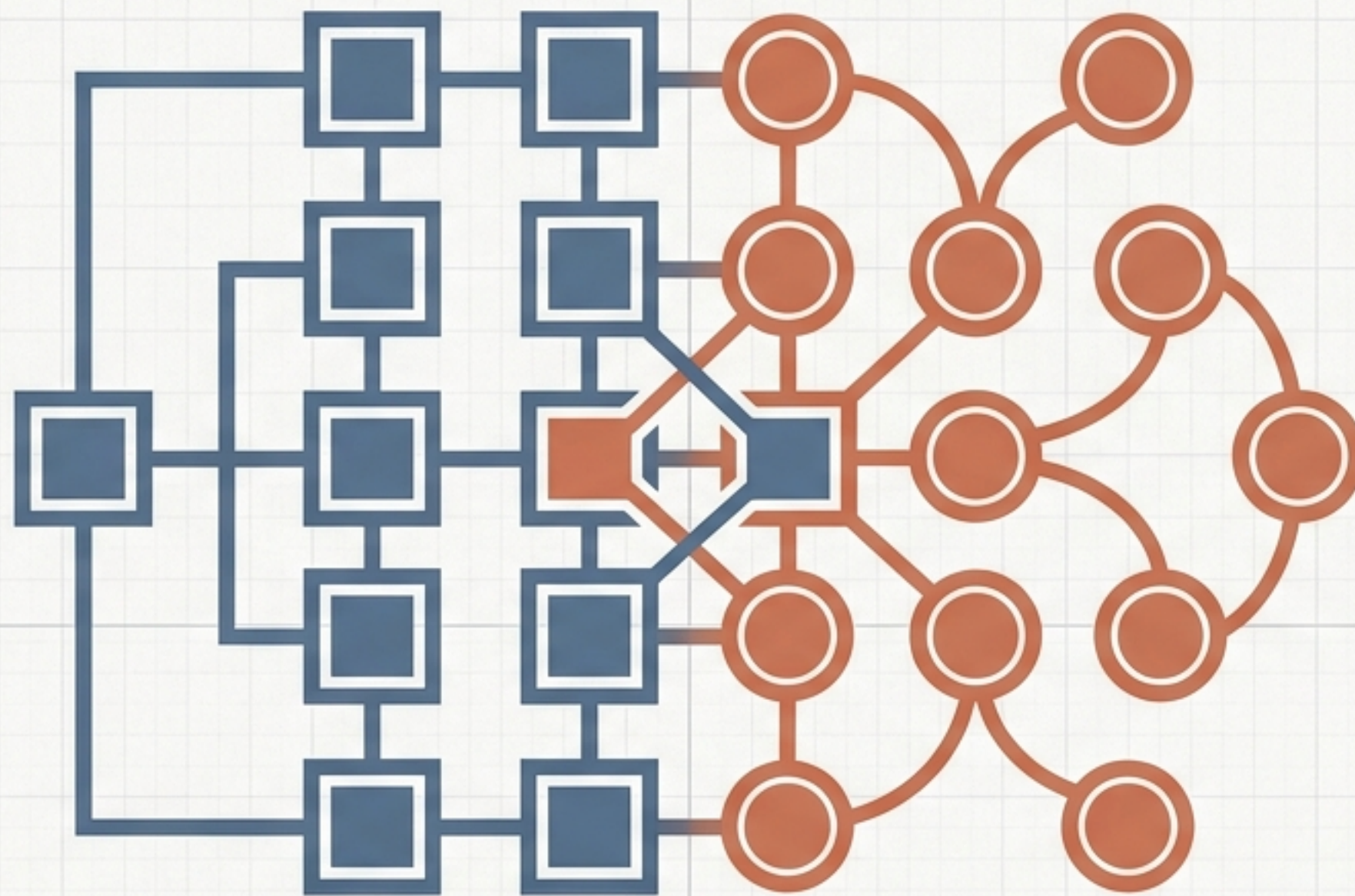


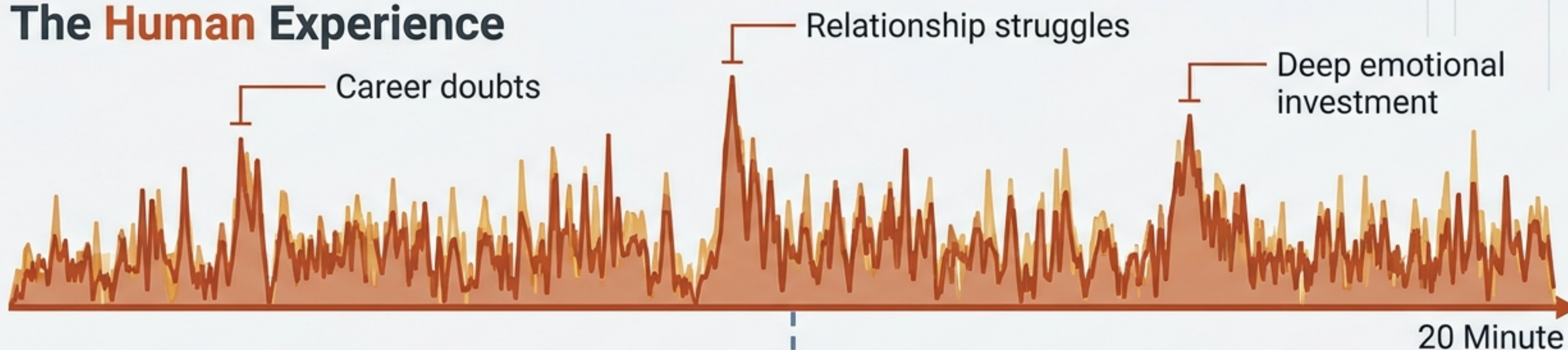


Architecture of a Mind

Building an Agentic Memory System with Claude Code (v0.1)

