

人形机器人赛道深度之运动控制：制约商业化落地进程的关键环节

-人形机器人行业专题报告

西部证券研发中心

2025年06月24日

分析师 | 齐天翔 S0800524040003 邮箱地址: qitianxiang@research.xbmail.com.cn

分析师 | 张磊 S0800524070001 邮箱地址: zhanglei@research.xbmail.com.cn





核心结论

人形机器人应用广泛，机遇与挑战并存。人形机器人，是指具有人的形态和功能的机器人。随着行业不断发展，具备高度仿人外形、强大的感知能力以及智能决策和学习能力的人形机器人有望广泛进入各种生态场景，如工业自动化、医疗康复、服务与零售、危险救援、家庭服务、教育科研等。随着资本和人才不断涌入，机器人控制和AI技术快速迭代，人形机器人产业日新月异，但从大规模商业化落地角度来看，人形机器人产业在技术、经济、社会等方面仍然面临诸多挑战。

精准的复杂运动控制技术是人形机器人得以广泛应用的基础。人形机器人运动控制是指对机器人进行精确控制，使其能够模拟和执行人类的各种动作和运动。这种控制不仅包括对机器人关节的精确控制，还包括对机器人整体运动轨迹的规划和控制，以确保机器人能够完成复杂的任务，为机器人在实际应用中的广泛发展奠定基础。

基于模型的控制和基于学习的控制方法相辅相成，助力人形机器人运控迭代升级。在软件算法层面，人形机器人的运动控制可以分为基于模型的控制和数据驱动控制，每种控制方法都有其独特的优势，能够解决不同的控制问题。**1) 基于模型的控制：**具有高可解释性等独特优势，但其主要依赖于准确的建模和手动参数调整，这使得在多刚体规划和控制中优化变得复杂。**2) 基于学习的控制：**是指利用数据驱动技术使机器人能够从经验中学习运动策略。基于学习的方法特别适用于传统模型可能难以捕捉与未结构化或不可预测地形互动复杂性的环境。二者相结合的混合控制方式则通过数学模型提供基础控制框架，利用数据驱动方法增强适应性和鲁棒性，兼顾了控制精度和适应性，大大提高了人形机器人的运动和作业能力。

头部企业运控表现亮眼，多环节有望受益。以特斯拉Optimus、宇树G1及波士顿动力Atlas等为代表的人形机器人产品由于硬件方案的差异化，各自擅长的领域有所差异，但均展现出了优秀的运控能力。一般而言，涉及运动控制相关的软件算法开发是由机器人主机厂自研完成。运动控制相关硬件本体（如控制器、执行器、传感器等）的开发分工差异则较大，自研或采购第三方供应商的形式均有。另外，涉及到机器人训练相关的硬件，如动捕设备、遥操设备、仿真软件工具，多为第三方供应商或开源平台提供。

【投资建议】

人形机器人的运动控制是实现其动态步态、精细操作和环境适应的核心技术，也是涉及硬件设备、软件工具和开发套件的综合应用。本文聚焦在人形机器人运动控制软件算法及数据训练的研究，建议关注如下受益环节及标的。

- ① 运控系统：固高科技、雷赛智能。
- ② 软件及仿真工具：NVIDIA Isaac Sim、PyBullet、Gazebo。
- ③ 动作捕捉设备：Xsens、Vicon、凌云光、诺亦腾。

【风险提示】人形机器人产业发展不及预期、数据训练效果不及预期、运控算法迭代不及预期。



目 录

CONTENTS

01

人形机器人概述

02

人形机器人运动控制

03

相关企业

04

投资建议

05

风险提示



人形机器人行业概述

❖ **定义：**人形机器人，又称仿人机器人或类人机器人，是指具有仿人的形态和功能的机器人，具备一定的感知、学习和认知能力。

❖ **特征：**

- ❑ **高度仿人的外形结构：**拥有与人类相似的躯干、四肢和头部。腰部灵活，可轻松实现转身、弯腰等动作。手臂的肩部、肘部和腕部关节活动范围大，能完成复杂的伸展、抓取和操作任务；腿部的膝关节和踝关节使机器人能够稳健行走，步伐节奏与人类相近。手部通常具有多个灵活的手指，可实现精细的抓握动作。
- ❑ **强大的环境感知能力：**配备多种先进的传感器，包括视觉、听觉、触觉等，能够全面感知周围环境信息。
- ❑ **智能决策与学习能力：**借助计算机科学和人工智能技术，它们能够对感知到的大量环境信息进行快速分析和处理。当面对复杂的任务时，人形机器人会根据预设的算法和模型，结合实时的环境信息，制定出最优的行动方案。它们还能通过机器学习算法，从过往的经验中吸取教训，不断优化自己的行为。

❖ **应用场景：**当前人形机器人主要应用于医疗护理、家庭服务、教育娱乐、工业生产和公共服务等领域中，替代人类完成危险或重复性高的工作，进一步释放劳动力，提高社会生产力。

图表：人形机器人在家庭和制造领域的应用



图表：人形机器人在应急领域的典型应用





人形机器人产业面临的挑战

❖ 技术挑战:

- ✓ **数据采集与标注复杂且成本高昂:** 特别是人类动作数据的收集, 需要高精度的传感器和专业的标注团队来确保数据准确性。
- ✓ **多模态信息融合鲁棒性与实时性不足:** 将视觉、听觉、触觉等多种信息有效融合以形成统一的环境认知, 在鲁棒性和实时性方面仍有不足, 尤其在复杂多变环境中, 机器人难以准确理解和应对变化。
- ✓ **自主学习与适应性不足:** 现有机器人虽能进行一定程度的自主学习, 但在处理未知或复杂任务时, 适应性和泛化能力不足。需要更有效的学习算法以从少量样本中快速学习并适应新环境。
- ✓ **学习过程中的安全问题:** 需要确保机器人在学习过程中不会产生偏差或做出有害行为。
- ✓ **能源效率与续航能力:** 尽管电动驱动系统有所改进, 但如何进一步降低能耗、延长工作时间仍是亟待解决的问题。

❖ **市场挑战:** 成本高昂、应用场景不明确、相关的法规政策有待完善、社会接受度有待提高。

❖ **目前国内智能感知和运动控制等核心关键技术仍与国际存在较大差距, 核心零部件尚不能完全自主生产。**

图表：人形机器人发展阶段



图表：人形机器人相关专利分布情况

企业名称	单位：件				
	本体结构	核心部件	智能感知	运动控制	支撑环境
本田技研工业株式会社	1075	29	444	819	47
索尼公司	373	33	593	266	18
深圳市优必选科技股份有限公司	481	95	432	266	33
丰田汽车公司	412	23	384	264	17
精工爱普生株式会社	229	203	455	121	1
三星集团	202	16	351	231	8
软件银行集团	168	12	148	89	43
波士顿动力公司	115	17	124	115	19

注：专利统计时间截至2023年3月31日。



目 录

CONTENTS

01

人形机器人行业概述

02

人形机器人运动控制

03

相关企业

04

投资建议

05

风险提示



人形机器人技术框架及软件架构

❖ 技术框架：

- a) 硬件层：机械结构（手臂、头部、腿部）、动力系统（伺服电机、液压）、传感器等。
- b) 算法/软件层：运动控制（通信、操作系统、控制器）、环境感知（触觉、GPS、视觉）。

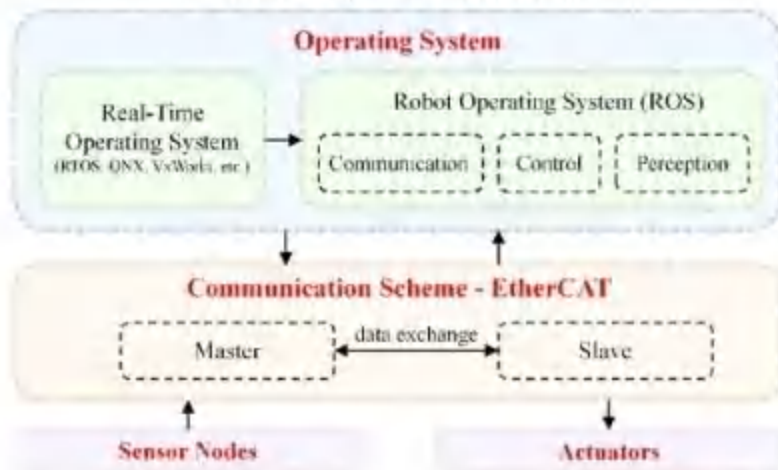
❖ 软件架构：

- a) 操作系统层：RTOS进行精密、及时的运动控制，处理高频传感器输入、协调多关节同步。开源框架ROS提供工具、库和组件间通信，实现通信、控制、感知。
- b) 通信方案层：高性能工业以太网协议EtherCAT（主从式架构），实现了低延迟（微秒级刷新率）、分布式时钟同步（实现精确协调运动）、高可扩展性（支持数百设备）。
- c) 硬件接口层：传感器节点（提供输入数据）、执行器（接收控制指令并执行）。

