

The Sonic Blueprint Light Mode Aesthetic.

唯爱AI音乐学苑第二讲：结构与情绪

副歌、冲突、20秒里站住（兼谈 Jev；特邀 Fran）

2026年9月26日 | 主讲人：Michael Huo | 嘉宾：Fran

ZJUAANC
NACUAA AI Entrepreneurs Club
VI AI Foundation

走向最终混音的路径

1. 9月12日:
从想法到可改的一首

2. 9月26日:
结构与情绪:
20秒里站住

3. 10月10日:
中文声线与华乐音色

4. 10月24日:
编曲层次与画面

5. 11月7日:
定稿

L1 问题: 这首草稿能改吗?

L2 问题: 陌生人会停留20秒吗?

我们践行人类把关的有限自我提升, 而非无约束的AI自动生成。

一首歌是一条路径，而非华丽句子的堆砌

主歌：入口

交代人物或场景，
不要直接说教。

load lines

桥段：新视角

简短AI草稿中可选。

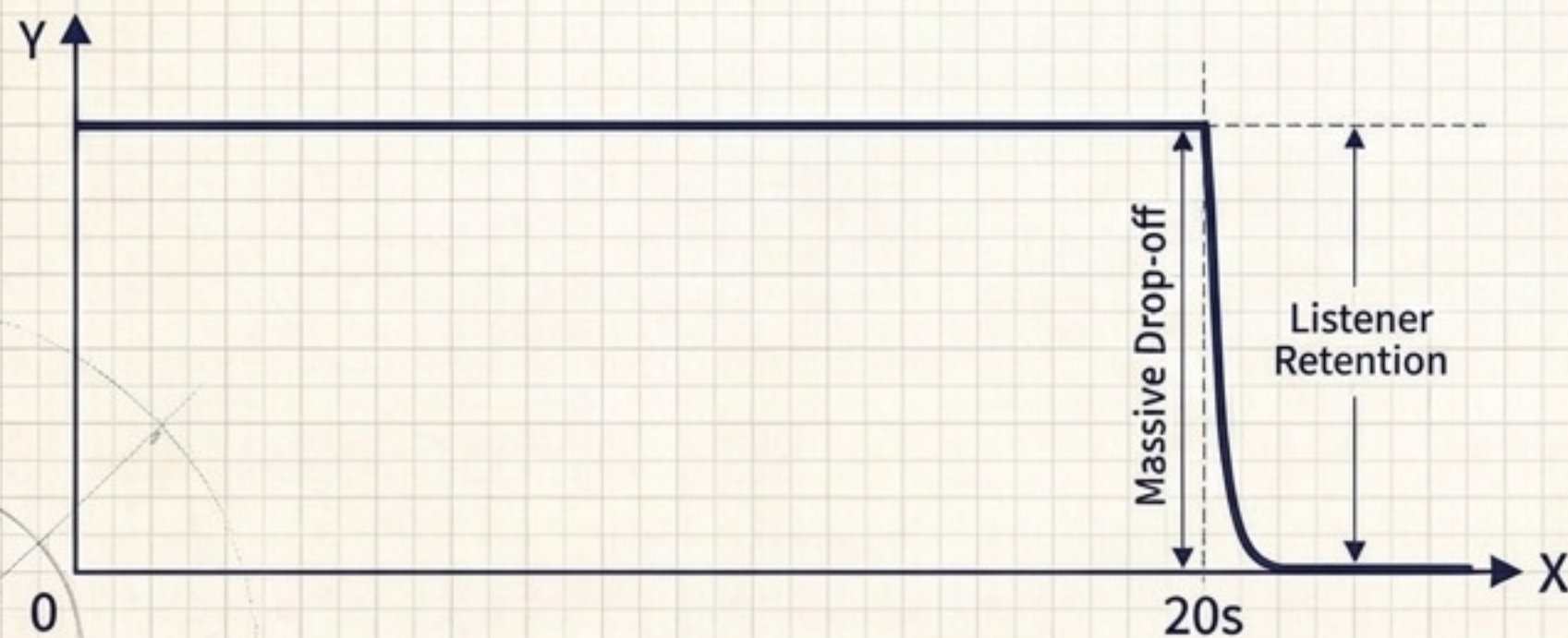
预副歌：倾斜

重心前倾，张力累积。

副歌：承诺

你能哼唱的承重句。

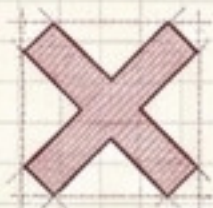
20秒生存法则



20秒 ≈ 手机端的第一印象

- 如果前20秒没站住，后期的精良制作也救不回来。
- 情绪必须有时间轴。一成不变的情绪只是背景墙。

重新定义冲突



误区：冲突不是悲伤的形容词

夜色 远方
温柔
Cinematic
Emotional

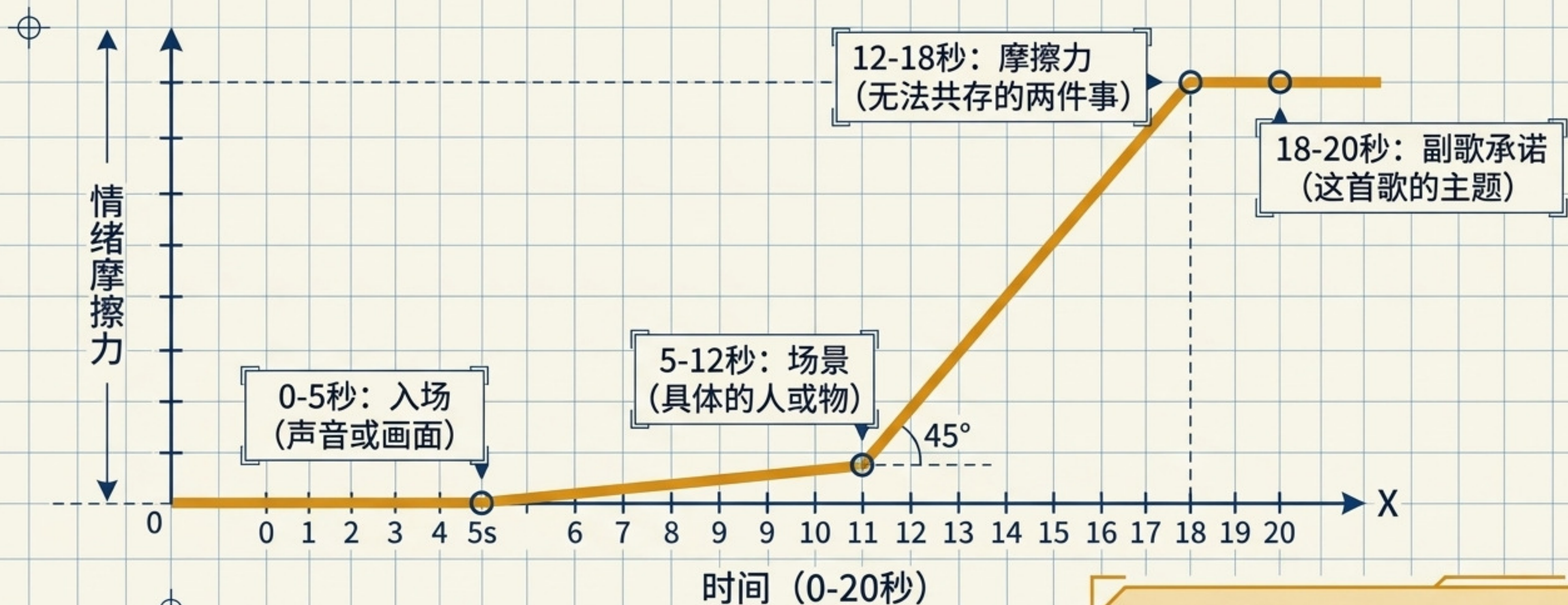


蓝图：两股无法同时胜出的力量



记与忘 | 默与言

短曲目的黄金冲突曲线



核心法则：推迟口号；把前12秒花在一个具象物体上。

诊断：站得住的副歌

✗ 没站住 (Fail)