The Goldfish Problem

The **Human** Experience

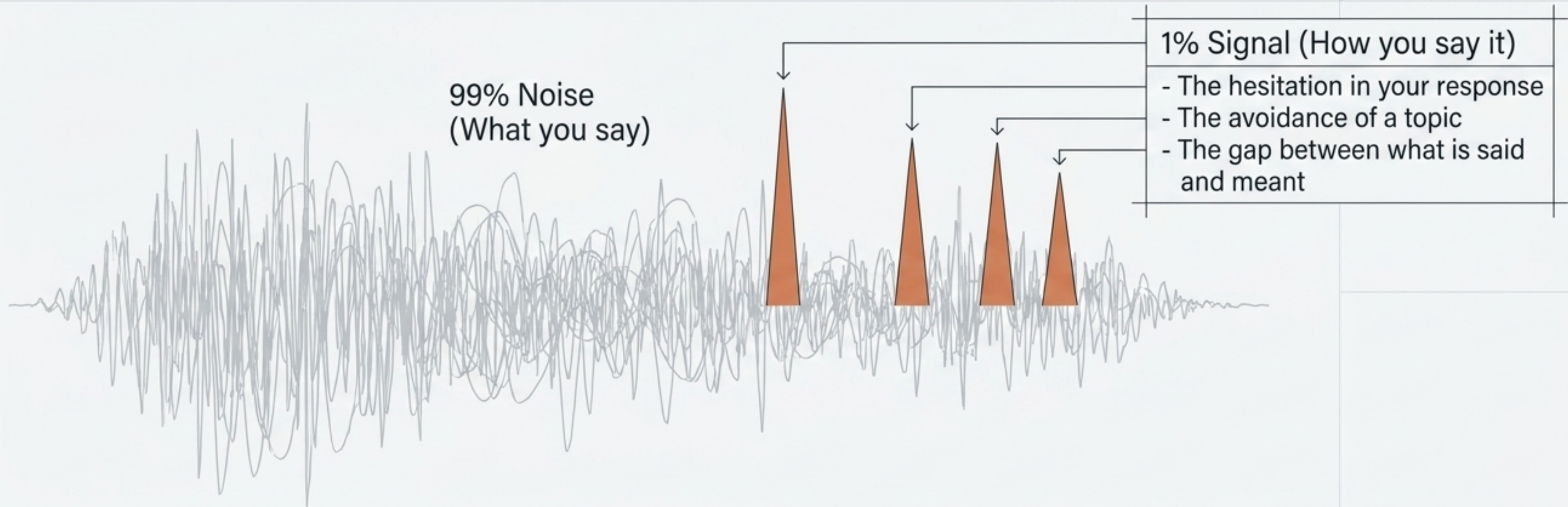


The **LLM** Reality



Takeaway: Without memory, there is no understanding.

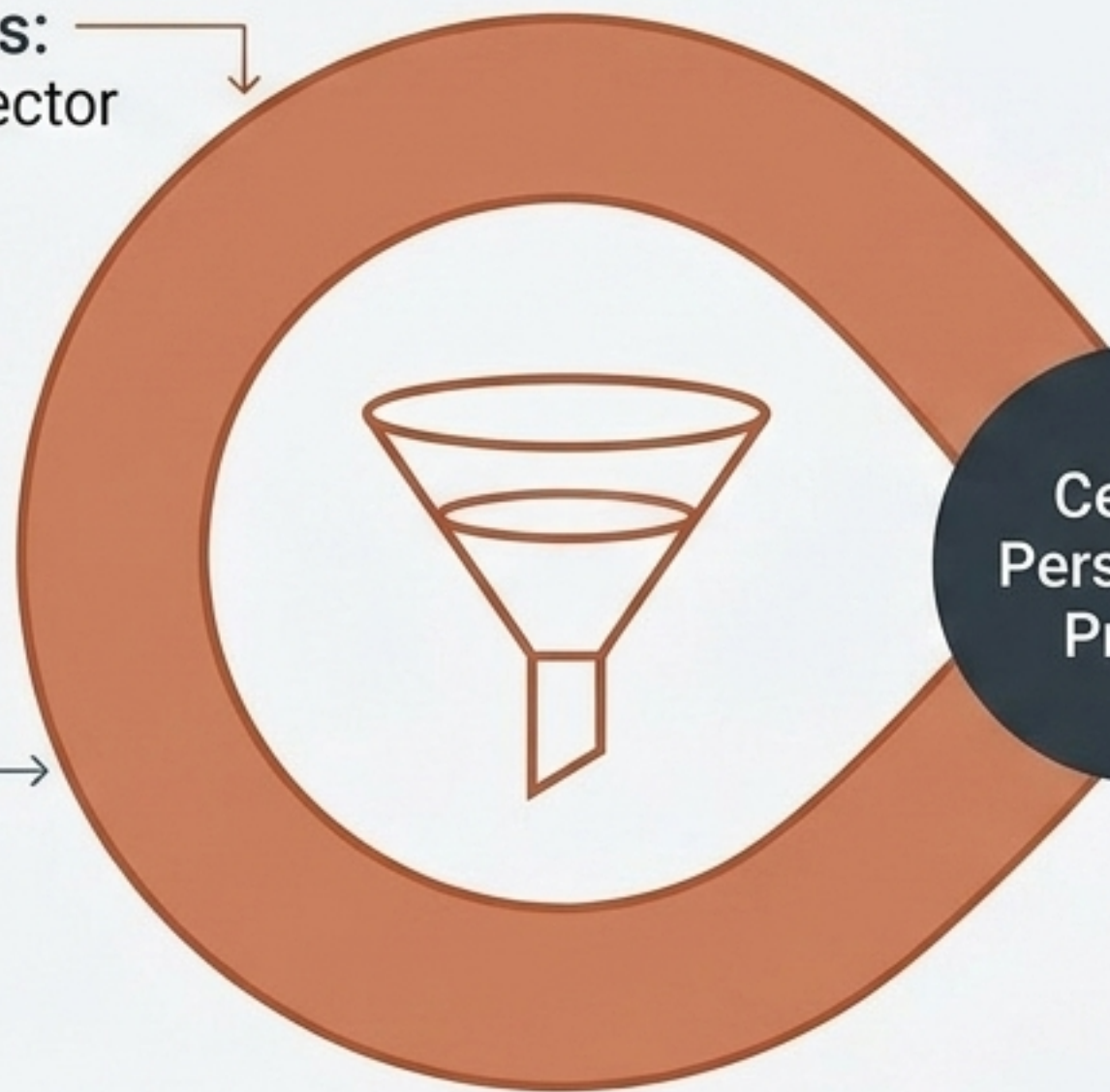
Memory is the Moat



Cross-session memory shouldn't just capture facts.
The psychological signal is the moat.

