

美股科技巨头纷纷入局， AI大模型将推动机器人商业化落地

—— 海外机器人系列跟踪报告（一）



分析师

付天姿

执业证书编号

S0930517040002

联系人

黄铮

2024年3月14日



证券研究报告

- 事件：机器人赛道火爆，产品迭代迅速，美股头部科技公司纷纷入局。** 2024年2月24日，英伟达宣布成立新的研究部门——“GEAR”，加码机器人赛道；2月29日，人形机器人初创公司Figure宣布，已从亚马逊创始人贝索斯、英伟达、OpenAI和微软等公司筹集了约6.75亿美元资金，使其公司估值达到26亿美元；3月13日，Figure最新发布了一段视频，展示了机器人Figure 01在OpenAI大模型的支持下，已能够与人类对话，理解并执行人类的指令和任务。当前美股机器人赛道火爆，人形机器人初创公司在一级市场上获得高估值，高科技公司纷纷加入，AI行业与人形机器人行业融合发展加速。
- 行业进展：人形机器人硬件上突破运动控制能力，结合AI大模型实现更多的功能。**

 - 1) 人形机器人厂商逐渐突破硬件瓶颈，机器人的操作能力和移动能力得到较大提升：Figure机器人具备行走能力；特斯拉Optimus-Gen2相比第一代运动性能提升。
 - 2) 人形机器人厂商通过自研或者与AI厂商合作的方式，结合AI训练人形机器人：Figure与OpenAI合作开发下一代人形机器人的AI模型；特斯拉采用自研FSD控制系统；1X与OpenAI合作开发具身学习模型。
- AI + 机器人：人形机器人是AI的重要应用场景之一，AI大模型成为人形机器人商业化落地的关键。**

 - 1) 机器人训练方式由“传统机器学习”向“端到端”神经网络转变：①机器人的“感知→规划→执行”流程缩短为“感知→执行”，减少人为写代码的过程，但对数据提出更高要求；②“具身学习”方法引入，从真实世界中获取数据，完成对机器人的训练。
 - 2) 面向机器人的生成式人工智能软件开发：主要由谷歌、Meta、微软和英伟达等头部AI厂商主导，目前重点为优化机器人任务的基础模型 + 机器人人工智能开发平台。
- 未来展望：硬件+软件+商业化，人形机器人发展潜力巨大。**

 - 1) 硬件层面：提升零部件精度&降低重量，突破运动与灵活度的性能瓶颈；
 - 2) 量产能力和商业化：核心为降本&控制良率，产品上市进程或将加快；
 - 3) 软件算法：“具身智能”有望为人形机器人通用性赋能。
- 风险提示：技术发展不及预期；商业化落地不及预期；相关技术被不当使用或存在缺陷等**

- 1、海外头部机器人：运动控制能力有所突破，将结合AI大模型拓展功能
- 2、美股科技巨头：纷纷入局，投资+优化开发机器人赛道的生成式AI平台
- 3、未来展望：硬件+软件+商业化，人形机器人发展潜力巨大
- 4、风险提示

1、海外人形机器人头部公司

- 1.1、Figure：人形机器人初创企业，具备端到端学习能力
- 1.2、特斯拉：Optimus研发进展迅速，FSD控制系统和Dojo算力提供技术底座
- 1.3、波士顿动力：深耕多年，机器人硬件技术积累深厚
- 1.4、1X Technologies：类人机器人硬件技术与“具身学习”AI技术相结合
- 1.5、Agility Robotics：致力于实现商业化人形机器人量产

1、海外机器人头部公司

表1：海外机器人头部公司人形机器人产品参数对比

	Figure	特斯拉	波士顿动力	1X Technologies	Agility Robotics
人形机器人产品	Figure 01	Optimus; Optimus-Gen 2	Atlas	NEO	Digit
运动能力和灵活性	中，具备行走能力，移动速度1.2m/s (4.32km/h)	中，移动速度第一代约为8.05km/h，第二代约为10.46km/h	高，可以后空翻、跑酷	中，行走速度4km/h，跑步速度为12km/h	较高，能适应不平整地面、上台阶、受到撞击使自己保持稳定
手部灵活性	五指	五指	三指	五指	
合作方	软件算法：OpenAI； 微软	无	软件算法：英伟达； 资金：美国军方（早期）	软件算法：OpenAI	资金：亚马逊； 制造：福特
AI融合程度	多模态大模型识别理解端到端训练	自有FSD和Dojo		具身学习+共享自治	较弱
适用场景	通用	通用	处在实验室研发阶段	通用	物流仓储
量产时间	未进入到量产阶段	预计2025年	Atlas暂未实现量产	未进入到量产阶段 (2024年产品发布)	2024年
定价	未公布	2万美元	未公布	未公布	25万美元

