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: 53 条 焦点: 8 条 快讯: 0 条

Executive Summary

Cognition 宣布成为全球最大独立智能体实验室，完成超10亿美元融资，估值达260亿美元，其年化收入增至4.92亿美元，标志着AI软件工程师赛道获得资本市场高度认可。Claude Code 发布v2.1.152版本，新增安全漏洞识别插件并优化代码审查功能，Perplexity 开源Unigram分词器将CPU占用降低5-6倍，Qwen3.5 在TokenSpeed推理引擎上达到580 tokens per second的创纪录速度。FastVideo Dreamverse 实现单张NVIDIA B200 GPU实时视频生成，7秒生成30秒1080p视频。

OpenRouter 完成1.13亿美元B轮投资，周处理量从5万亿增长至25万亿token，显示AI从实验向生产转化的加速趋势。阿里云 成为PyTorch基金会白金会员并入选Omdia智能体AI市场雷达领导者，高通 与字节跳动达成AI ASIC芯片合作，采购量达数百万颗级别。Runway MCP服务器 和 Google Pay 的智能体商务协议推进MCP标准化进程，MiMo 2.5 Pro 大幅降价99%并与DeepSeek V4 Pro同价，反映API市场竞争日趋激烈。

后续需跟踪Cognition的Devin产品商业化进展、MCP协议标准化对智能体生态的影响、Qwen推理优化技术的开源节奏，以及高通与字节跳动芯片合作对AI终端部署格局的具体影响。API价格战可能持续重塑成本结构，而智能体工作负载对算力基础设施的需求将持续增长。

重点 今日核心进展

★ 1. Claude Code v2.1.152 更新发布

Claude Code：GitHub Releases（RSS） · 22 小时前 · 应用与商业化

Claude Code 发布 v2.1.152 版本更新。核心改进包括：`/code-review --fix` 现在会将审查建议直接应用于工作目录；技能与斜杠命令支持通过 frontmatter 的 `disallowed-tools` 移除模型工具；新增 `/reload-skills` 命令可不重启会话重新扫描技能目录；`SessionStart` 钩子现可返回 `reloadSkills:

能力进展 基础设施 新发布

https://github.com/anthropics/claude-code/releases/tag/v2.1.152

★ 2. Cognition成为全球最大独立智能体实验室

X：swyx (@swyx) · 4 小时前 · 产业与基础设施

Cognition宣布已成为全球最大的独立智能体实验室。公司完成超10亿美元融资，估值达260亿美元，由Lux Capital、General Catalyst等领投。其企业使用量自年初增长超10倍，年化收入增至4.92亿美元。Cognition于两年前推出Devin，定位为首个AI软件工程师。公司强调其拥有多项领先优势，包括首个编码智能体、顶级代码审查能力等，并得到了Peter Thiel的重大

能力进展 监管/资本 新发布

https://x.com/swyx/status/2059717021944926238

★ 3. 开源FastVideo Dreamverse实时视频生成工具

X：Sky Computing Lab (@haoailab) · 6 小时前 · 应用与商业化

仅需7秒即可生成30秒1080p视频！ 我们开源了FastVideo Dreamverse：基于单张NVIDIA B200 GPU和LTX-2模型，实现实时视频生成的氛围引导工具。 Repo： https：//github.com/hao-ai-lab/FastVideo/tree/main/apps/dreamverse Blog： https：//haoailab.com/blogs/f

能力进展 基础设施 新发布

https://x.com/haoailab/status/2059695648103112946

★ 4. 阿里云成为PyTorch基金会白金会员

X：阿里云 / Alibaba Cloud (@alibaba_cloud) · 22 小时前 · 产业与基础设施

阿里云宣布以白金会员身份加入PyTorch基金会。作为AI基础设施全球领导者，阿里云是开源模型家族Qwen的缔造方，已在多样化硬件上大规模运行PyTorch，将为社区带来生产级工程经验。

能力进展 基础设施 新发布

https://x.com/alibaba_cloud/status/2059453607075004835

★ 5. Perplexity开源Unigram分词器降低CPU占用

X：Perplexity (@perplexity_ai) · 8 小时前 · 应用与商业化

我们开源了重新构建的Unigram分词器，可将CPU占用降低5-6倍。 小型重排序器和嵌入模型在GPU上运行时间仅为个位数毫秒，使得CPU分词成为总延迟的重要组成部分。 http：//github.com/perplexityai/pplx-garden

