

# Artificial Intelligence, Ethics, and Society

a course outline (C. Eliot, 15 July 2025)

Each main topic will normally have a theoretical text and an example case. At this stage, the listed readings are preliminary.

## 1 Artificial “intelligence”?

*A brief history of AI; the main varieties of AI*

*What does “intelligence” mean? In what sense are AIs intelligent?*

## 2 Introduction to value theory

*Major approaches to reasoning about rightness and wrongness, goodness and badness*

## 3 Data integration, surveillance, and privacy

Creel, Kathleen, and Tara Dixit. 2022. “Privacy and Paternalism: The Ethics of Student Data Collection,” <https://doi.org/10.21428/2c646de5.b725319a>.

Macnish, Kevin. 2012. “Unblinking Eyes: The Ethics of Automating Surveillance.” *Ethics and Information Technology* 14 (2): 151–67. <https://doi.org/10.1007/s10676-012-9291-0>.

Mittelstadt, Brent Daniel, and Luciano Floridi. 2016. “The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts.” *The Ethics of Biomedical Big Data*, 445–80. <https://doi.org/10.1007/978-3-319-33525-4>.

## 4 Opacity and responsibility

Creel, Kathleen A. 2020. “Transparency in Complex Computational Systems.” *Philosophy of Science* 87 (4): 568–89. <https://doi.org/10.1086/709729>.

Vaassen, Bram. 2022. “AI, Opacity, and Personal Autonomy.” *Philosophy and Technology* 35 (4): 1–20. <https://doi.org/10.1007/s13347-022-00577-5>.

## 5 Bias in decision systems

Creel, Kathleen, and Deborah Hellman. 2022. “The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems.” *Canadian Journal of Philosophy* 52 (1): 26–43. <https://doi.org/10.1145/3442188.3445942>.

Dressel, Julia, and Hany Farid. 2018. “The Accuracy, Fairness, and Limits of Predicting Recidivism.” *Science Advances* 4 (1): eaao5580. <https://doi.org/10.1126/sciadv.aao5580>.

## **6 Behavior manipulation**

- Ienca, Marcello. 2023. “On Artificial Intelligence and Manipulation.” *Topoi* 42 (3): 833–42. <https://doi.org/10.1007/s11245-023-09940-3>.
- Tarsney, Christian. 2025. “Deception and Manipulation in Generative AI.” *Philosophical Studies*, <https://doi.org/10.1007/s11098-024-02259-8>.

## **7 Autonomous systems, autonomous weapons**

- Müller, Vincent C. 2016a. “Autonomous Killer Robots Are Probably Good News.” In *Drones and Responsibility*. Routledge.
- Müller, Vincent C. 2016b. *Risks of Artificial Intelligence*. Vol. 5. CRC Press Boca Raton, FL.

## **8 Existential risks of super-intelligences**

- Chalmers, David J. 2016. “The Singularity: A Philosophical Analysis.” *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 171–224. <https://doi.org/10.1002/9781118922590.ch16>.
- Sparrow, Robert. 2024. “Friendly AI Will Still Be Our Master. Or, Why We Should Not Want to Be the Pets of Super-Intelligent Computers.” *AI and Society* 39 (5): 2439–44. <https://doi.org/10.1007/s00146-023-01698-x>.

## **9 Sentience and the criteria for moral status of non-human agents**

- Gunkel, David J. 2018. “The Other Question: Can and Should Robots Have Rights?” *Ethics and Information Technology* 20 (2): 87–99. <https://doi.org/10.3233/978-1-61499-480-0-13>.

## **10 Intellectual property**

- Cunningham, Joshua. 2025. “Painting in Gray: The Legal and Ethical Ambiguities of AI-Generated Art.” *Journal of Information, Communication and Ethics in Society* 23 (3): 384–91. <https://doi.org/10.1108/jices-01-2025-0011>.
- Nawar, Tamer. 2024. “Generative Artificial Intelligence and Authorship Gaps.” *American Philosophical Quarterly* 61 (4): 355–67. <https://doi.org/10.5406/21521123.61.4.05>.

## **11 Labor, employment, and agency**

- Waelen, Rosalie A. 2025. “The Desirability of Automizing Labor: An Overview.” *Philosophy Compass* 20 (1–2): e70023. <https://doi.org/10.1111/phc3.70023>.

## **12 Slop, fakes, deepfakes, and trust**

Rini, Regina. 2020. "Deepfakes and the Epistemic Backstop." *Philosophers' Imprint* 20 (24): 1–16. <https://doi.org/10.26556/jesp.v22i2.1628>.

Sahebi, Siavosh, and Paul Formosa. 2025. "The AI-Mediated Communication Dilemma: Epistemic Trust, Social Media, and the Challenge of Generative Artificial Intelligence." *Synthese* 205 (3): 1–24. <https://doi.org/10.1007/s11229-025-04963-2>.

## **13 Impact of AI on human relations and on interpersonal ethics**

Sahebi, Siavosh, and Paul Formosa. 2024. "Artificial Intelligence (AI) and Global Justice." *Minds and Machines* 35 (1): 1–29. <https://doi.org/10.1007/s11023-024-09708-7>.

Vallor, Shannon. 2024. *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. Oxford University Press.

## **14 Impacts of AI on our environment**

Moyano-Fernández, Cristian, and Jon Rueda. 2024. "AI, Sustainability, and Environmental Ethics." In *Ethics of Artificial Intelligence*. Springer.