



计算机行业研究

买入（维持评级）

行业专题研究报告

证券研究报告

计算机组

分析师：刘高畅（执业 S1130525120005）
liugaochang@gjzq.com.cn

分析师：郑元昊（执业 S1130525120004）
zhengyuanhao@gjzq.com.cn

分析师：孙恺祈（执业 S1130526060005）
sunkaiqi@gjzq.com.cn

AI 的下一站，物理 AI

行业观点

- 世界模型为代表的物理 AI 可视为 LLM 通向超智能机器人的中间站。LLM 擅长语言理解、知识调用和任务分解，已在 Coding、通用办公、金融、法律等白领工作领域实现体量不俗的商业化闭环，当然如今仍处于渗透率早期。我们认为，AI 的下一站是进入物理世界，物理 AI 还需理解三维空间、物体状态、接触关系和动力学约束，并预测不同动作可能带来的结果。世界模型通过学习环境状态及其随时间和动作发生的变化，为机器人建立可供推演的内部环境表征，使语言指令能够进一步转化为感知、预测、规划与控制。在这一框架下，VLA 将视觉和语言输入映射为机器人动作，世界模型则为动作选择提供未来状态预测和结果评估。两者逐步融合，有望推动机器人由单步操作走向多步规划、失败识别和自主恢复。随着模型能力向物理世界延伸，Physical AI 的竞争也将覆盖空间数据、物理仿真、机器人本体、训练算力和真实场景反馈等完整技术链条，从而为 AI 创造更大的商业化场景。
- AMD 收购 World Labs 推动世界模型进入产业投入期，模型与计算平台的协同进一步加深。2026 年 9 月 28 日，AMD 宣布拟以约 82 亿美元全股票收购 World Labs，交易预计于 2026 年底前完成，仍需满足监管审批等交割条件；交易完成后，李飞飞拟出任 AMD 执行副总裁兼首席科学家。World Labs 发布的 Atlas 能够原生处理文本、图像、视频、相机位姿和 3D 信息。世界模型持续处理视频、三维结构、环境状态和连续动作，对显存容量、内存带宽、视频编解码、推理时延及端云协同提出更高要求。当前技术大致形成三维空间生成、像素及视频预测、潜变量预测、物理及行动条件模型四类范式，评价标准也从画面质量延伸至空间一致性、物理合理性、动作可控性和长期预测能力。
- 具身智能的场景、数据和模型同步演进，高价值、低重复任务有望率先形成商业闭环。标准化、高重复任务已有通用或专用设备覆盖，灾害救援、复杂设备维护、柔性制造和家庭照护等非标任务更能体现人形机器人的跨场景价值；近期落地仍以付费主体和回报测算更清晰的 To B 场景为主。训练体系逐步形成真机、仿真和人类视频三线融合：真机数据校准物理误差，仿真数据扩充环境和失败恢复轨迹，人类视频提供任务语义与技能先验。Gemini Robotics 2、GR00T、V-JEPA 2 和 XPACE 已展现跨任务、跨本体及少样本适配能力，VLA 与世界模型开始支持多步规划和失败纠错。行业呈现“GPT-2 时刻”的早期特征，但部分公开多指灵巧手任务成功率仍处 32%至 44%，陌生环境泛化、长任务稳定性和人工接管频次仍是规模部署的核心约束。
- Optimus 由样机验证进入产线建设，制造爬坡、训练算力和真实数据开始形成迭代闭环。Tesla 在 2026 年 Q2 股东材料中确认，Fremont 已停用 Model S 和 Model X 产线并安装第一代 Optimus 生产线，首批产品将进入 Optimus Academy 开展数据采集和功能开发；超过 115MW 的 Cortex 2 已投入运行。The Information 报道称，Optimus 产量由二季度每周数十台提升至 8 月每周数百台，年底目标为建成周产超过 1,000 台的连续自动化产线。当前制造约束集中在灵巧手和前臂装配、触觉传感器可靠性、工装精度、标定效率及供应商一致性。更多机器人进入工位和仓库后，真实任务、失败轨迹和人工接管数据将回流至模型训练、仿真和评测体系，推动可靠性提升与场景复制，产业价值也由核心零部件延伸至自动化装配、检测标定、数据治理、仿真训练和部署运维平台。

相关标的

- 1) 本体：特斯拉、智元、上纬新材、宇树科技。
- 2) 北美链：科森科技、斯菱智驱、福赛科技、岱美股份、恒立液压、蓝思科技、北特科技、浙江荣泰、三花智控、拓普集团、科达利、万向钱潮。
- 3) 其他机器人链：领益智造、泛亚微透、奥比中光、绿的谐波、禾赛、速腾聚创、海康威视、大华股份、宇树科技、上纬新材等。
- 4) 其他相关：协创数据、道通科技、索辰科技、宝通科技、群核科技、智微智能、五一视界等

风险提示

- 世界模型技术路线快速变化的风险；仿真与真实世界存在偏差的风险；具身模型泛化和长任务能力提升不及预期的风险；Optimus 生产和供应链信息存在口径偏差的风险；灵巧手可靠性、产线良率和供应商扩产不及预期的风险；国内 Physical AI 产品商业化不及预期的风险；下游客户资本开支和场景 ROI 不及预期的风险。



内容目录

一、AMD 收购 World Labs，世界模型进入产业投入期	4
1.1 AMD 高价收购 World Labs，模型与算力平台进一步结合	4
1.2 世界模型沿四类范式演进，行动预测成为共同方向	6
1.3 国内企业从物理求解、空间数据和行业仿真三个方向切入	7
二、具身智能进入模型升级期，场景、数据与模型同步推进	8
2.1 场景应用：高价值低重复任务打开具身智能增量空间	8
2.2 数据训练：真机、仿真与人类视频形成三线融合	9
2.3 模型跃迁：VLA 能力持续提升，世界模型开始参与规划和纠错	11
三、Optimus 产量开始爬坡，制造和泛化能力仍是主要约束	13
3.1 产线建设进入新阶段，初始产量主要用于数据采集	13
3.2 灵巧手与精密装配成为量产爬坡的关键约束	13
3.3 Optimus Academy 承接首批产能，训练体系从单机采集走向平台化	14
四、相关标的	15
五、风险提示	15

图表目录

图表 1: AMD 拟以约 82 亿美元收购 World Labs，交易预计于 2026 年底前完成。	4
图表 2: World Labs 的 Atlas 可以生成单一图片的多角度视觉	5
图表 3: World Labs 的 Atlas 可根据用户选择的左右帧图片，构建出空间上下文	5
图表 4: 使用 World Labs 的 Atlas，可依据少量参考图像生成 1440p 分辨率、时长 1 分钟的视频，用户可手动规划摄像机在场景中的移动路径	5
图表 5: 世界模型的四类技术范式对比	6
图表 6: World Labs 按照功能将其划分为渲染器（Renderer）、模拟器（Simulator）和规划器（Planner）	6
图表 7: XPACE 是一个统一的具身世界模型，兼具世界动作模型和世界模拟器的功能	7
图表 8: 群核科技搭建“空间编辑工具-空间数据-空间大模型”的业务飞轮	8
图表 9: 五一视界由自动驾驶仿真向 Physical AI 训练平台延伸	8
图表 10: 具身职能应用场景四象限，人形机器人核心商业价值在于处理“高价值/低重复”任务	9
图表 11: 对高价值/低重复场景的进一步拆解，ToB/C×动脑/动手	9
图表 12: Gemini Robotics 2 从桌面操作延伸至全身移动和操作	9
图表 13: 具身 AI 生态系统 Molmo Spaces 的组成部分的流程图，使用 23 万以上室内场景、13 万以上物体模型、4,200 万条抓取标注。	10
图表 14: Molmo Spaces 用于训练模型所使用的部分仿真场景	10
图表 15: GR00T N1.6 预训练中的训练数据权重	10



