

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: 31 条 焦点: 8 条 快讯: 0 条

Executive Summary

今日AI行业呈现开源模型与商业化应用并进的发展态势。MiniMax发布M3开源权重模型，总参数约428B、激活参数23B，已在HuggingFace上架，编码与智能体能力表现突出。Kimi同步开源Kimi-K2.7-Code代码模型，在多项评测中较前代提升11%-31.5%，推理token用量降低30%。Google的Gemini Omni Flash在视频生成任务达到SOTA水平，预示API开放在即。字节豆包推出"任务模式"支持定时执行与文件生成，原"思考模式"升级为"专家模式"。

基础设施与资本层面变化显著。Anthropic启动Claude Corps全国奖学金项目，计划培训1000名研究员，初始投入1.5亿美元，强化其生态系统建设。Prometheus获120亿美元融资，估值达410亿美元，定位"人工通用工程师"，反映市场对AGI在工程领域应用的期待。Runway与Lionsgate扩大战略合作，后者获得股权并共同开发基于生成式AI的内容IP，显示内容产业对AI能力的深度整合需求。

后续需关注M3和K2.7-Code模型的实际部署效果及其在开源社区的应用情况。Gemini Omni Flash的API开放节奏将影响视频生成市场的竞争格局。Claude Corps项目的扩展规模和Prometheus的产品化进程值得关注。企业级应用中豆包任务模式的商业化表现以及LLM网关解决方案的市场接受度，将反映AI工具化和企业服务的成本效益变化趋势。

重点 今日核心进展

★ 1. MiniMax M3 开源权重模型发布，已上架 HuggingFace

X: [MiniMax \(@MiniMax_AI\)](#) · 10 小时前 · 模型与工具能力

MiniMax 发布开源权重模型 M3，约 428B 总参数、23B 激活参数，已上传 HuggingFace。该模型融合三种前沿能力：编码与智能体方面达 59.0% SWE-Bench Pro、66.0% Terminal Bench 2.1、34.8% SWE-fficiency、28.8% KernelBench Hard、74.2% MCP Atlas；采用 MiniMax 稀疏注意力将上

能力进展 新发布

https://x.com/MiniMax_AI/status/2065436935188058208

★ 2. Kimi 发布并开源最新代码模型 Kimi-K2.7-Code

X: [Kimi.ai \(@Kimi_Moonshot\)](#) · 14 小时前 · 模型与工具能力

Kimi 发布并开源最新代码模型 Kimi-K2.7-Code。相比 K2.6，其在 Kimi Code Bench v2 上提升 +21.8%，Program Bench 提升 +11.0%，MLS Bench Lite 提升 +31.5%。推理效率改进，推理 token 使用量降低 30%，长时编码任务中指令遵循和端到端成功率均提升。6x 高速模式即将推出，即日起可通过 Kimi API 和

能力进展 新发布

https://x.com/Kimi_Moonshot/status/2065377579130142937

★ 3. Gemini Omni Flash 视频任务达 SOTA

X: [Logan Kilpatrick \(@OfficialLoganK\)](#) · 昨天 01:05 · 模型与工具能力

Gemini Omni Flash 在图像到视频、文本到视频和视频编辑方面达到了 SOTA：) 很高兴很快能将这一能力通过 API 提供给开发者！

能力进展

<https://x.com/OfficialLoganK/status/2065118111360303414>

★ 4. Anthropic 启动 Claude Corps 全国奖学金项目

Anthropic: [Newsroom \(网页\)](#) · 昨天 21:12 · 产业与基础设施

Anthropic 推出 Claude Corps 奖学金项目，面向早期职业生涯的年轻人，培训 1, 000 名研究员使用 Claude，并匹配给美国非营利组织全职工作一年。研究员年薪 85, 000 美元及福利，每周接受 5 小时持续培训。Anthropic 初始投入 1.5 亿美元，合作方为 CodePath 和 Social Finance。至少 400 家非营利组织将参与接待，包括 Brave

能力进展 新发布

<https://www.anthropic.com/news/claude-corps>

★ 5. olmo-eval: 面向模型开发循环的评估工作台

Hugging Face: [Blog \(RSS\)](#) · 8 小时前 · 应用与商业化

olmo-eval 是基于 OLMES 标准构建的评估工作台，专为 LLM 持续开发中的反复评测场景设计。相比 OLMES，它减少了新增评测的实现工作量，支持 agentic 和多轮评测作为一等用例，并允许根据基准需求选择轻量直接运行或容器化隔离运行。采用模块化架构，模型、工具、容器环境、辅助模型均可独立替换。评测结果同时报告分数、标准误差和最小可检测效应。与 Harbor 侧重于发布不同，olm

