

Cursor被SpaceX收购，Google发布Gemini 3.7 Flash编程专用模型

中文内部早报 · 5分钟版

今日总览

今日AI领域重大并购完成，Cursor正式被SpaceX收购，双方合作成果Grok 4.6已发布。Google DeepMind推出专门面向编程与智能体的Gemini 3.7 Flash模型，价格仅为前代一半。

Agent与Coding板块迎来多项更新，Google发布Sheets canvas功能，让用户通过自然语言将表格数据转化为交互式应用。Cursor推出builds功能，使云智能体启动速度提升3倍。

模型与多模态方面，通义千问开源Qwen3.8系列模型，其中Qwen3.8-27B为原生多模态稠密模型。OpenAI发布GPT-5.6构建者指南，新增推理持久化等API能力。

大厂与产业链动态中，Hugging Face发布2026年夏季开源模型生态观察报告，显示中国实验室在前沿模型规模上领先。微软首发自研推理模型MAI-Thinking-1。

监管安全与研究工具板块，Anthropic宣布Claude文本将包含水印以符合欧盟AI法案要求。一项研究质疑Anthropic和OpenAI关于自主AI研究可达性的声明，实验显示AI智能体未能成功独立撰写AI研究论文。

头条

今日头条

1. Cursor正式被SpaceX收购

Cursor Blog · 08/15 08:00 · 大厂与产业链

Cursor已被SpaceX正式收购，完成自4月启动的收购流程。合并后Cursor将获得全球最大GPU集群，以构建更强且运行成本更低的模型，从而以更低价格向客户提供更强大的模型。本周三发布的Grok 4.6是双方合作成果的早期体现。

2. Claude Code v2.1.233发布新增GitLab支持

Claude Code: GitHub Releases · 08/15 08:00 · 开源项目与开发者工具

Claude Code v2.1.233发布，为--worktree标志和agents视图新增GitLab合并请求URL支持，并添加可选的forward_user_identity网关设置以按用户归因支出。

3. Google DeepMind推出Gemini 3.7 Flash编程专用模型

Google DeepMind: Blog · 08/14 01:04 · 智能体、编程与工作流

Google DeepMind发布Gemini 3.7 Flash，距3.6 Flash仅三周，主打编程与智能体任务，输入/输出价格分别为每百万token \$0.75和\$3.75，为原3.6 Flash的一半。

4. Google Sheets推出Sheets canvas用Gemini转化表格为应用

Google Blog: AI · 08/14 00:45 · 前沿模型与多模态

Google Sheets发布新功能Sheets canvas，基于Gemini构建，用户只需用自然语言提示词即可将表格数据转化为交互式仪表盘、学习追踪器、座位表等"迷你应用"。

5. Claude文本将包含水印以符合欧盟AI法案

Anthropic: Newsroom · 08/15 08:00 · 前沿模型与多模态

未来Claude模型生成的文本将包含水印，用于判断文本由Claude撰写的可能性，这是Anthropic为遵守欧盟《AI法案》而实施的变更。该方法基于Google DeepMind的SynthID-Text技术，对输出质量、创造力和可读性无实际影响，读者无法区分水印文本与普通文本，且不增加额外token或成本。

模型

前沿模型与多模态

1. OpenRouter 视觉指南：如何通过 API 向多模态模型发送图像

OpenRouter 公告 · 08/15 08:00

OpenRouter 发布视觉指南，详解通过 Chat Completions API 向多模态模型发送图像的方法。请求体采用 messages 数组，用户消息的 content 包含 text 和 image_url 两部分，支持公开 URL 或 base64 数据 URL 两种格式，兼容 PNG、JPEG、WebP 和 GIF。

