

AI 行业日报

模型与工具能力 · 产业基础设施 · 应用商业化 · 研究开源 · 资本监管 | Data Source: aihot.virxact.com

API耗时: 1s 精选条目: 63 条 焦点: 8 条 快讯: 0 条

Executive Summary

微软发布首款高级推理AI模型MAI-Thinking-1，定位"中等规模"并在关键软件工程基准测试中达到领先水平，标志着其在自研AI模型领域的重要突破。Holo3.1计算机使用智能体系列更新，提供多种参数尺寸并首次发布量化检查点，优化本地推理性能。Miso One开源语音模型正式发布，8B参数支持一次语音克隆且推理延迟仅110ms。Grok Imagine 1.5预览版上线API体验，DeepSeek计划首轮融资筹集约500亿元人民币，投后估值预计达3500-4000亿元。

DigitalOcean AI云服务加入OpenRouter平台，在DeepSeek V3.2输出速度和延迟方面表现优异。NVIDIA与Microsoft深化合作推进智能体AI时代，Cursor Enterprise推出组织管理功能支持企业级部署。Anthropic扩展Claude Partner Network，建立三级服务体系。ChatGPT月活用户突破10亿，成为史上最快速达成此里程碑的应用，Claude月活达5600万。基础设施层面，OpenShell新增Vertex AI推理支持，OpenClaw发布Windows节点版本。

需持续关注MAI-Thinking-1模型的具体性能指标和开放节奏，DeepSeek融资进展及其对大模型市场竞争格局的影响。DigitalOcean在AI推理市场的定价策略和算力供给能力，以及各大厂商在智能体工具链方面的API开放程度和成本结构变化。开源语音模型Miso One的生态发展情况，以及企业级AI应用在组织管理功能完善后的商业化落地速度。

重点 今日核心进展

★ 1. 微软首款高级推理AI模型MAI-Thinking-1发布

The Verge: [AI \(RSS\)](#) · 昨天 02:12 · 模型与工具能力

微软在Build 2026上发布了其首款高级推理AI模型MAI-Thinking-1。该模型被定位为"中等规模"，能在"关键"软件工程基准测试中达到领先模型的水平。微软称其完全从头使用干净数据进行训练，未涉及从第三方模型进行知识蒸馏。这标志着微软在自研AI模型上迈出重要一步，此前其主要依赖OpenAI。近期两家公司已重新协商合作协议，关系有所松动。

[能力进展](#) [基础设施](#) [新发布](#)

<https://www.theverge.com/tech/941664/microsoft-ai-model-reasoning-mai-thinking-1-build-2026>

★ 2. Holo3.1: 快速本地计算机使用智能体

Hugging Face: [Blog \(RSS\)](#) · 昨天 22:13 · 模型与工具能力

Holo3.1是基于Qwen模型家族的计算机使用智能体系列，旨在提升在桌面、网页和移动环境中的鲁棒性。新模型提供0.8B、4B、9B和35B-A3B四种尺寸，并首次发布量化检查点，包括FP8、Q4 GGUF和NVFP4，以优化本地推理。在AndroidWorld基准测试中，35B-A3B模型得分从67%提升至79.3%。在DGX Spark上，NVFP4量化相

[能力进展](#) [新发布](#)

<https://huggingface.co/blog/Hcompany/holo31>

★ 3. Miso One 开源语音模型：8B 参数、110ms 延迟、一次语音克隆

X: [Kim \(@kimmonismus\)](#) · 7小时前 · 模型与工具能力

Miso One 正式发布，一个8B参数的开源权重语音模型(TTS)，旨在模拟真实人类朗读的温暖与节奏。它支持一次语音克隆(只需短样本)，推理延迟仅110ms。模型权重已开源至GitHub，无需API即可自托管，音频数据不离开本地。API访问即将推出。演示已上线，可先试听再克隆仓库。

[能力进展](#) [新发布](#)

<https://x.com/kimmonismus/status/2062210845308780639>

★ 4. 介绍 Claude Partner Network 的 Services Track 和 Partner Hub

Anthropic: [Newsroom \(网页\)](#) · 10小时前 · 应用与商业化

Anthropic 扩展 Claude Partner Network，推出 Services Track 分级体系和 Partner Hub 门户。Services Track 设 Select、Preferred、Global Premier 三级，按认证人数、投产客户数及客户推荐量化评定。Partner Hub 提供每日更新仪表盘和公开目录，方便合作伙伴查看进展、客户寻找供应商。该网络三月

[能力进展](#) [基础设施](#) [新发布](#)

<https://www.anthropic.com/news/services-track-partner-hub>

★ 5. OpenClaw 2026.6.1发布：新增Windows节点与技能工坊

X: [OpenClaw \(@openclaw\)](#) · 2小时前 · 应用与商业化

OpenClaw 2026.6.1 已上线 ☑️原生 Windows 节点主机 ☑️用于自主学习型智能体的技能工坊 (Skill Workshop) ☑️工作板 (Workboard) 编排 ☑️支持 MiniMax M3 Windows 加入集群。无需企鹅服。https://github.com/openclaw/openclaw/releases/tag/v2026.6.1

