

具身智能机器人技术演进与未来展望

徐俊^{1,2} 郭喆晨^{1,2} 王凯龙^{1,2} 杨天^{1,2} 丁志壮^{1,2} 梅雪松^{1,2}

(1.西安交通大学机器人研究院 西安 710049;2.陕西省智能机器人重点实验室 西安 710049)

摘要: 具身智能机器人作为人工智能(AI)与机器人技术深度融合的前沿方向,已成为发展新质生产力和“人工智能+”行动的重要抓手,是实现通用智能的关键载体。本文系统梳理了机器人近代发展历程与人工智能符号主义、行为主义、连接主义三大学派技术演进脉络。归纳了国内外代表性学者对具身智能的定义与核心内涵,提出了涵盖数字孪生(Digital Twin)、现实世界(Reality)、互动(Engagement)、感知(Awareness)、执行(Motion)五大核心要素的具身智能机器人DREAM技术框架,并总结其物理载体依赖、孪生映射、自成长三大核心特征。深入剖析了当前具身智能的主流技术路线,包括一体化端到端模型路线、分层式模型路线、世界模型路线。分类阐述具身智能机器人的载体形式,涵盖了固定底座式、轮式、足式、飞行式及水下、空间、软体等特殊类型。总结了具身智能机器人的分级标准,结合当前具身智能机器人在工业制造、家庭服务、特种环境等应用场景以及面临的挑战,对具身智能机器人的未来发展趋势进行展望,为具身智能机器人研究提供参考。

关键词: 具身智能; 机器人; 人工智能; 大模型; DREAM架构

中图分类号: TG156 **文献标志码:** AA

DOI: 10.7652/xjtuxb202208000 **文章编号:** 0253-987X(XXXX)XX-0001-29

Research Progress on Embodied Intelligent Robots

XU Jun^{1,2}, GUO Zhechen^{1,2}, WANG Kailong^{1,2}, YANG Tian^{1,2}, DING Zhizhuang^{1,2}, MEI Xue-song^{1,2}

(1. Institute of Robotics, Xi'an Jiaotong University, Xi'an, 710049;

2. Shaanxi Key Laboratory of Intelligent Robots, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, 710049)

Abstract: As a cutting-edge field representing the deep integration of artificial intelligence (AI) and robotics, embodied AI robots have become a key driver for developing new forms of productive capacity and advancing the 'AI+' initiative, and serve as a crucial vehicle for achieving general intelligence. This paper systematically reviews the modern history of robotics and traces the technological evolution of the three major schools of thought in artificial intelligence: symbolism, behaviorism and connectionism. It summarizes the definitions and core concepts of embodied AI proposed by leading scholars both domestically and internationally, and proposes the DREAM technical framework for embodied intelligent robots, which encompasses five core elements: Digital Twin,

收稿日期:2026-06-XX。

Reality, Engagement, Awareness, and Motion. It also summarizes its three core characteristics: physical carrier dependency, twin mapping, and self-growth. It provides an in-depth analysis of the current mainstream technical approaches in embodied AI, including the integrated end-to-end model approach, the hierarchical model approach, and the world model approach. It categorizes and describes the carrier forms of embodied AI robots, covering fixed-base, wheeled, legged, flying, and special types such as underwater, space-based, and soft robots. It summarizes the classification criteria for embodied AI robots. By examining their current applications in industrial manufacturing, domestic services and specialized environments, alongside the challenges they face, the paper outlines future development trends for embodied AI robots, thereby providing a reference for future research in this field.

Keywords: embodied AI; robot; artificial intelligence; large model; DREAM architecture

具身智能作为新一轮科技革命与产业变革的战略性技术方向，是培育新质生产力的关键力量，正深刻重塑生产生活的各领域图景。

国家已从战略全局高度对具身智能产业作出系统的政策部署。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；2026年，“十五五”规划纲要将具身智能纳入国家未来产业重点战略布局，与量子科技、生物制造等并列为六大未来产业之一，推动其成为新的增长点。连续两年政府工作报告对具身智能产业发展作出专项规划。

与此同时，相关部门构建技术攻关、产业培育、场景应用、行业标准规范等政策支撑体系，各地方同步推进落地示范，统筹推进、协同发力。

在人才培育支持方面，2026年4月，教育部正式将“具身智能”列入《普通高等学校本科专业目录（2026年）》，在“交叉学科”门类下首次增设具身智能本科专业，西安交通大学等9所高校获批首批建设，为产业可持续

发展提供坚实的人才支撑。

在此国家重大战略需求背景下，系统梳理具身智能机器人的技术演进脉络、厘清核心技术架构并研判未来发展趋势，对于抢占新一轮科技革命制高点具有重要的理论价值与现实意义。

具身智能的核心思想源于具身认知理论，认为认知是具身且嵌入环境的，大脑、身体和环境构成一个动态的认知系统^[1]。具身智能强调智能是在物理躯体与环境的动态交互、实时反馈、自主学习与适应中涌现的，通过“感知-决策-规划-执行-反馈修正”回路，对物理世界的感知以及行为来验证和调整对环境的理解，更新修正内部认知模型。具身智能为通用人工智能（AI）的最终实现，提供了一条极具潜力的可行路径，其核心载体是具身智能机器人，即具备物理身体，在身体与真实环境相互作用过程中，自主感知、理解、决策并执行任务的独立智能体。

近年来，大型语言模型（LLMs）、视觉-语言模型（VLMs）、多模态大型语言模型（MLLMs）、视觉-语言-动作

模型 (VLAs) 等多类大模型技术的快速迭代,为具身智能机器人注入了新的发展动能,推动其在跨模态感知、复杂场景推理、自然语言交互等关键能力上实现突破性进展。多款代表性具身智能机器人产品相继发布,正逐步从实验室技术验证向实际应用场景渗透。

当前国内外已发表的相关综述性研究多侧重于单一技术维度或局部应用场景,例如大模型研究综述^[2]、人形机器人技术现状^[3]、具身智能机器人导航方法^[4]、特定场景下的应用^[5]等,尚未形成对“机器人技术演进及人工智能历史沿革-具身智能定义与典型技术路线-具身智能机器人载体形式、分级标准及应用场景-现存挑战与未来展望”的全链

条、系统性梳理,难以全面呈现具身智能机器人领域的整体发展脉络与研究全貌。

本文梳理机器人的近代发展史,回顾人工智能发展的三次浪潮,归纳代表性学者对具身智能的定义,提出具身智能机器人 DREAM 技术框架,剖析了当前具身智能的主流技术路线,阐述具身智能机器人的载体形式与分级标准,介绍了具身智能机器人的应用场景,最后总结当前具身智能机器人面临的挑战,展望未来研究方向与发展趋势。

1 具身智能是机器人与人工智能发展交汇点



图1 机器人发展时间线

Fig. 1 Development history of robotics

机器人与人工智能历经长期迭代演进,技术体系日趋完善,应用边界持续拓宽。机器人作为物理执行载体,已从传统自动化装备逐步向具备感知与决策

能力的智能载体升级;人工智能则从理论探索迈入规模化落地阶段。二者协同发展、互促共进,机器人为人工智能提供物理实体,人工智能为机器人赋予

“大脑”，共同推动通用智能的实现。本节对机器人与人工智能的发展历程进行梳理，明晰技术演化脉络。

1.1 机器人近代发展史

现代机器人起源于工业自动化需求，发展时间线如图1所示，历经示教再现、传感自适应与多场景扩展、运动能力突破与智能化升级三个阶段。如表1所示，整体趋势从自动化到智能化，从单一功能到多任务集成，从结构化工厂的工业生产工具，迈向非结构化的、复杂的人类生活和工作环境，具备自主交互能力的智能体。

示教再现阶段以预编程复现动作为核心，Unimate（尤尼梅特）与VERSTRAN机器人共同成为早期工业机器人标杆^[7]。这一时期的机器人按事

先程序设定的步骤执行，复现人类的示教动作，特点是对外界的环境没有感知，以固定机械臂为主，仅能模拟运动功能，在预设的标准化场景中完成单一任务，初步替代了重复性的人力劳动。

20世纪60年代中期开始，随着计算机技术与传感器技术的发展，机器人开始突破纯预编程的局限，进入传感自适应与多场景扩展阶段。这一阶段的机器人特征为具备感知、对变化可作出反应，从被动执行转向主动感知，能够根据应用场景从工业领域视觉、力觉等传感器反馈，对环境变化做出初步响应，应用场景从工业领域延伸至服务、医疗、航天、家庭等多个领域，拓展了应用边界。但此时的机器人仍以专用任务为主，缺乏自主学习与泛化能力，无法适应开放、动态的非结构化环境。

表1 机器人主要发展阶段

Table 1 Key stages in the development of robotics

阶段	预编程/示教再现阶段	传感自适应与多场景扩展阶段	运动能力突破与智能化升级阶段
表现	· 按事先程序设定步骤执行或单纯复现人类的示教动作 · 模拟运动功能，重复执行特定动作	· 计算机与传感器技术推动其具备初步环境感知能力 · 根据传感器反馈实时调整动作的自适应控制能力	· 互联网、计算能力提升及人工智能技术发展，具备更强的环境理解、决策、规划和自然交互能力 · 人形、协作机器人兴起
特点	初步替代重复性人力劳动	拥有“感觉”，对变化作出初步反应	智能化、自主化、多样化，改变生产生活方式

2010年后，高性能伺服电机、液压驱动系统以及运动控制算法的突破，使得机器人在关节灵活性、动态平衡与运动稳定性接近生物级水平。波士顿动

力Atlas、Spot等机器人展现出优异的动态运动与地形适应能力。2020年以来，大语言模型、多模态大模型技术的突破性进展，推动机器人与人工智能深

度融合,具身智能成为机器人领域的核心发展方向^[6]。在这一浪潮下,国内外多款人形机器人相继面世。这一阶段的核心特征在于,机器人正从执行算法的机械躯体,转变为通过物理交互自主学习的智能体,其智能不再局限于固定程序的重复执行,而是通过物理载体与真实环境的持续交互、试错与反馈逐步涌现,从而形成开放环境下的通用感知、自主决策与泛化执行能力。

