

人形机器人步入软件定义和 AI 驱动时代

证券研究报告

2024 年 04 月 17 日

人工智能系列报告（三）

• 核心结论

基于对人形机器人发展阶段及核心竞争力的分析，我们认为人形机器人已经跨入了一个全新的软件定义加 AI 驱动的时代。软件不仅承载着人形机器人的基础操作系统和应用算法，而且通过不断迭代升级，为人形机器人赋予了超越传统机械设备的智能特性。软件和 AI 不仅开启了人形机器人在各行各业广泛应用的可能性，更定义了人形机器人的功能边界。

- 基于人形机器人所展现出的高度机械化特性以及在视觉感知方面对自动驾驶算法的成功复用，我们认为人形机器人处于由高动态向高度智能化发展的阶段。高动态是指机器人在运动能力上表现优秀，特别是平衡性、越障碍能力等。当前人形机器人机械化程度较高且具备较强的运动控制及环境感知能力。运动控制方面，波士顿动力的 Atlas 人形机器人身体协调性非常出色，除了行走，还完成翻滚等高难度动作；感知方面，特斯拉 Optimus 基于完全端到端训练的神经网络架构，底层感知、识别算法等等已经和 FSD 打通复用。
- 我们认为人形机器人的核心竞争力在于其通用性和泛化性，即智能化程度，也就是在面对各种未知环境、任务和交互情境时的高度适应性和自我学习能力。人形机器人的设计初衷是通过模仿人类的形态和行为能力，让其能无缝使用人类所有的基础设施和工具，以便在多种环境中执行任务，提高生产效率，尤其是在替代人类执行一些可能危险、困难或不适宜的任务。人形机器人的商业化进程中深受应用场景局限性的影响，提高智能程度能够显著拓宽人形机器人应用场景，打破现有局限，提升其对复杂环境的适应性和任务执行的灵活性。

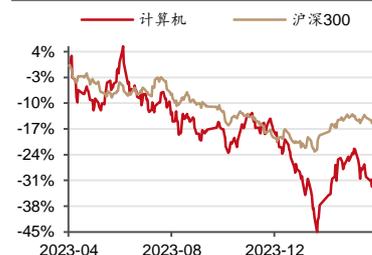
大模型成为“AI 大脑”，通用机器人曙光已现。1) “决策”+“认知”是智能机器人智能化水平的高层次判断标准。根据当前技术形态研判，国内外的人形机器人目前已普遍处在 L3 阶段，并随着大模型时代的到来逐渐向 L4 和 L5 阶段进行发展，而实现人形机器人智能化等级的跃升，提升其决策和认知能力是必经之路更是核心难点之一。2) 大模型+机器人是 AI 重要落地场景，大模型的泛化能力为通用机器人的发展带来曙光。大模型凭借其庞大的知识库和强大的理解能力所带来的泛化能力，赋予了人形机器人更高的通用性，使其能够满足不同场景下的多样化任务需求；同时大模型的辅助编程等功能还能有效降低软件开发成本，有望加速人形机器人商业化落地。

建议关注：1) 视觉感知：海康威视、大华股份、虹软科技、智微智能；
2) 机器人+大模型：科大讯飞、商汤-W、云从科技-UW。

风险提示：人形机器人技术发展不及预期、政策法规环境不确定风险、供应链与制造成本上升风险、国际竞争加剧与技术替代风险。

行业评级	超配
前次评级	超配
评级变动	维持

近一年行业走势



相对表现	1 个月	3 个月	12 个月
计算机	-16.01	-9.47	-37.28
沪深 300	-1.65	8.73	-15.38

分析师

郑宏达 S0800524020001
13918906471
zhenghongda@research.xbmail.com.cn

相关研究

计算机：大模型企业级 B 端收入有望高速增长——计算机行业周观点第 7 期 2024-04-14
计算机：广州数科集团，国企市值管理标杆，打造新质生产力——2024-04-09
计算机：继续看好端侧智能——计算机行业 2024 年 4 月研究观点 2024-04-07

