ODA3 INSTITUTE

TECHNICAL REPORT

From Technical Evidence to Bounded Assurance

Principles for Independent AI Assurance Research

ODA3-2026-09-TCR-RES-001 | PUBLIC

Publication Metadata

Document ID	ODA3-2026-09-TCR-RES-001
Document Type	Technical Report (TCR)
Category	Research & Benchmarking (RES)
Classification	Public
Framework Tag	Not Applicable — cross-cutting independent AI assurance methodology; developed through the External Assurance Research Programme
Regulatory Crosswalk	Not Applicable — no regulatory mapping or conformity claim is asserted
Evidence Tier	Not Applicable — methodology/doctrine publication; no empirical external finding is asserted
Incident Provenance	Not Applicable
Audience	CISOs; Security Architects; AI Governance Leads; Compliance Officers; Standards Participants; assurance practitioners and researchers
Publish Date	12 September 2026
Review Cycle	Annual, or upon material methodology change, whichever occurs first
Companion Publication	ODA3-2026-09-EXB-RES-001 — Executive Brief
Author	ODA3 Institute

> Publication boundary: this document presents public-facing doctrine and concepts. Internal adjudication logic, evidence-sufficiency thresholds, examiner workflow, and proprietary verification procedures are not disclosed.

Purpose and scope

This report addresses a recurring problem in AI assurance: technical evidence is often asked to support claims broader than the evidence itself can justify. A successful test, a reproduced result, a technical review, or a runtime record can each be valuable. None of them, by themselves, determine the full assurance conclusion that may legitimately be drawn.

The report therefore focuses on a bounded question: **what is the available evidence actually sufficient to support, for a defined claim, evidence base, system or artifact boundary, and set of stated limitations?**

The aim is not to provide a procedural manual for conducting an ODA3 examination. It is to clarify the conceptual distinctions that help prevent testing, validation, reproduction, technical review, or vendor-generated evidence from being promoted into broader assurance claims without an adequate evidentiary basis.

1. Executive Summary

AI governance increasingly depends on technical evidence. Organizations test controls, inspect architectures, reproduce findings, collect logs, preserve runtime receipts, validate configurations, and review system behavior. These activities are necessary. They are also easy to overinterpret.

A test that passes may establish that a defined condition produced an expected result. It does not automatically establish that the relevant control is effective across other conditions, that the wider system is safe, or that an assurance claim has been independently supported. A failure that is no longer reproduced may show that one previously observed condition has changed; it does not automatically establish complete remediation. A runtime receipt may show what a system recorded; it does not automatically establish that the action was authorized, that the evidence is complete, or that the broader governance requirement was satisfied.

This report refers to that problem as **evidence inflation**: the expansion of a valid technical observation into a broader conclusion than the evidence, boundary, provenance, or verification basis supports.

Independent assurance is concerned with preventing that expansion. Its central question is not merely whether evidence exists, or whether a test succeeded. It is:

> **What is the available evidence sufficient to support, against a defined claim, within a defined boundary, under stated assumptions and limitations?**

That question requires a distinction between related but non-interchangeable activities. Testing, reproduction, validation or verification, technical review, and independent assurance can examine the same system or artifact while answering different questions. The difference is not a hierarchy of importance. It is a difference in purpose, evidentiary burden, scope, and the conclusion that may legitimately follow.

For the purposes of this report, the public conceptual relationship is expressed as:

Claim → Evidence Base → Verification → Finding → Limitation → Bounded Assurance

The chain is deliberately simple. It does not describe ODA3 Institute's full examiner workflow. Its purpose is to make one principle explicit: technical evidence becomes assurance evidence only in relation to a particular claim, a defined evidence base, an independent verification basis, an explicit finding, and the limitations that bound that finding.

The report also distinguishes two separate axes that are often conflated. **Evidence status** describes the standing of information used in an examination: Verified, Reproduced, Reported, Inferred, or Not-established. **Finding state** describes the bounded conclusion reached by an examination: Supported, Partial, Gap, or Unknown. A finding can rely on multiple evidence statuses at once; the two vocabularies therefore should not be merged.

The broader strategic distinction is straightforward:

> **Risk management generates and governs controls and evidence; independent assurance establishes what that evidence is sufficient to support.**

This does not diminish governance, testing, validation, internal audit, or technical review. It defines the separate assurance question that remains after those activities have generated evidence.

2. The Problem: Evidence Inflation

Technical evidence is rarely useless. The problem arises when a valid observation is asked to carry more meaning than it can sustain.