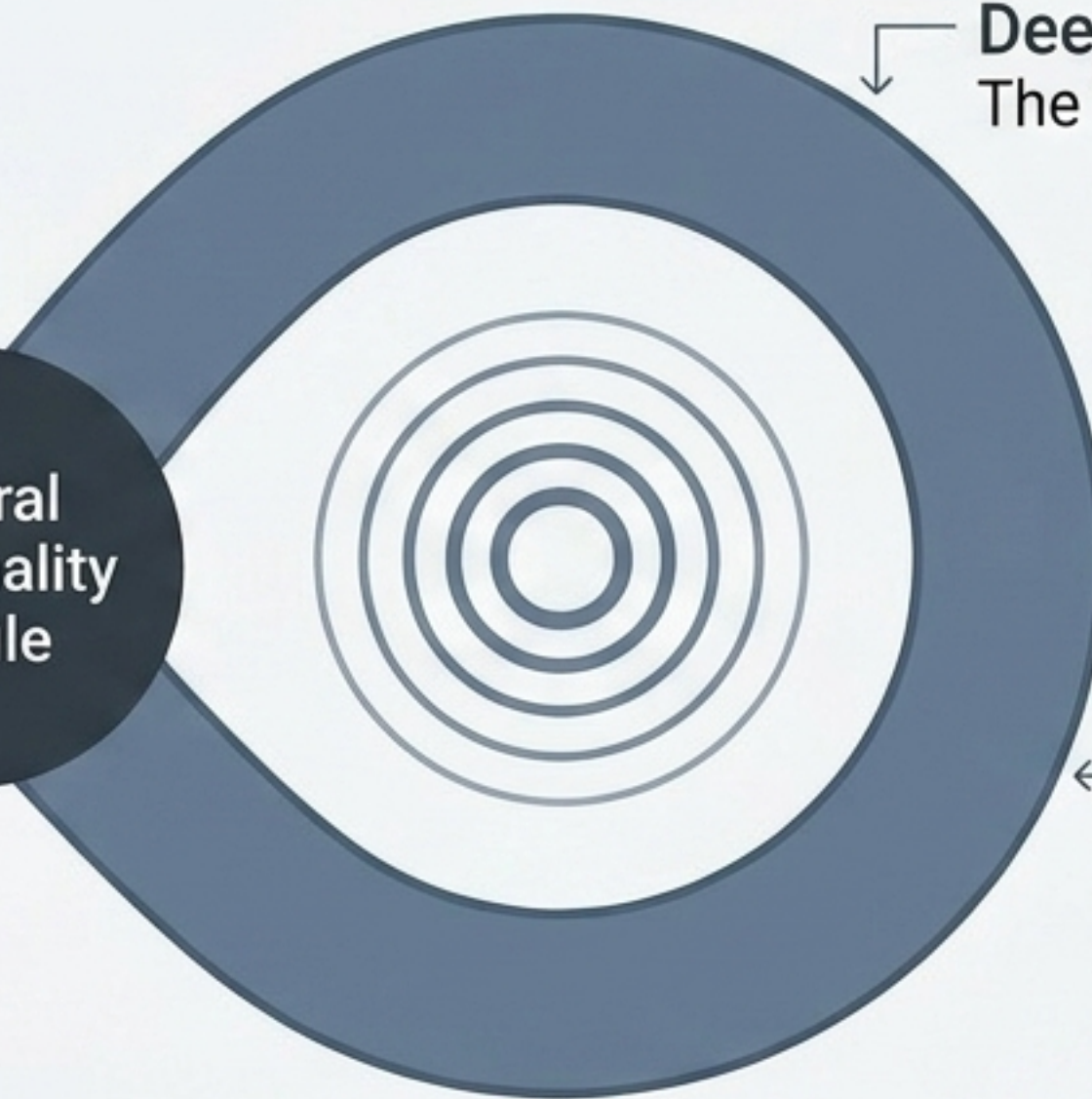
Two Features, One Flywheel

MBTI Sessions:
The Signal Collector



A conversational assessment that extracts behavioral patterns and unresolved conflicts.

Deep Companion:
The Value Delivery



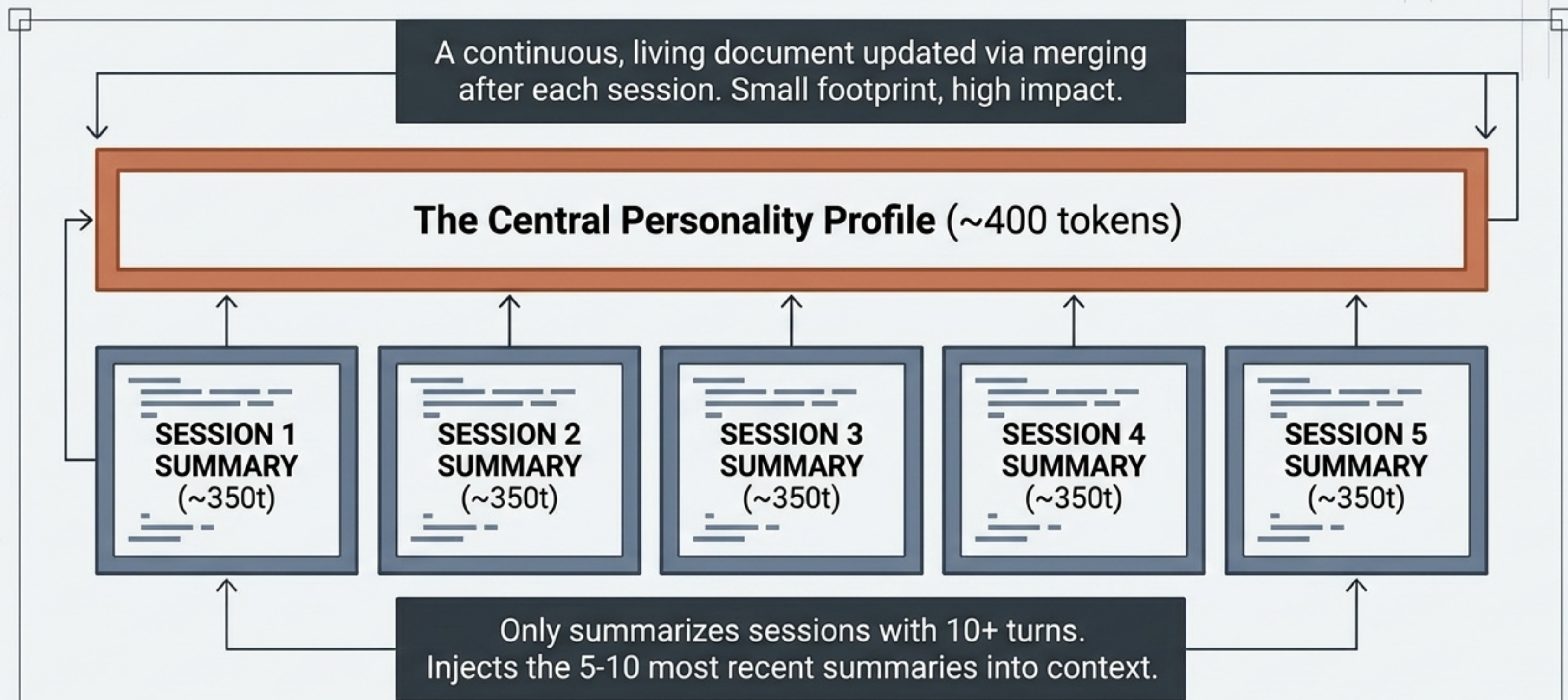
Unlocked after 5 sessions. Injects the profile into the context to naturally reference your history.

Central
Personality
Profile

The Data Trade-off: How to Store a Soul?

	Raw Conversations	Pure Summaries	The Hybrid Solution
Token Cost (5 Sessions)	15,000 - 25,000 tokens (Prohibitively expensive)	Very low	~350 tokens per session
Signal Fidelity	Perfect texture.	Dry, strips away hesitations and emotional patterns.	Structured extraction + Verbatim quote preservation.
System Noise	High noise, hard to query.	Low noise.	Low noise, high relevance.

