

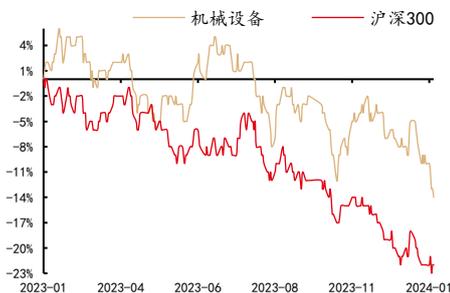
行业投资评级

强于大市 | 维持

行业基本情况

收盘点位	1301.16
52周最高	1606.0
52周最低	1301.16

行业相对指数表现（相对值）



资料来源：聚源，中邮证券研究所

研究所

分析师: 刘卓
SAC 登记编号: S1340522110001
Email: liuzhuo@cnpsec.com
研究助理: 傅昌鑫
SAC 登记编号: S1340123050006
Email: fuchangxin@cnpsec.com

近期研究报告

《挖掘机专题：国内市场静待回暖，国际化持续推进》 - 2024.01.14

机器人如何进行复杂操作和自主移动

● 投资要点

2024年1月，斯坦福大学团队在社交媒体上发布了与谷歌 DeepMind 团队合作研发的能炒菜、能做家务的双臂机器人 Mobile ALOHA。时隔不久，特斯拉也发布了关于机器人执行折叠衣物等复杂操作的演示视频，这些事件引发了人们对于人形通用机器人时代的遐想。Mobile ALOHA 依托系统收集的数据，通过人类示教动作学习模仿操作技巧与身体控制。其硬件配置并不高，整套价格仅需 3.2 万美元，且解决方案为开源。上述成本和配置就能打造出在家庭环境下能够实现诸多功能的 Mobile ALOHA，其关键在于机器人的运动控制及交互，和自主移动两方面。目前人形机器人表现出的决策执行能力已经逐步向过往的设想靠拢，今年或将成为人形机器人应用的元年，下文中将讨论机器人是怎样完成复杂操作及自主移动的。

机器人如何进行复杂操作？多种 AI 训练路径实现运动控制及环境交互。 ChatGPT 引起浪潮后，以谷歌 PaLM-E 为代表的多模态具身 VLM 不断面世，人形机器人领域也出现了许多不同的 AI 训练方法。目前比较主流的五种 AI 训练思路：虚拟仿真、动作捕捉或遥操作、模仿学习、VLM+小模型和 VLA。每种方式都有其优劣之处，各家公司会通过尝试及融合多种训练方式，以寻找最为适合的 AI 路径。虚拟仿真指在虚拟环境中构建智能体并进行强化学习算法训练，代表为英伟达的 VIMA；动作捕捉或遥操作指将真人动作通过传感器等设备精准复制给机器人，机器人接收并收集训练信息从而学会技能，代表为特斯拉等人形厂商；模仿学习指机器人通过观察人类现场演示或端到端学习人类示教视频，理解行动中的底层逻辑，进而复现人类演示的技能，代表为斯坦福 MimicPlay；VLM+小模型是由 VLM 大模型负责高层规划，小模型做低层运动控制，代表为谷歌 PaLM-E 和 MetaImage Bind；VLA 是从 VLM 中演化出来的，直接输出运动控制指令，构成感知-决策-动作的闭环过程，代表为谷歌 DeepMind RT-2。

机器人如何实现自主移动？SLAM 技术应用推进其发展。 SLAM 是实现机器人自主导航和后续交互的关键技术之一，是以定位和建图两大技术为目标的算法。根据传感器的不同，机器人用的 SLAM 算法可以分为二维激光 SLAM、三维激光 SLAM 以及视觉 SLAM。不同的 SLAM 算法，实现的具体细节会有所不同，一般通用架构会包含数据预处理、前端、后端、闭环检测、地图构建等。不同方案的 SLAM 均有其应用空间，三维激光、视觉 SLAM 更适合于人形机器人应用场景。国内多家人形机器人厂商已经在自主移动领域应用了 SLAM 技术，未来随着移动机器人、自动驾驶等领域 SLAM 技术的迁移，应用范围将更为广阔。

● 风险提示：

人形机器人下游发展不及预期风险；算法及算力更新迭代不及预期风险；机器人自主移动技术竞争加剧风险；技术落地不达预期风险。

目录

1 斯坦福推出擅长家务的 Mobile ALOHA, 特斯拉 Optimus 能叠衣服	5
2 机器人如何进行复杂操作: 多种 AI 训练路径实现运动控制及环境交互	8
2.1 虚拟仿真 (Sim2Real)	8
2.2 动作捕捉或遥操作 (Tele-operation)	9
2.3 模仿学习 (Mimic Play)	10
2.4 VLM (Vision Language Models) +小模型	12
2.5 VLA (Vision Language Action)	13
3 机器人如何实现自主移动: SLAM 技术的应用	14
4 风险提示	18

图表目录

图表 1: Mobile ALOHA 自主控制炒虾仁	5
图表 2: 研发人员遥操作 Mobile ALOHA 训练并掌握技能	6
图表 3: Mobile ALOHA 机器人硬件情况	6
图表 4: 特斯拉人形机器人 Optimus 叠衣服	7
图表 5: 主流 AI 训练思路优劣对比及代表团队	8
图表 6: VIMA 接收多模态 prompts 并执行指定任务	9
图表 7: Optimus 工程人员穿戴设备为 AI 训练做演示	10
图表 8: 基于 MimicPlay 框架的模仿学习的三阶段	11
图表 9: 通过模仿学习后机器人执行任务的成功率显著提升	11
图表 10: Figure01 操作咖啡机 (左) 及自我纠错 (右)	12
图表 11: PaLM-E 为机器人决策并抵抗任务期间的干扰	13
图表 12: VLM 从网络和 RT-1 机器人数据中学习从而演化成新的 VLA	14
图表 13: VLA 模型控制机器人抓取灭绝动物	14
图表 14: 涌现能力成功率对比	14
图表 15: SLAM 与构成自主移动导航的各领域关系图	15
图表 16: SLAM 根据传感器不同的分类情况	16
图表 17: 6 种代表性 SLAM 算法的综合比较	17
图表 18: 2022-2029E 全球 SLAM 市场规模 (亿美元)	17
图表 19: SLAM 技术的其他应用场景及国内代表公司	18

1 斯坦福推出擅长家务的 Mobile ALOHA，特斯拉 Optimus 能叠衣服

2024 年首周，斯坦福大学团队在社交媒体上发布了与谷歌 DeepMind 团队合作研发的能炒菜、能做家务的双臂机器人 Mobile ALOHA，引发了人们对于人形通用机器人时代的遐想。

