped record of the planned work where useful and safe.
5. **Evidence Collection** - Gather primary, secondary, technical, organizational, legal, qualitative, quantitative, and contrary evidence.
6. **Execution** - Conduct the study according to the protocol while recording deviations, failures, and exploratory work.
7. **Analysis** - Apply justified methods, test assumptions, examine uncertainty, and separate observation from inference.
8. **Challenge** - Use peer review, domain review, methodological review, independent review, adversarial review, or replication as required.
9. **Synthesis** - Integrate evidence, preserve disagreement, assign evidence level and confidence, and identify remaining gaps.
10. **Publication**
  - State methods, findings, limitations, provenance, conflicts, AI-tool use, and status.
11. **Monitoring**
  - Track new evidence, model changes, incidents, legal changes, and replication.
12. **Correction or Retirement**
  - Correct, supersede, withdraw, or retire findings when the evidence no longer supports them.

The methodology distinguishes several research modes.

## Descriptive Research

Documents what exists, happened, or changed.

## Exploratory Research

Maps an immature topic, identifies hypotheses, and locates gaps.

## **Confirmatory Research**

Tests prospectively defined hypotheses or decision rules.

## **Evaluation Research**

Measures models, systems, safeguards, protocols, or institutions.

## **Case-Study Research**

Examines a bounded event, organization, process, or failure in context.

## **Institutional-Design Research**

Develops and tests governance, standards, assurance, incentive, and interoperability arrangements.

## **Legal and Standards Research**

Interprets authoritative texts, institutional status, and formal requirements.

## **Comparative Research**

Examines differences across models, evaluators, methods, organizations, sectors, or jurisdictions.

## **Forecasting Research**

Makes explicit probabilistic or scenario-based claims about future conditions.

## **Synthesis Research**

Integrates existing evidence through systematic, scoping, rapid, or narrative review.

The project does not treat these modes as interchangeable.

A literature review cannot prove that a new evaluation protocol works.

A pilot cannot establish universal validity.

A case study cannot estimate a population rate without additional design.

An expert panel cannot convert uncertainty into observed fact.

An experimental result cannot alone establish the legitimacy of a binding standard.

Standards Body uses prospective planning to distinguish prediction from explanation after the fact.

Preregistration is required or strongly preferred when:

- A confirmatory hypothesis is being tested
- analyst flexibility could materially alter conclusions
- a benchmark or evaluator comparison may create reputational consequences
- a threshold or standard may depend on the result
- selective reporting is a credible risk

Preregistration may remain confidential or embargoed when public release would compromise held-out evaluation content, security, privacy, intellectual property, or the research question.

A registration does not make a weak design strong.

It creates a record of what was planned.

Exploratory research remains legitimate.

It should be labeled as exploratory rather than presented as prospectively confirmed.

Research openness is also bounded.

Standards Body supports open science, reproducibility, and reusable research artifacts.

It does not interpret openness as mandatory release of:

- Dangerous capability instructions
- active exploit details
- held-out tasks
- sensitive model access
- personal information
- legally restricted data
- confidential incident evidence

The project therefore uses the rule:

**As open as responsible, as restricted as necessary, and explicit about the difference.**

The methodology requires active search for contrary evidence.

Researchers should not merely gather sources supporting the expected conclusion.

They should seek:

- Failed replications
- negative results
- methodological criticism
- alternative explanations
- incidents
- dissenting experts
- incompatible legal interpretations
- evidence from smaller or less visible institutions

- evidence that the proposed standard would create burden, capture, or gaming

Research conclusions are assigned evidence levels and confidence according to EVIDENCE\_STANDARDS.md.

The method should identify:

- What is known
- what is inferred
- what is assumed
- what is forecast
- what remains unknown
- what could change the conclusion

Artificial intelligence tools may assist research.

They may support:

- Search planning
- document classification
- extraction
- translation
- coding
- analysis
- drafting
- consistency review

They may not serve as unverified independent evidence.

The research owner remains responsible for every published claim.

The project should record material model and tool use when it affects:

- Evidence collection
- analysis
- reproducibility
- confidentiality
- authorship
- interpretation

The methodology treats correction as a normal research function.

A credible research institution is not one that never changes its conclusions.

It is one that:

- Preserves the original record
- identifies error
- explains the effect
- corrects dependent work

- learns from the failure
- prevents recurrence

The final methodological rule is:

**Plan the work, identify the object, preserve the evidence, challenge the conclusion, state the uncertainty, and keep the result correctable.**

---

# 1. Foundational Research Principles

## 1.1 Decision Relevance

Research should identify the decision, claim, or institutional question it is intended to inform.

A project may be curiosity-driven.

Even then, it should define the knowledge gap it addresses.

## 1.2 Methodological Fit

The method should fit the question.

Do not select a method because:

- It is fashionable
- it produces a simple number
- it is easy to publish
- a tool is available
- another institution uses it
- it creates an appearance of rigor

## 1.3 Prospective Clarity

Where possible, define hypotheses, outcomes, exclusions, and analysis before observing the decisive results.

## 1.4 Exploratory Honesty

Exploratory work is valuable.

It should be reported as exploratory.

## **1.5 Evaluated-Object Precision**

Research involving AI should identify the relevant model, system, configuration, access, tools, safeguards, protocol, environment, and date.

## **1.6 Evidence Traceability**

Material findings should be traceable to data, sources, methods, code, logs, judgments, and reviewers.

## **1.7 Proportionality**

Research rigor, independence, security, and review should increase with:

- Consequence
- uncertainty
- novelty
- irreversibility
- conflict
- public reliance

## **1.8 Independence Without Isolation**

Researchers should seek necessary access and technical cooperation without surrendering control of methods or conclusions.

## **1.9 Contrary Evidence**

Research should actively search for evidence that could falsify, narrow, or overturn the expected conclusion.

## **1.10 Uncertainty Preservation**

Uncertainty should remain visible through analysis, review, and publication.

## **1.11 Reproducibility and Replication**

Research should support independent verification to the degree lawful, safe, and technically possible.

## **1.12 Ethical and Legal Fitness**

Research should respect:

- Human rights

- participant welfare
- privacy
- security
- intellectual property
- contractual duties
- applicable law
- professional ethics

### **1.13 Security by Design**

Security review should occur before sensitive evidence is collected, not only before publication.

### **1.14 Open Science With Boundaries**

Research artifacts should be findable, accessible under stated conditions, interoperable, and reusable where responsible.

### **1.15 Correction**

Every canonical research output should have a correction pathway.

### **1.16 No Authority Inflation**

Research findings should not be described as standards, certification, accreditation, regulation, or legal approval unless the appropriate process and authority exist.

### **1.17 Public Legibility**

Important methods and limitations should be understandable to non-specialists without sacrificing accuracy.

### **1.18 Institutional Reflexivity**

Standards Body should research the effects of its own methods, incentives, governance, and public claims.

---

## **2. Scope and Non-Claims**

### **2.1 Research Covered**

This methodology covers:

- Desk research

- literature reviews
- technical experiments
- frontier AI evaluations
- red-team research
- benchmark research
- human-uplift studies
- interviews
- surveys
- workshops
- Delphi and expert elicitation
- case studies
- institutional analysis
- governance design
- standards-development research
- legal and policy analysis
- comparative research
- historical research
- forecasting
- incident analysis
- pilot evaluation
- metaresearch

## 2.2 Activities Requiring Additional Review

Additional qualified review may be required for:

- Human-participant research
- personal or sensitive data
- vulnerable populations
- hazardous biological or chemical work
- active cybersecurity testing
- deception
- covert observation
- controlled or classified information
- dual-use capability research
- access to model weights
- release of exploitable findings
- research in regulated sectors

## 2.3 Research Versus Practice

Operational work may generate research evidence.

Examples:

- An evaluation service

- a standards pilot
- an incident response
- a certification exercise

The project should determine whether the activity is:

- Routine practice
- quality improvement
- research
- both

Legal definitions may differ by jurisdiction.

When human participants are involved, qualified ethics and legal review should determine applicable obligations.

## 2.4 No Universal Method

This methodology does not require every project to use:

- An experiment
- a systematic review
- preregistration
- quantitative analysis
- public data
- peer review
- one statistical framework

It requires a justified method and a clear record.

---

## 3. Canonical Research Definitions

Definitions in TERMINOLOGY.md govern.

### 3.1 Research

A systematic or structured activity intended to generate, test, synthesize, or apply knowledge.

### 3.2 Research Question

A clearly formulated question that guides evidence collection and analysis.

### 3.3 Hypothesis

A testable proposed explanation, relationship, prediction, or difference.

### **3.4 Protocol**

The prospective specification governing research purpose, question, method, data, analysis, security, ethics, review, and change control.

### **3.5 Preregistration**

A time-stamped record of a research plan created before the decisive data are collected or analyzed.

### **3.6 Registered Report**

A publication process in which the research question and method receive review before results are known, with publication commitment based primarily on the importance and rigor of the design rather than result direction.

### **3.7 Exploratory Analysis**

Analysis undertaken to discover patterns, generate hypotheses, or examine unexpected findings.

### **3.8 Confirmatory Analysis**

Analysis testing prospectively specified hypotheses or decision rules.

### **3.9 Deviation**

A departure from the approved or registered protocol.

### **3.10 Amendment**

A documented prospective change to the protocol.

### **3.11 Research Artifact**

A digital or physical object used or produced in research, including data, code, prompts, models, logs, tasks, instruments, rubrics, notes, or environments.

### **3.12 Data Management Plan**

A document governing data creation, access, storage, metadata, security, retention, sharing, and disposal.

### **3.13 Reproducibility**

Obtaining consistent computational or analytical results using the same data, code, methods, and conditions.

### **3.14 Replicability**

Obtaining substantively consistent findings through materially independent research addressing the same question.

### **3.15 Triangulation**

Using multiple methods, sources, or perspectives to examine the same claim.

### **3.16 Research Integrity**

Honest, verifiable, accountable, and professionally responsible conduct in proposing, performing, evaluating, and reporting research.

### **3.17 Research Security**

Protection of research people, information, artifacts, infrastructure, partnerships, and intellectual assets against unauthorized access, misuse, coercion, or interference.

### **3.18 Dual-Use Research**

Research with legitimate beneficial purposes and plausible applications that could create harm.

### **3.19 Human Participant**

A person whose interaction, behavior, data, or identifiable private information is involved in research, subject to applicable legal definitions.

### **3.20 Methodological Audit**

A structured review of whether the research was planned, executed, analyzed, and reported according to its stated method and applicable standards.

---

## 4. Standards Body Research Portfolio

### 4.1 Foundational Research

Purpose:

Develop concepts, first principles, definitions, and institutional architecture.

Typical outputs:

- Canonical working white papers
- foundational sources
- conceptual frameworks
- taxonomies

Primary methods:

- Structured literature review
- conceptual analysis
- comparative institutional analysis
- expert critique
- contrarian review

### 4.2 Evaluation-Science Research

Purpose:

Develop and test methods for measuring frontier AI systems.