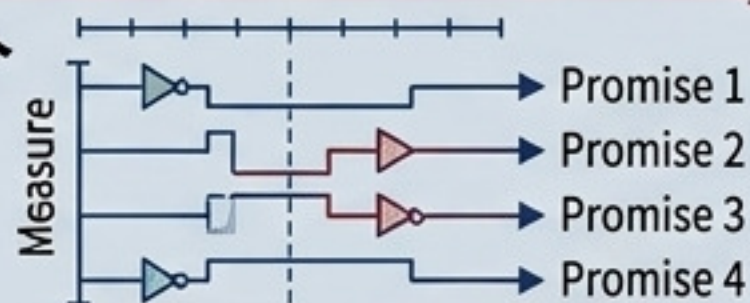
1. 只是鼓声更大的第二段主歌



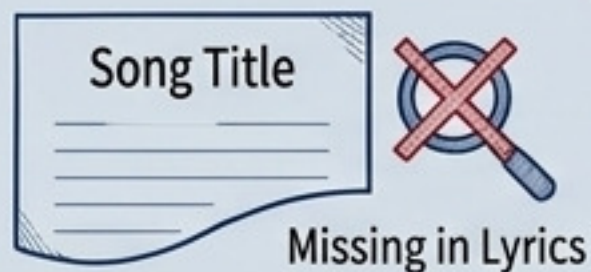
2. 记忆点是情绪形容词 (如：温柔)



3. 八个小节里塞了四个不同的承诺

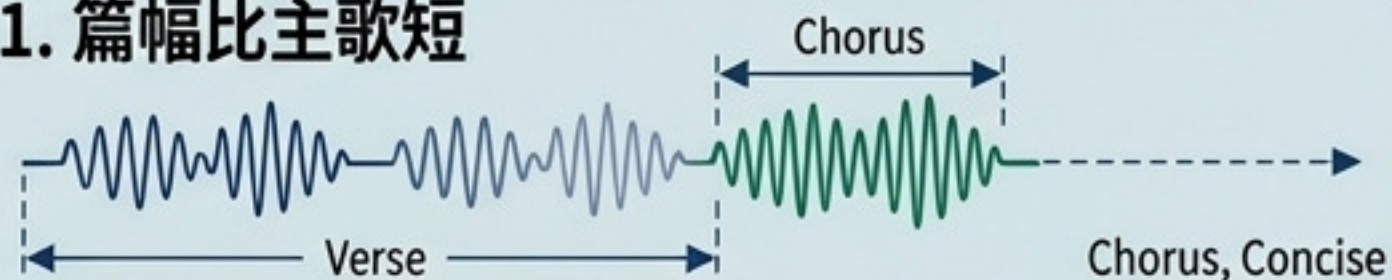


4. 歌名从未在演唱中出现

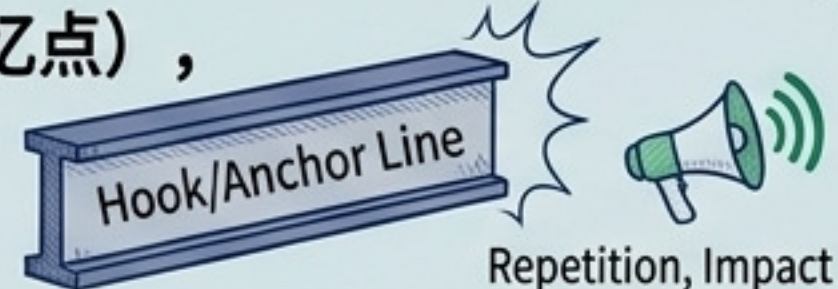


✓ 站得住 (Pass)

