Introduction to AI Ethics -
An Interdisciplinary Approach
COMSCI 88SB | Spring 2020

Instructor
Aaron Hui | aaronhui98@ucla.edu

Office Hours
Online | By Appointment Only

Course Description

As technology advances at an exponential rate, it is imperative that students begin thinking about how Artificial Intelligence (AI) will interact with and impact society at different facets of daily life, with the focus on ethical implication and its implementation within AI systems. The foresight of how important it is to regulate AI through the lens of ethical discussion and implementation is of paramount significance within the setting of the rapid development of AI and how it has already permeated our daily lives.

This course will allow students to gain an interdisciplinary introduction to classical and modern ethical theory and their implications on emerging autonomous technologies. We will cover a variety of issues, but they will all address one of the six ethical principles of AI identified by Microsoft:

1. **Fairness:** AI systems should treat all people fairly.
2. **Inclusiveness:** AI systems should empower everyone and engage people.
3. **Reliability and Safety:** AI systems should perform reliably and safely.
4. **Transparency:** AI systems should be understandable by people.
5. **Privacy and Security:** AI systems should be secure and respect privacy.
6. **Accountability:** AI systems should have algorithmic accountability.

This course will include introductions to the AI Robotics Ethics Society (AIRES), the UCLA Law AI Pulse program, the AI Ethics Lab, the USC Center for Artificial Intelligence in Society (CAIS), and more.
Spring 2020 Adjusted Syllabus

1. Nobody signed up for this.
   a. Not for the sickness, not for the social distancing, not for the sudden end of our collective lives together on campus.
   b. Not for an online class, not for teaching remotely, not for learning from home, not for mastering new technologies, not for varied access to learning materials.

2. The humane option is the best option.
   a. We are going to prioritize supporting each other as humans.
   b. We are going to prioritize simple solutions that make sense for the most.
   c. We are going to prioritize sharing resources and communicating clearly.

3. We cannot just do the same thing online.
   a. Some assignments are no longer possible.
   b. Some expectations are no longer reasonable.

4. We will foster intellectual nourishment, social connection, and personal accommodation.
   a. Accessible asynchronous content for diverse access, time zones, and contexts.
   b. Optional synchronous discussion to learn together and combat isolation.

5. We will remain flexible and adjust to the situation.
   a. Nobody knows where this is going and what we'll need to adapt.
   b. Everybody needs support and understanding in this unprecedented moment. PEOPLE COME FIRST.

This adjusted syllabus was adapted from Professor Brandon L. Bayne.
Learning Objectives

By the end of this course, you should be able to:

1. Understand the current issues in the field of AI Ethics today.
2. Summarize arguments of modern and classical philosophy as it applies to AI ethics.
3. Construct unique rational arguments for and against ethical perspectives as they pertain to AI technology.

Course Materials

There is no textbook for this course as such a textbook does not exist yet. All materials will be provided to you on CCLE.

Technology

This course will be using Zoom for all in lectures and discussions during the COVID-19 pandemic. Please turn video on and let me know asap if you have any technical difficulties. If you have Internet connectivity issues, the option to dial-in via phone is always available. If you still can’t access the Zoom, all meetings will be recorded and uploaded to YouTube and CCLE, so view them when you can.

Grading Policy

This course is a 1-unit P/NP seminar. As with all UCLA courses, a minimum of 70% is required in order to receive a passing grade. Grading is based upon three components:

1. Class participation: 33.3%
2. Weekly forum post and response: 33.3%
3. Final Paper: 33.3%
Attendance/Participation Policy

Participation will be based on active contribution to in-class discussions or activities. Only one unexcused absence will be allowed, along with a make up assignment on the week’s discussion. Per UCLA and USIE policy, more than one unexcused absence from the course will result in a failing grade. Special exceptions may be granted for emergencies given the unprecedented times we are living in. Please contact me for more information.

Forum Posts

Readings are to be completed before class. All readings will be provided on CCLE. You must complete at least 5 of the 9 post sets starting Week 2 in order to receive this portion of the grade.

Post 1: Before the beginning of each class beginning week 2, pose a question about the week’s assigned reading assignments on CCLE. This question should be open-ended and not able to be answered with a simple “yes” or “no.”

Post 2: After each class and before the start of the next class, answer a classmate’s question on CCLE!

Final Paper

In light of the adjusted syllabus, the poster project has been cancelled. Instead, there will be one two-page double spaced final paper that will be due on Week 10. The paper will be on a potential solution to one of the issues in AI ethics presented throughout the course. Completion of the final paper is a necessary requirement to pass the course. You will have the option to do it with a partner or solo if you are feeling ambitious. Amazing papers showing great potential may be selected for further review and future publication in the AI Ethics Journal (aiethicsjournal.org).
Academic Integrity

Academic misconduct in any form (including plagiarism, note-selling, multiple submissions, and cheating) will be dealt with according to UCLA’s policy and procedures regarding academic honesty. This includes reporting suspected violations to the Dean of Students. If you are uncertain as to what constitutes plagiarism, the library has a helpful guide: (http://guides.library.ucla.edu/citing/plagiarism/avoid).

