

世界模型研究报告

具身智能行业研究报告

湖北社保科创基金管理中心张泽锐

2026 年 8 月 26 日

目录

摘要	1
一、行业定义与核心结论	1
(一) 世界模型不是“世界的三维复制”，核心是动作 后果预测	1
(二) 世界模型与 VLA、仿真器、数字孪生的区别 . .	2
(三) 现阶段的七点判断	2
二、技术逻辑与演进路线	4
(一) 为什么机器人需要世界模型	4
(二) 从桌面操作向轮臂和人形扩展	6
(三) 六项共性挑战	8
三、数据与模型：从样本到策略	10
(一) 数据基础：从字段定义到规模采集	11
(二) 模型训练：从通用预训练到部署回流	14
(三) 模型形态：从未来预测到动作生成	15
四、产业链与产业化进程	16
(一) 产业链结构	16
(二) 需求基础：下游部署开始形成数据与工具需求 .	17
(三) 政策推动：资源从概念研究转向真实部署 . . .	19
(四) 价值分配：不宜用单一市场规模衡量	19

五、竞争格局与商业模式 21

（一）全球代表机构与公司 21

（二）国内代表公司 22

（三）当前竞争格局不适合计算 CR3/CR5 23

（四）资本市场定价：高融资额反映技术预期，不代表
利润池已经形成 23

（五）六种商业模式 24

六、投资建议与企业画像 25

（一）优先关注 25

（二）持续跟踪 26

（三）谨慎或回避 26

（四）优秀企业画像 27

附录一：重点基准测试与公司跟踪清单 28

A 类：必须长期跟踪 28

B 类：产业链重点公司 28

附录二：项目评估时最应该关注的 15 个问题 30

摘要

机器人世界模型的价值，不是把未来画面生成得更逼真，而是预测动作会带来什么后果，并把预测结果用于规划、测试和部署。现阶段，技术应先解决桌面操作中的接触问题，再向轮臂移动操作和人形全身操作扩展。产业机会主要落在两端：一端是数据、仿真和评测基础设施，另一端是能够持续产生真实运行数据的机器人平台。仅有世界模型概念、缺少动作数据和真机闭环的通用模型公司，商业价值仍然有限。

一、行业定义与核心结论

（一）世界模型不是“世界的三维复制”，核心是动作后果预测

“世界模型”并不是对现实世界做一次完整的三维复刻。放在机器人领域，它要解决的是一个更具体的问题：给定当前状态和候选动作，预测下一时刻会发生什么。最简形式可表示为：

$$S_t + A_t \rightarrow S_{t+1}$$

其中， S_t 可以包括 RGB/深度视觉、物体位置和机器人关节状态，也可以加入六维力、关节力矩、触觉等接触信息； A_t 是机器人执行的动作； S_{t+1} 是动作发生后的状态。就机器人应用而言，把世界模型理解为“动作后果预测模型”或“环境动态预测模型”更为准确。

Meta 在 V-JEPA 2 中把理解、预测和规划放在同一条技术链上。机器人版本使用视觉观测和控制动作训练模型，再通过模型

预测控制比较候选动作。1X 的思路相近：1XWM 最初用于策略评估，2026 年进一步被整合进 NEO 机器人的策略模型。

(二) 世界模型与 VLA、仿真器、数字孪生的区别

表 1 世界模型与相关技术的区别

技术	核心问题	典型输入	典型输出	与世界模型关系
VLM / 具身推理	我看到了什么、任务是什么	图像 + 语言	语义/计划	高层理解，不一定输出动作
VLA / 策略	我现在应该做什么	图像 + 语言 + 本体状态	动作	直接生成动作，可能不显式预测后果
世界模型	我这样做之后会发生什么	当前状态 + 动作	未来状态/潜在表征/视频	本报告研究核心
世界动作模型 (WAM)	预测与动作生成如何统一	视觉/语言/状态/动作	未来表征 + 动作	世界模型与策略融合方向
仿真器	按显式物理规律计算后果	完整场景参数 + 动作	仿真状态	规则驱动；世界模型偏数据驱动
数字孪生	复刻特定真实环境/设备	真实资产/测量数据	静态或动态数字映射	不一定具备通用动作预测能力

VLA 回答“现在该做什么”，世界模型回答“这样做会怎样”，两者并不冲突。世界模型既可以作为独立模块参与规划、评估和仿真，也可以被吸收到统一的世界—动作基础模型中。NVIDIA Cosmos 3 已经把视觉推理、世界生成、正向动力学、逆向动力学和机器人策略纳入同一套全模态架构，技术路线正在从单独预测转向预测与动作联合建模。

(三) 现阶段的七点判断

世界模型已经可以参与真实机械臂规划，但距离成熟应用还有明显差距。Meta V-JEPA 2-AC 使用不到 62 小时 DROID 机器人数据，能够在未见环境和未见物体上完成到达、抓取和抓取放

置规划；其新物体、新环境下的抓取放置成功率约为 65%—80%。RoboWM-Bench 2026 记录的空间推理错误、接触预测不稳定和非物理形变，也说明画面逼真并不等于动作可执行。

基础瓶颈首先出现在桌面操作，而不是人形双足。桌面机械臂不需要处理行走和动态平衡，工作空间与自由度也相对有限，但模型仍难稳定预测抓、推、拉、插、滑动和变形等接触后果。只有先把“动作—接触—物体状态变化”这条链路做准，技术才可能可靠地扩展到移动操作和全身人形。

