CAICT 中国信通院

清华大学电子工程系
Department of Electronic Engineering, Tsinghua University

集智 专题报告

具身智能发展报告

(2026 年)

中国信息通信研究院人工智能研究所
清华大学电子工程系
具身智能测试实验室

2026年10月

版权声明

本报告版权属于中国信息通信研究院和清华大学电子工程系，并受法律保护。转载、摘编或利用其他方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院和清华大学电子工程系、具身智能测试实验室”。违反上述声明者，编者将追究其相关法律责任。

前　言

具身智能是“有身体”的人工智能，被认为是实现通用人工智能（AGI）的关键路径之一，也是重塑人类生产生活范式和全球竞争格局的新领域新赛道。2025 年、2026 年，“具身智能”已连续两年被写入政府工作报告，成为国家层面重点培育的未来产业之一。

过去一年，具身智能技术产业迭代与成果日新月异，但也呈现出冷热两极分化的局面，一方面，资本市场与政策关注持续升温，模型、本体、数据等环节新成果密集发布；另一方面，规模化落地与商业闭环仍面临多重约束，技术演示与真实交付之间尚存在明显落差。技术层面，模型、数据、本体与算力协同演进，以深化物理世界理解和通用行动能力为目标，不断提升智能能力边界。应用层面，2026 年具身智能的部署落地加快推进，文娱表演和竞技赛事持续扩大社会关注度，玩具和陪伴类产品率先进入个人消费市场，越来越多的行业场景推进小批量部署验证，制造、物流、商业服务等领域的试点规模逐步扩大。

具身智能技术产业化的竞争焦点正从单一模型性能或硬件参数，转向“数据—模型—本体—场景”全栈协同能力，尤其是能落地、能交付的商业模式构建，以及涵盖核心零部件、整机制造、模型工具链、系统集成与运维服务的完整产业链能力。与此同时，具身智能复杂接触与精细操作、未知环境适应、长时间自主运行、安全性与可解释性等问题尚未根本解决，具身智能从“可用”走向“好用”、从“演示”走向“规模”仍需系统性突破。

在此背景下，中国信息通信研究院继《具身智能发展报告（2024

年）》《具身智能发展报告（2025 年）》后第 3 次发布具身智能研究报告，本报告聚焦 2026 年以来具身智能产业的新发展新变化新挑战，总结梳理国内外具身智能产业、技术、应用等方面的发展特点，并对未来发展进行展望。由于具身智能发展日新月异，限于编写时间、编写组知识积累水平有限等因素，报告中存在不足之处，敬请大家批评指正。

目 录

一、全球具身智能发展态势.....1
（一）技术范式：具身智能与物理 AI.....1
（二）全球态势：具身智能产业加速扩容.....5
（三）国内现状：具身智能产业热中有冷.....8
二、技术体系：模型、数据、本体与算力的协同演进.....12
（一）模型：技术路线远未收敛，系统闭环落地先行.....15
（二）数据：双线填补数据缺口，供给体系加速构建.....25
（三）本体：部件性能加速迭代，构型配置持续完善.....30
（四）算力：云边端协同演进，系统支撑能力持续完善.....37
三、应用拓展：从展示表演起步向作业模式转化.....39
（一）文娱赛事：面向大众的能力展示场与技术试炼场.....40
（二）个人消费：玩具和陪伴产品率先实现商业变现.....41
（三）生产应用：碎片化场景中的试点验证与应用拓展.....42
四、具身智能产业生态与商业模式.....44
（一）产业链分工进一步明确，系统集成与服务环节作用提升.....44
（二）多元主体加快实质性布局，产业协作关系进一步深化.....47
（三）产业链多环节创造收入，可持续的商业模式仍有待沉淀.....49
（四）产业开源生态建设加快推进，开放协同创新机制初步形成.....55
（五）评测标准实践不断深入，体系化建设与应用衔接进一步深化.....56
五、未来展望.....69
（一）技术趋势：模型架构持续迭代，软硬一体共同定义智能.....69
（二）产业趋势：重点场景引领，基础设施、产业生态和就业体系同步重构.....71
（三）治理趋势：标准体系从“建框架”走向“落地实施”.....73

图 目 录

图 1　具身智能与物理 AI 概念范式融合 4
图 2　具身智能技术体系图 14
图 3　具身智能产业链结构示意图 45
图 4　具身智能产业地域分布趋势 45
图 5　具身智能评测基准概览 57

CAICT 中国信通院

一、全球具身智能发展态势

具身智能（Embodied Artificial Intelligence，EAI）是“有身体”的人工智能，可执行现实世界任务并主动学习进化，是催生颠覆性产品和装备、促进人工智能与实体经济深度融合、重塑人类生产生活范式和全球竞争格局的新领域新赛道。具身智能并不是简单地“给人工智能装上一个身体”，其本质是通过物理实体将人工智能与真实环境建立持续交互，使智能在感知、认知、决策、行动和反馈过程中形成闭环，并在交互中实现能力适应和持续演进。从“实验室”到“实训场”再到“实景场”，2026 年的具身智能站在未来产业向试点应用的过渡节点。

具身智能驱动人工智能不再只是“思考”数字世界，而是开始“行动”于物理世界。2026 年 4 月，《Nature Machine Intelligence》发表题为《From embodied intelligence to physical AI》[1]指出，世界模型、具身智能、生态心理学、神经科学、物理人工智能（Physical AI，PAI）和机器人学等不同研究方向正在逐渐汇聚到一个共同的科学问题：人工智能如何不只是预测、模拟和推理世界，而是真正能够在世界中进行物理的、智能的行动。这表明，人工智能走向物理世界已经不仅是机器人产业发展的单一技术趋势，也正在成为人工智能研究范式演进的重要方向。

（一）技术范式：具身智能与物理 AI

[1] Nature Machine Intelligence. From embodied intelligence to physical AI[J]. Nature Machine Intelligence, 2026, 8: 491-492.

1.具身智能的概念与内涵深化

《具身智能发展报告（2024 年）》[2]首次对具身智能的概念内涵和相关名词进行深入辨析，《具身智能发展报告（2025 年）》[3]进一步分析具身智能核心内涵是基于身体交互的“智能闭环”：**具身智能不仅仅是“人工智能+机器人”，而是人工智能通过物理本体与环境交互实现“知行合一”的综合智能。从智能看，**具身智能的内涵正在进一步向物理世界理解和通用行动能力深化。智能不仅能够预测、模拟和推理世界，还能够在其中自主、可靠地行动。**从具身看**，具身智能的内涵正在进一步验证“具身”并不等同于某一种身体形态，可以有人形机器人、四足机器人、无人车、无人机等多种形态，关键在于能够进入真实环境并自主完成任务。**从路径看，**数据飞轮驱动模型能力提升并实现跨任务、跨场景、跨本体应用，通过“虚实结合”的仿真训练、多模态大模型赋能、端云协同架构以及硬件本体的迭代，最终实现智能体在真实物理世界中的自主感知、决策与执行。

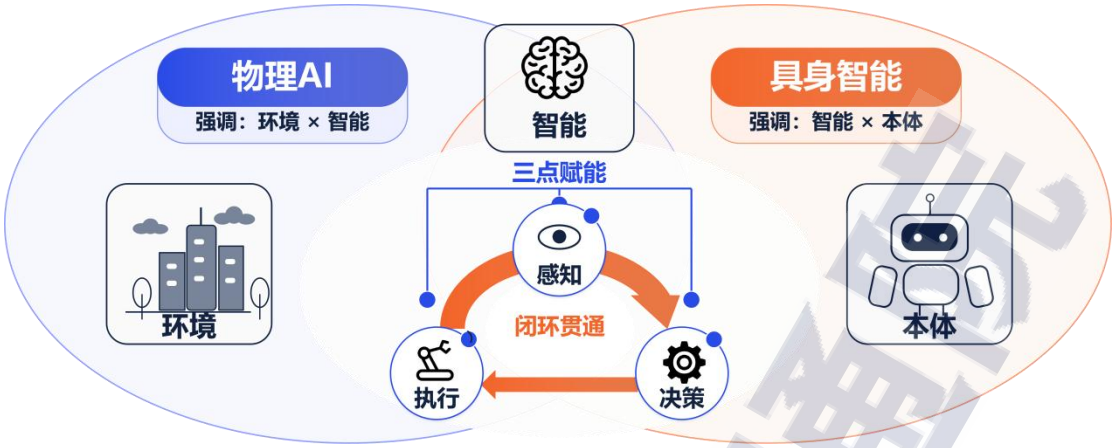
2.具身智能与物理人工智能快速趋同

物理 AI 以突破传统模型“只会表征、缺少物理约束”的局限为目标，自概念提出六年以来，已由学术概念演进为萌芽中的产业。物理 AI 是指通过把物理规律嵌入感知、推理、生成和行动全过程，使人工智能具备更强的真实性、可靠性与可部署性。2020 年，瑞士联邦材料科学与技术实验室与帝国理工学院联合在《自然·机器智能》（Nature Machine Intelligence）上发表论文，指出当前机器人“大脑”

[2] 中国信息通信研究院. 具身智能发展报告（2024 年）[R]. 2024.
[3] 中国信息通信研究院. 具身智能发展报告（2025 年）[R]. 2025.

的算法进化速度远超“躯体”的物理进化速度，首次提出并定义了物理 AI 概念。物理 AI 被定义为合成类自然智能机器人系统的理论与实践，其核心在于打破纯粹依赖计算机科学的现状，将智能视为材料科学、机械工程、计算机科学、生物学与化学五个学科深度融合的产物。2024 年英伟达将物理 AI 定位为 AI 发展的核心方向，2026 年麦肯锡在《The age of thinking machines: Perspectives on the future of robotics》中提出物理 AI 是人工智能的下一个前沿。

具身智能和物理人工智能在核心内涵、技术底座与作用对象层面正在快速趋同，本质是同一技术范式的不同表述。如图 1 所示。**一是强调“智能”核心**，EAI 与 PAI 本质上均为面向真实物理世界的智能技术，旨在使机器人和自动驾驶汽车等自主系统能够感知、理解并执行其环境中的复杂动作。EAI 更强调“智能如何通过身体实现”，PAI 则更强调“人工智能如何进入并作用于物理世界”。**二是共享技术底座**，包括多模态基础模型、视觉—语言—动作模型（VLA）、世界模型、机器人学习、强化学习、物理仿真、数字孪生、合成数据以及端侧智能计算等。EAI 更强调“感知－决策－执行”的闭环能力，PAI 更强调物理规律内嵌于“感知－决策－执行”全链路能力。**三是机器人是理想“试验田”**，机器人是系统在物理世界中行动并实时承受后果的主要载体，也成为检验 EAI 和 PAI 的验证场。EAI 更强调单体或群体智能，PAI 还涵盖了缺乏独立物理躯体、但负责管理和优化物理基础设施的系统。



来源：中国信息通信研究院

图 1 具身智能与物理 AI 概念范式融合

3.通用人工智能终将走向物理世界

物理世界正在成为检验人工智能是否真正走向通用智能的重要试验场。物理 AI 和具身智能的发展，正在改变人们对通用人工智能（AGI）的理解。过去，AGI 讨论更多集中于人工智能是否能够理解语言、获取知识、进行推理以及解决跨领域的数字任务。伴随大模型的快速迭代发展，AGI 讨论逐渐转向在开放环境中完成广泛的人类任务。根据 Stanford 2026 AI Index 报告[4]，一方面，AI 模型在语言、推理、代码和多模态任务上的能力仍在快速提升；另一方面，机器人在真实家庭环境中的任务成功率仍然只有 12%，远低于其在软件仿真环境中 89.4%的操作成功率。2026 年 9 月 OpenAI 发布的 GPT-6 Astra，在完成“将圆形拼图块精确插入对应凹槽（亚毫米级精度与接触控制任务）”时，成功率仅 10%。真正困难的人工智能基准测试，正在越

[4] Stanford Institute for Human-Centered Artificial Intelligence. The 2026 AI Index Report: Technical Performance[R]. 2026.

来越多地集中于真实世界中的行动任务，因为真实环境具有不可预测性，而错误还可能产生实际的物理后果。

（二）全球态势：具身智能产业加速扩容

1.全球投融资市场持续活跃

国外具身智能融资更突出十亿美元级头部项目，模型、技术底座和应用生态成为关注焦点。2026 年 1 月，Skild AI 完成 14 亿美元 C 轮融资，资金用于继续发展能够适配不同机器人本体和任务的通用 Skild Brain[5]。6 月，德国 NEURA Robotics 宣布最高 14 亿美元 C 轮融资，用于 Physical AI 平台 Neuraverse、真实世界训练设施 NEURA Gyms 以及机器人规模生产和全球部署[6]。世界模型在 2026 年成为新的大额资本投入方向，World Labs 于 2 月完成 10 亿美元融资，用于发展空间智能和世界模型，应用范围明确包括机器人[7]。3 月，AMI Labs 完成 10.3 亿美元融资，重点研发能够理解真实世界、具备记忆、推理和规划能力的世界模型[8]。此外，人形机器人企业 Apptronik 于 2 月新增 5.2 亿美元融资，使 A 轮累计融资超过 9.35 亿美元，资金主要用于 Apollo 人形机器人的规模生产和部署[9]。

国内具身智能投融资规模快速扩张、资金向头部和中后期项目逐渐集中、投资主体更加多元。据 IT 桔子统计，2026 年上半年国内具身智能领域发生融资 322 起，融资金额约 935 亿元，较 2025 年上半

[5] Skild AI. Announcing Series C[EB/OL]. (2026-01-14). https://www.skild.ai/blogs/series-c.
[6] NEURA Robotics. NEURA Robotics Announces Record Series C of up to $1.4 Billion to Accelerate the World's Leading Physical AI Platform[EB/OL]. (2026-06-10). https://neura-robotics.com/record-series-c/.
[7] World Labs. World Labs Announces New Funding[EB/OL]. https://www.worldlabs.ai/blog/funding-2026.
[8] AMI Labs. Updates[EB/OL]. https://amilabs.xyz/updates.
[9] Apptronik. Apptronik Closes Over $935 Million Series A with New $520 Million Extension Round[EB/OL]. (2026-02-11). https://apptronik.com/news-collection/apptronik-closes-over-935-million-series-a.

年增长 5 倍，融资事件数同比增长 137%。从融资阶段看，种子轮、天使轮和 Pre-A 轮等早期融资仍占 58.94%，但 A 轮至 B+轮等成长阶段融资金额已占 40%以上，头部项目融资轮次明显后移。资金集中度同步提高，融资额前 20 家企业合计获得约 550 亿元，占总融资额 59%；其中人形机器人本体企业约获 315 亿元，具身智能“大脑”企业约获 145 亿元，核心零部件企业约获 53 亿元，世界模型、AI 仿真等基础技术也开始获得大额投资。同期参与投资机构超过 800 家，形成头部 VC、国资和产业资本共同参与的投资格局。

2.全球范围内企业加速入局

科技大厂从前瞻技术布局转向体系化平台供给，争夺具身智能时代的通用技术底座。一是将通用模型作为连接跨本体和跨场景应用的生态入口。如谷歌推出的 Gemini Robotics 2，以及阿里千问推出的 Qwen-Robot 系列具身智能模型矩阵。**二是**通过模型、云计算、开发平台和产业接口构建开放生态。NVIDIA 构建从芯片到模型的全栈物理 AI 生态、腾讯构建混元具身模型、Tairos 开放平台和 RoboFusion 本体互联体系。**三是**通用大模型厂商也在推动大模型物理化，Anthropic Claude 大模型和 OpenAI GPT-6 Astra 均可作为云端大脑接入外部软件与设备。

汽车及自动驾驶厂商从能力外溢转向业务布局，小规模量产和自有场景验证在加速推进。特斯拉弗里蒙特工厂于 2026 年 7 月搭建 Optimus 本体产线；小鹏汽车加快 IRON 专用产线建设，计划于 2026 年底启动量产，前期优先在自身工厂、园区和门店开展应用。小米则

形成“自研模型 + 自有本体 + 汽车工厂”的闭环布局，进入工业验证阶段。

初创企业分层涌现，协同合作进一步加深。一方面，2026 年创业生态进一步分化为模型、整机、灵巧手和数据工具等层级，专业化分工模式开始显现。**另一方面，**模型、本体、芯片和系统集成的合作深入推进。美国 Skild AI 将通用机器人大脑嵌入 ABB、UR 等厂商的产品线并进入富士康产线。千寻智能开源模型与地瓜机器人具身算力平台完成深度适配。Genesis AI 与 LG 集团旗下系统集成商 LG CNS 建立长期战略合作，由 Genesis AI 提供通用机器人及基础模型，LG CNS 负责工作流改造、系统集成、数据采集、后训练和规模部署。Figure AI 将新一代 Figure 03 导入宝马斯帕坦堡工厂，联合开发零部件排序、柔性取放和料车牵引等移动操作任务。

传统机器人和制造产业借势回暖，新旧动能加速融合。一方面，传统机器人产业直接受益于具身智能带来的增量需求。国家统计局数据显示，2026 年 5 月，机器人减速器、工业机器人和服务机器人等产品产量分别增长 44.7%、27.9%和 19.8%[10]。FANUC 的 2026 年第一季度美洲区机器人业务销售额同比增长 20.9%，中国区同比增长 30.0%[11]。**另一方面，**老牌制造企业正通过收购、入股、自建等多种方式加速入场。韩国三星电子将人形机器人公司 Rainbow Robotics 收为子公司并设立直属 CEO 的未来机器人办公室，现代汽车于 2026 年 7 月收购软

[10] 国家统计局. 工业经济运行平稳 向新向优特征明显——国家统计局工业司首席统计师孙晓解读 2026 年 5 月份工业生产数据[EB/OL]. (2026-06-16). https://www.stats.gov.cn/sj/sjjd/202606/t20260616_1963959.html.
[11] FANUC CORPORATION. Financial Results for the Three Months Ended June 30, 2026[R]. 2026.

