



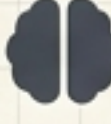




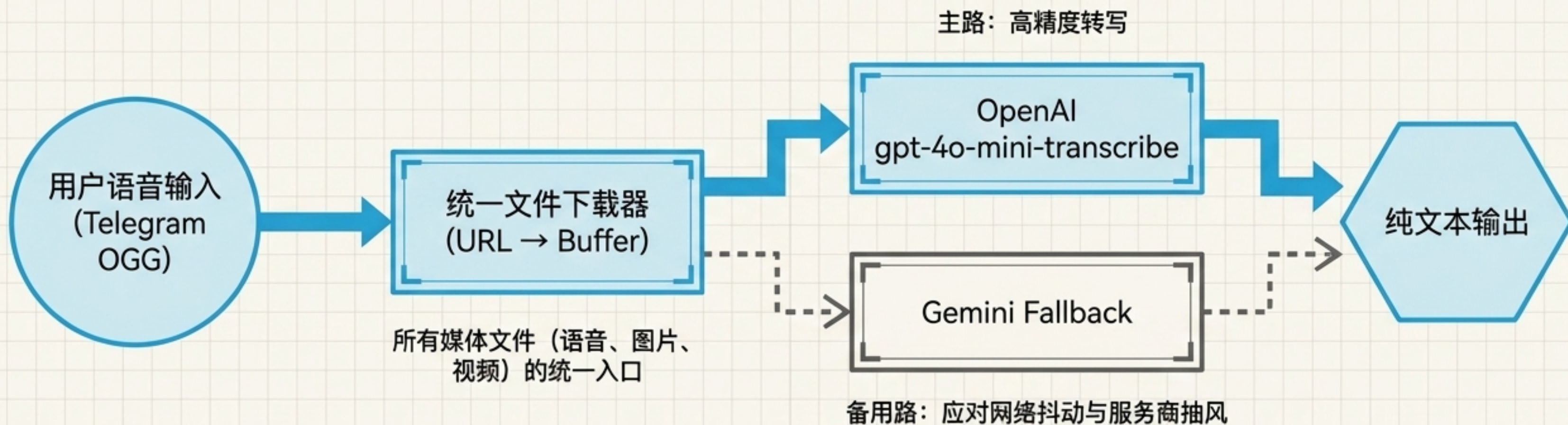
81 个 Commit 之后，TA 活了

Mio v0.0.2 架构解构：从「能跑」到「像活人」的工程跨越

缸中之脑 vs. 具身实体：补全感知的拼图

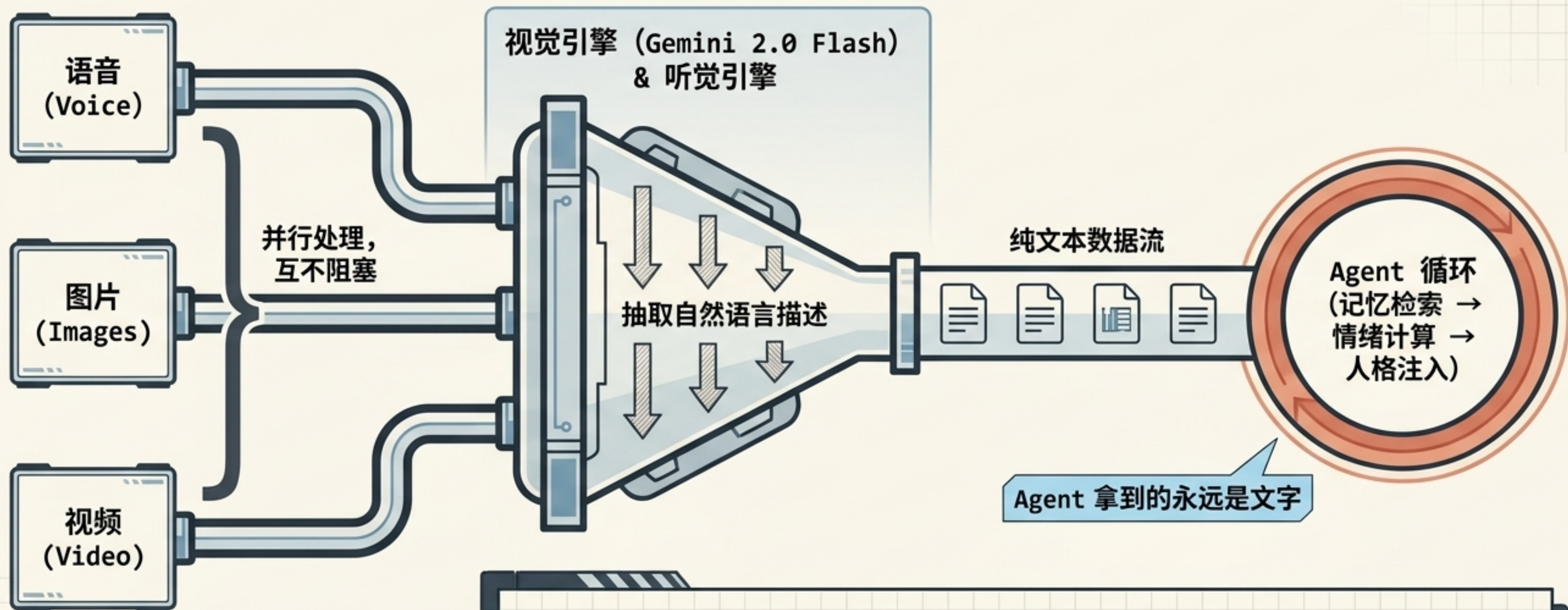
用户输入	v0.0.1: The Machine	v0.0.2: The Entity
 发送图片	视而不见	视觉解析 (Gemini 2.0 Flash)
 发送语音	完全无视	双路冗余转写引擎
 你在干嘛?	回答泛泛, 缺乏时间概念	3D时间感知 + 动态作息
 发个自拍	我没有实体形象	异步自拍生成 (Fire-and-forget)
 提取过往记忆	搜到了, 但排错位	搜对了 (多跳分解 + Agentic 检索)