资料来源：各公司官网，海量观世界，澎湃新闻，第一财经，CyberDaily，新智元，界面新闻，中国机器人网，搜狐网，前瞻网，新华社客户端，硅星人，财联社，维科网，机器人大讲堂公众号，cnBeta，华尔街见闻，极客公园，MetaPost元宇宙资讯，证券时报，爱范儿公众号，光大证券研究所整理

1.1、Figure：人形机器人初创企业，具备领先理解与学习能力

Figure：人形机器人初创企业，产品已经具备多模态理解、端到端学习能力和较为流畅运动能力。Figure是一家人形机器人初创企业，于2022年成立于美国硅谷，公司官网显示公司的目标是在全球范围内部署自主的人形工人机器人，以解决劳动力空缺的问题。公司推出的首款人形机器人产品命名为Figure 01，已经具备多项先进能力。

- 1) 高级视觉和语言智能：2024年3月13日，公司最新发布了一段视频，展示了Figure 01在OpenAI强大模型的支持下，通过与人类对话交互来执行命令。
- 根据视频，**Figure 01**具有以下能力：**a) 视觉识别和理解**：当人类问Figure 01看到了什么，Figure 01回答：“我看到了桌子中央的盘子上有一个红苹果、一个装满杯子和盘子的晾碗架，以及你站在附近，手放在桌子上。”**b) 语言识别和理解**：在人类说出需要食物的命令时递上苹果；**c) 语言解释动作原因**：解释递苹果是因为“这是我可以从桌子上为您提供的唯一可食用物品”；**d) 流畅的任务执行**：能够意识到垃圾应该放到筐子里，杯子和盘子应该归置放在沥水架上，并做出相应的动作；**e) 自我评价与反馈**。
 - **核心特点**：**a)** 在OpenAI的加持下，机器人已经能够理解任务内容，并非机械执行；同时能够与人类交互对话。**b)** 全程自主控制，无需任何远程操作；**c)** 机器人执行的速度和流畅度有了显著的提升，开始接近人类的速度。

图1：Figure 01给人类递上苹果



图2：Figure 01把垃圾放到筐子里



图3：Figure 01把盘子放到沥水架上



1.1、Figure：人形机器人初创企业，具备领先理解与学习能力



Figure：人形机器人初创企业，产品已经具备端到端学习能力和行走能力。

2) 端到端的学习能力：2024年1月，公司发布了Figure 01无需人为远程操控，实现自主冲咖啡的视频。据公司介绍，Figure 01只需看一段人类演示的视频，再加上10小时的端到端的训练，就能够学会视频里的操作技巧，自主完成任务；

3) 行走能力：2023年11月，公司展示了Figure 01的行走能力，能够从许多塑料箱子中拿起一个，搬运一定距离，并放置到传送带上。这说明此时Figure 01已拥有自主导航、识别箱子和任务优先级排序的能力。

多个高市值科技公司投资，公司在一级市场广受青睐。2024年2月29日，公司宣布，已从亚马逊创始人贝索斯、英伟达、OpenAI和微软等公司那里筹集了约6.75亿美元的资金，使其公司估值达到了26亿美元。

图4：Figure 01冲咖啡



图5：Figure 01搬箱子

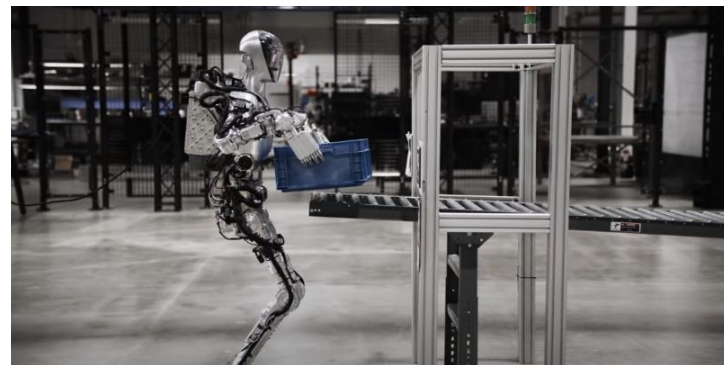
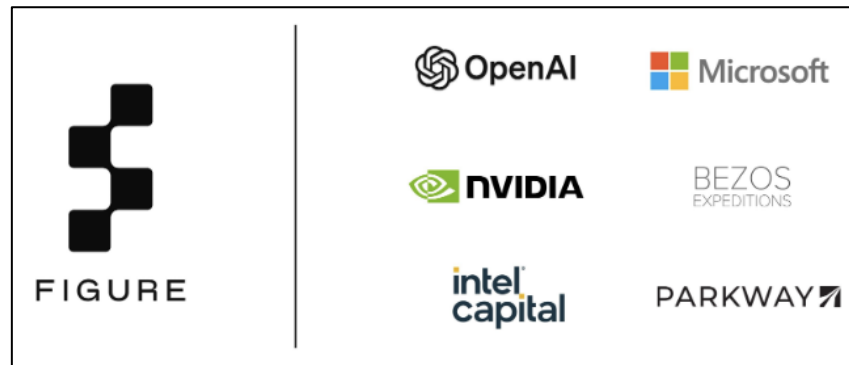


