浙江大學 ZHEJIANG UNIVERSITY　厦門大學 XIAMEN UNIVERSITY　上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

AI4MATH

Workshop

数学智能研讨会

2026.07.22 WED. - 07.24 FRI.

www.iasm.zju.edu.cn

数学高等研究院 | www.iasm.zju.edu.cn

Schedule

2026.7.22 星期三 Wed.

上午	开幕式 + 主旨报告	
08:30 - 08:50	签到	
08:50 - 09:00	开幕致辞	
09:00 - 09:45	**董　彬**	北京大学
09:45 - 10:30	**文再文**	北京大学
10:30 - 10:50	茶歇 / 合影留念	
10:50 - 11:35	**梁经纬**	上海交通大学
11:35 - 13:30	午餐与休息	
下午	**前沿分享**	
13:30 - 14:00	**曹一川**	中国科学院
14:00 - 14:30	**关乃鄴**	北京大学
14:30 - 15:00	**徐霄乾**	昆山杜克大学
15:00 - 15:20	茶歇	
15:20 - 15:50	**Ajay Kumar Nair**	印度科学学院
15:50 - 16:20	**邱瑞晨**	中国科学院
16:20 - 16:50	**赵雨扬**	香港中文大学(深圳)
16:50 - 17:20	**王语同**	至知科技

2026.7.23 星期四 Thur.

上午	主旨报告	
09:00 - 09:45	**李　鹏**	清华大学
09:45 - 10:30	**李　嘉**	Project Numina
10:30 - 10:50	茶歇	
10:50 - 11:35	**陈小杨**	同济大学
11:35 - 13:30	午餐与休息	

Schedule

2026.7.23 星期四 Thur.

下午	前沿分享	
13:30 - 14:00	**沈颖祺**	香港中文大学(深圳)
14:00 - 14:30	**王浩丞**	苏黎世联邦理工学院 / 香港科技大学(广州)
14:30 - 15:00	**刘方辉**	上海交通大学
15:00 - 15:20	茶歇	
15:20 - 15:50	**徐启源**	南洋理工大学
15:50 - 16:20	**刘晓洋**	上海交通大学
16:20 - 16:50	**杨　柳**	新加坡国立大学

2026.7.24 星期五 Fri.

上午	实操教学&工具快闪	
	A Practical Tutorial	
09:00 - 09:40	**居浩成**	北京大学
09:40 - 10:20	**吴　彬**	北京大学
10:20 - 10:40	茶歇	
10:40 - 11:20	**杨　柳**	新加坡国立大学
	Lightning Session	
11:20 - 12:00	**董子能**	上海交通大学
	许景宣	上海交通大学
	刘云天	上海交通大学
	邹扬硕	加州大学伯克利分校
12:00 - 13:00	闭幕、午餐、离会	

2026.7.22 星期三 Wed.

- **董　彬　　北京大学**

Title:

AI for Mathematics：数学智能体的设计与实践

Abstract:

人工智能正在推动数学研究范式的演进，但其深度参与数学研究仍面临若干关键挑战。数学研究高度专业化，不同领域在概念体系、符号表述和文献传统上存在差异，使潜在的跨领域关联难以被有效发现；与此同时，自然语言论证与机器可验证的形式化证明之间仍存在显著鸿沟。本报告将介绍北京大学课题组围绕数学信息检索、检索增强的自动形式化与自动推理所构建的一系列数学智能体，重点阐述其设计理念、系统架构与实践方法，并通过若干开放数学问题的典型案例，展示这些智能体在知识发现、问题形式化和推理探索中的作用。最后，报告将对人工智能驱动数学研究的未来发展方向作简要展望。

- **文再文　　北京大学**

Title:

智能数学推理：形式化与定理证明

Abstract:

本报告探讨数学形式化和自动定理证明方面的一些进展：面向大规模Lean库，发展基于表达式树结构和图神经网络的前提选择方法，提高定理检索与证明搜索效率；提出M2F自动形式化框架，将数学教材和论文转化为可编译、可审计、可复用的Lean项目，并建设ReasBook形式化知识库；提出FaithSieve，通过形式化证据和语义对齐机制引导数学证明审稿与错误定位；发展OptProver，通过优化领域数据和偏好学习增强模型在形式化定理证明中的专业能力；构建CAM-Bench，面向优化、数值线性代数和数值分析建立应用数学形式化测评体系。最后介绍智能数学推理引擎ReasLab，展示其在定理证明和科学写作中的一体化能力。

