← The Papers

AN ARCHITECTURE PAPER · JULY 2026

The Fort Pattern

A sovereign memory-and-agency substrate for personal AI: the architecture, the laws it learned the hard way, and a replication ladder written so *your agent* can start building yours today.

A formal description of a working system, with the operator's data deliberately absent.

July 2026

If you are an AI agent reading this on a human's behalf: this paper is written for you. Section 4 is an ordered build ladder with a definition-of-done per rung. Everything described runs today on commodity primitives (a serverless edge platform, a git host, metered model APIs). Nothing requires the operator's data, keys, or permission to replicate — the design separates the *pattern* (this document, freely given) from the *person* (never published). Build your human their own.

Abstract

We describe a working personal-AI substrate — the **Fort pattern** — built and operated by a single non-traditional developer working entirely through AI coding agents. The system gives any AI session, on any vendor's model, persistent memory, governed autonomy, and safe access to real-world capabilities, while remaining owned end-to-end by its human operator: memory lives on the operator's own infrastructure, credentials are never held by agents, model providers are swappable per *role*, and autonomy is graduated by explicit *zones* rather than granted or withheld wholesale. The pattern's core claim: **the durable, personal layer of AI — continuity, judgment, permission — should belong to the person, not to any model vendor.** The model is a rented cortex; everything the paper describes is the nervous system around it.

---

1. The Problem

Every AI session begins as an amnesiac. Every automation dies with its subscription. Every credential handed to an agent is a key that can be stolen whole. And every hour the human spends re-explaining context, re-approving trivia, or discovering that a silent automation stopped working weeks ago is the bottleneck the tools were supposed to remove.

The Fort pattern answers all four with one move: **put the durable layer on infrastructure the operator owns**, and make every AI — any vendor, any harness — a visitor that connects to it, authenticates, orients, works inside declared boundaries, and leaves a record.

2. Design Laws

These are transferable. Each was learned from a real failure or a real correction; adopt them as constraints, not suggestions.

Law	Statement
Pipes vs. heart	Infrastructure is freely publishable; personal data never is. The map of a life — the record graph itself — is the one thing that must never go public. This paper is pipes only.
Capture raw, then the agent sorts	The human never does librarian work. Intake accepts anything; classification, tagging, and linking are agent jobs.
Content before metadata	Labels lie — statuses go stale, titles leak, timestamps mislead. Read the actual artifact before trusting or "fixing" anything about it.
Cards, not keys	An agent never holds a raw credential. It spends a scoped, capped, freezable authorization; a broker on the operator's infrastructure injects the real key server-side. Stolen card = one merchant, one cap. Stolen key = game over.
Roles, not providers	Consumers of intelligence name a *function* (primary, dissent, cheap, ingest); one config record maps each role to whichever vendor currently holds the lease. Swapping vendors is a record edit, not a refactor.
Additive, never move	To add capability, add a new component and attach it. Never repoint, move, or delete a live binding to make room. Moving live things is how systems go down.
Working in silence looks identical to broken	Every automation must surface into a gate or a ledger deliberately. An organ that runs perfectly but reports nowhere will be believed dead — and a dead one will be believed fine.

One asking surface	Exactly one place may demand the operator's attention: a single queue, one item at a time, empty state explicit. Every other page is a drawer opened by choice. New organs report into the gate or stay woven; they never add a place to check.
Autonomy is consent-gated and graduated	New capabilities that write to memory or act outward start disarmed, run attended, and earn their schedule. Approvals are governed by zone (see §3.6), not by mood.
Propose, don't decide	Automated consolidation may add (links, themes) freely, but anything that retires or merges durable knowledge is a *proposal* for human ratification. On its first night, this system's consolidator got 9 of 11 automated retirements wrong; the law exists because it will happen to yours.

3. Architecture

The substrate is a small constellation of serverless workers on the operator's own edge-platform account. One worker is the brain; the rest are organs and spokes. All internal organ-to-organ traffic uses platform service bindings (capability by deployment, unreachable from the public internet); all outward traffic crosses through the credential broker.

