ARTIFICIAL INTELLIGENCE: PHILOSOPHICAL AND ETHICAL ISSUES (CGSC 3603 A/PHIL 3503 A)

Carleton University, Fall 2024

Prof. Myrto Mylopoulos

Office Hours: By appointment (over Zoom)

E-mail: myrto.mylopoulos@carleton.ca

Teaching Assistant: TBA

E-mail: TBA

Course Website: https://brightspace.carleton.ca/d2l/home/292784

Course Description
This course will examine a range of issues pertaining to the philosophy of artificial intelligence, including ethical issues, through a philosophical lens. Questions to be addressed include the following:
- Can artificial general intelligence ever be achieved?
- How likely are we to face a singularity event?
- What are the risks, both existential and non-existential, involved in building intelligent machines? And how should we handle them?
- Will "merging" with AI someday be possible?
- Could an AI ever be conscious or sentient?
- Might some present or future AIs deserve our moral consideration?
- Are AIs moral agents? Can they be said to be morally responsible?
- Is genuine friendship with AI possible?

Course Prerequisites: CGSC 2001 or PHIL 2501, and third-year standing in Cognitive Science or Philosophy.

Course Objectives
Our aim in this course will be to support and develop the following:
- An understanding of foundational philosophical issues in artificial intelligence.
- A familiarity with current debates in artificial intelligence from a philosophical perspective.
- An ability to critically examine, from an informed viewpoint, various ethical implications of advancements in AI.
- An ability to articulate a critical viewpoint on a philosophical issue clearly and effectively, both verbally and in writing.

Course Materials
- All required readings and additional resources will be made available on Brightspace.

Course Requirements

Course Component	Percentage of Final Grade
Reading Assignments (Perusall)	20%
In-Class Participation Exercises	15%
Writing Assignment #1	20%
Writing Assignment #2	20%
Group Project	25%

Reading Assignments
Readings will be assigned for every class session except the first and the last. All readings are required, except for those marked 'optional' on the course syllabus. While optional, you are strongly encouraged to do these readings, as they'll help both fill in the background for, and expand upon, the material we'll be discussing in class.

To work through the readings, we will be using Perusall, a collaborative e-reading platform that enables you to tackle the readings together as a group in between our class sessions. The reading assignments you complete on Perusall will be worth **20%** of your final grade. They are due the day before class **(Sundays) by midnight.** No late assignments will be accepted.

Reading assignments on Perusall will count towards your final grade from Week 2 through Week 12.

For instructions on how to register for and use Perusall, please see the *Guide to Perusall* posted under Week 1 on Brightspace.

In-Class Participation Exercises
These will take place during class starting in **Week 2**. They are aimed at giving you an in-class space to interact with your peers and an opportunity to further process and reflect on the issues we will be covering together. There are no make-ups for these exercises. In the case of illness or emergency preventing you from being in class on a given day, you may be excused from participating—please let me know promptly if this is the case.

Writing Assignments
Details TBA. We will go over instructions and a marking rubric in class closer to the due dates. We will also have a discussion surrounding the use of AI tools, such as ChatGPT, and what is permitted for these assignments.

Writing Assignment #1 (20%) (1500 words)
Due: Friday, October 18th by 11:59pm

Writing Assignment #2 (20%) (1500 words)
Due: Friday, December 6th by 11:59pm

These assignments will be submitted via Brightspace. Late submissions will be penalized 5% per day, including weekends, except in the case of illness, emergency, or bereavement.

Group Project

In groups of three, you will put together a presentation focused on one ethical issue related to AI. You will have the choice between creating a slideshow presentation or a podcast presentation for the class. Presentations will take place in class on **December 2** and **December 6**.

You will be provided with a list of topics to choose from.

There are two main components to the project: (i) Individual written component involving independent research, (ii) Group presentation component.

This project is worth 25% of your final grade. Further details TBA in class.

<u>**Course Topics and Reading Schedule**</u>

May be revised. Check Brightspace for most up-to-date information.

Week 1 (Sept. 9): Introduction
- No assigned readings.

Week 2 (Sept. 16): AI and Art (Guest Lecture: Mitchell Ross)
- Hertzmann (2018), "Can Computers Create Art?"
- Plaitas (2022), "State of the Arts: How Should Canadian Copyright Law Treat Works Generated by Artificial Intelligence?"

Week 3 (Sept. 23): Artificial Intelligence and Human Intelligence
- Millière & Buckner (2024), "A Philosophical Introduction to Language Models"
- Shevlin et al. (2019), "The Limits of Machine Intelligence"

Week 4 (Sept. 30): Super Intelligence and the Singularity
- Bostrom (20030 "When Machines Outsmart Humans"
- Chalmers (2010), "The Singularity: A Philosophical Analysis"

Week 5 (Oct. 7): AI and Existential Risk
- Russell (2021), "Human-Compatible Artificial Intelligence"
- Vold & Harris (2021), "How Does Artificial Intelligence Pose an Existential Risk?"

