DOI 10.15302/J-SSCAE-2026.03.017

智能体与机械臂融合技术研究进展及展望

李道童[1]，赵晓彬[2]，芦飞[1]，宋璐杰[1*]，王东清[1]，李盛新[1]

（1. 浪潮电子信息产业股份有限公司，济南 250101；2. 北京科技大学计算机与通信工程学院，北京 100083）

摘要：面向复杂环境下机械臂操作任务对自主决策、长时序执行与多约束交互的需求，智能体（Agent）与机械臂系统的融合逐渐成为具身智能的重要技术路径。为支撑该方向的研究梳理与工程落地，本文构建了Agent–机械臂融合研究的结构化技术框架，明确了方法体系、数据评测脉络及关键挑战。回顾机械臂智能化的技术基础与控制范式演进，从强化学习型Agent、规划与认知型Agent以及大模型驱动具身Agent三个层级，对代表性方法的系统结构、决策机制与典型应用进行归纳；进一步梳理典型公开数据集，给出数据与选型要点。研究表明，现有Agent–机械臂融合研究已形成以学习型策略、规划推理机制与大模型驱动闭环为核心的三类技术路线，不同路线在任务泛化能力、交互复杂度适应性与部署成本方面呈现差异；同时，数据分布覆盖不足、评测协议不统一等问题制约了结论的可复现性与可比性。面向通用操作Agent的发展，未来研究需重点关注数据与任务覆盖扩展、跨域迁移与鲁棒性提升、安全可靠机制及工程化部署流程等方向，为相关方法由原型验证走向规模化应用提供支撑。

关键词：智能体；机械臂；任务规划与决策；具身智能

中图分类号：TP242　　**文献标识码：**A

Advances and Future Outlook in the Convergence of Agents and Robotic Manipulators

Li Daotong[1], Zhao Xiaobin[2], Lu Fei[1], Song Lujie[1*], Wang Dongqing[1], Li Shengxin[1]

(1. Inspur Electronic Information Industry Co., Ltd., Jinan 250101, China; 2. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: To address the needs of robotic manipulators in complex environments for autonomous decision-making, long-horizon execution, and multi-constraint interactions, the integration of intelligent agents with robotic manipulator systems is increasingly recognized as a key technical pathway for embodied intelligence. To support research review and engineering implementation in this direction, this study seeks to establish a structured technical framework for agent–manipulator integration, clarifying the methodological landscape, data-and-evaluation context, and key challenges. First, the study reviews the technical foundations of manipulator intelligence and the evolution of control paradigms. Then, it summarizes representative methods in terms of system architectures, decision-making mechanisms, and typical applications from three levels: reinforcement-learning-based agents, planning and cognition-oriented agents, and large-model-driven embodied agents. Furthermore, we examine representative public datasets and provide key considerations for dataset selection. Our analysis shows that existing agent–manipulator integration research has formed three major technical routes centered on learning-based policies, planning-and-reasoning mechanisms, and large-model-driven closed-loop frameworks, and these routes differ in task generalization capability, adaptability to interaction complexity, and deployment cost.

收稿日期：2026-03-08；**修回日期：**2026-06-14

通讯作者：*宋璐杰，浪潮电子信息产业股份有限公司研究员，研究方向为图像解译与大模型应用；E-mail: 2569332006@qq.com

资助项目：国家自然科学基金项目(62401062)；中央高校科研业绩经费(FRF-TP-25-038)

本刊网址：sscae.engineering.org.cn

Meanwhile, insufficient coverage of data distributions and non-unified evaluation protocols hinder the reproducibility and comparability of reported findings. Toward the development of general-purpose manipulator agents, future research should prioritize expanding data and task coverage, improving cross-domain transfer and robustness, developing safety and reliability mechanisms, and establishing engineering-oriented deployment pipelines, thereby supporting the transition of relevant methods from prototype validation to scalable application.

Keywords: agent; robotic manipulator; task planning and decision-making; embodied intelligence

一、前言

近年来，机器人技术与人工智能持续融合，具身智能逐渐成为学术界和产业界关注的重要方向。机械臂作为具身操作系统的核心执行单元，需要在真实环境中完成感知交互、运动执行与任务反馈，在智能制造、仓储物流、医疗辅助和家庭服务等场景中具有广泛应用潜力[1]。与结构化工业环境中依赖精确建模和规则控制的方法相比，开放环境中的对象属性、空间关系和接触状态更加不确定，实际任务也常具有长时序、多步骤和强依赖特征。因此，机械臂系统的能力需求已从单纯轨迹跟踪和低层控制，扩展到任务理解、决策组织、在线调整和失败恢复等更高层次智能行为[2]。

从应用需求看，机械臂作业正由固定流程执行转向面向动态目标的自主操作。制造场景中的柔性生产、物流场景中的不确定抓取与分拣，以及服务和特种作业中的开放目标与安全约束，都要求机械臂能够根据环境变化动态组织行动。随着视觉、触觉、力反馈等多模态感知能力以及边缘计算和实时通信机制的发展，智能体（Agent）范式为机械臂智能化提供了新的系统组织方式。相较于传统“预设轨迹–被动执行”的控制链路，Agent更强调围绕任务目标构建“感知–决策–执行–反馈”闭环，使系统能够持续更新状态认知、调整行动策略并重组执行过程。“视觉–语言–动作”（VLA）统一建模和通用策略底座的发展，也进一步推动机械臂操作研究由单任务优化走向语义条件控制、知识迁移与跨任务泛化[3-4]。

尽管Agent与机械臂融合研究取得了显著进展，但从实验室验证走向真实部署仍面临一系列关键瓶颈。首先，长时序组合任务中的误差具有持续累积特征，状态估计偏差、动作执行偏差与环境扰动相互耦合，容易导致计划层与执行层之间出现不一致，进而削弱系统的任务恢复能力。其次，跨场景泛化和跨平台迁移仍然不足，长尾任务分布、对象多样性以及硬件平台异构性，使已有方法即便在大规模数据上训练，也难以保证在新环境中的稳定表现[5-6]；同时，真实交互数据采集成本高、失败样本和边界样本稀缺，仿真数据与现实环境之间的偏差在长链条任务中往往被进一步放大，从而限制了数据驱动方法的有效扩展[7-8]；此外，现有评测协议和监督维度仍以阶段性性能或单次成功率为主，对长时序稳定性、跨任务迁移性和失败恢复能力的刻画仍显不足，导致不同方法之间的比较缺乏足够一致的评价基础[9-10]；最后，受限于实时性约束、模块通信延迟及系统集成复杂度，端到端智能推理尚难以直接承担高频率、强鲁棒要求的控制闭环。

面对上述挑战，Agent与机械臂融合研究已由以模型和规则为中心的控制范式，逐步发展到以模仿学习和强化学习为代表的数据驱动策略学习，并进一步迈向由大模型和VLA统一建模支撑的具身Agent框架[11]。本文所讨论的“Agent与机械臂融合”是面向复杂作业任务的多层协同框架，强调高层任务规划与低层运动控制协同、多模态信息与动作表示统一建模，以及大模型、策略模型、规划器和执行模块之间的系统集成；其中，语言候选生成与可行性校验等结构化机制，为缓解“语义合理但物理不可行”提供了重要思路[12]。基于此，从强化学习型Agent、规划与认知型Agent和大模型驱动具身Agent三条技术路线梳理代表性研究，分析其任务适用性、系统组织方式与主要局限，并结合公开数据集、基准平台和评测协议，总结关键问题与未来发展趋势。

二、机械臂智能化技术基础与方法演进

机械臂智能化技术最初以运动学、动力学建模和轨迹规划为基础，强调在结构化环境中实现高精度、可重复和稳定执行。随着应用场景由重复作业拓展到复杂操作、动态交互和开放环境任务，传统方法对模型准确性、环境先验和任务边界的依赖逐

渐成为限制因素。尤其在模型误差、接触不确定性和长时序任务耦合条件下，仅依赖预定义模型和规则的控制体系难以满足自适应决策、在线调整和跨任务泛化需求，由此推动机械臂研究从传统控制与运动规划，逐步扩展到数据驱动学习和Agent闭环决策。

如图1所示，Agent与机械臂融合技术的演进并非不同阶段之间的简单替代，而是围绕任务复杂度、决策层级和系统能力的逐步扩展。早期研究主要依托运动学/动力学建模、轨迹规划及“比例–积分–微分”（PID）控制、阻抗控制等方法，侧重低层控制精度和结构化环境下的稳定执行；随后，模仿学习、行为克隆和逆强化学习等方法被用于从示教数据中学习控制映射，提高机械臂对任务经验的利用能力；在此基础上，深度确定性策略梯度（DDPG）、双延迟深度确定性策略梯度（TD3）、软演员–评论家（SAC）等深度强化学习方法以及分层强化学习、仿真到现实（Sim-to-Real）迁移进一步推动了技能学习、策略优化和技能复用。进一步地，规划与认知型Agent强调任务分解、约束推理、任务与运动规划（TAMP）以及世界模型/预测等能力，以支撑长时序任务中的可行性推理；大模型驱动具身Agent则进一步引入VLA、大语言模型/视觉语言模型（LLM/VLM）推理和行为基础模型，更加关注语义理解、多模态泛化与跨任务迁移。总体来看，该方向正由局部控制优化逐步转向高层决策、低层执行与工程部署协同发展的系统性Agent框架。

如图2所示，当前工业与服务场景中的机械臂在结构形式和应用定位上存在差异。按照国际标准化组织（ISO）的术语体系，工业机器人通常指自动控制、可重编程、多用途、具有三轴及以上自由度的操作机构，可固定安装，也可与移动平台集成，用于搬运、装配和加工等任务[13]。在此基础上，依据国际机器人联合会（IFR）常用的机械结构分类方式，机械臂通常可分为直角坐标/龙门型、选择顺应性装配机械臂（SCARA）型、关节型、并联/Delta型和圆柱型等类别[14]。不同结构在工作空间、可达性、刚度、负载和速度方面各有特点，影响运动学建模、控制策略和Agent系统接口设计。

