

# How Can We Make Trustworthy AI?

Mateja Jamnik   

Department of Computer Science and Technology, University of Cambridge, UK

---

## Abstract

Not too long ago most headlines talked about our fear of AI. Today, AI is ubiquitous, and the conversation has moved on from whether we should use AI to how we can trust the AI systems that we use in our daily lives. In this talk I look at some key technical ingredients that help us build confidence and trust in using intelligent technology. I argue that intuitiveness, interaction, explainability and inclusion of human domain knowledge are essential in building this trust. I present some of the techniques and methods we are building for making AI systems that think and interact with humans in more intuitive and personalised ways, enabling humans to better understand the solutions produced by machines, and enabling machines to incorporate human domain knowledge in their reasoning and learning processes.

**2012 ACM Subject Classification** Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning

**Keywords and phrases** AI, human-centric computing, knowledge representation, reasoning, machine learning

**Digital Object Identifier** 10.4230/LIPIcs.FSCD.2023.2

**Category** Invited Talk



© Mateja Jamnik;

licensed under Creative Commons License CC-BY 4.0

8th International Conference on Formal Structures for Computation and Deduction (FSCD 2023).

Editors: Marco Gaboardi and Femke van Raamsdonk; Article No. 2; pp. 2:1–2:1

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany