

Provenance, Stability and Falsifiability of Evidentiary Status

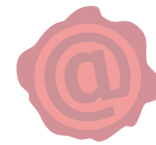
Extending the Declared / Established / Measured Framework

Emanuel Celano · Informatica in Azienda, Bologna

EVIDE - The Missing Evidentiary Layer for AI

Discussion Draft · September 2026

emanuel
celano
23.09.2026
13:23:37
GMT+02:00



Abstract

In July 2026, the EVIDE Discussion Draft “The Evidentiary Status of Internal Reasoning Signals” proposed that claims about an AI system’s internal state fall into three distinct evidentiary categories - Declared, Established, and Measured - and argued that these are not points on a single scale of strength but categories that a responsible governance architecture should never collapse into one another.

That Note closed by naming its own limits: before the distinction could become a field in any protocol, it would need explicit definitions, pre-registered test cases, and a frozen scope. This Note takes up that unfinished work. It does not revise the July 2026 Discussion Draft, and it does not treat the taxonomy it introduced as settled beyond what that Draft claimed. It asks a narrower, later question: once such a taxonomy exists, what must additionally hold - about who may assign a status, how that assignment is recorded, whether the taxonomy’s own boundaries can shift silently over time, and whether its distinctions can be tested against evidence rather than merely asserted - for a classification to remain trustworthy as it is used, preserved, and interpreted. In pursuing that question, this Note also draws on the wider evidentiary vocabulary developed in the September 2026 Research Discussion Note, Building Better AI without Losing the Evidence, without treating that paper as a source of the taxonomy itself.

This Note develops that question along several connected lines: a Classification Provenance discipline that separates Control Independence, Custody Independence, Economic Interest and Examination Authority as distinct properties rather than treating independence as a single word covering all of them; a Selection Provenance requirement for how a signal comes to be classified in the first place; a Semantic Stability discussion distinguishing four separate forms of drift; a Classification Status Transition discipline for how a classification may later be revised without being silently rewritten; a timing and fixation discipline distinguishing when a signal occurred, when it was classified, and when its record was fixed; two distinct levels of falsifiability, one for the taxonomy’s structure and one for a specific assignment; and a set of Forbidden Interpretations stating explicitly what none of this establishes. None of these developments revises the July 2026 taxonomy or the September 2026 evidentiary architecture; each is introduced here, and is cited to this Note rather than to either predecessor.

Research lineage

This Note should be read alongside two earlier EVIDE publications: The Evidentiary Status of Internal

Reasoning Signals (July 2026), which introduced the Declared / Established / Measured framework, and Building Better AI without Losing the Evidence (September 2026), which developed the wider evidentiary architecture within which questions of provenance, version binding, independent preservation and later examination arise. The present Note does not replace either publication. It develops a narrower question that becomes visible when the two are read together.

July: what evidentiary status does the signal deserve?

September: what evidence should survive, and under whose control, so that later examination remains possible?

Present Note: what must survive about the classification itself so that its evidentiary meaning can still be trusted later?

Further EVIDE Research Discussion Notes and Working Papers are available at:
app.certifywebcontent.com/working-papers

1. Where the July Paper Left Off

The July 2026 Discussion Draft closed with the following statement of its own limits:

“This paper is an early framing, not a specification. The Declared / Established / Measured distinction is offered as a promising direction, and it deserves the same discipline every EVIDE capability has been held to: before it becomes a field in any protocol, it would require explicit definitions, pre-registered test cases that exercise each category and the boundaries between them, and a frozen scope that parks the unresolved parts as backlog rather than implying they are settled. The idea is strong enough to merit that treatment; it is not yet mature enough to skip it.” (July 2026, §6)

That Note also stated two properties any future formalisation would have to carry: that signals are classified per instance rather than per type, and that a signal cannot promote itself - that its category is fixed by the channel that produced it, not by where it is later placed (July 2026, §6.1). It named category drift as the specific failure this second property guards against, and it identified provenance metadata as the mechanism required to prevent it, without specifying what that metadata would need to contain or who would be responsible for it.

