A Bridge over Deep Space: Astronomy and AI

Anna Jacyszyn [1*]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein Leopoldshafen, Germany
[*]Corresponding author: Anna.Jacyszyn@fiz-Karlsruhe.de.

Interdisciplinary Studies and AI

Leibniz Science Campus *Digital Transformation of Research* (DiTraRe)[1] [1] is an interdisciplinary consortium which studies AI as an additional dimension to digitalisation [2,3]. My task within DiTraRe is to connect computer scientists working on applied AI methods (FIZ Karlsruhe Knowledge Driven AI group[2]) with use cases originating from different research areas (institutes at Karlsruhe Institute of Technology). My experience proves that AI experts are crucial to a successful implementation of various AI methods in multiple cases [4–6]. This is valid especially for the more sophisticated tools like knowledge graphs (KGs). KGs add semantic layer to the data making it machine-understandable and thus, enabling interconnections with other semantically enriched data systems. On the other hand, these clear advantages come at the expense of expertise requirements, as only a trained professionals can install and sustain an ontology and build a KG.

It is worth noting that computer scientists involvement in research activities of other disciplines should not only be limited to supporting with AI methods which are more complicated in implementation. Using generative AI (genAI) and particularly large language models (LLMs) is made very easy and despite being highly profitable [7, 8], it may lead to severe complications [9–11], i.e., ethical concerns [12, 13], hallucinations [14, 15].

Computer scientists as the ones behind development of AI techniques are well aware of both gains and risks posed by application of AI tools. Thus, as the efforts within DiTraRe also prove, computer experts should be often engaged in studies involving AI.

[1]DiTraRe website `https://www.ditrare.de/en`

[2]FIZ Karlsruhe KDAI group website `https://www.fiz-karlsruhe.de/en/bereiche/information-service-engineering`

AI in Astronomy

Publication Culture. Astronomy is a shining example in the fields of research data management (RDM) and open science (OS). Upon the rise of the large data digital surveys, astronomers have already recognised the need for communication protocol, hierarchical metadata, high performance streaming interface, efficient object interchange format, and source cross-identification as reported already in 1999 [16]. Moreover, most of the projects in astronomy, and especially all large projects, are funded by large agencies (like ESA and NASA) and/or international consortia (i.e. Gaia mission [17]). With these government-based funding lines, for many years there has been a requirement on making the data public and results open access (see i.e. ROSES Open Science and Data Management Plan OSDMP[3]). This developed into an entire "culture" of open data publishing, since these practices are very popular across the field. Astronomers have developed common tools and standards, including catalogues (i.e. SIMBAD[4], NED[5]), file types (i.e. open standard Flexible Image Transport System FITS), archives (i.e. ADS[6], astro-ph[7]), software packages (i.e., Astropy[8], courses by Rubin Observatory Collaboration Members[9]), which tend to unify the RDM

[3]NASA OSDMP `https://science.nasa.gov/researchers/sara/faqs/osdmp/`

[4]SIMBAD Astronomical Database `https://simbad.u-strasbg.fr/simbad`

[5]NASA/IPAC Extragalactic Database (NED) `https://ned.ipac.caltech.edu/`

[6]Astrophysics Data System Abstract Service (ADS) `https://ui.adsabs.harvard.edu/`

[7]astro-ph (arXiv for astronomy) service `https://arxiv.org/archive/astro-ph`

[8]Astropy – Astronomy for Python `https://www.astropy.org/`

[9]Intermediate Python for Astronomical Software Development `https://shrra.github.io/`

practices across the field [18, 19].

Virtual Observatory. One of the globally recognised platforms for RDM and OS in astronomy is the Virtual Observatory (VO)[10] operated by the International VO Alliance (IVOA), formed in 2002. The VO aims at gathering databases originating from different instruments and creating a transparent FAIR (Findable, Accessible, Interoperable, Reusable) system [20,21], while the IVOA is a body deciding on interoperability standards. The workflow is modelled after the World Wide Web Consortium [22]. The VO Semantics Working Group concentrates on incorporating vocabularies and ontologies [23, 24], which are methods of symbolic AI, relying on explicit rules and logic.

