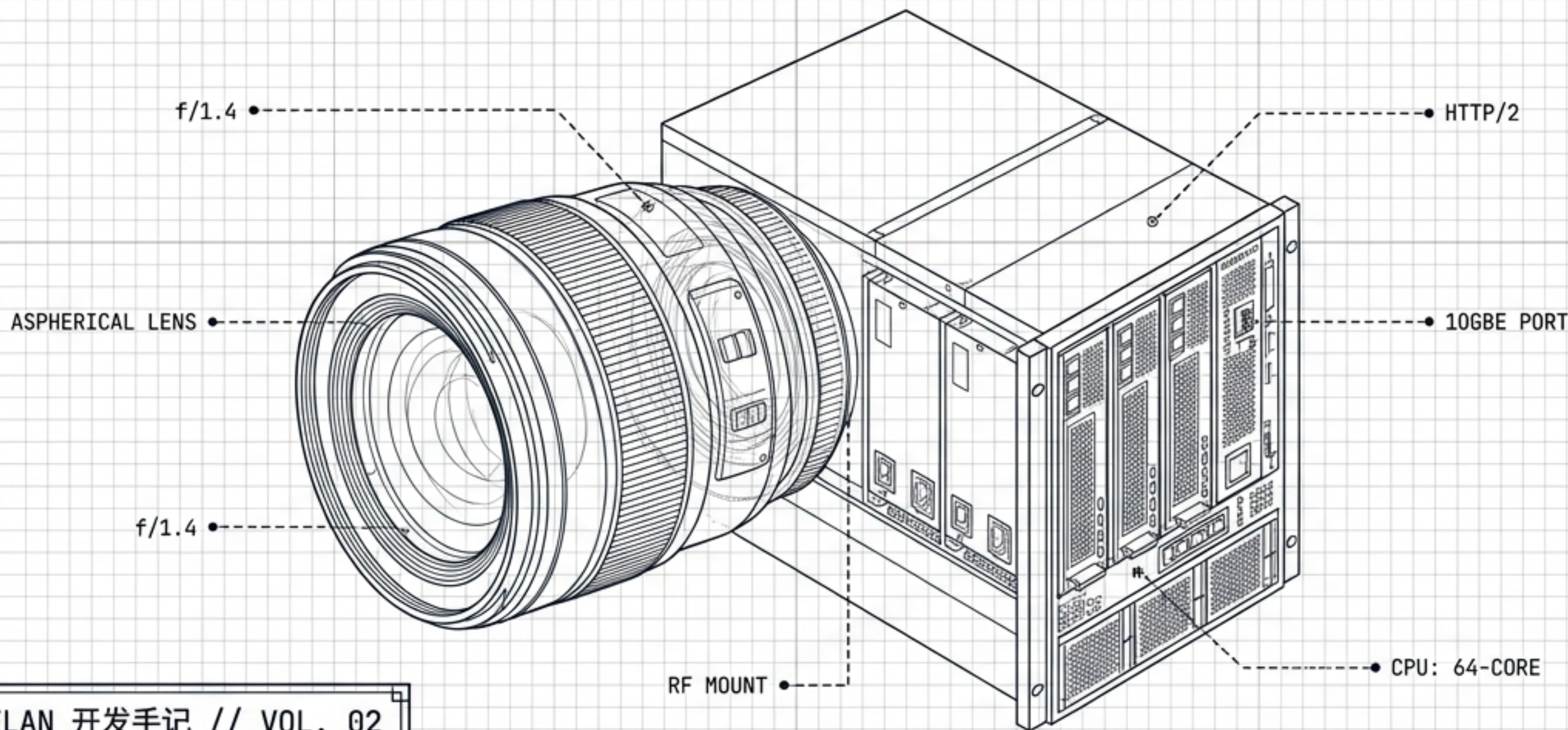


打开引擎盖：ÉLAN 的精密蓝图

ENGINEERING TEARDOWN
(10-PART PROMPT / SSE
/ COST STRUCTURE)

从幻觉到现实 / 逆向工程“不经意的高级感”