索引

内容目录

一、为什么我们认为人形机器人步入软件定义时代？	4
1.1 人形机器人现在处于什么发展阶段？	5
1.2 人形机器人的核心竞争力在于什么？	6
1.3 为什么我们认为人形机器人步入软件定义时代？	7
二、大模型成为“AI 大脑”，通用人形机器人曙光已现	8
2.1 人形机器人走向通用，“决策”+“认知”是必经之路	8
2.2 具身智能技术持续突破，通用人形机器人新纪元将至	9
2.2.1 ChatGPT for Robotics：大模型赋能机器人的初步探索	10
2.2.2 PaLM-E：具身多模态视觉语言模型，是具身智能领域的一次重大飞跃	11
2.2.3 RT-2：全球首个 VLA 多模态大模型，能够理解视觉输入，机器人模型里程碑	12
2.2.4 VoxPoser：实现零样本机器人任务轨迹规划	13
2.2.5 RoboAgent：实现通用机器人小数据集的快速训练	14
2.2.6 RT-X：基于多个数据集的机器人通用大模型	14
2.2.7 Eureka：基于 GPT-4 的人工智能代理	15
2.2.8 RobotGPT：ChatGPT 运用于机器人的框架	16
2.2.9 EVE：依赖端到端神经网络实现自主工作	16
2.2.10 Figure 01：能听、能说、能自主决策的人形机器人	17
2.2.11 GR00T——通用人形机器人基础模型	18
三、建议关注	19
四、风险提示	19

图表目录

图 1：汽车产业发展历程	4
图 2：人形机器人发展大事记	5
图 3：波士顿动力 Atlas 俯身下蹲抓起工具包	5
图 4：波士顿动力 Atlas 蹦上高台	5
图 5：波士顿动力 Atlas 推完箱子后保持身子平衡，不前倾摔倒	6
图 6：波士顿动力 Atlas 前空翻落地	6
图 7：Tesla Optimus Gen2 用手指抓取鸡蛋（屏幕右侧实时显示了手指压力）	6
图 8：Tesla Optimus Gen2 叠衣服（目前尚不能自主执行叠衣服操作）	6
图 9：优必选人形机器人发展历程及展望	7
图 10：泛化能力包括跨语言泛化、跨任务泛化、跨领域泛化、结构泛化、组合泛化、鲁棒性等	9
图 11：2023 年至今“大模型+机器人”发展进程	10
图 12：传统机器人任务需要工程师在循环中不断改进	11
图 13：接入 ChatGPT 后工程师无需在循环中调整	11
图 14：ChatGPT for Robotics 的设计原则	11
图 15：PaLM-E 主要架构	12
图 16：PaLM-E 在复杂的移动操作任务表现出色	12
图 17：RT-2 架构	13
图 18：Voxposer 系统流程	14
图 19：RoboAgent 采用的多任务动作分块 MT-ACT 架构	14

