

策略周评 20260802

月之暗面全面开源 Kimi K3

2026 年 08 月 02 日

本周 AI 要闻

- **事件 1:** 7 月 31 日-8 月 3 日，在 2026 ChinaJoy 上，AI 与游戏技术深度融合成为全场焦点，AI 产业在底层硬件与内容范式上迎来关键进展。芯片端，Arm 与 Unity 中国达成合作，将专用神经加速器集成至新一代 GPU；高通预告 9 月推出搭载新一代神经渲染技术的骁龙旗舰平台，从硬件底层重塑 AI 图形计算能力。产业应用层面，Unity 中国指出 AI 正重构内容创作与渲染流程，超分辨率及插帧技术显著降低算力成本。
- **事件 2:** 7 月 27 日，月之暗面全面开源迄今最强模型 Kimi K3，同步公开模型权重、技术报告及三项核心训练 Infra 技术（MoonEP、FlashKDA、AgentEnv）。Kimi K3 是全球首个开源的 3 万亿参数 MoE 模型，具备原生视觉理解与百万 token 上下文能力，整体扩展效率较 K2 提升约 2.5 倍。开源当日，华为昇腾、阿里云真武超节点等国内厂商，以及 Nebius、Cursor 等海外平台均实现 Day0 适配，生态响应迅速。
- **事件 3:** 7 月 30 日，腾讯发布 WorkBuddy 最新版本，打通腾讯文档，推出面向 Agent 时代的“人机双写”协同编辑能力。记者实测显示，用户通过自然语言指令即可实现对 Word、Excel、PPT 等文档的实时批量修改，无需手动查找入口或记忆复杂公式，AI 直接调用并就地保存。该功能覆盖文字、表格、演示全场景，同时支持一键上传腾讯文档实现多人在线 AI 共创，将“人人协同”延伸至“人机协同”乃至“机机协同”。目前该产品正探索个人与企业付费模式，标准版至旗舰版月费 70 至 700 元不等。

（信息来源：财联社、AI daily、上海证券报等）

周度观点

■ 月之暗面全面开源 Kimi K3，AI 进入 3T 模型时代

（1）AI 产业链持续围绕“模型能力跃迁—算力架构升级—应用生态扩展”展开演进。模型端，Kimi K3 开源推动开放模型迈入 3 万亿参数规模，带动模型能力与生态适配加速提升；算力端，LPDDR6 等新一代存储技术推进，AI 硬件向高性能、低功耗方向升级；应用端，WorkBuddy、千问办公等智能体产品加速落地，推动 AI 从辅助工具向 workflow 重构演进。

（2）算力与基础设施方面，本周 AI 产业竞争焦点进一步向硬件架构优化与终端计算能力提升方向延伸，底层芯片、存储以及边缘侧算力设施成为支撑 AI 规模化应用的重要环节。一方面，长鑫存储 LPDDR6 研发验证接近尾声，有望推动国产低功耗内存实现新一轮迭代。该产品设计速率达到 12800Mbps，在性能、功耗及可靠性方面较 LPDDR5X 进一步提升，有望满足 AI 手机等终端设备对高带宽、低功耗存储的需求。另一方面，2026 ChinaJoy 期间，Arm 与 Unity 中国围绕神经技术展开合作，通过将专用神经加速器引入新一代 GPU，提升 AI 图形处理效率；高通也计划推出搭载神经渲染技术的旗舰移动平台，推动 AI 超分、插帧等技术落地。

