

# Claude完成费马大定理形式化证明，OpenAI发布GPT-6 Astra

中文内部早报 · 5分钟版

今日总览

AI领域迎来重大突破，Anthropic利用Claude在11天内完成费马大定理的Lean形式化证明，成为首个完整经计算机验证的版本。这一成就标志着AI在数学推理和形式化验证能力上的显著进展。

Agent和编程 workflow 方面，Hugging Face发布开源工具funes为编码智能体提供本地记忆层，Meta推出Muse Spark 1.3提升智能体科学推理能力。xAI设计Grok Bot的持久化智能体交互界面，推动长期智能体应用发展。

前沿模型竞争激烈，OpenAI发布GPT-6 Astra面向付费用户开放，Google DeepMind推出WeatherNext 3全球天气AI模型。IFM发布K2 Horizon六款开源模型，xAI让Grok Bot承担采购工作实现超10万美元节省。

大厂产业链动态中，NVIDIA宣布以129.3亿美元收购Hugging Face，Cursor推出Self-Hosted Machines支持企业自有机器执行。Google Cloud介绍低成本搭建常驻Agent方案，显示云服务在AI应用部署上的创新。

安全监管层面，GPT-6 Astra虽幻觉减少但仍易受隐藏提示词注入攻击。OpenAI强调其网络安全能力达到Critical级，但基准测试结果存在分歧，显示模型评估标准有待统一规范。

头条

今日头条

1. Anthropic用Claude在11天内完成费马大定理首个机器验证的Lean形式化证明

Anthropic: Research (发表成果 · 网页)

· 09/05 08:00 · 智能体、编程与工作流

Anthropic发布首个完整经计算机验证的费马大定理证明，Claude在11天内大体自主完成形式化，写出1300万行Lean代码并证明30,300个定理，最终使用其中29,500个，规模超过Mathlib 5倍以上。

2. OpenAI发布GPT-6 Astra，面向Pro、Enterprise和Business Premium用户开放

X: OpenAI (@OpenAI)

· 09/05 08:00 · 大厂与产业链

据X: OpenAI (@OpenAI)，OpenAI 发布 GPT-6 Astra，面向 Pro、Enterprise 和 Business Premium 用户开放。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. Google DeepMind发布WeatherNext 3全球天气AI模型，hourly更新且分辨率较上一代提升约5倍

Google DeepMind: Blog

· 09/03 23:02 · 前沿模型与多模态

Google DeepMind与Google Research发布WeatherNext 3，称其为迄今最先进的全球天气AI模型，经Brightband独立实时评估，hourly更新且分辨率较上一代提升约5倍。

4. OpenAI 发布 Daybreak 安全工具

TechCrunch AI

· 09/04 02:01 · 监管、安全与版权

OpenAI 推出 Daybreak 系列安全工具，包括 Codex Security 和 GPT-5.5-Cyber，面向组织的大规模漏洞发现、验证和修补流程。

5. Google 推出 AI 产业部署新进展

Google AI: DEV 作者专属

· 09/03 23:54 · 大厂与产业链

Shir Meir Lador在Google AI开发者博客介绍如何用Cloud Run instances以每月\$5.70（1 vCPU、1Gi内存、共享CPU）在云端24/7运行常驻Agent，为低成本部署提供方案。

模型

前沿模型与多模态

1. GPT-6 Astra 幻觉更少但仍易受隐藏提示词注入攻击

The Decoder: AI News

· 09/05 08:00

The Decoder 报道，OpenAI 新模型 GPT-6 Astra 幻觉少于前代 GPT-5.6 Sol，直接提示词注入防御率达 99.99%，但多轮自适应攻击下防御率降至约 67%。

2. IFM 发布 K2 Horizon 六款开源模型，覆盖 0.9B 到 375B-A23B 并开放完整训练生命周期

Hacker News

· 09/04 01:35

IFM 发布 K2 Horizon 模型系列，共六个模型：375B-A23B、36B-A4B、32B、7B、3.7B 和 0.9B，均以 Apache 2.0 开源，其中 0.9B、3.7B 和 7B 宣称在其规模上达到 SOTA，36B-A4B 采用新提出的稀疏注意力架构 MoVA。

