Towards an Ontology-Driven Approach to Document Bias (Abstract Reprint)

Mayra Russo[1,2], **Maria-Esther Vidal**[3,1,2]

[1]L3S Research Center, Germany
[2]Leibniz University Hannover, Germany
[3]TIB Leibniz Information Centre for Science and Technology, Germany

Abstract Reprint. This is an abstract reprint of the journal article by Russo and Vidal (2025).

Abstract

Machine learning (ML)-powered systems are capable of reproducing and often amplifying undesired biases embedded in society, emphasizing the importance of operating under practices that enable the study and understanding of the intrinsic characteristics of ML pipelines. This supports the emergence of documentation frameworks with the idea that any remedy for bias starts with awareness of its existence. However, a resource that can formally describe ML pipelines in terms of detected biases is still missing. To address this gap, we present the Doc-BiasO ontology, a resource that sets out to create an integrated vocabulary of biases defined in the Trustworthy AI literature and their measures, as well as to incorporate relevant domain terminology and relationships between them. Overseeing ontology engineering best practices, we reuse existing vocabularies on machine learning and AI to foster knowledge sharing and interoperability between the actors concerned with its research, development, regulation, and others. In addition, we demonstrate the potential of Doc-BiasO with an experiment on an existing benchmark and as part of a neuro-symbolic system. Overall, our main objective is to contribute towards clarifying existing terminology on bias research as it rapidly expands to all areas of AI and to improve the interpretation of bias in data and downstream impact through its documentation.

References

Russo, M.; and Vidal, M.-E. 2025. Towards an Ontology-Driven Approach to Document Bias. *Journal of Artificial Intelligence Research*, 83: 38:1–38:35.