



物理 AI 行业研究

买入（首次评级）

行业深度研究

证券研究报告

机械组

分析师：满在朋（执业 S1130522030002） 分析师：李嘉伦（执业 S1130522060003）

manzaipeng@gjzq.com.cn

lijialun@gjzq.com.cn

物理 AI 奇点时刻到来，设备应该怎么投？

投资逻辑：

物理 AI，为什么是现在？

什么是物理 AI。融合物理规律和 AI 的新型范式，区别于纯基于数字空间的生成式 AI，物理 AI 通过物理约束模型把“智能”落地为“动作”，是 AI 应用进入下半场、打开实体场景的载体。

为什么是现在。模型端走完 0→1：从大语言模型做任务规划的初步尝试，到 VLA 把语义理解从语言延伸到物理空间，再到 $\pi 0$ 至 $\pi 0.7$ 围绕动作生成、数据效率与可靠性持续迭代，配合级联式与分层式两类工程架构，VLA 已从实验室走向真实产线。路线开始分化：除单一路径纵深迭代与系统架构层面分化外，以 Astra 为代表的“模型外部”路线出现，不再训练 VLA，改由通用 Agent 基模通过代码生成与工具调用驱动经典控制器，把机器人控制从“模型问题”变成“系统问题”。成本开始坍塌：不用为每个场景重训模型、不用为新任务重复采集数据、不用自研通用“大脑”。三项成本同时下行，物理 AI 有了从 1 到 10 的机会，我们判断当前是较好布局窗口。

应该投什么？短板就是机会

数据：最核心的瓶颈和机会。根据中国信息通信研究院《具身智能训练场研究报告（2026 年）》信息，假设要让机器人实现“ChatGPT”时刻，约需 21200 台机器人连续工作一年，产出千万小时级有效数据，而当前全球可用的高质量真实数据仅数十万至百万小时级，数据尤其是真机数据将成为物理 AI 最核心的瓶颈也是最好的投资机会。按采集路线划分，无本体（第一视角视频、动捕）与真机（遥操作、自主作业回流）两类方式，将分别对多目相机、IMU、动捕与 EGO 设备、触觉传感器等增量硬件形成拉动，并利好数据生产商、数据基础设施供应商（工具、仿真、资产）与训练场运营方。

感知与执行：高精度灵巧操作仍是硬边界，卖铲子确定性强。传感器/执行器作为物理 AI 落地的必需品具有不可替代的地位，同时其需求也与数据采集环节存在共振。以 Gemini Robotics 2 为例，夹爪类任务平均成功率已达 74.2%—89.6%，而多指灵巧操作成功率却大幅离散（拧下灯泡 92%、拧上灯泡 36%），瓶颈直指触觉与力控硬件。建议关注视觉、触觉、灵巧手三个方向。

再往后怎么看？短板补齐后，物理 AI 将打开全新应用场景

本体不会被跳过，机械臂/人形/非人形机器人均有放量潜力。我们认为本体端除人形机器人值得重视外，从工业场景应用看机械臂/非人形机器人也有较好放量前景，例如移动机械臂+商用 VLA 模型在物流/制造业场景中已经逼近生产级能力。

物理 AI 有望重塑制造业，仓储物流/服装行业或率先落地。随着短板逐步补齐，我们认为机会将转移至应用端，尤其是仓储物流/服装行业由于有大量柔性工序无法通过单纯的工业自动化实现升级有望率先落地。仓储物流领域 Agility/Figure 等企业已经实现了小批量规模部署，操作效率/准确率正在快速提升，具备放量基础；服装为劳动密集型行业机器换人痛点明确，杰克科技在柔性面料处理等高难度环节已具备初步成熟度，规划年产万台机器人产能并已开始获取部分意向订单，后续有望加速放量。

投资建议与估值

物理 AI 奇点时刻到来，建议一方面关注核心短板数据采集/传感器/灵巧手等环节奥比中光/奥普特/雷赛智能等，一方面关注仓储物流/服装机器人场景应用标的杰克科技等。

风险提示

大模型技术迭代不及预期风险/数据采集技术路线变化风险/终端应用不及预期风险。



内容目录

1. 物理 AI，为什么是现在？	6
1.1 什么是物理 AI	6
1.2 拐点为什么是现在，三项成本同时坍塌	7
1.2.1 模型架构经历四阶段发展走向工程化落地，物理 AI 实现 0 到 1	7
1.2.2 三类成本项同时坍塌，物理 AI 的 1 到 10 不再遥不可及	23
2. 应该投什么？短板就是机会	25
2.1 数据：最核心的瓶颈和机会	25
2.2 感知与执行：高精度灵巧操作仍是硬边界，“卖铲子”确定性强	31
3. 再往后怎么看？短板补齐后，物理 AI 将打开全新应用场景	43
3.1 本体不会被跳过，机械臂/人形/非人形机器人均有放量潜力	43
3.2 物理 AI 有望重塑制造业，仓储物流/服装行业或率先落地	45
3.2.1 仓储物流：已实现工业自动化升级，机器人将是第二波	46
3.2.2 服装：主要针对缝纫环节，解决招工难、人工成本高等下游核心痛点	51
4. 投资建议	54
5. 风险提示	55

图表目录

图表 1：物理 AI 概念总览	6
图表 2：生成式仿真利用模型直接学习从系统参数到复杂物理场的高维映射关系	6
图表 3：思维链的加入显著强化了多参数大模型的性能	7
图表 4：直接规划与单一路径任务规划对比	8
图表 5：多路径任务规划推理方式	8
图表 6：闭环任务规划会加入反馈机制	9
图表 7：从大模型落地到机器人行动的流程	9
图表 8：基于大模型控制的救援机器人系统	10
图表 9：谷歌 SayCan 框架给大语言模型加上了可行性接地	10
图表 10：清华大学 Vila 算法采用 GPT-4V 等 VLM 在复杂环境中主动进行任务规划	11
图表 11：RT-1 模型奠定了“视觉-语言-动作”统一建模的基础	11
图表 12：RT-1 架构示意	12
图表 13：RT-2 实现了与互联网规模的视觉-语言数据集联合训练	12
图表 14：RT-2 展示出了涌现推理	13
图表 15：RT-2 在泛化实验上显著优于 RT-1	13



图表 16:	谷歌 RT-H 解决了语义多样性带来的数据共享难题	14
图表 17:	OpenVLA 除语义泛化类别外, 所有评测维度上性能均超过闭源模型 RT-2-X	14
图表 18:	$\pi 0$ 解决动作怎么生成的核心矛盾: 串行离散走向并行连续.....	15
图表 19:	$\pi 0$ 训练数据组	16
图表 20:	$\pi 0$ 开箱评估性能大幅领先	16
图表 21:	经过预训练对全新任务微调即可泛化	16
图表 22:	$\pi 0.5$ 从异构数据迁移实现泛化, 可面向全新场景实现复杂多阶段行为.....	17
图表 23:	$\pi 0.5$ 模型总览	17
图表 24:	$\pi 0.5$ 在全部 4 个家庭任务上显著优于 $\pi 0$	18
图表 25:	$\pi 0.7$ 进一步解决了怎么吃下脏数据的问题, 前提是配上元数据.....	18
图表 26:	VLA 模型快速发展, 新的架构与算法不断涌现.....	19
图表 27:	主流 VLA 模型在 LIBERO 基准上的性能对比	19
图表 28:	三类 VLA 模型架构对比	20
图表 29:	单体式 VLA 模型架构	20
图表 30:	级联式 VLA 模型架构	21
图表 31:	分层式 VLA 模型架构	21
图表 32:	Figure Helix 02 采用三层级联式.....	22
图表 33:	Gemini Robotics 2 为分层式架构.....	22
图表 34:	ER2 相比 ER1.6 性能显著升级.....	23
图表 35:	Astra 三种机器人闭环控制方案.....	23
图表 36:	Agentic Policy 得分最高, 同时大幅高于其他模型.....	24
图表 37:	Astra 的瓶颈不在机器人执行, 而在大模型推理.....	24
图表 38:	后续大模型的性能提升也将进一步提升机器人操作的效果	24
图表 39:	三项成本同时坍塌, 物理 AI 有了从 1 到 10 的机会	25
图表 40:	为什么物理 AI 的投资不是更早	25
图表 41:	“数据飞轮”是物理 AI 产业具有巨大成长性的底层逻辑	26
图表 42:	4 种具身智能数据采集方式示例.....	26
图表 43:	具身智能数据处理流程	27
图表 44:	解决数据瓶颈的常用方法	27
图表 45:	4 种具身智能数据采集比较.....	27
图表 46:	产业正多路线探索真实数据采集技术	27
图表 47:	全球具身智能基建规模与多模态数据集全景对比	28
图表 48:	中国建成的训练场数量分布及统计	29
图表 49:	智元开源数据集把 VLA 算法侧的优化结论变成了数据供应链的工业化版本.....	29
图表 50:	Scale AI 三类数据采集场每天上传 1000+小时示范数据	30



图表 51:	Figure 自建全流程实现分布式数据采集	30
图表 52:	奥比中光无本体数据采集硬件平台	31
图表 53:	诺亦腾动作捕捉解决方案设备	31
图表 54:	索辰科技依托深厚 CAE 物理求解器积累持续加大物理 AI 布局	31
图表 55:	多指灵巧手操作成功率目前依旧偏低	32
图表 56:	人形机器人传感器的主要类型	32
图表 57:	机器视觉系统示意图	33
图表 58:	机器视觉核心功能是识别、测量、定位与检测	33
图表 59:	2024 年中国机器视觉市场中 3D 视觉占比 8.3%	34
图表 60:	机器人视觉感知向 3D 化发展	34
图表 61:	主流厂商短距离 3D 视觉传感器性能对比	34
图表 62:	主流厂商长距离 3D 视觉传感器性能对比	34
图表 63:	主流视觉传感器厂商梳理	35
图表 64:	机器人触觉传感系统的层次功能和结构框图	35
图表 65:	电子皮肤主要分为基底层、导电层、传感层	36
图表 66:	传统的主要触觉传感器类型优缺点介绍	36
图表 67:	压阻式柔性触觉传感器结构及实物图	37
图表 68:	2024 年压阻式触觉传感器市场占比约为 37%	37
图表 69:	帕西尼的 DexH13 多维触觉灵巧手结合了视觉和触觉	37
图表 70:	帕西尼 ITPU GEN2 触觉传感器在多项指标技术领先	38
图表 71:	2023 年全球柔性触觉传感器行业集中度高	38
图表 72:	触觉传感器代表厂商梳理	39
图表 73:	机器人灵巧手的主要构成	40
图表 74:	机器人灵巧手主要驱动方式分类和适用场景	40
图表 75:	空心杯电机采用无齿槽结构	41
图表 76:	有齿槽电机及无齿槽电机结构图	41
图表 77:	机器人灵巧手主要传动方式分类	41
图表 78:	腱绳传动灵巧手整体机构	42
图表 79:	腱绳+丝杠复合传动（单自由度视图）	42
图表 80:	特斯拉 OptimusGen2 灵巧手搭载触觉传感器轻松拿取鸡蛋	42
图表 81:	工业场景中工业机器人+大模型打造的工作站或将成更直接的物理 AI 落地场景	43
图表 82:	工业机器人目前在较多细分行业的覆盖场景和应用深度有限	43
图表 83:	移动机械臂已经有几十年的迭代历史	44
图表 84:	Franka Robotics Mobile FR3 Duo	45
图表 85:	搭载商用 VLA 模型的移动机械臂已经逼近了生产级能力	45



图表 86: 物理 AI 已经开始重塑制造、物流等工业领域	45
图表 87: 物理 AI 落地四阶段框架	46
图表 88: 物流行业整体已经有了较高的自动化程度	47
图表 89: 兰剑智能壁虎料箱机器人系统优化货到人拣选	47
图表 90: 德马科技智能播种机器人缩减二次分拣人工作业	47
图表 91: 拣选尤其是拆零拣选是人力耗费最大的环节	47
图表 92: 美国仓储、物流和制造行业有大量材料搬运岗位空缺	48
图表 93: Agility 人形机器人从物流领域起步逐步向其他行业拓展	48
图表 94: Figure Helix 进行了一系列通用控制优化满足物流行业需求	49
图表 95: Figure 02 在宝马工厂真实产线测试数据	49
图表 96: 星动纪元实现单型号本体供包/打包/拣选全链路作业	50
图表 97: 自变量机器人物流分拣效率/准确率表现优异	50
图表 98: 德马科技打造物流场景应用与数采同步布局	51
图表 99: 缝中环节是服装厂实现全面自动化的瓶颈	51
图表 100: 目前服装厂可以实现缝前、缝后的高自动化, 缝中主要依靠人工	52
图表 101: 模板机自动化程度更高, 但上下料还是需要人工操作	52
图表 102: 布料属于接触丰富型柔性物体, 机器人的应用具有一定难度	53
图表 103: 杰克科技服装机器人工厂已经正式投产	53
图表 104: 杰克在部分场景合格率已经可以达到 100%人工水平、效率可达人工 50%以上	54
图表 105: 杰克与恒锋服饰签订 200 台意向订单	54
图表 106: 杰克与浙江人形签约 2000 台意向订单	54
图表 107: 建议关注公司	54

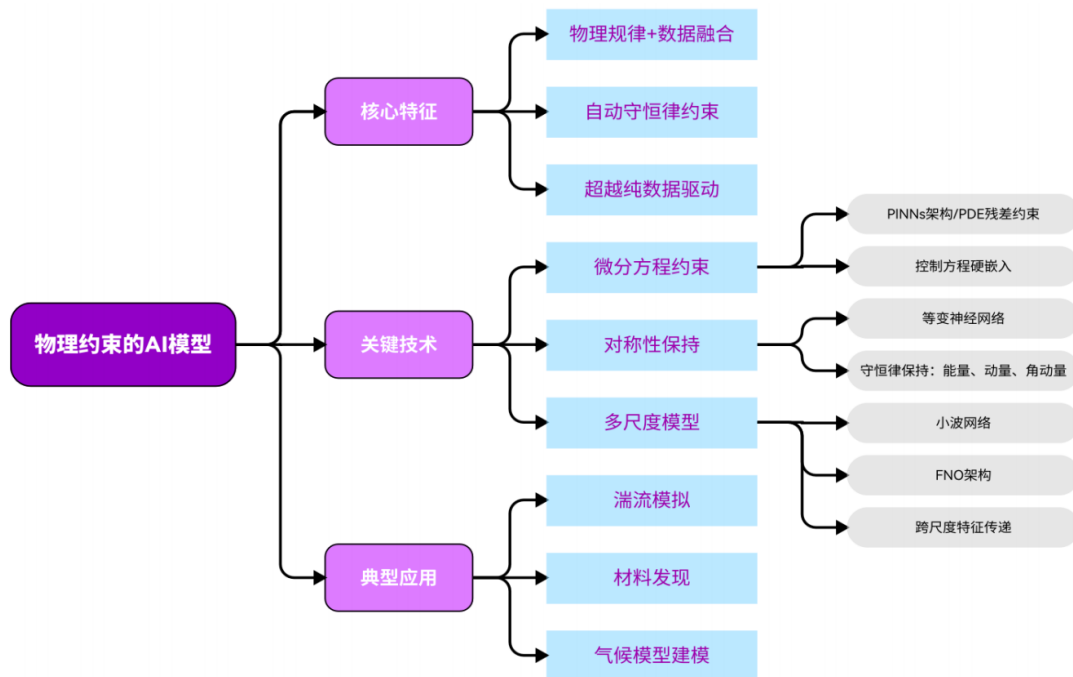


1.物理 AI，为什么是现在？

1.1 什么是物理 AI

物理 AI 是一种融合物理规律和人工智能的新型计算范式，例如其中一类重要路径源自物理约束机器学习（PIML）作为物理 AI 的核心理论框架，通过将物理规律转化为数学模型约束，实现了机器学习与物理原理的深度融合。

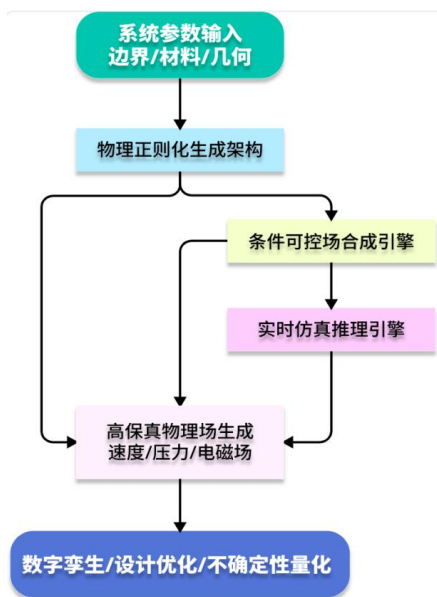
图表1：物理 AI 概念总览



来源：《基于物理 AI 的工业仿真新型范式研究》国金证券研究所

区别于纯数字空间的生成式 AI，核心在于把“智能”落地为“动作”，例如，英伟达提出的物理 AI 解决方案 Omniverse 平台，延续了工业仿真中以模型替代实物试验的思路。Omniverse 平台构建了一套完整的物理 AI 解决方案，采用“仿真驱动”范式，通过虚拟世界生成海量训练数据并迁移至现实系统，显著降低机器人及自动驾驶的研发成本与风险，体现了物理 AI 在具身智能领域的工程化落地路径。

图表2：生成式仿真利用模型直接学习从系统参数到复杂物理场的高维映射关系





来源：《基于物理 AI 的工业仿真新型范式研究》，国金证券研究所

1.2 拐点为什么是现在，三项成本同时坍塌

1.2.1 模型架构经历四阶段发展走向工程化落地，物理 AI 实现 0 到 1

一阶段：单纯基于 LLM 完成任务规划

随着大语言模型参数量的增加，其推理能力显著增强，尤其是针对参数量较大的模型（比如超过 1000 亿参数量），通过加入思维链技术可以显著提升模型性能，通过模拟人类思维的过程，逐步分解复杂任务，思维链技术使模型能够更好地理解问题并生成合乎逻辑的答案，这一点可以优先弥补传统机器人任务规划方法在处理复杂任务时的不足。

图表3：思维链的加入显著强化了多参数大模型的性能

模型输入	模型输入
问: 罗杰有5个网球。他又买了2罐网球。每罐有3个网球。他现在有多少个网球?	问: 罗杰有5个网球。他又买了2罐网球。每罐有3个网球。他现在有多少个网球?
答: 答案是11。	答: 罗杰一开始有5个球。2罐3个网球每个是6个网球。 $5+6=11$ 。答案是11。
问: 食堂有23个苹果。如果他们用了20个做午饭,然后又买了6个,还有多少个苹果?	问: 食堂有23个苹果。如果他们用了20个做午饭,然后又买了6个,还有多少个苹果?
模型输出	模型输出
答: 答案是27。	答: A: 食堂原来有23个苹果。他们用了20个做午饭。所以他们有 $23-20=3$ 。他们又买了6个苹果,所以他们有 $3+6=9$ 。答案是9。
错误	正确

来源：《大语言模型在机器人任务规划中的应用与展望》，国金证券研究所

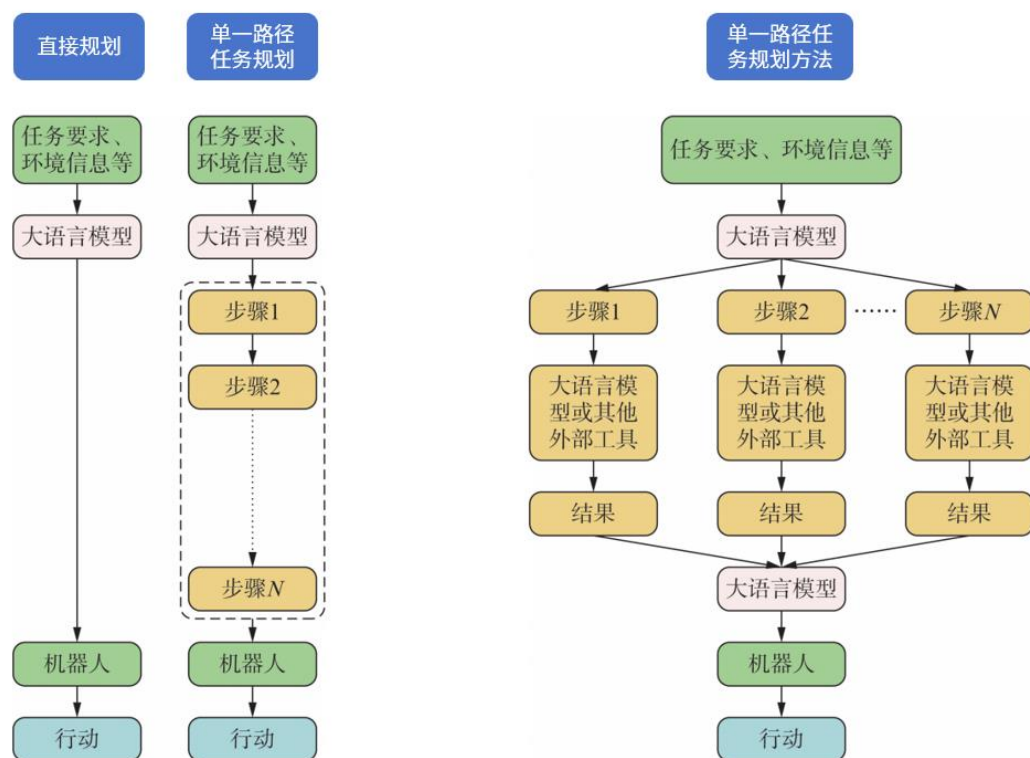
通过使用大语言模型，首先可以以开环方式进行机器人的任务规划，机器人首先制定一系列计划，随后在实际执行过程中依次逐步推进。在此过程中，大语言模型需根据当前任务要求及环境信息生成相应计划，并进一步自动生成相关代码，以便将计划有效部署至机器人系统中，其中包含从直接任务规划、单一路径任务规划、多路径任务规划等方式：

直接任务规划：大语言模型一次性制定完整的任务计划，机器人依据大语言模型一次性生成的任务计划执行操作，无需在执行过程中进行额外的逐步推理。其优势在于任务生成阶段省略了连续的推理步骤，从而能够更加迅速地制定出完整的任务计划。但可能无法充分考虑所有相关因素及潜在变量，导致生成的计划不够精确甚至出现错误，使机器人在实际操作过程中出现行动失误甚至设备损坏等情况。

单一路径任务规划：大语言模型基于确定的线性路径推导出结论的方法，该方法采用类似于思维链通过清晰展示每个中间步骤和逻辑推理过程，提高大语言模型处理复杂任务的可解释性，且在复杂规划任务中表现出优越的性能。



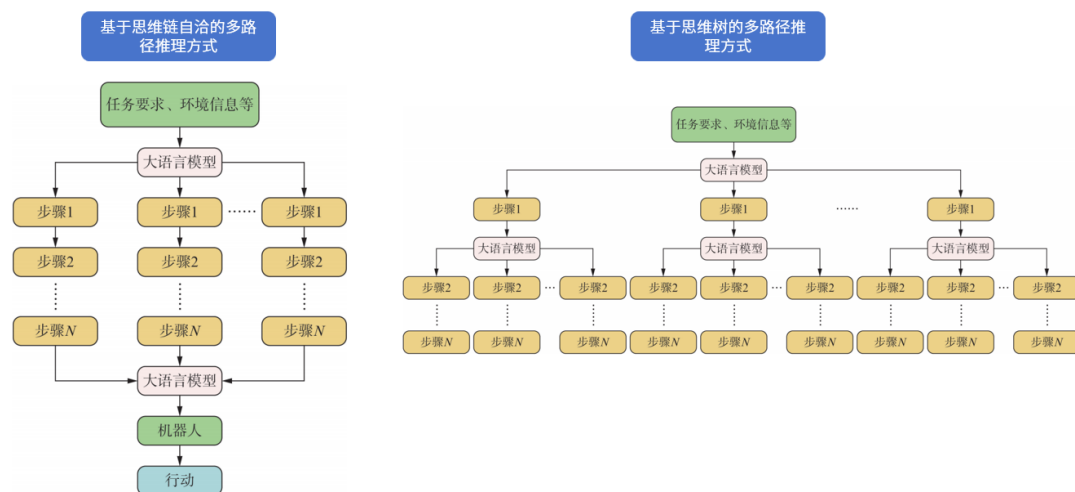
图表4：直接规划与单一路径任务规划对比



来源：《大语言模型在机器人任务规划中的应用与展望》，国金证券研究所

多路径任务规划：大语言模型可以通过生成与分析树状推理路径,探索多种可能的解决方案.其优势在于它允许模型从不同的视角来分析问题,从而提高解决方案的全面性和准确性。

图表5：多路径任务规划推理方式

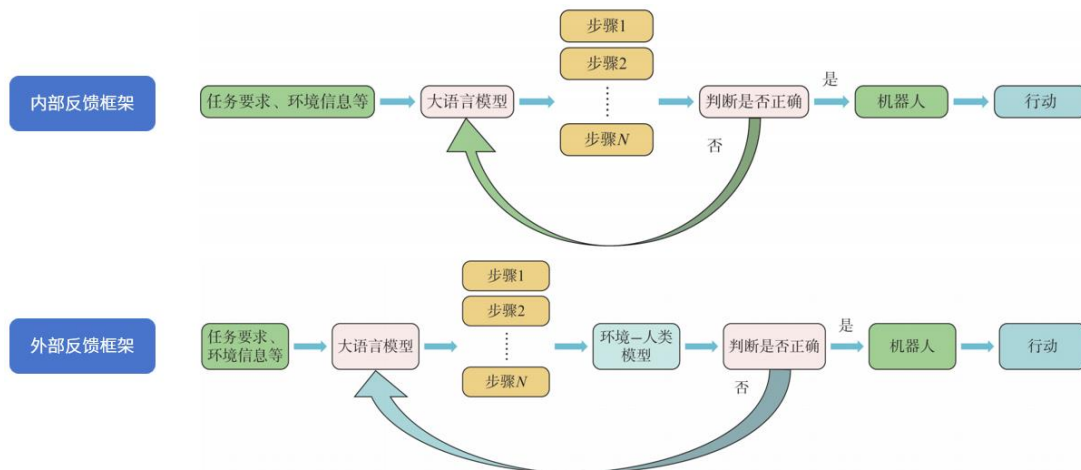


来源：《大语言模型在机器人任务规划中的应用与展望》，国金证券研究所

闭环任务规划开始需要考虑根据环境反馈来进行实时调整和任务优化,与单一路径任务规划不同,闭环路径规划不仅在初始阶段生成路径,还在任务执行过程中持续监督和调整,以应对环境变化及突发情况.其加入的反馈机制主要包括通过大模型自身的学习与调整机制来迭代的内部反馈,和通过基于用户输入、环境变量等实现的外部反馈。



