

AISD 2025

**Annual Conference of the Nations of the Americas Chapter of
the Association for Computational Linguistics**

**Proceedings of the 1st Workshop on AI and Scientific
Discovery (AISD): Directions and Opportunities**

May 3, 2025

The AISD organizers gratefully acknowledge the support from the following sponsors.

Gold



©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 979-8-89176-224-4

Introduction

Welcome to AISD, the First Workshop on AI & Scientific Discovery, co-located with NAACL 2025 in Albuquerque, New Mexico.

Just as coding assistants have dramatically increased productivity for coding tasks over the last two years, researchers in the NLP community have begun to explore methods and opportunities ahead for creating scientific assistants that can help with the process of scientific discovery and increase the pace at which novel discoveries are made. Over the last year, language models have been used to create problem-general scientific discovery assistants that are not restricted to narrow problem domains or formulations. Such applications hold opportunities for assisting researchers in broad domains, or scientific reasoning more generally. Beyond assisting, a growing body of work has begun to focus on the prospect of creating largely autonomous scientific discovery agents that can make novel discoveries with minimal human intervention. These recent developments highlight the possibility of rapidly accelerating the pace of scientific discovery in the near term. Given the influx of researchers into this expanding field, this workshop proposes to serve as a vehicle for bringing together a diverse set of perspectives from this quickly expanding subfield, helping to disseminate the latest results, standardize evaluation, foster collaboration between groups, and allow discussing aspirational goals for 2025 and beyond. This workshop welcomes and covers a wide range of topics, including (but not limited to): Literature-based Discovery, Agent-centered Approaches, Automated Experiment Execution, Automated Replication, Data-driven Discovery, Discovery in Virtual Environments, Discovery with Humans in the Loop, and Assistants for Scientific Writing.

A total of 7 papers appear in the proceedings. 24 papers were presented at the workshop itself, with the rest being submitted under two archival options: cross-submissions (Findings papers or those already presented at other venues, such as ICLR, EMNLP, NeurIPS, or the NAACL main conference), and regular non-archival submissions (unpublished work). The latter went through a normal peer review process. These papers can be found on the AISD website: <https://ai-and-scientific-discovery.github.io/>

Six papers were featured as oral presentations. These papers represented a selection of strong work that the organizers felt would be of broad interest to workshop participants. In addition, we featured four invited talks: Heng Ji, Jure Leskovec, Peter Clark, and Marinka Zitnik. We are thankful to all reviewers for their help in the selection of the program, for their readiness to engage in thoughtful discussions about individual papers, and for providing valuable feedback to the authors. We would also like to thank the NAACL workshop organizers for all the valuable help and support with the organizational aspects of the conference. Finally, we would like to thank all our authors and presenters for making this such an exciting event!

Peter Jansen, Bhavana Dalvi Mishra, Harsh Trivedi, Bodhisattwa Prasad Majumder, Tom Hope, Tushar Khot, Doug Downey, Eric Horvitz
AISD organizers

Organizing Committee

Organizers

Bhavana Dalvi Mishra, Allen Institute for Artificial Intelligence

Doug Downey, Allen Institute for Artificial Intelligence

Tom Hope, Allen Institute for Artificial Intelligence / Hebrew University of Jerusalem

Eric Horvitz, Microsoft

Peter Jansen, University of Arizona / Allen Institute for Artificial Intelligence

Tushar Khot, Google DeepMind

Bodhisattwa Prasad Majumder, Allen Institute for Artificial Intelligence

Harsh Trivedi, Allen Institute for Artificial Intelligence

Program Committee

Program Chairs

Bhavana Dalvi Mishra, Allen Institute for Artificial Intelligence
Doug Downey, Allen Institute for Artificial Intelligence
Tom Hope, Allen Institute for Artificial Intelligence and Hebrew University of Jerusalem
Eric Horvitz, Microsoft
Peter Jansen, University of Arizona and Allen Institute for Artificial Intelligence
Tushar Khot, Google DeepMind
Bodhisattwa Prasad Majumder, Allen Institute for Artificial Intelligence
Harsh Trivedi, Allen Institute for Artificial Intelligence

Reviewers

Jinheon Baek, Chandrayee Basu

Srinivasarao Daruna

Shivanshu Gupta

Naoya Inoue

Uri Katz

Arun Balajiee Lekshmi Narayanan, Daniel J. Liebling

Enrique Noriega-Atala

Zayne Rea Sprague

Rosni Vasu

Table of Contents

| | |
|--|----|
| <i>Variable Extraction for Model Recovery in Scientific Literature</i> | |
| Chunwei Liu, Enrique Noriega-Atala, Adarsh Pyarelal, Clayton T Morrison and Mike Cafarella | 1 |
| <i>How Well Do Large Language Models Extract Keywords? A Systematic Evaluation on Scientific Corpora</i> | |
| Nacef Ben Mansour, Hamed Rahimi and Motasem Alrahabi | 13 |
| <i>A Human-LLM Note-Taking System with Case-Based Reasoning as Framework for Scientific Discovery</i> | |
| Douglas B Craig | 22 |
| <i>Towards AI-assisted Academic Writing</i> | |
| Daniel J. Liebling, Malcolm Kane, Madeleine Grunde-McLaughlin, Ian Lang, Subhashini Venugopalan and Michael Brenner | 31 |
| <i>Evaluating and Enhancing Large Language Models for Novelty Assessment in Scholarly Publications</i> | |
| Ethan Lin, Zhiyuan Peng and Yi Fang | 46 |
| <i>LLM-Assisted Translation of Legacy FORTRAN Codes to C++: A Cross-Platform Study</i> | |
| Nishath Rajiv Ranasinghe, Shawn M. Jones, Michal Kucer, Ayan Biswas, Daniel O'Malley, Alexander Most, Selma Liliane Wanna and Ajay Sreekumar | 58 |
| <i>FlavorDiffusion: Modeling Food-Chemical Interactions with Diffusion</i> | |
| Junpyo Seo | 70 |