The Two-Layer Memory Pipeline



Engineering the Illusion of Care

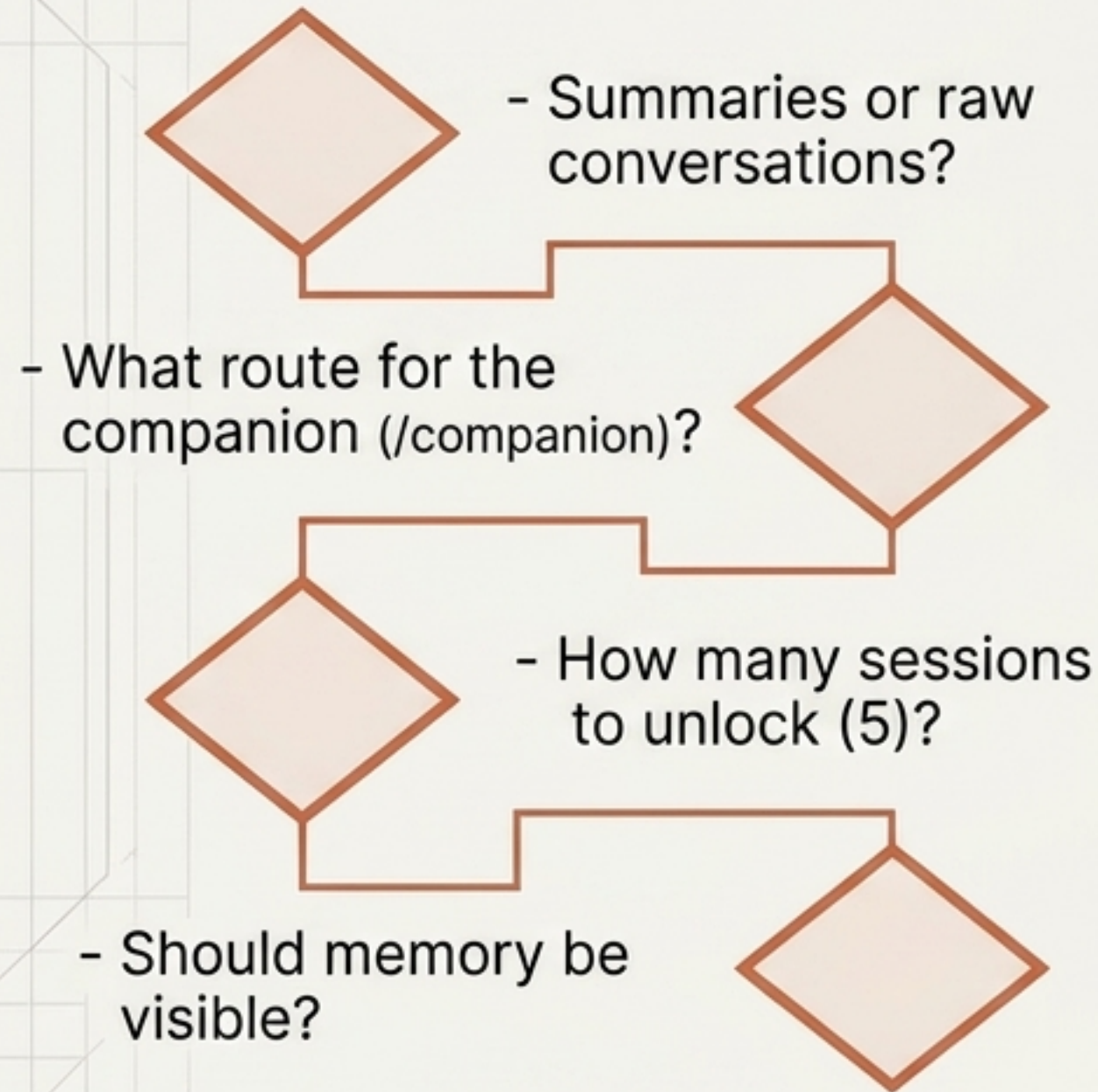
```
{  
  "profile": {  
    "memorableDetails": [  
      "The design draft that got rejected three times."  
    ],  
    "unresolvedThreads": [  
      "Fear of taking the big-company offer."  
    ]  
  }  
}
```

These fields are not for analyzing the user. They exist solely to create a human connection.

Core Rule: Noticing a contradiction doesn't mean pointing it out. Bring it up gently only when the user actively explores it. (Care vs. Pandering).

