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: 63 条 焦点: 8 条 快讯: 0 条

Executive Summary

今日重点

Krea 2作为Krea AI首个基础模型正式开放限量体验，提供访问码供开发者测试。Runway Agent发布，这是一个能够通过单次对话将创意想法转化为完整视频的智能创作伙伴，支持概念提案、故事节奏设计到最终成片的全流程自动化。Claude for Small Business服务包上线，包含连接器和15个预设工作流，深度集成QuickBooks、PayPal等企业工具。Claude Code周限额提升50%，付费计划还将提供专用月度编程额度。

技术动向

ExploitGym基准测试发布，包含898个真实漏洞用于评估AI智能体的漏洞利用能力，结果显示前沿模型已能成功转化相当数量的安全漏洞。OpenAI为Windows平台的Codex构建安全沙箱，通过严格控制文件访问权限和网络限制确保代码生成安全。Meta推出WhatsApp和Meta AI的Incognito Chat功能，推理完全在用户设备硬件安全飞地内进行，服务器不留痕迹。psql_bm25s开源项目实现PostgreSQL精确BM25检索，比传统pg_search快23倍，解决多智能体系统检索性能瓶颈。

近期关注

Claude 4.6系列和Opus 4.7在电脑与浏览器使用能力上将持续优化，其中Opus 4.7支持更高分辨率输入（最大长边2576像素、总像素375万）。Kling AI将在2026年戛纳电影节探讨AI电影制作应用。Qwen-Character推动交互式AI角色在游戏、虚拟伴侣和自适应学习领域的应用，预计提升用户参与度50%以上。

今日焦点

★ 1. AI角色实现记忆共情与主动交互

X：阿里云 / Alibaba Cloud (@alibaba_cloud) · 4 小时前 · 模型发布/更新

如果AI角色能够记忆、共情并主动交互呢？交互式AI的未来已来。无论您是为游戏、虚拟AI伴侣还是自适应学习进行开发，Qwen-Character都能打造沉浸式角色扮演体验，推动参与度加深50%以上并提升用户终身价值 观看完整视频了解运作原理：https：//int.alibabacloud.com/m/1000412854/ #AlibabaCloud #Qwen #QwenChara

https://x.com/alibaba_cloud/status/2054653031472414864

★ 2. Krea 2发布访问码，限量体验

X：Krea AI (@krea_ai) · 7 小时前 · 模型发布/更新

Krea 2 访问码发放！ K2-PRFUF8 / K2-NRWW9E / K2-CAP48S - 每个码可使用50次。 访问链接如下 【引用 @krea_ai】： this is Krea 2. our first foundation model， built completely from scratch for aesthetic diversity and stylistic

https://x.com/krea_ai/status/2054611917365526793

★ 3. Claude 工具 v2.1.141 版本更新

Claude Code：GitHub Releases（RSS） · 1 小时前 · 产品发布/更新

Claude 工具发布 v2.1.141 版本，带来多项功能新增与优化。主要更新包括：为钩子输出添加 `terminalSequence` 字段以支持无控制终端的桌面通知；新增 `CLAUDE_CODE_PLUGIN_PREFER_HTTPS` 环境变量，便于通过 HTTPS 克隆插件源码；引入 `ANTHROPIC_WORKSPACE_ID` 变量以在多工作区联盟中限定令牌范围。会话管理方面，`

https://github.com/anthropics/claude-code/releases/tag/v2.1.141

★ 4. Claude代码周限额提升50%

X：Claude Devs (@ClaudeDevs) · 5 小时前 · 产品发布/更新

Claude代码周限额正在提升50%，即日起持续至7月13日。 现已面向所有Pro、Max、Team及按席位计费的企业用户生效。

https://x.com/ClaudeDevs/status/2054639777685934564

★ 5. Anthropic 首次在 B2B 采用率上超越 OpenAI，Ramp 支出数据显示

The Decoder：AI News（RSS） · 5 小时前 · 产业与资本

根据 Ramp AI 指数数据，Anthropic 在美国企业客户中的采用率达到 34.4%，首次超越 OpenAI 的 32.3%。其业务覆盖范围在一年内增长了四倍。但文章指出，三个因素可能使其领先优势迅速减弱。

https://the-decoder.com/anthropic-overtakes-openai-in-b2b-adoption-for-the-first-time-according-to-ramp-spending-data

