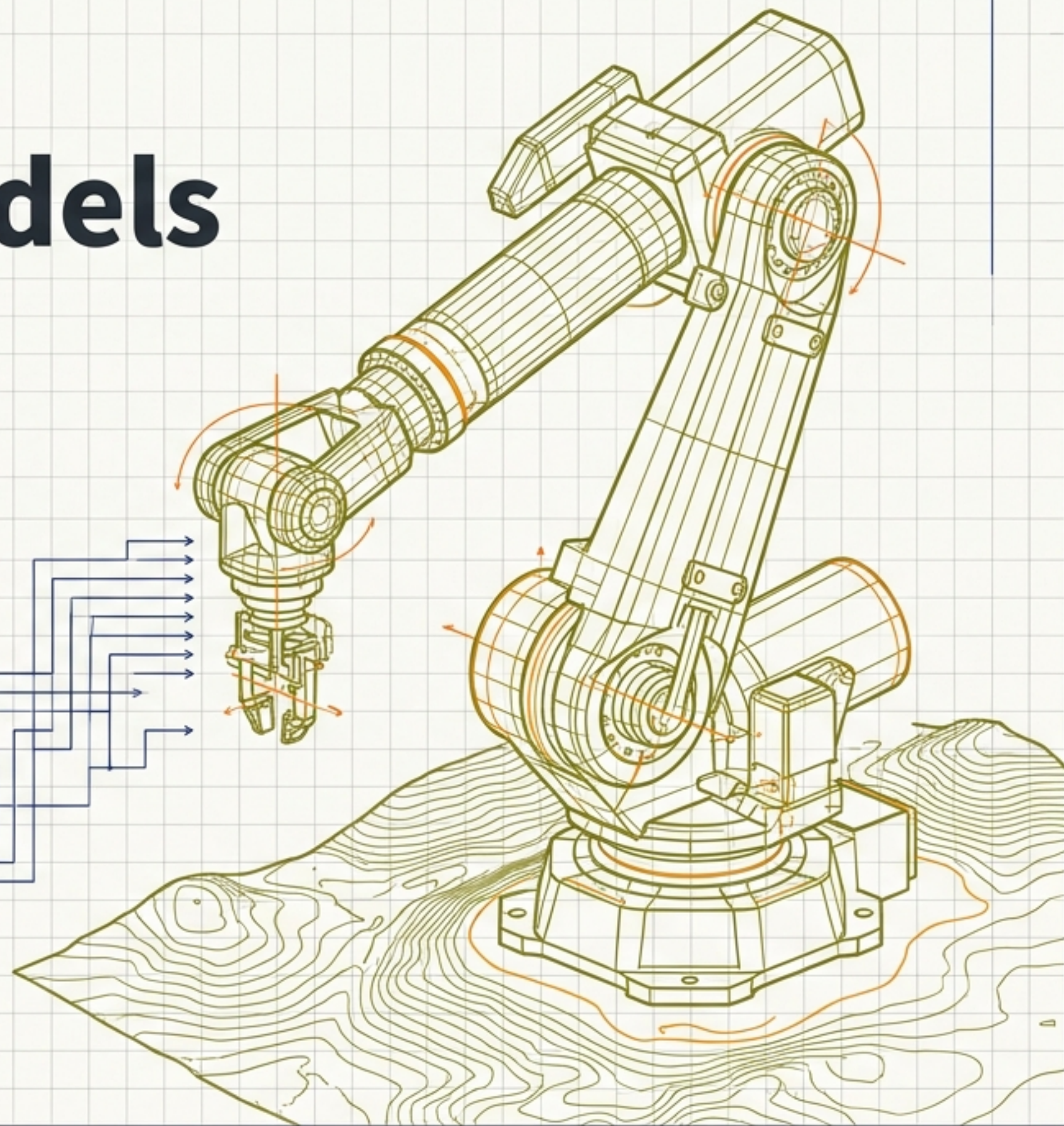
