Overview of the Third Workshop for Artificial Intelligence for Scientific Publications

Kelly Lockhart‡, Alberto Accomazzi‡, Felix Grezes‡, Tirthankar Ghosal§
‡ Astrophysics Data System, Harvard-Smithsonian Center for Astrophysics, US
§Oak Ridge National Laboratory, US
kelly.lockhart@cfa.harvard.edu, accomazzi@cfa.harvard.edu
felix.grezes@cfa.harvard.edu, ghosalt@ornl.gov

Abstract

The Workshop for Artificial Intelligence for Scientific Publications (WASP), formerly Workshop on Information Extraction from Scientific Publications (WIESP), started in 2022 to provide a platform for researchers to discuss research on information extraction, mining, generation, and knowledge discovery from scientific publications using Natural Language Processing and Machine Learning techniques. The third WASP workshop was held at the 14th International Joint Conference on Natural Language Processing & 4th Asia-Pacific Chapter of the Association for Computational Linguistics in Mumbai, India on December 23rd, 2025, as a hybrid event. The WASP workshop saw great interest, with 29 submissions, of which 16 were accepted. The program consisted of the contributed research talks, 2 keynote talks, a panel discussion, and one shared task, Telescope Reference and Astronomy Categorization Shared task (TRACS).

1 Workshop description

The rise in scientific paper publications has greatly contributed to scientific advancement but has also complicated the ability of researchers to stay up-to-date in their fields. To navigate this vast amount of data and facilitate discovery, incorporating the metadata, full text, and citations into search engines is crucial. A popular and open example is the Astrophysics Data System (ADS; Kurtz et al., 2000), which offers many ways to discover research articles of interest within a curated collection of over 26 million records. However, navigating through this vast amount of data presents considerable challenges. To overcome them, extracting structured and semantically meaningful information from scientific publications becomes imperative.

The Workshop for Artificial Intelligence for Scientific Publications (WASP) was started to provide a platform for researchers to discuss research on information extraction, mining, generation, and knowledge discovery from scientific publications using Natural Language Processing and Machine Learning techniques.

The first WASP workshop was held under the name Workshop on Information Extraction from Scientific Publications (WIESP; Ghosal et al., 2022) in conjunction AACL-IJCNLP 2022. The second edition of WIESP was held along with IJCNLP-AACL 2023 (Ghosal et al., 2023). Much technological change has occurred since the first Workshop, especially around Generative Artificial Intelligence research. The Workshop's scope has expanded, along with the technology, and this year the inclusion of AI was cemented along with the new workshop name.

2 Program

The WASP 2025 workshop consisted of two keynote talks, contributed talks, a shared task, and a panel discussion. The main workshop received 29 submissions for contributed talks, of which 16 were accepted (55% acceptance rate). Since the workshop will be hybrid, there will be both in-person and virtual presentations at the conference venue and online. The papers accepted to the workshop cover a diverse array of research topics primarily centered on automating scholarly workflows, enhancing information extraction from scientific literature, ensuring the reliability of large language models (LLMs) in research, and advancing data management for open science initiatives.

Compared to the previous workshops in this series, the collection of research activities described in these works demonstrates a movement toward AI-assisted critical curation, where LLMs are employed not just to process and generate information, but are architecturally constrained and verified using external knowledge and validation signals derived from the scholarly ecosystem itself. Figure

Proceedings of the Third Workshop for Artificial Intelligence for Scientific Publications, pages 1–4

