Reading List

The following are the topics, questions, and sources of the readings linked from the course's Canvas site. Many of the reading assignments are brief, targeted excerpts from articles for which the full citations follow. Many of the readings are available by direct link through Hofstra Library (requiring a Hofstra login) or on the open web. For the remainder, the files are posted to Canvas.

Week 1: What do we mean by "artificial intelligence"?

- A brief history of AI; the main varieties of AI
- What does "intelligence" mean? In what sense are AIs intelligent (and not)?

Bringsjord, Selmer, and Naveen Sundar Govindarajulu. 2024. "Artificial Intelligence." In *The Stanford Encyclopedia of Philosophy,* Fall 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/artificial-intelligence/.

Week 2: Introduction to ethics and value theory

- Major approaches to reasoning about rightness and wrongness, goodness and badness: (a) consequentialist approaches, (b) right-action approaches, (c) character approaches, (d) justice and fairness approaches

Blackburn, Simon. 2001. *Being Good: An Introduction to Ethics.* Oxford University Press.

Week 3: Privacy and big data sets

- Ways of thinking about the value of privacy
- Case: high school student data collection
- Case: automated surveillance

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.

Week 4: Responsibility when AI reasoning is hidden

- Ways of thinking about responsibility

- What aspects of AI activity are unobservable, even to the creators?
- How do we think about responsibility when unobservable processes yield objectionable results?

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.

Week 5: Self-controlled systems

- Case: Are self-driving cars safer or a menace?
- Case: Do armed-robot battles save soldiers' lives?
- Case: AI out of control

Müller, Vincent C. 2016a. "Autonomous Killer Robots Are Probably Good News." In *Drones and Responsibility.* Routledge. https://doi.org/10.4324/9781315578187.

Müller, Vincent C, ed. 2016b. *Risks of Artificial Intelligence.* Vol. 5. CRC Press. https://doi.org/10.1201/b19187.

Week 6: Automated decision-making, fairness and bias

- Ways of thinking about fairness and avoiding bias
- How can AI systems introduce biases?
- Case: automated student admissions
- Case: bail hearings and automated assessment of recidivism probability

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.

Week 7: AIs manipulating and deceiving people

- Ways of thinking about the problems with manipulation and deception
- Case: AI systems evading our cognitive defenses
- Measures for avoiding AI deception and 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.

Week 8: Should we create or ban super-intelligences?

- What are the risks and benefits of independent, super-human intelligences?
- Examples of technology and research bans and their histories

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.

Week 9: What would give an AI rights?

- What are the criteria for being morally considerable?
- Ways of thinking about rights
- What would a robot or AI need to do to earn rights in our society?

Gunkel, David J. 2018. "The Other Question: Can and Should Robots Have Rights?" *Ethics and Information Technology* 20 (2): 87–99. https://doi.org/10.1007/s10676-017-9442-4.

Week 10: How does AI challenge authorship and intellectual property?

- Approaches to creators' rights in ethics and law
- Case: training large language models on text and images
- Case: original works produced by generative AI

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.

Week 11: How might AI change labor and employment?

- Arguments about the possibility of a right to work
- Arguments for automating work as the key to the future
- Case: industries being replaced by automation

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

Week 12: How AI changes communication: slop, fakes, deepfakes, and trust

- Ways of thinking about the moral significance of trust in communication
- Ways of thinking about freedom of speech and expression
- Case: news slop
- Case: political deepfakes

Rini, Regina. 2020. "Deepfakes and the Epistemic Backstop." *Philosophers' Imprint* 20 (24): 1–16. https://hdl.handle.net/2027/spo.3521354.0020.024.

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.

Week 13: Should we work to preserve our humanity in a world with AI?

- Ways of thinking about justice, agency, and dignity
- What are possible challenges to global justice from AI?
- Does AI offer ways to rethink our human potential?

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.

Week 14: AI and our environment

- Why would we care about our environment?
- Kinds of environmental risks of AI activity
- AI in the service of environmental preservation
- Visions of sustainable AI

Moyano-Fernández, Cristian, and Jon Rueda. 2023. "AI, Sustainability, and Environmental Ethics." In *Ethics of Artificial Intelligence,* edited by Franisco Lara, 219–36. Springer. ISBN: 978-3-031-48135-2. https://doi.org/10.1007/978-3-031-48135-2_11.