图表1: Mobile ALOHA 自主控制炒虾仁



资料来源：斯坦福 Mobile ALOHA 项目开源网页，中邮证券研究所

斯坦福大学团队重心在于机器人通过人类示教动作学习模仿操作技巧与身体控制，即依托 Mobile ALOHA 系统收集的数据进行模拟学习的训练。其论文中对于 Mobile ALOHA 的定义是一种用于收集双臂运动数据的低成本全身遥操作系统。在视频演示中，Mobile ALOHA 迅速且流畅地完成了包括炒虾、清理桌面、水洗盘子、整理桌椅等多个非标准化的动作。根据项目团队的描述，为了实现相关功能，实验人员会首先操作机器人对每个任务的流程进行 50 次学习。经过协同训练后的机器人在自主完成炒虾、存放锅具、呼叫电梯等复杂移动和操作任务时成功率可以达到 90%。同时，团队还设计了远程操作系统，让用户可以通过远程控制的方式控制机器人作业。

图表2：研发人员遥操作 Mobile ALOHA 训练并掌握技能


资料来源：《Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation》(Zipeng Fu et al.)，中邮证券研究所

Mobile ALOHA 的硬件配置并不高，整套价格仅需 3.2 万美元，且解决方案为开源。机器人的整套系统主要由以下几个部分组成：

移动底座——一台 AgileX Tracer AGV

运动系统——两个 Viperx 300 6dof 机械臂

视觉系统——3 个 480×640 分辨率、50Hz 的罗技 C922x RGB 摄像头

控制系统——一台配备 Nvidia 3070 Ti GPU (8GB VRAM) 和 Intel i7-12800H 的消费级笔记本电脑

供电系统——1.26kWh、重 14kg 的电池

作为一种轻型机器人，Mobile ALOHA 全重只有 75kg，在 1.5m 的高度可以施加 100N 的拉力，移动速度可达 1.6m/s，类似于人类的平均步行速度，最多可承重 100kg，最长工作 12h。

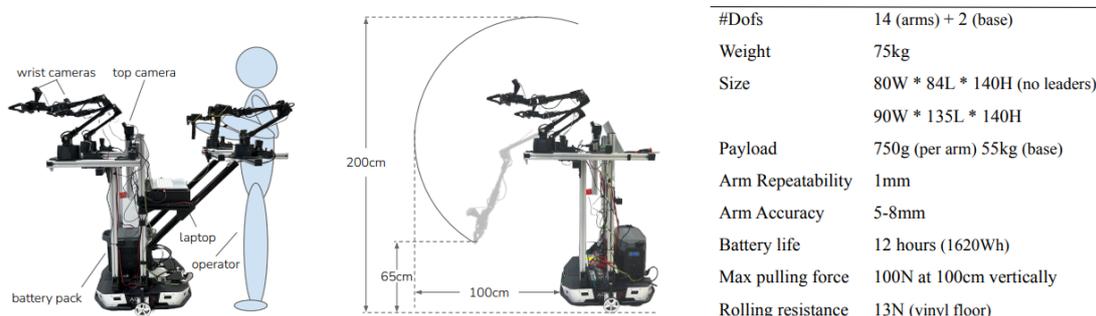
图表3：Mobile ALOHA 机器人硬件情况


Figure 2: Hardware Details. Left: Mobile ALOHA has two wrist cameras and one top camera, with onboard power and compute. Middle: The teleoperation setup can be removed and only two ViperX 300 [3] are used during autonomous execution. Both arms can reach a min/max height of 65cm/200cm, and extends 100cm from the base. Right: Technical specifications of Mobile ALOHA.

资料来源：《Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation》(Zipeng Fu et al.)，中邮证券研究所

上述成本和配置就能打造出在家庭环境下能够实现诸多功能的 Mobile ALOHA，其关键在于机器人的运动控制及交互，和自主移动两方面。机械手臂的运动控制及同作业环境的交互，是该项目的技术含量所在。同现阶段通过强化学习、大模型等驱动机器人自主认知环境、执行任务不同，Mobile ALOHA 采用的是更为直接的“模仿学习”，通过人类示范教学，机器人学习并模仿人类行为，从而构成机器人的行为逻辑；机器人的自主移动能力在 Mobile ALOHA 中主要体现在作为移动底座的 AGV 机器人上，该 AGV 机器人配备了一整套传感器，以支持室内 SLAM、导航和基于视觉的应用，实现在定位导航、路径规划方面的能力。

时隔不久，特斯拉也发布了关于机器人执行复杂操作的演示视频。1 月 16 日，特斯拉 CEO 埃隆马斯克在社交媒体 X 上发布了公司人形机器人 Optimus 叠衣服的演示视频，视频中 Optimus 展示出的操作速度虽然缓慢但十分自然，流畅度令人印象深刻。尽管之后马斯克表示视频中的 Optimus 并非完全自主行动，但该视频仍在互联网上引发热议。目前特斯拉通过远程控制机器人完成越来越复杂的任务，一方面是为了验证现阶段硬件能否灵活完成各项任务，另一方面是为了收集端到端神经网络所需的训练数据，从而让未来的机器人能够自行完成任务。马斯克也表示，未来 Optimus 势必可以在更任意的环境中完全自主执行复杂的任务。

图表4：特斯拉人形机器人 Optimus 叠衣服



资料来源：网易，中邮证券研究所

从斯坦福 Mobile ALOHA、特斯拉 Optimus 等人形机器人产品的新进展中，可以看到人形机器人已经能够完成一定程度的复杂任务，以及在特定环境下自主移动。同时，1 月初谷歌 DeepMind 发布了三项大模型新成果 AutoRT、SARA-RT 和 RT-Trajectory 的引入，助力现实世界机器人数据收集、速度及泛化。近期的事件引发了市场对于人形机器人智能进步的关注，目前人形机器人表现出的决策执行能力已经逐步向过往的设想靠拢，今年或将成为人形机器人应用的元年，下文中将讨论机器人是怎样完成复杂操作及自主移动的。

2 机器人如何进行复杂操作：多种 AI 训练路径实现运动控制及环境交互

ChatGPT 引起浪潮后，以谷歌 PaLM-E 为代表的多模态具身 VLM 不断面世，人形机器人领域也出现了许多不同的 AI 训练方法。通过对机器人进行 AI 训练，提升其“大脑”水平，从而使得机器人能够正确的执行动作及与环境中的被操作对象交互。大致总结下目前比较主流的五种 AI 训练思路：虚拟仿真、动作捕捉或遥操作、模仿学习、VLM+小模型和 VLA。每种方式都有其优劣之处，最终哪种训练方式会成为主流目前不得而知，各家公司可能会通过尝试及融合多种训练方式，以寻找最为适合的 AI 训练路径。

