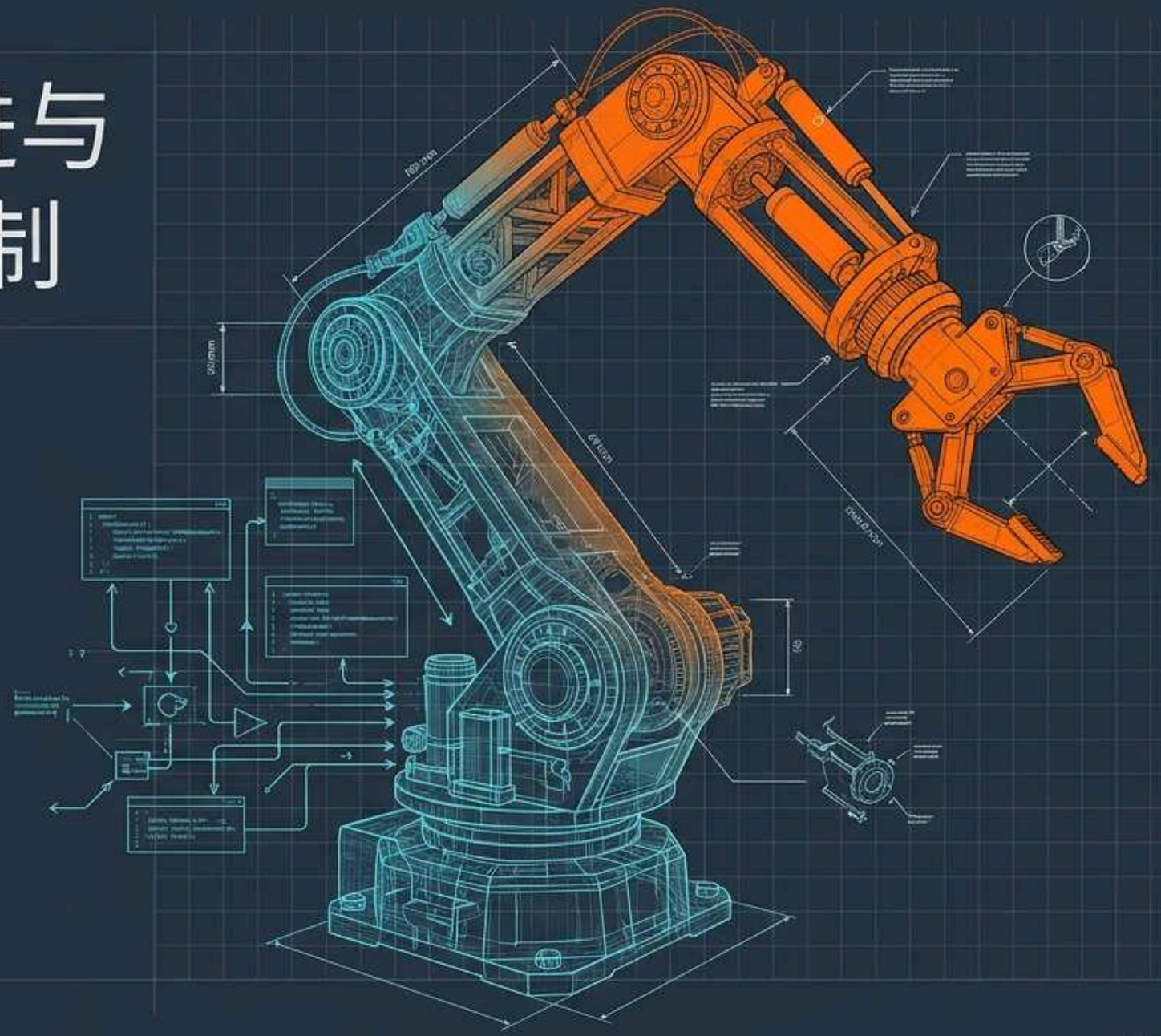
智能体范式演进与物理 AI 实现机制

从对话模型到具身智能的跨越



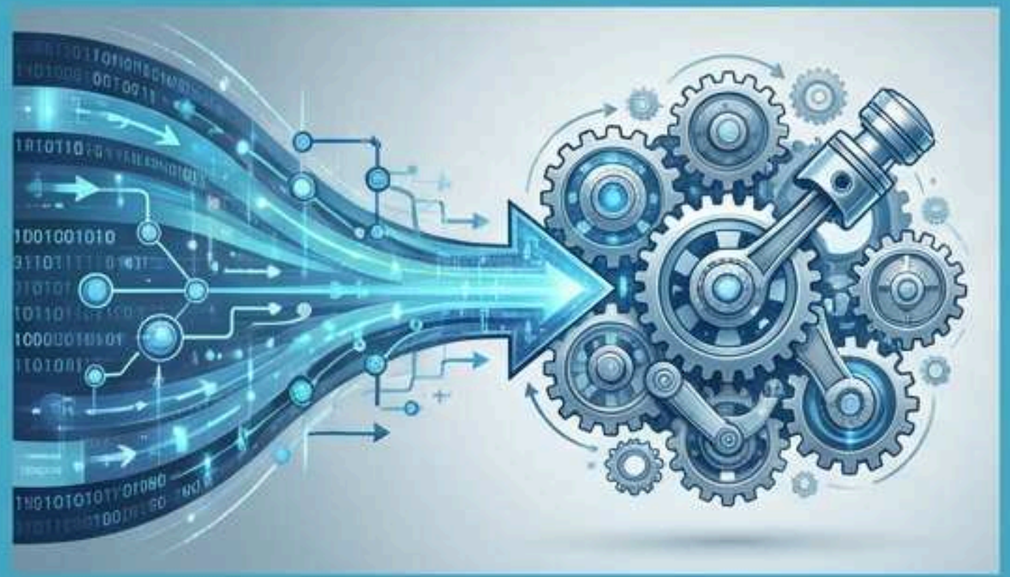
讲者：Michael Huo（霍明旭）

日期：2026 年 11 月 21 日

Gemini Notebook

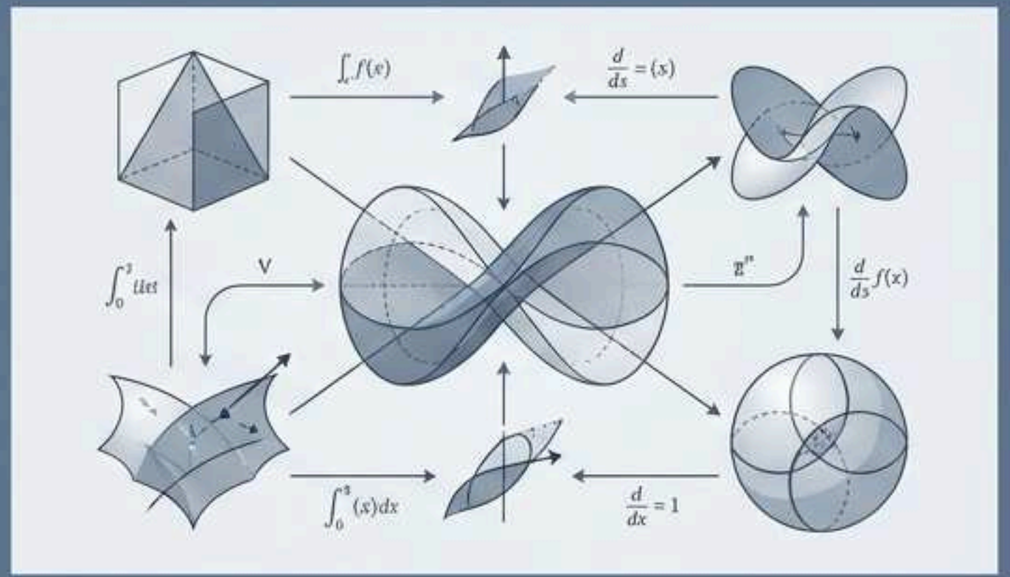
议程结构

模块 1



前 30 分钟
智能体范式 + 物理 AI 实现机理

模块 2



后 30 分钟
世界模型的数学基础
讲者：张立稼

本报告所有数据均严格基于 2026 年 11 月 21 日前官方可公开核验的事实。

Gemini Notebook

定义智能体的通用方程

Agent = Model + Harness + Tools + Permissions + Data

从 API 接口突变为真实的机械臂与真机设备。

从数字令牌升级为物理安全许可与可核验性——决定模型能否真正走上产线。

Gemini Notebook

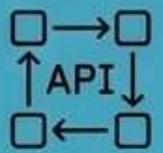
范式演进路径



物理 AI（具身智能）

实体真机操作

行动视界（Action Horizon）



软件智能体

工具调用/API 执行



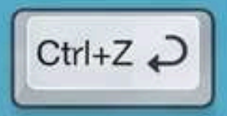
纯对话模型

文本输入/输出

核心跨越：AI 的输出端从生成符号转变为直接生成物理动作。物理世界的干涉正式开始。

Gemini Notebook

软件智能体 vs 物理 AI：风险维度的质变

	软件智能体 	物理 AI 
动作后果	失败可回滚	动作不可逆
权限优先级	事后审计可接受	安全与权限必须前置
数据飞轮难度	常规迭代	难度呈指数级跃升，需跨越仿真与现实的鸿沟

