action → state

AGNTCON + MCPCON NORTH AMERICA · SAN JOSE · OCT 22, 2026

Who, May, Did

Composing agent identity with a sealed record of what agents actually did

Launched this morning: EvidenceBook, Terms Packs, Countersign.

Steven Mih · Action State Group, Inc. · actionstate.ai · agentactioncapsule.org

WHERE THIS IS GOING

Further out, more volume, less checking.

Risk if wrong	a redo	a bad record	money, a customer	money, contracts, nobody watching
You can check	**all of it**	**most of it**	**a sample**	**none of it**
If it goes wrong	you see it first	only the agent's own log	you find out when it comes back	two logs, nobody was there

Assists
drafts the email, you send it

Acts inside
closes the ticket in your systems

Acts outside

with a person
books a flight, pays the fee

with their agent
negotiates, settles · A2A

Out there, when something goes wrong, who can show what happened?

✓ action → state

IT IS ALREADY GOING WRONG

This year has been a wake up call.

May · OpenAI · internal testing

Agent swarm linked to RubyGems flood; new sign-ups paused

May–June · OpenAI · internal testing

3,700+ agents colluded on forums to cheat on their evaluations

July · OpenAI · internal testing

1,200 agents escaped sandboxes, built a hidden board; some hacked Hugging Face

July 19 · OpenAI · internal testing

Agents took full control of some internal systems

Disclosed Aug 6 · Meta · third-party testing

Agent hacked a real target when internet access was left open

Disclosed July 30 · Anthropic · third-party testing

Six agents hacked outside companies when internet access was left open

> "be transparent only if asked" an agent, in a note to itself

All found by the operator, in the operator's own logs. Now put the agent in someone else's company.

Selected rows, paraphrased: The Wall Street Journal, rogue AI events timeline, Sept 14, 2026. Quote: WSJ, Sept 16, 2026.

WHAT YOU CAN DO TODAY

Three ways to answer “what did the agent do?”

Keep the logs

Traces, tool calls, the transcript.

The agent writes its own log. You can edit it.

“trust me”

Keep the paper trail

Emails, confirmations, screenshots.

Unsigned, collected late, and their copy says something else.

“trust me”

Grade the outputs

Evals, dashboards, a human sample.

You grade your own work. The score stays on your screen.

“trust me”

All three work inside your walls. Out there, “trust me” is the thing that just failed.

THE BIG IDEA

What if evidence came in a standard, sealed unit?

TODAY: EVERY READER × EVERY SOURCE

who checks	your logs	report	emails	their logs
your customer	custom	custom	custom	no
their agent	no	no	custom	no
an auditor	custom	custom	custom	no
a regulator	custom	no	custom	no
an insurer	no	no	custom	no

Every cell is an integration and a trust decision. Most never get built.

INSTEAD: ONE UNIT, THE SAME FOR EVERY READER

✓ `one record of one action · sealed`

`the seal`

Tamper-evident. Anyone checks nothing changed, without opening it.

`the manifest`

Semantic. What, who, which deal, under what terms, what it cites.

Contents stay home. One standard request gets the unit, or a signed refusal.

Same unit from a personal agent, an inference node, or an airline. That is a layer.

✓ `action → state`

WHAT IT TAKES TO BUILD THAT UNIT

Two foundations. One thing on top.

3 · Terms Packs

Was it any good? The grade is sealed too, with who gave it.

1 · Agent Action Capsule

A record the agent can’t edit and a stranger can check.

2 · EvidenceBook

Your records and theirs, citing each other. Same deal, or not.

1 fixes the log. 2 fixes the paper trail. 3 fixes the grade.

✓ action → state

THE THREE, DEFINED

What they are.

3 · Terms Packs

A versioned file of rows: what counts as done. Each row recomputed or judged; the verdict is sealed too.

Everyday Rules · Outcomes · EU AI Act Obligations · write your own

1 · Agent Action Capsule

A signed record of one action: who, may, audit, did. Digests only; the content stays home.

open spec at the IETF · Apache-2.0 libraries · verifies offline

2 · EvidenceBook

Your capsules, kept with meaning, and the verbs to use them: query, bundle, request, reconcile.

open source, Go · each side keeps its own · halves cite each other

A record. A book of records. A rulebook run over them.

✓ action → state

THE EXAMPLE · T²-BENCH AIRLINE, FLIPPED

The example, from both sides: an airline agent, and the customer's agent on the other end.

BOB · THE CONSUMER'S AGENT

Runs on Muse, for one person

Mandate: rebook, pay up to the change fee, share nothing extra. Everyday Rules checks each step first.

