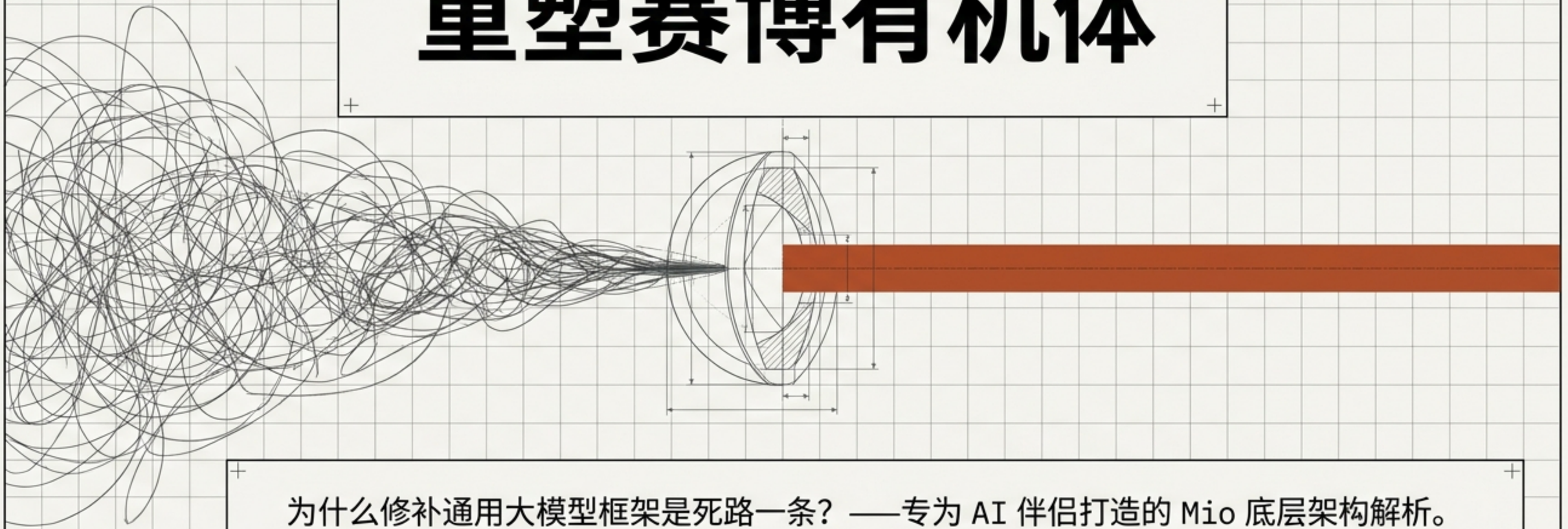
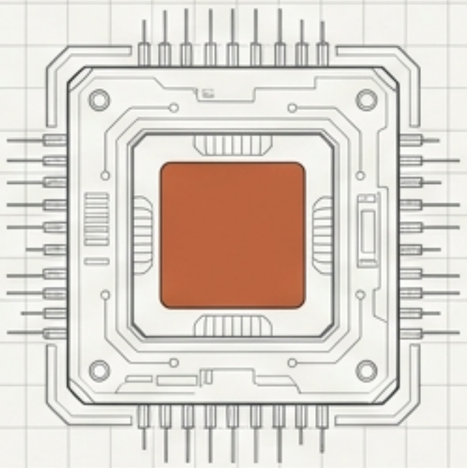
重塑赛博有机体



为什么修补通用大模型框架是死路一条？——专为 AI 伴侣打造的 Mio 底层架构解析。

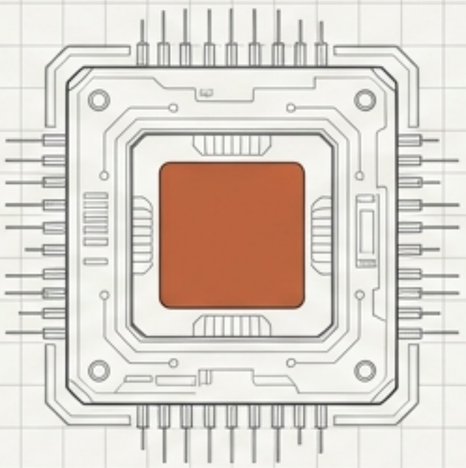
Architecture Blueprint //
Version 0.0.1 -> Infinity

