

具身智能行业调研

目录

| | |
|-------------------------|----|
| 1. 行业综述 | 3 |
| 1.1 行业演进 | 3 |
| 1.2 具身智能的核心架构 | 3 |
| 1.2.1 感知层 | 3 |
| 1.2.2 决策层 | 3 |
| 1.2.3 执行层 | 3 |
| 1.3 技术实现 | 4 |
| 1.3.1 多模态感知 | 4 |
| 1.3.2 VLA 模型 | 4 |
| 1.3.3 强化学习与仿真训练 | 4 |
| 2. 产业链拆解 | 4 |
| 2.1 产业链拓扑 | 4 |
| 2.2 上游核心零部件 | 5 |
| 2.3 中游本体集成 | 5 |
| 2.4 下游应用与场景落地 | 6 |
| 3. 智能应用前景与商业化价值演进 | 6 |
| 3.1 核心驱动力 | 6 |
| 3.2 商业化落地的主战场 | 6 |
| 3.3 需求侧的商业模式创新 | 7 |
| 3.4 未来应用前景表 | 7 |
| 4. 流程优化与商业规划 | 7 |
| 4.1 生产流程解构与 AI 赋能 | 7 |
| 4.2 交付路径 | 8 |
| 4.3 workflow 效率与成本收益分析 | 8 |
| 4.4 技术落地演进 | 8 |
| 5. 演进与愿景 | 9 |
| 5.1 物理交互与边界探索 | 9 |
| 5.2 具身智能产品的核心价值维度 | 9 |
| 5.3 结语 | 10 |

1. 行业综述

1.1 行业演进

在 2026 年的今天，我们正处于人工智能发展的第三次浪潮——具身智能的爆发期。

回顾 AI 的进化史，我们经历了从“规则驱动”到“数据驱动”的转变。早期的机器人（如传统的工业机械臂）本质上是“非具身”的自动化工具，它们运行在高度受限的结构化环境中，依赖工程师预设的 If-Then 逻辑。而具身智能的出现，标志着 AI 终于获得了“物理载体”。

具身智能的核心定义不仅是一个能够理解语言的聊天机器人，也不仅仅是一个能走路的机械装置。而是一个能够通过传感器感知现实世界，通过大脑进行逻辑推理与任务规划，并最终通过执行器（身体）产生物理反馈，从而改变环境状态的智能闭环系统。

1.2 具身智能的核心架构

具身智能的架构设计模仿了人类的神经生理系统，通常被划分为感知层（感官）、决策层（大脑/小脑）和执行层（身体）。

1.2.1 感知层

感知层是智能体获取外界信息的入口。与传统机器人仅依赖红外或超声波避障不同，具身智能的感知是多模态融合的：

视觉感知：利用 RGB-D 相机和激光雷达进行空间计算，同时需要拥有语义分割能力——机器人不仅要看到前方有物体，还要识别出那是“一个盛有半杯热咖啡的陶瓷杯”，以便后续决策。

触觉感知：2026 年的高端具身产品已标配电子皮肤或六维力传感器。这使得机器人能感知物体的硬度、纹理和滑动趋势，实现如“单手抓取生鸡蛋”或“拧开松动的瓶盖”等精细动作。

听觉与语音交互：集成大模型的音频输入，使智能体可以实现自然语言指令的实时接收，并具备声源定位能力。

1.2.2 决策层

这一层是具身智能最核心的技术壁垒，分为长程规划和实时控制：

大脑（任务规划层）：基于大模型，当用户给出模糊指令，如“我饿了”时，大脑负责将其拆解为一系列子任务：搜索厨房 -> 识别食物 -> 抓取 -> 加热 -> 递送。

小脑（运动控制层）：负责解决物理世界的动力学问题。通过强化学习训练出的运动策略，确保机器人在受到外力推搡或行经崎岖路面时，能够通过关节扭矩的调整保持平衡。

1.2.3 执行层

执行层决定了智能体能力的上限：

高功率密度关节：包括电机、谐波减速器和行星丝杠。通常在定义硬件产品时，需在扭矩密度、响应速度与 BOM 成本之间做权衡。

末端执行器：如灵巧手，目前的行业趋势是从三指夹持器向五指灵巧手进化，以实现更接近人类的操作自由度。

1.3 技术实现

具身智能的实现依赖于一个完整的数据与动作闭环，其核心技术路径如下：

1.3.1 多模态感知

机器人通过摄像头、激光雷达和触觉传感器获取环境数据。2026 年的主流趋势是空间计算的深度集成，使机器人能构建高精度的语义地图。

1.3.2 VLA 模型

这是具身智能的“灵魂”。VLA 模型打破了传统感知与控制的界限：

Input：视觉图像+自然语言指令（如：“帮我把桌上那个蓝色的杯子递给左边的人”）。

Process：模型直接预测动作序列（Action Tokens）。

Output：输出关节扭矩或坐标指令。

1.3.3 强化学习与仿真训练

由于物理世界采集数据成本极高且具危险性，技术实现高度依赖大规模并行仿真。在虚拟世界中训练百万次抓取，再将算法迁移至物理实体，解决“小脑”的灵活性问题。

2. 产业链拆解

2.1 产业链拓扑

2026 年具身智能产业链已从最初的实验室组装阶段，演进为精密的分工协作体系。上游核心零部件不仅决定了智能体的运动性能上限，更直接锁定了产品的 BOM 成本基调；中游本体厂商则从单纯的“集成商”转型为“软硬耦合的系统开发”，其核心竞争力在于如何将复杂的 VLA 模型算法与动力学硬件进行最优匹配。

与手机行业的开发逻辑类似，具身智能的产业链也遵循“模块化、标准化、国产化”的演进路径。目前，上游环节中的高功率密度执行器、六维力传感器及高性能算力芯片占据了整机产值的 70% 以上；中游本体端则分化出人形、四足、轮式等多种形态，以适应工业、商用及特种场景；下游应用端则通过“任务数据采集-模型训练-场景部署”形成数据飞轮。根据 Goldman Sachs (2025-2026 Robotics Report) 的分析，这种纵向一体化的产业链结构使得具身智能产品的迭代周期从过去的以“年”为单位缩短至以“季度”为单位，产业链的成熟度已支撑起从千台级向万台级规模化交付的跃迁。

2.2 上游核心零部件

执行器作为具身智能的“肌肉”，其成本与性能的平衡是当前产业链最大的挑战。在 2026 年的主流方案中，线性关节与旋转关节的选型直接影响了本体的爆发力与续航。其中，行星丝杠由于其高负载能力和长寿命，已成为人形机器人线性关节的标配。根据中信证券《机器人产业链专题报告》数据，单个高精度行星丝杠的成本虽较 2023 年下降了约 30%，但仍占整机执行器成本的 15% 左右，其工艺难点在于高精度的磨削与热处理，目前仍主要由海外巨头如 Rollvis、瑞典 SKF 主导，但国产厂商如恒立液压、贝斯特已在 2025 年后实现了在 20% 精度公差内的国产替代。与此同时，旋转关节中的谐波减速器已进入高度竞争阶段，国内绿的谐波与双环传动通过优化齿形设计和材料配方，已将单体成本压低至 1500-2500 元人民币区间，这为万元级具身智能产品的普及奠定了硬件基础。

在感知层，2026 年产业链的增量重心已从单一的“视觉感知”转向“多模态触觉融合”。与手机屏幕精密触控逻辑异曲同工，具身智能对末端执行器（灵巧手）的要求已进化到需实时反馈压力、滑动趋势及纹理感知的程度。目前的行业标准已广泛采用基于 MEMS 技术的电子皮肤或基于光路形变的视触觉传感器（如 GelSight 方案）。参考 IEEE Robotics & Automation Society (2025) 的技术综述，视触觉传感器通过内部相机捕捉弹性体受力后的微小形变，能够提供比特级精度以上的力反馈数据，这对于执行“抓取易碎品”或“精密装配”等长尾任务至关重要。感知系统的供应链整合，正在借鉴智能手机摄像头模组的 CMOS 集成经验，向着体积小小型化、接口标准化的方向快速演进。

2.3 中游本体集成

中游本体商在 2026 年的核心任务是通过工程化手段降低整机成本。一份典型的人形机器人 BOM 显示，动力系统（含电机、减速器、驱动器）占比约 50%，感知系统占比约 15%-20%，算力平台占比约 15%，其余为电池与结构件。以特斯拉 Optimus Gen-3 及国内智元为代表的先导机型，其整机成本控制已进入 3 万美元（约 20 万人民币）的临界点。下表详细列出了 2026 年主流人形机器人的核心 BOM 构成及参考供应商：

| 模块名称 | 核心器件 | 成本权重 | 2026 年主流供应商 |
|-------|----------------------|------|-----------------------------|
| 执行器 | 无框力矩电机、谐波/行星减速器、行星丝杠 | 52% | 三花智控、拓普集团、绿的谐波、Moog |
| 感知系统 | 深度相机、LiDAR、六维力传感器 | 18% | 华为（MDC 传感器）、奥比中光、柯力传感、帕西尼感知 |
| 控制与算力 | 高性 SoC、端侧计算模块 | 15% | NVIDIA、地平线、华为昇腾 |
| 末端执行器 | 灵巧手 | 10% | 宇树科技、腾讯 |

| | | | |
|-------|-------------------------|----|------------------|
| | | | Robotics X、智元机器人 |
| 能源与骨架 | 高能量密度锂电池、铝镁合金/ 碳纤维骨架 | 5% | 宁德时代、航天科技 |

这种成本结构的优化路径主要依赖于“核心件集成化”和“算法补偿硬件”两条路线。例如，通过更先进的强化学习算法，可以在一定程度上容忍电机输出扭矩的波动，从而降低对超高性能电机的过度依赖。

2.4 下游应用与场景落地

2026年，具身智能的下游应用正经历从“单一功能测试”向“复杂工序替代”的转型。根据 IFR (International Federation of Robotics) 发布的《2026 机器人与自动化趋势报告》，汽车总装线是目前最具商业确定性的落地场景。具身机器人被用于执行发动机线束插接、内饰件精密卡扣等传统刚性机械臂无法胜任的非结构化任务。在这一环节，智能体的核心价值在于“泛化能力”，即通过少样本学习快速适配不同型号的车型。除了 B 端工业，C 端康养场景在 2026 年也出现了突破口，重点在于处理简单的家庭取物与陪护协助。然而，下游场景的拓宽依然受限于“场景数据”的匮乏。目前，领先的本体厂商正通过部署大规模的远程操作系统采集人类专家的操作数据，并利用物理引擎进行百万倍速的 Sim-to-Real 训练，以解决复杂物理交互中的稳定性难题。

3. 智能应用前景与商业化价值演进

3.1 核心驱动力

2026 年具身智能的爆发源于全球制造业“结构性劳动力短缺”的迫切需求。相较于传统工业机械臂的“预设逻辑”，具身智能的核心优势在于其基于 VLA 大模型的泛化执行能力，能够通过少样本学习快速迁移至不同生产线，显著降低了重构成本。然而，在向 C 端家庭场景渗透时，具身智能遭遇了迥异于工业环境的挑战。家庭环境的高度非结构化、成员（儿童/宠物）的不可预测性，以及用户对隐私与安全近乎严苛的阈值，构成了其规模化落地的核心局限。

目前，行业共识在于，家庭场景的突破点不在于“全能型人形机器人”，而在于将 AI 大脑模块化植入现有家电，通过分步替代琐碎家务（如智能式收纳、精密洗涤）来逐步建立用户信任。

3.2 商业化落地的主战场

工业与物流领域凭借高 ROI 和受控环境，依然是 2026 年具身智能商业化的主战场。在汽车总装、危化品搬运及精细物流分拣中，机器人通过空间语义感知与灵巧手微操，成功填补了传统自动化设备在“动态非标工序”上的空白。根据 IDTechEx 的预测，工业及物流机器人预计将贡献 2026-2030 年间约 60%-65% 的具身智能市场增长。与之形成鲜明对比的是，家庭场景的商业化进程缓

慢。即便在硬件成本日益下降的趋势下，通用型人形管家在家庭中仍面临“投入产出比极低”的质疑——用户往往倾向于使用专业化设备（如扫地机器人、洗碗机）解决单一需求，而非购买一台昂贵且维护复杂、在处理日常杂乱环境（如杂乱桌面、衣物堆叠）时可靠性不足的通用机器人。

3.3 需求侧的商业模式创新

随着部署规模扩大，基于 ROI 的 RaaS 模式正成为主流。厂商通过“本体+模型+数据”的联合服务，不仅提供实体，更交付场景下的持续优化能力，这种模式有效分摊了客户的初始资本支出。然而，在家庭应用中，这种服务模式面临着严峻的“安全性悖论”：机器人必须具备完全的自主性以降低用户监督成本，但任何一次感知偏差导致的碰撞或隐私泄露，都可能摧毁用户对该技术的信任。因此，2026 年的具身智能仍处于淘汰赛初期，在进行产品定义时，必须在“模型实时响应速度”与“物理安全约束”之间进行权衡。当前，那些率先在特定狭窄场景（如老年人辅助行走、康养监测）中做到“极致可靠”的产品，正逐渐跑通商业闭环。

