

Course proposal

AI Ethics vs. AI Justice: Different approaches for assessing bias in machine learning

Teacher

Antonio Carnevale (<https://www.uniba.it/it/docenti/antonio-carnevale>)

Course description

Recent studies reveal a convergence in the ethical guidelines of AI, emphasizing the emergence of 'fundamental principles' for responsible AI. However, dissenting voices argue that these principles are insufficient to address the social impacts of AI, revealing a disconnect between ideals and implementation. This explains because new accounts are emerging more oriented to "AI Justice" instead of "AI Ethics".

Throughout the course, we will engage with the intricate process of identifying and cataloging discriminatory biases that can arise during the lifecycle of ML systems. We will analyze various forms of causal reasoning related to discrimination, including technical, counterfactual, and constructivist/genealogical perspectives. This exploration will help us understand the multifaceted nature of AI biases and the limitations of viewing them through a deterministic lens.

By the end of this course, you will have a deep understanding of the ethical considerations surrounding AI and be equipped to critically assess the social impacts of ML-driven technologies. You will be prepared to contribute to the development of more inclusive and responsible AI systems, ensuring that technological advancements benefit all members of society.

Course period

January 2025-March 2025

SSD

INF/01, M-FIL/03, M-FIL/06

Credits and Hours

3 credits, two of lectures (16 hours) and one of practice (15 hour), for a total of 31 hours.

- **First Module** (16 hours of lectures): AI Ethics vs. AI Justice: Defining the differences.
- **Second Module** (15 hours of practice): Workshops on case studies (LLM, ChatGPT, recommendation systems, robotics, exoskeletons, VR etc.)


Exam Modality


- 1) Students present and discuss the main topics covered in the course, also referring to case studies.

- 2) Alternatively, Paper presentation. Students present the content of a number of papers suggested by the teacher. No groups are allowed.

Teacher CV

Personal Info **Antonio Carnevale**

 DIRIUM Department, Palazzo Ateneo, Piazza Umberto I – 70121, Bari
University of Bari

 +39 3475537200

 antonio.carnevale@uniba.it / antonio.carnevale@dexai.eu

 [LINKEDIN](#)

 [ORCID](#)

Gender M | Date and place of birth 04/01/1975, Flörsheim am Main (D) | Citizenship Italy

Languages: Italian (native), English (good), German (intermediate), Spanish and Portuguese (basic)

Areas of specialization:

- My most recent fields of scientific specialization: Ethics and Philosophy of Technologies; Research ethics; Energy & Sustainability; Ethics & AI; Ethics & robotics; Robotics & healthcare; Critical studies of technology; STS; Technology and gender studies.
- Among my research topics, I would also mention German Idealism, Hegel, critical theory, theory of recognition, emotion theory, shame, and vulnerability.

Current position:

- Researcher in Moral Philosophy at University of Bari, Italy (FAIR PNRR project – Future Artificial Intelligence Research, an extended partnership promoted by the National Recovery and Resilience Plan)
- Ethicist, Consultant in Philosophy, Ethics and Emerging Technology
- Co-founder of [DEXAI – Etica Artificiale](#) Srls

Key professional experiences

- HORIZON-MSCA-2023 expert evaluator
- Co-author of a study funded by the Panel for the Future of Science and Technology (STOA) of the EU Parliament
- Ethics advisor of AIMPC (Eurostars), BEEYONDERS (Horizon Europe)
- Member of the EABs of PALANTIR; TANGO, TRUMPET, RELEVIMUM
- Ethics partner in several EU projects

ACADEMIC EXPERIENCE

April 2022 – **Postdoc Fellow in Environmental Psychology**

March 2024 Postdoc fellow at the **Norwegian University of Science and Technology** NTNU in Trondheim, Norway, on the H2020 research project ACCTING

- June 2016 – **Research Fellow in Ethics**
- December 2016 Research Fellow in Philosophy within the project "Reliable – Regulating Ethical and Legal Issues of Advanced Bionic Limbs and Exoskeletons". Scientific manager: Dr. Andrea Bertolini, **Sant'Anna School of Advanced Studies**, Pisa (Italy).
- March 2012 – **Research Fellow in Ethics and Philosophy**
- March 2016 Position within the European project "RoboLaw: Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics", funded by FTP7 (Theme: Regulating emerging scientific and technological developments). Faculty of Social Sciences, University **Sant'Anna School of Advanced Studies**, Pisa (Italy).
- October. 2011 – **Research Fellow in Political Philosophy**
- March 2012 Position at Faculty of Social Sciences, University **Sant'Anna School of Advanced Studies**, Pisa (Italy).
- September 2009 – **Postdoctoral Fellow in Political Philosophy**
- August 2011 Position at Faculty of Social Sciences, University **Sant'Anna School of Advanced Studies**, Pisa (Italy).

