



计算机行业研究

买入（维持评级）

行业点评
证券研究报告

计算机组

分析师：刘高畅（执业 S1130525120005） 分析师：李可夫（执业 S1130525120009）

liugaochang@gjzq.com.cn

likefu@gjzq.com.cn

开源模型跃迁会带来什么影响

一、Kimi K3 引领开源模型进入 3T 新阶段。国产开源模型持续实现技术突破，Kimi K3 作为全球首个开放 3T 级模型拉开开源大模型 3T 时代序幕。该模型拥有 2.8T 参数，搭载专属注意力技术，原生支持视觉输入与百万 Token 超长上下文，综合评测成绩仅落后头部海外闭源大模型；英伟达黄仁勋发布公开信，明确开源模型具备安全、创新、AI 自主主权多重价值，行业将同步发展前沿闭源与开源模型，持续拉动云、算力基础设施需求。即便基座开源，大模型厂商仍具备底层调优优势，配套开源组件同步加速行业渗透。

二、模型二次开发空间扩大，软件公司价值量提升。开源模型拓宽企业深度开发边界，推动软件行业价值重构。相较于闭源 API 仅能做基础检索与提示词调试，开源模型可覆盖数据治理、模型微调蒸馏、行业定制对齐、多环境部署全流程；软件厂商核心竞争力从 API 集成能力转向行业数据、业务流程与模型工程能力，公有云、私有云、边缘混合部署模式也进一步拉高软件企业模型治理服务的附加值。

三、算力需求环节增加，云厂商 AI Infra 经验重要性凸显。开源模型全生命周期运营催生更多算力负载，具备完善 AI 底层基础设施的云厂商优势显著。企业针对不同行业、客户定制专属模型版本后，算力消耗不再局限于终端推理，还延伸至数据处理、微调训练、评测验证、长期运维全环节；云平台业务从单一 API 入口升级为 GPU 供给、模型托管、全链路调优一体化服务，完整模型生命周期带来持续扩张的算力需求。

四、模型进入 3T 时代，部署需求超节点，利好国产芯片。3T 级别大模型推动行业部署架构转向超节点集群，国产 AI 芯片迎来发展机遇。Kimi K3 官方要求至少 64 卡集群才能实现高效推理，2026 WAIC 展会中华为昇腾、沐曦等国产芯片厂商集中推出大规格超节点产品；3T 模型将竞争从单卡性能延伸至芯片、高速互联、整机集群、推理软件栈一体化系统能力，具备完整超节点产品与生态适配能力的国产芯片厂商充分受益。

投资建议

相关标的：

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、首都在线、网宿科技、云赛智联、青云科技等。

海外算力：东山精密、工业富联、胜宏科技、江海股份、中际旭创、中钨高新、蓝思科技、东阳光、火炬电子、三环集团、欧科亿、天孚通信、鼎泰高科、新易盛、领益智造、兆易创新、鹏鼎控股、唯科科技、天岳先进、大普微、源杰科技、麦格米特、景旺电子、英维克、京东方等；英特尔、SK 海力士、Lumentum、闪迪、铠侠、美光。

国内算力：寒武纪、海光信息、长鑫科技、东阳光、利通电子、中芯国际、华虹半导体、华达科技、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、杰华特、京基智农、中科曙光、浪潮信息、利扬芯片、胜蓝股份、华勤技术、国科微、中国长城、晶科科技、罗曼股份、盈峰环境、芯原股份、晶科科技、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、神州数码、润泽科技、大位科技、润建股份、奥飞数据、瑞晟智能、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子。

风险提示

行业竞争加剧的风险；技术研发进度不及预期的风险；AI 资本开支压力过高的风险。



内容目录

一、Kimi K3 引领开源模型进入 3T 新阶段	3
1.1 国产开源模型突破，持续追赶海外闭源模型	3
1.2 黄仁勋发表开源模型公开信	3
二、模型二次开发空间扩大，软件公司价值量提升	4
三、算力需求环节增加，云厂商 AI Infra 经验重要性凸显	4
3.1 更大的二次开发空间最终转化为更多算力工作负载	4
3.2 云承接完整模型生命周期，算力调优对 AI Infra 需求增大	5
四、模型进入 3T 时代，部署需求超节点，利好国产芯片	5
五、开源模型对产业影响总结	5
风险提示	6

