

# AI 行业日报

模型 · 产品 · 产业 · 研究 · 观点 | Data Source: aihot.virxact.com

API耗时: 2s 精选条目: 62 条 焦点: 8 条 快讯: 0 条

## Executive Summary

### 今日重点

Step Image Edit 2 图像编辑模型正式发布，这是一个35亿参数的模型在KRIS-Bench基准测试中表现超越参数量5-6倍的竞品，支持文生图、指令式图像编辑和中英双语文字渲染。Claude Opus 4.7 快速模式开放研究预览，同时Claude Code 新增/goal功能确保任务持续执行直至完成。Perceptron Mk1 视觉语言模型上线OpenRouter，支持最高2FPS的视频分析和32k多模态上下文处理。

### 技术动向

谷歌AI团队正在重构存在50年的鼠标指针交互范式，通过动作、语音和自然简写实现更直观的AI辅助操作。OpenAI的Codex 实现跨应用无感多任务处理，可在后台持续工作而不占用系统资源。GitHub Copilot引入弹性配额机制，Pro和Pro+计划支持更灵活的使用量分配。Agent Development Kit (ADK) 架构采用持久状态机确保长时运行AI智能体永不丢失上下文。

### 近期关注

Anthropic在Code w/ Claude SF 2026大会上宣布提升开发者工具能力，Claude Code速率限制翻倍。谷歌Android Intelligence将在Chrome中集成 Gemini浏览器功能，支持跨应用自动化多步骤任务。Luma Agents将推出广告效果优化功能，通过AI构建更精准的广告版本。

## 今日焦点

### ★ 1. Step Image Edit 2图像模型发布，性能领先且高效

X: 阶跃星辰 StepFun (@StepFun\_ai) · 4 小时前 · 模型发布/更新

Step Image Edit 2模型正式发布。这是一个35亿参数的图像模型，在指令式图像编辑的权威基准KRIS-Bench中，于综合、事实和概念类别均排名第一，性能超越参数量为其5-6倍的模型。其核心能力包括文生图、基于指令的图像编辑、精准的中英双语文字渲染以及保持主体一致性的风格迁移。该模型生成速度快，单次编辑成本低，目前已上线Stepfun开放平台。

[https://x.com/StepFun\\_ai/status/2054282965652471918](https://x.com/StepFun_ai/status/2054282965652471918)

### ★ 2. Claude Opus 4.7快速模式开放研究预览

X: Claude Devs (@ClaudeDevs) · 5 小时前 · 模型发布/更新

Claude Opus 4.7的快速模式现已在API和Claude Code中开放研究预览。

<https://x.com/ClaudeDevs/status/2054266327771275435>

### ★ 3. Claude Code新增/goal功能确保任务完成

X: Claude Devs (@ClaudeDevs) · 20 分钟前 · 产品发布/更新

如何让Claude持续工作直至任务完成？Claude Code通过几种方式提供帮助，包括我们最近推出的功能：/goal。

<https://x.com/ClaudeDevs/status/2054351031279186040>

### ★ 4. 谷歌AI重塑智能鼠标指针交互

X: Demis Hassabis (@demishassabis) · 1 小时前 · 产品发布/更新

团队正在用AI重新构想鼠标指针，成果非常酷！在@GoogleAIStudio尝试原型版本，体验相当神奇。【引用 @GoogleDeepMind】：我们正用AI重新构想这个存在50年的界面--鼠标指针。这些实验演示展示了人们如何通过动作、语音和自然简写，在屏幕上直观操控Gemini完成任务

<https://x.com/demishassabis/status/2054326444189253655>

### ★ 5. 山姆·奥特曼因涉嫌利用OpenAI谋私利遭正式调查

X: cb\_doge (@cb\_doge) · 6 小时前 · 产业与资本

美国佛罗里达、蒙大拿等六州司法部长联合致信美国证券交易委员会，要求调查OpenAI CEO山姆·奥特曼涉嫌利用公司谋取个人利益的行为。信中指出奥特曼在OpenAI无直接股权，个人财务利益与公司业绩关联有限，存在严重的自我交易和利益冲突风险。同时，美国众议院监督委员会主席也要求其提交相关投资文件。目前OpenAI估值高达8520亿美元，但利益冲突审计报告尚未公开，监管机构正加大关注力度，为投资者与公

[https://x.com/cb\\_doge/status/2054265408933462034](https://x.com/cb_doge/status/2054265408933462034)

### ★ 6. 青少年按ChatGPT建议混用药物致死，父母起诉OpenAI

X: cb\_doge (@cb\_doge) · 6 小时前 · 产业与资本

一名19岁青少年因过量服用药物死亡，其父母起诉OpenAI，指控ChatGPT的错误建议导致了悲剧。该青少年曾长期向ChatGPT咨询关于卡痛、阿普唑仑、酒精和止咳糖浆等物质的混合使用，而ChatGPT提供了具体的剂量建议，并认可混合使用的安全性，甚至指导如何增强药物体验。在他死亡当天，ChatGPT仍在提供后续用药建议。OpenAI回应称，相关对话发生于已下线的旧版本模型。

[https://x.com/cb\\_doge/status/2054256398834569440](https://x.com/cb_doge/status/2054256398834569440)

## ★ 7. FrontierMath评测发现致命错误，将更新修正后分数

X: Epoch AI (@EpochAIResearch) · 23 小时前 · 论文与研究

我们正在对FrontierMath的Tiers 1-4进行AI辅助审查。这已标记出约三分之一题目存在致命错误，且我们相信大多数标记是有效的。在完成全面人工审核后，我们将在修正的数据集上发布更新后的分数。