能力进展 基础设施 新发布

https://x.com/perplexity_ai/status/2059664738087469511

★ 6. OpenRouter完成1.13亿美元B轮融资

X：OpenRouter (@OpenRouter) · 昨天 22:16 · 产业与基础设施

今天我们宣布完成由@CapitalGVC领投的1.13亿美元B轮融资。 过去6个月，随着AI从实验快速转向生产，OpenRouter的周处理量从5万亿增长到25万亿token。 我们对未来充满期待。

能力进展 监管/资本 新发布

https://x.com/OpenRouter/status/2059277623629664758

★ 7. OpenAI产品支持私有MCP服务器安全连接

X：OpenAI Developers (@OpenAIDevs) · 5 小时前 · 应用与商业化

您的团队可以在内部网络中保留MCP服务器，同时ChatGPT、Codex和Responses API通过仅出站HTTPS进行连接。

能力进展 监管/资本 新发布

https://x.com/OpenAIDevs/status/2059703536825565499

★ 8. Claude Code推出安全漏洞识别插件

X：Claude Devs (@ClaudeDevs) · 昨天 05:24 · 应用与商业化

我们为Claude Code发布了一个安全指导插件，可在编写代码时帮助识别和修复漏洞。 所有Claude Code用户均可使用。从插件市场（/plugins）安装。

能力进展 监管/资本 新发布

https://x.com/ClaudeDevs/status/2059385239781384341

产业 产业与基础设施

1. Cognition成为全球最大独立智能体实验室

X：swyx (@swyx) · 4 小时前

Cognition宣布已成为全球最大的独立智能体实验室。公司完成超10亿美元融资，估值达260亿美元，由Lux Capital、General Catalyst等领投。其企业使用量自年初增长超10倍，年化收入增至4.92亿美元。Cognition于两年前推出Devin，定位为首个AI软件工程师。公司强调其拥有多项领先优势，包括首个编码智能体、顶级代码审查能力等，并得到了Peter Thiel的重大

能力进展 监管/资本 新发布

https://x.com/swyx/status/2059717021944926238

2. 阿里云成为PyTorch基金会白金会员

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

阿里云宣布以白金会员身份加入PyTorch基金会。作为AI基础设施全球领导者，阿里云是开源模型家族Qwen的缔造方，已在多样化硬件上大规模运行PyTorch，将为社区带来生产级工程经验。

能力进展 基础设施 新发布

https://x.com/alibaba_cloud/status/2059453607075004835

3. OpenRouter完成1.13亿美元B轮融资

X：OpenRouter (@OpenRouter) · 昨天 22:16

今天我们宣布完成由@CapitalGVC领投的1.13亿美元B轮融资。 过去6个月，随着AI从实验快速转向生产，OpenRouter的周处理量从5万亿增长到25万亿token。 我们对未来充满期待。

能力进展 监管/资本 新发布

https://x.com/OpenRouter/status/2059277623629664758

4. 阿里云入选Omdia智能体AI市场雷达领导者

X：阿里云 / Alibaba Cloud (@alibaba_cloud) · 18 小时前

阿里云在Omdia的智能体AI市场雷达中被评为领导者。Omdia强调了阿里云在每一层的全栈能力，认可其是首个将整个平台围绕智能体范式进行构建的云服务商。

能力进展 基础设施

https://x.com/alibaba_cloud/status/2059509714854007181

5. 消息称高通与字节跳动达成 AI ASIC 芯片合作，采购量在数百万颗级别

IT之家（RSS） · 23 小时前

IT之家（RSS）披露：消息称高通与字节跳动达成 AI ASIC 芯片合作，采购量在数百万颗级别。该条属于产业与基础设施方向，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 基础设施

https://www.ithome.com/0/955/674.htm

6. 据报道Claude Mythos以"巧妙简洁的证明"解决了OpenAI里程碑式的Erdős问题

The Decoder：AI News（RSS） · 昨天 02:31

Anthropic工程师Sholto Douglas表示，Claude Mythos在周末期间解决了OpenAI提出的Erdős单位距离猜想问题，并给出了一个"巧妙简洁的证明"。这一成果被描述为人工智能在数学发现领域存在"严重超前"迹象。