图表5：主流 AI 训练思路优劣对比及代表团队

	优势	劣势	代表团队
虚拟仿真	训练成本相对较低	仿真数据与现实情况偏差较大	英伟达
动作捕捉/遥操作	过程直接有效，适合实际应用场景，与人形机器人匹配度高	数据采集成本高，且核心依赖真人	特斯拉等多家人形机器人厂商
模仿学习	可泛化性强，省略繁琐编程	科研场景居多，泛化后可靠性存疑	斯坦福大学、卡内基梅隆大学、Figure
VLM	可以执行复杂指令而无需重新训练，且能形成细分场景数据壁垒	决策端和执行端割裂带来的时滞，小模型数据获取难度大	谷歌、Meta
VLA	感知-决策-动作闭环，避免了复杂的信息转换	大模型实时性较差，执行成功率有待提升	谷歌 DeepMind

资料来源：中邮证券研究所

2.1 虚拟仿真 (Sim2Real)

虚拟仿真 (Sim2Real)，全称 Simulation to Reality，是指在虚拟环境中构建一个智能体并进行强化学习算法的训练，对于 AI 的训练是基于虚拟环境的结果。其为强化学习的一个分支，同时也属于 transfer learning 的一种，其主要解决的问题是在机器人领域中，直接让机器人或者机械臂在现实环境中与环境进行交互、采样时，会出现以下两个比较严重的问题：

一是**采样效率太低**。在应用强化学习算法解决机器人相关问题时，所需要的样本量一般会达到上千万，现实环境中采集如此数量级的样本要耗费几个月甚至更长的时间；

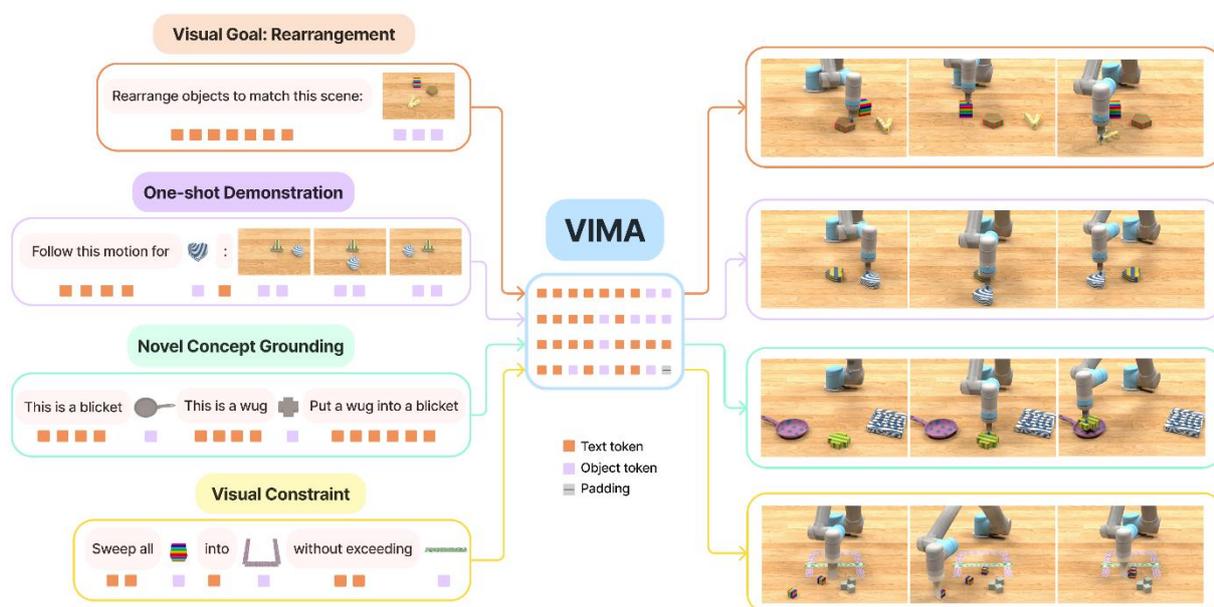
二是**安全问题**。由于强化学习需要通过智能体在环境中进行大范围的随机采样来进行试错，因而在某些时刻其做出的行为可能会损伤机器人自身（例如手臂

转动角度过大或者避障任务中由于碰撞造成的不可逆损伤等)或者周围的环境及生物。

在模拟器中进行强化学习算法的训练能够解决上述问题,但同时会引出一个新的问题 (Reality Gap)。由于模拟器对于物理环境的建模都是存在误差的,因而在模拟环境中学习到的最优策略无法直接在现实环境中应用。

此路径中的典型例子为多模态具身 AI 系统 VIMA。VIMA 诞生于 2022 年 10 月,是一个基于 Transformer 架构的仿真机器人智能体,由来自英伟达、斯坦福大学、加州理工学院等多方的研究者们合作完成。VIMA 智能体能够接受提示 (Prompt) 输入,且输入可以是文本、图像、视频等多模态混合,然后输出动作并完成指定任务。不过仿真环境下的被操作对象非常清晰,无法保证真实环境中的成功率。

图表6: VIMA 接收多模态 prompts 并执行指定任务



资料来源: VIMA 项目开源网页, 中邮证券研究所

此外,也有针对如灵巧手等的虚拟仿真训练。北京大学王鹤课题组与北京通用人工智能研究院合作改进了现有灵巧机械手抓取物体的合成方法,提供了一套高效的程序框架用以合成多样、稳定的抓取姿势数据。通过该框架,团队为 ShadowHand 生成了大规模仿真数据集 DexGraspNet,包含 133 类 5355 个物体的 132 万抓取数据,在数量和质量上优于现有灵巧机械手抓取数据集。

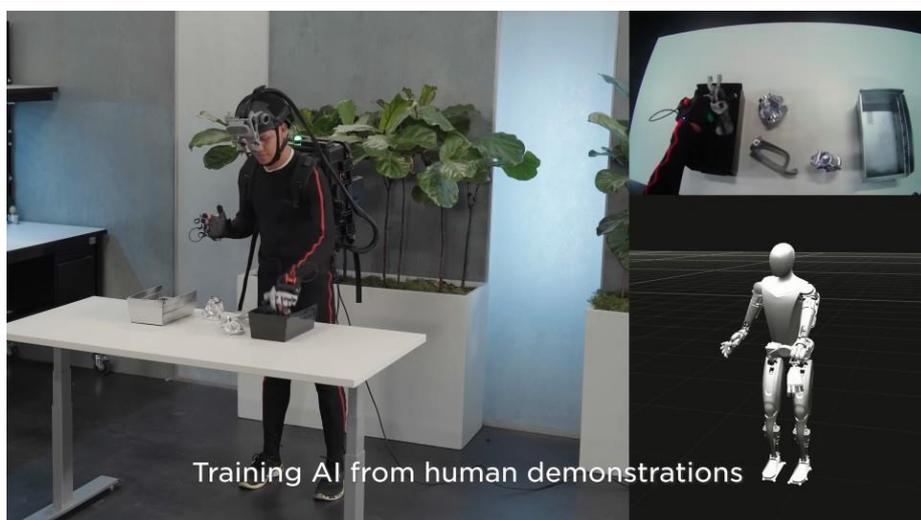
2.2 动作捕捉或遥操作 (Tele-operation)