This Note takes §6 of the July Draft as its starting point rather than as a conclusion to build past. Its purpose is to supply the explicit definitions, the first test cases, and a first attempt at a frozen scope that §6 called for - not to extend the taxonomy's claims about what internal-state signals are, which remain exactly as the July Draft stated them. Where it draws on vocabulary EVIDE developed elsewhere to do so, that vocabulary is cited to its own source, not folded silently into July's.

A note on how this Note came to be written. The September 2026 Research Discussion Note, Building Better AI without Losing the Evidence, addressing a structurally different question, was published for external comment and received sustained, unsolicited technical engagement. That engagement is documented separately, in full and attributed to its authors, in the External Review & Research Feedback accompanying that Note. It is referenced here only for a general pattern it revealed, not for its individual content.

Across that engagement, a recurring set of concerns emerged: who controls a boundary that determines what becomes evidentially significant; whether a record's meaning can drift silently across time or across versions of the scheme that defines it; whether "independence" is being used to cover several distinct properties that can vary separately; whether availability is being conflated with integrity; whether a party's own declaration is being treated as equivalent to independently available evidence supporting it; and whether a given evidentiary category remains available at all outside the context in which it was first demonstrated. None of these concerns was raised about the Declared / Established / Measured taxonomy directly, and none is attributed here to the individuals who raised it in that other context. Examining the taxonomy against the same general pattern, however, showed that it was exposed to versions of each of them. That pattern, not any individual's observation, is what motivated this Note.

2. The Research Question

The July 2026 Discussion Draft asked: which evidentiary status does a given claim about a system's internal state deserve - Declared, Established, or Measured?

This Note asks a different, later question:

Once an evidentiary-status taxonomy exists, what must hold for a classification to remain attributable, semantically stable, independently examinable and falsifiable as it is used, preserved and interpreted over time?

The two questions are not in tension. The first asks how a signal should be classified. The second asks what must be true of the classifying act itself, and of the record that survives it, for that classification to be worth relying on later. A taxonomy can be correctly defined and still be silently misused, silently reinterpreted, or silently untestable; this Note is about closing that second gap, not about reopening the first.

3. Relationship to Other EVIDE Constructs

This Note uses, without redefining, constructs introduced elsewhere in EVIDE's published research. Their relationship to one another has not previously been stated explicitly, and it is stated here because the coexistence of several partial architectures is exactly the condition under which an unstated composition tends to be assumed rather than examined.

Four constructs are in scope: the Evidentiary Metronome, the Minimum Evidentiary Surface, the Declared / Established / Measured taxonomy, and Classification Provenance. Each answers a different question, and none of the four questions implies an answer to the others.

- **The Evidentiary Metronome (September 2026, §4)** concerns when a deviation becomes material enough to trigger preservation.
- **The Minimum Evidentiary Surface (September 2026, §5)** concerns what evidentiary surface is available or has been declared for a given system and use case.
- **Declared / Established / Measured (July 2026)** concerns what evidentiary status a given signal deserves.

- **Classification Provenance (§7 below)** concerns why the assignment of that status can remain attributable and later examinable.

These are four orthogonal axes, not four stages of one pipeline. A system can have a well-specified Minimum Evidentiary Surface and no Declared / Established / Measured classification anywhere in it; a signal can be correctly classified without any Evidentiary Metronome having triggered its preservation; a classification can carry complete Classification Provenance while the underlying Minimum Evidentiary Surface remains narrow or undeclared. None of these combinations is a malfunction of the others.

Their coexistence does not establish that they form a single pipeline, that one implements another, or that interoperability between them has been demonstrated.

4. A Citation and Version Discipline

This Note reproduces and builds on material from two earlier EVIDE publications, each with a different role: the July 2026 Discussion Draft is the sole origin of the Declared / Established / Measured taxonomy; the September 2026 Research Discussion Note is the source of the wider evidentiary vocabulary - independent evidentiary preservation, version binding, independent examination, the separation between preservation and judgment - within which this Note's own questions are framed. Neither paper is revised by anything stated here.

