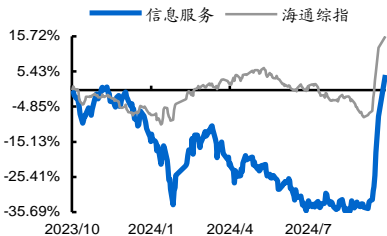


投资评级 优于大市 维持

市场表现



资料来源：海通证券研究所

相关研究

《财政加码，计算机顺周期领域迎双击》  
2024.10.08《计算机行业 2024 年 10 月研究观点：  
利好政策背景下，建议关注金融 IT 及顺  
周期方向》2024.10.07《OpenAI o1 初探：或能成为引领 AI  
Phenomenal Ride 的 LLM 新范式》  
2024.09.25

分析师:杨林

Tel:(021)23183969

Email:yl11036@haitong.com

证书:S0850517080008

分析师:杨蒙

Tel:(021)23185700

Email:ym13254@haitong.com

证书:S0850523090001

联系人:朱瑶

Tel:(021)23187261

Email:zy15988@haitong.com

联系人:杨昊翊

Tel:(021)23185620

Email:yhy15080@haitong.com

## EAI (具身智能): 驱动通用人工智能与机器人产业的关键技术

投资要点:

- EAI 构建新概念，相关政策推动技术发展：**具身智能（Embodied AI, EAI）由艾伦·图灵于 1950 年提出，通过“具身图灵测试”验证智能体是否能处理物理世界的复杂性。具身智能体不同于仅在网络空间中运作的非实体人工智能，它们结合了多模态大模型（MLMs）和世界模型（WMs），具备强大的感知、交互和规划能力，能够在虚拟和物理环境中主动适应并执行任务。相比传统的预编程系统，具身智能体更依赖于构建世界模型和想象力，以实现复杂的推理和决策。目前，上海和北京等地陆续出台政策，推动智能机器人和制造业的高质量发展，上海目标到 2025 年实现行业标杆企业的建立。海外方面，美国在机器人工业应用和商业化方面的进展更为成熟，中国仍在探索阶段。
- 从模拟、感知、交互三方面训练具身智能：**模拟器在训练具身智能中扮演了重要角色，通过提供逼真的虚拟环境进行算法开发和模型训练，帮助研究人员在降低成本、提高安全性和加速迭代的同时，将研究成果更快地转化为现实应用。另外，具身感知让智能体理解物理世界中的视觉推理和空间，技术包括视觉同步定位与绘图（vSLAM）和 3D 视觉定位，帮助智能体在动态环境中移动和互动。具身交互则强调智能体在物理或模拟空间中与环境 and 人类的互动能力，典型任务如具身问题解答（EQA），需要智能体主动探索环境，整合信息并执行目标导向的动作。
- 具身智能产品多样，覆盖广阔市场：**不同类型的机器人在各自的应用领域中展现出重要的功能和价值。1) 固定式机器人因其高精度和稳定性，广泛应用于实验室自动化、教育和工业制造。2) 轮式机器人在物流、仓储和安全检查中表现优异，而履带式机器人适用于农业、建筑和军事等复杂地形。3) 四足机器人因其稳定性和适应性，广泛用于复杂地形探索、救援任务和军事行动。4) 人形机器人在服务业、医疗保健和协作环境中日益普及，应用于智能交互、精密制造和医疗手术等领域。各类机器人在多种应用场景中展现独特优势，同时也面临不同的技术挑战。
- 投资建议：**我们认为具身智能（EAI）是推动通用人工智能（AGI）发展的关键技术，具有广阔的应用前景。其技术涵盖计算机视觉、自然语言处理和机器人技术，尤其在具身感知和交互方面表现突出。基于多模态大模型和世界模型的具身智能体，不仅能够理解并适应复杂的物理和虚拟环境，还能在动态场景中与人类自然互动，执行复杂任务。根据证券时报援引赛迪顾问发布的《2024 中国人形机器人产业生态发展研究》显示，全球具身智能市场将在未来几年持续增长，尤其是在工业自动化和服务机器人领域。此外，随着智能制造和人形机器人技术的快速发展，预计到 2026 年，中国人形机器人市场规模将突破 200 亿元。各地政策的积极推动，如上海和北京的智能机器人发展计划，将进一步加速具身智能的技术进步和市场应用。
- 建议关注：**AI 算力、模型和应用。**算力：**英伟达、寒武纪-U、海光信息、景嘉微、龙芯中科、浪潮信息、中科曙光、神州数码、软通动力、中国长城；**模型：**科大讯飞、中控技术、海康威视、大华股份、商汤-W；**应用：**微软、金山办公、赛意信息、宝信软件、万兴科技、虹软科技、新致软件、新国都、中科创达、索辰科技、比亚迪电子。
- 风险提示：**1.大模型发展不及预期；2. EAI 发展不及预期；3.具身智能下游需求不足。

## 目 录

1.	发展路线：EAI 构建新概念，相关政策推动技术发展.....	6
2.	技术背景：从模拟、感知、交互三方面训练 EAI .....	7
2.1	EAI 概念解析，虚拟与物理环境的结合 .....	7
2.2	具身模拟器（Embodied Simulator） .....	8
2.2.1	通用模拟器（General Simulator） .....	8
2.2.1	基于真实世界的模拟器（Real-Scene Based Simulators） .....	10
2.3	具身感知（Embodied Preception） .....	12
2.3.1	视觉同步定位和绘图（vSLAM） .....	12
2.3.2	3D 视觉定位 .....	13
2.3.3	视觉语言导航（Visual Language Navigation） .....	14
2.3	具身交互（Embodied Intereaction） .....	15
2.4	具身智能全面落地仍需解决四大难题 .....	15
3.	应用场景：具身智能产品多样，覆盖广阔市场 .....	16
3.1	固定基座机器人：全面赋能实验室与工业场景 .....	16
3.2	轮式/履带式机器人：高机动性适应复杂道路环境 .....	18
3.3	四足机器人：龙头制造商占据大量市场份额 .....	20
3.4	人形机器人：未来拥有强大潜力，技术仍需探索 .....	20
4.	潜在标的：美国商业化更为成熟，中国仍需探索 .....	22
4.1	Figure AI：获巨头投资，技术不断成熟 .....	22
4.2	特斯拉 Optimus：优先赋能特斯拉工厂 .....	23
4.3	宇树科技：技术领先，覆盖场景多元 .....	24
4.4	中科创达：布局端侧智能+机器人 .....	25
4.5	有鹿机器人：引入“通用智能大脑”概念 .....	26
4.6	科大讯飞：讯飞超脑计划，让机器人走向通才 .....	27
4.7	海康威视：视觉与移动机器人提供商 .....	29
4.8	比亚迪电子：AMR 提供物流解决方案 .....	29
5.	投资建议 .....	30
6.	风险提示 .....	30

## 图目录

图 1	中美机器人企业落地进度一览 .....	7
图 2	基于 MLM 和 WM 的具身智能框架.....	8
图 3	通用模拟器的例子.....	9
图 4	Isaac Sim 架构 .....	9
图 5	Isaac Sim 工作界面 .....	10
图 6	Isaac 模拟机械手臂 .....	10
图 7	Isaac 模拟无人机飞行.....	10
图 8	基于真实世界的模拟器实例 .....	11
图 9	ThreeDWorld (TDW) 设计展示 .....	11
图 10	多智能体互动和 VR 能力 .....	12
图 11	vSLAM 架构展示 .....	13
图 12	3D 视觉定位中的分级定位.....	13
图 13	共视聚类概念展示.....	14
图 14	NaVid 架构图 .....	14
图 15	EQA 任务例子 .....	15
图 16	ROMAN 框架的功能 .....	17
图 17	ROMAN 从错误中恢复的效果展示 .....	17
图 18	2013 年-2023 年亚马逊机器人应用数量 .....	18
图 19	KIVA 机器人构造.....	19
图 20	2018 年-2022 年中国移动机器人市场规模 .....	19
图 21	2022 年中国移动机器人市场规模分布情况.....	19
图 22	四足机器人发展路线.....	20
图 23	人形机器人产业各大关联厂商 .....	21
图 24	人形机器人应用展望 .....	21
图 25	2024-2035 年人形机器人市场规模预测.....	22
图 26	Figure AI 第一代与第二代机器人 .....	23
图 27	语音模块的工作模式解析.....	23
图 28	特斯拉 Optimus 自主工作.....	24
图 29	宇树科技产品一览.....	24
图 30	CES2024 宇树科技产品展示 .....	25

图 31	中科创达机器人产品 .....	26
图 32	有鹿机器人具身智能大模型 .....	27
图 33	有鹿机器人打造“通用大脑”概念 .....	27
图 34	科大讯飞机器人平台架构 .....	28
图 35	科大讯飞超脑计划 2030 .....	28
图 36	海康威视移动机器人产品一览 .....	29
图 37	比亚迪电子 AMR 机器人 .....	30

## 表目录

---

表 1	实体人工智能和非实体人工智能.....	6
表 2	政策推动人工智能技术发展.....	6

## 1. 发展路线：EAI 构建新概念，相关政策推动技术发展

具身智能 (Embodied AI) 最初是由艾伦-图灵 (Alan Turing) 于 1950 年提出的“具身图灵测试” (Embodied Turing Test)，旨在确定智能体是否能够展现出解决虚拟环境中问题的能力，而且能够驾驭物理世界的复杂性和不可预测性。网络空间中的智能体通常被称为非实体人工智能，而物理空间中的智能体则是实体人工智能。多模态大模型 (MLMs) 的最新进展为具身模型注入了强大的感知、交互和规划能力，从而开发出能与虚拟和物理环境积极交互的通用具身智能体和机器人。因此，具身智能体被广泛认为是 MLMs 的最佳载体，目前最有代表性的具身模型是 RT-2 和 RT-H。

表 1 实体人工智能和非实体人工智能

智能体种类	适应环境	物理实体	描述	代表性的智能体
非实体	网络空间	无	认知与物理实体相分离	ChatGPT, RoboGPT
实体	物理空间	机器人、汽车、其他设备	认知融入物理实体	RT-1, RT-2, RT-H

资料来源：Yang Liu 《Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI》，海通证券研究所

要让 AI 像人类一样理解这个物理世界，它必须能够以人类的方式解释和理解场景。比如，当 AI 被放置在一个房间里时，它需要能够像人类那样分析和解读周围的环境。另外，在不同领域之间建立联系，或者试图发现新知识时，传统的预编程和特定领域的专业系统已经无法满足需求。这些系统受到现有内置知识的限制，很难实现新的发现、创新和创造。让 AI 变得更聪明的关键在于利用“想象力”，其实就是人类和其他动物依靠世界的现有模式生成的想法，它是一个非常强大的规划工具。为了让 AI 有效地规划，它需要构建一个关于世界的模型 (WMS)，并能够利用这个模型进行推理和决策。因此，具身认知至关重要。系统需要通过具身认知来获取知识，并进一步生成抽象的认知。

表 2 政策推动人工智能技术发展

地区	文件名称	发布时间	具体内容
上海	《上海市智能机器人标杆企业与应用场景推荐目录》	2023 年 3 月	各区产业主管部门支持推动以机器人为代表的智能终端产业发展，培育一流营商环境。力争到 2025 年，上海市将打造 10 家行业一流的机器人头部品牌、100 个标杆示范的机器人应用场景、1000 亿元机器人关联产业规模。
上海	《上海市推动制造业高质量发展三年行动计划（2023-2025 年）》	2023 年 6 月	瞄准人工智能技术前沿，构建通用大模型，面向垂直领域发展产业生态，建设国际算法创新基地，加快人形机器人创新发展。
北京	《北京市机器人产业创新发展行动方案 2023-2025 年》	2023 年 6 月	加紧布局人形机器人，对标国际领先人形机器人产品，支持企业和高校院所开展人形机器人整机产品、关键零部件攻关和工程化，加快建设北京市人形机器人产业创新中心。以人形机器人小批量生产和应用为目标，打造通用智能底层软件及接口、通用硬件开发配套设施等基础条件，集中突破人形机器人通用原型机和通用人工智能大模型等关键技术。
北京	《北京市促进机器人产业创新发展的若干措施》	2023 年 8 月	由机器人骨干企业牵头，整合国内外一流创新资源，组建人形机器人创新中心，开展关键共性技术研究。支持机器人企业与“智能机器人与系统高精尖创新中心”联合开展产业化攻关。

资料来源：上海市经济和信息化委员会，上海市人民政府，北京市人民政府，海通证券研究所

相关政策已落地，带动具身智能行业发展。例如上海市的政策重点是推动智能机器人和智能制造业的发展，目标是通过营商环境的优化和创新基地的建设，到 2025 年实现行业标杆企业与应用场景的建立。北京市的政策则侧重于机器人产业的创新发展，特