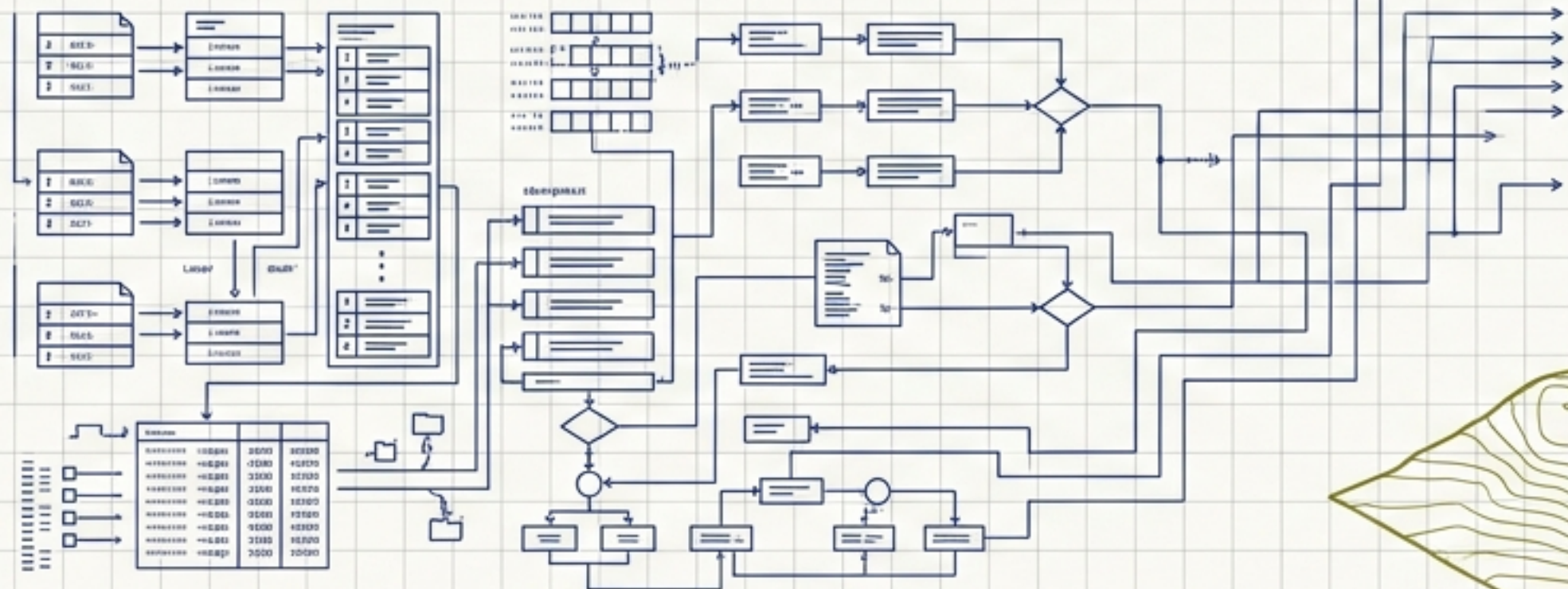


