Ontologies and Explainable AI

Luiz Olavo Bonino da Silva Santos [1] , Giancarlo Guizzardi [1]

[1] Semantics, Cybersecurity and Services group, University of Twente, The Netherlands
{l.o.boninodasilvasantos,g.guizzardi}@utwente.nl

As Artificial Intelligence (AI) systems become increasingly integrated into diverse domains, the demand for transparent and interpretable decision-making processes has grown significantly. Explainable AI (XAI) has emerged as a pivotal research area to address this need, aiming to elucidate the inner workings of complex AI models and provide comprehensible insights into their decision-making processes [2].

Explicit symbolic knowledge, in particular ontologies, can play different roles in Explainable AI and in the development of human-centric explainable systems and intelligible explanations.

By explicitly defining concepts, relationships, and hierarchies within a domain, ontologies facilitate a clearer understanding of the subject matter. These conceptualisations can be used to provide semantic grounding to data used and produced by AI algorithms, they enhance the transparency of AI models, allowing stakeholders to get a better comprehension of the underlying context in which the algorithms operate [3].

Ontologies can act as a bridge between human and machine understanding by providing a common semantic ground and representing knowledge in a format that is interpretable by both. This shared understanding fosters trust and ensures that the decisions made by AI systems align with human expectations and reasoning.

We consider three main perspectives in which ontologies can contribute significantly, namely reference modelling, common-sense reasoning, and knowledge refinement and complexity management [1].

In conclusion, ontologies play a multifaceted role in advancing the field of Explainable AI. From providing a structured representation of knowledge to facilitating contextual understanding and bridging communication gaps, ontologies emerge as tools in the quest for more transparent, trustworthy and comprehensible AI systems.

References

[1] Confalonieri, R. & Guizzardi, G. On the Multiple Roles of Ontologies in Explainable AI. (2023)

[2] Kapoor, S. & Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns.* **4** pp. 100804 (2023)

[3] Nakagawa, P., Pires, L., Moreira, J., Silva Santos, L. & Bukhsh, F. Semantic Description of Explainable Machine Learning Workflows for Improving Trust. *Applied Sciences.* **11** (2021), https://www.mdpi.com/2076-3417/11/22/10804