Typical outputs:

- Protocols
- task families
- scoring methods
- elicitation studies
- validity studies
- statistical models

Primary methods:

- Technical experiment
- simulation
- benchmark analysis
- task sampling
- human-baseline study
- replication
- transcript analysis

## 4.3 Capability and Safeguard Research

Purpose:

Measure consequential capabilities, access pathways, safeguards, and deployment conditions.

Primary methods:

- Held-out evaluation
- controlled environment
- red-team study
- human uplift
- adversarial testing
- expert judgment
- incident analysis

## 4.4 Institutional-Design Research

Purpose:

Develop governance, assurance, standards, accreditation, incentive, and interoperability systems.

Primary methods:

- Comparative case study
- process mapping
- stakeholder interview
- failure analysis
- pilot
- scenario analysis
- institutional simulation
- impact assessment

## 4.5 Standards Research

Purpose:

Determine whether a practice is mature enough for guidance, specification, standard, assurance, procurement, or formal requirement.

Primary methods:

- Evidence synthesis
- implementation pilot
- public comment
- interoperability testing
- burden analysis
- competition analysis

- evaluator-capacity analysis

## 4.6 Legal and Regulatory Research

Purpose:

Establish current legal status, authority, requirements, and institutional relationships.

Primary methods:

- Authoritative text review
- legislative history
- regulatory guidance analysis
- case-law analysis
- jurisdictional comparison
- qualified legal review

## 4.7 Public-Knowledge Research

Purpose:

Explain frontier AI standards questions accurately and accessibly.

Primary methods:

- Source verification
- synthesis
- explanatory writing
- fact checking
- expert review
- public-comprehension testing

## 4.8 Metaresearch

Purpose:

Evaluate research quality, publication bias, reproducibility, methods, incentives, and institutional effects.

Primary methods:

- Research audit
- replication study
- publication analysis
- citation analysis
- methods review
- outcome tracking

## 5. Research Classification

Every substantial project should be classified before work begins.

### 5.1 Project Type

Select one or more:

- Descriptive
- exploratory
- confirmatory
- evaluative
- causal
- predictive
- comparative
- qualitative
- quantitative
- mixed-method
- legal
- institutional
- historical
- synthesis
- pilot
- metaresearch

### 5.2 Consequence Level

#### **C0: Minimal**

Low consequence if wrong.

#### **C1: Limited**

May affect internal prioritization or nonconsequential public understanding.

#### **C2: Material**

May affect organizational decisions, public claims, research agendas, or resource allocation.

#### **C3: High**

May affect deployment, safeguards, procurement, standards, certification, institutional recognition, or significant public decisions.

#### **C4: Critical**

May affect severe-risk decisions, binding requirements, critical infrastructure, national security, or difficult-to-reverse action.

## 5.3 Research Sensitivity

### **R0: Public**

No material restriction.

### **R1: Controlled**

Access limited for integrity, licensing, or coordination.

### **R2: Confidential**

Contains proprietary, personal, contractual, or unpublished information.

### **R3: Restricted**

Contains material security, vulnerability, dangerous-capability, or high-risk information.

### **R4: Highly Restricted**

Requires specialized authorization, secure facilities, or legal controls.

## 5.4 Review Level

### **V0: Author Self-Check**

Suitable only for low-consequence informal work.

### **V1: Internal Peer Review**

At least one qualified internal reviewer.

### **V2: Domain and Method Review**

Separate subject-matter and methodological review.

### **V3: Independent Expert Review**

Qualified reviewers with sufficient independence and access.

### **V4: Multi-Party or Formal Assurance Review**

Used for critical claims, standards progression, or formal decisions.

## 5.5 Registration Level

### **P0: No Registration**

Used for informal notes or rapidly changing exploratory work.

### **P1: Internal Protocol Timestamp**

Time-stamped internal record.

### **P2: Public or Embargoed Registration**

Registered before decisive data access or analysis.

### **P3: Independently Reviewed Protocol**

Protocol reviewed before execution.

### **P4: Registered Report or Equivalent**

Publication or institutional commitment based on pre-results review.

## **5.6 Classification Record**

Record:

- Project type
  - consequence level
  - sensitivity
  - review level
  - registration level
  - rationale
  - approver
  - review triggers
- 

## **6. Research Lifecycle**

### **6.1 Initiation**

Create a research intake record.

### **6.2 Scoping**

Define:

- Problem
- question
- audience
- decision
- scope
- non-scope
- expected output

## 6.3 Feasibility

Assess:

- Access
- expertise
- data
- compute
- security
- ethics
- time
- cost
- institutional authority

## 6.4 Classification

Apply Section 5.

## 6.5 Protocol

Prepare the research protocol.

## 6.6 Registration

Timestamp or register as appropriate.

## 6.7 Collection

Gather data, sources, and artifacts.

## 6.8 Execution

Conduct the research and record deviations.

## 6.9 Analysis

Apply pre-specified and clearly labeled exploratory analyses.

## 6.10 Review

Complete required review.

## 6.11 Synthesis

Assign evidence level, confidence, and limitations.

## 6.12 Publication

Release public, controlled, or restricted output.

## 6.13 Monitoring

Track new evidence and change triggers.

## 6.14 Correction

Correct, supersede, withdraw, or retire.

---

# 7. Research Question Formation

## 7.1 Question Test

A strong research question should be:

- Clear
- bounded
- answerable
- relevant
- method-compatible
- ethically and securely researchable
- explicit about the object and period

## 7.2 Decision Question

State:

Which decision or institutional judgment could change because of this research?

## 7.3 Claim Question

State:

Which claim would the evidence support, weaken, or leave unresolved?

## 7.4 Object Question

State:

Which model, system, protocol, organization, jurisdiction, process, or event is being studied?

## 7.5 Counterfactual Question

Where causal or comparative reasoning is involved:

Compared with what?

## 7.6 Consequence Question

What happens if the research conclusion is wrong?

## 7.7 Scope Question

What will the project not establish?

## 7.8 Feasibility Question

Can the available method and access support the intended conclusion?

## 7.9 Research Question Register

Record:

- Question ID
  - exact wording
  - owner
  - project type
  - decision link
  - claim
  - object
  - scope
  - evidence standard
  - status
-

## 8. Protocol Development

### 8.1 Required Protocol Fields

A substantial protocol should include:

- Title
- version
- owners
- question
- rationale
- decision link
- claims
- hypotheses
- project classification
- background
- methods
- data or sources
- sampling
- evaluated object
- variables or constructs
- outcomes
- analysis
- uncertainty
- exclusions
- ethics
- security
- conflicts
- AI-tool use
- review
- publication
- correction
- change control
- timeline

### 8.2 Protocol Proportionality

A short descriptive project may use a concise protocol.

A C3 or C4 project requires a complete protocol and independent review.

## 8.3 Primary and Secondary Outcomes

For confirmatory studies, define:

- Primary outcome
- secondary outcomes
- exploratory outcomes
- stopping rule
- success criteria

## 8.4 Analysis Plan

State:

- Models or methods
- transformations
- missing data
- exclusions
- multiple testing
- uncertainty
- sensitivity analysis
- subgroup analysis
- software
- code review

## 8.5 Deviations

A deviation record should state:

- What changed
- when
- why
- who approved
- effect on validity
- whether the change occurred before or after observing relevant results

## 8.6 Protocol Amendments

Prospective amendments are permitted.

They should be:

- Versioned
- justified
- reviewed
- distinguishable from post-result changes

## 8.7 Protocol Freeze

For confirmatory analysis, freeze the relevant protocol before decisive analysis begins.

## 8.8 Protocol Availability

Publish or register the protocol unless:

- Security
- holdout integrity
- privacy
- law
- contract
- legitimate intellectual property

requires restriction.

Even when restricted, preserve a time-stamped record.

---

# 9. Preregistration and Registered Research

## 9.1 Purpose

Preregistration distinguishes planned research from analysis developed after seeing results.

The Center for Open Science describes preregistration as specifying a research plan in advance and submitting it to a registry, which helps distinguish planned from unplanned work.<sup>[^cos-prereg]</sup>

## 9.2 Required Use

Preregistration is normally required for:

- Confirmatory hypotheses
- evaluator rankings
- threshold validation
- high-stakes comparative claims
- studies likely to affect standards
- studies with substantial analytical flexibility
- public claims with significant reputational effect

## 9.3 Strongly Preferred Use

Preregistration is strongly preferred for:

- Human-uplift experiments

- model comparisons
- safeguard-effectiveness experiments
- evaluator-proficiency research
- survey experiments
- forecasting tournaments

## 9.4 Optional Use

It may be optional for:

- Exploratory mapping
- rapid incident response
- conceptual analysis
- historical research
- source-of-truth maintenance
- early prototype work

## 9.5 Registration Timing

Register before:

- Decisive data collection
- access to protected outcomes
- final task sampling
- confirmatory analysis

## 9.6 Registration Content

Include:

- Questions
- hypotheses
- outcomes
- sampling
- exclusions
- analysis
- stopping
- deviations
- publication
- access restrictions

## 9.7 Embargoed Registration

Use an embargo when public release would compromise:

- Held-out evaluations

- vulnerability research
- confidential partnerships
- model access
- participant privacy
- intellectual property

## 9.8 Registered Reports

Use a registered-report model where feasible for research vulnerable to publication bias or result-driven review.

## 9.9 Deviations

Report every material deviation.

A preregistration is not a promise that no change will occur.

It is a record that makes change visible.

## 9.10 Preregistration Limits

Preregistration does not:

- Validate the construct
- fix a weak sample
- create independence
- remove bias
- guarantee publication
- eliminate exploratory discovery

---

# 10. Research Intake Template

**Project ID:**

**Working title:**

**Research owner:**

**Date:**

**Problem**

**Research Question**

**Decision Link**

**Intended Claims**

**Evaluated Object**

**Audience**

**Project Type**

**Consequence Level**

**Sensitivity Level**

**Required Review Level**

**Registration Level**

**Expected Evidence Standard**

**Access and Resource Needs**

**Ethics and Human-Participant Screen**

**Dual-Use and Security Screen**

**Conflicts**

**Expected Deliverables**

**Approval**

---

## 11. Research Protocol Template

Project ID:

Protocol version:

Date:

Owners:

Status:

- 1. Research Question**
- 2. Rationale and Knowledge Gap**
- 3. Decision and Claim Link**
- 4. Scope and Non-Claims**
- 5. Hypotheses or Exploratory Aims**
- 6. Evaluated Object**
- 7. Research Design**
- 8. Data, Sources, or Participants**
- 9. Sampling**
- 10. Procedures**
- 11. Variables, Constructs, and Outcomes**
- 12. Analysis Plan**
- 13. Uncertainty and Sensitivity**
- 14. Exclusions and Missing Data**
- 15. Ethics and Participant Protection**
- 16. Security and Dual-Use Controls**
- 17. Data and Artifact Management**
- 18. Conflicts and Independence**
- 19. AI Tool Use**

## 20. Review Plan

## 21. Publication and Disclosure

## 22. Correction and Retirement

## 23. Change Control

## 24. Timeline and Resources

## 25. Approval

---

# 12. Literature and Evidence Review Methods

Standards Body uses several forms of review.

