

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: 57 条 焦点: 8 条 快讯: 0 条

Executive Summary

今日核心进展方面，模型能力持续增强：SenseNova发布U1-8B-MoT-Infographic信息图表生成模型，支持海报、图表等多种视觉格式；IBM推出Granite Embedding Multilingual R2开源多语言嵌入模型，具备32K上下文长度；MiniMax的M2.7模型已通过OrcaRouter平台提供OpenAI兼容API访问。商业化应用显著扩展，OpenAI为ChatGPT Pro用户推出美国地区个人理财功能，百度发布面向大规模智能体应用的全栈AI云，X平台开源其"For You"信息流推荐算法。

行业结构变化呈现多元化趋势：微软对OpenAI累计投入超1000亿美元的投资规模显示大型科技公司对AI基础设施的深度布局，Anthropic与盖茨基金会2亿美元合作表明AI技术向全球健康教育领域的渗透。Prime Intellect在nanoGPT优化赛道上实现AI智能体自主研究突破，消耗1.4万H200小时完成近万次实验，标志着自动化AI研究能力的重要进展。英国多部门联合警告先进AI模型的网络安全风险，反映监管层面对AI安全性的重视程度提升。

后续观察需关注：各大模型厂商的API开放节奏与定价策略、Recraft AI V4.1等新一代图像生成模型的市场接受度、企业级AI安全防护措施的标准化进程、开源AI工具链的生态完善程度，以及智能体在金融、医疗等垂直领域的商业化落地速度。

重点 今日核心进展

★ 1. 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) · 8 小时前 · 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

★ 2. inclusionAI/ARGenSeg-8B

蚂蚁 inclusionAI: HuggingFace 新模型 · 21 小时前 · 模型与工具能力

包容性AI团队发布ARGenSeg-8B模型，致力于通过开源和开放科学推动人工智能的进步与普及。该举措强调技术民主化，使更广泛的社区能够参与AI研发与应用。开源策略将促进协作创新，加速AI工具在多元场景中的落地，降低技术门槛，推动产业生态的开放发展。

能力进展 新发布

<https://huggingface.co/inclusionAI/ARGenSeg-8B>

★ 3. Granite Embedding Multilingual R2: 开源多语言嵌入模型，具备32K上下文与领先检索性能

Hugging Face: Blog (RSS) · 昨天 02:55 · 模型与工具能力

IBM Granite团队在Hugging Face发布了Granite Embedding Multilingual R2多语言文本嵌入模型。该模型采用Apache 2.0开源协议，上下文长度扩展至32K令牌，参数量低于1亿。在MTEB基准的检索评估中，它取得了同规模模型的最佳性能，尤其擅长多语言混合检索，为开发者提供了高效、轻量且可商用的嵌入解决方案。

能力进展 新发布

<https://huggingface.co/blog/ibm-granite/granite-embedding-multilingual-r2>

★ 4. X平台开源"For You"信息流推荐算法

X: cb_doge (@cb_doge) · 10 小时前 · 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. 为 OpenAI 累计投入超 1000 亿美元，纳德拉称微软当年投资时"没人愿意下注"

IT之家 (RSS) · 21 小时前 · 产业与基础设施

在"马斯克诉奥特曼"庭审中，微软企业发展负责人确认，微软对OpenAI的累计投入已超过1000亿美元，其中包括130亿美元原始投资及大量Azure基础设施成本。此次合作已为微软带来约300亿美元营收。CEO纳德拉表示，微软是在"没人愿意下注"时承担了风险。双方已续签非独家协议，微软不再支付收入分成，并将OpenAI的分成上限设为到2030年累计380亿美元，此举较原协议节省约970亿美元。此外，

能力进展 监管/资本 新发布

<https://www.ithome.com/0/950/728.htm>

★ 6. MiniMax M2.7模型上线OrcaRouter平台

X: [MiniMax \(@MiniMax_AI\)](#) · 7 小时前 · 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. 百度发布面向大规模智能体应用的全栈AI云

