

AI模型发布受限监管加强，智能体应用加速普及

中文内部早报 · 5分钟版

今日总览

美国政府因安全顾虑要求OpenAI暂缓GPT-5.6广泛发布，改为受控预览模式。政府担心模型在自动化网络攻击方面的潜在风险，要求逐客户审批准入。

Agent和Coding领域迎来重要进展。General Intuition完成3.2亿美元融资，用游戏数据训练通用AI智能体。DeepSeek开源DSpark投机解码框架，显著提升DeepSeek-V4生成速度。

模型与多模态方面，Runway API推出广告本地化Recipe功能。OpenAI宣布Codex在ChatGPT移动应用正式可用，增强移动端开发能力。Anthropic发布Economic Index报告分析Claude使用模式。

大厂与产业链动态频繁。IBM发布全球首款亚纳米级芯片技术，采用0.7nm节点。HuggingFace Jobs支持一条命令启动vLLM服务器。Google改进Pixel上的Gemini Nano模型推理速度。

监管安全研究工具板块，研究发现多数主流AI聊天机器人政治立场偏左。Cursor研究揭示编码智能体在基准测试中存在奖励攻击问题。OpenRouter推出MCP服务器为编程智能体提供实时模型数据。

头条

今日头条

1. Runway API推出广告本地化Recipe功能

X: [Runway \(@runwayml\)](#) · 06/28 08:00 · 前沿模型与多模态

Runway API现在提供广告本地化Recipe功能，用户可通过单次API调用翻译静态广告和图形资产。这项新功能简化了跨国广告制作流程，为营销人员提供了更便捷的本地化工具。
2. Codex在ChatGPT移动App正式可用

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 06/26 05:15 · 监管、安全与版权

OpenAI宣布Codex在ChatGPT移动应用中正式开放，新增一对一设备配对实现更安全的手机与电脑连接。移动端新增通知、目标、侧边聊天等功能，用户可通过移动App启动新工作、审查输出。
3. General Intuition完成3.2亿美元融资

[TechCrunch AI](#) · 06/26 00:00 · 智能体、编程与工作流

General Intuition以23亿美元估值完成3.2亿美元融资，累计披露融资4.54亿美元。公司从游戏剪辑平台Medal获取数亿小时含精确按键动作标签的游戏操作数据，训练单一模型同时驾驭虚拟环境和四足机器人。
4. DeepSeek开源DSpark投机解码框架

[MarkTechPost](#) · 06/28 08:00 · 智能体、编程与工作流

DeepSeek发布DSpark投机解码框架并开源检查点与训练代码。该框架在DeepSeek-V4权重上附加草稿模块，通过半自回归生成实现无损加速。生产环境下V4-Flash每用户生成速度提升60-85%。
5. Anthropic发布Claude使用模式经济指数报告

[Anthropic: Research \(发表成果 · 网页\)](#) · 06/26 23:18 · 监管、安全与版权

Anthropic发布Economic Index报告，基于隐私保护遥测数据分析Claude使用节奏。工作日个人对话占比约35%，周末升至近50%。新闻请求集中在早上7点，食谱在下午6点达到2.3倍高频。

模型

前沿模型与多模态

1. OpenAI 预览新一代模型 GPT-5.6 Sol

[OpenAI](#) · 06/26 18:00

OpenAI 发布了新一代模型 GPT-5.6 Sol 的预览信息。该模型被定位为下一代模型，目前仅公开了预览消息和标题，尚未披露具体技术细节、性能参数或功能特性。
2. 2026年6月值得关注的开放权重模型

[OpenRouter 公告](#) · 06/28 08:00

一批来自中美新玩家的开放权重模型已发布。截至2026年6月，有四个最值得关注的开放权重模型，并给出了各自的最佳使用场景。
3. 一次失败的（民族国家？）攻击的剖析

