

AI for Mathematics 从入门到前沿的学习指南

(面向数学专业学生)

董彬，北京大学，2025 年 9 月 28 日

(一) 绪论

AI for Mathematics (AI4M) 顾名思义就是 AI 辅助数学的研究，为数学前沿探索降本（主要是时间成本）增效。在传统的数学研究中，无论是寻找相似结论、快速学习新理论（**知识导航**），还是建立证明、验证反例（**证明与验证**），亦或是从复杂理论中发掘新模式、新关联（**洞察与连接**），都极其耗费精力。这三者环环相扣，共同驱动数学的进步，也正是 AI 可以大显身手的领域。因此，AI4M 的核心任务就是要围绕这些痛点，铸造三大核心能力：

- **知识导航**：AI 能帮助数学家快速学习和掌握核心理论与工具。
- **证明与验证**：AI 需要能够辅助证明、进行绝对严格的验证，甚至启发全新的证明思路。
- **洞察与连接**：AI 不仅能找到相似的概念与定理，更能从复杂的理论体系、证明和数据中发掘有趣的模式与观点，为数学家带来全新的灵感。

在具体给出入门学习指南之前，同学们先要问一问自己的兴趣主要是在数学还是在算法设计与编程。当然要做好 AI4M，两类同学都需要对 AI 和编程有基本的理解，所以对**机器学习、深度学习、强化学习、大模型**有一定的理解是必要的。网上有很多比较好的教材以及在线学科课件，我就不作太具体推荐，同学们可以根据自己的喜好来选择，但 LLM 我还是推荐一下人民大学文继荣老师团队写的那本《**大语言模型**》，以及 **Dimitri Bertsekas** 的强化学习在线课程和教材（<http://www.mit.edu/~dimitrib/RLbook.html>）。

此外，**智能体 (Agent)** 也是一个极为重要的概念，可以阅读一下 **Prompt Engineering Guide** 的网站上的博文（<https://www.promptingguide.ai/research/llm-agents>），以及网站上关于 Prompt Engineering (PE, 提示词工程) 的其他知识和相关代码工具。如果你觉得 PE 看起来和数学没有什么关系，可以阅读一下我们的一篇文章，从**最优控制角度刻画 PE 以及基于 PE 的智能体构建** (arXiv:2310.14201)。其实 PE 是今年由 Andrej Karpathy 提出来的软件 3.0 的基础（<https://www.bilibili.com/video/BV17u3LzQEYg/>）。

(二) AI4M 具体学习科研指南

目前 AI4M 主要任务是构建 AI 工具来辅助数学研究，具体分为两种模式：

- 1) 构建**专用工具**来激发数学家的灵感；
- 2) 构建**通用模型**来更加系统化的解决一类数学问题或者任务。

总的说来前者更适合对数学本身特别有兴趣的同学，而后者更适合对算法设计与编程更感兴趣的同学。

专用工具

近几年数学家发现数学研究中有不少问题是可以通过构建有针对性的 AI 工具来从数

据中挖掘特征，从而启发数学家提出更好的猜想，积极推动一些重大数学难题的解决。这一类型的工作大多都被总结到了 <https://seewoo5.github.io/awesome-ai-for-math/> 这个列表中，是一个不错的学习材料。其中我觉得最有代表性的就是 Geordie Williamson 和 DeepMind 在 2021 年发表的 Nature 封面文章，同学们可以看我在知乎上对该文章的解读，以及如何启发我和香港大学何旭华合作的一篇关于机器学习推动 ADLV 维数公式研究的工作 (<https://www.zhihu.com/question/503185412/answer/2256015652>)。

注：如果感兴趣这类研究，需要对数学本身有很敏锐的感知，且对机器学习和人工智能方法有足够的了解，从而可以判断数学问题中的哪些具体环节可以通过数据驱动的手段来推动，如何生成数据、选择模型、训练，以及如何对训练结果做数学上有意义的解读。

🌈 通用模型

这类 AI4M 方法更接近当前大模型的前沿与核心，需要同学们在了解大语言模型的基础上，理解数学作为一种语言的特殊性，这也是为什么形式化数学如此重要。我们现在会认为把数学从自然语言表述（也就是咱们通常意义上的数学表述方式）转换到某种形式化的表述的这个过程为**数学的数字化**。形式化语言有不只一种，其中目前生态最好，最受数学家欢迎的就是 Lean。所以感兴趣这个大方向的同学，可以先学习 Lean，一个推荐的讲义就是 Jeremy Avigad 等人写的《**Theorem Proving in Lean**》(https://leanprover.github.io/theorem_proving_in_lean4/)。

即便你对 AI 没有那么有兴趣，就是纯粹喜欢数学，Lean 本身对你的数学研究也会有极大的助力，毕竟可以帮你 100% 的验证定理证明的正确性。此外，你还可以用 Lean 来构建高难度评测集，难倒那些推理能力最强的大语言模型，这对 AI 模型本身的发展也是极为有帮助的。这类数据集包括北大-九坤 AI4M 团队推出的高难度代数评测集 **FATE** (<https://frenzymath.com/blog/fate/>)，其中 FATE-X 的难度显著超过了代数的博资考。

