

具身智能与智能机器人 技术演进、挑战与未来趋势研判

文 ◆ 中国工程院院士 于海斌



一、人工智能的历史脉络与具身智能的诞生

人工智能（AI）自 1956 年达特茅斯会议正式提出以来，经历了数次技术浪潮与寒冬。于海斌院士指出，AI 的发展可分为四个阶段：符号逻辑推理阶段（1950s–1980s）：以知识表达和

专家系统为核心，但受限于知识获取的困难，最终因实用性不足进入第一次寒冬。神经网络连接阶段（1980s–2000s）：辛顿（Geoffrey Hinton）等人提出反向传播算法，推动神经网络发展，但受限于算力和数据，应用场景有限。深度学习阶段（2010s 至今）：以卷积神经网络（CNN）和斯坦福大学李飞飞团队构建的 ImageNet 数据库为标志，AI 在图像识别等领域实现突破，识别率从 60% 提升至超越人类水平。具身智能（Embodied AI）阶段（2020s 起）：AI 从纯数据驱动转向与物理世界交互，强调智能体通过身体、环境与任务的协同实现认知与行为进化。

图灵早在计算机理论初期即提出两种智能路径——“离身智能”（如 ChatGPT 依赖纯数据推理）与“具身智能”（需与物理设备结合）。然而，具身智能因技术复杂度高长期滞后，直到机器人学、神经科学和心理学的交叉融合为其提供了新思路。

二、具身智能的核心内涵与科学依据

环境交互与智能发育的必然性：以一项科学实验来印证环境交互与智能发育呈现正相关性，该实验反馈了智能的成长依赖与环境的动态互动作用。来自麻省理工学院两位研究人员，把两只小猫放进了一个圆桶内，两只小猫都在圆筒内部绕圈运动。第一只小猫是自己走的；第二只小猫则被放在与柱体中心轴相连的小盒子里。两只小猫看到的東西完全相同。结果显示只有靠自己身体运动的小猫发育出了正常的视力。这进一步印证了“心灵手巧”的逆向逻辑——肢体动作的灵活性（如操作工具）会反向促进认知能力的提升。

具身智能三位一体的智能框架：具身智能强调“大脑-身体-环境”三位一体。其中大脑负责高层决策与意图理解（如大语言模型），身体通过传感器与执行器实现物理交互（如机械臂、仿生关节），环境提供动态反馈与训练场景（如仿真平台、真实物理空间）。三者协同构成闭环，使智能体能够通过试错学习适应复杂任务。

从“感知智能”到“行动智能”的跨越：传统 AI 擅长静态感知（如图像识别），但缺乏对物理世界的动态响应能力。波士顿动力（Boston Dynamics）的 Atlas 机器人通过深度学习与强化学习结合，实现了翻越障碍、适应地形变化等复杂动作，标志着具身智能在运动控制领域的突破。

三、围绕具身智能四大热点的思考与技术挑战

数字人：中国通用人工智能研究院开发的“数字人通通”模拟人类从婴儿期开始的成长过程，通过与虚拟环境的交互（如听觉、触觉反馈）实现认知进化。这一尝试为研究人类智能发育提供了新范式，但也引发伦理争议——若机器人通过环境交互自主进化，人类是否可能失去控制？

机械臂与大语言模型的融合：谷歌 RT-X 平台将大语言模型（LLM）与机器人操作结合，使机械臂能够理解自然语言指令（如“整理房

间”），并自主分解任务步骤（识别杂物、分类归置）。这一技术已在叠衣服、精细装配等场景中展现潜力，但其泛化能力仍受限于物理规则与数据多样性。

人形机器人的争议与前景：

马斯克力推的 Optimus 人形机器人引发两极评价。支持者认为人形是通用化的终极形态，可适配人类环境（如楼梯、工具），而质疑者指出其成本高昂、技术冗余（如多关节驱动难题）。

