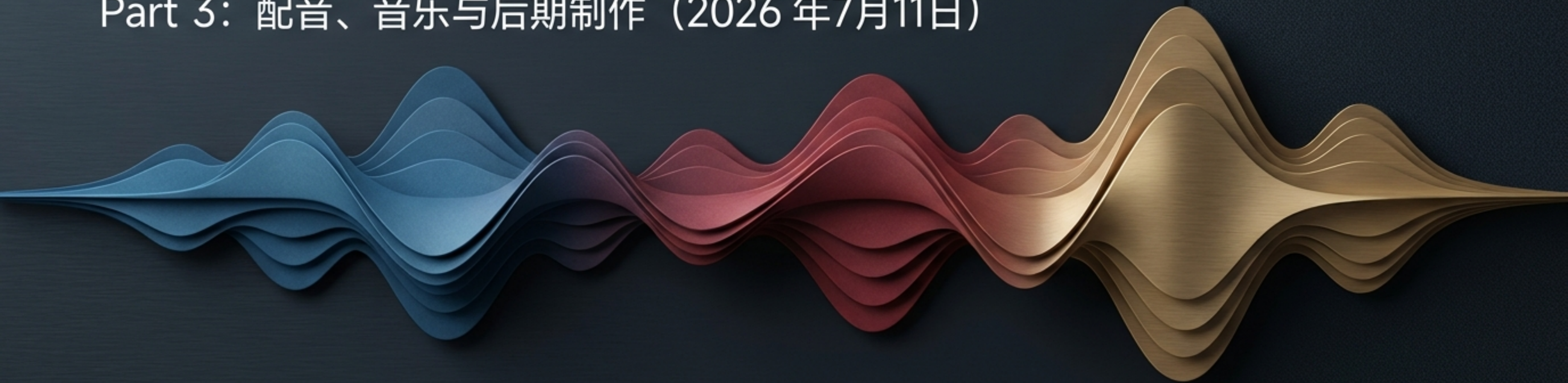


用AI和智能体高效制作 高质量短视频与短剧

Part 3: 配音、音乐与后期制作 (2026 年7月11日)



