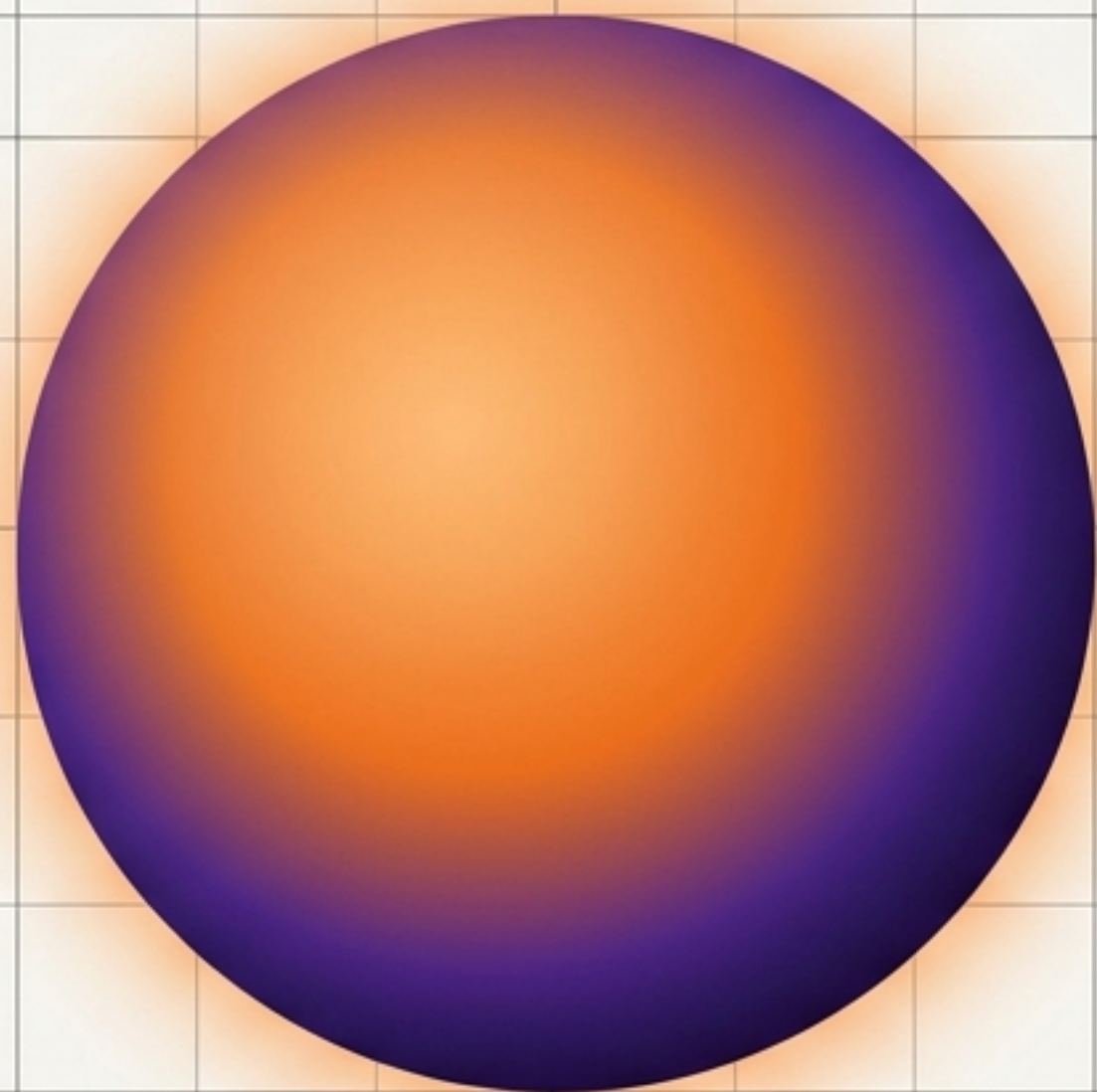
