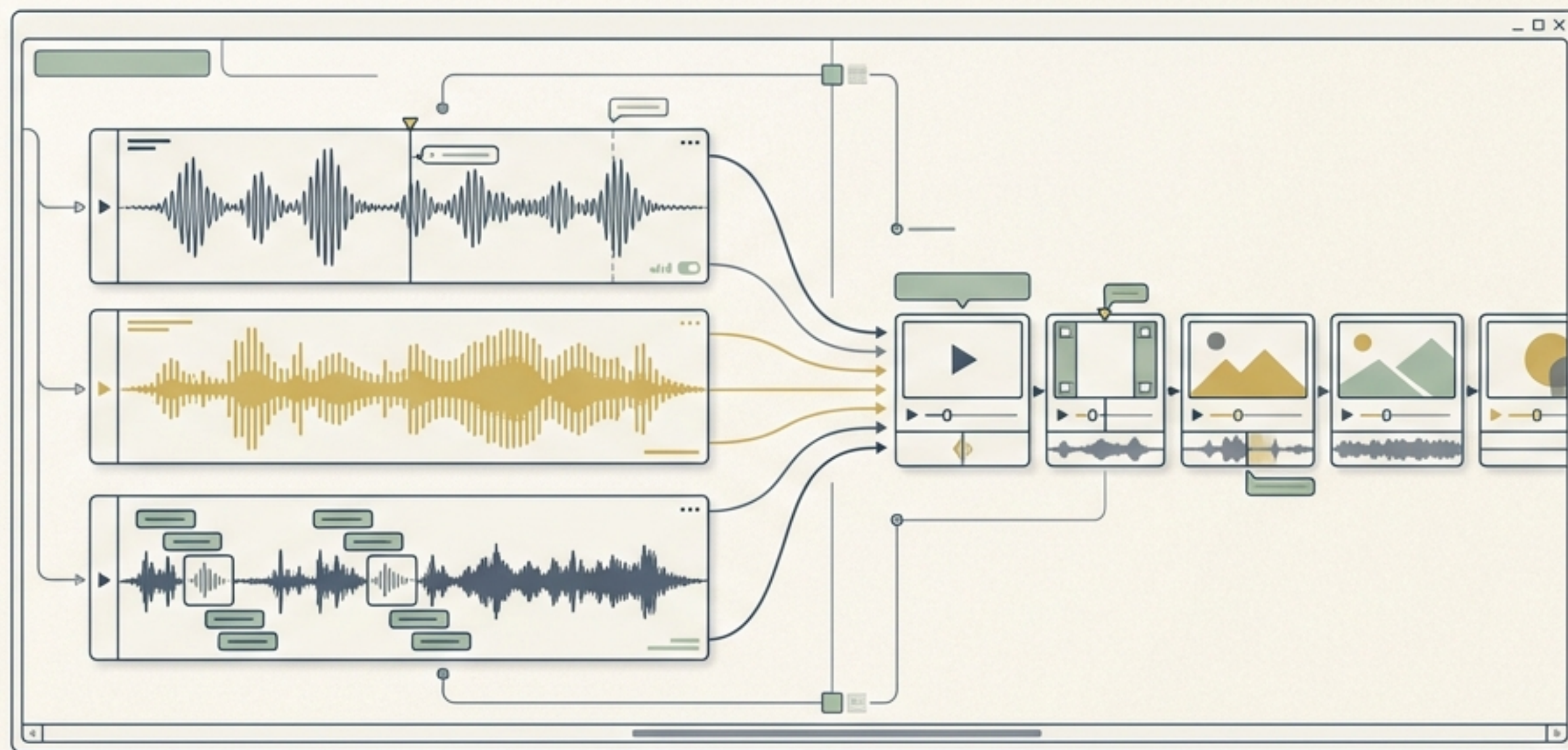


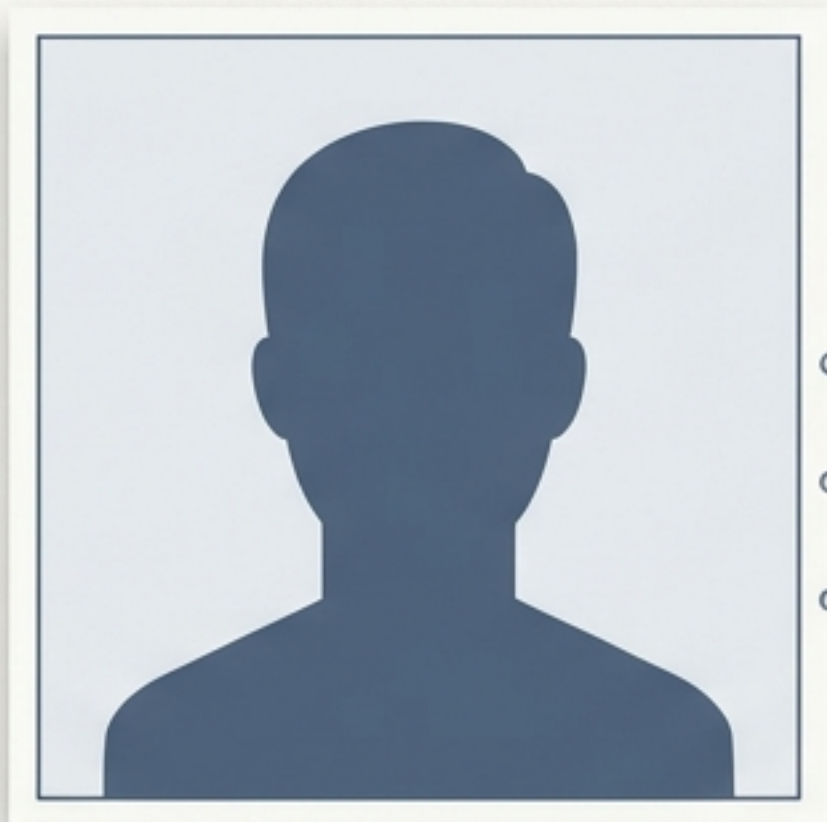
用AI和智能体高效制作高质量短视频与短剧

Part 3: 配音、音乐与后期制作



2026年7月11日

讲者档案



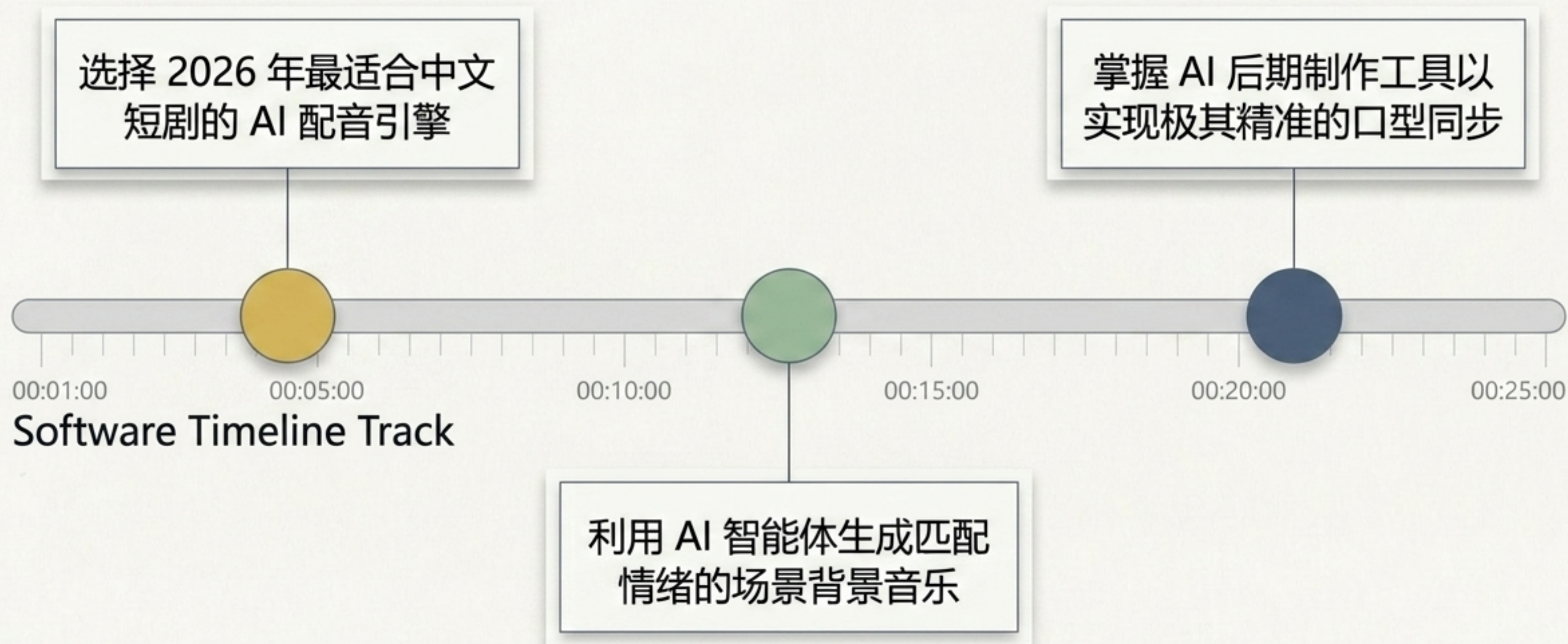
北美高校联盟 AI 创业俱乐部主席

北卡浙大校友会创会理事副会长

唯爱AI基金会及唯爱AI公益基金会主席