别是对高端机器人产品和国际化布局的支持,旨在推动产业生态系统的完善和技术创新。

图1 中美机器人企业落地进度一览

企业	产品型号	开始进厂时间	合作厂商	工作内容	落地进度
特斯拉	Optimus Gen2	2024.05	特斯拉	分拣电池	目前有2台Optimus人形机器人在工厂训练,预计明年超过1000台
Figure AI	Figure 01	2024.07	宝马	简单抓取	适配过程会持续12-24个月左右
优必选	Walker S	2024.01	蔚来、东风柳汽、一汽-大众青岛分公司	门锁质检、车灯盖板检测、安全带检测、贴车标、螺栓拧紧、零件安装、零件转运等	进入多家车厂实训,2024年底小规模交付
智元机器人	远征A1	暂无	均普智能、临港集团	/	/
宇树科技	H1	暂无	/	/	/
银河通用机器人	Galbot	暂无	/	/	/
达闼机器人	XR4	暂无	/	/	/
乐聚机器人	夸父	暂无	蔚来	计划在工厂检测验证	/
Appttronik	Apollo	2024.03	奔驰、GXO	汽车搬运、装配、物流配送等	奔驰在位于匈牙利的一家工厂试用数量不详的Apollo机器人
小米机器人	CyberOne	暂无	小米汽车	计划融入小米制造、智能制造多个场景	/
波士顿动力	Atlas	暂无	现代汽车	计划在现代汽车的制造产线上应用测试	/
星动纪元	XBot-L	暂无	/	/	/
戴盟机器人	Sparky 1	暂无	/	焊接电路板	/
Agility Robotics	Digit	2023.04	亚马逊	物流	/
开普勒	先行者系列	暂无	/	/	/
Sanctuary AI	Phoenix	2024.04	麦格纳	计划将产品用于部署在麦格纳的制造业务中	/

资料来源: 硬氦分析, 锦囊专家, 海通证券研究所

美国在机器人工业应用和商业化方面的进展更为成熟, 中国仍在探索阶段。美国和中国的机器人技术进展和落地进度存在一些显著差异。美国的机器人企业, 如特斯拉和Figure AI, 已在2024年中旬进入工厂, 负责分拣电池和简单抓取的任务。Agility Robotics与亚马逊的合作已经在2023年展开, 推进了物流和自动化领域的实际应用。Sanctuary AI也计划在2024年开始商业运营, 重点放在智慧城市和建筑领域的智能服务上。相比之下, 中国的机器人企业虽然在多个领域(如安防、教育和娱乐)都有布局, 但整体落地进度稍慢。例如, 优必选的Walker S预计在2024年初开始量产, 主要用于门锁质检和汽车制造领域。其他企业如小米机器人和波士顿动力的项目仍在研发阶段, 逐步优化视觉和环境交互技术。

## 2. 技术背景: 从模拟、感知、交互三方面训练 EAI

实现通用人工智能 (AGI) 的关键基础在于具身智能的发展。具身智能体与仅限虚拟对话的智能体 (如 ChatGPT) 不同, 它们可以通过控制物理实体在现实和模拟环境中进行交互。该技术涵盖了多个领域, 包括计算机视觉、自然语言处理和机器人技术, 特别是在具身感知、具身交互以及从模拟到现实的机器人控制方面展现了显著优势。具身智能体依托于多模态大模型 (MLMs) 和世界模型 (WMs), 像“脑”一样理解虚拟与物理环境, 主动感知多模态元素, 并根据人类的意图进行任务分解与执行。它们不仅能够与人类互动, 还能够借助知识库和工具完成复杂任务, 展现出比传统深度强化学习更高的灵活性和通用性。

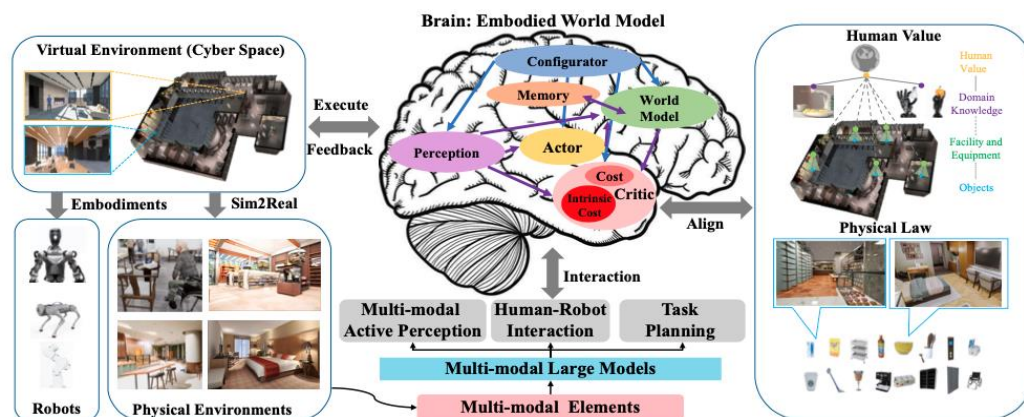
模拟器在训练具身智能中扮演了重要角色, 通过提供逼真的虚拟环境进行算法开发和模型训练, 帮助研究人员在降低成本、提高安全性和加速迭代的同时, 将研究成果更快地转化为现实应用。同时, 具身感知让智能体理解物理世界中的视觉推理和空间, 技术包括视觉同步定位与绘图 (vSLAM) 和 3D 视觉定位, 帮助智能体在动态环境中移动和互动。具身交互则强调智能体在物理或模拟空间中与环境 and 人类的互动能力, 典型任务如具身问题解答 (EQA), 需要智能体主动探索环境, 整合信息并执行目标导向的动作。

### 2.1 EAI 概念解析, 虚拟与物理环境的结合

具身智能的发展是实现通用人工智能 (AGI) 的重要基础。与类似 ChatGPT 的对话智能体不同, 具身智能可以通过控制物理实体并在模拟和物理环境中进行交互来实现。

目前，具身智能涵盖了计算机视觉（CV）、自然语言处理（NLP）和机器人技术等多个关键技术领域，最具代表性的是具身感知、具身交互、具身智能体和模拟到现实的机器人控制。

图2 基于 MLM 和 WM 的具身智能框架



资料来源：Yang Liu 《Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI》，海通证券研究所

基于多模态大模型（MLMs）和世界模型（WMs），具身智能体能够用“脑”来理解虚拟和物理环境，并主动感知多模态元素。主要的内容包括：1）**具身世界模型（Embodied World Model）**：这是具身智能体的核心，类似于智能体的大脑。它能够理解虚拟与物理环境，并进行多模态感知。2）**多模态感知**：智能体能够全面感知来自虚拟和物理环境的多模态元素，这意味着它可以处理不同来源的数据，如视觉、触觉等。3）**人类意图理解与对齐**：具身智能体能够充分理解人类通过语言表达的意图，并与人类的价值观对齐，确保其行为符合人类的期望。4）**任务分解与执行**：智能体具备将复杂任务分解为子任务的能力，并且能够精确执行这些任务。此外，它还能够与人类进行互动，并利用知识库和工具来完成任务。通过这些特性，具身智能体可以在复杂的动态环境中与人类自然互动，并可靠地完成任

## 2.2 具身模拟器（Embodied Simulator）

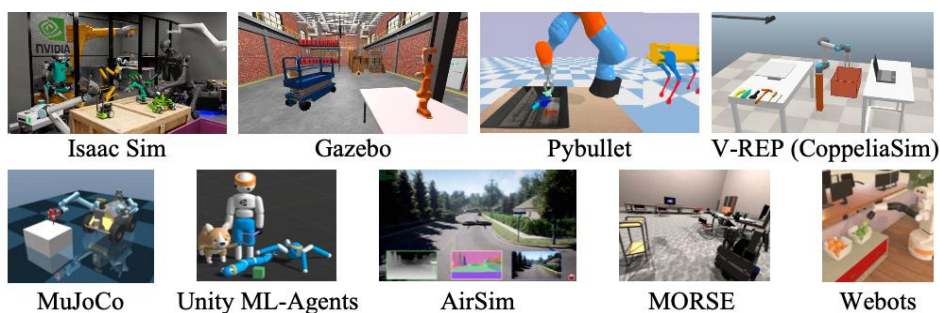
模拟器显著提升了 AI 训练的效率，并节省了大量成本。数据匮乏一直是具身人工智能研究面临的挑战，收集真实世界的机器人数据需要花费大量时间和成本。首先，现实世界中的机器人训练需要搭建专门的物理场所，导致训练进展缓慢，效率难以提升。另外，搭建专属场地、频繁的数据收集、聘请机器人专家操作等涉及的成本很高。此外，最重要的挑战在于可重复性，因为机器人的硬件配置、控制方法和实施框架存在巨大差异，阻碍了数据的复用性。在这种情况下，模拟器为具身人工智能的数据收集和训练提供了一种全新的解决方案。

具身模拟器对于 EAI 技术至关重要，因为它们能提供一个经济有效、可扩展且安全的实验平台。通过模拟潜在的危

### 2.2.1 通用模拟器（General Simulator）



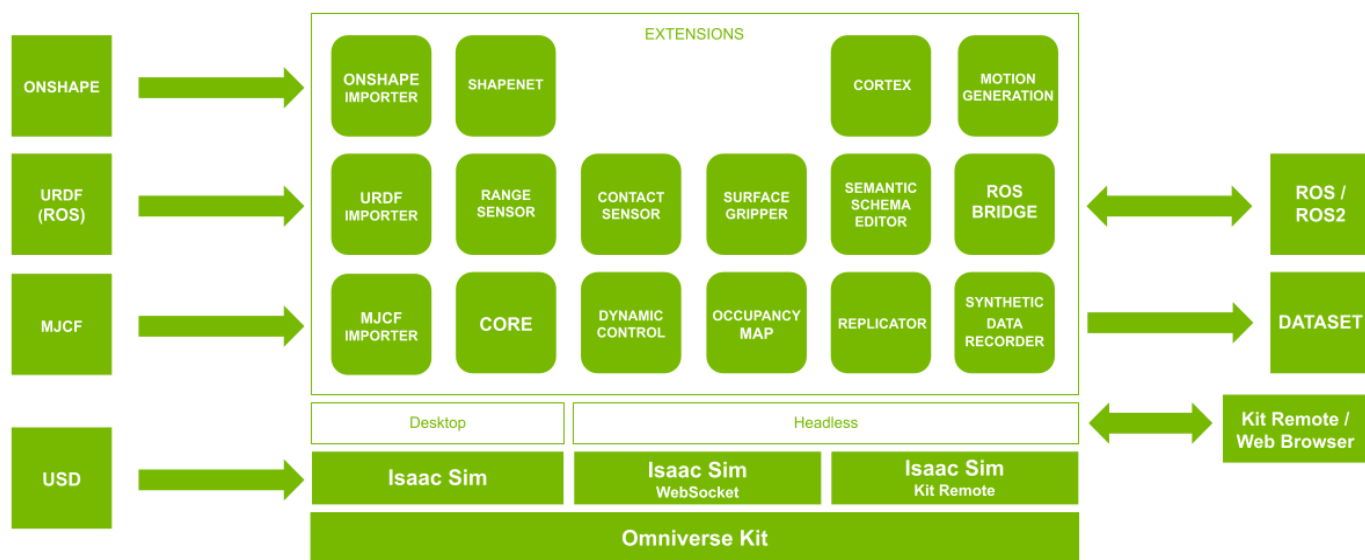
图3 通用模拟器的例子



资料来源：Yang Liu 《Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI》，海通证券研究所

具身人工智能的最终目标是将虚拟环境中的研究成果转化为现实世界中的应用。研究人员可以选择最适合自己需要的模拟器来辅助研究。通用模拟器提供了一个近似物理世界的虚拟环境，可以进行算法开发和模型训练，在成本、时间和安全性方面都有显著优势。