1.2 人工智能历史沿革

人工智能学科自诞生以来,历经七十余年的迭代演进,先后经历三次发展浪潮与两次研究寒冬。在这一演进过程中,符号主义、行为主义与连接主义三大主流学派逐步形成,它们从不同视角探究智能的本质、探索技术发展路径,共同推动人工智能学科的整体发展,如表2所示。

学科起源与早期探索 (1943-1956): M-P 神经元模型、控制论、图灵测试等理论相继诞生,为学科发展奠定基础。1956 年达特茅斯会议正式提出人工智能概念,标志该学科正式诞生^[7]。

第一次浪潮与第一次寒冬 (1956-1980): 符号主义占据主导,专家系统、早期自然语言处理系统等成果不断涌现。符号主义依靠符号、规则与逻辑推理实现知识表征与问题求解,优势是可解释性强,但高度依赖人工规则,难以适配复杂开放环境,叠加早期算力不

足,推理效率极低^[8],其核心缺陷被概括为“只懂规则,不懂世界”。1974 年起,各国政府与企业大幅缩减相关研发投入,人工智能学科由此陷入长达六年的第一次研究寒冬。

第二次浪潮与第二次寒冬 (1980-1997): 进入 20 世纪 80 年代,人工智能迎来第二次发展浪潮。符号主义依托商用专家系统短暂复苏,却因规则维护难度大、迁移性差再次陷入困境;布鲁克斯提出“行为主义智能”的概念,主张智能无需复杂的内部表征,仅需构建“感知-行动”的直接闭环即可实现^[9]。《强化学习导论》著作发表,标志着行为主义从分散的工程实践跃升为完整的理论体系^[10];连接主义在这一时期完成了关键的技术积累。Hopfield 网络的提出使得神经网络重新受到人们的关注^[11],反向传播 (BP) 算法推动神经网络研究迎来第二次兴起^[12]。此后,神经网络模型不断演化发展和创新。但受限于当时的数据规模不足与资金支持匮乏的双重制约,理论虽持续发展,却缺乏充足的实践应用与数据集来支撑模型的训练与验证,神经网络研究陷入短暂停滞,人工智能研究陷入第二次寒冬。

第三次浪潮 (1997-今): 人工智能进入第三次发展浪潮,以神经网络为核心的连接主义逐渐成为学科发展的主要方向。深度学习^[13]、Transformer 架构^[14]相继诞生,推动计算机视觉、自然语言处理快速发展;2018 年起,算力、数据与算法的融合突破使得训练超大规模

神经网络成为可能。大语言模型、多模态模型等成果持续迭代，进一步打通了文本、图像、语音等跨模态感知与生成能力。

表2 三大学派总结

Table 2 Summary of the three major schools

学派	符号主义	行为主义	连接主义
起源	逻辑学、哲学	控制论	神经科学
主要观点	认为智能产生于大脑抽象思维的过程中	认为智能源于环境反馈，复杂行为可分解为若干个简单行为	认为智能产生于大脑神经元之间的相互作用及信息往来的过程中
核心思想	使用符号、规则和逻辑来表征知识和进行逻辑推理	环境刺激的应变，通过接收环境的奖励或惩罚信号优化自身行为策略	通过权重矩阵与加权神经元的动态协同，实现对数据模式的识别与归纳
代表算法	逻辑推理、专家系统	进化算法、强化学习	神经网络、深度学习

纵观符号主义、行为主义、连接主义三大学派，它们分别从逻辑推理、环境交互与仿生神经网络三个不同视角探究智能的本质，并在不同历史时期取得了具有里程碑意义的实践成果^[15]。符号主义率先开启了人工智能研究的大门，如今已不再占据主流，但其理念并未消失，而是作为底层逻辑的基石，融入到当代人工智能系统中。行为主义则凭借跨领域融合和渐进式渗透的策略，将自身价值聚焦于真实环境的感知与响应，弥补了符号主义与连接主义在处理复杂动态环境下的不足。连接主义历经算法革新（如反向传播）、硬件革命（如GPU）与架构突破（如Transformer）的三重跃迁^[16]，最终凭借深度学习与大模型的突破，成为当下人工智能主流范式。

尽管连接主义的发展仍处于高潮阶段，但神经网络技术在实际应用中仍面临诸多突出问题，如“黑箱”不可解释性、模型幻觉、能耗过高以及泛化能力

不足等突出问题。通用人工智能的困境在于如何在复杂、动态、未知的真实环境中发挥效用，具身智能为当前的困境提供了一条可能的技术路径。

1.3 具身智能：机器人与人工智能发展的汇聚点

人工智能与机器人技术的发展始终呈现双向赋能、协同演进的良好互动格局。人工智能为机器人赋予了“大脑”，依托多模态感知、逻辑推理与自主决策等核心能力的突破，推动机器人逐步具备环境感知、动态思考与精准决策的高阶能力；而机器人则为人工智能提供了具象化的物理“身体”，使其通过与真实物理世界的直接交互、反馈，积累知识、沉淀经验，实现智能水平的迭代升级，具身智能发展路线如图2所示。

具身智能的兴起并非偶然的技术现象，而是人工智能与机器人技术各自发展至成熟阶段后的必然产物，是两者深

度融合、协同突破的核心具象体现。从发展逻辑来看,人工智能的持续迭代亟需依托与物理世界的交互来突破现有技术瓶颈、完善智能体系,而机器人的未

来发展也离不开更高级别的智能化水平作为核心支撑,二者的深度融合既是技术发展的内在需求,也是推动具身智能机器人领域持续进步的核心动力。

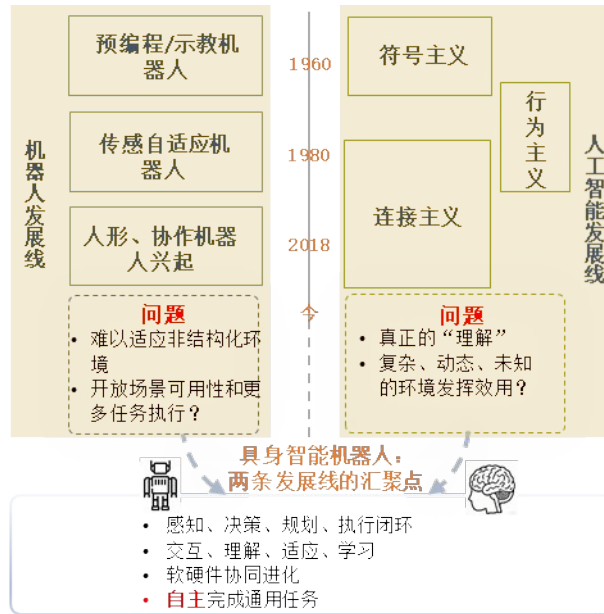


图2 具身智能发展历程

Fig. 2 Development history of embodied AI

2 具身智能的概念及主流技术路线

近年来,具身智能作为人工智能、机器人与认知科学深度交叉的领域,受到了学术界、产业界的广泛关注,被认为是“AI的下一个浪潮”。国内外学者围绕具身智能的定义与核心内涵从不同视角展开了深入探讨,涌现出多种代表性定义。与此同时,随着大语言模型、多模态感知、世界模型等技术的飞速发展,具身智能的技术路线也呈现出多元

化、交叉融合的姿态。

2.1 具身智能的概念

1) 具身的内涵

“具身”是理解具身智能的哲学基础与理论起点。受德国哲学家海德格尔“存在”(Being-in-the-world)概念的启发,具身认知理论主张将认知置于大脑、大脑置于身体、身体置于环境,构建一种具身-嵌入的认知研究模式。法国身体现象学的代表人物梅洛-庞蒂进

一步指出,身体是知觉的承载主体,嵌入世界之中,如同心脏嵌入身体,知觉、身体与世界构成统一体。人正是通过身体与世界的持续互动、作用而产生知觉、认识世界^[1]。认知语言学家莱可夫等人扩展了梅洛-庞蒂的思想,从语言学角度论证了思想范畴与概念隐喻均根植于身体的感觉运动经验,系统阐明了心智与身体的一体化特征。

基于上述哲学与认知科学理论,提炼具身的基本涵义:认知不仅是基于身体的,也是根植于环境的,身体与环境的互动“建构”出了认知^[17]。具身强调心智、认知根植于身体,并且由躯体物理特性、感知运动系统及环境交互行为共同塑造。

具身是贯通哲学、认知心理学、人工智能等多学科交叉的概念,其核心内核始终凸显躯体与环境的整体性,以及动态交互对认知与智能涌现的决定性作用。落脚到智能系统层面,具身的本质在于:智能依托实体物理躯体,在持续物理交互中完成环境认知的生成迭代,强调“身体”与“环境”的不可分割性,以及“交互”在生成意义和智能中的决定性作用。

2) 国内外学者对具身智能的定义

西安交通大学郑南宁院士指出,具身智能通过构建具有本体感知与行动能力的智能体,利用多模态传感器实时捕获环境状态,利用执行机构施加物理作用,并在连续时空维度中形成“感知-认知-决策-行动”的闭环学习系统,从

而实现对非确定性环境的动态建模与策略优化^[18]。其核心观点是,智能体通过感知环境并采取适当的行动来解决问题。

美国国家工程院院士李飞飞提出,“具身”的含义并非单纯指代物理实体,而是与环境交互以及在环境中执行的整体需求和功能^[19]。在2026斯坦福HAI讲座中李飞飞提出,空间智能是AI理解三维物理世界、掌握物理规则、感知物体关系、规划空间路径、并在真实环境中安全行动与交互的能力,其认为具身智能需要通过世界模型解锁空间智能,实现感知-想象-行动的闭环。

