

## AI for Math Summer Fellowship – Constraining LLMs for Theorem Proving

### Project Details

- **Project Title:** Constraining LLMs for Theorem Proving: A Neurosymbolic Approach to Guaranteed Autoformalization
- **Eligibility Criteria:** Open to PhD students.
- **Location:** In-person at Imperial College London. Renaissance Philanthropy will cover travel costs if necessary for fellows who need to relocate temporarily for the placement.
- **Host Organization / Institution:** Imperial College London
- **Project Lead:** Dr. Eleonora Giunchiglia
- **Project Mentor:** Mihaela Cătălina Stoian

**About the Project.** The goal of this project is to advance autoformalization—the task of translating informal mathematics into formal statements that can be automatically checked by state-of-the-art proof assistants like Lean. As even state-of-the-art models only achieve a modest success rate in generating syntactically correct Lean statements, this project will develop novel neural modules that can produce correct Lean code by construction. These modules integrate seamlessly into existing LLM architectures and can be jointly trained, thus enhancing overall performance.

**About Imperial College London.** As a world-leading institution, Imperial College London fosters a multi-disciplinary research culture dedicated to fundamental research and global technological leadership. Within this environment, our research group focuses on the integration of complex constraints directly into neural architectures with the goal of guaranteeing constraint satisfaction while actively improving the performance of the underlying models. Backed by a strong publication record in top-tier AI venues such as NeurIPS and ICLR, our team bridges neural networks and reasoning to develop reliable and trustworthy AI systems for real-world applications.

### Role and Responsibilities

The fellow will support ongoing research on constraining large language models for theorem proving and autoformalization using context-free grammars derived from formal systems.

Responsibilities will include:

- Implementing grammar-constrained decoding methods for large language models
- Integrating these methods into existing LLM-based theorem proving pipelines
- Running experiments to evaluate how well the constraints improve the generation of syntactically valid formal statements and proofs
- Assisting with analysis and documentation of experimental results

## Key Qualifications

### Essential

- Strong programming skills in Python
- Experience with machine learning or large language models

### Desirable

- Familiarity with formal languages or context-free grammars
- Background knowledge in logic or theorem proving systems (e.g., Lean, Coq)

## Deliverables

By the end of the fellowship, the fellow is expected to contribute:

- Contribute to codebase: a fully documented, open-source implementation of the contributed method
- Experimental benchmarks and analysis: a comprehensive evaluation report detailing the performance of the constrained model against baseline autoregressive models
- Research paper contribution: a formal write-up of the methodology, experimental setup, and findings, intended to serve as a foundational draft or core contribution to a paper submission for top-tier AI venues (e.g., NeurIPS, ICLR)
- End-of-summer presentation: a final presentation summarising the project's technical challenges, solutions, and overall impact on automated theorem proving

## How to Apply

Submit the following materials through [this application link](#):

- CV / Resume
- Statements of interest and relevant expertise (1–2 paragraphs)
- Optional: links to GitHub, research papers, or prior technical work