具身智能与物理AI： 从VLA到World Models 的范式演进

第二讲 - Part 1



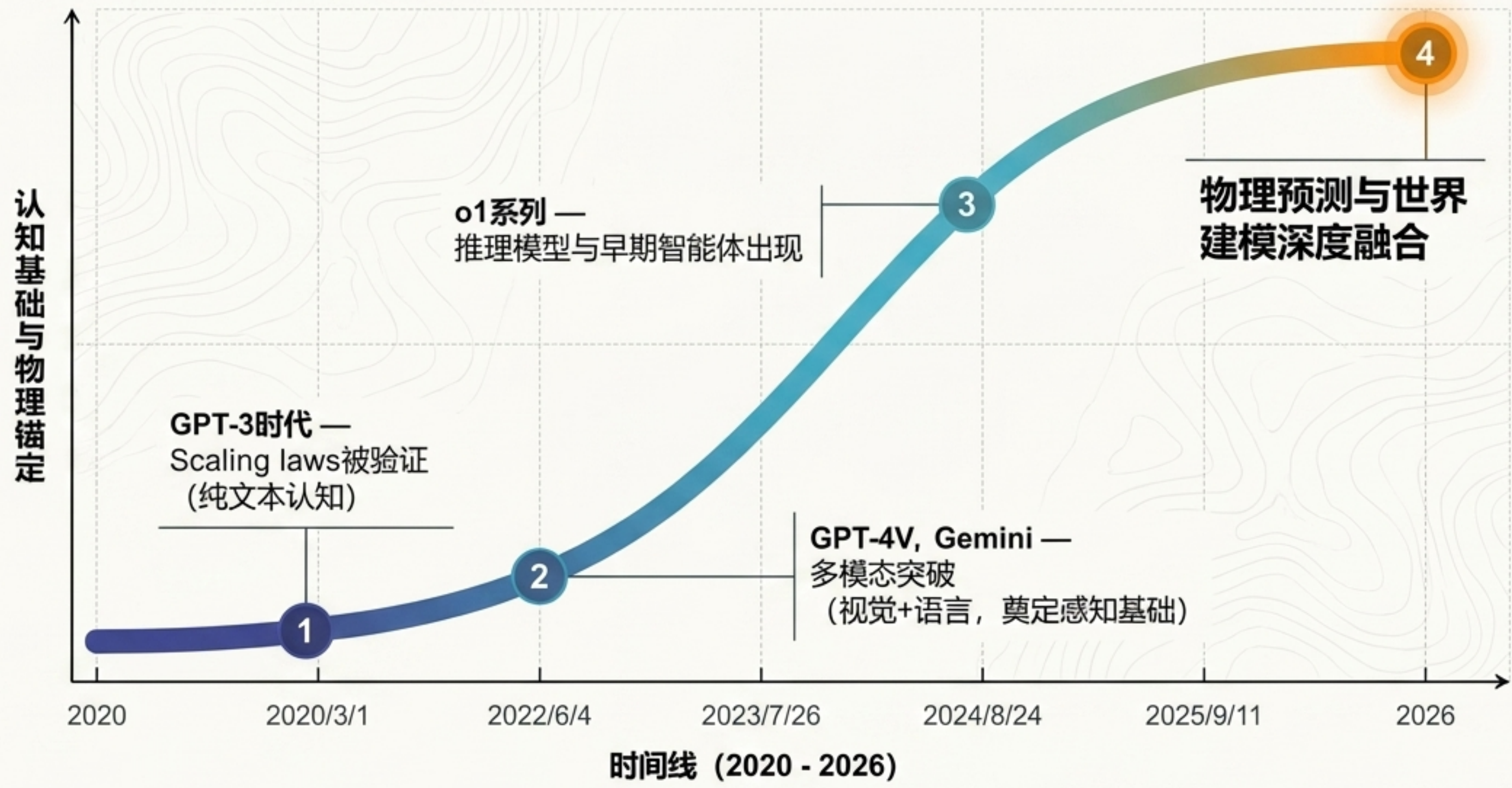
讲者档案与研究视界



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

研究方向：物理AI、具身智能与AI智能体系统



多模态模型让AI通过图像和视频“看见”物理世界，但这仅仅是具身智能的基础。

VLA模型：当前的工程标准



语言理解与任务拆解：极强。能够精准将复杂指令转化为子任务。