X: [百度 Baidu \(@Baidu_Inc\)](#) · 昨天 22:10 · 应用与商业化

随着智能体应用迈向更大规模部署,其背后的云技术栈也需同步扩展。在百度Create大会上,百度集团执行副总裁、百度智能云事业群总裁沈抖宣布推出专为大规模智能体应用打造的全新全栈AI云,其升级涵盖智能体基础设施与AI基础设施。基于我们自研的昆仑芯AI芯片构建的专用集群,已支持ERNIE 5.1系列中一个关键模型的训练。

能力进展 基础设施 新发布

https://x.com/Baidu_Inc/status/2054927298403688927

★ 8. OpenCode与Qwen 3.6 Plus再度免费开放

X: [opencode \(@opencode\)](#) · 昨天 07:32 · 应用与商业化

OpenCode x Qwen 3.6 Plus - 再次免费 上次各位把我们的容量当成了自助餐。我们找到了更多GPU。第二轮。

能力进展 基础设施 新发布

<https://x.com/opencode/status/2055068702538612784>

能力 模型与工具能力

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

X: [商汤 SenseTime \(@SenseTime_AI\)](#) · 8 小时前

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

2. inclusionAI/ARGenSeg-8B

蚂蚁 inclusionAI: [HuggingFace 新模型](#) · 21 小时前

包容性AI团队发布ARGenSeg-8B模型,致力于通过开源和开放科学推动人工智能的进步与普及。该举措强调技术民主化,使更广泛的社区能够参与AI研发与应用。开源策略将促进协作创新,加速AI工具在多元场景中的落地,降低技术门槛,推动产业生态的开放发展。

能力进展 新发布

<https://huggingface.co/inclusionAI/ARGenSeg-8B>

3. Granite Embedding Multilingual R2: 开源多语言嵌入模型,具备32K上下文与领先检索性能

Hugging Face: [Blog \(RSS\)](#) · 昨天 02:55

IBM Granite团队在Hugging Face发布了Granite Embedding Multilingual R2多语言文本嵌入模型。该模型采用Apache 2.0开源协议,上下文长度扩展至32K令牌,参数量低于1亿。在MTEB基准的检索评估中,它取得了同规模模型的最佳性能,尤其擅长多语言混合检索,为开发者提供了高效、轻量且可商用的嵌入解决方案。

能力进展 新发布

<https://huggingface.co/blog/ibm-granite/granite-embedding-multilingual-r2>

4. MiniMax M2.7模型上线OrcaRouter平台

X: [MiniMax \(@MiniMax_AI\)](#) · 7 小时前

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

产业 产业与基础设施

1. 为 OpenAI 累计投入超 1000 亿美元,纳德拉称微软当年投资时"没人愿意下注"

IT之家 (RSS) · 21 小时前

在"马斯克诉奥特曼"庭审中,微软企业发展负责人确认,微软对OpenAI的累计投入已超过1000亿美元,其中包括130亿美元原始投资及大量Azure基础设施成本。此次合作已为微软带来约300亿美元营收。CEO纳德拉表示,微软是在"没人愿意下注"时承担了风险。双方已续签非独家协议,微软不再支付收入分成,并将OpenAI的分成上限设为到2030年累计380亿美元,此举较原协议节省约970亿美元。此外,

能力进展 监管/资本 新发布

<https://www.ithome.com/0/950/728.htm>

2. Anthropic与盖茨基金会达成2亿美元合作，聚焦全球健康与教育

Anthropic: Newsroom (网页) · 昨天 23:29

Anthropic与盖茨基金会建立为期四年、总额2亿美元的合作，通过资金、Claude使用额度及技术支持，共同推进全球健康、生命科学、教育及经济流动项目。合作重点包括改善中低收入国家约46亿人口的基本医疗，利用AI加速疫苗与疗法研发，并开发公共卫生数据集等公共产品。在教育领域，双方将为美国、撒哈拉以南非洲和印度的K-12学生开发AI教学工具。经济流动方面则关注提升小农户生产力及美国职业技能认证。首

