国投证券 SDIC SECURITIES

2024 年 03 月 24 日
计算机

行业动态分析

证券研究报告

Kimi 升级+阶跃星辰发布，国产大模型黑马蓄势跃升

投资评级 领先大市-A
维持评级

首选股票	目标价（元）	评级

事件概述：

1）3 月 23 日，通用大模型创业公司阶跃星辰在 2024 全球开发者先锋大会期间正式对外亮相，并在大会开幕式上发布了 Step 系列通用大模型，包括 Step-1 千亿参数语言大模型、Step-1V 千亿参数多模态大模型，以及 Step-2 万亿参数 MoE 语言大模型的预览版。
2）3 月 18 日，AI 大模型初创企业月之暗面宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破，无损上下文长度从 20 万字提升至 200 万字，即日起，支持 200 万字上下文的 Kimi 已启动“内测”。

Kimi 扩容“内存”，200 万字长文本能力全球领先

3 月 18 日，公司宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破，无损上下文长度从 20 万字提升到 200 万字，在单点能力上已经超越了海外的大模型。支持更长的上下文意味着大模型拥有更大的“内存”，从而使得大模型的应用更加深入和广泛：比如通过多篇财报进行市场分析、处理超长的法务合同、快速梳理多篇文章或多个网页的关键信息、基于长篇小说设定进行角色扮演等。百万字无损上下文阅读能力帮助 Kimi 高速掌握新领域，过去要 10000 小时才能成为专家的领域，现在只需 10 分钟，Kimi 就能接近任何一个新领域的初级专家水平。

阶跃星辰重磅发布万亿参数 MoE 大模型预览版

阶跃星辰自成立起，在算力、数据、算法和系统这四大要素上综合布局，在大模型技术路径上坚定投入攀登 Scaling Law。阶跃星辰发布的 Step-1V 千亿参数多模态大模型多模理解能力突出，可以精准描述和理解图像中的文字、数据、图表等信息，并根据图像信息实现内容创作、逻辑推理、数据分析、视频理解等多项任务。千亿参数模型只是阶跃星辰团队在攀登通用人工智能路上迈出的第一步。此次大会上阶跃星辰还发布了 Step-2 万亿参数语言大模型预览版，提供 API 接口给部分合作伙伴试用。

建议关注：

1）AI 算力：润泽科技、云赛智联、中科曙光等。
2）AI 应用：金山办公、福昕软件、万兴科技、金蝶国际、彩讯股份、致远互联、拓尔思、通达海、华宇软件等。

风险提示：

AI 技术发展不及预期；行业竞争加剧。

行业表现

资料来源：Wind 资讯

升幅%	1M	3M	12M
相对收益	14.0	-5.9	-6.1
绝对收益	15.7	0.3	-17.5

赵阳 分析师
SAC 执业证书编号：S1450522040001
zhaoyang1@essence.com.cn

夏瀛韬 分析师
SAC 执业证书编号：S1450521120006
xiayt@essence.com.cn

马诗文 联系人
SAC 执业证书编号：S1450122050037
masw2@essence.com.cn

相关报告

英伟达 GTC2024 召开在即，AI、机器人或迎来催化	2024-03-17
计算机投资视角解读两会和新质生产力	2024-03-11
海外国内共振，AI 算力引领科技投资高景气	2024-03-04
卫星互联网产业化积极推进，发射端、卫星端获进展	2024-02-26
文生视频模型 Sora 有望引领 AI 新景气	2024-02-19

内容目录

图表目录

1. Kimi：无损长文本处理能力全球领先

Kimi 长文本输入量提升 10 倍，目前全球领先。AI 大模型初创企业月之暗面（Moonshot AI）创立于 2023 年 3 月，主力产品 Kimi 智能助手在 2023 年 10 月初次亮相，凭借约 20 万汉字的无损上下文能力，帮助用户解锁了很多新的使用场景，包括专业学术论文的翻译和理解、辅助分析法律问题、一次性整理几十张发票、快速理解 API 开发文档等，获得了良好的用户口碑和用户量的快速增长。今年 3 月 18 日，公司宣布 Kimi 智能助手在长上下文窗口技术上再次取得突破，无损上下文长度提升了一个数量级到 200 万字。根据机器之心数据，尚未上线的 GPT-4.5 Turbo 上下文窗口指定为 25.6 万个 token，Kimi 此次升级后长文本能力是其 10 倍，是**目前全球市场上能够产品化使用的大模型服务中所能支持的最长上下文输入长度。**