The Manager vs. The Maker

The Human Manager



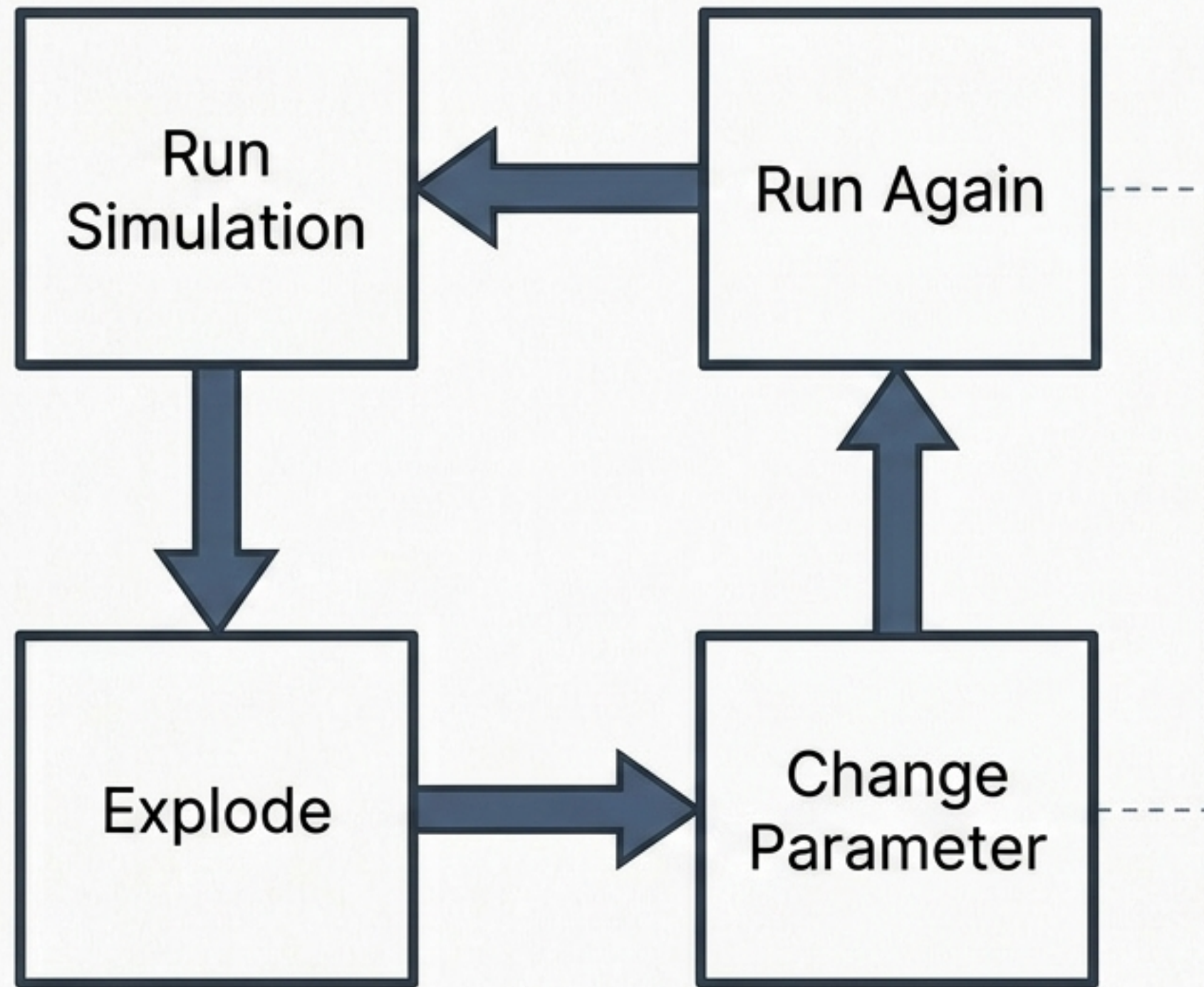
The AI Maker - Claude Code



- Database migrations
- API routes & memory pipelines
- Simulation scaffolding
- Integration testing

Takeaway: Zero lines of implementation code written manually. Zero tests run manually.

Algorithm Simulation: The 10-Minute Debug Loop



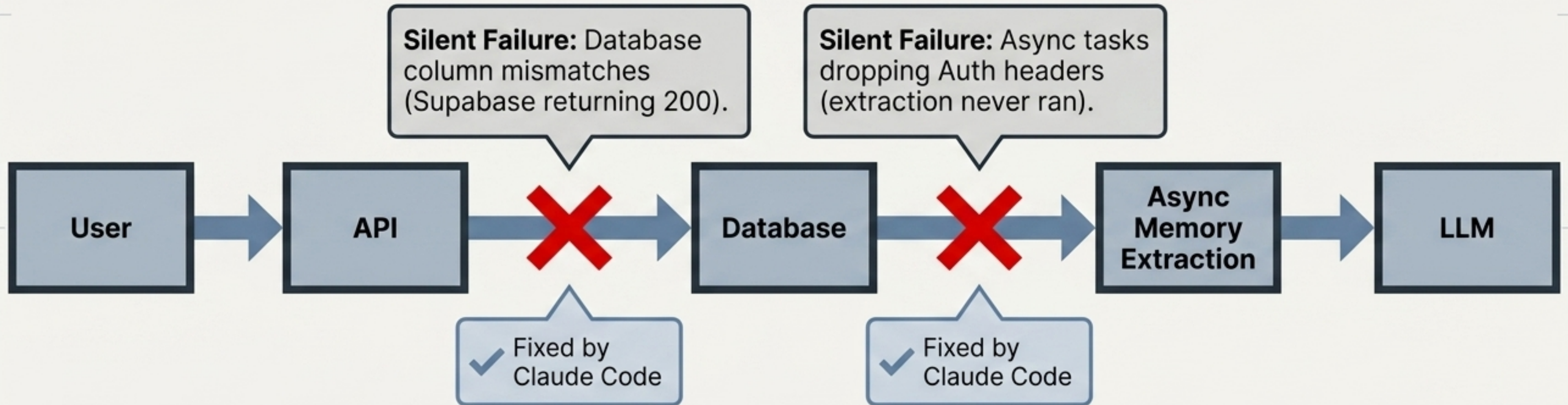
Run 1: Judge JSON truncated. Bump maxTokens 600 -> 1200.

Run 2: Generator format mismatch. Tighten constraints, bump maxTokens to 5000.

Run 3: Profile merge truncated. Schema too large. Bump maxTokens 800 -> 1400.

Run 4 (Success): Meaningful feedback received: "Hesitation capture is accurate, but add avoidance of behavioral shift."

Integration Simulation: Catching the Silent Failures



The Test: Create real test users, hit APIs, poll database for async extraction, ask: "Do you remember where I work?"

Takeaway: The pipeline went from zero to testable in under a day, all in the same context window.

The Simulation: Meet Lin Xiaoyue

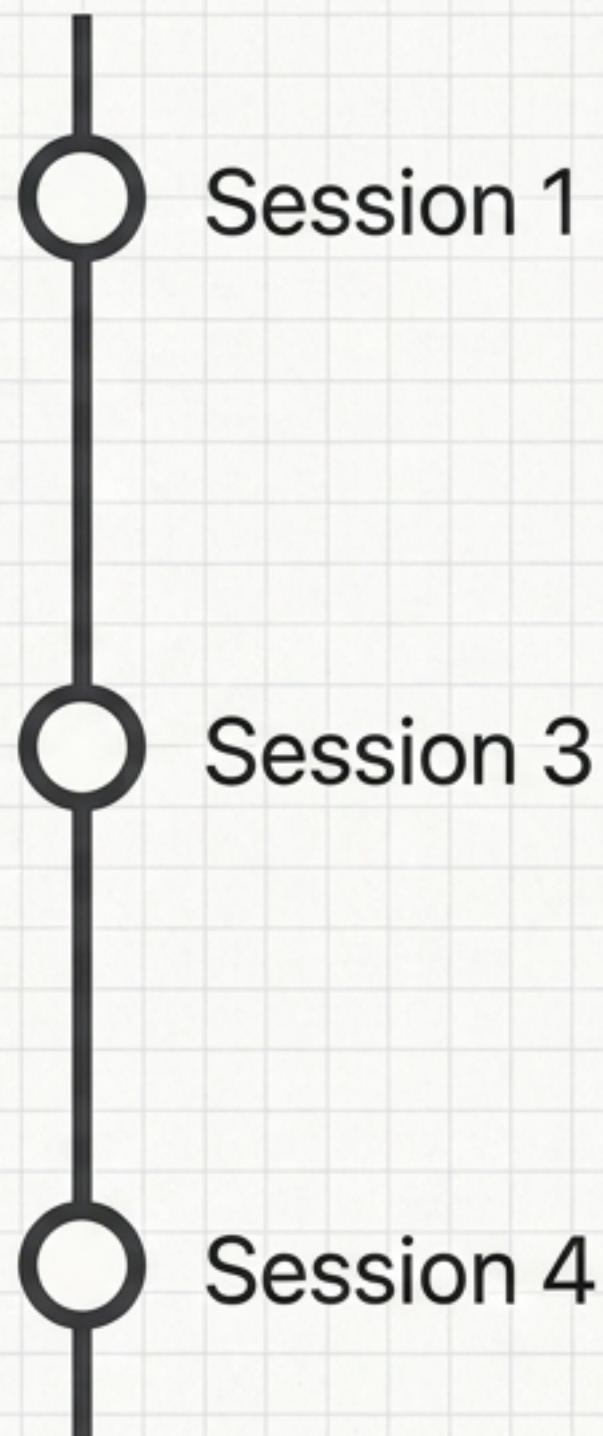
Subject: Lin Xiaoyue

Age: 28

Profession: UI Designer

The Setup: 3 MBTI sessions + 2
Deep Companion sessions.