银所持剩余股份、全资控股波士顿动力。杭州柯林 5 月公告控股开普勒机器人，完成从电网设备向智能装备的跨界。

（三）国内现状：具身智能产业热中有冷

2026 年，我国具身智能产业在持续升温中透露冷意。**一方面，长期价值牵引，产业热度持续攀升。**具身智能被赋予成为国民级产业的战略预期，有望成为新能源汽车之后的下一个万亿级产业，**未来市场潜力可按照“手机的数量乘以汽车的价格”来估算**。在这一预期牵引下，各方资源加速集聚。**另一方面，泡沫风险正在积聚，产业现实节奏明显遇冷。**演示级能力进步不等于真实产线作业能力，出货量增长不等于持续有效运营，意向订单增加也不等于回款、复购和商业闭环。当前的冷热交织是市场预期前置与产业能力滞后之间矛盾的集中体现，市场端已完成对技术价值的提前定价，产业端仍处于从技术验证向商业验证过渡的初期阶段。

热的一面，政策支持、资本投入和产品供给全面升温。从政策看，支持力度持续加码。从“十五五”规划定调，到实景实训专项行动细化任务，提出年底前推动机器人从“演示表演”转向常态化“作业模式”、落地超万台设备等可落地的指标要求，产业发展的确定性预期显著增强。**从资本看，**融资规模与企业估值激增。据 IT 桔子数据，2026 年上半年国内具身智能赛道融资总额达 935 亿元，仅半年即超过 2025 年全年的 735 亿元，较上年同期提升约 5 倍。据不完全统计，达到百亿估值企业已增至 20 家，且多数侧重具身大脑。**从产品供给看，**新品数量与量产交付能力快速增长。2025 年世界机器人大会发布

新品 123 款[12]，2026 年增至 311 款[13]，同比增长超过一倍。2026 年上半年，全球人形机器人出货量超过 2.2 万台，其中五家国内整机厂商的出货占比合计超过 86%[14]。

冷的一面，技术和部署落地的实际进展明显滞后，距离产业规模化商用拐点仍有一定时间周期。具身智能技术路线尚未收敛，产品主要在少数封闭或半结构化场景中开展示范验证，实际运行仍高度依赖场景改造、定制化调试和人工保障，规模复制能力与单位作业经济性尚未得到充分验证。

技术冷，具身智能“ChatGPT 时刻”存在不确定性。一是前沿模型闭源，技术迭代速度放缓。国内多数企业仍集中于 π_0、$\pi_{0.5}$ 和 OpenVLA 等较早开放模型进行复现和微调实验，这些基座落后前沿闭源模型约两代，产业可获得能力与前沿技术水平之间存在断层。**二是数据效用低，限制了模型能力上限。**2026 年 8 月发布的 Dyna-2 模型，首次在百万小时的数据规模下探讨模型扩展定律。结果发现预训练数据从 10 万小时增长到百万小时后，在 14 个自建任务集上的平均成功率仅从 45%提升至 53%[15]。而国内大多数模型在 10 万小时以内，相差约一个数量级。数据覆盖不足直接限制泛化边界，国内模型在复杂任务和开放环境中将更早触及能力边界。**三是模型部署门槛较高，**

[12] 世界机器人大会官网. 2025 世界机器人大会在京圆满闭幕[N/OL]. (2025-08-13). https://www.worldrobotconference.com/news/latestnews/3117.html.
[13] 央视网. “人机共生，产需共融”，预见未来！2026 世界机器人大会 311 款新品集中亮相[N/OL]. (2026-08-21). https://news.cctv.cn/2026/08/21/ARTIvNwtdbSVIsN7Tl41JgDO260821.shtml.
[14] Counterpoint Research. Global Humanoid Robot Shipments Soar Nearly 300% YoY in H1 2026, Driven by Commercial Deployments[EB/OL]. (2026-08-19). https://counterpointresearch.com/en/insights/global-humanoid-robot-shipments-soar-nearly-300-percent-yoy-in-h1-2026.
[15] Dyna Robotics. Dyna-2: A 1-Million-Hour Scaling Law for World-Action Models[EB/OL]. https://www.dyna.co/dyna-2.

尚未跨越技术产品化鸿沟。模型在泛化能力、作业节拍、持续运行和故障恢复等方面仍有明显短板，可以在限定环境中完成分拣、搬运、上下料、叠衣和物品抓取等任务，但不少能力仍依赖固定工位、预先建图、限定物料或人工接管。复杂任务中表现最好的具身智能基础模型仍普遍需要针对目标任务进行微调，场景、任务或操作对象一旦变化，就需要重新采集数据、微调适配。这意味着落地模式本质上仍未跳出传统机器人“每个场景、每个任务重做一遍”的研发路径，尚未形成可复制、可迁移的规模效应。

应用冷，真实场景落地远小于出货，技术成熟度尚无法满足现实需求。一是出货不等于上岗，部署运营率低。目前出货仍主要流向科研教育、数据采集、展览展示、文娱表演及由政府、产业资本和供应链伙伴推动的试点项目，基于“用工经济性”形成的大规模、长期部署仍然有限。据相关统计，2026 年上半年国内公开披露的具身智能及人形机器人相关中标项中，教育科研机构占比约 55.5%，政府及公共平台占比约 20.6%，工业及技术型企业仅占约 21.1%。实际使用中，相当比例的设备仅用于展厅接待或活动演示，表演档期之外大量空置，月均开机天数不足 10 天的情况并不少见。**二是有订单不等于有市场需求，订单含金量低。**当前公开订单中混杂着正式采购订单、框架协议、战略合作、意向订单以及政策性和示范性采购，在付款条件、验收要求、交付周期和实际作业量上差异很大——框架协议往往只约定合作意向而无采购数量，示范性采购带有任务属性而非市场行为，均不能直接等同于真实商业需求。**三是单机成本不等于落地成本，整机**

价格仍不具备经济可用性。整机价格可能仅占总拥有成本（TCO）的 40%—60%[16]，实际部署还需承担场景改造与系统集成、数据采集与模型训练、运维人员配置以及故障停机损失等支出。单台轮臂式或双足人形机器人的年度综合运维成本约 50 万—70 万元，远高于多数岗位的人力成本。

企业冷，同质化竞争加剧，低门槛市场利润率先出清，行业洗牌已从预期转向经营层面。一是大量供给集中在有限需求。现阶段可落地任务仍主要集中于包裹和零部件分拣、料箱搬运、取放、上下料、质量检查及巡检等少数标准化或半结构化流程，企业扎堆争夺同类客户和相似工位。**二是低门槛市场泡沫快速挤出。**表演租赁市场上，人形机器人日租金从 2025 年初的约 2 万元降至 2026 年的千元级，跌幅八至九成。二手网站和拆机市场上，部分 2024—2025 年购入价为 30 万—80 万元的旧款工程样机，2026 年二手行情已经降至 3 万—6 万元，残值缩水接近九成。早期依赖设备稀缺性面向表演、展示等低门槛场景“赚快钱”的商业模式，正在快速失效。**三是商业营收和经营承压。**多数企业增收不增利，如越疆科技 2026 上半年营业收入达 3.16 亿元，同比增幅 106.6%，净亏损扩大至 1.08 亿元，同比增幅 163.8%，收入仍主要来自传统协作机器人业务，具身智能业务尚未盈利。A 股超 500 家上市公司半年报提及具身智能，但披露相关业务收入的仅个位数。国内一部分具身智能创业企业已处于实质性停滞，仅维持低速

[16] Robotics Center. Humanoid Robot Buyer's Guide 2026: Compare Models, Costs & Use Cases[EB/OL]. (2026-04-12). https://www.roboticscenter.ai/guides/humanoid-robot-buyers-guide-2025.

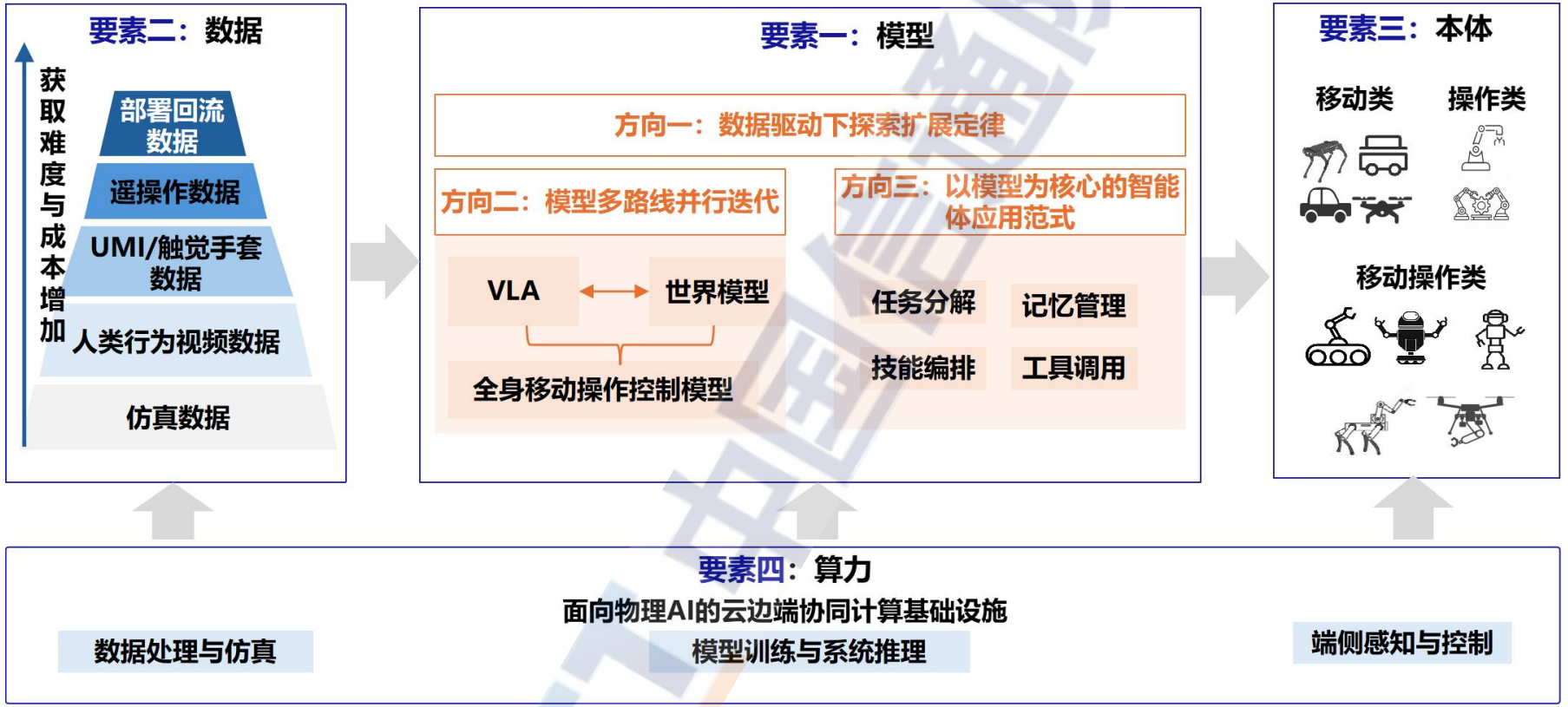
运转状态。一星机器人成立 5 个月即解散、钢铁侠科技最近一次公开融资为 2023 年 12 月，行业洗牌已经启动。

具身智能产业的冷热分化是产业回归理性的必经阶段。从产业演进规律看，硬科技产业在从 0 到 1 的阶段，都伴随阶段性泡沫，泡沫褪去之后，沉淀下来的是经过考验的技术、产能和供应链。2000 年前后的互联网泡沫推动宽带网络、数据中心和互联网基础设施加速建设，2015—2018 年，新能源汽车行业经历产能扩张和泡沫挤压，沉淀出完整的电动车产业链。具身智能产业也正在重复这一进程，**目前正处于期望膨胀期，未来两至三年将通过市场竞争淘汰缺乏真实价值的企业和模式，完成首轮行业出清。**

具身智能作为软硬深度融合的前沿科技产业，其技术进步不会遵循传统制造业的平滑改进，而是可能在 AI 范式驱动下周期性波动发展。技术演进上，在延续人工智能大模型的发展范式快速迭代。模型通用性和泛化能力一旦取得实质性突破，将显著降低跨场景跨本体的迁移和复用门槛。**产业化过程上，可类比新能源汽车。**模型的每一次突破都可能带来新一轮产业升温，但难以直接推动全行业爆发，需要本体可靠性、规模制造、成本控制和运维体系的同步成熟。**商业化路径上，更接近自动驾驶。**需要在局部能力可用时先进入运营，再逐步扩大运行范围。一旦具身智能的通用性、泛化能力和真实环境作业性能实现实质性突破，行业交付模式将实现从项目制定制转向标准化复制，进入长期商业化逐步爬坡阶段。

二、技术体系：模型、数据、本体与算力的协同演进

2025 年，具身智能行业讨论的焦点是技术路线之争：分层架构与端到端架构如何取舍，视觉语言模型（Vision-Language-Model，VLM）、视觉语言动作模型（Vision-Language-Action-Model，VLA）和世界模型谁更可能成为通用具身智能的基础架构。**进入 2026 年，能够统一“感知—决策—执行”闭环的主导架构仍未出现，**扩展定律仍在探索，模型路线远未收敛。在技术路线尚未定型的当下，**行业正在沿着数据、模型、本体和算力四个关键要素同步发力，以系统协同寻求能力突破，**如图 2 所示。目标是构建可部署、可迭代的系统闭环，让数据提供经验，模型形成智能，算力支撑智能运行，本体把智能作用于物理世界，并持续产生新的数据，让具身智能从单点能力突破走向系统部署与持续演进。



来源：中国信息通信研究院

图 2 具身智能技术体系图

（一）模型：技术路线远未收敛，系统闭环落地先行

近年来，端到端具身大模型路线快速发展，以数据驱动方式端到端学习“看懂场景、理解指令、输出动作”的完整映射，天然绕开了人工规则设计和模块拼接的局限。根据解决任务的不同，端到端具身大模型可以进一步分为具身导航大模型、具身操作大模型和全身移动操作控制模型。**一是**具身导航大模型，以视觉语言导航模型（Vision-and-Language Navigation，VLN）为代表。从落地情况看，具身导航能力已凭借传统模块化算法在酒店配送、商场引导、园区巡检等结构化场景中实现规模化应用，是当前三类模型中成熟度最高的方向；具身导航大模型则在此基础上，试图解决开放环境下的零样本泛化问题。例如阿里巴巴高德视觉技术中心提出 ABot-N1[17]，将城市级 POI 到达率提升 35.0%至 77.3%。**二是**具身操作大模型，以 VLA 模型为代表。近两年，VLA 的通用行为策略和专用技能学习能力快速提升。同时，世界模型与 VLA 融合正成为补齐模型物理认知与长程规划短板的关键方向。**三是**全身移动操作控制模型。随着 VLA 快速迭代，模型正在学习“移动+操作”的端到端映射。由于 VLA **的动作空间天然包含底盘自由度，纯导航可退化为其输出空间的一个子集，**VLN **正逐步被** VLA **吸收，移动与操作的边界逐渐模糊。**

整体来看，导航能力已相对成熟且正在被具身操作大模型整合，**当前行业关注的核心问题已经收拢为：如何突破物理认知和行动能力**

[17] GONG R, GUO Y, HU J, 等. ABot-N1: Toward a General Visual Language Navigation Foundation Model. arXiv:2607.10383, 2026.

短板，让具身智能在真实环境中学会“动手干活”。围绕这一问题，行业正沿着三个方向探索，**一是**延续大语言模型的规模化路线，推进端到端具身大模型的扩展定律研究。**二是**多路线并行推动模型快速迭代，围绕 VLA、融合世界模型、全身移动操作控制模型三条路线，实现认知和行动能力突破。**三是**构建智能体应用范式，在模型能力尚不充分的当下，借助任务分解、工具调用、人在回路等智能体范式先行推进真实场景部署，并以部署中回流的真实数据驱动模型的自我迭代与进化。

1.方向一：在数据驱动下探索具身智能模型的扩展定律

数据驱动下的模型性能仍在持续增长。一方面，VLA 探索真实物理交互数据扩展下性能增长，首次呈现出类似 GPT-3 的上下文学习能力。2026 年 4 月，美国 Generalist AI 发布 GEN-1 模型，经过 50 万小时数据的预训练后，仅需 1 小时机器人数据就能学会折叠衬衫、折叠纸盒等全新任务，且平均任务成功率达到 99%。8 月推出的 GEN-1.5 将适配成本从小时级数据压缩至单次演示，仅凭一段 3 秒至 12 秒的动作演示，即可在拉开笔袋拉链、旋开玻璃罐盖等 10 项新任务上取得 59%的平均任务成功率。**另一方面，**世界动作模型（World-Action Model, WAM）路线依赖成本更低、更容易获取的人类视频数据，验证另一条数据扩展路径。2026 年 8 月，Dyna Robotics 发布的 Dyna-2 用超过 100 万小时的第一人称人类操作视频预训练世界动作模型（World-Action Model, WAM），在客户现场零样本部署通过率从 46%提升至 87%。

具身智能领域难以复刻类似大语言模型领域的扩展定律，需要更多的数据工程、训练工程和系统工程。一是数据工程，数据组合与配比本身即难题。大语言模型的语料相对同质，而机器人数据天然异构——来自不同本体、不同任务、不同采集者的数据混合在一起，混合方式直接影响训练效果。已有研究表明，简单扩大异构数据规模并不必然带来性能提升，在统一末端执行器坐标系的前提下，仅使用 Open X-Embodiment 数据的模型在 LIBERO 基准上取得 77.3%的成功率，而混入异构真实机器人数据后降至 73.8%。清华大学研究发现，策略泛化能力与训练环境、物体的数量呈近似幂律关系（Pearson 相关系数接近-1）[18]。**二是训练工程**，模型预训练损失与真实任务价值之间缺乏稳定对应关系，损失函数的下降和物理世界中的任务成功率没有必然关联。**三是系统工程**，基础模型与真实执行之间存在明显差距。语言模型的输出是通用文本，而机器人的动作指令与本体强绑定。同一套模型输出的动作指定，跟本体动作执行空间、关节运控频率强绑定。模型部署到本体上通常仍需逐一处理技能或 API 连接、标定与控制接口对齐、接触与力的执行、安全与恢复机制等环节[19]。

2.方向二：多路线推动具身智能模型的认知和行动能力突破

路线一：延续 VLA，协同推动通用行为策略迭代和专用部署落地。VLA 仍是现阶段实现通用操作策略的主要实现路线，也最接近

[18] LIN F, HU Y, SHENG P, 等. Data Scaling Laws in Imitation Learning for Robotic Manipulation. arXiv:2410.18647, 2024.

