Semantic Web 00 (20xx) 1
IOS Press

Special Issue Introduction

The Role of Ontologies and Knowledge in Explainable AI

Editorial

Roberto Confalonieri [a,*], Oliver Kutz [b], Diego Calvanese [b,c], Jose Maria Alonso-Moral [d], Shang-Ming Zhou [e]
[a] *Department of Mathematics 'Tullio Levi-Civita', University of Padua, Padova, Italy*
E-mail: roberto.confalonieri@unipd.it
[b] *KRDB Research Centre for Knowledge and Data, Faculty of Engineering, Free University of Bozen-Bolzano, Bolzano, Italy*
E-mails: oliver.kutz@unibz.it, diego.calvanese@unibz.it
[c] *Department of Computing Science, Umeå University, Umeå, Sweden*
E-mail: diego.calvanese@unibz.it
[d] *Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain*
E-mail: josemaria.alonso.moral@usc.es
[e] *Centre for Health Technology, Faculty of Health, University of Plymouth, Plymouth, UK*
E-mail: shangming.zhou@plymouth.ac.uk

Keywords: Explainable AI, Symbolic Knowledge, Applied Ontologies

1. Motivations for and development of this special issue

Explainable Artificial Intelligence (XAI) has been identified as a key factor for developing trustworthy AI systems [1]. The reasons for equipping intelligent systems with explanation capabilities are not limited to user rights and acceptance. Explainability is also needed for designers and developers to enhance system robustness and enable diagnostics to prevent bias, unfairness, and discrimination, as well as to increase trust by all users in why and how decisions are made [1].

Defining and measuring interpretability of AI systems has been a matter of research for years [2], but it is still a hot topic in the computer science community due to the advances of big data, the more recent dramatic performance gains of large language models, and the evolution of AI regulation policy. For example, according to the European General Data Protection Regulation (GDPR), citizens have the legal right to an explanation of decisions made by algorithms that may affect them (see Article 22). This policy highlights the pressing importance of transparency and interpretability in algorithm design. Moreover, art.68c of AI Act is named "right to explanation" and emphasizes the

*Corresponding author. E-mail: roberto.confalonieri@unipd.it.

need to ensure the development and deployment of human-centered AI that is lawful, ethical and robust regarding both technical but also socio-economic perspectives.

XAI focuses on developing new explainable-by-design systems or generating and evaluating explanations of black-box models [1], thus achieving good explainability without sacrificing system performance. One typical approach is the extraction of local and global post-hoc explanations [3]. Other approaches are based on hybrid or neuro-symbolic systems [4], advocating a tight integration between symbolic and non-symbolic knowledge, e.g., by combining symbolic and statistical methods of reasoning.

The construction of hybrid systems is widely seen as one of the grand challenges facing AI today [5]. However, there is no consensus regarding how to achieve this, with proposed techniques in the literature ranging from knowledge extraction and tensor logic to inductive logic programming and other approaches. Knowledge representation—in its many incarnations—is a key asset to enact hybrid systems, and it can pave the way towards the creation of transparent and human-understandable intelligent systems.

This special issue is related to the Data meets Applied Ontologies Workshop (DAO-XAI)[1], an event that took place in co-location with the Bratislava Knowledge September (BASK) in September 2021[2]. BASK is a joint meeting of researchers, students, and industry professionals dealing with various aspects of knowledge processing. BAKS 2021 brought together the 30th International Conference on Artificial Neural Networks (ICANN 2021) and the 34th International Workshop on Description Logics (DL 2021), two events with a long tradition of research contributions related to sub-symbolic and symbolic reasoning respectively. The DAO-XAI Workshop was focused on the integration of sub-symbolic and symbolic reasoning, particularly, on the role played by explicit and formal knowledge, such as ontologies, knowledge graphs, knowledge bases, etc., in XAI. Authors of a selection of the papers presented at the workshop were invited to submit an extended version of their work to the special issue. In addition, we issued a world-wide open call for papers. We called for outstanding contributions dedicated to the role played by knowledge bases, ontologies, and knowledge graphs in XAI, in particular with regard to building *trustworthy* and *explainable* decision support systems. Knowledge representation plays a key role in XAI. Linking explanations to structured knowledge, for instance in the form of ontologies, brings multiple advantages. It does not only enrich explanations (or the elements therein) with semantic information—thus facilitating evaluation and effective knowledge transmission to users—but it also creates a potential for supporting the customisation of the levels of specificity and generality of explanations to specific user profiles or audiences. However, linking explanations, structured knowledge, and sub-symbolic/statistical approaches raise a multitude of technical challenges from the reasoning perspective, both in terms of scalability and in terms of incorporating non-classical reasoning approaches, such as defeasibility, methods from argumentation, or counterfactuals, to name just a few.

