Syllabus – Law, Ethics and AI – 2023/24

Prof. Federico Faroldi

Prerequisites

The student should have basic knowledge of the main tools and techniques used in AI, and an introductory knowledge of formal methods.

Objectives

The course aims at introducing and discussing some of the main current problems and approaches to the ethics and law of artificial intelligence, including, for example, problems of definition of AI techniques in legal texts, actual and projected uses of AI in the civil and criminal domain, the proposed EU AI regulation, the control and alignment problems, normative uncertainty and normative risk, and the human compatible approach. One part of course will be devoted to some of the main ethical issues raised by artificial intelligence, among which are the problem of the incorporation of biases by artificial intelligence, and the questions of the moral status and moral responsibility of AI.

The expected results are the following:

An in-depth understanding of the main claims made by each of the theories we consider.
The ability to identify the structure of arguments and theories.
The ability to present focused objections to arguments and theories.
The ability to rationally defend a point of view, possibly original, and to communicate effectively.

Methods

Lectures. Discussion sessions. Seminars. Guided readings of research papers. Talks by invited experts.

References

For students who attend at least 75% of the classes:

Slides of the lectures and selected papers.

For all other students:

For students who attend at least 75% of the classes:

Slides of the lectures and selected papers.

For all other students:

Julia Driver, Ethics: The Fundamentals (2006), only chapters 1, 2 (pp. 31–39), 3, 4, 5, 7, 8, 10.

Howell, R.J. 2014. Google Morals, Virtue, and the Asymmetry of Deference. Noûs.
48(3), pp. 389-415.

Stuart Russell, ``Human-Compatible Artificial Intelligence." In Stephen Muggleton and Nick Chater (eds.), *Human-Like Machine Intelligence*, Oxford University Press, 2021

Federico Faroldi, Lecture Notes on Law, Ethics, and AI (available at the end of the course).

Richard Ngo, AI Safety from First Principles

John-Stewart Gordon and Sven Nyholm, "Ethics of Artificial Intelligence", Internet Encyclopaedia of Philosophy, https://iep.utm.edu/ethics-of-artificial-intelligence/

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

Johnson, D.G. (2006). Computer systems: Moral entities but not moral agents. Ethics Inf Technol 8, 195-204;

Gunkel DJ. *The machine question: critical perspectives on AI, robots, and ethics*. Cambridge: The MIT Press; 2017 (chap. I).

Redaelli, R. (2023). Different approaches to the moral status of AI: a comparative analysis of paradigmatic trends in Science and Technology Studies. *Discov Artif Intell* 3, 25 (2023).

van de Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds & Machines* 30, 385-409;

Exam

Multiple-choice written test. Sample questions will be discussed during the course.
The test will include questions from all modules, and the vote will be unique.