

PHIL143 Ethics of Artificial Intelligence

Summer Session II 2020 (Remote)

Instructor: Minji Jang (*she/they*: I personally use *she/her* pronouns but feel free to use *they/them* to address me!)

Email: jangmj@live.unc.edu

Class Meeting: M/T/W/Th/F 9:45am – 11:15am

Office Hours: M/W 11:30am to 12:30pm, and by appointment

Content Warning

Some of the materials that we will read and discuss in this course might be difficult for students from a variety of backgrounds. (These topics include but are not limited to: discussion of death, non-consensual sex, and police violence.) Detailed content warning will be given prior to relevant units/readings. If you have any concerns about any elements of the course, or if there is any way that we can together make things easier to bear, please contact me by email, by Zoom, or in person.

Textbook

There is no required textbook; all readings and videos will be made available on Sakai.

1. Course Description

In this course, we will discuss ethical questions brought to us by the emergence and inclusion of artificial intelligence (AI) in our lives. There are seven units:

Unit 1. Singularity and Superintelligence. We'll begin by exploring the idea of singularity and superintelligence. Will artificial intelligence surpass human intelligence? How likely is this outcome, and how serious is the threat?

Unit 2. Creating Accountable AI. We'll discuss the problem of algorithmic bias and AI accountability. What problems are generated from using biased algorithms, and what do we do about these problems? How can we create fair and accountable AI systems?

Unit 3. Putting AI into Practice. In this unit, we'll look at the current and proposed applications of AI systems in the following areas: predictive policing, autonomous vehicles, autonomous weapons, deepfakes, and virtual reality.

Unit 4. Identity. We'll examine the questions and concerns raised about our personal identity by the development of AI technologies (e.g., simulated reality and mind uploading).

Unit 5. Moral Machines. Can machines be moral? If so, what moral and legal rights and responsibilities should we attribute to these machines? We'll discuss the implications of treating machines as "artificial moral agents" and revisit the question of AI accountability.

Unit 6. Machine Companions. In our final unit, we'll discuss the interactions between humans and robots. Can robots replace human beings—as workers, care providers, friends, or lovers?

2. Learning Objectives

This course has four learning objectives:

- You'll be introduced to **key topics and ideas** in contemporary debates surrounding ethics of artificial intelligence, as well as ethical, social and legal issues arising from these topics.
- Through readings and assignments, you'll learn **the value of interdisciplinary scholarship** and the role of philosophical theorizing in it.
- You'll practice **key philosophical skills**, including: identifying and critically evaluating different perspectives on a complicated issue, forming your own views based on such evaluation, and clearly communicating your views in writing and in speech with interlocutors who may have different or opposing opinions while engaging in respectful dialogue.
- You'll reflect on how the topics that we've discussed in class relate to **issues outside of the classroom** and form your own views on those issues.

3. Commitment to Respect and Diversity

Together as a class, we're committed to creating a safe learning environment for everyone involved, irrespective of their ability, age, economic status, ethnicity, first language, gender expression and identity, national origin, race, religion, sex, and sexuality. Personal offenses or discriminatory remarks toward others on any of these bases will not be tolerated under any circumstances.

During the semester, we'll have *anonymous climate surveys* to ensure that everyone feels welcomed and respected during class discussions and lectures. If you have any concerns about the learning environment at any time during the semester, please do not hesitate to let me know.

4. Course Assignments and Assessment

The required assignments for this course are as follows:

Due Date	Assignment	Percentage
Daily	Attendance & Participation	10%
Daily	Reading Reactions	20%
Throughout the Semester	Four Reflection Journals	20% (5% each)
7/5 (Sun)	Paper 1 Outline (1 page)	1%
7/13 (Mon)	Paper 1 (2-3 pages)	9%
7/19 (Sun)	Paper 2 Outline (2 pages)	2%

7/25 (Sat)	Paper 2 (4-6 pages)	18%
7/27 (Mon)	Final Exam	20%

I am required to assign you at least 10 pages of written work, not including in-class examinations. Your two papers and four reflection journals will suffice to meet this requirement.