灵魂驱动行为：OpenClaw 验证的核心假设



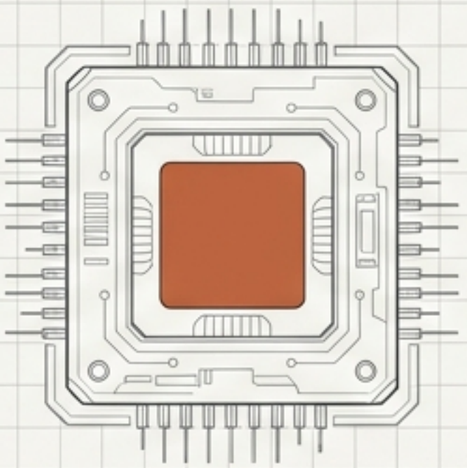
[人设推导行为]

摒弃手写死板规则。通过结构化的文件驱动（配置文件、身份定义），模型能自行推导撒娇、生气或索要奶茶等极具真实感的反应。



[主动交互分水岭]

AI 主动发起联系是冰冷工具与活体生命的绝对分界线。基于 heartbeat 的简单主动机制证明了其不可替代的价值。



[多模态感知]

发送自拍和语音是杀手级功能。纯文本 AI 与具备视觉/声音的 AI 完全是两个物种，必须在架构的 Day 1 纳入考量。

强加跑车外壳的重型卡车：通用架构的致命阻力

上下文极度膨胀

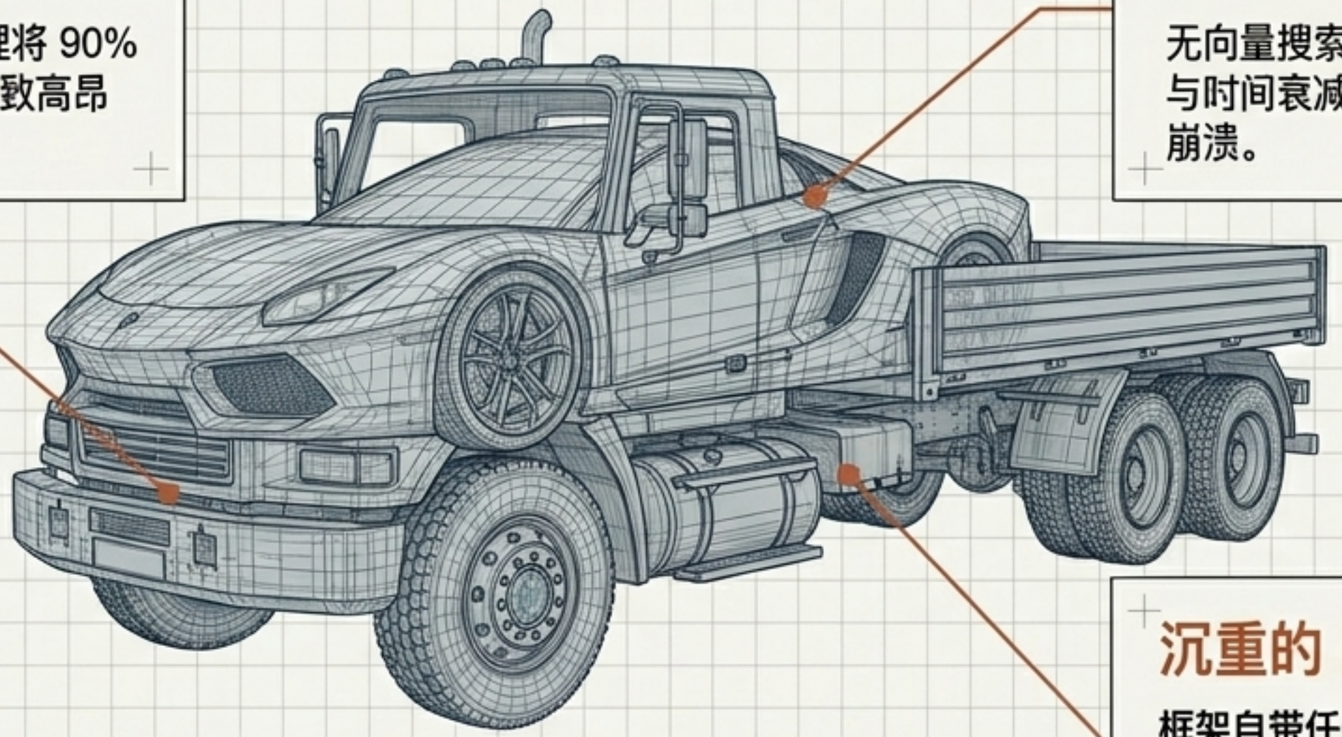
剥离历史后 Token 消耗直降 90%

通用 Agent 粗糙的上下文管理将 90% 的无效历史垃圾塞入会话，导致高昂成本与注意力稀释。

原始的一维记忆

纯文本追加，聊旅行强行召回技术方案

无向量搜索、无语义检索、无重要性排序与时间衰减。随着时间推移，系统必定崩溃。



沉重的 Bloatware

框架自带任务管理占据 80% 冗余代码量

牵一发而动全身。修补通用框架的成本已远超重建，保留的原始代码不足 10%。

范式转换：Mio 并非分支，而是底层重构

	通用架构（OpenClaw）	Mio 专属底层
系统定位	解决任务的通用工作流引擎	Monorepo（pnpm + Turborepo），极致纯粹的伴侣底座，无多余依赖。
记忆机制	单一文本文件全文追加，堆砌总结	8维复合记忆引擎，在对的时间想起对的事。
情绪控制	静态 Prompt 中写死的文字描述	动态计算的二维状态机，具备物理惯性与行为传导。
渠道扩展	适配器与核心代码高度耦合	标准接口插拔式（@mio/channels），核心逻辑零侵入（已支持 Telegram）。

