

From epistemic opacity to ethical opacity?

A workshop on non-transparent models and ethical uncertainty in science, technology, and engineering

Co-organizers: Viola Schiaffonati, Andreas Kaminski

Location: High-Performance Computing Center (HLRS), University of Stuttgart, room: "Kino"

Date: July 7, 2022

Computer models such as those used in advanced simulations and machine learning are often epistemically opaque. Individuals encounter a limit to explaining, understanding, or even justifying their results. Nevertheless, scientific, economic, and political decisions are to be made on their basis. Values necessarily enter into models. What are the consequences of epistemic opacity for questions of justification? How is the ethical evaluation of models and their results affected by it? Does epistemic lead to ethical opacity?

This workshop will explore novel theoretical frameworks to tackle the inherent uncertainty in different fields of science and engineering. The idea is to discuss how different disciplinary approaches can interact to deal with these issues. The focus will be mainly on issues related to epistemic opacity and ethical uncertainty and their relation to decision-making.

The workshop will include invited speakers, a discussion on relevant literature with (post-)graduate students from different research areas, and a final roundtable summarizing some results and possible future pathways.

Speakers: Viola Schiaffonati (Politecnico di Milano), Philip J. Nickel (TU Eindhoven), Markus Pantsar (Helsinki/Aachen), Andreas Kaminski (Aachen/HLRS)

Schedule

- 10:00** Viola Schiaffonati: Explorative Experiments:
A Paradigm Shift to Deal with Uncertainty in Autonomous Robotics
- 11:00** Philipp Nickel: Trusting AI systems under opacity
- 12:00** Lunch
- 13:30** Markus Pantsar: Theorem proving AI:
The prospect of genuine mathematical artificial intelligence
- 14:30** Andreas Kaminski: From epistemic to ethical opacity
- 15:30** Round-table

Abstracts

Trusting AI systems under opacity

Philip J. Nickel
Dept of Philosophy and Ethics
Eindhoven Artificial Intelligence Systems Institute
TU Eindhoven

Abstract: What is it to trust AI systems? In this presentation I put forward an account in terms of giving such systems *discretionary authority*. Professionals normally have discretionary authority for their own decisions within a work context (e.g., professors have authority to assign grades to students). AI systems disrupt this authority, requiring professionals to reconsider their own authority (e.g., AI systems that suggest student grades can reduce or augment the authority of professors.) Opacity of AI, or the lack of insight about the outputs of AI systems, may seem to threaten trust, but on the contrary, it is one of the reasons why we speak of trust in such contexts. Professionals often do not know in advance how and when to trust AI systems; working through this problem is an important dimension of the disruption.

Explorative Experiments: A Paradigm Shift to Deal with Uncertainty in Autonomous Robotics

Viola Schiaffonati
Department of Electronics, Information and Bioengineering
Politecnico di Milano

Abstract: With autonomous and intelligent systems uncertainty emerges from the complexity of the systems and their interaction with unknown environments. The management of uncertainty is not only a technical problem, but a philosophical one. To address this philosophical problem in this presentation I introduce the novel framework of explorative experiments. This framework presents a suitable context in which some of the issues relative to uncertainty, both at the epistemological level and at the ethical one, in this field should be reframed. The case of autonomous robot systems for search and rescue is used to make the discussion more concrete.

Theorem proving AI: The prospect of genuine mathematical artificial intelligence

Markus Pantsar
University of Helsinki
Department of Philosophy, History and Art Studies
Käte Hamburger Kolleg RWHT Aachen: Cultures of Research

Abstract: Computer assisted theorem proving is an increasingly important part of mathematical methodology, as well as a long-standing topic in artificial intelligence (AI) research. However, the current generation of theorem proving software shows little or no intelligence. Importantly, they are not able to discriminate interesting theorems and proofs from trivial ones. In order for computers to develop intelligence in theorem proving, there would need to be a radical change in how the software functions.

Recently, machine learning results in solving mathematical tasks have shown promise that deep artificial neural networks can learn symbolic mathematical processing. In this paper, I analyze the prospects of such neural networks in proving mathematical theorems, including the possibility that they could develop genuine intelligence. I argue that their intelligence may be hard to establish due to the “black box” problem of explaining the processing of artificial neural networks. A more feasible way to assess artificial mathematical intelligence is through the role that such an AI could play in the mathematical community.

From epistemic to ethical opacity?

Andreas Kaminski

RWTH Aachen, Chair for Philosophy of Science and Technology

University Stuttgart, HLRS, Department for Philosophy of Computational Science

Abstract: The fact that computer models can be opaque has so far been discussed primarily from an epistemic point of view. It was about how opacity makes understanding, explaining or even justifying processes and results problematic. Less focus has been placed on the ethical dimension. The lecture will examine the possible connection between epistemic and ethical opacity and on this basis explain why ethical opacity raises the question of the trustworthiness of computer models.

The Workshop is a Scientific Activity of the Project BRIO (2020SSKZ7R) awarded by the Italian Ministry of University and Research (MUR) (sites.unimi.it/brio)