下载日志已记录，仅供内部参考

汽车

报告日期：2026 年 10 月 28 日

具身智能的核心“燃料”：数据采集产业链梳理

——具身智能系列（1）：数据采集

投资要点

- **数据采集：具身智能时代的“卖铲人”。**机器人数据采集是通过其搭载的各类传感器、摄像设备、雷达装置，结合人工操控、仿真模拟等方式，全方位捕捉真实场景中的多维度原始数据，为后续智能训练储备基础素材，搭建起机器人认知世界的“原始素材库”，是机器人从“机械执行”走向“自主智能”的关键基石。当前机器人尚未普及的核心痛点在于机器人大脑智能化泛化能力不足，其深层原因则是机器人在进行训练时缺乏足够体量和高质量的数据，2026 年初全球合规可用的真实物理交互数据距离训练单个通用具身大模型尚有 20 倍缺口。当前行业正受益于标准体系加速构建、数据要素市场化、数据规模工业化、资本快速涌入四大利好因素驱动，带动具身智能数据采集快速发展。

- **数据采集技术路线主要为遥操作、EGO、UMI、仿真数据。**具身智能数据需要满足全链路闭环、多场景+高泛化、高质量+高精度、高兼容+灵活适配、 规模化与效率平衡五大需求，目前主要技术路径为遥操作、EGO、UMI、仿真数据。其中遥操作质量最好但成本较高，主要用于高质量模仿学习数据；UMI 兼具成本与质量，在真机难以大规模铺开时快速采集高质量演示；EGO 是当前主流的应用，成本较低，用于基座模型预训练；仿真数据理论上拥有最高的产能和最低的成本，在大规模预训练中潜力较大。随着数采商业模式的清晰，未来多条技术路径预计将不同场景中均有应用，外采与自采分工明确，共同受益于通用大模型的突破带来的需求增长。

- **数据采集产业链主要分为上游硬件厂、中游软件厂以及下游具身智能本体/大模型厂。**上游硬件设备中性预期下预计到 30 年市场规模达 153 亿元，其多为具身智能本体复用，产业链已较为成熟，壁垒和集中度相对更高的同时作为上游卖铲人，预计头部公司的业绩的确定性更大，建议优先关注具备刚需属性的相机和 IMU 环节，以及部分竞争格局较好&壁垒较高的环节；中游数据采集中性预期下市场规模预计到 30 年达 1741 亿元，26-30 年 CAGR203%，行业正处于从 0 到 1 的快速发展期，远期成长空间明确，弹性更大，建议关注其中的数据集供应商、具备软硬一体的数据采集商、数据后处理厂商以及仿真平台厂商。

- **投资建议**

  上游硬件关注：刚需&单机价值量较大，且竞争格局较好&壁垒较高的奥比中光、华依科技、福莱新材、日盈电子、岱美股份、兆威机电、东华测试等；中游关注：数据采集规模快速增长，已有明确订单落地的利亚德、凌云光、海天瑞声、天奇股份等；以及布局相关物理 AI，有望实现仿真数据从零到一的五一视界、索辰科技、凡拓数创等；下游本体厂关注：宇树科技、上纬新材等。

- **风险提示**

  具身智能落地不及预期；技术路线迭代风险；数据标准不统一风险；数据质量较差风险；估值风险较高风险；价格竞争风险；回款风险；政策不及预期风险。

行业评级：看好(维持)

分析师：文姬
执业证书号：S1230526080006
wenji@stocke.com.cn

相关报告

1 《政策扰动不改出海长期趋势，全球化布局加速进行》 2026.10.21

2 《特斯拉发布新一代 Roadster，产业链底部迎新催化》 2026.10.19

3 《政策端靴子落地，关注摩托出海确定性方向》 2026.10.14

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

正文目录

1 数据采集：具身智能时代的“卖铲人” 4
1.1 数据采集已成为当前制约具身智能发展的关键因素 4
1.2 四大因素驱动具身智能数据采集快速发展 5
2 数据采集技术路线对比 7
3 数据采集产业链及相关公司 11
3.1 中游软件商：新兴需求，弹性最大 11
3.2 上游硬件商：具身智能本体复用，产业链已较为成熟 12
4 投资建议 14
5 风险提示 15

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

图表目录

图 1： 具身智能数据处理流程示意图 4
图 2： 仿真与真实环境下机器人操控成功率对比 5
图 3： 具身智能存量缺口（万小时） 5
图 4： 觅蜂科技数据市场 5
图 5： 京东云具身智能数据交易平台 5
图 6： 数据采集训练场进化路线图 6
图 7： 具身智能数采五大核心需求 7
图 8： 遥操作数采案例 8
图 9： UMI 数采案例 9
图 10： EGO 数采案例 9
图 11： 索辰科技仿真数据物理 AI 平台 10
图 12： 具身智能数据采集产业链图谱 11

表 1： 26H1 部分具身智能数据赛道融资情况 6
表 2： 具身智能数据采集技术路径对比 7
表 3： 具身智能数据采集市场规模 11
表 4： 具身智能数采硬件市场规模 13
表 5： 关注标的概览 14

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

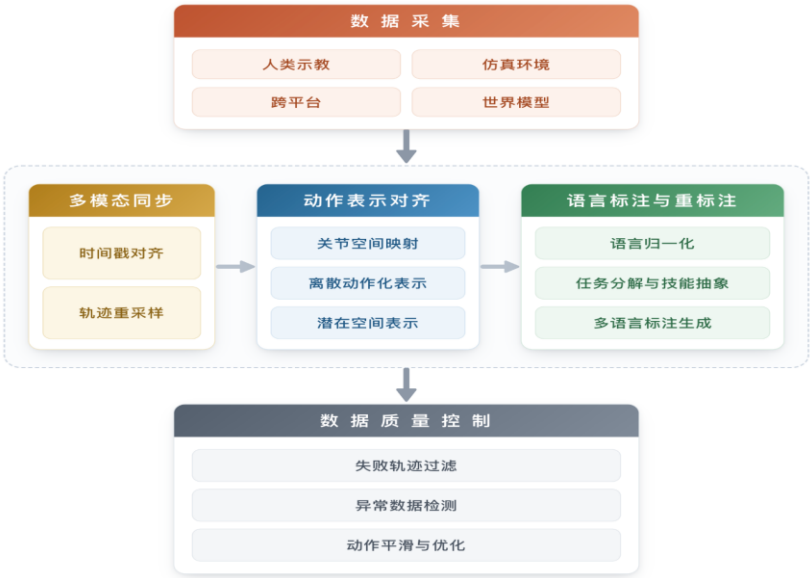
1 数据采集：具身智能时代的“卖铲人”

1.1 数据采集已成为当前制约具身智能发展的关键因素

数据采集是机器人从“机械执行”走向“自主智能”的关键基石。由于机器人无法像人类一样依靠五官自主认知世界，其所有对外界环境、作业场景、动作逻辑的感知，都依赖于全方位的数据采集。所谓机器人数据采集，就是通过机器人搭载的各类传感器、摄像设备、雷达装置，结合人工操控、仿真模拟等方式，全方位捕捉真实场景中的多维度原始数据，为后续智能训练储备基础素材，搭建起机器人认知世界的“原始素材库”。

与自然语言处理和计算机视觉领域可以直接利用互联网海量数据不同，具身智能数据通常涉及真实机器人与环境之间的物理交互，数据采集成本高、获取过程复杂。如何高效获取、处理并组织这些数据，已成为制约具身智能发展的关键问题。对于机器人而言，未经处理的原始数据只是杂乱的信号碎片，唯有通过专业采集、精细标注、规范梳理，才能转化为可被模型学习、可支撑智能决策的高质量数据，让机器人真正看懂环境、读懂指令、精准作业，是机器人从“机械执行”走向“自主智能”的关键基石。

图1：具身智能数据处理流程示意图



资料来源：《具身智能数据采集与处理综述》，浙商证券研究所

在具身智能时代，数据采集的重要性愈发凸显。在具身智能过去的发展历程中，行业将大量资本和人才投入到硬件本体与模型算法。然而，随着 VLA 模型架构的收敛和硬件供应链的成熟，制约机器人规模化落地的核心瓶颈，已从硬件和底层运控转向了数据。一方面，从训练时长来看，截至 2026 年初，全球合规可用的真实物理交互数据总量仅约 50 万小时，不足大语言模型训练数据的两万分之一。而训练一个通用具身大模型，至少需要 1000 万小时，仅距离训练单个大模型的数据缺口就高达 20 倍。另一方面，根据斯坦福 HAI 发布的《AI Index Report 2026》报告显示，机器人操控在仿真环境中的成功率高达 89.4%，但在真实家庭场景中骤降至 12%，77 个百分点的"仿真到现实迁移鸿沟"（Sim-to-Real Gap），让数十亿资金堆出的"机器人大脑"在真实环境中无法做出相匹配的行为。

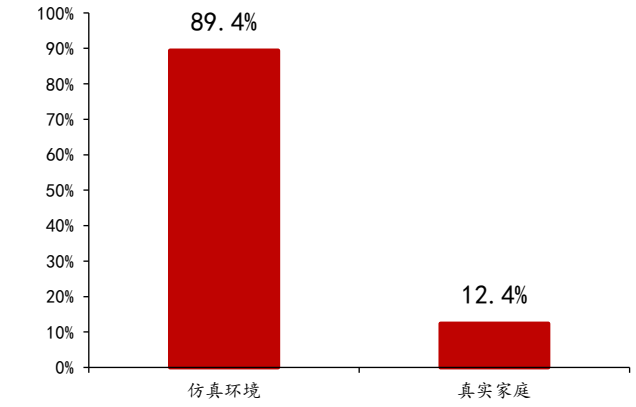
机器人从展示的功能向专业落地场景渗透、真正产生经济价值的阶段，**核心痛点为机器人大脑智能化泛化能力不足，深层原因是缺乏足够量级的高质量数据用于机器人模型训练**，甚至数据总量都未达可观规模，并且 VLA 模型的训练数据通常需要通过机器人与环境的真实物理交互获得，这一过程不仅涉及多模态传感器的同步、复杂环境的构建，还需精

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

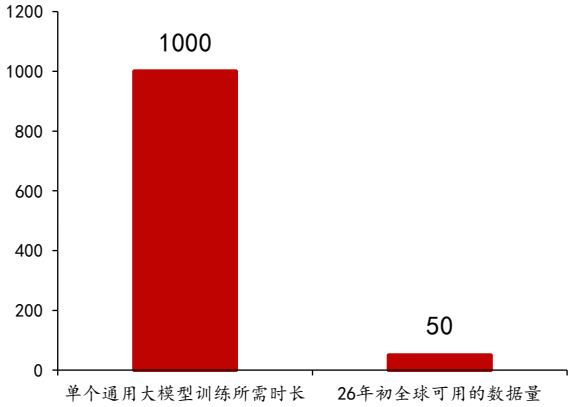
确记录动作细节，导致数据采集成本高昂、流程繁琐，从而严重制约了高质量大规模训练数据的获取。随着人形机器人的发展，数据采集的重要性愈发凸显。

图2：仿真与真实环境下机器人操控成功率对比



资料来源：《AI Index Report 2026》，浙商证券研究所

图3：具身智能存量缺口（万小时）



资料来源：虎嗅，浙商证券研究所

1.2 四大因素驱动具身智能数据采集快速发展

2026 年至今，中国具身智能数采行业正在多个维度发生结构性变化，具体体现在：

1）**标准体系加速构建：**国家标准计划《高质量数据集具身智能面向训练基地的数据采集与模型训练规范》已正式发布，填补了我国具身智能训练基地数据治理与模型训练的标准空白，为产业高质量发展筑牢数据底座。此外，今年 11 月 1 日《人工智能关键基础技术具身智能数据集质量要求及评价方法》即将正式实施，规定了质量要求框架及评价方法，包括数据生产操作规范、组织机制保障和质量评价指标三方面，质量评价指标涵盖完整性、一致性、多样性、真实性、易用性、实用性、可扩展性和安全性八个维度，带动具身智能数据集建设从「规模导向」向「质量导向」转变。

2）**数据要素逐步市场化：**随着国家层面对数据要素市场化的持续改革，今年以来百度智能云、京东先后上线具身智能数据交易平台，觅蜂科技首发 467 个数据集，数据从成本项变为独立商品，降低企业研发门槛，加速推动具身智能相关产业规模化普及，商业模式逐步清晰。

图4：觅蜂科技数据市场



资料来源：觅蜂科技官网，浙商证券研究所

图5：京东云具身智能数据交易平台



资料来源：京东云具身智能数据交易平台官网，浙商证券研究所

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

3）**数据规模工业化：**我国已建成启用 70 多家具身智能训练场，在建、规划 40 余家，其中综合类训练场占比 60% 以上，行业垂类约 18%，大量实训数据加速归集的同时，已经投入真机训练应用。具身智能训练场在发展初期侧重数据采集，但随着真实场景开放，实景实训功能凸显，部分训练场还会拓展中试功能，验证机器人在真实场景下的实际工作能力，数采的生产方式已逐步迈入工业化生产，有望带动未来的采集规模快速增长。

图6：数据采集训练场进化路线图



资料来源：中国信通院，浙商证券研究所

4）**资本快速涌入：**据 IT 桔子统计，2026 年上半年中国具身智能数据创业公司合计拿到超过 170 亿元的融资，占总融资的比重约 40%，其中多家公司如光轮智能累计融资金额已超 10 亿，估值已超 20 亿美元。资本的快速涌入反映了行业增长的确定性。

表1：26H1 部分具身智能数据赛道融资情况

公司	技术路线	成立时间	融资次数	融资估算金额	最新融资轮次
极佳视界	视频蒸馏/世界模型	2023-01	3	35 亿元	B+轮
它石智航	便携采集+本体	2024-07	1	近 32 亿元	Pre-A 轮
光轮智能	仿真合成+人类视频数据+评测	2023-01	3	23 亿元	战略投资
灵初智能	便携采集/触觉手套	2024-09	2	20 亿元	Pre-A 轮
跨维智能	仿真合成/3D VLA	2021-06	1	10 亿元	B 轮
帕西尼感知	触觉传感+采集工厂	2021-06	1	10 亿元	B 轮
深度机智	第一人称视频	2025-05	3	6 亿元	A 轮
觅蜂科技	无本体+仿真	2026-02	2	6 亿元	天使轮
穹彻智能	便携采集/口袋机	2023-11	2	6 亿元	A+轮
渊澈太初	无本体/便携采集	2025-08	2	数亿元	Pre-A 轮
LiberAI	UMI 数据+世界模型	2025-12	3	数亿元	Pre-A 轮
简智机器人	无本体采集+数据基础设施	2025-05	1	数亿元	Pre-A 轮
鹿明机器人	UMI 采集	2024-09	2	数亿元	A+轮
诺亦腾机器人	动作捕捉+全模态数据	2025-03	1	数亿元	Pre-A 轮
灵御智能	真机遥操作	2025-02	2	超亿元	天使轮
戴盟机器人	触觉感知+真机遥操作数采	2021-12	1	1 亿元	A 轮

资料来源：IT 桔子，浙商证券研究所整理

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

2 数据采集技术路线对比

五大数采需求催生不同技术路径诞生。具身智能数据需要满足全链路闭环、多场景+高泛化、高质量+高精度、高兼容+灵活适配、 规模化与效率平衡五大需求，并由此诞生了多个数据技术采集路线。

1）全链路闭环需求：具身数据需覆盖“任务发起-采集-处理-存储-复用”全流程，避免数据在某一环节断层。

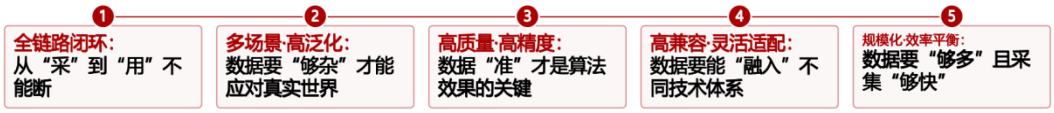
2）多场景+高泛化需求：通用具身智能需适配不同行业、不同任务，这要求数据能覆盖多样化场景与动作组合。

3）高质量+高精度需求：具身数据需兼顾多模态同步性与数据准确性，避免因数据偏差影响模型训练。

4）高兼容+灵活适配需求：不同企业的算法框架、设备类型不同，需数据在格式、适配性上具备灵活性。

5）规模化与效率平衡需求：通用具身智能需要海量数据支撑，但大规模采集易导致效率低下，需在两者间找到平衡。

图7： 具身智能数采五大核心需求



资料来源：机工智能 AI 科研，浙商证券研究所

目前具身智能数据采集的主要技术路径为遥操作、EGO、UMI、仿真数据。具身模型参数的增量极大，训练具备通用能力的模型估计至少需要百万小时级真实交互数据，而现有成熟数据集仅几十万小时量级，供需存在数量级落差 。此外，数据与机体型号深度绑定：自由度、臂长、传感器分布的差异，使面向某一机械臂采集的记录难以迁移到其他构型，数据零散分布、不易汇聚成规模。在数据采集层面高质量、大规模和低成本三者通常无法兼顾，由此诞生出了遥操作、EGO、UMI、仿真数据多个不同的技术路径，在数据采集的不同阶段分别有各自的应用场景。

表2： 具身智能数据采集技术路径对比

	遥操作	EGO	UMI	仿真数据
数据质量	高	中	中	低
规模化潜力	低	高	中	高
成本	高	低	中	低
跨机型复用	低	高	中	高
优点	动作质量高、人机同构	规模大且成本极低，适合做通用先验	无需依赖真实机器人，可带到任意场景，可复制性强	规模大、成本低
缺点	速度慢、成本高，难以大规模扩展	人机差异，视频质量与标注参差	采集的数据非机器人控制器，缺少触觉/力反馈	仿真与真实在动力学、视觉、接触上的差距
适用场景	高质量模仿学习数据	基座模型预训练	在真机难以大规模铺开时快速采集高质量演示	大规模预训练、强化学习训练

资料来源：亿欧智库，xiaolong 的技术笔记，浙商证券研究所整理

遥操作：质量最好但成本较高。遥操作是最早流行的方案，在 2024 年以前的多数垂类模型均通过该方式训练，其原理即由人实时操控机器人完成动作，并全程同步记录关节轨迹、

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

力反馈、触觉信息和视觉画面，以及完整保留人类在复杂环境中的决策过程——包括如何判断、取舍并不断修正动作。设备包含主从式机械臂、VR 手柄、3D 鼠标、示教器、穿戴式/动作捕捉、iPad/AR 眼镜触控等。

遥操作是当前具身智能数据采集中**数据质量最高的方案**，其记录的是机器人本体的动作指令，无需动作重定向即可用于训练，可覆盖双手协同、动态与灵巧操作；可同步采集语言指令。

但与此同时遥操作也是**成本最高、效率天花板最低的路线**，在考虑具身智能本体的情况下其设备的采购价格在 20 万元以上；由于需要驻场作业，人力成本较高，单人单天的采集成本约 300 元；受人类操作速度与疲劳限制，数据规模难以大规模扩展。

目前遥操作在数据采集中的占比约 5%，其适配场景多为高精度精密装配、复杂长程任务，是模型落地 “最后 10%” 高质量资料来源。

图8：遥操作数采案例



资料来源：搜狐网，浙商证券研究所

UMI：兼具成本与质量。UMI（Universal Manipulation Interface，通用操作接口）于 2024 年底开始兴起，其核心思路是通过一个手持式夹爪装置，让人类直接操作夹爪采集数据。与遥操作不同，UMI 不追求实时高精度远程操控，而是专注于高效生成可泛化的机器学习数据，并配合相对轨迹动作表示，把策略部署到不同的机器人本体上。