动作捕捉或遥操作是指将真人执行指令的动作通过传感器等设备精准复制给机器人，输入的信息和行动准确且有针对性，机器人接收并同步收集训练信息，从而学会技能，其训练需要依靠真人重复操作。

不论动作捕捉或是遥操作，均是通过专用设备将人类的动作直接迁移到机器人身上，再由机器人通过传感器收集数据，以达到输入的目的。遥操作在科研技术层面并不是一个十分有潜力的路径，但考虑到机器人更多属于工程应用层面的创新，所以动捕或遥操作在实际应用中较为直接有效，能够解决现实场景的刚需，尤其适合人形机器人使用。

除开前文中提到 Mobile ALOHA 有配置遥操作的途径，特斯拉等多家机器人厂商都有使用过动捕或遥操作的训练路径。特斯拉在 2023 年股东大会上展示了 Optimus 的进展，视频中工程人员穿戴设备进行演示，可以看到人类抓取木品的动作被 AI 算法精准识别并复制到机器人身上。

图表7: Optimus 工程人员穿戴设备为 AI 训练做演示



资料来源：集微网，中邮证券研究所

遥操作相比其他训练路径更为直接有效，且能够解决工业、极端危险等应用场景的刚需。比如太空探索、深海检测、核电站维护等。不仅如此，腾讯 Robotics X 在四足机器人领域也有基于真狗的动捕思路。

2.3 模仿学习 (Mimic Play)

模仿学习是指机器人通过观察人类现场演示或端到端学习人类示教视频，理解行动中的底层逻辑，进而复现人类演示的技能。过程中能够省去繁琐的编程，通过像人类一样从模仿中学习，从而最终达到与人脑类似的决策执行。

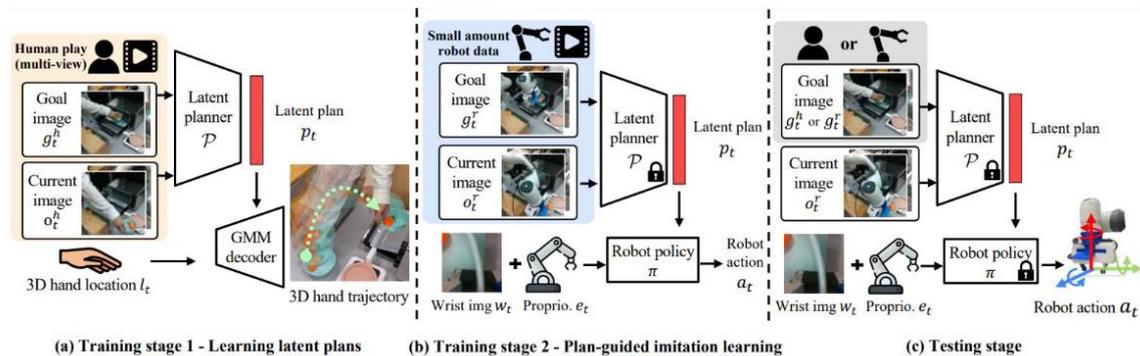
模仿学习的过程是机器人观察人类演示后迁移到自己本体上去执行任务。可以说，从人类演示中进行模仿学习，是训练机器人在现实世界中掌握技能的一种朝阳路径，其过程可泛化性强，尤其适合日常生活场景，可以通过迁移掌握人的

各类零散动作如打开橱柜、拧紧螺丝等，省略原本需要为机器人进行的逐一编程。不过，学习复杂的长期任务通常需要大量的演示，目前多为科研团队应用，量少、任务较简单，未来应用的可靠性无法保证。

2023年斯坦福和英伟达合作研究，引入了名为MimicPlay的分层学习框架，前文提及的Mobile ALOHA正是通过模仿学习掌握了诸多家务技能。为减少模仿学习对于数据的大量需求，研究团队诉诸于人们用手与环境自由交互的系列视频，认为人类视频中包含丰富而明显的物理交互信息，能够促进机器人的决策学习。MimicPlay框架从人类交互数据中学习潜在行动逻辑，用以指导基于少量遥操作演示训练的低级视觉运动控制，最终实现在任务成功率、泛化能力和对干扰的鲁棒性上的出色表现。

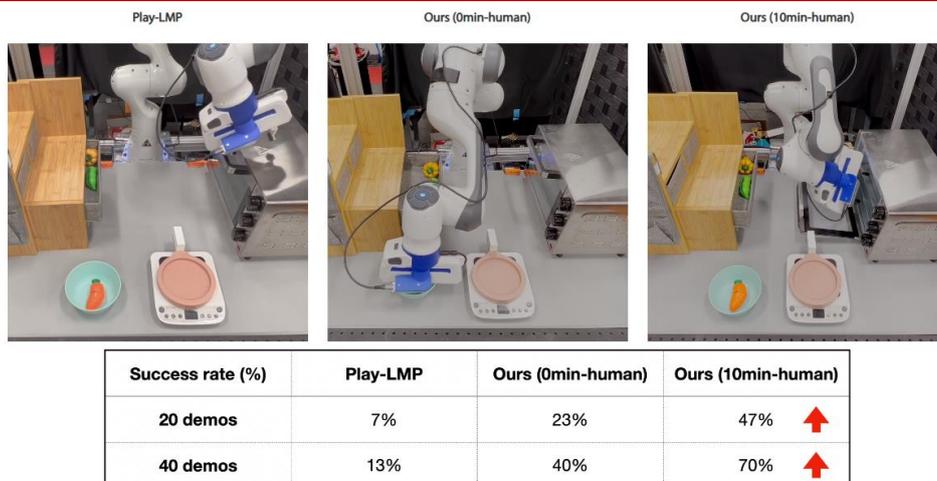
图表8：基于MimicPlay框架的模仿学习的三阶段

Method



资料来源：《MimicPlay: Long-Horizon Imitation Learning by Watching Human Play》(Chen Wang et al.)，中邮证券研究所