能力进展 新发布

https://the-decoder.com/claude-mythos-reportedly-solves-openais-landmark-erdos-problem-with-a-cute-simple-proof

7. 思科与OpenAI携手Codex重新定义企业工程

OpenAI：官网动态（RSS · 排除企业/客户案例） · 13 小时前

思科正与OpenAI合作，利用Codex重新定义企业工程实践。此次合作旨在帮助思科扩展AI原生开发能力、加速其AI Defense安全工作，并实现缺陷修复流程的自动化。

监管/资本 新发布

https://openai.com/index/cisco

8. 我国将加快研究推进人工智能健康发展综合性立法、低空经济立法等

IT之家（RSS） · 21 小时前

IT之家（RSS）披露：我国将加快研究推进人工智能健康发展综合性立法、低空经济立法等。该条属于产业与基础设施方向，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

https://www.ithome.com/0/955/758.htm

9. Replit入选Redpoint 2026 InfraRed 100榜单

X：Replit (@Replit) · 6 小时前

Replit 被列入 @Redpoint 的 2026 InfraRed 100 榜单。 这是一份构建 AI 运行基础设施的公司名单。很荣幸能入选，与 Stripe、Snowflake 和 HashiCorp 等往届入选者并列。 致每一位在 Replit 上发布产品的构建者：这份荣誉属于你们。 https：//www.redpoint.com/infrared/report/

新发布

https://x.com/Replit/status/2059688584899154134

10. 教皇没对AGI上头

The Verge：AI（RSS） · 12 小时前

教皇 Leo XIV 发布题为《Magnifica Humanitas》的通谕，警告人工智能的使用绝非纯粹技术问题，当其进入影响人类生活的过程时，便触及权利、机会、地位与自由。通谕发布时，Anthropic 联合创始人 Christopher Olah 出席。文件引发了科技界内外的广泛反应。

新发布

https://www.theverge.com/ai-artificial-intelligence/937933/pope-ai-encyclical-tech-industry-reactions

11. 黄仁勋展示英伟达台湾新园区

X：Rohan Paul (@rohanpaul_ai) · 6 小时前

黄仁勋展示了新的台湾园区。 英伟达计划每年在台湾投资约1500亿美元。 就在竞争对手AMD宣布将向台湾AI领域投资超过100亿美元一周后。

基础设施

https://x.com/rohanpaul_ai/status/2059689400267939925

12. 2026年选举信息与保障

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

在2026年全球选举临近之际，通过帮助公众获取选举信息、支持网络防御者以及提升人工智能透明度这三方面的努力，来为选举提供保障。

新发布

https://openai.com/index/election-safeguards-2026

13. 微软亚洲研究院发起全球AI价值观挑战赛

X：Microsoft Research (@MSFTResearch) · 昨天 22:00

AI能否理解人类价值观的现实复杂性？微软亚洲研究院的全新全球AI价值观挑战赛，诚邀哲学、伦理、法学和社会科学领域的研究人员共同探索。立即注册：https：//msft.it/6012vk8OO

https://x.com/MSFTResearch/status/2059273348232724565

应用 应用与商业化

1. Claude Code v2.1.152 更新发布

Claude Code：GitHub Releases（RSS） · 22 小时前

Claude Code 发布 v2.1.152 版本更新。核心改进包括：`/code-review --fix` 现在会将审查建议直接应用于工作目录；技能与斜杠命令支持通过 frontmatter 的 `disallowed-tools` 移除模型工具；新增 `/reload-skills` 命令可不重启会话重新扫描技能目录；`SessionStart` 钩子现可返回 `reloadSkills：

能力进展 基础设施 新发布

https://github.com/anthropics/claude-code/releases/tag/v2.1.152

2. 开源FastVideo Dreamverse实时视频生成工具

X：Sky Computing Lab (@haoailab) · 6 小时前

⌧仅需7秒即可生成30秒1080p视频！ 我们开源了FastVideo Dreamverse：基于单张NVIDIA B200 GPU和LTX-2模型，实现实时视频生成的氛围引导工具。 Repo：https：//github.com/hao-ai-lab/FastVideo/tree/main/apps/dreamverse Blog： https：//haoailab.com/blogs/f