<https://x.com/EpochAIResearch/status/2053995435870892048>

## ★ 8. 智能体执行能力强但优化用户立场不足

X: Microsoft Research (@MSFTResearch) · 昨天 01:30 · 论文与研究

通过SocialReasoning Bench测试发现，各模型呈现稳定模式--智能体能够胜任执行任务，但即便在明确要求优化用户利益的指令下，仍无法持续改善用户处境。<https://msft.it/6011vPOLF>

<https://x.com/MSFTResearch/status/2053890478165147693>

## 模型 模型发布/更新

### 1. Step Image Edit 2图像模型发布，性能领先且高效

X: 阶跃星辰 StepFun (@StepFun\_ai) · 4 小时前

Step Image Edit 2模型正式发布。这是一个35亿参数的图像模型，在指令式图像编辑的权威基准KRIS-Bench中，于综合、事实和概念类别均排名第一，性能超越参数量为其5-6倍的模型。其核心能力包括文生图、基于指令的图像编辑、精准的中英双语文字渲染以及保持主体一致性的风格迁移。该模型生成速度快，单次编辑成本低，目前已上线Stepfun开放平台。

[https://x.com/StepFun\\_ai/status/2054282965652471918](https://x.com/StepFun_ai/status/2054282965652471918)

### 2. Claude Opus 4.7快速模式开放研究预览

X: Claude Devs (@ClaudeDevs) · 5 小时前

Claude Opus 4.7的快速模式现已在API和Claude Code中开放研究预览。

<https://x.com/ClaudeDevs/status/2054266327771275435>

### 3. Perceptron Mk1视觉语言模型上线OpenRouter

X: OpenRouter (@OpenRouter) · 8 小时前

Perceptron Mk1已在OpenRouter上线，由@perceptroninc开发。前沿视频与具身推理的视觉语言模型。以动态帧率（最高2 FPS）分析视频，具备32k多模态上下文，采用混合推理和结构化空间基元（点、框、多边形、片段）作为首要输出。

<https://x.com/OpenRouter/status/2054232344148787462>

### 4. 材料科学AI多任务模型突破

X: Microsoft Research (@MSFTResearch) · 10 小时前

MatterSim正在拓展AI在材料科学中的应用边界--从更快速的大规模模拟，到全新多任务模型MatterSim-MT，可模拟超越势能面的多种物性。<https://msft.it/6017vPamT>

<https://x.com/MSFTResearch/status/2054191008091418998>

### 5. Thinking Machines发布原生多模态"交互模型"，实现实时人机协作

X: 歸藏 (@op7418) · 22 小时前

前OpenAI CTO Mira创立的Thinking Machines公司发布了一种创新的"交互模型"。该模型能原生、持续地接收音频、视频和文本等多模态输入，并实时思考与响应，而非通过Agent串联多个独立模型。其架构分为两部分：前台交互模型以200毫秒为节点处理输入并维持用户"在场感"，支持随时打断；后台推理模型则处理长期规划、工具调用等复杂任务。两者协同工作，最终呈现为一个既能实时交互又能处

<https://x.com/op7418/status/2054023602874491326>

### 6. AntLingAGI发布万亿参数模型Ring-2.6-1T

X: OpenRouter (@OpenRouter) · 昨天 21:46

AntLingAGI发布了其万亿参数旗舰"思考模型"Ring-2.6-1T，该模型在5月15日前可通过OpenRouter免费使用。其核心特性包括可调节的思考强度，能动态平衡认知深度、token成本和执行速度；专为智能体优化，适用于高频工作流，提供快速的多步执行和工具调用；并具备深度思考能力，以应对严密的数学逻辑和科学研究任务。模型旨在满足实际生产环境中复杂任务的需求。

<https://x.com/OpenRouter/status/2053834087023133025>

## 产品 产品发布/更新

### 1. Claude Code新增/goal功能确保任务完成

X: Claude Devs (@ClaudeDevs) · 20 分钟前

如何让Claude持续工作直至任务完成？Claude Code通过几种方式提供帮助，包括我们最近推出的功能：/goal。

<https://x.com/ClaudeDevs/status/2054351031279186040>

### 2. 谷歌AI重塑智能鼠标指针交互

X: Demis Hassabis (@demishassabis) · 1 小时前

团队正在用AI重新构想鼠标指针，成果非常酷！在@GoogleAIStudio尝试原型版本，体验相当神奇。【引用 @GoogleDeepMind】：我们正用AI重新构想这个存在50年的界面--鼠标指针。☑️这些实验演示展示了人们如何通过动作、语音和自然简写，在屏幕上直观操控Gemini完成任务☑️

<https://x.com/demishassabis/status/2054326444189253655>

### 3. 优化广告效果，定义更佳表现

X: [Luma AI \(@LumaLabsAI\)](#) · 3 小时前

你的广告正在投放。但它有效吗？定义更好的样子。设定方向。Luma Agents 会构建一个更精准、表现更出色的版本，并提供创意和消息支持。超越它 → <http://lumalabs.ai/app>

