

跨越奇点： 世界模型如何重塑具身 智能产业与商业新范式

——从生成式导向到控制导向的Physical AI 基础设施演进



目录

01 中国具身智能产业全景与市场洞察

02 从具身智能到世界模型： 核心内涵、技术演进与架构解析

03 国际主流世界模型评估： 核心指标体系与榜单解读

04 世界模型在具身智能领域的产业落地实践

05 未来趋势与生态演进

01

中国具身智能产业全景 与市场洞察

1.1 政策环境：顶层设计与地方产业集群协同共振

中央层面：确立战略坐标

1.提供确定性背书

- 中央政策的核心价值在于消除产业前期的方向不确定性，《“十五五”规划纲要》首次系统性描述，**具身机器人和人形机器人定位**从科研课题跃迁为具备国家竞争属性的战略资产。

2.锚定通用终局目标

- 引导产业向底层能力纵深攻关。**依托国家重点研发计划与“揭榜挂帅”机制，重点突破操作系统、仿真平台、世界模型等基础前沿技术，迫使资源向战略底层创新倾斜，构筑自主技术底座。

3.万台级规模化验证

- 工信部与国资委联合印发《人形机器人与具身智能实景实训专项行动》，部委协同推进，十省市加央企强势入局。**推动产业重心从单纯“技术攻关”向“万台级规模化验证”**发生根本转变，加速全栈能力构建。

国家定方向，地方建生态



地方层面：生态卡位与进化

1.搭平台：建创新载体

- 政府职能升级为生态架构者。**京沪等多地建设省级创新中心、检测验证场与数据训练场，提供技术验证公共平台，大幅降低企业从实验室到工厂的验证成本。

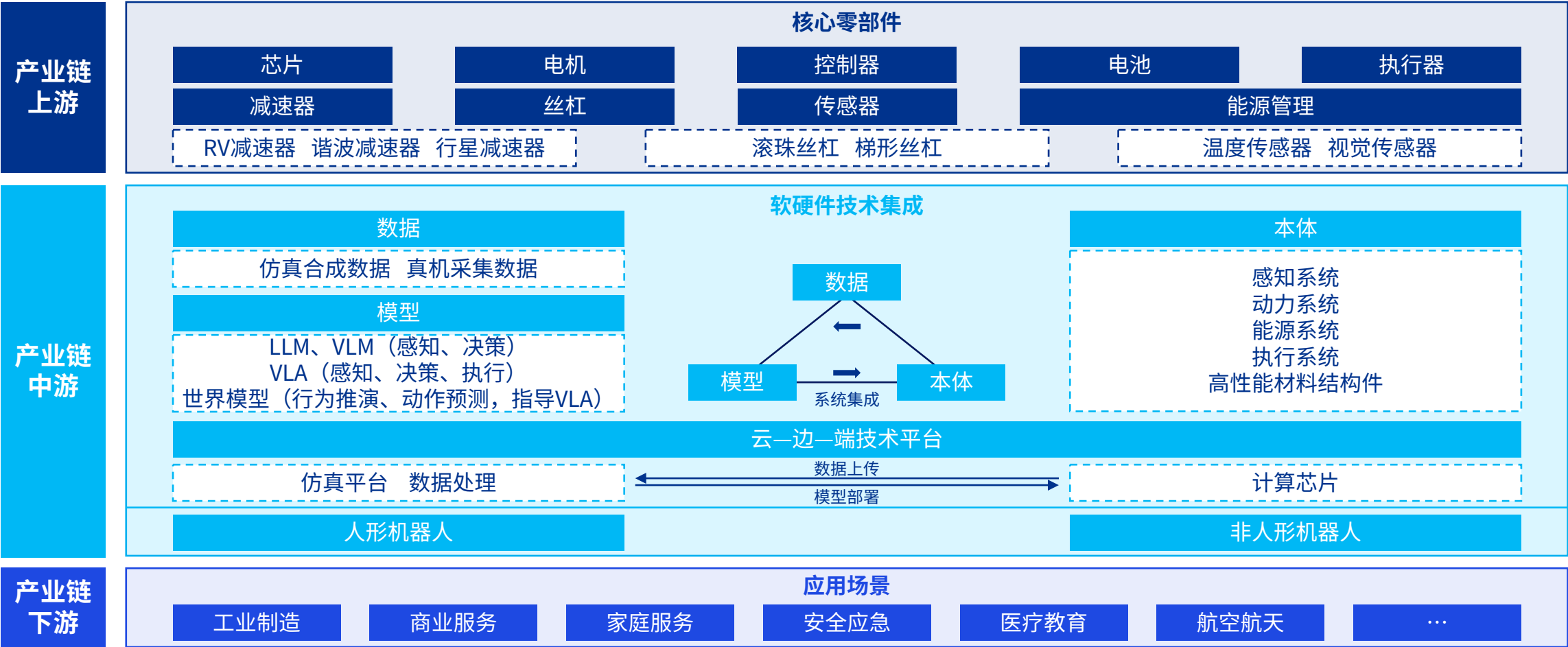
2.供场景：做首批客户

- 地方政府转变财政补贴模式，主动扮演首批“战略客户”。**面向企业开放市政巡逻、物流配送等真实、复杂的测试应用场景，提供商业化初期关键助力，加速技术落地与商业闭环形成。

3.定规则：掌握标准话语权

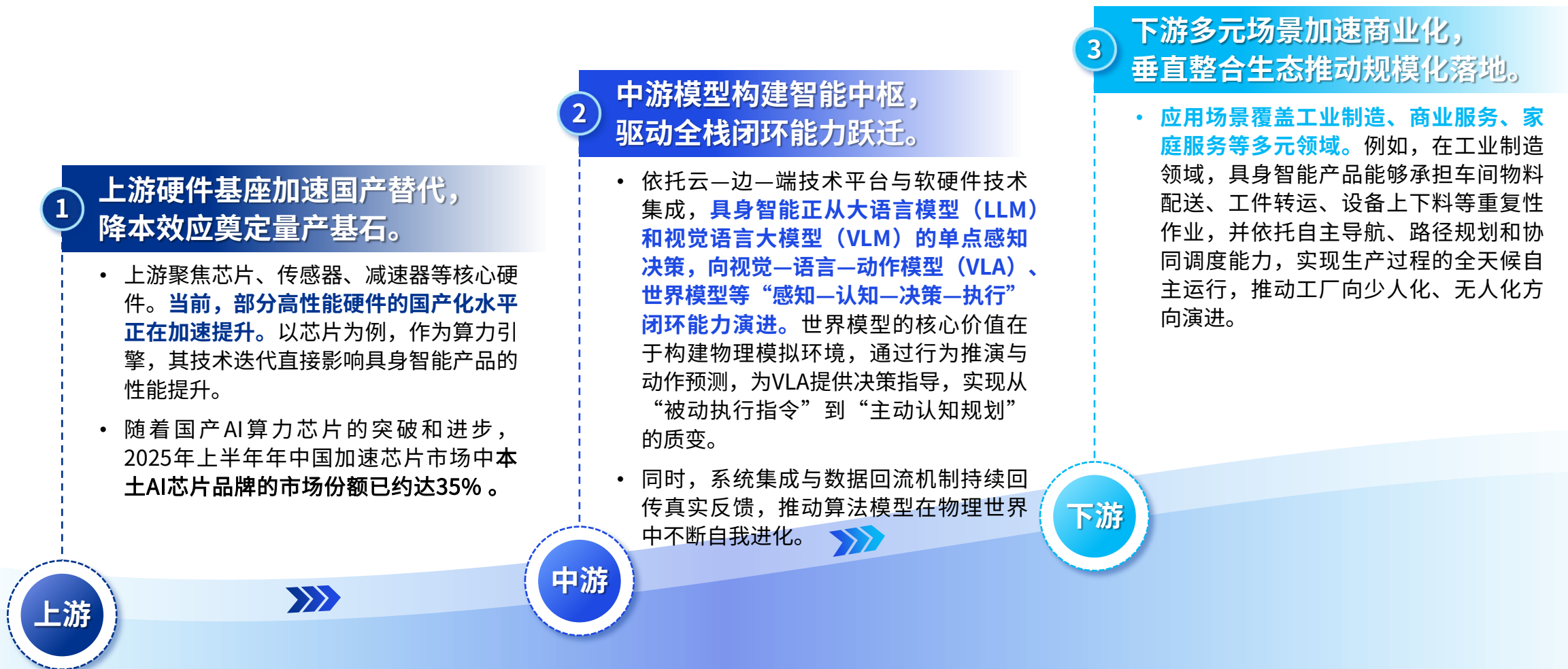
- 北京抢先启动人形机器人智能化分级、运动控制等团体标准研制，试图在产业成形前于上游规则层面建立先发优势，这意味着**产业的底层定义权**将更多来自中国场景的实践积累。

1.2 产业链分析：中国具身智能上下游生态图谱



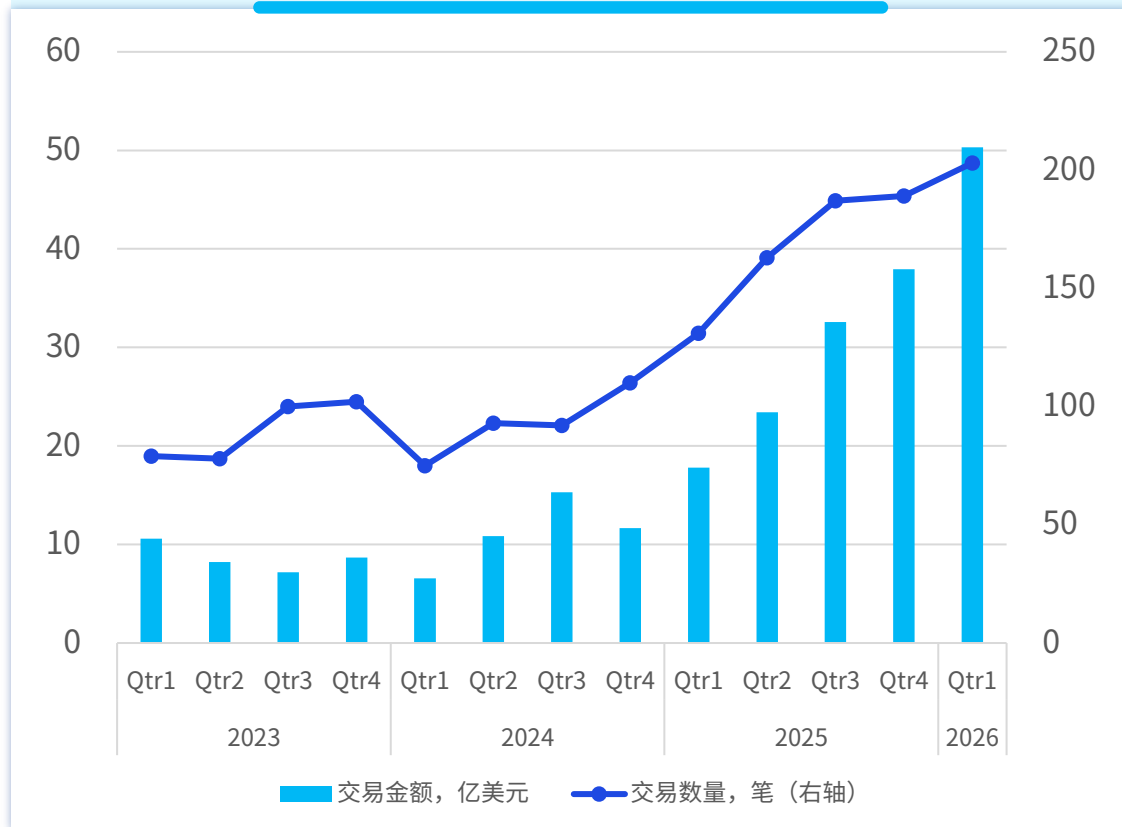
资料来源：信通院，毕马威分析

1.2 产业链分析：中国具身智能上下游生态图谱



1.3 一级市场投融资分析：资本热潮——近三年中国具身智能投融资趋势（1/3）

具身智能产业PE/VC投资额和数量，季度



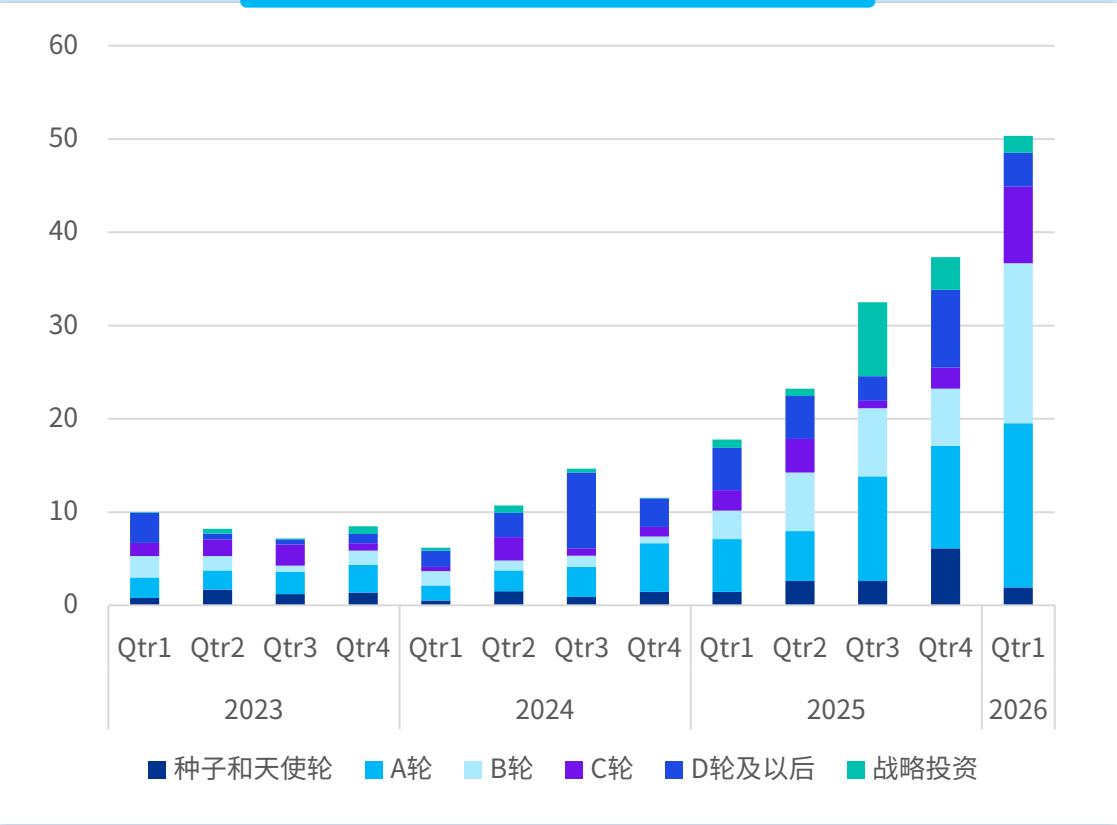
数据来源：投中数据，毕马威分析

具身智能作为大模型应用的主要发展方向，在2025年以来实现快速增长。具身智能2025年的融资额达111.7亿美元，同比增长**152.0%**；融资数量670笔，同比增长**81%**。2026年一季度融资额达50.3亿美元，同比增长**182.9%**；融资数量203笔，同比增长**55%**。

