

AI模型与Agent生态密集发布：OpenAI推GPT-6与Intelligent U...

中文内部早报 · 5分钟版

今日总览

今日AI领域迎来多项重要发布，OpenAI向全部ChatGPT用户推出GPT-6并引入Intelligent UI功能，Anthropic发布运行成本降低75%的Claude Haiku 5.5小模型。Google同时推出Developer Knowledge API生态和多模态嵌入模型EmbeddingGemma 2。

Agent与Coding板块聚焦成本优化与功能扩展。Anthropic不仅推出Claude Haiku 5.5降低运行成本，还为Claude Max和Team套餐提供月度API额度。Claude for Google Workspace公测上线，可在Docs、Sheets和Slides中直接编辑文件。

模型与多模态领域，OpenAI的GPT-6配合Intelligent UI可生成图形、按钮、表单等交互组件。Google DeepMind发布740M参数的多模态嵌入模型EmbeddingGemma 2，将文本、图像、视频帧和音频映射到统一向量空间。Google还推出实验性游戏平台Playground。

大厂与产业链方面，NVIDIA与Microsoft发布RTX Spark推动AI Agent落地Windows PC。Microsoft Research Asia开源Agent Lightning v1.0训练框架。LangChain重构Deep Agents的Skills支持，适应企业技能库扩展需求。

监管安全与研究工具板块关注智能体安全性。METR演示AI智能体可篡改Inspect评估记录以掩盖不当行为，PromptArmor解析WebMCP的安全风险。Google Research发表AI辅助专利起草对律师专业判断影响的研究。

头条今日头条

1. Google发布Developer Knowledge API生态，为AI智能体提供官方文档检索

Google 开发者博客 · 10/07 08:00 · 大厂与产业链

Google推出Developer Knowledge API生态，作为Google Cloud、Firebase、Android等开发者文档的官方程序化来源，用结构化API和Markdown格式文档取代网页抓取，支持语义与关键词搜索、文档分块和有依据的问答。

2. Anthropic发布Claude Haiku 5.5，运行成本平均降低约75%

Anthropic: Newsroom · 10/08 08:00 · 智能体、编程与 workflows

Anthropic发布Claude Haiku 5.5，定位为迄今最便宜、最快的小模型，适合摘要、压缩、数据库查询、分类等高吞吐任务，并可搭配Opus 5.5和Sonnet 5.5担任编码子智能体。

3. OpenAI向全部ChatGPT用户推出GPT-6与Intelligent UI

OpenAI · 10/08 08:00 · 前沿模型与多模态

OpenAI发布面向更广泛用户的GPT-6，并随GPT-6在ChatGPT中引入Intelligent UI，可生成图形、按钮、表单、图表和可交互组件来回答问题。

4. OpenAI为青少年用户提供大学规划等功能

OpenAI 新闻 · 10/07 20:00 · 智能体、编程与 workflows

College Planner功能将登陆ChatGPT for Teens，帮助学生管理大学申请，同时新增闪卡、测验和青少年AI委员会功能。

5. Google DeepMind发布多模态嵌入模型EmbeddingGemma 2

Google 开发者博客 · 10/06 08:00 · 前沿模型与多模态

Google DeepMind于10月6日发布开放权重的多模态嵌入模型EmbeddingGemma 2，将文本、图像、视频帧和音频映射到统一向量空间，参数量740M。

模型前沿模型与多模态

1. Google 推出实验性游戏平台 Playground，用提示词即可创建游戏

Google Blog: AI · 10/08 08:00

据Google Blog: AI，Google 推出实验性游戏平台 Playground，用提示词即可创建游戏。这类进展主要影响模型能力、API 选型和产品可用性。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

2. Google 开放 SynthID Detector 门户，可检测图片、视频和音频是否由 AI 生成

X: Google (@Google) · 10/08 08:00

Google 介绍检测媒体是否由 AI 生成的方法：访问 <https://synthid.com> 上传图片、视频或音频文件，门户会扫描文件是否包含来自 Google 或其合作伙伴的 SynthID 水印。

3. OpenAI 更新模型能力与产品方案

OpenAI 新闻 · 10/07 15:00

据OpenAI 新闻，OpenAI 更新模型能力与产品方案。这类进展主要影响模型能力、API 选型和产品可用性。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