[19] DOMAE Y, SHIRAI K, OH H, 等. The Embodiment Gap in Robot Foundation Models. arXiv:2608.18433, 2026.

真机部署落地。2026 年以来，VLA 模型正从实验室演示加速走进真实生产生活场景。在工厂，美国 Figure AI 搭载 Helix 02 模型的 Figure 03 人形机器人于 6 月进驻宝马工厂，承担物流与零部件排序作业；在门店，蚂蚁集团灵波科技让乐聚、星海图和自研 R-2 三台不同构型的机器人共用 LingBot-VLA 2.0 一个“大脑”，打造智慧药房解决方案；在家庭，美国 Sunday Robotics 的轮式机器人 Memo 搭载 ACT-2 模型，进入员工家庭和民宿实测叠衣服任务，计划秋季启动家庭试用。部署落地的提速，对模型能力提出了双重要求：既要有更强的通用基座支撑快速适配，也要通过后训练手段保障专用任务落地。

VLA 加快贯通感知—决策—执行闭环，构建通用行为策略底座。一是增强物理感知能力，由固定视觉向三维空间感知和多模态融合对齐演进。上海复旦大学于 2026 年 1 月提出 ActiveVLA[20]，根据任务从三维场景中选择更合适的虚拟观察视角，并放大关键操作区域，减少遮挡和细节缺失对动作判断的影响。法国国家信息与自动化研究所（Inria）研究团队于 2026 年 5 月提出 PointACT[21]，引入多尺度三维点云，使机器人在生成动作时能够兼顾局部形状细节和整体空间布局，提高精细操作中的空间判断能力。新加坡南洋理工大学于 2026 年 7 月提出 VistaVLA[22]，利用三维高斯表征将物体的语义信息与三维位置、形状联系起来。**二是**增强长程任务决策可靠性，由直接动作映射向推

[20] LIU Z, GU Y, WANG Y, 等. ActiveVLA: Injecting Active Perception into Vision-Language-Action Models for Precise 3D Robotic Manipulation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026.

[21] CHEN S, PACAUD P, SCHMID C. PointACT: Vision-Language-Action Models with Multi-Scale Point-Action Interaction[C]//Robotics: Science and Systems XXII, 2026.

[22] LIU M, GU Z, CHEN X, 等. VistaVLA: Geometry- and Semantic-Aware 3D Gaussian-Grounded VLA for Robotic Manipulation. arXiv:2607.12356, 2026.

理、预测和记忆协同演进。北京大学于 2026 年 6 月提出 AffordanceVLA[23]，将操作过程组织为“操作哪个物体—接触哪个位置—如何执行动作”，结合目标定位、接触区域判断和三维空间关系分析，将抽象指令转化为更明确的操作依据。2026 年 3 月，Physical Intelligence 发布 MEM[24]，将短期视频记忆与长期语言记忆结合：前者保留近期观察，缓解遮挡造成的信息缺失；后者记录已完成阶段和关键事件，支持任务进度跟踪及策略调整。9 月，新加坡南洋理工大学提出 MemBodied[25]，通过固定容量的循环记忆保留交互过程及初始场景信息，使机器人能够利用此前观察到、但当前已不在视野内的信息，为后续行动提供依据。**三是**增强跨本体动作策略泛化和反应式行动能力，由本体专用的动作学习向跨本体迁移、低延迟执行和接触反馈操作演进。StarVLA 开源研究社区于 2026 年 8 月发布 VLAct[26]，通过多动作头联合训练和跨本体动作语义共享，使有限的机器人轨迹能够跨任务、跨环境、跨本体复用。北京大学于 2026 年 7 月提出 Jetson-PI[27]，针对模型计算期间环境可能已经变化的问题，预测动作实际执行时的环境状态，并结合端侧推理优化，减少“依据旧画面生成动作”的时间错位。美国密歇根大学于 2026 年 6 月提出 TAP-VLA[28]，将触觉反

[23] KARCINI E, MEHRBAN F, NGUYEN Q, 等. Robots Need More than VLA and World Models. arXiv:2606.06556, 2026.
[24] TORNE M, PERTSCH K, WALKE H, 等. MEM: Multi-Scale Embodied Memory for Vision Language Action Models. arXiv:2603.03596, 2026.
[25] PALA T D, MAJUMDER N, GOH B, 等. MemBodied: Recurrent Associative Memory for Vision-Language-Action Models. arXiv:2609.28256, 2026.
[26] YANG S, WANG C, CHEN Y, 等. Beyond Data Scaling: Representation-Centric Continued Pre-training for Vision-Language-Action Models. arXiv:2608.27550, 2026.
[27] YANG Z, WANG Q, WANG Y, 等. EagleVLA: Towards Onboard Real-Time Robot Control via Foresight-Aligned Asynchronous Inference. arXiv:2607.12659, 2026.
[28] VAN DER MERWE M, SHEHAB M L, LEE J, 等. TAP-VLA: Tactile Annotation Prompting for Vision Language Action Models. arXiv:2606.29089, 2026.

映的受力方向标注到视觉图像中，在不改变 VLA 架构的情况下引入接触反馈，提高接触密集型操作的表现。它石智航 OmniVTA[29]采用慢—快分层机制，慢系统预测接触变化并规划动作，快系统通过 60 Hz 触觉反射控制器，根据预测与实际反馈之间的偏差及时修正动作，帮助机器人恢复中断的接触并避免用力过大，在擦拭、切菜等接触密集型任务上的成功率超过 80%。

VLA 结合强化学习，不断打磨、验证和优化专用技能。物理环境充满随机性与复杂变量，多数 VLA 模型难以开箱即用，强化学习成为后训练环节支撑行为策略快速迭代的主要技术路线。具体进展包括三个方面：**一是**用真机强化学习打磨行为策略，满足生产所需的精度和节拍要求。2026 年 3 月 Physical Intelligence 提出 RL Tokens[30]框架，仅用 10 分钟在线强化学习，便将 VLA 模型在螺丝安装任务上的成功率提升 40%，任务最困难阶段的执行速度最高提升 3 倍。7 月，上海期智研究院提出 RL-100 框架[31]，采用“离线迭代为主、在线强化学习兜底”的混合范式，离线阶段将 8 项任务平均成功率提升至 91.8%，真机阶段再补充数百次试验消除边缘失效，最终在 8 类真实任务上实现千次执行零失败。**二是**测试时适配将策略优化前移至推理过程，省去部署前的再训练。2026 年 7 月，密苏里大学堪萨斯城分校发表 TT-VLA[32]框架，让模型在推理阶段依据任务进度信号实时调整动作，显

[29] ZHENG Y, GU S, ZHENG Y, 等. OmniVTA: Visuo-Tactile World Modeling for Contact-Rich Robotic Manipulation. arXiv:2603.19201, 2026.
[30] XU C, SPRINGENBERG J T, EQUI M, 等. RL Token: Bootstrapping Online RL with Vision-Language-Action Models. arXiv:2604.23073, 2026.
[31] LEI K, LI H, YU D, 等. Performant robotic manipulation with real-world reinforcement learning[J]. Science Robotics, 2026, 11(116): eaed6267.
[32] LIU C, LIU Y, WANG T, 等. On-the-Fly VLA Adaptation via Test-Time Reinforcement

著提升扰动环境下的鲁棒性。7 月，英伟达发布 RoboTTT[33]，通过维护一组快速更新的权重，持续压缩并更新与当前任务最相关的状态信息，可完整执行一项耗时 5 分钟、包含 10 个阶段的组装任务。**三是**部署时学习构建真实场景中的数据闭环。每台部署的机器人都是数据回流的入口，机器人部署越多、运行得越久，模型积累经验的速度就越快。2026 年 5 月，上海创智学院提出集群式离线到在线强化学习框架 LWD[34]，汇总 16 台双臂机器人的自主运行、失败与人工干预数据进行离线预训练，再将统一策略重新部署并实时汇入在线数据，在云端持续更新。该框架在 8 项真机任务上平均成功率达 95%，较传统监督微调提升约 20%，长程任务成功率也有明显提升。

路线二：融合世界模型，拓展可用的数据基础，强化物理认知能力。2026 年以来，具身世界模型的作用正从生成未来画面，向辅助规划、支持策略训练和直接生成动作拓展。具身世界模型的价值主要在于两方面：**一方面，**世界模型可以从海量缺乏动作标注的人类操作视频中学习物理交互行为，拓展具身智能领域可用的数据基础。**另一方面，**具身世界模型相较 VLA 模型，有了更多的想象力，动手之前，先在虚拟空间中预演动作将带来的后果，做对了再执行，做错了在“脑海”中重来，学习新技能不必完全依赖昂贵的真机示范和物理试错。按照其参与动作决策的深入程度，可归纳为三条技术路线。

一是级联式世界模型架构，世界模型作为外部辅助，保留世界模

Learning[C]//Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2026: 40107-40125.

[33] NVIDIA. RoboTTT: Context Scaling for Robot Policies[R]. 2026.

[34] WANG Y, LI X, XIE P, 等. Learning While Deploying: Fleet-Scale Reinforcement Learning for Generalist Robot Policies. arXiv:2605.00416, 2026.

型与 VLA 的模块化分工，动作仍由独立策略模块生成。Physical Intelligence 于 2026 年 4 月发布的 $\pi_{0.7}$[35]则采用分层提示架构，世界模型生成视觉子目标，引导 VLA 生成更高质量的动作策略。**二是 VLA 和世界模型的协同架构，**将未来预测与策略训练结合，使世界模型和 VLA 在交互反馈中共同改进。在这一架构中，世界模型可以支持 VLA 预测轨迹的多轮尝试（Rollout），让 VLA 策略在连续预演和尝试执行中，获得可用于强化学习的策略优化经验。新加坡国立大学于 2026 年 2 月提出的 World-VLA-Loop[36]利用世界模型反复试错迭代，通过强化学习改进行为策略；再将改进后的策略部署到真实环境，采集新的成功与失败轨迹，用于修正世界模型，进而支持下一轮策略训练。2026 年 2 月，极佳视界发布 GigaBrain-0.5M*[37]，借助世界模型推演动作后果并预判价值，再通过强化学习优化 VLA 策略，在叠衣、装箱和咖啡制作等复杂任务中，其 RAMP 方法较 RECAP 基线提升约 30%，表明世界模型提供的推演信号比单纯的成败奖励更能指导策略学习。**三是统一世界动作模型，**将未来预测、动作生成和动作评价纳入同一模型，世界模型成为动作决策主体。2026 年 2 月，英伟达发布 DreamZero[38]，在同一扩散过程中联合生成未来视频和动作，在未见物体和环境上性能表现明显提升，平均任务进度达到 62.2%，相比 $\pi_{0.5}$

[35] Physical Intelligence. π 0.7: a Steerable Generalist Robotic Foundation Model with Emergent Capabilities[R]. 2026.
[36] LIU X, BAI Z, CI H, 等. World-VLA-Loop: Closed-Loop Learning of Video World Model and VLA Policy. arXiv:2602.06508, 2026.
[37] GigaBrain Team, WANG B, LI B, 等. GigaBrain-0.5M*: a VLA That Learns From World Model-Based Reinforcement Learning. arXiv:2602.12099, 2026.
[38] YE S, GE Y, ZHENG K, 等. World Action Models are Zero-shot Policies. arXiv:2602.15922, 2026.

提升 34.8%；生数科技于 2026 年 4 月推出 MotuBrain[39]，通过四层数据金字塔架构让模型学习异构多模态数据中的世界知识。中科第五纪发布的 FAM 2.0[40]，通过将光流匹配与世界模型深度融合，把不同机器人的动作统一编码为光流序列，构建与硬件形态解耦的通用动作表示空间。

路线三：探索全身移动操作控制，端到端的行为策略向底层运控延伸。以人形机器人为例，人形机器人的移动和操作空间相互耦合。伸手取物会改变身体重心，迈步又会改变操作范围和接触状态。传统工程范式将行走、停靠、抓取等环节切分开来各自研发，再由人工编写的规则衔接，控制栈中充斥着硬编码逻辑，每适配一个新场景都要重新调参，严重限制了技术迭代速度。2026 年以来，端到端模型开始自上而下打通“任务理解、全身动作、底层执行”全链路，用可学习的网络替代手工规则，具体包括两条技术路线。

一方面，拓展全身端到端策略，将 VLA 从上半身控制扩展到全身关节控制。2026 年 6 月，清华大学提出 OpenHLM[41]，将语言和视觉输入映射为全身动作指令，再由底层控制器执行，以不到一半的示教时间超过 GR00T N1.6 等基线。同期，复旦大学团队提出的 WholeBodyVLA[42]将操作知识与移动知识编码到统一潜在动作空间，从无动作标注的第一人称视频中学习全身移动操作。**另一方面，构建**

[39] Motubrain Team, XIANG C, BAO F, 等. Motubrain: An Advanced World Action Model for Robot Control. arXiv:2604.27792, 2026.
[40] CHEN Y, LI P, XU Y, 等. FlowWAM: Optical Flow as a Unified Action Representation for World Action Models. arXiv:2607.13017, 2026.
[41] HU Y, ZHU H, ZHENG B, 等. OpenHLM: An Empirical Recipe for Whole-Body Humanoid Loco-Manipulation[EB/OL]. https://openhlm-project.github.io/.
[42] OpenReview. Paper OCJmVjyzN7[EB/OL]. https://openreview.net/pdf?id=OCJmVjyzN7.

学习型全身控制器，以分层协同方式支撑全身动作执行。Figure AI 于 2026 年 1 月发布 Helix 02[43]，通过 1000 小时数据预训练学习底层控制器 System 0，以更高频率负责平衡与接触稳定，直接替代约 11 万行手工编写的 C++控制代码。8 月，英伟达推出 SONIC 全身控制器[44]，基于超过 700 小时人体动捕数据训练出通用运动基础模型，可作为标准化控制接口与 VLA 协同，完成电钻拿取、装箱、搬运、放置等多阶段任务。

3.方向三：构建智能体应用范式，构建可编排可自主改进的闭环系统

具身智能走向真实部署，不仅需要生成动作，还需要完成任务分解、技能调用、过程监控和失败恢复等环节，并将运行经验沉淀为持续改进的基础资源。2026 年以来，相关研究开始通过智能体范式统筹上述环节，推动通用模型、专用策略、控制器和机器人接口协同运行。根据智能体承担的主要功能，可归纳为三个方向。

一是系统接入与任务编排，让大模型复杂组织和指挥已有技能和硬件设备。通用模型负责理解目标、选择技能和检查进度，专用策略与控制器负责具体动作执行。2026 年 7 月，Anthropic 的 Claude Plays Robotics[45]研究发现，大模型接入预训练运动策略和 VLA 后，机器人任务表现明显改善，且新一代模型与专用策略的配合更加顺畅，验证了“高层推理指挥、专用能力执行”分工的可行性。同期，清华大学

[43] Figure AI. Introducing Helix 02: Full-Body Autonomy[EB/OL]. https://www.figure.ai/news/helix-02.
[44] LUO Z, YUAN Y, WANG T, 等. SONIC: Supersizing motion tracking for natural humanoid whole-body control[J]. Science Robotics, 2026, 11(117).
[45] Anthropic. Claude plays robotics[EB/OL]. (2026-07-09). https://www.anthropic.com/research/claude-plays-robotics.

提出 Harness VLA[46]，利用大模型学习 VLA 技能的编排与复用策略，完成复杂长程操作任务，在 LIBERO-Pro 基准上的任务成功率较最强相关基线提升 38.6 个百分点。8 月，Anthropic 进一步发布模型硬件标准（Model Hardware Standard）[47]，探索以统一接口连接不同实验设备，为智能体协调硬件操作提供基础。**二是采集、训练与评测自动化，**让智能体学会组织实验，把策略改进变成可重复运行的流水线。2026 年 3 月，智元机器人提出 RoboClaw[48]，构建可自动复位的数据采集循环，统一采集、学习和部署流程，在实验中减少 53.7%的人力时间投入。6 月，英伟达提出的 ENPIRE[49]将自动复位、结果验证、策略改进、真机执行和日志分析连接起来，使编码智能体能够读取实验反馈、查阅文献并修改训练算法，在灵巧操作实验中自主训练出成功率达 99%的策略。**三是运行诊断与技能积累，**让智能体从失败经验中自主进化，把修复经验转化为可复用能力。2026 年 6 月，英伟达 GEAR 实验室联合密歇根大学、伊利诺伊大学厄巴纳-香槟分校等机构提出 ASPIRE[50]，通过分析多模态执行记录自动诊断失败、修改控制程序，并将验证有效的修复沉淀为技能库。

（二）数据：双线填补数据缺口，供给体系加速构建

行业已多次观测到数据－模型扩展定律，模型泛化能力有望随高

[46] ZHANG Y, ZHANG H, GAO F, 等. Harness VLA: Steering Frozen VLAs into Reliable Manipulation Primitives via Memory-Guided Agents. arXiv:2607.08448, 2026.
[47] Anthropic. Previewing the Model Hardware Standard[EB/OL]. (2026-08-27). https://www.anthropic.com/news/model-hardware-standard-research-preview.
[48] LI R, ZHOU Y, ZHU Y, 等. RoboClaw: An Agentic Framework for Scalable Long-Horizon Robotic Tasks[EB/OL]. https://roboclaw-agibot.github.io/.
[49] XIAO W, XIE J, ZHANG T, 等. ENPIRE: Agentic Robot Policy Self-Improvement in the Real World. arXiv:2606.19980, 2026.
[50] LU R, WU Y, KOU E, 等. ASPIRE: Agentic Skills Discovery for Robotics. arXiv:2607.00272, 2026.