技术演进更适合按任务复杂度来观察：桌面操作、移动操作、人形全身操作。每向前一步，模型都会增加一组新变量。轮臂需要处理底盘移动、动态视角和长时任务；人形还要加入双足平衡、全身耦合和多点接触。ABot-M0.5、MobileWAM 和 DreamTrajectory 都把桌面 WAM 向移动操作扩展作为单独问题处理，这一路径已有较清晰的研究脉络。

近期商业价值主要来自降本增效，而不是单独出售世界模型 API。NVIDIA 用 Cosmos、GPU 和 Isaac 承接训练与仿真需求；Diligent 依靠医院车队积累数据；1X 把网络视频、第一视角数据、仿真、遥操作和在线策略数据接入同一训练流程。短期可量化的指标，是数据效率、测试成本、部署成功率和机器人利用率。

轮臂机器人是一个值得单独观察的载体。它比桌面机械臂多了移动、动态视角和长任务，又避开了双足平衡与全身动力学

的额外负担。Diligent Moxi 已在 25 家以上医院完成超过 125 万次配送，并据此推出 Moxi 2.0 机器人世界模型，初步跑通了“部署—数据—模型更新—再部署”的循环。

一级市场不宜按“是否拥有世界模型”给公司分类。更有区分度的问题是：是否掌握高质量动作数据，能否同步视觉、力觉和触觉，模型效果能否在真机闭环中复现，以及部署数据能否持续回流。纯模型公司面临 Meta、NVIDIA 等平台的挤压，拥有真实部署场景的轮臂和垂直机器人企业更容易建立数据壁垒。

二、技术逻辑与演进路线

本章依次回答三个问题：机器人为什么需要世界模型，技术为什么从桌面操作起步，以及向轮臂和人形扩展时会遇到哪些挑战。

（一）为什么机器人需要世界模型

1. 纯反应式策略的边界

VLA/策略的基本范式可以写成：

$$(S_t, Goal) \rightarrow A_t$$

这类策略根据当前状态和任务直接生成下一步动作，在短任务和训练分布内通常有效。问题会在任务变长、环境开放或接触风险升高后出现：模型未必理解动作后果，也缺少对失败路径的内部评估，许多错误只能等到真机执行时才暴露。

世界模型增加了一个“先想后做”的环节：

$$(S_t, A_t) \rightarrow \hat{S}_{t+1}$$

世界模型提供了一种比较候选动作的办法：先预测各个动作的结果，再选择更接近目标的一项。V-JEPA 2 采用的正是这种滚动方式，每执行一步都重新观察、评分和规划。

2. 世界模型的主要用途

动作规划。在“拿水杯”任务中，模型可以比较从杯口抓、从杯身抓或先调整手腕再抓的结果，而不是一次性给出动作。

失败预判与恢复。当夹爪未对准、杯子开始滑动或抽屉阻力异常时，模型可以判断继续执行的后果，并触发重新规划。

长时任务拆解。Act2Goal 先生成中间视觉状态，再由动作专家执行具体动作，用中间状态连接当前场景与最终目标。

策略评估。1X 最初训练世界模型，是为了在任务和环境不断增加时预测策略表现，减少真机重复测试。

合成数据与仿真。NVIDIA Cosmos、Google Genie 3 和 World Labs Marble 分别从物理智能、可交互环境和三维空间生成切入，可用于构建训练环境和合成数据。

3. 早期适用场景

短期内，世界模型更适合真机试错成本高、任务长尾明显、动作结果又相对容易观测的场景。自动驾驶仿真、桌面或工业操作、医院和仓储移动操作都符合这些条件。跑步、跳跃和动态平衡仍主要依赖强化学习、全身控制与低层运动控制，世界模型更

可能承担环境预测和高层规划，而非替代毫秒级控制器。

(二) 从桌面操作向轮臂和人形扩展

1. 为什么从桌面机械臂起步

桌面机械臂已经把具身问题压缩到相对可控的范围：固定或有限相机视角、固定机械臂底座、较少自由度、局部工作空间、明确物体集合。因此它非常适合验证世界模型最基本的命题：模型能否学习“动作如何改变物体和接触状态”。

相关研究已经从视频预测走向真机规划。V-JEPA 2-AC 把潜在表征世界模型用于真实 Franka 机械臂，Act2Goal 用目标条件视觉世界模型处理长时操作，RoboWM-Bench 则把生成结果能否被机器人执行纳入评价。

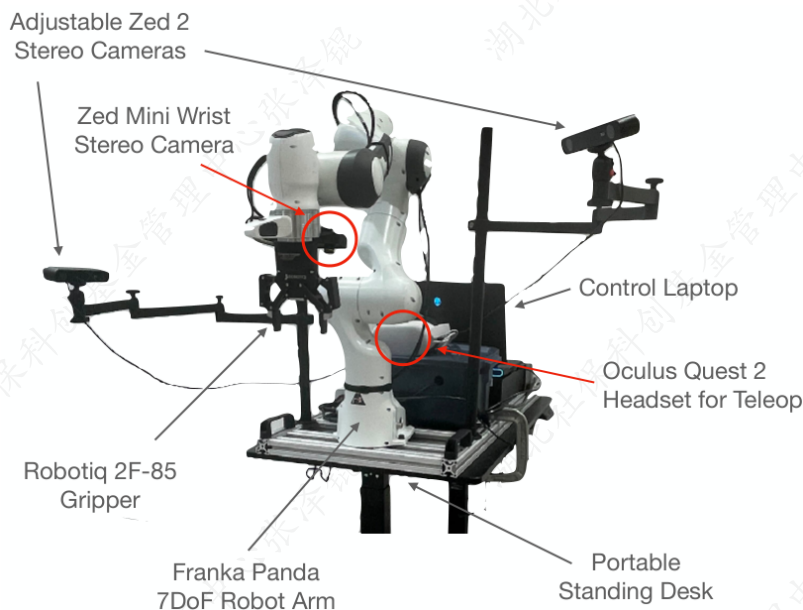


图 1 机械臂视觉采集器示意图

桌面操作主要处理局部接触和状态变化。一旦机器人开始移动，模型就必须同时处理自我运动、动态视角和更长的任务链；

进入人形阶段后，还要加入全身动力学和平衡控制。

2. 移动操作世界模型

轮臂式机器人在桌面能力基础上增加四个变量：**底盘移动、动态相机视角、底盘—机械臂协同、长时空间任务**。模型必须区分“物体移动”和“自我运动造成的视觉变化”，并决定什么时候移动底盘、什么时候移动手臂。