2. Contributions

The special issue attracted 10 submissions covering relevant areas of research. Six papers were finally accepted after two review rounds. Each paper was reviewed by 3 expert reviewers.

The accepted papers leveraged ontologies, knowledge graphs, and knowledge representation and reasoning in diverse ways. They can be classified into two distinct groups. One set of papers focused on proposing ontology specifications and extensions to enhance the conceptualization of user-centered explainable systems across various application domains, including chemistry, cyberbullying, finance, and data science. These papers introduced domain-specific ontologies, providing a structured framework to facilitate understanding and explanation of the systems within each domain. The other group of papers took a more foundational approach by presenting logic-based methodologies that fostered the development of explainable-by-design systems. These papers emphasized the use of logical reasoning techniques to achieve explainability and offered frameworks for constructing systems that inherently prioritize interpretability. In summary, the accepted papers demonstrated the utilization of ontologies, knowledge graphs, and knowledge representation and reasoning in advancing the field of XAI. In the following, we provide a broad overview of all the accepted papers.

[1]https://daoxai.inf.unibz.it/
[2]https://dai.fmph.uniba.sk/events/baks2021/

In the paper ‘Interpretable Ontology Extension in Chemistry’ by Martin Glauer, Adel Memariani, Fabian Neuhaus, Till Mossakowski, and Janna Hastings [6], the authors present a methodology for automatic ontology extension for domains in which the ontology classes have associated graph-structured annotations, and apply it to the ChEBI ontology, a prominent reference ontology for life sciences chemistry. Authors train Transformer-based deep learning models on the leaf node structures from the ChEBI ontology and the classes to which they belong. The models are then able to automatically classify previously unseen chemical structures, resulting in automated ontology extension. Visualization of the model’s attention weights support the explanations of the results by providing insight into how the model made its decisions.

In the paper ‘Explanation Ontology: A General-Purpose, Semantic Representation for Supporting User-Centered Explanations’ by Shruthi Chari, Oshani Seneviratne, Mohamed Ghalwash, Sola Shirai, Daniel M. Gruen, Pablo Meyer, Prithwish Chakraborty, and Deborah L. McGuinness [7], the authors proposed an explanation ontology and its extension to support user-centered explanations that make model recommendations more explainable. The explanation ontology is a general-purpose representation that is designed to help system designers connect explanations to their underlying data and knowledge. The ontology supports the specification of 15 literature-backed explanation types. Example of explanation type descriptions are described to show how to utilize the explanation ontology to represent explanations in five use cases spanning the domains of finance, food, and healthcare. The ontology has been released at https://purl.org/heals/eo.

In the paper ‘Data journeys: explaining AI workflows through abstraction’ by Enrico Daga and Paul Groth [8], the authors focus on the extraction and representation of data journeys from data science workflows involving multiple datasets, models, preparation scripts, and algorithms. A data journey is a multi-layered semantic representation of data processing activities linked to data science code and assets that provide a high level of abstraction. They propose an ontology to capture the essential elements of a data journey and an approach to extract such data journeys. The approach is evaluated using a corpus of Python notebooks from Kaggle.

In the paper ‘Engineering User-centered Explanations to Query Answers in Ontology-driven Socio-technical Systems’ by Juan Carlos L. Teze, Jose Nicolas Paredes, Maria Vanina Martinez, and Gerardo Ignacio Simari [9], the authors develop a line of research and development towards building tools that facilitate the implementation of explainable and interpretable hybrid intelligent socio-technical systems focusing on user-centered explanations. The implementation of a recently-proposed application framework for developing such systems is presented, and user-centered mechanisms are explored. The approach is validated with use cases of cyberbullying scenarios.