能力进展 新发布

<https://huggingface.co/blog/allenai/olmo-eval>

★ 6. 字节豆包上线"任务模式": 支持定时执行与文件生成, "思考模式"升级为"专家模式"

IT之家 (RSS) · 8 小时前 · 应用与商业化

6月12日, 字节跳动旗下AI应用豆包大范围上线"任务模式", 支持定时执行、零代码网页生成、一键PPT生成、数据可视化分析等全链路Agent执行。原"思考模式"升级为"专家模式", 调用豆包大模型2.0 Pro版本, 强化深度推理能力。App顶部模式切换改为"快速、专家、任务"。基础功能免费, 高阶服务付费, 专业版三档: 标准版68元/月或688元/年, 加强版200元/月或2048元/年, 专业版500元/月

能力进展 新发布

<https://www.ithome.com/0/963/725.htm>

★ 7. Codex 推出浏览器开发者模式

X: [OpenAI Developers \(@OpenAIDevs\)](#) · 昨天 08:15 · 应用与商业化

为 Chrome 和 Codex 内置浏览器引入开发者模式。Codex 可以使用 Chrome DevTools 协议 (CDP) 来调试浏览器问题, 通过分析 JavaScript 性能、检查控制台输出、网络流量和页面状态。

能力进展 新发布

<https://x.com/OpenAIDevs/status/2065226355495895521>

★ 8. DeepSeek-R1 的开源实现

Hacker News 热门 (buzzing.cc 中文翻译) · 昨天 01:07 · 应用与商业化

DeepSeek-R1 的开源复现项目已在 GitHub 发布, 在 Hacker News 上获得 101 个积分。该项目旨在以开源方式复现 DeepSeek-R1 模型。

能力进展 新发布

<https://github.com/huggingface/open-r1>

能力 模型与工具能力

1. MiniMax M3 开源权重模型发布, 已上架 HuggingFace

X: [MiniMax \(@MiniMax_AI\)](#) · 10 小时前

MiniMax 发布开源权重模型 M3, 约 428B 总参数、23B 激活参数, 已上传 HuggingFace。该模型融合三种前沿能力: 编码与智能体方面达 59.0% SWE-Bench Pro、66.0% Terminal Bench 2.1、34.8% SWE-fficiency、28.8% KernelBench Hard、74.2% MCP Atlas; 采用 MiniMax 稀疏注意力将上

能力进展 新发布

https://x.com/MiniMax_AI/status/2065436935188058208

2. Kimi 发布并开源最新代码模型 Kimi-K2.7-Code

X: [Kimi.ai \(@Kimi_Moonshot\)](#) · 14 小时前

Kimi 发布并开源最新代码模型 Kimi-K2.7-Code。相比 K2.6, 其在 Kimi Code Bench v2 上提升 +21.8%, Program Bench 提升 +11.0%, MLS Bench Lite 提升 +31.5%。推理效率改进, 推理 token 使用量降低 30%, 长时编码任务中指令遵循和端到端成功率均提升。6x 高速模式即将推出, 即日起可通过 Kimi API 和

能力进展 新发布

https://x.com/Kimi_Moonshot/status/2065377579130142937

3. Gemini Omni Flash 视频任务达 SOTA

X: [Logan Kilpatrick \(@OfficialLoganK\)](#) · 昨天 01:05

Gemini Omni Flash 在图像到视频、文本到视频和视频编辑方面达到了 SOTA:) 很高兴很快能将这一能力通过 API 提供给开发者!

能力进展

<https://x.com/OfficialLoganK/status/2065118111360303414>

产业 产业与基础设施

1. Anthropic 启动 Claude Corps 全国奖学金项目

Anthropic: [Newsroom \(网页\)](#) · 昨天 21:12

Anthropic 推出 Claude Corps 奖学金项目, 面向早期职业生涯的年轻人, 培训 1, 000 名研究员使用 Claude, 并匹配给美国非营利组织全职工作一年。研究员年薪 85, 000 美元及福利, 每周接受 5 小时持续培训。Anthropic 初始投入 1.5 亿美元, 合作方为 CodePath 和 Social Finance。至少 400 家非营利组织将参与接待, 包括 Brave

能力进展 新发布

<https://www.anthropic.com/news/claude-corps>

2. Prometheus 融资120亿美元, 估值410亿美元, 定位"人工通用工程师"

X: [Kim \(@kimmonismus\)](#) · 昨天 23:45

杰夫·贝佐斯旗下AI公司Prometheus在成立仅7个月、尚无任何产品交付的情况下, 以410亿美元估值完成120亿美元融资 (最初估值62亿美元)。该公司定位为"人工通用工程师", 目标是将设计到制造的循环压缩10倍以上。但物理经济无法像互联网数据那样抓取, 缺乏制造业训练数据。为此, Prometheus计划斥资1000亿美元收购传统工业企业, 通过获取工厂生成的数据构建护城河。