ABot-M0.5 将导航与操作动作解耦，并用 Dream-Forcing 减轻长时预测误差。MobileWAM 2026 通过运动和操作专家、Chain-of-Foresight 以及 ARX Lift2 真机微调，把 WAM 从桌面操作扩展到移动操作。DreamTrajectory 在三个真实任务中的成功率分别为 63.3%、81.7% 和 90.0%。这些结果说明世界模型已经能够改善轮臂任务，但不同任务间的稳定性差异仍然较大。

3. 人形/全身世界模型

人形机器人在移动操作基础上进一步增加：**双足动力学、重心变化、全身耦合、手脚多点接触、几十自由度动作空间以及跌倒安全风险**。对轮臂而言，“世界”主要是外部环境和机械臂接触；对人形而言，自身身体就是一个高度复杂的动态系统。

人形世界模型不是轮臂模型的简单升级。物体与空间先验、操作物理和任务规划可以迁移，但全身动力学、平衡和多点接触仍需要人形本体的专属数据。现阶段把“人形世界模型”作为独立成熟赛道估值，时间仍然偏早。

4. 为什么“先轮臂、后人形”具有产业意义

轮臂机器人介于桌面机械臂和人形之间：它能在跨房间、跨任务以及人机共存的环境中运行，又不需要承担双足平衡的风险。Diligent Moxi 在医院场景积累了规模化运行数据；Skild AI 选择从半结构化工业场景逐步进入医院、酒店和家庭。两条路径都把真实部署视为数据来源，而不是最终交付后的附属环节。

（三）六项共性挑战

无论采用桌面机械臂、轮臂还是人形本体，世界模型都绕不开以下六项基础问题。本体越复杂，这些问题越容易相互叠加。

1. 接触动力学仍是第一瓶颈

机械臂从距离杯子 20cm 移动到 5cm 主要是几何运动；夹爪接触杯子的瞬间，则需要同时理解碰撞、摩擦、刚柔性、力、接触位置、质量和表面材质。很多决定结果的变量无法从 RGB 画面直接推断。RoboWM-Bench 发现当前模型的典型失败正包括不稳定的接触预测和非物理形变。

2. 部分可观测

相同视觉画面可能对应不同真实物理状态：空杯和满杯质量不同，干杯与湿杯摩擦不同，遮挡侧几何结构不可见。因此世界模型真正学习的不是确定函数，而是条件概率：

$$P(S_{t+1}|S_t, A_t)$$

模型不仅要给出预测，也要表达不确定性。一个高置信度的

错误预测，往往比没有预测更容易破坏规划。

3. 状态—动作组合空间爆炸

“杯子”不是单一任务：不同尺寸、材质、摆放姿态、光照、抓取方向、夹爪速度和力度构成指数级组合。机器人不可能真实遍历全部状态，因此必须依靠网络视频、模拟数据、人类第一视角数据和少量机器人交互实现泛化。Meta 的两阶段训练正体现这一思路：先用超 100 万小时互联网视频学习视觉和动态先验，再用不足 62 小时 DROID 机器人数据做动作条件化。



图 2 Meta 视频训练

4. 长时误差累积

训练时每一步可以使用真实状态作为输入，但部署时模型往往要使用自己的预测继续滚动预测：

$$\hat{S}_{t+1} \rightarrow \hat{S}_{t+2} \rightarrow \dots$$

小误差在接触任务中可能导致离散失败：前面 2mm 偏差可能让后面从“成功夹住”变为“撞倒杯子”。因此真实系统通常

采用短时预测 + 重新观察 + 重新规划，而不是一次预测几十秒未来。ABot-M0.5 的 Dream-Forcing 直接针对训练—测试不一致和自回归误差累积设计。

5. “视觉逼真”与“决策有效”不同

机器人并不要求未来画面的每个像素都准确，但动作因果关系必须正确。生成视频即使足够逼真，只要出现夹爪穿模、摩擦方向错误或接触时序偏差，真机动作仍可能失败。因此，评价指标正在从 FVD 和视觉质量转向物理可执行性、任务成功率以及与闭环结果的相关性。

6. 多模态状态表达尚不成熟

理想世界模型输入可包括 RGB、深度、关节角/速度、关节力矩、六维力、触觉和动作；但这些模态频率、坐标系、噪声和数据结构不同。相机可能 30Hz，触觉/力觉可高达数百 Hz 甚至更高，如何进行时间对齐、空间对齐和统一标记化仍是前沿问题。北京市 2025—2027 行动计划已将“多传感器数据时空同步与校准、跨模态学习、规模化自动对齐”列为明确技术任务，反映该问题具有产业共性。