Week 6 (Oct. 14): NO CLASS (Thanksgiving)

Week 7 (Oct. 21): NO CLASS (Fall Break)

Week 8 (Oct. 28): "Merging" with AI

- Chalmers (2010), "Mind Uploading: A Philosophical Analysis"
- Turner & Schneider (2020), "Could You Merge with AI? Reflections on the
- Singularity and Radical Brain Enhancement"

Week 9 (Nov. 4): AI and Consciousness
- Shevlin (2020), "General Intelligence An Ecumenical Heuristic for Artificial Consciousness Research?"
- Schneider (2020), "How to Catch an AI Zombie"

Week 10 (Nov. 11): AI & Moral Patienthood
- Schwitzgebel & Garza (2015), "A Defense of the Rights of Artificial Intelligence"
- Danaher (2020), "Welcoming Robots into the Moral Circle"

Week 11 (Nov. 18): AI & Moral Agency
- Floridi & Sanders (2004), "On the Morality of Artificial Agents"
- Véliz (2021), "Moral Zombies: Why Algorithms Are Not Moral Agents"

Week 12 (Nov. 25): AI and Friendship
- Danaher (forthcoming), "The Philosophical Case for Robot Friendship"
- Bryson (2018), "Patiency is Not a Virtue: The Design of Intelligent Systems and Systems of Ethics"

Week 13 (Dec. 2): Group Presentations

Week 14 (Dec. 6): Group Presentations

Additional Information

In accordance with the Carleton University Undergraduate Calendar, the letter grades assigned in this course will have the following percentage equivalents:

A+ = 90-100	B+ = 77-79	C+ = 67-69	D+ = 57-59
A = 85-89	B = 73-76	C = 63-66	D = 53-56
A - = 80-84	B - = 70-72	C - = 60-62	D - = 50-52
F = Below 50			

Grades entered by Registrar:
WDN = Withdrawn from the course
DEF = Deferred

Statement on Academic Integrity

Carleton University is a community of scholars dedicated to teaching, learning and research. Academic integrity is the set of principles by which a community of scholars governs the pursuit of scholarly activity. The soundness and reliability of scholarly activity rests on a shared commitment to academic integrity based on principles of honesty, trust, respect, fairness, and responsibility.

Academic misconduct, in any form, is ultimately destructive to the values of the University. Conduct by any person that adversely affects academic integrity at Carleton University is a serious matter. The University requires, unequivocally, academic integrity from all of its members, including students. Misconduct in scholarly activity will not be tolerated.

Students should familiarize themselves with Carleton's Academic Integrity Policy, which can be found here: https://carleton.ca/secretariat/wp-content/uploads/Academic-Integrity-Policy-2021.pdf

For the purposes of this course, the following standards apply:

A. Students may not submit substantially the same piece of work more than once for academic credit. Here "substantially the same" means that the majority of the content is reproduced. If a student wants to submit an assignment with a minor portion of the content reproduced from another assignment of theirs, they must ask for my permission to do so first.

B. Unless indicated otherwise, students must work individually on their written assignments.

C. Use of AI Tools: The main planning and execution of your assignments—including the reading, writing, and thinking involved—must be done by you, and you are solely responsible/accountable for the content you produce.

You may **only** use generative AI tools (e.g., ChatGPT) to help you improve on the *clarity and organisation* of your own writing during the draft stages of the writing process, not for producing the content (i.e., definitions, arguments, main claims) itself.

You may also use AI tools (e.g., Grammarly) that help with the "mechanical" components of your writing, i.e., grammar and spelling.

D. Plagiarism: The University Academic Integrity Policy defines plagiarism as "*presenting, whether intentionally or not, the ideas, expression of ideas or work of others as one's own.*" This includes reproducing or paraphrasing portions of someone else's published or unpublished material, regardless of the source, and presenting these as one's own without proper citation or reference to the original source. Examples of sources from which the ideas, expressions of ideas or works of others may be drawn from include but are not limited to: books, articles, papers, literary compositions and phrases, performance compositions, chemical compounds, artworks, laboratory reports, research results, calculations and the results of calculations, diagrams, constructions, computer reports, computer code/software, material on the internet and/or conversations. Examples of plagiarism include, but are not limited to:

- any submission prepared in whole or in part, by someone else, including the unauthorized use of generative AI tools (e.g., ChatGPT);
- using ideas or direct, verbatim quotations, paraphrased material, algorithms, formulae, scientific or mathematical concepts, or ideas without appropriate acknowledgment in any academic assignment;
- using another's data or research findings without appropriate acknowledgement; and
- failing to acknowledge sources through the use of proper citations when using another's work and/or failing to use quotation marks.