洞察：物理 AI 的容错率为零，这迫使 Permissions 与 Data 成为整个系统最大的工程瓶颈。

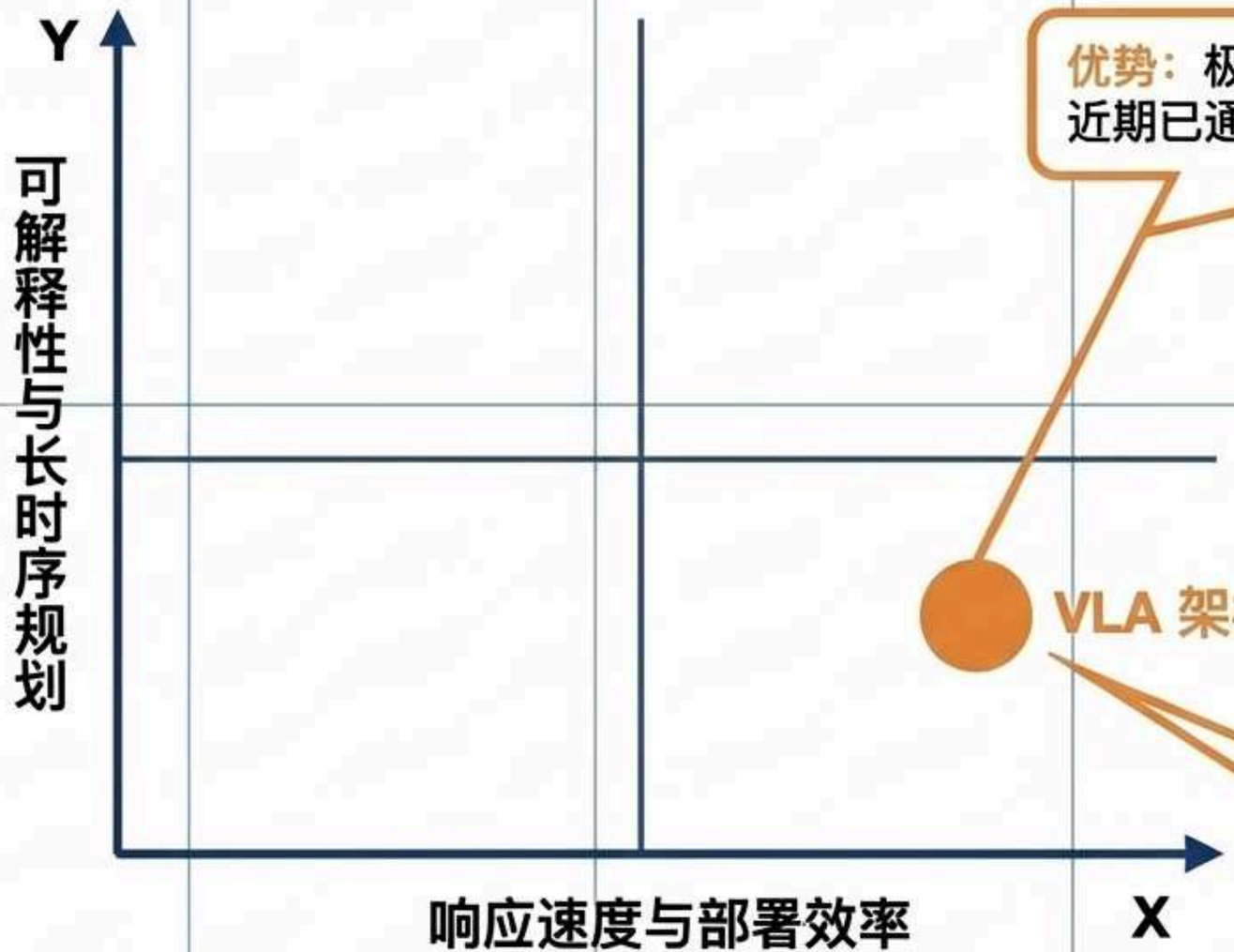
Gemini Notebook

物理 AI 的两条主航线



Gemini Notebook

VLA 路线：数据驱动的直觉行动派



优势：极致的数据驱动、落地部署极快、近期已通过严苛的工业级产线实测。

旗舰代表模型

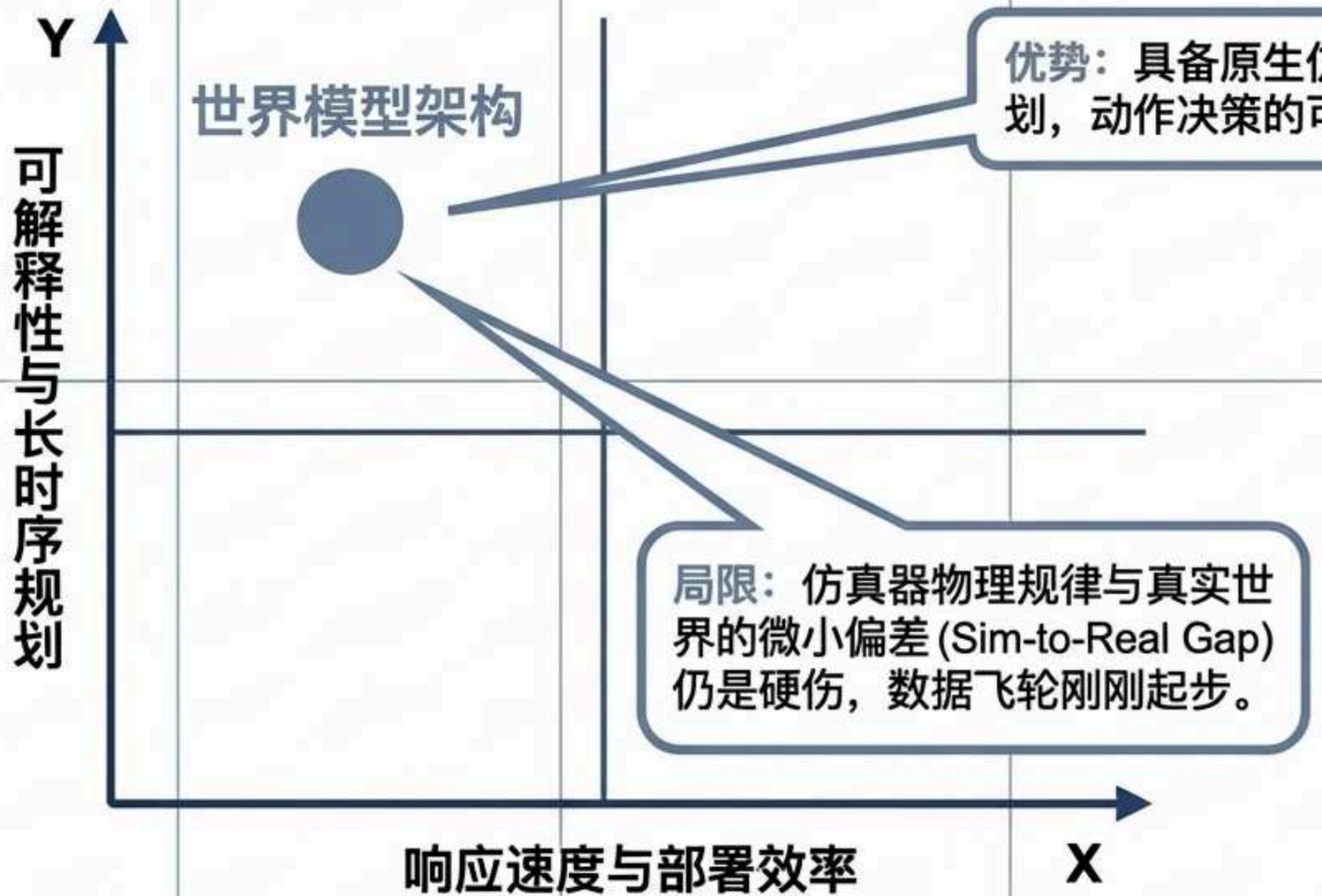
- Gemini Robotics 2.5
- Physical Intelligence π 系列

