The Tacit Layer Doctrine: Working Principles for Explicit Systems

Verification discipline, where it stops, and the knowledge that doesn't survive being written down.

How to read the numbering. Principles **1–4** are one class — "silence is not evidence" — and are best read together. **5–8** are the rest of the workflow discipline. **9–11** are the tacit layer.

Two halves and a bridge. **Part I (1–8)** is discipline for work whose errors are silent — parsers, caches, pipelines, joins, and language models alike; **Human Jurisdiction** states where it stops. **Part II (9–11)** is one claim about tacit knowledge seen three ways. They are in the same document because they share a root: an explicit system cannot tell you what it doesn't know, and treating its quiet as assurance is the same error in both halves.

Failure examples throughout are given as **shapes**, not incidents. Each is drawn from a real one; if a shape sounds contrived, it isn't.

What this does not apply to. Eleven principles all of which add friction, with no stated exemption, is a discipline that gets abandoned wholesale — which is the same argument that rescoped principle 2. So state the boundary: **this document runs unverified by default on work nobody downstream acts on.** Exploratory analysis, thinking out loud, single-reader notes, a throwaway cut of the data to see whether a question is worth asking. No provenance apparatus, no adversarial pass, no ledger entry. The cost of the discipline is only worth paying where being wrong propagates. The boundary is the one this document draws everywhere else — *does it leave the room* — with one addition, because the boundary moves without announcing itself: **the exemption travels with the artifact, not with its origin.** The moment a throwaway gets forwarded, cited, pasted into a deck or built on by anyone including you next month, it stops being exempt, and that transition is invisible by construction. Practical form: when you reach for an exploratory artifact a second time, that is the trigger to go back and do the verification you skipped. *The guard on the guard:* if most of your work is landing in the exempt class, the boundary has moved, not the work.

PART I · 1–8	PART II · 9–11
Workflow discipline	**The tacit layer**
1–4 silence is not evidence	9 value hides in the tacit layer
5–6 ambiguity, separated roles	10 tacit-to-tacit can't be automated
7–8 context and memory	11 don't let automation own relevance
Breaks loudly enough to leave a trace → earns corrections	*Generates no error signal at all → falsifiers must be stated in advance*

An explicit system cannot tell you what it doesn't know — and treating its quiet as assurance is the same error in both halves

Figure 1 — Two halves, one root

Part I — Workflow discipline

The genus: silence is not evidence

Principles 1–4 are one insight. Stating the genus is what makes them predictive instead of retrospective, so it comes first.

A parser coerces unreadable input to zero instead of raising, and a total understates for

months. An inferred value sits in a column with nothing marking it as inferred, and every downstream reader takes it as fact. A cached image is skipped by a build that reports success, and the delivered page shows figures from a month ago.

In each case **the system had no channel through which to signal wrongness, and the absence of a signal was read as a positive one.** Nothing looked wrong because nothing *could* look wrong. That is why failures of this class survive for weeks or months, while an exception surfaces in seconds.

The predictive form — a thing belongs to this class before you have found the bug in it. Anything with:

THE SEVEN SILENT PATHS	
1	A fallback, or an exception swallowed into a default
2	A default value standing in for a missing one
3	A cache, or a skip-if-exists path
4	A glob, or a “latest file” selection
5	A join on an unvalidated key
6	An inferred field stored without its derivation
7	A model’s fluent answer where it had no way to know it was missing something
No channel through which to signal wrongness. Absence of a signal read as a positive one — survives weeks or months.	
The only exit: a second channel that can fail differently An independent re-computation · A different read path · A reader who didn’t build the thing	
Two agreeing sources that descend from a common ancestor are one source. A cross-check that returns zero disagreements, run against a sibling of the thing being checked, has told you nothing at all.	

Figure 2 — The seven silent paths, and the only thing that converts silence into evidence

…cannot report its own failure. Treat *"we've never had a problem with it"* as uninformative about all of the above, because that sentence is exactly what this failure mode produces.

The only thing that converts silence into evidence is a second channel that can fail differently — an independent re-computation, a different read path, a reader who didn't build the thing. Two agreeing sources that descend from a common ancestor are one source. A cross-check that returns zero disagreements, run against a sibling of the thing being checked, has told you nothing at all.

The positive form, which the negative one implies and nobody states. If agreement between sources sharing an ancestor is worth nothing, agreement between sources that demonstrably do *not* share one is worth a great deal — and it is the only form of corroboration this document recognises. **Convergence is evidence exactly in proportion to the independence of the roads.** Two people reaching the same conclusion from a literature review and from a decade of things going quietly wrong have produced two observations; two who both read the same paper have produced one. When you find someone else already holding your conclusion, the question is not whether you were first — it is whether you got there by a different route. If you did, the overlap strengthens both positions rather than diminishing yours.

1. Use code for deterministic outcomes; verify against the source.

Where a verifiably-correct answer exists, build it as code rather than computing it in reasoning — then check the result against the source, with a human in the loop. The dividing line is **checkability, not difficulty**: a trivial exact sum belongs in code, a hard judgment call stays in reasoning. Verify against the source, never against the model's own restatement of it — self-consistent is not correct.

"Code errors are loud" is false as a general claim, and the belief is load-bearing in the wrong direction. Code errors are loud *only where you reconcile against an independent source*. Coercion, parse, and default-value paths are silent by construction: they return a plausible value instead of raising. A numeric conversion that maps malformed input to zero is confident paraphrase in code form — same failure as a fabricated quotation, wearing a lab coat.

The working pattern is an independent re-computation that halts the run — a verifier that re-derives the result from source data by a different route and stops the pipeline on any drift. Assert on the silent paths. Don't trust a number because the code ran.

2. Force provenance.

Distinguish what came from a source, what is inference, and what is genuinely uncertain. Quote before interpreting. The target is **silent confident paraphrase** — the failure that doesn't announce itself, and which survives repeated readings precisely because it sounds like something that would be true.

Scale the discipline to the stakes, not to the sentence. Tagging every clause inline destroys prose, and the clutter becomes the reason people quietly abandon the practice. Apply it to durable claims and to anything that leaves the room.