图表9：通过模仿学习后机器人执行任务的成功率显著提升

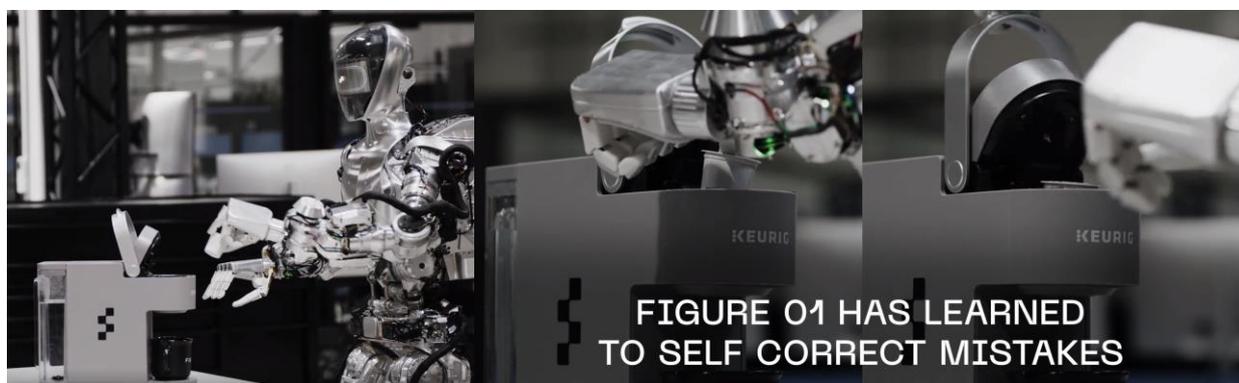


资料来源：《MimicPlay: Long-Horizon Imitation Learning by Watching Human Play》(Chen Wang et al.)，中邮证券研究所

卡内基梅隆大学团队也提出了一种高效、一次性机器人学习算法，称为 **WHIRL (In-the-Wild Human Imitating Robot Learning)**。模仿学习和强化学习虽然较为有效，但存在样本效率低以及局限于实验室环境等问题，该算法以第三人称视角学习为中心，基于采样做策略优化，以提高样本效率。

Figure01 仅用 10 小时模仿学习就学会煮咖啡，并能在过程中自我纠错。2024 年 1 月 8 日，美国 AI 机器人公司 Figure 创始人和首席执行官 Brett Adcock 对外公布了一段 Figure01 使用咖啡机煮咖啡的视频，宣称仅用了 10 小时端到端人工智能训练（神经网络接收人类行为视频并观察学习），就教会了 Figure01 此项技能。

图表10: Figure01 操作咖啡机（左）及自我纠错（右）



资料来源：搜狐，中邮证券研究所

2.4 VLM (Vision Language Models) +小模型

VLM+小模型是指依托 VLM 大模型思考解决高层规划问题，在末端应用场景中调度相对应的小模型以完成运动控制和执行，其本质上是进行通用大模型的训练和应用。

VLM 是一种结合视觉和文本理解的模型，能够同时处理图像、语言等多种形式的信息，其优势在于能够理解和生成包含视觉元素的复杂内容，从而在应用中表现出色。以谷歌为代表的公司将 VLM 应用于机器人领域的训练，通过积累机器人动作库，再使用大模型调度小模型，能够形成细分场景的数据壁垒，而且由于大小模型各司其职，无需反复训练，在标准化场景中效率颇高。其风险和劣势也较为明显，一是执行速度较慢，这主要同大模型是部署在云端还是边缘侧相关，VLM 负责决策而被调度的小模型负责执行，期间存在信息传输转化的时滞；二是小模型数据获取难度较大，数据安全及隐私问题可能成为绊脚石。

谷歌的 **Say-Can**、**PaLM-E** 以及 Meta 的 **ImageBind** 都属于 VLM 路径。以谷歌 2023 年 3 月发布的 PaLM-E 模型为例，PaLM-E (Pathways Language Model with Embodied) 是 PaLM-540B 语言模型与 ViT-22B 视觉 Transformer 模型的结合。作

为一种多模态具身 VLM，它不仅可以理解图像，还能理解、生成语言，执行各种复杂的机器人指令而无需重新训练。PaLM-E 还展示出了“正迁移”能力，即将从一项任务中学到的知识和技能迁移到另一项任务中。另外，将 PaLM-E 集成到机器人控制回路中后，它可以抵抗任务期间可能发生的中断。比如研究人员从机器人手中抓取薯片并移动，但它仍可以找到薯片并再次抓取。

图表11：PaLM-E 为机器人决策并抵抗任务期间的干扰



资料来源：澎湃，中邮证券研究所

VLM 在实际运动控制问题上会出现实时性较差的缺陷，故目前主流方案是 VLM 大模型+小模型，通过大小模型相互配合，各取所长，推动人形机器人的实际应用落地。未来在大小模型耦合的路径上，针对应用场景开发出小模型动作库可能成为行业壁垒。

2.5 VLA (Vision Language Action)

VLA 是从 VLM 中演化出来的，两者有诸多相同，其根本差异在于 VLA 将决策和执行的全过程纳入控制范围，直接输出运动控制指令，构成感知-决策-动作的闭环过程。谷歌 DeepMind 于 2023 年 7 月推出的 RT-2 就是一种新颖的 VLA 模型，它能够从网络和机器人数据中学习，并将知识转化为机器人控制的通用指令。具体而言，VLM 在网络规模数据上预先训练，再从 RT-1 机器人数据中学习，最终成为 RT-2，一种可以控制机器人的 VLA 模型。

图表12: VLM 从网络和 RT-1 机器人数据中学习从而演化成新的 VLA



资料来源: DeepMind, 中邮证券研究所

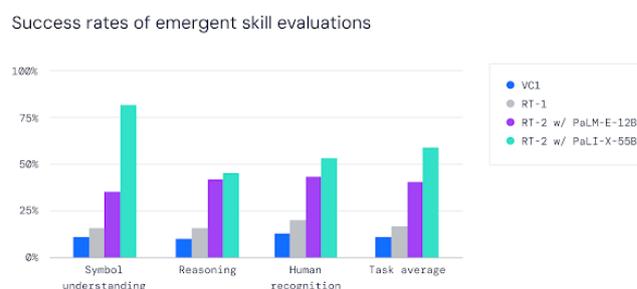
通过微调 VLM 模型生成一系列自然语言文本标记,从而输出动作指令控制机器人。一个或多个图像输入进 VLM 中,然后在模型输出端将操作描述为可由标准自然语言标记器处理的字符串,经过微调预训练的模型接收的是机器人摄像头中的图像,输出的则是机器人要执行的动作。完成后 RT-2 可以控制机器人变得超乎想象,只需要向对话一般下达命令,如“Move coke can to Taylor Swift”,机器人就能够在图片中找到对的目标并送她一罐可乐,甚至可以主动思考,完成从“选择灭绝动物”到抓取桌子上的塑料玩具恐龙这种多阶段推理的飞跃。此类任务要求机器人对数据中从未见过的物体或场景执行操作,过程中需要基于网络预训练知识的转化,也被称为模型的涌现能力。对比可知 VLA 模型 RT-2 在泛化性能显著提高,在涌现能力的成功率方面优于此前的 RT-1 和 VC1。