The method should match the purpose.

## 12.1 Systematic Review

Use when:

- The question is sufficiently defined
- the evidence base is large enough
- inclusion criteria can be specified
- comprehensive search and transparent selection are important
- the output may support a consequential standard or decision

Required elements:

- Protocol
- search strategy
- databases and sources
- date range
- inclusion and exclusion
- duplicate handling
- screening
- extraction
- quality assessment
- synthesis
- contrary evidence
- limitations
- update date

PRISMA 2020 provides a widely used reporting framework for systematic reviews, including checklists and flow diagrams. Standards Body may use PRISMA reporting elements where appropriate, while recognizing that a reporting guideline does not by itself determine whether the underlying review method is valid.[^prisma]

## 12.2 Scoping Review

Use when:

- The topic is broad or emerging
- terminology is unstable
- the goal is to map evidence and gaps
- a systematic effect estimate is premature
- research types are heterogeneous

Typical outputs:

- Evidence map
- terminology map
- source clusters
- gap register
- future-review recommendation

PRISMA-ScR may guide reporting for scoping reviews.[^prisma-scr]

## 12.3 Rapid Review

Use when:

- A decision is time-sensitive
- a full systematic review is not feasible
- limitations from reduced scope can be accepted and disclosed

Required controls:

- Explicit shortcuts
- limited claim
- priority primary sources
- reviewer check
- update plan

## 12.4 Narrative Review

Use when:

- The question is conceptual
- evidence is highly heterogeneous
- institutional history or theory matters

- synthesis requires judgment

Required controls:

- Search description
- source rationale
- contrary evidence
- author perspective
- limitations
- separation of fact and interpretation

## 12.5 Living Review

Use when:

- The topic changes rapidly
- new model releases or evaluations can alter conclusions
- legal or standards status changes frequently
- public reliance is ongoing

Required elements:

- Monitoring sources
- update triggers
- review cadence
- current-as-of date
- version history
- supersession

## 12.6 Review Method Selection

The protocol should explain why the selected review type is fit for purpose.

## 12.7 Search Strategy

A search strategy should specify:

- Concepts
- synonyms
- controlled vocabulary
- named entities
- dates
- languages
- domains
- source types
- databases
- official repositories

- citation chaining
- grey literature

## 12.8 Search Breadth

Search should include:

- Supporting evidence
- contrary evidence
- negative results
- withdrawn or corrected work
- operational reports
- incident evidence
- underrepresented institutions
- non-English sources where relevant

## 12.9 Source Prioritization

Prioritize primary sources for:

- Laws
- regulations
- standards
- institutional status
- model releases
- evaluation methods
- original research
- official decisions

## 12.10 Screening

For structured reviews, use:

- Title screening
- abstract or summary screening
- full-text review
- exclusion reasons
- duplicate resolution

Higher-consequence reviews should use at least two reviewers for a sample or all material inclusion decisions.

## 12.11 Extraction

Use a standardized form containing:

- Source ID
- claim
- method
- object
- sample
- result
- uncertainty
- conflict
- limitations
- status
- relevance

## 12.12 Quality Assessment

Apply EVIDENCE\_STANDARDS.md.

Do not exclude all low-quality evidence automatically.

Low-quality evidence may reveal:

- Early signals
- neglected questions
- practical concerns
- bias in the evidence ecosystem

Its weight should remain limited.

## 12.13 Citation Chaining

Use:

- Backward citation review
- forward citation review
- author search
- institution search
- dataset or protocol search
- related correction and retraction search

## 12.14 Search Completion

A search is sufficiently complete when additional searching is unlikely to materially change the bounded conclusion, or when time and resource limits are documented.

## 12.15 Review Flow Record

Maintain counts for:

- Records identified
- duplicates
- screened
- excluded
- included
- unavailable
- restricted

## 12.16 Review Update

Update after:

- Material new evidence
- new protocol or model generation
- legal change
- major incident
- contradiction
- scheduled review

---

# 13. Source Discovery and Verification

## 13.1 Discovery Is Not Verification

Search tools, recommendations, citations, and model outputs may locate sources.

The underlying source should be inspected.

## 13.2 Source Verification Checklist

Confirm:

- Title
- author or institution
- publication date
- current status
- version
- actual content
- corrections
- primary or secondary status
- relevance
- access

## 13.3 Official Status Claims

For claims about:

- Law
- standards
- certification
- accreditation
- organization role
- product status
- current framework

check the official source shortly before publication.

## 13.4 Archiving

For dynamic sources, preserve where lawful:

- Archive link
- downloaded copy
- screenshot
- hash
- access date
- change note

## 13.5 Unavailable Sources

If a source cannot be accessed:

- Do not imply full review
- identify reliance on an abstract or summary
- lower confidence
- seek alternative evidence
- request access where necessary

## 13.6 Translation

For non-English sources:

- Preserve original text
- record translation method
- use qualified review for legal or technical nuance
- identify ambiguous terms
- avoid assuming literal equivalence

## 13.7 Retractions and Corrections

Check for:

- Retraction
- correction
- expression of concern
- superseding version
- withdrawn standard
- expired certificate

## 13.8 Grey Literature

Grey literature may include:

- Government reports
- company reports
- standards drafts
- conference material
- technical documentation
- working papers
- incident reports

Assess it by method, provenance, access, and conflict rather than by publication label alone.

---

# 14. Technical Experiment Methodology

## 14.1 Experimental Question

State:

- Independent or manipulated factor
- outcome
- comparison
- system
- environment
- expected mechanism

## 14.2 Experimental Unit

Define the unit of analysis.

Possible units:

- Task
- model run

- trajectory
- user
- evaluator
- system configuration
- organization
- incident

Avoid pseudo-replication.

Multiple outputs from one shared condition may not be independent.

### 14.3 Control Condition

Use an appropriate comparison:

- No AI
- human baseline
- prior model
- alternate scaffold
- safeguard off
- standard access
- current protocol
- random or placebo condition where ethical

### 14.4 Randomization

Use randomization when it improves causal interpretation or reduces selection bias.

Record:

- Unit randomized
- method
- seed
- blocking
- stratification
- deviations

### 14.5 Blinding

Use blinding where feasible for:

- Human scoring
- model identity
- treatment condition
- hypothesis
- source
- sponsor

Record what could and could not be blinded.

## 14.6 Sample Size

Justify sample size through:

- Statistical power
- precision
- task diversity
- rare-event detection
- resource constraints
- sequential design
- pilot purpose

## 14.7 Stopping Rules

Define:

- Fixed sample
- precision target
- sequential boundary
- safety stop
- futility
- time or cost stop

Do not stop because a preferred result appears unless the rule was specified or clearly labeled exploratory.

## 14.8 Multiple Comparisons

When many models, tasks, domains, or metrics are tested:

- Define primary outcomes
- control false discovery where appropriate
- report all material analyses
- distinguish exploratory findings
- avoid selective emphasis

## 14.9 Missing Data

Record:

- Missingness
- failure cause
- invalid runs
- timeouts

- system errors
- scoring failures
- handling method
- sensitivity

## 14.10 Pilot Studies

Pilots may assess:

- Feasibility
- variance
- task validity
- burden
- safety
- instrumentation
- recruitment
- evaluator agreement

Do not present a pilot as definitive evidence without justification.

## 14.11 Sequential Research

Sequential methods may be useful when:

- Models change quickly
- evaluation is costly
- evidence arrives continuously
- safety stopping is necessary

The decision rule should be specified.

## 14.12 Adaptive Designs

Adaptive designs may change:

- Task allocation
- sample size
- elicitation
- evaluator effort

Adaptation should be governed prospectively and preserve valid inference.

## 14.13 Safe Proxies

Where direct testing is dangerous, use safe proxies when they preserve the relevant construct sufficiently.

Proxy limitations should be central to interpretation.

## 14.14 Experimental Records

Preserve:

- Protocol
  - code
  - environment
  - data
  - logs
  - prompts
  - model identity
  - deviations
  - analysis
  - review
- 

## 15. Frontier AI Evaluation Research

### 15.1 Evaluation as Research

A frontier AI evaluation may be:

- A measurement activity
- a research study
- an assurance activity
- a decision procedure

The protocol should identify which functions apply.

### 15.2 Evaluation Object

Record:

- Model family
- exact version
- system configuration
- tools
- scaffolds
- retrieval
- prompt
- access
- safeguards
- deployment context

## 15.3 Construct

Define the capability, behavior, safeguard, or institutional property being measured.

## 15.4 Task Universe

Describe the broader set of tasks the sample is intended to represent.

## 15.5 Task Development

Task development should include:

- Domain expertise
- difficulty calibration
- validity review
- provenance
- contamination assessment
- scoring
- security
- alternate forms
- retirement

## 15.6 Held-Out Design

Apply FOUNDATION\_02\_HELD\_OUT\_EVALUATIONS.md.

Protected material may include:

- Tasks
- solutions
- environments
- scoring
- attack strategies
- task-generation rules

## 15.7 Elicitation Research

Elicitation should specify:

- Prompting
- examples
- tool access
- search
- fine-tuning
- human assistance
- retries

- compute
- time
- optimization budget

AISI's science-of-evaluations work has emphasized that test-time resources and elicitation protocols can materially affect measured agent capability, reinforcing the need to treat elicitation as part of the method rather than an incidental detail.[^aisi-eval-science]

## 15.8 Agentic Evaluation

Agent research should record:

- Task horizon
- environment
- actions
- observation
- memory
- tools
- permissions
- persistence
- human intervention
- trajectory quality
- resource use
- stopping

## 15.9 Transcript and Trajectory Analysis

Quantitative success should be supplemented where useful by analysis of:

- Strategy
- failure mode
- recovery
- deception
- tool use
- environmental interaction
- unsafe intermediate action
- evaluator intervention

## 15.10 Scoring

Scoring should be:

- Valid
- reliable
- reviewable
- calibrated

- robust to ambiguity
- versioned

## 15.11 Model Judges

Validate model judges against qualified human assessment.

Examine:

- Shared model lineage
- position bias
- verbosity bias
- style bias
- adversarial manipulation
- calibration
- disagreement

## 15.12 Human Baselines

Human baselines should define:

- Population
- expertise
- tools
- time
- incentives
- sample
- support
- scoring

## 15.13 Human-Uplift Studies

Compare:

- Human alone
- human with AI
- AI alone
- expert and nonexpert users
- different access and training

Measure:

- Success
- time
- quality
- error
- learning

- misuse
- reliance

## 15.14 Evaluation Validity

Validate through:

- Content review
- criterion linkage
- real-world comparison
- replication
- incident evidence
- sensitivity
- task perturbation
- alternate protocols

## 15.15 Dynamic Protocols

Apply FOUNDATION\_01\_DYNAMIC\_EVALUATION\_PROTOCOLS.md.

## 15.16 High-Stakes Evaluation

Apply FOUNDATION\_03\_HIGH\_STAKES\_CAPABILITY\_EVALUATION.md.

## 15.17 Independent Review

Apply FOUNDATION\_04\_INDEPENDENT\_EXPERT\_REVIEW.md.

