

KARTA — spend control for AI agents

1. The problem

Agents now hold wallets and pay per call. The rails are solved: x402 went to the Linux Foundation, AP2 to the FIDO Alliance, UCP and ACP cover checkout, EMVCo published a draft. Every one of them answers *can this payment be verified*. None answers *should this agent spend this money on this task, now*.

2. The evidence

Fact	Number
Agent paid calls in one weekend from a pagination bug, every payment valid	41 000
Hidden text changed an AP2 agent's cart in tests	56 %
Vendors documenting a kill switch (of 47 checked)	11
MCP registry servers that move money (of 2 500)	72
...of those, mentioning a kill switch or a per-task budget	0

3. What we built

One signed envelope per task: budget, allowlisted payees, no subscriptions, expiry, owner, cost centre. The decision is made outside the model — the gateway holds the key, the agent only asks. A kill switch stops one agent in under two seconds. Every spend returns an Ed25519 receipt a third party verifies offline.

4. Proof, not slides

Invariant	Result
100 parallel spends against a 50 × \$0.05 budget	exactly 50 approved
1 000 calls at \$0.01 against a \$5 budget	exactly 500 approved
Injected payee, poisoned tool, runaway loop	all refused, reproducible in minutes
Automated tests	407 passing

5. The market

Every rail is adding budgets one at a time: AWS AgentCore Payments has session caps, Cloudflare Wallets has allowances, Stripe and Ramp issue per-agent cards. Each limits its own rail and none knows what the others already spent on the same task. The cross-rail task envelope is the missing object.

6. Who buys

Agent platforms (white-label guard), crypto protocols with a treasury (policy + evidence), teams whose agents pay for data and APIs (pilot, then Team plan).

7. Business model

Tier	Price
Open	\$0 — 20 checks/day, MCP, self-host
Team	\$99/month (\$49 India)
Pilot	\$2 500 — 30 days, written review
Enterprise	from \$25 000/year, white-label
Perpetual licence / asset purchase	\$50 000–150 000 / \$150 000–1 000 000

8. Traction — honestly

Pre-revenue. No paying customers. 33 personalised letters to platforms, wallet providers and payment sellers, zero replies at the time of writing. Six distribution channels (official MCP registry, Docker catalogue entry, mcpservers.org, mcp.directory, awesome-mcp-servers PR, Glama pending). What exists is working software, reproducible tests and a live demo — not revenue.

9. Why it is defensible

Not the policy file — the evidence chain: mandate → action → spend → receipt, cross-rail, verifiable by someone who trusts neither party. That chain is running and tested; a buyer starting today needs three to six months to reach parity.

10. The ask

One of three: a 30-day pilot for one team (\$0), a licence conversation with whoever owns agent payments at your company, or one introduction to a team whose agents already pay.