



计算机行业研究

买入（维持评级）

行业深度研究

证券研究报告

计算机组

分析师：刘高畅（执业 S1130525120005） liugaochang@gjzq.com.cn
 分析师：郑元昊（执业 S1130525120004） zhengyuanhao@gjzq.com.cn
 联系人：孙恺祈 sunkaiqi@gjzq.com.cn

具身智能逼近临界点，人形机器人商业化有望揭开序幕

行业观点

- **从央视春晚 Bot 火热看人形机器人四大潜在商业化场景。**马年央视春晚“含 Bot”量再度提升，松延动力、宇树科技、魔法原子、银河通用&Sharpa 为全国人民贡献了四台兼顾科技属性与人文关怀的精彩节目，也展示出人形机器人产业面向商业化落地的最新尝试。以星海图前首席科学家许华哲博士的思考为蓝本，我们认为人形机器人所面向的工作环境必然是“高价值+低重复性”的场景，上述场景又会在 ToB/C 客群差异、大脑/小脑层面泛化性要求的区分下形成四类买单方画像清晰且现阶段具身智能足以胜任的商业化场景——1) ToB+大脑泛化：导览、导购、导巡场景，比如小鹏 Iron 机器人的展厅导览、汽车门店导购、公司前台导流与巡逻场景，底层能力是流利流畅的语言智能沟通+亲切的人形姿态。2) ToB+小脑泛化：特种行业场景，比如央视春晚《武 Bot》节目中宇树科技在高动态集群控制下，实现 16 台机器人毫秒级协同响应的醉拳、后空翻、双节棍、跳马、队列穿插变阵等高难度动作的能力，底层能力是类人的运动控制+蜂群协同作战。3) ToC+大脑泛化：个人陪伴场景，比如央视春晚《奶奶的最爱》节目中，松延动力机器人展示出的情绪价值拉满的唠嗑、玩梗能力，底层能力是基于语言智能的情绪价值供给，用户数据、情感投射、品牌心智形成正向飞轮。4) ToC+小脑泛化：家庭照护场景，比如央视春晚贺岁微电影中，银河通用机器人使用 Sharpa 的灵巧手完成盘核桃、叠衣服等任务，底层能力是触视觉协同的长程精细操作能力。
- **模型：具身智能泛化的曙光。**泛化是机器人区别于自动化机器设备的本质区别，也是解锁上述商业化场景的钥匙。1) 一段式学习系统取得阶段性突破。26 年初，Figure AI 发布 Helix 02，新增运动直觉 S0 系统，能够在整个房间范围内实现“灵巧、长时程的自主行为”；同期，特斯拉选择将 Optimus 与无人驾驶共享一套闭环仿真世界模型，并希望将 Grok 大模型作为 Optimus 机器人集群的高级调度中枢。2) 大脑侧模型的预训练已跨越 ScalingLaw 门槛。25 年 11 月，北美创企 Generalist AI 发布 GEN-0，首次验证了激活具身智能的参数门槛大约在 7B。我们认为，由于 Scaling 路径已基本被 LLM 摸清，所以具身智能大脑侧的卡点更多是一个数据与工程问题而非复杂的科学问题。3) 小脑侧，25 年 9 月底北美创企 Skild AI 发布通用小脑，在相对于传统单一模型训练量提升 500 倍的情况下意外训练得到了能够跨本体实现运动控制的通用小脑能力——比如在被锯断了小腿的四足机器狗、锁定了一个电机关节只剩下三条腿的四足机器狗以及锁定了轮足的机器狗上，通用小脑均展现出快速的泛化适应能力。同期，李飞飞博士在发布 ImageNet 的十五年后启动首届 BEHAVIOR 挑战赛，开放总计 1,200+小时的 10,000 条专家遥操作示范，细致标注物体状态、空间关系和自然语言描述。ImageNet 曾推动计算机视觉进入大模型时代，我们相信 BEHAVIOR 也将为机器人时代设定第一个统一的 Benchmark，推动大小脑协同进化。
- **数据：三线交织，虚实融合，加速演绎。**我们认为，具身训练中，真机数据是刚需，而仿真与人类视频可以提供数据扩展必要的数量级补充，以上三条路线均在过去一年中取得突破性进展。1) 真机数据：UMI 作为一种轻量化、无本体、低成本的真实数采方案，由 3D 打印的平行夹爪+软指+扳机构成，整套成本约 400 美金。2) 仿真合成数据：英伟达 Isaac&Sim 与索辰开物机器人训练平台持续迭代，北美 Skild AI+国内银河通用都证明足够多样性的仿真数据可以一定程度上弥补 Sim2Real 的 gap 问题。3) 人类视频数据：Gemini 3.0/3.1 Pro 在屏幕理解、抽象推理层面取得显著能力提升，有望加速解锁机器人理解真实世界的的能力，北美 Tesla 与 Figure 都在转向人类视频数据进行模型的训练。
- **相关标的：**智能逼近临界点有望推动人形机器人进入大规模量产前奏，建议关注人形机器人核心供应链企业，如斯菱智驱、科森科技、福赛科技、新泉股份、三花智控、恒立液压、泛亚微透、唯科科技、领益智造、蓝思科技、海康威视、大华股份等；具身智能 AI 大脑和世界模型，建议关注协创数据（旗下奥佳科技 Fcloud 与 NV 合作提供 Ominibot 具身智能大脑训练平台）、索辰科技（中国物理 AI 稀缺资产）、群核科技、智微智能等。

风险提示

- 北美人形机器人量产节奏不及预期的风险；通用机器人 Day1 L4 路线缺乏商业化基础的风险；仿真合成数据质量不及预期的风险；模型及软件解决方案三方公司长期产业链话语权较低的风险。



内容目录

1. 引子：从央视春晚 Bot 火热看人形机器人商业化提速.....	4
2. 模型：具身智能泛化的曙光，一段式学习系统取得突破.....	6
2.1 端到端：Figure AI 布局全身推理系统，特斯拉 Optimus 与无人驾驶共享一套世界模型.....	6
2.2 大脑：预训练跨越 Scaling Law 门槛，激活参数或为 7B	8
2.2 小脑：跨本体通用小脑问世，Behavior 牵动具身智能 ImageNet 时刻.....	9
3. 数据：三线交织，虚实融合，加速演绎.....	10
3.1 真实数采：刚需，UMI 解锁轻量化、无本体、低成本方案.....	10
3.2 仿真合成：增广，不断逼近真实世界的性价比方案.....	11
3.3 人类视频：潜力，AI 多模态能力突飞猛进.....	18
4. 风险提示.....	19

图表目录

图表 1：春晚“含 Bot”量浓度升高，能力展示程度逐步深化.....	4
图表 2：人形机器人核心商业价值在于处理“高价值/低重复”任务.....	5
图表 3：对高价值/低重复场景的进一步拆解，ToB/C×动脑/动手.....	6
图表 4：Helix 02：一个能够同时对全身进行推理的单一学习系统.....	7
图表 5：特斯拉 Optimus 机器人与无人驾驶共享一套世界仿真模型.....	8
图表 6：激活具身智能模型的参数门槛或许是 7B	8
图表 7：2023 年 GPT-4 相对 3.5 扩展了 10X 参数量.....	9
图表 8：2025 年上下文窗口的 10X 扩容.....	9
图表 9：推理相对非推理解锁了 10X Token 消耗.....	9
图表 10：Skild AI 通用小脑使机器人可实现跨本体运动控制.....	10
图表 11：Behavior 为机器人提供日常家庭长时序任务综合仿真基准与训练环境.....	10
图表 12：UMI 设备由平行夹具+软指+扳机构成.....	10
图表 13：多家具身智能大模型所用训练数据引入仿真合成数据.....	11
图表 14：银河通用 GraspVLA 大模型工作展示中，在不同光照条件下，模型都能精准执行抓取任务.....	12
图表 15：NVIDIA 机器人三大计算平台协同解决方案.....	13
图表 16：三台计算机之间闭环 workflow 与数据协同.....	13
图表 17：Omniverse 验证机器人步态，支持机器人学习.....	13
图表 18：机器人在 Isaac Sim 中的不同仿真环境下训练.....	13
图表 19：NVIDIA Isaac Sim 生态系统合作伙伴遍布海内外.....	14
图表 20：Omniverse 免费+订阅模式促进生态建设，形成软件驱动硬件销售模式.....	14
图表 21：协创数据旗下 OmniBot 平台基于 NV Cosmos 搭建具身智能开发服务平台.....	15



图表 22: Omnibot 输入视频	15
图表 23: cosmos transfer1 模型输入视频	15
图表 24: 索辰开物平台涵盖四层架构	16
图表 25: 索辰开物平台实现设计仿真训练优化一体化界面	16
图表 26: 群核科技 Spatial Verse 空间智能平台基于海量室内 3D 设计数据, 提供高保真仿真合成数据	17
图表 27: 谷歌发布通用世界模型 Genie 3 支持生成多样化交互环境	18
图表 28: LingBot-World 在高动态环境下仍能保持主体一致性	18
图表 29: Gemini 3.1 Pro 实现抽象推理能力翻倍跃升	18
图表 30: Gemini 3.1 Pro 视觉表现力实现显著提升	18



