

The 2nd International Workshop on Artificial Intelligence and the Science of Science

Jian Wu¹, Sarah Rajtmajer² and Yi He³

¹Computer Science, Old Dominion University, Norfolk, VA, United States

²College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, 16802

³Department of Data Science, College of William & Mary, Williamsburg, VA 23185

Abstract

This preface document summarizes the theme, organization, review, schedules, and accepted papers of the 2nd International Workshop on Artificial Intelligence and the Science of Science (AI4SciSci'25). The workshop welcomes papers that report original works and published works. There were 10 papers reporting original works submitted for peer-review to this workshop. Out of these, 7 papers were accepted for this volume, 3 as regular papers, 2 as short papers, and 2 as extended abstracts. The workshop also accepted 2 published papers for presentation, but they are not included in the proceedings.

1. Preface

The 2nd International Workshop on Artificial Intelligence and the Science of Science (AI4SciSci 2025) was co-located with the 25th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), a premier conference on digital libraries and information retrieval. The AI4SciSci workshop brought together researchers and practitioners from artificial intelligence, data mining, and science of science to explore innovative solutions and ideas for emerging and challenging problems in this interdisciplinary field. The workshop welcomed high-quality submissions centered on several main themes in scholarly graph analytics, scholarly data mining, scientific document processing and understanding, and computational social sciences. The call for papers (CFP) accepted two types of submissions: published work, which included papers accepted or published in peer-reviewed journals and conferences, and original work, which had not yet been officially published. Original work was welcomed in three lengths: full papers (up to 8 pages), short papers (up to 4 pages), and extended abstracts (up to 2 pages). References and appendices were not counted towards page limits.

In response to the CFP the workshop received 12 submissions, including 2 published works and 10 original works. Each original paper was reviewed by at least three program committee members. Given the quality and topical alignment of submissions, we accepted a total of 9 papers: 2 published works and 7 original works. Original works include 3 long papers, 2 short papers, and 2 extended abstracts. The authors of accepted papers come from 5 countries: the USA, Belgium, Great Britain, China, and Korea. The full program, including the schedules, can be found on the AI4SciSci website: <https://ai4sciscsci.github.io/2025/>. All accepted papers are listed in Section 2. Only the 7 original works are included in this proceeding.

The 1st AI4SciSci workshop was co-located with the 2023 IEEE International Conference on Data Mining (ICDM) in Shanghai, China. The workshop was divided into two sessions. The morning session featured a keynote speech by Dr. Daniel Acuña, Associate Professor in the Department of Computer Science at the University of Colorado at Boulder, 2 long papers, 1 published paper, and 1 short paper. The afternoon session features 1 long paper, 1 published paper, 1 short paper, and 2 extended abstracts.

AI4SciSci'25: Workshop on the Artificial Intelligence and the Science of Science, December 15, 2025, DeKalb, Illinois, USA

✉ fanchyna@gmail.com (J. Wu); smr48@psu.edu (S. Rajtmajer); yhe15@wm.edu (Y. He)

🌐 <https://www.cs.odu.edu/~jwu/> (J. Wu); <https://www.rajtmajerlab.net/> (S. Rajtmajer); <https://yhe15.people.wm.edu/> (Y. He)

🆔 0000-0003-0173-4463 (J. Wu); 0000-0002-1464-0848 (S. Rajtmajer); 0000-0002-5357-6623 (Y. He)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Accepted Papers

Accepted papers are briefly introduced below. Keywords were provided by the authors.

2.1. Published Works

- **Transforming Role Classification in Scientific Teams using LLMs and Advanced Predictive Analytics** (*Wonduk Seo and Yi Bu*)
Keywords: Author Role Classification, Large Language Models (LLMs), Predictive Analytics, SHAP Interpretability
- **Scientific Productivity and Practice in the Era of Large Language Models** (*Keigo Kusumegi, Xinyu Yang, Paul Ginsparg, Mathijs de Vaan, Toby Stuart and Yian Yin*)
Keywords: Large Language Models, Productivity, Citation Analysis

