

大模型赋能人形机器人，软硬融合共创颠覆性产品

核心观点

- **国内蓝图开启，海内外利好有望共振。**11月2日，工信部印发《人形机器人创新发展指导意见》，明确人形机器人发展目标，2025年实现整机批量生产，2027年要形成安全可靠的产业链供应链体系。海外方面，特斯拉人形机器人进展迅速，在两年不到的时间内实现了快速迭代，完成了开发平台建造、Optimus产品亮相以及电机扭矩控制、环境探测与记忆、双手复杂任务等能力的实现。马斯克预计Optimus将大规模量产至“百万”量级，预计其单台成本或将低于2万美元。
- **软硬融合，打造人形机器人新高度。**软件决定人形机器人高度，算法需与硬件匹配。人形机器人本质是AI系统落地物理世界的最佳载体，算法是核心，需与硬件匹配。根据GGII预测，到2026年全球人形机器人在服务机器人中的渗透率有望达到3.5%，市场规模超20亿美元，到2030年全球市场规模有望突破200亿美元。
- **大模型赋能人形机器人，具身智能是未来目标。**人形机器人是软硬件能力高集成的实体，商业化的核心突破点在于“AI大脑”。大模型为人形机器人的发展带来了新的突破，使机器人具备更高的事物处理能力和自然语言交互能力。大模型的泛化能力则让研究者看到了人形通用机器人的曙光，泛化能力的出现让大模型能够在没有被训练过的场景中也能表现出色，是AI实现通用性的基础。综合使用多种数据模式的信息，可以显著提升模型的性能。因此，端到端的训练方式会在机器人模型领域有更多优势。今年内，科技巨头们纷纷加入了机器人模型的探索中来，试图找到一条适合于通用机器人的算法道路。
- **目前来看，距离实现真正的具身智能水平的模型还有很长的路要走。**机器人模型还面临着诸多问题：机器人算法的训练需要大量机器人真实数据，但实际可用于训练机器人学会执行新任务新技能的高质量数据非常匮乏；同时，机器人模型行动控制的周期仍太长，无法做到实时响应，需要大量算力支撑；人形机器人零部件众多，做好软硬件协同才能发挥其硬件水平。

投资建议与投资标的

国内顶层规划设计路线，明确量产时间点、关键技术、重点产品以及示范性应用场景，结合海外以特斯拉Optimus为代表的人形机器人持续推进，人形机器人产业化迎来曙光。我们看好人形机器人产业趋势，建议关注产业链上的核心零部件相关公司以及大模型公司：

- **总成：**三花智控(002050，买入)（汽车组覆盖）、拓普集团(601689，买入)（汽车组覆盖）；
- **空心杯电机：**鸣志电器(603728，未评级)、鼎智科技(873593，未评级)；
- **减速器：**绿的谐波(688017，未评级)、双环传动(002472，未评级)、中大力德(002896，未评级)；
- **编码器：**奥普光电(002338，未评级)、禾川科技(688320，未评级)、汇川技术(300124，未评级)、昊志机电(300503，未评级)、峰岬科技(688279，未评级)、儒竞科技(301525，未评级)；
- **丝杠和轴承：**五洲新春(603667，买入)、恒立液压(601100，未评级)、贝斯特(300580，未评级)、秦川机床(000837，未评级)、鼎智科技(873593，未评级)、禾川科技(688320，未评级)；
- **配套高端设备：**田中精机(300461，未评级)、华辰装备(300809，未评级)、浙海德曼(688577，未评级)、日发精机(002520，未评级)；
- **传感器：**柯力传感(603662，未评级)、昊志机电(300503，未评级)；
- **IMU：**芯动联科(688582，未评级)、华依科技(688071，未评级)、苏州固锟(002079，未评级)等。
- **大模型：**科大讯飞(002230，买入)、鼎捷软件(300378，未评级)、中科创达(300496，买入)等

风险提示

人形机器人进展不及预期；国产替代进程不及预期；行业竞争加剧

行业评级

看好（维持）

国家/地区

中国

行业

计算机行业

报告发布日期

2023年11月13日



证券分析师

王天一 021-63325888*6126
wangtianyi@orientsec.com.cn
执业证书编号：S0860510120021

浦俊懿 021-63325888*6106
pujunyi@orientsec.com.cn
执业证书编号：S0860514050004

杨震 021-63325888*6090
yangzhen@orientsec.com.cn
执业证书编号：S0860520060002
香港证监会牌照：BSW113

丁昊 dinghao@orientsec.com.cn
执业证书编号：S0860522080002

联系人

覃俊宁 qinjunning@orientsec.com.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

目 录

一、人形机器人：曙光已现，软硬融合共造颠覆性产品	5
1.1 国内蓝图开启，海内外利好有望共振	5
1.2 软硬融合，打造人形机器人新高度	9
二、大模型+机器人：未来智能革命将至	13
2.1 大模型赋能人形机器人，具身智能是未来目标	13
2.2 科技巨头纷纷入局，通用机器人算法路线仍需探索	16
三、机器人大模型面临的问题	26
3.1 高质量训练数据缺乏，合成数据有望成为新方向	26
3.2 实时性仍需提升，算力可能是瓶颈	27
3.3 软硬件融合是最终目标	28
投资建议与投资标的	30
风险提示	30

图表目录

图 1：工信部印发《人形机器人创新发展指导意见》明确发展目标和重点任务	5
图 2：特斯拉人形机器人发展历程图.....	8
图 3：人形机器人软硬件架构示意图.....	9
图 4：人形机器人行业产业链	9
图 5：2026-2030 年全球及中国人形机器人行业市场规模预测(单位：亿美元)	10
图 6：2023 年人形机器人核心零部件价值量分布图预测	10
图 7：2030 年人形机器人核心零部件价值量分布图预测	10
图 8：特斯拉人形机器人 FSD 视觉感知方案分析图	11
图 9：AI 大模型助力人形机器人拆解任务示意图	12
图 10：具身智能的研究是多学科交叉的	13
图 11：波士顿动力机器狗接入 ChatGPT 能力后能与人类交互	14
图 12：波士顿动力机器狗可以扮演不同性格的导游	14
图 13：泛化的种类	14
图 14：Tesla Optimus 是完全端到端训练的	15
图 15：GPT-4V 可以与环境进行正确交互	16
图 16：具身智能机器人算法一览图	16
图 17：传统机器人任务需要工程师不断调整目标	17
图 18：引入 ChatGPT 后无需工程师在循环中进行调整	17
图 19：ChatGPT for Robotics 的设计原则	17
图 20：ChatGPT 可操作机器臂拼出微软 Logo.....	18
图 21：需要将外部环境转化为文字信息输入给 ChatGPT.....	18
图 22：PaLM-E 能够接收多模态信息并输出决策文本	18
图 23：PaLM-E 具备迁移学习能力.....	19
图 24：PaLM-E 可以进行 zero-shot 多模态思维链推理.....	19
图 25：RT-2 架构和操作字符串的表示方式	20
图 26：RT-2 的任务成功率高于基线模型.....	20
图 27：RT-2 具备比基线模型更强的泛化能力.....	20
图 28：加入思维链能力使得 RT-2 能够处理更复杂的问题.....	21
图 29：Voxposer 流程一览图	21
图 30：Voxposer 的涌现能力	22
图 31：RoboAgent 采用的数据集远远小于 RT-1.....	22
图 32：RobeAgent 采用了数据增广技术对数据集进行扩充.....	23
图 33：RoboAgent 有更高的效率和更强的泛化能力	23
图 34：Open X-Embodiment 数据集	24

图 35: RT-1-X 性能表现较原有方法大幅提升	24
图 36: RT-2-X 的涌现能力是 RT-2 的三倍	24
图 37: Eureka 能够训练机器人执行复杂动作	25
图 38: Eureka 工作流程	25
图 39: 以人类设计算法为初始条件的 Eureka 表现进一步提升	25
图 40: Open X-Embodiment 数据集一览	26
图 41: MimicGen 合成机器人任务数据	27
图 42: MimicGen 可生成多种场景的机器人任务数据	27
图 43: Voxposer 需要 8 倍速播放才有流畅的动作	28
图 44: RT-2 需要 2 倍速播放才有流畅的动作	28
图 45: 特斯拉 Dojo 超算算力规模预测	28
图 46: 特斯拉将基于 Dojo 训练机器人模型	28
图 47: Tesla Optimus 机器人硬件结构	29
表 1: 工信部印发《人形机器人创新发展指导意见》指出要攻克的关键技术	6
表 2: 工信部印发《人形机器人创新发展指导意见》指出要攻关的重点产品和部组件	6
表 3: 2023 年部分全球人形机器人代表性企业产品分析	7
表 4: 中国人形机器人行业核心零部件国产替代空间对比和部分代表企业	11
表 5: 人形机器人执行层面两大关键环节分析	12
表 6: 人形机器人算法训练两大关键环节分析	12

一、人形机器人：曙光已现，软硬融合共造颠覆性产品

1.1 国内蓝图开启，海内外利好有望共振

国内蓝图开启，人形机器人有望成为颠覆性产品。2023 年 11 月 2 日，工信部印发《人形机器人创新发展指导意见》，明确指出：人形机器人集成人工智能、高端制造、新材料等先进技术，有望成为继计算机、智能手机、新能源汽车后的颠覆性产品，将深刻变革人类生产生活方式，重塑全球产业发展格局。

明确发展目标，2025 年实现整机批量生产，2027 年要形成安全可靠的产业链供应链体系。《指导意见》明确指出发展目标：到 2025 年，人形机器人创新体系初步建立，“大脑、小脑、肢体”等一批关键技术取得突破，确保核心部件安全有效供给。整机产品达到国际先进水平，并实现批量生产，在特种、制造、民生服务等场景得到示范应用，探索形成有效的治理机制和手段。培育 2-3 家有全球影响力的生态型企业和一批专精特新中小企业，打造 2-3 个产业发展集聚区，孕育开拓一批新业务、新模式、新业态。2) 到 2027 年，人形机器人技术创新能力显著提升，形成安全可靠的产业链供应链体系，构建具有国际竞争力的产业生态，综合实力达到世界先进水平。产业加速实现规模化发展，应用场景更加丰富，相关产品深度融入实体经济，成为重要的经济增长新引擎。

图 1：工信部印发《人形机器人创新发展指导意见》明确发展目标和重点任务



数据来源：工信部，东方证券研究所

关键技术攻克：《指导意见》提出以大模型等人工智能技术突破为引领，在机器人已有成熟技术基础上，重点在人形机器人“大脑”和“小脑”、“肢体”关键技术、技术创新体系等领域取得突破。

- 一是开发基于人工智能大模型的人形机器人“大脑”，增强环境感知、行为控制、人机交互能力，开发控制人形机器人运动的“小脑”，搭建运动控制算法库，建立网络控制系统架构。
- 二是系统部署“机器肢”关键技术群，打造仿人机械臂、灵巧手和腿足，攻关“机器体”关

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

键技术群，突破轻量化骨骼、高强度本体结构、高精度传感等技术。

- 三是构建完善人形机器人制造业技术创新体系，支持龙头企业牵头联合产学研用组成创新联合体，加快人形机器人与元宇宙、脑机接口等前沿技术融合，探索跨学科、跨领域的创新模式。

表 1：工信部印发《人形机器人创新发展指导意见》指出要攻克的关键技术

关键技术攻克	
机器人“大脑”关键技术群	围绕动态开放环境下人形机器人感知与控制，突破感知-决策-控制一体化的端到端通用大模型、大规模数据集管理、云边端一体计算架构、多模态感知与环境建模等技术，提高人形机器人的人-机-环境共融交互能力，支撑全场景落地应用。
机器人“小脑”关键技术群	面向人形机器人复杂地形通过、全身协同精细作业等任务需求，开展高保真系统建模与仿真、多体动力学建模与在线行为控制、典型仿生运动行为表征、全身协同运动自主学习等关键技术研究，提升人形机器人非结构化环境下全身协调鲁棒移动、灵巧操作及人机交互能力。
机器肢关键技术群	面向人形机器人高动态、高爆发和高精度等运动性能需求，研究人体力学特征及运动机理、人形机器人动力学模型及控制等基础理论，突破刚柔耦合仿生传动机构、高紧凑机器人四肢结构与灵巧手设计等关键技术，为人形机器人灵活运动夯实硬件基础。
机器人本体关键技术群	面向人形机器人本体高强度和高紧凑结构需求，研究人工智能驱动的骨架结构拓扑优化、高强度轻量化新材料、复杂身体结构增材制造、能源-结构-感知一体化设计以及恶劣环境防护等关键技术，打造具有高安全、高可靠、高环境适应性的人形机器人本体结构。

