



EIS 370: Ethical Issues in Artificial Intelligence

Fall 2024

Instructor: Elisa Warford, Ph.D.

warford@usc.edu

Section 28400 and 28401

MW 10:00-12:00 and 12:00-2:00

Office: OHE 106N

In-person drop-in hours: MW 2:00-3:00 and by appointment

Zoom office hours by appointment

Zoom Personal Meeting ID: <https://usc.zoom.us/my/elisawarford>

Course Overview and Objectives

Artificial intelligence is transforming virtually every domain of human endeavor: science and healthcare, business and the economy, politics, entertainment and the arts, our relationships, and our bodies themselves. The promises of AI are tantalizing, but it is essential that we reflect on the technological ends and means we are pursuing. How do we develop AI that supports or helps bring about flourishing human lives, a just society, and a thriving natural environment? In this course, we explore artificial intelligence from a practical ethics perspective—one that aims to answer real-life ethical questions surrounding AI.

We'll explore questions about AI's impact on issues related to liberal democratic values, such as equality and equity, free speech, and privacy. Then we'll look at AI's impact on our humanity—our relationships, our work, and our conceptions of the good life. The work of philosophers will inform our discussion, but so will creative works of fiction and film, current news events, and the views of people in the tech industry. Throughout, we'll discuss the moral responsibilities of software engineers/programmers and tech companies who are developing ever-stronger AI systems. The aim of this course is for students planning to work in AI to develop critically informed views on these thorny yet important issues.

In this course, students will develop their abilities to

- Explore the connections between ethics, society, and engineering;

- Understand and engage with a range of views on ethical issues related to AI and to ethics more broadly;
- Understand and reflect on the responsibilities of programmers/software engineers and tech companies regarding the ethical development and use of AI;
- Develop critical reading and reasoning skills through close reading and interpretation of philosophical and other texts;
- Use these skills to develop reasoned positions on the ethical issues;
- Communicate reasoned positions in clear, concise prose, oral presentations, and classroom discussion.

Prerequisite(s): None

Co-Requisite(s): None

Concurrent Enrollment: None

Recommended Preparation: None

Catalog Description

Ethical issues in artificial intelligence including algorithmic bias and interpretability, data privacy and control, autonomous systems and relationships with AI entities.

Contact Me

Don't hesitate to drop by my drop-in office hours to discuss the course, your progress in it, or any other concerns. Outside of office hours, email is the best way to contact me (warford@usc.edu). I do my best to respond to emails within 24 hours, but expect a longer turnaround over the weekend.

Course Modality and Attendance

The class is offered in-person. There is no official attendance policy for this course. However, I will take roll, and it is not a lecture course for which you can easily download notes from lecture slides or watch a lecture online. Rather, the class format is largely based on live class discussion. Therefore, for you to succeed in the course, it is important that you attend the sessions; you will be at a disadvantage—and the class will not be as engaging—if you are absent.

If you must miss class due to illness or other extenuating circumstances, please consult with a classmate to see what you missed, or meet with me during drop-in hours.

Brightspace

This syllabus (including updates to it as necessary), class readings, the paper assignment prompts, course handouts, your grades, announcements, and other course materials will be posted on our course Brightspace page. Please check Brightspace often.

Assignments and Grading

The assignments for the course are designed to provide a range of ways for you to develop positions and exchange your ideas about the course topics, and to demonstrate your critical thinking and communication proficiency.

Perusall

We'll be using the online platform Perusall to engage with course readings both individually and collaboratively. Perusall promotes active reading and discussion of the texts. To access Perusall through Brightspace, go to Content and click Perusall. Then click the Assignments tab in Perusall to access the course readings and raise comments, questions, and interact with other students' contributions. Students will be graded on their level of engagement with the text and their level of interaction with other students' posts. Perusall contributions are due by 9:00 a.m. the day the reading is scheduled. The lowest two Perusall grades will be dropped.

Case study paper and poster presentation

Students will partner with each other to write a case study (2-3 pp. each, single-spaced) related to course topics. Case studies will present a researched case narrative that demonstrates a critical selection of the morally relevant details of the case. Case studies will also include discussion questions that evoke critical analysis of the case. For guidelines on writing a case study, see Green and Raicu, "A Template for Technology Ethics Case Studies." Rubrics will be provided for case studies.

Case studies will be presented to the class as poster presentations for analysis and feedback for further revision. Then, students will individually write a 1-2 page (single-spaced) analysis of their case study for a separate grade.

The case studies will be given three grades, one group grade for the poster presentation, one group grade for the accompanying case study paper, and one for the individual student's analysis of their own case study. Case studies may be submitted to a case study archive for use in future classes and Viterbi-wide or public use.

Learning objectives of the case studies are to

- Deepen students' understanding of the complex ethical issues raised by the technology: the moral dilemmas, the multiple ways of framing the problems, and the interdisciplinarity of the problems;
- Develop students' ability to distinguish morally relevant details and to analyze an ethical issue;
- Allow students to examine in greater depth selected topics of the course;
- Develop students' written and oral communication skills.

