key technical contributions include: (1) training an accurate and faithful LLM-based verifier for mathematical proofs, (2) using meta-verification to largely reduce hallucinated issues and ensure verification quality, (3) incentivizing the proof generator to maximize proof quality through self-verification, and (4) scaling verification compute to automatically label increasingly hard-to-verify proofs to improve the verifier without human annotation. DeepSeekMath-V2 demonstrates strong performance on competition mathematics. With scaled test-time compute, it achieved gold-medal scores in high-school competitions including IMO 2025 and CMO 2024, and a near-perfect score on the undergraduate Putnam 2024 competition. This work establishes that LLMs can develop meaningful self-evaluation abilities for complex reasoning tasks. While significant challenges remain, we hope this research direction contributes to the goal of creating self-verifiable AI systems that can solve research-level mathematics.

References

AlphaProof and A. teams. Ai achieves silver-medal standard solving international mathematical olympiad problems, 2024. URL https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level.

L. Chen, J. Gu, L. Huang, W. Huang, Z. Jiang, A. Jie, X. Jin, X. Jin, C. Li, K. Ma, C. Ren, J. Shen, W. Shi, T. Sun, H. Sun, J. Wang, S. Wang, Z. Wang, C. Wei, S. Wei, Y. Wu, Y. Wu, Y. Xia, H. Xin, F. Yang, H. Ying, H. Yuan, Z. Yuan, T. Zhan, C. Zhang, Y. Zhang, G. Zhang, T. Zhao, J. Zhao, Y. Zhou, and T. H. Zhu. Seed-prover: Deep and broad reasoning for automated theorem proving. CoRR, abs/2507.23726, 2025. doi: 10.48550/ARXIV.2507.23726. URL https://doi.org/10.48550/arXiv.2507.23726.

Y. Chervonyi, T. H. Trinh, M. Olsák, X. Yang, H. Nguyen, M. Menegali, J. Jung, V. Verma, Q. V. Le, and T. Luong. Gold-medalist performance in solving olympiad geometry with alphageometry2. CoRR, abs/2502.03544, 2025. doi: 10.48550/ARXIV.2502.03544. URL https://doi.org/10.48550/arXiv.2502.03544.

L. M. de Moura, S. Kong, J. Avigad, F. van Doorn, and J. von Raumer. The lean theorem prover (system description). In A. P. Felty and A. Middeldorp, editors, Automated Deduction - CADE-25 - 25th International Conference on Automated Deduction, Berlin, Germany, August 1-7, 2015, Proceedings, volume 9195 of Lecture Notes in Computer Science, pages 378–388. Springer, 2015. doi: 10.1007/978-3-319-21401-6_26. URL https://doi.org/10.1007/978-3-319-21401-6_26.

G. DeepMind. Gemini 2.5 pro, 2025. URL https://deepmind.google/models/gemini/pro.

DeepSeek-AI. Deepseek-v3.2-exp: Boosting long-context efficiency with deepseek sparse attention, 2025.

J. Dekoninck, I. Petrov, K. Minchev, M. Balunovic, M. T. Vechev, M. Marinov, M. Drencheva, L. Konova, M. Shumanov, K. Tsvetkov, N. Drenchev, L. Todorov, K. Nikolova, N. Georgiev, V. Kalinkova, and M. Ismoldayev. The open proof corpus: A large-scale study of llm-generated mathematical proofs. CoRR, abs/2506.21621, 2025. doi: 10.48550/ARXIV.2506.21621. URL https://doi.org/10.48550/arXiv.2506.21621.

D. Guo, D. Yang, H. Zhang, J. Song, P. Wang, Q. Zhu, R. Xu, R. Zhang, S. Ma, X. Bi, X. Zhang, X. Yu, Y. Wu, Z. F. Wu, Z. Gou, Z. Shao, Z. Li, Z. Gao, A. Liu, B. Xue, B. Wang, B. Wu, B. Feng,