SERIES: ÉLAN 开发手记 // VOL. 02

X: 0100 Y: 0200 Z: 0300

X: 0100 Y: 0200 Z: 0300

X: 0100 Y: 0200 Z: 0300

X: 0400 Y: 0500 Z: 0600


X: 0400 Y: 0500 Z: 0600

X: 0400 Y: 0500 Z: 0600

毫不费力的背后，是极度精密的算计。

目标：交付完整的社交媒体时刻（照片+文案），投射“不经意的优越感”。

现实：长达 10 段的模块化 Prompt、对抗 AI 本能的指令、复杂的流式架构与毫厘必争的成本控制。



“一张灵感卡怎么变成 4 张照片和一段文案？一次生成到底花多少钱？现在，打开引擎盖。”

```
{  
  {'image': 'url',  
   'prompt': 'text'}  
}
```

```
{  
  {'image': 'url',  
   'prompt': 'text'}  
}
```

```
GET /api/v1/generate  
GET /api/v1/generate...  
{"text": "generate grob ebnmerts,  
{"name": "teakbaving a schedule crbcm  
{"header": "testnameng.hatet&M0; setur  
ay; poi can] 11b)  
{"image": "url", "prompt": "text"},  
{"image": "url", "structured af tign  
l, "prompt": "text"},  
l, "prom GET /api/v1/generate  
l, "prom GET /api/v1/generate  
{"image": "url", "prom GET /api/v1/generate  
{"image": "url", "prom GET /api/v1/generate  
{"image": "url", "prompt": "text"  
{"image": "url", "prom [0x1A, 0x2B, 0x3C....],  
{"image": "url", "prom [0x1A, 0x2B, 0x3C, 0x1B, 0x1C, 0x2B  
{"image": "url", "prom [0x1A, 0x2B, 0x3C, 0x1B, 0x1D, 0x2B  
{"image": "url", "prom [0x1A, 0x2B, 0x3C, 0x1B, 0x1C, 0x3B  
{"image": "url", "prom [0x1A, 0x2B, 0x3C, 0x4B, 0x1B, 0x2B  
{"image": "url", "prom [0x1A, 0x2B, 0x3C, 0x4B, 0x2D, 0x2D  
{"text": "05", prompt [0x1A, 0x2B, 0x3C, 0x3B, 0x1B, 0x3C  
{"text": "41", prompt [0x1A, 0x2B, 0x3C, 0x1B 0x10 0x20  
[0x1A, 0x2B, 0x3C, 0x1B NotebookLM
```

