

# REMAP

MEMORANDUM

DATE: June 10, 2026  
FROM: Nicholas F. Casal  
TO: REMAP Partners  
SUBJECT: AI Operating Framework

**Overview.** We run AI across the core of our business, part on hardware we own and part on frontier models we rent. Our agents source acquisition opportunities from records open to all and synthesized by almost none; they aggregate the fragments into a single picture and interrogate the whole for signal no record holds alone. We have built a skills library that encodes our underwriting and memo conventions, so the scaffolding that once took a morning now takes minutes, and we spend the difference on judgment. For sensitive documents, we run open-source models locally and keep the material off servers we do not control.

**Hybrid Infrastructure.** We route work by where it must run: the routine and the sensitive stay on a local headless Mac Studio, running closer to the metal; we direct hard reasoning out to frontier models, which we pay for by the token. We have designed the system so the compute node cannot reach back into the operator’s workstation, and every agent runs sealed in a least-privilege container, discarded at session close. Containment bounds what broad, unattended permissions can touch.

| Layer              | What it is                                                                     | Role                                                                               |
|--------------------|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Control node       | Windows desktop PC (operator workstation)                                      | Directs the agents and reviews output; reaches the compute node over SSH.          |
| Compute node       | Headless Mac Studio, M3 Ultra, 96 GB unified memory, always on                 | Hosts the local model runtime and the agent containers, unattended.                |
| Local network      | 2.5 GbE managed switch; encrypted WireGuard mesh (Tailscale) for remote access | Keeps node-to-node traffic on the LAN; no public ingress.                          |
| Trust boundary     | One-directional ACL, control node to compute node only                         | The compute node cannot initiate back to the workstation. Bounds the blast radius. |
| Local inference    | Open-weight models on the GPU via a local runtime (Ollama), bound to loopback  | Runs routine and sensitive work on-box at no per-call cost.                        |
| Frontier reasoning | Claude, via the Claude Code CLI; subscription plus token-metered API           | Hard reasoning we deliberately route off-box.                                      |
| Containment        | Per-agent isolated container, least-privilege mounts, ephemeral                | Torn down at session close; bounds what one agent can reach.                       |

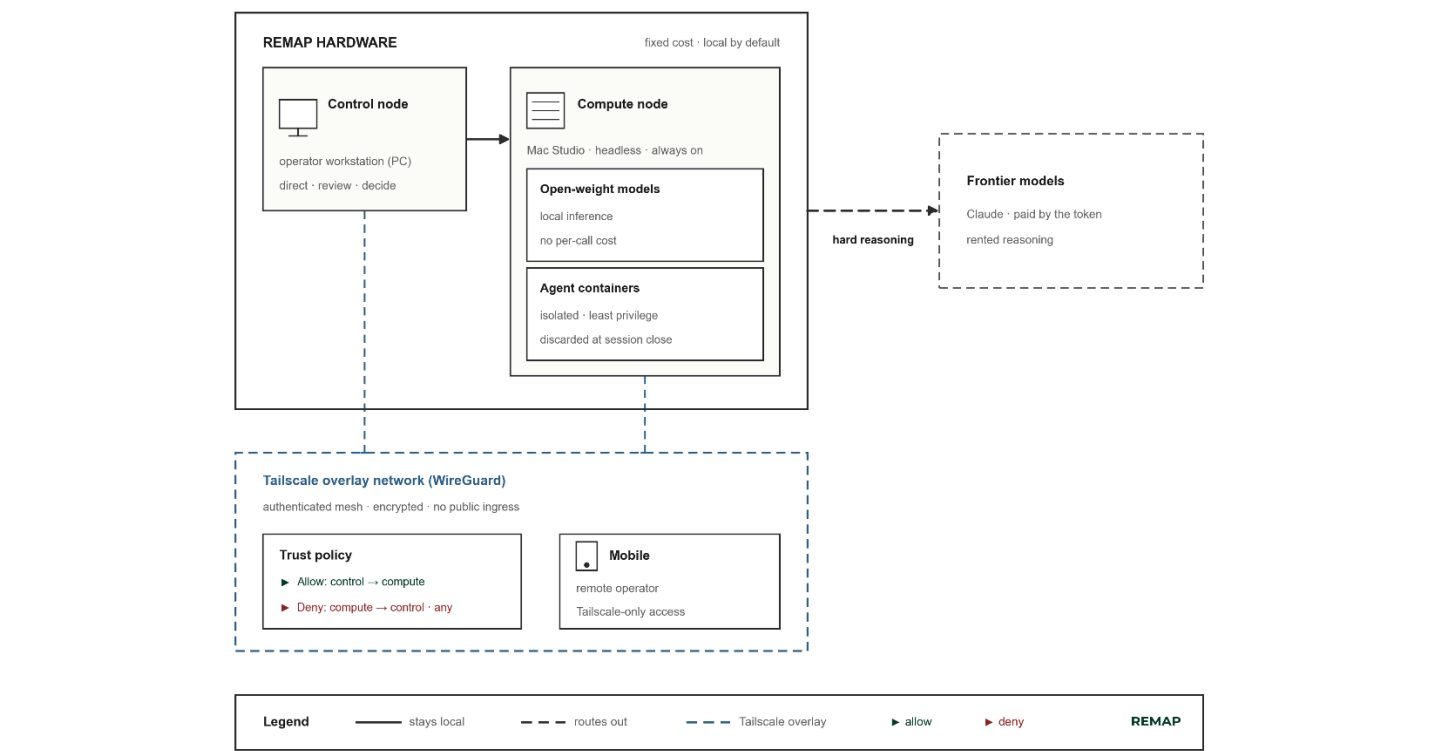
**Agent Deployment.** Our agents assemble and scrutinize property intelligence, which we pair with classical sources of alpha. This work was technically feasible before AI, but the ROI of doing it by hand never penciled. AI commoditizes intelligence. In the regime that follows, advantage accrues to high-agency firms that wield silicon intelligence with superior execution and judgment.

The stack we are currently utilizing is standard and available to anyone; nothing in it is proprietary.

| Tool        | Role                    |
|-------------|-------------------------|
| Claude Code | Agentic coding platform |
| VS Code     | Code editor             |
| Warp        | Terminal                |
| SQLite      | Embedded database       |
| Git         | Version control         |
| GitHub      | Code repository         |
| Obsidian    | Knowledge management    |

**Skills.** We deliberately built skills for recurring work that consumes the most analyst hours: our underwriting models and memos. Each skill encodes our conventions once and applies them in full to every model and memo that follows, so the work takes less time and the output is consistent.<sup>1</sup>

**Local models.** We run open-source models on our own hardware for three reasons. First, control: sensitive documents stay inside our environment and never leave. Second, cost: the hardware is a fixed, one-time outlay, so routine high-volume work runs at no per-call charge. Third, and less obvious: for bounded, well-specified tool calls, the open models we run ourselves are already frontier-class, and the gap opens only on long-horizon, multi-step reasoning, which we route to a frontier model by design.



**Building.** Every future REMAP hire will spend a meaningful share of their time in Claude Code, Codex, or whatever succeeds them. You learn what a tool can do by building with it, and the lessons compound. We are in the arena, learning as we build; we build to win.

[ End ]

<sup>1</sup> A skill is a modular, reusable package of instructions, with optional code and reference resources, that an agent invokes when a task calls for it, so a defined procedure runs the same way every time instead of being reconstructed from scratch.