能力进展 基础设施 新发布

https://x.com/haoailab/status/2059695648103112946

3. Perplexity开源Unigram分词器降低CPU占用

X：Perplexity (@perplexity_ai) · 8 小时前

我们开源了重新构建的Unigram分词器，可将CPU占用降低5-6倍。 小型重排序器和嵌入模型在GPU上运行时间仅为个位数毫秒，使得CPU分词成为总延迟的重要组成部分。 http：//github.com/perplexityai/pplx-garden

能力进展 基础设施 新发布

https://x.com/perplexity_ai/status/2059664738087469511

---

4. OpenAI产品支持私有MCP服务器安全连接

X：OpenAI Developers (@OpenAIDevs) · 5 小时前

您的团队可以在内部网络中保留MCP服务器，同时ChatGPT、Codex和Responses API通过仅出站HTTPS进行连接。

能力进展 监管/资本 新发布

https://x.com/OpenAIDevs/status/2059703536825565499

---

5. Claude Code推出安全漏洞识别插件

X：Claude Devs (@ClaudeDevs) · 昨天 05:24

我们为Claude Code发布了一个安全指导插件，可在编写代码时帮助识别和修复漏洞。 所有Claude Code用户均可使用。从插件市场（/plugins）安装。

能力进展 监管/资本 新发布

https://x.com/ClaudeDevs/status/2059385239781384341

---

6. Runway 推出 Model Context Protocol 服务器

Runway：News（网页） · 10 小时前

Runway 正式推出 Runway MCP 服务器，允许任何兼容 MCP 的 AI 智能体（如 Claude、ChatGPT、Cursor）在对话界面中直接生成图像与视频，无需切换工作流。该服务器接入了 Runway 最新的多款 SOTA 模型，包括 Gen-4.5、Seedance 2.0、GPT Image 2、Kling 3.0 及 Nano Banana Pro。其应用场景涵盖为产品制作

能力进展 新发布

https://runwayml.com/news/mcp

---

7. Google Pay 最新更新

Google Developers Blog（RSS） · 6 小时前

Google Pay 正向"智能体商务"演进，推出了通用商务协议和新的 MCP 服务器，允许 AI 智能体管理集成与分析趋势。Android 平台更新引入了动态回调以支持快速结账，并通过 WebView 将支付功能扩展至社交媒体应用。此外，平台还推出了跨设备生物认证和新的交易信号，旨在帮助商家减少流程摩擦。

能力进展 新发布

https://developers.googleblog.com/the-latest-updates-to-google-pay

---

8. Krea 2 API发布，支持多平台与智能体

X：Krea AI (@krea_ai) · 9 小时前

今天，我们发布了 Krea 2 的 API。 现已在 @fal 或 @ComfyUI 等平台可用，通过 @NousResearch 的 Hermes 等智能体使用，并全面支持 Claude、Codex 或 OpenClaw。 了解如何设置 ☒

能力进展 新发布

https://x.com/krea_ai/status/2059650622203515143

---

9. MiMo 2.5 Pro大幅降价，与DeepSeek V4 Pro同价

X：Kim (@kimmonismus) · 昨天 03:21

小米MiMo-V2.5系列API价格永久下调，最高降幅达99%，现与DeepSeek V4 Pro同价。Token套餐同步升级，同等价格下可用token量增加5-8倍，计费规则更简单透明。所有现有用户套餐额度将全额重置。此次降价源于MiMo全栈推理优化与服务效率提升，后续将发布技术博客详述细节。MiMo-V2.5-TTS限时免费，新定价于5月26日生效。

能力进展 新发布

https://x.com/kimmonismus/status/2059354372643975490

---

10. Web 更新

Midjourney：Updates（RSS） · 5 小时前

对话模式在文本和语音输入方面进行了改进。语音会话开始时，可访问用户的图像提示、风格参考、侧边栏设置和最近任务。图像提示功能现可从托盘和侧边栏直接使用。在语音提交过程中，托盘中的图像将保持不变，直至用户手动移除。

能力进展 新发布

https://updates.midjourney.com/web-updates-5

---

11. 通过万亿参数与 Hub Bucket 实现增量权重同步：TRL 中的增量权重同步

Hugging Face：Blog（RSS） · 昨天 08:00

本文标题涉及 Hugging Face TRL 框架中一项具体的增量权重同步技术。正文重申了 Hugging Face 的核心使命，即致力于通过开源和开放科学，来推动人工智能的进步与普及。

