

PHIL 3350: Roboethics
(a.k.a. Philosophy, Ethics and Public Affairs)

Course Outline: Winter 2012

Prof: Jason Millar

Department of Philosophy – Carleton University

1. *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
2. *A robot must obey orders given to it by human beings except where such orders would conflict with the first law.*
3. *A robot must protect its own existence as long as such protection does not conflict with the first or second law.*

Isaac Asimov (*Runaround*, 1941)

Logistics

Class Times: Mondays, 6:05 - 8:55 pm (Jan 4 – April 5, 2012)

Location: Southam Hall 313

Instructor's Office: 329B Paterson Hall

Instructor's Office hours: by appointment on Mondays 4:30-5:30pm

Email: jlmillar@connect.carleton.ca

Course Description

Context.

The government of South Korea has a vision: a robot in every household by 2020. To make that vision reality their Ministry of Information and Communication has enlisted no fewer than 30 companies and 1,000 scientists to help research and develop robotics technologies and policies. The desire to expand and study the role of robotics in society is widespread, with similar projects underway in Europe, North America and other parts of Asia. Internationally, research in artificial intelligence (AI) and robotics has acquired a new sense of promise over the past decade. That is due in part to recent synergies between mathematics, neuroscience, computer science and psychology, and also to recent field “successes” with military robots, autonomous online agents and service bots. Consider Watson, IBM’s artificial Jeopardy! contestant. Watson recently beat two of Jeopardy!’s most winning human players in a match they called “Man vs. Machine”. Robots are beginning to assume humanlike roles in society: establishing personal relationships with humans; teaching students; patrolling warzones; and caring for the elderly. Indeed, the pace of innovation suggests that future generations of robots will be quite unlike their “simple-minded” ancestors. Robots are increasingly autonomous by design and, according to many scientists and philosophers, they will surpass human intelligence within the next half-century. Eventually it might be extremely difficult to distinguish robots from humans.

Robots are challenging many of our philosophical and ethical notions while suggesting the need for new public policy. Along with their vision, South Korea announced they were working on a draft of an official Robot Ethics Charter (drawing largely from Asimov’s three laws) that would regulate both the design of robots and human-robot interactions. Similarly, the European Commission has sponsored research into “roboethics” as a means of anticipating the ethical challenges associated with pursuing various robot technologies and deploying them in, and across, societies.

What is a robot? Can robots follow rules? What role does “code” play in the ethical programming of robots and in society? Can robots have moral agency? What would that agency look like? What are the current directions in roboethics, and robot design, and are they adequate/justifiable? What kinds of relationships ought we to encourage between robots and humans, or robots and robots? 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?

Course Objectives:

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, public policy, and a newly emerging area called “roboethics”. Students will be challenged to consider how philosophical concepts including identity, consciousness, moral agency, personhood, artificial intelligence, and rationality apply to roboethics. As this course focuses on ethics and policy issues, an emphasis will be placed on examining and developing normative claims from the 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 science debates.

Pedagogy:

This course does not require any previous training in technology or robotics. The course will consist of a combination of lectures and regular class 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 as a means of providing thought experiments for framing discussions. Students will be encouraged to draw on all sources to participate in regular class discussions and are expected to have completed all readings prior to class.

Student Evaluation:

- 1) **Blog Posts (2 x 15% of overall course grade = 30%):** Students will be required to post at least 2 short blog entries during the term, each between 600 – 750 words, to the course blog. Details on how to blog and sample topics will be provided in class. Blog postings are due on the dates indicated in the detailed reading schedule.
- 2) **Argument Exposition (30% of overall course grade):** Students will be required to write one detailed argument exposition, max. 750 words in length, drawn from one of the course readings. The argument exposition is due on the date specified in the detailed reading schedule.
- 3) **Final Essay (40% of overall course grade):** Students will be required to write a final essay between 1800 - 2000 words in length. A selection of topics will be assigned in class. If you are interested in writing on a topic of your choosing you must submit a short, one-page outline of the paper to the instructor for approval **at least one week prior to submitting the essay for grading**. Final essays dealing with non-assigned topics that are unapproved will not be graded. Final essays are due on the last day of classes.

All assignments (except for blog postings) must include a standard cover sheet (see the attached document titled “Department of Philosophy and Carleton University Policies”). Assignments must be **double spaced, using a standard 10-12 pt. font (Times New Roman, Arial, Calibri or Cambria)**. Assignments that exceed the maximum word count will be subject to a grade reduction of 5%.

Students must use proper, consistent citations in all essay assignments. The preferred styles are APA or Chicago, and all citations must include: Author; Date; Title; Journal Title (if applicable); Journal Volume and Number (if applicable); City and Publisher (for books); Page(s) (for ALL quotes).

Students should use hyperlinks and attribute all original sources in blog posts.

As per the Carleton University guidelines (see the attached document titled “Department of Philosophy and Carleton University Policies”) assignments must be handed in by 4:15 on the day they are due, or they will be considered late. Please refer to the “Department of Philosophy and Carleton University Policies” document for guidelines regarding “Deferrals for Term Work”, “Plagiarism”, and “Academic Accommodation”. The university guidelines will be strictly enforced in this course.

