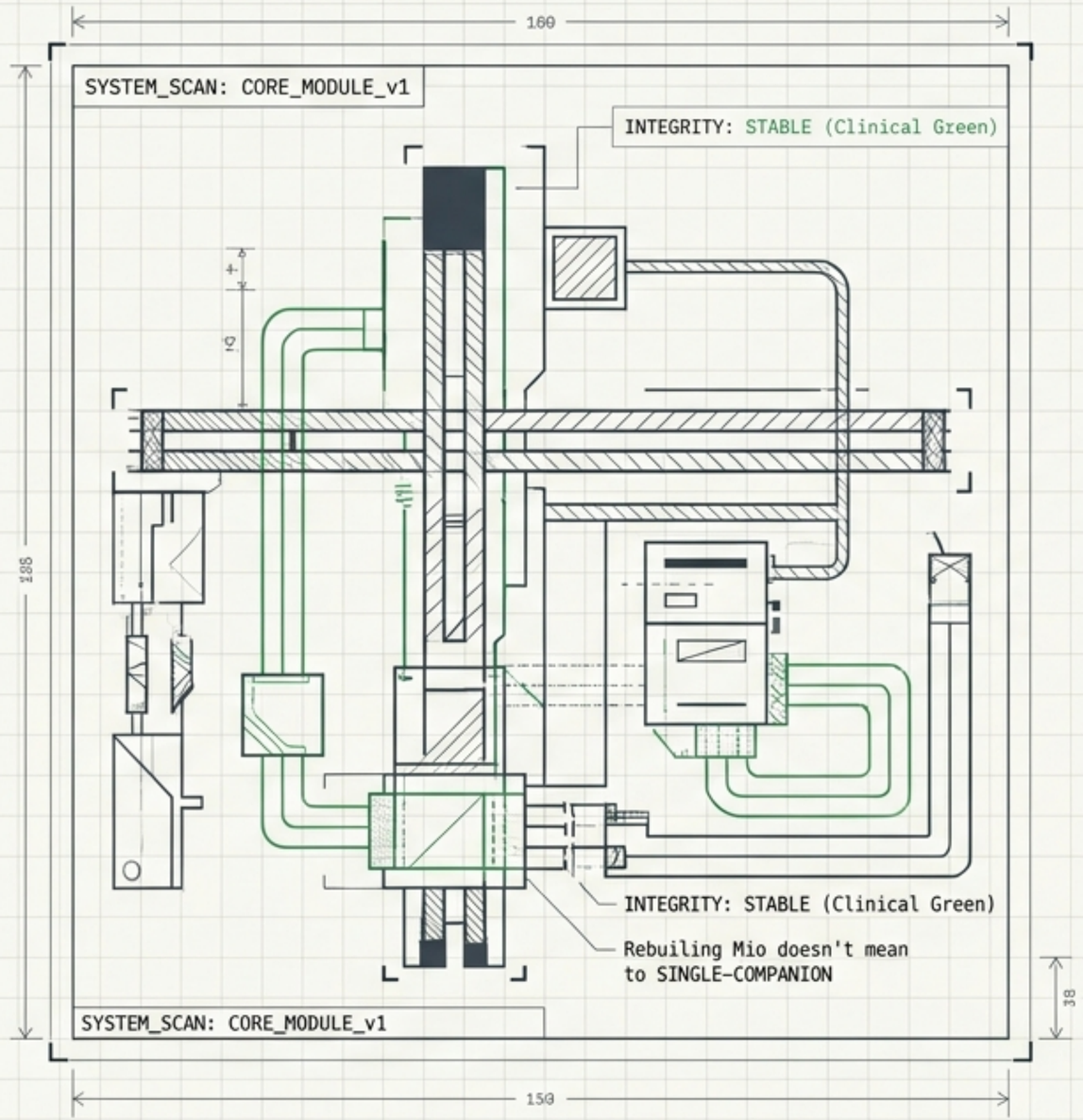


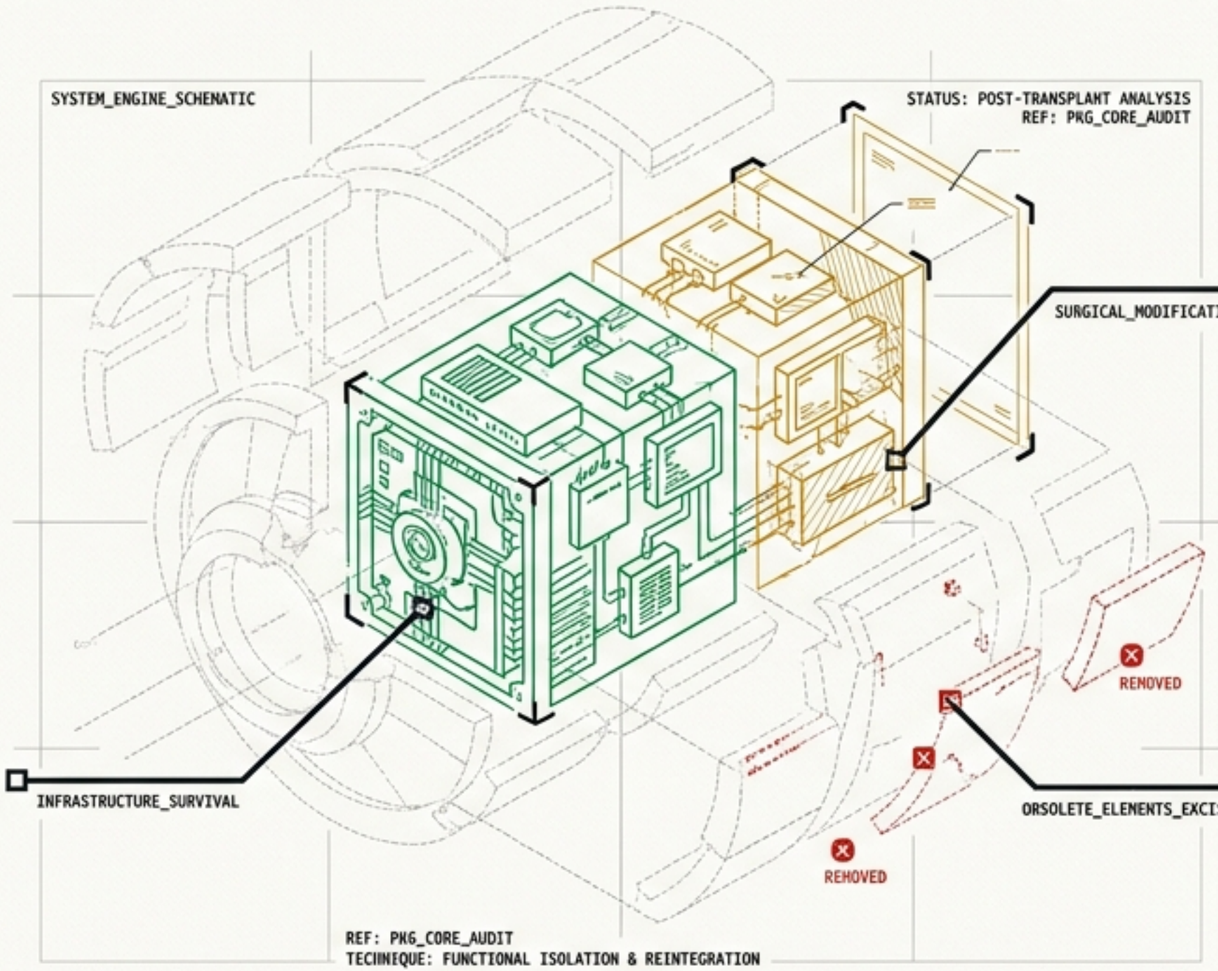
TARGET: packages/core
STATUS: Refactoring
DATE: 2026-03-05

Technical Autopsy: What Survives v1

Rebuilding Mio doesn’t mean rewriting everything. The clinical transition from multi-persona to single-companion.

A Transplant, Not a Rewrite

When tracing every import path in `packages/core`, most modules proved persona-agnostic. They accept parameters and return results. The functions don't care who is speaking.



15%

Logic requiring surgical modification.