[能力进展](#) [基础设施](#) [新发布](#)

<https://x.com/openclaw/status/2062288421406785710>

★ 6. OpenShell v0.0.55 发布：新增 Vertex AI 推理支持

X: [NVIDIA AI \(@NVIDIAAI\)](#) · 7 小时前 · 应用与商业化

OpenShell v0.0.55 修复 Google Vertex AI 推理提供者 基于配置文件的策略可见性 修复关中更好的 Podman 检测 修复 GPU procs 基准行为 修复 CI 与文档修复 运行智能体对接 Vertex AI，同时拥有改进的策略可见性以及更可靠的 Podman 和 GPU 沙箱行为。 <https://github.com/NVIDIA/OpenS>

能力进展 基础设施 新发布

<https://x.com/NVIDIAAI/status/2062210034109677665>

★ 7. 消息称 DeepSeek 首轮融资拟筹集 500 亿元，腾讯、宁德时代等参投

IT之家 (RSS) · 19 小时前 · 产业与基础设施

DeepSeek 计划在首轮融资中从腾讯、宁德时代等投资方筹集约 500 亿元人民币。融资落地后，其投后估值预计在 3500 亿至 4000 亿元。创始人梁文峰将出资 200 亿元，腾讯和宁德时代分别拟投资 100 亿元和 50 亿元，成为最大外部投资者。DeepSeek 去年凭借 V3 大模型与 R1 推理模型获得广泛认可。此次融资凸显了中国正打造从大模型到算力能源基建的全链条 AI 产业。

能力进展 基础设施 监管/资本

<https://www.ithome.com/0/959/249.htm>

★ 8. DigitalOcean AI 云服务上线 OpenRouter

X: [OpenRouter \(@OpenRouter\)](#) · 昨天 00:00 · 应用与商业化

新增服务商：DigitalOcean 的 AI-Native Cloud 现已在 OpenRouter 上线。提供高性能推理，覆盖热门开源权重模型。在 DeepSeek V3.2 的输出速度和延迟方面排名第一（数据来自 @ArtificialAnlys）。查看其数据并试用模型：<https://openrouter.ai/provider/digitalocean>

能力进展 基础设施 新发布

<https://x.com/OpenRouter/status/2061840338973806961>

能力 模型与工具能力

1. 微软首款高级推理 AI 模型 MAI-Thinking-1 发布

The Verge: AI (RSS) · 昨天 02:12

微软在 Build 2026 上发布了其首款高级推理 AI 模型 MAI-Thinking-1。该模型被定位为“中等规模”，能在“关键”软件工程基准测试中达到领先模型的水平。微软称其完全从头使用干净数据进行训练，未涉及从第三方模型进行知识蒸馏。这标志着微软在自研 AI 模型上迈出重要一步，此前其主要依赖 OpenAI。近期两家公司已重新协商合作协议，关系有所松绑。

能力进展 基础设施 新发布

<https://www.theverge.com/tech/941664/microsoft-ai-model-reasoning-mai-thinking-1-build-2026>

2. Holo3.1：快速本地计算机使用智能体

Hugging Face: Blog (RSS) · 昨天 22:13

Holo3.1 是基于 Qwen 模型家族的计算机使用智能体系列，旨在提升在桌面、网页和移动环境中的鲁棒性。新模型提供 0.8B、4B、9B 和 35B-A3B 四种尺寸，并首次发布量化检查点，包括 FP8、Q4 GGUF 和 NVFP4，以优化本地推理。在 AndroidWorld 基准测试中，35B-A3B 模型得分从 67% 提升至 79.3%。在 DGX Spark 上，NVFP4 量化相

能力进展 新发布

<https://huggingface.co/blog/Hcompany/holo31>

3. Miso One 开源语音模型：8B 参数、110ms 延迟、一次语音克隆

X: [Kim \(@kimmonismus\)](#) · 7 小时前

Miso One 正式发布，一个 8B 参数的开源权重语音模型（TTS），旨在模拟真实人类朗读的温暖与节奏。它支持一次语音克隆（只需短样本），推理延迟仅 110ms。模型权重已开源至 GitHub，无需 API 即可自托管，音频数据不离开本地。API 访问即将推出。演示已上线，可先试听再克隆仓库。