图6：Figure的投资方



资料来源：公司官网，财联社，华尔街见闻，量子位，腾讯新闻，光大证券研究所整理

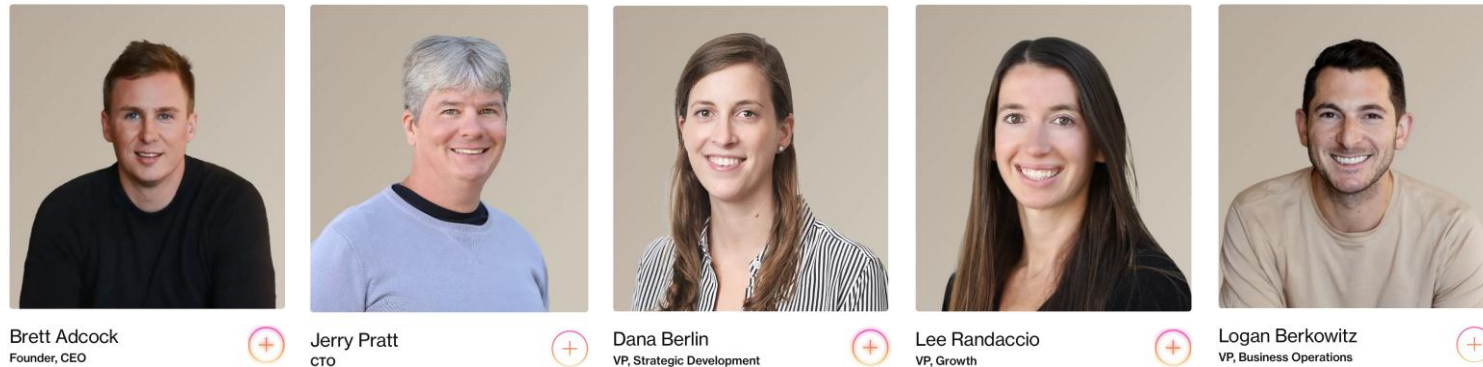
1.1、Figure：人形机器人初创企业，具备领先理解与学习能力



核心竞争力1：创始人和团队具备丰富的创业经验、技术能力和运营能力。

- 1) 公司创始人Brett Adcock：20年三次创业，曾经创办的企业包括基于AI的在线人才市场Vettery和eVTOL（电动垂直起降机）公司Archer，前者于2018年被全球最大的招聘公司The Adecco Group以1.1亿美元收购，后者则以27亿美元估值在纽交所上市，并与联合航空签署了15亿美元的商业协议。2022年，Brett开始专注于创办Figure；
- 2) 公司CTO Jerry Pratt：在加入Figure前是人类和机器认知研究所的首席研究员，专注于对步行和平衡的理解和建模，长期投身于机器人技术的研究；
- 3) 战略发展副总裁Dana Berlin：此前任职于巴克莱资本投行业务能源和可持续发展领域；
- 4) 其他细分模块：聚集了曾在波士顿动力、特斯拉、Lucid Motors等公司任职的专业人员。创始人和团队具备丰富的创业经验、技术能力和运营能力，将推动Figure持续稳健成长。

