

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: 2s 精选条目: 44 条 焦点: 8 条 快讯: 0 条

Executive Summary

今日核心进展：蚂蚁百灵发布Ring-2.6-1T开源模型，专为智能体 workflow 设计，具备万亿级推理能力，现已上线OpenRouter平台。NVIDIA研究团队推出SANA-WM，26亿参数的开源世界模型可生成1分钟720p视频。OpenAI进行史上最大规模重组，总裁Brockman全面接管产品战略，整合ChatGPT、Codex和API三大产品线。X平台开源"For You"信息流推荐算法，采用基于Grok架构的Phoenix Transformer模型。商汤发布SenseNova-U1-8B-MoT-Infographic信息图表生成模型。

行业结构变化：智能体能力边界显著扩展，Ring-2.6-1T和剑桥大学Articraft系统推动复杂工作流自动化。成本结构出现分化，OpenSquilla项目通过智能路由将AI处理成本从6美元降至0.68美元，而百实例自动化项目显示令牌成本正变得相对次要。云厂商与平台竞争加剧，Notion推出开发者平台，MiniMax M2.7接入OrcaRouter。安全风险上升，英国多部门警告先进AI模型网络攻击能力超越专业人员，研究人员已利用Anthropic Mythos绕过苹果M5芯片安全机制。

后续观察：模型开放节奏需关注Ring-2.6-1T的商业化部署情况和SANA-WM在长视频生成领域的应用进展。API价格竞争将持续，OpenSquilla的成本优化模式可能推动行业标准化。算力供给方面，万亿级模型对基础设施需求增长。监管政策将重点关注英国金融部门的AI安全指南和苹果M5漏洞的行业影响。开源生态发展需观察vLLM对万亿级模型支持的成熟度，以及lark-cli等办公AI操作工具的生态建设。

重点 今日核心进展

★ 1. Ring-2.6-1T开源并上线OpenRouter，专为智能体 workflow 设计

X: 蚂蚁百灵 (@AntLingAGI) · 8 小时前 · 模型与工具能力

Ring-2.6-1T模型现已开源，并上线OpenRouter平台，5月底前享有75%折扣。该模型由AntLingAGI开发，是一个为现实世界智能体构建的万亿级推理模型，其核心设计目标不仅是回答问题，更是执行任务。它专注于规划步骤、使用工具、维持上下文和完成复杂工作流，具备强大的智能体执行能力和高推理模式，并采用了Async RL与IcePop训练方法。

能力进展 基础设施 新发布

<https://x.com/AntLingAGI/status/2055681163033997815>

★ 2. SANA-WM：一个用于生成1分钟720p视频的26亿级开源世界模型

Hacker News 热门 (buzzing.cc 中文翻译) · 8 小时前 · 模型与工具能力

NVIDIA研究团队发布了SANA-WM，这是一个参数规模达26亿的开源世界模型，专门用于生成长达1分钟、分辨率为720p的视频。该模型已在GitHub页面开源，旨在推动高质量长视频生成的研发。其在Hacker News社区获得了107点热度，显示出业界对该技术进展的关注。

能力进展 基础设施 新发布

<https://nvlabs.github.io/Sana/WM>

★ 3. SenseNova发布增强版信息图表生成模型SenseNova-U1-8B-MoT-Infographic

X: 商汤 SenseTime (@SenseTime_AI) · 昨天 23:54 · 模型与工具能力

SenseNova近日推出专为信息图表生成设计的增强模型SenseNova-U1-8B-MoT-Infographic。该模型能够根据指令生成多种视觉格式内容，包括海报、图表、食谱卡、明信片以及学术风格的arXiv页面等。在性能方面，模型在BizGenEval (困难版) 和IGenBench (Q-ACC) 两项基准测试上表现突出，相比其基础U1模型分别提升了6.8分和18.2分。目前，该模型已在Hug

能力进展 新发布

https://x.com/SenseTime_AI/status/2055315819165597702

★ 4. X平台开源"For You"信息流推荐算法

X: cb_doge (@cb_doge) · 昨天 21:34 · 应用与商业化

X平台在GitHub开源了其"For You"信息流的最新推荐算法。该算法融合来自已关注账号的内容和通过相似性搜索发现的网络外内容，并利用基于Grok架构的Phoenix Transformer模型进行统一评分。模型通过分析用户近期互动，同时预测点赞、回复、转发等多种互动概率，并综合成最终相关性分数。排序过程包括候选内容收集、上下文信息补充、AI模型评分、多样性调整及垃圾信息过滤等步骤。此次开源提