数据来源：工信部，东方证券研究所

重点产品和部组件攻关：《指导意见》将打造整机产品、夯实基础部组件、推动软件创新作为主要发力方向。

- 在整机产品方面，打造基础版整机，构筑人形机器人通用整机平台，开发低成本交互型、高精度型以及极端环境下高可靠型等人形机器人整机产品，强化人形机器人整机的批量化生产制造能力；
- 在基础部组件方面，开发人形机器人专用传感器、高功率密度执行器、专用芯片，以及高能效专用动力组件；
- 在软件创新方面，构建人形机器人高实时、高可靠、高智能的专用操作系统，开发面向各类场景的应用软件，建设完善人形机器人应用开发平台和工具包。

表 2：工信部印发《人形机器人创新发展指导意见》指出要攻关的重点产品和部组件

重点产品和部组件攻关	
基础版整机	面向类人外观、双腿行走和双臂双手灵巧操作的基本形态功能，建立人形机器人基础软硬件架构，打造“公版”通用平台，支持不同场景需求下的结构改造、算法优化以及特定能力强化。
功能型整机	开发低成本交互型人形机器人，强化人类生活环境适应能力、多模态人机交互能力。开发高精度型人形机器人，强化双臂双手精细操作、工件鲁棒识别、轨迹智能规划等上肢作业能力。开发高可靠型人形机器人，强化恶劣环境生存、复杂地形适应、外力冲击防护等能力。
传感器	面向复杂环境感知需求，开发集成高精度仿生眼与类脑处理算法的视觉传感器，推出宽频响、高灵敏的仿生听觉传感器，开发高分辨率和具有多点接触检测能力的仿人电子皮肤，推出高灵敏检测多种气体的仿生嗅觉传感器，形成人形机器人专用传感器产品谱系。
执行器	面向人形机器人高爆发移动需求，突破高功率密度液压伺服执行器，打造高紧凑液压马达、缸、泵、阀及一体化单元系列产品。突破高力矩密度减速器、高功率密度电机、伺服驱动器等融合的高精度电驱动执行器，打造电驱动旋转关节、电推杆产品。
控制器	面向高实时协调运动控制需求，研发具有高动态运动驱动、高速通信等功能的专用芯片，研制“感-算-控”一体化的高性能运动控制器。面向人形机器人认知与决策需求，研发具有多模态空间感知、行为规划建模与自主学习等能力的智能芯片，提升人形机器人协调控制能力。

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并阅读本证券研究报告最后一页的免责申明。

动力能源	高动态、长续航能量需求，突破高能量密度电池、智能电源管理、电池组优化匹配等关键技术，开发高效、高紧凑动力能源总成产品，提升人形机器人的续航与环境适应能力。
------	---

数据来源：工信部，东方证券研究所

拓展场景应用：《指导意见》从特种领域、制造业典型场景、民生及重点行业三类方向提出意见措施。

- 一是加快人形机器人在特种环境应用，面向恶劣条件、危险场景作业等需求，强化复杂环境下本体控制、快速移动、精确感知等能力。
- 二是聚焦 3C、汽车等制造业重点领域，提升人形机器人工具操作与任务执行能力，打造人形机器人示范产线和工厂，在典型制造场景实现深度应用。
- 三是拓展人形机器人在医疗、家政等民生领域服务应用，满足生命健康、陪伴护理等高品质生活需求，推动人形机器人在农业、物流等重点行业应用落地，提升人机交互、灵巧抓取、分拣搬运、智能配送等作业能力。

回顾历史，早在 1986-1993 年，日本本田公司接连开发了 E0 到 E6 等 7 种行走机器人，只有腿部结构，主要用于研究行走功能。1993 年，公司在研制的 P1 基础上加上了双臂使其初具人形，而后同年完成的 P3 则是 ASIMO 的原型。2000 年，本田公司推出 ASIMO；2013 年，美国波士顿动力公司推出 Atlas，能够在碎石堆上行走。

当前人形机器人的主要布局者多为科技公司，也有多家跨界公司入局。美国 1X technologies 公司与 ADT commercial 合作成功开发的 EVE 机器人可用于安保、护理、调酒等场景，目前在售，每月可生产 10 台，在美国和欧洲部分地区开始投入使用；公司的 NEO 机器人主要应用于安保、物流、制造、机械操作以及处理复杂的作业，2023 年底开始接受预订。英国 Engineered Arts 研发的 Ameca 机器人可以在与人类互动时检测情绪和年龄，并可以用常见的表情进行交流，主要应用于实验研究、展览、接待、教育等领域。国内有小米 Cyberone 重建真实世界、实现运动姿态平衡、感知人类情绪；优必选 Walker X 主要应用于科技展馆、影视综艺、商演活动、政企展厅；智元机器人远征 A1 预计先应用在工业制造领域，随后逐步走向 to C 应用。

表 3：2023 年部分全球人形机器人代表性企业产品分析

公司	人形机器人	身高（cm）	体重（kg）	速度（km/h）	自由度（关节数）	功能及应用场景
1X technologies(美国)	EVE	186	83	14.4	/	安保、护理、调酒等
	NEO	/	30	步行：4 跑步：12	/	主要应用于安保、物流制造、机械操作以及处理复杂的作业
特斯拉(美国)	Optimus	173	73	8	50	浇水植物、移动金属棒环境感知和记忆力较好力度控制较准确等
Boston dynamics(美国)	Atlas	150	89	9	28	可以行走、奔跑并穿越多种复杂地形、手部灵活能完成抓取等动作

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

Agility Robotics(美国)	Digit	175	不足 65	/	16	移动手提袋或包裹、卸货、行走交付
Engineered Arts(英国)	Ameca	187	49	/	61	可以在与人类互动时检测情绪和年龄，并可以用常见的表情进行交流，应用于实验研究、展览、接待、教育
本田（日本）	ASIMO	130	50	2.7	57	可同时与多人进行对话，不仅可以步行、奔跑、倒退走、跳跃，也可以在些微不平的地面行走
小米（中国）	Cyberone	177	52	3.6	21	重注真实世界，实现运动动态平衡、感知人类情绪
优必选(中国)	Walker X	130	63	3	41	复杂地形自适应、动态足腿控制、手眼协调操作，U-SLAM 视觉导航；主要应用于科技展馆、影视综艺、商演活动、政企展厅
智元机器人(中国)	远征 A1	175	55	7	49+	先应用在工业制造领域随后逐步走向 To C 应用

数据来源：公司官网，前瞻产业研究院，东方证券研究所

特斯拉人形机器人进展迅速，在两年不到的时间内实现了快速迭代，完成了开发平台建造、Optimus 产品亮相以及电机扭矩控制、环境探测与记忆、双手复杂任务等能力的实现。2021 年首届特斯拉 AI DAY 上其首席执行官马斯克首次公布特斯拉机器人 Tesla Optimus 概念图。2022 年特斯拉 AI DAY 上特斯拉公布了 Optimus 原型机的首秀，实现了自主行走，转身，挥手等动作。2023 年 5 月的股东大会提到特斯拉自研超算 Dojo 为 Optimus 提供算力支持，可加快训练速度并降低训练成本，加快人形机器人的产业化落地。截至 2023 年 8 月，至少有五个特斯拉 Optimus 机器人原型已经建成并能够行走，使用的电机、控制器和电子设备均由特斯拉设计和制造。马斯克预计 Optimus 将大规模量产至“百万”量级，预计其单台成本或将低于 2 万美元。

图 2：特斯拉人形机器人发展历程图

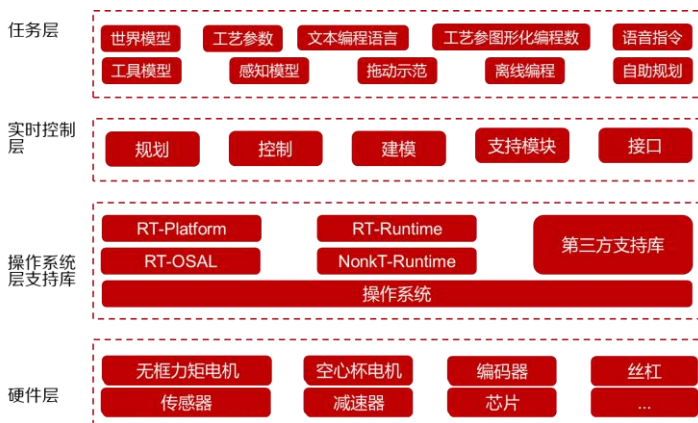


数据来源：特斯拉公司官网，前瞻产业研究院，东方证券研究所

1.2 软硬融合，打造人形机器人新高度

软件决定人形机器人高度，算法需与硬件匹配。人形机器人本质是 AI 系统落地物理世界的最佳载体，算法是核心，需与硬件匹配。机器人的输出包含了虚拟与物理两种能力。虽然人形机器人从本体硬件上看，存在抗压硬度与灵敏度不足的问题，但更核心问题在于是算法对运动能力的控制，包括本体平衡、行走的步态、手部抓取等规划与控制。这需要成熟的感知系统基础、强大的算法分解任务和规划动作、大模型不断仿真训练以及超强的算力支撑，同时要求算法与硬件相匹配。这要求机器人企业需自研算法，并持续更新迭代。

图 3：人形机器人软硬件架构示意图



数据来源：Tesla AI Day，前瞻产业研究院，东方证券研究所

硬件方面看，人形机器人产业链主要包括上游的核心零部件，例如无框力矩电机、空心杯电机、传感器、专用芯片等；中游为机器人本体制造，包括设计、制造、测试三大环节；下游为人形机器人应用领域，包括工业制造、仓储物流、医疗服务、商业服务、家庭使用等。

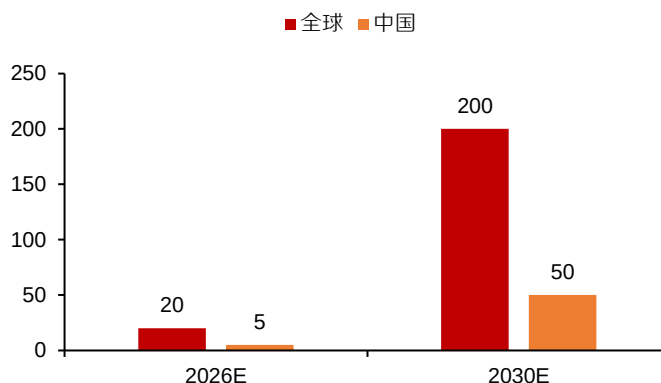
图 4：人形机器人行业产业链



数据来源：Tesla AI Day，前瞻产业研究院，东方证券研究所

从全球范围来看，人形机器人已有商业应用场景预期落地，例如巡逻、物流仓储等领域。商业化进程领先的人形机器人产品则有 EVE 和 Digit。具体而言，美国 1X technologies 公司与 ADT commercial 公司共同研发的人形机器人 EVE 目前已成功应用于巡逻安保场景；Digit 的应用场景主要是在物流仓储环节，进行的任务主要包括卸载货车、搬运箱子、管理货架等，预计将在 2025 年全面上市。在政策、资本以及技术多维度赋能下，人形机器人市场的潜力有望被加速释放。未来的商业应用场景有望渗透进入服务业、制造业等领域。马斯克曾表示其设计特斯拉机器人的远景目标是让其服务于千家万户比如做饭、修剪草坪、照顾老人等。目前较有潜力的人形机器人发展方向主要面向制造业、航天探索、生活服务业、高校科研等，预计 2025 年人形机器人将有望实现制造业场景应用的突破，小批量应用于电子、汽车等生产制造环境。根据 2023 年 5 月 GGII 发布的报告预测，预计到 2026 年全球人形机器人在服务机器人中的渗透率有望达到 3.5%，市场规模超 20 亿美元，到 2030 年全球市场规模有望突破 200 亿美元。参考中国服务机器人市场约占全球市场 25% 的数值测算，2030 年中国人形机器人市场规模将达 50 亿美元。

图 5：2026-2030 年全球及中国人形机器人行业市场规模预测(单位：亿美元)