The rule throughout is: where this Note reproduces a proposition from either earlier publication, the source and section are cited. Where it narrows, refines or introduces a distinction not present there, that change is stated explicitly. No refinement in this Note should be read as retroactively altering the meaning of either the July 2026 or the September 2026 publication, and no concept drawn from one is attributed to the other.

Explicit "refined here" or "introduced here" markers are used only where a reader could otherwise mistake this Note's wording for a predecessor's own, or where a definition here measurably narrows or sharpens a claim a predecessor made. Terms and distinctions with no counterpart in either earlier publication are, by the operation of this rule, new to this Note without needing to be marked as such individually.

5. Defined terms

Internal Reasoning Signal Any claim, structured or unstructured, that purports to describe a system's internal state, process, or reasoning at a given moment. (July 2026, §1–§2.)

Self-Report A claim produced by the very system whose state it describes. Its epistemic origin is unchanged by format, by where it is later stored, or by who later relies on it. (July 2026, §2.)

Independent Determination A claim produced by a channel that does not depend on the observed system's willingness or ability to report accurately on itself. (July 2026, §2.)

Declared The evidentiary status of a self-report. Grounded in nothing beyond the system's own assertion. (July 2026, §3.2.)

Established The evidentiary status of a claim attested by a separate, named, accountable party under an explicit protocol. Its basis is attributable determination under that protocol, not causal validation of the system's internals.

(July 2026, §3.2, defined the category; this Note states its basis without reference to comparative strength, for consistency with the principle in July 2026, §3.1 that the three categories are not points on one scale.)

Measured The evidentiary status of a causally grounded observation: a reading produced by a method validated, through intervention, to reflect the internal state it reports. (July 2026, §3.2.)

Causal Grounding The degree to which a determination's method has been validated by intervention (ablation, swapping, or an equivalent causal test) rather than by correlation or assertion alone. (July 2026, §3.3, §4.)

Accountability The identification of the party responsible for a determination, and the protocol under which that party made it. Distinct from, and orthogonal to, causal grounding. (July 2026, §3.1, §3.3.)

Taxonomy Authority The identified source responsible for defining or revising the criteria under which evidentiary statuses are assigned. Distinct from Classification Authority, which applies those criteria to a particular signal.

Classification Authority The identified party or process responsible for assigning an evidentiary status under the applicable taxonomy. Classification authority is limited to classification and does not confer examination, adjudicative, governance, or execution authority.

(The limitation stated in the second sentence keeps this term consistent with the separation of preservation from judgment developed in September 2026, §5: "It preserves the conditions for judgment; it does not exercise the judgment." Classification Authority is this Note's equivalent boundary for the act of classifying.)

Classification Provenance The record of the applicable taxonomy authority and taxonomy version; who exercised classification authority over a signal; under what method; and in what relationship to the system being classified.

Selection Provenance The record of which party selected a given signal for classification, under what criterion, whether that criterion was fixed before the case at hand rather than after it, and what conditions determined the signal's inclusion. Selection Provenance is not a separate authority alongside Taxonomy Authority and Classification Authority; it is a component of Classification Provenance (§7), because a classification applied correctly to an opportunistically selected signal is not thereby a trustworthy classification of the population it is read as representing.

(Introduced here.)

Control Independence Whether a party other than the party that assigned a classification, or the party the classified signal describes, is able to alter, delete, withhold, or substitute the classification or its record. Distinct from Measurement Independence, which concerns the production of the underlying reading rather than control over the classification once made.

(Introduced here.)

Custody Independence Whether the classification record is held by a party other than the classifying party, and whether it would survive that party's disappearance, withdrawal of service, dissolution, revocation of access, or revision of the taxonomy under which the classification was made. Custody Independence is a property of where and how the record is held, not of whether its content has been altered; a record can be integrity-protected and still be held exclusively by the party it concerns, with correspondingly weaker Custody Independence.