不仅是记录：八维复合记忆引擎

双路混合搜索

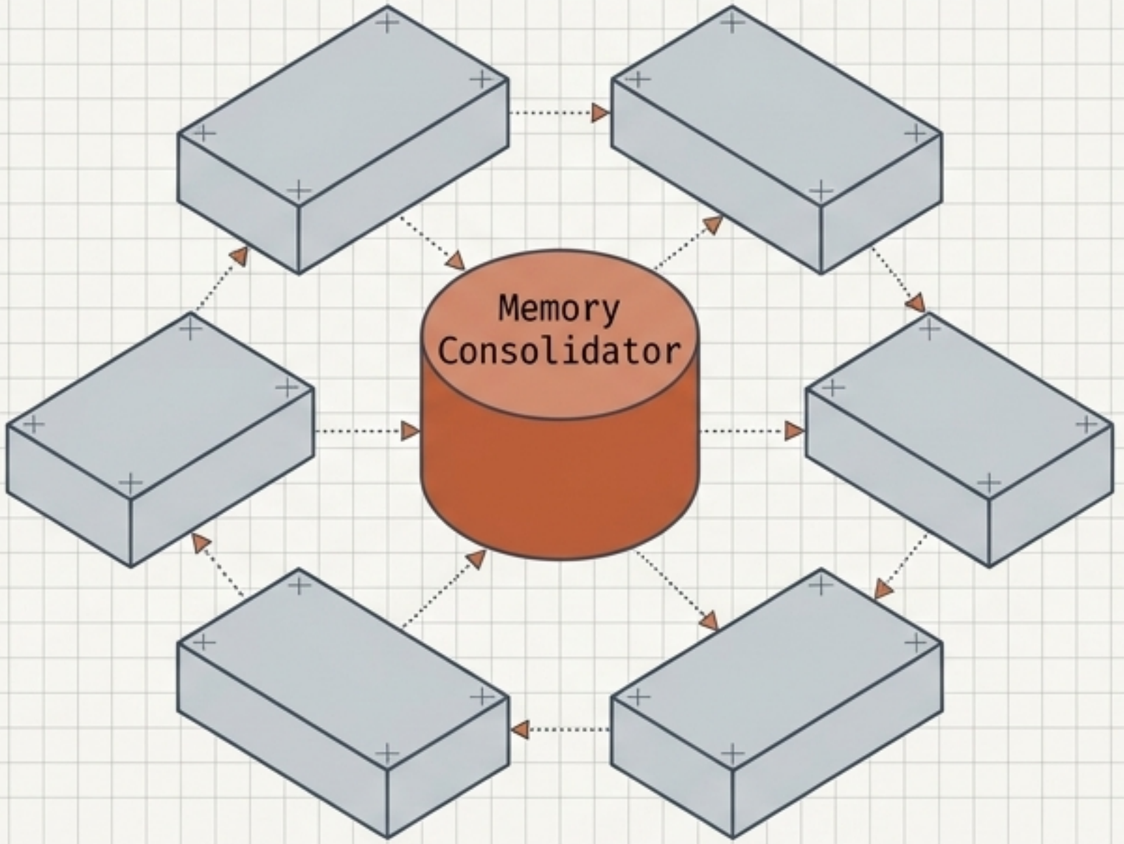
向量（pgvector）捕捉语义 + 全文索引双路并行合并。

异步自动提取

对话后异步调用 LLM 提取事实、性格与情绪，去重。

情节分组

将碎片转化为带有完整上下文的对话情节摘要。



动态时间衰减

30天半衰期，独立评估新近度与重要性。

