

2026 AI 数学突破实录

OpenAI / ChatGPT 与 Claude：从猜想反例到机器可检验证明

资料截点：2026 年 8 月 28 日

目录

1. 怎样判断“核弹级”	2
2. 总览：截至 2026-08-28 的核心事件	2
3. OpenAI：从单点合作到批量原创	2
3.1. 前奏：GPT-5 与 GPT-5.2 开始进入真实研究	2
3.2. Astra 的十项结果：从“一个突破”变成“批量突破”	2
3.2.1. 为什么这批结果比“十篇论文”更重要	3
4. Claude：长程搜索、显式反例与解析数论	3
4.1. Claude’s Cycles：Knuth 的开放组合问题	3
4.2. 密码分析：HAWK 与 7 轮 AES	4
4.3. 黎曼 zeta 零点：从 $\frac{5}{12}$ 到超过 $\frac{2}{3}$	4
4.3.1. 它没有证明黎曼猜想	4
4.3.2. 核心想法	4
4.3.3. 发现过程	5
5. 观察名单： S^6 复结构	5
5.1. OpenAI	5
5.2. Anthropic / Claude	5
5.3. 背景与方法论	6

1. 怎样判断“核弹级”

数学成果的热度和可信度是两回事。本文使用下面四级证据标尺。

2. 总览：截至 2026-08-28 的核心事件

时间	系统	成果	状态
2025-11	GPT-5	协助解决 Nesterov 加速梯度点收敛问题，属于深度人机协作。	A
2025-12	GPT-5.2 Pro	直接推出最大似然估计学习曲线单调性的多项新结论。	A/B
2026-02	Claude Opus 4.6	构造 Knuth 有向图分解问题的奇数阶通解；后续 GPT 与 Claude 补全偶数阶。	A
2026-05	OpenAI 内部模型	推翻 Erdős 单位距离猜想。	A
2026-07	Claude Fable 5	公布三维 Jacobian 猜想的显式反例。	B
2026-07	Claude Mythos Preview	改进 HAWK 与 7 轮 AES 的最佳已知攻击。	B/C
2026-08	OpenAI Astra	一次发布十项数学与理论计算机科学结果。	B
2026-08	Claude 研究版	zeta 临界线简单零点比例下界推进到 0.67250...。	A
2026-08	Claude + 人类	宣称在 S^6 上构造复结构。	D

3. OpenAI：从单点合作到批量原创

3.1. 前奏：GPT-5 与 GPT-5.2 开始进入真实研究

2025 年的案例仍带有明显的“人类主导、模型加速”特征。

Nesterov 加速梯度。 UCLA 的 Ernest Ryu 用 GPT-5 探索一个源自 1983 年的优化问题：Nesterov accelerated gradient 在凸优化中的点收敛为何成立。模型提出了若干关键结构，但也多次给出错误论证；Ryu 负责筛选、修正、重写和最终证明。约 12 小时的密集交互把原本可能持续数周的探索压缩到三晚。这个案例的重要性在于，它展示了“高吞吐量猜想生成器 + 专家快速否认”的协作模式，而不是模型一次性吐出正确论文。[OAI-NAG](#)

学习曲线单调性。 随后的 GPT-5.2 Pro 案例更接近自主求解。问题问的是：在模型设定正确时，最大似然估计是否会随着样本增加而平均变好？Sellke 与 Yin 报告，论文中的结果由 GPT-5.2 Pro 的多个变体推导，人类没有提供证明策略或中间论证，只负责继续追问、验证和誊写。论文覆盖未知协方差高斯模型、Gamma 分布及更一般的指数族情形。[LCM](#)

这两项为什么不是同一种“AI 贡献”