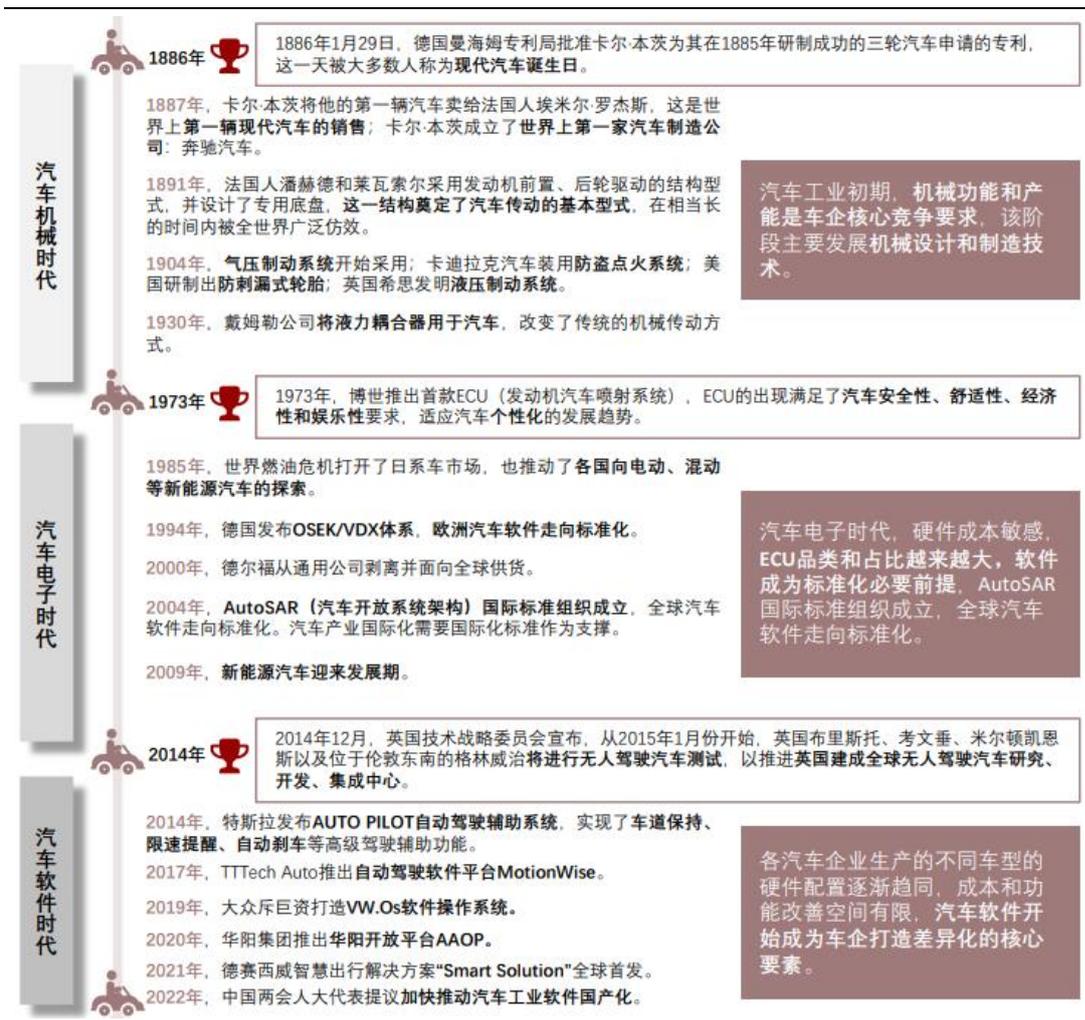
图 20: RT-X 架构	15
图 21: Eureka 训练机械手臂完成高难度的转笔动作	15
图 22: Robot-GPT 架构	16
图 23: EVE 可自主进行物品搬运收纳	17
图 24: EVE 可自主进行充电	17
图 25: Figure01 技术原理	17
图 28: 优必选 Walker S 根据语音指令完成叠衣服任务	18
图 26: GR00T 能够理解语言、视频和演示	18
图 27: GR00T 基于 NVIDIA 深度技术堆栈开发	18
表 1: 机器人智能化信息模型要素	8
表 2: 通用智能化等级判断依据	9

一、为什么我们认为人形机器人步入软件定义时代？

由于技术与产业关联性、以及未来成本下降和规模化生产路径相似性，人形机器人经常与汽车类比。人形机器人和汽车都是高度工程化的产物，需要集成多种技术和部件。特斯拉等公司在开发人形机器人时，强调了与汽车业务的共享技术路径，比如使用类似的电池技术、电机技术、AI 算法以及自动驾驶技术。汽车集合了动力系统、传动系统、电子控制系统等多种复杂组件，人形机器人同样融合了机械设计、运动控制、感知系统（目前多数汽车也有传感器和雷达）等；而汽车的自动驾驶技术则可以对应于目前人形机器人自主决策与动作规划。

复盘汽车产业“机械时代→电子时代→软件时代”发展历程，本质是汽车的核心竞争力在发生变化。在“机械时代”，汽车的核心竞争力主要体现在机械结构和制造工艺上，这一时期的汽车更多依赖于物理结构和机械性能来满足市场需求。而后汽车产业进入“电子时代”，电子控制单元（ECU）的应用开始提升车辆的功能性，此时的核心竞争力转向了电子技术和系统集成能力，这些技术的应用不仅提高了汽车的安全性和舒适性，也使得汽车能够实现更复杂的控制功能。随后，硬件功能及成本改善升级空间逐步变小，汽车产业进入“软件时代”，汽车的核心竞争力发生了根本性的变化，软件定义汽车（SDV）的概念应运而生，汽车从高度机电一体化的机械终端，逐步转变为一个智能化、可拓展、可持续迭代升级的移动电子终端。