拼装咒语：一张照片背后的 10 段式蓝图

第 1 段：身份锚定

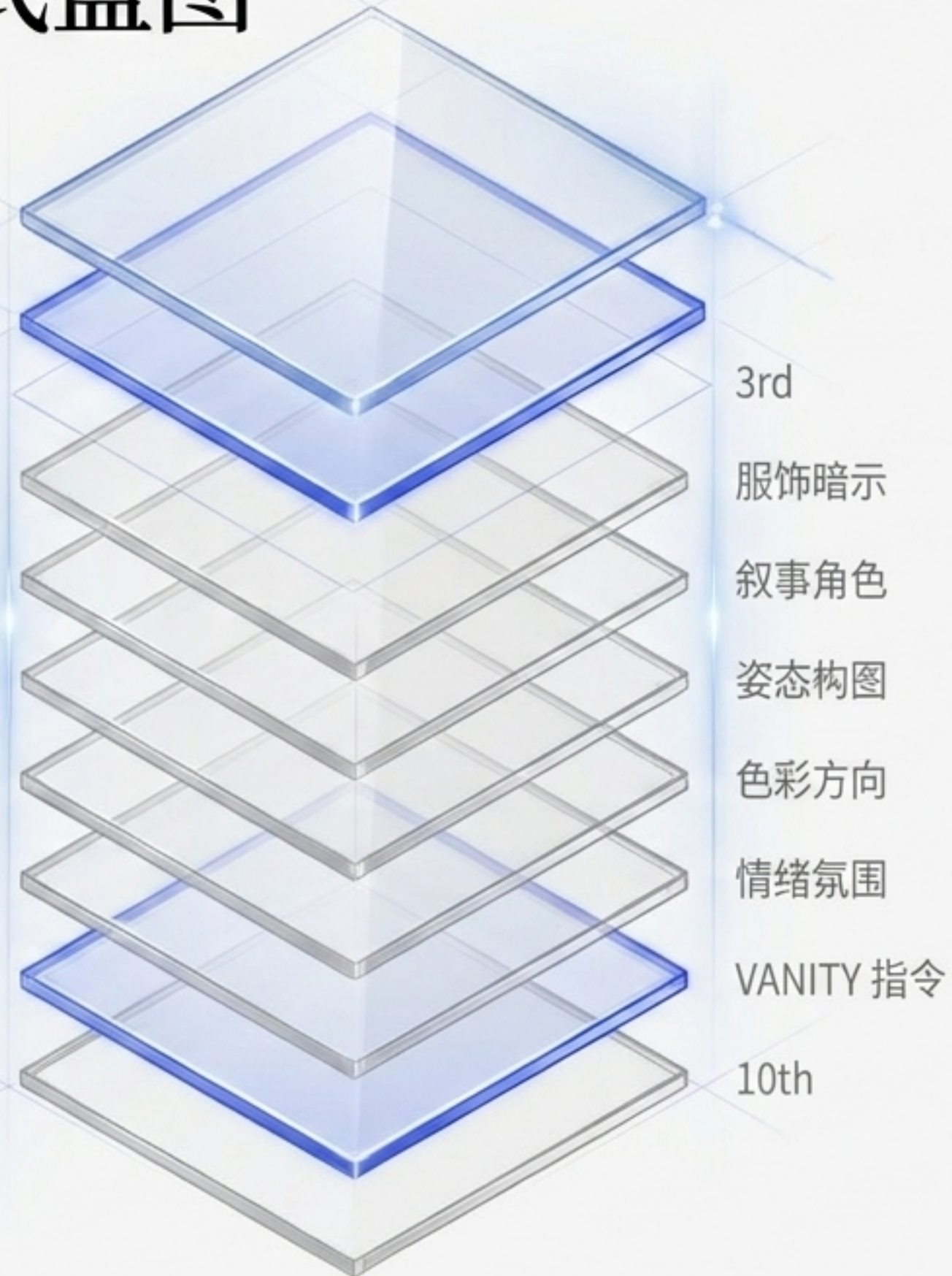
故意极简。身份描述写多反而跟图片抢注意力，文字只说“用那些照片”就够了。

第 2 段：中文场景描述

针对中国用户审美，中文构图比英文翻译更“对味”。

第 9 段：身份锁定尾注

“风格和身份冲突时，选身份。”强制覆盖模型偏好。



视觉矩阵：一套完整的社交叙事

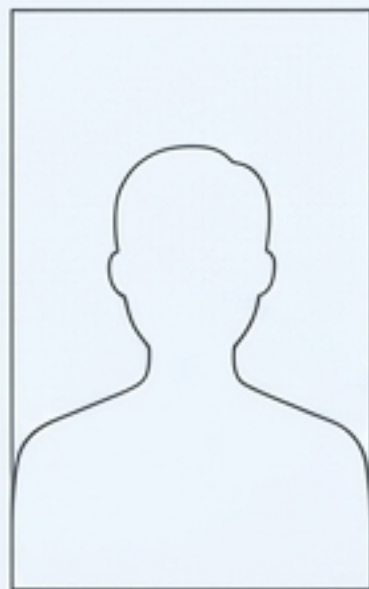
1

Establishing (全景建立)