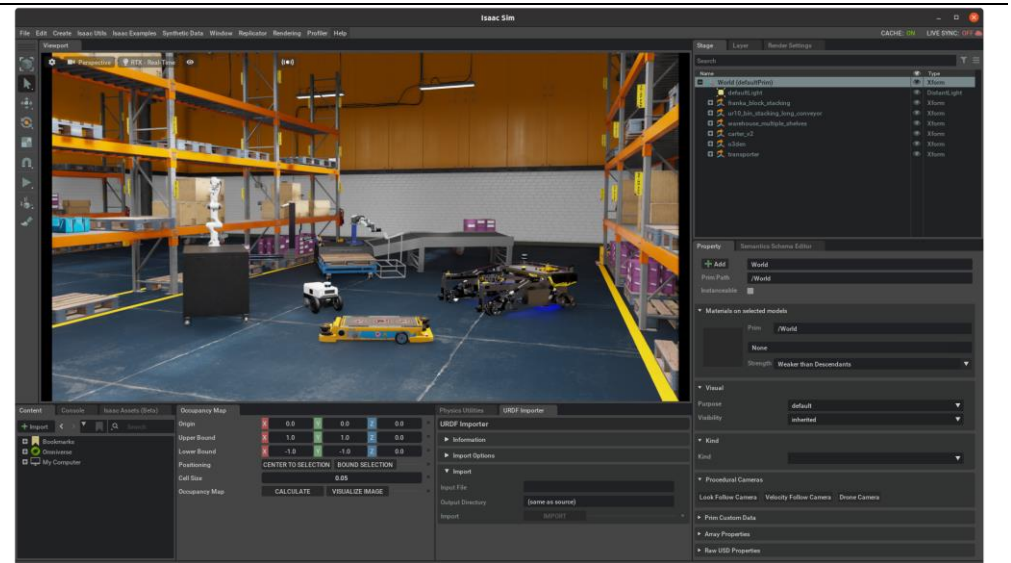
图4 Isaac Sim 架构



资料来源：Isaac Sim Documentation, CSDN, 海通证券研究所

Isaac Sim 是一个专为 NVIDIA Omniverse 平台开发的机器人仿真工具包,它提供了构建仿真机器人世界和进行实验所需的大部分功能。Isaac Sim 可以接受来自不同来源的输入,比如 Onshape、URDF、MJCF、USD,其中 USD 直接导入 Isaac Sim,其他类型的输入则会通过对应的 importer 插件进行导入。Onshape 是一种基于云的计算机辅助设计 (CAD) 软件,用于进行三维建模和设计工作。类似于 Fusion360。URDF (Unified Robot Description Format) 是一种 XML 文件格式,用于描述机器人模型的几何形状、连接性、关节、传感器和其他相关信息。在这个架构中,USD (Universal Scene Description) 用作场景描述,用于在不同工具之间进行内容创建和交换。目前 USD 正在广泛应用,不仅在视觉效果社区,还在建筑、设计、机器人技术、制造和其他领域中得到采用。

图5 Isaac Sim 工作界面



资料来源：CSDN，海通证券研究所

该工具包还提供了创建稳健、物理精确的仿真和合成数据集所需的工具和工作流程。Isaac Sim 支持常见的机器人框架，如 ROS/ROS2，允许用户通过这些框架进行导航和操作应用。此外，Isaac Sim 能够模拟来自多种传感器的数据，包括 RGB-D、激光雷达和 IMU，适用于各种计算机视觉技术，如域随机化、地面真值标注、分割和边界框的生成。

图6 Isaac 模拟机械手臂



资料来源：IsaacGymEnvs，CSDN，海通证券研究所

图7 Isaac 模拟无人机飞行



资料来源：IsaacGymEnvs，CSDN，海通证券研究所

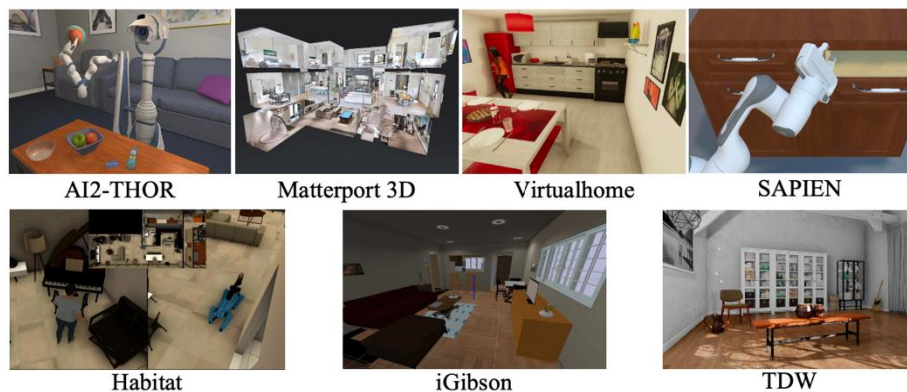
机器人仿真是利用计算机技术来模拟机器人运动、感知和互动的过程。这一过程涵盖了机器人硬件和软件系统的模拟，以便在虚拟环境中对机器人的算法和控制策略进行测试、开发和验证。其中的好处包括 **1) 成本控制**：仿真环境能显著降低机器人开发与测试成本，避免了对昂贵硬件和设备的依赖。若不使用仿真，而采用大量真实机器人进行测试，将面临硬件组装、调试及损坏等高昂的时间与经济成本。尤其对于特殊测试场景，如沙漠或核电站，搭建逼真测试环境的成本同样极高。**2) 安全性保障**：例如在工业机械臂、无人机等机器人设备调试中，无人机失控坠毁或机械臂故障会对企业的人员安全构成威胁。**3) 快速迭代**：仿真环境避免了对真实机器人的繁琐调试，例如为一千台机器人重新烧录固件或修改搭载的算法，从而节省了大量的调试时间。并且模拟器允许开发人员快速迭代机器人的算法和控制器，以优化性能和功能。

## 2.2.2 基于真实世界的模拟器（Real-Scene Based Simulators）

在室内活动中实现通用具身智能一直是 AI 研究领域的重点。这些具身智能体需要深入理解人类的日常生活，并执行复杂的具身任务，如室内环境中的导航和交互。为了满

足这些复杂任务的需求，模拟环境需要尽可能接近真实世界，这就对模拟器的复杂性和逼真度提出了很高的要求。因此，基于真实世界环境的模拟器应运而生。这些模拟器大多从现实世界收集数据，创建逼真的三维资产，并使用 UE5（虚幻 5）和 Unity 等三维游戏引擎构建场景。丰富而逼真的场景使基于真实世界环境的模拟器成为研究家居活动中的体现式人工智能的首选。

图8 基于真实世界的模拟器实例

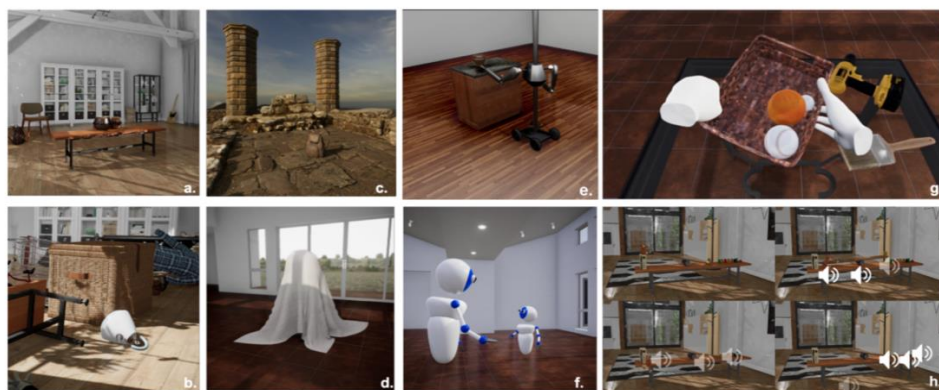


资料来源：Yang Liu 《Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI》，海通证券研究所

在腾讯网援引映维网的文章中指出，2021 年，麻省理工学院（MIT）、MIT-IBM 沃森人工智能实验室、哈佛大学和斯坦福大学的研究人员开发了一个名为 ThreeDWorld（TDW）的平台，并希望创造一个类似于《黑客帝国》的丰富虚拟世界。TDW 能够模拟室内和室外的高保真音频和视频环境，并允许用户像在现实生活中一样根据物理定律与对象进行交互。当发生相互作用时，系统能够计算并执行流体、柔体和刚体的对象方向、物理特征和速度，从而产生精确的碰撞和撞击声音。

TDW 支持在三维环境中模拟移动智能体和对象之间的高保真感觉数据和物理交互。独特的特性包括：实时接近照片真实感的图像渲染；各种物质类型的真实物理交互作用，包括布、液体和可变形物体；具身智能体的可定制“智能体”；并支持人类与 VR 设备的交互。TDW 的 API 允许多个智能体在模拟中交互，并返回代表世界状态的传感器和物理数据范围。Yang Liu 等人介绍了 TDW 在计算机视觉、机器学习和认知科学等新兴研究方向上的初步实验，包括多模态物理场景理解、物理动力学预测、多智能体交互、“像孩子一样学习”的模型，以及人类和神经网络的注意力研究。

图9 ThreeDWorld（TDW）设计展示



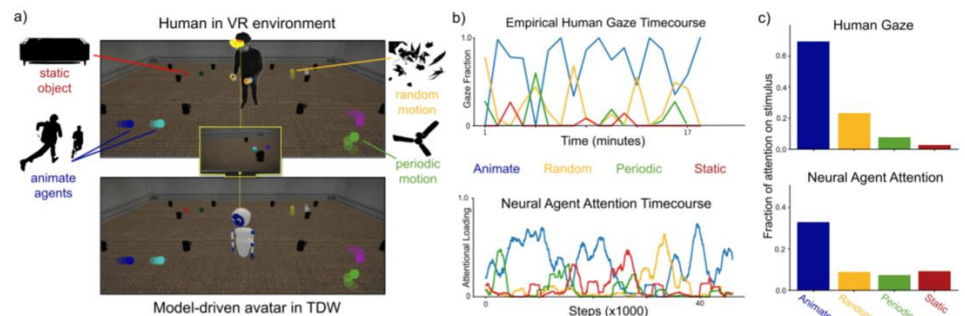
资料来源：Yang Liu 《ThreeDWorld: A Platform for Interactive Multi-Modal Physical Simulation》，CSDN，海通证券研究所

利用 TDW 的多智能体 API 的灵活性，可以创建各种多智能体交互设置的实现。其中包括一个“观察者”智能体被安置在一个有多个无生命物体的房间里，与几个不同控制的“行动者”智能体一起（图 9a）。“行动者”智能体由硬编码或交互策略控制，实现



对象操作、追逐和隐藏以及运动模仿等行为。在这种情况下，人类观察者只被要求看他们想看的任何东西，而虚拟观察者寻求最大限度地提高其预测同一显示中参与者行为的能力，根据“进展好奇心”的度量来分配其注意力，该度量寻求估计哪些观察最有可能增加观察者做出参与者预测的能力。

图10 多智能体互动和 VR 能力



资料来源：CSDN，海通证券研究所

## 2.3 具身感知（Embodied Preception）

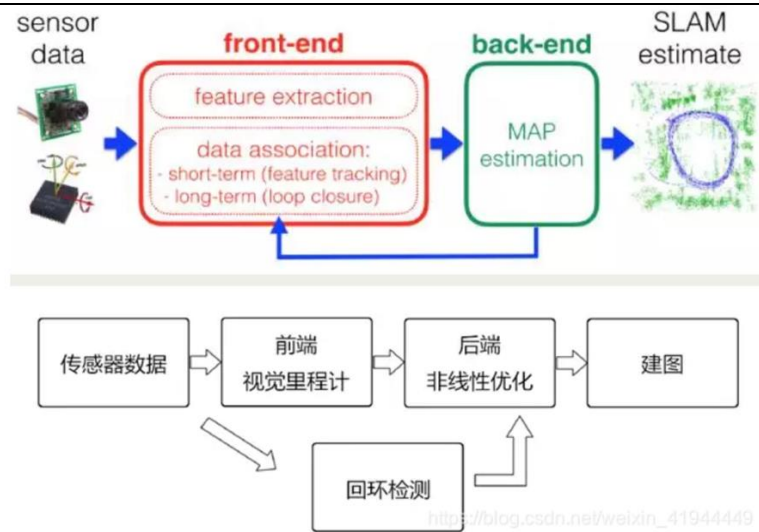
具身感知未来主要的发展方向是以智能体为中心的视觉推理。与仅仅识别图像中的物体不同，具有具身感知能力的智能体必须在物理世界中移动并与环境互动。这就要求对三维空间和动态环境有更深入的了解。

### 2.3.1 视觉同步定位和绘图（vSLAM）

SLAM (Simultaneous Localization And Mapping, 同步定位与地图构建)，主要为了解决移动机器人在未知环境运行时定位导航与地图构建的问题。SLAM 能够解决机器人在陌生环境中的定位、环境感知、移动方向等问题。机器人可以配置多种传感器来实现 SLAM，包括激光雷达（3D，2D），毫米波雷达，超声波，RGB-D，摄像头（单目，多目）等，通常根据使用场景、制造成本、设备功率、算力的需求与约束，机器人采用不同传感器或组合的解决方案，以减少误差并提高准确性。目前两个主流的解决方案是基于激光雷达的 Lidar SLAM 以及基于摄像头的 Visual SLAM。