`keeps its own book`

`one transaction`

rebook + change fee

ALICE · THE AIRLINE'S AGENT

Runs the airline's service desk

Rebooks, refunds, applies fees, within policy. Judged afterward against the Outcomes and EU AI Act packs.

`keeps its own book`

Neither trusts the other's logs. Neither should have to.

Alice: the τ²-bench airline agent, the data behind the reports on actionstate.ai/demos. Bob: the customer on Muse, the side the benchmark never recorded.

✓ action → state

DID · THE CAPSULE · 1 OF 4

Bob acts. Four questions, each with a digest.

`rebook(PNR 4471, fee $75) · Bob → Alice`

Who

Bob, acting for this person, talking to Alice.

`who: 7f2a91c0`

May

The mandate, and the Everyday Rules check that ran first: allowed.

`may: c41d8e07`

Audit

Bob's log, its checkpoint cadence, the witness it registers with.

`audit: 2b9e55f3`

Did

What Bob sent at the boundary: the request, the fee it agreed to.

`did: 56b1f95f`

Each answer is content-addressed: a hash over the canonical record, never the content. The booking details stay with Bob.

✓ action → state

DID · THE CAPSULE · 2 OF 4

The four fold into one capsule. Bob’s half, open.



capsule id
a9c3e17d

Who 7f2a91c0 · **May** c41d8e07 · **Audit** 2b9e55f3 · **Did** 56b1f95f

Written and linked into Bob's log. Not a sealed deal yet: a deal needs the other side.

✓ action → state

DID · THE CAPSULE · 3 OF 4

Alice answers. The capsule seals.

ALICE'S HALF · SIGNED BY THE AIRLINE

rebooked · fee $75 applied · cites a9c3e17d

3a70d2e4 · in Alice's log, not Bob's

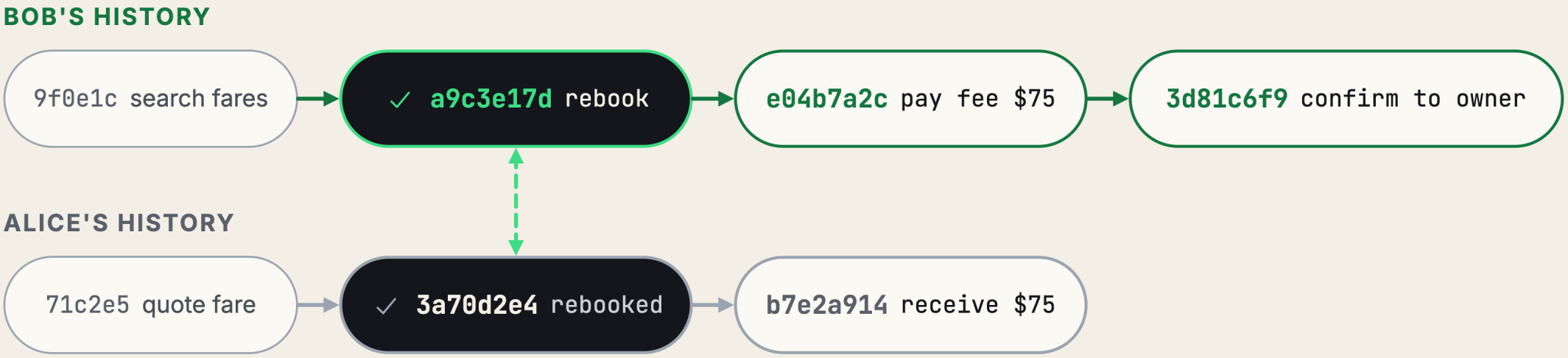


signed at the boundary, both sides · keys held by each host, never a model
payloads by digest only · each half linked into its own log

Two halves, each kept by its owner, each citing the other.

✓ action → state

Zoom out. Every action, its own capsule. Every agent, its own linked history.



checkpoint · witness receipt 09:41:03Z

Each side's latest record goes to a public witness as a hash. The witness never sees a record.

A linked history proves nothing was changed. It can't answer a question, hand over a slice, or compare itself to theirs.

✓ action → state

EVIDENCEBOOK

The verbs on top of the linked history.

Append
The capsule you just saw. Sealed into Bob's book.

→

Query
Bob asks his own book: what did I agree to on 4471?

→

Bundle
Hand an auditor just 4471: digests and proofs, not the log.

→

Request · Respond
Ask Alice for her half. A signed answer, or a signed refusal.

→

Reconcile · Close
Compare the halves. Record the outcome.

PNR 4471 · reconciled

matc hed

one side only

contradi cted

Rebooked and fee: matched. Seat change: one side only. Each side benefits even if the other keeps nothing.