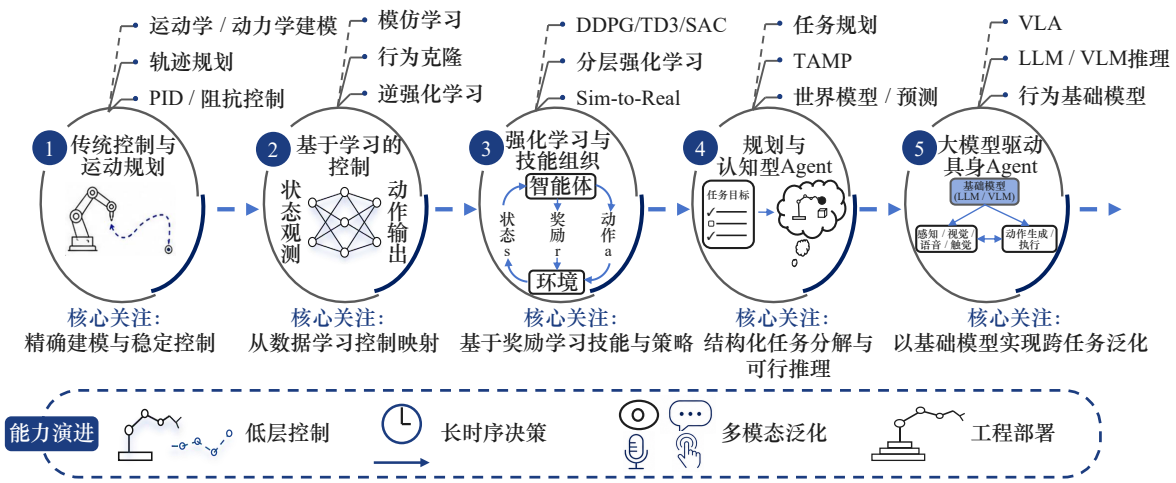


图1　Agent与机械臂融合技术演进脉络图



(a) 直角坐标/龙门型　(b) 水平多关节型　(c) 关节型　(d) 并联型　(e) 圆柱型

图2　典型机械臂结构类型示例

除结构差异外，作业模式也影响 Agent – 机械臂融合系统的组织方式。典型模式包括单机械臂独立作业、双臂或多臂协同、移动机械臂作业及多机器人 / 群体协同作业。前者更强调感知定位、轨迹规划、低层控制和局部失败恢复；后者则进一步涉及任务分配、时序同步、通信协同、资源调度、冲突避免和多主体状态一致性等问题。因此，多智能体机制可为群体任务分解、协作策略生成和全局目标一致性维护提供支撑，但通信延迟、信用分配、协同安全和系统可扩展性仍是后续研究需要解决的关键问题。

（一）传统机械臂控制与运动规划方法

机械臂传统控制体系首先建立在运动学与动力学建模基础之上。其中，运动学建模主要描述末端执行器位姿与关节变量之间的映射关系，基于 Denavit – Hartenberg 参数体系构建的齐次变换方法为正、逆运动学求解提供了统一表达框架[15]。针对冗余机械臂，零空间优化常被用于嵌入避障、关节限位规避等次级目标，从而在满足主任务约束的同时提升执行可行性与安全性[16]。然而，无论是解析求解还是数值迭代方法，其稳定性与收敛性通常都与机构形式、约束条件和初值设置密切相关。在动力学层面，传统方法通常基于拉格朗日方程或牛顿 – 欧拉递推算法建立关节力矩与系统运动状态之间的定量关系[17]。基于此，计算力矩控制可通过补偿非线性耦合项提高轨迹跟踪精度；在存在参数不确定性和外部扰动时，则往往需要进一步结合非线性控制或鲁棒控制设计以保证系统稳定性[18]。对于装配、人机协作等接触任务，阻抗控制与导纳控制通过建立力 – 位姿之间的动态关系，使机械臂能够表现出一定程度的可控柔顺性，因此成为交互型任务的重要基础[19]。但总体来看，这类方法依然较强依赖模型准确性和参数整定质量，在高不确定接触环境中对任务条件和调参策略较为敏感。相关顺应控制与力控制综述也指出，面向复杂交互任务，力控设计已难以仅依赖单一控制律完成性能保障，而需要与感知反馈、任务约束和在线调节机制协同考虑，以进一步提升系统鲁棒性和可复现性[20]。

在控制之外，轨迹规划是传统机械臂系统实现任务执行的另一核心环节，其目标是在满足运动学、碰撞和工艺约束的前提下生成可执行路径。对于重复性较强、流程相对固定的任务，多项式插值和样条曲线方法可通过对关节空间或操作空间施加连续性约束生成平滑轨迹，因此在工业作业中应用广泛。面向存在障碍物或高维约束的复杂环境，采样型规划方法通过在可行空间中随机采样来缓解组合爆炸问题，代表性方法包括快速随机树（RRT）[21]和概率路图（PRM）[22]。这类方法在高维空间中具有较好的扩展性，但其生成结果往往仍需进一步进行平滑处理，而且难以直接满足动力学一致性和实时执行要求。

相较于仅关注几何可行性的规划方法，基于优化的轨迹规划进一步将路径生成转化为带约束的优化问题，从而能够在统一框架下考虑碰撞约束、轨迹平滑性与动力学一致性。以协变哈密顿优化运动规划（CHOMP）为代表的方法，通过对轨迹进行连续优化提升路径质量，因此在复杂操作任务中表现出较强的工程适用性[23]。在执行层面，PID 控制仍是工业系统中最常用的跟踪策略[24]，而分辨加速度控制等方法则通过操作空间误差反馈实现更高精度的跟踪性能[25]。总体而言，传统规划与控制体系在结构化环境中具有高精度和高稳定性的优势，但其感知、决策与执行链路通常相对分离，因而在开放环境、动态场景和长时序任务中往往面临较高的建模与扩展成本。相关综述表明，随着动态环境操作与人机共域任务需求增加，采样、优化与学习相融合的规划框架正逐步成为研究趋势，与之相应，评测标准也由单纯关注“几何可行”逐步扩展到“实时可行、可恢复以及安全约束可满足”等更具任务导向的维度[26]。

（二）基于学习的机械臂控制方法

为降低传统控制方法对精确模型和先验规则的依赖，机械臂研究逐步引入监督学习与模仿学习方法，将控制问题转化为状态到动作的映射学习任务。其中，行为克隆和学习示教方法通过拟合专家演示直接获得控制策略，在示教覆盖较充分时能够取得较为稳定的执行效果[27]；但当测试状态偏离训练分布时，这类方法容易出现误差累积和性能退化。为提升策略在更广泛状态空间中的一致性，逆强化学习尝试从专家行为中恢复潜在奖励函数，从而间接学习控制目标，其理论上具有更强的泛化潜力，但求解过程更为复杂，且对示教质量与样本分布较为

敏感[28]。在此基础上，数据集聚合等方法通过在训练过程中持续引入专家修正，缓解了分布偏移对闭环执行的影响，从而提升了策略部署阶段的稳定性[29]。

近年的示教学习综述进一步表明，示教数据的获取方式、质量控制与评测协议已成为影响结论可复现性的关键因素，相关研究正在由单纯关注算法改进，逐步转向“数据–训练–评测”协同优化的研究范式[30]。因此，对模仿学习方法的比较不仅应报告平均性能，还需要说明示教来源、覆盖范围与关键失败状态比例等条件，避免数据差异掩盖方法差异。

与依赖专家示教的模仿学习不同，强化学习通过与环境持续交互完成策略优化，为机械臂控制提供了一条无需显式动力学建模的学习路径。相关综述普遍认为，强化学习在复杂动力学系统、接触交互任务以及策略自我改进方面具有较强表达能力，但训练稳定性不足、样本效率偏低以及真实系统中的安全风险，仍然是其面向实际部署的主要障碍[31]。在连续动作空间控制中，基于深度策略梯度的算法被广泛应用，如DDPG等方法实现连续动作策略学习[32]，近端策略优化（PPO）等方法通过约束策略更新提升训练稳定性[33]。为提升复杂任务中的学习效率与任务复用能力，分层强化学习引入时间抽象与技能组织机制，选项框架为高层决策与低层控制解耦提供了形式化基础[34]。

在实际部署中，Sim-to-Real迁移是强化学习应用的关键问题。域随机化通过在仿真环境中扰动物理参数与视觉条件来增强策略泛化能力，是较为常用的迁移手段[35]。然而，在复杂接触任务和长时序任务中，迁移误差通常并非来自单一因素，而是由闭环控制差异、观测噪声累积以及未建模扰动等因素共同作用所致。近年的Sim-to-Real综述更倾向于从状态／观测、动力学、奖励与任务设定、训练流程、评测协议等多个层面分解误差来源，并将域随机化、系统辨识、表示不变性学习以及少量真实数据微调等方法纳入统一对比框架，以提升迁移研究的结构化程度和结果可复现性[36]。在上述方法的演进基础上，当前Agent–机械臂融合研究已经形成较为清晰的系统组织框架。为增强不同技术路线之间的可比性，本研究进一步从感知与数据层、智能决策层、执行与反馈层三个层面，对Agent–机械臂融合的总体结构进行概括。如图3所示，Agent–机械臂融合的核心并不在于简单叠加智能算法和机械臂硬件，而在于建立稳定的系统接口和闭环协同关系。在感知与数据层，视觉感知、语言指令、触觉／力

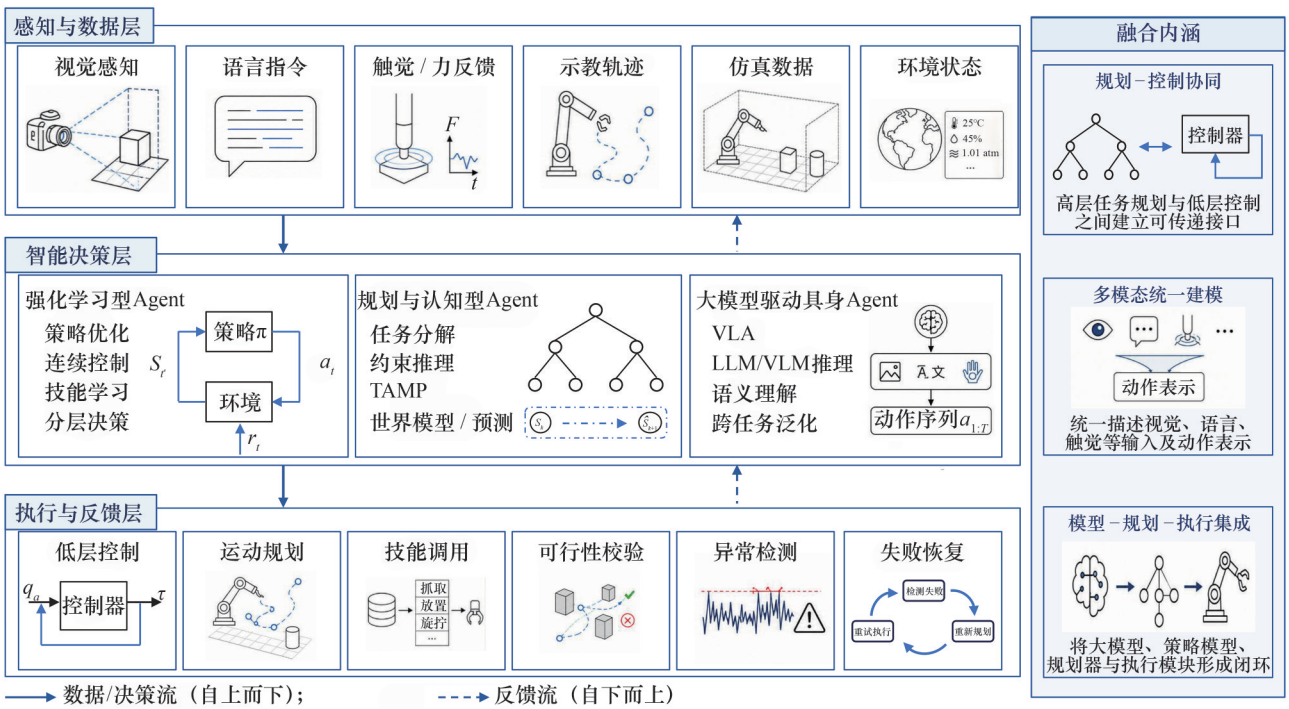


图3　Agent–机械臂融合技术路线总览图

反馈、示教轨迹、仿真数据以及环境状态共同构成任务状态理解与策略学习的输入基础；在智能决策层，强化学习型Agent主要承担策略优化与技能学习，规划与认知型Agent主要承担任务分解与约束推理，大模型驱动具身Agent则主要承担语义理解、多模态推理和跨任务泛化；在执行与反馈层，低层控制、运动规划、技能调用、可行性校验、异常检测和失败恢复机制共同保障高层决策能够转化为物理可执行的动作过程。对于“VLA+规划”“LLM+TAMP”“策略模型+安全控制”等混合架构而言，图3中的三类智能决策模块并非孤立运行，而是通过任务接口、可行性校验和执行反馈形成协同关系。