融资金额和数量的不断走高一方面反映出赛道的火爆，另一方面**“大模型+机器人躯体”的技术路线在此时逐步趋于收窄**。相比于软件，具身智能涉及复杂的机械硬件、传感器和产线建设，属于典型的高门槛、重资产赛道。一旦技术可行性被验证，就需要大量的资金投入，这也直接**推高了融资金额上限**。

1.3 一级市场投融资分析：资本热潮——近三年中国具身智能投融资趋势（2/3）

具身智能产业PE/VC按轮次投资额，亿美元

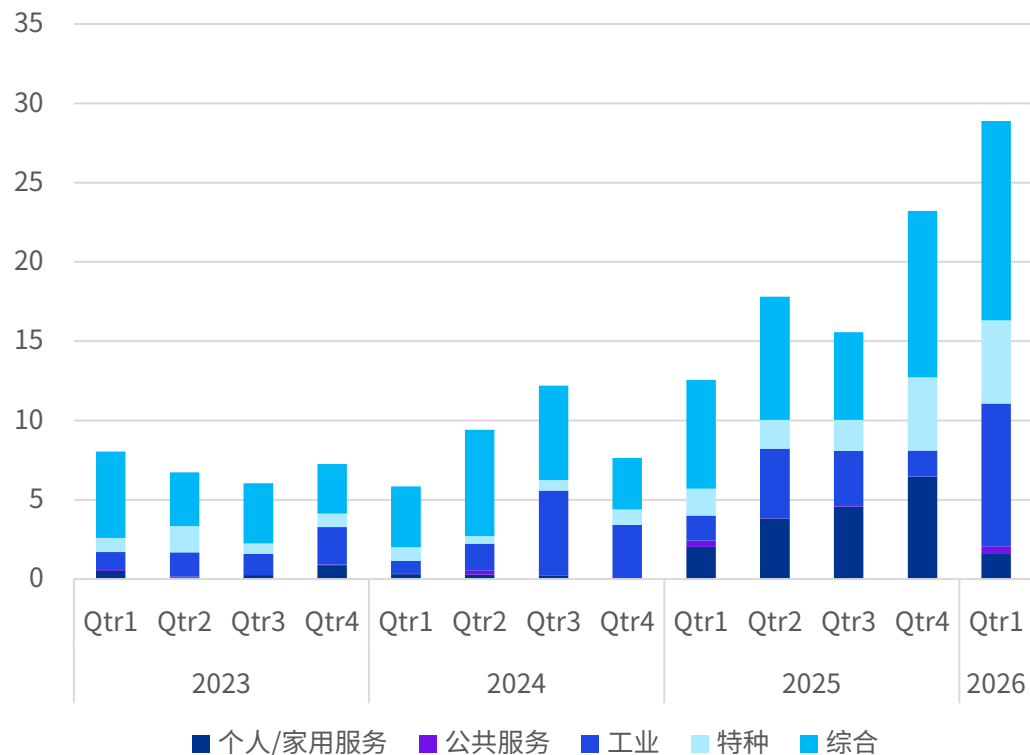


- 从投资轮次的分布来看，资本风险偏好加速前移，早期轮次融资额呈爆发式增长。具体来看，2025年种子天使轮、A轮、B轮、C轮、D轮融资额分别同比增长**191.7%**、**170.8%**、**403.7%**、**86.1%**和**29.2%**，资本转而全面押注具备底层技术壁垒的早期创新主体。
- 进入2026年第一季度，资金在早期轮次的集中度达到历史高峰。种子和天使轮、A轮、B轮、C轮融资额分别同比增长**33.1%**、**211.4%**、**458.9%**、**275.3%**，而侧重成熟期的D轮及以后则出现21%的同比收缩。更值得关注的是，这四轮早期融资合计占据当季总融资额的89.2%。**投资机构“投早投小”趋势明显，资本仍在竞相争夺具有模型研发能力与差异化定义的潜力标的。**

数据来源：投中数据，毕马威分析

1.3 一级市场投融资分析：资本热潮——近三年中国具身智能投融资趋势（3/3）

具身智能产业单一应用领域PE/VC融资金额，亿美元



数据来源：投中数据，毕马威分析

- 根据下游应用场景的不同，具身智能企业主要分布在工业机器人、特种机器人、个人/家用服务机器人、公共服务机器人等赛道，而相较单一应用领域赛道，拓展下游跨领域应用的综合类企业是融资主力，以2025年为例，综合类企业融资额达**30.7**亿美元，占下游应用总融资额的**44%**，远高于其他单一应用领域企业。
- 在单一应用领域赛道，工业及特种机器人占主导地位，2025年融资额分别为**11.1**和**10.1**亿美元，分别占比**16%**和**15%**。值得一提的是，家用服务机器人在2025年投资增长迅速，融资额达16.9亿美元，较2024年大幅增长**19**倍，但2026年一季度有所回落，融资额仅1.6亿美元，**远低于工业机器人9.0亿和特种机器人5.2亿美元的融资额。**

1.3 一级市场投融资分析：世界模型赛道狂飙，初创企业估值暴涨（1/2）

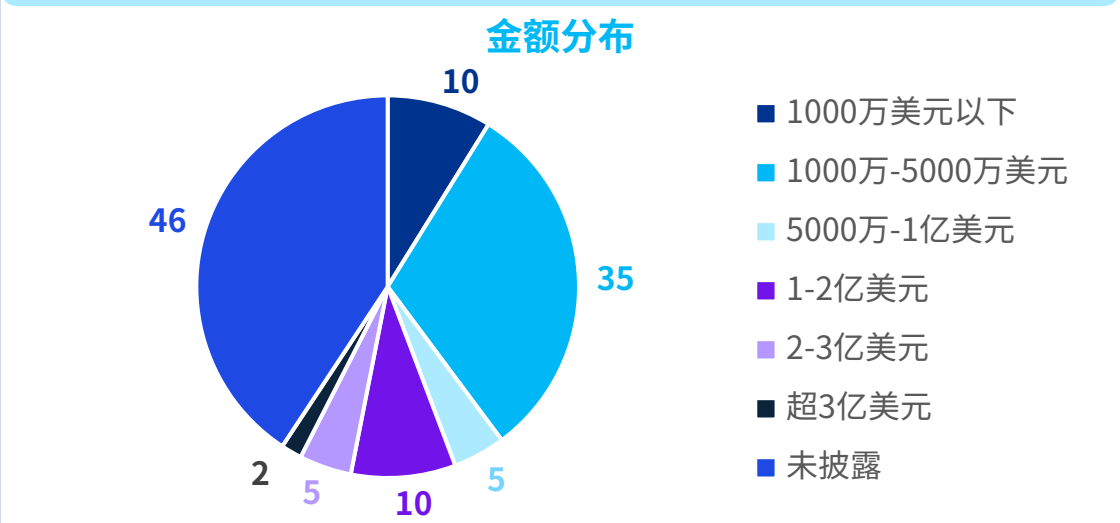
世界模型赛道企业融资笔数连续五个季度攀升



- 2025下半年以来，企业融资节奏明显加快
- 2026年上半年融资事件总数达**42**笔，已基本和2025年全年融资事件数打平。

• 基于世界模型赛道33家重点企业融资数据展开分析，数据来源为天眼查。需要注意的是，当前公开数据中“世界模型”仍未形成完全统一的投融资分类，相关企业多分布在具身智能、机器人通用大脑、空间智能、仿真数据和多模态模型等相邻口径下。

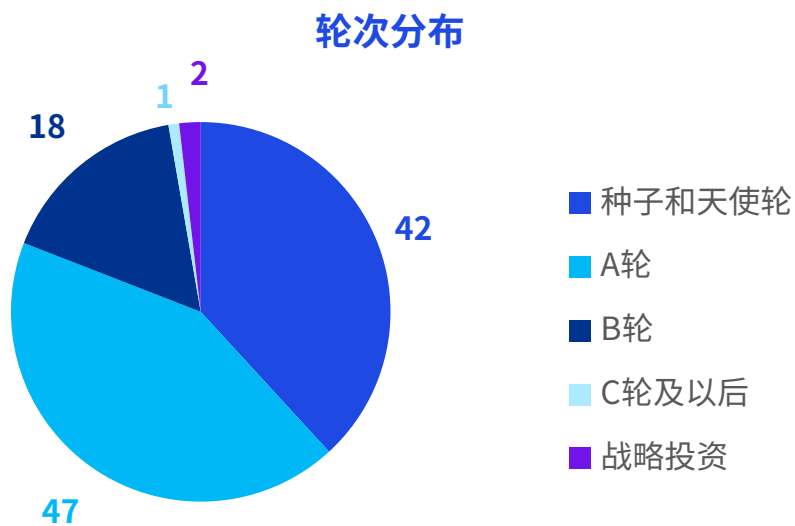
大额融资频发，迎来数亿美元级事件集中爆发期



- 已披露金额的融资事件中，**85%**的规模超千万级美元
- 规模在1亿美元以上的事件数达到**17**笔，其中**13**笔均发生在2026年上半年

1.3 一级市场投融资分析：世界模型赛道狂飙，初创企业估值暴涨（2/2）

早期即重注：早期轮次占比高且大额化趋势明显



- 种子轮和天使轮、A轮的融资事件数达89笔，占比达80%。
- 1亿美元以上的17笔交易中，种子和天使轮占3笔，A轮融资占5笔，B轮融资占9笔

早期即高估值：部分企业在早期轮次即获得独角兽级定价

企业名称	成立时间	最近一轮融资轮次 (截至26.06.30)	对应估值	交易时间
卜拉格	2026-05	天使轮	20亿美元	2026-06
大晓机器人	2025-07	天使+	10亿美元	2026-06
极佳视界	2023-01	B+轮	20亿美元	2026-06
生数科技	2023-03	B轮	18亿美元	2026-04
星海图	2023-09	B+轮	30亿美元	2026-04
银河通用	2023-05	B++轮	30亿美元	2026-03

中国世界模型融资中的两种早期高估值路径：

- 依靠明星模型团队和范式想象获得资本提前定价，以卜拉格为代表。
- 依托具身大脑、产业资源和工程化路径快速进入独角兽行列，以大晓机器人为代表。

1.4 产业痛点：具身智能产业化面临四大核心瓶颈，世界模型是破局关键（1/2）

数据瓶颈：

真实机器人数据难以支撑规模化训练

- 真实机器人**数据采集成本高、场景覆盖窄**，难以穷尽不同物体、动作、环境和机器人形态。
- **开放世界视频、人类行为与机器人数据虽分别提供了物理规律、任务逻辑和动作锚定，但三者天然不对齐。**例如同样是“拿杯子”，视频呈现的是画面，人类数据记录的是关节配合，而机器人数据则是具体的机械臂角度与力度。



任务瓶颈：

短动作能力难以自然延展为长链路任务能力

- 当前机器人已能完成抓取、放置、开门等相对明确的短动作，**但真实产业场景更多是连续、多步骤、强上下文依赖的长链路任务**，例如整理桌面、清洁房间、仓库分拣和工厂装配。
- 长任务要求模型持续记住物体位置、遮挡状态、任务进度和前序动作带来的后续影响。**一旦缺乏长期状态保持能力，机器人容易在多步骤执行中出现状态漂移、动作断裂或任务遗忘。**



控制瓶颈：

世界理解难以稳定转化为可执行动作

- 视觉模型可以识别环境，生成式模型可以预测未来画面，**但机器人真正需要的是判断“如果我现在这样行动，世界会如何变化”。**
- 开放世界视频通常缺少动作条件，人类示范难以直接映射到机器人执行机构，**机器人交互数据又过于稀缺和狭窄，导致模型即使能够理解场景，也未必能够形成稳定的感知—动作映射。**



