

PHIL 3350A: PHILOSOPHY, ETHICS AND PUBLIC AFFAIRS

THE ETHICS AND GOVERNANCE OF ROBOTICS AND ARTIFICIAL INTELLIGENCE

FALL 2016

INSTRUCTOR INFORMATION

Prof: Jason Millar

Email: jason.millar@carleton.ca

Office: Paterson 3A54

Office Hours: Wednesdays 3:00-4:00

Contacting Me (your prof): I encourage you to contact either myself, or the course TAs, with any questions or concerns. The best way to get in touch with me is email. **Be sure to include the following text in ALL emails or I cannot guarantee your email will get through my filters: "PHIL 3350"**.

COURSE INFORMATION

Time: Wednesdays, 6:05-8:55pm

Location: Southam 316

Course Website: I'll be using cuLearn for this course. You should check it regularly (more than once a week) for important messages, scheduling information, readings, assignment or exam information, and other important information. You will be responsible for keeping current with the information on cuLearn.

COURSE DESCRIPTION

Recent headlines are awash with warnings of the existential threat that Robotics and Artificial Intelligence (AI) pose to humanity. Stephen Hawking, Bill Gates, and Elon Musk are among the most prominent technology and robotics experts that have publicly warned of such a threat. In a recent open letter more than 1,000 robotics/AI experts called for a ban on offensive autonomous weapons (drones), more commonly termed "killer robots", arguing that their development would lead to a new robot arms race with potentially dire consequences for humanity. The recent explosion in robotics/AI research and development underscores their concerns. Global investment in robotics/AI is growing rapidly in all areas of the economy including: healthcare, manufacturing, consumer electronics, entertainment, military, transportation, and public safety. MIT roboticist Cynthia Breazeal recently broke crowdfunding records with the announcement of her new "humanized" social robot, Jibo, which she is calling "the world's first family robot". Google and most major automobile manufacturers are designing autonomous (driverless) cars, some of which are already sharing the roads with us. No longer science fiction or the stuff of speculation, the robots are here. But a future in which we find ourselves living alongside robots/AIs raises new ethical uncertainties and governance issues ripe for investigation.

Robots are challenging many of our philosophical and ethical notions while suggesting the need for new public policy. Recent revelations regarding the United States' military drone program has prompted widespread international discussions on "robot ethics" as a means of anticipating the ethical challenges

associated with pursuing various robot technologies and deploying them in, and across, societies. A recent international Campaign to Stop Killer Robots was launched with wide press coverage following the release of a UN report on lethal autonomous robots. Academic institutions worldwide have held workshops (e.g. robots.law.miami.edu) and are publishing more and more on the social implications of robotics. It is an exciting time to study the ethics and governance of robotics and AI!

What is a robot? What role does design ethics play in the making of robots and of society? Can robots be persons? What would that agency look like? What are the current directions in robot ethics, and robot design, and are they adequate/justifiable? What kinds of relationships ought we to encourage between robots and humans, or robots and robots? Should robots kill? What norms, if any, ought to be considered with respect to governing the design of, and interaction with, robots? What might we owe robots in the future?

The goal of this course is to investigate the above questions through an examination of classic and contemporary texts in ethics, philosophy of mind, science and technology studies, human-robot interaction, law, public policy, and robot ethics. Students will be challenged to consider how philosophical concepts including identity, consciousness, moral agency, personhood, artificial intelligence, and rationality apply to robots and artificial intelligence. As this course focuses on ethics and policy issues, an emphasis will be placed on examining and developing normative claims for use in policy contexts from the course literature. By engaging public policy documents students will also have an opportunity to examine and critique the application of philosophical concepts in the context of interdisciplinary technology debates.

COURSE LEARNING OBJECTIVES

Upon successful completion of this course, students will be able to:

1. Explain some of the key issues in the ethics and governance of robotics and artificial intelligence.
2. Critique course readings and the arguments they contain using clear argumentation.
3. Justify particular ethical positions using clear argumentation.
4. Compare course readings with reference to key issues in robot ethics.
5. Compare policy options in order to formulate policy recommendations.