图 1：汽车产业发展历程

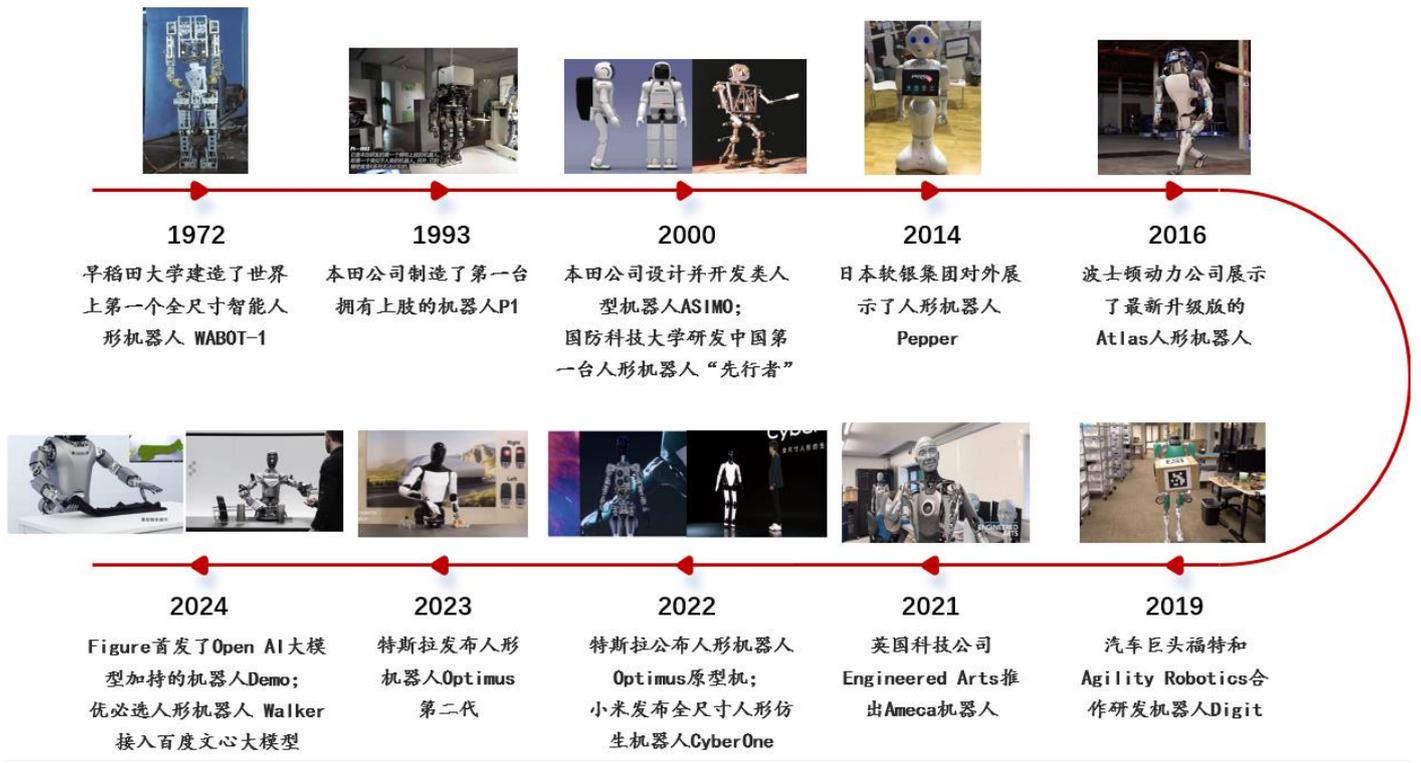


资料来源：头豹研究院、西部证券研发中心

1.1 人形机器人现在处于什么发展阶段？

基于人形机器人所展现出的高度机械化特性以及在视觉感知方面对自动驾驶算法的成功复用，我们认为人形机器人处于由高动态向高度智能化发展的阶段。高动态是指机器人在运动能力上表现优秀，特别是平衡性、越障碍能力等。当前人形机器人机械化程度较高，体现强大的运动性能、仿生结构设计、先进材料应用和精细操作能力等层面；且具备较强的运动控制及环境感知能力，体现在多传感器融合感知、无线通信、模块化设计以及能源管理等多个方面。展望未来，人形机器人将在现有基础上，进一步强化自主决策与深度学习能力，在复杂场景下能够独立思考、适应环境并持续优化自身行为。