（3）模型技术方面，本周 AI 领域核心变化集中于开源大模型能力持续逼近前沿水平，以及智能体应用能力加速成熟。随着高参数规模模型和低成本推理模型不断涌现，模型竞争正从单纯追求参数规模转向性能、开放生态与应用效率的综合比拼。当前 AI 正处于前沿模型能力加速开放、智能体从实验探索走向实际应用阶段。月之暗面正式开源 Kimi K3 模型权重及相关训练基础设施，成为全球首个开放 3 万亿参数级模型。该模型采用 MoE 架构，支持百万 Token 上下文窗口，面向长程编程、知识工作和复杂推理场景，并获得华为昇腾、阿里云等生态适配，推动国产模型与算力基础设施协同发展。DeepSeek 开放 DeepSeek-V4-Flash 正式版 API 公测，重点提升 Agent 任务处理能力。在智能体测试中，其性能大幅提升，同时通

证券分析师 陈刚

执业证书：S0600523040001

cheng@dwzq.com.cn

证券分析师 谢立昕

执业证书：S0600525080001

xielx@dwzq.com.cn

相关研究

《反弹的时间和空间研判》

2026-08-02

《十大未来产业系列之七——原子级制造》

2026-07-31

过低成本 API 定价降低开发者使用门槛，加速模型在代码、长文本处理等场景中的应用。整体来看，开源模型能力提升正在降低 AI 应用开发门槛，推动产业竞争重点由“模型可用性”进一步转向“生态构建与场景落地能力”。

(4) 应用方面，本周 AI 应用端持续向智能体驱动的生产力场景落地方向推进，办公、企业服务等高频场景成为 AI 商业化探索的重要入口。相比此前以辅助问答为主的应用形态，当前 AI 正加速向任务执行和流程协同演进，处于从单点工具应用向 Agent 重构 workflow 阶段过渡的阶段。腾讯 WorkBuddy 完成版本升级，打通腾讯文档并推出“人机双写”协同编辑能力，进一步推动 AI 与办公软件深度融合。与此同时，阿里“千问办公”开启内测，整合 Qoder Work、悟空、Mule Run 等多款智能体产品，布局从个人办公到企业服务的 AI 办公生态。此外，字节跳动启动 ToB 业务组织调整，将飞书产品团队并入豆包体系，并整合飞书销售与火山引擎团队，强化 AI 模型、云服务与企业应用之间的协同。

(5) 我们认为 A 股风格切换已经是“进行时态”，但并不意味着科技行情彻底结束，只是“全面押注科技”不再是最优解。从产业趋势角度，中长期来看需要增加对 AI 中下游环节的重视度；随资金面矛盾边际改善，上游硬件方向在下半年有望迎来分化修复机会，优选量增逻辑主导的方向，以及供给壁垒高且难找到替代技术路线的细分，“涨价线”中部分环节高估了远期供需矛盾，需要适当降低预期、去伪存真。

本周报告推荐

- 【东吴有色】AI 金属乘风破浪（二）：钨&稀土，配额限制下的供需错配
- 【东吴地产】地产+AI 工具系列报告之七：从 AI 投委会到可用的投研平台——AlphaChain 产品化
- 【东吴宏观】如何用 AI 自动写研报？——搭建 Hermes 架构的证券分析师“超级助理”
- 【东吴汽车】AI 系列深度 01 期：从数字 AI 通用底座到物理 AI 真实世界闭环
- 【东吴汽车】AI 系列深度 02 期：AGI 是什么？
- 【东吴汽车】AI 系列深度 03 期：数字 AI 与物理 AI 的边界在哪里？
- 【东吴汽车】AI 系列深度 04 期：什么是表征？
- 【东吴汽车】AI 系列深度 05 期：模型如何学习表征并实现对齐？
- 【东吴汽车】AI 系列深度 06 期：Transformer 如何从序列任务走向基座模型？
- 【东吴汽车】AI 系列深度 07 期：CNN 与 ViT 两类视觉骨干有何差异？
- 【东吴汽车】AI 系列深度 08 期：GAN 与 Diffusion 两类生成范式有何差异？
- 【东吴汽车】AI 系列深度 09 期：从 LLM 到 VLM 再到 VLA 意味着什么？
- 【东吴汽车】AI 系列深度 10 期：预训练与后训练如何塑造大模型？
- 风险提示：美联储降息不及预期，AI 技术发展不及预期