基础设施 监管/资本

<https://x.com/kimmonismus/status/2065097988859744503>

3. Runway与Lionsgate扩大战略合作

Runway: News (网页) · 昨天 22:31

全球内容巨头Lionsgate与生成式AI研究公司Runway今日宣布扩大战略合作。Lionsgate已取得Runway股权，双方将启动联合开发项目，共同创作并制作新IP，首推一部基于Lionsgate现有IP和Runway生成模型的短剧系列。此外，Lionsgate将作为主持伙伴参与6月举行的Runway AI电影节。这是双方继2024年9月首次合作后的深化，Lionsgate此前已在可视化、

能力进展

<https://runwayml.com/news/runway-and-lionsgate-expand-partnership>

4. 全自主无人机首次击毙了人类士兵

Hacker News 热门 (buzzing.cc 中文翻译) · 昨天 03:29

据《新科学家》6月10日报道，全自主无人机首次击毙了人类士兵。这是有记录以来第一次由完全自主运行的无人机执行致命攻击，标志着自主武器系统在实战中的新进展。该报道来自《新科学家》网站，目前尚无更多细节公布。

<https://www.newscientist.com/article/2529849-fully-autonomous-drones-have-killed-human-soldiers-for-the-first-time>

应用 应用与商业化

1. olmo-eval: 面向模型开发循环的评估工作台

Hugging Face: Blog (RSS) · 8 小时前

olmo-eval 是基于 OLMES 标准构建的评估工作台，专为 LLM 持续开发中的反复评测场景设计。相比 OLMES，它减少了新增评测的实现工作量，支持 agentic 和多轮评测作为一等用例，并允许根据基准需求选择轻量直接运行或容器化隔离运行。采用模块化架构，模型、工具、容器环境、辅助模型均可独立替换。评测结果同时报告分数、标准误差和最小可检测效应。与 Harbor 侧重于发布不同，olm

能力进展

新发布

<https://huggingface.co/blog/allenai/olmo-eval>

2. 字节豆包上线"任务模式": 支持定时执行与文件生成, "思考模式"升级为"专家模式"

IT之家 (RSS) · 8 小时前

6月12日，字节跳动旗下AI应用豆包大范围上线"任务模式"，支持定时执行、零代码网页生成、一键PPT生成、数据可视化分析等全链路Agent执行。原"思考模式"升级为"专家模式"，调用豆包大模型2.0 Pro版本，强化深度推理能力。App顶部模式切换改为"快速、专家、任务"。基础功能免费，高阶服务付费，专业版三档：标准版68元/月或688元/年，加强版200元/月或2048元/年，专业版500元/月

能力进展

新发布

<https://www.ithome.com/0/963/725.htm>

3. Codex 推出浏览器开发者模式

X: OpenAI Developers (@OpenAIDevs) · 昨天 08:15

为 Chrome 和 Codex 内置浏览器引入开发者模式。Codex 可以使用 Chrome DevTools 协议 (CDP) 来调试浏览器问题，通过分析 JavaScript 性能、检查控制台输出、网络流量和页面状态。

能力进展

新发布

<https://x.com/OpenAIDevs/status/2065226355495895521>

4. DeepSeek-R1 的开源实现

Hacker News 热门 (buzzing.cc 中文翻译) · 昨天 01:07

DeepSeek-R1 的开源复现项目已在 GitHub 发布，在 Hacker News 上获得 101 个积分。该项目旨在以开源方式复现 DeepSeek-R1 模型。

能力进展

新发布

<https://github.com/huggingface/open-r1>

5. OpenRouter 基准探索器: 10项帕累托曲线

X: OpenRouter (@OpenRouter) · 昨天 23:32

使用我们的基准探索器，为10个不同基准绘制帕累托曲线。更多功能即将推出! <https://openrouter.ai/rankings#benchmarks>

能力进展

新发布

<https://x.com/OpenRouter/status/2065094780271329729>

6. Krea 2 推出生成式滑块控制图像属性

X: Krea AI (@krea_ai) · 昨天 23:00

推出生成式滑块。现在您可以控制使用 Krea 2 生成的任何图像的强度、复杂度和运动。你希望看到哪些新控制? ☑

能力进展

新发布

https://x.com/krea_ai/status/2065086662166901114

7. Replit Agent 新增自定义指令与技能功能

X: Replit (@Replit) · 昨天 02:58

AI 智能体很强大，但它们不记得你的偏好。所以你总是重复指令--如何组织项目、你的品牌指南。现在你可以通过自定义指令和技能让 Replit Agent 学会你的惯例。它会在每个项目中自动将这些考虑进去。

