Artificial Intelligence in Society:
New Challenges, Ethical Futures

PHI 040, Spring 0000

Professor:
Office:
Office hours:
Email:
Texts: readings will linked from, or distributed through, a Canvas reading list
Course time & place:

What we're doing here

Emerging artificial intelligence technologies create new challenges for traditional ideas about human ethics. "Ethics," here, means the study of what to do—including what is good, bad, right, and wrong—and why we should do one thing rather than another. By exploring how artificial intelligence creates specific challenges you will learn how to think about possible solutions. We will do that by bringing traditional ways of thinking about ideas like right action, goodness, personal character, rights and responsibilities, and fairness and justice to confront a world being changed by artificial intelligence. We will observe how these ideas interact with artificial intelligence in specific cases and practice reasoning logically about what to do with AI, using those concepts. We will explore where our society might need new ethical ideas, and we will think about how to build a promising future society we will want to live in, that includes AI.

Philosophical Ethics, which offers an intellectual backbone for this course, is a core area of investigation in English-speaking ("anglophone") Philosophy. One way of understanding what "Philosophy" is as analysis of questions that cannot be answered through observation alone ("empirically"). Because answers to philosophical questions cannot be proved by observation alone, philosophers use and analyze logically-structured *arguments* to gain understanding. In this course we will be examining important and influential arguments about artificial intelligence and also developing some arguments of our own. We will try to understand a variety of possible positions we could come to adopt for ourselves and some of their strengths and weaknesses.

Important dates

First quiz: Sep 29
First test: Oct 15
Second quiz: Nov 12
Second test: Dec 15 1:30–3:30

Course requirements

1.   You will need to **check Canvas** for reading assignments, writing assignments, and other essential information. It serves as an extension of this syllabus and our dynamic schedule.

2.   You will need to **access the texts** which will be linked from or uploaded to Canvas and figure out your strategy for reading them: print, tablet, screen?

3.   You need to **bring the texts** to class (or in some way have them readily available to you in class).

4.   You will be expected to **read** and **be prepared to discuss** the assigned texts. You will receive a participation grade which figures in both your contributions to class and the apparent preparedness reflected in them. Participation is not mere attendance; it involves volunteering contributions. I may even call on you and expect you to have something relevant to say. Though presentations will help explain the readings, you cannot expect

to understand our discussion without reading yourself. Much of the point of our meeting synchronously is that you bring you be able to engage *your independent scholarship* with that of other scholars—your student peers and me. Obviously, that requires having both read and digested the text. Do it.

5. You are expected to **attend** the class sessions. I will record attendance. Two absences will be automatically excused. Third through sixth absences, if unexcused, drop the attendance score, which starts at 4, by one point (which is also 1%). Therefore, at 6 total absences, with third–sixth unexcused, the attendance score reaches 0. More absences than those six will diminish your ability to participate sufficiently, and will necessarily therefore also be reflected in that grade. You *do not* need to contact me about being absent. However, if you have doctors' notes or excuses from Hofstra officials, keep them on file until the end of the semester, in case you need to show why you needed to be absent more than twice.

6. You will need to **participate** in our class discussion by asking and responding to questions. See the Participation Rubric on Canvas to understand how participation will be assessed.

7. **Use electronic devices only when permitted**, during some portions of class sessions, typically during presentations. During other portions of class sessions, you will be asked to put away all devices out of sight.

8. You will **submit four short writing assignments** through Gradescope. Instructions for accessing Gradescope are posted to Canvas.

9. You will take two **quizzes**, handwritten in class. You will need to bring a writing implement. Each of the quizzes will involve writing in paragraph form similar to what you will have done in your out of class assignments.

10. You will need to attend and **take a first/midterm test** and a **second test/final exam** at the end of the term. The second test will be comprehensive, but will emphasize the second half of the course. More information about the test content will be posted to Canvas. For each test, I will circulate a list of questions in advance, from which the actual test will be drawn. There will therefore be no questions on the tests which you have not seen in advance. The second test will cover the entire course but will strongly emphasize the second half. No bathroom breaks will be permitted during tests. Make-up tests will not be permitted except when you are formally excused. Make travel plans accordingly.