图表目录

图表 1: Artificial Analysis 模型评测榜单	3
图表 2: 黄仁勋在 X 首帖发布开源模型公开信	3
图表 3: 部分美股软件业务	4
图表 4: 曦景 S 系列超节点产品	5



一、Kimi K3 引领开源模型进入 3T 新阶段

1.1 国产开源模型突破，持续追赶海外闭源模型

Kimi K3 拥有 2.8T 参数，基于 Kimi Delta Attention (KDA) 与 Attention Residuals (AttnRes) 构建，具备原生视觉能力和 100 万 Token 上下文窗口，并将其称为全球首个开放 3T 级模型。在 Artificial Analysis 评测基准上，kimi k3 得分仅次于 GPT-5.6/Claude Fable 5 和 Claude Opus 5。

图表1: Artificial Analysis 模型评测榜单



来源: artificial analysis, 国金证券研究所

我们认为大模型全面普及背景下，即便基座对外开放，头部大模型厂商依旧保有不可替代的核心优势：底层架构迭代、大规模参数调优等硬核技术门槛持续拉开差距；同时模型厂商配套开源开发套件、微调工具、推理优化组件持续下沉渗透，进一步巩固自身生态话语权。

1.2 黄仁勋发表开源模型公开信

黄仁勋在 X 账号首帖中分享了英伟达参与联署的开源模型公开信，并表示，AI 将改造每个行业、赋能每家公司、由每个国家建设；开源模型能够强化安全与网络安全、加速创新与扩散并支持主权化，世界同时需要前沿闭源模型与前沿开源模型。

图表2: 黄仁勋在 X 首帖发布开源模型公开信





来源：X，国金证券研究所

我们认为，黄仁勋将开源模型与创新扩散、主权化并列，说明开源模型的产业价值不只在成本，更在于让企业和国家能够在自己的数据、云和硬件环境中建设 AI。多主体、多区域部署有望扩大云平台、数据中心与推理算力的需求边界。

二、模型二次开发空间扩大，软件公司价值量提升

我们认为，开源模型的进步，使企业、软件公司和云平台可以进入过去被闭源 API 遮蔽的模型深层：

在数据与知识层，闭源 API+RAG 主要围绕知识库、检索和提示词展开，开源模型则可以把开发范围扩展到数据治理、合成、训练与持续回流。企业对模型交互数据的沉淀可以更深层次地优化开源模型。

在模型层，企业也由调用固定能力进一步进入微调、蒸馏、对齐、路由与评测，同一基座模型可以形成更多行业版本、岗位版本和任务版本，应用丰富度明显增加。软件公司的差异化从“谁更会接 API”转向“谁拥有行业数据、工作流理解和模型工程能力”。软件公司的二次开发价值量因此提升。

在部署和产品层，开源模型支持公有云、私有云、本地、边缘及混合部署。也为软件公司的治理能力提出更高要求，进一步带来价值量提升。随企业级 Agent 进入部署阶段，数据库查询、日志监控、工作流调度等任务从“人使用软件”转向“软件调用软件”，系统复杂度上升后，企业对于数据治理、实时监控、权限体系、Workflow orchestration 等方面的治理需求进一步增强，因此具备可观测、安全、身份、Workflow 类的平台能力的软件公司，将重新获得估值溢价。

图表3：部分美股软件业务

公司	定位/业务情况
Snowflake	全球领先云数据平台，提供数据仓库、数据共享及 AI 数据基础设施服务，Cortex Code 等 AI 产品渗透率持续提升
ServiceNow	企业级工作流与 IT 服务管理龙头，推动企业流程自动化与 Agentic AI 落地，Now Assist ACV 显著增长
CRWD	全球领先云原生网络安全平台，以 Falcon 平台覆盖终端、安全运营及身份安全，Falcon Flex 持续放量
Atlassian	聚焦企业协同办公与研发管理，Rovo 快速增长、云业务加速、AI 驱动客单价提升
Datadog	云监控与可观测性平台领军者，显著受益于 AI 原生应用扩张带来的算力与监控需求增长
Palantir	通过 AIP 平台为客户提供数据分析与 AI 应用服务，AIP 需求持续超预期
PANW	覆盖网络安全、云安全与 SOC 运营，Next-Gen Security 业务高增长
FTNT	网络安全基础设施龙头，核心产品包括防火墙、安全网络与 SASE 平台
Salesforce	涵盖销售、营销、客服及企业 AI Agent 平台，Agentforce 商业化加速、企业 AI 应用需求增长

