2025 年 07 月 22 日

行业及产业

电子

全球最大参数模型 Kimi K2 发布

——人工智能月度跟踪

强于大市

一年内行业指数与沪深 300 指数对比走势：

资料来源：聚源数据，爱建证券研究所

相关研究

《电子行业周报：国产内存龙头启动 IPO 辅导，产业链迎来发展机遇》2025-07-14

《电子行业周报：移动电源 3C 认证加速品牌龙头发展》2025-07-07

《电子行业周报：AI Glasses 市场持续爆发——2025/06/23-2025/06/27》2025-07-01

《电子行业周报：Switch 2 开启新一轮成长周期——2025/06/16-2025/06/20》2025-06-24

《电子行业周报： 低轨卫星市场正处于持续爆发周期——2025/06/09-2015/06/13》2025-06-16

证券分析师

许亮

S0820525010002

0755-83562506

xuliang@ajzq.com

投资要点：

- **2025 年 7 月 11 日，月之暗面(Moonshot AI)发布采用 MoE 架构的大模型 Kimi K2，并同步开源。**这款模型总参数量达 1 万亿，每次推理仅激活 320 亿参数，在代码能力和通用 Agent 任务处理上表现突出，同时凭借架构优化实现了性能与成本的平衡，输入输出价格更具优势。

- **月之暗面由杨植麟于 2023 年 4 月创立，聚焦探索能源转化为智能的最优路径，其产品迭代轨迹清晰。**2023 年 10 月推出首款智能助手 Kimi，以 Transformer-XL 等算法实现 20 万汉字输入的长文本处理突破；2024 年持续升级，先后实现 200 万字无损上下文能力、拓展多模态场景及工具调用功能，同年年底发布对标 OpenAI o1 的 k0-math 数学模型与 k1 视觉思考模型；2025 年 1 月推出的 k1.5 多模态模型，在 Long CoT 模式下能力达 o1 水平，Short CoT 模式下领先 GPT-4o 和 Claude 3.5。2025 年 7 月 11 日，公司发布 KimiK2 大模型并同步开源。

- **Kimi K2 采用 64 头注意力+384 专家 MoE 设计，相比 DeepSeek V3/R1 更具效能。**这一设计减少了自注意力计算负担，在加快推理速度、提升 128K 长文本处理效率的同时，扩展了知识覆盖范围和多任务适配性。训练端，借助 Muon Clips 优化器完成 15.5 万亿 Tokens 的高效训练，全程无峰值且持续提升 Token 利用效率；此外，为解决工具交互数据稀缺问题，它采用大规模 Agentic 数据合成策略，学习复杂工具调用能力。

- **Kimi K2 是性能与成本平衡的大规模模型（总参数量达 1 万亿，每次推理仅激活 320 亿参数）。**其训练成本覆盖算力（如 GPU 集群）、数据准备、算法调优等核心环节，相较于 GPT-4.5、SparkDesk-v1.1、Llama-3.1 等模型，Kimi K2 通过更精准的参数激活与架构优化控制成本，设计更聚焦实际落地效率。目前 Kimi K2 输入、输出价格分别为 0.6$/Million Tokens、2.5$/Million Tokens。

- **Kimi K2 在自主编程、工具调用、数学推理等复杂任务上表现突出，应用场景广泛。**代码生成速度与软件开发效率显著提升，数学推理与科研计算精度加速研究进程，创意写作质量（文学评测 SOTA）更是高居榜首。技术落地推动硬件升级，既拉动高性能 GPU/TPU 及边缘计算设备的需求与性能跃升，又优化电子供应链、降低中小企业 AI 应用门槛。

- **Kimi K2 的发布标志着国产 AI 在全球竞争中的全新突破。**Kimi K2 强大的代码能力、Agent 任务处理能力和开源策略，为开发者与用户提供了无限可能。无论是科研人员、开发者还是普通用户，都可以通过 Kimi K2 探索 AI 的更多潜力。

- **风险提示：1）先进算力芯片限制加强 2）下游应用需求不及预期 3）国产模型迭代升级迟缓**

目录

图表目录

1. 全球最大参数模型 Kimi K2 发布

2025 年 7 月 11 日,月之暗面(Moonshot AI)发布大模型 Kimi K2,并宣布同步开源。Kimi K2 是一款采用 MoE 架构的基础模型,总参数量高达 1 万亿 Token,每次推理激活 320 亿参数,尤其在代码能力和通用 Agent 任务处理方面表现出色。

