

人工智能大模型推动具身智能无人系统发展

随着人工智能技术的快速发展,具身智能无人系统逐渐成为研究热点。人工智能大模型技术的发展为具身智能无人系统带来了新的机遇。具身智能强调智能体通过与环境的交互来实现智能行为,而无人系统则摆脱了人类直接操控的限制,能够自主完成任务。人工智能大模型的出现,为具身智能无人系统的发展注入了强大动力。大模型具有强大的学习能力和泛化能力,能够处理复杂的多模态数据,为无人系统的感知、决策和控制提供更高效、更智能的解决方案。在当今科技竞争日益激烈的背景下,研究人工智能大模型驱动的具身智能无人系统,对于提升国家的科技实力和推动产业升级具有重要意义。

具身智能无人系统的载体形态丰富,涵盖无人机、无人地面车辆(UGV)、无人水面/水下航行器(USV/UUV)以及各种类型移动机器人。在军事应用领域,察打一体无人机可执行战场侦察、精确打击任务,有效提升作战效能并降低人员伤亡风险;无人运输车辆能在生化污染、炮火覆盖等危险环境中完成物资投送,保障后勤补给链路畅通。在民用领域,无人配送车通过路径规划算法与环境感知模块实现“门到门”货物自动配送,大幅提升物流末端效率;农业植保无人机结合多光谱成像技术与精准喷洒系统,可完成农田长势监测、病虫害防治等作业,助力农业生产向精准化、集约化转型;水下无人航行器则可搭载声学探测、水质分析设备,开展深海资源勘探、海洋生态环境监测等任务。这些应用场景充分验证了无人系统的价值潜力,而具身智能技术的融入,将进一步突破其功能边界,

实现更复杂场景下的自主任务执行。

具身智能,其核心思想是智能体需要通过身体与环境进行交互来获取知识和实现智能行为。早期研究主要集中于结构化环境中的专用控制算法与运动规划方法,实现诸如定点移动、物体抓取等基础性任务。随着深度学习、强化学习等人工智能分支的发展,具身智能的应用范围逐步从工业机器人拓展到智能驾驶车辆、无人机等更为复杂的无人系统,实现了从“结构化环境中的规则性任务”向“半结构化环境中的多目标任务”的重大跨越。近年来,以 Chat GPT、DeepSeek、LLaMA 等为代表的大模型技术爆发式发展,推动具身智能进入一个全新时代。基于深度强化学习与环境交互自学习的智能体,能够通过

与环境的持续互动自主提炼任务策略,并

借助视觉、语言、力触等多模态感知系统,实现对动态环境的深度理解与快速自适应,从而显著提升在未知、非结构化环境中的系统鲁棒性与任务完成率。

通用基础大模型是具身智能无人系统的“底层引擎”。通用基础大模型是 AI 大模型体系的核心基座,其技术特征在于通过大规模无标注数据预训练,构建跨领域的语言理解、逻辑推理与知识存储能力。该类模型以 Transformer 架构为核心,通过千亿至万亿级参数规模实现“知识图谱+语言理解”的双重能力,典型代表包括 ChatGPT 系列、LLaMA 系列、DeepSeek 及文心一言等。在技术范式上,通用基础大模型采用“预训练-微调”模式:预训练阶段通过海量文本数据学习人类语言规律与世界知识,形成通用语义理解与逻辑推理框架;微调阶段则输入少量任务相关标注

人工智能大模型的出现,为具身智能无人系统的发展注入了强大动力。大模型具有强大的学习能力和泛化能力,能够处理复杂的多模态数据,为无人系统的感知、决策和控制提供更高效、更智能的解决方案。



王耀南

机器人技术与智能
控制专家
中国工程院院士
中国图象图形学学会
理事长

现任湖南大学机器人视觉感知与控制技术国家工程研究中心主任。研究方向为智能机器人感知与控制技术及工程应用。

湖南大学,长沙
410082

数据,快速适配文本生成、智能问答、代码编写等特定任务需求。在具身智能无人系统中,通用基础大模型承担“大脑中枢”职能:例如,无人车决策系统可基于大模型的逻辑推理能力,解析交通法规文本、路况描述信息,生成符合人类认知的驾驶决策逻辑;服务机器人可通过大模型理解复杂自然语言指令,完成任务拆解与执行指令转化。

多模态大模型是具身智能无人系统复杂环境跨模态信息交互的“融合桥梁”,其能够统一处理文本、图像、音频、视频以及各类传感器(如激光雷达、毫米波雷达、惯性测量单元(IMU)等)产生的异构数据,打破不同模态之间的语义隔阂。在模型结构上,多模态大模型通常引入模态对齐模块,例如跨注意力机制、特征映射网络等,将不同模态的数据嵌入到统一的语义表征空间中,从而实现真正的跨模态理解与生成能力,例如图像描述生成、语音指令理解、基于多传感器融合的环境状态判断等。这一能力对无人系统在真实世界中的环境感知与语义建模具有至关重要的作用。

