

A Little of That Human Touch: Achieving Human-Centric Explainable AI via Argumentation

Antonio Rago

King's College London, United Kingdom


Abstract

As data-driven AI models achieve unprecedented feats across previously unthinkable tasks, the diminishing levels of interpretability of their increasingly complex architectures can often be sidelined in place of performance. If we are to comprehend and trust these AI models as they advance, it seems clear that symbolic methods, given their unparalleled strengths in knowledge representation and reasoning, can play an important role in explaining AI models. In this talk, I discuss some of the ways in which one branch of such methods, computational argumentation, given its human-like nature, can be used to tackle this problem. I first outline a general paradigm for this area of explainable AI, before detailing a prominent methodology therein which we have introduced. I then illustrate how this approach has been put into practice with diverse AI models and types of explanations, before looking ahead to challenges, future work and the outlook in this field.

Declaration on Generative AI

The author has not employed any Generative AI tools.

23rd International Workshop on Nonmonotonic Reasoning, November 11-13, 2025, Melbourne, Australia

 0000-0001-5323-7739 (A. Rago)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).