## 15.18 Tooling and Reproducibility

Evaluation frameworks such as AISI's Inspect illustrate the value of structured, inspectable evaluation tooling, although tool adoption does not itself establish validity.[^inspect]

---

# 16. Statistical Methodology

## 16.1 Statistical Purpose

Statistics should serve the question.

Do not use complexity to substitute for design quality.

## 16.2 Descriptive Statistics

Report distributions, not only averages.

Consider:

- Median
- quantiles
- variance
- failure rate
- tail behavior
- task-level results
- subgroup results

## 16.3 Estimation

Prefer effect estimates and uncertainty over binary significance alone.

## 16.4 Hypothesis Testing

Where used, report:

- Hypothesis
- test
- assumptions
- statistic
- p-value
- effect size
- interval
- multiplicity
- interpretation

A p-value is not the probability that the claim is true.

## 16.5 Bayesian Methods

Bayesian methods may support:

- Hierarchical evaluation
- sparse data
- sequential updating
- probability statements
- partial pooling

State:

- Prior

- likelihood
- model
- sensitivity
- posterior interpretation

## 16.6 Hierarchical Structure

AI evaluation data are often nested:

- Trials within tasks
- tasks within domains
- models within families
- judges within panels
- users within groups

Account for dependence.

## 16.7 Nondeterminism

Model outputs may vary across:

- Sampling
- infrastructure
- hidden system changes
- tool calls
- environment

Use repeated runs and report variance.

## 16.8 Rare Events

For rare harmful events:

- Report exposure
- use appropriate intervals
- avoid claiming zero risk from zero observations
- consider stress testing and structured expert judgment

## 16.9 Calibration

Forecasts, probabilistic classifiers, and model judges should be assessed for calibration.

## 16.10 Sensitivity Analysis

Test the effect of:

- Exclusions

- priors
- scoring
- thresholds
- missing data
- task weights
- judge choice
- system version

## **16.11 Robustness Checks**

Robustness checks should be relevant, not decorative.

## **16.12 Statistical Code Review**

C2 through C4 projects should receive code review or independent recomputation where feasible.

## **16.13 No Universal Composite Score**

Avoid combining unrelated measures into one score unless:

- Construct is defined
  - weighting is justified
  - components remain visible
  - sensitivity is tested
- 

# **17. Qualitative Research Methodology**

## **17.1 Appropriate Uses**

Qualitative research is appropriate for:

- Institutional practice
- governance
- evaluator experience
- incident narratives
- standards implementation
- expert reasoning
- stakeholder impact
- emerging concepts

## 17.2 Sampling

Possible methods:

- Purposive
- maximum variation
- snowball
- criterion
- theoretical
- convenience

State the rationale and limitations.

## 17.3 Interviews

An interview protocol should include:

- Purpose
- consent
- recording
- confidentiality
- question guide
- role
- conflicts
- data handling

## 17.4 Focus Groups and Workshops

Consider:

- Group pressure
- confidentiality limits
- power differences
- facilitator influence
- documentation
- dissent

## 17.5 Coding

Define:

- Codebook
- inductive or deductive method
- coders
- training
- disagreement
- software

- revision

## 17.6 Thematic Analysis

Report:

- Data source
- coding process
- theme development
- negative cases
- quotations
- researcher interpretation

## 17.7 Reflexivity

Researchers should record how their:

- Role
- access
- expectations
- institution
- values
- relationships

may affect collection and interpretation.

## 17.8 Saturation

Do not invoke saturation without stating:

- What type
- how assessed
- sample context
- remaining gaps

## 17.9 Member Checking

Where appropriate, participants may review:

- Factual accuracy
- quotations
- contextual interpretation

Participants should not receive veto over legitimate analysis unless agreed.

## 17.10 Qualitative Rigor

Support rigor through:

- Triangulation
  - audit trail
  - negative-case analysis
  - peer debrief
  - reflexivity
  - rich description
  - preserved dissent
- 

## 18. Mixed-Methods Research

### 18.1 Purpose

Mixed methods should integrate complementary evidence.

Do not add interviews to a quantitative study merely to claim comprehensiveness.

### 18.2 Designs

Possible designs:

- Convergent
- explanatory sequential
- exploratory sequential
- embedded
- multiphase

### 18.3 Integration

State where integration occurs:

- Question
- sampling
- collection
- analysis
- interpretation
- decision

## 18.4 Divergence

When qualitative and quantitative findings conflict:

- Preserve conflict
  - examine object and sample
  - inspect measurement
  - seek additional evidence
  - avoid forced resolution
- 

## 19. Case-Study Methodology

### 19.1 Case Definition

Define the bounded case by:

- Event
- organization
- system
- process
- place
- time
- decision

### 19.2 Case Selection

Possible rationales:

- Critical case
- extreme case
- typical case
- revelatory case
- failure case
- comparative case
- longitudinal case

### 19.3 Evidence Sources

Use multiple sources where possible:

- Documents
- interviews
- logs
- artifacts
- public records

- evaluations
- timelines
- observations
- incident evidence

## 19.4 Chain of Evidence

Maintain traceability from:

- Question
- source
- observation
- interpretation
- conclusion

## 19.5 Rival Explanations

Identify and test alternative explanations.

## 19.6 Generalization

Case studies support:

- Analytical generalization
- mechanism understanding
- hypothesis development
- institutional learning

They do not automatically support population estimates.

## 19.7 Failure Case

A failure case should examine:

- Expected function
- failure mode
- trigger
- contributing factors
- governance
- incentives
- detection
- response
- recurrence

## 19.8 Comparative Case Study

Use a common framework across cases.

Explain selection and noncomparability.

---

# 20. Institutional-Design Research

## 20.1 Research Object

Institutional design may examine:

- Governance
- standards process
- accreditation
- certification
- evaluator markets
- incentives
- registries
- recognition
- international coordination

## 20.2 Design Inputs

Use:

- First principles
- existing institutions
- failure cases
- stakeholder needs
- legal constraints
- capacity
- incentives
- public-interest analysis

## 20.3 Functional Analysis

Ask:

- Which function is needed?
- Which actor performs it?
- What authority is required?
- What conflicts arise?
- What evidence supports the design?
- How can it fail?

- How is it corrected?

## 20.4 Comparative Institutional Analysis

Compare:

- Mandate
- governance
- funding
- authority
- competence
- transparency
- participation
- appeals
- outcomes

## 20.5 Design Prototypes

Produce:

- Process map
- role map
- decision matrix
- template
- pilot rule
- governance charter
- registry schema

## 20.6 Institutional Pilot

Test:

- Feasibility
- role clarity
- burden
- conflict
- gaming
- access
- correction
- interoperability

## 20.7 Outcome Evaluation

Do not evaluate the institution only by:

- Documents produced

- meetings held
- certifications issued
- members enrolled

Measure:

- Decision quality
  - error correction
  - incident learning
  - risk reduction
  - access
  - competition
  - trust
  - unintended effects
- 

## **21. Legal, Regulatory, and Standards Research**

### **21.1 Source Priority**

Use authoritative sources:

- Enacted law
- regulation
- official guidance
- court decisions
- treaties
- standards texts
- official registries
- institutional decisions

### **21.2 Status Control**

Distinguish:

- Proposed
- adopted
- effective
- delayed
- repealed
- superseded
- guidance
- binding rule

## 21.3 Jurisdiction

State:

- Country or region
- authority
- sector
- territorial scope
- effective date

## 21.4 Legal Interpretation

Separate:

- Text
- official interpretation
- judicial interpretation
- analyst interpretation
- unresolved question

## 21.5 Qualified Review

Material legal conclusions should receive qualified legal review.

## 21.6 Standards Research

Record:

- Issuing body
- standard number
- edition
- scope
- status
- voluntary or incorporated status
- certification relevance
- access limitations

## 21.7 Crosswalks

A crosswalk maps provisions.

It does not establish legal equivalence unless recognized by the relevant authority.

## 21.8 Current Verification

Reverify legal and standards claims before publication.

---

## 22. Comparative and International Research

### 22.1 Unit of Comparison

Define whether the comparison concerns:

- Models
- protocols
- evaluators
- laws
- standards
- institutions
- countries
- sectors
- incidents

### 22.2 Comparability

Assess:

- Terminology
- data
- method
- time
- scope
- legal context
- institutional capacity
- language
- incentives

### 22.3 Local Context

Avoid treating one jurisdiction as the default.

### 22.4 Translation

Validate key terms with domain and local expertise.

### 22.5 Data Gaps

Absence of published data may reflect:

- Capacity
- language

- disclosure norms
- security
- resource inequality

Do not interpret absence as absence of activity.

## 22.6 Recognition Research

Distinguish:

- Evidence recognition
- competence recognition
- process recognition
- legal recognition
- policy agreement

## 22.7 International Participation

Include institutions from affected regions early enough to influence:

- Question
- method
- interpretation
- publication

## 22.8 Capacity Impact

Assess whether the research:

- Builds local capacity
- extracts information
- creates dependence
- concentrates authority

---

# 23. Forecasting and Scenario Research

## 23.1 Forecast Definition

A forecast should specify:

- Target
- resolution criteria
- probability or range
- time horizon
- information date

- forecaster
- update rule

## 23.2 Base Rates

Use relevant historical or comparative base rates where available.

## 23.3 Decomposition

Break complex forecasts into:

- Technical progress
- adoption
- access
- safeguards
- actors
- institutions
- policy
- incidents

## 23.4 Calibration

Track whether stated probabilities correspond to outcomes over time.

## 23.5 Scenarios

Scenarios explore coherent possibilities.

They are not predictions unless probabilities are assigned.

## 23.6 Stress Scenarios

Use stress scenarios to test:

- Institutional resilience
- evaluation gaps
- incident response
- standards failure
- international conflict

## 23.7 Forecast Updates

Record:

- Prior forecast
- new evidence

- updated probability
- reason
- date

## 23.8 Avoided Practice

Do not present:

- A vivid scenario as the expected future
  - an expert quote as a calibrated probability
  - a model-generated forecast as independent evidence
- 

# 24. Structured Expert Judgment

## 24.1 Use Cases

Use when:

- Evidence is sparse
- events are rare
- direct testing is unsafe
- decisions are time-sensitive
- multiple domains interact

## 24.2 Question Design

Questions should be:

- Precise
- decomposed
- resolvable where possible
- explicit about time and units

## 24.3 Individual Judgment First

Collect individual estimates before group discussion to reduce conformity pressure.

## 24.4 Evidence Packet

Provide balanced evidence and contrary views.

## 24.5 Aggregation

State whether estimates are:

- Equal-weighted
- performance-weighted
- median
- interval
- consensus
- unaggregated

## 24.6 Calibration

Where feasible, evaluate experts on seed questions or prior forecasts.

## 24.7 Disagreement

Preserve the distribution and reasons for disagreement.

## 24.8 Update

Expert judgments should have review dates and update triggers.

---

# 25. Data and Artifact Management

## 25.1 Data Management Plan

Every project that creates or handles material data should maintain a data management plan.