1. 引子：从央视春晚 Bot 火热看人形机器人商业化提速

春晚作为国内前沿科技商业化落地的集中展示窗口，其历年展示的机器人形态变化反映了行业从预编程执行向具身智能迈进的发展轨迹。回顾历年春晚，从 1996 年科幻小品中的服务概念萌芽，到哈工大与优必选展示的小型机器人阵列，再到优必选 Walker 与宇树四足机器人的双足稳定行走与动态起舞，早期阶段仍以预编程序的机械同步与基础动态平衡为主，整体处于运动控制与硬件底座的打磨阶段。2025 年成为关键技术拐点，宇树 H1 完成人机协同、手绢抛接等动作，标志着国产机器人已跨越复杂步态与基础精细力控门槛。2026 年实机展示进一步体现出多场景商业化能力跃迁：宇树 G1/H2 集群与魔法原子产品集中验证了低延迟协同与极限动态抗冲击控制；银河通用 GalbotG1 凭借灵巧手完成盘核桃、叠衣服等柔性物料处理，攻克了复杂物理环境下的高精度力控难点；松延动力 E1/N2 则依托大模型实现了自然语义交互与情绪陪伴。上述能力集中突破，印证了大模型大脑与通用运动小脑的融合已初见成效，标志着人形机器人正式具备了向真实复杂商业场景渗透的物理与智能基础。

图表1：春晚“含Bot”量浓度升高，能力展示程度逐步深化

年份	节目名称	出演机器人 (厂商、型号)	展示的能力
1996	小品《机器人趣话》	演员蔡明扮演，非物理机器人	演员以科幻喜剧的方式模拟服务型机器人的“温柔型”、“奔放型”等性格模式切换，以及做饭、跳舞等家政服务。
2012	儿童节目《机器人总动员》	哈尔滨工业大学创新基地 - 小型机器人	简单的机械关节动作控制，随着音乐节拍进行基础的集体舞蹈动作。
2016	歌曲《冲向巅峰》	优必选 - Alpha 1S (小型人形机器人)	540 台机器人通过高精度控制实现毫秒级同步，集体展示倒立、劈叉、整齐划一的舞蹈等动作。
2018	歌舞《狗年旺旺》	优必选 - Jimu 汪汪 (编程机器狗)	大规模阵列变型、集体舞蹈、精确走位，模拟宠物小狗的拜年姿态。
2019	歌舞《青春畅想》	优必选 - Walker (大型双足人形机器人)	双足直立稳定行走、伴舞、随音乐挥手互动、双手抓取并持有道具进行表演。
2021	创意表演《牛起来》	宇树科技 - 犇犇 (四足机器人)	灵活奔跑、跳跃、后空翻、随节奏起舞。
		优必选 - 拓荒牛 (大型四足机器人)	负重稳定行走、抬头互动。
		ABB - 工业机械臂	执笔写书法、配合演员精准舞蹈。
2024	舞台舞美支持	新松机器人 - 工业机器人	作为幕后与边缘支持，举重若轻地支撑大型屏幕显示与灯光设备，完成高精度的舞台动态调度。
2025	舞蹈《秧BOT》	宇树科技 - H1 (人形机器人)	人机协同跳东北秧歌。展示双足复杂步态行走、手臂灵活转动手绢、高精度协同甚至将手绢抛空中的精细力控与平衡力。
2026	武术《武BOT》	宇树科技 - G1/H2 (人形机器人)	抗冲击力与动态控制：20 多台机器人低延迟、高精度协同，完成连续跑酷、弹射空翻、打醉拳、舞剑、耍双节棍等高难度赛博功夫。
	歌曲《智造未来》及其他特技节目	魔法原子 - MagicBot Z1 / Gen1 (人形机器人)	高难度特技与交互：MagicBot Z1 展示“托马斯 360°街舞特技；Gen1 进行手势互动。
	贺岁微电影《我最难忘的今宵》	银河通用 - Galbot G1 (人形机器人)	家庭照护精细力控：搭载 Sharpa 灵巧手，展示应对柔性物料及质地较软物品的高难度处理能力，如盘核桃、叠衣服、拿起水杯等。
	小品《奶奶的最爱》	松延动力 - E1/N2 (双足人形机器人/仿生人形机器人)	互动陪伴与情绪价值：在小品中与演员（如蔡明）进行实时大模型语音交互，展示翻跟头、头部伸缩等特技。

来源：界面文化，中经 e 商圈，新京报，投中网，国金证券研究所

人形机器人的商业化价值几乎必然面向“高价值/低重复”任务。我们将机器人面向的商业场景抽象为四大象限。第一象限（高重复、高价值）是典型的专用设备领域，如汽车组装车间的机械臂或医疗领域的达芬奇手术机器人，客户愿意为特定高价值任务买单，但设备泛化性弱；第二象限（高重复、低价值）多由通用型自动化设备主导，如仓储 AGV 或家用扫地机，以标准化方案解决低附加值劳动；我们认为，人形机器人由于初期研发成本较高，其必须面向具有高额支付意愿的“高价值”领域；同时，其通用形态与泛化智能的溢价，只有在面对非标准化、



高度复杂的“低重复性”任务时才能得到真正释放。因此，人形机器人真正的星辰大海在于第四象限（低重复、高价值）。此外，第三象限（低重复、低价值）中长尾且琐碎的日常杂务（如捡拾零散纸团），目前并不具备独立的商业化土壤，我们预计这类场景将在第四象限核心技术成熟后，受益于技术外溢被顺带解决。

图表2：人形机器人核心商业价值在于处理“高价值/低重复”任务



来源：国金证券研究所

在“高价值、低重复”的核心市场中，基于机器人能力侧重点（大脑思考泛化/小脑动作泛化）与客群属性（ToB/ToC）的进一步解耦，人形机器人的商业化落地可划分为四大细分场景。

- 其一，多动脑×ToB端的营销服务场景，主要聚焦商业环境下的逻辑推理与信息交互。该场景对机器人物理运动能力要求较低，仅需完成平地行走、手势引导等基础动作，却对模型“大脑”的动态认知能力要求颇高。小鹏汽车全新人形机器人IRON率先落地自有门店“导览、导购、导巡”核心商业场景，在展厅动态复杂环境中，可实时调用VLM/MLA架构的底层物理大模型，除完成引路导览外，还能精准解析客户非标准化需求、自主开展产品答疑，并在多轮对话中挖掘客户痛点、输出定制化营销话术，核心商业价值体现在提升门店获客转化率、优化终端运营效率、实现服务人力成本替代上。
- 其二，多动脑×ToC端的个人陪伴场景，核心在于提供情绪价值与非结构化对话。该场景对机器人的物理动作复杂度要求有限，基础的肢体回应与表情反馈即可满足需求，但高度依赖大语言模型的语义共情、意图识别与长程多轮交互能力。以春晚小品中亮相的松延动力机器人为例，其面向老年群体的陪伴功能颇具代表性，不仅能流畅完成日常聊天、兴趣互动等对话任务，还可实时感知情绪变化并给予针对性回应。该场景核心是通过高频次的情感交互建立深度用户粘性，将机器人转化为家庭中的智能陪伴终端，从而挖掘长期的服务价值。
- 其三，多动手×ToB端的特种行业场景，对机器人的动态平衡、抗冲击力及全地形机动能力提出高要求。以宇树科技在2026年春晚《武Bot》节目中的表现为例，其技术底座已具备向特种场景落地的清晰可行性：一是具备较强的高动态运动控制能力，机器人在连



续后空翻、武术动作中展现的姿态控制、落地缓冲与动态稳定能力，可为复杂地形通行、应急越障、高危环境作业提供运动基础；二是具备人机协同与精细力控能力，可完成精准器械操作、高速动作协同，并与人员安全无碰撞交互，感知与操控精度已满足协同作业、制式装备操作等场景的硬件条件；三是具备集群协同作业能力，20 余台机器人在直播环境下完成高速编队与队形变换，验证了低延迟通信、路径规划与避障算法的成熟度，形成了多机协同作业的技术闭环。更为关键的是，依托全栈自研，宇树将单机成本压降至 10 万元以内，而海外竞品波士顿动力 Spot 的售价约 53 万元且未实现规模化量产，其成本优势为规模化部署提供了商业基础，其早期产品甚至曾获美国海军陆战队的批量采购用以侦察运输。这种兼具高机动性能、集群协同潜力与低成本量产能力的产业化进展，已引发战略层面关注。2026 年 2 月 14 日，美国国防部将宇树科技列入“1260H”清单，实施投资与采购限制。我们认为，此类管制措施从侧面印证了高机动性小脑机器人在特种行业及关键领域的战略价值与应用潜力。

- 其四，多动手×ToC 端的家庭照护场景，需要应对高度非结构化且随时变动的家庭物理环境。该场景需要机器人具备稳定的底盘移动能力，以适配居家多样的空间布局，同时要求针对各类柔性物料的精细力控与灵巧操作能力，精准把控施力幅度与动作精度。如银河通用机器人搭载 Sharpa 灵巧手，可灵活完成盘核桃、折叠衣物、平稳端水杯等精细化操作，精准适配居家日常各类操作需求，以此高效替代人类从事兼具精细度与实用性的高价值家政劳动，释放家庭照护中的人力成本。