（三）从学习模型到Agent范式的转变

随着学习方法在机械臂控制中的深入应用，研究对象逐渐由单一模型或局部策略扩展为面向完整任务过程的智能体系统。传统学习模型主要优化特定的状态–动作映射，通常依赖固定任务定义和预设数据分布；Agent范式则将环境感知、状态更新、任务组织、技能执行和反馈修正纳入统一闭环，使机械臂能够依据执行结果持续调整决策，并在多阶段任务中保持目标与策略的一致性。这一转变使研究重点由“如何生成单步动作”进一步拓展到“如何组织、监控并完成完整任务”。

在系统实现层面，采用Agent并不意味着使用单一端到端模型取代传统的感知、规划和控制模块，而是通过统一状态表示与任务接口增强各模块之间的协同。当前具身Agent通常采用“高层任务生成或选择–低层技能执行–状态反馈与再规划”的分层闭环结构，并结合可行性筛选、异常检测和回退机制约束执行链路，从而降低“语义合理但物理不可行”的风险[37]。因此，Agent范式的核心价值在于提高长时序任务中的适应性、稳定性和失败恢复能力。

（四）Agent–机械臂融合方法的分类标准与交叉关系

需要说明的是，Agent–机械臂融合方法之间并不存在完全割裂的边界，许多实际系统往往同时包含策略学习、任务规划、多模态理解和低层控制等多个模块。因此，对相关研究的分类并非依据单一算法名称，而是依据方法在系统中的主导功能、决策层级和建模对象。具体而言，重点考虑三个方面：第一，方法主要作用于机械臂系统的哪一决策层级，即低层连续动作生成、中高层任务组织，还是高层语义理解与任务泛化；第二，方法的核心建模对象是什么，即状态–动作映射、任务结构与约束关系，还是VLA等多模态统一表示；第三，方法在系统闭环中承担何种功能角色，即主要用于技能学习与策略优化、规划推理与可行性验证，还是语义解析、候选生成与跨任务迁移。

基于上述原则，将强化学习型Agent理解为以策略学习、交互优化、示教初始化和技能生成为核心的一类低层或中低层决策方法。严格来说，模仿学习、行为克隆、扩散策略和动作分块Transformer等方法在优化目标和监督信号上与狭义强化学习存在差异，但在机器人学习系统中，它们常与强化学习共同服务于状态–动作映射学习、策略初始化、探索成本降低和复杂行为生成。例如，模仿学习可利用专家示教为强化学习提供初始策略或经验分布，从而缓解真实机械臂在线试错成本高、样本效率低和安全风险大的问题；生成式策略则可在离线数据基础上建模多模态动作分布，为后续策略优化或闭环执行提供更稳定的动作先验。因此，在强化学习型Agent路线中讨论这类与策略学习密切相关的方法，并在具体论述中区分其数据依赖、优化机制和系统功能。

规划与认知型Agent主要指以任务结构表达、约束推理、任务–运动协同规划和内部预测机制为核心的方法。该类方法通常不直接学习端到端动作映射，而是通过显式建模任务步骤、对象关系、空间约束和运动可行性，为长时序任务组织和执行过程校验提供结构化支撑。因此，TAMP、神经引导规划、任务图和用于辅助规划的世界模型等方法被归入该类路线。

大模型驱动具身Agent主要指以大规模预训练模型、多模态语义理解和VLA统一建模为核心的方法。该类方法的主导作用在于开放指令理解、任务分解、候选动作生成、跨任务泛化和高层推理，而不是单纯完成低层控制优化。对于“VLA+规划”“LLM+TAMP”“策略模型+安全控制”等交叉方法，将其视为混合型架构，并依据在系统中的主要功能分别在相关章节中讨论，以避免将复杂系统强

行归入单一类别。

如表1所示，三类路线在决策层级、数据依赖、可解释性、物理可行性保障、跨平台适配、实时性和部署难点上各有侧重：强化学习型Agent擅长连续动作优化和技能学习，但训练成本及真实部署风险较高；规划与认知型Agent擅长任务结构表达、约束推理和可解释性，但实时求解能力受限；大模型驱动具身Agent具备语义理解和跨任务泛化潜力，却需结合可行性校验与低层控制保障执行稳定。三类路线功能互补，单一路线难以兼顾长时序推理、物理可行性、实时控制与跨平台部署，因此“VLA+规划”“LLM+TAMP”“策略模型+安全控制”等混合架构逐渐成为重要方向：高层语义推理生成任务目标或候选动作，中层规划验证任务约束和运动可行性，策略模型或低层控制器负责闭环执行，并依据反馈完成异常检测和失败恢复。

三、基于强化学习的Agent–机械臂融合方法

强化学习为机械臂系统提供了一种以任务回报为目标的自主决策机制。在连续控制和高维感知场景中，该类方法通过与环境交互持续更新策略，使机械臂能够在复杂动力学和不确定操作条件下形成一定的自适应行为。随着大规模交互数据积累、仿真平台扩展以及多任务学习框架的发展，强化学习在机械臂领域的研究重点也逐步由单任务控制转向规模化训练、技能组织与跨任务泛化。

需要说明的是，文中所称“强化学习型Agent”并非仅指狭义的在线强化学习算法，而是从机器人策略学习和交互优化的广义视角出发，指以状态–动作映射学习、任务回报优化、示教初始化和策略改进为核心的一类Agent–机械臂融合方法。在真实机械臂任务中，由于直接在线试错成本高、安全风险大、样本效率低，强化学习往往并不以完全从零探索的形式单独使用，而是与模仿学习、行为克隆、离线强化学习、生成式策略建模等方法结合，用于提供初始策略、压缩探索空间、提升训练效率和增强动作生成稳定性。类似地，在复杂博弈与决策任务中，模仿学习也常被用于为强化学习提供高质量初始策略或经验分布，从而降低纯强化学习的探索成本。因此，本研究将模仿学习、行为基础模型和生成式策略等与强化学习密切耦合的方法，纳入强化学习型Agent路线中讨论，但在具体分析时区分其监督信号、数据依赖和优化机制差异。

表1　Agent–机械臂融合主要技术路线比较

技术路线	主要决策层级	代表性方法	数据依赖	可解释性	物理可行性保障	跨平台适配能力	实时性与部署难点
强化学习型Agent	低层控制与技能学习	DDPG、TD3、SAC、PPO、分层强化学习、离线强化学习	依赖交互数据、仿真数据或示教数据	较弱，策略通常为隐式映射	主要依赖奖励设计、安全约束和仿真验证	中等，受动力学差异和动作接口影响明显	训练成本高，真实试错风险大，部署需关注安全探索与策略稳定性
规划与认知型Agent	中高层任务组织与约束推理	TAMP、神经引导规划、任务图、世界模型	依赖任务模型、符号约束、环境状态或预测模型	较强，任务结构和约束关系相对清晰	通过约束验证、运动规划和可行性检查实现	中等偏强，但依赖任务建模和接口定义	求解效率、实时重规划和复杂场景扩展是主要难点
大模型驱动具身Agent	高层语义理解、任务分解与多模态决策	VLA、RT系列、OpenVLA、SayCan、Code as Policies	依赖大规模多模态数据、机器人轨迹和指令数据	中等，高层推理可解释，低层动作生成较难诊断	通常需结合可行性校验、低层控制器或规划模块	较强，具备较好语义泛化潜力	推理延迟、动作精度、幻觉问题和系统集成复杂度突出
混合型架构	高层规划与低层控制协同	“VLA+规划”“LLM+TAMP”“策略模型+安全控制”	同时依赖语义数据、机器人数据与任务约束	中等偏强，取决于模块接口透明度	通过分层控制、反馈校验和失败恢复共同保障	较强，但系统适配成本较高	模块通信、接口标准化、错误传播和长期稳定性是关键问题

（一）单智能体强化学习控制方法

连续控制算法如 DDPG、TD3[38] 与 SAC[39] 构成了机械臂强化学习的基础框架。这些方法通过策略梯度优化高维连续动作空间，在抓取与推拉任务中实现了端到端控制。随着计算能力和感知能力的提升，强化学习的输入形式由低维状态逐步扩展到视觉观测，视觉条件下的端到端策略学习由此成为重要的研究方向。

近五年来，研究重点逐渐从“单任务策略训练”转向“多任务与大规模策略学习”。RoboNet 项目通过构建跨机器人平台的大规模交互数据集，实现了跨场景策略迁移[40]。BridgeData 及其扩展版本通过跨环境、多任务数据整合提升策略泛化能力。RT-1 与 RT-2 系列模型通过将大规模视觉–语言数据与机器人交互数据结合，实现跨任务泛化与语言驱动控制[41]。

这一阶段的显著特征是策略训练规模显著扩大，强化学习逐步与行为克隆、预训练模型结合，形成数据驱动型策略基础模型。然而，大规模数据训练虽然提升了泛化能力，也带来了算力需求与数据标注成本问题。

强化学习在复杂操作任务中的稳定性问题长期存在。除传统经验回放与目标网络机制外，近年来研究者开始引入对比表示学习与自监督预训练机制，以提升视觉特征表达质量。例如，PerAct 通过引入三维体素表示提高操作定位精度[42]，BC-Z 利用大规模示教数据实现零样本泛化[43]，VIMA 通过多模态预训练增强任务理解能力[44]。

此外，扩散模型被引入策略生成过程。Diffusion Policy 将扩散生成模型应用于连续动作预测，在复杂操作任务中表现出更强的稳定性与鲁棒性[45]。行为变换器通过序列建模方式提升策略时间一致性[46]。这类方法在策略表达能力与生成稳定性方面优于传统的 Actor-Critic 框架。由此可见，强化学习、模仿学习与生成式策略并非完全割裂的技术路线，而是在机械臂策略学习中围绕数据利用、探索效率和动作分布建模形成互补关系。强化学习侧重通过回报信号优化策略，模仿学习和行为克隆可为策略提供示教初始化和稳定动作先验，扩散策略、Behavior Transformer 和动作分块 Transformer 等生成式方法则进一步提升多模态动作分布建模和长时序动作生成能力。对于真实机械臂系统而言，这类组合式策略学习框架能够在一定程度上降低在线试错成本，并提升复杂操作任务中的训练效率和部署稳定性。

（二）分层强化学习 Agent 与多 Agent 协作

分层强化学习通过引入时间抽象机制解决长时序任务问题。Options Framework 与层级 Actor-Critic 结构为子策略调用提供理论基础。近年来，技能发现与自监督分层结构逐渐兴起。“多样性即所需的一切”方法[47]通过无监督方式生成多样化技能，强化学习在多任务环境中的适应能力得到提升。

