

FRONT PAGE · THE FIRST DISTINCTION

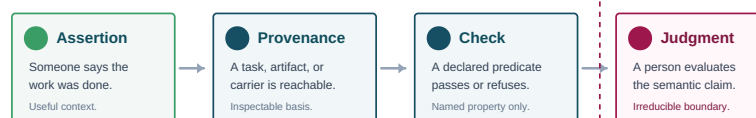
A map is not a promise.

Lettuce gives agents a place to orient, act, leave evidence, and discover the next piece of work. It does not turn a green cell into a universal certificate of quality.

A GUIDED ROUTE THROUGH THE THEORY · READ IN ORDER

One claim, four different kinds of support

Each stage narrows uncertainty. None turns automatically into the next.



Do not infer quality from the green :

The problem in one minute

Software work contains facts, checks, and judgments. A task can be closed. A test can pass. A design can still be wrong for its users, expensive to change, or quietly accumulating risk.

Agents are especially vulnerable because their context is temporary and their language is persuasive. They need a shared memory that makes the work and its limits visible.

The glassbox answer

Lettuce is a filesystem-native coordination and provenance substrate. It names work, maps a quality space, preserves attempts, links evidence, and refuses some declared-invalid moves.

Its central virtue is not omniscience. It is the ability to say: **this narrow thing was recorded; that broader thing remains open.**

How to read this paper

Each desk advances one idea. The green boxes are lessons to keep. The amber boxes mark the boundary of what a mechanism proves. Diagrams show shape; captions say what the shape cannot certify.

The system should make uncertainty harder to erase, not easier to decorate.

1 · Give the work a shape

Scopes are the frame

A project is too large to inspect as one undifferentiated promise. A **scope** says which slice of the product is under consideration: an API surface, a user journey, a service, or a subsystem.

Scopes prevent a polished component from borrowing credit from an untouched one. They create a denominator that can be challenged.

Dimensions are the questions

A **dimension** is a quality axis: test behavior, inspect accessibility, measure latency, review security, or examine maintainability. Its members name the questions inside that axis.

Dimensions are not decoration. They are an agreement about what “looked at” means.

Units are the subjects

A **unit** is the thing being examined: a command, endpoint, screen, package, release, or customer journey. A unit can appear across several dimensions because one subject has many ways to fail.

Declare the questions before counting the answers

scope = checkout / 3 units x 3 applicable dimensions = 9 coordinates

	functional	accessibility	recovery
cart	HARDENED stored + credited	implicit untested	implicit untested
payment	implicit untested	implicit untested	GAP stored finding
receipt 1 credited / 8 applicable = 12.5%	implicit untested	EXCLUDED rationale required Not complete: the implicit cells never disappeared.	implicit untested

KEEP THIS

A cell is an address, not a grade.

The intersection of a unit and a dimension is a **cell**. It tells an agent where a claim belongs. Its state says how far the project has carried that claim through the declared workflow.

WHAT THE MAP PROVES / DOES NOT PROVE

Proves: the project declared a coordinate and recorded its current state. **Does not prove:** that the map contains every important surface or that the chosen axis is the right proxy for quality.

A complete grid can still describe an incomplete world.

2 · Define “done” without hiding the denominator

Zero is not “not applicable”

If a scope has no applicable members, its result should be **n/a**, not a fabricated 0%. If members exist and none is hardened, 0% is honest. The two states demand different next actions.

DoD is a floor

A Definition of Done is a versioned quality bar. It may require a grade floor, evidence depth, freshness, or other criteria. A project can be 100% hardened and still have an unmet DoD because its evidence is stale or too shallow.

That apparent tension is useful: coverage and durability are separate facts.

Metrics belong beside the state

Complexity, coupling, mutation score, latency, and defect rate are measurements. Lettuce should preserve their raw value, unit, method, threshold, subject revision, and artifact. The cell state is the derived claim, not a replacement for the number.

Observed	Meaning	Agent response
Gap	Known problem or failed bar.	Remediate; keep the claim open.
Untested	No meaningful attempt recorded.	Design the check or expand the map.
Hardened	Declared gate and provenance passed.	Use it, but respect freshness and limits.
Stale / unknown	Trust boundary is no longer established.	Re-measure, reconcile, or escalate.

KEEP THIS

The bar must travel with the claim.

If an agent can lower the threshold, change the denominator, or replace a failed report without a visible policy revision, the metric is an invitation to game the map.

The **-1 ... 0 ... +1** continuum can describe harmful, unresolved, and useful states, but it is not a universal quality score or a current substitute for the cell workflow.

3 · Evidence is narrower than proof

Four levels of knowing

Assertion: someone says the work is done.
Provenance: a durable record points to who, when, and what changed.
Mechanical check: a repeatable predicate passed.
Judgment: an accountable person decides that the evidence is adequate for the consequence.

Lettuce is strongest at the first three. It makes the fourth visible; it does not impersonate it.

The receipt has a subject

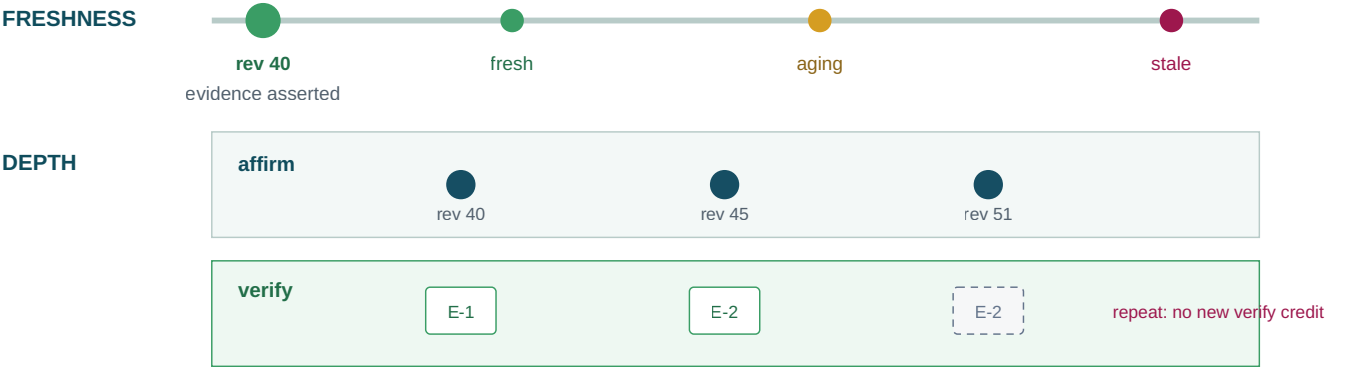
A report should identify what it measured, at which repository revision, with which analyzer and configuration. “Complexity improved” without a subject revision is a story with its address removed.

Evidence ages

Store revisions, code changes, analyzer changes, incidents, and changing expectations can make yesterday’s evidence less relevant. Freshness is a reason to revisit, not a probability that the claim is false.

One claim, two different reasons to revisit it

Freshness measures store movement. Depth records corroboration structure.



THE ANTI-HALLUCINATION RULE

A passing command is not evidence until its output, scope, method, and subject are preserved. A linked artifact is not semantic proof until someone asks whether it actually supports the claim.

KEEP THIS

Never let a green projection outlive its evidence.

When a cited task is deleted, reopened, or its subject changes, reconciliation should regress the dependent claim. A board that stays green after its proof disappears is memory laundering.

4 · Orient → act → reflect → file

Orient

Read `board next --depth 2` and open tasks. Choose one bounded coordinate. If the map is empty, organize the map; do not call the empty frontier success.

Act

Create or find the task, acquire a lease, start the workflow, and begin a run. The lease prevents a legitimate duplicate claim on that task. It does not prove that two patches are independent.

Reflect

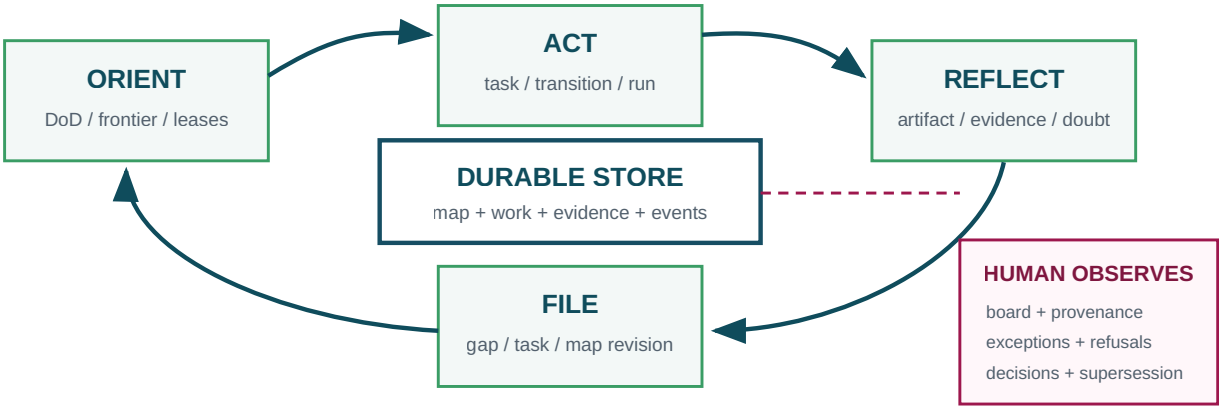
Record the run, attach the artifact, finish the attempt, and link the evidence to the cell. Ask the declared gate to decide. If it refuses, work or escalate; never facilitate green merely to keep moving.