图 2：人形机器人发展大事记



资料来源：中国机器人网、新智元、智东西、机器之心、优必选科技视频号、西部证券研发中心

运动能力方面，以波士顿动力的人形机器人 Atlas 为例。

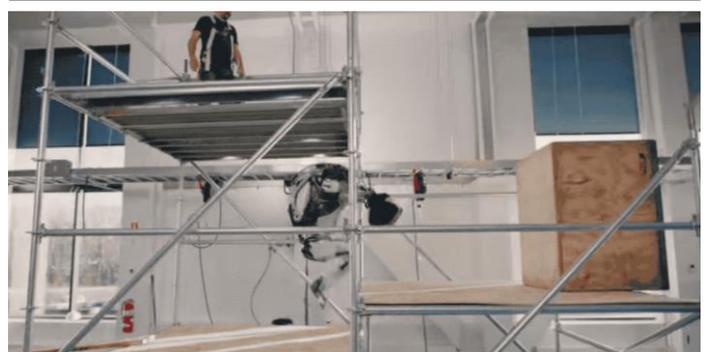
Atlas 有多达 28 个自由度，涵盖双足、躯干、手臂等，确保了全方位、精细化的动作模拟。这种高自由度赋予 Atlas 出色的身体协调性与灵活性，使其能精准模仿人类复杂的行走步态、跑动、跳跃，甚至完成翻滚、360 度旋转等高难度动作。

图 3：波士顿动力 Atlas 俯身下蹲抓起工具包



资料来源：量子位微信公众号、西部证券研发中心

图 4：波士顿动力 Atlas 蹦上高台



资料来源：量子位微信公众号、西部证券研发中心

图 5：波士顿动力 Atlas 推完箱子后保持身子平衡，不前倾摔倒



资料来源：量子位微信公众号、西部证券研发中心

图 6：波士顿动力 Atlas 前空翻落地



资料来源：量子位微信公众号、西部证券研发中心

感知方面，以 Optimus——特斯拉基于端到端训练的神经网络架构的人形机器人为例。

2023年12月，特斯拉发布其最新一代人形机器人 Optimus Gen2 的最新进展，步行速度提升 30%，整体重量减轻 10kg，且具有触觉感知灵巧手。与之前的版本相比，Optimus Gen2 的稳态行走能力有了较大提升，步行速度加快 30%；手有 11 个自由度，手指搭载触觉传感器能够精准抓握鸡蛋等易碎物品；且能够完成多个深蹲动作并保持身体平衡等。

Optimus 基于完全端到端训练的神经网络架构，底层感知、识别算法等等已经和 FSD 打通复用。马斯克表示，特斯拉目前已经打通了 FSD 与 Optimus 的底层模块，实现了一定程度的算法复用。FSD（Full Self-Driving，全自动驾驶）算法主要依赖于神经网络和计算机视觉技术，其核心是神经网络模型：通过对实时传感器获取的数据进行处理和分析并从中提取有关道路、车辆、行人和障碍物等信息，可以实现车辆的环境感知和物体识别，而 FSD 算法在人形机器人的感知、决策和控制等方面也同样起到重要作用。Optimus 的神经网络是完全的端到端训练架构，即可以做到视频信号输入和控制信号输出，能够直接从原始输入端到输出端进行训练而无需进行手动特征工程或中间阶段处理，有效缩短研发周期、实现快速的产品迭代。