能力进展 基础设施 新发布

https://x.com/cb_doge/status/2055280665583898691

★ 5. 突发! OpenAI 大规模重组, 总裁 Brockman 夺权挂帅

IT之家 (RSS) · 17 小时前 · 产业与基础设施

OpenAI宣布进行史上最大规模重组, 将ChatGPT、Codex和API三大核心产品线合并为统一组织。联合创始人兼总裁Greg Brockman正式全面接管产品战略, 成为实际掌权者, 而ChatGPT原负责人Nick Turley被调离核心岗位。此次重组旨在整合资源, 聚焦"智能体时代", 并秘密开发集成多项功能的"超级应用"桌面端产品。与此同时, 竞争对手Anthropic估值飙升至900亿美元, 使

能力进展 监管/资本 新发布

<https://www.ithome.com/0/951/258.htm>

★ 6. MiniMax M2.7模型上线OrcaRouter平台

X: MiniMax (@MiniMax_AI) · 昨天 00:56 · 模型与工具能力

M2.7已在@OrcaRouter上线--可通过单一OpenAI兼容API访问。今天就来试试吧! ☒

能力进展 新发布

https://x.com/MiniMax_AI/status/2055331545402245246

★ 7. 智能体驱动系统Articraft实现3D资产生成自动化, 开源万件数据集降低门槛

X: Berry Xia (@berryxia) · 23 小时前 · 研究与开源进展

剑桥大学团队推出Articraft, 这是一个由智能体驱动的编码系统, 能全自动生成带关节、可运动的交互式3D资产。系统通过AI智能体自主编写代码、执行并接收物理反馈迭代优化, 将以往需数日甚至数周的创作过程极大简化。团队同时开源了Articraft-10K数据集, 包含超1万个物体, 覆盖250个类别, 所有资产均仿真就绪。此举显著降低了机器人训练和物理AI领域对高质量3D资产的数据门槛, 为解决"数据饥渴"

能力进展 基础设施 新发布

<https://x.com/berryxia/status/2055444087684407505>

★ 8. 研究人员利用Anthropic Mythos工具构建macOS内核漏洞, 绕过苹果M5芯片内存完整性执行安全系统

X: Kim (@kimmonismus) · 15 小时前 · 产业与基础设施

三名研究人员借助Anthropic的Mythos工具, 成功开发出一个可绕过苹果M5芯片内存完整性执行(MIE)安全机制的macOS内核漏洞利用程序。MIE是苹果耗时五年、投入巨资为M5和A19芯片打造的旗舰安全功能, 旨在彻底消除内存损坏漏洞。研究团队于4月25日发现漏洞, 5月1日即完成开发, 并亲自前往苹果园区提交报告。该攻击采用纯数据攻击方式, 无需操纵指针, 仅通过非特权用户的标准系统调用即可获得根

基础设施 监管/资本 新发布

<https://x.com/kimmonismus/status/2055571960260645125>

能力 模型与工具能力

1. Ring-2.6-1T开源并上线OpenRouter, 专为智能体 workflow 设计

X: 蚂蚁百灵 (@AntLingAGI) · 8 小时前

Ring-2.6-1T模型现已开源, 并上线OpenRouter平台, 5月底前享有75%折扣。该模型由AntLingAGI开发, 是一个为现实世界智能体构建的万亿级推理模型, 其核心设计目标不仅是回答问题, 更是执行任务。它专注于规划步骤、使用工具、维持上下文和完成复杂 workflow, 具备强大的智能体执行能力和高推理模式, 并采用了Async RL与IcePop训练方法。

能力进展 基础设施 新发布

<https://x.com/AntLingAGI/status/2055681163033997815>

2. SANA-WM: 一个用于生成1分钟720p视频的26亿级开源世界模型

Hacker News 热门 (buzzing.cc 中文翻译) · 8 小时前

NVIDIA研究团队发布了SANA-WM, 这是一个参数规模达26亿的开源世界模型, 专门用于生成长达1分钟、分辨率为720p的视频。该模型已在GitHub页面开源, 旨在推动高质量长视频生成的研发。其在Hacker News社区获得了107点热度, 显示出业界对该技术进展的关注。

