

THREE WITNESS FAMILIES, FIVE EVIDENCE CHANNELS

The Three-Witness Evidence Matrix

Bring one consequential run you are responsible for. Weigh its three witness families across five evidence channels until they reconcile.

BRING one consequential agent run, ideally a disputed one _____

1. The evidence-weight matrix

Five rows are still three witnesses: the self-report, the action log, and independent judgment in its three forms. The third column is the discipline - write the question this witness is not allowed to answer.

Witness	What it says	What it proves	What it cannot prove
Self-report (the system's story)			
Action log / harness trace			
External ledger (record the agent does not control)			
Independent verifier (deterministic or off-family check)			
Human domain review			

2. Can you produce the witnesses at all?

A system that can produce all three witnesses can be audited; two can often be reconstructed; one is only a story. Tick what your system can actually deliver for the run.

- We can retrieve the self-report - the rationale or status message the system produced.

- We have a machine-readable action log: prompts, tool calls, observations, retries, errors, timestamps.

- The log is portable - exportable, replayable, mapped to a schema, legible outside the vendor dashboard.

- An external ledger exists to reconcile against (bank statement, CRM mutation history, code diff, mail server, document access log).

- The judgment layer is structurally independent of the generator - not the same model family grading its own homework.

- We have logged coverage gaps explicitly, so a missing span reads as "not instrumented" rather than "did not happen".

3. Reconcile, and name the missing claims

If a row above is empty because that witness does not exist, name the claim you can no longer make without it. If two rows disagree, decide who has authority to resolve it - before the disagreement is live.

4. Five questions for the vendor

Put these to any vendor before signing. A vendor who can show only final outputs is selling a black box with a friendly transcript stapled to it.

- Can you produce all three witnesses for a single disputed run?

- Are the logs machine-readable, replayable and portable enough to leave your dashboard?

- Which trace fields are retained, which redacted, which simply unavailable - and under whose policy?

- Which of your checks are deterministic, which model-judged, which human-reviewed?

- When the layers disagree, who has the authority to decide?

Companion worksheet to **Essay 17 · The Three Witnesses to a Run**, in the series **Architecting the AI Coworker**. · Dr Peter McCann Strain · Fill this in against one real agent, action class or vendor. © 2026 Peter McCann Strain.

Series Companion + all 22 worksheets: **Release_v12/Series_Companion.pdf**