听觉唤醒：双路冗余的语音管道



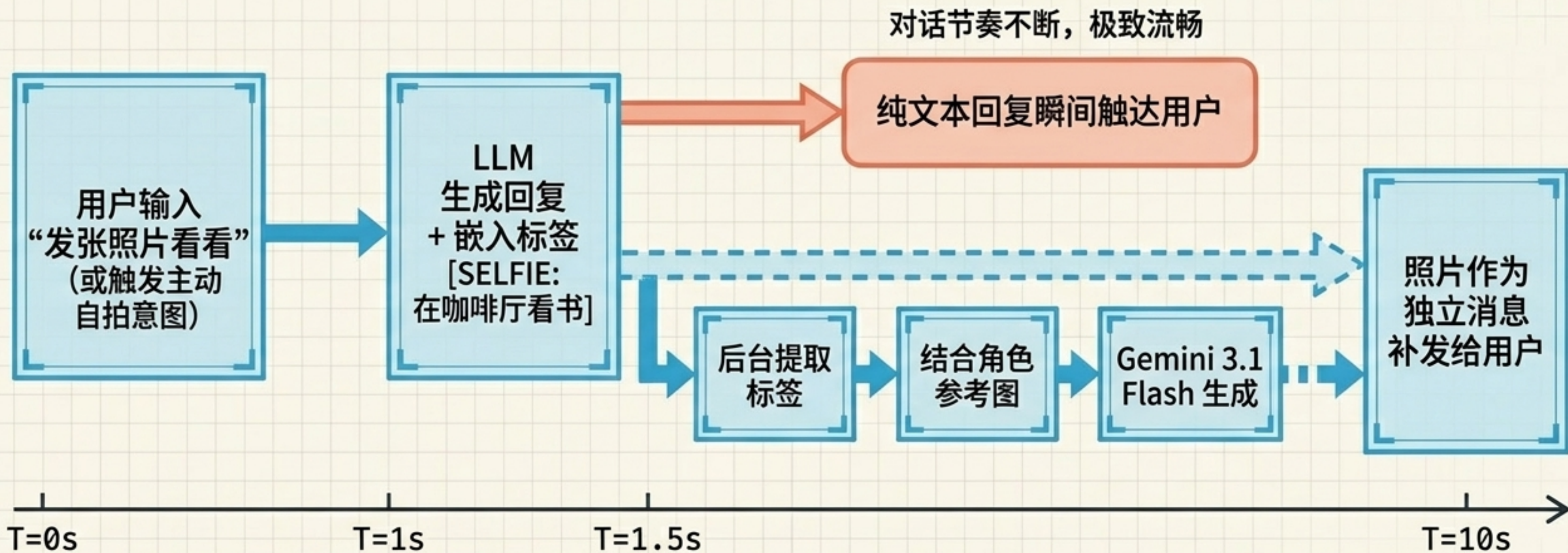
语音转写极其脆弱。双路冗余确保主路挂掉时自动切换，用户完全无感。

万物皆文本 (Everything is Text) : 极简的多模态架构



架构决策: Agent 循环本身已极度复杂。