大模型赋能人形机器人智能交互与自主决策的技术架构及应用研究

牟多铎*

（马来西亚理工大学，马来西亚 柔佛州 士古来 81310）

摘 要：随着人工智能技术的快速发展，大模型在自然语言理解、多模态融合、知识推理与生成式交互方面展现出显著优势，为机器人系统从“功能型自动化”迈向“通用型智能体”提供了新的技术基础。人形机器人作为智能装备领域的重要发展方向，因其类人结构和适配人类环境的优势，在智能制造、公共服务、医疗康养与应急救援等场景中具备广阔应用前景。然而，传统人形机器人系统在开放环境中的语义理解、复杂任务规划、跨场景泛化与安全可控等方面仍存在不足，难以满足高动态、高不确定性的真实应用需求。基于此，本文围绕大模型赋能人形机器人智能交互与自主决策的关键问题，系统分析大模型在多模态感知融合、语义理解、意图识别、任务分解、策略生成与行为执行中的作用机制，提出一种面向工程落地的总体技术架构，并进一步探讨边云协同部署、实时控制约束、可信安全治理及典型应用场景的实现路径。研究认为，大模型赋能人形机器人智能化升级的核心在于构建“多模态感知—语义理解—知识推理—任务规划—执行反馈”的闭环体系，同时通过约束机制、对齐策略与安全控制提升系统可靠性与可控性。本文研究可为人形机器人产业化应用与智能体系统工程设计提供参考。

关键词：大模型；人形机器人；多模态交互；自主决策；任务规划；边云协同；可信安全

DOI：https://doi.org/10.65196/zxmm8035

1 引言

近年来，人形机器人逐渐成为全球智能制造与人工智能产业竞争的重要方向。相比传统工业机器人，人形机器人具有更接近人类的运动结构与交互形态，能够适配人类生活与生产环境，具备在非结构化场景中执行多任务的潜力。随着高性能伺服驱动、力矩控制、轻量化材料以及传感器技术的发展，人形机器人在双足行走、动态平衡、抓取操作等方面取得显著突破，逐步从实验室样机走向产业化验证阶段[1]。然而，在实际应用中，人形机器人不仅需要具备稳定运动控制能力，更需要在复杂环境中实现自主感知、理解、规划与协作，从而满足真实世界任务的动态性与不确定性需求。

传统机器人系统多依赖预设规则、有限状态机或小规模机器学习模型，其控制逻辑往往 针对特定场景优化，缺乏跨任务迁移能力。当任务环境发生变化或指令表达具有模糊性时，系统容易出现行为失效或决策错误。此外，人形机器人在与人交互过程中面临语言理解、意图推断、情境识别以及安全约束等综合挑战，使得传统方法难以实现高水平的人机协同。

大模型技术的突破为机器人智能化发展带来了新的机遇。大模型具有大规模参数、海量知识储备与强推理能力，可在自然语言交互、多模态语义融合与复杂任务规划方面提供通用能力支撑。通过引入大模型，人形机器人可以将自然语言指令转化为可执行任务序列，并在执行过程中动态调整策略，实现更接近人类的交互与决策能力[2]。与此同时，多模态大模型能够融合视觉、语音、文本及触觉信息，使机器人具备更强的环境理解能力，从而提升其在开放场景中的自主性。

尽管大模型赋能机器人具有广阔前景，但其工程落地仍面临关键难题。首先，大模型推理计算资源消耗大，与机器人实时控制需求存在矛盾；其次，大模型输出存在不确定性与“幻觉”风险，可能导致错误决策；再次，人形机器人涉及高自由度运动与复杂动力学约束，语言推理与物理控制之间存在较大鸿沟；此外，机器人应用场景往往具有高安全要求，需要构建可解释、可验

作者简介：牟多铎（1991-），男，博士研究生，研究方向为人工智能、计算机视觉、增强现实等。
通讯作者：牟多铎

证、可追溯的可信机制[3]。

基于上述背景，本文围绕“大模型赋能人形机器人智能交互与自主决策”展开研究，提出一种系统化技术架构，并结合典型应用场景分析其落地路径与挑战，为未来人形机器人智能化升级提供参考。