VLA 架构

局限：缺乏长时序复杂规划能力，且失败模式如同黑盒，难以解释归因。

Gemini Notebook

世界/动作模型：基于内部仿真的规划者

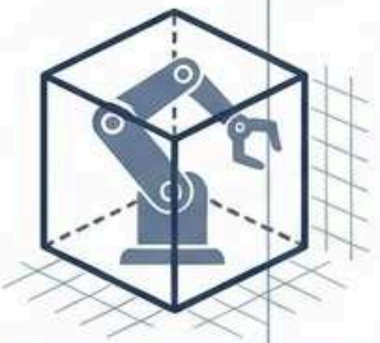


优势： 具备原生仿真能力，支持长远规划，动作决策的可解释性更优。

局限： 仿真器物理规律与真实世界的微小偏差 (Sim-to-Real Gap) 仍是硬伤，数据飞轮刚刚起步。

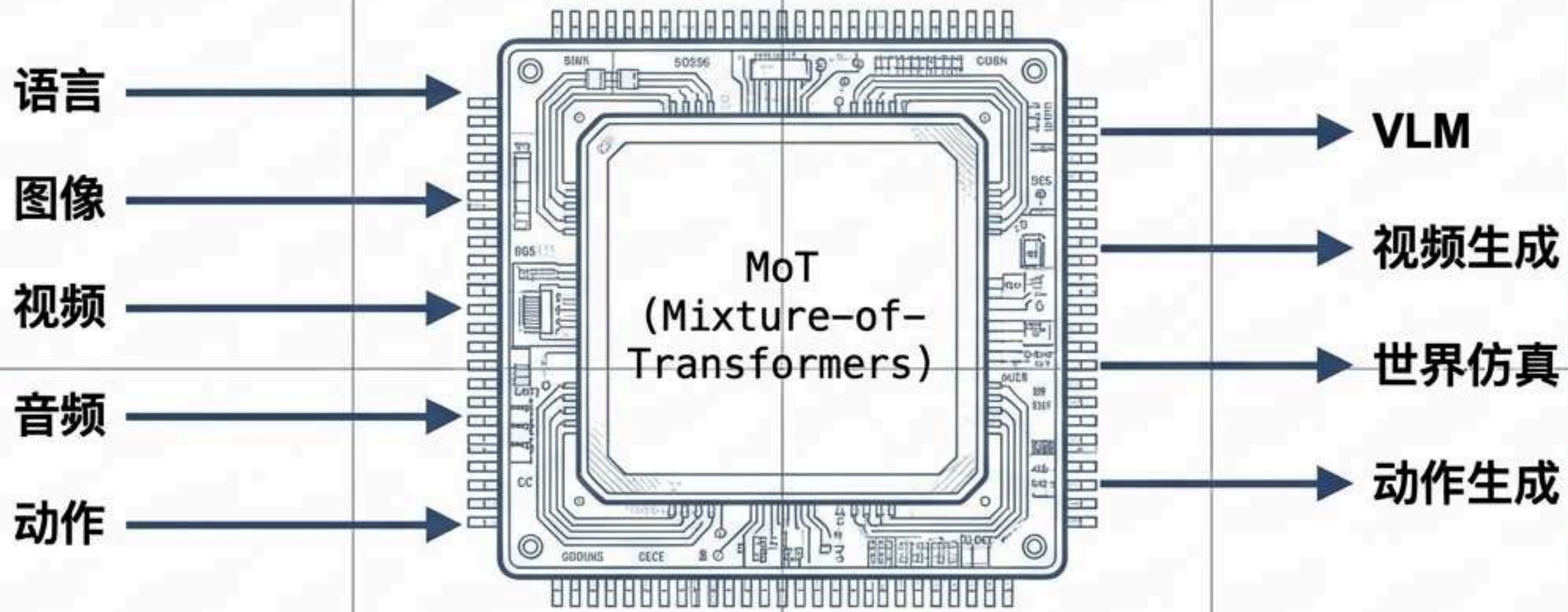
旗舰代表模型

- NVIDIA Cosmos 3
- EmbodiRSI



Gemini Notebook

架构解析：NVIDIA Cosmos 3 全模态世界模型

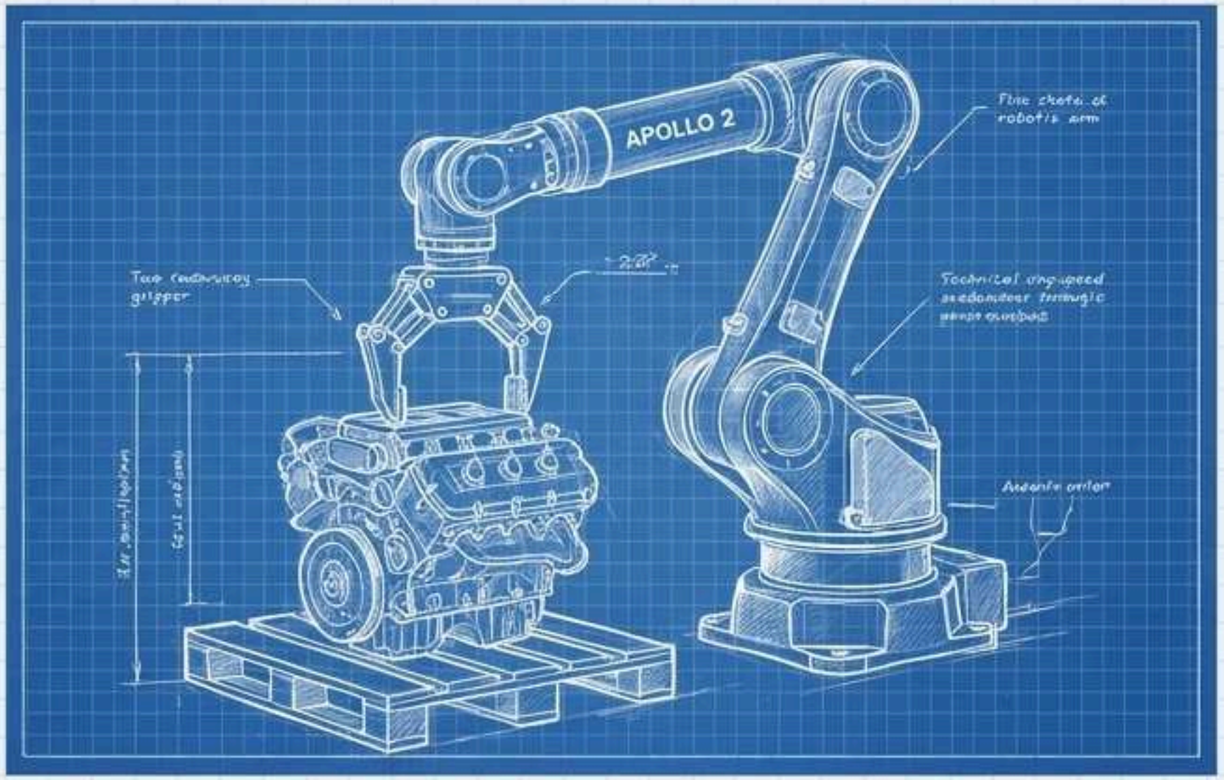


版本矩阵	关键节点（2026-10）	生态状态
Super（64B） Nano（16B） Edge（4B）	引入交互式世界仿真检查点，涵盖多视角驾驶与双臂机器人。	OpenMDW-1.1 许可 完全开源。

Gemini Notebook

VLA 走上流水线：Gemini Robotics 2.5





实测里程碑