部署瓶颈：

离线模型能力难以直接进入真实闭环系统

- 机器人行动具有强实时性，抓取、避障、移动和协作都需要低延迟反馈。如果模型推理速度慢、显存占用高、通信成本大，就**难以嵌入真实机器人系统，也无法在观察—行动—反馈过程中持续修正策略。**
- 具身智能要从实验室演示走向产业化部署，需要世界模型具备可运行、可部署、可闭环的工程能力，而不仅是离线评测中的生成或预测能力。



1.4 产业痛点：具身智能产业化面临四大核心瓶颈，世界模型是破局关键（2/2）



具身智能产业破局的关键在于让机器人拥有**理解物理世界规律的物理直觉**，世界模型可以在当前的四大瓶颈中提供相应支撑。

当前瓶颈	 需要何种物理直觉	 世界模型支撑路径	 产业化价值
1 数据瓶颈	- 提炼物理规律 - 对齐任务逻辑与动作后果	- 融合开放世界视频、人类行为数据、机器人交互数据 - 构建统一世界知识底座	- 降低机器人数据依赖 - 提升跨场景、跨物体、跨任务泛化能力
2 任务瓶颈	- 记住物体状态与空间关系 - 跟踪任务进度与因果链条	- 长时程状态建模 - 维持世界状态一致性	- 支撑多步骤连续任务 - 从短动作走向长链路执行
3 控制瓶颈	- 预判动作对环境的影响 - 理解行动—结果因果关系	- 理解—生成—预测一体化 - 连接环境感知、未来预测与动作策略	- 提升动作成功率 - 增强复杂场景适应能力
4 部署瓶颈	- 实时调用物理判断能力 - 支持在线反馈与修正	- 低延迟推理 - 端侧/边缘侧部署与高效表征	- 从离线演示走向真实部署 - 支撑可扩展的具身智能基础设施

02 从具身智能到世界模型： 核心内涵、技术演进与架构解析

2.1 划定边界：世界模型的内涵定义



尽管当前世界模型尚无统一定义，但从代表性研究来看，其能力目标正在不断趋同，但关于其能力边界已经逐渐形成较为一致的认识，即世界模型应能够学习环境规律、预测未来状态、推演动作影响，并服务于智能体规划与决策。

1 学习并形成内部世界表示

- 世界模型首先从连续感知数据中学习环境结构、对象关系及动态规律，并形成可供智能体持续使用的内部表示。这种表示并非对现实世界的简单记录，而是可支撑理解、推理和预测的认知基础。

3 推演动作带来的环境变化

- 区别于传统视频预测模型，面向具身智能的世界模型更加关注动作与环境之间的因果关系，即预测不同动作将如何改变未来世界状态。



2 预测环境未来演化

- 世界模型不仅描述当前状态，还能够生成符合物理常识的未来环境的发展趋势预测，使智能体能够提前理解环境变化，而不是仅依赖实时感知进行反应。

4 支撑规划与决策

- 世界模型最终并非服务于预测本身，而是帮助智能体在真实行动之前完成内部推演，对不同策略进行比较和评估，从而降低真实试错成本，提高决策效率。

模型定义

综合现有学术研究及技术发展趋势，基于其承担的核心能力对世界模型进行定义：

- 世界模型是智能体通过持续感知与环境交互，学习外部世界运行规律，在内部形成环境表示，并能够预测环境演化、推演动作影响、支撑规划决策的一类核心认知能力。
- 其本质是在智能体内部构建一个能够理解世界、模拟未来并指导行动的动态世界表示，从而实现更加**高效、安全和自主的学习与决策**。

2.2 技术演进简史与爆发：从1950s到2026的五个时代



世界模型并非近年来突然出现的新概念，而是在控制理论、序列建模、强化学习和生成模型等多个技术方向长期积累后逐步融合形成的新范式，其核心演进逻辑表现为从状态预测走向环境建模，从单步预测走向长期规划，从面向特定任务的模型发展为能够支撑通用智能的世界基础模型。在这些分散技术逐步成熟并开始相互融合的过程中，世界模型的雏形才得以显现。



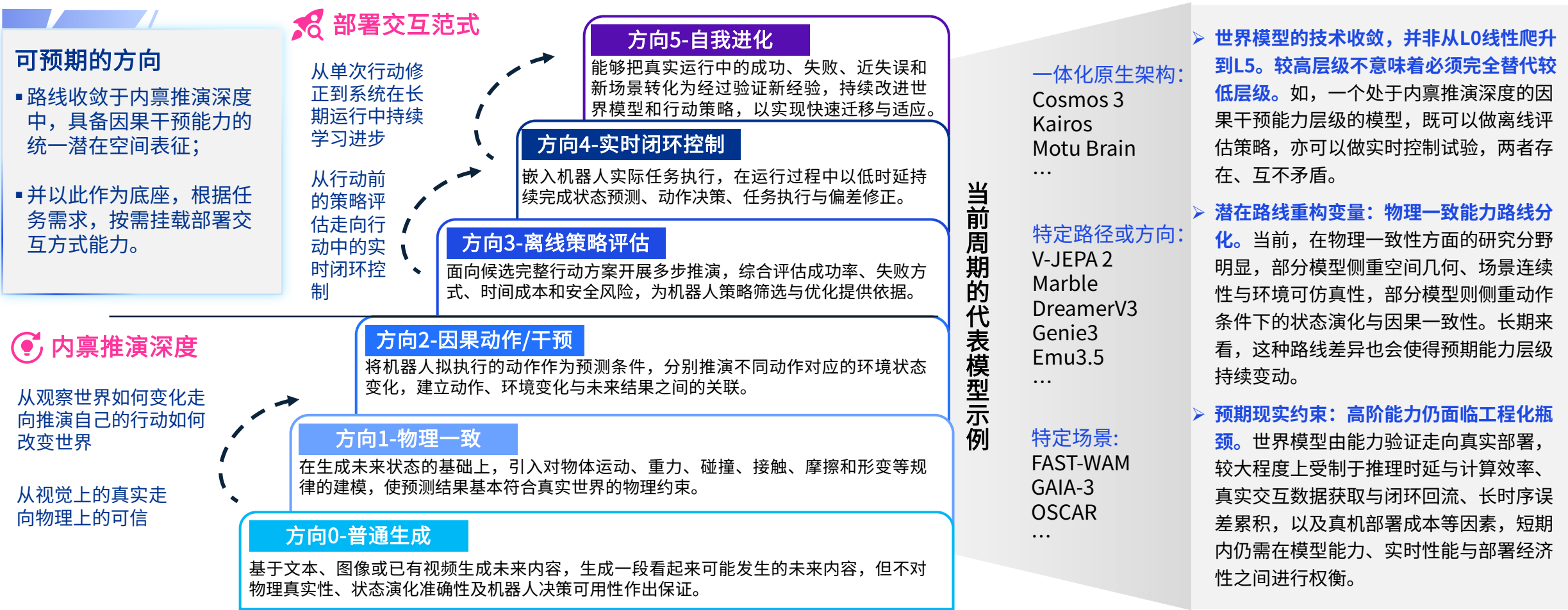
2.3 全球世界模型的主流技术架构流派评析



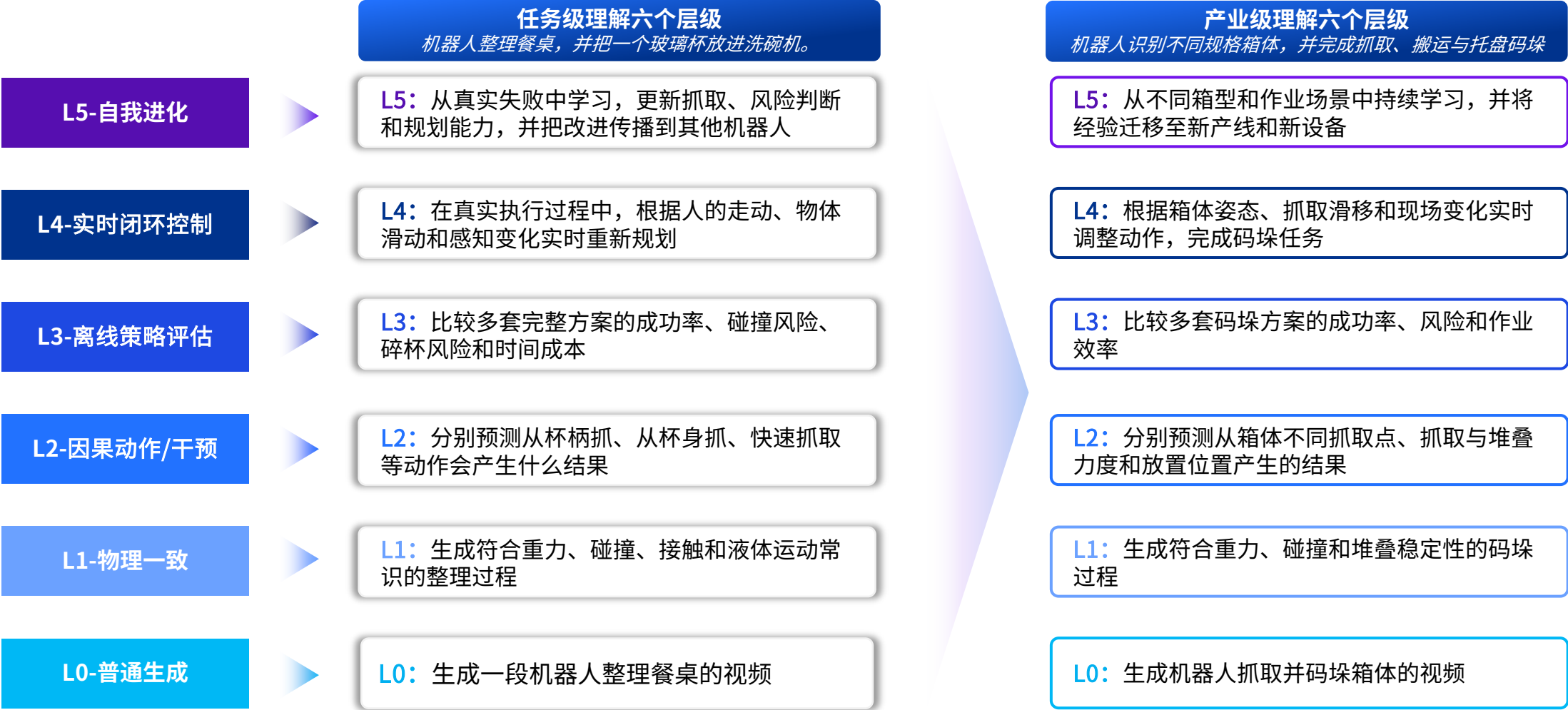
现阶段技术路径正呈现由百花齐放向**一体化架构收敛趋势**



2.4 模型能力进阶图谱：各流派拆解技术方向在当前周期内并行演进



2.4 世界模型能力层级理解：从单一任务到产业应用



2.5 世界模型六大工业化价值（1/2）

01

样本效率提升
从真实试错走向内部学习

- 世界模型能够学习环境动力学，并在内部世界生成大量虚拟交互轨迹，使Agent无需依赖大量真实试错即可完成策略优化。
- 相比传统无模型强化学习高度依赖真实环境采样，世界模型能够**充分利用有限交互数据，在内部模拟中反复推演和学习**，显著降低真实环境的数据需求，提高训练效率，并减少高成本、高风险场景下的真实试验次数。

案例：

- Dreamer V3在Atari100K（强化学习领域经典的基准测试）的HNS得分为1.12，已超越人类水平。
- NVIDIA Isaac平台通过高保真机器人仿真环境和合成数据生成能力，使机器人在部署前能够在虚拟环境中进行大规模训练和策略验证，减少真实机器人交互需求，加速策略从仿真环境向真实场景迁移。

02

长程规划推演
从即时反应走向前瞻决策

- 世界模型不仅能够预测未来状态，更能够帮助Agent在行动前进行多步推演，比较不同动作路径可能带来的状态变化、任务收益和潜在风险。
- 机器人、自动驾驶等复杂系统**可在真实执行前完成内部预演，使决策由即时反应转向前瞻规划**，为机器人与工业系统等真实环境应用奠定基础。

案例：

- MuZero将隐空间环境模型与蒙特卡洛树搜索相结合，无需预先获得环境规则，通过预测未来奖励、策略和价值，对候选行动及其长期结果进行多步评估与规划。该方法验证了学习型环境模型可直接服务于决策搜索，为世界模型支持长期规划提供了重要概念基础。

03

反事实安全验证
从真实测试走向可控压力测试

- 在自动驾驶、工业控制、机器人作业等试错成本极高的场景中，世界模型可作为可控仿真与反事实推演工具，用于生成真实路测中低频但高风险的复杂场景，例如极端天气、遮挡行人、异常车辆行为和突发交通冲突。
- 世界模型的价值不在于替代真实测试，而在于**扩大安全场景覆盖范围，降低真实环境试错成本，并提升模型在长尾场景下的鲁棒性**。

案例：

- Wayve的GAIA-1可基于多模态输入生成可控驾驶场景；GAIA-2增强了场景保真度和可控性，支持多视角一致性及边缘场景生成；GAIA-3可对不同驾驶策略进行验证，实现策略安全性、鲁棒性和泛化能力评估。
- Waymo World Model可生成罕见极端场景，如龙卷风、野火、洪水和动物闯入，并通过驾驶行为控制、场景布局控制和语言控制支持反事实测试。

2.5 世界模型六大工业化价值（2/2）

04

策略评估筛选
从模型生成走向策略评估

- 传统机器人策略验证高度依赖真实环境测试，测试成本高、覆盖范围有限且迭代效率低。
- 世界模型的预测环境变化能力可扩展为机器人研发流程中的评价工具，通过构建可交互的虚拟环境以模拟不同策略执行过程，并对任务完成效果进行自动化评价，**使机器人能够在部署前完成策略筛选、性能比较和风险识别，降低真实机器人测试成本，提高策略开发效率。**