2 大模型赋能人形机器人智能交互与自主决策的机理分析

2.1 大模型的能力特征与机器人需求匹配

大模型在自然语言处理领域取得突破，核心在于其具备语义理解、上下文建模与生成式推理能力。通过预训练与指令微调，大模型能够理解复杂语言表达并生成具有逻辑连贯性的回答。同时，大模型具备较强的知识储备与推理能力，可在缺乏明确规则的情况下完成任务规划与决策支持。这些能力与人形机器人在开放环境中的智能需求高度匹配。

从人形机器人系统需求角度来看，其核心能力可概括为：环境感知、语义理解、意图识别、任务规划、运动控制、执行反馈与安全约束。传统方法往往采用模块化设计，将感知、规划与控制分离，导致信息传递效率低、语义表达能力弱，难以形成整体智能闭环。大模型能够在语义层面整合多源信息，实现从自然语言到任务序列的映射，从而提升系统的统一性与泛化能力[4]。

此外，大模型具有一定的“世界知识”与常识推理能力，可帮助机器人理解场景背景并生成合理行动策略。例如在家庭服务场景中，机器人接到“帮我准备早餐”指令时，需要理解早餐包含的物品、操作流程与安全注意事项，并在执行中根据实际环境动态调整。大模型在此类复杂任务中能够提供较强的推理支持。

然而，大模型能力并不等同于机器人可执行能力。机器人执行需要将高层语义规划映射到低层运动控制命令，这一过程涉及动力学约束、实时反馈与安全策略，因此必须构建合理的架构以实现语言推理与物理控制的有效衔接。

2.2 多模态交互的语义融合机制

人形机器人在真实环境中运行，需要处理多种类型信息，包括视觉图像、语音输入、文本指令、空间位置、力觉触觉等。多模态交互的关键在于实现不同模态信息的对齐与融合，形成统一语义表示。传统多模态融合方法多依赖特征拼接或注意力机制，但在复杂开放场景下仍存在语义漂移与信息不一致问题。

多模态大模型的发展为机器人多模态交互提供了新的解决方案。通过大规模跨模态预训练，多模态大模型能够在视觉与语言之间建立语义映射，实现图像理解、物体识别、场景描述与指令对齐。例如，当机器人识别到桌面上有杯子、面包与牛奶时，多模态模型可以将这些视觉信息转化为语义描述，并与用户语言指令结合形成任务意图。该机制能够显著提升机器人对场景的理解深度与交互自然度[5]。

在工程实现中，多模态交互通常包含三类融合：第一是感知层融合，将视觉、语音、触觉等原始数据整合为统一特征；第二是语义层融合，将多模态信息映射到统一语义空间；第三是决策层融合，将语义理解结果转化为行为策略。大模型主要在语义层与决策层发挥作用，而感知层仍需要依赖传统传感器融合算法与实时处理模块。

值得注意的是，多模态融合并非单纯的信息叠加，而是需要实现时间同步与空间对齐。例如机器人通过摄像头看到物体位置，同时通过语音听到用户指向“那个杯子”，系统必须将语音指代与视觉目标绑定，并进一步规划抓取动作。这要求多模态模型不仅具备语义理解能力，还需具备指代消解与空间推理能力。

2.3 自主决策的任务分解与规划机制

自主决策是人形机器人实现复杂任务执行的关键。传统机器人规划方法通常采用路径规划、行为树或强化学习策略，适用于结构化环境或固定任务流程，但难以应对开放场景下的多目标、多约束任务。大模型能够将自然语言任务转化为多步计划，并通过推理生成任务分解序列，从而提升规划能力[6]。

大模型赋能的任务规划机制通常包括：任务理解、子任务分解、资源与约束分析、动作序列生成、执行监控与反馈调整。具体而言，机器人接收用户指令后，大模型首先进行意图识别，将指令解析为目标与约束条件；随后基于知识推理生成子任务列表，并按逻辑顺序排列；接着结合

环境感知信息判断可执行性，并输出可执行动作计划。执行过程中，机器人通过感知反馈监控任务进度，当出现异常情况时，大模型可重新规划并调整策略。

这种规划方式的优势在于灵活性与泛化能力。即使面对从未训练过的新任务，大模型仍可通过语言推理生成合理步骤。但其缺陷在于缺乏物理可行性验证，可能生成超出机器人能力范围的计划，因此需要引入约束规划与执行验证机制，以确保输出策略的安全性与可执行性。