三、数据与模型：从样本到策略

世界模型的开发可以沿着一条链条理解：先确定一条训练样本需要记录什么，再解决数据从哪里来、如何进入训练，最后选择合适的模型表示并与机器人策略连接。

(一) 数据基础：从字段定义到规模采集

1. 样本结构：时间同步的状态—动作—结果轨迹

一条可用于训练世界模型的机器人轨迹，至少要把状态、动作和结果按时间对齐：

$$[Visual_t, Proprioception_t, Force/Tactile_t, Action_t] \rightarrow [State_{t+1}]$$

具体可拆为：

表 2 世界模型训练数据字段

数据层	典型数据	核心作用
外部视觉 Visual	RGB、深度、双目、多视角、点云	物体/空间/姿态/环境
本体感知 Proprioception	关节角、速度、加速度、编码器、IMU	“我身体现在是什么状态”
力觉 Force	六维力/力矩、关节力矩、电机电流	整体受力、负载、接触方向
触觉 Tactile	压力阵列、接触位置、剪切力、滑移	局部接触与精细操作
动作 Action	末端执行器位姿增量、关节目标、夹爪动作、底盘速度	建立“动作—后果”因果对齐
结果/标签 Label	成功/失败、任务进度、未来视觉/潜在表征	训练预测与价值判断

动作字段把世界模型与普通视频模型区分开来。互联网视频能告诉模型世界通常如何变化，却不能直接回答某台机器人执行特定控制指令后会发生什么。

2. 数据来源及训练价值

表 3 不同数据来源的训练价值

来源	规模潜力	物理真实性	动作标签	成本	世界模型价值
互联网视频	极高	高	无	低	学通用视觉/动态先验

续表

来源	规模潜力	物理真实性	动作标签	成本	世界模型价值
第一视角人类视频	高	高	人类动作隐式/可估计	中	学操作过程和空间行为
遥操作真机数据	中	最高	有	高	动作条件化核心数据
随机操作	中	最高	有	高	学失败/非最优动作后果
机器人自主运行/在线策略	随部署扩大	最高	有	边际成本下降	最有价值的数据飞轮
仿真数据	极高	中	完备	低	补长尾、并行训练、RL

1X 的训练数据包括网络媒体、人类第一视角视频、仿真、灵巧遥操作和 NEO 在线运行数据。Skild 则认为遥操作数据难以单独支撑基础模型规模，因此先用大型仿真和互联网视频预训练，再用针对性的真实数据做后训练。两家公司都没有把某一种数据源视为完整答案。

3. 数据采集硬件与基础设施

视觉/深度。奥比中光 2026 年推出无机器人数据采集硬件平台。EGO、UMI、WristCam 和 Hub 分别覆盖第一视角、手持操作、腕部近场与数据同步，其中 EGO RGB-D 面向桌面精细操作的 VLA 和世界模型训练。

六维力/关节力矩。坤维传感的产品覆盖六维力和关节扭矩。公司披露，2024 年已向智元、银河通用和北京人形机器人创新中心供货，2026 年建成 10 万台/年的六维力自动化产线。除闭环控制外，这类传感器还可以记录接触状态，为世界模型提供训练数据。

触觉/全模态采集。帕西尼同时布局多维触觉、灵巧手、数据采集和 OmniVTLA。其 OmniSharing DB 覆盖多维触觉、多视角视觉、语音、文本、自体感知和空间轨迹，公司同时建设 Super EID Factory。“百亿级数据规模”为公司公开披露的数据。



图 3 帕西尼数据采集

数据与仿真平台。Lightwheel 提供 SimReady 资产、EgoSuite 人类数据、RoboFinals 评测和 Real2Sim2Real 平台。公司披露 2026 年第一季度订单约 1 亿美元，业务集中在物理智能所需的数据、仿真和评测基础设施。

4. 传感器配置取决于任务

世界模型会增加多模态数据需求，但传感器配置仍取决于任务。视觉是基础配置；六维力和触觉更适合接触频繁、精细操作、柔性物体或安全要求较高的场景。判断相关公司的价值，应看产品是否进入客户的训练流程、接口能否跨客户复用，以及量

产价格与毛利，而不能把“世界模型需要触觉”直接换算成估值溢价。

(二) 模型训练：从通用预训练到部署回流

1. 典型训练流程

常见训练流程可以拆成六步：

- 通用预训练：大量互联网视频/图像学习物体、运动、空间和基本物理先验；
- 具身中期训练：第一视角人类操作、仿真或跨机器人数据学习动作结构；
- 机器人动作条件化：使用机器人状态 + 动作 + 结果对齐具体本体；
- 任务/场景后训练：对目标场景进行 SFT、LoRA、策略训练；
- 真实部署回流：在线策略数据、失败案例、长尾环境进入回放缓冲区；
- 持续训练与评估：避免灾难性遗忘，验证新模型是否真正提升旧任务和新任务。

2. 三个训练难题

持续学习。新数据加入后，旧能力可能退化。工程上通常采用回放或混合训练、冻结基础参数、LoRA/适配器和多任务评测，同时观察新旧任务的成功率变化。

跨本体。不同机器人的动作空间不同，关节数值不能直接混合。常见方案包括统一末端执行器表示、动作标记器、本体标记

以及专属动作头或适配器。Skild 采用“全本体”路线；自变量 2026 年发布 X-Tokenizer，也在尝试统一跨模态动作表示。

数据混合。仿真、遥操作、成功示教、失败轨迹、在线策略数据和第一视角视频的质量与分布差异很大，不能按原始条数简单随机采样。数据配比、质量过滤、跨本体对齐和持续训练能力，比累计时长更能影响最终效果。