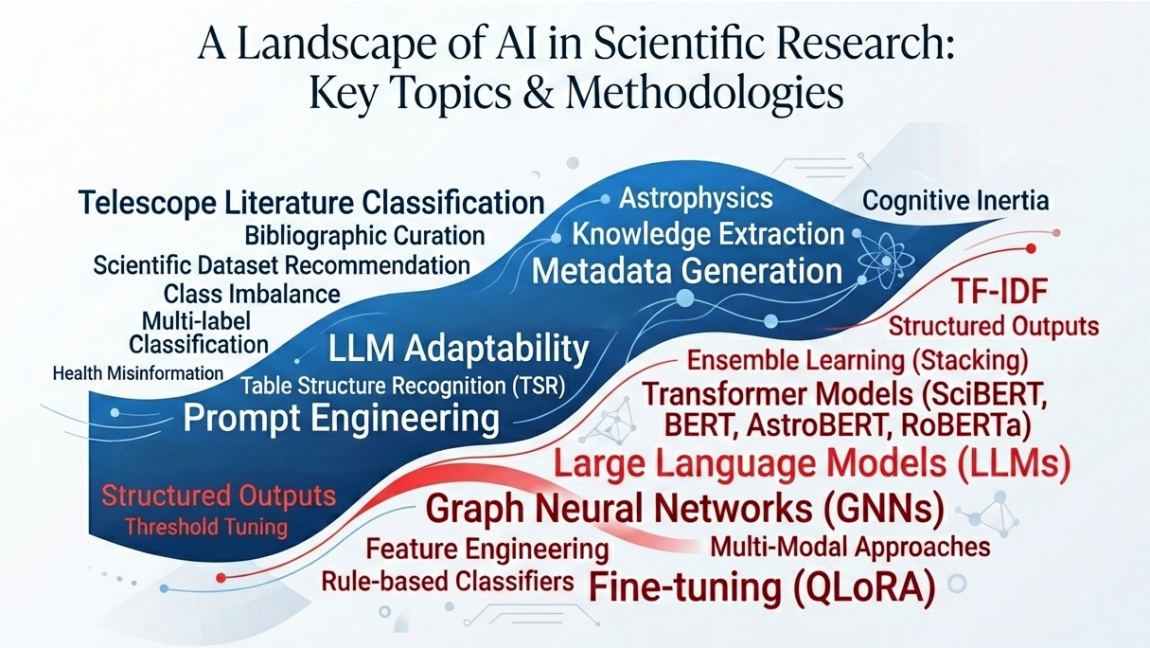

Figure 1: A representation of the scientific topics (blue) and methodologies (red) in the WASP papers.

1 provides a graphical representation of key topics and methodologies.

WASP 2025 also includes a panel discussion, *LLMs for "Trustworthy and Grounded" Scientific Discovery*. The panelists include our two keynote speakers, Karin Verspoor and Kartheik Iyer, along with Prasanna Balaprakash (Director of AI Programs, ORNL), and ChatGPT.

The full program, with links to papers, is available at `https://ui.adsabs.harvard.edu/WIESP/2025/schedule`.

3 Keynotes

This year we had two keynote lectures from researchers working in AI for scientific publications:

- Karin Verspoor, Dean, School of Computing Technologies, Royal Melbourne Institute of Technology, Australia
- Kartheik Iyer, NASA Hubble Fellow, Columbia University, USA

Speaker Karin Verspoor

Title "Impacts of AI on the Scientific Ecosystem"

Abstract Artificial Intelligence, in both predictive and generative forms, is increasingly being adopted to support — and in some cases, entirely perform — scientific research. In this talk, I will discuss both the significant opportunities that AI brings to science and the questions that AI raises for science. The talk will be grounded in some of my own work in use cases including bio-curation and literature-based discovery, as well as ongoing work exploring the limitations of LLMs, that may have particular impacts in the scientific arena.

Speaker Kartheik Iyer

Title "Wandering through the Cosmic Library: Harnessing the embedding spaces of large language models for astronomical research and discovery"

Abstract Astronomical literature is expanding at an unprecedented rate, with thousands of papers added every month to preprint servers like arXiv.org and indexed by the NASA Astrophysics Data System (ADS). For academics and students, staying current with relevant work while keeping track of shifting trends therefore represents a critical challenge. This talk presents lessons learned from working with the UniverseTBD collaboration to develop Pathfinder, a complement to systems like ADS that uses large language models combined with retrieval-augmented generation (RAG) to enable semantic search and question-answering across the astronomy literature. I will discuss some of the unique challenges of applying NLP and LLMs to scientific publications in astronomy, including: (1) handling domain-specific terminology and mathematical notation, (2) grounding LLM responses in archival data to minimize hallucinations, and (3) leveraging embeddings to create inter-

pretable semantic spaces for literature exploration. Drawing from Pathfinder's deployment (pfdr.app) and user feedback from the astronomy community, I will highlight how interpretable intermediate representations such as semantic embeddings and citation graphs can lend interpretability and rigor to otherwise black-box models, and help their adoption in research pipelines. Beyond astronomy, the development of these methods have broader implications for AI-assisted scientific discovery across disciplines. I will conclude by discussing open challenges in adapting large models in scientific contexts, the importance of retrieval mechanisms that preserve provenance, and the potential for LLM-powered tools to not just assist with literature review, but to help generate testable hypotheses and identify research gaps. As scientific publishing continues to accelerate across all fields, developing trustworthy and grounded systems for navigating the literature becomes increasingly essential.

4 Telescope Reference and Astronomy Categorization Shared task (TRACS)

WASP 2025 hosted a shared task, Telescope Reference and Astronomy Categorization Shared task (TRACS; Grezes et al., 2025). The organizing committee was: Alberto Accomazzi, Tirthankar Ghosal, Kelly Lockhart, and Felix Grezes. The detailed overview paper is referred to and included in the proceedings. TRACS is available publicly on HuggingFace [1]. The scoring evaluation was run on the Kaggle platform[2].

TRACS Description: Astronomers typically gauge the scientific influence of observational facilities by examining publications that use the facilities' data. This depends on bibliographies that explicitly annotate and link data products to the relevant literature, enabling bibliometric analysis of data impact. Compiling such bibliographies is time-intensive and requires experts to comb the literature for names, acronyms, and identifiers, and then assess whether and how observations were used. Beyond impact assessment, these data-literature links are vital for researchers, as they form an important route to discovering and accessing data. By capitalizing on the expertise of librarians and archivists, telescope bibliographies can therefore directly support the scientific research workflow. In this context, we present the Telescope Reference and Astronomy Categorization Shared task (TRACS) and dataset, comprising more than 89,000 publicly available English texts drawn from space telescope bibliographies. These texts are labeled with a new, streamlined taxonomy developed in collaboration with experienced bibliographers. TRACS is intended as training material for contemporary Machine Learning and Artificial Intelligence methods that can assist data curators in building bibliographies. As an initial benchmark, we assess how existing Large Language Models perform on automatic bibliography curation. Both baseline and participant results underscore the difficulty of the problem and highlight the need for specialized tools. The TRACS shared task attracted 9 participating teams on Kaggle, of which 6 submitted system papers to WASP 2025.

