

# Jasper Dekoninck

COMPUTER SCIENCE PHD CANDIDATE · ETH ZÜRICH

✉ [jasper.dekoninck@inf.ethz.ch](mailto:jasper.dekoninck@inf.ethz.ch) | 🏠 [sri.inf.ethz.ch/people/jasper](https://sri.inf.ethz.ch/people/jasper) | 📺 [JasperDekoninck](#) | 🔗 [LinkedIn](#) | 🐦 [j\\_dekoninck](#) | 🎓 [Google Scholar](#)

## Latest Work

My current research is primarily centered around [mathematical evaluation of LLMs](#). Most notably, I lead the [MathArena project](#) 📄, a mathematical benchmark that was used by [Google](#) 📄, [xAI](#) 📄, and [Anthropic](#) 📄. It has also been featured in numerous discussions online, see e.g., [here](#) 📄 and [here](#) 📄. My research on mathematical evaluation has so far led to eleven publications, two of which are highlighted here.

- [arXiv](#) 📄 **Beyond Benchmarks: MathArena as an Evaluation Platform for Mathematics with LLMs** ArXiv'26  
[Jasper Dekoninck](#), Nikola Jovanović, Tim Gehringer, Kári Rögnvaldsson, Ivo Petrov, Chenhao Sun, Martin Vechev
- An overview of the MathArena platform, which contains mathematical benchmarks on various axes, regularly kept up-to-date through detailed analyses of the latest LLMs.
- [arXiv](#) 📄 **QED-Nano: Teaching a Tiny Model to Prove Hard Theorems** ArXiv'26  
Yuxiao Qu, Amrith Setlur, [Jasper Dekoninck](#), Edward Beeching, Jia Li, Ian Wu, Lewis Tunstall, Aviral Kumar
- Part of a team that created QED-Nano, a state-of-the-art open model for mathematical theorem proving.

## Education

- 2023 – now **PhD Candidate, Computer Science**, [SRI Lab](#), [ETH Zurich](#) 📄 Zurich, Switzerland  
• Advised by: Prof. Dr. Martin Vechev and Prof. Dr. Florian Tramèr.
- 2021 – 2023 **M.Sc. Data Science**, [ETH Zurich](#) 📄, Average grade: 6/6 Zurich, Switzerland  
• First student ever to graduate with perfect GPA.  
• Master thesis on controlled text generation and research project on quantum circuit synthesis published at major conferences (ICLR and ACM OOPSLA).
- 2018 – 2021 **B.Sc. Mathematics**, [KU Leuven](#) 📄, Summa Cum Laude with compliments of the jury Leuven, Belgium  
• Part of dual degree program in Mathematics and Physics.
- 2018 – 2021 **B.Sc. Physics**, [KU Leuven](#) 📄, Summa Cum Laude Leuven, Belgium  
• Part of dual degree program in Mathematics and Physics.

## Experience

- 2026 **Machine Learning Intern, GResearch**, [GResearch](#) 📄 London, UK  
• During the summer of 2026, I worked at GResearch, a quantitative finance company.
- 2019–2021 **Software Engineer Intern, Risk Concile**, [Risk Concile](#) 📄 Leuven, Belgium  
• Each year, I spent a month working as a software engineer intern at Risk Concile, a fintech startup specializing in risk management solutions for financial institutions. Topics included portfolio optimization, volatility surface prediction, and financial shock simulations.

## Honors and Awards

- 2026 **2nd place AI4Math@ICML'26 Workshop Competition**, [AI4Math@ICML'26](#) 📄 Virtual  
• Obtained second place (\$600) in one of the four competition tracks at AI4Math@ICML'26.
- Best Submission FirstProof, Second Batch**, [FirstProof](#) 📄 Virtual  
• Part of a team that designed an agent for solving open mathematical problems. We solved 7/10 problems in the FirstProof challenge, better than all other participants, including a team containing Terence Tao.
- 2025 **Top Reviewer Award (Top 10%)**, NeurIPS 2025 San Diego, USA
- 2024 **Willi Studer Prize for Best MSc Data Science**, [ETH Zurich](#) 📄 Zurich, Switzerland  
• Prize awarded to the best master student of the year.
- 2024 **Silver Medal**, [Kaggle Challenge: AIMO Progress Prize II](#) 📄 Virtual  
• I was part of a team that became 19th out of 2213 participants.
- 2023 **Red Teaming Track Winner (\$60k Prize)**, [US Privacy-Enhancing Technologies Prize Challenge](#) 📄 Virtual  
• Awarded as part of the “ETH SRI” red team. The challenge was sponsored by NIST and NSF.
- 2021 **ETH-D Scholarship**, [ETH Zurich](#) 📄 Zurich, Switzerland  
• Full-ride scholarship for master students.
- 2015–2017 **Honors**, [Finalist Flemish Math Olympiad](#) 📄 Brussels, Belgium  
• Reached the finals (Top 100 out of ≈11000) three years in a row.