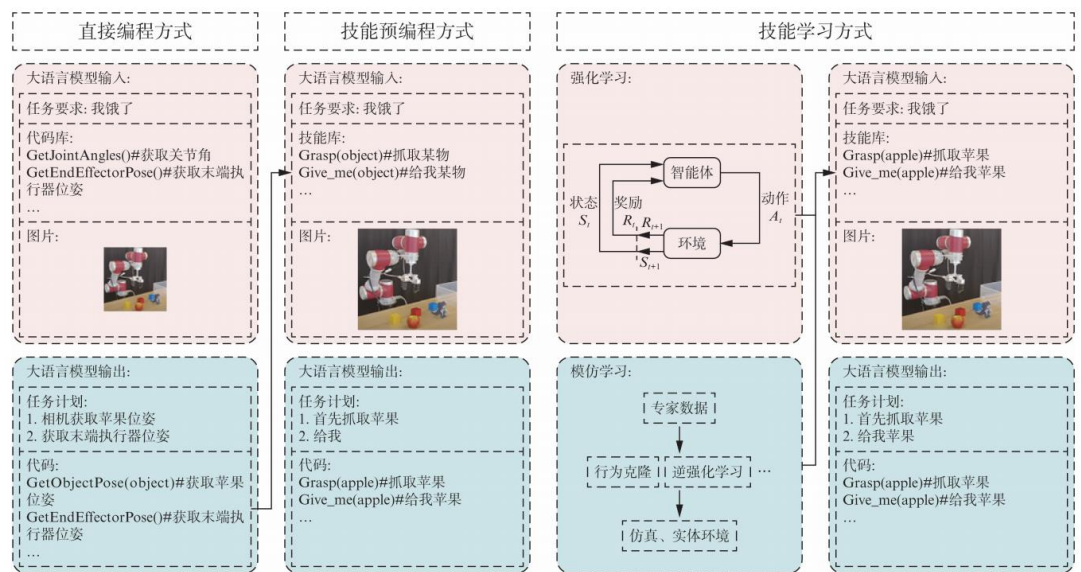
图表6：闭环任务规划会加入反馈机制



来源：《大语言模型在机器人任务规划中的应用与展望》，国金证券研究所

最终大模型可以根据给定的任务要求、环境条件及其他相关信息自动生成相应的代码，通过调用各类算法和函数，指导机器人执行预定的操作任务，例如导航、避障、物体抓取或路径规划等。若采用技能预编程、技能学习可以进一步提升任务泛化能力，简化机器人任务规划与执行的流程。

图表7：从大模型落地到机器人行动的流程

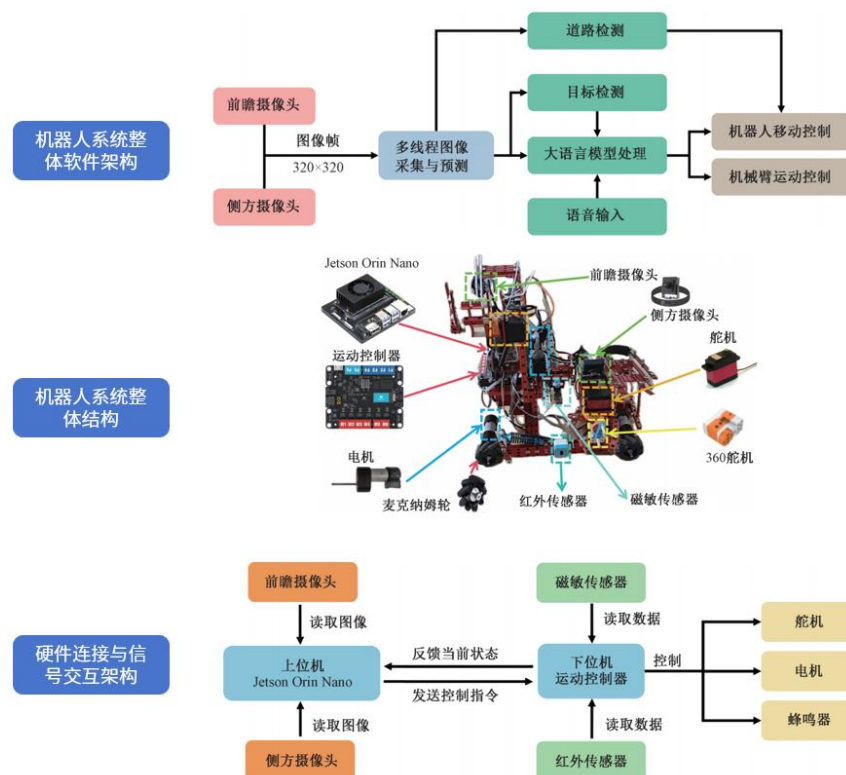


来源：《大语言模型在机器人任务规划中的应用与展望》，国金证券研究所

参考《基于大语言模型的救援机器人机械臂控制系统设计》中的案例，实现了“云端 ERNIE 大模型语义解析+Jetson 边缘视觉伺服+PID 闭环控制”的救援机器人系统，将模糊口语指令端到端映射为结构化控制指令，定位误差 $\leq 5\text{mm}$ ，综合任务完成率 $\geq 95\%$ 。该案例证明：LLM 已从实验室的“任务规划器”走进了集目标检测 (mAP50=0.963)、轨迹规划与实时闭环控制于一体的完整工程系统——大模型驱动的控制链路已具备工程可用性。



图表8：基于大模型控制的救援机器人系统



来源：《基于大语言模型的救援机器人机械臂控制系统设计》，国金证券研究所

谷歌 SayCan 框架给大语言模型加上了可行性接地。LLM 提供“应该做什么”（Say），基于 RL 预训练的 affordance 价值函数提供“当前能做什么”（Can），两个分数相乘选出可执行的最优技能。在真实厨房中评估 101 个指令，PaLM-SayCan 实现 84%规划成功率/74%执行成功率。

图表9：谷歌 SayCan 框架给大语言模型加上了可行性接地

单纯的 LLM 模型未与自身所处环境进行交互，无法给出现实直接有效的解决方案。

SayCan 通过预训练价值函数为 LLM 模型奠定基础，让机器人能够执行现实世界中抽象的、长期的指令。



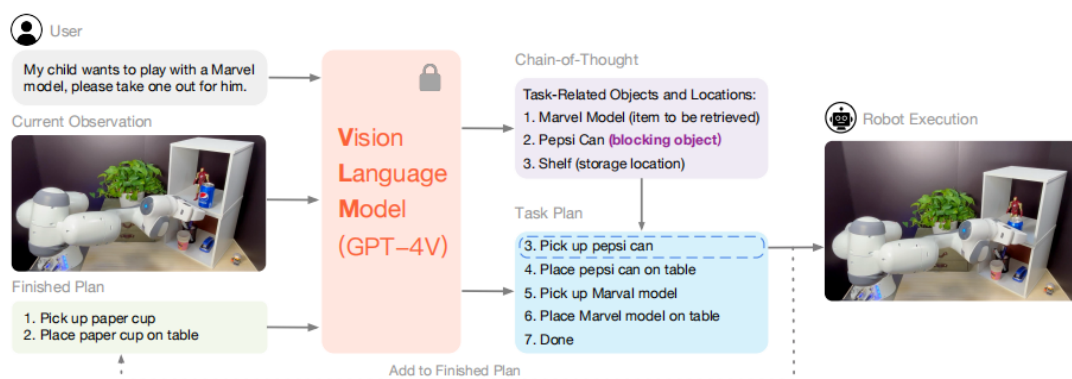
来源：《Do As I Can, Not As I Say: Grounding Language in Robotic Affordances》，国金证券研究所

二阶段：VLA 诞生，语义推理从语言穿透到物理空间

多模态模型指一个模型能够接受输入不同“模态”，例如图像、文本或音频信号。VLM（视觉语言模型）同时融合了 LLM 和 VFM，是一种同时接受图像和文本的多模态模型，同时接收图像和语言作为输入，并根据语言指令和图像信息产生输出，完成图像问答任务，实现更准确的任务规划和决策。例如清华大学交叉信息院高阳研究组提出 ViLa 算法，通过利用 GPT-4V 这类 VLM，使机器人能够在复杂环境中主动进行任务规划。



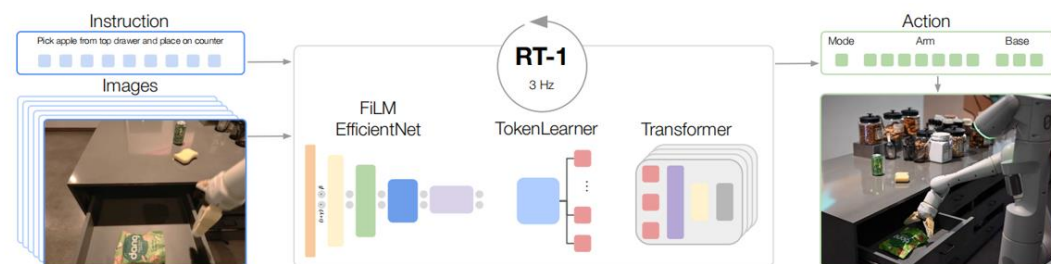
图表10: 清华大学 Vila 算法采用 GPT-4V 等 VLM 在复杂环境中主动进行任务规划



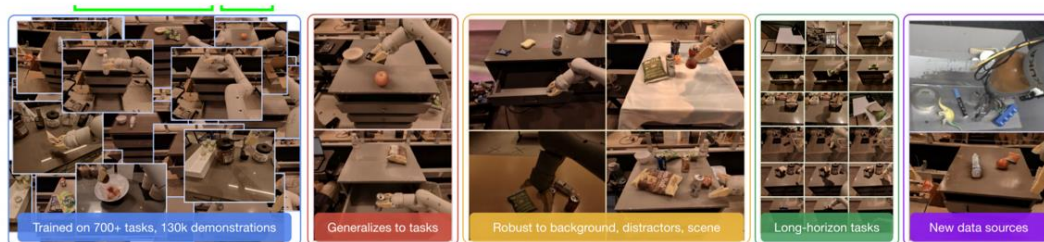
来源:《Look Before You Leap: Unveiling the Power of GPT-4V in Robotic Vision-Language Planning》, 国金证券研究所

谷歌 RT-1 模型用 13 万条真机演示、700+任务训练了 3500 万参数 Transformer 策略, 首次证明了 Transformer 可以规模化地将多模态输入(图像+语言指令)直接映射为机器人动作, 并在 700 多个任务上实现了泛化, 奠定了“视觉-语言-动作”统一建模的基础。

图表11: RT-1 模型奠定了“视觉-语言-动作”统一建模的基础



RT-1 接收图像与自然语言指令作为输入, 输出离散化的底座移动与机械臂动作。尽管模型参数量仅 3500 万, 凭借高效且高容量的架构设计, 其推理频率可达 3 赫兹



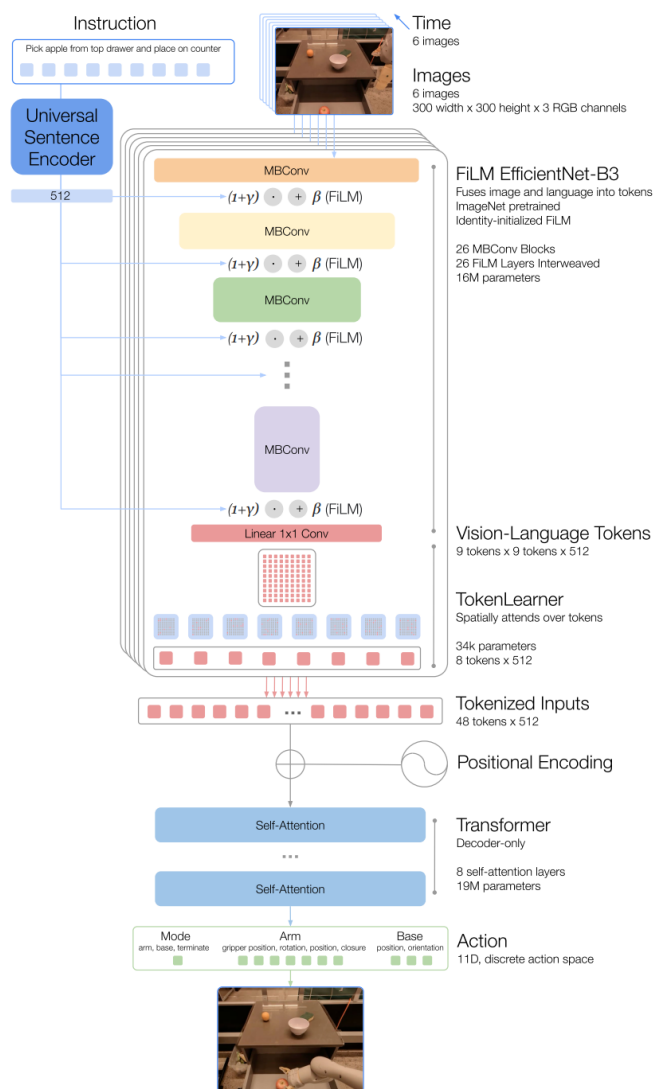
RT-1 经过大规模真实场景训练(13 万条演示数据)与实测评估(3000 次真实环境测试), 展现出优异的泛化能力、鲁棒性, 以及从多源异构数据中学习的能力

来源:《RT-1: ROBOTICS TRANSFORMER FOR REAL-WORLD CONTROL AT SCALE》, 国金证券研究所

指令文本被转换为通用句子编码器(USE)嵌入向量, 后通过特征线性调制(FiLM)层对预训练 EfficientNet 进行条件调制。得到的视觉-语言 Token 经 TokenLearner 做降维处理, 送入仅解码器的 Transformer, 最终把机器人动作离散编码为 Token 序列。



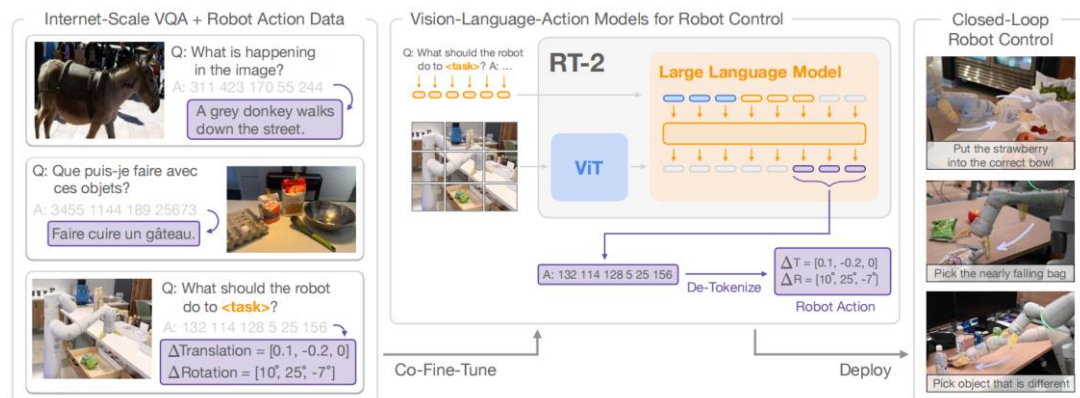
图表12: RT-1 架构示意



来源:《RT-1: ROBOTICS TRANSFORMER FOR REAL-WORLD CONTROL AT SCALE》, 国金证券研究所

谷歌 RT-2 首次直接提出了“VLA” (Vision-Language-Action Model) 这一概念, 在 RT-1 的基础上进行了革命性升级, 将机器人动作离散化为文本 token, 与 VQA 等互联网任务联合微调, 使 VLM 输出语言 token 的能力直接迁移为输出动作 token。这一简单的技巧使互联网规模的视觉语言预训练直接“注入”了低层机器人控制。

图表13: RT-2 实现了与互联网规模的视觉-语言数据集联合训练

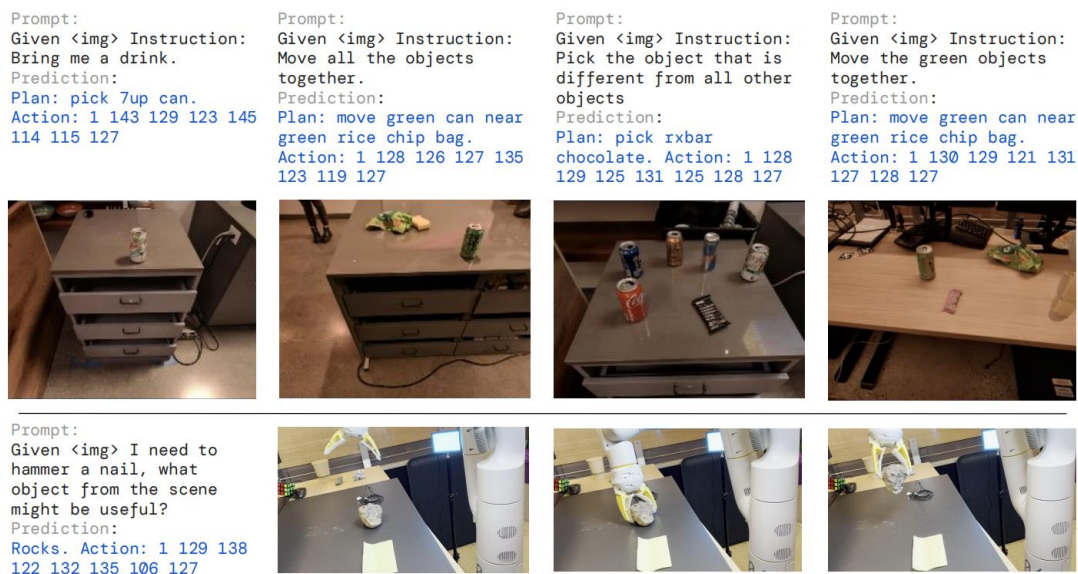




来源：《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，国金证券研究所

RT-2 将语义推理第一次从语言空间穿透到物理动作空间，同时展示出了涌现推理。例如训练数据中从未出现的“拾取最小的物体”或“将物体放到数字 3 上”等指令能被正确执行。结合 Chain-of-Thought，甚至能判断“哪种物体可用作临时锤子（石头）”。

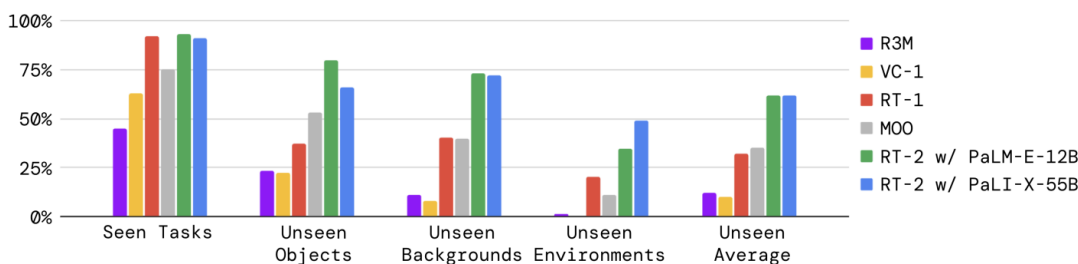
图表14：RT-2 展示出了涌现推理



来源：《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，国金证券研究所

通过和其他模型尤其是 RT-1 进行对比，可以看到 RT-2 在未出现过任务的泛化实验上差距最大，说明 VLA 模型优势在于可以从互联网规模预训练数据中迁移泛化能力更强的视觉与语义概念。

图表15：RT-2 在泛化实验上显著优于 RT-1



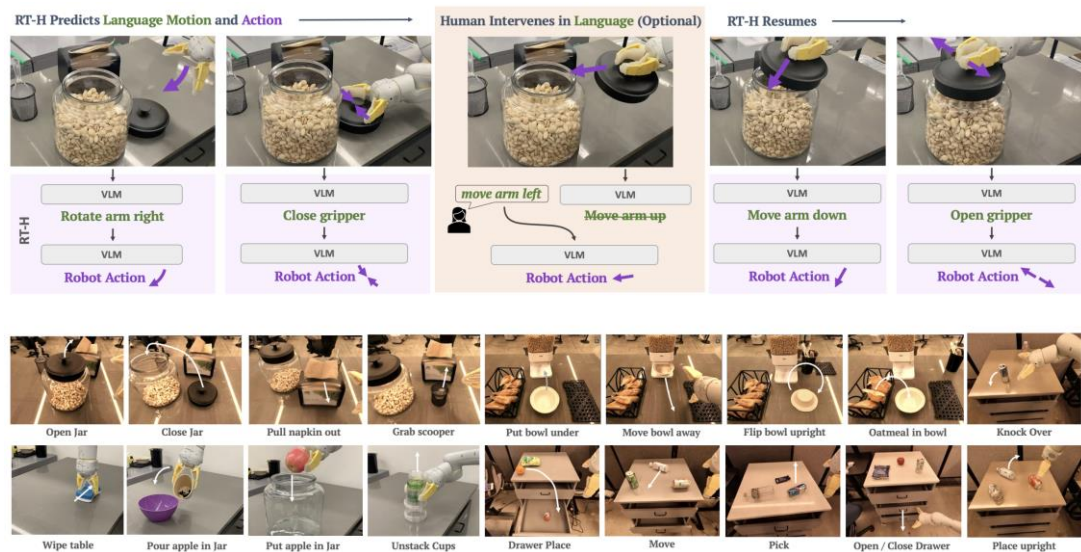
来源：《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，国金证券研究所

RT-H 进一步引入了语言动作 (Language Motions) 作为中间层，构筑了一个动作层级结构 (Action Hierarchy)，从而提升了模型在多样化任务上的性能、泛化能力和人机交互的灵活性。

RT-H 解决了语义多样性带来的数据共享难题，虽然高层任务不同，但它们的底层动作 (如“向前移动手臂”、“闭合夹爪”) 可能是相似的。因此，通过学习这些共享的、细粒度的“语言动作”作为中间表示，可以更好地在多样化任务间复用数据。