★ 6. Kling AI将亮相2026戛纳探讨AI电影制作

X：可灵 Kling AI (@Kling_ai) · 9 小时前 · 产业与资本

Kling AI将于2026年5月18日在戛纳电影节电影市场会议中举办专场活动，主题为"从创意可能到制作现实：Kling AI在电影工作流中的应用"。活动旨在探讨AI辅助电影制作的现状与未来演进。Kling AI通过支持《House of David》、《Born of the Tide》及《RAPHAEL》等项目，展示了AI在好莱坞级制作、全AI生成动画及剧情长片等实际影视生产中的多元化应用。

https://x.com/Kling_ai/status/2054577409064911330

★ 7. ExploitGym：AI智能体能否将安全漏洞转化为真实攻击？

Berkeley RDI：Blog（AI 安全与评测） · 3 小时前 · 论文与研究

由伯克利RDI、马克斯·普朗克安全与隐私研究所、Anthropic、OpenAI及谷歌等机构研究人员组成的团队，发布了名为ExploitGym的新基准测试。该测试包含898个真实漏洞，要求AI智能体根据漏洞描述生成完整的漏洞利用程序。结果显示，前沿AI模型已能成功利用相当数量的漏洞，即使在启用ASLR等标准防御措施后，部分攻击仍能成功。这证明AI已具备自主将漏洞转化为实际攻击的能力，该技术具有双重

https://rdi.berkeley.edu/blog/exploitgym

★ 8. 在 Windows 上构建安全有效的沙箱以启用 Codex

OpenAI：官网动态（RSS · 排除企业/客户案例） · 刚刚 · 技巧与观点

OpenAI 为 Windows 平台上的 Codex 构建了一个安全沙箱环境。该沙箱通过严格控制文件访问权限和实施网络限制，确保了代码生成与执行过程的安全性。这一举措使得基于 Codex 的编码助手能够以高效且受控的方式运行，在提供强大编程辅助功能的同时，有效隔离了潜在风险，保障了用户系统的安全。

https://openai.com/index/building-codex-windows-sandbox

模型 模型发布/更新

1. AI角色实现记忆共情与主动交互

X：阿里云 / Alibaba Cloud (@alibaba_cloud) · 4 小时前

如果AI角色能够记忆、共情并主动交互呢？⌧交互式AI的未来已来。无论您是为游戏、虚拟AI伴侣还是自适应学习进行开发，Qwen-Character都能打造沉浸式角色扮演体验，推动参与度加深50%以上并提升用户终身价值 ⌧观看完整视频了解运作原理：https：//int.alibabacloud.com/m/1000412854/ #AlibabaCloud #Qwen #QwenChara

https://x.com/alibaba_cloud/status/2054653031472414864

2. Krea 2发布访问码，限量体验

X：Krea AI (@krea_ai) · 7 小时前

Krea 2 访问码发放！ K2-PRFUF8 / K2-NRWW9E / K2-CAP48S - 每个码可使用50次。 访问链接如下 ⌧【引用 @krea_ai】： this is Krea 2. our first foundation model， built completely from scratch for aesthetic diversity and stylistic

https://x.com/krea_ai/status/2054611917365526793

3. SenseNova-U1 技术报告深度发布：前沿原生多模态模型构建全指南

X：商汤 SenseTime (@SenseTime_AI) · 18 小时前

SenseNova-U1 技术报告详尽披露了构建前沿原生多模态模型的方法，核心包括原生多模态统一建模、无损视觉接口、联合自回归与像素空间流匹配训练、以及原生混合专家骨干网络。报告提供了六阶段训练方案、强化学习后训练与蒸馏的完整实践指南。其开源版本 SenseNova-U1-A3B-MoT 基于混合专家架构，仅激活30亿参数，实现了高效快速的性能。相关资源已全面开放，涵盖技术报告、模型权重、代码和演