图表 16: 数据质量最高且获取成本最高的机器人真机数据极为匮乏.....	11
图表 17: Gemini Robotics 2 覆盖全身控制、灵巧操作和多机器人协同.....	11
图表 18: Gemini Robotics 2 不同任务成功率对比.....	12
图表 19: Optimus 位于费蒙市的第一代产线.....	13
图表 20: Optimus 产线和产量信息.....	13
图表 21: Optimus 灵巧手量产瓶颈及受益环节.....	14
图表 22: Optimus 形成“部署-数据-模型-场景”的迭代循环	15

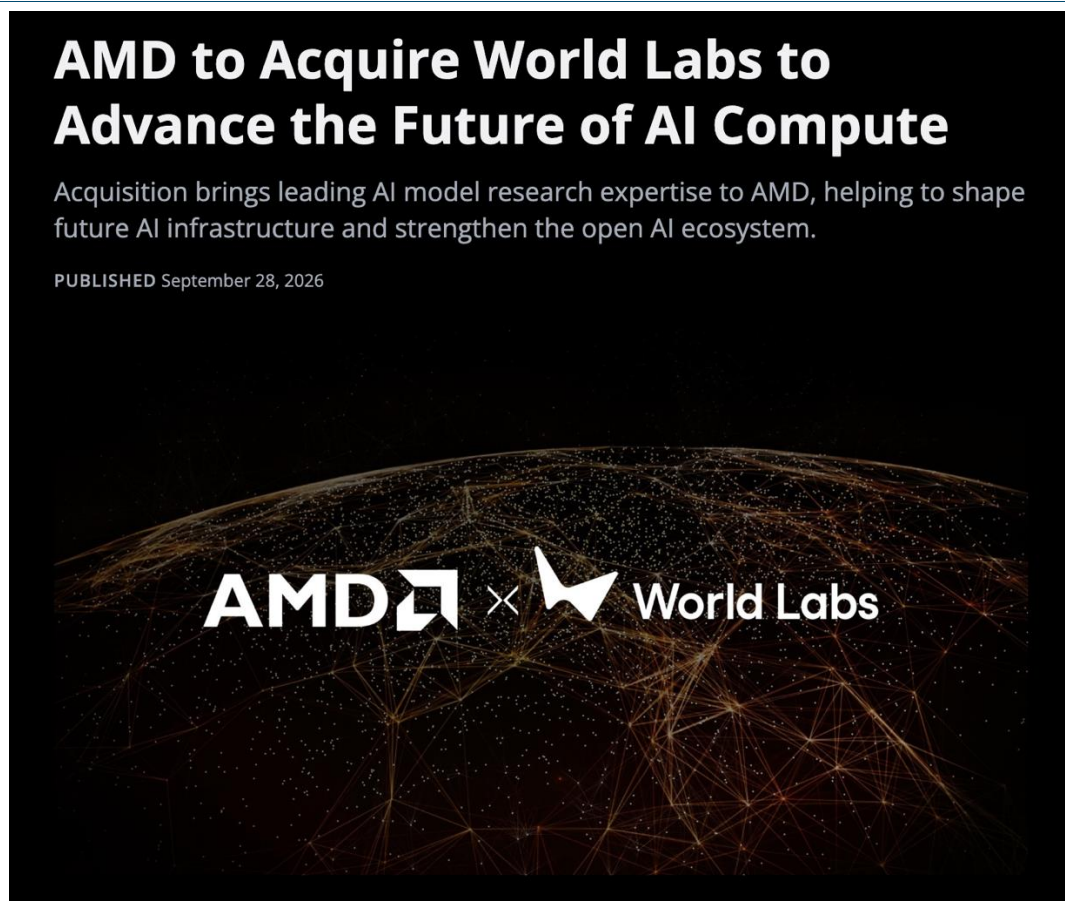


一、AMD 收购 World Labs，世界模型进入产业投入期

1.1 AMD 高价收购 World Labs，模型与算力平台进一步结合

世界模型成为本周 Physical AI 产业最强催化。2026 年 9 月 28 日，AMD 宣布与 World Labs 签署最终收购协议，交易对价约 82 亿美元，采用全股票方式支付，预计于 2026 年底前完成，尚需通过监管审批及其他常规交割条件。交易完成后，李飞飞将加入 AMD 并担任执行副总裁兼首席科学家，World Labs 团队继续开展模型研究。按照 AMD 的官方表述，World Labs 的模型与研究能力将帮助 AMD 理解推理、机器人、仿真及 Physical AI 工作负载的变化，并据此规划未来硬件、软件和系统。

图表1：AMD 拟以约 82 亿美元收购 World Labs，交易预计于 2026 年底前完成。



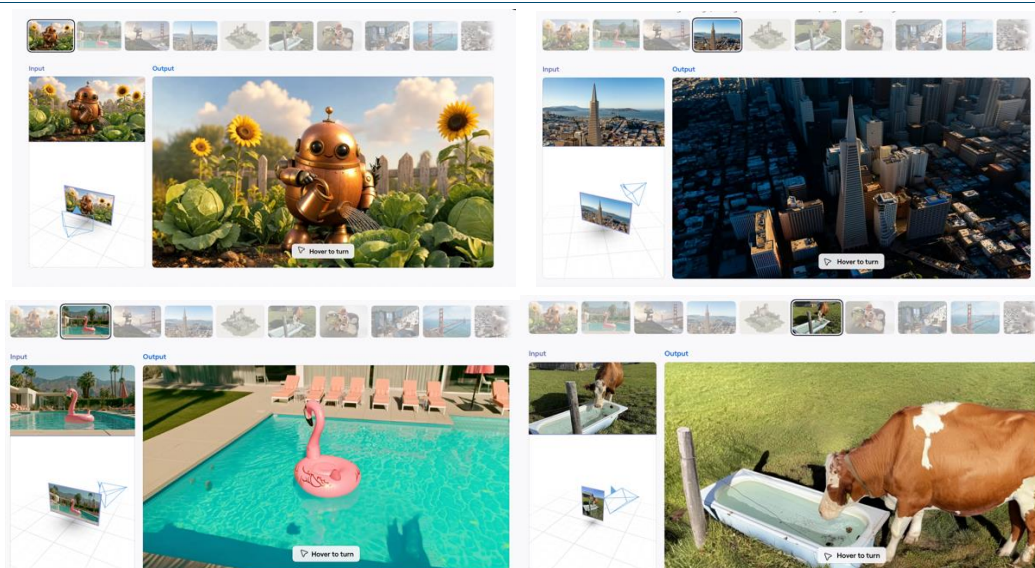
来源：AMD，国金证券研究所

这笔交易的产业意义来自两个层面：

- 世界模型的计算需求与传统大语言模型存在明显差异。大语言模型主要处理离散文本 Token，世界模型需要同时处理高分辨率视频、3D 结构、相机位姿、深度信息、机器人状态和连续动作。模型既要保持空间及时间一致性，也要满足机器人实时控制对低时延的要求。显存容量、内存带宽、视频编解码、并行推理和边缘部署的重要性随之提升。
- 模型与硬件的协同将更早进入芯片研发流程。World Labs 披露，其团队自 2025 年起已与 AMD 围绕 GPU 上的模型训练和推理优化开展深入合作。Atlas 采用多模态自回归扩散 Transformer，既能利用 LLM 领域的 KV Cache、缓存路由和解耦式服务，也能利用图像及视频模型领域的扩散蒸馏、VAE 和无分类器引导等技术。不同计算环节对算力、带宽和存储的需求差异较大，模型团队进入 AMD 后，有望缩短工作负载变化向芯片和系统设计传导的周期。

