

从VLA到World Models的范式演进

具身智能与物理AI第二讲 (Part 1)

Michael HUO

北美高校联盟 AI 创业俱乐部主席 / 唯爱AI基金会主席

2026年7月18日

讲者视角：2026年正在发生的认知重构



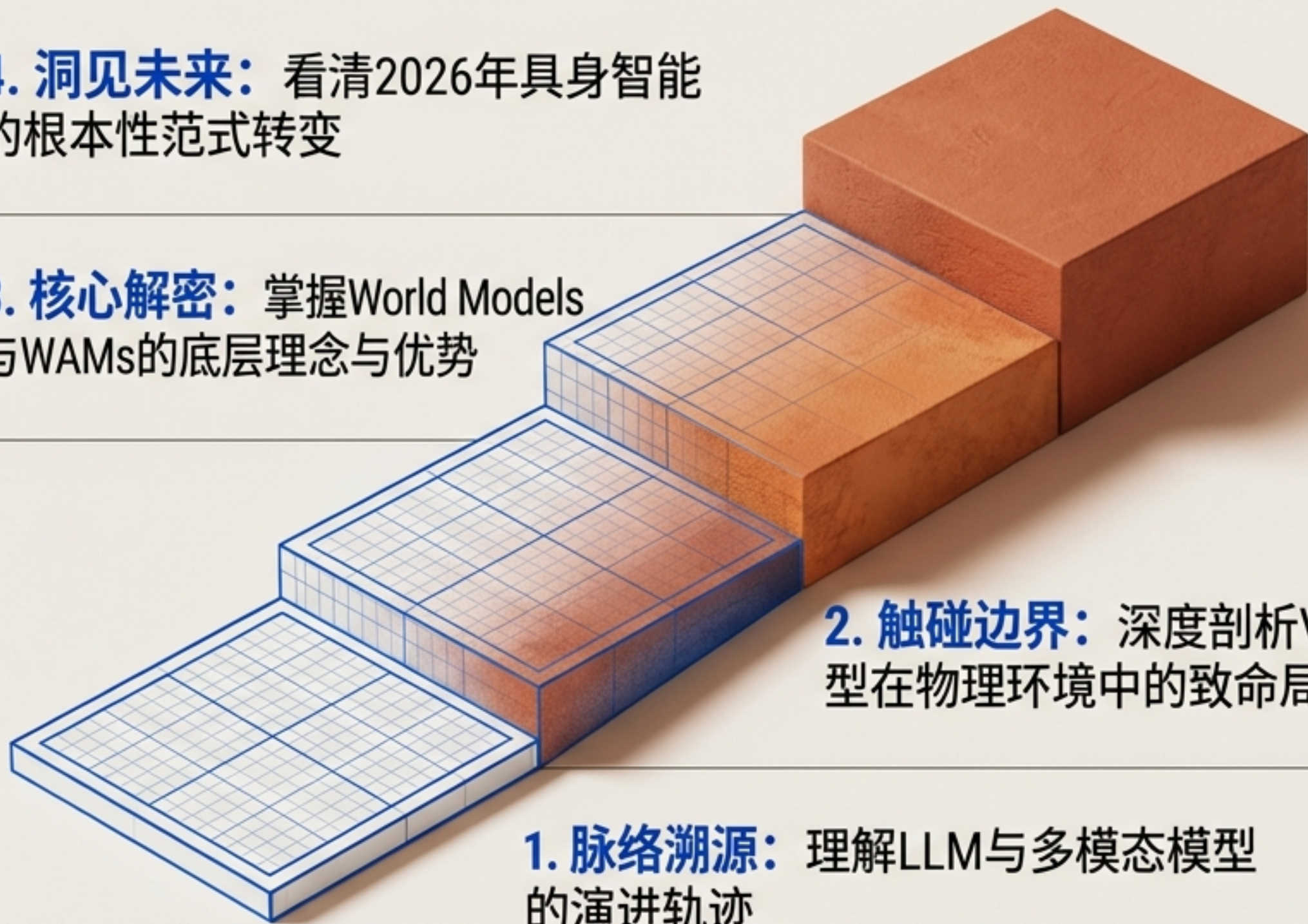
专注领域：物理AI、具身智能与AI智能体系统
(北卡浙大校友会创会理事副会长)