然而 Lean 是一个学起来有一定难度的语言，这方面可以去关注一下**陶哲轩的 blog**，他对 Lean 有很多的讨论和思考（关于 Lean 的学习体会，可以参看 <https://mathstodon.xyz/@tao/111252123637176099>）。因此，近两年 AI 辅助形式化变得越来越重要，特别是如何加快 mathlib（Lean 中目前最大的形式化数据库）的完善。首先，用户在基于 mathlib 进行形式化的时候，一个占去近一半时间的事情就是频繁的在 mathlib 里面进行检索，因此我们构建了一个目前被 Lean 社区广泛使用的语义检索工具 **LeanSearch** (<https://github.com/frenzymath/LeanSearch> & <http://leansearch.net/>)，该工具也后续帮助我们构建了后文提到的自动形式化工具（Herald）和形式化证明系统（REAL-Prover）。该工作也启发了一些近期的工作，比如 LeanExplorer (arXiv:2506.11085, <https://www.leanexplore.com/>)。

为了进一步辅助形式化，近几年出现了**自动形式化（autoformalization）**这个重要的 AI4M 子任务，最早的工作包括 Y. Wu et al. 对于形式化语言 Isabelle/HOL 的**自动形式化** (arXiv:2205.12615)，以及面向 Lean 的很多自动形式化数据集和模型的构建，包括 ProofNet (arXiv:2302.12433)、A. Jiang et al. 的**多语言模态的自动形式化** (arXiv:2311.03755,)、**TheoremLlama** (arXiv:2407.03203) 和 **Herald** (arXiv:2410.10878) 等等。近期初创企业 math.inc 推出的辅助形式化系统 **GAUSS** (<https://www.math.inc/gauss>) 更是把这个方向推到了一个新的高度。

数学的数字化对于 AI 提升科研水平的推理能力是极为重要的，因为可以很方便的在模型训练和推理过程中高效的给出高质量的反馈信号。这类研究工作近两年越来越多，

最早利用 Lean 形式化验证系统来提升模型推理能力的工作包括 **GPT-f**

(arXiv:2009.03393) 和 **HyperTree Proof Search** (arXiv:2205.11491), 之后还有在 2024 年获得数学奥林匹克竞赛银牌的 **AlphaProof** 和 **AlphaGeometry** (AlphaGeometry 用的是一个仅适用于一部分平面几何问题的独立的形式化系统, 但究其方法本质也属于借助形式化验证提升推理能力的这一范畴)。

今年开始, 越来越多的基于 Lean 构建和训练的推理模型被提出, 比如 **Goedel-Prover** (arXiv:2502.07640), **Kimina Prover** (arXiv:2504.11354), 以及北大-九坤 AI4M 团队近期推出的 **REAL-Prover** (arXiv:2505.20613), 当然还有著名的 **DeepSeek-Prover** 系列 (arXiv:2405.14333, arXiv:2504.21801) 和 **Qwen2.5-math** (arXiv:2409.12122) 等等 (论文太多不可能覆盖全, 同学们可以持续关注 AI 顶会相关论文)。

这里多提一句, 我们团队已经把 REAL-Prover 做成了一个 Lean tactic, 叫做 **reap**, 可以帮助 Lean 用户自动填充一些证明片段, 目前已经在 Lean 社区逐渐推广 (可以关注 <https://frenzymath.com/blog/reap/> 以及后续更新)。

当然, 要说 AI 形式化推理这个方向最具热度的新进展, 应该首推今年在 IMO 获得金牌的这些模型和 Agent 系统, 可以系统的阅读一下这些推文/论文 (**DeepMind-DeepThink, Harmonic-Aristotle, ByteDance-Seed Prover** 等)。更有意思的是, 也有学者展示, 仅仅通过调用 Gemini 2.5 pro 来构建智能体 (Agent), 也可以获得 2025 的 IMO 金牌 (arXiv:2507.15855), 这也表明当基座模型越来越强, 构建智能体可能是在解决很多问题时候的第一选择。

此外, 类似 DeepMind 推出的 **AlphaEvolve** (arXiv:2506.13131) 这样的智能体也很有意思, 属于 AI Scientist 的研究范畴, 可以阅读 AlphaEvolve 对应的 blog, 也可以具体上手魔改一下类似 **OpenEvolve** (<https://github.com/codelion/openevolve>) 和最近的 **ShinkaEvolve** (arXiv:2509.19349) 等开源版。这类工具也特别适合那些在数学上目标定义比较明确且可以量化, 而且已经有了一个还可以的初始化方案, 之后用这种 Evolve 的智能体来改善初始方案, 时常会收获令人惊艳的结果。

(三) 写在后面的话

数学研究中, 证明并不是最重要的 (参考本文第一章开头)。也正如著名数学家 **William Thurston** 在他经典文章《**On Proof and Progress in Mathematics**》(arXiv:math/9404236) 中指出的那样: 很多时候, “理解”远比严谨的证明本身更为重要。Thurston 在文中指出, 我们从事数学研究的根本目的是提升人类对数学的理解和洞察力。

所以我认为, 定理的证明虽然重要, 但它们其实更像是一种工具, 是帮助我们获得深刻理解的手段, 而不是最终的目的。因此, 当 AI 技术逐渐成熟, 能够高效地替我们完成那些繁琐的、机械化的定理证明与验证工作时, 我们人类数学家就可以从这些耗时的工作中被解放出来, 真正把精力集中在探索数学内在的深刻联系与规律, 去寻找更深层次的数学理解。

最后, 陶哲轩最近在维护一个很大的和 AI+数学有关的网站, **Crowdsourced Math Projects**, 涵盖了很多相关项目、论文列表、数据集、代码等等, 可以持续关注。

(<https://mathoverflow.net/questions/500720/list-of-crowdsourced-math-projects-actively-seeking-participants>)