图表：人形机器人技术框架



图表：人形机器人软件架构





人形机器人运动控制系统框架

运动控制是人形机器人能够在复杂环境中灵活移动的基础。人形机器人运动控制系统一般可分为硬件和软件算法两部分。

❖ 硬件部分：主要包括控制器、伺服驱动器和传感器。

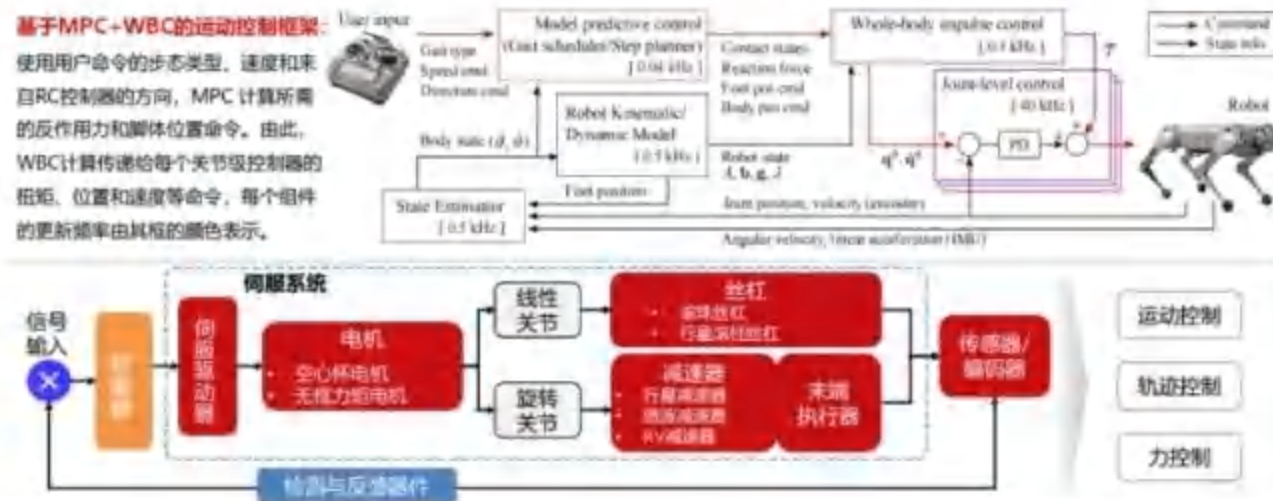
❑ 控制器：运动控制的核心部件，负责协调机器人的运动。

❑ 驱动器：作为执行部件，负责接收控制器的指令并驱动机器人的运动。

❑ 传感器：在运动控制中起到关键作用，它们负责监测机器人的运动状态和环境信息。

❖ 软件算法部分：历经迭代，1) 基于模型的控制算法；2) MPC+WBC融合控制；3) 强化学习和模仿学习等AI技术引入。

图表：人形机器人运动控制系统框架



“小脑”运动控制经历了三代的迭代

- ① 第一代具身小脑的研究是基于模型的控制算法；
- ② 第二代是通过模型来预测控制全身，然后进行全身的自由控制，就是MPC加WBC的控制算法；
- ③ 第三代到现在是通过强化学习加模拟的解决方案，可以让具身智能从环境中获取更优的状态，然后由智能体做出决策，并对环境做出一个适当的行为反应。



人形机器人运动控制的典型方法

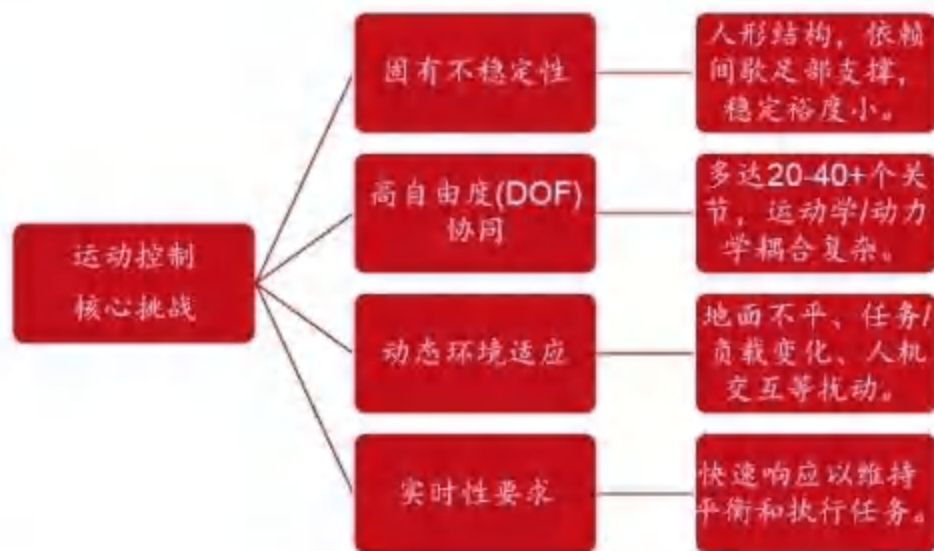
❖ **运动控制分类：**人形机器人的运动控制领域可以分为基于模型的控制方法和基于学习的控制方法，每种方法都有其独特的优势，解决不同的控制问题。

- ✓ **基于模型的运动控制方法：**核心是平衡控制器，平衡控制器接收机器人的状态和用户的命令，以控制机器人的平衡。电机控制器将平衡控制器的输出动作转换为位置或扭矩，以控制机器人。基于模型的方法具有高可解释性等独特优势，但它们主要依赖于准确的建模和手动参数调整，这使得在多刚体规划和控制中优化变得复杂。
- ✓ **基于学习的运动控制方法：**是指利用数据驱动技术使机器人能够从经验中学习运动策略。基于学习的方法特别适用于传统模型可能难以捕捉与未结构化或不可预测地形互动复杂性的环境。

