

ACAI 2021

The advanced course on AI on Human Centered AI

Tutorial Outline

Ethics and AI: An Interdisciplinary Approach

Guido Boella, Università di Torino; Maurizio Mori, Università di Torino

This half day tutorial on Ethics and AI intends to provide a brief overview of the main ethical issues concerning AI from an interdisciplinary perspective, involving a leading professor of Bioethics, Prof. Maurizio Mori and a professor of AI, Prof. Guido Boella, with a background in philosophy.

The starting point of Prof. Mori will be the clarification of what an “Ethics” is, so as to offer the reference coordinates of the discourse, which is then articulated in the analyzes by way of examples of specific issues in the different directions, with an interacting approach promoting debate with the students. In particular, Prof. Mori will present what is Ethics, as a coordination instrument among people, and how Bioethics from the 70s changed Ethics as the result of the introduction of new technologies: the technical possibility of abortion and of postponing end of life in terminal patients thanks to machines lead to the revision of the so called Ethics of absolutes, both religious ones and layman ones such as the Kantian Ethics, introducing notions such as prima facie ethical conclusions.

Guido Boella will consider how many problems such as bias are not only technical problems, how it originates from the lack of curation of datasets used by Machine Learning algorithms, how these algorithms risk to reflect the status quo ante of reality and create dangerous feedback loops, as explained in the Weapons of Math Destruction of Cathy O’Neil. He will consider the case of the firing of Timnit Gebru and Margareth Mitchell from Google due to the denunciation they made in On the Dangers of Stochastic Parrots paper. He will consider the issue of explainability starting from the analysis of Everyday Chaos of David Weinberger looking at what are explanations and challenging the possibility of explainable AI. He will describe the limitation of emotion recognition approaches due to their shaky scientific foundations in the work of Paul Ekman, as discussed in the Atlas of AI of Kate Crawford.

The work of Kate Crawford will also be the basis for passing to the discussion of the impact of AI in the society and how looking at broader economic and political structures and social systems, we can escape the notion that artificial intelligence is a purely technical domain. The books of Tim Wu will be the basis for trying to understand what the history of technology in the past century has to teach us about the market of attention and monopolistic tendencies.

He will discuss the predictions of Emil Durkheim and Hannah Arendt about the repeating cycle of accumulation under the light of the Shoshana Zuboff’s The Age of Surveillance Capitalism.

Finally, the impact of AI on ethics itself will be examined, with a warning that Ethics as we know it will need to be restructured. First he will show how the approach of applying the four

principles of Beauchamp and Childress (autonomy, non-maleficence, beneficence and justice) from the Bioethics field to the Ethics of AI is not feasible, then he will discuss how the impact of AI on the construction of social reality will lead these ethical principles as we know them to become circular.

Slides of Guido Boella:

<https://www.dropbox.com/s/ag06twmch6o7ts0/Ethics%20and%20AI%20tutorial.pdf?dl=0>

Similar lessons from the PhD course on Ethics and AI organized by SIPEIA:

Maurizio Mori <https://youtu.be/dYSSzYhgNBg>
<https://youtu.be/tvGcD5Xt3a0>

Guido Boella <https://youtu.be/adNfsZFousQ>

Bibliography:

Principles of Biomedical Ethics - Tom L. Beauchamp, James F. Childress

The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power

Shoshana Zuboff (QUESTO LIBRO CONTA PER DUE)

The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence
Kate Crawford

Weapons of Math Destruction
Cathy O'Neil

Everyday Chaos Technology, Complexity, and How We're Thriving in a New World of Possibility
David Weinberger

21 Lessons for the 21st Century
Yuval_Noah_Harari

The Master Switch: The Rise and Fall of Information Empires
Tim Wu

The Attention Merchants: How Our Time and Attention Are Gathered and Sold
Tim Wu

The Curse of Bigness - Antitrust in the new Gilded Age
Tim Wu

Who Controls the Internet? Illusions of a Borderless World
Jack Goldsmith e Tim Wu

Zucked waking up to the facebook
Roger McNamee

Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor
Virginia Eubanks

The Culture of AI: Everyday Life and the Digital Revolution.
Anthony Elliott