2.4 大模型“幻觉”风险与安全可控问题

大模型在生成式推理中存在“幻觉”现象，即输出看似合理但实际错误的信息。在机器人系统中，幻觉风险可能直接导致错误决策或危险动作。例如机器人误判物体属性、错误识别危险源或生成不合理动作序列，都可能引发安全事故。因此，大模型赋能机器人必须建立安全可控机制[7]。安全可控问题主要体现在三个方面：第一是输出可信性，即如何验证大模型生成计划的正确性；第二是行为可控性，即如何约束机器人执行过程不越界；第三是责任可追溯性，即如何记录决策过程以便审计与纠错。为此，需要构建多层安全架构，包括规则约束、可信推理验证、行为监控与紧急制动机制。

同时，机器人系统还涉及隐私安全与伦理问题。在家庭服务或公共服务场景中，机器人可能采集大量用户数据，若缺乏隐私保护机制，将引发社会风险。因此，大模型赋能机器人不仅是技术问题，更需要制度与治理框架的配套支持。

3 大模型赋能人形机器人智能交互与自主决策的总体技术架构

3.1 总体架构设计思路

基于大模型能力特征与人形机器人系统需求，本文提出一种面向智能交互与自主决策的总体技术架构，其核心目标是实现“感知—理解—规划—执行—反馈”的闭环系统。该架构采用分层设计思想，将机器人系统划分为感知层、语义理解层、任务规划层、执行控制层与安全治理层，并通过边云协同实现计算资源优化。

该架构的关键在于：将大模型作为高层语义推理与任务规划核心模块，同时保留传统控制算法在低层运动控制中的实时性优势，形成“高层智能+低层稳定”的协同模式。这种架构能够避免大模型直接输出底层控制指令带来的风险，并提升系统整体可靠性[8]。

3.2 多模态感知与环境建模模块

感知层是机器人理解环境的基础，主要包括视觉感知、语音感知、空间定位与触觉力觉感知。视觉感知通过 RGB-D 摄像头、激光雷达等获取环境信息，实现目标检测、语义分割与三维重建。语音感知模块负责语音识别与声源定位，为人机交互提供输入通道。空间定位模块通过 SLAM 算法构建地图并定位机器人位置。触觉力觉感知模块通过力矩传感器与触觉阵列获取接触信息，用于抓取与操作反馈。

在大模型赋能框架下，感知层输出的结果不仅包括传统的目标类别与坐标信息，还需要生成语义化场景描述。例如“桌面上有一个红色杯子，杯子在右侧 30 厘米处”。这种语义化输出能够更好地与大模型的语言推理能力结合，提高任务规划准确性[9]。

此外，环境建模需要实现动态更新与记忆机制。人形机器人在复杂环境中运行时，场景会不断变化，例如物体被移动、障碍出现或人类行为干预。系统应构建短期工作记忆与长期知识记忆，以支持持续交互与长期任务执行。

3.3 语义理解与交互生成模块

语义理解层主要负责自然语言指令解析、意图识别、对话管理与多模态语义融合。大模型在该层发挥核心作用，通过对话上下文与环境语义信息生成任务意图表示，并输出交互回应。例如用户提出“把杯子递给我”，系统需要理解“杯子”指代的具体目标，并结合视觉感知确定杯子位置，再生成动作规划。

交互生成模块不仅需要回答用户问题，还需具备主动交互能力。例如当机器人无法执行任务时，应主动询问用户补充信息或提出替代方案。这种交互能力能够显著提升用户体验，并减少误操作风险[10]。

语义理解层还应具备情境识别与情感交互能力。在服务场景中，机器人需要判断用户语气与

情绪状态，并生成适当回应。虽然情感交互不是自主决策的核心，但在长期人机共存场景中具有重要价值。

3.4 任务规划与自主决策模块

任务规划层是大模型赋能人形机器人自主决策的关键。该模块将用户目标转化为可执行的任务序列，并结合环境约束与机器人能力生成策略。任务规划通常包括高层规划与低层规划两部分。高层规划由大模型完成，输出子任务序列；低层规划由传统规划算法完成，包括路径规划、抓取规划与动作轨迹生成。