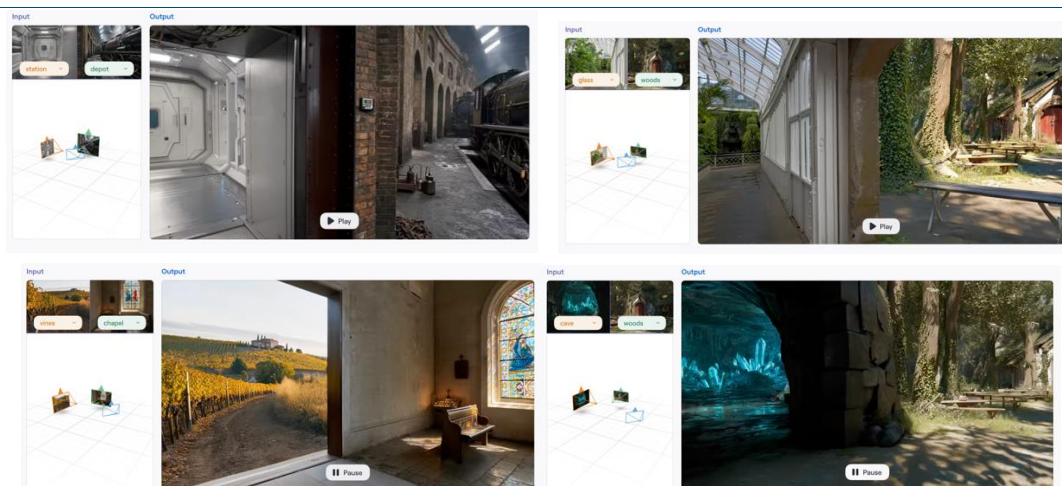


图表2: World Labs 的 Atlas 可以生成单一图片的多角度视觉



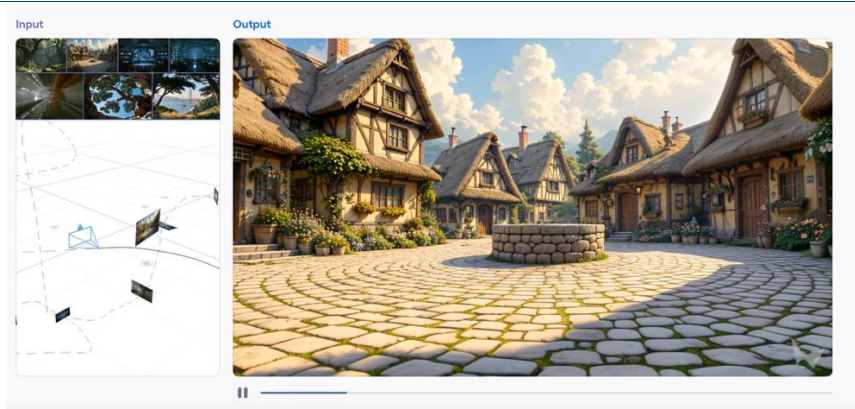
来源: World Labs, 国金证券研究所

图表3: World Labs 的 Atlas 可根据用户选择的左右帧图片, 构建出空间上下文



来源: World Labs, 国金证券研究所

图表4: 使用 World Labs 的 Atlas, 可依据少量参考图像生成 1440p 分辨率、时长 1 分钟的视频, 用户可手动规划摄像机在场景中的移动路径



来源: World Labs, 国金证券研究所



1.2 世界模型沿四类范式演进，行动预测成为共同方向

“世界模型”在当前产业讨论中包含多种技术路线。World Labs 按照功能将其划分为渲染器（Renderer）、模拟器（Simulator）和规划器（Planner）：Renderer 强调生成可供人观看的像素，Simulator 用于根据状态和动作推演环境变化，Planner 则利用内部模型选择行动路径。从技术表示方式看，当前主流方案可以进一步归纳为三维空间生成、像素及视频预测、潜变量预测和物理行动模型四类。不同路线的边界正在相互渗透，评价标准也从单一视觉质量扩展至空间一致性、物理合理性、动作可控性及长期预测能力。

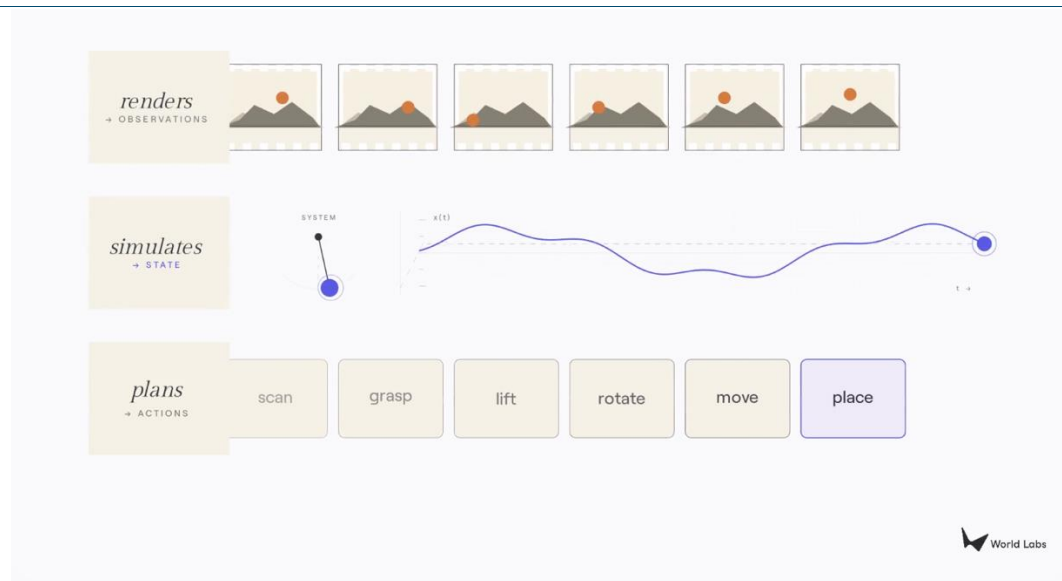
图表5：世界模型的四类技术范式对比

对比维度	三维空间生成	像素及视频预测	潜变量预测	物理及行动条件模型
世界表示	显式3D几何、深度、相机位姿与空间上下文	像素序列或视频帧，依据文本、图像、视频预测未来视觉状态	在抽象表征空间预测与任务相关的状态变化，不逐像素生成	联合建模观察、机器人状态、动作及执行后的视觉或物理结果
代表模型	World Labs Marble、Atlas	NVIDIA Cosmos Google DeepMind Genie	Meta V-JEPA 2	XPENG XPACE World Action Model
主要优势	几何一致性较强；适合场景重建、数字孪生和Real-to-Sim；便于提取深度、可通行区域与物体位置	训练素材丰富；适合合成数据、长尾场景扩充和策略测试；视觉覆盖面广	减少纹理、光影等非关键细节开销；聚焦运动、交互与规划；推理效率较高	可直接评估动作后果；支持策略改进、失败恢复和闭环训练；与机器人控制结合更紧密
主要限制	对动力学、接触关系和长期环境变化刻画仍有限；高质量3D数据获取成本较高	视觉逼真不等于物理正确；摩擦、质量、碰撞等仍需物理引擎或真机校准	表征可解释性较弱；精确控制、长时规划和跨任务稳定性仍待验证	动作与本体强相关；数据采集和对齐复杂；Sim-to-Real及实时推理要求较高