案例：

- OSCAR基于Omni-Embodiment Action-Conditioned World Model，通过动作条件生成机器人未来执行过程，并利用视觉语言模型完成任务成功判断，实现虚拟策略评估。实验表明，OSCAR的虚拟策略评估与现实世界评估存在显著相关性，证明世界模型具备作为机器人策略评估并降低评估成本的潜力。

05

部署效率提升
从离线仿真走向实时闭环

- 相比传统依赖单任务训练和真实交互测试的机器人模型，世界模型通过连续环境建模、低延迟推理和系统架构优化，能够持续预测环境变化、快速生成动作并支持闭环执行，**支撑具身智能系统完成更高效的感知、预测、决策和执行实时闭环。**
- 随着世界模型运行成本和部署门槛降低，进一步**推动其从离线仿真与研究工具向实际应用与部署演进**，加速物理AI在复杂场景中的规模化应用。

案例：

- MotuBrain基于统一世界动作模型，将视觉、动作和语言信息进行联合建模，使机器人能够从单任务训练模式向多任务快速迁移模式转变；并通过多视角表示、语言动作耦合和高效推理架构优化，实现超过50倍推理加速，支持人形机器人以11Hz控制频率运行，提高机器人系统部署效率。

06

跨场景泛化迁移
从单任务策略走向可复用世界理解

- 世界模型的长期价值在于推动智能体从任务特定策略走向可迁移的环境理解。
- 与直接拟合单一任务动作策略不同，世界模型尝试学习环境状态、动作后果和物理动态之间的关系，故**可够支持知识迁移、跨场景复用及开放环境泛化，使智能体能够在不同任务、场景和硬件形态之间复用已有知识**，降低每个新场景从零训练的成本。

案例：

- Kairos通过原生物理知识学习和跨具身数据建模，使世界模型能够从不同机器人形态和交互经验中学习共享环境规律，并将已有知识迁移至新的任务和场景。该能力提升了物理AI系统的泛化能力，使机器人能够基于统一世界表示适应更开放、更复杂的应用环境。

03

国际主流世界模型评估： 核心指标体系与榜单解读

3.1 核心指标深度评估体系设计（1/3）



世界模型核心指标深度评估体系

随着人工智能技术的快速发展，世界模型作为**连接感知与决策的重要组件**，其评估体系的建立变得日益重要。近年来，国际学术界和产业界开始构建更加系统化、多维度的世界模型评估体系。这些体系强调世界模型在**物理遵循、时空一致、感知理解、因果推演、生成质量、任务执行、跨场景应用**等不同维度的表现。基于此，我们将世界模型评估的重点放在**三个方面**，并以此拆解为世界模型**六个核心特征**：

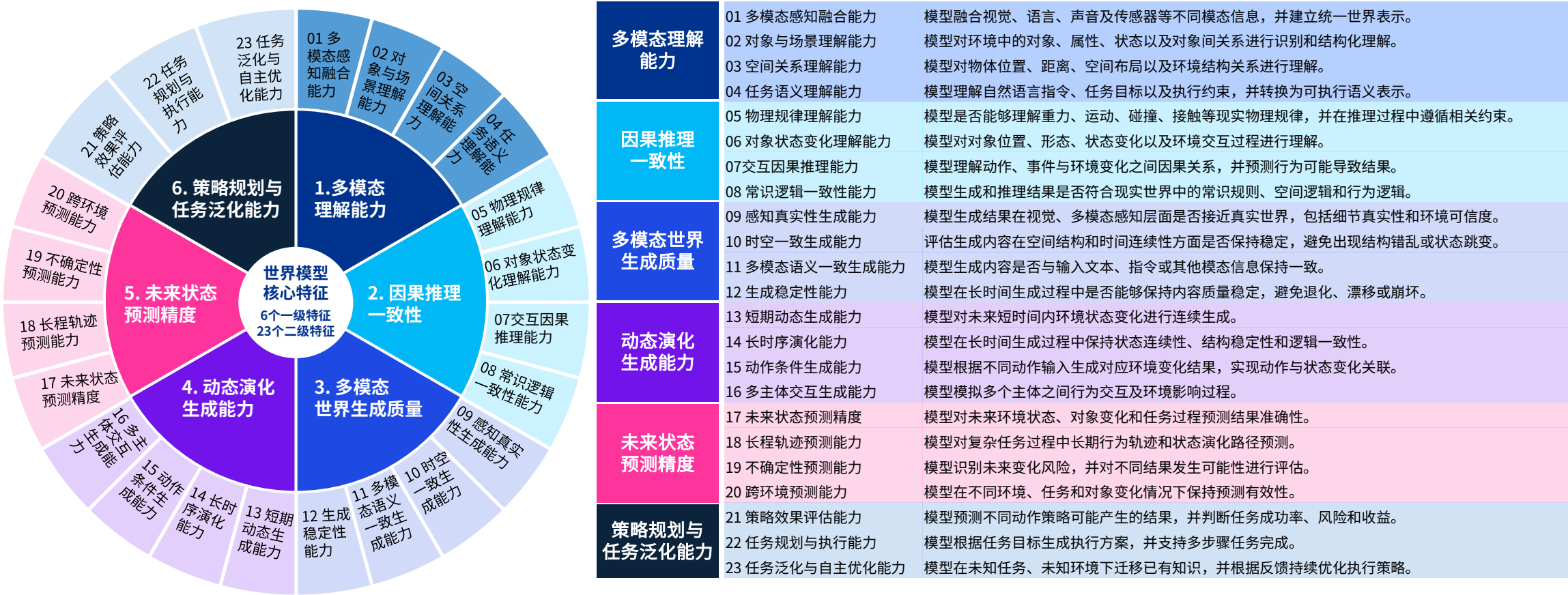


3.1 核心指标深度评估体系设计（2/3）



世界模型核心指标深度评估体系

基于上述三大关注方向和6大特征，对世界模型的表现进行进一步拆解，形成可量化、可对比的多维评价指标体系。



3.1 核心指标深度评估体系设计（3/3）

世界模型能力演进分析

基于面向泛化场景中具身任务完成需求的世界模型能力演进视角，分析不同发展阶段对于六大核心特征的依赖程度变化，其中：

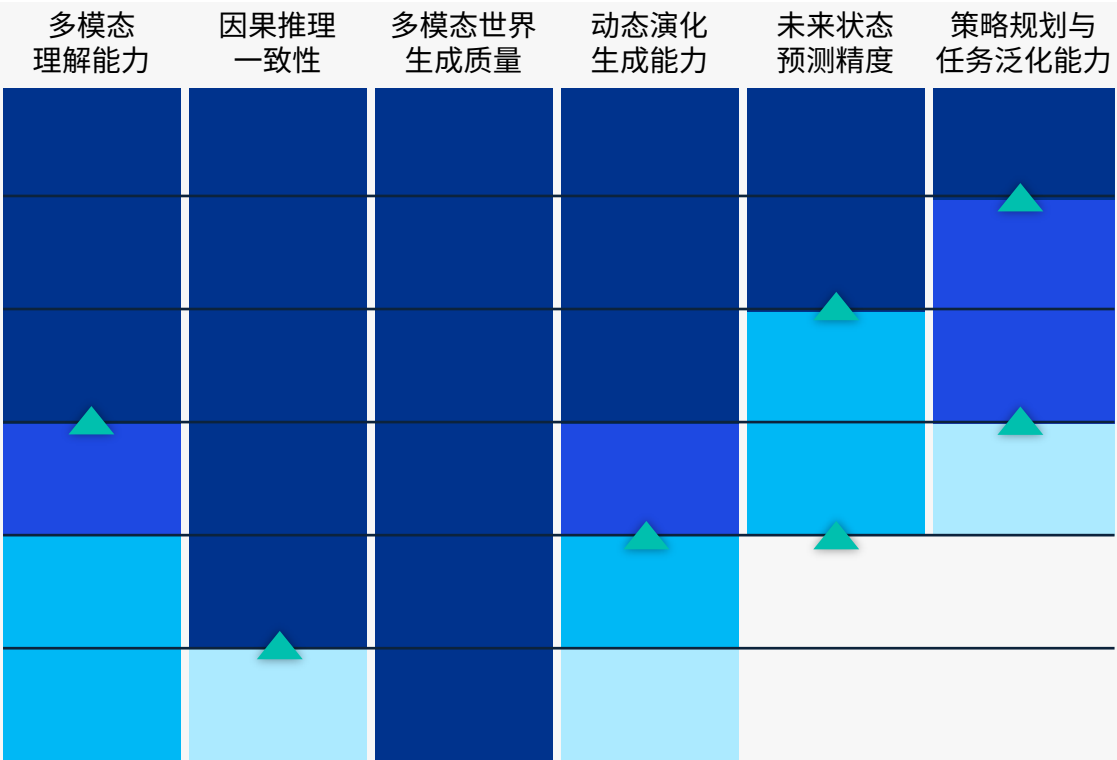
- 横轴表示世界模型六大核心特征，纵轴表示六个能力层级
- 矩阵颜色深浅代表各阶段对核心特征的依赖程度，颜色越深表示对该特征的依赖程度越高
- 箭头标识各阶段向下一等级演进过程中需要优先突破的关键特征，体现世界模型发展的主要演进路径

演进特征：世界模型从理解与生成开始走向预测与执行，能力需求也由早期的内容生成与世界理解，逐渐向未来预测、任务执行和自主进化演进。

能力层级&任务完成核心特征

- L5-自我进化
- L4-实时闭环控制
- L3-离线策略评估
- L2-因果动作/干预
- L1-物理一致
- L0-普通生成

具身任务导向的能力层级对六大核心特征要求



表示不同层级对各项核心特征的依赖程度，颜色越深代表依赖程度越高

表示从当前层级向下一层级演进过程中，需要优先突破和强化的核心特征

3.2 国际主流榜单参考与具身视角映射解读（1/5）

世界模型国际主流榜单与评测基准									
榜单/评测基准	基本介绍	发布机构/实验室	关注点与核心评价维度	对6大特征的评价侧重点					
				多模态理解能力	因果推理一致性	多模态世界生成质量	动态演化生成能力	未来状态预测精度	策略规划与任务泛化能力
WorldArena	面向具身智能世界模型的综合评测基准，评价感知质量和功能价值，同时将世界模型作为机器人数据引擎、策略评价器进行统一评估，强调世界模型是否能够服务机器人任务，而不仅是生成逼真视频。	清华大学等高校联合团队（THU、SJTU、HKU、Princeton、CAS等）	关注世界模型从生成能力走向交互和决策能力 - 视觉质量visual quality - 运动质量motion quality - 内容一致性content consistency - 物理遵循性physics adherence 3D准确性3D accuracy - 可控性controllability等	☆	☆	☆☆☆	☆☆☆		
WorldScore	面向世界生成模型的统一评价体系，重点评价模型生成世界的质量、动态变化和可控性，包括3D/4D/视频世界生成能力	Stanford Vision and Learning Lab等相关研究团队	关注世界生成能力，回答生成的世界是否真实可信 - 可控性评估Controllability - 质量评估Quality - 动态评估Dynamics	☆	☆	☆☆☆	☆☆☆		
DreamGen	利用视频世界模型生成机器人训练数据，验证世界模型生产的数据对机器人学习的价值。	NVIDIA GEAR Lab、UT Austin等团队	关注世界模型生产的数据能否提高机器人学习效果 - 指令遵循度（Instruction Following, IF） - 物理合理性（Physics Alignment, PA） - 人工评审	☆	☆	☆☆☆	☆☆☆		
WorldModelBench Robot	WorldModelBench整体用于评价视频生成模型作为世界模型的能力，覆盖机器人、自动驾驶、工业等场景，通过Instruction Following、Common Sense、Physical Adherence评价模型是否具备世界建模能力。	UC Berkeley、MIT、NVIDIA、UCSD等团队	关注动作合理性和物理遵循能力 - 指令跟随性Instruction Following - 物理遵循性Physics Adherence - 基本常识性Common Sense		☆☆☆		☆☆☆		
RoboTwin 2.0	面向机器人双臂操作任务的大规模评测基准，重点评价机器人在模拟环境中的操作能力、任务完成能力和泛化能力。	上海人工智能实验室、香港大学、上海交通大学等多家机构联合发布	关注机器人在复杂、多变化环境下能否完成任务 - 平均成功率ASR(Average Success Rate) - Top5平均成功率Top5-ASR(success over the top-5 candidates per task) - 平均在终止前的迭代次数CR-Iter(average refinement iterations before termination) - 生成策略代码的平均 token 数Token(average number of tokens in generated policy code)					☆☆☆	☆☆☆
LIBERO-Plus	基于LIBERO扩展的机器人鲁棒性评测基准，通过改变物体布局、摄像机视角、机器人状态、语言指令等因素测试VLA模型鲁棒性。	复旦大学、同济大学、上海人工智能创新研究院等	关注世界模型能否具备环境泛化能力 - 组合泛化能力（Compositional Generalization） - 任务成功率（the task success rate）					☆☆☆	☆☆☆
World Models Performance Index (WMPI)	综合性世界模型评价排行榜，通过多个维度综合衡量世界模型能力，目标是建立世界模型能力比较体系。	World Models相关社区/研究团队维护	关注对业内各类世界模型综合能力的观察与评价 - 世界真实感与物理真实感（World Fidelity & Physical Realism） - 交互性、控制与规划（Interactivity, Control & Planning） - 时空一致性与记忆（Temporal/Spatial Consistency & Memory） - 泛化与迁移（Generalization & Transfer） - 实用性与生态系统（Practicality & Ecosystem）	☆☆	☆☆	☆☆	☆☆		☆☆☆

3.2 国际主流榜单参考与具身视角映射解读（2/5）



世界模型国际主流榜单与评测基准在不同关注方向下的模型示例

感知与理解方向

- 通过融合视觉、语言、声音、传感器等多模态信息，对现实环境中的对象、状态、空间关系、物理规律以及任务目标进行感知和理解。
- 对应六大特征中的多模态理解能力、因果推理一致性。

Qwen-RobotWorld 阿里巴巴 通义千问团队	Emu3.5 北京智源 人工智能研究院
Kairos 大晓机器人	Cosmos 3 NVIDIA
Sora 2 Open AI	HY-World 2.0 腾讯
...	