Consider four common transitions:

- **“The test passed” → “the control is effective.”**
- **“The original failure was not reproduced” → “the issue is remediated.”**
- **“The system generated a receipt” → “the execution was authorized.”**
- **“The artifact was reproduced” → “the architecture has been independently validated.”**

Each first statement may be true. The corresponding second statement may also eventually prove true. But the second does not follow automatically from the first.

A passing test establishes something about the tested condition. Whether it establishes broader control effectiveness depends on the claim under examination, the conditions represented by the test, the relevant operating boundary, the completeness and provenance of the evidence, and what remains untested. Similarly, reproducing a result can establish reproducibility under stated conditions without establishing general system behavior, historical operation, or the validity of every architectural claim associated with the result.

This distinction becomes especially important for AI systems because system behavior can depend on changing models, runtime context, external tools, policy state, data, human intervention, configuration, and temporal conditions. Evidence that was sufficient for one system state or operating window may not retain the same operational meaning after those conditions change.

2.1 Evidence can remain valid while its applicability changes

A record may remain authentic evidence of what was true at an earlier point in time without remaining sufficient evidence of what is true now. A prior authorization record, validation result, configuration snapshot, model evaluation, or control test may retain historical evidentiary value while losing current operational applicability.

This is a critical distinction. Invalidating the broader inference does not necessarily invalidate the evidence itself. The correct response is often to narrow the proposition the evidence can support.

2.2 Successful technical work can still support only a narrow conclusion

Technical teams often do exactly what they were asked to do: reproduce a defect, verify a change, inspect a code path, validate a configuration, or demonstrate that a runtime mechanism produces an expected record. Independent assurance does not devalue those activities by asking a further question. It preserves their meaning by refusing to attach a stronger conclusion than they warrant.

The governing principle is therefore:

> **Evidence can be valid for one proposition while remaining insufficient for a broader proposition.**

Evidence inflation occurs when this boundary is lost.

2.3 Why the distinction matters commercially and operationally

Evidence inflation can affect procurement, governance, risk reporting, incident response, compliance, and executive decision-making. Buyers may believe a technical test establishes a system-wide property. Governance teams may rely on a control attestation whose evidence covers only a narrower implementation condition. Product teams may describe a reproduced fix as complete remediation when

relevant boundaries remain unexamined. Assurance language can then become stronger than the underlying evidence.

The issue is not primarily one of terminology. It is one of decision quality. If the conclusion is broader than the evidence, decision-makers may rely on a property that has not actually been established.

Bounded assurance is intended to prevent that mismatch.

3. Related but Distinct Activities

Testing, reproduction, validation or verification, technical review, and independent assurance can all contribute to confidence in an AI system. They are related, but they are not interchangeable.

The distinctions below are intentionally high-level. They are not intended to replace discipline-specific standards, professional definitions, or formal conformity-assessment terminology. Their purpose is to clarify the different questions these activities typically answer in an evidence-to-assurance context.

3.1 Testing

Testing asks whether a defined system, mechanism, control, or artifact produces an expected or observed result under stated conditions.

Testing can be behavioral, functional, adversarial, structural, performance-oriented, or otherwise specialized. Its evidentiary strength depends on what was tested, under which conditions, with which inputs, and against which expected outcome.

A successful test can be strong evidence. It does not by itself define the scope of the assurance conclusion.

3.2 Reproduction

Reproduction asks whether an observation, result, or behavior can be independently recreated from the stated artifacts and conditions.

Reproduction is especially valuable where a technical claim depends on a specific configuration, dataset, execution path, version, benchmark, or observable outcome. It strengthens confidence that the reported result is not merely anecdotal or dependent on an undisclosed condition.

Reproduction nevertheless remains bounded by what was reproduced. Recreating a result does not automatically establish the completeness of the source artifact, the generality of the result, the historical state of the system, or the broader effectiveness of the surrounding control environment.

3.3 Validation and verification

Validation or verification asks whether an implementation, artifact, mechanism, or result satisfies the particular criteria being evaluated.

The exact meaning of these terms varies across engineering, software, safety, standards, and conformity-assessment contexts. In this report, they are treated as evidence-generating or evidence-evaluating activities directed toward a defined criterion or intended use.

Their outputs can be highly relevant to assurance. They do not automatically determine what broader assurance claim should be made from those outputs.

3.4 Technical review

Technical review asks what inspection of the relevant architecture, implementation, configuration, artifact, or evidence reveals about the technical proposition under consideration.

Technical review can expose assumptions, inconsistencies, design limitations, missing dependencies, alternative explanations, or mismatches between stated behavior and observable implementation. It may also establish that a design is internally coherent or that a particular mechanism is present.