跨界融合：连接技术与数字叙事



Michael HUO

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

解构沉浸感的三个声音支柱

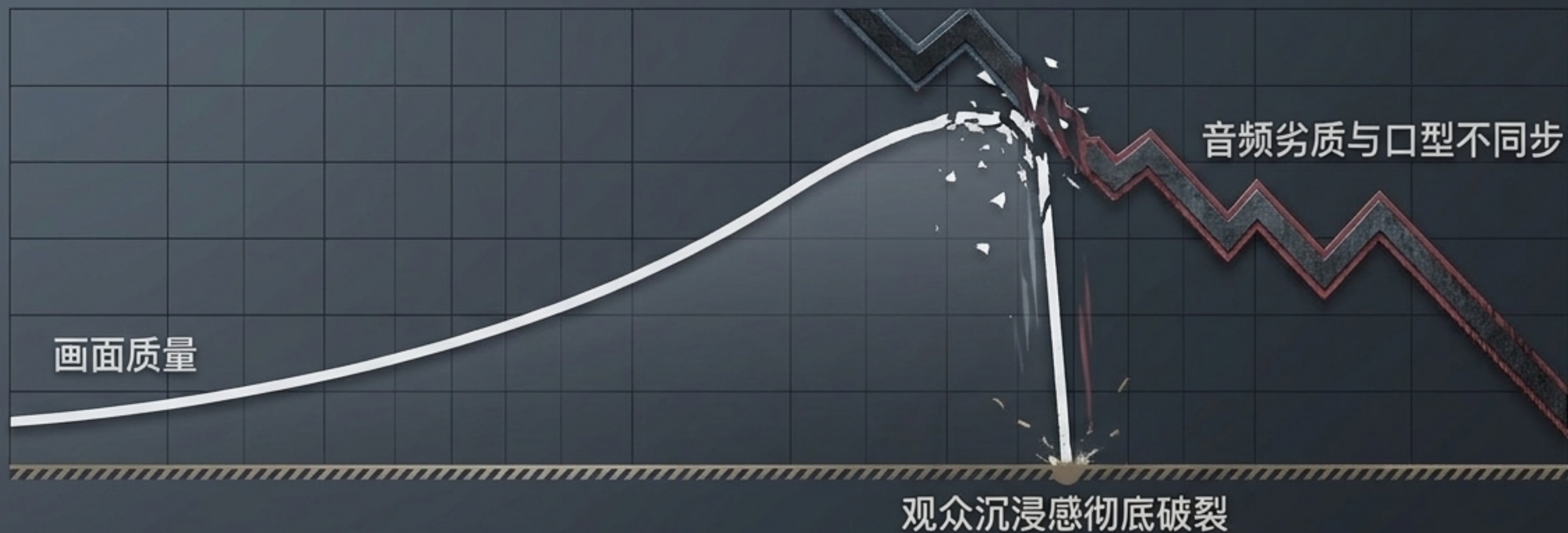
赋予角色灵魂：2026年最佳中文AI配音引擎的选择与应用。

注入情感共鸣：基于AI的动态背景音乐生成与情绪匹配。

缝合真实幻觉：后期自动化处理与微表情、精准口型同步。

AI Agent 自动协同

沉浸感悬崖：当视觉完美遭遇声音失步



即使画面达到院线级标准，配音品质的瑕疵与口型的轻微错位，都会在瞬间摧毁观众的沉浸感。声音，才是决定数字影视真实感的最终防线。

2026年数字影音的四大声音引擎



ElevenLabs

核心优势：情感表达与顶级音质的行业标杆。