在该模块中，应引入“可执行性检查”机制，对大模型输出计划进行验证。例如大模型生成“先抓起杯子，再打开冰箱”，系统需判断机器人是否具备打开冰箱的能力、杯子是否可抓取、路径是否可行。可执行性检查可通过规则库、仿真验证或强化学习模型实现，从而提升决策可靠性[11]。

此外，自主决策模块需要具备实时反馈调整能力。当执行过程中出现异常情况，如目标物体掉落或路径受阻，系统应快速重新规划。大模型可以基于异常描述生成新计划，而低层控制模块则负责快速响应与稳定控制。

3.5 行为执行与运动控制模块

执行控制层负责将规划结果转化为机器人动作，包括步态控制、抓取控制、姿态平衡与运动轨迹跟踪。人形机器人运动控制具有高自由度、强耦合与动态稳定性要求，因此必须采用实时控制算法，如模型预测控制、全身控制（Whole-Body Control）与力矩控制策略。

大模型通常不直接参与低层控制，而是通过输出“动作意图”或“行为策略”与控制层交互。例如输出“走到桌子旁并伸手抓取杯子”，控制层则根据动作意图调用路径规划与抓取控制算法生成具体控制指令。这种方式能够保持控制系统的实时性与稳定性，并降低大模型不确定性带来的风险[12]。

在执行过程中，控制层还需要持续向大模型反馈状态信息，如任务完成情况、异常事件与环境变化。反馈信息应以结构化形式表示，以便大模型进行推理与再规划。

3.6 安全治理与可信约束模块

安全治理层是大模型赋能机器人系统不可或缺的组成部分。该模块主要包括权限管理、行为约束、风险评估、异常检测与紧急制动。对于高风险动作，如接触人体、搬运重物或使用工具，系统应设置严格的权限与安全阈值，避免大模型生成不安全指令。

可信约束机制可以采用多策略组合：第一，基于规则的安全约束，如禁止机器人进入危险区域；第二，基于模型的风险预测，通过学习模型评估动作风险；第三，基于仿真验证，对高风险计划进行虚拟测试；第四，人类监督机制，在关键决策环节引入人类确认。通过多层约束，可显著提升系统安全性[13]。

此外，安全治理还需涵盖数据安全与隐私保护，包括语音数据、图像数据与用户行为数据的加密存储与访问控制，防止数据泄露。对机器人决策过程的日志记录与可追溯机制也是重要内容，以便出现问题时能够定位原因并优化系统。

4 边云协同与工程化部署关键技术

4.1 边云协同的必要性与部署模式

大模型推理通常需要大量计算资源，而人形机器人本体受限于体积、功耗与散热条件，难以完全本地部署超大规模模型。因此，边云协同成为大模型赋能机器人落地的关键技术路径。边云协同主要包括三种模式：本地轻量模型+云端大模型、本地中等模型+边缘计算节点、以及多级分布式协同推理。

在家庭服务或公共服务场景中，云端大模型能够提供更强推理能力，但网络延迟与数据安全问题较突出；在工业制造或应急救援场景中，网络可能不稳定，因此更适合采用边缘计算节点部署，确保实时性与可靠性。总体而言，边云协同的核心在于根据任务实时性需求与数据敏感性选择合适部署模式[14]。

此外，模型压缩与蒸馏技术能够将大模型能力迁移到轻量模型，使机器人在本地实现基础交

互与决策能力，云端则负责复杂推理与知识更新。这种分工能够在性能与成本之间取得平衡。

4.2 实时控制约束与推理延迟优化

人形机器人控制系统对实时性要求极高，尤其在步态控制与动态平衡中，控制周期往往需要达到毫秒级。大模型推理延迟通常在百毫秒甚至秒级，难以直接嵌入控制环路。因此，大模型与控制系统必须采用异步协同机制：大模型负责高层决策与规划，控制系统负责实时执行与稳定控制。

为降低推理延迟，可采用以下优化策略：第一，采用模型量化、剪枝与推理加速框架提升计算效率；第二，采用缓存机制存储常见任务规划模板，减少重复推理；第三，采用分阶段推理，将复杂任务拆分为多个小推理步骤；第四，利用边缘计算节点分担推理负载。通过这些方法，可在一定程度上满足机器人实时交互需求[15]。

