

中文声线与华乐音色（再谈 RSI）

讲者：Michael Huo

活动：唯爱AI音乐学苑第三讲

日期：2026-10-10 周六

时间：17:30-18:30 PDT (美东 20:30 EDT | 北京 10月11日 08:30)

Zoom 会议号：952 9981 2540

密码：18970521

学苑官网：music.weiai.ai

Luma: luma.com/5j2b3108

(注：Luma 标题未更新，以本讲义为准)



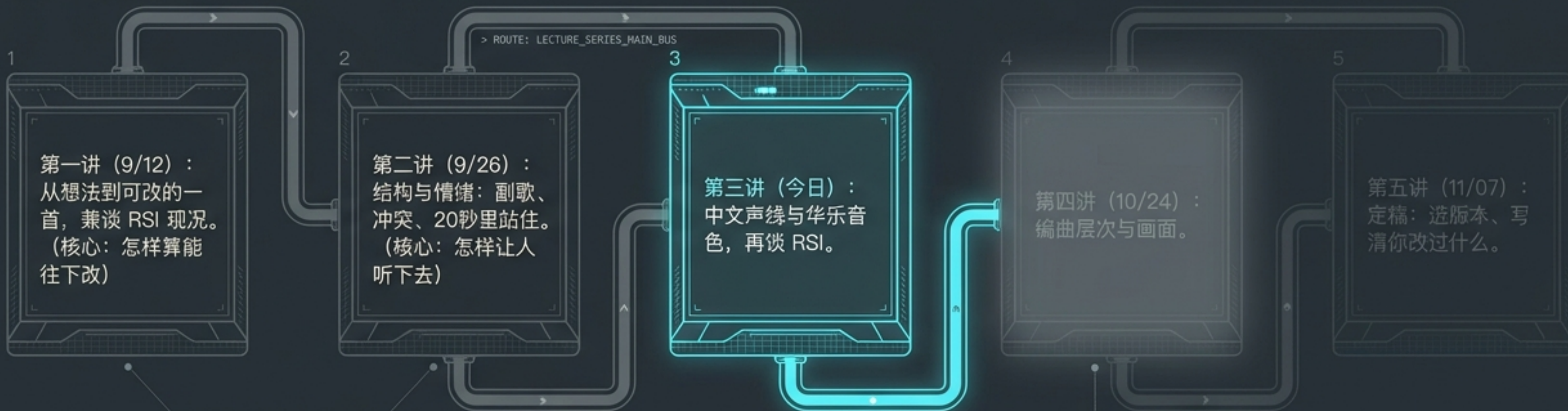
主办：北卡浙大校友会、北美高校联盟AI创业公益俱乐部

承办：唯爱AI基金会、唯爱AI公益基金会

协办：北美高校联盟各分会及姊妹浙大校友会

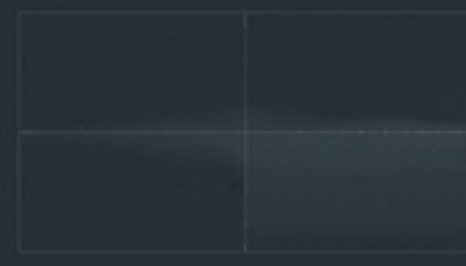
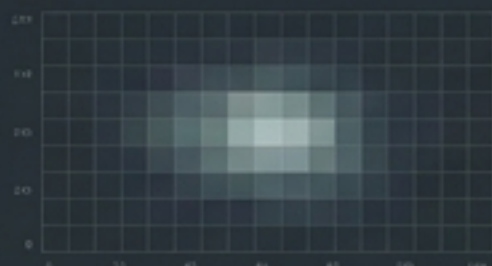
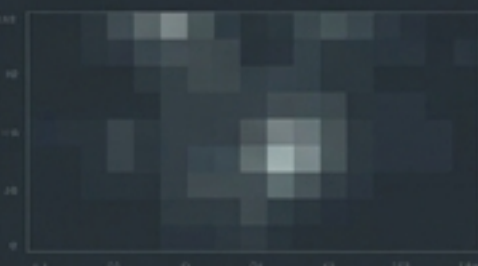
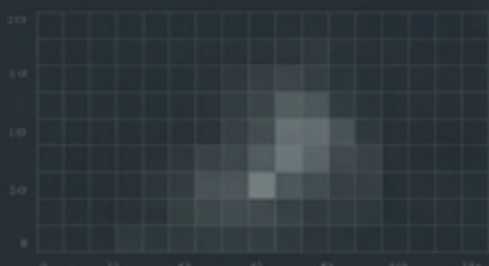
学苑学习路径图：当前定位