When you submit an assignment with your name on it, you are signifying that the work contained therein is yours, unless otherwise cited or referenced. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged. Penalties for academic misconduct may include a failing grade on the assignment and/or a failing grade in the course, among other possibilities. If you are unsure about the expectations for completing an assignment or taking a test exam, be sure to seek clarification beforehand.

DISCLAIMER

This syllabus is intended to give students guidance in what may be covered this quarter. The instructor reserves the right to make modifications to this information as the course progresses, especially in light of the novel applications of AI in light of the COVID-19 pandemic.

The AI Robotics Ethics Society (AIRES)

The AI Robotics Ethics Society is the largest AI related ethics organization in the United States. Founded by Aaron Hui at UCLA in 2018, AIRES is now a nonprofit organization with chapters at USC, Caltech, Stanford, Cornell, and Brown. At the AI Robotics Ethics Society, we focus on educating tomorrow’s AI leaders in ethical AI principles to ensure AI is created ethically and responsibly. Our initiatives include the AI Ethics Journal (AIEJ), Beyond the Turing Test Podcast, highschool outreach, and our AIRES conferences. Join and learn more at theaires.org.
CAMPUS RESOURCES FOR STUDENTS

Academic Achievement Program (AAP):

AAP advocates and facilitates the access, academic success, and graduation of students who have been historically underrepresented in higher education; informs and prepares students for graduate and professional schools; and develops the academic, scientific, political, economic, and community leadership necessary to transform society. Learn more at http://www.aap.ucla.edu/

Undergraduate Writing Center:

The Undergraduate Writing Center offers UCLA undergraduates one-on-one sessions on their writing. The Center is staffed by peer learning facilitators (PLFs), undergraduates who are trained to help at any stage in the writing process and with writing assignments from across the curriculum. PLFs tailor appointments to the concerns of each writer. Multiple locations and hours available. For more information or to schedule an appointment, visit http://wp.ucla.edu/wc/

Center for Accessible Education (CAE):

Students needing academic accommodations based on a disability should contact the Center for Accessible Education (CAE) at (310)825-1501 or in person at Murphy Hall A255. When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations. For more information visit www.cae.ucla.edu.

Counseling and Psychological Services (CAPS):

CAPS supports student mental health needs as they pursue their academic goals. Their services are designed to foster the development of healthy
well-being necessary for success in a complex global environment. CAPS offers a variety of services to meet student needs including: crisis counseling available by phone 24 hours a day/7 days a week 310-825-0768, emergency Intervention, individual counseling and psychotherapy, group therapy, psychiatric evaluation and treatment, psychoeducational programs and workshops, and campus mental health and wellness promotion. Please visit http://counseling.ucla.edu for more information.
TENTATIVE Schedule (Subject to Substantial Change)

Readings and videos will be posted on CCLE under each week. Refer to the designations listed below for readings

Required: (R: approx reading time)
Highly recommended: (HR)
Optional: (O)

Week 1: Introduction to Ethics in AI and Machine Learning

- This unit will cover the structure of the seminar, expectations, and goes over the syllabus. We will introduce the main topic of the course and review the current debates in AI ethics and the fundamentals of machine learning. Discussion topics will focus on why AI ethics is necessary in developing AI, and whether ethics can be hard coded into AI systems.

Week 2: How to Analyze and Create Arguments

- This unit will cover the basics of common argument forms involving inductive and deductive arguments. Discussion topics will include debates in various topics raised by students but utilizing proper argument forms instead of resorting to violence and rhetoric or emotional appeals.

Week 3: The Different Ethics

- This unit will cover the basics of Utilitarianism, Deontological, and Virtue Ethics and how they pertain to AI ethics. Discussion topics will compare the three ethical standpoints and see how each one can contribute to AI ethics.

Week 4: DEBATE WEEK

- This until will be spent on reviewing all the covered ethical standpoints and the class will be split into groups advocating for different ethical
perspectives. They will debate on which one of the ethical standpoints they find best suited for implementing into AI systems. Students will be randomly assigned to one of the three ethical perspectives.

Week 5: Issues in Fairness

- This unit will cover the issues of fairness in AI in terms of bias towards certain groups of people. Discussion topics will include a case study on the problems of racial bias for algorithms designed to determine prison sentences for criminals.

Week 6: Issues in Inclusiveness

- This unit will cover the issue of how humans remain useful in a world of autonomous machines. Discussion topics will include a case study of the relationship between the American trucker and AI.

Week 7: Issues in Transparency

- This unit will cover the issue of the problem with understanding how AI fundamentally makes their decisions. Discussion topics will include a debate on a case study in the Foreseeability Problem of AI regulation.

Week 8: Issues in Accountability

- This unit covers how it could be possible to make AI systems accountable. Discussion topics will include a debate on why it’s necessary to ensure that AI systems are designed ethically and responsibly, and a case study on the self driving car.

Week 9: Issues in Reliability and Safety

- This unit will cover how we can create AI in a safer and more reliable manner. Discussion topics will include a special emphasis on the usage of AI in medicine and how diagnostic accuracy and reliability is a
primary concern, as well as a secondary case study on the self driving car.

**Week 10: Issues in Privacy and Security**

- This unit will cover the dangers of allowing AI to invade the privacy of individuals. Discussion topics will include a special emphasis and discussion on China’s upcoming AI controlled Social Credit System, and a case study on Facebook Privacy.