Semantics in Astronomy. As Cecconi et al. (2025) [25] discuss in their comprehensive article, astronomy, even with its mature RDM practices, still stays behind i.e. environmental or life sciences in terms of the adoption of semantic methods. To advance the semantics environment in astronomy and go beyond the limited categories which are covered by the IVOA, OPAL – Ontology Portal for Astronomy Linked-data was recently started. It is an interdisciplinary collaboration with OntoPortal Alliance and aims at expanding the use of IVOA vocabularies across not only astronomy but also neighbouring fields. OntoPortal Astronomy[11] will combine the existing vocabularies like Unified Astronomy Thesaurus (UAT)[12] [26] and FAIRify them.

Machine Learning in Astronomy. Astronomy has been long time friends with one specific AI branch – machine learning (ML). It enabled multiple milestone discoveries and led to significant advancements in the field (see [27–29] and references therein). Diverse ML methods are deeply incorporated into astronomy research: there also exists an astronomy dedicated Python module (astroML[13]. Many astronomers became experts in ML and utilise various methods fluently in their research work.

GenAI in Astronomy. A subdivision of ML is genAI, based on deep learning algorithms. With the current hype on genAI, multiple ideas on applications have been developed very recently also in the field of astronomy [30]. Astronomy-specific LLMs have been trained, astroBERT [31] and AstroLLaMA [32,33]. A group dedicated to advancing the applications of LLMs in astronomy has been created, AstroMLab[14]. GenAI is now being used for tasks such as: research workflow assistant [34], interacting with publications [35], working with images [36], creating a Slack chatbot [37], teaching via ChatGPT application [38]. Recently, it was proven that the latest models are capable of winning a gold medal at the International Olympiad on Astronomy and Astrophysics [39].

Towards Neurosymbolic AI in Astronomy. If researchers noncritically base their analysis on genAI outputs, their conclusions may be simply incorrect [9, 11]. To avoid that, genAI may be combined with symbolic AI into neurosymbolic AI to improve performance and reasoning abilities [40–43]. Its rapid growth in the last few years has already proven its usefulness in many disciplines, e.g.: collecting and organising scientific knowledge in the Open Research Knowledge Graph (ORKG)[15] [44], adding interpretability to diagnosis prediction in healthcare [45], advancing cognitive systems in natural language processing [46], ensuring transparency and traceability [47]. Given all the advancements in the field of both symbolic and sub-symbolic AI in astronomy, collaborations between astronomers and computer scientists resulting in neurosymbolic methods are soon to be expected to influence the field.

Summary

This contribution is a combination of a review paper and a collection of reflections on my experiences as a coordinator of the Leibniz Science Campus DiTraRe. The goal of this article is manifold: to computer scientists, it presents the current state-of-the-art of AI utilisation use cases in astronomy, while to astronomers, it introduces AI techniques currently being applied in the field. It also aims at encouraging closer and more intense interdisciplinary collaborations.

python-intermediate-development/

[10] VO website https://ivoa.net/

[11] OntoPortal Astronomy https://ontoportal-astro.eu/

[12] UAT https://astrothesaurus.org/

[13] astroML https://www.astroml.org/

[14] AstroMLab on Hugging Face https://huggingface.co/AstroMLab

[15] ORKG website https://orkg.org/

Acknowledgments

The Leibniz Science Campus "Digital Transformation of Research" (DiTraRe) is funded by the Leibniz Association (W74/2022).

References

[1] M. Razum, F. Bach, S. Brünger-Weilandt *et al.*, "Proposal for a Leibniz ScienceCampus – Digital Transformation of Research (DiTraRe)," 2023. doi: 10.5281/zenodo.11109406 ., project proposal.

[2] A. Jacyszyn, H. Sack, D.-S. Group, F. Bach, and M. Razum, "DiTraRe: AI on a Spider's Web. Interweaving Disciplines for Digitalisation," in *4th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K)*, vol. 3780. CEUR, 2024. [Online]. Available: https://ceur-ws.org/Vol-3780/paper5.pdf

[3] A. M. Jacyszyn, F. Bach, H. Sack *et al.*, "Interim Report of the Leibniz Science Campus 'Digital Transformation of Research (DiTraRe)'," 2025. doi: 10.5281/zenodo.14941634 .