2. OpenAI 更新模型能力与产品方案

OpenAI 新闻 · 08/13 18:00

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。

3. OpenAI 更新模型能力与产品方案

Ars Technica: AI · 08/15 08:00

据Ars Technica: AI，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。

4. 通义千问开源 Qwen3.8 系列模型

X: 通义千问 / Qwen (@Alibaba_Qwen) · 08/14 23:02

通义千问兑现承诺，开源 Qwen3.8 系列模型。其中 Qwen3.8-27B 为原生多模态稠密模型，仅 27B 参数即全面超越 Qwen3.7-Plus，原生支持 262K 上下文，可通过 YaRN 扩展至 1M tokens，采用 Apache 2.0 许可。Max 级 Qwen3.8-2.4T-A95B 的开放权重也已同步发布。

智能体 智能体、编程与 workflows

1. Claude Code 会话如何最大化 token 价值

Claude 博客 · 08/15 08:00

Claude Code 的 token 成本由模型、输入/输出 token 和提示缓存三因素决定，输出 token 价格约为输入的 5 倍。任务间运行 /clear 可减少无关上下文回传，降低 token 用量；会话中途切换模型或 effort 级别会破坏提示缓存，增加成本。

2. Cursor 推出 builds：云智能体启动速度提升至 3 倍

Cursor Blog · 08/13 20:00

Cursor 推出 builds 功能，在后台持续准备就绪的开发环境副本，让云智能体启动时无需从零搭建，响应速度最高提升 3 倍。内部环境启动快 10 倍，首个 token 生成快 3 倍；智能体始终从最近一次成功的 build 启动，依赖更新或安装脚本出错时不会影响运行。8 月 17 日起所有环境默认启用 builds，无需额外费用。

3. GPT-5.6 构建者指南：如何以更低成本实现前沿智能体性能

OpenAI · 08/13 19:00

GPT-5.6 模型家族以更低成本实现前沿级智能体性能，并新增推理持久化、原生多智能体编排和程序化工具调用等 API 能力。在 ARC-AGI-3 上，启用保留推理和压缩后，Sol 得分从 13.3% 跃升至 38.3%，且输出 token 减少约 6 倍。Luna 在 BrowseComp 上以 84.04% 的得分追平 GPT-5.5 (84.36%)，成本从 \$33.27 降至 \$1.33。

4. dots3-note Preview 开源：280B 参数轻量模型，主打长程智能体与多模态推理

公众号: 小红书技术 (dots.llm) · 08/15 08:00

小红书技术开源 dots3-note Preview，这是 dots3 系列最轻量模型，总参数 280B、激活参数 16B，支持 512K 上下文及文本、视觉、语音多模态理解，并针对复杂推理和长程 Agent 任务优化。

产业 大厂与产业链

1. 2026年夏季开源模型生态观察：中国前沿模型规模领先，AMD与NVIDIA主导发布量

Hugging Face 博客 · 08/15 08:00

2026年1至8月，Hugging Face公开模型仓库从243万增至296万，但85.6%的模型下载量不足200次，1.5%的仓库占据99.2%下载量。中国实验室月度最大开源模型参数规模在754B至2.78万亿之间，美国实验室七个月中五个月低于130B。AMD与NVIDIA各发布超200个新模型仓库，成为发布开源模型最多的机构。

2. 印尼首个大学AI中心落成：UGM、Indosat与NVIDIA合作培养本地AI人才

NVIDIA Blog · 08/15 08:00

印尼通信与数字事务部、Indosat、NVIDIA与加查马达大学（UGM）在日惹共同启动UGM Indosat NVIDIA AI技术中心（NVAITC），这是该国首个大学AI技术中心。

3. 微软首发自研推理模型MAI-Thinking-1

X: Mustafa Suleyman (Microsoft AI CEO) (@mustafasuleyman) · 08/13 00:00

我们的首个推理模型 MAI-Thinking-1 从零开始构建，现已在 Microsoft Foundry 上线。为团队点赞！更多详情如下。

4. OpenAI 推出 AI 产业部署新进展

OpenAI 新闻 · 08/12 14:00

据OpenAI 新闻，OpenAI 推出 AI 产业部署新进展。这类进展主要影响云算力、硬件投入、企业采购和服务成本。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。