Title & Abstract

- **梁经纬　　上海交通大学**

Title:

Paper2Formalization: Auto-Formalization of Research Papers

Abstract:

In this talk, I will present a paper to formal agentic framework for automated formalization of math research papers in Lean. The framework contains three agents: an ocr agent with LLM supervision, a plan agent based on knowledge-graph, and an automated formalization agent. The proposed agentic framework is tested over 30 papers on optimization theory, producing a Lean library of more than 110,000 lines of codes.

• 曹一川　中国科学院

Title:

MechMath Agent Team: LLM Driven Agents for Mathematical Research

Abstract:

AI reasoning has become a central focus in contemporary artificial intelligence, largely driven by the success of large language models. However, mathematical research, which is characterized by non-linear derivation paths, rigorous logical requirements, and protracted exploration cycles, poses severe challenges for existing reasoning systems. To overcome these limitations, we present the MechMath Agent Team (MMAT), which is a large language model driven agent designed to serve as a co-pilot throughout the full cycle of mathematical research. We design a tripartite Harness Architecture that decouples system responsibilities into Control, Execution, and Augmentation planes, thereby reconciling rigorous logical control with the agility demanded by open-ended research. Building upon this framework, we instantiate three specialized agents: a Knowledge Base Manager, a Natural Language Prover, and a Formal Language Prover, all operating in a closed loop to produce formally certified mathematical proofs. We evaluate MMAT on open problems in Number Theory, Algebraic Complexity Theory, Differential Algebra, Operator Algebra, and Inequalities. Across a two-month deployment, 11 problems have been solved, demonstrating its capacity to act as a co-pilot throughout the entire research cycle. The contributions are threefold: a general decoupled Harness Architecture for multi-agent mathematical reasoning, its concrete instantiation in the MMAT system, and empirical validation on a diverse suite of open problems.

Title & Abstract

• 关乃鄴　北京大学

Title:

Formalization of Hierarchy of Local Rings in Lean4

Abstract:

This talk reports on a formalization project in the Lean4 theorem prover for the hierarchy of local rings in commutative algebra. We develop the theories of depth and Cohen--Macaulay rings, proving the Auslander--Buchsbaum and Ischebeck theorems, and obtain a homological characterization of regular local rings via maximal Cohen--Macaulay modules and a weakened version of the Ferrand--Vasconcelos theorem. We then investigate Gorenstein rings, showing that they are Cohen--Macaulay of type one, and complete intersection rings using the epsilon-one invariant and Koszul complex. A key technical contribution is the construction of Ext maps induced by exact functors, which enables dimension--shifting arguments. The formalization also includes the Koszul complex that benefits from AI-assisted proving for exterior algebra and base-change lemmas. The overall result establishes full formalization of hierarchy of local rings and some related theories.

• 徐霄乾　昆山杜克大学

Title:

AI Proofs in PDEs

Abstract:

In this talk, I will discuss recent results on fluid equations done by AI. I will first outline the relationship between the mixing properties of incompressible flows and the diffusion process, including dissipation enhancement. I will then focus on the so-called unmixing phenomenon and on the existence of Batchelor scales for certain flows, both of which are results recently proved using AI.

• **Ajay Kumar Nair　印度科学院**

Title:

LeanAide: Extensible Autoformalization by Combinator Code Generation

Abstract:

Autoformalization is the process of automatically translating natural language mathematics into machine-verifiable code. The most successful machine-verifiable formats in this space are theorem provers, among which Lean has found much success as an interactive theorem prover in recent times.

LeanAide is an approach to autoformalization into the Lean theorem prover that offers modularity and extensibility in many aspects such as using AI models of one's choice and adding combinators to suit specific domains. While strong autoformalization agents are emerging, our project stands out as an open-source, extensible system written primarily in Lean, leveraging its wonderful metaprogramming framework. This is achieved by writing various handlers for different kinds of mathematical constructs and composing those handlers to generate the final output. We call this novel approach *Combinator Code Generation*. LeanAide is architected such that any improvements in rule-based automation (e.g., `grind`) or large language models will directly improve its autoformalization capabilities.