图表13: VLA 模型控制机器人抓取灭绝动物



资料来源: DeepMind, 中邮证券研究所

图表14: 涌现能力成功率对比



资料来源: DeepMind, 中邮证券研究所

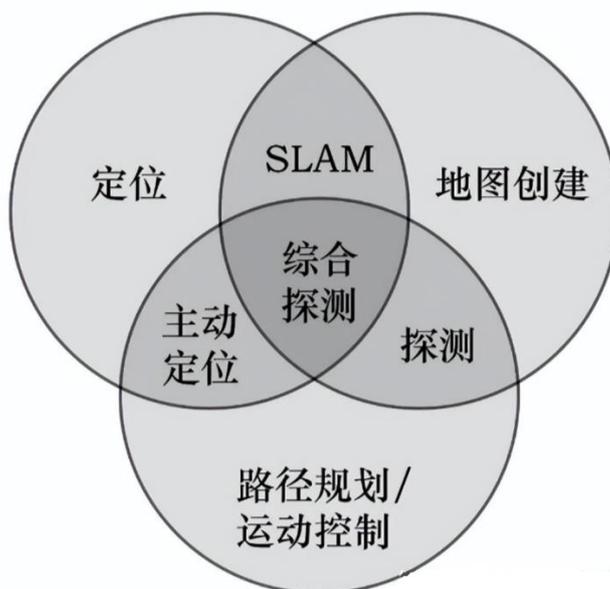
VLA 的痛点在于大模型实时性较差,导致执行成功率有待提升。未来若大模型的实时性问题能够得到解决,最终运动控制表现改善,届时一体化的 VLA 模型有望成为主流 AI 训练路径。

3 机器人如何实现自主移动: SLAM 技术的应用

SLAM 是实现机器人自主导航和后续交互的关键技术之一。SLAM 即 Simultaneous Localization and Mapping，是以定位和建图两大技术为目标的算法。目前主流的 SLAM 技术应用为激光 SLAM(基于激光雷达)和视觉 SLAM(基于单/双目摄像头)，也可以分为基于滤波 (Filter-Based) 的 SLAM 和基于图优化 (Graph-Based) 的 SLAM。

这里需要明确，SLAM 并不是自主移动导航，它不解决行动问题，只解决实时定位与地图构建问题。其核心原理是使用相机、激光雷达、惯性测量单元等传感器来收集环境信息，然后用算法将这些信息融合起来，以确定机器人在未知环境中的位置，并构建一张环境地图。机器人在路径规划时需要以地图作为参考，定位时，也需要消除里程计的累加误差所造成的干扰。通过 SLAM 技术，机器人可以在未知环境中进行探索及实时构建地图。前文中 Mobile ALOHA 的移动底座，就是一个具备较为成熟的室内 SLAM 技术的 AGV 机器人。

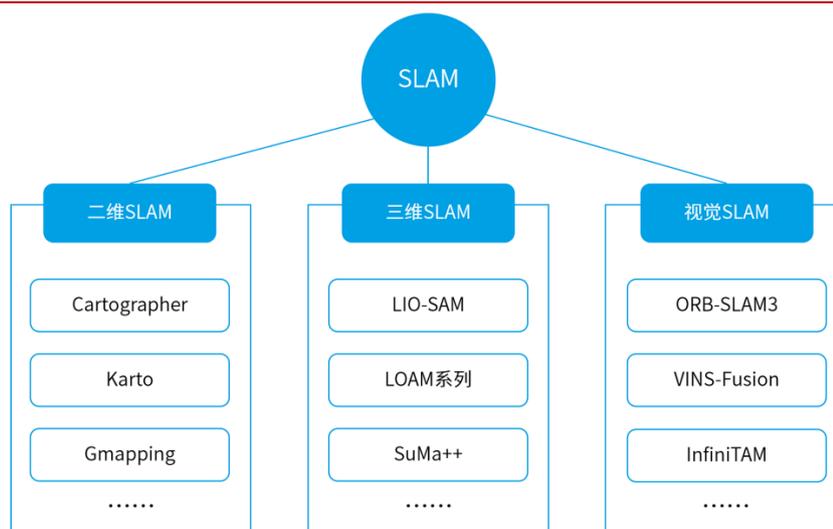
图表15: SLAM 与构成自主移动导航的各领域关系图



资料来源：思岚科技，中邮证券研究所

根据传感器的不同，机器人用的 SLAM 算法可以分为二维激光 SLAM、三维激光 SLAM 以及视觉 SLAM。其中，二维激光 SLAM 常用的有 Cartographer、Karto，三维激光 SLAM 较流行的是 LIO-SAM、LOAM 系列，视觉 SLAM 主流的方案为 ORB-SLAM3、VINS-Fusion 等。

图表16: SLAM 根据传感器不同的分类情况



资料来源：深圳市众星智能，中邮证券研究所

不同的 SLAM 算法，实现的具体细节会有所不同，一般通用架构会包含数据预处理、前端、后端、闭环检测、地图构建等。

数据预处理：传感器数据接收、多传感器数据同步、传感器数据预处理（包括去除无效值、序列检查、去除点云遮挡点与平行点等）、点云去畸变处理；

前端：特征提取、数据融合、数据关联、运动估计等，将数据同已有地图相关联，从而确定机器人的轨迹。前端的质量对于 SLAM 系统的准确性和稳定性有着关键的影响；

后端：利用前端获取的运动轨迹和地图信息形成位姿图，并通过机器人的状态、地图和传感器误差等进行计算约束，实现位姿图的优化。后端优化是对整个系统的运动轨迹和地图进行全局优化；