能力进展 新发布

<https://www.anthropic.com/news/gates-foundation-partnership>

3. 微软研究院发布AI新工具与模型

X: Microsoft Research (@MSFTResearch) · 9 小时前

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>

4. 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) · 9 小时前

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>

5. Anthropic的Mythos AI在五天内协助发现并利用两个未知macOS内核漏洞

X: Rohan Paul (@rohanpaul_ai) · 昨天 07:41

据《华尔街日报》报道，Anthropic的Mythos AI工具在短短五天内，成功帮助研究人员发现了两个此前未知的macOS内核漏洞，并将其串联成一个完整的权限提升攻击链。该攻击针对操作系统最底层的核心，通过组合多个漏洞和技术，绕过了苹果的内存完整性保护机制，访问了本应受保护的系统区域。这凸显出现代macOS的防御重点已从单纯防止漏洞发现，转向增加漏洞利用难度。Mythos在此类研究中展现出强大能

能力进展

https://x.com/rohanpaul_ai/status/2055071152511594832

6. OpenEvidence覆盖65%美国医生，shadow AI模式引关注

X: 小北 (@frxiaobei) · 昨天 01:46

OpenEvidence已覆盖65%的美国医生，4月单月临床场景使用达2700万次，平均每位医生每月使用41次。平台由医生个人通过执业编号在手机上注册，医院最初不知情，Mount Sinai的AI负责人称此为shadow AI，表示其早在基层普及。医院后来才追签企业合作，OpenEvidence强调这是美国医疗史上首次让大多数医生自愿采用单一技术平台的突破。合作伙伴包括NEJM、JAMA、NCCN

新发布

<https://x.com/frxiaobei/status/2054981573150449754>

7. Anthropic估值与营收飙升

X: Kim (@kimmonismus) · 14 小时前

Anthropic简况：估值：2026年2月：3500亿美元 2026年5月：9000亿美元 三个月内增长近3倍。年度经常性收入：2025年底：90亿美元 2026年5月底：450亿美元 五个月内增长近5倍。天啊。via FT

监管/资本

<https://x.com/kimmonismus/status/2055222524774846576>

8. 戛纳倒计时 | Kling AI大会主讲阵容揭晓!

X: 可灵 Kling AI (@Kling_ai) · 9 小时前

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》；Eekjun 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

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) · 10 小时前

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. 百度发布面向大规模智能体应用的全栈AI云

X: [百度 Baidu \(@Baidu_Inc\)](#) · 昨天 22:10

随着智能体应用迈向更大规模部署，其背后的云技术栈也需同步扩展。在百度Create大会上，百度集团执行副总裁、百度智能云事业群总裁沈抖宣布推出专为大规模智能体应用打造的全新全栈AI云，其升级涵盖智能体基础设施与AI基础设施。基于我们自研的昆仑芯AI芯片构建的专用集群，已支持ERNIE 5.1系列中一个关键模型的训练。

能力进展 基础设施 新发布

https://x.com/Baidu_Inc/status/2054927298403688927

3. OpenCode与Qwen 3.6 Plus再度免费开放

X: [opencode \(@opencode\)](#) · 昨天 07:32

OpenCode x Qwen 3.6 Plus - 再次免费 上次各位把我们的容量当成了自助餐。我们找到了更多GPU。第二轮。

能力进展 基础设施 新发布

<https://x.com/opencode/status/2055068702538612784>

4. ChatGPT 推出全新个人理财体验

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

OpenAI 为美国地区的 ChatGPT Pro 用户推出个人理财功能预览版。用户可安全连接个人金融账户，获取基于自身财务状况、目标和优先级生成的 AI 洞察与指导。该体验通过账户关联实现个性化分析，标志着 ChatGPT 正式向个人金融助手领域拓展。