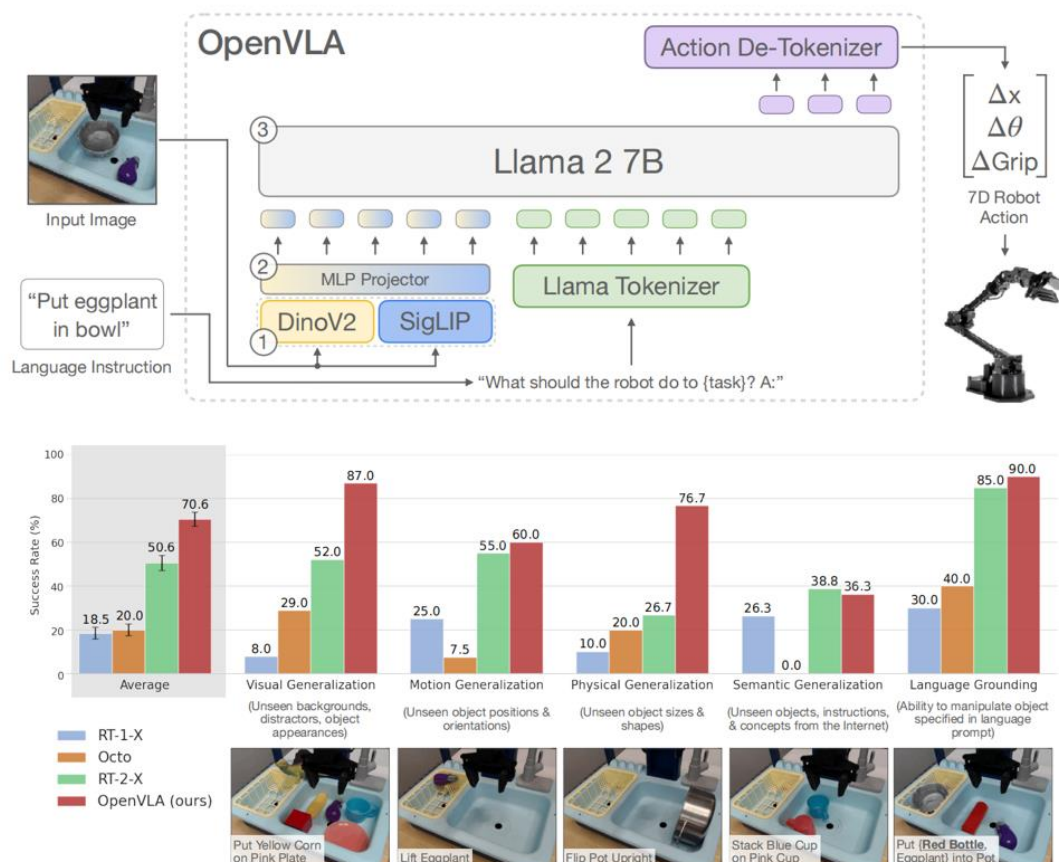
图表16: 谷歌 RT-H 解决了语义多样性带来的数据共享难题



来源:《RT-H: Action Hierarchies Using Language》, 国金证券研究所

OpenVLA 实现 7B 开源 VLA 超越 55B 闭源。基于 Llama-2+DINOv2/SigLIP 视觉编码器, 在 97 万条真机轨迹上训练 7B 参数开源 VLA, 以 1/7 参数量超越 RT-2-X, 并提供 LoRA 微调与量化部署方案——成为开源 VLA 的事实基线。

图表17: OpenVLA 除语义泛化类别外, 所有评测维度上性能均超过闭源模型 RT-2-X



来源:《OpenVLA: An Open-Source Vision-Language-Action Model》, 国金证券研究所

三阶段: 从 $\pi 0$ 到 $\pi 0.7$, 动作生成/数据采集/可靠性等核心矛盾被逐步解决

我们认为在第二阶段的发展中, OpenVLA 完成了开源化的收官, 但其离散动作 token 范式

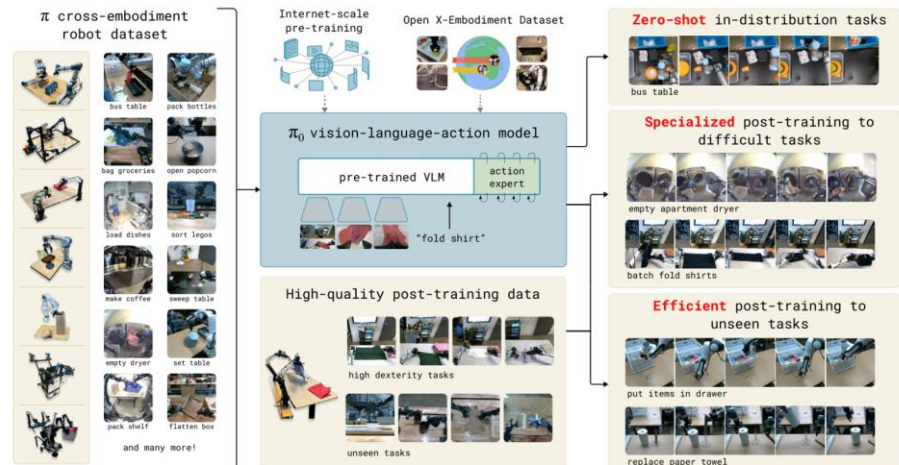


的天花板较低。其动作生成方式本身，自回归逐 token 串行输出使控制频率最高仅约 6Hz，远低于机械臂 50Hz 的实时闭环需求；动作被离散到 256 个桶造成精度损失；且仅接收单帧图像、缺失本体状态与历史记忆。

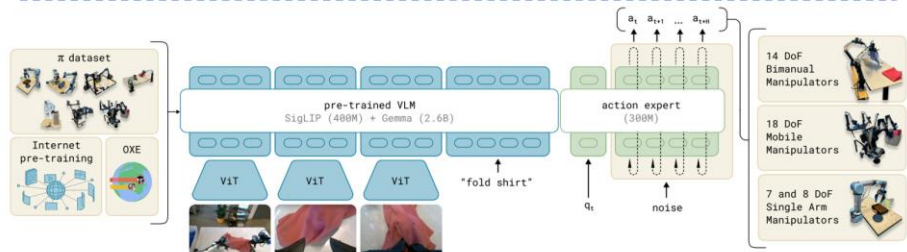
$\pi 0$ 则实现了接近 50hz 的高频控制，同时完成了三次跃迁，成为 VLA 技术演进的分水岭：

1) 架构跃迁：从串行生成到并行动作块。 $\pi 0$ 以 PaliGemma 为视觉-语言主干，并在其上追加一组同构但独立的权重作为动作专家，两者实现单一 Transformer、两套权重的 MoE 式结构，通过条件流匹配对连续动作分布建模，一次性输出长度 $H=50$ 的动作块，从而将控制频率提升至最高 50Hz，因而特别适合高频灵巧任务。

图表18: $\pi 0$ 解决动作怎么生成的核心矛盾：串行离散走向并行连续



采用预训练视觉-语言模型 (VLM) 作为主干网络，并使用包含各类灵巧操作任务的跨本体多样化数据集。通过增设独立的动作专家模块适配机器人控制，该模块基于流匹配生成连续动作，实现精准、流畅的操作能力。模型既可以依据提示词执行任务，也可基于高质量数据做微调，从而完成复杂多阶段任务，例如叠多件衣物、组装纸箱。



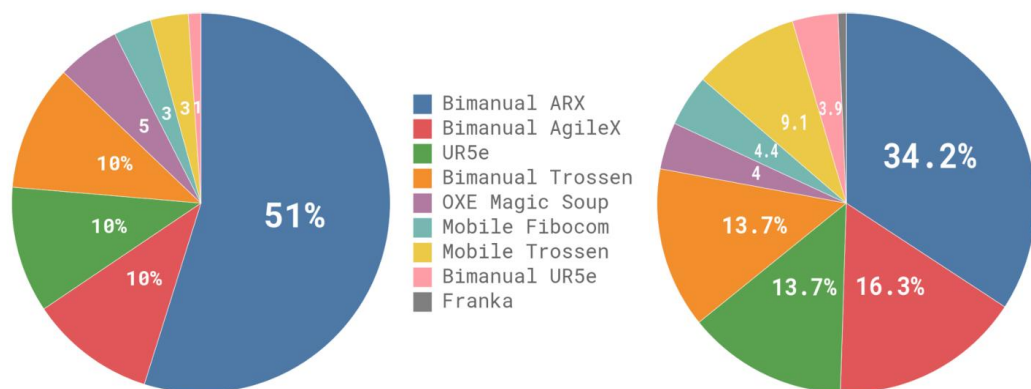
训练使用混合预训练数据集，包含自研灵巧操作数据集与开源公开数据。基于该混合数据集训练流匹配 VLA 模型；该模型由规模较大的 VLM 主干网络，以及规模较小、用于处理机器人状态与动作的动作专家模块两部分构成。VLM 主权重由 PaliGemma 初始化，复用互联网大规模预训练学到的表征能力。最终得到的模型可控制多种动作空间互不相同的机器人本体，完成种类丰富的任务

来源：《 $\pi 0$: A Vision-Language-Action Flow Model for General Robot Control》，国金证券研究所

2) 数据跃迁：规模上了一个量级。 $\pi 0$ 的自采数据达 903M 时间步（单臂 106M+双臂 797M）、68 类任务；完整预训练混合约 10000 小时，覆盖 7 种机器人构型与 68 类任务，并叠加 OXE (22 种本体)、DROID、Bridge 等开源数据，开源部分占混合的 9.1%。



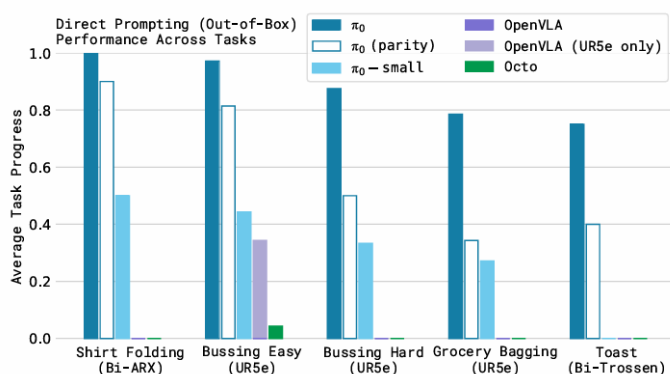
图表19: $\pi 0$ 训练数据组



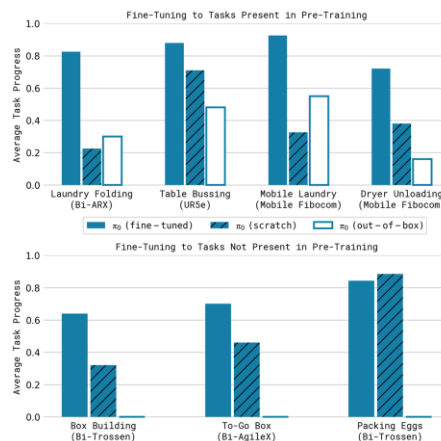
来源:《 $\pi 0$: A Vision-Language-Action Flow Model for General Robot Control》, 国金证券研究所

3) 训练范式跃迁: 系统移植预训练/后训练两段式。对应大语言模型的预训练/后训练分离, 预训练使用规模大但质量较低、行为多样的数据, 使模型获得广泛的物理能力与错误恢复能力, 后训练再以少量高质量、策略一致流畅的精选数据微调, 使执行熟练流畅。可以看到 $\pi 0$ 在开箱评估中性能大幅领先, 尤其是经过预训练的 $\pi 0$ 机器人模型, 在下游任务微调时, 相比从零开始训练有巨大优势; 哪怕是预训练里完全没见过的新任务, 它也能快速学会。

图表20: $\pi 0$ 开箱评估性能大幅领先



图表21: 经过预训练对全新任务微调即可泛化



来源:《 $\pi 0$: A Vision-Language-Action Flow Model for General Robot Control》, 国金证券研究所

来源:《 $\pi 0$: A Vision-Language-Action Flow Model for General Robot Control》, 国金证券研究所

$\pi 0.5$ 最大的进步则是改写了数据配方, 例如其第一阶段 97.6% 的训练样本不来自实际执行任务的场景, 而是来自其他机器人、实验室任务、子任务标注、人类语言指令和网络多模态数据。也就是说, 目标场景的真机数据 2.4%, 作为解决真机数据不足的方法之一, 主要将异质数据把泛化能力嫁接过来。



图表22: $\pi 0.5$ 从异构数据迁移实现泛化, 可面向全新场景实现复杂多阶段行为

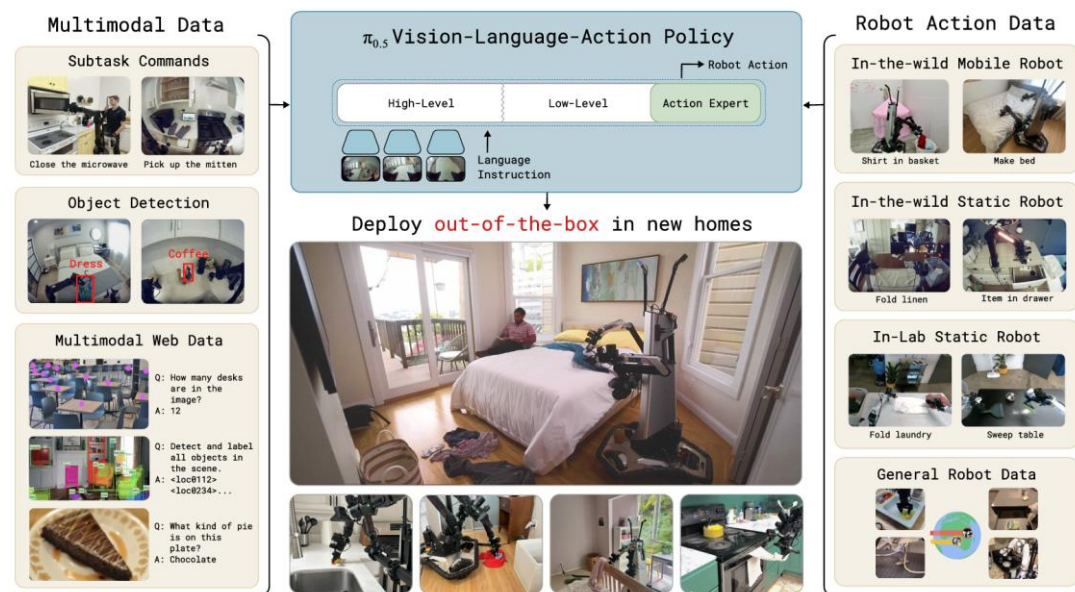
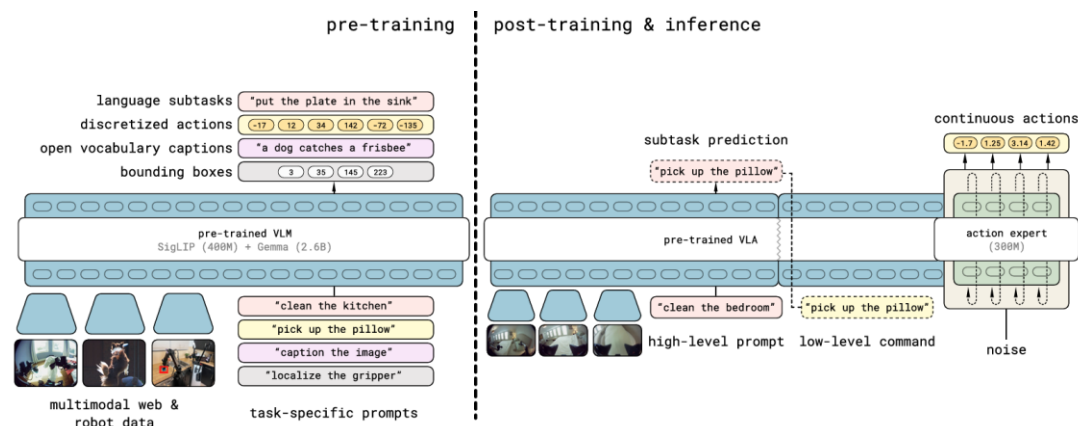


Fig. 1: The $\pi_{0.5}$ model transfers knowledge from a heterogeneous range of data sources, including other robots, high-level subtask prediction, verbal instructions, and data from the web, in order to enable broad generalization across environments and objects. $\pi_{0.5}$ can control a mobile manipulator to clean kitchens and bedrooms in new homes that were not present in the training data, performing complex multi-stage behaviors with durations of 10 to 15 minutes.

来源:《 $\pi 0.5$: a Vision-Language-Action Model with Open-World Generalization》, 国金证券研究所

$\pi 0.5$ 分为两个训练阶段, 第一阶段为预训练: 融合全部异构数据源, 得到基于离散 token 的初始 VLA 模型。该阶段使用多类机器人平台数据、高层语义动作预测数据以及互联网数据; 机器人动作数据通过 FAST 动作分词器将动作编码为离散 token。第二阶段为后训练: 利用与任务高度相关的数据 (含人类监督者的语言指令), 让模型专门适配移动操控任务的低层、高层推理。该阶段采用流匹配 (flow matching) 建模动作分布, 既支持高效实时推理, 又可以表征细粒度的连续动作序列。推理时, 模型先输出高层子任务规划, 再基于该子任务预测具体动作。

图表23: $\pi 0.5$ 模型总览

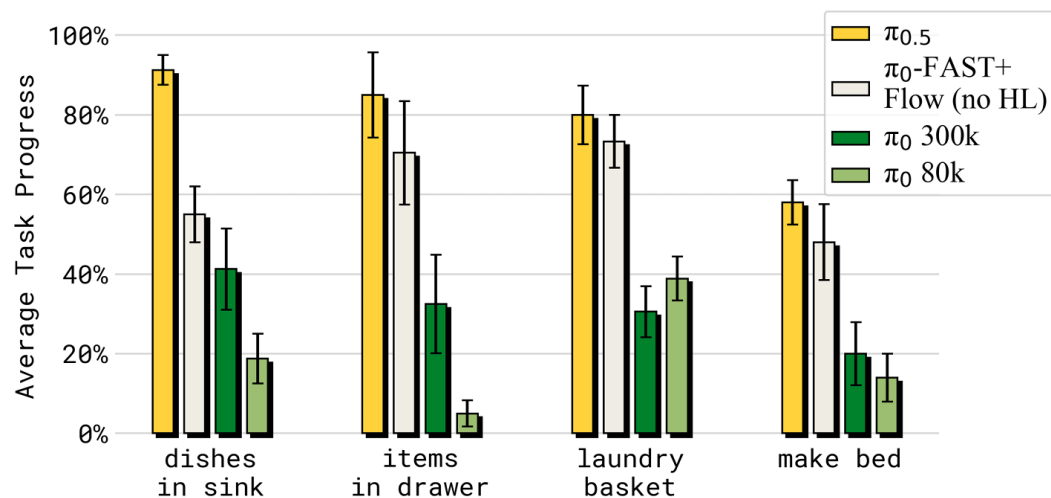


来源:《 $\pi 0.5$: a Vision-Language-Action Model with Open-World Generalization》, 国金证券研究所

通过加入高层子任务监督+网页数据泛化, $\pi 0.5$ 在四个家庭任务中均超越了 $\pi 0$, 即使是同样采用 FAS



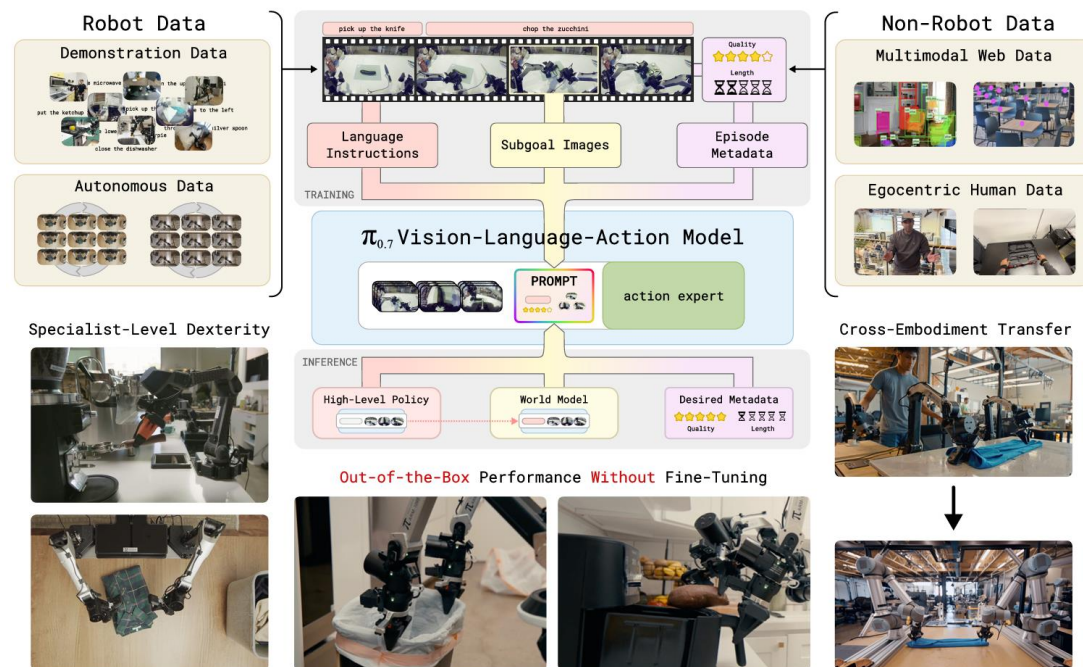
图表24: $\pi 0.5$ 在全部 4 个家庭任务上显著优于 $\pi 0$



来源:《 $\pi 0.5$: a Vision-Language-Action Model with Open-World Generalization》, 国金证券研究所

$\pi 0.7$ 使用多样化提示进行训练, 提示信息不仅包含任务描述, 还包含详细文本、生成的子目标图像以及轨迹元数据。这为模型提供了更丰富的上下文, 既明确要做什么, 也说明该怎么做, 让 $\pi 0.7$ 能够充分利用机器人与非机器人两类海量数据, 并以全新方式组合已有技能, 完成未知新任务。但前提是脏数据需要配上元数据, 单纯的加入更多低质量数据时 $\pi 0.7$ 的形象反而会下降, 决定模型能力的不是数据小时数, 而是每条数据带了多少可用的上下文。

图表25: $\pi 0.7$ 进一步解决了怎么吃下脏数据的问题, 前提是配上元数据

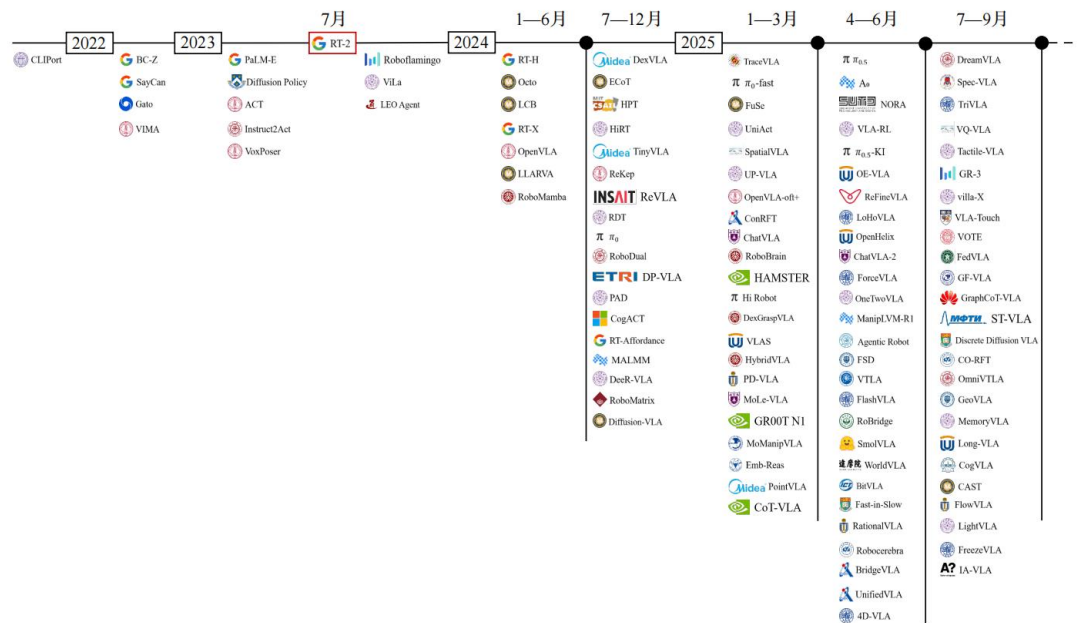


来源:《 $\pi 0.7$: a Steerable Generalist Robotic FoundationModel with Emergent Capabilities》, 国金证券研究所

四阶段: 模型架构走向工程化落地

VLA 模型当前正处于前所未有的快速发展阶段, 新的架构与算法不断涌现。

图表26: VLA 模型快速发展, 新的架构与算法不断涌现



来源：《视觉-语言-动作模型研究综述：迈向通用机器人》，国金证券研究所

可以看到 VLA 模型性能的提升不再单纯依赖参数规模的扩张,而是开始更多地关注架构和推理机制的改进,例如在并行解码方向上,OpenVLA-OFT 已将动作生成速度提升 26 倍,突破自回归瓶颈。在轻量化上,SmoI VLA 仅以主流模型约 1/15 参数量实现 88.8%平均成功率——证明"小模型亦能承载 VLA"。