4. 洞见未来：看清2026年具身智能的根本性范式转变

3. 核心解密：掌握World Models与WAMs的底层理念与优势

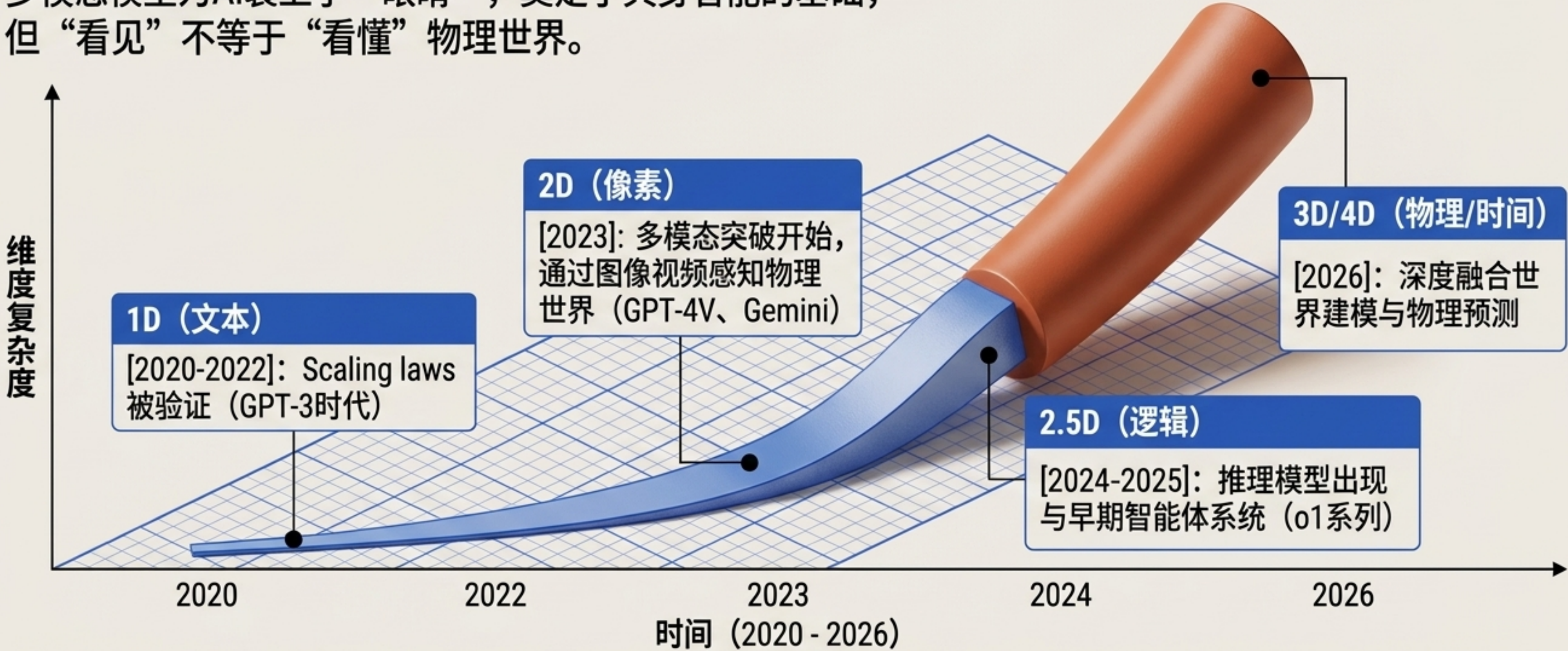
2. 触碰边界：深度剖析VLA模型在物理环境中的致命局限