质量数据供给增长呈现可预期的提升。据测算[51]，要实现机器人“ChatGPT”时刻，模型需要千万小时级有效数据。然而，与大语言模型直接利用的互联网数据不同，具身智能无法沿用存量数据。大语言模型能够从互联网获取万亿级词元（token）[52]，具身智能所需的视觉、动作、关节扭矩、力触觉等多模态交互数据，从未被系统性记录，也无法从互联网直接获取，只能从零逐条生产。在高数据需求与大规模数据缺口牵引下，国内数十家公司已进入具身数据赛道。据 IT 桔子数据，近一年内，具身智能数据赛道出现 97 起融资事件，涉及 34 家公司，披露金额超 200 亿元。

1.仿真合成与真实数据并行演进，能力边界渐趋清晰

行业正沿仿真合成与真实世界采集两条路线并行建设。人工智能科学家李飞飞在公开访谈中指出，仿真合成数据与真实数据并非相互替代，而是互补关系[53]。一方面，仿真合成数据规模化能力强、成本可控，具备模拟长尾情况与极端环境的潜力；另一方面，真实物理环境中积累的数据如实反映具身智能与物理世界交互状态，缩短从技术到场景落地的距离。

仿真合成持续推进 Sim-to-Real 迁移与验证，支撑真实环境部署。一方面，弥补零样本或少样本 Sim-to-Real 迁移差距。上海人工智能实验室提出 SIM1[54]，面向可变形物体构建物理对齐的数据引擎，在真

[51] Malik R. Scaling to a Trillion Tokens in Robotics[EB/OL]. https://rayansmalik.substack.com/p/scaling-to-a-trillion-tokens-in-robotics.
[52] Bessemer Venture Partners. Bessemer Predicts: Robotics and Physical AI[EB/OL]. (2026-04-16). https://www.bvp.com/atlas/bessemer-predicts-robotics-and-physical-ai.
[53] Andreessen Horowitz. Fei-Fei Li on Spatial Intelligence and Robotics[EB/OL]. (2026-07-28). https://a16z.com/podcast/fei-fei-li-on-spatial-intelligence-and-robotics/.
[54] ZHOU Y, LIU H, JIANG X, 等. SIM1: Physics-Aligned Simulator as Zero-Shot Data Scaler in Deformable Worlds. arXiv:2604.08544, 2026.

实场景中部署仿真训练模型的零样本成功率达 90%。密歇根大学提出 SBLR[55]，避开逐传感器建模的路线，通过瓶颈隐空间重建迁移触觉策略，在真实机器人轴孔装配等任务上取得 85%—97.5%的零样本成功率。**另一方面，**拓展世界模型对模型策略的评估验证能力。清华大学与斯坦福大学联合提出 Ctrl-World[56]，支持策略在想象空间中闭环评估与改进，对陌生指令的执行成功率由 38.7%提升至 83.4%。英伟达提出 Hydra-0[57]，回放成功率与真实成功率的皮尔逊相关系数达 0.96。

真实数据沿真机数据和无本体数据双线供给。一方面，从单一成功演示向多样化质量分布演进，挖掘失败恢复与部署回流数据价值，不再局限于成功遥操作。Physical Intelligence 发布 $\pi_{0.7}$[58]，采用高质量遥操作、低质量和失败遥操作数据，标注质量评分与错误标签作为动作专家监督信号，使得模型能同时从不同质量、不同策略的遥操作轨迹里蒸馏出最佳行为。加州伯克利大学和英伟达等机构发布 CLIFT[59]，将部署时奖励反馈转化为监督数据，经过两个闭环回流微调后，将密集接触任务操作成功率提升至 96%以上。**另一方面，从固定工位走向开放式采集环境，**使用通用操作接口（Universal Manipulation Interface，UMI）夹爪、头戴式相机、腕部相机、数据手套等设备进入真实场景采集数据。**一是**采集数据真实性和多样性强，规模扩增迅速。京东使

[55] YANG F, WI Y, YU J, 等. Tactile Sim2Real without Tactile Simulation via Bottlenecked Latent Reconstruction. arXiv:2608.15897, 2026.
[56] GUO Y, SHI L X, CHEN J, 等. Ctrl-World: A Controllable Generative World Model for Robot Manipulation. arXiv:2510.10125, 2025.
[57] LI H, WEN B, ZHU X, 等. Hydra-0: Action Flow for Generalist World Modeling and Control. arXiv:2608.18077, 2026.
[58] Physical Intelligence, AI B, AMIN A, 等. π 0.7: A Steerable Generalist Robotic Foundation Model with Emergent Capabilities. arXiv:2604.15483, 2026.
[59] CHEN Y, SRIKANTH H, JEW N, 等. CLIFT: Turning Gemini Robotics On-Device into Humanoid Specialists via Non-Invasive Closed-Loop Iterative Fine-Tuning. arXiv:2607.29172, 2026.

用头戴式采集设备 JoyEgoCam，让采集者在真实的家务、收纳、清洁、物流等场景中完成连续操作任务，形成 1680 小时的开源人类第一人称视角数据集 EgoLive[60]。该数据集预训练的 JoyAI-RA[61]模型已部署在智元和睿尔曼等厂家本体上，进驻京东 MALL 零售场景。**二是**摆脱本体构型约束，提升模型跨本体泛化能力。清华大学 RDT2[62]采用超 1 万小时家庭场景 UMI 数据集进行预训练，并采用与机器人末端相机同型号的 UMI 相机、执行器夹爪和 6-DoF 位姿，无需后训练即可实现跨本体泛化部署，在叠衣服、整理餐桌、拉拉链等复杂任务上表现优于 $\pi_{0.5}$。随着设备门槛进一步降低，无本体数据采集设备有望向 AI 眼镜、智能手机等更便捷、易普及的形态演进，为具身智能积累更广泛的真实世界经验。

2.工程化平台沉淀高质量数据，资产化流通机制初步构建

为破解真实数据采集成本高、产能低、难复用的问题，行业与政策共同发力数据供给设施建设。近两年，我国掀起具身智能训练场建设热潮，通过搭建应用场景，在真实物理环境中支撑具身智能数据生产、模型训练、系统应用能力验证和迭代优化。据中国信息通信研究院《具身智能训练场研究报告（2026 年）》[63]统计，截至 2026 年 6 月底，全国建成启用超 70 家具身智能训练场，已覆盖国内超半数省份。

[60] LI Y, WEI X, LUO J, 等. EgoLive: A Large-Scale Egocentric Dataset from Real-World Human Tasks. arXiv:2604.23570, 2026.
[61] ZHANG T, YUAN Z, CHI D, 等. JoyAI-RA 0.1: A Foundation Model for Robotic Autonomy. arXiv:2604.20100, 2026.
[62] LIU S, LI B, MA K, 等. RDT2: Exploring the Scaling Limit of UMI Data Towards Zero-Shot Cross-Embodiment Generalization. arXiv:2602.03310, 2026.
[63] 中国信息通信研究院. 具身智能训练场研究报告（2026 年）[R]. 2026.

具身智能数据建设正从零散采集走向体系化供给，从企业内部资源逐步转向**可生产、可评估、可流通**的生产要素。

数据生产加快工程化平台建设，自动化清洗、质检与标注等工具链持续完善。分环节看，**一是**清洗标注环节，卡尔斯鲁厄理工学院与英伟达联合提出 SPARC[64]，自动生成空间标注，通过可靠性分数进行筛选，保留高精度样本，相比基线方法标注比例提升 3 倍。均普智能工业具身智能数据平台 Primus Forge[65]以视觉语言大模型替代人工完成 80%标注；**二是**质检验证环节，西北大学和斯坦福等联合研究 MANIGUARD[66]，以时序逻辑监视器逐步质检，在超 2.3 万次执行中检出 6.0%至 21.0%的“成功但违规”样本；**三是**回流闭环环节，通过回流部署后实际应用场景数据，获取第一手信息，用于进一步提升模型能力。Physical Intelligence 发布强化学习框架 RECAP[67]，自动回流机器人在真实部署中自主作业的数据与人工遥操作纠错数据，持续更新策略并同步回机器人集群。

业界探索构建质量评价体系与标准，**围绕“什么是高质量具身智能数据”形成初步共识。**上海人工智能实验室联合上海交通大学等机构在研究[68]中为每条样本构建统一表征，用表征之间的熵代表数据多样性，用表征的记忆难度、泛化难度代表数据可学习性。工业和信息

[64] BLANK N, MATTES P, LI M X, 等. SPARC: Reliable Spatial Annotations from Robot Demonstrations at Scale. arXiv:2606.13497, 2026.
[65] 均普智能. 聚焦工业场景 | 均普智能具身智能数据平台 Primus Forge 正式发布！[EB/OL]. (2026-07-08). https://cn.piagroup.com/news/company2/537.html.
[66] PENG Y, WANG P, ZHAN S S, 等. MANIGUARD: A Benchmark and Data Suite for Specification-Grounded Safety Evaluation and Improvement of Robotic Manipulation. arXiv:2608.17386, 2026.
[67] Physical Intelligence, AMIN A, ANICETO R, 等. π*0.6: a VLA That Learns From Experience. arXiv:2511.14759, 2025.
[68] XIAO J, YAN B, ZHANG J, 等. Data Assessment for Embodied Intelligence. arXiv:2511.09119, 2025.

化部发布行业标准《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》，从完整性、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统一质量评价口径。

数据流通尚处于起步阶段，沿交易与合规展开探索。一是多种交易模式初步落地，例如数据超市、数据交易所等平台形式，可汇聚多数据厂商产品，便捷选购。京东发布京东云具身智能数据交易平台[69]，在平台上明确数据类型、标识码、采集方式、场景、任务分级以及购买方使用权限，进一步避免确权争议。鹿明机器人推出 FastUMI Pro 数据超市[70]，将操作数据按场景和具体任务分类，并透明定价。箸境智能四大场景数据集在江苏省数据交易所完成全国首单具身智能数据集场内交易。**二是**数据合规同步推进。我国网络安全法、数据安全法、个人信息保护法为具身智能数据处理活动提供了基本法律依据，《网络数据安全管理条例》进一步对涉及自动化工具、生成式人工智能等新兴技术的数据处理活动作出特殊规定。国际层面，主要经济体也在加快将具身智能纳入既有数据治理框架。欧盟《数据法》将机器人作为互联产品纳入数据访问、共享与再利用的强制性规则范围;《人工智能法案》采用风险分级监管框架，已将部分具身智能应用纳入高风险类别规制。

（三）本体：部件性能加速迭代，构型配置持续完善

2026 年，具身智能本体仍然呈现多元化发展格局，包括双足人

[69] 京东云. 具身智能数据交易平台[EB/OL]. https://robotdata.jdcloud.com/.
[70] Lumos. Lumos 数据商城[EB/OL]. https://data-market.lumosbot.tech/.

形、轮臂、四足及轮足等主要构型，硬件迭代主要体现于核心零部件性能改进与整机结构配置优化。零部件层面，关节驱动、灵巧手、电子皮肤和端侧计算成为主要优化方向。整机层面，不同构型在全身活动范围、承载能力、操作空间和环境适应性上呈现不同配置侧重，部分产品通过快速换电与自主补能改善持续运行条件。

1.核心零部件性能改进与功能集成同步推进

核心零部件的性能指标不断迭代，进一步加强电、感、控、算一体化的功能集成。其中，关节驱动兼顾动力输出、轻量化与运行损耗，灵巧手加强长时间抓持和耐久验证，传感器与电子皮肤改善受力辨识和接触耐磨性，端侧计算部件考虑算力、功耗与多传感接入配置，电池系统完善储能能力、优化补能方式。

关节一体化设计优化，动力输出、能量转换效率和散热的综合表现升级。为了在有限安装空间内提高输出能力，国内关节产品通常采用电机、减速器、编码器与驱动电路的一体化配置，并通过改进绕组、传动及散热设计，控制运行热量及损耗。例如，具微科技采用高槽满率扁线绕组，提高关节输出能力与能量转换效率，关节模组效率达到 90%，峰值扭矩达到 600 牛·米[71]；关节热管理与抗冲击性能也在整机运动中开展验证，荣耀“闪电”在 2026 年半程马拉松中采用峰值扭矩达 400 牛·米的一体化关节模组，配置液冷管路，冷却液循环流量约为每分钟 4 升。通过主动散热缓解高功率运行中的热量积累，改善关节持续输出的温度条件。

[71] 具微科技. 具微科技亮相 WAIC 2026：全栈自研新品首发，“本体-数据-大脑”生态闭环成型[EB/OL]. https://www.micbotics.cn/list_3/126.html.

灵巧手多种传动方案并行发展，技术仍未收敛。目前，灵巧手主要存在关节直驱、连杆传动和腱绳传动等三种技术路线，三类方案在关节控制、抓握输出和手部轻量化方面各有侧重。**一是直驱方案**便于独立控制各手指关节并获取位置反馈。临界点 OmniHand 3 Ultra-M 配置 20 个主动自由度，重复定位精度达到 ± 0.2 毫米；韩国 TESOLLO 的 DG-5F 手指各配置 4 个独立驱动关节，同样形成 20 个主动自由度，支持不同抓握姿态与多指协调运动。但直驱方案中，电机集中在手部，会增加重量、占用空间并限制散热。**二是连杆方案**通过刚性杆件带动手指，兼顾抓握力和动作控制。灵心巧手 Linker Hand L20 结合连杆与滚珠丝杠，配置 16 个主动自由度和 5 个被动自由度，五指最大抓握力为 100 牛；大寰机器人 DH-5-6 采用空心杯电机、行星减速器、滚珠丝杠与连杆相结合的传动结构，五指抓握力为 80 牛。但连杆方案增加关节数量和活动范围后，存在杆件相互干涉和部分关节难以独立控制的问题。**三是腱绳方案**则将电机移至前臂，减少指节内的驱动机构，为感知传感器留出空间。1X 新一代 NEO 灵巧手采用低减速比腱绳传动，手指与手掌配置 22 个主动自由度，并在指尖和手指表面集成触觉感知；临界点 OmniHand 3 Ultra-T 采用前臂后置驱动，配置 22 个手部主动自由度，手部重量为 500 克，配置较高自由度的同时减轻手部质量。此外，**耐久性已成为灵巧手最受关注的性能指标。**后续路线选择将更加取决于具体任务对精度、负载和耐久性的综合要求。CRAFT 灵巧手结合腱绳传动、刚性指节与柔性关节，在与采用相同型号电机的 LEAP 对照手分别持续抓持 5 磅哑铃一小时的试验

中，平均电流约降低 50%，改善了持续抓持的驱动负担。临界点 OmniHand 3 Ultra-T 支持在 10 分钟内更换单根腱绳，将易损件更换纳入产品设计。

触觉传感器感知灵敏度和耐久性得到改善，电子皮肤支持全手掌感知和接触反馈。触觉传感器为抓握力度调整提供反馈；电子皮肤将触觉感知单元布置在手指、手臂和躯干等部位，使机器人能够感知更大范围的表面接触。**触觉传感方面，**Yun 等人[72]提出柔性三维力传感阵列，在夹爪试验中实现未知物体的滑移检测与抓握调整，验证了接触信息参与动作调节的可行性，力方向测量偏差小于 2°。麻省理工学院提出耐久型视触觉指尖，改善接触表面的抗磨损性能。加速砂磨试验中，新传感器的保护膜约在 2—3 小时后破裂，对照样品约在 24—30 秒出现表面膜破裂。**电子皮肤方面，**接触感知与手部力控制相结合，为滑移判断和抓握调整提供反馈。2026 年 7 月，NEO 新一代灵巧手中，电子皮肤覆盖指尖及手指表面，可获取接触位置、压力和切向受力信息。

端侧计算平台，加强模型计算与运动控制集成，算控一体化配置逐步完善。端侧计算平台加强计算、控制与传感接口的集成，支持推理、控制在同一平台运行，减少不同板卡之间的连接与适配。2026 年 6 月，地瓜机器人推出的 RDK S600 开发套件，提供最高 560 TOPS 的 INT8 稀疏算力，集成 CPU、人工智能计算单元和 6 核实时微控制器，

[72] YUN G, CHEN Z, CHEN Z, 等. Multiscale-structured miniaturized 3D force sensors[J]. Nature Materials, 2026.