与遥操作相比，UMI 的**成本优势明显**，一是其无需依赖真实机器人即可采集数据，免去了较多设备采购成本，二是其不需要驻场作业，支持室外场景大规模采集，且操作门槛低，人工成本相对较低。此外，UMI **可实现跨机型数据复用**，一套采集数据可适配多种机器人形态。

UMI 的缺点主要在于数据可使用率仅 30%-40%，存在无效/脏数据；跨本体迁移能力弱，通常仅能锚定单一款机械臂/末端执行器；并且通常缺少触觉/力反馈。

UMI 在数据采集中的占比约 5%，通常用于中期训练，在真机难以大规模铺开时快速采集高质量演示。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

图9：UMI 数采案例



资料来源：智在无界公众号，浙商证券研究所

EGO：当前成本最低的主流应用。EGO（Ego-Centric Multimodal Data Collection，第一人称视角多模态数据采集）兴起于英伟达于 26 年 2 月发表的 EgoScale 论文，并在这之后迅速成为行业主流，目前已占据近 90%的数据量。其核心思路是采集者佩戴头盔或颈挂式设备，在日常工作环境中自然完成任务，模型方通过算法还原将数据映射到机器人末端执行器用于训练，因此不依赖特定的机器人本体。

EGO 的优势在于成本的进一步降低，与 UMI 相同其支持室外作业，且进一步免去了机器人本体的成本；并且可实现跨场景、跨任务、跨对象训练，多样性高。

EGO 的缺点在于其精度低于遥操作和 UMI，且人类与机器人存在人机差异，若视频无配对机器人动作标签，需动作重定向/统一动作表示对齐后才能用于训练。

EGO 当前主要用于早期预训练，已成为基座模型预训练阶段的"主食"。

图10：EGO 数采案例



资料来源：新浪新闻，浙商证券研究所

仿真数据：产能上限最高。与其他数据采集通过真实数据不同，仿真数据通过在各类仿真软件中批量生成任务场景与动作轨迹，从而实现在虚拟世界中的训练，并泛化到真实世界的具身智能中。

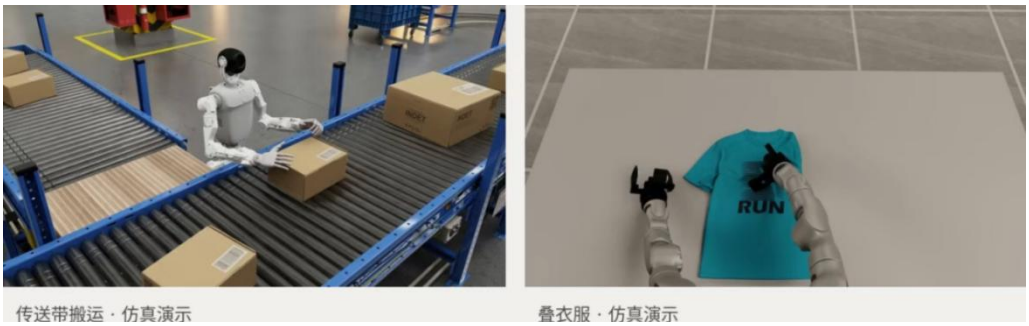
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

仿真数据的最大优势在于是产能上限最高、成本最低的方案——理论上只要有足够的算力，就能无限产出标准化训练数据，且边际成本趋近于零，仅需承担算力开销。天然适合强化学习大规模训练与"先把模型喂饱"的预训练。

仿真数据当前存在的最大问题即仿真到现实的鸿沟（Sim-to-Real Gap），由于物理引擎难以高保真还原形变、物理摩擦、传感器噪声与电机延迟，仿真中的策略迁移到真机往往表现明显退化。如何在未来实现 Sim-to-Real 的数据质量提升仍是当前阻止仿真数据大批量应用的核心。

图11：索辰科技仿真数据物理 AI 平台



资料来源：索辰科技公众号，浙商证券研究所

具身智能数据采集发展趋势展望：

商业模式逐渐清晰：具身智能数据与数据行业的商业化突围，依赖于数据服务商、整机/应用企业、政府端三侧的协同共进。其中**数据服务商**将逐渐升级为基础设施和合作伙伴，实现数据集标准化复售/订阅、数据清洗与定制化采集、模型调优与评测服务等多元模式的盈利模式；**整机/大模型企业**逐渐成为数据闭环的发起者与最终价值的实现者，除整机销售外，实现服务订阅、数据资产变现、生态分成的盈利模式；**政府端**成为生态构建师，为产业健康、安全、快速发展提供土壤，最终获得产业主导权、税收、就业和高科技产业集群的综合回报。