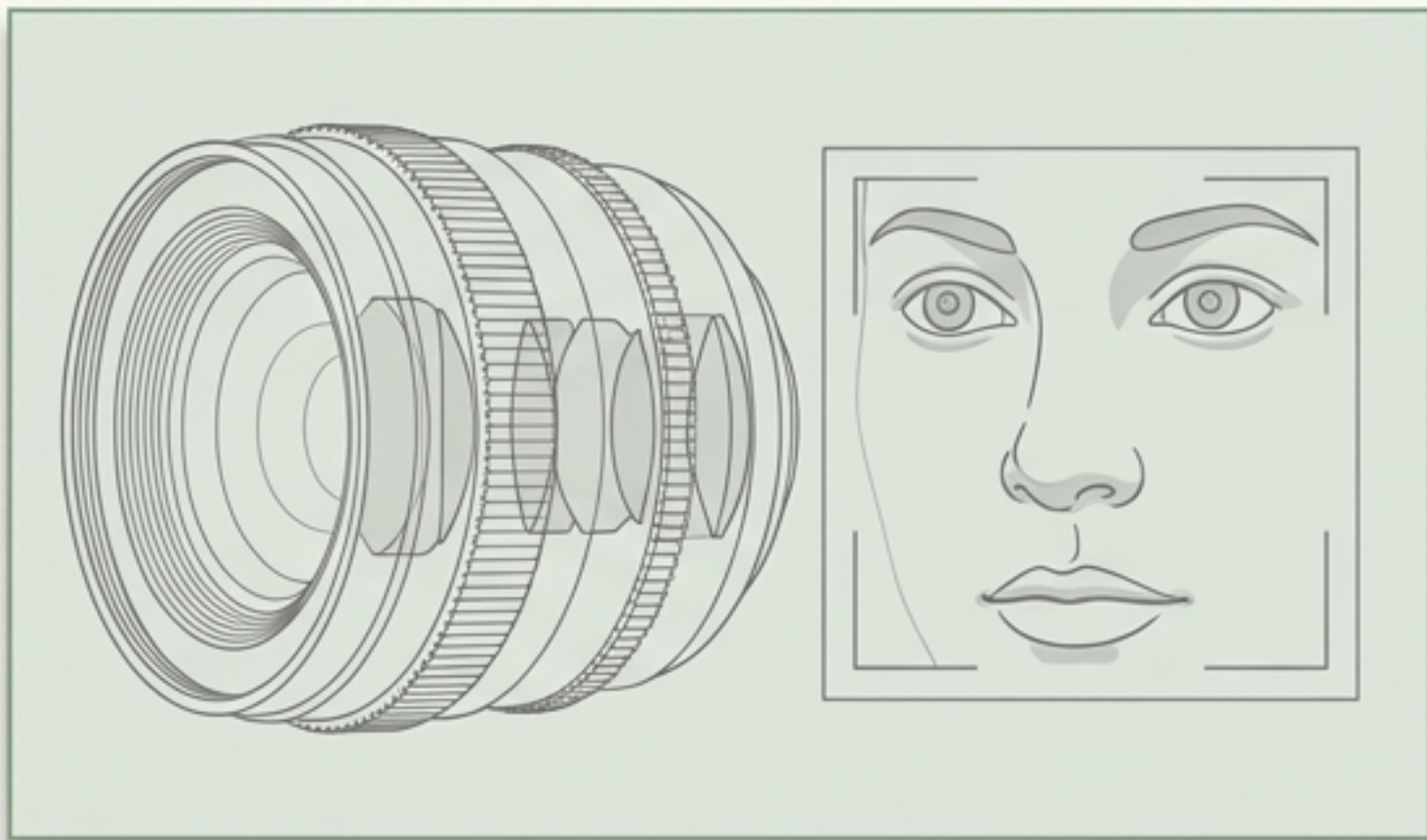
Michael HUO

核心目标与技术解锁



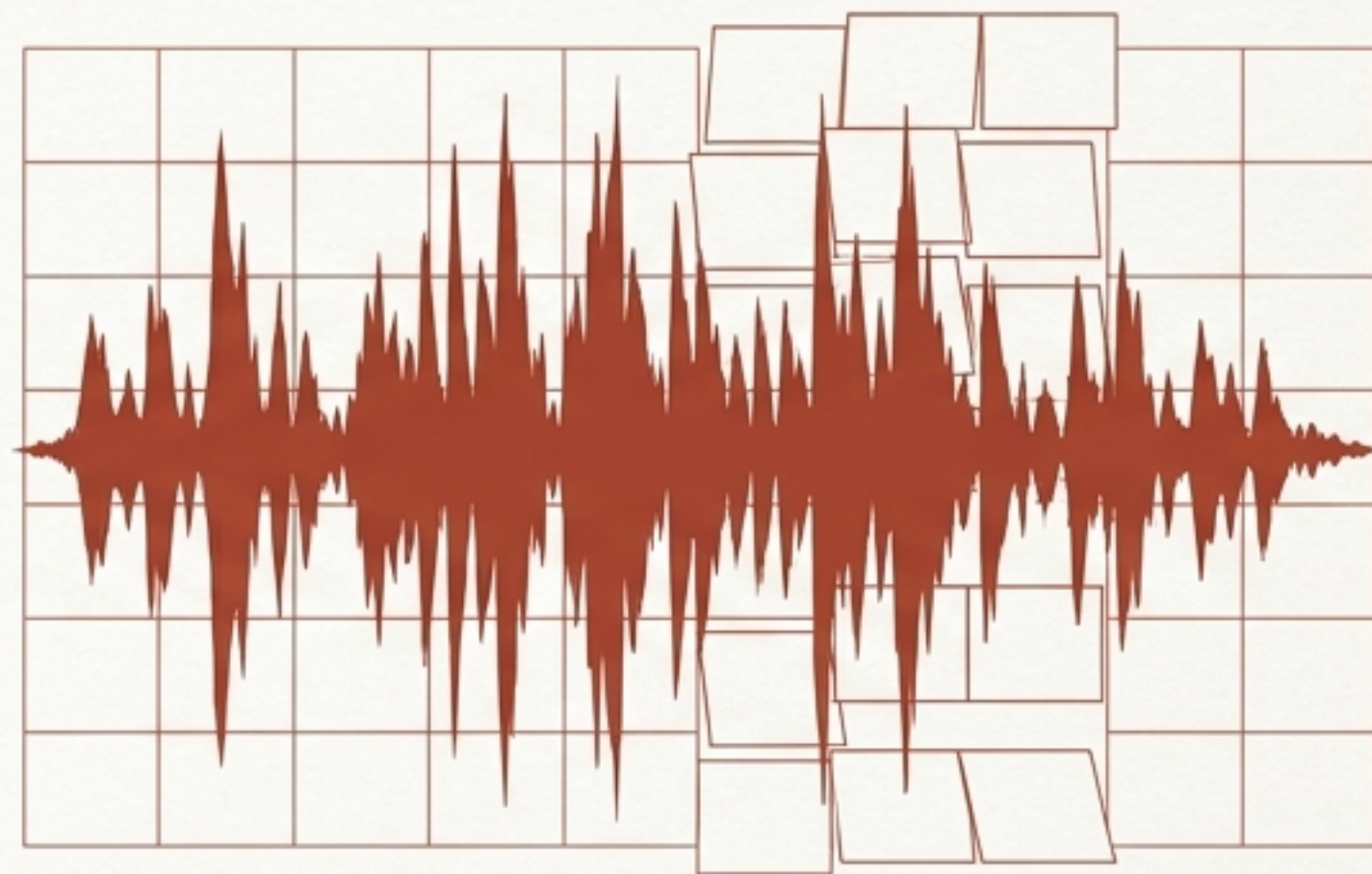
沉浸感断层：为什么音频与后期是成败关键

Optimal State



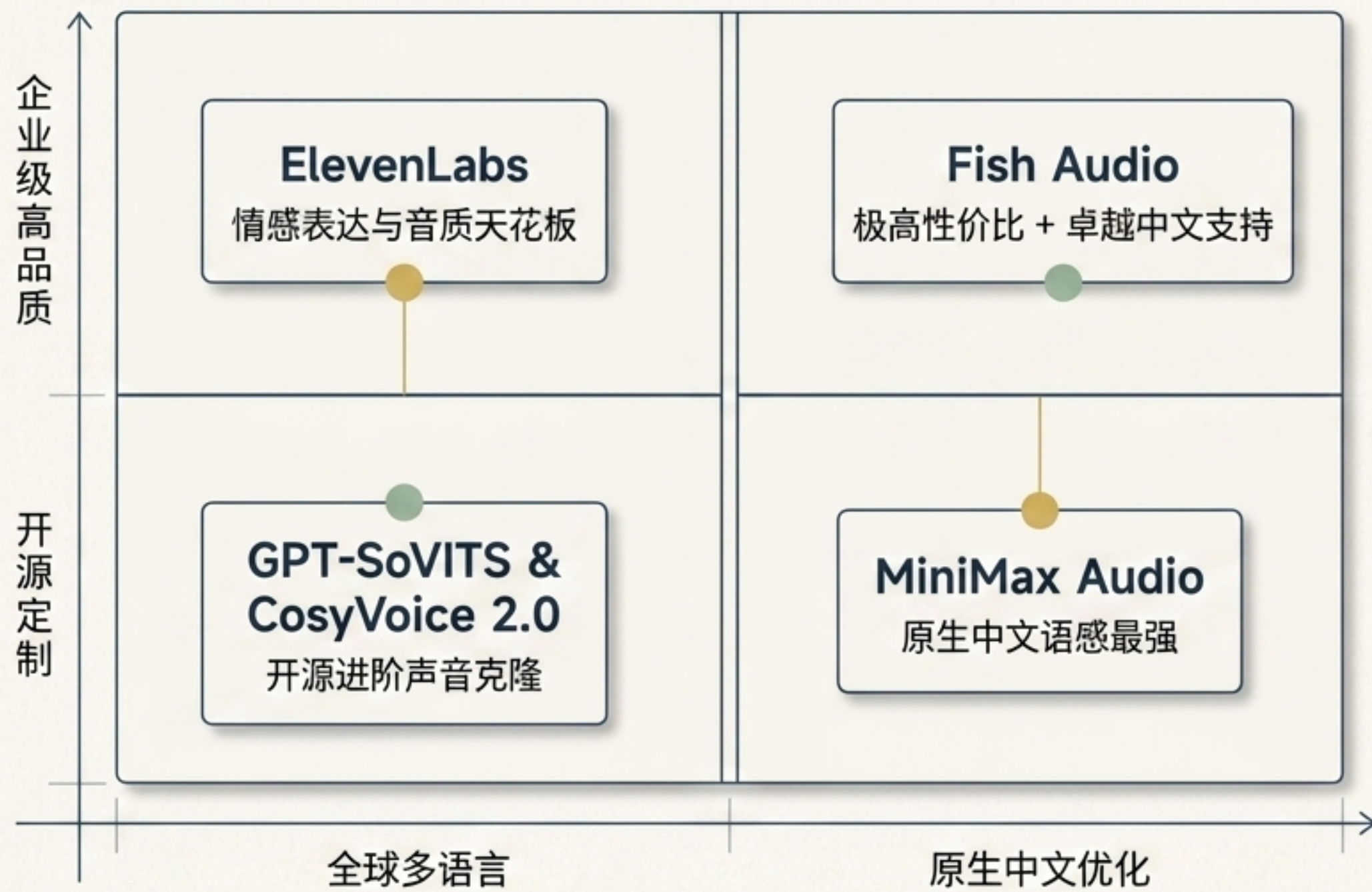
视觉吸引眼球，但音频维系沉浸。

Critical Failure



即使画面达到院线级标准，劣质的配音或微弱的口型不同步都会瞬间打破观众的沉浸感，导致完播率暴跌。

2026 AI 配音生态引擎概览



配音引擎决策矩阵

配音名称	高束密度诊断对比	场景应用
ElevenLabs	情感表达 (Emotion) Max Emotion 中文支持 (Chinese) Medium Chinese 成本 (Cost) High Cost	Use Case 场景应用：高情感张力戏份、爆发戏
Fish Audio	情感表达 (Emotion) Medium Emotion 中文支持 (Chinese) High Chinese 成本 (Cost) Low Cost	Use Case 场景应用：高性价比常规中文配音、大体量流水线
MiniMax Audio	情感表达 (Emotion) High Emotion 中文支持 (Chinese) Max Chinese 成本 (Cost) Medium Cost	Use Case 场景应用：追求自然原生中文语感与本土化表现
GPT-SoVITS / CosyVoice	自定义调优与开源灵活性	Use Case 场景应用：自定义专属角色声音克隆与私有化部署

