

## Invited talk

# Bias in AI-systems: A multi-step approach

Eirini Ntoutsis

ntoutsis@kbs.uni-hannover.de

*Leibniz Universität Hannover & L3S Research Center*

## Abstract

Algorithmic-based decision making powered via AI and (big) data has already penetrated into almost all spheres of human life, from content recommendation and healthcare to predictive policing and autonomous driving, deeply affecting everyone, anywhere, anytime. While technology allows previously unthinkable optimizations in the automation of expensive human decision making, the risks that the technology can pose are also high, leading to an ever increasing public concern about the impact of the technology in our lives. The area of responsible AI has recently emerged in an attempt to put humans at the center of AI-based systems by considering aspects, such as fairness, reliability and privacy of decision-making systems.

In this talk, we will focus on the fairness aspect. We will start with understanding the many sources of bias and how biases can enter at each step of the learning process and even get propagated/amplified from previous steps. We will continue with methods for mitigating bias which typically focus on some step of the pipeline (data, algorithms or results) and why it is important to target bias in each step and collectively, in the whole (machine) learning pipeline. We will conclude this talk by discussing accountability issues in connection to bias and in particular, proactive consideration via bias-aware data collection, processing and algorithmic selection and retroactive consideration via explanations.