清华大学刘华平教授认为,具身智能更侧重关注“交互”,即智能受脑、身体与环境协同影响,并由身体与环境相互作用中,通过信息感知与物理操作过程连续、动态地产生^[20]。他提出基于形态的具身智能的体系架构,综合分析了形态与行为、学习之间的关系。

上海交通大学卢策吾教授指出,具身智能是指一种基于物理身体进行感知和行动的智能系统,其通过智能体与环境的交互获取信息、理解问题、做出决策并实现行动,从而产生智能行为和适应性。在具身智能的技术体系探索中,卢策吾教授在多次学术演讲中系统阐释了具身智能的PIE方案,该方案系统整合了具身感知(Perception)、具身想象(Imagination)与具身执行(Execution)三大核心模块,三者关系是具身感知是行为依据,具身想象是行为决策的过

程,具身执行是行为执行的过程。

中国科学院大学特聘教授蒋树强提出,具身智能是指通过物理智能本体与外界环境的互动来实现智能的理论与技术研究,相比于静态、离身的人工智能,具身智能具有涉身性、情境性、主动性和交互性等特点^[21]。

安徽大学孙长银教授提出ABCDE框架来理解具身智能,即通过认知引擎(AI)、物理载体(Body)、控制枢纽(Control)、决策协同(Decision)驱动具身智能(Embodied AI),以多维协同机制解决传统AI在物理交互、环境适应和群体进化中的局限^[22]。

国际标准《具身智能系统框架及能力要求》ITU-T F.748.66在ITU-T第21专业组(SG21)2025年第二次全体会议期间正式冻结,给出定义:具身智能指与物理实体融合的人工智能,能自主与物理世界交互并适应环境。

《中国人工智能学会系列白皮书■具身智能(2026版)》中提出,具身智能专注于智能体通过物理本体与外界环境的互动来实现智能的理论与技术研究,涵盖环境感知、记忆推理、对话交互、自主学习、决策规划、动作执行等综合性技术,从而在真实物理世界中展示出类人的智能行为。具身智能具有涉身性、情境性、主动性和交互性等特点。

中国信通院发布的《具身智能发展报告(2025年)》中提到,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。报告指出,基于身体交互的“智能闭环”是具身智能的核心内涵。

3) 本文关于具身智能的理解

综合国内外学者的定义,本文认为具身智能是基于物理实体进行感知和行动的智能系统,其通过物理实体与真实环境的持续交互,获取信息、理解问题、做出决策并生成行动,从而涌现出智能行为与自适应能力。本质上,具身智能也可视为AI的物化。

具身智能系统一般包含感知、决策、规划、控制与交互模块,如图3所示。感知模块通过多模态传感器获取动态环境与自身状态信息,并实现对象的识别和空间理解;决策模块基于任务目标与环境理解进行高层策略选择与任务拆解;规划模块生成无碰撞路径与满足动力学约束的运动轨迹;控制模块将规划结果转化为驱动指令,实现对机械臂、灵巧手、移动底盘等执行器的精准控制;交互模块则通过自然语言、手势、意图识别等方式实现人机以及多智能体之间的协同通信。

4) 具身智能机器人DREAM技术框架

基于以上对具身智能的理解,本文提出了具身智能机器人DREAM技术框架,如图5所示。DREAM框架是对具身智能机器人核心要素的概括:Digital twin(数字孪生):在虚拟空间中构建

与现实世界一一映射的孪生模型,完成任务仿真、动作预演与策略优化; Reality (现实世界): 具身智能体所处的真实物理环境与价值锚点,所有感知、决策与执行等活动均以现实物理世界的规则与任务需求为约束; Engagement (互动): 具身智能体与现实世界之间的双向交互过程,搭建虚实双向传输通道,涵盖环境感知、物理作用及反馈接收等,涉及到的交互类型包括: 智能体与环境的交互、人机界面(如指令输入)的交互、自然语言的交互以、多智能体间的交互等,是持续动态演进的; Awareness (感知): 通过多模态传感器对自身状态、环境信息及交互结果进行实时检测与理解,形成对现实世界的态势认知; Motion (执行/动作): 智能体基于规划与决策结果,经由物理载体产生的一系列可控运动与实体操作动作,是达成任务目标的最终输出环节。动作执行后环境产生新状态,再次经由感知、互动反馈更新数字孪生模型,五要素持续循环迭代,构成具身智能运行体系。

该框架五要素构成了以孪生(D)为全局中枢、现实(R)为边界锚定、交互(E)为虚实纽带、感知(A)为信息之源、执行(M)为落地实现的层次化结构,形成了感知为始、执行为终、交互贯通、孪生统筹的逻辑闭环,如图4所示。在真实物理世界端,存在若干个形态不限的具身智能体,而在虚拟空间端一侧,每个智能体拥有具身智

能“脑”,对物理空间进行映射,并具备记忆、仿真、预测、推演等功能。具身智能体通过多模态融合的主动感知获取环境信息,实时感知到的环境状态以及交互信息同时传递给“脑”,以构建模拟引擎,即基于数字孪生技术形成的物理世界虚拟映射模型。

在此基础上,系统通过任务策略分解、路径轨迹规划与仿真执行预演,验证策略可行性并生成动作序列指令,下发控制指令,驱动实体执行。具身智能体执行动作后,对环境产生交互作用,感知系统再次采集反馈信息并传递给“脑”,“脑”基于观察反馈以及任务目标,对执行行为进行校准修正,动态调整行为。同时,通过从反馈中学习、从交互中学习的机制,不断迭代优化内部模型与更新知识,实现能力的自主进化。

虚拟空间的推演与真实世界的执行形成双向闭环: 前者为后者提供先验知识、环境与任务理解、策略预演; 后者则为前者提供反馈数据,驱动持续学习与模型更新。感知-决策-规划-执行-交互反馈-学习-优化的迭代过程,构成了具身智能实现自适应行为的根本逻辑,体现了具身智能通过物理交互与虚实融合实现智能涌现的核心内涵。

另一方面, DREAM 框架取意于“梦”的生物智能预演机制,其借鉴了人类在梦境中预演现实行动的自然智能启示,即在数字孪生映射的虚拟空间中,对具身智能体的物理执行过程进行

全面仿真、推演与预测。该框架以数字孪生为梦境预演场，使智能体能够在虚拟环境中先行经历任务执行的全流程，验证策略可行性、规避潜在风险，并将优化后的指令下发给真实物理载体，最终实现“以虚驭实、先演后行”的闭环智能。

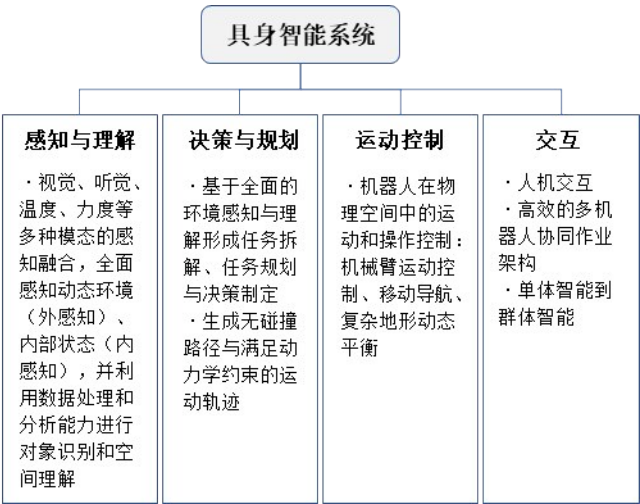


图3 具身智能系统中的感知、决策、规划、控制与交互模块

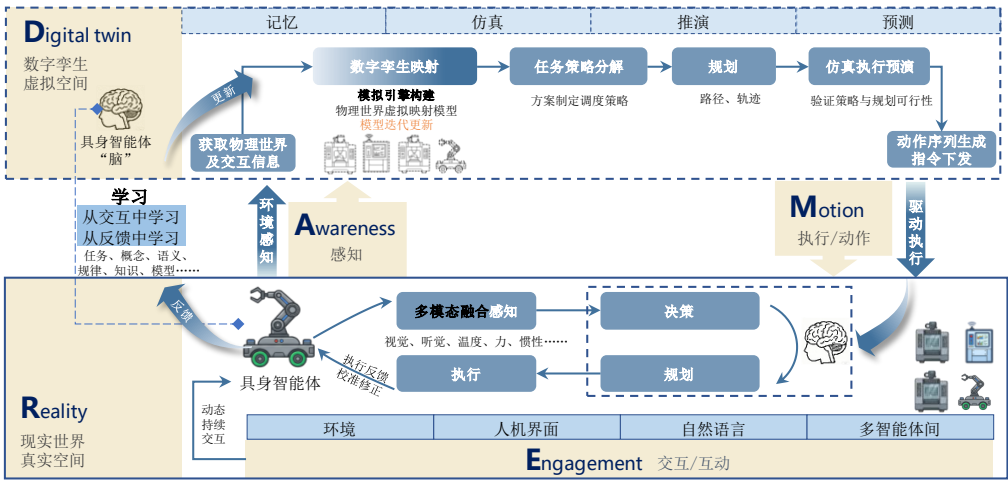


图4 具身智能机器人DREAM技术框架

Fig. 4 The DREAM framework of Embodied Intelligent robots

Fig. 3 Perception, decision-making, planning, control, and interaction modules in embodied intelligent systems

基于前述定义与技术框架，具身智能的核心特征可概况为以下三点：

①物理载体依赖

智能行为无法脱离物理实体而独立存在。具身智能体以物理实体为核心载体与真实环境进行持续交互,其感知、决策与执行能力均需通过物理实体得以实现。物理载体的形态结构、传感器配置与性能参数、运动控制能力,直接影响智能体的环境交互方式与所能达成的智能水平。

②孪生映射

具身智能系统在虚拟空间中构建与物理世界相对应的数字孪生映射模型,该模型具备记忆、仿真、预测与推演等核心能力。智能体基于孪生映射模型,完成任务策略分解、路径规划与仿真预演,在验证行动可行性后生成控制指令,为物理实体的执行提供先验知识与策略支撑;同时,真实世界的执行反馈持续驱动虚拟映射模型的迭代更新,从而形成虚实双向赋能、闭环协同的运行机制。

③自成长

具身智能系统具备持续学习与迭代优化的能力。智能体依据环境反馈与任务目标,自适应调整与校准行为、优化内部模型、更新知识库、从而实现能力的自主成长与持续进化。这一特性赋予具身智能体自主学习的能力,使其能够适应动态变化的环境与多样化的任务场景,逐步从特定封闭场景走向开放世界的泛化应用。

基于上述三种核心特征,具身智能

体以物理载体为依托、以孪生映射为技术框架、以自成长性为演进动力,构建起“感知-决策-规划-执行-交互反馈-学习-优化”的闭环智能体系。

2.2 具身智能的主流技术路线

当前具身智能领域呈现多技术路线并行探索的发展格局,学术界与产业界的主要技术路线可归纳为:一体化端到端模型路线、分层式模型路线、世界模型路线。

1) 端到端式路线

端到端模型路线由一个大模型包办所有,将感知、决策与执行统一于单一模型中,直接将多模态输入(视觉、语言、本体感知等)映射为动作输出,让机器人完成识别环境、分解任务、执行操作等任务^[23]。

该技术路线的典型代表是视觉-语言-动作(VLA)模型。2023年7月,谷歌DeepMind发布具身智能VLA模型RT-2(Robotics Transformer 2)^[24]。RT-2以预训练的视觉-语言模型为核心骨干,将机器人动作视为另一种语言,将机器人动作映射为文本标记并与互联网规模的视觉-语言数据集共同训练,通过动作Token化将机器人控制任务统一到自然语言生成框架中,形成端到端单一模型架构。

端到端模型路线是大小脑深度耦合的,它的映射方式不拆分中间环节,这种方式的优点是系统架构简洁,无需设计独立的感知、规划、控制模块,简化

了系统的层次结构^[25];依托海量数据驱动自适应环境变化,具备一定的场景泛化潜力。但模型的黑盒化显著,决策过程可解释性差,故障难以溯源,推理透明度不足;对数据需求以及依赖度高,训练需要更多来自物理世界即真实世界动态环境中的交互数据^[26],难度大,数据量和成本将会指数级增加,算力与真机采集成本高昂。

2) 分层式模型路线

分层式模型路线是将大小脑分区,解耦配合,通过不同层次模型协作,将指令拆解为高层系统与低层系统两个层级执行,高层系统负责任务理解、语义推理和策略规划,将复杂任务拆解为可执行的子任务序列;低层系统负责连续动作生成、运动控制和执行任务,分层协作,不同层级由多个神经网络分别训练优化再联合^[27]。

分层式模型路线的代表有 Figure AI 发布的 Helix 02 以及银河通用的三层级模型系统。Helix 02^[28]的架构中, System 2 (S2) 负责慢速推理目标:理解场景、解析语言、规划行为序列; System 1 (S1) 反应迅速,以 200 Hz 的频率将感知转化为全身关节目标,实时执行精细动作控制; System 0 (S0) 以 1 kHz 运行,底层基于 1,000+小时人类动作数据的学习模型,处理全身的平衡、接触和协调。银河通用提出三级大模型系统,顶层大模型,可以调度中间技能 API,来实现完整的从任务的感知、规划到执行的全流程;中间层通过仿真

合成数据形成的自主建图、自主导航、物体抓取、开门开抽拉开冰箱、移动操作、挂衣服叠衣服柔性物体操作的技能;底层是硬件层。

分层式模型路线各层级解耦,不同层级由多个神经网络分别训练优化再联合,无需纯端到端路径的海量长任务数据,降低了各模块的训练难度和数据需求,并提升了系统的可解释性,更易于工程化部署^[29]。但该种路线高层规划与低层执行之间的信息传递可能存在延迟或断层,影响实时性^[30];独立的各个模块间存在错误累计问题,且在跨多样任务泛化时表现不佳;同时,底层模型缺乏语义理解。

3) 世界模型路线

世界模型路线的核心思想是构建对物理世界的完整建模,通过预测未来状态来优化动作决策^[31]。世界模型强调对世界的理解,通过规划、推理、搜索和优化的过程,设计出一系列动作去完成任务、达成目标,与大语言模型自回归式逐 Token 预测的方式在底层架构上不同^[32]。

该路线的核心推动者与代表性成果包括:杨立昆提出基于联合嵌入预测架构 (JEPA) 架构的世界模型^[33],“世界”是一个可用于预测和规划的抽象状态空间,强调预测应发生在抽象表征空间中,模型不必也不应该去重建观测信号的全部细节,将输入信息压缩成高度概括的数学表征,模型直接在抽象层面上预测事物的运动规律和发展趋势。

李飞飞创立 World Labs，核心研究方向是“显式世界模型”，聚焦“空间智能”，核心目标是让 AI 理解三维世界^[34]。李飞飞定义了世界模型的生成性、多模态、交互性三个核心能力，提出要让机器人具备能在真实的三维、甚至是四维物理世界中，理解结构、保持一致性、进行推理，并与环境交互。发

布 Marble 世界模型^[35]，可将多模态的输入转化为一个可导航、可交互的三维世界。当前，世界模型路线正处于研究与探索阶段。

具身智能的三种主流技术路线在架构逻辑、响应速度、数据需求、可解释性与产业化维度进行横向对比，如表 3 所示。

表 3 三种技术路线对比

Table 3 Comparison of three technical routes

对比维度	端到端路线	分层式路线	世界模型路线
架构逻辑	以单个模型建立多模态输入至动作的映射	解耦架构，高层认知与规划，低层控制与执行，分别训练再联合	环境建模，三维物理世界数学表征，在抽象层面进行状态预测与动作规划
响应速度	较快，无层级间信息传递延迟	受限于模块间延迟	需推理与预测时间
数据需求	高，规模真实交互数据，训练算力与数据采集成本指数增长	分离独立训练，降低了训练难度和数据需求	依赖仿真/视频数据
可解释性	弱，黑箱	模块可拆解	预测可验证
产业化角度	被视为长期方向，现阶段受限于高昂的数据与算力成本，仅少数头部企业有能力布局	合理的资源需求、良好的解释性与成熟度，成为当前商业产品的首选架构	研究与探索阶段，接近“理解物理世界”的通用智能愿景

整体而言，三条路线的核心差异在于系统架构逻辑的根本不同，这种差异并非简单的优劣之别，而是在泛化能力、实时性能、部署成本与运行安全性等关键指标上各有侧重与权衡，多路线融合架构或将成为未来的重要发展趋势。

2.3 具身智能机器人的载体形式

具身智能机器人是指具备物理身体，在身体与真实环境相互作用过程中，自主感知、理解、决策并执行任务的独立智能体。

具身智能的物理载体是智能体与环境交互的直接媒介，是智能体实现环境交互、动作执行的硬件基础^[36]。当前，面向不同应用场景与功能需求，具身智能的物理载体已发展出多种形态，既能独立完成单项作业任务，也可依据场景需要进行组合使用（如 AGV 搭载机械臂、移动底盘协同人形机器人等），拓展多场景复合作业能力。按照本体构型与运动方式，现有具身智能实体可划分为固定底座式、轮式、足式、飞行式及其他特殊类型五大类别，如图 5。

1) 固定底座式载体



图5 具身智能主要载体形态

Fig. 5 Main carrier forms of embodied AI

固定底座式载体以工业机械臂、手术机械臂及智能机床为典型代表,其底座固定于工作台或地面,凭借多自由度关节/末端执行器实现在有限工作空间内的精确定位、力控制与高精度精细化操作^[37]。该类载体的核心优势在于高精度、高刚度,适合执行焊接、装配、打磨、精密加工、手术等对轨迹精度与力控制要求严苛的作业任务,代表性应用领域包括智能制造与医疗手术等领域,是工业与医疗场景下精细化作业的主流载体。

2) 轮式载体

轮式载体主要包括无人驾驶汽车、自主移动机器人 AMR 及自动导引车 AGV 等,依靠轮系驱动实现平坦地面高效平稳移动,凭借成熟的底盘控制技术适配标准化环境^[38]。轮式载体现已广泛应用于物流仓储如仓库内的 AGV 拣选系统、服务机器人如酒店配送机器

人、商场导购机器人及智能交通等领域。

3) 足式载体

足式载体包括四足机器人和双足人形机器人。该类载体依托仿生的腿足结构实现行走、奔跑、跳跃、攀爬等复杂运动,具备优异的地形适应能力、高运动自由度与环境灵活性,可翻越障碍、适配崎岖路面,在楼梯、碎石、草地、斜坡等地形中稳定移动^[39-40]。足式载体的核心价值在于突破轮式结构对平坦地面的依赖,适用于应急搜救与勘探、公共安全与巡检如变电站、化工厂巡检以及通用服务场景如家庭陪护、养老助残、工业作业任务。其中,人形机器人被视为具身智能最具代表性的载体形态^[41],承载着通用物理智能的终极愿景。

4) 飞行式载体

飞行式载体以无人机为代表,通过旋翼、固定翼或倾转旋翼等方式实现空中机动。该类载体的突出优势在于空中的高机动性与广域覆盖能力,突破地面空间约束,能够快速抵达高空、偏远等地面机器人难以进入的区域^[42-43]。飞行式载体在航拍与测绘、安防与救援、环境监测及物流配送等领域得到规模化应用。

5) 其他特殊类型载体

针对极端环境或特殊介质作业需求,水下机器人、空间在轨机器人、软体机器人归为其他特殊类型载体,通过定制化本体结构适配水下、太空、受限狭小空间等极端特殊工况,面向特种介质与极端环境完成专属作业任务。水下机器人可在海洋、湖泊等水下环境中执行勘探、管线检测、水下焊接等任务^[44-45];空间机器人需在微重力、高辐射、极端温差条件下完成在轨服务、星球探测等作业^[46-47];软体机器人采用柔性材料与驱动方式,具备连续变形能力与安全的人机交互特性,适用于狭窄空间探测、医学内窥镜、柔性抓取等场景^[48-49]。这些特殊载体突破了传统刚性结构的局限,将具身智能的应用边界拓展至深海、深空等领域。

6) 多形态组合

实际应用中,不同形态的物理载体可依据任务需求进行组合与协同,典型模式包括:轮式底盘与机械臂组合,形

成移动操作机器人,可实现在仓储环境中自主导航与精细作业;移动底盘与人形上半身组合形成轮式人形机器人,兼顾移动效率与操作灵巧性,也是具身智能载体的重要形式之一。

具身智能的物理载体呈现形态多样、场景适配、组合灵活的发展特征。不同载体在运动能力、环境适应性、操作精度与作业范围等方面各有侧重,共同构成了具身智能落地的硬件基础。相同的具身智能模型可向下兼容多类物理载体形式;而多形态载体的异构组合模式,进一步打破单一机器人的作业边界,成为现阶段具身智能产业化落地的重要发展方向。

2.4 具身智能机器人的关键技术

具身智能机器人整机硬件与软件系统可按照身、手、腿、能、感、脑六大模块拆解,完整覆盖机械本体、驱动能源、感知硬件与智能决策软件全链条,各模块协同支撑机器人环境感知、自主决策与物理执行全流程,如表4所示。

1) 躯体执行硬件

身:本体驱动单元,作为机器人本体的动力核心,由减速器、伺服电机以及丝杠等关键部件构成^[50-51]。这些部件的精度、响应速度与可靠性直接决定了机器人的运动品质与负载能力。减速器是连接动力源和执行装置的中间件,将转速降到各关节所需速度并增加扭矩,以满足对动力输出的要求,是精准传动的关键,保证机器人运动稳定性和精

度^[52]。电机作为动力源,主要作用是产生驱动转矩。电机的额关键性能要求是高转矩密度、低转矩波动、强过载能力、优异动态响应、高可靠性能、高效散热设计^[53]。丝杠是一种将旋转运动转化为直线运动的传动部件,在机器人关节中实现精准直线运动,是实现复杂动作和高负载操作的关键^[54]。其中,行星滚柱丝杠是人形机器人实现高精度直线运动的核心传动部件,另外在人形机器人的关节中也有利用机构来实现转动和移动的相互转换。

表4 具身智能机器人关键技术分解图
Table 4 Decomposition diagram of key technologies for embodied intelligent robots

板块	具体技术	
身	减速器、伺服电机、丝杠等	
手	灵巧手、夹爪等	
腿	轮式、四足、双足、复合式等	
能	电池、其他能源等	
感	视觉、触觉、听觉、惯性、力/力矩等	
脑	机器人操作系统	
	决策:	任务拆解、策略选择等
	规划:	路径规划、轨迹规划等
	控制:	机械臂、灵巧手移动导航等
	交互:	人机交互、多智能体交互等

手:末端执行单元,以灵巧手、夹爪为主流构型,是机器人和作业对象直接交互的执行末端^[55-56]。夹爪结构相对简单,通过两指或三指的平行开合实现快速、可靠的夹持,广泛应用于工业抓取与分拣。灵巧手通常具有多个独立驱动的手指,可模仿人手的抓取与执行复

杂动作的能力,像人手一样精准力控感知操作。当前灵巧手的技术路线以欠驱动+电机为主流路线,国内外五指方案占据绝对主流,腱绳传动方案的占比提升,驱动电机多数以空心杯电机为主,搭载触觉逐步成为标配。国外灵巧手自由度普遍偏高,国内灵巧手更注重本体轻量化设计^[57]。

腿:移动行走单元,包含轮式、四足仿生、双足人形以及轮足复合式等多种移动方案,对应不同地形通行需求,决定机器人空间机动能力,机器人下肢形态的演进,本质上是移动效率、地形适应性与控制复杂度三者之间的权衡,如图6所示。

能:能源供给单元,机器人的电池是集“储能-管理-防护”于一体的复杂系统,其性能直接决定了机器人的续航、充电频率和整体工作效率。以锂电池为当前主流供电方案,同时布局氢燃料电池、半固态电池、全固态电池等能源技术^[58]。高能量密度、高功率密度、快速充放电以及安全可靠的电源管理是核心挑战。人形机器人和四足机器人通常需要支撑数小时的连续作业,续航能力是制约其实用化的关键瓶颈之一。

感:多模态感知单元,是机器人获取环境与自身状态的关键通道,常用的传感器包括视觉传感器、听觉传感器、触觉传感器、测距传感器、力/力矩感知、惯性感知等多类传感硬件^[59-60],从视觉、力觉、听觉、位姿等多维度采集环境与本体状态数据,为上层“大脑”

提供输入信息，是实现复杂环境适应性 与精细操作的前提，如表5-表7所示。

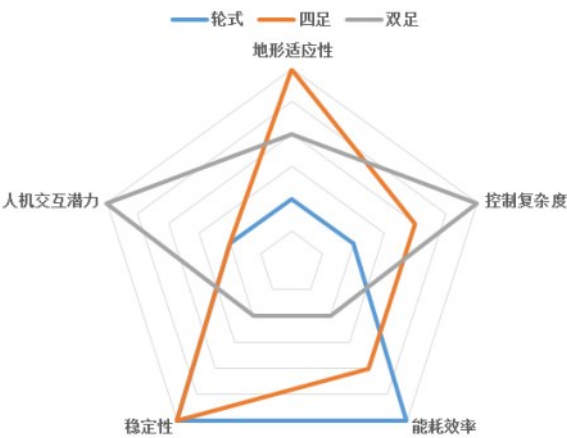


图6 机器人不同下肢形态的特性分析雷达图

Fig. 6 Radar chart comparing the key characteristics of various robot lower-limb configurations.

表5 (a) 二维视觉传感器
Table 5 (a) 2D vision sensor

相机	特点
线阵相机	高分辨率高速检测连续平面物体
面阵相机	静态或运动不剧烈的场景进行直观的检测、定位、识别与质量检查

表5 (b) 三维视觉传感器
Table 5 (b) 3D vision sensor

相机	原理	特点	应用场景
双目	模拟人眼，计算视差获取深度信息	户外效果佳、成本低但运算量大、特征缺失效果差	机器人导航、三维重建
结构光	条纹/栅格投射物体表面通过变形图案获取深度	一定范围内精度高、易受环境光干扰	工业高精度三维检测、人脸识别
TOF	测量调制光脉冲从发射到返回的飞行时间	快速反应、测量距离较远，精度和分辨率相对较低	体感游戏、手势识别
激光雷达	激光三角测距法或脉冲测距法进行扫描	探测距离远、范围广、精度高但成本较高	自动驾驶、地形测绘

2) 智能软件系统—“脑”

具身智能机器人“脑”依托机器人操作系统实现软硬件资源调度，并划分

为决策、规划、控制、交互四个模块。

决策模块负责根据任务目标与环境理解进行任务拆解与策略选择，并结合

当前状态形成从初始的作业状态到目标状态的动作序列。

规划模块分为路径规划与轨迹规划,路径规划是根据任务目标和环境信息如障碍物分布、地形结构等,结合空间可行性以及几何最优性,在机器人可达的空间中生成一条从起点到终点的无碰撞几何路径;轨迹规划是在路径规划的基础上,生成机器人连续可执行运动轨迹,保证轨迹的动态可行性、平滑性与精确性的机械臂各轴角位姿^[61-62]。

控制模块是将规划结果转化为各硬件可执行的驱动指令,实现对机械臂、灵巧手、移动底盘等的精确控制^[63]。通过机械运动控制协调机器人多关节的运动,实现复杂动作和姿态;灵巧手精细夹持控制完成具体的操作任务;移动底盘的导航控制实现在空间场景中移动导航、自主避障。

交互模块承担人机交互、多智能体集群协同交互功能,人机交互涵盖语音对话、手势识别、界面指令输入、意图理解等接口^[64];多智能体交互则涉及多机协同中的通信、任务分配与协调控制,其中,多智能体包含工业母机、工业机器人、人形机器人等多种载体形态^[65]。

整体而言,硬件“身、手、腿、能、感”负责物理空间的感知与动作落地,软件“脑”完成信息处理、任务决策与指令下发,软硬件双向闭环协同,构成完整的具身智能机器人技术体系。

3 具身智能机器人分级

具身智能机器人领域正处于技术快速演进与标准体系加速构建的并行阶段,目前已形成若干条并行推进的分级标准与分类体系,下面从标准体系以及学术研究分别展开。

国家标准层面,由全国信息技术标准化技术委员会(TC28)归口,中国电子技术标准化研究院牵头起草的《人工智能 具身智能智能化分级》(计划号:20252043-Z-469),目前处于在研阶段。

在团体标准层面,2024年10月,上海市人工智能行业协会发布T/SAIAS 016-2024《具身智能 智能化等级分级指南》。该标准按照感知、认知、决策、自主等核心能力,并结合功能性、自主性等原则,将具身智能划分为G1到G5五个阶段:G1阶段具备基础智能,依赖预设规则或简单感知能力,满足限定场景的基本操作需求;G2-G3阶段实现了中级智能,具备多模态感知、动态适应能力,可通过数据驱动的学习实现持续进化;G4-G5阶段则标志着通用智能的实现,具备自主决策、复杂任务执行能力。

中国电子学会于2025年5月发布《人形机器人智能化分级》标准(T/CIE 298-2025),该标准采用“四维五级”评价框架,从感知认知(P)、决策学习(D)、执行表现(E)、协作交互(C)四个维度,构建了L1至L5的五级智能化能力分级体系,从L1至

L5, 智能化能力水平逐级递增。其中, 感知认知 P1-P5 的核心能力为单模态感知→多模态融合→场景理解→跨领域认知→自主知识构建; 决策学习 D1-D5 的核心能力为规则执行→简单推理→任务规划与学习→知识迁移→自我演进; 执行表现 E1-E5 的核心能力为基础运动→多任务协调→工具运用→复杂操作→类人灵活执行; 协作交互 C1-C5 的核心能力为单模态响应→多模态理解→情绪识别→个性化交互→群体协同。

在学术研究层面, 南京大学等9家单位联合撰写的英文综述中, 依据智能机器人在不同环境中的任务完成能力、自主决策深度、交互复杂性和伦理认知能力进行分类, 提出了一个涵盖自主性、任务处理能力、环境适应能力与社会认知能力四个关键维度的智能机器人五级能力分级体系。IR-L0: 基础执行级别、IR-L1: 程序化响应级别、IR-L2: 基础感知与适应性级别、IR-L3: 人形认知与协作级别以及 IR-L4: 完全自主级别^[66]。

Yequan Wang 与 Aixin Sun 的论文^[67]中, 从全模态能力、类人认知能力、实时响应能力和泛化能力四个核心维度评估具身 AGI 能力, 提出具身 AGI 的 L1~L5 的五阶段路线图。L1 单任务完成, 机器人可靠地执行单一、定义明确的任务; L2 组合任务完成, 机器人能够通过将高级人类指令分解为简单行动序列来处理组合任务; L3 条件通用任务完成, 机器人能够处理广泛的任务类

别, 在任务、环境和人类指令之间展现条件泛化; L4 高度通用机器人, 对广泛未见任务表现出稳健泛化; L5 全能机器人, 能够满足人类日常需求的广泛谱系, 代表具身 AGI 的终极目标。

王烽、毛思通参考自动驾驶的分级标准, 将具身智能能力划分为五个等级: Level 1 (预编程/遥控), 无感知能力, 仅能重复执行固定轨迹或完全依赖人类遥控; Level 2 (单一任务自动化), 具备基础感知和避障能力, 可在结构化环境中执行单一特定任务; Level 3 (多任务/条件泛化), 能理解自然语言指令, 执行多种任务, 但在未见过的环境中需要人类偶尔接管; Level 4 (通用自主), 具备长程规划能力, 适应非结构化环境, 具备物理常识和自主纠错能力, 仅在极端情况下需要干预; Level 5 (完全具身智能), 具备人类水平的感知精度、灵巧操作能力和情感交互能力, 完全自主^[68]。

当前, 具身智能机器人的分级研究在标准规范与学术研究两个层面已涌现出多套分级框架、指标体系及评价维度。然而, 由于具身智能机器人技术仍处于快速演进阶段, 现有分级体系在分级依据、能力维度划分、评价指标设定以及判定逻辑等关键问题上尚未形成广泛共识。因此, 围绕具身智能机器人分级体系的深化研究与持续探索仍有待推进, 以期技术评估、产品对标及产业应用提供更为统一的参照框架。

表6 触觉传感器

Table 6 Tactile sensor

种类	优势	劣势
压阻式	过力的作用改变导电材料电阻值	迟滞性大、温漂高
电容式	分辨率和灵敏度较高	易受干扰、负载能力差
压电式	无需外部电源、动态特性佳	不适用于静态压力、难测切向力
视触觉	分辨率可达微米级，最接近人手触觉	依赖算法、成本高
磁电式	动态响应范围宽、支持三维力感知	结构复杂、易受磁场等干扰

表7 测距传感器

Table 7 Distance sensor

种类	原理	优势	劣势	应用场景
超声波	TOF（声波）	成本低、环境适应性较好、测量范围（厘米 - 数米）	精度较低（厘米级），响应慢	机器人简易避障，泊车
激光雷达	TOF（激光）	精度高（毫米 - 厘米级）、测量距离远（可达数百米）、速度快	成本高，怕恶劣天气	机器人SLAM、自动驾驶、3D测绘、高精度导航
毫米波雷达	FMCW（电磁波）	极端环境适应性强、探测距离远、可达数百米）、穿透能力强	分辨率低，点云稀疏	自动驾驶中前向感知、盲区检测、自适应巡航
红外距离	三角测量	成本低、体积小、响应速度快	易受光干扰，范围短	扫地机器人沿边清扫、近距离避障
视觉	视差、光学编码、TOF（光）	信息丰富、精度与量程匹配场景	不同方法存在各自劣势	物体识别与抓取、三维重建、SLAM、人脸识别、手势交互

4 具身智能机器人的挑战与展望

4.1 应用场景

具身智能机器人逐步落地工业场景、家庭服务、特种环境三大应用领域^[69]，不同场景依据工况特征匹配差异化的机器人本体形态与智能方案，实现智能化替代与效能升级。

工业场景：工业制造领域是具身智能机器人较为成熟的应用场景。具身智

能机器人突破传统工业机器人固定程序、执行重复性固定任务的局限，在非标零部件装配、产线物料自主转运与上下料、焊接加工、智能机床加工等环节，通过多模态感知，可自主识别工件偏差、实时校正、动态调整作业路径，适配多品种柔性生产线^[70-72]，替代人工完成高危、高强度、高精度的作业任务，提升生产效率、降低生产成本、保障作业安全，推动制造业向智能化、柔性化转型升级，契合新质生产力发展

需求。

家庭服务：家庭服务是具身智能机器人具有市场潜力的应用方向。相较于工业场景，家庭环境具有高度的非结构化、开放式、动态变化及人机共融等特点，对机器人的安全性、交互自然性与任务泛化能力提出了更高要求^[73-74]。具体应用涵盖医疗康复、日常家务劳作、养老助残、儿童教育陪伴、居家安保、物品配送、迎宾接待等细分方向。随着通用智能算法的迭代提升，家用服务领域的具身机器人逐步向全场景综合服务机型演进，满足人们日常生活中的服务需求，提升生活质量，是未来民用智能化的重要落地方向。

特种环境：针对人类难以近身的高危极端工况，具身智能机器人可替代或

辅助人类执行任务，主要包含军事任务、极限环境作业、灾害应急救援^[75-77]。具身智能特种机器人可深入易燃易爆、高温高压、深海、太空、地震坍塌废墟等恶劣环境，完成军事侦查、管线探伤、火场搜救、地质勘探等危险作业，规避人员伤亡风险。凭借环境自适应与自主决策能力，突破传统遥控机器人依赖人工远程操控的短板，在突发灾害、极端工况下实现自主探测与抢险作业。

具身智能机器人的渗透路径为：优先在结构相对受限、场景发挥可预测的偏工业类场景落地，然后逐步走向特种作业、医疗康复等领域，最终进入人类的日常生活空间，完成家庭服务，如图7所示。

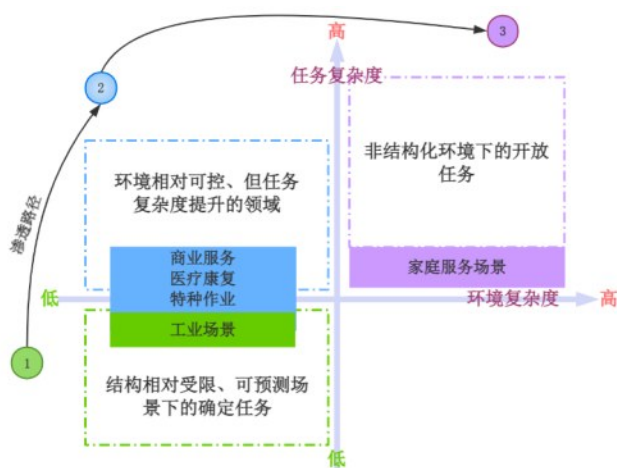


图7 具身智能机器人渗透路径图

Fig.7 Penetration pathway of embodied AI robots

总体来看，工业场景已进入商用落地阶段，家庭服务处于样机试点与产业化培育期，特种环境应用依托政策与刚

需持续稳步迭代，三大场景共同构成当前具身智能产业化落地的主体版图。

4.2 发展挑战

尽管具身智能机器人在理论算法与样机研发层面取得显著进展,并在工业场景、家庭服务、特种环境等领域展现出广阔应用前景,但其产业化落地仍面临核心零部件自主不足、高质量数据稀缺、通用智能技术路线尚未收敛以及实际场景落地等瓶颈问题。

1) 核心零部件自主可控能力不足

高端算力 GPU、激光雷达等精密传感元器件对外依存度较高,供应链自主可控能力不足,存在关键技术“卡脖子”风险。减速器等核心传动部件长期高负荷工况下易出现柔轮疲劳损耗,大幅降低整机使用寿命与运行稳定性^[78-79]。电机在高负载或长时间运行时温升显著,导致能耗增加、效率下降,甚至影响控制精度^[80-81]。智能体执行部件的设计与制造如灵巧手、多自由度关节模组工程成熟度不足。此外,数据集集成基础设施建设相对滞后,限制了模型训练的效率和效果^[82-83]。

2) 高质量物理数据稀缺

模型能力的持续进化高度依靠海量、优质、多样的数据供给。数据采集方式呈现多元化,各有优劣:真实采集数据保真度高,但人工采集模式成本高、产出效率有限,难以满足规模化训练需求;仿真合成数据虽可低成本快速生成,但与真实环境之间仍存在显著虚实域迁移鸿沟。可用于训练的有效数据规模远跟不上模型迭代的需求。须着力

破解高质量物理世界数据稀缺的瓶颈。

3) 通用智能的技术路线尚未收敛

现阶段具身智能领域形成端到端一体化、分层模块化、世界模型三类技术路线并行探索的格局,三者架构逻辑存在根本差异:端到端侧重全局优化^[84],分层强调模块化解耦^[85],世界模型聚焦环境建模^[86]。技术路线分立导致产业界与学术界在资源投入上呈现分散态势,难以形成协同攻关合力。更关键的是,目前尚无任何一条路线能够在泛化性、数据效率、可解释性、实时性与成本之间取得令人满意的平衡。

4) 实地场景落地适配挑战

当前国内具身智能领域标准体系尚未完备,各类产品评测指标、分级维度缺乏统一规范,造成不同厂商产品横向对标难度大、产业链上下游协同壁垒突出。技术研发、产品迭代、场景验证脱节,产业标准与协同体系缺失延缓技术落地速度。

受现有软硬件技术水平制约,现阶段具身智能机器人样机距离全场景商业化落地仍存在明显差距。面对工业离散制造领域、复杂居家环境、多变野外工况等差异化场景的多元化需求,现有系统的通用化适配能力明显不足,技术成熟度尚无法支撑大范围规模化落地应用。

4.3 展望

具身智能机器人正处于从实验室技术验证向规模化商业应用跨越的关键阶

段。这一进程仍受到硬件、数据、算法架构与产业生态的耦合制约,亟需构建多方协同、系统推进的正向发展格局。

在核心硬件层面,未来需统筹推进核心硬件攻关,重点突破高可靠性关节模组、精密传感器、长寿命精密传动部件及AI算力芯片等关键方向的研发迭代,逐步实现供应链自主可控,夯实机器人硬件技术底座。

在数据层面,需进一步推动具身智能机器人实景实训,搭建跨企业、跨场景的数据共建共享平台,统一异构机器人的数据规范与交互接口。持续积累真实物理交互样本,为模型物理常识习得与环境自适应能力的持续提升提供高质量数据支撑。

在算法架构层面,未来应打破单一路线割裂发展思维,推动多技术路线融合创新,探索类人认知机制的混合架构。各模块的融合方式与功能边界仍有待在大规模真实部署中持续探索与收敛。

在标准化与产业生态层面,亟需以国家标准为顶层指引,加速统一接口协议、数据格式与存储规范、量化评测体系,打通技术研发、产品迭代与场景落地之间的协同链路,降低系统集成成本,构建开放共赢的产业生态。

具身智能机器人产业落地仍面临诸多严峻挑战。关键在于构建以标准规范为引领,技术底座为支撑、多技术融合为路径的协同创新体系,推动具身智能机器人向通用自主智能体演进,充分释

放其赋能新质生产力、重塑生产生活模式的战略价值。

5 结论

具身智能是实现通用人工智能(AGI)的重要抓手。本文系统梳理了机器人发展史与AI演进脉络,归纳了具身智能的概念、技术路线、具身智能机器人的载体形态、分级标准、应用场景及发展挑战,得出以下主要结论:

1) 技术溯源与融合关系:具身智能是机器人与人工智能发展的交汇点,是两者深度融合的体现,是破解传统人工智能落地困境、实现通用智能的可能路径。

2) 技术框架:本文提出具身智能机器人DREAM技术框架,总结出物理载体依赖、孪生映射、自成长三大核心特征,阐释了“感知-决策-规划-执行-反馈-学习-优化”的智能运行闭环。

3) 技术路线呈多元化并行格局:当前领域并存一体化端到端模型、分层式模型、世界模型三类技术路线,各路线架构差异显著,各有优劣,尚未形成统一范式。

4) 载体形态与技术模块:具身智能机器人的物理载体形态多元、组合方式灵活。整机技术拆解为身、手、腿、能、感、脑六大模块,软硬件协同构成具身智能机器人技术体系。

5) 分级标准现状:具身智能机器人分级研究在标准规范与学术层面已涌现多套评价体系,但分级依据、能力维

度与判定逻辑尚未达成共识。

6) 行业整体态势:具身智能机器人呈现工业先行试点、服务与特种场景稳步探索的应用格局。现阶段受制于核心硬件短板、高质量物理数据稀缺、算法路线割裂、标准体系不完善等多重约束,制约规模化落地推广。对应需从硬件攻关、数据共建、多技术路线融合、构标准生态等多维度协同发力,构建技术攻关与产业应用协同演进的正向循环,持续提升具身智能机器人通用适配能力与运行稳定性。

展望未来,随着上述瓶颈逐步破解,具身智能机器人将逐步进化为具备在开放环境中自主完成多样化任务的通用智能体,成为人类生活和工作中的得力伙伴,深刻推动社会生产力的跃升与人类生活方式的变革。

参考文献

- [1] 叶浩生. 有关具身认知思潮的理论心理学思考[D], 2011.
- [2] 王文晟, 谭宁, 黄凯, et al. 基于大模型的具身智能系统综述[J]. 自动化学报, 2025, 51 (1): 1-19.
- [3] 陶永, 万嘉昊, 王田苗, et al. 构建具身智能新范式: 人形机器人技术现状及发展趋势综述[J]. 机械工程学报, 2025, 61 (15): 121-147.
- [4] 陈铂垒, 康嘉绪, 钟萍, et al. 面向具身人工智能的物体目标导航综述[J]. 软件学报, 2024, 36 (4): 1715-1757.
- [5] 袁依格, 何卓丰, 李威, et al. 具身智能驱动的智能制造应用发展研究[J]. 中国工程科学, 2025, 27 (3): 67.
- [6] 高超, 杨莹, 陈世超, et al. 多模态模型驱动的具身智能研究综述[J]. 智能感知工程, 2025, 2 (2): 1-12.
- [7] Cordeschi R. AI turns fifty: revisiting its origins[J]. Applied Artificial Intelligence, 2007, 21 (4-5): 259-279.
- [8] 林艳. 人工智能的符号主义纲领及其困境[J]. 求索, 2019, 6: 186-193.
- [9] Brooks R. A robust layered control system for a mobile robot[J]. IEEE journal on robotics and automation, 2003, 2 (1): 14-23.
- [10] Sutton RS, Barto AG. Reinforcement learning: An introduction[M]: MIT press Cambridge, 1998.
- [11] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the national academy of sciences, 1982, 79 (8): 2554-2558.
- [12] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors[J]. Nature, 1986, 323 (6088): 533-536.
- [13] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313 (5786): 504-507.
- [14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [15] 蔡恒进. 行为主义, 联结主义和符号主义的贯通[J]. 上海师范大学学报(哲学社会科学版), 2020, 49 (4): 87-96.
- [16] 赵冬斌, 邵坤, 朱圆恒, et al. 深度强化学习综述: 兼论计算机围棋的发展[J]. 控制理论与应用, 2016, 33 (6): 701-717.
- [17] 叶浩生. 具身认知的原理与应用[M]. 北京: 商务印书馆, 2017.
- [18] 郑南宁, 杨劼, 姜维周, et al. 具身智能发

- 展趋势与展望[J]. 中国工程科学, 2026, 28 (2): 1.
- [19] Fei-Fei L, Krishna R. Searching for computer vision north stars[J]. *Daedalus*, 2022, 151 (2): 85-99.
- [20] 刘华平, 郭迪, 孙富春, et al. 基于形态的具身智能研究: 历史回顾与前沿进展[J]. *自动化学报*, 2023, 49 (6): 1131-1154.
- [21] 蒋树强, 宋新航, 张思贤. 具身智能系统的“大脑”里都有什么?[J]. *计算*, 2025, 1 (4): 35-42.
- [22] 孙长银. “试错”驱动的具身智能学习及进化[J]. *计算*, 2025, 1 (4): 43-48, 63.
- [23] Ahn M, Brohan A, Brown N, et al. Do as i can, not as i say: Grounding language in robotic affordances[J]. *arXiv preprint arXiv:220401691*, 2022.
- [24] Zitkovich B, Yu T, Xu S, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control[C]: *PMLR*, 2023: 2165-2183.
- [25] O' Neill A, Rehman A, Maddukuri A, et al. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0[C]: *IEEE*, 2024: 6892-6903.
- [26] Bu Q, Cai J, Chen L, et al. Agibot world colosseo: A large-scale manipulation platform for scalable and intelligent embodied systems[J]. *arXiv preprint arXiv:250306669*, 2025.
- [27] Liang J, Huang W, Xia F, et al. Code as policies: Language model programs for embodied control[C]: *IEEE*, 2023: 9493-9500.
- [28] Lorente A. Value Creation and Value Capture in AI: A Triple Helix Model[C], 2025: 1637-1646.
- [29] Belkhal S, Ding T, Xiao T, et al. Rt-h: Action hierarchies using language[J]. *arXiv preprint arXiv:240301823*, 2024.
- [30] Huang W, Xia F, Xiao T, et al. Inner monologue: Embodied reasoning through planning with language models[J]. *arXiv preprint arXiv:220705608*, 2022.
- [31] Ha D, Schmidhuber J. Recurrent world models facilitate policy evolution[J]. *Advances in neural information processing systems*, 2018, 31.
- [32] Wu P, Escontrela A, Hafner D, et al. Daydreamer: World models for physical robot learning[C]: *PMLR*, 2023: 2226-2240.
- [33] Assran M, Duval Q, Misra I, et al. Self-supervised learning from images with a joint-embedding predictive architecture [C], 2023: 15619-15629.
- [34] Bommasani R, Arora S, Chayes J, et al. Advancing science-and evidence-based AI policy[J]. *Science*, 2025, 389 (6759): 459-461.
- [35] Yang J, Yang S, Gupta AW, et al. Thinking in space: How multimodal large language models see, remember, and recall spaces [C], 2025: 10632-10643.
- [36] 兰沅卜, 赵文博, 朱凯, et al. 基于具身智能的移动操作机器人系统发展研究[J]. *中国工程科学*, 2024, 26 (1): 139.
- [37] 胡杰, 张铁. SCARA 机器人控制系统的设计与研究[J]. *机械设计与制造*, 2012, (5): 162-164.
- [38] 朱磊磊, 陈军. 轮式移动机器人研究综述 [J]. *机床与液压*, 2009, (8): 242-247.
- [39] 陈致远, 涂群章, 张详坡, et al. 足式爬行机器人研究进展与发展趋势[J]. *Journal of Ordnance Equipment Engineering*, 2020, 41 (9): 1-12.
- [40] Huang G-Y, Li W-F. Humanoid Robot Walking and Grasping Method Using

- Similarity Reward-Augmented Generative Adversarial Imitation Learning[J]. *Sensors*, 2026, 26 (9): 2756.
- [41] 刘伟龙, 李彬, 侯兰东, et al. 基于深度强化学习的四足机器人研究综述[J]. 2022.
- [42] Cao H, Shen J, Zhang Y, et al. Proximal cooperative aerial manipulation with vertically stacked drones[J]. *Nature*, 2025, 646 (8085): 576-583.
- [43] Gupta H, Guo X, Ha H, et al. UMI-on-Air: Embodiment-Aware Guidance for Embodiment-Agnostic Visuomotor Policies[J]. *arXiv preprint arXiv: 251002614*, 2025.
- [44] 李硕, 吴园涛, 李琛, et al. 水下机器人应用及展望[J]. *中国科学院院刊*, 2022, 37 (7): 910-920.
- [45] 李硕, 刘健, 徐会希, et al. 我国深海自主水下机器人的研究现状[J]. *中国科学: 信息科学*, 2018, 48 (9): 1152-1164.
- [46] 张辉, 左孝中, 张伟, et al. 空间双臂机器人漂浮基座扰动最小化轨迹规划[J]. *Electronics Optics & Control*, 2025, 32 (2): 24-31.
- [47] 高胜, 常玉连, 张瑞杰, et al. 基于自由漂浮空间机器人的空间焊缝跟踪[J]. *焊接学报*, 2008, (11): 65-68.
- [48] 管清华, 孙健, 刘彦菊, et al. 气动软体机器人发展现状与趋势[J]. *中国科学: 技术科学*, 2020, 50 (7): 897-934.
- [49] 王国彪, 陈殿生, 陈科位, et al. 仿生机器人研究现状与发展趋势[J]. *机械工程学报*, 2015, 51 (13): 27-44.
- [50] Kaminaga H, Otsuki S, Nakamura Y. Development of high-power and backdrivable linear electro-hydrostatic actuator[C]: *IEEE*, 2014: 973-978.
- [51] Pratt JE, Krupp BT. Series elastic actuators for legged robots[C]: *SPIE*, 2004: 135-144.
- [52] Wensing PM, Wang A, Seok S, et al. Proprioceptive actuator design in the mit cheetah: Impact mitigation and high-bandwidth physical interaction for dynamic legged robots[J]. *Ieee transactions on robotics*, 2017, 33 (3): 509-522.
- [53] Guo K, Guo Y, Fang S, et al. Design and analysis of a permanent magnet frameless motor[J]. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2024, 12 (3): 3124-3134.
- [54] Hollander KW, Sugar TG. Design of lightweight lead screw actuators for wearable robotic applications[J]. 2006.
- [55] Townsend W. The BarrettHand grasper – programmably flexible part handling and assembly[J]. *Industrial Robot: an international journal*, 2000, 27 (3): 181-188.
- [56] Grebenstein M, Chalon M, Friedl W, et al. The hand of the DLR hand arm system: Designed for interaction[J]. *The International Journal of Robotics Research*, 2012, 31 (13): 1531-1555.
- [57] Xue H, Ren J, Chen W, et al. Reactive diffusion policy: Slow-fast visual-tactile policy learning for contact-rich manipulation[J]. *arXiv preprint arXiv: 250302881*, 2025.
- [58] Mao Z, Xie Z, Fang K, et al. In-plant autonomous mobile robot scheduling and routing problem considering battery consumption model[J]. *Transportation Research Part C: Emerging Technologies*, 2026, 183: 105488.
- [59] Wisth D, Camurri M, Fallon M. Vilens: Visual, inertial, lidar, and leg odometry for all-terrain legged robots[J]. *Ieee transactions on robotics*, 2022, 39 (1): 309-326.

- [60] Cheng N, Xu J, Guan C, et al. Touch100k: A large-scale touch-language-vision dataset for touch-centric multimodal representation[J]. *Information Fusion*, 2025, 124: 103305.
- [61] Karaman S, Frazzoli E. Sampling-based algorithms for optimal motion planning[J]. *The International Journal of Robotics Research*, 2011, 30 (7): 846-894.
- [62] Chi C, Feng S, Xu Z, et al. Visuomotor policy learning via action diffusion. Google Patents, 2025.
- [63] Kwon T, Di Palo N, Johns E. Language models as zero-shot trajectory generators[J]. *IEEE Robotics and Automation Letters*, 2024, 9 (7): 6728-6735.
- [64] Beer JM, Fisk AD, Rogers WA. Toward a framework for levels of robot autonomy in human-robot interaction[J]. *Journal of human-robot interaction*, 2014, 3 (2): 74.
- [65] Erden MS, Tomiyama T. Human-intent detection and physically interactive control of a robot without force sensors[J]. *Ieee transactions on robotics*, 2010, 26 (2): 370-382.
- [66] Long X, Zhao Q, Zhang K, et al. A survey: Learning embodied intelligence from physical simulators and world models[J]. *arXiv preprint arXiv:250700917*, 2025.
- [67] Wang Y, Sun A. Toward embodied agi: A review of embodied ai and the road ahead[J]. *arXiv preprint arXiv:250514235*, 2025.
- [68] 王烽, 毛思通. 具身智能: 核心挑战, 关键技术演进与产业前景展望[J]. *自动化博览*, 2026, 43 (1): 50-57.
- [69] Xu J, Sun Q, Han Q-L, et al. When embodied AI meets Industry 5.0: Human-centered smart manufacturing[J]. *IEEE/CAA Journal of Automatica Sinica*, 2025, 12 (3): 485-501.
- [70] Ren L, Dong J, Liu S, et al. Embodied intelligence toward future smart manufacturing in the era of AI foundation model[J]. *IEEE/ASME Transactions on Mechatronics*, 2024, 30 (4): 2632-2642.
- [71] Fan H, Liu X, Fuh JYH, et al. Embodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics[J]. *Journal of Intelligent Manufacturing*, 2025, 36 (2): 1141-1157.
- [72] Ren Y, Liu Y, Ji T, et al. AI Agents and Agentic AI – navigating a plethora of concepts for future manufacturing[J]. *Journal of Manufacturing Systems*, 2025, 83: 126-133.
- [73] Lee I. Service robots: a systematic literature review[J]. *Electronics*, 2021, 10 (21): 2658.
- [74] Lisondra M, Benhabib B, Nejat G. Embodied ai with foundation models for mobile service robots: A systematic review[J]. *Robotics*, 2026, 15 (3): 55.
- [75] 孙长银, 袁心, 王远大, et al. 具身智能自主无人系统技术[J]. *自动化学报*, 2025, 51 (4): 762-777.
- [76] 东鹏欧, 欣丁, 晶茗杨, et al. 基于立体环境智能检测的快速应急响应机器人设计[J]. *智能城市应用*, 2025, 8 (3): 107-109.
- [77] 相辉玉, 许创, 张宝海. 双臂救援机器人现状及发展综述[J]. *Construction Machinery & Equipment*, 2026, 57 (4): 106.
- [78] Guo S, He Y, Shi L, et al. Modal and fatigue analysis of critical components of an amphibious spherical robot[J]. *Microsystem Technologies*, 2017, 23 (6): 2233-2247.
- [79] Bi Z, Meruva K. Modeling and prediction of fatigue life of robotic components in in-

- telligent manufacturing[J]. *Journal of Intelligent Manufacturing*, 2019, 30 (7): 2575-2585.
- [80] Ishiki A, Neeves K, Marr D. Thermally assisted microbot transport through high-viscosity media[J]. *Scientific Reports*, 2025.
- [81] Li G, Svogor I, Beltrame G. Long-term pattern formation and maintenance for battery-powered robots[J]. *Swarm Intelligence*, 2019, 13 (1): 21-57.
- [82] Almohamade SS, Clark JA, Law J. Dataset for collaborative robotics[J]. *Data*, 2026, 11 (5).
- [83] Mei H, Peng L. On multi-robot data collection and offloading for space-aerial-surface computing[J]. *IEEE Wireless Communications*, 2023, 30 (2): 90-96.
- [84] Jiang S-Q, Zhang S-X, Tao S-D, et al. Self model for embodied artificial intelligence [J]. *Journal of Computer Science and Technology*, 2026: 1-16.
- [85] Li Q, Liang Y, Wang Z, et al. Cogact: A foundational vision-language-action model for synergizing cognition and action in robotic manipulation[J]. *arXiv preprint arXiv:241119650*, 2024.
- [86] Zhang Z, Chen R, Ye J, et al. Whale: Towards generalizable and scalable world models for embodied decision-making[J]. *arXiv preprint arXiv:241105619*, 2024.