来源：World Labs, NVIDIA, Meta, XPENG Robotics, 国金证券研究所

- 第一类是三维空间生成与重建。World Labs 的 Marble 和 Atlas 属于这一方向。Atlas 将文本、图像、相机位姿及 3D 深度图等信息放入统一空间上下文，通过自回归方式逐步生成多模态结果。显式空间结构有利于保持相机运动下的几何一致性，也便于机器人获得深度、可通行区域和物体位置。该路线适合三维内容生成、数字孪生、场景重建和 Real-to-Sim。当前仍需加强动力学、接触关系以及长时间尺度下的环境变化建模。

图表6：World Labs 按照功能将其划分为渲染器（Renderer）、模拟器（Simulator）和规划器（Planner）



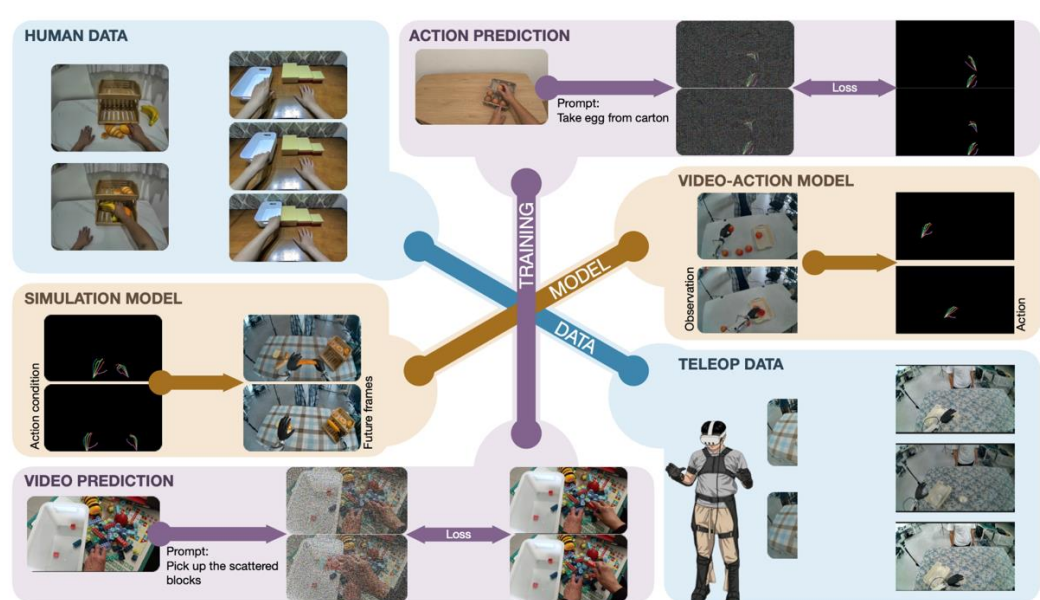
来源：World Labs, 国金证券研究所

- 第二类是像素及视频预测。NVIDIA Cosmos 通过 Text2World、Image2World 和 Video2World 生成未来状态视频，并与 Omniverse、Isaac 等平台结合，用于合成数据、机器人策略测试和自动驾驶场景扩充。Google DeepMind Genie 系列也以可交互视频世界为核心。视频数据规模大、训练素材丰富，适合覆盖现实中难以采集的危险场景



- 和长尾情况；其约束在于，视觉表现符合人类直觉并不必然代表接触、摩擦、质量和动力学参数准确。模型输出进入机器人训练前，通常还需要物理引擎或真实数据校准。
- 第三类是潜变量预测。Meta V-JEPA 2 不直接生成每一个像素，而是在抽象表征空间预测环境状态变化。模型通过大规模自然视频学习物体运动和交互，再使用 DROID 数据集中的 62 小时机器人视频进行行动条件后训练。Meta 披露，V-JEPA 2 可以在新环境中完成抓取、到达和取放等零样本机器人规划任务。潜变量路线减少了对纹理、光影等非关键细节的计算，把资源集中在与动作和状态相关的信息上；精确控制、长时规划和跨任务稳定性仍需更多验证。
 - 第四类是物理及行动条件世界模型。该路线同时预测机器人动作和动作执行后的视觉或状态变化，常被称为 World Action Model。小鹏 XPACE 将世界行动模型和世界模拟器放入统一架构：一条路径根据观察、机器人状态和任务指令预测可执行动作及未来视频，另一条路径根据历史、骨架控制和相机位姿生成未来视频。XPACE 披露使用 5,000 小时具身视频，数据包括第一视角视频、人类视频—动作数据、桥接数据和 IRON 遥操作数据，并通过世界模拟器生成偏离—恢复轨迹。世界模型由此承担策略改进和失败恢复的训练职能。

图表7: XPACE 是一个统一的具身世界模型，兼具世界动作模型和世界模拟器的功能



来源：XPENG Robotics，国金证券研究所

从产业角度看，四类范式可能长期共存。三维生成负责场景和空间结构，视频模型负责扩充视觉变化，潜变量模型负责提取与任务相关的状态，物理引擎及行动条件模型负责验证动作结果，机器人公司可根据任务精度、实时性和成本选择组合。

1.3 国内企业从物理求解、空间数据和行业仿真三个方向切入

国内世界模型相关公司已形成三类基础能力：物理求解器、空间数据和行业仿真场。三类能力对应的客户、产品形态和商业化节奏不同。索辰科技从 CAE 和多物理场求解切入，群核科技侧重大规模室内三维空间及虚拟训练场，五一视界则依托自动驾驶仿真和数字孪生构建行业级 Physical AI 平台。

索辰科技依托 CAE 核心求解器、多物理场算法和工程算例，2026 年发布“大圣”智能体、“行者”合成数据及“营造·万象”世界物理模型，推动 AI 能力由视觉感知延伸至流体、结构、电磁等物理场，可在工业及能源、具身智能和低空经济中提供物理建模、合成数据与仿真验证。群核科技截至 2025 年底已积累超过 5 亿个 3D 场景和 4.79 亿个 3D 模型，并推出 SpatialVerse 机器人虚拟训练场及 SpatialGen 空间生成模型。大规模室内空间资产可以降低训练环境搭建成本、增加场景多样性。五一视界从自动驾驶仿真和数字孪生延伸至 51World Model、51Claw 及具身智能训练平台，具备场景重建、合成数据、仿真测试和部署验证能力，重点覆盖交通、矿山、港口和工厂等行业场景。后续关注具身智能项目落地、场景复用率和平台化收入。三家公司分别覆盖物理规律、空间资产和行业场景，构成国内 Physical AI 基础设施的三类映射。短期主要催化来自新品发布、客户验证和项目落地；中期若具身智能训练由小规模真机采集转向虚实结合，上述能力的调用频率和软件价值量有望抬升。