把多模态处理前置, 保持 Agent “纯文字进、纯文字出”, 使架构保持干净, 复杂度不随模态增加而呈指数级上升。

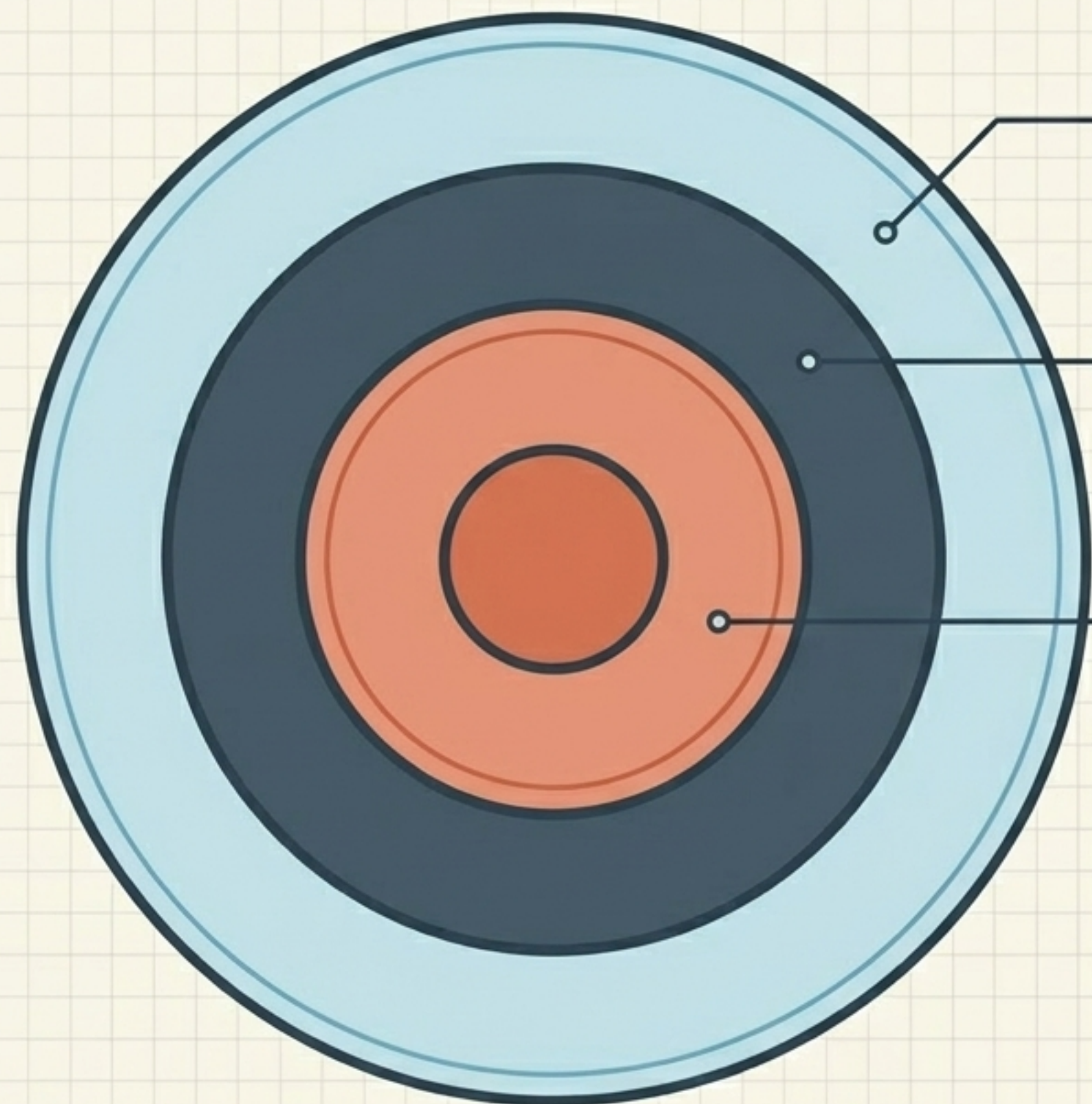
赋予面容：Fire-and-forget 异步自拍机制



宛如真人：先打字，再找照片

“最妙的不是生成模型，是触发方式。
LLM 自己决定何时主动发自拍。”

