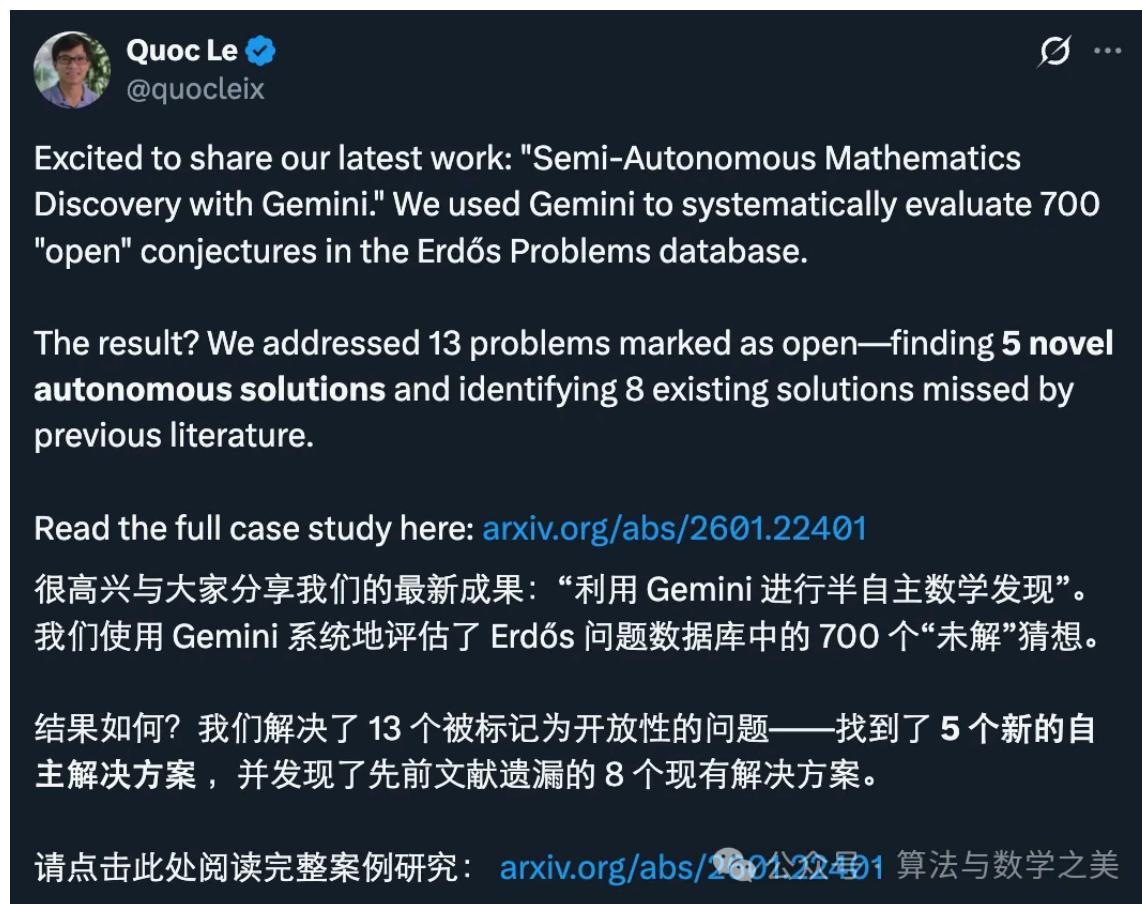


谷歌给「AI解数学题」神话降温：能摘低垂果实，但过程依然痛苦

算法与数学之美 2026年2月6日 22:33 安徽

刚刚，谷歌发布了一项新的研究进展：他们用 Gemini 做了一次系统性的数学攻关实验，把目标对准了著名的 Erdős Problems 数据库里 700 个仍被标注为 open（未解决）的猜想。



Quoc Le @quocleix

Excited to share our latest work: "Semi-Autonomous Mathematics Discovery with Gemini." We used Gemini to systematically evaluate 700 "open" conjectures in the Erdős Problems database.

The result? We addressed 13 problems marked as open—finding **5 novel autonomous solutions** and identifying 8 existing solutions missed by previous literature.