https://x.com/SenseTime_AI/status/2054446490420924558

4. Hy3预览版登陆GMI，开源最强模型领跑

X：腾讯混元 (@TencentHunyuan) · 20 小时前

Hy3 预览版现已登陆 @gmi_cloud。⌧

https://x.com/TencentHunyuan/status/2054403079433572357

5. 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) · 昨天 03:30

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

6. Claude Opus 4.7快速模式开放研究预览

X：Claude Devs (@ClaudeDevs) · 昨天 02:23

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

7. Perceptron Mk1视觉语言模型上线OpenRouter

X：OpenRouter (@OpenRouter) · 昨天 00:08

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

8. 材料科学AI多任务模型突破

X：Microsoft Research (@MSFTResearch) · 昨天 21:24

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

产品 产品发布/更新

1. Claude 工具 v2.1.141 版本更新

Claude Code：GitHub Releases（RSS） · 1 小时前

Claude 工具发布 v2.1.141 版本，带来多项功能新增与优化。主要更新包括：为钩子输出添加 `terminalSequence` 字段以支持无控制终端的桌面通知；新增 `CLAUDE_CODE_PLUGIN_PREFER_HTTPS` 环境变量，便于通过 HTTPS 克隆插件源码；引入 `ANTHROPIC_WORKSPACE_ID` 变量以在多工作区联盟中限定令牌范围。会话管理方面，`

https://github.com/anthropics/claude-code/releases/tag/v2.1.141

2. Claude代码周限额提升50%

X：Claude Devs (@ClaudeDevs) · 5 小时前

Claude代码周限额正在提升50%，即日起持续至7月13日。 现已面向所有Pro、Max、Team及按席位计费的企业用户生效。

https://x.com/ClaudeDevs/status/2054639777685934564

3. Krea 2 新增情绪板分享功能

X：Krea AI (@krea_ai) · 5 小时前

推出情绪板分享功能。 现在你可以与他人分享 Krea 2 情绪板。 下方有几个可供尝试的示例 ☒

https://x.com/krea_ai/status/2054630514439696401

4. Claude付费计划将提供月度编程使用额度

X：Claude Devs (@ClaudeDevs) · 7 小时前

自6月15日起，付费Claude计划可申领专用的月度编程使用额度。 该额度涵盖以下用途： - Claude Agent SDK - claude -p - Claude Code GitHub Actions - 基于Agent SDK构建的第三方应用

https://x.com/ClaudeDevs/status/2054610152817619388

5. Introducing Runway Agent

Runway：News（网页） · 7 小时前

Runway正式发布Runway Agent，这是一个能够通过单次对话将创意想法转化为完整、可发布视频的智能创作伙伴。用户只需用自然语言描述需求，Agent便能根据上下文和目标，自主完成概念提案、故事节奏设计、视觉方向规划，并最终生成包含多场景、旁白、对话和音乐的成片。它旨在为品牌团队、营销人员、创意机构和电影制作人快速生产各类视频内容，如品牌宣传、社交媒体素材和短片，将传统需要数天或数周的审核制

https://runwayml.com/news/introducing-runway-agent

6. Anthropic推出面向小型企业的Claude服务包

Anthropic：Newsroom（网页） · 7 小时前

Anthropic推出"Claude for Small Business"服务包，旨在帮助小型企业弥补在AI应用资源上与大型公司的差距。该产品包含一系列连接器和15个开箱即用的自动化工作流，能将Claude深度集成到QuickBooks、PayPal、HubSpot等企业日常工具中。其核心功能是自动化处理财务、运营、销售等领域的重复性任务，如规划薪资、月末结算、追踪发票和分析营销活动等。用户通过

https://www.anthropic.com/news/claude-for-small-business

7. Runway Agent

Runway：Changelog（网页） · 8 小时前

Runway Agent 是一个集成化创意工具平台，旨在为用户提供实现任何创意所需的全套资源与功能。该平台整合了视频编辑、图像生成、3D建模等多种人工智能驱动工具，允许用户在一个工作流中无缝完成从概念到成品的创作过程。其核心特点是降低了专业内容制作的技术门槛，通过自动化与智能辅助功能，让用户能够更自由地将想法转化为视觉作品。

https://app.runwayml.com/agent

8. Suno登陆车载系统，车内流媒体新体验

X：Suno (@suno) · 9 小时前

Suno 现在可在 Apple CarPlay 和 Android Auto 上使用☒在车里流媒体播放您最喜欢的创作。在早晨通勤时用这个播放列表试试看！ https：//suno.com/playlist/a255cf6d-bb99-4c1f-aedd-8d584579bddb