But do not replace inline tags with a disposition. "I could produce the source if asked" has no trigger and no artifact; it is knowable right up until nobody asks. Move the tag **off the sentence and onto the artifact**:

Artifact	Where provenance lives
Prose	a sources block at the foot
Data	a _source / _derivation column travelling with the field
Generated pages	a header or footer naming the input and its date
Persistent notes	the origin named in the entry itself

The test is not *"could I produce it"* but **"will the next reader see it without asking"** — because the next reader is often a future session, or a future you, that doesn't know there's anything to ask about.

3. Provenance has to survive the write.

Principle 2 marks a claim at the moment of assertion. It does not survive **derivation**. The moment an inference is written into a data file as a field value, its tag is gone: the next reader sees a bare value and treats it as fact. Repetition across derived artifacts then launders it. Nobody lies at any step — the uncertainty simply falls off in transit.

HOW AN INFERENCE BECOMES AN ESTABLISHED FACT		
1	Inference from whatever evidence was to hand	
2	Written into a table as a bare field value	*The tag falls off here.*
3	Other files built from that table	
4	Reports built from those files	
5	A year later, cited as established	

Nobody lies at any step; the uncertainty simply falls off in transit. A cross-check against any derived file agrees — they all descend from the original guess.

Figure 3 — How an inference becomes an established fact

The rule: a derived field carries its derivation. If a value's origin is inference, the artifact storing it says so. And the discipline that goes with it: before a field becomes load-bearing for anything that leaves the room, ask **"where does this actually come from?"** and follow it to the bottom. *"Everyone has always used this column"* is not an answer.

4. Verify the artifact, not just the pipeline.

Principles 7 and 8 govern staleness on the **input** side. This is the **output** side, where the logic is correct, the source data is correct, and the thing that shipped is old.

The shape: a generator skips regeneration when an output already exists. A build selects its input by glob and quietly takes the alphabetically last file. A manual copy step is skipped. In each case the run reports success and the delivered artifact is wrong.

The rule: caches are staleness-aware or they are regenerated, and **the final check is to look at what was delivered, not at the run log.** A pipeline reporting success is not evidence that what shipped is current.

Corollary, general form: **confirm through a channel that can fail differently.** When a file is read through a syncing or caching layer, a single read can return a stale snapshot — and re-reading it the same way returns the same stale snapshot, confidently. Size, line count, and modification time work as a check precisely because they don't share the read path that produced the error.

This is the same instruction as principle 1's "verify against the source, not the model's own output," and the same one as the genus above. Three faces of one rule.

5. Pin down ambiguity before executing, not after.

On underspecified work a model silently resolves every ambiguity to the most probable reading and builds on those quiet assumptions. It does not experience the ambiguity as ambiguity. Surface the assumptions up front, flag anything with more than one reading, then proceed. Catch divergence at the assumption stage, not after a thousand tokens of confidently-wrong output built on a term that meant two things.

6. Separate the roles — generate, then adversarially critique, then revise.

A model doing all three in one pass grades its own homework. The critical pass must be *"what is wrong here,"* not *"does this look right."* Same model, switched stance, gets you the second look.

GENERATE → *first pass*	CRITIQUE → *"what is wrong here"*	REVISE *retraction is the success condition*
TRIGGER (A)	**Anything that leaves the room carrying a number.**	
TRIGGER (B)	**Any claim about what a field means or where it came from.**	

(b) is not redundant — trigger (a) alone misses the whole class of principle-3 failures, because categorical fields carry no numbers. Point the critic at the load-bearing step, not the whole document.

Figure 4 — Three stances, and the two triggers that fire the middle one

Give it a trigger, or it becomes an exhortation and gets skipped. The pass is expensive, so "do this when it matters" means "do this when someone remembers." Two triggers, both chosen because they're unambiguous: **(a)** anything that leaves the room carrying a number, and **(b)** any claim about what a field means or where it came from. (b) is not redundant. Trigger (a) alone misses the entire class of principle-3 failures — categorical fields carry no numbers — and it misses the most common form of analytical error, which is not arithmetic but a claim about what a column represents.

Two notes on running it well:

- **Point the critic at the load-bearing step**, not at the whole document. In most analyses the arithmetic is fine and the inference from *measured thing* to *claimed thing* is where the work fails.
- **A retraction is the success condition**, not evidence the first pass was bad. A critical pass that returns "looks sound" on a complex claim usually means the stance didn't actually switch.

The shape worth knowing: a validation check that turns out to be tautological — the quantity doing the validating is computed from the quantity being validated, so the agreement was guaranteed and carries no information. This is the same error as two sources with a common ancestor, and it is very hard to see from inside the generating pass.

The control on review. Adversarial review is not a channel that only subtracts error, and treating it as one is how a second stance quietly becomes a second author. A reviewer produces two kinds of claim, and they arrive in the same voice, at the same temperature, in the same message. **Settleable** — the reviewed artifact can decide it. **Unsettleable** — nothing in the artifact can decide it; that is a new claim about the world, wearing review's clothes. **Only the first is review.** The second inherits this entire document — provenance, both triggers, all of it. The sorting has to be mechanical, because fluency is uniform across the two: **for each claim, name the document that settles it. If you cannot name one, it is not review.** Notice what this does *not* require — that the reviewer be honest, self-aware, or accurate about their own reasoning. It is a property of the claim, not of the reviewer's confidence, which is why it survives a reviewer who is bluffing, tired, partisan, or a language model that is fluent by construction. **Any control routed through a reviewer's self-report is worthless**, a model's account of its own interior included. Log accordingly: an unsettleable claim that turns out wrong is not a catch, it is a defect the reviewer introduced.

7. Manage context deliberately — what's in the window is all it knows.

No memory between turns, no privileged access to the wider project, no awareness of what was omitted. Missing facts get silently reconstructed into something plausible; long windows dilute attention across everything in them. Curate the working set as a deliberate act, and instruct the model to say **"not in context"** rather than fill the gap. An admission of ignorance is a signal; a reconstruction is silence wearing the costume of an answer.

