

# PHIL. 058: AI & ETHICS

Fall 2025: M&W 4:15–5:30

Bernard 105

Kyle Scott

Email: [kyle\\_scott@pitzer.edu](mailto:kyle_scott@pitzer.edu)

Office Hrs: W 5:45–6:45pm and by appt., in Broad Hall 212 & via Zoom

## COURSE DESCRIPTION

This course introduces students to a rapidly growing branch of philosophy concerned with the ethical, social, and political challenges of artificial intelligence and data-driven technologies. AI is often described as creating both unprecedented opportunities and profound risks. It can enrich science and medicine, eliminate mundane tasks, improve technology, and even provide forms of interpersonal support. Yet it can also entrench social injustice, erode democracy through misinformation, violate privacy, exploit workers, and undermine human values and self-determination. For instance, many AI systems display discriminatory bias: what is the nature of this bias, and how might it be addressed? The effectiveness of AI often depends on collecting vast amounts of personal data: who should be allowed to collect it, and what rights do individuals have to privacy and anonymity? The opacity of algorithmic systems also raises questions of governance and accountability. What explanations are people owed for automated decisions that affect them?

Although there is an expanding interdisciplinary literature on these questions, this course will focus on issues of distinctively philosophical interest. In particular, we will consider:

- The values and biases embedded in technological systems.
- The opacity of algorithmic decision-making.
- Problems of surveillance and privacy.
- The erosion or replacement of distinctively “human” values, such as in relationships or creative activity.
- The impact of AI on democracy, misinformation, and work.

This is a philosophy course, specifically in value theory, which concerns the nature and application of social, political, and moral norms and values. Accordingly, our focus will be on understanding the ethical and social problems raised by AI: why they are problems, what values they implicate, and how they illuminate broader philosophical concerns. By contrast, we will not primarily be concerned with

the technical workings of AI systems (though some background will be needed), nor on specific policy or engineering solutions (important as those may be).

## COURSE OBJECTIVES

Beyond introducing students to core questions in the ethics of AI, this course is also designed to cultivate the fundamental skills of philosophical inquiry. Students will gain practice in analyzing and evaluating arguments, developing their own positions, and communicating ideas clearly and persuasively in both writing and discussion. By the end of the course, students should be able to:

- **Formulate and communicate arguments clearly** — producing cogent arguments and expressing them with clarity and precision in both oral and written form.
- **Engage opposing views constructively** — identifying and articulating objections to one's own views, applying the principle of charity to others' positions, and responding to disagreement respectfully and effectively.
- **Interpret and analyze philosophical texts** — reading primary sources (both historical and contemporary) with care, extracting their central arguments, and situating them in broader debates.
- **Refine ideas through writing** — transforming initial insights into polished, well-structured pieces of philosophical prose.
- **Develop and defend original positions** — constructing one's own informed and reasoned responses to enduring problems in political philosophy.

## TEXTS

All readings will be made available on the course website, which you can access here:  
<https://pitzer.instructure.com/courses/2241>.

## COURSE REQUIREMENTS

The course requirements are designed to help you build three types of skills relevant to philosophy (and general life): (1) how to clearly and productively discuss a difficult text or complex topic, (2) how to read a difficult text, and (3) how to clearly and compellingly express yourself in writing.

## **Attendance and Participation (20%).**

Philosophy is not simply a set of facts or questions that one could passively memorize; it is rather a way of approaching important problems and thinking them through. As such, philosophy is something that is “done.” It is thus crucial to learning philosophy that it be actively practiced. Your attendance and informed participation in the class is therefore expected and required. This generally entails regular attendance and engagement in discussion during class, but can also be accomplished by taking part in other fora, such as office hours. More specifically, all of the following counts as active participation:

- Asking questions and engaging with the instructor and other class participants in a courteous and charitable dialogue about the week’s required reading.
- Asking a classmate to clarify or repeat a point they made or asking them a question about something they said.
- Explaining that you are confused or unsure about some aspect of the course material.
- Visiting office hours or scheduling an appointment with me to discuss the material.

For further examples of effective participation, please see **“But How Do I Participate? A Sampling of Ways to Contribute to a Philosophical Conversation,” by Olivia Bailey**, which is posted to the course website.