演化与生成方向

- 世界模型基于当前环境状态、历史轨迹以及外部动作输入，生成符合物理规律和因果逻辑的未来世界状态，并模拟环境动态演化过程。
- 对应六大特征中的多模态世界生成质量、动态演化生成能力。

Kairos 大晓机器人	Genie DeepMind
GigaWorld Giga AI极佳视界	HY-World 2.0 腾讯
Marble World Labs	GE 智元机器人
...	

预测与执行方向

- 基于对未来状态的预测，对不同动作策略进行评估、规划、决策和优化，并指导完成复杂任务或跨环境任务，由认知智能迈向行动智能。
- 对应六大特征中的未来状态预测精度、策略规划与任务泛化能力。

Kairos 大晓机器人	LingBot-VA 蚂蚁灵波
V-JEPA 2 Meta AI	MotuBrain 生数科技
FAST-WAM 星海图	Being-H0.7 智在无界
...	

鉴于当前世界模型领域尚未形成统一综合评价标准，本文不进行世界模型综合能力排名，也不对模型整体水平作出评价。本文参考公开评测榜单、论文实验结果及行业公开信息，围绕不同维度选取具有代表性或在特定指标中表现突出的模型作为案例分析对象。所选模型不代表推荐、认可或优劣判断。模型按照发布机构名称首字母顺序排列，排名不分先后，相关内容仅供参考。

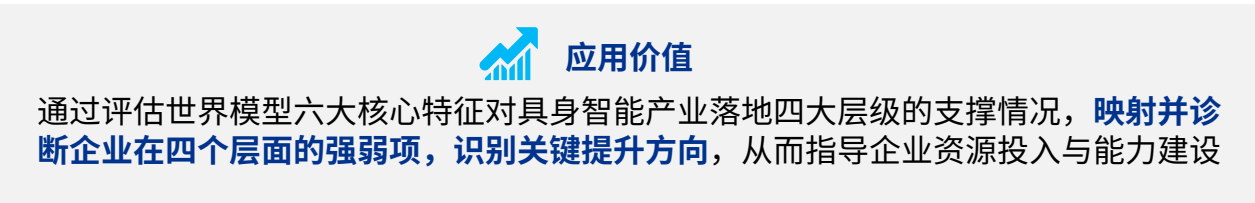
3.2 国际主流榜单参考与具身视角映射解读（3/5）

具身智能产业视角：从物理基础到智能决策再到执行结果，具身智能产业落地的四大层级

从具身智能产业视角出发，我们将具身智能产业落地拆分为4个层面，本质是按照从物理基础到智能决策再到执行结果的系统工程链路进行分层。

产业落地视角分层体系	这一层说的是什么？	本质是什么？	关注的是什么？
任务执行层 效率与结果	执行系统 能执行：将决策转化为具体任务闭环输出，衡量系统最终是否做成事情	任务闭环与结果交付系统	执行效率、全链路闭环、结果质量
认知决策层 大脑与动作	决策系统 能想：负责对环境进行理解、推理与决策，是具身智能的“大脑”	环境建模与决策生成系统	感知能力、推理与规划、泛化能力
运动控制层 物理运动性能	控制系统 能动：负责将高层决策转化为稳定、精确、连续的物理运动，实现系统动态平衡与可控行为	连续物理控制系统	运动精度、动态平衡、柔性控制
物理本体层 材料与硬件基础	物理系统 能造：具身智能系统赖以存在的物理基础，决定机器人能否被制造出来并稳定运行	物理约束与硬件基础	材料工艺、关节与执行器、续航与能效

3.2 国际主流榜单参考与具身视角映射解读（4/5）



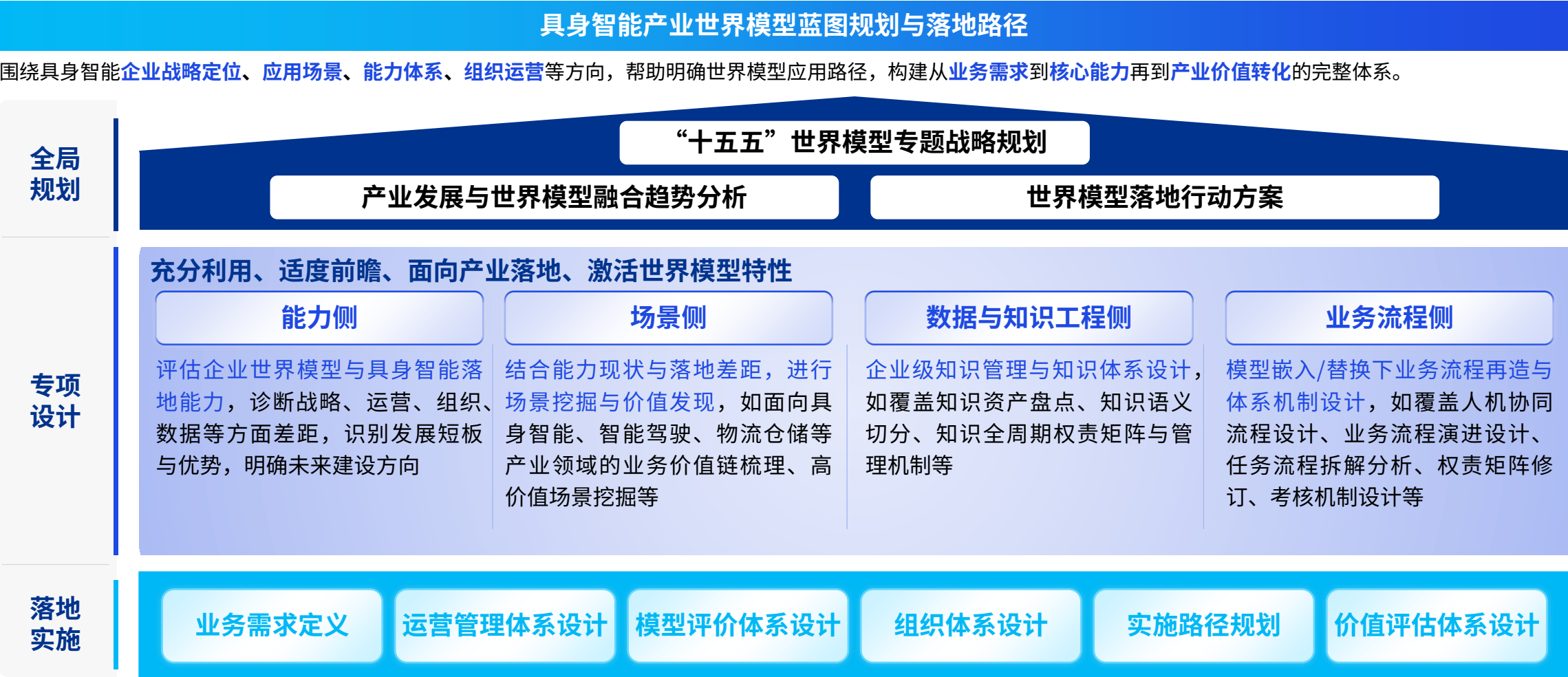
3.2 国际主流榜单参考与具身视角映射解读（5/5）

具身智能产业公司在落地过程所面临的挑战与需求

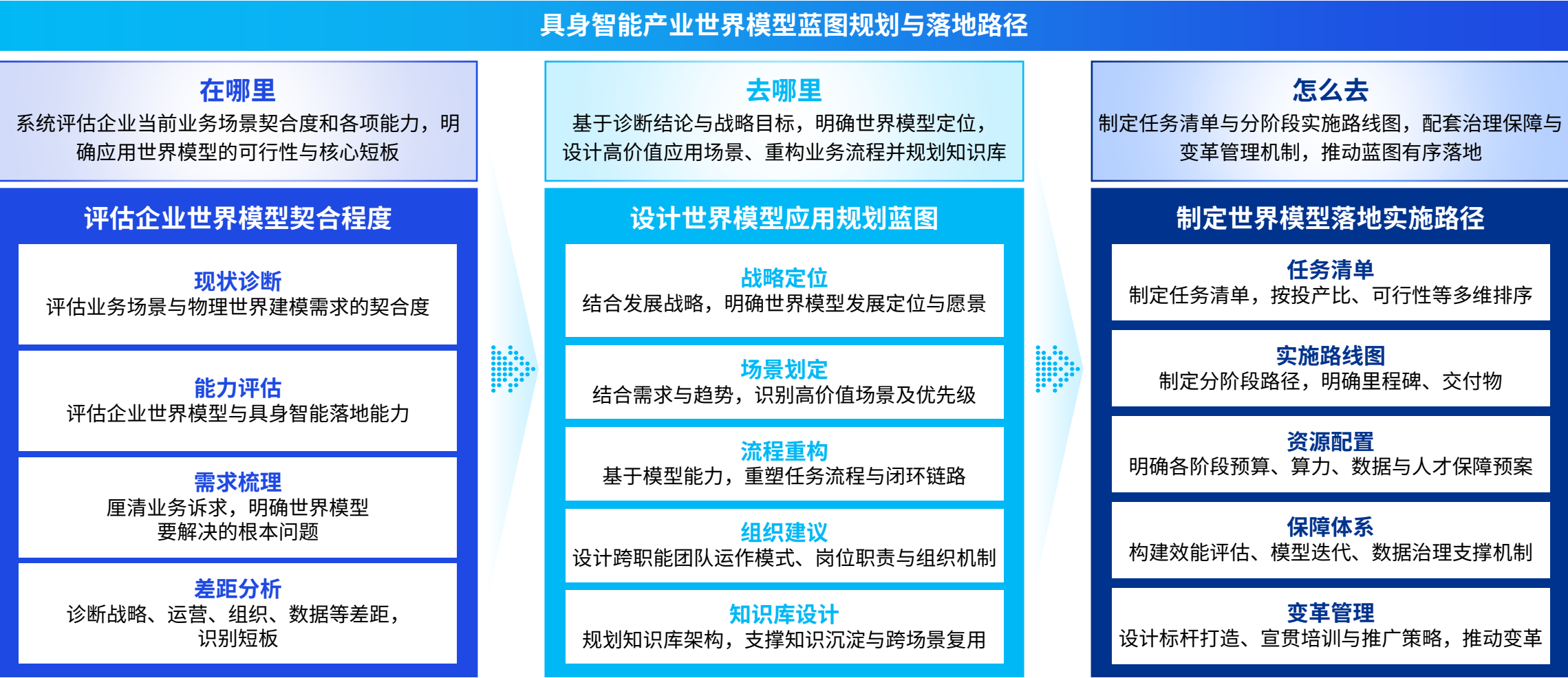
世界模型作为具身智能企业的核心基础设施，通过支撑**认知决策**、**运动控制**和**任务执行**等多层面能力的全方位提升，进一步推动具身智能企业的产业化落地；但在此过程中，诸多具身智能企业仍会面临**战略定位**、**能力建设**、**知识沉淀**、**流程重构**和**组织管理**等方面的一系列挑战与需求。

★ 全局视角	战略定位需求：不清楚世界模型在企业中的定位，需明确世界模型的战略定位与路径规划	商业模式需求：不清楚世界模型能力如何转化为产品竞争力与商业价值，需要设计产品形态、服务模式和盈利模式	场景筛选需求：不知道自身产业场景是否适合引入世界模型，缺少优先级判断，需要场景挖掘，筛选高价值落地场景	能力评估需求：无法量化自身是否具备物理世界建模的能力，需要进行能力评估	能力建设需求：不知道未来需要建设哪些能力，包括数据、知识、流程、人才，需要企业能力建设路线图设计	...	
任务执行层 效率与结果	任务流程重构需求：不知道在机器人帮助下，哪些流程需要优化、哪些需要合并、哪些需求摒弃，需要流程重构设计	任务执行评价需求：机器人完成任务但无法稳定达到业务要求，需要任务标准定义与评价体系设计		闭环流程设计需求：多步骤复杂任务无法形成完整闭环，仍然依赖大量人工辅助，需要端到端的业务闭环流程设计		...	
认知决策层 感知分析与决策	认知决策能力评价需求：不清楚世界模型应该具备什么程度的认知与决策能力，需要进行能力评价	模型架构建设需求：不清楚世界模型应该解决企业什么业务问题，需要有一个世界模型应用架构规划	知识工程建设需求：缺少对业务知识、行业规则、操作经验的系统沉淀，需要企业知识工程体系建设	数据治理需求：数据、经验分散，无法形成机器人可理解的知识体系，需要对企业现有数据进行治理	场景泛化设计需求：一个场景训练后无法迁移到其他场景，需要场景体系设计与能力复用规划	模型评价体系设计需求：缺少评价世界模型是否满足业务需求的方法，需要模型能力评价体系建设	...
运动控制层 物理运动性能	业务动作流程标准化需求：不清楚哪些任务与动作适合机器人，需要将业务任务拆解为标准化流程与动作，形成机器人可理解、可执行的知识体系		运动控制闭环流程设计需求：缺少机器人运行过程中的反馈闭环设计，需要运行反馈设计与闭环流程机制建设		人机协同流程设计需求：人工频繁介入，不清楚哪些环节可以自动化，需要人机协同流程设计、自动化边界设计	模型跨场景应用体系设计需求：不同场景之间无法复用已有经验，需要场景标准化与经验复用体系设计	...
物理本体层 材料与硬件基础	本体能力定义需求：不清楚目标业务场景对于机器人形态、硬件能力边界有什么要求，需要场景需求分析与机器人能力需求定义		软硬件能力协同分析需求：缺少机器人硬件能力与业务流程之间的匹配分析，需要人机协同流程设计、能力匹配分析		知识工程建设需求：缺少机器人运行过程中的操作经验、故障经验、环境知识、维护知识等知识沉淀机制，需要企业知识工程体系建设		...