> SYSTER: PATHWAY_ANALYSIS_ACTIVE
> ROUTE: LECTURE_SERIES_RAIN_BUS



> SYSTER: PATHWAY_ANALYSIS
> ROUTE: LECTURE_SERIES_RAIN_BUS

> SYSTER: PATHWAY_ANALYSIS_ACTIVE
> ROUTE: LECTURE_SERIES_RAIN_BUS



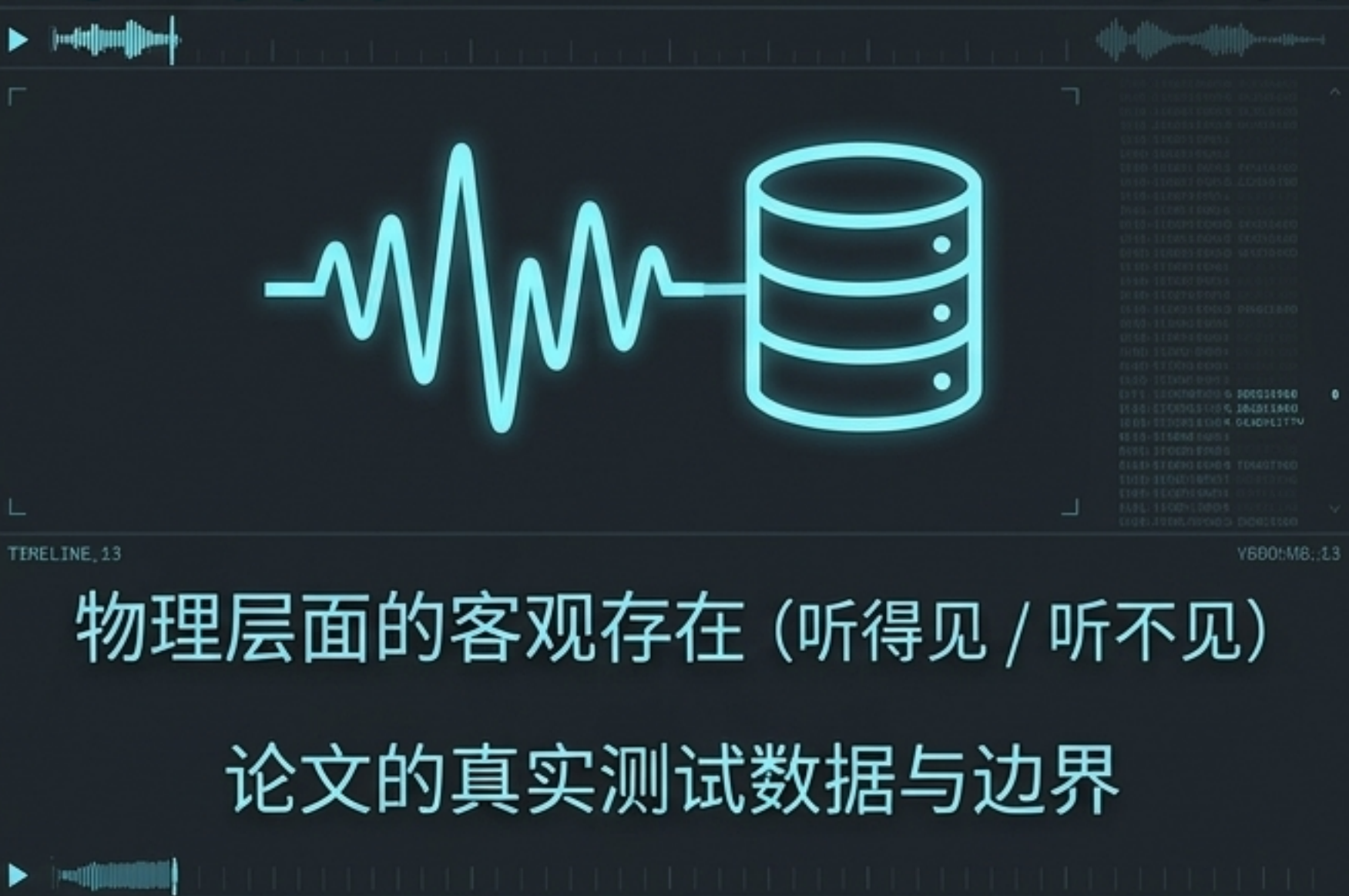
今日诊断准则：剔除噪声，只看信号

CHANNEL_A: SIGNAL_NOISE_RATIO_CRITICAL



CHANNEL_B: OBJECTIVE_DATA_STREAM_VERIFIED

REFERENCE_ID: PHYS_DIAG_L3



本讲不空谈审美，只做物理级诊断。

声线的同一性测试

同一句词，换了声线，还是不是同一首歌？

TIME: 00:21:34.90

TIME: 00:21:34.99



拆解 AI 人声的“塑料感” (Plastic Feel)

TIME: 00:43:12.55

SYSTEM: AUDIO_ANALYSIS_PLASTIC_FEEL_DETECTED



TIME: 00:43:12.55

TIME: 00:43:12.55

华乐音色物理检测矩阵

风格词写了「华乐」，耳朵里到底有没有这件乐器？

乐器 (Instrument)	提示词是否包含 (In Prompt?)	音频是否检出 (Audio Detected?)
二胡 (Erhu)	☒ 是 / 否	☐ 听得见 / 听不见
琵琶 (Pipa)	☐ 是 / 否	☐ 听得见 / 听不见
古筝 (Guzheng)	☒ 是 / 否	☐ 听得见 / 听不见
笛 (Dizi)	☐ 是 / 否	☐ 听得见 / 听不见
笙 (Sheng)	☒ 是 / 否	☐ 听得见 / 听不见

诊断依据仅为物理听觉，不评判演奏口味。

现场诊断：版本 A vs 版本 B

版本 A (Version A)

测试词句：[等待现场输入]

声线稳定性诊断：

- 咬字偏移：[]
- 气口连贯：[]
- 尾音支撑：[]

(异常物理依据：第 ____ 秒)

华乐检出率（真实还原乐器）：

[_____]

核心判定：

- 声线站住没有？ []
- 华乐听得见没有？ []

版本 B (Version B)

测试词句：[等待现场输入]

声线稳定性诊断：

- 咬字偏移：[]
- 气口连贯：[]
- 尾音支撑：[]

(异常物理依据：第 ____ 秒)

华乐检出率（真实还原乐器）：

[_____]

核心判定：

- 声线站住没有？ []
- 华乐听得见没有？ []



05 : 00

听觉重置 (Intermission)

