Leibniz Universität Hannover

Conference Program

Bias and Discrimination in Big Data and Algorithmic Processing

April 27/28, 2023

Thursday, April 27

From 8:30 *(local time in Hannover: CEST)*	Conference Opening
9:15 - 9:30	**Welcome and Introduction** Dietmar Hübner
9:30 - 10:45	**John Danaher** Will large language models spark a moral revolution? *Chair: tba*
10:45 - 11:00	Coffee
11:00 - 12:15	**Toon Calders** The fairness-accuracy tradeoff revisited *Chair: tba*
12:15 - 13:30	Lunch
13:30 - 14:45	**Jan Horstmann, Arjun Roy** Expanding fairness in machine learning: Beyond single-task learning & single protected attributes *Chair: tba*
14:45 - 15:00	Coffee

15:00 - 16:15	**Gianclaudio Malgieri** Pre-discrimination law: Debiasing AI through vulnerable subjects' participation and contestation *Chair: Hannah Ruschemeier*
16:15 - 17:30	**Uljana Feest** Algorithmic bias in psychological research *Chair: Mathias Frisch*
18:30	Dinner at Meier’s Lebenslust

Friday, April 28

9:30 - 10:45	**Atoosa Kasirzadeh** Artificial general intelligence in a structurally unjust world *Chair: Lucie White*
10:45 - 11:00	Coffee
11:00 - 12:15	**Christian Heinze, Caroline Gentgen-Barg, Jan Horstmann** The EU AI Act and its consequences for the regulation of bias in AI systems *Chair: Andreas Sesing-Wagenpfeil*
12:15 - 13:30	Lunch
13:30 - 14:45	**Symeon Papadopoulos** Bias in computer vision and multimodal settings *Chair: Wolfgang Nejdl*
14:45 - 15:00	Coffee

15:00 - 16:15	**Raphaële Xenidis** Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law *Chair: Timo Rademacher*
16:15	**Closing Remarks**
	Coffee

Funded by: VolkswagenStiftung

Additional Funding: GAP - Gesellschaft für analytische Philosophie

Location: The conference will be held in the historic old town of Hannover at the Leibnizhaus:

Holzmarkt 4 - 6
30159 Hannover

Conference Dinner: A conference dinner will take place at *Meiers Lebenslust* on the evening of the 27th:

Osterstraße 64
30159 Hannover

Contact: If you have any questions, please contact Dietmar Hübner (dietmar.huebner@philos.uni-hannover.de).

Institute for Information Processing

Institute of Philosophy

L3S Research Center

Institute for Legal Informatics

Abstracts:

April 27 – 9:30
John Danaher – University of Galway
Will large language models spark a moral revolution?
The idea that technologies can change, possibly even revolutionise, moral beliefs and practices is an old one. But how, exactly, does this happen? This talk builds on an emerging field of inquiry by developing a synoptic taxonomy of the mechanisms of techno-moral change. It argues that technology affects moral beliefs and practices in three main domains: decisional (how we make morally loaded decisions), relational (how we relate to others) and perceptual (how we perceive situations). It argues that across these three domains there are six primary mechanisms of techno-moral change: (i) changing options; (ii) changing decision-making costs; (iii) enabling new relationships; (iv) changing the burdens and expectations within relationships; (v) changing the balance of power in relationships; and (vi) changing data, mental models and metaphors. If changes across these six domains are sufficiently widespread, rapid and longlasting, they could prompt a 'moral revolution'. Using the specific case study of large language models, particularly the various iterations of GPT, the talk considers how this technology might transform, and potentially, revolutionise our social morality in the near future.

April 27 – 11:00
Toon Calders – University of Antwerp
The fairness-accuracy tradeoff revisited

Demographic parity, equality of opportunity, calibration, individual fairness, direct and indirect discrimination: these are just a few of the many measures for bias in data and algorithms. Although for each of these measures strong arguments in favor can be found, it has been shown that they cannot be combined in a meaningful way. What is the right measure is hence commonly accepted to be "situation-dependent" and in the eye of the beholder. Nevertheless, unfortunately, surprisingly little guidelines for selecting the right measure are available for practitioners. A second issue in fairness-aware machine learning is the perception that we need to give up something in order to attain fair models: the so-called "fairness-accuracy trade-off". Arguably, this assumption is in many situations counter-intuitive given the goal of fair machine learning of undoing unfair bias. Thirdly, I believe that for many fairness-aware algorithms we do not properly understand and subsequently ignore *how* they satisfy the fairness constraints, which, as I will argue, may lead to even more unfair decision procedures. In this talk I will go deeper into these issues and end with proposing an alternative way of looking at fairness-aware machine learning as optimizing accuracy in a theoretical fair world.

April 27 – 13:30
Arjun Roy – University of the Bundeswehr Munich
Learning to teach fairness-aware deep multi-task learning

Fairness-aware learning mainly focuses on single task learning (STL). The fairness implications of multi-task learning (MTL) have only recently been considered and a seminal approach has been proposed that considers the fairness-accuracy trade-off for each task and the performance trade-off among different tasks. Instead of a rigid fairness-accuracy trade-off formulation, we propose a flexible approach that learns how to be fair in a MTL setting by selecting which objective (accuracy or fairness) to optimize at each step. We introduce the L2T-FMT algorithm that is a teacher-student network trained collaboratively; the student learns to solve the fair MTL problem while the teacher instructs the student to learn from either accuracy or fairness, depending on what is harder to learn for each task. Moreover, this dynamic selection of which objective to use at each step for each task reduces the number of trade-off weights from 2T to T, where T is the number of tasks. Our experiments on three real datasets show that L2T-FMT improves on both fairness

(12–19%) and accuracy (up to 2%) over state-of-the-art approaches.