3.4 未来应用前景表

| 应用场景 | 典型任务 | 当前痛点 | 具身智能赋能价值 |
|--------|-------------|----------------|-------------------|
| 高端精密制造 | 非标组装、内饰安装 | 换线成本高，精度要求极高 | 少样本学习，快速切换生产任务 |
| 仓储物流 | 非结构化拆码垛 | 环境动态变化，物体语义识别难 | 视觉-语义深度融合，提升分拣准确率 |
| 康养陪护 | 辅助起身、跌倒预警 | 交互生硬，家庭环境复杂 | 多模态自然交互，实时安全辅助 |
| 家庭生活 | 整理杂乱桌面、折叠衣物 | 物理交互可靠性低，隐私风险高 | 基于大模型的任务拆解与小步迭代落地 |

4. 流程优化与商业规划

具身智能在制造业的落地，标志着工业自动化正经历从“刚性程序执行”向“语义智能决策”的本质蜕变。传统工业机器人依赖离线编程与高度结构化的固定工装，这种模式在面对多品种、小批量的制造需求时，显现出极高的换线成本与响应滞后性。将具身智能引入生产 workflow，不仅仅是引入一种新设备，更是通过 AI 大模型重构制造流程的算力与决策逻辑，实现“物理执行能力”与“智能感知决策”的深度协同。

4.1 生产流程解构与 AI 赋能

在现有的制造流程中，物料、形态多变及环境干扰是导致产线中断的主要原因。具身智能的赋能核心在于赋予系统“感知-推理-执行”的闭环能力：

视觉语义化与定位纠偏：传统自动化需要昂贵的视觉定位系统配合高精度夹具，以保证抓取坐标无偏差，而基于 VLA 大模型的智能体，能够将零件识别与空间语义计算同步进行。系统不再依赖于零件被精确放置在固定位置，而是通过对复杂语义环境的理解，实时校准抓取姿态。这一技术路径直接减少了对硬性定位装置的需求，显著降低了产线硬件部署的固定资产投入。

作业流程的自适应：在复杂装配环节，智能体通过多模态触觉传感器实时反馈装配阻力，类比人类神经系统，系统能够根据装配受力状态实时调整动作轨迹。这种具备“触觉反馈”的智能执行，有效解决了传统机械臂因刚性碰撞导致的损耗与组装失效问题。

4.2 交付路径

商业化规模扩大的核心障碍在于非标准化的安装与调试。参考消费电子领域的集成经验，具身智能的交付需要遵循“硬件模组化”原则：

模组化设计：生产环境中，关节电机、减速器与末端执行器被定义为标准化模组，这一设计不仅优化了后期的维护，也使得本体制造商能够通过通用化的模组供应链，极大压低单台机器人的整机 BOM 成本。

算法补偿硬件：在保证成本的前提下，对于采用中低端力矩传感器或减速器的硬件执行终端，通过端侧部署强化学习算法，实现对硬件运动偏差的主动补偿。在不更换昂贵硬件的前提下，通过软件调优达到工业级执行精度，这种“软件换空间”的工程策略，是实现万元级具身智能设备的核心可行性路径。

4.3 workflow 效率与成本收益分析

从企业长周期运营视角评估，具身智能的价值在于对生产效率的系统性提升：

运维数据闭环：部署一套“数据回流-模型微调”的边缘计算架构，可以使产线作业过程中出现的抓取失败、路径偏移等长尾场景被标记为关键训练数据回流至本地服务器，通过小步迭代的训练，确保智能体在特定生产工序中的泛化能力随时间呈指数级增强。这种机制极大地减少了因模型失效导致的停机时间，保障了产线综合效率的稳健增长。

隐性成本管控：AI 驱动的在线全流程质检能够取代人工离线抽检，在消除质检环节瓶颈的同时，大幅度降低由于质检滞后导致的返工成本。针对多品种小批量的定制需求，智能体的快速换产能力允许企业承接更高溢价的制造订单，实现从“规模依赖型”制造向“范围经济型”制造的结构跨越。

全生命周期成本的优化：通过 RaaS 商业模式，将重资本支出转化为经营性支出，结合智能体带来的持续自动化率提升，使得企业在 18-24 个月内实现投资平衡。

4.4 技术落地演进

具身智能产品的商业化落地应保持节奏的可控性与场景的连贯性：

第一阶段：特定工序渗透，聚焦于危化品作业、高频高危操作及复杂线束插接等刚需场景，重点解决执行的稳定性与可靠性。通过在单一任务场景下实

现 99.9%以上的成功率，建立产品在工业端的口碑。

第二阶段：场景互联，推进不同类型智能体的协同作业。例如，将具备自主移动能力的底盘与精密装配机械臂进行组合，实现物料物流与装配任务的云端协同调度。此时，重点在于构建跨场景的数据资产积累，通过扩大部署数量，驱动核心算法的泛化水平进一步提升。