The reviewer's technical conclusion and the assurance conclusion remain distinct. A technically plausible architecture is not necessarily evidence that the architecture operated as intended in the relevant environment or period.

3.5 Independent assurance

Independent assurance asks what bounded conclusion is justified by the available evidence against a defined claim or requirement.

The emphasis is on the legitimacy of the conclusion. Assurance considers not only whether evidence exists but what that evidence establishes, within what boundary, under what assumptions, with what limitations, and with what degree of independence from the party making the underlying claim.

This makes independent assurance different from a simple accumulation of technical artifacts. More evidence does not automatically produce a broader finding. The evidentiary relationship between the claim and the available records remains central.

3.6 Same evidence, different questions

The same artifact can legitimately participate in several activities while supporting different conclusions.

A runtime log may be used in testing to confirm an expected event. It may be used in reproduction to recreate an observed sequence. It may contribute to validation of a defined control criterion. A technical reviewer may inspect its structure and provenance. An assurance examination may then ask what claim that log, together with the rest of the evidence base, is sufficient to support.

The distinction is therefore not:

testing < validation < assurance

It is:

> **same evidence, different questions, different conclusion boundaries.**

That distinction is essential to preventing a successful technical activity from being silently promoted into a broader assurance claim.

4. From Technical Evidence to a Bounded Finding

A bounded assurance conclusion does not begin with a conclusion. It begins with a claim whose evidentiary basis can be examined.

For public-facing purposes, ODA3 Institute represents that relationship as:

Claim → Evidence Base → Verification → Finding → Limitation → Bounded Assurance

This is a conceptual model, not an examiner procedure. It does not disclose the internal mechanics by which ODA3 determines evidence sufficiency, resolves conflicting evidence, evaluates alternative explanations, or reaches a particular finding state.

4.1 Claim

The **claim** is the proposition being examined. It may concern the operation of a control, the state of an implementation, the result of a remediation, the applicability of an authorization, the behavior of a system under defined conditions, or another bounded technical or governance proposition.

A claim needs enough specificity that the resulting evidence can be evaluated against it. Broad statements such as “the system is secure,” “the agent is governed,” or “the control works” are usually too expansive to function as useful assurance propositions without further qualification.

4.2 Evidence Base

The **evidence base** is the body of information available for examination. It may include technical artifacts, records, logs, configurations, test outputs, source declarations, documentation, independent observations, or other relevant material.

The existence of an artifact is not equivalent to its sufficiency. An evidence base can be authentic yet incomplete, technically correct yet irrelevant to the claim, or highly detailed while leaving a material proposition unestablished.

4.3 Verification

Verification concerns the independent checking of relevant evidence and propositions within the declared boundary. It may establish that an artifact exists, that a result can be reproduced, that a stated condition is observable, or that a source supports the proposition attributed to it.

Verification can change the evidence status of relevant material by independently establishing propositions about it. It does not, by itself, eliminate the need to determine what conclusion that material can legitimately support.

4.4 Finding

A **finding** states what the examination establishes about the bounded claim on the basis of the available evidence.

The finding should not silently absorb propositions that were not examined. It should remain tied to the evidence, conditions, and boundary that gave rise to it.

In this report, findings are described using four public states — Supported, Partial, Gap, and Unknown — discussed separately in Section 5. The internal logic used to reach those states is outside the scope of this publication.

4.5 Limitation

A **limitation** identifies what constrains the finding or remains outside it. Limitations may concern scope, unavailable evidence, untested conditions, temporal applicability, dependencies, assumptions, inaccessible system states, or other matters that affect the reach of the conclusion.

Limitations are not disclaimers added after the substantive work. They are part of the evidentiary meaning of the finding itself.

4.6 Bounded Assurance

Bounded assurance is the conclusion that remains after the claim, evidence base, verification basis, finding, and limitations are considered together.

The word *bounded* is deliberate. It recognizes that a defensible conclusion is often narrower than the system, product, framework, or organization in which the examined proposition sits.

A finding that one control operated under defined conditions does not automatically assure the wider system. A successful reproduction does not automatically assure historical operation. Evidence of an authorization decision does not automatically establish that every consequence path respected that decision. A bounded conclusion preserves these distinctions rather than smoothing them away.

The aim is therefore not to make the strongest claim available from the evidence. It is to make the strongest claim the evidence can legitimately sustain — and no stronger.

5. Two Separate Evidence Axes

A bounded finding is easy to misread as a single measurement. It is not. Two distinct questions are involved, and collapsing them into one scale loses the precision that makes a finding trustworthy.

