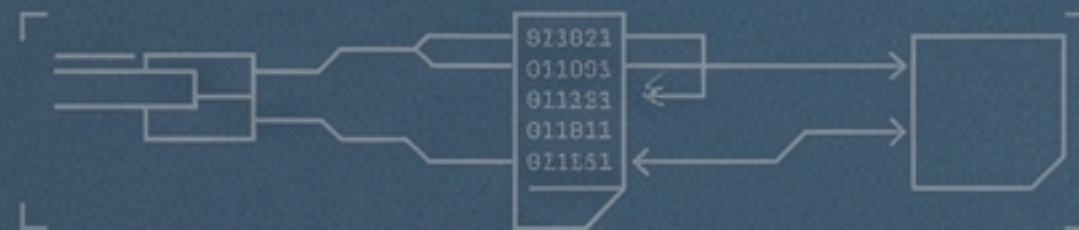




STATUS: NOMINAL
SYS CHK: OK
ORBITAL PARAMETERS: CALIBRATED
...

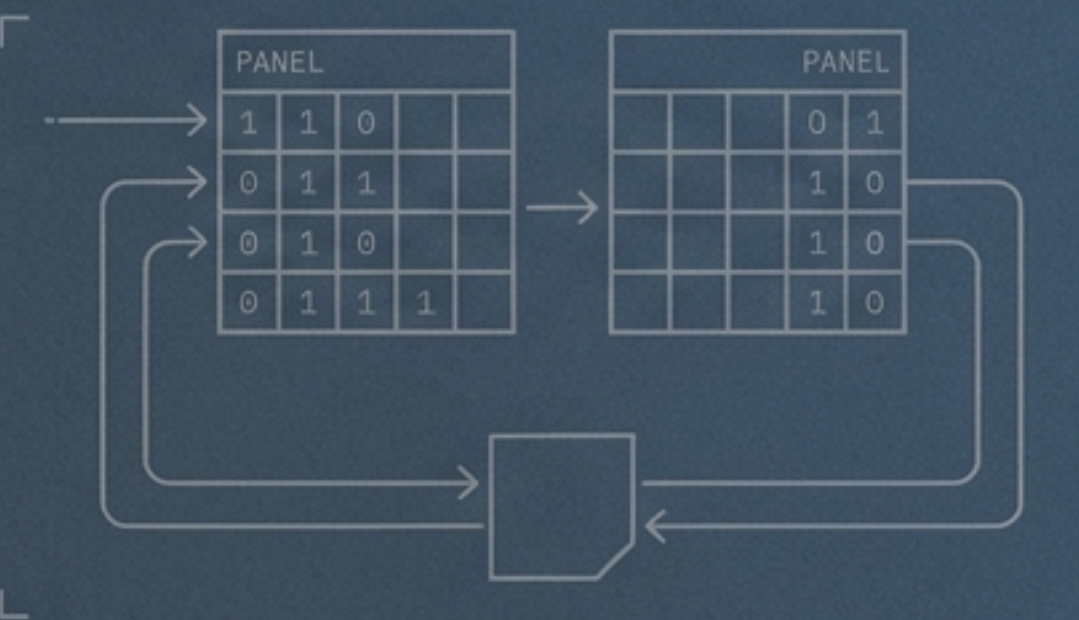


The Longer Leash

What an agent has to earn,
and who decides that it has.



STATUS: NOMINAL
SYS CHK: OK
ORBITAL PARAMETERS: CALIBRATED



Dale Martin

Dale Martin

The agent is stuck for a reason that has nothing to do with whether it can be trusted: the leash has no stated condition.

No stated condition (no sentence anywhere that says the boundary moves).



The Theater

- Approve every action by hand until the approval becomes a reflex.
- The agent has a supervisor who is no longer looking.

DATE: _____
TIME: _____
ACTION: _____
SUPERVISOR: _____

APPROVED BY: _____

APPROVED BY: _____

The Security Review

- Quietly stop approving.
- The agent has no supervisor at all.