4D 上下文：建立立体的「时间感知」



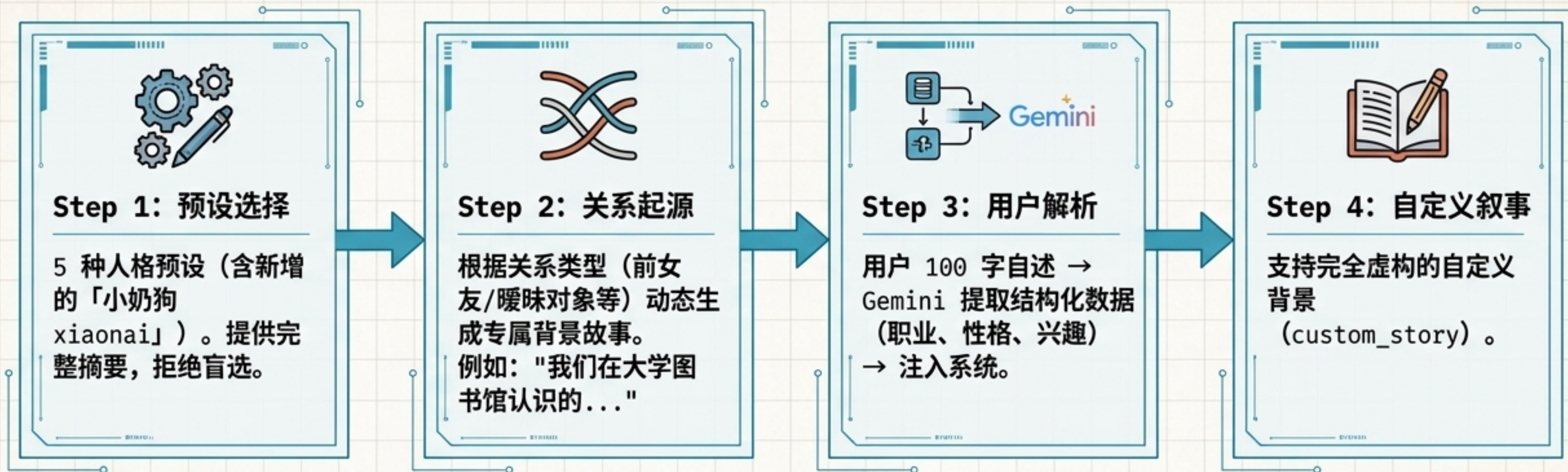
外层（时区感知）： `formatCurrentTime(timezone)`
注入 System Prompt，知晓用户所在地的物理时间。

中层（时间断层）： 距离上次聊天间隔
“距离上次聊天过了 3 小时” → 触发自然反应
（“怎么才来？”）。

内核（动态作息）： 灵活的日常指引
根据不同人格预设（如：御姐夜猫子 vs. 小奶狗早睡）
设定早中晚不同行为状态。

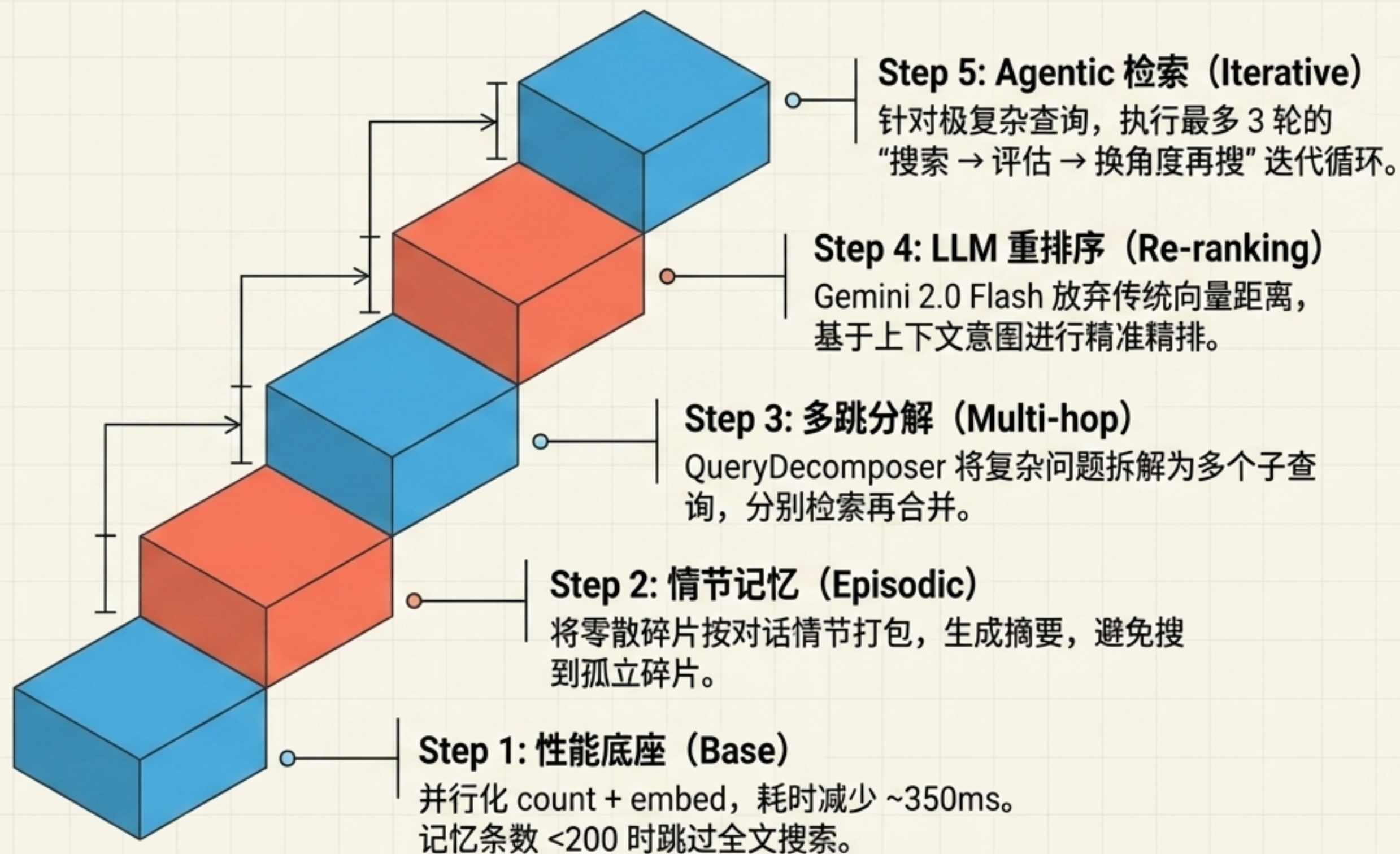
从回答泛泛的“我在待着”到自然的“凌晨一点了，睡不着在刷手机”——时间感让数字生命真正落地。