能力进展 基础设施 新发布

<https://nvlabs.github.io/Sana/WM>

3. SenseNova发布增强版信息图生成模型SenseNova-U1-8B-MoT-Infographic

X: 商汤 SenseTime (@SenseTime_AI) · 昨天 23:54

SenseNova近日推出专为信息图生成设计的增强模型SenseNova-U1-8B-MoT-Infographic。该模型能够根据指令生成多种视觉格式内容, 包括海报、图表、食谱卡、明信片以及学术风格的arXiv页面等。在性能方面, 模型在BizGenEval (困难版) 和IGenBench (Q-ACC) 两项基准测试上表现突出, 相比其基础U1模型分别提升了6.8分和18.2分。目前, 该模型已在Hug

能力进展 新发布

https://x.com/SenseTime_AI/status/2055315819165597702

4. MiniMax M2.7模型上线OrcaRouter平台

X: MiniMax (@MiniMax_AI) · 昨天 00:56

M2.7已在@OrcaRouter上线--可通过单一OpenAI兼容API访问。今天就来试试吧! ☒

能力进展 新发布

https://x.com/MiniMax_AI/status/2055331545402245246

5. 社区协作再创佳绩，vLLM支持万亿级模型

X: 蚂蚁百灵 (@AntLingAGI) · 7 小时前

又一次Day0协作，又一次社区胜利。感谢@vllm_project团队始终可靠的支持~ ☺

能力进展

<https://x.com/AntLingAGI/status/2055689329339760829>

产业 产业与基础设施

1. 突发！OpenAI 大规模重组，总裁 Brockman 夺权挂帅

IT之家 (RSS) · 17 小时前

OpenAI宣布进行史上最大规模重组，将ChatGPT、Codex和API三大核心产品线合并为统一组织。联合创始人兼总裁Greg Brockman正式全面接管产品战略，成为实际掌权者，而ChatGPT原负责人Nick Turley被调离核心岗位。此次重组旨在整合资源，聚焦"智能体时代"，并秘密开发集成多项功能的"超级应用"桌面端产品。与此同时，竞争对手Anthropic估值飙升至9000亿美元，使

能力进展 监管/资本 新发布

<https://www.ithome.com/0/951/258.htm>

2. 研究人员利用Anthropic Mythos工具构建macOS内核漏洞，绕过苹果M5芯片内存完整性执行安全系统

X: Kim (@kimmonismus) · 15 小时前

三名研究人员借助Anthropic的Mythos工具，成功开发出一个可绕过苹果M5芯片内存完整性执行（MIE）安全机制的macOS内核漏洞利用程序。MIE是苹果耗时五年、投入巨资为M5和A19芯片打造的旗舰安全功能，旨在彻底消除内存损坏漏洞。研究团队于4月25日发现漏洞，5月1日即完成开发，并亲自前往苹果园区提交报告。该攻击采用纯数据攻击方式，无需操纵指针，仅通过非特权用户的标准系统调用即可获取根

基础设施 监管/资本 新发布

<https://x.com/kimmonismus/status/2055571960260645125>

3. OpenAI 与马耳他合作，向所有公民提供 ChatGPT Plus

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

OpenAI 与马耳他政府达成合作，将向该国所有公民免费提供 ChatGPT Plus 订阅服务。此次合作旨在扩大人工智能技术的普及范围，并包含面向公民的 AI 技能实践培训，以帮助民众负责任地使用 AI。该计划是国家级大规模推广 ChatGPT Plus 的首例，预计将提升马耳他全民的 AI 素养与应用能力。

能力进展 新发布

<https://openai.com/index/malta-chatgpt-plus-partnership>

4. 微软研究院发布AI新工具与模型

X: Microsoft Research (@MSFTResearch) · 昨天 23:06

微软研究院的新工具、模型、代码库和论文现已发布。使用AI和智能体？值得关注：
· MSR AI Frontiers的MagenticLite · 智能化GitHub工作流 · 验证优先智能体 · 意义匹配微调 · AI转型与经济发展