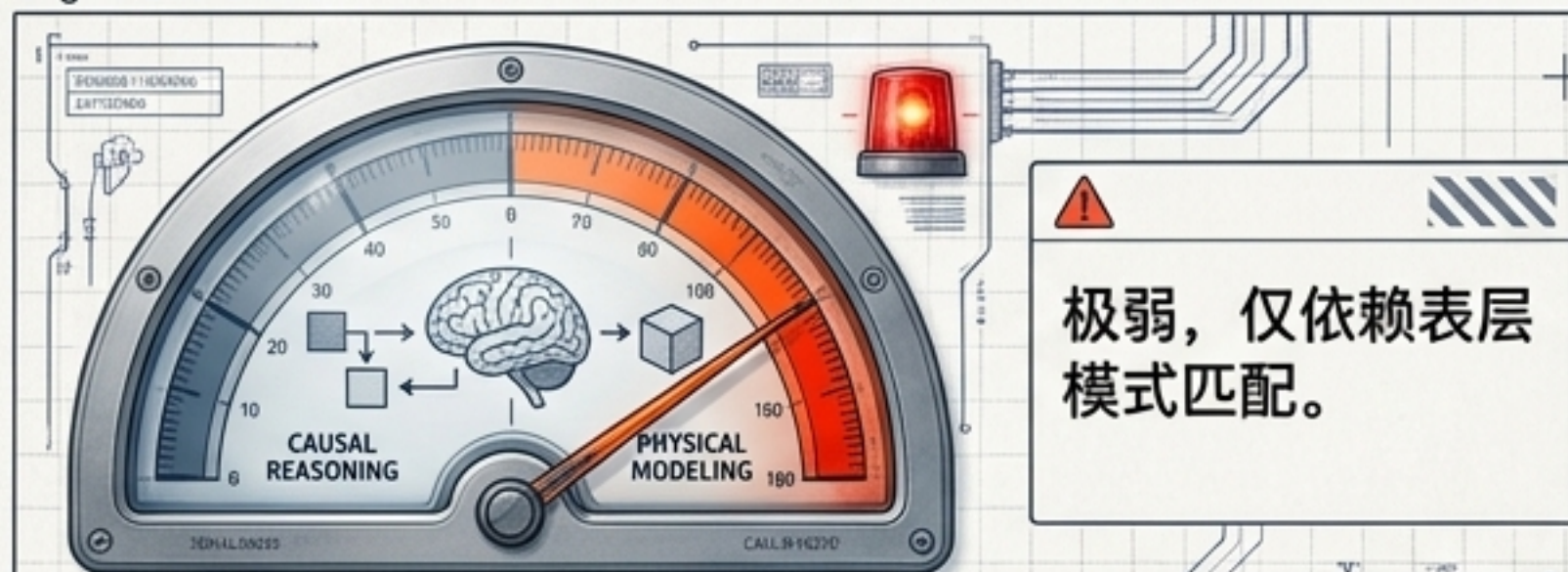
示教数据驱动：在拥有充足、高质量的机器人轨迹数据时表现优异。



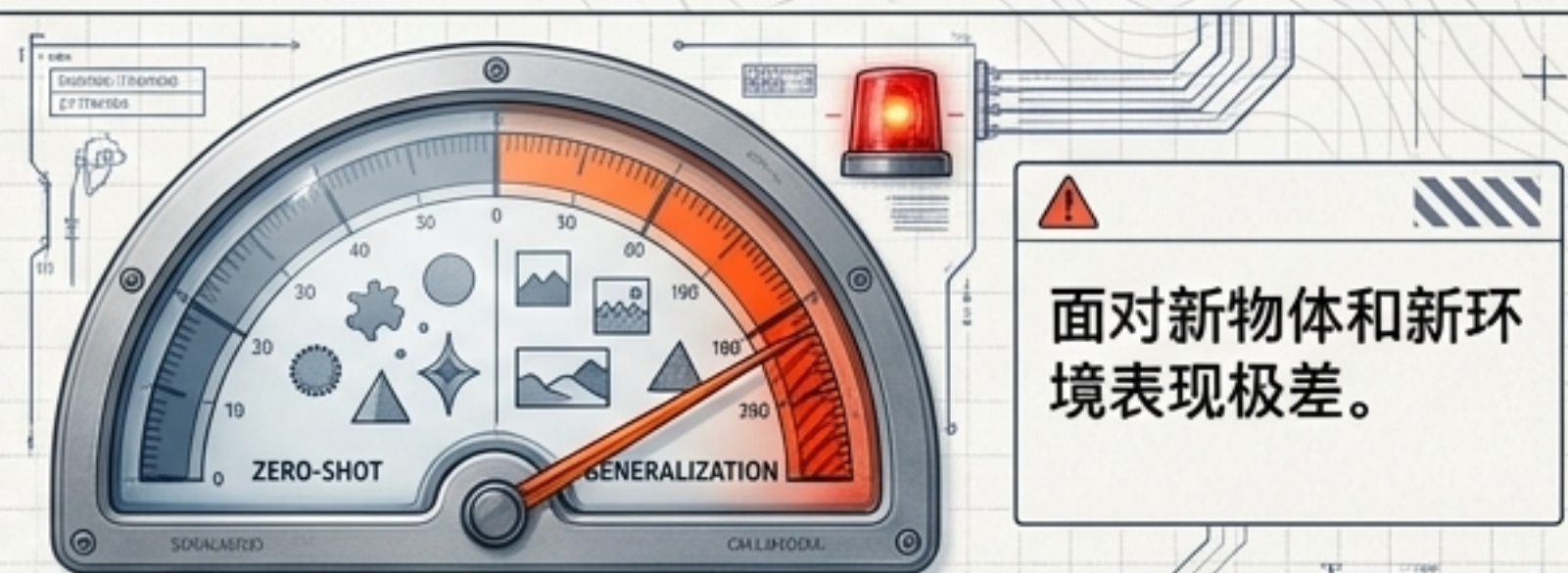
部署准备度：拥有相对成熟的工程流程与管线。

物理环境中的VLA瓶颈（2026视角）

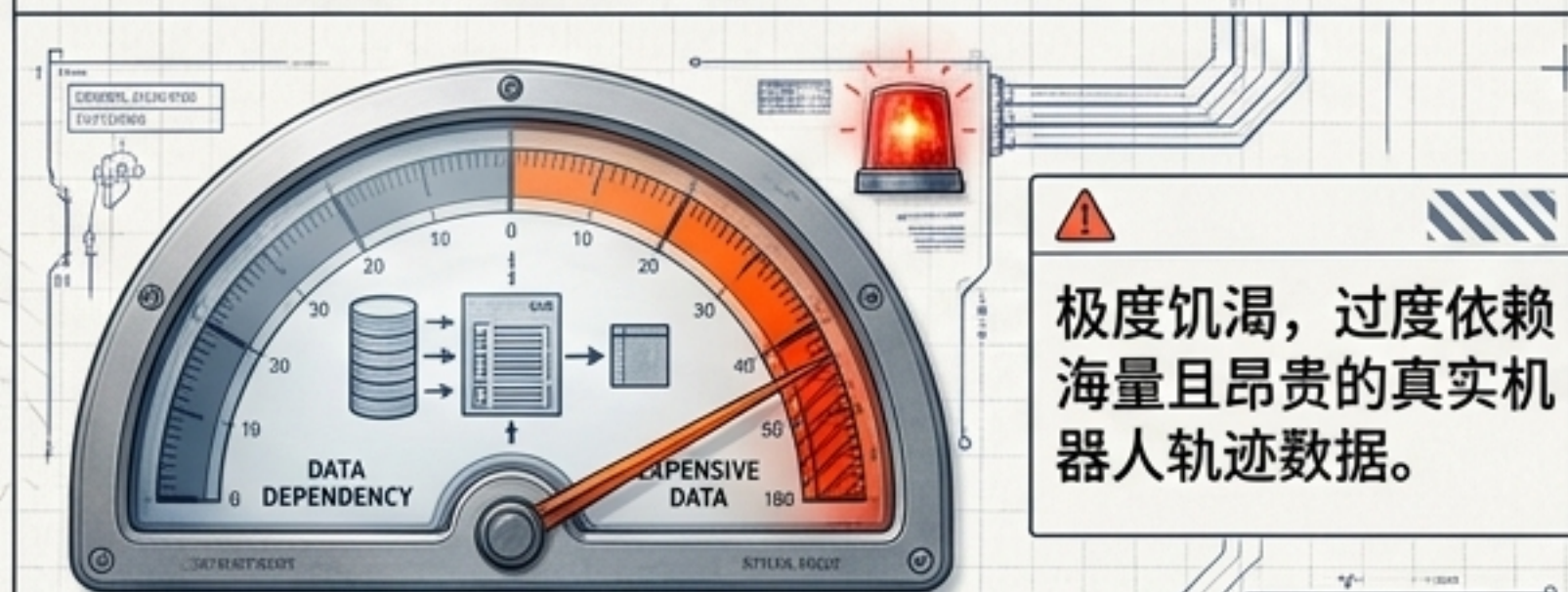
Engine Dashboard



物理推理与因果理解



零样本泛化能力



数据依赖度



长时序规划

结论：VLA的局限性源于架构本身，无法单纯通过Scaling定律解决。