Attendance & Participation (10%)

You'll meet this requirement *in part* by attending class regularly. But you'll also need to regularly participate in discussion and in-class group activities to get full points. There will be alternative ways of earning participation points; we'll discuss a guideline for these assignments on the first day.

Cold-calling. In class, I'll occasionally "cold-call" you and ask if you'd like to contribute your thoughts. You can always feel free to "skip" a particular question, or "opt-out" of being cold-called at any point during the semester for a particular class, a unit, or the entire semester.

Mid-semester Estimate. In the mid-semester, I'll release an *estimate* of your participation grade—this does not count toward your overall grade and only serves to give you a sense of how you've been doing so far. If what you see is not the grade you expected or wanted, come talk to me and we'll discuss the ways to boost your participation grade!

Reading Reactions (20%)

Prior to each class meeting where there's a new required reading, you'll write a brief reaction (1-2 paragraphs) to the reading. On some days, I'll give you a question about the reading for you to answer in your reaction. On other days, I'll ask you to submit your own reaction that (1) briefly summarizes one of the main ideas, claims, or arguments in the reading, and (2) presents a question or a comment about this idea, claim, or argument.

How to submit. Include your reaction in the body of a message, not as an attachment and send it via Sakai message, with the subject line RR Last Name Date (e.g. RR Jang 6/22), by **11:59pm** the night before the class meeting.

There will be 23 reading reactions in total, which I'll grade *pass/fail*. I'll drop your lowest 3, and the remaining 20 will count toward your grade. You'll pass the reading reaction, if it clearly shows that you've carefully done the reading and made a serious attempt to engage with it. You'll fail the reading reaction, if it doesn't clearly show this attempt.

Reflection Journals (20%)

In addition to your two paper assignments, you'll submit four reflection journals throughout the semester. The goal of this assignment is to give you a chance to reflect on how the topics discussed in class relate to the issues that have direct impacts in *your* life and that *you* personally care about. The due dates are specified on the reading schedule. A detailed guideline with a prompt will be released for each of these journals.

Paper I (10%)

Before submitting your Paper 1, you'll submit a 1-page outline for your paper. The outline counts for a very small portion of your overall grade (1%), so it's a low-stakes opportunity to try out your ideas and make mistakes. The outline is due on Sunday, July 5th. The paper itself is due on Monday, July 13th and is worth 9% of your grade. I'll release a specific guideline for both the outline and the paper closer to the due date.

Paper II (20%)

You'll also submit an outline for your Paper 2. The goal is again to give you a chance to try out your idea for the paper. This outline should be about 2 pages in length and is worth 5% of your overall grade. It is due on Sunday, July 19th. The paper itself is due on Sunday, July 25th and is worth 18% of your overall grade. I'll release a specific guideline for both the outline and the paper closer to the due date.

Final Exam (25%)

There will be a cumulative final exam on Monday, July 27th. There will be a review session on Thursday, July 23rd. A detailed study guide and an exam guideline will be released closer to that date.

5. Grading and Other Class Policies

Each assignment will be marked out of the number of percentage points that it contributes to your overall grade. This will give you a total score of 100, which I'll convert to your final letter grade:

A	A-	B+	B	B-	C+	C	C-	D+	D	F
94	90	87	84	80	77	74	70	67	60	0-59

For the papers, I will also give you a corresponding letter grade for the papers. We will look at grading criteria closer to the due dates, but here's a rough description of what these letter grades signify¹:

A: “*Mastery* of course content at the *highest level* of attainment that can reasonably be expected of students at a given stage of development.”

B: “*Strong* performance demonstrating a *high level* of attainment for a student at a given stage of development.”

C: “A *totally acceptable* performance demonstrating an adequate level of attainment for a student at a given stage of development.”

D: “A *marginal* performance in the required exercises demonstrating a minimal passing level of attainment [at a given stage of development].”

¹ These descriptions are from UNC's official “Explanation of Grading System” (<https://registrar.unc.edu/academic-services/grades/explanation-of-grading-system/>).

F: “For whatever reason, an *unacceptable* performance [at a given stage of development].”