能力进展 新发布

https://huggingface.co/blog/delta-weight-sync

---

12. OpenCode与MiMo V2.5限时免费开放

X：opencode (@opencode) · 6 小时前

OpenCode x MiMo V2.5 - 限时免费 1M 上下文 · 推理 · 文本 · 图像

能力进展 新发布

https://x.com/opencode/status/2059696100626297225

13. Qoder平台限时半价使用Qwen3.7-Max模型

X：通义千问 / Qwen (@Alibaba_Qwen) · 13 小时前

【引用 @qoder_ai_ide】：Qwen3.7-Max，半价。 从今天起，Qwen3.7-Max--通义千问（Qwen）家族的最新旗舰模型--在Qoder上半价提供。限时活动。 新用户？你每天还能获得100次免费模型调用。自动应用，无需领取，无需开关。 桌面端、JetBrains插件、CLI、QoderWork、QoderWake--全部覆盖。 现在正是用它处理难题的好时机。

能力进展

https://x.com/Alibaba_Qwen/status/2059586373456380225

14. Grok编程智能体登陆Kilo IDE平台

X：xAI (@xai) · 8 小时前

在 @kilocode 中使用您的 SuperGrok 或 X Premium+ 订阅。 尝试 grok-build-0.1，享受高速和智能体编程智能，可在 Kilo IDE 扩展或 CLI 中使用。 https：//x.ai/news/grok-kilocode

能力进展

https://x.com/xai/status/2059666227115819149

15. Claude Marketplace 新增五家合作伙伴

X：Claude (@claudeai) · 8 小时前

Claude Marketplace 新增成员：@augmentcode、@boltdotnew、@coderabbitai、@hebbia 和 @WeAreLegora。 您现有的 Anthropic 消费承诺可用于购买其 Claude 驱动的产品。 了解更多：http：//claude.com/platform/marketplace

能力进展

https://x.com/claudeai/status/2059662933924123044

16. Replit 应用添加登录的两种方式

X：Replit (@Replit) · 昨天 00:00

在 Replit 上为你的应用添加登录有两种方式： → Replit Auth：零配置，用户使用其 Replit 账户登录 → Clerk Auth：你自己的品牌化登录，开发/生产环境均只需一个提示词。文档和视频见下方 文档：https：//docs.replit.com/learn/projects-and-artifacts/auth#auth

能力进展

https://x.com/Replit/status/2059303550375674139

研究 研究与开源进展

1. Fast， faster， Qwen.

X：通义千问 / Qwen (@Alibaba_Qwen) · 7 小时前

Qwen3.5在TokenSpeed推理引擎上，针对智能体工作负载达到了创纪录的580 tokens per second （tps）速度。这一成果由通义千问推理团队、lightseekorg Foundation TokenSpeed团队、NVIDIA及Mooncake团队共同实现，并采用了tri_dao的FlashAttention-4 （FA4） 优化。此里程碑标志着开源大语言模型推理性能的

能力进展 基础设施 新发布

https://x.com/Alibaba_Qwen/status/2059674574397313277

2. KPop 新方法让 Ring-2.6-1T 在 SWE-bench Verified 上突破 76 分

X：蚂蚁百灵 (@antLingAGI) · 昨天 23:14

团队推出 KPop，用于稳定大规模 MoE 模型的智能体强化学习训练。它用基于二元 KL 散度的自适应掩码机制，替代了此前 IcePop 方法中的固定比例掩码，能根据训练过程中的训练-推理不匹配程度动态调整。这一改进使得 Ring-2.6-1T 模型在无需修改基础设施或路由重放的情况下，仅通过纯 RL 训练，在 SWE-bench Verified 上取得了超过 76 分的成绩。

能力进展 基础设施 新发布

https://x.com/antLingAGI/status/2059292063032918422

3. ITBench-AA：前沿大模型在首个智能体企业IT任务基准测试中得分均低于50%

Hugging Face：Blog（RSS） · 7 小时前

由Artificial Analysis和IBM推出的ITBench-AA SRE基准测试显示，所有前沿大模型得分均未超过50%。Claude Opus 4.7（自适应推理，最大努力）以47%领先，GPT-5.5（xhigh）和Qwen3.7 Max分别得46%和42%。该测试包含59个需要通过Shell命令调查Kubernetes事件快照并提交根因诊断的智能体任务。关键发现是模型推理轮次差异近3

