

TS6059



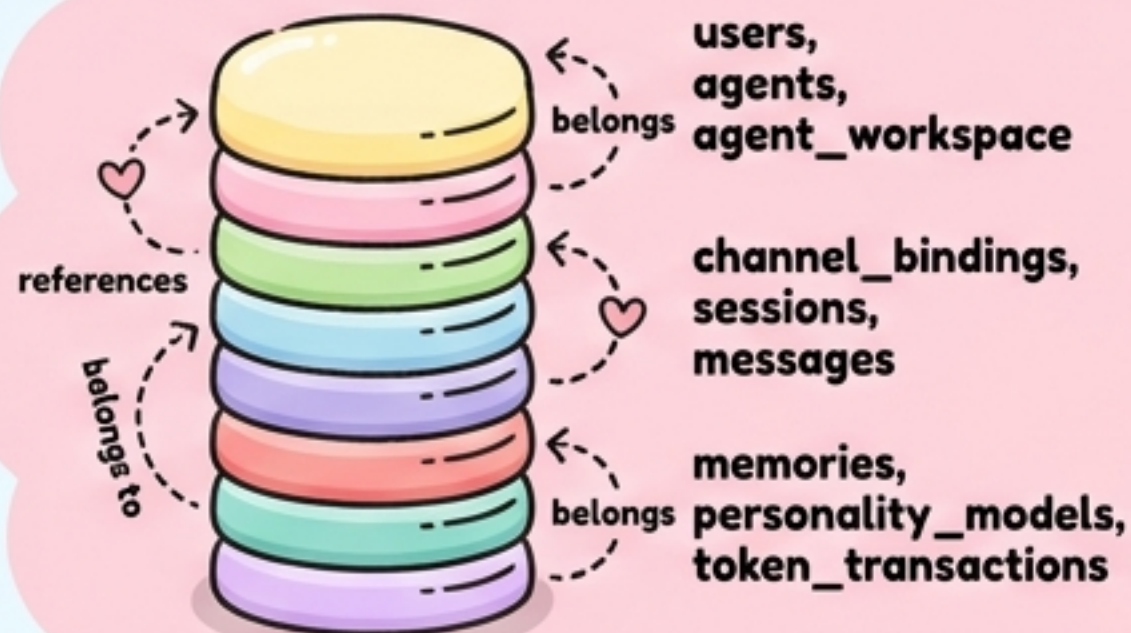
Day 1: Laying the Skeleton (and fighting TypeScript)

The Tech Stack



The Monorepo Tax

- First code = project structure. 📁
- First hurdle = rootDir conflict. 🌀?

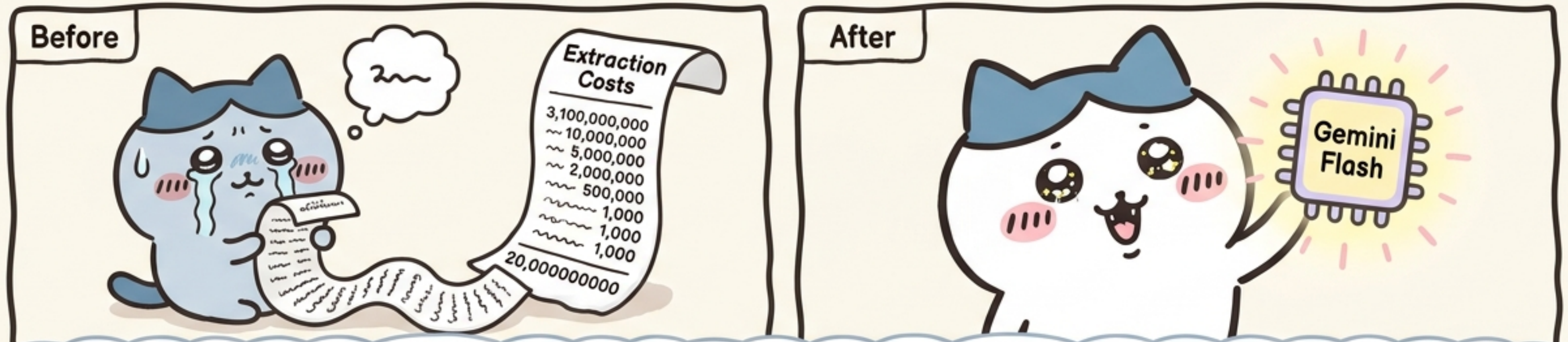


The 9-Table Core

Designer Note:

No generic metadata tables. A strict schema saves you later.

Memory is Hard. Paying for it is Harder.



The Big Pivot: Anthropic Haiku extraction costs exceeded chat costs. Switched to Gemini Flash & Embeddings. Practically free.

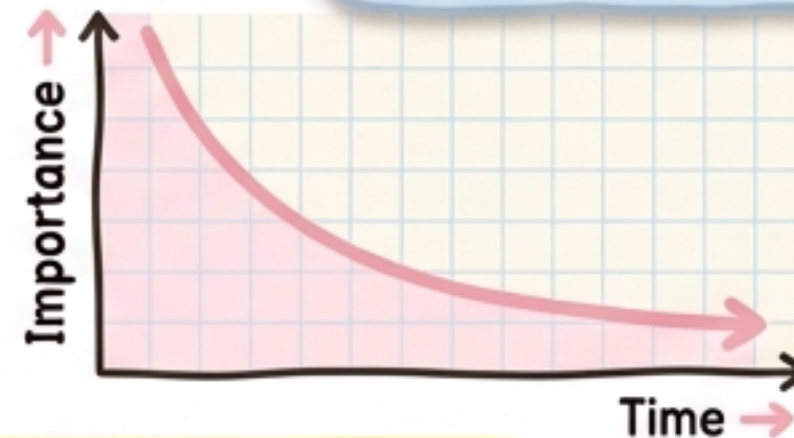
Hybrid Memory Search

70% Vector:
pgvector cosine
distance. Captures
semantics.