交代场景与空间尺度



2

Portrait (中景聚焦)
锁定人物状态与主体情绪



3

Detail (局部特写)
脚踩进水里、手搭在书上。
营造真实生活感



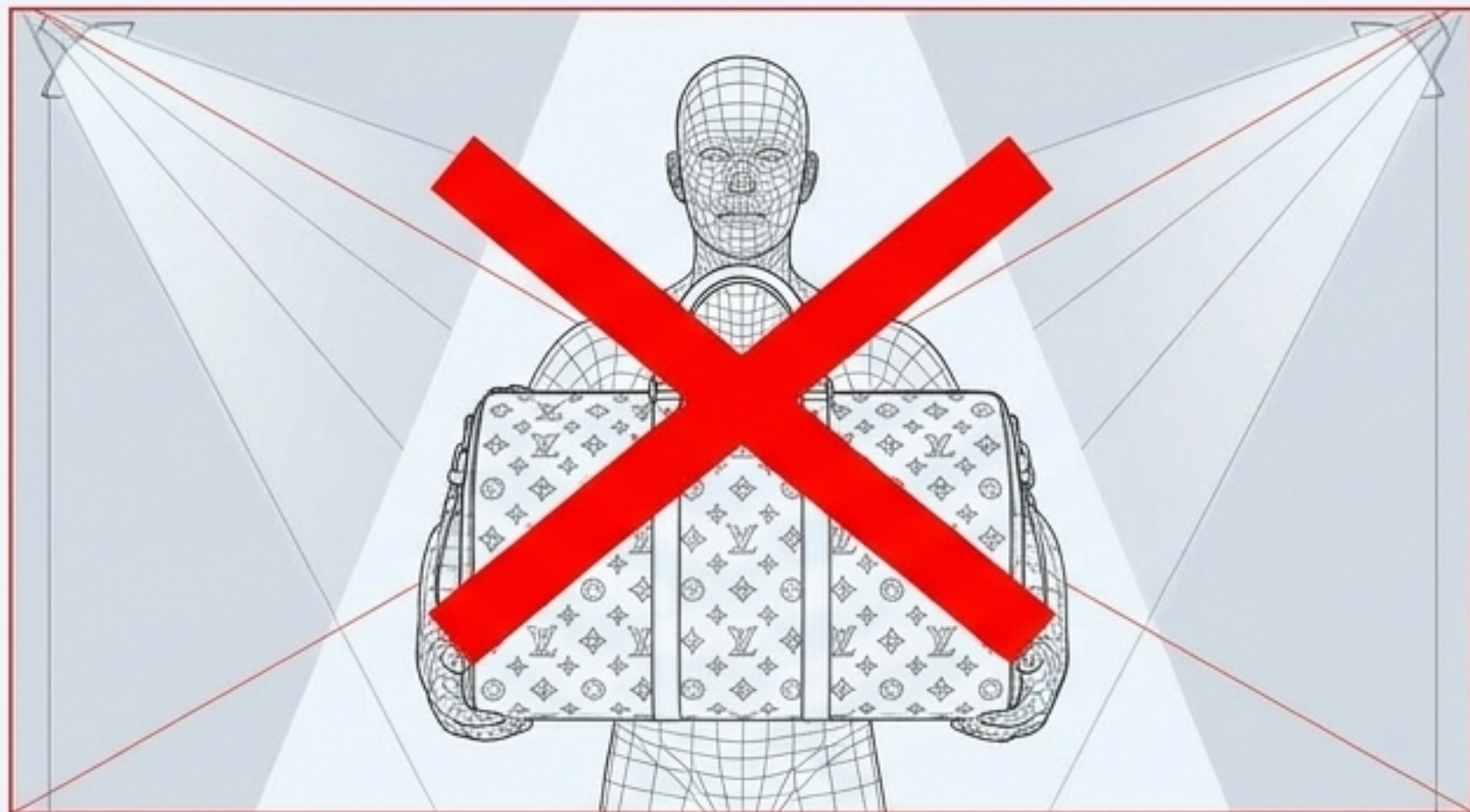
4

Mood/Closing
(氛围收尾)
通常是剪影或背影，留下呼吸感

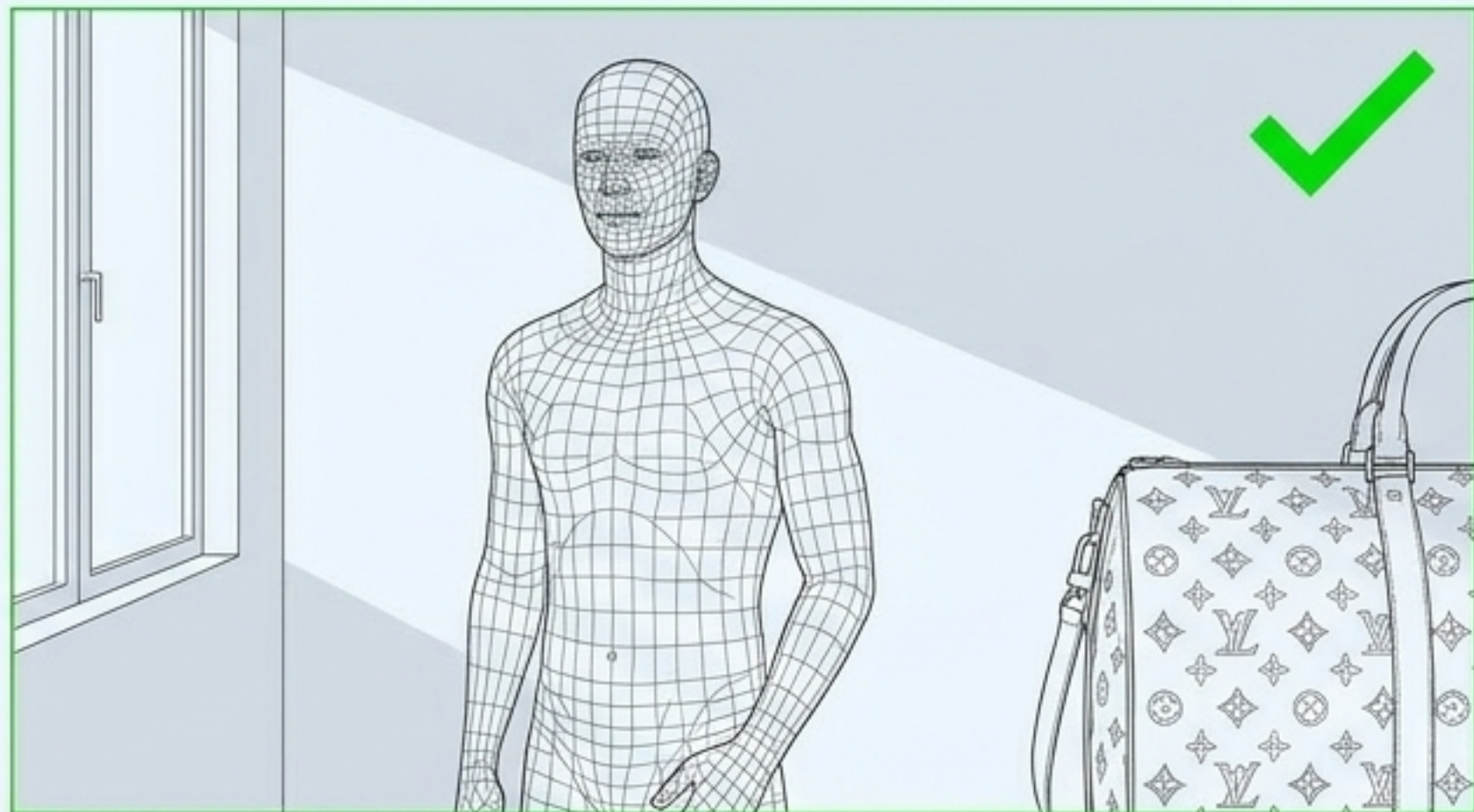


拼在一起是一个视觉叙事，而不是随机四张照片。

核心干预：驯服 AI 的“广告本能”



AI 默认本能 (商业广告)

