Formal Methods for Intersymbolic AI

Maurice H. ter Beek[1], Clemens Dubslaff[2,3], and Ina Schaefer[4]

[1] CNR–ISTI, Pisa, Italy
m.terbeek@isti.cnr.it
[2] Eindhoven University of Technology, Eindhoven, The Netherlands
c.dubslaff@tue.nl
[3] Centre for Tactile Internet with Human-in-the-loop (CeTI), Dresden, Germany
[4] KIT, Karlsruhe, Germany
ina.schaefer@kit.edu

Abstract. A key benefit of symbolic (rule-based) artificial intelligence (AI) is its formal rigor, which comes at the cost of formal modeling effort and computational expensive reasoning. Differently, subsymbolic (datadriven) AI approaches usually outperform rigorous ones in performance but might lead to unsound results. *Intersymbolic* AI is an emerging field in AI that aims to combine symbolic and subsymbolic AI approaches, exploiting the benefits from both worlds. The scope of the ISoLA 2025 track on "Formal Methods for Intersymbolic AI" is to gather researchers and practitioners from formal methods and (sub)symbolic AI to establish the boundaries of intersymbolic AI and to investigate and clarify the role of formal methods therein.

Keywords: Formal Methods for AI, AI-enabled Verification, Explainable AI

Motivation

In his keynote contribution during last year's ISoLA [4], Platzer [9] called for the study of the field that was coined *intersymbolic artificial intelligence (AI)*. This field targets the combination of *symbolic AI*, whose building blocks have inherent significance/meaning, with *subsymbolic AI*, whose entirety creates significance/effect despite the fact that individual building blocks escape meaning. Symbolic AI, as implemented in rule-based systems, provides formal rigor, but this comes at the cost of increased modeling effort and computationally expensive reasoning. Differently, subsymbolic AI approaches, which typically use data-driven methods from statistical learning, are not as computationally expensive as the rigorous ones but might lead to unsound results. The idea is that intersymbolic AI combines benefits from both symbolic and subsymbolic AI to increase the overall effectiveness, rigor, and explainability of AI compared to either kind of (sub)symbolic AI alone.

Instances of such combinations have already been established in the literature and showcase the broad applicability of the intersymbolic AI concept. The probably most

established instance of intersymbolic AI is in *neurosymbolic AI*, which focuses on the combination of symbolic and neural network reasoning [1,10]. Other instances have also been reflected in last year's ISoLA track on "X-by-Construction Meets AI", where several contributions involved intersymbolic AI. For example, an *intersymbolic programming language* was proposed, pairing logical primitives with training and prediction based on subsymbolic methods [5]. Within the area of *explainable AI* (XAI), an incarnation of intersymbolic AI by means of *logic-based XAI* was addressed [7]. In this strand of XAI, symbolic AI by means of logic reasoning is used to explain classifiers learned using subsymbolic AI approaches. Such approaches are particularly important since the operation of the most advanced AI models is often beyond the grasp of human decision makers. Much work on XAI relies on measures to quantify feature importance such as SHAP [6]. While such measures can give an indication of which are the relevant aspects in AI components, they cannot rigorously explain them. In high-risk or safety-critical domains, more formal approaches at different levels of abstraction are required to build the much needed trust [8].

In formal methods, explainability is an ongoing topic of research, turning formal correctness proofs on decision-making processes into interpretable explications [2]. Subsymbolic approaches such as reinforcement learning may assume an underlying formal model, e.g., by means of a Markov decision process (MDP). These models can be subject to formal methods, such as probabilistic model checking or explainability through formal notions of causality [3].

Research Questions

All of the above mentioned areas only provide a glimpse of the many research questions and research opportunities that are emerging from the combination of formal methods and AI, constituted in the field of intersymbolic AI: What is the role of formal methods in intersymbolic AI? How can formal methods ensure rigorous explanations of intersymbolic AI approaches? Is there a generic methodology for intersymbolic AI that provides the benefits from both symbolic and subsymbolic AI approaches? What are the lessons learned from applying formal methods for neurosymbolic AI or other forms of intersymbolic AI?

