

具身智能数据采集：10 家案例分析

性质：原创分析（基于 2026-08 WRC 世界机器人大会及多家媒体公开报道交叉整理） 日期：2026-08-23 标注说明：[已核实] = 多源交叉验证；[推断] = 方向可靠、属分析归纳；[估算] = 逻辑外推，无直接来源

引言：数据墙下的"卖水人"生意

在前文《家用具身智能：从「模型崇拜」到「数据觉醒」》中，我们判断家用具身智能的瓶颈"不是模型，是数据"。这个判断背后有一个硬数据：

贵州省大数据局 2026 年 7 月数据：国内真实物理交互场景合规数据仅 50 万小时，而机器人商业化落地需要数千万小时——缺口超 99% [已核实]。

这个"99% 缺口"，催生了一个确定性的"卖水人"赛道——具身智能数据采集。本文分析 10 家代表性企业，按技术路线分为四类。

核心发现（先行结论）：

- 数据采集已分化成三条路线——真机数据、仿真/合成数据、动捕数据，且尚未收敛；
- 睿尔曼的 GLN 模式是"数据 + 劳务"双轮的最优解——数据不再是烧钱成本，而是"真干活"时免费沉淀的资产；
- "真机 vs 仿真"的双轨制是长期常态——纯真机或纯仿真都可能被淘汰，打通两者、持续产生增量数据者才是赢家。

一、真机数据采集：真干活 / 遥操作 / 数采方案（4 家）

1. 睿尔曼智能 —— "先真干活，再学干活"的行业标杆

- 路线：真机数采 + GLN（Global Labor Network）远程劳动力网络 [已核实]
- 模式：RealBOT 机器人在 24 小时智慧药店补货取药、配电室巡检、与稻香村师傅协作做京式月饼、异地远程操控（北京操作、常州执行） [已核实]
- 商业闭环：用"真劳务收入"反哺数据采集，数据是干活的副产品而非成本 [推断]

- 资本状态：八年深耕，关节模组年产能超 50 万台，已启动创业板 IPO（国泰海通辅导）[已核实]
- 分析 [推断]：睿尔曼是把"数据采集"从"烧钱苦力"升级为"劳务平台 + 数据飞轮"的标杆，是这条赛道最接近跑通商业闭环的一家。

2. 自变量机器人 —— 无本体数采方案

- 路线：WALL-B 世界统一模型 + 无本体数据采集方案 [已核实]
- 进展："机器人住家"计划落地（家庭服务 + 物流分拣双场景），四家大厂押注 [已核实]
- 分析 [推断]：用世界模型降低数据采集对"特定本体"的依赖，让采集数据可跨本体复用——这是对"数据资产化"的早期探索。

3. 章鱼动力（SynapX） —— "脑-手-数据"一体化

- 路线：SYNWorld 世界基础模型 + OctoH-Hand 高自由度仿生灵巧手 + OctoSense 数采方案 [已核实]
- 分析 [推断]：把数据采集与"灵巧手 + 世界模型"绑定，走的是"硬件 + 数据 + 大脑"垂直一体化，赌的是"手"作为数据入口的稀缺性。

4. 灵初智能（PsiBot） —— 人类操作数据的旗帜

- 路线：10 万小时人类操作数据（工业装配、生活操作、物体抓取），其中 1000 小时开源 [已核实]
- 标志：SynData 数据集冲上 HuggingFace Trending 全球第一，下载量约 1.46 万次 [已核实]
- 观点：CEO 王启斌"具身智能的胜负手，不在整机，在人类数据"；联创陈源培"世界模型不是核心方向，是数据迁移的工具" [已核实]
- 分析 [推断]：灵初是"人类数据路线"最纯粹的旗手，用开源构建生态、用数据规模构筑壁垒，与"世界模型派"形成鲜明对照。

二、仿真 / 合成数据 + 世界模型（3 家）

5. 光轮智能 —— "物理 AI 数据与评测基础设施"

- 路线：人类行为数据（EgoSuite-Open100K）+ 物理仿真 + 规模化评测（RoboFinals）+ 部署反馈（RoboStack）全闭环 [已核实]
- 标志：发布全球首个十万小时级全模态人类行为开源数据集（7 大类环境、128 类场景）[已核实]
- 生态：已与黎曼动力、源策未来、德塔智能等多家合作 [已核实]
- 分析 [推断]：光轮做的是"数据 + 评测 + 反馈"的基础设施，用"评测标尺"补上具身智能缺失的"MMLU 时刻"——这是比单纯卖数据更高级的定位。

6. 五一视界 (HK06651) —— "Physical AI 第一股"的降维打击

- 路线：智驾仿真市占率超 50%，跨界做具身智能仿真 [已核实]
- 资本状态：市值 256.92 亿港元，"国内 Physical AI 第一股" [已核实]
- 观点：CEO 李熠"当前市场对具身智能、对物理 AI 的期望过高" [已核实]
- 分析 [推断]：从智驾仿真降维切入具身仿真，有成熟的物理仿真技术积累，是"仿真路线"里资本最厚、技术最成熟的一家。

7. 黎曼动力 —— 世界模型 + "仿真真机双 SOTA"

- 路线：Riemann-1.0 具身世界动作模型 + Matrix-Game 3.5 交互式世界模型，仿真、真机双 SOTA [已核实]
- 动作：联合光轮智能、诺亦腾，剑指 2026 年底 100 万小时具身数据建设 [已核实]
- 分析 [推断]：黎曼是"世界模型 + 仿真合成"路线的代表，与光轮（评测）、诺亦腾（动捕）形成"大脑 + 基建 + 数据"的互补联盟，典型的技术联盟打法规避单点瓶颈。

三、动捕 / 人体运动数据 (1 家)

8. 诺亦腾机器人 (Noitom Robotics) —— 给机器人"言传身教"

- 路线：HiPhi 数据集，开源 610+ 小时高精度人体运动数据 [已核实]
- 独特性：同时具备大规模、高精度、丰富动作覆盖，规模超此前最大开源数据集（德国马普所） [已核实]
- 分析 [推断]：动捕数据是"把人类动作教给机器人"的关键，属于稀缺的"高质量示范数据"。诺亦腾用动捕积累卡住了"人体运动数据"这个细分入口。

四、数据闭环 / 真实数据一体化 (2 家)

这两家不是"纯数据公司"，但它们的核心竞争力高度依赖数据闭环——是"数据驱动的具身大脑"样本。

9. 千寻智能 (Spirit AI) —— 数据闭环 + 长程任务

- 路线：Spirit v1.6 具身模型 + Moz 系列机器人 + "客厅整理"长程任务 [已核实]
- 资本状态：30 天融资 30 亿（顺为资本、云锋基金联合领投——雷军、马云罕见"同框"） [已核实]

- 观点：联创高阳"数据闭环的跑通效率，直接决定规模化进程"；郑灵茵"2026—2027 是从讲故事到交付的分水岭" [已核实]
- 分析 [推断]：千寻把"长程任务"当作数据闭环的磨刀石——任务越长，沉淀的数据越有价值。它证明了"数据"正从"训练素材"变成"核心竞争壁垒"。

10. 星海图 —— 真实数据的坚持者

- 路线：坚持用真实数据，反对"仿真替代论" [已核实]
- 资本状态：北京亦庄首只百亿独角兽、10 亿 B 轮、累计近 11 轮融资近 30 亿 [已核实]
- 观点：创始人高继扬"具身智能无瞬间拐点，坚持用真实数据" [已核实]
- 分析 [推断]：星海图是"真实数据派"的旗帜，与"仿真派"（五一视界/光轮/黎曼）形成路线对峙。这种对峙本身说明：真机数据 vs 仿真数据的价值之争，远未定论。

结论：三个关键判断

1. 数据采集已分化成三条路线，且各有拥趸。真机派（睿尔曼/灵初/星海图）赌"真实不可替代"，仿真派（五一视界/光轮/黎曼）赌"合成可规模化"，动捕派（诺亦腾）卡"人体运动"细分。三线并行、尚未收敛 [推断]。
2. 睿尔曼的 GLN 是"数据 + 劳务"双轮的最优解。它把数据采集从"成本中心"变成"利润来源"（真干活赚钱 + 数据沉淀），这是所有纯数据公司最该学的模式 [推断]。
3. "真机 vs 仿真"双轨制是长期常态，打通者赢。黎曼动力联盟（大脑 + 仿真 + 动捕）、光轮智能（数据 + 评测 + 反馈）都在尝试打通多条路线。终局赢家大概率是"真机数据 + 仿真数据 + 数据资产运营"三者打通、并能持续产生增量数据的"数据基础设施公司" [推断]。

附：证据分级与主要信源

- [已核实]：睿尔曼 GLN 与 IPO（环球网/雷科技/和讯网）、灵初 10 万小时数据与 SynData（钛媒体/机器之心/星河频率）、光轮 EgoSuite-Open100K（北京商报/环球网）、五一视界（每经）、黎曼动力联盟（猎云网/界面/北京商报）、诺亦腾 HiPhi（阿尔法公社）、千寻融资（每经/贝果财经）、星海图（证券时报/21 世纪经济报道）、数据缺口（澎湃新闻，转引贵州省大数据局）。
- [推断]：三条路线划分、GLN 最优解判断、双轨制长期并存——方向可靠，属分析归纳。
- [估算]：本文无外推数据。

本文为原创分析，全文证据分级标注。