8. Run the persistent memory like a knowledge base.

Once a memory file is load-bearing it inherits every knowledge-base problem, so it inherits the solutions too: hot and cold tiers, capture driven by demand rather than completeness, currency and staleness hygiene, and an index so depth loads on demand instead of all at once. **Separate settled doctrine from volatile current state** — they have different half-lives and mixing them means the stable material rots at the speed of the volatile material.

Don't over-articulate prematurely; promote a pattern once it's real. And apply the rule to the doctrine itself: **a change resolved in conversation gets written into the file in the same session, or it doesn't count as resolved.** A decision that lives only in a transcript will be re-litigated, and in the meantime the file is actively wrong.

Keeping a denominator

A doctrine like this is assembled from the failures that were **caught**. That is a sample selected on the outcome — the same reasoning nobody would accept in any other setting. "These principles have caught a lot" is compatible with "these principles miss most of what goes wrong," and nothing in the catch record distinguishes the two.

The cheap fix is a ledger. One line per correction, appended **at the moment of correction**, recording what was wrong, which principle bore on it, and — the load-bearing column — **who caught it**:

WHO CAUGHT IT — THE LOAD-BEARING COLUMN	
internal	verification fired before it shipped — the process worked
review	a second, equally-informed stance caught it before it shipped — still inside the room, and still able to introduce defects of its own
downstream	a reader caught it after circulation — a hole in the process, however gracefully fixed
late	found long afterward, by accident or unrelated work
The denominator. Without downstream and late, "the doctrine works" is unfalsifiable. The internal/review line is the same boundary drawn everywhere else: did it leave the room.	
Append at the moment; never reconstruct. *A ledger rebuilt from memory re-selects on the catches.*	
A stale ledger must look empty, not full. *Date every line and date the file.*	

Figure 5 — The ledger, and where the denominator lives

Downstream and late entries are the denominator. Without them, "the doctrine works" is unfalsifiable. Keep a companion list of **standing gaps**: things known to be wrong and not yet fixed, each with the date it was recognized. A gap with a date on it is honest; a gap nobody wrote down is the exact failure this document is about.

Two disciplines, or the ledger becomes its own principle-4 failure:

- **Append at the moment; never reconstruct.** A ledger rebuilt from memory re-selects on the catches and reproduces the bias it exists to correct.
- **A stale ledger must look empty, not full.** Date every line and date the file. *"We track this"* is worse than not tracking it when it's false.

Human Jurisdiction

This is the bridge between the two halves: Part I says what the method can catch, this says what it cannot, and Part II is what lives in the space that leaves. It is not a disclaimer and not a gap awaiting a future revision. It is a structural feature of the method, stated with the same specificity as the principles.

Why it has to be here. A verification discipline that does not state its own jurisdiction is dangerous in one specific way: it invites the assumption that anything it does not flag is therefore sound. **That assumption is itself a silent failure — the genus, aimed at this document.** A framework built to detect false confidence can manufacture false confidence about its own coverage, and it is the one failure it is least equipped to catch in its own operation.

The boundary condition

The whole mechanism rests on one precondition:

> **There must be an artifact, and the claim must be one the artifact can settle.**

Every principle in Part I is a refinement of that question — what counts as an artifact, what counts as settling, how to tell a settled claim from one that merely sounds settled. The method's power is that the question is answerable. Its boundary is that the question only exists once an artifact does.

Four categories fall outside, and they are four failures of the same precondition — which is what makes this a claim about the method's shape rather than a list of its weaknesses:

	Status of the artifact	
Selection	none **yet** — one is coming	what is worth checking, before anything exists
Interpretation	exists, but cannot settle the claim	whether the state is *good*, not whether it *is*
Synthesis	two exist, both settle, and they conflict	which governs

Origination	none, and there will not be one	judgment that was never externalized

That table carries more weight than it looks. **Selection and Origination are both pre-artifact tacit judgment and are distinguished only by whether an artifact is coming.** Keep that line sharp or the two collapse into each other.

Which category am I in? — the procedure

The categories are only worth having if you can place a live piece of work in one. Four questions, in order. The first one you answer badly is your category.

1 — Is there an artifact yet? *No, and one should exist* → **Selection.** The judgment is *what to make*. *No, and forcing one into existence would change or destroy the judgment* → **Origination.** The judgment is *what you already know and cannot yet say*.

2 — There is an artifact. Can it settle the claim you actually care about? *It settles a claim, but not that one* → **Interpretation.** The state is checkable; whether the state is good is not.

3 — Is there more than one artifact, each settling its own claim, pointing different ways? *Yes* → **Synthesis.** No third artifact ends this; see the regress below.

4 — Artifact exists, settles the claim, nothing conflicts. → **Delegable.** Not *easy*, not *unimportant* — **delegable**: a mechanism can do it, and human attention spent here is attention not spent on the other four. This is the answer with the most operational consequence in the whole document, and the one people are most reluctant to reach.

On multiple answers. The common real result is that a piece of work sits in two or three categories at once — an article that has to be worth writing (Selection), has to land with a reader (Interpretation), and has to not contradict the policy page (Synthesis). That is not the procedure failing. **The count is the measurement:** how many categories a piece of work touches is how much judgment it is carrying, and therefore how badly it repays being rushed or handed to a mechanism.

1 — Selection

Deciding what is worth checking, before any artifact exists.

By the time the doctrine engages, someone has decided that a particular thing was worth

documenting, testing, measuring or externalizing. That decision is invisible to the method, because the method begins with the artifact and the decision precedes it.

This is not a small omission. The most consequential failures in most knowledge systems are not wrong answers to the questions asked — they are **right answers to questions that did not matter, produced while the question that mattered went unasked.** A pipeline can verify every claim it is given and still be pointed at nothing important.