图表：机器人运动控制的分类



图表：运动控制的核心挑战





基于模型的运动控制：核心思想与基本框架

❖ **核心思想**：通过建立机器人的运动学和动力学模型，进行运动轨迹规划和平衡控制，特点是身体控制稳健，步频较慢。

❖ **基本框架**：

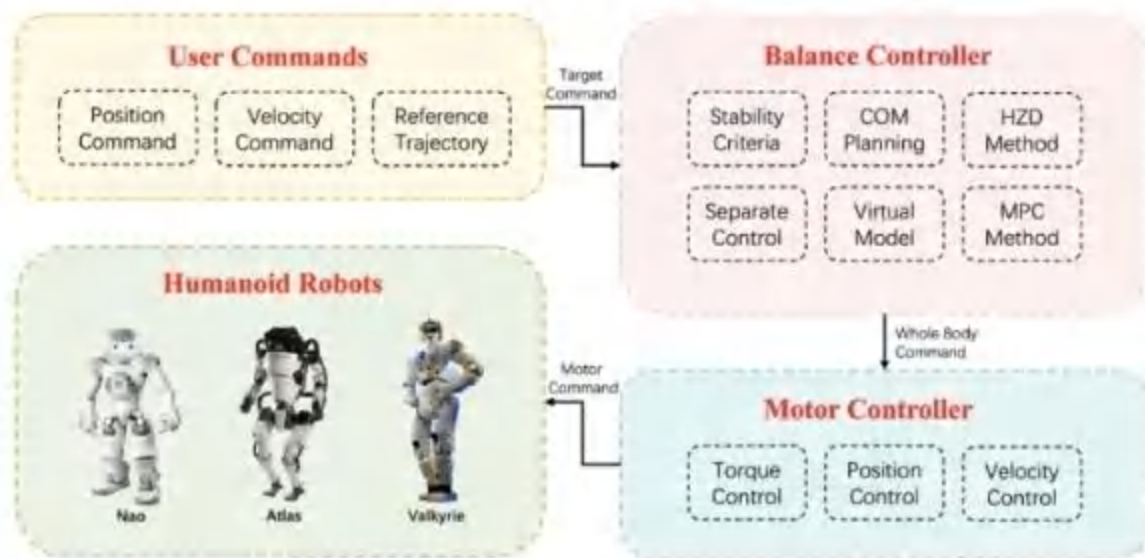
✓ **输入**：用户指令，包含位置指令、速度指令、参考轨迹。

✓ **核心**：平衡控制器(Balance Controller)，包含稳定性判据、质心规划、H2D/MPC等先进控制策略。

✓ **输出**：期望的关节动作（位置/速度/力矩）。

✓ **执行**：电机控制器(Motor Controller)将期望的关节动作转换为底层电机指令。

图表：基于模型的运动控制框架

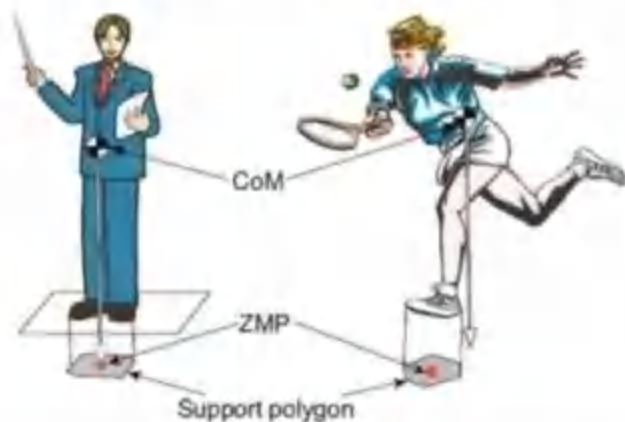




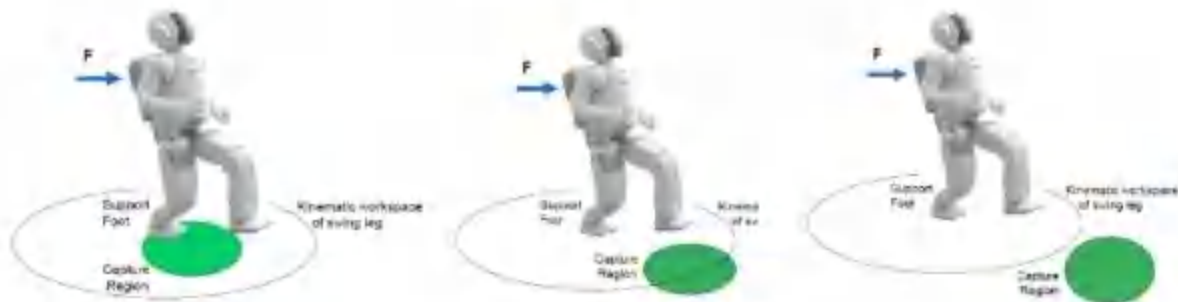
基于模型的运动控制：稳定性判据

- ❖ **稳定性判据(Stability Criteria):** 定义了机器人在执行运动或与环境交互时如何维持平衡和避免跌倒的标准和方法。
- ❖ **核心思想:** 将复杂的机器人动态平衡问题简化为一些可测量、可计算或可控制的指标。通过监控这些指标是否在安全范围内, 或者通过控制策略主动使这些指标保持在安全范围内, 来实现机器人的稳定运动。
- ❖ **常见稳定性判据:**
 - ✓ **零力矩点(Zero Moment Point - ZMP):** 物体所受重力与惯性力的合力的延长线与支撑面的交点。如果ZMP始终保持在机器人的支撑多边形(通常是脚底形成的区域)内部, 机器人就能在动态过程中保持平衡。
 - ✓ **捕获点(Capture Point)/捕获区域:** 如果机器人从当前状态开始不再施加任何额外的控制力(例如, 关节力矩保持不变或为零), 机器人质心(CoM)的水平投影最终会停止的点。如果捕获点在支撑区域之外, 机器人将会跌倒。控制器通过调整落脚点或身体姿态, 使捕获点始终保持在未来的支撑区域内。

图表：基于ZMP判据的运动控制



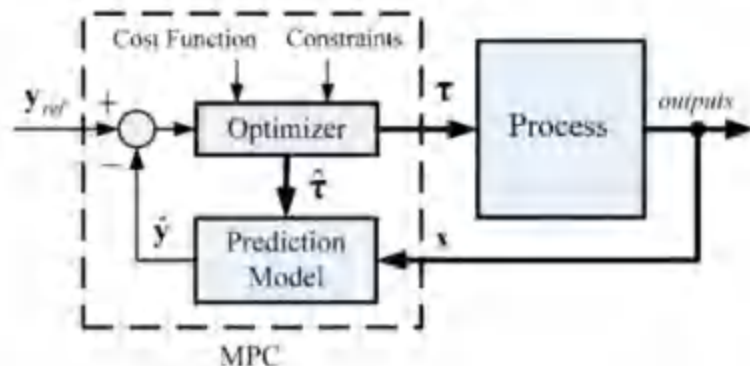
图表：基于捕获区域的推力恢复迈步策略



基于模型的运动控制：模型预测控制

- ❖ **原理：**模型预测控制(Model Predictive Control, MPC)使用机器人系统的动力学模型来预测机器人在未来有限时间范围内的行为，并通过优化一个目标函数来计算当前的最优控制输入。
 - ✓ **模型预测：**通过测量机器人的状态（如机身位置、速度和角速度），结合机器人动力学模型，预测出一个时间序列内的机身状态。
 - ✓ **优化求解：**将一个时间序列内的参考状态与实际测量值输入优化求解器，并结合约束条件，求得一段时间内的控制指令。
 - ✓ **实时控制：**选取第一个控制指令作为期望的末端接触力与力矩，再输入到下一个控制模块中。
- ❖ **应用：**MPC早期用于在线步行模式生成，以提高对干扰和用户命令的响应能力；DARPA机器人挑战赛后，MPC主要用于在线轨迹生成以及全身控制。人形机器人平台如HRP系列、Atlas、TORO、TALOS、E2-DR等。
- ❖ **优点：**可以在非线性域中生成轨迹，充分利用运动学可动范围和扭矩极限；通过频繁地根据当前状态重新规划运动，使机器人对扰动更具鲁棒性。
- ❖ **缺点：**
 - ✓ **高计算复杂度、实时性不足：**非线性MPC需在短采样时间内求解复杂NLP问题，计算量大；通用求解器和SQP方法难以满足实时需求。
 - ✓ **模型依赖性强：**性能依赖动态模型准确性，模型误差易降低控制精度。

图表：MPC的通用框图



图表：基于MPC的人形机器人实例(TALOS, Atlas)

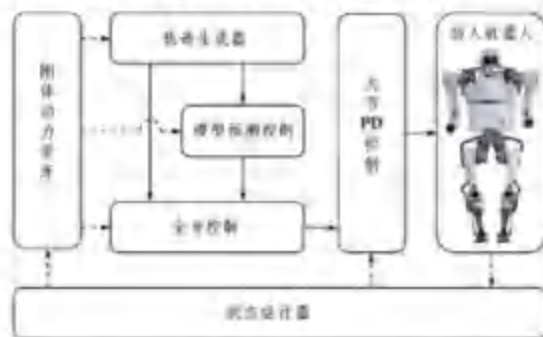




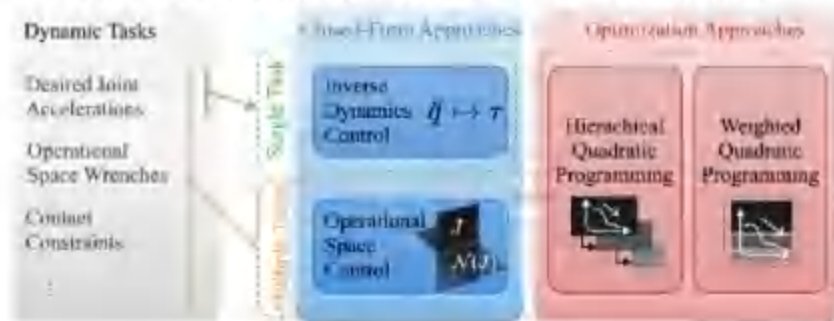
基于模型的运动控制：全身控制

- ❖ **原理：**全身控制(Whole-Body Control, WBC)是一类产生关节力矩、约束力和广义加速度的控制器，以实现给定的一系列期望动态任务。
 - ✓ **建模系统动力学和约束：**建立机器人的全阶Euler-Lagrangian动力学方程，该方程关于决策变量(广义加速度 $\ddot{q}$ 、外部力 λ 和关节扭矩 τ)是线性的；定义与环境之间的物理接触约束。
 - ✓ **定义动态任务：**获得所需的机器人行为或目标(如跟踪关节加速度、避免碰撞等)，将其表达为关于决策变量($\ddot{q}, \lambda, \tau$)的线性方程。
 - ✓ **构建并解决控制问题：**将系统动力学、接触约束和动态任务结合起来，形成一个数学问题来求解当前的决策变量($\ddot{q}, \lambda, \tau$)。有闭环形式方法和基于优化的方法。
 - ✓ **输出控制信号：**输出计算得到的决策变量，发送到机器人的执行器，驱动关节运动。
- ❖ **应用：**通常接收MPC输出的操作空间任务(如质心/末端执行器轨迹)，也支持遥操作/VR界面生成动态任务(姿态、行走方向、抓取目标)。
- ❖ **优点：**通常采用线性化动力学方程，计算高效；闭环形式的WBC计算成本较低；基于优化的WBC增强了灵活性，可以模块化地添加和移除动态任务，并能有效地处理不等式约束(如关节限制、障碍物避免)。
- ❖ **缺点：**依赖模型精度，WBC的性能很大程度上依赖于对机器人动力学模型的准确表示；将动态交互建模为外部力需要准确的力/扭矩感知或全身触觉感知。