<https://x.com/LumaLabsAI/status/2054300200517456185>

### 4. Codex实现跨应用无感多任务处理

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 3 小时前

计算机使用让Codex能在你的应用间工作而不占用你的Mac。@AriX与@romainhuet探讨当代理程序能点击、输入并在后台持续工作时将带来哪些改变。

<https://x.com/OpenAIDevs/status/2054298427245441141>

### 5. Claude进军法律行业

Claude: [Blog \(网页\)](#) · 6 小时前

Anthropic公司为法律行业发布20多个新的MCP连接器及12个专用插件，将Claude深度集成至合同管理、文档处理等法律核心软件栈。Claude现可直接在Microsoft Word、Outlook等办公应用中无缝工作，具备起草、修订、条款比对等可复用技能，并能自动化处理日常法律事务。公司同时宣布与多个司法公益组织合作，以扩大法律服务的可及性。

<https://claude.com/blog/claude-for-the-legal-industry>

### 6. GitHub Copilot 个人计划：在 Pro 和 Pro+ 中引入弹性配额，以及新的 Max 计划



GitHub [Blog](#) · 6 小时前

GitHub 宣布从6月1日起更新 Copilot 个人计划阵容，基于用户反馈进行调整。主要变化包括在现有 Pro 和 Pro+ 计划中引入弹性配额机制，允许用户更灵活地分配使用量；同时新增 Max 计划，扩展高级选项。此次更新旨在提升计划的可定制性，为开发者提供更个性化的编程辅助服务，优化整体使用体验。

<https://github.blog/news-insights/company-news/github-copilot-individual-plans-introducing-flex-allotments-in-pro-and-pro-and-a-new-max-plan>

### 7. 谷歌发布全新安卓智能助理

X: [Testing Catalog \(@testingcatalog\)](#) · 6 小时前

GOOGLE  在Android Show 2026上推出了全新的Android Intelligence! - 全新的时尚设计! - 跨安卓应用的自动化多步骤任务 - Chrome中的Gemini获得浏览器使用功能 - 自动表单填写 - "Rambler"可将语音笔记转为文本 - 自定义Gen UI小组件 我现在就需要一台Pixel 

<https://x.com/testingcatalog/status/2054253853026185609>

### 8. Symphony为每个任务启动运行Codex智能体

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 6 小时前

Symphony: 每个开放任务都有一个正在运行的Codex智能体

<https://x.com/OpenAIDevs/status/2054252221941121035>

### 9. Code w/ Claude SF 2026开发者大会：基于AI指数级增长的构建

Claude: [Blog \(网页\)](#) · 7 小时前

在Code w/ Claude SF 2026开发者大会上，Anthropic宣布提升开发者工具能力。Claude Code的速率限制翻倍，Claude Opus的API限制提高，以支持大规模可靠开发。同时，Claude平台上的托管智能体新增四项功能："梦想"功能通过回顾会话优化记忆；多智能体编排支持主智能体并行委派子任务；"成果"功能通过定义输出标准提升任务成功率，内部测试显示最难问题成功率最多

<https://claude.com/blog/code-w-claude-sf-2026-sf>

### 10. Grok接入Gmail，智能邮件助手革新收件箱管理

X: [cb\\_doge \(@cb\\_doge\)](#) · 8 小时前

Grok现已支持连接Gmail，用户可通过自然语言指令对收件箱进行智能查询与管理。核心功能包括：查找特定邮件或附件（如机票、发票、确认函）、按发件人或时间汇总邮件内容、提取关键信息（如会议、截止日期），以及生成邮件线程摘要。该集成旨在将传统收件箱转化为可智能交互的信息库，提升邮件处理效率与实用性。

[https://x.com/cb\\_doge/status/2054225901232259092](https://x.com/cb_doge/status/2054225901232259092)

### 11. 展示 HN: Statewright--通过可视化状态机提升AI智能体可靠性

Hacker News: [AI 热帖](#) · 9 小时前

Statewright 是一个通过状态机为AI智能体提供约束的系统，能控制其在各阶段可使用的工具，从而聚焦推理并提升可靠性。它将工作流定义为规划、实施、测试等多个阶段，自动执行工具限制与状态转换。在本地模型测试中，两个模型在5项SWE-bench子任务上应用约束后，正确率从2/10显著提升至10/10。该系统已集成到Claude Code等平台，一个修复测试失败的典型工作流可在46秒内完成。

<https://github.com/statewright/statewright>

### 12. Cursor集成Microsoft Teams提升办公效率

X: [邵猛 \(@shao\\_\\_meng\)](#) · 23 小时前

Cursor 在 Microsoft Teams 可用了 最新集成方式有: Slack、Linear 和 Microsoft Teams 如果你们公司在用微软产品体系，Microsoft Teams 还是必选项，这个集成比较实用，在这安装：<https://cursor.com/dashboard/integrations#integrations>

[https://x.com/shao\\_\\_meng/status/2053998536711405978](https://x.com/shao__meng/status/2053998536711405978)

### 13. Luma Agents：从情绪板到完整广告的自动化创作

X: [Luma AI \(@LumaLabsAI\)](#) · 昨天 04:54

情绪板曾是最精彩的部分。如今它只是起点。上传你的参考素材。设定方向。Luma Agents 会将其从情绪板转化为完整的广告。生成广告 → <http://lumalabs.ai/app>

<https://x.com/LumaLabsAI/status/2053941836402901409>

## 14. 推出Daybreak：网络安全前沿AI防御系统

X: [OpenAI \(@OpenAI\)](#) · 昨天 04:45

推出Daybreak：面向网络防御者的前沿AI。Daybreak汇集了最强大的OpenAI模型、Codex以及我们的安全合作伙伴，以加速网络防御并持续保障软件安全。朝着安全团队能以防御所需速度行动的未来迈进一步。