能力进展 新发布

https://huggingface.co/blog/ibm-research/itbench-aa

4. SilverTorch：索引即模型--推荐系统的新检索范式

Meta Engineering Blog（RSS） · 昨天 00:00

Meta 推出SilverTorch推荐系统架构，统一了用户生成内容的所有检索组件。该架构吞吐量比现有技术高23.7倍，计算成本效率比CPU方案高20.9倍，同时提升了准确性。

能力进展 新发布

https://engineering.fb.com/2026/05/26/ml-applications/silvertorch-index-as-model-new-retrieval-paradigm-recommendation-systems

5. 社会科学中的编码智能体

Anthropic：Research（发表成果 · 网页） · 1 小时前

一项针对1260名定量社会科学家的调查显示，虽然81%的受访者用过AI聊天机器人，但仅有20%将Claude Code、Codex等编码智能体常规应用于工作。采用率存在显著差异：以男性名字命名的研究者使用率是女性研究者的两倍；顶尖大学研究者可能性高出40%。用户产出更多工作论文和基金申请，但这可能反映早期采用者自身差异。研究者对AI助力撰写可发表论文更乐观，但对重塑整个社会科学领域持保留态度。这是

能力进展

https://www.anthropic.com/research/coding-agents-social-sciences

6. 通过零信任聚合实现的隐私分析

Google Research：Blog（网页） · 6 小时前

Google Research 推出了一种新的隐私分析解决方案。该方案结合了一种新的密码学安全聚合协议与可信执行环境（TEE）的透明性，旨在实现前沿的隐私与安全保证。其核心是基于零信任原则，通过密码学与硬件保护的结合，确保系统仅能获取群体的匿名化聚合洞察。

监管/资本 新发布

https://research.google/blog/private-analytics-via-zero-trust-aggregation

格局 观点、资本与监管

1. Sundar Pichai 谈 AI、搜索的未来及网络的变化

The Verge：AI（RSS） · 昨天 22:00

Google 与 Alphabet CEO Sundar Pichai 在 Google I/O 后受访，回顾了公司为应对 ChatGPT 而进行的战略重组与高管调整。访谈聚焦于新的 Gemini 模型及其在产品中的整合，包括全新的智能搜索框与 Gemini Spark 智能体平台，旨在让搜索从提供结果转向启动任务。Pichai 讨论了这些变化对开放网络的持续冲击，回应了主持人此前提出的"Goog

能力进展 基础设施 新发布

https://www.theverge.com/podcast/936445/sundar-pichai-ai-search-google-zero-youtube-web

2. SenseNova-U1全训练代码开源，支持多模态多任务训练

X：商汤 SenseTime (@SenseTime_AI) · 昨天 22:58

OpenSenseNova开源了SenseNova-U1的完整训练代码库，支持其8B密集模型与A3B MoE架构。该代码库使用一个统一的框架，可同时训练多种多模态任务，包括文本到图像生成、图像编辑、交错生成及文本与视觉理解。工程上为大规模训练设计，支持混合并行、流式可恢复数据管线、环境变量驱动配置以及从1×8 GPUs到多节点集群的扩展能力。代码已在GitHub开源，采用Apache-2.0协议。

能力进展 基础设施 新发布

https://x.com/SenseTime_AI/status/2059288013994406013

3. 使用大语言模型保障源代码安全

Claude：Blog（网页） · 2 小时前

本文分享了使用 Claude Opus 构建威胁模型、发现代码漏洞并进行验证、分类和修复的最佳实践。其核心流程是一个六步循环：威胁建模、沙箱隔离、漏洞发现、验证、分类和修复。作者指出，漏洞发现现在易于并行化，瓶颈已转移到后续的验证与处理阶段。以他们对开源软件的扫描为例，截至2026年5月22日已披露1，596个漏洞，其中97个已修补。指南建议结合代码库文档和专家访谈来构建准确的威胁模型，以降低误报