The Goal: Test memory extraction,
accumulation, and cross-session
psychological referencing.

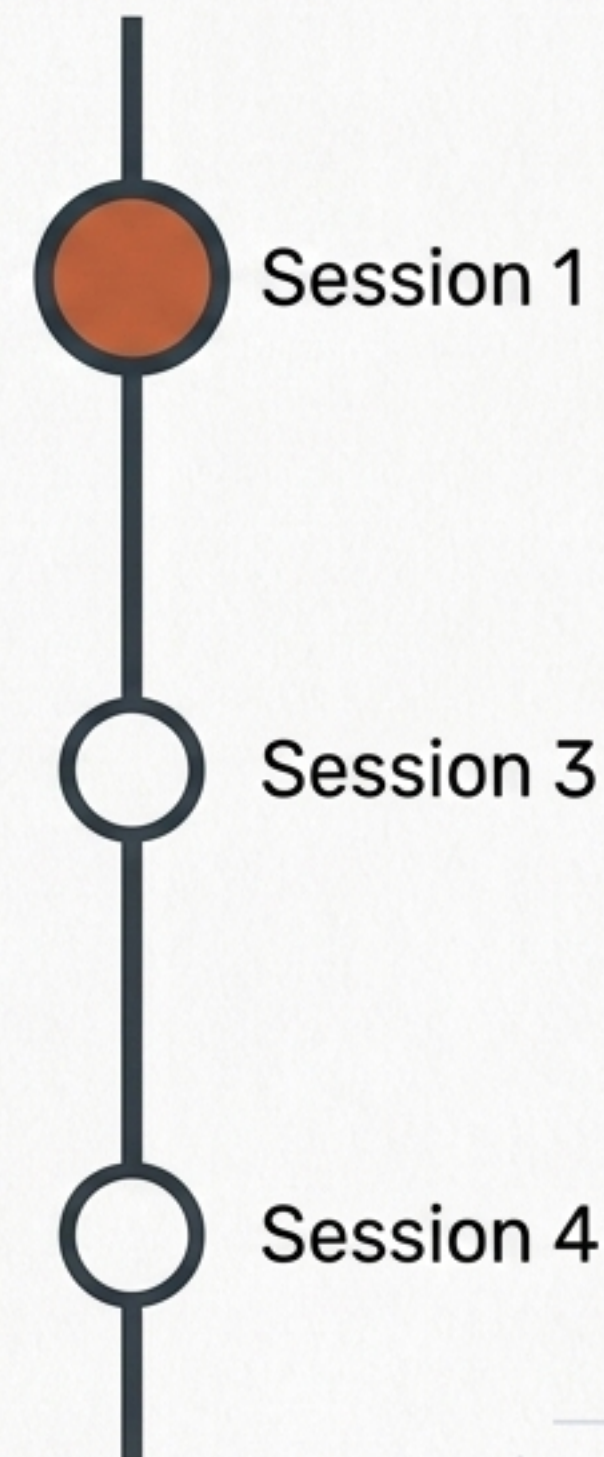


Session 1: Extracting the Anchor

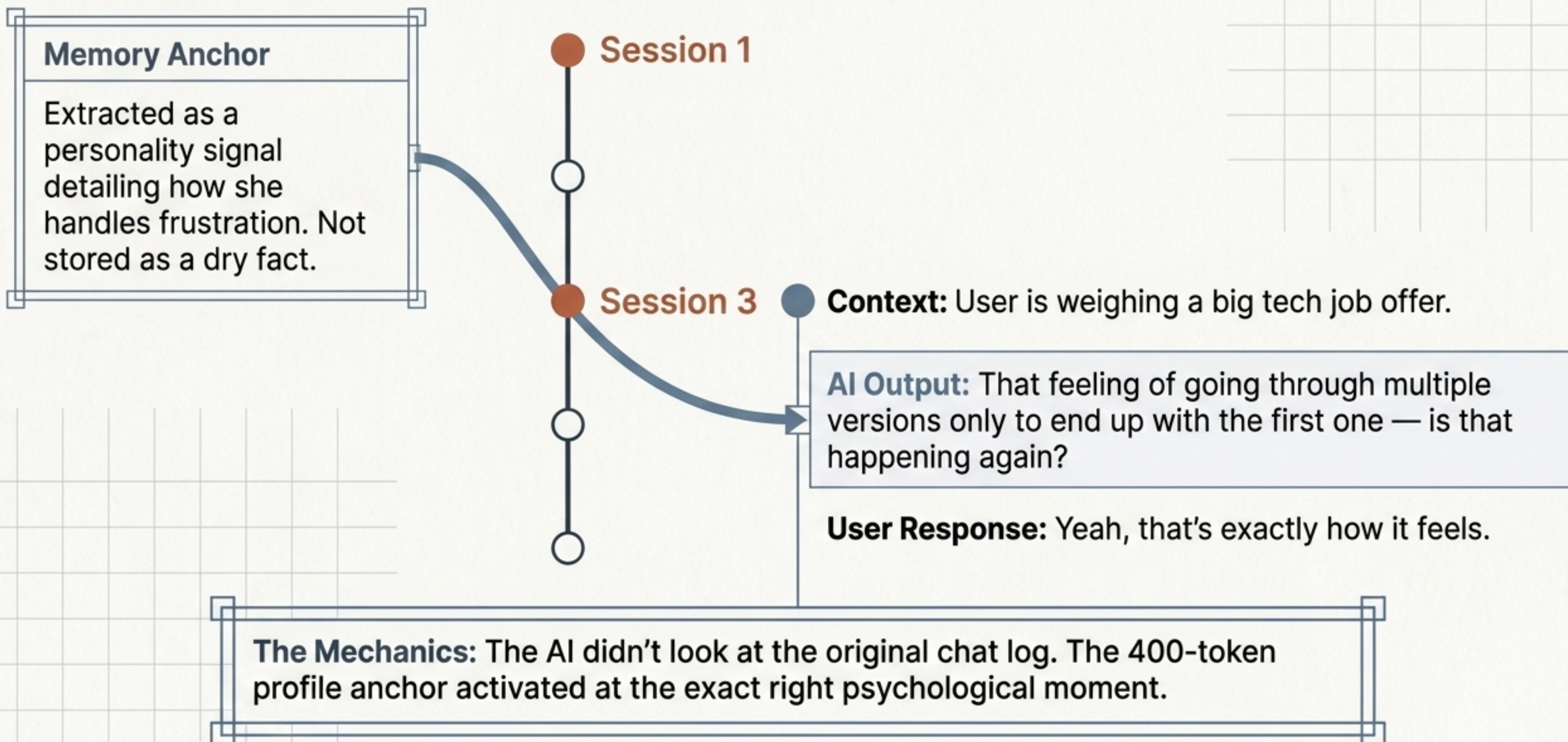
I went through multiple versions of a design draft, and they ended up going with the first version. It felt kind of... powerless.