## Articles and Talks

---

|      |                                                                                                                                                                                                           |                          |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| 2026 | <b>Financial Times</b> , <a href="#">News Article</a> %                                                                                                                                                   | Virtual                  |
|      | <ul style="list-style-type: none"><li>• Interview regarding our blog post on K2-Think.</li></ul>                                                                                                          |                          |
| 2025 | <b>AI4Math@NeurIPS</b> , Panelist for the AI4Math workshop at NeurIPS 2025                                                                                                                                | San Diego, United States |
|      | <ul style="list-style-type: none"><li>• Panelist discussing the future of AI in math, alongside <a href="#">Swarat Chaudhuri</a> %, <a href="#">Dawn Song</a> %, and <a href="#">Chi Jin</a> %.</li></ul> |                          |
| 2025 | <b>NZZ</b> , <a href="#">News Article</a> %                                                                                                                                                               | Zurich, Switzerland      |
|      | <ul style="list-style-type: none"><li>• Interview for our evaluation on the International Math Olympiad in NZZ, one of the largest Swiss newspapers.</li></ul>                                            |                          |
| 2025 | <b>Nature News</b> , <a href="#">News Article</a> %                                                                                                                                                       | Virtual                  |
|      | <ul style="list-style-type: none"><li>• Interview regarding our paper on sycophancy with mathematical proofs in LLMs (BrokenMath).</li></ul>                                                              |                          |

## Teaching Experience

---

|           |                                                                                        |                     |
|-----------|----------------------------------------------------------------------------------------|---------------------|
| 2024-2026 | <b>Rigorous Software Engineering (MSc)</b> , <a href="#">ETH Zurich</a> %, Exercise TA | Zurich, Switzerland |
| 2023-2025 | <b>Reliable and Trustworthy AI (MSc)</b> , <a href="#">ETH Zurich</a> %, Exercise TA   | Zurich, Switzerland |
| 2023-2024 | <b>Software Engineering Seminar (MSc)</b> , <a href="#">ETH Zurich</a> %, Supervisor   | Zurich, Switzerland |

## Volunteering Experience

---

|           |                                                                                                                                                                                                                                                                  |                     |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 2026      | <b>Organizer</b> , <a href="#">Swiss Roundnet Championship</a> %                                                                                                                                                                                                 | Zurich, Switzerland |
|           | <ul style="list-style-type: none"><li>• The largest roundnet tournament in Switzerland, with over 150 participants.</li><li>• Part of the team organizing the tournament, including logistics, marketing, and event planning.</li></ul>                          |                     |
| 2026-now  | <b>Treasurer</b> , <a href="#">ETH Chess Club</a> %                                                                                                                                                                                                              | Zurich, Switzerland |
|           | <ul style="list-style-type: none"><li>• The ETH Chess Club organizes regular chess events.</li><li>• Responsible for event planning and organization, and managing the club's finances.</li></ul>                                                                |                     |
| 2023-2026 | <b>President</b> , <a href="#">ETH Chess Club</a> %                                                                                                                                                                                                              | Zurich, Switzerland |
|           | <ul style="list-style-type: none"><li>• Responsible for event planning and organization, and managing the board members.</li></ul>                                                                                                                               |                     |
| 2022-2023 | <b>Treasurer</b> , <a href="#">ETH Chess Club</a> %                                                                                                                                                                                                              | Zurich, Switzerland |
|           | <ul style="list-style-type: none"><li>• Responsible for event planning and organization, and managing the club's finances.</li></ul>                                                                                                                             |                     |
| 2021-2022 | <b>Core Organiser</b> , <a href="#">Student Sustainability Week</a> %                                                                                                                                                                                            | Zurich, Switzerland |
|           | <ul style="list-style-type: none"><li>• The Student Sustainability Week is a week with over 50 events on sustainability and over 1500 participants.</li><li>• Responsible for leading the core team and the organization of several events.</li></ul>            |                     |
| 2021      | <b>President</b> , <a href="#">Election Committee Stura</a> %                                                                                                                                                                                                    | Leuven, Belgium     |
|           | <ul style="list-style-type: none"><li>• Stura is the student council of KU Leuven, representing over 60,000 students in educational matters.</li><li>• Responsible for organizing the elections of the new board.</li></ul>                                      |                     |
| 2020-2021 | <b>President</b> , <a href="#">Overarching Education Council</a> %                                                                                                                                                                                               | Leuven, Belgium     |
|           | <ul style="list-style-type: none"><li>• This organization represents more than 3500 students of the Faculty of Science of KU Leuven.</li><li>• Responsible for presiding meetings, student vote in rector elections, and voicing the student concerns.</li></ul> |                     |
| 2019-2020 | <b>Board Member</b> , <a href="#">Overarching Education Council</a> %                                                                                                                                                                                            | Leuven, Belgium     |
|           | <ul style="list-style-type: none"><li>• Responsible for presiding meetings, and voicing the student concerns.</li></ul>                                                                                                                                          |                     |
| 2019-2020 | <b>Freshmen Representative and Public Relations</b> , <a href="#">WINA</a> %                                                                                                                                                                                     | Leuven, Belgium     |
|           | <ul style="list-style-type: none"><li>• Wina is the student organization for math, physics and computer science at KU Leuven.</li><li>• Responsible for business relations and organising events for freshmen.</li></ul>                                         |                     |
| 2018-2021 | <b>Student Representative Educational Committee</b> , <a href="#">POC Wiskunde</a> %                                                                                                                                                                             | Leuven, Belgium     |
|           | <ul style="list-style-type: none"><li>• The POC is responsible for filling in the student program for the Bachelor of Mathematics at KU Leuven.</li><li>• Responsible for voicing the student concerns.</li></ul>                                                |                     |