Having said that, I do of course understand that life sometimes gets in the way of attending class. Each student will thus be permitted **three absences from the course, no questions asked. You do not need to email me or provide any excuse for missing any of those three classes.** However, if you miss *more* than three classes (for any reason), then we may need to determine some way to make up for the missed class time, since attending class is an integral part of the learning process for this course.

## **Protocols (20%)**

A protocol is a 1–2 page (double-spaced) reading response, which should be submitted to the course website within two weeks of the day that reading is due. The protocol can take various forms including:

- A summary of key points in the reading with one or more critical questions
- A working through of one or two issues in the reading
- A close reading of one or more quotes from the reading with analysis of the argument

- A reflection on the reading in relation to personal experience, beliefs, current events, or examples from media
- Any combination of the above

Grading of the protocols will depend on effort, depth of the engagement with the text, and stylistic clarity. The grading scale is:

- ✓+: “Excellent,” 20/20 (100%)
- ✓: “Satisfactory,” 17/20 (85%)
- ✓: “Unsatisfactory,” 14/20 (65%)
- 😕: “Incomplete,” (you didn’t submit a protocol), 0/20 (0%)

You must submit at least **4 protocols throughout the semester**. *When* you choose to submit them is up to you. The only requirement is that any protocol must be submitted within two weeks of the class session for which the reading was assigned. For example, if the reading is due on September 3, you may submit a protocol on it any time up to September 17.

### Three Essays

The majority of your grade will be determined by how well you do on the three required papers you must complete for the course. Having to explain our ideas to others in writing forces us to be clear, and to be thoughtful and honest about the positions we hold. We will talk about what is expected of the problem essays, and discuss tips for writing well in philosophy, at a number of points before the papers are due.

**First essay (15%).** Your first essay will be a short paper of 2–3 double-spaced pages with 1-inch margins and 12 pt. font. It will ask you to address a question concerning readings we have discussed in class. You will be able to select from among a choice of different questions.

**Second essay (20%).** The second essay will be a slightly longer and somewhat more complex paper of 3–4 double-spaced pages (with the same formatting as above). It will ask you to address a question concerning readings or thinkers you have not yet written on. Again, you will be able to select from among a choice of different questions.

**Third essay (25%).** The third essay will be a paper of 5-6 double-spaced pages (with the same formatting as above) that will ask you to address the arguments of two or more theorists you have not yet written on. Because the third paper will be longer and more complex than the others, you will have more time to complete it, and it will be worth a higher portion of your overall course grade. Again, you will be able to select from among a choice of different questions.

## Extra Credit Opportunities

At various points in the course, I may offer extra credit assignments. These will typically involve short writing exercises that connect a reading or idea from class to another medium (such as a news article, film, or current event). I'll announce these opportunities as they become available.

## Summary Grade Breakdown and Paper Due Dates

Assignment	Length	Due	Percent of Grade
Participation	-----	-----	20%
Protocols (4)	1–2 pages each	-----	20%
First Essay	2–3 pages	9/24	15%
Second Essay	3–4 pages	11/01	20%
Third Essay	5–6 pages	12/12	25%

## GRADING AND LATE POLICY

Unless otherwise stated, all assignments will receive a letter grade based on the Pitzer 4.0 scale. Papers and presentations are due at the time and date specified. Work turned in late is considered to have been turned in the next day and will be penalized 1/3 of a grade (A to A-, B- to C+, etc.) for each day late (including weekend days). Email submissions will not normally be accepted. Under appropriate circumstances, short extensions on papers may be granted so long. To arrange an extension please email me at least 36 hours in advance of the due date. In some rare instances, such as in last-minute emergencies, an extension may be granted after the deadline has passed.

## ELECTRONICS POLICY

Phones, tablets, laptops, and other electronic devices are not permitted in class, unless prior permission has been arranged. Our classroom is meant to be a space for sustained, focused engagement. Research consistently shows that taking notes by hand improves comprehension compared to laptop note-taking, and that multitasking not only lowers academic performance but also increases susceptibility to distraction. In short, even though we may feel adept at multitasking, the evidence is clear: we learn better when we give our full attention to the task at hand. Respect for each other's focus and presence is also an important part of our shared classroom environment.

Recording class sessions (audio or video) is likewise not permitted.