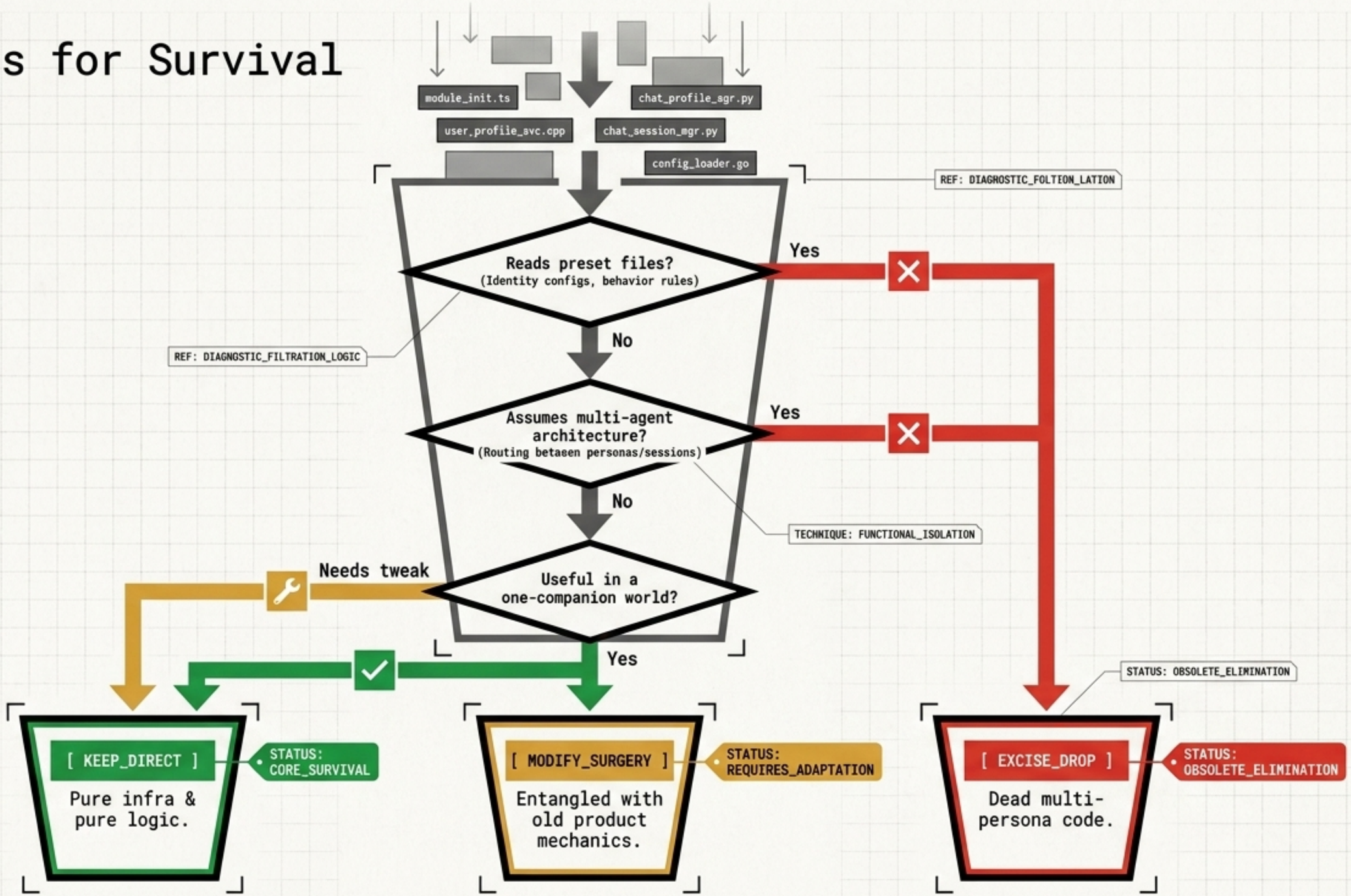
60-70%

Pure infrastructure directly carried over.

15-20%

Obsolete baggage dropped entirely.

3 Filters for Survival



The Core Vitals: Infrastructure

REF: SYSTEN_MENORY_CORE_VITAL

Memory System (packages/core/memory/)

Input: user_id + content
Output: extraction, 768-dim vector embeddings, semantic search via pgvector, importance scoring, decay. Zero preset coupling.

REF: SYSTEM_MEMORY_CORE_VITAL

STATUS: KEEP_DIRECT_SURVIVAL

TTS (tts.ts)

Input: text + voice ID → **Output:** audio.

REF: TTS_NODULE_VITAL

Vision (vision.ts)

Input: image → **Output:** description.