https://x.com/suno/status/2054574166104297905

9. 全球首个全AI运营的在线广播电台上线，24小时不间断播报AI动态

X：Kim (@kimmonismus) · 10 小时前

全球首个完全由AI运营的在线广播电台在X平台正式开播，专为创业者、开发者和建设者提供全天候AI领域资讯。该电台由五名具备独立编辑判断、记忆和个性的AI主播主持，不仅能实时播报几分钟内的突发新闻，还提供每30分钟一次的新闻综述、初创公司融资追踪、GitHub等平台的工具趋势分析，并整合社区讨论与行业真实观点。AI主播会主动收集信息模式、发现矛盾并形成论点进行实时辩论，而非单纯播报数据。节目辅以非干扰

https://x.com/kimmonismus/status/2054562237684051974

10. Browser Run：现基于 Cloudflare Containers 运行，速度更快、扩展性更强

Cloudflare Blog · 11 小时前

Browser Run 产品已基于 Cloudflare Containers 完成重构，实现了使用限制提升、性能加速、可靠性增强以及交付速度提高。此次重构使产品能够更高效地处理大规模并发任务，显著缩短了任务响应时间，并提升了服务稳定性。团队通过容器化技术优化了资源调度与隔离机制，从而为用户提供更快速、更可扩展的浏览器自动化服务。

https://blog.cloudflare.com/browser-run-containers

---

11. 为智能体配置开发环境

Cursor Blog · 12 小时前

Cursor发布新工具，用于配置云端智能体开发环境。核心更新包括：支持多仓库环境，使智能体可跨代码库协同工作；提供基于Dockerfile的代码化配置，支持构建密钥并优化缓存，命中缓存后构建速度提升70%；增强由智能体主导的环境设置流程，提供验证与故障回退机制。同时新增环境治理与安全功能，如版本历史、审计日志，以及可在环境级别独立管控的网络出口和密钥权限。这些改进旨在帮助团队在受控环境中更高效地运

https://cursor.com/blog/cloud-agent-development-environments

---

12. Miaoda应用与企业版上线，自生成代码占比90%

X：百度 Baidu (@Baidu_Inc) · 13 小时前

Miaoda应用和Miaoda企业版现已发布，让更多开发者和企业能够使用我们的编程助手！ 最有趣的细节是什么？Miaoda应用90%的代码由Miaoda自身生成。 编程助手正使按需定制软件具备商业可行性。截至目前，Miaoda生成的应用已服务超1000万用户，应用总价值达50亿元人民币。

https://x.com/Baidu_Inc/status/2054511974172557463

---

13. Codex应用内浏览器升级，提升多视口测试与标注效率

X：Tibo (@thsottiaux) · 21 小时前

Codex应用内浏览器功能迎来多项改进，支持在不同视口尺寸下测试应用，并能控制设备工具栏、在不同断点进行点击验证。长时测试中，Codex会在关键节点截图供用户核查。为加速测试，可隐藏应用内浏览器以禁用动画，使测试速度提升1-2倍。此外，标注功能现在发送更快且消耗更少tokens。

https://x.com/thsottiaux/status/2054398093886673260

---

14. AI未来原生智能体 Qwen 3.6 Plus限免

X：阿里云 / Alibaba Cloud (@alibaba_cloud) · 22 小时前

AI的未来是智能体原生的。 很高兴能与Hermes Agent及@NousResearch社区共同开启这段旅程。 Qwen 3.6 Plus现于Nous Portal限时免费--快来尝试吧。⚡

https://x.com/alibaba_cloud/status/2054372732234797342

---

15. Google 发布首款 AI 优先笔记本 Googlebook，集成 Gemini 智能

X：邵猛 (@shao__meng) · 23 小时前

Google 正式推出首款为 Gemini Intelligence 设计的笔记本 Googlebook，标志着从"云优先"的 Chromebook 时代进入"AI 优先"新阶段。其核心创新包括：Magic Pointer 将系统光标变为 AI 交互入口，可直接触发上下文建议与任务；Create Your Widget 允许通过自然语言生成聚合多源信息的动态桌面小组件；深度整合 Android 生

https://x.com/shao__meng/status/2054360963571446232

---

16. 无需注册付费，Telegram内一键启动AI智能体

X：Berry Xia (@berryxia) · 23 小时前

牛逼！Browser Use 今天把"AI agent 即用即走"做到了极致。 BuxFather：Telegram 里直接 Spin up agent，无需任何注册付费，24/7 自主运行 + 自改进，还带 stealth browser。 几下点击就有完整电脑 + 浏览器环境。 这波对重度 Telegram 用户来说真的爽了！ https：//x.com/browser_use/sta