图 7：Tesla Optimus Gen2 用手指抓取鸡蛋（屏幕右侧实时显示了手指压力）



资料来源：机器之心微信公众号、西部证券研发中心

图 8：Tesla Optimus Gen2 叠衣服（目前尚不能自主执行叠衣服操作）



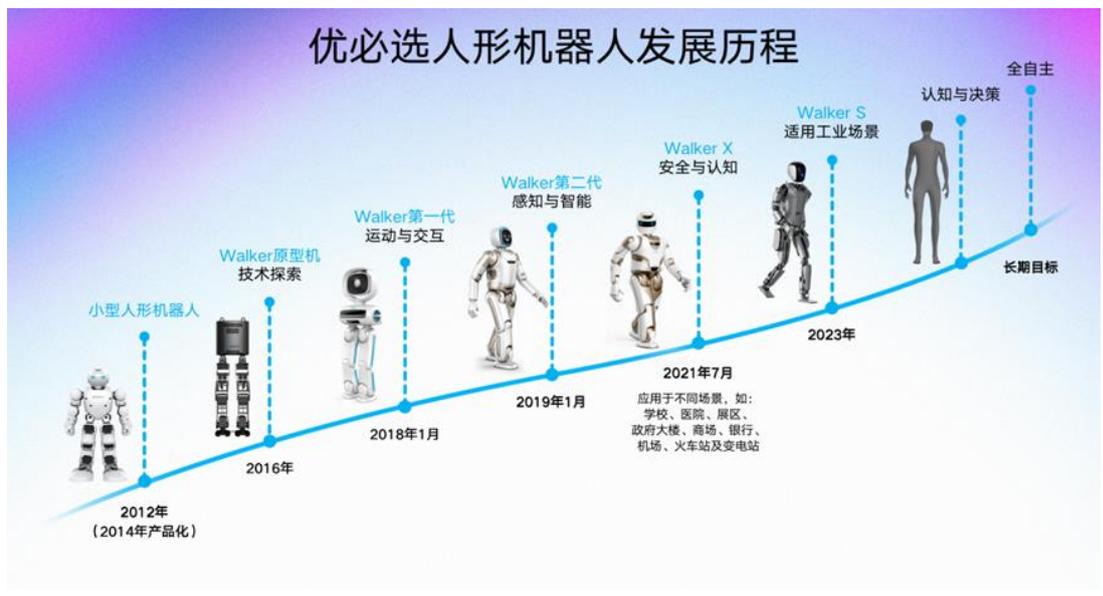
资料来源：机器之心微信公众号、西部证券研发中心

1.2 人形机器人的核心竞争力在于什么？

“拟人化”不只是身体和行为拟人，更重要的是大脑和思维拟人，人形机器人不是单纯地执行预先编写程序的机械设备，而是拥有自主决策和学习能力的智能终端。人形机器人的研究和发展不仅仅只是为了模仿人类的外观形态，更重要的是要赋予它们类似于人类的认知能力和情感表达能力，使其能够理解和响应人类的需求，在不同的环境和场景

中自主地完成任务。

图9：优必选人形机器人发展历程及展望



资料来源：雷峰网、优必选官网、西部证券研发中心

人形机器人的商业化进程中深受应用场景局限性的影响，通用性和智能化或将成为破局关键。回顾历史，可以发现人形机器人商业化进程非常艰难折：波士顿动力公司多次易主；曾经引领时代的本田 ASIMO 机器人退役；软银缩减机器人业务,停止 Pepper 生产。尽管波士顿动力、本田 ASIMO 和软银 Pepper 等人形机器人在技术演示中展现出卓越的运动能力、互动性能或情感交流特性，但在实际应用中，它们往往难以找到足够广泛且具有经济效益的落地场景。

所以，我们认为人形机器人的核心竞争力在于其通用性和泛化性，即智能化程度，也就是在面对各种未知环境、任务和交互情境时的高度适应性和自我学习能力。汽车是一种代步工具，而人形机器人更多地聚焦于替代或辅助人类执行任务。人形机器人的设计初衷是通过模仿人类的形态和行为能力，让其能无缝使用人类所有的基础设施和工具，以便在多种环境中执行任务，提高生产效率，尤其是在替代人类执行一些可能危险、困难或不适宜的任务。

提高智能程度能够显著拓宽人形机器人应用场景，打破现有局限，提升其对复杂环境的适应性和任务执行的灵活性。高度智能的人形机器人将具备更强的学习能力，能在与环境的互动中不断优化行为策略，适应多样化的操作任务和未知情境。同时，自然语言处理和情境理解的进步将使人形机器人能够更好地与人类进行沟通，理解并响应复杂指令，甚至预测用户需求。此外，智能化还体现在自主决策与故障诊断能力的提升，使人形机器人能够在无人干预下独立完成任务，减少对远程操控或现场人工指导的依赖，大大增强了其在远程作业、危险环境作业等领域的应用潜力。