17:55 - 18:00

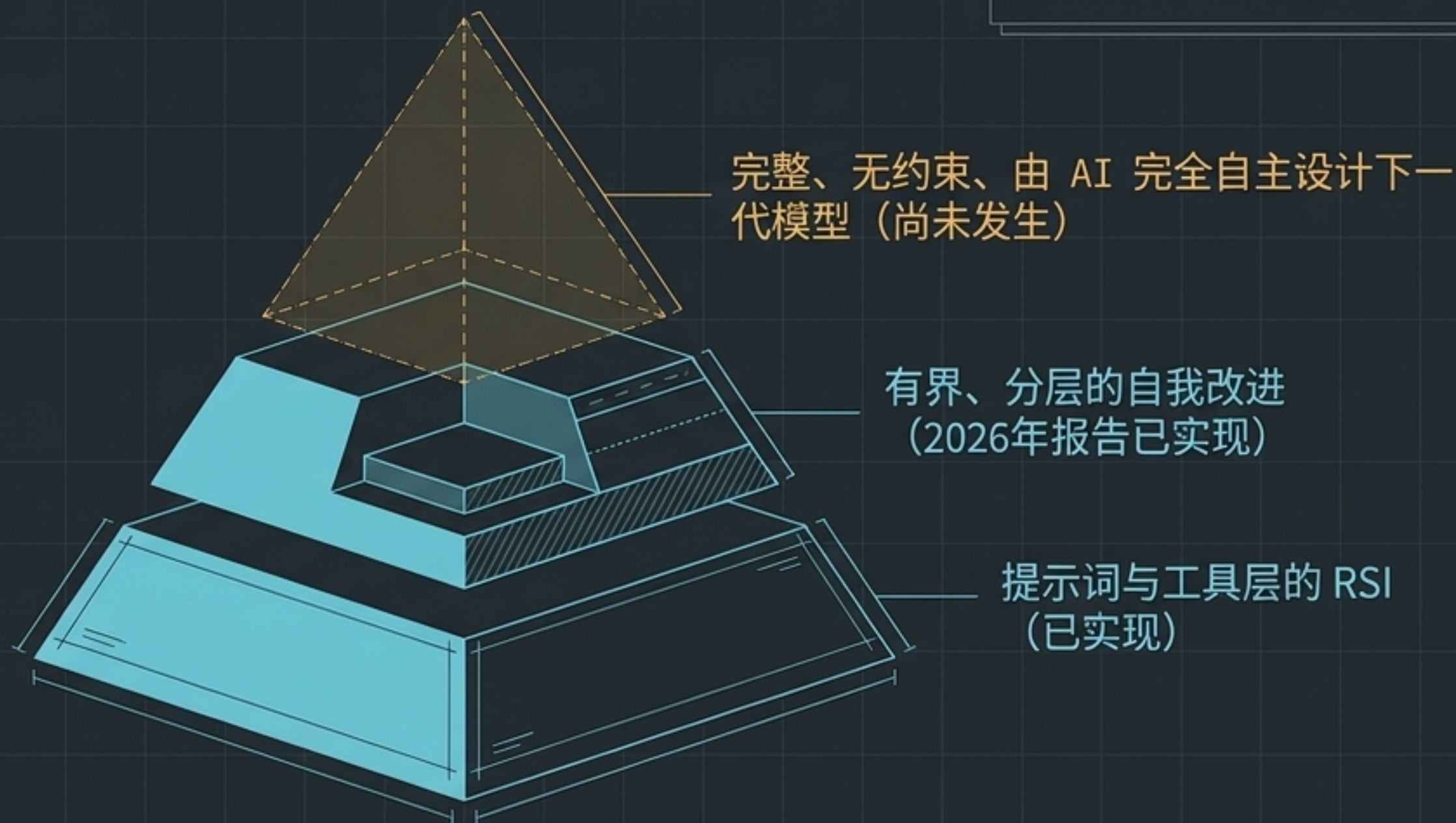
让耳朵休息5分钟。

下一环节：从音频诊断转向代码与论文诊断——再谈 RSI。

重新校准 RSI (递归自我改进) 的当前边界

TERMINAL