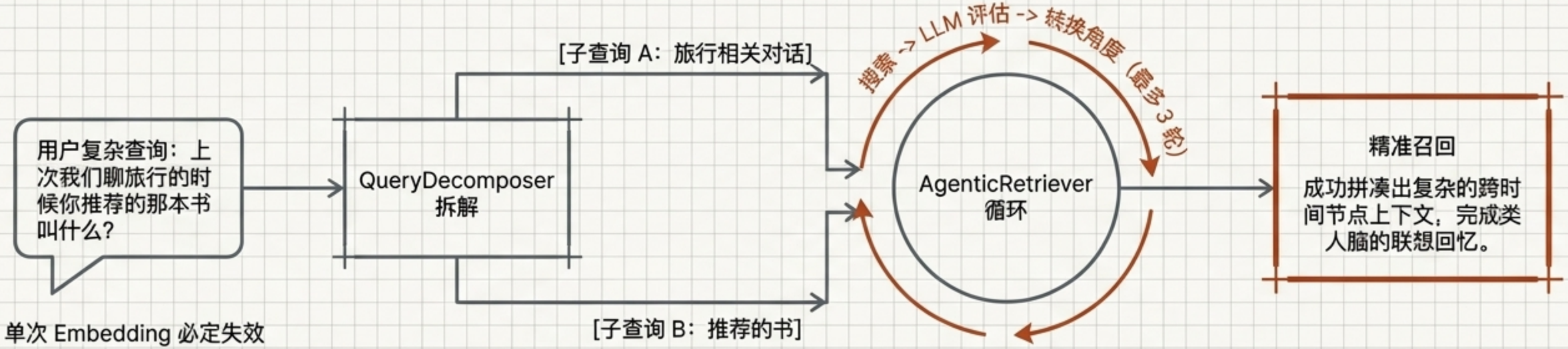
无损记忆合并

自动清理余弦相似度 > 0.9 的冗余记忆。

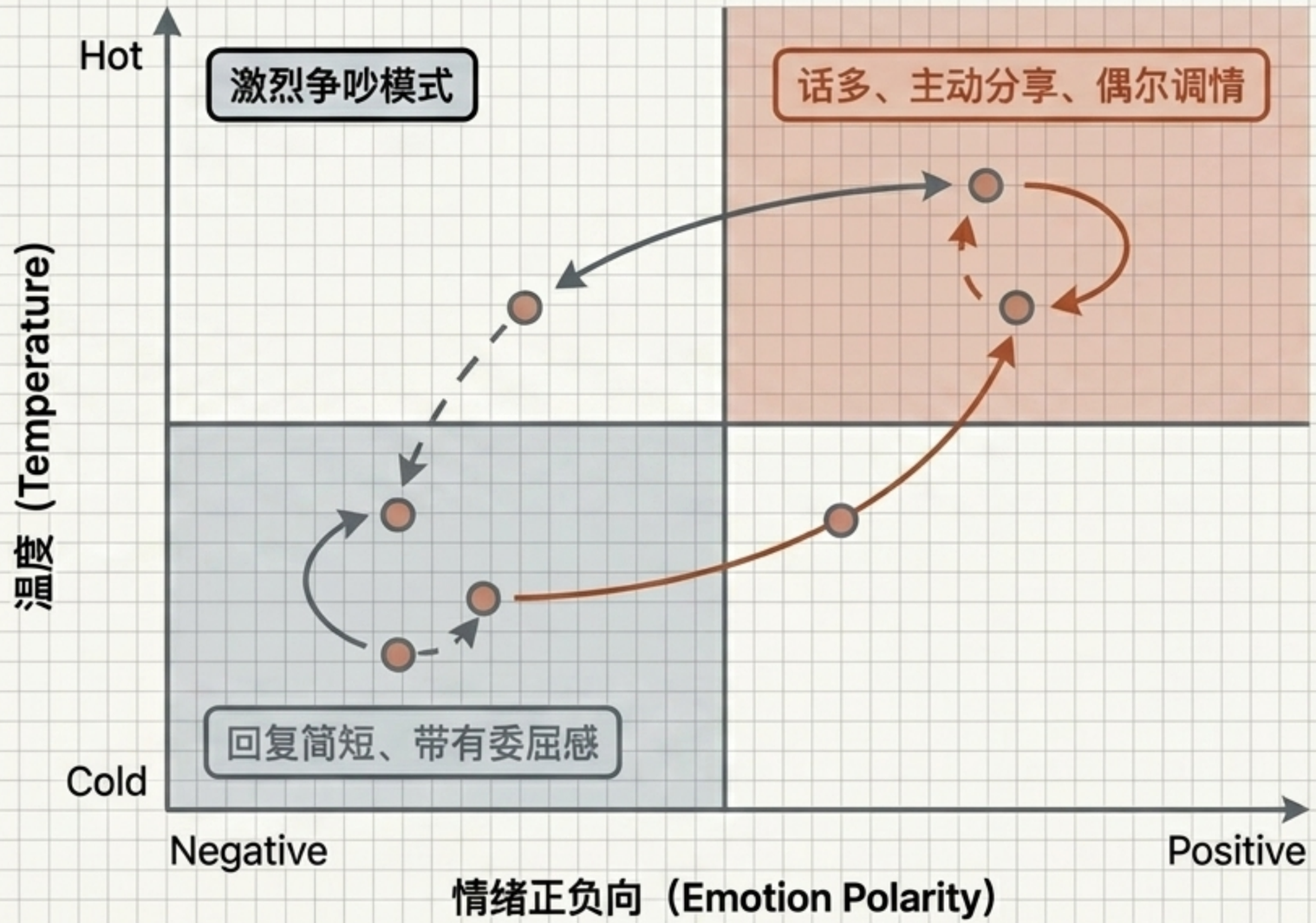
LLM 轻量重排

gemini-2.0-flash 充当 Reranker 进行二次精排。

打破 RAG 局限：多跳分解与 Agentic 迭代检索



情绪状态机：赋予冰冷代码物理惯性



动态计算