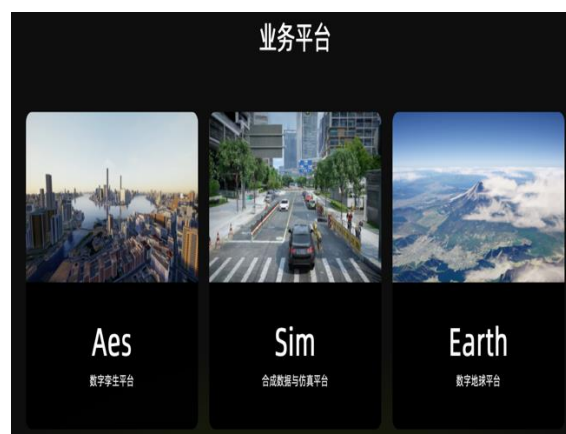


图表8：群核科技搭建“空间编辑工具-空间数据-空间大模型”的业务飞轮

业务层	主要内容
空间编辑工具层	公司开发了图像&视频解析、CAD 识别、物理仿真、3D 渲染器等丰富的工具。
空间数据层	工具的使用沉淀了海量、物理正确的三维可交互数据。公司拥有全球最大的室内场景认知深度学习数据集 InteriorNet，截至 2025 年末积累超过 5 亿个 3D 场景和 4.79 亿个 3D 模型。
空间大模型层	基于海量数据，公司自主研发了百亿级参数的多模态空间大模型，为业界首个专注于 3D 室内场景认知与生成的模型，在真实感全息漫游、可交互性、复杂室内空间处理方面具有独特优势。

来源：群核科技官网，国金证券研究所

图表9：五一视界由自动驾驶仿真向 Physical AI 训练平台延伸



来源：趣解商业，五一视界官网，国金证券研究所

二、具身智能进入模型升级期，场景、数据与模型同步推进

2.1 场景应用：高价值低重复任务打开具身智能增量空间

高重复任务已有成熟设备。商场清洁机器人、仓储 AGV、自动贴标机以及扫地机、洗衣机等通用设备，能够以较低成本完成标准化动作；汽车焊接和喷涂机械臂、半导体天车、高精度数控机床及手术机器人，则依靠专用结构和固定流程处理高价值任务。此类场景强调节拍、稳定性和单位成本，人形机器人短期替代优势有限，机会主要存在于产线仍按人体工学设计、任务频繁切换或专用改造成本较高的环节。

人形机器人的近期重点落在高价值、低重复的非标任务。To B 端包括灾害救援与排爆、非标复杂设备的户外维修，以及柔性制造中的多任务操作；To C 端包括养老和医疗护理、全屋整理、需要判断与灵巧操作的家务，以及高端个性化服务。人形形态可以复用人类工具和作业空间，通用模型则有望降低新任务的编程、调试与部署成本，二者共同提升跨场景复用价值。

进一步按照 To B、To C 以及“动脑、动手”拆分，人形机器人可形成四类方向：To B 端的动手任务对应特种行业，重点解决高风险、高复杂度作业；To B 端的动脑任务对应商业智能服务，覆盖接待、导览和运营服务；To C 端的动手任务对应家庭助手，主要承担家务和照护操作；To C 端的动脑任务对应个人陪伴，核心价值来自交互、看护和情绪支持。现阶段 To B 特种作业及商业服务的付费主体、数据边界和回报测算相对清晰，有望率先验证；家庭助手和个人陪伴空间更大，但仍受安全性、可靠性、隐私、价格及使用频率约束。

低价值、低重复场景是人形机器人能力成熟后的外溢方向，包括非结构化环境中的碎片化长尾任务、物体归位、复杂柔性连续处理和突发干预。商业化路径有望沿“高价值非标任务验证—跨任务复用—硬件与部署成本下降—低价值长尾场景扩展”推进。从投资角度看，建议重点跟踪任务成功率、人工接管频次、单项技能导入时间、每小时作业成本和连续运行时长；当同一机器人能够跨多个流程稳定复用时，模型、数据和软件的价值占比有望上升。



图表10：具身职能应用场景四象限，人形机器人核心商业价值在于处理“高价值/低重复”任务



来源：国金证券研究所

图表11：对高价值/低重复场景的进一步拆解，ToB/C X 动脑/动手



来源：国金证券研究所

图表12：Gemini Robotics 2 从桌面操作延伸至全身移动和操作



来源：Google DeepMind，国金证券研究所

2.2 数据训练：真机、仿真与人类视频形成三线融合

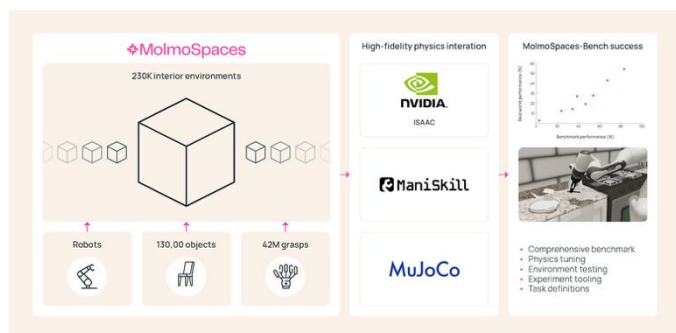
数据仍是具身智能规模化的核心约束。互联网为大语言模型提供了大量文本，汽车保有量为自动驾驶持续产生道路数据；机器人数据通常需要通过遥操作、真机自主运行或仿真生成。机器人本体数量较少、形态差异较大，动作数据还需要与相机画面、关节状态、力反馈和任务指令对齐，采集和清洗难度明显更高。

- 真机数据负责物理接地和失败校准。真实机器人可以记录接触、摩擦、物体形变、传感器噪声、执行器回差和控制时延，这些信息直接决定动作能否落地。其成本来自设备占用、遥操作人员、安全防护和失败后的场景复位。真机采集还存在任务分布偏窄的问题：大量数据来自实验室桌面取放，工业现场中的脏污、遮挡、振动和突发情况覆盖不足。未来需要提高失败数据和恢复轨迹占比，并建立跨本体、跨任务的数据标准。
- 仿真数据负责规模和长尾覆盖。仿真环境可以并行运行大量机器人，系统化调整光照、材质、物体位置、质量、摩擦和相机参数，并安全生成碰撞、跌倒和异常操作等现实中成本较高的样本。Ai2 于 2026 年发布 MolmoSpaces，平台包含超过 23 万个室内场景、13 万个物体模型和 4,200 万条带物理约束的机器人抓取标注；其 MolmoBot 数据



集进一步包含 170 万条专家操作轨迹，覆盖 1.1 万多个物体、9.4 万多个程序化环境和两类机器人平台。Ai2 披露，基于该体系完全在仿真中训练的操作模型可以零样本迁移至真实机器人。该案例表明，场景丰富度和物理参数覆盖提高后，仿真数据的应用边界正在扩大。

