

A I 前沿发展日报 | 2 0 2 6 - 0 7 - 2 0 (A s i a)

日期：2 0 2 6 - 0 7 - 2 0 ；覆盖窗口：2 0 2 6 - 0 7 - 1 7 0 0 : 0 0 至 2 0 2 6 - 0 7 - 2 0 a i) ，周末低发布密度窗口，纳入经实时复核的关键新增。

今日要点

- 月之暗面发布 K i m i K 3 : 2 . 8 万亿总参数、原生视觉、1 0 0 万 t o k e n 上下文，并于 7 月 2 7 日放出完整权重。它把中国开放权重模型的竞争重点从“够用且便宜”推向前沿编码与长程任务。
- M e t a 与 A n t h r o p i c 据报洽谈两年最高 1 0 0 亿美元的算力租赁。若落地，M e t a 算力将变成对外云收入，模型公司与基础设施供应商的边界进一步模糊。
- 欧盟正式要求 G o o g l e 向竞争性 A I 助手开放 A n d r o i d 系统能力；这不是普通反垄断罚款，而是直接重写移动端 A I 入口的分配方式。
- 2 0 2 6 世界人工智能大会把开放生态、普惠可及和发展中国家能力建设写入联合声明。短期看是政策信号，长期看会影响开放模型、算力服务与行业解决方案的全球南方市场路线。

今日结论

- 1 . 前沿模型竞争新增了“开放权重 + 可直接采购的 A P I ”这一强变量；企业选型不应再把开放模型默认视为二线能力，但仍要等完整权重和独立评测落地。
- 2 . 算力正在从单纯成本中心变成可交易资产。大厂自建集群、模型公司租用能力、云厂商撮合需求三种角色开始重叠，采购方更应关注合同期限、容量保障和模型迁移成本。
- 3 . 移动操作系统的 A I 权限正在变成监管对象。未来的助手竞争不只比模型效果，还比唤醒、屏幕理解、系统动作和默认入口能否获得同等访问权。

A I 产品与应用

阿里云百炼披露 5 0 0 万用户、1 2 0 万付费用户

新华社在 W A I C 前夕披露，阿里云百炼推理平台已服务全球 5 0 0 万用户，其中 1 2 0 万为付费用户。这组数据比展会上的功能演示更重要：开发者平台开始用付费账户规模证明需求，而非只用调用量或注册量描述采用情况。新华社 (<https://www.news.cn/tech/6/8e6a2e74f0cd4e888b9873d5b26901c3/c.html>)

商业含义是，模型平台竞争会进一步转向推理价格、工具链兼容、企业数据连接和稳定交付。对采购方而言，下一步应核对活跃付费率、留存与单位任务成本，而不是把累计用户数直

接等同于生产采用。

D a t a b r i c k s 把分析对话推进为可复用 A g e n t

D a t a b r i c k s 7 月 16 日更新 G e n i e O n e：加入记忆测试版、历史对话检索和按需任务；同时允许将聊天线程保存为 G e n i e A g e n t，并开放单次 A g e n t 对话分享预览。D a t a b r i c k s 发布说明 (<https://docs.databricks.com/aws/en/>)

新增价值不在“又一个聊天入口”，而在分析过程可被保存、共享并再次执行。企业数据团队应同步设置记忆内容的保留规则、计划任务审批和共享范围，否则便利性会先于控制能力扩张。

模型与技术进展

K i m i K 3 把开放权重模型推到 3 T 级别

月之暗面称 K i m i K 3 拥有 2.8 万亿总参数，采用稀疏专家架构，具备原生视觉与 100 万 token 上下文；A P I 与 K i m i C o d e 已上线，完整权重计划在 7 月 27 日发布。博客 (<https://www.kimi.com/fr-fr/blog/kimi-k3>) K i m i (www.kimi.com/code/docs/en/kimi-code/whats-new.html)

A P 报道其在前端编码评测中进入领先区间，但当前仍需区分厂商自报、早期竞技场结果与可复现实验。A P (<https://apnews.com/article/0d8a5e268deb>) 对企业的现实影响是：高性能开放模型正在增加议价空间；真正的采购检查点将是完整权重许可证、私有部署所需显存、长上下文稳定性与单位有效任务成本。