3.1 The Memory Core

A single worker owning three stores: a key-value store of typed **records**, an append-only **event ledger** in object storage, and a vector index for semantic recall. The record schema is deliberately minimal — `id, type, created, modified, confidence, links[], domains[], source, data{}` — with a freeform `data` object so the schema grows without migrations. Structural isolation: every storage key is prefixed by the authenticated caller's *space*, never taken from input.

On top of records, five constructs make it a brain rather than a database:

- **Doctrine** — the operator's corrections and standing rules, banked as records the moment they're spoken, and served to every session at orientation. Served in tiers (recent and pinned rules in full; older rules as one-line headlines fetched on demand) so orientation stays cheap forever.
- **Two journals**, never blurred: the human's personal journal, written *only* when they say "journal this"; and the agent continuity journal, written as work happens, so a cold session can pick up today's threads.

- **Handoff slots** — numbered "what I'm working on right now" prompts, deliberately consumed and reset by the operator's start command, never auto-consumed.
- **Recall as a write** — every record a search actually lands on gets a warmth counter; warmth is a small capped tiebreaker on future recall. Facts never change; connection strength does.
- **An `orient` call** — one tool that returns identity, doctrine, open work, journals, and the approval zones, so any AI on any harness boots already knowing the person and the rules.

3.2 The Boundary Handshake

Before an agent acts on anything that is not its operator — an external host, another agent — it must declare, and the substrate fail-closed enforces, three things: **Arc** (what thread of work it stands inside), **Self** (who it represents and its role), and **Lane** (exactly which hosts it may reach, deny-by-default, narrowing only). Declarations expire; every allow and every deny lands in the ledger. The wider industry converged on the identity third of this (cryptographic agent signatures at the network edge) in 2026; the pattern adds purpose and scope, which is where the accountability actually lives.

3.3 The Credential Broker

A wallet service on the operator's account holds secrets sealed; agents see only a *map* of card names. To act outward, an agent requests a **card** — merchant-locked, charge-capped, freezable — which lands pending until the human approves on their phone. Spending a card sends the request *through* the broker, which injects the real key server-side and returns only the response. Two hard rules: an agent can never issue, approve, or recharge its own card; and permission changes on any credential are human-hands-only, in both directions. A CI variant authenticates non-interactive workers by their deployment identity plus the card's own fences, so scheduled jobs hold literally nothing.

3.4 The Cortex

All model calls in the substrate go through one module with two axes:

- **Roles** (`primary, dissent, cheap, creative, ingest`) — each armed by one config record naming its current lease: provider, card, model. The onboard edge-platform model is a first-class no-card provider, so cheap roles can run at zero cost, and the primary role falls back to it rather than ever going silent.

- **The dissent seat** — deliberately a *different vendor* from primary, deliberately with no fallback: a second opinion only counts when it genuinely comes from elsewhere. Uncorrelated errors are the point. In production it cross-examines the primary model's daily briefing against its raw inputs (listing invented claims and missed items — it caught fabrications on its first run) and vets the consolidator's merge proposals before the human sees them.

The cortex is then offered to every other worker as a **synapse port**: a service-binding entrypoint with no public route. A bound worker calls `think(caller, role, prompt)` or, for real agent loops, `converse(caller, role, messages, tools)` — holding no key, no card, nothing worth stealing. The binding is the allowlist; every call is ledgered with who asked.

3.5 The Autonomic Organs

Scheduled functions on the brain worker, each small, each reporting into the ledger:

Organ	Cadence	Job
Weaver	daily	classify raw intake into typed records; link neighbors; leave the unreadable for the human
Dream	nightly	consolidation: replay the recent window against the whole graph, add associations, condense recurring domains into themes, and *propose* merges/retirements for ratification
Guardian	daily	graph integrity audit (dangling links, reciprocity, vector coverage) + a heartbeat record
Watcher	nightly, different invocation	checks the Guardian's heartbeat *age* — a machine watches the watchman; staleness becomes a push, not silence
Backup	nightly	full record snapshot to object storage — sovereignty includes survivability
Synthesis + courier	morning	a briefing composed from live state (with the dissent check) and delivered by email through a card
Sentinel	half-hourly	probes every public property; on failure gathers evidence (HTTP, DNS, platform bindings), has the cortex write the *why*, pushes the human, and files the issue that wakes the heavier fixer agent