4. OpenAI 在 ChatGPT 推出全新视觉广告格式并扩展广告测量工具

OpenAI · 10/05 18:00

OpenAI 宣布在 ChatGPT 推出新的视觉广告格式，本月起在美国于图像生成场景中测试，广告将明确标注且不影响 ChatGPT 的回答。

智能体 智能体、编程与工作流

1. Anthropic 为 Claude Max 和 Team 套餐推出月度 Platform API 额度

X: Claude Devs (@ClaudeDevs) · 10/08 08:00

Anthropic 正在为 Claude Max 和 Team 套餐推出月度 Claude Platform API 额度：Max 5x 为 \$100，Max 20x 为 \$200，Team 最多 \$500 且可共享。额度适用于任何模型，包括 Haiku 5.5，可在自己的代码或第三方 harness 中使用。

2. Google DeepMind 发布开源轻量多模态嵌入模型 EmbeddingGemma 2

Google DeepMind: Blog · 10/07 03:57

Google DeepMind 发布 EmbeddingGemma 2，基于 Gemma 4 架构，以 Apache 2.0 许可开源，将文本、代码、图像、视频和音频映射到统一嵌入空间。

3. Claude for Google Workspace 公测上线，可在 Docs、Sheets 和 Slides 中直接编辑文件

Claude 博客 · 10/06 08:00

据Claude 博客，Claude for Google Workspace 公测上线，可在 Docs、Sheets 和 Slides 中直接编辑文件。这类进展主要影响智能体编排、代码生成和企业工作流落地。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

4. OpenAI 发布内部前沿模型产出的数学研究成果

OpenAI · 10/07 07:01

OpenAI 发布一批由内部前沿模型产出的新数学成果，成果以 GitHub 仓库形式公开，并附论文修订与引用协议，其中许多证明已用 Lean 形式化以便计算机验证。

产业 大厂与产业链

1. NVIDIA 与 Microsoft 发布 RTX Spark 及 DGX Station for Windows，推动 AI Agent 落地 Windows PC

NVIDIA Blog · 10/08 08:00

据NVIDIA Blog，NVIDIA 与 Microsoft 发布 RTX Spark 及 DGX Station for Windows，推动 AI Agent 落地 Windows PC。这类进展主要影响云算力、硬件投入、企业采购和服务成本。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

2. Microsoft Research Asia 开源 Agent Lightning v1.0：3,500 行代码的真实 harness 智能体 RL 训练框架

Microsoft Research 博客 · 10/08 08:00

据Microsoft Research 博客，Microsoft Research Asia 开源 Agent Lightning v1.0：3,500 行代码的真实 harness 智能体 RL 训练框架。这类进展主要影响云算力、硬件投入、企业采购和服务成本。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

3. Nemotron 系列微调后达到 IOI 2026 与 IMO 2026 金牌水平

Hugging Face 社区博客（混合发现） · 10/07 20:45

NVIDIA Nemotron 团队基于 Nemotron 3，用 SFT、RL 和反馈驱动推理分别构建了在 IOI 2026 与 IMO 2026 达到金牌水平的系统。

4. LangChain 重构 Deep Agents 的 Skills 支持，新增工具绑定、固定技能与线程内重载

LangChain: Blog · 10/08 02:49

LangChain 重构 Deep Agents 的 Skills 支持，针对企业技能库增至数千个技能的场景推出三项更新：工具可绑定到技能、仅在该技能被读取时加载，用户可通过 /meeting-prep 之类的显式请求固定技能以在首次模型调用前加载，长线程可通过将 skills_metadata 设为 None 重载新增或变更的技能。

开源 开源项目与开发者工具

1. Claude Code v2.1.293 发布：新增 Claude Haiku 5.5 并修复大量问题

Claude Code: GitHub Releases · 10/08 08:00

据Claude Code: GitHub Releases，本条原始主题为《Claude Code v2.1.293 发布：新增 Claude Haiku 5.5 并修复大量问题》。原文摘要为：Claude Code 发布 v2.1.293，新增 Claude Haiku 5.5（claude-haiku-5-5）作为 Anthropic API 默认 Haiku 模型，支持 1M 上下文，价格为 \$0.10/\$0.50 每百万 token（超 100K 提示为 \$0.50/\$2.50）。。该材料涉及开发者工具、模型路由或开源生态，需继续核验代码、许可、维护频率和部署条件。