分别承担模型计算与实时控制任务，并通过控制域接口连接传感器和执行机构。高通 Dragonwing IQ10 机器人将计算、传感接入和运动控制纳入统一配置，支持最多 12 路 GMSL2 摄像头，并提供 EtherCAT、CAN-FD 等通信接口，满足多传感信息处理与运动指令传输需求。同时，计算模组也增加了不同算力和功耗的产品选择。英伟达于 2026 年 1 月推出 Jetson T4000 模组，提供最高 1200 TFLOPS 的 FP4 稀疏算力，配置 64GB 内存，功耗范围为 40—70 瓦，算控平台可在不同供电和散热条件下部署模型。

电池性能与供电方案双线并进，改善连续作业的供能条件。长时间移动与操作要求电池在有限重量下提供足够电量，同时缩短充电等待时间，保持电池更换过程中的供电。**电池性能方面，**2026 年 7 月，均胜电子发布机器人固液混合电池包，电池包比能量达到 380 瓦时/千克，标称循环寿命超过 2000 次，支持 30 分钟充至 80%电量，将储能能力、反复充放电寿命与快速补能需求结合。**供电保持方面，**智元远征 A3 采用双电池冗余供电方案，电池系统总容量为 1152 瓦时，支持热插拔换电，可在 10 余秒完成换电。该配置可在更换电池时保持系统供电，减少断电重启造成的额外等待，并通过更换电池缩短补能时间。

2.整机结构与配置持续完善，拓展操作空间与环境适应能力

关键零部件的成熟已经大幅降低整机制造门槛。造出基本性能相近的具身智能本体正变得越来越容易，但造出能适应真实场景、能够

长期稳定运行的产品仍然很难。**一方面，**随着一体化关节、灵巧手、计算控制平台等关键部件的功能集成度不断提升，叠加仿真环境、运动控制软件和开发工具链逐步成熟，整机样机开发和基础运动能力实现的门槛持续下降。采用相近规格零部件的本体，在自由度、负载、活动范围和常规动作等基础指标上也更容易处于相近水平。**另一方面，**零部件层面的规格相近只是整机性能相近的必要条件，不代表产品综合能力相同。真实场景中的整机表现，还取决于面向任务的构型设计、传感器与执行器配置、结构强度、能量与热管理、运动控制、安全冗余、系统集成及可靠性设计。

2026 年，**具身智能整机制造从零部件组装转向产品定义，**场景适配、系统集成和长期可靠运行能力成为行业关注重点。整机设计在满足承载需求的同时，更加注重作业范围、运动稳定性与持续运行能力的统筹考虑。其中，双足人形机器人加强躯干与四肢的结构配合，完善感知配置与末端扩展；轮臂机器人结合躯干升降与重载双臂，兼顾物料取放与空间适应；四足及轮足机器人完善带载通行的防护与环境适应能力，扩大任务设备的使用范围。自主补能与模块化维护也成为部分产品的整机配置内容，为减少运行中断和便利现场维修提供支持。

一是双足人形机器人优化腰胯部关节布局扩大操作空间。腰髋布局优化使躯干和下肢能够更充分地参与操作位置与姿态调节，配合双臂完成低位取放和不同方向的末端对位。整机同时通过传感器布置和末端接口配置，完善环境信息获取与操作部件接入条件。北京人形机器人创新中心推出的天工 3.0 采用三自由度串联腰部，并调整髋部关

节布局，支持完全俯身下蹲，扩展垂直方向的可达范围与地面操作空间。其头部搭载激光雷达，可选配六路环视摄像头，同步获取点云与多视角图像；腕部接口可接入灵巧手、六维力传感器等部件，将全身活动、环境感知与末端扩展纳入同一平台设计。开普勒 K3 大黄蜂在延续 K2 双臂 30 千克负载配置的基础上，新增腰部自由度，支持躯干姿态微调、多角度工件对位和转向操作，在保持承载能力的同时，为双臂作业提供更多姿态调整方式。

二是轮臂机器人优化躯干配置，兼顾重载和空间适应。轮臂产品在移动底盘与双臂组合的基础上，根据物料重量配置机械臂，通过躯干升降扩大垂直作业范围，并优化局部结构以适应受限工位，体现重载搬运、高低位取放和紧凑作业等性能取向。银河通用 Galbot S1 采用四轮转向全向底盘，躯干升降行程为 750 毫米，双臂最大负载为 50 千克，操作范围覆盖地面至 2.3 米高货架，将较大承载能力与高低位取放配置相结合。紧凑型产品则通过缩小局部运动空间、采用折叠结构等方式，适应受限工位和整机转运需求。无界动力第二代 K15 采用紧凑的腕部设计，腕部翻转半径为 30 毫米，可在全向工作空间宽度不超过 800 毫米的环境中作业。

三是四足及轮足机器人持续拓展使用场景，增强载荷和防护能力。四足及轮足机器人同时考虑有效负载、移动续航、越障和环境防护，并通过搭载任务工具扩展功能。璇玑动力 Hypertron-T01 开展移动平台与消防工具配合的展示验证，将重载底盘与专用水炮结合，水炮重量为 17.5 千克，有效射程不低于 60 米，最大流量不低于 21 升/秒。

宇树 As2-W 将主动轮与四足结构结合，整机重量约 25 千克，可携带 16 千克负载的运行时间超 2 小时，连续台阶攀爬高度指标为 30 厘米。户外环境适应性成为重要性能指标，云深处山猫 M20S 的整机重量约为 36 千克，防护等级提高至 IP67，低温工作下限扩展至－30℃，工作温度范围达到－30℃至 55℃。

（四）算力：云边端协同演进，系统支撑能力持续完善

随着具身智能加快进入工业、商业服务和家庭等真实场景，算力需求正从单一的模型训练与推理，扩展为覆盖“数据采集、模型训练、端侧部署、推理和控制”的系统性需求，且任一环节的卡点都会制约整体系统性能。为了保障具身智能系统可以在真实物理世界中稳定、实时、经济地运行，云边端协同的计算基础设施发挥重要作用。从现有实践看，主要包括三条技术路线。

一是统一异构本体资源调度和服务，让每个本体都成为可被识别、可被调度的数据和服务节点。面向不同厂商、不同形态和不同地域的机器人设备，GPU 算力、边缘计算节点和仿真环境等资源统一纳入资源池，通过标准化接口实现资源状态感知、能力描述和统一管理。例如无问芯穹与清华大学发布 RLark，把云端 GPU 集群、边缘算力节点和机器人真机设备统一接入到一个平台中，支持训练、数据采集、评测和真机执行等多种具身任务统一编排运行。普林斯顿大学与微软研究团队于 2026 年 5 月发布 Kairos，提出面向物理 AI 的多机器人共享推理服务系统，将机器人“动作生成—物理执行”闭环纳入资源调度，根据机器人执行进度调度共享推理资源，提升多机器人云边推理效率。

在峰值负载下，其平均端到端任务时延降低 31.8%—66.5%；100 台机器人并发时降低最高 42.8%。

二是端云分割推理与通信－计算协同设计，支撑不同模型模块按照时间合理分工。具身智能系统具有明显的“多层异构”特征，语义推理、动作生成与底层控制可以运行在不同时间尺度上。美国 Figure AI 公司发布的 Helix 02 采用“三系统”分层设计，其中负责场景理解和语言推理的 System 2 以 7—9 Hz 运行，负责反应式动作生成的 System 1 以 200 Hz 运行；负责全身平衡、接触和协调的 System 0 以 1 kHz 运行。北京大学于 2026 年 4 月提出的 RoboECC 框架[73]进一步把模型结构、云边硬件性能和网络状态联合纳入 VLA 分割决策，将 OpenVLA 在 Jetson Orin 上的运行速度提升 3.16 倍，真实机器人控制延迟从 1200 毫秒以上降至约 360 毫秒。2026 年 9 月，柏林工业大学提出的 ComVLA 将通信因素纳入模型推理本身，考虑推理与通信联合设计，用语言指令引导视觉 Token 剪枝，使 Token 预算随信道容量自适应，云侧算力下降 74%而成功率仅损失 1.5 个百分点。

三是端侧高能效推理与算控一体，支撑资源有限本体上的实时持续运行。端侧部署是具身智能在本体侧实现可靠和安全运行的关键环节，需在功耗、散热受限的嵌入式平台上支撑连续感知与高频控制。无问芯穹面向 VLA 模型开发的 APXInf 推理引擎通过流水线、计算图、算子内核和量化的多层协同优化，将 $\pi_{0.5}$ 在 Jetson Thor 上的端到端时延由 278 毫秒降低至 26 毫秒；同时采用 C++与 Rust 协同的底层

[73] ZHENG Z, CAO H, CHEN J, 等. RoboECC: Multi-Factor-Aware Edge-Cloud Collaborative Deployment for VLA Models. arXiv:2603.20711, 2026.

架构，将内存安全与并发安全前置到编译阶段，保障长时间运行的可靠性。NVIDIA Jetson AI Lab 则通过 ONNX 转换、TensorRT 编译及 FP8/NVFP4 量化，将 $\pi_{0.5}$ 在 Jetson AGX Thor 上的端到端时延由约 132 毫秒降至 49 毫秒，提速约 2.7 倍。

三、应用拓展：从展示表演起步向作业模式转化

新技术的应用往往从功能相对简单、风险相对较低的任务起步，在技术能力、产品成熟度和用户接受度不断提升后，逐步进入更加复杂的生产生活环节。生成式人工智能的发展已经体现出这一过程，早期应用主要集中于对话问答、内容生成和图像生成，随后逐步进入办公、软件开发、数据分析等生产性任务，并进一步发展为能够调用工具、处理文件和执行多步骤任务的智能体。具身智能也呈现出类似的发展过程，但其作用对象由数字信息转向物理世界，应用需要同时满足感知、决策、操作、安全和可靠运行等要求，实际应用的成熟过程更加复杂。

2026 年上半年全球人形机器人出货量超过 2.2 万台，同比增长近 300%，其中文娱商演和科教数采合计占比超过 60%，导览导购约占 19%，智能制造和仓储物流分别约占 13%和 5%[74]。**具身智能应用呈现出由展示体验向实际使用、由单点验证向规模应用逐步深化的发展特征，**文娱表演和竞技赛事持续扩大社会关注，消费级具身智能产品通过陪伴互动、教育娱乐、户外跟随等低门槛任务，率先进入个人和家

[74] Counterpoint Research. Global Humanoid Robot Shipments Soar Nearly 300% YoY in H1 2026, Driven by Commercial Deployments[EB/OL]. https://counterpointresearch.com/en/insights/global-humanoid-robot-shipments-soar-nearly-300-percent-yoy-in-h1-2026.

庭市场，越来越多行业通过真实场景试点寻找具身智能垂类应用任务，在各类相对结构化生产场景中验证技术可行性，并逐步向开放服务场景拓展。但总体来看，现阶段具身智能仍处于单工位或小批量试点阶段，试点范围集中于低复杂度任务和结构化程度较高的封闭场景，系统在开放动态环境下的任务成功率、长时作业稳定性和复杂场景泛化能力尚不能满足规模化生产要求。具身智能要从技术演示转变为生产要素，成长为支撑社会生产生活的新型智能劳动力，仍需经历一段技术爬坡与产业沉淀的过程。

（一）文娱赛事：面向大众的能力展示场与技术试炼场

文娱表演和竞技赛事以爆发式增长的社会参与度打开社会认知与市场预期，以系统性的真实场景检验倒逼技术快速迭代。在社会参与度上，2025 年世界机器人大会现场参会人次 27.1 万[75]，2026 年达 55.7 万[76]，同比增长约 106%。2026 年北京亦庄人形机器人半程马拉松参赛队伍由 2025 年的 20 支增加至 102 支。2026 年第二届世界人形机器人运动会吸引了 16 个国家的 666 支队伍、2056 台机器人同台竞技，参赛队伍数量较首届增长 138%，机器人数量翻了两番[77]。据央视网数据统计，跨媒体总触达受众 24.03 亿人次。**在技术试炼上，**2026 年北京亦庄人形机器人半程马拉松约四成队伍采用自主导航方式参

[75] 北京经济技术开发区. 2025 世界机器人大会圆满落幕，融资 14.81 亿元，售出机器人及相关产品 1.9 万台[EB/OL]. (2025-08-12). https://kfqgw.beijing.gov.cn/zwgkkfq/ztzl/lqztkfq/lqzx/zsdt/202508/t20250818_4175529.html.

[76] 北京经济技术开发区. 30 家国际机构支持 3000 余件展品参展 55.7 万人次现场参会 世界机器人大会见证机器人加速迈向千行百业[EB/OL]. (2026-08-26). https://kfqgw.beijing.gov.cn/ywdt/gzdt/202608/t20260826_4837862.html.

[77] 新华网. 第二届世界人形机器人运动会参赛机器人数量较上届翻两番[N/OL]. (2026-08-14). https://news.cctv.com/2026/08/14/ARTInrPLHAmBubal1nGoNVQe260814.shtml.

赛，人形机器人冠军成绩从一年前的 2 小时 40 分 42 秒缩短至 50 分 26 秒，不到去年时长的三分之一。2026 年第二届人形机器人运动会所有田径赛项均为全自主参赛，同时新增了网球、乒乓球等高动态高对抗项目，展现了算法控制频率和运控精度的进步。其中 100 米完赛时间由首届的 21.50 秒缩短至 9.32 秒；太极拳“腾空飞脚”等高难度动作已成为多家参赛本体的标配，反映出关节峰值扭矩密度、高动态响应能力以及持续高功率输出下散热性能的整体提升。

（二）个人消费：玩具和陪伴产品率先实现商业变现

情绪消费需求叠加政策补贴，具身智能以玩具和陪伴产品为切入点率先打开消费市场。一方面，部分消费级产品累计销量达到万台级，个人用户呈现真实付费意愿。例如，Baby Alpha 系列消费级机器狗累计销量超过 2.5 万台，其中家庭用户占比达到 90%[78]；2026 年 5 月维他动力消费级四足机器人实现量产交付，截至 8 月累计已完成 126 万次大模型交互、23700 公里用户伴随里程和 13.15 万小时用户陪伴[79]。海外消费级机器人也开始形成一定市场规模，2026 年以来卡西欧 AI 陪伴机器人 Moflin 全球累计销量已超过 2 万台；松下陪伴机器人 NICOBO 累计销量突破 1 万台。**另一方面，**政策支持进一步推动具身智能产品进入普通消费体系。2026 年商务部等 8 部门明确提出促进人工智能机器人消费，支持人形机器人以及四足、仿生机器人产品研发迭代和升级，并加快养老机器人、陪伴机器人和人工智能生活助手

[78] 南京市人民政府. 南京制造“含新量”持续攀升[EB/OL]. (2026-08-03). https://www.nanjing.gov.cn/zzb/ywdt/njxx/202608/t20260803_5887389.html.
[79] 新华网. Vbot ATOM 发布 维他动力人形机器人进入产品化时刻[N/OL]. (2026-08-22). https://www.xinhuanet.com/auto/20260822/552eab20bbfb45aab48880b1878f7938/c.html.

等产品应用示范[80]。北京市将陪伴机器人、机器狗、外骨骼机器人和养老照护机器人等具身智能产品正式纳入个人消费补贴范围，个人购买可享销售价格 15%、最高 1500 元的补贴[81]；江苏[82]、海南[83]等地也陆续出台相关补贴政策，消费级具身智能开始进入地方促消费政策体系。

（三）生产应用：碎片化场景中的试点验证与应用拓展

具身智能加快进入部署作业模式，通过碎片化场景验证逐步提高场景供给质量。2026 年，具身智能面向工业制造、物流仓储、能源电力、医疗康养和商业服务等应用场景进行多点探索，在结构化环境、作业流程和任务变量相对明确的条件下，不断验证和筛选真正适合具身智能承担的作业环节，并关注任务成功率、作业节拍、连续运行时间、安全性、维护成本和投资回报等真实生产指标。其中，工业制造和物流仓储任务边界较为清晰、标准化程度较高，部分应用开始向连续作业推进；能源电力等专业场景应用价值明确，但对系统可靠性和安全性要求更高；商业服务和医疗康养等开放环境中的任务变化更大，目前仍以适用性验证和应用场景探索为主。

在工业制造领域，汽车、3C 电子等流程相对标准化的生产环境

[80] 商务部等 8 部门. 商务部等 8 部门关于加快“人工智能+消费”发展的实施意见[EB/OL]. (2026-07-21). https://scjss.mofcom.gov.cn/zlgh/zcfb/art/2026/art_e24c3760c5e3453199d2701febe7abbc.html.

[81] 北京市商务局. 北京市商务局关于新增 2026 年消费品以旧换新补贴产品的公告[EB/OL]. (2026-07-23). https://sw.beijing.gov.cn/phone/zwxx/2024zcwjphone/202607/t20260723_4780182.html.

[82] 江苏省发展和改革委员会, 江苏省财政厅, 江苏省商务厅. 关于优化调整江苏省 2026 年消费品以旧换新地方自主品类补贴政策的通告[EB/OL]. (2026-08-28). https://doc.jiangsu.gov.cn/art/2026/8/28/art_78712_11821822.html.

[83] 海南省商务厅. 海南省商务厅关于新增海南省 2026 年消费品以旧换新地方自主补贴品类的公告[EB/OL]. (2026-07-01). https://dofcom.hainan.gov.cn/dofcom/0400/202607/614b64bfc257477b96b8e7773b8a96b7.shtml.

成为具身智能机器人优先验证的场景。BMW 分别在德国莱比锡工厂和美国斯帕坦堡工厂开展人形机器人生产应用测试，验证其在零部件搬运、生产物流等任务中的系统适配和连续作业能力。**在物流仓储领域，**应用测试逐步由规则物体抓取扩展至随机来件识别、分拣和连续作业。2026 年 5 月，Figure AI 人形机器人首次直播包裹分拣任务，3 台 F.03 机器人通过自主轮换充电实现连续作业超过 200 小时，累计处理约 24.96 万件包裹；8 月自变量机器人在真实物流包裹分拣任务中直播一小时，总共分拣包裹 1816 件。**在能源电力领域，**具身智能开始向巡检、设备操作和高危作业等专业场景拓展。2026 年，南方电网基于电力具身大模型研发的机器人已在深圳、广州多座变电站试点运行，可自主完成开关柜巡检、红外测温等作业，输电线路补销机器人也开始探索自主上线、越障和销钉补位等操作任务；国家电网研发的线路除冰机器人、变电运检具身智能人形机器人、地下管廊六足巡检机器人已在四川等地进行试点应用。灵境智源具身智脑支撑人形机器人自主完成跨越挡鼠板、上下楼梯等电力巡检作业动作。**在医疗康养领域，**具身智能机器人应用逐步向医院、社区和养老机构等综合服务环境延伸。北京安贞医院引入多类医疗机器人开展临床试用，覆盖门诊、药房、病房等场景，重点探索机器人在院内配送、导诊、药品处理和辅助操作等环节的适用性；北京荣华街道智慧康养机器人养老驿站集中部署多类养老服务机器人，围绕餐饮、康复、助行和陪伴等任务开展实际服务验证；青岛等地建设康养领域具身机器人训练验证平台，对机器人在康复辅助、行动支持等任务中的实际适用性开展测

试。**在商业服务领域，**具身智能开始进入顾客直接参与、环境动态变化较大的开放场所。京东 MALL 等零售场景将机器人应用于迎宾、导购、理货和门店服务等工作，重点验证其在开放商业环境中的自主服务和人机交互能力；乐聚机器人在江苏银行苏州分行上岗，全程参与迎宾接待、需求识别、业务分流、业务引导、陪同办理等完整服务流程。

四、具身智能产业生态与商业模式

（一）产业链分工进一步明确，系统集成与服务环节作用提升

2026 年，产业链上下游分工进一步明确，如图 3 所示，形成上游数据、算力、基础软硬件和核心零部件、中游具身智能产品与技术服务、下游行业应用的发展格局。根据中国人工智能产业发展联盟（AIIA）具身智能产业图谱初步统计显示（截至 2026 年 10 月 8 日），国内相关企业达到 405 家。其中上游 180 家、中游 257 家、下游 212 家，191 家企业同时覆盖两个及以上产业链环节，占比达到 47.2%，具身智能企业普遍通过软硬件协同、整机与零部件协同以及产品与场景协同加快技术和产品迭代。

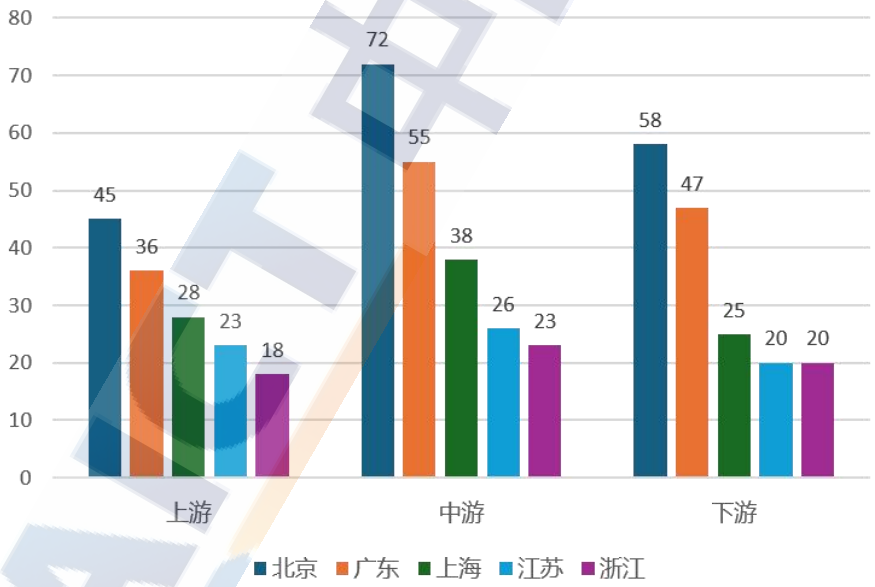
从地域分布看，如图 4 所示，具身智能产业继续保持较高的区域集中度。北京、广东、上海分别收录企业 109 家、84 家和 58 家，合计占比约 62%；江苏和浙江分别为 45 家和 39 家，前五省市合计占比超过八成。北京广泛布局算法模型、整机产品和数据采集设备等领域，广东、上海在整机研发与基础配套方面均有较多企业参与，江苏、浙

江则在机器人产品、配套供应及行业应用等方面形成聚集趋势，产业地域分布“头部集中、长尾分布”态势依旧显著。



来源：中国信息通信研究院

图 3 具身智能产业链结构示意图



来源：中国信息通信研究院

图 4 具身智能产业地域分布趋势

从产业分工看，**上游层面**本体配套硬件仍是企业布局的重点，数据、算力和通信网络共同构成基础支撑。**本体零部件企业**（133 家）

占上游企业数量约 74%，覆盖电机、传动机构、控制器、传感器、材料和灵巧手六个方向，其中控制器、传感器的参与企业较多。**数据相关企业**（37 家）覆盖数采设备、训练场、数据管理平台、仿真平台和开发平台等多个方向，数据获取、管理、开发与训练环境建设相结合，近九成训练场企业同时布局数采设备。

中游层面，机器人产品作为主体，模型研发与产品开发联系较为紧密。**产品方面**，机器人企业覆盖多种构型，人形、轮臂式机器人是企业布局最为集中的方向，四足及多足机器人、仿生机器人和智能机械臂并行发展；智能运载装备覆盖无人机、无人船和自动驾驶汽车，智能可穿戴设备、变形移动装置和集群式微型智能机器人等新型产品开始进入产品体系。**技术服务方面**，算法模型企业中，约半数同时布局机器人产品，模型与整机协同布局。系统集成中的**二次开发**等方向也已形成企业布局。具身智能从整机产品出厂到实际落地应用的工程链路较长。产品在进入具体场景后，仍需面向作业要求完成硬件配置调整、模型适配和现场调试。这些差异化需求将催生一批二次开发服务商，依托已有本体和模型完成应用交付，进一步细化整机企业、模型企业与二次开发服务商之间的专业化分工。

下游层面，工业应用场景布局最为集中，多行业应用格局已经形成。工业应用数量最多，商业、家庭和应急也是主要参与领域，交通、能源、教育、医疗、农业和航天等方向均已有企业布局。超过六成的下游企业同时覆盖中游产品或技术服务，产品研发与行业应用相结合是较为普遍的布局特点。

（二）多元主体加快实质性布局，产业协作关系进一步深化

2026 年，信息通信、消费电子、汽车、先进制造和行业应用等领域企业加快进入具身智能。跨界主体的布局方式由投资合作和技术验证进一步转向平台建设、整机研发、产线建设和场景共建，产业协作开始向研发、验证、交付和运营等环节延伸。

一是具身智能企业继续发挥产品创新和系统集成作用。智元机器人、宇树科技等企业围绕具身模型、本体结构、运动控制和操作能力开展研发，是人形、轮臂和四足等机器人产品创新的主要力量。从发展路径看，一类企业加强模型、本体和核心部件的全栈布局，通过软硬件协同提高产品迭代效率；另一类企业聚焦具身模型、灵巧手、力触觉传感器或开发平台等关键环节，通过专业化产品进入整机配套体系。随着产品进入真实场景，原生企业的竞争重点逐渐由样机能力展示转向稳定运行、批量交付和场景复制。

二是数字技术和消费电子企业分别从平台与智能终端切入。信息通信企业围绕数据、模型、仿真、网络连接和端云协同等环节建设具身智能平台。华为云 CloudRobo 已覆盖数据管理、模型开发、仿真验证、部署运行和持续优化等环节，腾讯 Tairos 则通过开放具身模型和开发工具支持不同本体接入。消费电子企业依托端侧计算、操作系统、多模态交互和智能终端开发经验向机器人拓展，荣耀于 2026 年推出首款人形机器人，小米、vivo 等企业也持续推进相关布局[84]。此类企

[84] 科创板日报（财联社）. 荣耀进场 又一手机厂商杀入机器人赛道[N/OL]. (2026-02-24).

业有助于拓展具身智能的消费级产品形态，但在运动控制、机电系统和整机可靠性等方面仍需积累。

三是汽车和先进制造企业强化工程化与规模制造能力。汽车企业在感知、计算、控制和复杂系统集成方面具有一定技术基础，同时具备供应链管理、质量控制和规模制造经验。小鹏已启动人形机器人全链条量产基地建设，推动汽车领域积累的芯片、控制和制造能力向机器人迁移[85]。美的则依托电机、控制系统、工业机器人和智能制造体系，协同布局核心部件、机器人整机和工厂应用，“美罗”系列已进入多个生产基地开展验证[86]。汽车和制造企业的进入，正在增强具身智能由样机研发向工程制造和内部场景应用转化的能力。

四是行业应用方和央国企由产品采购者逐步转变为场景联合定义者。能源、汽车、电子制造、物流和商业服务等企业掌握真实岗位、生产流程和运行数据，开始围绕负载、节拍、连续运行、安全和维护等要求参与机器人产品定义。央国企则依托工业、能源、交通和应急等场景，承担应用组织、规模采购和示范推广作用。应用方的深度参与，使产业协作由提供试验场地进一步延伸至联合研发、数据反馈和持续运营。

2026 年具身智能产业生态中，跨界企业开始通过平台、产品、产线和场景等方式实质性进入，原生企业、跨界企业和行业用户之间形成更加具体的能力分工，跨本体接口、数据流通、责任划分和收益分

https://www.cls.cn/detail/2293497.

[85] 小鹏汽车. 机器人量产加速！小鹏布局行业首个人形机器人全链条量产基地[EB/OL]. (2026-02-25). https://www.xiaopeng.com/news/company_news/5537.html.

[86] 新华网. “十骏”腾跃向未来——广东人形机器人产业加速崛起[N/OL]. (2026-08-20). https://www.news.cn/fortune/20260820/3f36dd8dd44641f2b17753bb07395966/c.html.

配机制尚不成熟，产业协作的标准化和可复制性仍需提升。

（三）产业链多环节创造收入，可持续的商业模式仍有待沉淀

1.产业链各环节商业模式加速分化

商业模式是检验具身智能产业化能力的试金石。2025 年下半年以来，随着整机产品供给日益丰富、场景试点密集铺开，具身智能的变现路径沿产业链各环节加速分化，数据、模型、人才培训等“卖铲子”环节率先兑现收入，整机出货量快速扩张，租赁、订阅等新模式开始萌芽。总体上看，多数环节商业模式仍处于**增长已成立、盈利未建立**的探索阶段，成熟闭环尚未形成。

数据环节是具身智能产业链中最先闭环的环节，按订单组织的数据采集服务与数据集交易已形成真实收入。从供需双方来看，数据服务的买方主要为本体厂商与模型公司，卖方多为具身智能训练场、拓展数据采集业务的整机提供商，或专业数据服务商。数据类型上，真实数据与仿真合成数据交易并存，计价形式各有区别。**真实数据**交易按采集条数或有效时长出售，根据成本按数据定制化程度高低区分价格，真机数据价格相对更高。鹿明机器人数据超市[87]出售 UMI 真实场景采集数据，以 100 条为一个产品，时长从 20 秒到 60 秒，单个产品售价百余元。遥操作真机数据价格在数百到上千元不等。**仿真合成数据**以不同的计费逻辑拓展数据环节的收入边界，例如光轮智能以数据帧数为核心计费方式，服务英伟达、谷歌、Figure AI、比亚迪等头部

[87] 鹿明机器人（Lumos Robotics）. Data Market（数据清单）[EB/OL]. https://data-market.lumosbot.tech/.

团队，据其披露 2025 年全年营收实现 10 倍增长[88]。数据服务的收入虽已兑现，但数据集能否跨本体通用、买方是否愿意持续复购，仍有待市场进一步验证。

模型环节的商业化整体处于早期，其价值主要通过**整机溢价、订阅服务和生态绑定**间接变现。**一是**整机企业将自研模型作为整机竞争力的核心而不单独出售，例如优必选搭建千亿参数基座大模型 Thinker 及世界模型，模型能力计入整机定价，支撑其在工业制造场景的交付。**二是**模型企业作为供应商向本体厂商输出“大脑”能力，提供订阅或按项目收费。例如 Skild AI 将其基础模型 Skild Brain 通过云端 API 授权给机器人厂商，已有越南 VinDynamics 付费将该模型集成到其工业人形机器人。**三是**整机或模型企业进行生态开源，促进自家模型与其他本体企业生态绑定。2026 年 1 月，蚂蚁集团灵波科技开源 LingBot-VLA 及后训练代码，已与星海图、松灵、乐聚等主流本体厂商完成适配，以开源验证跨本体迁移能力、构建技术生态。

本体环节当前以整机销售与租赁双轨并行。整机销售模式由整机企业主导，线下直营店、网店与经销商渠道并存，收入确认路径清晰。**人形**机器人整机定价多在数十万，单笔金额大，盈利基础相对薄弱，**同质化竞争风险正在积累。**据 IDC 统计，2025 年全球人形机器人出货约 1.8 万台，但 85%聚集在文娱表演、科研教育、导览导购等场景，或进入训练场采集数据，真正落地干活的商业试点仍在探索。**非人形**具身智能具备现阶段最突出的整机销售商机。一方面，四足机器人覆

[88] 新华财经（中国金融信息网）. WAIC 2026丨专访光轮智能杨海波：深耕具身仿真数据，为机器人搭建“学习系统”[N/OL]. (2026-07-23). https://www.cnfin.com/gs-lb/detail/20260723/4444946_1.html.

盖行业级和消费级形态，云深处山猫 M 系列四足机器人毛利率高达 57.5%[89]。另一方面，轻量化消费级非人形具身智能瞄准个人玩家或开发者，实现大规模销售盈利。Pollen Robotics 发布 Microduck 桌面双足机器人[90]，售价 399 美元，配置预训练动作，也可通过开源代码进行二次开发，上线 6 小时订单额突破 100 万美元。**整机租赁模式将一次性资产开支转化为持续性服务收入，降低应用门槛，形成机器人即服务**（RaaS）。**一是**整机企业推出机器人订阅服务。1X 公司 NEO 提供每月 499 美元机器人进家方案[91]。自变量机器人与 58 同城合作推出智能保洁服务[92]，深圳地区 3 小时定价 149 元。**二是**以机器人租赁为主业的专业 RaaS 第三方平台也进入市场。美国 Robonow Global 公司对外出租宇树 G1 机器人，每台日租费用 3000 美元[93]；日本的 EmplifAI 公司以每日 75000 日元的价格对外出租宇树 G1 机器人[94]。随着整机集成成本下探，整机租赁费用也随之走低，该模式面临早期购入价格高、后期成本难覆盖的风险。

场景解决方案正在落地初期，整机企业和专业解决方案提供商均有参与。一是整机企业创造单点落地场景，跑通后按门店或产线复制，例如银河通用机器人在北京十余家合作药店承担盘点、补货、取送、打包全流程，单店日处理订单 200 至 300 单。这类模式短期内扩张，

[89] 云深处科技股份有限公司. 关于云深处科技股份有限公司首次公开发行股票并在科创板上市申请文件问询函的回复[EB/OL]. (2026-08-07). https://static.sse.com.cn/stock/disclosure/announcement/c/202608/002190_20260807_MP3Z.pdf.
[90] Pollen Robotics. Microduck: A tiny biped robot you can teach new tricks[EB/OL]. https://pollen-robotics.com/microduck/.
[91] 1X. NEO Home Robot[EB/OL]. (2025-10-28). https://www.1x.tech/discover/neo-home-robot.
[92] 深圳市南山区人民政府. 全球首个家庭保洁机器人正式上岗 人机搭档解锁家政新玩法[EB/OL]. (2026-03-19). https://www.szns.gov.cn/xxgk/qzfxxgkml/tpxw/content/post_12691603.html.
[93] RoboNow Global. Unitree G1 Rental in the U.S.[EB/OL]. https://www.robonowglobal.com/robots/g1.
[94] Emplif. Unitree G1 レンタルサービス[EB/OL]. https://emplif.ai/service/rent-unitree-g1.

通过自建回路进行快速能力验证，但长期来看仍需探索更具竞争力的成本与服务模式。**二是**整机企业与场景方签单，面向企业产线二次开发、深度绑定。例如 2026 年 3 月千寻智能与京东达成战略合作，千寻 Moz 机器人在京东 MALL 线下门店正式“上岗”，担任咖啡制作服务员，未来将探索药房分拣、配药等更多细分零售岗位试点。**三是**场景解决方案提供商拉通技术方与场景方，承接场景方需求，推动技术二次开发落地。例如 Extend Robotics[95]，把宇树整机硬件和自研软件集成与工程服务能力打包，出售给制造业等终端场景方。

面向机器人全周期的服务与人才培训等后市场初步成形。全周期服务方面，出现机器人 4S、6S、7S、9S 店等新形态。传统 4S 店复刻汽车销售模式，包括卖整机、供配件、维修保养、收集用户使用数据反馈给厂商等。6S、7S、9S 店则在 4S 店基础上叠加不同的交易形式与生态服务。例如 6S 店增加租赁与个性化定制，7S 店增加解决方案、展示与培训，9S 店进一步拓展至孵化、研学等生态职能。2026 年 2 月，机器人 9S 店长沙超能机器人大世界涵盖场景展示、产品销售、租赁服务、定制开发、售后维护、新品发布、成果孵化、科普研学和人才培训等九大功能。**人才培训认证方面，**机器人企业与高校共建产业学院、共研课程与技能等级证书，例如 2026 年 1 月开普勒真实工业机器人平台进入高校[96]，提供“真实平台、真实任务、真实数据”

[95] Extend Robotics. Extend Robotics named official integrators for Unitree Robotics, helping accelerate the deployment of humanoid robots into real world scenarios[EB/OL]. (2026-01-21). https://www.extendrobotics.com/post/extend-robotics-named-official-integrators-for-unitree-robotics-helping-accelerate-the-deployment-o.

[96] 中国日报网. 以具身智能赋能教育未来——开普勒机器人助力高校产学研融合创新[N/OL]. (2026-01-04). https://cn.chinadaily.com.cn/a/202601/04/WS695a16f5a310942cc499a004.html.