11. Since you will be responsible on tests for knowing what happened in class even if you did not attend: **if absent, coordinate** with one or more other students in the class. I can tell you what topic we discussed in class, but I can't repeat the lecture or discussion or email you notes. I don't have notes that would make sense to you. Others might.

12. Finally, I always hope this goes without saying, but: I ask you to **respect the business of the class** and of other students' efforts to focus. Respect doesn't always mean solemn silence. See the participation rubric on the Canvas site for more details.

Evaluation

I would prefer to teach entirely without grades, but their existence has various kinds of value for you. So we have them. Here is where yours come from:

To the degree it is practically feasible, I evaluate your work anonymously, to eliminate unconscious biases and approach objectivity. (Obviously, under certain circumstances, it is not practically feasible, but the course's use of Gradescope makes anonymity possible on nearly all assignments.) Your grades will be calculated according to the University's standards, relative to course expectations, and relative to other members of the class.

Your grade will *not* be "curved" to a mean score, but it may be adjusted upwards depending on overall class performance. That is, you will earn at least the score you deserve according to University standards, but also one related to how other students performed.

Your grade will be calculated as follows:

Participation	20%
Attendance	4%
Out-of-class assignments	24%
First quiz	4%
Second quiz	4%
First test	20%
Second test	24%

I will circulate partial/preview grades for participation at the one-third and two-thirds marks, so you have a clear sense of how I think you're doing. The participation rubric is posted to Canvas. Grades for writing assignments, quizzes, and tests will be converted to Hofstra's 4-point scale before calculation, as summarized below. (The scale is described in the annual *Undergraduate Bulletin* section "Classes, Attendance and Grading Policies," under "Cumulative Grade Point Average." More details on grade definitions appear in the same *Bulletin* section under "Grades".)

4 = **A**   represents honors level work.
3 = **B**   represents performance distinctly above that required.
2 = **C**   represents achievement of the objectives of the course.
1 = **D**   represents performance sufficient for degree credit but not major or minor credit.

That entails that your final grade will be calculated as the sum of the multiples of each of these percentages by the associated four-point grade posted to Canvas for that item. The resulting 4-point grade will be converted back to a letter grade for filing, using Hofstra's 4-point scale. A final grade of "Incomplete" will not be assigned automatically, nor in the absence of a compelling, written request.

Academic integrity policy

Familiarize yourself with the university's Academic Integrity policy. Representing someone else's work as your own, or any other form of academic dishonesty (as defined by the University), will automatically earn you an F for the course, beyond the required report to the University. So this is the key point you should internalize now: If you ever find yourself in circumstances where you feel motivated to be dishonest, please come talk to me about what we can do about the circumstances instead. You will find that while *after* I've detected academic dishonesty, the outcome is severe and automatic, *beforehand* I try to be as helpful as possible.

Generative artificial intelligence policy

This course does not have a simple, across-the-board rule against the use of generative artificial intelligence (including ChatGPT and other large language models, among other software) in your coursework. The course presumes that, like spelling-correction software and search engines, generative AI is a widely used tool, and that as with all tools, students benefit from understanding their strengths and limitations.

Students should consider that here, just like in the broader world, using generative AI but not understanding its limitations can produce serious, negative consequences. Specifically, students using generative AI must be mindful of the following:

1. You are paying here, in part, for help with developing understanding and skills. If you use generative AI in a way that undermines your learning, you are not getting what you paid for and you are wasting your time.

2. When using assistive technologies, you remain responsible for everything you turn in. By turning something in, you assert authorship of it, no matter what tools you have used to write it.

3. This course's academic integrity policy does not stop applying just because you've used generative AI. Submitting false citations, for example, is an academic integrity violation. As of 2025, ChatGPT and other LLMs generate false citations *frequently*.

4. For the purposes of this course, including text generated by generative AI in work for which you assert authorship is not on its own an academic integrity violation. However, if the text you submit happens to use a source without appropriate attribution or required quotation marks, that *is* a clear academic integrity violation.