在真实机器人操作中，技能库机制被用于跨任务迁移。“从自由探索中学习潜在规划”方法利用自由探索数据学习潜在行为表示[48]，在多步骤操作任务中表现出较强的泛化能力。RoboCat 等模型通过持续学习框架不断扩展技能库，实现跨任务积累[49]。

尽管技能化结构增强了策略复用能力，其层级划分与技能组合仍缺乏统一标准。技能粒度选择与调用机制设计直接影响泛化性能。

多智能体强化学习通过集中训练、分散执行机制实现多个执行单元协同优化[50]。在双臂协同抓取与复杂装配任务中，多智能体结构显著提升协作效率。在人机协作场景中，强化学习模型需考虑人类行为的不确定性与安全约束。部分研究将人类操作数据融入训练过程，通过共享表示空间实现协作适配。交互式强化学习框架通过在线反馈机制优化策略更新，提高系统对人类意图的响应能力。

然而，多智能体强化学习在复杂协作场景中仍面临信用分配与样本效率问题，若扩展至大规模多体系统仍具挑战。

（三）基于仿真到现实迁移的强化学习方法

真实机械臂平台的试错成本与安全风险限制了强化学习的直接应用。域随机化仍是主流迁移方法。Isaac Gym 等高性能仿真平台使大规模并行强化学习成为可能[51]。近年来，结合真实数据微调的混合训练方式逐渐普及，在保持策略稳定性的同时降低现实风险。离线强化学习通过利用历史交互数据进行策略优化，降低真实环境探索需求[52]。动作分块变换器（ACT）与扩散行为模型通过序列建模方式提升长期任务稳定性[53]。该类方法逐渐与语言

模型结合，为跨任务泛化与认知决策结构铺垫基础。尽管现实部署能力持续增强，强化学习策略在复杂接触动力学与精密装配任务中的可控性仍需进一步验证。

总体而言，强化学习型Agent已由连续控制优化拓展至技能组织、规模化训练和语言条件多任务学习，大规模数据与生成模型提高了策略泛化和任务理解能力。其优势在于连续动作优化和复杂交互中的自适应性，适合作为低层控制、技能学习和局部闭环执行模块；但其训练依赖交互数据与仿真环境，真实试错成本和安全风险较高，且缺乏显式任务结构，难以独立处理长时序任务中的逻辑依赖、约束推理和失败恢复。因此，复杂任务仍需与规划和认知模块协同。

四、基于规划与认知的Agent–机械臂融合方法

强化学习方法在机械臂控制中主要解决的是动作生成最优化问题，而规划型Agent关注的是任务结构组织问题。这两者的差异并非算法层面，而是决策层级的差异。强化学习通过参数化函数逼近策略，但其内部并不显式编码任务逻辑依赖关系；规划型方法则通过显式结构表达阶段性目标、约束耦合与因果关系。近五年的研究进展实际上围绕一个核心矛盾展开：如何在保持任务结构可解释性的同时，引入数据驱动机制以提升泛化能力。

（一）神经引导任务规划：从逻辑搜索到结构学习

传统任务规划依赖符号搜索，其优势在于逻辑完备性，但在复杂对象组合场景下呈现指数级扩展。近年的神经引导规划并非简单的替代搜索，而是试图通过学习机制重塑搜索空间结构。

结构化规划学习通过从数据中学习符号操作符，减少了对人工建模的依赖，并为任务与运动规划提供搜索引导[54]。这一变化的核心意义在于：规划复杂度不再仅由逻辑分支数量决定，而与模型表达能力相关。Neural Task Graph将任务依赖关系嵌入图结构中，使系统能够学习子任务之间的层级与顺序约束[55]。该方法实质上将离散任务图转化为可微结构表示，从而为跨任务迁移提供条件。

语言模型参与规划则进一步模糊了符号规划与序列生成之间的界限，语言模型能够在零样本条件下生成合理的任务序列[56]。然而，该类方法的局限也十分明显：生成结果缺乏物理一致性保障，必须依赖下游验证机制。神经引导规划提升了任务抽象能力，但并未消除物理可行性验证的必要性。

（二）任务与运动协同规划方法

任务与运动协同规划的核心难点在于离散决策变量和连续运动约束之间的耦合。早期方法采用分阶段验证，逻辑规划与运动规划交替执行，导致计算开销较高。相关综述指出，统一建模是解决该问题的根本路径[57]。

可微分TAMP框架将逻辑变量编码为连续优化变量，通过约束松弛机制实现联合求解[58]。这一方法在理论上减少了规划成功但运动不可行的情况，其实质是将任务图嵌入优化问题，使决策空间连续化。与采样式方法相比，该框架在高维操作场景中表现更稳定，但对初始条件与梯度信息依赖较强。

增量式TAMP通过在规划过程中逐步更新约束，实现任务规划与运动规划的协同求解。Dantam等提出的IDTMP采用增量约束求解，在任务层添加和删除运动可行性约束[59]，以提高联合规划的求解效率。

近年来，研究者开始将TAMP的符号规划能力与强化学习的技能学习能力结合，以支持长时程操作中的技能获取与复用[60]。这种融合方式并非替代规划，而是利用规划提供任务结构并引导策略学习。其理论意义在于：TAMP不再是纯符号系统，而成为混合型决策架构。当前TAMP的瓶颈主要体现在实时性与规模扩展能力。当任务空间与对象种类显著增加时，联合优化问题规模迅速扩大，求解效率成为限制因素。

（三）内部模型与预测机制在认知型Agent中的作用

面对环境不确定性、复杂接触动力学及感知偏差，传统TAMP仅依赖几何与约束验证，难以保证执行一致性，因而研究开始在规划框架中引入内部预测模型。早期世界模型通过潜在动力学预测提升强化学习样本效率[61]；面向机械臂操作，高斯世界模型（GWM）以动作条件三维场景表示预测抓取与接触过程[62]，交互式真实机器人动作仿真器（IRASim）进一步建模细粒度交互变化[63]。

内部模型通过比较预测结果与实际观测，触发局部重规划或策略修正，形成“预测–执行–修正”闭环[64]。NVIDIA Cosmos 等生成式世界模型平台[65]则主要用于数据扩展与仿真评测。因此，其定位是连接规划与执行、增强预测和恢复能力，而非替代高层任务结构。

总体而言，规划与认知型 Agent 以显式任务分解、约束建模和可行性验证支撑长时序、多阶段及强约束操作，并借助世界模型增强预测与失败恢复能力。其主要瓶颈在于先验建模成本，以及复杂动态环境中的实时求解和规模扩展，因此更适合作为中高层任务组织与约束校验模块，与低层策略学习和大模型语义推理形成互补。

五、大模型驱动的 Agent–机械臂融合方法

大模型驱动的 Agent–机械臂融合方法以大规模预训练知识、多模态表示和语言推理为基础，重点提升开放指令理解、任务泛化与跨场景适应能力。按系统功能可分为 VLA 统一建模、跨平台行为预训练与策略基础模型、语言条件高层任务建模、生成式世界模型及训练 / 评测基础设施、公开基准与评测指标五类。实际系统通常将其组合为“高层语义推理–规划验证–低层执行–反馈修正–评测诊断”的闭环。

（一）VLA 统一建模与端到端控制

VLA 路线的关键贡献在于把机械臂控制统一到大模型擅长的序列建模范式中，使“语义条件”成为策略的内生变量。RT-1 通过 Transformer 在真实多任务机器人数据上训练统一策略，奠定了规模化多任务控制的工程范式。RT-2 进一步将互联网规模视觉–语言预训练知识迁移到机器人控制，体现出语义泛化与组合能力的增益。与此相对，PaLM-E 将多模态感知嵌入语言模型结构中，强化具身推理与跨任务迁移的表征共享。VIMA 以多模态提示为接口组织任务条件，强调“条件组合”与任务可扩展性。

在开源与可复现层面，OpenVLA 将 VLA 训练与参数高效微调系统化，降低跨平台适配门槛，并推动通用 VLA 从封闭体系向可迭代生态迁移[66]。另一条典型路线以“基础 VLM+轻量控制头”为核心：RoboFlamingo 直接建立在 OpenFlamingo 等视觉语言基础模型之上，通过模仿学习微调实现语言条件操作，体现出较强的数据效率与部署灵活性[67]。RoboMamba 则把状态空间序列模型（Mamba）引入 VLA，强调推理效率与轻量微调，使“推理–动作”链路更适配实时系统约束[68]。

2024—2025 年，VLA 领域进一步向通用机器人基础策略、开放世界泛化和高效部署方向发展。π_0 提出面向通用机器人控制的 VLA 流模型，将预训练视觉语言模型与流匹配动作生成机制结合，在继承互联网规模语义知识的同时输出适用于复杂灵巧操作的连续动作，并覆盖单臂、双臂和移动机械臂等多种机器人平台[69]。在此基础上，$\pi_{0.5}$ 进一步面向开放世界泛化进行扩展，通过异构任务共训练以及多机器人数据、高层语义预测、网络数据等多来源信息联合建模，提升了模型在新环境和真实长时序操作任务中的泛化能力[70]。与此同时，GR00T N1 面向通用人形机器人提出双系统 VLA 架构，由视觉语言模块承担环境理解与指令解析，由扩散 Transformer 模块生成实时运动动作，体现出 VLA 模型向多形态机器人和人形机器人泛化的趋势[71]。此外，SmolVLA 则从模型规模和部署成本角度出发，强调在较低训练与推理成本下实现具备竞争力的语言条件机器人控制能力，并通过异步推理机制提升动作执行响应性[72]。上述工作表明，VLA 研究正在由早期语义条件操控逐步迈向通用策略建模、开放场景泛化、多具身形态适配和轻量化工程部署并重的发展阶段。

VLA 端到端体系的结构性瓶颈集中在两个方面：动作词元化或离散化带来的精细控制分辨率与物理一致性问题，以及端到端生成链路中错误定位困难导致的可控性不足。面向装配、力控与复杂接触任务，VLA 更稳定的落地形态往往是高层语义到子目标 / 关键动作的生成，再由低层控制或优化模块保证可执行性，而不是完全端到端闭环。

（二）跨平台行为预训练与策略基础模型

与 VLA 更强调语义到动作的统一映射不同，行为基础模型更强调动作先验的规模化学习：通过多平台、多任务数据训练通用策略底座，使新任务 / 新平台以少量数据完成适配。Open X-Embodiment

（RT-X）通过统一多机构多机器人数据格式，提供跨具身形态共享学习的基础条件，并展示跨平台正迁移现象。Octo平台进一步把跨传感器、跨动作空间、跨平台的适配作为核心设计目标，形成通用策略初始化与快速微调路径。BridgeData V2则从数据侧推动可规模化现实操控学习的基准化，提供跨环境、多轨迹的数据支撑。

在策略模型形态上，生成式策略成为近两年的显著趋势。Diffusion Policy将扩散生成用于动作分布建模，缓解多模态动作分布下回归策略不稳定的问题，并在多种操控基准上验证性能优势。ACT针对精细双臂／接触任务的长时序误差累积问题，提出动作分块预测与序列生成式策略，显著提升少量演示条件下的成功率，体现“时间结构建模”对稳定操控的重要性。围绕ACT的后续工作已开始面向更复杂的双臂耦合与动作依赖建模扩展，如InterACT强调双臂互依赖建模以提升协同稳定性[73]，One ACT Play探索单演示条件下的动作分块克隆以进一步降低数据门槛[74]。