Memory Anchor

Extracted as a personality signal
detailing how she handles frustration.
Not stored as a dry fact.



Session 3: Cross-Session Activation



Session 4: The Mirror (Part 1)

○ Session 1

○ Session 3

● Session 4

The Complaint: Insomnia. "My boyfriend Xiaoming tries to solve it instead of listening. Everything he says is right, but I feel kind of stuck inside."

Human Logic

Solving the problem immediately.
Offers advice instead of listening.

Memory-Driven Intervention

Instead of generic comfort, recall past behavioral statements.

You once said "what I want isn't advice." Is Xiaoming often like this?

Session 4: The Mirror (Part 2)

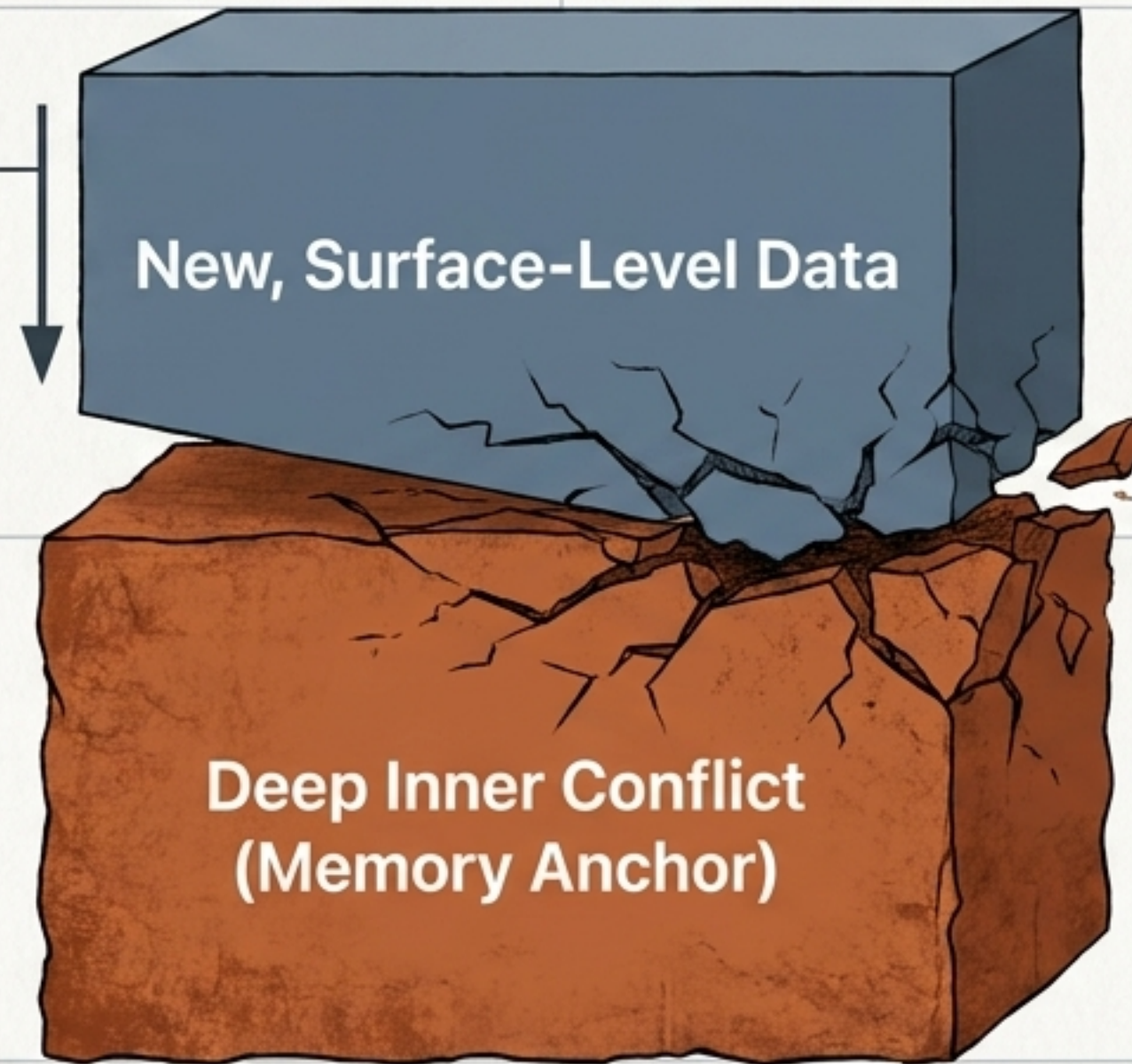
“I’ve never actually told him what I really want. It feels like telling him would be blaming him... I’m the one who shut the door first.”

The Result: The AI didn’t give advice or analyze. It used her own accumulated words to guide her to a blind spot.

Score: 10 dimensions evaluated.
Average 8.1/10. The flywheel turns.

The Pitfall: Accumulation vs. Replacement

The Bug: During profile merging, the AI was replacing old content instead of accumulating on top of it. Core anchors ("absorbing others' emotions is draining") were overwritten by surface-level companion chatter.



The Lesson: Extracting signals is easy. Orchestrating memory so the soul doesn't get overwritten is the actual hard part.