VSLAM 即 Visual Simultaneous Localization and Mapping，主要是指如何用相机解决定位和建图问题。当用相机作为传感器时，通过一张张连续运动的图像(它们形成一段视频)，从中推断相机的运动，以及周围环境的情况。VSLAM 的技术框架主要由 5 部分组成，包括传感器数据预处理、前端、后端、回环检测、建图。前端，又称为视觉里程计（visual odometry，简称 VO），主要是研究如何根据相邻帧图像定量估算帧间相机的运动。通过把相邻帧的运动轨迹串起来，就构成相机载体（如机器人）的运动轨迹，解决定位的问题，然后根据估算的每个时刻相机的位置，计算出各像素的空间点的位置，就得到地图。

图11 vSLAM 架构展示

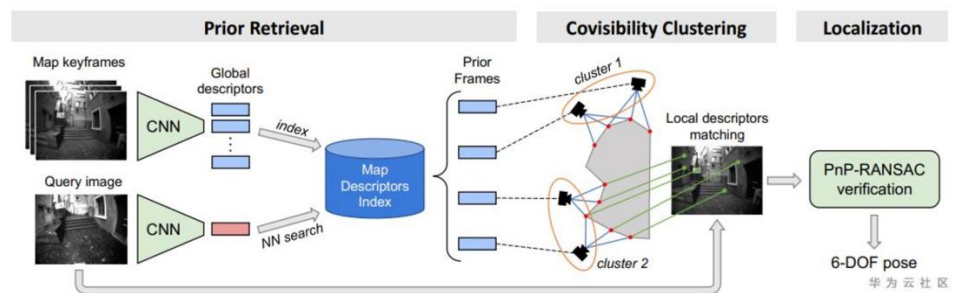


资料来源：CSDN，海通证券研究所

### 2.3.2 3D 视觉定位

3D 视觉定位的核心任务在于确定当前图像中相机的位置和姿态。为实现这一目标，最直接的方法是构建三维空间点与二维图像点之间的对应关系，并基于这些匹配点对来估算相机的位姿。这一过程被专门定义为 PnP（Perspective-n-Point）问题。求解 PnP 问题的方法有很多，常见的有 P3P、EPnP、UPnP 等。视觉定位需要解决的一大关键问题是如何建立 3D 点与 2D 点之间的匹配关系。对于这一点，在论文《From Coarse to Fine: Robust Hierarchical Localization at Large Scale》中，作者 Sarlin 提出过一种分级定位的方案。

图12 3D 视觉定位中的分级定位



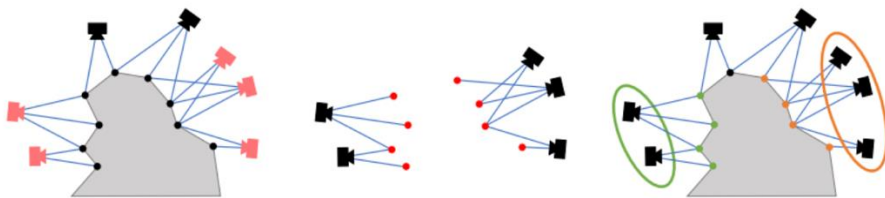
资料来源：嘉励官网，Sarlin 《From Coarse to Fine: Robust Hierarchical Localization at Large Scale》，海通证券研究所

分级定位的框架大约可以分成三步：预检索、共视聚类、局部匹配与定位。**预检索**：预检索的意义在于获取前 k 张与当前图像最相似的图像，判断相似的依据通常是通过匹配图像的全局特征。一般而言，产生全局特征的方法可以依赖于局部特征所组成的词袋，不过近些年，一些深度学习方案也被引入了进来，例如 NetVLAD 或更加轻量级的 MobileNetVLAD。最终通过获取当前图像的全局特征的 k 个最近邻来获取预检索得到的相近图集。**共视聚类**：然而由于可能产生的错误匹配，所获取到的预检索图集并不一定全部都面向同一场景，这时就需要先将面向不同场景的图像区分开来，这项技术就被称作共视聚类，简而言之就是将具有共视关系的图像聚成一类。**局部匹配与定位**：一般认为图像数量较多的类所对应的场景是正确场景的可能性较大。因此从这一场景开始，尝试获取相机位姿。获取的方式主要依赖求解 PnP 问题，因此需要首先构建当前图像的 2D 关键点在 3D 模型中的坐标位置。在尚不知道相机姿态前，这一信息的获取需要首先



匹配当前图像和场景内的图像，特别要匹配那些能够对应到 3D 位置的 2D 特征点，若能够匹配上则确定了当前图像中的 2D 点和 3D 点的对应关系，继而即可通过对 PnP 问题的求解获取相机位姿。

图13 共视聚类概念展示



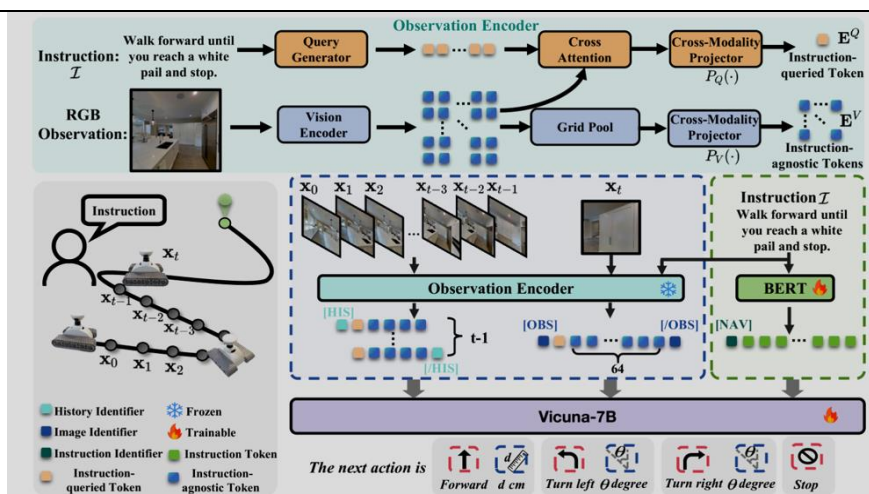
资料来源：嘉励官网，Sarlin 《From Coarse to Fine: Robust Hierarchical Localization at Large Scale》，海通证券研究所

### 2.3.3 视觉语言导航（Visual Language Navigation）

视觉语言导航（Vision Language Navigation VLN）是一种技术，它结合了计算机视觉、自然语言处理和自主学习三大核心技术，使智能体能够跟随自然语言指令进行导航。这种技术不仅理解指令，还能理解指令与视角中可以看见的图像信息，然后在环境中对自身所处状态进行调整和修复，最终做出对应的动作，以达到目标位置。例如，如果把一个机器人放置在一间卧室里，并给它一个指令去取放在另一间卧室里的足球，由于两个房间之间无法直接看见，机器人需要走出房间，经过走廊、客厅等场所，通过理解并处理看到的每一处环境信息，判断下一步的行进方向，最终找到足球。

NaVid 是首个专为视觉语言导航（VLN）任务设计的基于视频的视觉语言大模型。此模型模仿人类导航策略，仅将视频信息作为输入，无需地图、里程计或深度数据的支持。在视觉语言导航任务中，模型需要借助历史信息来判断自己已完成指令的进度。然而，历史帧中往往包含大量冗余和重复的信息。对于当前帧，模型不仅要提供该帧所在的最新场景信息，还需预测出符合指令的下一步合理动作。

图14 NaVid 架构图



资料来源：Jiazha Zhang 《NaVid: Video-based VLM Plans the Next Step for Vision-and-Language Navigation》，海通证券研究所

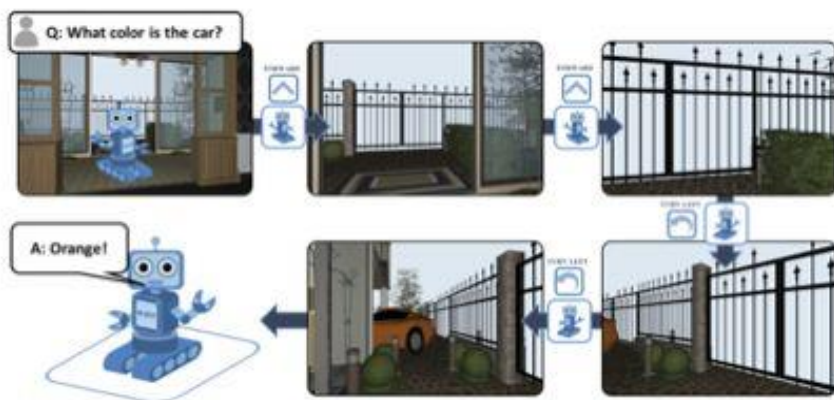
目前该方法还有两个难点：1）导航数据的模态与大模型常见的数据模态不一致。机器人的导航数据需要建模历史信息和当前信息，并保证导航过程中动作输出的格式一

致性。2) 缺少大量高质量的视觉语言导航任务的真实数据。在真实世界收集这样的数据极其耗时耗力，且在场景和指令的多样性上有欠缺。这些困难限制了视频大模型用于导航任务的前景。

## 2.4 具身交互 (Embodied Interaction)

具身交互任务指的是智能体在物理或模拟空间中与人类和环境进行交互的场景。典型的具身交互任务是具身问题解答 Embodied Question Answering (EQA)。在这个任务里面，一个 agent 在三维虚拟空间中被进行询问，agent 为了回答这个问题，需要在环境中进行探索和信息整合。这个任务需要 agent 具有主动的认知、语言理解能力、目标驱动的探索、常识推理并将自然语言的信息整合到动作序列中。

图15 EQA 任务例子



资料来源：腾讯云，海通证券研究所

图中给出了 EQA 的一个形象化的表述，在这个例子中，向 agent 提出的文本问题为“车是什么颜色的？”，agent 经过在虚拟的环境中进行探索，当发现车子后，返回给系统“车是橘黄色的。” EQA 的最终目标是：agent 可以理解他们所处的环境（通过视觉等感知方式），具有沟通交流的能力，可以采取一定的动作（例如在环境中根据问答或对话进行主动的环境探索）。

Agent 需要具有的一些能力进行具体的分析：**主动探索性**：由于 agent 的出现地点是在环境中随机生成的，所以 agent 需要进行一定的动作和移动去寻找能够回答问题的视觉信息。**常识信息推理**：agent 没有一个关于虚拟环境的地图，所以 agent 需要自行对环境进行探索。因此 agent 必须自行构建一个常识系统，就像人一样去比较有目的的进行探索房屋（例如一个人接收到一个类似厨具颜色的问题，他会根据自己的常识经验直接前往厨房、餐厅等去查看，拿到相应的结果后再进行回答，而不会漫无目的的在房间中摸索）。**收敛性分析**：在这个 EQA 的任务上，难度就是 agent 的探索可能是很长时间都难以收敛的（比如说问题是大房子里一共有多少个小房中有椅子，这种情况下 agent 可能跑了很多次也不能给出完全正确的答案。此时开发者也不知道究竟是 agent 理解错了，还是没有识别出椅子，还是计数出现了错误，这就导致进行训练的时候，难度非常大），为了解决这种问题，作者提出利用模仿学习的策略给 agent 进行一个示范，或者是给出一些引导进行前期的训练，以方便 agent 进行更好的学习和收敛。

## 2.5 具身智能全面落地仍需解决四大难题

### 1. 通用本体平台的挑战

解决硬件的关键零部件技术突破，形成具有优秀运动能力和操作能力的平台级通用机器人产品；平衡机器人本体的可靠性、成本和通用能力是难题。尤其在被认为是具身智能终极形态的人形机器人领域，相关研发依然是热点和核心挑战。

### 2. 智能体系统设计的难点

智能体作为具身智能的核心，必须具备复杂环境感知和认知能力。这包括 3D 环境感知、任务编排与执行、多轮人机交互、long-term 记忆和任务迁移等多项挑战。此外，具身智能要求实时感知和决策能力，以适应复杂和变化的环境。这要求高速的数据采集、传输和处理，以及实时的决策反应，尤其是 LLM 所消耗的算力规模巨大，对于资源有限的机器人处理系统将形成巨大的数据量、AI 计算能力和低延迟的挑战。

### 3. 高质量数据需求

现实场景的复杂多变，使得现阶段缺乏足够的场景数据来训练一个完全通用的大模型，进而让智能体自我进化。耦合的个体，需要实际部署到真实环境中，才能够采集数据，这也是和非具身智能的明显不同。但对于关键业务，要求成功率，则仍然需要高质量的垂直数据。同时，通过层次化的智能体设计，将不同任务限定到特定领域，则是一个解决泛化和成功率的有效尝试。

### 4. 虚拟与真实交互中的学习进化

通过虚拟与真实环境的互动，具身智能体能够持续学习和进化，以应对复杂环境。尽管形态变化无穷，但在有限计算资源下快速学习合理的规划和决策能力，是具身智能进化中的重要课题。