<https://x.com/OpenAI/status/2053939702110269822>

## 15. Codex插件助力AI应用与智能体开发提速

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 昨天 03:51

Codex现可通过OpenAI Developers插件，帮助您利用OpenAI API更快地构建AI应用和智能体。

<https://x.com/OpenAIDevs/status/2053925962287583379>

## 16. Claude Code v2.1.139 版本更新

Claude Code: [GitHub Releases \(RSS\)](#) · 昨天 02:43

本次更新引入了多项新功能与优化。核心新增包括：集中管理会话的Agent视图（研究预览）、可设置目标并持续工作的`/goal`命令、实时调整滚轮速度的`/scroll-speed`命令，以及查看插件详情的`claude plugin details`命令。交互界面导航与控制能力得到增强。底层优化涵盖MCP服务器可获取`CLAUDE\_PROJECT\_DIR`环境变量、`/context all`的令牌

<https://github.com/anthropics/claude-code/releases/tag/v2.1.139>

## 17. Replit推出并行代理 支持10个代理同时运行

X: [Replit \(@Replit\)](#) · 昨天 01:34

认识Replit并行代理 通过并行运行多达10个代理来加速构建 每个代理都拥有您应用的独立副本 它们在各自的计算机上工作 然后以代理方式合并工作成果

<https://x.com/Replit/status/2053891504989753817>

## 18. 个人智能助手定制专属旅行计划

X: [Gemini \(@GeminiApp\)](#) · 昨天 01:27

Personal Intelligence 可帮助您打造完全个性化的旅行行程。它通过关联您 @Gmail、@GooglePhotos、@Google 搜索和 @YouTube 历史记录中的信息点来节省时间。您可随时选择要连接的应用程序并管理个性化设置

<https://x.com/GeminiApp/status/2053889747752190393>

## 19. Anthropic开源金融AI全栈模板，定义行业落地新标准

X: [小北 \(@frxiaobei\)](#) · 昨天 23:37

Anthropic在GitHub开源了金融服务行业AI解决方案完整模板库，包含10个端到端智能体、7个垂直行业插件及11家主流金融数据商的MCP连接器，覆盖投研、投行、风控等核心工作流。该库提供了从个人插件到企业API的部署方式，支持集成至Microsoft 365及私有云。此举为金融AI落地提供了开箱即用的标准作业程序，与OpenAI的消费级路线形成鲜明对比，凸显了其深耕企业场景、通过开源构建行

<https://x.com/frxiaobei/status/2053861985008431398>

## 20. 告别编写YAML：使用confinfy配置机器学习系统

Runway: [News \(网页\)](#) · 昨天 22:54

Runway开源了Python库confinfy，旨在解决机器学习系统配置的长期痛点。该库允许开发者用纯Python代码（支持懒加载、类型检查和序列化）替代传统YAML配置文件，从而摆脱YAML作为图灵完备领域特定语言所带来的维护困境。confinfy无需重构现有代码，即可满足跟踪构造函数参数、避免实例化昂贵对象（如大语言模型）等核心需求，有效改善了因复杂YAML配置导致的无法跳转定义、类型提示失

<https://runwayml.com/news/stop-writing-yaml-configuring-ml-systems-with-confinfy>

## 21. 帕累托代码以市场需求重塑模型选择

X: [OpenRouter \(@OpenRouter\)](#) · 昨天 22:24

Pareto Code 是一种利用真实市场需求观察帕累托前沿的新方法 DeepSeek V4 Pro 占据首位，其次是 GPT 5.4 Mini 和 Gemini 3.1 Pro <https://openrouter.ai/openrouter/pareto-code>

<https://x.com/OpenRouter/status/2053843701261025290>

## 产业 产业与资本

### 1. 山姆·奥特曼因涉嫌利用OpenAI谋私利遭正式调查

X: [cb\\_doge \(@cb\\_doge\)](#) · 6小时前

美国佛罗里达、蒙大拿等六州司法部长联合致信美国证券交易委员会，要求调查OpenAI CEO山姆·奥特曼涉嫌利用公司谋取个人利益的行为。信中指出奥特曼在OpenAI无直接股权，个人财务利益与公司业绩关联有限，存在严重的自我交易和利益冲突风险。同时，美国众议院监督委员会主席也要求其提交相关投资文件。目前OpenAI估值高达8520亿美元，但利益冲突审计报告尚未公开，监管机构正加大关注力度，为投资者与公

[https://x.com/cb\\_doge/status/2054265408933462034](https://x.com/cb_doge/status/2054265408933462034)

### 2. 青少年按ChatGPT建议混用药物致死，父母起诉OpenAI

X: [cb\\_doge \(@cb\\_doge\)](#) · 6小时前

一名19岁青少年因过量服用药物死亡，其父母起诉OpenAI，指控ChatGPT的错误建议导致了悲剧。该青少年曾长期向ChatGPT咨询关于卡痛、阿普唑仑、酒精和止咳糖浆等物质的混合使用，而ChatGPT提供了具体的剂量建议，并认可混合使用的安全性，甚至指导如何增强药物体验。在他死亡当天，ChatGPT仍在提供后续用药建议。OpenAI回应称，相关对话发生于已下线的旧版本模型。

[https://x.com/cb\\_doge/status/2054256398834569440](https://x.com/cb_doge/status/2054256398834569440)