Fish Audio

核心优势：极致性价比与卓越的原生中文支持。



MiniMax Audio

核心优势：强大的原生中文自然语境表现力。



GPT-SoVITS & CosyVoice 2.0

核心优势：开源生态下的进阶级自定义声音克隆。

AI配音选择矩阵：为特定场景寻找完美发声者



声音克隆的跨集一致性法则



样本提纯

准备零底噪、无混响的
纯净干声样本

锚点锚定

在 AI 引擎中建立角色的
基准声纹特征库

情绪微调

基于场景需求微调参数，
但不偏离核心声纹

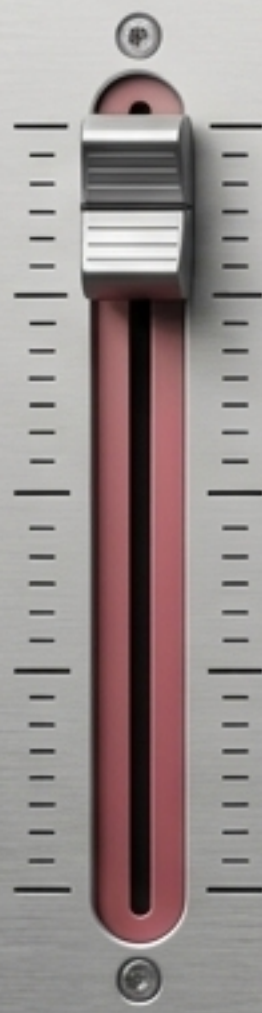
多集一致

确保在50集短剧内，角
色的声音辨识度始终如一

AI音乐引擎均衡器：精准控制戏剧张力

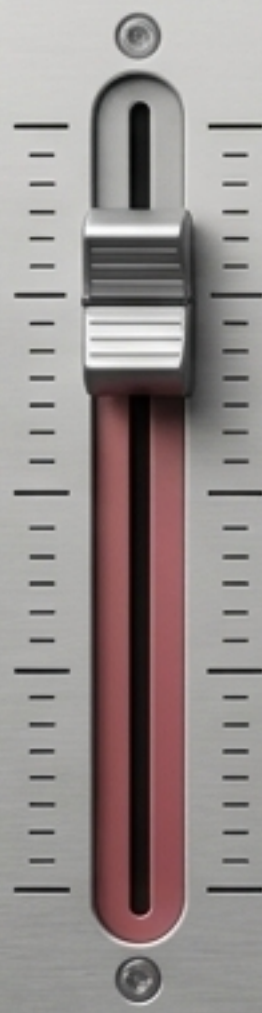