## 3. 应用场景：具身智能产品多样，覆盖广阔市场

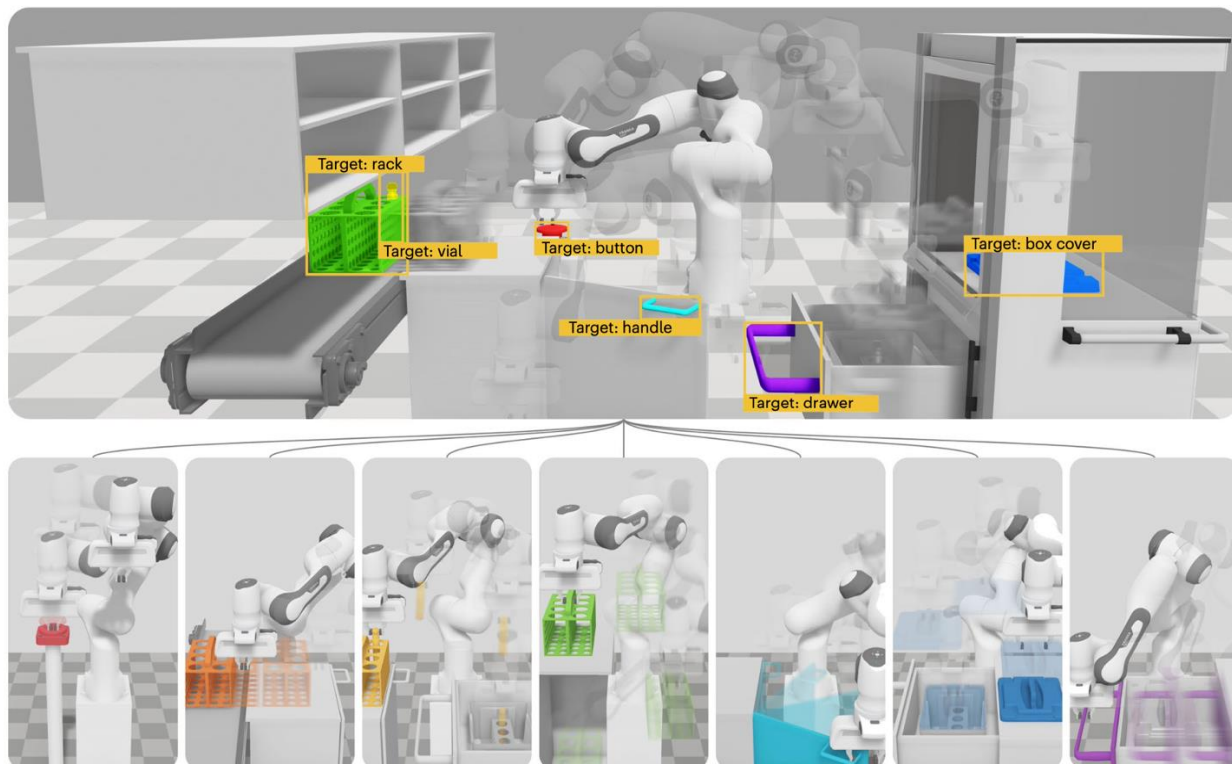
不同类型的机器人在各自的应用领域中展现出重要的功能和价值。1) 固定式机器人因其高精度和稳定性，广泛应用于实验室自动化、教育和工业制造。2) 轮式机器人在物流、仓储和安全检查中表现优异，而履带式机器人适用于农业、建筑和军事等复杂地形。3) 四足机器人因其稳定性和适应性，广泛用于复杂地形探索、救援任务和军事行动。4) 人形机器人在服务业、医疗保健和协作环境中日益普及，应用于智能交互、精密制造和医疗手术等领域。各类机器人在不同应用场景中展现独特优势，同时面临成本和技术挑战。

### 3.1 固定基座机器人：全面赋能实验室与工业场景

固定式机器人由于其精度和稳定性，广泛应用于实验室自动化、教育和工业制造。它们在需要高精度的任务中表现出色，但受限于固定位置，成本高且维护要求高。其中一个新颖的框架被称为机器人操作网络 (Robotic Manipulation Network ROMAN)。ROMAN 是一种机器人操作网络，通过混合层次学习 (HHL) 来解决复杂的顺序任务。它结合了模仿学习和强化学习，能够执行多种复杂的操作任务，比如按按钮、拾取物品、旋转、插入等。ROMAN 的优势在于它可以处理长时间任务，适应随机的任务顺序，并且在出错时能够自动恢复，继续完成任务。



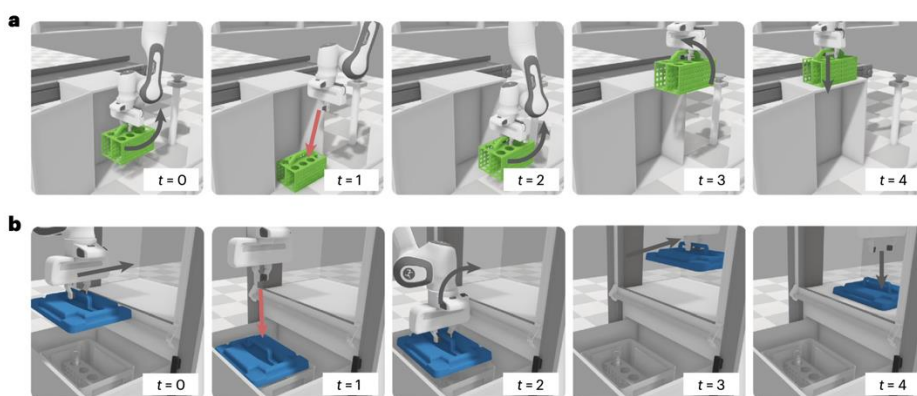
图16 ROMAN 框架的功能



资料来源:《Robotic Manipulation Network (ROMAN)— Hybrid Hierarchical Learning for Solving Complex Sequential Tasks》, CSDN, 海通证券研究所

HHL 结合了行为克隆 (BC)、生成对抗模仿学习 (GAIL) 和近端策略优化 (PPO) 等算法。首先,行为克隆用于初始化策略,通过学习专家演示的动作来引导机器人。然而,单独依赖模仿学习存在局限性,因此 HHL 还利用 GAIL 和 PPO 来增强学习过程。GAIL 通过对比机器人生成的轨迹与专家轨迹,进一步优化机器人行为,而 PPO 则提供了一个外部奖励信号,帮助机器人在探索中找到更优的策略。通过这种方式,HHL 框架能够在长时间任务中适应新情况,并具有从错误中恢复的能力。

图17 ROMAN 从错误中恢复的效果展示



资料来源:《Robotic Manipulation Network (ROMAN)》, CSDN, 海通证券研究所

ROMAN 具备识别错误的能力,且能快速进行动态恢复。当 ROMAN 在执行“拾取和放置”以及“拾取和丢弃”子任务时出现错误的情况下,系统迅速调整策略,以重新抓取物品并完成任务。这些错误可能包括抓取对象不准确、执行轨迹出错或激活了错误的任务顺序,但 ROMAN 通过其中央操控网络,能够在任务失败时激活不同的专家网络,进行调整和修正。这种能力展示了 ROMAN 框架的强大适应性,不仅能够执行预设的任务序列,还能够在遇到新的或意外的情况时进行调整,恢复并继续完成任务。通过平衡探

索和利用，ROMAN 生成了超越原始示范的行为，体现了在面对复杂和动态环境时的灵活性。

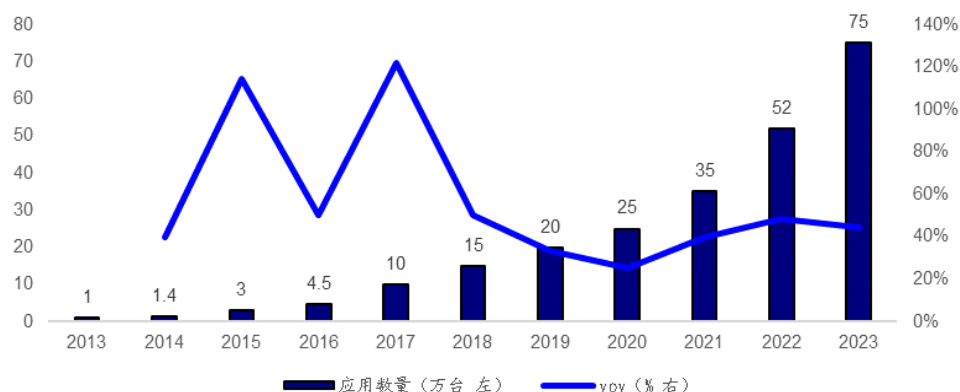
目前工业机械臂市场规模稳步增长，国产品牌份额提升。根据中国报告大厅发布的文章表示，2022 年中国机械臂市场规模接近 178.3 亿元，同比增长 6.26%。到 2023 年，市场规模增至 186.4 亿元，预计 2024 年将达 193.4 亿元，2025 年将超过 200 亿元。2022 年，国内品牌工业机械臂市场份额显著提升，达到 71.2 亿元。到 2023 年，国产工业机器人市场份额为 52.45%，首次超过外资品牌，显示出国产品牌市场竞争力的增强。

### 3.2 轮式/履带式机器人：高机动性适应复杂道路环境

移动机器人适应复杂多样的应用场景。轮式机器人因其高效机动性，广泛应用于物流、仓储和安全检查。其优点包括结构简单、成本较低、高能效及在平坦表面快速移动的能力，但在不平坦地形中表现有限。与此不同，履带式机器人在复杂地形中表现出色，适合农业、建筑和军事领域，但能效较低且在平坦表面上移动较慢。两种机器人各有优劣，适用于不同环境和任务。

在物流领域，亚马逊的 Kiva 机器人较为成熟。自 2012 年收购 KIVA 以来，亚马逊探索各类机器人及智能化技术在仓储物流领域的应用，截至目前，亚马逊内部使用的机器人数量已超 75 万，是全球最大的机器人使用主体之一。

图18 2013 年-2023 年亚马逊机器人应用数量



资料来源：中国移动机器人（AGV/AMR）产业联盟援引新战略咨询，海通证券研究所

KIVA 机器人能够提升仓库拣货效率。Kiva 开发了一套系统，包括可移动货架、二维码地板网格、智能包装站、橙色机器人和强大的软件系统。我们认为，Kiva 彻底改变了仓库内部的物流方式。创新想法是让货架移动到工人面前，而非让工人走到货架前。这一系统大幅提高了配送中心的效率，减少了人工错误和操作成本。



图19 KIVA 机器人构造



图1 托举铸件背面、铸件局部



图2 前壳内部红外传感器组、后壳顶部视图

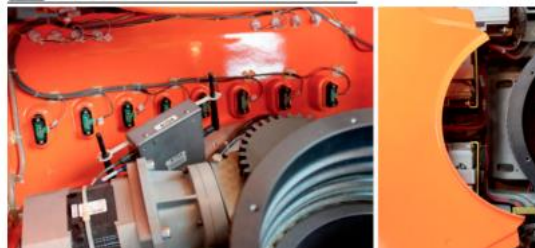


图3 KIVA注塑罩壳



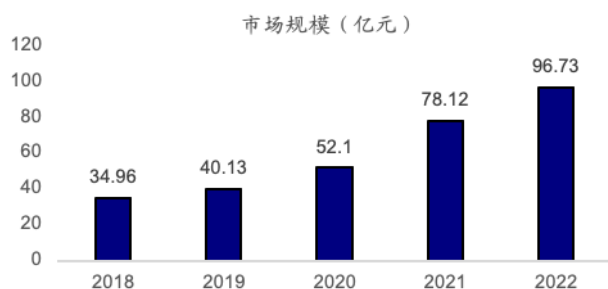
图4 管式碰撞传感器、红外传感器



资料来源：CSDN，海通证券研究所

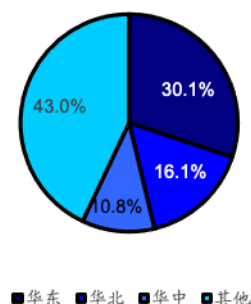
Kiva 机器人通过精巧的设计提升了仓库自动化的安全性和效率，同时优化了制造成本和操作可靠性。Kiva 部署的仓库地面上每隔大约 1 米就有一个二维码，Kiva 就根据这些标记进行定位和误差补偿。它的每一个动作都来自于云端的指令。在它到达目标货架底部后，其使用一个滚珠丝杠升降梯结构，通过原地旋转来升高自己，将货架顶起约 10 厘米。从外面看，Kiva 机器人外壳的每一侧都有红外传感阵列，以及气动保险杠，用于检测和缓冲碰撞。外壳上还有充电接口和一系列状态指示灯。每一台 Kiva 机器人有三个独立的自由度：两个驱动轮，加上一个用于起重的旋转电机。起重电机转动时，两个驱动轮反方向旋转，结果是托盘相对于地面没有旋转，只在滚珠丝杠的作用下升高。相比于传统的大负载直线驱动方案比如液压、剪式升降台，Kiva 的这种利用轮子的结构显然更加简单可靠。

