

## Editorial

# Editorial for the Special Issue “New Discoveries in Astronomical Data”

Yanxia Zhang <sup>1,\*</sup>  and A-Li Luo <sup>1,2</sup> 

<sup>1</sup> CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: zyx@bao.ac.cn

Over the past decade, astronomy has shifted from a data-starved to a data-drenched science [1,2]. The Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST), ESA’s Euclid, and the Square Kilometre Array (SKA) pathfinders are already delivering terabytes of images, spectra, and time-series each night. Traditional analysis pipelines—calibrated for megabyte-scale data sets—cannot keep pace, and a new generation of artificial intelligence (AI) and machine learning (ML) techniques has become indispensable [3–6]. Yet the very power of these methods has revealed a critical gap: the “black-box” nature of many state-of-the-art algorithms threatens the reproducibility, interpretability, and ultimately the credibility of the discoveries they enable.

This Special Issue was conceived to bridge three intertwined gaps:

**Trustworthy AI.** How do we exploit deep neural networks, self-supervised learning, and real-time inference without sacrificing the transparency required by the scientific method?

**Domain-specific bottlenecks.** Solar radio spectroscopy, wide-field meteor monitoring, CO clump cataloguing, and RR Lyrae parameterisation each suffer from small training sets, strong selection effects, or instrument-specific artefacts that generic ML toolkits ignore.

**Environmental integrity.** As facilities push into previously “quiet” frequency bands, radio-frequency interference (RFI) threatens both current and future observations; rigorous site-protection modelling is now mission-critical.

The eight articles collected here address these challenges head-on and chart a coherent path forward.

Lieu’s comprehensive review opens the volume by framing the interpretability crisis [Contribution 1]. She distils the vocabulary—transparency, interpretability, explainability—and showcases astronomy-specific tools (symbolic regression, physics-informed neural networks, and SHAP/LIME visualizations) that convert predictive power into physical insight. Her call for interdisciplinary benchmarks sets the tone for every contribution that follows.

The next contribution, on solar physics, demonstrates the power of self-supervised learning when labeled examples are scarce [Contribution 2]. By adapting BERT-style masking to Vision Transformers, the authors achieve 99.5% accuracy in classifying solar radio bursts—outperforming supervised baselines—while revealing that a 75% masking rate (far above NLP standards) is optimal for redundant spectral images. The study establishes a new transfer-learning paradigm for small-sample astronomical data.

Accurate environmental protection starts with accurate propagation models [Contribution 3]. The RFI analysis software presented here integrates global terrain, meteorology, and satellite ephemerides into an ITU-R-compliant framework that has already guided site-selection for the Five-hundred-meter Aperture Spherical Telescope (FAST) and the



Received: 28 August 2025

Accepted: 1 September 2025

Published: 3 September 2025

**Citation:** Zhang, Y.; Luo, A.-L. Editorial for the Special Issue “New Discoveries in Astronomical Data”. *Universe* **2025**, *11*, 299. <https://doi.org/10.3390/universe11090299>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

planned Qitai Telescope (QTT) [Contribution 4]. Freely available binaries and an interactive map-based GUI invite the community to extend these protections to next-generation arrays.

Single-site meteor detection is notoriously plagued by false positives from aircraft, satellites, and sensor noise. GWAC's new algorithm exploits temporal differencing, probabilistic Hough transforms, and light-curve morphology to reach 90% accuracy on  $4K \times 4K$  frames sampled every 15 s—a three-fold speed-up over prior work [Contribution 5]. The open-source pipeline is already ingesting real-time data from the Geminid stream and will inform the forthcoming Chinese Space Station Telescope transient key project.

Cataloguing molecular clouds has traditionally required human-tuned thresholds and dozens of free parameters [Contribution 6]. MCD-YOLOv5 replaces this tedium with a two-stage deep-learning workflow: an attention-augmented YOLOv5 detector slices Galactic plane images, while Density Peak Clustering stitches velocity channels into 3D clumps. Trained on 10,000 synthetic but realistic MWISP cubes, the network attains 98% recall with only two tunable parameters—an order-of-magnitude reduction compared with FellWalker or ClumpFind.

The RR Lyrae study shows how neural emulators can invert light-curve morphology into fundamental stellar parameters [Contribution 7]. A dense hydrodynamic grid and a four-layer ANN recover mass, luminosity, effective temperature, and metallicity from TESS single-band photometry alone. The derived period–luminosity–metallicity relation is consistent with theoretical predictions and demonstrates a fast, scalable route to precision stellar astrophysics for LSST's multi-band light curves.

Finally, the study on H II regions in NGC 2403 concludes that these regions are primarily star-forming with a clear inside-out evolution pattern [Contribution 8]. The  $N_2O_2$  diagnostic is identified as the most reliable for estimating metallicity due to its insensitivity to the ionization parameter and age. The study highlights limitations in reproducing certain emission lines, particularly [O II], suggesting future work should consider the effects of stellar rotation or binary populations to improve model accuracy. This research provides valuable insights into the properties and evolution of H II regions in NGC 2403, contributing to the broader understanding of star formation and galaxy evolution.

