AI ETHICS NETWORK LLC

Research Allocation Synthetic Verification Framework

Legal Attestation, Technical Provenance, Cognitive Defense & Post-Allocation Governance

Working Standard · Version 1.1 · September 2026

Governing principle. RASVF does not ban generative tools. It bans the conversion of unverified synthetic output into disbursed public or private capital.

This standard establishes intake, review, and disbursement controls for research proposals and technical manuscripts submitted to funding bodies that adopt RASVF. Use of generative AI is not, by itself, disqualifying. Undisclosed material synthesis, inability to independently verify or reproduce claimed results, and inability to defend those results under unassisted examination are disqualifying.

1. Scope and Defined Terms

This document applies to primary investigators (PIs), co-investigators, and institutional applicants seeking capital, in-kind compute, or formal endorsement from an adopting body. For purposes of this standard:

- "Generative AI" means a model capable of producing text, code, proofs, figures, or data that could reasonably be presented as original research output.
- "Material synthesis" means AI-generated content on which a funding claim, scientific conclusion, or eligibility representation depends.
- "Independent capacity" means the named human authors can reconstruct, verify, and defend the relevant derivation, experiment, or argument without the generating model present.
- "Adopting body" means any funder, publisher, consortium, or review board that incorporates RASVF into its intake and award instruments.

2. Tier 1 — Legal Attestation and Misrepresentation Safeguards

Prior to technical review, every PI and applicant must execute an Attestation of Authorship and AI Utilization. The attestation is an intake disclosure. Contractual remedies for non-compliance reside in the funding agreement, not solely in the intake form.

2.1 Categorization of Generative AI Usage

Applicants must categorize and document the scope of generative AI use in the submitted work.

Category	Definition and disclosure duty
A — Administrative	Grammar, prose styling, formatting, or translation. Requires basic acknowledgment of tools used.
B — Technical assistance	Boilerplate code, routine plotting, search indexing, or similar assistance, provided all outputs were audited and integrated by human authors. Requires tool disclosure and repository or lab-notebook logging.

C — Material synthesis	AI used to generate core hypotheses, mathematical proofs, non-executed synthetic data, or technical arguments where the applicant lacks independent capacity to verify or reproduce the output. Not eligible for funding unless declared and subjected to mandatory Tier 2 audit and, where triggered, Tier 3 examination.

Safe harbor. Documented human-in-the-loop review, with named reviewers and timestamps, does not automatically place work in Category C. Category C is triggered by dependence on unverified synthetic output, not by the mere presence of an AI tool in the workflow.

2.2 Legal Misrepresentation and Remedies

The funding agreement—not only the intake attestation—shall include the following reserved remedies:

- **Failure to disclose.** Presenting Category C synthetic text or AI-generated output as original human research constitutes a material misrepresentation of fact.
- **Cure window.** Good-faith omissions limited to Category A or B that do not conceal Category C material may be cured within a stated period after notice, without automatic termination.
- **Capital clawback.** Allocations disbursed under fraudulent pretenses are subject to termination and recovery of funds, together with reasonable administrative costs, as specified in the award instrument.
- **Statutory referral.** Where wire communications or payment systems were used to solicit funds under falsified credentials or synthetic research claims, the adopting body may refer the matter to relevant authorities under applicable false-claims and wire-fraud statutes. Referral is reserved power, exercised after internal finding and notice.

Applicants shall be given plain-language notice of these categories before they sign. Nuclear remedies shall not be the only text on the first page of an application portal.

3. Tier 2 — Technical Provenance and Artifact Auditing

Manuscripts and proposals that pass Tier 1 screening undergo automated and manual provenance checks before committee evaluation.

3.1 Version Control and Commit Analytics

For computational, software, or algorithmic submissions, applicants must provide access to a version-controlled repository demonstrating an organic development history.

- Red-flag indicators include single-commit uploads of monolithic codebases, absence of commit-message variance, and commit timestamps that do not align with claimed research timelines.
- A red flag triggers audit; it is not, by itself, a finding of fraud.

Theoretical, formal, or pure-mathematical submissions with no executable artifact are exempt from repository requirements and shall instead supply derivation notebooks, proof sketches, or equivalent work product sufficient to establish human authorship over time.

3.2 Containerized Execution and Replication Gates

- Where the submission claims experimental or computational results, applicants must provide a containerized or pinned environment (for example, a Dockerfile or a fixed-dependency notebook) capable of running the scripts that produce the figures or tables in the manuscript.
- Reviewing bodies will run smoke tests in that environment. Failure to execute, or material variance from published results, invalidates the computational claims unless the applicant supplies a documented, reproducible explanation.
- Proprietary, dual-use, or export-controlled code may be audited under a restricted-access protocol rather than public container deposit. The adopting body shall state who funds audit infrastructure.