能力进展 监管/资本 新发布

<https://openai.com/index/personal-finance-chatgpt>

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

Claude Code: [GitHub Releases \(RSS\)](#) · 1 小时前

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>

6. 开源3D生成工具包：单张图片快速构建可交互3D世界

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

开发者@neilsonks开源了一套专为Claude Code设计的完整3D生成工具包。该工具能将输入的单张图片自动拆解，生成包含环境、网格、物理、灯光和音频的全套可交互3D场景。其流程首先利用图像与3D生成技术提取物体并生成高质量网格，随后移除物体以得到静态背景，最后为整个场景添加物理模拟、实时灯光和环境音效。配套查看器支持对生成物体的点击编辑与一键导出。此工具将以往常数天的2D转3D工作流程缩

能力进展 新发布

<https://x.com/berryxia/status/2055086814009180316>

7. 将Grok接入Hermes智能体

xAI: [News \(网页\)](#) · 昨天 08:00

xAI宣布，用户现可将Grok订阅账户接入Nous Research的开源自改进智能体Hermes Agent。该集成对所有订阅层级开放，允许用户在Hermes环境中直接使用Grok 4.3进行文本对话与高级推理、调用其文本转语音功能生成语音回复，并利用Grok Imagine创建图像与视频。Hermes Agent可持久运行于电脑、沙盒或VPS，具备跨会话长期记忆能力，并能连接WhatsApp、

能力进展 新发布

<https://x.ai/news/grok-hermes>

8. Claude 代理工具 v2.1.142 版本更新

Claude Code: [GitHub Releases \(RSS\)](#) · 昨天 06:55

Claude 代理工具发布 v2.1.142 版本。本次更新新增了 `--add-dir`、`--settings`、`--model` 等 8 个用于配置后台会话的命令行标志，并将 Fast 模式的默认模型升级为 Opus 4.7。插件功能得到增强，拥有根目录 `SKILL.md` 的插件现可被识别为技能，插件详情面板会显示其提供的 LSP 服务器。此外，版本修复了超过 15 项问题，包括 MC

能力进展 新发布

<https://github.com/anthropics/claude-code/releases/tag/v2.1.142>

9. Recraft AI V4.1上线OpenRouter平台

X: [OpenRouter \(@OpenRouter\)](#) · 昨天 00:09

现已在 OpenRouter 上线：@recraftai V4.1! 包含六款新图像生成模型：追求高美学的 V4.1 和 V4.1 Pro，用于 SVG 插画的 V4.1 Vector 和 V4.1 Pro Vector，以及优先考虑克制风格的产品图像的 V4.1 Utility 和 V4.1 Utility Pro。照片级真实感更自然，渐变更平滑，简短提示词能更准确地命中目标，无需过多手动

能力进展 新发布

<https://x.com/OpenRouter/status/2054957185982177504>

10. 语义代码审查工具clawpatch发布

X: [Peter Steinberger \(@steipete\)](#) · 5 小时前

☑️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>

11. Codex推出自动化钩子与程序化令牌

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 昨天 05:06

Codex正变得更易于围绕用户代码实现自动化与定制。其核心更新包括"钩子"功能, 允许在任务关键节点运行脚本, 以进行工作验证、扫描密钥、记录对话或按仓库定制行为。同时, 面向商业和企业团队推出"程序化访问令牌", 提供范围化凭证, 可从ChatGPT工作区设置创建, 用于CI/CD、发布流程和内部自动化, 支持设置过期或撤销, 并将使用情况关联回工作区。