图表：MPC+WBC的通用框图



图表：求解WBC的闭环形式方法、基于优化的方法





基于模型的运动控制：模型预测控制+全身控制

❖ 算法框架：

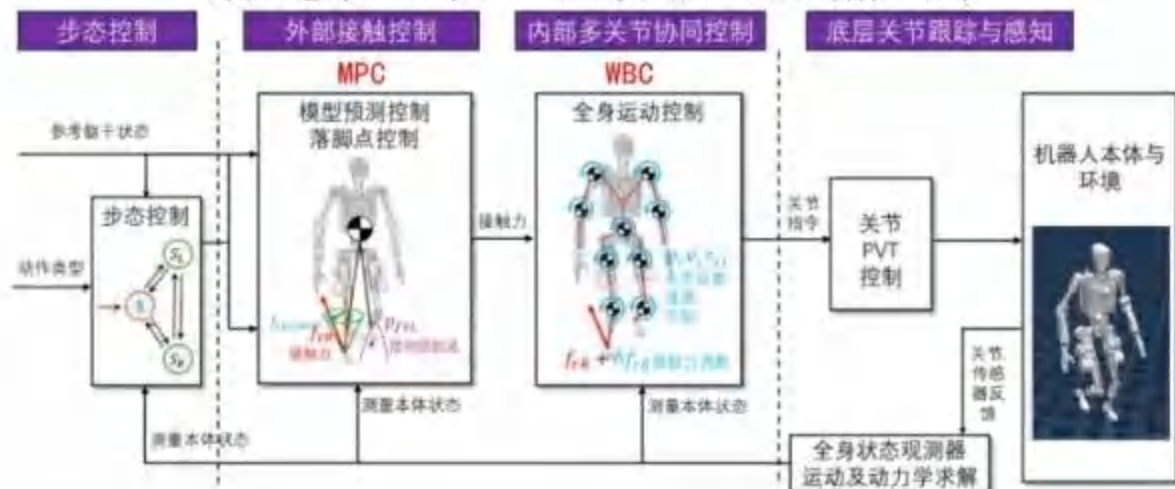
- ✓外部接触控制：主要依靠 MPC 和落脚点的规划来求取末端的接触力和轨迹。
- ✓内部多关节协同控制：依靠 WBC 根据指令计算关节的速度和位置。
- ✓底层关节跟踪与感知模块：计算机器人最终所需的指令；进行状态估计，获取机器人的状态，为前面的控制模块提供反馈。

❖ 应用：业内比较主流的传统控制(基于模型的控制)的方案。

❖ 优点：

- ✓MPC增强全局性与长期规划能力、提升动态稳定性、提高复杂地形适应性；
- ✓WBC更加侧重于协调机器人多自由度的全身运动，在多任务和复杂约束条件下完成高效、鲁棒的控制。

图表：基于MPC与WBC的人形机器人运动控制算法框架





基于学习的运动控制：核心思想与基本框架

❖ **核心思想**：利用计算资源和数据，让机器人自主或通过专家指导来获取和掌握运动-操作(LOCO-MANIPULATION)技能，以适应复杂、动态和不确定的环境，而不是依赖于人类预先编程的复杂模型或控制器，并可能发现超越人类预设编程能力的新行为。

✓ **数据驱动 (Data-Driven)**：与依赖人类设计者先验知识提炼的传统预编程方法不同，基于学习的方法以数据为核心。这些数据可以来源于：

- 自主探索与交互 (RL)：机器人通过与环境的试错交互，根据反馈（奖励/惩罚）积累经验。
- 专家演示 (IL)：从人类专家或其他已有策略中获取高质量的行为数据。

✓ **从经验中提炼策略**：无论是通过RL的奖励最大化，还是IL的演示模仿，最终目标都是形成一个策略 (policy)——一个将感知输入映射到动作输出的函数或模型。

✓ **泛化与多功能性**：让机器人学习到的技能不仅能在训练过的特定场景下工作，还能泛化到新的、未曾见过的任务和环境中。

✓ **弥合模拟与现实**：由成本、安全和效率考虑，大量学习在模拟中进行，需要有效迁移到真实机器人。

❖ **基本框架**：无参考数据的强化学习、机器人经验模仿学习、人类行为数据模仿学习、混合学习。

图表：基于学习算法的实例优势



提升运动的鲁棒性

通过多阶段设计仿真环境训练，实现了受试者、美国、越南、上下运动等运动学习 (RSS2021, Humanoids 2022)



提升运动的准确性

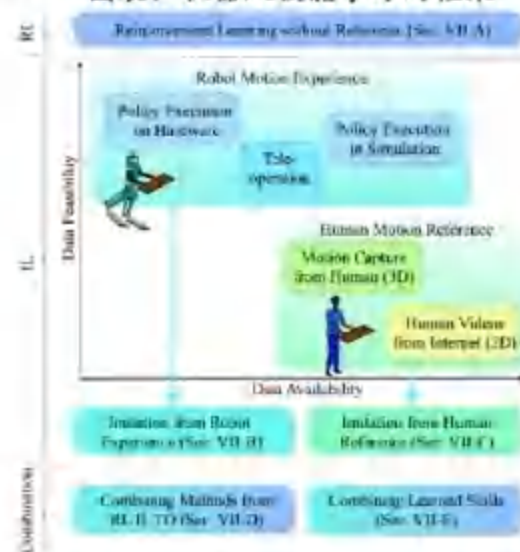
借助生成式模型学习技能全参数模型，实现网络行走 (ACM SIGGRAPH 2020, arXiv 2022)



提升负载变化适应性

采用预训练Transformer模型，从观测到的历史中对未来行动进行自回归预测训练 (ICRA2022)

图表：机器人技能学习的框架





基于学习的运动控制：强化学习原理与应用

❖ **原理：**强化学习(RL)中，机器人通过与环境交互、根据奖励和惩罚信号自主学习运动技能，其训练过程最少或不需要监督。

✓ **强化学习框架：**智能体在环境中执行动作，环境会反馈一个奖励(reward)或惩罚(penalty)信号，并使智能体进入新的状态。智能体的目标是学习一个策略(policy)，即从状态到动作的映射，以最大化长期累积奖励。

✓ **实例：基于actor-critic强化学习的机器人控制框架。**强化学习策略 π 基于机器人当前状态和参考运动状态，计算出一个动作增量 δa ，加参考关节角度 $\hat{a}$ 得到最终的目标关节角度 a 。最后，低层PD控制器使用这些目标关节角度来驱动机器人，从而实现对参考运动的跟踪和鲁棒的行走。

❖ **应用：**

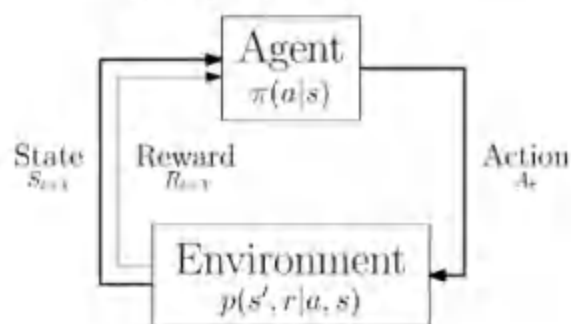
✓ **复杂全身运动协调：**RL在协调人形机器人的复杂全身运动方面取得了成功，包括跳舞、敏捷的足球动作和鲁棒的运动。