声音克隆实战管线：多集一致性控制

准备纯净样本



确保无背景底噪、无混响干声。

特征提取与训练



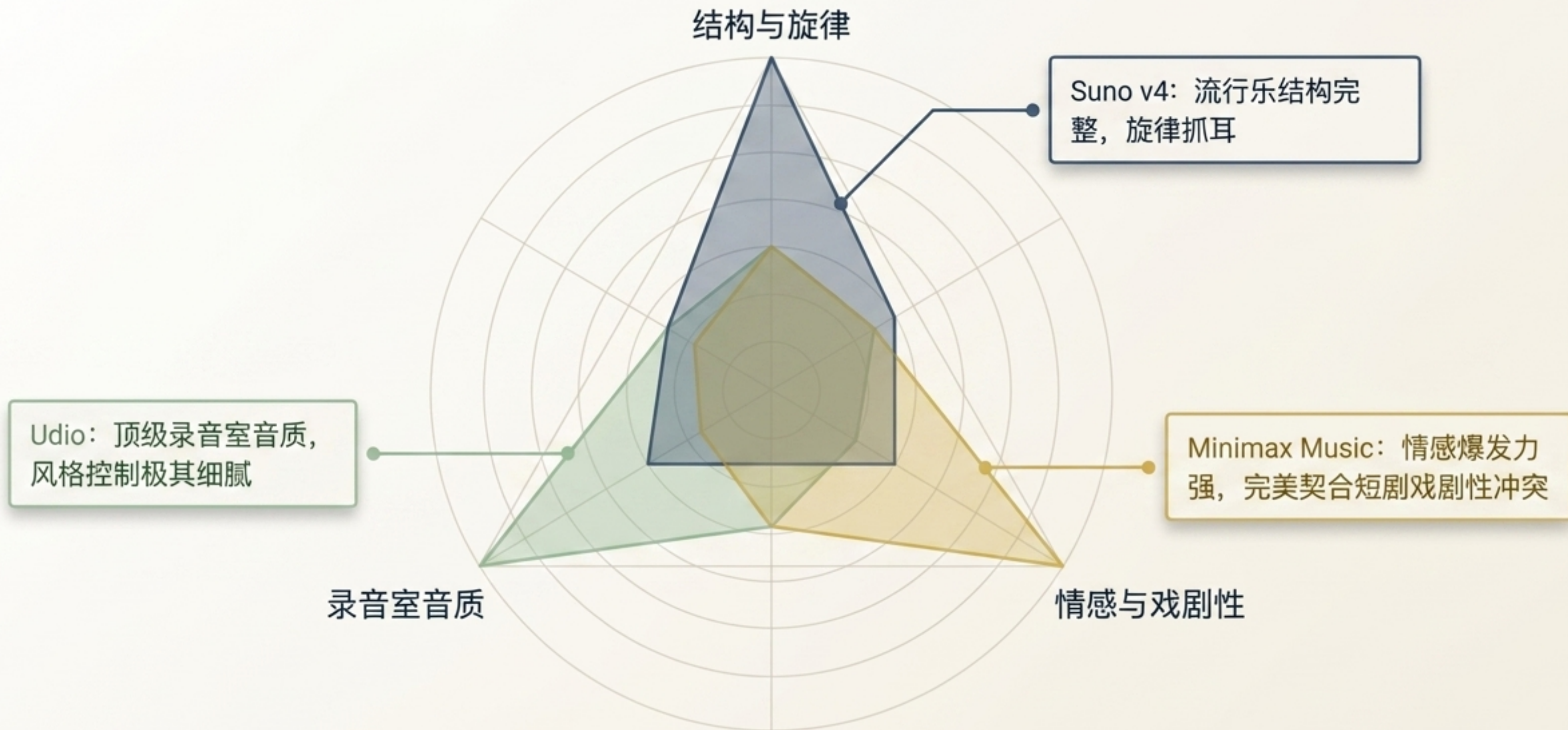
使用 GPT-SoVITS/CosyVoice 锁定音色特征。

剧集一致性校验



建立角色声音预设库，确保第1集与第100集音色绝对统一。

2026 AI 音乐生成雷达



情绪到音频的自动化映射



2026 核心后期技术栈

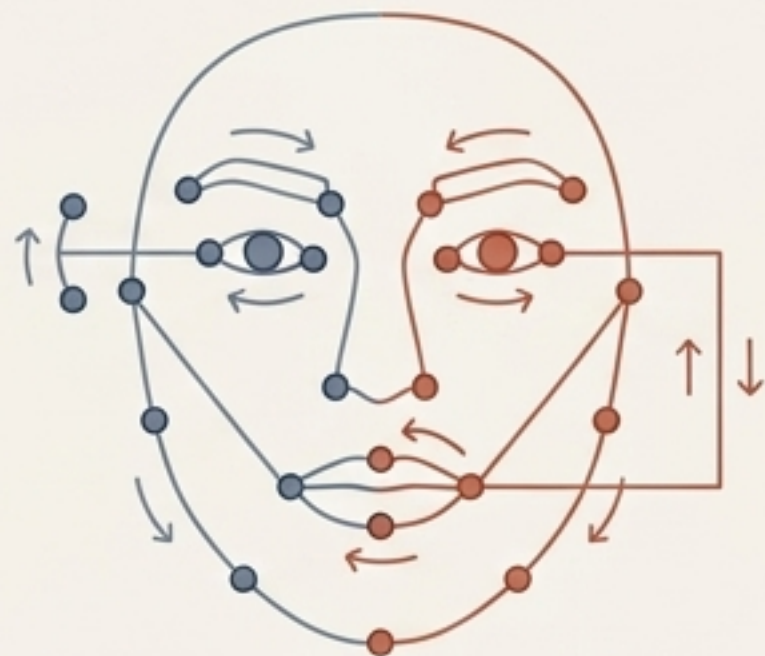
剪映专业版 **综合效能中枢**

