

LREC-COLING 2024

**MathNLP: The 2nd Workshop on Mathematical Natural  
Language Processing  
@LREC-COLING 2024**

Workshop Proceedings

Editors  
Marco Valentino

21 May, 2024  
Torino, Italia

**Proceedings of MathNLP: The 2nd Workshop on Mathematical Natural Language Processing @ LREC-COLING 2024**

Copyright ELRA Language Resources Association (ELRA), 2024  
These proceedings are licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-22-7  
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association  
and the International Committee on Computational Linguistics

## Message from the Workshop Organizers

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. MathNLP welcomed contributions of previously unpublished papers which could be either long (8 pages) or short (4 pages). MathNLP welcomed both archival and non-archival submissions. Only archival submissions have been included in the proceedings. All submissions have been peer-reviewed by 2 independent reviewers. A total of 5 papers have been accepted for presentation at the workshop. MathNLP is particularly interested in (but is not limited to) works related to the following topics:

- Neural/Neuro-symbolic architectures to support mathematical natural language inference;
- Large Language Models for Mathematics;
- Equational embeddings;
- Autoformalisation and translation from natural language to formal languages (and vice-versa);
- Linguistic analysis of mathematical discourse and argumentation relations in the context of mathematical text;
- Probing mathematical understanding of state-of-the-art models;
- Adaptation of NLP tasks for mathematical discourse;
- NLP applied to mathematics education;
- Premise selection over mathematical text;
- Understanding and typing of variables in mathematical text;
- Retrieval of equations/formulas/expressions based on textual queries;
- Retrieval of textual context based on equational queries.

MathNLP is partially funded by the Swiss National Science Foundation (SNSF) project NeuMath (200021\_204617)



## **Organizing Committee**

- Marco Valentino, Idiap Research Institute
- Deborah Ferreira, The MathWorks
- Mokanarangan Thayaparan, The MathWorks
- Andre Freitas, Idiap Research Institute & The University of Manchester



## Table of Contents

<i>An Approach to Co-reference Resolution and Formula Grounding for Mathematical Identifiers Using Large Language Models</i> Aamin Dev, Takuto Asakura and Rune Sætre .....	1
<i>Fluid Dynamics-Inspired Emotional Analysis in Shakespearean Tragedies: A Novel Computational Linguistics Methodology</i> Davide Picca .....	11
<i>Math Problem Solving: Enhancing Large Language Models with Semantically Rich Symbolic Variables</i> Ali Emre Narin .....	19
<i>Data Driven Approach for Mathematical Problem Solving</i> Byungju Kim, Wonseok Lee, Jaehong Kim and Jungbin Im .....	25
<i>Exploring Internal Numeracy in Language Models: A Case Study on ALBERT</i> Ulme Wennberg and Gustav Eje Henter .....	35





# Tutorial Program

*An Approach to Co-reference Resolution and Formula Grounding for Mathematical Identifiers Using Large Language Models*

Aamin Dev, Takuto Asakura and Rune Sætre

*Fluid Dynamics-Inspired Emotional Analysis in Shakespearean Tragedies: A Novel Computational Linguistics Methodology*

Daide Picca

*Math Problem Solving: Enhancing Large Language Models with Semantically Rich Symbolic Variables*

Ali Emre Narin

*Data Driven Approach for Mathematical Problem Solving*

Byungju Kim, Wonseok Lee, Jaehong Kim and Jungbin Im

*Exploring Internal Numeracy in Language Models: A Case Study on ALBERT*

Ulme Wennberg and Gustav Eje Henter