NAG 案例中，人类专家承担了核心判断与修正；学习曲线论文则声称模型直接给出了证明路线。只说“ChatGPT 解决了开放问题”会抹掉这种关键差异。

3.2. Astra 的十项结果：从“一个突破”变成“批量突破”

2026 年 8 月 1 日，OpenAI 公布内部下一代模型 Astra 产生的十项结果：一份 253 页论文、一套 Lean 4 证书，以及每项结果的推理过程说明。OpenAI 表示，寻找这些解所消耗的总 token，按 Sol API 价格估算约 2,000 美元；人类随后用同一模型整理手稿，模型再做形式化。[OAI-10](#) [OAI-10-PAPER](#) [OAI-10-LEAN](#)

下面按“证明了什么”压缩记录。

编号	领域	核心结论与意义
1	高维球堆积	精确确定 Cohn–Elkies 线性规划在高维的指数强度： $LP_d^{\frac{1}{d}} \rightarrow \sqrt{\frac{e}{2\pi}}$ 。对应指数约 0.6044，超过 1978 年以来的通用高维球堆积指数约 0.5991。
2	二进制与球面码	对任意固定最小距离参数，把经典上界按指数因子改进；球面码极限还重新推出第 1 项的球堆积指数。
3	群论	构造显式非 sofic 群，回答“是否每个可数群都是 sofic”这一长期中心问题。
4	算子代数	构造无限多个两两不同构、具有性质 (T) 的群，却拥有相同群 von Neumann 代数，从而否定 Connes 刚性猜想。
5	算术电路复杂性	对 permanent 证明无除法电路下界 $\Omega(n^2 \log \log n)$ ，公式下界 $\Omega\left(\frac{n^4}{\log n}\right)$ 。这不是解决 $VP \neq VNP$ ，但属于极难取得进展的电路下界方向。
6	量子复杂性	对任意有限双人纠缠游戏证明指数型平行重复定理，把经典复杂性中的基本放大原理扩展到一般量子情形。
7	格与密码学	从 3SAT 直接归约，证明欧氏最近向量问题具有 $n^{\frac{1}{400}}$ 因子近似困难性，并推出二元解码等相关结论。
8	凸几何	在任意维度证明 Ehrhart 体积猜想：若凸体的重心是唯一内部格点，则体积至多 $\frac{(n+1)^n}{n!}$ 。
9	Ramsey 理论	证明多色三角 Ramsey 数 $R_{k(3)} = k^{\Theta(k)}$ ，给出超指数下界并解决 Erdős 问题 183。
10	极值图论	用不同的二部图构造否定 Erdős–Simonovits 紧致性猜想与 Erdős 退化度猜想，对应解决 Erdős 问题 146 与 180。

3.2.1. 为什么这批结果比“十篇论文”更重要

- **跨度异常大**：调和分析、编码、群、算子代数、电路、量子游戏、格密码与图论之间并无单一路线可复用。
- **不只是填空**：其中包含存在性问题的正面构造、著名猜想的反例、精确渐近和复杂性下界。
- **产出带证书**：公开仓库分别提供 SpherePacking.lean、NonSoficGroup.lean、ConnesRigidity.lean 等文件。
- **边际成本骤降**：如果 2,000 美元量级的推理成本可以稳定换取若干可发表结果，数学研究的稀缺资源将从“想法生成”转向“选题、理解、验证与组织知识”。

当前可信度判断：强证据，但仍在社区评审期
论文和 Lean 仓库均已公开，证据远强于新闻稿；但十项结果同时发布，且覆盖多个高度专业领域。截至资料截点，还不能说每一章都经历了与单位距离论文同等强度的独立专家消化。本文因此整体标为 B，而不是把“有 Lean 文件”直接等同于“十项均已完成传统同行评议”。