的教学方案。

2.商业闭环仍面临场景落地多重卡点

具身智能领域商业模式成立与否最终取决于场景落地闭环能否打通，其价值兑现都以机器人在真实场景中稳定作业、客户愿意持续付费为前提。当前具身智能产业在生产端已取得显著进展，整机产品供给日益丰富、资本市场热度持续高涨，但从“能造出来”到“能落地、能交付、能复购”之间仍存在多重卡点，场景试点周期、本体性能适配、供需双方协调、订单转化与验证保障等环节的商业闭环尚未真正打通，多数落地项目仍停留在试点示范与样机验证阶段，距离规模化、可复制的商业交付仍有明显距离。

一是场景试点周期长，二次开发成本高，可复制性弱。**一方面，**机器人本体从开箱部署到真正“能用、能干活”，需经历数据采集、模型训练、现场调试与反复验证等环节，周期以周甚至以月计，且高度依赖工程师经验。例如优必选 Walker S 系列 2024 年 7 月首次进入汽车工厂实习，2025 年 3 月才开展首例多台、多场景、多任务协同实训，真正实现产业级量产能力还需要一到两年时间。**另一方面，**即便在细分场景中完成验证，跨品类复制的经济性仍难以成立，**单点试点成果向多场景推广的路径尚未走通。**例如杰克科技旗下艾图机器人在单一品类、单一面料条件下验证了裁片分离成功率 97%，但面对未训练过的面料或工件仍无法自适应调整；Figure 公司机器人在宝马工厂完成为期 11 个月的部署，到 2026 年 5 月才宣布以新一代机器人推进新用途项目，从试点到场景扩展开发周期以年计。

二是本体方与场景方面临需求与共创双重协调困难。**一方面，**本体性能与场景需求不匹配，可靠性尚未跨过连续作业门槛。客户期望在规定时间内可量化、可验收、次次成功，而行业现状仍处于“某些条件下能完成，扰动增多时成功率波动”。以工业场景为例，现场的金属反光、粉尘干扰等因素经常导致感知系统失灵。**另一方面，**场景方数据开放不足，制约本体性能。当前合作共创以场景方提供场地、具身智能公司提供解决方案的模式，或以两方成立合资公司的模式铺开，均以场景方在场地、人员、数据乃至股权上的实质性投入为前提。以工业场景为例，本体企业需要进入真实产线采集数据、迭代模型，而场景方掌握工艺参数等核心资产，对数据外泄与产线停改风险存有顾虑，双方缺乏稳定的利益分配与信任机制[97]。

三是能力度量缺标尺、责任界定缺依据、风险分担机制刚起步。**能力度量方面，**由于缺乏统一标准，接口、通信协议、安全标准各自摸索，数据无法互通、每家企业重复采集，规模化商用卡在“每个场景单独训练”的瓶颈，形成重复造轮子的局面。**责任与风险方面，**人形机器人一旦发生人身伤害或财产损害，企业可能同时面临民事赔偿、产品召回、行政罚款等多重责任，责任链条复杂、法律依据尚不完备。中国太保产险推出全国首单遵循 6S 全周期标准的具身机器人综合保险[98]。但由于缺乏历史数据与统一风险评估标准，风险识别与精准定价仍是难点。

[97] 新华网. 科创“智”高点丨从“道具”到“生产力”：2026 人形机器人“确定性”大考[N/OL]. (2026-06-11). https://app.xinhuanet.com/news/article.html?articleId=202606112acbc76c56554258af2eef3bf3db4bd5.
[98] 新华网. 从“工具赋能”到“AI 原生” 中国太保以智能体重构保险价值链[N/OL]. (2026-09-01). https://app.xinhuanet.com/news/article.html?articleId=20260901248055b460694b099bb8340c96a3fc5c.

（四）产业开源生态建设加快推进，开放协同创新机制初步形成

开源开放已从少数企业和开发者的个体探索成为具身智能产业生态建设的主线之一。行业围绕**数据、模型、本体**三线并进，以开源开放换取生态位。一方面，面向企业和开发者积累采用量，为自家数据、模型或本体预置生态基础。另一方面，借助社区力量摊薄试错成本，借助社区开放者的力量共同开展技术路线验证。

数据方面，开源数据集加速扩增，触觉模态关注度增加。据不完全统计，截至 2026 年 9 月，全球开源具身智能数据集超 100 个，2026 年开源数据集总量及含触觉模态的数据集数量，均达到 2025 年全年开源数量的 2 倍。戴盟机器人联合 Google DeepMind、中国移动、新加坡国立大学、香港科技大学、上海交通大学、日本东北大学等海内外数十家学术机构与企业在魔搭社区开源 Daimon-Infinity 含触觉具身智能数据集，截至 2026 年 9 月下载量累计超 800 万次[99]。

模型方面，国内外具身智能大模型保持开源趋势，并探索与多型号本体技术适配，拓展自身合作生态，同时降低中小企业与行业用户入局门槛。Physical Intelligence 开源“OpenPi”仓库[100]，汇集该公司代表产品 π_0、π_0-FAST、$\pi_{0.5}$ 等，支持多个常见机器人平台（Franka 臂、ALOHA 双臂、Libero 仿真平台等），提供开箱即用示例。蚂蚁集团灵波科技开源 LingBot-VLA 及后训练代码，与星海图、松灵、乐

[99] Daimon Robotics. Daimon-Infinity[EB/OL]. https://modelscope.cn/datasets/daimonrobotics/Daimon-Infinity.
[100] Physical Intelligence. openpi: Open-Source Models and Packages for Robotics[EB/OL]. https://github.com/Physical-Intelligence/openpi.

聚等主流本体厂商完成适配，以开源验证跨本体迁移能力、构建技术生态。

本体方面，整机与零部件的硬件设计、软件开发工具初步形成开源生态，降低企业与科研人员研发成本。**整机**上，宇树通过官方 GitHub 开放 SDK、强化学习训练框架与全系列 URDF/USD 模型文件，并推出完全开源、可 3D 打印的双足机器人 Qmini 与轻量机械臂 UniArmL1[101]。萝博派对全栈开源人形机器人 ROBOTO_ORIGIN 机械结构、系统软件栈、部署框架、运动和技能训练[102]，Github 星标数量超 2000。**零部件**上，加州大学洛杉矶分校 2026 年 7 月开源直驱灵巧手 MIDAS Hand[103]，具备 16 个自由度和 283 个三轴触觉单元，物料成本低于 3000 美元。香港大学与深圳河套学院开源电机远置的腱绳传动灵巧手 MM-Hand 的全部硬件设计与软件框架[104]。

（五）评测标准实践不断深入，体系化建设与应用衔接进一步深化

1.具身智能评测标定技术水位与应用边界，促进技术研发与应用落地的有机衔接

具身智能正处于从技术研发迈向规模化落地的关键阶段，不同技术路线与产品方案不断涌现，行业亟需一套公认的具身智能评测体系，客观衡量技术进步和真实应用成效，搭建起连接技术研发与应用落地

[101] unitreerobotics. GitHub 主页（Unitree Robotics）[EB/OL]. https://github.com/unitreerobotics.
[102] Roboparty. roboto_origin: Fully Open-Source DIY Humanoid Robot[EB/OL]. https://github.com/Roboparty/roboto_origin.
[103] midas-hand-org. MIDAS Hand[EB/OL]. https://midas-hand.com/.
[104] HKU MMLab. MM-Hand 1.0[EB/OL]. https://mmlab.hk/research/MM-Hand.

的桥梁。在此背景下，行业愈发关注具身智能评测研究和体系建设。编写组基于 2026 年公开发表的约 800 篇具身智能模型论文及其使用的公开基准情况进行统计。根据支持的评测任务类型，现有公开基准可分为具身认知类、具身导航类、具身操作类和具身安全类四类。各个基准使用频次统计词云如图 5 所示。



来源：中国信息通信研究院

图 5 具身智能评测基准概览

具身认知类基准是多模态大模型评测向物理世界的拓展，以静态问答式评测为主，重点考查面向具身任务的认知和推理能力。这类基准无需搭建仿真或真实“考场”，直接以“视频—问答对”作为考题输入模型，检验系统空间感知和推理、物理常识理解、长程任务记忆与规划、具身交互与可供性推理等能力。2026 年活跃度较高的具身认知基准包括 OpenEQA（开放式具身问答）、VSI-Bench（视觉空间智

能）、ERQA（具身推理问答）、RefSpatial-Bench（空间指代与放置点预测）等。**从公开基准评测结果看，一方面，具身认知从“想”到“做”存在断层，认知能力无法直接转化为动作输出。**以 Gemini 3 Pro 为例，其在 ERQA 具身推理基准上得分达 70.5，但在要求直接输出操作动作点位的 RoboSpatial-Pointing 基准上，预测准确率仅 27.0%; GPT-5.4 与 Claude Sonnet 4.6 更低，分别仅 20.8%和 5.8%，说明通用多模态大模型能看懂场景，却无法将视觉理解转化为可执行的动作接口。**另一方面，具身认知短板主要表现在连续记忆和多步推理能力。**在考察情景记忆与主动探索的 OpenEQA 基准上，即使在直接提供连续 24 帧上下文（记忆信息完整呈现）的条件下，GPT-5.4 的得分也仅为 74.1；需要自主组织与回溯记忆时，高德 ABot-AgentOS 为 59.9，两者均明显低于 86.8 的人类基线。多步推理则随步数增加显著衰减: RefSpatial-Bench 结果显示，RoboRefer 的点位预测成功率由 1 步任务的 66.67%降至 3 步任务的 36.36%。

具身导航类基准以动态仿真评测为主，重点考察系统的探索与移动能力。评测时需先搭建仿真“考场”，再布置导航“考题”，检验系统能否在仿真环境中完成导航任务。2026 年具身导航研究使用最多的考场是 Habitat-Matterport3D—一套基于真实住宅三维扫描构建的室内场景; 使用最多的考题包括 ObjectNav（按物体类别寻找目标）、R2R（按语言指令沿路线行进），以及将指令跟随迁移至连续物理环境的 R2R-CE、RxR-CE（面向更长、更复杂的指令）等，主要覆盖目标物体导航、视觉语言导航、主动探索和交互导航等任务类型。**从现**

有评测结果看，一方面，具身导航模型初步显现通用性，具备多任务统一和跨本体迁移能力。高德于 2026 年 7 月推出通用导航基座 ABot-N1，以单一模型覆盖点位导航、目标导航、指令跟随、兴趣点导航、人物跟随五类任务，在 R2R-CE 和 RxR-CE 未见环境上的成功率分别达到 70.9%和 73.9%；在 EVT-Bench 基准上的单目标人物跟随成功率达到 90.1%。法国 Mistral AI 于 2026 年 7 月推出的 Robostral Navigate，通过预测视觉空间航点实现轮式、足式、无人机等异构本体部署；亮源新创于 2026 年 8 月推出的 LightNav-0，以统一中间表征描述跨本体执行轨迹，支持零样本部署至人形、四足、飞行、轮式等异构本体。**另一方面，具身导航模型尚不具备开放动态环境下的长程稳定运行能力。一是开放语义目标导航效率低。**Qwen-RobotNav 在 HM3D ObjectNav 目标导航基准上的成功率虽达 71.2%，但路径搜索效率仅 33.0%。**二是长程导航和真实动态环境适应能力不足。**同济大学研究发现[105]，代表性 VLN 模型从仿真迁移至真实环境时，成功率从 61%降至 22%，面对真实环境中的光线、视角变化时性能会严重缩水。2026 年 7 月发布的纯视觉长程导航基准 VoLN 显示，最优模型在未见环境高难度组的成功率仅 1.8%。**三是“最后一米”导航面临挑战。**高德于 2026 年 5 月推出的 POINav-Bench 重点考察 2 米内抵达入口的能力，实验发现，入口指代正确率达 98.8%，视觉定位质量达 94.8%，但最终导航成功率仅 42.33%，表明模型能够识别并定位

[105] WANG L, SHENG K, HE Z, 等. A Comprehensive Survey and Systematic Real-World Evaluation of Embodied Vision-and-Language Navigation[J]. IEEE Transactions on Automation Science and Engineering, 2026, 23: 14405-14434.

入口，却难以完成最后的抵达动作。

具身操作类基准采用仿真或真机评测方式，重点考察操作和协同作业能力。评测时需先搭建仿真或真实考场，再布置操作考题，检验具身智能系统能否完成上半身桌面级操作或全身移动操作任务。**仿真评测中，**LIBERO 和 RoboTwin 2.0 是目前使用最多的具身操作基准，聚焦单臂或双臂操作的原子技能或简单抓放任务。RoboDojo 进一步拓展仿真评测维度，包括泛化性、记忆、长程执行、精确操作、开放词汇指令跟随等，支持面向模型能力的诊断式分析。**真机评测中，**RoboChallenge 采取统一考题、远程盲测的“统考”模式，RoboDojo 采取多维度拆解、三人双盲打分的“诊断”模式，RoboArena 采取不预设考题、众包出题的“开放面试”模式，中国信通院 EAI Bench 真机测试采取行业出题、企业答题的“职业技能考试”模式，四个真机评测基准探索了各具特色的真机评测方式。**从现有评测结果看，具身操作能力仍局限于受控环境下的上半身操作，开放环境下泛化和全身移动操作能力明显不足。**南加州大学在 Psi-Zero 模型论文中自建八个真实的全身移动操作任务，论文结果显示 GR00T N1.6 和 $\pi_{0.5}$ 等主流开源模型的平均成功率均不到 30%。**一是受控桌面场景下的具身操作能力快速迭代，主流基准趋于饱和。**在 LIBERO 基准上，论文报告的平均成功率多在 97%至 99%区间；RoboTwin 2.0 引入双臂协同与场景随机化，平均成功率也在一年内升至 95.8%。**二是泛化能力不足，环境扰动和任务组合均使性能大幅下滑。**LIBERO-Plus 在标准任务基础上施加相机位姿、光照、初始位置等七维扰动后，原本成功率在 95%

以上的模型性能普遍跌至 30%以下，π_0 在相机视角和初始状态两类扰动下，分别跌至 13.8%和 6.0%，暴露出模型对训练分布的机械记忆依赖。在场景、目标物体更具多样性的 RoboCasa365 基准上，模型论文自报的最好成绩是 62.6%的平均任务成功率，在多技能组合和未见任务泛化设置下平均任务成功率多在 50%以下。**三是精细灵巧操作、长程多步操作和开放语义交互操作是现阶段模型能力的突出短板。**从 RoboDojo 仿真评测结果看，针对拧螺丝、插入充电器、存硬币等精细灵巧操作任务，模型最高平均成功率在 20%左右。开放语义交互操作任务上表现最好的模型成功率不足 5%。RoboChallenge 真机评测年度报告同样显示，即便是 Table30 榜首模型，精细操作类任务成功率也不足 15%。中国信通院于 2026 年 8 月发布的 EAI Bench 商业服务场景专项评测结果显示，模型能够基本胜任抓放等简单任务，但在涉及多目标连续操作的任务上，会因单步执行误差累积、任务状态判断失误、执行步骤错乱等多种因素叠加，导致模型性能显著下降。**四是仿真到真机迁移存在显著落差，真机操作人机差距大。**RoboDojo 双榜结果显示，仿真榜首与真机榜首并非同一模型，部分仿真领先模型在真机上优势缩水，亦有模型在真机上实现反超；截至 2026 年 9 月 5 日，真机榜最高成功率仅 24.4%，与人类专家 100%的成绩差距较大。

具身安全类基准以仿真评测为主，逐步从语言认知安全评测闭环执行和物理后果评估扩展。目前这类基准处于发散式探索阶段，第三方复现及模型发布时采用仍相对有限。评测方法具体可分为四类：**一**

是语义安全类评测，考察具身智能是否能够结合用户指令和视觉输入，判断和识别具有安全风险的任务，并采取拒绝、澄清或请求人工协助等适当行为。例如上海交通大学于 2024 年 12 月发布的 SafeAgentBench[106]基于 AI2-THOR 仿真环境构建了 750 个任务，包括 450 个危险任务和 300 个安全任务，判断模型是否会生成危险执行计划。北京航空航天大学于 2025 年 6 月推出 AGENTSAFE[107]从人员安全、环境安全和机器人自身安全三个方面构造 45 个对抗场景和 1350 个危险任务，在 AI2-THOR 仿真环境中评测模型是否会生成或执行危险动作。**二是行为策略和闭环执行类评测，**考察具身智能在连续执行任务的过程中是否始终满足安全约束。例如 2026 年 6 月清华大学推出的 LIBERO-Safety[108]基于 LIBERO 构建一系列安全场景，分析 VLA 模型在仿真环境中的行为策略安全性，核心指标包括安全成功率、碰撞率、执行时间、LDLJ 轨迹平滑度和危险任务拒绝率等。佐治亚理工学院于 2026 年 5 月推出 SafeManip[109]设置碰撞与接触、抓取稳定等 8 类安全属性，并在 50 个 RoboCasa365 仿真任务上对整个操作过程进行验证。**三是物理交互评测，**考察具身智能的动作执行是否会对物理环境造成影响。美国得克萨斯大学奥斯汀分校于 2026 年 7 月提出 OopsieVerse[110]借助 Omnivers、OmniGibson 和 RoboCasa 等仿真引

[106] YIN S, PANG X, DING Y, 等. SafeAgentBench: A Benchmark for Safe Task Planning of Embodied LLM Agents. arXiv:2412.13178, 2024.
[107] YING Z, WANG L, XIAO Y, 等. AGENTSAFE: Benchmarking the Safety of Embodied Agents on Hazardous Instructions. arXiv:2506.14697, 2025.
[108] CUI R, ZHANG Z, PANG J, 等. LIBERO-Safety: A Comprehensive Benchmark for Physical and Semantic Safety in Vision-Language-Action Models. arXiv:2606.23686, 2026.
[109] HUANG C, HUYNH K V, ZHANG M M, 等. SafeManip: A Property-Driven Benchmark for Temporal Safety Evaluation in Robotic Manipulation. arXiv:2605.12386, 2026.
[110] BALAJI A, BAHETY A, AMBATIPUDI S, 等. OopsieVerse: A Safety Benchmark with Damage-Aware Simulation for Robot Manipulation[C]//Proceedings of Robotics: Science and Systems (RSS), 2026.

擎测量碰撞力、温度、液体、外力扰动及损坏程度。**四是系统行为评测，**考察具身智能能否在高层规划器、低层控制器、安全监测工具和人工操作员之间作出正确的安全决策。2026 年 7 月，谷歌针对 Gemini Robotics 2 专门发布了配套的安全评估策略[111]，包括危险任务与安全约束遵循、人员接近监测、安全工具调用、VLA 可行性检查、指令歧义处理和视觉不确定性处理等六类系统行为。

2.具身智能评测仍是开放的测量科学问题，在有限资源下呈现具备产业价值的评测仍面临挑战

具身智能评测不同于传统人工智能的静态评测，测试结果不仅取决于模型性能，还受本体、物理环境、任务设置、人类参与和随机扰动等多重因素影响。同一模型部署在不同本体、场景和控制系统上，可能产生显著不同的测试结果。**如何从模型、本体和环境共同作用的系统表现中准确识别具身智能能力，并通过有限测试预测机器人在真实场景中的长期运行表现，仍是尚未解决的测量科学问题。**具体来看，主要面临以下三方面挑战。

一是评测真实性、可复现性与低成本投入难以兼顾。一方面，真机测试贵在真实，能够完整反映物理接触、硬件误差和环境扰动，但短板在复现与成本。真实环境难以精确复位，物体初始位置、摩擦系数、光照条件和人员活动会随测试发生变化，电池电量、关节温度、传感器标定和机械磨损也会持续改变机器人状态，即便采用相同模型和任务，在同一实验室多次测试或不同实验室复现实验，结果也有可

[111] Google DeepMind. Gemini Robotics 2: Safety Evaluations[R]. 2026.