投融资与商业动态

M e t a 或向 A n t h r o p i c 出租最高 100 亿美元算力

R e u t e r s 援引知情人士称，M e t a 正与 A n t h r o p i c 洽谈一项两年最高 100 亿美元租赁交易；磋商尚未完成。M e t a 此前已被报道筹备对外销售超额算力、托管第三方模型的云业务。R e u t e r s 转引 (<https://www.streetinsider.com/Reuters/2Bin2Btalks2Bfor2Bpotential2B%2410%2BbillionC2BNYT%2Breports/26783337.html>)

若成交，这会同时验证两件事：A n t h r o p i c 的容量需求仍在快速上升，M e t a 也愿意把自建 A I 基础设施从内部战略资产转成外部收入。对市场的更大影响是，传统云厂商将面对拥有自研模型、广告现金流和大规模集群的新型竞争者。验证状态：交易洽谈获一级媒体确认，但金额、期限和最终签约仍未完全验证。

政策、伦理与安全

欧盟要求 Android 对竞争性 AI 助手开放系统能力

欧盟委员会依据《数字市场法》完成针对 Alphabet 的措施决定，要求 Google 给予竞争性 AI 服务与 Gemini 同等的 Android 功能访问。措施包括语音唤醒等能力；并发热测试最迟须在 Android 19、即 2028 年 8 月 1 日前实现。欧盟委员会决定与开发者讨论 <https://digital-markets-act.ec.europa.eu/development/alphabet-specification-proceedings-interoperability> : <https://apnews.com/article/184b3067120e56d858cb8c81aee>

这会把 ChatGPT、Claude 等第三方助手从“打开一个 App”推向系统级候选入口。商业上，手机厂商与 AI 服务商的合作价值上升；同时，权限授权、默认设置和跨应用数据访问将成为新的合规与用户信任焦点。

WAIC 联合声明把开放生态与普惠可及列为共同方向

7 月 17 日至 20 日举行的 2026 世界人工智能大会联合声明提出，负责任地共同发展开放源码生态、建设开放包容的国际社区，并在知识产权保护基础上提高 AI 技术与服务的可及性。中国外交部：大会主席声明 (https://www.fmprc.gov.cn/eng/xt/t20260717_11984715.html)

对企业而言，最值得跟踪的不是口号本身，而是后续是否转化为开放模型许可、公共算力、跨境合作项目和发展中国家采购计划。政策方向已经明确，执行机制与资金规模仍待后续文件验证。

X 平台高信号

1 .

信号标题：Kimi K3 同步给出 API 价格与完整权重时间表

类型：定价与访问变化

来源或链接：Kimi K3 官方博客 (<https://www.kimi.com/fr-fr/blog>)
页（需以官方资料校验） (<https://x.com/i/trending/20770400707>)

核心观点：Kimi K3 已通过 API 提供服务，公开价格为每百万 token 输入 3 美元、15 美元；完整权重计划于 7 月 27 日发布。当天新增变量是“可购买的服务”和“明确的权重发布日期”，不只是模型预告。

为什么重要：价格、可用性和权重交付决定开放模型能否真正进入企业评测与私有化部署。

影响：美国闭源前沿模型将承受新的价格锚点；企业可立即做 API 对照测试，但本地部署结论应等权重发布。

验证状态：价格与时间表已由官方博客核验；独立成本—效果比较仍待扩大样本。

2 .

信号标题：Android AI 助手的系统级权限被要求对竞争者开放

类型：平台规则变化

来源或链接：欧盟委员会开发者门户 (https://digital-markets-act.ec.europa.eu/developer-portal/interoperability/alphabet-specific-compatibility-ai-services_en)

核心观点：欧盟最终决定要求 Google 让竞争性 AI 服务获得与 Gemini 相同的 Android 功能访问，并把并发语音唤醒的最晚交付节点锁定为 Android 19、2028 年 8 月 1 日。

为什么重要：系统权限而非单纯模型排名，决定助手能否成为默认入口。

影响：OpenAI、Anthropic 和独立助手厂商在欧洲获得新的产品分发机会；Google 需调整 API、隐私和安全控制。

验证状态：欧盟委员会正式决定，已核验。

3 .

信号标题：百炼披露 120 万付费用户

类型：硬采用数据

来源或链接：新华社 (<https://www.news.cn/tech/20260716/8e6a5b26901c3/c.html>)

核心观点：阿里云百炼当前披露 500 万全球用户、120 万付费用户，付费用户约占披露用户数的 24%。

为什么重要：付费账户规模比注册用户或模型下载量更接近商业化真实度。

影响：国内模型平台的竞争会更集中到续费、推理成本、企业连接器与交付质量；24% 仅为静态比例，不能替代活跃率和收入指标。

验证状态：数据由新华社报道并注明平台口径；活跃度、收入与去重规则未披露。

4 .