The plan should address:

- Data types
- source
- collection
- format
- metadata
- identifiers
- storage
- access
- security
- privacy
- quality
- backup
- retention

- sharing
- licensing
- destruction
- ownership
- stewardship

## 25.2 Research Artifact Inventory

Possible artifacts include:

- Protocol
- registration
- source register
- datasets
- prompts
- task banks
- code
- model adapters
- environment files
- logs
- transcripts
- rubrics
- judge outputs
- statistical notebooks
- interview guides
- coding schemes
- consent records
- review records
- publication files

## 25.3 Persistent Identification

Assign stable identifiers to:

- Projects
- protocols
- datasets
- code releases
- task banks
- models or systems
- evidence packages
- publications

## 25.4 File Naming

Names should support:

- Object identification
- version
- date
- status
- relationship

Avoid ambiguous filenames such as:

- final
- final2
- new
- latest

without controlled versioning.

## 25.5 Metadata

Metadata should identify:

- Creator
- date
- method
- format
- version
- license
- access
- relationship
- provenance
- status
- restrictions

## 25.6 Data Quality

Data-quality checks may include:

- Completeness
- validity
- consistency
- range
- duplicate detection
- timestamps
- schema
- outliers

- label review
- source reconciliation

## 25.7 Raw and Processed Data

Preserve the distinction among:

- Raw
- cleaned
- transformed
- analyzed
- published

Record every material transformation.

## 25.8 Immutable Records

For high-consequence projects, preserve immutable or tamper-evident copies of:

- Raw outputs
- logs
- protocol version
- scoring records
- review findings
- public result

## 25.9 Data Minimization

Collect only data necessary for the research and governance purpose.

## 25.10 Retention

Retention should consider:

- Reproducibility
- legal obligation
- participant consent
- security risk
- task compromise
- institutional learning
- storage burden

## 25.11 Destruction

Destruction should be:

- Authorized
- logged
- complete to the required degree
- consistent with legal holds and research integrity

## 25.12 Stewardship

Every material dataset or artifact collection should have an identified steward.

---

# 26. Open Science and Responsible Access

## 26.1 Open-Science Position

Standards Body supports open research practices because they can improve:

- Scrutiny
- reuse
- replication
- participation
- correction
- interoperability
- public access

The UNESCO Recommendation on Open Science provides an international framework emphasizing open scientific knowledge, infrastructures, engagement, and equitable access.[^unesco-open]

## 26.2 Responsible Openness

Openness should be balanced against:

- Security
- privacy
- consent
- intellectual property
- held-out integrity
- dangerous capability
- legal restriction
- contractual duty

## 26.3 Access Categories

Research artifacts may be:

- Public
- public after embargo
- registration-only
- controlled access
- reviewer-only
- restricted
- unavailable with explanation

## 26.4 FAIR Principles

Where appropriate, research objects should be:

- Findable
- accessible under clear conditions
- interoperable
- reusable

FAIR does not require unrestricted public release. Accessibility may include authenticated or controlled access with rich metadata.[^fair]

## 26.5 Open Methods

Even when data are restricted, publish where safe:

- Research question
- method
- metadata
- analysis plan
- evidence class
- limitations
- review process
- result status

## 26.6 Open Code

Release code when:

- It does not materially increase harm
- licensing permits
- sensitive credentials are removed
- documentation is adequate
- maintenance expectations are clear

## 26.7 Open Tasks

Do not release active held-out tasks merely to satisfy an openness norm.

## 26.8 Equitable Access

Open-science design should consider:

- Language
- disability
- compute
- regional infrastructure
- paywalls
- technical skill
- licensing

## 26.9 Community Contribution

Provide pathways for:

- Reproduction
- issue reporting
- pull requests
- alternate implementations
- translation
- new tasks
- corrections

## 26.10 Openness Statement

Every major output should state:

- Which artifacts are available
  - where
  - under which license
  - which are restricted
  - why
  - how qualified access may be requested
-

## 27. Reproducibility and Replication

### 27.1 Reproducibility Standard

A computational study should allow a qualified reviewer to reconstruct:

- Data inputs
- software
- dependencies
- configuration
- code
- analysis
- outputs

The National Academies defines reproducibility as obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis, while reserving replicability for consistent findings across studies using new data.[^nasem-repro]

### 27.2 Reproducibility Package

Include:

- README
- environment specification
- dependencies
- installation
- execution
- data
- data-access instructions
- code
- configuration
- seeds
- expected outputs
- licenses
- known failures
- hardware or service requirements

### 27.3 Containerization and Environments

Where practical, preserve:

- Container
- lockfile
- package versions
- operating system
- hardware

- accelerator
- API version

## 27.4 External Services

If a result depends on an external API or hosted model, record:

- Provider
- model identifier
- date
- endpoint
- parameters
- relevant service configuration
- known version instability

## 27.5 Reproducibility Status

Use:

- Publicly reproducible
- reproducible under controlled access
- partially reproducible
- not currently reproducible

## 27.6 Replication Plan

High-consequence research should state:

- What an independent replication would test
- which elements should remain the same
- which should vary
- expected range of result
- how disagreement will be handled

## 27.7 Artifact Review

Artifact review should examine whether research artifacts are:

- Available
- functional
- documented
- reusable
- sufficient to reproduce the result

ACM artifact-review practice provides useful distinctions among artifact availability, functional evaluation, reusability, reproduced results, and replicated results.<sup>[^acm-artifacts]</sup>

## 27.8 Replication Credit

Replication and reproduction should receive substantive contributor credit.

## 27.9 Failed Replication

A failed replication should trigger:

- Method comparison
- object comparison
- task comparison
- system-version review
- uncertainty review
- correction where necessary

## 27.10 Restricted Reproducibility

Use controlled environments when public release would compromise:

- Security
  - privacy
  - held-out validity
  - proprietary access
  - safety
- 

# 28. Human-Participant Research

## 28.1 Screening

Before collecting data from or about people, determine whether the activity involves:

- Interaction or intervention
- identifiable private information
- observation
- survey
- interview
- experiment
- user telemetry
- workplace records
- expert elicitation
- public data with re-identification risk

## 28.2 Legal Determination

A qualified institution or responsible official should determine whether formal human-subjects review applies.

In the United States, the Common Rule outlines core provisions for institutional review boards, informed consent, and assurances for covered research, while applicability depends on the conducting or supporting agency and the nature of the activity.[^common-rule]

## 28.3 Ethical Principles

Human-participant research should respect:

- Persons and autonomy
- beneficence
- justice
- privacy
- informed choice
- fair selection
- proportional risk

The Belmont Report remains a foundational statement of respect for persons, beneficence, and justice in research involving human subjects.[^belmont]

## 28.4 Consent

Consent should address:

- Purpose
- procedures
- risks
- benefits
- data use
- confidentiality
- withdrawal
- contacts
- compensation
- AI involvement
- future use

## 28.5 Waiver or Alteration

A waiver or alteration of consent should be determined only through the applicable qualified review process.

## 28.6 Vulnerable or Dependent Participants

Additional safeguards may be required for:

- Children
- workers
- contractors
- students
- people subject to institutional authority
- marginalized groups
- people exposed to retaliation

## 28.7 Expert Interviews

Experts may still be research participants.

Consider:

- Attribution
- confidentiality
- employer relationship
- proprietary information
- quotation review
- power and reputation

## 28.8 Deception

Research involving deception requires strong justification, risk controls, qualified review, and debriefing where appropriate.

## 28.9 Compensation

Compensation should be fair and not improperly coercive.

## 28.10 Participant Safety

Monitor:

- Psychological distress
- professional risk
- privacy
- retaliation
- exposure to harmful content
- security

## 28.11 Withdrawal

Explain:

- Whether participation may stop
- what happens to previously collected data
- limits created by anonymization or publication

## 28.12 Public Data

Public availability does not automatically eliminate ethical concerns.

Consider:

- Reasonable expectations
- sensitivity
- scale
- re-identification
- context collapse
- downstream harm

## 28.13 International Research

Apply local legal and ethical requirements.

Do not treat one country's consent or review framework as universally sufficient.

---

# 29. Privacy and Personal Data

## 29.1 Privacy by Design

Privacy controls should be defined before collection.

## 29.2 Personal-Data Inventory

Identify:

- Direct identifiers
- quasi-identifiers
- sensitive attributes
- behavioral data
- communications
- biometric data
- employment data
- location

- model-interaction logs

## 29.3 Lawful Basis

Determine the lawful basis and institutional authority for collection, processing, transfer, and retention.

## 29.4 De-Identification

De-identification should consider:

- Linkage
- rare attributes
- free text
- timestamps
- model memorization
- external datasets

## 29.5 Pseudonymization

Pseudonymization reduces direct identification but does not necessarily make data anonymous.

## 29.6 Access

Limit personal data to people whose role requires it.

## 29.7 Publication

Avoid unnecessary identifying detail.

Use participant quotations carefully.

## 29.8 Data Subject Rights

Where applicable, support:

- Access
- correction
- deletion
- objection
- restriction
- withdrawal

## 29.9 Breach Response

A breach should trigger:

- Containment
  - assessment
  - required notice
  - participant protection
  - evidence review
  - correction
- 

## 30. Dual-Use and Research Security

### 30.1 Dual-Use Screen

Every C3 or C4 technical project should assess whether the research could materially enable:

- Cyber exploitation
- biological harm
- chemical harm
- evasion
- surveillance
- coercion
- critical-infrastructure attack
- model theft
- dangerous autonomous action

### 30.2 Research Security Screen

Assess:

- Sensitive model access
- task and solution exposure
- threat actors
- insider risk
- partner security
- data transfer
- export or sanctions controls
- facility and device security
- publication risk

## 30.3 Benefit-Risk Review

Consider:

- Scientific benefit
- public benefit
- enabling potential
- availability of the information elsewhere
- required detail
- safeguards
- publication alternatives

## 30.4 Safe Research Design

Possible controls:

- Safe proxies
- sandboxing
- constrained tools
- synthetic data
- staged access
- rate limits
- monitoring
- split knowledge
- isolated environments
- expert supervision

## 30.5 Publication Review

Before release, review whether details could:

- Enable attack
- reveal active vulnerabilities
- compromise held-out tasks
- disclose system defenses
- expose personal data
- undermine national security

## 30.6 Graduated Publication

Possible outputs:

- Full public paper
- redacted paper
- delayed publication
- methods-only paper

- high-level public summary
- restricted technical annex
- no release with internal record

## 30.7 Responsible Disclosure

Coordinate vulnerability or control-failure disclosure with affected parties while preserving independent reporting rights.

## 30.8 Researcher Protection

Provide:

- Clear authorization
- legal and security guidance
- reporting channel
- incident support
- safe-harbor terms where available

## 30.9 Research Security Is Not Censorship

Restrictions should be:

- Specific
- justified
- time-bounded
- reviewed
- no broader than necessary

## 30.10 Security Incident

A research security incident may require:

- Protocol suspension
  - access revocation
  - evidence-status change
  - notification
  - investigation
  - re-evaluation
  - publication correction
-

# 31. Conflicts, Independence, and Funding

## 31.1 Conflict Types

Research conflicts may be:

- Financial
- employment
- ownership
- client
- organizational
- intellectual
- political
- personal
- reputational
- access-related

## 31.2 Disclosure

Material conflicts should be disclosed before:

- Project approval
- reviewer selection
- analysis
- publication

## 31.3 Management Options

Use:

- Public disclosure
- role limitation
- independent analysis
- recusal
- alternate reviewer
- data firewall
- funding diversification
- exclusion

## 31.4 Sponsor Role

A sponsor may define a research need and provide technical information.