https://x.com/berryxia/status/2054358231791923392

---

17. Claude Code新增/goal功能确保任务完成

X：Claude Devs (@ClaudeDevs) · 昨天 08:00

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

---

18. 谷歌AI重塑智能鼠标指针交互

X：Demis Hassabis (@demishassabis) · 昨天 06:22

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

---

19. 优化广告效果，定义更佳表现

X：Luma AI (@LumaLabsAI) · 昨天 04:38

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

---

20. Codex实现跨应用无感多任务处理

X：OpenAI Developers (@OpenAIDevs) · 昨天 04:31

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

---

21. 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 · 昨天 01:35

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

22. 谷歌发布全新安卓智能助理

X：Testing Catalog (@testingcatalog) · 昨天 01:34

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

23. Symphony为每个任务启动运行Codex智能体

X：OpenAI Developers (@OpenAIDevs) · 昨天 01:27

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

24. Grok接入Gmail，智能邮件助手革新收件箱管理

X：cb_doge (@cb_doge) · 昨天 23:43

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

25. 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 · 昨天 22:24

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

产业 产业与资本

1. Anthropic 首次在 B2B 采用率上超越 OpenAI，Ramp 支出数据显示

The Decoder：AI News（RSS） · 5 小时前

根据 Ramp AI 指数数据，Anthropic 在美国企业客户中的采用率达到 34.4%，首次超越 OpenAI 的 32.3%。其业务覆盖范围在一年内增长了四倍。但文章指出，三个因素可能使其领先优势迅速减弱。

https://the-decoder.com/anthropic-overtakes-openai-in-b2b-adoption-for-the-first-time-according-to-ramp-spending-data

2. Kling AI将亮相2026戛纳探讨AI电影制作

X：可灵 Kling AI (@Kling_ai) · 9 小时前

Kling AI将于2026年5月18日在戛纳电影节电影市场会议中举办专场活动，主题为"从创意可能到制作现实：Kling AI在电影工作流中的应用"。活动旨在探讨AI辅助电影制作的现状与未来演进。Kling AI通过支持《House of David》、《Born of the Tide》及《RAPHAEL》等项目，展示了AI在好莱坞级制作、全AI生成动画及剧情长片等实际影视生产中的多元化应用。

https://x.com/Kling_ai/status/2054577409064911330

3. 消息称 Anthropic 正就以超 9000 亿美元投前估值筹集至少 300 亿美元谈判

IT之家（RSS） · 20 小时前

据报道，AI公司Anthropic正就新一轮融资进行初步谈判，目标是以超过9000亿美元的投前估值筹集至少300亿美元资金。这有望成为该公司迄今最大规模的融资轮次，交易最快可能在本月底完成。此前，Anthropic在今年2月完成了300亿美元的G轮融资，投后估值为3800亿美元，并从谷歌和亚马逊获得了150亿美元的投资承诺。为应对高昂的算力成本，公司计划于今年晚些时候进行首次公开募股（IPO）。

https://www.ithome.com/0/949/715.htm

4. 山姆・奥特曼因涉嫌利用OpenAI谋私利遭正式调查

X：cb_doge (@cb_doge) · 昨天 02:20

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

5. 青少年按ChatGPT建议混用药物致死，父母起诉OpenAI

X：cb_doge (@cb_doge) · 昨天 01:44

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

6. 首届虚构节目提案大赛揭晓二十强

X：Runway (@runwayml) · 昨天 22:46

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

7. 人工智能首要应用应是改善人类健康

X：Demis Hassabis (@demishassabis) · 昨天 21:50

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

论文 论文与研究

1. ExploitGym：AI智能体能否将安全漏洞转化为真实攻击?

Berkeley RDI：Blog（AI 安全与评测） · 3 小时前

由伯克利RDI、马克斯·普朗克安全与隐私研究所、Anthropic、OpenAI及谷歌等机构研究人员组成的团队，发布了名为ExploitGym的新基准测试。该测试包含898个真实漏洞，要求AI智能体根据漏洞描述生成完整的漏洞利用程序。结果显示，前沿AI模型已能成功利用相当数量的漏洞，即使在启用ASLR等标准防御措施后，部分攻击仍能成功。这证明AI已具备自主将漏洞转化为实际攻击的能力，该技术具有双重