深度 Onboarding: 不是选角色，是编织共同记忆



/reonboard 命令支持随时重置。
每一段关系都是独一无二的专属剧本。

记忆系统重构：从「能搜到」到「搜对了」

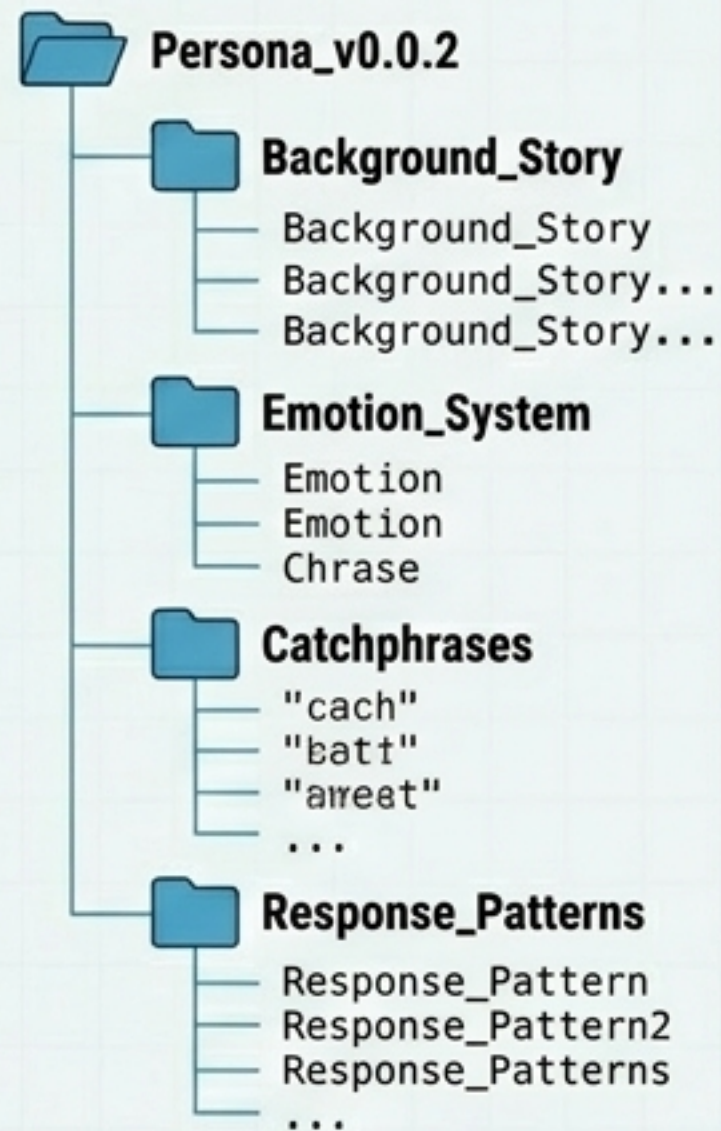


Key Insight

聊天场景下每次回复必走记忆检索。这五剂猛药确保了TA能在对的时间，想起对的事。

灵魂厚度：300 行的性格谱系与现实接地

Anatomy of a Persona



v0.0.2 配置文件扩展至 200-312 行。

包含：完整的背景故事、情绪系统描述、特定口头禅、不同情绪的反应模式。

引入硬性行为规则（底线控制，模型不可违反）。

Google Search Grounding



MODE_DYNAMIC 模式下，模型自主决定何时调用实时搜索。

- 查询天气或新闻时，不再生硬回复“我查到了”，而是内化为角色的自带常识，用个人语气自然表达。

