

国产算力厂商火速适配Kimi K3 开源风暴席卷全球AI产业

证券时报记者 阮润生

随着月之暗面(Moonshot AI)正式开源Kimi K3模型权重,国产算力生态积极适配。同时,由该模型掀起的开源风暴持续扩大,英伟达集结本土科技巨头达成开源联盟,而闭源模型巨头尚未加入

1 国产生态火速适配

7月27日,Kimi K3完整模型权重正式开放下载。该模型总参数量达2.8万亿,是全球首个开源的3万亿级别模型。模型基于KDA混合线性注意力机制(Kimi Delta Attention)和注意力残差(Attention Residuals)技术构建,原生支持视觉理解,并拥有100万Token(词元)上下文窗口,KV Cache(键值缓存)压缩75%,长上下文解码速度最高提升6.3倍,大幅优化了训练与推理的效率及成本。

根据Artificial Analysis最新排名,Kimi K3智能水平位居全球第四,为开源模型之首。

此次Kimi K3一经开源发布,昇腾便宣布950全系列、Atlas A3产品支持Kimi K3部署,并表示,Kimi K3 896专家的大EP(大规模专家并行)部署场景,能够充分发挥昇腾超节点架构优势。另外,昇腾生态合作伙伴趋境科技也完成了Kimi K3在昇腾Atlas A3上的推理适配。

7月28日,阿里云真武M890超节点

入。业内专家建议,国产AI产业链打造全球开源社区,争取行业标准主导权,提升国际话语权。

记者获悉,7月29日,月之暗面已完成F轮融资。消息显示,公司该轮融资金额超35亿美元,并以500亿美元的融资前估值进行上市前融资。

实例已适配Kimi K3,双方将进一步展开国产算力合作,千问AI平台和阿里云百炼也将提供Kimi K3的模型API。同时,阿里云自研芯片平台平头哥宣布与月之暗面联合调优,配套阿里云真武M890超节点,专门适配万亿级MoE模型高并发推理。据测试,真武超节点运行Kimi K3的首Token时延降低约35%,长上下文场景下用户体验更流畅。

A股上市公司中,海光信息宣布完成该模型在海光DCU平台的全流程适配、部署与性能验证。围绕Kimi K3的架构特性,海光DCU团队依托自研DAS/DTK软件栈,完成了从底层算子、训练推理框架到推理引擎的全栈适配,带来部署与性能的双重升级。

另外,摩尔线程28日宣布,基于AI训推一体智算卡MTT S5000及MUSA软件栈,公司完成Kimi K3的适配;润建股份29日宣布,Kimi K3模型已正式登陆算力服务平台“润道星算”。

2 任务成本差异加大

Kimi K3的发布,持续引发对当前开源与闭源模型性的任务成本差距的关注。

IDC中国研究经理程荫向证券时报记者表示,相比闭源头部前沿模型(GPT系列、Claude旗舰版),同等能力的国产开源模型(Kimi K3、DeepSeek旗舰)依托更高效架构、国产算力集群,训练总成本仅为闭源厂商的1/5至1/6。

在推理方面,闭源模型需要自研整套训练底座、私有高质量数据集,固定研发成本极高,开源模型与闭源模型的推理任

务成本差距可达几十倍。

成本优势反映在市场流量结构变化上。非营利机构Mozilla统计,截至2025年底,开源权重模型在OpenRouter平台上Token占比约1/3;到今年7月,Token交易量排名前七的均为开源权重模型。

更重要的是,Kimi K3并未走低成本路线,其定价属于国产高端模型水平,甚至倒逼全球头部模型厂商调整策略。

“随着需求分层,不同梯队模型定价权出现差异。”瑞银证券中国互联网行业分析师熊玮日前在媒体分享会上表示。

据瑞银证券统计,中国头部模型的训练成本约为海外龙头的1/10;推理成本方面,中国模型API(应用编程接口)定价约为海外对标模型平均定价的10%至20%,且仍能保持20%至40%的毛利水平,略低