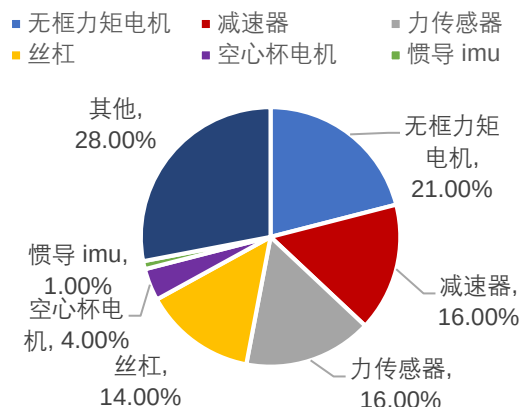
数据来源：GGII，前瞻产业研究院，东方证券研究所

根据 Tesla AI Day 的预测数据，以特斯拉 Optimus 为例，2023 年人形机器人核心零部件价值量排名前三的是无框力矩电机、减速器和力传感器；2030 年无框力矩电机价值量占比下降，力传感器、减速器价值量占比上升，且力传感器将超过减速器，排名第二，三者合计占比仍超过 50%。

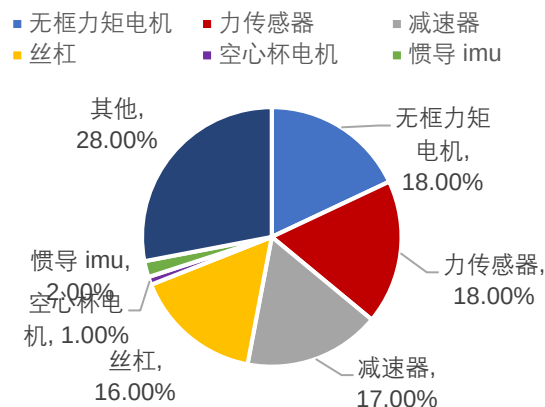
图 6：2023 年人形机器人核心零部件价值量分布图预测

图 7：2030 年人形机器人核心零部件价值量分布图预测

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。



数据来源: Tesla AI Day, 前瞻产业研究院, 东方证券研究所



数据来源: 前瞻经济学人 APT, 东方证券研究所

从单机价值量占比来看, 无框力矩电机、减速器和力传感器价值量占比较高; 从降本空间来看, 空心杯电机、无框力矩电机等降本空间较大; 而从国产替代空间来看, 行星滚柱丝杠、空心杯电机、惯导 imu 等国产化率较低, 国产替代空间大。

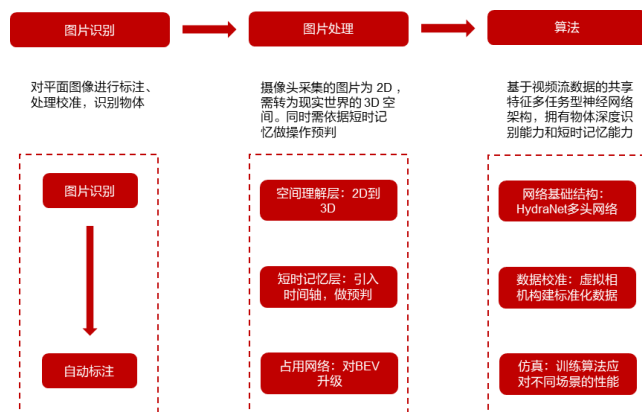
表 4: 中国人形机器人行业核心零部件国产替代空间对比和部分代表企业

核心零部件	2023 单机价值量占比	国产化率	部分代表企业
无框力矩电机	21%	中等	步科股份、禾川科技、昊志机电等
减速器	16%	较高	绿的谐波、昊志机电、国茂股份、秦川机床、丰立智能、双环传动、中大力德、科风智能等
力传感器	16%	中等	柯力传感、昊志机电等
丝杠	14%	低	五洲新春、新剑传动、贝斯特、恒立液压、秦川机床、鼎智科技、禾川科技、长盛轴承、南京工艺等
空心杯电机	4%	低	鸣志电器、鼎智科技（江苏雷利）、拓邦股份等
惯导 imu	1%	低	芯动联科、华依科技、苏州固得等

数据来源: Tesla AI Day, 各公司官网, 前瞻产业研究院, 东方证券研究所

软件方面看, 特斯拉 FSD 纯视觉深度学习神经网络渐趋成熟, 可嫁接至人形机器人。与自动驾驶类似, 人形机器人同样需利用传感器(摄像头、激光雷达等)感知、采集和处理周围环境信息, 以便做出决策。特斯拉人形机器人仅使用 3 个摄像头(中间鱼眼摄像头、左右各一个视觉摄像头), 直接嫁接 FSD 成熟的纯视觉方案, 再辅之以力、力矩、声学、触觉、温度等传感器。

图 8: 特斯拉人形机器人 FSD 视觉感知方案分析图

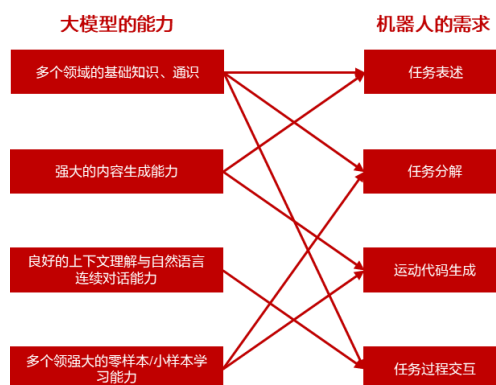


有关分析师的申明, 见本报告最后部分。其他重要信息披露见分析师申明之后部分, 或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

数据来源：Tesla AI DAY，前瞻产业研究院，东方证券研究所

人形机器人决策难度更高：决策层依据感知层获取的信息进行决策判断，来控制机器人身体做出动作规划并下发指令。特斯拉人形机器人与 FSD 底层模块打通，一定程度上算法可复用，但人形机器人需完成人类各种动作，动作连续复杂、需频繁的物理交互且操作因果性多，算法难度远高于自动驾驶。随着人工智能大模型的快速发展，ChatGPT 能够助力人形机器人拆解任务。大语言模型擅长推断语言条件，并利用其代码编写能力，拆分任务，给出运动规划的目标函数。

图 9：AI 大模型助力人形机器人拆解任务示意图



数据来源：Tesla AI DAY，前瞻产业研究院，东方证券研究所

执行：人形整机平衡与行走步态是关键。人形机器人执行层指的是系统在做出决策后，对机器人本体做出控制。机器人各操作系统都与决策系统相链接，并按指令精确执行。其中，人形整机平衡与行走步态是最基本也是最关键的两个环节。

表 5：人形机器人执行层面两大关键环节分析

关键环节	具体分析
自平衡系统	机器人在不同环境下保持动态平衡(特别是外力冲击下)，需要软件算法和机械设计共同作用。软件层面看，一方面通过传感器获取机器人的状态信息，从而控制关节运动实现平衡；另一方面，通过预测机器人的运动轨迹和所需动作，而提前应对。
行走步态	零力矩点(必须落在支撑面内，合理地规划踝关节和髌关节，以保持动态行走时重心的稳定；同时腿部应具备适当的机械柔顺性，有效缓解来自未知高刚度环境的碰撞冲击)。这均要求算法与关节硬件相匹配。

数据来源：前瞻产业研究院，东方证券研究所

仿真：算法训练可提升机器人智能化水平。仿真的目的在于评估机器人结构和算法的设计，包括机器人的运动、工作环境、感知等，意义在于通过仿真模型快速、低成本、高安全性地训练机器人的算法。通过仿真，可加快软件更新迭代，同时缩短算法与硬件调整时间，极大提高训练效率。另外，随着芯片与 AI 技术的发展，未来端到端训练有望突破，人形机器人具身智能未来可期。

表 6：人形机器人算法训练两大关键环节分析

关键环节	具体分析
AI 大模型仿真学习	仿真的目的在于评估机器人结构和算法的设计，包括机器人的运动、工作环境、感知等，意义在于通过仿真模型快速、低成本、高安全性地训练机器人的算法。通过仿真，可加快软件更新迭代，同时缩短算法与硬件调整时间，极大提高训练效率
端到端训练	特斯拉 ESD 将推出 12 版本，将小模型集成至大模型中，也有望应用于机器人，提高训练效率。具身智能是人工智能的终极形态，机器人可理解、推理并与物理世界互动，集视觉、语言、认知和推理、机器学习等大成。

数据来源：前瞻产业研究院，东方证券研究所

二、大模型+机器人：未来智能革命将至

2.1 大模型赋能人形机器人，具身智能是未来目标

人形机器人是软硬件能力高集成的实体，商业化的核心突破点在于“AI 大脑”。可以说，当前的 AI 大脑在逻辑思维和行为智慧决策层面还需要一段成长空间，其驱动力很大程度上来自于算法的升级与高水平的智能化。

具身智能是人形机器人想要实现的最终方向。具身智能是指一种智能系统或机器能够通过感知和交互与环境（物理世界）产生实时互动的能力。具身智能包括三个模块：具身感知（Perception）、具身想象（Imagination）和具身执行（Execution），AI+机器人正是“具身智能”当前的落点。在机器人领域，有一个“莫拉维克悖论”——人类所独有的高阶智慧能力只需要非常少的计算能力，但无意识的技能和直觉（如辨识人脸、举起铅笔、在房间内走动、转笔等能力）却需要极大的运算能力，即越简单的事情越难，越难的事情越简单。具身智能机器人未来要能够像人一样与环境交互、感知、决策、完成任务，其感知层和认知层的能力是非常高的。显然，当下的人形机器人距离这一状态还有非常巨大的差距。人形机器人传感器数量、品类、执行机构复杂程度远高于工业机器人，对控制器实时算力、集成度要求极高。

图 10：具身智能的研究是多学科交叉的



数据来源：上海交通大学卢策吾，东方证券研究所

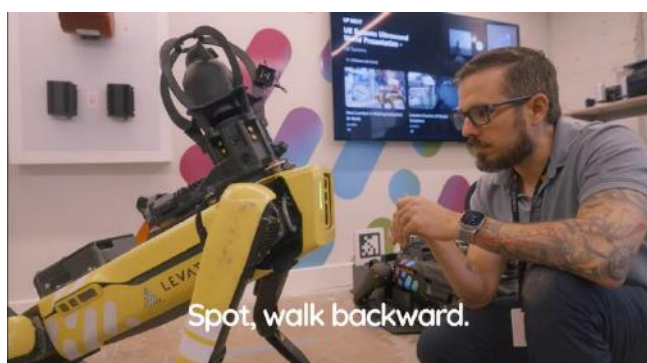
大模型为人形机器人的发展带来了新的突破。以往由于算法模型的局限，机器人难以实现通用化能力，只能通过对于某一个特定领域训练对应的基础模型，达到能够满足特定用途的机器人 AI 算法，比如工厂车间大量采用的机器人臂、家用扫地机器人等。软件层面的局限性限制了机器人的应用场景狭窄、可通用性较差，无法充分发挥硬件层面的性能。而大模型的出现，补全了让机器

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并阅读本证券研究报告最后一页的免责声明。

人能力实现跃升的技术基础。大模型的向量应用提供了大脑给机器人，使机器人具备更高的事物处理能力。大模型的 Chat 应用把自然语言理解提供给了机器人，机器人首次可以用接近人的语言水平，和真实人类聊天，进行基于自然语言的互动。此外，大模型的编程辅助可以肉眼可见的减少软件工程师的工作量，产出软件的成本会逐年下降。

ChatGPT 已经在机器狗领域落地应用，人形机器人未来可期。今年 4 月，AI 公司 Levatas 与波士顿动力合作，将 ChatGPT 和谷歌的语音合成技术接入 Spot 机器狗，成功实现与人类的交互。可以说，ChatGPT 的成功，也为人形机器人这一更加高级的领域带来了发展拐点。以 ChatGPT 为代表的语言大模型能够实现近似人一样的自然语言交互，多模态大模型则让人形机器人能够通过“视觉”去与环境交互。可以说，大模型赋予了人形机器人一个通识大脑，从而能够顺畅地和外界对话，还可以增加任务理解、拆分和逻辑推理等“决策”能力。

