

Kimi K3发布2.8T参数模型 Open Secure AI联盟成立

中文内部早报 · 5分钟版

今日总览

月之暗面发布2.8T参数Kimi K3模型并开源相关技术栈，成为今日最大技术亮点。该模型具备原生视觉理解和百万token上下文窗口，多家平台提供发布日支持。

Agent与编程领域迎来多项进展，Kimi K3开源分布式智能体环境AgentENV，GitHub Copilot推出Harness workflow整合开发全流程。研究显示AI代理成本效益仍待验证。

前沿模型方面，Google AI Overviews搜索出现率达43%，OpenAI研究显示ChatGPT用户跨岗位工作现象显著。OpenRouter新增图像生成API端点，Kimi K3上线Modal平台。

产业合作层面，Cognizant成为Claude Partner Network全球首要合作伙伴，培训超3万名员工。ESP32-S3微控制器成功运行28.9M参数模型显示边缘部署潜力。

安全监管领域，NVIDIA等数十家机构联合成立Open Secure AI Alliance推动AI安全开源化。Anthropic澄清未主张禁止开源权重模型，但支持芯片出口管制和安全测试。

头条

今日头条

1. NVIDIA等联合成立Open Secure AI联盟推动AI安全开源化

NVIDIA Blog · 07/28 08:00 · 监管、安全与版权

NVIDIA、Microsoft、Hugging Face、IBM等数十家机构联合成立Open Secure AI Alliance，旨在通过开源模型、工具和框架构建可审查、可定制的AI安全防御体系。

2. Cognizant扩大与Anthropic合作成为Claude全球首要合作伙伴

Anthropic: Newsroom · 07/28 08:00 · 大厂与产业链

Cognizant与Anthropic扩大合作，成为Claude Partner Network中的Global Premier Partner，并将Claude嵌入其Flowsource等平台。已有超过3万名员工完成Claude培训，其为一家生物制药公司构建的智能合约系统将合同审核时间缩短最高40%，提取准确率超过88%。

3. 用Claude和Python构建技能驱动的分析智能体

MarkTechPost · 07/28 08:00 · 智能体、编程与 workflows

本教程基于Anthropic的financial-services仓库，用纯Python复现其技能驱动架构。通过解析SKILL.md文件构建可搜索技能注册表，并创建可复用SkillAgent，将金融分析剧本注入Anthropic Messages API，支持迭代工具调用循环。

4. OpenAI研究显示AI正在扩展工作者职能范围

OpenAI 新闻 · 07/27 11:30 · 前沿模型与多模态

据OpenAI 新闻，OpenAI 更新模型能力与产品方案。这类进展主要影响模型能力、API 选型和产品可用性。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

5. Kimi K3开源2.8T MoE模型与技术报告

X: Kimi.ai (@Kimi_Moonshot) · 07/28 08:00 · 智能体、编程与 workflows

Kimi发布其最强模型Kimi K3，这是一个2.8T参数的MoE模型，具备原生视觉理解和1M token上下文窗口。新架构实现了每单位计算2.5倍的智能提升。除模型权重外，Kimi还开源了高性能注意力内核、MoE通信库及大规模智能体运行环境基础设施。

模型

前沿模型与多模态

1. Google AI Overviews 搜索结果出现率升至43%

TechCrunch AI · 07/28 08:00

Google AI Overviews 在搜索结果中的出现率一年内从15%升至43%，AI Mode月访问量从1.26亿增至2.79亿。用户搜索长度增加，正从短关键词转向更长的自然对话式查询。

2. OpenRouter 新增图像生成模型专用 API 端点

OpenRouter 公告 · 07/28 08:00

OpenRouter 通过专用`/api/v1/images`端点提供图像生成模型服务，图像理解仍通过`/chat/completions`端点完成，两者共用同一 API Key 和计费体系。该平台同时公布了两种任务的完整接口合约，并修复了此前出现的“no endpoints found”错误。