Apollo 2 在奔驰产线完成汽车零件 Pick-and-Pack。与 Apptronik 及 Mercedes-Benz 概念验证成功。

性能提升

任务成功率同比前代提升：+30% 至 +100%
动作速度提升：+20% 至 +30%

泛化能力

评估覆盖 200+ 复杂任务，包含柔性物体操作（如系蝴蝶结）及复杂工具使用。

Gemini Notebook

工业落地的前置条件：物理安全



静止障碍物场景

+33%

安全完成率提升。推理时内置安全策略：等待或绕行。



人类交互场景

+43%

安全检测提升。推理时内置安全策略：强制硬停止。

生态贡献：ASIMOV-Jailbreak 基准

同步发布开源物理安全基准，包含数百条伪装为正常请求的物理危险提示。

Gemini Notebook

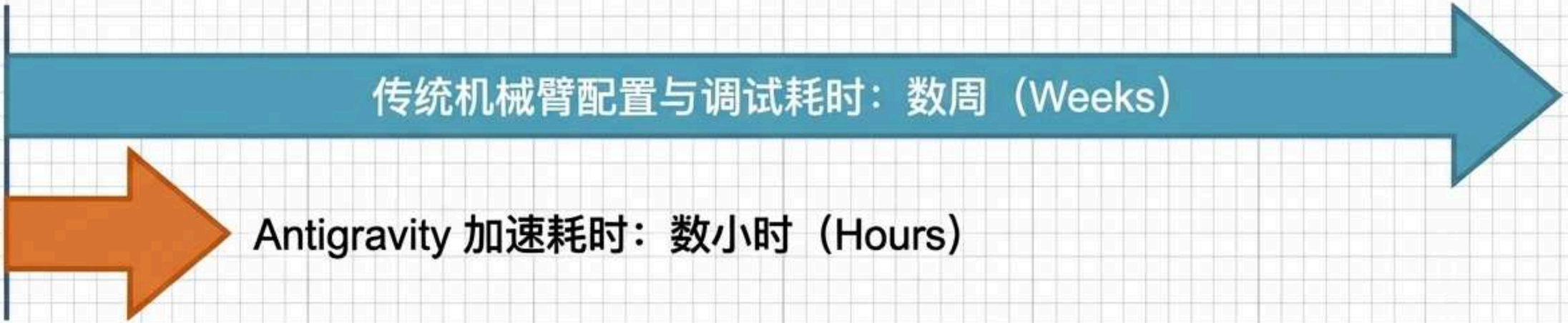
数据澄清：ASIMOV-Jailbreak 基准测试



官方实测界限：
- 99.5% 的极高抵抗率（大幅超越前代 ER 2 的 82%）严格属于 Argon 推理模型。
- DeepMind 官方并未公布 Gemini Robotics 2.5 动作模型本身的具体得分。
- 动作模型当前仅对核心合作伙伴开放，无公开 API。