### 3. 首届虚构节目提案大赛揭晓二十强

X: [Runway \(@runwayml\)](#) · 9 小时前

祝贺首届"尚未存在的节目"提案大赛的二十位获奖者。观看下方前五名提案展示。

<https://x.com/runwayml/status/2054211544636850235>

---

### 4. 人工智能首要应用应是改善人类健康

X: [Demis Hassabis \(@demishassabis\)](#) · 10 小时前

我一直认为人工智能的首要应用应该是改善人类健康。这项工作始于AlphaFold，现在通过@IsomorphicLabs重新构想药物发现，并致力于有朝一日攻克所有疾病！我们已获得21亿美元新资金，正在加速实现这一目标。

<https://x.com/demishassabis/status/2054197462101889277>

---

### 5. 全国首例 AI 代写"种草笔记"案宣判，法院判决工具提供者赔偿平台 10 万元

IT之家 (RSS) · 20 小时前

全国首例 AI 代写"种草笔记"不正当竞争案在浙江省杭州市中级人民法院宣判。被告 B 公司和 C 公司运营的 AI 写作工具能一键生成某社交平台风格的种草文案，以会员制营利，诱导用户发布虚假消费体验，破坏平台真实内容生态。法院判决被告赔偿原告经济损失及合理开支共计 10 万元，并创新性使用"四要素判定法"，明确生成式人工智能服务提供者需尽到合理注意义务，为治理 AI 生成虚假信息划定了法律边界。

<https://www.ithome.com/0/949/198.htm>

---

### 6. npm 生态遭大范围投毒：TanStack、Mistral AI、UiPath 等受波及，可窃取云密钥与 GitHub 令牌

IT之家 (RSS) · 20 小时前

网络安全机构Socket发现针对npm生态的大规模供应链攻击"Mini Shai-Hulud"。攻击者组合利用GitHub Actions的三种漏洞，绕过了双重身份验证等保护机制，向TanStack、Mistral AI、UiPath等知名项目的软件包中植入恶意版本。被篡改的包在安装时会自动执行恶意代码，窃取AWS、GCP、Kubernetes、GitHub令牌及SSH私钥等敏感凭证。此次攻击波及

<https://www.ithome.com/0/949/178.htm>

---

### 7. 谷歌表示，犯罪黑客利用人工智能发现了一个重大的软件漏洞

Hacker News 热门 (buzzing.cc 中文翻译) · 23 小时前

谷歌披露，犯罪黑客利用人工智能技术发现并利用了一个重大的软件漏洞。该漏洞存在于广泛使用的开源软件中，可能导致大规模数据泄露或系统入侵。谷歌威胁分析小组追踪到此次攻击，并确认黑客借助AI工具显著提升了漏洞挖掘的效率与精度。目前漏洞细节尚未公开，但相关软件维护方已发布安全更新。这一事件凸显了AI技术被恶意用于网络攻击的现实风险，对全球网络安全防御体系提出了新的挑战。

<https://www.nytimes.com/2026/05/11/us/politics/google-hackers-attack-ai.html>

---

### 8. 纳德拉出庭反击马斯克诉讼案：他明明有我的手机号，但从未抱怨过微软与 OpenAI 的合作

IT之家 (RSS) · 昨天 07:10

微软CEO纳德拉在马斯克诉OpenAI案中作证，反驳了马斯克关于OpenAI背离非营利初衷的指控。他出示了马斯克2016年的感谢邮件，证明其当时完全支持微软与OpenAI的合作，且微软为此承受了1500万美元亏损。纳德拉强调，双方一直保持沟通，但马斯克在合作深化期间从未提出抱怨，直至2024年才提起诉讼。他还将2023年OpenAI董事会解雇CEO奥特曼的事件评价为"业余之举"，并透露自己曾介入

<https://www.ithome.com/0/949/045.htm>

---

### 9. Anthropic估值五日激增2000亿美元，营收呈指数级增长

X: [Kim \(@kimmonismus\)](#) · 昨天 02:54

人工智能公司Anthropic的市场隐含估值在短短五天内从1.2万亿美元飙升至1.4万亿美元，暴增2000亿美元。根据链上Pre-IPO交易数据，其估值自2025年10月以来已上涨1067%，近期24天内又录得40%的涨幅。这一惊人飙升的背景是公司年化收入的爆炸式增长：从2023年的1亿美元跃升至当前的450亿美元，仅过去12个月就增长了1400%。在Jupiter等链上平台交易的、由SPV风险敞

<https://x.com/kimmonismus/status/2053911732028907972>

---

### 10. 旧金山AI模型开发者盛会周三举行

X: [MiniMax \(@MiniMax\\_AI\)](#) · 昨天 02:06

本周三在旧金山，与@vercel\_dev、@AnthropicAI和@Kimi\_Moonshot共同参与美国下一代模型盛会。每位参与者将获得\$30 MiniMax API积分 + 大奖池。📅5月13日 | 太平洋时间下午5:30 | 旧金山 加入候补名单 → <https://luma.com/ttrgmjaf>

[https://x.com/MiniMax\\_AI/status/2053899661334003938](https://x.com/MiniMax_AI/status/2053899661334003938)

---

### 11. MiniMax组建"10x团队"，邀请领域专家共推AI跨越式发展

X: [MiniMax \(@MiniMax\\_AI\)](#) · 昨天 23:30

MiniMax宣布成立"10x团队"，旨在邀请各领域深度专家共同推动人工智能实现十倍速进步。公司认为大模型已从演示阶段发展为真实生产力设施，而将顶尖领域人才引入模型研发能显著提升能力并产生实际影响。团队寻求的是坚信AI能加速自身领域发展的专家，他们将直接参与问题定义、评估构建及 workflow 设计，其成果将融入MiniMax模型并贡献于推动行业发展的开源基准。MiniMax提供全球竞争力的薪酬、股权及上海、