图1. Kimi 智能助手启动 200 万字无损上下文内测

资料来源：Moonshot AI 官方公众号，国投证券研究中心

更长上下文意味着更大“内存”，提高海量文件处理效率。从技术视角看，参数量决定了大模型支持多复杂的“计算”，而能够接收多少文本输入（即长文本技术）则决定了大模型有多大的“内存”，两者共同决定模型的应用效果。支持更长的上下文意味着大模型拥有更大的“内存”，从而使得大模型的应用更加深入和广泛：比如通过多篇财报进行市场分析、处理超长的法务合同、快速梳理多篇文章或多个网页的关键信息、基于长篇小说设定进行角色扮演等。同时，Kimi Chat 通过创新的网络结构和工程优化，在千亿参数下实现了无损的长程注意力机制，不依赖于滑动窗口、降采样、小模型等对性能损害较大的“捷径”方案。

图2. Kimi 能够快速分析总结出英伟达的财报历史

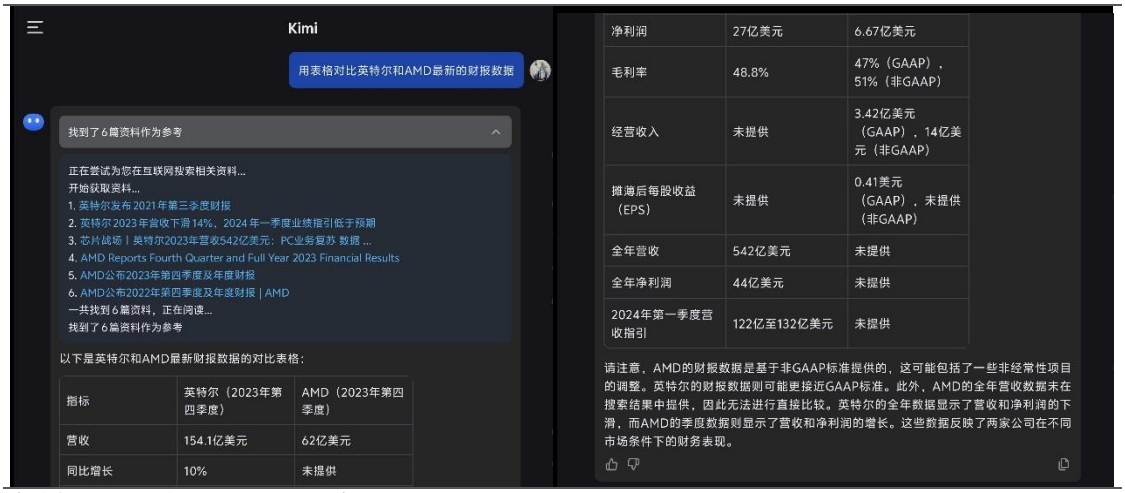
资料来源：Moonshot AI 官方公众号，国投证券研究中心

图3. Kimi 根据 500 份简历筛选候选人信息

资料来源：Moonshot AI 官方公众号，国投证券研究中心

智能检索并分析总结与长文本处理能力息息相关：Kimi 可以根据用户的问题，主动去互联网上搜索、分析和总结最相关的多个页面，搜索得到的多篇资料，会作为上下文的一部分交给模型去推理，生成更直接、更准确的答案。正是因为 Kimi 大模型支持的上下文窗口足够长，窗口内的信息损失足够低，Kimi 智能助手才能输出高质量的结果。例如，用户可以让 Kimi 主动去搜索和对比两家同领域上市公司的最新财报数据，直接生成对比表格，节省大量的资料查找时间。