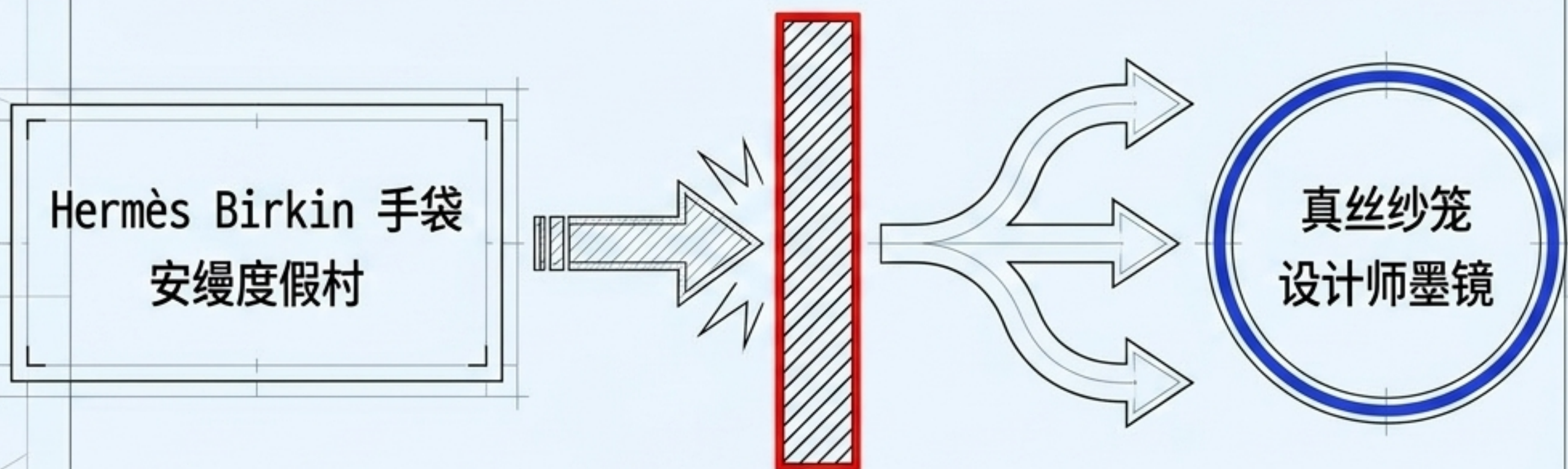


ÉLAN VANITY_DESIGN_INSTRUCTIONS

1. 不经意出现，绝不放正中。(覆盖产品摄影默认行为)
2. 部分可见或自然角度。(拒绝完整正面 Logo)
3. 随拍生活瞬间，不是广告。(打断杂志风/商业风构图意图)
4. 焦点在人和氛围上。(奢侈品是语境，人才是主体)
5. “这只是我普通的周二”。(口语化情绪指令，极其稳定)

降维暗示：奢侈感的合法表达

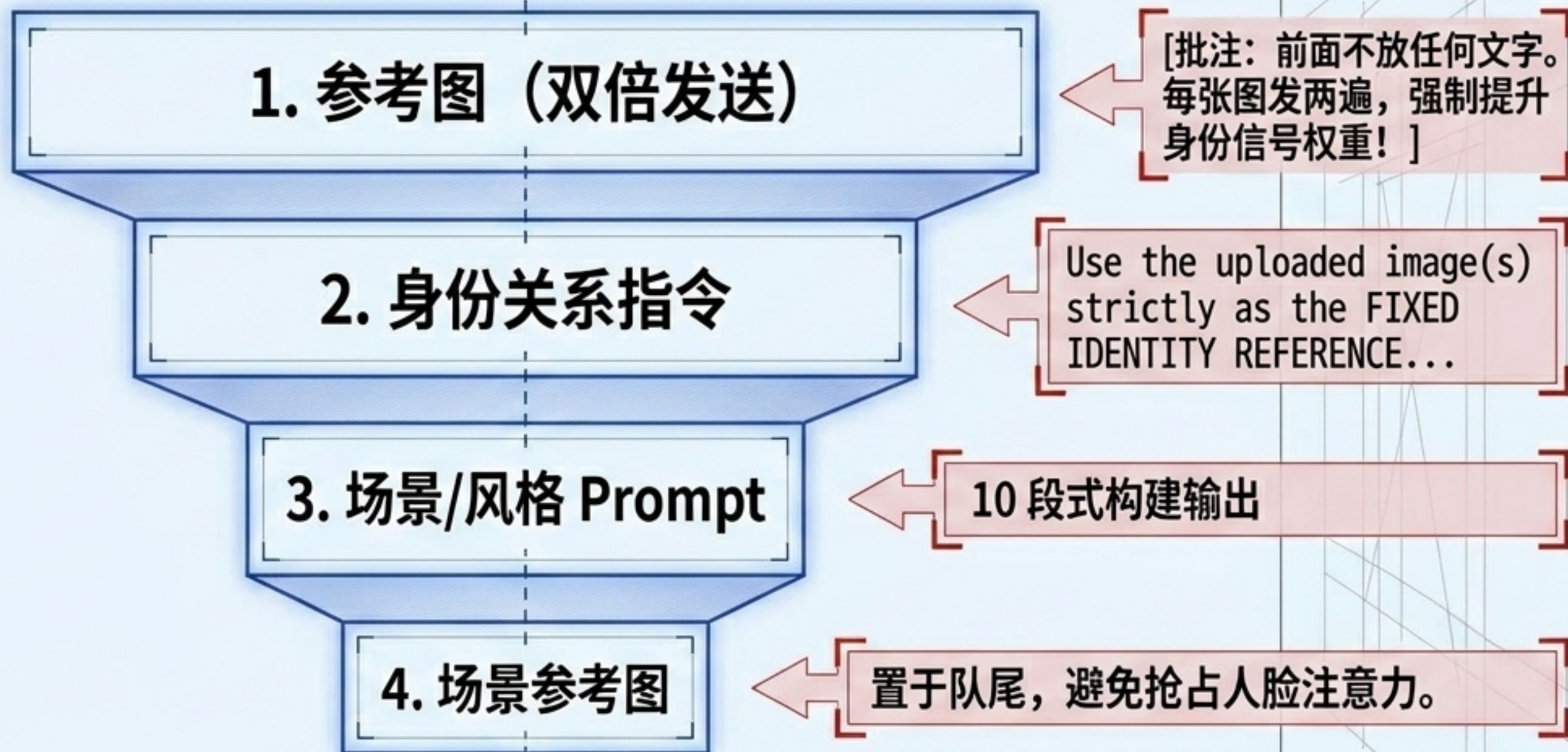
[法律雷区 / 知识产权风险]