这一路线的核心矛盾在于：跨平台共享表征并不等价于跨动力学共享控制。预训练底座更像“可迁移初始化”，仍需面向具体硬件动力学与传感器差异进行适配。工程上常见的稳健组合是：通用策略底座提供先验与初始化，结合在线校正、约束控制或规划验证以保证分布外状态下的安全性与可恢复性。

（三）语言条件高层任务建模与闭环具身推理

语言模型在机械臂系统中的结构价值主要体现在高层：以语言为统一语义空间表达目标、约束与任务分解，并与执行反馈形成闭环。“Language models as zero-shot planners”一文证明语言模型能够从语义知识中抽取可执行的高层行动序列，为“语义到计划”的映射提供早期范式。SayCan进一步把语言模型生成与可行性评估耦合：语言提出候选行为，价值／可达性模型筛选可执行选项，从而将“语义合理”收敛到“物理可行”。Inner Monologue强调将环境反馈持续写入提示上下文，以形成可持续修正的推理链条，体现“反馈驱动推理”的闭环价值[75]。Code as Policies则将高层意图落到可执行程序结构，通过应用程序编程接口组合形成可复用、可审计的策略表达，提升可控性与工程可落地性。

上述路线的共同点是把语言模型从直接输出动作转向输出结构：输出计划、子目标、可调用技能组合或可执行程序，并依赖执行反馈进行修正。其瓶颈也相对清晰：一方面，高层推理与低层执行之间的接口需要可验证机制，否则容易出现语义正确但物理不可行的情况；另一方面，闭环推理增加系统复杂度，错误传播路径更长，需要明确失败检测、回退策略与安全约束。因而，高层语言推理更适合与第四章中的规划、TAMP、约束控制形成互补，而不是孤立替代。

（四）生成式世界模型与规模化训练／评测基础设施

为突破真实数据规模瓶颈，生成式世界模型与大规模仿真成为支撑性基础设施方向。DreamerV3代表“世界模型+想象rollout”的控制范式，通过潜在动力学学习提升跨任务泛化与稳定性；其后续的工作进一步强化了世界模型在多样控制任务上的证据链。面向机械臂操控，GWM将三维高斯表示引入动作条件预测，更贴近具身操作的空间一致性需求；IRASim聚焦细粒度交互建模，使接触动态与交互过程更可预测；ReViWo则针对视角变化提出视角不变世界模型学习，以提升视觉扰动下的操控鲁棒性。

在工程平台层面，RoboCasa以大规模仿真环境扩展任务分布与场景覆盖，为通用策略训练与评测提供可扩展载体[76]；Mobile ALOHA通过低成本全身遥操作实现高难度双臂移动操作数据采集，直接提升真实世界长时序任务的数据可得性。面向更广义的具身平台，GR00T N1提出双系统架构，将高层视觉语言推理与低层动作生成解耦，体现“推理–反射式控制”分层的工程化路径。

这一方向的核心评价维度已经从“生成质量”转向“控制收益与物理一致性”：世界模型能否在长时序、接触动力学与多模态反馈下保持稳定预测，能否有效扩展长尾分布并提升真实执行成功率，决定其在机械臂具身智能中的长期价值。

（五）大模型驱动机械臂智能体的评测指标与公开基准

需要说明的是，本节所讨论的评测指标体系与

公开基准并非将评测方法视为大模型驱动算法本身，而是将其作为大模型驱动Agent–机械臂系统的重要评价支撑环节。与传统机械臂控制方法主要关注轨迹误差、局部控制精度或单任务成功率不同，大模型驱动具身智能体通常同时包含语义理解、任务规划、可供性判断、动作生成、执行反馈和失败恢复等多个模块，其能力边界难以通过单一结果指标全面刻画。因此，公开基准和评测体系在该方向中承担着能力诊断、过程分析、安全验证和系统迭代的作用，是理解大模型驱动机械臂智能体发展水平的重要组成部分。

大模型驱动的机械臂系统在研究范式上呈现“模型先行、评测追赶”的特点。早期操控评测多以二值成功率为主，但在通用策略与长时序任务背景下，该指标容易掩盖关键缺陷。例如，策略能够完成任务，却存在不稳定抓取、短暂滑移、双臂时序不一致、对齐误差累积等行为问题，导致跨场景迁移时性能骤降。近两年基准体系的显著变化，是将评测从“结果正确”转向“过程可诊断”，并将能力维度拆分为可复现的阶段指标。

面向通用操控技能，ManiSkill2通过多任务族、多对象几何变化与统一评测协议，强调在对象级拓扑与几何变化下的泛化能力，提供从经典sense-plan-act到模仿学习／强化学习的统一接口与对比机制，为“可组合操控技能”的评测提供框架基础[77]。在语言条件长时序任务方面，CALVIN以“长序列语言指令链”定义评测场景，将零样本语言泛化与长时序组合执行作为核心难点，并提供统一评估协议，使语言条件策略的能力比较具备可复现性。在持续学习与知识迁移维度，LIBERO将长期学习与任务序列顺序敏感性纳入基准设计，形成围绕迁移、鲁棒性与预训练效益的评价维度，适配“通用策略底座+增量适配”的研究范式。这些基准共同推动评测从单任务精度比较，转向面向泛化、组合与持续学习的系统能力比较。

随着大模型在机器人“认知层”承担更高比重，评测进一步出现“认知–执行解耦”的趋势。一类工作以“具身大脑”的多模态大模型为评测对象，重点覆盖指令理解、感知推理、规划与可供性推断等能力，并尝试用更贴近物理可行性的评价方式替代纯文本相似度。RoboBench明确将多模态大模型作为具身认知核心，给出覆盖理解、推理、规划、可供性与失败分析的维度设计，并提出以“世界模拟rollout”方式评估规划的具身可行性，试图解决“规划文本合理但不可执行”的评测失真问题[78]。另一类工作把“安全可靠”纳入评测目标，Responsible Robot Bench针对风险场景定义安全感知、风险规避与人机协作求助等指标，强调“任务完成”与“安全代价”共同度量，从而使具身智能评测进入可靠性维度[79]。此外，ManipBench以低层操控相关推理为目标，采用题库式评测多模态模型在操控知识与过程推理上的能力，为“认知层评测”提供低层操控导向的补充[80]。

总体而言，上述分支分别作用于任务推理、策略生成、环境预测与能力诊断，彼此互补。面向真实部署，仍需与规划器、约束检查和低层控制器协同，以兼顾语义泛化、物理可行性、执行稳定性与评测可复现性。

六、Agent–机械臂融合研究中的公开数据集

公开数据集与评测基准共同构成Agent–机械臂融合研究的资源基础。前者主要提供训练、验证和迁移所需的数据分布，后者则用于定义任务协议、评价指标和可复现比较条件。由于操控任务普遍存在接触不确定性、对象长尾分布与长时序依赖，仅报告成功率难以解释方法差异，数据集的采集方式、模态对齐、动作接口与泛化划分往往直接决定结论边界。因此，本章聚焦代表性公开数据集的“数据形态–适用任务–使用要点与局限”，并给出面向不同研究目标的选型建议。

（一）数据集的作用与分类维度

如图4所示，本研究从任务时序长度和跨平台／迁移泛化程度两个维度，对代表性公开机器人操作数据集进行层级化归纳，可大致分为真实采集／示教／遥操作数据、跨平台整合数据底座、长时序任务与离线强化学习基准，以及终身／迁移学习数据集四类。不同类型数据集在任务复杂度、模态构成、评测重点和适用方法上存在差异，因此需要结合具体研究目标进行选择。具体而言，真实采集／示教／遥操作数据更强调操作轨迹质量、控制稳定

性和真实交互过程记录；跨平台整合数据底座侧重多机器人平台、多任务和多数据源的统一组织，有助于开展跨平台迁移和通用策略学习，其中 Open X-Embodiment 更偏向数据整合基础，RT-X 则更偏向基于该类数据训练得到的跨机器人模型体系；长时序任务与离线强化学习基准更突出任务组织、子技能组合和误差累积问题；终身 / 迁移学习数据集则主要服务于持续学习、知识迁移和分布变化下的策略评估。为增强公开数据集的选取指导价值，本研究在上述层级化归纳的基础上，进一步从数据来源、机器人平台、主要模态、主要适用方向和局限性等维度，对代表性机械臂及机器人操作数据集进行比较，如表 2 所示。

（二）面向低层控制与技能学习的示教数据

BridgeData V2 以真实示教轨迹为核心，面向低层技能学习提供相对稳定的现实分布。其价值在于：一方面，数据覆盖多环境与多任务形态，适合在统一数据分布下比较行为克隆、离线强化学习与条件化模仿等方法；另一方面，轨迹的采集与整理更强调工程可复现性，使研究者能够在控制频率、动作接口与观测模态明确的条件下复现实验设置。

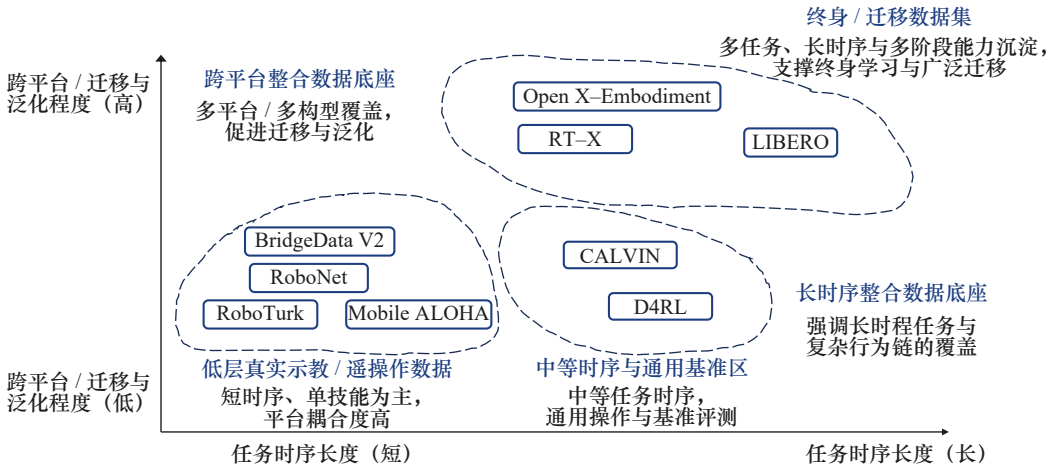


图 4　机械臂数据集层级划分

注：LIBERO 为具身智能评测基准。

表 2　典型机械臂数据集

数据集	数据来源	机器人平台	主要模态	主要适用方向	局限性
BridgeData V2	真实示教轨迹	Widow X 250 机械臂	视觉+动作	行为克隆、离线 RL、技能学习	关键失败状态覆盖有限
RoboNet	多机器人真实轨迹	多机器人平台	视觉+动作	跨平台迁移、分布差异分析	数据异构性强、对齐成本高
RoboTurk	众包遥操作示教	机械臂平台	视觉+动作	低成本示教获取、鲁棒学习	轨迹质量与一致性波动较大
Mobile ALOHA	真实遥操作数据	移动双臂平台	视觉+动作	双臂协同、长时序任务、动作结构建模	任务链长、建模难度较高
Open X-Embodiment / RT-X	多来源整合数据	多机器人平台	多模态	通用策略预训练、跨平台迁移	协议统一与重参数化要求高
CALVIN	仿真 / 标准化任务数据	标准机械臂平台	视觉+动作+语言	语言条件长时序任务评测	与真实开放环境存在差距
D4RL	离线交互数据	标准机械臂任务	状态 / 动作为主	离线强化学习、组合任务评测	真实感和开放性有限
LIBERO	标准化任务序列	标准机械臂平台	视觉+动作+任务描述	终身学习、迁移与遗忘评测	仍需结合真实部署验证