多个技术路径有望共存，外采与自采分工明确：当前具身智能在进行一组训练时整体用量比例约为 Ego:UMI:遥操=10000:1000:100 小时，而由于当前具身智能训练动辄数万小时，对于下游尤其是轻资产的模型厂商而言缺乏足够的人力与物力进行重资产大规模采集，并随着数据集的规模持续扩张，我们预计 95%左右数据将通过外部采购，仅 5%的特种任务、高精尖、高价值垂类数据会通过遥操、UMI 等方式进行选择自采。

大模型的突破将带动行业需求爆发增长：据宇树科技王兴兴在 2026 世界物理 AI 大会暨宇树合作伙伴生态大会上表示，随着 GPT-6 的发布，目前物理 AI 已经看到被击穿的苗头，很可能最近 1-2 年或 2-3 年，目前具身智能迈过 ChatGPT 时刻的最核心挑战是 AI 模型的输入输出和真实机器人的对齐。当前的具身智能的训练依旧以单一场景的垂类模型为主，缺少不同场景间的泛化能力。若未来通用大模型实现跨越，我们认为一方面将带动行业数采需求呈爆发式增长，另一方面物理 AI 的突破也有望带动仿真平台实现大规模应用，进一步降低数据采集的成本。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

3 数据采集产业链及相关公司

数据采集产业链图谱：数据采集产业链的上游主要为硬件设备，涵盖相机、灵巧手、动捕设备、惯性 IMU、力觉/触觉传感器、电子皮肤、力矩传感器、算力芯片等核心产品，中游则为相配套的软件，包含数据集、仿真平台、具身智能大模型等，二者共同实现数据采集后交给进行数据后处理与标注，最终交付给下游具身智能本体/模型厂商。

图12： 具身智能数据采集产业链图谱



资料来源：亿欧智库，浙商证券研究所

3.1 中游软件商：新兴需求，弹性最大

由于针对具身智能的数采需求从近年来才开始得到行业重视，因此行业呈快速增长态势，如海天瑞声已经已与国内某头部具身厂商签署真机数据采集合约，并与多家头部具身厂商、科技大厂及地方政府启动订单对接；凌云光表示 25 年开始关于人形机器人的动捕需求显著加快；乐聚的夸父系列机器人 25 年收入中数据采集同比增长近 77 倍。作为从 0 到 1 的行业，其需求有望快速放量。根据我们的测算，中性/乐观/悲观情形下，**预计到 2030 年具身智能数据采集的规模将达到 1741/2296/1076 亿元，中性情况下 26-30 年 CAGR 达 203%**。此外，由于大部分的数据为预训练通用环境，**其具备复用的特性有望带动相关公司保持较好的盈利能力**。

表3： 具身智能数据采集市场规模

	2026E	2027E	2028E	2029E	2030E
总需求（万小时）	3000	50000	200000	500000	1000000
EGO 占比	85%	85%	83%	78%	68%
UMI 占比	10%	9%	9%	9%	9%
遥操作占比	4%	3%	3%	3%	3%
仿真合成占比	1%	3%	5%	10%	20%
EGO 需求（万小时）	2550	42500	166000	390000	680000
UMI 需求（万小时）	300	4500	18000	45000	90000

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

遥操作需求（万小时）	120	1500	6000	15000	30000
仿真合成需求（万小时）	30	1500	10000	50000	200000
EGO 单价（元/小时）	50	30	21	17	15
UMI 单价（元/小时）	100	60	42	34	30
遥操作单价（元/小时）	400	240	168	134	121
仿真合成单价（元/小时）	30	18	11	6	4
EGO 规模（亿元）	13	128	349	655	1028
UMI 规模（亿元）	3	27	76	151	272
遥操作规模（亿元）	5	36	101	202	363
仿真合成规模（亿元）	0	3	11	32	78
中性数据采集合计规模（亿元）	**21**	**193**	**536**	**1040**	**1741**
乐观数据采集合计规模（亿元）	**34**	**676**	**1254**	**1976**	**2296**
悲观数据采集合计规模（亿元）	**21**	**116**	**346**	**717**	**1076**

资料来源：亿欧智库，公开资料整理，浙商证券研究所整理

软件端我们建议从以下几个投资维度入手：

1、数据集供应商：目前行业正处于快速发展的早期，行业玩家多为初创公司。由于采集数据集本身并非壁垒较高的板块，我们认为行业更倾向于选择数据质量高且可用数据更多的厂商，因此过往在其他数据集领域具备数据采集经验和规模，以及由在具身智能领域深耕多年的企业所孵化的数采厂商更有望在未来的发展中脱颖而出。

2、具备软硬一体的数据采集商：具身智能的真机数据采集高度依赖采集硬件与数据的协同，纯软件路线难以保证采集数据的一致性与可追溯性。我们认为具备自研采集硬件、能打通"硬件销售—数据采集—模型反哺"数据飞轮的厂商，其数据规模与质量壁垒更难被纯软件或纯外包模式追赶。