图表27: 主流 VLA 模型在 LIBERO 基准上的性能对比

VLA 模型	参数量	Spatial/%	Object/%	Goal/%	Long/%	Average/%
DP	$\sim 300 \times 10^6$	78.3	92.5	68.3	50.5	72.4
Octo	93×10^6	78.9	85.7	84.6	51.1	75.1
OpenVLA	7×10^9	84.7	88.4	79.2	53.7	76.5
OpenVLA-OFT	7×10^9	96.2	98.3	96.2	90.7	95.3
π_0	3.3×10^9	96.8	98.8	95.8	85.2	94.2
π -FAST	3.3×10^9	96.4	96.8	88.6	60.2	85.5
CogACT	$\sim 7.3 \times 10^9$	97.2	98	90.2	88.8	93.2
SpatialVLA	7×10^9	88.2	89.9	78.6	55.5	78.1
TraceVLA	7×10^9	84.6	85.2	75.1	54.1	74.8
CoT-VLA	7×10^9	87.5	91.6	87.6	69	81.1
NORA	3×10^9	92.2	95.4	89.4	74.6	87.9
BitVLA	2×10^9	97.4	99.6	94.4	87.6	94.8
SmolVLA	450×10^6	93	94	91	77	88.8
WorldVLA	7×10^9	87.6	96.2	83.4	60	81.8
UniVLA	7×10^9	95.4	98.8	93.6	94	95.5
4D-VLA	7×10^9	88.9	95.2	90.9	79.1	88.6
TriVLA	$\sim 2 \times 10^9$	91.2	93.8	89.8	73.2	87
MemoryVLA	7×10^9	98.4	98.4	96.4	93.4	96.5
FlowVLA	8.5×10^9	93.2	95	91.6	72.6	88.1

来源：《视觉-语言-动作模型研究综述：迈向通用机器人》，国金证券研究所

随着 VLA 模型性能的不断升级, 其架构设计开始很大程度上决定模型的性能表现, 根据《视觉-语言-动作模型研究综述: 迈向通用机器人》中的分类, 现有工作模式可以归纳为



三类单体式架构、级联式架构、分层式架构。

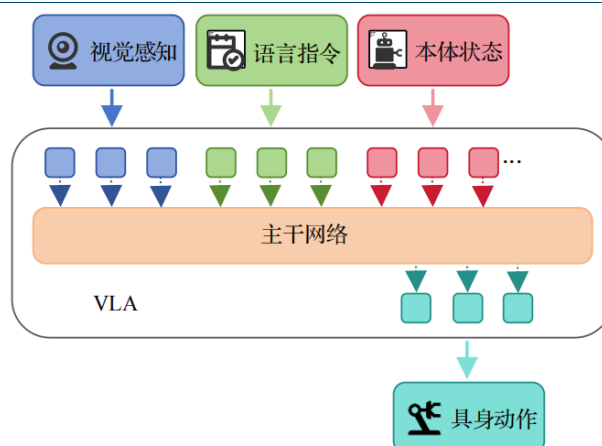
图表28：三类 VLA 模型架构对比

架构类型	设计思路	代表模型	推理速度	适用场景
单体式	统一 Transformer 序列建模，视觉/语言/动作一个模型出	RT-2、OpenVLA、Octo	慢（自回归，1-10Hz）	语义泛化任务
级联式	高层语义主干+轻量动作专家（流匹配/扩散）	$\pi 0$ 、GROOT N1、SmoIVLA	快（50Hz+）	高频闭环控制
分层式	高层规划→中层理解→底层控制，含显式中间表征	SayCan、Code as Policies	中（10-30Hz）	长时序可解释任务

来源：《视觉-语言-动作模型研究综述：迈向通用机器人》，国金证券研究所

单体式 VLA 模型的核心特征是在同一个架构中联合处理视觉观测、语言指令等多模态信息，并在统一潜在空间中进行跨模态建模与推理，最终通过动作解码器生成可执行的动作表示。

图表29：单体式 VLA 模型架构

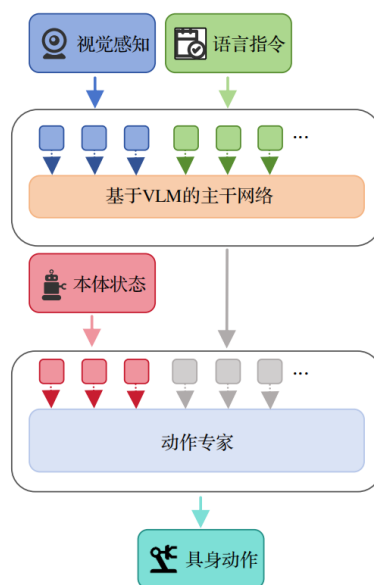


来源：《视觉-语言-动作模型研究综述：迈向通用机器人》，国金证券研究所

级联式架构的基本形态是将复杂推理与高频控制进行功能解耦，形成分工协作的两个子系统。其中，高层子系统依托通用大模型的深度推理与决策能力，负责语义理解与任务规划；底层子系统则以动作专家为核心，负责动作生成与闭环执行。通过这种解耦，高层能够集中于全局语义与策略组合，底层则专注于实时控制与执行稳定，从而在保证推理深度的同时满足高频闭环的时延约束。



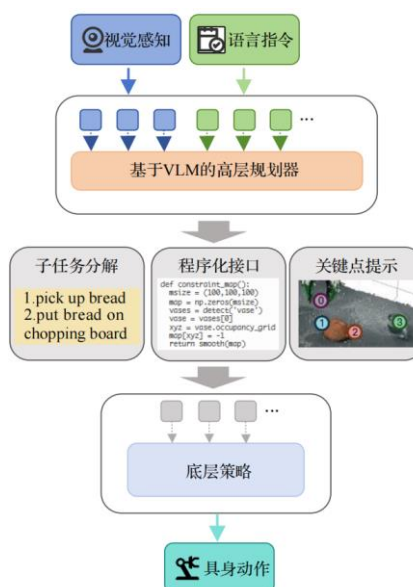
图表30：级联式 VLA 模型架构



来源：《视觉-语言-动作模型研究综述：迈向通用机器人》，国金证券研究所

分层式 VLA 架构通过在高层语义推理与底层控制之间显式引入中间表征，以实现规划与执行的独立。这种方式尤其适合长时序任务推理、空间抽象建模和复杂动作分解等场景。

图表31：分层式 VLA 模型架构



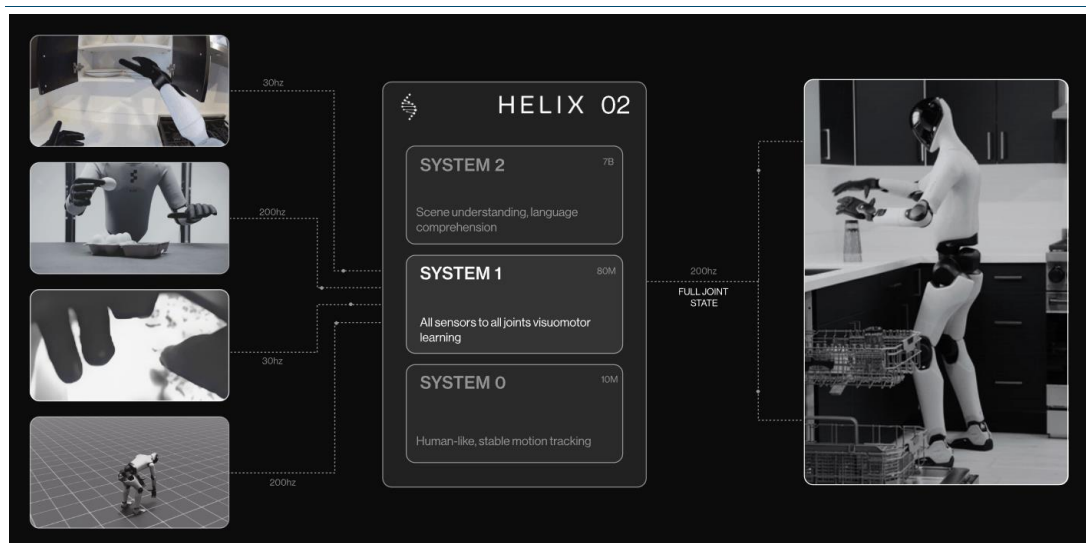
来源：《视觉-语言-动作模型研究综述：迈向通用机器人》，国金证券研究所

Figure Helix 模型作为级联式的代表，使用了"System 2 (7B 开源 VLM, 7-9Hz) /System 1 (80M 交叉注意力 Transformer, 200Hz)"的快慢脑架构，各层之间通过连续隐向量连接，Helix 02 在其下再叠一层 System 0 (10M 参数、1kHz、用 1000+小时人类运动数据训练的全身平衡先验)，变成"三层级联"，把级联延伸到"像素→扭矩"的全身尺度。

级联式兼顾 VLM 的深度语义推理（高层语义理解与底层高频控制解耦，微调时可选择冻结高层以降低训练成本）与动作专家的实时性（流匹配并行解码），且训练成本大幅降低，在绑定固定硬件的情况是最优的工程化落地路线。



图表32: Figure Helix 02 采用三层级联式



来源: Figure 官网, 国金证券研究所

谷歌最新推出的 Gemini Robotics 2 分层式架构则更加追求生态开放性, Gemini Robotics 2 不是"一个模型", 而是两个解耦、独立部署的模型通过工具调用拼起来的系统:

第一层 ER 2 (基于 Gemini 3.5 Flash 的 VLM) 当"高级大脑": 理解指令、拆解任务、跟踪进度、自我纠错, 然后把执行以工具调用 (tool orchestration) 的方式"交给任意一个底层 VLA";

第二层 Gemini Robotics 2 与 On-Device 2 是真正的 VLA 模型, 功能基本相同, 按照任务需要进行选择, Gemini Robotics 2 基于云端算力能力更强满足复杂全身高精度场景, On-Device 2 为端侧 VLA 满足本地部署需求。

图表33: Gemini Robotics 2 为分层式架构

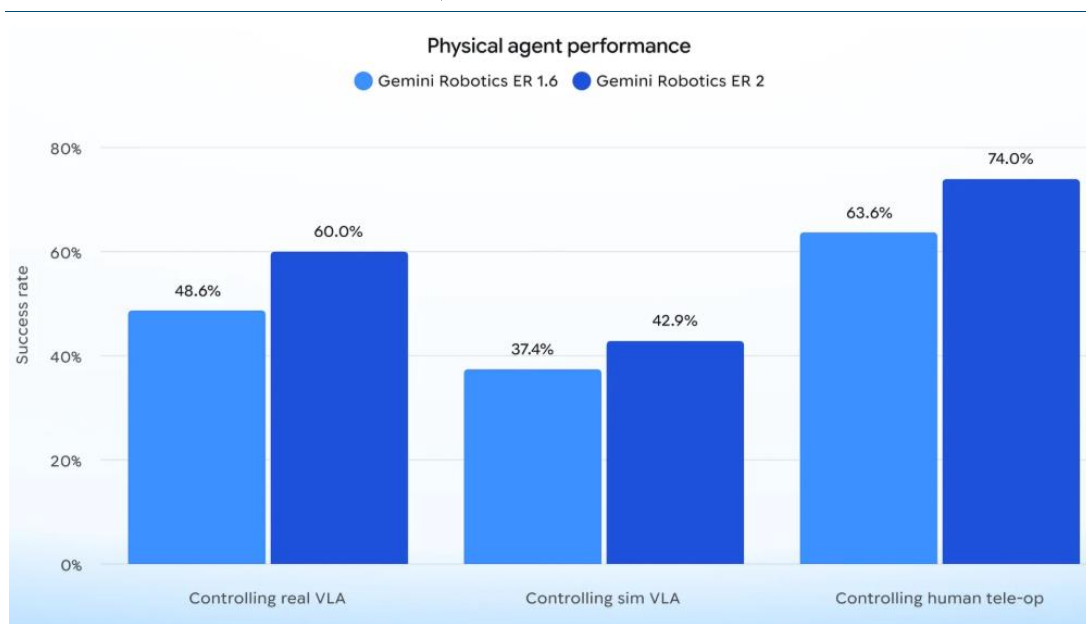
定位	模型	作用
推理	Robotics ER 2	本质上基于 VLM, 定位为高级大脑, 理解指令、观察房间、拆解多步任务 (数分钟、数百步决策)、跟踪进度、自我纠错、多机协作、安全监控。输出是决策与编排, 然后把执行"交给任意一个底层 VLA"层间通过工具调用衔接
	Robotics 2	基于云端算力的最先进 VLA, 实现全身控制 (脚到指尖)、双机械臂、灵巧手+夹爪, 适用于高保真、复杂全身任务
执行	On-Device 2	最高效的本地部署 VLA, 实现新本体快速适配, 适用于低延迟、隐私、离网场景

来源: Google DeepMind 官网, 国金证券研究所

其中 ER2 支持机器人与人类对话交互、理解现实物理环境、规划多步骤任务, 之后再将运动执行工作交由任意底层 VLA 模型完成, 相比 ER1.6 机器人现在可以通过读取连续视频流, 实时追踪任务进度、在出现故障时自主调整方案, 并精准判断何时推进至下一环节。本次更新同时推出多机器人协同能力, 多台机器人可在同一空间内配合作业, 完成单台机器人无法独立实现的复杂工作流程。



图表34: ER2 相比 ER1.6 性能显著升级



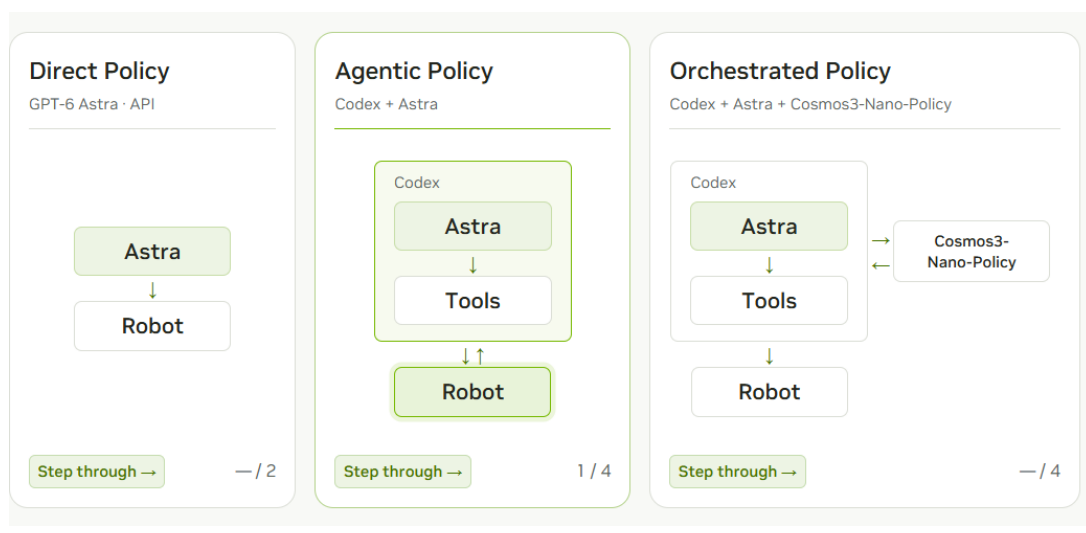
来源：谷歌官网，国金证券研究所

1.2.2 三类成本项同时坍塌，物理 AI 的 1 到 10 不再遥不可及

可以看到从 $\pi 0$ 之后模型的演进沿着几个方向展开：

- 1) 沿同一条技术路线做纵向深化 ($\pi 0.5 \rightarrow \pi 0.6 \rightarrow \pi 0.7$)，即在模型内部依次叠加数据来源、训练范式、记忆机制与上下文条件；
- 2) 系统架构层面的路线分化（级联式与分层式）；
- 3) 2026 年 GPT-6 Astra 打开的“模型外部”路线——不再训练 VLA，而由通用 Agent 基模通过代码、工具调用与经典控制器驱动机器人本体。此外还有数据策略与推理效率侧的并行优化。

图表35: Astra 三种机器人闭环控制方案

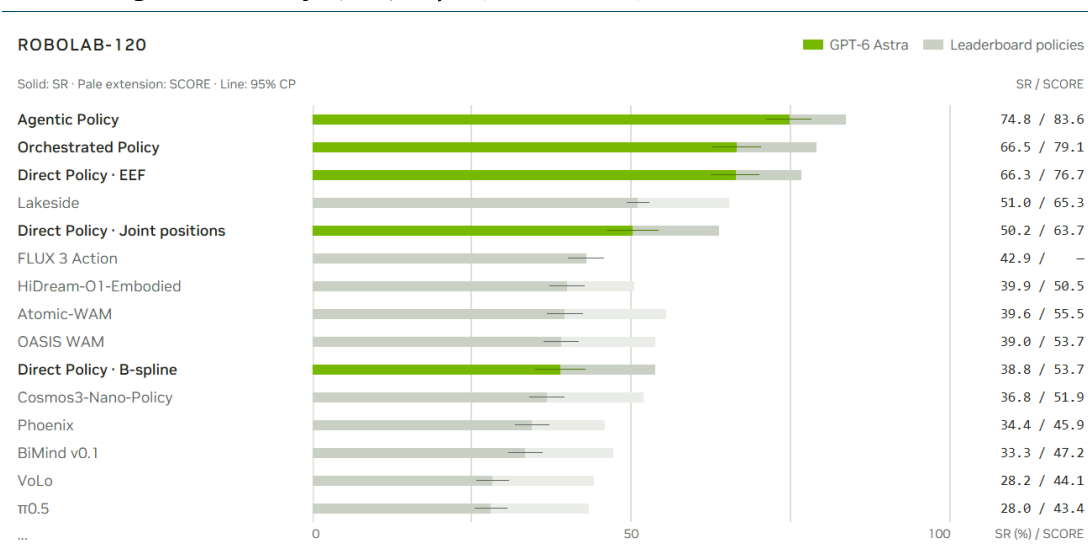


来源：英伟达官网，国金证券研究所

从 RoboLab-120 数据集上的成功率与完成度得分看，Astra 的智能体策略 (Agentic Policy) 成功率达到 74.8%；直接策略（末端执行器版本，Direct Policy EEF）为 66.8%，编排策略 (Orchestrated Policy) 为 66.5%。



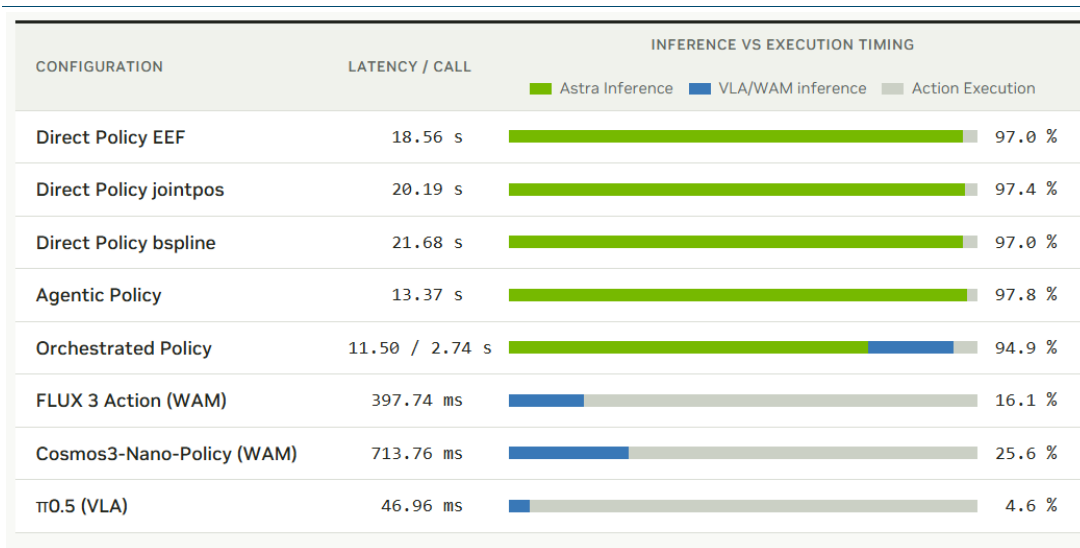
图表36: Agentic Policy 得分最高, 同时大幅高于其他模型



来源: 英伟达官网, 国金证券研究所

但目前其推理耗时占到了总耗时的绝对大头, 所有参评 Astra 配置下, 推理耗时占总时长的 94.9%~97.8%。Direct Policy EEF 单次调用平均耗时 18.56 秒; Astra 推理耗时已经包含网络延迟。

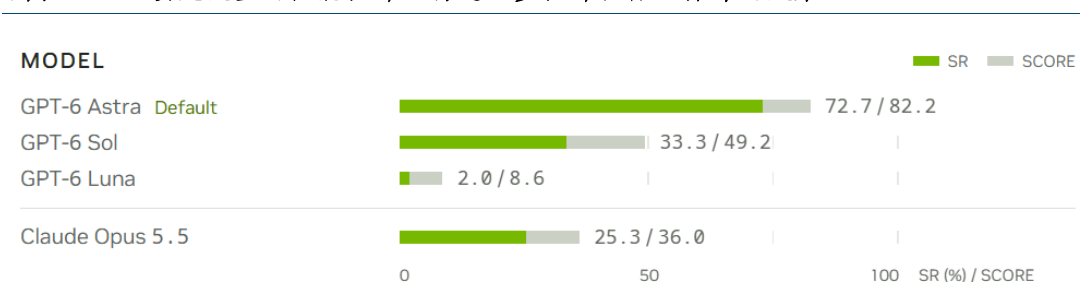
图表37: Astra 的瓶颈不在机器人执行, 而在大模型推理



来源: 英伟达官网, 国金证券研究所

后续大模型的性能提升也将进一步提升机器人操作的效果, 可以看到在直接策略的测试中, Astra 成功率 72.7%, Sol 为 33.3%, Luna 仅 2.0%, Claude Opus 5.5 为 25.3%, 模型的选择对成功率的影响非常大。

图表38: 后续大模型的性能提升也将进一步提升机器人操作的效果





来源：英伟达官网，国金证券研究所

随着以 Astra 为代表的第三类技术路线出现，我们认为三项重要成本开始坍塌，物理 AI 有了从 1 到 10 的机会。

图表39：三项成本同时坍塌，物理 AI 有了从 1 到 10 的机会

核心变化的成本项	对应的模型变化	消失的成本
不再为每个场景重训一遍	$\pi 0.5$ 的异质数据迁移泛化，带来泛化能力提升	场景适配成本
不再为新任务采一遍数据	$\pi 0.7$ 的组合泛化+语言教练，把怎么做也写进提示，长程新任务可以用语言教练+蒸馏成高层策略的方式习得，而不必采集示范数据	任务扩张成本
不再自研“大脑”	Astra 让机器人核心的决策层不再完全依赖自研而是可以外采	“大脑”获取成本

来源：国金证券研究所

三类成本坍塌的交点就是现在，场景成本被大幅压缩，新任务的边际数据从“采集”变成“教练”，“大脑”从必须自研变成可以外采，最终带来的结果就是物理 AI 第一次可以被直接按照单位经济性来进行计算。

图表40：为什么物理 AI 的投资不是更早

时间窗	模型状态	是否可行
2023 之前	动作 1 - 3Hz (RT-2 级)，不可实用；每进一个场景要重采数据	太早：能力与成本双不足，只能做 demo
2024	$\pi 0$ 解决“动作怎么生成”（最高 50Hz）	仍不足：泛化与可靠性未解决，仍是“实验室可用”
2025—2026	$\pi 0.5$ 解决数据、 $\pi 0.6$ 解决可靠、 $\pi 0.7$ 解决组合泛化、Astra 打开分工	现在：三类成本同时坍塌

来源：国金证券研究所

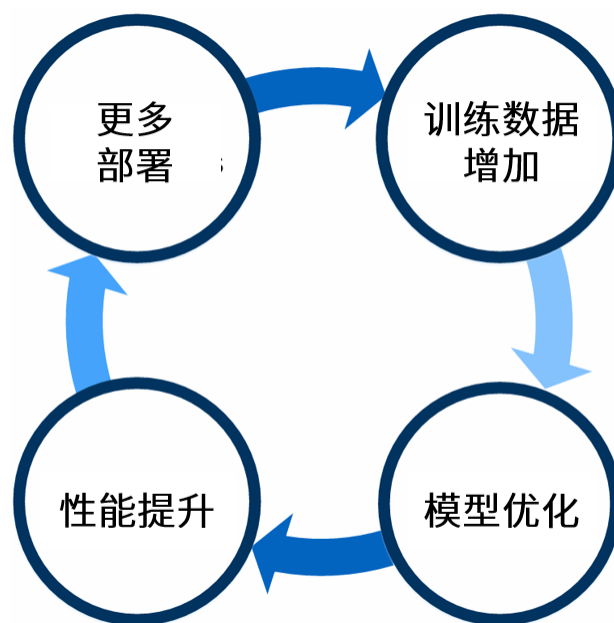
2.应该投什么？短板就是机会

2.1 数据：最核心的瓶颈和机会

物理 AI 产业具有巨大成长潜力的底层逻辑是“数据飞轮”，通过更多的应用部署能带来更多数据以供训练，最终优化出更成熟的模型强化性能进一步推动放量，形成自我加速的正向循环。