[4] A. M. Jacyszyn, S. Jiang, G. A. Gesese *et al.*, "AI4DiTraRe: Towards LLM-Based Information Extraction for Standardising Climate Research Repositories," in *1st Bridge on Artificial Intelligence for Scholarly Communication (AI4SC)*, 2025. [Online]. Available: https://zenodo.org/records/14872359

[5] E. Norouzi, N. Jung, A. M. Jacyszyn *et al.*, "AI4DiTraRe: Building the BFO-Compliant Chemotion Knowledge Graph," in *5th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K)*, vol. 4065. CEUR, 2024. [Online]. Available: https://ceur-ws.org/Vol-4065/paper04.pdf

[6] S. R. Ondraszek, J. Waitelonis, K. Keller *et al.*, "Ontologies in Motion: A BFO-Based Approach to Knowledge Graph Construction for Motor Performance Research Data in Sports Science," in *5th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K)*, vol. 4065. CEUR, 2024. [Online]. Available: https://ceur-ws.org/Vol-4065/paper08.pdf

[7] S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, vol. 381, no. 6654, pp. 187–192, 2023. doi: 10.1126/science.adh2586 . [Online]. Available: https://www.science.org/doi/abs/10.1126/science.adh2586

[8] M. Uddin, S. U. Arfeen, F. Alanazi *et al.*, "A Critical Analysis of Generative AI: Challenges, Opportunities, and Future Research Directions," *Archives of Computational Methods in Engineering*, 2025. doi: 10.1007/s11831-025-10355-z . [Online]. Available: https://doi.org/10.1007/s11831-025-10355-z

[9] Editorial, "Using Large Language Models Wisely," *Nature Astronomy*, vol. 9, no. 315, 2025. doi: 10.1038/s41550-025-02524-8 . [Online]. Available: https://doi.org/10.1038/s41550-025-02524-8

[10] M. Fouesneau, I. G. Momcheva, U. Chadayammuri *et al.*, "What is the Role of Large Language Models in the Evolution of Astronomy Research?" 2024. [Online]. Available: https://arxiv.org/abs/2409.20252

[11] Editorial, "Why scientists trust AI too much — and what to do about it," *Nature Astronomy*, vol. 627, no. 243, 2024. doi: https://doi.org/10.1038/d41586-024-00639-y .

[12] Y. Ning, S. Teixayavong, Y. Shang *et al.*, "Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist," *The Lancet Digital Health*, vol. 6, no. 11, pp. e848–e856, 2024. doi: https://doi.org/10.1016/S2589-7500(24)00143-2 . [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589750024001432

[13] M. Al-kfairy, D. Mustafa, N. Kshetri *et al.*, "Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective," *Informatics*, vol. 11, no. 3, 2024. doi: 10.3390/informatics11030058 . [Online]. Available: https://www.mdpi.com/2227-9709/11/3/58

[14] Y. Sun, D. Sheng, Z. Zhou, and Y. Wu, "AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content," *Humanities and Social Sciences Communications*, p. 1278, 2024. doi: 10.1057/s41599-024-03811-x .

[15] S. Farquhar, J. Kossen, L. Kuhn, and Y. Gal, "Detecting hallucinations in large language models using semantic entropy," *Nature*, vol. 630, pp. 625–630, 2024. doi: 10.1038/s41586-024-07421-0 . [Online]. Available: https://doi.org/10.1038/s41586-024-07421-0

[16] A. Szalay and R. Brunner, "Astronomical archives of the future: a virtual observatory," *Future Generation Computer Systems*, vol. 16, no. 1, pp. 63–72, 1999. doi: https://doi.org/10.1016/S0167-739X(99)00036-9 . [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X99000369

[17] Gaia Collaboration, Prusti, T., de Bruijne, J. H. J. *et al.*, "The Gaia mission," *Astronomy and Astrophysics*, vol. 595, p. A1, 2016. doi: 10.1051/0004-6361/201629272 . [Online]. Available: https://doi.org/10.1051/0004-6361/201629272

[18] C. L. Borgman, P. T. Darch, A. E. Sands, and M. S. Golshan, "The Durability and Fragility of Knowledge Infrastructures: Lessons Learned from Astronomy," in *2016 Annual Meeting of the Association for Information Science and Technology, October 14-18, 2016, Copenhagen, Denmark*, 2016. [Online]. Available: https://arxiv.org/abs/1611.00055

[19] A. E. Sands, "Managing Astronomy Research Data: Data Practices in the Sloan Digital Sky Survey and Large Synoptic Survey Telescope Projects." 2017, PhD thesis, UCLA. [Online]. Available: https://escholarship.org/uc/item/80p1w0pm

[20] S. Djorgovski, "Virtual astronomy, information technology, and the new scientific methodology," in *Seventh International Workshop on Computer Architecture for Machine Perception (CAMP'05)*, 2005, pp. 125–132. doi: 10.1109/CAMP.2005.53 .