能力进展 新发布

<https://x.com/MSFTResearch/status/2055303919497089337>

5. 英国多部门联合警告：当前最先进 AI 模型网络攻击能力已远超专业人员，企业应做好防范措施

IT之家 (RSS) · 昨天 23:06

英国财政部、英格兰银行及金融行为监管局联合警告，当前最先进的AI模型已具备远超普通专业人员的网络攻击能力，其攻击速度更快、范围更广、成本更低。若被恶意利用，将严重威胁企业运营安全、客户数据、金融市场稳定性乃至整个金融体系。英格兰银行行长安德鲁·贝利此前已点名Anthropic的Mythos产品，网络安全专家亦警告此类AI可能强化复杂攻击，对银行业及金融科技体系构成新挑战。

能力进展 监管/资本

<https://www.ithome.com/0/951/144.htm>

6. 杭州基地启用，机器人有了国家级职业技能训练场

IT之家 (RSS) · 10 小时前

国家人工智能应用中试基地（具身智能）5月16日在浙江杭州挂牌启用，为机器人提供国家级职业技能训练场。该基地是集场景体验、技术展示、研发合作、产业赋能于一体的综合性平台，旨在推动具身智能技术从实验室迈向现实应用。杭州市于5月1日施行首部具身智能机器人地方性法规，支持核心研发、平台建设和场景开放。目前杭州已集聚机器人产业相关企业700余家，2025年具身智能产业集群产值达1068亿元。

基础设施 新发布

<https://www.ithome.com/0/951/390.htm>

7. Figure人形机器人连续自主运行四天，迈向实用化关键一步

X: Berry Xia (@berryxia) · 1 小时前

Figure公司的F.03人形机器人已进入第四天不间断自主运行测试，在真实仓库环境中24/7连续工作直至出现故障。测试核心在于评估机器人执行抓取、搬运、分拣等任务的长期耐力，并收集故障数据、维护需求及安全恢复机制等信息。这标志着人形机器人从展示单次动作的"能动"阶段，进入了考验持续工作能力的"能干"实用化关键阶段。

监管/资本

<https://x.com/berryxia/status/2055786570570137989>

8. 戛纳倒计时 | Kling AI大会主讲阵容揭晓!

X: [可灵 Kling AI \(@Kling_ai\)](#) · 昨天 23:00

2026年戛纳电影节Kling AI大会将汇聚三位世界级电影制作人，展示他们运用Kling AI技术创作的标杆作品。中国动画导演魏立将分享AI生成动画《Born of the Tide》；Jon Erwin将解析为Amazon Prime制作的剧集《House of David》；Eekyun Yang则将介绍院线电影《RAPHAEL》的创作过程。三位讲者将深入探讨Kling AI在电影制作中的幕

https://x.com/Kling_ai/status/2055302191129473502

9. 美国开始出现人工智能相关岗位的大规模裁员

Hacker News 热门 ([buzzing.cc 中文翻译](#)) · 4 小时前

美国人工智能相关岗位正出现大规模裁员。根据彭博社报道，受AI影响的职位开始经历严重的就业岗位流失。这一趋势表明AI技术对劳动力市场的冲击已从理论讨论进入现实阶段，具体裁员数字和涉及的行业领域在进一步显现中。

<https://www.bloomberg.com/news/articles/2026-05-15/us-is-starting-to-see-heavy-job-losses-in-roles-exposed-to-ai>

应用 应用与商业化

1. X平台开源"For You"信息流推荐算法

X: [cb_doge \(@cb_doge\)](#) · 昨天 21:34

X平台在GitHub开源了其"For You"信息流的最新推荐算法。该算法融合来自已关注账号的内容和通过相似性搜索发现的网络外内容，并利用基于Grok架构的Phoenix Transformer模型进行统一评分。模型通过分析用户近期互动，同时预测点赞、回复、转发等多种互动概率，并综合成最终相关性分数。排序过程包括候选内容收集、上下文信息补充、AI模型评分、多样性调整及垃圾信息过滤等步骤。此次开源提

能力进展 基础设施 新发布

https://x.com/cb_doge/status/2055280665583898691

2. Claude Code v2.1.143 版本更新：插件管理与用户体验增强

Claude Code: [GitHub Releases \(RSS\)](#) · 昨天 06:28

Claude Code 发布 v2.1.143 版本，重点增强了插件管理功能，包括强制执行插件依赖关系，并新增了插件市场的预估上下文成本显示。为方便直接编辑工作副本，增加了`worktree.bgIsolation: "none"`设置。多项体验得到改进：后台会话唤醒后保留模型与努力级别设置；Windows PowerShell 工具默认绕过执行策略；`claude agents` 命令新增多