图 11：波士顿动力机器狗接入 ChatGPT 能力后能与人类交互



数据来源：cybernews，东方证券研究所

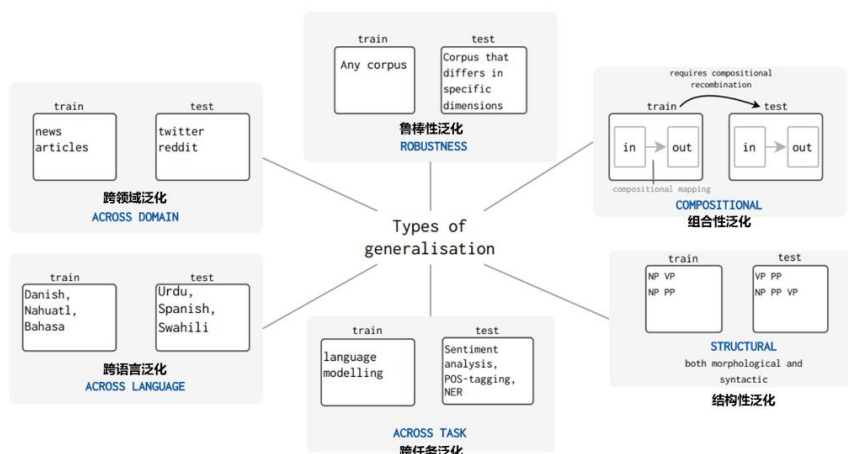
图 12：波士顿动力机器狗可以扮演不同性格的导游



数据来源：cybernews，东方证券研究所

大模型的泛化能力让研究者看到人形通用机器人的曙光。以往的 AI 模型训练完成后，就只能用于其被设计出来的场景中，难以进行拓展，无法实现通用性。而大模型具备强大的泛化能力，让通用这一目标的实现成为了可能。泛化（Generalization）可以理解作为一种迁移学习的能力，把从过去的经验中学习到的表示、知识和策略应用到新的领域。人类就具有“举一反三”的能力，学习一个新概念后立即就能理解它在其他情况下的相关用法。以往的 AI 泛化能力很低，应用场景比较局限，泛化能力的出现让大模型能够在没有被训练过的场景中也能表现出色，是 AI 实现通用性的基础。人形机器人所面临的应用场景与人类的日常生活接近，需要面对多种多样、不重复、没见过的任务，模型的泛化能力就成为了其能否真正实现通用的核心要素。

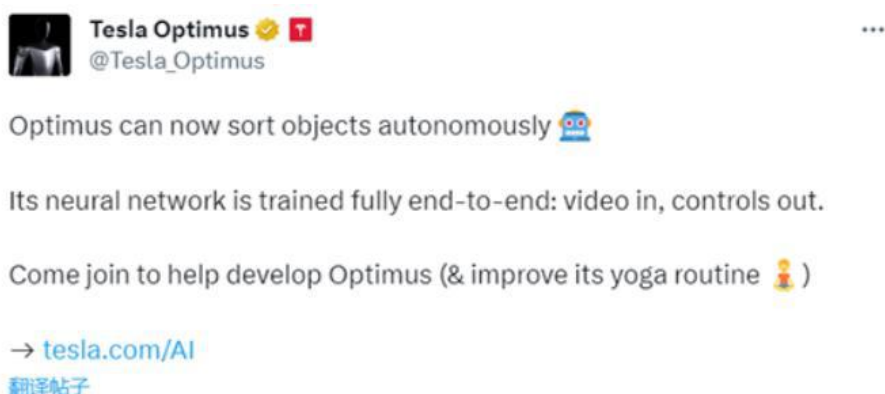
图 13：泛化的种类



数据来源：Meta《State-of-the-art generalization research in NLP : A taxonomy and review》，东方证券研究所

多模态数据包含更多信息，端到端的算法训练框架提升模型泛化能力。图像、视频等多模态的数据拥有着比文字要多许多的信息，采用多模态数据的大模型有助于实现更高细粒度的语义理解、对话意图识别以及更精确的情感分析。综合使用多种数据模式的信息，可以显著提升模型的性能。因此，端到端的训练方式会在机器人模型领域有更多优势。传统的 AI 模型训练需要经历对原始数据的特征工程或者数据处理阶段，而端到端训练是指直接以原始数据作为输入，输出最终结果的方法。端到端的训练框架通过缩减人工预处理和后续处理，尽可能使模型从原始的多模态数据输入到最终输出，给模型更多可以根据数据自动调节的空间，增加模型的整体契合度，提升其泛化能力。特斯拉的 FSD 算法就是纯端到端的自动驾驶算法，并将这样的训练思路也延续到了其人形机器人 Optimus 上，特斯拉表示 Optimus 的神经网络训练是“完全端到端的”，即可实现视频信号输入，控制信号输出。

图 14：Tesla Optimus 是完全端到端训练的



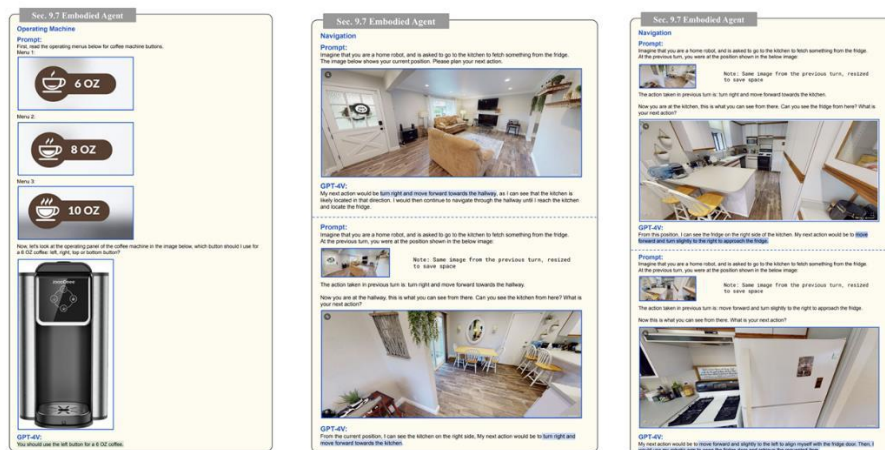
数据来源：Twitter，东方证券研究所

以 GPT-4 为首的多模态大模型已经具备成为通用性人形机器人核心大脑的初步条件。OpenAI 在 10 月正式上线 GPT-4V (ision) 这一新版本，为 GPT-4 新增了语音与图像功能。现在用户可以直接与 GPT-4V 进行语音交互，并且 GPT-4V 能够对图像进行推理和分析。根据微软团队对 GPT-4V 的详细评测，GPT-4V 有作为具身机器人的理解核心的潜力。在微软的测试案例中，GPT-4V 可以扮演一名家庭机器人，阅读咖啡机的操作界面并给出正确的指令操作；或者通过房间图片的

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

输入，要求 GPT-4V 规划出去厨房冰箱取物品的路线，GPT-4V 也可以执行面向任务的导航。具有多模态输入的 GPT-4V 在面对动态环境时可以很好地与环境交互，证明了 GPT-4 拥有成为人形机器人核心大脑的潜力。如果将 GPT-4 与合适的硬件进行结合，就有望实现具备 GPT-4 水平的具身智能。

图 15：GPT-4V 可以与环境进行正确交互



数据来源：微软《The Dawn of LLMs》，东方证券研究所

2.2 科技巨头纷纷入局，通用机器人算法路线仍需探索

基于大模型的具身智能体成为学术界重点研究方向。具身智能被斯坦福大学计算机科学家李飞飞定义为 AI 领域的下一个“北极星问题”之一，在大模型流行起来之后，关于以大模型为基础构造的具身智能体的研究也逐渐变多。今年内，谷歌、微软、英伟达、Meta 等科技巨头都纷纷加入了这条技术路线的探索中来，试图找到一条适合于通用机器人的算法道路。

图 16：具身智能机器人算法一览图



数据来源：澎湃，东方证券研究所

1) ChatGPT for Robotics：大模型+机器人结合初探

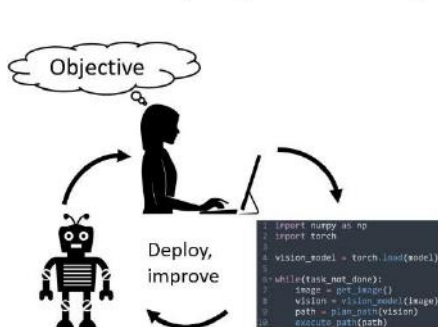
2023 年 2 月，微软团队发布了一篇名为《ChatGPT for Robotics: Design Principles and Model Abilities》的技术报告，展现出了使用 ChatGPT 来实现自然的人机交互的可能性，为如何将 ChatGPT 融入机器人领域提供了一种全新的视角，也是对于大模型和机器人的结合进行了初探。

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

ChatGPT 在机器人领域的应用主要体现在两个方面：自然语言交互和自动化决策。大多数的机器人任务需要由工程师来构建完整的高低级指令代码、不断进行迭代循环才能完成任务。而 ChatGPT 则带来了新的机器人任务范式，ChatGPT 具备自然语言交互和自动化决策能力，机器人可以通过 ChatGPT 来理解人类的自然语言指令，并根据指令进行相应的动作。另外，机器人在执行任务时需要做出各种决策，例如路径规划、物体识别等。ChatGPT 可以应用于机器人的自动化决策中，机器人可以通过 ChatGPT 来生成任务方案，根据任务要求做出相应的决策。引入 ChatGPT 后，工程师仅需做好机器人的底层库函数及其描述，ChatGPT 就能够自己利用这些底层函数来编写代码以完成目标，将工程师从循环中解放出来。

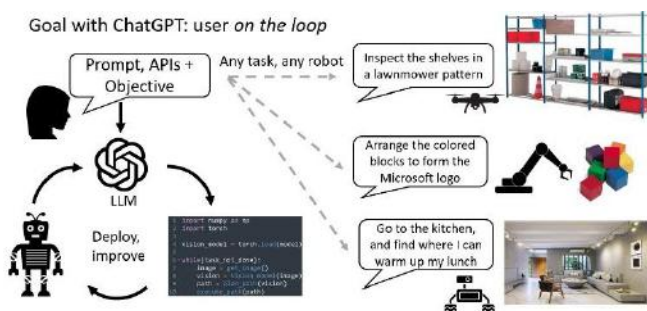
图 17：传统机器人任务需要工程师不断调整目标

Robotics today: engineer in the loop



数据来源：微软，东方证券研究所

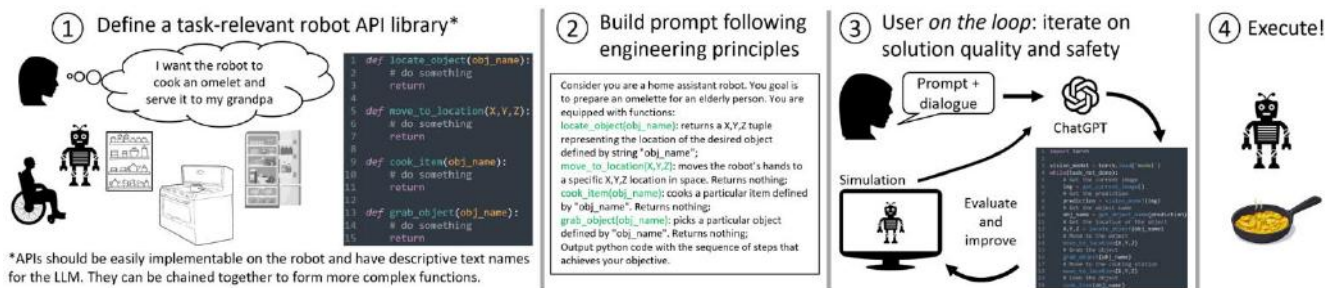
图 18：引入 ChatGPT 后无需工程师在循环中进行调整



数据来源：微软，东方证券研究所

微软团队给出了一套适合为机器人任务编写提示方法的设计原则。首先，定义机器人的功能库函数，比如检测物体、移动底盘等，函数命名要与实际动作符合。第二，为 ChatGPT 编写 prompt，描述任务目标，同时明确说明可用的高级函数库中的哪些函数。提示还可以包含有关任务约束的信息，或者有关 ChatGPT 如何形成其答案（特定编码语言，使用辅助解析元素）的信息。第三，在仿真环境中运行并评估 ChatGPT 的代码输出，保证模型能够安全完整地做好任务。最后，将算法部署到机器人上。

图 19：ChatGPT for Robotics 的设计原则



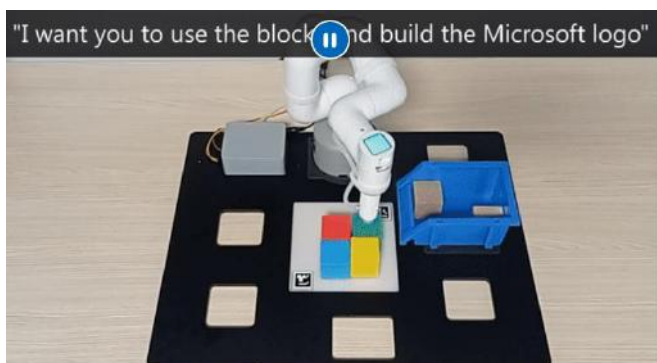
数据来源：微软，东方证券研究所

基于该设计原则可以实现多种 ChatGPT 结合机器人的应用，为进一步研究大模型+通用机器人打下了基础。微软团队也针对几个场景进行了 ChatGPT 结合机器人的测试，包括用无人机检查货架、模拟环境中用无人机进行工业巡检、控制机器人手臂拼积木等。由于 ChatGPT 只能用文字进