COURSE TEXT

Required: Leiber, J. (1985). *Can Animals and Machines Be Persons? - A Dialogue*. (Hackett). Available at the campus bookstore.

Other **required** course readings (listed in the Class Schedule below) will be posted and available on cuLearn, either in document form (e.x. PDF), or via hyperlink.

PEDAGOGY

This course does not require any previous training in technology, robotics, or philosophy. The course will consist of a combination of interactive lectures and regular in-class activities and discussions. In addition to the readings a variety of sources will be engaged throughout the course including sci-fi literature, movies, print media and television for framing discussions. Students will be encouraged to draw on all sources to participate in regular in-class discussions and are expected to have completed all readings prior to class.

My goal is to create a highly interactive learning environment. Formal lectures will be discouraged in place of discussions and in-class peer teaching/learning activities such as discussions, debates, and very short presentations. The course has been designed to provide you opportunities to engage the readings critically on a regular basis, in order to get the most out of the course.

STUDENT EVALUATION

- 1) **One-page Argument Exposition** (2 x 20% of overall course grade): Students will be required to provide a one-page argument exposition of one of the course readings. Details to be distributed in class.
- 2) **Policy Brief** (25% of overall course grade): Students will be required to submit a written policy brief. Details will be distributed in class.
- 3) **Final Take-Home Exam** (35% of overall course grade): Details will be distributed in class.

NOTE: Assignment due dates/times and any associated late penalties will be discussed in class and specified on assignment instructions sheets distributed on cuLearn.

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.

CLASS SCHEDULE:

Sept 7: Setting the Scene: An Intro to Robot Ethics

1. Class Introduction: Course Syllabus and Reading List
2. Lin, P., Abney, K., Bekey, G. (2011). "Robot Ethics: Mapping the Issue for a Mechanized World." *Artificial Intelligence* 175(5-6):942-949.

Sept 14: What's So Special About Robots?

3. Calo, R. (2015). "Robotics and the Lessons of Cyberlaw." *California Law Review* 103. Online: http://papers.ssm.com/sol3/papers.cfm?abstract_id=2402972
4. Calo. Video Presentation from *We Robot 2014*.

Sept 21: The Imitation Game

5. Turing, A.M. (1950). "Computing Machinery and Intelligence." *Mind* 59:433-460. Available at: <http://loebner.net/Prize/TuringArticle.html>
6. Searle, J.R. (1980). "Minds, Brains, and Programs." *The Behavioral and Brain Sciences* 3:417-457. Online at <http://journals.cambridge.org/action/displayFulltext?type=1&fid=6573588&jid=BBS&volumeld=3&issuelid=03&aid=6573580>

Sept 28: Social Robots I - Anthropomorphism

7. **One-page argument exposition #1 due by midnight (hand in on cuLearn).**
8. Jibo: Jibo.com
9. Duffy, B.R. (2003). "Anthropomorphism and the Social Robot." *Robotics and Autonomous Systems* 42:177-190.
10. Darling, K. (2015). "'Who's Johnny?' Anthropomorphic Framing In Human-Robot Interaction, Integration, and Policy." *We Robot 2015 Proceedings*. Online: http://papers.ssm.com/sol3/papers.cfm?abstract_id=2588669

Oct 5: Social Robots II – Personhood

11. Leiber, J. (1985). *Can Animals and Machines Be Persons? – A Dialogue*. (Hackett Publishing Company).

Oct 12: Social Robots III – Robot Rights

12. Petersen, S. "Designing People to Serve." In P. Lin, G. Bekey, and K. Abney (Eds.). *Robot Ethics: The Ethical and Social Implications of Robotics*. Available at: <http://stevepetersen.net/professional/petersen-designing-people.pdf>
13. Darling, K. (2015). "Extending Legal Protection to Social Robots." In R. Calo, M. Froomkin, I. Kerr (eds.) *Robot Law*. (Northampton: Edward Elgar).