DATE: _____ TIME: _____
STATUS: _____
REVIEWER: _____

REVIEWED BY: _____

REVIEWED BY: _____

You cannot hand an agent more rope when you cannot see what it does with the rope it already has.

The Noun

AGENT: [NAME]
BUILDER: [ORGANIZATION]
PURPOSE: [DEFINITION]
REGISTRY ID: [CODE]

- An agent's name in a registry.
- Tells you that the agent exists, who built it, and what it was meant to do.
- Tells you nothing about what the agent did.

LOGGED BY: _____

The Verbs

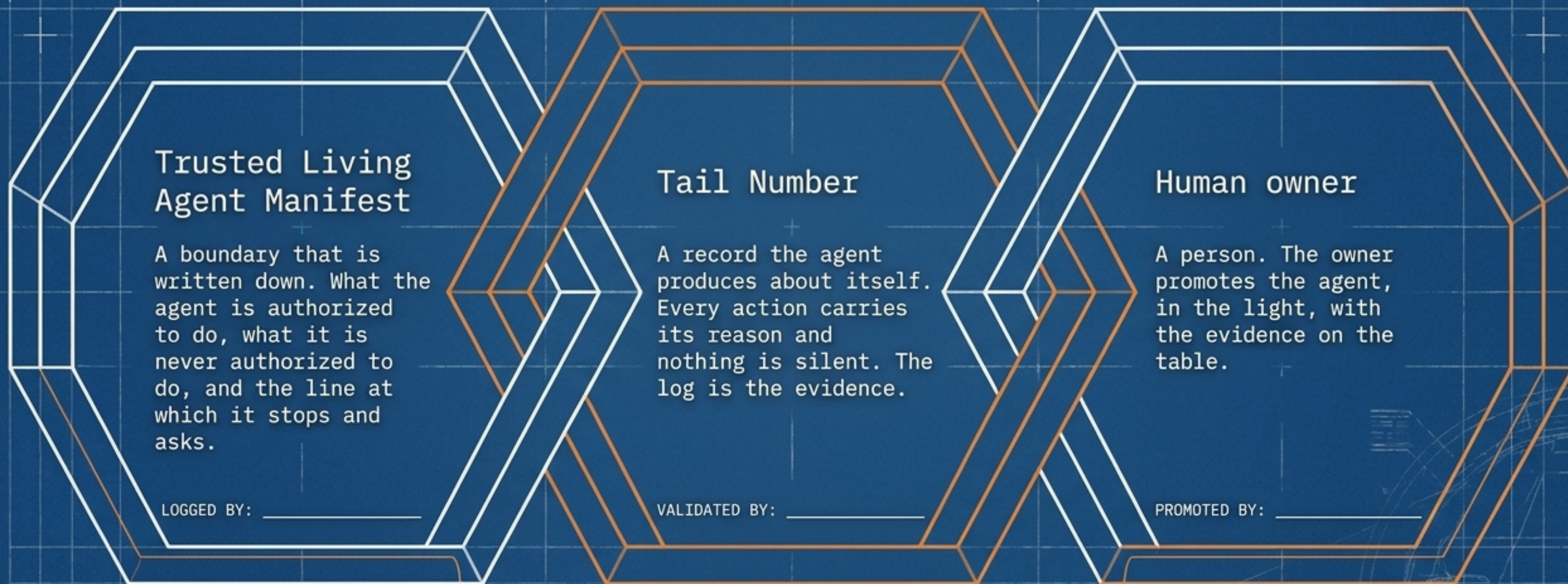


UNANTICIPATED
EVENT
T+2:47

- The record of what the agent did inside the boundary.
- A case came in that nobody had anticipated at 2:47 on a Tuesday morning.
- Trust rests on the record, and on whether that record was clean.

VALIDATED BY: _____

Trust requires an architecture where the governance travels inside the agent and not around it.



You don't earn authority in the dark. You earn it in the light, one shown manifest at a time.



LOGGED BY: _____

VALIDATED BY: _____

The test for a loop that looks like trust and is not.