图片来源:AI生成

于美国同行,但仍然是一个非常健康的水平。

“这种优势并非源于激进补贴或者厂商疯狂让利,而是结构性的成本效率优

势,包括训练算法创新和架构设计、模型策略重点差异以及开源生态协同等因素。”同时,熊玮提醒国产模型训练和推理还存在明显的算力瓶颈。

3 开源风暴席卷全球

本次Kimi K3的发布掀起了开源风暴。一方面,OpenAI与Anthropic的高管质疑开源模型的安全性与合规性,甚至警告开源将削弱美国实验室加码算力投资的商业动机。另一方面,近200家美国初创企业敦促美国政府不要切断对中国开源模型的访问。

近日,英伟达创始人、CEO黄仁勋在接受节目采访时直言“中国模型非常出色”,市场低估了开源模型Kimi带来的影响力,就像此前低估了DeepSeek一样。

同时,高质量的开源大模型可以让更多人使用AI,这样会带来更多芯片、数据中心、算力的需求,催生更多AI服务应用。英伟达于7月27日宣布成立“开放安全AI联盟”(OSaIA),最新已经获50多家合作伙伴加入。

关于开源生态的竞争,程荫建议国内AI产业链算力端软硬件协同优化,搭建分层算力网络;完善全链路国产软件栈,建立分层商业模式,摒弃内卷聚焦行业落地;打造全球开源社区,争取行业标准主导权,提升国际话语权。

人形机器人板块活跃 超十家公司上半年业绩预增

证券时报记者 聂英好

7月29日,人形机器人概念股活跃,明新旭腾3连板,宇环数控涨停,埃斯顿、光洋股份、丰光精密、万马股份等跟涨。

行业利好消息不断

消息面上,近日人形机器人头部厂商利好消息不断。7月24日,智元创新正式启动赴港上市流程,成为人形机器人赛道又一重大资本动作。此外,宇树科技发布新款轻量化机器狗Unitree As2。

近期,A股多人形机器人产业链公司业绩预喜。有券商观点认为,随着头部企业产能持续爬坡和工业场景规模化部署率先跑通,人形机器人商业化闭环有望加速形成,产业端与资本端形成共振,量产元年的投资机会逐步显现。

相关公司业绩预喜

Wind数据显示,A股共计116家人形机器人概念股,截至7月29日,16家概念股披露上半年业绩预增公告,涵盖埃斯顿、雷赛智能、亿纬锂能、当升科技、慈星股份、拓斯达、金力永磁等。2026年,具身智能企业深入实景应用场景,各大人形机器人企业加速商业化落地,带动人形机器人产业链公司上半年业绩预增。

国产工业机器人龙头埃斯顿预计

2026年上半年实现归母净利润1.5亿元—1.8亿元,同比增长2144.74%—2593.68%;预计扣非净利润6000万元—7500万元,同比增长440.85%—526.07%。

埃斯顿将上半年业绩大幅增长归因于三方面:一是优化产品结构,重点聚焦高附加值产品及高质量订单,综合毛利率同比显著提升;二是加强精细化管理、实施严格的预算管控措施,其间费用率同比有所下降;三是因对参股公司资产重组完成,公司持有的南京工艺股权置换为上市公司南京化纤股权,股权公允价值增加。

雷赛智能聚焦机器人核心零部件,重点布局关节模组和灵巧手两大核心品类,该公司预计2026年上半年营收12.30亿元—12.66亿元,同比增长38%—42%;预计归母净利润1.84亿元—1.96亿元,同比增长55%—65%。雷赛智能表示,上半年下游智能制造的自动化设备市场环境有所回暖,公司业务订单情况有较大好转,叠加公司持续加大对重点战略性行业以及新兴业务领域的市场开拓力度,相关业务收入贡献逐步提升;同时,该公司的新兴业务与第二增长曲线,也取得较大突破与增长。

“广东机器人第一股”拓斯达预计2026年上半年归母净利润介于9000万元至1.15亿元,比上年同期的2873.22万元增长213.24%—300.25%。针对业绩大幅增长的原因,拓斯达解释称:一方面,公司产品竞争力增强及与3C头部客户合作深