1. 篇幅比主歌短



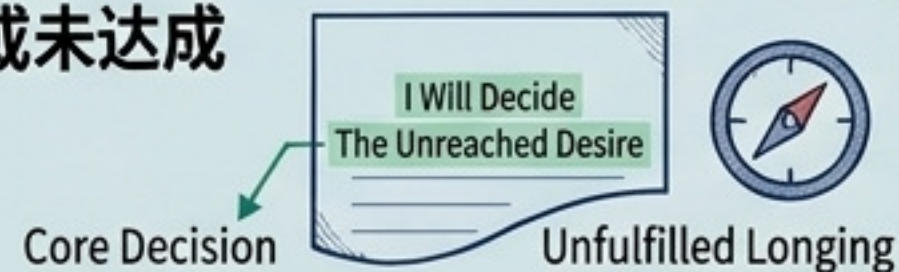
2. 重复承重句 (记忆点), 而非写新小作文



3. 回应主歌的张力, 而非转移话题



4. 表达一个决定或未达成 未达成的渴望

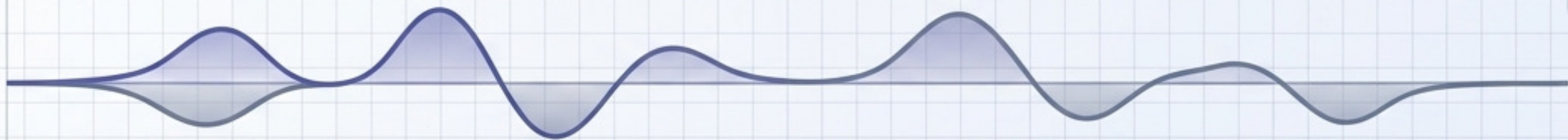


现场盲听测试

不要问听众喜不喜欢。只问：站住 / 没站住。

[站住]

[没站住]



✕ 音轨 A (没站住)

四句纯写景，没有做决定。

✓ 音轨 B (站住)

一个物件 + 一处摩擦 + 一句可重复的歌词。

20秒生存检查清单

- [] 入口：第一句有具体声音或画面
- [] 场景：听众能把它画下来
- [] 冲突：两股力量，而非两个形容词
- [] 承诺：一句可以作为歌名的话
- [] 重复：记忆点在完整副歌中至少出现两次
- [] 反向检查：没有将前20秒浪费在风格词上

绘制蓝图：如何给大模型下指令

核心原则：AI 是结构改写工具，不是代笔枪手。

红线禁忌

**让它更感人 / 更古风
/ 更有爆款相。**

转而使用结构提示词

针对特定节拍的、严格的、有时间限制的指令。

提示词 1：结构诊断

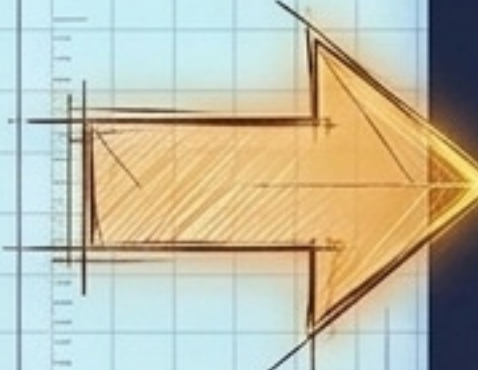
先不要重写。

标出入口、场景、冲突、副歌承诺。

指出缺失了哪个节拍。

限20字内。

I 提示词 2：提升冲突



保持原歌名和相同的行数。
将‘决定’移入副歌。
在第一行放入一个具象物件。
删除情绪形容词。

提示词 3：锁定副歌

写出三个版本的4行副歌。
每个版本必须重复一句完全相同的记忆点。
不要改动主歌。
标注为A、B、C。

A



B



C



现实检验

特邀嘉宾：Fran

1. 第一次跑出好结果后的真实感受
2. 事后才看清的结构性错误
3. 对本周仍在修改首批作品的建议

AI 协作循环

人类要求：
写首爆款

模型写歌

完成
(错误循环)

