

Knowledge Discovery IG Session

Chair: Raffaele D'Abrusco

Vice-Chair: Yihan Tao

22 May 2024

Sydney, Australia

Talks

<i>Alberto Accomazzi</i>	BiblioPile : Building a Dataset to Support AI-enabled Bibliography Curation efforts	15'+5'	A well-established way to assess the scientific impact of an observational facility in astronomy is the quantitative analysis of the studies published in the literature which have made use of the data taken by the facility. A requirement of such analysis is the creation of bibliographies which annotate and link data products with the literature, thus providing a way to use bibliometrics as an impact measure for the underlying data. An automated assistant able to emulate some of the associated activities would provide a valuable contribution to the human effort involved. LLMs have shown flexibility in interpreting and classifying scientific articles which are the basis for this curation activity. They have also been successfully used for information extraction tasks, which would help identify the specific datasets mentioned in the papers. In this talk I will describe our effort to create the BiblioPile , a contributed dataset consisting of open access fulltext papers and annotated bibliography from institutions that maintain them in order to help train AI/ML bibliographic annotation pipelines.
<i>Yan Shao</i>	Generative Named Entity Normalization for Astronomical Facilities	15'+5'	Named entity normalization for astronomical facilities is crucial in the related academic research. Unlike the majority of the previous work, we model named entity normalization as a sequence generation problem via utilizing large language models, without assuming a comprehensive set of predefined normalized forms for any entities. Four entity normalization scenarios that are likely to occur in real-world application are discussed specifically, depending on whether the explicit normalization rules as well as the corresponding annotated instances are available. Moreover, we propose respective generative normalization methods and evaluate on datasets compiled from the standard telescope name lists maintained by the American Astronomical Society (AAS) and the Astrophysics Data System (ADS). The empirical findings demonstrate that the analytical, inductive, and generative capabilities of LLMs empower generative entity normalization to achieve commendable performances, even under very stringent conditions. The generative normalization effectively remedies the shortcomings of the retrieval-based methods.
<i>Kai Polsterer</i>	Spherinator & HIPster & Jasmine	15'+5'	Simulations are the best, and often the only, approximation to experimental laboratories in Astrophysics. However, the complexity and richness of their outputs severely limits the interpretability of their predictions. We describe a new conceptual approach to obtaining useful scientific insights from a broad range of astrophysical simulations. These methods can be applied to state-of-the-art simulations and will be essential to automate the data exploration and analysis of the next-generation exascale simulations and the extreme data challenges they will present. Our concept is based on applying the latest advances in unsupervised deep learning algorithms to efficiently represent the multidimensional datasets produced by Astrophysics simulations and to learn compact but accurate representations of the data in a low-dimensional manifold that naturally describes the data in an optimal feature space. The data can seemingly be projected onto this latent space for interactive inspection, visual interpretation, and quantitative analysis, including the option of deriving symbolic expressions to build interpretable models. We present a working prototype of the pipeline using an autoencoder trained on galaxy images from SDSS (or equivalently simulated galaxies) as well as the Illustris simulations, to produce a natural 'Hubble tuning fork' similarity space that can be visualized interactively on the surface of a sphere by exploiting the power of HiPS tilings in AladinLite . Besides Spherinator and HIPster to do this spherical projections, we are working on Jasmine a tool to explore the rich data from simulations in detail.

Panel Discussion

- **Panelists:** Alberto Accomazzi, Rafeal Martinez-Galarza, Kai Polsterer, Yihan Tao

- **Topics:**

1. How can state-of-the-art AI technologies, such as LLMs, foundation models and agents enhance the VO?
2. What are the potential applications of these AI technologies within the VO framework?
3. What are the best practices and strategies for integrating AI agents and models with VO tools and science platforms that can help user efficiently access to and analyse astronomical data? What are the challenges?

Astronomical Techniques and Instruments

Call for papers



Special Issue: AI for astronomy

About this issue:

Given the ongoing and evolving debate surrounding AI, it's necessary to shed light on how AI can enhance the astronomy field effectively and ethically. This special issue aims to promote the integration of massive astronomical data with advanced AI technologies, to facilitate cross-field cooperation and exchanges between experts in the fields of astronomy and information processing, and to expand the application of artificial intelligence algorithms in the field of astronomy. We invite experts and scholars in related research fields to submit rigorous original research papers and reviews that address AI in astronomical fields. The special issue aspires to contribute to a more nuanced and informed discourse on leveraging AI for more sustainable, knowledgeable, and advanced research through this comprehensive examination.

List of topics:

- AI for Time Domain Astronomy
- AI for Multi-messenger Astronomy
- AI in Astronomy Big Data
- AI with Intelligent Computing Platform
- AI in Intelligent Observation
- AI and Galaxy Cosmology
- AI and Stellar Physics
- AI and Galaxy
- AI and Solar Physics
- AI and Planetary Science
- AI for Fundamental Astronomy
- AI in Extraterrestrial Civilization Search

Guest editors



Prof. Chenzhou Cui
National Astronomical
Observatories, CAS,
China



Prof. Feng Wang
Guangzhou University



Prof. Hailong Zhang
Xinjiang Astronomical
Observatory, CAS,
China

Manuscript submission

<https://mc03.manuscriptcentral.com/ati>

Submission deadline

Submission closing: Aug 31, 2024

Contact information

ati@ynao.ac.cn