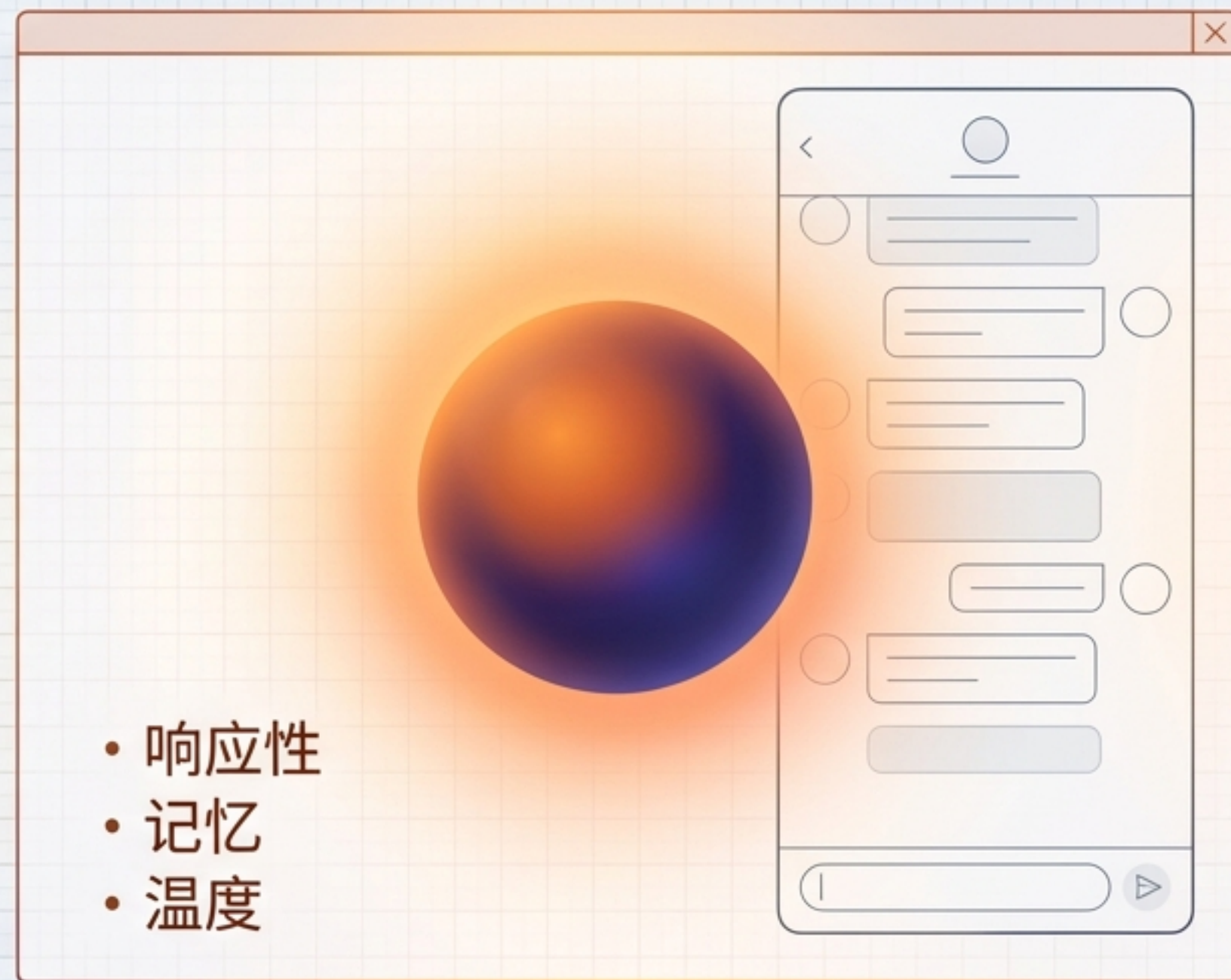