> **Narrowed in integration, because the first draft overstated it.** The draft said selection failures are *structurally undetectable* from inside the artifact set, since nothing in it records what was excluded and there is no negative space to inspect. That is true of the artifact set **alone**, and false as a general claim — and practice falsifies it. A corpus audit that finds the articles nobody has ever used, a coverage check that finds the teams no training reached, a demand analysis that finds the questions no document answers: each is an inspection of negative space. What makes it possible in every case is an **independent census** — a roster, a directory, a full record of the demand — something that enumerates the population the selection was drawn from. So the accurate statement, and the useful one: **Selection failure is undetectable from the artifact set alone, and becomes partly detectable exactly when you can obtain an independent enumeration of what could have been selected.** Where such a census exists, go get it; that is territory the method can take back. Where none exists, the boundary is real.

Selection is principle 11 seen from the other side. *Don't let automation own relevance as the goal moves* is a claim about what happens when selection is mechanized: the layer optimizes against a fixed proxy for "what matters" and drifts confidently. Same claim, opposite face — and worth saying plainly, because otherwise the obvious objection ("recommender systems already automate selection") looks unanswered. They automate it. That is the problem, not the refutation.

In practice: choosing which processes get documented at all; deciding which metric gates a phase transition; judging that one customer pattern is worth an article and another is not.

Characteristic error: treating a complete, verified artifact set as evidence that the right things were captured. Completeness *within* a set says nothing about the set's boundaries.

What would move the boundary: a reliable method for identifying high-value omissions with no independent census to compare against. Nobody has one; if one arrives, this category shrinks rather than disappears.

2 — Interpretation

Meaning beyond state.

The doctrine reports state: the artifact exists, is current, is findable, passes, matches. State is checkable, and checkable things are cheap. What state cannot report is whether the state is **good**.

An article can be findable and incomprehensible. A test can pass and cover the wrong condition. A process can be adhered to and be the wrong process. In each case the doctrine returns a clean result, and the clean result is *accurate* — it is answering a different question than the one that matters.

The asymmetry, which is the most useful idea in this section. State-checking is mechanical and delegable, so its cost trends toward zero as tooling improves. Interpretation does not compress: the effort to judge whether an artifact actually does its job for a human reader is roughly constant regardless of how good the surrounding machinery gets. The consequence is worth stating directly — **as the doctrine succeeds, the proportion of remaining human effort that is interpretive goes up, not down.** Freed bandwidth does not disappear; it relocates into exactly the work that resists mechanization. Any account of the doctrine's value that stops at "less checking" has told half the story, and the less important half.

(That consequence is argued, not measured. Nothing here tracks the interpretive share of effort over time, and it would be straightforward to be wrong about the slope.)

In practice: reading an article as the customer would read it; judging whether a passing test tests the thing it claims to; recognizing that a technically-correct answer will not land with its audience.

Characteristic error: reading a green state signal as a quality signal. They are different signals occupying the same visual position.

What would move the boundary: a system that reliably predicts human judgments of "does this artifact do its job for this reader," validated against held-out human ratings rather than against its own outputs. **And note the tension with this category's own retrospective test.** That test — did the reader use it, finish it, act on it — is behavioural, measurable and loggable. Which means that if the test is real, **the training signal for the boundary-mover already exists**: usage logs are held-out human judgments in everything but name. Interpretation may therefore be

more compressible than the asymmetry argument above implies, and the honest position is that this is the category most likely to move. That is a real research programme, not an impossibility — and if it landed, the asymmetry claim weakens and this category partially compresses.

3 — Synthesis

Reconciling artifacts that individually pass.

Two artifacts can each settle their own claims correctly and stand in tension anyway — in emphasis, in framing, in what they imply a reader should do next. The doctrine has nothing to say here, and the reason is structural rather than incidental.

To adjudicate between two verified artifacts you would need a third establishing which governs — and that third is subject to the same problem the moment a fourth appears. The regress does not terminate. There is no artifact-level resolution to an artifact-level conflict.

What resolves it is judgment held by people who understand the purpose both artifacts serve — context that exists relationally rather than documentarily. Part II has a name for that context: **Ba**, the shared place in which knowledge can be created at all. Synthesis is the one category with a pre-existing theory behind it, and it is Nonaka's. **This is the closest point of contact between the doctrine and the tacit layer it is named for: synthesis is performed in shared understanding, and shared understanding is not a document.** It is the hand-off into Part II.

In practice: deciding which of two accurate framings should lead; noticing that two correct policies pull a team in different directions; holding a coherent position across a body of work when no single document establishes it.

Characteristic error: escalating a synthesis problem into a documentation problem — writing a new artifact to resolve the conflict, which produces a third thing to reconcile rather than a resolution.

What would move the boundary: a governance rule that resolves conflicts without itself requiring interpretation. Precedence hierarchies are the usual attempt; they work until two artifacts sit at the same level, which is when they were needed.

4 — Origination

Tacit judgment never externalized at all. The permanent edge.

The doctrine requires an artifact to interrogate; judgment that was never externalized presents no surface. A coach's read that a trainee is not ready. An engineer's hesitation on an approach they cannot yet argue against. The sense that a customer's stated problem is not their actual problem.

These are not pre-artifacts awaiting capture — which is the whole distinction from Selection. Some will never be captured, because externalizing them changes or destroys what made them useful: the judgment is fast, holistic, and grounded in accumulated experience that does not decompose into stated criteria. The doctrine's silence here is not an oversight to be corrected in a future version. It is the shape of the method.

The practical implication is a **discipline of restraint**: when a judgment call cannot be reduced to a checkable claim, the correct response is *not* to force it into artifact form so the doctrine can operate on it. Forcing it produces a checkable artifact that verifies something other than the original judgment — a false positive dressed as rigor.

In practice: the read that precedes the analysis; the objection raised before the reason is available; the decision made on pattern recognition that survives later scrutiny.

Characteristic error: mistaking "not yet externalized" for "not yet rigorous," and demanding artifact form as the price of taking a judgment seriously.