人形机器人的核心价值在于“本体硬件”与“智能算法”的协同突破。例如，开源算法（如波士顿动力的运动控制模型）大幅降低了本体研发门槛，而黄仁勋（英伟达 CEO）布局的具身智能训练平台，或将推动行业标准化。

智能驾驶与低空经济：尽管全无人驾驶面临地面环境的极端复杂性（如突发行入、不规则道路），但网联汽车的辅助驾驶（如车道保持、自动泊车）和低空经济（无人机物流、巡检）已成为更可行的商业化方向。

四、技术路径：肢体、小脑与大脑的协同进化

在机器人技术路径未来发展趋势探讨方面，于海斌院士表

示未来人形机器人肢体部分有很大的提升空间，如基于高能量密度的仿生驱动肢体。仿照人类心脏的高效供能机制，研发微型液压驱动与仿生肌肉，如德国费斯托的气动机械臂就是一个很好的参考方向，该机械臂核心优势在于其高精度控制与仿生设计的融合，拥有 12 个自由度，动作柔性且稳定，部分型号搭载 AI 技术实现自我学习优化，电磁阀寿命超 3000 万次，具备良好的性能和耐用性。

此外，生物融合也是未来机器人发展趋势。未来，我们可以通过培养生物细胞构建具有感知与驱动能力的“类器官”产品，尽管其生存环境要求苛刻，但为软体机器人提供了新思路。

小脑主要作用在于提升人形机器人的运动协调能力，运动控制依赖海量物理数据，而真实环境训练成本过高。目前国内已经有相对成熟的平台进行机器人的模拟训练，如国家地方共建人形机器人创新中心开发的，异构人形机器人训练场通过“虚拟-现实”迁移学习（Sim2Real），使机器人能在低成本场景中预训练动作（如行走、跳跃），再迁移至真实环境微调。

大脑提供决策模型与机器

人的认知能力，目前多模态感知技术是主要方案，该方案融合视觉、触觉、力觉传感器，提升环境理解的全面性。在具体案例方面，北京大学电子学院程翔教授团队提出的“机器联觉”系统，通过智能融合通信与多模态感知信息，有效提升了机器人在复杂环境中的感知、决策与通信能力。

如何解决大模型“灾难性遗忘”问题，实现机器人的终身学习方式。爱丁堡大学提出的增量强化学习框架是一个很好的解决“灾难性遗忘”的解决方案。它能够在保持已有知识的基础上，通过逐步增量式地更新策略，快速适应新环境或新任务，从而显著提升机器人在动态变化环境中的学习效率和性能表现，无需从头开始重新训练整个模型。

五、未来十年技术趋势研判与产业启示

未来十年，具身智能的算法与训练平台将快速发展，而硬件（如高精度传感器、仿生关节）受限于材料与工艺，进步速度相对缓慢。企业需优先布局软件生态（如仿真平台、数据链），而非盲目投入硬件研发。

此外具身智能缺乏统一理论体系（如认知科学的世界模型），但场景驱动的工程化应用（如仓储机器人、医疗外骨骼）将率先落地。国家需推动跨机构协作，解决数据确权与模型共享难题。

在通用终端的终极形态方面，人形机器人未必是唯一答案，具身智能可能通过“一脑多机”模式赋能多样化设备（如工业机床、家用电器）。例如，同一 AI 核心可同时调度无人机群与地面机器人，实现任务协同。

在伦理与安全方面，若具身智能体通过环境交互自主进化，可能超出人类预设的边界。行业需提前建立伦理规范（如行为约束算法、人机权责界定），避免技术失控风险。

六、结语

具身智能正在让 AI 从“虚拟大脑”向“物理实体”迭代，在技术演化的道路上，需要更加关注技术创新的有效性与产业化的可落地性。正如于海斌院士所言：“蓝海的意义在于未知，而未知需要勇气与智慧并存。”对产业从业者而言，打破学科壁垒、深化理论突破，聚焦场景刚需、参与生态共建或将成为抢占先机的关键。■

（文章来源：沈阳“创新聚沈·向阳而行”大会演讲实录）