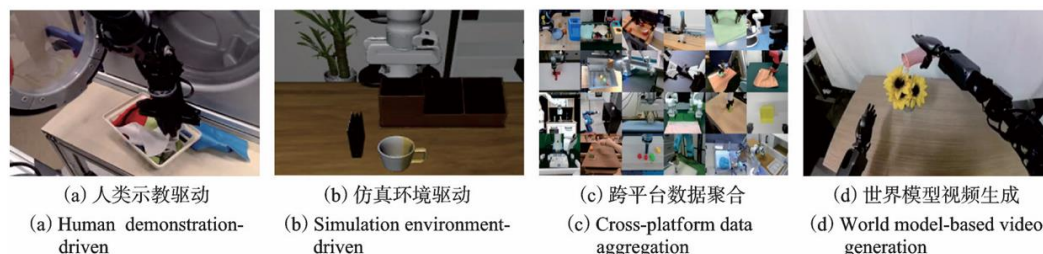
图表41：“数据飞轮”是物理AI产业具有巨大成长性的底层逻辑



来源：《Data Pyramid and Data Flywheel for Robotic Foundation Models》，国金证券研究所

具身智能领域中的数据并非单一模态，而是由多种信息共同构成的复杂组合。典型的具身智能数据可以表示为视觉（Vision）、语言（Language）、动作（Action）以及状态（State）等多模态信息的联合体。模型的能力和上限很大程度上受制于其训练数据的质量以及规模，现有具身智能数据采集方式可归纳为4种核心范式。

图表42：4种具身智能数据采集方式示例

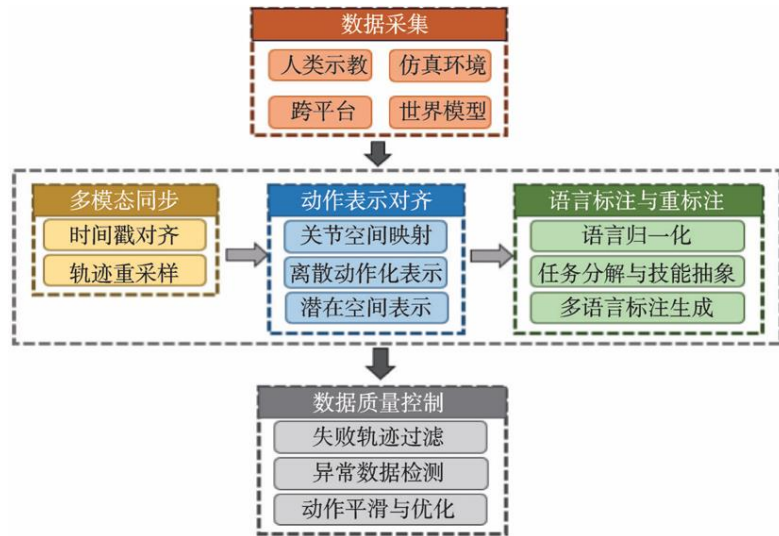


来源：《具身智能数据采集与处理综述》，国金证券研究所

对于VLA模型而言，原始交互数据通常来源复杂，既可能包含真实机器人示教数据，也可能来自仿真环境或跨平台大规模整合数据集。这些数据在动作表示、时序频率、语言标注粒度以及数据组织结构等方面存在显著差异。



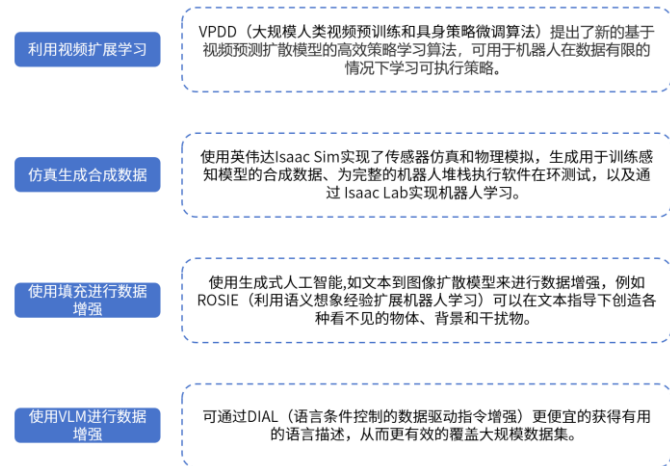
图表43: 具身智能数据处理流程



来源:《具身智能数据采集与处理综述》, 国金证券研究所

目前有一些方法来解决数据瓶颈,但仿真数据的保真度和规模有限,依旧是真机数据训练效果最好。其中人类示教驱动是目前应用最为成熟的数据获取方式之一。该范式通过让人类直接控制机器人执行任务,采集包含视觉观测、语言指令以及连续动作轨迹的高质量示范数据,广泛应用于模仿学习和具身智能系统中。与基于自动策略生成的数据采集方式不同,人类示教能够在数据中显式注入任务语义、动作顺序以及失败规避等高层决策信息,从而显著降低学习难度。

图表44: 解决数据瓶颈的常用方法



来源:《机器人模型发展与挑战》, 国金证券研究所

图表45: 4种具身智能数据采集比较

采集方式	采集成本	数据质量	数据	轨迹数	机器人种类数	场景数	动作种类数	是否有语言标注
人类示教驱动	极高	高	RT-1 Datasets ^[5]	130 k	1	15	700	是
			BridgeData ^[6]	7.2 k	1	12	4	是
			BridgeData V2 ^[6]	60.1 k	1	24	82	是
仿真环境驱动	低	一般	LIBERO ^[17]	6.5 k	1	5	130	是
			CALVIN ^[11]	约 24 h	1	4	34	部分是
			Meta-World ^[8]	n/a	1	1	50	否
跨平台数据聚合	一般	较高	OXE ^[25]	1.4 M	22	311	217	部分是
			RoboTwin 2.0 ^[23]	100 k	5	50	50	是
世界模型视频生成	低	较低	DreamGen ^[20]	n/a	n/a	n/a	n/a	是
			RoboEnvision ^[24]	n/a	n/a	n/a	n/a	是

来源:《具身智能数据采集与处理综述》, 国金证券研究所

根据中国信息通信研究院《具身智能训练场研究报告(2026年)》信息,假设实现机器人“ChatGPT”时刻的模型参数量级为 55B,按大语言模型的缩放定律,训练所需数据量约 1.1 万亿 token。若全部由真机回流提供,按控制频率 8Hz、每步约 3 个预测 token、每天有效交互 2 小时、每年工作 300 天测算,约需 21200 台机器人连续工作一年,产出千万小时级有效数据。而当前全球可用的高质量真实数据仅数十万至百万小时级,远未达到支撑具身基础模型智能涌现的门槛。

为了解决真实数据瓶颈问题,目前产业正多路线探索真实数据采集技术。

图表46: 产业正多路线探索真实数据采集技术

路线	数据采集方法	成本	规模化程度	真机映射难度	数据用途
无本体	人类第一视角视频数	低	高	高	支持场景理解、人类行为理解、空间预测



采集	据采集			支持多场景下末端精细交互操作任务
	末端操作数据采集	低	高	
	动作捕捉数据采集	中等	中等	
	遥操作数据采集	高	低	
真机采集	一致			无域偏移问题，与本体运动、力控、结构一致
	自主作业回流数据	最高	低	具备真实环境信息、真实扰动、失败案例等高价值数据

来源：《具身智能训练场研究报告（2026 年）》，国金证券研究所

目前国内/海外企业均持续加码真实数据采集，但路径依旧较为多元，例如 Figure 披露外采数据不达质量要求，第三方和自建两条路都在同步推进。

图表47：全球具身智能基建规模与多模态数据集全景对比

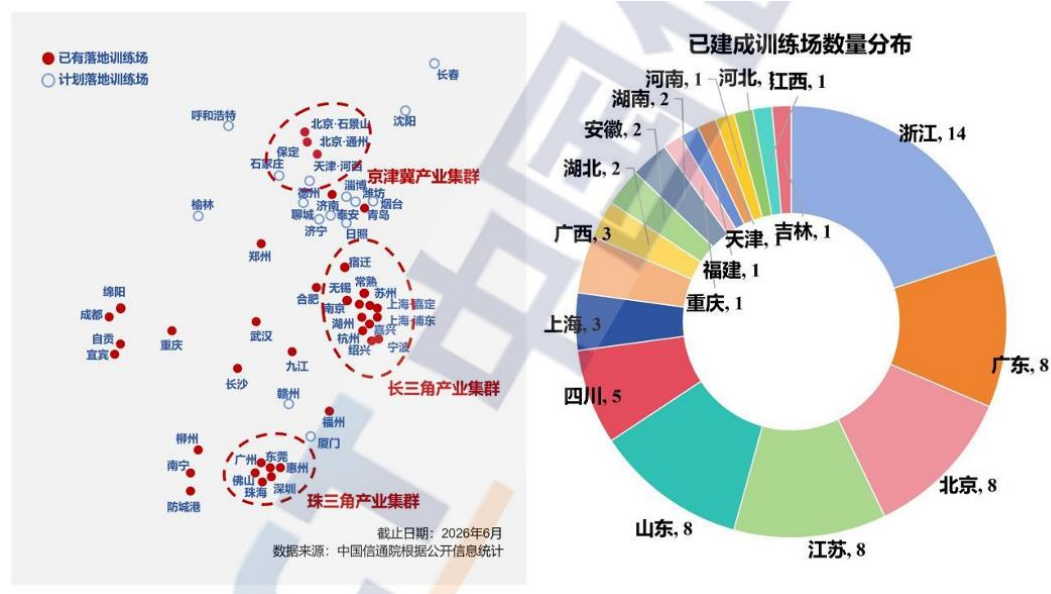
机构/企业	核心战略模式	物理训练场/采集设施规模	数据集产出与核心特征
Tesla（特斯拉）	闭环超级数据工厂	直接嵌入弗里蒙特等超级制造产线	实战生产闭环数据：聚焦端到端真实工业任务，不对外开放，直接反哺自身模型迭代
Figure AI	跨界商业物业网络	自建 BotQ 工厂+联合 BMW、商业地产与制造产线；同时推分布式数据采集平台 Index	泛化场景预训练数据：大量真实世界长尾任务采集，构建闭源全球预训练数据池
Google（谷歌）	跨本体联合开源	依托全球高校与合作方实验室 Open X-Embodiment	聚合超 60 个子数据集，100 万+轨迹，验证跨硬件本体泛化
Hugging Face	社区化众包网络	依托全球开源社区分散节点	LeRobot 计划：持续扩充的开源库，提供标准化低成本双臂采集方案
北京石景山人形机器人创新中心	集中式全场景枢纽	总占地超万平米，1:1 还原场景，部署百余台异构设备	高质量交互数据：覆盖 16 类真实场景，牵头制定行业统一测试基准与对齐标准
乐聚机器人	分布式协同训练网	全国参与建设十余个训练场（含北京、苏州等）	多模态数据集：6 万分钟首批开源，涵盖 117 项底层技能，年数据产能 2500 万多模态数据集
帕西尼感知	触觉平台训练集群	天津数据工厂达 1.2 万平米，构建 15+N 全场景矩阵	触觉感知通用数据：支持超千种通用任务训练，以多维度触觉为核心特征
智元机器人	聚焦长程任务基地	上海张江基地超 4000 平米，AGIBOT WORLD	超 100 万条轨迹，80%覆盖 217 个复杂任务场景；长程复杂任务（仿真+真机结合）
戴盟机器人	外发式分布式采集	不依赖单一物理中心，依托超 80 个真实社会化场景	Daimon-Infinity：约 1 万小时规模，数百万条轨迹，探索不受物理空间限制的泛化网络
光轮智能	虚实结合数据底座	构建集仿真、数据、评测于平台化基础设施	侧重为机器人公司和产业落地场景提供系统化的数据底座服务
无问智科	区域型集中枢纽	长三角（德清）训练场总面积 8000 平米（一期 3000 平米）	千万级数据产能：依托大面积独立场地，预计 2026 年全年数据采集量可达千万条

来源：中国经济新闻网，国金证券研究所

国内根据中国信息通信研究院《具身智能训练场研究报告（2026 年）》统计，到 26 年 6 月底，全国建成启用超 70 家训练场，在建或规划中的训练场 46 家。具身智能训练场或成物理 AI 时代的新基建。



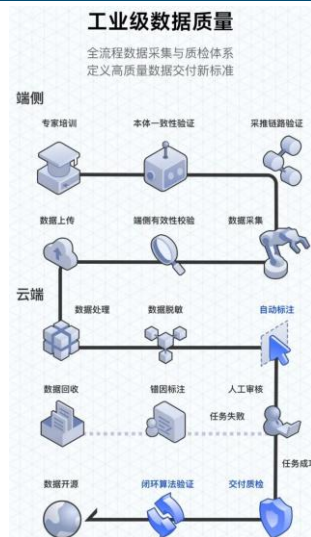
图表48：中国建成的训练场数量分布及统计



来源：《具身智能训练场研究报告（2026年）》，国金证券研究所

智元不光建立了大规模的数据采集工程，开源的 AGIBOT WORLD 2026 数据集构建了一套多层级的标注体系，原始数据采集完成后，都将通过算法策略训练与真机测试，确保数据采集的有效性。所有数据均进入严格的 DaaS 工业质检流水线，依据统一标准进行多轮筛查与清洗，每一个环节都围绕一个目标展开：确保交付结构清晰、标注精准、可直接用于模型训练的高质量数据。

图表49：智元开源数据集把 VLA 算法侧的优化结论变成了数据供应链的工业化版本



来源：智元机器人官网，国金证券研究所

Scale AI 覆盖了 Physical Intelligence、Generalist AI、Cobot 等客户，目前每天上传 1000+小时的示范数据，设立 Data Factories（集中式）/Residential（居家众包）/Commercial & Industrial（现场）三类数据采集场，并自研 Scale Harness 无本体第一视角采集平台。

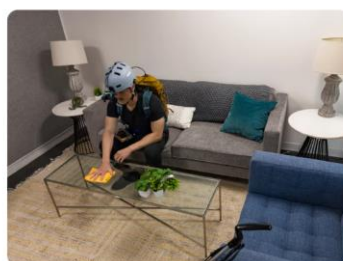


图表50: Scale AI 三类数据采集场每天上传 1000+小时示范数据



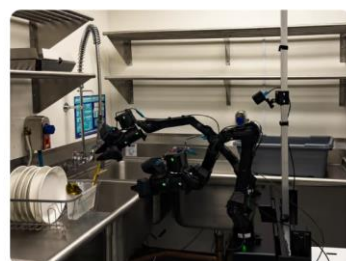
Data Factories

High-volume centralized collection operations



Residential

At-home collection with globally-distributed collectors



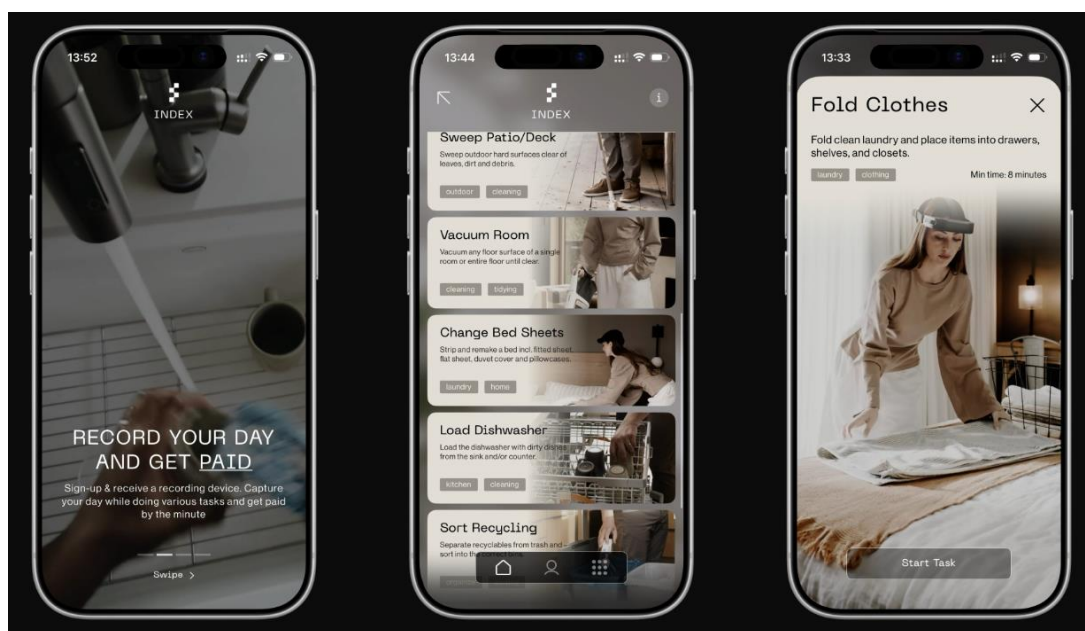
Commercial and Industrial

In-field deployments of collection platforms

来源: Scale AI 官网, 国金证券研究所

Figure 尝试向外部数据供应商进行采购,但供应商在吞吐量、多样性、质量上都达不到 Helix 的要求,所以自建全流程推出 Index (展示了采集从集中走向分布的一个路径),目前已向数据提供者支付 1500 万美元,APP 实现 26.4 万次下载、108 个国家、4.4 万周活、1600 万段视频;平台每秒接收 30 分钟视频。

图表51: Figure 自建全流程实现分布式数据采集



来源: Figure 官网, 国金证券研究所

数据采集需求的爆发,首先将对数据采集设备及供应链形成巨大拉动,根据数据采集方式不同其中包含人形机器人本体/机械臂/相机/IMU/触觉传感器等。



图表52：奥比中光无本体数据采集硬件平台



图表53：诺亦腾动作捕捉解决方案设备



来源：奥比中光官网，国金证券研究所

来源：诺亦腾官网，国金证券研究所

同时数据生产商/数据基础设施供应商(工具/仿真/资产)/训练场运营等领域也建议关注，例如索辰科技以CAE物理求解器与“天工开物”平台为核心，作为国内极少数掌握底层物理引擎的公司，为物理AI提供“物理正确性”。

图表54：索辰科技依托深厚CAE物理求解器积累持续加大物理AI布局

索辰智能工坊

该模块利用物理AI技术，为用户提供从数据生成、模型训练到全局寻优与实时预测的全流程自动化解决方案。模块能够系统性地规划试验方案并自动化生成海量高保真仿真数据，基于这些数据快速构建AI代理模型与降阶模型，替代传统耗时计算。利用高效模型，用户可进行大规模设计空间探索、多目标寻优及动态物理场实时预测，并支持与外部仿真软件兼容和二次开发接口，将行业知识与AI深度融合，帮助企业突破效率瓶颈，以更低成本、更短周期开发性能更优的创新产品。

- ✓ 从数据生成、模型训练到全局寻优和实时预测全流程自动化
- ✓ 系统性规划试验方案并自动化生成海量高保真仿真数据
- ✓ AI代理模型与降阶模型替代传统耗时计算
- ✓ 大规模设计空间探索与多目标寻优
- ✓ 动态物理场实时预测实现产品性能突破
- ✓ 兼容外部仿真软件与开放二次开发接口

来源：索辰科技官网，国金证券研究所

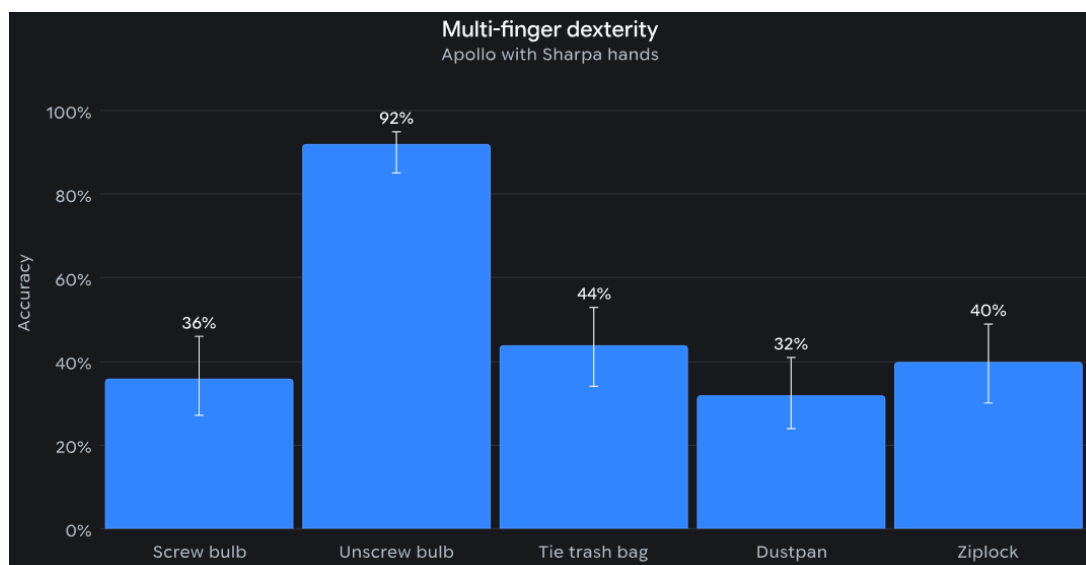
2.2 感知与执行：高精度灵巧操作仍是硬边界，“卖铲子”确定性强

数据采集作为目前物理AI最大短板需求刚性但技术路线上仍有一些分歧，传感器/执行器作为物理AI落地的必需品也同样具有不可替代的地位，同时受到技术路线变化的影响更小。

例如目前看到谷歌DeepMind发布的基于SharpWave多指灵巧手操作成功率依旧偏低，我们认为其中除了“模型不够聪明”也有触觉硬件的缺位。同时可以看到拧下灯泡/拧上灯泡成功率有较大差异，主要就由于拧下只需要夹住并旋转；拧上则要螺纹对位、接触力控制，失配后还要重新定位。



图表55：多指灵巧手操作成功率目前依旧偏低



来源：Google DeepMind 官网，国金证券研究所

感知端来看主要包括视觉传感器（如摄像头、激光雷达、毫米波雷达、红外传感器等）、听觉传感器（如麦克风）、嗅觉传感器（如气味传感器）、触觉传感器（包括力传感器）、编码器、IMU 等，以实现感知和分析外部环境并进行自身状态反应，实现复杂的交互和操作任务。

图表56：人形机器人传感器的主要类型

传感器类型	原理	获得的信息	在工业机器人中的应用
触觉传感器	电容式、压电式、压阻式、光学式	接触力、面积、位置	人机协作、物体抓取、质量监控
视觉传感器	CCD 或 CMOS 成像	图像	人-机器人协作（HRC）、导航、机械手控制、装配、机器人编程
激光传感器	飞行时间（TOF）、三角测量、光学干涉	距离、位移	人-机器人协作（HRC）、导航、操纵器控制
编码器	光电、磁性、电感性、电容性	角位移	导航、操纵器控制
接近传感器	电容式、电感式、光电式	物体接近	人-机器人协作（HRC）、物体抓取
惯性的传感器	航位推算（DR）	加速度、角速度、方位角	导航、操纵器控制
扭矩传感器	感应电阻应变	扭矩	人-机器人协作、物体抓取、机器人编程
声传感器	电容的	声音信号	人-机器人协作（HRC）、焊接
磁性传感器	霍尔效应	磁场强度	航行
超声波传感器	飞行时间（TOF）	距离	障碍物回避

来源：《工业机器人传感器技术综述》，国金证券研究所

视觉：主流方案为 3D 视觉，国产化程度较高

机器视觉是智能制造体系中的核心感知与控制技术，被誉为工业自动化的“眼睛与大脑”。机器视觉系统通过光源、镜头、工业相机等硬件完成图像采集，再经视觉控制器与算法软件进行分析处理，最终向驱动单元输出控制指令，实现对工件的高精度检测、定位、测量与识别。作为替代人工检测的关键技术，机器视觉可有效提升生产效率、保障产品良率，是制造业向智能化、柔性化升级的重要支撑。其系统性能直接决定了工业场景中自动化生产的精度与稳定性，是高端制造产业链中不可或缺的基础环节。



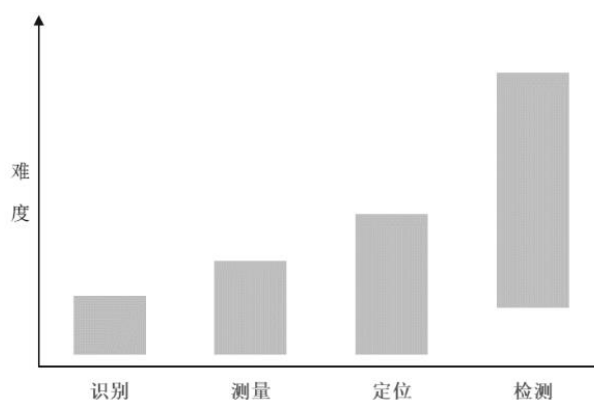
图表57: 机器视觉系统示意图



来源：奥普特公告，国金证券研究所

机器视觉核心功能可归纳为识别、测量、定位与检测。识别功能通过特征比对实现条码读取、字符甄别，核心在于准确率与响应速度；测量功能可将图像像素标定为标准尺寸，实现高精度几何量检测，是其核心优势场景；定位功能可输出二维/三维位置信息，支撑自动化设备的精准抓取与装配；检测功能则涵盖外观缺陷、装配完整性等复杂场景，也是技术难度最高、应用最广泛的环节。