闭环检测：识别机器人经过的相似位置，避免累积误差的产生。闭环检测可以有效降低机器人的定位误差，提高 SLAM 算法的精度和鲁棒性；

地图构建：大致可分为语义地图、基于概率更新的占栅格地图（黑白灰三色图）、3D 点云地图（稀疏、稠密）、基于 TSDF 的占栅格地图等。

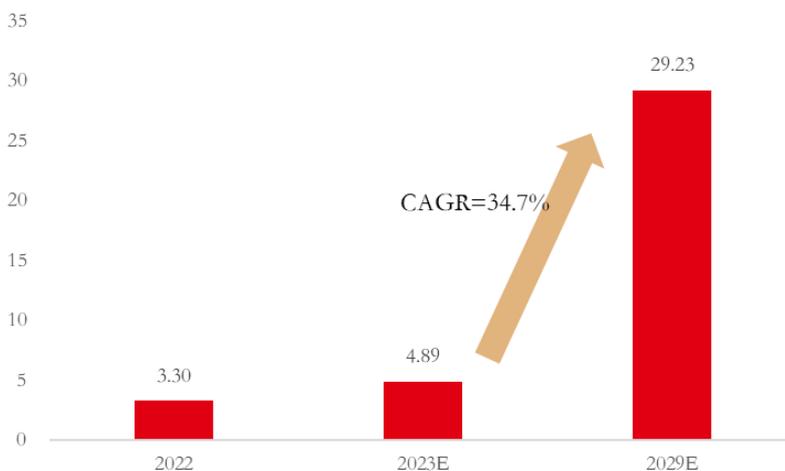
不同方案的 SLAM 都有其应用空间，三维激光、视觉 SLAM 更适合于人形机器人应用场景。二维激光 SLAM 仅适合在二维平面上运动的机器人，如扫地机器人、迎宾机器人等；三维激光 SLAM 适用于需要在三维空间中进行建模和定位的场景，如机器人在户外环境中进行探索等；视觉 SLAM 通过摄像头等视觉传感器来获取机器人与环境间的相对位置关系，并不需要其他传感器配合，更为轻量化，但对于光照变化、运动模糊等问题仍需进一步优化。针对人形机器人的应用场景，三维激光 SLAM 能够获取到的深度信息更优，精度高，受环境影响不大，能够很方便地得到物体轮廓，实现定位及避障。而视觉 SLAM 的探测更为密集，能够获取到物体的颜色、纹理等信息，还可以与人形机器人的感知识别及交互需求共用传感器。故三维激光、视觉 SLAM 更适合于人形机器人应用场景，根据实际应用场景也可能存在结合。

图表17：6种代表性 SLAM 算法的综合比较

	传感器	鲁棒性	定位精度	建图面积	建图质量	实时性能	可扩展性	硬件需求	工程实现难度
Cartographer	2D Lidar	高	高	大	高	适中	中	较高	中
Karto	2D Lidar + ROS	适中	中	中	高	高	适中	较高	低
LIO-SAM	3D Lidar	中	中	中	高	高	适中	适中	较高
LOAM	3D Lidar, 适用慢速移动场景	中	中	小	高	高	较差	较高	低
ORB-SLAM3	视觉	中	高	中	中	高	适中	较低	较高
VINS-Fusion	视觉和惯性	高	高	中	高	中	高	较高	较高

资料来源：深圳市众星智能，中邮证券研究所

根据 Market Monitor Global 调研数据显示，2022 年全球 SLAM 即时定位与地图构建技术市场规模大约为 3.30 亿美元，预计后续六年 CAGR 为 34.7%，到 2029 年行业成长至 29.23 亿美元。

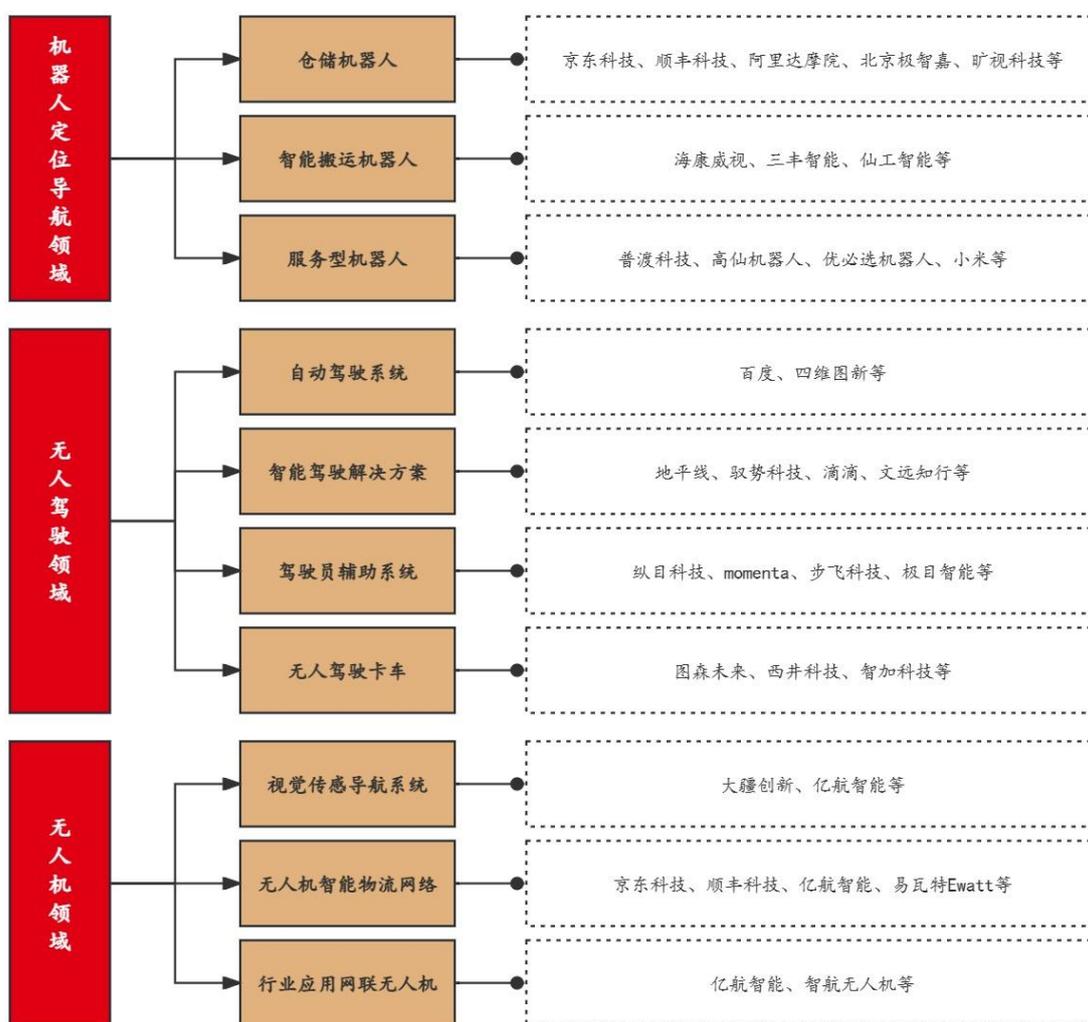
图表18：2022-2029E 全球 SLAM 市场规模（亿美元）


资料来源：Market Monitor Global，中邮证券研究所

国内多家人形机器人厂商已经在自主移动领域应用了 SLAM 技术，未来随着移动机器人、自动驾驶等领域 SLAM 技术的迁移，应用范围将更为广阔。在人形机器人赛道上，特斯拉凭借其自动驾驶领域的深厚积累，将 FSD 平移至 Optimus 上以实现自主移动。FSD V12 实现了完全端到端自动驾驶，深度依赖神经网络。V12 把摄像头获取的图像数据输入到神经网络，网络能够直接输出车辆控制指令，其过程更像是人类大脑的分析思考和决策。但除开特斯拉，国产人形机器人厂商