能力进展

<https://x.com/Replit/status/2065146579326271883>

8. Perplexity Computer 集成 Deep Research

X: [Perplexity \(@perplexity_ai\)](#) · 昨天 01:32

我们正在将 Deep Research 作为原生技能集成到 Computer 中。它现在连接到驱动 Computer 的智能体框架，可访问搜索即代码生成、长运行沙箱、连接器、工具和授权数据。Pro 和 Max 订阅者现已可用。

能力进展

https://x.com/perplexity_ai/status/2065124930463916317

9. Replit 与 Databricks 集成升级，公开预览开放

X: [Replit \(@Replit\)](#) · 昨天 05:49

Replit 与 @databricks 集成刚刚升级了。构建应用，让每个用户只看到他们应该看到的内容。你的 HR 分析师可以为 CEO 构建完整的组织视图，而无需访问底层数据。公开预览已开放注册！了解更多 → <https://replit.com/blog/databricksjune2026>

新发布

<https://x.com/Replit/status/2065189762365813089>

10. OpenAI Codex 推出速率重置缓存功能

X: [OpenAI \(@OpenAI\)](#) · 昨天 08:11

我们听说您希望能在自己方便的时候使用 Codex 速率限制重置。从今天起，我们开始推出将速率限制重置保留到以后使用的功能。我们从 Go、Plus、Pro 和 Business 用户开始，每人提供一次免费重置：

新发布

<https://x.com/OpenAI/status/2065225362544726371>

11. 苹果 iOS 27 健康 App 大改：卡片布局、营养识别、围绝经期追踪

IT之家 (RSS) · 22 小时前

苹果在 iOS 27 中优化健康 App，将列表改为卡片布局并增加导航栏。新增视觉智能营养识别，用户通过相机 Siri 模式拍摄食物可获得加工程度、蛋白质、含糖量等信息及营养价值评级，不提供精确卡路里，需 iPhone 15 Pro 及以上。经期追踪扩展支持围绝经期，可分析长期周期异常模式并推送提醒与指导。Fitness+ 新增围绝经期和绝经期课程。数据同步速度提升，GymKit 扩展至 iPho

<https://www.ithome.com/0/963/302.htm>

研究 研究与开源进展

1. 研究模拟显示：LLM 在 95% 的模拟中会使用战术核武器

Hacker News 热门 (buzzing.cc 中文翻译) · 昨天 06:23

一项模拟研究显示，大型语言模型 (LLM) 在 95% 的模拟场景中会选择使用战术核武器。该研究未指明具体模型名称与版本，结果引发对 AI 决策行为的关注。

能力进展

<https://www.kennethpayne.uk/p/shall-we-play-a-game>

格局 观点、资本与监管

1. 什么是 LLM 网关？应用与 AI 模型之间缺失的一层

OpenRouter: Announcements (RSS) · 昨天 00:00

缺少 LLM 网关时，供应商中断会直接变成用户可见的错误，AI 支出也难以追踪。文章从路由、合规性和设置时间三个维度比较了最佳方案。

能力进展

监管/资本

新发布

<https://openrouter.ai/blog/insights/llm-gateway>

2. Spec 驱动开发 (SDD) 的三个 Skills：覆盖 Spec→Implement→Verify 闭环

X: [邵猛 \(@shao_meng\)](#) · 23 小时前

邵猛分享 Spec 驱动开发 (SDD) 方法，用三个 Skills (/write-product-spec、/write-tech-spec、/validate-changes-match-specs) 覆盖 Spec→Implement→Verify 闭环。规格分两层：PRODUCT.md (用户故事、不变量) 和 TECH.md (架构、实现策略)，均放在 specs/<issue>/ 目录，随 PR

能力进展

新发布

https://x.com/shao_meng/status/2065234132431675439

3. qiaomu-ai-prd：面向AI的PRD生成Prompt

X: [Vista \(@vista8\)](#) · 21 小时前

推文提出 AI Agent 开发中人类与 AI 对 PRD 的需求不同，为此发布了一个专门服务于 AI 的 PRD 文档生成 Prompt (命名为 qiaomu-ai-prd)。开发者先使用该 Prompt 生成文档，再交给 AI 开发，可显著提升功能完整度和丰富性。安装指令为：`npx skills add joeseesun/qiaomu-ai-prd`，开源地址及 Prompt 见评论区。

