

MathNLP 2026

**MathNLP 2026: The 4th Workshop on Mathematical Natural
Language Processing**

Proceedings of the Workshop

October 29, 2026

©2026 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 979-8-89176-468-2

MathNLP 2026

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.

For additional details about MathNLP 2026, please visit the website: <https://sites.google.com/view/mathnlp2026>.

Organizing Committee

General Chair

Marco Valentino, University of Sheffield

Program Chairs

Deborah Ferreira, The MathWorks

Mokanarangan Thayaparan, The MathWorks

Leonardo Ranaldi, University of Edinburgh

Jo-Ku Cheng, University of Sheffield

Andre Freitas, Idiap Research Institute and University of Manchester

Program Committee

Program Committee

Joshua Ong Jun Leang
Feng He
Zhonglin Xie
Mingzi Cao
Yanyu Chen
Yinjie Cheng
Yanwen Peng
Jiajun Chen
Jingyun Wang
Jingyang Deng
Hikaru Matsuoka
Constantinos Karouzos
Koji Mineshima
Jordan Meadows
George Chrysostomou
Carlos Toxtli
Xin Quan
Xingwei Tan
Lan Zhang
Jasivan Alex Sivakumar
Lei Xu
Somak Aditya
Surangika Ranathunga
Dhairya Dalal
Qianying Liu
Fajri Koto
Nischal Reddy Chandra
Pierre Beckmann
Cheuk Lun Ko

Table of Contents

<i>Beyond Better Helpers: A Shared-Kernel Study of Symbolic Interfaces for Auditable LLM Algebra</i> Norio Kosaka, Genki Osada, Shinichi Higashino and Shuji Yamaguchi	1
<i>Hard or Just Unreached? Diagnosing the Sampling Blind Spot in Math-Reasoning Difficulty Estimation</i> Luca Zhou, Sajel Shah, Emanuele Rodolà and Roberto Dessi	13
<i>BlueprintRepair: Typed Local Edits for Failed Lean Proof Blueprints</i> Ruslan Khrulev	30
<i>From Recognition to Reconstruction: Towards Robust Procedural Mathematical Reasoning</i> Fateme Mazdarani and Carlos Toxtli	42
<i>Elision Markers in Mathematical Prose: Language Models Respond to Claimed Effort Rather Than Hidden Derivation Depth</i> Md. Faisal Sheikh	57
<i>AMIR-GRPO: Inducing Implicit Preference Signals into GRPO</i> Amir Hossein Yari and Fajri Koto	69
<i>The Cost of Incoherent Context Grows with Capability: A Control for Exemplar Retrieval</i> Srinivas Harish	86
<i>InvariantSpec: Benchmarking LLM Autoformalization of Conservation and Dissipation Laws into Verifiable State-Space Constraints</i> Hikaru Matsuoka	101
<i>MathQueue: Reasoning Scheduling Helps Only Within a Narrow Budget, and Answer Extraction Decides Whether It Helps at All</i> Jonghyun Lee and Byungjun Kim	112
<i>Typed Neuro-Symbolic Autoformalization of Quantitative Clinical Language: Executable Constraints and Selective Verification</i> Aleksandra Beliaeva	126
<i>Which Mathematical Context Supports Target-Aligned Conjecture Generation? A Paired Ablation Study</i> Austin LaHue, Michael Burr, Luis David Garcia Puente, Vinita Gangaram Jansari, Benjamin E Nye and Carlos Toxtli Hernandez	136
<i>The Connective but in Informal Mathematical Proof: A Corpus-Based Study</i> Rinta Yamada, Sakiko Yamasaki, Risako Ando, Takayuki Amamoto, Ryota Akiyoshi and Koji Mineshima	152
<i>NaturalPRISM: A Natural Language Premise Retrieval Task for Research-level Intermediate Proof States</i> Harris Proctor, LI AN, Austin LaHue, Michael Burr, Vinita Gangaram Jansari, Carlos Toxtli Hernandez, Luis David Garcia Puente and Benjamin E Nye	167
<i>Improving Natural-Language Combinatorial-Optimization Accuracy in Resource-Constrained Language Models via Formal Abstractions</i> Shrenil Shaun Sharma and Avi Sharma	177
<i>Structural Signatures of Correct Mathematical Reasoning Traces</i> Maria Marina and Alexander Panchenko	194

<i>Simulating Student-Tutor Interactions with LLMs for Multilingual Math Feedback</i>	
Junior Cedric Tonga, Kv Aditya Srivatsa, Kaushal Kumar Maurya, Fajri Koto and Ekaterina Kochmar	201
<i>Beyond Accuracy: Diagnosing Algebraic Reasoning Failures in LLMs Across Nine Complexity Dimensions</i>	
Parth Shrikant Patil, Yash Sinha, Murari Mandal and Dhruv Kumar	216
<i>Same Text, Different Transitions: Action Insufficiency in Lean Provers</i>	
Edison Yang, Neel Marripalapu, Akshaj Agadi and Andrew Kim	240
<i>Scoring-Location Sensitivity in Process Reward Models</i>	
Jai Kumar Sharma	254
<i>Embedding-Based Educational Taxonomies</i>	
Kia Karbasi, Kaveh Karbasi and Gregory Pottie	268
<i>Numbers Through Signs: How LLMs Connect Counts, Number Words, and Numerals</i>	
J. Albert Yoo	285
<i>Definition Modeling for Mathematics: Understanding Audience Bias and its Problem Solving Relevance</i>	
Carolina Dias-Alexiou, Edison Marrese-Taylor, A. Solano and Yutaka Matsuo	291
<i>Who Should Decide Mathematical NLI? Auditing Evidence Selection, Solver Gating, and Benchmark Shortcuts</i>	
Soumyadeep Nag and Suparnojit Sarkar	305
<i>Proof Identity and Divisible Correspondence in Multi-Derivation Reasoning Evaluation</i>	
Toprak Seda Karaosmanoğlu	312