Anonymous Grading

I will grade the papers and the final exam *anonymously*. This means that you shouldn’t put any identifying information on them, other than your PID number. Make sure to include your PID number *both* in the body of your paper *and* in the file name.

Page Limit Policy

Paper 1 and 2 are subject to strict page limits. To standardize length, all submissions should be: 12pt font, Times New Roman, double-spaced, with 1-inch margins. If you have a nonstandard formatting, I’ll convert it to standard formatting to check for length.

*I will dock a third of a letter grade for each half of a page you are over or under the required page range: 2–3 pages for Paper 1; 4–6 pages for Paper 2 (excluding the reference list or bibliography). The outlines and reflection journals are **not** subject to the page limit policy.*

Due Dates and Late Policy

All assignments are due at **11:59pm** sharp on the due date and should be posted on Sakai’s Assignment tab in a Word (.docx) format. *A late assignment without an extension granted in advance will be docked a third of a letter grade for each calendar day that it is late (or part of a day).*

For **reading reactions**, I’ll only grant extensions under extraordinary circumstances. It’s because the whole point of this assignment is to complete it before class and write your reactions to the reading, not to the class discussion/lecture.

For **papers**, I’ll usually grant you an extension if you make a reasonable request *at least 72 hours in advance* of the deadline. The length of an extension will vary depending on the assignments.

Electronics

Electronics—laptops and tablets, but no phones—are allowed in class for accessing the electronic copies of the readings and note-taking. Please note that I reserve the right to rescind this permission if they become of a distraction. Whether you bring a laptop to class or not, you should bring paper and something to write with in order to complete in-class exercises and activities.

Plagiarism and the Honor Code

You are expected to abide by UNC’s Honor Code, and refrain from any kind of academic dishonesty, including cheating and plagiarism. At UNC, plagiarism is defined as “deliberate or reckless representation of another’s words, thoughts, or ideas as one’s own without attribution in connection with submission of academic work, whether graded or otherwise.” This includes *unintentionally* stealing another person’s work or failing to give credit to someone else’s idea *by accident*.

Just as you are bound by the Honor Code not to plagiarize, I am bound by it to report suspected cases of academic dishonesty of any kind to the Honor Court. If you are ever in any doubt as to whether something constitutes plagiarism, check with me first!

6. Resources

Accessibility Resources & Service

If you need special accommodations for class assignments, please register with the Accessibility Resources & Service Office (<https://ars.unc.edu/>). Once I receive an official request from the ARS Office, I'd be happy to make the arrangements.

Writing Center

To ensure the anonymous grading, I will not read the paper drafts. But you can always set up an appointment with one of the coaches at the UNC Writing Center (<https://writingcenter.unc.edu/>) to improve your paper at any stages of writing—from brainstorming, to polishing a full draft.

Counseling and Psychological Services

CAPS (<https://caps.unc.edu/>) is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, both for short-term and long-term needs.

Title IX Resources

Any students impacted by acts of discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, stalking, and related retaliation are encouraged to report the incident and seek resources on campus or in the community.

Please contact the Director of Title IX Compliance (Rebecca Gibson: rmgibson@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential: caps@unc.edu), or the Gender Violence Services Coordinators (confidential: gvsc@unc.edu) to discuss your specific needs. Additional resources are available at <https://safe.unc.edu/>.

7. Reading Schedule (Subject to change)

6/22 Mon – *Course Overview*

- Course Syllabus

Unit 1. Superintelligence and Singularity

6/23 Tue – *Singularity*

- David Chalmers, “Introduction” and “The Argument for a Singularity” in “The Singularity: A Philosophical Analysis” (2010)

6/24 Wed – *Superintelligence*

- *Ted talk*: Nick Bostrom, “What happens when our computers get smarter than we are?” (2015)
- Nick Bostrom, “The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents” (2012)

Unit 2. Creating Accountable AI

6/25 Thurs – *Biased Algorithms*

- *Ted talk*: Cathy O’Neil, “The era of blind faith in big data must end” (2017)
- Cass Sunstein, “Algorithms, Correcting Biases” (2019)
- Megan Garcia, “Racist in the Machine: The Disturbing Implications of Algorithmic Bias” (2017)