✓ AI 智能配音 ✓ 自动字幕美化 ✓ 画质一键提升

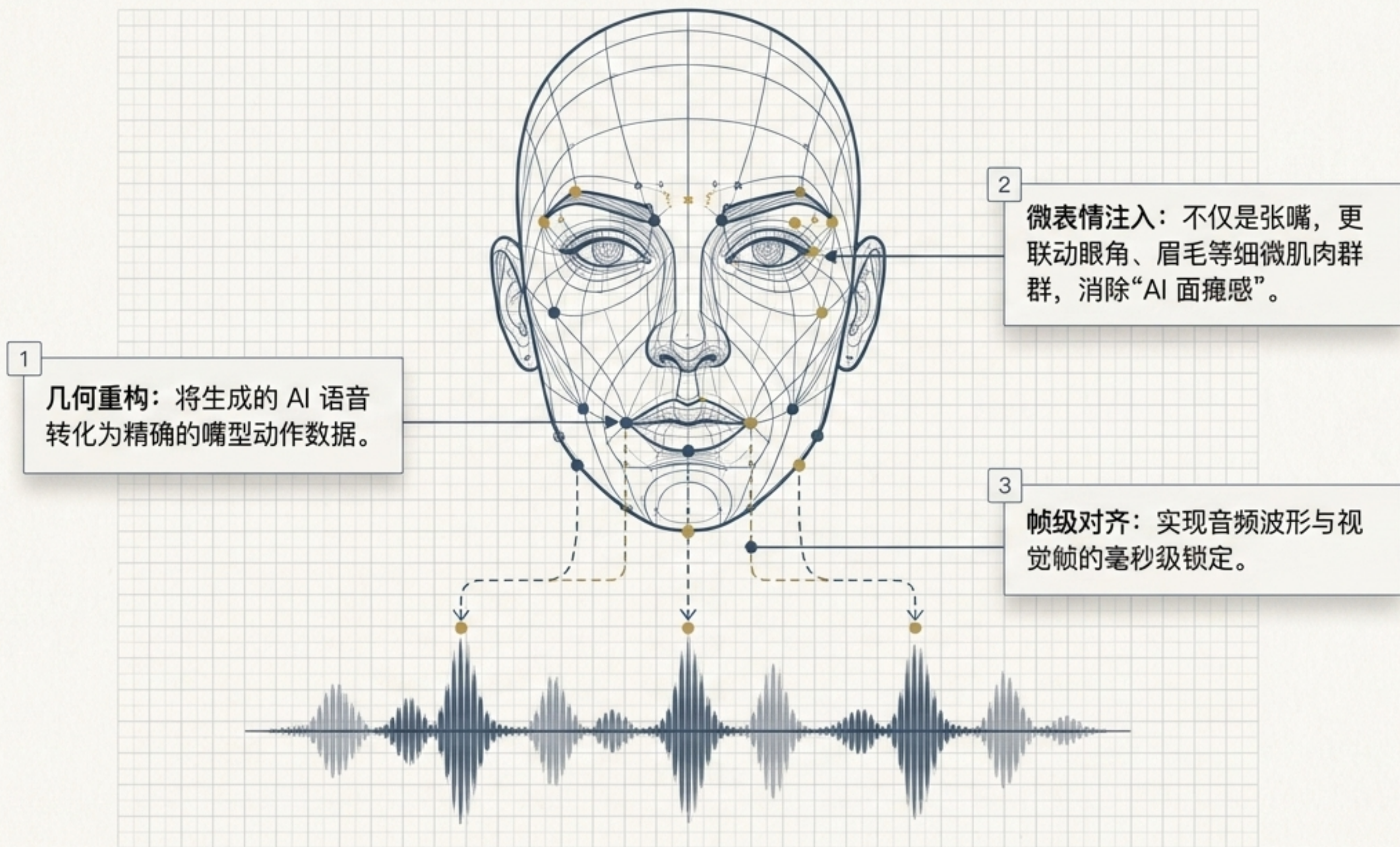


LivePortrait **极致细节引擎**

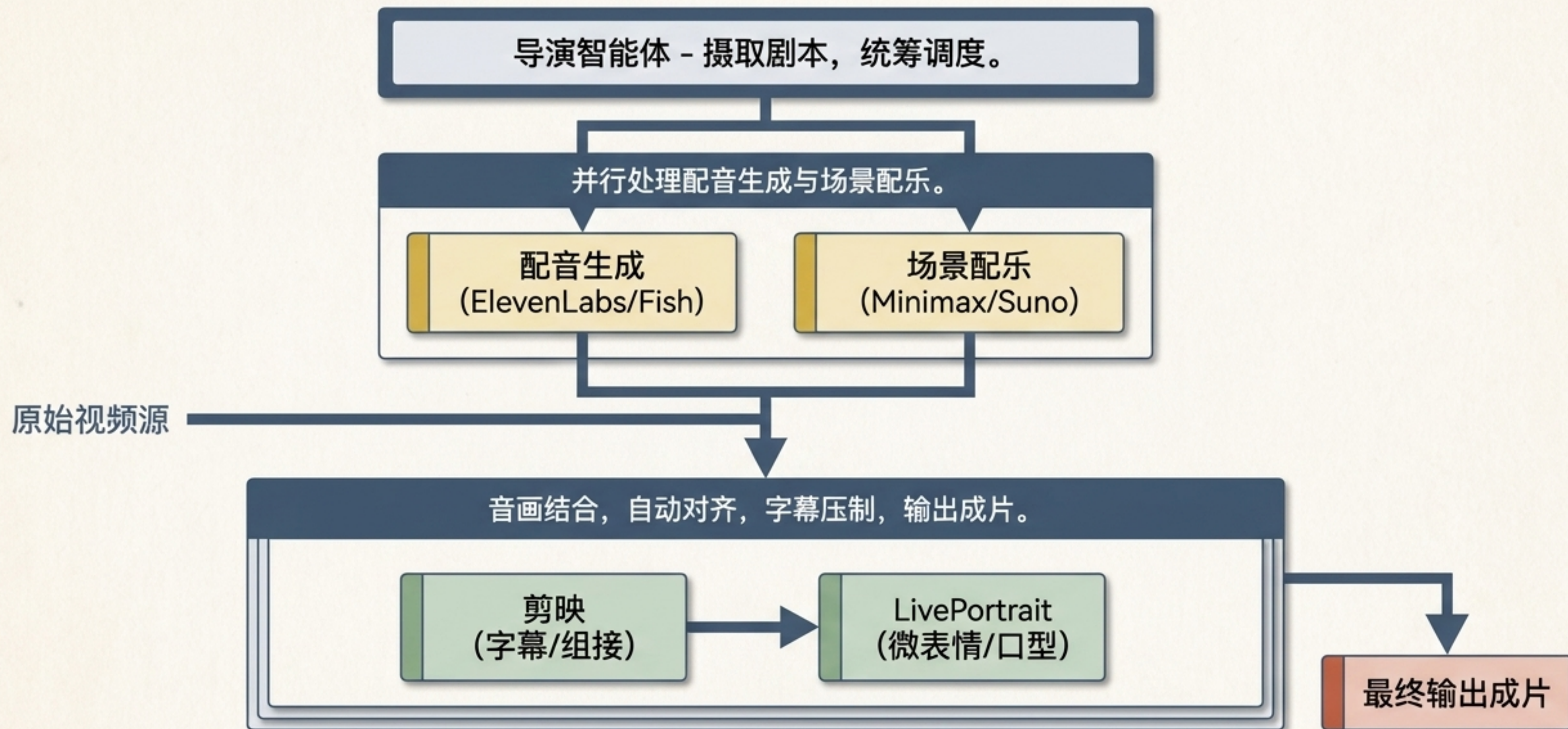
🎯 电影级微表情控制 🎯 像素级精准口型同步



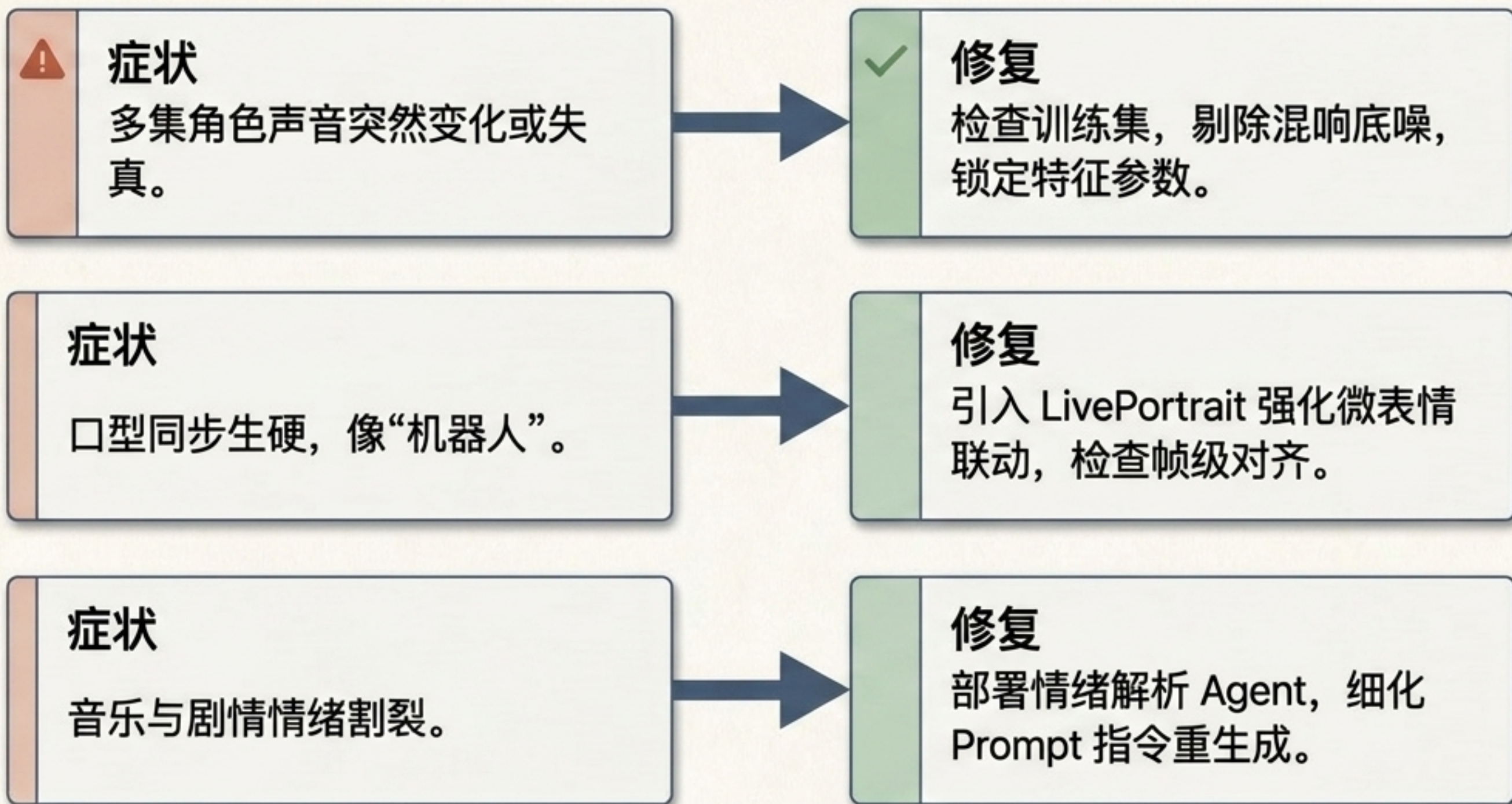
深度解构：LivePortrait 微表情与口型同步



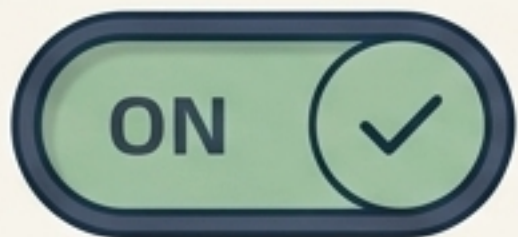
终极形态：AI 智能体全自动化后期管线