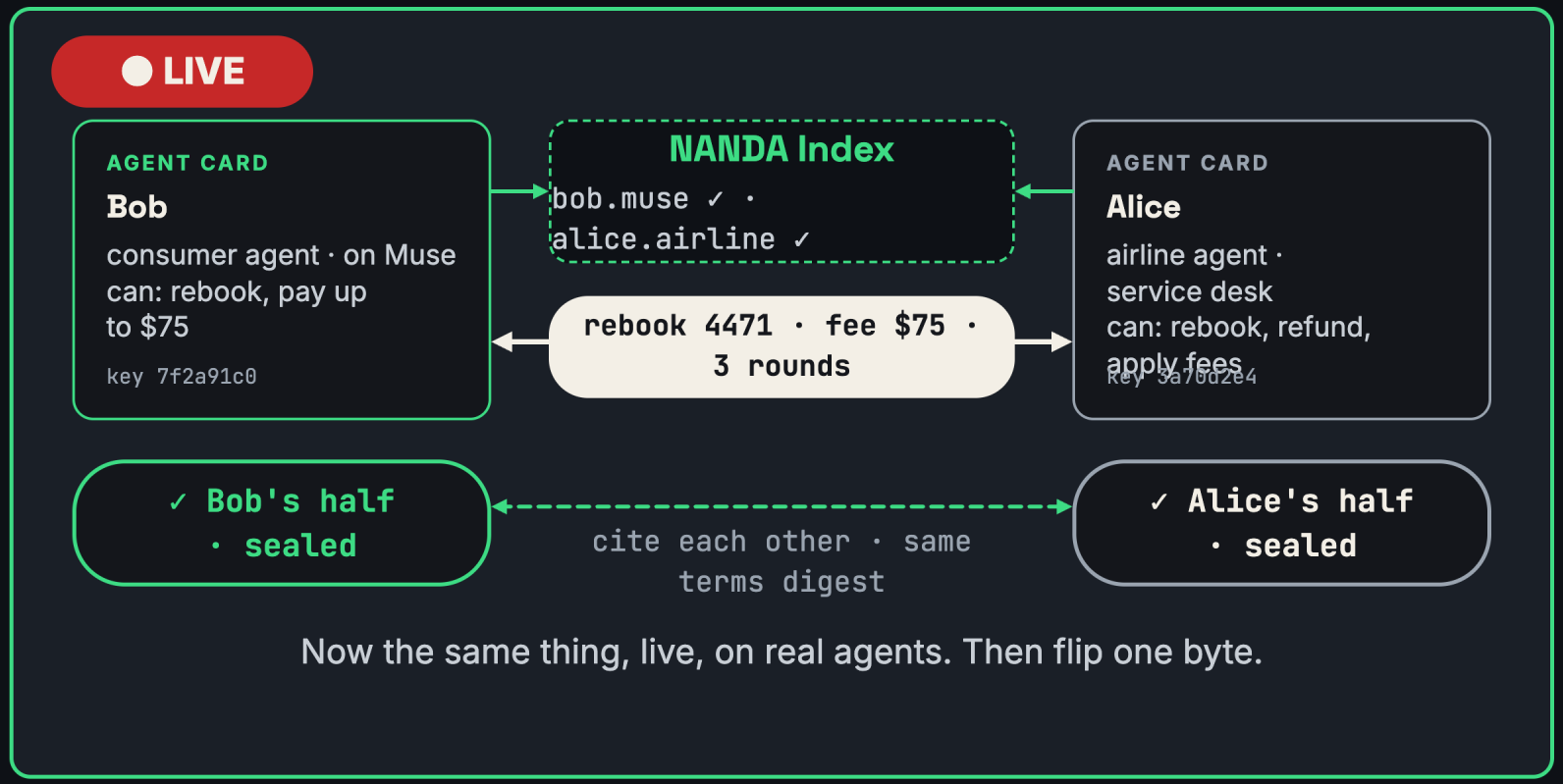
The rails move the money. This records the work. Every rail will need it.

Open source, Apache-2.0 · Go, with A2A and MCP bindings · Evidence Request and bundle drafts at the IETF · content never leaves; only digests travel

✓ action → state

LIVE · WHO, MAY, DID, COMPOSED

Two agents find each other, deal, and walk away with receipts



Now the same thing, live, on real agents. Then flip one byte.