此外，应建立实时优先级机制。当机器人处于动态平衡或危险状态时，应优先执行控制系统的安全策略，而非等待大模型推理结果。

4.3 数据闭环与持续学习机制

人形机器人在真实环境中运行会产生大量数据，包括视觉、语音、运动轨迹、触觉反馈与任务日志。这些数据是系统持续学习与优化的重要资源。通过构建数据闭环机制，机器人可以将执行结果反馈给大模型，并用于后续模型微调与知识更新。

持续学习机制主要包括在线学习与离线学习两类。在线学习可在任务执行中通过强化学习或自监督学习更新局部策略，提高适应性；离线学习则通过集中数据训练优化大模型能力，提升长期性能。由于机器人数据具有隐私与安全敏感性，持续学习必须建立严格的数据治理机制，包括匿名化处理、权限控制与安全存储[16]。

此外，应建立场景知识库与经验库，将常见任务流程、失败案例与安全规则结构化存储，以支持大模型推理与规划。经验库的引入能够提升机器人决策的稳定性与可解释性。

5 典型应用场景分析与落地模式研究

5.1 智能制造与柔性生产场景

在智能制造领域，人形机器人可用于装配、搬运、检测与协作生产。相比传统工业机械臂，人形机器人能够适配现有工厂布局，无需大规模改造生产线，并能执行需要双足移动与双手操作的任务。大模型赋能可使机器人理解工艺指令、自动规划装配步骤，并在异常情况下进行调整。

例如在装配任务中，机器人需要识别零件类型、理解装配顺序并执行精细操作。大模型可根据装配手册生成操作步骤，并结合视觉识别结果判断零件匹配情况，从而实现柔性生产。与此同时，通过多模态交互，工人可用自然语言指导机器人完成任务，降低编程门槛，提高生产效率[17]。

然而，制造场景对精度与可靠性要求极高，机器人必须具备高精度定位与力控能力，并通过安全机制防止误操作。因此，大模型规划结果必须经过严格验证，并与工业控制系统深度集成。

5.2 家庭服务与智慧康养场景

家庭服务是人形机器人最具潜力的应用场景之一，包括清洁整理、物品递送、陪护照料与健康监测等。大模型赋能能够显著提升机器人交互自然度，使其能够理解复杂家庭指令并执行多步任务。例如用户提出“帮我把药拿过来并倒一杯水”，机器人需要理解任务组合并规划执行顺序。

在智慧康养场景中，人形机器人可为老年人提供陪护、健康提醒与紧急求助服务。大模型能够根据对话判断用户需求，结合传感器数据识别异常情况，并主动发起交互或报警。此外，多模态感知可用于跌倒检测与行为分析，提高安全保障能力[18]。

但家庭场景涉及高度隐私数据，如家庭环境图像与用户语音信息，因此必须建立严格的数据安全机制。同时，机器人在与老人或儿童互动时必须确保行为可控，避免因误判导致伤害风险。

5.3 公共服务与智慧城市场景

在智慧城市建设中，人形机器人可应用于导览咨询、公共巡检、安防协助与设施维护等领域。例如在机场、医院或政务大厅，机器人可提供智能问答、路线指引与信息查询服务。大模型能够支持更复杂的自然语言交互，并通过知识库调用实现实时信息服务。

在城市巡检中，人形机器人可进入狭窄或危险区域进行检测，结合视觉识别与传感器数据判断设施状态。大模型可根据巡检任务生成路线规划与检测流程，并在发现异常时自动生成报告。相比传统巡检机器人，人形机器人具有更强的环境适应性与操作能力，可执行开门、按按钮、搬运工具等任务[19]。

但公共场景人流密集，机器人必须具备高可靠的避障与安全控制能力，同时需要符合公共安全与伦理规范，避免出现误识别与误决策导致的社会风险。

5.4 应急救援与危险作业场景

应急救援是人形机器人重要的高价值应用领域。在火灾、地震、矿难等灾害现场，人形机器人可代替救援人员进入危险区域执行侦察、搬运、破拆与生命探测任务。大模型赋能能够使机器人在复杂环境中理解救援指令并动态规划行动路径，提高自主性。