1. 重要数据跟踪

表1：美国主要科技股周度表现（2026/07/27-2026/07/31）

代码	证券简称	收盘价美元	市值亿美元	周涨跌幅	2026 年初至今涨跌幅
NVDA.O	英伟达(NVIDIA)	200.75	48,582	-2.94%	7.77%
MSFT.O	微软(MICROSOFT)	464.72	34,508	21.75%	-3.48%
AAPL.O	苹果(APPLE)	308.91	45,083	-7.24%	13.84%
GOOGL.O	谷歌(ALPHABET)-A	356.13	43,583	11.38%	13.93%
AMZN.O	亚马逊(AMAZON)	271.58	29,293	17.00%	17.66%
META.O	脸书(META PLATFORMS)	556.71	14,182	-6.47%	-15.51%
TESLA.O	特斯拉(TESLA)	311.21	12,291	-0.58%	-30.80%

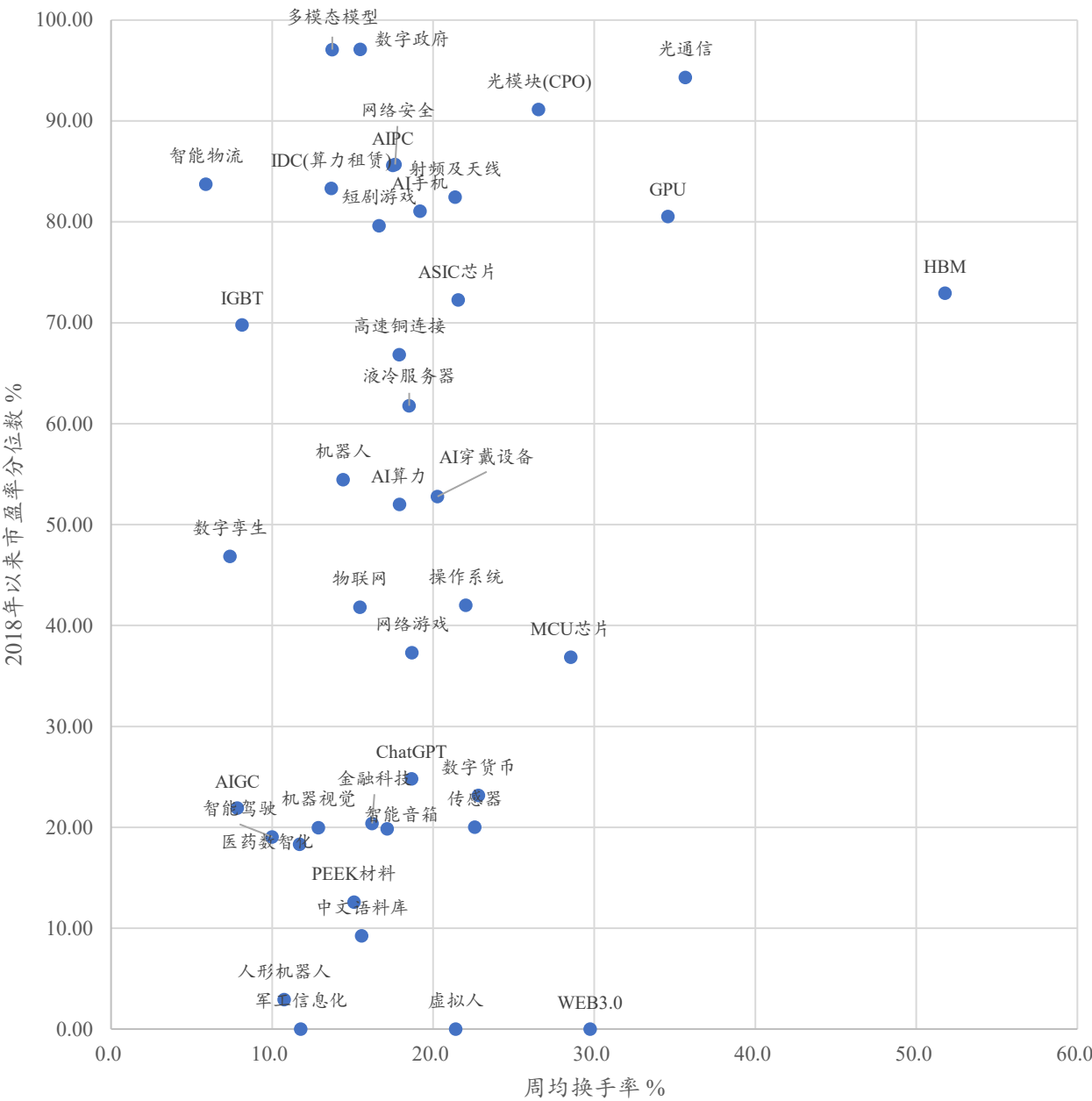
数据来源：WIND，东吴证券研究所

图1：主要科技指数的收益表现（归一化）



数据来源：WIND，东吴证券研究所

图2：A股主要AI概念指数情绪（截至2026年07月31日）



数据来源：WIND，东吴证券研究所

2. 人工智能报告推荐

表2：东吴人工智能报告推荐一览

发布时间	报告名称	核心内容	相关标的
2026/7/29	【东吴有色】AI 金属乘风破浪（二）：钨&稀土，配额限制下的供需错配	钨&稀土，我国唯二受开采配额管理的优势金属，出口管制打开国产替代空间。算力需求催生半导体制造材料体系升级，拉动钨&稀土景气需求。	钨：中钨高新，厦门钨业，章源钨业，翔鹭钨业，佳鑫国际资源等。稀土：中国稀土、盛和资源、华宏科技等。
2026/7/31	【东吴地产】地产+AI 工具系列报告之七：从 AI 投委会到可用的投研平台——AlphaChain 产品化	AI 投研产品化的关键约束，主要来自深度、效率与问责三方面。结构化决策卡有助于降低长报告的使用门槛。决策跟踪与战绩榜是平台形成长期可信度的重要基础。机构化落地仍需数据质量、安全与合规机制共同支撑。	