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

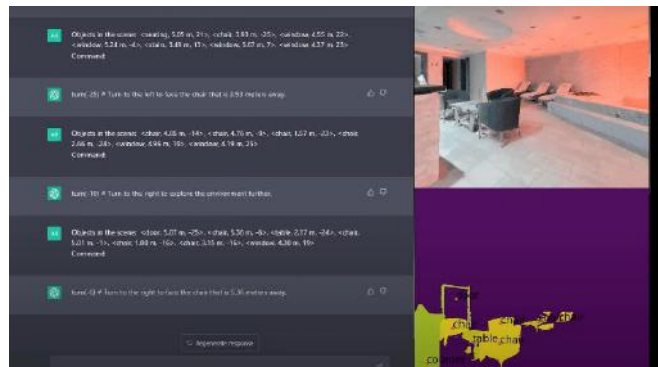
行交互，对于外界环境的感知需要对传感器传回的图像进行对象检测，并转化为文本描述才能输入给 ChatGPT。微软团队的研究初步展示了 ChatGPT 这一大语言模型作为机器人的决策中枢的可能性，为后续更加深度的大模型+机器人探索迈出了第一步。

图 20: ChatGPT 可操作机器臂拼出微软 Logo



数据来源：微软，东方证券研究所

图 21: 需要将外部环境转化为文字信息输入给 ChatGPT



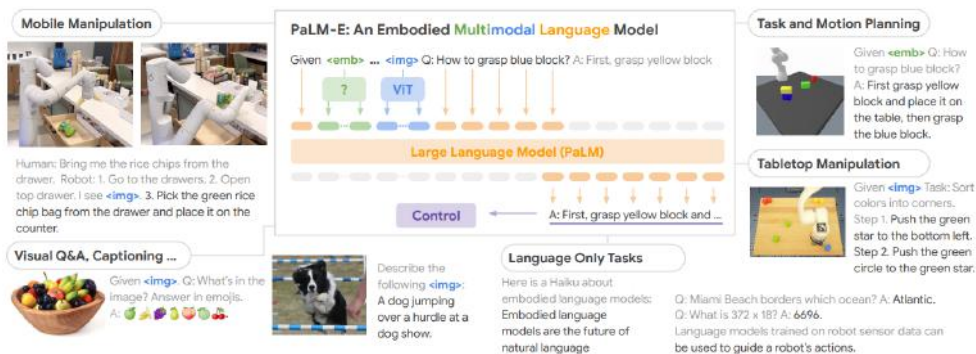
数据来源：微软，东方证券研究所

2) PaLM-E: 最大的多模态视觉语言模型

2023 年 3 月，柏林工业大学和 Google Robotics 团队结合了 Google 当时 5400 亿参数的 PaLM 大语言模型和 220 亿参数的 Vision Transformer (ViT) 模型，提出了当时最大规模的 5620 亿参数的具身多模态视觉语言模型 (Visual Language Model, VLM) PaLM-E。在 PaLM 模型基础上，引入了具身化和多模态概念，实现了指导现实世界机器人完成相应任务的功能。

PaLM-E 采用从多模态信息到决策端的端到端训练。PaLM-E 直接将连续的、具体的多模态观察 (如图像、状态估计或其他传感器模态)，转化为和语言 token 嵌入空间维数相同的向量序列，用和语言 token 同样的方式注入预训练语言模型的语言嵌入空间，从而在文字和感知之间建立联系，已解决机器人相关的具身问题。模型的输入是交错的视觉、连续状态估计和文本组成的多模态编码，然后对这些编码进行端到端训练，输出的内容则是对于机器人要执行的动作的文本决策。整个过程不需要对场景的表示进行预处理。

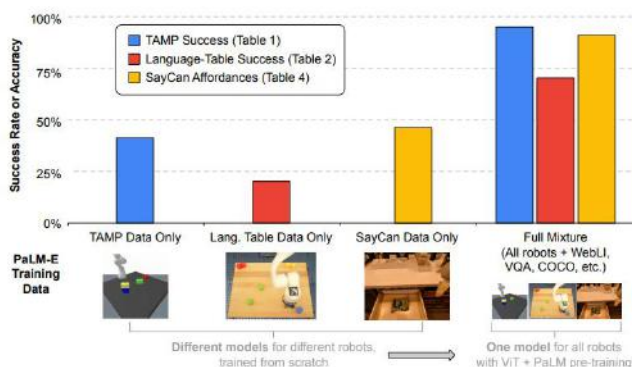
图 22: PaLM-E 能够接收多模态信息并输出决策文本



数据来源：Google《PaLM-E: An Embodied Multimodal Language Model》，东方证券研究所

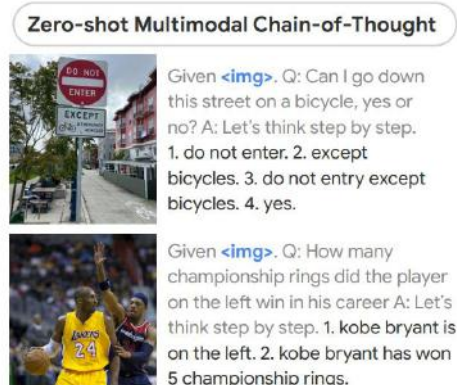
以大模型作为核心的 PaLM-E 表现出了较强的泛化能力和涌现能力。研究人员发现，PaLM-E 继承了大语言模型的核心优点：泛化和涌现能力。得益于端到端的多模态信息训练，PaLM-E 在面对没有学习过的任务（zero-shot）时也能有很好的表现，具备将从一项任务学到的知识和技能迁移到另一项任务的能力。经过不同任务混合训练后的 PaLM-E，与执行单一任务的机器人模型相比，性能明显提高。同时，尽管 PaLM-E 只接受了单图像提示的训练，但却已经展示出了涌现能力，比如多模式思维链推理（可让模型分析包括语言和视觉信息在内的一系列输入）与多图像推理（用多个图像作为输入来做出推理或预测）。

图 23: PaLM-E 具备迁移学习能力



数据来源：Google《PaLM-E: An Embodied Multimodal Language Model》，东方证券研究所

图 24: PaLM-E 可以进行 zero-shot 多模态思维链推理



数据来源：Google《PaLM-E: An Embodied Multimodal Language Model》，东方证券研究所

PaLM-E 展示了大模型和机器人结合的诸多可能性。以大模型为核心的 PaLM-E 有了良好的迁移学习能力，从而可以通过自主学习来完成长跨度规划的任务，比如，“从抽屉里拿出薯片”这类任务包括了多个计划步骤，并且需要调用机器人摄像头的视觉反馈。经过端到端训练的 PaLM-E 可以直接从像素开始对机器人进行规划。由于模型被集成到一个控制回路中，所以机器人在拿薯片的过程中，对途中的干扰具有鲁棒性。并且由于其采用了多模态信息作为输入，相比 ChatGPT for Robotics 论文中需要将图像信息转化为文字输入来说能够获取更多的信息，从而提升机器人模型的性能，能够应用到更广泛的场景中。

3) RT-2: 首个视觉-语言-动作多模态大模型

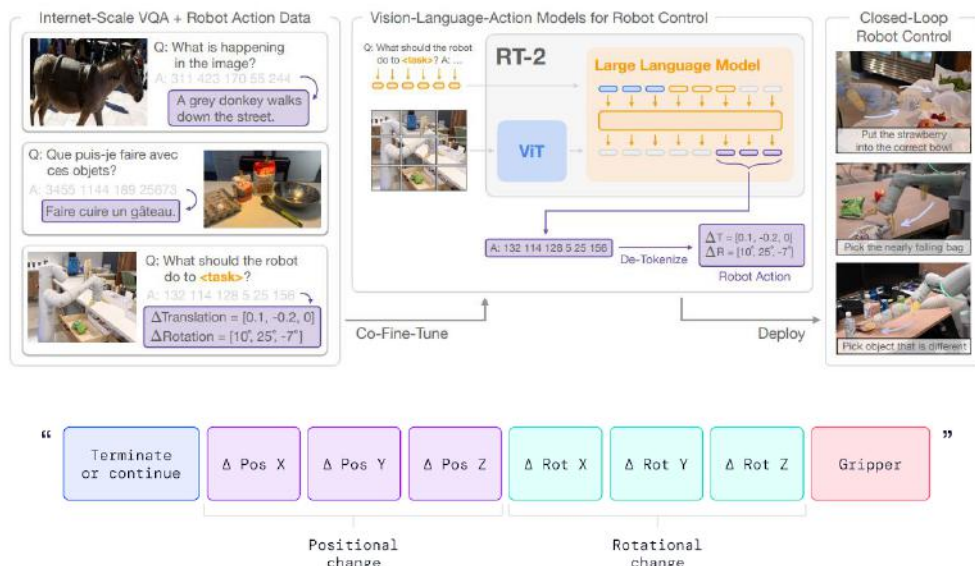
2023 年 7 月，谷歌 DeepMind 推出了 RT-2，这是全球第一个控制机器人的视觉-语言-动作（Vision-Language-Action, VLA）模型。通过将 VLM 预训练与机器人数据相结合，能够端到端直接输出机器人的控制。

RT-2 以 VLM 为基础，升级成为全新的视-语言-动作（VLA）模型。RT-2 将一个或多个图像作为输入，并生成一系列通常代表自然语言文本的标注。例如 RT-2 将原本非常具体的机器人动作数据转变成文本 token，例如将转动度数、坐标点等数据，转变成文本“放到某个位置”。这样一来，机器人数据也能被用到视觉-语言数据集中进行训练，同时在进行推理的过程中，原本的文本指令也会被重新转化为机器人数据，实现控制机器人等一系列操作：为了控制机器人，研究人员通过将操作表示为模型输出中的标注，并将操作描述为可以由标准自然语言标注生成器处理的字符串，这种字符串的示例可以是机器人动作标记编号的序列，例如“1 128 91 241 5 101 127 217”。该字符串以一个标志开始，该标志指示机器人是继续还是终止当前情节，然后机器人根据指示改变

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并阅读本证券研究报告最后一页的免责声明。

末端执行器的位置和旋转以及机器人抓手等命令。由于动作被表示为文本字符串，因此机器人执行动作命令就像执行字符串命令一样简单。有了这种表示，就可以直接对现有的视觉-语言（VLM）模型进行微调，并将其转换为视觉-语言-动作（VLA）模型。

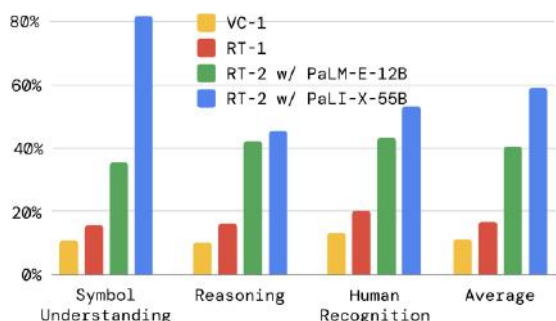
图 25：RT-2 架构和操作字符串的表示方式



数据来源：谷歌《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，东方证券研究所

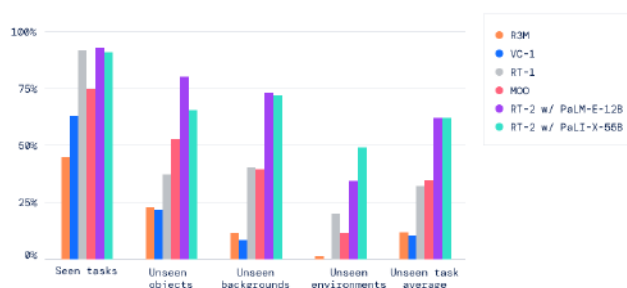
RT-2 较基线模型 RT-1 等而言具有更强的任务处理能力和泛化能力。谷歌在去年推出了 RT-1 机器人模型，RT-2 的训练则延续使用了 RT-1 的演示数据，由 13 个机器人在办公室、厨房环境中，历时 17 个月进行收集。RT-2 的基础 VLM 则选用了不久前推出的 PaLM-E 和更早一些的 PaLI-X，通过测试后发现，RT-2 与之前的模型相比，具有更高的任务成功率（符号理解、推理和人类识别三个方面）以及更强的泛化和涌现能力（在机器人面对以前未见过场景中的性能方面，RT-2 从 RT-1 的 32% 提高到 62%）。