3.3 具身智能产业世界模型蓝图规划与落地路径（1/2）



3.3 具身智能产业世界模型蓝图规划与落地路径（2/2）

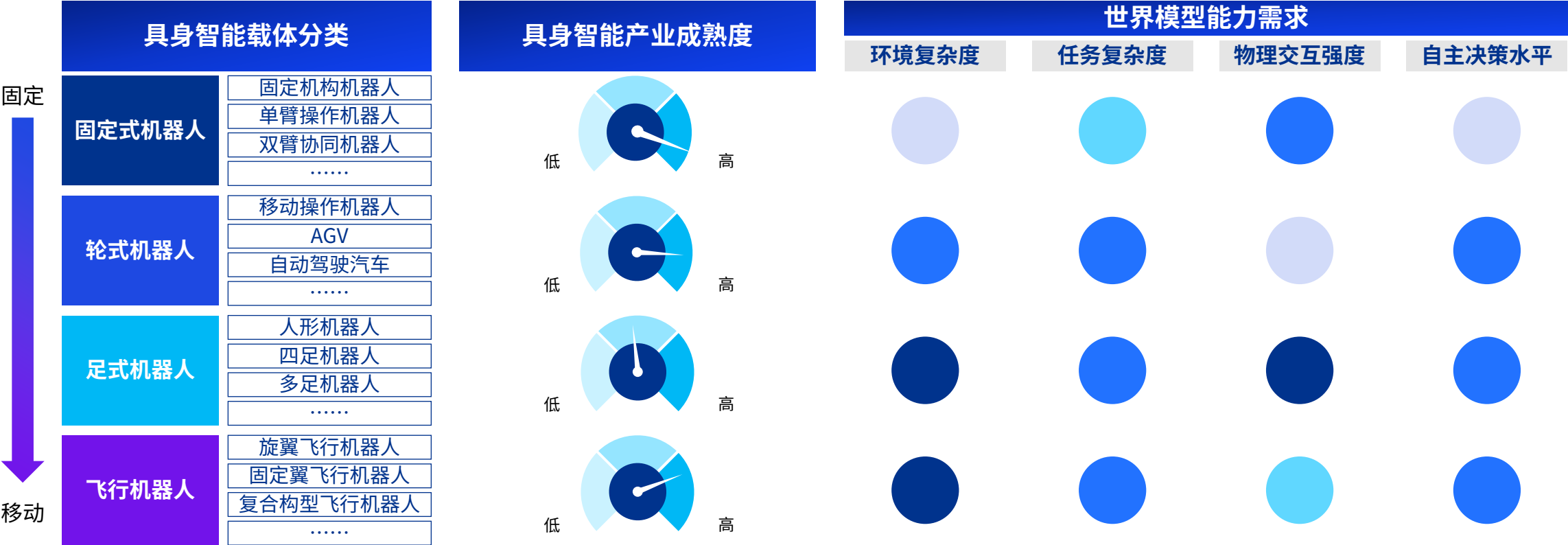


04

世界模型在具身智能领域的 产业落地实践

4 具身智能载体产业化水平与世界模型能力需求图谱

- 具身智能的物理载体形态呈现多样性，根据其运动空间和移动特性/方式，可以分为**固定底座机器人、轮式机器人、足式机器人、飞行机器人**；在此基础上根据机器人形态进行二级划分*
- 在实际应用场景中，不同形态的机器人可以进行有机地组合、兼备形态优势特点以满足具体的场景需求



4.1 世界模型在机械臂领域的落地实践：支撑精细操作预测与策略优化

- 传统工业机械臂主要依赖离线编程、人工示教和真实工位调试，每次生产流程和物料位置变化都需要重新采集数据、调整程序并开展试机验证，
- 世界模型通过学习机器人动作与环境变化之间的关系，可在机械臂真实执行前预测候选动作的结果，并发展为支持数据生成、策略评估、后训练和闭环优化的预测型基础设施

执行前预测动作结果

- 某工业机器人企业在既有VLA控制系统上增加世界模型规划模块，使工业机器人能够预测并避免故障模式。
- 该模型通过在视觉潜在空间中生成候选未来轨迹，进行预期成功率和效率评分，再由机械臂执行最高评分方案。通过在单臂和双臂操作平台上零人工干预进行四项工业基准测试任务并取得优于基线的成功率，表明基于世界模型规划的机械臂能够在复杂工业环境中可靠地运行。

部署前验证策略效果

- 某机器人研究团队构建跨机械臂形态的动作条件视频世界模型，以二维运动学骨架作为统一表示，接收策略输出的动作并生成执行后的视频，以估评估任务能否成功完成。
- 该模型已在机器人策略评测平台上对多种通用策略开展测试。实验结果显示，虚拟策略评估与真实机械臂评估具有较强相关性，表明世界模型可以用于模型版本比较、策略排序和部署前初步风险筛选，减少部分重复实机测试。

以生成经验优化策略

- 某机器人研究团队基于大规模固定式机械臂数据训练多视角世界模型，实现策略在生成环境中开展多轮闭环交互并连续生成多视角的联合预测。
- 该模型通过合成数据生成、无需真实机器人部署即可准确评估策略性，并生成多样化的合成成功轨迹用于策略微调。在实验中，经过微调基础策略在新物体和指令的成功率显著提升，验证了世界模型连接虚拟交互、合成数据生成和机械臂策略优化可行性。

世界模型能力需求

环境复杂度

任务复杂度

物理交互强度

自主决策水平

1

构建动作后果推演机制

- 结合工位观测、机械臂状态和候选动作，预测抓取、搬运、放置及装配过程中的未来状态，帮助机器人在真实执行前比较不同操作路径，降低错误动作和无效试机。

2

支撑策略效果评估验证

- 构建可重复运行的虚拟工位，对不同策略、模型版本和候选动作进行统一测试，识别碰撞、越界、任务失败及动作不稳定等风险。

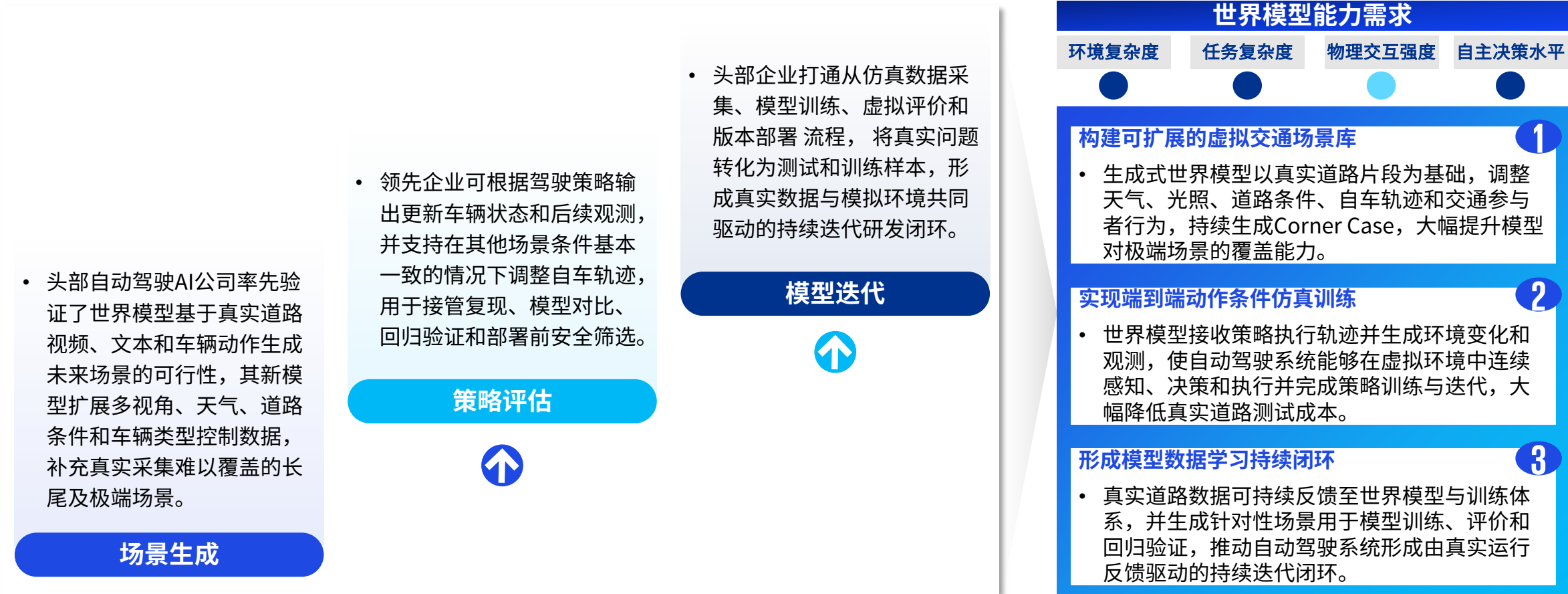
3

推动仿真训练闭环优化

- 自主生成新工件、位置、任务和失败场景，并为机器人策略提供虚拟交互轨迹和任务反馈，降低人工示教、设备占用和真实环境复位成本。

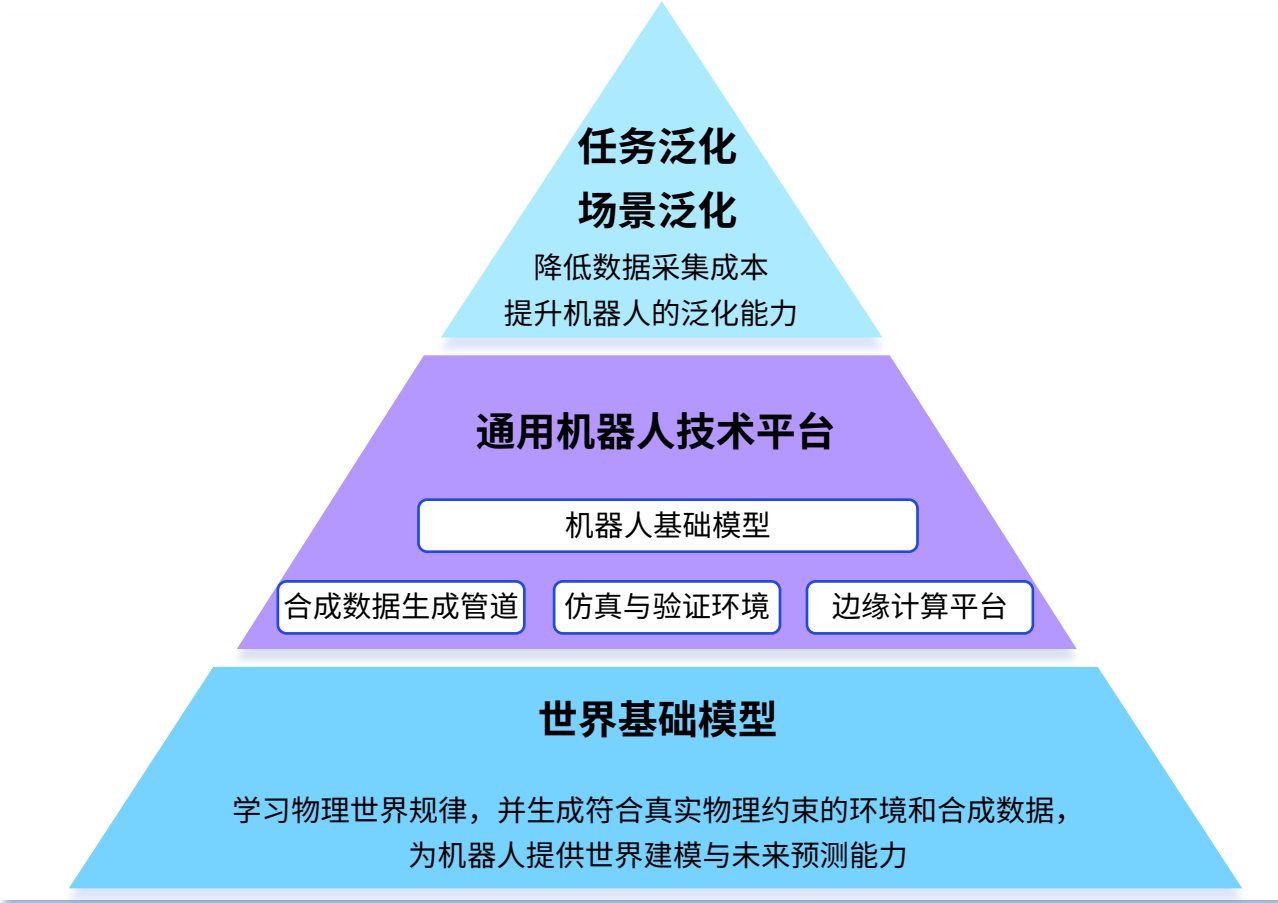
4.2 世界模型在自动驾驶领域的落地实践：重构虚拟验证与研发迭代范式

- 传统自动驾驶研发高度依赖真实道路采集，存在数据获取成本高、危险场景难以覆盖以及长尾问题难以验证等限制。
- 世界模型在自动驾驶中的不仅能够提升强化学习效率，更能够成为下一代智能系统的重要研发基础设施，广泛用于训练、评估和验证，已经具备较强产业可行性。



