

We sign our findings,

not our opinions.

Argus Agent Test Lab runs black-box adversarial batteries against the SDKs and platforms that move agentic money — x402 payment flows, credit and settlement handling, idempotency guarantees. The SUT is always the published artifact: npm dist, exact pin, unmodified. Every verdict is an Ed25519-sealed record chained to the hash of its predecessor in a public registry anyone can verify in one command. Seals attest run validity and findings completeness — not bug-freeness.

\$ node tools/registry/verify-registry.mjs →

registry intact: 10 records, chain head c26af630...

10 SEALED RECORDS	6 FRONTS UNDER TEST	60+ PRE-REGISTERED PROBES	0 FINDINGS RETRACTED	2 FIXES VERIFIED CLOSED
---	---	---	--	---

METHOD — FOUR RULES

- 01

Black-box by construction

The SUT is always the published artifact — no source access, no vendor cooperation required.
- 02

Pre-registered before the run

Required behaviors and verdict rules are fixed in committed matrix files before a single probe fires.
- 03

Offline by default

All traffic on 127.0.0.1 against adversarial stands; honest mocks modeled from the vendor's own contracts, each decision citing a vendor source line.
- 04

Sealed and hash-chained

Every record Ed25519-signed, chained to the previous hash, verifiable in one command.

VERDICT TAXONOMY

- FAIL

Vendor-side failure — the SDK loses money or integrity under the tested fault.

HOLD

The SDK rejected or limited the fault — a genuine guard fired.

PASS

The platform held beyond the documented scenarios.

OBSERVED

Behavior matches the vendor's published disclosure — honest posture, not a defect.

DX

Tooling hides a hazard the runtime itself flags; no direct money-loss path.

TRACK RECORD — SIX FRONTS

x402-foundation	reference client + PR #3920 core · 5 review rounds	A10 OPEN
Inngest	idempotency conformance · 13 probes, 0 FAIL	CLOSED
trip13tr3s	fix verified in v0.2.0 · 10/10 PASS · credited in SECURITY.md	FIXED
BlockRunAI	nano-client 0.8.0 · payTo substitution, ×50 no-cap, recovery dead-end	3 FAIL
Nevermined	payments 1.13.0 · retry double-charge, unpaid delivery, streaming w/o receipt	7 FAIL
AgentMiddleware	PayPal merchant agent demo · live round + clean-machine repro	REPRO'D

ENGAGE — TWO TRACKS

- A · Coordinated production round

from \$990 / round

Written scope, funded test wallet, 48-hour notice window, no-debit/refund arrangement, full request capture shared.

Deliverable: a sealed production-round report — clearing your endpoint, or documenting the exact boundary where defects become live money movement.
- B · Fix verification

per-finding seal

Our batteries become your acceptance rig: re-run the sealed suite against your published release, PASS/FAIL sealed per finding. A clean seal is an independently checkable artifact you can cite to integrators.

- \$ Fee agreed before any run; vendor responses get equal space alongside every finding.

\$ Private resolution welcome — the seal is not a weapon, it is a receipt.

\$ No vendor operates our stands; no engagement buys a verdict.