CSCI 375: Ethics and Epistemology of Artificial Intelligence

Michael McCarrin

Spring, 2025

E-mail: mmccarri@oberlin.edu   Web: **https://cs.oberlin.edu/~mmccarri**
Office: 223C   Class Room: King 325
Student Hours: Class Times: MF 3:00-4:15pm
MF 9:30-10:30am or by appointment

Course Description

As machine-learning systems become more prevalent and more powerful, they play an ever larger role in determining what we believe and how we decide to act. This seminar will investigate the underlying assumptions that inform work on artificial intelligence, as well as practical implications for issues of social justice and equity. Because the root problems often extend beyond the field of computer science, we will take an interdisciplinary approach that incorporates philosophical, historical and social context. Coursework will combine reading and discussion of academic scholarship with hands-on programming exercises.

Course Objectives

1. Learn to analyze claims regarding potential benefits and harms of developments in artificial intelligence.

2. Acquire an intuitive understanding of core technical concepts used in artificially intelligent systems, especially those pertaining to deep learning and large language models.

3. Identify key concepts at stake in contemporary discussions about AI (e.g., intelligence, bias, subjectivity, understanding, knowledge, etc.), and consider the cultural and historical context in which these concepts have been formulated.

4. Practice explaining theoretical and technical concepts in language accessible to an educated general reader.

Textbook

You do not need to purchase a textbook for this course. Links to required readings will be provided on the course Blackboard page. Some assignments may refer to the following book, which is available through Oberlin's subscription to the O'Reilly Safari Learning Platform:

- Raschka, Sebastian, Yuxi Hayden Liu, and Vahid Mirjalili. *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python.* Packt Publishing Ltd, 2022.

See the Blackboard page for instructions about how to access this textbook.

Content and Schedule

Course content falls into three related categories:

- Contemporary academic articles addressing problems of ethics and bias in artificial intelligence and machine learning systems.
- Canonical texts in philosophy and critical theory relating to concepts that inform the current moment.
- Textbook and video materials providing technical explanations of core components of AI systems.

See the course Blackboard page for a detailed list of source materials and the assignment schedule.

Activities and Assignments

Weekly Readings / Videos

Readings and videos will be assigned each week in Blackboard. The dates given for these assignments indicate the date of class in which we will discuss them. Readings are ungraded but will be necessary to complete your assignments and participate meaningfully in class discussion. I reserve the right to give in-class reading quizzes if conditions call for them.

Class Participation (20% of grade)

Your participation in class discussions and activities will be an important part of your work in this seminar. Most days will involve activities like individual generative writing, pair work and group discussion. We will use this time to go over questions and share ideas about assigned reading and other homework.

You may miss up to three classes without penalty. Should you need to miss more than three classes, please get in touch with me so we can discuss your situation.

Special Discussion Roles (15% of grade)

In addition to regular discussion activities, each class will have designated **note takers** who contribute to our shared collaborative notes for the course. This role will typically last for half a class period and be shared by two students at a time (so we will generally rotate through four note takers in a class period).

Everyone will also lead at least one discussion over the course of the semester. The **discussion leader** will be expected to prepare materials briefly summarizing the topic or assignment to be discussed, and some questions or activities to guide discussion. The materials and questions will be submitted in Blackboard before the start of class.

A sign-up sheet for both roles will be available in Blackboard. To get full credit in this category, you will need to lead one discussion and serve as a note taker five times over the course of the semester. If a situation arises in which you are unable to attend on a day you signed up for, it is your responsibility to find another classmate to switch with you.

Homework (30% of grade)

Weekly assignments will be posted on Blackboard. These may include technical exercises in Google Colab Notebooks, worksheets, and other questions or activities related to your reading. Generally, these assignments will be due on Thursday night, and will be graded on completion. We will discuss them and give feedback in class.