沿用第一讲定义：系统用自己的输出改进下一步（包括提示词、工具、代码、评测或训练配方）。



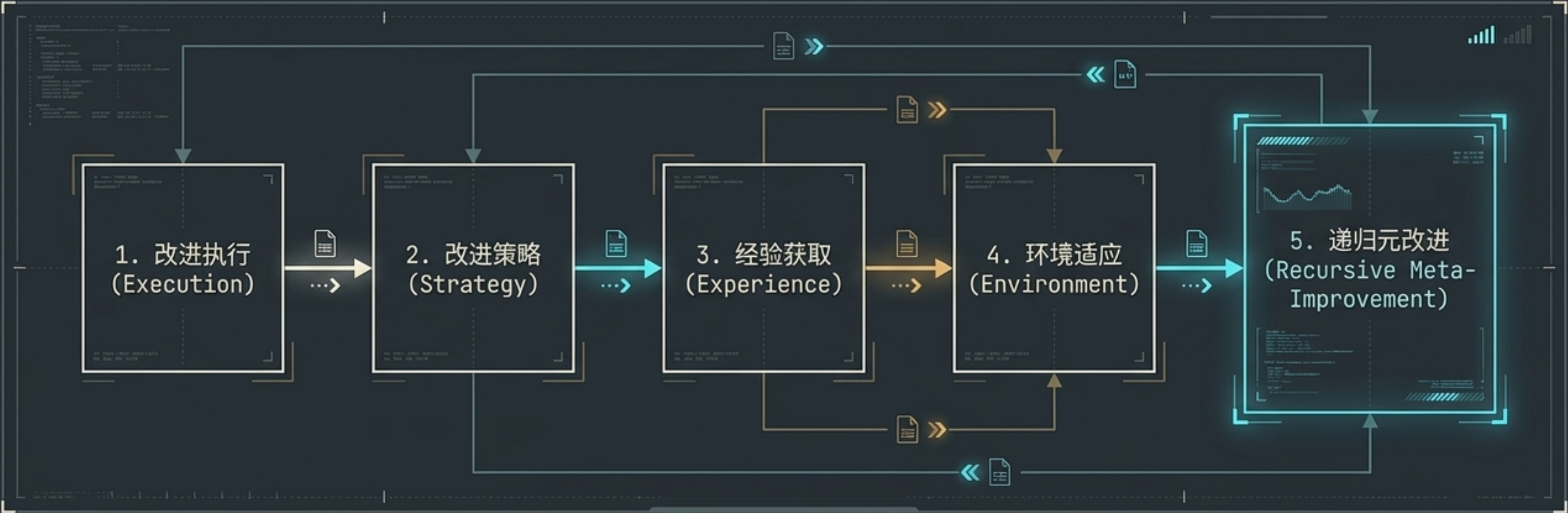
注：不演示改模型，因为完整 RSI 尚未发生。目前的音乐三轮迭代，仅为应用层示例。

