

自动化设备行业点评报告

人形机器人思考-专题 2：我们离真正的具身智能大模型还有多远？

增持（维持）

2025 年 02 月 27 日

证券分析师 周尔双

执业证书：S0600515110002

021-60199784

zhouersh@dwzq.com.cn

投资要点

■ **背景：**2.20 日 Figure AI 发布 Helix VLA 大模型引爆市场关注，但我们发现市场对于具身智能大模型的认知仍待提升，本文希望通俗易懂的讲明白，我们需要什么样的具身智能大模型？我们离真正的具身智能大模型还有多远？

■ 问题一：具身智能大模型（VLA）是什么？

VLA (Vision-Language-Action) 大模型指的是视觉-语言-动作大模型，能够让机器人通过理解环境和语言指令，并通过执行模块输出为动作。

■ 问题二：VLA 大模型中分层和端到端模式的区别？业界目前的选择？

VLA 模型在执行过程中，一般有三个步骤——1) 接收并理解语音和图像输入；2) 根据接收的信息做推理决策；3) 根据决策生成动作指令并控制机器人运动。简单来说，如果这三个步骤是在一个模型中完成的，则是端到端大模型，如果这三个步骤分别是调用三个不同的模型完成，则是分层模型。

端到端模式的优缺点：1) 优点在于反应速度快+规模化+能够实现智能涌现；2) 缺点在于难度高+需要大量训练数据+短期难以落地。

结论&现实情况：短期视角来看，目前国内初创人形机器人企业多采用分层模式，主要是为了快速商业化落地；仅有少数企业例如特斯拉、星动纪元等坚持端到端模型，但从长期视角来看，端到端模式是最终实现具身智能涌现的必要条件。

■ 问题三：训练出好用的端到端大模型难点在哪里——瓶颈在于数据。

1) 数据量差距巨大：相较于 VLM 大模型亿条级别的数据量，目前机器人实际单一场景的训练数据量仅仅在千条和万条级别，差距百倍。

2) 机器人数据获取难度极高：相较于互联网上常见的语料供 VLM 大模型训练，机器人训练数据获取难度极高。目前有三种数据获取模式：

①真实数据遥操采集：问题在于成本极高，目前动捕设备一套价格在几十万区间，初创企业如果要是靠动捕设备遥操采集数据，成本非常高；

②虚拟生成数据：例如银河通用发布的 GraspVLA，通过虚拟仿真技术生成数据，用于机器人训练，但目前难以解决 sim-to-real gap。简单来说，就是用虚拟仿真数据训练机器人效果很差，如果是简单的抓放搬运场景，虚拟数据相对可行，但如果涉及到柔性场景，比如说衣服被子等柔性物体，就很难运用。因为涉及到柔性物体形变仿真，在物理层面本身就很难建模。

③真人数据映射：UMI 和 DexCap（斯坦福机器人团队）等正在探索真人数据映射（即采集真实人的数据，通过某种映射关系转化为机器人数据），但目前还比较早期。

3) 遥操采集的数据本身存在毒性：**①人在运动过程中会有额外的运动轨迹：**例如简单的搬箱子，人在遥操录制过程中，可能会因为外界干扰停顿几秒，但这个停顿对于机器人来说就存在毒性，因为其无法理解人为什么要停顿。**②人的运动轨迹和机器人不一致：**目前市面上大量机器人都是以旋转关节为主，而人的上下肢是直线关节，因此同样是搬箱子的动作，人和机器人的运动轨迹就是不一致的，这时候用真人的数据去训练机器人本身就存在毒性。

行业走势



相关研究

《Figure 01 超快迭代，大模型转变机器人决策逻辑》

2024-03-15

《新质生产力成为两会热门话题，人形机器人有望引领智能制造浪潮》

2024-03-06

4) 机器人本体方案未收敛导致数据难以复用：例如用特斯拉本体采集的数据很难给智元的机器人来训练，因为本体方案不同。

■ 问题四：数据端存在这么多问题，业界如何解决？

真实的情况是，业界目前还无法解决上述数据端的问题。但目前各家都选择在自己的方案上努力收集数据，先在单一场景下实现一定程度的泛化，从而让更多人形机器人投入实际运用。我们认为可能是 3-5 年后，当市场上有足够多的人形机器人数据，并且硬件方案逐步收敛，具身智能基础模型呈现出一定的智能涌现，才有可能实现真正的端到端具身智能大模型。

■ 问题五：能否采用 Deepseek 的范式来加速具身智能大模型的发展？

Deepseek 是通过 pre-train+post-train（强化学习）的模式，并且导入高质量数据，来降低大模型的算力和数据量需求。但目前来看，这种范式对具身智能大模型来说，路径是正确的，但基本要素还不具备。一方面，具身智能大模型并没有一个强大的基础模型；另一方面，也不存在一个完善的强化学习流程。学界一直在推行所谓的模仿学习+后训练强化学习的方案（类似于 deepseek 的路线），就是通过模仿学习达成 0-1，然后通过强化学习达成 1-10，但目前来看必要条件还未达成。

■ 问题六：Figure Helix 大模型详解和局限点解析：

Helix 的特点是准分层架构，用一个开源的 70e 参数的 VLM 作为大脑，然后下面结合一个 Transformer 架构的动作策略快系统，这个快系统实际只要用 8000 万参数去吸收 500h 的数据，然后使其泛化性变得足够强。

PS：简单理解就是大脑的思考完全交由 VLM 大模型，因为互联网上有很多家居的视频和数据，因此可以通过 VLM 大模型去进行解析，本身 VLM 大模型就已经有很好的泛化性。再将解析出来的指令通过快系统去执行。

优劣点解析：Helix 的优点在于：快速商业化能力，能够用很小的数据量达到很好的泛化水平；Helix 的劣势在于：1) Helix 是一种纯模仿学习，还没有上强化学习；2) 没法处理突发情况，例如一些碰撞和避障；3) 互联网上海量的数据还是集中在生活场景多一些，工业数据很少，因此 helix 可能短期内更适用于家居场景，工业场景无法使用。

■ 风险提示：机器人行业竞争加剧，新品推出速度不及预期。

免责声明

东吴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。

会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；

增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；

中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；

减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持：预期未来 6 个月内，行业指数相对强于基准 5%以上；

中性：预期未来 6 个月内，行业指数相对基准-5%与 5%；

减持：预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号
邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>