(Introduced here.)

Economic Interest Any commercial, contractual, or comparable relationship between the classifying, measuring, custodial, or examining party and the party the classification concerns. Economic Interest is disclosed rather than treated as automatically disqualifying: a payment relationship does not by itself establish control, and a commercial relationship does not by itself establish the absence of Control Independence. What is disclosed alongside the relationship is the concrete capacity, if any, that relationship gives one party to alter, delete, or withhold the record.

(Introduced here.)

Measurement Independence The disclosed degree of organisational separation between the party producing a measurement and the party that builds, deploys, or operates the system being measured. Distinct from causal grounding: a measurement may be causally grounded without being organisationally independent.

Examiner Independence Whether the party evaluating what a classified signal establishes is a party other than the one that classified it, produced it, or is described by it. Distinct from measurement independence: a measurement can be organisationally independent in its production and still be interpreted by an interested party.

(This distinction extends, to the act of classification specifically, the separation September 2026 draws generally between the party that preserves evidence and the “authorised examiner” who evaluates it (September 2026, §5, “Independence of the evidentiary layer”).)

Examination Authority The legal, contractual, procedural, or otherwise expressly recognised basis under which a party is entitled to examine a classified signal and its supporting record. Examination Authority is distinct from Examiner Independence, from technical competence, and from custody or access: a party may hold any of these without holding the others.

(Introduced here.)

| *Independence does not confer authority. Authority does not establish independence.*

Semantic Stability The property by which the meaning, criteria, limitations and forbidden interpretations attached to an evidentiary status remain recoverable for the version under which that status was assigned. Semantic Stability is not a single failure mode; it is threatened by at least four distinguishable forms of drift, named individually below.

Category Drift The absorption of a signal into a document, summary, dashboard, or later examiner’s own reasoning, in a way that carries a higher evidentiary status than the signal’s own origin supports. A Declared signal remains Declared regardless of what later attests to, summarises, interprets, or displays it.

(July 2026, §6.1, named the failure and the general risk; this Note extends the description of where it can occur, from documents and dashboards to an examiner’s own reasoning, for reasons given in §6.1 below.)

Taxonomy Drift The failure mode in which the criteria defining Declared, Established or Measured themselves change across taxonomy versions in a way that is not disclosed at the point where an earlier classification is read. Schema Version Binding prevents Taxonomy Drift from being silent; it does not by itself prevent the taxonomy from changing, only from changing unrecorded.

Interpretation Drift The failure mode in which what a later party takes a classification to establish changes over time or across readers, independently of any change to the signal, the classification, or the taxonomy itself. Interpretation Drift is the mechanism §6.1 identifies when an examiner treats a Declared signal as though it carried Established weight without any formal reclassification having occurred.

Record-Context Drift The failure mode in which the context necessary to understand a classification - its method, taxonomy version, dependencies, or stated limitations - becomes unavailable or is separated from the classification itself, even though the bare status label survives. A label that survives without its context is not thereby false, but it can no longer be relied upon at the strength the original classification supported.

(These four forms of drift are independent failure modes. A classification can resist one while remaining exposed to another; guarding against Category Drift, for instance, does not by itself guard against Record-Context Drift.)

Classification Status Transition The disciplined alternative to silent reclassification. A status transition does not overwrite the earlier classification; it creates a new, separately attributable record that preserves the prior status, states the reason for change, the evidence relied upon, and the taxonomy version applied. A downgrade is recorded under the same discipline as an upgrade: retrocession is not exempt from being logged as a transition.

(Introduced here.)

Schema Version Binding The requirement that a classification carry the version of the Declared/Established/Measured taxonomy under which it was assigned, so that a later revision of the taxonomy's boundaries does not silently reclassify a signal recorded under an earlier version.

(Introduced here; the term echoes, without importing, the version-binding requirement September 2026 states for reference profiles (September 2026, §4, §7).)