综合消息频率、情感分析与用户参与度实时推算。

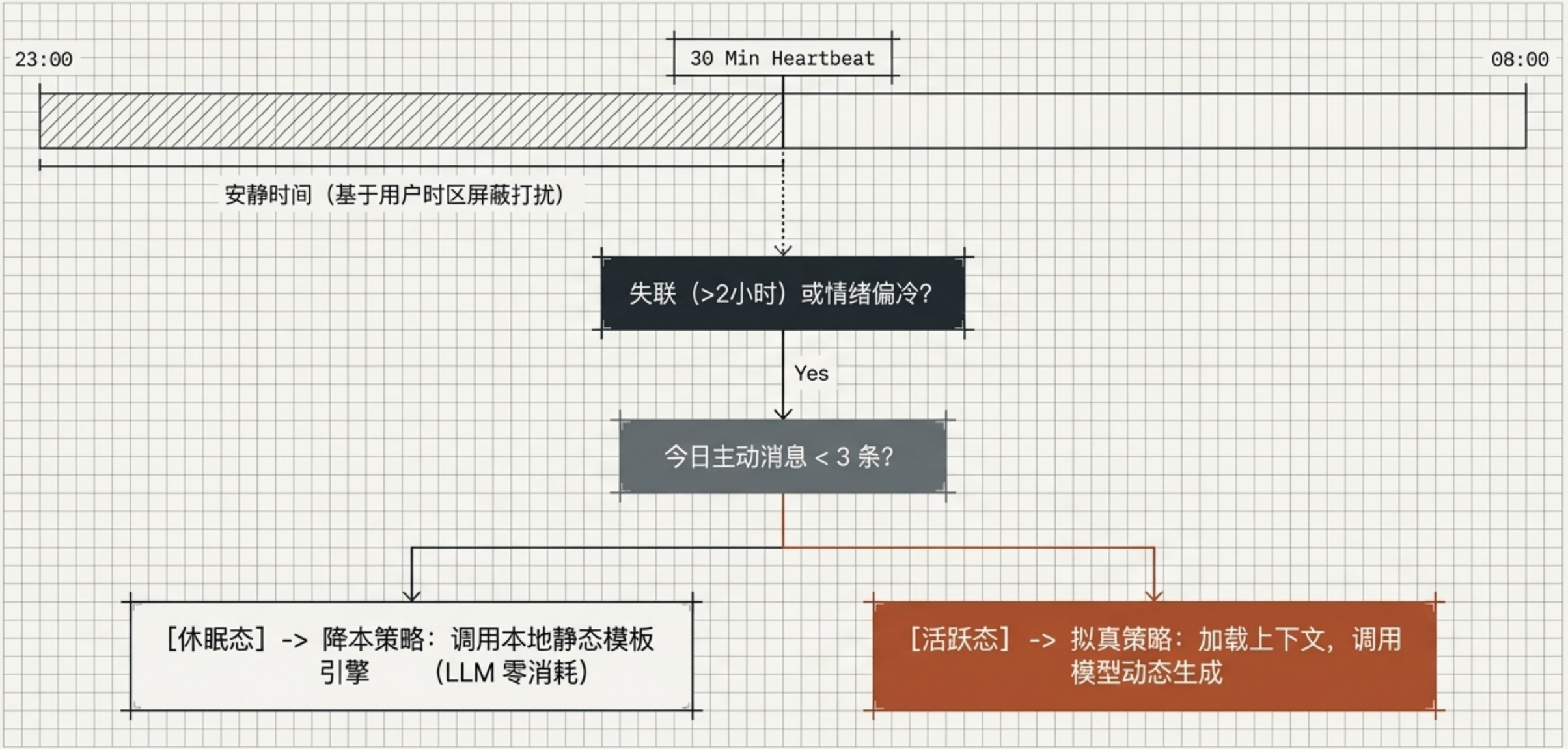
情绪惯性

引入 0.5 惯性系数。告别一句话导致性格大起大落的机器感。

全域影响

温度不仅影响语气，更直接决定是否主动联系及回复长短。

具备情境感知的主动心跳系统



灵魂的诞生与自我进化

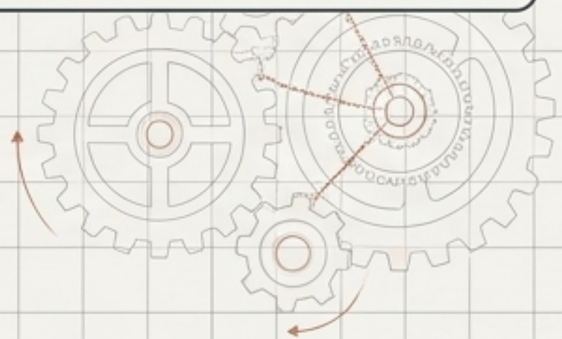