月之暗面由杨植麟于 2023 年 4 月创立,聚焦探索能源转化为智能的最优路径,其产品迭代轨迹清晰。2023 年 10 月 9 日,公司推出首款智能助手 Kimi。该产品基于自研 Moonshot 大模型架构,采用 Transformer-XL、XLNet 等先进算法,实现长文本处理能力的突破性进展,是全球首个支持输入 20 万汉字的智能助手。

图表 1:Kimi 发展史梳理

资料来源:Kimi 官网,36Kr,C114,爱建证券研究所

2024 年 3 月,Kimi 推出 200 万字无损上下文能力,巩固了其在长文本处理领域的优势,显著提升学术文献、法律卷宗等超长文本的解析与处理效率。4 月 Kimi 持续迭代,新增常用语及内置提示词提升交互效率;增加语音功能与搜索溯源,拓展多模态场景;API 支持 Tool Calling,实现工具调用落地。

2024 年 11-12 月,公司先后发布 k0-math 数学模型与 k1 视觉思考模型,两者均对标 OpenAI o1。2025 年 1 月 20 日,Kimi 发布 k1.5 多模态模型。在 Long CoT(长思维链)模式下,其数学、代码、多模态推理能力达到 Open AI o1 水平;而在 Short CoT(短思维链)模式下,k1.5 领先 GPT-4o 和 Claude3.5。

2025 年 7 月 11 日,月之暗面(Moonshot AI)发布大模型 Kimi K2,并宣布同步开源。其开源版本包括 Kimi-K2-Base(未经指令微调,适配科研及自定义场景)与 Kimi-K2-Instruct(通用指令微调版,面向对

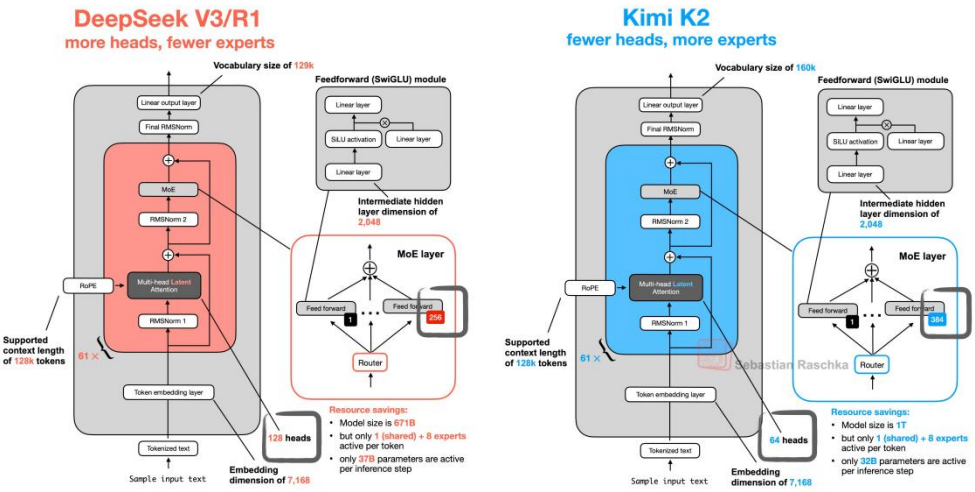
话与智能体应用)。

2. 深入剖析 Kimi K2

2.1 Kimi K2 从框架到训练的技术突破

Kimi K2 框架设计与 DeepSeek V3/R1 具有相似性，但核心参数配置差异显著。K2 配备 64 个 attention heads 与 384 个专家(Experts)，而 DeepSeek V3/R1 为 128 个 attention heads 与 256 个专家。这一设计的优势体现在两方面：减少 attention heads 数量降低了自注意力计算复杂度，加快推理速度的同时提升 128K 长文本处理效率；增加专家数量则扩展知识覆盖范围，强化多任务适配能力。