Read the full case study here: arxiv.org/abs/2601.22401

很高兴与大家分享我们的最新成果：“利用 Gemini 进行半自主数学发现”。我们使用 Gemini 系统地评估了 Erdős 问题数据库中的 700 个“未解”猜想。

结果如何？我们解决了 13 个被标记为开放性的问题——找到了 **5 个新的自主解决方案**，并发现了先前文献遗漏的 8 个现有解决方案。

请点击此处阅读完整案例研究：arxiv.org/abs/2601.22401 算法与数学之美

结果相当亮眼：Gemini 在这批问题中一共推进了 13 个——其中 5 个是模型自主给出的全新解法，另外 8 个则是模型在文献中挖出了早已存在、但此前被遗漏的解答。

Semi-Autonomous Mathematics Discovery with Gemini: A Case Study on the Erdős Problems

Tony Feng^{†*}, Trieu Trinh^{*}, Garrett Bingham^{*}, Jiwon Kang[†], Shengtong Zhang[†], Sang-hyun Kim[†], Kevin Barreto[†], Carl Schildkraut[†], Junehyuk Jung[†], Jaehyeon Seo[†], Carlo Pagano[†], Yuri Chervonyi^{*}, Dawsen Hwang^{*}, Kaiying Hou[†], Sergei Gukov[†], Cheng-Chiang Tsai[†], Hyunwoo Choi[†], Youngbeom Jin[†], Wei-Yuan Li[†], Hao-An Wu[†], Ruey-An Shiu[†], Yu-Sheng Shih[†], Quoc V. Le[°], Thang Luong[°]

[†]Mathematical contribution, ^{*}Engineering contribution, [°]Principal Investigators

公众号·算法与数学之美

- 论文标题: Semi-Autonomous Mathematics Discovery with Gemini: A Case Study on the Erdős Problems
- 论文链接: <https://arxiv.org/pdf/2601.22401>

Erdős Problems 数据库以数学家 Paul Erdős 的名字命名。他是 20 世纪最多产的数学家之一，留下了大量论文和未解决的猜想，涵盖数论、组合数学、图论等。2023 年，数学家 Thomas Bloom 推出了 ErdosProblems.com 网站，这是一个集中式数据库，旨在整理这些猜想并跟踪其研究进展。目前，该数据库共收录 1179 个问题，其中 483 个（41%）被归类为已解决。

然而，该数据库中标注「open」的问题并不一定代表问题真的未被解决，而是意味着至少有一位专业数学家尝试通过网络搜索寻找已发表的解决方案，但以失败告终。

事实证明，很多问题并非「未解决」，而是答案被淹没了。去年 10 月份，OpenAI 宣布 GPT-5 在该网站上发现了 10 个标记「open」的问题，但其实它们的答案已经存在于相关文献，只是之前未被搜到。

这一发现使得 Bloom 的数据库受到了广泛关注，同时促使陶哲轩近期创建了一个社区维基，专门跟踪人工智能辅助解决 Erdős 问题的相关动态。

如今，谷歌的研究把 Erdős 问题的解决又往前推了一步。但他们也坦言，这并不意味着 AI 已经能「自动做数学研究」了，背后的脏活累活远超普通人想象。

研究方法

作者团队在 2025 年 12 月 2-9 日部署了一个基于 Gemini Deep Think 的定制数学研究智能体 Aletheia，对 Bloom 数据库中当时仍标注为「Open」的约 700 个 Erdős 问题进行半自动探索。Aletheia 内置自然语言验证器（verifier），用于在大规模生成后先做第一轮筛选，将候选问题从 700 个快速收敛到 212 个「看起来可能正确」的回答。

接下来进入人工评估阶段。研究团队先由非该领域专家的数学家进行快速过滤，尽可能在可控时间内剔除明显错误解，从而把候选规模压缩到 27 个，再交由内部领域专家逐一严审；当解法的正确性明确但新颖性存疑时，还会咨询外部专家核对文献。

最终统计显示，在可明确判定的约 200 个候选解中，137 个（68.5%）存在根本性错误；63 个（31.5%）在形式上成立，但**其中只有 13 个（6.5%）真正回答了 Erdős 原本想问的问题**。其余 50 个虽然「技术上正确」，却因为误读题意而导致数学意义有限，作者计划对这些问题提出更严谨的修订表述；此外还有 12 个回答因问题本身开放或表述不清而被标记为「歧义」。

根据陶哲轩的建议，作者着重列出了上述数据以保证透明度。这也是为了更完整地呈现 AI 辅助数学研究的真实成本：除了少数正例之外，大量时间会消耗在核验、纠错、排查细微错误，以及检索文献以排除「无意重复」上。

这表明，业内广为流传的「**AI 正在加速科学**」的论断有一定片面性：人们通常只展示少数成功案例，强调 AI 在某个任务上比人类更快，从而声称 AI「加速」了这一结果；但这类叙事很少把负例纳入计算。