3.3 Raw Data and Compute Auditing

- Submissions relying on empirical data or model training must furnish raw or minimally processed datasets, logs, or execution traces, subject to lawful privacy and export constraints.
- For machine-learning models, applicants must disclose compute infrastructure used (hardware class, approximate GPU- or TPU-hours, and provider logs where available).
- Claims of training complex models without corresponding compute evidence constitute a failure of Tier 2 verification.

4. Tier 3 — Cognitive Defense and Real-Time Technical Examination

Tier 3 exists to separate genuine subject-matter command from fluent recitation of synthetic manuscripts. It is value-triggered, not universal.

4.1 Trigger Conditions

Unassisted technical defense is required when any of the following apply:

- Requested award or in-kind value exceeds the adopting body’s published Tier 3 threshold;
- Tier 2 produced an unresolved red flag; or
- The applicant declared Category C use and seeks eligibility despite that declaration.

Small awards, first-time independent researchers, and submissions that cleared Tier 2 without flags are not required to sit a full panel defense. Adopting bodies shall publish the threshold and may offer a shorter written or recorded alternative for accessibility.

4.2 Unassisted Whiteboard Derivation

When triggered, the candidate completes a 45-to-60-minute defense before an independent panel of subject-matter experts. Slide decks, notes, and digital prompts are not permitted. The candidate must derive core equations, sketch system architecture, and explain algorithmic logic on a physical or digital whiteboard in real time.

Reasonable accommodations (extended time, written supplement, captioned remote session) shall be available on the same terms the adopting body uses for other high-stakes evaluations. Accommodation is not evidence of deficient expertise.

4.3 Dynamic Perturbation Testing

Reviewers may introduce real-time parameter changes to the proposed model—for example, altering a foundational assumption, changing a loss function, or modifying an input variable.

Candidates with genuine command adapt and explain downstream effects. Candidates dependent on an unrehearsed synthetic manuscript typically cannot reconcile the change. Failure is a finding about the submission under review, not a lifetime ban, and is subject to the appeals process in Section 6.

5. Post-Allocation Governance and Threat Prevention

5.1 Milestone-Gated Tranche Disbursement

Capital shall not be released in an unconditioned lump sum. Award instruments adopting this framework shall structure release across predefined, verifiable technical gates, for example:

- Gate 1 — Execution replication of the core experiment or proof artifact;
- Gate 2 — Benchmarked performance against the claims in the proposal;
- Gate 3 — Prototype, preprint, or other contracted deliverable.

Gates may be tailored to theoretical work that has no prototype. The principle is verification before the next tranche, not uniformity of artifact type.

5.2 Cross-Institutional Incident Registry

Adopting bodies agree to contribute to a shared, access-controlled registry of verified incidents involving:

- Fabricated academic credentials or identity fraud;
- Materially deceptive synthetic submissions;
- Retaliatory conduct, cyber threats, doxxing, or targeted harassment directed at reviewers or organizations following proposal rejection.

Registry entries require: defined retention periods; role-based access; notice to the subject when an adverse entry is created; and a correction or appeal path. Contribution is limited to verified incidents, not mere allegations. Adopting bodies remain responsible for their own defamation, privacy, and data-protection compliance.

6. Appeals, Exemptions, and Operational Responsibilities

- An applicant who fails Tier 2 or Tier 3 may appeal in writing within the period stated in the award or review policy. The appeal is heard by reviewers who did not make the original finding.
- The adopting body shall publish who pays for container and smoke-test infrastructure, how proprietary or dual-use code is handled, and the current Tier 3 dollar or risk threshold.
- Nothing in this standard treats neurodivergence, non-native English fluency, or unconventional research style as a proxy for synthetic authorship.

7. Adoption Blueprint

1. **Phase 1 (Day 1–30) — Legal integration.** Place the Tier 1 attestation and a plain-language category guide in application portals. Place clawback, cure, and referral clauses in the funding agreement.
2. **Phase 2 (Day 31–60) — Technical gates.** Require repository links or equivalent work product, and containerized or pinned runtimes where computational claims are made. Stand up internal or outsourced audit environments and state who funds them.
3. **Phase 3 (Day 61–90) — Panel capacity.** Train panels on perturbation testing and unassisted defense for triggered cases only. Publish the threshold, accommodations policy, and appeals path before the first live session.

8. Document Control

This working standard is issued by AI Ethics Network LLC for adoption, adaptation, or comment by funding organizations. It is not legal advice. Adopting bodies should have counsel review award instruments, registry protocols, and statutory-referral language before use.

Classification: Working standard, Version 1.1, September 2026.

Contact for adoption inquiries: AI Ethics Network LLC, Bellevue / Redmond, Washington.