1. 脉络溯源：理解LLM与多模态模型的演进轨迹



维度跃升：从文本理解到物理预测的AI进化史

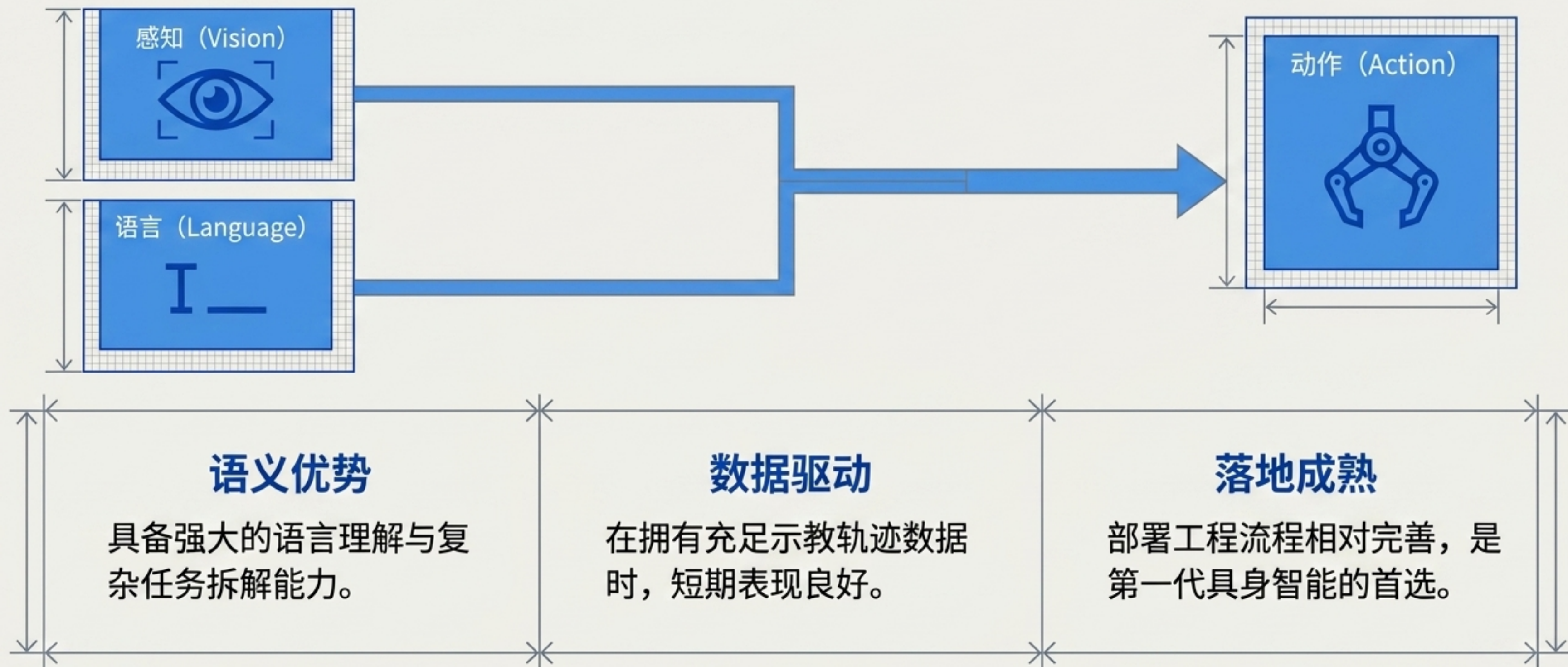
多模态模型为AI装上了“眼睛”，奠定了具身智能的基础，但“看见”不等于“看懂”物理世界。