图表3: 对高价值/低重复场景的进一步拆解, ToB/C × 动脑/动手



来源：新京报，NE 时代智能体，国金证券研究所

2. 模型：具身智能泛化的曙光，一段式学习系统取得突破

2.1 端到端：Figure AI 布局全身推理系统，特斯拉 Optimus 与无人驾驶共享一套世界模型

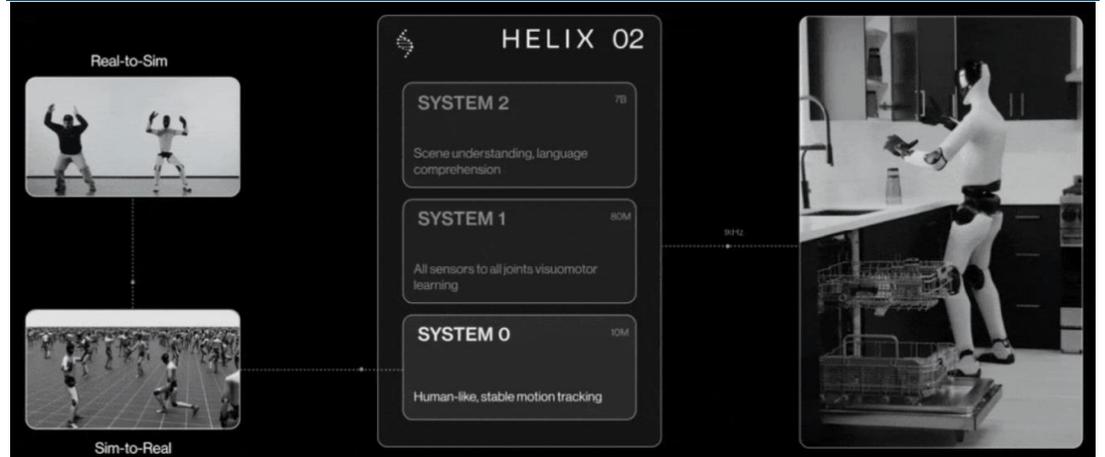
2026 年初，Figure AI 发布 Helix 02，采用了单一的神经系统，新增 S0 系统，能够在整个房间范围内实现“灵巧、长时程的自主行为”：

- S2（大脑）负责理解复杂的指令，缓慢地进行目标推理：解读场景、理解语言并规划行为序列。
- S1（脊髓）负责敏捷的动作执行，快速思考，以 200 Hz 的频率将感知转化为全身关节的目标指令。
- S0（小脑）负责稳健的身体控制，以 1 kHz 的频率执行，负责全身的平衡、接触与协调。



新增的 S0 系统为机器人提供“运动直觉”，它是一个用于机器人全身控制的基础模型，是 Helix 02 物理具身的核心基础。当更高层级负责任务推理与规划时，S0 确保每个动作都能平稳、安全且稳定地执行。S0 并非为行走、转向、蹲下或伸手等动作分别设计奖励函数，而是直接从大量且多样化的运动数据中学习追踪人体动作。Figure 投喂了超过 1,000 小时的关节级重定向人类运动数据，而 S0 则是在不断的仿真模拟中，模仿这些人类的动作。

图表4: Helix 02: 一个能够同时对全身进行推理的单一学习系统



来源: 新智元, 国金证券研究所

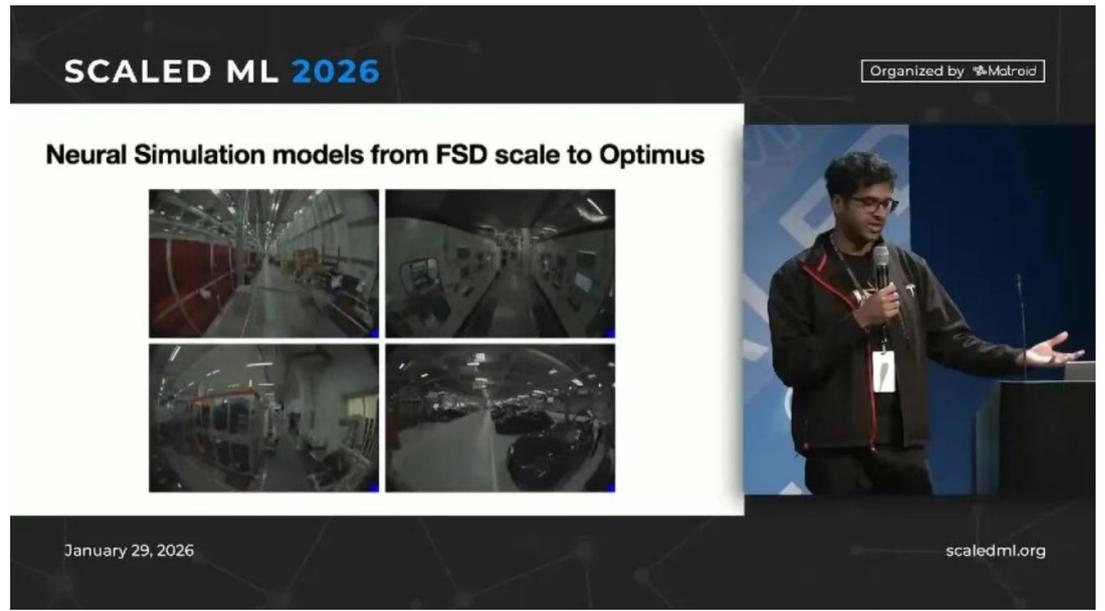
在 2026 年 Scaled ML 大会上，特斯拉人工智能软件副总裁阿肖克·埃卢斯瓦米发表主题演讲，提及将奥斯汀上线的无人驾驶 robotaxi 服务与即将量产的 Optimus 串联起来。在此背景下，特斯拉研发出世界仿真模型——这一模型基于状态与动作的配对数据训练而成，只需输入当前的摄像头画面和执行动作，就能精准生成摄像头的下一帧画面，构建出高度逼真的虚拟场景。当世界仿真模型达到预期效果后，便可与策略神经网络对接，形成闭环仿真系统：模型生成后续视频帧，策略神经网络依据画面规划下一步动作，动作再反馈至模型进行迭代，最终构建出连贯、多样的仿真环境。值得注意的是，这款世界仿真神经网络并非专为自动驾驶设计，由于采用通用数据训练，它同样可完美适配 Optimus 人形机器人。

目前，Optimus 能在这一虚拟环境中流畅完成移动、操作等任务，而画面中的所有像素均由模型生成——借助这一工具，机器人智能体可在精准还原的极端场景及其各类变体场景中反复测试，大幅降低现实测试的成本与风险。马斯克透露，特斯拉正基于这一模型，打造一座“Optimus 学院”，未来将让数万台机器人在这一实体训练基地中，通过现实场景的“自训练”掌握各类任务技能。

此外，为了打通“仿真到现实的鸿沟”，特斯拉计划在现实场景中部署 1 万至 3 万台 Optimus 机器人，一方面用于验证虚拟训练的效果，另一方面通过采集现实场景中的海量数据，持续优化虚拟训练所使用的“现实生成器”，实现虚拟与现实的双向迭代。同时，马斯克也提到 xAI 研发的 Grok 大模型，或将成为 Optimus 机器人集群的高级调度中枢。



图表5: 特斯拉 Optimus 机器人与无人驾驶共享一套世界仿真模型

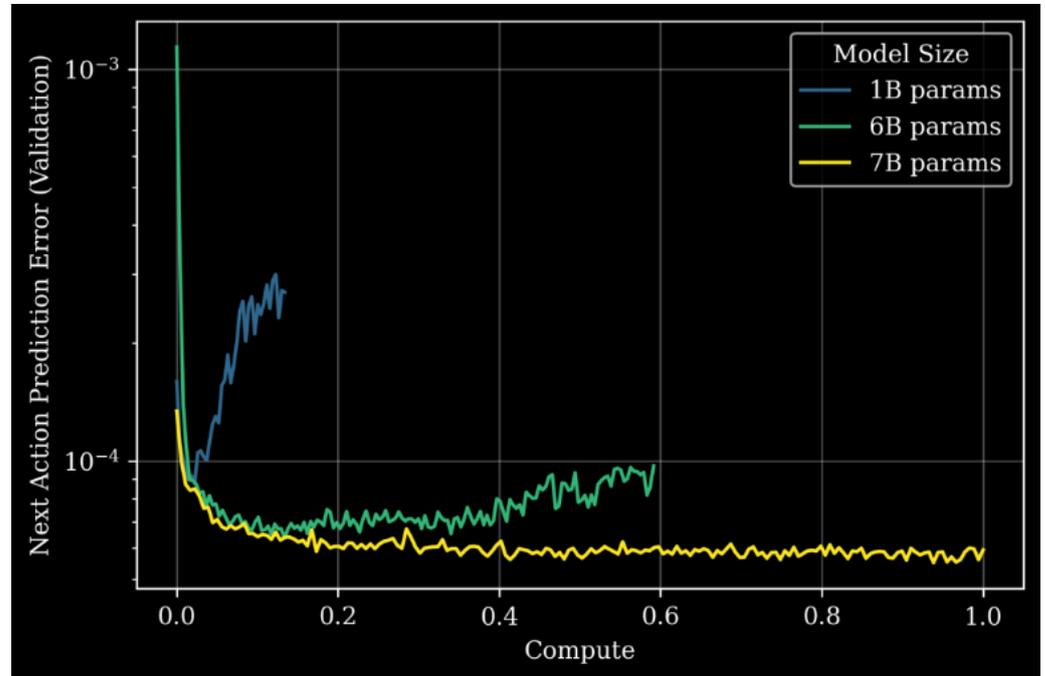


来源: NE 时代智能体, 国金证券研究所

2.2 大脑: 预训练跨越 Scaling Law 门槛, 激活参数或为 7B

具身智能大脑模型的预训练阶段正加速跨越 Scaling Law 的激活门槛, 行业面临的卡点已从科学探索转向工程与数据堆叠。2025 年 11 月, 北美初创企业 Generalist AI 发布 GEN-0 模型, 首次在业内验证了具身智能模型存在类似大语言模型 (LLM) 的参数激活现象, 且其门槛当量约为 7B, 而当前市面现有模型参数普遍停留在 4B 左右。