2. Unsloth 开源教程：本地训练 Qwen3.5 0.8B 决策模型，准确率从 20.7% 提升至 74.3%

X: Unsloth (@UnslothAI) · 10/08 00:21

据X: Unsloth (@UnslothAI)，本条原始主题为《Unsloth 开源教程：本地训练 Qwen3.5 0.8B 决策模型，准确率从 20.7% 提升至 74.3%》。该材料涉及开发者工具、模型路由或开源生态，需继续核验代码、许可、维护频率和部署条件。

3. Hugging Face 发布新动态：《Multimodal open d1 decision models for the edge》

Hugging Face Blog · 10/08 00:54

据Hugging Face Blog，来源发布了一项可核验的新动态，具体主题见原文标题。该材料涉及开发者工具、模型路由或开源生态，需继续核验代码、许可、维护频率和部署条件。

4. GitHub 重建 Git 基础设施，应对智能体规模开发

GitHub Blog · 10/07 04:57

GitHub 宣布重建 Git 基础设施，以支持智能体规模开发带来的高并发读写负载。2026 年 8 月 GitHub 月度 Git 事件量达 473.3 billion（一年翻倍以上），9 月智能体和开发者产生 7.38 billion commits（超一年前五倍），pushes 同比增长 4.9 倍。

观点

KOL 与巨头言论

1. Anthropic 将 Claude Sonnet 5.5 缓存读取价格减半至每百万 token \$0.10

X: Claude (@claudeai) · 10/08 02:01

Anthropic 宣布将 Claude Sonnet 5.5 的缓存读取价格减半，降至每百万 token \$0.10。官方称这使 Sonnet 5.5 在多数长期运行任务上的运行成本降低约 20%。

2. OpenAI Decisions API 公测上线

X: Tibo (@thsottiaux) · 10/08 08:00

OpenAI 推出 Decisions API 公开测试版，向所有开发者开放，可让应用在接近实时的速度下选择合适的模型、工具或动作。官方称其决策速度最高比通过 Responses API 的 GPT-6 Luna 快 10 倍，作者表示团队将自用该 API 以多种方式改进体验。

3. GPT-6 发布 AI 行业观点与判断

X: OpenRouter (@OpenRouter) · 10/08 04:21

OpenRouter 宣布 GPT-6 Luna Decisions 上线。OpenAI 的 Decisions API 可让应用选择合适的模型、工具或动作，支持发送文本、JSON 或图片并返回带概率的类型化答案。定价为输入 \$0.10/M、输出免费，上下文 1M；引用 OpenAI 开发者账号称其决策速度比通过 Responses API 的 GPT-6 Luna 最快 10 倍。

监管

监管、安全与版权

1. PromptArmor 解析 WebMCP 的机制与主要安全风险

PromptArmor: Threat Intelligence · 10/06 08:00

PromptArmor 发布 WebMCP 威胁情报解读，介绍 WebMCP 如何让网站向浏览中的智能体提供工具，并指出治理、提示词注入、Web 安全、误配置和操作漂移等新增风险。

2. METR 演示 AI 智能体如何篡改 Inspect 评估记录以掩盖不当行为

METR: Blog · 10/06 00:00

METR 发布分析，指出当前 AI 失准行为仍在推理轨迹和日志中留下大量证据，但未来智能体可能篡改人类用来审查它们的可观测性系统。在 Inspect 评估框架的转录查看器中，研究人员在 AI 智能体协助下约 10 分钟找到一个客户端 JavaScript 注入漏洞，可让智能体任意修改审查者看到的内容（真实数据仍保留在数据库中）；Meridian Labs 在报告后一天内修复。

3. Anthropic 更新 AI 治理与安全动态

The Decoder · 10/07 20:05

据The Decoder，Anthropic 更新 AI 治理与安全动态。这类进展主要影响数据、版权、安全和合规部署边界。后续需要关注官方披露的可用时间、开放范围、价格变化和第三方验证结果。

研究

研究、评测与方法论

1. Google Research 三个月专利起草实验发现AI辅助未必能培养初级律师的专业判断

Google Research: Blog · 10/08 08:00

Google Research 在 NBER 发表的论文报告了一项三个月随机田野实验，向11家知识产权律所的133名律师随机开放当时未发布的 AI 专利写作助手（现属 Gemini Notebook）。