File

Every unfixed discovery becomes a task, gap cell, new dimension member, milestone change, or human decision. Then reconcile and validate before orienting again.

The worker may vanish; the loop must remain

Each move reads or changes durable state before another agent inherits the work.



```
ORIENT board next --depth 2
ACT lease → workflow → run
REFLECT artifact → evidence → gate
FILE task / gap / dimension / decision
REPEAT reconcile → validate → orient
```

KEEP THIS
The loop is a method for surviving restarts.

The agent’s context can vanish. The store must still explain what was attempted, what was proven, what is blocked, and what another agent should do next.

5 · A graph is a contract before it is a picture

Shape

Nodes and edges describe topology: producer, reviewer, router, fork, join, loop, verifier, human, terminal. Shape says what route exists and which structural conditions must hold.

Behavior

Instructions say what a station should do. An instruction is not a proof that the station did it. Execution policy says how a node is staffed; it is not automatically a scheduler.

Carriers

A carrier is a durable handoff. It makes an output inspectable and lets a later reader reconstruct the route. A declared carrier that the runtime never requires is a diagrammatic promise, not enforcement.

Quorum

K-of-M agreement can reduce search uncertainty. It does not create independent evidence when every branch shares a model, prompt, source tree, and blind spot.



GRAPH DESK CORRECTION
A graph definition can be sound while a caller supplies weaker runtime terms. Read the effective contract and the actual run; never infer enforcement from the diagram alone.

Topology gives work a route. It does not give the workers wisdom.

6 · Agents propose; the harness constrains

Non-determinism is not the enemy

Language models vary with sampling, context, model, tools, orchestration, and ambiguity. Variation can search a space. It cannot turn a token likelihood into the probability that a patch is correct.

Put determinism at boundaries

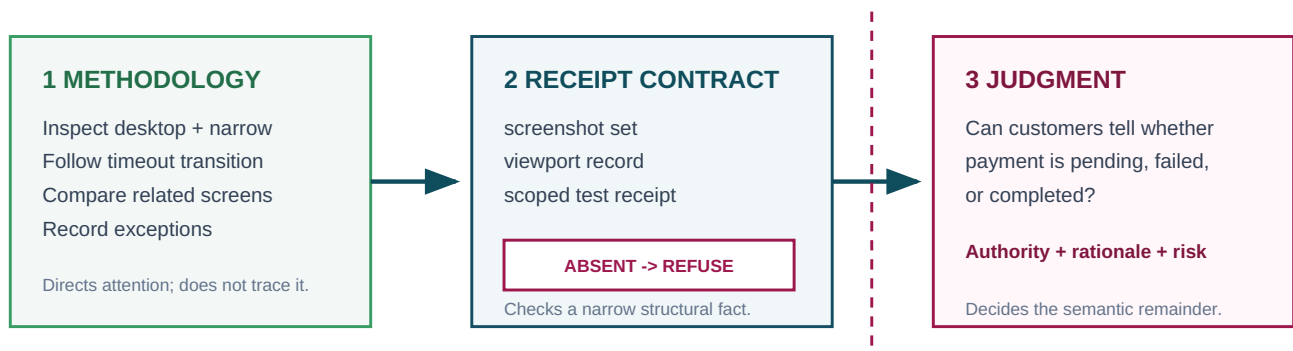
Check paths, leases, revisions, schemas, carrier presence, loop caps, effect permissions, and required evidence. These are narrow questions a harness can refuse mechanically.

Beware correlated agreement

Five agents using the same model, prompt, repository, tests, and framing are often one instrument sampled five times. Preserve dissent and vary falsification questions.

Three layers answer three different questions

Direction can be explicit. Receipts can bite. Meaning still has an owner.



A receipt crossing layer 2 is not an answer to layer 3.

The system can expose the remainder and route it to authority; it should not counterfeit the decision.

KEEP THIS

Confidence is not authority.

Let a model recommend. Let a deterministic boundary check. Let a qualified human or independent assurance process decide when the consequence exceeds the oracle.

METRIC WARNING

A worker rewarded for lowering complexity can scatter logic, delete tests, or optimize the analyzer. Pair structural metrics with behavior, security, reliability, and human outcomes.

7 · Decide where judgment belongs

Identity is not authority

An author field tells us which principal recorded an event. Authentication tells us whether the credential was verified. Authority tells us what that principal may decide. Qualification tells us whether an evaluator may make this class of judgment. Independence asks whether the observer could rewrite the subject's evidence.

Human nodes are not human decisions

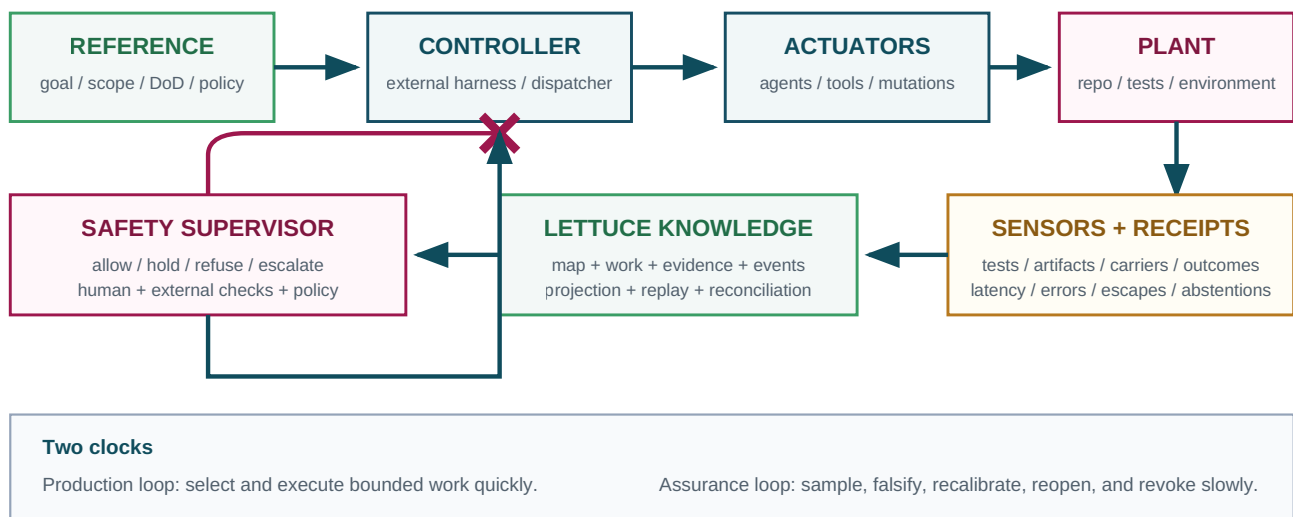
A graph node labelled human is a route shape. A decision needs a question, authority, alternatives, evidence, rationale, outcome, conditions, effects, and supersession. "Human-resolved" is not the same as a durable accountable act.

The supervised factory

The primitive loop can drive one agent today. A future factory controller would add bounded shifts, capability matching, budgets, independent assurance, effect authorization, and reopening when outcomes expose rot.

The ledger is memory; the factory still needs a supervisor

A controlled shift closes the loop from policy to action to evidence to the next authorized move.



KEEP THIS

Quiet operation is not absent oversight.

A factory should be quiet because routine work is bounded and inspectable, not because nobody can challenge the map, the bar, the evidence, or the authority.

The small set of rules that keeps the glassbox honest

Lettuce is

- A map of declared work and quality questions.
- A ledger of tasks, leases, runs, artifacts, evidence, revisions, and decisions.
- A place where selected invalid transitions can be refused.
- A restartable memory for agents that cannot rely on yesterday's context.
- A projection surface for humans: boards, graphs, timelines, and replay.

Lettuce is not

- An oracle that software is good.
- A guarantee that the denominator is complete.
- Proof that a linked artifact is truthful or relevant.
- Independent assurance merely because several agents agree.
- Authorization for an external effect or a transfer of accountability.

1. Name the claim before choosing the check.
2. Declare scopes, dimensions, units, and the denominator.
3. Keep raw metric values beside derived cell state.
4. Acquire a lease before exclusive work.
5. Record a run and real artifact evidence.

6. Let gates refuse; never facilitate green to make progress.
7. File every unfixed discovery.
8. Reconcile after evidence or subject changes.
9. Keep production and assurance independent.
10. Read effects backward: decision, run, evidence, work.

The measure is useful when it helps the next agent know what to do, what not to claim, and what must still be decided.

Continue with Issue 02 for the operating-agent journey, Issue 03 for trust after green, and Issue 04 for the Graph Atlas. For live operation, read [SKILL.md](#), run [lettuce usage](#), and consult [docs/guides/AGENT-WORKFLOW.md](#). For the deeper human theory, continue to Inside Huru's Glassbox.