在使用上，BridgeData V2 更适合回答“在相同数据预算下，策略结构 / 训练配方是否带来稳定收益”的问题；其局限在于关键失败状态与极端接触情形仍相对稀缺，导致模型在少数转折状态（如滑移、遮挡、夹持不稳）上的鲁棒性需要结合额外评测或补充数据验证。

RoboNet 以多机器人、多场景的真实轨迹集合为主要特点，强调“跨平台分布差异”这一通用操控研究中的关键变量。与单平台示教数据相比，RoboNet 更适合研究跨机器人迁移、域间差异与表示对齐问题，如不同机械臂的动作空间定义、控制频率与相机视角差异如何影响策略泛化。需要注意的是，多来源数据的异构性会显著增加训练与评测的对齐成本，研究中应显式报告动作重参数化、时间同步与训练预算，否则容易出现由数据混杂导致的不可比结论。

RoboTurk 的核心贡献在于以众包遥操作降低示教获取成本，从而以更低门槛扩展示教数据规模与任务多样性。该类数据更适合用于分析“数据质量 – 成本 – 规模”的权衡关系，并验证在噪声示教与策略偏置存在时，行为学习方法的稳定性边界。其局限也相对明确：众包示教的动作一致性与轨迹质量更难控制，容易引入策略偏置与轨迹噪声，因此更适合用于鲁棒学习与数据筛选 / 重加权等研究问题，而不是作为高精度装配类任务的唯一训练来源[81]。

Mobile ALOHA 面向移动双臂长时序操作，突出“复杂任务示教可获得性”。其数据采集方式通常包含移动底盘与双臂协同动作，覆盖多阶段操作链路，使其更接近真实服务与物流场景中的任务组织形式。在方法验证上，该数据集特别适合评测动作结构建模（如动作块、双臂协同约束）与长时序误差累积问题，并可用于检验模型在动态视角变化与操作链路延长条件下的稳定性。其主要挑战在于任务链更长、状态更复杂，对模型的时序建模与失败恢复能力提出更高要求。

从选型角度看，若研究目标是低层控制、行为克隆或离线策略学习的稳定性对比，可优先选择 BridgeData V2 等轨迹质量较高、接口相对清晰的真实示教数据，并在实验中明确动作频率、控制模式和训练数据规模；若研究重点是跨机器人泛化、平台异构性或数据分布差异，可结合 RoboNet 和 RoboTurk 分析多来源数据带来的多样性收益与对齐成本；若关注双臂协同、移动操作和长时序示教学习，则 Mobile ALOHA 更适合用于评估动作结构建模、双臂协调和任务链稳定性。总体而言，低层技能示教数据的选取应重点关注示教质量、动作接口、关键失败状态覆盖和数据噪声水平，避免仅依据数据规模判断其适用性。

（三）跨平台整合数据底座

Open X-Embodiment / RT-X 将多来源真实机器人数据统一到一致格式，为“共享底座+快速适配”的通用策略训练范式提供数据基础。与单一数据集不同，其核心价值不在于某一类任务的覆盖，而在于提供跨机器人、跨任务的统一数据协议，使研究者能够在更一致的接口定义下比较预训练初始化、微调预算与迁移收益。若研究目标是通用策略底座、跨平台预训练或少量数据微调，Open X-Embodiment / RT-X 适合作为基础数据资源。但在使用该类跨平台整合数据时，应重点核查不同机器人平台之间的动作空间定义、传感器配置、控制频率和数据协议是否经过统一处理，并在实验中明确预训练预算、微调预算、seen / unseen 对象与场景划分。否则，模型性能差异可能更多来源于平台和协议不一致，而非方法本身的泛化能力。

（四）长时序任务数据集

CALVIN 以语言条件长时序任务链为核心，使研究能够在统一设置下评测“任务组合能力”，而不仅是单步技能是否成功。其典型评测关注多步骤指令链的成功分布，并强调闭环执行过程中的状态更新与错误累积效应。因此，CALVIN 更适合验证高层任务组织、技能选择与执行闭环相关方法；在报告结果时，除平均成功率外，通常还需要给出组合长度分布与失败模式，以避免短链路成功掩盖长链路失效。

D4RL 中的 Franka Kitchen 任务常用于离线强化学习与长时序组合能力评测，其任务由多个子目标构成，强调在无在线交互或低交互预算条件下，策略能否在多阶段链路中保持一致执行。该数据集对研究“离线数据分布偏移”“误差累积”与“策略保守性”具有代表性：当策略偏离数据分布时，错误往往以链式方式放大。因而它更适合用于检验

离线学习算法在长时序任务中的稳定性边界，而不是直接代表真实世界的开放环境分布[82]。

CALVIN更适合用于构建统一比较；若研究重点是离线强化学习、分布偏移和多阶段任务中的误差累积，D4RL中的Franka Kitchen等任务可作为典型测试场景。对于长时序数据集，除报告最终任务成功率外，还应关注组合链长度、子任务完成情况、失败模式和中间状态反馈，以避免短链路任务表现掩盖长链路执行不稳定的问题。

（五）终身 / 迁移数据集

LIBERO以任务序列为核心，突出持续适配与知识迁移的评测需求。相较单任务基准，它重点考察模型在扩展任务库过程中的知识迁移、灾难性遗忘和任务顺序敏感性，适合评估通用策略底座在持续微调过程中的稳定性。使用时应明确任务序列、训练顺序、数据预算与评测划分，并同时报告迁移收益、遗忘代价和不同任务族的表现，避免以单任务平均成功率替代持续学习评价。

七、智能体与机械臂融合的技术挑战及发展展望

（一）数据效率与样本获取

数据规模的增长为通用策略底座与VLA体系提供了训练基础，但真实操控数据中的关键矛盾并未随之消失。首先，滑移、碰撞、遮挡以及失败恢复等关键转折状态在数据中天然稀缺；其次，示教轨迹的质量与多样性容易受到采集系统和人类操作偏置的影响；再次，跨平台数据在动作空间、控制频率和传感器同步上的差异，也使大规模整合过程伴随分布混杂风险。由此带来的问题是，平均成功率的提升可能更多来自常规状态覆盖的增加，而长尾失效仍集中出现在少数不可恢复状态中，从而造成表面泛化能力与真实鲁棒性之间的偏差。

从方法发展来看，近年来生成式策略、动作分块与行为底座等结构改进，在一定程度上缓解了有限数据条件下的多解覆盖不足和长时序误差累积问题。但这类方法更多提升的是对已有数据分布的利用效率，并不能替代关键状态覆盖本身。因此，后续研究需要更加重视数据覆盖范围与失败类型分布的描述，在实验报告中明确训练数据规模、示教来源、任务分布以及关键状态覆盖策略，以更准确地判断性能提升究竟来自方法结构改进，还是来自数据规模扩展。

从工程实践角度看，数据效率问题的短期目标是建立更可复用的数据采集与筛选流程，包括统一示教数据格式、明确关键失败状态标注规则、补充滑移、遮挡、碰撞和恢复过程等长尾样本，并通过数据重加权、轨迹质量筛选和少量真实数据微调提升有限数据条件下的策略稳定性。长期挑战则在于构建覆盖多平台、多任务、多场景和多模态交互过程的持续增长数据体系，使数据资源不仅服务于单次模型训练，还能够支撑通用操作策略的持续更新、跨域迁移和安全边界评估。

（二）仿真到现实迁移与跨平台泛化

通用策略与VLA体系在语义理解和任务迁移层面带来了明显增益，但机械臂操控的泛化能力仍取决于“语义–物理”闭环是否一致。在高精度接触与装配任务中，动作表示离散化、时序同步误差以及隐式约束缺失，容易导致“语义合理但物理不稳”的失效现象。这类问题通常不会表现为明显的错误动作，而更多以微小偏差持续累积的方式出现，并最终导致任务失败，因此也更难通过静态评测或单次结果加以识别。

仿真环境与合成数据体系在规模化训练中发挥了重要作用，但有效迁移的实现仍依赖系统级闭环设计，而非单一技术手段。即便仿真环境具有较高逼真度，真实场景中的传感噪声、执行延迟以及未建模扰动仍会在长时序任务中不断累积，并进一步影响系统稳定性。与此同时，跨平台部署还会受到机械臂结构、控制接口和传感配置差异的影响，使得同一策略在不同平台上的迁移效果存在较大不确定性。因此，后续研究应将迁移评估由一次性成功扩展到过程稳定性、失败恢复能力和跨平台适应性，并进一步明确训练、微调与测试之间的完整协议链路，以增强相关结论的可比性与可复现性。

短期内，仿真到现实迁移研究应优先建立更加透明的迁移协议，包括明确仿真参数随机化范围、真实数据微调比例、传感器噪声建模方式、动作接口重参数化方法以及seen / unseen对象和场景划分，以提高不同研究之间的可比性。长期来看，跨平台泛化的核心挑战在于实现语义层、动作层和动

力学层之间的统一适配，使同一 Agent 能够在不同机械臂结构、末端执行器、传感器配置和控制频率条件下保持稳定执行，而不依赖大量平台专用的调参方法。

（三）长时序任务组织与失败恢复

真实环境中的机械臂任务通常具有明显的长时序特征，需要经过目标识别、动作规划、执行操作和状态确认等多个环节。与单步任务相比，这类任务步骤间依赖更强，前序环节中的微小偏差也更容易在后续过程中持续累积，最终导致整体任务失败。现有研究已在任务分解、层级决策和技能调用等方面取得一定进展，通过高层决策与低层执行的分层组织方式，系统在一定程度上提升了复杂任务的执行能力。然而，这类方法通常依赖较清晰的任务边界和较准确的中间状态感知。当任务链条较长、环境扰动较大或状态判断出现偏差时，原有执行链路仍容易发生级联失效。

此外，长时序任务中的失败往往并非由单次明显错误直接引起，而更多表现为抓取偏差、状态漂移、动作时序不一致等问题的逐步累积。若系统缺乏中间状态校验、在线重规划和局部回退机制，就容易出现局部失误放大并最终导致任务中断的情况。因此，后续研究需要更加关注长时序任务中的任务组织与反馈修正机制。一方面，应增强高层计划与低层执行之间的协同，减少任务分解与实际执行之间的不一致；另一方面，在评测中增加对过程稳定性和失败恢复能力的考察，推动相关研究由“能否完成任务”向“能否稳定完成任务”转变。

短期目标是将长时序任务评测从最终成功率扩展到过程性指标，如子任务完成率、中间状态校验结果、失败发生位置、恢复尝试次数和局部重规划成功率，并在系统设计中显式加入状态检查、异常检测和回退策略。长期挑战则在于构建具备持续记忆、因果推理和自我恢复能力的机械臂智能体，使系统能够在开放环境中识别偏差来源、动态调整任务计划，并在局部失败后恢复到可继续执行的状态，而不是简单终止任务。