能力进展 新发布

<https://x.com/kimmonismus/status/2062210845308780639>

4. Grok Imagine 1.5 预览版发布

X: [cb_doge \(@cb_doge\)](#) · 5 小时前

Grok Imagine 1.5 预览版已发布，即日起可在 API 中体验。SpaceXAI 正在发力。

能力进展 新发布

https://x.com/cb_doge/status/2062242490745594085

5. Ideogram v4.0 发布：2K 分辨率和 JSON 提示支持

X: [Krea AI \(@krea_ai\)](#) · 6 小时前

介绍 Ideogram v4.0。原生 2K 分辨率，出色的文字渲染，支持 JSON 提示词。立即在 Krea 中体验。

新发布

https://x.com/krea_ai/status/2062227837130887567

1. 消息称 DeepSeek 首轮融资拟筹集 500 亿元，腾讯、宁德时代等参投

IT之家 (RSS) · 19 小时前

DeepSeek 计划在首轮融资中从腾讯、宁德时代等投资方处筹集约 500 亿元人民币。融资落地后，其投后估值预计在 3500 亿至 4000 亿元。创始人梁文峰将出资 200 亿元，腾讯和宁德时代分别拟投资 100 亿元和 50 亿元，成为最大外部投资者。DeepSeek 去年凭借 V3 大模型与 R1 推理模型获得广泛认可。此次融资凸显了中国正打造从大模型到算力能源基建的全链条 AI 产业。

能力进展 基础设施 监管/资本

<https://www.ithome.com/0/959/249.htm>

2. Sensor Tower: OpenAI 旗下 ChatGPT 月活已破 10 亿，史上最快

IT之家 (RSS) · 22 小时前

据市场情报机构 Sensor Tower 估计，OpenAI 旗下 ChatGPT 全球月活跃用户在 2025 年 5 月突破 10 亿，成为达成此里程碑最快的应用，增速超过 Google Maps、TikTok 等。同期，其竞争对手 Anthropic 的 Claude 月活达 5600 万，同比增幅约显著提升，增长势头迅猛。数据显示，部分用户已开始在 Claude 与 ChatGPT 间切换

能力进展 新发布

<https://www.ithome.com/0/959/083.htm>

3. Qwen Cloud全球AI黑客马拉松启动

X: 阿里云 / Alibaba Cloud (@alibaba_cloud) · 14 小时前

首届Qwen Cloud全球AI黑客马拉松现已启动！参与5大高级赛道，在全球舞台上挑战AI智能体极限。总奖金池超过70,000美元（赛道冠军1万美元）。立即在Devpost报名：<https://click.qwencloud.com/m/20000000281/>

能力进展 基础设施

https://x.com/alibaba_cloud/status/2062113338994172169

4. Nathan Lambert离开Ai2，结束2.5年OLMO等项目工作

X: Nathan Lambert (@natolambert) · 昨天 22:13

Ai2 (Allen Institute for AI) 研究员Nathan Lambert宣布离职。他在Ai2工作超过2.5年，期间主导或参与了OLMO和Tulu等开源模型项目，称其为职业生涯的巅峰。他表示将暂时休息，未来仍会继续深耕开源模型与开放科学领域。

能力进展 新发布

<https://x.com/natolambert/status/2061813361848029631>

5. Alphabet拟融资800亿美元 Anthropic提交IPO申请

Bloomberg: Technology (RSS) · 昨天 02:39

Alphabet宣布拟通过股权融资800亿美元，用于扩展AI基础设施。Anthropic已秘密提交IPO申请，在上市竞赛中领先于竞争对手OpenAI。此外，SpaceX正与华尔街机构协商其IPO的承销费用，HPE则因AI基础设施需求旺盛，年度销售预期超出市场估计。

监管/资本 新发布

<https://www.bloomberg.com/news/videos/2026-06-02/bloomberg-tech-6-2-2026-video>

6. 欧盟公布全面技术主权计划，推动芯片与AI自主发展

Bloomberg: Technology (RSS) · 12 小时前

欧盟公布一项全面技术主权计划，旨在扩大本土半导体、人工智能和云计算供应链，以减少对美国和亚洲的依赖。该计划覆盖芯片制造、AI基础设施及云计算服务等多个关键领域，目标是增强欧洲在核心技术上的自主能力。

基础设施 监管/资本

<https://www.bloomberg.com/news/articles/2026-06-03/europe-unveils-sweeping-tech-sovereignty-plan-to-boost-chips-ai>

7. 微软与OpenAI分道扬镳--如今双方准备正面交锋

The Verge: 订阅版科技 (RSS) · 10 小时前

微软与OpenAI的合作关系已彻底破裂，双方进入正面竞争态势。前DeepMind高管、现任微软AI主管Mustafa Suleyman明确表示，微软现在必须从头证明自己能独立完成所有必需的任务。这标志着两家科技巨头从紧密合作伙伴转变为直接竞争对手。

新发布

<https://www.theverge.com/ai-artificial-intelligence/942242/microsoft-build-ai-agents-openai-competition>

8. 宏利香港与阿里云达成AI战略合作

X: 阿里云 / Alibaba Cloud (@alibaba_cloud) · 21 小时前

宏利香港宣布与阿里云建立战略合作伙伴关系，旨在构建一个专注于推进负责任的AI创新，并加速AI技术在业务中部署的合作框架。

基础设施

https://x.com/alibaba_cloud/status/2062006591377829922

9. Suno完成4亿美元D轮融资

X: Suno (@suno) · 9 小时前

我们激动地宣布Suno的新篇章：4亿美元D轮融资，估值54亿美元！我们的使命一直很简单：让更多人能体验制作音乐的乐趣。非常感谢我们不可思议的社区和投资者与我们共同建设。点击此处阅读Mikey的博客：<https://suno.com/blog/series-d-announcement>

