

具身智能前沿技术发展与通用人工智能路径探索

林琦¹ 郑霖¹ 苏旭云² 郑钦匀³ 林其发⁴ 孙威⁵

1 福建人形机器人创新中心有限公司,福建福州,350000;

2 烨琦科技(福建)有限公司,福建福州,350000;

3 烨琦科技(福州)有限公司,福建福州,350000;

4 福厦金新材料研究院(厦门)有限公司,福建厦门,361000;

5 烨琦科技(厦门)有限公司,福建厦门,361000;

摘要: 具身智能被视为实现通用人工智能(AGI)的关键路径,其前沿技术正经历爆发式发展。本文围绕“感知-学习-想象-协同”四个核心方向,系统分析具身智能的前沿技术进展,包括多模态融合感知、世界模型与想象机制、社会性交互协同等关键技术,探讨具身智能从专用型向通用型演进的技术路径,并展望其在实现通用人工智能过程中的作用与挑战。

关键词: 具身智能;通用人工智能;多模态融合;世界模型;技术前沿

DOI: 10.64216/3104-9680.25.02.017

引言

具身智能作为人工智能领域的前沿方向,强调智能体通过感知、理解和交互来适应环境,并执行任务的能力。与传统的基于规则或符号的人工智能不同,具身智能强调感知和行动相结合,使智能体能够更好地理解其周围的环境和与环境的互动。近年来,随着算力提升、算法创新和数据积累,具身智能技术取得显著进展,被广泛认为是通往通用人工智能的重要途径。具身智能的核心在于构建完整的“感知-理解-决策-执行”闭环系统,使智能体能够在真实或虚拟环境中完成复杂任务。这一过程涉及多模态感知、环境理解、任务规划、动作生成等多个环节,需要集成感知计算、认知科学、机器人学等多学科知识。本文从技术前沿视角,分析具身智能的关键技术进展与发展趋势,探讨其实现通用人工智能的可行路径。

1 具身智能前沿技术发展分析

1.1 多模态融合感知技术

多模态融合感知是具身智能与环境交互的基础。前沿研究正从被动感知向主动感知发展,智能体能够自发改变观察角度或执行交互行为来获得更高的信息增益。例如,主动视觉感知方法如 Neu-NBV、Fan 等通过智能体主动选择观测点位,提升环境建模与理解效率。在三维视觉定位方面,ScanRefer、3DVG-Transformer、

ReGround3D 等方法将自然语言与三维点云数据结合,在复杂环境中实现精准的目标对象定位与推理。这些技术为具身智能体在服务、救援、无人仓储等场景下的自主操作奠定基础。非视觉感知技术也取得重要进展。GelSight、DIGIT、9DTact 等触觉感知技术可捕捉表面纹理、压力分布与滑动等信息,为智能体在精细装配或柔性抓取等复杂操作任务中提供高分辨率的物理反馈。多模态感知的融合与协同,正成为提升具身智能环境理解能力的关键方向。

1.2 世界模型与想象机制

世界模型(WorldModel)是具身智能体对环境规律和物理法则的内化表征,使其能够在行动前预测结果,进行推理和规划。前沿研究如 DeepMind 的 DreamerV3、E3B 等在虚拟环境中构建带有时序与物理一致性的世界模型,帮助智能体预测环境状态。具身智能的“想象”能力体现在其对可能未来的模拟和评估上。例如,斯坦福大学李飞飞团队的研究表明,在复杂环境中进化的虚拟生命能更快学习新任务,这表明智能体可以通过内部模拟加速学习。Helix 框架首创的“S1 快系统, S2 慢系统”的 VLA 模型,通过快慢系统分工实现反应性行动与深思熟虑决策的平衡。世界模型与想象机制的发展,使具身智能体能够进行离线推理和因果推断,减少实际交互中的试错成本,提高决策质量。这一方向被认为是实现高效强化学习和泛化能

力的关键。

1.3 社会性交互与协同技术

社会性交互是具身智能体融入人类环境的重要能力。前沿研究关注智能体与人类之间的自然、高效、语义对齐的互动方式。例如，具身问答系统（如 EQAv1、iGQA）使智能体在视觉环境中自主移动、感知并回答问题。多智能体协同是另一重要方向。DiscussNav 框架通过多专家讨论来帮助代理进行导航，将具有不同能力的大模型作为领域专家，让代理在每一步移动之前与这些专家积极讨论，收集关键信息。这种方法有效纠正了意外错误，并筛选出不一致的运动决策。随着自然语言理解和多模态融合技术的升级，具身交互不断拓展智能体在服务机器人、人机协作和日常生活中的应用深度，努力实现“所见即所得、所说即所行”的人机互动模式。

2 具身智能通向通用人工智能的路径分析

2.1 从专用到通用的技术演进

当前具身智能主要针对特定任务和环境设计，属于“专用型”智能。向通用人工智能演进需要解决任务泛化、环境适应和自主学习三大挑战。技术演进路径包括：通过大规模预训练积累通用知识和技能；采用模块化架构实现能力的组合和复用；建立自监督学习机制减少对标注数据的依赖。例如，Google 的 RT-X 项

目试图构建通用机器人基础模型，通过大规模多样化数据训练，使机器人能够适应不同任务和环境。这种基础模型方法借鉴了大语言模型的成功经验，试图在机器人领域实现类似的能力跨越。

2.2 具身智能与抽象思维的结合

通用人工智能需要结合具身体验和抽象思维能力。当前研究尝试将具身智能与大语言模型相结合，例如 PaLM-E 模型将真实世界的连续传感器模态直接融入语言模型中，实现了单词和感知之间的联系。这类尝试使智能体既能理解抽象概念和符号，又能将其与物理现实联系起来。具身智能与抽象思维的结合面临表征对齐难题：如何将连续、高维的感官输入与离散、组合性的符号概念对应起来。解决这一难题可能需要新的表示学习方法和推理架构。

2.3 自主性与价值对齐挑战

通用人工智能需要具备一定程度的自主性，能够主动设定目标并追求目标。但这带来了价值对齐问题：如何确保智能体的目标与人类价值观一致。具身智能由于与物理世界直接交互，其行为后果更加直接和显著，价值对齐问题尤为关键。当前研究开始探索具身智能的价值对齐方法，包括通过人类反馈进行强化学习、价值学习框架、可解释性与透明度提升等。这些研究旨在确保具身智能的发展符合人类利益和安全需求。

表 1：具身智能实现通用人工智能的关键挑战与应对方向

挑战类别	具体挑战	前沿研究方向	代表性工作
感知与交互	多模态信息融合、主动感知	多模态融合、主动感知算法	NeU-NBV、3DVG-Transformer
推理与规划	因果推理、长期规划、物理推理	世界模型、分层强化学习	DreamerV3、Helix 框架
学习与适应	样本效率低、仿真到现实差距	元学习、域自适应、进化算法	DERL 框架、PaLM-E
社会与伦理	人机交互、价值对齐、安全	人类反馈学习、可解释 AI	DiscussNav、EQAv1

3 讨论与展望

具身智能作为通向通用人工智能的重要路径，其发展将经历从专用到通用、从简单到复杂、从封闭环境到开放世界的渐进过程。当前技术已在多模态感知、世界模型构建和社会交互等方面取得显著进展，但距通用人工智能仍有相当距离。未来发展趋势包括：首先，感知-行动闭环学习将更受重视，智能体通过主动与环境交互获取知识，而非仅从静态数据学习；其次，模拟与真实世界的边界将进一步模糊，通过仿真到现

实迁移技术实现高效训练和安全部署；最后，社会性智能将成为重点，使智能体能够与人类和其他智能体协同完成复杂任务。从哲学视角看，具身智能的发展也促使我们重新思考智能的本质。传统的“离身智能”观念将智能视为抽象算法，而具身智能强调智能与身体、环境的不可分割性。这种观念转变不仅具有理论意义，也指导着技术创新路径的选择。实现通用人工智能是具身智能的长期目标，需要跨学科合作和持续投入。随着技术进步，具身智能有望在智能制造、家庭服务、医疗康复、太空探索等领域发挥重要作用，

全面提升人类生活品质。

4 结论

具身智能前沿技术正快速发展，在多模态融合感知、世界模型与想象机制、社会性交互协同等方面取得显著进展。这些技术为具身智能从专用型向通用型演进奠定了基础，也为实现通用人工智能提供了可行路径。当前具身智能仍面临环境适应、任务泛化、自主学习等多重挑战，需要从算法、数据、平台等多个层面进行突破。未来研究应更加重视感知与行动的深度融合、仿真与现实的协同演化、以及安全性与社会价值的对齐。具身智能的发展不仅是技术创新的过程，也是对人类智能本质的深入探索。通过构建具有身体、能够与环境交互的智能系统，我们或许能更深入地理解智能的起源和本质，最终实现人工智能与人类社会的和谐共处。

参考文献

- [1] 张岩,罗旭,国玉霞,等.智能技术具身下审美教育锚定人工智能素养的理路演进[J].中国电化教育,2024(5):120-127.
- [2] 曾昭睿,周康林,柯罗马.具身智能如何重塑制造业

企业工作流与产业格局[J].清华管理评论,2024(9):25-33.

[3] 黄小明.具身智能的有效规制路径[J].检察风云,2024(23):13-15.

[4] 王亚坤,方勇,江昊,等.2023年生成式人工智能技术主要发展动向分析[J].无人系统技术,2024,7(2):101-112.

[5] 肖仰华.生成式语言模型与通用人工智能:内涵,路径与启示[J].社会科学文摘,2024(4).

[6] 康薇薇.媒体融合视域下虚拟新闻主播创新路径研究[J].新闻文化建设,2024(19):58-60.

[7] 郭亚军,庞义伟,周家华,等.具身智能知识服务:元宇宙与AIGC等新技术赋能下的知识服务新图景[J].情报理论与实践,2024,47(4):90-97.

[8] 马欣,王振雄.身体演变,技术嵌入与关系建构:智媒时代主持传播的主体迷思与重归[J].2024(1):57-7.

[9] 佚名.第二届中国具身智能大会(CEAI2025)[J].智能系统学报,2024,19(6):1448-1448.

[10] 钱鹏,张尧,唐盛.人工智能与苏剧表演姿态的交融:技术与传统的碰撞[J].科技视界,2024(35).