If you are concerned about keeping track of everything discussed in class without a device, keep in mind two things. First, this course has no midterm or final exam, and no assignment will test rote

memorization of facts. Detailed transcripts of our discussions will thus not be necessary for such purposes. Second, while I do not plan to record class sessions (except when pedagogically necessary), I will frequently provide handouts on the day's readings. These will serve both to guide our discussions and to give you a record to take away, so you can focus your energy on participating rather than transcribing.

You should bring all assigned readings to class in hard copy, either printed out or as the physical book itself. If printing is a challenge for you, please let me know and we can find a solution.

If you require accommodations or if this policy poses difficulties for personal reasons, I encourage you to meet with me so we can work out an arrangement that supports your participation.

## ACCOMMODATIONS

If you are a student with academic accommodation needs based on learning differences, chronic illness or other needs please contact the Pitzer Office of Academic Support Services (PASS) at [academicsupport@pitzer.edu](mailto:academicsupport@pitzer.edu) or (909) 607-0213. In cases where a need may conflict with the course policies, students are also invited to discuss this with me.

## ACADEMIC INTEGRITY

All student work must be of original creation. It is always better to turn in an essay of your own creation than to plagiarize. Cases of plagiarism will be reported to the administration for disciplinary action. Please review the Academic Policies and Procedures, as well as the Guide to Student Life, for information on what counts as plagiarism. <https://www.pitzer.edu/tlc/academic-integrity/>

If you have any question about what counts as a violation of academic integrity or how to cite your sources, please consult with me. Err on the side of caution. Any standard citation style will do (Chicago, APA, MLA, etc.). If you aren't sure which to use, consider Chicago style author-date format: [https://www.chicagomanualofstyle.org/tools\\_citationguide/citation-guide-2.html](https://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-2.html)

## AI POLICY

The use of generative AI systems, such as ChatGPT, Gemini, or Copilot, for completing written assignments in this course is not permitted. Yes, even thought this is a course on AI! This is not simply because I am "old-fashioned." Rather, it is because the central purpose of these assignments is to help you develop your own skills in philosophical thinking and writing. Outsourcing these skills to a machine undermines that goal.

I recognize that AI can sometimes be used in less direct ways, such as for correcting grammar or refining your writing. But learning to write clearly, which is itself an extension of learning to *think* clearly, is a key objective of this course. If you rely on AI to perform that work for you, you deprive yourself of an essential part of the learning process. (Note as well that you will not lose points for grammatical or spelling errors, so there is little need to worry on that front.)

That said, there *may* be justified uses of AI that do not conflict with these aims: for example, brainstorming ideas or mapping out possible directions for a paper. If you are unsure whether a particular use is acceptable, I encourage you to ask me. To be clear: what is not permitted is the use of AI to generate or refine any portion of the written work you submit for credit.

If I suspect that AI has been used, I may compare your work with your in-class writing assignments or employ AI-detection tools. I hope not to do this. I have no interest in policing your work. But I will do so if it becomes a problem.

Finally, note that the assignments are deliberately designed to be poor fits for AI. For instance, you may be asked to incorporate insights from our in-class discussions or to develop an original philosophical argument, which are things an AI system cannot do particularly well on your behalf. So, in addition to violating course policy, relying on AI is far from a sure-fire way of earning a strong grade.

## WRITING CENTER

I strongly encourage you to utilize the Pitzer Writing Center, which is an invaluable resource for this course and beyond. The Writing Center is located in 131 Mead Hall, just across from the fountain, and offers virtual and in-person consultations with peer Fellows trained to work with writers on assignments in any discipline and at any stage of the writing process, from brainstorming ideas to polishing a final draft. The Writing Center is one of Pitzer's most popular academic resources, holding close to 2,000 one-on-one consultations per year. To book a 50-minute session or learn more about workshops and other resources and events, visit <https://www.pitzer.edu/writing-center>. All Pitzer students may use the Writing Center for any writing need. Students from the other 5Cs are welcome to book appointments for courses taken at Pitzer.

## COURSE READINGS AND SCHEDULE

As you may already know, reading philosophy is challenging. Here are some resources students report they've found useful in improving their reading experience (but feel free to contact me for more guidance about reading):

- <https://www.blogs.ppls.ed.ac.uk/2017/02/28/read-philosophy-step-step-guide-confused-students>
- <https://philosophy.arizona.edu/sites/philosophy.arizona.edu/files/Rosati%20How%20to%20Read%20a%20Philosophical%20Article%20or%20Book.pdf>
- <http://melissajacquart.com/wp-content/uploads/2014/09/How-to-Read-Philosophy-Handout.pdf>

You will see that some materials below are marked as “optional.” These are truly optional: you are not expected to engage with them, and if I reference them in class, I will do so without assuming prior familiarity. That said, they can enrich your understanding of the material and deepen class discussions, and I encourage you to explore them if you have the time and interest. Optional materials may also play a role in extra credit assignments offered during the semester, and they are fair game for you to reference in your own assignments and papers if you find them useful.