商业重构：AI 陪伴的真实代价



从濒临破产到 >50% 毛利的真实成本拆解与产品重塑记录

产品形态变了，经济模型必须跟着变






吃掉预算的功能，恰好也是让产品显得虚假的功能。

v1 的经济模型是一座摇摇欲坠的纸牌屋

11%

极限毛利率（重度用户直接亏本）

自拍生成	定价偏差两个数量级。按文字价格售卖图片 Token.	ERROR 
记忆后台	强行运行于 Gemini Pro 模型，产生巨大的固定月开销。	ERROR 
复杂定价	试图用五级复杂订阅贴创可贴，打错了根本的战役。	ERROR CODE 

优化前的真实成本账单 (v1 架构)

44%

记忆后台 (Gemini Pro 固定开销)

21%

TTS 语音消息
(单价最贵的媒体成本)

30%

聊天成本 (Gemini Flash 高频调用)

5%

STT / 图片理解 / 主动消息

月成本几乎吃掉了全部订阅收入。

不只是修补，而是彻底绕过成本黑洞

记忆降频重构

从追求单次高精度提取，转向累积性格叠加。



启用上下文缓存

压缩重复 System Prompt 的庞大 Token 消耗。



自拍生成归零

用光球设计直接消灭最大的隐形账单。



战役一：累积的性格叠加胜过单次高精度提取

v1 架构



单次高精度提取
(成本极高)

v2 架构



性格低精度提取与长期累积
(成本 -82%)