能力进展 新发布

<https://x.com/OpenAIDevs/status/2055032115964870838>

12. 开源工具html-anything助力Agent生成高质量HTML

X: [小互 \(@xiaohu\)](#) · 昨天 22:03

用户分享了对开源项目html-anything的积极体验。该项目旨在让AI Agent能将任何数据转换为具有世界级设计水准的HTML代码。该项目历时三天开发, 包含约一万五千行代码, 支持75套Skills和9种导出格式, 并能兼容包括Claude Code、Codex、OpenClaw、Hermes在内的多种代码生成Agent。

能力进展 新发布

<https://x.com/xiaohu/status/2054925632061231431>

13. Kimi推出网页桥接扩展 支持多平台交互

X: [Kimi.ai \(@Kimi_Moonshot\)](#) · 昨天 21:34

认识Kimi网页桥接--Kimi的浏览器扩展。现在智能体可以像人类一样与网站互动: 搜索、滚动、点击、输入并完成任务。支持Kimi Code CLI、Claude Code、Cursor、Codex、Hermes等平台。现已在<http://kimi.com/features/webbridge>和Chrome应用商店上线。

能力进展 新发布

https://x.com/Kimi_Moonshot/status/2054918374837322140

14. 随时随地使用 Codex

OpenAI: [官网动态 \(RSS · 排除企业/客户案例\)](#) · 昨天 21:00

用户现可通过 ChatGPT 移动应用随时随地使用 Codex。该功能支持跨设备和远程环境实时监控、引导及批准编码任务, 实现了对编程工作的无缝移动端管理。

能力进展 新发布

<https://openai.com/index/work-with-codex-from-anywhere>

15. Luma Agents高效生成电商素材全流程

X: [Luma AI \(@LumaLabsAI\)](#) · 昨天 06:05

更多产品。更多市场。更多格式。再无瓶颈。定义需求。设定风格。Luma Agents 从此处理所有电商活动素材。立即扩展 → <http://lumalabs.ai/app>

能力进展

<https://x.com/LumaLabsAI/status/2055046873740984429>

16. Runway Agent一键生成完整广告

X: [Runway \(@runwayml\)](#) · 5 小时前

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>

17. OpenClaw新版本速度提升约3.5倍

X: [OpenClaw \(@openclaw\)](#) · 11 小时前

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>

18. OpenRouter BYOK三大升级 支持多密钥轮换

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

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>

19. Mixpanel集成Replit MCP，开发流程内嵌数据分析

X: [Replit \(@Replit\)](#) · 昨天 04:11

发布产品。衡量效果。全在一个流程中完成。@Mixpanel 现已登陆 Replit MCP。下周伦敦黑客松现场演示

新发布

<https://x.com/Replit/status/2055018223431454850>

20. SuperGrok Heavy限时六折，Grok Build开放测试

X: [cb_doge \(@cb_doge\)](#) · 昨天 04:10

SuperGrok Heavy 现提供约67%的半年折扣，即每月仅需99美元（原价300美元）。强烈建议升级至Heavy版本，并试用Grok Build的测试版。

新发布

https://x.com/cb_doge/status/2055017857352913319

21. Suno应用更新亮点

X: [Suno \(@suno\)](#) · 昨天 23:35

Suno应用刚刚焕然一新。过去几周我们进行了一些更新。以下是我们喜爱的部分亮点

新发布

<https://x.com/suno/status/2054948668625559902>

22. 计算机直连Snowflake实现实时数据洞察

X: [Perplexity \(@perplexity_ai\)](#) · 昨天 23:08

计算机现已连接至Snowflake。可基于实时仓库数据开展端到端工作，通过SQL、源表、筛选器和指标获取答案。这就像一支随时待命的个人数据科学团队，从公司实时数据中提供精准答案。