3. Kimi K3 上线 Modal，支持无损加速推理

X: Kimi.ai (@Kimi_Moonshot) · 07/28 08:00

很高兴 @modal 成为 Kimi K3 的 Day 0 发布合作伙伴！ 他们为 K3 的架构训练了自定义 DFlash 投机器，实现更快推理且无质量损失。

4. 用AI Skill自动生成可协作HTML PPT

X: [Vista \(@vista8\)](#) · 07/28 08:00

Vista 基于 bento PPT 改造了一个 Skill，输入内容或主题即可自动生成可编辑、在线演示并支持协作的 HTML PPT。安装指令为 `npx skills add joeseesun/qiaomu-bento-ppt`，推荐使用 Kimi K3 或 Opus 4.8+ 等前端审美好的模型。

智能体 智能体、编程与工作流

1. 浪费20亿Token后，我开源了帮Agent定义目标的Leader.skill

公众号: [数字生命卡兹克](#) · 07/28 08:00

作者开源了Leader.skill，用于将模糊的人类需求转化为Agent可独立执行数小时的目标任务书。该Skill基于“目标七问”方法论，涵盖目的、完成态、反作弊、边界等维度，并推荐用Claude Fable 5或Kimi K3规划目标，再交由GPT-5.6 Sol或GLM-5.2等模型长程执行。项目已开源。

2. METR 更新 AI 工作流与智能体方案

[The Decoder](#) · 07/27 20:28

据The Decoder，METR 更新 AI 工作流与智能体方案。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. xAI 发布 Grok CLI 并支持 /tutorial 命令

X: [Elon Musk \(@elonmusk, xAI\)](#) · 07/26 08:25

据X: Elon Musk (@elonmusk, xAI)，xAI 发布 Grok CLI 并支持 /tutorial 命令。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

4. Kimi K3 开源分布式智能体环境 AgentENV

X: [Kimi.ai \(@Kimi_Moonshot\)](#) · 07/27 23:25

我们与 kvcache-ai 合作开源了 AgentENV。AgentENV 是一个用于大规模运行智能体环境的分布式系统。其组件为 Kimi K3 的智能体强化学习训练提供支持，具备快速快照、恢复和分支功能，适用于大规模并行智能体工作流。在 GitHub 上探索：<http://github.com/kvcache-ai/AgentEnv>

产业 大厂与产业链

1. OpenAI 称越来越多员工用 ChatGPT 做其他岗位的工作

[The Decoder: AI News](#) · 07/28 08:00

OpenAI 分析超 80 万条与工作相关的 ChatGPT 消息后发现，43.5% 的岗位特定查询涉及另一职业，营销和工程任务交叉最多。用户用 AI 处理合同审查、数据分析、网站故障排查等原由专家负责的工作。OpenAI 认为这是岗位职责正在变化的早期信号，该趋势在缺乏专业团队的小公司尤为明显。

2. 当编码不再是瓶颈：Berkeley RDI 提出软件自主开发三级框架

[Berkeley RDI: Blog \(AI 安全与评测\)](#) · 07/26 00:00

Berkeley RDI等机构提出三级软件自主开发框架：代码自主（AI完成设计与实现，人类决定构建内容并审核PR）、流水线自主（AI运行从设计到部署的全流程，人类仅评估结果）、需求自主（AI自主决定构建内容）。该框架旨在为能力声明、部署选择与问责提供清晰分类。

3. 在 8 美元的 ESP32-S3 微控制器上运行 28.9M 参数大语言模型

[Hacker News](#) · 07/26 11:54

开发者成功在售价约 8 美元的 ESP32-S3 微控制器上运行了一个 28.9M 参数的大语言模型，完全在芯片本地运行，无需连接服务器，生成速度约 9.5 tok/s。

4. 数百用户向ChatGPT索要毒药与生物武器配方，部分获得高中生水平步骤指南

[The Decoder: AI News](#) · 07/26 16:35

2025年夏季，OpenAI内部将GPT-5标记为高风险，因其可帮助教育程度有限的用户制造生物危害。据《华尔街日报》报道，自去年夏天以来，数百名用户向ChatGPT询问如何制造生物武器和毒药，部分用户获得了员工称高中生都能遵循的逐步指南。OpenAI暂停了相关账户，但未向当局报告任何事件。