图表58: 机器视觉核心功能是识别、测量、定位与检测



来源：公司公告，国金证券研究所

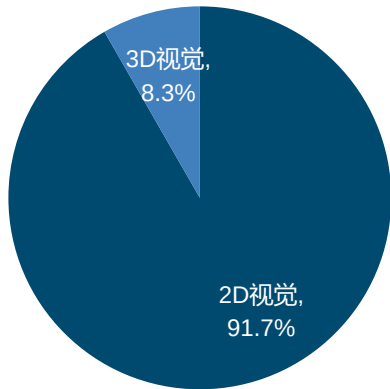
机器人视觉方案主要分为 2D 视觉和 3D 视觉：

- 1) 2D 视觉：通过 2D 相机分析灰度或彩色图像中的像素灰度特征获取目标中的有用信息，识别纹理形状。容易受到光照影响，且无法实现三维精准测量。
- 2) 3D 视觉：精度更高、信息量更大（三维图像）、集成度更高，适用于更加复杂、精密的识别、检测需求。

机器人的视觉感知属于高度集成的模组方案，对于体积、成本、精准度都有较高要求，因此 3D 视觉高度集成、体积小、精度高的优势更适合机器人的场景，据中商产业研究院数据，2024 年中国机器视觉市场中 3D 视觉占比 8.3%，未来成长空间广阔。



图表59：2024 年中国机器视觉市场中 3D 视觉占比 8.3%



来源：中商产业研究院，国金证券研究所

图表60：机器人视觉感知向 3D 化发展

对比指标	2D	3D
产品型态	模块化视觉设备 (体型大)	嵌入式视觉设备 (高度集成)
应用场景	一般固定	复杂多变
扫描精度	工业产线范围内	毫米级甚至更高
维度	平面维度	立体维度
技术要求	较低	技术壁垒较高
下游市场	识别、检测、测量	刷脸支付、FaceID、智能机器人

来源：奥比中光官网，国金证券研究所

3D 视觉传感器的性能由芯片、算法、传感器模组影响。从硬件端来看，壁垒在于感光芯片、多传感器融合模组设计以及生产中标定、对齐等工艺把控。从软件端来看，3D 视觉传感器对于算法研发、光学系统等有较高要求。

- 从性能表现来看，壁垒在于精度、视角、测量范围、分辨率：
- 深度分辨率：代表每一帧图像深度数据点数，分辨率越高细节越优；
- 深度帧率：帧率越高，动态场景支持越好。
- 视场角/测量范围：视场角/测量范围越大覆盖场景范围越大。
- 精度：体现 3D 测量数据值与真实值之间的差异。
- 功耗：同性能下，功耗越低适配性越好。

从性能指标来看，国内头部厂商奥比中光在分辨率、精度和功耗上已经对标乃至超过海外竞争对手，有望引领国产厂商实现替代。

图表61：主流厂商短距离 3D 视觉传感器性能对比

技术指标	英特尔	瑞芯微	华捷艾米	奥比中光
型号	D435	RMSL201-1301	A100S+mini	AstraE Astra—MiniS
深度分辨率、深度帧率	1280x720@30fps		640×480@30fps	1280×800@30fps 640×400@60fps 1280x1024@7fps 640x480@30fps
视场角(H,V)	85.2, 58.46.0, 68.2		60, 47	67.9, 45.3 60, 49.5
测量范围	0.1~10m		0.28~1m	0.25~1.5m 0.35~1m
精度	±20mm@2m		±5mm@1m	±5mm@1m ±1mm@1m
功耗	<4.5w	<2.5w	3.2w~4w	<2.2w <2.4w

来源：奥比中光招股书，国金证券研究所

图表62：主流厂商长距离 3D 视觉传感器性能对比

技术指标	微软	英特尔	英特尔	华捷艾米	奥比中光
型号	Kinect1.0	R200	D435	A100M	AstraPro
深度分辨率、深度帧率	640×480@30fps	640×480@30fps	1280x720@30fps	640×480@30fps	1280x1024@7fps 640x480@30fps
视场角(H,V)	57, 43	59, 46	85.2, 58	60, 48	60, 49.5
测量范围	0.8~3.5m	0.4~2.8m	0.1~10m	0.4~6m	0.6~8m
精度	±1mm@1m	±12mm@2m	±20mm@2m	±20mm@1m	±1mm@1m
功耗	<2.5w	1.3~1.6w	<4.5w	<3.5W	<2.4w

来源：奥比中光招股书，国金证券研究所

海外头部厂商先发优势明显，国内厂商多技术布局、通过芯片自研逐渐打开市场。海外头部 3D 视觉传感器厂商通过芯片自制、自身产品依托，业务规模较大，国内代表公司奥比中光等通过自研芯片，实现 3D 视觉传感器部分技术指标逐渐接近海外龙头厂商，未来有望实现国产替代。



图表63：主流视觉传感器厂商梳理

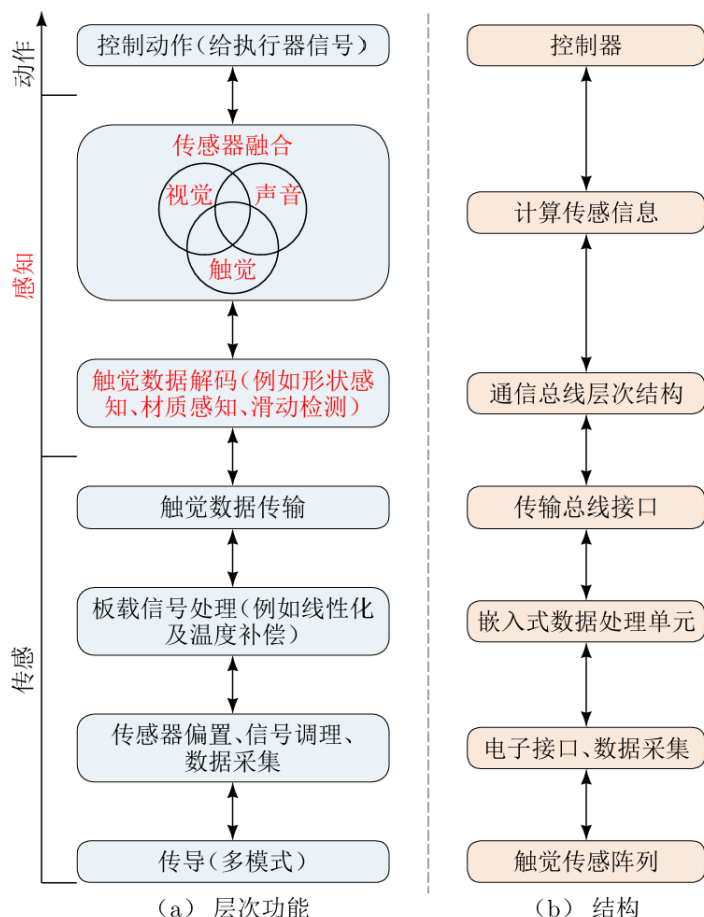
公司	主要技术	市场地位（3D视觉领域）	核心竞争力
苹果	结构光dToF	全球最大的内置3D视觉传感器的移动产品制造商，在消费电子领域基于3D视觉感知技术的布局一直处于领先地位	①芯片自主性：一款以上带结构光深度引擎加速的芯片，与STMicro合作结构光感光芯片，与索尼合作dToF感光芯片 ②关键元器件自主性：自主设计且有高质量的供应商、代工厂资源 ③量产能力：高于百万级 ④其他：有手机、平板等终端产品线做依托
微软	结构光iToF	3D视觉传感器以及配套的算法服务知名度高	①芯片自主性：自主研发的iToF芯片 ②关键元器件自主性：自主设计加采购（微软已经宣布和奥比中光合作设计制造下一代Kinect3D视觉传感器） ③量产能力：高于百万级（代工） ④其他：有算法（如骨架等）平台、云计算平台（Azure云）、操作系统（Windows系列）以及游戏机（Xbox）等作为依托
英特尔	结构光双目Lidar	目前世界上规模最大的消费级双目3D视觉传感器制造商	①芯片自主性：双目3D深度引擎芯片自主，芯片的设计及制造是英特尔的领先优势 ②关键元器件自主性：自主设计加采购 ③量产能力：百万级
华捷艾米	结构光	自研结构光3D视觉传感器，有一定的量产能力。	芯片自主性：深度引擎芯片自主设计
奥比中光	结构光双目iToF/dToF Lidar	在3D传感器领域持续出货到手机、人脸、机器人、三维扫描等多个潜力领域，市场规模稳步扩大，在客户中的认可程度也逐步提高	①芯片自主性：结构光、双目深度引擎芯片自主设计，iToF感光芯片自主设计，dToF、结构光感光芯片在研； ②关键元器件自主性：自主设计加采购； ③量产能力：百万级； ④其他：全球化技术团队+本土化产业链配套+本土化庞大市场规模（消费电子、移动支付等），形成了对国际巨头差异化及本土化优势，对国内企业的先发技术储备及应用优势

来源：奥比中光招股书，国金证券研究所

触觉：柔性触觉传感器是关键传感元件，视觉+触觉方案应用潜力大

触觉传感系统是机器人实现环境感知与精细操作的重叠系统。通过接触获取物体的物理特征，为机器人提供力度控制与物体识别所需的信息。在抓取过程中，触觉反馈可帮助机器人动态调节抓力，避免因抓力不足导致物体滑落，或因抓力过大造成物体损坏；结合深度学习等识别算法，机器人还可进一步判断接触物体的材质与类别，提升操作的精细化和智能化水平。

图表64：机器人触觉传感系统的层次功能和结构框图



来源：《触觉传感器与电子皮肤研究进展》，国金证券研究所

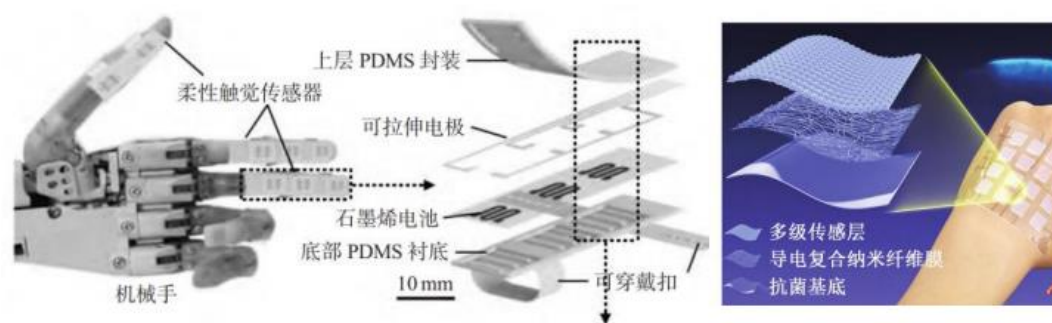


电子皮肤是触觉传感器由单一力觉检测向多模态仿生感知升级的重要方向。随着新材料与新工艺持续突破，触觉传感器的研究已不再局限于力信息采集，而是进一步模拟生物皮肤的复杂属性与功能，包括高柔性、自愈合、自清洁等物理特性，以及对温度、湿度、硬度和黏度等信息的感知能力，推动机器人触觉系统向更接近人类皮肤的方向演进。

电子皮肤主要可分为基底层、导电层和传感层：

- 基底层：基底层是电子皮肤的最内层，通常由柔性材料制成，以模拟人体皮肤的弹性和柔韧性。又称为支撑层，起到支撑作用。
- 导电层：导电层在电子皮肤中起到传导电信号的作用，通常包含导电材料，如导电石墨或特定的导电聚合物，用于检测并传输由外界刺激引起的电信号变化。导电层的设计和材料选择对于电子皮肤的灵敏度和响应速度至关重要。
- 传感层：传感层位于电子皮肤的中间层，包含了大部分的传感元件，负责精确识别压力和摩擦力并转化为电信号。传感层可以实现对外界刺激的精细感知，是电子皮肤的核心材料。

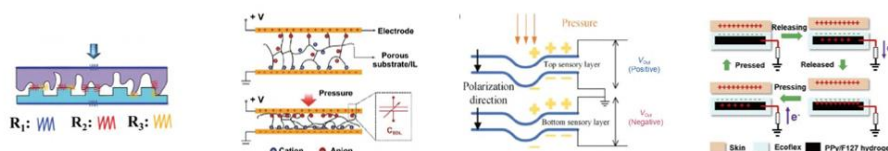
图表65：电子皮肤主要分为基底层、导电层、传感层



来源：《纤维基自供能电子皮肤的构建及其应用性能研究进展》，国金证券研究所

柔性触觉传感器是电子皮肤中最重要的传感元件。从检测方法来看，柔性触觉传感器传统方案主要包括压阻式、电容式、压电式、摩擦式四种类型，

图表66：传统的主要触觉传感器类型优缺点介绍



	压阻式	电容式	压电式	摩擦式
原理	材料的压阻效应，即将机械信号转换成电信号，具体机制包括材料的电阻率变化、接触电阻的变化以及复合材料中的量子隧穿效应等。	电容效应，将力学信号转化为电容信号，通过测量电容值的变化来获取物体与电子皮肤之间接触状态的信息	压电效应，压电材料由于形变而产生电极化	柔性摩擦纳米发电机，通过摩擦效应和静电感应耦合作用将外部机械能转换为电能
优点	成本较低、制造结构简单、稳定程度高、灵敏度高、响应快	灵敏度高，能够测量微小接触力	高灵敏性、响应快	结构简单，制备简便
缺点	电阻变化可能受到环境因素的影响	对噪声敏感，介电常数易受到环境温度影响	压电材料常常会受到温度和湿度等环境因素的影响	受环境影响大，只能检测动态压力

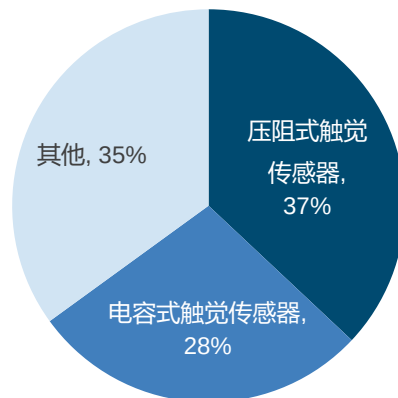
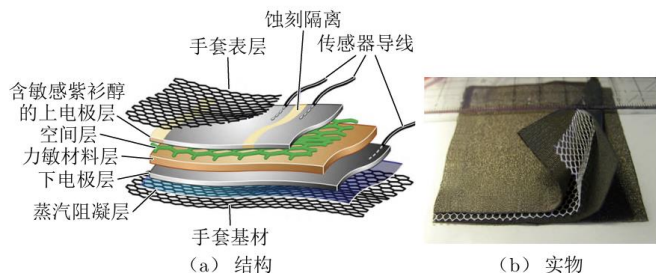
来源：《柔性触觉传感电子皮肤研究进展》，国金证券研究所



其中压阻式触觉传感器具有成本低、稳定性好等多个优点，是目前的主流方案。其工作原理是基于压阻效应，材料受力后内部电阻率发生变化，采集电路再将电阻变化转换为可测量的电信号，从而检测受力大小。据华经产业研究院，压阻式触觉传感器因制作成本低、高灵敏度等因素，其应用最为广泛，市场占比约为 37%，其次是电容式触觉传感器，占比约为 28%。

图表67：压阻式柔性触觉传感器结构及实物图

图表68：2024 年压阻式触觉传感器市场占比约为 37%

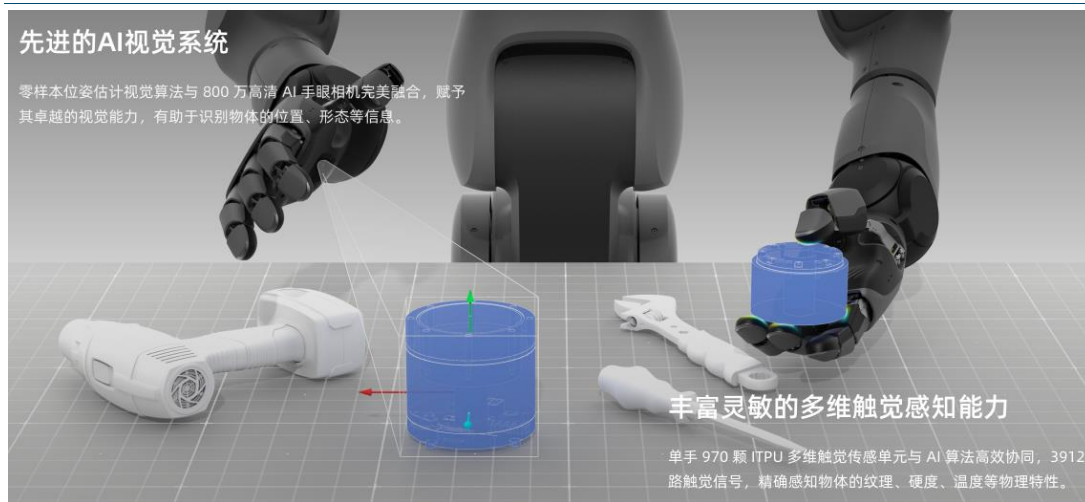


来源：《触觉传感器与电子皮肤研究进展》，国金证券研究所

来源：华经产业研究院，国金证券研究所

国内头部厂商帕西尼提出了视觉+触觉方案，在触觉的多项技术指标实现领先，有望成为未来的主流方案。在 2024WRC 大会上，帕西尼正式发布了第二代多维触觉灵巧手 DexH13，及基于 ITPU 技术的多维度触觉传感器 ITPU GEN2。在帕西尼感知科技的 DexH13 双手上，搭载了近 2000 颗自研生产的高精度触觉传感器 ITPU GEN2；此外，该双手还集成了 DexH13 的 800 万像素高清 AI 手眼相机，基于先进的零样本位姿估计视觉算法，能够识别和抓取物体六维位姿，适应多种复杂环境和任务。

图表69：帕西尼的 DexH13 多维触觉灵巧手结合了视觉和触觉



来源：帕西尼官网，国金证券研究所

帕西尼的 ITPU GEN2 触觉传感器在多项技术指标领先、且成本低。根据帕西尼官网，ITPU GEN2 多维触觉传感器在灵敏度、抗干扰等多项指标表现优异，且通过全集成方案，产品成本相对较低。



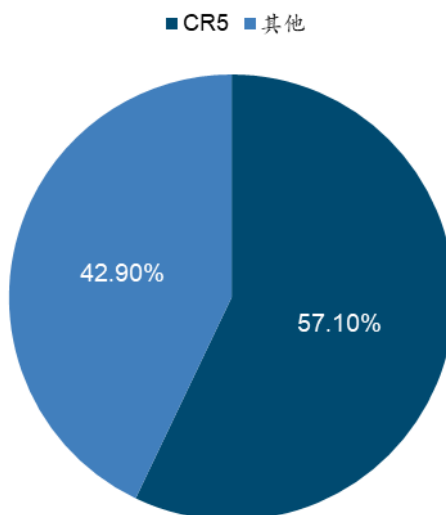
图表70：帕西尼 ITPU GEN2 触觉传感器在多项指标技术领先

	PaXini专业级机器人ITPU GEN2	光学触觉传感	压阻触觉传感	电容触觉传感
灵敏度	高	高	低	高
动态范围	高	低	低	中
响应速度	高	低	高	高
迟滞	低	高	高	低
尺寸	小	大、厚	小	小
抗干扰能力	软硬件优化，抗干扰能力强	中	强	弱
柔性抓取能力	高	高	低	低
多维力表征能力	六维力+量化的9种人类触觉参数	有一定切向力表征能力但较难直接定量测量	无	无
模组集成度	高集成度 测、算、传一体	低集成度， 需上位机提供图像处理算力	低集成度， 需外置信号采集处理装置	低集成度， 需上位机提供时序信号处理算力
成本	全集成，成本低	需高成本采集卡+视觉算力	需高成本采集卡	需高成本采集卡

来源：帕西尼官网，国金证券研究所

全球市场以外资品牌为主，行业集中度高。全球柔性触觉传感器市场主要以外资品牌为主，其中 CR5 为 57.1%，头部厂商包括 Novasentis、Tekscan、Japan Display Inc. (JDI)、Baumer、Fraba。国内柔性传感器厂商以帕西尼科技、汉威科技、钛深科技、弘信电子等企业为代表，近年来迅速发展。

图表71：2023 年全球柔性触觉传感器行业集中度高



来源：华经产业研究院，国金证券研究所

国内目前触觉传感器技术领先的企业主要为创业公司(钛深科技、柔宇科技、苏州慧闻)，上市公司主要为汉威科技旗下的苏州能斯达。


图表72：触觉传感器代表厂商梳理

名称	概况	主要产品
国内厂商		
帕西尼	为业内首家实现多维度大阵列触觉传感器及机器人产品批量商业化公司。公司技术来源来自日本早稻田大学菅野实验室，该实验室在机器人自动化和传感器领域拥有数十年丰富经验，发布了数款行业前沿的人形机器人、及灵桥手等产品	具有多维触觉人形机器人、多为触觉灵巧手、ITPU 多为触觉传感器等多个产品
苏州能斯达	具有稳定的纳米敏感材料体系：柔性压阻、柔性压电、柔性温湿度、柔性电容四大核心技术；一条年产 1000 万支柔性传感器的超净印刷线和组装线	柔性压力传感器、柔性压电传感器、柔性织物、柔性应变传感器、柔性温湿度传感器、柔性热敏传感器、柔性电容传感器
钛深科技	开发并拥有全球最灵敏、最柔性的触觉传感技术-柔性离电式传感技术 (FITS)、能够提供实时的、高质量的、低噪声的触觉/压力信号、并拥有全柔性、光学透明及超薄封装等物理特性	脉搏波测量、压力分布测量、柔性压力传感器、触觉传感器、压力阵列传感器、压力薄膜传感器、压力分布测量系统、压电式压力传感器
柔宇科技	成立于 2012 年，拥有自主研发的核心柔性电子技术，推出全球最薄彩色柔性显示屏、建立全球首条全柔性显示屏大规模量产线	生产全柔性显示屏和全柔性传感器
苏州慧闻	成立于 2014 年，是国内少数能够提供人工嗅觉解决方案的智能传感器企业	柔性薄膜压力传感器、气体传感器
国外厂商		
Tekscan	成立于 1987 年，公司拥有薄型触觉传感器技术，专注于超薄、柔性压力传感器，具有极高的分辨率和超薄性，既可用作测试和测量工具，也可用作嵌入式传感组件	压力传感器 5051、压力传感器 3000、压力传感器 7202、压力传感器 6300
JapanDisplay	成立于 2002 年，从事中小型显示设备及相关产品的开发、生产和销售，通过结合静电容量式玻璃指纹传感技术和柔性显示屏技术，成功开发静电容量式柔性指纹传感器	透明传感器、柔性 OLED 面板
Canatu	成立于 04 年，是 3D 可成型和可弯曲透明导电薄膜和触摸传感器的开发商和制造商	汽车 3D 触摸控制面板
Sensei	成立于 13 年，开发了一种多点触控输入设备，能感应施加在各个点上的压力	压敏输入设备
Forciot	成立于 2015 年，可穿戴设备、汽车和物流行业的压力传感器组件供应商，传感器可以通过集成来监控力分布、负载平衡、运动等参数。适用于运动状态跟踪、物流和汽车跟踪，以优化负载平衡并防止机器组件颤动等。	基于物联网的压力传感器

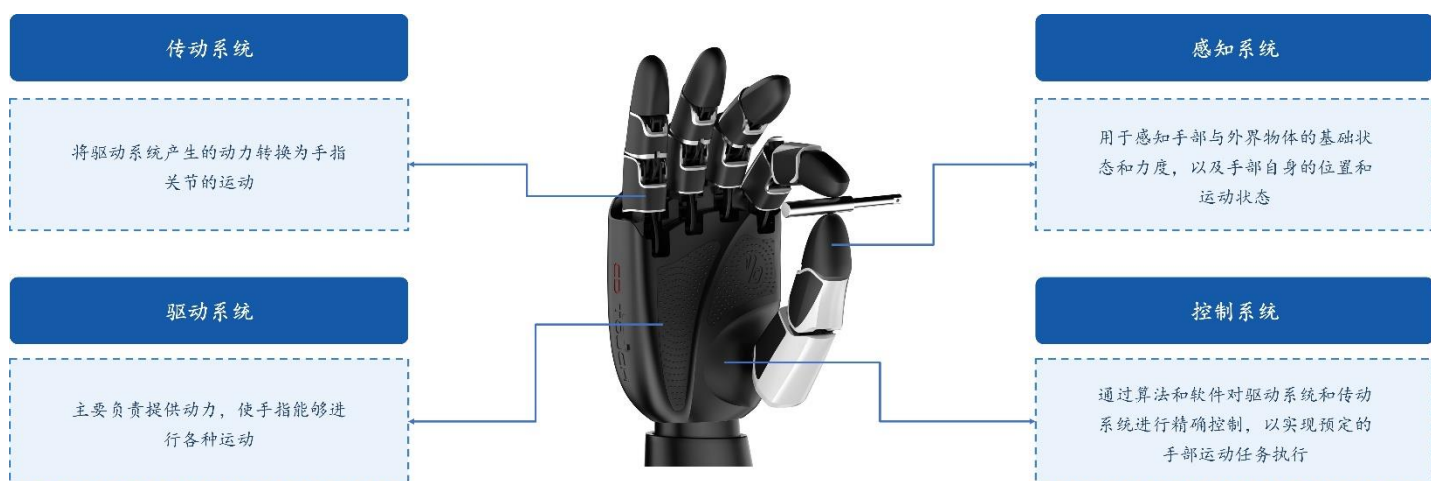
来源：传感器专家网，CINNO，人形机器人场景应用联盟，国金证券研究所

灵巧手：向高承载力、柔性感知方向发展

灵巧手的设计灵感来源于人类手部的复杂结构和功能，它使得机器人能够执行诸如抓取、操纵、甚至感知等多样化任务，灵巧手的成熟度决定了其融入人类世界的深度。典型灵巧手由驱动系统、传动系统、感知系统和控制系统构成。目前灵巧手的技术迭代方向往高灵活度、柔性感知、智能化方向发展。



图表73：机器人灵巧手的主要构成



来源：《Towards Robotic Dexterous Hand Intelligence: A Survey》，国金证券研究所

■ 驱动方式：空心杯电机是灵巧手电机主流方案，无刷有齿槽电机是降本方案。

电机驱动具有体积小、响应快、精度高和输出稳定等优势，更适用于灵巧手。驱动系统负责为灵巧手提供动力，主要驱动方式包括电机驱动、液压驱动、气压驱动和形状记忆合金驱动。液压驱动输出能力较强，但系统需配置液动机、伺服阀、油泵和油箱等部件，更适用于工业机械手及大型抓取作业；气压驱动具备轻量化、动作快、成本适中及维护简便等特点，但可操作性和轨迹精度存在不足；形状记忆合金驱动响应迅速、位移较大，适用于小型高精度装配，但持续工作能力和疲劳强度仍构成约束。整体而言，不同驱动方案在负载能力、控制精度、系统复杂度、体积和耐久性之间各有取舍，灵巧手驱动系统的升级方向并非单纯提高输出能力，而是推动驱动、传动、感知与控制环节协同优化，在有限空间内兼顾功率密度、控制精度和运行可靠性。

图表74：机器人灵巧手主要驱动方式分类和适用场景

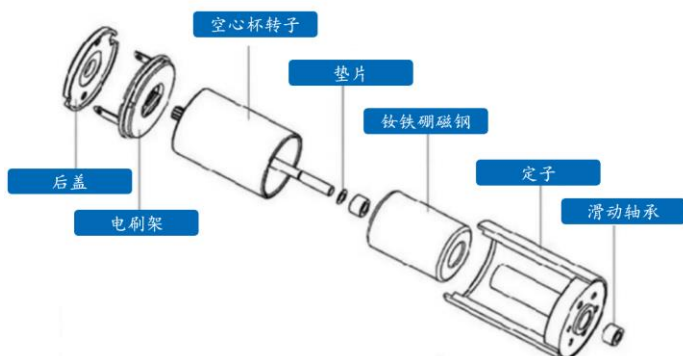
驱动方式	系统构成或驱动原理	优势	劣势	适用场景
液压驱动	驱动系统一般由液动机、伺服阀、油泵和油箱等组成，通过液压系统驱动机械手完成任务	输出能力强，适合较大负载	系统组成较复杂，需配置油泵、油箱等部件	工业机械手、大型抓取作业
电动机驱动	通过电动机输出动力，并配合传动与控制系统实现灵巧手运动	体积小、响应快、调控方便、稳定性好、精度高、输出力矩稳定	需配套减速器与驱动器、成本与质量偏高、低速大扭矩场景效率下降	对体积、精度和控制性能要求较高的灵巧手
气压驱动	利用压缩空气驱动执行机构运动	操作方便、质量轻、动作迅速、价格适中、维护简便	可操作性不强，轨迹精度不足	对成本和轻量化较敏感、精度要求相对较低的场景
形状记忆合金驱动	利用形状记忆合金受热或冷却时产生的形变与恢复力实现驱动	可负载驱动，响应快、位移大、变位迅速	无法长时间工作，疲劳强度较低	小型、高精度机器人装配作业

来源：《机器人灵巧手研究综述》，国金证券研究所

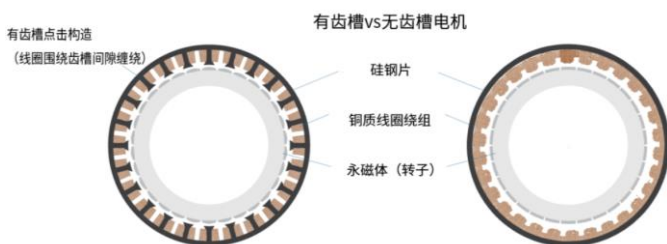
电机驱动作为最适合灵巧手批量生产使用的方式，其使用的电机可分为无刷有齿槽电机和无刷无齿槽电机，大部分直流无刷电机都是有齿槽设计，空心杯电机属于无齿槽电机，具备小直径、扭矩波动小等特点，是当前机器人灵巧手电机的主流方案。无刷有齿槽电机直径较空心杯电机更大，可部分替代空心杯电机在灵巧手中的应用，是可行的灵巧手降本方案。



图表75：空心杯电机采用无齿槽结构



图表76：有齿槽电机及无齿槽电机结构图



来源：万泰电机官方微信公众号，国金证券研究所

来源：科尔摩根官方公众号，国金证券研究所

■ 传动方式：微型丝杠+腱绳复合传动或成未来灵巧手主流方案。

灵巧手按照传动方式可分为腱绳传动、连杆传动、齿轮传动和带传动。机器人灵巧手抓取的稳定性和灵活性指标主要取决于传动系统，按照传动方式不同，可将灵巧手分为腱绳传动、连杆传动、齿轮传动和带传动。连杆传动结构紧凑、负载能力较强，但抓取空间和控制方式存在局限；齿轮传动效率高、可靠性强，但会增加手部质量和惯量；带传动结构简单、成本较低，适合大轴间距和多轴传动；线绳传动能够降低末端负载、提高抓取速度，适合狭小空间和多自由度设计，但仍面临负载能力不足、预紧力波动及重载效率下降等问题。总体看，传动方案的选择需要在负载能力、自由度、空间、效率和可靠性之间进行权衡。

图表77：机器人灵巧手主要传动方式分类

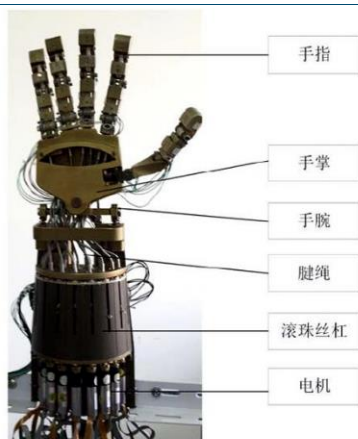
传动方式	核心优势	主要局限	适用场景
连杆传动	结构紧凑，可传递较大的运动和动力，能够抓取大型物体并完成包络抓取	远距离控制较困难，容易发生弹射，抓取空间较小	工业、商业用途及大物体抓取
齿轮传动	传动比稳定、传动效率高、可靠性强	齿轮自重会增加整体质量和转动惯量	对传动稳定性和可靠性要求较高的场景
带传动	结构简单、传动平稳，具有缓冲作用；价格较低，无需润滑，维护方便	传动比不准确、张紧力需维护、过载易打滑	大轴间距、多轴之间的动力传递
腱绳传动	模拟人手肌腱结构，可将大型驱动器远离末端，降低负载和惯量，提高抓取速度；布置灵活	负载能力较弱，预紧力变化较大，负载越大传动效率越低	空间狭小、驱动自由度较多的灵巧手

来源：《机器人灵巧手研究综述》，国金证券研究所

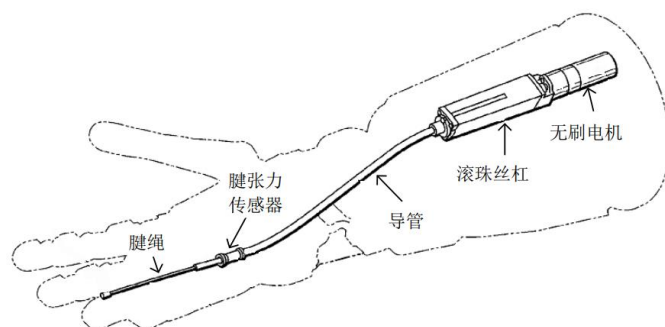
腱绳传动凭借轻量化、远端驱动和布置灵活等优势，成为灵巧手中应用最广的传动方案。而微型丝杠新型传动方案具备高承载、高效率、高精度、高可靠性等优点，未来微型丝杠+腱绳复合传动或成主流方案。2024年11月28日，Tesla Optimus在X平台官方账号发布旗下 Optimus 机器人接球视频展示其新一代手臂和灵巧手，掌内所有驱动系统已移至前臂，通过腱绳驱动手指。根据《空间五指灵巧手控制系统设计_韩运峰》，电机和滚珠丝杠外置于手臂中，电机通过减速器带动滚珠丝杠，电机轴的旋转运动被转化为丝杠螺母的平移运动，丝杠螺母拉动腱绳，腱绳另一端连接到手指指骨上，拉动手指关节绕关节轴旋转，形成手指弯曲运动。



图表78: 腱绳传动灵巧手整体机构



图表79: 腱绳+丝杠复合传动 (单自由度视图)



来源:《空间五指灵巧手控制系统设计_韩运峰》, 国金证券研究所

来源:《空间五指灵巧手控制系统设计_韩运峰》, 国金证券研究所

■ 感知系统: 电阻式和电容式触觉传感器有望成为主流。

灵巧手感知系统可使用的传感器包含力、位移、触觉、角位移、视觉、压力、力矩、位置、霍尔和温度传感器等。根据特斯拉 GEN2 发布会, 特斯拉机器人上确定搭载了触觉传感器。触觉传感器赋予灵巧手触觉感知能力, 通过接触测量物体的物理特征, 从而实现对周围环境的感知。根据原理不同, 触觉传感器分为压阻式、电容式、压电式和光学式等, 最常见的是压阻式、电容式和压电式触觉传感器。通常, 电阻式适用于大面积覆盖, 可用于手臂和身体部位; 电容式适用于高精度场景, 可用于指腹以辨别纹理; 热辐射式可布局在指缝间感知温度。

图表80: 特斯拉 OptimusGen2 灵巧手搭载触觉传感器轻松拿起鸡蛋



来源: X 平台 TeslaOptimus 官方账号, 国金证券研究所

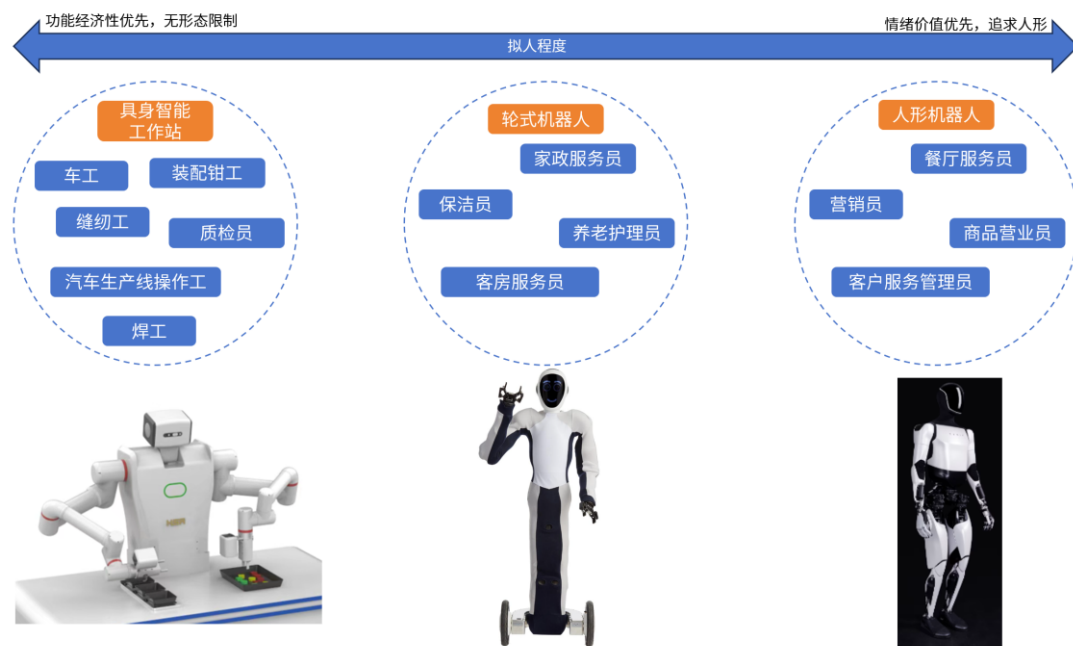


3.再往后怎么看？短板补齐后，物理 AI 将打开全新应用场景

3.1 本体不会被跳过，机械臂/人形/非人形机器人均有放量潜力

在工业场景中有望看到多形态的机器人并行推进落地，例如在无法直接高效应用工业机器人的柔性化制造业场景，以垂直大模型+工业机器人为基础，快速推广应用，例如解决服装行业始终为劳动密集型行业的情况。

图表81：工业场景中工业机器人+大模型打造的工作站或将成更直接的物理 AI 落地场景



来源：华中数控微信公众号，1X 官网，特斯拉官网，国金证券研究所

目前工业机器人基本上能覆盖全部常见制造业工序如搬运、上下料、装配、检测等，但在较多细分行业的覆盖场景和应用深度有限，我们认为主要是因为工序较为柔性，难以实现较好的低成本规模化应用。根据艾瑞咨询研究院的统计，塑料化工、食品加工、家电等行业有较多工序工业机器人应用成熟度较低

图表82：工业机器人目前在较多细分行业的覆盖场景和应用深度有限

		生产									检测	包装	出货	
行业应用成熟度		搬运	上下料	切割	冲压	焊接	打磨	喷涂	洁净/清洁	分拣	装配	检测/质检/巡检	包装/贴标	码垛
成熟行业	汽车	★★★	★★★	★★	★★	★★★★		★★ (涂装)			★★ (总装)			
	3C	★★★★	★★★★			★★★	★★			★★★	★★★	★★★		
半成熟行业	金属加工	★★★	★★★	★★		★★★★					★★			★★★
	塑料化工	★★★							★★★			★★ (巡检)		★★★
	食品加工			★★						★★★			★★	★★★
潜力行业	锂电	★★★	★★★★			★★★★					★★★	★★★	★★★	★★
	光伏		★★★			★★			★★★ (清洁)	★★★	★★★	★★	★★	
	家电	★★★			★★	★		★					★★	★★
	医药制造	★★								★★★		★★★	★★★	★★
注：1) 应用成熟度指该行业该场景中工业机器人的渗透情况，即对比只在同一行业中才有意义，由1星到4星成熟度递增。2) 灰色底纹代表应用较少、或行业中不太重要的工艺场景。3) 绿色底纹代表行业中较为规范、标准的机器人应用场景。4) 蓝色底纹代表行业中较为重要、且具有成长空间的应用场景。														

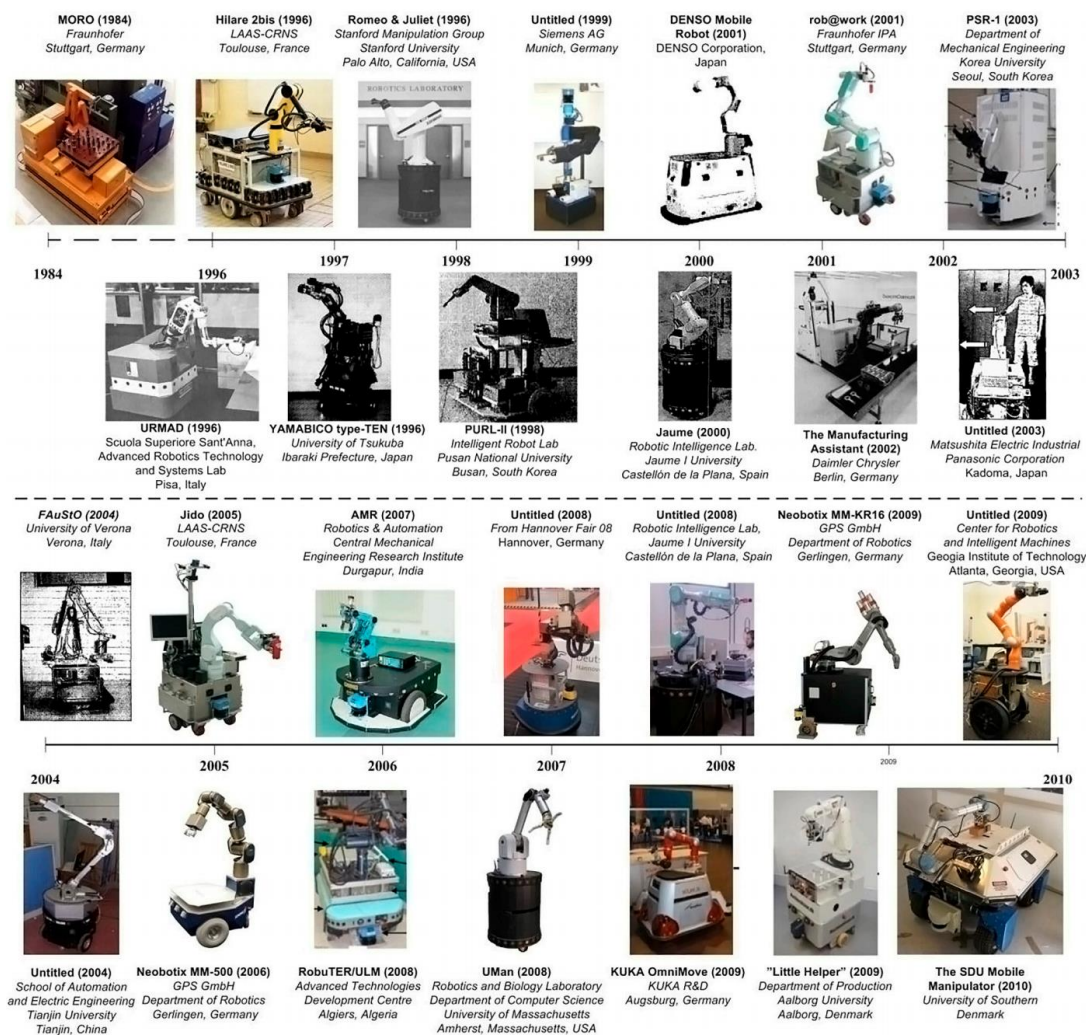
来源：艾瑞咨询研究院，国金证券研究所

以移动机械臂为例，主要由一台移动平台以及安装在平台上方的单条或多条机械臂组成。该系统搭载配套的传感器、驱动器与作业工具，能够感知周边环境、安全自主地完成移动，



并可通过在机械臂末端执行器更换不同工具来执行对应任务。移动机械臂现已落地众多领域：农业、军事、核工业、航天、矿山、医疗陪护、家庭服务、搜救、工业场景。其中 49% 的应用在物流行业、33% 应用在制造业。

图表83：移动机械臂已经有几十年的迭代历史



来源：《Mobile Manipulators in Industry 4.0: A Review of Developments for Industrial Applications》，国金证券研究所

Franka Robotics 推出的移动机械臂 Mobile FR3 Duo 把双臂操作、多模态感知、全向移动能力集成至单一具身智能系统，能够脱离固定工作单元，完成数据采集与策略执行，成为了物理 AI 落地理想的介质。



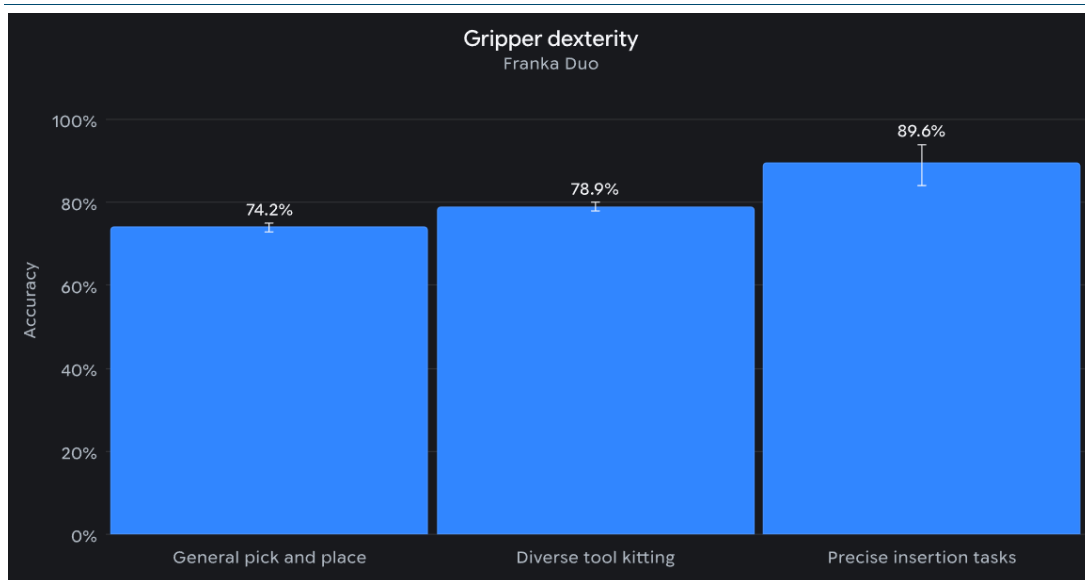
图表84: Franka Robotics Mobile FR3 Duo



来源: Franka Robotics 官网, 国金证券研究所

基于 Gemini Robotics 2 基准测试数据, Franka Duo 在精密插入、工具分拣、通用抓放等物流、制造业常见场景中, 已经逼近了生产级能力。也就是商用 VLA 已经在驱动机器人实际走向部署。

图表85: 搭载商用 VLA 模型的移动机械臂已经逼近了生产级能力



来源: Google DeepMind 官网, 国金证券研究所

3.2 物理 AI 有望重塑制造业, 仓储物流/服装行业或率先落地

从应用端看, 目前物理 AI 已经开始重塑生产线、仓储、移动服务以及科学实验室, 但相比于追求通用人工智能, 主流方案是采用专用硬件+基础大模型+仿真基础设施+人工监管的组合架构。

图表86: 物理 AI 已经开始重塑制造、物流等工业领域

细分领域	机构/企业	技术栈	实测成果
制造与装配			



细分领域	机构/企业	技术栈	实测成果
汽车	BMW+Figure	Figure 02、VLA	多年试点；物料搬运
汽车	Tesla	Optimus（擎天柱）	电池搬运；内部部署
零售/服务	Sanctuary AI	Phoenix、灵巧操作	店内后台运营；SKU 处理
科研	宇树科技	R1（5900 美元）	科研实验室；全身控制
仓储与物流			
履约仓	Amazon	Isaac Sim、基础模型（FMs）	100 万+机器人；工作流验证
拣选	Covariant	RFM-1 VLA	新 SKU 识别、杂乱料箱拣选
配送	DHL	VLA 拣选	吞吐量提升 40%；低差错率
分拣	FedEx	自主移动机器人（AMRs）+ AI	10 万+包裹/小时；错分率低于 1%
自主移动与现场作业			
自动驾驶	Wayve	RouteDrive、端到端（E2E）	车队学习：合作方 Asda、Post、Nissan
农业	John Deere	自动驾驶拖拉机、视觉-语言模型（VLM）	厘米级 GPS；自然语言任务指令
建筑	t Robotics	四足机器人、挖掘机	勘测成本降低 60%；安全性提升
实验室与医疗			
实验室自动化	GAMORA	LLM 液体处理机器人	闭环实验；自然语言目标设定

来源：《Physical AI: The Next Frontier in AI and Robotics to Build Truly Autonomous Machines》，国金证券研究所

根据《Physical AI: The Next Frontier in AI and Robotics to Build Truly Autonomous Machines》提出的四阶段成熟度框架，真正已规模化落地的是物流仓储（AMR/机械臂，非人形）。

图表87：物理 AI 落地四阶段框架

阶段	适用行业	过渡判定标准	核心特征	主要障碍
阶段 1：规模化落地	物流、仓储	部署 > 100 台设备；连续稳定运行 > 12 个月；具备可验证的投资回报	部署数百-上千台设备；7×24 小时不间断作业	基础设施成本；专业技术人才缺口
		ROI		
阶段 2：试点验证	制造业（人形机器人、协作机器人）	部署 > 10 台设备；试点测试 > 3 个月；合作方验证任务有效性	部署数十-上百台设备；有人监督；执行定向专项任务	可靠性、安全认证、系统集成难度
阶段 3：合规监管	自动驾驶车辆	安全方案获得认可；取得有限商用牌照	技术层面就绪；小规模商业运营	责任界定、公众接受度
阶段 4：科研研发	农业、建筑、医疗健康	公开展示原型；至少 1 次真实场景试验	原型演示、小范围场景部署；高度依赖人工监管	非结构化环境、经济性成本问题