来源：iFind，国金证券研究所

三、算力需求环节增加，云厂商 AI Infra 经验重要性凸显

3.1 更大的二次开发空间最终转化为更多算力工作负载

我们认为，软件公司为每个行业、客户和业务流程维护不同模型版本后，算力不再只发生在用户点击应用的时刻，还发生在数据处理、适配训练、版本评测、灰度验证和持续运维阶段。项目数量和版本数量的增加会扩大算力消耗。具备客户入口与行业数据的软件公司可以把模型二次开发打包为订阅、专属托管、按效果付费和持续运维服务。算力由成本项转化为产品交付的一部分，软件公司与云、算力供应商形成更紧密的长期合作。



3.2 云承接完整模型生命周期，算力调优对 AI Infra 需求增大

我们认为，开源模型对云的需求在于：云从闭源 API 入口扩展为 GPU 供给、模型托管、数据治理、训练调优、推理服务与运维平台。企业同时管理多个模型和多个适配版本后，云平台的调度、治理和规模化交付价值更高。

算力需求也从单次 API 推理扩展为“训练/微调—评测—专属部署—持续推理—再训练”的完整链条，多阶段工作负载和更大的模型规模对算力调优需求增大，利好具备更强 AI Infra 能力的云厂商。开放门槛会带来更多模型版本、更多场景和更高调用频率，总算力需求有望扩张。

四、模型进入 3T 时代，部署需求超节点，利好国产芯片

Kimi 官方技术博客建议，为获得更大的高带宽通信域、提升推理效率，Kimi K3 应部署在配备 64 张及以上加速卡的超节点配置上。

WAIC 上 2026 多家国产 AI 芯片厂商集中发布或展示超节点：

华为官 WAIC 2026 期间，昇腾 950 超节点 (Atlas 950 SuperPoD) 首次线下展出，真机采用 1024 卡集群部署。

沐曦股份官网显示，公司在 WAIC 2026 正式发布曦景 S 系列超节点，其中曦景 S600 实现单机柜 64 卡高密度部署和柜内全互连。

图表4：曦景 S 系列超节点产品



来源：沐曦股份官网，国金证券研究所

我们认为，接近 3T 的开源模型正在把推理优化从单卡、单机问题推向大规模高带宽互连与系统级协同。国产 AI 芯片厂商有望通过芯片、互连、整机和软件栈协同提升大模型部署能力，利好具备超节点产品与生态适配能力的国产芯片厂商。

五、开源模型对产业影响总结

我们认为，随着国产开源模型持续进步，追赶行业前沿闭源模型：

软件公司有望由 API 集成商升级为模型资产与持续运营服务商，可以基于自己的行业 know how 和数据优势形成领域专属微调模型，迎来价值量上升。

云厂商有望承接模型选择、调优、托管、运维与治理需求，受益逻辑在于工作负载由单点 API 推理扩展到训练、评测、专属部署和持续推理，具备 AI Infra 经验的云厂商更能发挥推理成本上的技术优势。



开源大模型参数规模持续扩大,部署需求向超节点演进:更大通信域和更强系统协同要求,使竞争维度由单卡性能扩展到芯片、互连、整机、编译与推理软件栈,国产 AI 芯片厂商有望受益。

风险提示

行业竞争加剧的风险:

众多新兴玩家参与到市场竞争之中,若市场竞争进一步加剧,竞争优势偏弱的企业或面临出清,某些中低端品类的毛利率或受到一定程度影响。

技术研发进度不及预期的风险:

技术开发需投入大量资源,如果相关厂商新品研发进程不及预期,表观层面将呈现出投入产出在较长时期的滞后特征。

AI 资本开支压力过高的风险:

AI 资本开支过高可能导致云厂商自由现金流持续承压,并带来折旧压力。



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。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、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其他任何损失承担任何责任。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

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国内大街 26 号

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究