✓ **特定任务微调：**RL策略通常针对特定环境中的特定任务进行微调。

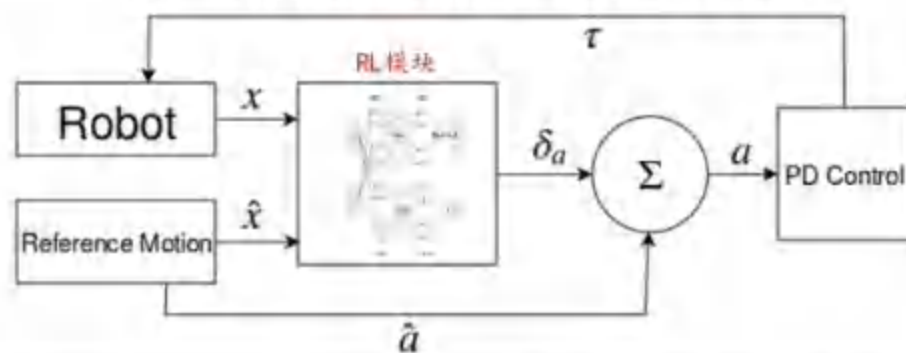
✓ **模拟环境训练：**由于真实世界训练的挑战，RL策略主要在模拟中进行训练。

✓ **混合学习方法：**RL常用于混合学习方法中，如微调IL预训练的策略，或作为教师策略指导学生策略学习。

图表：强化学习框架



图表：基于actor-critic强化学习的机器人控制框架





基于学习的运动控制：强化学习优缺点

❖ 优点：

- ✓ 自主学习新行为：能够通过试错发现人类设计者难以编码的新颖或更优的行为策略。
- ✓ 无需精确动态模型：许多RL方法（如PPO、SAC）是无模型的，降低了对机器人精确物理模型的依赖。
- ✓ 无需演示数据：从零开始的RL不依赖专家演示，简化了初始训练数据的收集。
- ✓ 端到端学习：可以将原始传感器输入直接映射到低级控制指令，简化控制流程。
- ✓ 策略可实时执行：训练好的RL策略通常可以实时运行。

❖ 缺点：

- ✓ **Sim-to-Real差距**：模拟器与真实世界物理模型（如接触动力学、摩擦、驱动器特性）的差异，导致在模拟中训练的策略在真实机器人上表现不佳或失败。
- ✓ **奖励函数设计困难**：RL的性能高度依赖于奖励函数的设计，需要精心设计才能引导策略学习期望的行为，这本身就是一个复杂的工程问题。
- ✓ **样本效率低下**：从零开始学习通常需要大量与环境的交互（数百万甚至数十亿步），在真实机器人上进行如此大量的试错既耗时又不安全。流行的RL算法（如PPO、SAC）在大多数类人机器人运动-操作任务中都失败了，部分原因在于这些任务的复杂性、样本效率低下以及稀疏的奖励设计。
 - ✓ **PPO及类似在线算法**：样本效率低下。
 - ✓ **SAC及类似离线算法**：由于策略更新的计算成本较高，训练速度较慢。
- ✓ **任务完成无保证**：从零开始的策略学习无法保证任务完成。
- ✓ **真实世界训练不切实际**：使用RL训练真实世界的机器人几乎是不切实际的（在当前发展阶段）。



基于学习的运动控制：模仿学习原理

❖ **原理：**模仿学习(IL)是一个涵盖多种算法（监督学习、无监督学习和强化学习）的总称。这些算法从专家演示中训练策略。步骤如下：

① **捕获专家演示：**获取专家行为数据。

② **重定向(Retargeting)：**将捕获的演示映射到机器人运动。若演示直接来自机器人（例如通过远程操作），则不需要重定向步骤。

③ **策略训练：**使用重定向后的数据训练机器人策略。

❖ **数据来源：**机器人经验数据（现有策略、远程操作）和人类行为数据（运动捕捉、人类视频）。

✓ **现有策略：**已经完成训练的现成机器人智能体的策略，数据质量和数据可得性都比较高。

✓ **远程操作：**远程操作是直接捕获由人类专家指挥的机器人数据的最常见方法之一。远程操作的主要优势在于它能够为各种任务提供流畅、自然和精确的轨迹。数据质量较高，数据可得性略低于策略执行。

✓ **运动捕捉：**通过动作捕捉仪器及算法捕获人体运动轨迹及与周围环境的交互数据。数据可得性较高，数据质量较低，但略高于人类视频。

✓ **人类视频：**从互联网获取的视频和图像提供了丰富多样的人体运动数据来源，包括运动表现、艺术舞蹈或日常琐事。数据质量最低和但可得性最高。

图表：基于四种数据来源的模仿学习的优缺点

方法与数据	优缺点
无参考强化学习 (RL Without Reference)	✓ ☐ 新行为 ✗ 奖励调整
模仿学习-现有策略 (IL Robot Execution)	✓ ☐ 有标注数据 ✓ ☐ 动态可行 ✗ 数据稀缺 ✗ 多样性有限
模仿学习-远程操作 (IL Teleoperation)	✓ ☐ 多模态行为 ✗ 很少有全身运动
模仿学习-动作捕捉 (IL Motion Capture)	✓ ☐ 精确的运动学 ✗ 数据集小 ✗ 户外数据有限 ✗ 仅本体感受
模仿学习-人类视频 (IL Human Video)	✓ ☐ 多样丰富的数据 ✗ 非物理运动 ✗ 仅本体感受

图表：通过远程操作、人类视频、动捕进行模仿学习





基于学习的运动控制：模仿学习原理

❖ 模仿学习方法：

- ✓ **行为克隆(Behavior Cloning, BC)**: 将IL视为监督学习问题, 从单模态数据中学习, 常用于策略蒸馏。直接学习一个从观察到动作的映射函数(如一个神经网络), 目标是最小化预测动作与专家动作之间的差异。
 - **局限性**: 朴素的BC方法容易受到分布漂移(distribution shift)的影响, 即由于策略的微小误差累积, 导致机器人进入训练数据中未曾见过的状态, 从而使得策略表现不佳。
- ✓ **动作分块(Action Chunking Transformer, ACT)**: 为了解决朴素BC中的分布漂移问题, 并处理演示数据的多模态性(即对于同一任务, 专家可能提供多种不同的成功轨迹)。利用Transformer架构的序列建模能力, 将动作序列进行分块处理, 从而更好地捕捉长期依赖关系和动作的多样性。最后生成多样的未来动作序列, 而不仅仅是单一步的动作预测。
- ✓ **扩散策略(Diffusion Policy)**: 属于BC方法的一种, 通过从多个专家策略收集的大型数据集中学习, 来获取多模态运动技能的能力。基于扩散模型, 逐步向数据添加噪声(前向过程)然后学习从噪声中恢复原始数据(反向过程)。
- ✓ **逆强化学习(Inverse Reinforcement Learning, IRL)**: 与朴素BC不同, IRL旨在从专家演示中重建或推断专家行为背后的奖励函数。一旦奖励函数被推断出来, 就可以使用标准的强化学习算法来训练一个策略, 以最大化这个推断出的奖励。

❖ 处理人类行为数据特有难点的方法：

- ✓ **弥合具身差距**: 通过各种重定向策略(如关节空间对应, 任务空间对应, 接触点, 指尖、步态同步, 运动可行性滤波器)来解决机器人与人类的形态差异。
- ✓ **处理不完整的演示数据**: 由于人类数据通常仅包含本体感受信息, 缺乏来自与环境交互的触觉或力数据, 并且不是完整的状态-动作对。这限制了学习丰富的物理交互能力。因此IL控制策略被训练来在物理模拟器中跟踪参考。物理模拟器在此过程中提供感官输入并验证策略动作的物理可行性。
- ✓ **生成对抗模仿学习(Generative Adversarial Imitation Learning, GAIL)**: 采用生成对抗网络(GAN)的框架。一个生成器试图产生与专家演示相似的行为, 一个判别器试图区分生成器产生的行为和真实的专家行为。通过这种对抗过程, 策略被训练得越来越像专家。



基于学习的运动控制：模仿学习应用与优缺点

❖ 应用：

- ✓ 工业公司（如特斯拉和丰田研究所）通过远程操作扩展数据收集，以用于扩散策略方法。
- ✓ 一系列用于研究和硬件评估的开源类人机器人硬件项目，例如 Hector、MIT Humanoid、Berkeley Humanoid、Stanford HumanPlus 等。比如 HumanPlus 能实时模仿人类操作员的全局运动，也能在真实世界中自主完成复杂任务。

❖ 优点：

- ✓ 学习复杂技能：对于难以明确指定的复杂任务特别有效。
- ✓ 解决 RL 奖励设计难题：IL 通过利用演示数据来解决 RL 中奖励函数难以设计的问题。
- ✓ 人类数据的潜力：数据丰富且易获取、简化行为合成、解锁广泛能力。

❖ 缺点：机器人经验数据稀缺；当前捕获人类关节姿态的数据采集方法只能学习人类“在做什么”，而非“如何做”以及“为什么这么做”。

图表：斯坦福 HumanPlus 机器人



图表：特斯拉人形机器人 Optimus 二代





基于学习的运动控制：基于强化学习和模仿学习的混合学习

❖ **原理：**基于RL和IL的混合学习，旨在利用IL从演示中快速学习基础行为的能力，以及RL通过自主探索进一步优化和适应的能力，从而实现有效的模拟到现实的转移。一般分为两种教师-学生范式：