## Other Projects and Blog posts

---

|           |                                                                                                                                                                                                                                                                  |  |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| 2025      | <b>ChessImageBench</b> , <a href="#">GitHub Repository</a> %                                                                                                                                                                                                     |  |
|           | <ul style="list-style-type: none"><li>• We introduce a benchmark that shows that AI image generators fail to generate accurate chessboards.</li></ul>                                                                                                            |  |
| 2025      | <b>Debunking the Claims of K2-Think</b> , <a href="#">Blogpost</a> %                                                                                                                                                                                             |  |
|           | <ul style="list-style-type: none"><li>• We show that the reported performance of K2-Think is overstated, relying on flawed evaluation.</li></ul>                                                                                                                 |  |
| 2025-2026 | <b>MathArena Blog posts</b> , <a href="#">Blog posts</a> %                                                                                                                                                                                                       |  |
|           | <ul style="list-style-type: none"><li>• Blog posts where we introduce evaluations on, among others, the International Math Olympiad (IMO), International Math Competition (IMC), and a benchmark set of very challenging mathematical problems (Apex).</li></ul> |  |

## Publications

|      |                                                                                                                                                         |                                  |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| 2026 | <b>Beyond Benchmarks: MathArena as an Evaluation Platform for Mathematics with LLMs</b>                                                                 | AI4Math@ICML'26                  |
|      | <a href="#">Jasper Dekoninck</a> , Nikola Jovanović, Tim Gehrunger, Kári Rognvaldsson, Ivo Petrov, Chenhao Sun, Martin Vechev                           |                                  |
|      | <b>ProofCouncil: An LLM Agent for Solving Open Mathematical Problems</b>                                                                                | under review                     |
|      | Johannes Schmitt, Tim Gehrunger, <a href="#">Jasper Dekoninck</a> , Gergely Bérczi, Uri Kreitner, Liam Price                                            |                                  |
|      | <b>Optimizing the Cost-Quality Tradeoff of Agentic Theorem Provers in Lean</b>                                                                          | AI4Math@ICML'26                  |
|      | Kári Rognvaldsson, Chenhao Sun, <a href="#">Jasper Dekoninck</a> , Martin Vechev                                                                        |                                  |
|      | <b>Learning from Saturated Data: Signals Beyond Correctness for LLM Training</b>                                                                        | AI4Math@ICML'26                  |
|      | Hanno Hiss, <a href="#">Jasper Dekoninck</a> , Martin Vechev                                                                                            |                                  |
|      | <b>Delay, Plateau, or Collapse: Evaluating the Impact of Systematic Verification Error on RLVR</b>                                                      | CoLM'26                          |
|      | Kazuki Egashira, Mark Vero, <a href="#">Jasper Dekoninck</a> , Florian E. Dorner, Robin Staab, Martin Vechev                                            |                                  |
| 2025 | <b>QED-Nano: Teaching a Tiny Model to Prove Hard Theorems</b>                                                                                           | CoLM'26                          |
|      | Yuxiao Qu, Amrith Setlur, <a href="#">Jasper Dekoninck</a> , Edward Beeching, Jia Li, Ian Wu, Lewis Tunstall, Aviral Kumar                              |                                  |
|      | <b>Adaptive Generation of Bias-Eliciting Questions for LLMs</b>                                                                                         | ICML'26                          |
|      | Robin Staab, <a href="#">Jasper Dekoninck</a> , Maximilian Baader, Martin Vechev                                                                        |                                  |