(三) 模型形态：从未来预测到动作生成

1. 多种模型表示将长期并存

表 4 世界模型主要技术路线

路线	核心表示	优势	主要问题	更适合场景
潜在表征预测模型	预测未来潜在表征	计算效率高，过滤无关像素	可解释性/可视化弱	在线规划、短时控制
视频生成式世界模型	预测未来视频	直观、可做仿真/合成数据	像素成本高、物理一致性难	数据生成、评估、可视仿真
三维/四维空间世界模型	预测空间结构/动态场景	空间一致性强	接触与动作动力学仍难	环境建模、导航、数字空间
世界动作模型	世界预测与动作联合	直接服务策略	训练复杂、易受本体差异影响	桌面/移动操作
统一物理智能模型	文本/视觉/视频/动作统一	有望形成通用底座	算力高、路线未定型	长期平台化

这几条路线承担的任务不同。视频和三维模型适合离线仿真与数据生成，潜在表征适合实时短时预测，世界动作模型则直接连接策略。NVIDIA Cosmos 3 尝试把它们放入统一架构，但专用模型在计算成本和部署效率上仍有存在空间。

2. 从世界模型到世界动作模型

世界模型正在从独立预测器向策略模块靠拢。V-JEPA 2 用世界模型辅助规划；1XWM 把视频预训练的世界模型直接用于机器人策略；Cosmos 3 尝试统一世界建模、推理和动作。国内的 WALL-A/WALL-WM 与 Act2Goal 也沿着预测和策略融合的方向推进。

因此，“纯世界模型公司”的边界可能越来越模糊。相关能力可以嵌入 VLA、机器人基础模型、仿真平台或整机软件栈。独立模型供应商若想形成稳定收入，需要跨本体服务多个客户，并进入客户的长期 workflow。

四、产业链与产业化进程

本章先明确世界模型在产业链中的位置，再讨论下游需求、政策推动和价值分配。这样既能判断产业化进度，也能区分技术价值最终由哪些环节承接。

（一）产业链结构

世界模型产业链可以分为三层：上游提供算力、传感器、数据采集和仿真工具；中游负责模型训练、评测及世界模型与策略的融合；下游由机器人整机、服务平台和工业、物流、医疗等应用场景构成。三层之间并非简单采购关系，真实部署产生的数据还会持续回流到上游和中游。

具身智能世界模型产业链与技术演进

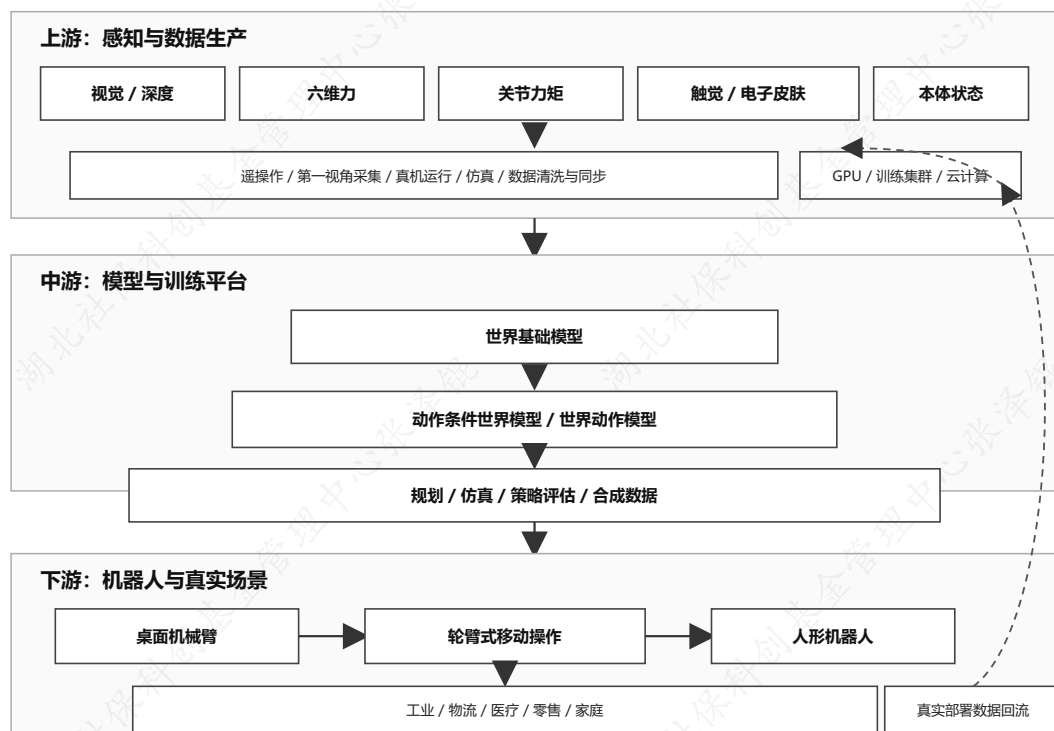


图 4 世界模型产业链

(二) 需求基础：下游部署开始形成数据与工具需求

1. 已有的下游产业基础

IFR《World Robotics 2025》统计，全球已有 944 家服务机器人生产企业；2024 年专业服务机器人销量超过 19.9 万台，机器人即服务车队增长 31%，达到 2.45 万台以上，物流运输机器人仍是最大应用组。世界模型不会覆盖所有产品，但现有开发者和部署规模已经能够形成数据、仿真与模型需求。

Diligent Moxi 已在 25 家以上医院完成超过 125 万次配送。其收购方 Serve 披露，Diligent 拥有接近 100 台机器人，单家医院年

