Can Ontologies help making Machine Learning Systems Accountable?

Iker Esnaola-Gonzalez

TEKNIKER, Basque Research and Technology Alliance (BRTA), Iñaki Goenaga 5, 20600 Eibar, Spain

1. Extended Abstract

Even though the maturity of the Artificial Intelligence (AI) technologies is rather advanced nowadays, according to McKinsey[1], its adoption, deployment and application is not as wide as it could be expected. This could be attributed to many barriers including cultural ones, but above all, the lack of trust of potential users in such AI systems.

The different factors that affect the users' trustworthiness on AI systems were studied in [1]. Some of these factors comprise the so-called Explainable Artificial Intelligence (XAI), which according to [2] refers to the "techniques that enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners". However, the explainability of AI systems is necessary but far from sufficient for understanding them and holding them accountable [3]. Therefore, in order to develop trustworthy AI systems, not only should they be explainable, but also accountable.

Accountability can be defined as the ability to determine whether a decision was made in accordance with procedural and substantive standards and to hold someone responsible if those standards are not met [3]. This means that with an accountable AI system, the causes that derived a given decision can be discovered, even if its underlying model's details are not fully known or must be kept secret.

Therefore, it seems reasonable to consider that the adequate representation of data, processes and workflows involved in AI systems could contribute to make them accountable. There are a variety of technologies that offer conceptual modelling capabilities to describe a domain of interest, but only ontologies combine this feature with Web compliance, formality and reasoning capabilities [4].

The usage of Semantic Technologies towards the achievement of Trustworthy AI has been researched in the literature [5, 6], but their full potential is not exploited yet. Towards the achievement of trustworthy AI systems, this article proposes an ontology-based approach aimed at providing Machine Learning systems with accountability. This approach consists of three phases as shown in Figure 1.

DAO-XAI 2021, 3rd International Workshop on Data meets Applied Ontologies
iker.esnaola@tekniker.es (I. Esnaola-Gonzalez)
0000-0001-6542-2878 (I. Esnaola-Gonzalez)

[1]https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain

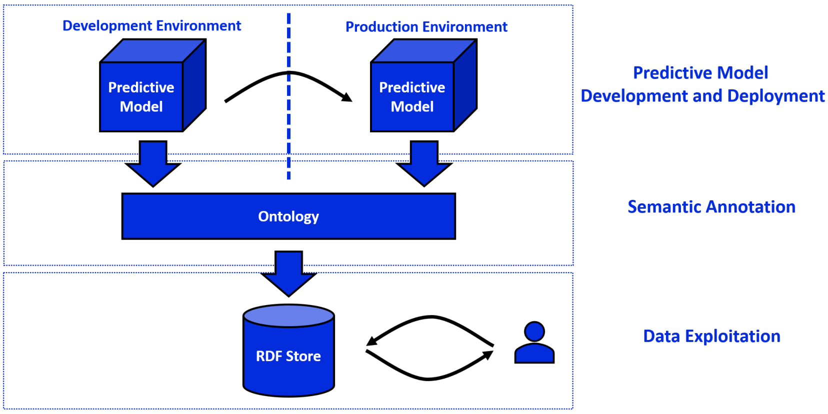

Figure 1: Outline of the proposed ontology-based approach.

The first phase is related to the development of the predictive model and its deployment in production where it will be executed. In the second phase both the procedure followed to develop the deployed predictive model and the results produced by the predictive model are annotated with the adequate ontology terms. As for the third phase, it is responsible for managing the annotations of the previous phase and facilitating their exploitation by users via SPARQL queries.

The topics that can be identified in Machine Learning system and that may be of interest to annotate with ontologies are: the forecast made by the predictive model, and the procedure followed for making such a forecast. Likewise, the latter procedure-related information can be divided in the information that addresses the training data and the information concerning the predictive model itself. After considering and evaluating the suitability of different ontologies, finally, three Ontology Design Patterns (the AffectedBy ODP[2], the Execution-Executor-Procedure (EEP) ODP[3] and the Result-Context (RC) ODP[4]) and the ML-Schema[5] have been chosen for representing this knowledge.

The full potential of Semantic Technologies to fill existing gaps and unsolved challenges towards trustworthy AI systems is yet to be unlocked. This article is aimed at paving the way for future research in this direction.

Acknowledgments

This work is partly supported by the project 3KIA (KK-2020/00049), funded by the SPRI-Basque Government through the ELKARTEK program and the AI-PROFICIENT project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 957391

[2]https://w3id.org/affectedBy
[3]https://w3id.org/eep
[4]https://w3id.org/rc
[5]http://www.w3.org/ns/mls

References

[1] B. Cahour, J.-F. Forzy, Does projection into use improve trust and exploration? an example with a cruise control system, Safety science 47 (2009) 1260–1270. doi:10.1016/j.ssci.2009.03.015.

[2] D. Gunning, Explainable artificial intelligence (xai), Defense Advanced Research Projects Agency (DARPA), nd Web 2 (2017).

[3] J. A. Kroll, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, H. Yu, Accountable algorithms, U. Pa. L. Rev. 165 (2016) 633–705.

[4] D. Oberle, How ontologies benefit enterprise applications 5 (2014) 473–491. doi:10.3233/SW-130114.

[5] I. Tiddi, F. Lécué, P. Hitzler, Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges, volume 47, IOS Press, 2020.

[6] A. Seeliger, M. Pfaff, H. Krcmar, Semantic web technologies for explainable machine learning models: A literature review, in: Joint Proceedings of PROFILES 2019 and SEMEX 2019, 1st Workshop on Semantic Explainability (SemEx 2019), co-located with the 18th International Semantic Web Conference (ISWC '19), volume 2465 of *PROFILES-SEMEX 2019*, CEUR-WS, 2019, pp. 30–45. URL: http://ceur-ws.org/Vol-2465/semex_paper1.pdf.