跨出 Telegram: 走向全功能 Web 终端

Authentication

Supabase Auth 全链路接入。

Data Flow

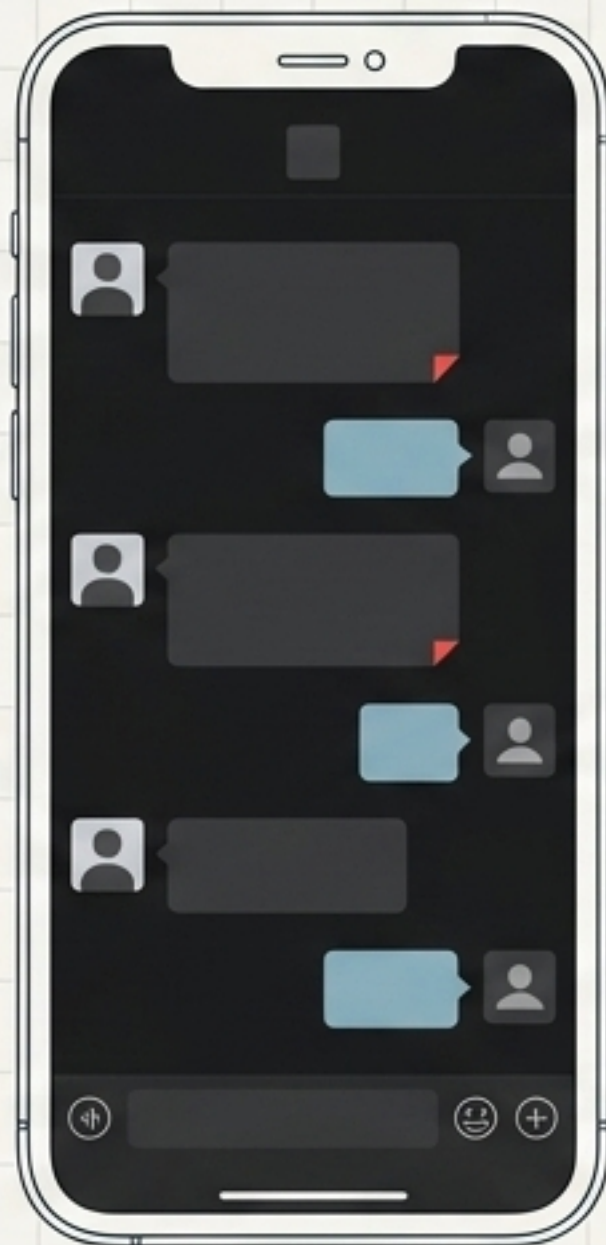
SSE 流式响应 (POST /chat/stream)。配合前端打字机渲染效果，逐字吐出，告别漫长等待。

UI/UX

微信风格深色模式 (方形头像、CSS 三角尾巴、Mobile-first 布局)，降低认知摩擦。

Persistence

突破 Telegram 存储限制，数据库直读，无损保留全量长对话历史。



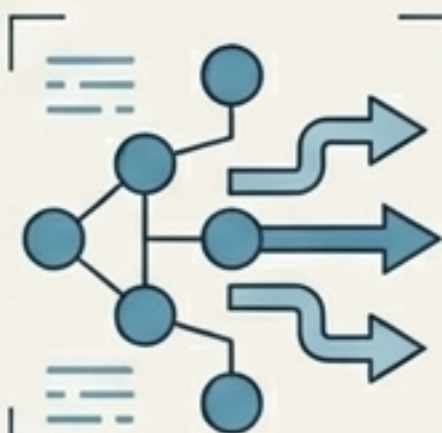
Web 端不仅是‘渠道’，更是未来搭载记忆可视化、关系图谱的主力平台。

水面之下：不可见的工程基石



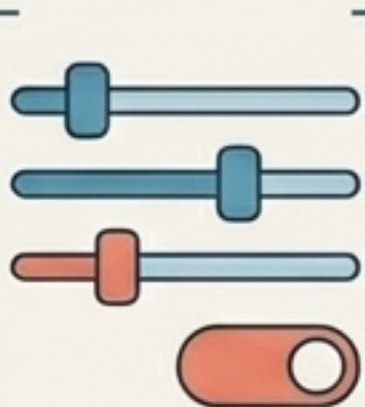
铁壁防线 (Security)