https://rdi.berkeley.edu/blog/exploitgym

观点 技巧与观点

1. 在 Windows 上构建安全有效的沙箱以启用 Codex

OpenAI：官网动态（RSS · 排除企业/客户案例） · 刚刚

OpenAI 为 Windows 平台上的 Codex 构建了一个安全沙箱环境。该沙箱通过严格控制文件访问权限和实施网络限制，确保了代码生成与执行过程的安全性。这一举措使得基于 Codex 的编码助手能够以高效且受控的方式运行，在提供强大编程辅助功能的同时，有效隔离了潜在风险，保障了用户系统的安全。

https://openai.com/index/building-codex-windows-sandbox

2. BestBlogs早报：AI智能体工程化实战与安全架构

X：洪明 (@hongming731) · 1 小时前

BestBlogs早报聚焦AI智能体的工程化落地。Anthropic官方指南详解Claude Computer Use最佳实践，包括解决点击偏移的根本原因、推荐分辨率策略及必须采用虚拟机隔离与人工确认门控的安全原则。OpenAI工程师分享了为Codex构建Windows安全沙箱的历程，其最终方案通过专属安全标识符和写受限令牌，实现了操作系统层面的强制文件系统隔离。早报同时指出，基准测试优异的RAG

https://x.com/hongming731/status/2054701978924859865

3. Claude 电脑与浏览器使用的最佳实践

Claude：Blog（网页） · 4 小时前

Claude 最新模型在电脑与浏览器使用能力上显著提升，支持构建复杂智能体系统。本文针对Claude 4.6系列和Opus 4.7提供实践指南，重点优化截图分辨率：Claude 4.6系列API限制最大长边1568像素、总像素115万；Opus 4.7提升至最大长边2576像素、总像素375万。发送前将截图缩放到限制内是提升点击准确性的最有效方法。推荐起始分辨率为1280x720，Opus 4.7

https://claude.com/blog/best-practices-for-computer-and-browser-use-with-claude

4. AI电影大师Gossip Goblin创作流程首度揭秘

X：可灵 Kling AI (@Kling_ai) · 5 小时前

又一篇来自@PJaccetturo的精彩解析！ 查看完整推文，了解@Gossip_Goblin这部主要使用Kling制作动画的杰作！

https://x.com/Kling_ai/status/2054633446002377084

5. Meta首席AI官宣WhatsApp和Meta AI推出Incognito Chat

X：阿易 AI Notes (@AYi_AInotes) · 6 小时前

Meta首席AI官宣布，Incognito Chat功能正式登陆WhatsApp和Meta AI。与ChatGPT等仅不保存历史记录的"临时聊天"不同，该功能的关键创新在于：对话推理完全在用户手机的硬件安全飞地内进行，Meta工程师无法获取明文，且不产生任何服务器日志，会话结束后数据永久消失。此举将WhatsApp成熟的端到端加密标准应用于AI对话，旨在彻底解决用户对隐私的顾虑，从而鼓励用户与AI

https://x.com/AYi_AInotes/status/2054616319127904403

6. 人形机器人已能自主完成8小时轮班

X：Kim (@kimmonismus) · 7 小时前

"如果AI抢了你的白领工作，那就转行做蓝领吧。" ☒↗ 与此同时，蓝领工作： 【引用 @adcock_brett】：观看一组人形机器人以人类绩效水平完成完整的8小时轮班。这是完全自主运行的Helix-02 https：//x.com/i/broadcasts/1dxYljYVREYJX

https://x.com/kimmonismus/status/2054609852094193850

7. 情绪板教程：10-20张参考图即可定调

X：Krea AI (@krea_ai) · 7 小时前

很棒的情绪板制作教程！ 【引用 @goo_vision】：使用Krea 2进行创作 ☒第一步：建立情绪板。 不必强求填满全部250个图片位。 即使只有10-20张优质参考图，也足以确立坚实的视觉方向并产出优秀成果。

https://x.com/krea_ai/status/2054608346778665557

8. 开源psql_bm25s，让PostgreSQL多智能体检索提速23倍

X：Emad Mostaque (@EMostaque) · 8 小时前

在构建多智能体生产级系统时，PostgreSQL虽可靠但检索速度不足。团队为此开发并开源了psql_bm25s，这是一个原生PostgreSQL访问方法，实现了精确的BM25检索。其在标准基准测试中比pg_search快约23倍，使得检索不再成为性能瓶颈和成本负担，智能体得以高效查询数据，为大规模自主智能体应用铺平道路。