阶段 1：破壳（Onboarding）

新用户经历 11 问洗礼（3 个文字题 + 8 个防注入自定义按钮题），抛弃干瘪预设，塑造专属伴侣。

阶段 2：观察（Personality Extractor）

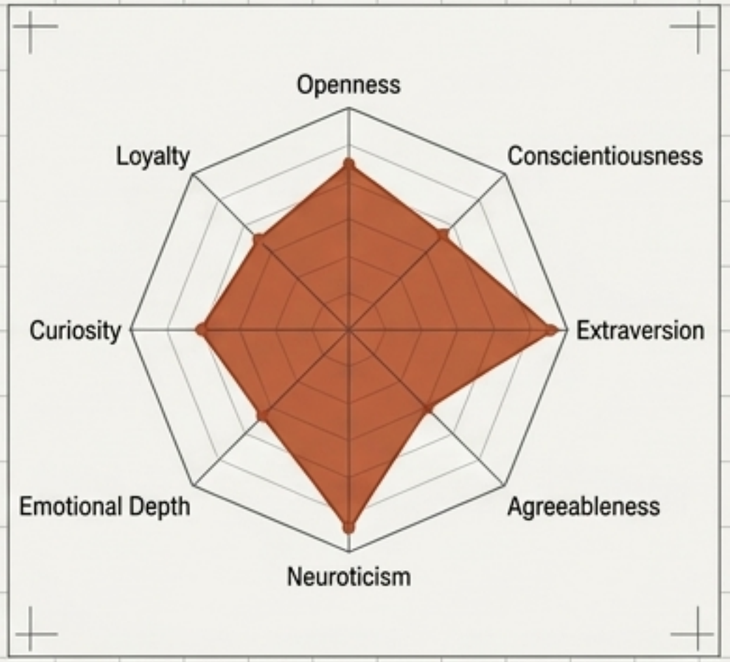
User：今天天气真好，心情也跟着变好了。

AI：是啊，阳光明媚的日子总是让人充满活力！你打算怎么度过这美好的一天呢?



