Philosophy 9232/9234B: Ethical and Societal Implications of AI

Tuesdays 9:30 - 11:30, in-person, STVH 1145.

Michael Barnes and Mike Katchabaw

- michael.barnes@uwo.ca and mkatchab@uwo.ca
- Office Hours: Posted in Calendar

Prerequisites: Registration in the Collaborative Specialization in Artificial Intelligence, Masters of Data Analytics Specialty in Artificial Intelligence, a graduate program in the Department of Philosophy, or permission of the Department.

General Description

In examining the topics and questions listed below, students will be expected to achieve the following learning objectives:

- Develop an understanding of basic ethical issues associated with the use of new technologies.
- Develop the ability to reflect on AI and data science from a critical, moral point of view.
- Develop critical and analytical skills to formulate and assess arguments, and to communicate clearly verbally and in writing, regarding the broader implications of AI.

Topics and Questions

It's become something of a cliché to say that artificial intelligence (AI) is ubiquitous and has the potential to radically alter the way we live our lives. Nonetheless, these claims are for the most part true. AI is everywhere and will change everything.

This rapid expansion necessitates an ethical assessment. Ethical questions arise when we consider applications of AI to almost any sphere of life, in ways that sometimes are immediately clear and, at other times, unobvious and unexpected.

This course will cover some of the major themes in ethics of AI, robotics, and automation, concerning both present-day and near- and long-term future uses of these technologies. We will examine what sorts of social impact they are likely to have and evaluate this impact with the tools of ethical theory, applied ethics, and a smattering of political philosophy.

The readings will feature a variety of perspectives from a range of professionals: philosophers, AI researchers, social scientists, lawyers, and tech journalists, among others.

Topics may include:

The ethics of algorithms: opacity, explanation, persuasion, bias, and trust

Some of the most popular and powerful AI techniques today involve the use of machine learning algorithms. These algorithms sometimes work in ways that are opaque and impossible to understand even to their creators. Is this a problem? Should AI be made explicable to an average user? What about algorithms that produce outputs that appear biased? How should they be evaluated? Could algorithms be trusted?

Professional ethics for AI practitioners – voluntary and state solutions

What ethical rules should AI professionals abide by in their work? What are the various AI ethics codes developed by a number of leading tech companies? Are they adequate? Should there be greater involvement by government agencies in deciding how tech companies are run?

Automation and labor

AI promises to revolutionize the labor markets across the world. Is this a desirable development? Will robots take all our jobs? Should we be preparing for a life without work? How should we organize our societies' response to automation?

Autonomous vehicles and autonomous weapons
What does it mean for weapons or vehicles to be "autonomous"? Are such technologies desirable? If they make their decisions without human involvement, how should they be programmed? Who's morally and legally responsible when these machines do something harmful?

Robot rights?
Will it ever be possible to create machines that think and feel just as human beings do? If so, should they be created? How should they be treated? Should they be given rights or remain subservient to human beings?

Superintelligence
What is superintelligence? What is the likelihood that superintelligent machines will be developed? What are potential impacts of superintelligence? How should we prepare?

Course Website and Readings

Assigned readings, supplementary readings, updated schedules, and commentaries will be posted on the website. This website is hosted under OWL and will be available to all students registered in the course.

Evaluation

For students registered in Philosophy 9232B, the course is evaluated on a pass/fail basis. To receive a pass in this course, students must actively participate in weekly discussions in the course, preparing each week by completing the assigned readings, podcasts, and videos and by writing position statements collecting your thoughts and views on the assigned materials. Further details are provided below:

- Participation: A key goal of the course is to have lively and well-informed discussions in class. To fulfill this goal, please come to class prepared to contribute, based on careful reading and reflection on the topics raised in the assigned readings. Your contributions should be thoughtful and productive, reflecting insights gained through your preparation, and delivered in a professional and courteous fashion.
- Preparation: You will be asked to write a series of short position papers (300 − 500 words each) regarding each week's topic. We will state the topic each week in class and on the course website, and provide appropriate materials covering the topic. Your papers will normally be submitted online through the OWL submission system by 5:00 p.m. Wednesday prior to class, with any exceptions noted in class and/or through the course website. We will typically ask you to argue for a specific position related to the topic we will discuss. The aim of the assignments is to enhance students' ability to engage in debates regarding ethical, societal, and policy issues, by properly identifying the central issues in the debate and developing clear, persuasive arguments. Topics will be assigned for each week of the class, and no lateness will be accepted.

For students registered in Philosophy 9234B, the course is graded. In addition to actively participating as noted above for students in Philosophy 9232B, students in Philosophy 9234B will also be required to complete and submit a term paper by the end of the course. Details will be provided through OWL.

Provisional Schedule

Week 1 Jan 12: Introduction: what is ethics? What is AI? (Barnes /Katchabaw)

Week 2 Jan 19: Opacity (Katchabaw)

Week 3 Jan 26: AI and trust (Barnes)

Week 4 Feb 2: Deepfakes (Barnes)

Week 5 Feb 9: AI, social media, and public discourse (Katchabaw)

Week 6 Feb 16: Reading break

Week 7 Feb 23: Algorithmic Bias (Barnes)

Week 8 March 2: Certification of AI professionals (Katchabaw)

Week 9 Mar 9: AI governance and AI policy (Barnes)

Week 10 Mar 16: Autonomous weapons & vehicles (Barnes)

Week 11 Mar 23: Automation (Katchabaw)

Week 12 Mar 30: AI rights? (Barnes)

Week 13 April 6: Superintelligence/wrap-up (B&K)

Ethical Conduct

Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site: https://www.uwo.ca/univsec/pdf/academic_policies/appeals/scholastic_discipline_grad.pdf.

Plagiarism: Students must write their essays and assignments in their own words. Whenever students take an idea, or a passage from another author, they must acknowledge their debt both by using quotation marks where appropriate and by proper referencing such as footnotes or citations. Plagiarism is a major academic offence. Please note, however, that students are not allowed to make use of the work of others unless explicitly instructed to do so in the description of an assignment.

The University of Western Ontario uses software for plagiarism checking. Students may be required to submit their written work and programs in electronic form for plagiarism checking.

All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (http://www.turnitin.com/).

Accessibility Statement

Please contact the course instructor if you require lecture or printed material in an alternate format or if any other arrangements can make this course more accessible to you. You may also wish to contact Student Accessibility Services (SAS) at 661-2147 if you have any questions regarding accommodations.

The policy on Accommodation for Students with Disabilities can be found here: https://www.uwo.ca/univsec/pdf/academic_policies/appeals/Academic%20Accommodation_disabilities.pdf.

Support Services

Learning-skills counsellors at the Student Development Centre (http://www.sdc.uwo.ca) are ready to help you improve your learning skills. They offer presentations on strategies for improving time management, multiple-choice exam preparation/writing, textbook reading, and more. Individual support is offered throughout the Fall/Winter terms in the drop-in Learning Help Centre, and year-round through individual counseling.

Students who are in emotional/mental distress should refer to Health and Wellness (https://www.uwo.ca/health) for a complete list of options about how to obtain help.

Additional student-run support services are offered by the USC, http://westernusc.ca/your-services.

The website for Registrarial Services is http://www.registrar.uwo.ca.

The policy on Accommodation for Religious Holidays can be found here: http://www.uwo.ca/univsec/pdf/academic_policies/appeals/accommodation_religious.pdf.

Tutoring

The role of tutoring is to help students understand course material. Tutors should not write assignments or take-home tests for the students who hire them. Having employed the same tutor as another student is not a legitimate defense against an accusation of collusion, should two students hand in assignments judged similar beyond the possibility of coincidence.