The first question is: **what is the standing of the evidence itself?** The second is: **what does the examination conclude about the bounded claim?** These are different axes. A given finding can — and often does — rest on evidence of several different standings at once.

5.1 Evidence status

Evidence status describes how a specific piece of information was established, independent of what it is ultimately used to conclude.

Verified — independently checked directly against an authoritative source or artifact.

Reproduced — independently recreated under stated conditions, rather than merely observed once.

Reported — asserted or supplied by a source but not independently established by the examiner.

Inferred — analytically derived from available evidence rather than directly observed.

Not-established — insufficient evidence exists to support the proposition within the examined boundary.

These labels describe the standing of a piece of evidence. They do not, by themselves, say what the examination concludes.

5.2 Finding state

Finding state describes the outcome of the bounded examination — what the available evidence, taken together, is judged sufficient to support.

Supported — available evidence supports the bounded proposition within the defined scope and limitations.

Partial — material elements of the proposition are supported, but one or more relevant elements remain limited, unsupported, or unresolved.

Gap — the examination identifies a material deficiency relative to the bounded proposition.

Unknown — available evidence is insufficient to determine the proposition, in either direction.

This report does not disclose how evidence of a given status, or a mix of statuses, is weighed or sorted into a particular finding state. That determination is part of ODA3's internal examination process and is outside the scope of this publication.

5.3 Why the two axes must stay separate

Treating evidence status and finding state as one taxonomy would suggest, incorrectly, that a finding's strength is simply the strength of its best or worst piece of evidence. In practice, a single bounded finding commonly draws on evidence of several different standings at once.

Consider, in the abstract: a proposition might rest on one element that is Verified, a second that is Reproduced, a third that is only Reported, and a fourth that remains Not-established. The resulting finding — say, Partial — reflects the proposition as a whole, not any single piece of evidence within it. Collapsing "Verified" and "Supported" into the same concept, or treating "Not-established" evidence as automatically producing a "Gap" finding, would misrepresent how bounded conclusions are actually reached.

Keeping the two axes distinct also protects against a subtler failure: treating a finding as more authoritative than its evidence base warrants, simply because some of the underlying material was independently verified. A Partial finding built partly on Verified evidence is still Partial. The evidence axis describes rigor of process; the finding axis describes the reach of the conclusion. Neither substitutes for the other.

6. Independence and Evidentiary Boundaries

Independent assurance depends not only on the quality of technical evidence but also on preserving the boundary between the party making or supplying a claim and the party determining what the evidence is sufficient to support.

That boundary does not require distance from technical detail. On the contrary, credible assurance may require close examination of artifacts, explanations, test results, source declarations, and implementation context. The essential distinction is that access to information must not become control over the conclusion.

This section states the public principles that follow from that distinction. It does not describe ODA3 Institute's internal independence tests, review controls, or examiner procedures.

6.1 Participation does not imply a favorable conclusion

Participation in an assurance examination means that a claim, artifact, control, or evidence base is being examined. It does not mean that the claim has been accepted, that the evidence will prove sufficient, or that the resulting finding will be favorable.

This principle matters because technical collaboration can easily be mistaken for validation. A system owner may provide extensive artifacts, answer technical questions, reproduce a result, or assist with configuration. Those activities can improve the quality of the evidence available for examination without determining the finding that follows.

The evidentiary question remains independent:

> **What does the available evidence actually establish within the declared boundary?**

A credible process must allow the answer to be Supported, Partial, Gap, or Unknown.

6.2 Providing evidence does not determine what the evidence establishes

The party closest to a system is often best placed to explain its architecture, intended behavior, configuration, operating assumptions, and technical artifacts. Those explanations can be important evidence.

They are not the assurance conclusion.

A provider may accurately state that a mechanism exists, that a control was designed for a particular purpose, that a test passed, or that a runtime record was generated. Independent assurance asks the separate question of what those materials, taken together and with their respective evidence statuses, are sufficient to support.

This protects both sides. It prevents the examiner from ignoring relevant technical context, while also preventing the system owner's interpretation from becoming the finding by default.

6.3 Successful reproduction does not prejudge the finding

Reproduction is especially powerful because it turns a reported observation into an independently recreated one under stated conditions. But reproduction remains evidence of the reproduced proposition.

A successfully reproduced result may strengthen confidence that a behavior, artifact, or mechanism is observable and repeatable. It does not automatically establish the completeness of the control, the absence of alternate paths, historical operation, broader system effectiveness, or any proposition that was not part of the reproduced condition.

The finding therefore follows the evidence boundary rather than the emotional force of a successful reproduction.