Argument expositions and blog posts submitted late will be penalized 5% per day. Unless students qualify for a “deferral” final essays submitted late will be penalized 10% per day.

Blog posts will be evaluated based on the following criteria:

- 10% spelling and grammar
- 10% student demonstrates the relevance to course readings and/or Roboethics
- 10% student draws from course readings to support argument
- 20% use of linked and embedded media
- 20% clarity of thesis
- 30% clarity and structure of argument

Argument Expositions will be evaluated based on the following criteria:

- 5% spelling and grammar
- 5% appropriate use of quotes and citations
- 10% clarity of introduction and conclusion
- 20% clarity and structure of exposition
- 10% explain how your chosen topic fits into roboethics
- 50% accurate exposition of source argument

Final Essays will be evaluated based on the following criteria:

- 5% spelling and grammar
- 5% appropriate use of quotes and citations
- 10% clarity of thesis statement and conclusion
- 10% depth of argument and analysis
- 10% demonstrates thorough understanding of essay topic
- 15% clarity and structure of argument
- 15% good use of related course readings (breadth and depth)
- 30% accurate use and representation of source materials

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.

Engaging Your Prof:

The best ways to get in touch with me are by email, by chatting after class, or by appointment.

I will make every effort to respond to emails on the same day they are sent (if sent during regular business hours) or by noon the next day. **Please include the following text in the subject line of ALL emails you send me: “PHIL 3350”.** If you do not include “PHIL 3350” in the subject line there is a good chance that your email will look like spam, and will be filtered by my email applications.

Feel free to make an appointment to discuss anything during my regular office hours. I’m here as a resource to you, and to help make this class, and your time at Carleton, as positive a learning experience as possible. In addition to discussing course related topics you can schedule time to

chat about other topics in philosophy or other academic interests/concerns, ask questions about grad school, etc..

Course Readings:

Required readings for the course will be drawn primarily from *Robot Ethics*, which is available from the campus bookstore. There are also required electronic resources available either through the Carleton Library, or online via the provided links.

Required Texts:

1. Robot Ethics: An Anthology, MIT Press (Campus Bookstore)
2. Online articles specified in the reading list.

Weekly Reading Schedule:

Week 1 (Jan 9): Setting the Scene: An Intro to Roboethics

1. Class Introduction: Course Syllabus and Reading List
2. Kurzweil, R. *The Age of Spiritual Machines: When Machines Exceed Human Intelligence – Chapter One*. Online at <http://www.nytimes.com/books/first/k/kurzweil-machines.html>
3. Lin, P. (2011). "Introduction to Robot Ethics." in Section I of Robot Ethics.

Week 2 (Jan 16): Overview of the Problems of Roboethics

4. Čapek, K. "R.U.R. (Rossum's Universal Robots)." Online at: <http://etext.library.adelaide.edu.au/c/capek/karel/rur/complete.html>.
5. Bekey, G. "Current Trends in Robotics: Technology and Ethics." In Section I of Robot Ethics.
6. Abney, K. "Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed." in Section I of Robot Ethics.

Week 3 (Jan 23): The Imitation Game I

7. Turing, A.M. Computing Machinery and Intelligence. *Mind* 59, 1950. 433-460. Online at: <http://loebner.net/Prizef/TuringArticle.html>
8. Searle, J.R. "Minds, Brains, and Programs." In John Haugeland (Ed.) *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. (Cambridge, Mass.: MIT Press, 1997). 183 – 204. Online at <http://pami.uwaterloo.ca/tizhoosh/docs/Searle.pdf>

Week 4 (Jan 30): The Imitation Game II

9. **First Blog Entry Due**
10. Proudfoot, D. (2011). "Anthropomorphism and AI: Turing's Much Misunderstood Imitation Game." *Artificial Intelligence* 175(5-6). 950-957.
11. Sparrow, R., (2011). "Can Machines be People? Reflections on the Turing Test." In Section VII of Robot Ethics.
12. Warwick, K. (2011). "Robots with Biological Brains." In Section VII of Robot Ethics.

Week 5 (Feb 6): Design and Programming (Section II of Robot Ethics)

13. Wallach, W., Allen, C. *Robot Ethics*.
14. Hughes, J. *Robot Ethics*.
15. Bringsjord, S., Taylor, J. *Robot Ethics*

Week 6 (Feb 13): Military (Section III of Robot Ethics)

16. **Second Blog Entry Due**

17. Sharkey, N. *Robot Ethics*.
18. Guarini, M., Bello, P. *Robot Ethics*.
19. Lokhorst, G., van de Hoven, J. *Robot Ethics*.

Break Week (Feb 20) – No Class

Week 7 (Feb 27): Law (Section IV of Robot Ethics)

20. O'Meara, R. *Robot Ethics*.
21. Asaro, P. *Robot Ethics*.
22. Calo, M. *Robot Ethics*.

Week 8 (Mar 5): Psychology and Sex (Section V of Robot Ethics)

23. **Argument Exposition Due.**
24. Scheutz, M. *Robot Ethics*.
25. Levy, D. *Robot Ethics*.
26. Whitby, B. *Robot Ethics*.