REF: VISION_MODULE_VITAL

Transcribe (transcribe.ts)

Input: audio → **Output:** text.

REF: TRANSCRIBE_MODULE_VITAL

Browse (browse.ts)

Input: URL → **Output:** page content.

REF: BRONSE_MODULE_VITAL

The whole media pipeline was designed as pure functions. That decision pays off now.

The Core Vitals: Business Logic

STATUS: VITAL_KEEP

REF: LOGIC_LEDGER

LEDGER_CARD_VIEW: CORE_LOGIC

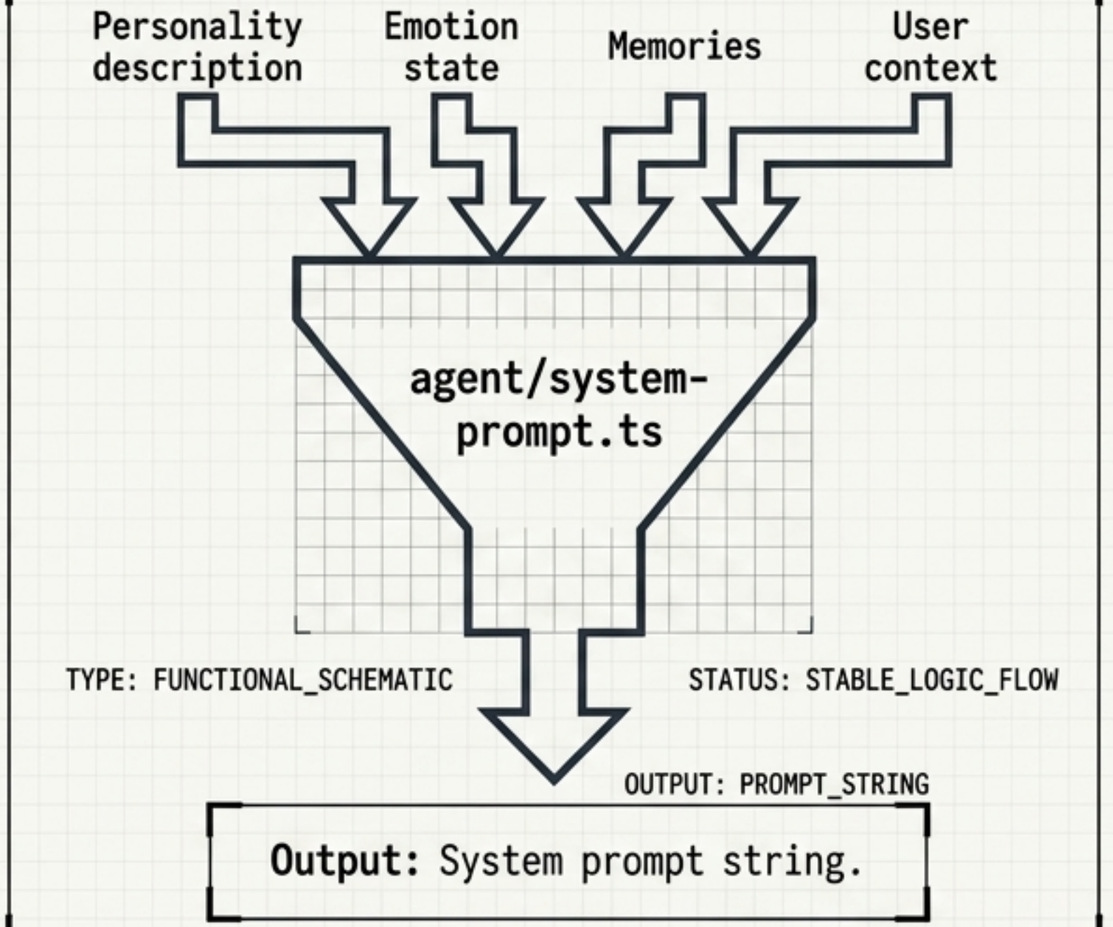
✓	**Cost Tracking (core/cost/):** Records LLM calls, TTS synthesis, vision analysis with token counts + USD cost.
✓	**Subscriptions (core/subscription/):** Free/starter/pro/max tier definitions and daily limits. Completely blind to personas.
✓	**Models Config (models.ts):** LLM pricing, context windows, routing rules.