监管/资本

<https://x.com/suno/status/2062183524887675243>

10. Anthropic支持美国AI行政令实施

X: [Anthropic \(@AnthropicAI\)](#) · 昨天 05:35

这项行政令是加强美国AI领导地位的重要一步。我们期待与白宫合作，支持其实施。 <https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/>

<https://x.com/AnthropicAI/status/2061924580222968183>

11. Satya Nadella 谈微软 Build 大会主旨演讲

X: [Satya Nadella \(@satyanadella\)](#) · 20 小时前

与我们共同构建前沿智能生态系统。今早我在 Microsoft Build 上的主旨演讲亮点。

<https://x.com/satyanadella/status/2062022060176801826>

应用 应用与商业化

1. 介绍 Claude Partner Network 的 Services Track 和 Partner Hub

Anthropic: [Newsroom \(网页\)](#) · 10 小时前

Anthropic 扩展 Claude Partner Network，推出 Services Track 分级体系和 Partner Hub 门户。Services Track 设 Select、Preferred、Global Premier 三级，按认证人数、投产客户数及客户推荐信量化评定。Partner Hub 提供每日更新仪表盘和公开目录，方便合作伙伴查看进展、客户寻找供应商。该网络三月

能力进展 基础设施 新发布

<https://www.anthropic.com/news/services-track-partner-hub>

2. OpenClaw 2026.6.1发布：新增Windows节点与技能工坊

X: [OpenClaw \(@openclaw\)](#) · 2 小时前

OpenClaw 2026.6.1 已上线 原生 Windows 节点主机 用于自主学习型智能体的技能工坊 (Skill Workshop) 工作板 (Workboard) 编排 支持 MiniMax M3 Windows 加入集群。无需企鹅服。 <https://github.com/openclaw/openclaw/releases/tag/v2026.6.1>

能力进展 基础设施 新发布

<https://x.com/openclaw/status/2062288421406785710>

3. OpenShell v0.0.55 发布：新增 Vertex AI 推理支持

X: [NVIDIA AI \(@NVIDIAAI\)](#) · 7 小时前

OpenShell v0.0.55 Google Vertex AI 推理提供者 基于配置文件的策略可见性 网关中更好的 Podman 检测 恢复 GPU procs 基线行为 CI 与文档修复 运行智能体对接 Vertex AI，同时拥有改进的策略可见性以及更可靠的 Podman 和 GPU 沙箱行为。 <https://github.com/NVIDIA/OpenS>

能力进展 基础设施 新发布

<https://x.com/NVIDIAAI/status/2062210034109677665>

4. DigitalOcean AI云服务上线OpenRouter

X: [OpenRouter \(@OpenRouter\)](#) · 昨天 00:00

✦ 新增服务商：DigitalOcean 的 AI-Native Cloud 现已在 OpenRouter 上线。提供高性能推理，覆盖热门开源权重模型。在 DeepSeek V3.2 的输出速度和延迟方面排名第一（数据来自 @ArtificialAnlys）。查看其数据并试用模型： <https://openrouter.ai/provider/digitalocean>

能力进展 基础设施 新发布

<https://x.com/OpenRouter/status/2061840338973806961>

5. Cursor Enterprise 推出 Organizations 组织管理功能

[Cursor Blog](#) · 12 小时前

Cursor Enterprise 正式推出 Organizations 结构，允许企业在统一面板中管理多个团队。每个团队可独立设置预算、安全策略、模型访问和功能控制。新增 Groups 作为跨团队或团队内的轻量级用户集合，用于分段管理模型访问、花销上限和智能体权限，不同设置取最宽松权限生效。管理员可创建沙箱团队预先测试新功能，再向全公司推送；也可按部门划分模型访问和预算。组织级仪表盘汇总所有团队

能力进展 监管/资本 新发布

<https://cursor.com/blog/organizations>

6. NVIDIA发布自进化Hermes智能体

X: [NVIDIA AI \(@NVIDIAAI\)](#) · 昨天 02:00

自进化Hermes智能体：随使用而改进的企业AI | Nemetron Labs

能力进展 基础设施 新发布

<https://x.com/NVIDIAAI/status/2061870499232190967>

7. 黄仁勋与纳德拉共议智能体AI时代

X: [NVIDIA \(@nvidia\)](#) · 6 小时前

智能体AI时代来了。从台北，黄仁勋与@satyanadella共同出席#MSBuild，展示NVIDIA与@Microsoft如何携手构建，从Windows设备到规模化AI工厂。▶ 观看对话： <https://nvda.ws/4uefQbs>

能力进展 基础设施

<https://x.com/nvidia/status/2062228974273716457>

8. Google DeepMind 开源科学智能体工具包

X: [Google AI for Developers \(@googleaidevs\)](#) · 昨天 05:34

构建用于科学发现的自主智能体? [@GoogleDeepMind Science Skills](#) 现已在 GitHub 上发布。我们已开源这个专用工具包, 以科学基础和更高的 token 效率加速您的智能体工作流。立即下载 ↓ <https://github.com/google-deepmind/science-skills>