过去范式
直接学习动作

**物理AI必须在
行动前预演未来。**

2026 新范式
先学习理解和预测物理世界

2026 新范式
先学习理解和预测物理世界

World Models: 构建内化的物理引擎

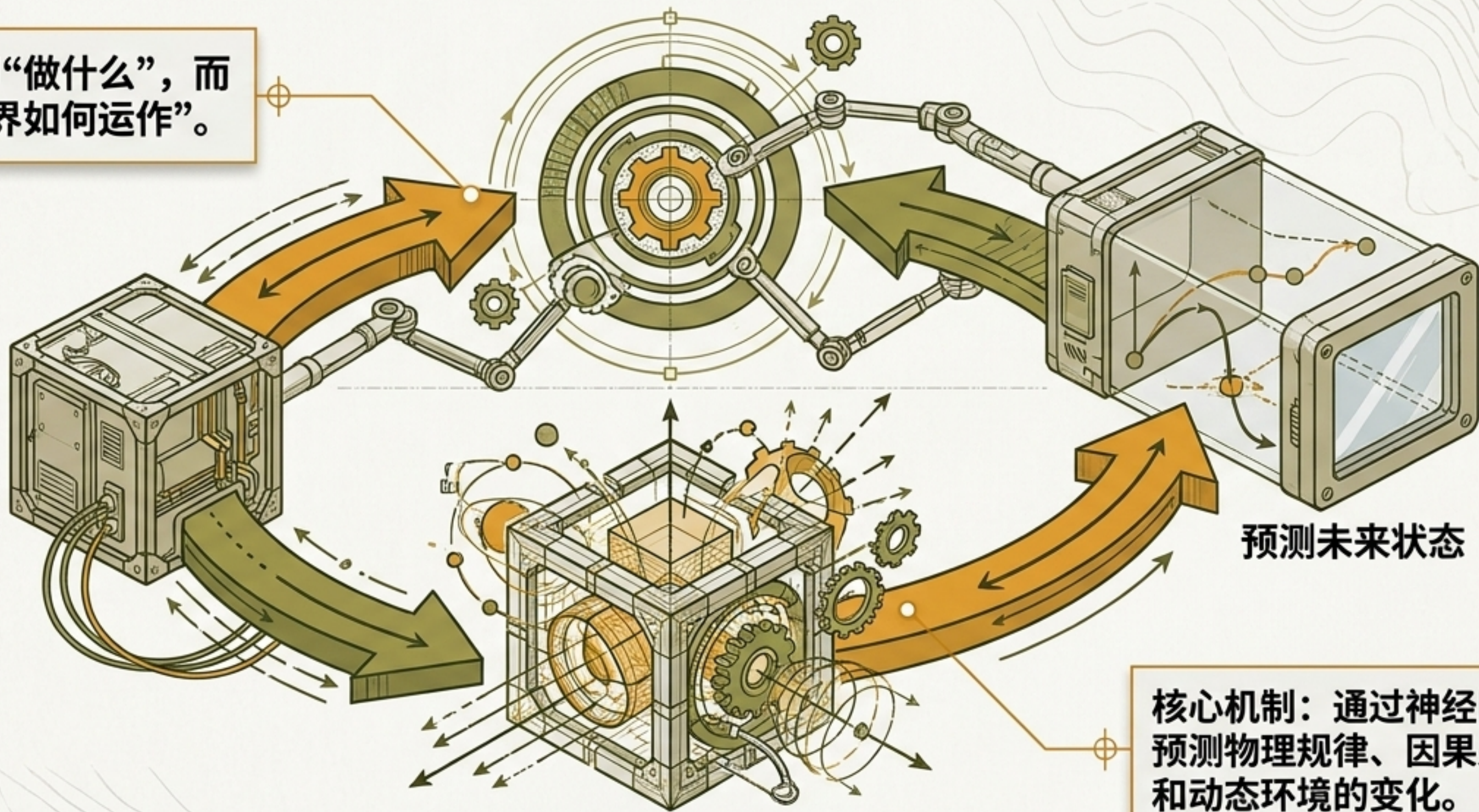
不只是学习“做什么”，而是学习“世界如何运作”。

当前观测
+
潜在动作

预测未来状态

物理与因果模拟器

核心机制：通过神经网络预测物理规律、因果关系和动态环境的变化。



World Action Models (WAMs) : 预测与生成的统一

整合于同一模型架构内。

动作生成执行

世界状态预测

泛化跃升

统一架构大幅提升跨场景、跨物体的泛化能力。

数据降维

显著降低对大规模、高成本机器人示教数据的依赖。

架构演进对比矩阵

	VLA (视觉-语言-动作)	World Models (世界模型)	WAMs (世界动作模型)
核心机制	视觉语言直出动作	预测未来物理状态	状态预测与动作生成结合
物理因果理解	弱（黑盒映射）	极强（内化物理规律）	强（基于物理预演）
数据依赖	极高（依赖海量动作轨迹）	依赖视频与状态数据（无需纯动作）	低（大幅降低动作数据需求）