*~ Note: this schedule below is tentative and liable to change as the class progresses ~*

---

### ***Unit 1: Introduction***

**M 8/25**

Introduction

**W 8/27**

**Required**

Langdon Winner, “Do Artifacts Have Politics?”

**M 9/01**

***Labor Day, No Class***

**W 9/03**

**Required**

Walter Sinnott-Armstrong, Jana Schaich Borg, and Vincent Conitzer, “What is AI?”

Matthew S. Liao, “A Short Introduction to the Ethics of AI”

**Optional**

Luke Munn, “The Uselessness of AI Ethics”

Kate Crawford, Vladan Joler, [“Anatomy of an AI System”](#)

**M 9/08**

**Required**

David Morrow, “The Values Built Into Technologies”

Justin B. Biddle, “Values in Artificial Intelligence Systems”

**Optional**

Joseph C. Pitt, “Value Free Technology?”

Shen-yi Liao, Bryce Huebner, “Oppressive Things”

***Unit 2: Automated Decisions and Algorithmic Bias***

**W 9/10**

**Required**

Claire Benn and Seth Lazar, “What’s Wrong with Automated Influence?”

**Optional**

Drew Harwell (*The Washington Post*), [“Hirevue’s AI Face-Scanning Algorithm Increasingly Decides Whether You Deserve The Job”](#)

Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorow, [“Physiognomy’s New Clothes”](#)

**M 9/15**

**Required**

Logic mini-lecture slides (parts 1 and 2)

## Writing advice slides

Claire Benn and Seth Lazar, “What’s Wrong with Automated Influence?” (continue)

**W 9/17**

### **Required**

Claire Benn and Seth Lazar, “What’s Wrong with Automated Influence?” (finish)

**M 9/22**

### **Required**

Renée Jørgensen, “Algorithms and the Individual in Criminal Law”

### **Optional**

Kasper Lippert-Rasmussen, “‘We Are All Different’: Statistical Discrimination And The Right To Be Treated as an Individual”

Ellora Israni (*New York Times*), [“When an Algorithm Helps Sends You to Prison”](#)

*Hi-Phi Nation* Podcast, [“The Pre-Crime Unit”](#) and [“Risky Business”](#)

**FIRST PAPER DUE WED. 9/24**

**W 9/24**

### **Required**

Gabrielle M. Johnson, “Algorithmic Bias: on the Implicit Biases of Social Technology”

### **Optional**

Gabrielle M. Johnson, “Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning”

Sina Fazelpour, David Danks, “Algorithmic Bias: Senses, Sources, Solutions”

Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner (*ProPublica*), [“Machine Bias: There’s Software Used Across The Country To Predict Future Criminals. And It’s Biased Against Blacks”](#)

**M 9/29**

**Required**

Kate Vredenburgh, “The Right to Explanation”

**Optional**

Warren J. von Eschenbach, “Transparency and the Black Box Problem: Why We Do Not Trust AI”

Samuel Dishaw, “The Right to a Justification”

Aaron Klein (*Brookings*), “[Credit Denial in the Age of AI](#)”

Lauren Kirchner and Matthew Goldstein (*New York Times*), “[How Automated Background Checks Freeze Out Renters](#)”

***Unit 4: Distinctively Human Values?***

**W 10/01**

**Required**

James Lenman, “On Becoming Redundant or What Computers Shouldn’t Do”

Robert Nozick, “The Experience Machine”

**M 10/06**

**Required**

Shannon Vallor, “Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character”

**Optional**

Joshua Schulz (*The New Atlantis*), “[Machine Grading and Moral Learning](#)”

Shannon Vallor, “Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century”

**W 10/08**

**Required**

Amy Kind, “Love in the Time of AI”

*Black Mirror* episode, “[Be Right Back](#)”

**Optional**

Kurt Vonnegut, “EPICAC”