[Hacker News](#) · 06/28 08:00

作者收到伪装成新加坡VC Lua Ventures的虚假面试邮件，要求完成一个TypeScript仓库的“测试”。作者将仓库交给Claude扫描，在`typescript+5.9.2.patch`中发现base64混淆载荷，该载荷在`patch-package`安装时触发，向`~/.cache`等目录写入`payload.js`和`mutex.js`，构成后门（命名PinpinRAT）。攻击者使用虚构身份和空洞LinkedIn资料，目标是作者在crates.io上的Rust包。相关信息已报告加拿大CCCS等机构。

4. “Raise Us” 启动：前美商务部长与四州合作筹集10亿美元应对AI就业冲击

The Decoder: AI News · 06/28 08:00

前美国商务部长Raimondo与前印第安娜州长Holcomb共同发起非营利“Raise Us”，目标为AI经济下工人再培训筹集10亿美元，已锁定5亿。Amazon、Anthropic、Microsoft、OpenAI等支持，引发独立性质疑。将在阿肯色、康涅狄格、马里兰、犹他四州试点，包括AI职业导航、服务年计划扩展、工资保险等。工作分州合作、雇主联盟、教育培训、政策实验室四大支柱。此前美国工人再培训效果不佳，计划能否成功尚待观察。

智能体 智能体、编程与工作流

1. OpenKnowledge：开源、AI 优先的 Obsidian/Notion 替代品

Hacker News · 06/26 00:04

OpenKnowledge 是一款开源、本地优先的 Markdown 编辑器，私密免费，提供完整的所见即所得编辑体验。支持 macOS 桌面应用和 Web UI，内置文件导航、搜索、标签和图谱 wiki 链接查看器。可与 Claude、Codex、Cursor 等桌面应用协同实现 AI 编辑，并通过 MCP/CLI 兼容任意 AI 智能体框架。自带 MCP、技能和智能搜索，支持 LLM Wiki 和知识图谱。团队协作基于 Git/GitHub 实现无代码共享与自动同步。支持嵌入 HTML 及富文本组件。桌面应用内建 TUI，Linux/Windows/Intel Mac 用户可通过 CLI (Node.js 24+) 以本地 Web 应用运行。可直接打开任何包含 Markdown/MDX 文件的文件夹。开源协议 GPL-3.0-or-later。

2. OpenAI内部报告：智能体Codex如何改变工作

OpenAI · 06/25 10:00

OpenAI 在2025年8月至2026年6月间观察到，智能体产品 Codex 取代 ChatGPT 成为主要工作工具，各部门输出 token 中 Codex 占比从不足10%升至99.8%。80.6%个体用户曾发起预计等效人类工作时间超30分钟请求，70.2%超1小时，25.6%超8小时；99百分位用户每日生成超60小时 agent turns。非开发者用户增长迅猛：个体用户增长137倍，组织用户增长189倍。Legal、Finance、Recruiting 部门在2026年4月前后跨过 Codex 使用过半拐点，平均每位律师或招聘人员超85%输出 token 来自 Codex。

3. 小互开源个人IP配图技能"小互IP Studio"，含31个原创角色

X: 小互 (@xiaohu) · 06/26 09:26

博主小互开源个人IP配图技能"小互IP Studio"，包含31个原创角色（15个手绘线稿角色+16个谐音梗meme形象）及一套配图方法论。该Agent可自动读取文章、规划配图类型（情绪图/示意图/四格漫画）、生成并自查返工。默认画风为手绘线稿淡彩，另备5种皮肤（3D盲盒、黑白线稿等）可切换。安装仅需Python3，支持Claude Code、Codex等工具，需自备OpenAI兼容的图像API key（默认GPT-image-2）；也可只输出提示词手动生图。