3. xAI 让 Grok Bot 承担采购工作，Haggle Bot 找出超 10 万美元直接节省

xAI 新闻

· 09/05 08:00

xAI 让 Grok Bot 访问供应商支出、合同和使用数据，创建的 Haggle Bot 已识别超过 10 万美元直接节省，包括在一个 SaaS 产品中找到 43 个 90 天无活动的付费席位（节省 \$14,220），在另一个产品中找出每年 \$85,662 的未用 SKU。

4. OpenAI 更新模型能力与产品方案

OpenAI 新闻 · 09/03 20:00

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。

智能体 智能体、编程与工作流

1. Hugging Face 发布开源工具 funes，为编码智能体提供可本地持有的记忆层

Hugging Face 博客 · 09/03 08:00

Hugging Face 发布开源工具 funes，为 Claude Code、Codex、pi、Hermes 等编码智能体提供本地记忆层，把已有会话记录索引成 Lance 数据集，一条 funes add 命令即可让 Agent 自主召回原始出处（Agent、时间戳、会话、轮次）。

2. Meta 发布 Muse Spark 1.3，智能体与科学推理能力提升

X: Alexandr Wang（Scale AI 创始人/Meta 首席 AI 官） (@alexandr\_wang) · 09/03 04:37

据X: Alexandr Wang（Scale AI 创始人/Meta 首席 AI 官） (@alexandr\_wang)，Meta 发布 Muse Spark 1.3，智能体与科学推理能力提升。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。

3. OpenAI 更新 AI 工作流与智能体方案

OpenAI 新闻 · 09/03 20:00

据OpenAI 新闻，OpenAI 更新 AI 工作流与智能体方案。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. xAI 设计 Grok Bot：为持久化智能体重构交互界面

xAI 新闻 · 09/03 08:00

xAI 发布设计文章，介绍 Grok Bot 如何为超越单次会话的持久化智能体设计界面。产品以 Bot 为主要对象而非会话，Bot 拥有身份、记忆、自己的计算机和工具；头像动效呈现空闲、工作、等待、阻塞、思考、完成等状态；工作区提供状态、预览、接管三级访问。

产业 大厂与产业链

1. Cursor 推出 Self-Hosted Machines，云智能体可在企业自有机器上执行

Cursor Blog · 09/02 20:00

Cursor 发布 Self-Hosted Machines，让云智能体的工具执行迁移到企业自有网络内的机器，智能体循环、推理和规划仍留在 Cursor 云端，通过 worker 的出站 HTTPS 连接对接，Cursor 不会主动连入企业网络。

2. NVIDIA 宣布以 129.303 亿美元收购 Hugging Face

NVIDIA Blog · 09/03 19:59

NVIDIA 宣布已同意以 12，930，300，000 美元收购 Hugging Face，黄仁勋在官方博客公布了这一消息。Hugging Face 目前有超过 1800 万开发者，托管超过 300 万个模型、50 万个数据集和 100 万个应用，服务超过 20 万家企业。

3. Tom Tunguz 解析 Meta Muse Spark 双轨定价背后的数据换算力逻辑

Tomer Tunguz 博客（VC 分析） · 09/03 08:00

据Tomer Tunguz 博客（VC 分析），Tom Tunguz 解析 Meta Muse Spark 双轨定价背后的数据换算力逻辑。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。

4. NVIDIA 以 129.3 亿美元收购 Hugging Face

Artificial Intelligence News · 09/03 00:00

NVIDIA 已同意以 129.3 亿美元收购 Hugging Face，以扩展该开源模型仓库的平台与基础设施。交易目标是平台增长和基础设施投资，旨在为全球企业开发者、软件工程师和研究机构扩大 AI 访问机会。Hugging Face 由 Clem Delangue、Julien Chaumond、Thomas Wolf 等人在过去十年间打造。