3、数据后处理厂商：对于 VLA 模型而言，原始交互数据通常来源复杂，既可能包含真实机器人示教数据，也可能来自仿真环境或跨平台大规模整合数据集。这些数据在动作表示、时序频率、语言标注粒度以及数据组织结构等方面存在显著差异。若缺乏统一处理流程，模型将难以实现跨平台泛化与规模化训练。随着模型规模不断扩大，数据处理与标准化质量逐渐成为影响性能上限的重要因素，具备较高的技术壁垒。

4、仿真平台厂商：真机采集成本高、长尾与危险场景难覆盖，仿真合成数据以低成本、可批量、可安全生成的方式补足训练数据缺口，随着 sim-to-real 技术成熟，仿真数据在训练中的占比有望持续提升。我们认为壁垒不在单一渲染能力，而在高保真物理引擎、海量场景资产库与"仿真—真机"数据闭环的叠加，因此具备自研物理仿真引擎、丰富场景库并能与真机数据形成迭代闭环的平台厂商更有望走出来。

3.2 上游硬件商：具身智能本体复用，产业链已较为成熟

硬件厂商经历了具身智能数年的发展，大部分产业链已较为成熟，壁垒和集中度相对更高，如奥比中光 25 年在全球机器人 3D 视觉感知市场的份额为 29.0%位居第一，其他力矩/触觉传感器等也已形成稳定的供货体系，头部公司的业绩的确定性更大。

根据我们的测算，中性/乐观/悲观情形下，**预计到 2030 年具身智能数采上游硬件设备的规模将达到 153/235/121 亿元，中性情况下 26-30 年 CAGR 达 230%**。虽然考虑到上有

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

硬件在具身智能本体上同样已经有广泛的应用基数，其综合的增速预计逊于软件厂商，但作为上游硬件的卖铲人**有望更早的兑现业绩**。

表4：具身智能数采硬件市场规模

	2026E	2027E	2028E	2029E	2030E
EGO 需求（万小时）	2550	42500	166000	390000	680000
UMI 需求（万小时）	300	4500	18000	45000	90000
遥操作需求（万小时）	120	1500	6000	15000	30000
对应 EGO 设备需求（万台）	3.2	49.6	172.9	357.5	566.7
对应 UMI 设备需求（万台）	0.4	5.3	18.8	41.3	75.0
对应遥操作设备需求（万台）	0.2	1.8	6.3	13.8	25.0
EGO 设备单价（元/台）	3000	2700	2430	2187	1968
UMI 设备单价（元/台）	5000	4500	4050	3645	3281
遥操作设备单价（元/台）	10000	9000	8100	7290	6561
EGO 设备规模（亿元）	0.96	13.39	42.02	78.19	111.54
UMI 设备规模（亿元）	0.19	2.36	7.59	15.04	24.60
遥操作设备规模（亿元）	0.15	1.58	5.06	10.02	16.40
中性数据采集设备规模（亿元）	**1**	**17**	**55**	**103**	**153**
乐观数据采集设备规模（亿元）	**2**	**54**	**122**	**198**	**235**
悲观数据采集设备规模（亿元）	**1**	**9**	**34**	**63**	**121**

资料来源：亿欧智库，公开资料整理，浙商证券研究所整理

硬件端我们建议着重关注以下几点：

1、**刚需**：从 EGO 到 UMI 再到遥操作，不同的技术路径对应有不同的增量设备，而 EGO 的核心设备即相机和 IMU。其中相机方面 EGO 通常搭载多目相机阵列，配备至少 2-4 颗全局快门 RGB 相机，覆盖超过 180 度的超广角视场，从而精确定位物体在三维空间中的位置。IMU 则为数据采集起到空间定位作用，让设备还知道自己在空间中的位置和姿态。二者作为数采硬件中不可或缺的部件，其需求量预计更大。

2、**竞争格局较好&壁垒较高**：部分数采硬件设备兼具较高技术壁垒与较强客户粘性，竞争格局明显优于通用标准件。以触觉传感器为例，其同时横跨柔性材料、阵列工艺与触觉信号处理算法三重壁垒，且需与采集手套/本体深度耦合，国内厂商已凭借先发量产经验建立优势。我们认为在数采需求从 0 到 1 放量的窗口期，被本体厂/数采商深度绑定、且具备"算法—硬件"一体化能力的环节更难被新进入者颠覆，如触觉传感器、力矩传感器等。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

4 投资建议

我们建议优先关注中游尚处于起步状态的数据集采集商、仿真平台等厂商，远期成长空间明确，弹性更大；上游硬件厂商相对而言集中度更高，业绩的确定性更大，且作为上游的卖铲人预计将更早的兑现业绩，同样建议积极关注。

上游硬件关注：刚需&单机价值量较大，且竞争格局较好&壁垒较高的**奥比中光、华依科技、福莱新材、日盈电子、岱美股份、兆威机电、东华测试等**；中**游关注：**数据采集规模快速增长，已有明确订单落地的**利亚德、凌云光、海天瑞声、天奇股份**等；以及布局相关物理 AI，有望实现仿真数据从零到一的**五一视界、索辰科技、凡拓数创**等；下游本体厂关注：**宇树科技、上纬新材**等。

表5： 关注标的概览

公司	代码	收盘价（10月28日）	总市值（亿元）	归母净利润（亿元）			PE		
				26E	27E	28E	26E	27E	28E
奥比中光	688322.SH	107.16	442	2.78	5.08	8.01	158.87	87.08	55.20
华依科技	688071.SH	57.00	48	0.45	0.80	/	107.41	60.41	/
福莱新材	605488.SH	48.61	146	1.15	1.70	2.43	126.89	85.79	60.00
日盈电子	603286.SH	54.45	64	/	/	/	/	/	/
岱美股份	603730.SH	11.45	246	10.08	11.13	12.48	24.41	22.11	19.71
兆威机电	003021.SZ	97.41	261	1.64	2.58	2.54	159.17	101.02	102.82
东华测试	300354.SZ	28.98	40	1.33	1.65	2.02	30.14	24.29	19.89
利亚德	300296.SZ	6.22	169	/	/	/	/	/	/
凌云光	688400.SH	45.42	216	5.21	3.86	5.07	41.64	56.19	42.80
海天瑞声	688787.SH	135.57	82	0.35	0.57	0.87	231.47	142.65	94.36
天奇股份	002009.SZ	17.62	71	/	/	/	/	/	/
五一视界	6651.HK	57.95（港元）	208	-1.51	-1.23	1.00	-138.29	-169.57	209.18
索辰科技	688507.SH	160.93	143	0.61	0.89	1.15	234.52	161.12	124.66
凡拓数创	301313.SZ	63.00	64	/	/	/	/	/	/
宇树科技	688836.SH	695.90	2815	6.58	10.27	14.54	427.75	274.11	193.53
上纬新材	688585.SH	175.00	706	-0.13	-0.92	3.57	-5401.23	-769.91	197.87

资料来源：Wind，浙商证券研究所 注：盈利预测出自 10 月 28 日 wind 一致预期

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

5 风险提示

具身智能落地不及预期：人形机器人、具身大模型整体研发迭代进度慢于预期，下游本体厂商资本开支、数据采购预算收缩，对物理交互数据需求下滑，数采业务订单与收入承压。

技术路线迭代风险：真机遥测采集、EGO、UMI、仿真合成数据等路线存在不确定性；若仿真数据技术快速成熟，行业对高价真机采集数据需求被替代，现有数采产能、采集平台资产面临贬值风险。

数据标准不统一风险：当前行业缺乏统一数据格式、标注规范、任务评价体系，不同机器人本体、传感器硬件的数据互通难度高，数据集跨平台复用能力弱，客户粘性、规模化交付受限。

数据质量较差风险：物理世界长尾场景复杂，采集得到的动作数据未必能够有效迁移到机器人本体控制；存在“数据量大但泛化增益有限” 的情况，下游客户采购后模型提升不达预期，复购意愿下降。

估值风险较高风险：当前具身智能数据采集相关公司的估值中已包含较高的远期成长预期，行业尚处发展早期，收入体量小、盈利兑现节奏存在不确定性，若行业进展不及预期或市场风险偏好回落，相关公司存在估值回调风险。

价格竞争风险：数据采集环节进入门槛相对较低，随着越来越多数据服务商、标注厂商及主机厂自建团队入场，行业竞争或逐步加剧。若采集与标注服务供给快速扩张而需求放量偏慢，数据产品及服务单价存在下行压力，行业整体及参与厂商的毛利率可能受到侵蚀。

回款风险：具身智能数据采集业务多以项目制交付为主，下游客户涵盖 AI 实验室、主机厂、机器人厂商等，部分客户结算账期较长。若客户预算执行节奏放缓或信用状况变化，数据服务商存在应收账款回收不及时、坏账计提增加的风险，进而影响现金流与业绩质量。

政策不及预期风险：具身采集多为多模态感知数据，包含人脸、步态、手眼动作、环境空间信息，场景多在工厂、居家等实景环境；知情授权、脱敏、数据权属界定规则尚不完善，存在隐私泄露、违规采集、数据流通合规等风险，若相关政策趋严，部分业务开展范围可能受限。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 买　入 ：相对于沪深 300 指数表现＋20%以上；
2. 增　持 ：相对于沪深 300 指数表现＋10%～＋20%；
3. 中　性 ：相对于沪深 300 指数表现－10%～＋10%之间波动；
4. 减　持 ：相对于沪深 300 指数表现－10%以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：

1. 看　好 ：行业指数相对于沪深 300 指数表现＋10%以上；
2. 中　性 ：行业指数相对于沪深 300 指数表现－10%～＋10%以上；
3. 看　淡 ：行业指数相对于沪深 300 指数表现－10%以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层
北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层
深圳地址：广东省深圳市福田区广电金融中心 33 层
上海总部邮政编码：200127
上海总部电话：（8621）80108518
上海总部传真：（8621）80106010
浙商证券研究所：https://www.stocke.com.cn

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster