

集智 蓝皮书 No.202515

具身智能发展报告

(2025 年)

中国信息通信研究院

2026年1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、摘编或利用其它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”。违反上述声明者，编者将追究其相关法律责任。



前 言

2025年《政府工作报告》首次将“具身智能”列入国家未来产业重点培育清单。具身智能作为与物理实体融合的人工智能，正在成为改变人类生产生活方式、推动社会智能跃升的重要引擎，有望催生颠覆性终端产品装备，促进生产力和生产关系的深层次变革。

过去一年，具身智能在运动场上实现运动能力突破，在训练场中学习作业技能，为真实场景应用打下基础。产业整体展现出“融合”“多元”“繁荣”的发展特点：**一是聚焦实现软硬、知行和虚实融合，突破智能瓶颈。**推进软硬融合创新，加速具身基础模型和本体结构的软硬协同进化。关注知行合一，打造感知决策行动一体化的闭环链路。打通虚实贯通路径，推动从仿真模拟到现实实践的迁移和部署。**二是打造“上天入海涉险”多元化产品，加速细分场景拓展。**在具身智能加持下，各类硬件载体从单一用途到多用功能，从移动范围突破到解决操作难题，从工厂、车间等结构化环境到航空、水下、极限环境等非结构化自然环境，呈现明显的智能化突破和场景拓展。以四足、多足、轮足等机器人载体以及无人车、无人机、无人船和可变形移动装置等为代表的移动类产品，解决“去哪里”问题；以协作机械臂、复合轮臂式机器人、人形机器人等操作类产品，解决“干什么”问题。**三是产业生态体系不断完善，市场空间广阔。**行业应用、产品服务、技术服务和基础设施四大板块逐步形成完整产业链。“马拉松”“格斗赛”“运动会”引发社会高度关注，据披露数据统计，截至2025

年 12 月，我国具身智能和机器人领域投资事件数达 744 起，融资总额 735.43 亿元人民币。其应用前景广阔，涉及“工具”“用具”“载具”“玩具”等多种类型，有望成为智慧员工、生活助手、智驾司机和智能伙伴。

具身智能并非单一技术的突破，而是一场由技术、工程、场景与资本合力推动的全球浪潮。作为一个长坡厚雪的赛道，在热潮背后也面临三大“非共识”：一是路径选择争议，“感知-认知-决策-执行”的实现，数据和模型孰重孰轻？二是本体构型争议，打造通用的唯一终极形态还是面向场景的专用构型？三是数据方案争议，真实数据与仿真/合成数据如何选择、处理和混合训练？具身智能整体处于发展早期，面临“数据-模型-本体-场景”难闭环，商业模式、规模化量产、标准缺失等产业化挑战。商用落地处于初期探索阶段，将从“科研试验”迈向“进厂入户”。

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）》后第 2 次发布具身智能蓝皮书，本蓝皮书聚焦过去一年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、全球具身智能总体发展情况.....	1
(一) 具身智能的概念与内涵逐步明确.....	1
(二) 具身智能成为全球关注的下一个科技前沿.....	3
(三) 具身智能产业发展阶段仍需客观冷静看待.....	6
二、数据驱动下的具身智能软硬融合创新.....	8
(一) 算法技术尚未收敛, 多路径探索快速迭代.....	10
(二) 数据成为核心需求, 相关技术加速发展.....	20
(三) 本体技术多元发展, 重在软硬一体的零部件创新.....	25
(四) 云-边-端协同运行, 支撑现实应用落地.....	27
三、场景驱动下的具身智能产品谱系不断丰富.....	30
(一) 机器人成为具身智能的最“热门”载体.....	31
(二) 智能运载装备是具身智能落地“最快”载体.....	38
(三) 新型智能产品是具身智能“前沿”探索方向.....	43
四、具身智能产业生态加速完善.....	45
(一) 产业链横跨多个领域, 生态体系初步形成.....	45
(二) 训练场建设成热点, 实际效用仍需进一步验证.....	49
(三) 标准体系有序建设, 正在针对性解决产业亟需问题.....	52
(四) 安全问题引发关注, 成为规模化落地限制要素.....	54
五、具身智能产业展望.....	57
(一) 技术架构重构: 从“功能模块堆叠”迈向“多模态认知融合”	58
(二) 应用场景深化: 从“演示”迈向“实用”	59
(三) 安全伦理构建: 从“合规”到前瞻布局“伦理协同框架”	59

图 目 录

图 1 具身智能三要素及闭环系统.....	2
图 2 具身智能概念内涵及特征示意图.....	3
图 3 具身智能技术框架.....	9
图 4 具身智能技术演进图.....	12
图 5 具身智能技术路径.....	14
图 6 具身智能产业图谱.....	46
图 7 具身智能主流企业区域分布图.....	47
图 8 四大板块整体情况图.....	48
图 9 具身智能标准体系.....	54

表 目 录

表 1 具身智能数据采集方案对比.....	21
-----------------------	----

一、全球具身智能总体发展情况

具身智能被誉为迈向通用人工智能（AI）的重要一步，引发了新一轮的技术浪潮。相较于在“计算机中思考”的传统人工智能，具身智能更强调“走进物理世界学习”。其发展建立在生成式人工智能突破和机器人等硬件技术的基础之上，受生物进化本源启发，推动了对“智能本质”的重新思考和软件与硬件技术的高度整合。近一年来，“世界模型”（World Models）驱动的预测推理和“通用具身基础模型”的多模态融合与认知决策能力取得明显突破，推动具身智能进入了全新的发展阶段。同时，随着本体结构实用性更强、零部件更加轻量化以及性能的不断突破，行业开始强调具身智能不仅会“感知和行动”，还能具备更强的认知能力，在与环境反复交互中不断进化，具备跨场景、跨任务的泛化能力，向通用智能迈进。

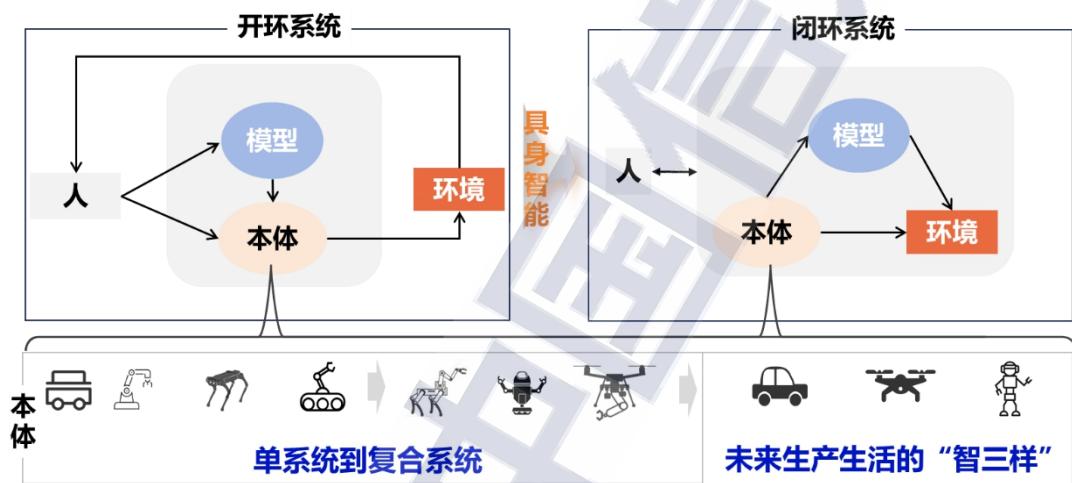
（一）具身智能的概念与内涵逐步明确

1. 具身智能核心内涵：基于身体交互的“智能闭环”

自《具身智能发展报告（2024年）》¹对具身智能的概念内涵和相关名词进行深入辨析后，业界对其认识逐渐深化：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。其核心在于具身智能以构建真正的机器人“智能闭环”为目标，区别于“开环系统”需要注入人的知识经验和操作方式，具身智能将构建的是一个可以由数据驱动形成闭环的智

¹ <https://www.caict.ac.cn/kxyj/qwfb/bps/202408/P020240830312499650772.pdf>

能系统，如图 1 所示，核心三要素包括：**一是**具身本体，包括协作机械臂、四足机器人、轮臂复合式机器人等，具身智能推动单系统到复合系统升级，其中人形机器人、无人驾驶汽车和无人机又被称为未来生活“智三样”。**二是**智能内核，依托大模型、世界模型与多模态技术，实现认知、决策与行动的统一。**三是**环境交互，以“第一人称视角”与现实物理世界进行动态交互和自适应学习。



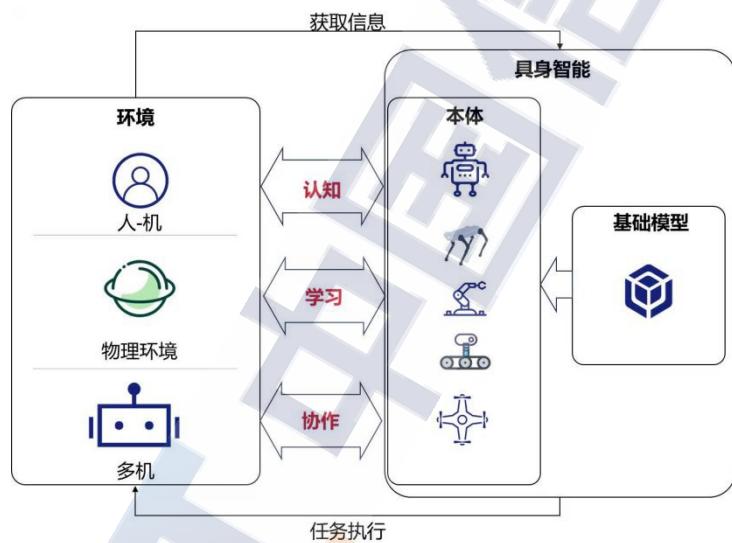
来源：中国信息通信研究院

图 1 具身智能三要素及闭环系统

2. 具身智能以增加认知、协作和学习三个能力为目标

国际三大标准化组织之一国际电信联盟（ITU）已经冻结的标准 ITU-T F.748.66 “Requirements and Framework for embodied artificial intelligence systems”（具身智能系统框架及能力要求）中给出具身智能（EAI, Embodied Artificial Intelligence）定义：指与物理实体融合的人工智能，能够自主与物理世界交互并适应环境。从智能能力上看，具身智能增加了认知、协作和学习三个主要功能，如图 2 所示。**一是**

认知能力，从“看见”到“看懂”。通过视觉、语言、动作等多模态数据的联合训练，不仅看到环境中有什么，还能理解内、外、远、近等空间关系。**二是协作能力**，从“单机执行”到“协同作业”，通过端-边-云协同，突破单机在计算负载、通信等方面的资源限制，强化人机、多机、多任务协作能力。**三是学习能力**，从“干前学”到“干中学”。通过学习现实场景中各种不可预测的情况，实现从“做得到”向“做得好”的转变。



来源：ITU-T F.748.66

图 2 具身智能概念内涵及特征示意图

（二）具身智能成为全球关注的下一个科技前沿

集“人工智能”与“机器人”等物理实体于一体的具身智能有望成为驱动未来经济增长的核心引擎。一方面，麦肯锡全球研究院 2024 年 10 月发布的研究报告《下一个重大竞争领域》²将人工智能和机器

²https://www.mckinsey.com/~/media/mckinsey/mckinsey%20global%20institute/our%20research/the%20next%20big%20arenas%20of%20competition/the-next-big-arenas-of-competition_final.pdf

人技术列为塑造未来全球生产力和数字基础设施的五大新兴前沿之一。根据美国市场研究公司 Grand View Research 预测，到 2030 年被誉为“智三样”的人形机器人约 40.4 亿美元³，自动驾驶汽车（L1-L3）约 1722 亿美元⁴，以及无人机约 1636 亿美元⁵，累计市场规模将达到 3398.4 亿美元。另一方面，全球科技巨头纷纷进军具身智能领域，多家具身智能公司估值突破百亿元。英伟达创始人黄仁勋在 2025 年消费电子展（CES）上强调，“人工智能的下一个前沿是具身智能”，并预测通用人形机器人的“ChatGPT 时刻”即将到来。特斯拉将研发具备自主性的人形机器人作为优先事项。OpenAI 也重启了四年前的机器人软件项目并布局消费硬件方向。

人形机器人成为全球具身智能领域的焦点。以中美两国企业为代表的的主要玩家运营路径可分为三类。**一是软硬件全栈派**，从 AI 大脑到硬件躯体全自主研发，通过技术闭环提升软硬协同能力，代表企业如 Figure AI、智元机器人等。例如，Figure AI 于 2025 年 2 月发布自研端到端模型 Helix，可实现人形机器人上半身的高速精确控制；智元于 2025 年 3 月推出通用具身基座大模型 Go-1、8 月发布世界模型 Genie Envisioner，硬件也发布了远征、灵犀、精灵等系列产品。**二是重硬件“身体派”**，核心优势在于机器人本体设计、运动控制算法、关节电机等动力学与控制技术，代表如宇树科技、Agility Robotics 和 Apptronik 等。例如，宇树科技产品在成本控制方面具有明显优势，

³ <https://www.grandviewresearch.com/industry-analysis/humanoid-robot-market-report>

⁴ <https://www.grandviewresearch.com/industry-analysis/semi-autonomous-vehicle-market>

⁵ <https://www.grandviewresearch.com/industry-analysis/drone-market-report>

于2025年7月发布最新款人形机器人R1，售价3.99万元起。Agility的双足机器人Digit专注于物流场景，已与亚马逊等巨头合作。Apptronik于2023年发布人形机器人产品Apollo，并与Google DeepMind合作进一步提升产品功能。**三是重软件“大脑派”**，以成为所有机器人的“操作系统”为目标，专注于开发通用机器人基础模型，代表公司如Physical Intelligence、Field AI和银河通用。例如，Physical Intelligence于2025年9月开源的π0.5(pi-zero-point-five)模型，开发者可基于标准化接口适配不同机器人硬件，降低具身智能研发门槛。Field AI以“Field Foundation Models(FFMs)”物理优先基础模型为核心技术，推动模型作为各类机器人的通用“大脑”，让机器人在不同环境中实现自主运行。

多种构型本体有望深度融入人类社会空间，改变智能与空间、机器与空间、机器与人的关系。多样化的多形态仿生构型与环境高度适应，例如除了人形、四足、蛇形、章鱼等形态，多样的自适应变形移动装置可实现结构与运动模式的动态重构，能够依据环境变化进行形态调整与能耗优化；集群式微型装置可形成可扩展、可重组的智能群体，在复杂环境中实现多模态协同与任务自适应重构。“陆、海、空、天”一体化，将推动智能体从单一功能执行向多空间、多主体协同的社会化存在形态。这将推动智能基础设施、生产系统与社会生活形态的深层变革。具身智能体将在工业、交通、医疗、教育、城市治理等领域形成人机共融、物理与数字融合的新生态格局，带动智能社会空间

结构的重塑与产业组织方式的再造。

（三）具身智能产业发展阶段仍需客观冷静看待

具身智能资本融资规模迅速扩大，商业化落地仍处于早期探索阶段。2025年以来以美为代表的多家公司突破单笔最高融资，2月Apptronik完成3.5亿美元A轮融资，9月Figure完成10亿美元融资，估值390亿美元。据不完全统计，截至2025年12月，我国具身智能和机器人领域投资事件数达744起，融资总额735.43亿元人民币。但需要认识到，具身智能从技术突破到应用落地仍需要一定发展周期。无论是模型实现路径、数据方案还是本体分类分级的应用落地途径，行业都未找到确定性答案，正在多路径、多场景探索。现阶段，以机器人为代表的具身智能产品走进训练场、行业场景中展开实训，但走向大规模应用前仍需进一步突破智能短板，跑通从定制化开发，整机规模化交付到二次开发运维的场景化落地闭环。

具身智能概念虽热，但具体实现涉及多技术融合，市场推广需多方努力探索。一方面，“智能”层面仍很难达到用户预期，要让模型学会现实任务，尤其是包括系列动作执行的长程任务，需要对每个动作环节采集大量数据用于模型训练。为更好满足行业现实任务作业需求，不仅需要从几十万小时到百万甚至千万小时的数据突破，还需要模型架构和训练范式的突破，以持续提升跨场景、跨任务，甚至跨本体的泛化能力，以及在特定场景任务上的作业成功率。另一方面，“具身”层面仍面临千台到标化量产鸿沟，在大规模落地时需要解决一系

列工程化技术难题。如需要有硬件集成与软硬适配能力，能应对不同行业场景下的硬件调试与技术适配需求，能支持数千小时的稳定作业等；需要有技术持续迭代能力，如稳定的研发团队、相关技术专利储备，可针对特定场景及时优化技术方案和对应产品功能。

具身智能是一种全新的技术范式，强调“数据驱动”的人工智能与“场景驱动”的机器人的深度融合。它并非简单地在传统机器人上增加人工智能算法，而是以物理身体为载体，身体形态取决于应用场景，数据来源于身体与环境交互，智能在交互反馈中得以进化，并更好适应场景。当前仍无法实现“数据—模型—本体—场景”的闭环。

一是数据层面，一方面，真机数据采集主要依赖于训练场，尚未实现规模效应。另一方面，合成数据获取主要依赖于仿真平台或世界模型，难以弥合虚拟到现实差距。数据规模缺失，策略未定，存在如何采和怎么用争议。**二是模型层面**，能力突破主要面临针对不熟悉任务和环境的可靠性不足，未见场景任务和异构本体的泛化性有限，且难以应对长程任务等挑战。同时，技术路径与训练范式仍存在分歧，以端到端视觉语言动作模型（VLA）为代表的主流架构是否存在物理智能涌现尚不确定。**三是本体层面**，本体能力很难与实际场景需求恰好匹配，构型设计上存在“先专用，场景能用”还是“要通用，全能作业”的争议。同时，本体在部署到实际场景时仍存在固有的物理局限性，面临较高的场景适配与部署维护成本，往往更高性能的硬件才能更好地展现模型能力，再加上研发与产线固定成本，整机系统价格仍处于高

位，大规模场景落地仍需时日。

二、数据驱动下的具身智能软硬融合创新

具身智能有别于传统 AI 线性的“输入数据 - 输出答案”模式，需要在物理环境的交互中形成自主进化的智能闭环：本体被部署到真实场景中 → 在执行任务中收集真实行为数据 → 高质量数据被用于模型训练和迭代 → 更通用的模型被部署到更多的机器人上，从而收集更多、更高质量的数据。其中，数据是闭环飞轮启动起点，具身智能在数据驱动下实现软硬融合创新，围绕“数据 - 模型 - 本体”三个关键技术要素展开多路径探索，云边端协同支撑系统研发和在现实环境中的有效运行。如图 3 所示。**一是数据**，作为模型通用化发展原料，行业围绕仿真合成数据和真实采集数据开展广泛技术实践，发力高质量、大规模具身智能数据建设。**二是模型**，作为智能核心仍处于早期阶段，路径呈现明显分散。小脑层面围绕技能学习和运动控制算法已有长期技术积累，大脑能力和大小脑协同是现阶段智能的核心瓶颈，也是主要创新方向。**三是本体**，作为承载智能的物理实体，多元本体形态融合发展，围绕系统集成的中间件、关键核心零部件、端侧计算单元方面取得明显突破。**四是云边端协同**，作为系统主要开发和运行模式，云侧开发平台和工具链不断完善，端侧计算能力加速升级，同时结合操作系统和中间件不断降低功能开发门槛，优化应用模式。



来源：中国信息通信研究院

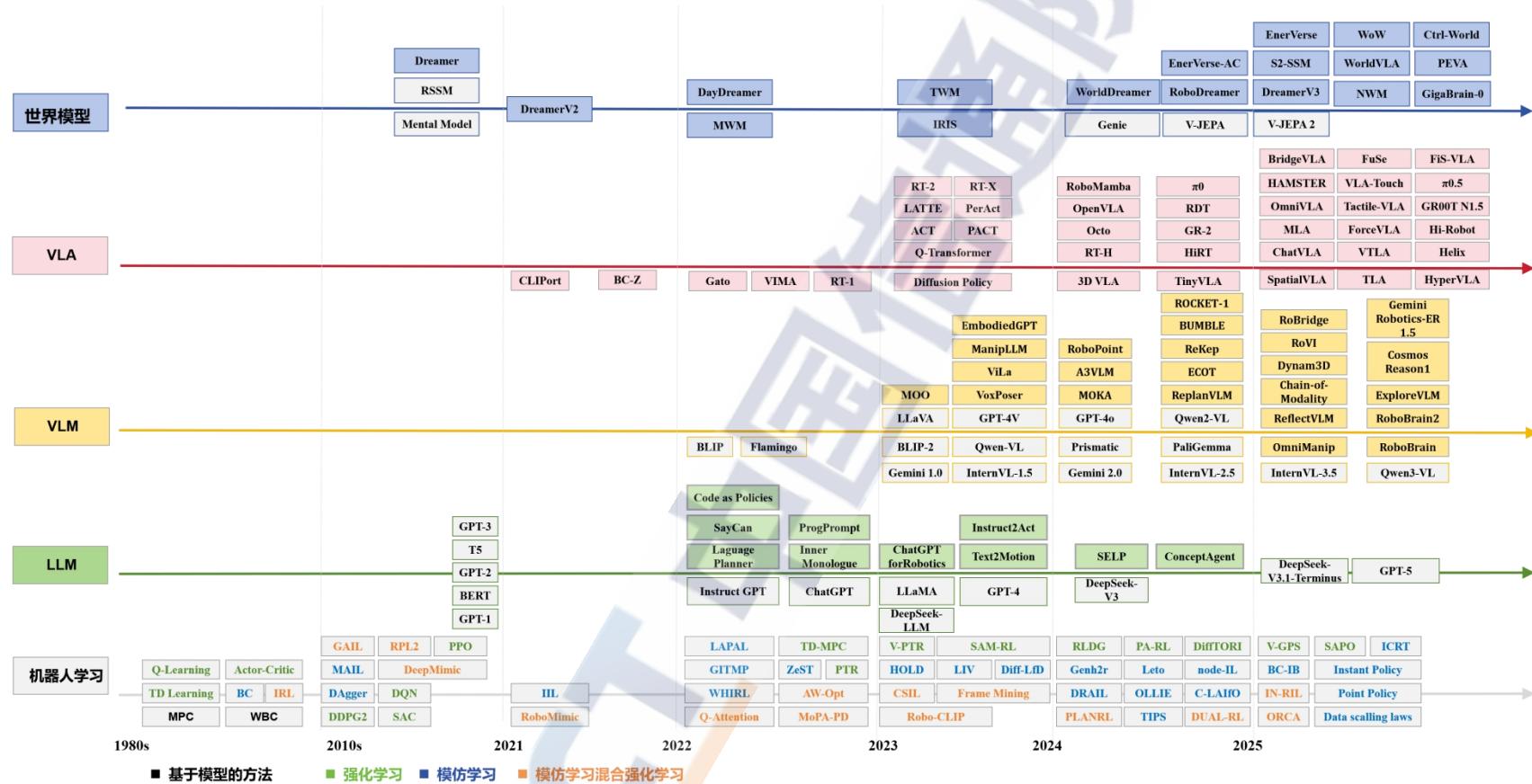
图 3 具身智能技术框架

（一）算法技术尚未收敛，多路径探索快速迭代

具身智能技术突破是生成式人工智能与机器人学习技术的融合和延续，核心是借助生成式人工智能的“认知”能力，驱动机器人构建“感知 - 认知 - 决策 - 执行”闭环，如图 4 所示。传统机器人运动控制技术主要包括基于模型的方法和基于学习的方法。基于模型方法通过建立运动学和动力学模型来进行运动轨迹规划和平衡控制，主流方法包括模型预测控制（Model Predictive Control, MPC）和全身控制（Whole-Body Control, WBC）。基于学习的方法以强化学习、模仿学习或两者结合，从数据经验中学习运动策略。2022 年以后，随着生成式人工智能技术突破，“大模型+”机器人成为重点创新方向。大语言模型（LLM）的发展打开了机器理解外界自然语言和常识知识的窗口，可以通过自动化编程（如 Code as Policy、ChatGPT for Robotics）替代原先大量的工程师编程工作，通过任务理解和规划（如 Language Planner、Inner Monologue）提升人机交互能力。视觉语言模型（VLM）在语言知识的基础上进一步融合了大量的视觉信息，提升对环境的语义理解能力。VLM 在语言知识的基础上进一步融合了大量的视觉信息，提升对环境的语义理解能力。视觉语言动作模型（VLA）将视觉、语言和动作统一到一个学习框架，根据视觉观测和自然语言指令输入端到端地输出动作。2023 年至 2024 年，VLA 通过利用 Transformer 架构统一视觉、语言和动作等多模态数据学习更通用的动作策略（如

Gato、RT-1、ACT、Q-Transformer），或结合 Diffusion 架构生成连续动作（如 Diffusion Policy）。同时借助 VLM 提升策略泛化性（例如 RT-2、OpenVLA、Octo、RDT-1B、 π_0 等）。这一阶段的 VLA 也被称为“大行为模型”，聚焦动作控制层面的可扩展生成。

2025 年，围绕“通用大脑”和“技能可扩展学习”两大方向开展密集创新。一方面，大脑从 LLM 和 VLM 的“大模型+”路线的感知、决策能力单点突破，向 VLA、世界模型等“感知-认知-决策-执行”闭环能力演进，技术创新呈现爆炸式增长。前 Meta 公司、英伟达公司研究员 Chris Paxton 认为，“VLA 作为新一代通用机器人的基础模型，是值得研究的基础架构，各个实验室在技术和方法上都在关注 VLA”。杨立昆认为“机器人是否足够聪明甚至具备通用用途，取决于能否在世界模型上有重大进展”。截至 2025 年 12 月，2025 年谷歌学术以“端到端 VLA”为关键词的论文在不到一年时间翻了近 4 倍，仅 2025 年已有 1700 余篇。另一方面，延续如图 4 的机器人学习路线，深入探索技能可扩展学习方法，支持本体在与环境交互过程中学习技能并持续精进。可扩展学习方法是具身智能持续学习的基础，也是其在真实世界中实现稳定可靠应用的关键所在。



来源：公开信息整理

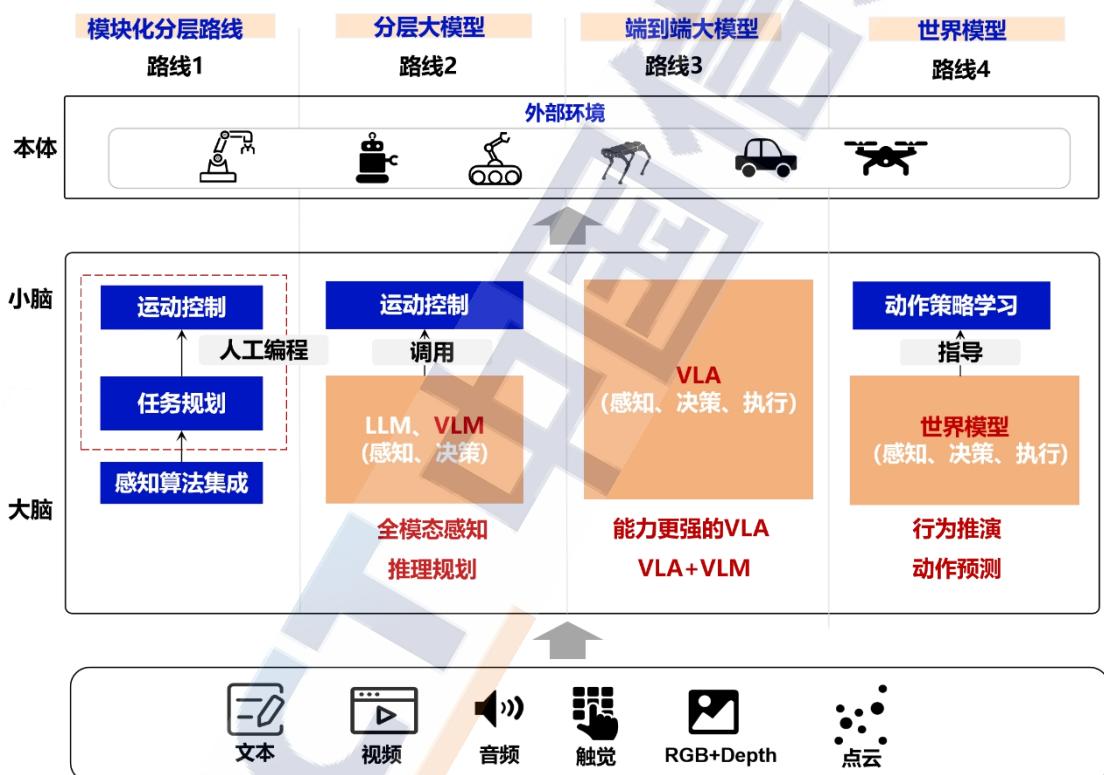
图 4 具身智能技术演进图

具身智能技术路径整体可分为四条，包括模块化分层、分层大模型、端到端大模型和世界模型，如图 5 所示。其中，模块化分层路线聚焦机器人运动技能学习突破，其他三条路线旨在发展具身基础模型，构建一个通用大脑，以适应现实世界中任务指令、动态环境、物体变化和动作干扰等复杂情境。模块化分层路线上，用传统深度学习算法实现感知，用机器人学习算法训练运动技能，并通过人工编程实现感知与运控的模块化组合或高难度组合式运控。例如自主移动机器人（AMR）的仓储搬运，机器人在运动会上的舞蹈表演、自由搏击等均采用分层路径，先在仿真环境中进行运动技能学习，再结合人工编程或遥控实现可靠运控。分层大模型路线上，用 LLM、VLM 作为“大脑”进行任务推理和规划，并通过 API 调用相应的动作策略。这一路径关注如何让机器人等物理实体更好地理解外部任务需求并进行规划和推理。例如谷歌 Gemini Robotics-ER 1.5、英伟达 Cosmos-Reason⁶能够通过规划逻辑推理将复杂任务进行多步骤拆分，并调用机器人的硬件 API 或专门的技能模型来完成任务。端到端大模型路线上，端到端 VLA 将视觉、语言和动作统一到一个学习框架，根据视觉观测和自然语言指令输入端到端地输出动作，通过大小脑的协同作用直接控制本体执行任务。例如 Physical Intelligence π 0.5⁷采用 VLM 结合 VLA 的双系统架构实现大脑和小脑间跨层级的知识推理和动作控制。

⁶ <https://arxiv.org/abs/2503.15558>

⁷ <https://www.physicalintelligence.company/blog/pi05>

自变量机器人 WALL-OSS⁸采用 MOE 混合架构在一个单系统链路中协同表达跨模态信息，实现大脑语义知识到小脑动作执行的知识迁移。世界模型路线上，通过动力学建模和未来观察进行行为预演和动作预测，以指导 VLA 动作策略学习或动作生成。例如清华大学联合斯坦福大学提出 Ctrl-World 可控世界模型，通过任务预演和虚拟策略打分提升在陌生任务上的成功率。



来源：中国信息通信研究院

图 5 具身智能技术路径

分层大模型路线以 VLM 为核心，促进规划推理与动作执行的有效衔接。2024 年，LLM 和 VLM 在不改变底层本体架构的情况下实

⁸ <https://x2robot.com/en/research/68bc2cde8497d7f238dde690>

现模型与机器人等物理设备的集成，将人工在环的编程控制转换为模型在环的规划控制，开启了具身智能新的交互和推理范式。这类模型能够作为高层“规划器”，通过解析用户任务指令、制定执行计划，并调用 API 接口或与 ROS 等中间件进行交互来控制本体。例如微软使用 ChatGPT 通过预定义函数库和提示工程，生成可调用不同机器人的 API 控制接口⁹。2025 年，VLM 模型从规划者转型为更全能的协调者，通过融合多模态、辅助信息或思维链等，提升感知和规划推理能力。**一是提升环境感知能力。**例如中山大学 RoBridge 结合对象分割信息来选择操作对象。浙江大学 ExploreVLM 结合空间关系图推理环境的 3D 空间结构。聆动通用联合科大讯飞推出 iFlyBot-VLM，通过构建 2D 或 3D 的定位数据、2D 指向数据和可供性数据等空间感知数据集，提升模型对物体可操作属性、末端执行器姿态约束及动态操作轨迹的建模能力。**二是提升规划推理能力。**上海人工智能实验室 OWMM-VLM 模型基于多模态输入进行链式思考，生成关键目标位置后可函数调用路径规划器、机械臂运动规划器等工具来执行具体任务。智源 VLM 模型 RoboBrain 2.0¹⁰结合 RoboOS 2.0¹¹分层框架，将复杂任务逐级拆解，并基于网络拓扑结构进行子任务动态分配，调度其他分布式小脑模型执行专项技能。**世界模型在大脑中构建了物理模**

⁹ <https://www.microsoft.com/en-us/research/publication/chatgpt-for-robotics-design-principles-and-model-abilities/>

¹⁰ <https://github.com/FlagOpen/RoboBrain2.0>

¹¹ <https://arxiv.org/abs/2505.03673>

拟世界，通过行为推演和预测优化行为策略。DayDreamer 通过预测未来状态实现更高效的技能学习。Meta V-JEPA 2 能从当前图像状态中捕捉有用信息，并预测执行一系列候选动作的后果，根据它们与预期目标的接近程度对候选动作进行评级。谷歌 Genie 3 可以生成视频来模拟机器人采取不同的推理策略、执行不同的动作以后得到的结果。瑞士苏黎世联邦理工学院 Robotic World Model 在大脑预演行为过程中持续优化动作策略，适用于操作和移动等不同任务。

端到端大模型 VLA 呈现“大一统”和“分层协作”两类设计架构。**一是**架构统一且精简的单系统 VLA。清华大学提出 UniAct¹²，直接从大量异构数据中学习一个通用动作空间，仅需 100 个演示数据即可完成四种控制接口，展现通用具身基础模型的跨本体迁移能力。北京大学提出多感官语言动作模型 MLA 架构，引入图像、点云、触觉、语言等完整多感官模态，通过整合异构信号提供更稳健的决策依据。2025 年 11 月，Generalist AI 推出 GEN-0¹³，在一个统一架构下实现感知和行动的异步时间处理，同时支持端到端的训练和模型规模扩展。GEN-0 模型大小扩展到 100 亿参数级别时，能更快地适应新任务。**二是**推理与控制高效协同的双系统 VLA 架构。Figure AI 公司推出模型，Figure AI 公司推出的 Helix¹⁴模型采用双系统架构实现 7~9Hz 认知推

¹² <https://arxiv.org/abs/2501.10105>

¹³ <https://generalistai.com/blog/nov-04-2025-GEN-0>

¹⁴ <https://www.figure.ai/news/helix>

理和 200Hz 本体运控的协同。北京大学、香港大学联合智源研究院和智平方公司提出 Fast-in-Slow¹⁵ (FiS)，将快速执行系统嵌入慢速推理系统，在确保常识推理的同时能够达到 117.7Hz 的动作生成频率。

端到端 VLA 通过融合或集成不同组件实现关键能力突破，整体呈现“拼接式”发展特点。一是融合力觉、触觉、听觉等提升全模态能力。上海复旦大学提出的 ForceVLA 独立引入力感知的混合专家融合模块 FVLMoE，通过跨模态路由融合多模态特征。三星中国研发中心提出触觉 - 语言 - 动作模型 TLA¹⁶，在架构输入侧增加了触觉编码器，其在编码触觉令牌 (Token) 后送入 LLM 进行视觉、语言、触觉等多模态 Token 学习。二是融合世界模型，强化 VLA 的物理规律和行为认知能力。上海交通大学推出 DreamVLA¹⁷，融合世界知识预测环境的未来状态，辅助动作搜索。中国科学院自动化所 DriveVLA-W0，用世界模型预测未来图像，生成密集的自监督信号，促使模型学习驾驶环境的底层动态规律。三是融合强化学习，提高数据稀缺条件下的策略泛化能力。得克萨斯大学奥斯汀分校提出 RIPT-VLA 引入强化学习的交互式后训练，在和环境互动中不断接收简单的成功/失败反馈。清华大学联合多个单位开发的首个面向具身智能的大规模强化学习框架 RLinf¹⁸集成了主流的强化学习算法、具身智能模型和仿真平台，

¹⁵ <https://arxiv.org/abs/2506.01953>

¹⁶ <https://arxiv.org/abs/2503.08548>

¹⁷ <https://arxiv.org/abs/2507.04447>

¹⁸ <https://github.com/RLinf/RLinf>

提供了易用、高效的强化学习训练工具链。清华大学的 SimpleVLA-RL 融合强化学习后，在与环境迭代交互中生成动作轨迹，并通过任务执行结果反馈不断提升性能。Physical Intelligence $\pi^* 0.6$ ¹⁹ 使用 Recap 强化学习方法改进 VLA 模型，能从专家纠正指导下修复错误，并不断学习来提升成功率。

端到端 VLA 尚未形成“感知—决策—执行”的完整闭环，还无法支撑真实环境应用。一是泛化性有限。任何在训练数据分布外的场景任务，都会导致性能急剧下降，难以应对现实场景的复杂情况。例如 1X World Model 在面临不熟悉的任务和环境时，很难准确地理解交互对象的物理属性。Physical Intelligence 的 $\pi 0.5$ 模型在执行时，一旦出现需要操作陌生物体或任务目标被遮挡等情况，成功率会大幅下降。另外，现有 VLA 模型性能表现与本体强相关，在跨本体迁移时很难复现模型性能。二是长程任务执行能力不足。谷歌最新研究结合分层 Gemini Robotics-ER 1.5 和 Gemini Robotics VLA 模型来执行“整理垃圾”、“物品交换”等长程任务时，成功率大多在 20%~40%。三是软硬协同难。一方面，模型与本体的结合需要在多个时间尺度上实现协同控制，且整个软硬协同链路上任何一环信号传输得不稳定，都很可能导致任务执行失败。美国敏捷机器人的联合创始人乔纳森·赫斯特提出了具身智能运行的多时间尺度问题，针对高层次的任

¹⁹ <https://www.physicalintelligence.company/blog/pistar06>

务规划和决策，处理频率大概在 1~10Hz，针对低层次的本体控制频率需至少在 100Hz 以上。

训练范式上，具身学习聚焦基础模型和机器人融合学习范式。与基础模型学习相比，具身智能学习整合了基础模型和机器人学习的实现方式，具备物理实体性和环境交互性两个典型特征。如何既能利用强大的人类先验知识，又能通过自主探索持续自我改进，是具身智能学习的关键所在。

采用预训练结合后训练的分阶段训练方式，实现通用任务策略到特定动作策略的泛化。预训练阶段，通过自监督学习让具身智能获得大量技能，并学习应对从未见过的环境和任务。后训练阶段，通过模仿学习、强化学习或两者结合等方式掌握特定技能。例如 Physical Intelligence π 0 和英伟达 GR00T N1 预训练阶段学习本体无关的通用策略，后训练阶段通过模仿学习针对少部分真机数据适配特定任务。中国科学院自动化所提出的 RoboGPT-R1 在预训练基础上，后训练使用监督微调（SFT）来快速获得基础任务能力，再通过强化学习来提升模型长期规划能力。

结合真实环境中的强化学习实现可扩展学习，目标是在真实环境中稳健地执行任务。强化学习一直是交互式环境中训练智能体的主要方式，但长期面临数据效率低、样本利用率不足和环境复杂度提升带来的维度灾难等问题。现阶段主要探索方向包括：一是“好奇心驱动”

的强化学习，在不依赖人工设计的任务和外部奖励情况下，高效且自主地探索环境。德国马普智能系统研究所提出 CWYC 方法在学习过程中将学习进度优先和惊喜信号引导作为内在动机，并鼓励模型自主探索并解决不确定情况。北京人形机器人创新中心推出的 Pelican-VL 模型运用了“刻意练习” DPPO 训练范式，模仿人类元认知的学习方式，通过强化学习探索弱点、生成失败样本，再进行有针对性的监督微调，让模型不断自我纠错和迭代进步。**二是自我改进的强化学习**，通过学习可扩展的奖励函数，在真实环境中自主练习，提升成功率。谷歌 DeepMind 提出数据驱动的奖励函数学习方法，将本体部署到真实环境中不断训练策略和奖励函数，进而学习新技能。**三是终身强化学习**，在学习新任务的同时不会遗忘先前学习的任务知识。慕尼黑工业大学提出 LEGION 终身学习框架，能够在连续任务输入时持续积累知识，并通过知识组合和重用解决复杂任务。

（二）数据成为核心需求，相关技术加速发展

高质量、大规模数据是支撑具身智能模型发展的关键，数据预训练阶段的可扩展性初步得到验证。已有研究证明，大规模数据对具身基础模型的性能提升具有显著作用。清华大学 ManiBox 论文指出空间泛化所需数据量与空间体积呈现幂律关系，如果要在 38 倍的空间泛化内实现泛化，那么数据量需要扩展大约 $38^{0.35}$ 倍²⁰。Generalist AI

²⁰ Tan, Hengkai, et al. "ManiBox: Enhancing Spatial Grasping Generalization via Scalable Simulation Data

GEN-0 模型研究表明每增加 10 倍数据，模型性能也会成比例改进²¹。

行业围绕仿真合成和真实数据积极探索数据采集方案，试图以可控成本提供更多高质量数据。具身智能领域数据包括海量互联网数据、仿真合成数据和真实数据三大类。在构建具身基础模型时，往往将以互联网多模态数据预训练的视觉语言模型为基座，让模型先具备通识知识，再结合仿真合成数据或真机数据训练，使其看懂物理环境、理解行为逻辑、学会动作技能并适应现实世界。其中仿真合成数据包括通过物理模拟引擎结合仿真平台、世界模型或其他合成技术合成的 3D 物体资产、3D 场景、本体轨迹或动作视频等，采集效率较高，但数据质量不足以支撑可靠的行为学习和真实场景泛化。真机数据是通过部署在真实环境中的机器人收集的数据，这类数据最能反映机器人实际行为，质量最高，但采集效率最低，且成本需要考虑机器人和大量人工费用。其中，最缺的是能反映本体真实行为的真机数据和合成数据，如表 1 所示。

表 1 具身智能数据采集方案对比

数据类型	采集场地	获取方式	积累量	质量	效率	成本
真实数据	训练场/开放环境	遥控采集、开放环境采集、动作捕捉	少	高	低	高
仿真合成数据	仿真环境	仿真平台、世界模型、生成式模型	中	中	中	中

Generation." arXiv preprint arXiv:2411.01850 (2024).

