

AI模型与智能体技术密集发布，大厂加速商业化落地

中文内部早报 · 5分钟版

今日总览

今日AI领域呈现技术迭代与商业应用并进态势。xAI发布Grok 4.6强化智能体能力，各大厂商围绕推理模型、数据库迁移、企业应用等场景推出新产品和服务。

Agent与Coding板块迎来重要进展，Scale AI开源Muse系列智能体模型，OpenRouter发布跨模型工具调用指南，ChatGPT桌面端支持导入其他智能体工作数据，OpenChamber推出基于代理的开发环境。

模型与多模态领域，Claude in Chrome升级为Cowork会话，Runway Seedance 2.5支持50角色参考，NVIDIA推出Nemotron 3.5 Lightning加速本地智能体任务，Qwen发布2.4T参数最大开源模型。

大厂与产业链方面，Google Cloud将Gemini集成到数据库迁移服务，微软首发自研推理模型MAI-Thinking-1，OpenAI研究企业智能体落地方式，NVIDIA研发万亿参数Nemotron 4模型。

监管安全与研究工具板块，WhatsApp推出Scam Alert诈骗提醒功能采用端到端加密，研究人员发现主流AI模型API存在推理过程泄露漏洞，AutoGPT通过技能门控管理AI生成PR。

头条

今日头条

1. Google 推出 AI 产业部署新进展

Google Cloud: Databases · 08/12 00:00 · 大厂与产业链

据Google Cloud: Databases，Google 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。

2. xAI发布Grok 4.6，强化长时运行智能体能力

xAI 新闻 · 08/13 08:00 · 研究、评测与方法论

xAI发布Grok 4.6，在Grok 4.5基础上重点强化长时运行智能体及更复杂的交互式与视觉工作能力。该模型在多项智能体编码与知识工作基准上达到前沿水平，在Artificial Analysis Intelligence Index上追平GPT-5.6 Sol。

3. WhatsApp用端到端加密构建Scam Alert诈骗提醒功能

Meta Engineering Blog · 08/13 08:00 · 监管、安全与版权

WhatsApp推出可选功能Scam Alert，通过端到端加密保护下在设备端运行机器学习模型，识别潜在诈骗消息，且消息内容不会离开设备或自动上报。该功能遵循仅设备端处理、无自动上报、用户控制三大原则，模型权重公开供独立验证，遥测数据经差分隐私聚合处理。目前已在Beta版有限推出，并邀请安全研究社区参与测试。

4. 微软首发自研推理模型MAI-Thinking-1

X: Mustafa Suleyman (Microsoft AI CEO) (@mustafasuleyman) · 08/13 08:00 · 大厂与产业链

微软首个推理模型MAI-Thinking-1从零开始构建，现已在Microsoft Foundry上线。该模型标志着微软在推理能力方面的自主研发成果。

5. OpenAI研究：企业如何用ChatGPT和Codex落地智能体AI

OpenAI · 08/13 08:00 · 大厂与产业链

OpenAI研究揭示企业采用智能体AI的方式，以及前沿企业如何在AI应用上拉开差距。企业正通过ChatGPT和Codex将AI从辅助转向执行，头部公司已率先将智能体投入实际业务流程。

模型

前沿模型与多模态

1. Claude 更新模型能力与产品方案

Claude 博客 · 08/13 08:00

Claude in Chrome 浏览器扩展的侧边栏现已升级为 Claude Cowork 会话，对话会保存至历史记录，技能和连接器可在浏览器中工作，且任务可在桌面、网页和移动端应用间无缝切换。

2. 我写了一本 AI 教科书——AI 还要多久才能写得更好？

Nathan Lambert: Interconnects · 08/13 08:00

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。

3. Runway Seedance 2.5 上线，支持50角色参考

X: Runway (@runwayml) · 08/12 02:53

完整阵容，完整曲目，一次生成一个。Seedance 2.5 已在 Runway 上线，支持 50 个独特角色参考，以及最长 30 秒、与音乐同步的片段。点击下方链接即可开始使用。

4. OpenAI 更新模型能力与产品方案

OpenAI 新闻 · 08/10 20:00

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。

智能体 智能体、编程与工作流

1. Scale AI 开源 Muse 系列模型

X: Alexandr Wang (Scale AI 创始人/Meta 首席 AI 官) (@alexandr_wang) · 08/10 18:06

1/ 今天有个重大消息：我们很快将发布 Muse Spark 1.2 的开源权重版本。同时，我们还将发布 Muse Glimmer--一个 30B 参数的智能体模型，采用 Apache 2.0 协议开源权重。Muse Glimmer 可在 24GB 显存上运行，且不损失智能体可靠性。☑

2. OpenRouter 工具调用指南：一次编写循环，切换模型字符串即可跨模型运行

OpenRouter 公告 · 08/13 08:00

OpenRouter 发布工具调用指南，展示如何用同一套代码在 Claude、GPT 和开源权重模型间切换，仅需更改模型字符串。指南涵盖定义工具、发送请求、读取 tool_calls 响应、执行函数并返回结果的完整循环，支持 OpenAI 兼容的 JSON schema，并提供 cURL、Python 和 JavaScript/TypeScript 示例。

3. ChatGPT 桌面端支持导入其他智能体工作数据

X: OpenAI Developers (@OpenAIDevs) · 08/12 02:20

你现在可以将其他智能体的工作内容与 ChatGPT Work 和 Codex 保持同步。可导入项目、聊天记录、技能和插件，查看导入历史，并可在设置中选择开启自动更新。现已在 ChatGPT 桌面应用中提供。