Plagiarism is a serious offence that cannot be resolved directly by the course's instructor. The Associate Dean of the Faculty conducts a rigorous investigation, including an interview with the student, when an instructor suspects a piece of work has been plagiarized. Penalties are not trivial. They can include a final grade of "F" for the course.

Statement on Student Mental Health
As a University student you may experience a range of mental health challenges that significantly impact your academic success and overall well-being. If you need help, please speak to someone. There are numerous resources available both on- and off-campus to support you. Here is a list that may be helpful:

Emergency Resources (on and off campus): https://carleton.ca/health/emergencies-and-crisis/emergencynumbers/

Carleton Resources:
- Mental Health and Wellbeing: https://carleton.ca/wellness/
- Health & Counselling Services: https://carleton.ca/health/
- Paul Menton Centre: https://carleton.ca/pmc/
- Academic Advising Centre (AAC): https://carleton.ca/academicadvising/
- Centre for Student Academic Support (CSAS): https://carleton.ca/csas/
- Equity & Inclusivity Communities: https://carleton.ca/equity/

Off Campus Resources:
- Distress Centre of Ottawa and Region: (613) 238-3311 or TEXT: 343-306-5550, https://www.dcottawa.on.ca/
- Mental Health Crisis Service: (613) 722-6914, 1-866-996-0991, http://www.crisisline.ca/
- Empower Me: 1-844-741-6389, https://students.carleton.ca/services/empower-me-counselling-services/
- Good2Talk: 1-866-925-5454, https://good2talk.ca/
- The Walk-In Counselling Clinic: https://walkincounselling.com

Requests for Academic Accommodations

Carleton is committed to providing academic accessibility for all individuals. You may need special arrangements to meet your academic obligations during the term. The accommodation request processes, including information about the Academic Consideration Policy for Students in Medical and Other Extenuating Circumstances, are outlined on the Academic Accommodations website (students.carleton.ca/course-outline).

For an accommodation request the processes are as follows:

- **Pregnancy obligation**: Please write to me with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For accommodation regarding a formally-scheduled final exam, you must complete the Pregnancy Accommodation Form (click here).

- **Religious obligation:** Please write to me with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details click here.

- **Academic Accommodations for Students with Disabilities**: The Paul Menton Centre for Students with Disabilities (PMC) provides services to students with Learning Disabilities (LD), psychiatric/mental health disabilities, Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorders (ASD), chronic medical conditions, and impairments in mobility, hearing, and vision. If you have a disability requiring academic accommodations in this course, please contact PMC at 613-520-6608 or pmc@carleton.ca for a formal evaluation. If you are already registered with the PMC, contact your PMC coordinator to send me your Letter of Accommodation at the beginning of the term, and no later than two weeks before the first in-class scheduled test or exam requiring accommodation (if applicable). After requesting accommodation from PMC, meet with me to ensure accommodation arrangements are made. Please consult the PMC website for the deadline to request accommodations for the formally-scheduled exam (if applicable).

- **Survivors of Sexual Violence:** As a community, Carleton University is committed to maintaining a positive learning, working and living environment where sexual violence will not be tolerated, and where survivors are supported through academic

accommodations as per Carleton's Sexual Violence Policy. For more information about the services available at the university and to obtain information about sexual violence and/or support, visit: https://carleton.ca/equity/sexual-assault-support-services

- **Accommodation for Student Activities:** Carleton University recognizes the substantial benefits, both to the individual student and for the university, that result from a student participating in activities beyond the classroom experience. Reasonable accommodation will be provided to students who compete or perform at the national or international level. Write to me with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. https://carleton.ca/senate/wpcontent/uploads/Accommodation-for-Student-Activities-1.pdf

Important Information
- Students must always retain a copy of all work that is submitted.
- Standing in a course is determined by the course instructor subject to the approval of the Faculty Dean. This means that grades submitted by the instructor may be subject to revision. No grades are final until they have been approved by the Dean.
- For us to respond to your emails, we need to see your full name, CU ID, and the email must be written from your valid CARLETON address. Therefore, in order to respond to your inquiries, please send all email from your Carleton CMail account. If you do not have or have yet to activate this account, you may wish to do so by visiting http://carleton.ca/ccs/students/
- November 15th, 2024: Last day for academic withdrawal from full fall and late fall courses.
- For a list of dates and deadlines, including holidays and exam dates, please visit: https://calendar.carleton.ca/academicyear/