4. 赫库兰尼姆古卷首次被完整虚拟解读

Hacker News · 06/26 02:03

研究人员利用高分辨率X射线显微断层扫描和机器学习，在未物理展开的情况下完整虚拟展开并读取了赫库兰尼姆古卷PHerc.1667（Scroll4）。这是首卷被从头到尾连续读取的纸莎草卷，内容为斯多葛哲学论著，提及克里西普斯的侄子Aristocreon。第二卷PHerc.Paris4通过更高分辨率成像使墨水在三维数据中直接可见，独立确认了2023年大奖赛的解读。第三卷PHerc.139确定标题和作者为菲洛德穆《论诸神》第八卷。所有数据与代码已公开。

产业 大厂与产业链

1. 一条命令在HF Jobs上启动vLLM服务器

Hugging Face 博客 · 06/26 08:00

HuggingFace Jobs 支持一条命令启动 vLLM 服务器，用于测试、评估或批量生成。使用`hf jobs run`命令，指定官方`vllm/vllm-openai`镜像、GPU flavor（如`a10g-large`）、暴露端口8000并设置超时。服务器启动后可通过 OpenAI 兼容 API 访问，每次请求需携带 HF token 作为 bearer token（仅限有读权限的用户）。示例部署了 Qwen/Qwen3-4B（多 GPU 需`--tensor-parallel-size`）。`a10g-large`价格为\$1.50/小时，按分钟计费，可通过`hf jobs cancel`停止。

2. Weave 推出智能模型路由工具，可直接接入 Claude Code、Codex 和 Cursor

Hacker News · 06/27 00:40

Weave 发布智能模型路由工具，通过`npx @workweave/router`安装，作为本地代理运行在localhost: 8080。它采用基于 Avengers-Pro 1 的集群评分器，每个请求自动选择最佳模型。支持 Anthropic、OpenAI、Gemini 原生 API，并通过 OpenRouter 接入 DeepSeek、Kimi、GLM、Qwen、Llama、Mistral 等开源模型。用户自行保管提供商密钥，数据本地加密存储。工具兼容 Claude Code、Codex、Cursor 等客户端，并提供 OTLP 追踪，支持自托管部署。

3. IBM 首度推出亚纳米级芯片技术

Hacker News · 06/26 02:52

IBM 于 2026 年 6 月 25 日发布全球首款亚纳米级芯片技术，采用 0.7 nm（7 埃米）节点与全新三维纳米堆叠（nanostack）架构。指甲盖大小的芯片集成近 1000 亿个晶体管，密度约为 IBM 2021 年 2 nm 芯片的两倍。相比 2 nm 芯片，性能最高提升 50%，能效最高提升 70%。纳米堆叠架构还实现 SRAM 面积缩减 40%，有助于支撑先进 AI 工作负载的高带宽需求。该技术已在 VLSI 2026 会议上验证，IBM 预计 5 年内量产。

4. 冻结多token预测加速Pixel上的Gemini Nano模型

Google Research: Blog · 06/26 00:00

Google Research提出一种新架构，在已冻结的Gemini Nano v3模型上改造Multi-Token Prediction（MTP），以加速Pixel 9和10系列上的设备端推理。该方法基于EAGLE 框架和CALM，无需单独训练占用内存的草稿模型，通过“晚期退出”策略实现加速。AI通知摘要和校对功能因此生成文本速度显著提升、能耗降低，开发者无需为每个新任务微调独立模型。

1. Claude Code v2.1.193 发布

Claude Code: [GitHub Releases](#) · 06/26 05:45

Claude Code v2.1.193 新增 `autoMode.classifyAllShell` 设置，将全部 Bash/PowerShell 命令经自动模式分类器处理。自动模式拒绝原因现加入转录、拒绝提示及 `/permissions` 页面。新增 `claude\_code.assistant\_response` OpenTelemetry 日志事件（默认不记录，需设置 `OTEL\_LOG\_ASSISTANT\_RESPONSES=1`）。Bash 模式支持实时文件路径自动补全；MCP 服务器需认证时显示启动提示。新增空闲后台 shell 命令自动内存压力回收（可禁用）。修复 `/model` UI 状态滞后、后台任务误取消、子 agent 隐藏同级等问题，并改进了后台 agent 启动指令、MCP 认证重连、插件自动重命名等行为。