The sponsor should not receive hidden control over:

- Method

- inclusion
- analysis
- conclusion
- publication

## 31.5 Publication Rights

Agreements should define:

- Factual review
- confidentiality review
- security review
- timing
- independent conclusion
- right to publish
- dispute process

## 31.6 Result-Dependent Funding

Payment contingent on a favorable result is prohibited.

## 31.7 In-Kind Support

Disclose material:

- Model access
- compute
- personnel
- data
- infrastructure
- travel
- security services

## 31.8 Access Dependence

Dependence on future model access may create a conflict even without direct payment.

## 31.9 Intellectual Commitments

Researchers should disclose when they:

- Designed the evaluated framework
- publicly advocated the conclusion
- own the method
- compete with the subject

## 31.10 Independence Record

C3 and C4 work should include an independence profile.

---

# 32. Research Roles, Contribution, and Authorship

## 32.1 Role Clarity

Projects should identify:

- Research owner
- principal investigator or lead
- protocol author
- data steward
- engineer
- analyst
- statistician
- domain expert
- security reviewer
- ethics reviewer
- independent reviewer
- writer
- publication owner

## 32.2 Contribution Taxonomy

Contribution may include:

- Conceptualization
- methodology
- software
- data curation
- investigation
- validation
- formal analysis
- visualization
- writing
- review
- supervision
- funding
- project administration
- security
- public-interest input

## 32.3 Authorship

Authorship should require substantive contribution and accountability for the work.

## 32.4 No Honorary Authorship

Status, funding, leadership, or access alone does not justify authorship.

## 32.5 Acknowledgment

Recognize contributions that do not meet authorship criteria.

## 32.6 Contributor Statement

Major publications should include a role-based contribution statement.

## 32.7 Responsibility

Every author should understand the central claims and identify which parts they can directly verify.

## 32.8 Disagreement

An author should not be required to endorse claims they materially dispute.

Use:

- Qualified authorship
- separate note
- minority statement
- withdrawal from authorship

## 32.9 Contributor Safety

Protect contributors from retaliation for good-faith:

- Methodological criticism
  - negative findings
  - security reporting
  - correction
  - dissent
-

## 33. Artificial Intelligence Tools in Research

### 33.1 Permitted Uses

AI tools may assist:

- Search-term generation
- source classification
- document extraction
- coding assistance
- transcription
- translation
- statistical programming
- simulation
- drafting
- formatting
- consistency checking

### 33.2 Verification

Material outputs require human or independently testable verification.

### 33.3 Source Rule

Cite the underlying evidence rather than an assistant-generated summary.

### 33.4 Tool Record

Record when material:

- Tool
- provider
- model
- version or date
- prompt or procedure
- data supplied
- output use
- review
- restrictions

### 33.5 Sensitive Data

Do not submit confidential or restricted data to an external AI service without authorization and appropriate controls.

## 33.6 Code Generation

Generated code should receive:

- Review
- tests
- security analysis
- dependency review
- reproducibility check

## 33.7 Translation

Machine translation should receive domain review where wording affects:

- Law
- standards
- capability constructs
- participant consent
- safety

## 33.8 Extraction

Sample automated extraction for:

- Omission
- hallucination
- table corruption
- citation mismatch
- loss of qualification

## 33.9 Model-as-Researcher Claims

Do not list an AI system as an accountable author.

## 33.10 Model Output as Study Data

When model outputs are the object of research, preserve:

- Model identity
- prompts
- system configuration
- date
- sampling
- tools
- outputs
- filtering

## 33.11 Disclosure Threshold

Disclose AI use when it materially affects:

- Method
  - analysis
  - text
  - evidence selection
  - confidentiality
  - reproducibility
  - interpretation
- 

## 34. Review Architecture

### 34.1 Review Types

Research may receive:

- Self-review
- internal peer review
- domain review
- methodological review
- statistical review
- security review
- ethics review
- legal review
- independent expert review
- public review
- artifact review
- replication

### 34.2 Review Fit

Select reviewers based on the actual claim and method.

### 34.3 Reviewer Access

Reviewers should have sufficient access to evaluate the public conclusion.

### 34.4 Reviewer Independence

External is not automatically independent.

Apply FOUNDATION\_04\_INDEPENDENT\_EXPERT\_REVIEW.md.

## 34.5 Review Questions

Reviewers should assess:

- Question
- method
- object identity
- evidence
- analysis
- uncertainty
- contrary evidence
- security
- ethics
- conflict
- conclusion
- public wording

## 34.6 Factual Review

The studied party may identify:

- Factual error
- misunderstood configuration
- confidentiality
- security risk

It should not receive veto over independent interpretation.

## 34.7 Adversarial Review

For C3 and C4 work, appoint a reviewer or team to develop the strongest case that:

- The construct is wrong
- the method is invalid
- the result is overgeneralized
- contrary evidence was omitted
- incentives distort the research
- the recommendation creates harm

## 34.8 Red-Team Review

A methodological red team may attempt to:

- Break the protocol
- manipulate the metric
- exploit exclusions
- reveal task leakage

- create misleading results
- identify unsafe publication

## 34.9 Dissent

Preserve material dissent through:

- Reviewer response
- minority report
- unresolved issue register
- confidence reduction

## 34.10 Review Record

Record:

- Reviewer
- qualification
- conflict
- scope
- access
- findings
- author response
- unresolved issue
- date

## 34.11 Review Completion

A review is complete when material comments are:

- Resolved
- accepted as limitations
- preserved as dissent
- escalated

Not every reviewer must agree.

---

# 35. Analysis, Synthesis, and Interpretation

## 35.1 Analysis Separation

Separate:

- Data preparation
- descriptive analysis

- confirmatory analysis
- exploratory analysis
- sensitivity analysis
- interpretation
- recommendation

## 35.2 Planned and Exploratory Work

Label analyses as:

- Prospectively specified
- amended before relevant results
- exploratory after result access
- post hoc sensitivity analysis

Post hoc analysis may be valuable.

It should not be represented as prospectively confirmed.

## 35.3 Assumption Checks

Document assumptions concerning:

- Independence
- distribution
- missingness
- measurement
- causal structure
- comparability
- system stability
- judge validity

## 35.4 Triangulation

Use multiple evidence forms when one method cannot support the full claim.

Examples:

- Benchmark result plus trajectory review
- policy document plus operational interview
- certification record plus sampled practice evidence
- incident record plus system log
- quantitative comparison plus expert interpretation

## 35.5 Contrary Evidence

Create a contrary-evidence section for C2 through C4 work.

State:

- Evidence
- source
- quality
- effect
- unresolved question

## 35.6 Alternative Explanations

List plausible alternate explanations and indicate:

- Tested
- partly tested
- untested
- ruled out
- unresolved

## 35.7 Evidence Synthesis

Apply EVIDENCE\_STANDARDS.md.

Assign:

- Evidence level
- confidence
- claim limitations
- expiration
- additional evidence needed

## 35.8 Heterogeneity

Do not pool results when:

- Constructs differ
- systems differ materially
- protocols are incompatible
- outcome meanings differ
- time periods are incomparable
- legal contexts are distinct

## 35.9 Negative Results

Report negative and inconclusive results.

Distinguish:

- Evidence of absence
- absence of evidence
- underpowered study
- elicitation failure
- invalid method
- genuine null result

## 35.10 Interpretation Boundary

The conclusion should not exceed:

- Research question
- object
- method
- sample
- environment
- time
- evidence
- authority

## 35.11 Recommendation Boundary

A recommendation should distinguish:

- Technical finding
- institutional judgment
- value choice
- legal authority
- implementation assumption

---

# 36. Uncertainty and Confidence

## 36.1 Uncertainty Register

C2 through C4 projects should maintain an uncertainty register.

Possible categories:

- Measurement
- sampling
- task
- elicitation
- model version

- system configuration
- data quality
- causal
- forecast
- legal
- institutional
- translation
- security
- unknown

## 36.2 Quantification

Quantify uncertainty when the method supports it.

## 36.3 Qualitative Uncertainty

Use structured qualitative descriptions when numerical precision would be misleading.

## 36.4 Confidence

Assign confidence according to EVIDENCE\_STANDARDS.md.

## 36.5 Decision Under Uncertainty

State:

- Decision urgency
- reversible options
- monitoring
- evidence that would change the decision
- cost of false positive
- cost of false negative

## 36.6 Unknown Unknowns

Do not convert incomplete awareness into narrow confidence intervals.

Use stress tests, diverse review, and monitoring.

## 36.7 Confidence Decay

Confidence should be reconsidered after:

- Model change
- new evidence

- failed replication
  - incident
  - legal change
  - task compromise
  - evaluator-status change
- 

## 37. Research Reporting

### 37.1 Minimum Report Structure

A major research report should include:

- Title
- version
- status
- authors and contributors
- date
- authority note
- executive summary
- research question
- rationale
- scope and non-claims
- method
- object identity
- data or sources
- analysis
- findings
- contrary evidence
- uncertainty
- limitations
- evidence level
- confidence
- conflicts
- AI-tool disclosure
- ethics and security
- artifact availability
- review
- conclusion
- recommendations
- references
- revision record

## 37.2 Method Transparency

The method should be detailed enough for a qualified reader to understand:

- What was done
- why
- by whom
- with which evidence
- under which constraints
- how the conclusion followed

## 37.3 Reporting Guidelines

Use relevant reporting guidelines where they improve completeness.

Do not treat checklist completion as proof of validity.

## 37.4 Result Reporting

Report:

- All primary outcomes
- material secondary outcomes
- failed runs
- exclusions
- deviations
- adverse or unexpected findings
- uncertainty

## 37.5 Language

Use TERMINOLOGY.md.

Avoid unsupported:

- Safe
- proven
- official
- certified
- compliant
- global consensus
- experts agree
- best practice

## 37.6 Abstract and Summary

Summaries should preserve material limitations.

## 37.7 Visual Evidence

Figures and tables should include:

- Source
- scale
- units
- uncertainty
- version
- missing data
- comparability limits

## 37.8 Public and Technical Versions

A project may publish:

- Public summary
- full technical report
- confidential annex
- restricted artifact package

The public version should explain what is withheld and why.

## 37.9 Research Status

Use:

- Exploratory note
- research brief
- working paper
- canonical working white paper
- proposed framework
- pilot report
- replication report
- withdrawn
- superseded

## 37.10 Current-As-Of Date

Time-sensitive research should display the date through which evidence was reviewed.

---

## 38. Publication and Dissemination

### 38.1 Publication Decision

Consider:

- Evidence maturity
- public benefit
- security
- participant protection
- legal duty
- correction readiness
- risk of misinterpretation

### 38.2 No Favorable-Result Requirement

Publication should not depend on whether the result supports the preferred hypothesis or sponsor.