What would move the boundary: demonstrating that a judgment of this kind survives externalization intact — that the captured version performs as well as the original in the hands of someone who lacks the experience behind it. That is Part II's claim restated, and it carries Part II's evidential status: **held on theory, not established.**

> **Terminology note.** "Origination" here means *judgment never externalized*. In the knowledge-creation literature the same word is used for *where novelty comes from*. Two senses; say which is meant.

What this is not claiming

Three misreadings are available, and each would weaken the argument.

Not a claim that humans verify better than machines. They do not. Within the doctrine's jurisdiction — checkable claims against inspectable artifacts — mechanized verification is faster, more consistent, and less prone to the confidence drift that makes human checking unreliable at volume. These four categories are not places where humans do the same work better. They are places where the work is a **different kind**.

Not a soft coda to a hard document. Each category carries a named characteristic error and a stated condition under which its boundary would move. They are meant to be usable: you should be able to look at a piece of work and say which category the human contribution falls into — **or notice that it falls into none, and is therefore delegable.**

Not an argument for less automation. The opposite. The case for pushing state-verification toward zero cost is precisely that it protects attention for the four categories, which cannot be automated and which degrade badly under the cognitive load of manual checking. **The doctrine is a filter, not a judge.** It removes false confidence from what can be checked, so that judgment is not spent verifying what a mechanism could have verified faster.

The alibi problem, and the guard against it

The most serious objection to this section is not that it is wrong. It is that it is useful to the wrong person.

Four named categories of work that cannot be verified are the best excuse for unaccountable judgment anyone has yet written down. **"That's Interpretation"** becomes *trust me*. **"That's Origination"** becomes *I can't explain it and you can't ask*. This is the elitism trap from the theory section, arriving with better vocabulary — the oldest unfalsifiable authority claim there is, now equipped with a taxonomy.

The distinction that defuses it is narrow and load-bearing:

> **These categories are unverifiable in advance. They are not unaccountable in retrospect.** A judgment that cannot be checked before it is made can almost always be checked after it has played out.

Category	The retrospective test
Selection	Did the thing you chose to build turn out to be

	the thing that mattered? Against a census where one exists; against downstream use where none does. **Nobody ever read it** is a selection failure reported late.
Interpretation	Did the reader use it, finish it, act on it? Claims about whether an artifact does its job are settled by whether it did.
Synthesis	Did the tension recur? A synthesis that actually resolved a conflict stops it coming back. One that papered over it does not.
Origination	Track record across a run of calls, never a single call. One hunch is not evidence. A person whose hunches survive scrutiny at better than chance is.

The four tests are not equally strong, and the section should not pretend they are. Interpretation and Synthesis have genuinely observable tests: did they use it, did the tension recur. **Selection and Origination do not — and for the same reason in both cases: acting on the judgment destroys the evidence.** You never observe the article you chose not to write. And if the coach's read that a trainee is not ready is acted on, the trainee never fails visibly, so the read is never tested. That is the negative-space problem from Selection resurfacing *inside the accountability guard* — and it means the guard is weakest in exactly the two pre-artifact categories, which are also the two where the alibi is most tempting. Two partial remedies, neither complete: **Selection** can sometimes be tested against a census, as above; and **Origination** can be tested across a *run* of calls where a base rate exists, which is why the test is track record rather than any single judgment. Where neither is available, say so rather than implying a test that cannot be run.

So invoking a category is not a defence. It is an entry in a record. If you claim Interpretation, you are accepting the test that goes with it — and the claim is falsified the same way any other is, just later and at coarser resolution.

The tell, borrowed from the elitism trap because it is the same tell. Someone genuinely operating in these categories can usually **develop the judgment in other people** — that is what coaching is, and it is why Part II treats the human channel as load-bearing rather than merely irreplaceable. Someone using the categories as cover cannot, and does not try. Inarticulacy is a fact about knowledge. **Refusing accountability for outcomes is a fact about the person.**

The operational claim

The doctrine tells you which of your confidence is unearned. It does not tell you what to be confident about. Selection, Interpretation, Synthesis and Origination are where that second question lives, and they remain human not by preference, tradition or sentiment but **by construction** — each fails the doctrine's own precondition in a different way.

Selection has no artifact yet. Interpretation has an artifact that cannot settle the claim. Synthesis has artifacts that settle their claims and conflict anyway. Origination has no artifact and will not have one.

Four failures of one precondition. Four categories of irreducibly human work.

Part II — The tacit layer

9, 10 and 11 are one claim seen three ways: keep tacit knowledge in the channel that carries it, and never let an explicit system — a document, a database, a model — mistake itself for the thing that knows what matters.

These three are inherited, not derived here — and the distinction is load-bearing. Principles 1–8 came out of failures: a parser that returned a plausible zero, an inferred column nobody had marked, a cached image under a successful build, a stale read of a synced file. They were arrived at by being wrong and noticing. **9–11 come from the knowledge-management literature** — Polanyi, Nonaka, and the practitioners who built on them. The evidence for them is theirs, not this document's. They are here because **they are why the first eight matter**: without a tacit layer, "an explicit system cannot report its own incompleteness" degrades into a rule about error handling. The silence is only expensive because of what it is failing to carry. **Two consequences, both practical.** *First*, hold 9–11 to a different evidentiary standard than 1–8 — see *What would disconfirm 9–11*, where for most of them no measurement channel exists and the claim is a design commitment held on theory. *Second*, **do not cite this document as independent support for that literature on these three.** It isn't independent; it descends from it, and two agreeing sources with a common ancestor are one source — this document's own genus, applied to itself. Principles 1–8 *can* be offered as corroboration, precisely because they were reached from failures rather than from theory and only afterwards found to agree.

9. The value hides in the tacit layer, and the recurring failure is forcing it through an explicit channel.

Documentation, training material, process automation, the memory file itself — the same error each time: **mistaking the articulable artifact for the substance.** Some knowledge transmits by demonstration, correction and shared work, and does not survive conversion to text. Forcing it through an explicit channel doesn't merely fail to capture it — it produces a confident artifact that *appears* to have captured it, which is worse than an obvious gap, because an obvious gap is still visible.