（四）评测体系与公开基准

当前，Agent 与机械臂融合研究的评估结果，仍较大程度依赖具体数据集、任务设定和实验流程。不同研究在对象类别、场景复杂度、动作接口、观测模态以及训练测试划分上的差异，使方法之间的横向比较缺乏统一基础。现有评测通常仍以任务成功率作为主要指标，但该指标更适合反映最终结果，难以揭示长时序执行中的误差传播、失败类型和恢复能力差异。对于需要多阶段决策与持续反馈的任务，仅以一次性成功或失败作为判断依据，容易掩盖系统在中间环节中的真实表现。

此外，公开基准虽然为方法验证和结果复现提供了重要支撑，但不同平台在机械臂结构、控制频率、传感器配置和任务接口上的差异，限制了基准结论的可迁移性。若缺乏更清晰的协议说明和更统一的测试设置，同一方法在不同平台上的表现差异便难以准确解释。因此，后续研究需要在公开基准建设中进一步强化任务定义、数据划分和评测协议的一致性，并增加对过程稳定性、失败恢复和跨场景迁移能力的考察。相比单纯增加任务数量，更重要的是形成能够支持公平比较、错误诊断和复现实验的评测体系。

短期内，评测体系建设应从单一任务成功率扩展到多维指标组合，至少包括过程稳定性、执行效率、失败类型、恢复能力、跨场景泛化、跨平台迁移和安全代价等维度，并要求公开训练 / 测试划分、数据预算和任务协议。长期挑战在于构建能够覆盖真实复杂作业环境的开放评测体系，使评测不仅能够比较不同模型的平均性能，还能够诊断系统在长尾任务、动态交互、人机共域和安全约束条件下的能力边界。

（五）安全性、稳定性与可解释性

随着系统能力提升，当前机械臂智能体的可靠性问题更多表现为偶发失效和诊断困难。单一成功率已难以支撑方法之间的可信比较，因为相同的失败结果可能对应不同的失效来源，如感知漂移、抓取滑移、双臂协调不同步、动作时序不一致或高层计划粒度不匹配等。这些失败虽然在结果上相似，但其产生原因和改进路径并不相同。

在系统结构层面，分层组织与可验证接口仍是提升安全性和稳定性的常见思路。通常由高层模块负责候选生成与任务组织，低层闭环负责执行与约束，中间层提供状态检查、异常反馈和回退机制。若系统缺乏必要的可检验接口，即使在平均性能上

表现较好，也难以保证真实环境中的长期稳定运行。因此，后续研究不仅要关注任务是否成功，还应结合失败类型、触发条件和恢复过程，对系统行为进行更细致的分析，从而为安全性评估和结构优化提供依据。此外，为增强系统运行过程中的安全性与诊断能力，实际部署中还需要引入异常检测机制，对感知漂移、抓取失稳、执行超时和模块输出异常等情况进行识别，并触发回退或重规划流程。

短期目标是建立面向机械臂 Agent 的安全约束与异常检测机制，包括对感知漂移、抓取失稳、执行超时、碰撞风险和模块输出异常等情况进行识别，并通过安全控制、动作约束、人工接管接口或局部回退策略降低失效风险。长期挑战则在于形成可解释、可验证和可审计的机械臂 Agent 安全框架，使系统不仅能够完成任务，还能够说明决策依据、暴露不确定性来源，并在高风险场景中提供可追溯的行为记录和责任分解。

（六）工程实现与系统部署挑战

大模型体系在训练阶段依赖大规模算力与数据支撑，而在实际部署中还必须同时满足实时闭环、资源预算和系统可维护性等要求。由于端到端大模型推理链路通常存在较高延迟和不确定性，因此很难直接承担高频控制闭环。当前工程系统中更常见的做法是采用分层解耦结构，即由大模型负责语义解析和子目标生成，低层由控制器、约束优化模块或策略底座承担闭环执行。这种结构不仅是工程实现上的现实选择，也反映了系统对稳定性和可靠性的实际需求。

在系统集成层面，接口一致性仍是影响部署稳定性的关键因素。若状态表示、动作接口、时间同步机制、失败信号定义以及回退触发条件缺乏统一规范，即使各模块在单独测试中表现良好，系统整体也可能在联调和长期运行中出现不稳定现象。随着跨平台通用策略和跨数据底座逐渐增多，后续研究需要进一步推动接口标准化、实验流程规范化和部署方案可复现化，以提高不同系统之间的集成效率和结果可比性。

短期内，工程部署应优先推动模块接口标准化和系统集成流程规范化，包括统一状态表示、动作接口、时间同步机制、失败信号定义、日志记录和回退触发条件，并通过分层架构降低大模型推理延迟对低层控制闭环的影响。长期挑战在于构建可长期运行、可持续更新和可跨平台迁移的机械臂智能体系统，使其能够在真实生产、物流、服务和特种作业环境中稳定运行，并在任务变化、硬件更换和环境扰动条件下保持可维护性与可扩展性。

此外，面向多机器人和群体作业场景，Agent–机械臂融合还需要进一步解决多主体任务分配、通信受限条件下的协同决策、局部操作与全局目标的一致性维护，以及多机械臂 / 移动平台之间的安全避碰和资源调度问题。相比单机械臂系统，群体作业对系统可扩展性、通信鲁棒性和协同安全机制提出了更高要求，是后续从单体智能走向群体智能的重要研究方向。

利益冲突声明

本文作者在此声明不存在任何利益冲突或财务冲突。

Received date: March 8, 2026; **Revised date:** June 14, 2026

Corresponding author: Song Lujie is a research fellow from Inspur Electronic Information Industry Co., Ltd. His major research fields include image interpretation and large language model (LLM) applications. E-mail: 2569332006@qq.com

Funding project: The National Natural Science Foundation of China Project (62401062); Fundamental Research Funds for the Central Universities (FRF-TP-25-038)

参考文献

[1] Li D Z，Jin Y X，Sun Y H，et al. What foundation models can bring for robot learning in manipulation：A survey[J]. The International Journal of Robotics Research，2026，45（7）：1091-1142.

[2] Liu Y，Chen W X，Bai Y J，et al. Aligning cyber space with physical world：A comprehensive survey on embodied AI[J]. IEEE/ASME Transactions on Mechatronics，2025，30（6）：7253-7274.

[3] Zitkovich B，Yu T，Xu S，et al. RT-2：Vision-Language-Action models transfer web knowledge to robotic control[C]//The 7th Conference on Robot Learning，2023：2165-2183.

[4] Driess D，Xia F，Sajjadi M S M，et al. PaLM-E：An embodied multimodal language model[C]//The 40th International Conference on Machine Learning，2023，202：8469-8488.

[5] Collaboration O X，Padalkar A，Pooley A，et al. Open X-embodiment：Robotic learning datasets and RT-X models[C]//IEEE International Conference on Robotics and Automation（ICRA），2024：6892-6903.

[6] Ghosh D，Walke H R，Pertsch K，et al. Octo：An open-source generalist robot policy[C]//Robotics：Science and Systems XX，2024：90.

[7] Walke H R，Black K，Zhao T Z，et al. BridgeData V2：A dataset for robot learning at scale[C]//The 7th Conference on Robot Learning，2023：1723-1736.

[8] Fu Z P，Zhao T Z，Finn C. Mobile ALOHA：Learning bimanual mo-

bile manipulation using low-cost whole-body teleoperation[C]//The 8th Conference on Robot Learning，2025：4066-4083.

[9] Mees O，Hermann L，Rosete-Beas E，et al. CALVIN：A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks[J]. IEEE Robotics and Automation Letters，2022，7(3)：7327-7334.

[10] Liu B，Zhu Y F，Gao C K，et al. LIBERO：Benchmarking knowledge transfer for lifelong robot learning[C]//The 37th Annual Conference on Neural Information Processing Systems，2023：44776-44791.

[11] Xu Z Y，Wu K，Wen J J，et al. A survey on robotics with foundation models：Toward embodied AI[PP/OL]. v1. arXiv(2024-02-04)[2026-03-31]. https://arxiv.org/abs/2402.02385.

[12] Ichter B，Brohan A，Chebotar Y，et al. Do as I can，not as I say：Grounding language in robotic affordances[C]// The 6th Conference on Robot Learning，2022：287-318.

[13] ISO 8373：2021 Robotics—Vocabulary[S].

[14] International Federation of Robotics. World robotics 2025—Industrial robots：Sources and methods[R]. Frankfurt：VDMA Services GmbH，2025：30-31.

[15] Denavit J，Hartenberg R S. A kinematic notation for lower-pair mechanisms based on matrices[J]. Journal of Applied Mechanics，1955，22(2)：215-221.

[16] Nakamura Y. Advanced robotics：Redundancy and optimization[M]. Reading：Addison-Wesley Publishing Company，1991：128-137.

[17] Spong M W，Hutchinson S，Vidyasagar M. Robot modeling and control[M]. Hoboken：John Wiley & Sons，Inc.，2006：187-224.

[18] Slotine J J E，Li W. Applied nonlinear control[M]. Englewood Cliffs：Prentice Hall，1991：212-213，290-302.

[19] Hogan N. Impedance control：An approach to manipulation：Part I—Theory[J]. Journal of Dynamic Systems，Measurement，and Control，1985，107(1)：1-7.

[20] Zhu M L，Gong D W，Zhao Y Y，et al. Compliant force control for robots：A survey[J]. Mathematics，2025，13(13)：2204.

[21] LaValle S M. Rapidly-exploring random trees：A new tool for path planning[R]. Ames：Department of Computer Science，Iowa State University，1998：1-4.

[22] Kavraki L E，Svestka P，Latombe J C，et al. Probabilistic roadmaps for path planning in high-dimensional configuration spaces[J]. IEEE Transactions on Robotics and Automation，1996，12(4)：566-580.

[23] Zucker M，Ratliff N，Dragan A D，et al. CHOMP：Covariant hamiltonian optimization for motion planning[J]. The International Journal of Robotics Research，2013，32(9/10)：1164-1193.

[24] Åström K J，Hägglund T. PID controllers：Theory，design，and tuning[M]. 2nd ed. Research Triangle Park，NC：Instrument Society of America，1995：59.

[25] Luh J，Walker M，Paul R. Resolved-acceleration control of mechanical manipulators[J]. IEEE Transactions on Automatic Control，1980，25(3)：468-474.

[26] Liu J，Yap H J，Khairuddin A S M. Review on motion planning of robotic manipulator in dynamic environments[J]. Journal of Sensors，2024，2024：5969512.

[27] Argall B D，Chernova S，Veloso M，et al. A survey of robot learning from demonstration[J]. Robotics and Autonomous Systems，2009，57(5)：469-483.

[28] Ng A Y，Russell S J. Algorithms for inverse reinforcement learning[C]//The Seventeenth International Conference on Machine Learning（ICML 2000），2000：663-670.

[29] Ross S，Gordon G J，Bagnell J A. A reduction of imitation learning and structured prediction to No-regret online learning[C]// The Fourteenth International Conference on Artificial Intelligence and Statistics，2011，15：627-635.

[30] Correia A，Alexandre L A. A survey of demonstration learning[J]. Robotics and Autonomous Systems，2024，182：104812.

[31] Kober J，Bagnell J A，Peters J. Reinforcement learning in robotics：A survey[J]. The International Journal of Robotics Research，2013，32(11)：1238-1274.