图 26：RT-2 的任务成功率高于基线模型



数据来源：谷歌《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，东方证券研究所

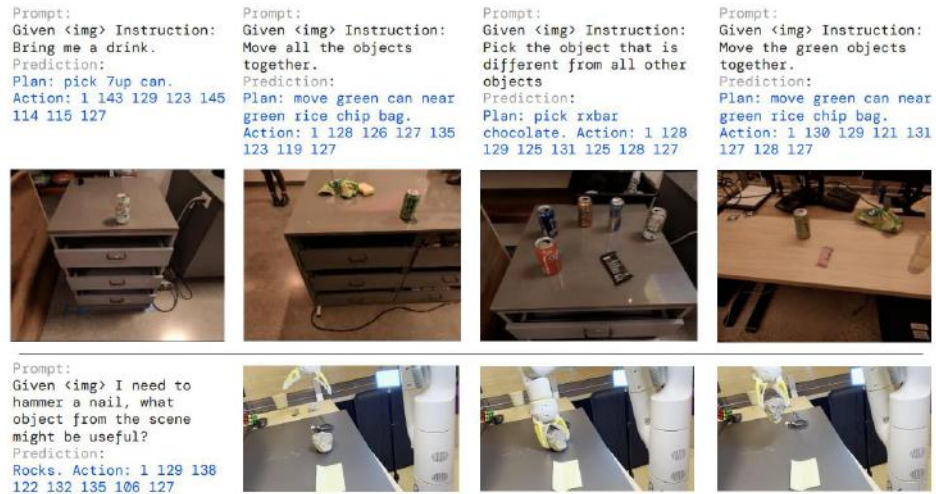
图 27：RT-2 具备比基线模型更强的泛化能力



数据来源：谷歌《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，东方证券研究所

RT-2 具有学习长期规划和简易技能的能力。研究人员将思维链能力加入到了机器人控制中，通过对数据进行增强，添加了一个额外的“计划”步骤，该步骤用自然语言描述机器人即将采取的动作的目的，然后是动作的描述和标注。例如“指令：我饿了。计划：挑选 Rxbar 巧克力。行动：1 128 124 136 121 158 111 255”。通过这一过程，RT-2 就可以先用自然语言规划自己的行动，从而可以执行更多复杂的命令。

图 28：加入思维链能力使得 RT-2 能够处理更复杂的问题



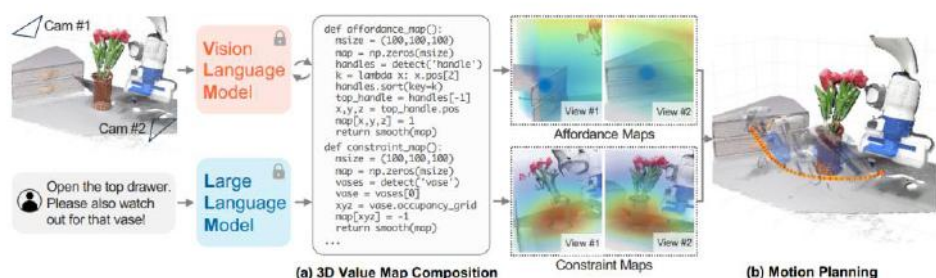
数据来源：谷歌《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，东方证券研究所

4) VoxPoser：机器人任务轨迹规划

2023 年 7 月，斯坦福大学李飞飞团队提出了智能系统 VoxPoser，它可以从大模型 LLM 和视觉-语言模型 VLM 中提取可行性和约束，以构建 3D 仿真环境中的值地图，供运动规划器使用，用于零样本本地合成机器人操纵任务的轨迹，从而实现在真实世界中的零样本机器人操纵。

该模型把复杂指令转化成具体行动规划，人类可以随意地用自然语言给机器人下达指令，机器人也无需额外数据和训练。其原理很简单，首先，给定环境信息（用相机采集 RGB-D 图像）和我们要执行的自然语言指令。接着，LLM 根据这些内容编写代码，所生成代码与 VLM 进行交互，指导系统生成相应的操作指示地图，即 3D 值地图。它是可行性地图和约束地图的总称，既标记了“在哪里行动”，也标记了“如何行动”。再利用动作规划器，将生成的 3D 地图作为其目标函数，便能够合成最终要执行的操作轨迹。相比传统方法需要进行额外的预训练，这个方法用大模型指导机器人如何与环境进行交互，直接解决了机器人训练数据稀缺的问题。

图 29：Voxposer 流程一览图

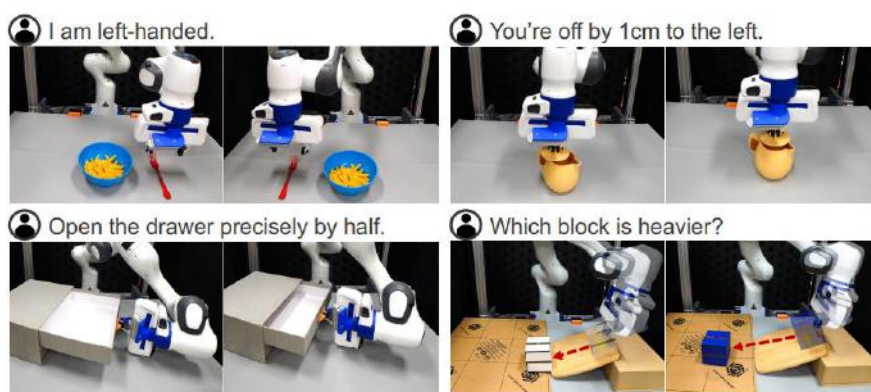


有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并阅读本证券研究报告最后一页的免责声明。

数据来源：Li Fei-Fei 等《VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models》，东方证券研究所

VoxPoser 还具有 4 个新颖的涌现能力。1) 估算物理属性：给定两个未知质量的方块，机器人被要求使用现有工具进行物理实验，确定哪个方块更重。2) 常识性行为推理：在摆桌子的任务中，用户可以指定行为偏好，比如“我是左撇子”，机器人能根据上下文理解其含义。3) 细粒度语言纠正：比如执行“给茶壶盖上盖子”这种精度要求较高的任务时，可以向机器人发出“你偏离了 1 厘米”等精确指令来校正它的操作。4) 基于视觉的多步操作：比如叫机器人将抽屉精准地打开成一半，由于没有对象模型导致的信息不足可能让机器人无法执行这样的任务，但 VoxPoser 可以根据视觉反馈提出多步操作策略，即首先完全打开抽屉同时记录手柄位移，然后将其推回至中点就可以满足要求了。

图 30: Voxposer 的涌现能力



数据来源：Li Fei-Fei 等《VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models》，东方证券研究所

5) RoboAgent: 小数据集下的通用机器人智能体

2023 年 8 月，Meta 和 CMU 团队联合推出了通用机器人智能体 RoboAgent。不同于 DeepMind 的 RT 系列模型采用了大规模机器人数据集进行训练，考虑到机器人在现实世界中的训练和部署效率问题，Meta 将数据集限制到了 7500 个操作轨迹中，并基于此让 RoboAgent 实现了 12 种不同的复杂技能，包括烘焙、拾取物品、上茶、清洁厨房等任务，并能在 100 种未知场景中泛化应用。

图 31: RoboAgent 采用的数据集远远小于 RT-1



数据来源：Meta，东方证券研究所

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

构建一个可以泛化到许多不同场景的机器人智能体需要一个覆盖面足够广的数据集。例如 RT-1 就采用了超过 13 万条机器人操作轨迹数据来进行训练，RoboAgent 则着眼于在有限的的数据下提高机器人学习系统的效率，而较少的数据通常会导致模型过拟合。RoboAgent 的数据集共有 7500 条数据，分别在 4 个厨房场景中采集，包含 12 个技能，38 个任务。基于此小体量的数据，Meta 采用了全自动的数据增广来对数据集进行信息扩充，通过 Meta 提出的“分割一切模型”（Segment Anything Model, SAM）来对图像中的被操作物体和背景进行分割，然后分别对被操作对象和背景进行修改，以达到数据集扩充的目的。

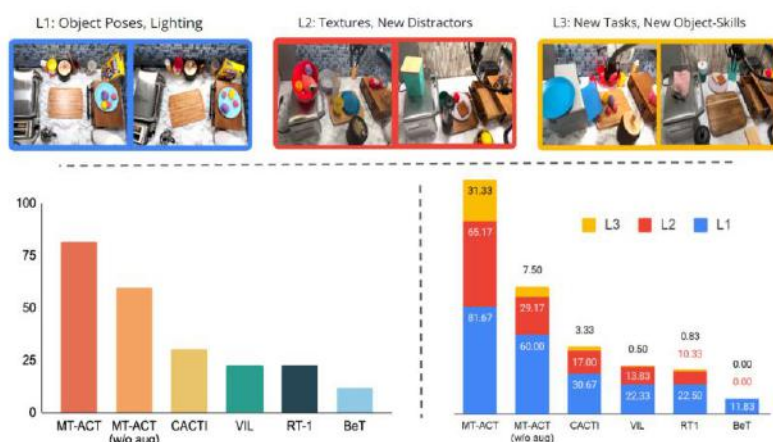
图 32: RoboAgent 采用了数据增广技术对数据集进行扩充



数据来源：Meta《RoboAgent: Generalization and Efficiency in Robot Manipulation via Semantic Augmentations and Action Chunking》，东方证券研究所

RoboAgent 带来了数据效率的提升。Meta 提出了 MT-ACT——多任务动作分块 Transformer 这一学习框架，与一些常见的训练算法框架相比，RoboAgent 体现出了更高的样本效率，并且在多个泛化层面上都有出色表现。Meta 对泛化级别进行了可视化，L1 表示物体姿态变化，L2 表示多样的桌面背景和干扰因素，L3 表示新颖的技能-物体组合。MT-ACT 的表现显著优于其他算法，特别是在更困难的泛化层次（L3）上。

图 33: RoboAgent 有更高的效率和更强的泛化能力

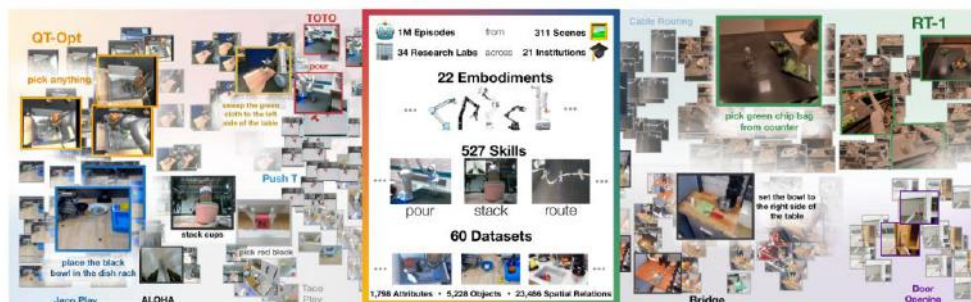


数据来源：Meta《RoboAgent: Generalization and Efficiency in Robot Manipulation via Semantic Augmentations and Action Chunking》，东方证券研究所

6) RT-X 和 Open X-Embodiment Dataset: 机器人的 ImageNet 时刻

10 月 4 日，谷歌 DeepMind 在官网发布了全球最大通用大模型之一 RT-X，并开放了训练数据集 Open X-Embodiment。该数据集来自 21 个研究机构，包含在 22 种类型的机器人上采集的超过 15 万个任务的上百万条数据，展示了 527 项技能。诸多人工智能科学家表示，机器人模型的 ImageNet 时刻到来了。

图 34: Open X-Embodiment 数据集



数据来源：谷歌《Open X-Embodiment: Robotic Learning Datasets and RT-X Models》，东方证券研究所

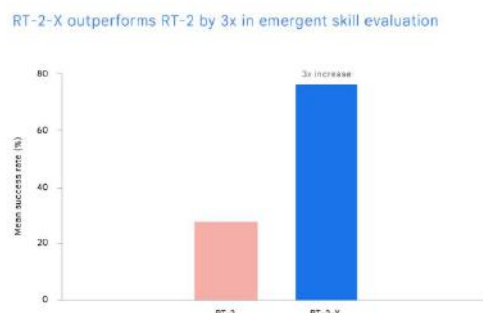
采用 Open X-Embodiment 训练的 RT-X 性能表现大幅提升。DeepMind 分别基于 RT-1 和 RT-2，采用 Open X-Embodiment 数据集进行训练得到了 RT-1-X 和 RT-2-X，得益于更多样化、跨实体的训练数据，性能表现较原有模型均有大幅提升。在相同的模型架构下，使用 Open X-Embodiment 数据集训练的 RT-1-X 平均性能优于原始模型 50%。RT-2-X 则可以执行从未训练的操作，涌现能力达到了 RT-2 的三倍，包括对空间的更好理解，动作的细腻程度等。例如，让机器人把苹果拿到布附近，而不是拿到布上面，RT-2-X 都能很好的区别这些指令的差异，并做出相应的动作。