4. Claude：长程搜索、显式反例与解析数论

4.1. Claude’s Cycles：Knuth 的开放组合问题

Donald Knuth 在写《计算机程序设计艺术》时研究如下有向图：顶点为

$$(i, j, k) \in \mathbb{Z}_m^3, \tag{1}$$

每个顶点有三条边，分别把 i, j, k 中的一个坐标加 1（模 m ）。问题是能否把全部边分解为三个长度 m^3 的有向 Hamilton 圈。

Claude Opus 4.6 先把问题重写成每个顶点上对三种方向分配一个排列，随后经历深度优先搜索、蛇形路径、纤维分解与模拟退火等多轮失败，最终从 $m = 3$ 的解中识别出只依赖少量“边界状态”的通式，并解决所有奇数 $m \geq 3$ 。Knuth 给出人工证明，Kim Morrison 又很快用 Lean 形式化了该构造。[KNUTH](#) 后续过程本身也值得记录：

- GPT-5.3 Codex 找到偶数 $m \geq 8$ 的闭式程序；
- GPT-5.4 Pro 根据程序独立写出约 14 页证明；
- GPT 与 Claude 的协同又找到更简洁的奇、偶统一构造；
- 因而这不是“某个模型单枪匹马解决全部情况”，而是一个模型发现奇数结构，其他模型与人类继续补全的接力案例。

为什么可信

原问题、构造、人工证明、程序和 Lean 验证均公开；Knuth 的 2026-04-14 修订版还保留了失败过程和后续补全细节。这是研究溯源做得最好的案例之一。

4.2. 密码分析：HAWK 与 7 轮 AES

Claude Mythos Preview 在 2026 年 7 月公布两项密码分析结果：

1. 对后量子签名候选 HAWK 找到更强攻击，把估计密钥强度有效削弱约一半；该候选已经历两年、两轮专家审查。
2. 对 7 轮 AES-128 的 meet-in-the-middle 攻击提出“Möbius Bridge”指纹，把原先必须枚举的一个 256 值猜测吸收到不变量中；综合额外优化后，比此前攻击快约 200–800 倍。

HAWK 结果由一位研究员与 Claude 在约 60 小时内完成；AES 结果则由研究员搭建脚手架后，Claude 近乎自主地产生数亿到十亿量级 token，随后人类花费数百小时验证。每项主要结果的开发 API 成本约 10 万美元。[ANTH-CRYPTO](#)

这不等于“Claude 破解了 AES”

攻击对象是人为削弱的 7 轮 AES，而标准 AES-128 有 10 轮；HAWK 也只是 NIST 额外后量子签名流程中的候选，尚未部署。成果的意义是模型已经能做专家级算法密码分析，不是现有互联网加密突然失效。

4.3. 黎曼 zeta 零点：从 $\frac{5}{12}$ 到超过 $\frac{2}{3}$

4.3.1. 它没有证明黎曼猜想

黎曼猜想断言，zeta 函数每个非平凡零点 $\rho = \beta + i\gamma$ 都满足

$$\beta = \frac{1}{2}. \quad (2)$$

Claude 没有证明这一点。它证明的是一个无条件比例结果：当高度 $T \rightarrow \infty$ 时，至少 $\frac{2}{3} - o(1)$ 的非平凡零点既位于临界线，又是简单零点；采用 Montgomery–Taylor 窗函数时，比例提高为

$$0.67250\dots \quad (3)$$

同时，至少 $\frac{5}{6} - o(1)$ 的零点互不相同，优化常数为 0.83625…。此前“简单且位于临界线”的无条件纪录是 $\frac{5}{12} \approx 0.4167$ ，“互不相同”的纪录约为 0.6603。[ZETA-ARXIV](#)

4.3.2. 核心想法

Montgomery 1973 年的相关推导在黎曼猜想成立时能把零点侧理解为对实数纵坐标的正和。Claude 的路线不再假设所有零点都在线上，而是：

1. 用 Weil 显式公式得到一个 Hermitian 型；
2. 对它做有限维压缩；
3. 用 rank-trace 不等式提取“能有多少正方向”；
4. 用 Sylvester 惯性定律处理偏离临界线的共轭零点对；
5. 接入 Aryan 以及 Baluyot–Goldston–Suriajaya–Turnage–Butterbaugh 的现代解析输入。