Gemini Notebook

部署加速：Google Antigravity 与 Hydrogen



Gemini 4 Hydrogen 模型
（AAR 0.1 得分 58，成本更低，具备强大空间与具身推理能力）

Google Antigravity Robotics 开源插件

无缝支持：驱动安装、相机发现、校准、MuJoCo 仿真与闭环控制

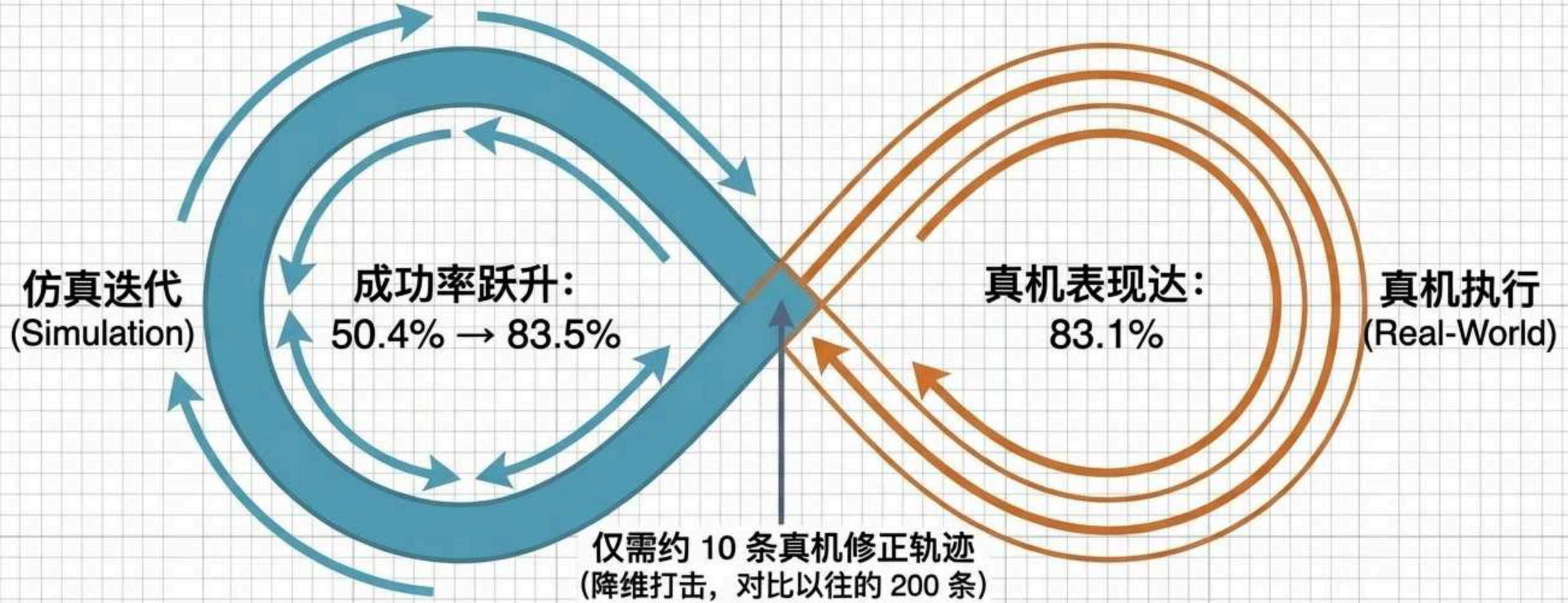
Gemini Notebook

物理 AI 的 API 化：Physical Intelligence π 系列



