

Replies to Reviewers' Comments

May 25, 2024

Journal:	Neurosymbolic AI
Manuscript Ref. #:	683-1663
Manuscript Title:	On the Multiple Roles of Ontologies in Explanations for Neuro-symbolic AI

We thank the reviewers for their time and effort in reviewing our manuscript. We applied all their valuable suggestions. All changes have been highlighted in blue in the revised version of the paper.

Reviewer 1:

This paper proposes an interesting overarching classification of approaches that use symbolic knowledge (in the form of ontologies) for ensuring explainability of AI systems, thus discussing an interesting and timely topic in the area of neurosymbolic systems. The material is thoroughly researched and points to a wide range of papers from various domains. The summaries of various representative works are easy to understand and supported by illustrative examples. The paper concludes with a number of challenges in using ontologies for explainable AI. The paper is suitable for publication as is. Possible, but not mandatory improvements could be:

(1) better highlight the relation and relevance to the neurosymbolic field in the abstract/introduction. While this relevance is clear once the paper is read, making it more prominent at the beginning of the paper would be important especially given that this paper will appear in an inaugural issue of the NAI journal.

Reply 1. *Thanks for the suggestion. In the revised paper, we updated the introduction to make the relationship of our paper with neuro-symbolic AI clearer and we added a sentence to the abstract. We also changed the title of the paper to 'On the Multiple Roles of Ontologies in Explanations for Neuro-symbolic AI'.* ■

(2) section 6 which discusses challenges could be more structured (e.g., include headings/bulleted list for each challenge) and potentially extended with additional challenges - in fact, we can assume that such understanding of challenges in an emerging field is of high interest to readers of the planned inaugural issue. Consider also renaming it to "Open Challenges".

Reply 2. *Thanks for the suggestion. We renamed Section 6 to 'Open Challenges' and we included a bulleted item for each challenge. Given the space limitation imposed by the editors, we were not able to add new challenges. Instead, we added a sentence making clear that these are some of the open challenges, and we pointed to [1, 2] where the interested reader can find other open challenges.* ■

Also, the summary of the challenges at the end of Section 7 does not fully overlap with what is discussed in Section 6, so a tighter alignment in content should be attempted.

Reply 3. *Thanks for the suggestion. We aligned the summary of the challenges of Section 7 with what we discussed in the perspectives and in Section 6.* ■

Reviewer 2:

This paper gives an overview of the different roles of ontologies in Explainable AI. They consider three roles for ontologies: (i) reference modelling, (ii) common-sense reasoning, (iii) knowledge refinement and complexity management. Along those roles, they discuss existing works. A very valuable overview of XAI methods that exploit ontologies. The paper gives indeed a nice overview. It would be nice to make the different categories more fine grained such methods can be more easily compared. However I do understand that it is also difficult. Main points:

(1) The most important suggestion is to include a paragraph in the introduction that explains why this explainability paper fits well in the new journal of Neurosymbolic. Include also adding a sentence in the abstract. In my opinion the paper fits the journal, because explainability methods are discussed that explain a black box (ML) by exploiting knowledge (reasoning). Include a more prominent link to neurosymbolic in your "summary" section as well.

Reply 4. *Thanks for the suggestion. Please see Reply 1.* ■

(2) The authors summarise a number of papers. Would it be possible to add motivation to the selection of papers?

Reply 5. *Thanks for the suggestion. We included a sentence summarizing the rationale behind the selection of the papers.* ■

(3) it would be nice to spend a paragraph on each challenge, explain and discuss the three challenges in more detail.

Reply 6. *Thanks for the suggestion. Each perspective is discussed in a dedicated section in details. We added a sentence summarizing the main challenge of each of the perspectives in the corresponding item list.* ■

include a more prominent link to neurosymbolic in your "summary" section.

Reply 7. *Thanks for the suggestion. We included a more prominent link to neurosymbolic AI in the summary. Furthermore, we aligned the content of the section with the precedent section (see also Reply 3).* ■

Smaller remarks: page 2, row 48: "AI systems" do you mean here ML systems? page 8, row 12: "depends on the specific requirements of each application". A concrete example might be helpful.

Reply 8. *Thanks for the suggestion. We applied all the remarks.* ■

References

- [1] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J.D. Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, R. Jiang, H. Khosravi, F. Lecue, G. Malgieri, A. Páez, W. Samek, J. Schneider, T. Speith and S. Stumpf, Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions, *Information Fusion* **106** (2024), 102301. doi:<https://doi.org/10.1016/j.inffus.2024.102301>. <https://www.sciencedirect.com/science/article/pii/S1566253524000794>.
- [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>.