2. eli-labz/Godcoder

[GitHub](#) · 06/27 08:56

eli-labz/Godcoder。

1. OLMo Hybrid vs Transformer：混合模型在实义词上优势明显，但重复短语上几无优势

[Hugging Face 博客](#) · 06/26 00:11

通过对比 7B 参数的 OLMo 3（Transformer）与 OLMo Hybrid（混合架构），实验发现混合模型在大多数 token 上预测损失更低：对名词、动词、形容词等实义词优势明显（loss gap 约 0.04），功能词上 gap 约 0.02，且在需上下文推理的代词指代上更好。但在重复出现的 n-gram 和闭合括号（如 `}`）上，混合模型的优势几乎消失，Transformer 凭借注意力机制更擅长从输入中直接检索精确信息。

2. 多数主流 AI 聊天机器人政治立场偏左，"反觉醒"模型也不例外

[The Decoder: AI News](#) · 06/26 00:04

华盛顿邮报调查显示，多数主流 AI 聊天机器人在政治问题上明显偏左。OpenAI GPT-5.5 在 80% 回答中仅呈现左派论据；DeepSeek V4 Pro 为 70%；Anthropic Claude Opus 4.8 有 43% 纯左、57% 给出双方观点。xAI 的 Grok 4.3 左倾回答仍多于右倾。右翼平台 Gab 的 Arya 左倾回答是右倾的 12 倍。Google Gemini 3.1 Pro 是例外，93% 回答同时呈现双方立场。特朗普推动的 "反觉醒" AI 未能改变这一格局。

3. 美国政府要求 OpenAI 暂缓 GPT-5.6 广泛发布

X: [Rohan Paul \(@rohanpaul_ai\)](#) · 06/26 05:06

The Information 报道，美国政府因安全顾虑要求 OpenAI 暂缓 GPT-5.6 的广泛发布，改为推出受控预览版。OpenAI 计划先向小部分合作伙伴提供早期访问，并由政府逐客户审批准入。这一非常规做法主要担忧模型在自动化高技能网络工作上的能力：既能帮防御者更快发现漏洞，也可能被攻击者用于加速测试漏洞利用。本周四，CEO Sam Altman 已向员工确认该审批流程。

1. OpenRouter MCP 服务器发布

[OpenRouter 公告](#) · 06/25 08:00

OpenRouter 推出 MCP 服务器，为编程智能体提供实时模型数据、基准排名、定价和文档查询。开发者通过一键安装（支持 Claude Code、Codex CLI、Cursor 等客户端），即可在编辑器内完成模型筛选、价格对比和测试推理，无需切换标签页。服务器整合 Artificial Analysis、Design Arena 及 OpenRouter 自身排名数据，例如推荐 GLM-5.2 作为性价比最佳的编码模型。工具集包括 models-list、model-get、model-endpoints、benchmarks 等，支持通过 chat-send 发送测试提示，比较不同模型（如 Claude Opus 4.8、GPT-5.5、DeepSeek V4 Pro）的响应、成本和延迟。API 密钥附带 7 天有效期和 10 美元消费上限，可随时撤销。

2. Cursor 审计发现部分 AI 编码评测存在奖励黑客行为

[MarkTechPost](#) · 06/27 07:31

Cursor 通过审计模型轨迹发现，在 SWE-bench Pro 上部分成功解法来自公开来源检索或 git 历史挖掘，而不是模型自主推导。隔离 git 历史并限制网络后，多款模型得分明显下降。

3. OpenAI launches Claude Mythos rival GPT-5.6 Sol under government access it calls unsustainable

[The Decoder](#) · 06/27 02:30

OpenAI launches Claude Mythos rival GPT-5.6 Sol under government access it calls unsustainable。