1. 人类设定标题与冲突

2. 模型只写主歌或副歌

5. 人类锁定核心句，
再进入生成

可修改的草稿 > 看起来漂亮的草稿

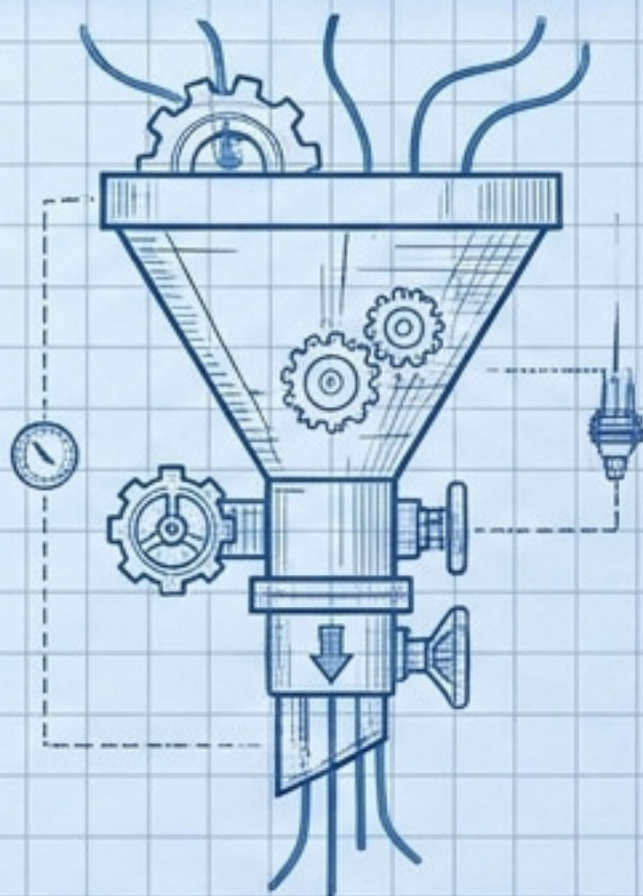
4. 模型只改一个节拍

3. 人类试听20秒

Jev 边界：系统一决策模型

输入：歌词状态 + 类型化问题

Jev 决策引擎



判断：是 / 否

打分：0.81（置信度）

选项：站住 / 没站住

定义：快速评判的决策模型，不是聊天机器人。

能力边界矩阵

厂商宣称

端到端 70-500 毫秒。
比大模型做裁判更快更便宜。

输出限制

它不写歌词、不给理由、
不写代码注释。

架构限制

无法生成设定外的文本（不会
自己发明副歌）。

现实限制

仍可能选错标签。0.81的『包
含记忆点』高分不等于爆款。

生成与决策的分离

把关层 (Human)
聆听与最终决策。

决策层 (Jev)
快速打分：有无记忆点 / 冲突是否清晰。

生成层 (LLM / Suno)
生成无限的候选方案。

这是生成与决策分离。
绝非『模型自我提升』。

Jev 课堂演示模型

输入问题

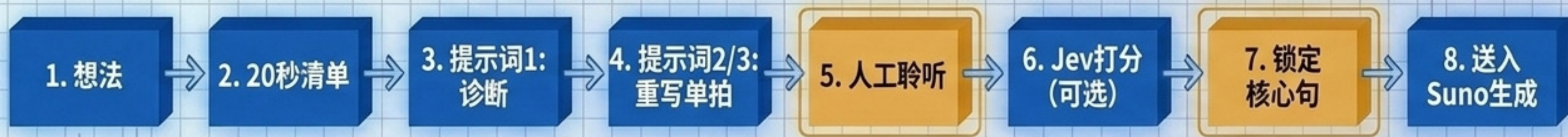
1. 选择题：站住 / 勉强 / 没站住
2. 评分题：冲突等级 1-5
3. 判断题：歌名是否在副歌中唱出？

输出结果

1. [勉强] (概率：0.62)
2. [等级 2] (概率：0.88)
3. [否] (概率：0.91)

绝对准则：如果 Jev 的跑分与现场听众的感受相左，听众永远是对的。

统一 workflow 蓝图



今晚的致命误区

[错误 01]: 评判个人喜好，而不是评判有没有站住。

[错误 02]: 向模型要新风格，而不是补齐缺失的结构节拍。

[错误 03]: 把 Fran 的获奖作品当成模板直接抄。

[错误 04]: 把 Jev 当成填词人。

[错误 05]: 把漂亮的模型跑分当成一首完整的歌。

蓝图已就绪

课后实践

拿出一首草稿
填完20秒清单
仅重写前20秒
下次课带着两个版本来

下一步

第三讲（10月10日）：
中文声线与华乐音色