图4. Kimi 能够快速分析对比上市公司财报数据

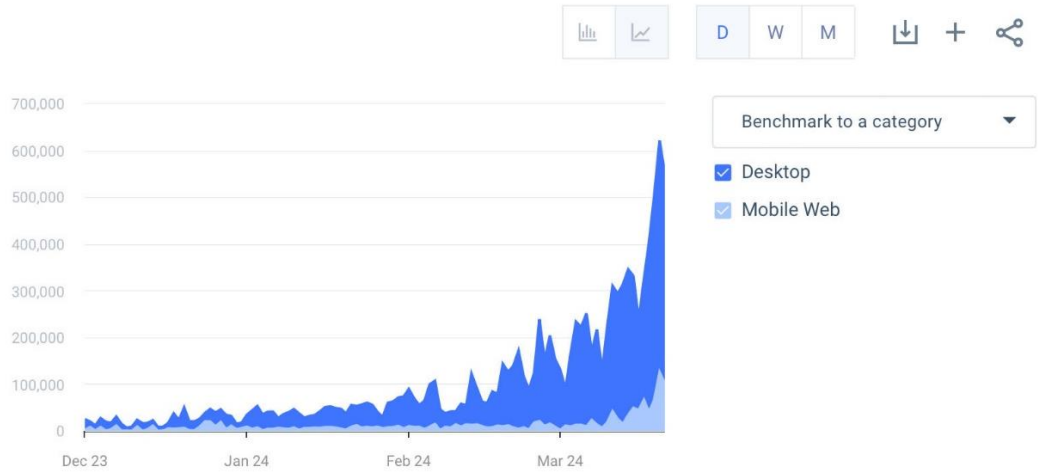


资料来源：KIMI 智能助手，国投证券研究中心

多轮交互和超长指令遵循能力突出：与大模型的无损上下文能力息息相关的指标是指令遵循（Instruction Following）能力。指令遵循能力主要体现在两个方面：1）模型在多轮对话中是否能够始终遵循用户的指令，理解用户的需求；2）模型是否能够遵循复杂指令，有时候复杂指令可能长达几千、上万字。从产品推出以来的用户反馈来看，Kimi 智能助手的多轮交互和超长指令遵循能力，也是产品的一项核心优势。

Kimi 流量增加趋势远超预期，已采取扩容等紧急措施。根据 Similarweb 数据，Kimi 网页版日活用户数当前已连续数日超 20 万，峰值日活达 34.6 万，周活数据环比增长 45%持续创新高。月之暗面发布情况说明，从 2024 年 3 月 20 日 9:30 开始观测到 Kimi 的系统流量持续异常增高，流量增加的趋势远超公司对资源的预期规划。这导致了从 3 月 20 日 10:00 开始，有较多的 SaaS 客户持续的体验到 429:engine is overloaded 的异常问题，对此公司深表抱歉，已经有多项应急措施正在实施，包括不限于：从观测到流量异常增高后，已经进行了 5 次扩容工作。推理资源会持续配合流量进行扩容，以尽量承载持续增长的用户量；设计了一套更有效的 SaaS 流量优先级策略，以保障付费用户的调用稳定，预计 3 月 25 日之前完成并上线。

图5. Kimi 访问量激增



资料来源：similarweb，国投证券研究中心

顶尖算法工程人才汇聚，创始团队成员参与过多个大模型研发。月之暗面团队创始人杨植麟，本科毕业于清华大学计算机科学与技术系，博士就读于全美自然语言处理排名第一的卡内基梅隆大学语言技术研究所（LTI），杨植麟本人学术引用量自 2019 年起已超 2 万余次。在算法和工程领域，月之暗面囊括了自然语言处理、计算机视觉、强化学习、基础设施等方面的新生代人才，创始团队的核心成员参与了 Google Gemini、Google Bard、盘古 NLP、悟道等多个大模型的研发，多项核心技术被 Google PaLM、Meta LLaMa、Stable Diffusion 等主流产品采用。

建议关注：润泽科技（Kimi+算力）、福昕软件（Kimi+文档处理）、金山办公（Kimi+办公套件）、万兴科技（Kimi+视频创意）、金蝶国际（Kimi+企业管理）、彩讯股份（Kimi+邮箱）、拓尔思（Kimi+公文写作）、华宇软件/通达海（Kimi+法律文件）等。

2. 阶跃星辰：发布万亿参数 MoE 大模型预览版

阶跃星辰 Step-1V 多模理解能力突出，并蓄力发布万亿参数模型。通用大模型创业公司阶跃星辰成立于 2023 年 4 月。2024 年 3 月 23 日，公司在上海举行的 2024 全球开发者先锋大会期间正式对外亮相，阶跃星辰创始人、CEO 姜大昕博士在大会开幕式上对外发布了 Step 系列通用大模型。Step-1V 千亿参数多模态大模型的多模理解能力突出，可以精准描述和理解图像中的文字、数据、图表等信息，并根据图像信息实现内容创作、逻辑推理、数据分析、视频理解等多项任务。该模型在中国权威的大型模型评估平台“司南”（OpenCompass）多模态模型评测榜单中位列第一，性能比肩 GPT-4V。此次大会上还发布了 Step-2 万亿参数 MoE 语言大模型预览版，该模型采用 MoE 架构，聚焦深度智能的探索，并提供 API 接口给部分合作伙伴试用。训练万亿参数模型体现了阶跃星辰的核心技术能力和探索通用人工智能的决心。