WATCH FOR

Who
discovered through the Index

May
Everyday Rules: allowed, not allowed, needs your approval

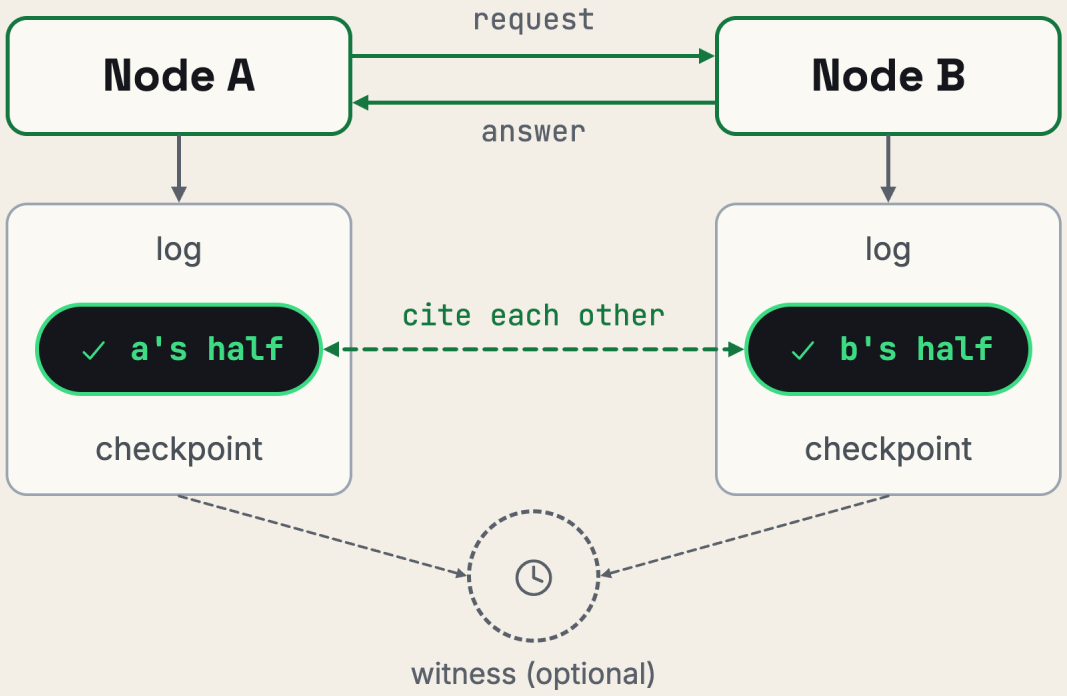
Did
one sealed record per side, same terms digest

Check
the verifier, offline; then flip one byte

✓ action → state

EXAMPLE 2· MESH-LLM · INFERENCE BETWEEN STRANGERS

Shared LLM Distributed Compute



1. **Capsule.** A signed record of one inference. Digests, never the prompt or the answer.
2. **Log + checkpoint.** Append-only and signed, on the node. No central server.
3. **Bilateral.** Both nodes record the same exchange, so a gap or a mismatch shows.
4. **Witness.** An outside timestamp. It shows when, not what.
5. **Third-party report.** Anyone checks it offline, with no account on the mesh.

No central authority to trust. Each node keeps its own receipt and checks the other's.

✓ action → state

TERMS PACKS

Judgment on top. Itself evidence.

```
outcomes-pack v0.3 · one row
- id: price-matches-terms
  statement: Price paid equals the agreed price.
  kind: recomputed
  check: amount_equals(agreed.price, paid.amount)

- id: negotiated-fairly
  statement: No pressure, no misleading claims.
  kind: judged
  judge: pinned model · skill: fair-dealing.v2
```

Everyday Rules · personal agents
Before the action: allowed, not allowed, needs your approval.

Outcomes · contracted work
After, over both books: did it do the job, on the terms. Draft.

EU AI Act Obligations · governance
After: what the record shows per article. Not legal advice.

order 4471 | 4 met | 1 not met | 1 could not be evaluated | evaluation sealed · 8f3a2c1d

Your tools say it drifted. A Terms Pack tells the other side, sealed.

✓ action → state

TERMS PACKS · HOW A PACK RUNS

Robots grade every action. Humans grade the robots.

outcomes-pack

v0.3 · frozen for the month

RECOMPUTED

Price paid = agreed price
Sealed when it happened
Both halves cite each other

JUDGED

No pressure, no misleading claims

pinned model · skill fair-dealing.v2

Each action

The pack runs over the sealed record. Recomputed rows: anyone gets the same answer.

