CAICT 中国信通院

X-HUMANOID 北京人形机器人创新中心

集智 蓝皮书 No.202405

具身智能发展报告

（2024 年）

中国信息通信研究院

北京人形机器人创新中心有限公司

2024年8月

版权声明

前　言

具身智能是人工智能（AI）与其他学科交叉融合发展的新范式，从字面可理解为“具身+智能”，通过给 AI 赋予“身体”，使其能够与物理世界产生交互，并在交互中主动探索世界、认识世界、改变世界。随着数字世界的 AI 算法开始展现出逼近甚至超越人类的思维能力，具身智能有望打开 AI 从数字世界到物理世界的窗口，在复杂的物理世界中进一步延伸和拓展 AI 边界，实现“知行合一”。

具身智能将在技术涌现式创新和突破下，实现“一脑多形”，即让一个智能系统适配各种形态的物理实体，如智能机器人、智能车辆等；实现“一机多用”，即让一个机器设备可以灵活地执行多种任务，适应多样化的场景需求。未来具身智能将从工业协作生产到柔性制造，从家务助手到医疗护理，从灾难救援到太空探索，深入融入人类社会。但当前其仍面临技术能力不足，数据短缺，以及工程实现复杂等一系列挑战。

本报告从 AI 视角切入，致力于厘清具身智能的概念内涵、演进历程、技术体系，通过梳理当前具身智能技术发展现状，研判分析具身智能应用潜力与可能影响，提出面临的问题挑战，展望思维智能和行动智能融合的未来发展趋势。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

图 目 录

一、全球具身智能发展态势

1950 年，图灵在其经典论文《Computing Machinery and Intelligence》[1]中探讨“机器是否能思考”这一根本问题，认为人工智能的终极形态是像人一样能与环境交互感知，自主规划、决策、行动和执行的机器人/仿真人（在虚拟环境中）。而有望实现的两条路径，一是聚焦抽象计算（比如下棋）所需的智能，二是为机器配备最好的传感器，使其可以与人类交流，像婴儿一样进行学习。后续，这两条路径逐渐演变成了离身智能（Disembodied Artificial Intelligence[2]）和具身智能（Embodied Artificial Intelligence，简称“EAI”）。

当前，依靠海量数据，结合算法和计算能力的提升，以 ChatGPT 为代表的离身智能实现智能涌现。自其推出之后，数字世界的 AI 技术逐步展现出逼近人类甚至超越人类的思维能力。加利福尼亚大学圣迭戈分校的研究团队在交互式双人图灵测试中发现，人们无法区分 GPT-4 与人类[3]。但在物理世界中，智能机器人仍然仅是智力有限的任务工具。在此背景下，人们的关注点转向如何让 AI 的认知从互联网的数字信息拓展到现实的物理概念，包括感官、空间、行动等信息，并将其更好地应用于物理世界。实际上，大模型对互联网上大量图文信息的处理和学习，本质上是“读万卷书”的过程，这

[1] https://phil415.pbworks.com/f/TuringComputing.pdf

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10020609/pdf/frai-06-1148227.pdf

[3] https://arxiv.org/pdf/2405.08007

可以增强智能体的感知、知识理解和思维能力，但无法取代“行万里路”所带来的体验。就像人类在真实世界中的亲身体验和劳动，无法仅通过阅读和观看视频来替代。具身智能可以赋予 AI 身体，并具备与物理世界的交互学习能力，这是不能通过看图、看文这些数字信息所能够弥补、习得的。2023 年，Nature 子刊刊登了由 Yoshua Bengio、Yann LeCun 等科学家联名发表的文章，提出下一代 AI 的终极挑战是通过具身图灵测试，即复现生物体的感觉运动能力，包括与世界互动、灵活的行为、高效的能源利用等[4]。具身智能被誉为迈向通用人工智能的重要一步，引发了新一轮的技术浪潮。

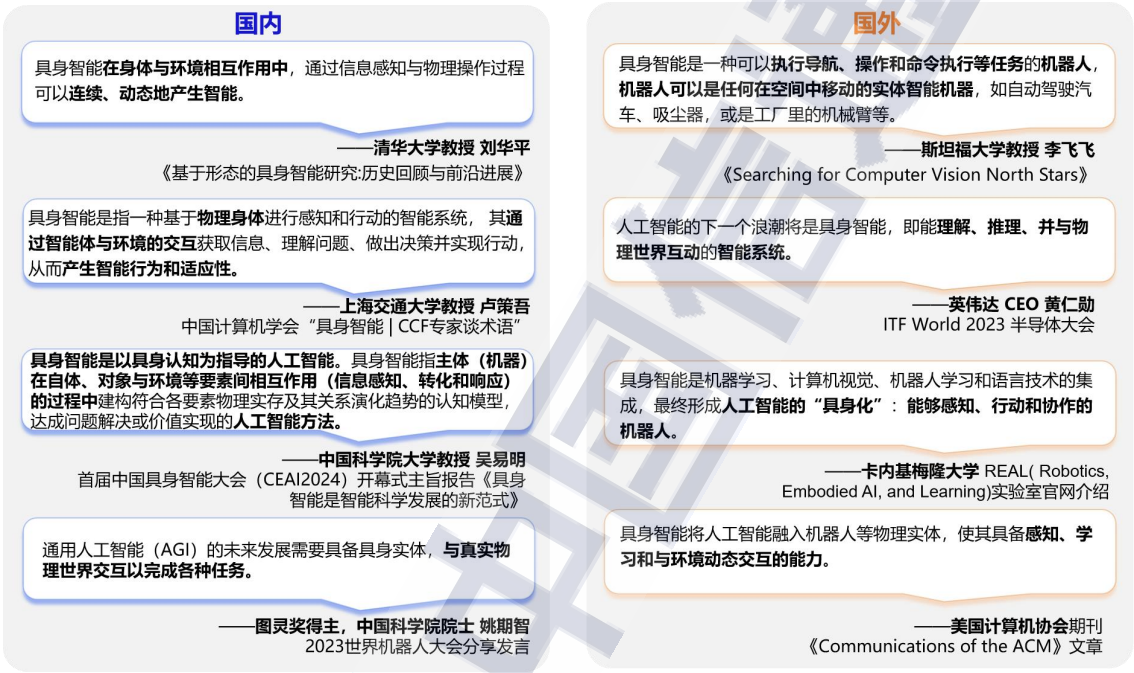
（一）具身智能的概念与内涵

1.具身智能：依靠物理实体通过与环境交互来实现智能增长的智能系统

具身智能从字面可理解为“具身化的人工智能”，**“具身”是前提，**即具有身体且能通过交互、感知、行动等能力来执行任务，具身本体的形态不必限制在外观上的“人形”，同时身体的形态也不能作为判断是否属于“具身智能”的依据。根据使用用途和场景的不同，具身智能可以有多种形态。例如，通用智能机器人，大型的工业设备加上 AI 系统，自动驾驶等多种具象化形态都属于具身智能。**“智能”是核心，**GPT-4o、Sora 等 AI 技术的最新进展，实现了对文本、视觉、语音等多模态信息的理解和转换。将这些 AI 技术

[4] https://www.nature.com/articles/s41467-023-37180-x

嵌入到物理实体如机器人上，可显著提升对环境的感知、交互和任务执行能力。先前的智能机器人，更侧重于执行特定的任务。而具身智能更强调在环境中交互能力，智能表现在物理实体能以“第一人称”主动进行感知、理解、推理、规划到移动和操作等任务。



来源：公开信息整理

图 1 国内外专家有关具身智能的观点

具身智能的发展主要来自于两个领域的交叉融合，一方面机器人的通用智能需要借助人工智能，另一方面人工智能走向物理世界需要一个身体，同时涉及到包括机械工程自动化、嵌入系统控制优化、认知科学、神经科学等多个学科的融合。这也导致了当前对具身智能这一概念的界定，不同专家的说法略有差异，**一类观点强调具身交互对智能的影响。**清华大学教授刘华平等在《基于形态的具

身智能研究：历史回顾与前沿进展》中总结：具身智能在身体与环境相互作用中，通过信息感知与物理操作过程可以连续、动态地产生智能。上海交通大学教授卢策吾曾表示通过智能体与环境的交互能够产生智能行为和适应性[5]。**另一类观点关注具身交互对解决实际问题的作用。**斯坦福大学教授李飞飞表示具身的含义在于与环境交互以及在环境中做事的整体需求和功能。中国科学院院士姚期智认为通用人工智能（AGI）的未来发展需要具备具身实体，与真实物理世界交互以完成各种任务。**但普遍认可：智能不仅体现在处理信息和解决问题的能力上，还体现在对其周围环境的感知、理解和操作能力上。**

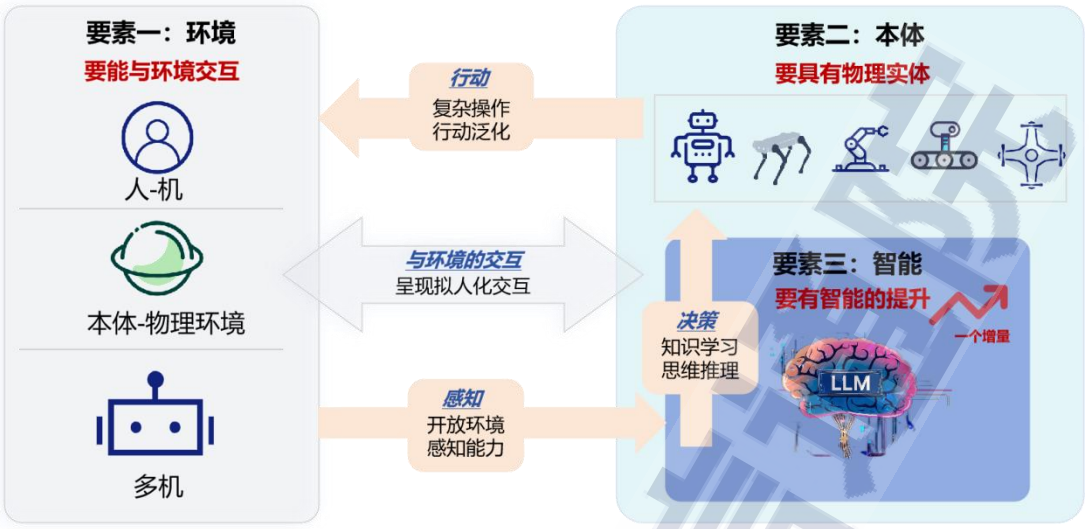
当前，针对具身智能各家观点百花齐放，但都明确了“智能”的核心地位。因此，本报告从 AI 的角度切入，认为**具身智能是指通过机器人等物理实体与环境交互，能进行环境感知、信息认知、自主决策和采取行动，并能够从经验反馈中实现智能增长和行动自适应的智能系统。**