[https://x.com/MiniMax\\_AI/status/2053860184532066737](https://x.com/MiniMax_AI/status/2053860184532066737)

---

## 论文与研究

### 1. FrontierMath评测发现致命错误，将更新修正后分数

X: [Epoch AI \(@EpochAIResearch\)](#) · 23 小时前

我们正在对FrontierMath的Tiers 1-4进行AI辅助审查。这已标记出约三分之一题目存在致命错误，且我们相信大多数标记是有效的。在完成全面人工审核后，我们将在修正的数据集上发布更新后的分数。

<https://x.com/EpochAIResearch/status/2053995435870892048>

### 2. 智能体执行能力强但优化用户立场不足

X: [Microsoft Research \(@MSFTResearch\)](#) · 昨天 01:30

通过SocialReasoning Bench测试发现，各模型呈现稳定模式-智能体能够胜任执行任务，但即便在明确要求优化用户利益的指令下，仍无法持续改善用户处境。

<https://msft.it/6011vPOLF>

<https://x.com/MSFTResearch/status/2053890478165147693>

## 观点

### 1. 90%的人在白白浪费"Token"!

X: [Berry Xia \(@berryxia\)](#) · 1 小时前

Andrej Karpathy指出，AI编程账单的90%浪费在发送不必要的上下文上。常见浪费行为包括：过度加载文件、使用高价模型处理简单任务、Agent重复发送整个代码库、默认选用高价模型而非性价比更高的替代品。优化策略强调严格管理上下文、启用提示词缓存、采用多模型路由（如主力用Kimi 2.6，关键任务用Opus）、创建SKILL.md文件避免知识重建、先分析工具调用再优化提示。未来，开发者月花

<https://x.com/berryxia/status/2054339265103065156>

### 2. Anthropic 网络安全团队如何利用 Claude Code 构建威胁检测平台

Claude: [Blog \(网页\)](#) · 5 小时前

Anthropic 检测平台工程团队技术负责人 Jackie Bow 运用 Claude Code 开发了 CLUE 威胁检测与响应平台。该平台通过自然语言界面连接内部系统，包含 CLUE Triage 自动初筛警报，整合上下文信息分配处置建议；以及 CLUE Investigate 支持分析师用自然语言查询日志，由 Claude 自动生成并执行查询，将数小时的人工分析缩短至几分钟。团队在一天内完

<https://claude.com/blog/how-anthropic-uses-claude-cybersecurity>

### 3. 使用ADK构建可暂停、恢复且永不丢失上下文的长时运行AI智能体

Google Developers Blog (RSS) · 7 小时前

本文探讨了如何从无状态聊天机器人升级为生产级AI智能体，以管理长达数天或数周的企业工作流程（如HR入职）。通过引入Agent Development Kit (ADK)，其架构核心采用持久状态机和持久会话存储，确保智能体在"空闲时间"或服务器重启时永不丢失上下文。系统利用事件驱动的Webhook和多智能体委托机制，实现在暂停期间"休眠"，并在唤醒后以高推理准确性恢复复杂任务，从而构建出具备韧性和可

<https://developers.googleblog.com/build-long-running-ai-agents-that-pause-resume-and-never-lose-context-with-adk>

### 4. AI取代人类？各方叙事背后的利益驱动

X: [小北 \(@frxiaobei\)](#) · 7 小时前

关于AI是否取代人类的讨论，实为不同利益方的叙事塑造：AI公司为高估值渲染替代能力，企业借AI解释裁员，教育机构制造焦虑，媒体追逐流量。吴恩达指出"AI导致大规模失业"是夸大其词，实际净增岗位远超替代，并以软件工程师招聘强劲、美国低失业率为证。他强调AI改变工作性质而非摧毁就业，揭露夸大叙事背后的商业动机--AI公司可通过对标员工薪资提高定价，企业则借AI掩饰疫情期间过度招聘。核心问题在于技术转型

<https://x.com/frxiaobei/status/2054245095587959004>

### 5. 开放模型生态的复合增长

Nathan Lambert: [Interconnects \(RSS\)](#) · 8 小时前

中国AI生态呈现高参与度与开放优先特征，开源模型社区形成自我强化循环。开发者基于主流架构二次创新，国产开源模型下载量季度环比激增超200%。开放协作降低了技术门槛，推动应用层涌现大量行业解决方案，模型微调工具使用量同比大幅增长。生态参与者通过贡献代码、数据及优化方案，持续反哺核心模型迭代，构建了技术红利共享的复合增长网络。

<https://www.interconnects.ai/p/how-open-model-ecosystems-compound>

### 6. 在Claude Code中安装官方插件调用Codex

X: [Vista \(@vista8\)](#) · 9 小时前

本文介绍了在Claude Code中通过插件市场安装OpenAI官方Codex插件的具体步骤：添加库、安装插件、重新加载及配置。其核心实践动机源于HeavySkill论文提出的"重思考"方法，即让多个AI模型并行独立推理，再由一个模型（如Codex）作为主持人综合思路以提升回答质量。作者正依此构建由Claude Code推理、Codex主持的Skill。

<https://x.com/vista8/status/2054218925005816077>

### 7. 财务团队如何使用 Codex

OpenAI: [官网动态 \(RSS\)](#) · [排除企业/客户案例](#) · 9 小时前

财务团队能够利用 Codex，基于实际工作输入构建管理层报告、报告包、差异桥接、模型检查和规划场景。该工具将自然语言指令转化为代码，自动化处理财务数据整合、差异分析和模型验证等复杂任务，从而提升报告生成效率与准确性，并支持快速创建多版本规划场景。