STATUS: VITAL_KEEP

AUDIT: VERIFIED_GREEN

REF: PROMPT_ASSEMBLY_FUNNEL

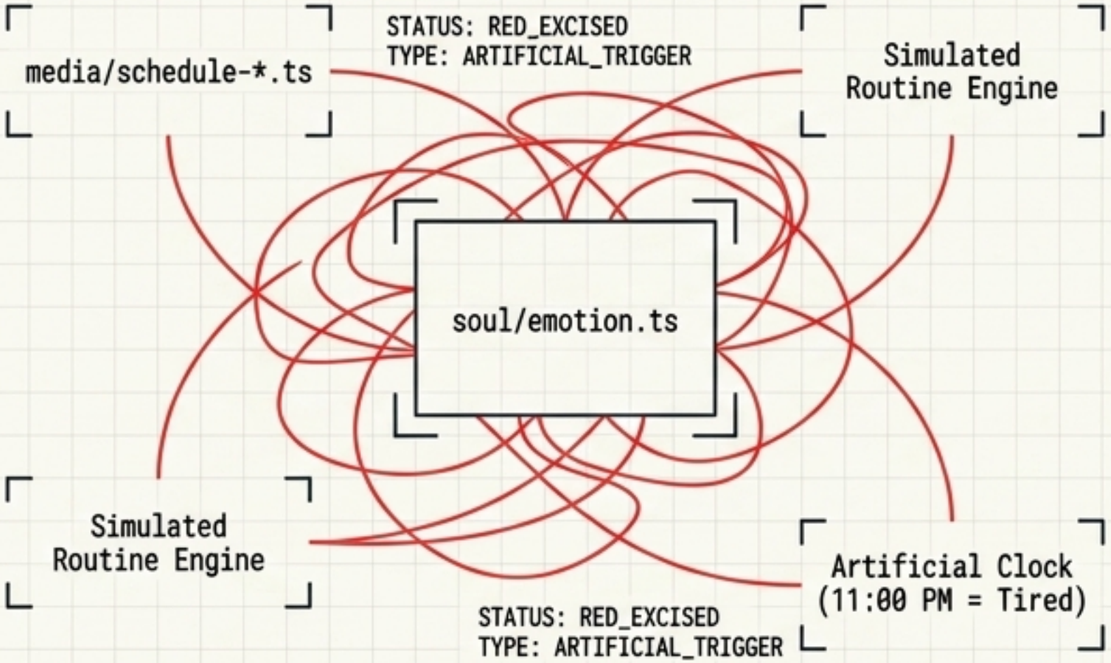
TYPE: FUNCTIONAL_SCHEMATIC



The function stays identical; only the injected parameters change.

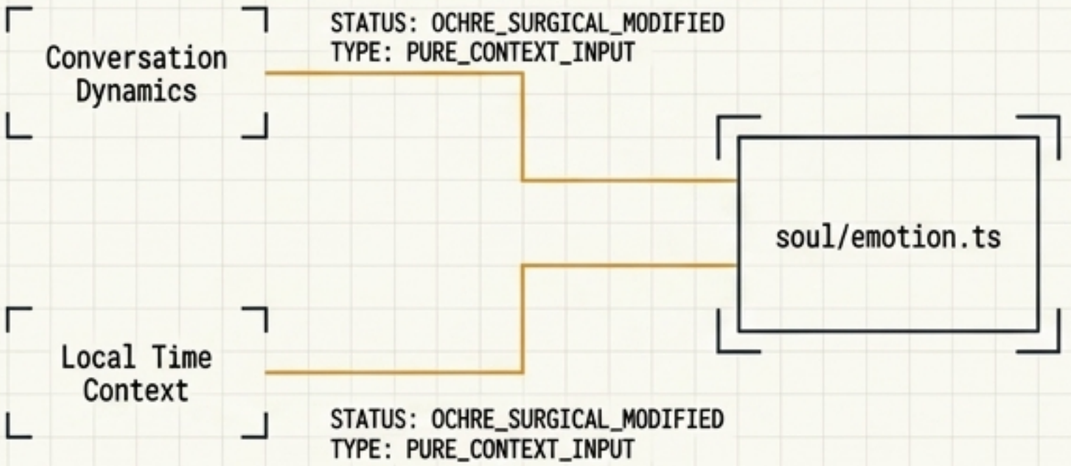
Reconstructive Surgery: The Emotion Engine

v1: Entangled & Fake



Mio feels tired because the hardcoded clock dictates an artificial daily routine.

v2: Pure & Contextual



The core emotion model (valence, arousal, discrete emotions) stays intact. The fake-life triggers are amputated.