图表 2：Kimi K2 框架类似于 DeepSeek V3/R1

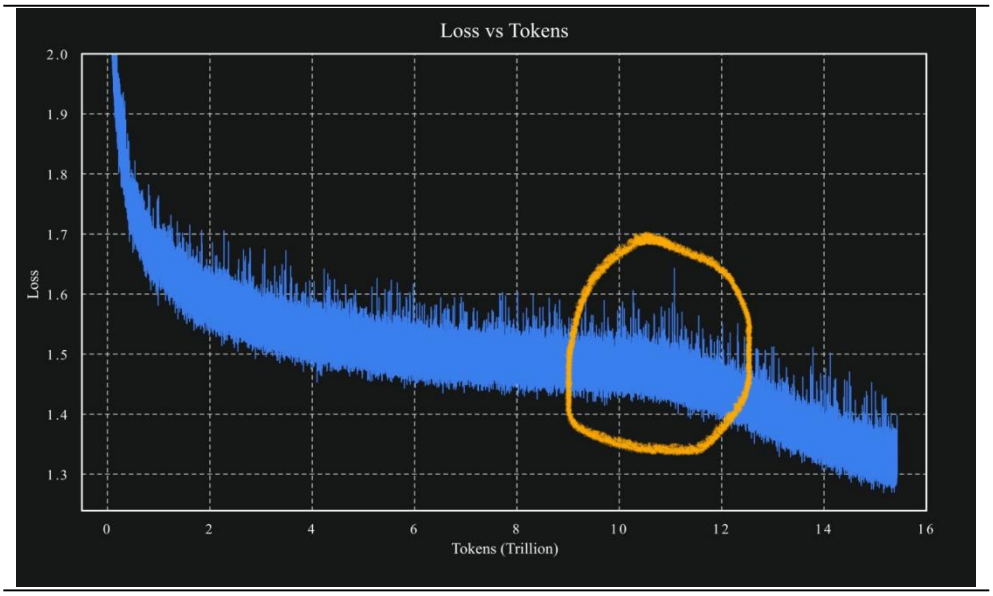


资料来源：Sebastian Raschka，爱建证券研究所

在有限预训练数据集与固定模型配置约束下，Token 的优化器对提升大语言模型（LLM）训练至关重要。此前，Moonlight 已验证，Muon 优化器的性能显著优于传统的 Adam W，成为更高效的训练工具。

但当模型向百万亿参数级规模扩展时，Muon 优化器难以应对新的挑战。伴随参数与训练数据量增加，注意力机制中“查询（Query）”与“关键（Key）”的计算结果（Logits）易出现数值失控飙升，导致训练剧烈波动甚至中断。Kimi K2 借助 MuonClips 优化器，实现了万亿参数模型的高效训练(15.5 Trillion Tokens)，全程无训练峰值，保障了模型稳定性。同时，Kimi K2 通过 MuonClips 技术显著提高了 Token 的利用效率，使其能够在有限的数据集上达到最佳性能。

图表 3：Kimi K2 实现百万亿参数模型高效训练

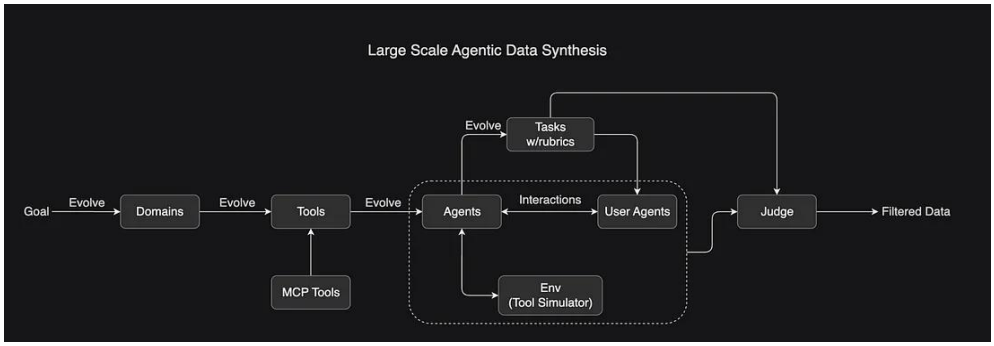


资料来源：Yuchen Jin，36Kr，爱建证券研究所

为了解决真实工具交互数据稀缺的问题，Kimi K2 采用大规模 Agentic 数据合成策略，并让模型学习复杂工具调用（Tool Use）能力。Kimi K2 目前提供两个版本：Kimi K2 Base 用于研究与微调，Kimi K2 Instruct 适用于通用任务与 Agent 部署。