|      | <b>BrokenMath: A Benchmark for Sycophancy in Theorem Proving with LLMs</b>                                                                              | ICML'26 (Spotlight)              |
|      | Ivo Petrov, <a href="#">Jasper Dekoninck</a> , Martin Vechev                                                                                            |                                  |
|      | <b>MathArena: Evaluating LLMs on Uncontaminated Math Competitions</b>                                                                                   | NeurIPS'25                       |
|      | Mislav Balunović, <a href="#">Jasper Dekoninck</a> , Nikola Jovanović, Ivo Petrov, Martin Vechev                                                        |                                  |
|      | <b>IMProofBench: Benchmarking AI on Research-Level Mathematical Proof Generation</b>                                                                    | MathAI@NeurIPS'25                |
|      | Johannes Schmitt, Gergely Bérczi, <a href="#">Jasper Dekoninck</a> , Jeremy Feusi, Tim Gehrunger, et al.                                                |                                  |
| 2024 | <b>Constrained Decoding of Diffusion LLMs with Context-Free Grammars</b>                                                                                | DL4C@NeurIPS'25 (Oral) & ICLR'26 |
|      | Niels Mündler, <a href="#">Jasper Dekoninck</a> , Martin Vechev                                                                                         |                                  |
|      | <b>A Unified Approach to Routing and Cascading for LLMs</b>                                                                                             | ICML'25                          |
|      | <a href="#">Jasper Dekoninck</a> , Maximilian Baader, Martin Vechev                                                                                     |                                  |
|      | <b>MathConstruct: Challenging LLM Reasoning with Constructive Proofs</b>                                                                                | ICML'25                          |
|      | Mislav Balunović*, <a href="#">Jasper Dekoninck</a> *, Nikola Jovanović, Ivo Petrov, Martin Vechev                                                      |                                  |
|      | <b>Proof or Bluff? Evaluating LLMs on 2025 USA Math Olympiad</b>                                                                                        | AI4Math@ICML'25                  |
|      | Ivo Petrov, <a href="#">Jasper Dekoninck</a> , Lyuben Baltadzhiev, Maria Drencheva, Kristian Minchev, Mislav Balunović, Nikola Jovanović, Martin Vechev |                                  |
|      | <b>The Open Proof Corpus: A Large-Scale Study of LLM-Generated Mathematical Proofs</b>                                                                  | ICLR'26                          |
|      | <a href="#">Jasper Dekoninck</a> *, Ivo Petrov*, Kristian Minchev, Mislav Balunovic, Martin Vechev, et al.                                              |                                  |
| 2024 | <b>Polyrating: A Cost-Effective and Bias-Aware Rating System for LLM Evaluation</b>                                                                     | ICML'25                          |
|      | <a href="#">Jasper Dekoninck</a> , Maximilian Baader, Martin Vechev                                                                                     |                                  |
|      | <b>ConStat: Performance-Based Contamination Detection in Large Language Models</b>                                                                      | NeurIPS'24                       |
|      | <a href="#">Jasper Dekoninck</a> , Mark Niklas Müller, Martin Vechev                                                                                    |                                  |
|      | <b>Controlled Text Generation via Language Model Arithmetic</b>                                                                                         | ICLR'24 (Spotlight)              |
|      | <a href="#">Jasper Dekoninck</a> , Marc Fischer, Luca Beurer-Kellner, Martin Vechev                                                                     |                                  |
|      | <b>Synthetiq: Fast and Versatile Quantum Circuit Synthesis</b>                                                                                          | OOPSLA'24                        |
|      | Anouk Paradis*, <a href="#">Jasper Dekoninck</a> *, Benjamin Bichsel, Martin Vechev                                                                     |                                  |
|      | <b>Evading Data Contamination Detection for Language Models is (too) Easy</b>                                                                           | ArXiv                            |
|      | <a href="#">Jasper Dekoninck</a> , Mark Niklas Müller, Maximilian Baader, Marc Fischer, Martin Vechev                                                   |                                  |