Two substances, one system. The mistake is not building the explicit machinery; it's believing the machinery is the whole thing because the machinery is the part that reports.

10. Tacit-to-tacit transfer can't be automated or self-served, and it's load-bearing.

Apprenticeship, mentoring and peer practice work because they keep transfer in its native mode. A self-serve course ships the shell with the judgment stripped out — and worse, it lets people believe the shell was the substance, which forecloses the question. **A fuzzy copy of the real thing beats a crisp copy of the wrong thing.**

This is also the answer to the persuasion problem. Belief moves through shared practice, not through information: you cannot argue someone into a conviction that is held tacitly by everyone who has it. Distributed peer practice is the honest fallback when time is short, because it is *the same mode* spread thinner — not a watered-down version of a different mode.

11. Don't let automation own relevance as the goal moves.

"What matters" is a moving target, and knowing where it has moved to is a tacit judgment. An automated layer optimizing against a fixed proxy for it will drift confidently and take everything downstream with it, because it has no tacit sense of wrongness to pull itself up mid-sentence. It cannot detect the relational signal that tells a person the ground has shifted.

"The system proposes, a human disposes" is not a courtesy — **the disposal step is where the relevance judgment lives.** Point automation at articulation and transformation, where it is genuinely strong, and protect the disposal step, which is where drift enters. Build the judgment layer before the automation that will need to carry it; do it in the other order and the drift gets paid for on the back end, with interest.

Note honestly: this principle rests on argument, not on a documented case of an organization that automated relevance, drifted, and rolled back. Predicted risks and observed ones are different evidentiary animals, and this is the former.

What would disconfirm 9–11

The most uncomfortable section, and the one that keeps this document from becoming the thing it warns about.

Part I earns corrections because it touches code and artifacts that break loudly enough to leave a trace. Part II generates no error signal at all — and by the genus stated at the top, silence from an unfalsifiable claim tells you nothing. A principle that cannot report its own

failure is in the same class as a cache and a swallowed exception. So state, in advance, what would count as disconfirming:

#	WHAT WOULD COUNT AS DISCONFIRMING	MEASUREMENT CHANNEL
9	A purely explicit intervention — document, standard, tooling change, no demonstration or coaching attached — producing durable capability change of the kind attributed to tacit transfer	*State it, per site. A falsifier with no measurement path is decoration.*
10	A self-serve arm producing outcomes comparable to live apprenticeship, on the same population, measured on practice rather than on completion	*Requires practice-level measurement, not completion data*
11	An automated relevance layer holding up across a genuine goal shift with the human disposal step removed, not merely present and unused	*Where the honest answer is "no channel exists," the principle is a design commitment held on theory, not a finding*

The trap. A falsifier defined loosely enough gets reinterpreted away when it fires — "that wasn't real tacit transfer," "that wasn't a genuine goal shift." **If a falsifier fires, the burden is on the doctrine, not on the definition.**

Figure 6 — Stated falsifiers, and whether a channel exists to observe them

And say which of these you actually have a channel to observe. A falsifier with no measurement path is decoration. Where the honest answer is "no channel exists," the principle is a **design commitment held on theory, not a finding** — which is a respectable thing to be, but must be described that way in anything consequential.

The trap while doing this. A falsifier defined loosely enough gets reinterpreted away when it fires: *that wasn't real tacit transfer*, *that wasn't a genuine goal shift*. The claim then becomes unfalsifiable by redefinition — which is the elitism trap below, arriving through the back door. **If a falsifier fires, the burden is on the doctrine, not on the definition.**

The theory underneath

The lineage, and why it matters: **Polanyi** (tacit/explicit — *"we know more than we can tell"*) → **Nonaka's SECI** (four modes converting between them: Socialization, Externalization,

Combination, Internalization — with the observation that Externalization is not Socialization, and instrumenting the former tells you nothing about the latter) → **Ba** (Nonaka's "place" — the shared context in which tacit knowledge surfaces at all: the mentoring relationship, the pair working side by side) → **phronesis** (what steers the whole thing and cannot be systematized).

THE LINEAGE	
Polanyi	tacit / explicit — "we know more than we can tell"
↓ **Nonaka · SECI**	Socialization · Externalization · Combination · Internalization
↓ **Ba**	the shared "place" where tacit knowledge surfaces at all
↓ **Phronesis**	what is called for here, now — steers the whole thing, cannot be systematized

Externalization is not Socialization — instrumenting the former tells you nothing about the latter

- Trained perception of particulars, not a rule — developed only among people who have it
- Cleverness (deinotes) plus knowing what is good — principle 11 in one line
- Distributed rather than concentrated; transmits by mentoring — principle 10

Figure 7 — The lineage: Polanyi → SECI → Ba → phronesis

Phronesis (Aristotle, *Nicomachean Ethics*) — practical wisdom, distinct from *episteme* (universal knowledge, knowing *that*) and *techne* (craft, knowing how to *make*). It is knowing **what is called for here, now**, in a specific contingent situation that the rules underdetermine. Three properties that matter here:

- It is **trained perception of particulars**, not a rule — which is why it is tacit, and why Aristotle held it cannot be taught, only developed through experience among people who have it. Tacit knowledge, twenty-three centuries before Polanyi.
- It is inseparable from a grasp of **the ends worth pursuing**. Phronesis is cleverness (*deinotes*) *plus* knowing what is actually good; cleverness alone hits any target, including bad ones. This is the whole of principle 11 in one line.

- Nonaka's late work makes phronesis the **driver** of the knowledge spiral, insists it be **distributed** rather than concentrated at the top, and holds that it transmits through on-the-job training, emulation and mentoring — which is principle 10.

The elitism trap — a live objection, keep it in view. *"The highest knowing is tacit and unarticulable"* is also the oldest unfalsifiable authority claim there is: *trust my judgment, I can't explain it*. Two failure modes: **mystify** it and you get unaccountable elitism; **deny** it and you flatten everything to checklists and lose the judgment that was real.