Oct 19: Social Robots IV – Gender

14. Dattaro, L. (2015). "Bot Looks Like A Lady: Should Robots Have Gender?" *Slate.com* (Feb 4). Online: http://www.slate.com/articles/technology/future_tense/2015/02/robot_gender_is_it_bad_for_human_women_single.html
15. Carpenter et al. (2009). "Gender Representation and Humanoid Robots Designed for Domestic Use." *International Journal of Social Robotics* 1(3):261-265.

Oct 26: Fall Break – No Class

Nov 2: Social Robots V - Love and Sex with Robots

16. **One-page argument exposition #2 due by midnight (hand in on cuLearn).**
17. Levy, D. Excerpt from *Love and Sex With Robots*
18. Gutiu, S. (2015). "The Roboticization of Consent." In R. Calo, M. Froomkin, I. Kerr (eds.) *Robot Law*. (Northampton: Edward Elgar).

Nov 9: Movie Night

19. Movie TBD.

Nov 16: Autonomous Cars I

20. Millar, J. (2015). "Technology as Moral Proxy: Autonomy and Paternalism by Design." *IEEE Technology & Society* 34(2):47-55.
21. Lin, Patrick. (2013). "The Ethics of Autonomous Cars." *The Atlantic*, October 8. <http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>.

Nov. 23: Autonomous Cars II

22. Millar, J. (forthcoming). "Ethics Settings for Autonomous Vehicles." In P. Lin, K. Abney & R. Jenkins (Eds.) *Robot Ethics 2.0*. (Oxford: Oxford University Press).
23. Lin, P. (2014). "Here's a Terrible Idea: Robot Cars With Adjustable Ethics Settings." *WIRED*. (Aug 18). Online: <http://www.wired.com/2014/08/heres-a-terrible-idea-robot-cars-with-adjustable-ethics-settings/>
24. Millar, J. (2014). "You Should Have a Say in Your Robot Car's Code of Ethics." *WIRED*. (Sept 2). Online: <http://www.wired.com/2014/09/set-the-ethics-robot-car/>

Nov. 30: Military Robots I – Setting the Scene

25. CBC. (2011). *Remote Control War*. Doc Zone Documentary.
26. International Humanitarian Law.
27. Human Rights Watch and ICRC. (2014). *Advancing the Debate on Killer Robots: 12 Key Arguments for A Preemptive Ban on Fully Autonomous Weapons*. (Human Rights Watch). Online: http://www.hrw.org/sites/default/files/related_material/Advancing%20the%20Debate_final.pdf

Dec 7: Military Robots II – Meaningful Human Control

28. **Policy Brief due by midnight (hand in on cuLearn)**
29. Article 36. (2015). *Killing By Machine*.
30. Horowitz & Scharre. (2015). *Meaningful Human Control in Weapons Systems*.
31. **Final Take-Home Exam Distributed**

Department of Philosophy and Carleton University Policies (2016-17)

Assignments:

Unless specifically told otherwise by their instructors, students:

- must not use a plastic or cardboard cover or paper clips
- must staple the paper (there is a stapler on the essay box)
- must include the following:
 - student name
 - student number
 - course number and section
 - instructor's name
- No assignments will be accepted after the last day for handing in term work – see dates in next column.
- Assignments handed in through the essay box (just inside the glass doors, Paterson Hall, Floor 3A) must be dropped into the box by 4:15 on a regular business day in order to be date-stamped with that day's date. Assignments handed in after 4:15 or on a non-business day will be stamped as having been handed in on the next business day.
- Students are required to keep copies of their assignments. If your paper is lost at any point, you will be considered not to have submitted it if you cannot produce a copy immediately on request.