信号标题：Meta 尝试把自建算力转成对外收入

类型：基础设施商业化

来源或链接：Reuters 转引 (<https://www.streetinsider.com/Anthropic%2Bin%2Btalks%2Bfor%2Bpotential%2B%2410%2Bbillion%2Bdeal%2C%2BNYT%2Breports/26783337.html>)

核心观点：Meta 与 Anthropic 洽谈两年最高 100 亿美元算力租赁；新增事实是潜在交易对象与金额上限，而非泛泛的“Meta 做云”。

为什么重要：大型模型实验室的算力需求开始为非传统云厂商创造直接收入。

影响：若签约，Meta 将同时扮演模型开发商、集群运营商和算力供应商；Anthropic 则进一步分散容量来源。

验证状态：Reuters 已确认磋商存在；尚未签约，金额上限不代表最终合同价值。

5 .

信号标题：Genie One 增加记忆并允许把聊天保存为 Agent

类型：产品访问变化

来源或链接：Databricks 2026 发布说明 (<https://docs.databricks.com/release-notes/2026>)

核心观点：7 月 16 日更新把记忆、历史对话检索、计划任务按需运行和“聊天转 Agent”放到同一分析产品内。

为什么重要：分析助手由一次性问答转向可持续保存、复用和执行的工作单元。

影响：数据团队能更快沉淀分析流程，但管理员需要把记忆保留、任务执行与会话分享纳入权限策略。

验证状态：官方发布说明已核验；记忆仍为 Beta，会话分享仍为公开预览。

6 .

信号标题：WAIC 声明把开放源码生态与技术可及性写入共同承诺

类型：国际政策信号

来源或链接：中国外交部：2026 WAIC 主席声明 (https://www.fmprc.gov.cn/xw/202607/t20260717_11984715.html)

核心观点：主席声明明确提出建设开放包容的国际开源社区，并在知识产权保护基础上分享研究成果与技术经验；新增变量是多边会议文本中的正式表述。

为什么重要：开放模型正在从开发者策略上升为国际合作与市场拓展工具。

影响：面向全球南方的低成本模型、培训与行业解决方案可能获得更多政策支持；具体项目和预算仍需逐项跟踪。

验证状态：官方会议声明已核验；后续执行安排尚未公布。

前沿研究速递

ClawBench：把网页 Agent 放到 144 个真实网站上测试

做了什么：ClawBench 设计了 153 个日常在线任务，覆盖购物、预约、求职等 15 类场景，并直接在 144 个生产网站上测试 Agent。([Hugging Face Daily Papering Face.co/papers?q=AI+Agents](https://huggingface.co/papers?q=AI+Agents))

新在哪里：它避开静态网页沙箱，把页面变化、跨站步骤、用户文档读取和长表单填写纳入同一评测。

潜在应用方向：采购前评估网页自动化 Agent，建立真实失败类型清单，并测算人工接管率。

一句话判断：真实网站评测比单一成功率更接近商业风险，但结果会受网站更新影响，复现必须记录时间与环境。

Multi-Head Latent Control：让模型从内部状态判断是否升级、答

做了什么：研究者在冻结的 L L M / V L M 上增加轻量控制头，根据隐藏状态轨迹判断当前模型能否完成任务，以及应直接回答、澄清、调用工具、升级到强模型或拒答。论文页面 (<https://HuggingFace.co/papers/2607.14277>)

新在哪里：控制信号不再只依赖输入提示或外部路由器；论文报告在 A n d r o i d W o r l d 上可减少最高 90.7% 的大模型调用，同时保留大部分性能。

潜在应用方向：多模型路由、客服升级、工具调用门控和高风险任务的自动停手。

一句话判断：这是把“何时不该继续生成”做成可训练组件的实用路线，但跨模型与线上漂移仍需验证。

A g e n t C o m p a s s : 统一 A g e n t 评测基础设施

做了什么：A g e n t C o m p a s s 提供开源、轻量、可扩展的 A g e n t 评测基础设施，试图减少不同任务套件之间的耦合与重复工程。论文页面 (<https://HuggingFace.co/papers/13705>)

新在哪里：重点不是提出新榜单，而是把能力评估的执行、扩展和复现做成统一基础设施。

潜在应用方向：企业内部 A g e n t 回归测试、供应商横评、版本升级前的任务级验收。

一句话判断：当 A g e n t 进入生产，稳定可复现的评测流水线会比单次榜单名次更有采购价值。