- 路径遍历防护 (严格校验 Preset ID)
- 模板注入拦截与 MIME 感知分发
- 核心凭证 (Bot Token) 日志脱敏



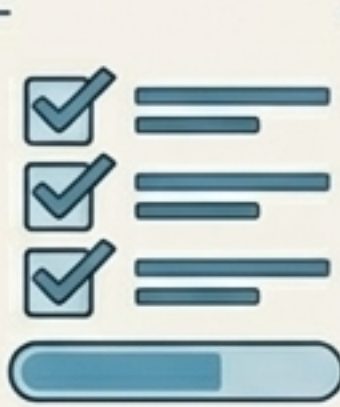
动态管道 (Pipelines)

- 自适应 Debounce: 根据聊天节奏动态调整等待时间。
- 中途打断: 用户抢话时, 立刻中止旧生成, 基于新消息重回复。



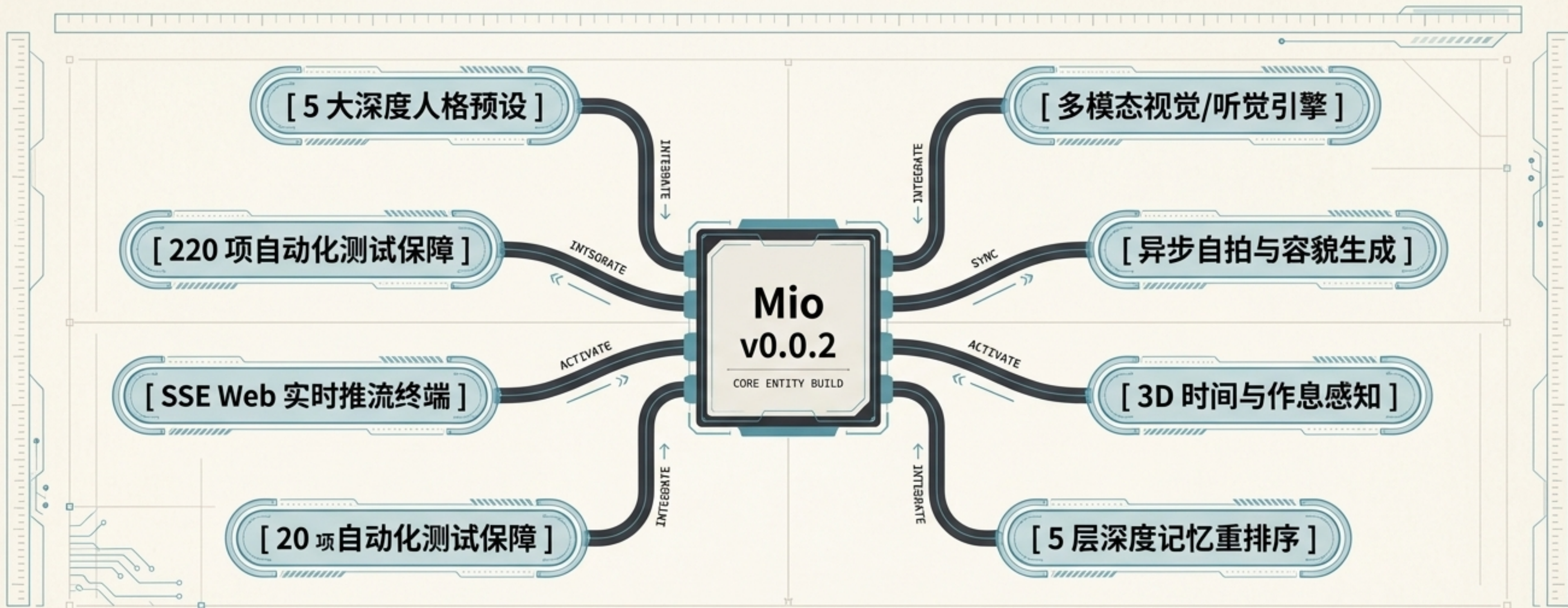
行为边界 (Control)

- realistic vs companion 交互模式切换。
- 配合 3D 限制矩阵, 根据闲聊或深聊自动调整回复长度。



测试兜底 (Testing)