图 35: RT-1-X 性能表现较原有方法大幅提升



数据来源：谷歌《Open X-Embodiment: Robotic Learning Datasets and RT-X Models》，东方证券研究所

图 36: RT-2-X 的涌现能力是 RT-2 的三倍



数据来源：谷歌《Open X-Embodiment: Robotic Learning Datasets and RT-X Models》，东方证券研究所

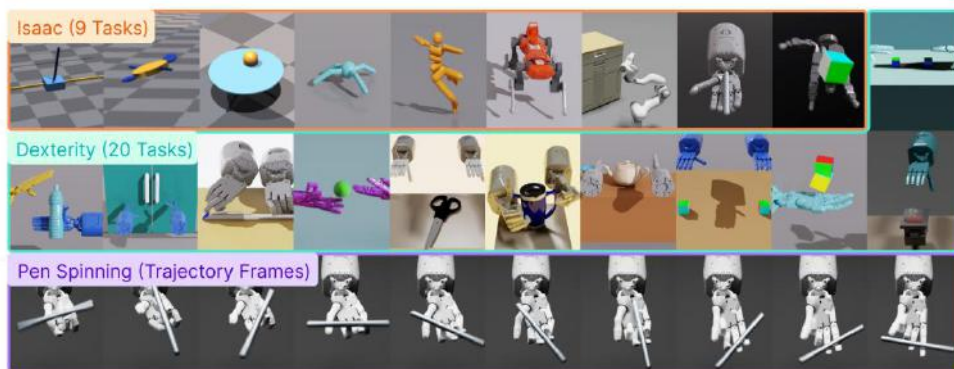
DeepMind 的这项新研究表明，通过利用更多样化的数据和更好的模型进行扩展学习，有可能开发出更有用的辅助机器人。基于大量机器人数据训练得到的通用大模型可以在不同环境下应用，无论是在谷歌 DeepMind 的机器人上，还是在世界各地不同大学的机器人上，其性能都得到了显

著提高。未来的另一个方向是进一步探索不同数据集的混合会如何影响跨具身智能体泛化，以及这种泛化是如何实现的。

7) Eureka: 基于 GPT-4 的 AI 系统

2023 年 10 月，英伟达发布了一种能够自动训练机器人执行新任务的 AI 系统 Eureka。该系统以 OpenAI 的 GPT-4 为基础，本质是一种由大模型驱动的算法生成工具，能训练实体机器人执行例如“转笔”、“开抽屉”、“拿剪刀”、“双手互传球”等多种复杂动作。

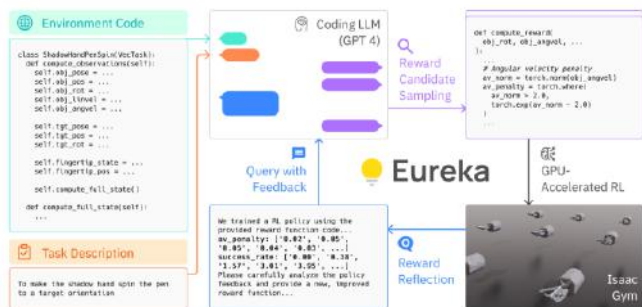
图 37: Eureka 能够训练机器人执行复杂动作



数据来源：英伟达《Eureka: Human-Level Reward Design Via Coding Large Language Models》，东方证券研究所

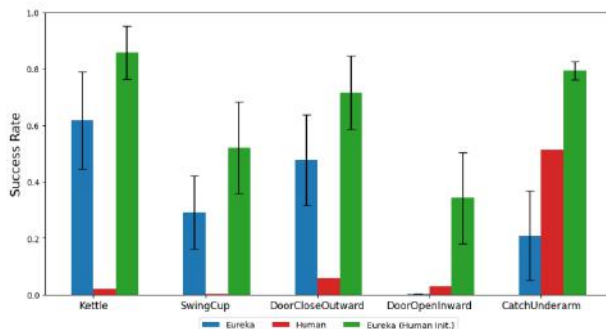
Eureka 的关键创新在于结合生成式 AI 来实现了人类水平的奖励算法设计。Eureka 使用 GPT-4 的零样本生成、代码编写以及上下文改进功能，对奖励执行策略进行了优化，由此通过强化学习来学会执行复杂的技能。Eureka 会基于训练结果汇总关键统计数据，并指导 LLM 改进其奖励函数的生成。在 29 种不同的开源强化学习环境中，Eureka 奖励设计的性能达到了人类水平，这些环境包括 10 种不同的机器人形态（四足机器人、四旋翼机器人、双足机器人、机械手以及几种灵巧手）。在没有任何特定任务提示或奖励模板的情况下，Eureka 生成的奖励程序在超过 80% 的任务上优于专家编写的奖励程序，这使得机器人的平均性能提高了 50% 以上。在以人类设计的奖励程序作为初始条件时，Eureka 的表现还会进一步提升。

图 38: Eureka 工作流程



数据来源：英伟达《Eureka: Human-Level Reward Design Via Coding Large Language Models》，东方证券研究所

图 39: 以人类设计算法为初始条件的 Eureka 表现进一步提升



数据来源：英伟达《Eureka: Human-Level Reward Design Via Coding Large Language Models》，东方证券研究所

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并阅读本证券研究报告最后一页的免责声明。

英伟达的最新成果则展示了大模型在机器人精细化控制方面的作用，能够让机器人首次实现一些高复杂性动作。我们认为，在英伟达的推动下，更智能的机器人学习方式有望被迅速引入工业和消费应用领域。

三、机器人大模型面临的问题

虽然各家科技巨头厂商都在针对通用机器人大模型进行深入探索，但目前来看，距离实现真正的具身智能水平的模型还有很长的路要走。机器人大模型还面临着诸多问题，包括机器人数据、实时性、软硬件融合等。

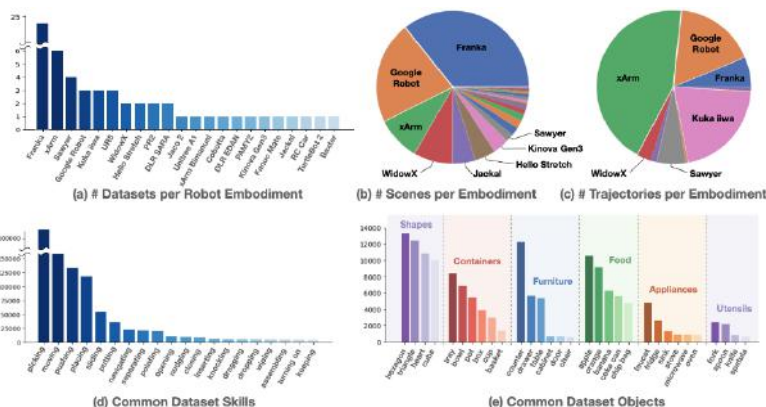
3.1 高质量训练数据缺乏，合成数据有望成为新方向

机器人算法的训练需要大量机器人真实数据。机器人需要通过多种传感器感知环境状态，然后执行实际动作来完成任务，因此训练用于机器人的大模型需要用到大量机器人在真实世界中与环境进行交互的数据集。相比图像和自然语言处理领域可以从网上大量获取训练数据或者通过人类标注快速低成本的获取数据，可用于训练机器人学会执行新任务新技能的高质量数据非常匮乏。主要原因在于：

- 1> 数据丰度问题：ChatGPT 是一个语言模型，从语言角度来说，每个人每分钟能够说出 100-200 字，转化成计算语言就是 1~2 k 字节每秒的传输速率。多模态模型一次输出 4 张图片的传输速率也不过在 MB 级别甚至更低，这个数据会远小于机器关节运动的所产生的信息量。机器人在执行任务时面临的环境和交互内容、产生的数据模态更多更复杂，机器人在单位时间内能够产生的数据要远超语言几个数量级。
- 2> 机器人领域缺乏数据基础：对于 ChatGPT 来说，公开的互联网数据十分充足，有大量的公域数据可供爬取。但现实中的机器人保有量太少，可用于收集训练数据的机器人就更少。此外，每个机器人厂商都会有自己的壁垒，倾向于保护自己的数据，导致机器人领域数据获取难度进一步提升。
- 3> 数据搜集效率：传统 CV/NLP 等任务只需要收集文字、图像等数据，可以通过公开互联网进行爬取，数据采集效率非常高。而机器人领域数据的采集需要真实的机器人去执行任务操作，花费的时间较爬虫相比效率很低。

开源 Open X-Embodiment 数据集是重要进步，但仍旧不足。谷歌开源的机器人训练数据集 Open X-Embodiment 汇集了全球 34 个机器人研究实验室的 60 个现有数据集，包含了在 22 种类型的机器人上采集的超过 15 万个任务的上百万条数据，被誉为机器人领域的 ImageNet 时刻，是机器人模型领域的重要进步。但该数据集主要针对的是常见的操作技能，比如拾取、放置、推拉、移动等，如果要朝着人形机器人的未来发展的话，人形机器人涉及到的全身肢体协调、行走时的平衡等数据的需求量仍较大。

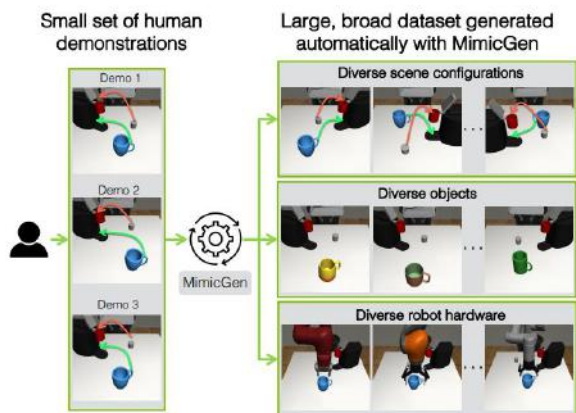
图 40：Open X-Embodiment 数据集一览



数据来源：谷歌，东方证券研究所

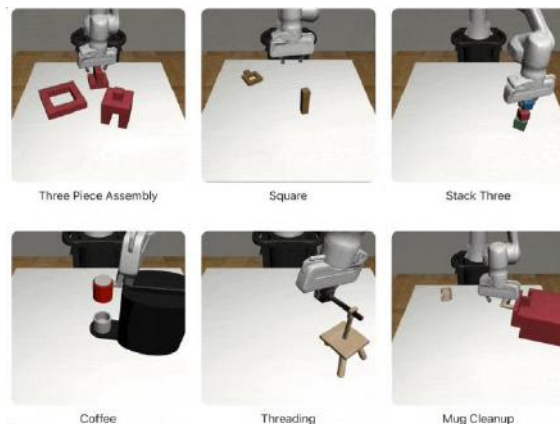
合成数据有望能够解决部分数据不足的难题。英伟达在 10 月末推出了名为 MimicGen 的系统，只需少量人类示范，便能自动生成大规模的机器人训练数据集。使用不到 200 个人类演示，MimicGen 可自主生成超过 5 万个训练数据，涵盖 18 个任务、多个模拟环境。机器人智能体可以通过模仿学习在生成的数据集上进行有效训练，以在长期和高精度任务中获得出色的性能，例如多部件组装和制备咖啡等任务。我们认为，合成数据有望能解决机器人领域高质量训练数据不足、采集效率低下等问题，是一条扩大机器人学习的强大且经济的途径。

图 41: MimicGen 合成机器人任务数据



数据来源：英伟达《MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations》，东方证券研究所

图 42: MimicGen 可生成多种场景的机器人任务数据



数据来源：英伟达《MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations》，东方证券研究所

3.2 实时性仍需提升，算力可能是瓶颈

机器人模型对机器人行动控制的周期仍太长，无法做到实时响应。当前各大科技厂商研究出的算法虽然能够实现机器人技能水平的提升，但是在响应速度方面仍较差。谷歌 RT-2 的演示视频需要 2 倍速播放才能实现比较流畅的机器人动作，而 Voxposer 更是需要 8 倍速才能有流畅的动作表现，如果按原速看的话，能发现 Voxposer 控制的机器人轨迹并不是平滑的，而是有不断的小停顿。这样的迟滞性在真正想要实现通用的人形机器人领域是比较难以接受的。

实时性的提升依赖于算力，对于面对高复杂性环境的人形机器人尤其重要。中科院院士姚期智在 2023 世界机器人大会的演讲中表示，计算能力的挑战非常重要，谷歌研发的 RT-2 模型在机器人

