Philosophy 9234: Ethical & Societal Implications of A.I.
(STATS 9940 / ECE 9660 / COMPSCI 9147)

Thursdays 11:30 - 2:30, in-person
Room: Will be posted in OWL

Instructor: Andrew Buzzell

abuzzel2@uwo.ca
Office Hours: Will be posted in OWL

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

The recognition that new technology can create ethical challenges and tensions is not new, but many think that there is something truly novel in the ethical and social impacts of Artificial Intelligence. The rapid and increasing prevalence of AI has brought with it a surge of interest in the ethical dimensions of this technology. Corporations laud new technology as ethical, publish articles on their ethical AI strategies, and even hire ethicists, while others complain that ethics has become an obstacle to progress. Academic and public discourse has increasingly attended to questions about ethical AI and technology, often drawing attention to cases where the consequences of AI technology have been harmful. Those who make, use, and interact with AI systems increasingly face questions about the extent to which these can be said to be ethical. How can we think productively about these questions?

This course will introduce some general ethical concepts and theories, situate them in context of applied technology ethics, and explore their application to current and emerging concerns about AI. There will be an emphasis on thinking through and discussing particular cases and application domains. Through these discussions we will clarify several distinctive features of AI – including opacity, sociotechnical complexity, the inability of an AI user to understand how it has reached a result, and the potential for autonomous decision-making – and consider their implications for the use of AI tools in a variety of settings. We will further consider governance of AI, such as whether tech companies or professional societies have developed adequate ethical guidelines, and how and whether government should develop novel regulatory approaches for AI. The readings will feature a variety of perspectives from a range of professionals: philosophers, AI researchers, social scientists, lawyers, and tech journalists, among others.

The course will be divided into roughly two halves - the first will provide an overview of the ethics of AI, the second will consider a series of case studies exploring the impact of AI in specific fields, with a number of guest lecturers.

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 analytical capacity to survey and assess technology in terms of impacts, risks, stakeholders, and relevant norms and values
- Develop critical and analytical skills to formulate and assess arguments, and to communicate clearly verbally and in writing, and collaborate with others to conduct ethical analysis of technology.

Course Website and Readings

Core and optional readings, as well as schedules, notices and other course communication will be hosted on OWL,and will be available to all students registered in the course.

Evaluation

The majority of students in the course are evaluated based on pass/fail basis. There is an option primarily intended for Philosophy graduate students to take the course for a percentage grade (see below).

Evaluation

10% Participation - AI and technology ethics are evolving and emerging issues that we confront in our personal and professional lives. The capacity to engage meaningfully with others to identify, analyze, and discuss ethical issues with technology and the ways in which we might respond to them is important. Ensure you have carefully read the assigned material and be prepared to contribute to in-person and online class discussion in a thoughtful courteous, professional, and polite manner. While students are expected to participate in in-class conversations, it is understood that not all students will be able to do so, especially given the class size. There will be a forum in OWL, and each week there will be a new thread devoted to continuing class discussions. Anyone who hasn't participated in the live discussion, either due to missing class or for some other reason, should instead participate in these.

40% Preparation - For the first part of the course, you will be asked to prepare for each meeting by writing short position papers (300 - 500 words each) regarding each week's topic. I 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. (Any exceptions to this schedule will be noted in class and/or through the course website.) I 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 late submissions will be accepted. For the second part of the course, the preparation for each week will vary, with the appropriate format determined in consultation with the guest lecturer. In some cases, for example, I will ask you to

form groups to analyze and assess a case study, as preparation for in class discussion. No late submissions will be accepted, but the two lowest scores will be dropped.

50% Group Projects - Most real-world applied ethics happens in collaborative contexts. You will prepare, in assigned groups of 4-5, a case study regarding legal and ethical implications of a particular use of AI technology. The groups will include students from at least two different disciplinary backgrounds. Meeting and discussion time will be scheduled during regular class hours. The assignment will involve clearly describing the AI technology and its implementation; identifying the relevant ethical issues; and considering how current legal and regulatory frameworks apply. (Details regarding the case study and how it will be evaluated will be distributed later in the term.)

For philosophy students or others who would like to receive a percentage grade, a seminar paper will take the place of the group project. The evaluation will then be: 10% participation, 20% preparation, and 70% seminar paper. Please contact me if you wish to choose this option.

Provisional Schedule

Week 1 - (January 11) - Introductory discussion of ethics and AI and case study discussion of ChatGPT

Part 1: Overview of Ethics of AI

Week 2 - (January 18) - Algorithmic Opacity: Are AI tools distinctively opaque, by contrast with other technologies; and what are the implications of this opacity?

Week 3 - (January 25) - Algorithmic Bias: What are the different technical sources of algorithmic bias? How can we mitigate harmful effects, and to what extent can we tell when we've done so effectively?

Week 4 - (February 1) - Trust and Interpretability: When are we justified in "trusting" an AI system? How does this notion of trust compare to trust in human relationships, and should we appeal to this idea of trust or fall back on something more like reliability?

Week 5 - (February 8) - Regulating Emerging Technologies such as Generative AI: New technologies often create "policy vacuums": they make a new kind of action possible that falls outside, or challenges, existing legal and regulatory frameworks.

Week 6 - (February 15) - Auditing Algorithms: Automated decision making is becoming more prevalent in a variety of institutions; how can we effectively audit the performance of these algorithms to assess / mitigate negative social impacts? How should the use of machine learning and automated decision making be regulated? How can we mitigate or prevent biased outcomes at a practical level?

Reading week no class (February 22)

Part 2: Impact of AI in Different Fields

Week 7 - (February 29) - AI and Social Media and/or Guest Lecture

Week 8 - (March 7) - Guest Lecture: AI in Hiring

Week 9 - (March 14) - Guest Lecture TBA

Week 10 - (March 21) - Guest Lecture: AI Ethics Leadership in Private and Public Institutions

Week 11 - (March 28) - Guest Lecture: AI and Intellectual Property

Week 12 - (April 4) - Group presentations

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/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.