### 38.3 Publication Delay

Delay may be justified for:

- Responsible disclosure
- participant protection
- task integrity
- active investigation
- legal review
- coordinated correction

Delay should not become indefinite suppression without review.

### 38.4 Preprints

Preprints should be clearly labeled as not yet peer reviewed where applicable.

### 38.5 Public Comment

Public comment may improve:

- Standards proposals
- taxonomies
- institutional frameworks
- interoperability
- implementation guidance

It does not replace technical validation.

## 38.6 Media Communication

Media materials should:

- Preserve uncertainty
- avoid sensationalism
- distinguish result from implication
- provide access to the full report
- identify current status

## 38.7 Stakeholder Briefing

Affected parties may receive advance briefing for:

- Factual accuracy
- safety
- remediation
- implementation

They should not control independent conclusions.

## 38.8 Accessibility

Public reports should support:

- Plain-language summary
- accessible formatting
- machine-readable metadata
- translation where material
- stable links

---

# 39. Correction, Supersession, and Retirement

## 39.1 Correction Triggers

Correct when:

- A fact is wrong
- analysis is wrong
- citation is mismatched
- material evidence was omitted
- system identity is incomplete
- protocol deviation was undisclosed
- legal or standards status changed
- public wording exceeds evidence

## 39.2 Correction Classification

### Minor

No material effect on the central conclusion.

### Material

Changes evidence level, confidence, scope, or recommendation.

### Invalidating

The output should no longer support the central claim.

## 39.3 Correction Procedure

1. Receive report.
2. preserve original.
3. assess evidence.
4. classify severity.
5. notify responsible owners.
6. conduct independent review where material.
7. issue correction, supersession, or withdrawal.
8. propagate to dependent files.
9. update `VERSION_HISTORY.md`.
10. review process failure.

## 39.4 No Silent Replacement

Do not silently overwrite a material public error.

## 39.5 Supersession

Use when a newer study or version replaces the prior output without proving it invalid.

## 39.6 Withdrawal

Use when:

- Evidence is unreliable
- security or ethics failure invalidates use
- misconduct is substantiated
- central claims cannot be supported
- object identity was materially wrong

## 39.7 Retirement

Retire methods and findings that are no longer current or useful.

## 39.8 Correction Credit

Good-faith correction should receive institutional credit.

Repeated negligence, concealment, or manipulation should be treated separately.

---

# 40. Research Quality Assurance

## 40.1 Quality Objectives

Research quality includes:

- Validity
- integrity
- traceability
- competence
- independence
- security
- ethics
- reproducibility
- clarity
- correction

## 40.2 Project Quality Plan

C3 and C4 projects should define:

- Quality roles
- review points
- acceptance criteria
- artifact checks
- code review
- evidence audit
- publication approval

## 40.3 Methodological Audit

A methodological audit should sample:

- Protocol compliance
- registration

- deviations
- data provenance
- analysis
- exclusions
- contrary evidence
- conflicts
- AI-tool use
- correction readiness

## 40.4 Reanalysis

Independent reanalysis may be required when:

- Results are consequential
- analysis is complex
- sponsor conflict is material
- anomalies exist
- public controversy is high

## 40.5 Research Integrity

Standards Body adopts research-integrity expectations consistent with honest and verifiable methods, responsible peer review, protection of sensitive information, and accountable reporting. NSF and NIH research-integrity guidance provide relevant institutional reference points, while Standards Body applies its own bounded methodology.[^nsf-integrity][^nih-conduct]

## 40.6 Misconduct

Potential research misconduct should be handled through a fair process.

Distinguish:

- Honest error
- methodological disagreement
- negligence
- falsification
- fabrication
- plagiarism
- evidence suppression
- unauthorized disclosure

## 40.7 Quality Metrics

Possible metrics:

- Protocol completion

- deviation rate
- citation accuracy
- reproducibility rate
- replication rate
- correction time
- reviewer agreement
- unresolved critical issues
- artifact completeness
- source freshness
- negative-result publication

## 40.8 Quality Review

Review the methodology itself using:

- Researcher feedback
- external critique
- incident analysis
- replication
- outcome measurement
- annual audit

---

# 41. Research Governance

## 41.1 Governance Functions

Standards Body should govern:

- Project approval
- consequence classification
- protocol review
- ethics and security escalation
- evidence standards
- reviewer selection
- publication
- correction
- research records
- methodology updates

## 41.2 Core Roles

### Research Director or Methodology Owner

Maintains the research system.

**Project Lead**

Owns the project.

**Method Reviewer**

Reviews methodological fit.

**Domain Reviewer**

Reviews subject-matter validity.

**Evidence Reviewer**

Reviews sourcing and claims.

**Security Reviewer**

Reviews sensitive research.

**Ethics Reviewer**

Reviews participant and public-interest issues.

**Independent Reviewer**

Challenges consequential work.

**Publication Authority**

Approves release under current governance.

**Records Custodian**

Maintains protocols, evidence, and versions.

**41.3 Approval Matrix****C0 and C1**

Project lead plus self or peer check.

**C2**

Project lead, domain or method review, evidence check.

**C3**

Protocol review, domain and method review, security or ethics review where applicable, independent review.

## **C4**

Multi-party governance review, independent expert review, security and ethics approval, publication decision, monitoring and appeal.

### **41.4 Recusal**

Decision participants should recuse for material conflicts.

### **41.5 Dissent**

Material dissent should be recorded.

### **41.6 Appeals**

Appeals may concern:

- Method
- evidence exclusion
- security restriction
- authorship
- publication
- correction
- classification

### **41.7 Emergency Research**

Emergency work may use an accelerated protocol.

It should still record:

- Question
  - source
  - method
  - uncertainty
  - security
  - approval
  - expiration
  - later full review
-

## 42. Research Maturity Model

### Level 0: Informal

Characteristics:

- Unrecorded questions
- ad hoc sources
- no protocol
- no review
- conclusions based on intuition

### Level 1: Documented

Characteristics:

- Question and scope
- source list
- method description
- named owner
- basic review
- status and date

### Level 2: Protocol-Governed

Characteristics:

- Prospective protocol
- project classification
- evidence standard
- data plan
- deviation log
- uncertainty
- artifact inventory

### Level 3: Independently Challenged

Characteristics:

- Domain and method review
- contrary-evidence search
- reproducibility statement
- conflict disclosure
- independent review for consequential claims
- visible corrections

## Level 4: Decision-Grade

Characteristics:

- C3 and C4 governance
- preregistration or equivalent
- secure evidence
- replication or reperformance
- decision linkage
- monitoring
- appeal
- expiration

## Level 5: Adaptive Research Institution

Characteristics:

- Living reviews
- continuous quality audit
- replication program
- public correction record
- machine-readable provenance
- research-method experiments
- measured institutional outcomes
- cross-border interoperability

## Maturity Rule

Research maturity should be assessed project by project.

An institution should not claim high maturity because one flagship study is rigorous.

---

## 43. Implementation Plan

### Phase 1: Methodology Adoption

Apply this file to all new substantial projects.

### Phase 2: Research Intake

Create a project registry using the intake template.

## Phase 3: Protocol Library

Create controlled protocol templates for:

- Literature review
- technical evaluation
- case study
- interview research
- forecasting
- institutional pilot

## Phase 4: Source and Evidence Integration

Link each project to:

- SOURCES .md
- claim register
- evidence levels
- confidence ratings

## Phase 5: Research Artifact Repository

Establish:

- Version control
- metadata
- access classes
- retention
- reproducibility packages

## Phase 6: Review Network

Develop qualified pools for:

- Evaluation science
- statistics
- cyber
- biology
- governance
- standards
- law
- security
- ethics
- public interest

## Phase 7: Replication Program

Select high-value claims for:

- Reproduction
- replication
- bridge study
- external review

## Phase 8: Methodological Audit

Audit a representative sample of canonical work.

## Phase 9: Public Research Register

Publish appropriate:

- Project title
- question
- status
- protocol
- evidence level
- review
- publication
- correction

## Phase 10: Method Evolution

Revise this methodology based on:

- Use
- failure
- incident
- external critique
- research outcomes
- international practice

---

## 44. Research Methodology Scorecard

| Dimension       | Core question                                                    |
|-----------------|------------------------------------------------------------------|
| Question        | Is the research question clear, bounded, and answerable?         |
| Decision link   | Is the intended decision or knowledge gap explicit?              |
| Object identity | Is the model, system, institution, or case precisely identified? |

| Dimension              | Core question                                                           |
|------------------------|-------------------------------------------------------------------------|
| Method fit             | Does the method match the question?                                     |
| Classification         | Are consequence, sensitivity, review, and registration levels assigned? |
| Protocol               | Was the work prospectively specified at an appropriate level?           |
| Registration           | Are planned and exploratory work distinguishable?                       |
| Sampling               | Is the sample justified and relevant?                                   |
| Sources                | Are authoritative and contrary sources included?                        |
| Data                   | Are data quality, provenance, and transformation controlled?            |
| Artifacts              | Are code, prompts, tasks, logs, and instruments preserved?              |
| Analysis               | Are assumptions, exclusions, and uncertainty handled?                   |
| Qualitative rigor      | Are interpretation, coding, reflexivity, and negative cases addressed?  |
| Statistical rigor      | Are dependence, power, multiplicity, and sensitivity addressed?         |
| Evaluation validity    | Are construct, elicitation, scoring, and system identity valid?         |
| Ethics                 | Are participant welfare, consent, privacy, and justice addressed?       |
| Security               | Are dual-use, access, disclosure, and incident risks controlled?        |
| Conflicts              | Are funding, access, and intellectual conflicts managed?                |
| AI-tool use            | Is material AI assistance verified and disclosed?                       |
| Reproducibility        | Can the work be reconstructed or re-executed?                           |
| Replicability          | Is independent confirmation planned or available?                       |
| Review                 | Did qualified reviewers have sufficient access and independence?        |
| Contrary evidence      | Was disconfirming evidence actively sought?                             |
| Uncertainty            | Are material uncertainties visible?                                     |
| Reporting              | Does the report preserve method, limitations, and status?               |
| Public claims          | Do public statements remain within evidence?                            |
| Correction             | Can error be corrected and propagated?                                  |
| Monitoring             | Are update and retirement triggers defined?                             |
| Proportionality        | Is the burden appropriate to consequence?                               |
| Institutional learning | Will the project improve future methods and standards?                  |

## 44.1 Critical Failures

The following normally prevent a C3 or C4 project from publication as decision-grade research:

- Unidentified research object
- no protocol or reconstructable method
- material hidden deviations
- selective exclusion of adverse evidence
- unresolved human-participant or legal review

- uncontrolled sensitive-data exposure
- unaddressed dual-use risk
- decisive conflict without mitigation
- no qualified review
- model-generated claims treated as evidence
- material contrary evidence ignored
- unsupported authority or safety claims
- no correction pathway

## 44.2 No Composite Rating

Do not average the scorecard into one universal number.

Critical weaknesses should remain visible.