图表13：具身 AI 生态系统 Molmo Spaces 的组成部分的流程图，使用 23 万以上室内场景、13 万以上物体模型、4,200 万条抓取标注。



图表14：Molmo Spaces 用于训练模型所使用的部分仿真场景



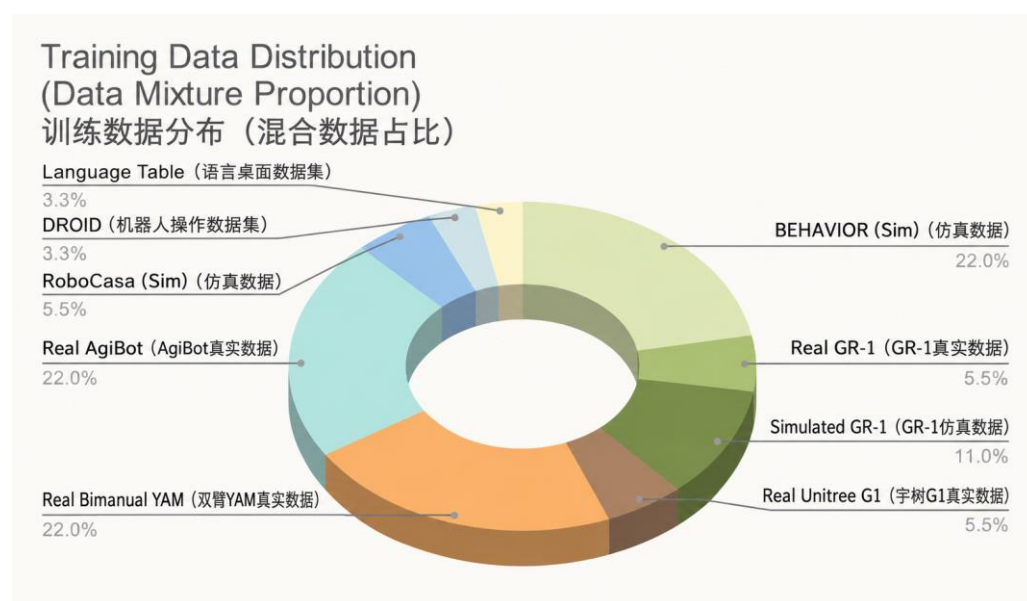
来源：Ai2，国金证券研究所

来源：Ai2，国金证券研究所

- 人类视频负责通用语义和技能先验。互联网视频、第一视角视频和人类示范包含丰富的人—物交互，可以帮助模型理解任务目标、物体用途和动作顺序。其主要难点是缺少机器人关节状态和精确动作标签，人手动作也无法直接映射至不同结构的灵巧手或夹具。当前行业通过人体姿态估计、相机空间动作表示、桥接数据和少量机器人后训练完成对齐。小鹏 XPACE 将第一视角视频、人类视频—动作数据、桥接数据和机器人遥操作数据放入同一数据体系；Meta V-JEPA 2 则先从自然视频学习世界表征，再利用 62 小时机器人视频适配行动规划。

三类数据将形成明确分工：人类视频学习任务和技能，仿真扩充环境和失败恢复，真机数据校准物理误差并验证策略。NVIDIA GR00T N1.6 已经加入数千小时来自双臂机械臂、智元机器人、Galaxea 仿真和宇树 G1 等不同本体的数据。模型在多种机器人上共享视觉和语言能力，再通过小规模任务数据完成后训练。随着数据体系由单本体、单任务扩展到多本体、多场景，仿真平台、空间资产、遥操作设备、数据清洗和评测工具的价值量有望提高。

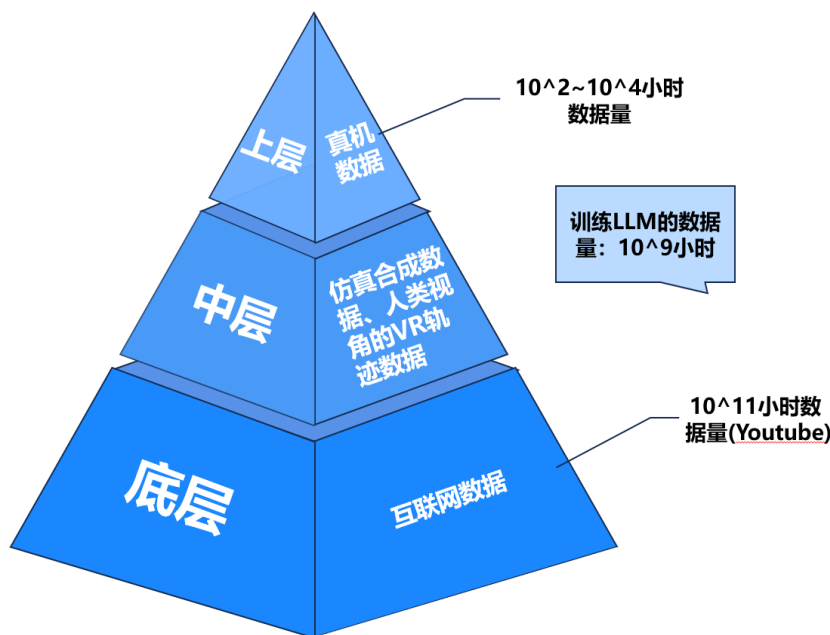
图表15：GR00T N1.6 预训练中的训练数据权重



来源：NVIDIA Research，国金证券研究所



图表16：数据质量最高且获取成本最高的机器人真机数据极为匮乏



来源：OpenLoong 开源社区公众号，世界人工智能大会（WAIC）“人工智能交叉科学论坛”主题活动，国金证券研究所

2.3 模型跃迁：VLA 能力持续提升，世界模型开始参与规划和纠错

VLA 将视觉、语言和动作纳入统一模型，推动机器人控制从任务专用策略走向通用基础模型。模型接收相机图像、语言指令和机器人自身状态，输出连续动作或动作片段。与传统模块化方案相比，VLA 可以利用视觉语言模型的语义知识理解新物体和新指令，并通过大规模预训练提高跨任务迁移能力。当前进展已经从桌面取放扩展至双臂协同、全身移动、灵巧手操作和多步骤任务。

Google DeepMind 于 2026 年 7 月发布 Gemini Robotics 2，包括 VLA 模型 Gemini Robotics 2、具身推理模型 Gemini Robotics ER 2 和端侧模型 Gemini Robotics On-Device 2。VLA 负责将视觉和语言转为机器人动作，ER 2 负责与人沟通、理解环境并规划持续数分钟的任务，端侧模型则在本地设备运行。DeepMind 披露，同一模型检查点可以控制配备不同灵巧手的 Apollo 2 以及使用夹爪的 Franka Duo；端侧模型适配新双臂本体通常只需数小时、少于 200 个示例。