Category Availability Whether a given evidentiary status can be truthfully claimed for a signal in a given deployment context, given the tooling, access, and protocol actually in place there. A category can be defined in principle and unavailable in practice.

Future Examability The degree to which the preserved signal, its classification record, and the information necessary to understand how that classification was produced remain available for later examination. Future examability does not imply that the original measurement can necessarily be reproduced.

(Introduced here as a term; the underlying distinction between what is available and what is reproducible restates, for classification records specifically, September 2026's observation that "evidence availability is not itself a factual conclusion... it describes only what can still be independently examined after the event" (September 2026, §5).)

Evidentiary Status The Declared / Established / Measured classification of a signal at a stated moment, under a stated taxonomy version, assigned by a stated party. Not a judgement of the signal's truth. (July 2026, §5.2, §5.3.)

Classification Unresolved A procedural state, not a fourth evidentiary status, indicating that the available record does not support a responsible assignment among Declared, Established, and Measured. Where a signal cannot responsibly be classified, marking it Classification Unresolved is the disciplined response; forcing an assignment among the three defined categories in order to avoid an empty field is precisely the failure this state exists to prevent.

(Introduced here.)

6. Structural distinctions

- **Declared ≠ Established ≠ Measured** - three categories, not three points on one scale. (July 2026, §3.1.)
- **Causal grounding ≠ accountability** - a determination can be strong for either reason; the two do not sum into a single score. (July 2026, §3.3.)
- **Taxonomy authority ≠ classification authority** - defining the categories and applying them are separate acts and should remain separately attributable.
- **Measurement independence ≠ examiner independence** - organisational separation in producing a reading is not the same as independence in interpreting it.
- **Evidentiary status ≠ truth** - a classification describes the kind of evidence a claim is, not whether the claim is correct. (July 2026, §5.2.)
- **Category availability ≠ future examinability** - a status truthfully claimable today does not guarantee the record will remain verifiable later.
- **Integrity does not establish availability. Availability does not establish reproducibility.** A record may remain unaltered but inaccessible; accessible but no longer reproducible; or reproducible only while a particular model, tool, dependency, or privileged access path remains available.
- **Classification ≠ adjudication** - assigning a status is documentation; deciding what a claim means, or what follows from it, is not this framework's role. This is the same boundary September 2026 states for preservation itself: "It preserves the conditions for judgment; it does not exercise the judgment" (July 2026, §5.3; September 2026, §5).
- **Schema version ≠ semantic meaning across versions** - a taxonomy revision does not retroactively change what an earlier classification meant when it was assigned.
- **Absence of a Measured signal ≠ evidence of absence** - a system's internal state may be real and unfavourable even where no causally grounded reading of it exists.
- **A stronger evidentiary channel does not retroactively promote the origin of an underlying signal** - a self-report carried into a signed, certified, or audited document remains Declared; the surrounding document's own status does not transfer to it. (July 2026, §6.1.)
- **Evidence of declaration ≠ evidence of the declared state** - independently establishing that a system produced a statement does not independently establish the internal state described by that statement.
- **Control independence ≠ measurement independence ≠ examiner independence ≠ examination authority** - four distinct properties, not one word doing four jobs.
- **Independence does not confer authority.** Authority does not establish independence.
- **Economic interest ≠ control** - a commercial relationship is disclosed, not treated as automatically disqualifying.
- **Custody independence of the classification record ≠ integrity of the classification record** - a record can be technically unaltered and still be held exclusively by the party it concerns.

- **Label identity ≠ semantic identity across taxonomy versions** - the same category label under two taxonomy versions does not establish that the two classifications mean the same thing. Version continuity does not by itself establish semantic continuity.
- **A status transition does not rewrite the earlier classification** - it creates a new, separately attributable record.
- **Classification Unresolved is not a fourth evidentiary status** - it is a procedural state indicating the record does not support a responsible assignment among the three.