In the paper ‘Separability and Its Approximations in Ontology-based Data Management’ by Gianluca Cima, Federico Croce, and Maurizio Lenzerini [10], the authors tackle with the logical separability task in the context of Ontology-based Data Management (OBDM). Given two set of examples, the logical separability task seeks finding a formula in a certain target query language that separates them. When the input datasets of examples are treated as instances classified as positive or negative by a black-box model, the derived separating formula can be employed to offer global post-hoc explanations for the model’s behavior. Since a formula that properly separates two input datasets does not always exist, they propose best approximations of the proper separation and they present a general framework for separability in OBDM. Furthermore, they study three natural computational problems associated with the framework, namely verification, existence, and computation of the logical separability task.

In the paper ‘Searching for explanations of black-box classifiers in the space of semantic queries’ by Jason Liartis, Edmund Dervakos, Orfeas Menis-Mastromichalakis, Alexandros Chortaras and Giorgos Stamou [11], the authors tackle the challenge of extracting explanation rules from a black-box classifier, approaching it as a semantic query reverse engineering problem. In their study, the extracted rules are represented using the terminology of a knowledge graph. To ensure the reliability of the extracted rules, the authors offer guarantees and subsequently delve into exploring the relationship between explanation rules and semantic queries for a particular class. To solve this inverse problem, the authors develop algorithms that employ heuristic search within the semantic query space. These algorithms aim to find solutions efficiently and effectively. To evaluate the performance of the algorithms, the authors conduct simulations across four distinct use cases, providing a comprehensive analysis of their efficacy.

Acknowledgements

The guest editors of this special issue would like to thank Prof. Pascal Hitzler and Krzysztof Janowicz, Editors-in-Chief of the Semantic Web journal, for their great support in initiating and developing this special issue together. Many thanks to all members of the editorial team for their kind support during the editing process of this special issue. Last but not least, we would also like to thank the authors for submitting their valuable research outcomes as well as the reviewers who critically evaluated the papers. We sincerely hope and expect that readers will find this special issue useful.

References

[1] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J.M. Alonso-Moral, R. Confalonieri, R. Guidotti, J.D. Ser, N. Díaz-Rodríguez and F. Herrera, Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence, *Information Fusion* (2023), 101805. doi:https://doi.org/10.1016/j.inffus.2023.101805. https://www.sciencedirect.com/science/article/pii/S1566253523001148.

[2] R. Confalonieri, L. Coba, B. Wagner and T.R. Besold, A historical perspective of explainable Artificial Intelligence, *WIREs Data Mining and Knowledge Discovery* **11**(1) (2021). doi:https://doi.org/10.1002/widm.1391.

[3] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti and D. Pedreschi, A Survey of Methods for Explaining Black Box Models, *ACM Comp. Surv.* **51**(5) (2018), 1–42.

[4] A.d. Garcez and L.C. Lamb, Neurosymbolic AI: the 3rd wave, *Artificial Intelligence Review* (2023). doi:10.1007/s10462-023-10448-w.

[5] H. Kautz, The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture, *AI Magazine* **43**(1) (2022), 105–125. doi:10.1002/aaai.12036. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/19122.

[6] M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski and J. Hastings, Interpretable ontology extension in chemistry, *Semantic Web* **Preprint** (2023), 1–22, Preprint. doi:10.3233/SW-233183.

[7] S. Chari, O. Seneviratne, M. Ghalwash, S. Shirai, D.M. Gruen, P. Meyer, P. Chakraborty and D.L. McGuinness, Explanation Ontology: A general-purpose, semantic representation for supporting user-centered explanations, *Semantic Web* **Preprint** (2023), 1–31, Preprint. doi:10.3233/SW-233282.

[8] E. Daga and P. Groth, Data journeys: Explaining AI workflows through abstraction, *Semantic Web* **Preprint** (2023), 1–27, Preprint. doi:10.3233/SW-233407.

[9] J.C.L. Teze, J.N. Paredes, M.V. Martinez and G.I. Simari, Engineering user-centered explanations to query answers in ontology-driven socio-technical systems, *Semantic Web* **Preprint** (2023), 1–30, Preprint. doi:10.3233/SW-233297.

[10] G. Cima, F. Croce and M. Lenzerini, Separability and Its Approximations in Ontology-based Data Management, *Semantic Web* **Preprint** (2023), 1–36, Preprint. doi:10.3233/SW-233391.

[11] J. Liartis, E. Dervakos, O. Menis-Mastromichalakis, A. Chortaras and G. Stamou, Searching for explanations of black-box classifiers in the space of semantic queries, *Semantic Web* **Preprint** (2023), 1–42, Preprint. doi:10.3233/SW-233469.