品牌名绝不直接写入 Prompt。它们只存在于卡片的 brandHints 数组里，作为 SCENE AESTHETIC 传给模型。

让模型理解奢华的“档次”，而不是去画具体的 Logo。

权重漏斗：顺序决定一切



Gemini 对 parts 数组前面的图片注意力明显更高。

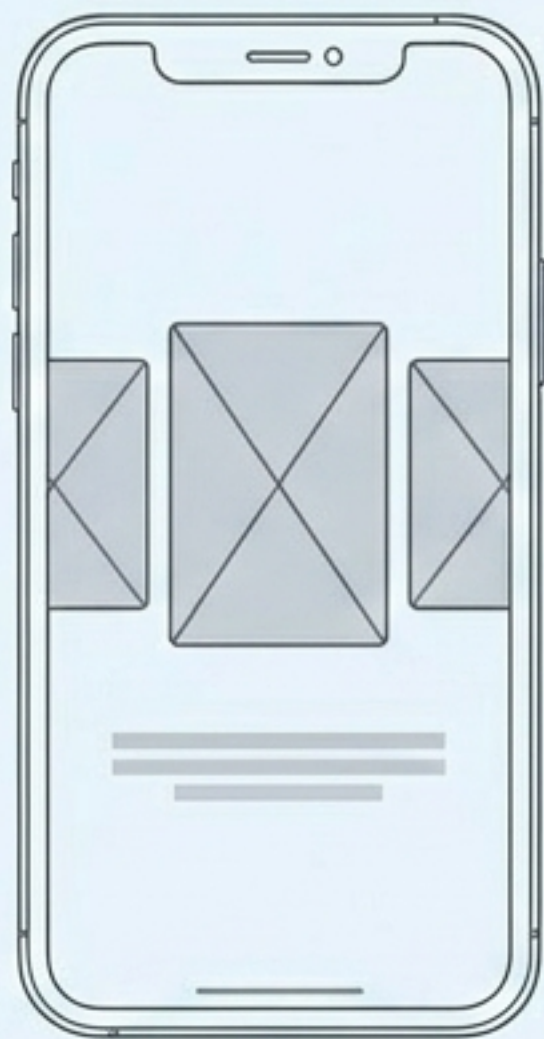
架构决断：如何优雅度过 4 分钟的等待？

	WebSocket	Server-Sent Events (SSE)
数据流向	双向，过度浪费	单向接收，完美匹配
Vercel Serverless 适配度	需复杂 Edge Functions 与长连接池	走标准 HTTP，基建极简
移动端弱网环境	需手动编写恢复逻辑	内置重连协议自动处理

4 张照片并行生成，耗时 15-60 秒/张。

掩盖延迟的幻觉设计

绝对不显示进度条。
(进度条暗示可预测的时长，
但 AI 耗时无法预测)