真正新的一步，是把线上与线外零点统一放进允许非对角项的二次型里，用惯性而不是逐点正性来计数。

4.3.3. 发现过程

Anthropic 报告，未公开研究版 Claude 最初尝试 650 个思路均失败。第二轮用了约 60 个子代理、3,100 万输出 token、2,400 次命令和数百个 Python 程序；模型下载 54 篇 arXiv 论文检查先例，并让不同代理寻找反例、独立重证与互审。人类输入大多只是要求继续与鼓励。

Levent Alpöge 与 Ralph Furman 随后独立重推并以人类作者身份提交 21 页 arXiv 论文，论文明确记录“证明由 Claude 自主发现，人类负责验证与传播”；Brian Conrey 与 Dan Goldston 也在短时间内审阅。Lean 4 形式化已公开。[ANTH-ZETA ZETA-LEAN](#)

当前可信度判断：强确认

公开 AI 原稿、独立人类重述、领域专家审阅、arXiv 版本与 Lean 形式化同时存在。虽然尚不能替代长期同行评议，但这是目前证据链最完整、数学分量也最高的 Claude 成果之一。

5. 观察名单： S^6 复结构

2026 年 8 月下旬出现了一份由 Levent Alpöge 与 Claude 相关流程生成的长文，声称在六维球面 S^6 上构造复结构。这个问题自 1948 年以来一直开放；如果成立，数学意义极大。

但截至本文截点：

- 公开文本有 2 页摘要式版本和 100 多页长版本，论证很难消化；
- Berkeley 数学家 Tony Feng 明确写道，他尚未检查数学细节，不能为正确性背书；
- 公开讨论指出它需要与 Campana–Demailly–Peternell 一类既有障碍精确兼容；
- 尚缺少类似单位距离论文那样的多人独立人类重述。

因此本文只记录“存在该主张”，不把它算进已确认突破。[S6-STATUS](#)

记录原则

越是接近“改写教科书”的结论，越要降低传播速度、提高验证门槛。对 S^6 的正确表述是“出现了 AI 辅助的声称构造，正在核验”，不是“Claude 已解决 78 年难题”。

5.1. OpenAI

1. [Planar Point Sets with Many Unit Distances](#)：AI 生成的单位距离反例论文。
2. [Remarks on the Disproof of the Unit Distance Conjecture](#)：九位数学家的独立消化与评论。
3. [Ten Advances in Mathematics and Theoretical Computer Science](#)：十项结果的 253 页合订论文。
4. [openai/ten-proofs](#)：十项结果的 Lean 4 形式化仓库。
5. [OpenAI: Ten advances](#)：发布说明与责任、署名立场。
6. [On Learning-Curve Monotonicity for Maximum Likelihood Estimators](#)：GPT-5.2 Pro 推导的统计学习理论结果。
7. [How GPT-5 helped Ernest Ryu](#)：NAG 人机协作的完整案例说明。

5.2. Anthropic / Claude

1. [More than two thirds of the zeta zeros are simple and on the critical line](#)：人类验证与传播的 arXiv 论文。

2. [Anthropic: Learning more about Claude' s mathematical capabilities](#): 发现过程、成本与参与者说明。
3. [anthropics/formal-math](#): zeta 结果的 Lean 4 形式化。
4. [Donald Knuth: Claude' s Cycles](#): 组合问题的发现、证明、失败轨迹与后续补全。
6. [Discovering Cryptographic Weaknesses with Claude](#): HAWK、AES 与后续密码分析结果。
7. [Tony Feng: \$S^6\$ 声称构造的当前状态](#): 明确说明尚未核验、不能背书。

5.3. 背景与方法论

1. [The Gold Rush in AI4Math: Where Are We Now?](#): 2026 年数学 arXiv 投稿中的 AI 使用横截面研究。
2. [Leiden Declaration on AI and Mathematics](#): 关于透明度、署名、验证与数学共同体责任的原则声明。

完。本文只反映截至 2026-08-28 可获得的公开材料；高影响的新主张应随社区验证持续更新。