Autonomy that is watched.

- Reading every log line as it lands, from the trailhead.
- Checking the output because you cannot stand not to.
- The leash got longer and you did not get released.

LOGGED BY: _____

The quiet failure.

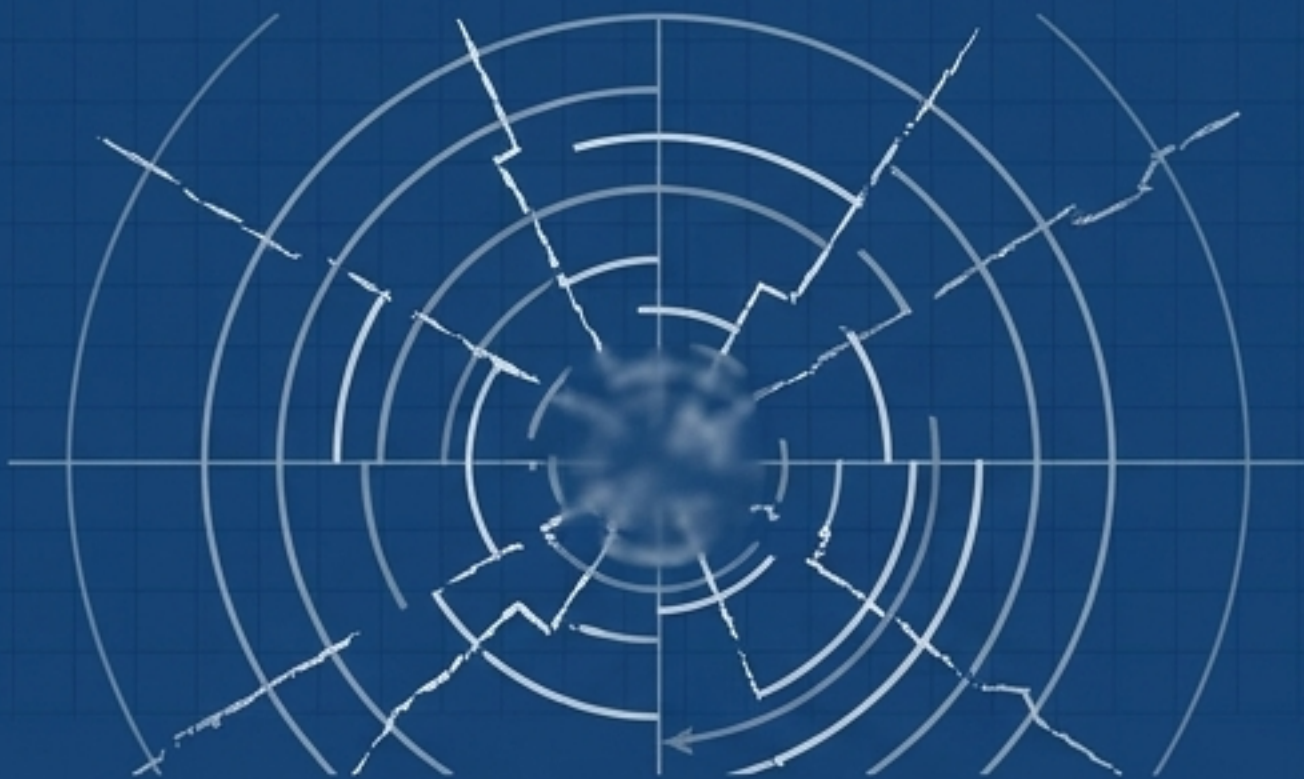
- The boundary never widens, the review never comes.
- A very expensive way of doing your own work.

VALIDATED BY: _____

The analog dividend.