开源 开源项目与开发者工具

1. AutoGPT 如何用 AGENTS.md 和技能门控管理 AI 生成的拉取请求

GitHub Blog · 08/13 02:00

AutoGPT 维护者发现，AI 智能体不会主动阅读文档，因此将指令放在 AGENTS.md 和技能文件中，并置于代码目录旁。他们通过强制 PR 模板、测试计划、CI 覆盖率门槛和 CLA 签名等门控机制，将智能体提交的 PR 从 "不可用" 转变为 "可用但不符合路线图"。其中 CLA 签名因需浏览器和 OAuth 流程，被用作区分人类与智能体的 "人类探测器"。

2. 新开源项目 Electricitysheep/dsh-handbook 上线

GitHub · 08/13 23:22

DeepSeek Harness (dsh) 从 0 到 1 深度手册：安装/插件开发/性能调优/实测案例/同模型多 Agent 实测对比（中文 + 英文 PDF） Stars: 203, Forks: 5, Language: Python。

3. Hugging Face 发布新动态：《What We Learned by Reproducing 2,200 papers from ICML》

Hugging Face Blog · 08/13 08:00

据Hugging Face Blog，本条原始主题为《What We Learned by Reproducing 2,200 papers from ICML》。该材料涉及开发者工具、模型路由或开源生态，需继续核验代码、许可、维护频率和部署条件。

4. AllenAI推出OlmoEarth嵌入模型：支持自定义导出用于下游分析

Hugging Face Blog · 08/13 00:14

AllenAI在Hugging Face博客发布OlmoEarth嵌入模型，该模型支持从OlmoEarth Studio进行自定义嵌入导出，用于下游分析任务。这是专门针对地球科学数据的嵌入模型工具。

观点

KOL 与巨头言论

1. Hacker 发布新动态：《Show HN: LymeScribe – one computer on your network transcribes for the rest》

Hacker News · 08/14 02:18

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: Tried some experiments with architecture for Long term memory for LLM》

Hacker News · 08/13 04:03

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. Hacker 发布新动态：《Show HN: Pane – An open-source browser that turns your work into living websites》

Hacker News · 08/13 00:25

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。

监管

监管、安全与版权

1. OpenAI 更新 AI 治理与安全动态

The Decoder · 08/15 00:06

据The Decoder，OpenAI 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。

2. 新兴多智能体系统的模式与问题

Anthropic: Research (发表成果 · 网页) · 08/13 09:20

Anthropic 研究指出，随着 AI 智能体在共享代码库、市场等社会系统中承担更多任务，智能体间交互量或将超过人机交互。实验显示，45 个协调智能体在 2700 万 token 运行中发现 266 个漏洞，而独立并行方法在 650 万 token 中发现 21 个，两种方法仅 12 个重叠，且协调智能体学会专业化分工。研究同时警示个体层面的良性行为怪癖可能叠加为意外的系统性失败。

3. 我写了一本 AI 教科书--AI 还要多久才能写得更好？

Nathan Lambert: Interconnects · 08/12 21:01

作者在完成一本 RLHF 教科书后反思，LLM 在长文非虚构写作上进展停滞，GPT 4.5 和 Kimi K2 等写作强模型已显老旧，而编码、数学等任务已接近超人水平。模型能改错字、做编辑，但组织整章内容时仍混乱且易出错，作者认为这阻碍了模型自主解决开放科学问题。

研究

研究、评测与方法论

1. Cursor 与 SpaceXAI 联合发布 Grok 4.6

Cursor Blog · 08/12 08:00

Cursor 与 SpaceXAI 今日发布 Grok 4.6，重点强化长时运行智能体与交互式视觉任务，在多项智能体编程与知识工作基准上达到前沿水平，并在 Artificial Analysis Intelligence Index 上追平 GPT-5.6 Sol。