Workshop

Responsibility & AI

fonti.univie.ac.at/workshops/responsibility-ai/

universität wien

Thursday, May 12, Hörsaal 3B

10:00 Welcome & Intro

10:30 Mark Coeckelbergh: Democracy, epistemic agency, and AII

11:45 Fiorella Battaglia: AI and Responsibility.
Reflections on recent accounts in philosophy and ethics of technology

12:45 Joint Lunch (Room change: Hörsaal 2H NIG)

15:00 Laura Crompton: The decision-point-dilemma:
yet another problem of responsibility in human-AI interaction

16:15 Andrea Bertolini: AI does not exist! Why a technology-neutral approach to liability regulation for advanced technology is bad policy

Friday, May 13, Hörsaal 3D

10:00 Atoosa Kasirzadeh: Responsible Algorithmic Fairness:
Insights from Feminist Political Philosophy

11:15 Harald Leitenmüller: From Principles to Practice

12:15 Joint Lunch

14:00 Mario Günther: When Should We Attribute Beliefs to AI Systems?

15:15 Vincent Müller: Digital Philosophy: A Programme

Neues Institutgebäude (NIG)
Universitätsstraße 7
1010 Vienna
2022

Democracy, epistemic agency, and AI

Mark Coeckelbergh (University of Vienna)

In this talk I introduce my recent work on the politics of AI, including my new book The Political Philosophy of AI. After giving an overview of the different themes of the book, I zoom in on AI's impact on democracy, in particular on the relation between democracy, epistemic agency, and AI.

AI and Responsibility. Reflections on recent accounts in philosophy and ethics of technology

Fiorella Battaglia (University of Salento)

In the last decade, the debate on agency and responsibility has taken a dramatic turn. Along with the efforts to conceive of a continuum of agency that lies between amoral and fully autonomous moral agents, serious concerns have arisen about the impossibility—even for humans—to meet the epistemic condition for responsibility (relevant information), let alone the question whether they have control over their actions mediated by technology. As a result, responsibility gaps have flourished anywhere in the literature. The most influential positions in this debate include either a relational perspective or an appeal for meaningful human control. The aim of this talk is to explore these distinctive proposals by articulating their thematic strands, but also by showing the actual relations between them.

The decision-point-dilemma: yet another problem of responsibility in human-AI interaction

Laura Crompton (University of Vienna)

AI as decision support supposedly helps human agents make 'better' decisions more efficiently. However, research shows that it can, sometimes greatly, influence the decisions of its human users. In this talk I aim to shed some light on the ethical and moral concerns that arise with AI influence. I argue that AI influence has important implications for the way we perceive and evaluate human-AI interaction. To make this point approachable from both the theoretical and practical side, and to avoid anthropocentrically-laden ambiguities, I introduce the notion of decision points. Based on this, the main argument of this talk will be presented in two consecutive

steps: i) unintended AI influence doesn't allow for an appropriate determination of decision points—this will be introduced as decision-point-dilemma, and ii) this has important implications for the ascription of responsibility.

AI does not exist! Why a technology-neutral approach to liability regulation for advanced technology is bad policy

Andrea Bertolini (Sant'Anna Scuola Superiore Pisa)

AI is typically referred to as a single technological object requiring regulation. However, the breadth of the spectrum of applications typically falling under this notion is such that no real common element may be found. Categorisation with respect to levels of risk is also insufficient in capturing such diversity. At the same time, the pervasive nature of AI will cause it to spread in most domains of human activity that, up until today, where considered and regulated separately by policymakers. Adopting a technology-neutral (one-size-fits-all) approach will therefore results in a bad, ineffective and potentially harmful policy, limiting innovation and not shaping desirable incentives for the different parties involved.

Responsible Algorithmic Fairness: Insights from Feminist Political Philosophy

Atoosa Kasirzadeh (University of Edinburgh)

Data-driven predictive algorithms are widely used to automate and guide high-stake decision making such as bail recommendation and medical resource allocation. Nevertheless, harmful outcomes biased against vulnerable groups have been reported. The growing research field known as Algorithmic Fairness aims to mitigate the harmful biases. The methodology consists in proposing mathematical metrics to address the social harms resulting from an algorithm's biased outputs. The metrics are typically motivated by – as well as substantively rooted in – concrete ideals of distributive justice, as formulated by political theorists and philosophers. The perspectives of feminist political philosophers on social justice, by contrast, has been largely neglected. Feminist political philosophers have critizised the paradigm of distributive justice and have proposed corrective amendments to surmount its limitations. The present paper brings key insights of feminist political philosophy to Algorithmic Justice. The paper has three goals. First, I show that the current scope of Algorithmic Fairness cannot accommodate the concerns of social justice

as identified by feminist political philosophy. Second, I defend the relevance of structural injustices – as pioneered in the contemporary philosophical literature by Iris Marion Young – to Algorithmic Fairness. Third, drawing upon Young's notion of social justice, I take some steps in developing the paradigm of Responsible Algorithmic Fairness to correct for errors in the current implementation of the Algorithmic Fairness paradigm.

From Principles to Practice
Harald Leitenmüller (Microsoft)

Let's have an overview of the why, the what and the how regarding the adoption of a Responsible AI strategy and approach for your AI-powered solutions, illustrated through the lenses of the Microsoft journey, from Microsoft's core Responsible AI principles to the way these principles translate into a framework of requirements, guidance, and governance through the Responsible AI standard and related practices.

When Should We Attribute Beliefs to AI Systems?
Mario Günther (LMU Munich)

How should we explain a decision made by an AI system to a layperson? It has been suggested that an AI system can be seen as a rational agent whose behavior is explainable in terms of beliefs and desires. This answer requires a theory that allows us to attribute beliefs to an AI system in a justified way. Here we propose such a theory of belief attribution which helps the layperson to understand how a given AI system works on a level that abstracts away from its underlying probability calculations. We argue that the resulting explanations engender trust in AI systems where trust is appropriate.

Digital Philosophy: A Programme
Vincent Müller (Technical University of Eindhoven)

Philosophy is methodologically in the air after the silent death of analytic philosophy and ordinary language philosophy – but we still do "conceptual analysis", mostly separate from empirical insight. Attempts to set foundations for the discipline

through epistemology, then logic, then philosophy of language, then philosophy of mind have failed. We pretend that there is such a thing as 'merely applying' philosophy to other areas such as technology or science. – There is a way out of all this: We can do classical conceptual analysis in a methodologically sound way, without the need for a new philosophical 'foundation'; while reviving a proper way of 'applied' philosophy. How? We can do this by find conceptual problems and trying out solutions in digital technology: Call it "digital philosophy". I will introduce the concept, then sketch some examples from the more distant or more recent past, and finally provide an outlook for the use of this method.