Black Mirror sketch

Students will partner with each other to develop a plot sketch for a *Black Mirror* episode related to AI. Sketches will be presented to the class in a poster session format. Students will submit a co-designed “preview” slide, a co-written plot sketch, and a co-written “Light Mirror” news story. Students will also submit an individually written analysis of the ethical issues related to the episode.

Reflective essay

A 2-3 page (single-spaced) essay at the end of the term in which students reflect on their views on the course topics and the role of AI in society. The prompt for this essay will be open-ended, allowing students to reflect on what they have learned and how their views have been challenged, changed, or better informed by encountering new and multiple perspectives; or, they may reflect on their broad view of artificial intelligence and our response to it. Essays will be assessed on their originality and rigor of thought (i.e., their synthesis or critique of the ethical views presented in the course) and their clarity of expression.

Participation

This includes attendance, preparation of the readings and reading quizzes, and engagement in class discussions and activities.

Grade Breakdown

Each assignment will receive a specified number of points, for a total of 1000 points.

Perusall comments	250
Case study #1 team grade (poster presentation, 50 points; and paper, 100 points)	150
Case study #1 individual grade (analysis of case)	125
<i>Black Mirror</i> team grade (<i>Black Mirror</i> slide and newspaper article slide, 50 points; and episode sketch, 100 points)	150
<i>Black Mirror</i> individual grade (ethical analysis of episode)	125
Final reflective essay	100
Participation	100

Total	1000
-------	------

There is no curve in this course: an A is 94% and above, A- 90-93.5%, B+ 87-89.5% and so on.

Rough Draft Feedback Policy

I am available to provide feedback on rough drafts for the case study papers, Black Mirror sketches, and the final reflective essay. Due to time constraints, I will provide general feedback on entire rough drafts only in a conference. If you have specific questions about a paper, such as its introduction, support for a particular paragraph, or citation questions, I will answer those questions over email.

Paper Submissions and Late Submissions

Unless otherwise directed, papers will be submitted through Turnitin as PDFs. Formatting should adhere to the guidelines specified for the assignment.

Perusal comments are set to receive partial credit on a linearly declining scale for two weeks after their due date. At my discretion, other assignments will be penalized one plus- or minus-letter grade for each additional class period they are late. I am generally flexible about deadlines, so please contact me if you need a reasonable extension. Any paper not turned in will result in a 0. *It is therefore always in your best interest to turn in a paper, even if it is very late. I will always accept late papers.*

Academic Integrity and AI Policy

The purpose of this class—and college more broadly—is to help you develop critical thinking skills, the ability to think for yourself. As we enter the age of AI (if that's what we're doing), the ability to think for yourself and to express your ideas clearly in writing will become more important than ever. In your academic writing, you are engaging in a dialogue with other humans about major ideas. I urge you to engage in this dialogue with your own ideas and in your own voice. The writing process, while difficult and time-consuming, helps you refine and clarify your ideas. This class offers you the opportunity to practice and develop these essential critical thinking skills. You will not benefit from it if you rely on chatbots to generate ideas and write papers for you.

For these reasons, while we may be using chatbots from time to time in class, the use of natural language generators such as ChatGPT is prohibited at any stage of the writing process unless specifically directed by me.

Of course, other types of plagiarism are also prohibited. This includes submitting someone else's work—in whole or in part—as your own, submitting your own work completed for another class without my permission, falsifying information, failing to document sources properly, and

paraphrasing too closely to the original source material. Depending on the violation, you may fail the paper, fail the course, and risk suspension from the school. For an overview of the University’s policies concerning academic integrity, including what constitutes plagiarism, see <http://web-app.usc.edu/scampus/>.

Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with the Office of Student Accessibility Services (OSAS) each semester. A letter of verification for approved accommodations can be obtained from OSAS. Please be sure the letter is delivered to me as early in the semester as possible.

Required Texts and Materials

Ted Chiang, *Exhalation*. 2020. We’ll be reading a novella from this collection of stories, *The Lifecycle of Software Objects*.

Available at Amazon:

https://www.amazon.com/Exhalation-Ted-Chiang/dp/1101972084/ref=sr_1_1?crid=1TZDOLTRK P8JD&keywords=exhalation+by+ted+chiang&qid=1704491272&srefix=exhalation%2Caps%2C 174&sr=8-1

All other readings and materials can be found on Perusall or online.

Course Schedule (subject to change)

Complete the readings and Perusall comments by 11:00 a.m. the day they are being discussed. Bring access to the readings to refer to during class.