2026/7/26	【东吴宏观】如何用 AI 自动写研报？——搭建 Hermes 架构的证券分析师“超级助理”	本套 AI 研报生成 Skill 通过多体系协作实现了研报撰写从数据准备到编辑检查的全链条自动化，实际案例证明其产出在专业水准上与真人研究员高度一致，同时显著释放了研究员在新闻采集与结构性写作环节的时间投入。	
2026/7/27	【东吴汽车】AI 系列深度 01 期：从数字 AI 通用底座到物理 AI 真实世界闭环	Transformer 推动数字 AI 技术发展从专用架构到统一底座。物理 AI 正从复用数字 AI 能力转向补齐自身底层能力。	
2026/7/28	【东吴汽车】AI 系列深度 02 期：AGI 是什么？	AGI 的核心研究问题，是如何定义、衡量并判断通用人工智能何时到来。从机构判断看，AGI 并非一个单一时间点，而更像能力阶梯。AGI 的衡量与经济价值仍缺统一答案。	
2026/7/28	【东吴汽车】AI 系列深度 03 期：数字 AI 与物理 AI 的边界在哪里？	数字 AI 与物理 AI 并非完全对称的二分概念，更适合沿“物理闭环程度”理解。两类 AI 最根本差异体现在部署形态和安全约束。从通往 AGI 的视角看，机器人正成为数字 AI 与物理 AI 的交汇点。	
2026/7/29	【东吴汽车】AI 系列深度 04 期：什么是表征？	表征是模型对现实世界信息形成的内部抽象表示，是人工智能进行理解、推理与预测的基础。落到物理 AI，表征视角可解释自动驾驶与机器人路线分歧。	
2026/7/29	【东吴汽车】AI 系列深度 05 期：模型如何学习表征并实现对齐？	嵌入的本质，是把文本、图像、商品、用户等对象映射为数字向量，并使语义相近的对象在向量空间中距离更近。自监督学习是通用表征规模化形成的关键机制。	
2026/7/30	【东吴汽车】AI 系列深度 06 期：Transformer 如何从序列任务走向基座模型？	RNN 的基本思路是把历史压缩进一个随时间更新的隐藏状态。Transformer 替代传统 RNN 主干，并不意味着循环思想消失。	