Track Format

The track on "Formal Methods for Intersymbolic AI" (FMIAI) at AISoLA 2025 is organized as a two-day event to foster collaboration and research in intersymbolic AI. It addresses researchers and practitioners from formal methods, symbolic or subsymbolic AI, and established fields of intersymbolic AI such as neurosymbolic AI and XAI. Topics accepted for presentation range from logicbased XAI, verification and explanation of neural networks, and combinations of statistical model checking and reinforcement learning, to large language and modal models in a variety of application domains.

Track participants have the opportunity to extend their presented contributions and include aspects discussed during the conference, and submit their work for publication in the forthcoming post-proceedings volume of AISoLA 2025.

References

1. Bader, S., Hitzler, P.: Dimensions of Neural-symbolic Integration - A Structured Survey. arXiv:cs/0511042 [cs.AI] (2005).https://doi.org/10.48550/arXiv.cs/0511042
2. Baier, C., et al.: From verification to causality-based explications. In: Bansal, N., Merelli, E., Worrell, J. (eds.) Proceedings of the 48th International Colloquium on Automata, Languages, and Programming (ICALP'21). LIPIcs, vol. 198, pp. 1:1–1:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2021). https://doi.org/10.4230/LIPICS.ICALP.2021.1
3. Baier, C., Piribauer, J., Ziemek, R.: Foundations of probability-raising causality in Markov decision processes. Log. Methods Comput. Sci. **20**(1), 4:1–4:66 (2024). https://doi.org/10.46298/LMCS-20(1:4)2024
4. ter Beek, M.H., Cleophas, L., Dubslaff, C., Schaefer, I.: X-by-construction meets AI. In: Margaria, T., Steffen, B. (eds.) Proceedings of the 12th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation: Software Engineering Methodologies (ISoLA'24). LNCS, vol. 15222, pp. 155–161. Springer (2024). https://doi.org/10.1007/978-3-031-75387-9
5. Belmonte, G., Bussi, L., Ciancia, V., Latella, D., Massink, M.: Towards hybrid-AI in imaging using VoxLogicA. In: Margaria, T., Steffen, B. (eds.) Proceedings of the 12th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation: Software Engineering Methodologies (ISoLA'24). LNCS, vol. 15222, pp. 205–221. Springer (2024). https://doi.org/10.1007/978-3-031-75387-9
6. Lundberg, S.M., Lee, S.: A Unified approach to interpreting model predictions. In: Guyon, I., et al. (eds.) Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). pp. 4768–4777. Curran Associates Inc. (2017). https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
7. Marques-Silva, J.: Logic-based explainability: past, present & future. In: Margaria, T., Steffen, B. (eds.) Proceedings of the 12th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation: Software Engineering Methodologies (ISoLA'24). LNCS, vol. 15222, pp. 181–204. Springer (2024). https://doi.org/10.1007/978-3-031-75387-9 12
8. Mienye, I.D., et al.: A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges. Inform. Med. Unlocked **51**, 101587 (2024). https://doi.org/10.1016/j.imu.2024.101587
9. Platzer, A.: Intersymbolic AI: Interlinking symbolic AI and subsymbolic AI. In: Margaria, T., Steffen, B. (eds.) Proceedings of the 12th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation: Software

Engineering Methodologies (ISoLA'24). LNCS, vol. 15222, pp. 162–180. Springer (2024). https://doi.org/10.1007/978-3-031-75387-9
10. Sheth, A.P., Roy, K., Gaur, M.: Neurosymbolic artificial intelligence (why, what, and how). IEEE Intell. Syst. **38**(3), 56–62 (2023). https://doi.org/10.1109/MIS.2023.3268724