1.3 为什么我们认为人形机器人步入软件定义时代？

人形机器人正逐步迈进由高动态运动性能向高度智能化跃升的新阶段，这一转变离不开 AI 的发展与深度融入。随着 AI 技术在感知、认知、决策与执行层面的全面赋能，人形机器人拥有了更强大的环境感知能力，可以通过视觉、听觉、触觉等多种传感器收集信息，利用深度学习、强化学习等先进算法解析复杂数据，进行实时决策，并通过高精度的动

力系统实现细腻、流畅且灵活的动作执行。

更重要的是，人形机器人的核心竞争力在于其依托 AI 实现的通用性和泛化性。借助强大的 AI 引擎，人形机器人将能在多种任务场景中实现快速学习与适应，无需硬件改造即可通过软件更新升级功能，这使得它们未来能够广泛应用于制造、教育、医疗、娱乐、养老、救援等多元领域，展现较强的普适性和延展性。

因此，我们认为人形机器人已经跨入了一个全新的软件定义、AI 驱动的时代。软件不仅承载着人形机器人的基础操作系统和应用算法，而且通过不断迭代升级，为人形机器人赋予了超越传统机械设备的智能特性。软件和 AI 不仅开启了人形机器人在各行各业广泛应用的可能性，更定义了人形机器人的功能边界。

二、大模型成为“AI大脑”，通用机器人曙光已现

2.1 人形机器人走向通用，“决策”+“认知”是必经之路

“决策”+“认知”是智能机器人智能化水平的高层次判断标准。我国 2022 年发布的《智能机器人智能化等级评价规范》从要素智能化等级、综合通用智能化等级和综合场景智能化等级三个方面对机器人智能化等级进行了权威的分类，并根据智能化程度不同从低到高划分为 1 级~5 级。其中，综合通用智能化评级主要从感知、执行、决策和认知四个方面对机器人的智能化综合能力进行评价，从低到高可分为 L1（基础型）、L2（半交互型）、L3（交互型）、L4（自主型）、L5（自适应型）。根据该标准，决策和认知能力是判断机器人智能化水平高低的最高层次标准，其中认知能力主要衡量其建模、理解和推理等方面的能力，执行能力衡量其运动执行和交互等方面的能力。根据当前技术形态研判，国内外的人形机器人目前已普遍处在 L3 阶段，并随着大模型时代的到来逐渐向 L4 和 L5 阶段进行发展，而实现人形机器人智能化等级的跃升，提升其决策和认知能力是必经之路更是核心难点之一。

表 1：机器人智能化信息模型要素

要素	子要素	具体项目
感知	视觉	人脸识别、字符识别、图像识别、定位测距
	听觉	声源定位、语言识别
	触觉	压觉、力觉、冷热觉
	嗅觉	对于气味的识别
	其他	环境感知、味觉
认知	建模	激光建图、视觉建图、语义建图、知识库构建
	理解	语义理解、多轮语义规划、多轮对话理解、情感理解、多模态意图理解
	推理	偏好推荐、知识推理
决策	规划	任务规划、动作规划、路径规划、不确定环境下的运动规划
	优化	自适应标定技术、可强化的物体操纵技术、机器学习技术、知识库扩充
执行	运动执行	移动能力、操作能力
	交互	语音交互、表情交互、文字交互、肢体语言交互、界面交互、多模态交互、多机交互、操作交互