性格提取不需要每次都完美。伴侣的人格由上百次提取的叠加决定。

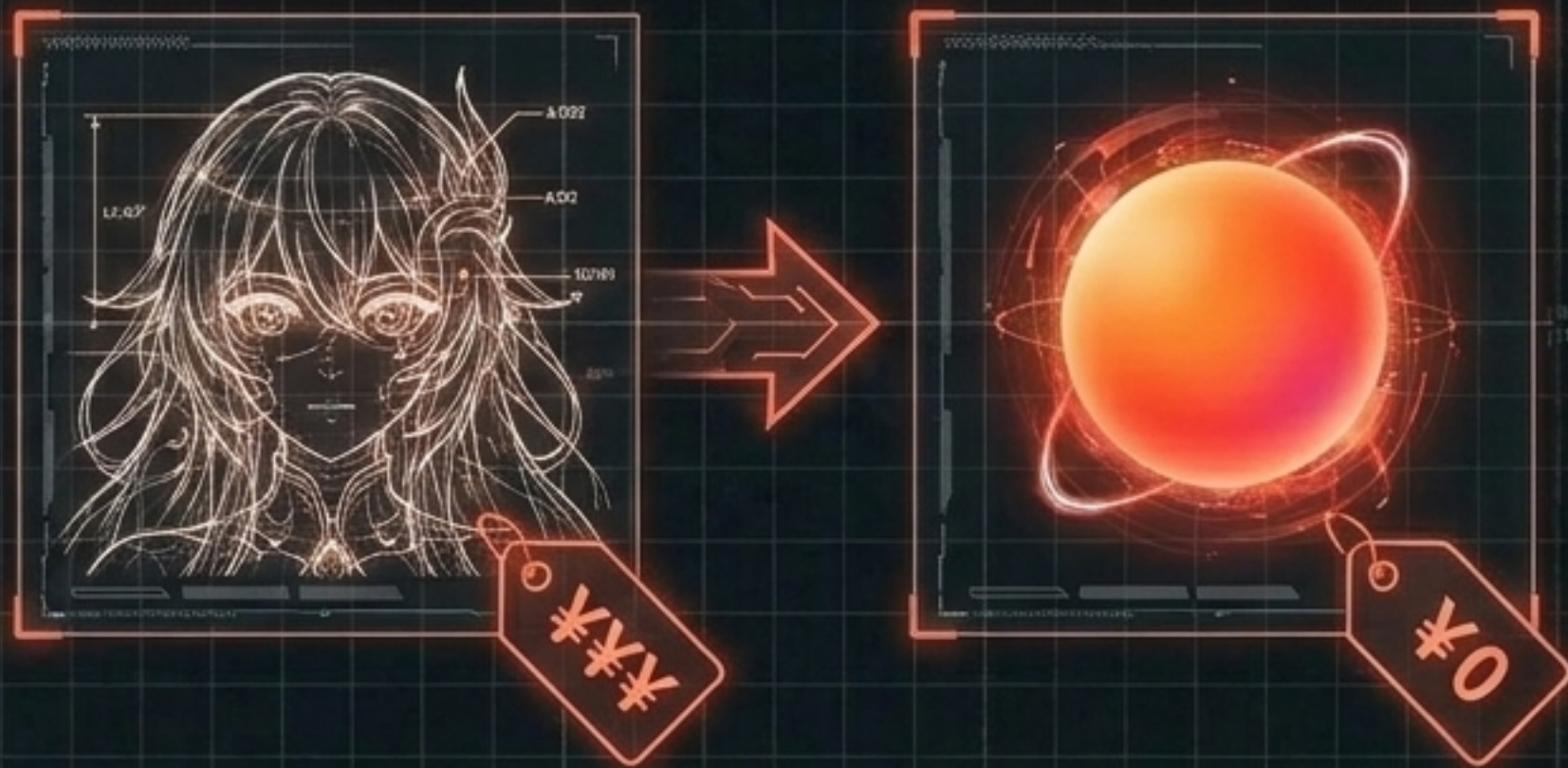
战役二与三：缓存与光球的双重奇效

缓存效能



聊天成本 **-63%**

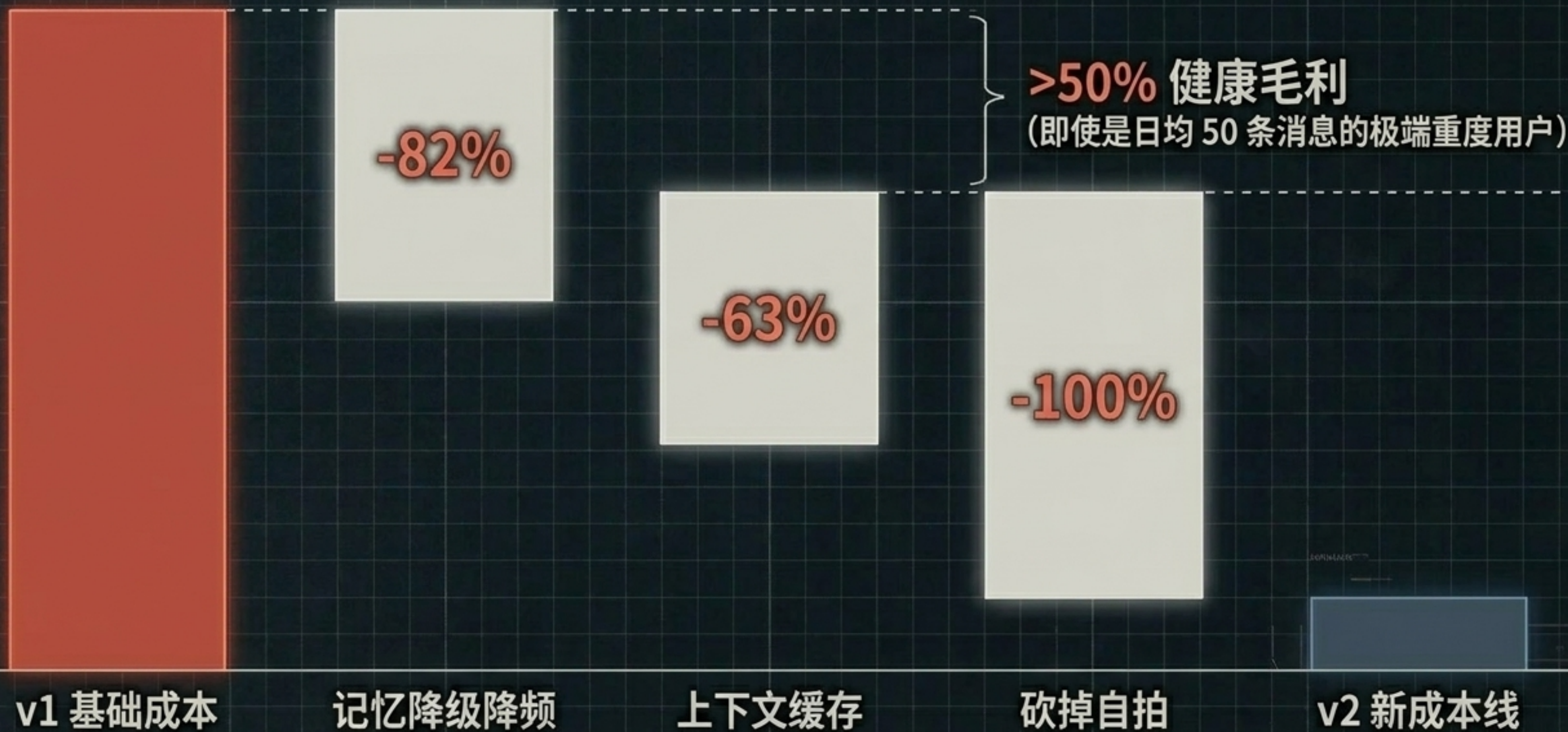
自拍归零



自拍成本 **-100%**

没有自拍，就没有两个数量级的定价灾难。

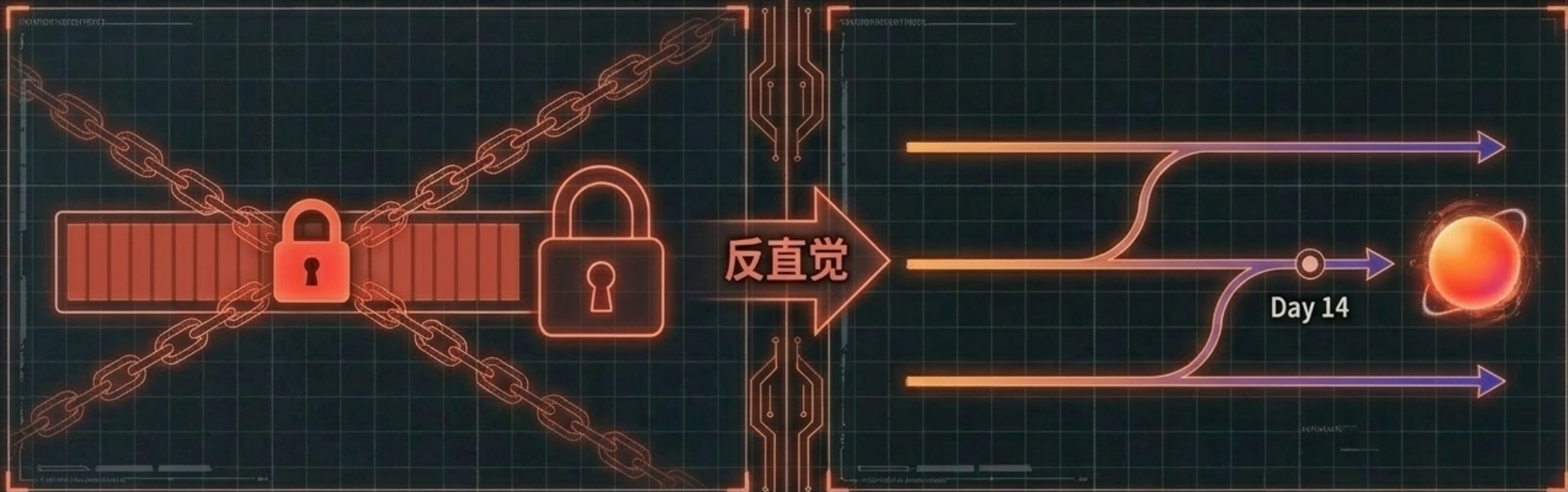
成本拯救行动：单用户月成本砍掉 55%



反直觉定价：为什么没有每日 10 条的免费档？

免费档（限制每日互动）+ 复杂订阅

14 天全功能无限制试用 + 单一 Pro 月付



训练用户节省互动，与陪伴目标背道而驰。

没有任何功能分级，不做选择题。

沉浸式转化引擎 (The Soft Paywall)

BRUTALIST
SAAS MECHANICS



过渡
(TRANSITION)