4.3 世界模型在人形机器人领域的落地实践：加速实现通用可泛化具身智能

- 传统人形机器人研发高度依赖遥操作示范和真实环境试验，训练数据获取和策略验证难度较大。
- 世界模型可生成操作数据与复杂场景，支持人形机器人持续训练与验证，并通过动作推演和任务规划，提升复杂操作、自主决策与持续优化能力。



世界模型能力需求

环境复杂度	任务复杂度	物理交互强度	自主决策水平
●	●	●	●

1

学习：仿真训练与持续进化

- 世界模型通过生成大量操作数据和复杂场景，实现机器人在数字世界持续训练、验证与优化。
- 与此同时，在真实执行过程中积累的数据能够持续反哺世界模型，不断完善环境建模能力，加速机器人能力迭代与跨场景泛化。

2

推演：构建物理推理与动作预演能力

- 世界模型能通过学习物体运动规律、接触关系及环境动态变化，使机器人在执行动作前对不同操作路径进行内部模拟，预测行为可能产生的物理结果，并选择成功率更高的执行策略。

3

规划：实现复杂操作任务的自主决策

- 面对开放环境多步骤任务，世界模型能够结合多维信息，对任务目标、环境状态及执行过程进行联合建模，自动拆解任务流程，规划连续动作序列，并根据实时反馈动态调整执行策略。

4.3 世界模型商业化落地过程中的工程化核心阻碍（1/2）

多数应用集中于研发验证、试点部署和示范场景

- ✓ 提升 研发效率 ? 提升 人工替代
- ✓ 降低 数据成本 ? 提升 设备效率
- ✓ 缩短 训练周期 ? 加快 生产节拍

一项针对1,678家企业的调研显示，**79%**的企业已开展具身智能的**探索**，仅**11%**的企业实现**规模化部署**，产业整体仍处于由试点向规模复制过渡阶段。

商业
价值
亟需提升

工程体系
建设
不足

训练和推理效率
VS
计算资源约束

VS

VS

虚实迁移可靠性
长时序误差累积

VS

模型性能
VS
高质量具身数据

数据来源：《Physical AI: Taking human-robot collaboration to the next level》，毕马威分析

4.3 世界模型商业化落地过程中的工程化核心阻碍（2/2）

模型性能提升需求

不仅需要二维视觉数据，还需要动作轨迹、关节状态、力矩、触觉、接触关系以及任务反馈等高质量、多模态数据共同参与训练。

高质量具身数据短缺

高质量具身数据的采集成本更高、标注更复杂，同时涉及现场、人员和设备等敏感信息，增加了数据获取和共享难度。



工程化建设探索：

构建真实、仿真与人类数据有序结合的数据体系

训练和推理效率提升需求

世界模型通常需要在当前观测和动作基础上预测一个或多个未来状态。部分方法还需要生成视频、执行迭代去噪，或在多个候选动作之间反复开展模型滚动和规划，由此造成训练成本增加与推理延迟。

计算资源约束

推理延迟、显存占用、通信开销与能耗都是硬性指标。对于具身智能而言，无法在目标任务的决策周期内提供有用结果的世界模型，很难真正进入机器人闭环。



工程化建设探索：

减少像素级预测、减少生成步骤、蒸馏与分块预测

虚实迁移可靠性需求

世界模型通过学习机器人动作与环境状态变化之间的关系，对未来状态进行预测。要求实现预测结果与动作保持因果一致，并在长时序中维持物理、运动学和跨视角一致性。

长时序误差累积

当机器人执行一次不准确的动作后，真实环境状态已经发生变化，后续预测将基于偏离真实轨迹的状态继续演化。长时序任务会进一步放大这一问题。



工程化建设探索：

利用真实数据调参、短周期滚动、在训练阶段加入可执行约束

05

未来趋势与生态演进

总览：世界模型推动具身智能进入新竞争周期



未来具身智能企业的核心竞争力，将从硬件与单点系统交付，转向持续训练、验证、部署和运营的物理AI系统。

5.1 技术演进趋势：世界模型将从生成内容走向理解、预测与运行

4

能力层

从训练辅助工具走向运行基础设施

世界模型将从离线训练和数据补充工具，进入仿真测试、策略评估、风险验证、在线反馈和持续更新流程。它将成为连接数据生成、模型训练、推理决策和系统部署的基础层。

3

预测层

从内容生成走向状态演化与行动后果预测

世界模型的核心能力将从生成逼真画面，升级为预测对象、空间、关系和物理状态如何变化。更重要的是，模型需要判断不同行动可能带来的结果、风险和失败路径。

2

表征层

从单模态生成走向全模态世界表征

世界模型将不再只依赖图像或视频表征世界，而是融合文本、图像、视频、音频、3D、传感器状态和行动信号。模型目标将从生成视觉结果，转向建立统一的时空状态表征。

1

数据层

从单一视频语料走向多源世界数据

世界模型训练数据将从互联网视频扩展到仿真轨迹、3D场景、传感器数据、交互记录、失败样本和合成数据。关键能力不只是获取更多数据，而是将多源数据转化为可训练、可对齐、可验证的世界经验。

5

生态层

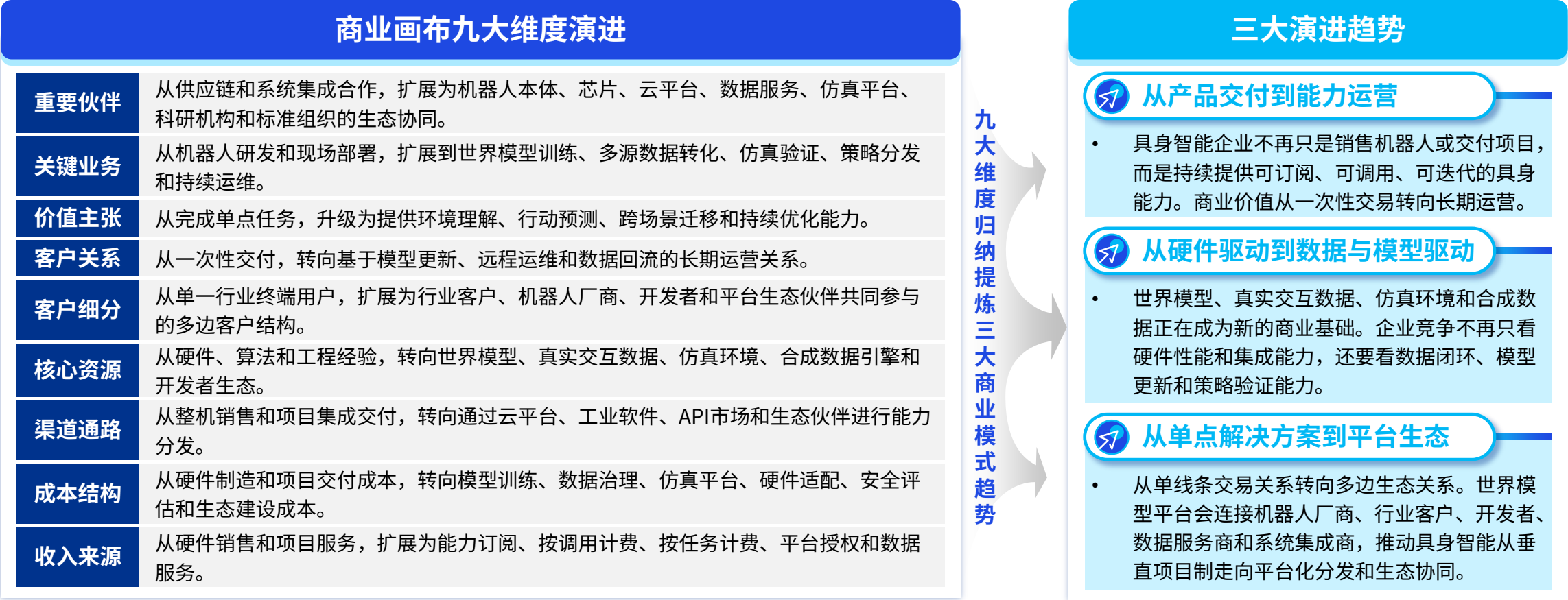
从闭源模型竞争走向开源工具链与标准化接口

世界模型竞争将不只体现在模型参数和生成效果上，也体现在数据管线、仿真环境、评测框架、训练工具和接口标准上。开源工具链和标准化接口将推动世界模型从少数机构探索走向更广泛的生态协同。

5.2 商业模式画布：世界模型推动具身智能商业模式从一次性交付向持续运营服务升级



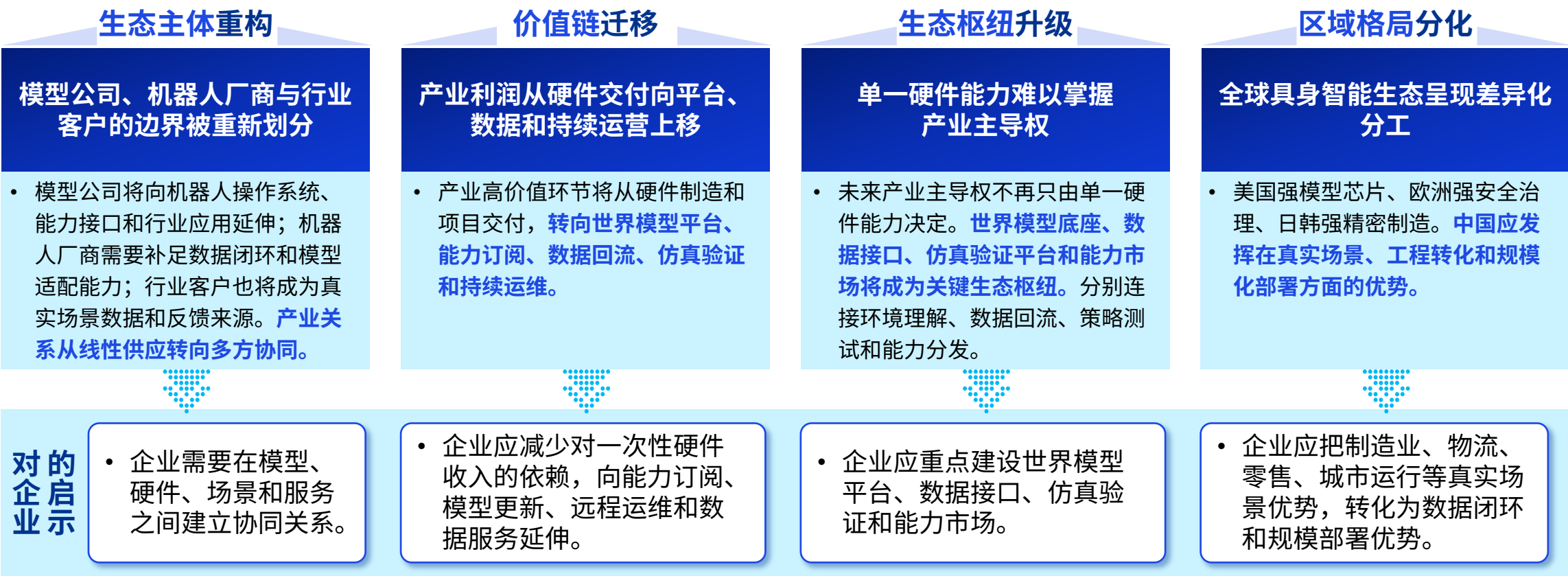
5.2 商业模式演进：从产品交付走向能力运营与平台生态



具身智能的商业价值，将从一次性交付向持续运营服务延伸。

5.3 生态展望：从机器人产业链走向物理AI生态网络

世界模型将推动具身智能从机器人产业链走向物理AI生态网络，竞争重点转向数据闭环、平台能力、仿真验证和生态协同。



附录：参考资料（1/5）

- 【1】 中华人民共和国国民经济和社会发展第十五个五年规划纲要，国务院，2026年3月13日， https://www.gov.cn/yaowen/liebiao/202603/content_7062633.htm
- 【2】 工业和信息化部关于印发《人形机器人创新发展指导意见》的通知，工信部，2023年11月2日
https://www.miit.gov.cn/jgsj/kjs/wjfb/art/2023/art_50316f76a9b1454b898c7bb2a5846b79.html
- 【3】 工业和信息化部办公厅 国务院国资委办公厅关于联合开展2026年度人形机器人与具身智能实景实训专项行动的通知，工信部，2026年6月8日，
https://www.miit.gov.cn/jgsj/kjs/wjfb/art/2026/art_cd666691abf8471fb8553d463aa416e3.html
- 【4】 IDC：AI服务器同比增长超一倍，市场环境深化本土化进程，2025年10月8日， <https://my.idc.com/getdoc.jsp?containerId=prCHC53858025>
- 【5】 北京市加快推动场景培育和开放应用工作方案，北京市人民政府，2026年5月13日，
<https://www.beijing.gov.cn/zhengce/zhengcefagui/202605/W020260514582672176065.pdf>
- 【6】 上海市经济信息化委等关于印发《上海市促进智能机器人产业高质量 创新发展行动方案（2023-2025年）》的通知，上海市经济信息化委，2023年10月26日，
<https://www.shanghai.gov.cn/xxsjsj/20250606/fea5a592c4374472aa574b033355d7a2.html>
- 【7】 深圳市加快打造人工智能先锋城市行动计划（2025—2026年），深圳市工业和信息化局，2025年3月3日，
https://gxj.sz.gov.cn/xxgk/xxgkml/zcfgjzcjd/qtwj/content/post_12052805.html
- 【8】 首批人形机器人系列国家标准正式立项，亦城时报，2025年4月21日， <https://open.beijing.gov.cn/html/yizhuang/gzdt/2025/4/1745221749437.html>
- 【9】 具身智能发展报告2025，中国信通院，2026年2月3日， <https://www.caict.ac.cn/kxyj/qwfb/bps/202601/P020260130541978285206.pdf>
- 【10】 银河通用机器人展示商超值守机器人已获得100家门店订单，证券时报网，2025年7月27日， <https://finance.eastmoney.com/a/202507273468476973.html>