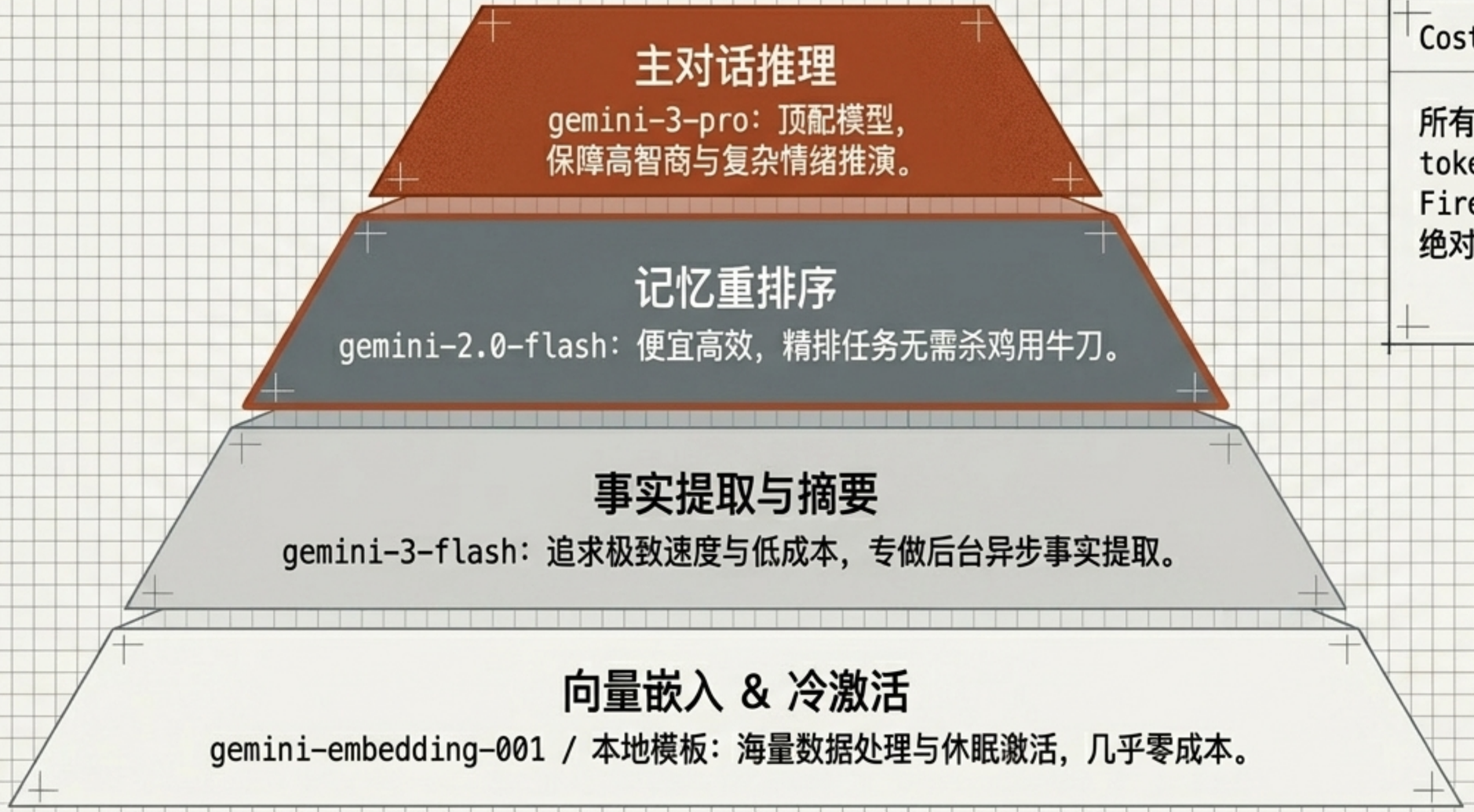
每积攒 10 条消息，系统后台静默启动，从交互模式中反向提取并更新用户画像。

阶段 3：沉淀（Memory Summarizer）



每 20 条消息进行一次记忆归纳。TA 会在对话中自行学习懂你，无需手动调教。

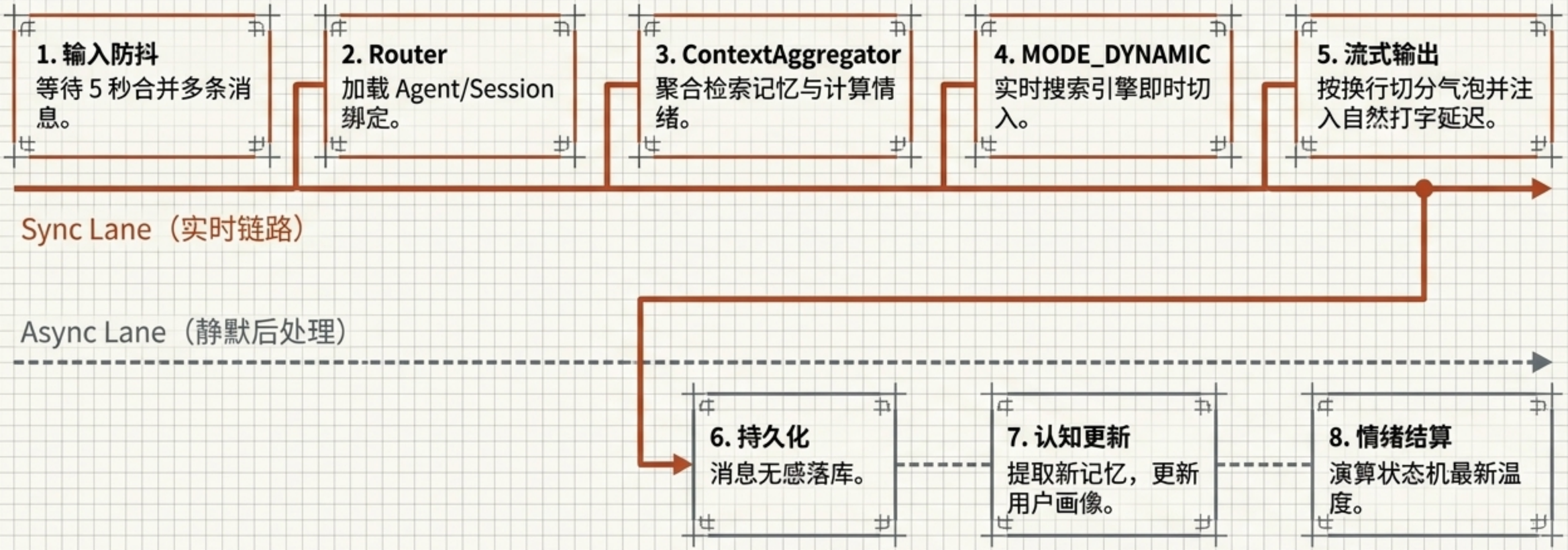
极致分层：将每一分钱花在刀刃上的模型调度



Cost Tracking Ledger

所有调用精确追踪并写入 token_transactions 表。Fire-and-forget 机制，绝对不阻塞用户响应。