The honest position: tacit expertise exists, but its **claims stay accountable** — checkable by outcomes, open to challenge, transmitted rather than hoarded. The tell for the real thing versus the robes: do the results show, and is the person **developing it in others** or using its inarticulacy as a moat?

On rhetorical superiority. *"You're more persuasive doesn't make you right"* is valid, not a dodge — it names *deinotes*, cleverness decoupled from truth. The honest form is **"I'm unmoved and can't rebut it yet, so I'm parking it to check myself"** — epistemic self-defense. The dishonest form is *"you're clever, therefore you're wrong,"* which is a genetic fallacy and terminates thought. The distinction matters most against a fast, fluent interlocutor — including a model, which is fast and fluent by construction and persuasive independent of being right.

Practitioner notes: working the tacit layer directly

FOUR MOVES FOR PULLING THE TACIT LAYER UPWARD	
Concrete case	"Walk me through the last time this went wrong." They supply episodes; you do the articulating.
Catch it live	Interrupt mid-action: "wait — what just made you do that?" Visible in the moment, gone in retrospect.
Offer a wrong answer	The strongest move. They can't state the rule but can reject a bad version of it.
Contrast cases	"What's the difference between one that works and one that doesn't?" The difference is the knowledge.

Coaching · downward ↕ **Extraction · upward**

Same recognition, opposite direction — in both you supply the articulation scaffolding without making them feel tested.

Figure 8 — Four moves for pulling the tacit layer upward

Extracting tacit knowledge from a low-articulacy expert

The person who has the knowing but can't say it. The signs: competence and articulacy come apart — rough talk, good outcomes; chronic underselling; verbal deference to people who are visibly worse; and "how did you know?" answered with "I just could tell."

- **Ask about the concrete case, not the principle.** Tacit knowledge is stored as episodes, not axioms. *"Walk me through the last time this went wrong."* You do the articulating; they supply the cases.
- **Catch it live.** Watch them work and interrupt mid-action: *"wait — what just made you do that?"* The tell is visible in the moment and gone in retrospect.
- **Offer a wrong answer to correct.** The strongest move available. People who cannot state the rule can almost always **reject a bad version of it**. Say the deliberately flat thing; their expertise fires on your error.
- **Use contrast cases.** *"What's the difference between one that works and one that doesn't?"* They can't define it, but they can tell two things apart — and the difference **is** the knowledge.

Getting dissent from the conflict-avoidant expert

Usually the same person: quiet competence tends to come with a disinclination to fight you, so

they withhold exactly when you need them most.

- **Remove the social cost.** Pre-authorize it: *"I need you to tell me why this is wrong — assume I've missed something."* Makes dissent compliance rather than defiance.
- **Reframe as scenario analysis.** Not *"do you disagree"* but *"under what conditions does this blow up?"* Technical and safe rather than confrontational.
- **Ask last, one-to-one, off the record.** Stating your view first anchors them; groups reliably kill tacit dissent. Have them red-team a proposal that isn't visibly yours, so there's no ego in the room to manage.
- **Dig on the hedge.** *"It could work, I guess"* contains a swallowed objection. *"You paused — what's the part you don't like?"* Make swallowing it more effortful than saying it.

The through-line: this is coaching in reverse. Coaching transmits the tacit layer downward; extraction pulls it upward from someone who has more than they can say. Same recognition, opposite direction — and in both you supply the articulation scaffolding they lack (a rule to correct, a case to narrate, a wrong answer to reject) without making them feel tested. The moment the quiet-competent feel tested, they go quieter.

Revision history

Companion to the doctrine. Current version: **v7**, 2026-08-17.

Why this page exists

A document that arrives whole invites the reading that it was reasoned out in advance and has been right ever since. This one wasn't and hasn't. Every principle in it was written after something went wrong, several were amended after the amendment went wrong, and one was retracted outright as a general claim.

Publishing the doctrine without its history would be a small instance of the failure it describes: an artifact asserting more settledness than it has, with nothing in it that could signal otherwise.

So this is the record — at the resolution a reader can use, which is not the resolution I keep privately. The working copy logs specific incidents, in a specific organization, with names and dates attached. Those stay where they are. What follows is the same history stated as shapes rather than incidents, which is the rule the doctrine already applies to its own failure examples: the shape teaches better than the episode, and it is what makes a failure recognisable before you have your own version of it.

How to cite this, and what a version number means

Cite the version and the date. Both appear at the top of the doctrine itself. The version moves under a stated rule, so that the number carries information:

> **The rule.** A version bump means scope, numbering, or a principle's substance changed — anything that makes a prior citation wrong. Evidence added, wording tightened, an example swapped, or a note appended are edits in place. Those move the revision date, not the version.

If you cited v5 and the document is now at v7, the two bumps in between are described below, and you can decide for yourself whether what you cited still holds. That is the whole purpose of numbering it.

The history

v1 — 2026-07-20

Nine principles, written down after roughly a year of working with language models on material that other people would act on. Assembled from failures, not derived from a theory. The tacit-layer principles (now 9–11) were present from the start; the workflow half was thinner and, in one place, wrong.

v2 / v3 — 2026-08-12

Three weeks of use, and the first version where the document was corrected by its own subject matter rather than extended.

One principle was retracted as a general claim. v1 asserted that code errors are loud — that a program which is wrong tends to say so. That is true of exceptions and false of the more dangerous class: silent coercion, a swallowed error resolving to a default, a parse that quietly yields a plausible wrong value. The amended principle names those paths and pairs them with a working pattern: re-compute independently and halt on drift, rather than trusting a run that reported success.

One principle was amended twice, the second time reversing the first fix. The original form of the provenance rule required inline marking of every claim, which made real prose unusable. The replacement — mark it on request instead — was then itself reversed, because a rule that fires only when someone thinks to ask is a rule with no trigger, which is precisely the failure the principle exists to prevent. Provenance moved to the artifact level, where it has a place to live.

Two principles were added, both from the same discovery: that provenance is lost at the moment of derivation rather than at the moment of assertion, and that verifying a process is not verifying its output. A derived value that travels without its derivation becomes an established fact within a few hops. A pipeline can report success while shipping something stale.