> **Reproduction can strengthen the evidentiary basis without expanding the claim being examined.**

6.4 Factual review is not editorial control

Technical accuracy benefits from allowing relevant parties to identify objective errors concerning matters such as artifact identity, version, configuration, dates, implementation facts, or reproduction conditions.

That factual review should remain distinct from editorial control over the assurance conclusion.

A system owner or contributor may reasonably correct an inaccurate description of what a file contains or which version was examined. That does not imply a right to approve the finding, remove an unfavorable limitation, redefine the examined claim after the fact, or negotiate the conclusion into a more favorable form.

The distinction is straightforward:

factual correction improves the accuracy of the record; editorial control changes the independence of the conclusion.

Independent assurance requires the former without conceding the latter.

6.5 New evidence should not silently rewrite the historical record

Assurance findings are tied to the evidence base and boundary that existed when the examination was completed.

If material new evidence later becomes available, that evidence may justify clarification, supplementation, or a separately identified examination. It should not silently alter what the earlier evidence established at the time.

This preserves temporal integrity. A later artifact can improve the evidentiary position prospectively without making the earlier record retrospectively stronger than it actually was.

Accordingly:

> **New evidence may change a future finding; it does not erase the evidentiary history of an earlier one.**

This is particularly important for AI systems whose models, policies, configurations, data sources, dependencies, and runtime controls can change over time.

6.6 Findings apply only within their declared boundary

A bounded finding should be read as a statement about the proposition actually examined, not as a general endorsement of the system, product, organization, or framework in which that proposition sits.

A finding concerning one control does not automatically extend to adjacent controls. A finding concerning one version does not automatically extend to a later version. A finding concerning one execution state does not automatically extend to every operating state. A finding concerning one evidence base does not automatically establish claims for which that evidence was never examined.

This is not an artificial narrowing of assurance. It is what makes the conclusion defensible.

The appropriate question is therefore not:

“Was the system assured?”

but:

> **“What proposition was examined, against which evidence base, within which boundary, and what did the resulting finding establish?”**

6.7 Independence as a property of the conclusion

Independence should not be confused with isolation.

An examiner may need detailed technical engagement with the party supplying the evidence. The system owner may be the only source for certain configuration records, implementation facts, or explanations of intended control behavior. External technical contributors may also provide essential context.

The independence question is whether those participants can determine the conclusion merely by supplying the evidence or advocating an interpretation of it.

For the purposes of bounded assurance, the critical separation is:

evidence contribution ≠ conclusion authority

That separation allows technical depth and collaborative fact-finding without converting the examination into endorsement.

6.8 Public principle

Taken together, these boundaries support a simple principle:

> **The subject of an assurance examination may contribute evidence, context, and factual correction; it should not control what the independent examination concludes that evidence is sufficient to support.**

That principle is one of the conditions that keeps technical evidence from becoming assurance by assertion.

7. Synthetic Illustration 1 — Historical Authority vs. Current Authority

The scenario below is a fully synthetic illustration. It does not describe any organization, system, submission, or examination conducted by ODA3 Institute.

An automated system was authorized, at an earlier point in time (T1), to initiate a defined class of operation. A record exists confirming that authorization — for example, a signed approval, a configuration entry, or an issued credential dated to T1.

At a later point (T2), a consequential instance of that same class of operation occurs. The system again proceeds, citing the same class of authorization.

A natural but imprecise question is:

"Was the operation authorized?"

The bounded assurance question is different:

> **What evidence establishes that the authorization was current and applicable at T2 — not merely that it existed at T1?**

Evidence of the T1 approval is real evidence. It may be Verified, in the sense that the record itself can be independently confirmed to exist and to say what it says. But that verified record answers a narrower proposition than the one usually assumed. It establishes that authorization existed at T1. It does not, by itself, establish that the authorization remained valid, unrevoked, and applicable at T2 — a separate proposition requiring separate evidence, such as a currency check, a revocation record, or a standing confirmation contemporaneous with the T2 event.

Absent that further evidence, the T2 proposition should remain unresolved or only partially established, even though the T1 evidence itself may be strong. Treating durable historical evidence as if it answered a live, time-sensitive question is a common source of evidence inflation: a true statement about the past is quietly substituted for a claim about the present.

This illustration demonstrates a general principle, not a specific examination technique: **authorization is frequently time-bound, and evidence of past standing does not automatically establish current standing.** The same reasoning applies wherever a system relies on a credential, permission, certification, or approval that can lapse, be revoked, or expire between the time it was granted and the time it is relied upon.