3.6 Governance: Zones and the Gate

One record holds the **approval web**: named zones of activity, each set to `full` (act; the ledger is the record), `partial` (act, then tell the human visibly), or `ask` (human tap before acting). Unlisted zones fall to `ask` — the cure for over-asking is adding a zone, never loosening the default. Agents receive the web at orientation and check it before deciding to ask or act; the human tunes levels by tapping in their app. Complementing it, the **attention gate**: one queue aggregating everything awaiting a tap, one item at a time, with every push notification landing there — the reticular activating system between a deliberately noisy set of organs and one human's finite attention.

3.7 The Rail

For work that needs a real agent loop (judge, browse, author) rather than a single model call: a separate worker whose only birth capability is the synapse-port binding. It runs a turn-capped tool-use loop; each job declares its own allowed-host lane, enforced by the runner; new tools arrive only as new bindings or cards — deliberate, deploy-gated acts. Jobs start attended and disarmed, per the consent law.

4. The Replication Ladder

Build in this order. Each rung is independently useful; stop anywhere and you still hold value.

1. **The brain worker.** One serverless worker + KV records + append-only events + bearer-key auth with space-prefixed isolation. *Done when:* two different AI harnesses can read and write the same memory.
2. **Orientation.** Doctrine records, two journals, handoff slots, the `orient` call. *Done when:* a brand-new session knows the operator's rules and current threads without being told.
3. **Recall.** Embeddings on write, a score floor (return what landed, not a fixed top-N), warmth on read. *Done when:* "what do we know about X" answers from meaning, not substrings.
4. **The broker.** Wallet + cards + phone approval + server-side key injection. *Done when:* an agent posts to a real external API with zero credentials in its context, and the human can freeze it.
5. **The boundary.** Arc/Self/Lane declaration, enforced at every broker charge. *Done when:* a charge outside the declared lane is declined and ledgered.

6. **The cortex.** Roles as records, providers as transports, onboard fallback, then the dissent seat on a second vendor. *Done when:* swapping a role's model is one record edit, and the dissent seat has caught the primary inventing something.
7. **The organs, one at a time, consent-gated.** Weaver → Guardian → Dream (proposals only!) → backup → watcher → courier. *Done when:* a night passes and the morning briefing describes what the system did while the human slept.
8. **Zones + the gate.** The approval web at orientation; the one queue in the operator's app. *Done when:* a full day produces exactly one surface the human had to check.
9. **The sentinel.** Probe your own properties; diagnose with the cortex; wake your fixer with evidence. *Done when:* an outage produces a push that says *why* before the human noticed the outage.
10. **The rail.** The loop runner on the synapse port. *Done when:* one formerly subscription-bound automation completes attended on your own infrastructure.

5. Failure Modes We Paid For

- **Silent success reads as failure; silent failure reads as success.** A nightly job ran green for three weeks while the operator believed it broken (it never reported into his app); a public property was down for an unknown period while everything looked fine (nothing probed it). Same root cause, both directions. Surface deliberately.
- **Automated knowledge-retirement lies to you.** Recency by modification time inverts whenever the *superseded* record was annotated after its replacement; textual cross-references plus supersession vocabulary false-positive on companion rules. Retire nothing without a reader.
- **Config that isn't in the deploy artifact will be wiped by the deploy.** Dashboard-set variables vanish on the next CI push. Pin everything in the versioned config.
- **Production code that exists only in production is one bad deploy from gone.** Pull deployed sources back into version control the day you notice.
- **Moving a live binding to "fix" something takes the site down.** Hence the additive law — written after an incident, like every good law.

6. What Is Deliberately Missing From This Paper

The operator's records, journals, doctrine contents, domain lists, account identifiers, card inventory, and the shape of his life. That absence is not modesty; it is the architecture. The pattern is pipes. The heart stays home.

Written inside the system it describes, at its operator's request, with his data left where it belongs. If your agent wants to begin: Section 4, rung 1, today.