2.2. Original Works – Full Papers

- **Humans vs. LLMs on Open Domain Scientific Claim Verification: A Baseline Study** (*Benjamin Curtis, Stefania Dzhaman, Matthew Maisonave and Jian Wu*)
Keywords: Scientific Claim Verification, Large Language Model, Large Reasoning Model, Prompt Engineering
- **A Gradio-Based Toolkit for Remote Sensing Data Fusion Literature** (*Caleb Cheruiyot, Leidig Jonathan P and Jiaxin Du*)
Keywords: Remote Sensing, Data Fusion, Knowledge Graph, Uncertainty Tagging, Gradio, BERT
- **Tracing Research Inequality in NLP: How Resource Disparities Shape Topic Trends and Methodological Diffusion via Citations** (*Lizhen Liang and Bei Yu*)
Keywords: Research Inequality, Citation Intent, Scholarly Text Processing,

2.3. Original Works – Short Papers

- **A Case For Clarity: Knowledge Engineering And Its Evolving Role In Fundamental And Applied Medical Sciences** (*Henri Van Overmeire and Patrick Wouters*)
Keywords: Physiology, Medicine, Knowledge Engineering, Epistemology, Central Venous Pressure, IEER
- **Quantifying Contextual Hallucinations in NLP Research Papers Before and After the LLM Era** (*Adiba Ibnat Hossain, Miftahul Jannat Mokarrama and Hamed Alhoori*)
Keywords: Hallucination, LLM, Scientific Writing, Context Inconsistency, Faithfulness, NLP, Academic Integrity

2.4. Original Works - Extended Abstracts

- **Geography of Medical Knowledge: Scientific Focus, Disease Burden, and Research Response** (*Hongyu Zhou, Prashant Garg and Thiemo Fezter*)
Keywords: Research Responsiveness, Geographic Disparities, Disease Burden, LLM-enabled Knowledge Graph
- **Applying LLM to Library Metadata: Mapping Geography and Language in the Library of Congress Collection** (*Kai Li, Hongyu Zhou, Raf Guns, Tim C.E. Engels and Brian Dobreski*)
Keywords: Library Metadata, Large Language Models, Knowledge Geography, Cataloging, Cultural Representation

3. Keynote

We invited Dr. Daniel Acuña, Associate Professor in the Department of Computer Science at the University of Colorado at Boulder to deliver the Keynote. Daniel leads the Science of Science and

Computational Discovery Lab. He works in science of science, a subfield of computational social science, and AI for science. He publishes research and builds web-based software tools to accelerate knowledge discovery. His current work aims to understand historical relationships, mechanisms, and optimization opportunities of knowledge production. Title and abstract of his talk follow.

Estimating the predictability of questionable open-access journals

Abstract: Questionable journals threaten global research integrity, yet manual vetting can be slow and inflexible. Here, we explore the potential of artificial intelligence (AI) to systematically identify such venues by analyzing website design, content, and publication metadata. Evaluated against extensive human-annotated datasets, our method achieves practical accuracy and uncovers previously overlooked indicators of journal legitimacy. By adjusting the decision threshold, our method can prioritize either comprehensive screening or precise, low-noise identification. At a balanced threshold, we flag over 1000 suspect journals, which collectively publish hundreds of thousands of articles, receive millions of citations, acknowledge funding from major agencies, and attract authors from developing countries. Error analysis reveals challenges involving discontinued titles, book series misclassified as journals, and small society outlets with limited online presence, which are issues addressable with improved data quality. Our findings demonstrate AI's potential for scalable integrity checks, while also highlighting the need to pair automated triage with expert review.

4. Acknowledgments

We thank the JCDL'25 Chairs for supporting the workshop. We thank Northern Illinois University and Old Dominion University for providing technical support. We also thank EasyChair for providing our review platform. Special thanks to Rochana R. Obadage, PhD candidate at Old Dominion University, for building and maintaining the website.

We also want to express our gratitude to the program committee members who dedicated time to reviewing papers. The following members are sorted in alphabetical order by last name.

- Sai Koneru, PhD Candidate, Pennsylvania State University, USA
- Dr. Meijun Liu, Associate Professor, Fudan University, China
- Dr. Jin Mao, Associate Professor, Wuhan University, China
- Dr. Staša Milojević, Professor, Indiana University, Bloomington, USA
- Rochana R. Obadage, PhD Candidate, Old Dominion University, USA
- Dr. Yifang Wang, Assistant Professor, Florida State University, USA
- Dr. Lingfei Wu, Associate Professor, University of Pittsburgh, USA
- Dr. Yian Yin, Assistant Professor, Cornell University, USA
- Dr. Chengzhi Zhang, Professor, Nanjing University of Science and Technology, China
- Dr. Yi Zhao, Lecturer, Anhui University, China

Declaration on Generative AI

During the preparation of this preface, the authors used Grammarly and OpenAI GPT for grammar and spell checking.