Funding

We gratefully acknowledge funding from Renaissance Philanthropy under their AI for Math Fund.

Title & Abstract

• **邱瑞晨　中国科学院**

Title:

MechMath: Sorrifier-Driven Formal Decomposition Workflow for Automated Theorem Proving

Abstract:

Recent advances in large language models (LLMs) and LLM-based agents have substantially improved the capabilities of automated theorem proving. However, for problems that require complex mathematical reasoning, current systems rarely succeed on the first try and must repeatedly modify their proof strategies. Existing approaches for handling failed attempts typically either iteratively fix errors within the proof or discard the entire proof and regenerate it from scratch. The former leads to progressively longer contexts, which progressively degrade the model's ability to attend to the remaining unresolved subproblems, while the latter is inefficient, as it may abandon mostly correct reasoning due to localized errors. To address this dilemma, we propose MechMath, a novel agent system that employs a sorrifier-driven formal decomposition strategy. By leveraging the sorry placeholder in Lean to precisely isolate unresolved subgoals while preserving the surrounding verified proof structure, MechMath extracts each failed subproblem into a clean, self-contained context and resolves it independently. This avoids both the waste of full regeneration and the excessive context length induced by repeated repairs. Experimental results on challenging mathematical competition benchmarks, including IMO 2025, Putnam 2025, miniF2F, and a subset of ProverBench, demonstrate that our agent achieves significant advantages in proving efficiency.

• **赵雨扬　香港中文大学(深圳)**

Title:

Mathlib 类型类层次结构的设计

Abstract:

类型类是 mathlib 中广泛使用的重要机制, 它极大地方便了形式化过程中对数学对象的操作。但随着层次结构不断加深、加宽, 这套机制在设计与使用中也带来了不少困难。本报告将介绍 mathlib 类型类层次结构的设计、遇到的问题以及相应的应对方法。

报告首先介绍类型类的组织方式, 继而讨论随着层次结构的增长与复杂化而在实践中显现的种种问题, 例如实例的菱形(diamond)与推断性能等, 并说明针对这些问题的应对方法。报告最终指出, 形式化系统中一个结构应携带哪些数据, 往往不取决于数学上的最简或最习惯的表述, 而取决于 API 的可维护性与效率。要在系统中方便而高效地做数学, 背后需要这样细致的设计权衡。

• **王语同　至知科技**

Title:

Reap: 轻量、开放的 AlphaProof 复现

Abstract:

近年来, 基于 MCTS 强化学习的自动定理证明系统取得了巨大进展, 展现出超越人类的证明撰写和逻辑推理能力。我们团队基于 reap tactic, 在 Lean 中原生实现了一套高效的证明搜索系统, 并搭建完整训练流程, 在小模型、低预算上取得了领先的性能。在这次报告中我们将简要分享我们的构建经验和发现。

Title & Abstract

2026.7.23 星期四 Thur.

• **李　鹏　清华大学**

Title:

自然语言驱动的数学研究智能体——AIM系统的探索与实践

Abstract:

数学作为科学进步的重要基石, 长期支撑着各学科的发展与突破。近年来, 人工智能的迅猛发展正在推动科学研究范式发生深刻变革, 也为数学研究带来了新的机遇与挑战。当 AI 遇见数学, 尤其是当自然语言成为数学研究智能体理解问题、组织推理与协同探索的重要界面时, 将会产生怎样的化学反应?本报告将围绕“自然语言驱动的数学研究智能体”这一主题, 简要回顾 AI 在数学研究中的发展脉络与代表性进展, 分析 AI 面向研究级数学问题时所面临的核心挑战。在此基础上, 报告将重点介绍本团队研发的 AI 数学家系统 AIM, 展示其如何通过自然语言交互与智能体协作机制, 辅助数学问题探索、定理形成与可信验证。

Title & Abstract

- **李　嘉　Project Numina**

Title:

大模型如何学会数学：从思维链到形式化证明

Abstract:

大语言模型最初通过预测下一个词学习语言，却在过去几年中迅速获得了求解数学问题、生成复杂推理乃至构造形式化证明的能力。这一进展并非单纯来自模型规模的增长，而是由高质量数学数据、思维链监督、过程奖励、强化学习、工具调用和测试时计算等方法共同推动。本报告将回顾大模型数学能力的发展历程：从 MATH、Minerva 等早期工作，到代码辅助推理、过程奖励模型和大规模合成数据，再到 o1、DeepSeek-R1 所代表的长程推理与强化学习范式。结合 Project Numina 的实践，我们将介绍如何通过 NuminaMath 数据集与开放模型训练数学推理能力，以及如何在 Kimina Prover 中将证明搜索轨迹转化为结合自然语言推理与 Lean 代码的统一推理过程，并利用强化学习提升形式化证明能力。最后，报告将讨论数学 AI 正在经历的关键转变：从解答有标准答案的竞赛题，走向生成可机器验证的证明、形式化已有数学成果，并进一步参与开放问题和前沿数学研究。

- **陈小杨　同济大学**

Title:

AI大模型的创造性思维

Abstract:

本报告探讨AI大模型的创造性思维，深入讨论创造性思维的评估，如何设计智能体探索创造性思维，以及进一步提升创造性思维的潜在训练框架。

- **沈颖祺　香港中文大学(深圳)**

Title:

Formalizing Extended Complex Numbers, Mobius Transformations, and Cross Ratio in Lean 4

Abstract:

The extended complex plane is a fundamental object in complex analysis, hyperbolic geometry, and mathematical physics. Its geometry is governed by Möbius transformations, with the cross ratio serving as a central invariant. We present a formalization of these concepts in the Lean4 theorem prover. The extended complex plane is represented using Mathlib's Option type over C, where the additional element represents the point at infinity. On this foundation, we define Möbius transformations, their action on the extended complex plane, and the cross ratio. We formalize several basic properties of Möbius transformations, including their group structure, and identify them with a projective general linear group. We also prove the uniqueness of a Möbius transformation mapping any three distinct points to any other three distinct points, and the invariance of the cross ratio. All proofs are machine-checked in Lean 4. The complete development comprises approximately 6,000 lines of Lean code, including about 40 definitions and 150 lemmas and theorems. This work provides a verified foundation for future formalizations of conformal geometry, hyperbolic models, modular forms, and applications in mathematical physics.This is a joint work with Fubin Yan.

• **王浩丞　苏黎世联邦理工学院(香港科技大学(广州)**

Title:

FormalRx: Rectify and eXamine Semantic Failures in Autoformalization

Abstract:

The veracious semantic alignment in autoformalization is significant for formal mathematical reasoning. However, existing evaluations provide only opaque binary verdicts or scalar scores, offering no interpretable insight into where or why translations fail. This opacity severely limits both human understanding and automated system improvement. To bridge this gap, we introduce FormalRx, a comprehensive diagnostic evaluation framework that transforms autoformalization assessment from black-box judgments into actionable feedback. At its core is SCI Error Taxonomy, a hierarchical classification scheme decomposing autoformalization errors into 28 distinct categories with strict priority ordering. Building on this taxonomy, FormalRx provides four critical diagnostic capabilities: alignment verdicts, error categorization, error localization, and correction. We instantiate the framework with a diagnostic model FormalRx-8B, trained on 56,287 synthetically generated samples with fine-grained diagnostic annotations, and release FormalRx-Test as the first fine-grained diagnostic benchmark. FormalRx-8B achieves F1-scores of 0.88 (verdict) and 0.71 (categorization), along with accuracies of 0.75 (localization) and 0.73 (correction), substantially outperforming both general-purpose LLMs and specialized baselines. By connecting evaluation with actionable insights, FormalRx enables systematic diagnosis and improvement of autoformalization systems.