图表17：Gemini Robotics 2 覆盖全身控制、灵巧操作和多机器人协同

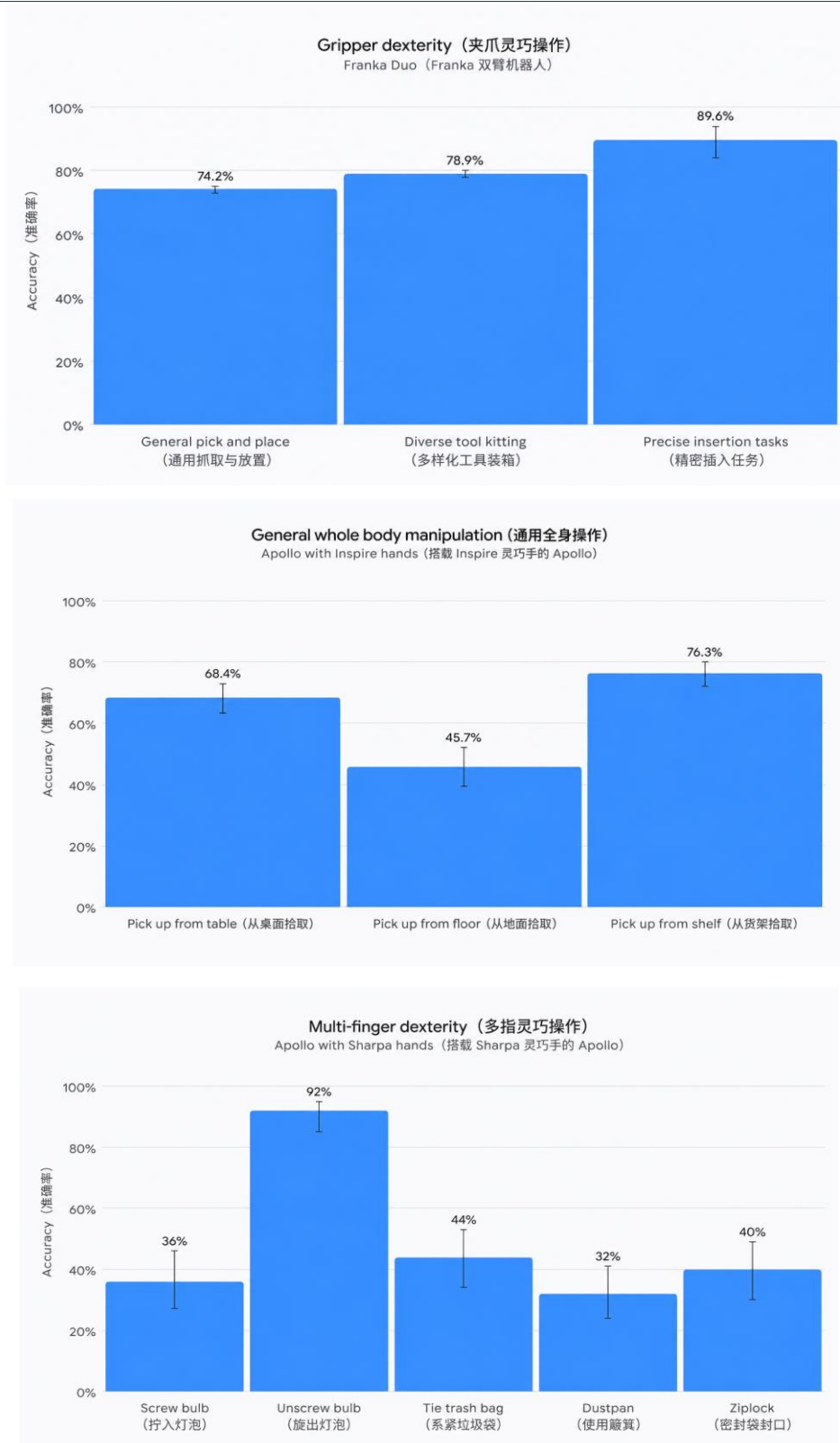


来源：Google DeepMind，国金证券研究所



公开评测同时说明灵巧操作仍有明显提升空间。Gemini Robotics 2 在 Apollo 配合 Inspire 手执行从桌面、地面和货架取物时，公开成功率分别为 68.4%、45.7%和 76.3%；使用 Sharpa 五指手执行旋开灯泡的成功率为 92%，拧入灯泡、系垃圾袋、使用簸箕和封闭密封袋等任务成功率在 32%至 44%之间。双指夹爪在通用取放、工具装箱和精密插入任务上达到 74.2%、78.9%和 89.6%。多指灵巧手能够覆盖更多人类工具，但控制难度和稳定性明显高于夹爪，硬件和模型需要共同迭代。

图表18: Gemini Robotics 2 不同任务成功率对比



来源: Google DeepMind, 国金证券研究所



世界模型补充了 VLA 在长任务中的规划和纠错能力。纯 VLA 根据当前观察直接生成动作，容易在多步骤任务中积累误差；当物体发生意外位移、动作未完成或环境出现新障碍时，模型需要识别进度、推演后果并生成恢复策略。世界行动模型可以针对多个候选动作预测未来画面或状态，再选择成功概率更高的路径。XPACE 通过模拟器生成偏离—恢复轨迹并回用于策略训练，体现了“策略产生动作—模型预测结果—合成恢复轨迹—策略再训练”的迭代方式。

我们将当前阶段概括为具身智能的“GPT-2 时刻”：基础模型开始展现跨任务和跨本体能力，数据量和计算量增加可以带来较为连续的性能提升，少量后训练数据能够适配新设备。该阶段距离广泛商业部署仍有距离，行业也缺少类似语言模型困惑度或统一基准的通用衡量标准。

三、Optimus 量产开始爬坡，制造和泛化能力仍是主要约束

3.1 产线建设进入新阶段，初始产量主要用于数据采集

Tesla 对 Optimus 的官方披露已经从概念和样机推进至产线建设。2026 年 1 月，Tesla 表示计划停产 Model S 和 Model X，将 Fremont 相关产线用于 Optimus，并为年产 100 万台的长期规划做准备。2026 年 Q2 股东材料进一步确认，Model S 和 Model X 产线已经停用，Fremont 正在安装第一代 Optimus 生产线，公司预计于年内启动生产。Texas 的后续工厂也在建设中。

初始产品将优先服务数据采集。Tesla 在 Q2 材料中表示，首批 Optimus 将用于 Optimus Academy，承担训练数据采集和功能开发。Tesla 还披露，2026 年上半年其 Texas 现场训练算力按 MW 计算增加超过一倍；Cortex 2 同时支持汽车和人形机器人的自主控制软件开发，并计划在下半年继续扩容。这意味着早期下线数量需要结合用途理解：内部训练、工程验证和客户测试均可能占据较高比例，短期产量与商业交付量之间存在差异。据 The Information 于 2026 年 9 月 25 日报道，Tesla 在 8 月每周生产数百台 Optimus，较二季度小批量测试阶段的每周数十台提升约十倍；管理层目标是在年底前建立每周超过 1,000 台的连续自动化产线。

图表19: Optimus 位于费蒙市的第一代产线



图表20: Optimus 产线和产量信息

时间	产线及产量进展	数量或目标
2026 Q2	停用 Fremont 工厂 Model S/X 产线，安装第一代 Optimus 生产线	预计 2026 年内启动生产
2026 Q2	首批 Optimus 用于 Optimus Academy	用于训练数据采集及功能开发
2026 年 8 月	Optimus 进入产量爬坡阶段	周产数百台，较 2026 年 Q2 的每周数十台提高约十倍
2026 年底	建设连续自动化生产线	目标周产超过 1,000 台

来源：Tesla，国金证券研究所

来源：Tesla，The Information，Electrek，国金证券研究所

3.2 灵巧手与精密装配成为量产爬坡的关键约束

Optimus 制造瓶颈集中于手部、前臂和精密装配。据 The Information 报道，Optimus V3 的手部和前臂包含超过 100 个螺丝及其他小型组件，目前仍由工人手工装配。产线工装对位、运行速度、触觉传感器可靠性和供应商批量一致性也需要继续改善。Tesla 正在同步调整产品设计、建设自动化产线并验证新供应商，部分供应商还需要配合区域化供应链安排。

- 灵巧手的工程难度来自空间、自由度和可靠性约束同时存在。人类手部可以完成抓取、捏取、旋转、按压和工具使用，人形机器人若希望进入既有人类工作环境，需要兼容门把手、按钮、包装、工具和异形零件。更多自由度能够提高任务覆盖，却会增加微型电机、传动、线束、编码器和传感器数量。驱动和传感单元集中在有限手掌与前臂空间内，还要兼顾重量、输出力、散热、冲击和维护便利性。



- 触觉传感器决定机器人能否稳定处理柔软、易碎和滑动物体。视觉可以判断物体位置，接触后的微小位移、夹持力和滑动趋势仍需触觉及力控补充。传感器需要在灵敏度、量程、耐久性、封装和成本之间平衡；材料批次、粘接工艺及温度变化也可能影响一致性。量产阶段的难点从单个样机“能够工作”转向数百至数千台产品保持相近性能。
- 装配和标定构成另一项约束。大量小型部件提高了人工工时和出错概率，多自由度系统还需要完成零位、力矩、触觉和视觉—动作坐标标定。若每台机器人均需要较长调试时间，产线节拍和制造成本将受到限制。因此，自动化装配、在线检测、标定软件和数字化质量管理可能随产量爬坡获得新增需求。供应链观察指标包括设计冻结时间、零部件定点数量、一次合格率、返修率、供应商扩产及交付周期。