能力进展 新发布

<https://x.com/googleaidevs/status/2061924472245153863>

9. Replit 上线 SEO Agent 助应用被发现

X: [Replit \(@Replit\)](#) · 7 小时前

你发布了你的应用。然后呢? 你的应用可能看起来很棒, 但如果没人能找到它, 它就依然不可见。发布只是开始。认识一下 SEO Agent。它会为你运行一次扫描, 并建议修复措施, 帮助你的应用在网页搜索和 AI 搜索中被发现。

能力进展 新发布

<https://x.com/Replit/status/2062211976995188871>

10. Meta 面向 WhatsApp Business 的 AI 智能体现已全球上线

TechCrunch: [AI \(RSS\)](#) · 10 小时前

Meta 为 WhatsApp Business 打造的 AI 智能体面向全球商家开放, 将按照模型 token 使用量向企业收费。

能力进展 新发布

<https://techcrunch.com/2026/06/03/metas-ai-agent-for-whatsapp-business-is-now-available-globally>

11. NousResearch 发布 Hermes Agent 桌面应用公测版

X: [硅基流动 SiliconFlow \(@SiliconFlowAI\)](#) · 18 小时前

官方 Hermes Agent 桌面应用现已推出!。该条来自 X: [硅基流动 SiliconFlow \(@SiliconFlowAI\)](#), 属于应用与商业化方向, 后续关注其对模型能力、产品形态或产业链节奏的影响。

能力进展 新发布

<https://x.com/SiliconFlowAI/status/2062042813852995899>

12. OpenRouter 上线微软三款新模型

X: [OpenRouter \(@OpenRouter\)](#) · 昨天 03:36

三款新的 @MicrosoftAI 模型现已在 OpenRouter 上线! 同步推出: MAI-Image-2.5、MAI-Transcribe-1.5 和 MAI-Voice-2。详情见下文 [📄](#)

能力进展 新发布

<https://x.com/OpenRouter/status/2061894672847671724>

13. OpenAI Codex 推出团队专属插件

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 昨天 03:11

Codex 中的角色专属插件围绕团队实际工作构建。数据分析、创意制作和产品设计插件为 Codex 提供了创建报告、创意方向和原型的工具与上下文。由 OpenAI 团队构建并使用。

能力进展 新发布

<https://x.com/OpenAIDevs/status/2061888366791246071>

14. 微软发布开源框架 Adaptive Spec-driven Scoring: 支持用文本描述创建 AI 评估测试

TechCrunch: [AI \(RSS\)](#) · 昨天 03:02

微软开源发布 Adaptive Spec-driven Scoring for Evaluation and Regression Testing 框架。开发者可通过文本描述快速生成 AI 行为测试, 用于模型评估与回归测试。

能力进展 新发布

<https://techcrunch.com/2026/06/02/new-microsoft-tool-lets-devs-spin-up-ai-behavior-tests-using-text-descriptions>

15. GitHub Copilot 应用: 智能体原生的桌面体验

GitHub Blog · 昨天 01:30

在微软 Build 2026 大会上, GitHub 发布了新的工具和更新, 并将 Copilot 应用定位为"智能体原生的桌面体验"。其核心目标是让 AI 智能体能够以用户已经习惯的方式进行工作。

能力进展 新发布

<https://github.blog/news-insights/product-news/github-copilot-app-the-agent-native-desktop-experience>

16. Google DeepMind 发布 Gemini 多智能体科研系统

X: [Google DeepMind \(@GoogleDeepMind\)](#) · 昨天 01:08

我们相信 AI 可以成为专属研究伙伴, 帮助发现下一个突破。隆重推出 Co-Scientist: 我们最新的基于 Gemini 的多智能体系统, 能够为复杂科学问题生成、辩论和演进新颖的假设 [📄](#)

能力进展 新发布

<https://x.com/GoogleDeepMind/status/2061857539977842793>

17. OpenAI Codex 发布 Python SDK, 可直接嵌入应用

X: [Vista \(@vista8\)](#) · 昨天 00:25

这个有点厉害, Codex 出 Python SDK了。安装指令: `pip install openai-codex` 整合到自己的代码中, 相当于直接内置了顶级编程和生图Agent? 最关键的是, 可以复用 Codex 登录态。

能力进展 新发布

<https://x.com/vista8/status/2061846741885018296>

18. Grok模型登陆Cloudflare AI Gateway

X: [xAI \(@xai\)](#) · 2 小时前

在 @Cloudflare 的 AI Gateway 上尝试 Grok 模型!