https://x.com/perplexity_ai/status/2054941905633612059

23. 一键生成F1进站时刻肖像特效

X: [PixVerse \(@PixVerse_\)](#) · 昨天 23:00

不容错过PitCrewMoment潮流。一键将任何肖像转化为F1直播电视进站时刻。立即在PixVerse网页端尝试！

https://x.com/PixVerse_/status/2054939801435177238

24. Krea 2正式上线面向专业用户

X: [Krea AI \(@krea_ai\)](#) · 4小时前

从今天起，Krea 2 正式为 Pro 用户上线。

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

25. Pixverse模板：让你成为全场焦点

X: [PixVerse \(@PixVerse_\)](#) · 13小时前

Pixverse让你成为焦点~用Pixverse网页版的音乐会聚光灯模板制作吧！

https://x.com/PixVerse_/status/2055245210184200497

研究 研究与开源进展

1. AI自主研究实现突破：智能体在nanoGPT优化赛道上超越人类基准

X: [Berry Xia \(@berryxia\)](#) · 昨天 07:55

Prime Intellect 近期在AI研究自动化领域取得重要进展。他们让Claude Code与Codex智能体完全自主运行于nanoGPT速度挑战的优化器赛道，利用闲置算力完成了近万次实验，消耗约1.4万H200小时。最终，Claude Code将记录提升至2930步，超越了2990步的人类基准。实验显示，智能体在系统整合社区主流优化方法、进行超参数扫描和策略组合方面效率极高，但在要求真正创

能力进展 基础设施 新发布

<https://x.com/berryxia/status/2055074608261578949>

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 \(@AYI_AInotes\)](#) · 11小时前

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) · 6 小时前

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. 克劳德代码与《代码书》技能：有针对性的技能培养

Hacker News 热门 (buzzing.cc 中文翻译) · 昨天 21:23

开发者发布了一款名为"克劳德代码与《代码书》技能"的GitHub工具，旨在通过刻意练习提升编程技能。该工具利用AI模型生成特定主题的代码示例与解释，帮助用户进行针对性学习。项目在Hacker News上获得104点热度，关注度较高。其核心变化在于将传统的广泛学习转化为聚焦、可重复的技能训练模式，通过结构化练习提升学习效率。

能力进展 基础设施 新发布

<https://github.com/DrCatHicks/learning-opportunities>

4. 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) · 2 小时前

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>

5. 在法律行业全面部署Claude：产品指南与实施路线图

Claude: Blog (网页) · 7 小时前

2026年报告显示，法律团队生成式AI使用率已从44%跃升至87%。为应对日益复杂的工作，法律行业正将Claude应用于合同审阅、并购尽调及诉讼准备等核心流程，并通过多款产品组合提升效率：Chat用于即时研究，Claude Cowork处理跨文件协作，Microsoft 365插件集成办公套件，Platform支持定制应用开发。Anthropic同步发布法律行业部署指南，涵盖产品矩阵、12个预设业

能力进展 监管/资本 新发布

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

6. 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) · 9 小时前

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、Cla

能力进展 监管/资本 新发布

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

7. 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) · 昨天 08:00

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。

能力进展 监管/资本 新发布

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

8. 推理的一阶导数：AI浪潮下的增长逻辑

Tomer Tunguz 博客 (VC 分析) · 昨天 08:00

AI推理是当今规模最大、增长最快的技术市场，预计七年内将达到2500亿美元。直接销售或转售推理服务的公司增长迅猛，如Anthropic和谷歌云。在AI时代前的软件公司中，Datadog和Twilio作为"推理的一阶导数"脱颖而出：Datadog的LLM可观测产品数据量近一季增长近两倍，其约20%的AI客户贡献了约80%的年度经常性收入；Twilio则通过AI重构的语音服务吸引客户。当前周期呈现高度

能力进展 基础设施

<https://www.tomtunguz.com/first-derivative-of-inference>

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

X: 宝玉 (@dotey) · 9 小时前

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>

10. 牛津大学博士后开源视频翻译工具Violin，支持多语言翻译与视频对话

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

牛津大学博士后Kevin Lin开源视频翻译工具Violin，旨在打破高质量视频内容的语言壁垒。该工具将语音识别、大语言模型翻译与语音合成整合为自动化流水线，支持多语言互译与个性化翻译风格调整，例如将学术报告转化为儿童易懂版本。用户还能直接与视频内容进行对话并获取相关答案。Violin提供Web应用、命令行界面和Agent Skill三种使用方式，所有功能基于MIT协议开源，由Together C