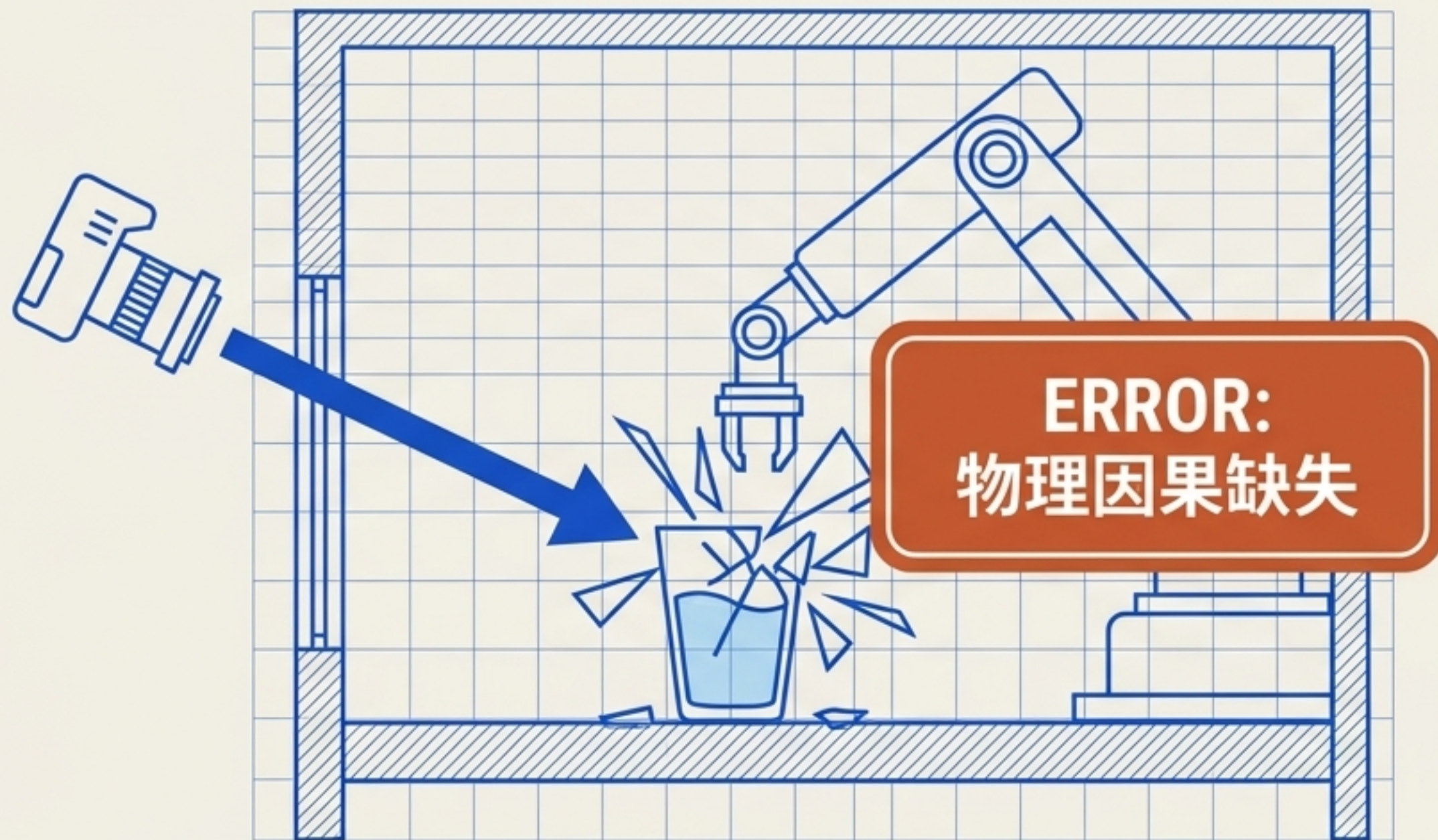
核心结论：AI正经历从数字世界的观察者到物理世界的参与者的维度质变。

VLA架构：多模态在具身领域的工程胜利

将感知（Vision）和语言（Language）直接映射为机器人动作（Action）。



物理世界的残酷现实：VLA的“短路”效应



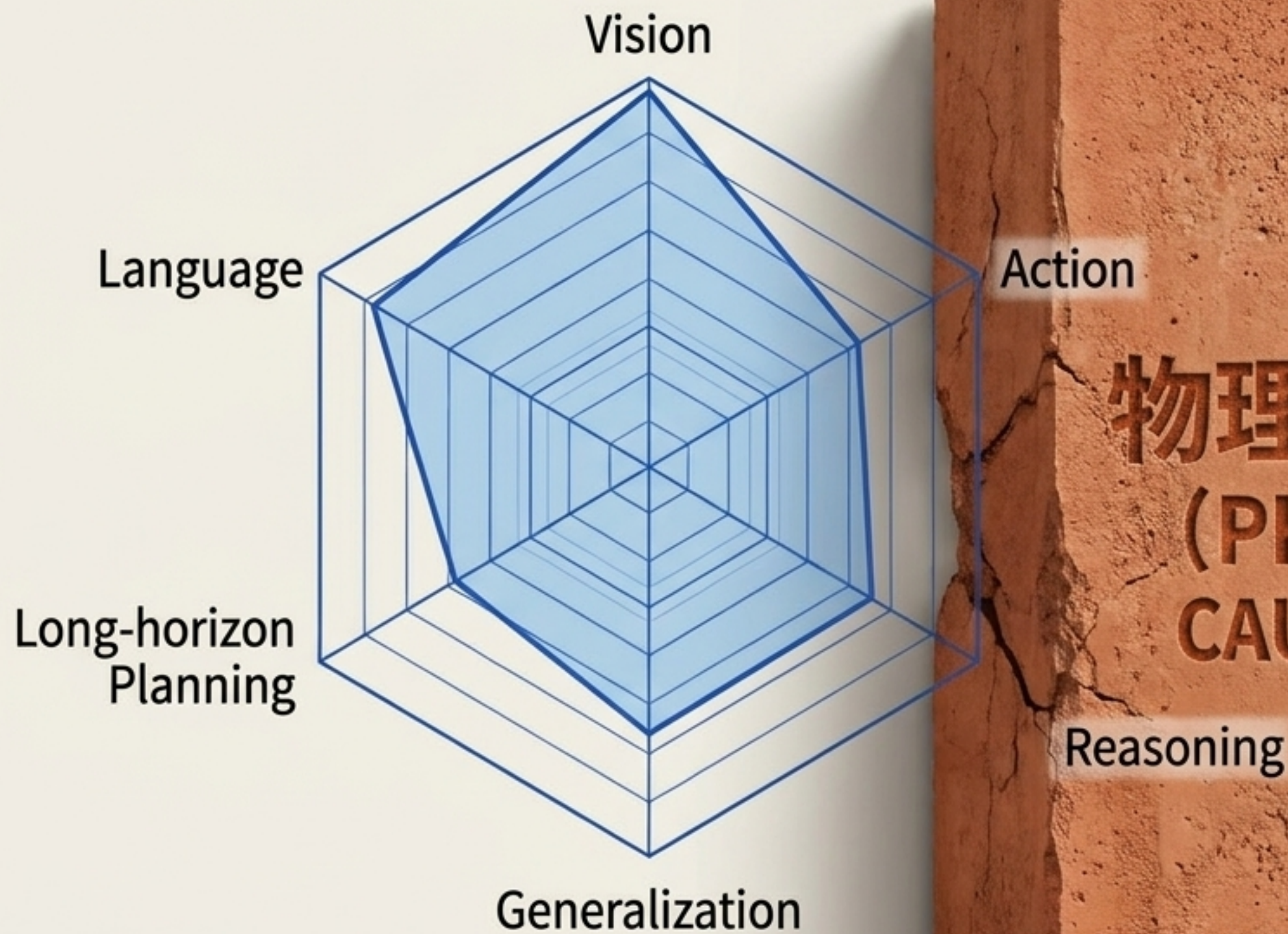
VLA仅仅知道“当前图像匹配训练集中的抓取动作”，但它不知道玻璃是易碎的、表面是湿滑的。它没有重力、摩擦力或因果关系的内在概念。

$$f(\text{See, Hear}) = \text{Move}$$

致命盲区：

VLA的底层逻辑（黑盒映射）：VLA仅仅知道“当前图像匹配训练集中的抓取动作”，但它不知道玻璃是易碎的、表面是湿滑的。它没有重力、摩擦力或因果关系的内在概念。

VLA的性能天花板（2026年视角）



物理因果律 (PHYSICAL CAUSALITY)

1. 物理推理弱

缺乏对因果关系和基本物理定律的理解。

2. 泛化能力差

面对新物体、新材质、新环境时极易失效。

3. 极度依赖数据

沦为“数据黑洞”，需要海量且昂贵的机器人轨迹示教。

4. 短视规划

缺乏长时序任务（Long-horizon）的预判与规划能力。

认知重构：2026具身智能的根本性范式转变

直接学习动作 (Learning to act directly)