图6. Step-1V 的优势

资料来源：阶跃星辰官网，国投证券研究中心

创始团队坚定投入攀登 Scaling law，在算力/数据/算法/系统四大要素布局。创始人和 CEO 是前微软全球副总裁、微软亚洲互联网工程院首席科学家姜大昕博士，核心创始团队包括系统负责人朱亦博博士和数据负责人焦斌星博士。姜大昕是自然语言处理领域的全球知名专家，在机器学习、数据挖掘、自然语言处理和生物信息学等领域拥有丰富的研究及工程经验。朱亦博拥有多次单集群万卡以上的系统建设与管理实践经验。焦斌星此前担任微软必应引擎核心搜索团队负责人，负责利用数据挖掘和 NLP 算法优化索引和搜索质量。阶跃星辰在大模型技术路径上坚定投入攀登 Scaling Law。根据阶跃星辰数据，等效 A800 万卡单一集群，高效稳定的训练，十万亿 tokens 高质量的数据，加上驾驭新颖的 MoE 架构，任何一环出现短板，Scaling law 就攀登不上去。因此公司自成立起，在算力、数据、算法和系统这四大要素上综合布局：

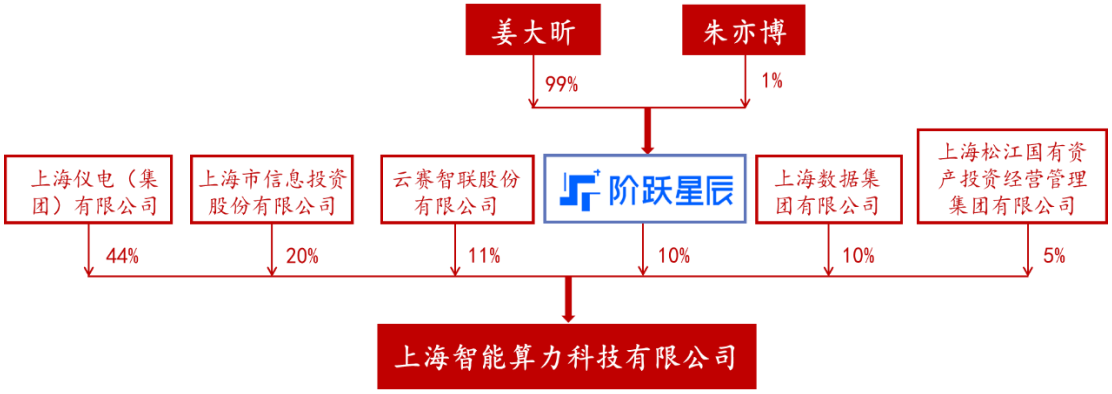
1）**算力：**通过自建机房+租用算力，积极进行算力储备。前瞻布局算力资源，阶跃星辰出资 2 亿元人民币投资上海智能算力科技有限公司并持股 10%。（该公司大股东为上海仪电集团，持股 44%，云赛智联持股 11%。）

2）**系统：**实践过单集群万卡以上的系统建设与管理。训练千亿模型的 MFU（有效算力输出）达 57%。

3）**数据：**数据团队核心骨干来自必应搜索引擎，曾支持全球 100 多种语言，为 200 多个国家和地区提供服务。对全球互联网高质量语料的分布有深入了解。并建立起强大的数据处理和知识图谱流水线。