常见错误排查与诊断面板



2026 高质量短剧后期黄金法则



情境决定工具

永远根据情绪浓度与语言需求动态选择配音与音乐引擎，切勿“一招鲜”。



智能体接管繁琐

使用 AI Agent 自动化处理音乐匹配与后期基础任务，将精力留给创意。



出品级质检

最终渲染前，必须进行口型同步与微表情的 100% 逐帧质检。

核心要点总结

“

音频质量与口型同步是维系观众沉浸感的绝对基石。

“

“精准工具 + 智能体编排”能带来指数级的效率飞跃。

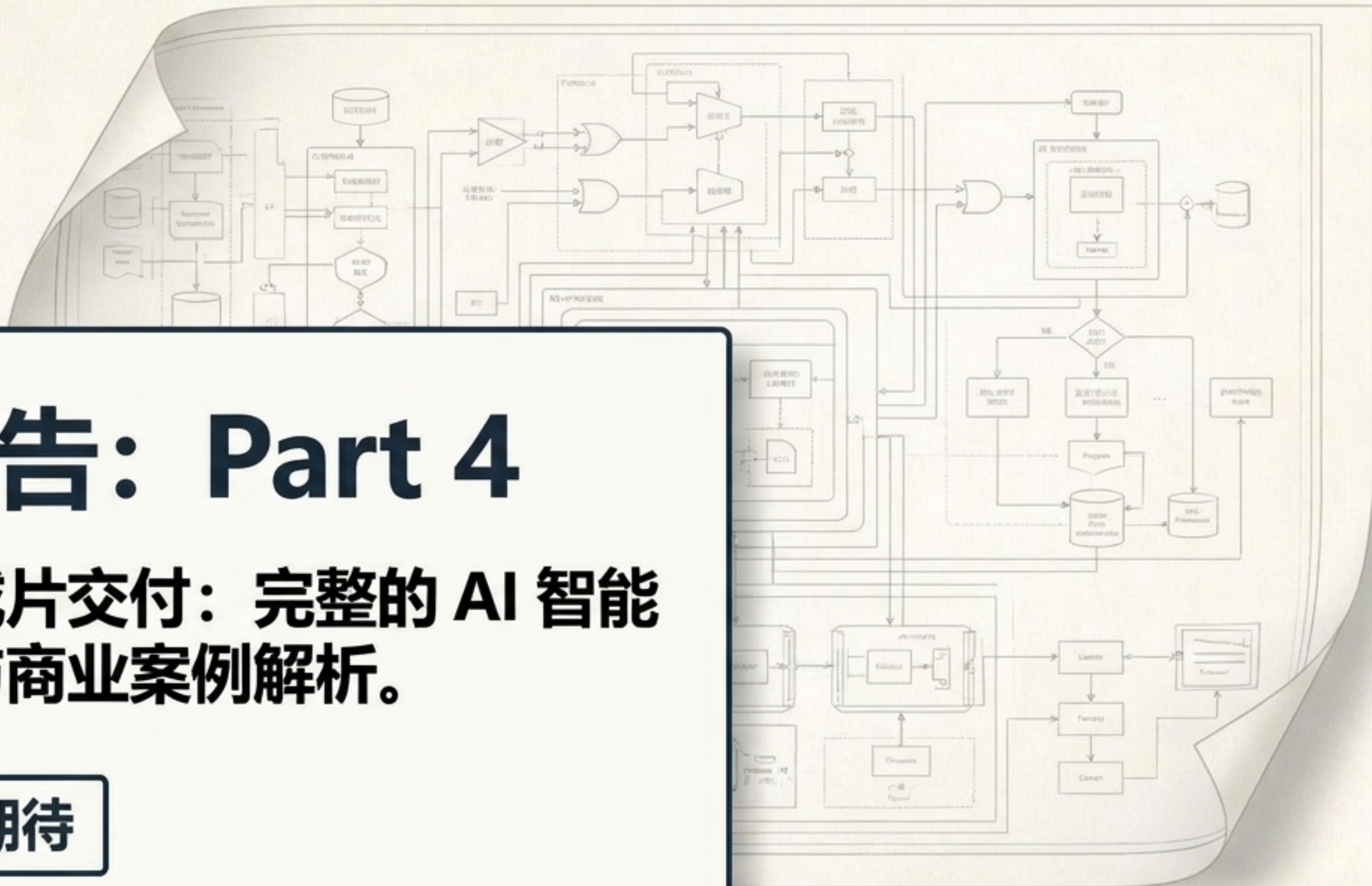
“

决定短剧最终上限的，往往是后期细节的打磨。

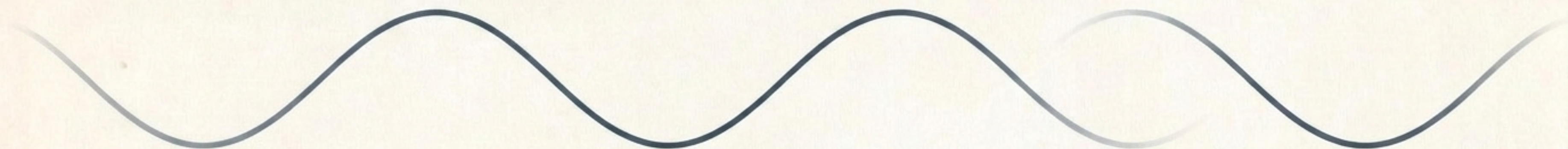
下一站预告：Part 4

从灵感火花到成片交付：完整的 AI 智能体实战工作流与商业案例解析。

系统就绪 - 敬请期待



感谢收看 Part 3

A decorative wavy line consisting of three connected sine waves, rendered in a dark gray color, spanning the width of the slide.

有问题欢迎提问。