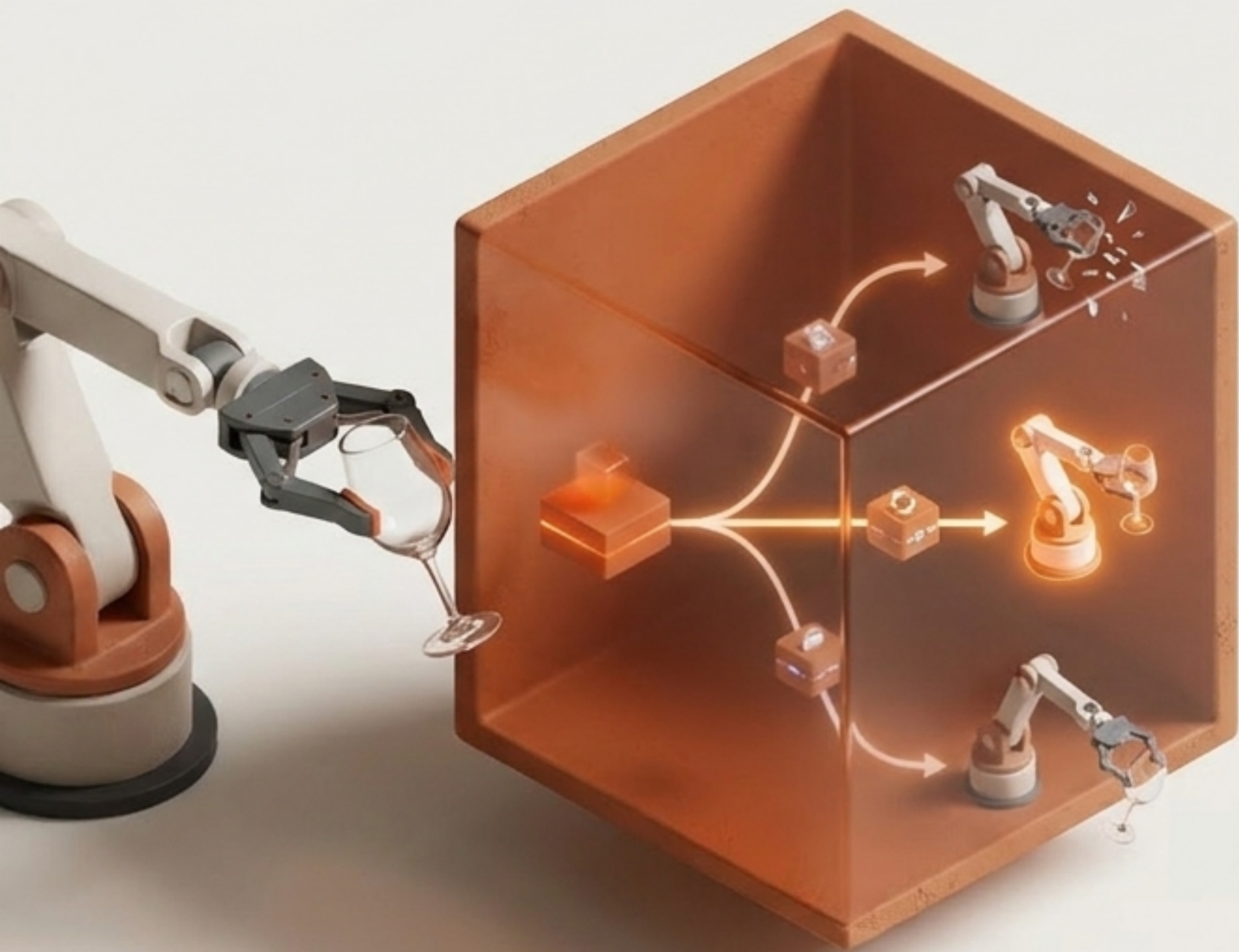
$f(\text{Observation}) \rightarrow \text{Action}$

先理解并预测物理世界 (Learning to understand the world first)

$f(\text{Observation}, \text{Potential Action}) \rightarrow \text{Predict Next State}$

业界已达成共识——实现物理世界通用智能的关键，不再是“预测下一个动作”，而是“预测下一个物理状态”。

什么是 World Models?



用力过猛
→ 玻璃碎裂

动态适中
→ 稳定抓取

抓取太松
→ 玻璃滑落

预测未来状态

基于当前观测和潜在动作，预测物理环境的未来走向。

学习物理内核

核心目标是内化物理规律、因果关系与动态变化，而非死记硬背动作轨迹。

“三思而后行”