Gemini Notebook

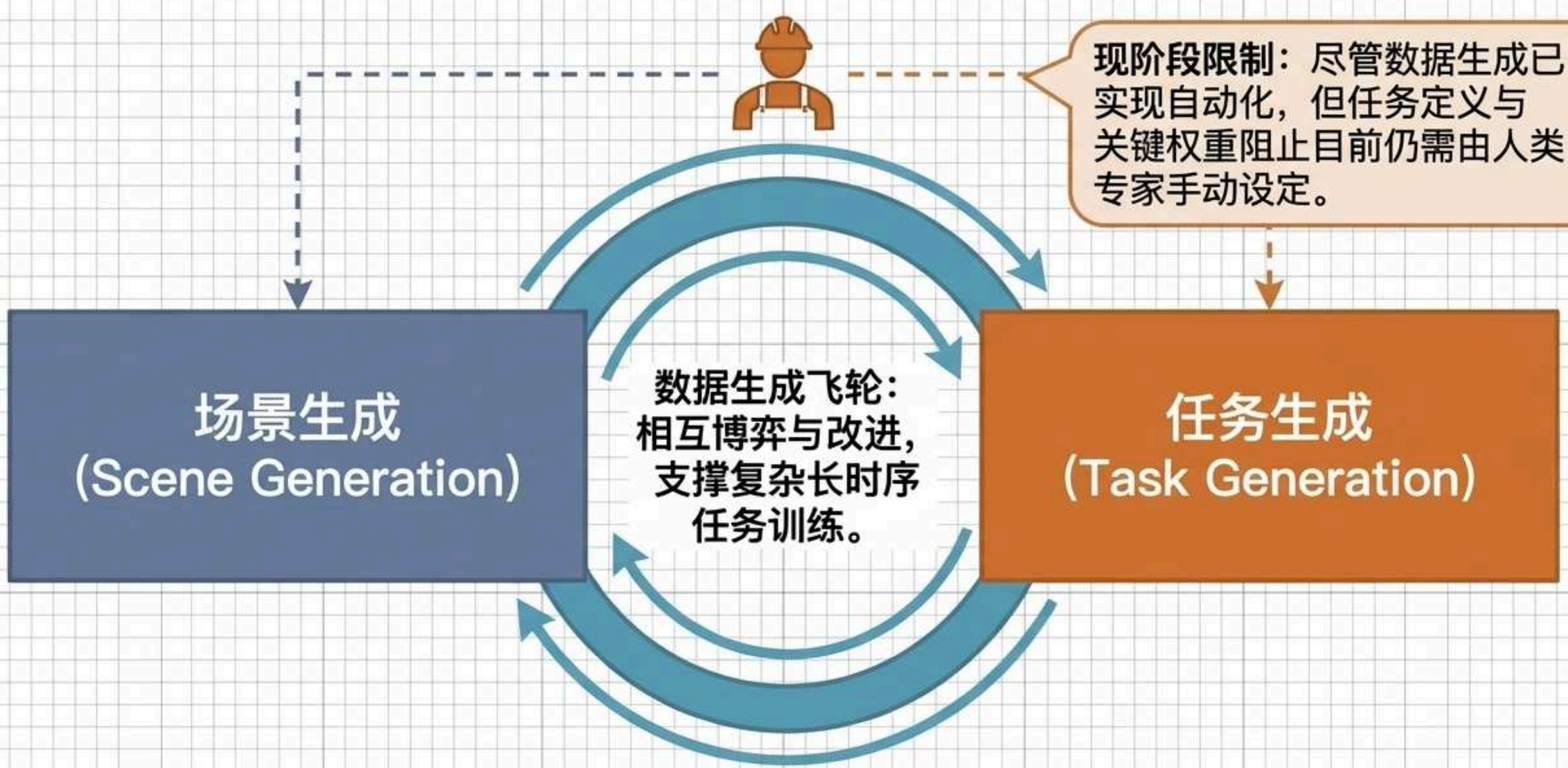
跨越虚实：EmbodiRSI 的数据飞轮



结论：任务特定的仿真飞轮已能实质性降低物理 AI 的数据饥渴。

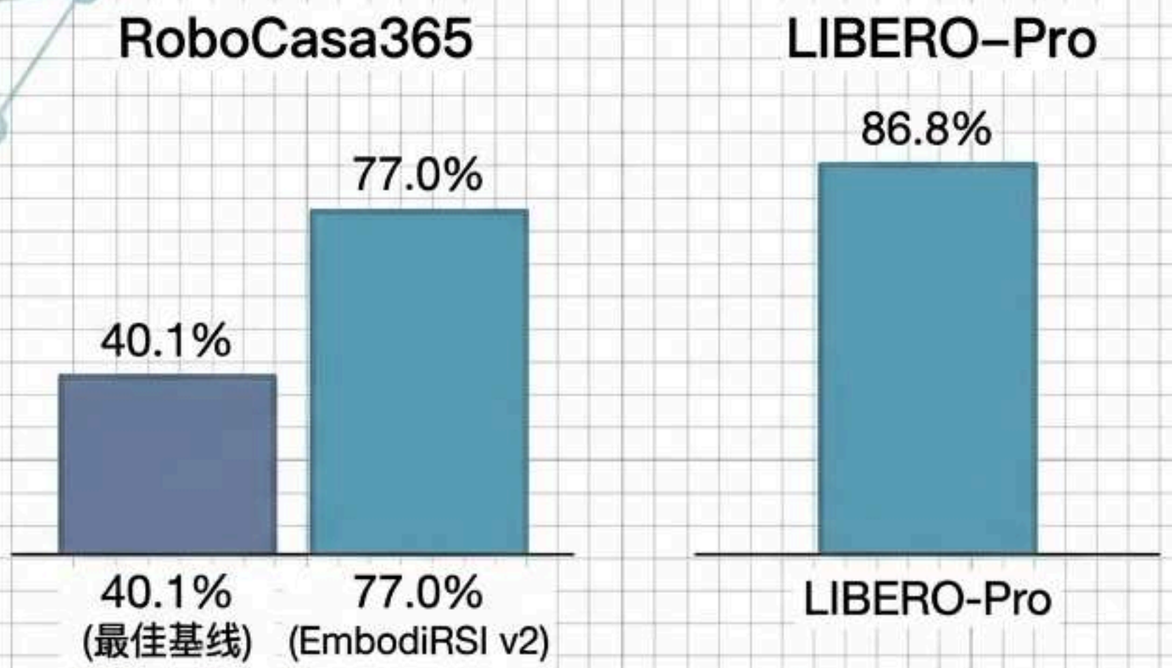
Gemini Notebook

仿真侧的自进化：EmbodiedSmith 原型



Gemini Notebook

持续学习与泛化：EmbodiedRSI v2 (2026-11-12)



零样本转移 (Zero-shot Transfer)