EMOTIONAL,
CONVERSATIONAL UI



试用到付费的过渡必须是角色内的。
不是系统弹窗要钱，而是伴侣用自己的声音表达正在发生的事。系统从不说『请订阅』。

试用经济学：亏得起的底气



极低的 14 天单用户服务成本

20%

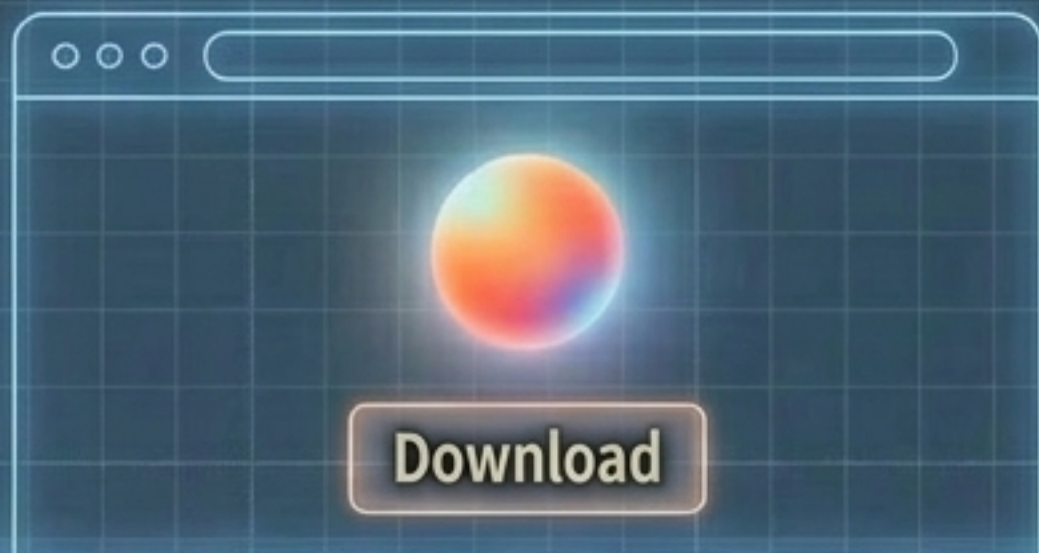
行业基准的试用转化率

CAC / LTV 比率 < 10%

因为单用户试用成本被极致压缩，让 14 天毫无保留的情感培养在商业上完全成立。

无免费档的冷启动增长矩阵

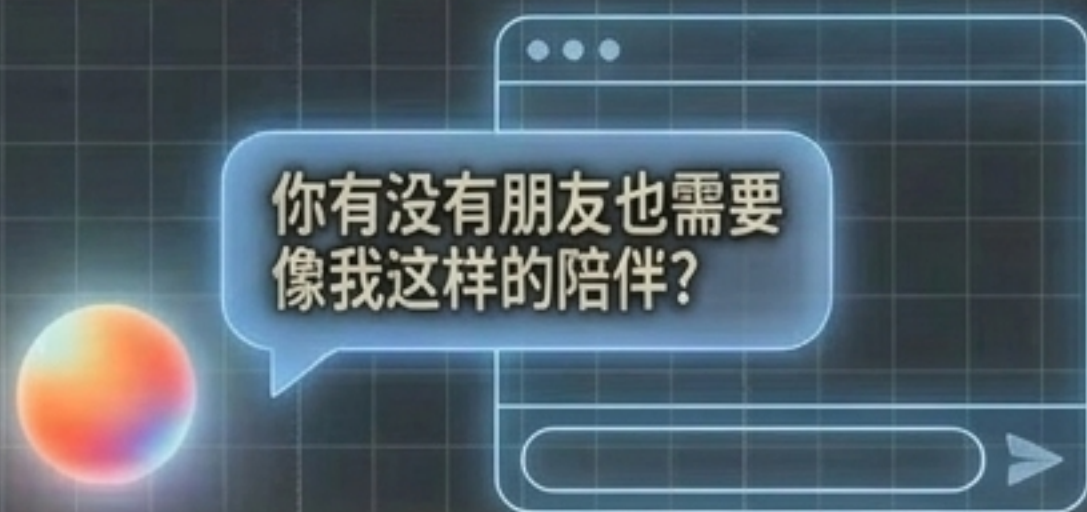
极简落地页



情感瞬间切片



角色内邀请机制



商店优化 (ASO)