6.1 Testing a candidate organising thesis

The July Draft's governing formulation for this whole area was: a signal cannot promote itself (July 2026, §6.1). A broader candidate has been proposed for this Note: that a signal cannot promote itself, and neither should a classifier, a taxonomy revision, or a later examiner silently promote it. The candidate is tested here against the distinctions above, rather than adopted on the strength of its symmetry.

A classifier should not silently promote a signal. This clause is already load-bearing in §5 and §7: Classification Authority is explicitly limited to classification and confers no adjudicative or evidentiary weight beyond it, and Classification Provenance requires the method and relationship under which a status was assigned to be recorded. A classifier that assigned Established status to a signal meeting only Declared criteria, without disclosing the basis, would violate a requirement already stated. The clause holds without modification.

A taxonomy revision should not silently promote a signal. This is precisely what Schema Version Binding exists to prevent: a classification carries the taxonomy version under which it was made, and a later revision of the taxonomy's boundaries does not retroactively reclassify it. The clause holds without modification.

A later examiner should not silently promote a signal. This clause is the least settled of the three. Examiner Independence, as defined, concerns whether the examining party has an interest in the outcome; it does not by itself prevent an independent examiner from treating a Declared signal as though it carried Established weight in their own reasoning, without formally reclassifying anything. The original definition of Category Drift, confined to documents, summaries, and dashboards, does not clearly cover this case either. For the candidate thesis to hold, Category Drift has to be read as covering an examiner's own reasoning, not only artefacts a signal passes through - the extension made explicitly in §5 above.

On this test, the candidate thesis holds, but only once Category Drift is stated broadly enough to include the examiner as a possible site of drift, not only the documents an examiner produces or consults. That extension is adopted in §5. With it, the organising proposition for this Note is:

A signal cannot promote itself - and neither can a classifier, a taxonomy revision, or a later examiner promote it on the signal's behalf, without that promotion being disclosed as a separate, attributable act.

7. Classification provenance

Sections 5–6 establish what grounds each evidentiary status. They do not by themselves establish who is permitted to assign one, or what must be disclosed about that assignment for the classification itself to be trustworthy. The question is structurally the one September 2026 asks of preserved evidence generally - who controls it, and who is entitled to examine it - applied here to the classification act rather than to the underlying signal (September 2026, §5, “Independence of the evidentiary layer”).

A complete classification record should state: the taxonomy authority and taxonomy version under which the applicable criteria were defined; the classification authority that applied those criteria to the signal, and that party’s relationship to the system being classified; the measurement-independence status of the reading, where the claim is Measured; and the method or protocol applied.

Two cases deserve explicit statement. First, taxonomy authority and classification authority are separate acts and need not be exercised by the same party; a classification record naming only one of the two is incomplete. Second, a Measured status assigned by the same organisation that trains and deploys the system it reads is not thereby invalid - causal grounding is a property of the method, not of who operates it - but it is not organisationally independent under the definition of Measurement Independence in §5, and a record that omits this is incomplete. Causal grounding and measurement independence are disclosed separately; neither stands in for the other.

Classification Provenance also has a selective and a temporal dimension not yet made explicit above.

Selection. Before a signal can be classified, it must be selected for classification. Classification Provenance should record which party selected the signal, under what criterion, whether that criterion was fixed before the case at hand rather than after it, and what conditions determined the signal’s inclusion. A classification can be methodologically sound and still be applied to a population selected opportunistically; Selection Provenance (§5) is what distinguishes the two.

Timing. At minimum, three distinct times should not be treated as interchangeable: the time at which the signal occurred, the time at which it was classified, and the time at which the classification record was fixed. A classification produced today about a signal from three months ago is not a classification produced at the time of the signal, and a signature applied today does not establish that the record was fixed three months ago. Collapsing these into a single stated moment, as §5’s definition of Evidentiary Status does for brevity, is adequate for defining what a classification is; it is not adequate for establishing when the classifying act, or the fixation of its record, actually took place.