[1] huggingface.co/datasets/adsabs/TRACS

[2] kaggle.com/competitions/tracs-wasp-2025

5 Conclusion

The rapid growth of scientific publishing presents both opportunity and difficulty for researchers who rely on accurate, interpretable, and well-structured access to scholarly literature. While significant progress has been made in information extraction, document understanding, and the responsible use of artificial intelligence, many challenges remain, particularly in reliably grounding automated systems in the scholarly ecosystem and in ensuring that AI-generated outputs can be trusted in research contexts. The contributions of WASP 2025 illustrate both the promise of current approaches and the continued need for rigorously validated, domain-aware methods.

By assembling researchers from NLP, information retrieval, and neighboring disciplines, WASP aims to advance these efforts and highlight emerging directions for AI-assisted scholarship. We hope that this workshop and its shared task will spur new collaborations, sharpen our understanding of open problems, and inspire the creation of robust tools that meaningfully support scientific discovery and the broader research community.

Program Committee

1. Akhil Pandey Akella, AllSci Corp
2. Akiko Aizawa, National Institute of Informatics
3. Alberto Accomazzi, Center for Astrophysics | Harvard & Smithsonian
4. Antoine Gauquier, Ecole Normale Supérieure

de Paris
5. Anurag Acharya, Pacific Northwest National Laboratory
6. Arne Binder, German Research Center for AI
7. Atilla Kaan Alkan, Center for Astrophysics | Harvard & Smithsonian
8. Bhargab Choudhury, Gauhati University
9. Biswadip Mandal, Amazon
10. Buse Sibel Korkmaz, Imperial College London
11. Daniel Acuna, University of Colorado at Boulder
12. Ekaterina Artemova, Toloka AI
13. Felix Grezes, Center for Astrophysics | Harvard & Smithsonian
14. Hamed Alhoori, Northern Illinois University
15. Iana Atanassova, Université de Franche Comté
16. Ibrahim Al Azher, Northern Illinois University
17. Jeyadev Needhidevan, New York University
18. Kelly Lockhart, Center for Astrophysics | Harvard & Smithsonian
19. Markus Zhang, Stanford University
20. Mengyang Liu, Independent Researcher
21. Michael Cochez, VU Amsterdam
22. Mina Basirat, University of Central Florida
23. Neil R. Smalheiser, University of Illinois at Chicago
24. Pawin Taechoyotin, University of Colorado at Boulder
25. Pierre Zweigenbaum, LISN, CNRS, Université Paris-Saclay
26. Rajesh Piryani, Akkodis
27. Sergey Feldman, Allen Institute for Artificial Intelligence
28. Sergi Blanco-Cuaresma, UniDistance Suisse
29. Shufan Ming, University of Illinois Urbana-Champaign
30. Shuntaro Yada, Tsukuba University, Tokyo Institute of Technology
31. Soham Chitnis, New York University
32. Sridevi Wagle, Pacific Northwest National Laboratory
33. Sujit Pal, Elsevier
34. Ted Pedersen, Nara Institute of Science and Technology
35. Thomas S Allen, Center for Astrophysics | Harvard & Smithsonian
36. Tim Schopf, Technische Universität Dresden
37. Tosho Hirasawa, Omron Sinic X
38. Wojtek Sylwestrzak, University of Warsaw
39. Yagmur Ozturk, Université Grenoble Alpes
40. Yuan Chang, University of the Chinese Academy of Sciences
41. Ziyue Li, University of the Chinese Academy of Sciences

References

Tirthankar Ghosal, Sergi Blanco-Cuaresma, Alberto Accomazzi, Robert M. Patton, Felix Grezes, and Thomas Allen, editors. 2022. *Proceedings of the first Workshop on Information Extraction from Scientific Publications*. Association for Computational Linguistics, Online.

Tirthankar Ghosal, Felix Grezes, Thomas Allen, Kelly Lockhart, Alberto Accomazzi, and Sergi Blanco-Cuaresma, editors. 2023. *Proceedings of the Second Workshop on Information Extraction from Scientific Publications*. Association for Computational Linguistics, Bali, Indonesia.

Felix Grezes, Jennifer Lynn Bartlett, Kelly Lockhart, Alberto Accomazzi, Ethan Seefried, and Tirthankar Ghosal. 2025. Overview of TRACS: the telescope reference and astronomy categorization dataset & shared task. In *Proceedings of the Third Workshop for Artificial Intelligence for Scientific Publications*, Online. Association for Computational Linguistics.

Michael J. Kurtz, Guenther Eichhorn, Alberto Accomazzi, Carolyn S. Grant, Stephen S. Murray, and Joyce M. Watson. 2000. The NASA Astrophysics Data System: Overview. , 143:41–59.