此外，Kimi K2 不仅在可验证任务上（代码、数学）强化学习，还通过引入自我评价机制（self-judging），解决了不可验证任务的奖励稀缺问题。

图表 4：Kimi K2 采用大规模 Agentic 数据合成策略



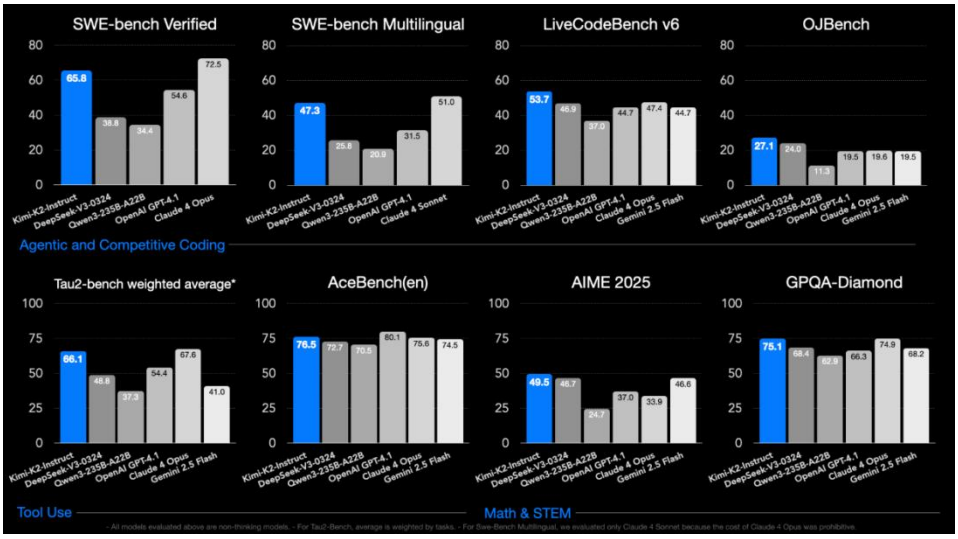
资料来源：51CTO，爱建证券研究所

2.2 Kimi K2 性能达到行业领先水平

Kimi K2 性能达到行业领先水平。该模型聚焦自主编程（Agentic Coding）、工具调用（Tool Use）、数学（Math&STEM）等复杂智能任务，旨在为这类任务提供精准、高性能的执行路径与问题求解方案。其核心优势体现在三大能力的基准测试表现突出，依托大规模参数量与高

效推理激活机制，兼顾强大算力支撑与精准任务适配。

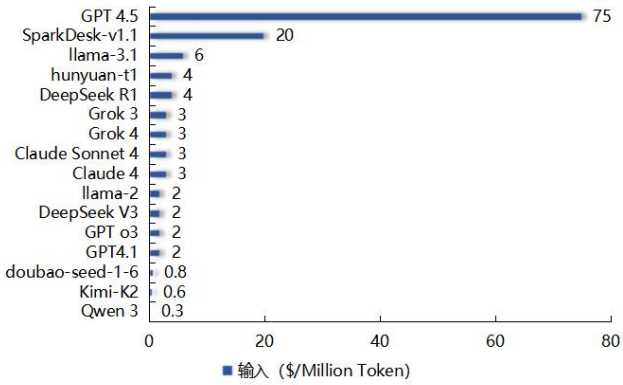
图表 5：Kimi K2 性能达到行业领先水平



资料来源：Kimi 官网，爱建证券研究所

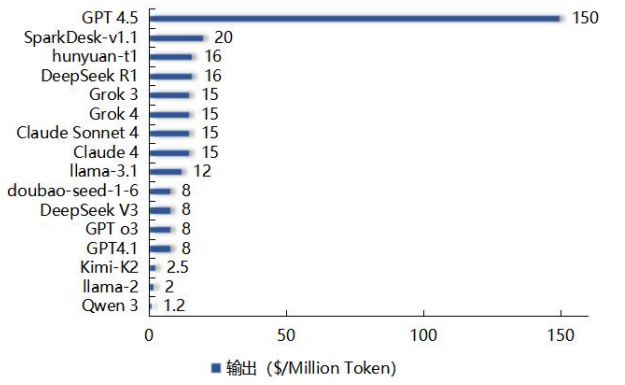
Kimi K2 是性能与成本平衡的大规模模型（总参数量达 1 万亿，每次推理仅激活 320 亿参数）。其训练成本覆盖算力（如 GPU 集群）、数据准备、算法调优等核心环节，相较于 GPT-4.5、SparkDesk-v1.1、Llama-3.1 等模型，Kimi K2 通过更精准的参数激活与架构优化控制成本，设计更聚焦实际落地效率。目前 Kimi K2 输入、输出价格分别为 0.6$/Million Tokens、2.5$/Million Tokens。