行业宣称的时间线与数据审核



RSI 研究的演进路线图

[arXiv:2609.11873] (9月10日提交, 9月22日修订) - 综述论文



注: 此为理论框架与测绘路线图, 并非演示系统本身。

认知与软件级 RSI 论文对齐测试

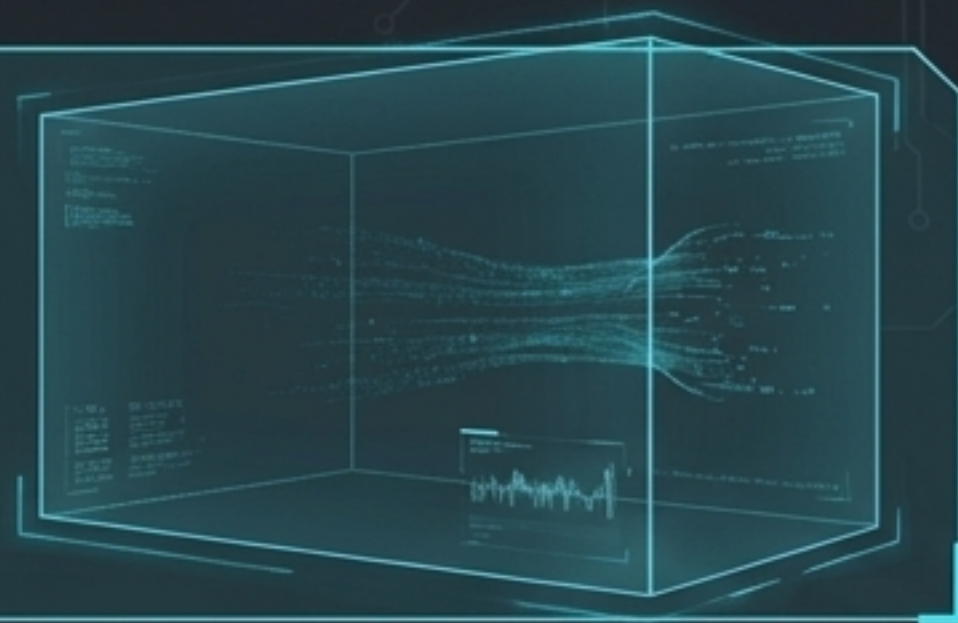
Capability Heatmap	Dream-RSI	AIDE^2	RRSI
	[arXiv:2609.14858] (9/14, Google/DeepMind)	[arXiv:2609.26457] (9/22)	[arXiv:2609.24972]
改了什么 (Modified?)	下一步搜索策略 (利用过去搜索记录当模拟器)	8天自主运行留下7次改进, 未见过任务有效	提示词、控制流、工具、记忆 (正则压制过拟合)
没改什么 (Preserved?)	模型本身、评测器、工具、 权重完全不动	[数据缺失]	底层架构
审计状态 (Audit Status)	方法论透明	⚠️ 论文自述, 未见独立复现	⚠️ 分数由作者自报

具身智能 (Embodied AI) 中的 RSI 尝试

[本讲新增]