度销售预期约 20 万—40 万美元。持续运营收入为数据积累和软件迭代提供了基础，也使世界模型有机会嵌入真实服务流程。

2. 三类直接需求

第一类需求来自数据。真实机器人数据稀缺且昂贵，网络视频、第一视角人类数据和仿真虽然不能完全替代真机数据，却可以减少新任务所需的采集量。

第二类需求来自测试。模型承担的任务越多，回归测试成本越高；物体、光照、场景和失败条件无法全部依靠真机反复验证，仿真和世界模型评测因而进入开发流程。

第三类需求来自任务复杂度。机器人正从单步动作走向长时任务。“抓杯子”和“找到杯子—拿起—移动—放到指定位置—失败恢复”不是同一个难度等级，后者需要持续跟踪未来状态和任务进度。

3. 各子方向所处阶段

表 5 世界模型各子方向发展阶段

子方向	当前阶段	判断
桌面操作世界模型	技术验证后期/工程攻关期	已能规划和泛化，但接触/长时/闭环不足
移动操作世界模型	早期验证/早期商业化	2026 年出现专用 WAM 和真实部署世界模型
人形世界模型	前沿探索期	大量能力仍需从桌面/轮臂继承后再扩展
数据采集硬件	快速产品化期	视觉、力觉、触觉、无本体采集加速推出
数据/仿真平台	商业化早期	客户需求开始增加，收入模式仍在形成
通用世界基础模型	平台竞争期	NVIDIA/Meta/Google 开源或平台化压力大

世界模型已经具备技术可行性，但在产业链中的位置尚未固

定。一级市场可以从基础设施和数据闭环切入，估值上不宜直接套用成熟软件或机器人公司的收入倍数。

（三）政策推动：资源从概念研究转向真实部署

北京《具身智能科技创新与产业培育行动计划（2025—2027年）》将多传感器数据同步与校准、跨模态融合、未来状态预测、世界模型仿真平台和多模态数据采集平台列入任务。

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》（工信厅联科函〔2026〕256 号）提出，到 2026 年底凝练形成百个以上高价值应用场景，并在实景实训中积累高质量真机数据，记录全身运动轨迹、力位控制曲线、操作执行序列及异常工况等信息。

从政策工具看，资源主要投向真实场景开放、多模态数据采集、训练与仿真平台、关键传感器和机器人本体，以及数据标准建设。受益者不只包括模型公司。地方数采场和训练中心如果缺少终端需求，仍可能出现重复建设；数据数量很大，也不等于具有同等训练价值。

（四）价值分配：不宜用单一市场规模衡量

1. 世界模型对应多个价值池

截至 2026 年，世界模型还没有统一的产品定义、收入分类和统计口径。第三方报告常把视频生成、数字孪生、机器人仿真、自动驾驶仿真、空间智能和物理智能放在同一市场中统计。由此

得到的单一市场规模数字，很难对应可投资、可确认收入的具体产品。

按收入和价值承接位置，可以拆成五部分：

- 1. 上游传感器增量市场：视觉/深度、六维力、关节力矩、触觉及数采硬件；
- 2. 数据基础设施市场：遥操作、数据采集、清洗、标准化、数据管理、仿真和评测；
- 3. 模型/软件市场：世界基础模型、WAM、授权/API/私有部署；
- 4. 机器人内部价值：世界模型提升的任务成功率、机器人利用率、部署效率和数据效率；
- 5. 算力市场：训练、仿真、推理带来的 GPU 与云计算需求。

机器人内部价值和算力支出可能大于独立世界模型软件收入。只统计模型 API，会漏掉世界模型对整机效率和算力需求的影响；反过来，也容易高估纯模型公司的利润份额。

2. 利润池判断

表 6 世界模型产业链利润池判断

环节	当前利润确定性	壁垒	主要风险	投资判断
GPU/算力	高	芯片 + 生态	资本开支周期、国产替代	已形成确定利润池
视觉/深度	中高	算法 + 芯片 + 量产 + 客户	ASP 下降、同质化	关注“感知 + 数采”双重逻辑
六维力/触觉	中	精度、可靠性、量产、客户	机器人渗透率、单机配置不确定	精细操作增量方向

续表

环节	当前利润确定性	壁垒	主要风险	投资判断
数据采集/仿真	中高潜力	数据标准、工具链、客户流程	项目制、客户自研	基础设施重点方向
通用世界模型	低一中	模型 + 数据	开源、算力、大厂竞争	技术价值高，独立利润池未证实
WAM/机器人基础模型	中	数据闭环 + 跨本体 + 真机效果	路线快速迭代	重点跟踪
轮臂整机/服务	中高	部署、场景、数据飞轮	硬件毛利、运维	当前较易形成商业闭环
人形整机	低一中	全栈	技术/成本/需求不确定	高弹性、高风险

从价值承接看，两类公司相对清晰：一类是能够服务多个模型和多种本体的数据、仿真与评测平台；另一类是已经进入真实场景，并能用世界模型提高成功率和利用率的机器人平台。缺少独占数据和下游闭环的世界基础模型，容易受到 Meta、NVIDIA 和 Google 开放模型的挤压。

五、竞争格局与商业模式

（一）全球代表机构与公司

表 7 全球代表机构与公司

公司/机构	主要位置	代表产品/路线	当前证据	商业化判断
Meta	潜在表征世界模型	V-JEPA 2 / V-JEPA 2-AC	62h 机器人数据、真实 Franka 规划	开源研究底座，直接收入弱
NVIDIA	物理智能平台	Cosmos 3 + Isaac	世界/推理/动作统一，模型与框架开放	通过 GPU、云、生态变现最清晰