Suno v4

叙事结构
与旋律



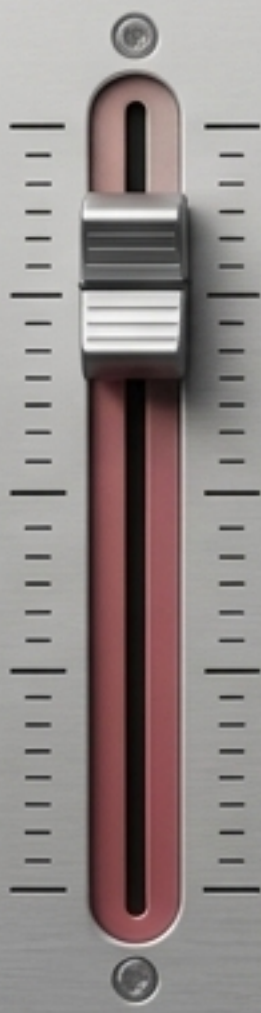
95%

戏剧张力



80%

音质精度



85%

Minimax Music

叙事结构
与旋律



80%

戏剧张力



98%

音质精度



80%

最强情感共鸣

Udio

叙事结构
与旋律



85%

戏剧张力



85%

音质精度与
风格控制



95%

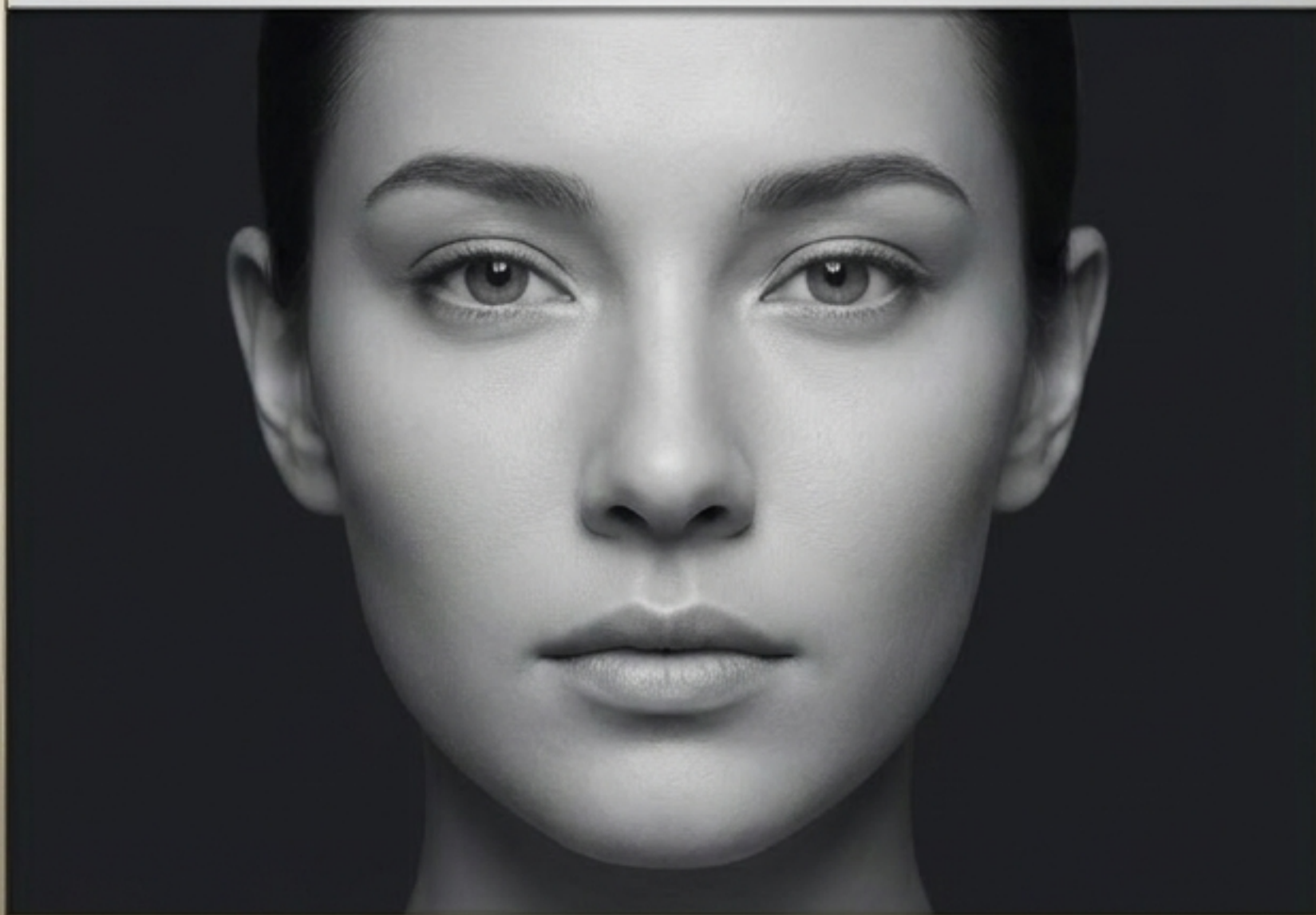
自动化情绪映射：让AI成为你的配乐指导



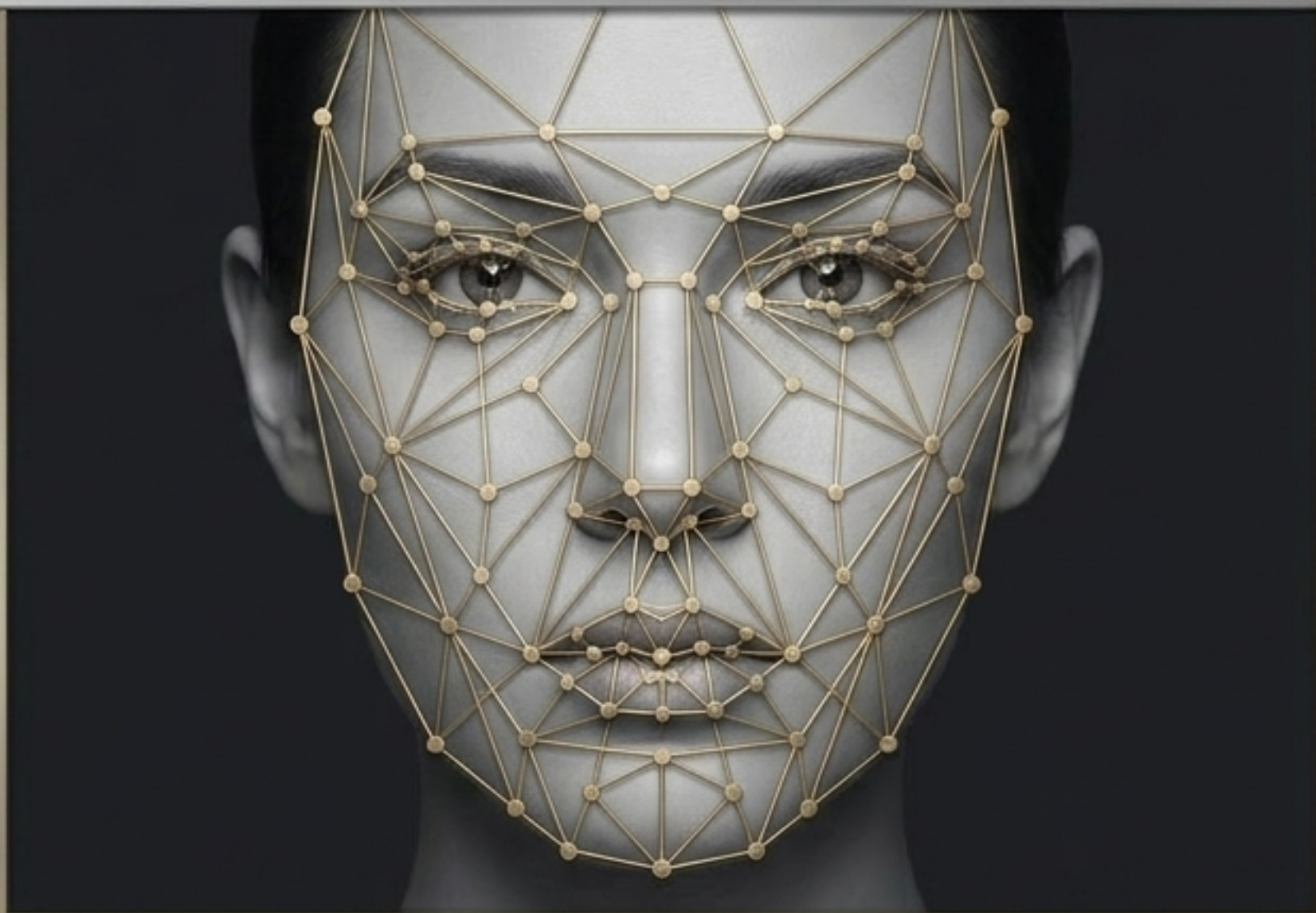
2026后期制作技术栈：宏观组装与微观雕琢



缝合现实：微表情引擎与绝对口型同步

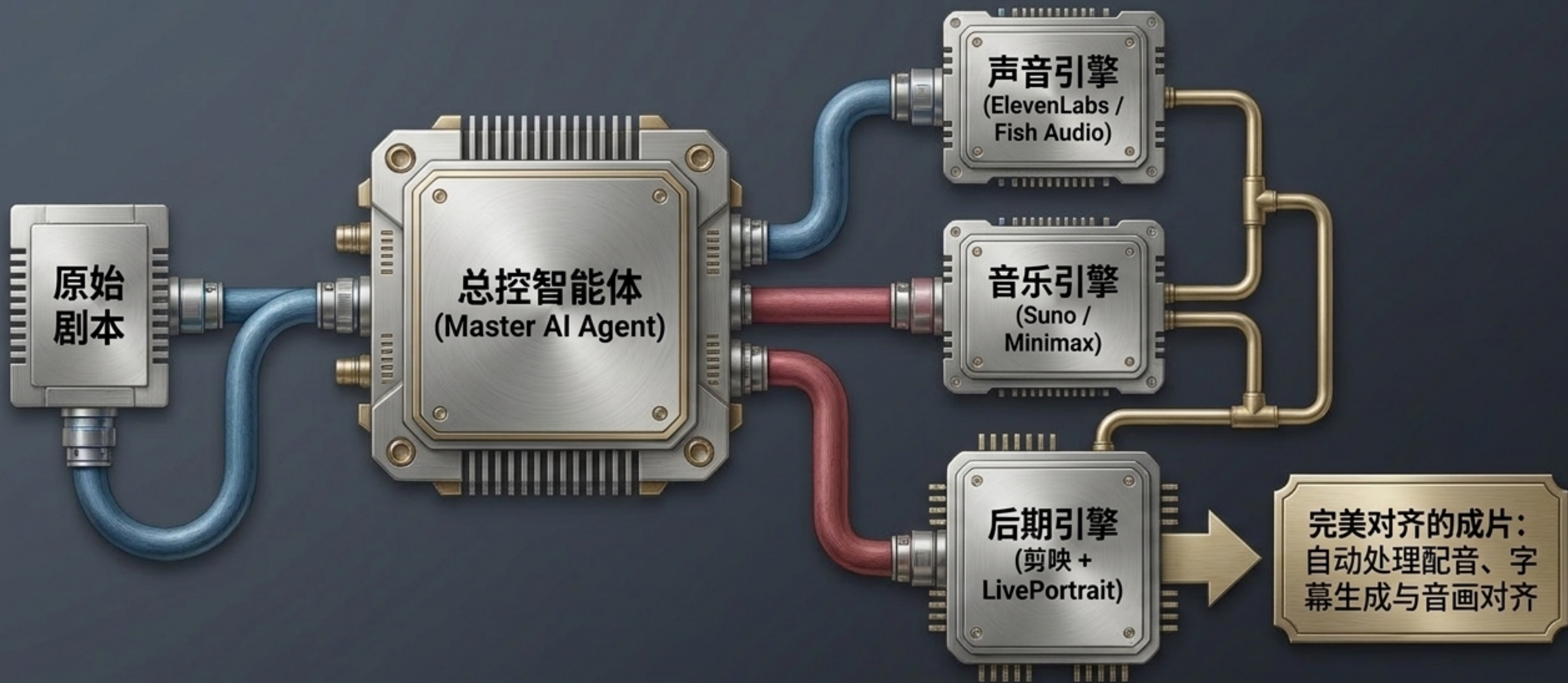