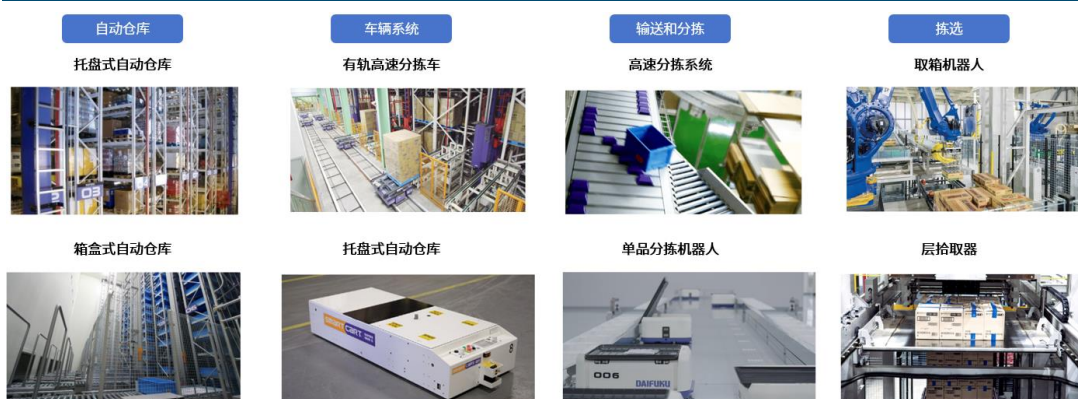
来源：《Physical AI: The Next Frontier in AI and Robotics to Build Truly Autonomous Machines》，国金证券研究所

3.2.1 仓储物流：已实现工业自动化升级，机器人将是第二波

物流行业经过多年自动化升级，整体已经有了较高的自动化程度，尤其是仓储、传送、分拣等环节。



图表88：物流行业整体已经有了较高的自动化程度



来源：大福官网，国金证券研究所

在此基础上，物流装备公司产品持续迭代，推出了壁虎料箱机器人系统、自动播种机等进一步提高自动化程度的产品。

图表89：兰剑智能壁虎料箱机器人系统优化货到人拣选



来源：《我国机器人产业最新变化发展纵览》，国金证券研究所

图表90：德马科技智能播种机器人缩减二次分拣人工作业



来源：德马科技官网，国金证券研究所

但在物流作业场景中，仍有大量环节需要人工参与，尤其是拆零拣选环节，在物流中心是作业成本最高、人力耗费最大、时间占用最多的工作，工作量约占整个仓储作业的 40%以上。

图表91：拣选尤其是拆零拣选是人力耗费最大的环节

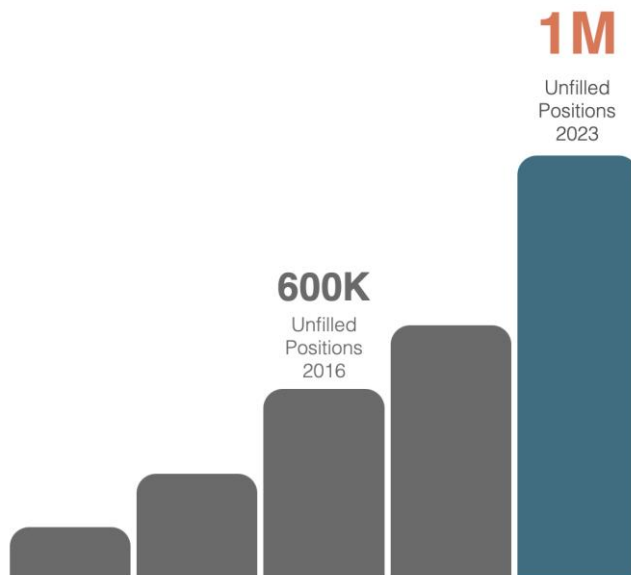




来源：大福官网，国金证券研究所

根据 Agility Robotics 数据，美国仓储、物流和制造行业有超过 100 万个材料搬运岗位空缺，机器人可以有效填充这些空缺。

图表92：美国仓储、物流和制造行业有大量材料搬运岗位空缺



来源：Agility Robotics 官网，国金证券研究所

Agility Robotics 人形机器人最早在物流领域与 GXO、亚马逊等企业开展合作，在真实仓储环境应用并产生了收入，亚马逊的机队规模已经扩大到了 75 台。后续进一步拓圈，与舍弗勒、丰田等企业开始机器人落地项目，Digit v5 机器人已经拿到了超 3 亿美元的多年期订单。

图表93：Agility 人形机器人从物流领域起步逐步向其他行业拓展

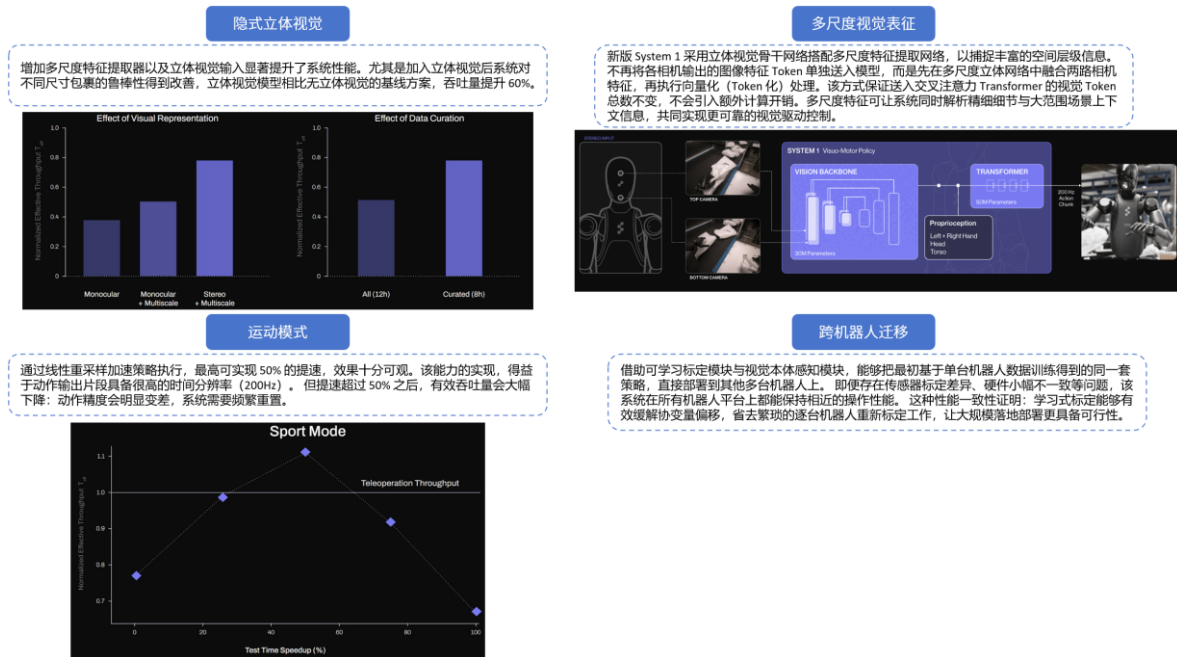
重点客户	进展
GXO Logistics	2024 年 6 月签订全球首个多年期人形机器人 RaaS 协议，到 2025 年 11 月宣布累计搬运 100000+料箱，任务主要包括从 AMR 取料箱、转送上输送线、地面位置堆叠容器
亚马逊	2023 年起在西雅图机器人研发设施测试，到 2026 年 4 月扩展部署至 75 台，主要用于空箱整理回收
舍弗勒	24 年宣布成为战略投资者兼客户，计划 2030 年前在全球约 100 家工厂部署
丰田	26 年签署了付费 RaaS 协议，在成功试点后计划部署机器人来协助制造、供应链和物流。

来源：singularity、aiwiki、agility 官网，国金证券研究所

Figure 一直在物流领域推动人形机器人模型的技术迭代，主要面向物流包裹抓取与分拣场景，其发布的 Helix VLA 模型统一视觉、语言理解与学习式控制。为了物流行业的针对性优化，Helix System 1 底层视觉运动控制策略进行了一系列通用控制优化，从而满足物流应用中面临的包裹材料形态各异、响应时间短、难以仿真、需要自我纠错等痛点。



图表94: Figure Helix 进行了一系列通用控制优化满足物流行业需求



来源：Figure 官网，国金证券研究所

除物流场景外，Figure 也正在进一步拓圈进入汽车领域，宝马和 Figure 的合作进一步加深，从 2025 年初 3 台试点，到 2026 年 5 月官方宣布进入“full production deployment”（100 台、两班制）。任务节拍周期 84 秒、零件装载耗时 37 秒；单班次放置成功率 $\geq 99\%$ ，目标实现零人工干预。

图表95: Figure 02 在宝马工厂真实产线测试数据

指标	详情
供应商（斯帕坦堡工厂）	Figure 02，后续更换为 Figure 03 / Figure AI
Figure02 部署周期	约 11 个月
累计运行时长	约 1250 小时
搬运零部件总数	90000 件以上
配套生产整车数量	30000 台以上宝马 X3
初始作业任务	焊接用钣金件上料/插入
Figure03 工作内容	物流序化分拣（分拣装入物料台车）

来源：industrialsage，国金证券研究所

星动纪元已推出具身物流解决方案，实现了一个型号的具身机器人一机通用，适配小件供包/智能打包/工作站拣选全链路物流作业环节。



图表96: 星动纪元实现单型号本体供包/打包/拣选全链路作业



来源: 星动纪元微信公众号, 国金证券研究所

2026年8月12日, 自变量机器人完成了一场持续1小时的全自主物流分拣直播, 机器人依托自研的世界统一模型 WALL-B, 采用双机械臂搭配夹爪的硬件方案, 在全程无人工干预的情况下, 对大小、形状、材质及重量随机的真实物流包裹进行连续分拣, 实现 1816 件/小时的分拣效率, 准确率超过 98%。同时, 机器人+“五指灵巧手”的复杂硬件方案, 自变量使用更加简洁的双机械臂和夹爪, 整套成本下降约 70%, 同时能够处理更加随机、复杂的真实包裹。更贴合真实的物流分拣场景。

图表97: 自变量机器人物流分拣效率/准确率表现优异

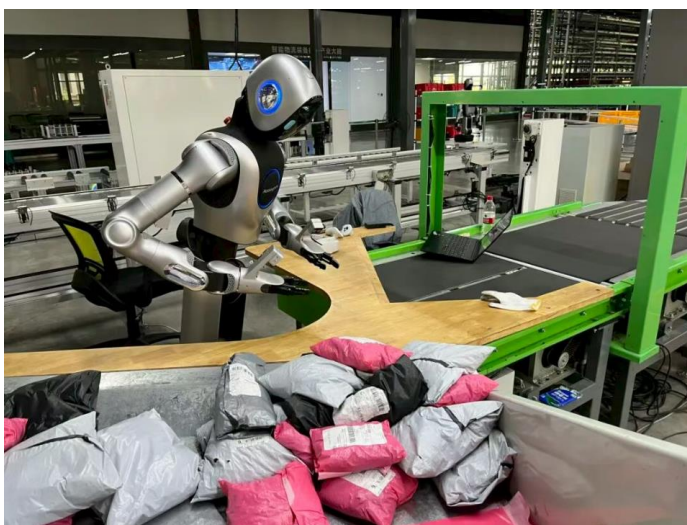


来源: 新华网, 国金证券研究所

德马科技目前一方面和优必选/蚂蚁灵波等企业合作, 共同推动人形机器人技术在物流等核心场景的规模化落地, 一方面依托“本体+EGO”技术体系, 持续采集、沉淀人形机器人在物流场景下的高质量具身数据集。相关数据一方面用于自研物流垂域模型迭代, 持续优化机器人场景作业能力; 另一方面稳步推进具身数据服务对外商业化, 已取得稳定的数据服务订单。



图表98：德马科技打造物流场景应用与数采同步布局

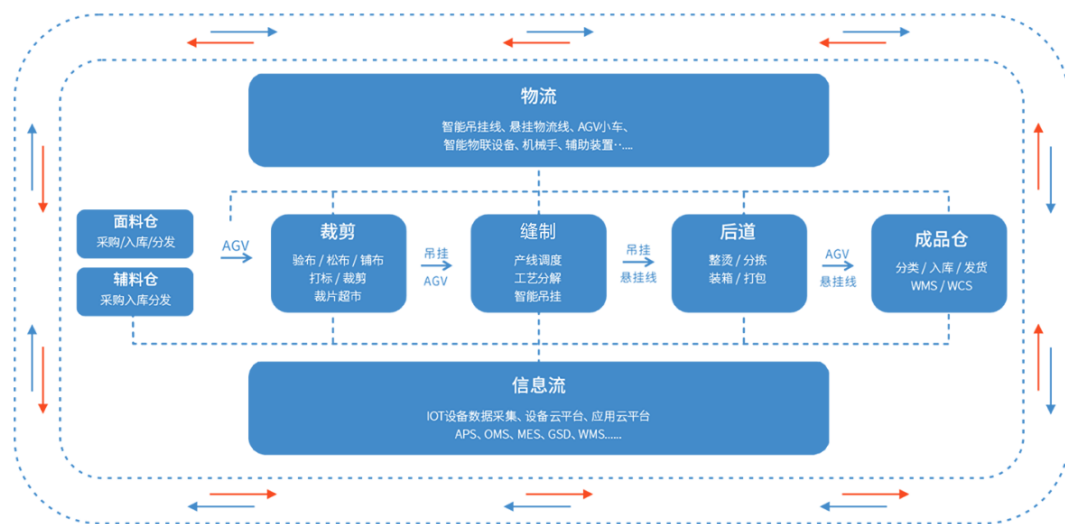


来源：AGV 网微信公众号，国金证券研究所

3.2.2 服装：主要针对缝纫环节，解决招工难、人工成本高等下游核心痛点

目前服装厂的成衣制造工序可大致归类为缝前、缝中、缝后环节，缝前主要通过使用铺布机、裁床完成开片，均可采用大型的自动化设备完成；缝中主要通过使用工业缝纫机进行缝纫加工，需要大量人工参与，也导致服装产业为劳动密集型产业，目前即使是使用模板机之类的自动化设备，上下料也需要较多人工参与；缝后主要是整烫、分拣、打包等，目前较多企业开发了智能吊挂线等，实现了较高的自动化水平。

图表99：缝中环节是服装厂实现全面自动化的瓶颈



来源：杰克科技官网，国金证券研究所

其中的主要痛点在于缝纫的柔性化程度太高，在衣服款式/尺寸、布料种类/厚度等变量影响下基本只能通过人工手动完成。



图表100：目前服装厂可以实现缝前、缝后的高自动化，缝中主要依靠人工



来源：杰克科技官网，国金证券研究所

即使是将工业缝纫机升级为模板机，上下料的环节还是涉及较多人工操作。

图表101：模板机自动化程度更高，但上下料还是需要人工操作

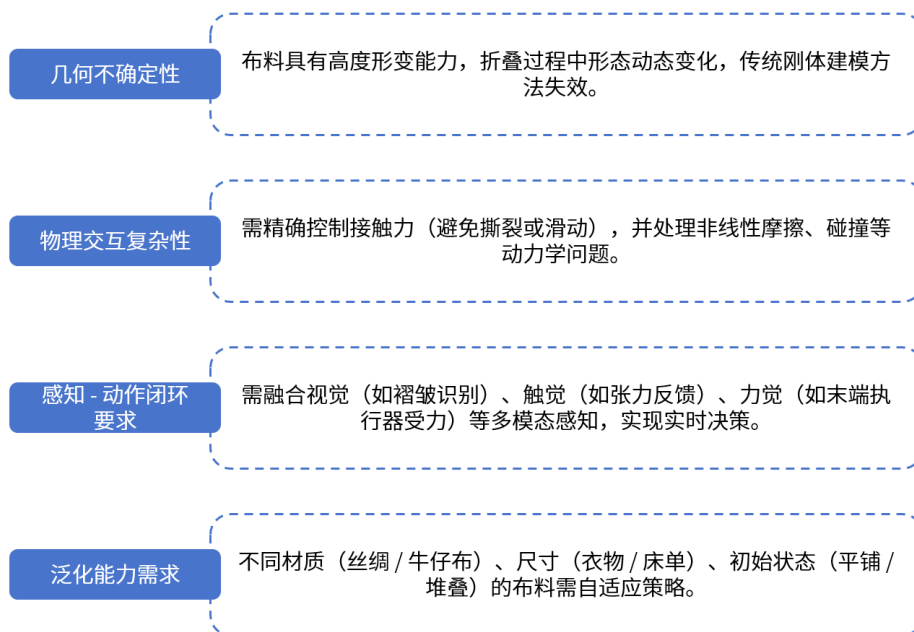


来源：杰克科技官网，国金证券研究所

从具体的机器人功能来看，主要需要的是实现布料的移动、折叠、平铺等操作，而布料属于可变形物体，物理交互复杂对大模型、传感器要求较高，同时物理特性很难精准建模。



图表102：布料属于接触丰富型柔性物体，机器人的应用具有一定难度



来源：《面向线状柔性物体的机器人操作研究进展与展望》，《Learning Foresightful Dense Visual Affordance for Deformable Object Manipulation》，国金证券研究所

杰克科技推出了 AI 缝纫机器人艾图机器人，旨在实现服装行业无人化最终目标，具备面料自适应、秒级换款、稳定可靠等优势，实现裁片分离、抓取、摆放、对齐的全流程自动化。根据艾图科技披露信息，其位于浙江台州的缝制机器人制造基地正式启动投产，这也是全球首个面向缝制机器人的专业化制造工厂，年产能规划可达 1 万台。

图表103：杰克科技服装机器人工厂已经正式投产

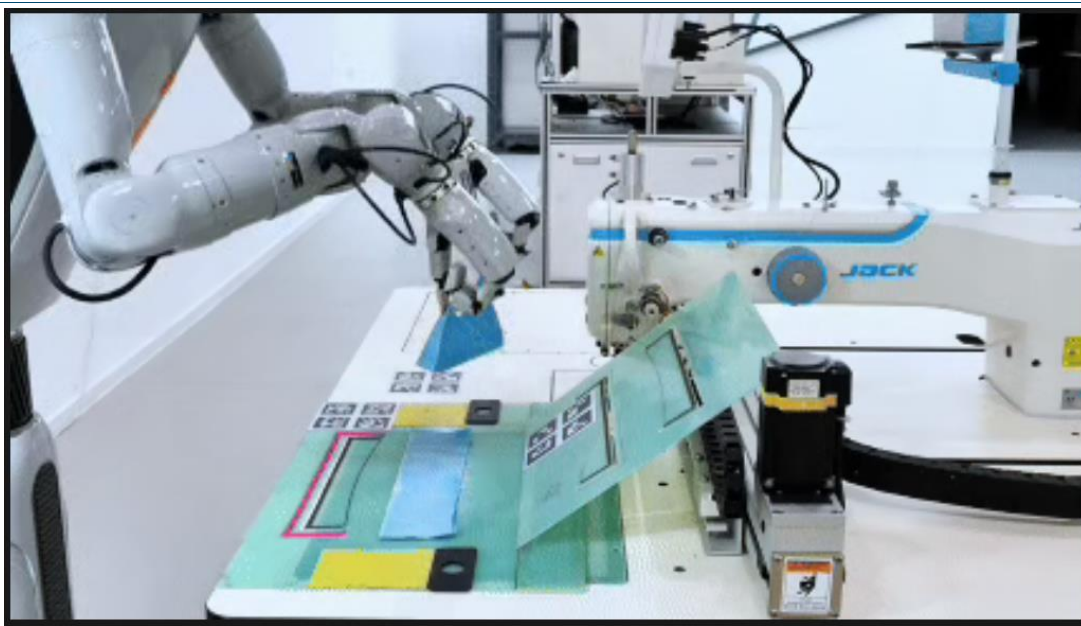


来源：Aitu 艾图微信公众号，国金证券研究所

面对服装面料材质多样、形态不定、易皱易偏，衣领、口袋等裁片对位偏差需控制在 $\pm 2\text{mm}$ 以内，而裁切、缝纫等核心工序对设备动作精度要求达 0.3-0.5mm 的高难度场景，目前部分场景合格率已经可以达到 100%人工水平、效率可达人工 50%以上。



图表104: 杰克在部分场景合格率已经可以达到100%人工水平、效率可达人工50%以上



来源: 浙江人形机器人创新中心微信公众号, 国金证券研究所

同时杰克科技目前已经获得了恒锋服饰等企业的意向采购订单, 产品正式发布后有望看到快速放量。

图表105: 杰克与恒锋服饰签订200台意向订单



来源: Aitu 艾图微信公众号, 国金证券研究所

图表106: 杰克与浙江人形签约2000台意向订单



来源: 浙江人形机器人创新中心微信公众号, 国金证券研究所

4.投资建议

物理AI 奇点时刻到来, 建议一方面关注核心短板数据采集/传感器/灵巧手等环节奥比中光/奥普特等, 一方面关注仓储物流/服装机器人场景应用标的杰克科技等。

图表107: 建议关注公司

环节	证券代码	股票名称	市值 (亿元)	归母净利润 (亿元)					PE				
				2024A	2025A	2026E	2027E	2028E	2024A	2025A	2026E	2027E	2028E
数据采集	688686.SH	奥普特	180	1.36	1.86	2.44	3.98	4.88	132	97	74	45	37
	688583.SH	思看科技	96	1.21	0.96	1.06	1.91	2.82	80	100	91	50	34
	688322.SH	奥比中光	441	-0.63	1.28	3.18	6.16	9.51		345	139	72	46
	605488.SH	福莱新材	133	1.39	0.77	1.16	1.72	2.45	95	172	114	77	54
	300007.SZ	汉威科技	119	0.77	1.59	1.51	1.80	0.00	155	75	79	66	
	688071.SH	华依科技	46	-0.47	-0.62	0.30	0.75	0.00			154	62	
	603662.SH	柯力传感	139	2.61	3.41	3.26	3.91	4.51	53	41	43	36	31



灵巧手	603341.SH	龙旗科技	196	5.01	5.85	7.32	10.76	14.72	39	33	27	18	13
	688306.SH	均普智能	150	0.08	0.12	0.76	1.10	1.75	1825	1224	197	137	86
	300307.SZ	慈星股份	81	2.84	0.97	3.02	4.26	5.77	29	83	27	19	14
	300115.SZ	长盈精密	386	7.72	5.98	9.32	13.44	15.39	50	65	41	29	25
	002009.SZ	天奇股份	70	-2.55	0.55	2.37	3.26	4.09		128	30	22	17
	688787.SH	海天瑞声	82	0.11	0.14	0.35	0.57	0.87	720	578	231	142	94
	002611.SZ	东方精工	163	5.01	7.25	38.27	3.01	3.46	33	22	4	54	47
	003021.SZ	兆威机电	231	2.25	2.54	3.08	4.05	4.86	102	91	75	57	47
	002979.SZ	雷赛智能	176	2.00	2.25	4.07	5.85	8.16	88	78	46	35	27
	002139.SZ	拓邦股份	121	6.71	3.64	5.84	6.55	7.51	18	33	21	18	16
本体/应用	603337.SH	杰克科技	211	8.12	8.56	10.70	13.02	15.56	26	25	20	16	14
	688836.SH	宇树科技	2588	0.95		6.68	10.88	16.24	2711	930	388	238	159
	688585.SH	上纬新材	623	0.89	0.41	-0.13	-0.92	3.57	703	1517			175
	688360.SH	德马科技	54	0.93	0.95				58	56			
	603611.SH	诺力股份	40	4.62	4.29	6.03	7.26	7.70	9	9	7	5	5
	688557.SH	兰剑智能	27	1.12	1.09	1.42	1.89	2.51	24	25	19	14	11
	603901.SH	永创智能	60	0.16	1.37	2.70	3.25	3.96	387	44	22	19	15
	002380.SZ	科远智慧	69	2.52	1.39	3.89	4.74	5.53	27	49	18	14	12

来源：iFinD，国金证券研究所；除奥普特/奥比中光/杰克科技/雷赛智能外盈利预测取自 iFinD 2026 年 10 月 26 日一致预期

5.风险提示

大模型技术迭代不及预期风险：VLA 的能力边界取决于模型/数据/算力协同，若主流技术路线迭代在长程任务成功率与跨本体泛化上迭代速度缓慢，或 Astra 这类“模型外部”路线最终在真实工况下经济性不及预期，均将导致物理 AI 产业化进展延后。

数据采集技术路线变化风险：若出现第一视角人类视频叠加世界模型导致真实动作采集占比下降/无本体可穿戴采集替代固定采集工位/合成或仿真数据有效性进一步提升，则高投入形成的采集资产存在利用率不足与减值风险。

终端应用不及预期风险：我们判断仓储物流和服装产业有望率先落地，但终端对投资回报依旧高度敏感，若整机价格下行/效率准确率提升不及预期，或下游资本开支因宏观需求走弱而收缩，放量节奏或低于预期。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国内大街 26 号

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究