5. A significant portion of the graded work for this course will not be submitted electronically, but rather hand-written in class, and you will not be able to use generative AI to produce it. If, when you are taking those assessments, you haven't developed skills and understanding because you relied on generative AI earlier, the assessments will probably not go well for you. And our goal—mine and presumably yours—is for them to go well!

Syllabus adjustments

The schedule may dynamically update on Canvas. Beyond that, events can lead to changes in the schedule and syllabus. However, it's important to me that you not feel that the rules have changed on you mid-stream, so, if any changes are necessary, I will make them as fair as possible.

Contacting me

The best way to contact me is by email; I check it regularly. If you have a quick question, often we can take care of it by email. But if you need more help, don't hesitate to visit my office hours, listed above, or ask (in person or by email) to set up another mutually-convenient time, in person or online. Phoning is not a reliable way to reach me.

General university policies applying to this course

Information about Class Attendance; Recording in Classes; Inclement Weather; Academic Integrity; Student Health and Wellness, Disability-Related Accommodations; Resources for Students who are Pregnant; Temporary Accommodations; Academic Leave of Absence; Absences for Religious Observance; University Deadlines; Grading Policy; and Sex-Discrimination, Sexual Assault, Dating and Domestic Violence, and Stalking is available on the Provost's webpage at the following URL:
`https://www.hofstra.edu/provost/policies-wording-syllabi.html`.

Department's learning goals and objectives for this course

Goals for Philosophy courses:

▷ Students understand major philosophical ideas accurately
▷ Students apply their understanding of ideas in novel contexts
▷ Students write effectively
▷ Students speak effectively
▷ Students argue with precision, balance, and insight
▷ Students read analytically, critically, and empathetically
▷ Students critically assess their own commitments and ideas

Specific learning objectives for Philosophy courses:

▷ Objective 1a: Students give accurate and relevant answers, complete with supporting details, to specific questions about philosophical ideas relevant to the course.
▷ Objective 1b: Students give accurate accounts of philosophical ideas relevant to the course in the context of criticizing or assessing those ideas.
▷ Objective 2a: Students speculate, in well-informed, well-supported, and plausible fashion, about what a given philosopher would say about a novel issue or problem.

▷ Objective 2b: Students extrapolate creatively and plausibly from their knowledge of philosophers or philosophical positions in developing their own related ideas.

▷ Objective 3a: Students write paragraphs that exhibit clarity, focus, a good command of the subject matter, and an orderly development of ideas.

▷ Objective 3b: Students write multi-paragraph pieces that exhibit clarity, focus, a good command of the subject matter and an ability to work with that subject matter creatively, and an orderly development of ideas both within and across paragraphs.

▷ Objective 4a: Students speak in clear, focused, well-informed, and orderly fashion.

▷ Objective 5a: Students state arguments accurately and clearly, and identify strengths and weaknesses of different arguments.

▷ Objective 5b: Students develop and defend their own arguments, taking into account a variety of philosophical positions but adding original insights or emphases.

▷ Objective 7a: Students explain difficult passages clearly, accurately, and thoroughly.

▷ Objective 7b: Students use apt quotations and creative, critical, plausible readings of texts in their writing.

▷ Objective 8a: Students are able to explain the weaknesses of their own present positions, and the strengths of competing positions.

▷ Objective 8b: Students are able to explain why their pre-theoretical commitments have or have not changed as a result of what they have learned in the course, and if they have changed how they have done so.

Final thought

Inevitably, grades are a function of performance, not of effort in itself. I can't reasonably assess effort. What's challenging varies from student to student. You will need to figure out what *you* need to do to perform well. I will try to help you with what's hard for you, if I know you need help. In the end, you are responsible for your education, however, and if you are confused, you should ask a question, or I will assume you understand. Unless you discuss them with me in person or by email, I will also likely not be aware of any dissatisfactions you have with any aspect of the course. I hope you will not be dissatisfied. I think this material is fun and useful, and believe an important part of my job is trying to show you why it is.

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.

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/10.1002/9781118 922590.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.