Weekly Reflections (10% of grade)

To help get a sense of our collective interests, questions and concerns, I ask you to complete short weekly reflections. The goal of these reflections is to give you an opportunity to say what ideas stood out to you, what you found challenging, what you would like to know more about, and how you're doing. They will be short (a few sentences) and should take about 5 minutes per week. You may miss up to three reflections without penalty.

Final Presentation (25% of grade)

Your culminating assignment in this course will be an 8-10 minute presentation describing the connection between a position taken in one of the assigned articles in the course, and a technical component of a machine learning system. To help you prepare, you will create and submit a short précis midway through the semester summarizing your intended content. See Blackboard for further details about how to complete this assignment, as well as an explanation of how it will be graded.

Grading Mechanics

Your overall score for the course will be calculated as follows:

Category	Percentage
Attendance and Participation	20%
Special Discussion Roles	15%
Weekly Reflections	10%
Homework and Quizzes	30%
Final Presentation	25%

Final Grades

At the end of the class, I will assign your final letter grade according to the following guide:

Points	Letter Grade
by discretion	A+
93 and above	A
90-92	A-
87-89	B+
83-86	B
80-82	B-
etc...	

Although I generally follow the table above, judgment calls are sometimes necessary and I reserve the right to apply a curve if it makes sense to do so.

Course Policies

This Syllabus

This syllabus is subject to change. I will communicate any substantial policy changes through multiple channels (Blackboard, in-class announcements, etc.)

Getting in Touch

You are always welcome to contact me by email. On workdays I will respond to your emails within 24 hours. Outside of workdays, I offer a "best effort" service, but I can't promise instant responses, so please make every effort to plan ahead.

What to do if you Miss Class

If you have to miss a class, please let me know as soon as you can. To get caught up, please contact a classmate and get notes about what we covered while you were out. This doesn't have to be a detailed transcript of everything that was said, but it should include at least a basic outline of the topics we touched on, as well as any discussion questions or activities. Make sure to check whether we talked about upcoming assignments or deadlines. It is helpful to make a plan in advance with your classmates in case this happens.

Late Assignments and Extensions

I strongly encourage you to turn work in by the course assignment deadlines. They are intended to help you get the most out of the class and also to encourage a long-term intellectual practice of steady engagement with your work.

Of course, I understand unexpected situations come up. If you need to request an extension, please email me as soon as possible, and no later than the day before the assignment is due, with the following information:

- a short explanation of your situation
- *your proposal for an alternative deadline*

For practical reasons, I make no guarantees about granting extension requests the day the assignment is due. Of course, if extenuating circumstances arise (such as illness, hospitalization, family emergencies, etc.) that interfere with your ability to submit assignments, please use your best judgment and let me know as soon as you can.

In the absence of an extension, late assignments will lose 5% for every 24 hours after the due date. Work that is more than one week late will not be accepted without explicit permission.

Academic Integrity and Honesty

You are expected to adhere to the Oberlin College Honor Code. Any violations will be reported to the Honor Code Committee. Please remember to write and sign the Honor Pledge on all assignments for this course. The new pledge reads: "I have adhered to the Honor Code in this assignment."

Because of the nature of the course, we may sometime use ChatGPT or other generative AI systems as part of the assignments, and I will indicate explicitly when I expect you to do so. Obviously, you are **not** allowed to submit text produced by ChatGPT *or any person or text-generating entity* as if it were your own work. To avoid any appearance of impropriety, please be sure to credit any sources you quote or paraphrase in any responses you turn in by properly citing them.

Accommodations for Disabilities

It is the policy and practice of Oberlin College and Conservatory to create inclusive learning environments. If there are aspects of the instruction or design of this course that present barriers to your inclusion or an accurate assessment of your needs to best complete course requirements (e.g., time-limited exams, inaccessible web content, use of videos without captions), please let me know and contact the Office for Disability and Access to request accommodations. Office: 440-775-5588, Mudd Center 205. Students can learn more about the accommodation process on the ODA's website: www.oberlin.edu/disability-access.