4. OpenChamber：一个基于代理的开发环境

Hacker News · 08/10 08:46

OpenChamber 是一个基于代理的开发环境，可跨桌面、浏览器、手机和 VS Code 使用，支持会话目标、多模型并行运行与融合、变更走查、从 issue 到 PR 的完整流程及定时任务。该工具基于 OpenCode SDK，完全开源且免费，代码和会话内容均保存在本地，远程访问可通过 UI 密码和端到端加密的 Private Relay 保护。

产业 大厂与产业链

1. NVIDIA 推出 Nemotron 3.5 Lightning，加速本地智能体任务

NVIDIA Blog · 08/11 21:00

NVIDIA 发布 Nemotron 3.5 Lightning，一款可定制的开源 30B 混合专家 (MoE) 模型，专为常驻智能体设计。相比同类开源模型，其 token 生成速度最高提升 4 倍，任务完成时间缩短 30%。该模型采用开放权重，支持用户微调以匹配特定任务，并可在 RTX PC、DGX Spark 及 Jetson 等设备上运行。

2. SGLang 宣布 Day-0 支持 NVIDIA Nemotron 3.5 Lightning

LMSYS: Blog (Chatbot Arena 团队) · 08/11 21:51

SGLang 宣布对 NVIDIA Nemotron 3.5 Lightning 提供 Day-0 支持，该开源模型为 30B 总参数、3B 激活参数的混合专家架构，支持最长 1M token 上下文，可从 Hugging Face 下载 BF16 和 NVFP4 权重。模型支持 MTP、DFlash、DSpark 三种投机解码技术，并可通过 OpenAI 兼容 API 接入智能体工作流。

3. OpenAI 推出 AI 产业部署新进展

OpenAI · 08/13 08:00

RingCentral 通过全员发放 ChatGPT Work 和 Codex，推动从工程到运营的 AI 原生开发，其 AI-Native Challenge 让数千名员工（含非技术人员）交付了可运行项目。

4. 消息称英伟达开发万亿参数开源 AI 模型 Nemotron 4，目标挑战全球顶级

IT之家 · 08/11 22:54

英伟达正在研发新一代开源 AI 模型系列 Nemotron 4，规模最大的模型预计至少拥有 1 万亿个参数，旨在与全球最先进的开源模型竞争。英伟达尚未确定发布日期，最终训练也未完成，员工认为该模型最早可能在今年秋末准备就绪。此举意在通过开放模型生态扩大 AI 应用范围，并推动市场对其 GPU 算力的需求。

开源 开源项目与开发者工具

1. Claude Code v2.1.229 发布：新增远程会话恢复、插件市场命令源及多项修复

Claude Code: GitHub Releases · 08/13 08:00

Claude Code v2.1.229 发布，新增远程控制会话恢复、自托管 runner 的服务器端 hook 支持，以及插件市场命令源。修复了长响应流式输出丢失、窄终端渲染崩溃、Windows 扩展路径崩溃等问题。改进工作流扇出以复用缓存提示前缀，并调整 /commit-push-pr 对危险 git/gh 命令不再自动批准。

2. SGLang 与 Miles 为 Qwen3.8-2.4T-A95B 提供 Day-0 支持

LMSYS: Blog (Chatbot Arena 团队) · 08/13 08:00

SGLang 与 Miles 在发布首日即支持 Qwen3.8-2.4T-A95B，这是 Qwen 最大的开源模型，总参数 2.4T，每 token 激活 95B，采用混合注意力架构。

3. 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 08:00

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”。

4. OlmoEarth推出定制嵌入导出功能用于下游分析

Hugging Face Blog · 08/13 00:14

OlmoEarth Studio现在提供定制嵌入导出功能，用户可以将嵌入数据导出用于下游分析任务。该功能来自Allen Institute for AI的研究成果，旨在支持更灵活的数据处理需求。

观点

KOL 与巨头言论

1. Show 发布 AI 行业观点与判断

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。

2. Here 发布 AI 行业观点与判断

Hacker News · 08/12 22:49

据Hacker News，Here 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 发布 Daybreak 安全工具

OpenAI · 08/10 18:00

OpenAI 推出 Daybreak 系列安全工具，包括 Codex Security 和 GPT-5.5-Cyber，面向组织的大规模漏洞发现、验证和修补流程。

2. Claude Code 自动模式默认开启原理

X: Claude Devs (@ClaudeDevs) · 08/10 23:58

我们最近将自动模式设为 Claude Code 的默认选项，这意味着你不再需要批准每一个操作。但什么决定某个操作是否可以安全运行？看看它是如何工作的：

3. 研究人员发现可读取ChatGPT等模型加密推理过程的API漏洞

The Decoder: AI News · 08/12 01:38

Alexander Panfilov团队发现OpenAI、Anthropic、Google等主要AI提供商API存在漏洞，可读取推理模型的加密思考过程。扫描约7000条公开会话发现62个API密钥、33个邮箱和33个密码。通过越狱，Anthropic的Haiku 4.5可逐字转写Opus 4.8的原始推理；解码10000条推理轨迹的API成本约720美元。

研究

研究、评测与方法论

1. Meta 发布开源模型 Muse Glimmer

X: AI at Meta (@AlatMeta) · 08/10 18:13

推出 Muse Glimmer，一款开放权重、300 亿参数的模型，专为本地、常驻运行的智能体工作流优化。与同尺寸领先模型相比，Muse Glimmer 在关键智能体用例和基准测试中表现出色，并设计为完全在消费级硬件（如 Mac 或配备高性能 GPU 的 PC）上运行。秉承我们长期分享基础 AI 研究的传统，我们以宽松的 Apache 2.0 许可证发布模型权重。

2. 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。