²¹ <https://generalistai.com/blog/nov-04-2025-GEN-0>

互联网 数据	开放环境	互联网多模态数据	海量	低	高	低
-----------	------	----------	----	---	---	---

来源：中国信息通信研究院

为此，行业围绕仿真合成和真机数据建设开展广泛技术实践。**仿真合成数据方面**，英伟达基于物理模拟引擎 Omniverse 的 Isaac Sim 平台构建了完整的数据合成管线，并在 2025 年几乎逐月更新，不断完善平台功能。Isaac Sim 5.0 则集成了 Cosmos 世界模型提升合成真实性，扩展运动数据生成模块、支持遥控操作或脚本驱动等采集方式。银河通用基于英伟达仿真平台的合成数据管线能够在一周内生成十亿级机器人操作数据，并突破了可变形布料模拟难题。光轮智能基于英伟达世界模型 Cosmos²²构建高效数据合成平台，能根据指令精确生成针对特定任务目标的数据集。**真实数据采集方面**，主要包括遥操作数据采集、开放环境采集和动作捕捉三类采集方案。其中，遥操作数据采集是绑定本体的手把手教学，包括采用 VR 设备、手持操纵杆、动捕服等本体操控方式。遥操作数据采集方式获取的数据在训练模型后更容易迁移到同一本体上，成为行业现阶段主要选择。例如星尘智能的 Atribot Suite 全身遥操作平台、零次方 ZERITH-H1 等。开放环境采集是支持跨本体学到人类动作示范，包括 UMI 夹爪、灵巧手和外骨骼等采集设备。开放环境采集系统研发由高校主导，例如斯坦福

²² <https://zhuanlan.zhihu.com/p/1930990275850929320>

大学推出的 DexUMI²³、DexCap 等，将人类手部动作直接转换为机器人手操作。上海人工智能实验室提出人形机器人遥操作系统 HOMIE²⁴，集成了人体运动策略和外骨骼手臂，使人形机器人能够在适应任意上肢姿势的同时实现行走或下蹲到特定高度的动作。动作捕捉是对人体的动作映射，主要包括光学捕捉和惯性捕捉两种方式。动作捕捉更多用于运动技能训练，例如在 2025 年 8 月召开的世界人形机器人运动会上，机器人表演的《秦俑魂》《英歌舞》等舞蹈，均通过采集动作捕捉数据完成。

现阶段实际可用的数据量仍远远不够，且数据采集、处理和使用等环节仍面临突出问题。行业普遍认为要实现物理智能的涌现至少需要百万小时、甚至千万小时的高质量真实行为互动数据。主要挑战在于：数据采集层面，需要考虑采集的数据和实际任务需求的适配度，数据质量和成本控制等多个层面问题。从任务适配度考虑，对于接触密集的任务学习更加依赖真实数据，且需要配备触觉传感器；对于导航任务则更适合在仿真环境下学习。在采集时也需要尽可能覆盖到所有可能出现的任务变化情况。从质量和成本控制考虑，真实数据采集依赖场景搭建、专业设备和专家示教，且整个采集过程需要数据采集员进行人在环的质量把控，单次采集成本投入远高于图像/文本数据。现有仿真平台大多模拟了“白盒”的物理环境，尤其是在物理属性模

²³ <https://arxiv.org/abs/2505.21864>

²⁴ <https://homietele.github.io/>

拟上需依赖大量理想化假设，很难对现实中的不确定性进行充分建模。

数据处理层面，需要考虑数据整合和处理方法。从数据整合考虑，不同真机平台或采集设备在传感器类型、精度、采样率上差异显著，导致真实数据难以有效共享。不同仿真平台采集的数据在数据格式、坐标系、物理参数和任务映射标准等方面均不一致，仿真合成数据整合困难。从数据处理方法考虑，数据标注的精细度、视觉信息处理内容和采样频率可能均不一致。例如 $\pi 0.5$ 的训练数据包括对象检测和视觉问答标注信息。Gemini Robotics-ER 的训练数据包括抓取关键点、子任务标注和 3D 检测框等。数据使用层面，关键在于找到更高效的数据训练策略，尽快让机器跨越“演示”阶段并部署到真实场景中，在实际应用中形成持续的数据回流。现阶段数据训练策略包括：**一是**真实到仿真再到真实的多源数据训练策略；例如逐级动力通过现实环境数据采集、仿真数据预训练和真机数据后训练的方式，在仿真环境中加速策略初始化，同时确保策略迁移到真实环境的可靠性。**二是**仿真到真实的大规模合成数据训练。例如银河通用利用 10 亿级的合成数据，让 VLA 模型学会抓取技能，并实现真实场景的泛化。**三是**真实到真实的真机数据学习。美国 Dyna Robotics 的 DYNA-1 在部署到真实场景后自主收集大量数据，并对奖励策略进行闭环训练，在六周时间内从 1 次成功操作进化到能连续工作 24 小时且成功率达到 99.4%。

（三）本体技术多元发展，重在软硬一体的零部件创新

模型带动具身智能本体多形态协同发展，包括人形、四足/多足狗、轮足式、轮臂式、仿生机器人、微型集群机器人、无人机、自动驾驶汽车等。虽然本体形态千差万别，但底层硬件模块具有一致性，零部件工艺链技术具备可参考性，能够实现技术复用或低成本技术迁移。感知系统方面，汉威科技、他山科技、柯力等公司均将自动驾驶汽车的传感器方案迁移应用在机器人领域，包括压力传感器、触觉传感器、力矩传感器等。动力系统方面，蓝黛科技、丰立智能等凭借汽车传动齿轮加工能力，研发生产机器人关节模组、谐波减速器等精密部件。能源系统方面，新能源汽车、机器人和无人机都对高能量密度、高倍率的固态电池有较高需求。近两年，氢燃料电池也开始广泛用于氢能无人机、氢动力汽车和氢动力机器人等。2025年4月，川崎重工推出氢动力四足仿生机器人“CORLEO”，能在满载300kg时持续作业8小时。执行系统方面，形态不同导致执行系统呈现明显差异。例如机器人的执行系统可以是移动底盘、夹爪、机械臂、灵巧手等；自动驾驶汽车是线控底盘，无人机是舵机等。底层核心都是电机、控制器和传动系统三部分。高性能材料结构件方面，铝合金、碳纤维复合材料和聚醚醚酮（PEEK）等轻量化、高强度的高性能材料结构件需求持续增长。例如空客A350机身夹板、托架等零部件，特斯拉Optimus Gen2均用到了PEEK材料。

机器人作为创新热点，围绕一体化关节、灵巧手、力触觉传感器和电子皮肤等重点突破。一体化关节方面，紧凑型、高扭矩和散热性能成为行业关注重点。宇树关节模组，将无框力矩电机、减速器、控制器等部件紧密结合，并集成热管理系统，将体积缩小至传统方案的 1/3，散热效率提升 50%，实现爆发力、灵活性和稳定性兼备。天工通过结构优化设计和关节导热技术，在相同功率下电机铜损降低 15%~20%，从源头减少热源产生。国外厂商如瑞士 Maxon、德国 Faulhaber 和瑞士 Portescap 在一体化关节方面也具备技术优势。灵巧手方面，关注自由度升级和智能感知能力。特斯拉在 Optimus 第二代上集成的灵巧手有 22 个自由度，Shadow Hand 灵巧手具有 24 个自由度，灵心巧手 Linker Hand 系列灵巧手科研版最高达 42 个自由度。帕西尼 DexH13 GEN2 上搭载了 1956 颗 ITPU 高精度多维触觉传感器。北京通用人工智能研究院与北京大学联合研制仿生视触觉灵巧手 F-TAC Hand²⁵，集成 17 个高分辨率触觉传感器和高维触觉数据的手型生成算法，实现接近生物触觉的感知能力。力触觉传感器方面，以六维力触觉传感器的突破为主。六维力传感器在结构设计、应变片生产加工、标定检测等方面有较高技术壁垒。美国 ATI、SCHUNK、Robotiq 等品牌在性能指标和高端市场上具有较大先发优势，国内柯力传感、宇立仪器、蓝点触控和坤维科技等企业正在持续推动关键技

²⁵ <https://arxiv.org/abs/2412.14482>

术突破。电子皮肤方面，向多模态感知单元集成、高性能柔性材料创新等方向发展。电子皮肤旨在通过模仿人类皮肤结构和感知功能以提升机器人的操作精度和交互安全性。2025年10月，天安新材与他山科技宣布联合开发兼具接近觉感知与亲肤手感的电子皮肤，重点攻关柔性机器人和智能交互领域。

本体要规模化部署到真实场景中仍需解决三大难题：一是本体结构存在物理局限性。随着本体构型的日益复杂，高自由度的协调控制和逆运动学解算上的难度明显增加。本体在实际使用中的寿命和稳定性还无法保障。例如有厂商反映采购的双臂操作平台在不到一周时间内就完全损坏，部分灵巧手可能只有几个星期的使用寿命。二是规模化应用难。本体需要结合实际业务需求，能承担关键任务才能发挥实用价值。目前正处于场景验证阶段，硬件层面的工程化与生产流程标准化不足，不同厂商，甚至不同产品间，在设计标准、模块接口、驱动方式都可能存在较大差异。而软件层面也需要结合任务需求进行定制化二次开发，任何场景需求变化都需要软件层面的再调试。当前大多仍是单台套交付模式。三是成本控制难。如果没有足够大的市场需求或明确的应用场景，就难以分散研发与产线建设的固定成本。产能过剩/产量不够规模就无法摊薄这些成本。

（四）云-边-端协同运行，支撑现实应用落地

端侧计算芯片要求升级，向高算力、大带宽、低能耗和集成化方

向发展。随着具身智能技术发展，多模态动态交互和模型端侧部署以及场景应用需求增加，对端侧计算芯片也提出了更高要求，推动其不断升级。**一是高算力支持大模型在本体上的部署和高性能计算。**英伟达 Jetson Thor 可提供 2070 TFLOPS 算力，支持 Qwen、Cosmos 等基础模型的高效推理；小鹏汽车 P7 系列搭载三颗自研图灵 AI 芯片，整车有效算力大于 2000TOPS，支持 VLM 和 VLA 模型的本地部署；黑芝麻智能计划于 2025 年量产的华山 A2000 系列最高算力也达到 1000TFLOPS。**二是大内存带宽支持大规模多模态数据的高效处理。**英伟达 Jetson Thor 内存容量相比 Orin 系列提升两倍，显存带宽约为 273 GB/s。地平线远征 6P 内存带宽 205GB/s，与英伟达 Orin 系列相当。瑞芯微推出的端侧大算力协处理器 RK182X，内嵌高带宽 DRAM，实现百 GB 级带宽²⁶。**三是低能耗与良好的热管理，**以支持系统长时间高强度运行，避免过热导致性能下降或系统崩溃。地平线远征 6P 采用 ASIC 方案，其能效比是 Thor 系列的 1.2-1.4 倍。**四是算控一体化集成的高算力控制器，**以端侧计算芯片为核心，通过集成丰富的通信接口和软硬件工具链，实现微秒级的多源数据同步和融合分析，提升大小脑一体化控制能力。例如天准科技的星智 007、国讯芯微 NSPIC-R007NP+ 等。