图表6: 激活具身智能模型的参数门槛或许是 7B



来源: Generalist AI, 国金证券研究所

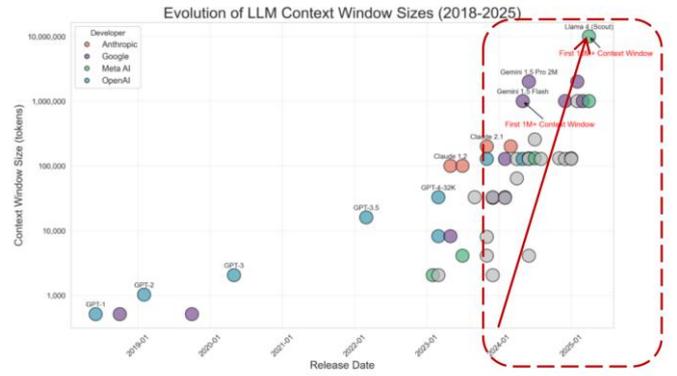
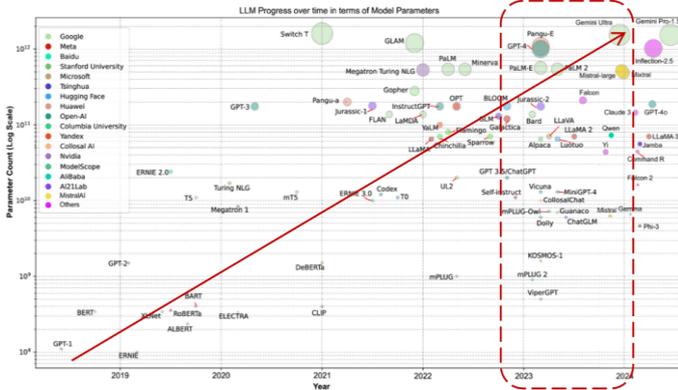
复盘 LLM 过去 3 年的演进轨迹, 每一年都是在特定技术维度上通过资源堆砌的暴力美学实现了 10 倍级以上的跃升。2023 年是预训练的 Scaling Law, 以 GPT-4 为代表, 行业将模型参数量从千亿级拉升至万亿级规模 (第一个 10 倍), 成功将海量互联网数据压缩进权重, 彻底解锁了模型的通用常识底座; 2024 年是后训练的 Scaling Law, 以 DeepSeekR1 及 GPT-o 为代表, 通过强化学习与思维链技术解锁了深度逻辑推理能力, 同一模型在响应复杂任务时的推理 Token 消耗量是单纯回答问题的 10 倍以上 (第二个 10 倍); 2025 年则是存储的 Scaling Law,



随着 Gemini3 等新一代模型的面世，通过堆高存储硬件与架构优化，将模型的上下文窗口长度从 100k 拉升至 1M+甚至 10M+级别（第三个 10 倍），解锁了模型的长期记忆与海量信息吞吐能力。

图表7: 2023 年 GPT-4 相对 3.5 扩展了 10X 参数量

图表8: 2025 年上下文窗口的 10X 扩容



来源: 《Survey of different Large Language Model Architectures: Trends, Benchmarks, and Challenges》, 国金证券研究所

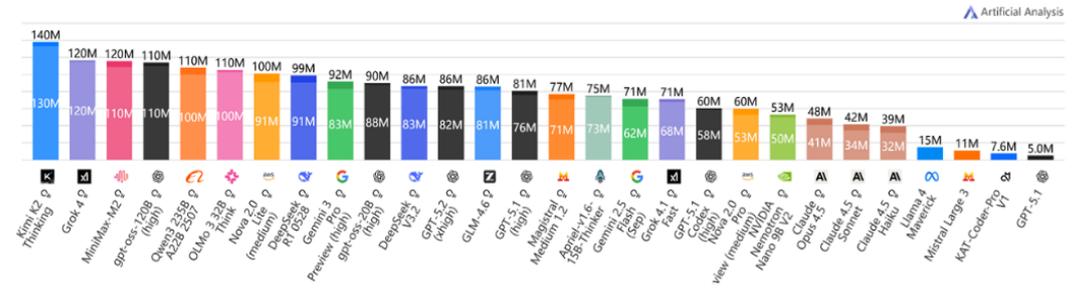
来源: 《Understanding the Impact of Increasing LLM Context Windows》, 国金证券研究所

图表9: 推理相对非推理解锁了 10X Token 消耗

Output Tokens Used to Run Artificial Analysis Intelligence Index

Tokens used to run all evaluations in the Artificial Analysis Intelligence Index

Answer Tokens Reasoning Tokens



来源: Artificial Analysis 官网, 国金证券研究所

我们认为具身智能正处于复刻该指数级跃升的前夕。随着模型跨过 7B 这一幂律定律的激活临界点，后续能力的提升将主要依赖于算力与多维空间数据的暴力美学灌溉。换言之，具身智能大脑的发展路径已相对明朗，其核心矛盾已由复杂的科学算法问题，降维为算力供给与数据获取的工程问题。

2.2 小脑：跨本体通用小脑问世，Behavior 牵动具身智能 ImageNet 时刻

跨本体通用小脑的问世与标杆数据集的开源，标志着机器人运动控制领域逐步迈向范式升级，有望迎来类似计算机视觉领域的“ImageNet 时刻”。

传统机器人运动控制高度依赖特定硬件本体的单一模型训练，泛化性极差。而在 2025 年 9 月，北美创企 Skild AI 突破性地发布了通用小脑模型，在相对于传统模型训练量提升 500 倍的算力加持下，该模型成功实现了跨本体的自适应运动控制，即无论机器狗是双足、四足还是存在肢体受损，同一套模型均能实现高效的自平衡与运动泛化。

Skild AI 的实践验证了大规模训练在机器人控制领域的有效性，但要实现从单点突破到行业规模化复制，亟需统一的数据基座与评价体系。2012 年，ImageNet 以千万级高质量标注图片及统一竞赛标准，为深度学习算法迭代提供了客观依据，成为计算机视觉技术爆发的起点；如今，BEHAVIOR 挑战赛正推动具身智能领域重现这一历程。该挑战赛向业界开放了总计超 1,200 小时、包含 10,000 条专家遥操作示范的高质量数据集，并对物体状态、空间关系及自然语言描述进行精细标注。BEHAVIOR 不仅缓解了机器人训练数据匮乏的痛点，更通过构建首个统一基准测试，为通用小脑泛化能力提供了公平度量标准，将有效推动相关算法快速收敛与迭代。



图表10: Skild AI 通用小脑使机器人可实现跨本体运动控制



来源: skild AI, 国金证券研究所

图表11: Behavior 为机器人提供日常家庭长时序任务综合仿真基准与训练环境



来源: Behavior 官网, 国金证券研究所

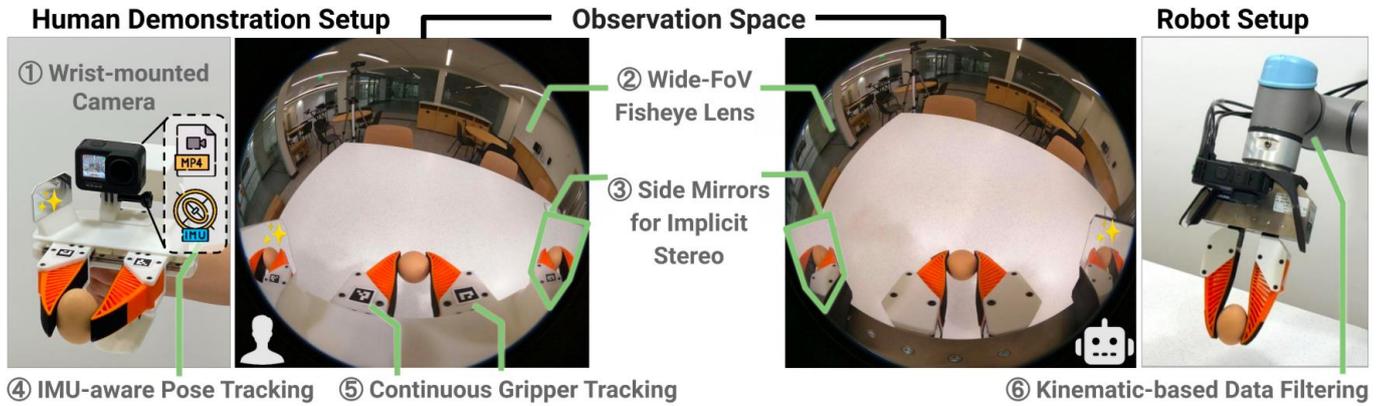
3. 数据：三线交织，虚实融合，加速演绎

3.1 真实数采：刚需，UMI 解锁轻量化、无本体、低成本方案

真实机器采集数据是具身智能迭代的核心刚需，轻量化采集设备的技术突破推动数据获取实现降本提效，UMI 方案则进一步破解传统采集模式的痛点，成为轻量化真实数据采集的关键路径。

UMI (Universal Manipulation Interface) 作为一种新型轻量化真实数据采集方案，其核心在于实现数据采集过程与机器人本体的完全解耦。该系统硬件极为精简，仅由一套手持式 3D 打印平行夹具与腕部视角相机构成，整套设备成本控制在 400 美元左右。采集过程中，相机同步记录相对末端轨迹，而非传统方式中机械臂的绝对关节角度。这种设计将数据采集设备从昂贵的工业机械臂简化为一套“即拿即用”的手持终端，本质上是将机器人操作数据的生产门槛从专业实验室级别降维至消费电子级别。