Taken together, these works deliver a clear message: **interpretable, domain-aware AI is no longer optional; it is the prerequisite for turning tomorrow's exabyte surveys into lasting scientific knowledge.**

Looking forward, five priorities emerge:

**Mechanistic interpretability.** We must move from post hoc saliency maps to architectures that embed conservation laws, radiative transfer, and gravitational dynamics directly into their loss functions.

**Causal inference.** Correlations uncovered by neural networks need to be interrogated with counterfactual simulations to disentangle astrophysical causation from observational bias.

**Scalable, uncertainty-aware pipelines.** As data volumes grow, even lightweight explainability tools become computationally prohibitive. Hardware-aware pruning and probabilistic neural networks will be essential.

**Cross-messenger consistency.** Joint electromagnetic, gravitational-wave, and neutrino data sets require federated learning frameworks that respect proprietary formats while preserving interpretability.

**Ethical stewardship of the spectrum.** RFI mitigation strategies must evolve in tandem with satellite mega-constellations; open-source tools like those presented here should become community standards endorsed by the IAU and national regulators.

We invite the readership to treat this Special Issue not as a collection of isolated advances but as a blueprint for the next decade of trustworthy, data-driven discovery.

The code, models, and data sets accompanying each paper are released under permissive licences precisely to encourage rapid iteration and broader adoption. Let us build on these foundations—together—to ensure that the coming avalanche of astronomical data yields insights that are as reliable and profound as the night sky itself.

As AI technology continues to evolve, its applications in astronomy are expected to expand and deepen, leading to even more groundbreaking discoveries [7–14]. The future of AI in astronomy holds several promising directions:

**Advanced Machine Learning Models:** The development of more sophisticated machine learning models will likely uncover new types of phenomena and optimize space missions in ways not yet imagined. These models will be capable of handling the increasing volume and complexity of astronomical data, leading to more accurate predictions and discoveries.

**Data Fusion and Multi-Messenger Astronomy:** The ability to effectively fuse data from multiple sources and instruments will become increasingly important. AI will play a crucial role in integrating and analyzing data from different telescopes, observatories, and missions, enabling a more comprehensive understanding of cosmic events. This will be particularly important for multi-messenger astronomy, where data from gravitational waves, electromagnetic radiation, and other sources are combined to study cosmic phenomena.

**Enhanced Observational Capabilities:** AI will continue to improve the capabilities of ground-based and space-based telescopes. This includes advancements in adaptive optics, which can correct for atmospheric distortions in real-time, and the development of new algorithms for image processing and data analysis. The integration of AI with upcoming observatories such as the SKA and the Rubin Observatory will enable unprecedented levels of detail and accuracy in astronomical observations.

**Citizen Science and Public Engagement:** AI can facilitate citizen science projects by automating tasks that would otherwise require significant human effort. This will allow more people to participate in astronomical research, contributing to the discovery and analysis of celestial objects. Additionally, AI-generated visualizations and educational tools will enhance public engagement and understanding of astronomy.

**Addressing Challenges:** While AI offers numerous benefits, it also presents challenges such as data quality, model interpretability, and computational resource requirements. Future research will focus on developing robust data quality enhancement algorithms, creating interpretable AI models, and optimizing computational resources to ensure the efficient and effective use of AI in astronomy.

In conclusion, AI has already transformed the field of astronomy by enhancing data processing, enabling new discoveries, and improving observational capabilities. As technology continues to advance, AI will play an increasingly vital role in unravelling the mysteries of the universe, making astronomy more accessible and engaging for both researchers and the public.

**Author Contributions:** Conceptualization, Y.Z. and A.-L.L.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z. and A.-L.L.; funding acquisition, Y.Z. and A.-L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (grant Nos. 12273076, 12273075 and 12133001).