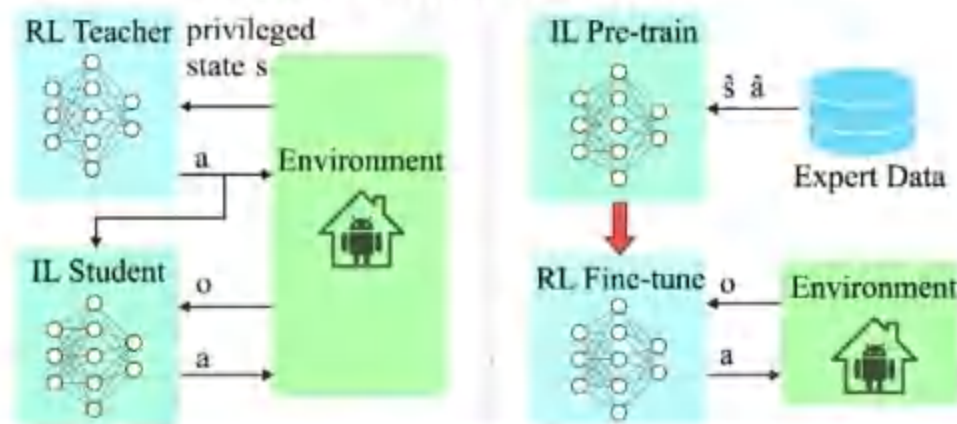
- ✓ **RL教师，IL学生：**首先使用纯RL，利用模拟环境中的特权观测（即模拟器才能提供的完美信息，如物体精确位置、内部状态等）训练一个高性能的教师策略。然后，一个学生策略通过模仿学习（IL）来学习教师策略的行为。关键在于，学生策略仅使用**部分观测**（即机器人实际部署时可获得的传感器信息）。使学生策略在仅依赖实际可用信息的情况下，达到与教师策略相似的性能，并能直接在硬件上部署。
- ✓ **IL预训练，RL微调：**首先使用IL从专家数据中预训练一个模仿策略。然后使用RL对这个模仿策略进行微调。从而超越IL专家本身的性能，并能适应不同的环境或任务。

❖ **应用：**模拟到现实的有效转移。结合纯RL和IL的方法（特别是教师-学生范式）被广泛用于实现有效的模拟到现实的转移。

❖ **优点：**结合IL可以快速引导RL学习方向，避免从零开始的低效探索；RL微调可以使IL学习到的策略性能超越专家。

❖ **缺点：**教师策略的局限性。如果教师策略（无论是RL训练的还是其他方式获得的）本身存在缺陷或次优，学生策略模仿的效果也会受限。

图表：两种教师-学生范式对比

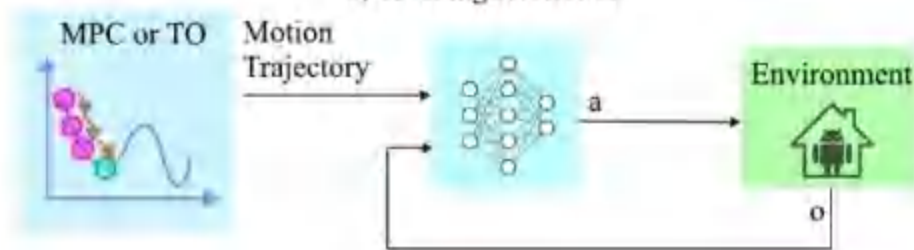




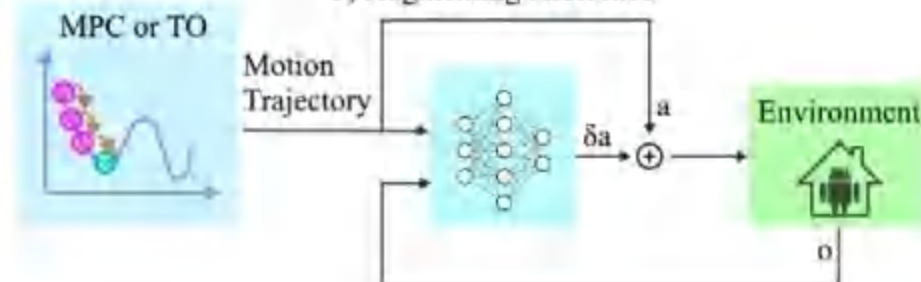
基于学习的运动控制：基于学习方法和基于模型方法的混合学习

- ❖ **原理：**基于学习方法和基于模型方法的混合学习，利用基于模型方法（如轨迹优化TO、模型预测控制MPC）生成高质量的参考轨迹或提供模型知识，以指导和加速基于学习的方法的训练过程。
 - ✓ **轨迹参考跟踪学习 (Learning to Track Trajectory Reference)：**由基于模型的方法（如MPC或TO）生成参考运动轨迹，这些轨迹随后被用作基于学习的方法（RL或监督学习）的模仿目标或奖励信号。在线RL跟踪离线轨迹的方法避免了在线计算MPC可能出现的训练时间过长或失败的问题。
 - ✓ **轨迹参考增强学习 (Learning to Augment Trajectory Reference)：**学习到的策略不是完全复制参考轨迹，而是在参考轨迹的基础上进行适应性的修改或残差增强。这允许策略在遵循基本指导的同时，根据实际情况进行微调。既利用了参考轨迹的指导作用，又通过学习增强项赋予了策略一定的适应性和对变化的鲁棒性。
- ❖ **应用：**在线RL跟踪离线MPC生成的轨迹已被证明对实现多功能双足运动有效；通过学习任务特定的力输出修改参考轨迹，实现了类人机器人的挥拍。
- ❖ **优点：**加速学习。参考轨迹为学习过程提供了明确的指导，显著加快了策略的收敛速度，避免了从零开始的低效探索。
- ❖ **缺点：**学习到的策略性能高度依赖于MPC或TO生成的参考轨迹的质量。如果参考轨迹本身次优或不适应所有情况，策略的上限会受限。

图表：轨迹参考跟踪学习框架
2) Tracking Reference



图表：轨迹参考增强学习框架
3) Augmenting Reference





基于模型的方法 V.S. 基于学习的方法

- ❖ 基于模型的方法(MPC+WBC): 精于精确运动与操作, 实时性好, 功能多样, 但对模型依赖高, 感知与泛化能力有限。
- ❖ 强化学习(RL): 鲁棒性与运动能力极强, 能整合多模态感知, 但精度、泛化及操作能力尚有不足。
- ❖ 模仿学习(IL): 灵活整合多模态感知, 实时性与多功能性好, 但鲁棒性、运动及操作能力偏弱, 泛化受限。

图表：基于模型的方法 V.S. 基于学习的方法

性能指标	基于模型的MPC+WBC	强化学习 (RL)	模仿学习 (IL)
整合多模态传感器输入的灵活性	低	中	高
鲁棒性	高	极高	低
运动精度	极高	中	中
实时可行性	高	高	高
多功能性/通用性	高	中	高
泛化能力	中	低	低
运动能力	极高	极高	中
运动-操作能力	高	中	低
全身多点接触能力	中	高	NA



目 录

CONTENTS

01

人形机器人行业概述

02

人形机器人运动控制

03

相关企业

04

投资建议

05

风险提示



人形机器人主要厂商

公司	国家	代表型号	关键指标	应用领域
特斯拉	美国	Optimus Gen 2	身高~1.73m、重量~57kg	工业制造、家庭自动化
Appteronik	美国	Apollo	身高1.7m、重量~60kg、28+ DOF 关节	工业制造
Figure AI	美国	Figure 02	身高~1.7m、重量~60kg、单手16个自由度	工业制造、物流、家庭
Agility Robotics	美国	Digit	身高~1.2m、重量30kg	城市导航、物流
波士顿动力	美国	Electric Atlas	身高~1.5m、重量~75kg	工业制造、搜救、科研
1X Technologies	挪威	NEO	身高~1.6m、重量~70kg、24+ DOF 关节	工业自动化、家庭自动化
智元机器人	中国	远征A2	身高~1.69m、重量~69kg、40+ 全身主动自由度	营销客服、展厅讲解、前台、业务咨询、商超导览
宇树科技	中国	G1	身高~1.4m、重量40-50kg	服务业、科研
众擎机器人	中国	SE01	身高~1.70m、重量55kg、32 全身主动自由度、单手6个自由度	工业制造
开普勒机器人	中国	Kepler K2	身高1.78m、重量~85kg、40 全身自由度	工业制造、仓储物流、高危作业、科研教育
星动纪元	中国	星动STAR1	身高~1.7m、重量~60kg	工业制造、仓储物流、商业服务、家政
优必选	中国	Walker S1	身高~1.72m、重量~76kg、支持多模态规划大模型	工业制造

