

MathNLP 2025

**MathNLP 2025: The 3rd Workshop on Mathematical Natural
Language Processing**

Proceedings of the Workshop

November 9, 2025

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-348-7

MathNLP 2025

The articulation of mathematical arguments is a fundamental part of scientific reasoning and communication. Across many disciplines, expressing relations and interdependencies between quantities is at the centre of scientific argumentation. Nevertheless, despite its importance, the application of contemporary NLP models for inference over mathematical text remains under-explored or subject to important limitations. MathNLP represents a forum for discussing new ideas to advance research on Mathematical Natural Language Processing, welcoming novel contributions on model architectures, evaluation methods and downstream applications.

Recent advances in Natural Language Processing (NLP) enabled by Deep Learning-based architectures bring the opportunity to support the interpretation of textual content at scale. The application of these methods can facilitate scientific discovery, reducing the gap between current research and the available large-scale scientific knowledge. Previous work has shown the potential of designing neural architectures for different mathematical natural language inference tasks, such as premise selection in natural language, expression derivation, and mathematical information retrieval.

However, there are still technical gaps that need to be addressed such as the availability of datasets and evaluation tasks, techniques for the joint interpretation of different modalities present in mathematical text (equational and natural language), the understanding of unique aspects of mathematical discourse and multi-hop models for mathematical inference.

We proposed this workshop as a continuation of our previous editions, with a new emphasis on the integration of Large Language Models (LLMs) and symbolic approaches with the goal of addressing these challenges and connect different experts in this field.

In this edition, MathNLP accepted a total of 40 papers, of which 16 are included in the proceedings. For additional details about MathNLP 2025, please visit the website: <https://sites.google.com/view/mathnlp2025>.

Organizing Committee

General Chair

Marco Valentino, University of Sheffield

Program Chairs

Deborah Ferreira, The MathWorks

Mokanarangan Thayaparan, The MathWorks

Leonardo Ranaldi, University of Edinburgh

Andre Freitas, Idiap Research Institute and University of Manchester

Program Committee

Program Committee

Kuan Lu
Lei Xu
Surangika Ranathunga
Pierre Beckmann
Xin Quan
Nisansa de Silva
Federico Ranaldi
Aishwarya Jadhav
Pinzhen Chen
Mardhiyah Sanni
Yingji Zhang
Giulia Pucci
Lan Zhang
Joshua Ong Jun Leang
Xingwei Tan
Matt Freestone
David L. J. Ho
Wenda Li
Gautam Pendse

Table of Contents

<i>Syntactic Blind Spots: How Misalignment Leads to LLMs' Mathematical Errors</i>	
Dane A Williamson, Yangfeng Ji and Matthew B. Dwyer	1
<i>Step-KTO: Optimizing Mathematical Reasoning through Stepwise Binary Feedback</i>	
Yen-Ting Lin, Di Jin, Tengyu Xu, Tianhao Wu, Sainbayar Sukhbaatar, Chen Zhu, Yun He, Yun-Nung Chen, Jason E Weston, Yuandong Tian, Arash Rahnema, Sinong Wang, Hao Ma and Han Fang	15
<i>BloomWise: Enhancing Problem-Solving capabilities of Large Language Models using Bloom's-Taxonomy-Inspired Prompts</i>	
Maria-Eleni Zoumpoulidi, Georgios Paraskevopoulos and Alexandros Potamianos	34
<i>Scalability of LLM-Based Multi-Agent Systems for Scientific Code Generation: A Preliminary Study</i>	
Yuru Wang, Kaiyan Zhang, Kai Tian, Sihang Zeng, Xingtai Lv, Ning Ding, Biqing Qi and Bowen Zhou	50
<i>FIRMA: Bidirectional Formal-Informal Mathematical Language Alignment with Proof-Theoretic Grounding</i>	
Maryam Fatima	62
<i>CHECK-MAT: Probing the Mathematical Reasoning and Rubric-Alignment of Vision-Language Models on Handwritten Solutions</i>	
Ruslan Khrulev	77
<i>RoMath: A Mathematical Reasoning Benchmark in Romanian</i>	
Adrian Cosma, Ana-Maria Bucur and Emilian Radoi	95
<i>Into The Limits of Logic: Alignment Methods for Formal Logical Reasoning</i>	
Francisco Fernando Lopez-Ponce and Gemma Bel-Enguix	112
<i>Formula-Text Cross-Retrieval: A Benchmarking Study of Dense Embedding Methods for Mathematical Information Retrieval</i>	
Zichao Li	124
<i>BanglaMATH : A Bangla benchmark dataset for testing LLM mathematical reasoning at grades 6, 7, and 8</i>	
Tabia Tanzin Prama, Christopher M. Danforth and Peter Dodds	134
<i>Logically Constrained Decoding</i>	
Franklin Ma and Alan J. Hu	150
<i>Modeling Tactics as Operators: Effect-Grounded Representations for Lean Theorem Proving</i>	
Elisaveta Samoylov and Soroush Vosoughi	168
<i>UniMath-CoT: A Unified Framework for Multimodal Mathematical Reasoning with Re-Inference Affirmation</i>	
Zhixiang Lu, Mian Zhou, Angelos Stefanidis and Jionglong Su	176
<i>An in-depth human study of the mathematical reasoning abilities in Large Language Models</i>	
Carolina Dias-Alexiou, Edison Marrese-Taylor and Yutaka Matsuo	186
<i>Synthetic Proofs with Tool-Integrated Reasoning: Contrastive Alignment for LLM Mathematics with Lean</i>	
Mark Obozov, Michael Diskin, Aleksandr Beznosikov, Alexander Gasnikov and Serguei Barannikov	195