**Conflicts of Interest:** The authors declare no conflicts of interest.

### List of Contributions

1. Lieu, M. A Comprehensive Guide to Interpretable AI-Powered Discoveries in Astronomy. *Universe* **2025**, *11*, 187. <https://doi.org/10.3390/universe11060187>.
2. Li, S.; Yuan, G.; Chen, J.; Tan, C.; Zhou, H. Self-Supervised Learning for Solar Radio Spectrum Classification. *Universe* **2022**, *8*, 656. <https://doi.org/10.3390/universe8120656>.
3. Wang, J.; Zhao, Y.; Yang, C.; Shi, Y.; Hao, Y.; Zhang, H.; Sun, J.; Luo, D. The Analysis and Verification of IMT-2000 Base Station Interference Characteristics in the FAST Radio Quiet Zone. *Universe* **2023**, *9*, 248. <https://doi.org/10.3390/universe9060248>.
4. Wang, Y.; Zhang, H.; Wang, J.; Huang, S.; Hu, H.; Yang, C. A Software for RFI Analysis of Radio Environment around Radio Telescope. *Universe* **2023**, *9*, 277. <https://doi.org/10.3390/universe9060277>.
5. Chen, Y.; Li, G.; Liu, C.; Qiu, B.; Shan, Q.; Li, M. A Meteor Detection Algorithm for GWAC System. *Universe* **2023**, *9*, 468. <https://doi.org/10.3390/universe9110468>.
6. Hu, J.-B.; Huang, Y.; Zheng, S.; Chen, Z.-W.; Zeng, X.-Y.; Luo, X.-Y.; Long, C. Molecular-Clump Detection Based on an Improved YOLOv5 Joint Density Peak Clustering. *Universe* **2023**, *9*, 480. <https://doi.org/10.3390/universe9110480>.
7. Kumar, N.; Singh, H.P.; Malkov, O.; Joshi, S.; Tan, K.; Prugniel, P.; Bhardwaj, A. Extraction of Physical Parameters of RRab Variables Using Neural Network Based Interpolator. *Universe* **2025**, *11*, 207. <https://doi.org/10.3390/universe11070207>.
8. Wu, Q.-M.; Mao, Y.-W.; Lin, L.; Zou, H.; Wang, S.-T. Spectroscopic Observations and Emission-Line Diagnoses for H ii Regions in the Late-Type Spiral Galaxy NGC 2403. *Universe* **2025**, *11*, 280. <https://doi.org/10.3390/universe11080280>.

### References

1. Khalil, M.; Said, M.; Osman, H.; Ahmed, B.; Ahmed, D.; Younis, N.; Maher, B.; Osama, M.; Ashmawy, M. Big data in astronomy: From evolution to revolution. *Int. J. Adv. Astron.* **2019**, *7*, 11. [\[CrossRef\]](#)
2. Mickaelian, A.M. Big Data in Astronomy: Surveys, Catalogs, Databases and Archives. *Commun. BAO* **2020**, *67*, 159. [\[CrossRef\]](#)
3. Wang, K.; Guo, P.; Yu, F.; Duan, L.; Wang, Y.; Du, H. Computational Intelligence in Astronomy: A Survey. *Int. J. Comput. Intell. Syst.* **2018**, *11*, 575. [\[CrossRef\]](#)
4. Das, S.; Dey, A.; Roy, N. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *Int. J. Comput. Appl.* **2015**, *115*, 9. [\[CrossRef\]](#)
5. Meher, S.K.; Panda, G. Deep learning in astronomy: A tutorial perspective. *Eur. Phys. J. Spec. Top.* **2021**, *230*, 2285. [\[CrossRef\]](#)
6. Thatikonda, V.K.; Golla, Y.; Mudunuri, H.R. Developments in Artificial Intelligence and Machine Learning: Recent Advances and Future Prospects. *ESP Int. J. Adv. Sci. Technol.* **2024**, *2*, 1–5.
7. Djorgovski, S.G.; Mahabal, A.A.; Graham, M.J.; Polsterer, K.; Krone-Martins, A. Applications of AI in Astronomy. *arXiv* **2022**, arXiv:2212.01493. [\[CrossRef\]](#)
8. Fluke, C.J.; Jacobs, C. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1349. [\[CrossRef\]](#)
9. Impey, C. How AI Is Helping Astronomers. EarthSky. Available online: <https://earthsky.org/space/artificial-intelligence-ai-is-helping-astronomers-make-new-discoveries/> (accessed on 23 May 2023).
10. Pruthi, M.N. Artificial Intelligence in Astronomy. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* **2019**, *7*, 903. [\[CrossRef\]](#)
11. Huang, K.; Hu, T.; Cai, J.; Pan, X.; Hou, Y.; Xu, L.; Wang, H.; Zhang, Y.; Cui, X. Artificial Intelligence in Astronomical Optical Telescopes: Present Status and Future Perspectives. *Universe* **2024**, *10*, 210. [\[CrossRef\]](#)
12. Smith, M.J.; Geach, J.E. Astronomia ex machina: A history, primer and outlook on neural networks in astronomy. *R. Soc. Open Sci.* **2023**, *10*, 221454. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Sharma, P.; Vaidya, B.; Wadadekar, Y.; Bagla, J.; Chatterjee, P.; Hanasoge, S.; Kumar, P.; Mukherjee, D.; Philip, N.S.; Singh, N. Computational Astrophysics, Data Science & AI/ML in Astronomy: A Perspective from Indian Community. *arXiv* **2025**, arXiv:2501.03876. [\[CrossRef\]](#)
14. Znamenskiy, V. Artificial Intelligence and Astronomy: Lab Manual for Generative AI-Based Learning Activities. Available online: <https://zenodo.org/records/15555802> (accessed on 28 August 2025).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.