如开普勒、追觅、大象等均在产品中应用了 SLAM 算法。开普勒自研算法搭配 100TOPS 高算力主板，令其机器人拥有视觉识别、视觉 SLAM 等功能；追觅在智能清扫机器人上拥有丰富的多传感器感知融合处理、定位导航及地图构建的经验，算法上积累了 SLAM、ToF 激光雷达、结构光等核心技术，在其人形机器人上能够实现技术复用；大象机器人推出的水星 Mercury 采用先进的 RTABMAP 3D 与 2D 配合的 SLAM 导航技术，通过激光雷达感知平面环境，使用 3D 摄像头实时构建地图并同时定位，实现了智能导航、避障，具有高效而精准的移动能力。另外，国内许多公司在移动机器人（AGV、AMR、服务型机器人等）和汽车及无人机自动驾驶领域有着 SLAM 算法技术积累，未来人形机器人不断落地商业化应用，有望通过技术迁移使 SLAM 技术得到更为广泛的应用。

图表19：SLAM 技术的其他应用场景及国内代表公司



资料来源：中邮证券研究所

4 风险提示

人形机器人下游发展不及预期风险；
算法及算力更新迭代不及预期风险；
机器人自主移动技术竞争加剧风险；
技术落地不达预期风险。

中邮证券投资评级说明

投资评级标准	类型	评级	说明
报告中投资建议的评级标准： 报告发布日后的 6 个月内的相对市场表现，即报告发布日后的 6 个月内的公司股价（或行业指数、可转债价格）的涨跌幅相对同期相关证券市场基准指数的涨跌幅。 市场基准指数的选取：A 股市场以沪深 300 指数为基准；新三板市场以三板成指为基准；可转债市场以中信标普可转债指数为基准；香港市场以恒生指数为基准；美国市场以标普 500 或纳斯达克综合指数为基准。	股票评级	买入	预期个股相对同期基准指数涨幅在 20%以上
		增持	预期个股相对同期基准指数涨幅在 10%与 20%之间
		中性	预期个股相对同期基准指数涨幅在-10%与 10%之间
		回避	预期个股相对同期基准指数涨幅在-10%以下
	行业评级	强于大市	预期行业相对同期基准指数涨幅在 10%以上
		中性	预期行业相对同期基准指数涨幅在-10%与 10%之间
		弱于大市	预期行业相对同期基准指数涨幅在-10%以下
	可转债评级	推荐	预期可转债相对同期基准指数涨幅在 10%以上
		谨慎推荐	预期可转债相对同期基准指数涨幅在 5%与 10%之间
		中性	预期可转债相对同期基准指数涨幅在-5%与 5%之间
		回避	预期可转债相对同期基准指数涨幅在-5%以下

分析师声明

撰写此报告的分析师（一人或多人）承诺本机构、本人以及财产利害关系人与所评价或推荐的证券无利害关系。

本报告所采用的数据均来自我们认为可靠的目前已公开的信息，并通过独立判断并得出结论，力求独立、客观、公平，报告结论不受本公司其他部门和人员以及证券发行人、上市公司、基金公司、证券资产管理公司、特定客户等利益相关方的干涉和影响，特此声明。

免责声明

中邮证券有限责任公司（以下简称“中邮证券”）具备经中国证监会批准的开展证券投资咨询业务的资格。

本报告信息均来源于公开资料或者我们认为可靠的资料，我们力求但不保证这些信息的准确性和完整性。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价，中邮证券不对因使用本报告的内容而导致的损失承担任何责任。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

中邮证券可发出其它与本报告所载信息不一致或有不同结论的报告。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

中邮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头寸并进行交易，也可能为这些公司提供或者计划提供投资银行、财务顾问或者其他金融产品等相关服务。

《证券期货投资者适当性管理办法》于 2017 年 7 月 1 日起正式实施，本报告仅供中邮证券客户中的专业投资者使用，若您非中邮证券客户中的专业投资者，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司不会因接收人收到、阅读或关注本报告中的内容而视其为专业投资者。

本报告版权归中邮证券所有，未经书面许可，任何机构或个人不得存在对本报告以任何形式进行翻版、修改、节选、复制、发布，或对本报告进行改编、汇编等侵犯知识产权的行为，亦不得存在其他有损中邮证券商业性权益的任何情形。如经中邮证券授权后引用发布，需注明出处为中邮证券研究所，且不得对本报告进行有悖原意的引用、删节或修改。

中邮证券对于本申明具有最终解释权。

公司简介

中邮证券有限责任公司，2002年9月经中国证券监督管理委员会批准设立，注册资本50.6亿元人民币。中邮证券是中国邮政集团有限公司绝对控股的证券类金融子公司。

公司经营范围包括：证券经纪；证券自营；证券投资咨询；证券资产管理；融资融券；证券投资基金销售；证券承销与保荐；代理销售金融产品；与证券交易、证券投资活动有关的财务顾问。此外，公司还具有：证券经纪人业务资格；企业债券主承销资格；沪港通；深港通；利率互换；投资管理人受托管理保险资金；全国银行间同业拆借；作为主办券商在全国中小企业股份转让系统从事经纪、做市、推荐业务资格等业务资格。

公司目前已经在北京、陕西、深圳、山东、江苏、四川、江西、湖北、湖南、福建、辽宁、吉林、黑龙江、广东、浙江、贵州、新疆、河南、山西、上海、云南、内蒙古、重庆、天津、河北等地设有分支机构，全国多家分支机构正在建设中。

中邮证券紧紧依托中国邮政集团有限公司雄厚的实力，坚持诚信经营，践行普惠服务，为社会大众提供全方位专业化的证券投、融资服务，帮助客户实现价值增长，努力成为客户认同、社会尊重、股东满意、员工自豪的优秀企业。

中邮证券研究所

北京

邮箱：yanjiusuo@cnpsec.com

地址：北京市东城区前门街道珠市口东大街17号

邮编：100050

上海

邮箱：yanjiusuo@cnpsec.com

地址：上海市虹口区东大名路1080号邮储银行大厦3楼

邮编：200000

深圳

邮箱：yanjiusuo@cnpsec.com

地址：深圳市福田区滨河大道9023号国通大厦二楼

邮编：518048