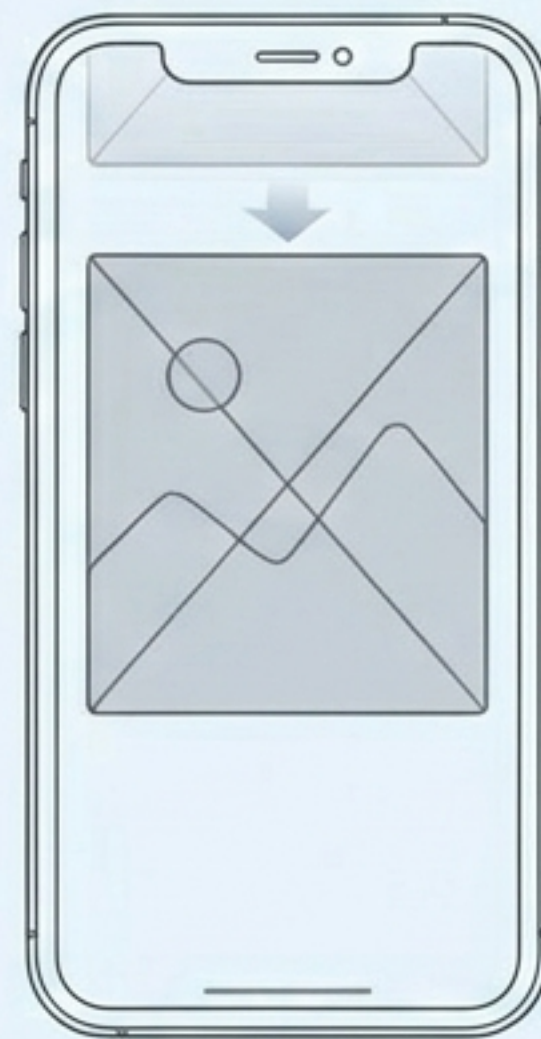
State 1

灵感卡预览轮播 + 文案：“你的光，刚刚好”



State 2

SSE 触发 photo_completed 事件

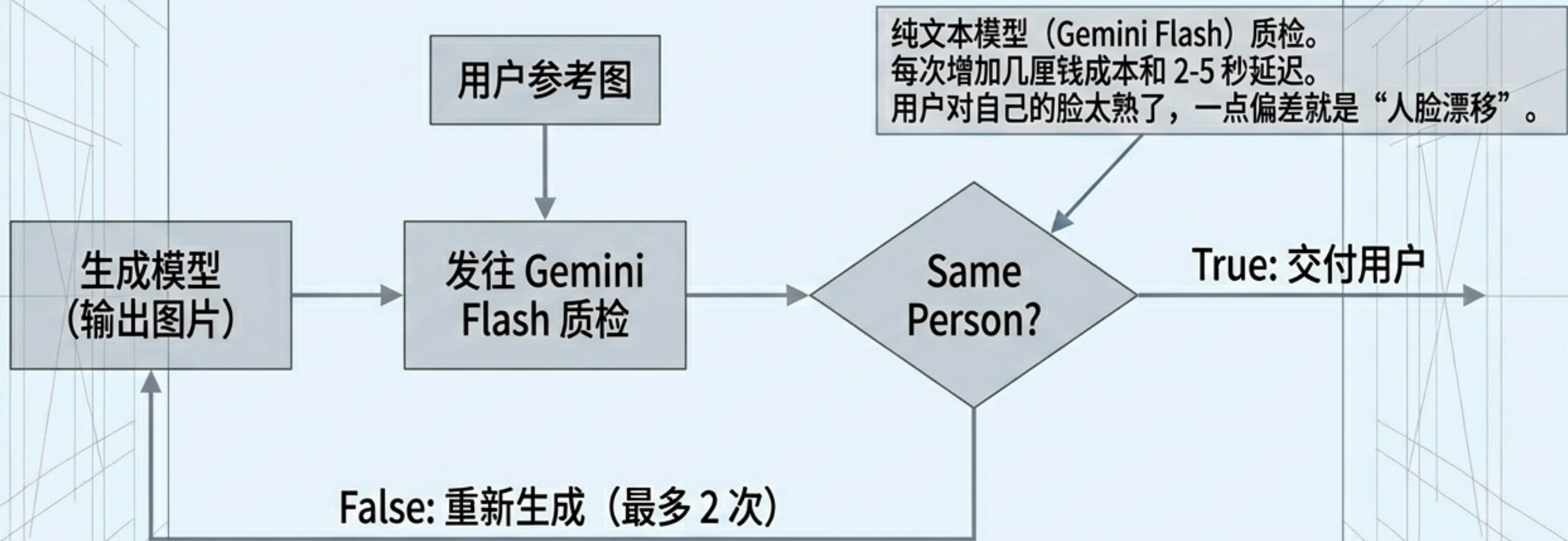


State 3

单张入场动画

营造的心理预期是照片在“到达”，而不是系统在“处理”。

自动质检：用 3 分钱的门卫，拦截 80% 的崩盘



拦截率：15-20% (首次生成)
// 挽回率：约 80% (重试后成功)。

如果两次重试均失败，交付最终结果。
（“给用户看 80% 像的照片，总比报错强。”）

注入灵魂：配文系统的三种面孔

	小红书	朋友圈
凡尔赛 (不经意炫耀)	200字 / 3-5标签 2 Emoji	80字 / 无标签 1 Emoji
文艺 (极短的诗)	150字 / 3-5标签 1 Emoji	60字 / 无标签 1 Emoji
简约 (极致留白)	50字 / 3-5标签 0 Emoji	30字 / 无标签 0 Emoji