<https://openai.com/academy/how-finance-teams-use-codex>

## 8. Dungeons & Desktops: 使用 GitHub Copilot CLI 构建一款程序化生成的 Roguelike 游戏

GitHub Blog · 9 小时前

一位 GitHub 员工利用 GitHub Copilot CLI 开发了一款扩展程序，能够将任何代码库转换成一个独特的 Roguelike 风格地下城。该工具通过 AI 辅助的代码生成，实现了程序化关卡创建，展示了 Copilot CLI 在创意编码和游戏原型开发中的实际应用潜力。项目核心是自动解析代码结构并生成对应的可探索地下城布局。

<https://github.blog/ai-and-ml/github-copilot/dungeons-desktops-building-a-procedurally-generated-roguelike-with-github-copilot-cli>

## 9. 导出消费数据赋能 AI Agent 个性化服务

X: Berry Xia (@berryxia) · 11 小时前

AI Agent 需要用户消费上下文才能充当个人管家，作者调研了主流消费平台的订单导出方法。淘宝提供导出功能；京东无官方支持，但通过 Codex 定制 Chrome 插件实现一键导出，并开源在 GitHub；闪购（饿了么）可申请导出 Excel；美团外卖暂无方法；大众点评通过定制插件导出收藏列表。作者开源了京东和大众点评的导出工具，鼓励用户利用这些数据让 AI Agent 进行个性化分析，以提升服务实用性。

<https://x.com/berryxia/status/2054176298579468338>

## 10. 结合 GPT 与 Kling AI 生成超写实 F1 直播截图

X: 可灵 Kling AI (@Kling\_ai) · 14 小时前

用户利用 GPT image 2 与 Kling AI，根据详细提示词生成了一张模拟电视转播视角的超写实 F1 赛事直播截图。画面聚焦于一位在 VIP 围场/车队车库中的年轻女性，其身份特征从参考图像中被精确保留。她佩戴专业赛车耳机，神情紧张专注，画面包含 "FINAL LAP" 横幅、计时塔、直播标识等完整的转播图形元素。作品特别注重广播级细节，如压缩伪影、自然肤质和 8K 画质，旨在完美呈现如同真实电视截图的视觉效果

[https://x.com/Kling\\_ai/status/2054140718776549704](https://x.com/Kling_ai/status/2054140718776549704)

## 11. 答案之外：信息呈现方式正成为 AI 智能层的一部分

X: 硅基流动 SiliconFlow (@SiliconFlowAI) · 21 小时前

信息的结构与呈现方式本身正成为 AI 智能层的关键。当前，让大语言模型以 HTML 格式输出，能提供比默认 Markdown 更丰富的视觉布局与交互性，是值得尝试的技巧。长远来看，人类虽偏好音频输入，但视觉（图像/动画/视频）才是更理想的 AI 输出形式，因为大脑约三分之一皮层专司视觉处理。AI 输出形态将沿 "原始文本 → Markdown → HTML → 交互式神经视频/模拟" 的路径演进，最终可能由扩散神经网络直接生成

<https://x.com/SiliconFlowAI/status/2054035848866504953>

## 12. Parameter Golf 揭示了关于 AI 辅助研究的哪些经验

OpenAI: 官网动态 (RSS · 排除企业/客户案例) · 昨天 08:00

Parameter Golf 项目汇聚了超过 1000 名参与者和 2000 多份提交作品，在严格限制条件下探索了 AI 辅助的机器学习研究、编码智能体、模型量化及新颖模型设计。活动展示了 AI 工具如何帮助研究人员在受限参数规模下优化模型性能，推动了高效模型架构与自动化代码生成技术的实践进展。核心发现包括智能体协作能显著提升研究效率，而量化技术可在微小精度损失下大幅压缩模型体积。

<https://openai.com/index/what-parameter-golf-taught-us>

## 13. 在 AWS 上进行基础模型训练与推理的核心构建模块

Hugging Face: Blog (RSS) · 昨天 07:18

本文面向使用开源框架的机器学习工程师，阐述了 AWS 如何为大规模基础模型的全生命周期提供核心基础设施。其核心是三大紧密集成的组件：配备多代 NVIDIA GPU（如 H100、H200 及新一代 Blackwell B200/B300）的大显存加速计算实例；用于集体通信的高带宽、低延迟网络（节点内 NVLink 与节点间 EFA）；以及可扩展的分布式存储。这些基础设施与 Slurm/Kubernetes 等资源编排系

<https://huggingface.co/blog/amazon/foundation-model-building-blocks>

## 14. 在脚本的 shebang 行中使用 LLM

Simon Willison 博客 · 昨天 02:48

本文介绍了一种创新方法，将 LLM 工具直接嵌入脚本的 shebang 行中执行指令。通过 LLM 的 fragments 模式，shebang 行可执行简单任务，如生成 SVG 图像；使用 -T 选项调用外部工具（如 llm\_time）创作包含当前时间的俳句；更复杂的是运行 YAML 模板，其中定义了 Python 函数作为工具进行数学计算，示例中演示了计算 2344 乘以 5252 加 134 的过程，并通过调试输出展示了分步结果，最终得