A genus was named. Principles 1, 3 and 4 turned out to be one insight — silence is not evidence — and stating it made them predictive rather than retrospective. You can now recognise a system as belonging to this class before you have found the bug in it: anything with a fallback, a default standing in for a missing value, or a cache that can be skipped has a channel through which it cannot report being wrong.

Two sections were added to keep the document honest about itself: a catch ledger, because

an assessment built only from the times verification worked is a sample selected on its outcome; and a disconfirmation section for the tacit-layer claims, because those emit no error signal at all and so their survival tells you nothing.

v4 — 2026-08-12

A restructuring version, and one that came from an unexpected direction: writing a de-contextualized edition of the doctrine — one that could be handed to a reader with no shared background — surfaced structural problems invisible in the original. Stripping the context made the shape visible.

The principles were renumbered so the three that form a class sit together. Generic failure shapes were added alongside the specific incidents. Two lines were promoted out of examples and into the genus, where they belong:

- Two agreeing sources that descend from a common ancestor are one source.
- A falsifier with no measurement path is decoration.

v5 — 2026-08-13

Driven entirely by a reader. All three changes came from outside the process.

A scope boundary was added, and it is the reason for the bump. Eleven friction-adding principles with no stated exemption is a discipline that gets abandoned wholesale — the same argument that had already forced the provenance rule to be rescoped. The doctrine now states plainly that it runs unverified by default on work nobody downstream acts on, with two guards: the exemption travels with the artifact rather than its origin, so reaching for a throwaway a second time is the trigger to go back and verify; and if most of your work is landing in the exempt class, the boundary has moved, not the work.

A dated alignment pin was added to the de-contextualized edition, which had been claiming present-tense agreement with the working copy with nothing that would fire when the working copy moved.

The version-bump rule quoted above was stated, because this was the fourth bump in two days and a number that moves for every edit stops carrying information.

The same day produced a clean illustration of principle 4 in the document's own tooling: a

cached date field in a template had been silently stamping every file built from it, for months, including documents already in circulation. The template reported success on every build. The fix went into the builder rather than into any one script, and the circulated copies were left uncorrected as an accepted, recorded exposure — which is a different fact from the problem never having happened.

v6 — 2026-08-13

One change, and it closed a hole the previous version had opened.

v5 had introduced a category for errors caught by an independent reviewer and treated that channel as one that only subtracts error. It doesn't. A reviewer also produces claims that nothing in the reviewed artifact can settle — and those are not review. They are new claims, inheriting the whole doctrine, and if one is wrong it is a defect the reviewer introduced.

The sort is mechanical, and deliberately not routed through the reviewer's own judgment of their own reliability: **name the document that settles it, or it isn't review.**

The same version added the reciprocal citation pin described below, and made the document's independence claim carry its own limit in the same breath — because "reached from failures, only afterwards found to agree" is self-attested, unfalsifiable from outside, and made by the party it benefits.

v7 — 2026-08-17

A new section, placed as a bridge between the two halves: **Human Jurisdiction, which states where the method stops.**

A verification discipline that does not state its own jurisdiction invites the assumption that anything it fails to flag is therefore sound — which is the genus aimed at this document. The section names four categories that fall outside the method's reach, and frames them not as a list of weaknesses but as four failures of a single precondition: there must be an artifact, and the claim must be one the artifact can settle.

Four things were corrected during integration and immediately after, and they are worth naming because the section is about the limits of checking:

- **An overstatement was narrowed.** The draft claimed one category of failure was structurally undetectable. It isn't — it is undetectable from the artifact set alone, and becomes partly

detectable wherever an independent enumeration of what could have been chosen exists. That turned a stated boundary into an instruction.

- **A falsifier was stated for each of the four,** because four strong irreducibility claims with no stated failure condition would fail this document's own test.
- **A placement procedure and an accountability guard were added.** The section claimed a reader could place live work in a category and supplied no way to do it; and four categories of unverifiable work are, stated carelessly, the best alibi for unaccountable judgment anyone has written down. The guard is narrow: these categories are unverifiable in advance and not unaccountable in retrospect.
- **Then the guard was itself corrected, twice, on the same day.** Its four retrospective tests are not equally strong — two of them describe judgments that destroy their own evidence when acted on, so the guard is weakest exactly where the alibi is most tempting. And one category's retrospective test turned out to be nearly the same signal as the thing that would let it be automated, which makes it more compressible than the section had implied.

The section's most quotable claim — that as verification improves, the share of remaining human effort that is interpretive goes up rather than down — is marked in the text as argued rather than measured. Nothing tracks it.

Two things that travel with the document

The honest limit on all of the above

The record only shows what verification caught. There is no complete log of what slipped through, so the evidence base is biased toward success by construction. Read this history as "these principles fired at least this often," not as a measured hit rate. A catch ledger exists and is kept privately; the bias is structural until it has run long enough to contain entries nobody wanted to write.

A careful process that caught eighteen things and a careless one that generated eighteen things needing catching produce the same record. That ambiguity is not resolved here.

One authorial lineage — nothing here corroborates anything else here

This doctrine, its de-contextualized edition, and the related book chapter are three registers of one person's argument. The doctrine has always said not to cite itself as independent support

for the theory it draws on. The reverse now also holds: the book descends from this document and must not be cited back as support for it.

Direction of descent is doctrine → essay → book, and it is stated here with dates rather than left to be reconstructed. Publication order fixes descent but not evidence: whichever artifact appears first becomes the citable source for a claim whose support still lives in the others. Order changes what a reader can trace. It does not create a second witness.

Incidents behind each change are logged privately and are not published. If a principle here reads as obvious, that is the intended end state — none of them were obvious in advance of the failure that produced them, and all of them were violated by people who would have endorsed them in the abstract.

Licence and citation

Suggested citation. Nick Morgan, *The Tacit Layer Doctrine: Working Principles for Explicit Systems*, v7, 17 August 2026. CC BY 4.0. Cite the version and the date — the revision history above explains what a version number means and what it does not.