图表12: UMI 设备由平行夹具+软指+扳机构成



来源: CSDN, 国金证券研究所

相较于传统采集方案，UMI 在采集场景、数据泛化、成本控制三个核心维度实现突破：

- 采集环境从封闭式向开放式拓展，摆脱对数采工厂的依赖，保障数据多样性与采集高效性。传统遥操作采集依赖昂贵的工业机械臂与动捕设备，仅能在封闭实验室或专业数采工厂开展采集工作。而 UMI 凭借手持终端的轻量化优势，无需固定场地及复杂配套设备，采集人员可便捷进入家庭、办公室、户外等各类非结构化真实场景完成采集，显著丰富了数据的场景多样性与真实度，使采集数据更贴合具身智能的实际应用需求。
- 数据泛化能力显著提升，可实现高质量数据的跨设备复用。传统采集模式高度依赖特定机器人的关节角度数据，导致数据通用性差、难以跨构型复用。而 UMI 以腕部第一视角相机作为唯一观测点，避免采集者体型差异带来的视觉误差，同时重点记录末端执行器的轨迹与姿态，而非具体关节参数。这种“重末端、轻本体”的表征方式，使采集数据具备极强的硬件无关性，可无缝迁移至不同构型、不同尺寸的机器人训练中，有效解决了传统数据复用性差的问题，实现数据价值最大化。
- 商业经济性优势突出，可将真实数据的采集成本击穿至消费级区间。传统真机采集属于



重资产投入，高额的场地租金、机械臂折旧及设备运维成本，制约了数据的规模化获取。而 UMI 凭借硬件精简、与机器人本体解耦的设计，有效剥离了高额固定成本。运营成本层面，其部署效率较高，无需搭建固定工装，也无需调试机械臂与动捕系统的联动参数，开机即可启动采集；同时，其采集速度达到传统遥操作模式的 3 倍，可在相同时间内完成更多场景的采集任务，提升单位时间样本产出量。双重优势叠加下，UMI 将采集成本击穿至消费级区间，为具身智能大规模数据采集提供了财务可行性。

3.2 仿真合成：增广，不断逼近真实世界的性价比方案

一般而言，真机数据的优势在于其高保真度（物理交互的真实性），能够精准反映复杂操作中的力学反馈和不确定性，劣势在于采集成本高昂、采集周期长且现阶段数量较少。仿真合成数据的优势在于低成本、高效率与场景的可扩展性，基于物理引擎可快速生成海量的多样化数据，尤其适合预训练数据积累与极端场景的模拟，劣势在于“仿真鸿沟”——虚拟环境与物理现实的偏差导致模型在真实任务中表现不稳定。目前，英伟达、Skild AI、银河通用三家厂商已明确在训练具身智能大模型时采用了仿真合成数据。

图表 13: 多家具身智能大模型所用训练数据引入仿真合成数据

模型名称	厂商	模型类别	模型架构	训练数据类型
RT-2	谷歌 DeepMind	VLA 大模型	单模型架构，以 VLM 模型（PaLI-X 或 PaLM-E）为主干网络	联合微调：大规模互联网数据 + 机器人真机数据
$\pi 0$	Physical Intelligence	VLA 大模型	分层双系统架构：1) 预训练 VLM: 3B 参数的 PaliGemma; 2) 动作专家模块: 300M 参数规模	1) 预训练: -VLM: 大规模互联网数据; -动作专家模块: 开源真机数据集和基于遥操作采集的真机数据; 2) 后训练: 高质量真机数据
Helix	Figure AI	VLA 大模型	分层双系统架构：1) 系统 S1: 80M 参数的 Transformer 模型; 2) 系统 S2: 7B 参数的预训练 VLM 模型	1) 系统 S1: 机器人真机数据; 2) 系统 S2: 大规模互联网数据
GR00T N1	英伟达	VLA 大模型	分层双系统架构：1) 系统 S1: 基于扩散变换器 (DiT) 的动作模块; 2) 系统 S2: 预训练 VLM——Eagle-2	预训练: 真实机器人演示数据、合成数据以及互联网上的人类视频数据
Skild Brain	Skild AI	基于视觉的端到端运动模型	分层架构: 用低频率的高层动作策略为高频率的低层动作策略提供输入, 而且适用于几乎所有机器人。	预训练: 仿真环境与人类操作视频; 后训练: 借助每台联网机器人的真实运行数据进行微调
智元启元大模型 (GO-1)	智元机器人	VLM 基础上发展的 ViLLA 大模型	ViLLA 架构由 VLM(多模态大模型) + MoE(混合专家)组成	AgiBot World 的大规模高质量真机数据+互联网大规模异构视频数据
Seed GR-3	字节跳动	VLA 大模型	采用 Mixture-of-Transformers (MoT) 的网络结构, 把“视觉-语言模块”和“动作生成模块”结合成了一个 40 亿参数的端到端模型	遥操作机器人收集的高质量真机数据+基于 VR 设备的人类轨迹数据+公开、可用的大规模视觉语言数据

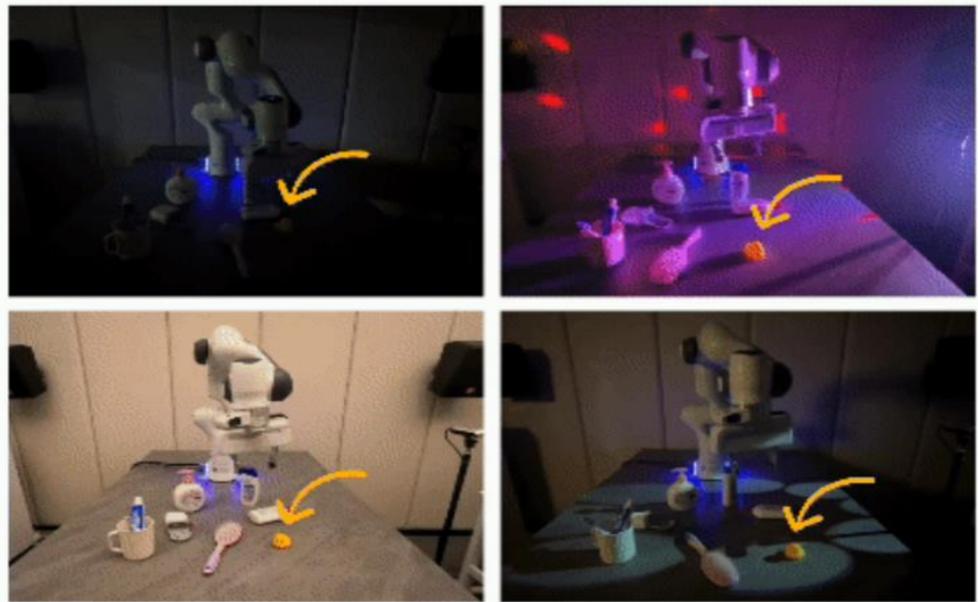


ERA-42	星动纪元	VLA 模型	分层系统架构; 1) 高层次规划: 7B 参数的 InstructBLIP 视觉语言模型; 2) 低层次控制: 40M 参数的 Transformer 架构模型	人类操作数据+互联网机器人数据+遥操作方式获取的数据+大规模视频数据+机器人数据
GraspVLA	银河通用	VLA 模型	\	1) 预训练: 大规模合成数据; 2) 后训练: 少部分真机数据

来源: Physical Intelligence 工作论文《 $\pi 0$: A Vision-Language-Action Flow Model for General Robot Control》, 星动纪元官方公众号, 42 号电波公众号, 智元机器人公众号, 字节跳动 Seed 公众号, 银河通用机器人公众号, 国金证券研究所

真机数据采集路线的典型代表是北美具身智能明星企业 Physical Intelligence (简称 PI), 有报道称 PI 租下了 Airbnb 在旧金山的许多公寓用于真实空间和场景的数据采集, 成本极为高昂。仿真合成数据的代表以银河通用为例, 其基于 NVIDIA Isaac 平台构建高效的数据生产管线, 通过升级物理真实性与渲染并行度, 仅一周即生成全球规模最大的十亿级机器人操作数据集。作为全球首个完全基于仿真合成大数据进行预训练的具身大模型, GraspVLA 展现出比 OpenVLA、 $\pi 0$ 、RT-2、RDT 等模型更强大的泛化能力: 光照泛化、干扰物泛化、平面位置泛化、高度泛化、背景泛化、物体类别泛化。