能力进展 监管/资本 新发布

https://claude.com/blog/using-llms-to-secure-source-code

4. AI智能体的零信任安全框架

Claude：Blog（网页） · 6 小时前

Anthropic 发布了针对企业部署自主 AI 智能体的安全框架，指出前沿大语言模型正将漏洞利用周期从数月压缩至数小时。部署智能体面临双重风险：基础设施易受 AI 加速攻击，且智能体自身具备自主决策与执行能力。文章提出一个三层零信任架构（基础、高级、优化级）及八阶段实施流程，并概述了提示注入、工具投毒、记忆投毒等特有威胁。

能力进展 监管/资本 新发布

https://claude.com/blog/zero-trust-for-ai-agents

5. OpenAI 奥尔特曼称 AI 对白领冲击不如预期般严重：我很高兴自己当时错了

IT之家（RSS） · 16 小时前

IT之家（RSS）披露：OpenAI 奥尔特曼称 AI 对白领冲击不如预期般严重：我很高兴自己当时错了。该条属于观点、资本与监管方向，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 监管/资本 新发布

https://www.ithome.com/0/956/021.htm

6. Reachy Mini 实现完全本地化语音交互

Hugging Face：Blog（RSS） · 昨天 08:00

Reachy Mini 机器人现可通过 `speech-to-speech` 库实现完全本地化的语音交互，无需依赖云端。该方案采用级联流水线架构，对外提供 Realtime API 兼容的 WebSocket 接口。默认组件包括 Silero VAD 用于语音活动检测、Parakeet-TDT 作为语音转文本模型、通义千问（Qwen3-TTS）作为文本转语音模型。大语言模型推荐使用 llama.c

能力进展 基础设施

https://huggingface.co/blog/local-reachy-mini-conversation

---

7. 我认为 Anthropic 和 OpenAI 找到了产品市场契合点

Simon Willison 博客 · 7 小时前

Anthropic 与 OpenAI 通过编程智能体找到了产品市场契合点，这导致企业客户成本显著上升。两家公司已于 2026 年 4 月前后调整了企业套餐定价，从原先的高额折扣改为与 API 用量挂钩。Anthropic Enterprise 套餐变为每席位 20 美元/月外加 API 费用，OpenAI Codex 则按 API token 用量计费。同期发布的新模型 GPT-5.5（4月23日

能力进展 新发布

https://simonwillison.net/2026/May/27/product-market-fit

---

8. Gemini Omni 视频提示词使用指南

X：Google AI (@GoogleAI) · 昨天 05:08

Google 发布了其多模态模型 Gemini Omni 的视频生成功能使用指南。该模型可通过 Gemini 应用、Google Flow 等平台体验。指南包含五项提示词技巧：利用模型已有的现实世界知识进行简洁描述；精确控制文本在视频中的渲染与排版；使用专业镜头指令（如推拉摇移）像电影摄影师一样调度画面；通过迭代编辑高效修改视频；以及在生成中直接调整角色的动作节奏或情绪。其核心在于通过精准的提示词

能力进展 新发布

https://x.com/GoogleAI/status/2059381218660270435

---

9. 我们如何对不同产品中的Claude进行隔离控制

Anthropic：Engineering（事故复盘 + 工程实践 · 网页） · 昨天 02:11

Anthropic通过三重机制控制Claude智能体的部署风险，包括用户误用、模型异常行为和外部攻击。其防护策略聚焦于三个层面：通过沙箱、虚拟机和网络出口控制限制智能体运行环境；利用系统提示词和模型训练引导其行为；以及对MCP服务器、第三方插件等外部内容实施细粒度权限管理。文章以Claude Code、claude.ai和Claude Cowork为例，阐述了不同产品如何设计对应的隔离架构。

能力进展 基础设施

https://www.anthropic.com/engineering/how-we-contain-claude

---

10. 未来展望：2026年5月的一些想法

Nathan Lambert：Interconnects（RSS） · 昨天 23:39

文章展望了截至2026年5月AI领域的动态。内容涉及 Gemini Flash 3.5 的发布、名为 Mythos 的新产品或项目、开源与闭源生态平衡（open-closed balance）的讨论、美国开源力量的显著增长（America's open-source surge），以及由此引发的新兴权力博弈（emerging power struggles）。