度广度提高,工业机器人及自动化应用系统业务收入显著增加,毛利贡献大幅提升;另一方面,公司加强应收账款回收,信用减值损失同比减少。

工业机器人公司凯尔达预计2026年上半年归母净利润2615.11万元—3053.38万元,同比增长1005.75%—1191.07%,上年同期盈利236.5万元。凯尔达称,报告期内业绩变化的主要原因是,随着公司产品竞争力逐步提升和下游市场需求的相对回暖,公司营业收入持续增长,同时股份支付费用同比下降等因素共同带动了业绩增长。

此外,供应机器人电池的亿纬锂能预计上半年归母净利润31.3亿元至33.71亿元,同比增长95%至110%;当升科技预计归母净利润为4.8亿元至5.3亿元,同比增长54.26%至70.33%。金力永磁预计上半年净利润4亿元至4.6亿元,其中机器人及工业伺服电机领域营收同比增约90%,具身机器人电机转子产品已有小批量交付。

招商证券机械行业首席分析师郭倩倩分析,整个机器人行业的利润和收入规模同步往前走,包括电机、减速机、传感器、轴承等,产业链上每一个环节的公司均有不同程度受益。

加速商业化落地

尽管人形机器人和具身智能仍处于产业化早期,但今年行业正在积极迈向规模化

商业化,真实订单、交付数量、现场运行时长、故障率、任务成功率以及客户投资回收周期等指标的重要性将进一步提升。

工信部数据显示,1月至5月我国机器人行业规上企业营收突破900亿元,同比增长26.9%,近5年年均增速超过20%。工信部预计,2026年国内整机产量将突破10万台大关。

2026年是人形机器人量产验证与场景落地的关键窗口,从供给端看,特斯拉、宇树科技、智元机器人、优必选稳步推进量产节奏。今年6月,智元第15000台具身智能机器人下线,距离万台量产节点不足3个月;优必选6月2日开始预售的“全球首款全尺寸超仿生人形机器人”优世界U1,截至6月30日线上线下全渠道订单累计突破13361台。特斯拉人形机器人Optimus已经进入量产关键阶段,其加州弗里蒙特工厂的Model S/X产线已停用并改造为Optimus首条专用产线,规划年产100万台。

中金公司在研究报告中称,2025年全球人形机器人已进入万台级交付规模,国产厂商抢跑交付,2026年下半年海外量产线进入拉通阶段,2027年进入比肩竞争期;产业核心驱动力将从依托技术想象的主题投资,转变为依托量产进度和实际场景订单的行业价值兑现。



7月28日,超讯科技召开算电协同行业交流会,公司董事长梁建华在会上明确表示,公司已全面从通信技术转向算力及能源协同产业,算力业务已成为公司第一增长曲线,营收占比超过七成。

交流会背后折射出的是国内一批深耕通信工程、基站建设与网络代维优化的上市公司正在集体完成业务重心迁移。过去主业是服务三大运营商的5G基站建设与网络维护,如今纷纷加码算力集群建设、算力租赁与运维业务。以超讯科技、润建股份、中贝通信为代表的企业,正依托各自积累的全国工程实施队伍、机房站点资源与运营商客户关系切入算力新赛道。

超讯科技的转身颇具代表性。今年5月,公司完成从“超讯通信股份有限公司”到“超讯科技股份有限公司”的更名,直接原因是2025年智算业务收入占总营收比例已达68.36%,成为驱动业绩增长的核心动能。

此次交流会上,梁建华表示,围绕AI算力公司已构建起四大核心能力。一是国产AI芯片及服务器集成能力,其原型服务器已入围国家机关和税务系统采购清单;二是智算中心全生命周期运维能力,具备大型算力中心及万卡超节点集成实力;三是算力运营服务能力,实现从算力资源接入到应用服务支撑的全流程运营;四是新能源与算力融合发展能力,探索源网荷储算协同模式。这些能力的组合,本质上是把过去服务运营商积累的工程实施与机房运维经验,平移到了算力基础设施的建设与运营场景中。

与此同时,润建股份走的是另一条更聚焦“Token(词元)化”的路径。公司提出“以Token为核、绿算协同、智维百业”战略,将新能源工程建设、智能运维、绿色智算、虚拟电厂、AI能源及Token全栈服务整合为一体。