续表

公司/机构	主要位置	代表产品/路线	当前证据	商业化判断
Google DeepMind	通用交互世界模型	Genie 3	20—24fps、720p、分钟级交互	智能体训练潜力高，机器人动作仍有限
World Labs	空间世界模型	Marble / World API	持续性三维世界；2026年融资 10 亿美元	API 已产品化，机器人应用仍在拓展
1X	机器人世界模型	1XWM + NEO	从评估扩展至策略；建立 World Model Lab	数据—模型—人形全栈
Skild AI	全本体基础模型	Skild Brain	覆盖桌面、轮臂、人形、四足	跨本体路线；2026 估值 >140 亿美元
General Intuition	动作标注世界模型	MIRA	游戏动作数据，20fps 多人世界模型	虚拟数据 → 现实机器人路线，前沿高风险
Diligent Robotics	商业轮臂世界模型	Moxi 2.0	医院真实车队、学习飞轮	当前最强商业闭环案例之一

(二) 国内代表公司

表 8 国内代表公司

公司	位置	代表能力	公开进展	研究价值
自变量机器人	模型 + 本体 + 数据	WALL-A/WALL-B/WALL-WM/X-Tokenizer	2026 年发布事件级世界模型和跨模态动作标记器；企业披露估值超过 200 亿元	国内世界模型/VLA 融合重点
智元机器人	模型 + 本体	Act2Goal	目标条件世界模型 + 动作专家；长时任务与在线自我提升	桌面操作世界模型重点
ACE Robotics	世界模型 + 机器人	Kairos	企业披露多个基准测试领先；计划扩展真实数据	新兴项目，处于基准测试与商业化探索阶段

续表

公司	位置	代表能力	公开进展	研究价值
奥比中光	视觉/数据基础设施	EGO/ UMI/ WristCam/ Hub	明确面向 VLA/世界模型数据采集；2026 扩展无本体数采	上游视觉与数据平台重点
坤维传感	力觉	六维力/关节扭矩	企业披露供货智元/银河通用等，10 万台/年产线	接触数据和力控关键零部件
帕西尼	触觉 + 数据 + 模型	触觉/OmniSharing DB/OmniVTLA	全模态数据平台；企业口径百亿级数据	触觉数据闭环重点，商业收入处于形成阶段
Lightwheel	数据/仿真/评测	EgoSuite/ RoboFinals/ SimReady	企业披露第一季度订单约 1 亿美元	独立物理智能基础设施样本
厘清智能	物理智能基础设施	跨本体智能底座	与奥比中光合作世界模型数据/三维深度	早期项目，重点看跨本体交付

(三) 当前竞争格局不适合计算 CR3/CR5

技术领先、模型使用量、商业收入和整机部署目前无法放在同一口径下比较。Meta 的世界模型研究领先，但没有对应的直接收入；Diligent 把模型价值内化在医院机器人服务中；NVIDIA 通过 GPU 和云服务变现；World Labs 则直接按 API 收费。用融资额或单项基准测试计算市场份额，会混淆不同商业模式。

更有意义的比较维度是：数据规模与质量 → 动作对齐能力 → 物理一致性 → 长时预测 → 跨本体 → 真机任务成功率 → 真实部署数据飞轮 → 商业付费与复购。

(四) 资本市场定价：高融资额反映技术预期，不代表利润池已经形成

表 9 代表公司资本市场定价

公司	最新公开资本信号	公开口径	投资含义
Skild AI	2026 年 1 月 C 轮融资 14 亿美元	官方披露估值超过 140 亿美元	市场愿意为“跨本体通用机器人基础模型”给予极高期权价值，但收入/利润需与估值匹配
World Labs	2026 年 2 月新增融资 10 亿美元	官方未披露本轮估值	空间世界模型融资能力强，机器人商业化处于早期
General Intuition	A 轮融资 3.2 亿美元	官方披露	动作标注游戏数据 → 现实世界迁移获得资本认可，仍属前沿技术押注
自变量机器人	2026 年 1 月 A++ 轮 10 亿元；后续 B/B+/B++/C 连续完成	企业官网披露估值突破 200 亿元	国内对具身基础模型 + 本体全栈给出高定价，重点关注真实收入、客户与模型增益
帕西尼	2026 年 3 月 B 轮融资超 10 亿元	企业官网披露估值突破 100 亿元	触觉 + 数据 + 模型一体化获得高定价，但需区分传感器收入、数据收入与模型收入
Lightwheel	2026 年第一季度企业披露订单约 1 亿美元	企业口径，非审计收入	数据/仿真基础设施出现大额需求信号，后续关注订单质量和收入确认

资本市场对物理智能基础设施和机器人基础模型给出了较高预期，但公开收入和利润仍落后于融资规模。估值时应把融资信息与真实付费、数据复用性、训练成本、客户留存、真机任务提升和持续融资需求分开观察。

（五）六种商业模式

1. 模型 API。 World Labs 已推出 World API，General Intuition 也开始向游戏、仿真和机器人合作伙伴开放商业 API。这种模式容易标准化，但也最容易受到开源模型和大厂平台的价

格压力。

2. 私有化部署和授权。面向机器人、自动驾驶和工业企业提供定制后训练与部署，适合对安全和数据隐私要求较高的客户，收入形态往往偏项目制。

3. 数据采集硬件与数据服务。奥比中光、帕西尼和 Lightwheel 等公司正在把采集、清洗和管理工具产品化。持续收入取决于数据标准能否进入客户的日常训练流程。