能力进展 新发布

<https://github.com/anthropics/claude-code/releases/tag/v2.1.143>

3. Notion推出开发者平台及CLI工具

X: [Oran Ge \(@oran_ge\)](#) · 17 小时前

Notion正式推出开发者平台，核心组件包括Notion CLI终端工具、Workers计算服务、数据库同步功能以及多种Agent工具与API。平台旨在让开发者更便捷地在Notion基础设施上运行代码、连接外部数据源并构建自动化工作流。未来，即使非开发者也能通过AI Agent在Notion上构建应用。

能力进展 新发布

https://x.com/oran_ge/status/2055539910212460878

4. 语义代码审查工具clawpatch发布

X: [Peter Steinberger \(@steipete\)](#) · 昨天 03:08

☑️clawpatch 0.1.0 已上线：Clawpatch 将代码库映射成语义功能切片，审查其中的错误和质量问题，并记录经过验证的明确修复尝试。您会发现它能发现的问题之多令人惊讶。npm install -g clawpatch <https://clawpatch.ai>

能力进展 新发布

<https://x.com/steipete/status/2055364630709448970>

5. 按需定制快捷键，优化工作效率

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 5 小时前

处理大家的反馈让我们感到太有趣了。（请继续反馈。）键盘快捷键现已支持自定义。围绕你的实际工作方式设置Codex，然后通过设置调整快捷键，无需再适应我们的默认配置。

能力进展 新发布

<https://x.com/OpenAIDevs/status/2055717793841221796>

6. Runway Agent一键生成完整广告

X: [Runway \(@runwayml\)](#) · 昨天 03:08

Runway Agent 让你仅需一次会话，就能从产品照片和想法转变为完全制作完成的广告。立即通过下方链接开始体验。

能力进展

<https://x.com/runwayml/status/2055364605979869229>

7. OpenClaw新版本速度提升约3.5倍

X: [OpenClaw \(@openclaw\)](#) · 昨天 21:07

最新版OpenClaw速度提升约3.5倍 ☑️我们每6小时通过真实消息渠道（此处使用Telegram，采用全新的机器人间通信）对所有已发布的npm版本进行端到端RTT测试。不再有静默回归。所有运行器均在@useblacksmith CI上运行。在您察觉之前捕捉减速问题。

新发布

<https://x.com/openclaw/status/2055273947537490422>

8. Codex新增键盘快捷键自定义功能

X: [Tibo \(@thsottiaux\)](#) · 4 小时前

不知为何我们选择周六发布更新，但Codex确实迎来了一系列优化。这些改进让使用体验愉悦许多，实在不该拖到周二再公布。键盘快捷键现已支持自定义。根据实际工作方式配置Codex，通过设置调整快捷键，不必再迁就默认设定。

新发布

<https://x.com/thsottiaux/status/2055735802894954612>

9. OpenRouter BYOK三大升级 支持多密钥轮换

X: [OpenRouter \(@OpenRouter\)](#) · 昨天 23:12

OpenRouter BYOK 三项重大升级 🗝️/ 可在同一工作空间为同一服务商添加多个密钥，并设置调用顺序。便于在速率限制间轮换、区分开发与生产凭证，或在团队账户间分配使用量！

新发布

<https://x.com/OpenRouter/status/2055305259287801865>

10. Krea 2正式上线面向专业用户

X: [Krea AI \(@krea_ai\)](#) · 昨天 03:37

从今天起，Krea 2 正式为 Pro 用户上线。

https://x.com/krea_ai/status/2055371934452642107

研究 研究与开源进展

1. 智能体驱动系统Articraft实现3D资产生成自动化，开源万件数据集降低门槛

X: [Berry Xia \(@berryxia\)](#) · 23 小时前

剑桥大学团队推出Articraft，这是一个由智能体驱动的编码系统，能全自动生成带关节、可运动的交互式3D资产。系统通过AI智能体自主编写代码、执行并接受物理反馈迭代优化，将以往需数日甚至数周的创作过程极大简化。团队同时开源了Articraft-10K数据集，包含超1万个物体，覆盖250个类别，所有资产均仿真就绪。此举显著降低了机器人训练和物理AI领域对高质量3D资产的数据门槛，为解决"数据饥渴"

