Artificial Intelligence & Human Autonomy
Philosophical Explorations

Sofia Bonicalzi (University of Roma Tre)

Autonomy in Action: Understanding Agency in the Context of AI

Simona Tiribelli (University of Macerata)

Relational Autonomy and Artificial Intelligence: A Missing Piece

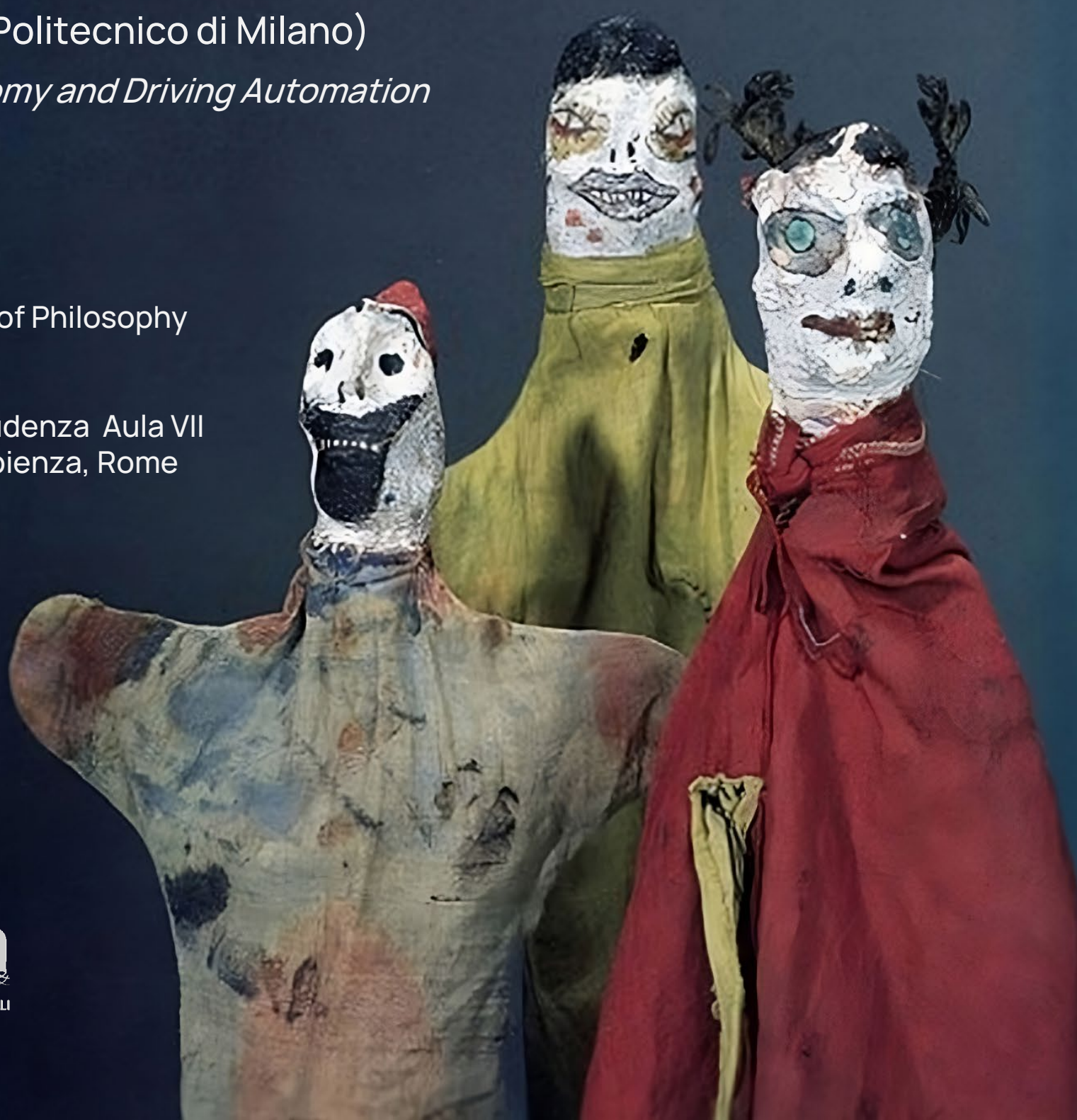
Radu Uszkai (University of Bucharest)

Show, Don't Tell! Designing AI Tools for Autonomy and Moral Enhancement

Fabio Fossa (Politecnico di Milano)

