

PHIL 3370

Ethics of Artificial Intelligence

Fall 2023

Section 1

Department of Philosophy

Credit Weight: 0.50

## \*\*Please check Syllabus and Announcements on CourseLink for the most up-to-date information\*\*

**Schedule:** Tuesday/Thursday, 11:30AM-12:50 PM. Please consult Courselink for the classroom location.

**Instructor:** Please see Web Advisor for instructor information.

**E-mail:** Please see Web Advisor for instructor information.

**Office Hours:** By appointment on Teams or in-person (MCKN 328)

# Calendar Description

Artificial Intelligence (AI) increasingly touches all aspects of our lives, from our economy, healthcare systems, and food production down to our social interactions. This course will explore some of the real and possible implications of hand over decision-making power to machines. Some of the questions explored may include individual privacy, algorithmic bias, social inequality and AI, and the moral status, both in terms of agency and responsibility, of AI.

# Course Description

Emerging technologies powered by Artificial Intelligence, Machine Learning and Data Science increasingly influence significant aspects of our lives including where we go to school, who we date, where and how we work, who we talk to, what news we read, how we are entertained, and more. The philosophical discipline of Ethics is concerned (among other things) with the question: How should we live a good life? Through a careful study of some ethical traditions applied to contemporary problems raised by AI, we will explore whether and how we can live well in the midst of rapid technological advancement.

# Required Resources

Zerilli, J. (2021). *A citizen's guide to artificial intelligence*. MIT Press. ISBN: 9780262044813. Note, all of the individual chapters are available for free through the UofG library and are posted on CourseLink. But you can also buy a hardcopy or e-book if you prefer. Additional readings will be posted as .pdfs on CourseLink

# Learning Outcomes

* To provide students with time and space to step back from their everyday practices to reflect on the nature of their interactions with technology.
* To provide students with familiarity in theoretical and empirical work in the ethics of technology.
* To provide students with examples of, and practice in, sustained analytical reflection on theoretical and empirical questions about technology.
* To provide students with the presentation skills (both written and spoken) necessary to meaningfully engage across disciplinary boundaries in the sciences and humanities.

# Assessments

## Weekly Responses (30% of final grade)

Each week, you will submit a Weekly Response about the readings for the **upcoming** week. The content of the response is up to you, but all responses should deal **directly** with the reading assigned for that week. You may want to choose a sentence or paragraph you found especially provocative, difficult, or remarkable, then explain why you found it provocative, difficult, or remarkable. Alternatively, you may want to argue that one of the authors is right or wrong in making some particular claim. Or you may wish to connect two passages in the reading that illuminate each other. You might relate a central concept from the reading to recent or current events. You may even just pick a passage and ask detailed questions about it. Your responses will *directly* guide what we address in class, and I will sometimes quote from them, so please take them seriously. In general, the responses should demonstrate that you have made a good effort to engage with the readings, and that you have raised a question worthy of consideration and discussion with your peers. While there is no strict requirement for length, the responses should be at least a paragraph long but probably shorter than a full page. I will drop your two lowest scores.

## Mid-term Exam (30% of final grade)

The take-home midterm will be distributed at 9:00 AM, October 13; **DUE 5:00 PM, Monday, October 16.** Please plan accordingly.

## Final Research Paper (30% of final grade)

The culmination of your work in the class is a final research paper. You will need to incorporate at least one peer-reviewed source from outside of the assigned readings. Your research paper should demonstrate high-level engagement with themes from the course. The paper should not be a mere summary of other people’s writing, but should also reflect your unique contribution to debates in technology ethics. The paper should be around five pages (~1,000 words). Strategies for successful writing will be discussed in class. The due date is **Monday, December 11 by 11:59 PM.**

## Engagement (10% of final grade)

Good philosophy happens through discussion. It is therefore of the utmost importance that you attend class and participate. Excused absences will be granted for documented emergencies, or matters discussed with me AHEAD of time. In addition to showing up, you need to be prepared to participate. This includes showing up on time, actively speaking or listening, paying attention to what the other students have to say, taking notes, asking questions, etc. And of course it includes not sleeping, texting, snapchatting, emailing, tweeting, tik toking, yik yaking, etc.

# Schedule (subject to change)

\*always check CourseLink for most up-to-date information on scheduling, readings, assignments, etc.