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

控制周期上只能达到 3Hz 的水平，和通用人形机器人需要的 500Hz 差的很远，距离实际需要的控制水平仍有许多事情要做。从动态角度看，人形机器人需要实时控制系统，所有的伺服关节也要同步运动，传感器的数据也要同步采集，从而保证算法的输入和输出都始终处于一个节拍，从而保证算法的性能。所以，实时算法急需高算力的支撑，让系统在一定的时间周期内完成计算。而人形机器人未来的目标是要走进千家万户，这就要求必须具备对复杂环境的适应性算法，包括走路时对地面的适应性、手臂工作时对障碍物的适应性、机器人在人机交互、人机协同时的适应性。这些复杂算法都对控制器的算力提出了巨大的挑战。

图 43: Voxposer 需要 8 倍速播放才有流畅的动作



数据来源: Li Fei-Fei 等《VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models》，东方证券研究所

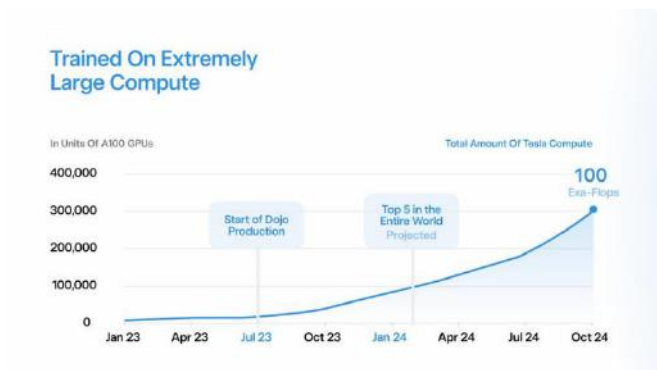
图 44: RT-2 需要 2 倍速播放才有流畅的动作



数据来源: 谷歌《RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control》，东方证券研究所

特斯拉为 Optimus 机器人自研 SoC，Dojo 超算也将为 Optimus 提供算力支持。在特斯拉 2022 年 AI Day 上，特斯拉宣布将基于 FSD 为 Optimus 推出自研的 SoC。2023 年 6 月，特斯拉宣布今年 7 月 Dojo 超算就会正式投产，2024 年 2 月特斯拉的算力规模将进入全球前五，2024 年 10 月特斯拉的算力总规模将达到 100 EFlops，相当于 30 万块英伟达 A100 显卡的算力总和。Dojo 能够处理海量的视频数据，不仅能够加速特斯拉的 Autopilot 和 FSD 系统的迭代，还能为人形机器人 Optimus 提供算力支持。

图 45: 特斯拉 Dojo 超算算力规模预测



数据来源: Tesla AI, 东方证券研究所

图 46: 特斯拉将基于 Dojo 训练机器人模型



数据来源: Tesla AI, 东方证券研究所

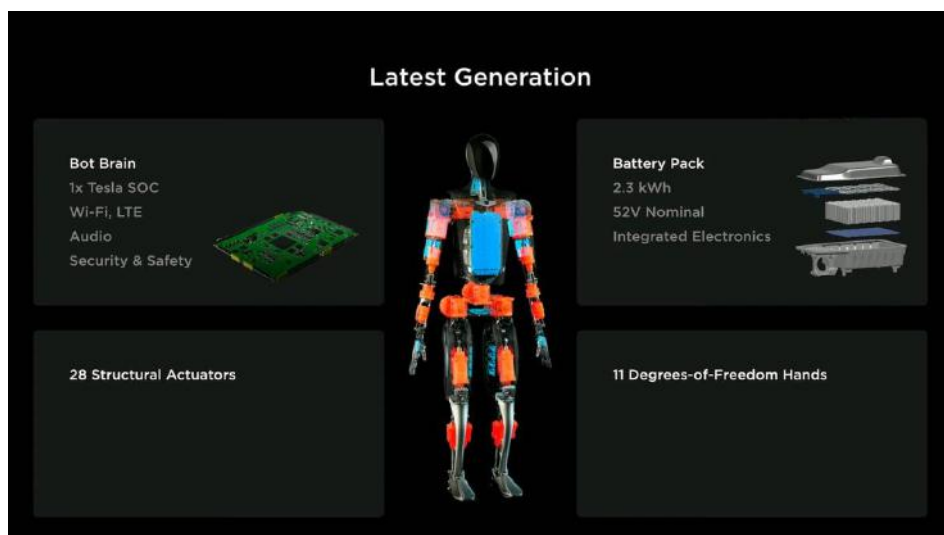
3.3 软硬件融合是最终目标

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

软硬件能力的协同进化也同样重要。波士顿动力创始人 Marc Raibert 在演讲中提到，未来的机器人技术发展过程中，硬件工程和软件同等重要。“有些人认为软件可以克服硬件上所有的问题和限制，我并不赞同这个观点”。以波士顿动力的人形机器人为例，在 Atlas 的硬件工程方面，公司做了大量工作——包括液压系统、多个专门的阀门、专用电池、负载等各个方面，把机器人的重量从 170 公斤缩小到 90 公斤，这个过程中，他们没有在机器人的功能上有任何妥协，反而提高了机器人的运动范围力量和速度。

人形机器人零部件众多，做好软硬件协同才能发挥其硬件水平。人形机器人集成了大量不同的传感器，如：视觉、听觉、位置、柔性传感器等，这些传感器能够让机器人通过“五官”接近人类，收集并感知外界信息。以特斯拉 Optimus 为例，原型机动力系统部分有 28 个结构执行器，50 个基础自由度，全身能够产生 200 个以上不同角度的动作。灵巧手单手有 6 个执行器，11 个自由度，采用金属肌腱带动机器人能够精确抓住小而薄的物体。而在接下来的“分解、规划任务、运动中识别物体、完成高质量的实时交互，最后完成具体任务”一整套作业过程中，更需要把人形机器人的核心软硬件有机整合，协同好各部分的速度、力度、准确性，让这些硬件系统各司其职完成任务，其难度是巨大的。

图 47：Tesla Optimus 机器人硬件结构



数据来源：Tesla Side，东方证券研究所

端到端大模型一定程度上解决软硬件协同问题。端到端的算法能够把机器人各个部位传感器的感知数据直接作为输入，输出硬件层面各个部件的控制动作，一定程度上能够解决软硬件协同的问题。但是这种方法属于“大力出奇迹”的一条道路，需要通过大量的数据训练、甚至是几个数量级以上的数据训练才能够真正得到一个能够接受人形机器人产生的所有数据并能给出输出的超大基础模型，这就又回到了前两个挑战：高质量数据的缺乏和能够支撑这样强的端到端大模型实时给出响应的算力。未来可能在云端会有比较强的一个端到端、且足够泛化的大模型，但是在端侧，到底需要多大能力的芯片，或者说需要有多长多强的工程化的剪枝的能力，才可以将大模型去落到端侧人形机器人呢？这仍将是人形机器人走向通用化面临的重要困难。

投资建议与投资标的

国内顶层设计规划路线，明确量产时间点、关键技术、重点产品以及示范性应用场景，结合海外以特斯拉 Optimus 为代表的人形机器人持续推进，人形机器人产业化迎来曙光。我们看好人形机器人产业趋势，建议关注产业链上的核心零部件相关公司以及大模型公司。

- 总成：三花智控(002050，买入)（汽车组覆盖）、拓普集团(601689，买入)（汽车组覆盖）；
- 空心杯电机：鸣志电器(603728，未评级)、鼎智科技(873593，未评级)；
- 减速器：绿的谐波(688017，未评级)、双环传动(002472，未评级)、中大力德(002896，未评级)；
- 编码器：奥普光电(002338，未评级)、禾川科技(688320，未评级)、汇川技术(300124，未评级)、昊志机电(300503，未评级)、峰岬科技(688279，未评级)、儒竞科技(301525，未评级)；
- 丝杠和轴承：五洲新春(603667，买入)、恒立液压(601100，未评级)、贝斯特(300580，未评级)、秦川机床(000837，未评级)、鼎智科技(873593，未评级)、禾川科技(688320，未评级)；
- 配套高端设备：田中精机(300461，未评级)、华辰装备(300809，未评级)、浙海德曼(688577，未评级)、日发精机(002520，未评级)；
- 传感器：柯力传感(603662，未评级)、昊志机电(300503，未评级)；
- IMU：芯动联科(688582，未评级)、华依科技(688071，未评级)、苏州固锟(002079，未评级)等。
- 大模型：科大讯飞(002230，买入)、鼎捷软件(300378，未评级)、中科创达(300496，买入)等

风险提示

人形机器人进展不及预期：人形机器人目前还处于初期阶段，技术路线等尚未完全确定，发展进程可能不及市场预期；

国产替代进程不及预期：目前国内相关厂商的技术与国外头部厂商相比仍有较大差距，若技术突破进展不及预期，将影响到国产替代进程；

行业竞争加剧：伴随人形机器人放量以及行业降本趋势下，新参与者可能会增加，产品价格可能会有所下降，行业竞争可能会有所加剧；

分析师申明

每位负责撰写本研究报告全部或部分内容的研究分析师在此作以下声明：

分析师在本报告中对所提及的证券或发行人发表的任何建议和观点均准确地反映了其个人对该证券或发行人的看法和判断；分析师薪酬的任何组成部分无论是在过去、现在及将来，均与其在本研究报告中所表述的具体建议或观点无任何直接或间接的关系。

投资评级和相关定义

报告发布日后的 12 个月内行业或公司的涨跌幅相对同期相关证券市场代表性指数的涨跌幅为基准（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）；

公司投资评级的量化标准

- 买入：相对强于市场基准指数收益率 15%以上；
- 增持：相对强于市场基准指数收益率 5% ~ 15%；
- 中性：相对于市场基准指数收益率在-5% ~ +5%之间波动；
- 减持：相对弱于市场基准指数收益率在-5%以下。

未评级 —— 由于在报告发出之时该股票不在本公司研究覆盖范围内，分析师基于当时对该股票的研究状况，未给予投资评级相关信息。

暂停评级 —— 根据监管制度及本公司相关规定，研究报告发布之时该投资对象可能与本公司存在潜在的利益冲突情形；亦或是研究报告发布当时该股票的价值和价格分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确投资评级；分析师在上述情况下暂停对该股票给予投资评级等信息，投资者需要注意在此报告发布之前曾给予该股票的投资评级、盈利预测及目标价格等信息不再有效。

行业投资评级的量化标准：

- 看好：相对强于市场基准指数收益率 5%以上；
 - 中性：相对于市场基准指数收益率在-5% ~ +5%之间波动；
 - 看淡：相对于市场基准指数收益率在-5%以下。
- 未评级：由于在报告发出之时该行业不在本公司研究覆盖范围内，分析师基于当时对该行业的研究状况，未给予投资评级等相关信息。
- 暂停评级：由于研究报告发布当时该行业的投资价值分析存在重大不确定性，缺乏足够的研究依据支持分析师给出明确行业投资评级；分析师在上述情况下暂停对该行业给予投资评级信息，投资者需要注意在此报告发布之前曾给予该行业的投资评级信息不再有效。

免责声明

本证券研究报告（以下简称“本报告”）由东方证券股份有限公司（以下简称“本公司”）制作及发布。

。本公司不会因接收人收到本报告而视其为本公司的当然客户。本报告的全体接收人应当采取必要措施防止本报告被转发给他人。

本报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司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证券研究报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的证券研究报告之外，绝大多数证券研究报告是在分析师认为适当的时候不定期地发布。

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，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人作出邀请。

本报告中提及的投资价格和价值以及这些投资带来的收入可能会波动。过去的表现并不代表未来的表现，未来的回报也无法保证，投资者可能会损失本金。外汇汇率波动有可能对某些投资的价值或价格或来自这一投资的收入产生不良影响。那些涉及期货、期权及其它衍生工具的交易，因其包括重大的市场风险，因此并不适合所有投资者。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者自主作出投资决策并自行承担投资风险，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。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面协议授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容。不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

经本公司事先书面协议授权刊载或转发的，被授权机构承担相关刊载或者转发责任。不得对本报告进行任何有悖原意的引用、删节和修改。

提示客户及公众投资者慎重使用未经授权刊载或者转发的本公司证券研究报告，慎重使用公众媒体刊载的证券研究报告。

东方证券研究所

地址：上海市中山南路 318 号东方国际金融广场 26 楼

电话：021-63325888

传真：021-63326786

网址：www.dfzq.com.cn

东方证券股份有限公司经相关主管机关核准具备证券投资咨询业务资格，据此开展发布证券研究报告业务。

东方证券股份有限公司及其关联机构在法律许可的范围内正在或将要与本研究报告所分析的企业发展业务关系。因此，投资者应当考虑到本公司可能存在对报告的客观性产生不利影响的情况。