Human Autonomy and Driving Automation

World Congress of Philosophy
August 8th, 2024
h. 09.00-11.00
CU002 Giurisprudenza Aula VII
University La Sapienza, Rome



supported by

POLITECNICO
MILANO 1863

META
UNITÀ DI STUDI UMANISTICI E SOCIALI
SU SCIENZA E TECNOLOGIA

Abstracts:

Sofia Bonicalzi (University of Roma Tre) - Autonomy in Action: Understanding Agency in the Context of AI

The sense of agency, which is a subject of lively debate in philosophy and neuroscience, refers to the subjective experience of controlling one's actions and their effects on the external world. As such, the sense of agency is thought to represent a key element of the feeling of autonomy and responsibility in decision-making and action. Research indicates that the sense of agency is influenced by factors such as the fluency of the action selection process, the outcomes of those actions, and the presence of other individuals. Recently, several studies have explored how interacting with AI may impact the sense of agency and autonomy, particularly when devices perceived as having their own intentionality are involved. While some research suggests that these interactions can diminish the sense of agency, raising important ethical considerations about our experience as authors of our actions, there is also evidence that when computer assistance is limited and allows for human initiative, the sense of agency can be preserved. In this paper, I will emphasize the importance of designing AI-based systems that support the agent's internal locus of control. In particular, I will argue that although interactions with artificial devices can effectively reduce the sense of agency, with potentially nefarious implications for autonomy, balanced interactions that allow for human initiative can sustain or even increase our agentive capabilities, enabling control over the environment.

Simona Tiribelli (University of Macerata) - Relational Autonomy and Artificial Intelligence: A Missing Piece

Benchmark AI ethics frameworks acknowledge the principle of autonomy as necessary to mitigate the harms that might result from the use of AI within society. These harms often disproportionately affect the most marginalized and historically oppressed. In this talk, I argue that the principle of autonomy, as currently formalized in AI ethics, is itself flawed, as it expresses only a mainstream mainly liberal notion of autonomy as rational self-determination and control, derived from Western traditional philosophy. In particular, I show that the adherence to such principle, as currently formalized, does not only fail to address many ways in which people's autonomy can be violated, but also to grasp a broader range of AI-empowered harms profoundly tied to the legacy of colonization, and which particularly affect the already marginalized and most vulnerable on a global scale. To counter such a phenomenon, I propose a framework for understanding autonomy based on a relational rethinking of the AI ethics principle of autonomy, drawing on theories on relational autonomy developed both in Western and non-Western ethical theories, and show how it helps us to operationalize by design a sounder, more inclusive, and globally sensitive ethical account of autonomy in transnationally deployed AI.

Radu Uszkai (University of Bucharest) - Show, Don't tell! Designing AI Tools for Autonomy and Moral Enhancement

As AI tools are offering more and more guidance and assistance, one common concern that some have raised against them is that an overreliance on such tools could lead to various types of deskilling. For instance, relying too much on AI assistance for orientation could lead to losing our orientation abilities. While the deskilling argument is relevant even in such cases where an action has just instrumental value, it is much more important to look into those where it has an intrinsic one. In the past decade, a couple of philosophers have pushed for the use of AI assistants dubbed "Artificial Moral Advisors" (AMAs) for the purpose of offering moral guidance and moral enhancement. Exploring the epistemic underpinnings of the proposed designs of such AMAs, the purpose of my talk is to provide a critical assessment of AMAs that rely on moral testimony (i.e. those that tell users what to do) instead of moral inspiration (i.e. those that show users what to do).The crux of my argument revolves around the claim that, for AMAs to contribute to moral enhancement and avoid moral deskilling, they should cultivate moral autonomy and the development of crucial critical skills needed to think and act in moral contexts. To achieve such a goal, AI tools should be designed in such a way so as to inspire users to act morally, not as tools for outsourcing moral decision making to an alleged algorithmic authority

Fabio Fossa (Politecnico di Milano) - Human Autonomy and Driving Automation: the Case of Automated Route Planning

In both industrial and scientific discourse, automated vehicles are often presented as autonomy-enhancing technologies. In my talk I intend to provide a criticism of the hype and technological solutionism underpinning such viewpoints. I aim to show that the exercise of human autonomy in the practice of driving is multifaceted, so that driving automation is bound to mediate it in ambiguous ways. I first consider crash-optimization algorithms for risk distribution during unavoidable collisions, where the automation of moral judgment might constrain individual moral autonomy. The abstractness and remoteness of the example, however, might suggest that risks are not pressing. To argue that this is not the case, I consider automated