行业验证：WAIC 2026 性能跃升

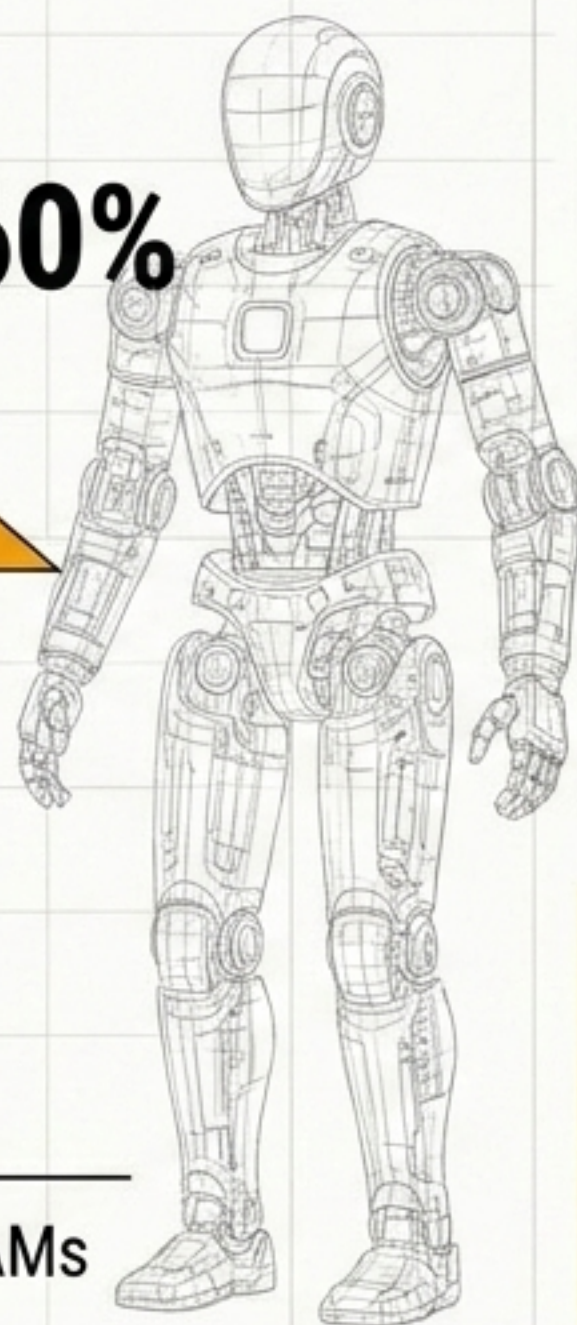
泛化基准测试

+30% to +60%

1.0x

VLA Baseline

World Models / WAMs

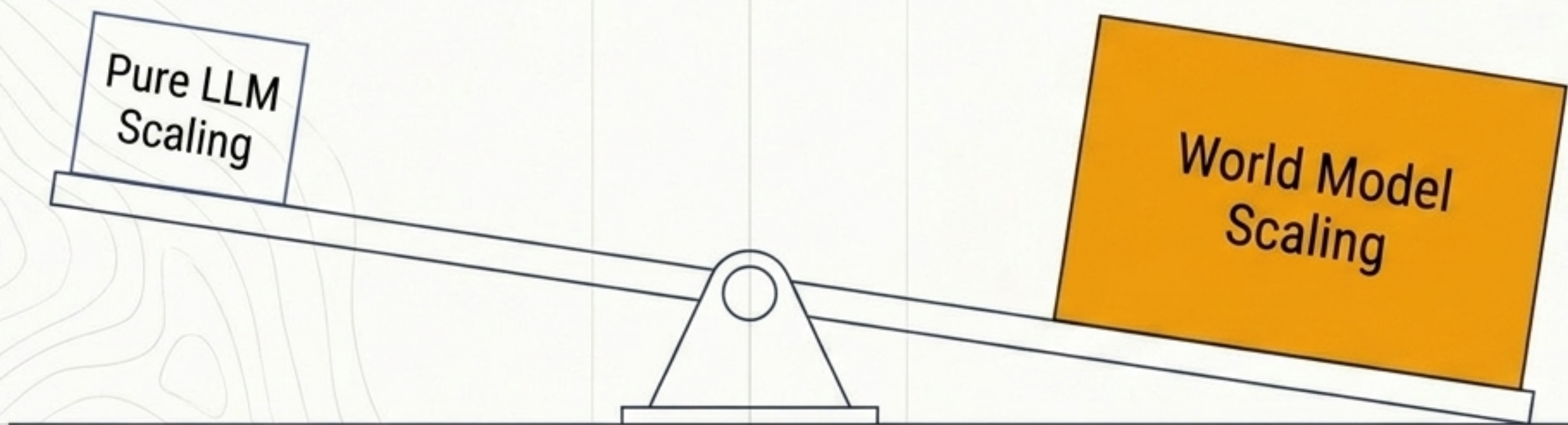


多支顶尖研究团队在2026年世界人工智能大会（WAIC）上实测证实：WAMs在泛化能力上彻底碾压传统VLA。

落地现状：多家人形机器人平台已正式将世界建模技术应用于真实物理环境部署。

AGI-2026 的核心共识

“在物理环境中实现通用人工智能（AGI），
扩展 World Models（世界模型）比单纯
扩展大语言模型（LLMs）更为关键。”