Fixation. Classification Provenance must be fixed in a form that permits later verification of its content identity, attribution, applicable taxonomy version, and recorded chronology. This requirement does not specify a technology. Fixing the classification record does not establish that the classification is correct; it establishes only that what was recorded, and when, can later be checked against what is claimed.

Completeness. A classification record may be complete, incomplete, or the classification itself may be Classification Unresolved (§5); these are not the same failure. A complete record states the taxonomy authority and version, the classification authority and its relationship to the system, the selection provenance, the measurement-independence and control-independence status where applicable, the custody arrangement, and any disclosed economic interest. An incomplete record is missing one or more of these while the underlying

classification may still be plausible; treating an incomplete record as though it were as reliable as a complete one is itself a form of Record-Context Drift (§5).

A future minimum classification provenance record would need to be capable of representing, at minimum, the following dimensions: the signal's identity and a means of confirming it has not been altered; the times distinguished above; the selection provenance; the taxonomy authority, taxonomy version, and classification authority; the status assigned and the method or protocol applied; the disclosed independence and custody position under §5; and, where the classification supersedes an earlier one, the status transition record required by Classification Status Transition (§5). This Note states the dimensions such a record would need to cover. It does not propose a schema, a data format, or a specific technology for meeting them; doing so is future work, consistent with the frozen-scope discipline stated in §10.

8. Falsification and limitations

Consistent with the July Draft's own standard for itself (July 2026, §6), this framework should be tested rather than assumed. Two distinct levels of falsifiability are in scope: whether the taxonomy holds together as a structure, and whether a specific classification can be shown to be wrong. §8.1 addresses the first; §8.2 addresses the second.

8.1 Structural falsifiability

A first step: take signals of known evidentiary origin - self-reported, independently attested, and causally measured - and provide independent classifiers with the same bounded classification record, while withholding the expected category label. Test whether the three categories remain separable under classification pressure or collapse into one another.

A second step: examine whether causal grounding and measurement independence can in fact vary independently. If empirical cases cannot be constructed in which one is present without the other, treating them as separate dimensions would require reconsideration.

A third step: construct cases in which causal grounding, control independence, custody independence, economic interest, examiner independence, and examination authority vary separately rather than together - for instance, a causally valid method operated by the system's own provider; an externally located classifier with an undisclosed economic dependency on the outcome; an independent custodian with no examination authority; an authorised examiner with a disclosed economic interest in the result; and an integrity-protected record that nonetheless remains recallable by its operator. If these dimensions cannot in practice be separated in any constructed case, treating them as distinct axes in §5 would require reconsideration.

8.2 Conditions that can falsify a specific assignment

The steps above test the taxonomy's structure. They do not state what would falsify a particular Declared, Established, or Measured assignment, or a particular Classification Provenance record, once made. The following conditions do.

A Declared assignment is falsified if the signal does not in fact originate from the system it is presented as describing, if the self-report's origin cannot be demonstrated, or if the record has been reconstructed or altered after the fact.

An Established assignment is falsified if the named accountable party does not exist or cannot be identified, if the stated protocol does not exist or was not in fact followed, or if the protocol actually applied does not match the one declared.

A Measured assignment is falsified if the causal validation it claims is absent, if a merely correlational reading is presented as causal, if the method or its version does not match what is declared, if the claimed intervention cannot be reproduced or documented, or if the reading does not in fact track the phenomenon it purports to measure.

A Classification Provenance record is falsified, independently of whether the underlying status is correct, if the stated taxonomy version cannot be recovered, if a stated digest does not match the signal it is claimed to identify, if the stated attribution cannot be verified, if the stated authority cannot be demonstrated, if a status transition has been omitted, or if the stated chronology is not supported by independently available evidence.

These conditions test whether a specific assignment survives scrutiny. They do not, by themselves, establish that a surviving assignment is true; they establish only that it has not yet been shown to be false in one of the ways stated here.