在神经网络内部进行物理沙盘推演。

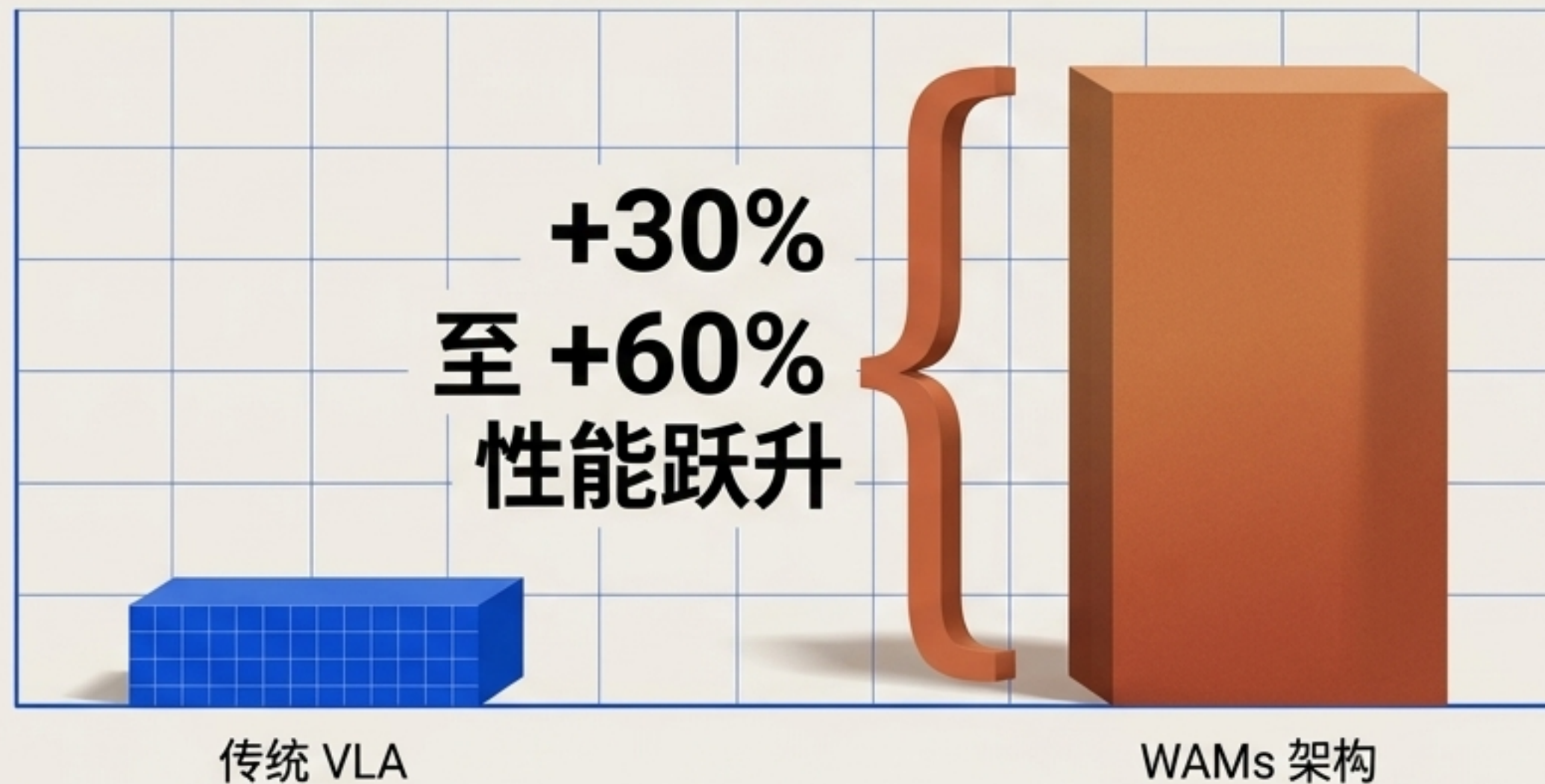
架构诊断：为何我们需要统一范式？

	VLA模型	World Models	World Action Models / WAMs
核心目标	直接预测动作	预测未来物理状态	预测世界 + 生成动作
物理因果理解	✗ 弱	✓ 强	✓ 强
动作直接生成	✓ 是	✗ 否（需外接策略）	✓ 是
泛化能力	✗ 差	✓ 强	✓ 极强（统一架构）

WAMs 将世界预测与动作生成整合在同一模型中，补齐了最后一块拼图。

2026经验验证：从理论走向规模化部署

泛化能力基准测试 (Generalization Benchmark)



WAIC 2026 最新进展展示，WAMs 大幅提升泛化能力，并显著减少对大规模真实机器人数据的依赖。多家人形机器人平台已经将世界建模技术应用于实际部署。

AGI-2026 核心论断

“在物理环境中实现通用智能，扩展 World Models (世界模型) 可能比单纯扩展大语言谄‘模型更为关键。”

新的蓝图：混合架构主导的未来

- LLM与多模态奠定认知基石。
- VLA触碰物理边界与数据天花板。
- World Models 引领物理预测的根本性转变。
- 混合架构（WAMs）正以惊人速度重塑行业。



Part 2 预告

理论已经成型，现实世界表现如何？

下期揭秘：2026年重大技术突破与物理AI机器人的实际部署案例。敬请期待。