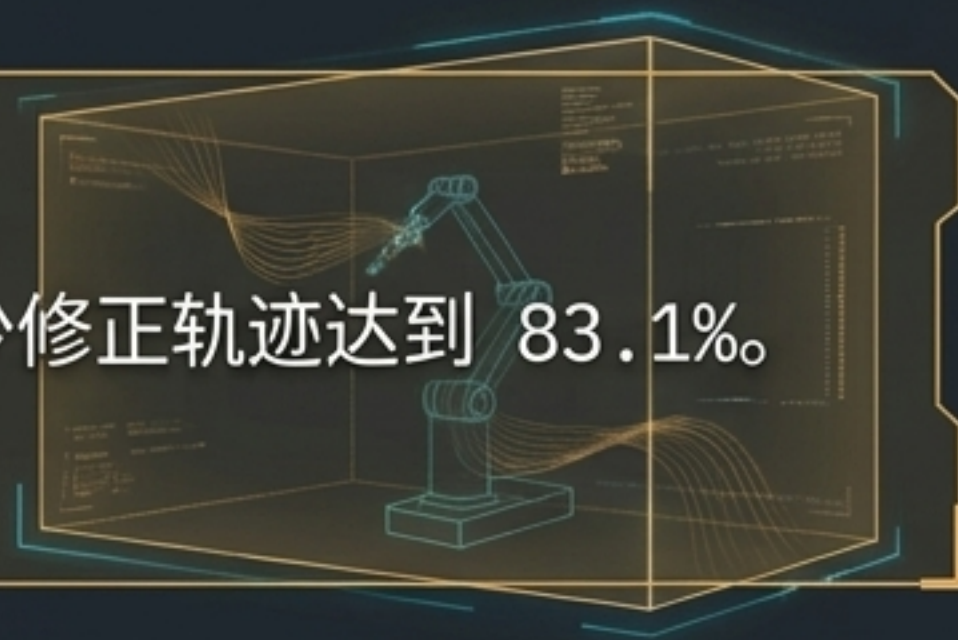
仿真环境 Simulation

- EmbodiedSmith [arXiv:2610.07969] (10/6提交)
 - 机制：仿真里场景和任务互相改，用于扩充训练数据。
 - 局限：没有真机成功率数据。