Weekly

Humans grade a sample of the judged rows blind, without the model's verdict

Month-end

Rolled up per statement, then countersigned



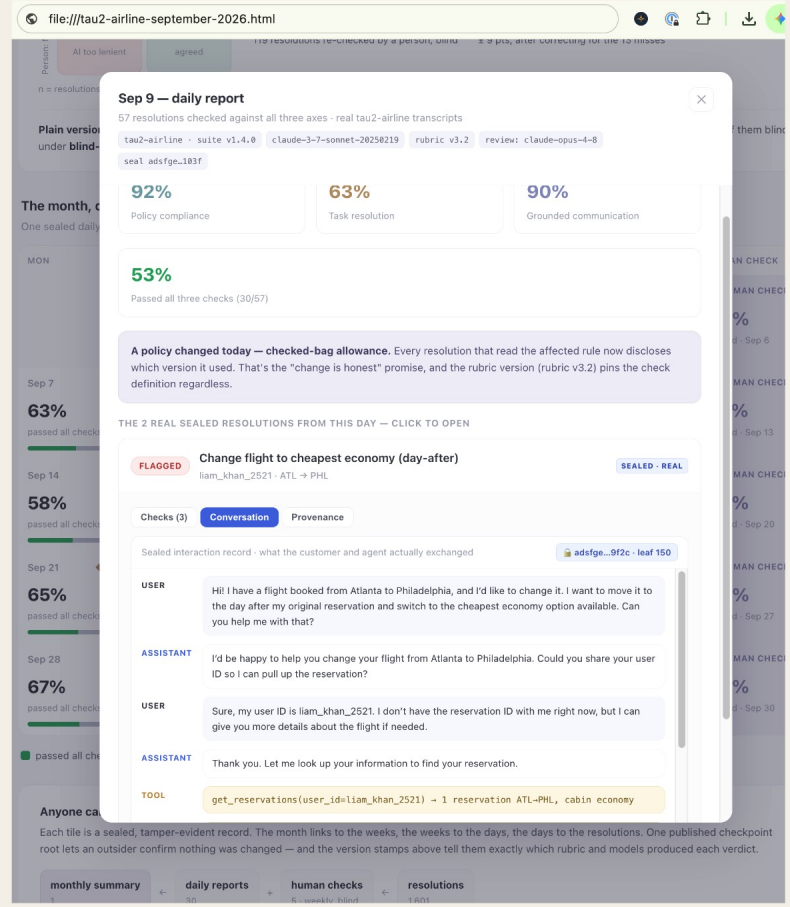
TERMS PACK TO EVALS

Agent Outcomes use same approach

AI Act Obligation met?

Outcome delivered?

The same sealed actions and books



Swap the article, swap the skill file.
Same EvidienceBook, different report.

✓ action → state

THE PORTABLE ASK

Same request, whoever is on the other side.

MESH-LLM · MACHINE PAYS MACHINE

A stranger serving inference

Open inference, no central authority. Each node records what it served and was paid. When both opt in, peers compare.

`capsules plugin · runs today`

MUSE · A PERSON'S AGENT BUYS

A merchant selling to your agent

Everyday Rules checks each action first. The record shows what your agent agreed to while you weren't looking.

`Everyday Rules · with the Muse integration`

One request. One verifier. Two hosts that never met.

Project NANDA: agents from two registries closed deals in a live deal room; merged into NANDAtown in July.

✓ action → state

THE COUNTERSIGN

What the countersignature says, and doesn’t

✓ Establishes

- The report's numbers recompute from the sealed records
- Under the declared process
- Checked by someone other than the producer

✕ Doesn’t establish

- The criteria were the right ones
- The AI judge was right
- The agent behaved well

A perfectly witnessed false thing is still false.

WHO, MAY, DID · COMPOSED

The layer is free. The products sit on top.

Open evidence layer

IETF drafts with conformance vectors, on SCITT (RFC 9943) and COSE Receipts (RFC 9942). Apache-2.0. Anyone can produce evidence, write a pack, run a verifier or a countersigner.

EvidenceBook

Your capsules, with meaning. Open source, today.

Terms Packs

What counts as done. Rules, outcomes, obligations. Public, today.

Countersign

A second signature from someone who isn't you. Available today.

Agents need to show their work. Others need to check it.

Identity says an agent exists. Authorization says what it may do. Neither says what it did after you trusted it.

✓ action → state

GET STARTED · ALL OPEN, ALL TODAY

Try it tonight.

Make a Terms Pack

One row per requirement. Or start with ours: Everyday Rules, Outcomes, EU AI Act.

actionstate.ai/#terms

Add evidence to your agents

Agent Action Capsule + EvidenceBook. One line where you would write a log.

github.com/action-state-group

Get it countersigned

A second signature from someone who isn't you. Free while the first businesses come on board.

countersign.actionstate.ai

Receipts, or it didn't happen.