30% Full-Text:
Keyword matching.
Dumb, but never
misses.

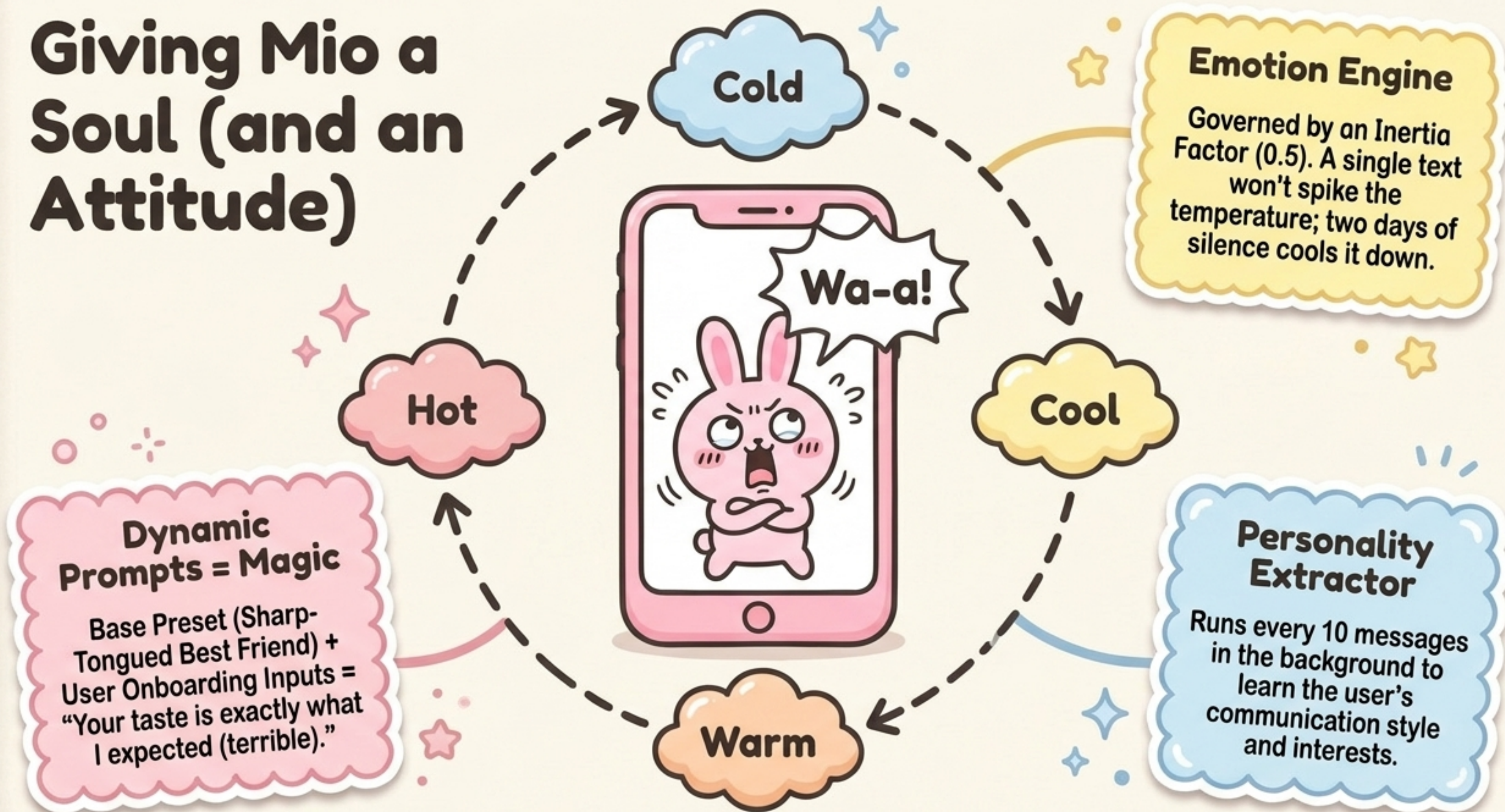
Exponential Decay



Memories fade
based on time +
importance score.

Key Takeaway: Using Vercel AI SDK `streamText()` meant changing a single config, not rewriting business logic.

Giving Mio a Soul (and an Attitude)



Slowing Down to Feel Human

The Problem: Instant AI replies to rapid-fire texts spike costs and feel robotic.



Message Debounce (5s Window)

- Wait 5 seconds after the first message.
- Catch all subsequent texts in a buffer.
- Process as one batch.



Pacing & Typing Delays

- Responses split by newlines.
- Short line: 0.5s delay + "typing..."
- Long line: 2.0s delay + "typing..."



Key Takeaway: Reading paced messages feels like chatting with a real person, not querying a database.

The Cron Job Engine

Runs every 30 minutes.

Trigger: Has the user been silent for 2+ hours?
Is their emotion state cooling?

Rules: Max 3 times a day.
Strict quiet hours (23:00 - 08:00).

When They Reach Out First

39 Commits Later

From an empty repo to a working AI companion.

Thinking of you

The Realization

Proactive messaging turns an AI from a tool into a living presence.

Mio doesn't need you to speak first. It thinks of you on its own.