能力进展 基础设施

<https://x.com/xai/status/2062294202625696081>

19. xAI Grok语音模型上线Vapi平台

X: [xAI \(@xai\)](#) · 7 小时前

试试 @Vapi_AI 上最自然的TTS和性价比最高的STT API。来自 @xai 的Grok STT和Grok TTS现已在企业语音AI平台Vapi上线。基于Vapi构建自定义语音智能体, 让它们用客户的语言交流、在受监管的工作流中捕捉重要细节, 并在每次通话中明显更具人性化。

能力进展 监管/资本

<https://x.com/xai/status/2062209374039499178>

20. Claude Platform 新增 CLI 工具

X: [Claude Devs \(@ClaudeDevs\)](#) · 昨天 02:27

我们为 Claude Platform 添加了一个 CLI, 使每个 API 端点都可以从你的终端运行。调用 Messages API, 启动 Claude 托管智能体, 并将结果直接管道传输到你的 shell。ant CLI 被使用 `claude-api` 技能的编码智能体 (Claude Code) 很好地理解。

能力进展

<https://x.com/ClaudeDevs/status/2061877343078244459>

21. Replit 与微软合作发布 Fabric 集成

X: [Replit \(@Replit\)](#) · 昨天 03:26

宣布与 @Microsoft 的新合作 组织现在可以在 Replit 中构建内部工具、工作流或数据仪表盘, 并直接发布到 Microsoft Fabric, 内置安全、身份验证和治理功能。

监管/资本 新发布

<https://x.com/Replit/status/2061892255028486435>

22. OpenClaw 与微软合作进入企业生态

X: [OpenClaw \(@openclaw\)](#) · 昨天 01:56

"你现在可以在公司内部运行 OpenClaw 了。" 宣布我们与 @Microsoft 的合作, 将 OpenClaw 带入微软和 Windows 生态系统。Claws 现在可以在企业环境中安全运行。

监管/资本 新发布

<https://x.com/openclaw/status/2061869633624580452>

23. NVIDIA DGX Station 开始交付开发者和研究人员

X: [NVIDIA \(@nvidia\)](#) · 昨天 04:15

NVIDIA DGX Station 系统正式开始送达开发者和研究人员的办公桌。☑️企业团队可以本地部署数据中心级性能, 搭载 GB300 的系统正从华硕、戴尔、技嘉、惠普、微星和超微等合作伙伴处发货。☑️阅读博客: <https://nvidia.ws/4x3VdBr>

基础设施

<https://x.com/nvidia/status/2061904438478970985>

24. OpenAI Codex Sites 功能发布

X: [OpenAI \(@OpenAI\)](#) · 昨天 00:22

构建应用从未如此简单。通过 Sites, Codex 可以将你的工作、想法和计划转化为一个交互式网站或应用, 你的团队可以通过一个 URL 进行探索、使用和分享。该功能将首先向 Business 和 Enterprise 计划推出, 之后会更广泛地扩展。

新发布

<https://x.com/OpenAI/status/2061845949170045346>

25. MiniCPM-V 4.6 现已支持 vLLM v0.22.0

X: [面壁智能 OpenBMB \(@OpenBMB\)](#) · 昨天 22:02

MiniCPM-V 4.6 现已完全支持 vLLM v0.22.0! 无需自定义分支, 无需额外编译。只需拉取预构建包即可运行。非常感谢 @vllm_project 的顺畅集成! ☑️<http://huggingface.co/openbmb/MiniCPM-V-4.6>

新发布

<https://x.com/OpenBMB/status/2061810723169415205>

26. Replit Canvas 推出多项新功能

X: [Replit \(@Replit\)](#) · 昨天 00:00

Replit Canvas 有一些更新! 了解更多请访问: <http://replit.com/canvas> 展开讨论

新发布

<https://x.com/Replit/status/2061840304534102036>

27. Perplexity Personal Computer 登陆 Windows

X: [Perplexity \(@perplexity_ai\)](#) · 9 小时前

Personal Computer 即将登陆 Windows。面向 Windows 的 Personal Computer 在你的机器上运行, 并协调你每天使用的应用和文件。我们将首先向等候名单上的付费 Max 和 Enterprise Max 订阅用户推送。

https://x.com/perplexity_ai/status/2062189045728596080

研究 研究与开源进展

1. NVIDIA Research 在 CVPR 2026 发表三篇论文: 规模化训练实现抓取、自动驾驶与智能体泛化

NVIDIA AI Blog · 9 小时前

NVIDIA Research 在 CVPR 2026 上发表三篇论文, 展示规模化训练带来的泛化能力。GraspGen-X 是首个零样本抓取基础模型, 基于 20 亿次模拟抓取训练, 可为任意末端执行器生成抓取姿态。LCDrive 用紧凑潜在表示替代文本推理, 让自动驾驶在嵌入式硬件上更快推理。NitroGen 基于 Isaac GR00T 架构, 在大量虚拟环境中训练具身智能体。此外还发布了新的物理 A

能力进展 基础设施 新发布

<https://blogs.nvidia.com/blog/cvpr-research-grasping-driving-agent-training>

2. 世界模型与语言模型: 论具体推理与抽象推理的互补性

HuggingFace Daily Papers (社区热门论文) · 昨天 21:07

本研究探讨了世界模型与多模态大语言模型在预测未来状态时的互补性。世界模型可生成具体的视觉未来轨迹, 但可能视觉合理却任务错误; 多模态大语言模型则擅长抽象推理。为此, 研究提出了"受控的具体推理"框架, 并构建了VRQABench和OpenWorldQA两个基准。同时, 提出了Privileged-Future On-Policy Self-Distillation (PF-OPSD) 方法, 该方法在训练时利用

能力进展 基础设施 新发布

<https://arxiv.org/abs/2606.03603>

3. Anthropic 分析 832 个 AI 恶意账户: 中高风险攻击者半年从 33% 跃至 56%

Anthropic: Research (发表成果 · 网页) · 14 小时前

Anthropic 分析 2025 年 3 月至 2026 年 3 月间 832 个被封禁的恶意账户, 映射至 MITRE ATT&CK 框架。67.3% 使用 AI 编写恶意软件, 6.5% 用于横向移动。六个月间中高风险攻击者占比从 33% 升至 56%。AI 用于账户发现增长 8.9%, AI 辅助钓鱼下降 8.6%。传统基于技术数量或平台 (Claude Code、API、聊天界面) 的威胁评估失效

能力进展

<https://www.anthropic.com/news/AI-enabled-cyber-threats-mitre-attack>

4. 微软研究: 装瓶厂 AI 从聊天到决策

X: [Microsoft Research \(@MSFTResearch\)](#) · 8 小时前

一份在中西部装瓶厂进行的三个月试点显示, 当 AI 超越聊天进入决策领域时会发生什么--约束条件变化、风险真实、答案必须可靠。https://msft.it/6015vjYUN

<https://x.com/MSFTResearch/status/2062204914223169635>

5. 斯坦福大学法学院研究: 人工智能的表现优于法学教授

Hacker News 热门 (buzzing.cc 中文翻译) · 21 小时前

斯坦福大学法学院的一项研究表明, 人工智能的表现优于法学教授。该发现引发了广泛关注, 在 Hacker News 平台获得了 104 个 Points。

<https://law.stanford.edu/press/ai-outperforms-law-professors-in-stanford-law-study>

6. 微软研究: Aurora 天气预报速度超传统超算数千倍

X: [Microsoft Research \(@MSFTResearch\)](#) · 昨天 05:45

天气预报速度比传统超级计算机快数千倍。听听 Kenji Takeda 在 #MSBuild 微软研究实验室关于 Aurora 的分享。了解更多: https://msft.it/6018vjGUA

<https://x.com/MSFTResearch/status/206192718997727450>

格局 观点、资本与监管

1. 洪水韧性新篇章: Google 开源水文建模框架

Google Research: Blog (网页) · 5 小时前

Google Research 在 GitHub 开源了其水文建模框架, 这是一个基于 PyTorch 的 Python 包, 采用与 Flood Hub 河段洪水预报相同的模型架构和训练数据。该框架允许研究者和预报员训练 AI 洪水预报模型、添加新模型与数据, 并已与捷克水文气象研究所 (CHMI) 等合作测试。开放模型架构和训练管线旨在让各国气象水文部门在保留数据完全控制权的同时, 整合基于 AI 的洪水

能力进展 基础设施 新发布

<https://research.google/blog/the-next-chapter-in-flood-resilience-open-sourcing-googles-hydrology-framework>

2. 智能性价比

Tomer Tunguz 博客 (VC 分析) · 昨天 08:00

微软在模型发布卡中首次加入平均token使用量指标。其模型在SWE-Bench Verified上达71.6分，仅消耗约Claude Haiku 4.5三分之一的token。Artificial Analysis的Intelligence Index显示GPT 5.5与Claude Opus 4.8得分相近（约60分），但Opus 4.8运行成本高出40%（\$4, 685 vs \$3, 357）。Ub

能力进展 新发布

<https://www.tomtunguz.com/tokens-per-result>

3. 商汤开源SenseNova-Skills AI办公技能套件

X: 商汤 SenseTime (@SenseTime_AI) · 昨天 22:48

商汤开源了AI办公技能套件SenseNova-Skills。这是一个为任何技能兼容智能体（如OpenClaw与HermesAgent）设计的开源技能集合，提供四大核心功能：图像信息图表生成（可镜像参考风格）、数据分析（支持多表解析、清洗与可视化）、PPT创建（生成大纲内容并智能排版，输出可编辑文件）以及深度研究（跨学术、技术、社交等多源搜索并生成报告）。该技能套件现已完全开源。

能力进展 新发布

https://x.com/SenseTime_AI/status/2061822148076093625

4. Karpathy 的 llm-wiki 项目获超五千星

X: 硅基流动 SiliconFlow (@SiliconFlowAI) · 18 小时前

@karpathy 的 llm-wiki 在几周内获得了 5, 000+ 颗星。其理念是：停止在每个会话中重新发现知识。让一个大语言模型构建并维护一个维基，每次使用时它都会变得更智能。以下是如何使用 @opencode + @justsisyphus OMO + SiliconFlow 构建你自己的版本