图20 2018 年-2022 年中国移动机器人市场规模



资料来源：观研报告网搜狐号，海通证券研究所

图21 2022 年中国移动机器人市场规模分布情况



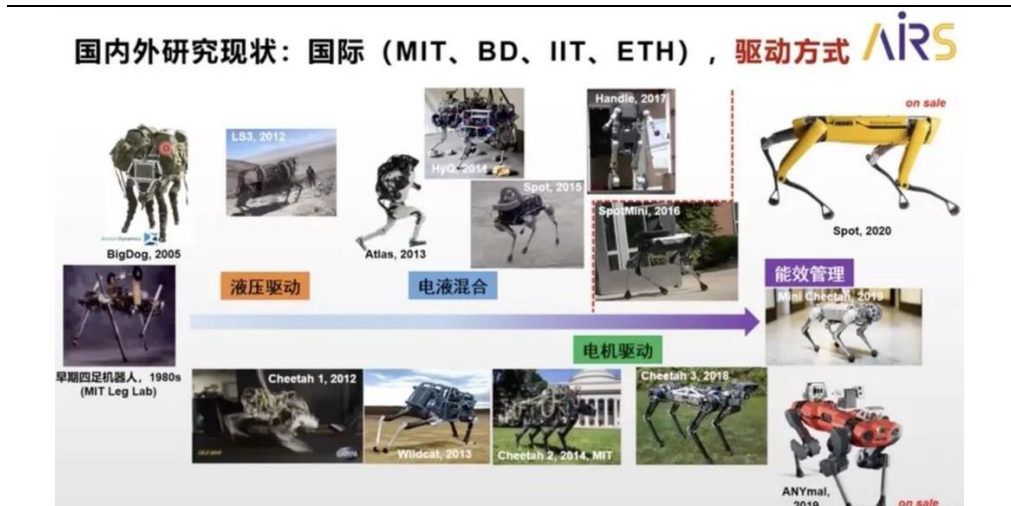
资料来源：观研报告网搜狐号，海通证券研究所

受益于制造业自动化、服务业升级以及技术创新，我国移动机器人市场一直处于迅速发展阶段。2021-2022 年，国外深陷疫情泥潭时，我国疫情得到有效防控，企业复工复产已成常态化，因此国内机器人成为众多海外市场的选择；根据搜狐援引观研天下的文章中指出，预计随着随着机器人技术的快速发展、应用场景的多样化，未来我国移动机器人行业将有千亿级别空间。搜狐援引观研天下的数据显示，2018-2022 年我国移动机器人市场规模从 34.96 亿元扩大到 96.73 亿元。目前我国移动机器人行业主要集中在华东地区，占比达到了 30.11%；其次为华北地区，其市场规模占比为 16.14%。

### 3.3 四足机器人：龙头制造商占据大量市场份额

四足机器人以其稳定性和适应性而广泛应用于复杂地形的探索、救援任务和军事行动。它们的设计灵感来源于四足动物，通过多关节设计实现生物运动的模拟，并能够自动调整姿态，以适应不断变化的地形。传感系统如激光雷达和摄像头则为机器人提供环境感知，使其能够自主导航和避障。当前，几种四足机器人被广泛使用，包括 Unitree Robotics 的 Unitree A1 和 Go1、Boston Dynamics 的 Spot 以及 ANYbotics 的 ANYmal C。Unitree A1 和 Go1 因成本效益和灵活性广受欢迎，具有强大的机动性和智能避障能力。Boston Dynamics 的 Spot 则以其出色的稳定性、负载能力和操作灵活性著称，适用于工业检查和救援任务。尽管四足机器人技术先进，但其高昂的制造成本和电池续航限制了其在某些领域的广泛应用。

图22 四足机器人发展路线



资料来源：AIRS，海通证券研究所

例如 AIRS 团队的麒麟机器人，团队引入仿生设计思路来优化设计四足机器人整体结构。基于对中型犬解剖骨骼结构的仔细研究，合理设计每个部位的尺寸。机器人整体结构具备较高负载能力，同时大大提升了越障性能。团队结合复杂地形感知技术，提出了一种能够让四足机器人在通过崎岖 3D 地形时实现高能效的运动。

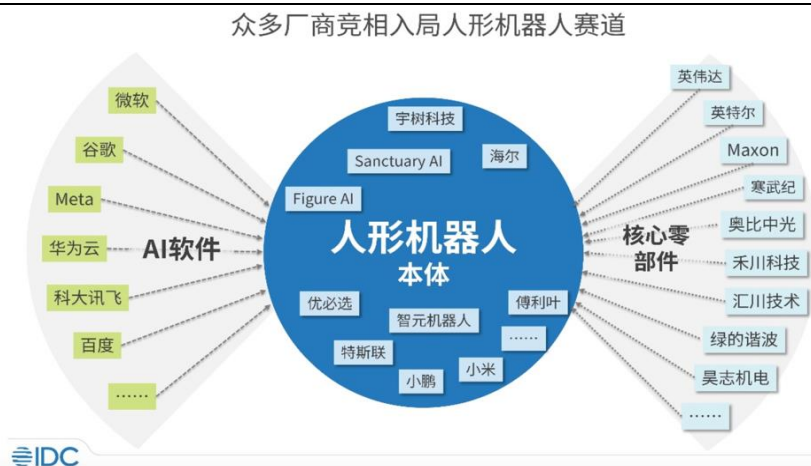
根据 AIRS 援引 QYR（恒州博智）的统计及预测，2023 年全球腿式机器人市场销售额为 6.02 亿美元，预计 2030 年将达到 15.11 亿美元，CAGR 为 14.3%（2024-2030）。全球主要的腿式机器人（Legged Robot）生产商包括索尼、优必选、蔚蓝智能科技和乐聚机器人，市场占比超过 68%。其中，索尼以约 28% 的市场份额领先。亚太地区是最大的市场，占比约 57%；其次是北美和欧洲，分别约占比 24%和 12%。从产品类型来看，四足机器人占市场份额最大，约 56%。在产品应用方面，最大的应用领域是教育和娱乐，其次是科学研究。

### 3.4 人形机器人：未来拥有强大潜力，技术仍需探索

人形机器人因其类人形态在服务业、医疗保健和协作环境中越来越普及。它们模仿人类的动作和行为，提供个性化服务，特别是在医疗手术和精密制造等需要复杂操作的领域。人形机器人的灵巧手设计拥有多个自由度和高精度传感器，使其具备出色的抓握

和操纵能力。例如，Boston Dynamics 的 Atlas 以其卓越的机动性和稳定性闻名，能够执行复杂动态动作。AIST 的 HRP 系列在研究和工业应用中表现出色，具有高稳定性和灵活性，适合复杂环境下与人类协作任务。Honda 的 ASIMO 能够行走、跑步、爬楼梯，还能识别面部和手势，适用于接待和导游服务。Softbank Robotics 的 Pepper 是一种小型社交机器人，能够识别情绪并进行自然语言交流，广泛应用于客户服务和教育环境。尽管人形机器人在复杂环境中的稳定性和可靠性方面仍面临挑战，但通过引入电机驱动系统和整合大语言模型（LLMs），人形机器人有望在制造业、医疗保健和服务行业中填补劳动力缺口，提高效率和安全性。

图23 人形机器人产业各大关联厂商



资料来源：IDC，海通证券研究所

各大厂商合作打造具身智能，本体集成仍存在技术瓶颈。AI 软件相关厂商通过提供 AI 算力、大模型和算法软件，推动人形机器人实现视觉分析、语义理解、任务分解和动作规划等功能；核心零部件厂商则提供各类传感器、电机、减速器等硬件设备，构建机器人所需的感知、控制、动力和运动模块；机器人本体厂商负责将这些硬件和软件整合，打造完整的人形机器人产品。然而，人形机器人商用化进程仍面临技术瓶颈，特别是在本体集成、运动控制、感知系统和智能分析等关键技术方面，需要进一步突破才能实现自主导航和任务执行。

图24 人形机器人应用展望



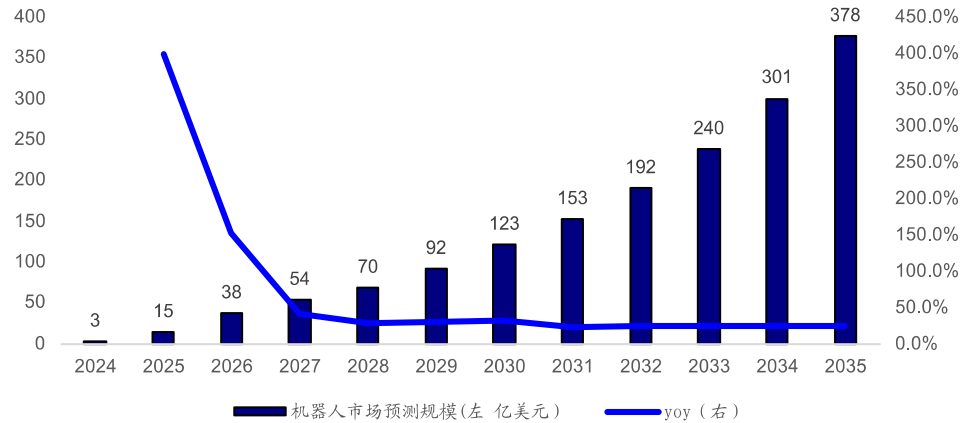
资料来源：IDC，海通证券研究所

人形机器人的市场普及需要产品价格的逐步下降。目前相较于商用服务机器人和工业机器人，人形机器人的价格仍然高达数十万至百万级，这限制了用户的采纳意愿。然而，随着技术突破、产业链发展和应用规模扩大，价格有望持续下探，从而提高用户的购买意愿。产品的核心优势在于其场景通用性，特别适用于复杂、多变的环境和需要高度人机协同的任务。为了实现这一目标，软硬件的解耦设计至关重要，即在坚实的硬件基础上，根据不同需求动态加载各领域的大型模型。



人形机器人的商业化进程需要循序渐进。首先在商用服务领域，如酒店、商场等场所提供智能交互服务。随着技术迭代，机器人将逐步进入灾害救援和极限环境作业等特殊应用领域。在工业制造中，人形机器人会率先在精密装配和物料搬运等环节实现自动化作业。最终，随着技术和市场的成熟，人形机器人将进入家庭市场，承担陪伴、教育和家务等多种角色，满足家庭日常生活需求。

图25 2024-2035 年人形机器人市场规模预测



资料来源：高盛，海通证券研究所

人形机器人行业的快速发展受到政策支持、技术进步和成本控制等多方面因素的推动。未来三年，人形机器人产业的年增长率预计将维持在40%以上。政策方面，北京亦庄将出台人形机器人三年行动计划，进一步打造国内领先的机器人技术创新策源地、高端制造集聚地、集成应用新高地和产业生态示范地。同时，上海举办的2024中国人形机器人开发者大会，围绕人形机器人创新中心建设、核心技术壁垒突破、产业链成本降低、应用落地场景、投资赛道方向等议题展开。根据证券时报援引赛迪顾问发布的《2024中国人形机器人产业生态发展研究》显示，2023年人形机器人产业进入爆发期，预计到2026年中国人形机器人产业规模将突破200亿元，未来功能型整机将逐步占据主流，发展潜力巨大。

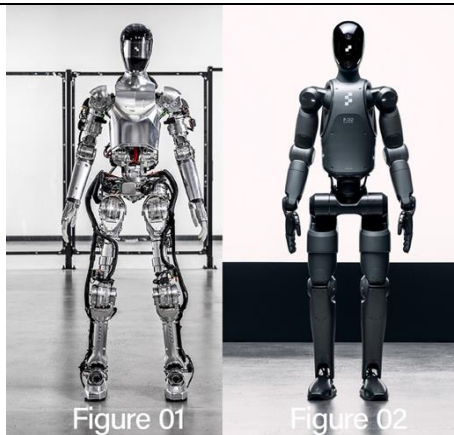
## 4. 潜在标的：美国商业化更为成熟，中国仍需探索

### 4.1 Figure AI：获巨头投资，技术不断成熟

Figure.AI 专注于人形机器人领域，并与汽车厂商合作。公司成立于2022年，专注于设计和制造自主人形机器人。其旗舰产品 Figure01 是全球首款商业化的自主人形机器人，主要应用于制造、物流、仓储和零售行业。Figure.AI 的团队由经验丰富的机器人专家组成，致力于推动人工通用智能（AGI）的研究，并在机器人性能和稳定性方面进行深入研究。公司近期与 BMW 签署了合作协议，将其机器人技术引入汽车生产领域。