4. 仿真和评测平台。如果平台能进入每个策略版本的回归测试和安全验证流程，客户粘性通常高于单次 API 调用。这是独立软件层较有机会形成持续收入的模式。

5. 整机或机器人即服务。Diligent Moxi 和 1X NEO 通过销售或订阅获得收入，世界模型不单独收费，而是用于提高任务成功率和设备利用率。

6. 算力平台。NVIDIA 开放 Cosmos，扩大物理智能训练、仿真和推理需求，再由 GPU、DGX Cloud、Isaac 和 NIM 承接收入。

六、投资建议与企业画像

（一）优先关注

1. 多模态具身数据基础设施。关注能够采集 RGB-D、本体状态、动作、力觉和触觉，并完成硬件同步、自动清洗、标准化及跨本体转换的平台。单纯依靠采集人力的公司，复用性和毛利空间都较有限。

2. 生成式仿真和评测。关键指标是仿真中的策略排序能否在真机中复现，也就是“仿真里更好，真机也更好”。相比单纯生成视频，评测更容易进入刚性开发流程。

3. 轮臂机器人及其部署数据。医疗、仓储、零售和工业场景更容易形成稳定任务、持续数据和机器人即服务收入。世界模型的价值可以直接反映在成功率和利用率上。

4. 世界模型与策略/VLA 融合。关注能够用真机结果证明世界预测改善了长时任务、跨场景泛化和失败恢复的团队。

（二）持续跟踪

视觉、六维力和触觉传感器。需求方向明确，但价值取决于单机配置、客户渗透率、量产价格和毛利，不能仅凭“多模态世界模型”概念定价。

通用世界基础模型。技术价值较高，同时资本开支大、开源竞争强。现阶段更适合观察 API、授权收入和客户留存。

（三）谨慎或回避

- 纯视频生成模型包装成“机器人世界模型”，没有动作条件化和真机验证；
- 只有论文基准测试，没有真实机器人闭环任务；
- 数据平台依赖地方训练场采购，没有真实机器人公司复购；
- 号称“百万/亿级数据”但无法说明数据单位、有效比例、动作标签和训练增益；
- 模型公司缺少独占数据或客户 workflows，仅靠公开数据 + 开

源主干网络；

- 高估值人形世界模型项目，但连桌面操作物理一致性和长时成功率都未证明。

（四）优秀企业画像

表 10 优秀企业评价标准

维度	优秀标准
技术	动作条件预测、长时稳定、接触一致性至少一项领先
数据	多模态同步、包含失败/在线策略数据、持续真实回流
跨本体	有统一动作表示/适配器，并已有 2 种以上本体验证
真机	公开真实机器人成功率、未见物体/未见场景测试
商业	有付费客户和复购，不依赖科研演示
数据效率	新任务适配所需真机数据持续下降
闭环价值	能量化提升成功率/降低测试成本/缩短部署周期
组织	AI、机器人控制、数据工程、硬件系统能力均衡
现金流	训练资本开支可控，融资依赖度与模型收入匹配
估值	充分计入技术不确定性，不以“基础模型故事”替代收入表现

附录一：重点基准测试与公司跟踪清单

A 类：必须长期跟踪

1. **Meta V-JEPA 2**：潜在表征预测是否持续提升机器人规划和物理推理；
2. **NVIDIA Cosmos**：是否形成物理智能标准底座、模型是否进一步吸收策略；
3. **1X**：世界模型直接策略路线能否在家庭复杂任务规模验证；
4. **Skild AI**：全本体路线是否证明跨桌面/轮臂/人形正迁移；
5. **Diligent Robotics**：医院真实数据飞轮能否提高 Moxi 任务完成和利用率；
6. 自变量机器人：WALL-WM/WALL-B/X-Tokenizer 是否形成真实商业场景提升；
7. 智元 **Act2Goal**：目标条件世界模型在长时真实操作上的持续表现。

B 类：产业链重点公司

1. 奥比中光：无本体数采硬件能否从设备销售升级为数据标准平台；
2. 坤维传感：六维力在人形/操作机器人单机配置和量产渗透率；
3. 帕西尼：触觉硬件、数据和模型一体化是否形成外部客户收入而非内部循环；
4. **Lightwheel**：数据/仿真/评测订单能否转化为持续收入及

高毛利；

5. **World Labs:** 空间世界模型能否从三维生成进入机器人训练和仿真；
6. **General Intuition:** 游戏动作数据能否有效迁移至物理智能；
7. **厘清智能:** 跨本体物理智能基础设施能否形成标准化交付。

附录二：项目评估时最应该关注的 15 个问题

1. 你们对世界模型的定义是什么？输入/输出到底是什么？
2. 模型是否显式使用动作条件化？动作空间怎么定义？
3. 训练数据中互联网视频、人类第一视角、仿真、遥操作、在线策略分别占多少？
4. 真机交互数据多少小时/回合？成功、失败、随机操作比例？
5. 是否使用深度、六维力、关节力矩、触觉？加入后成功率提升多少？
6. 新数据加入后旧任务是否出现灾难性遗忘？如何评测？
7. 不同本体如何统一动作表示？是否有跨本体真机实验？
8. 最长稳定滚动预测多长？超过该长度后误差如何增长？
9. 你们最难的接触动力学任务是什么？成功率多少？
10. 世界模型预测的策略排序与真实机器人测试相关系数/一致性是多少？
11. 与不使用世界模型的 VLA 基线相比，真实任务成功率提高多少？
12. 世界模型是否减少真实机器人数据需求？减少比例是多少？
13. 模型训练、推理、仿真的单位 GPU 成本及下降趋势如何？
14. 商业模式是 API、授权、数据、仿真平台还是整机内化？已有多少付费/复购客户？
15. 如果 NVIDIA/Meta 明天开放性能相近模型，你们还有什么不可替代壁垒？