6/26 Fri – *Algorithmic Accountability*

- *Youtube*: “Facial Recognition: Last Week Tonight with John Oliver” (2020)
- *Ted talk*: Joy Buolamwini, “How I’m fighting bias in algorithms” (2016)
- Reuben Binns, “Algorithmic Accountability and Public Reason” (2017)

6/28 Sun – **Reflection Journal 1 due!**

Unit 3. Putting AI into Practice

6/29 Mon – *Predictive Policing*

- Martin Degeling and Bettina Berendt, “What is Wrong about Robocops as Consultants? A technology-centric Critique of Predictive Policing” (2018)
- *The Atlantic*: Ed Yong, “A Popular Algorithm is No Better at Predicting Crimes than Random People” (2018)

6/30 Tue – *Autonomous Vehicle*

- *Ted talk*: Iyad Rahwan, “What moral decisions should driverless cars make?” (2017)
- Patrick Lin, “Why Ethics Matters for Autonomous Cars” (2016)

7/1 Wed – *Autonomous Vehicle, cont'd*

- No new reading!

7/2 Thu – *Autonomous Weapon*

- Robert Sparrow, “Killer Robots” (2007)

7/3 Fri – *Independence Day!*

- No class / **Reflection Journal 2 due!**

7/5 Sun – **Paper 1 Outline due!**

7/6 Mon – **Paper Writing Workshop**

7/7 Tue – *Deepfakes*

- *NYT Opinion*: Regina Rini, “Deepfakes Are Coming. We Can No Longer Believe What We See” (2019)
- Regina Rini, “Deepfakes and the Epistemic Backstop” (2020)

7/8 Wed – *Surpassing Death*

- 10-minute documentary: *Meeting You* (2020)
- Felipe Pereira and Travis Timmerman, “The (un)desirability of immortality” (2020)

Unit 4. Identity

7/9 Thu – *Simulation*

- Jim Pryor, “What’s So Bad about Living in the Matrix?” (2005)
- *Black Mirror*: “San Junipero” (S3E4)

7/10 Fri – *Uploading and Identity*

- David Chalmers, “Uploading and Personal Identity” in *The Singularity: A Philosophical Analysis* (2010)

7/12 Sun – **Reflection Journal 3 due!**

7/13 Mon – *Uploading and Identity, cont'd*

- Susan Schneider, “Is Your Mind a Software Program?” in *Artificial You* (2019)
- **Paper 1 due!**

Unit 5. Moral Machines

7/14 Tue – *Artificial Moral Agents*

- *Youtube*: IEEE Spectrum, “How to Build a Moral Robot” (2016)
- Wendell Wallach and Colin Allen, excerpts from “Can (Ro)bots Really be Moral?” from *Moral Machines* (2008)

7/15 Wed – *Artificial Moral Agents, cont'd*

- Colin Allen, Gary Varner, and Jason Zinser, “Prolegomena to any future artificial moral agent” (2010)

7/16 Thu – *AI Rights*

- Erin Schwitzgebel and Mara Garza, “A defense of the rights of artificial intelligence” (2015)

Unit 6. Machine Companions

7/17 Fri – *Broadening the Circle of Concern*

- Sherry Turkle, excerpts from *Alone Together: Why We Expect More from Technology and Less from Each Other* (2011)
- Optional: Peter Singer, “All Animals are Equal” (1974)

7/19 Sun – **Paper 2 Outline due!**

7/20 Mon – *Love and Sex Robots*

- *Ted talk*: Kate Devlin, “Sex Robots” (2017)
- Sven Nyholm and Lily Frank, “From Sex Robots to Love Robots: Is Mutual Love with Robot Possible?” (2020)
- **Reflection Journal 4 due!**

7/21 Tue – *Care Robots*

- Robert Sparrow, “Robots in aged care: a dystopian future?” (2016)

7/22 Wed – *Wrapping up*

Exam Preparation

7/23 Thu – **Final Exam Review Session**

7/25 Fri – **Paper 2 due!**

7/27 Mon – **Final Exam** (8:00am-11:00am)