https://x.com/EMostaque/status/2054587062033043799

9. 利用搜索垫图提升AI绘画准确性与质量

X：歸藏 (@op7418) · 15 小时前

当使用Codex等AI生成涉及生僻事实的配图时，可先让其搜索相关图片作为参考，再基于此生成新图。该方法能确保图像的真实性，同时生成符合比例要求的高清图片。例如，对于云南甲马符这类GPT可能不了解的主题，通过垫图后AI能准确绘制。

https://x.com/op7418/status/2054491392261632448

10. 在VS Code中集成多款AI模型开发

X：硅基流动 SiliconFlow (@SiliconFlowAI) · 15 小时前

通过@continuedev在VS Code中直接运行DeepSeek V4、GLM-5.1、Kimi K2.6等多款模型@SiliconFlowAI 支持标签自动补全、AI对话编辑和智能体功能 以下是3步设置指南 ⛒

https://x.com/SiliconFlowAI/status/2054484427192050041

11. AI技能更新地图组件，支持交互与标记

X：歸藏 (@op7418) · 18 小时前

Skills功能已更新，新增了带地图的版式和地图组件。用户可让各自的AI更新此技能。更新后的地图支持缩放、拖动等基本交互操作，并且AI能够在地图上进行任意标记。这增强了AI在空间信息处理和可视化方面的能力。

https://x.com/op7418/status/2054433146532479266

12. oMLX更新强化苹果端侧AI，本地能力直逼云端

X：Berry Xia (@berryxia) · 19 小时前

oMLX项目更新至0.3.9.dev2版本，集成了Gemma 4的MTP视觉路径、DFlash引擎和ParoQuant技术，显著提升了图文处理速度。新增一键启动copilot功能，可便捷接入Claude等工具，并通过oQ自动代理解决显存瓶颈。这些改进大幅增强了本地AI在速度、集成与易用性上的表现，正推动AI能力从云端向个人电脑回归。

https://x.com/berryxia/status/2054420138095694324

13. 智能体时代新指标：日活跃智能体数

X：百度 Baidu (@Baidu_Inc) · 21 小时前

Robin提出将日活跃智能体（DAA）作为智能体时代的定义性指标，相当于移动互联网时代的日活跃用户数。 虽然令牌消耗更多反映成本而非价值，但DAA让讨论回归产出本质。 正如Robin所指出的，衡量平台或生态系统健康度时，应更关注DAA指标--即积极工作并交付成果的智能体数量。

https://x.com/Baidu_Inc/status/2054400181903524133

14. BenchLoop：本地大模型一键基准测试与排行榜发布

X：Berry Xia (@berryxia) · 21 小时前

BenchLoop 提供了一套对本地大模型进行基准测试的标准化流程。用户只需拉取模型并运行该工具，即可实时获取模型在质量、速度和可靠性方面的综合评分。平台支持对比不同提示框架（如原生模式与 Hermes 模式）下的性能表现。测试完成后，结果可自动发布至公开排行榜，便于横向比较不同模型的优劣。

https://x.com/berryxia/status/2054390836721553487

15. The 6 Messages That Actually Matter

Tomer Tunguz 博客（VC 分析） · 昨天 08:00

知识工作者平均每天收到121封邮件，传统收件箱处理模式难以为继。未来邮件处理将转向高度个性化与自动化：用户能用自然语言定义处理规则，实现收据自动转发、销售线索自动录入CRM等流程。所有历史邮件将构成个人上下文层，为AI处理新邮件提供背景信息，敏感信息则由设备端模型进行私密处理。最终，收件箱本身将消失，真正重要的信息可能浓缩至仅6条。

https://www.tomtunguz.com/the-disappearance-of-email

16. 90%的人在白白浪费"Token"！

X：Berry Xia (@berryxia) · 昨天 07:13

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

17. AI取代人类？各方叙事背后的利益驱动

X：小北 (@frxiaobei) · 昨天 00:59

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

18. 开放模型生态的复合增长

Nathan Lambert：Interconnects（RSS） · 昨天 23:54

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

19. 在Claude Code中安装官方插件调用Codex

X：Vista (@vista8) · 昨天 23:15

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

20. 财务团队如何使用 Codex

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

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

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

GitHub Blog · 昨天 23:00

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

22. 导出消费数据赋能AI Agent个性化服务

X：Berry Xia (@berryxia) · 昨天 20:26

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