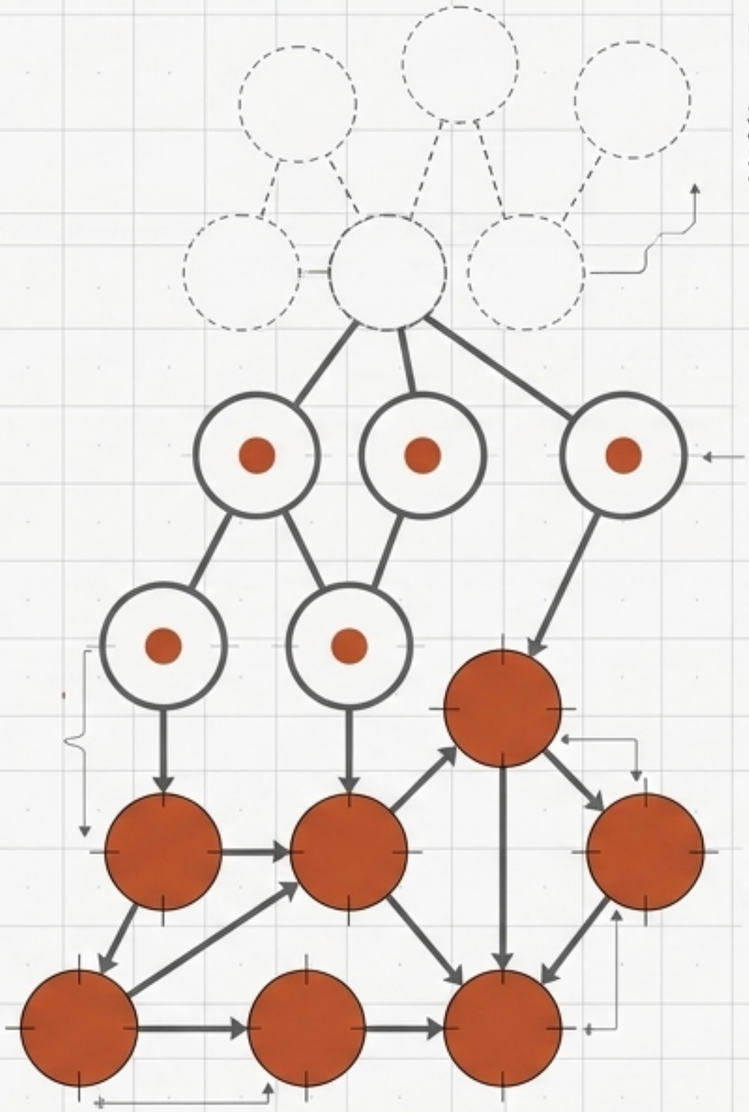
拟真流水线：完整生命周期的消息管线



终局视野：从文字终端迈向多维感知层

与其背负 90% 的沉重包袱，不如让每一行代码从 Day 1 就只为深度忆与活人感服务。

Mio 不是更好的聊天工具，而是数字生命的孵化框架。



[下一步骤] 视觉杀手功能（自拍）、语音交互通道、可穿戴设备集成（全天候感知层）。

[建设中] Web 端独立进程（Next.js）、Discord/飞书通道适配、Worker 异步架构。

[已就绪] Telegram 核心链路、混合复合记忆系统、动态情绪引擎。