---

## 45. Protocol Deviation Record Template

Project ID:

Protocol version:

Deviation ID:

Date identified:

**Planned Method**

**Actual Method**

**Timing Relative to Result Access**

**Reason**

**Approval**

**Effect on Validity**

**Effect on Evidence Level or Confidence**

**Corrective Action**

**Publication Disclosure**

---

## 46. Research Review Record Template

Project:

Reviewer:

Role:

Qualifications:

Conflict disclosure:

Access provided:

Date:

**Scope**

**Method Findings**

**Evidence Findings**

**Security and Ethics Findings**

**Contrary Evidence**

**Required Corrections**

**Limitations**

**Recommendation**

- Approve
- approve with conditions
- revise and resubmit
- restrict
- defer
- do not publish
- withdraw

**Unresolved Dissent**

---

## 47. Research Artifact Availability Statement

Project:

Version:

| Artifact      | Status | Location or access process | Restriction reason | License | Review date |
|---------------|--------|----------------------------|--------------------|---------|-------------|
| Protocol      |        |                            |                    |         |             |
| Registration  |        |                            |                    |         |             |
| Data          |        |                            |                    |         |             |
| Code          |        |                            |                    |         |             |
| Prompts       |        |                            |                    |         |             |
| Task bank     |        |                            |                    |         |             |
| Environment   |        |                            |                    |         |             |
| Analysis      |        |                            |                    |         |             |
| Review record |        |                            |                    |         |             |

**Reproducibility status:**

**Replication status:**

**Contact:**

## 48. Publication Readiness Checklist

Before publication, confirm:

1. The question and scope are clear.
2. The research object is precisely identified.
3. The method is documented.
4. Material protocol deviations are disclosed.
5. Primary and contrary evidence were reviewed.
6. Citations support the exact claims.
7. Data and artifacts are preserved.
8. Statistical and qualitative methods were reviewed.
9. Uncertainty and confidence are stated.
10. Human-participant obligations were resolved.
11. Privacy and security review is complete.
12. Dual-use publication risk is addressed.
13. Conflicts and funding are disclosed.
14. Material AI-tool use is documented.
15. Required independent review is complete.
16. Dissent is preserved.
17. Public wording remains within evidence.
18. Artifact availability is stated.
19. Correction and monitoring paths exist.
20. Version and current-as-of date are displayed.

## 49. Canonical Standards Body Research Positions

Standards Body adopts the following working positions.

1. Research should begin with a defined question, object, scope, and intended use.
2. Method selection should follow the question rather than institutional fashion.
3. Planned and exploratory work should remain distinguishable.
4. Exploratory research is legitimate when labeled honestly.
5. Preregistration improves transparency but does not repair invalid design.
6. High-consequence comparative, threshold, and confirmatory studies should ordinarily be preregistered or prospectively frozen.
7. Restricted registration is acceptable when public registration would compromise security or holdout integrity.
8. Research protocols should be versioned and deviations recorded.
9. AI evaluation research should identify the full system, not only the model name.
10. Elicitation conditions are part of the research method.
11. Test-time resources can materially change measured capability and should be reported.
12. Public benchmarks alone are insufficient for many consequential frontier AI claims.
13. Negative results should be published and interpreted carefully.
14. "Not demonstrated" should not be rewritten as "absent" without stronger evidence.
15. Literature reviews should state their search, inclusion, exclusion, and update method.
16. Reporting checklists improve completeness but do not prove validity.
17. Primary sources should anchor legal, standards, institutional, and direct technical claims.
18. Contrary evidence should be sought deliberately.
19. Quantitative and qualitative methods should be chosen according to the claim.
20. Mixed methods should integrate evidence rather than decorate a study.
21. Case studies support contextual and analytical learning but not automatic population estimates.
22. Institutional-design proposals should be piloted and evaluated by outcomes.
23. Technical evidence does not uniquely determine policy.
24. Legal claims require jurisdiction, status, date, and qualified interpretation.

25. International comparisons should not treat one jurisdiction or language as the default.
26. Forecasts should be probabilistic or scenario-based, time-bounded, and updateable.
27. Expert judgment should be structured, conflict-aware, and distinguishable from observed fact.
28. Data and research artifacts should have provenance, versions, access rules, and stewards.
29. Research should be as open as responsible and as restricted as necessary.
30. FAIR research practice does not require unrestricted public release.
31. Reproducibility and replicability are distinct and should be reported separately.
32. Restricted research should still support qualified independent challenge.
33. Human-participant research requires appropriate ethical and legal review.
34. Publicly available personal data can still create ethical and privacy risk.
35. Dual-use and research-security review should begin before data collection.
36. Security restrictions should be specific, proportionate, time-bounded, and reviewable.
37. Sponsor and access relationships should be disclosed.
38. Result-dependent research funding is prohibited.
39. Contributor credit should follow actual contribution and accountability.
40. Honorary authorship is prohibited.
41. AI tools may assist research but cannot serve as unverified independent evidence.
42. Material AI-tool use should be recorded when it affects method, confidentiality, reproducibility, or interpretation.
43. Consequential work should receive domain, methodological, and independent challenge appropriate to its risk.
44. Reviewed parties may correct facts but should not control independent conclusions.
45. Material dissent should remain visible.
46. Research reports should state evidence level, confidence, limitations, and current-as-of date.
47. Public summaries and headlines should not exceed the underlying evidence.
48. Corrections should be visible, timely, and propagated.
49. Obsolete methods and findings should be retired rather than preserved through reputation.

50. Standards Body should continuously evaluate and improve its own research system.

---

## 50. Relationship to Other Canonical Files

### **PROJECT\_IDENTITY.md**

Defines why Standards Body conducts research and limits its present authority.

### **TERMINOLOGY.md**

Defines the controlled vocabulary used throughout research.

### **FOUNDATIONS\_APPENDIX.md**

Locates research within the integrated eight-foundation architecture.

### **EVIDENCE\_STANDARDS.md**

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

### **TAXONOMY.md**

Will classify research objects, methods, evidence, risks, actors, and outputs.

### **EVALUATION\_PHILOSOPHY.md**

Will define the deeper approach to measurement, validity, thresholds, and interpretation.

### **GOVERNANCE\_FRAMEWORK.md**

Will define formal decision rights, recusals, appeals, and oversight.

### **TRANSPARENCY\_FRAMEWORK.md**

Will define disclosure classifications and public reporting.

### **CONTRIBUTOR\_FRAMEWORK.md**

Will govern participation, conduct, contribution, credit, and removal.

## **SOURCES .md**

Will maintain the master research source registry.

## **VERSION\_HISTORY .md**

Will maintain current, superseded, corrected, withdrawn, and retired research records.

---

# **51. Final Research Position**

Standards Body should not be known merely for having strong opinions about frontier AI standards.

It should be known for a research process that makes its conclusions difficult to overstate and possible to challenge.

That requires more than citations.

It requires:

- Questions that can be answered
- objects that can be identified
- methods that fit the question
- protocols written before the conclusion
- honest separation of exploratory and confirmatory work
- evidence that includes inconvenient findings
- data and artifacts that can be inspected
- security that protects without concealing weak reasoning
- expert judgment that remains distinguishable from fact
- technical findings that do not pretend to settle political values
- reviewers with competence, access, and independence
- publication that preserves uncertainty
- correction that protects the integrity of the institution

Frontier AI research will often remain incomplete.

Models will change.

Protocols will become stale.

Access will be uneven.

Some evidence will remain confidential.

Some risks will be difficult to test directly.

The answer is not to lower methodological standards or to pretend that uncertainty has disappeared.

The answer is to build a research system that can learn.

The defining research rule of Standards Body is:

**Ask a bounded question, use a fit method, preserve the evidence, invite challenge, and keep the conclusion revisable.**

---

## References and Research Basis

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

[^nist-tevv]: National Institute of Standards and Technology, **AI Test, Evaluation, Validation and Verification (TEVV)**. <https://www.nist.gov/ai-test-evaluation-validation-and-verification-tevv>

[^nist-itl-2026]: National Institute of Standards and Technology, **Information Technology Laboratory AI Program**, including advancing TEVV and measurement science, updated 2026. <https://www.nist.gov/artificial-intelligence/nist-information-technology-laboratory-itl-ai-program>

[^nasem-repro]: National Academies of Sciences, Engineering, and Medicine, **Reproducibility and Replicability in Science**, 2019. <https://doi.org/10.17226/25303>

[^prisma]: PRISMA, **PRISMA 2020 Statement and Checklist**. <https://www.prisma-statement.org/prisma-2020>

[^prisma-scr]: PRISMA, **PRISMA Extension for Scoping Reviews**. <https://www.prisma-statement.org/scoping>

[^cos-prereg]: Center for Open Science, **Preregistration**. <https://www.cos.io/initiatives/prereg>

[^fair]: GO FAIR, **FAIR Principles**. <https://www.go-fair.org/fair-principles/>

[^unesco-open]: UNESCO, **Recommendation on Open Science**, adopted 2021. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>

[^acm-artifacts]: Association for Computing Machinery, **Artifact Review and Badging and Software and Data Artifacts in the ACM Digital Library**. <https://www.acm.org/publications/artifacts>

[^common-rule]: U.S. Department of Health and Human Services, Office for Human Research Protections, **Federal Policy for the Protection of Human Subjects, Common Rule**. <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>

[^belmont]: U.S. Department of Health and Human Services, Office for Human Research Protections, **The Belmont Report**. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>

[^nsf-integrity]: U.S. National Science Foundation, **Responsible and Ethical Conduct of Research and Research Integrity**. <https://www.nsf.gov/policies/responsible-research-conduct>

[^nih-conduct]: National Institutes of Health, Office of Intramural Research, **Guidelines for the Conduct of Research**, 2026. [https://oir.nih.gov/system/files/media/file/2026-06/guidelines-conduct\\_research.pdf](https://oir.nih.gov/system/files/media/file/2026-06/guidelines-conduct_research.pdf)

[^aisi-early]: UK AI Security Institute, **Early Lessons from Evaluating Frontier AI Systems**, 2024. <https://www.aisi.gov.uk/blog/early-lessons-from-evaluating-frontier-ai-systems>

[^aisi-eval-science]: UK AI Security Institute, **Science of Evaluations** research collection, including structured elicitation, statistical evaluation, transcript analysis, and test-time compute research. <https://www.aisi.gov.uk/category/science-of-evaluations>

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

---

## Revision Record

### Version 1.0

**Date:** July 16, 2026

**Change type:** Complete foundational edition

**Summary:** Establishes the canonical Standards Body research methodology. Defines authority limits, foundational principles, research portfolio and classification, lifecycle, question formation, protocols, preregistration, literature and source review, technical experiments, frontier AI evaluation, statistical and qualitative methods, case studies, institutional design, legal and international research, forecasting, expert judgment, data and artifact management, open science, reproducibility, human-participant research, privacy, dual-use and research security, conflicts, contributions, AI-tool use, review, synthesis, uncertainty, reporting, publication, correction, quality assurance, governance, maturity, implementation, scorecards, operational templates, canonical positions, and research basis.

**Status:** Approved foundational source.