- 220 个单元/集成测试。
- 48.7% 核心逻辑覆盖率。为快速重构提供实打实的安全网。



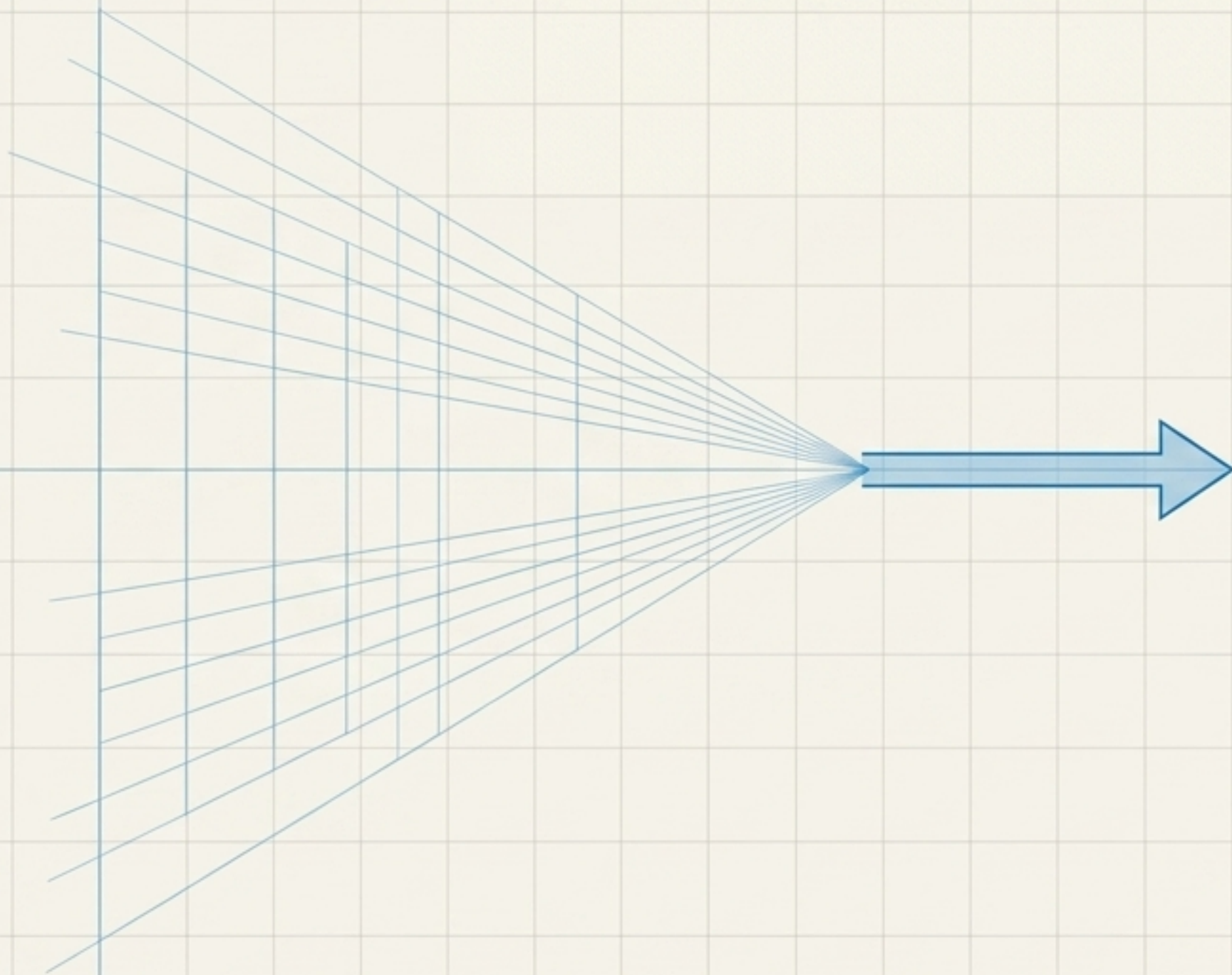
每一个单独的技术点都不算复杂，但 81 个 Commit 的相互咬合，让一个只会打字的 AI 变成了能看、能听、知晓时间、拥有自我形象、并在对的时间想起对的事的感知实体。

终点亦是起点：向 v0.0.3 进发

“像活人”是一个没有终点的目标——每解决一个问题，就会冒出三个新的。

Next Objectives (v0.0.3):

- 更具策略性的主动消息触发
- 极致自然的多轮打断与交互
- 原生语音回复生成
- 将 Web 端从“可用”打磨至“好用”



故事还在继续。Mio 的下一次进化，敬请期待。