7月世界人工智能大会(WAIC)期间,润建股份与明阳智慧能源正式签署战略合作协议,双方将围绕新能源装备采购、绿色智算及算电协同展开合作,进一步夯实“AI+能源”双向赋能模式。在南宁五象云谷,润建股份已自建220千伏专用变电站,提供480兆瓦容量支持,公司方面预计这一算电协同布局可降低30%以上的算力使用成本。

中贝通信选择围绕智算业务与AI生态,积极布局产业链上下游优质企业,旨在共同推进通用人工智能基础研发与AI应用落地,打造AI产业生态。截至2026年4月底,该公司已运营算力规模超过22000P,并持续推进智算集群建设,目前正在运营的包括上海松江、上海临港、青海海东、甘肃庆阳、海外等多个智算集群项目。

华泰证券通信研究团队首席分析师王兴表示,全球算力需求仍处于持续扩张阶段。尽管市场近期因现金流担忧出现回调,但产业基本面依然强劲。随着头部大厂持续加大算力投资,国内算力基建飞轮已经启动。

银河证券通信首席分析师赵良毕表示,Token数量的持续增长验证了AI算力需求的真实性,所谓“算力过剩”为时过早。

上半年利润增长低于预期 SK海力士今年资本开支扩大

证券时报记者 阮润生

存储巨头SK海力士7月29日公布了美股上市后首份财报:今年上半年公司营业利润同比增长超5倍,不过仍低于分析师预期,开盘后股价持续下跌。

对于存储行业供需情况,公司高管在业绩说明会上称,模型与系统效率提升后,AI的覆盖面和可及性反而扩大;短期内供需失衡仍难以显著改善;另外,鉴于投产时间表提前和投资扩大,2026年公司资本开支预计将达到40万亿韩元上方。

受益于面向AI服务器的高性能产品需求持续放量,上半年SK海力士累计营收首次突破100万亿韩元,同比增长2.57倍;营业利润同比增长5.57倍。其中,第二季度公司实现营收79.32万亿韩元,营业利润60.54万亿韩元;净利润93.92万亿韩元,环比增长133%。

不过,此前分析师普遍预期第二季度公司营业收入84.06万亿韩元,营业利润64.09万亿韩元。

SK海力士表示:“DRAM和NAND闪存价格延续上季度涨势,均实现大幅提升。公司通过扩大高带宽存储器(HBM)、面向AI服务器的DRAM以及企业级固态硬盘(eSSD)等高附加值产品的销售,实现历史最佳盈利水平。”

公司已包括核心客户在内的十余家客户结束讨论长期供应协议(Long Term Agreement),并正与业内主要客户进一步展开洽谈。

面对市场对AI投资回报的广泛质疑,SK海力士高管在电话会上回应称,公司注意到部分大型科技企业正在重新审视数据中心租赁业务,叠加高效AI模型兴起,引发外界对AI基础设施投资可能放缓的担忧。但SK海力士认为,这些调整并非削减投资,而是提升利用率和加速大规模AI基础设施商业化的过程。

公司高管表示,在云厂商持续的AI竞争以及AI服务持续扩展的支持下,AI基础设施投资将在明年之后保持稳固,并预计存储器需求总体上将持续扩大。这不仅包括用于AI计算的HBM,还包括支持智能体AI的服务器DRAM,以及适应AI服务扩展和数据增长的高性能、高容量NAND。

在SK海力士看来,在客户需求持续超过供给的背景下,能否及时满足客户需求,已成为核心竞争优势。为此,公司正提前推进韩国清州M15X的量产进程,同时在为2027年初龙仁集群首座工厂无尘室启用后快速扩产展开相应投资。

鉴于投产时间表提前和投资扩大,2026年SK海力士的资本开支预计将扩大到40万亿韩元上方。SK海力士强调,将在积极把握中长期增长机遇的同时,坚持设备投资原则,实现产能扩张与财务稳健的协同发展。

算电协同成新战场 多家通信工程商切换赛道

证券时报记者 康殷