在救援场景中，机器人需要处理高不确定性环境，如烟雾遮挡、地形破坏与通信中断。大模型可结合多模态感知生成环境语义地图，并通过推理判断最佳行动策略。同时，机器人可通过自然语言与救援人员协作，实现高效指挥与任务分配[20]。

然而，救援场景对实时性与可靠性要求极高，任何决策错误都可能导致重大损失。因此，大模型赋能救援机器人必须采用高冗余安全机制，并在关键决策环节引入人类监督确认。

6 关键挑战与发展趋势分析

6.1 技术挑战：从语言推理到物理执行的鸿沟

尽管大模型具备强大推理能力，但其输出通常以语言或符号形式表达，难以直接转化为物理可执行动作。人形机器人执行涉及动力学约束、实时控制与环境不确定性，因此需要构建“语义—动作”映射机制。未来研究应加强基于行为树、技能库与强化学习的中间层表示，使大模型输出能够转化为可验证的动作策略。

同时，多模态感知误差会影响大模型推理准确性。视觉识别错误或空间定位偏差可能导致机器人执行失败。因此，需要提高感知鲁棒性，并建立感知与推理的协同纠错机制。

6.2 工程挑战：算力、能耗与成本约束

人形机器人产业化必须考虑硬件成本与能耗限制。大模型推理需要高算力芯片与存储资源，若完全本地部署将显著增加成本与功耗。未来需要发展更高效的边缘推理芯片、模型压缩技术与轻量化多模态模型，以实现低成本部署。同时，边云协同架构需要稳定网络支持，在弱网环境下仍需保持基本智能能力。

6.3 安全挑战：可信对齐与责任治理

大模型输出的不确定性与幻觉风险是机器人安全应用的核心挑战。未来需要发展面向机器人系统的可信对齐方法，使大模型决策符合安全规则与伦理约束。同时，需要建立机器人行为审计与责任追溯机制，确保决策过程可解释、可记录、可验证。

在社会治理层面，机器人应用涉及隐私保护、数据安全与伦理规范，未来应制定相关标准与法规，推动产业健康发展。

6.4 发展趋势：通用智能体与群体协作

未来人形机器人发展将从单体智能走向群体协作。通过多机器人协同与云端知识共享，机器人可实现任务分工与集体学习，提高整体效率。同时，通用智能体框架将成为关键方向，即机器人具备跨场景迁移能力，可在不同任务中快速适应并持续学习。大模型作为通用智能底座，将在其中发挥核心作用。

7 结论

本文围绕“大模型赋能人形机器人智能交互与自主决策”展开研究，分析了大模型在语义理解、多模态融合、任务规划与决策生成方面的能力优势，并提出面向工程落地的总体技术架构。研究表明，大模型赋能人形机器人实现智能交互与自主决策的关键在于构建“多模态感知—语义理解—任务规划—执行控制—反馈优化”的闭环体系，并通过边云协同解决算力与实时性矛盾。

同时，为应对大模型幻觉风险与安全隐患，必须引入多层可信约束机制，涵盖输出验证、行为监控、异常检测与紧急制动等策略。在典型应用场景中，智能制造、家庭服务、智慧城市与应急救援均具备较强落地潜力，但也对系统可靠性、隐私保护与安全治理提出更高要求。未来研究应进一步加强语义推理与物理执行的融合机制，推动轻量化模型部署与可信对齐技术发展，并完善标准化与治理体系，为人形机器人规模化应用奠定基础。