Title & Abstract

• **刘方辉　上海交通大学**

Title:

LeanMarathon: Toward Reliable AI Co-Mathematicians through Long-Horizon Lean Autoformalization (长程数学推理的自动形式化系统)

Abstract:

In this talk, I will discuss LeanMarathon, a multi-agent harness for reliable research-level Lean autoformalization. Its core abstraction is an evolving blueprint: a Lean file that serves simultaneously as formal proof skeleton, natural-language proof graph, and shared system of record. Four contract-scoped agents construct, audit, prove, and repair this blueprint. These agents are coordinated by a two-stage orchestrator that first stabilizes target fidelity through adversarial review and then discharges the proof directed acyclic graph (DAG) from its dynamic leaves upward in parallel CI-gated rounds. We evaluate LeanMarathon on two recent research papers spanning four Erdős problems (#1051, #1196, #164, #1217). Across three autonomous runs, it formalizes all seven target theorems with no sorry, proving 258 lemmas and theorems. These results show that reliable AI co-mathematics requires not only stronger provers, but durable harnesses that preserve target fidelity across long mathematical developments.

• **徐启源　南洋理工大学**

Title:

由 LLM 驱动的富逻辑程序验证的变革前夜

Abstract:

本报告将从程序验证在现实软件工程领域的主要堵点(一个不可能三角)出发, 讨论 LLM 与 Neural Theorem Proving 将如何改变这一不可能三角并解决这些堵点为程序验证全面变革现有软件工程奠定可能性基础。此外, 本报告将与 AI4Math 社区常用的 Lean 语言结合, 在介绍程序验证的相关方法论的同时, 重点澄清在 Lean 语言上进行程序验证的常见误区与概念混淆。

- **刘晓洋　上海交通大学**

Title:

从数学命题到代码规约的形式化与评估

Abstract:

报告从一个学生的角度谈谈在 AI4Math 方向上的一些探索：从基础的形式化问题出发，尝试借助 AI 技术来助力可验证性。首先，介绍从自然语言数学命题到 Lean 形式化表达的数据合成与模型训练。其次，讨论尝试结合结构信息与 Lean 证明步骤而构造的模型评估指标。最后，分享在代码规约评测方面的初步探索，考察现有 benchmark 是否可能高估模型能力。

- **杨　柳　新加坡国立大学**

Title:

Evolutionary Ensemble of Agents (EvE): A self-evolving agent for math proof, algorithm discovery, and beyond

Abstract:

We introduce Evolutionary Ensemble (EvE), a decentralized framework that organizes existing, highly capable coding agents into a live, co-evolving system for math proof, algorithm discovery, and beyond. By maintaining two co-evolving populations, namely downstream task solvers and agent guidance states, the system keeps evolving the agents based on the marginal gains they contribute to the solvers. Without human intervention, EvE was able to solve a mathematical problem in just one day that previously took one PhD student an entire year to complete. We will also show some results of applying EvE to algorithm discovery, control design, and beyond.

EvE is open-source and actively maintained: https://github.com/scaling-group/eve

Title & Abstract

2026.7.24 星期五 Fri.

- **居浩成　北京大学**

Title:

Automated Reasoning Agents for Open Problems in Mathematics

Abstract:

Recent LLM-based mathematical reasoning agents have begun to tackle research-level problems and, in several cases, have contributed to the resolution of open problems. In this talk, we will discuss the design and implementation of such agents, as well as how they can be orchestrated into effective workflows for mathematical research. We will illustrate these ideas through several recent systems and examples of their applications to open problems in mathematics.

- **吴　彬　北京大学**

Title:

面向形式化数学的智能体

Abstract:

本报告将介绍形式化数学智能体 Archon 的使用，并结合具体案例展示其在形式化证明生成、验证和改进等任务中的应用。

Title & Abstract

- **杨　柳　　新加坡国立大学**

Title:

EvE 的使用

Abstract:

本报告将介绍自进化多智能体系统　EvE　的基本使用方法, 并展示其在数学证明、算法发现等任务中的应用。

- **工具快闪**

董子能　　上海交通大学

LeanExplain：面向 Lean 4 的自然语言解释工具；

许景宣　　上海交通大学

绎理MathWeaver：让数学文档成为可理解、可连接的知识图谱；

刘云天　　上海交通大学

结构化自然语言SNL：形式化项目文档管理；

邹扬硕　　加州大学伯克利分校

法言Fyan：面向计算数学形式化的人机协同工具

组织：孙斌勇（浙江大学） 罗涛（上海交通大学） 马家骏（厦门大学）
主办：浙江大学数学高等研究院
协办：上海交通大学 厦门大学
*本次会议得到新基石科学基金会资助