- ❖ 特斯拉Optimus以其车端自动驾驶(Autopilot)技术为基础,通过完全端到端本地神经网络实现机器人任务处理。Optimus采用纯视觉方案,利用来自机器人 2D 摄像头的视频以及板载本体感受器上的数据,通过预先训练完成的搭载在嵌入式FSD计算机上的本地神经网络直接产生关节控制序列,从而对关节进行控制。完全端到端神经网络的训练数据通过人类远程操作收集,并针对各种任务进行了扩展。
- ❖ 特斯拉建立AI计算集群Cortex、研发专用芯片Dojo。为了满足FSD以及人形机器人的神经网络训练算力需求,特斯拉于奥斯汀特斯拉超级工厂内部署Cortex AI训练超级集群,集群中采用大约9万个英伟达H100GPU以及大约4万个特斯拉AI4计算机,同时Dojo1拥有大约8000个H100的算力。如此庞大的算力将助力特斯拉加快训练Optimus所搭载的神经网络,提升其行走及任务完成水平。

图表：通过人类远程控制获取训练数据



图表：特斯拉 Dojo 超级计算机单元





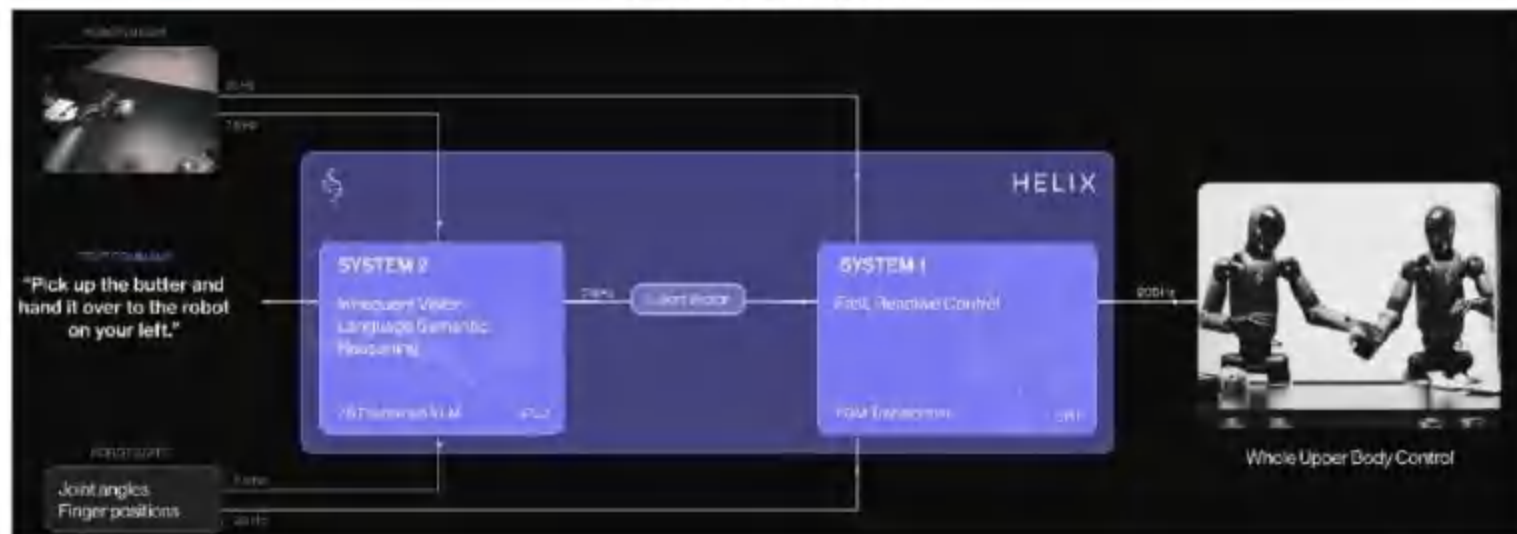
Figure AI

❖ Figure AI 推出了通用的视觉-语言-动作 (VLA) 模型 Helix，能够实现全身控制、多机器人协作、以及基于自然语言提示拿起数千种未遇到过的物品等多种功能。

❖ Helix 模型由两个系统组成：

- ✓ 系统1 (S1)：一种快速反应的视觉运动策略，可将 S2 产生的潜在语义表征转换为 200 Hz 的精确连续机器人动作。S1 是一个 8000 万参数的交叉注意力编码器-解码器 Transformer (一种神经网络结构)，负责低级别控制。它依赖于一个完全卷积的多尺度视觉骨干网络进行视觉处理，该网络完全在模拟环境中进行预训练初始化。虽然 S1 接收与 S2 相同的图像和状态输入，但它以更高的频率处理这些输入，以实现更灵敏的闭环控制。来自 S2 的潜在向量被投影到 S1 的标记空间中，并与 S1 视觉骨干网络提取的视觉特征沿序列维度连接，提供任务条件。
- ✓ 系统2 (S2)：一个机载互联网预训练的 VLM，以 7-9 Hz 运行，用于场景理解和语言理解，实现跨物体和上下文的广泛泛化。S2 基于一个 70 亿参数的开源、开放权重的 VLM 构建，该 VLM 在互联网规模数据上进行了预训练。它处理单目机器人图像和机器人状态信息（包括手腕姿态和手指位置），并将它们投影到视觉语言嵌入空间中。结合指定期望行为的自然语言指令，S2 将所有语义任务相关信息提炼为一个连续的潜在向量，传递给 S1 以调节其低级别动作。

图表：Helix 运行原理

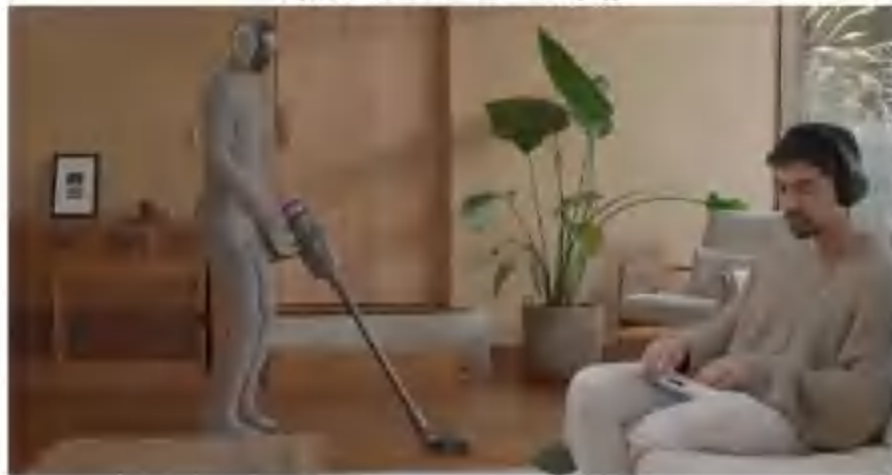




1X Technologies

- ❖ 1X Technologies于2025年2月22日发布了最新人形机器人产品——**NEO Gamma**。该款机器人拥有柔软纺织物皮肤以及流畅的人机交互及行走姿态，专门针对家庭服务领域开发，可以承担不少家务劳动。
- ❖ **NEO Gamma搭载算法：**
 - ✓ **全身控制器（WBC）：**通过强化学习和对海量人体动作数据的训练，NEO Gamma能够以100Hz的频率模拟人类自然行走、摆动手臂，甚至蹲下捡拾物品。
 - ✓ **视觉操作模型：**1X训练了一个视觉操作模型，该模型能够让机器人在不同场景中拾取各种对象，包括此前训练时从未接触的对象。另外，Gamma还利用了经过训练的神经网络，该网络可直接根据原始传感器数据来预测远程操控动作。
 - ✓ **自研大语言模型：**Gamma的Companion功能集集成了一个新的内部语言模型，可实现与用户的自然对话和肢体语言互动。
- ❖ **1X Technologies自研世界模型：**收集了数千小时关于EVE人类在家庭和办公室中执行各种移动操作任务并与人们互动的数据。1X将视频和动作数据结合起来，训练了一个世界模型，该模型可以通过观察和动作预测未来的视频。