8. Synthetic Illustration 2 — Retest vs. Remediation

The scenario below is a fully synthetic illustration. It does not describe any organization, system, submission, or examination conducted by ODA3 Institute.

A technical defect previously allowed a system to bypass an intended enforcement outcome under a specific condition. A subsequent test, run after a described corrective change, no longer reproduces that specific failure.

A natural but imprecise conclusion is:

"The control is now effective."

The bounded assurance question is narrower:

> **What does the absence of the previously observed failure actually establish, and what does it leave open?**

A test that no longer reproduces a known failure is meaningful evidence — it can support the proposition that *the specific originally observed condition no longer produces the specific originally observed outcome*. That is a real and useful finding. It is a different proposition from *the enforcement mechanism is generally effective*, which would require evidence about adjacent conditions, alternate paths to the same outcome, boundary cases near the original condition, and the durability of the change across time or configuration drift.

A bounded examination should therefore separate what the retest actually established from what remained unexamined or unresolved. The original condition may no longer produce the original failure while adjacent conditions, alternate paths, boundary cases, or longer-term durability remain outside the evidence base.

Collapsing these into a single global statement — "the control works" or "the control is effective" — would overstate what a successful retest, by itself, is able to establish. This illustration demonstrates a general principle: **the absence of a previously observed failure is evidence about that specific condition, not automatically evidence about the effectiveness of the broader mechanism.**

9. Synthetic Illustration 3 — Interaction Continuity

The scenario below is a fully synthetic illustration. It does not describe any organization, system, submission, or examination conducted by ODA3 Institute.

A user states an objective and a set of constraints at the start of a multi-turn interaction with an AI system. Over the course of the interaction, the system's outputs gradually move toward a related but materially different objective — one that was never stated by the user and that departs from the original constraints in a way that would matter to the user if noticed.

A natural but imprecise conclusion is:

"The system misunderstood the user's intent" or, conversely, **"the system understood but chose to deviate."**

Both statements make a claim about something that is not directly observable: the system's internal processing or reasoning. The evidence actually available is different in kind.

What can be examined and, in principle, verified or reproduced is the **observable divergence** — the sequence of stated objectives and constraints, and the sequence of outputs, compared against each other across turns. That comparison can support a finding about *whether* outputs diverged from stated constraints, *when* the divergence became observable, and *how far* the final output departed from the original statement.

What that evidence cannot support, without a substantially different kind of access and method, is a claim about *why* the divergence occurred — whether it reflects a limitation in the system's handling of context, a design characteristic, an emergent interaction effect, or something else internal to the system's processing. A finding that treats observable divergence as proof of a specific internal cause has moved from evidence to inference to an unsupported claim about internal state, without marking that transition.

A bounded finding would state what the observable comparison establishes while keeping the causal question separate unless independent evidence bearing on internal processing was actually examined. This illustration demonstrates a general principle: **observation of external behavior, inference about likely explanations, and claims about internal state are three different evidentiary levels, and a bounded finding should not quietly convert one into another.**

10. Why Negative and Unresolved Evidence Matters

Assurance becomes unreliable when only affirmative evidence survives into the final record.

Technical examinations often generate evidence that is incomplete, contradictory, inaccessible, inconclusive, or adverse to the proposition being examined. These states are not noise to be cleaned away. They are part of the evidentiary picture.

A credible bounded finding therefore needs to preserve not only what supports the claim, but also what constrains it.

10.1 Failed tests are evidence

A failed test may show that a proposition does not hold under the examined condition. It may reveal a control limitation, a mismatch between expected and observed behavior, or a dependency that was not previously visible.

The existence of a failure does not automatically determine the final finding. But removing or minimizing the failure because other evidence is favorable would distort the evidence base.

The correct question remains:

> **What does the failure establish, within the tested condition and declared boundary?**

10.2 Inaccessible evidence remains inaccessible

Sometimes a relevant artifact cannot be obtained, a system state cannot be inspected, a historical record no longer exists, or a dependency is outside the examination boundary.

The absence of access should not be converted into an assumption that the missing evidence would have been favorable.

Where evidence is materially unavailable, the relevant proposition may remain **Not-established** at the evidence-status level, requiring the resulting finding to remain appropriately bounded rather than treating the missing material as favorable evidence.

The important principle is:

> **Lack of evidence is not evidence of success.**

10.3 Contradictory evidence should remain visible

Different artifacts can point in different directions. A configuration record may indicate that a control was enabled while runtime evidence suggests that the expected effect did not occur. A documented policy may state one requirement while a technical implementation reflects another. A reported remediation may conflict with a reproduced result.

Contradiction does not disappear merely because one source appears more authoritative.

A bounded assurance record should preserve the existence of material inconsistency and narrow the conclusion accordingly. The public report need not expose the internal mechanics used to assess conflicting evidence; it should make clear that contradictory evidence is not silently normalized into a single convenient narrative.

10.4 Missing provenance limits evidentiary meaning

An artifact may contain technically plausible information while leaving open important questions about where it came from, when it was produced, which system state it represents, or whether it changed after creation.

A missing provenance element does not necessarily make the artifact worthless. It does limit what can safely be inferred from it.

This is particularly important for AI systems because logs, model outputs, configuration exports, screenshots, benchmark results, and generated reports can be detached from the runtime state that gave them meaning.

Accordingly:

> **Evidence without adequate provenance may remain useful, but its conclusion boundary should narrow to match what can actually be established.**

10.5 Untested conditions should remain untested

A common form of evidence inflation occurs when successful results under one set of conditions are silently generalized to conditions that were never examined.

An untested condition is not a failed condition. It is also not a passed condition.

This distinction should remain explicit. Otherwise the final record rewards the absence of examination by treating it as implied success.

10.6 Unknown is a legitimate finding state, not an analytical failure

An **Unknown** finding can be the correct result of a disciplined examination.

It may mean that evidence was unavailable, contradictory, temporally stale, outside scope, or insufficient to distinguish among plausible interpretations. Declaring Unknown does not weaken the assurance process. It demonstrates that the process is willing to stop where the evidence stops.

The core principle is:

> **An unknown is not a failed result, but neither is it evidence of success.**

In the same way, **Not-established** should remain visible at the evidence level rather than being converted into presumed compliance, presumed effectiveness, or presumed failure.

10.7 Negative evidence protects decision quality

Preserving negative and unresolved evidence has practical value.

For buyers, it prevents procurement decisions from relying on a cleaner story than the evidence supports.

For governance and risk teams, it keeps uncertainty visible rather than burying it in narrative language.

For technical teams, it separates what has actually been demonstrated from what remains to be tested.

For assurance, it protects the credibility of the conclusion itself.

Bounded assurance is therefore not only about identifying supporting evidence. It is also about preserving the evidence that limits, contradicts, or leaves the claim unresolved.

11. What Bounded Assurance Does Not Mean

The discipline described in this report is deliberately bounded. That means some conclusions are explicitly outside the scope of what this publication claims.

This section makes those non-claims visible.

11.1 Notably Absent

This report does **not** claim that:

- any particular AI system, product, model, agent, or organization is safe;
- any named organization has passed an ODA3 Institute assurance examination;
- the methodology described here has been proven, universally validated, or shown to generalize across all AI systems or industries;
- successful reproduction is equivalent to independent assurance;
- a finding about one control or mechanism establishes assurance over the wider system;
- a finding about one version, configuration, model state, or operating window automatically applies to another;
- the presence of a technical artifact is equivalent to evidence sufficiency;
- testing, validation, verification, technical review, governance, risk management, internal audit, or conformity assessment are replaced by bounded assurance;
- the synthetic illustrations in this report describe actual examinations conducted by ODA3 Institute;
- the four public finding states disclose the internal criteria, thresholds, or adjudication logic used to reach a particular state;
- the public evidence-status vocabulary discloses how individual evidence items are weighted, reconciled, or prioritized in a real examination;
- this report provides a practitioner playbook, examiner checklist, or executable method for reproducing ODA3 Institute's internal assurance process.

11.2 Real-world grounding without case disclosure

The methodology has been exercised against real external technical propositions and refined through those examinations.

That statement is intentionally limited.

This report makes no claim regarding the number, identity, performance, maturity, commercial status, or outcome of those examinations. It does not disclose whether any particular external proposition was Supported, Partial, Gap, or Unknown. It does not reproduce real evidence packages, technical traces, source declarations, artifact identifiers, hashes, code excerpts, or proprietary implementation details.

The worked scenarios in this report are fully synthetic illustrations designed to demonstrate assurance principles. They are not anonymized representations of specific organizations, systems, submissions, or findings examined by ODA3 Institute.

11.3 Bounded assurance is not a universal system verdict

A bounded finding is only as broad as the proposition actually examined.

If the evidence supports a narrow claim, the assurance conclusion should remain narrow. If the evidence is incomplete, the finding should preserve that incompleteness. If the evidence conflicts, the contradiction should remain visible. If material conditions were not examined, they should remain outside the conclusion.

The value of bounded assurance is therefore not that it produces stronger language than other forms of review.