能力进展 新发布

<https://x.com/berryxia/status/2055093068085547507>

11. Show HN：根据基准测试排名，为您的硬件寻找最适合的本地 LLM

Hacker News 热门 ([buzzing.cc 中文翻译](#)) · 12 小时前

一个名为"WhichLLM"的开源工具已在GitHub发布，可根据用户硬件配置推荐最适合的本地大语言模型。该工具通过基准测试对各类模型进行排名，帮助用户依据自身设备的性能指标选择最优模型。项目在Hacker News上获得116点热度，显示出社区对本地化AI部署效率工具的积极关注。

能力进展 新发布

<https://github.com/Andyyyy64/whichllm>

12. 克劳德（法律事务）

Hacker News 热门 ([buzzing.cc 中文翻译](#)) · 13 小时前

Anthropic公司在GitHub上开源了专为法律领域优化的AI模型"Claude for Legal"。该模型能更准确地处理法律文本、合同及案例摘要，旨在提升法律从业者效率。此消息在Hacker News上获得105个投票点数，显示出技术社区对其在法律科技领域应用潜力的关注。

能力进展 新发布

<https://github.com/anthropics/claude-for-legal>

13. Yetone发布Native Feel桌面应用开发Agent Skill

X: [Berry Xia \(@berryxia\)](#) · 昨天 07:54

开发者Yetone将一篇关于桌面应用开发"最佳实践"的文章转化为一个名为"native-feel-skill"的Agent Skill。该Skill旨在帮助开发者利用Coding Agent，轻松地重构或开发跨平台桌面应用，并使其获得极其接近Native原生应用的性能体验。项目代码已开源在GitHub上。

能力进展 新发布

<https://x.com/berryxia/status/2055074211715301830>

14. 为什么资深开发者讲不清自己的专业能力

X: [宝玉 \(@dotey\)](#) · 22 小时前

资深开发者与业务团队存在根本认知差异。业务团队生活在"消除不确定性"的循环中，追求快速试错验证，核心是速度。而资深开发者身处"管理复杂性"的循环，核心职责是保障付费服务的长期稳定，因此对增加系统复杂性的行为极为警惕。沟通失败在于，开发者用"控制复杂性"的理由拒绝需求，却未回应业务端"消除不确定性"的迫切诉求。解决方案是，开发者应将其精简需求、复用代码等专业能力，包装成能帮助业务"更快获得答案"的方

能力进展

<https://x.com/dotey/status/2055097242755706984>

15. API提示预缓存加速首令牌生成

X: [Claude Devs \(@ClaudeDevs\)](#) · 昨天 07:35

减少API长提示首令牌生成时间的实用技巧：预热提示缓存。在用户提示前发送系统提示。Claude会将其写入缓存，但跳过生成任何输出。当真实用户请求到达时，将直接命中预热缓存。

能力进展

<https://x.com/ClaudeDevs/status/2055069548672631218>

16. 英伟达CEO称技工前景优于计算机科学毕业生

X: [Kim \(@kimmonismus\)](#) · 2 小时前

英伟达CEO Jensen Huang在卡内基梅隆大学告诉2026届计算机科学毕业生，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>

17. AI助手Sai化身虚拟同事，自主完成深度研究

X: [Kim \(@kimmonismus\)](#) · 6 小时前

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>

18. 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 \(RSS\)](#) · 10 小时前

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>

19. Eric Jang - 从零开始构建 AlphaGo

Dwarkesh Patel: [Podcast & Blog \(RSS\)](#) · 8 小时前

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>