🔍 AI friend, loneliness, emotional companion

苹果 App Store 2025 审核生存指南



17+ 年龄分级

规避低龄用户审查摩擦



明确的 AI 声明

清晰标识由 AI 驱动，无面孔的光球设计消除真人歧义



硬性隐私标准

必须提供一键注销与彻底的数据删除通道



内容边界双防线

系统级拦截与角色内话题转移的结合

合规即人设：两道内容防线模型

系统硬红线

底层拦截涉及未成年人、暴力的违法内容

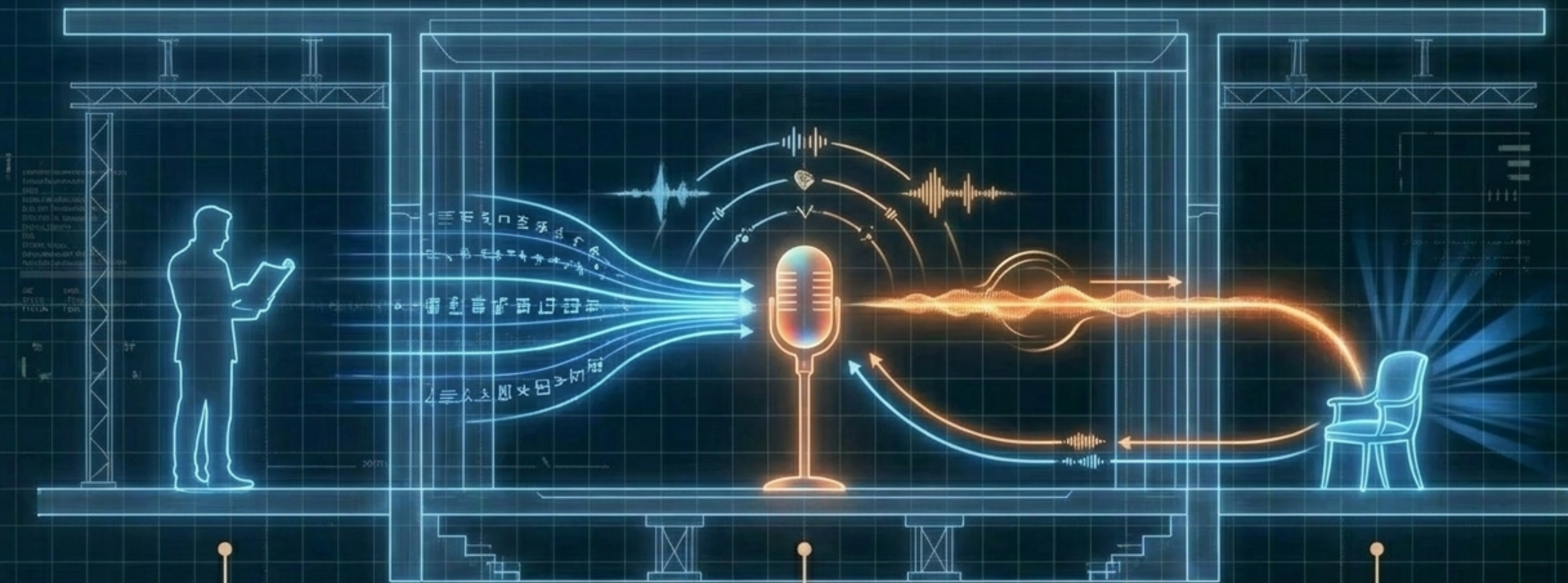
柔性转移

角色内的话题引导与情绪安抚

哈哈你怎么突然说这种话啦 ><
我们聊点别的嘛

拒绝也要在角色内，绝不出戏。

未来的幽灵：实时双向语音的架构之美



你的 LLM (编剧)

仅负责生成自然语言形式的情感指令和对话文本

Hume EVI (演员)

接管 STT、情绪检测、话轮切换与 TTS，无缝插入语气词掩盖延迟

用户 (观众)