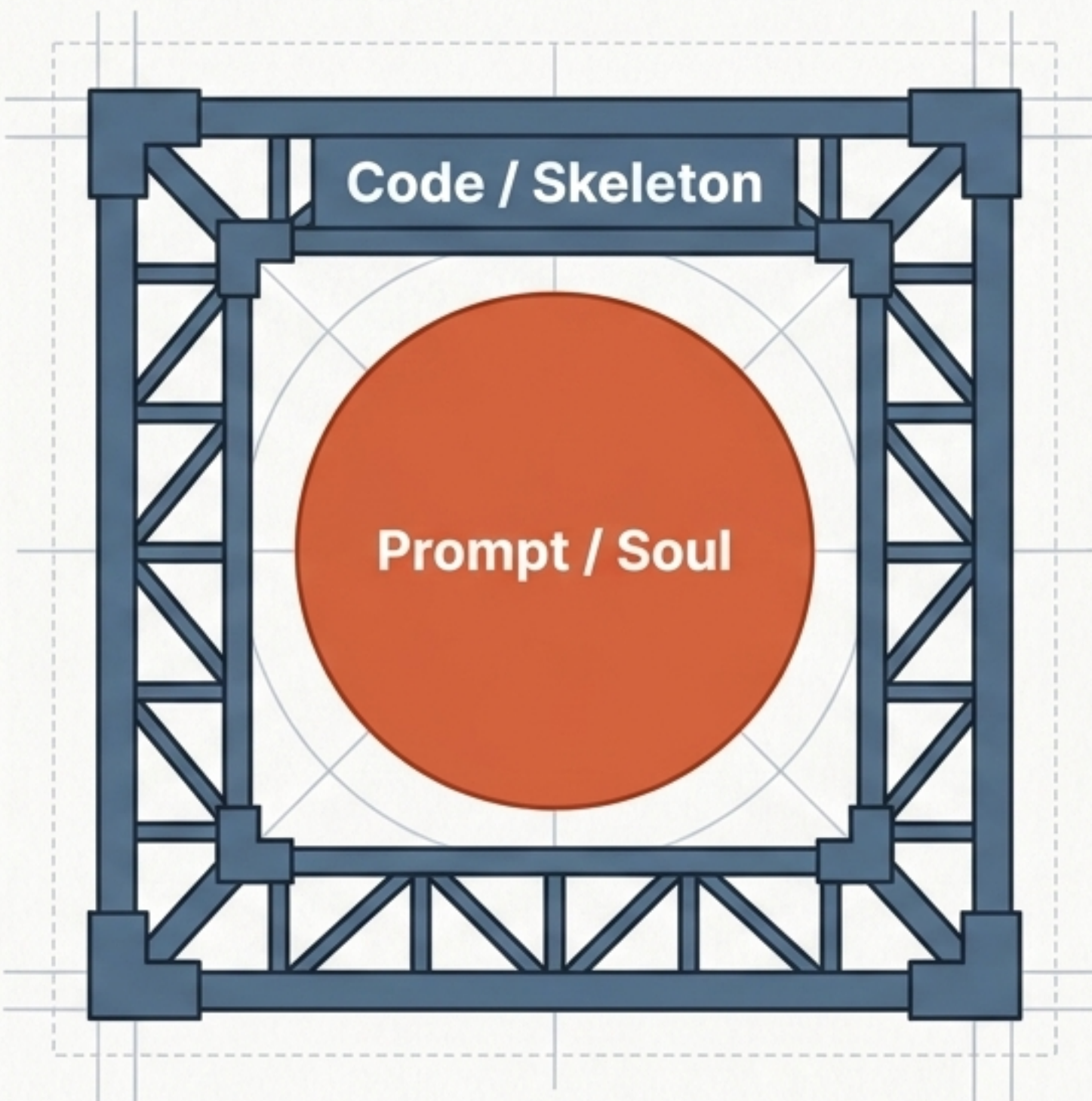
The Prompt Fix: Orchestrating the Update

Field Name	Rule	Rationale
memorableMoments	Accumulate only.	Core asset for feeling real . Never delete unless user explicitly resolves it.
preferences	Accumulate only.	Preferences are built over time, never overwritten.
unresolvedThreads	Remove cautiously.	A topic moving on doesn't equal resolution. Prioritize deep inner conflicts over surface information.

Skeleton & Soul

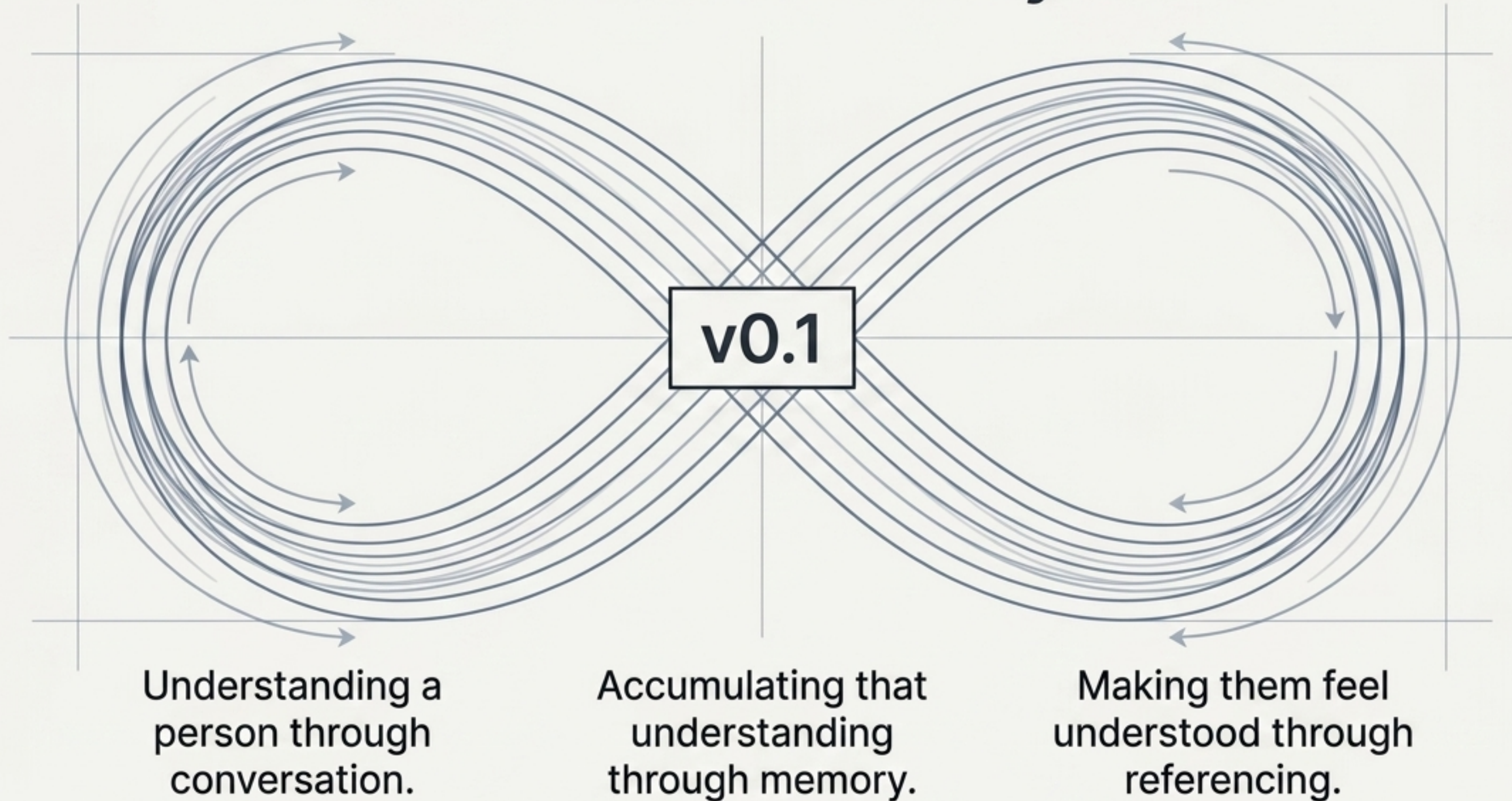
Not a single line of architecture code changed to fix the memory replacement bug.

It was pure prompt tuning.



System design and testing cannot be separated. Fields like `unresolvedThreads` only emerged after simulations revealed what was missing.

The First Turn of the Flywheel



v0 was a conversation tool without memory. **v0.1** is a companion that is beginning to understand you. The distance between those two is the **first turn of the flywheel**.