更具挑战性的是最后一步——确认解答是否已在文献中出现、以及是否真正契合 Erdős 的原始意图。许多问题的困难不在数学推导，而在题面细节的抄录误差、遗漏、以及符号与定义约定的歧义；模型若不了解 Bloom 网站的定义惯例，往往会在多个「各自合理」的解释之间混淆。

作者指出，在深入做文献核查与语义对齐后，「真正有意义的正确解」数量会显著下降，这也提醒未来的 AI 数学发现工作必须对题意一致性与文献溯源保持高度谨慎。

关键结果

作者将 13 个有意义的正确结果分为四类：

1、AI 自主解决。对于这些问题，Aletheia 找到了首个正确的解决方案，且解决方案具有实质性的数学意义。其中包括 Erdős-652 和 Erdős-1051，但需要说明的是，Erdős-652 的解决是通过直接引用现有文献中的结果实现的。

2、部分由 AI 解决。对于这些包含多个子问题的复杂问题，Aletheia 找到了其中一个子问题的首个正确解决方案。其中包括 Erdős-654、Erdős-935 和 Erdős-1040。

3、独立重发现。对于这些问题，Aletheia 找到了正确的解决方案，但人类审核者随后发现文献中已存在独立的解决方案。其中包括 Erdős-397、Erdős-659 和 Erdős-1089。这些解决方案似乎是模型独立重发现的：作者仔细检查了 Aletheia 的推理过程日志，确保该解决方案并非直接从文献中提取。当然，**该解决方案也有可能是通过中间来源或预训练过程间接从文献中获取的**。这凸显了 AI 生成数学内容所伴随的一个新风险：模型可能会再现预训练过程中习得的文献知识，却不注明来源，即存在「潜意识抄袭」的风险。

4、文献识别。对于这些问题，尽管在模型部署时 Bloom 网站将其标记为「open」，但 Aletheia 识别出文献中已明确存在相关解决方案。其中包括 Erdős-333、Erdős-591、Erdős-705、Erdős-992 和 Erdős-1105。

需要明确的是，研究团队并未声称后两类结果具有创新性。上述提到的 5 个自主生成的解决方案分别对应 Erdős-652、Erdős-654、Erdős-935、Erdős-1040 和 Erdős-1051。**根据专家的评估，这 5 个解决方案均未达到学术论文的水平。事实上，其中一些解决方案仅相当于研究生习题的难度**（基于现有文献）。

他们初步认为，Aletheia 对 Erdős-1051 的解决方案是 AI 系统自主解决具有一定普遍性（温和）数学意义的重要开放 Erdős 问题的早期案例——虽然存在关于密切相关问题的过往文献，但这些文献均未完全解决 Erdős-1051。

此外，与许多之前讨论的案例不同，作者认为 Aletheia 的解决方案并非直接受任何先前人类论证的启发，但该方案确实采用了经典思路：转向级数尾部并应用马勒准则（Mahler's criterion）。在 Aletheia 与人类数学家以及 Gemini Deep Think 的协作下，Erdős-1051 的解决方案得到了进一步推广，并形成了研究论文。

研究意义

研究结果表明，**Erdős 问题中存在「低垂的果实」，而 AI 已发展到能够摘取这些果实的水平。**尽管这为 AI 研究人员提供了一种新的、有趣的数学基准，但作者提醒**人们不应过度夸大其数学意义。本文解决的所有开放问题，任何相关领域的专家都能轻松完成。**另一方面，人类专家的时间有限。如果能够提高 AI 的可靠性，它已展现出加速数学发现中注意力瓶颈环节的潜力。

在本文的案例研究中，作者遇到了一些最初未预料到的困难。绝大多数技术正确的自主生成解决方案都源于对问题陈述的误解或解读缺陷，而诊断这些问题有时需要花费大量精力。

此外，人类专家面临的最具挑战性的步骤并非验证解决方案的正确性，而是确定这些解决方案是否已存在于文献中。随着人工智能生成数学内容的增多，**学术界必须警惕「潜意识抄袭」**，即 AI 再现训练过程中习得的文献知识，却未给予适当引用。需要注意的是，形式化验证无法解决这些问题。

尽管 AI 自主解决 Erdős 问题的尝试取得了一定成功，但也引发了误导性的炒作和彻头彻尾的虚假信息，并在社交媒体平台上被放大，这对数学界造成了损害。除了 Erdős 问题，未来可能还会有许多其他数学猜想列表成为（半）自主研究的目标。作者恳请相关研究人员关注本文提出的这些问题。

更多信息请参考原论文。

© THE END

文章来源：机器之心。