Week 9 (Mar 12): Social Robots and Human-Robot Interaction

27. Mori, M. "The Uncanny Valley." *Energy* 7:4, 1970. 33 – 35. Online at <http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html>
28. Duffy, B. "Anthropomorphism and the Social Robot." *Robotics and Autonomous Systems* 42(3-4), 2003. 177 - 190.
29. Levy, D. "The Ethical Treatment of Artificially Conscious Robots." *International Journal of Social Robotics* 1, 2009. 209 – 216.

Week 10 (Mar 19): IBM's Watson: Expert Systems and the Californication of Machines

30. Dworkin, R. (1967). "The Model of Rules." *The University of Chicago Law Review* 35(1): 14-46.
31. Kerr, I. (2004). "Bots, Babes, and the Californication of Commerce." *University of Ottawa Law & Technology Journal* 1(1-2): 285-324. Available online at: <http://www.uoltj.ca/articles/vol1.1-2/2003-2004.1.1-2.uoltj.Kerr.285-324.pdf>.
32. In-Class viewing of IBM videos and an episode of "Jeopardy! Man vs. Machine", followed by an in-class exercise.

Week 11 (Mar 26): Medicine, Care, and Robot Servitude (Section VI of Robot Ethics)

33. Borenstein, J., Pearson, Y. *Robot Ethics*.
34. Sharkey, N., Sharkey, A. *Robot Ethics*.
35. Petersen, S. *Robot Ethics*.

Week 12 (Apr 2): Engineering Ethics and the Sociology of Technology

36. **Final Essay Due.**
37. Verbeek, P. (2008). "Morality in Design: Design Ethics and the Morality of Technological Artifacts." In *Philosophy and Design*. (SpringerLink): 91-103. (available on the Carleton Library as an electronic resource).
38. Latour, B. "Where are the Missing Masses?" In Wiebe E. Bijker and John Law (Eds.). *Shaping Technology/Building Society: Studies in Sociotechnical Change*. (Cambridge, Mass.: MIT Press): 225-258.

Other sources of useful Roboethics information:

I will be adding posts, links, RSS feeds, and Twitter hash tags to the course blog, hosted at <http://jasonmillar.ca/roboethics>.

Department of Philosophy and Carleton University Policies

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 in the lower right corner of the cover sheet:
 - student name
 - student number
 - course number and section
 - instructor's name
- The Philosophy Department does not accept assignments by FAX. You may send them by courier, if necessary.
- 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://www4.carleton.ca/calendars//ugrad/current/regulations/acadre/suniv14.html#14.3>)

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 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 visit the Equity Services website <http://www2.carleton.ca/equity/accommodation/>

Religious obligation: 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 visit the Equity Services website <http://www2.carleton.ca/equity/accommodation/>

Students with disabilities requiring academic accommodations in this course must register with the Paul Menton Centre for Students with Disabilities (PMC) for a formal evaluation of disability-related needs. Documented disabilities could include but are not limited to mobility/physical impairments, specific Learning Disabilities (LD), psychiatric/psychological disabilities, sensory disabilities, Attention

Deficit Hyperactivity Disorder (ADHD), and chronic medical conditions. Registered PMC students are required to contact the PMC, 613-520-6608, every term to ensure that I receive your Letter of Accommodation, no later than two weeks before the first assignment is due or the first in-class test/midterm requiring accommodations. If you only require accommodations for your formally scheduled exam(s) in this course, please submit your request for accommodations to PMC by the deadlines published on the PMC website: <http://www2.carleton.ca/pmc/new-and-current-students/dates-and-deadlines/>

Important Dates:

Sept. 8	Classes start (after Orientation events).
Sept. 21	Last day for registration and course changes in Fall term and two-term courses.
Sept. 30	Last day for entire fee adjustment when withdrawing from fall term or two-term courses.
Oct. 7	University Day – no classes.
Oct. 10	Thanksgiving Day – university closed.
Nov. 21	Last day for tests or examinations in courses below 4000 level before the Final Examination period.
Dec. 5	Last day of classes, Fall term. Last day for handing in term assignments, subject to any earlier course deadline.
Dec. 5	Last day to withdraw from Fall term courses (academic purposes only).
Dec. 8-21	Final examinations for Fall courses, mid-terms for Fall/Winter courses.
Jan. 4	Winter term classes begin.
Jan. 17	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, classes suspended.
Mar. 22	Last day for tests or examinations in courses below 4000 level before the Final Examination period.
Apr. 5	Last day of Fall/Winter and Winter term classes. Last day for handing in term assignments, subject to any earlier course deadline.
Apr. 5	Last day to withdraw from Fall/Winter and Winter term courses (academic purposes only).
Apr. 6	Good Friday – university closed.
Apr. 11-24	Final Examinations.

Addresses:

Department of Philosophy:	3A46 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
Paul Menton Centre:	500 University Centre www.carleton.ca/pmc 520-6608
Writing Tutorial Service:	4 th Floor, Library www.carleton.ca/wts 520-6632
MacOdrum Library	http://www.library.carleton.ca/ 520-2735