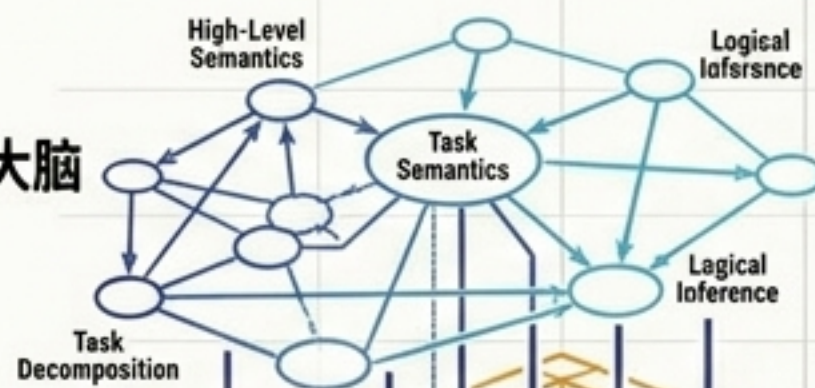


2026 终极形态：混合架构的诞生

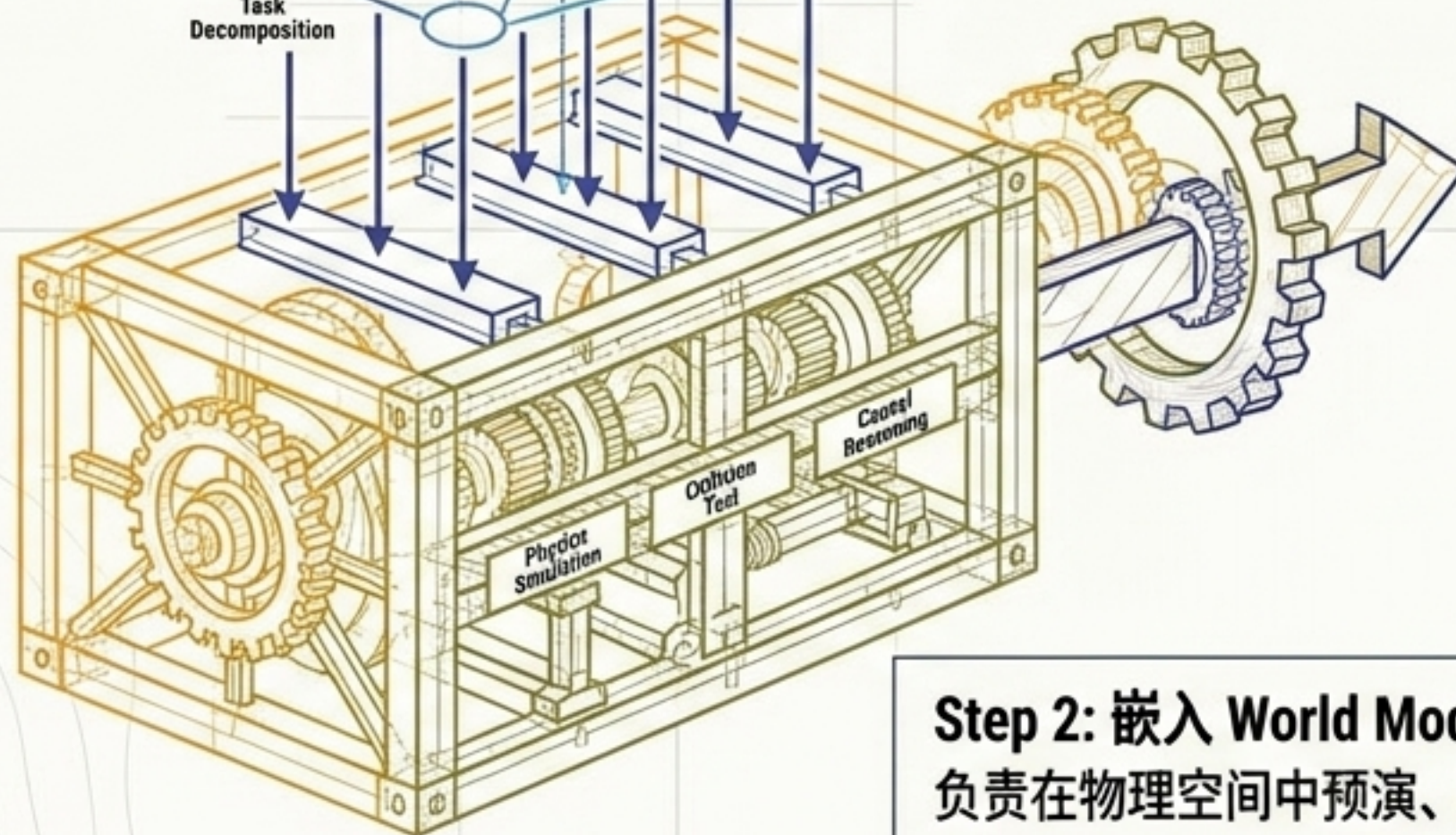
Step 1: VLA 推理大脑

负责高层语义理解与任务拆解。

VLA 推理大脑



World Models
预测引擎



WAMs 动作输出

Step 3: WAMs 动作输出
执行经过物理定律验证的
最优解。

Step 2: 嵌入 World Models 预测引擎

负责在物理空间中预演、碰撞测试和因果推演。

VLA与World Models并非非此即彼。领域正在不可逆转地走向混合架构，融合逻辑推理与物理直觉。

从认知到落地：物理AI的新纪元

Part 1 核心要点

◆ LLM与多模态构筑了感知底座。

◆ VLA触及了物理理解与泛化的天花板。

◆ World Models 与 WAMs 引发了2026年的根本性范式转变。

**NEXT STEP:
2026年的重大
技术突破与真
实世界机器人
部署。**

我们将探讨这些“物理大脑”如何驱动具身实体进入真实工业与生活场景。