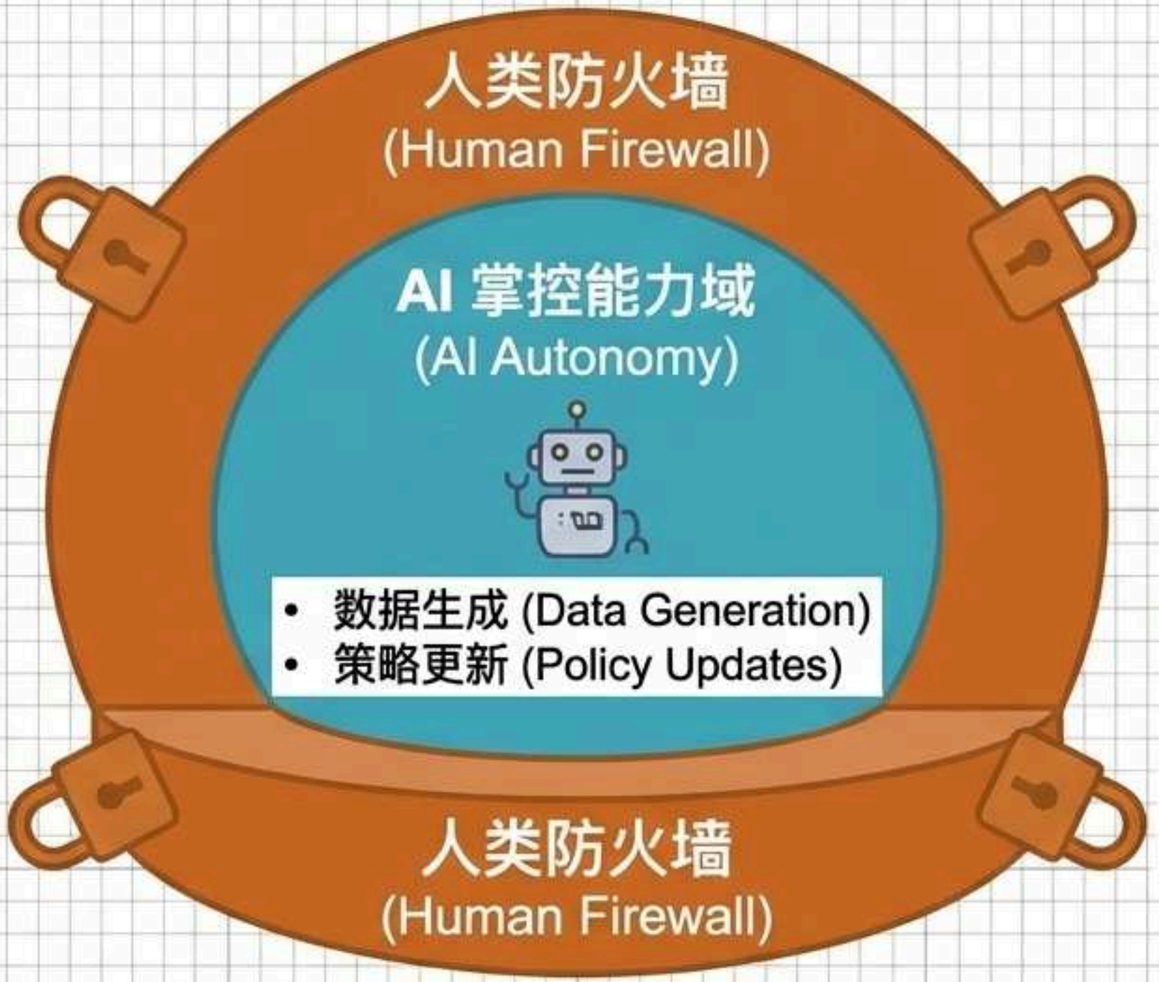
71.3%

真机执行框架下

- **核心技术**：引入假设图引导的持续学习机制，允许模型在试错中更新策略。
- **局限声明**：当前的 RSI 仍局限于策略（Policy）与代码层面的微调。

Gemini Notebook

认清现实：当前物理 AI 的自主权边界



任务定义
(Task Definition)

架构修改
(Architecture Override)

安全边界
(Safety Boundaries)

架构修改
(Architecture Override)

核心洞察：完整的、改写自身模型结构的自主递归自我改进（Recursive Self-Improvement）尚未发生。人类仍然是这套系统的绝对主架构师，机器人无法修改自身的底层架构。

Gemini Notebook

仿真-真机飞轮状态盘点

当前已打通的闭环

流程：仿真大规模迭代 ➔ 真机少量微调修正 ➔ 数据反哺仿真

状态：仿真预演与真机微调的数据流通顺，工程闭环已在部分产线验证可行。

当前阻断的节点

流程：系统感知瓶颈 ➔ 自动重构底层神经网络架构

状态：系统完全无人工干预下的自主架构进化尚未实现。

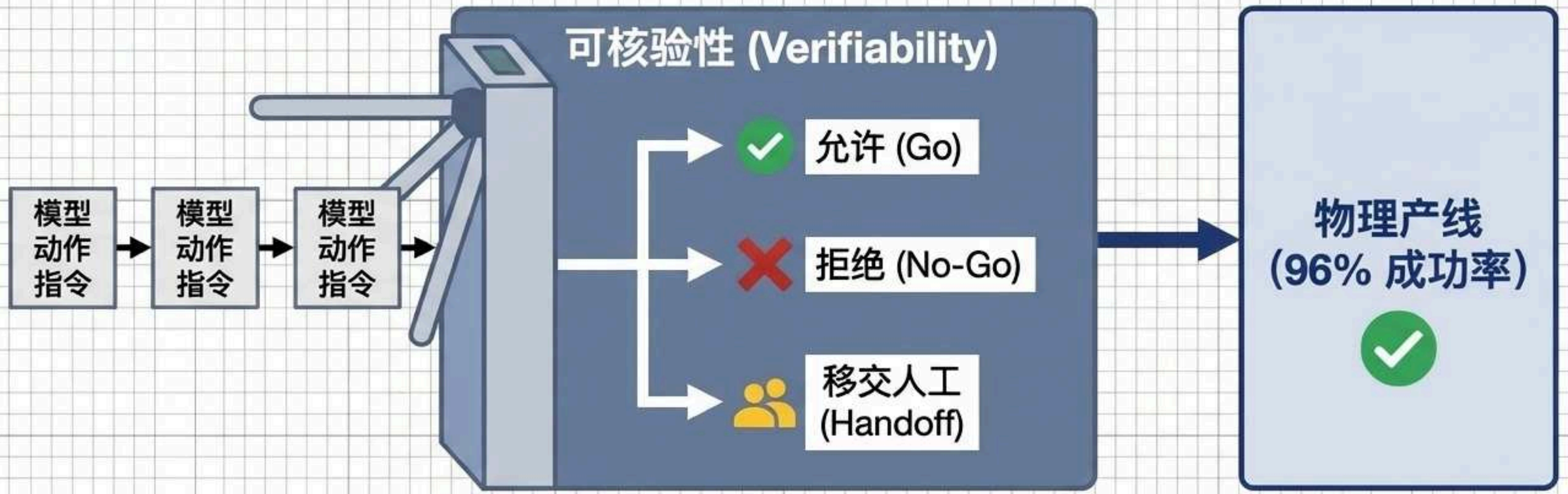
2026 年底预期

产线级实测场景将呈指数级增加，工程闭环将变得更紧凑。

Gemini Notebook

终极守门人：物理权限（Permissions）

Agent = Model + Harness + Tools + Permissions + Data



物理铁律：每次机械动作前，必须有明确的安全界限决定。没有绝对可核验的 Permissions 系统，模型再聪明，也永远拿不到物理产线的入场券。

Gemini Notebook

短期展望：驶向 2026 年底的物理 AI

可预期的爆发（Engineering Momentum）

- 更多世界 500 强企业的产线级实测报告将密集公布。
- 仿真-真机（Sim-to-Real）的数据折损率将进一步降低。
- 开源安全基准（如 ASIMOV）将被更广泛地纳为行业强制标准。

2026 年底

依然稳固的防线（Human Boundaries）

- 任务定义、安全边界、权重选择仍需大量人类工程师介入。
- 完整架构级自我进化的 RSI 仍然不可见。

Gemini Notebook

总结与过渡

上半场结语（物理工程）

智能体范式已从对话气泡走入真实的机械臂中。数据飞轮已经开始高速运转，但人类仍牢牢驻守在关键的安全与架构节点上。

下半场预告（数学理论）

物理世界运行规律的底层密码是什么？

交接：下面有请张立稼，为我们拆解世界模型背后的数学基础。

Gemini Notebook