Deferrals for Term Work:

If you miss a final examination and/or fail to submit a final assignment by the due date because of circumstances beyond your control, you may apply for a deferral of examination/assignment. For deferred examinations, you must apply within 5 working days after the scheduled date of your exam. To apply for deferral of a final assignment, you must apply within 5 working days of the last scheduled day of classes. Visit the Registrar's Office for more information.

Plagiarism:

It is the responsibility of each student to understand the meaning of 'plagiarism' as defined in the Undergraduate or Graduate Calendars, and to avoid both committing plagiarism and aiding or abetting plagiarism by other students. (Undergraduate Calendar Academic Regulations, section 14.3, or <http://calendar.carleton.ca/undergrad/regulations/academicregulationsoftheuniversity/acadregsuniv14/>)

Academic Accommodation:

You may need special arrangements to meet your academic obligations during the term. For an accommodation request the processes are as follows:

Pregnancy obligation: write to your professor with any requests for a academic accommodation during the first two weeks of class, or as soon as possible after the need for a accommodation is known to exist. For more details visit the Equity Services website: <http://www2.carleton.ca/equity/>

Religious obligation: write to your professor with any requests for a academic accommodation during the first two weeks of class, or as soon as possible after the need for a accommodation is known to exist. For more details visit the Equity Services website: <http://www2.carleton.ca/equity/>

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 a 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 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 a accommodation (*if applicable*). After requesting a accommodation from PMC, meet with your professor to ensure

a accommodation arrangements are made. Please consult the PMC website for the deadline to request accommodations for the formally-scheduled exam (*if applicable*) at <http://www2.carleton.ca/pmc/new-and-current-students/dates-and-deadlines/>

You can visit the Equity Services website to view the policies and to obtain more detailed information on a academic accommodation at <http://www2.carleton.ca/equity/>

Important Dates:

Sept. 7	Classes start.
Sept. 20	Last day for registration and course changes in Fall and Fall/Winter courses.
Sept. 30	Last day for entire fee adjustment when withdrawing from Fall term or two-term courses.
Oct. 10	Thanksgiving Day – University closed.
Oct. 24-28	Fall Break – no classes.
Nov. 25	Last day for tests or examinations in courses below 4000-level before the Final Examination period.
Dec. 9	Last day of classes, Fall term. Classes follow a Monday schedule.
Dec. 9	Last day for handing in term work and the last day that can be specified by a course instructor as a due date for Fall term courses.
Dec. 9	Last day to withdraw from Fall term courses (academic purposes only).
Dec. 10-22	Final examinations for Fall courses, mid-terms for Fall/Winter courses. Exams are normally held all seven days of the week.
Dec. 22	Take-home exams are due.
Jan. 5	Winter term classes begin.
Jan. 18	Last day for registration and course changes in Winter term classes.
Jan. 31	Last day for entire fee adjustment when withdrawing from winter courses or winter portion of two-term courses.
Feb. 20	Family Day – University closed.
Feb. 20-24	Winter Break, no classes.
Mar. 24	Last day for tests or examinations in courses below 4000-level before the Final Examination period.
Apr. 7	Last day of Fall/Winter and Winter term classes. Last day for handing in term work and the last day that can be specified by a course instructor as a due date for term work for Fall/Winter and Winter term courses.
Apr. 7	Last day to withdraw from Fall/Winter and Winter term courses (academic purposes only).
Apr. 10-25	Final Examinations. Exams are normally held all seven days of the week.
Apr. 14	Good Friday – University closed.
Apr. 25	Take-home exams are due.

Addresses:

Department of Philosophy:	3A35 Paterson Hall www.carleton.ca/philosophy 520-2110
Registrar's Office:	300 Tory www.carleton.ca/registrar 520-3500
Student Academic Success Centre:	302 Tory www.carleton.ca/sasc 520-7850
Writing Tutorial Service:	4 th Floor, Library http://www1.carleton.ca/sasc/writing-tutorial-service/ 520-6632
MacOdrum Library	http://www.library.carleton.ca/ 520-2735