开源 开源项目与开发者工具

1. GitHub 发布 Project HydraFusion 研究预览，用多模型运行时编排降低 Copilot 成本

GitHub Blog · 09/05 08:00

GitHub 推出 Project HydraFusion 研究预览，通过运行时多模型编排，在 Single、Cascade、Critique 三种执行模式间为每个任务选择工作流，以平衡质量、成本和延迟。

2. Hugging Face 发布新动态：《NeoMME: an efficient Multimodal-native and Multilingual Encoder》

Hugging Face Blog · 09/03 21:13

据Hugging Face Blog，本条原始主题为《NeoMME: an efficient Multimodal-native and Multilingual Encoder》。该材料涉及开发者工具、模型路由或开源生态，需继续核实验码、许可、维护频率和部署条件。

3. GitHub Copilot 如何在不牺牲任务质量的前提下降低 AI 编码成本

GitHub Blog · 09/03 02:00

GitHub 工程师 Erik Kristensen 分享了 Copilot 降本的四项改动：选择性压缩工具输出、移除 view 工具行号前缀（线下推理成本降约 5%，线上用户日均推理成本降约 3%）、压缩 task-tool 提示词（每轮省约 1300 token，每活跃小时归一化成本降 2.9%）、后台任务完成后直接交付结果（AI Credits 用量降约 2.3%）。

4. Hugging Face 发布新动态：《Fine-tuning a 350M Model for Better Structured Outputs in 100 GRPO Steps》

Hugging Face Blog · 09/03 08:00

据Hugging Face Blog，本条原始主题为《Fine-tuning a 350M Model for Better Structured Outputs in 100 GRPO Steps》。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。

观点

KOL 与巨头言论

1. Hacker 发布新动态：《Show HN: Votito helps you decide what to build next with confidence》

Hacker News · 09/04 22:04

据Hacker News，Show 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。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. One-person 发布 AI 行业观点与判断

Hacker News · 09/04 05:07

据Hacker News，One-person 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。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. Hacker 发布新动态：《Show HN: Pangolin – SSO and WireGuard instead of API keys for LLM access》

Hacker News · 09/03 21:35

据Hacker News，Show 发布 AI 行业观点与判断。这类信息反映行业参与者对技术路线和商业节奏的判断。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。

监管

监管、安全与版权

1. OpenAI 发布 GPT-6 Astra 并公布安全概览，称其网络安全能力达到 Preparedness Framework 的 Critical 级

OpenAI · 09/03 08:00

OpenAI 发布 GPT-6 Astra，称这是其部署过的最强模型，也是首个在其 Preparedness Framework 下达到网络安全能力 Critical 级的模型，可在无人逐步引导的情况下发现未知安全漏洞并开发利用方式。

2. 用 TRL 和 OpenEnv 训练编码模型画水彩：Hugging Face 全流程开源复现

Hugging Face 博客 · 09/03 08:00

Hugging Face 博客作者基于 Surya Narreddi 的原始想法，用 TRL、OpenEnv 和 Qwen/Qwen3.5-35B-A3B 复现了让语言模型通过 p5.brush 写 JavaScript 画水彩的 RL 训练流程，所有数据集、环境、脚本和模型均开源在 Hub。

3. Google DeepMind 发布 Gemini 3.8 Flash 与 3.8 Flash Cyber 两款新模型

Google DeepMind: Blog · 09/03 00:18

据Google DeepMind: Blog，Google DeepMind 发布 Gemini 3.8 Flash 与 3.8 Flash Cyber 两款新模型。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. GPT-6 Astra 基准表现分歧，ARC-AGI-3 效率超人类令 Chollet 提前 AGI 预测

The Decoder: AI News · 09/05 08:00

GPT-6 Astra 的基准结论相互矛盾：Epoch AI 以 169 分将其排在 267 个模型之首，Artificial Analysis 给出 61 分，仅与前代 Sol 持平、落后 Claude Fable 5.1 的 66 分。