Reconstructive Surgery: Proactive Messaging

```
{ trigger: 'just got off work',
  response: 'heading to yoga' }
```

~~{ trigger: 'just got off work', response: 'heading to yoga'}~~ ← proactive.json [REMOVED]

Time-Aware

"It's late, how was your day?"

Knows the clock, doesn't pretend to have a schedule.

Memory-Driven

"You mentioned that interview — how did it go?"

Pulls directly from extracted memories.

Emotion-Continuing

"You seemed down yesterday, feeling better?"

Reads the last recorded valence state.

Simple Care

"Haven't talked in a while, thinking of you."

Pure gap detection.

Excisions: Shedding the Persona Baggage

STATUS: RED_EXCISED TYPE: FILE_PATH

~~[DELETED] /personas/*/*.json~~

Who 'Xiaomeng the barista' is. Personality now emerges from conversation, not files.

STATUS: RED_EXCISED TYPE: FILE_PATH

~~[DELETED] /media/selfies/*~~

A companion that doesn't pretend to be human doesn't need a face.

STATUS: RED_EXCISED TYPE: FILE_PATH

~~[DELETED] /media/schedule-*.ts~~

Simulated daily routines. The hardest part to build, least valuable for retention.

STATUS: RED_EXCISED TYPE: FILE_PATH

~~[DELETED] /relationship/evolution-*.ts~~

Over-engineered state machines. Warmth + memory deepens relationships naturally.

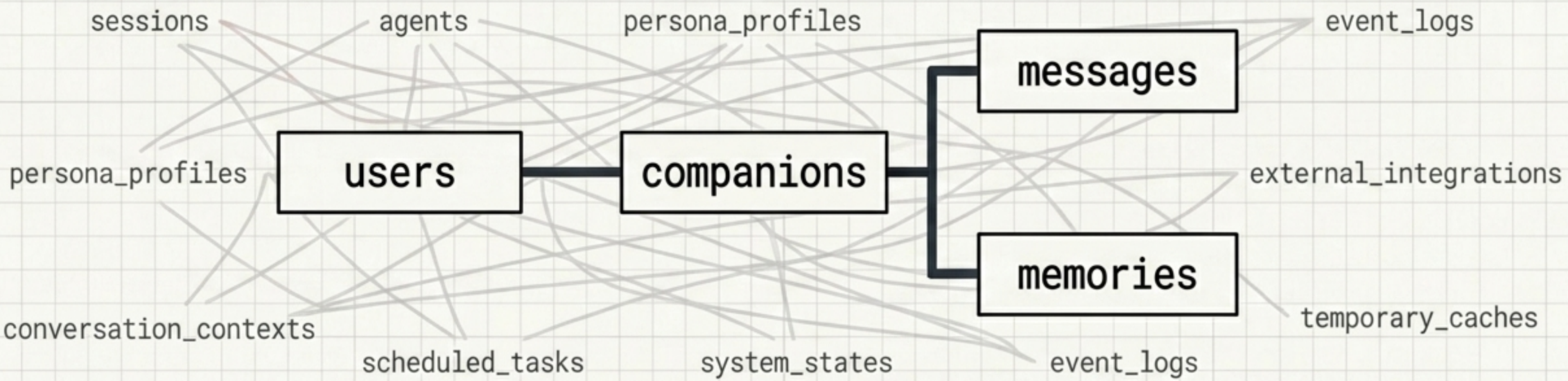
STATUS: RED_EXCISED TYPE: FILE_PATH

~~[DELETED] /media/persona-style.ts~~

Hardcoded relationship dynamics (friend, romantic).

Architectural Collapse: 10 Tables to 4

The entire multi-agent infrastructure collapses into a single constraint: `UNIQUE(companions.user_id)`. One user, one companion.



(v1) Message → Session → Agent → User (3 joins to ask “What did the user say?”)

EXCISED

(v2) Message → (1 join. Done.)