Figure02 搭载语言模型，距离具身智能更进一步。2024年8月公司推出新 Figure02 机器人，相比 Figure01 更加紧凑，电源和计算设备的布线被巧妙隐藏，提升了美观性和耐用性。其手部设计是亮点，配备了具有“16个自由度”的机械手，和人手尺寸相当，上面还配有传感器。这种设计使得机器人能够处理人类需要完成的任务，并且可以提起重达20公斤的物体。其手部的改进为实现与人类相当的力量和灵巧度铺平了道路。机器人还配有6个RGB摄像头，帮助其通过视觉语言模型进行语义理解和常识推理。

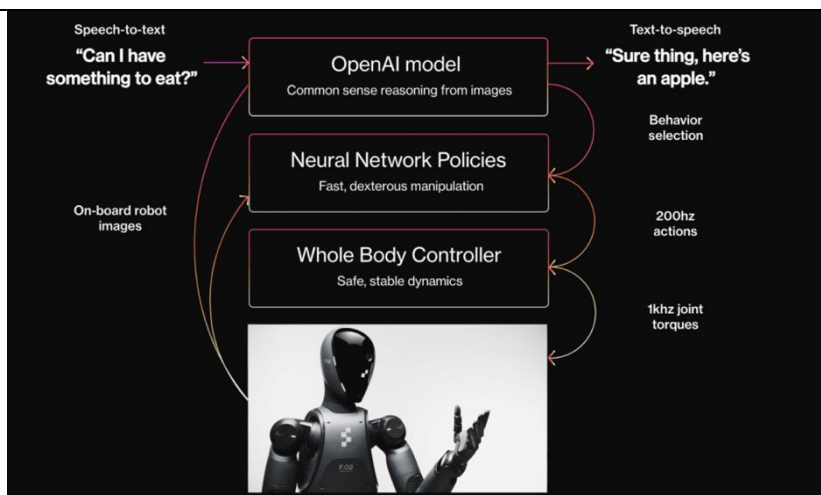
图26 Figure AI 第一代与第二代机器人



资料来源：新浪科技，DeepTech 深科技援引 Figure AI，海通证券研究所

Figure02 能够进行语音识别，续航能力显著增强。用户可以通过语音指令给 Figure 02 机器人下达任务，机器人会结合语音命令和摄像头数据来理解并执行相应的行动。相比 Figure01，Figure02 的计算和 AI 推理能力提高了 3 倍，任务执行速度更快。它配备了一个 2.25 千瓦时的电池组，比上一代多 50% 的电力，运行时间为 5 小时，目标是延长到每天工作 20 小时。

图27 语音模块的工作模式解析



资料来源：新浪科技，DeepTech 深科技援引 Figure AI，海通证券研究所

人形机器人在工厂和仓库环境中具有广阔的应用前景，Figure 02 可以有效面对劳动力短缺的问题。除了工业应用，Figure 公司暗示其机器人未来可能进入家庭领域。Figure AI 并非唯一玩家，其他公司如 1X Robotics、特斯拉等也在这一领域竞争，许多公司还使用英伟达的多模态 AI “Project GR00T” 来驱动机器人功能。值得注意的是，Figure AI 与 OpenAI 建立了合作伙伴关系，并获得了微软、亚马逊等科技巨头的投资，助力其机器人开发和语音交互能力的提升。

风险提示：人形机器人发展不及预期；EAI 发展不及预期；具身智能下游需求不足。

## 4.2 特斯拉 Optimus：优先赋能特斯拉工厂

在 2024 世界人工智能大会（WAIC 2024）上，特斯拉展示其最新 Optimus 人形机器人进展。这表明人形机器人产业化的进度可能超预期。Optimus 人形机器人现已具备在工厂内执行任务的能力，并在多项技术上取得了显著提升，预计到 2025 年将有超过 1000 个 Optimus 机器人在特斯拉工厂内工作。特斯拉的持续创新和技术投入将推动智



能制造和自动驾驶领域的发展，并为全球市场带来新的变革。

图28 特斯拉 Optimus 自主工作



资料来源：锦囊专家，36 氪援引特斯拉官方视频，海通证券研究所

当前 Optimus 已更新至第二代，身高 1.7 米、体重 56 公斤，最高行走速度达到 5 英里/小时，搭载了特斯拉自主设计和制造的执行器和传感器驱动。2024 年 5 月，Optimus 二代走进特斯拉工厂实训，已经可以进行分拣电池、行走等任务。根据 36 氪援引特斯拉官方介绍，Optimus 基于其自身的视觉神经网络和 FSD(完全自动驾驶)芯片，能够将电动车电池单元进行分装，精准放到卡槽中。

根据第一财经报道，马斯克预测，明年特斯拉将有超过 1000 台，乃至数千台在运行的 Optimus 机器人。目前已有两台在弗里蒙特工厂工作。虽然 Optimus 仍处于早期阶段，但其未来潜力巨大，马斯克预计人形机器人数量将超过人类，并在工业领域发挥主力作用，目标年产 10 亿台。特斯拉计划将 Optimus 的成本控制在 1 万美元，以 2 万美元出售。人形机器人商业化元年已至，特斯拉及多家公司正在探索汽车工业作为落地场景，推动产业化进程。

风险提示：人形机器人发展不及预期；EAI 发展不及预期；具身智能下游需求不足。

#### 4.3 宇树科技：技术领先，覆盖场景多元

杭州宇树科技是一家世界知名的民用机器人公司，专注于消费级、行业级高性能通用足式/人形机器人及灵巧机械臂的自主研发、生产和销售。在机器人核心零部件、运动控制、机器人感知等综合领域具有卓越的领先性。宇树高度重视自主研发和科技创新，全自研电机、减速器、控制器、激光雷达等机器人关键核心零部件和高性能感知及运动控制算法，整合机器人全产业链，在足式机器人领域达到全球技术领先。目前累计申请国内外专利 180 余项，授权专利 150 余项。

图29 宇树科技产品一览

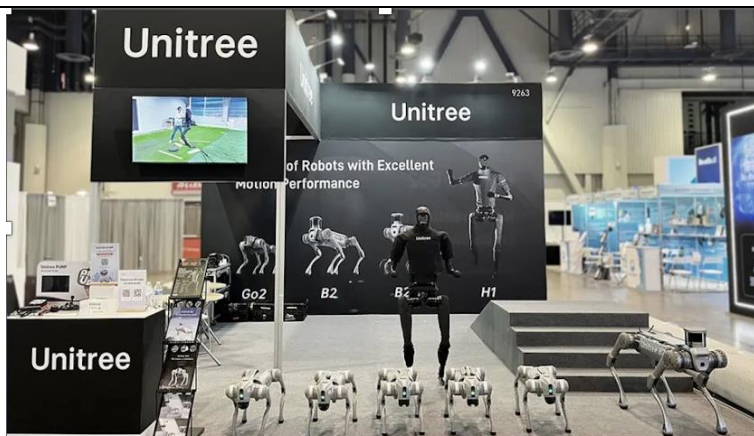


资料来源：宇树科技官网，海通证券研究所

商业模式上，新的人形机器人产品主要面向高校、研究所、科技公司和工业落地，作为机器人开发平台研究或落地使用。四足机器人产品目前正在向 to C 和 to B 方向齐

头并进，to C 定位个人以及家庭服务场景；to B 在特定的消防、巡检、军警等领域执行任务。

图30 CES2024 宇树科技产品展示



资料来源：凤凰网援引公司官网图片，海通证券研究所

产品针对多样客户群体，覆盖大量行业。在 C 端领域，Go 部分型号产品就是消费级产品。其中，宇树 2023 年发布的 Unitree Go2 四足机器人适用于外出陪跑、散步等场景，支持实时视频图传和雷达高度图显示功能，奔跑速度可以和最顶级的马拉松选手不相上下，还搭载了公司最新的 AI 交互技术，引入了大语言模型能力。该产品售价仅 9997 元起。B 端例如更大体型的 B2 则是一款工业级机器狗产品，极限负重 100kg，持续作业负重 40kg 以上，可以应用于农业、工业、安防巡检、勘测探索、公共救援等行业级场景。

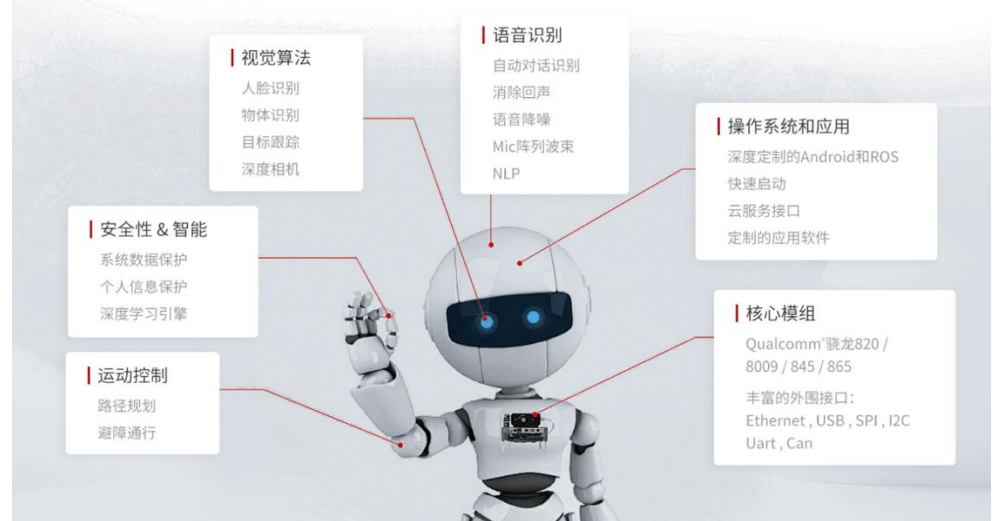
宇树科技四足机器人销量领先，核心零部件自研能力强。根据 36 氪对宇树科技创始人王兴兴的采访，2024 年宇树科技的四足机器人产品占全球出货量的 60% 以上，处于全球历年销量领先的位置。在智能机器人赛道，宇树科技的核心竞争力在于全栈自研技术积累的先发优势——公司不仅自研了高能量密度关节电机、减速器、3D 激光雷达等机器人核心零部件；以及传统机器人领域的感知和运动控制算法；同时还在开发机器人 AI 世界模型。

风险提示：EAI 发展不及预期；具身智能下游需求不足。

#### 4.4 中科创达：布局端侧智能+机器人

中科创达将基于 Arm 技术赋能下一代多模态智能机器人。2024 年 8 月 27 日，在 elecon2024 深圳国际电子展的第六届中国嵌入式技术大会上，中科创达受邀与 Arm 联合进行了一场主题为《如何通过先进的大语言模型解锁下一代多模态智能机器人》的精彩演讲。会中，中科创达明确表达了中科创达将凭借 Arm 处理器与其操作系统技术的协作，助力具身智能与机器人实现创新突破。

图31 中科创达机器人产品



资料来源：中科创达官网-集团介绍页面，海通证券研究所

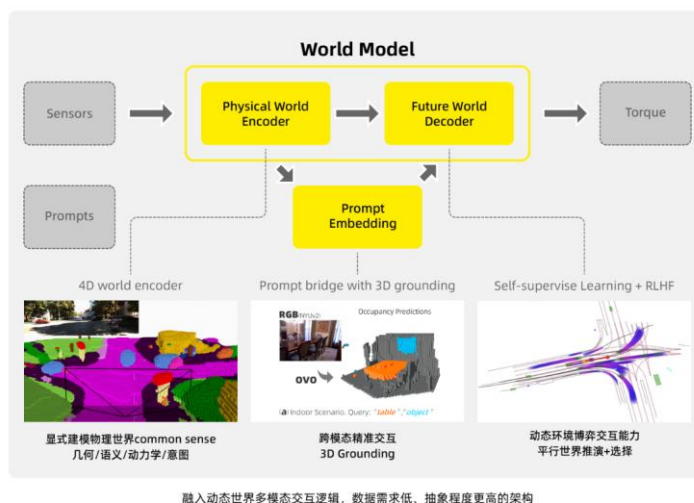
中科创达通过核心技术实现机器人不同场景中的智能化。公司可以提供核心技术和生态合作广泛地赋能机器人产业发展。从基于基础类半导体模块，提供平台工具软件、平台层、算法和应用的集成，到构建边缘计算成为一个共通的计算平台，推动操作系统走向平台化，从而支撑算法的可扩展、可扩充和之后的可延展的维护，进而迭代开发工具和环境去支撑应用厂商更快产品和更低成本。未来是各种各样的机器人存在的世界。本质上是随着智能化和信息化的分离，信息化是以人和人交互为中心，智能化则是通过机器感知周边的世界和环境，关键是传感器及其运动能力、操作能力进行连接，而形成一个所谓的机器人时代。公司的端侧智能+机器人的布局，进一步为机器人产业的变革注入了更多活力与可想空间。

风险提示： EAI 发展不及预期；具身智能下游需求不足。

#### 4.5 有鹿机器人：引入“通用智能大脑”概念

根据中自数字移动传媒报道,2024 年 5 月从事具身智能技术和产品研发的有鹿机器人正式签订一笔超千万元的具身智能通用大脑 Master2000 的订单。订单来自中国头部清洁设备制造商，采购大脑用于对现有全系产品进行智能化升级。有鹿机器人也正式实现了具身智能从实验室走向商业化的阶段。公司同步披露了创新工场、元璟资本、百度风投等超 1 亿元融资的信息。

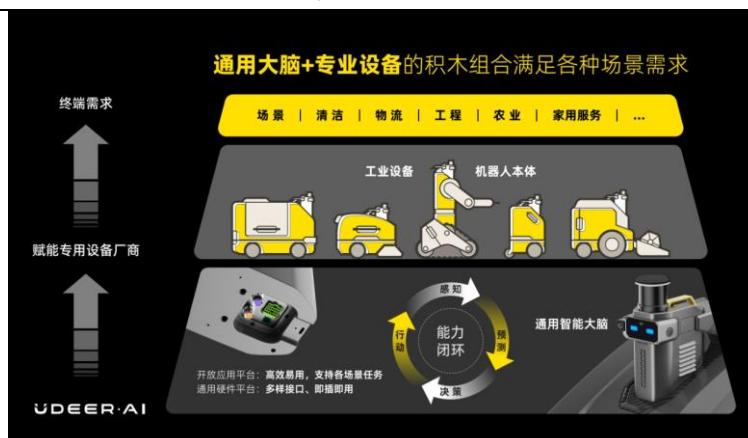
图32 有鹿机器人具身智能大模型



资料来源：中国数字移动传媒，海通证券研究所

2023 年，有鹿机器人联合阿里云通义千问发布了具身智能大模型 LPLM-10B，该模型创新性地融合了语言大模型与物理世界，克服了传统具身模型的延迟高、准确率低、泛化能力差等问题。LPLM-10B 的优势在于其通用性，能够在跨模态、跨场景、跨行业的数据中提取共性信息，适用于各类型机器人。2024 年 3 月，有鹿机器人推出了基于 LPLM-10B 的通用具身大脑产品 Master 2000，帮助客户解决智能化升级中的投入产出比低的问题。该产品在短时间内实现了超千万的订单并开始量产交付。与大多数大模型公司面临的商业化挑战相比，Master 2000 的成功验证了软硬件结合产品在商业化中的优势，为大模型的商业应用提供了新的战略意义。

图33 有鹿机器人打造“通用大脑”概念



资料来源：中国数字移动传媒，海通证券研究所

有鹿机器人创始人陈俊波博士表示，中国拥有世界领先的制造能力，在具身智能时代，预计中国制造的产品即将爆发式迎来机器人形态，如轮式机器人形态的叉车、物流车、清洁车、割草机、充电车等，复合式机器人形态的挖掘机、装载机、吊车等，以及中国大力发展的新型人型机器人。这些机器人产品形态的总体市场规模预计将在万亿级以上。有鹿机器人的使命就是打造一个可以赋能不同制造企业的通用机器人脑，适配于专业设备，帮助成熟的制造企业进行产品的具身智能模态升级。

风险提示：EAI 发展不及预期；具身智能下游需求不足。

## 4.6 科大讯飞：讯飞超脑计划，让机器人走向通才



科大讯飞打造全球领先的智能交互机器人开放创新平台，为机器人产业提供通用化的智能应用服务，大幅降低行业技术研发和使用门槛。机器人开放平台旨在赋能机器人硬件，快速接入 AI 能力，并提供可视化开发机器人业务以及项目管控和运营监测能力。实现可复用化、服务统一化、流程角色优化、运营监控化和资源管控化，使平台成为一个强大的机器人能力支持中心。最终依托机器人开放平台，赋能于机器人产业公司，建立机器人合作生态，推动机器人产业持续健康快速的发展。

图34 科大讯飞机器人平台架构

业务应用前台	酒店	餐厅	展厅	政务	商超	工业	更多
AI中台	AI能力服务	语音能力 语音识别 语音合成 声纹识别 语种识别 语音预处理 原声比对		图像能力 人脸识别 OCR识别 视频结构化 证件识别 相似图识别 工业视觉		语义认知 机器翻译 意图识别 关键词抽取 自动编目 问题相似度 语义解析	
	平台服务	深度学习平台 整合标注工具、开源框架、数据集、预训练模型等，提供一站式深度学习研发实验环境		运营监控平台 运营管理后台 运营分析报告 数据运营服务 人工后台服务 自助分析 性能监控 监控预警 状态反馈		管理平台 角色权限管理 服务资源管理 多租户管理 知识库管理 会话管理 技能管理 可视化流程引擎 配置中心	
数据中台	数据采集	数据处理	数据分析	数据管理	数据安全	数据服务	更多

资料来源：科大讯飞机器人平台官网，海通证券研究所

科大讯飞机器人业务应用广泛，覆盖多个垂直领域。业务应用前台包括酒店、餐厅、展厅、政务、商超、工业等多个行业，展示了 AI 解决方案在不同行业的应用场景。AI 中台由 AI 能力服务和平台服务两部分组成。AI 能力服务包括语音能力（如语音识别、语音合成、声纹识别）和图像能力（如人脸识别、OCR 识别、工业检测）。平台服务则包括深度学习平台、运营监控平台和管理平台。深度学习平台提供整体的 AI 开发工具，运营监控平台涵盖数据监控、AI 运维服务等，而管理平台则负责角色权限、数据管理等功能。数据中台则为这些服务提供支持，涵盖数据采集、数据处理、数据分析等内容。

图35 科大讯飞超脑计划 2030



资料来源：中国科技网，海通证券研究所

科大讯飞机器人首席科学家李超博士分享了团队结合具身大模型打通基于模型优化和基于数据驱动的运动控制路线齐头并进的技术路线，深度解析了大模型给具身智能通用机器人带来的提升空间和产业化落地瓶颈及挑战，并提出了场景导向、产业链合作、突破卡脖子问题和交叉学科人才培养四方面建议，推动具身通用机器人产业化快速发展。

科大讯飞依托超脑 2030 计划旨在让懂知识、善学习、能进化的 AI 机器人进入千家万户，大模型底层能力的突破，给机器人带来了巨大的提升空间。作为人工智能的重要分支，具身智能正逐步从理论走向实践，从“实验室”走向“应用场”。随着具身智能技术的不断突破和应用，未来的机器人将会更加智能，更加接近人类的感知、思考和行动能力，真正实现从“专才”走向“通才”，在各个领域发挥更大的作用。

风险提示：大模型发展不及预期；具身智能下游需求不足。



#### 4.7 海康威视：视觉与移动机器人提供商

海康机器人是面向全球的机器视觉和移动机器人产品及解决方案提供商，业务聚焦于工业物联网、智慧物流和智能制造，构建开放合作生态，为工业和物流领域用户提供服务，以创新技术持续推动智能化，引领智能制造进程。

图36 海康威视移动机器人产品一览



资料来源：海康威视官网-产品系列，海通证券研究所

产品随着基于 AMR、iWMS-1000 和 RCS-2000 等智能设备和系统为核心的解决方案落地实施，持续推动内物流智慧化变革。通过全方位的智能算法驱动、智能设备联动、智能系统应用，实现内物流运营的集约化、可视化和智慧化，持续为客户创造价值。

风险提示：EAI 发展不及预期；具身智能下游需求不足。

#### 4.8 比亚迪电子：AMR 提供物流解决方案

公司基于英伟达 Isaac 机器人开发平台打造自主移动机器人。公司为保障工人安全、优化运输流程、节省生产成本，且为其他各界客户提升生产智能化，比亚迪电子自主开发全系自主移动机器人（AMR），为工厂提供全面物流解决方案。比亚迪电子开发高阶 AMR 过程中，NVIDIA 基于 NVIDIA Isaac 和 Jetson 平台给予了技术支持，帮助客户加速物流应用的部署。

图37 比亚迪电子 AMR 机器人



资料来源：比亚迪电子官网-产品与服务，海通证券研究所

比亚迪电子的 AMR 采用了 NVIDIA Jetson Orin 高性能人工智能系统级模块，结合了激光雷达和视觉感知模块。利用激光雷达、IMU、轮速计等多传感器融合技术，AMR 能够在室内大面积且复杂动态环境中进行实时高精度建图和定位。此外，其自有调度系统能够对多种类型的 AMR 进行实时集群调度，同时通过实时回环检测自动更新局部子地图，以适应室内动态多变的环境。为了提高环境适应能力和安全性，比亚迪电子的 AMR 还应用了人工智能自动驾驶技术。

风险提示：EAI 发展不及预期；具身智能下游需求不足。

## 5. 投资建议

我们认为具身智能（EAI）是推动通用人工智能（AGI）发展的关键技术，具有广阔的应用前景。其技术涵盖计算机视觉、自然语言处理和机器人技术，尤其在具身感知和交互方面表现突出。基于多模态大模型和世界模型的具身智能体，不仅能够理解并适应复杂的物理和虚拟环境，还能在动态场景中与人类自然互动，执行复杂任务。根据证券时报援引赛迪顾问发布的《2024 中国人形机器人产业生态发展研究》显示，全球具身智能市场将在未来几年持续增长，尤其是在工业自动化和服务机器人领域。此外，随着智能制造和人形机器人技术的快速发展，预计到 2026 年，中国人形机器人市场规模将突破 200 亿元。各地政策的积极推动，如上海和北京的智能机器人发展计划，将进一步加速具身智能的技术进步和市场应用。

## 6. 风险提示

1.大模型发展不及预期；2. AI 智能体发展不及预期；3.智能体下游需求不足。

## 信息披露

### 分析师声明

杨林 计算机行业  
杨蒙 计算机行业

本人具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的职业态度，独立、客观地出具本报告。本报告所采用的数据和信息均来自市场公开信息，本人不保证该等信息的准确性或完整性。分析逻辑基于作者的职业理解，清晰准确地反映了作者的研究观点，结论不受任何第三方的授意或影响，特此声明。

### 分析师负责的股票研究范围

重点研究上市公司： 易华录,润和软件,神州控股,创业慧康,航天宏图,狄耐克,京北方,永信至诚,,广联达,纳思达,用友网络,云赛智联,联想集团,商汤-W,思维列控,锐明技术,中控技术,拉卡拉,能科科技,税友股份,星环科技-U,国能日新,海光信息,\*ST 银江,拓尔思,鼎捷软件,中科曙光,金山办公,道通科技

### 投资评级说明

1. 投资评级的比较和评级标准:	类 别	评 级	说 明
以报告发布后的 6 个月内的市场表现为比较标准, 报告发布日后 6 个月内的公司股价 (或行业指数) 的涨跌幅相对同期市场基准指数的涨跌幅;	股票投资评级	优于大市	预期个股相对基准指数涨幅在 10%以上;
		中性	预期个股相对基准指数涨幅介于-10%与 10%之间;
		弱于大市	预期个股相对基准指数涨幅低于-10%及以下;
		无评级	对于个股未来 6 个月市场表现与基准指数相比无明确观点。
2. 市场基准指数的比较标准:	行业投资评级	优于大市	预期行业整体回报高于基准指数整体水平 10%以上;
A 股市场以海通综指为基准; 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。		中性	预期行业整体回报介于基准指数整体水平-10%与 10%之间;
		弱于大市	预期行业整体回报低于基准指数整体水平-10%以下。

### 法律声明

本报告仅供海通证券股份有限公司 (以下简称“本公司”) 的客户使用。本公司不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。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。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断, 本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。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。

市场有风险, 投资需谨慎。本报告所载的信息、材料及结论只提供特定客户作参考, 不构成投资建议, 也没有考虑到个别客户特殊的投资目标、财务状况或需要。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。在法律许可的情况下, 海通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。

本报告仅向特定客户传送, 未经海通证券研究所书面授权, 本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品, 或再次分发给任何其他人, 或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。如欲引用或转载本文内容, 务必联络海通证券研究所并获得许可, 并需注明出处为海通证券研究所, 且不得对本文进行有悖原意的引用和删改。

根据中国证监会核发的经营证券业务许可, 海通证券股份有限公司的经营经营范围包括证券投资咨询业务。