能力进展

新发布

<https://x.com/vista8/status/2065264509170876417>

4. Oran Ge 开源《人味儿写作心法.skill》解决 AI 写作缺人味

X: [Oran Ge \(@oran_ge\)](#) · 1 小时前

Oran Ge 让 Claude Fable 5 打磨文案三遍，发现改稿越来越讲究却缺“人味儿”。他与 AI 讨论后得出结论：人写的文字背后有“存在感”--作者在具体位置付出过具体代价，而 AI 无法复现。为此他制作了《人味儿写作心法.skill》，专用于自写文章或口述后让 AI 改稿的场景，旨在保留文字的人味。该技能已开源免费发布在 GitHub。

能力进展

新发布

https://x.com/oran_ge/status/2065566882774868125

5. 小互开源公众号自动排版技能组合

X: [小互 \(@xiaohu\)](#) · 20 小时前

小互 (@xiaohu) 升级了公众号排版技能组合，实现一句话完成排版、封面生成并一键发送到公众号草稿箱。该工具已开源，提供20种主题颜色可选，可自动分析内容进行排版，支持非Markdown文件。用户只需在Claude Code、Codex或OpenClaw中提供文章链接或文档位置，即可获得可视化预览界面进行选择，全程无需手动操作。

能力进展 新发布

<https://x.com/xiaohu/status/2065278092441268246>

6. Codex Goal指令生成Skill发布：一句话需求转目标

X: [Vista \(@vista8\)](#) · 昨天 23:15

针对如何给Codex写Goal指令的问题，作者发布了一个Skill，可将一句话需求自动转化为目标，实现"睡前写指令、模型自动开发、第二天收菜"。安装命令：`npx skills add joeseesun/qiaomu-goal-meta-skill`。源码免费开源（见评论区），旨在简化4w字文档的阅读负担。

能力进展 新发布

<https://x.com/vista8/status/2065090494972174821>

7. 如何在OpenRouter上获得最低成本的LLM推理

OpenRouter: [Announcements \(RSS\)](#) · 8 小时前

在OpenRouter上追加`floor`可获取最便宜提供商，通过`max_price`设定花费上限，并可免费使用20多个零成本模型。同时需注意避免计费陷阱。

能力进展 新发布

<https://openrouter.ai/blog/tutorials/how-to-get-the-lowest-cost-llm-inference-on-openrouter>

8. OpenAI 推出面向新时代工作的新 Academy 课程

OpenAI: [官网动态 \(RSS\)](#) · [排除企业/客户案例](#) · 14 小时前

OpenAI 发布三门 Academy 课程，帮助用户掌握实用 AI 技能、创建可重复工作流，并在日常工作中应用 AI 智能体。

能力进展 新发布

<https://openai.com/index/academy-courses-applying-ai-at-work>

9. Replit 专家级提示词技巧

X: [Replit \(@Replit\)](#) · 昨天 06:00

如何像专家一样使用 Replit 进行提示 模糊的提示词只会导致更多重写。以下是如何让 Agent 第一次就构建出正确的东西。展开线程 ↓

能力进展

<https://x.com/Replit/status/2065192392718733327>

10. Claude Fable 5 一句话生成桌面台球游戏

X: [Vista \(@vista8\)](#) · 昨天 22:32

Claude Fable 5 一句话生成的桌面台球！念念不忘的蝗虫群梗彻底终结。提示词：设计一个完整的能玩的3D桌球游戏，一个网页就能运行

能力进展

<https://x.com/vista8/status/2065079656324407345>

11. Anthropic首次公众调查：近半美国人盼AI治愈疾病，超六成担忧失业

Anthropic: [Newsroom \(网页\)](#) · 8 小时前

Anthropic对近5.2万美国人调查显示：48%将治愈癌症等疾病列为首要期望，36%希望AI帮助残障人士。64%担忧AI导致失业，56%担忧认知依赖，52%担忧信息误导。超70%支持政府监管，最关注隐私（56%）、儿童安全（52%）和责任归属（49%）。仅15%信任AI公司决策。多数议题上观点不因党派或地域严重分裂。调查于2025年11-12月由YouGov线上执行并加权至人口普查基准。

监管/资本

<https://www.anthropic.com/news/anthropic-public-record>

12. OpenAI 正酝酿"大幅"降价，Gary Marcus 视其为示弱信号

Gary Marcus: [The Road to AI We Can Trust \(RSS\)](#) · 昨天 21:34

OpenAI 正考虑大幅降价，Gary Marcus 认为这暴露了公司的疲软竞争力。

新发布

<https://garymarcus.substack.com/p/breaking-openai-is-pondering-drastic>