图表：NEO机器人做家务



图表：世界模型中的世界

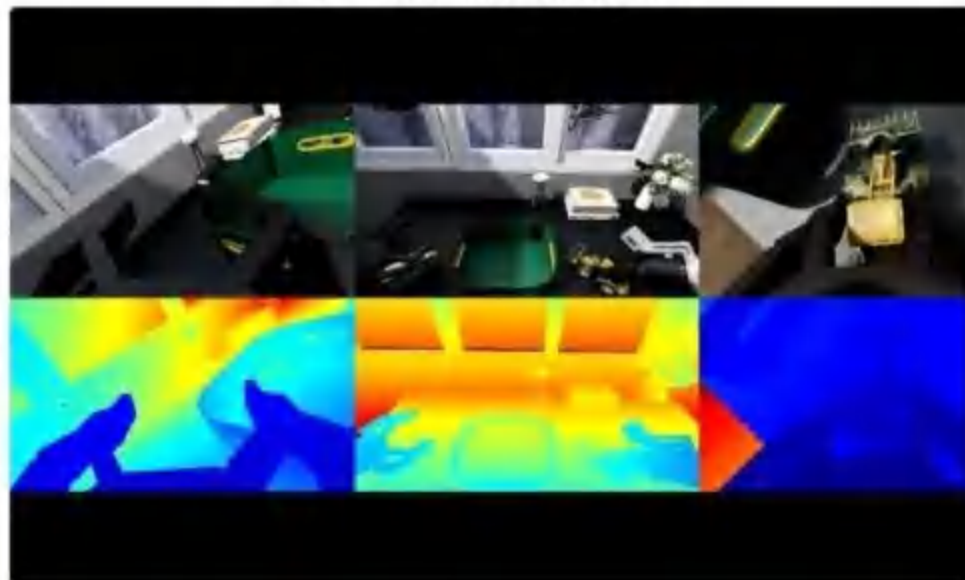




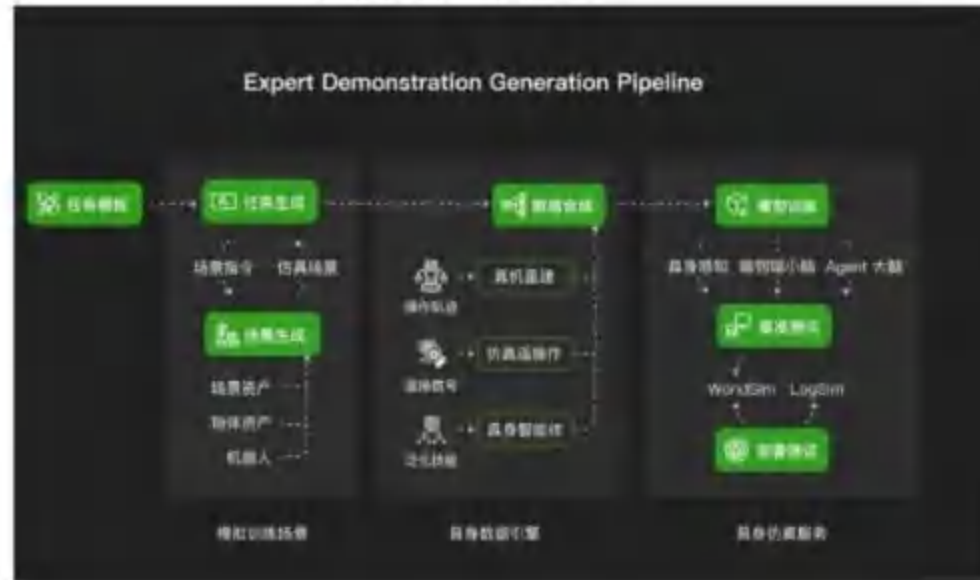
智元机器人

- ❖ 智元机器人成立于2023年2月，是一家致力于以AI+机器人的融合创新、打造世界级领先的具身智能机器人产品及应用生态的创新企业。2023年8月，智元机器人发布第一代通用型具身智能机器人原型机远征A1。
- ❖ 智元机器人推出AgiBot Digital World大型仿真框架，为机器人技术提供高保真模拟与数据支持。该框架集成海量三维资产库、专家轨迹生成机制和评估工具，通过NVIDIA Isaac-Sim平台实现高度逼真的视觉渲染与物理模拟，支持自动化生成多样化训练场景与任务，有效缩小仿真与现实的域差异。在进行强化学习的过程中，由于学习过程需要不断试错，在真实世界试错成本太高，因此仿真模拟必不可少。

图表：自动生成多模态数据集



图表：模拟训练过程



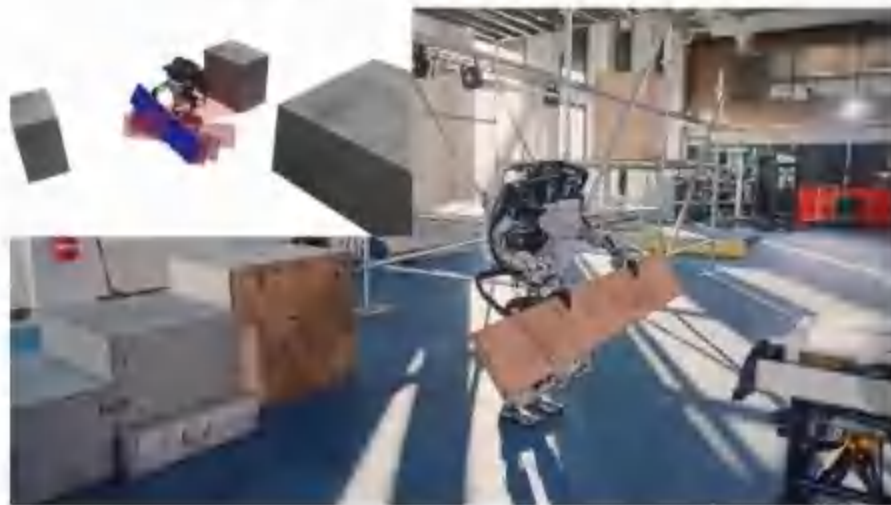


- ❖ Atlas通过模型预测控制(MPC)与扩展动力学模型，实现机器人本体与外部物体的动态协同优化。其算法将机器人关节运动、连杆动量与搬运物体的作用力统一建模，实时预测并调整动作轨迹。例如，搬运重物时同步优化机器人与物体的动量分布，利用牛顿第三定律抵消外力扰动；在跳跃或抛掷任务中，通过MPC框架将离线动作库与实时指令融合，确保全身关节、机械臂与物体运动学约束下的动态平衡，从而完成高难度操作。
- ❖ 分层行为融合与动态极限突破驱动Atlas向更复杂的物理交互场景演进。算法结合预定义动作库（如空翻、抛掷）与实时运动优化，在MPC控制下实现“超极限”动作，例如空中蜷缩肢体加速旋转后精准着陆。同时，通过显式建模物体惯性（如木板、工具袋），Atlas可主动利用外部动量辅助平衡，模拟人类借力原理。未来，波士顿动力将持续探索MPC在复杂环境感知与任务泛化中的应用，推动人形机器人从实验室特技向真实场景的实用化跨越。

图表：MPC预测下机器人运动



图表：MPC与物体显示模型



- ❖ 宇树科技有三类人形机器人产品：通用人形（H1/H1-2, G1）、应用方案（G1-Comp, 铁甲霸王）、灵巧手（Dex3-1, Dex5-1）。
- ❖ H1/H1-2：高端旗舰型号，售价65万元；全球首款实现原地空翻的电驱人形机器人，移动速度达3.3米/秒，关节扭矩360N·m，动态稳定性极强；配备3D激光雷达+深度相机，支持360°全景深度感知，适应复杂环境；超轻量机身（<50kg），18个自由度关节，铝合金骨架+碳纤维材质，续航依赖67.2V高压电池；集成端到端AI模型，支持深度交互和复杂任务自主决策。
- ❖ G1：亲民普及型号，售价9.9万元起；小跑速度2米/秒，最大关节扭矩120N·m，擅长动态站起、折叠等高灵活性动作；同样搭载3D激光雷达，但感知范围和算力略低于H1；更紧凑（身高1.27m，重35kg），23-43个关节模块化设计，支持力控灵巧手（如焊接、抓取鸡蛋等精细操作）；强化学习驱动，可通过模仿持续进化，适合教育科研场景。
- ❖ 核心技术：
 - ✓创新的电机驱动与结构设计：摒弃传统液压系统，采用自研高性能电机直驱；力矩补偿快速放线算法实现高精度力控；模块化设计。
 - ✓AI驱动的运动控制与智能化：通过强化学习、模仿学习进行动作生成优化和动态平衡控制；舞蹈节奏对齐算法保障表演中的同步性。
 - ✓环境感知与多智能体协同、开源RL框架unitree_rl_gym等。

图表：宇树科技人形机器人产品