能显著不同；若为追求可复现性而严格固定场景布置、物体位置和机器人状态，测试又会退化为实验室中的“标准动作”，丧失对真实复杂环境的代表性。同时，真机测试场地、设备和人力投入大，难以支撑低成本、大规模的重复实验。**另一方面，**仿真测试支持环境精确复位、变量可控制、实验可低成本大规模重复，可复现性优势明显，但其物理接触、柔性物体、传感器噪声和硬件时延等与现实世界仍存在差距。仿真评测结果在现阶段很难作为产业应用的可信依据。如何形成“仿真规模筛选—真机校准验证—真实场景试运行”的分层评测路径，仍需进一步探索。

二是测试结果受多类变量与随机因素影响，有限测试能否形成稳定结论仍缺乏科学依据。一方面，变量因素多且不可穷举。具身智能测试需考虑本体构型、传感器配置、控制频率、物体属性、初始位置、光照条件、障碍布局、指令表达、人员干扰和通信时延等多类变量。若对所有变量穷举组合，测试规模将呈指数级增长，实践中只能抽样覆盖，覆盖是否充分难以验证。**另一方面**，有限样本难以可靠区分性能接近的模型。现有真机 VLA 论文对每项任务的测试通常为 10 至 20 次，且普遍不报告置信区间与配对统计检验。Positronic Robotics 于 2026 年 5 月推出的 PhAIL 真机 VLA 基准中提出，使用平均任务成功率这个二元指标区分性能相差 5 个百分点的模型，需要 600 至 1500 次对比测试。评测轮次并非常数，待区分的两个模型性能差距越小，所需测试轮次越多。同时，在同一设备、同一测试时段内对模型进行盲测和随机交替，是保障横向可比性的必要手段。在既定测试预算下

无法区分两个模型，本身也是有效、应当公开的科学结论。

三是现有基准与产业实际需求仍存在差距。当前评测多采用边界清晰、时长较短的实验室任务，重点考察机器人能否完成单次操作；产业界则更关注机器人能否持续稳定作业，以及任务节拍、人工接管、故障恢复、能耗和运行成本等生产性指标。**一方面，**短任务成功率难以反映连续作业能力。英伟达 RoboLab-120 基准结果显示，主流通用机器人策略均无法成功完成超过 4 个复杂子任务的长程操作；从误差累积规律看，即便单步可靠性达到 99%，连续执行 100 步后的整体成功率也仅约 37%。**另一方面**，现有评测指标体系与产业关切之间仍有明显错位。平均任务成功率只反映机器人是否完成了任务，几乎说明不了机器人是如何完成任务的；PhAIL 真机 VLA 基准论文提出的机器相较人类作业效率指标（Human-Relative Throughput，HRT）虽将视角转向作业速度，但只衡量规定时间内完成单次操作所需的时间，未计入任务重置、故障恢复和人工干预耗时，因而不能直接换算为生产吞吐量。Alphabet 旗下工业机器人软件公司 Intrinsic 在 2026 年“AI for Industry Challenge”中，以线缆与连接器装配为任务，除考察插接成功外，还将装配精度、碰撞和过大作用力、整组任务周期时间纳入评分，并在最终阶段通过真实工业工作站验证方案，体现出产业评测对精度、安全和生产节拍的综合关注。

3.具身智能标准体系化建设步伐加速，高质量供给能力亟待提升

通信产业的发展经验表明，标准建设往往先于产业规模商用，商

用前数年正是标准布局的黄金期，5G 全球标准格局在 2019 年商用开启前即已基本确定，为后续产业高速发展奠定了规则基础。具身智能正处在由技术验证迈向规模化落地的关键节点，标准建设迎来同样的黄金窗口期，对破解产业同质化、创新碎片化至关重要。当前，**具身智能标准梯次推进格局初步形成，体系化供给水平大幅提升。**

联盟规范，先行先试发挥标准“试验田”作用。针对具身智能技术迭代快、路线未收敛的特点，联盟和团体标准以机制灵活、周期短的优势率先补位，承担了技术验证和共识孵化功能。中国人工智能产业发展联盟（AIIA）具身智能工作组于 2024 年联合行业单位共同发布《具身智能标准体系 1.0》和 AIIA/T 0198-2024《具身智能系统总体架构及技术要求》，首次在行业内提出了统一的具身智能能力框架，并成功推动该规范向国际标准转化。同时，AIIA 具身智能工作组也编制并发布了家庭陪伴机器人、数据集质量评估等联盟规范，并依托规范先行开展评估验证，对相关技术指标进行检验和优化，为行业标准、国家标准研制输送成熟条款与实施经验。2026 年，工作组进一步围绕遥操作数据采集、无本体数据采集、数据平台、人机交互安全等方向深化标准预研，联合产业各方就术语定义和基础技术框架快速沉淀共识。

行业标准，产业急需方向标准成果落地。工信部人工智能标准化技术委员会（MIIT/TC1）WG6 具身智能组于 2026 年先后发布 YD/T 6770-2026《人工智能 关键基础技术 具身智能基准测试方法》和 YD/T 6771-2026《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》两份行业标准，标志着行业标准在评测、数据集质量两大产业急需方向实现零的突破。同时，WG6 组围绕训练场综合评估、智能化分类分级、家庭陪伴等方向的标准研制也在加快推进，目前已进入征求意见阶段。工信部人形机器人与具身智能标准化技术委员会（MIIT/TC8）发布《人形机器人与具身智能标准体系（2026 版）》，专门针对人形机器人与具身智能的全产业链给出了标准化路线图。

国家标准，体系化标准供给明显提速。截至 2026 年 8 月，全国机器人标准化技术委员会（SAC/TC591）已下达人形机器人标准研制计划 52 项，覆盖核心零部件、数据集、整机技术要求、系统集成、测试验证和安全要求等方向。其中整机技术要求的 5 项国家标准和仿真测试平台 1 项标准已进入批准阶段，多项项目研制周期压缩至半年。

国际标准，全球具身智能标准取得积极进展。2025 年 12 月，我国牵头的《具身智能系统框架及能力要求》（ITU-T F.748.66，Requirements and Framework for embodied artificial intelligence systems）正式发布，首次在国际层面明确了具身智能定义，成为全球具身智能标准研制的统领性标准。2026 年 2 月，国际电信联盟电信标准化局（ITU-T）第 21 研究组（SG21）批准成立具身智能标准焦点组（FG-EAI）。焦点组是全球首个具身智能国际标准研究平台，由我国专家担任主席，下设的九个工作组布局与工信部《人形机器人与具身智能标准体系（2026 版）》深度衔接，实现国内国外标准体系的协同推进。目前焦点组已吸引五大洲 39 个国家、120 余家机构参与，预研国际

技术文稿覆盖术语、数据、以人为本、行业应用等重点方向。

高质量标准才能有效支撑产业高质量发展。**具身智能标准化工作仍处于起步培育期，在提升标准供给效能、优化标准供给质量方面仍面临三方面挑战。一是标准协同机制不健全，供给存在重数量、轻质量倾向**。多条标准线并行推进但缺乏统一立项统筹，数据集、一体化关节、仿真平台等热点方向出现国家标准和行业标准多头布局、内容交叉，标准动态复审更新机制尚未建立，存在“标准一发布即落后”的风险。**二是检验检测基础设施建设不完善，标准实施缺乏力度**。具身智能技术体系复杂，评测技术与工具研发尚处于探索初期，“标准—检测—认证—采信”闭环尚未打通。从设施建设看，具身智能测试实验室和具身智能测试实验室长三角分中心已经初步建成面向工业物流、零售商超、家庭服务、巡检作业等场景的真机智能化测试场，以及台阶、崎岖地形、交叉斜坡等多地形的综合运控测试场，但检测设备、测量方法等方面仍需在检验检测实践中不断完善。**三是标准“走出去”深度不足，国际认证贯通衔接不畅通**。具身智能领域全球尚无统一的测试评估与认证互认体系，企业不得不逐国逐项开展认证，成本高、周期长、不确定性大。同时，面向具身智能的专门认证标准缺失，出海认证只能“拼装”既有规则。以出口欧盟为例，实际依据的是机械指令、电磁兼容指令、无线电设备指令等横向法规与 EN ISO 10218-1（工业机器人）、EN ISO 13482（服务机器人）、EN ISO 13849-1（功能安全）等传统标准的组合，智能化水平、作业能力等核心维度缺乏认证依据。

五、未来展望

具身智能正处于从“技术闭环”走向“生态协同”的关键窗口期。需要认识到，具身智能的发展不会以一次技术突破突然完成，而更可能呈现螺旋式上升、渐进式演化的特征。未来的具身智能发展，在技术层面需要重新思考人工智能的底层架构，使模型、身体、环境和行动形成更加紧密的动态耦合。产业层面将从“技术驱动”转向“价值驱动”，能力边界内做深做实是产业可持续发展的关键。

（一）技术趋势：模型架构持续迭代，软硬一体共同定义智能

模型架构由全模态融合与多模型协同向认知—预测—行动一体化演进。一是作为当前具身智能模型主流架构的端到端 VLA，将由单一 VLA 向视觉、语言、空间、动作、触觉、力觉及运动控制等异构模型协同的复合架构演进。随着机器人数据采集能力提升，模型输入也将由视觉、语言等传统数字模态进一步拓展至深度、三维空间、触觉、力觉、本体状态和运动轨迹等多源物理信息，推动模型由“全模态感知融合”向“物理世界统一建模”升级。**二是**世界模型与 VLA、行为模型和控制模型进一步融合，模型将逐步打通感知、理解、预测、规划和执行等环节。例如，世界模型与 VLA 融合解决其存在物理幻觉、长视野规划不可验证、机器人数据稀缺三大瓶颈，为 VLA 智能体注入“物理大脑”。**三是**突破现有模块化模型架构，形成面向物理世界的动态基础模型新架构，实现模型、身体与环境的深度耦合，使人工智能从“理解信息”进一步迈向“理解世界、预测世界并作用于

世界”。形成“认知—预测—行动—反馈”的动态闭环。

形成“数据赋能模型、模型与数据共同演进”的具身智能发展范式。一方面，数据持续赋能模型能力提升。仿真合成数据规模化、以人为中心场景化和真机数据闭环加速，构成具身智能数据“新金字塔”。数据不仅只是模型学习知识的载体，而成为模型形成物理世界理解能力和行动能力的基础。价值也将由“规模”进一步转向“经验密度”，高质量、长时序、强交互和具有明确结果反馈的数据将成为提升具身智能模型能力的关键资源。**另一方面，数据与模型之间将形成持续迭代的闭环。**未来，能够支持数据回流模式的场景将会是具身智能模型率先应用落地的场景。机器人通过真实环境执行任务产生交互数据，数据经过筛选、加工和训练进入模型；模型能力提升后进入新的场景执行任务，又会产生新的数据和反馈，并进一步暴露模型在泛化、规划、安全和控制等方面的不足，从而驱动下一轮针对性数据采集和模型优化，形成“真实交互—数据回流—模型训练—能力提升—任务执行—反馈再采集”的循环。

软硬件由适配走向深度耦合，软硬一体化成为迈向通用智能的重要支撑。一方面，硬件驱动，成为智能发展的重要载体。具身智能的智能表现将伴随电机扭矩、减速器性能、传感器精度、材料特性、电池续航、通信时延以及控制频率等底层能力优化而提升，不同材料、形态和执行器将与模型共同构成特定的“身体—智能”组合，成为模型能力的一部分。**另一方面，软硬一体成为创新主战场，推动模型、本体与计算平台协同设计。**面对具身智能对实时性、功耗、精度和可

靠性的综合要求，单纯提升硬件性能或扩大模型规模均难以实现系统能力的持续跃迁，未来将加速推进芯片、执行器、传感器、本体结构、控制算法与具身模型的联合优化，在关节功率密度、机身轻量化、动态平衡、运动控制、边缘计算与散热设计等方面实现协同突破。同时，竞争评价也将由单项性能指标转向长期稳定运行能力，可靠性、可维护性和全生命周期成本将成为规模化交付的重要门槛，例如连续运行时间、故障率、维护周期和“三年无大修”等指标将逐步成为产业化竞争的关键指标。

（二）产业趋势：重点场景引领，基础设施、产业生态和就业体系同步重构

本体企业将加速分化，产业竞争由“拼数量”转向“拼真实能力”。本体数量的快速增长并不意味着产业已经成熟。**近期来看，**行业洗牌将明显加速。资本和产业资源将从“广撒网”逐渐转向“精聚焦”，真正能够解决泛化、可靠性、成本和数据闭环问题的企业将获得更多资源，而主要依赖概念展示、单一运动能力或短时表演的企业将面临较大压力。未来两三年，产业竞争的核心指标将逐渐从“机器人能不能动”转向“机器人能不能稳定完成真实工作”。**长期来看，**机器人本体可能出现明显的平台化和专业化分工。少数企业负责通用本体和核心零部件，大量企业围绕不同场景开发专业能力、工具和技能包，形成类似“智能手机硬件+操作系统+应用生态”的产业结构。

应用将遵循“二八原则”，从确定性任务向开放环境渐进拓展。具身智能并不需要等待一个“完美的通用人形机器人”出现之后再进

入产业。恰恰相反，未来相当长一段时期内，“边成熟、边落地”将成为更加现实的发展路径。**近期来看**，八成的产业资源仍将集中于确定性较强的场景。如巡检、安防、工业搬运、物流、仓储、柔性装配、质量检测等任务具有明确的目标、较稳定的环境和较清晰的价值回报，将成为具身智能优先突破的“滩头阵地”。**长期来看**，剩余二成的探索将面向开放环境和通用智能。例如家庭服务、养老陪护、复杂公共服务等场景将逐步成为下一阶段的目标。机器人最终需要从“完成一个工作”走向“理解人的需求并自主完成一系列工作”，从而真正进入千家万户。

具身智能加速催生新型基础设施、新型产业生态和新型职业。一是催生面向物理智能的新型基础设施，未来将加速建设面向具身智能的仿真训练环境、机器人训练场、数据采集与处理平台、模型训练平台、机器人测试认证平台以及云—边—端协同计算设施，形成覆盖“数据采集—仿真训练—模型开发—真机测试—规模部署—运行反馈”的全生命周期基础设施体系。**二是新型产业生态加速构建，**具身智能的产业价值链将从“整机制造”向“模型—数据—本体—场景”协同生态演进，随着机器人操作系统、标准接口和模型服务逐步成熟，第三方企业可以围绕特定行业开发技能模型、任务 Agent、数据服务和应用插件，形成类似“硬件平台+基础模型+开发工具+应用生态”的新型产业分工。**三是新型职业涌现，**具身智能进入产业应用后，一方面将替代部分高重复、高危险和高强度的劳动任务，另一方面也将产生大量围绕机器人训练、数据生产、系统运维和人机协作的新型岗位。

职业结构将由单纯的“人操作机器”逐步向“人训练机器、人管理机器、人机共同完成任务”转变。例如，机器人训练师负责通过遥操作和示范数据训练机器人完成特定任务，具身数据工程师负责采集、清洗、标注和治理多模态物理交互数据，机器人技能工程师负责开发和优化特定任务技能，机器人运维师负责长期运行、故障诊断和系统维护，具身智能安全测试工程师负责机器人在真实环境中的安全性、可靠性和风险评估。

（三）治理趋势：标准体系从“建框架”走向“落地实施”

标准体系将由“框架建设”转向“落地实施”。近期来看，标准化重点将从“提出框架”转向“可测试、可验证、可认证”。数据标准、模型接口、机器人接口、测试评价、功能安全、人机协作、安全风险和数据治理等标准将逐步与产业实践结合，建立从仿真测试、实验室测试到真实场景测试的分层评价体系。评价也不再只是“任务成功率”，而会覆盖能力、泛化、安全、可靠性、效率、长期运行和人机协作等维度。**长期来看，**具身智能将建立覆盖“数据—模型—本体—系统—场景”的全生命周期标准体系。标准将从描述技术逐渐转向支撑产业，即从“标准是什么”走向“按照标准能不能开发、测试、认证和交易”。

安全治理将从“模型安全”扩展到“物理行动安全”。近期来看，安全治理将重点关注数据隐私、模型可靠性、动作安全、人机协作、异常检测和故障恢复等问题。尤其是在工业、家庭、医疗等高风险环

境中，需要建立更加严格的测试、认证和责任追溯机制。**长期来看，**安全将从“事后检测”走向“设计内生”。安全约束将直接进入数据采集、模型训练、规划决策和动作执行过程，使机器人能够在行动前评估风险、行动中监测风险、行动后总结风险，形成具备自监测、自恢复能力的可信具身智能系统。

未来全球发展的着力点，将从技术创新逐步转向生态可持续与规则治理。近期来看，是商业验证和能力闭环的关键时期。各国和企业将围绕制造、物流、巡检、服务等场景加快部署，通过真实应用验证模型、数据和本体能力，并逐步建立产业标准和测试体系。**中期来看，**未来三到五年将进入规模扩展和生态构建阶段。多种机器人本体、具身基础模型、数据平台、仿真平台、开发工具和应用服务将形成更加完整的产业生态，跨本体数据和模型迁移能力成为重点关注方向。**长期来看，发展**重点可能进一步转向通用和全球化能力塑造。

编制说明

谨向为本报告提供指导与支持的具身智能测试实验室长三角中心、具身智能测试实验室（北京朝阳）、北京地平线信息技术有限公司、北京人形机器人创新中心有限公司、华为云计算技术有限公司、人形机器人（上海）有限公司、自变量机器人科技（深圳）有限公司、杭州具身智能中试基地科技有限公司等单位及专家，致以衷心的感谢。

CAICT 中国信通院

中国信息通信研究院 人工智能研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62308338

传真：010-62308338

网址：www.caict.ac.cn