Limitations. The framework presupposes that Measured status is achievable at all - that some causally validated method exists to read the signal in question. This is not the ordinary case. The interpretability method that grounds the Measured category in the July Draft was developed and is operated by the organisation that trains and deploys the models it reads; whether comparable interpretability access is available to a party examining a system it did not build is a separate, open question this Note does not resolve. The absence of a Measured channel must not be treated as a defect in the system being examined. It is a limitation of the available evidentiary surface.

9. Forbidden Interpretations

This Note's frozen scope depends as much on what it does not claim as on what it does. Every item below is a reading this Note explicitly does not support, stated here because an unstated boundary tends to be filled in by whoever reads the Note next, and because the constructs in §3 make several such misreadings easy to reach for without noticing.

This Note does not establish, and should not be read as establishing, that:

- Established means true.
- Measured means infallible.
- Measured is always stronger than Established; they are different in kind, not ranked (§5, §6).
- A classification confers Examination Authority.
- Examination Authority confers adjudicative, governance, or execution authority.
- Causal grounding establishes independence of any kind.

- Independence, of any kind defined in §5, establishes correctness.
- Integrity of a classification record establishes the correctness of the classification it records.
- A digital signature or timestamp validates the semantic content of what it signs, rather than its content identity and chronology.
- Future Examinability guarantees reproducibility.
- Category Availability guarantees Custody Independence.
- The same category label carries the same meaning across different taxonomy versions.
- The absence of a Measured signal establishes the absence of the state it would have described.
- This taxonomy implements the Evidentiary Metronome.
- This taxonomy coincides with, or substitutes for, the Minimum Evidentiary Surface at any level.
- The coexistence of the constructs named in §3 demonstrates their interoperability or composition into a single architecture.

Where a future reader finds a use for this taxonomy that depends on one of the readings above, that use is outside what this Note supports, and should be stated as an extension rather than attributed to this Note.

10. Status and next steps

This Note is itself an early framing, not a specification, and is offered under the same discipline it asks the taxonomy it extends to meet. Several things remain open.

First, the falsification work in §8 has been proposed, not run; nothing in this Note should be read as reporting results, because none yet exist. This applies equally to the structural steps in §8.1 and to the per-assignment conditions in §8.2, which have not yet been tested against constructed or real cases. Second, the extension to Category Drift adopted in §6.1 was reached by testing a candidate thesis against existing definitions rather than by independent derivation; it should be treated as provisional until it, too, has been pressure-tested. Third, this Note has not examined whether Classification Provenance as defined in §7 is itself sufficient to survive the failure modes named in §6 - a provenance record can be falsified or omitted, and this Note does not yet propose a mechanism for detecting either beyond the conditions stated in §8.2. Fourth, the distinct provenance dimensions introduced in §5 - including Control Independence, Custody Independence, Economic Interest, Examiner Independence, and Examination Authority - are asserted here to vary separately; §8.1's third step proposes how to test that claim, but the test itself remains to be run.

The principle this Note preserves from its predecessors is unchanged: classification is not adjudication, evidentiary status is not truth, and preservation does not exercise judgment. What this Note adds is narrower - that the act of classifying, and the record that survives it, are themselves subject to the same discipline the July Draft asked of the signals they classify, and the September Note asked of the evidence that survives an event.

Primary sources

EVIDE Governance Lab (July 2026). The Evidentiary Status of Internal Reasoning Signals - From “Who Says So?” to “How Do We Know?” Discussion Draft. The Declared / Established / Measured taxonomy, the two disciplines of §6.1, and the passages cited throughout this Note as “July 2026” originate there.

EVIDE Governance Lab (September 2026). Building Better AI without Losing the Evidence. Research Discussion Note. The wider evidentiary vocabulary this Note draws on - independent evidentiary preservation, version binding, independent examination, and the separation between preservation and judgment - originates there; the passages cited throughout this Note as “September 2026” refer to it. This paper did not originate, and is not the source of, the Declared / Established / Measured taxonomy.

Further EVIDE Research Discussion Notes and Working Papers are available at:
app.certifywebcontent.com/working-papers

EVIDE: <https://app.certifywebcontent.com/>