图7：Figure公司领导层



资料来源：公司官网，光大证券研究所整理

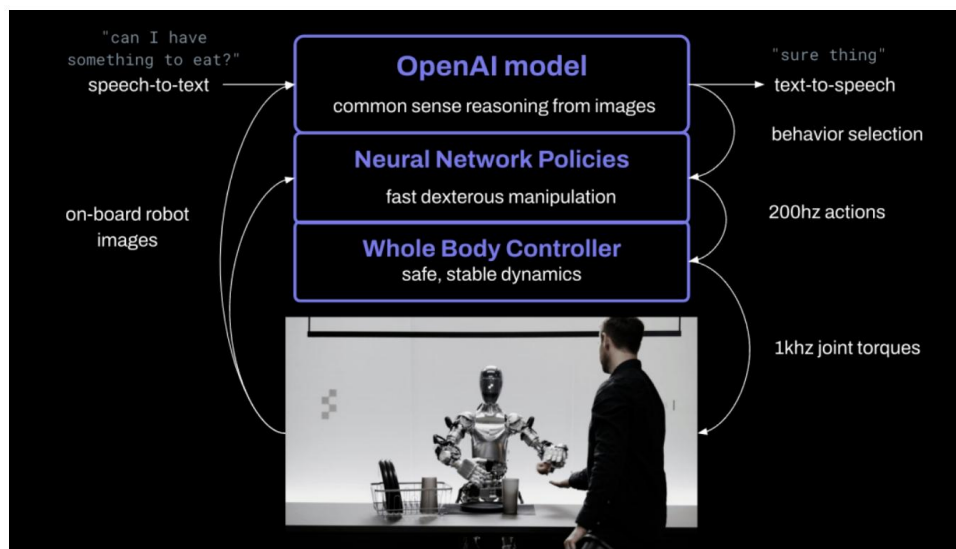
1.1、Figure：人形机器人初创企业，具备领先理解与学习能力



核心竞争力2：多模态大模型赋能，机器人加速迭代。

- **多模态模型同时理解输入的图像和文本信息：**Figure AI 将机器人摄像头拍摄的图像和通过板载麦克风记录的语音转录文本输入到一个由 OpenAI 训练的多模态模型中，这个模型能够同时理解图像和文本信息。
- **模型负责生成语言响应：**该模型会处理整个对话的历史记录，包括以往的图像，以生成语言响应，并通过文本到语音的方式向人类回话。
- **模型对于给定命令进行动作决策执行：**同一个模型还负责决定执行哪种已学习的闭环行为来响应给定的命令，它将特定的神经网络权重加载到 GPU 上，并执行相应的策略。

图8：Figure机器人操作高级AI工程师Corey Lynch介绍机器人技术原理



资料来源：华尔街见闻

1.1、Figure：人形机器人初创企业，具备领先理解与学习能力



核心竞争力3：自主完成任务——具备端到端的学习能力

表象：机器人训练省时省力。Figure 01只需看一段人类操作的视频，再用上10小时训练，就能实现无需人为远程操控，自主冲咖啡；在咖啡胶囊位置没摆正时还可以自动调整。

底层逻辑：端到端神经网络(End to End Neural Networks)

- **传统机器学习：**过程包括，（1）输入原始数据；（2）引入特定领域的知识，手动提取特征；（3）采用传统机器学习算法生成预测。尽管能够达成目标，但程序非常耗时，需要大量特定领域的知识。
- **端到端深度学习：**将数据直接通过一个大型神经网络，该神经网络处理输入的数据并自动提取相关特征，然后用于生成预测。

“端到端”的特点：（1）神经网络技术使得无需手动提取特征，省去传统机器学习第二个过程，直接从数据中进行输入-输出映射的学习，有效减少训练所需的时间和精力；（2）技术的前提是训练期间需要使用大量数据。

图9：传统机器学习流程图

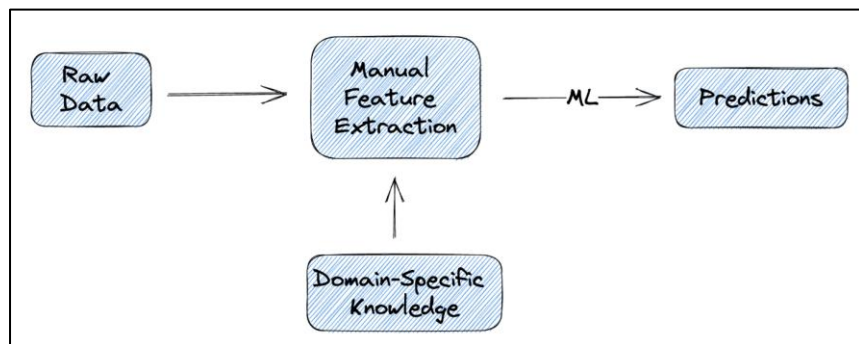
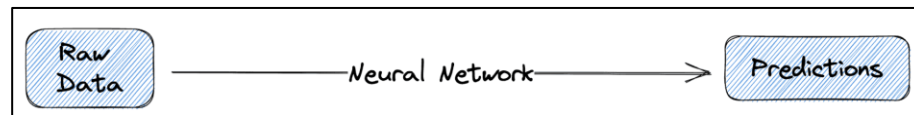


图10：端到端深度学习流程图



资料来源：公司官网，华尔街见闻，Baeldung，光大证券研究所整理

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/21711126144006052>