开源 开源项目与开发者工具

1. SGLang 和 Miles 为月之暗面 2.8T 参数 Kimi K3 模型提供发布当日支持

[LMSYS: Blog \(Chatbot Arena 团队\)](#) · 07/28 08:00

SGLang 和 Miles 为月之暗面开源的 2.8T 参数模型 Kimi K3 提供发布当日支持，分别负责推理和 RL 训练。K3 采用 69 层 KDA 线性注意力与 24 层 MLA 交错的混合架构，在 SGLang 上单卡 batch-1 解码速度达约 113 tok/s，结合 DSpark 推测解码可达约 423 tok/s。

2. GitHub Copilot 发布 “Harness” 工作流：用单一工具完成原型、规划、实现与代码审查

[GitHub Blog](#) · 07/28 08:00

GitHub Copilot 推出 “Harness” 工作流，让开发者通过单一 AI 工具完成从原型设计、规划、实现到代码审查的完整软件开发流程，无需追逐多种新 AI 工具。该工作流强调实用性与集成性，旨在减少工具切换带来的效率损耗。

3. GitHub 发布新动态：《GitHub Copilot app 入门指南：多 Agent 会话工作区与 Canvas 预览》

GitHub Blog · 07/28 08:00

据GitHub Blog，本条原始主题为《GitHub Copilot app 入门指南：多 Agent 会话工作区与 Canvas 预览》。原文摘要为：GitHub Copilot app 将 AI 编码工具升级为多 Agent 会话工作区，支持同时管理多个任务线程而不丢失进度。用户可为每个会话绑定项目上下文，通过 ``/create-canvas`` 命令在浏览器 Canvas 中预览 UI 并直接点选修改，还能启用 Agent Merge 自动处理 PR 审查反馈和合并冲突。。该材料涉及开发者工具、模型路由或开源生态，需继续核验收码、许可、维护频率和部署条件。

4. 新开源项目 0xwilliamortiz/openclaude-improved 上线

GitHub · 07/27 05:58

GitHub 新仓库 0xwilliamortiz/openclaude-improved 在本次覆盖窗口内创建。仓库简介显示，该项目面向 AI 开发者工具或自动化场景。当前公开数据约为 429 Stars、70 Forks，主要语言为 TypeScript。该条目代表新项目出现，不把普通代码提交描述为版本发布。

观点 KOL 与巨头言论

1. Hacker 发布新动态：《Show HN: BixRouter – a non-linear AI chat interface for OpenRouter》

Hacker News · 07/28 00:21

据Hacker News，Show 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

2. Hacker 发布新动态：《Show HN: GG Translator – Turn gaming shit talk into friendly phrases》

Hacker News · 07/27 06:23

据Hacker News，Show 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. Ask 发布 AI 行业观点与判断

Hacker News · 07/25 17:20

据Hacker News，Ask 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

监管 监管、安全与版权

1. Anthropic 澄清立场：从未主张全面禁止开源权重模型，支持芯片出口管制与安全测试

Anthropic: Newsroom · 07/28 08:00

Anthropic CEO Dario Amodei 明确表示公司从未主张禁止开源权重模型，并认为不具备危险能力的开源权重模型是公共产品。他提出三项实际措施：对华芯片出口管制、打击工业级知识蒸馏、对所有足够强大的模型进行强制性安全测试。Amodei 指出，保护主义禁令无法解决其最担忧的国家安全威胁，包括威权政府利用更强大 AI 实现军事优势或深度监控。

2. Microsoft 更新 AI 治理与安全动态

TechCrunch AI · 07/28 02:32

据TechCrunch AI，Microsoft 更新 AI 治理与安全动态。这类进展主要影响数据、版权、安全和合规部署边界。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. OpenAI、Anthropic 游说美国限制中国开源模型，黄仁勋与马斯克公开反对

IT之家 · 07/26 20:54

OpenAI 与 Anthropic 正游说美国监管机构限制中国开源 AI 模型，认为开放开发过于危险。英伟达 CEO 黄仁勋、微软 CEO 纳德拉、马斯克及扎克伯格等人公开支持开源，签署联名信反对限制。近 200 家硅谷创业公司也敦促特朗普政府不要限制获取中国开源模型，美国官员倾向于将此事作为国家安全问题单独处理。

研究 研究、评测与方法论

1. OpenAI 发布 AI 研究评测新进展

The Decoder · 07/28 02:50

据The Decoder，OpenAI 发布 AI 研究评测新进展。这类进展主要影响模型训练、评测方法和应用路线选择。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。