[21] R. Hanisch, G. Berriman, T. Lazio *et al.*, "The Virtual Astronomical Observatory: Re-engineering Access to Astronomical Data," *Astronomy and Computing*, vol. 11, pp. 190–209, 2015. doi: https://doi.org/10.1016/j.ascom.2015.03.007 ., the Virtual Observatory: II. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213133715000256

[22] E. J. Etemad and F. Rivoal, "W3C Process Document," 2025. [Online]. Available: https://www.w3.org/policies/process/20250818/

[23] L. Cambrésy, S. Derriere, P. Padovani *et al.*, "Ontology of Astronomical Object Types Use Cases, Version 1.1," IVOA Documents, 2010. [Online]. Available: https://www.ivoa.net/documents/Notes/AstrObjectOntologyUseCases/20100117/

[24] M. Demleitner, N. Gray, and M. Taylor, "Vocabularies in the VO, Version 2.1," IVOA Recommendation, 2023. [Online]. Available: https://www.ivoa.net/documents/Vocabularies/20230206/

[25] B. Cecconi, L. Debisschop, S. Derrière *et al.*, "OntoPortal-Astro, a Semantic Artefact Catalogue for Astronomy," *Astronomy and Computing*, vol. 53, p. 100991, 2025. doi: https://doi.org/10.1016/j.ascom.2025.100991 . [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213133725000642

[26] K. Frey and A. Accomazzi, "The Unified Astronomy Thesaurus: Semantic Metadata for Astronomy and Astrophysics," *The Astrophysical Journal Supplement Series*, vol. 236, no. 1, p. 24, may 2018. doi: 10.3847/1538-4365/aab760 . [Online]. Available: https://doi.org/10.3847/1538-4365/aab760

[27] D. Baron, "Machine Learning in Astronomy: a practical overview," 2019. [Online]. Available: https://arxiv.org/abs/1904.07248

[28] S. Djorgovski, A. Mahabal, M. Graham *et al.*, *Artificial Intelligence for Science.*

A Deep Learning Revolution. World Scientific, 2023, ch. Applications of AI in Astronomy, pp. 81–93. doi: https://doi.org/10.1142/13123 . [Online]. Available: https://www.worldscientific.com/worldscibooks/10.1142/13123

[29] Dutta, Samya and Paul, Prithwineel, “Deep space insights: Machine learning revolutionizing astrophysical discoveries,” *EPJ Web Conf.*, vol. 325, p. 01012, 2025. doi: 10.1051/epjconf/202532501012 . [Online]. Available: https://doi.org/10.1051/epjconf/202532501012

[30] M. J. Smith and J. E. Geach, “Astronomia ex machina: a history, primer and outlook on neural networks in astronomy,” *Royal Society Open Science*, vol. 10, no. 5, p. 221454, 2023. doi: 10.1098/rsos.221454 . [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsos.221454

[31] F. Grezes, S. Blanco-Cuaresma, A. Accomazzi *et al.*, “Building astroBERT, a language model for Astronomy Astrophysics,” 2021. [Online]. Available: https://arxiv.org/abs/2112.00590

[32] T. D. Nguyen, Y.-S. Ting, I. Ciucă *et al.*, “AstroLLaMA: Towards Specialized Foundation Models in Astronomy,” 2023. [Online]. Available: https://arxiv.org/abs/2309.06126

[33] R. Pan, T. D. Nguyen, H. Arora, A. Accomazzi, T. Ghosal, and Y.-S. Ting, “AstroMLab 2: AstroLLaMA-2-70B Model and Benchmarking Specialised LLMs for Astronomy,” 2024. [Online]. Available: https://arxiv.org/abs/2409.19750

[34] Roland, Benoit, Giffels, Manuel, Khalatyan, Arman, Sacchi, Elena, Enke, Harry, Huebner, Michael, Freyermuth, Oliver, Wissing, Christoph, Achkar, Baida, Drabent, Alexander, Hoeft, Matthias, and Gupta, Prateek, “Latest Developments of the PUNCH4NFDI Compute and Storage Infrastructures,” *EPJ Web Conf.*, vol. 337, p. 01264, 2025. doi: 10.1051/epjconf/202533701264 . [Online]. Available: https://doi.org/10.1051/epjconf/202533701264