## 1. Week 1: September 7 – Introduction to Technology Ethics

Readings

1. .PDF on Courselink: Vallor, *Technology and the Virtues*, Introduction & Ch. 1 (pp. 1-34)

## 2. Week 2: September 12 & 14 – What is AI?

Readings

1. Zerilli, Prologue and Ch. 1 (35 pgs.)
2. Browse some of the blog posts, “Philosophers on Next Generation Large Language Models”:

https://dailynous.com/2023/03/16/philosophers-on-next-generation-large-language-models/

Assignments

1. Weekly Response 1 by 11:59 PM on Sunday, September 10

## 3. Week 3: September 19 & 21 – Transparency & Bias

Readings:

1. Zerilli, Ch. 2 (22 pgs.)
2. Zerilli, Ch. 3 (18 pgs.)

Assignments

1. Weekly Response 2 by 11:59 PM on Sunday, September 17

## 4. Week 4: September 26 & 28 – Responsibility & Control

Readings:

1. Zerilli, Ch. 4 (18 pgs.)
2. Zerilli, Ch. 5 (14 pgs.)

Assignments:

1. Weekly Response 3 by 11:59 PM on Sunday, September 24

## Week 5: October 3 & 5 – Privacy & Surveillance; MIDTERM

Readings:

1. Zerilli Ch. 6 (14 pgs)
2. .PDF on Courselink: Amnesty International (2019) *Surveillance Giants: How the business model of Google and Facebook threatens human rights* (51 pages)

Assignments

1. Weekly Response 4 by 11:59 Sunday, October 1

## Week 6: October 12 – Autonomy & Government

Readings:

1. Zerilli Ch. 7 (20 pgs)
2. Zerilli Ch. 8 (22 pgs)

Assignments:

1. Weekly Response 5 by 11:59 Tuesday, October 10;
2. Midterm distributed at 9:00 AM, Friday, October 13; **DUE 5:00 PM, Monday, October 16**

**\*MIDTERM DUE 5:00 PM, Monday, October 16\***

## Week 7: October 17 & 19 – ChatGPT & LLMs

Readings:

1. .PDF on Courselink: Bowman, S.R. (2023). Eight things to known about large language models.
2. .PDF on Courselink: Chiang, T. (2023) ChatGPT is a blurry JPEG of the web. *The New Yorker*.

Assignments:

1. Weekly Response 6 due 11:59 PM on Sunday, October 15

## Week 8: October 24 & 27 – Deepfakes & Disinformation

Readings:

1. .PDF on Courselink: Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosopher’s*

*Imprint*, 1-16.

1. .PDF on Courselink: NATO Association of Canada (2019) - Disinformation and digital democracy (52 pgs.)

 \*\*we will focus on Rini’s essay “Social Media Disinformation and the Security Threat to Democratic Legitimacy” pp. 10-15, but feel free to read other articles of interest).

Assignments:

1. Weekly Response 7 due 11:59 PM on Sunday, October 22

## Week 9: October 31 & November 2 – Automation & Future of Work

Readings:

1. Zerilli, Ch. 9 (10 pgs)
2. .PDF on Courselink: Danaher, J. (2017). Will life be worth living in a world without work? Technological unemployment and the meaning of life. *Science and engineering ethics, 23*(1), 41- 64.

Assignments:

1. Weekly Response 8 due 11:59 PM on Sunday, October 29

## 10. Week 10: November 7 & 9 – Autonomous Vehicles

Readings:

1. .PDF on Courselink: The ethics of crashes with self-driving cars: A roadmap, 1 & 2. *Philosophy Compass.*
2. .PDF on Courselink: Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352* (6293), 1573-1576.

Assignments:

* 1. Weekly Response 9 due 11:59 PM on Sunday, November 5

## 11. Week 11: November 14 & 16 – Moral Consideration for AI Agents

Readings:

1. .PDF on Courselink: Heller, N. (2016). If animals have rights, should robots? *The New Yorker*.
2. .PDF on Courselink: Sebo, J., and Long, R. (2023). Moral consideration for AI systems by 2030.

Assignments:

* 1. Weekly Response 10 due 11:59 PM on Sunday, November 12

## 12. Week 12: November 21 & 23 – AI Safety & Existential Risk

Readings:

1. .PDF on Courselink: Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies, 17*(9), 7-65.
2. .PDF on Courselink: Yudkowsky, E. (2023). Pausing AI development isn’t enough. We need to shut it all down. *Time*.
3. .PDF on Courselink: Eisikovits, N. (2023). AI is an existential threat – just not the way you think. *The Conversation*.

Assignments:

* 1. Weekly Response 11 by 11:59 PM on Sunday, November 19

## 13. Week 13: November 28 & 30 – Oversight & Regulation

Readings:

1. Zerilli, Ch. 10 (16 pgs)
2. .PDF on Courselink. Chiang, T: (2023). Will A.I. become the new McKinsey? *The New Yorker*.

Assignments:

* 1. Weekly Response 12 due 11:59 PM on Sunday, November 26

 **Final Research Paper is due Monday, December 11 by 11:59 PM**

Late Policy

WEEKLY RESPONSES must be submitted on time unless an extension has been granted BEFORE the due date. Extensions will not be granted after the fact. Weekly responses that are not submitted on the due date when an extension has not been granted will receive zero for that week.