Spike Jonze’s film, *Her*

*Hi-Phi Nation* Podcast, “[Love in the Time of Replika](#)”

John Danaher, “The Philosophical Case for Robot Friendship”

Apryl Williams (*Time*), “[When Love and the Algorithm Don’t Mix](#)”

Shannon Vallor, “Flourishing On Facebook: Virtue Friendship & New Social Media”

**M 10/13**

*Fall Break, No Class*

***Unit 5: Democracy and AI***

**W 10/15**

**Required**

John Danaher, “The Threat of Algocracy: Reality, Resistance, and Accommodation”

Optional

Evgeny Morozov (*MIT Technology Review*), “[The Real Privacy Problem](#)”

**M 10/20**

**Required**

Thomas Cristiano, “Algorithms, Manipulation, and Democracy”

**Optional**

Fiona McEvoy, “Political Machines: Ethical Governance in the Age of AI”

Jonathan Zittrain (*Harvard Business Review*), “Engineering an Election  
Digital gerrymandering poses a threat to democracy”

**W 10/22**

**Required**

C. Thi Nguyen, “Echo Chambers and Epistemic Bubbles”

**Optional**

C. Thi Nguyen (*Aeon*), [“Escape the Echo Chamber”](#)

Kashmir Hill, Dylan Freedman (*New York Times*), [“Chatbots Can Go Into a Delusional Spiral. Here’s How It Happens.”](#)

**M 10/27**

**Required**

Regina Rini, “Deepfakes and the Epistemic Backstop”

**Optional**

Adam Sartariano, Paul Mozur (*New York Times*), [“The People Onscreen Are Fake. The Disinformation Is Real.”](#)

*Tech Won’t Save Us* Podcast, [“Is Social Media Fueling Far-Right Riots?”](#)

**W 10/29**

**Required**

Regina Rini and Leah Cohen, “Deepfakes, Deep Harms)

**Optional**

Samantha Cole, (*Vice*) [“AI-Assisted Fake Porn Is Here and We’re All Fucked”](#)

Samantha Cole, (*Vice*), [“Deepfakes Were Created as a Way to Own Women’s Bodies—We Can’t Forget That”](#)

**SECOND PAPER DUE SATURDAY 11/01**

***Unit 6: Work***

**M 11/03**

**Required**

John Danaher, “In Defense of the Post-Work Future: Withdrawal and the Ludic Life”

**Optional**

Kate Wells (NPR), [“National Eating Disorders Association Phases Out Human Helpline, Pivots To Chatbot”](#)

**M 11/05**

**Required**

Pegah Moradi, Karen Levy, “The Future of Work in the Age of AI: Displacement or Risk-Shifting?”

**Optional**

Karen Levy, (*Wired*), [“Robo Truckers and the AI-Fueled Future of Transport”](#)

**M 11/10**

**Required**

Kate Vredenburgh, “Freedom at Work: Understanding, Alienation, and the AI-Driven Workplace”

**Optional**

Jan Kandiyali, “The Importance of Others: Marx on Unalienated Production”

**W 11/12**

**Required**

James Muldoon and Paul Raekstad, “Algorithmic Domination in the Gig Economy”

**Optional**

Karen Hao, Nadine Freischlad (*MIT Technology Review*), [“The Gig Workers Fighting Back Against The Algorithms”](#)

*Hi-Phi Nation* Podcast, [“The Problem with Gig Work”](#)

*More Perfect Union*, [“How Uber Robs Its Drivers”](#)

***Unit 7: Creating Ethical Agents***

**M 11/17**

**Required**

Peter Railton, “Ethical Learning, Natural and Artificial”

**Optional**

Regina Rini (*Aeon*), [“Creating Robots Capable Of Moral Reasoning Is Like Parenting”](#)

**W 11/19**

**Required**

Nick Bostrom, “The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents”

**Optional**

Michael Anderson, Susan Leigh Anderson, “Machine Ethics: Creating an Ethical Intelligent Agent”

**M 11/24**

***Catch-Up Day***

**W 11/26**

***Thanksgiving Break, No Class***

**M 12/01**

**Required**

Brett Karlan, “Human Achievement and Artificial Intelligence” (Winner of student vote on open-day topic)

**W 12/03**

**Required**

Robert Sparrow, “Killer Robots” (Winner of student vote on open-day topic)

**FINAL PAPER DUE FRIDAY 12/12**