LAST SCIENTIFIC ACTIVITIES, COMMENTARIES

-
- 2021
- (January – July): Scientific co-director of "Corso di perfezionamento in Etica e Informatica". University of Tor Vergata, Rome.
 - (September): Co-organizer of Panel 10: "La differenza polarizzata e il rovesciamento del tutto nelle parti: Ripensare l'universale politico nell'era della post-verità?", Festival *La filosofia, il castello, la torre*, VII edizione, Ischia – Napoli.
 - (October): Invited lecturer for "Artificial Intelligence: Law and Ethics of Emerging Technologies" of the European University of Rome.
- 2020
- (June): speaker at DECON 2019: International Conference on Decision Economics, Avila, Spain, June 26-28. Title of paper (with C. Occhipinti): Ethics and Decisions in Distributed Technologies: A Problem of Trust and Governance Advocating Substantive Democracy.
- 2019

SERVICE TO PROFESSIONAL ORGANIZATIONS

-
- 2004 – today
- Ethics advisor of AIMPC (Eurostars)
 - Member of the External Advisory Board (EAB) of EU projects [Palantir](#); TANGO, TRUMPET, RELEVIVUM (Horizon 2020 / HORIZON Europe);
 - Ethics advisor for the EU projects: [Persist](#), [Impulse](#), [BEEYONDERS](#)
 - Member of Scientific Board of the *European Society of Moral Philosophy* (ESMP)
 - Peer Reviewer for *Elsevier journals*
 - Peer Reviewer for *Ethical Theory and Moral Practice*
 - Founder and member of the Editorial Board for the book series *Politica, Critica, Società* (*Politics, Critique, Society - PCS*) published by Pensa Multimedia, Lecce (Italy)
 - Peer Reviewer for *Journal of Classical Sociology*
 - Member of the Scientific Board of *Revue de la Société Internationale d'Ethnographie*
 - Member of the *International Research Laboratory on Conflict, Development and Global*

Politics, at Sant'Anna School of Advanced Studies, Pisa

- Reviewer for *Jura Gentium – Journal of Philosophy of International Law and Global Politics*
- Member of the *Italian Society of Philosophy (SFI)*

RECENT PUBLICATIONS

1. Delgado, C. F., Ferretti, M. T., & Carnevale, A. (2024). Beyond one-size-fits-all: Precision medicine and novel technologies for sex- and gender-inclusive COVID-19 pandemic management. In S. Mellino, D. Motti, E. Santus, & A. S. Chadha (Eds.), *Innovating Health Against Future Pandemics* (pp. 133–156). Academic Press. <https://doi.org/10.1016/B978-0-443-13681-8.00012-6>.
2. Carnevale A, Falchi Del Gado C., Bisconti Lucidi P. (2023). “Hybrid Ethics for Generative AI: Some Philosophical Inquiries on GANs”. In: *_Humana.Mente_*, vol. 16, n. 44, pp. 33-56. <https://www.humanamente.eu/index.php/HM/article/view/434> [1]
3. Occhipinti, C., Carnevale, A., Briguglio, L., Iannone, A., Bisconti, P. (2023), “SAT: a methodology to assess the social acceptance of innovative AI-based technologies”. *_Journal of Information, Communication and Ethics in Society_*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/JICES-09-2021-0095>.
4. Luigi Briguglio, Carmela Occhipinti, Antonio Carnevale, Riccardo Santilli (2022) *Privacy and security aspects of the 5G technology*. Study funded by the Panel for the Future of Science and Technology (STOA) of EU Parliament.
5. Antonio Carnevale, Piercosma Bisconti, Federica Russo, Carmela Occhipinti (eds.) (2023). *Blockchain for Trustworthy Digital Identity Systems: Socio-Technical Perspectives*. Special issue of *Digital Society* (Springer).
6. Antonio Carnevale and Emanuela Tangari (eds.) (2021). *Ethics, Health Data, and Bio-Citizenship: From the Politics of Body to the Politics of Knowledge*. *Philosophical News*, no. 22.
7. Fiorella Battaglia and Antonio Carnevale (eds.) (2014). *Reframing the Debate on Human Enhancement*. Special issue of *Humana.Mente*, vol 7 no 26.
8. Occhipinti, C., Carnevale, A., Briguglio, L., Iannone, A. and Bisconti, P. (2022), SAT: a methodology to assess the social acceptance of innovative AI-based technologies. *Journal of Information, Communication and Ethics in Society*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/JICES-09-2021-0095>.
9. Arturo Cuesta; Daniel Alvear; Antonio Carnevale; and Francine Amon (2022). Gender and Public Perception of Disasters: A Multiple Hazards Exploratory Study of EU Citizens. In: *Safety*. 2022; vol 8, no 3, issue 59. DOI: <https://doi.org/10.3390/safety8030059>
10. Bisconti P. and Carnevale A. (2022). “Alienation and Recognition – The Δ Phenomenology of Human-Social Robot Interactions”. In: *Techné: Research in Philosophy and Technology*, vol. 29, no.1, pp 147-171. DOI: <https://doi.org/10.5840/techné202259157>