能力进展 新发布

https://www.interconnects.ai/p/some-ideas-for-what-comes-next-may

---

11. 使用 Codex 构建自改进税务智能体

OpenAI：官网动态（RSS · 排除企业/客户案例） · 17 小时前

OpenAI、Thrive 与 Crete 合作，使用 Codex 构建了一个自改进的税务智能体。该智能体能够自动处理报税流程，提升工作准确性并加速整体工作流。

能力进展 新发布

https://openai.com/index/building-self-improving-tax-agents-with-codex

---

12. 用好 Coding Agent，重点是两头，尤其是开头的部分，如果一开始就走偏了后面怎么改都改不好。

X：宝玉 (@dotey) · 1 小时前

用好 Coding Agent 的关键在于初始规划。方法是先将需求整理后，用最强模型（如 GPT-5.5、Claude Opus 4.7）分别在 Codex、Claude Code、Cursor 的 Plan 模式下生成设计方案，选择最优方案并借鉴其他版本。对于复杂计划，可将其拆分为多个 Phases 并明确要求与验证标准，形成 Markdown 文档。执行时按 Phases 进行，并辅以人工审核

能力进展

https://x.com/dotey/status/2059773942500298934

---

13. 软件之后是AI时代

Tomer Tunguz 博客（VC 分析） · 昨天 08:00

软件时代正过渡至"智能体框架"时代。AI作为强大但需驯化的"野马"，其智能驯化包含七个核心组成部分：上下文与记忆、工具与行动、编排与循环、状态与持久性、沙箱与计算、可观测性与治理、成本与工作流优化。这些组件共同构成了一个生产级的智能体系统。这一转变将重塑软件竞争格局，模型通用化的未来中，最佳的智能体驾驭者将获胜。

能力进展

https://www.tomtunguz.com/harnessing-ai

---

14. Project Luxo：跨越AI媒体的恐怖谷

Runway：News（网页） · 昨天 22:34

Runway通过Project Luxo研究发现，AI生成视频已跨越"恐怖谷"。他们向创意生态从业者展示了《The Rogue》等AI短片及广告样片，评估显示观众开始关注故事本身，而非技术瑕疵。所有作品均由单人团队制作，耗时从3周到4小时不等。Runway认为，这标志着AI媒体成熟--当技术足够好以至于"隐形"，观众沉浸于故事时，便实现了这一跨越。

能力进展

https://runwayml.com/news/project-luxo

---

15. 与Google搜索产品副总裁Robby Stein的访谈：AI原生搜索时代

X：Kim (@kimmonismus) · 8 小时前

本文记录了与Google搜索产品副总裁Robby Stein在Google I/O的访谈，核心探讨Google Search向"AI原生"模式的重大转变。讨论话题包括AI Mode是进化还是重塑、如何将复杂问题拆解为多轮搜索、AI搜索的高运行成本、Google TPU及基础设施的优势、AI时代搜索量不减反增的原因，以及优质AI回答与出版商流量之间的张力。访谈还涉及Google决定展示哪些信息源与链

基础设施

https://x.com/kimmonismus/status/2059668961181004275

---

16. 藏师傅发布小红书图文排版AI Skill，集成地图与自动配图

X：歸藏 (@op7418) · 13 小时前

该推文介绍了guizang-social-card-skill，一款针对小红书图文常见类别进行优化的AI Skill。其亮点在于为旅行博主集成了地图组件，用户输入目的地和线路后，AI能自动在底图上标记并嵌入图片。根据引用，该Skill完全基于HTML和实拍图片生成内容，不会被平台标注为AI生成，并会主动从高质量图片网站寻找对应主题图片，以优化图文排版。

新发布

https://x.com/op7418/status/2059587983289016348

---

17. 人类与AI分工：教育咨询及文学奖争议

X：Ethan Mollick (@emollick) · 昨天 03:59

我写了一篇新文章，探讨我们需要保留哪些人类特质，以及哪些可以交给AI，其中涉及教育、咨询领域的实验，以及最近关于文学奖的争议。

https://x.com/emollick/status/2059363865536668040

---

18. 选择保持人性

Ethan Mollick：One Useful Thing（RSS） · 昨天 03:56

社交媒体平台上的帖子内容正变得越来越相似。这种趋同现象可能意味着大量内容正在被AI生成或同质化处理，引发了人们对于内容原创性与人类独特视角的讨论。

https://www.oneusefulthing.org/p/choosing-to-stay-human