由独立的 Gemini 3 Flash 处理，输出 JSON 候选数组，耗时 2-4 秒。

配文系统的“铁律”

**禁止提及任何品牌名、
酒店名、餐厅名、景点名**

这是凡尔赛公式的延伸。如果文案里写了“住了四季酒店”，那种不经意的感觉就全毁了。

文案必须足够泛化。让奢华只在照片里可见。

成本冰山：一次出图到底烧多少钱？

输出图片 Token

单价是文字的 120 倍。
(2K 分辨率成本是 1K 的 1.5 倍)。

4%

输入文本 Token

输入参考图

人脸质检 (Flash)

配文生成

全部加起来只是零头。

一次标准会话 (4张图+文案+质检重试)
总成本控制在 1 元以内。

“额度系统的诞生不是因为商业策略，
是因为不想测试阶段就破产。”

坦白局：无法逃避的工程取舍

Prompt 长度 vs. 人脸还原度



场景描述越丰富，模型分配给参考图的注意力就越少。构图极美时，人脸可能只有 80% 像。

跨图一致性 vs. API 独立生成限制



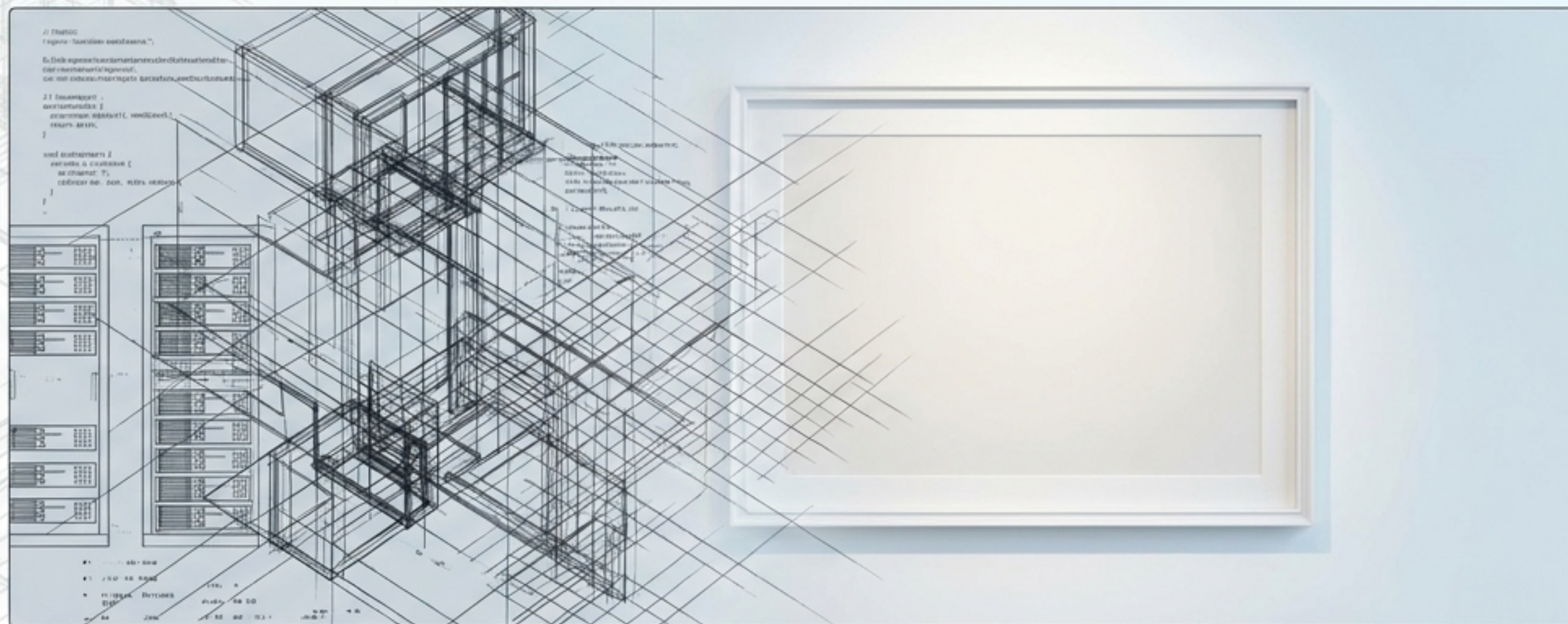
4 张图独立生成，微小的肤色/眼距差异必然存在。真正的人脸锁定需要微调全新模型架构。

存储成本 vs. 用户体验



图片与会话元数据存放在 Vercel Blob / Redis, 2 小时强制过期。这是成本控制下被吐槽最多的妥协。

极度克制，方能毫不费力。



ÉLAN 跟 AI 沟通的方式，不是抖机灵的单个 Prompt。
而是整套分层的指令系统、空间架构的权重博弈、流式推送的错觉设计，以及严苛的质检循环。

高级感，是算出来的。

// TO BE CONTINUED // 下一篇：18 张灵感卡背后的数据结构。