图表14: 银河通用 GraspVLA 大模型工作展示中, 在不同光照条件下, 模型都能精准执行抓取任务



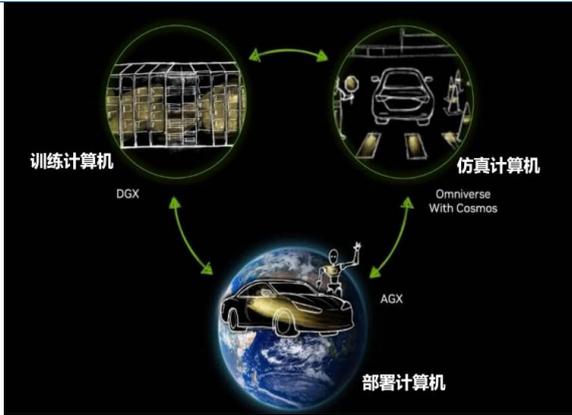
来源: 新智元公众号, 国金证券研究所

NVIDIA CEO 黄仁勋于 2025 年 CES 大会上指出, 每家机器人公司最终都必须构建三台基础计算机协同的解决方案, 形成从训练到优化再到执行的完整体系。对于英伟达而言, “第一台计算机”是 DGX AI 超级计算机, 用于训练。“第二台计算机”是基于 NVIDIA RTX PRO 服务器的 NVIDIA Omniverse 和 Cosmos, 用于生成合成数据、进行强化学习, 同时也是 DGX 和 AGX 的桥梁。“第三台计算机”是 NVIDIA 为边缘计算和自主系统设计的嵌入式平台 NVIDIA Jetson AGX Thor, 用于端侧部署。从训练到仿真再到部署, “三台计算机”打造了英伟达在通用机器人完整框架, 可以缩短人形机器人的开发周期、降低开发成本和风险, 构建开放的生态系统。

第一步, 在“第一台计算机”上训练模型, 开发者可以在 NVIDIA DGX 平台上使用 NVIDIA NeMo 来训练和微调 AI 模型, 还可以利用 NVIDIA Project GR00T 使人形机器人能够理解自然语言, 并能够通过观察人类行为来模仿动作。第二步, 开发者可利用 Omniverse 和 Cosmos 生成海量符合物理特性的多样化合成数据, 例如 2D 或 3D 图像、分割掩码、深度图或运动轨迹数据, 为模型训练和性能优化奠定基础; 开发者在基于 Omniverse 构建的 NVIDIA Isaac Sim 的仿真环境中零风险验证其机器人策略; 还可以使用“第二台计算机”的 Isaac Lab (一个开源的机器人学习框架), 赋能机器人的强化学习和模仿学习, 加速完善机器人的训练策略。第三步, 训练好的 AI 模型被部署到端侧计算机上。

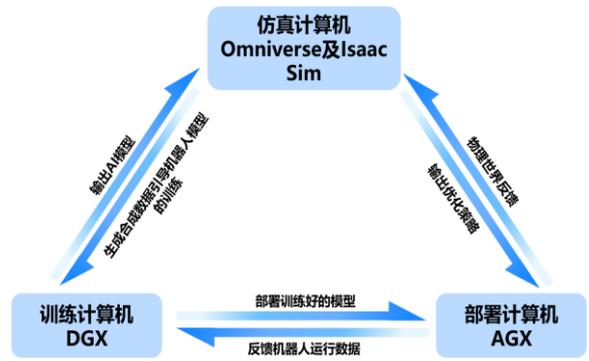


图表15: NVIDIA 机器人三大计算平台协同解决方案



来源: NVIDIA 官网博客, 国金证券研究所

图表16: 三台计算机之间闭环 workflow 与数据协同



来源: NVIDIA 官网博客, 国金证券研究所

近年来, 英伟达不断强调旗下图形部门中仿真条线的重要性, 即“第二台计算机”的重要性。美国科技媒体《The Information》深入报道了英伟达创始人兼 CEO 黄仁勋子女在该公司任职的情况。报道指出, 黄仁勋现年 34 岁的女儿黄敏珊 (Madison Huang) 于 2020 年加入英伟达, 现任 Omniverse 和机器人技术领域的资深产品营销主管。其 35 岁的儿子黄胜斌 (Spencer Huang) 于 2022 年加入英伟达, 目前担任机器人项目的产品经理。

黄仁勋的管理理念与传统企业不同, 他相信企业架构的扁平化结构与信息的极度透明可以提高企业运转效率, 因此公司高管直接向黄仁勋汇报工作。仿真技术副总裁和英伟达其他重要业务部门副总裁 (如 GPU 工程高级副总裁、DGX Cloud 业务副总裁等) 一并直接向黄仁勋汇报工作。从黄仁勋家族成员在公司的角色配置和岗位安排, 再到仿真技术副总裁的直接汇报, 不难看出黄仁勋本人对于英伟达仿真业务的重视程度之深。

Omniverse 是一个用于大规模构建和运行 3D 应用程序及服务的开发平台, 该平台基于 OpenUSD (通用场景描述) 和 NVIDIA RTX 渲染技术, 支持实时协作、物理精准仿真和生成式 AI 集成, 主要应用于工业数字化、机器人训练、自动驾驶仿真和元宇宙等领域, 已成为工业设计、数字孪生、AI 开发等领域的核心工具。

NVIDIA Isaac Sim 是一款基于 Omniverse 构建的开源应用, 使开发者能够在基于物理的虚拟环境中模拟和测试 AI 机器人解决方案。Isaac Sim 有以下三大功能:

- 生成合成数据: Isaac Sim 支持大规模合成数据生成, 包括感知、移动、基于物理的抓取等, 并提供写实渲染和自动生成真实标签, 用于训练和微调机器人基础模型。
- 执行软件在环测试: Isaac Sim 可通过与真实机器人软件集成, 为完整的机器人堆栈实现软件在环 (software-in-the-loop) 测试, 从而验证机器人控制与感知系统。
- 支持机器人学习: Isaac Lab 是基于 Isaac Sim 平台构建的开源轻量级应用, 专为大规模机器人学习进行优化。通过 Isaac Lab 支持机器人学习流程, 可加速仿真中的训练, 助力模型在现实场景中的快速部署。

图表17: Omniverse 验证机器人步态, 支持机器人学习



来源: NVIDIA 英伟达企业解决方案公众号, 国金证券研究所

图表18: 机器人在 Isaac Sim 中的不同仿真环境下训练



来源: NVIDIA 英伟达企业解决方案公众号, 国金证券研究所

目前, Agility Robotics、波士顿动力、傅利叶、Mentee Robotics、Neura Robotics 和小鹏机器人等公司正在使用 Isaac Sim 和 Isaac Lab 对其人形机器人进行仿真和验证。Skild AI 正在



使用该仿真框架开发通用机器人智能，General Robotics 正在将其集成到其机器人仿真平台中。此外，中国台湾的电子和机器人制造商，比如威刚科技 (Adata)、研华科技 (Advantech)、台达电子 (Delta Electronics)、Foxconn、Foxlink、所罗门 (Solomon)、达明机器人 (Techman) 和纬创 (Wistron) 同样也在使用 Isaac Sim 和 Isaac Lab 开发下一代 AI 机器人。

图表19: NVIDIA Isaac Sim 生态系统合作伙伴遍布海内外



来源: NVIDIA 官方公众号, 国金证券研究所

目前，个人创作者、设计师和开发者可免费使用 NVIDIA Omniverse 的基础版本，目的是降低设计技术门槛，吸引用户参与生态建设，通过培养用户群体，推动内容创作和技术创新，进而间接带动企业端需求增长。针对企业用户，NVIDIA Omniverse 专门推出了 Omniverse Enterprise 软件，企业用户可拥有 90 天试用期，试用期过后的订阅费用为每 GPU 每年 4,500 美元。但该软件试用版要求拥有搭载 NVIDIA RTX 的工作站或服务器(官网建议要求具有 16GB VRAM 的支持 RTX 的 GPU)。这种销售模式将在一定程度上带动硬件产品 (GPU 等) 销量增长，形成“软件驱动硬件”的协同效应。

图表20: Omniverse 免费+订阅模式促进生态建设，形成软件驱动硬件销售模式

客户类型	使用版本	付费规则	目的
C 端	NVIDIA Omniverse 基础版本	免费使用	吸引用户参与生态建设，推动内容创作和技术创新，进而间接带动企业端需求增长
B 端	Omniverse Enterprise 软件	客户拥有 90 天试用期，试用期过后的订阅费用为每 GPU 每年 4,500 美元	软件试用版要求拥有搭载 NVIDIA RTX 的工作站或服务器，进而带动硬件产品销量增长，形成“软件驱动硬件”的协同效应

来源: NVIDIA 官网, 国金证券研究所

2025 年 7 月，协创 Fcloud Omnibot 平台与 NVIDIA Cosmos Transfer 模型的集成为机器人训练提供解决方案。通过云端一键部署的先进多模态模型，开发者可以分钟级生成物理精确的合成数据，实现跨模态数据增强与域适应，构建包含边缘案例的多样化训练集，建立持续自我优化的数据闭环系统。

OmniBot 依托奥佳的 GPU 云技术底座，集成 NVIDIA Isaac Sim 和 NVIDIA Isaac Lab，并结合丰富的 AI 大模型，为机器人开发者提供更好、更高效的开发方式、训练、仿真、部署、运营以及机器人系统优化等功能。



图表21: 协创数据旗下 OmniBot 平台基于 NV Cosmos 搭建具身智能开发服务平台

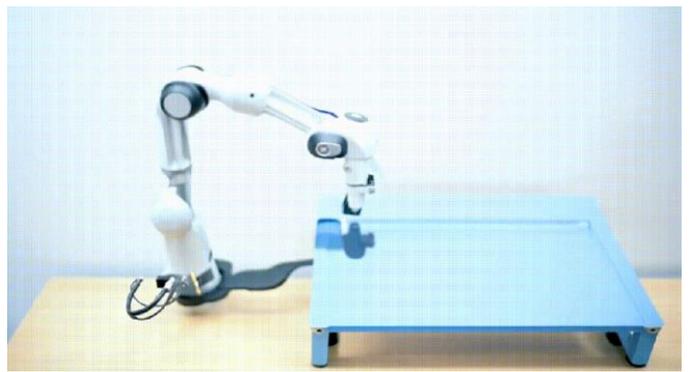
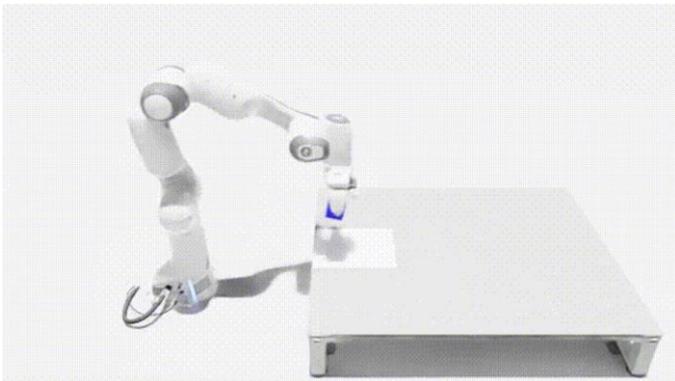