图表21: Optimus 灵巧手量产瓶颈及受益环节

量产瓶颈环节	主要难点	重点受益环节
灵巧手结构	手部及前臂包含超过 100 个螺丝及小型组件，当前仍有较多环节依赖人工装配；零件数量多、装配空间紧凑，对结构精度和产线节拍要求较高	精密结构件、紧固件、灵巧手总成、自动化装配设备
驱动系统	微型驱动单元集中于手掌和前臂，需要兼顾体积、重量、输出力、回差、散热及使用寿命	微型电机、空心杯电机、无框力矩电机、丝杠及精密传动
触觉感知	触觉传感器的灵敏度、耐久性和批量一致性仍需改善，同时需要适应高频接触、冲击和长期磨损	触觉传感器、力传感器、力矩传感器及相关材料
标定检测	多自由度系统需要完成零位、力矩、触觉及视觉—动作坐标标定；个体误差可能影响整机动作一致性和良率	在线检测设备、自动标定系统、测试设备及质量管理软件

来源：Tesla, The Information, Electrek, 国金证券研究所

3.3 Optimus Academy 承接首批产能，训练体系从单机采集走向平台化

Optimus 的制造爬坡开始与数据训练直接衔接。Tesla 在 2026 年二季度股东材料中披露，Fremont 工厂已停用 Model S/X 产线并安装第一代 Optimus 产线，首批产品将进入 Optimus Academy，用于训练数据采集和功能开发。首批机器人由此成为真实世界数据节点：本体数量增加带来更多任务轨迹，模型升级又扩大单台机器人的任务范围，制造、部署和训练逐步形成同一条迭代链路。

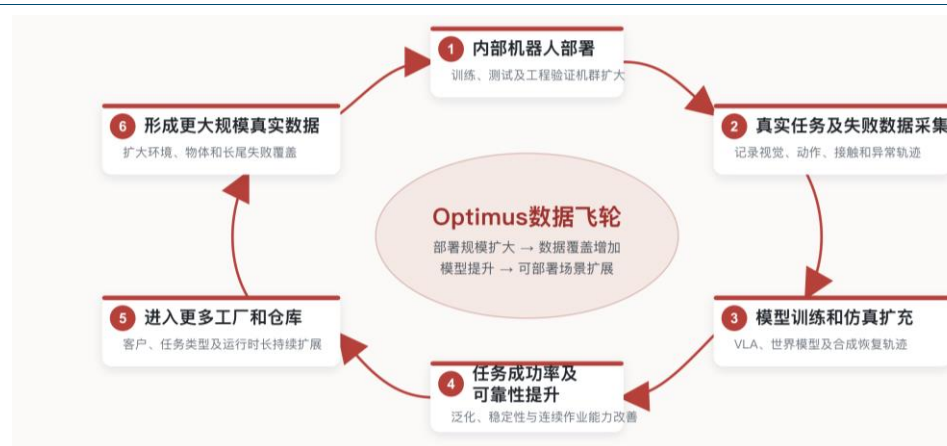
训练方式正在从实验室遥操作走向工厂数据运营。机器人进入真实产线后，需要处理物料位置变化、遮挡、人员干扰、工具差异和任务中断，训练数据也从标准成功轨迹扩展至失败与恢复过程。Tesla 的 Optimus 数据运营岗位已覆盖采集站点建设、日常运营、物流调度和数据质量管理，工厂集成岗位则负责筛选部署任务、调整工位及物料流，并将现场数据回传训练。机器人还需同步记录视觉、语言、关节位置、力矩、触觉和人工接管信号，数据价值因此更多取决于任务覆盖度、困难样本密度和多模态对齐质量。自动标注、轨迹去重、困难样本挖掘和质量评估逐步成为数据引擎的核心环节。

模型侧正在形成“模仿学习+强化学习+仿真+自动评测”的组合架构。真机示范提供可靠的动作起点，仿真环境并行扩充不同物体、光照、摩擦和布局下的轨迹，并反复测试跌倒、碰撞和任务中断后的恢复策略；真机数据再用于修正接触、执行器误差和传感器误差相关岗位信息显示，Optimus 团队正在建设覆盖训练、验证和模型发布的自动化流程，并推进视觉、语言、音频和触觉融合，以及三维理解和世界建模。算力基础设施同步扩张：Tesla 披露，2026 年上半年 Texas 本地 AI 算力按兆瓦计较年初增加一倍以上，装机规模超过 115MW 的 Cortex 2 已投入运行，同时服务汽车和人形机器人自治软件。大规模集中训练与端侧量化、剪枝和蒸馏结合，使复杂模型能够在机器人本体上保持实时运行。

工厂闭环正在改变人形机器人产业的竞争重点。Tesla 内部产线拥有明确工序、稳定数据接口和集中的工程团队，可以连续完成部署、故障复现和模型更新；机器人在搬运、分拣、上下料和简单装配中积累数据，能力稳定后再扩展至更多工位和仓库。跨任务迁移能力提升后，新工位所需的示范、训练和调试时间随之缩短，更多部署又会带来更丰富的数据分布。遥操作、数据治理、仿真训练、VLA 与世界模型、评测系统和运维软件由此连接成完整技术栈，行业竞争也从单机功能展示转向数据运营效率、模型迭代速度和场景复制能力。Optimus Academy 的启动意味着训练平台开始从研发辅助工具升级为连接制造、模型与商业部署的基础设施。



图表22: Optimus 形成“部署-数据-模型-场景”的迭代循环



来源：Tesla、国金证券研究所

四、相关标的

- 1) 本体：特斯拉、智元、上纬新材、宇树科技。
- 2) 北美链：科森科技、斯菱智驱、福赛科技、岱美股份、恒立液压、蓝思科技、北特科技、浙江荣泰、三花智控、拓普集团、科达利、万向钱潮。
- 3) 其他机器人链：领益智造、泛亚微透、奥比中光、绿的谐波，禾赛、速腾聚创、海康威视、大华股份、宇树科技、上纬新材等。
- 4) 其他相关：协创数据、道通科技、索辰科技、宝通科技、群核科技、智微智能、五一视界等。

五、风险提示

■ 世界模型技术路线快速变化的风险。

三维生成、视频预测、潜变量预测及物理行动模型均处快速迭代阶段，主流架构和商业模式尚未稳定，现有产品可能面临技术替代。

■ 仿真与真实世界存在偏差的风险。

合成数据可能无法准确覆盖接触、摩擦、材料形变、传感器噪声和复杂长尾情况，仿真训练效果向真机迁移可能低于预期。

■ 具身模型泛化和长任务能力提升不及预期的风险。

当前公开演示多集中于有限环境和特定任务，陌生场景、长任务、失败恢复及跨本体迁移能力仍需验证。

■ Optimus 生产和供应链信息存在口径偏差的风险。

周产数百台、年底周产超过 1,000 台、灵巧手组件数量及租赁模式等信息主要来自第三方报道，可能与实际进度存在差异。

■ 灵巧手可靠性、产线良率和供应商扩产不及预期的风险。

精密装配、触觉传感器、微型驱动和标定环节可能限制量产节拍，并推高单机成本和维护费用。



■ 国内 Physical AI 产品商业化不及预期的风险。

国内相关产品多处于研发、场景验证和客户导入阶段，订单、收入和利润贡献可能低于预期，较高研发投入也可能继续影响短期盈利。

■ 下游客户资本开支和场景 ROI 不及预期的风险。

若机器人单小时成本、可靠性和部署周期无法满足客户要求，工厂、仓库及行业客户可能推迟采购和扩大部署。



行业投资评级的说明:

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；
增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；
中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；
减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国门内大街 26 号

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服



【公众号】
国金证券研究