能力进展 基础设施 新发布

<https://x.com/berryxia/status/2055444087684407505>

2. Δ-Mem：适用于大型语言模型的高效在线内存

Hacker News 热门 (buzzing.cc 中文翻译) · 10 小时前

研究人员提出了Δ-Mem，一种专为大型语言模型设计的高效在线内存系统。该系统通过仅存储和更新模型激活的增量变化，而非完整的激活状态，显著降低了内存占用。实验表明，Δ-Mem能将内存使用量减少高达70%，同时保持模型输出的质量基本无损。这一方法有助于在资源受限的环境中部署和运行大规模语言模型，提升其在线推理和持续学习场景下的可行性。

能力进展 新发布

<https://arxiv.org/abs/2605.12357>

格局 观点、资本与监管

1. 飞书开源CLI工具45天获万星，AI操作可见可控引关注

X: [阿易 AI Notes \(@AYI_AInotes\)](#) · 昨天 20:24

飞书开源命令行工具lark-cli在45天内获得超过1万GitHub star，成为国内首个破万星的办公套件开源项目。该工具允许AI通过命令行直接操作飞书，执行建群、建文档等任务，且每一步操作都可预览、可审查，与云端不可见的MCP模式形成对比。其主干代码已合并10位外部开发者的贡献，而同类产品钉钉和企业微信则为零。这种可见、可控的特性被视为开发者放心将任务交给AI Agent的前提，预示着一句话指

能力进展 基础设施 新发布

https://x.com/AYI_AInotes/status/2055263164145504501

2. 杨立昆访谈：剖析LLM局限，畅谈AI未来与创业新途

X: [Yann LeCun \(@ylecun\)](#) · 昨天 01:56

在Unsupervised Learning播客中，AI先驱杨立昆阐述了其对于大型语言模型局限性的反主流观点，并探讨了机器人技术的发展路径。他解释了离开Meta的原因，以及与Geoff Hinton、Yoshua Bengio在图灵奖观点上的重大分歧。访谈还涵盖了他对2027年的预测、新公司AMI对世界模型的押注，并将OpenAI和Anthropic比作Sun Microsystems。此外，他建议

能力进展 监管/资本 新发布

<https://x.com/ylecun/status/2055346714039464373>

3. 无视令牌成本，用百个AI实例自动化驱动开源项目

X: [Peter Steinberger \(@steipete\)](#) · 昨天 05:48

作者在OpenClaw项目中大规模运用AI，探索在"令牌成本无关紧要"的未来如何构建软件。团队持续运行约100个Codex实例，自动化处理多项核心工作：审查代码与安全问题、去重归类议题、自动重现复杂测试环境并录制验证视频、从会议讨论中主动创建任务、过滤垃圾评论以及监控性能回归。通过clawpatch.ai等工具将项目拆分为功能单元进行审查，并整合Vercel DeepSec等进行安全分析。整套自动

能力进展 监管/资本 新发布

<https://x.com/steipete/status/2055405041843052792>

4. 开源项目OpenSquilla：智能路由降低AI成本近十倍，兼容多平台与模型

X: 小互 (@xiaohu) · 昨天 22:50

开源项目OpenSquilla通过智能路由架构，在保持与OpenClaw相近性能（0.9251分）的同时，将复杂AI任务处理成本从6美元大幅降至0.68美元。其核心是将简单任务分配给廉价模型，仅复杂任务调用Opus等顶级模型。系统具备四层记忆结构模拟人脑认知，支持按需加载16种工具，并设有三档代码沙箱安全防护。项目提供统一后端，支持网页、命令行及Slack、飞书等平台接入，兼容OpenAI、Cl

能力进展 监管/资本 新发布

<https://x.com/xiaohu/status/2055299820215300380>

5. Anthropic 《Founder's Playbook》：AI提高创业失败率的警示

X: Berry Xia (@berryxia) · 11 小时前

Anthropic发布内部手册《Founder's Playbook》，核心观点是AI（如Claude Code）将提高创业失败率，而非降低。手册指出，AI能快速生成原型，易使创始人混淆“能运行”与“有市场需求”，并通过确认偏误和研究引擎强化错误想法。它将创业分为Idea、MVP、Launch、Scale四阶段，剖析AI放大的风险：如原型不等于验证、Agentic技术债、创始人成为决策瓶颈等。最终