Jan Horstmann – Leibniz University Hannover, Arjun Roy – University of the Bundeswehr Munich
Multi-dimensional concepts of discrimination in law and machine learning

The vast majority of proposed methods in fairness-aware machine learning assess fairness based on a single protected attribute, e.g. only gender or race. In reality, though, human identities are multi-dimensional, and discrimination can occur based on more than one protected characteristic. Taking inspiration from legal concepts to analyse multi-dimensional discrimination, we survey if and how these have been transferred/operationalized in fairness-aware machine learning.

April 27 – 15:00
Gianclaudio Malgieri – Leiden University
Pre-discrimination law: Debiasing AI through vulnerable subjects' participation and contestation

The regulation of Automated Decision-Making is a key factor of the GDPR and the proposed AI Act. The real focus should be on the impact of biased algorithms on vulnerable populations. Discrimination law has proven ineffective in dealing with new forms of (induced and often even unconscious) vulnerability. Ex-ante design tools should help, in particular, contestability and participative design of AI.

April 27 – 16:15

Uljana Feest – Leibniz University Hannover

Algorithmic bias in psychological research

This talk will examine the use of machine learning models in personality research. I will note that such models should be regarded as measurement tools, and I will ask how they fare with regard to standard criteria of test evaluation, such as validity. I will begin by (1) providing an overview of the big-five model of personality, followed by (2) an overview of the notion of construct validity. I will then (3) show that even though the big five model of personality has long been regarded as having construct validity, there remain open questions concerning the theoretical interpretation of those factors. I will then (4) argue that while ML models add to the construct validation of the big five, they have similar shortcomings as traditional measures of the big five. This will prompt me to (5) consider questions about possible positive contributions ML models might make to personality research, while also (6) taking a closer look at potential problems of algorithmic biases that might be said to arise from the data that ML models are trained on.

April 28 – 9:30

Atoosa Kasirzadeh – University of Edinburgh

Artificial general Intelligence in a structurally unjust world

Discussing artificial general intelligence (AGI) is no longer limited to the realm of science fiction. At least two major tech players, DeepMind and OpenAI, assert that they are working towards building AGI with the potential to "benefit all of humanity." This paper highlights a significant challenge faced by these endeavours: the development of AGI is taking place within a structurally unjust world. By examining the potential negative implications of AGI, I will argue that certain social and political complexities must be considered to ensure that AGI development aligns with benefiting all of humanity. Otherwise, some of humanity will be worse off.

April 28 – 11:00

Christian Heinze – University of Heidelberg

The EU AI Act and its consequences for the regulation of bias in AI systems

Caroline Gentgen-Barg – Leibniz University Hannover

AI Act and anti-discrimination regulation

Jan Horstmann – Leibniz University Hannover

AI Act and human oversight

The proposed EU AI Act and its neighbouring instruments - which are currently debated by the European legislative institutions - will provide an EU-wide legal framework that specifically addresses AI systems. The presentations will highlight general issues of the proposed Act (such as the interpretation of the terms of the Act, the definition of AI systems, an overview of its main rules, and the rights of possible subjects of AI systems) and focus on its consequences for bias in AI systems and human oversight

April 28 – 13:30

Symeon Papadopoulos – Centre for Research and Technology Hellas

Bias in computer vision and multimodal settings

Abstract: AI bias is a well known issue affecting the performance of data-driven AI systems and an increasing number of incidents are reported about AI systems and applications that exhibit discriminatory behaviour against vulnerable or underrepresented groups of people. This talk will attempt to move beyond the typical AI bias setting involving tabular data and a well-defined objective and to provide an overview of the emerging problem of visual and multimodal AI bias, especially in connection with Computer Vision algorithms and applications, and discuss recent developments, methodologies and challenges.

April 28 – 11:00

Raphaële Xenidis – Science Po Law School

Two round holes and a square peg: An alternative test for algorithmic discrimination in EU equality law

Algorithmic bias pervades numerous areas of life and worsens inequalities. Yet EU equality law only offers limited legal remedies. On the one hand, its central doctrinal categories – direct and indirect discrimination – are analytically ill-suited to capturing algorithmic types of discrimination. On the other hand, the prevailing system of ex post individual redress does not effectively address the reality of AI predictions, which turn past discrimination into a self-fulfilling prophecy. To remedy these fundamental inadequacies, this article proposes an alternative legal test, seeking resilience in the margins of the current regulatory regime. It argues that conceptualising discriminatory algorithms as ‘instructions to discriminate’ fosters legal certainty and equality by shifting responsibility for algorithmic discrimination away from society and to the users who draw profit from AI systems. This alternative test centres positive action by creating a ‘duty to reasonably debias’, which in turn stimulates the creation of prevention ecosystems.