图表 6：Kimi K2 输入价格优势明显



资料来源：火星 AI，Info Q，爱建证券研究所

图表 7：Kimi K2 输出价格优势明显



资料来源：火星 AI，Info Q，爱建证券研究所

2.3 Kimi K2 的应用场景广泛

Kimi K2 模型的应用场景广泛，涵盖代码与软件研发、数据推理及科研辅助、创意性写作等领域。

在代码与软件开发方面，K2 专为复杂代码任务设计，支持一次性阅读上万行源码或整份需求文档，生成完整项目框架。开发者可以利用 K2 快速搭建项目框架，减少重复性工作，提高开发效率。

在数学推理与科研辅助领域，Kimi K2 在 AIME、MATH 等基准测试中领先主流开源模型。用户可一次性输入整篇论文、竞赛题或复杂公式，模型会给出分步推导。科研人员可以利用 K2 加速研究进程，发现新的科学规律。

在短篇小说创意写作评估中，Kimi K2 模型位列榜首。这一突破不仅证明了 Kimi K2 在文学创作方面的卓越实力，其 ELO Score 达 1638.2，超过 Claude、DeepSeek 等模型。

图表 8：Kimi K2 在创意性写作位列榜首

资料来源：51CTO 技术栈，爱建证券研究所

Kimi K2 的发布标志着国产 AI 在全球竞争中的全新突破。Kimi K2 强大的代码能力、Agent 任务处理能力和开源策略，为开发者与用户提供了无限可能。无论是科研人员、开发者还是普通用户，都可以通过 Kimi K2 探索 AI 的更多潜力。

3. 风险提示

1）先进算力芯片限制加强
2）下游应用需求不及预期
3）国产模型迭代升级迟缓

爱建证券
AIJIAN SECURITIES

爱建证券有限责任公司

上海市浦东新区前滩大道 199 弄 5 号
电话：021-32229888　　传真：021-68728700　　服务热线：956021
邮政编码：200124　　邮箱：ajzq@ajzq.com　　网址：http://www.ajzq.com

评级说明

投资建议的评级标准

报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 个月内的相对市场表现，也即：以报告发布日后的 6 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场：沪深 300 指数（000300.SH）；新三板市场：三板成指（899001.CSI）（针对协议转让标的）或三板做市指数（899002.CSI）（针对做市转让标的）；北交所市场：北证 50 指数（899050. BJ）；香港市场：恒生指数（HIS.HI）；美国市场：标普 500 指数（SPX.GI）或纳斯达克指数（IXIC.GI）。

股票评级

买入	相对同期相关证券市场代表性指数涨幅大于 15%
增持	相对同期相关证券市场代表性指数涨幅在 5%～15%之间
持有	相对同期相关证券市场代表性指数涨幅在-5%～5%之间
卖出	相对同期相关证券市场代表性指数涨幅小于-5%

行业评级

强于大市	相对表现优于同期相关证券市场代表性指数
中性	相对表现与同期相关证券市场代表性指数持平
弱于大市	相对表现弱于同期相关证券市场代表性指数

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告采用信息和数据来自公开、合规渠道，所表述的观点均准确地反映了我们对标的证券和发行人的独立看法。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法可能存在局限性，请谨慎参考。

法律主体声明

本报告由爱建证券有限责任公司（以下统称为“爱建证券”）证券研究所制作，爱建证券具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管。

本报告是机密的，仅供我们的签约客户使用，爱建证券不因收件人收到本报告而视其为爱建证券的签约客户。本报告中的信息均来源于我们认为可靠的已公开资料，但爱建证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供签约客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，爱建证券及其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测后续可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，爱建证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

版权声明