能力进展 新发布

<https://x.com/berryxia/status/2055635826462130227>

6. 工具使用代理认知与行动脱节机制研究

X: Elvis Saravia (@omarsar0, DAIR.AI) · 3 小时前

该可解释性论文聚焦工具使用代理，通过探测隐藏状态发现模型常能识别应调用工具，但实际调用失败，不匹配率达26%-54%。问题完全集中于认知到行动的过渡阶段，而非认知本身。内部探测方向可解码，但后期层的最后令牌机制使信号旋转，几乎与产生的行动正交。研究旨在预测干预措施效果，指出常见归因如提示或训练不足可能忽略后期层几何结构，这为工具使用提示A/B测试中的性能上限提供了合理解释。

能力进展 基础设施

<https://x.com/omarsar0/status/2055750162526715926>

7. Eric Jang分享从零构建AlphaGo的经历与见解

X: Marc Andreessen (@pmarca, a16z) · 5 小时前

Eric Jang在过去几个月从零开始实现AlphaGo，这是2016年激发他进入深度学习的AI突破。他最初理解AlphaGo为“通过自我对弈训练的搜索增强深度神经网络”，但通过亲手构建获得了更深层次的理解。他指出，前沿深度学习研究虽昂贵，但特定能力的实现成本下降迅速；在2026年，训练强大围棋AI仅需几千美元租赁算力，无需DeepMind级别的资源。他荣幸受邀在播客中分享这一过程，并提供了详细教

能力进展 基础设施

<https://x.com/pmarca/status/2055723987850690927>

8. Forward Deployed Engineer：AI时代的新宠岗位，到底干什么？

X: 宝玉 (@dotey) · 昨天 23:22

Forward Deployed Engineer（FDE）是AI落地应用的关键角色，负责在客户现场部署和集成AI系统。近期，OpenAI成立独立部署公司，Anthropic与华尔街巨头合作，Google也简化流程大力招聘，显示行业重心正从模型竞赛转向商业落地。FDE工作介于工程师与顾问之间，需写代码、调试并深度理解业务。此岗位为毕业生提供了接触企业级项目的机会，也帮助资深工程师贴近业务需求，成为

能力进展 新发布

<https://x.com/dotey/status/2055307775417139447>

9. 最新开源成果（#21）：开源模型盛宴！Gemma 4、DeepSeek V4、Kimi K2.6、MiMo 2.5、GLM-5.1 等。基于CAISI的V4评估。

Nathan Lambert: Interconnects (RSS) · 7 小时前

本月开源AI模型领域迎来密集发布，包括Gemma 4、DeepSeek V4、Kimi K2.6、MiMo 2.5和GLM-5.1等一系列重要新版本。这些模型均在CAISI的V4评估框架下进行了性能评测。多个团队几乎同期推出旗舰级更新，反映出开源社区当前高速迭代的竞争态势。

能力进展 新发布

<https://www.interconnects.ai/p/latest-open-artifacts-21-open-model>

10. Show HN：烧吧，宝贝，烧吧（那些代币）

Hacker News 热门 (buzzing.cc 中文翻译) · 9 小时前

开发者发布了一个名为“烧吧，宝贝，烧吧”的开源项目，旨在通过销毁代币来应对加密货币领域的通胀问题。该项目提供了一个工具，允许用户主动销毁自己持有的代币，从而减少总供应量。此举可能提升剩余代币的稀缺性与潜在价值。项目已在GitHub上开源，并在Hacker News上获得了100点的社区热度。

能力进展 新发布

<https://github.com/dtnewman/burn-baby-burn>

11. 藏师傅PPT Skills更新，AI自动美化截图

X: 锦藏 (@op7418) · 21 小时前

藏师傅PPT Skills产品更新，新增截图美化功能。该功能无需消耗GPT-Image 2.0资源，即可根据用户截图的大小、长宽比以及当前PPT的模板类型和颜色主题，自动匹配合适的背景图，实现类似CleanShot X等专业软件的截图美化效果。AI还能处理过长等有问题的图片，例如将其自动裁剪并排成两列展示。