Similarly, the MIDTERM must be handed in on-time unless an extension has been granted.

FINAL RESEARCH PAPERS that are not submitted on the due date when an extension has not been granted will be penalized 3% for every day that it is late.

If you require an EXTENSION, you must communicate with me before the due date with reasonable timeliness. An extension requested hours before an assignment is due or over the weekend when I may not be available to respond are unlikely to be granted.

Large Language Model Policy

WEEKLY RESPONSES must be submitted without the help of large language models (LLMs) such as ChatGPT, Claude, etc. These responses are meant to reflect your personal thoughts and reflections about the reading, unmediated by further research.

Similarly, the MIDTERM must be submitted without the help of LLMs unless the question directly asks you to use ChatGPT or a similar tool to complete your answer. Successfully completing the midterm will depend on you regularly attending class and being privy to class discussions, something the LLM cannot do.

LLMs may be used help the process of writing your FINAL RESEARCH PAPERS. LLMs can be helpful tools for generating ideas, outlining papers, editing, etc. They CANNOT and SHOULD NOT be used to substitute the writing process but can be used to support the writing process. If you choose to use LLMs to help any aspect of writing your final papers, you must submit a 1-2 page document explaining your choice to use LLMs, how you used them, how they were helpful, and how they were unhelpful. This document will be a critical assessment of the pros and cons of using LLMs to write academic papers, as well as a reflection of your own personal thoughts and experiences with LLMs.

Generally speaking, my policy on LLMs is that their use must be justified, transparent, and documented. Throughout the course, we will explore the various uses and misuses of LLMs, examining what they are and may be helpful for, and what they are not. The course assessments are designed with this in mind. This means that if you are asked to compete an assignment without the aid of LLMs and you do, you are likely to perform poorly, whether or not you are found out.

# University Standard Statements

## Email Communication

As per university regulations, all students are required to check their <uoguelph.ca> e- mail account regularly: e-mail is the official route of communication between the University and its students.

## When You Cannot Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons, please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact.

Undergraduate Calendar - Academic Consideration and Appeals <https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml>

Graduate Calendar - Grounds for Academic Consideration <https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml>

## Drop Date

Courses that are one semester long must be dropped by the end of the last day of classes; two-semester courses must be dropped by the last day of classes in the second semester. The regulations and procedures for course registration are available in their respective Academic Calendars.

Undergraduate Calendar - Dropping Courses [https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08- drop.shtml](https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-%20drop.shtml)

Graduate Calendar - Registration Changes [https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg- regchg.shtml](https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-%20regchg.shtml)

## Copies of Out-Of-Class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments: you may be asked to resubmit work at any time.

## Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.

When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required, however, interim accommodations may be possible while that process is underway.

Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to make a booking at least 14 days in advance, and no later than November 1 (fall), March 1 (winter) or July 1 (summer).

Similarly, new or changed accommodations for online quizzes, tests and exams must be approved at least a week ahead of time.

More information: [www.uoguelph.ca/sas](http://www.uoguelph.ca/sas)

## Academic Misconduct

The University of Guelph is committed to upholding the highest standards of academic integrity and it is the responsibility of all members of the University community – faculty, staff, and students – to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff and students have the responsibility of supporting an environment that discourages misconduct. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

Undergraduate Calendar - Academic Misconduct [https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08- amisconduct.shtml](https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-%20amisconduct.shtml)

Graduate Calendar - Academic Misconduct <https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/index.shtml>

## Recording of Materials

Presentations which are made in relation to course work—including lectures—cannot be recorded or copied without the permission of the presenter, whether the instructor, a classmate or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

## Resources

The Academic Calendars are the source of information about the University of Guelph’s procedures, policies and regulations which apply to undergraduate, graduate and diploma programs.

## Disclaimer

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings, changes in classroom protocols, and academic schedules. Any such changes will be announced via Courselink and/or class email.

This includes on-campus scheduling during the semester, mid-terms and final examination schedules. All University-wide decisions will be posted on the COVID-19 website (https://news.uoguelph.ca/2019-novel-coronavirus-information/) and circulated by email.

## Illness

Medical notes will not normally be required for singular instances of academic consideration, although students may be required to provide supporting documentation for multiple missed assessments or when involving a large part of a course (e.g., final exam or major assignment).

## COVID-19 Safety Protocols

For information on current safety protocols, follow these links:

* [https://news.uoguelph.ca/return-to-campuses/how-u-of-g-is-preparing-for-your- safe-return/](https://news.uoguelph.ca/return-to-campuses/how-u-of-g-is-preparing-for-your-%20safe-return/)
* <https://news.uoguelph.ca/return-to-campuses/spaces/#ClassroomSpaces>

Please note, that these guidelines may be updated as required in response to evolving University, Public Health or government directives.