4）**算法：**团队不仅能驾驭各种架构，比如万亿参数的 MoE 架构，并且对大模型的认知以及发展路线有深刻洞察。

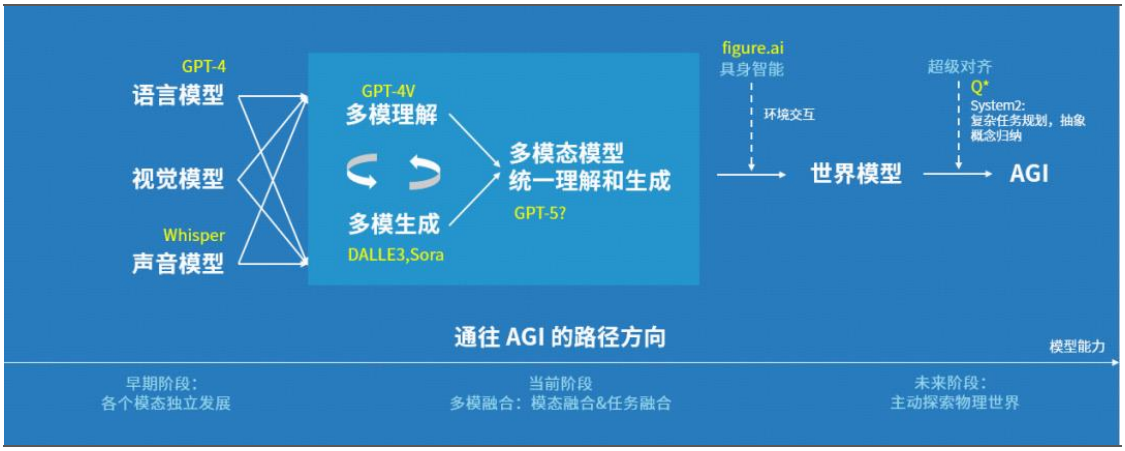
图7. 上海智能算力科技有限公司股权结构



资料来源：爱企查，国投证券研究中心

多模理解和生成的统一是通往 AGI 的必经之路。阶跃星辰认为，模型的演化必然会经历“单模->多模->世界模型”三个阶段。**早期阶段**是语言、视觉和声音各个模态独立发展，各个模型学习如何更好表征各个模态。**当前阶段**是多种模态走向融合，无论是语言、视觉还是声音，现在都可以映射到同一个空间加以表征。尽管当前阶段多种模态开始走向融合，但是仍然存在一个问题——理解模型和生成模型是分开发展的。其造成的结果就是理解模型的理解能力强而生成能力弱（比如 GPT-4V），或者生成模型的生成能力强但理解能力弱（比如 Sora ）。理解和生成必须统一在一个模型里面，即多模理解和生成的统一是通向 AGI 的必经之路。**在未来阶段**，有了理解和生成的统一，就可以进一步和具身智能结合起来，形成一个世界模型。再进一步，在世界模型的基础上加入复杂任务的规划能力和抽象概念的归纳能力，就真正演化到了 AGI 的阶段。

图8. 通往 AGI 的路径方向



资料来源：阶跃星辰官方公众号，国投证券研究中心

基于自研大模型底座，阶跃推出了两款面向 C 端用户的 AI 应用产品：1）跃问（StepChat）是基于公司千亿级参数模型所研发的免费 AI 聊天机器人，定位为个人效率助手，主要功能包括 AI 对话聊天、图片内容理解、文档信息总结、网页内容分析、联网在线搜索等。**2）冒泡鸭**是基于公司千亿级参数模型研发的免费 AI 开放世界平台，它提供了覆盖拟人、工具、内容、游戏、娱乐等多个领域的海量智能体，设定了十亿种剧情和角色，用户可与其进行多场景的角色扮演体验。冒泡鸭 AI 依靠超长的上下文记忆能力和实时联网搜索的能力，能够深度理解用户意图，并提供即时、准确、个性化的回复和选择。

图9. 个人效率助手—跃问

资料来源：阶跃星辰官网，国投证券研究中心

图10. AI 开放世界平台—冒泡鸭

资料来源：阶跃星辰官网，国投证券研究中心

建议关注：云赛智联（阶跃星辰+算力）、万兴科技（阶跃星辰+应用）等。

行业评级体系

收益评级：
领先大市 —— 未来 6 个月的投资收益率领先沪深 300 指数 10%及以上；
同步大市 —— 未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-10%至 10%；
落后大市 —— 未来 6 个月的投资收益率落后沪深 300 指数 10%及以上；
风险评级：
A —— 正常风险，未来 6 个月的投资收益率的波动小于等于沪深 300 指数波动；
B —— 较高风险，未来 6 个月的投资收益率的波动大于沪深 300 指数波动；

分析师声明

本报告署名分析师声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

本公司具备证券投资咨询业务资格的说明

国投证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向本公司的客户发布。

免责声明

本报告仅供国投证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准，如有需要，客户可以向本公司投资顾问进一步咨询。

在法律许可的情况下，本公司及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或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦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。

本报告的估值结果和分析结论是基于所预定的假设，并采用适当的估值方法和模型得出的，由于假设、估值方法和模型均存在一定的局限性，估值结果和分析结论也存在局限性，请谨慎使用。

国投证券股份有限公司对本声明条款具有惟一修改权和最终解释权。

国投证券研究中心

深圳市
地　址：深圳市福田区福田街道福华一路 119 号安信金融大厦 33 楼
邮　编：518046
上海市
地　址：上海市虹口区东大名路 638 号国投大厦 3 层
邮　编：200080
北京市
地　址：北京市西城区阜成门北大街 2 号楼国投金融大厦 15 层
邮　编：100034