参考文献：

[1] 李德毅，王飞跃. 智能机器人发展趋势与关键技术[J]. 中国科学：信息科学，2020，50(9)：1301-1315.
[2] 刘宏，马会娟. 大规模预训练模型的技术演进与应用前景[J]. 软件学报，2021，32(10)：3010-3025.
[3] 马会娟，刘宏. 生成式人工智能的安全风险与治理框架研究[J]. 情报杂志，2023，42(7)：12-20.
[4] 马会娟，张立. 人形机器人运动控制研究进展[J]. 机器人，2020，42(6)：785-798.
[5] 周志华，刘知远. 多模态深度学习研究综述[J]. 计算机学报，2021，44(7)：1253-1270.
[6] 王飞跃，李德毅. 面向复杂任务的智能体规划方法研究[J]. 自动化学报，2022，48(3)：521-534.
[7] 刘伟，张铭. 大语言模型幻觉问题及其可信控制研究[J]. 计算机研究与发展，2023，60(11)：2411-2424.
[8] 赵鑫，王田苗. 人形机器人系统架构与集成技术研究[J]. 机械工程学报，2021，57(18)：1-14.
[9] 马会娟，李勇. 机器人视觉感知与场景理解技术研究进展[J]. 电子学报，2020，48(12)：2501-2512.
[10] 张尧学，朱松纯. 人机交互智能化发展趋势与关键问题[J]. 中国科学基金，2021，35(4)：521-528.
[11] 李群，马会娟. 机器人任务规划与行为树方法研究综述[J]. 控制与决策，2022，37(9)：2141-2152.
[12] 王田苗，赵鑫. 人形机器人全身控制与动态平衡关键技术[J]. 机器人，2021，43(5)：577-590.
[13] 刘宏，李强. 智能机器人安全控制与可信决策机制研究[J]. 自动化学报，2023，49(6)：1123-1136.
[14] 马会娟，张浩. 边缘计算在智能机器人中的应用研究[J]. 计算机科学，2021，48(10)：23-31.
[15] 陈杰，王飞跃. 大模型推理加速与边缘部署技术研究[J]. 计算机工程与应用，2023，59(18)：1-10.
[16] 李勇，周波. 机器人持续学习与数据闭环优化机制研究[J]. 软件学报，2022，33(8)：2635-2650.
[17] 张立，赵鑫. 人形机器人在智能制造中的应用模式与关键技术[J]. 制造业自动化，2022，44(12)：15-22.
[18] 马会娟，陈晓. 服务机器人在智慧养老中的应用现状与发展路径[J]. 科技管理研究，2021，41(16)：180-187.
[19] 王强，李宏. 智慧城市巡检机器人技术体系与应用研究[J]. 城市发展研究，2022，29(5)：67-74.
[20] 赵鑫，王田苗. 应急救援机器人关键技术与发展趋势[J]. 工程科学学报，2020，42(9)：1120-1128.

Large Model–Empowered Technical Architecture and Application Study for Intelligent Interaction and Autonomous Decision-Making in Humanoid Robots

MOU Duoduo

(Universiti Teknologi Malaysia, Skudai, Johor 81310, Malaysia)

Abstract: With the rapid advancement of artificial intelligence, large-scale models have demonstrated notable strengths in natural language understanding, multimodal fusion, knowledge reasoning, and generative interaction, providing a new technical foundation for robotic systems to evolve from “function-oriented automation” toward “general-purpose intelligent agents.” As a key direction in intelligent equipment, humanoid robots—owing to their human-like morphology and compatibility with human environments—show broad application potential in intelligent manufacturing, public services, healthcare and

eldercare, and emergency rescue. However, conventional humanoid-robot systems still fall short in semantic understanding, complex task planning, cross-scenario generalization, and safe controllability in open environments, making it difficult to meet the practical demands of highly dynamic and uncertain real-world settings. In response, this paper focuses on the key issues of large model–empowered intelligent interaction and autonomous decision-making in humanoid robots. It systematically analyzes the roles of large models in multimodal perception fusion, semantic understanding, intent recognition, task decomposition, policy generation, and behavioral execution, proposes an engineering-oriented overall technical architecture, and further discusses implementation pathways for edge–cloud collaborative deployment, real-time control constraints, trustworthy safety governance, and typical application scenarios. The study suggests that the core of large model–driven upgrading of humanoid-robot intelligence lies in building a closed-loop framework of "multimodal perception–semantic understanding–knowledge reasoning–task planning–execution feedback," while enhancing system reliability and controllability through constraint mechanisms, alignment strategies, and safety controls. This work provides a reference for the industrial deployment of humanoid robots and the engineering design of intelligent-agent systems.

Keywords: Large-scale models; Humanoid robots; Multimodal interaction; Autonomous decision-making; Task planning; Edge–cloud collaboration; Trustworthy safety