来源: Fcloud 公司官网, 国金证券研究所

基于 Omnibot 开发平台进行数据增强模型实测显示, 使用 NVIDIA Research 官网上的视频案例作为模型输入, 视频中一个机械臂正在移动物体, 我们给 cosmos transfer 1 模型第一个的问题是: “给视频中的机械臂加一个底座”, 可以看到模型可以快速理解机械臂的底部位置在哪里并完成生成动作; 给 cosmos transfer 1 模型第二个的问题是: “给视频中的桌子变成蓝色的, 并给背景加上光线”, 模型基于用户的提示词可以快速执行并生成可用的工业化结果。

图表22: Omnibot 输入视频

图表23: cosmos transfer1 模型输入视频



来源: FCloud-协创数据微信公众平台, 国金证券研究所

来源: FCloud-协创数据微信公众平台, 国金证券研究所

2025 年 2 月, 索辰科技在上海、北京、广州等六地举行索辰物理 AI 开发平台“天工·开物”产品发布会, 推出索辰开物平台。索辰开物平台基于生成式物理 AI 技术和实景渲染技术, 实现真实场景下的四维时空耦合多物理场设计、仿真、优化和训练, 并应用于工业装备的研制和部署。索辰开物平台具有以下四大功能:

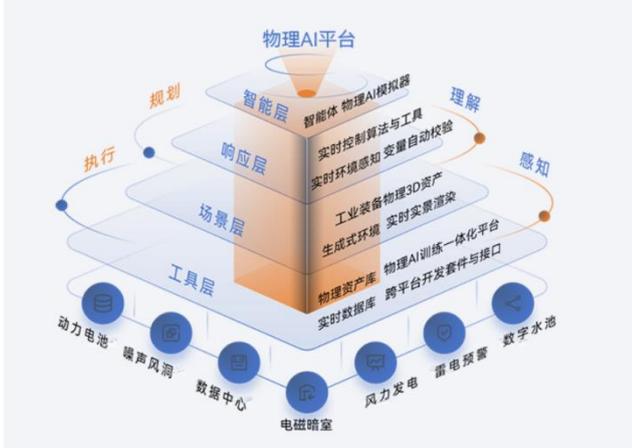
- 环境感知: 通过传感器硬件和企业级实时历史数据库 pSpace, 物理 AI 可实时感知现实环境, 支持国内外主流的 1,000 多种厂家设备, 支持 1,000 客户端并发访问、高速数据读写。通过 3D 渲染场景中的环境感知器, 物理 AI 可实时感知虚拟环境, 获取计算所需的物理参数。
- 设计仿真训练优化一体化: 全参数化的 CAD-CAE 一体化设计使得设计参数贯穿整个工



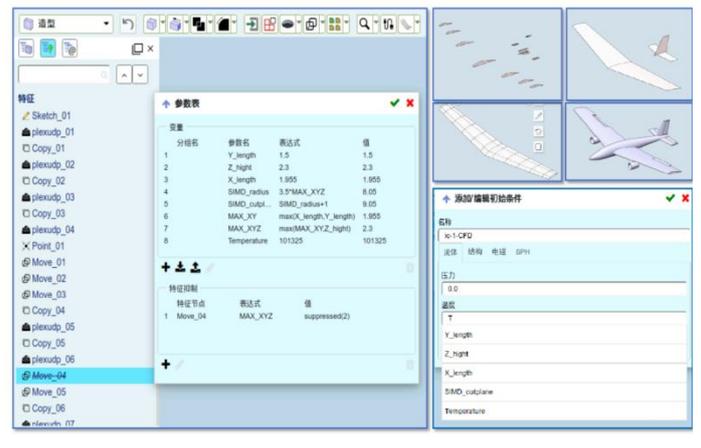
作流。通过特征抑制可快速切换不同学科不同设计阶段的模型，对设计参数的变更可立刻反馈到所有学科的几何模型以及分析参数。在参数化几何上直接赋予仿真分析设置，无需转换数据格式或者在不同软件中导入导出，简化操作，规避数据丢失、冗余，或者几何拓扑错误等问题。通过自动化的工作流实现仿真、训练、优化的一体化运行。

- 数据互联：支持导入常见 3D 资产格式、CAD 格式、网格格式。支持表面物理场、空间物理场、特殊物理场等结果可视化输出。支持导入 URDF 格式的机器人模型。
- 拓展开发：支持对物理仿真和 AI 模块的前后端以及应用平台进行二次开发。

图表24：索辰开物平台涵盖四层架构



图表25：索辰开物平台实现设计仿真训练优化一体化界面



来源：公司官网，国金证券研究所

来源：公司官网，国金证券研究所

2025年7月，索辰科技在索辰开物平台的基础上推出了索辰具身智能虚拟训练平台。该虚拟平台如同一个无限延展的训练场，能够构建出远超真实世界场景库能提供的海量数据量级，打破真实世界的限制，让机器人在各种复杂情境下都能得到充分训练。

2025 CES 英伟达发布 Cosmos 世界模型，迭代 Omniverse 平台与 Isaac Sim 4.5。Isaac Sim 为 AI 机器人开发提供物理虚拟环境，覆盖训练、测试到部署全流程的机器人仿真工具。从技术架构看，索辰科技构建了完整的物理 AI 生态。索辰科技基于自主研发的仿真求解器，开发出覆盖流体、结构、电磁、声学、光学等多学科的核心算法，不仅能够实现传统 CAE 仿真，而且能够支撑物理 AI 的训练、测试和部署全流程。从这个角度看，我们认为索辰科技的物理 AI 实质上是中国版的 Isaac Sim。

群核科技于 2020 年推出空间智能平台 Spatial Verse，这是一个面向室内环境 AI 开发的解决方案，该平台通过构建庞大且物理正确的数据集库，利用高精度 3D 设计数据、渲染引擎与空间编辑工具，生成高度逼真、符合真实世界物理特性（如材质反射与重力模拟）的合成环境，使开发者能训练 AIGC 模型并增强智能机器人、AR/VR 系统的空间认知能力；同时，通过多传感器兼容性与 NVIDIA Isaac Sim 的 Open JSD 框架对接，支持高保真 RTX 渲染和工业级虚拟仿真，SpatialVerse 架起了数字世界与物理现实的桥梁，为空间智能算法提供降低测试成本、加速落地的产业级工具。