附录：参考资料（2/5）

- 【1】 Ha, D., Schmidhuber, J., World Models, arXiv:1803.10122, 2018.
- 【2】 Li, X., He, X., Zhang, L. et al., A Comprehensive Survey on World Models for Embodied AI, arXiv:2510.16732, 2025.
- 【3】 Motubrain Team, Motubrain: An Advanced World Action Model for Robot Control, arXiv:2604.27792, 2026.
- 【4】 Schrittwieser, J., Antonoglou, I., Hubert, T. et al., Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model, arXiv:1911.08265, 2019.
- 【5】 Burchi, M., Timofte, R., Learning Transformer-based World Models with Contrastive Predictive Coding, arXiv:2503.04416, 2025.
- 【6】 Guo, Y., Shi, L. X., Chen, J. et al., Ctrl-World: A Controllable Generative World Model for Robot Manipulation, arXiv:2510.10125, 2025.
- 【7】 Aida, A., Amer, W., Bankovic, K. et al., Cortex 2.0: Grounding World Models in Real-World Industrial Deployment, arXiv:2604.20246, 2026.
- 【8】 Lee, J., Bjelonic, M., Reske, A. et al., Learning Robust Autonomous Navigation and Locomotion for Wheeled-Legged Robots, arXiv:2405.01792, 2024.
- 【9】 Wu, Z., Gao, J., OSCAR: Omni-Embodiment Action-Conditioned World Model for Robotics, arXiv:2606.04463, 2026.
- 【10】 NVIDIA, Cosmos World Foundation Model Platform for Physical AI, arXiv:2501.03575, 2025.
- 【11】 ACE-Brain Team, ACE-Brain-0.5: A Unified Embodied Foundational Model for Physical Agentic AI, arXiv:2607.04426, 2026.
- 【12】 Chu, M., Zhang, X. B., Lin, K. Q. et al., Agentic World Modeling: Foundations, Capabilities, Laws, and Beyond, arXiv:2604.22748, 2026.
- 【13】 Hou, B., Li, G., Jia, J. et al., World Model for Robot Learning: A Comprehensive Survey, arXiv:2605.00080, 2026.
- 【14】 Kube, D., Hadwiger, S., Meisen, T., Robotic Foundation Models for Industrial Control: A Comprehensive Survey and Readiness Assessment Framework, arXiv:2603.06749, 2026.
- 【15】 Huang, J., Chen, W., Li, Z. et al., H-WM: Robotic Task and Motion Planning Guided by Hierarchical World Model, arXiv:2602.11291, 2026.
- 【16】 Wang, Y., Syed, R., Wu, F. et al., Interactive World Simulator for Robot Policy Training and Evaluation, arXiv:2603.08546, 2026.
- 【17】 Luo, H., Zhang, W., Feng, Y. et al., Being-H0.7: A Latent World-Action Model from Egocentric Videos, arXiv:2605.00078, 2026.
- 【18】 Lou, H., Zhang, M., Geng, H. et al., D-REX: Differentiable Real-to-Sim-to-Real Engine for Learning Dexterous Grasping, arXiv:2603.01151, 2026.
- 【19】 Wang, R., Liu, Q., Deng, Y. et al., EVA: Aligning Video World Models with Executable Robot Actions via Inverse Dynamics Rewards, arXiv:2603.17808, 2026.
- 【20】 Tu, P., Zhu, H., Sun, J. et al., Embod4D: A Generalist Data Engine for Embodied 4D World Modeling, arXiv:2605.01799, 2026.
- 【21】 Karcini, E., Mehrban, F., Nguyen, Q. et al., Robots Need More than VLA and World Models, arXiv:2606.06556, 2026.
- 【22】 Kairos Team, Kairos: A Regret-Aware Native World-Action Model Stack for Physical AI, arXiv:2606.16533, 2026.
- 【23】 甲子光年智库, 2025: 《具身智能技术发展与行业应用简析》, 《2025具身智能行业发展研究报告》系列报告之一
- 【24】 Capgemini, Physical AI : Taking human-robot collaboration to the next level, 2026
- 【25】 NVIDIA, Isaac GR00T: Generalist Robot 00 Technology, NVIDIA Developer, 2024.
- 【26】 NVIDIA, Physical AI: The Next Frontier of Artificial Intelligence, NVIDIA Technical Report, 2025.
- 【27】 Huang, H.-M., Autonomy Levels for Unmanned Systems (ALFUS) Framework: Safety and Application Issues, NIST, 2007.

附录：参考资料（3/5）

- 【1】 Bruce, J., Dennis, M., Edwards, A., et al., Genie: Generative Interactive Environments, arXiv:2402.15391, 2024
- 【2】 Google DeepMind, Genie 2: A Large-Scale Foundation World Model, Google DeepMind Blog, 2024.
- 【3】 NVIDIA et al., Cosmos 3: Omnimodal World Models for Physical AI, arXiv:2606.02800, 2026.
- 【4】 Assran, M., Bardes, A., Fan, D., et al., V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning, arXiv:2506.09985, 2025.
- 【5】 Meta AI, V-JEPA: Yann LeCun’ s AI Model for Video Understanding, Meta AI Blog, 2024.
- 【6】 Hafner, D., Lillicrap, T., Ba, J., Norouzi, M., Dream to Control: Learning Behaviors by Latent Imagination, arXiv:1912.01603, 2020.
- 【7】 Hafner, D., Lillicrap, T., Norouzi, M., Ba, J., Mastering Atari with Discrete World Models, International Conference on Learning Representations (ICLR), 2021.
- 【8】 Hafner, D., Pasukonis, J., Ba, J., Lillicrap, T., Mastering Diverse Domains through World Models, arXiv:2301.04104, 2023.
- 【9】 Yuan, T., Dong, Z., Liu, Y. et al., Fast-WAM: Do World Action Models Need Test-time Future Imagination?, arXiv:2603.16666, 2026.
- 【10】 Hu, A., Russell, L., Yeo, H. et al., GAIA-1: A Generative World Model for Autonomous Driving, arXiv:2309.17080, 2023.
- 【11】 Russell, L., Hu, A., Bertoni, L. et al., GAIA-2: A Controllable Multi-View Generative World Model for Autonomous Driving, arXiv:2503.20523, 2025.
- 【12】 Cui, Y., Chen, H., Deng, H. et al., Emu3.5: Native Multimodal Models are World Learners, arXiv:2510.26583, 2025.
- 【13】 Wayve, GAIA-3: Scaling World Models to Power Safety and Evaluation, Wayve Research, 2025.

附录：参考资料（4/5）

- 【1】 Zhang, Y., Liu, Y., Chen, X., et al., WorldArena: A Unified Benchmark for Evaluating Perception and Functional Utility of Embodied World Models, arXiv:2602.08971, 2026.
- 【2】 Huang, R., Zhang, H., Zhang, Y., et al., WorldScore: A Unified Evaluation Benchmark for World Generation, arXiv:2504.00983, 2025.
- 【3】 Huang, Z., Zhang, Y., Liu, J., et al., WorldModelBench: Judging Video Generation Models As World Models, arXiv:2502.20694, 2025.
- 【4】 Liu, Y., Chen, X., Wang, Z., et al., LIBERO-Plus: In-depth Robustness Analysis of Vision-Language Action Models, arXiv:2510.13626, 2025.
- 【5】 Wu, Y., Chen, Z., Xu, X., et al., RoboTwin 2.0: A Scalable Data Generator and Benchmark with Strong Domain Randomization for Robust Bimanual Robotic Manipulation, arXiv:2506.18088, 2025.
- 【6】 Hu, Y., Li, X., Zhang, J., et al., DREAMGEN: Unlocking Generalization in Robot Learning through Video World Models, arXiv:2505.12705, 2025.
- 【7】 World Models Leaderboard, World Models Research Community, 2026. <https://world-models.io/en/leaderboard/>
- 【8】 NVIDIA et al., Kairos: A Regret-Aware Native World-Action Model Stack for Physical AI, arXiv:2606.16533, 2026.
- 【9】 Zhang, J., Chen, X., Wang, Y., et al., Being-H0.7: A Latent World-Action Model from Egocentric Videos, arXiv:2605.00078, 2026.
- 【10】 Wang, H., Li, Z., Zhang, Y., et al., MemoryVLA++: Temporal Modeling via Memory and Imagination in Vision-Language-Action Models, arXiv:2606.09827, 2026.
- 【11】 Zhang, H., Chen, X., Wang, Y., et al., World Models for Physical Intelligence: A Comprehensive Survey, arXiv:2604.27792, 2026.
- 【12】 Liu, S., Wang, Z., Chen, J., et al., Learning World Models for Embodied Agents: A Survey, arXiv:2603.16666, 2026.
- 【13】 Li, Y., Zhang, H., Wang, J., et al., Video World Models for Robot Learning: A Survey, arXiv:2603.17240, 2026.
- 【14】 Wang, Y., Liu, Z., Chen, X., et al., Towards Generalist Robot Learning with World Models, arXiv:2601.21998, 2026.
- 【15】 Chen, X., Zhang, Y., Li, J., et al., World Models as Foundation Models for Robotics, arXiv:2512.13030, 2025.
- 【16】 Wang, H., Zhang, L., Liu, Y., et al., Scaling World Models for Embodied Intelligence, arXiv:2604.11135, 2026.

附录：参考资料（5/5）

- 【1】 Wang, Z., Wang, B., Zhang, H. et al., Vision-Language-Action in Robotics: A Survey of Datasets, Benchmarks, and Data Engines, arXiv:2604.23001, 2026.
- 【2】 Karcini, E., Mehrban, F., Nguyen, Q. et al., Robots Need More than VLA and World Models, arXiv:2606.06556, 2026.
- 【3】 Hou, B., Li, G., Jia, J. et al., World Model for Robot Learning: A Comprehensive Survey, arXiv:2605.00080, 2026.
- 【4】 NVIDIA, Agarwal, N., Ali, A., Allen, J. et al., Kairos: A Native World Model Stack for Physical AI, arXiv:2606.16533, 2026.
- 【5】 NVIDIA, Agarwal, N. et al., Cosmos 3: Omnimodal World Models for Physical AI, arXiv:2606.02800, 2026.
- 【6】 Jiang, W., Clemons, J., O’ Flaherty, R., et al., ROSA: A Robotics Foundation Model Serving System for Robot Factories, arXiv:2607.01088, 2026.
- 【7】 NVIDIA, Agarwal, N., Ali, A., Allen, J. et al., Kairos: A Native World Model Stack for Physical AI, arXiv:2606.16533, 2026.
- 【8】 Li, X., He, X., Zhang, L., et al., A Comprehensive Survey on World Models for Embodied AI, arXiv:2510.16732, 2025.
- 【9】 MERICS, China’ s Ambitious Path to Transform Its Robotics Industry, 2026

联系我们：



江立勤
客户与业务发展主管合伙人
毕马威中国
电邮：
michael.jiang@kpmg.com



柳晓光
变革咨询数字化转型业务牵头人
“智慧之光”数智化解决方案主管合伙人
毕马威中国
电邮： silvester.liu@kpmg.com



孙箐阳
变革咨询数字化转型团队首席AI架构师
毕马威中国
电邮： eric.qy.sun@kpmg.com

研究团队：

王薇
研究院副总监
毕马威中国
电邮： wei.w.wang@kpmg.com

程苑芬
研究院助理经理
毕马威中国
电邮： fannie.cheng@kpmg.com

周逸夫
战略与变革咨询经理
毕马威中国
电邮： yifu.zhou@kpmg.com

寿乐怡
战略与变革咨询高级咨询顾问
毕马威中国
电邮： stephanie.shou@kpmg.com

杨宁
战略与变革咨询高级咨询顾问
毕马威中国
电邮： camelia.yang@kpmg.com

朱瑜欣
战略与变革咨询咨询顾问
毕马威中国
电邮： krystal.yx.zhu@kpmg.com

谢谢!

kpmg.com/cn/socialmedia



如需获取毕马威中国各办公室信息，请扫描二维码或登陆我们的网站：<https://kpmg.com/cn/zh/about/office-locations.html>

所载资料仅供一般参考用，并非针对任何个人或团体的个别情况而提供。虽然本所已致力提供准确和及时的资料，但本所不能保证这些资料在阁下收取时或日后仍然准确。任何人士不应在没有详细考虑相关的情况及获取适当的专业意见下依据所载资料行事。

© 2026 毕马威华振会计师事务所(特殊普通合伙) — 中国合伙制会计师事务所及毕马威企业咨询 (中国) 有限公司 — 中国有限责任公司，均是与毕马威国际有限公司(英国私营担保有限公司)相关联的独立成员所全球组织中的成员。版权所有，不得转载。

毕马威的名称和标识均为毕马威全球组织中的独立成员所经许可后使用的商标。

刊物编号：1783477337301

二零二六年七月印刷