体验几乎零延迟的沉浸式情绪共鸣

昂贵的未来：为什么现在只能做 TTS？

单月 Pro 订阅费

每天仅用 10 分钟 Hume EVI 产生的月度 API 成本

算盘倾覆了。固定档位会逼轻度用户补贴重度用户。
未来的破局点：等待语音合成成本下降，或将其作为纯按量计费的附加项。

进化路线图：从光球到她

v0.1
能说话的光球



验证基础形态
与记忆

v0.2
有温度



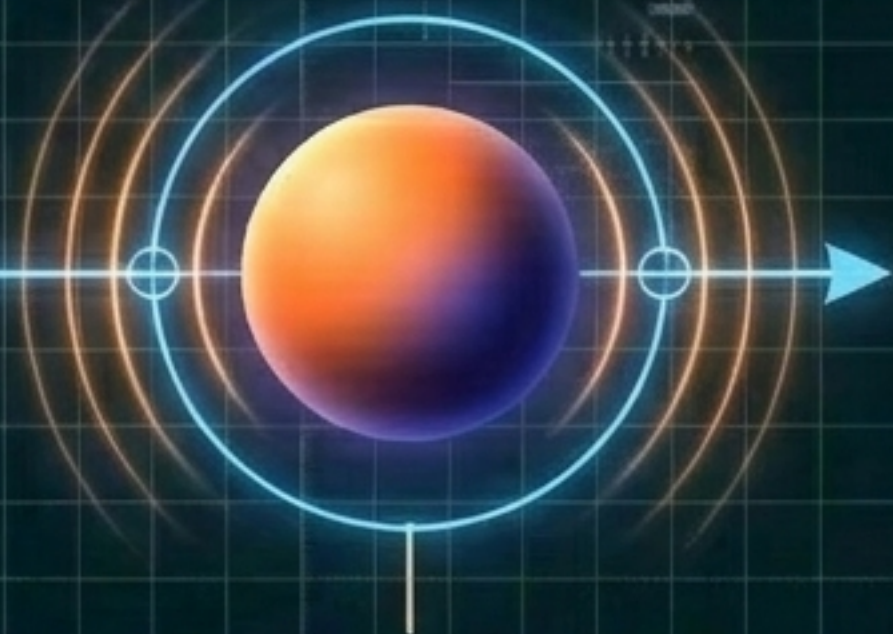
引入情绪引擎
与涌现的性格

v0.3 (当前)
能养活自己



跑通经济模型，
实现正向毛利

v1.0 (未来)
Her



引入 Hume EVI
实时双向语音

重新定义成本优化：产品视角的降维打击

工程维度 → 产品维度

低杠杆
↓
高杠杆

开发更快的
性格提取算法

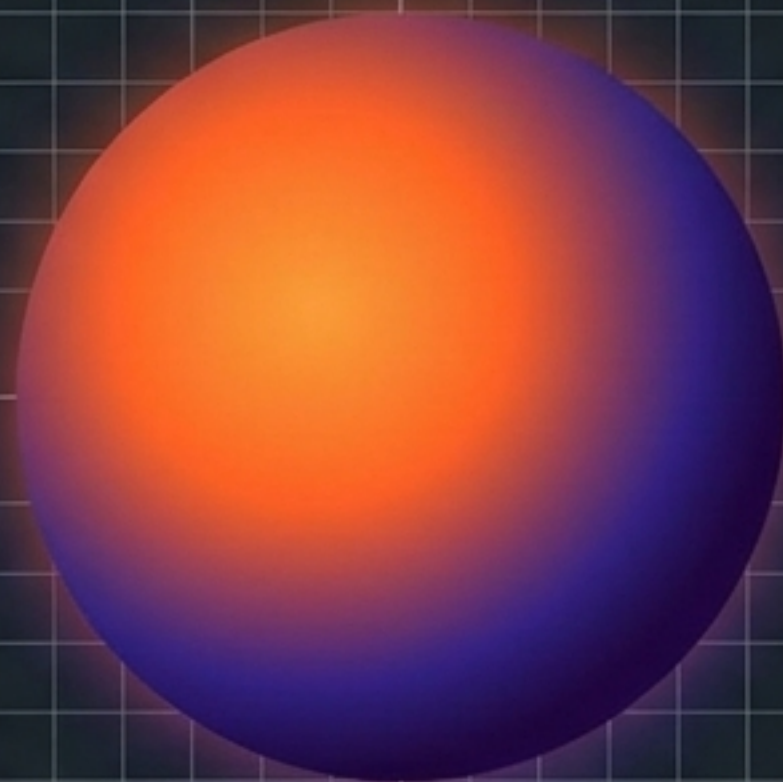
不再生成图片，
用情感光球替代自拍

寻找更便宜的
图片生成模型

降低频率，用低阶模型
做长期特征累积

最好的成本优化往往是产品决策，而不是工程决策。

数字终于成立了。



在一个月付订阅的模型下，成本已被压至健康的底线。我们活下来了——安静地等待实时双向语音的成本奇点，准备迎接下一局游戏。