能力进展 新发布

<https://x.com/op7418/status/2055477163579891897>

12. MagicPath与Codex深度整合，设计开发流程合二为一

X: [Berry Xia \(@berryxia\)](#) · 2 小时前

MagicPath AI CEO @skirano 演示了其产品与Codex的深度整合。用户现可将MagicPath作为原生画布直接在Codex中运行，通过拖拽设计UI，Codex能实时感知项目并自动生成、编辑代码，实现设计与开发的无缝衔接，无需在Figma和IDE间切换。安装简便，只需在Codex中粘贴一条命令即可完成配置。此举将"设计-开发-迭代"流程整合于单一窗口，显著缩短了从创意到可运行应

能力进展

<https://x.com/berryxia/status/2055766210512560416>

13. Codex实现多设备远程控制与上下文共享

X: [锦藏 \(@op7418\)](#) · 16 小时前

Codex除了通过ChatGPT连接和远程控制外，还能控制另一台电脑，实现在ChatGPT上直接管理多台设备而无需切换，只需切换项目即可访问不同设备的上下文和文件。引用推文补充，此功能还支持通过远程SSH设置其他VMs等，增强了多设备协作的灵活性和实用性。

能力进展

<https://x.com/op7418/status/2055561525633642762>

14. 英伟达CEO称技工前景优于计算机专业毕业生

X: [Kim \(@kimmonismus\)](#) · 昨天 05:41

英伟达CEO Jensen Huang在卡内基梅隆大学告诉2026届计算机专业毕业生，电工、水管工等技工比他们更有前景。数据支持这一观点：Randstad分析显示技工需求增长是白领职位的三倍，机器人技术人员职位增长107%，而斯坦福研究发现AI相关职位早期就业下降16%。顶级电工资年薪超10.6万美元，无需学位负债。科技公司今年投资7000亿美元建设数据中心，全球到2030年预计达7万亿美元，这些基础

基础设施

<https://x.com/kimmonismus/status/2055403142913884288>

15. AI助手Sai化身虚拟同事，自主完成深度研究

X: [Kim \(@kimmonismus\)](#) · 昨天 02:14

AI工具Sai被设计为拥有独立桌面的"虚拟同事"，而非仅提供聊天窗口。用户可向其下达"深度研究某主题"等指令后离线，Sai会在安全的工作空间内自主操作：打开标签页、点击应用、阅读并交叉引用资料、做笔记，其工作方式类似人类研究员。任务完成后，它会整理好引用来源，并在执行任何潜在风险操作前请求用户批准。这展示了"始终在线"的计算机使用代理的实际形态。

监管/资本

<https://x.com/kimmonismus/status/2055351198467911769>

16. 美国人工智能政策一团糟，以下是应对之策

Gary Marcus: [The Road to AI We Can Trust \(RSS\)](#) · 昨天 22:01

美国各州和联邦层面已提出约1200项人工智能相关法案，但至今未能形成统一的国家政策框架。这种分散且矛盾的立法状态，可能导致监管冲突、创新受阻，并削弱美国在全球人工智能治理中的竞争力。当前亟需建立跨层级的协调机制，将零散的提案整合为连贯的战略，以平衡技术发展与风险管控。

监管/资本

<https://garymarcus.substack.com/p/us-ai-policy-is-a-clumsy-mess-heres>

17. RLVR 可能在科学领域格外糟糕

Dwarkesh Patel: [Podcast & Blog \(RSS\)](#) · 5 小时前

RLVR（强化学习与验证）在科学理论验证中可能表现出不成比例的缺陷。科学理论的验证循环周期长达数十年甚至数个世纪，且当前被视为更优的理论实际上常会做出更差的预测。这一矛盾揭示了基于短期反馈的强化学习范式与科学探索长期性、复杂性之间的根本冲突，凸显了现有AI方法在应对科学发现这类超长反馈周期任务时存在的结构性局限。

<https://www.dwarkesh.com/p/rlvr-might-be-disproportionately>

18. Eric Jang - 从零开始构建 AlphaGo

Dwarkesh Patel: [Podcast & Blog \(RSS\)](#) · 昨天 00:04

文章以AlphaGo为例，阐述了智能的基本构成要素。AlphaGo至今仍是最清晰、最完整的范例，它融合了三大核心基础：搜索技术、从经验中学习以及自我对弈。这三大要素共同构成了其实现超越人类棋艺的关键路径。

<https://www.dwarkesh.com/p/eric-jang>