具身智能大模型是智能无人系统交互控制的“专用大脑”。具身智能大模型是专为“智能体与环境交互”设计的技术,其核心特征是将“环境感知-动作决策-反馈学习”融入模型架构,实现“感知即决策、决策即控制”的端到端能力。与前2类大模型相比,具身智能大模型在训练数据上增加了“交互数据”(如机器人的运动轨迹数据、无人车的驾驶行为数据、人类与环境的交互数据),在架构上增加了“动作预测模块”和“反馈优化模块”,能够直接输出控制信号(如关节角度、电机转速、行驶速度),而非仅生成文本或图像。具身智能大模型的关键技术突破在于“实时交互与动态适应”:例如,人形机器人在行走时遇到障碍物,模型可在 50 ms

内完成“感知障碍物(视觉+触觉)—调整步态(动作决策)—输出关节控制参数(控制)”的全流程,且通过实时反馈(如脚底压力传感器数据)不断优化动作,避免摔倒,通过具身大模型,可在凹凸不平的地面行走、上下楼梯、搬运 10 kg 重物,动作流畅度接近人类。

近年来,人工智能大模型不断发展,在具身智能无人系统中也发挥着越来越多的作用。首先,在环境感知方面,大模型能够融合并处理来自摄像头、激光雷达、麦克风阵列、IMU等多源异构传感器数据,通过其在大规模数据预训练中获得的强大表征学习能力,提取出更具判别性和泛化能力的特征表示,从而显著提升系统在遮挡、光照变化、天气干扰等复杂条件下的感知精度与鲁棒性。其次,在决策与规划层面,依托其深厚的常识知识库和强大的符号推理能力,大模型能够基于多模态感知信息,结合任务目标与上下文语境,快速生成多种可行的行为方案,并对各方案的成本、风险与预期收益进行综合评估,从而输出最优决策序列,实现高效任务分解与运动规划。第三,在系统适应与迁移学习方面,大模型凭借良好的泛化特性和少样本学习能力,能够帮助无人系统快速适应新的任务环境与操作要求,仅需少量新场景数据即可完成模型微调,大幅提升了部署效率与应用灵活性。

尽管大模型为具身智能无人系统带来了革命性进展,但是进一步发展仍面临诸多技术挑战。数据安全与隐私保护是首要问题,无人系统在运行过程中可能采集包括个人身份信息、精确定位数据、医疗记录在内的敏感信息,必须构建覆盖数据加密、访问控制、匿名化处理和合规审计的全链路安全治理体系。此外,模型的可解释性与可靠性同样是亟待突破的瓶颈,由于大模型固有的“黑盒”特性,其决策逻辑往往

难以追溯和解释,这在自动驾驶、手术机器人等高可信要求场景中可能带来难以估量的风险,因此需大力发展可解释人工智能(XAI)技术,实现决策过程的可视化、可溯源与可验证。另一方面,大模型训练与推理所需的高性能计算资源与能源消耗,也限制了其在微型无人机、便携机器人等终端设备中的部署。为应对这一挑战,“边缘-云”协同计算架构逐渐成为主流解决方案,通过将实时性要求极高的感知与控制任务部署于边缘计算节点,将大规模数据预处理、模型训练与复杂推理任务卸载至云端,从而实现计算资源的高效分配与系统能效的优化平衡。

展望未来,人工智能大模型仍将持续驱动具身智能无人系统迈向更高水平。首先,多模态感知与端到端控制模型的泛化能力将得到进一步加强,从而显著提升系统在高度非结构化、动态开放环境中的适应性与生存能力。其次,融合模型预测控制、强化学习与仿生控制机制的新型智能控制框架,将有效增强系统在复杂动态场景中的行为泛化能力和抗干扰能力。第三,“虚实协同”训练范式将得到更广泛应用,通过高保真物理仿真环境与世界模型技术,智能体可在虚拟空间中完成大规模预训练和策略优化,再通过真实世界数据进行快速微调,从而大幅提升任务执行效率并降低实体训练风险。最后,多智能体协同与分布式智能将成为重点发展方向,通过构建标准化的智能体协作协议,并结合联邦学习、同态加密等隐私计算技术,实现智能体间的高效通信、协同决策与安全协作,从而全面拓展具身智能无人系统在低空经济、智慧城市、智能制造、智慧农业、太空探索、国防安全等重大场景中的应用潜力。