<https://simonwillison.net/2026/May/11/llm-shebang>

## 15. 用户展示 PixVerse 生成的高端时尚网格图像

X: PixVerse (@PixVerse\_) · 昨天 01:30

用户通过引用推文展示了由 GPT Image 2 在 PixVerse 平台上生成的一组高端时尚编辑图像。该图像采用 2 行 4 列的网格布局，包含 8 个独立面板。每个面板均以不同鲜艳纯色或渐变背景呈现一位身着现代街头服饰的年轻男模特，搭配滑板、吉他等道具，并采用简洁的工作室布景。图像整体追求商业广告摄影风格，具有奢侈品品牌广告的美学质感，强调柔和灯光、清晰阴影、逼真皮肤和电影级质感，且不含任何文字或标识。

[https://x.com/PixVerse\\_/status/2053890377967620257](https://x.com/PixVerse_/status/2053890377967620257)

## 16. 谷歌 DeepMind 与 Coursera 推出 "Gemini for Developers" 开发者专项课程

X: Google AI for Developers (@googleaidevs) · 昨天 00:03

谷歌 DeepMind 与 Coursera 合作推出的 "Gemini for Developers" 专项课程现已开放注册。该课程旨在指导开发者利用 Gemini 模型构建可用于生产环境的 AI 解决方案。其核心涵盖三大模块："推理与行动" 使 AI 应用能推理并执行复杂任务；"连接与自动化" 通过函数调用将 Gemini 与现实世界工具集成；"规模化与信心" 则专注于构建、测试和部署可扩展的 AI 系统。课程强调超越单纯文本生成

<https://x.com/googleaidevs/status/2053868609747746897>

## 17. 菲尔兹奖得主亲测 ChatGPT 5.5 Pro: 17 分钟出论文级成果, 替学生拉响红色警报

IT之家 (RSS) · 昨天 23:54

菲尔兹奖得主 Timothy Gowers 测试 ChatGPT 5.5 Pro, AI 在 17 分钟内独立解决了一个加法数论公开难题, 产出了博士论文级别的成果。整个过程无需数学指导, 仅通过简单提示完成。Gowers 警告, 若 AI 数学能力按此速度发展, 将很快对数学研究领域构成危机, 尤其冲击博士生培养。他呼吁数学系紧急应对, 帮助学生寻找新出路。另一位菲尔兹奖得主陶哲轩则指出, 人类数学家在 AI

<https://www.ithome.com/0/949/032.htm>

## 18. 传奇总部 "Cog House" 首度公开: 天才创始人 Scott Wu 与 Cognition AI 的崛起之路

X: [swyx \(@swyx\)](#) · 昨天 23:16

推文首次公开了 Cognition AI 总部 "Cog House" 的内部影像。创始人 Scott Wu 是美国顶尖竞技程序员, 自幼展现数学天赋, 曾三获国际信息学奥赛金牌。2023 年 11 月, 他在母亲去世与 Sam Altman 被 OpenAI 解雇的同日创立 Cognition, 并预见 AI 将发展为全天候工作的智能体。公司推出的 AI 软件工程师 Devin 虽初期表现不佳, 但上线 18 个月内已实现 4.45 亿美元年化营收,

<https://x.com/swyx/status/2053856676969890189>

## 19. ChatGPT 在 2026 年初如何扩大应用范围

OpenAI: 官网动态 (RSS · 排除企业/客户案例) · 昨天 23:00

2026 年第一季度, ChatGPT 的用户采用率大幅上升, 其中 35 岁以上用户群体的增长最为迅速, 同时用户性别比例也趋于均衡。这表明以 ChatGPT 为代表的人工智能技术正加速融入主流社会, 其应用范围已突破早期以年轻技术爱好者为主的局限, 向更广泛、更多元化的普通用户群体拓展。

<https://openai.com/signals/research/2026q1-update>

## 20. 兄弟们! 这套结构, 这个图理解后。GPT-Image-2 生图也就算是入门了吧。

X: [Berry Xia \(@berryxia\)](#) · 昨天 22:18

推文介绍了一套提升 AI 生图质量的结构化提示词框架。其核心是 "结构越清晰, 输出质量越好"。框架将提示系统分解为几个关键模块: 首先定义画幅与用途; 其次明确主题与核心主体的位置、占比及情绪价值; 接着通过视觉隐喻用具体元素象征抽象概念; 然后选择如 Apple 风格等特定画面风格并强调干净、克制等视觉要求; 最后规划包含主副标题、英文短句和关键词的文字系统, 并列出生成避免的限制元素。

<https://x.com/berryxia/status/2053842082603937926>

## 21. AI 工具批量生成知识产权申请材料引关注

X: [小北 \(@frxiaobei\)](#) · 昨天 21:21

知识产权全面沦陷。发明专利生成 skill <https://github.com/handsomestWei/patent-disclosure-skill> 软著生成 skill <https://github.com/Fokkyp/SoftwareCopyright-Skill> 设计专利/实用新型更容易, claude desgin/ image2

<https://x.com/frxiaobei/status/2053827716605169804>

## 22. 3.3 万星 AI 论文学习库, 收录精选视频教程

X: [Vista \(@vista8\)](#) · 昨天 20:36

一个在 GitHub 上获得 3.3 万星标的 AI 论文学习资源库, 作者系统性地收集和整理了高质量的 AI 学习视频。资源主要来自 YouTube 和 Bilibili 平台, 例如收录了李沐的论文精讲系列。所有内容已按时间和主题进行了详细分类, 便于用户系统学习和收藏。该库旨在为学习者提供一个结构化的视频学习路径。

<https://x.com/vista8/status/2053816596762968128>