2.具身智能与人形机器人、智能体等的概念辨析

实际上，人工智能领域的快速发展使得大模型、智能体等技术名词不断涌现，也导致关于具身智能的概念有许多容易混淆的表述。**首先，具身智能不等于“大模型+机器人”，准确来说是人工智能+机器人等物理实体。**大模型具备思维推理、计划决策、语言和视觉

[5] https://www.ccf.org.cn/Media_list/gzwyh/jsjsysdwyh/2023-07-22/794317.shtml

理解等能力，这仅能模拟大脑皮层部分功能分区的智力表现。2024 年 5 月，斯坦福大学教授李飞飞在《时代周刊》撰文写道，“大模型不存在主观感觉能力，多少亿参数都不行”。脑、身体和环境的深度耦合是产生高级认知的基础。这需要构建新一代人工智能算法，结合了脑神经、运控控制等复杂理论，推动具身智能实现认知涌现。**其次，具身智能不等于人形机器人，从载体看具身智能可以是搭载到任意形态的机器人**。人形机器人只是具身智能的一种形态，也被广泛认为是最理想的应用形态。但除此之外，比如能在家庭中行驶并与人简单交互的宠物机器人、比如 L4 自动驾驶，本质上都同时具备具身和智能两种属性。**再者，具身智能不等于智能体，两者各有交叉和侧重**。智能体（Agent）是指能自主感知环境并在该环境中采取行动以实现特定目标的实体，更强调自主性和目标导向性。智能体既可以是虚拟世界中的计算机程序（软件智能体），如聊天机器人 ChatGPT、虚拟助手苹果 Siri 等；也可以存在于物理世界的智能实体，如智能机器人。具身智能则强调智能体的具体形态和环境之间的交互作用，通过行动的物理交互能够感知和改变环境，通过行动反馈能不断学习和适应环境。具身智能的主要存在形式是物理世界中的各种物理实体。



来源：中国信息通信研究院

图 2 具身智能的“三要素”概念内涵示意图

对具身智能可以用“三个要素”来对其概念内涵进行理解。如图 2 所示，具身智能同时需要具备“本体+环境+智能”三要素，**首先强调要有具身本体**，通常是机器人等物理实体，可以有多种形态，如人形机器人、四足机器人、无人车、无人机等。本体具备环境感知、运动和操作执行等能力，是连接数字世界和物理世界的载体，同时本体的能力边界会限制智能体的能力发挥。**其次强调与环境的交互能力，**具身智能不仅能感知环境，还能通过行动来影响环境，并在与环境的交互中不断学习和适应。以“第一人称”视角去自主感知物理世界，用拟人化的思维路径去学习，从而做出人类期待的行为反馈。**最后强调一个增量，主要是智能的提升**，具身智能利用大模型的知识理解和表达能力，赋能多种形态的物理实体实现智能增长。在数据驱动的算法学习下，不断增强感知、决策以及行动能

力，并让感知与行动更紧密地连接在一起。**强调不仅通过算法和计算实现智能，还通过本体与物理世界的交互来展现和发展智能。**“展现智能”在于依赖具身本体与环境的交互行为来解决实际问题，例如机器人在通用智能的加持下将本体的行动价值最大化。“发展智能”可理解为在具身本体与环境的交互中实现可持续的智能进化。

（二）具身智能发展历程

具身智能与离身智能相互补充、协作发展共同促进了对智能的理解、模拟与扩展，从具身智能与离身智能两类研究范式在历史上多次交锋的角度出发，整体发展历程如图 3 所示。

具身智能从字面上可以拆分为“具身”+“智能”，天然具备“机器人”和“人工智能”两种属性，同时链接物理和虚拟两个世界。**从人工智能视角看，**自 1956 年 AI 概念诞生以来，智能的发展主要由符号主义与连接主义主导，两种范式从不同的侧面模拟人类的大脑，在以互联网信息处理为代表的领域取得了极大的成功。与符号主义强调“表示”和连接主义强调“计算”的离身智能不同，基于行为主义的“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[6]。**从机器人视角看，**早期机器人无需与人协同，关注点主要集中在替代人力和工业场景自动化上，以工业机器人的应用为典型代表。当前，机器人与人的交互能力和广泛

[6] http://www.aas.net.cn/cn/article/doi/10.16383/j.aas.c220564?viewType=HTML

的通用性成为发力点，探索机器人的自适应性和智能性成为重点，伴随着硬件制造和软件技术等方面的进步，以及产业链各环节的相互促进，具身智能将赋予机器人更多的智慧，不断拓宽机器人的智能边界和自主行动能力，使其更好地理解世界、自然化人机交互和高效执行任务，引领机器人进入通用智能新代际。

结合人工智能的演进历程，具身智能的发展大致可以分为三个阶段，即：早期萌芽阶段（1950s-1990s）、技术积累阶段（1990s-2022），以及技术突破阶段（2022 年至今）。

早期萌芽阶段（1950s-1990s），在对智能的激烈争论和分立研究中，形成 AI 三大学派，尚未形成成熟的智能理论。1956 年达特茅斯会议之后的一段时期内，符号主义主导了 AI 早期发展，试图用逻辑规则、符号、知识工程来模拟人类思维。这一阶段的研究集中在逻辑抽象、逻辑运算和逻辑表达等方面，如逻辑理论家、通用问题求解器、专家系统等。连接主义则强调通过神经网络模拟人类大脑的学习和计算能力，但早期的连接主义模型是简单的、浅层的网络，如感知机，难以处理复杂任务。直到 1986 年反向传播算法让多层网络的训练成为可能，重新激发了研究者们对神经网络的研究热情。然而以符号主义和连接主义为代表的计算智能的局限性很快显现出来。1988 年“莫拉维克悖论”提出人类认为困难的任务对机器来说很容易，而人类容易做到的事情对机器来说却非常困难。可以通俗地表述为：要让电脑如成人般地下棋是相对容易的，但要让电

脑有如一岁小孩般的感知和行动能力却是相当困难甚至是不可能的。1980 年代，罗德尼·布鲁克斯（Rodney Brooks）发现传统的逻辑程序在机器人导航方面显得非常缓慢和笨拙，开始直接关注通过感知和动作驱动的环境交互来设计智能机器。自此，行为主义 AI 开始发展，主张通过身体与环境的交互来产生智能。

该阶段“具身”机器人进行早期实验性尝试，关注“逻辑规则算法+机器人”实现特定应用功能。1954 年麻省理工学院生产第一台能够预先编程控制的机械臂，具备了机器人的雏形。1960s 机器人学诞生。1960 年首台工业机器人 Unimate 投入使用，在美国通用汽车公司（General Motors）的一条生产线上进行焊接工作。这一时期，开始将以符号主义为基础的逻辑规则算法与控制论结合，实现移动、对话等功能。例如 1968 年，斯坦福研究院（SRI）人工智能中心研制了世界上第一台移动机器人 Shakey。1973 年，日本早稻田大学研发了会对话的人形机器人 WABOT-1。1970s 工业机器人开始在制造业领域广泛应用。1980s 计算机硬件和传感器等技术取得突破性进展，服务机器人进入人们的视野，例如 1985 年，日本公司 Epson 推出了第一款家庭机器人“AIBO”。1990 年，麻省理工学院制作一款模仿人头部的机器人 Kismet，具有听觉，视觉和本体感受等能力。

技术积累阶段（1990s-2022），随着智能理论的完善、底层数学理论的深耕，AI 三大学派从各自突破，逐步走向取长补短的综合性研究，为具身智能发展奠定理论和算法基础。**一方面，行为主义在**

反思计算智能的局限中获得发展。布鲁克斯在 1980 年代对计算智能的根本性思考，推动了一系列以“底层智能”（即从简单的感知反应机制逐渐累积到复杂行为的生成）为基础的研究，试图参考生物的结构设计和行为方式模仿生物感官和运动能力。1991 年由布鲁克斯发表研究论文《没有表征的智能》提出智能行为可以直接从自主机器与其环境的简单物理交互中产生，而这种交互不依赖于预先设定的复杂算法。**另一方面，底层数学理论的深耕研究让 AI 算法逐渐打破桎梏，三大学派在相互补充中协作发展。**深度学习、强化学习、形态计算等理论及算法模型快速突破。与具身智能紧密相关的算法理论突破主要有三方面。**一是深度强化学习（强化学习+深度学习）**，2016 年，基于深度强化学习和蒙特卡罗树搜索的 AlphaGo 击败了人类顶尖职业棋手。**二是模仿学习（强化学习+监督学习）**，1999 年提出模仿学习，聚焦让机器人模仿人类行为的研究，通过让机器人直接模仿专家行为，可以快速、稳定地使其掌握技能，而不依赖于过多探索。**三是形态计算，**将物理形态的影响引入对智能体感知、学习、控制的作用分析，探索基于形态计算的行为生成。2004 年 C Paul 提出形态计算，聚焦双足运动形态和控制研究。

该阶段“具身”机器人快速发展，关注“行为主义”架构的仿生机器人研发和“人工智能+机器人”的智能化水平提升。1990 年，麻省理工学院制作一款模仿人头部的机器人 Kismet，具有听觉，视觉和本体感受等能力。1991 年由布鲁克斯基于“感知—行动”框架，