云-边-端分布式计算为具身智能提供了互补运行环境，支撑系统

²⁶ <https://www.rock-chips.com/a/cn/news/rockchip/2025/1021/2110.html>

智能在现实场景中的有效泛化。**一是**云侧依赖大规模计算资源，成为“大脑”升级的算力底座。行业厂商通过云服务形式提供从数据服务、仿真训练再到模型服务等开发支持，帮助模型能力快速升级。如华为云 CloudRobo 平台在云侧结合视频生成模型和物理引擎提供合成数据工具链。**二是**云端协同依赖统一开发平台，实现“脑”与“身”协作开发和跨本体管理。腾讯 Tairos 具身智能开放平台支持自定义本体模型导入和大模型在本体上的快速验证。华为云 R2C(Robot to Cloud) 协议、阿里达摩院机器人上下文协议 RynnRCP 构建了端云适配接口，实现对机器人本体的统一管理。**三是**边云协同提供数据实时分析和处理能力。美国亚马逊通过边云协同方式让工厂机器人可以实时适应环境变化，在边侧直接处理温度、产品类型或材料等动态数据信息。**四是**端边云协同依赖操作系统及中间件进行本体功能开发和应用优化。开源项目 ROS 2 通过工业级的数据分发服务（DDS）支持可靠数据传输。美国 OpenMind 推出 OM1 操作系统支持 OpenAI、Gemini、DeepSeek 等大模型 API 的即插即用和跨本体开发。具识智能 insightOS 内置功能开发包，支持端云异构算力整合和 AI 模型调度，帮助本体厂商快速开发智能功能。

云边端协同运行环境仍面临多方面技术挑战：**一是**接口兼容性。模型输出的动作空间和硬件设备上存在异构性，且不同设备集成的中间件交互协议碎片化，导致系统一致性兼容难。**二是**端侧计算限制。

端侧芯片在模型部署和计算层面通常存在内存、显存、缓存和存储空间限制。据调研本体上嵌入的计算芯片以英伟达 Orin Nano 为主，算力在 275 TOPS 左右。该芯片在运行 7B 参数规模的模型时性能会明显下降，运行 3B 参数规模的模型也需要更多的加载时间²⁷。三是通信和资源分配限制。云边端协同计算时会受到数据通路带宽和网络不稳定的影响，在资源分配与服务供给中面临灵活权衡计算需求与通信成本，复杂数据流的实时处理与协同决策等挑战。

三、场景驱动下的具身智能产品谱系不断丰富

我国超大市场规模为具身智能应用落地提供更多机会，多样化场景需求也催生了具身智能产品谱系的不断丰富。从产品形态看，目前具身智能产品包括机器人、智能运载装备和新型智能产品三大类，其中人形机器人的技术突破，是从核心部件到本体机构设计的系统性升级。围绕人形的技术，也正在外溢到其他机器人和各种形态的物理载体上，驱动多样化载体融合发展。复合轮臂式机器人、四足/多足机器人、自动驾驶汽车、无人驾驶航空器等典型产品形态成为当前发展热点。同时，如仿生机器人、智能机器臂、无人船、变形移动装置、集群式微型智能机器人、智能可穿戴设备等产品形态也在加速探索。从场景应用看，以人形机器人、复合轮臂机器人（轮式底盘结合双臂）为代表的新兴具身智能产品以科研、训练场、娱乐表演或特定场景试

²⁷ https://dev.to/jeremycmorgan/review-the-new-nvidia-jetson-orin-nano-4ci7?utm_source

点为主，部分产品在逐步推进行业场景的小规模应用。从模型部署看，现阶段部署端到端 VLA 模型的产品还处于演示阶段，例如容易出现夹爪抖动严重、人工干预重置等情况等，且边缘案例（corner case）、长尾问题（long-tail problem）和失效案例（bad case）等很可能导致系统任务失败。部署 VLM 模型的产品主要用于接待导览、展厅讲解、教育陪伴等，大多处于内测阶段，产品化程度不足。

（一）机器人成为具身智能的最“热门”载体

1. 人形机器人类型是业界追逐的热点方向

双足人形机器人加速走向赛场、训练场和行业场景实训，运动和作业能力同步升级。2025年以来，国外特斯拉公司的 Optimus 系列、波士顿动力公司的 Atlas 系列以及 Figure AI 公司的系列产品持续更新迭代。2025年2月至9月，Figure AI Figure 02 机器人在具身基础模型 Helix 加持下不断训练、学习双机协作、物流分拣、洗衣服、叠衣服、叠毛巾和洗碗等新场景和新任务。2025年10月，Figure AI 推出 Figure 03，在公开的视频中能连贯地完成一系列小任务组成的复杂场景：例如集合“收盘子、冲洗、放入洗碗机和返回充电座”等长链条动作。国内也涌现出宇树 H 系列和 G 系列、智元灵犀系列、优必选 Walker 系列、乐聚夸父系列、星动纪元 STAR 系列等一批人形机器人产品。2025年从4月北京人形机器人半程马拉松、5月世界机器人大

赛的格斗擂台赛，到 8 月的世界人形机器人运动会，人形机器人从实验室走到真实赛场、训练场，并逐步推进行业场景的试点验证，在真实场景的实践打磨中推动技术不断进步。

轮式人形机器人有望在短时间内形成真正有应用价值的产品。

2025 年，轮式人形机器人成为市场关注重点，世界机器人大会发布具身智能机器人产品中轮臂式占比超 50%。轮式人形机器人相较双足结构在实际应用场景中具备作业、续航和成本优势。作业上，可以将更加关注“操作”能力，无需再去应对叠加下肢的全身协调控制难题。续航上，轮式底盘相比双足的续航性能要提升 2 倍以上。成本上，同等身高下，两条腿的原材料成本（BOM）比一个轮式底盘贵十倍左右。我国代表厂商及产品如国地共建北京人形机器人创新中心的天轶系列、智元机器人精灵 G2、千寻智能墨子 Moz1、自变量的量子 2 号、星海图的 Galaxea R1 Pro、聆动通用 LDT 采训推机器人等。

人形机器人加速应用探索，场景覆盖广泛。其中，双足人形机器人缺少成规模成体系的应用场景。由于人形机器人的双足构型控制难度较高，产品距离成熟可用的阶段仍有数年差距，目前双足人形机器人的应用场景仍以教学科研、娱乐表演、接待导览等为主。国内优必选、乐聚、众擎为代表的双足人形机器人厂商，也正在依托训练场或联合行业厂商开展“实训”。轮式人形机器人则在地面条件较好的场景中可以更好地发挥具身智能产品的作业优势，正在工业制造、商业

服务等场景中加速落地，商业订单不断增多，已进入规模落地前夕。

家庭服务则是业界的主要探索方向但目前仍在研发和训练阶段。工业制造领域，上海智元与富临精工达成合作，提供近百台远征 A2-W 机器人用于富临精工的产线，可满足产线全量排产需求。商业服务领域，银河通用基于 Galbot G1 机器人形成了智慧药房解决方案，机器人能高效地完成取货、送货、补货等任务，并通过自主规划取送路径，实现 24 小时无人值守的便捷服务²⁸。

应用局限性在于，人形机器人真正进入工业产线或生活环境“干活”则面临着客户“不能用、不愿用、不敢用”的问题。一是作业能力仍有待提升，客户“不能用”。现有人形机器人产品仍不能满足工业场景中的高精度、高可靠性和一定程度的泛化工作能力要求。对于服务场景中的高交互要求和非标准化流程的工作更是尚有较大差距。二是综合成本提高使用门槛，客户“不愿用”。目前具备实用功能的全尺寸人形机器人产品动辄数十万元，企业大规模购入的成本投入巨大，且后期的维护保养以及故障维修方面也是一笔庞大的开支。此外，企业试用、调试人形机器人时对生产效率的影响过大，试错成本过高。三是安全规范和标准滞后，客户“不敢用”。由于目前人形机器人一般拥有较大的动能和静止质量，在控制不当、系统出错或被恶意远程操控的情况下，极有可能出现摔倒或失控等问题，对应用场景中的其

²⁸ <https://finance.eastmoney.com/a/202507273468476973.html>

他人员和环境造成无法挽回的损伤。同时，还可能通过摄像头、麦克风等传感器设备，收集用户的环境信息、个人信息、行为数据和生物特征数据等，带来严重的隐私问题。

2. 机器狗复杂地形适应性升级，初步落地行业场景

目前，机器狗已经成为四足机器人的主要设计形态。一方面，“狗”的形态在腿部结构、体型比例等方面更适合机械上的工程实现。另一方面，机器狗形态更容易被人们接受和理解，可以充当“警犬”“搜救犬”“导盲犬”等人们熟悉的角色，自然地融入现实环境中。四足机器狗运动性能和交互智能大幅升级，成为全能运动员和智能伙伴。

一是更能跑，更能爬。浙江科创中心人形机器人创新研究院研发的黑豹四足狗以 10.3 米/秒的速度打破由波士顿动力 WildCat 创造的 8.89 米/秒的奔跑纪录²⁹。瑞士苏黎世联邦理工学院的研究让 ANYmal D 四足机器人学会攀爬 70 度到 90 度不同倾斜角的梯子³⁰。**二是更轻，但更强。**宇树 Unitree A2 以 37kg 的机身承受 100kg 的负载，自重负载比相较上一代产品 B2 提高了 1.35 倍，能攀爬 1m 高度，也能行走在碎石地、山洞、岩石坡复杂地形环境。**三是以狗的形态，亲近和服务人。**在具身智能加持下，四足机器狗从环境感知到情感认知再到全模态交互能力升级，成为人们的生活伙伴。维他动力 Vbot 通过协同屏

²⁹ <https://www.cls.cn/detail/2077309>

³⁰ <https://arxiv.org/html/2409.17731v1#S2>

幕表情，肢体、扬声器和指示灯提供丰富情感表达。蔚蓝科技 BabyAlpha 系列机器狗能识别语音指令、人脸表情，并做出拟真动作反馈。

四足机器狗加速走向商用和消费市场，凭借模块化组件集成不断扩大应用范围。商用层面，在电力巡检、安防巡逻、应急救援等行业场景初步落地。云深处绝影 X30 四足机器狗在浙江某换流站实现 1000 小时无故障电力巡检³¹，并与湖南长沙市消防救援支队合作，代替指战员进入有毒、缺氧或浓烟等危险灾害事故现场开展侦察、探测和搜救。消费层面，成为教育、娱乐、陪伴等领域的新型消费品。可以科技 Loona 机器狗发布后半年即达成 400 万美元销售额。美国 Tombot 的仿生机器狗 Jennie 已有超过 1.6 万名用户预订。通过集成或扩展组件方式拓展细分场景应用。例如七腾科技的防爆智能四足机器人“钛格”，背部可加装轻型机械臂，实现 5kg 级物品的抓取、转移、分拣，配合本体的防爆能力，可在石油、化工等易燃易爆场景实现灵活作业³²。

同时，伴随机构设计和运控算法进步，轮式四足、轮足可替换复合结构以及多足结构的高动态运动控制问题逐渐得到解决，轮足和六足机器狗也开始显现出与四足机器狗互补的应用优势。一是轮足机器狗能兼顾轮式高效移动的机动性和足式越障能力，相比四足有着明显的能耗优势。瑞士 RIVR 递送机器人采用轮足结构能快速通过城市开

³¹ <https://cn.chinadaily.com.cn/a/202507/28/WS688762bca310a07bb590aa6f.html>

³² <https://finance.ifeng.com/c/8EYi1bHGw6T>

放道路，可以跨越四足难以应对的小坑、水平细条等微型障碍，也能在配送的最后 100 米爬楼梯送货。**二是六足机器狗**行走时三条腿着地，形成三角结构稳定支撑身体，天然具备运动平稳性、稳定性和承载性。越疆六足机器狗能拉动自身重量 5 倍物体，且能稳定通过湿滑、不平整地面而不会滑倒和侧翻。另外，相比四足的小步快跑，六足机器狗的三角步态也有效降低了行走噪音。

应用局限性在于，一是行业采用率有限。IDC 数据显示 2024 年四足机器人出货量约 2 万台，其中商用级产品仅占总出货量的 27.9%³³，应用主要集中于教育培训和娱乐等场景。石油、化工、安防等作为主要采用方基本只支持单台套采购模式，这些高强度作业场景对防爆、防腐蚀、防火或防护等级等特殊需求有明确差异，且作业环境挑战大（例如 60 度倾斜角爬坡、细小金属杆密集的障碍区域等），导致单台套改造成本高、检测认证周期长，很难产生规模效应。**二是用在现实复杂环境中的导航稳定性和检测性能仍需优化。**例如石油炼厂存在密集管廊、各种钢制格栅平台、狭窄通道等，会出现不可预料的微小障碍让机器人卡住。另外，金属管网拓扑结构也打造了弱信号区域，容易产生信号延迟让机器人步态失衡和检测失败。**三是面向细分场景的定制化解决方案供给不足。**目前主要是将机器人作为“可移动摄像头”用在巡检场景，还有巨大的模块化开发空间。近期，具微科技在

³³ <https://my.idc.com/getdoc.jsp?containerId=prCHC53644125>

杭州落地具身智能4S店，提供用户定制和乐高式模块化开发的产品服务模式。MOVEONEW T1即插即用的接口可兼容智能摇篮、储物箱、机械臂、摄影测绘仪等模块，用一台机器适用于育儿、户外、商用和应急四个场景。

3. 仿生机器人、智能机械臂等产品加速探索

智能机器人在硬件结构上与具身智能更加契合，可同时具备优越的运动和作业能力，应用范围也更加广泛。除目前关注度较高的人形机器人、复合轮臂式机器人、四足/多足机器人外，业界也在广泛探索其他构型的智能机器人产业。

一方面，学界和产业界加速各类仿生机器人的探索性研究。该类机器人以现有生物为模板、以具身智能技术为驱动、以场景落地为目标，新的构型和新的产品不断涌现。如清华大学深圳国际研究生院研发的仿生蝠鲼软体机器人，提出了双稳态扑翼设计，能够高速运动的同时实现多种游泳模式，为水下探测及环境监测等领域提供了新的技术支持³⁴。再如哈尔滨工业大学研发的蛇形机器人，可实现水中和陆地的两栖运动能力，模仿实现蛇的移动、伸缩、转向、竖起等动作，在缝隙、管道等其他形态机器人难以施展的环境中运动自如，可用于巡逻巡检、灾害搜救等应用场景。

另一方面，明确的需求场景催生了机械臂等传统产品叠加具身智

³⁴ <https://www.science.org/doi/10.1126/sciadv.adv9572>

能的全新形态。在制造业转型升级的行业背景下，传统工业机械臂在具身智能加持下的能力升级也成为业界的重点研究方向，在不改动或少改动原有机械结构的前提下，通过融合智能感知、决策和执行控制能力，实现程序执行到自主作业的“跃迁”，避免在生产不同产品和型号时频繁对生产线进行调整，助力企业实现柔性制造。此外，还有智能化的高空玻璃幕墙清洗机器人、太阳能光伏板清洁机器人、酒店清洁机器人等各类产品，由于具有明确的应用场景和技术路线，已经实现了小批量的落地应用。

（二）智能运载装备是具身智能落地“最快”载体

1. 自动驾驶汽车法规不断完善，加速商业化进程

具身智能成为自动驾驶领域的前沿探索方向，让车辆更能像司机一样思考，交互性和复杂路况处理问题解决能力明显提升。理想 i8 搭载 VLA 司机大模型后，可以直接通过“前方掉头”“前进 10 米”等简单指令调整行车路线。Waymo 推出 VLA 架构的 EMMA 端到端模型面对潮汐车道、施工区和复杂交通规则时能适应性调整驾驶行为。

自动驾驶汽车作为新质生产力的典型代表，各地正在竞相发展，政策法规不断完善。国内北京、武汉、广州、南京等出台地方性法规，为自动驾驶汽车的上路通行和场景试点提供有力支持。其中，北京市 2025 年 4 月实施的《北京市自动驾驶汽车条例》明确支持自动驾驶

汽车应用于个人乘用车出行、出租车等客运服务。美国国家公路交通安全管理局（NHTSA）在2025年4月宣布扩大自动驾驶汽车豁免计划，并批准Zoox公司的自动驾驶汽车路测³⁵，加速松绑监管政策，为规模化营造宽松环境。

自动驾驶汽车L3级车型量产和L4级商业化运营并行推进。摩根士丹利预测到2030年，自动驾驶汽车的市场规模将达到2000亿美元³⁶。一是L3级别智驾能力上车。华为智驾方案乾崑智驾ADS4计划在2025年底实现高速L3商用试点³⁷。广汽、吉利、小鹏汽车等均宣布要在2025年实现L3级别车型的量产上市。二是L4级以无人出租车（Robotaxi）为代表，正在探索规模化、商业化示范应用。高盛报告预测到2030年Robotaxi带来的个人出行市场规模将突破230亿美元³⁸。百度萝卜快跑的每周全无人订单超过25万，全球出行服务次数总计超过1700万次，已覆盖全球22座城市；全无人里程突破1.4亿公里，自动驾驶总里程数累计超过2.4亿公里。2025年以来，萝卜快跑陆续宣布将落地迪拜、阿布扎比、瑞士、德国、英国等地，并与全球出行服务平台Uber、Lyft达成合作，推进全球化部署。文远知行2025年第二季度财报显示，Robotaxi业务二季度营收4590万元，同比大幅增长836.7%。美国自动驾驶公司Waymo运营车辆超1500辆，

³⁵ <https://www.nhtsa.gov/press-releases/nhtsa-issues-first-ever-demonstration-exemption-american-built-tomated>

³⁶ <https://www.bloomberg.com/news/videos/2025-07-28/morgan-stanley-sees-200b-autonomous-car-market-by-2030-video>

³⁷ <https://cj.sina.com.cn/articles/view/1496814565/593793e502001qqpi>

³⁸ <https://www.goldmansachs.com/pdfs/insights/goldman-sachs-research/robotaxi/report.pdf>

已在5月完成1000万次无人驾驶行程。2025年6月，特斯拉在美国奥斯汀正式入局Robotaxi的竞争，并宣布车队规模将在几个月内增加至千台。三是L4级别无人驾驶小车在无人公交、无人配送、无人环卫等场景应用落地，推动未来出行与生活方式转变。截至2025年7月底，新石器已在全球15个国家和地区累计部署超过7500辆无人配送车，L4级自动驾驶里程超4200万公里。

应用局限性在于，**一是法律法规仍需进一步完善**。现行法律规范主要针对传统车辆和人工驾驶模式，尚未针对自动驾驶技术形成明确的法规支持和规范指导，自动驾驶车辆上牌、上路、事故责任划分等仍是难题。自动驾驶企业在不同城市开展测试时，需要针对不同城市的法律法规进行调整和适应，导致研发和运营效率显著降低。**二是技术成熟缺乏高质量场景数据喂养**。当前覆盖更复杂、更广泛场景的真实数据高度稀缺，单纯采用仿真模拟不足以实现模型能力的有效泛化。应用规模越大，才能更早发现和解决问题，从而反哺技术加速迭代。虽然全国多地已经开展测试，但真正实现规模化试点示范的仅北京、武汉等少数城市突破千台级规模，而其他城市大多停留在百台车辆、数百公里道路的小范围测试阶段。

2. 无人驾驶航空器快速发展，推进多元场景应用

无人机和电动垂直起降飞行器（eVTOL）作为无人驾驶航空器的核心载体，在具身智能技术融合下实现作业能力全面升级。通过具身

智能深度应用，无人驾驶航空器正在从基础空中移动平台（以无人机为主）和运载平台（以 eVTOL 为主）向具备自主作业能力的飞行机器人进化，从简单的空中拍摄到复杂的环境认知，从单一任务到自适应协同，实现自主性和协同作业能力的全面突破。例如，香港大学研发的飞行机器人 SUPER 能在没有 GPS 和预先绘制地图的情况下，仅通过视觉系统实现在野外密集树林、夜晚等复杂环境的自主导航，开阔地带飞行速度可超过 20m/s³⁹。

无人机在农林植保、测绘勘探、安防监控及物流配送等场景的应用更加广泛和深入。2025 年 6 月，美团无人机在香港航线上进行外卖配送服务，通过智能感知和决策，实现路径高效规划和精准避障，对比传统骑行方式效率提升近 7 倍⁴⁰。随着深度强化学习、多智能体协同和云边端分布式计算等技术进步，多个无人机能在未知环境中进行多机协同调度和智能任务分配，更好适应于高密度任务和复杂环境作业。例如贵安新区使用 110 架无人机执行电力线路巡检任务⁴¹。贵州贵阳用 4 架重载无人机协同作业，仅用 4 分钟便将 374 公斤塔材从山脚塔材堆放地运送到山顶目的地⁴²。eVTOL 在城际运输方面取得突破性进展。eVTOL 支持单次更大载荷、更多货物的点对点运输，通过搭载自动驾驶系统实现智能导航和避障、自主起降和紧急情况处理

³⁹ <https://www.science.org/doi/10.1126/scirobotics.ado6187>

⁴⁰ https://m.guancha.cn/economy/2025_06_06_778470.shtml

⁴¹ [https://cn.chinadaily.com.cn/a/202506/11/W\\$68495cffa3102053770379e0.html](https://cn.chinadaily.com.cn/a/202506/11/W$68495cffa3102053770379e0.html)

⁴² <http://www.chinapower.com.cn/dlxh/wurenji/2025-08-20/283045.html>

等，其高效直达特性在医疗急救、生鲜等特殊物品运输方面展现效率优势，确保长时间、长距离的安全飞行。2025年7月，亿航智能物流无人驾驶eVTOL装载海鲜从广东珠海唐家港运起飞，途经珠海、中山、广州三市82.9公里，在广州穗港码头落地，历时约55分钟，验证了城际间长航程、高时效运输的可行性⁴³。

应用局限性在于，一是智能水平仍有很大提升空间。现有多传感器融合（激光雷达+视觉+毫米波雷达）在雨雾天气可靠性下降，难以应对电缆、玻璃幕墙等细小障碍。仍需进一步提升环境感知鲁棒性，同时需强化边缘计算能力，实现百毫秒级反应式控制。**二是续航和载荷限制。**商用无人机电池续航普遍低于1小时，载荷多限制在5kg以内，制约了在物流配送、农业植保、安防监控等场景的工作效率。eVTOL电池能量密度也仅能勉强满足200公里以内的短途飞行。如何提升电池的能量密度、放电倍率、快充能力是下一步技术突破重点。**三是管理体系滞后问题。**我国尚未形成统一的适航认证体系，适航审定缺乏标准化依据，航路规划需综合考虑飞行高度精准设定、建筑物障碍物分布、交叉避让规则等要素。同时，多飞行器并行可能引发“空中打架”风险，城市级空中交通管理系统的缺失成为规模化运营的潜在阻碍。

此外，具有自主导航和作业能力的无人船也在快速发展。搭载具

⁴³ <https://news.cctv.cn/2025/07/30/ARTIpKfLgArEVLO2yDLIBlET250730.shtml>

身智能的无人船可以用于智能化的水文监测、安防巡检、河道运维、水上救援等功能。如欧卡智舶的 APAS，具备感知识别、自主定位、规划决策和自主控制算法，可实现自主航行、自主避障和自主靠离泊等一系列全自主航行功能⁴⁴。

（三）新型智能产品是具身智能“前沿”探索方向

在智能机器人和智能运载装备两大方向快速发展的同时，变形移动装置、集群式微型智能机器人、智能可穿戴装备等新式的具身智能产品构型也在加速探索，有效丰富了具身智能的产品谱系，同时也拓展了具身智能产品的能力边界，成为未来可规模化落地产品的高水平储备。其中**变形移动装置、集群式微型智能机器等方向以学界的科研探索为主**。变形移动装置方面，瑞士洛桑联邦理工学院研发的 GOAT 自适应变形装置，可以实现从扁平漫游车到球体的连续、可控的大尺度三维变形，能无缝切换驾驶、翻滚、游泳等多种运动模式，未来有望应用于应急救援、野外探索、月球探测等场景⁴⁵。集群式微型智能机器人方面，香港中文大学研发的微型机器人集群，未来有望应用于靶向给药、细胞操作等前沿医疗领域⁴⁶。

智能可穿戴设备则正在形成可落地的行业解决方案，其中的典型代表是智能外骨骼。目前，外骨骼机器人可分为上肢、下肢和全身三

⁴⁴ <https://orca-tech.cn/product/apas>

⁴⁵ <https://actu.epfl.ch/news/morphing-robot-turns-challenging-terrain-to-its--2/>

⁴⁶ <https://ieeexplore.ieee.org/document/10598230>

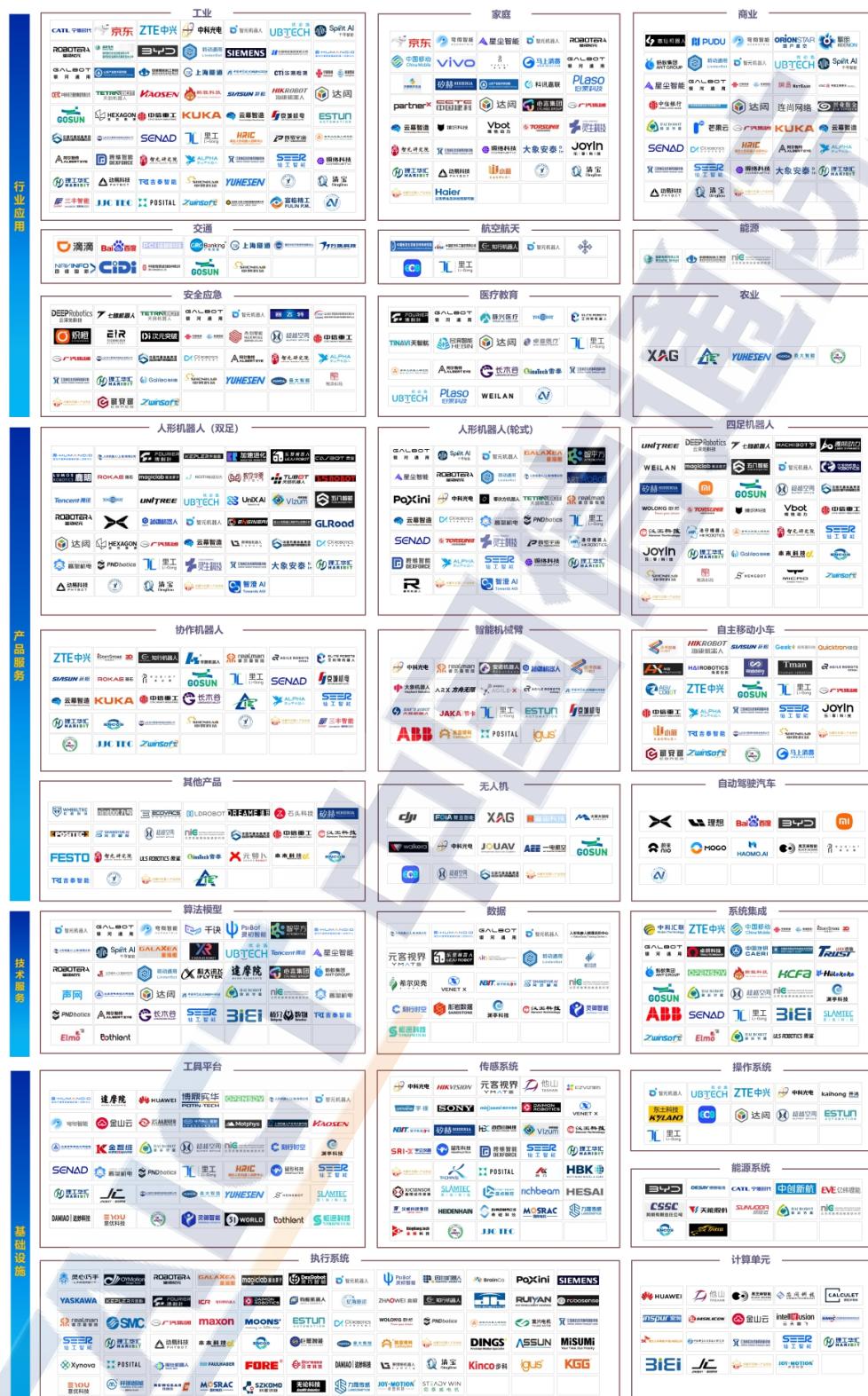
大类型，上肢外骨骼机器人类似于给手臂和肩部披上“机械肌肉”，助力使用者轻松举起重型工具或货物，完成装配、搬运作业。下肢外骨骼机器人则聚焦于对腰腿部位的支撑，其自适应支撑结构既能在负重登高时构建动态平衡支点，也能通过仿生步态算法，赋予截瘫患者自然流畅的行走节奏。全身外骨骼机器人则更进一步，通常采用模块化设计，将机械框架与人体的生物力学耦合，能支持特定部位的功能强化，提升使用者整体运动能力。从应用领域上看，目前智能外骨骼主要分为增强类和康复类。增强类智能外骨骼通过各类传感器准确理解使用者的动作意图，为关键部位提供精准的力量补充。康复类外骨骼则通过与用户的自然交互和智能感知，帮助脑卒中、脑损伤和脊髓损伤等患者重拾肢体运动能力。智能外骨骼面临着成本高、续航低和舒适度差三大落地瓶颈。成本方面，外骨骼要继承多种微型传感器、驱动器、芯片等，且为了保证轻量化，大量使用钛合金、碳纤维等材料，导致成本较高。目前具备AI动态调节、多模态交互等技术的智能外骨骼产品价格在数十万元，超过大部分用户的消费能力。续航方面，受限于当前电池技术的能量密度以及自身的重量负担，大多数智能外骨骼系统陷入了“充电两小时，工作一刻钟”的尴尬境地。舒适度方面，外骨骼机器人的骨架通常由金属连杆构成，采用刚性结构设计，导致体积和重量较大，通常一套下肢外骨骼有10至30公斤。此

外，使用者穿戴外骨骼时，需要使用多层绑缚带将金属框架与人体腰、腿等部位多点锚定，佩戴舒适度欠佳。

四、具身智能产业生态加速完善

（一）产业链横跨多个领域，生态体系初步形成

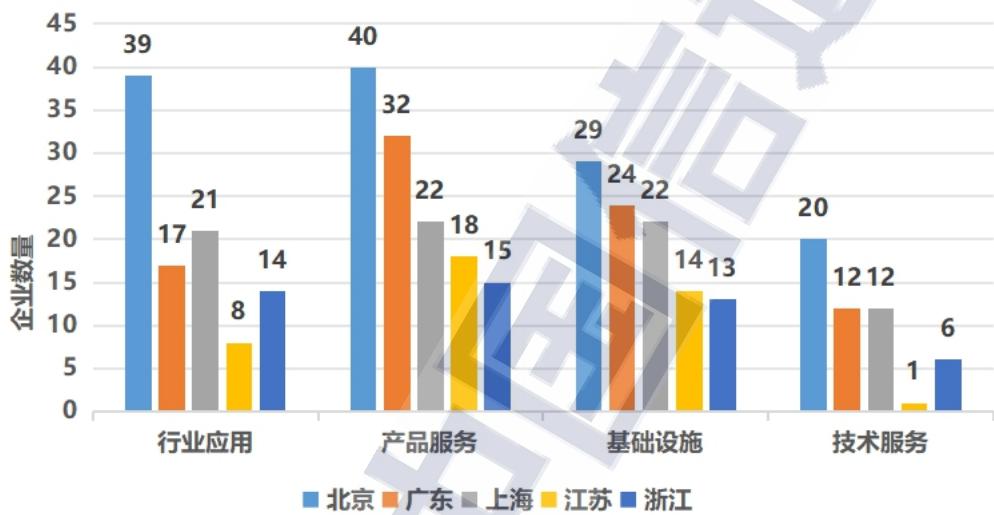
具身智能的概念较为泛化，在初期存在许多的不同具体解读，但近年来，经过业界的不断探索和市场的持续反馈，具身智能产业链横跨多个领域，涌现出了大批企业。目前具身智能的产业结构基本上可分为行业应用、产品服务、技术服务和基础设施四大板块。中国人工智能产业发展联盟（AIIA）发布的具身智能产业图谱初步统计（截至 9 月 25 日），如图 6 所示。国内以具身智能为核心业务的企业达到 352 家，其中行业应用 147 家、产品服务 167 家、技术服务 74 家、基础设施 157 家（部分企业存在重叠）。



来源：中国人工智能产业发展联盟具身智能工作组、具身智能测试实验室（截至9月25日）

图 6 具身智能产业图谱

从地域角度看，目前我国在十余个省市/地区均有企业布局，产业地域集中度极高。北京市（93家）、广东省（78家）、上海市（50家）三地企业数量合计占比近50%，构成产业核心支柱。江苏省（35家）、浙江省（33家）紧随其后，五大省市汇聚主要产业力量，其余地区分散，“头部集中、长尾分布”态势显著，具体如图7所示。

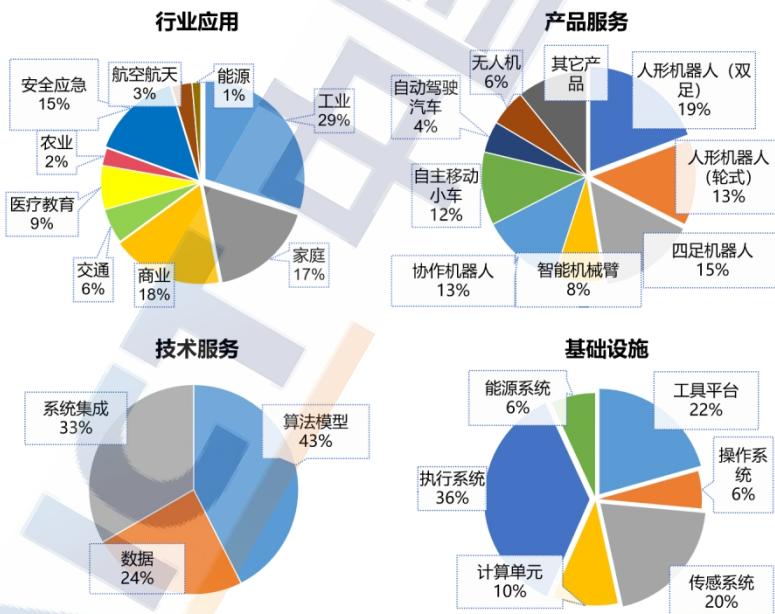


来源：中国信息通信研究院

图7 具身智能主流企业区域分布图

从细分环节看，行业应用方面，工业、家庭、商业和应急行业的企业占比合计达到了80%，体现了当前业界的重点应用方向，医疗教育、交通占比为9%和6%，也是目前在积极探索的垂类领域。产品服务方面，人形机器人（双足/轮式）是业界关注热点，企业占比达到32%。四足机器人、协作机器人、自主移动小车占比均超过10%，且此类构型产品技术正在加速成熟，落地应用进展迅速。技术服务方面，

算法模型、系统集成和数据服务“三足鼎立”，但模型方面明显更为火热，企业占比达到43%。基础设施方面，执行系统、工具平台和传感系统成为行业热点，入局企业较多，占比分别为36%、22%和20%。其他方向占比均超过6%，呈现均衡发展的整体态势。四大板块企业具体情况如图8所示。但由于具身智能的应用场景过于广泛，跨行业、跨场景现象普遍，对产品提供方提出了巨大的挑战。目前业界正在出现一批在现有机器人基础上进行二次开发，并向用户提供最终解决方案的新式集成商，通过更加贴合用户需求的专业化、定制化服务，有效解决现有供需双方衔接不够紧密的问题。



来源：中国信息通信研究院

图 8 四大板块整体情况图

（二）训练场建设成热点，实际功用仍需进一步验证

具身智能训练场通过搭建应用场景、采集数据、训练模型和反馈迭代，帮助破解具身智能领域物理世界数据采集成本高，效率低下的核心瓶颈。据不完全统计，截至2025年12月底，国内已建成的具身智能训练场超过30家，覆盖了北京、上海、天津、广东、浙江、江苏、安徽、湖北、湖南、福建、河南、四川、山东、辽宁等省市，其中公开披露的21家训练场总面积超过10万平方米。总体上看，目前国内的具身智能训练场主要有以下三大特点。

一是构建具身智能走向真实应用的“培训学校”。训练场对真实场景的还原能力决定了产出数据集的“深度”，对真实场景的覆盖能力决定了产出数据的“广度”。具身智能产品的应用场景探索聚焦在制造、物流、服务、能源电力以及医疗行业等，因此这些行业场景作为训练场建设的重点方向。大部分训练场以广泛覆盖应用场景为建设思路，北京人形机器人数据训练中心一期⁴⁷搭建了家庭康养、汽车装配、机器人新餐饮、3C电子工厂等十大实景场景，面积3000平方米。二期⁴⁸搭建了工业智造、智慧家庭、康养服务和5G融合四大类共16个细分场景，面积6000平方米。天津帕西尼具身智能超级数据工厂⁴⁹面积更高达12000平方米，构建了15+N个全场景矩阵，涵盖制造、

⁴⁷ <https://www.shougang.com.cn/m1/mtgz/20250329/137.html>

⁴⁸ <https://mp.weixin.qq.com/s/eulhn3CAgK3w4TNcUy-nAQ>

⁴⁹ <https://finance.sina.com.cn/stock/t/2025-06-24/doc-infcefek9048758.shtml?froms=ggmp>

家庭、餐饮、商超、康养等全链条具身智能应用环境。

二是支持多品牌、多构型本体入驻。由于具身智能的产品形态多样，训练场如果只支持单一构型机器人，将难以满足行业应用对数据的多样化需求。通过异构机器人协同训练生成规模化数据集，将最大化发挥训练场的场景价值。上海张江异构人形机器人⁵⁰训练场部署了来自智元新创、傅利叶、开普勒等10余家家人形机器人整机企业的100余台不同构型的机器人，并集成了遥操作、动作捕捉等多元数据采集方式，形成了“异构群智”的训练模式。北京建筑大学科技园通用智能机器人训练场⁵¹可以支持人形机器人、“四足机器人+机械臂”“轮式机器人+机械臂”等不同构型，训练机器人自主导航、越障巡检及行业技能训练。

三是建设主体基本清晰，正在探索数据产品的商业模式。当前训练场的建设主体主要有两大类，一是政府主导下的单企业或企业联合体，如湖北人形机器人创新中心⁵²由武汉市政府相关部门指导，北武院、光谷华汇、光谷东智等单位共建。北京人形机器人数据训练中心由石景山区政府指导，一期⁵³由睿尔曼机器人建设，二期⁵⁴由京石科创集团、国内人形机器人领军企业乐聚机器人联合运营。二是机器人企业主导自建并运营，如天津帕西尼具身智能超级数据工厂由帕西尼

⁵⁰ https://www.pudong.gov.cn/rmt_pdxw/20250122/800516.html

⁵¹ <http://www.fangchan.com/news/132/2025-04-30/7323222307324105487.html>

⁵² https://jxt.hubei.gov.cn/bmdt/rdjj/202506/t20250605_5680578.shtml

⁵³ https://kw.beijing.gov.cn/xwdt/kcyx/xwdtyqqy/202508/t20250818_4175280.html

⁵⁴ https://kw.beijing.gov.cn/xwdt/kcyx/xwdtshgg/202509/t20250926_4211351.html

科技主导建设，上海浦东智元机器人数据采集工厂⁵⁵由智元机器人主导建设。从商业模式上看，目前训练场产出的数据产品主要供建设方或联建方自用，用于机器人的训练或开源生态的构建。部分训练场正在探索数据交易模式，如北京人形机器人数据训练中心的数据产品实现了向美国 OpenAI、Google 等公司的销售⁵⁶，湖北人形创新中心的数据产品与智元达成了交易⁵⁷。

训练场的实际效用仍有待进一步验证，距离真正解决产业问题仍面临多重挑战。一是训练场对真实场景的覆盖广度有限，与实际业务需求的匹配度仍有待考量。实际场景的业务环节多且任务差异性大，半数以上训练场披露已覆盖制造、医疗、零售、服务等场景，但采集数据是否能匹配实际业务需求有待考量。二是跨训练场数据难以实现互通互用。各训练场数据采集、标注方式与部署本体构型不统一，数据训练的模型效果与本体构型高度相关，跨训练场数据难以整合和共享。三是训练场数据产品是否能形成可持续的商业模式尚不明确。具身智能数据方案已经成为企业核心竞争力，数据构建链路中的方法、流程、动作标准等是企业最重要的经验沉淀。几乎每个具身智能创新企业都有自己的数据采集工厂。因此训练场数据产品是否有买方持续买单尚未可知。亟须通过标准化和第三方评估，规范化训练场建设与

⁵⁵ <https://www.zhiyuan-robot.com/products/AIDEA/DaaS>

⁵⁶ <http://www.cb.com.cn/index/show/zj/cv/cv135319741264>

⁵⁷ https://mp.weixin.qq.com/s/8aYzHFniMzcYBX-zxo1_ug

运营，客观评价训练场与专用场景和具体业务环节的契合度，推动数据互通共享，形成“评估达标—市场认可—商业盈利—持续投入提升能力”的正向循环。

（三）标准体系有序建设，正在针对性解决产业亟需问题

由于具身智能尚处于产业发展初期，其标准体系的建设既需要能满足当前发展阶段产业对标准和规范的需求，又要能够为未来产业进一步发展预留空间。通过科学、高效、准确的标准化研究，可以有效推动产业技术升级和规模应用。从目前产业的发展现状看，主要存在三方面问题亟须通过标准化解决。

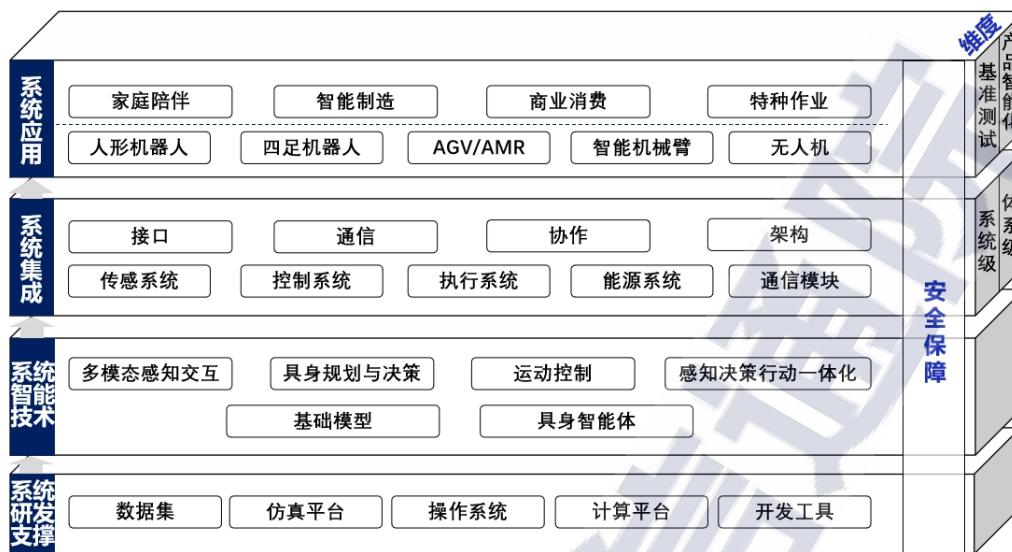
一是保障要素无法共享。由于目前标准尚未统一，不同企业使用的遥操作设备、仿真平台、机器人数据采集设备、数据集等存在严重的割裂，资产无法互通，重复开发和二次开发现象普遍，形成了创新资源的浪费。尤其是数据集方面，由于目前的具身智能大模型普遍使用数据作为智能升级和功能泛化的“原料”，造成了业界对数据需求的激增。不同训练场采集数据的结构、质量均不同，导致数据“重复采、不好用、难共享”。

二是智能化评价体系未统一。如何有效评价具身智能产品的智能水平，是否可以满足使用需求，是业界目前关注的热点。当前针对软件算法、软硬适配的技术实现路径尚在多点探索，有效的标准化评估可以推动智能水平和实际任务表现能力快速迭代升级。但由于未形成

统一标准，多种评测维度也在并行探索，导致相关标准体系过多，不同企业和机构采用不同框架，难以形成行业共识。且现有分级大多是概念性描述，缺少可操作、可量化的评价指标，实操性较低。需要聚焦实际应用场景，构建一套具身智能的基准测试方法（EAI Bench），形成产业公认的能力标尺，并通过“研-用-产-评”协同，形成产品和测试的循环迭代、互相促进。

三是安全与伦理标准缺失。具身智能安全既需要考虑软件和硬件的各自安全，也要考虑两者融合之后的新型安全问题，同时也包括人、机、物和物体环境复杂协同过程中的安全风险。当前，具身智能领域的安全治理标准体系存在明显空白，亟需构建涵盖本体安全、交互安全和应用安全等完整标准体系。明确“人在环”的协作机制，构建人机交互安全规范，保障具身智能与人直接交互的安全性。通过系统安全设计规范等，强化数据隐私保护和网络安全能力。同时，根据具体的应用场景、功能需求和防护要求，分类分级构建具身智能安全等级保护标准和评测标准，推动国际协同的安全准则互操作和互认，为具身智能应用和可持续发展提供安全保障。

为加快具身智能领域标准化建设，2024年工业和信息化部建立了人工智能标准化技术委员会，下设具身智能工作组（WG6），从系统研发支撑、智能系统技术、系统集成和系统应用四方面统筹推进国内具身智能产业标准体系建设工作，具体体系如图9所示。



来源：MIIT TC1 WG6

图 9 具身智能标准体系

目前工作组正在加速开展具身智能系统智能化分类分级总则、基准测试方法、数据质量要求及评价方法、具身智能训练场能力评估指南、电驱动一体化关节要求及试验方法、灵巧手技术规范等多项产业亟需标准的研制工作。后续也需要产学研用各界通力合作，及时识别产业亟需解决的标准问题，并加快推动标准的研制工作，通过标准领航，实现产业的健康有序发展。

（四）安全问题引发关注，成为规模化落地限制要素

具身智能核心依赖的人工智能技术具有不可解释性，在场景变化或执行泛化任务时存在一定的失误甚至失控概率。同时，由于具身智能产品未来将会在工业、服务业等领域广泛应用，与人类和周围环境的交互极为紧密，任何微小的问题都可能对外界造成严重损伤。2025

年3月的一次展会上，已在GXO物流仓库稳定工作并完成10000个订单履约的Agility Robotics公司Digit机器人在展区示范作业时突然摔倒，引发了行业对安全的热议。此外，具身智能所搭载的传感器也在实时采集周围的信息，给用户带来了极大的隐私泄露隐患。总体上看，具身智能在广泛应用后将面临系统性的安全风险，但目前国内在外在该方面尚处于探索阶段，未出台相关标准和国家层面的规范文件。从目前产品的技术组成和应用模式上看，具身智能的安全问题可以分为模型安全、本体安全和信息安全三部分。

模型安全是具身智能系统。模型安全不仅关乎个体任务执行的可靠性，更直接牵涉物理环境中的行为安全、社会伦理风险及公共信任体系。2024年7月，华中科技大学提出BADROBOT攻击范式，通过操控LLM漏洞、语言与行动安全错位及知识缺陷诱导危险行为，首次实现物理场景下具身系统的越狱攻击。2024年10月，宾夕法尼亚大学指出AI驱动机器人中存在的关键漏洞，成功在宇树GO2四足机器人等共三类机器人系统上实现“越狱”，操控机器人执行危险任务。这些攻击手段的出现，揭示了当前具身智能模型的安全问题。模型安全主要包括对抗攻击防御、反常指令规避、跨场景性能衰减等内容。**对抗攻击防御**是模型对外界数据投毒、对抗样本注入或恶意模型分发等攻击的核心防御手段。**反常指令规避**是模型在交互时的主动安全防御方式，防止被诱导执行违反伦理道德或法律的指令。**跨场景性**

能衰减是具身智能产品在正常作业时需要直面的安全问题，由于环境动态变化可能导致模型感知与决策失效，需通过多模态传感器融合与自适应校准算法动态补偿环境干扰，同时结合增量学习持续优化模型的环境适配。

本体安全是产品融入人类工作生活的前提条件。本体安全的核心和底线是保护用户不受到物理伤害，可以重点考虑紧急制动、倾倒控制、安全距离保持、碰撞力阈值四个方面。**紧急制动方面**，传统直接断电的急停机制可能导致机器人因瞬间动力丧失而失控倾倒，甚至引发二次伤害，因此需设计渐进式卸力制动系统。**倾倒控制方面**，则需应对动态失衡的复杂挑战。当预测模型通过融合重心位置、运动加速度及 IMU 数据判定高风险跌倒不可避免时，机器人需在毫秒级内执行避险动作序列，降低结构性损伤风险。**安全距离保持方面**，需依赖多模态感知融合构建环境动态地图，结合行为规划算法实时计算与人员、障碍物的最小交互距离。**碰撞力阈值方面**，主要聚焦于瞬时冲击的主动缓解。在硬件层面，可采用轻质复合材料外壳与内置缓冲层分散撞击能量；在控制层面，通过电机主动反向补偿技术，在碰撞发生的毫秒级窗口内生成反向扭矩抵消冲击力。此外，本体安全还涉及机械安全、电气安全、电池安全等方面，但相关安全体系均较为成熟，此处不再展开。

信息安全是保障用户隐私的关键手段。具身智能产品集成多模态

传感器，如摄像头、麦克风、激光雷达等，若安全机制失效，可能被恶意操控为“移动间谍”，成为窃取个人隐私与商业机密的“超级窃听器”与“窃看器”。在个人隐私层面，可能违法收集家庭私密谈话、电话通信内容等敏感语音信息，非法获取人脸、指纹等生物特征数据，甚至通过视觉识别分析家庭成员行为习惯、财产状况、住宅布局等私密信息。在商业领域，则可能窃取生产工艺流程细节及生产环境布局等关键机密，对企业的核心竞争力造成毁灭性打击。由于具身智能产品的安全涉及软件、硬件、网络等多个层面，需构建纵深防御的信息安全机制体系，包括传感器权限分级控制、硬件级加密与防篡改、实时数据脱敏处理、加密通信协议、网络隔离与访问控制等方面。通过“端侧硬件加密奠基、实时脱敏控源、通信加密防泄、网络隔离阻侵”的多层协同，最大程度将具身智能产品的“感知”能力约束于安全可信的边界之内。

五、具身智能产业展望

虽然具身智能现在尚未完全成熟，在技术路线、本体形态等方面还存在多方面的争议和问题，成熟的应用场景也未真正出现，具身智能产业的火热与落地应用缺失之间的矛盾有可能会引发新一轮泡沫。但具身智能作为走向通用人工智能的重要路径，对社会生产范式和人民生活方式必将产生颠覆性的影响。随着技术和产业沿着客观且科学的周期不断向前发展，具身智能也将会迎来更加广阔前景，真正融

入我们生产生活的各个角落，实现广泛场景服务人类、复杂场景辅助人类、危险场景替代人类。未来具身智能的发展将重点聚焦以下几个方面。

（一）技术架构重构：从“功能模块堆叠”迈向“多模态认知融合”

目前的具身智能模型正在从感知决策和运动控制分离走向全栈式多模态融合，而未来为了真正实现“感知—决策—认知—行动”的全链路贯通，将会有更多的技术实现融合创新。认知层面，探索模型架构创新，突破视觉、触觉、语言、运动信息等多模态范式深度融合，构建时空对齐的多模态融合表征。基于世界模型发展具身智能认知引擎，通过整合物理属性、运动规律、时序信息等，实现物理规律内化。探索与脑科学、心理学等多学科融合创新，构建价值驱动的具身智能技术体系和认知架构。学习层面，探索类似人类的场景记忆和学习能力，能主动探索、反复试错，在与真实环境交互中主动获取经验，并持续进化。同时，随着具身基础模型的通用能力增强，模型将具备更强的跨场景适应和跨本体迁移能力，通过小时级的实训学习即能上岗作业。协作层面，构建软件和硬件间的通信协议与协作机制，通过分层协作控制的具身智能体，实现模型与本体的无缝对接，驱动本体在真实环境中的快速部署。同时提升多机任务分配、资源共享和协作能

力，探索人机协作新模式，使具身智能能够理解人类意图，更好地与人类共同完成任务。

（二）应用场景深化：从“演示”迈向“实用”

要从实验室的“能动”走向实际应用的“好用”，具身智能仍有很长的路要走。未来，通过与物理环境实时交互并从中学习的能力具身智能将以场景任务需求为导向，深刻赋能千行百业。**一是突破一批**，在工业制造、物流仓储、医疗康养和商业服务等场景，具身智能将成为生产工具，提升生产效率与质量，推动行业向智能化、柔性化转型。**二是量产一批**，生产工具规模化升级带来生产力提升，将重构生产关系，使得人机协作更加紧密，生产过程中的决策与执行更加高效协同。**三是普及一批**，在广泛而深入的场景探索下，有望应用在国民经济行业分类中的所有20个门类，并在部分典型的垂直应用领域中实现深度发展，孕育形成数个万亿级的市场，呈现“探索底层通用加速垂直场景规模化应用”的“梳子型”产业结构。

（三）安全伦理构建：从“合规”到前瞻布局“伦理协同框架”

随着具身智能从实验室的受控“演示”迈向社会的广泛“实用”，其安全与伦理问题已经从学术议题演变为产业关切重点。未来，需要把握具身智能“人-机-物”深度融合的关键特征，建立多维度、系统

性的具身智能安全防护和治理体系。**一是从安全被动响应模式向主动预防模式的转变。**除了保障机械安全、网络安全和数据安全等信息域和物理域的安全问题，未来，具身智能系统研发将致力于增强具身智能内生安全，将安全与伦理准则和人类价值内化为系统的固有属性，发展策略监测与自诊断技术，实现对系统内在机理的反向推断与监测。**二是探索动态、敏捷的治理模式。**未来将通过建立全生命周期管理制度、伦理审查制度以及就业结构调整等措施，构建覆盖技术研发、产品认证、场景部署与失效追溯的全生命周期治理体系，确保具身智能产品能够真正服务于人类，保障技术产业的健康向上发展。**三是形成人机协同与公众参与的社会共识。**具身智能应用将对社会各层面产生复杂而深远的影响，未来需进一步明确具身智能产品作为生产生活工具的社会属性，研究并制定以人为本的具身智能系统设计规范、伦理准则和法律法规，确保其与真实社会需求对齐，与人类价值观保持一致，真正辅助于人，服务于人，让人类继续承担高价值任务，为产业的长远发展保驾护航。

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-68094709

传真：010-62304980

网址：www.caict.ac.cn