第三阶段：平台化协同，开放标准化的 API 接口，使不同厂商的具身智能体能够接入企业的统一生产管理系统。在这一阶段，具身智能将作为企业基础设施的一部分，实现生产现场的动态感知与全自动化作业闭环，最终推动制造企业迈向自主化、高韧性的无人制造新阶段。

这一系列策略不仅回应了制造业对降本增效的迫切需求，也为具身智能的开发者提供了从单一硬件制造转向全流程智能服务商的商业逻辑框架。核心竞争力在于将不可控的物理操作，转化为可由 AI 持续优化的数据驱动逻辑，从而实现生产力的跨越式迭代。

5. 演进与愿景

在 2026 年的产业背景下，具身智能的崛起不仅是算法模型的单向进化，更是物理法则、能量管理与精密执行系统的全方位集成。当 AI 获得了物理载体，产品定义的内核也随之发生了根本性的位移：从单一的信息处理逻辑，演变为对“物理执行-环境感知-语义理解”这一深度闭环。

5.1 物理交互与边界探索

具身智能的核心挑战在于物理世界的随机性。相比于规则明确的数字环境，现实场景中充斥着物理规则带来的种种不确定性。将具身智能引入产业流程，本质上是试图通过数字化的“大脑”去驯服极其复杂的物理反馈。

在这一认知维度下，我们需要探索的任务不再仅仅是设计一个交互界面或一组功能逻辑，而是要定义智能体与物理世界交互的边界。每一个关节模组的力矩选择、每一个触觉传感器的采样频率、每一次基于 VLA 模型的语义解析，实际上都是在为物理世界建模。具身智能的演进路径，是将那些原本无法被数字化的物理接触，转化为可被模型理解、可被算力优化的参数集。这种从“模糊感应”到“精确建模”的跨越，正是具身智能商业化落地的第一级台阶。

与此同时，具身智能体必须在有限的电池容量、受限的端侧算力以及严苛的机械寿命约束下完成任务。这意味着，优秀的产品定义必须在算法模型与工程现实之间寻找平衡点。这种平衡能力，决定了具身智能是停留在实验室还是能真正走向产线的资产。

5.2 具身智能产品的核心价值维度

具身智能产品的成功不应由算法的“炫技”来定义，而应由其解决真实痛点的效率来衡量。未来的产品闭环需要聚焦于以下三个核心维度：

- 1、场景洞察：产品定义不应执着于通用型全能机器人，而应深入制造业与服务业。例如，在危化品处理或极端环境下的精密组装场景，人类面临极高的安全风险或生理极限，这正是具身智能发挥“高 ROI”的黄金地带。通过精确识别这些高频、刚需且人力成本高昂的作业单元，将其转化为标准化的智能执行方案，才能实现从技术到商业的闭环。

2、自适应能力：具身智能的核心价值在于“泛化”。一个优秀的智能体不仅要在实验室的平整地面上表现优异，更要在充满油污、振动和随机光影的工厂车间里保持作业精度。因此，产品的核心竞争力在于其“自适应闭环”的能力——即通过多模态感知实时纠偏的能力。

3、持续进化：产品交付不再是交易的终点，而是数据采集与迭代的起点。通过构建科学的边缘数据回流闭环，确保每一台落地的终端设备都能在真实的作业场景中收集长尾样本。这些来自物理现场的真实反馈回传至模型端进行训练，再通过 OTA 推送至全球终端，实现“群体学习”。这种不断进化的能力，才是具身智能产品最核心的技术壁垒。

5.3 结语

2026 年是具身智能行业从技术发散期进入商业深水区的关键节点。随着硬件模组化水平的提升与模型参数规模在端侧的优化，具身智能正在从实验室走向千行百业的生产一线。

从精密器件开发到具身智能系统的跨越，不仅是技术领域的横跨，更是对生产力本质理解的升级。物理世界的每一个细微动作，背后都是算力、感知与机械结构的精密耦合。未来的产品经理不仅要懂 AI 模型的层级结构，更要懂齿轮间的啮合精度，懂传感器的信号噪声，懂产线工人的作业习惯。

作为这场变革的参与者，应当始终坚持以务实的工程视角去解构需求，以严谨的商业逻辑去推动落地。具身智能的未来不在于它能跳多好的舞，而在于它能多稳定地拧好一颗螺丝，多智能地分拣一盒零件。在这场迈向物理-智能融合的新纪元中，我们不仅是在制造机器，更是在定义人类与 AI 协作的新秩序，让智能真正走出屏幕，成为支撑现代工业文明持续进化的物理基石。