注：上述测试项目是基于目前技术现状，后续根据技术发展会出现新的测试项目。

资料来源：《智能机器人智能化等级评价规范》，西部证券研发中心

表 2: 通用智能化等级判断依据

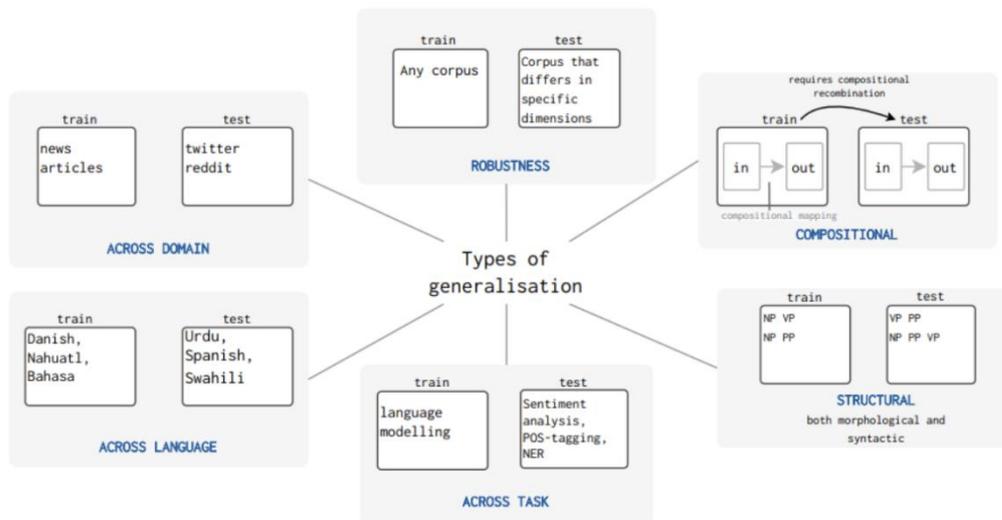
等级	判断依据			
	感知	执行	决策	认知
L1 (基础型)	有	部分有	/	/
L2 (半交互型)	有	有	/	/
L3 (交互型)	有	有	部分有	/
L4 (自主型)	有	有	有	/
L5 (自适应型)	有	有	有	有

资料来源:《智能机器人智能化等级评价规范》, 西部证券研发中心

正如我们前文所说, 人形机器人以解放人体、自主完成任务为目标, 其长期的核心价值在于通用性、泛化性。目前工业机器人的技术虽已相对成熟, 但其仅适用于汽车制造、纺织、包装等行业中单一重复性的生产工作。而人形机器人作为智能机器人的具象化, 其在“拟人”之下更为核心的价值在于通用性, 更高的通用性则要求其要具备更高的感知、执行、决策和认知能力, 以减少对人工指令的依赖, 提升对更复杂的非结构化环境的理解和适应能力, 从而能够在更广泛的应用场景中发挥作用。

大模型+机器人是 AI 重要落地场景, 大模型的泛化能力为通用机器人发展带来曙光。模型泛化能力 (Generalisation) 是机器学习和人工智能领域评判模型性能的重要指标之一, 其可以理解为一种迁移学习的能力, 即把从过去的经验中学习到的表示、知识和策略应用到新领域的的能力。以往的算法模型泛化能力较低, 即便经过大量训练也难以覆盖所有小概率边缘场景, 因此通常仅被用在特定的应用场景, 难以进行应用场景的拓展。而大模型凭借其庞大的知识库和强大的理解能力所带来的泛化能力, 赋予了人形机器人更高的通用性, 使其能够满足不同场景下的多样化任务需求; 同时大模型的辅助编程等功能还能有效降低软件开发成本, 有望加速人形机器人商业化落地。

图 10: 泛化能力包括跨语言泛化、跨任务泛化、跨领域泛化、结构泛化、组合泛化、鲁棒性等



资料来源:《State-of-the-art generalisation research in NLP: A taxonomy and review》, 西部证券研发中心

2.2 具身智能技术持续突破, 通用机器人新纪元将至

大模型掀起了具身智能新兴领域的研究热潮, 具身智能或将成为 AI 领域的下一个“北极星”问题。自大模型流行至今, 谷歌、微软、英伟达等国外科技巨头等纷纷投入到基于大模型的具身智能的研究赛道中, 相关理论和技术得到持续突破。我们梳理了 2023 年至今国内外“大模型+机器人”的最新研究成果, 从初步技术框架再到具体应用落地, 基于

大模型的机器人技术路线愈发明晰，通用人形机器人新纪元即将到来。

图 11：2023 年至今“大模型+机器人”发展进程



资料来源：各公司官网、西部证券研发中心整理

2.2.1 ChatGPT for Robotics：大模型赋能机器人的初步探索

2023 年 2 月，微软发布了一篇名为《ChatGPT for Robotics: Design Principles and Model Abilities》的技术报告，阐述了 ChatGPT 赋能机器人的方法。

ChatGPT 在机器人领域的应用主要体现在自然语言交互和自动化决策。传统上，机器人任务需要工程师在循环中编写复杂的代码并不断进行迭代改进，但 ChatGPT 使得机器人能够基于自然语言指令采取响应行动或是完成自动化决策。接入 ChatGPT 后，工程师只需构建底层库函数及其描述，机器人就能基于底层函数来编写代码完成相关任务，从而减轻工程师的工作负担。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/567122013164006065>