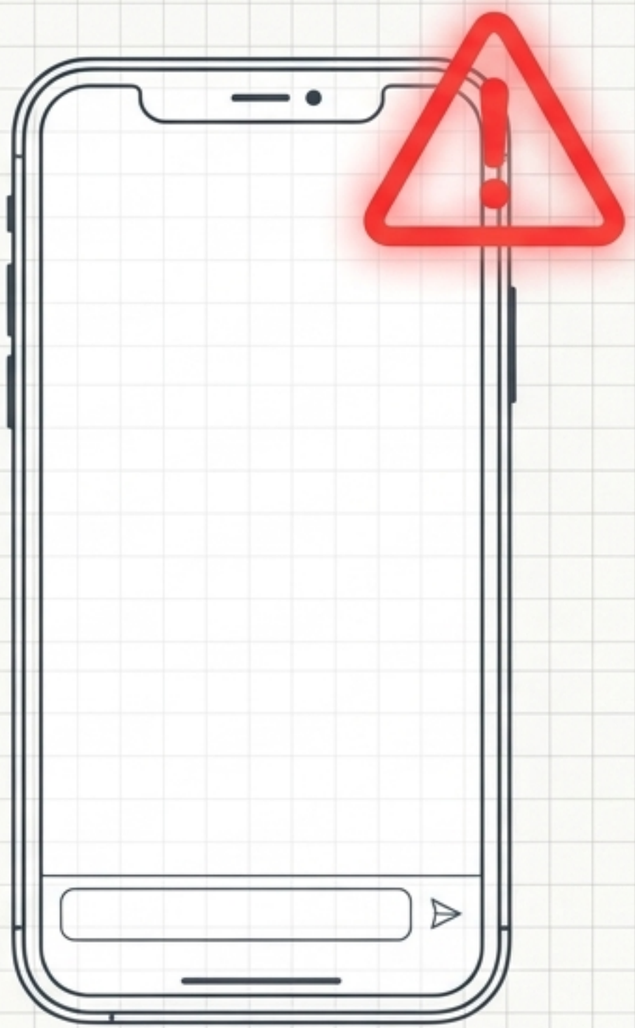
Eliminating the Middlemen

~~Eliminated Table~~	v1 Reason	v2 Architecture ✅
~~agents~~	Multi-personas	One companion per user (companions table). ✅
~~sessions~~	Multi-agent/channel routing	Flat stream per user. ✅
~~channel_bindings~~	Telegram + Web per agent	Single native app. ✅
~~onboarding_states~~	Complex state machine	Conversational (state lives in chat). ✅
~~telegram_allowlist~~	Bot access control	No Telegram. ✅
~~account_link_tokens~~	Cross-platform linking	Single auth system. ✅

Protocol Shift: The Real-Time Layer

	SSE (Server-Sent Events)	**WebSocket**
Bidirectionality	Server push only. Bad for future real-time voice. ✗	Fully bidirectional. Native support for voice chat. ✓
Proactive Messaging	Requires client polling or separate channel. ✗	Server pushes cleanly on the exact same channel. ✓
React Native / Expo Support	Requires polyfills, edge-cases on reconnection. ✗	Mature, robust libraries like reconnecting-websocket. ✓
Connection Health	Fragile app-level keepalive on mobile. ✗	Native ping/pong frames built-in. ✓

Platform Focus: The Telegram Kill



The Fatal Flaw

Chat history doesn't sync to Telegram's UI. A user can talk for weeks in the native app, open Telegram, and see an empty chat log. The companion knows everything, but the UI is a blank slate.

