



PRODUCT CASE STUDY • PERSONAL SOFTWARE • SELF-HOSTED AI

personal-os

An operating system for one human.

A native iOS super-app backed by a private AI cloud — local language models, a personal RAG assistant, and ~15 self-hosted services. Designed, built, and operated by one person, for one user.

Why build your own cloud?

● Privacy

Journals, health data, finances, conversations — the most intimate data a person produces. Here it never leaves owned hardware.

● Capability

Frontier workflows — retrieval-augmented chat, speaker diarization, vision, voice — running on local models, tuned to one life.

● Craft

Commercial apps average their millions of users. Software for exactly one user can be shaped with a tailor's precision.

The bet: in the AI decade, the most personal computer is one you run yourself.

THE ARCHITECTURE

Three tiers, one system.

THE APP

● Native iOS client

16 native features in a SwiftUI super-app — opens into an ambient, time-of-day 'Lounge', not a grid of icons.

THE CLOUD

● VPS · public tier

~15 containerized services behind a reverse proxy with single sign-on — reachable anywhere, owned entirely.

THE BRAIN

● Home GPU server · 'Helios'

Local LLMs (Ollama, 14B), speech-to-text, speaker diarization, embeddings — the private AI tier. No API keys to anyone.

Docker Compose · Caddy · SSO · EventKit on-device · SSE streaming — every layer hand-assembled.

THE CENTERPIECE

An assistant that has actually read your life.

YOU what did I decide about the school website copy?

ASSISTANT You chose the 'to kindle' draft as the primary voice — warmer, parent-first. Cited: Website Copy v0.1, brand-voice notes.

YOU remember: revisit fee-framing copy after conjoint results

ASSISTANT Saved to memory.

- **Streaming RAG chat**

Token-by-token answers over a personal knowledge vault, with citations back to the source notes.

- **Local 14B model**

qwen2.5:14b via Ollama on the home GPU — zero cloud AI touches private data.

- **Persistent memory**

“remember: ...” writes durable memories the assistant carries forward.

THE CONSTELLATION

Sixteen native features. One life, fully instrumented.

● Meeting recorder + live speaker labels	● Nutrition + Whoop recovery	● Voice-to-calendar (speak a plan)	● Streaming vault assistant
● Expense tracker	● IPO tracker	● Personal CRM (people)	● Book scanner (shoot a cover)
● RBI + markets dashboard	● Journal pipeline	● Reflection templates	● Coffee brew guides
● Piano practice	● Tasks (Vikunja)	● Articles archive	● The Lounge (ambient home)

Each one replaces a subscription, a spreadsheet, or a habit that never stuck — unified under one design language.

THE DESIGN IDEA

Software that greets you like a room, not a grid.

● DAWN

cream → peach

● DAY

light → sage

● DUSK

amber → coral

● NIGHT

deep teal → ink

The app opens into a **Lounge** — an ambient home whose gradients, greeting, and suggestions shift with the hour. It surfaces the unhurried things: coffee, a calm minute, reflection, piano.

The philosophy runs through the whole system: no badge counts, no streaks, no guilt mechanics. **Nothing here ages, nothing is late.**

THE OPERATING MODEL

One builder, plus an AI workforce he manages.

- **The owner**
Product decisions, architecture, taste, deploys — and every line reviewed before it ships.
- **AI build partners**
Claude-class models as pair engineers across iOS, Python services, and infrastructure.
- **A background agent**
“computer-ji” — an autonomous coding agent running on the home server, handling deploys, smoke tests, and pipelines while the owner sleeps.
- **The paper trail**
Every session logged in handoff documents so any future collaborator — human or AI — starts with zero context loss.

This is what an AI-native, one-person engineering org looks like in practice.

WHAT THIS PROVES

The super-app for one user is the new personal computer.

Full-stack range

iOS, Python services, GPU inference, networking, design systems — one person, production quality.

Private-AI conviction

A working demonstration that frontier AI workflows don't require surrendering personal data.

Compounding asset

Every service added makes the assistant smarter and the system more valuable. It appreciates.