传统生成：面部僵硬，口型漂移



LivePortrait技术：通过底层面部节点映射，
让生成的音频与肌肉运动实现绝对咬合。

全自动数字片场：智能体的核心调度网络



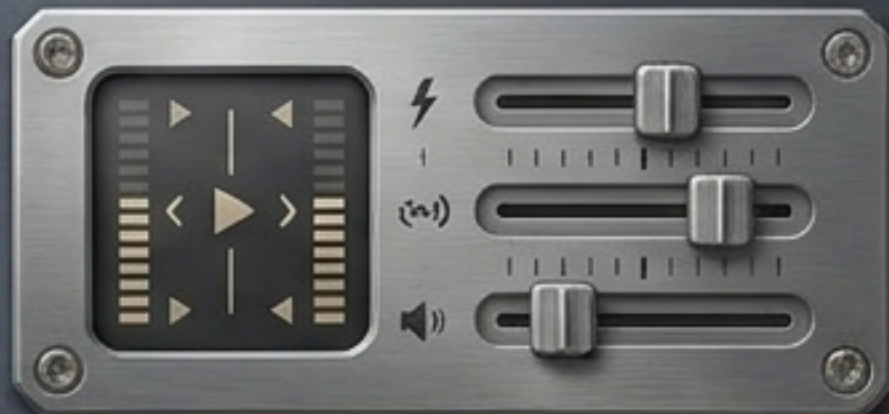
摧毁沉浸感的三个致命警报



警报 1：声纹漂移

现象：多集中角色声音不一致

对策：建立严格的基准音频样本库



警报 2：视听脱节

现象：粗糙的口型同步品质

对策：最终输出前，必须通过LivePortrait级工具进行口型查验



警报 3：情绪错位

现象：音乐风格与场景内在情绪完全背离

对策：引入智能体进行剧本情绪的深度预分析

2026数字影视的声音与后期法则

1. 需求决定引擎

永远根据具体的情绪爆发力与语言环境来倒推选择配音工具。

2. 让智能体充当指挥家

停止纯手动拼凑。利用 AI Agent 将音乐生成、字幕挂载与后期任务串联为自动化流水线。

3. 守住沉浸感底线

无论工期多紧，最终渲染输出前，口型同步的逐帧检查是不可逾越的红线。

The Final Cut: 重塑数字视听的边界



沉浸感是唯一指标：

音频品质与完美的口型同步，是连接观众情感的最关键桥梁。

工具+智能体=生产力核爆：

正确的生成工具结合智能体的工作流，能够实现指数级的效率飞跃。

后期的降维打击：

卓越的自动化后期制作，能够化腐朽为神奇，全面拉升内容的最终质感。

下一讲预告：从概念到成片的完整工业流

在 Part 4，我们将揭开最终的底牌。

解锁从“闪念萌芽”到“全片交付”
的完整 AI Agent 闭环工作流。

深度拆解爆款短剧的实战案例。

Q&A

感谢收看 Part 3
欢迎提问与交流