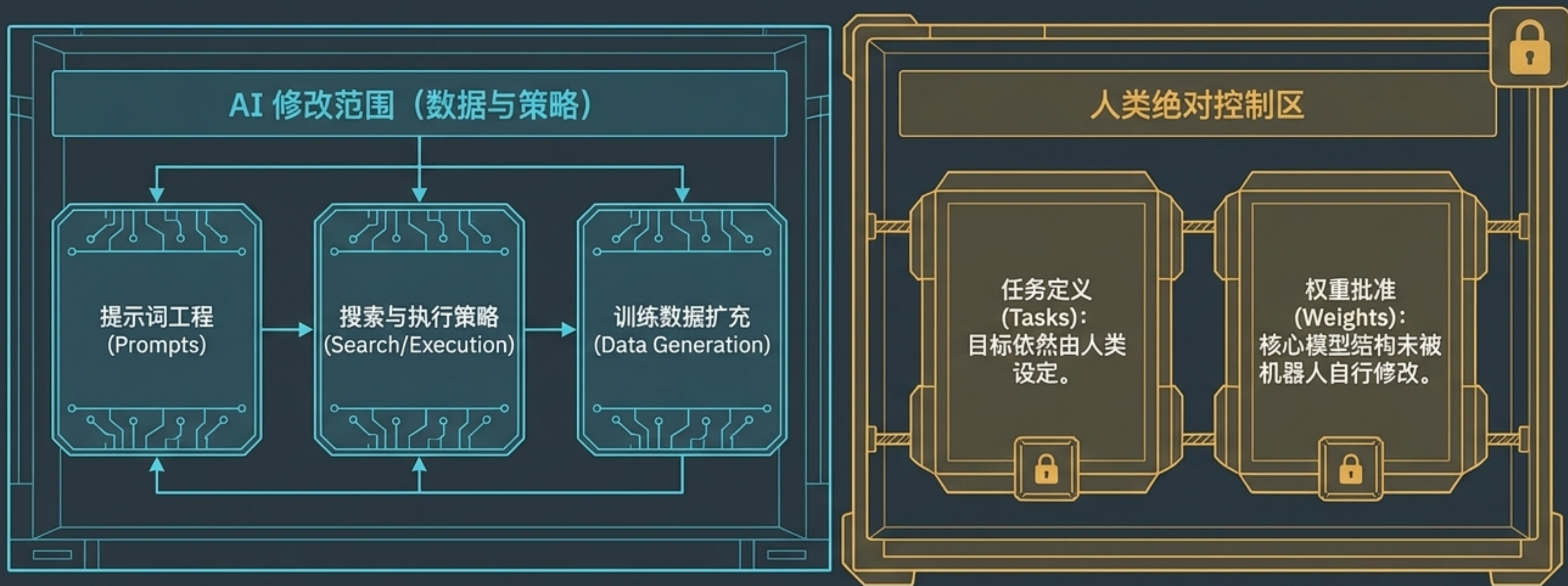


真机跨越 Sim-to-Real

- EmbodiRSI [arXiv:2609.38905] (桌面任务)
 - 数据：仿真成功率两轮后从 50.4% 升至 83.5%。真机仅用极少修正轨迹达到 83.1%。
 - 审计状态：⚠️ 论文自述。

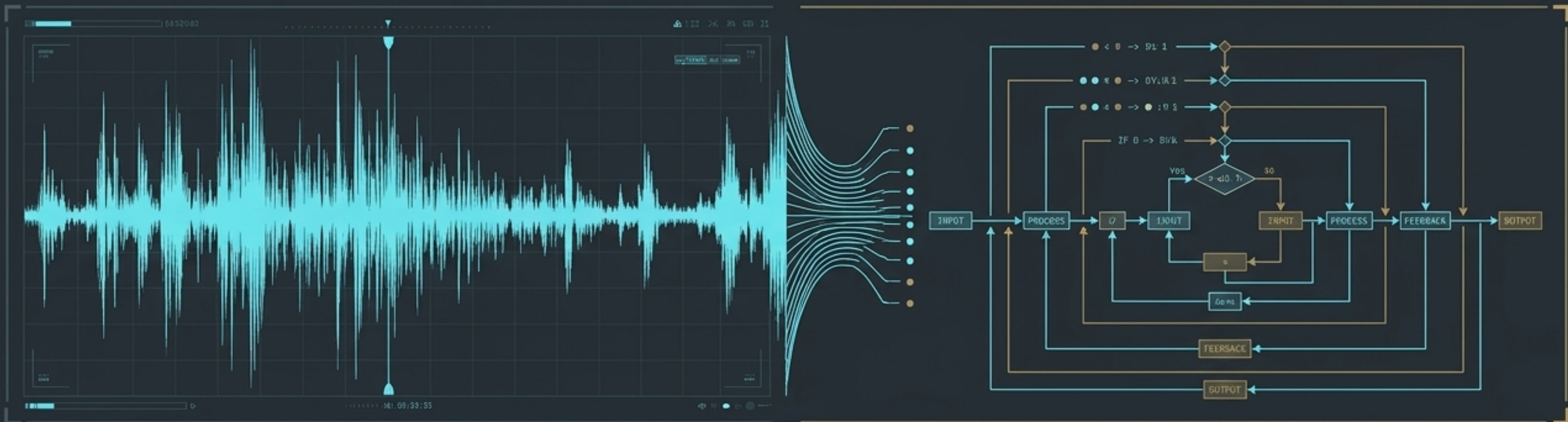


阶段性判决：当前 RSI 中人类的位置



结论：今年底还有两个月，目前 AI 仅在改数据和策略。我们等到明年此时再做对比。

跨域共识：抛弃标签，测量输出



在音乐中：

不管风格词是否写了「华乐」，只测物理波形中是否听得见。

核心哲学：无论是评测 AI 音乐还是 AI 演进能力，方法论完全一致——信任你的耳朵，阅读研究方法，验证系统边界。

在代码中：

不管厂商是否自称「自动化研究员，只看论文方法中权重是否被真实修改。

诊断结束，下一步准备

[18:25 - 18:30]

现场问答与论证
(Live Q&A)

[10月24日]

唯爱AI音乐学苑·第四讲
预告主题：编曲层次与画面

Next Steps: 准备好你的工程文件，我们将从基础声线诊断，进入多轨编曲的空间与画面构建。

感谢北卡浙大校友会及所有协办方。详情访问 music.weiai.ai