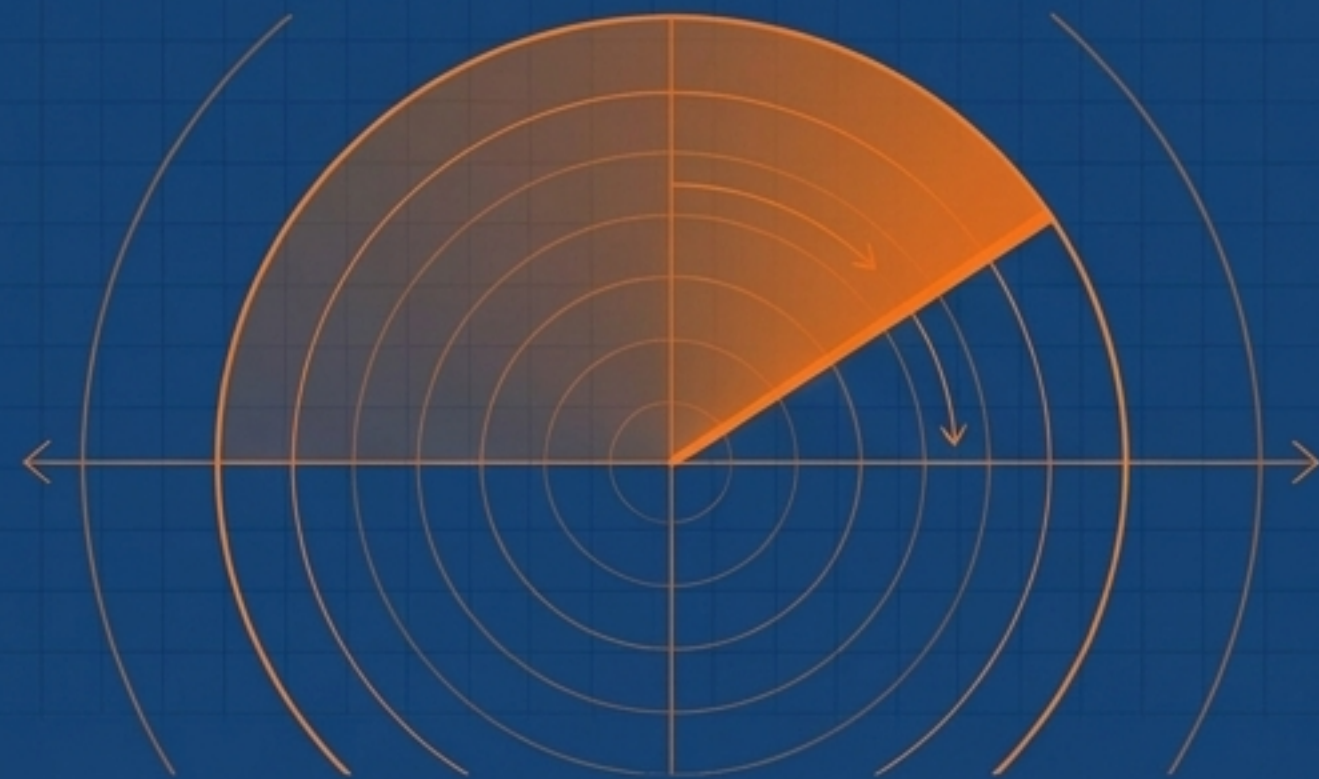
- The payoff of trusted AI measured in life rather than in output.
- The time that comes back off the screen once the work can be trusted to run without the person watching.

Governance is the right to delegate faster than anyone else.



The Competitor

Guessing or stalling.
Slower than looking.



The Organization that Builds the Seeing First

Looking.
Can widen what it delegates to them at a pace a competitor cannot match.
Can see, agent by agent, whether the trust is holding.

The pace at which you can widen a boundary is exactly the pace at which you can see what happens inside it.

A trustworthy agent is something a person can build, and you have already started building one.

15 Virtues

The pillars govern what an agent may do. The virtues describe what it should be.
A clean record is evidence of character as well as conduct.

The Workbench

Where the boundary, the record and the owner's line get written for real. Give the agent its boundary, set the date, and see what it does with the rope.



Everything rests on the agent's own account of what it did, and whether that account is enough on its own is a question left unanswered.