2026/7/30	【东吴汽车】AI 系列深度 07 期：CNN 与 ViT 两类视觉骨干有何差异？	视觉骨干网络是机器视觉系统的核心模块，其作用是把原始像素转化为高层语义特征，并决定分类、检测、分割等下游任务上限。CNN 的优势来自三项视觉归纳偏置：局部连接、权值共享和平移等变性。	
2026/7/31	【东吴汽车】AI 系列深度 08 期：GAN 与 Diffusion 两类生成范式有何差异？	生成模型的核心目标，是学习数据背后的整体分布，并据此生成此前不存在的新样本。GAN 与 Diffusion 的路线差异集中在四个维度：采样速度、训练稳定性、样本多样性和可控性。	
2026/7/31	【东吴汽车】AI 系列深度 09 期：从 LLM 到 VLM 再到 VLA 意味着什么？	LLM、VLM、VLA 可理解为同一技术脉络上逐层扩展模态和输出能力的三类大模型。LLM 以文本为输入输出，主要解决语言理解、生成和推理；VLM 在语言底座上加入视觉理解，使图像和文字在同一语义空间交互；VLA 进一步把动作作为输出模态，使模型能够驱动机器人或车辆在物理世界中行动。三者存在继承关系，但不能简单归结为同一种 next-token 机制。	
2026/8/1	【东吴汽车】AI 系列深度 10 期：预训练与后训练如何塑造大模型？	预训练与后训练是理解大模型能力形成的一组关键框架。数字 AI 与物理 AI 在训练框架上同构、在数据和反馈上分立。	

数据来源：东吴证券研究所整理（报告筛选标准为：本周公开已发布，对相关标的进行长期跟踪）

免责声明

东吴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。

本研究报告仅供东吴证券股份有限公司(以下简称“本公司”)的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下,本报告中的信息或所表述的意见并不构成对任何人的投资建议,本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下,东吴证券及其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并进行交易,还可能为这些公司提供投资银行服务或其他服务。

市场有风险,投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息,本公司力求但不保证这些信息的准确性和完整性,也不保证文中观点或陈述不会发生任何变更,在不同时期,本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有,未经书面许可,任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的,应当注明出处为东吴证券研究所,并注明本报告发布人和发布日期,提示使用本报告的风险,且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的,应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期(A 股市场基准为沪深 300 指数,香港市场基准为恒生指数,美国市场基准为标普 500 指数,新三板基准指数为三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的),北交所基准指数为北证 50 指数),具体如下:

公司投资评级:

- 买入:预期未来 6 个月个股涨跌幅相对基准在 15%以上;
- 增持:预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间;
- 中性:预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间;
- 减持:预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间;
- 卖出:预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级:

- 增持:预期未来 6 个月内,行业指数相对强于基准 5%以上;
- 中性:预期未来 6 个月内,行业指数相对基准-5%与 5%;
- 减持:预期未来 6 个月内,行业指数相对弱于基准 5%以上。

我们在此提醒您,不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系,表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况,如具体投资目的、财务状况以及特定需求等,并完整理解和使用本报告内容,不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号

邮政编码: 215021

传真: (0512) 62938527

公司网址: <http://www.dwzq.com.cn>