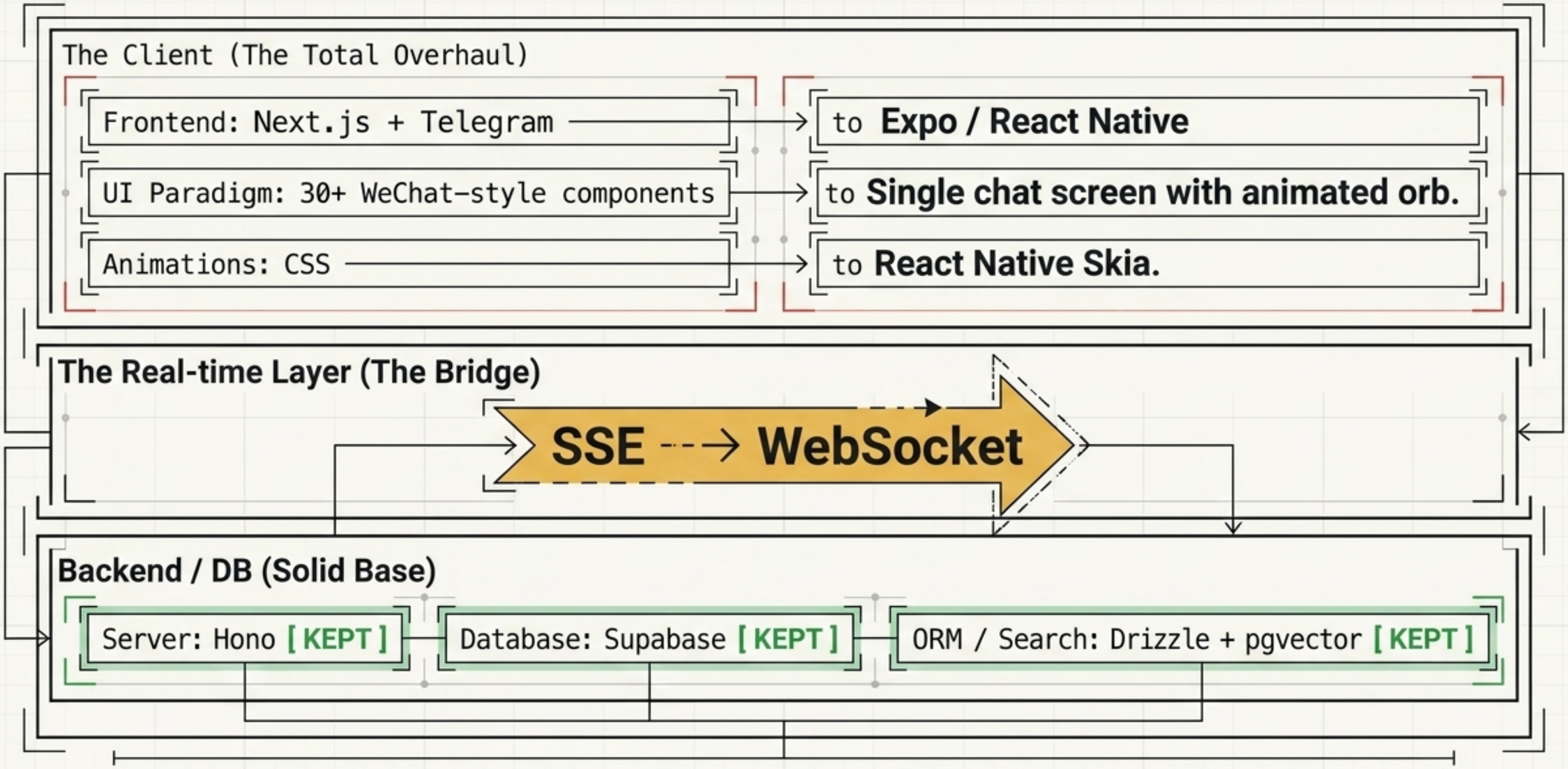
The Onboarding Problem

v2 features conversational onboarding—the companion is born through dialogue. This requires the native app. If users download the app to onboard, they have no reason to return to Telegram.

Conclusion

No Telegram in v0-v1. Focus solely on the native application.

Tech Stack Evolution



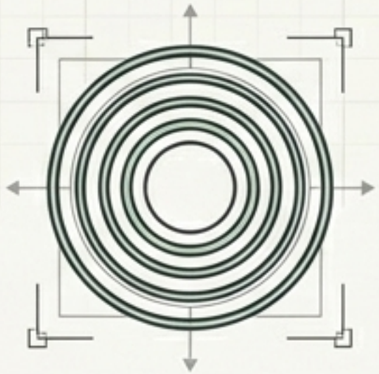
The Starting Line for v2

First commit isn't from zero.

We inherit a working **semantic memory** system, full **media pipeline**, **billing**, **cost tracking**, and **dynamic prompt builders**.

Clinical metadata

Clinical metadata



1. Implement **4-table database schema**.
2. Wire new **WebSocket server**.
3. Stand up **Expo client + animated Skia orb**.
4. Implement **conversational onboarding**.
5. Hook up modified **emotion/proactive engines**.

The hardest part isn't the code. It's resisting the urge to rebuild what already works.