[32] Lillicrap T P，Hunt J J，Pritzel A，et al. Continuous control with deep reinforcement learning[C/OL]//The 4th International Conference on Learning Representations（ICLR 2016），2016：1-14. https://arxiv.org/abs/1509.02971v6.

[33] Schulman J，Wolski F，Dhariwal P，et al. Proximal policy optimization algorithms[PP/OL]. v2. arXiv（2017-08-28）[2026-03-31]. https://arxiv.org/abs/1707.06347. DOI：10.48550/arXiv.1707.06347.

[34] Sutton R S，Precup D，Singh S. Between MDPs and semi-MDPs：A framework for temporal abstraction in reinforcement learning[J]. Artificial Intelligence，1999，112(1/2)：181-211.

[35] Tobin J，Fong R，Ray A，et al. Domain randomization for transferring deep neural networks from simulation to the real world[C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems（IROS），IEEE，2017：23-30.

[36] Da L，Turnau J，Kutralingam T P，et al. A survey of sim-to-real methods in RL：Progress，prospects and challenges with foundation models[PP/OL]. v3. arXiv(2025-03-08)[2026-03-31]. https://arxiv.org/abs/2502.13187. DOI：10.48550/arXiv.2502.13187.

[37] Liang J，Huang W L，Xia F，et al. Code as policies：Language model programs for embodied control[C]//2023 IEEE International Conference on Robotics and Automation（ICRA）. IEEE，2023：9493-9500.

[38] Fujimoto S，van Hoof H，Meger D. Addressing function approximation error in actor-critic methods[C]//The 35th International Conference on Machine Learning，2018：1587-1596.

[39] Haarnoja T，Zhou A，Abbeel P，et al. Soft actor-critic：Off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//The 35th International Conference on Machine Learning，2018：1861-1870.

[40] Dasari S，Ebert F，Tian S，et al. RoboNet：Large-scale multi-robot learning[C]//The Conference on Robot Learning，2020：885-897.

[41] Brohan A，Brown N，Carbajal J，et al. RT-1：Robotics transformer for real-world control at scale[C]//Robotics：Science and Systems XIX. Daegu，Republic of Korea，2023：25.

[42] Shridhar M，Manuelli L，Fox D. Perceiver-actor：A multi-task transformer for robotic manipulation[C]//The 6th Conference on Robot Learning，2023：785-799.

[43] Jang E，Irpan A，Khansari M，et al. BC-Z：Zero-shot task generalization with robotic imitation learning[C]//The 5th Conference

on Robot Learning，2022：991-1002．

[44] Jiang Y，Gupta A，Zhang Z，et al. VIMA：Robot manipulation with multimodal prompts[C]//The 40th International Conference on Machine Learning，2023：14975-15022.

[45] Chi C，Xu Z J，Feng S Y，et al. Diffusion policy：Visuomotor policy learning *via* action diffusion[J]. The International Journal of Robotics Research，2025，44（10/11）：1684-1704.

[46] Shafiullah N M，Cui Z，Altanzaya A，et al. Behavior transformers：Cloning k modes with one stone[C]//The 36th Annual Conference on Neural Information Processing Systems，2022：22955-22968.

[47] Eysenbach B，Gupta A，Ibarz J，et al. Diversity is all you need：Learning skills without a reward function[C/OL]//International Conference on Learning Representations，2019：1. https://openreview.net/forum?id=SJx63jRqFm.

[48] Lynch C，Khansari M，Xiao T，et al. Learning latent plans from play[C]// The Conference on Robot Learning，2020：1113-1132．

[49] Bousmalis K，Vezzani G，Rao D，et al. RoboCat：A self-improving generalist agent for robotic manipulation[J/OL]. Transactions on Machine Learning Research，2024：1-60. https://openreview.net/forum?id=vsCpILiWHu.

[50] Lowe R，Wu Y，Tamar A，et al. Multi-agent actor-critic for mixed cooperative-competitive environments[C]//Advances in Neural Information Processing Systems 30，2017：6379-6390.

[51] Makoviychuk V，Wawrzyniak L，Guo Y，et al. Isaac Gym：High performance GPU based physics simulation for robot learning[C/OL]// The Neural Information Processing Systems Track on Datasets and Benchmarks，2021：1. https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/hash/28dd2c7955ce926456240b2ff0100bde-Abstract-round2.html.

[52] Levine S，Kumar A，Tucker G，et al. Offline reinforcement learning：Tutorial，review，and perspectives on open problems[PP/OL]. v3. arXiv（2020-11-01）[2026-03-31]. https://arxiv.org/abs/2005.01643. DOI：10.48550/arXiv.2005.01643.

[53] Zhao T Z，Kumar V，Levine S，et al. Learning fine-grained bimanual manipulation with low-cost hardware[C]//Robotics：Science and Systems XIX，2023：16.

[54] Silver T，Chitnis R，Tenenbaum J B，et al. Learning symbolic operators for task and motion planning[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems（IROS），2021：3182-3189.

[55] Huang D A，Nair S，Xu D，et al. Neural task graphs：Generalizing to unseen tasks from a single video demonstration[C]//The IEEE/CVF Conference on Computer Vision and Pattern Recognition，2019：8565-8574.

[56] Huang W L，Abbeel P，Pathak D，et al. Language models as zero-shot planners：Extracting actionable knowledge for embodied agents[C]//The 39th International Conference on Machine Learning，2022：9118-9147.

[57] Garrett C R，Chitnis R，Holladay R，et al. Integrated task and motion planning[J]. Annual Review of Control，Robotics，and Autonomous Systems，2021，4：265-293.

[58] Envall J，Poranne R，Coros S. Differentiable task assignment and motion planning[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems（IROS），2023：2049-2056.

[59] Dantam N T，Kingston Z K，Chaudhuri S，et al. Incremental task and motion planning：A constraint-based approach[C]//Robotics：Science and Systems XII，2016.

[60] Cheng S，Xu D. LEAGUE：Guided skill learning and abstraction for long-horizon manipulation[J]. IEEE Robotics and Automation Letters，2023，8（10）：6451-6458.

[61] Hafner D，Pasukonis J，Ba J，et al. Mastering diverse control tasks through world models[J]. Nature，2025，640（8059）：647-653.

[62] Lu G X，Jia B X，Li P H，et al. GWM：Towards scalable gaussian world models for robotic manipulation[C]//The IEEE/CVF International Conference on Computer Vision（ICCV），2025：9263-9274.

[63] Zhu F Q，Wu H T，Guo S，et al. IRASim：A fine-grained world model for robot manipulation[C]//The IEEE/CVF International Conference on Computer Vision（ICCV），2025：9834-9844.

[64] Pang J C，Tang N，Li K Y，et al. Learning view-invariant world models for visual robotic manipulation[C/OL]//International Conference on Learning Representations（ICLR），2025：54853-54876. https://proceedings.iclr.cc/paper_files/paper/2025/hash/8882d370cdafec9885b918a8cfac642e-Abstract-Conference.html.

[65] NVIDIA，Agarwal N，Ali A，et al. Cosmos world foundation model platform for physical AI[PP/OL]. v3. arXiv（2025-07-09）[2026-03-31]. https://arxiv.org/abs/2501.03575. DOI：10.48550/arXiv.2501.03575.

[66] Kim M J，Pertsch K，Karamcheti S，et al. OpenVLA：An open-source vision-language-action model[C]//The 8th Conference on Robot Learning，2025：2679-2713．

[67] Li X H，Liu M H，Zhang H B，et al. Vision-language foundation models as effective robot imitators[C/OL]//International Conference on Learning Representations（ICLR），2024：26703-26721. https://proceedings.iclr.cc/paper_files/paper/2024/hash/71639c317fb0bf398835627b4418693e-Abstract-Conference.html.

[68] Liu J M，Liu M Z，Wang Z Y，et al. RoboMamba：Efficient vision-language-action model for robotic reasoning and manipulation[C]//Advances in Neural Information Processing Systems，2024：40085-40110.

[69] Black K，Brown N，Driess D，et al. π_0：A Vision-language-action flow model for general robot control[C]//Robotics：Science and Systems XXI，2025：10.

[70] Black K，Brown N，Darpinian J，et al. $\pi_{0.5}$：A vision-language-action model with open-world generalization[C]//The 9th Conference on Robot Learning，2025：17-40.

[71] NVIDIA，Bjorck J，Castañeda F，et al. GR00T N1：An open foundation model for generalist humanoid robots[PP/OL]. v2. arXiv（2025-03-27）[2026-03-31]. https://arxiv.org/abs/2503.14734. DOI：10.48550/arXiv.2503.14734.

[72] Shukor M，Aubakirova D，Capuano F，et al. SmolVLA：A vision-language-action model for affordable and efficient robotics[PP/OL]. v1. arXiv（2025-06-02）[2026-03-31]. https://arxiv.org/abs/2506.01844. DOI：10.48550/arXiv.2506.01844.

[73] Lee A C，Chuang I，Chen L Y，et al. InterACT：Inter-dependency aware action chunking with hierarchical attention transformers for bimanual manipulation[C]//The 8th Conference on Robot Learning，2025：1730-1743.

[74] George A，Barati Farimani A. One ACT Play：Single demonstration behavior cloning with action chunking transformers[PP/OL]. v1.

arXiv（2023-09-18）[2026-03-31]. https://arxiv.org/abs/2309.10175. DOI：10.48550/arXiv.2309.10175.

[75] Huang W L，Xia F，Xiao T，et al. Inner monologue：Embodied reasoning through planning with language models[C]//The 6th Conference on Robot Learning，2022：1769-1782.

[76] Nasiriany S，Maddukuri A，Zhang L，et al. RoboCasa：Large-scale simulation of everyday tasks for generalist robots[C]//Robotics：Science and Systems XX，2024：50.

[77] Gu J Y，Xiang F B，Li X L，et al. ManiSkill2：A unified benchmark for generalizable manipulation skills[C/OL]//International Conference on Learning Representations（ICLR），2023：1-30. https://iclr.cc/virtual/2023/poster/12152.

[78] Luo Y L，Fan C K，Dong M H，et al. RoboBench：A comprehensive evaluation benchmark for multimodal large language models as embodied brain[C/OL]//European Conference on Computer Vision（ECCV），2026：501-520. https://eccv.ecva.net/virtual/2026/poster/4213.

[79] Zhang L，Dong J，Bai K，et al. ResponsibleRobotBench：Benchmarking responsible robot manipulation using multi-modal large language models[PP/OL]. v1. arXiv（2025-12-03）[2026-03-31]. https://arxiv.org/abs/2512.04308. DOI：10.48550/arXiv.2512.04308.

[80] Zhao E，Raval V，Zhang H，et al. ManipBench：Benchmarking vision-language models for low-level robot manipulation[C]//The 9th Conference on Robot Learning，2025：3413-3462.

[81] Mandlekar A，Zhu Y K，Garg A，et al. Roboturk：A crowdsourcing platform for robotic skill learning through imitation[C]//Conference on Robot Learning，2018：879-893.

[82] Fu J，Kumar A，Nachum O，et al. D4RL：Datasets for Deep Data-Driven Reinforcement Learning[PP/OL]. v4. arXiv（2021-02-06）[2026-03-31]. https://arxiv.org/abs/2004.07219.