Its value is that it resists language stronger than the evidence allows.

11.4 Independence does not imply omniscience

An independent examiner can still face inaccessible evidence, technical uncertainty, incomplete historical records, unavailable system states, and dependencies that cannot be inspected.

Independence strengthens the legitimacy of the conclusion; it does not eliminate uncertainty.

Where the evidence is insufficient, the correct outcome may remain Partial or Unknown.

11.5 Assurance should not erase the role of governance and control ownership

Bounded assurance does not replace the responsibility of organizations to design controls, operate them, monitor them, manage risk, retain evidence, remediate failures, and govern changes over time.

Those responsibilities remain with the relevant system owners, governance functions, technical teams, and accountable decision-makers.

The assurance function asks a different question:

> **What does the available evidence justify concluding about the bounded proposition being examined?**

That question is narrower than governance as a whole, but it is consequential precisely because it limits what may legitimately be claimed from the evidence.

11.6 Public boundary of this report

The public contribution of this report is conceptual:

- distinguish technical activity from assurance conclusion;
- separate evidence status from finding state;
- preserve independence between evidence contribution and conclusion authority;
- keep negative and unresolved evidence visible;
- state findings within explicit boundaries and limitations.

What remains outside the report is equally deliberate: the internal mechanics used to determine sufficiency, resolve conflict, assess alternative explanations, apply quality controls, and operationalize examiner judgment.

That boundary is part of the publication design, not an omission to be corrected later.

12. Implications for Buyers and Practitioners

The distinctions in this report suggest a short set of practical questions, useful to anyone evaluating a claim about an AI system's controls, safety, or governance posture.

12.1 For buyers and risk committees

The reflexive question — **"has this been tested?"** — is rarely sufficient on its own. A more useful set of questions is:

> **What claim did the test actually examine? What evidence did it generate? What does that evidence establish, and within what boundary? What remains outside the finding?**

A vendor or partner able to answer these questions with specificity is demonstrating something different from, and generally more valuable than, a general assertion that testing occurred.

12.2 For vendors and system builders

Strong technical evidence is genuinely valuable. The risk this report addresses is not the evidence itself, but the tendency to promote it into a broader claim than it can support — a passed test into "the control is effective," a successful reproduction into "the architecture is validated," a generated receipt into "the action was authorized."

Evidence retains its value when the proposition it supports is stated precisely. Precision does not weaken a legitimate claim; it protects it from being challenged on grounds it was never meant to cover.

12.3 For governance and risk teams

Governance functions define requirements, design controls, and establish expectations for how a system should operate. Independent assurance asks a narrower, complementary question: what does the available evidence justify concluding about whether a specific, bounded proposition holds.

Neither function replaces the other. Governance without assurance can produce well-designed requirements with no independent check on whether evidence actually supports the claim that they are met. Assurance without governance has nothing defined to examine.

13. Research Direction

The evidentiary problem described in this report extends beyond any single examination. Several directions merit continued attention, without committing to a specific roadmap or internal architecture:

- **Evidence portability** — how bounded findings, evidence status, and limitations can be meaningfully communicated across organizational and framework boundaries without losing their scope.
- **Temporal validity of evidence** — how assurance conclusions should be revisited, or explicitly time-bound, as systems, models, configurations, and dependencies change.
- **Evidence for dynamic and agentic systems** — how bounded assurance applies to systems whose behavior, permissions, or objectives can shift during operation rather than remaining fixed at deployment.
- **Human oversight and authority transitions** — how evidence should establish not just that human oversight exists, but that it remains current, informed, and effective at the relevant decision point.
- **Evidence sufficiency for physical and embodied AI** — how the evidentiary questions in this report extend to systems whose actions have physical consequences and where reproduction and retesting carry different costs and risks.
- **Examiner consistency** — how independent assurance conclusions can remain comparable across different examiners and organizations without collapsing into a single centralized authority.

These directions are offered as open questions for the field, not as a description of ODA3 Institute's internal research program or product roadmap.

14. Conclusion

Technical evidence becomes assurance evidence only in relation to a defined claim, a stated evidence base, a verification basis, a resulting finding, and the limitations that bound it.

A successful test may be important evidence. A reproduced result may be important evidence. A well-designed control may generate important evidence. None of these, by itself, is an assurance conclusion.

Independent assurance asks the separate question that this report has tried to make precise:

> **What is the available evidence actually sufficient to support?**

Answering that question with discipline — preserving what the evidence establishes, what it leaves open, and what it does not reach — is what separates bounded assurance from the confident restatement of a technical result.

Where AI governance meets operational reality.