		<i>Part 1: What is artificial intelligence?</i> <i>Definitions, history, terms, technologies</i>
Week 1	M 8/26	Introduction to course How do large language models work?
	W 8/28	<i>History and Definitions: Defining AI</i> Readings: Aguera y Arcas and Norvig, “Artificial General Intelligence is Already Here” https://www.noemamag.com/artificial-general-intelligence-is-already-here/ Chomsky, Roberts and Watumull, “The False Promise of ChatGPT” https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatg

		<p>pt-ai.html</p> <p>Bender and Koller, “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data”</p> <p>And/or, Marcus, “Deep Learning is Hitting a Wall”</p>
Week 2	M 9/2	Labor Day holiday–no class
		<p><i>Part 2: Liberal democracy and AI</i> <i>How do we ensure AI helps advance well-being, autonomy, and justice?</i></p>
	W 9/4	<p><i>Value alignment</i></p> <p>Reading: O’Neil and Gunn, “Near-Term Artificial Intelligence and the Ethical Matrix”: pp. 1–11 (Sections 8.1 and 8.2)</p>
Week 3	M 9/9	<p><i>Predictive algorithms and fairness</i></p> <p>Reading: Angwin, “Machine Bias”</p>
	W 9/11	<p><i>Predictive algorithms and fairness, continued</i> Other cases</p>
Week 4	M 9/16	<p><i>Algorithmic opacity and trust</i> Guest speaker: Ulubilge Ulusoy (tentative)</p> <p>Reading: Burrell, “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms”</p>
	W 9/18	<p><i>Recommender systems and misinformation: polarization, content moderation, and Section 230</i></p> <p>Reading: Selections from Sunstein, <i>#Republic</i>, “The Public Forum”</p>
Week 5	M 9/23	<p><i>Deepfakes and misinformation: epistemic problems and harm</i></p> <p>Reading: Rini, “Deepfakes and the Epistemic Backstop”</p>
	W 9/25	<i>The Panopticon: Big data privacy and surveillance</i>

		Reading: Kantor and Sundaram, " The Rise of the Worker Productivity Score " (not on Perusall–no annotations, but do the scoring in the article)
Week 6	M 9/30	<i>Privacy, continued</i> Reading: Barocas and Nissenbaum, "Big Data's End Run Around Anonymity and Consent" Introduction to case study assignment
	W 10/2	<i>Prof. Warford travel–no class</i> Work on case studies
Week 7	M 10/7	<i>The responsibility gap: Self-driving cars</i> Reading: "Ethics, Safety and Autonomous Vehicles" How to write a case study
		<i>Part 3: AI and the human condition</i> <i>How do we ensure AI will help us live flourishing lives?</i>
	W 10/9	<i>Employment: A post-work world?</i> Reading: Thompson, "A World Without Work" In-class work on case studies: topics
Week 8	M 10/14	<i>Employment: The Mechanical Turk: Pulling back the curtain on "artificial artificial intelligence"</i> Reading: Williams, Miceli and Gebru, "The Exploited Labor Behind Artificial Intelligence" https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/
	W 10/16	<i>Employment: The perils of automation</i> Reading: Acemoglu, "Redirecting AI" FAQ from the future: How we solved the work problem

Week 9	M 10/21	<p>Case study poster presentations</p> <p>Case study papers due Friday, October 25, 11:59 p.m.</p> <p>Start reading “The Lifecycle of Software Objects.” As you read, think about the ethical questions it raises about how the humans treat the VR creatures and the commercial system under which they operate.</p>
	W 10/23	<p><i>Employment: The data market</i></p> <p>Reading: Arrieta-Ibarra et al., “Should We Treat Data as Labor? Moving Beyond ‘Free’”</p>
Week 10	M 10/28	<p><i>Relationships with AI: Are they valuable?</i></p> <p>Reading: Chiang, <i>The Lifecycle of Software Objects</i> Watch NYT doc, Opinion My A.I. Lover - The New York Times Replika</p>
	W 10/30	<p><i>Relationships with AI: Moral status</i></p> <p>Reading: Chiang, <i>The Lifecycle of Software Objects</i></p>
Week 11	M 11/4	<p><i>AI moral agents</i></p> <p>Reading: Nyholm, “Can a Machine Be a Moral Agent? Should a Machine Be a Moral Agent?”</p>
	W 11/6	<p><i>AGI/Superintelligence</i></p> <p>Reading: Floridi, “Singularitarians, Altheists, and Why the Problem with Artificial Intelligence is H.A.L., Not HAL”</p>
Week 12	M 11/11	Veterans’ Day holiday–no class
	W 11/13	<p>AGI/superintelligence continued</p> <p>The <i>Black Mirror</i> Writers’ Room</p>
Week 13	M 11/18	<p><i>Thanksgiving holiday–no class</i></p> <p><i>Control over technology: Is AI inevitable?</i></p>

		Lanier, "There is No A.I."
	W 11/20	<i>Control over technology: Responsibility of developers, whistleblower ethics</i> Reading: Thomson, Glover
Week 14	M 11/25	In-class work on <i>Black Mirror</i> Writers' Room project
	W 11/27	Thanksgiving holiday—no class
Week 15	M 12/2	In-class work on <i>Black Mirror</i> Writers' Room project
	W 12/4	<i>Black Mirror</i> pitches

Black Mirror slides and episode sketch, newspaper articles, and analyses due: Friday, April 26, 11:59 p.m. Submit on Blackboard

Final reflective essays due: Friday, May 3 11:59 p.m.