[35] I. Ciucă and Y.-S. Ting, “Galactic ChitChat: Using Large Language Models to Converse with Astronomy Literature,” *Research Notes of the AAS*, vol. 7, no. 9, p. 193, sep 2023. doi: 10.3847/2515-5172/acf85f . [Online]. Available: https://doi.org/10.3847/2515-5172/acf85f

[36] S. T. Erukude and L. Shamir, “Galaxy Image Simplification Using Generative AI,” *Astronomy and Computing*, vol. 53, p. 100990, 2025. doi: https://doi.org/10.1016/j.ascom.2025.100990 . [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213133725000630

[37] K. McCormick, A. Hyk, I. Ciucă *et al.*, “Evaluating Large Language Models in Astronomy Research,” *Bulletin of the AAS*, vol. 57, no. 2, feb 28 2025, https://baas.aas.org/pub/2025n2i356p09.

[38] J. Pinochet, “Networked Observatory for Virtual Astronomy (NOVA): Teaching Astronomy with AI,” *Physics Education*, vol. 60, no. 4, p. 045034, Jun. 2025. doi: 10.1088/1361-6552/ade358 . [Online]. Available: http://dx.doi.org/10.1088/1361-6552/ade358

[39] L. C. D. Pinheiro, Z. Chen, B. C. Piazza, N. Shroff, Y. Liang, Y.-S. Ting, and H. Sun, “Large Language Models Achieve Gold Medal Performance at the International Olympiad on Astronomy Astrophysics (IOAA),” 2025. [Online]. Available: https://arxiv.org/abs/2510.05016

[40] A. d. Garcez and L. C. Lamb, “Neurosymbolic AI: the 3rd Wave,” *Artificial Intelligence Review*, vol. 56, pp. 12 387–12 406, 2023. doi: 10.1007/s10462-023-10448-w . [Online]. Available: https://doi.org/10.1007/s10462-023-10448-w

[41] B. C. Colelough and W. Regli, “Neuro-Symbolic AI in 2024: A Systematic Review,” in *LNSAI 2024: First International Workshop on Logical Foundations of Neuro-Symbolic AI, August 05, 2024, Jeju, South Korea*, vol. 3819. CEUR, 2024. [Online]. Available: https://ceur-ws.org/Vol-3819/paper3.pdf

[42] A. Sheth, K. Roy, and M. Gaur, " Neurosymbolic Artificial Intelligence (Why, What, and How) ," *IEEE Intelligent Systems*, vol. 38, no. 03, pp. 56–62, May 2023. doi: 10.1109/MIS.2023.3268724 . [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/MIS.2023.3268724

[43] X.-W. Yang, J.-J. Shao, L.-Z. Guo *et al.*, "Neuro-Symbolic Artificial Intelligence: Towards Improving the Reasoning Abilities of Large Language Models," in *Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25), Survey Track*, 2025. [Online]. Available: https://www.ijcai.org/proceedings/2025/1195.pdf

[44] S. Auer, J. D'Souza, K. E. Farfar *et al.*, *Open Research Knowledge Graph: A Large-Scale Neuro-Symbolic Knowledge Organization System*, ser. Frontiers in Artificial Intelligence and Applications. IOS Press, 2025, ch. Volume 400: Handbook on Neurosymbolic AI and Knowledge Graphs, pp. 385–420. doi: 10.3233/FAIA250216 .

[45] Q. Lu, R. Li, E. Sagheb *et al.*, "Explainable Diagnosis Prediction through Neuro-Symbolic Integration," *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, pp. 332–341, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12150699/

[46] U. Nawaz, M. A. ur Rahaman, and Z. Saeed, "A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems," *Intelligent Systems with Applications*, vol. 26, p. 200541, 2025. doi: https://doi.org/10.1016/j.iswa.2025.200541 . [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667305325000675

[47] O. Bougzime, S. Jabbar, C. Cruz, and F. Demoly, "Unlocking the Potential of Generative AI through Neuro-Symbolic Architectures: Benefits and Limitations," 2025. [Online]. Available: https://arxiv.org/abs/2502.11269