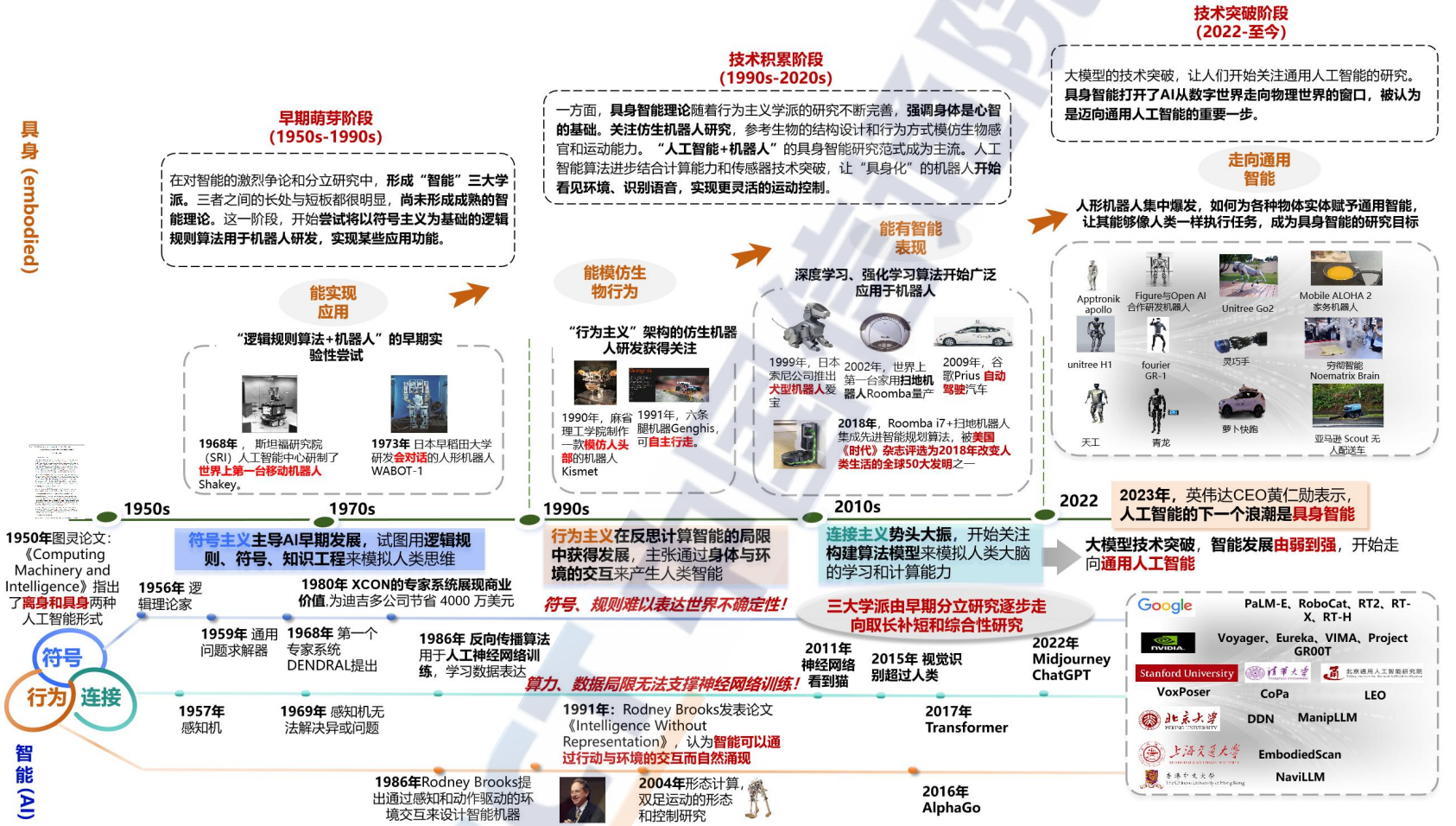
研发六条腿机器人 Genghis，可以自主行走。1999 年，日本索尼公司推出犬型机器人爱宝（AIBO）。2002 年，丹麦 iRobot 公司推出第一款家用扫地机器人 Roomba，获得当时的市场认可。2010 年代，出现了众多消费级机器人，例如扫地机器人、智能音响等。同期，无人驾驶技术取得了显著进展，特斯拉、谷歌等企业推出了自动驾驶汽车，此外无人机在物流、航拍、监测等领域也得到了广泛应用。在医疗、养老、家政等领域服务机器人逐渐成为标配。此外，在如今机器人行业的发展中，常常能看到生物学的身影，因为仿生能够帮助机器人更好地适应自然。例如 Boston Dynamics 的“大狗”、会飞的蜻蜓机器人、软体章鱼机器人等。

技术突破阶段（2022-至今），具身智能时代有望加速来临。2022 以来，以 ChatGPT 为代表大模型的通用知识和智能涌现能力为机器人实现智能感知、自主决策乃至拟人化交互方面带来巨大潜力。**大模型让具身智能的新进展井喷式涌现，大幅提高机器人的语言交互、环境感知和任务决策等关键能力。**例如，2023 年提出的 VoxPoser 模型利用 ChatGPT 理解任务语言描述并进行任务步骤分解。PaLM-E 具身多模态语言模型，将真实世界的连续传感器模态融入大语言模型（Large Language Models，LLMs）中，构建了文本和其他感知数据之间的语义联系，实现更全面的环境感知。2024 年，NaviLLM 为导航任务中语言描述、视觉观察对象以及运动轨迹等不同阶段的任务需求设计了统一的指令输入方案，让 LLMs 能够直接生成运动方

向、对象位置等行动信息。

探索具备通用智能，能够像人类一样执行任务的具身机器人成为业界共同目标。“2023 半导体大会”上，英伟达创始人黄仁勋表示 EAI 是能理解、推理、并与物理世界互动的智能系统，是人工智能的下一个浪潮。2024 年，人形机器人集中爆发，其他形态的本体如协作机械臂、移动操作机器人、仿生灵巧手、无人驾驶出租车等也显现出智能升级趋势。2024 年 3 月 OpenAI 与人形机器人初创公司 Figure 合作推出了 Figure 01 机器人，能听、会说、能与人类对话交流并且可以执行多样化任务。8 月推出的 Figure 02 凭借 GPT-4o 的大脑升级和本体的巧妙设计，如配有全方位摄像头、仿生灵巧手等，在感知、移动和操作能力上取得进一步突破。7 月世界人工智能大会（WAIC2024）上，有超过 25 款人形机器人亮相，同时在该大会上，加持了 Noematrix Brain 穹彻具身大脑的双臂协作系统展现了叠衣、削黄瓜皮等能力。百度萝卜快跑无人驾驶出租车进入商业化运营阶段，有数据显示曾单日单车峰值超过 20 单，与出租车司机的平均日单量相当[7]。

[7] https://www.lifeweek.com.cn/h5/article/detail.do?artId=231170



来源：中国信息通信研究院

图 3 具身智能发展历程

（三）全球具身智能提速发展

全球主要经济体均高度重视具身智能发展，不断提升细分领域关注度。美国紧抓人工智能基础研究，保持具身智能领域的前沿领先地位。2024 年 4 月，美国高校联合发布新版“国家机器人路线图”，旨在重振机器人技术领先地位。**日本**正在将机器人纳入社会并使机器人成为其社会基础的关键部分[8]。在人口老龄化的背景下持续聚焦机器人应用以升级制造业生产和替代人类服务。2024 年丰田研究所推出软机器人 Punyo 定位于服务人类日常生活，配备内置传感器结合柔软肢体实现全身协同操作。**韩国**出台多项政策推动以机器人和自动驾驶为核心的具身智能技术创新。2023 年发布机器人产业发展战略，擘画有关行业中长期发展蓝图。在战略中提出到 2030 年在各领域推广使用百万台的目标。**我国**加快推进新型工业化，具身智能作为新质生产力的典型代表，成为各省布局产业规划的关注重点。2024 中关村论坛年会“未来人工智能先锋论坛”上，北京市海淀区发布了《打造全国具身智能创新高地三年行动方案》。

具身智能有望成为迈向通用人工智能的重要驱动力，巨头纷纷布局，产业融合加速推进。具身智能将可以充分利用大模型的优势，在新任务上实现少样本和零样本学习，有效推动“具身化”机器人向跨任务学习和多任务迁移发展。2023 年 5 月，英伟达发布多模态具身智能系统 VIMA，能在视觉文本提示的指导下，执行复杂任务、获取概念和理解边界。2023 年 8 月谷歌 DeepMind 推出机器人模型 Robotics

[8] 《东方法学》2024 年第 3 期(人形机器人法治专刊)(总第 99 期）

Transformer 2（RT-2），是全球第一个控制机器人的视觉-语言-动作大模型（Vision Language Action Models，VLAs），10 月发布 RT-X 机器人大模型。2024 年 2 月，英伟达宣布成立通用具身智能体研究实验室 GEAR，标志着英伟达正式入局具身智能领域的研究，加速人工智能具身化进程。2024 年 4 月，优必选人形机器人 Walker S 通过百度智能云千帆 AppBuilder 平台接入百度文心大模型进行任务调度应用开发，共同探索 AI 大模型+人形机器人应用。2024 年 4 月份起，北京具身智能机器人创新中心围绕具身智能基础模型、具身智能仿真应用以及大规模具身智能数据集等，开展具身智能体母平台“开物”的研发。特斯拉宣称将推进 Optimus 人形机器人的进一步应用，预计 2025 年 Optimus 正式部署到工厂[9]。

二、具身智能技术突破，重塑智能边界

具身智能技术的发展从前期模块化的 AI 算法集成，逐渐转向大模型驱动的统一技术框架，在通用性和泛化性上取得明显突破。早期实现通过集成多个“小模型”结合人工介入方式，根据场景或用途按需调用模型，来完成相应任务，如视觉层面采用目标检测算法用于识别物体、控制层面凭借强化学习、模仿学习和形态计算等传统机器人学习技术，让机器人能够在没有人为干预情况下做出最优行动决策。这一阶段的技术发展主要是为了满足日益增长的机器人应用需求，试图为机器人赋予智能化元素，使其不再局限于固定的自动化机械操作。**大模型出现后，**具身智能逐渐将不同模块的功能融合到一个统一框架

[9]https://www.iotworldtoday.com/robotics/tesla-optimus-humanoid-robot-draws-crowds-at-world-ai-conference#close-modal

下，利用大模型潜在的知识理解和表达能力，实现了自然的语言交互，无感的多模态信息处理与转换，甚至可以对语言、视觉、触觉、听觉等各种感官信息进行统一处理，并通过融合机器人轨迹数据等运动经验，可以执行具体行动操作。

来源：中国信息通信研究院

图 4 具身智能技术体系

具身智能技术体系如图 4 所示，可分为“感知—决策—行动—反馈”四个模块，四个模块形成一个闭环，在与环境的不断交互中，实现对环境的重构映射、自主决策和自适应行动，并从经验反馈中不断学习进化。**具身智能的技术尚处于多条路径探索发展阶段，可以类比于自然语言处理领域的“BERT”发展时期。**BERT 和 GPT 的出现让自然语言理解能力有了里程碑式突破，但仍有多条技术路线在并行发展，直到 ChatGPT 的出现。目前具身智能也正在围绕“感知+决策”、“感知+决策+行动”等并行探索多条有潜力的技术路径，探索如何打

造具备通用智能的具身智能基础模型。

（一）感知模块—赋予机器感官，实现多模态感知泛化

感知模块是具身智能的“信息采集和处理器”，建立对外部环境的感知和理解，为可靠的决策和成功完成行动提供支持。感知模块主要任务包括对象识别、位置定位、场景理解、环境重建和状态监测等。**感知实时性和精度将直接影响决策的可靠性和行动的准确度。**例如在仓储物流场景，对象识别即识别不同的包装箱、货架、托盘和环境中的其他设备，当一批新货物到达仓库时，可快速完成分拣。场景理解即理解仓库内物体布局、货物堆放以及人员活动等情况，用于分析仓库内的货物存储情况、货架占用率等。环境重建即生成仓库的三维模型，用于规划货物导航方案。位置定位即确定自身和货物运输的目标位置。引导机器人从指定位置取货，并准确送到目标位置。状态监测即通过机器人运作中不断接收的传感数据，监测仓库内的温度、湿度、照明、障碍物、设备运行状态等，帮助及时发现并处理故障问题。

感知模块的具体实现从集成不同的 AI 算法，逐渐转向使用多模态模型来处理和融合多维传感数据。感知模块需要对来自 RGB 摄像机、激光雷达、深度摄像机、重力传感器等多种外接传感设备的输入数据进行处理，进而从不同模态的数据中获得多维环境信息。由于不同模态的数据存在格式差异性、时间和空间的不一致性以及干扰噪声等问题，多模态数据的融合以及统一的环境概念表达面临挑战。

先前，通过组合各个 AI 算法来执行不同的感知任务，实现针对特定场景的环境感知和理解。这一阶段，通常在空间有限、场景结构

相对固定、且动态变化相对可控的封闭场景下，预先构建目标检测、姿态估计、3D 重建等 AI 算法模型，组合用于识别环境中的对象，理解场景和环境状态变化。例如移动机器人在导航时至少需要理解有什么物体和目标位置在哪里。常见解决方案是采用计算机视觉技术如 YOLO 负责物体的识别和定位，采用 SLAM 技术生成环境的三维地图，帮助规划导航路径。

当前，大模型通过对多模态信息的统一处理与灵活转换，实现对环境的多模态感知泛化。视觉基础模型（Vision Foundation Models，VFMs），如 CLIP、MVP、R3M 等，帮助大模型获取预训练好的视觉表达，提供视觉辅助信息。EmbCLIP、CLIPort、RoboFlamingo 等均采用这一方法。视觉语言大模型（Vision Language Models，VLMs）支持处理图像、3D 数据、状态信息等多模态数据，将现实世界数据转化为可被 LLMs 理解的表达，弥合了语言符号指令与视觉感知信息间的差距，例如直接根据语言指令中的“苹果”一词识别环境中苹果区域和位置信息等。**动态学习作为 VFMs、VLMs 等的学习策略，可以为模型注入时间维度的动态变化信息，提升模型视觉表达的丰富度。**Vi-PRoM[10]在对比预训练基础上联合动态学习，通过捕捉时间上的视觉变化，来理解视觉的语义信息。大模型结合世界模型能够实现感知预测，模拟环境的动态变化。3D-VLA[11]在 VLM 之上结合 3D 世界模型的视觉生成能力，能够想象和预演环境动态变化与行动后果间的关联。随着多模态处理能力的演进，具身智能将融合语言与视觉、听觉、

[10] https://arxiv.org/pdf/2308.03620.pdf

[11] https://arxiv.org/abs/2403.09631

触觉等感官信息，更容易实现可变环境的自适应和未见任务的行动泛化。2024 年 1 月 UCLA 提出多模具身智能大模型 MultiPLY 具备包括视觉、听觉、触觉在内的多模态感知能力，能够与 3D 环境全面交互。

（二）决策模块—提升机器脑力，实现人类思维模拟

决策模块是具身智能的“指挥中心”，接受环境感知信息后，完成高级任务规划和推理分析，并生成逐步决策指令来控制行动。决策模块的主要任务包括任务规划和推理分析等。**可靠的决策依赖于感知模块对环境的准确理解。**尤其在动态变化的环境中，丰富的感知信息能带来明显增益。北京大学提出的视觉导航技能 PixelNav 利用多模态大模型提取环境中的视觉语义、物体线索等多视角的感知信息，实现了对任意类别物体的导航任务规划和策略推理[12]。**精细决策可以增强行动的精准度和可控性。**例如，中国人民大学提出了可泛化铰链物体操纵的具身智能框架，其中的决策模块在基于运动学信息推理操纵步骤后，可进一步生成精确的 3D 操纵关键点，解决了复杂铰链物体的底层操纵难题。

决策模块的具体实现从依靠人工知识的编程决策、专用任务的算法设计，转为以大模型为核心的机器智能决策。决策模块负责接收来自感知模块的各种信息，并结合任务目标做进一步处理后，制定具体的行动策略。决策模块的灵活性和适应性直接影响着具身智能系统的智能化水平。一个高度智能化的具身智能系统，能够根据环境和任务需求的变化，实时调整决策；能够不断获取感知信息和行动经验，学习和优化决策；能够有效协调和控制其他各个模块，确保决策效率。

先前，人工编程决策和强化学习算法设计在环境状态变化可控的

[12] https://arxiv.org/abs/2309.10309

条件下，能够完成简单任务决策。尤其在一些明确、可定义的任务场景中，人工编程决策可以发挥作用。例如人工编写的 A*算法和 Dijkstra 算法，广泛用于完成简单的导航和路径规划任务。通过预编程的任务脚本用于完成工业产线任务的顺序执行决策。但这类完全定制化的算法很难应对动态变化的环境和未知情况。随着强化学习方法发展，基于近端策略优化算法、Q-learning 算法的强化学习方法在具身智能自主导航、避障和多目标收集等任务中[13]，可以获取运动序列样本进行策略更新，展现更好的决策灵活性。但对复杂环境的适应能力、决策准确度和效率仍然受限。

当前，大模型在环境动态变化的条件下，能够模拟人类思维完成复杂任务决策。大模型在大规模的互联网数据上进行预训练后展现出强大的思考和推理能力，能够像人类一样做出更加智能和适应性的决策。**一是**利用 LLMs 的语言理解能力，弥合了自然语言和机器指令间的语义鸿沟。俄亥俄州立大学推出的 LLM-Planner[14]提出了高级和低级两层的任务规划策略，其中高级规划器利用 LLM 对用户的任务描述生成自然语言规划，低级规划器将子任务转化为行动指令。LLM+P[15]利用 GPT-4 能直接将任务规划转化为机器能够理解的规划领域定义语言（PDDL）描述。**二是**利用 LLMs 的代码生成能力，替代人类的复杂编程环节。Code as Policies[16]利用 LLMs 生成任务策略代码，调度其他模块或底层 API 函数。**三是** LLMs 结合其他辅助信息，更好地适应实际环境的复杂性和动态变化。Inner Monologue[17]将视觉

[13] http://kzyjc.alljournals.cn/kzyjc/article/pdf/20220203

[14] https://dki-lab.github.io/LLM-Planner/

[15] https://arxiv.org/abs/2304.11477

[16] https://arxiv.org/abs/2209.07753

[17] https://arxiv.org/abs/2207.05608

的检测结果整合到 LLMs 的提示词中进行规划或重新规划。PHYSOBJECTS[18]利用 LLMs 生成初始规划，并通过查询日常物体的物理概念（如材料、易碎性），在 VLMs 的帮助下进行下一步决策。3D-VLA[19]整合了 3D 空间信息，能够完成 3D 空间推理和交互决策，如把最远的杯子放在中间的抽屉里。

（三）行动模块—提升机器自主行动能力，实现精细动作执行

行动模块是具身智能的“执行单元”，负责接收决策模块指令，并执行具体动作。行动模块的主要任务包括导航、物体操作和物体交互。导航任务即通过四处移动，寻找目标位置，例如把客厅里的椅子放到第二个阳台上[20]，在物流运输、车间搬运、家庭清洁、家庭伴随等场景中都有涉及。物体操作需要接触物体并通过操作改变物体状态，如简单操作扔、推、滑等，复杂操作炒菜、转笔等。物体交互指通过交互才能完成的操作任务，如拉开抽屉、按按钮、旋转阀门等。物体操作和物体交互常见于家务劳动、工业分拣等场景。

行动模块要实现精细的动作控制面临很大挑战，具体实现可分为三条主要技术路线。在真实环境中，机器人行动能力受到复杂环境以及环境动态变化的限制。环境中温度、湿度、摩擦力、障碍物、部件磨损等环境属性和条件的动态变化，均会导致感知观测误差和决策准确性，进而影响任务执行的成功率。当前，仅依赖大模型仍难以很好应对操作对象的变化和复杂的操作要求，需要考虑优化奖励策略，以及整合环境、运动等多样化信息。

[18] https://arxiv.org/abs/2309.02561

[19] https://m.thepaper.cn/newsDetail_forward_26788704

[20] https://arxiv.org/pdf/2108.04097

一是强化学习与主流 Transformer 架构结合，应对泛化性挑战。强化学习范式一直主导了机器人行动学习技术的研究，让机器人在与环境的交互中，不断试错、学习和优化策略，并依据奖励策略不断优化动作执行结果。然而，强化学习方法在面对未知环境时存在泛化差距，难以将学习到的行动经验迁移到新的、以前未见过的环境中[21]。最近，一些研究工作利用主流 Transformer 对多模态数据的通用表达和转换能力，驱动强化学习方法实现多任务泛化。例如 Q-Transformer 采用强化学习方法在大规模多样化的真实世界数据集上训练 Transformer 模型，能够自动积累经验，快速适应不同任务。

二是大模型作为强化学习的辅助工具，突破强化学习发展瓶颈。一方面，利用 LLMs 设计或塑造深度强化学习的奖励策略，避免了人工费力设计策略函数的过程。EUREKA 利用 GPT-4 自主设计的奖励函数在 83%的任务中优于人类专家设计的奖励。这种奖励能够让具身智能完成很多之前不容易完成的任务，如转笔、打开抽屉和柜子、抛球接球和盘球、操作剪刀等[22]。**另一方面，**大模型的先验知识和多模态信息提取能力解决了强化学习方法的低样本效率问题。例如多模态大模型能够处理语言提示、目标图像、轨迹规划策略、3D 热力图等各种类型的数据，并将其转化为监督且能够灵活地将其纳入反馈机制来优化策略。

三是视觉语言动作大模型实现了从语言到可执行动作指令的直接转换。VLAs 是对 LLMs 和 VLMs 的进一步扩展，将互联网知识、物理世界概念与运动信息融合到统一框架中，能够直接依据自然语言描述生成可执行的动作指令。Prompt2Walk[23]将语言与运动信息结合，

[21] https://arxiv.org/abs/2010.10814

[22] https://www.jiqizhixin.com/articles/2023-10-23-5

[23] https://prompt2walk.github.io/

使用 LLMs 通过收集的少量运动数据提示直接输出关节角度。英伟达发布 VIMA[24]可以通过多模态的输入提示来学习操作动作。RT-2[25]采用模仿学习的范式将 VLMs 融合机器人运动数据，能够直接生成可被机器人识别的操作指令。然而，这类解决方案仍面临较大的成本挑战。谷歌 RT-1 的数据收集使用了 13 个机器人且耗时 17 个月[26]。

（四）反馈模块—拓展机器交互通道，实现自主学习演进

反馈模块是具身智能的“调节器”，通过多层交互不断接收来自环境的反馈经验并进行调整和优化，以提高对环境的适应性和智能化水平。反馈模块将环境交互的经验用于优化感知、决策和行动模块，实现感知增强，策略优化和行动适应。对感知模块而言，环境交互中能够持续反馈视觉、触觉、听觉等各种感官数据，从而提高对外部环境变化的敏感度，实现更准确且更细致的环境感知。例如配备了摄像头和触觉传感器的机器人，通过不断接收和处理视觉图像和触觉反馈，可以更准确地识别物体的形状、位置和材质。对决策模块而言，环境交互中能够持续反馈行动结果、获取语言指令等，从而快速识别有效和无效策略，做出更智能的决策。例如在家庭服务中，通过持续收集用户的生活习惯和偏好等反馈信息，来优化照明、温控和安防策略，为用户提供更舒适和智能的居住体验。对行动模块而言，接收反馈信息后，会根据决策模块的指令灵活调整动作，确保在不确定和多变环境中也能高效运转。例如调整运动轨迹、改变力量输出或改变动作顺序，以应对实时的环境变化和任务需求。

反馈模块主要依赖大模型来加速反馈经验的学习，形成闭环的优

[24] https://vimalabs.github.io./

[25] https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/

[26] https://www.sohu.com/a/617629740_129720

化过程。一是通过大模型处理收集到的真实交互数据，实现更细致的环境感知。环境交互层面，大模型在与环境交互的过程中，持续收集对象位置、动态和空间关系等细节物理概念信息，并将其转换为奖励信号，实现高保真的动态环境模拟。剑桥研究实验室的 LanGWM[27]将不同时间段的观察、语言和行动纳入记忆反馈模块，增强对环境状态的动态感知。**二是通过大模型处理交互信息，实现模仿人类反馈的决策。人机交互层面**，LLMs 及 VLMs 大模型允许以更自然的方式将环境属性、状态或各种模态的输入提示信息转化为特定的行动指令信号，降低了从交互经验到决策优化间的反馈链路复杂性。斯坦福大学最新的具身智能系统 YAY Robot[28]能够基于人类语言反馈及时调整策略。例如在“清洗盘子”任务中，通过口头反馈使清洁力度明显更强。**多机交互层面**，大模型在具身智能中主要用于解决单智能体的任务规划问题。然而，由于大模型知识和特定的具身环境不对齐，大模型产生的规划往往难以在环境中执行。中国电信李学龙教授团队提出了一种通过多智能体强化学习的大模型反馈方式，大幅提升群体沟通和环境反馈的效率[29]。**三是大模型获取交互行动经验，学习最佳行为策略。**当 LLMs 生成行动决策后，可以通过强化学习反馈，根据价值函数对行动进行重新排序，以最大化行动的累计奖励。谷歌的 SayCan[30]利用操作完成程度的价值度反馈来不断优化行动选择。

[27] https://arxiv.org/abs/2311.17593

[28] https://m.thepaper.cn/newsDetail_forward_26967077

[29] https://arxiv.org/pdf/2405.14314

[30] https://say-can.github.io/

（五）支撑要素—本体、数据和软硬件底座共同构成具身智能发展基础

本体作为具身智能的任务执行机构，负责对环境的主动感知并执行具体动作。本体配有的传感器和核心零部件等硬件组件，以及自身形态对具身智能的能力发挥有直接影响。**短期来看，硬件的基础能力足以支撑具身智能的研究和落地验证。在运动层面**，电机、丝杠、减速器等执行器不仅能够支撑机器人的稳定运动，且成本可控，移动机器人和四足机器人的研发费用只需万元左右。人形机器人“天工”每条胳膊上的 3 个关节、每条腿上的 6 个关节里有机组合了电机、减速器、编码器、控制器四大关键零部件，实现了 6 公里/小时的拟人化稳定奔跑。**在操作层面**，机械臂技术较为完善和成熟，被谷歌、清华、斯坦福大学等机构广泛用于科研实验中，也在工业场景里得到了落地验证。灵巧手的进展相对缓慢，目前更多关注通过增加关节自由度和传感器配置来提高操作灵活性，在仿生人手结构的精巧度、类似皮肤的柔性感知方面仍有较大挑战。**不同形态的本体适用于不同的环境和任务需求。**例如轮式机器人在平坦地面上移动效率高，四足、双足机器人在不平坦地形上具有更好的适应性，空中无人机适用于高空检测，多关节机械臂、仿生灵巧手等可以执行更复杂的动作，而人形机器人在手的操作能力和脚的移动能力上具有最高的自由度和最强的通用性，作为更容易被人类接受的本体形态，不仅可以执行复杂的抓取和操作任务，也常用于社交互动、情感陪伴以及交互服务等场景。**长期来看，**具身智能从落地验证走向商业化的过程中，需要本体硬件能力

的持续提升和应用形态的恰当设计，实现研发成本、执行效率和通用性的平衡。硬件的抗冲击能力、灵巧手的操作能力、触觉和力觉传感器的集成等仍需不断地提升，例如英国 Shadow Robot 公司推出的 Shadow dexterous hand 是目前最成熟的商品化多指灵巧手之一，拥有 24 个自由度，配备指端触觉传感器，但仍然不能实现与人手相当的自由灵活程度和操作能力。

数据对具身智能的能力提升和应用探索至关重要。在能力提升上，高质量的多模态数据驱动具身智能感知、决策及行动控制能力快速提升。上海人工智能实验室在研究工作 EmbodiedScan 中提出更大、更真实的数据集、更多样的场景和更详尽的标注可以显著提升具身智能的 3D 感知能力。北京大学构建了涵盖 132 万条的灵巧机械手抓取数据集 DexGraspNet，在规模、稳定性和多样性上明显优于现有数据集。已有算法在该数据集上训练后能提升抓取成功率，最高可达 10%[31]。**在应用探索上，**数据是具身智能快速适应新的环境和任务的关键。谷歌联合全球机构汇集了 22 种不同机器人类型的数据，构建了最全面的具身智能数据集 Open X-Embodiment，并用于训练通用具身智能大模型 RT-X。RT-X 可以在无需任何训练数据或极少训练的情况下，泛化到特定任务上，如仓库搬运、防爆救险、家庭护理等。北京具身智能机器人创新中心正在组织建设大规模的高质量具身智能数据集，支持机器人实现长行程的任务规划能力。

具身智能数据按采集方式主要分为真实数据和仿真数据两大类。

[31] https://arxiv.org/abs/2210.02697

短期来看，仿真数据用于解决简单任务，助力具身智能实现 0 到 1 的突破。尤其针对跑步、跳跃或跳舞等简单的运动任务，仿真数据已经足够支撑。仿真数据的优势在于获取快、成本低且数据量大。然而，实际研发过程对仿真效率和成本投入的综合考量，导致现在仿真数据的模拟质量仍然粗糙。例如为保证仿真效率，会简化和近似处理对环境中的物理属性和三维场景的建模。同时高逼真的环境模拟也需要高性能的 GPU 显卡和大量的计算资源支持。**长期来看，真实数据对处理复杂任务不可或缺，推动具身智能实现 1 到 N 的深度应用。**例如炒菜、装配等复杂任务涉及复杂操作和动态变化，仿真和现实之间的微小差异都会影响策略的有效性。斯坦福大学家务机器人 Mobile ALOHA 推椅子的任务成功率有 80%，而炒虾只有 40%，在执行这类复杂家务活动时，仍需要收集人类操控机械臂的动作数据来模仿相似的动作[32]。

软件工具驱动具身智能系统的灵活开发和高效测试。数据准确阶段，数据采集、生成、处理和分析等全链路工具让复杂的数据工程化任务变得简单高效。LabVIEW 传感器编程软件通过丰富的硬件接口和驱动程序，支持接入各种传感器进行数据采集。Unity3D、Omniverse、Gazebo 等 3D 仿真引擎可以产生大量的仿真数据，缓解真实数据的获取难题。**技术研发阶段，**强大的软件生态系统显著提升技术研发效率。ROS 和 ROS 2 是目前广泛使用的机器人操作系统，通过标准化的接口能快速集成各种传感器、执行器和其他软件工具，简化了复杂具身

[32] https://mobile-aloha.github.io/resources/mobile-aloha.pdf

智能系统的开发和测试过程。例如在 ROS 2 集成英特尔的 OpenVINO 视觉推理工具，可使具身智能系统具备实时人脸识别、目标检测和人体姿态估计等能力。**技术验证阶段，**具身智能仿真测试平台提供了一个安全、高效且低成本的测试环境。英伟达 Isaac Sim 和斯坦福大学的 BEHAVIOR-1K 等仿真测试平台，能够真实地模拟多样化任务活动，创建高保真 3D 环境，准确再现具身智能在真实世界应用时可能遇到的情况。**落地部署阶段，**为了让物理实体更好地承载 AI 模型的推理和计算，需要并行计算、低比特量化、模型压缩、3D 空间计算等配套算法支持，优化端侧的实时性、多模融合和 3D 空间计算能力。

通用计算平台为具身智能系统的复杂计算和可靠运行提供有力支持。具身智能对计算系统的灵活性、计算效率和可扩展性方面有着严苛要求[33]。在传感数据处理计算层面，具身智能需要依赖不同硬件模块同步处理多个传感数据，才能有效融合各个传感器的环境感知信息。NVIDIA Jetson Nano 计算模组支持来自多个高分辨率传感器的数据并行计算。在模型决策推理层面，需要高性能的端侧计算芯片支持大规模推理计算和实时决策。英伟达 Jetson AGX Orin 模组在边缘端的计算能力，可与内置 GPU 的服务器相比。在数据流处理层面，需要分布式数据处理满足不同应用场景下的通信计算需求。英伟达 Isaac 机器人平台通过适配 ROS 2 软件生态，引入数据分布服务（Data Distribution Service，DDS）通信协议，实现低延迟的数据通信计算。

[33] https://airs.cuhk.edu.cn/article/1119

（六）安全与隐私保障—确保具身智能执行安全可信

安全和隐私保障能力是具身智能成功应用和推广的关键。具身智能系统在真实世界中执行任务时，需要遵守道德规范、保护用户隐私不受侵犯、确保用户的数据安全以及系统可靠运行。在道德规范方面，具身智能系统的设计和应用需要遵循伦理原则，确保其行动不会对人类产生不利影响。在隐私保护方面，真实数据收集的过程中要做好数据脱敏和匿名化处理，并制定清晰透明的隐私政策，让用户了解系统如何收集、使用和保护他们的数据。在数据安全方面，应采用数据加密和隐私计算技术，保障数据在存储、传输、使用以及处理过程中的安全性。

在系统可靠运行方面，系统部署前可以进行大量的仿真模拟测试，在仿真环境中再现真实世界应用时潜在的安全问题。但很难完整模拟所有情况。以将盘子放入洗碗机这一任务为例，从初始状态“找到盘子”到实现最终状态“盘子在洗碗机里”，有无数种状态变化，很容易存在潜在风险，例如机器人在行动中撞倒障碍物导致任务失败。系统运行中，可以通过对话、指令输入等方式进行干预，纠正错误决策。清华大学提出人机协同框架 HumanTHOR，该框架使人类可以通过虚拟现实设备在虚拟 3D 环境中与机器人协同工作，解决用户信任问题。此外，系统迭代时可以利用收集到的真实任务数据进行反馈学习，让系统对齐真正的任务需求。

三、具身智能在各领域的应用前景

具身智能通过模拟人类大脑的“智能”和不同形态的机器人“身

体”，将在多个领域释放出巨大的应用潜力，成为迈向通用人工智能的重要一步。相比于传统基于 AI 视觉及特定场景预训练的机器人，**具身智能具体表现在**：一是不再依赖预定义的复杂逻辑来管理场景；二是能形成学习进化机制，持续获取交互反馈来实现环境自适应；三是能通过身体与环境交互产生新的交互数据，并用于实现智能增长。目前的最新进展仅是基本具备三个表现，**尚未出现功能完善的商业化产品。但可以预见的是，**随着技术的不断突破，**具身智能将使得各种物理实体显现出四个能力增长点**，即对环境动态变化的自适应能力、多任务行动的泛化能力、交互方式的拟人化表现和更高的任务执行效率。这些能力增长点**有望带来更高的应用价值和广阔的市场空间**。

（一）工业制造领域：打破人机协作瓶颈，实现智能化柔性适配

工业制造领域具身智能有望成为新型工业化的关键核心和有效抓手。具身智能将使得机器人从“能动”到“能干活”转变，以此来为工业制造业的智能化升级提供强大支持。以机器人和机械臂等为载体的具身智能应用，将使得工业制造过程更加智能化、灵活和高效。微软正计划将 ChatGPT 的能力扩展到机器人领域，通过自然语言和 ChatGPT 交流，使用 ChatGPT 来控制机械臂、无人机、移动机器人等。阿里巴巴也在将千问大模型接入工业机器人，为机器人提供了推理决策的能力，从而有望让机器人的灵活性和智能性大幅提升。西安中科光电推出智能焊接机器人，目标是替代焊接工人在工厂自主进行焊接作业。

具体来说，具身智能将变革人机协作模式，实现更安全、智能化的柔性制造流程。一方面，具身智能从根本上打破人机交互的语义隔离，以高效的人机沟通方式提高整个协作过程的安全性。人类可以用自身习惯的方式与具身智能工业机器人沟通，如自然语言、肢体语言、动作示范等。机器能够更及时、更好地理解人类意图，提前做出适应性的安全控制动作，降低错误发生概率。例如，香港理工大学利用 LLMs 让机器人实现更直观、灵活的人机交互，可以适应工业场景中非结构化的作业环境，如环境的频繁变化、不同类型的操作任务等。发那科 CRX 系列协作机器人在感知到机器人本体与人类或其他物体轻微接触时，便会立即停止运动，从而防止伤害的发生。**另一方面，具身智能工业机器人将能够替代人类成为工业生产线上最柔性的执行机构。**具身智能将使工业机器人实现智能化的柔性制造，能够不断观察周围环境，并在执行任务过程中自动更新决策和优化行动，让工业产线需要的人工干预程度降低。这种生产模式上不仅具备高度适应性，还具有更高的生产效率和制造精度。例如，特斯拉 Optimus 人形机器人在特斯拉电池工厂工作时，能够提高电池生产效率，降低人为因素对产品质量的影响。

（二）自动驾驶领域：适应开放交通环境，实现安全可靠智能驾驶

自动驾驶领域，具身智能有望通过提升开放交通环境适应性实现安全可靠驾驶。自动驾驶汽车不仅要能感知周围环境，还需要根据感知到的信息做出快速且准确的决策，并通过执行系统来实现车辆的操

控。具身智能能够将这些环节紧密结合，形成一个高效、协同的工作流程。例如，特斯拉的自动辅助驾驶系统 Autopilot 通过车载传感器和摄像头收集数据，实现自适应巡航控制、车道保持辅助和自动变道等功能，显著提高了驾驶的安全性和便捷性。以谷歌 Waymo 自动驾驶技术融合感知、定位、规划、控制等，能够在行驶过程中实时识别行人、车辆、交通信号灯等关键信息。通过这些信息，系统能够预测潜在的风险并提前做出规避动作，大大提高了行车的安全性。

具体来说，具身智能通过融合感知、决策和执行等功能，将提升自动驾驶系统的整体性能。一是实现对动态环境的全面感知和高度泛化。具身智能自动驾驶系统能够理解环境中广泛的物理世界概念，并在与环境的实时互动中，适应不同的光照、天气等条件。**二是实现可靠的智能决策和可控行动**，具身智能自动驾驶系统具备高效的多模态信息提取能力，将最大化利用车辆搭载的各种传感器优势，综合考虑各种信息实现合理可靠的驾驶决策和及时的行动控制。**三是实现高度智能的自主学习适应。**车辆在与环境交互的过程中，不断收集新的数据和经验，通过学习和适应，不断提高在复杂开放交通环境下驾驶性能和智能水平。特斯拉创始人马斯克宣布将推出无人驾驶出租车 Robotaxi，或将引领具身智能自动驾驶的跨越式发展。

（三）物流运输领域：优化仓储物流产线，实现高效货物运转

物流运输领域，具身智能有望降低流通成本，成为形成高效、快捷、智能化的物流体系的关键因素。当前物流领域包括拣选机器人、

叉取机器人、搬运机器人、料箱机器人等。具身智能技术的赋能，可以在仓储、装卸、搬运、分拣、包装、配送等环节提升工作效率和管理水平。物流机器人将更加智能化，具备更强的自主决策和学习能力，能够适应更复杂、多样化的任务，不仅局限于传统的仓储和物流行业，还将渗透到制造业、农业、医疗、教育等领域，提高各行各业的智能化水平和生产效率。例如，亚马逊近期在其仓库运营中，已经在测试由其投资的公司 Agility Robotics 开发的人形双足机器人 Digit，综合全面完成主要包括卸载货车、搬运箱子、管理货架等任务，大幅提高了仓库作业的效率。

具体来说，具身智能将助力仓储物流产线的智能化升级，实现安全、高效且可持续的物流运输作业。一是更好的环境适应性。在具身智能的加持下，物流移动机器人对环境感知、路线规划和运动导航能力将明显增强，更好地适应可变环境，识别多点目标，自主调整路径并能够及时避障。丹麦 Capra Robotics 公司最新推出的 Hircus 移动机器人平台，实现厘米级的位置精度定位，并首次能够同时适用室内室外两种环境。**二是更灵活高效的工作模式。**物流移动机器人可以凭借具身智能同时执行多点、多任务甚至多层任务。美国 Brightpicks 公司推出的自动移动机器人（AMR）可以无缝进行商品订单识别和拣选任务，整个过程无需员工人工推车拣选。**三是低成本的连续作业。**具身智能移动机器人可以连续 24 小时待机，随时投入生产，同时凭借其高度的智能化水平，将避免作业过程中的人工监督成本。美国初创公司 Agility Robotics 的 Digit 人形机器人在亚马逊仓库打工连续工作

长达 7.5 小时，并在任务执行时实现了 100%的自主性，据称其工作效率已达到人类速度的 75%，任务完成成功率高达 97%。

（四）家庭服务领域：解放人类双手束缚，实现全场景的智能家务服务

家庭服务领域，具身智能通过高级的认知和行动能力实现真正意义的定制化服务。家庭服务机器人的发展已经从基础的扫地机器人演变到现在可以进行地面清洁、物品搬运和基本家务的多功能机器人。未来，通用具身智能机器人能够（拟人化）感知、使用传统工具、在非确定环境下自主执行任务，属于全场景家庭助手，像汽车一样走进千家万户，成为每个家庭不可或缺的生活伙伴和帮手，如康复、家务类任务等。例如，1X 公司与 OpenAI 公司已经在深度合作，开发一款具身智能类人机器人 EVE，可以实现对人类日常工作环境的认知理解，在与环境交互的过程中学习、纠正、收集数据，完成自主居家、办公帮手任务。

具体来说，具身智能使得家庭服务机器人真正成为人类友好的智能助手，提供智能化、人性化的全场景家庭服务。家庭服务机器人在具身智能的不断发展下，已经从基础的扫地机器人演变到可以进行全面清洁、基本家务和餐饮服务多功能机器人。**一是模拟人类执行多种家务。**2024 年 2 月，美国谷歌和斯坦福联合推出家务服务机器人 Mobile ALOHA 2，通过移动底座在大的空间范围内实现长距离操作，同时能够模拟人类双手进行备菜、翻炒、出锅，洗衣、逗猫、浇花等。据智元机器人官网介绍看，其推出的智元绝尘 C5，集扫地、洗地、

尘推等多种清洁能力于一身，并且能够在人工最小干预的情况下，在复杂的环境中自主完成充电、加排水、清洁污水箱等任务。2024 年 4 月，星尘智能在发布视频中展示了 Astribot S1 家务服务机器人的能力，能够执行叠衣服、物品分类、烹饪、吸尘清洁以及叠杯子等家务活动。**二是替代人类进行体力劳作。**2024 年 4 月，越疆科技发布 X-Trainer 具身智能机器人，在演示视频中自主完成了刷盘子任务，利用视觉语言大模型从带有红色食物残留物的盘子、放在黄色盘子上的海绵，以及后面挂着碟子的金属架等复杂任务描述中，推理出清洗盘子并收纳到金属架的任务。**三是陪伴人类满足情感需求。**日本家庭陪伴机器人 LOVOT 主打情感陪伴功能，每台全新的 LOVOT 会呈现不同的性格特点，有的热情和主人聊天，有的害羞不敢说话，甚至会根据与人们相处的时间、互动的过程展现不同的情感状态。截至 2023 年，在日本的售卖量超过 1 万台。

（五）医疗康养领域：应对老龄化问题，实现拟人化交互服务

医疗康养领域，具身智能正逐渐成为应对老龄化挑战、提供高质量医疗服务的关键技术。当前，具身智能技术已被应用于自动化手术机器人，这些机器人能够执行精确的切割和缝合操作，极大地提高了手术的安全性和效率。达芬奇手术系统是此类技术的典型代表，它允许外科医生通过高度精确的机器臂进行远程手术操作。未来，具身智能有望极大地改善医疗服务的质量和效率。不仅能够提供拟人化的交互服务，提高老年人的生活质量和幸福感，还能有效减轻医护人员的

负担，提高医疗服务的效率和质量。例如，日本公司 AIST 已推出外形像海豹的 Paro 治疗机器人，被用于老年护理和儿童医院，提供情感支持和陪伴，帮助缓解焦虑和孤独感。美国公司捷迈邦美推出用于机器人辅助肩关节置换手术的 ROSA® Shoulder 系统，能够帮助外科医生灵活地使用解剖或反向技术进行全肩关节置换术，并实现精确放置以改善手术结果。

具体来说，具身智能让医疗康养机器人实现拟人化的交互模式，可以提供人性化的服务体验。一是实现个性化的情感社交互动。迪士尼推出情感互动型机器人“瓦力”，在儿童大小的身体中融入具有情感表达的肢体动作。韩国公司 Hyodol 推出专门为老年人服务的 AI 伴侣娃娃，旨在缓解老年人孤独感和阿尔茨海默病问题，在大模型支持下能够与老人进行完整对话。**二是提供人性化的服务体验。**美国 Glidance 开发的导盲机器人 Glide，采用完全人性化的设计，它不会主动拉动用户，而是根据用户的动作做出响应，确保用户在导航过程中保持控制和主动性。2024 年 2 月，日本丰田研究所发布软体人体机器人 Punyo 为缓解老年人劳动力不足问题提供了解决方案，帮助搬运大型、重型和笨重的物品，例如搬抬箱子、堆叠两个收纳箱以减少空间占用、搬起水桶等操作。

（六）其他领域：从赋能到变革，推动各行各业创新与转型

除以上领域之外，科研、应急等领域的具身智能应用也将带来深刻变革。**科研探索领域，具身智能能够自主执行科研实验，进行长时**

间连续工作，从而加速科研进程。在极端环境中，如深海和太空，具身智能机器人可以代替人类进行探索，发现未知的科学奥秘。同时，它们还能承担高危或繁琐任务，确保人员安全并提高工作效率。此外，具身智能机器人还能高效地收集和分析数据，为科研人员提供重要信息。通过机器学习和人机协作，具身智能机器人正成为科研领域的重要助手，推动科学研究的深入发展。例如，美国宇航局（NASA）的毅力号探测器在火星表面不仅采集到岩石，还收集到火星空气样本，将帮助了解地球以外的其他星球气候是如何演化的[34]。

应急领域，具身智能通过执行高风险或人类难以承受的任务，成为保障人员安全和优化作业流程的关键技术。在搜索与救援场景，具身智能机器人可以用于搜索失踪人员、运送医疗物资和执行救援任务，减少人员伤亡。在爆炸物处理和排雷场景，具身智能机器人可以携带爆炸物探测器，探测和处理爆炸物，同时也可以在雷区进行排雷作业，避免人员伤亡。乌克兰国防技术发展集群 Brave1 推出 ST1 扫雷无人机，能够在野外条件下执行任务，排雷速度是人类的 4 倍[35]。七腾机器人的防爆四足机器人能够实现在楼梯、台阶、缝隙、狭小空间等复杂路面上进行防爆巡检。在核、危、化、害等恶劣环境，具身智能机器人有望完整替代人类进行复杂危险作业。例如，后端通信中断的情况下，无人机自主导航与作战目标达成、全自主无人化作战系统的大规模应用（基于视觉+惯性）。因为这些场景中因作业对象不确定（多品种、小批量）、作业环境不确定且恶劣，当前相关产品在复杂环境

[34] https://www.ithome.com/0/777/446.htm

[35] https://www.sohu.com/a/737423622_121494821

下的运动能力已有突破，如星动纪元的人形机器人小星，可以爬长城、过雪地，在多种地形上稳定行进。云深处科技的四足机器人绝影 X30 能够在-20℃到 55℃的极端环境下作业。但整体来看，还未实现对复杂人工作业过程的完整替代。

总体来看，具身智能正迅速发展成为与各行各业深度融合的创新驱动力，其相关应用正快速扩展至社会经济的各个层面，将推动着生产力的跃升和生活方式的变革。

四、具身智能发展所面临的挑战

具身智能被誉为是实现通用人工智能的重要路径。具身智能在感知与认知、学习与泛化、计算能力、多任务处理、安全性、隐私保护以及人机关系等多个方面都面临着挑战。

（一）技术挑战

算法层面：具身智能在实现通用智能时面临两大根本性挑战。具身智能的目标是具备通用智能，即能够自主学习如何在各种场景和任务要求下执行任务。然而，现在的具身智能研究大多是将大模型的智能塞进机器人中，这仍是学习人类知识和经验的过程，缺乏自主产生意图的能力，也难以快速适应环境变化。**一是系统需要人类智能的介入。**目前的学习系统本质上仍是一个开环系统，需要人类根据学习结果，有针对性地采集更多更好的数据，调整数据的概率分布，反复迭代优化奖励函数等来实现闭环，Yann Lecun 将目前的机器学习系统描述为“辅助智能（Assisted Intelligence）”，而实现通用具身智能需

要的是“自主智能（Autonomous Intelligence）”[36]。**二是尚未实现感知到行动间的认知映射。**感知和行动需要紧密相连，才能快速应对不断变化的环境。《Thinking，Fast and slow》这本书中提到了人类思维的两种模式，即系统 1（快思考）和系统 2（慢思考）。系统 1 负责实现快速的反应式自主控制，而系统 2 负责实现需要慎重思考、推理分析的有意识的决策。人脑高效运作的原因在于，95%的时间在调度系统 1，只有很少的任务需要调度系统 2。而目前具身智能的智能增益主要在于系统 2，也是由大模型主导实现的思维推理能力。从感知到行动的认知映射涉及物理概念理解、感知预测、行为推理等，也需要构建感知输入与行为输出的关联。目前业界从世界模型、扩散策略、脑神经科学等角度开展了相关研究，但仍未完全解决这一难题。

数据层面：缺乏数据成为具身智能能力突破的重要壁垒。与大模型所依赖的互联网数据不同，EAI 所依赖的数据涉及动态环境中的复杂交互，这使得收集数据成为一项昂贵且具有挑战性的工作。EAI 的数据来源，**一方面**，通过真实数据收集，例如遥操作、观察学习人类等技术路线，面临**一是获取广泛、高质量和多样化的数据挑战。**机器人在不同环境中的适应和泛化能力取决于其处理数据的多样性。例如，家庭服务机器人必须适应各种家庭环境和任务，要求它们从广泛的家庭环境数据中学习，以提高其泛化能力。**二是获取大量真实数据成本过高。**例如，为自动驾驶汽车捕获一小时的多模式机器人数据的成本

[36] LeCun， Yann. "A path towards autonomous machine intelligence version 0.9. 2， 2022-06-27." Open Review 62.1 (2022).

为 180 美元，是模拟相同数据的成本的 100 倍[37]。**另一方面**，则是合成数据，例如通过提供虚拟仿真环境，机器人可以在各种条件下进行模拟操作；或通过算法和数学模型创建的，模拟真实数据中的统计模式和关系。合成数据主要面临**“现实差距”——即模拟环境与现实世界之间的差异挑战**，包括物理、光照和意外交互的差异，在需要高度真实交互的场景中，如精密操作、复杂环境导航等，仿真环境通常无法满足需求。

软件层面：软件生态与硬件结合成为具身智能能力提升的关键挑战。具身智能系统的软件不仅要能够高效地处理和解释由硬件传感器收集的数据，还要能够与硬件平台紧密集成。**一是缺乏统一的操作系统和标准化软件开发工具链，**目前市场上存在多种机器人操作系统，如 ROS 或基于 Linux 自行开发等，由于采用了大量开源组件，常会出现兼容性或版本升级导致系统不可用的情况，增加了开发难度，带来开发时间和成本的增加。**二是算法成熟度不高，**尽管 AI 算法有所进步，但在实际应用中仍面临挑战。例如，在 3D 场景中的情景问答（SQA3D）任务中，当前最先进的模型也只能达到约 47.20%的准确率，远低于人类的 90.06%[38]。**三是软硬件解耦难题，**硬件在移动空间需要做到厘米级别，手眼协调的空间做到毫米级别，具身智能模型才能够实现动作控制算法与产品形态的紧密耦合。

硬件层面：耐用性和能源效率以及与软件的深度集成需求构成了

[37] https://cacm.acm.org/blogcacm/the-value-of-data-in-embodied-artificial-intelligence/

[38] https://arxiv.org/pdf/2210.07474.pdf

具身智能硬件发展的主要障碍。具身智能硬件的发展不仅需要技术上的突破，还需要考虑成本效益、维护升级等多方面因素。**一是耐用性和可靠性挑战。**具身智能硬件载体需要在多变的环境中稳定运行，这对机械部件的耐用性提出了高要求。当前机器人在复杂环境下的故障率仍然较高，维护成本也相对较大。**二是能源效率问题。**电池技术的能量密度和充电速度限制了机器人的持续工作时间。当前，机器人可能仅能连续工作数小时，之后就需要充电。例如，Figure01 续航时间 5 小时，优必选 Walker X 装续航时间 2 小时。**三是硬件需要与软件系统深度集成**，以实现高效的数据处理和精确控制。例如，自动驾驶汽车需要将传感器收集的数据实时传输给控制系统，这要求硬件具备高速数据传输能力和与软件的无缝对接。

（二）应用挑战

产品层面：产品形态的合理性和内部硬件系统结构，会影响具身智能的行动能力边界。具身智能在真实世界中的落地应用，需要构型合理、兼容性高、接口丰富、运动能力良好且可靠性高的机器人产品。**一是通用且强大的具身本体挑战。**具身智能的产品研发需要兼顾芯片算力供给和经济性、通信总线的交换效率、运动功耗等各项指标。例如在需要连续工作的场景，本体的电池续航能力很重要。在实时性和可靠性要求高的场景，对云端通信的效率和本体侧芯片推理能力有更高要求。在执行操作任务的场景，需要本体形态有着更高的灵活度和自适应调节能力。在野外等复杂环境中，可能会遇到滑倒或从高处跌

落的情况，要求本体具备更高的抗击打和抗冲击能力。而实现这些不仅需要对具体场景的需求有深入理解，也面临将本体的执行可靠性、任务效率和成本控制做到平衡这一巨大挑战。**二是内部软硬系统的紧密耦合挑战。**随着具身智能基础模型的多模态和泛化能力提升，具身智能的行动能力也获得改善，但大多仍需结合复杂动作控制算法执行复杂任务。动作控制算法与产品硬件是紧密耦合的关系。产品内部硬件系统结构，会影响具身智能的行动能力边界。例如，波士顿动力 Spot 四足机器人搭载先进动作控制算法，使其能够在复杂地形中行走，但它的硬件设计限制了它在需要精细操作或与人交互时的能力，使得 Spot 机器人擅长在户外巡检，但使用工具灵活不足。

商业场景层面：市场需求的明确性和用户接受度会影响具身智能的商业应用进程。具身智能虽然潜力巨大，但具体应用场景和商业模式不够清晰，面临：**一是场景差异化和开放度挑战，**服务、生产、消费等各种场景都可能成为具身智能的潜在应用领域。然而当前的大规模商用还需要选择容错度较高的环境，且用户买单能力比较强的场景，市场需求的甄别和预测成为商业落地的首要难题。**二是用户接受度和信任建立的挑战，**用户对具身智能技术的接受程度和信任感需要建立和维护，这对于技术的成功商业化至关重要。例如，在医疗领域，尽管机器人手术系统如达芬奇手术系统能提供高精度手术操作，但患者和医生对机器人手术的接受度和信任仍在逐步建立过程中，这限制了其广泛应用。**三是安全与隐私问题，**在数据隐私方面，通过机器人的摄像头、麦克风等传感器设备，收集用户的个人信息和行为数据，如

语音指令、生物特征数据等，带来数据安全隐私问题；在物理安全方面，机器人具有较高的动力和运动能力，因此可能对周围人员和环境造成伤害。系统安全方面，入侵者可能通过篡改指令、控制机器人、窃听敏感信息等方式对机器人进行远程操控，从而对用户造成威胁。

来源：中国信息通信研究院

图 5 具身智能产业链示意图

产业链层面：产业链条的完整性和各环节之间的协同效率，影响具身智能产业的持续发展。如图 5 所示，**上游：**硬件迭代周期与成本跟不上软件或算法模型的迭代速度。在具身智能本体技术的关键领域和价值链条中，核心技术壁垒主要围绕三大核心组件展开：减速器、伺服系统以及控制器，在机器人整体成本结构中占比六到七成。三大核心组件行业面临精度、稳定性、计算能力等挑战，影响上层软件的运动控制指令以及对更多精准大规模数据的收集能力。**中游：**挑战在于如何开发出高效、可靠的软件系统，以及如何实现软硬件的深度集成。比如，开发能够适应复杂环境和任务的控制算法是一个技术难点，同时需要大数据、大模型和大算力的加持，且三个‘大’互相关联，缺一不可，还需要不断更新，适应新的任务与环境。**下游：**跨界融合成

为应用新挑战。随着具身智能在家庭服务、教育培训、休闲娱乐、医疗保健、生物制造、物流运输、制造业、低空经济、航空航天等行业的广泛应用，个性化定制将成为机器人生成的新模式，跨界融合突破单一领域的应用将成为新的趋势。需要垂直场景探索与通用泛化兼顾。

（三）标准与合规挑战

具身智能产业在发展和培育的过程中，面临促发展与安全监管并重挑战。在标准化层面，具身智能技术、评测、安全伦理等标准缺失。因涉及跨人工智能、机械自动化等交叉学科技术，安全和伦理问题突出，标准化工作面临系列挑战和难度。在技术评测标准方面，虽然已有国外 softGym、Habitat 3.0、BEHAVIOR-1K 以及国内 AIIA EAI Bench 等工作，但具身智能基准测试标准体系仍建设面临数据规模有限和质量不高、需要构建任务活动知识库，模拟真实任务活动情况等问题。在安全标准方面，因机器人能与现实世界直接进行互动，盗窃或误用可能会产生直接的物理后果，具身智能技术的安全问题包括传统网络安全中不存在的漏洞，安全标准也必须不断发展。**法律与伦理规范层面，**具身智能机器人的出现，不仅要考虑生命安全风险，还面临信息安全、个人隐私等一系列伦理和社会学问题。当机器人与人类伦理发生冲突时，如何规范、合理地开发 AI 技术、使用 AI 产品，以及如何应对人机交互过程中可能出现的社会问题，成为当今时代下必须重视的问题。需要有相关的监管标准和规范，明确机器人在各个应用场景中的边界和限制。同时，人工智能与机器人技术的进步将带来劳动力变化，扩大技能差距和人才短缺。2023 年 3 月高盛发布报告

称，人工智能可能取代相当于 3 亿个全职工作岗位[39]，新技术驱动的工作所需技能与当前劳动力所拥有技能之间的不匹配，需要监管和政策更好地应对行业构成和就业模式的转变。

五、迈向未来，具身智能迎来无限可能

具身智能使信息域和物理世界深度融通，进一步拓展人工智能发展边界，使机器人等物理实体更好地理解世界、更自然地与人类交互和更高效地执行任务。思维智能和行动智能的有机融合将推动人类社会进一步迈向智能化新时代，加速通用人工智能（AGI）的到来。

（一）技术创新发展，推动具身智能持续进化

具身智能将进一步加深对智能本质的深刻理解。通过感知、决策、行动、反馈的循环，具身智能可以实现持续地智能进化。**未来，**智能不再是先验设计的结果，而是在开放环境中涌现的产物；不再局限于中央处理器，而是分布在感知、思维、行动的动态网络之中。**一是数据驱动下的“感知—决策—行动—反馈”闭环，**具身智能需要具备跨模态（如视觉、听觉、触觉）感知和认知能力，以能够更好地理解复杂场景，并在其中做出更加精确和灵活的响应，获得更全面和深入的环境理解。未来，能够理解、预测、做出决策并适应变化的世界模型是实现通用具身智能的关键。**二是形态涌现，**将通过强化学习、进化算法等技术，实现具身智能形态和行为的自适应和优化，提升自主决策能力和行为执行的精确性。未来。探索如何减少人类干预，使控制系统更加自主成为重要发力点。**三是多体协同，**如何构建多个智能体之

[39] https://www.bbc.com/news/technology-65102150

间的协作框架，实现集体优化是多体协同关注的重点。未来，这一方向可能发展出更加高级的群体智能算法和多智能体系统，使得具身智能体能够协同完成复杂任务。

（二）产业跨界整合，开辟更广阔的市场空间

大模型的快速突破让具身智能在各个行业的应用优势不断释放。**未来**，具身智能将突破数据瓶颈和产品形态限制，以经济、灵活且高效的方式实现规模化应用。**工业制造领域**，具身智能有望成为新型工业化的关键核心和有效抓手，使得机器人从“能动”到“能干活”转变，以此来为工业制造业的自动化和智能化升级提供强大支持。未来，以机器人和机械臂等为载体的具身智能应用，将使得工业制造过程更加自动化、灵活和高效。**自动驾驶领域**，自动驾驶的感知、决策与行动能力与具身智能天然契合。将从简单的导航到全面的环境交互和决策的转变，为未来智能交通和智慧城市的建设提供坚实的基础。**航空航天领域**，具身智能将凭借强大的环境感知和自主决策能力实现更好的飞行规划和太空探索。通过探索各类飞行仿真场景，有望将任务规划周期从几天缩短到几分钟。**医疗康养领域**，具身智能提供辅助和服务。具身智能有望极大地改善医疗服务的质量和效率。它将推动医疗服务从传统的被动治疗向主动预防、个性化护理和智能化康复转变。**交通物流领域**，具身智能降低流通成本助力形成高效、快捷、现代化、智能化的物流体系。应用在包括提升仓储、装卸、搬运、分拣、包装、配送等环节的工作效率和管理水平。**家庭消费领域，**具身智能将为高端家庭服务，家庭服务机器人可以实现对人类日常工作环境的认知理

解，在与环境交互的过程中学习、纠正、收集数据，完成自主居家、办公帮手任务。未来的机器人应用将更加多样化、个性化、智能化，跨界融合成为机器人应用的新趋势。

（三）体系重构加速，引发更深层次社会思考

具身智能代表着人工智能发展的一个新的里程碑，预示着我们即将进入一个“知行合一”的新时代。在这个时代，智能将不再局限于冰冷的算法和数据，而是与现实世界紧密交织、共生共进。**未来**，其发展和应用将对社会的各个层面产生复杂而深远的影响。**劳动就业层面，**随着具身智能在制造业、服务业、医疗健康等各个行业的深入应用，许多传统的工作岗位可能会被自动化技术取代，这要求社会对就业结构进行调整，并为劳动力提供再培训和转岗的机会。**人机关系层面**，具身智能的发展意味着机器将更加深入地融入人类生活的各个方面，从日常辅助到高级决策支持，将使人类与机器的关系更加紧密。这可能导致人类对机器的过度依赖，甚至失去一些基本的技能和能力。我们需要思考如何在享受技术带来便利的同时，保持人类的自主性和独立性。**社会关系层面，**一些人可能会对新技术持怀疑态度，担心其带来的不确定性和风险。社会需提高公众对具身智能技术的认识和理解，同时，企业也需要在推广新技术时，充分考虑用户的接受度和心理反应。**伦理和法律层面，**具身智能的自主性和决策能力将引发对伦理和法律问题的讨论。**一是伦理决策和责任归属，**具身智能的伦理问题和责任归属需要明确，包括是否应赋予机器人某些权利，如何防止其被滥用，如何设定具身智能系统的行动准则以确保其决策符合人类伦理

标准，以及如何在具身智能发生故障或者导致事故发生时判定责任归属等，这涉及制定相应的伦理标准和法律法规。**二是数据隐私和安全，**具身智能系统会加大数据滥用和泄露风险，如何确保数据的安全和合法使用，具身智能的全球性影响要求国际社会共同参与，制定全球性的政策和标准，以实现平衡创新与监管的目标。

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：13552276063

传真：010-62304980

网址：www.caict.ac.cn