University of Lisbon

School of Arts and Humanities - Department of Philosophy

2025/2026 (S2)

Course: 920479 – *Ética*

Teachers: Felipe León (Part I) and Mariana Teixeira (Part II)

E-mails: felipe.leon@edu.ulisboa.pt; mariana.o.teixeira@edu.ulisboa.pt

Course description (Part I):

The first half of the course will be devoted to the ethics of Artificial Intelligence (AI). Smartphones, chatbots (such as ChatGPT), self-driving cars, autonomous weapons, and social robots are some examples of an increasing number of AI artifacts that have become widespread in the contemporary world. The use of these and other AI systems raises pressing ethical and philosophical questions: Is it plausible to assume that some AI systems are or will ever become conscious? Do we have special responsibilities towards AI systems? Can some of them, in turn, be held responsible? Are decision-making AI systems by default more neutral than human beings, or can those systems also be discriminatory and biased? Can we trust some AI systems, and perhaps even fall in love with them? We will cover these and other questions by focusing on six main topics in connection with the ethics of AI: (1) Consciousness, (2) Responsibility, (3) Bias, (4) Trust, (5) Grief, and (6) Love.

Programme:

Session	Date	Main topic	Assigned readings
1	10/2	Introduction	Brooks (2017), Donath (2020)
2	12/02		
3	13/02		
4	19/02	Consciousness	Birch (2025), León & Zahavi (2023)
5	20/02		
6	24/02	Responsibility	Noorman (2023), Himmelreich & Köhler (2022)
7	26/02		
8	27.02		
9	3/3	Bias	Fazelpour & Danks (2021), Pozzi (2023)
10	5/3		
11	6/3		
12	10/3	Trust	Coeckelberg (2011), Lahno (2020), Reinhardt (2023)
13	12/3		
14	13/3		

15	17/3	Grief	Krueger & Osler (2022), Fabry & Alfano (2024)
16	19/3		
17	**20/3**	**First in-person exam**	
18	24/3	Love	Lopez-Cantero (2025), Kind (2021), Klonschinski & Kühler (2021)
19	26/3		
20	27/3		
...	...	...	...

Assessment:

Two in-person exams: 50% (each one 25%), one final essay: 45%, and class participation: 5%. The final essay should be on a topic previously approved by one of the teachers. More information about the requirements for the final essay, as well as guidance for how to write it, will follow. *Important dates*: the first in-person exam will be on **March 20**, the second on **April 30**. Students are requested to send an essay proposal to one of the teachers by **May 19**. Deadline for submitting the final essay: **June 3**.

Mandatory course bibliography for Part I:

Birch, J. (2025). "AI consciousness: A centrist manifesto". Manuscript available at:

https://philpapers.org/archive/BIRACA-4.pdf

Brooks, R. (2017). “The Seven Deadly Sins of AI Prediction”. *MIT Technology Review*. Available at: https://www.technologyreview.com/2017/10/06/241837/the-seven-deadly-sins-of-ai-predictions/

Coeckelbergh, M. (2012). “Can we trust robots?”. *Ethics and information technology*, *14*(1), 53-60. https://link.springer.com/article/10.1007/s10676-011-9279-1

Donath, J. (2020). “Ethical Issues in Our Relationship with Artificial Entities”. In Dubber, M. D., Pasquale, F. and Das, S. (eds), *The Oxford Handbook of Ethics of AI* (pp. 53-73). New York: OUP.

Fabry, R.E. & Alfano, M. (2024) “The Affective Scaffolding of Grief in the Digital Age: The Case of Deathbots”. *Topoi* 43, 757–769. https://doi.org/10.1007/s11245-023-09995-2

Fazelpour, S., & Danks, D. (2021). “Algorithmic bias: Senses, sources, solutions”. *Philosophy Compass*, 16(8), e12760, pp. 1-16. https://compass.onlinelibrary.wiley.com/doi/full/10.1111/phc3.12760

Himmelreich J. & Köhler, S. (2022) “Responsible AI Through Conceptual Engineering”. *Philosophy & Technology* (2022) 35: 60. https://link.springer.com/article/10.1007/s13347-022-00542-2

Kind, A. (2021). “Love in the Time of AI”. In Dainton, B., Slocombe, W., and Tanyi, A., (eds.) *Minding the Future: Artificial Intelligence, Philosophical Visions and Science Fiction* (pp. 89-106). Cham: Springer International Publishing.

Klonschinski, A., & Kühler, M. (2021). “Romantic Love Between Humans and AIs: A Feminist Ethical Critique”. In Cushing, S. (ed.) *New Philosophical Essays on Love and Loving* (pp. 269-292). Cham: Springer International Publishing.

Krueger, J. & Osler, L. (2022). “Communing with the Dead Online: Chatbots, Grief, and Continuing Bonds”. *Journal of Consciousness Studies 29 (9-1l0):222-252.* https://philarchive.org/archive/KRUCWT

Lahno, B. (2020). “Trust and emotion”. In Simon, J. (ed.) *The Routledge handbook of trust and philosophy* (pp. 147-159). New York: Routledge.

León, F., & Zahavi, D. (2023). “Consciousness, philosophy, and neuroscience”. *Acta Neurochirurgica*, *165*(4), 833–839. https://doi.org/10.1007/s00701-022-05179-w

Lopez-Cantero, P. (2025). “The ethics of break-up chatbots”. *Phenomenology and the Cognitive Sciences*, 1-20. https://doi.org/10.1007/s11097-025-10120-4

Noorman, M. (2023). “Computing and Moral Responsibility”, In Zalta, E. N., Nodelman, U. (eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition). Available at: https://plato.stanford.edu/entries/computing-responsibility/

Pozzi, G. (2023). Automated opioid risk scores: a case for machine learning-induced epistemic injustice in healthcare. *Ethics and Information Technology* 25, 3. https://doi.org/10.1007/s10676-023-09676-z

Reinhardt, K. (2023). Trust and trustworthiness in AI ethics. *AI and Ethics*, *3*(3), 735-744. https://link.springer.com/article/10.1007/s43681-022-00200-5