

Viola Schiaffonati

Politecnico di Milano

Preliminary steps for experimentally evaluating the impact of AI and robotic technologies

The fears about the future consequences of Artificial Intelligence are monopolizing the ethical debate, while the introduction of relatively untested AI systems into social institutions is carried out without a rigorous analysis about the impact of these technologies on their social, cultural and political settings. Recently, an interesting social-systems analysis has been proposed (Crawford and Calo 2016) to overcome the limitations of existing approaches, such as compliance, values in design and thought experiments. Moreover, the conceptualization of new technologies (including AI and robotics) as social experiments, stressing their experimental character and the role of uncertainty together with the need of learning (van de Poel 2016), contributes to shift the focus of attention to the problem of rigorously evaluating the introduction of these technologies into society. Along this line we have recently proposed the notion of *explorative experiment* as a form of investigation carried out in the absence of a proper theory or theoretical background (Amigoni and Schiaffonati 2016), by investigating experimentation in autonomous robotics. In this talk we plan to expand further along this line, and in particular in the direction of framing explorative experiments in AI and robotics as a special case of social experimentation, where a particular form of experimental control (*a posteriori control*) is carried out. This is the case in which the experimenter is not in full control of the experimental setting due to the impossibility of anticipating plausible outcomes for the lack of a proper theory or theoretical background that make impossible to clearly state hypotheses. In this situation, establishing a priori dependent and independent variables is problematic so that to precisely track what aspects will be controlled during the experiment becomes impossible. We claim that this reframing is beneficial for working on a different notion of experimentation with new technologies in which a trial-and-error learning takes place about an on-going intervention thought the experimental introduction of these technologies into society. We believe this is a first step to ask questions about the impact of AI within a conceptual framework that takes into account not only the ethical consequences of these technologies, but the radical epistemological shifts that strongly impact on these consequences themselves.

[Joint work with Francesco Amigoni]