图表26: 群核科技 Spatial Verse 空间智能平台基于海量室内 3D 设计数据, 提供高保真仿真合成数据

	解决方案详情	解决方案优势
智能体感知	室内智能设备诸如扫地机器人, 家庭看护机器人, 室内无人机等, 需要面对复杂的室内环境, 并具备环境建图, 导航规划, 物体识别能力, 群核空间智能平台旨在通过大规模的室内环境数据, 比如带标注的 2D 图片数据集以及 3D 环境数据, 帮助智能设备厂商以高性价比的方式实现数据采集和使用。	①质量与精度: 全面的数据管理和审核机制, 保障数据交付过程中的质量把控, 不合格数据的清洗筛查, 利用自动化工具结合人工质检流程, 确保数据交付即可用。②性价比: 结合酷家乐数据库的存量优势以及大规模渲染能力, 将合成数据集的制作成本大大压缩, 相对于真实数据, 综合成本下降到其三分之一甚至更低。③数据安全: 在数据源头即采取合规审查措施, 数据生成过程中不接触任何真实数据, 数据集不包含任何敏感信息, 保障数据使用方合规性需求。
AIGC	AIGC 技术爆发的时代, 诞生了多种生成式大模型, 通常这些模型是基于大量数据进行预训练, 群核空间智能平台可以提供包括图像, 视频, 模型等多种要素的训练数据资源, 并具备完整的标签体系, 帮助 AIGC 研究者实现大模型技术突破。	①数据形态全面: 在酷家乐庞大数据库的支持下, 公司开发了众多数据支持和拓展工具, 实现多样化数据生成能力, 满足图片、视频、以及建模类大模型的数据训练需求, 可以一站式采购, 统一化标准。②标注能力全面: 依靠自动化工具以及人工标注能力结合, 可以实现数据在语义, 材质, 位置, 状态等多种主观和客观标签信息的提供, 满足 AIGC 的训练需求。③数据安全: 在数据源头即采取合规审查措施, 数据生成过程中不接触任何真实数据, 数据集不包含任何敏感信息, 保障数据使用方合规性需求。
机器人仿真	机器人仿真可以快速, 低成本, 安全的对产品进行验证, 群核空间智能平台通过提供高渲染和物理真实性的仿真环境, 并结合具体仿真平台打造仿真环境数据库, 为企业机器人仿真提供助力。	①渲染真实性高: 逼真性和真实感是仿真训练中一个因素, 在酷家乐数据专家团队的支持下, 在 3D 数据渲染效果进行了有效的提升, 在 UE、isaac sim 等平台下, 3D 数据的呈现效果与真实世界做到了极大的接近, 可以更好的促进仿真训练泛化到真实世界中。②数据具备物理真实性: 利用丰富的材质库, 赋予了数据真的物理参数, 比如密度, 摩擦力, 弹性, 阻尼等, 同时还支持对于活动部件的物理约束, 比如抽屉, 门等, 让机器人与仿真世界的交互研究更加真实和便利。③适配多种仿真平台: 包括市面上常见的 gazebo, UE, isaac sim 等平台, 均打通了数据转换通道, 数据流程经过了精心设计和优化, 确保了数据转化的完整性和正确性, 让客户可以快速获取可用格式的数据。
产品可视化推广	数字化展示是一种可以让客户直观感受产品魅力的手段, 群核空间智能平台通过 3D 数据资源库以及高逼真的环境展示能力, 通过数据赋能, 让用户在虚拟环境中感受产品魅力。	①覆盖全国的室内户型资源: 酷家乐积累了众多户型数据, 可以一键搜索和匹配用户家庭户型数据, 实现点到点的直观感受体验。②多类型的室内模型: 数据库包含数量众多的家居、家电、生活、厨卫等模型。③全面的技术支持能力: 提供场景 AI 自动生成能力, 支持户型, 风格一键生成, 支持局部模型挑战, 为企业的营销工具提供全面技术支持。
XR	高质量数据集能显著提升 AR/VR/MR 系统的环境适应能力和用户体验, 如精确的虚实结合效果、顺畅的交互操作等。同时, 数据集也赋能内容创作者获取必要的 3D 模型、材质和动画资源, 推动内容创新, 从而有力推进 XR 技术的整体进步与广泛应用。	①数据定制交付方案: 数据采集, 清洗, 标注一体化解决方案, 用户可以直接使用公司的数据产品, 告别繁琐的数据处理流程。以高效的方式解决 XR 设备对于数据的高需求。②多样化数据资源: 数据资产具备 3D 模型, 3D 点云, 各种环境交互数据等资源, 可以为 XR 设备的内容生成提供数据支撑。③数据安全: 在数据源头即采取合规审查措施, 数据生成过程中不接触任何真实数据, 数据集不包含任何敏感信息, 保障数据使用方合规性需求。

来源: 群核科技官网, 国金证券研究所

世界模型的持续迭代, 使机器人可在虚拟环境中精准预判物理规律与动态变化。北美方面, 谷歌 DeepMind 发布的 Genie 3 首次实现了仅凭文本提示即可生成 720P 分辨率、24FPS 的高动态三维交互世界, 该模型不仅具备长达数分钟的物理一致性与空间记忆, 还能对导航、环境干预等外部操作做出符合物理定律的实时响应。国内方面, 蚂蚁灵波科技全面开源了 LingBot-World 及其衍生模型, 实现了极低延迟的键盘/鼠标实时操控与长达 10 分钟的无损连续生成, 克服了长时漂移与细节塌陷问题, 且首创了自回归“视频-动作”一体化框架, 支持机器人同步推演未来物理状态并输出动作指令。这种内化了物理因果律、空间结构与长时交互能力的新一代世界模型, 为具身智能提供了可靠的虚拟仿真与交互推演支撑, 其生成的海量高质量交互合成数据, 可缓解机器人产业的数据供给瓶颈。



图表27: 谷歌发布通用世界模型 Genie 3 支持生成多样化交互环境



图表28: LingBot-World 在高动态环境下仍能保持主体一致性



来源: 新智元, 国金证券研究所

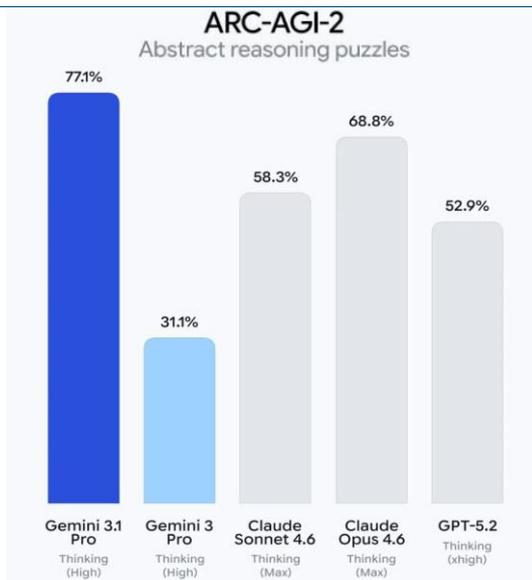
来源: 智东西, 国金证券研究所

3.3 人类视频: 潜力, AI 多模态能力突飞猛进

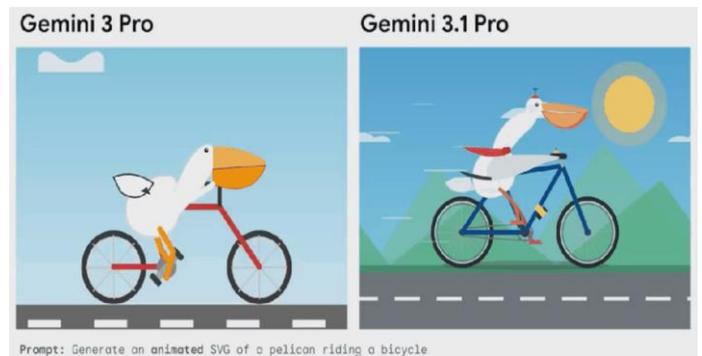
AI 大模型多模态理解与生成能力的快速进步, 使得海量人类第一人称及第三人称视频, 正高效转化为机器人训练数据。此前, 受人手与机器夹爪的跨本体形态差异限制, 人类视频难以直接用于机器人训练。随着多模态大模型持续迭代, 该技术瓶颈已被快速突破, 模型可将人类视频中的动作语义精准映射至机器本体。目前, Tesla Optimus、Figure 等北美头部企业已大规模采用人类视频数据开展模型训练。

谷歌此前发布的 Gemini 3.0 Pro 已展现出强劲的视觉解析能力: 在多模态屏幕理解任务 (Screenshot-Pro) 中得分达 72.7%, 显著高于 Gemini 2.5 Pro (11.4%) 和 Claude 4.5 (36.2%); 在高阶视觉逻辑推理指标 ARC-AGI-2 上, 得分由 Gemini 2.5 Pro 的 4.9% 提升至 31.1%, 同样领先 GPT-5.1 (17.6%)。在此基础上, 最新迭代的 Gemini 3.1 Pro 实现能力再次跃升, 其 ARC-AGI-2 抽象推理指标得分高达 77.1%, 较 3.0 版本实现翻倍提升, 并大幅领先同期的 Claude Opus 4.6 (68.8%) 与 GPT-5.2 (52.9%)。核心指标的快速提升, 印证了多模态大模型在空间抽象与逻辑推理能力上的高速进化。这一底层能力突破, 将显著提升人类视频向机器人训练数据的转化与映射效率, 充分盘活全网海量存量视频资源, 为机器人理解真实物理世界提供持续、高质量的数据支撑。

图表29: Gemini 3.1 Pro 实现抽象推理能力翻倍跃升



图表30: Gemini 3.1 Pro 视觉表现力实现显著提升



来源: Google DeepMind, 国金证券研究所

来源: Google, 国金证券研究所



4. 风险提示

■ 北美人形机器人量产节奏不及预期的风险。

目前产业聚焦在北美量产指引的兑现节奏上，若新版本机器人发布效果不及预期或量产指引持续 miss，或对产业链放量节奏产生一定扰动。

■ 通用机器人 Day1 L4 路线缺乏商业化基础的风险。

目前通用机器人产业发展尚处早期，人形机器人各项软硬件环节成熟度参差不齐，若押注 Day1 L4 路线在展业过程中缺乏足够容量的买单客群，或存在商业闭环到数据闭环都无法建立的风险。

■ 仿真合成数据质量不及预期的风险。

目前“缺数据”仍是机器人落地的核心卡点之一，假如基于 Sim2Real 模式的仿真合成数据质量不及预期，或对英伟达 Isaac Sim、索辰机器人虚拟平台等主体的发展带来扰动。

■ 模型及软件解决方案三方公司长期产业链话语权较低的风险。

参考部分智能驾驶算法厂商与主机厂之间的角力关系，若未来机器人产业渐趋成熟，在国内商业生态普遍对于软件厂商不算友好的环境下，或出现三方数据商、模型算法商产业链话语权较低的风险。



行业投资评级的说明:

- 买入: 预期未来 3-6 个月内该行业上涨幅度超过大盘在 15%以上;
- 增持: 预期未来 3-6 个月内该行业上涨幅度超过大盘在 5% - 15%;
- 中性: 预期未来 3-6 个月内该行业变动幅度相对大盘在 -5% - 5%;
- 减持: 预期未来 3-6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明:

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的任何其他损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应当视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级(含 C3 级)的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话: 021-80234211	电话: 010-85950438	电话: 0755-86695353
邮箱: researchsh@gjzq.com.cn	邮箱: researchbj@gjzq.com.cn	邮箱: researchsz@gjzq.com.cn
邮编: 201204	邮编: 100005	邮编: 518000
地址: 上海浦东新区芳甸路 1088 号 紫竹国际大厦 5 楼	地址: 北京市东城区建内大街 26 号 新闻大厦 8 层南侧	地址: 深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806