能力进展 新发布

<https://x.com/SiliconFlowAI/status/2062054848762450324>

5. Claude Code团队实践：智能体编程如何重塑工程组织与流程

Claude: Blog (网页) · 昨天 00:45

在Code w/ Claude SF 2026活动上，Claude Code工程团队分享了将智能体编程设为默认工作方式后带来的流程与结构变革。核心变化包括：规划转向即时（JIT）模式，强调快速原型与反馈；上下文收集变为“先问Claude”；代码审查中Claude处理风格与测试，人工专注于法律、安全等专业判断。新范式下，工程瓶颈从编写代码转向验证、审查与安全维护。

能力进展 监管/资本

<https://claude.com/blog/running-an-ai-native-engineering-org>

6. Reachy Mini 添加 MCP 工具

Hugging Face: Blog (RSS) · 昨天 08:00

Reachy Mini 推出了公开的 MCP canary Space，用于远程工具调用。

能力进展 新发布

<https://huggingface.co/blog/adding-mcp-tools-to-reachy-mini>

7. 世界模型的功能分类

X: Fei-Fei Li (@drfeifei, World Labs) · 5 小时前

World Labs团队与李飞飞发文，梳理“世界模型”这一被滥用的术语。对比语言模型学习文本统计，世界模型学习空间与时间统计（如光照、物理规律）。基于部分可观马尔可夫决策过程（POMDP）框架，智能体通过动作影响世界状态，观测是部分视图。当前被称为“世界模型”的不同系统本质上是同一循环的不同投影：第一类为渲染器，输出给人眼看的像素，以视觉保真度为核心。文章着重于概念分层，未给出具体模型名、参数或基

能力进展

<https://x.com/drfeifei/status/2062247238143996275>

8. Anthropic 用 Claude 赋能自助数据分析

Claude: Blog (网页) · 6 小时前

Anthropic 使用 Claude 自动化了 95% 的业务分析查询，整体准确率约 95%。其关键在于构建智能体分析栈（agentic analytics stack），通过数据基础层、维护验证流程和技能（skills）分别解决概念-实体歧义、数据过时和检索失败三大错误来源。相比编码场景，数据分析的难点在于将用户问题映射到正确的数据实体，而执行 SQL 反而是简单的。Anthropic 的数据

能力进展

<https://claude.com/blog/how-anthropic-enables-self-service-data-analytics-with-claude>

9. 智能体工程实战窍门全录

X: 邵猛 (@shao__meng) · 23 小时前

该内容源自@mvanhorn的分享，介绍了“智能体工程”如何重塑软件开发。其核心是从“人主导编码”转向“人主导方向、智能体执行”，中心从IDE变为终端与计划文件。方法论遵循Research → Plan → Work循环，核心是让plan.md约束智能体行为。分享者总结了22条实战技巧，涵盖规划、并行执行、输入方式、远程控制等方面，并列出了完整的工具栈。

能力进展

https://x.com/shao__meng/status/2061974983094755575

10. 超越聊天机器人的直接偏好优化

Hugging Face: Blog (RSS) · 11 小时前

Dharma-AI 在 Hugging Face 博客发文，探讨直接偏好优化（Direct Preference Optimization, DPO）在聊天机器人之外的更广泛应用场景。

能力进展

<https://huggingface.co/blog/Dharma-AI/direct-preference-optimization-beyond-chatbots>

11. Claude Code 自我检查与反馈闭环技巧

X: [Claude Devs \(@ClaudeDevs\)](#) · 昨天 03:59

如何让 Claude Code 在交回工作前检查自己的成果？看看如何编码你的手动检查，让 Claude 自己关闭反馈循环：

能力进展

<https://x.com/ClaudeDevs/status/2061900434722496604>

12. Gemini Spark：最令人印象深刻也最可怕的AI体验

[The Verge](#): [订阅版科技 \(RSS\)](#) · 昨天 20:38

Google DeepMind的AI模型Gemini Spark提供了一次极为深刻但同时令人感到不安的用户体验。该模型展现的强大能力令人印象深刻，但其带来的影响和潜力也引发了深刻的恐惧感。

能力进展

<https://www.theverge.com/ai-artificial-intelligence/941388/gemini-spark-ai-agent-trip-planning>

13. Gary Marcus：为什么事情终将崩塌

[Gary Marcus](#): [The Road to AI We Can Trust \(RSS\)](#) · 昨天 23:41

知名人工智能批评者Gary Marcus在其关于可信AI的专栏中，探讨了人工智能发展面临的根本性挑战。文章开篇即指向问题的核心，指出相关数学理论的局限性与人类心理的复杂性，是导致AI系统最终可能出现问题的根源。

<https://garymarcus.substack.com/p/why-things-will-eventually-fall-apart>

14. 优步每月 1,500 美元的 AI 使用上限为 AI 工具定价提供参考

[Hacker News 热门 \(buzzing.cc 中文翻译\)](#) · 5 小时前

优步将 AI 工具每月使用上限定为 1500 美元，这一做法为行业 AI 工具定价提供了有价值的参考信号。

<https://simonwillison.net/2026/Jun/3/uber-caps-usage>
