Spring 2019

FYS: Ethics And Technology (PHIL 07/CPSC 15) Syllabus

Ameet Soni
Swarthmore College, soni@cs.swarthmore.edu

Krista Karbowski Thomason
Swarthmore College, kthomas2@swarthmore.edu

FYS 007: ETHICS AND TECHNOLOGY
Sci Center 246
MWF 10:30-11:20
Spring 2019

Prof. Thomason
Office location: 209 Beardsley
Email: kthomas2@swarthmore.edu

Prof. Ameet Soni
Office location: 253 Science Center
Email: asoni1@swarthmore.edu

Office Hours:
Jointly Held: 3 PM – 4 PM Friday, Science Center Coffee Bar
Prof. Soni: 2 PM – 4 PM Monday, 253 Science Center
Prof. Thomason: 2 PM – 4 PM Wednesday, 209 Beardsley

Contact Information: You don't need to set up an appointment for office hours; they are a drop-in, first-come-first-serve affair. If you can't make it during office hours, please email to make an appointment for another day or time. Do bear in mind that we have other work, meetings, functions, etc., which might mean we cannot chat long if you drop by unplanned outside of office hours. Email is the best way to get in touch with us. We check email regularly, but we are also people with lives outside of school. Please give us adequate time to respond.

Course description: There has been an accelerated shift in the influence of computing technology and the use of algorithms in our daily lives. With this technology comes serious ethical questions. Philosophers are often well-equipped to wrestle with ethical questions, but less well-equipped to wrestle with questions of technology itself. Computer scientists are well-equipped to deal with the problems and challenges of technology, but less well-equipped to deal with the ethical problems and challenges that technology can pose. In this co-taught course, we bring together the two fields to address ethical questions involving social media, data mining, self-driving cars, artificial intelligence, and other topics.

Required Texts

- Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*
- Sheila Jasanoff, *The Ethics of Invention: Technology and the Human Future*
- Articles and book chapters posted on the course Moodle page

Assignments

In this course, you will write two formal papers, four reading responses, and one critical algorithm analysis.

Formal philosophy papers: The midterm and final papers for this course are formal papers that will follow the traditional philosophical format: explaining a position or question, developing your own argument, and anticipating objections (we will discuss this format during class time so you know exactly what it looks like). We will give you a list of topics to choose from for both the midterm and final. These papers will not require outside sources. You are **specifically forbidden** to consult Wikipedia, Sparknotes, and online sources of a similar nature. If you have trouble with the material, come talk to us during office hours—that's what they are for! We are looking for **your** critical engagement with the material presented in class and in the text. The midterm paper will be 4 to 6 pages and the final paper will be 6 to 8 pages.

Reading responses: You will write four reading responses during the semester. These are very short (no more than two pages) papers in which you (1) explain part of an argument in one of the readings and (2) respond in some way. Responding to the reading includes any of the following:

- Agree with the argument (or part of the argument) and explain why
- Disagree with the argument (or part of the argument) and explain why
- Pose a related question and explain why it's important
- Show how the argument (or part of the argument) is related to a contemporary example or problem

You'll do two response papers before spring break and two response papers after spring break. These are due in class hard copy and will be used to help open the discussion for the class meetings. You will not be assigned to particular readings; you will choose the readings you want to respond to. However, your response papers are due the date the reading is covered in class. Once the readings are covered, response papers on those readings will not be accepted.

Lab practicum: In addition, you will do a critical algorithm analysis. For this assignment you will work with software related to algorithms we will be discussing in class. This software builds mathematical representations of language, and using it will allow you to explore the way in which biases can be encoded in these models. The graded portion of the assignment will be a write-up of this exercise. Further details about this assignment and the due date will be forthcoming.

Grading: This course uses a point system for grading rather than a percentage system. You can monitor your progress in the course by simply adding up the number of points you have received for the assignments and checking the total against the conversion chart below:

Midterm paper: 100 points
Final paper: 150 points

Four response papers: 25 points each (100 point total)
Critical algorithm analysis: 50 points
Course total: 400 points

A = 400-360
B = 359-320
C = 319-280
D = 279-240
F = 239-000

Late work policy: The late work policy for this course is evil and draconian. All work is due in class on the day designated on the syllabus – not by mail slot, office box, or email. *Any assignment handed in after the due date – be it one hour or one day – will not be accepted*. Printer or computer problems are no excuse. This is college; plan ahead. If you are very sick, you can email the assignment ahead of time, but it must be BEFORE class begins. Early work is always welcome if you know you will be absent, but please make arrangements beforehand.

Course Policies

Things we shouldn't have to say: You are legally an adult. You signed up for this course voluntarily. You know when it meets and you know what time it begins. Routine lateness is unacceptable. Routine absence is unacceptable. Routine playing with electronic things during class is unacceptable. You get our full attention for 50 minutes; reciprocity dictates that we get yours. We reserve the right to deduct points from any assignment due to excessive lateness or excessive absences. We will do so with NO prior warning and we will do so at any point during the semester. It's your responsibility to read the syllabus and familiarize yourself with the course expectations and policies.

Classroom Environment: Professors strive for different kinds of classroom environments, so it may be helpful to know what to expect. We look for a lot of verbal and visual feedback from students. If you look confused or unhappy, we might ask you why – not to put you on the spot, just because we want to help. It's cliché, but chances are good that one of your classmates has the same question or puzzle you do. We not only welcome but also encourage you to think aloud, even if it doesn't come out sounding polished or erudite (it doesn't matter how smart you sound; it matters how much you're thinking). Feel free to preface your comments with "I'm just formulating this thought" or something similar so that we know you're testing something out. We will likely ask students follow-up questions or ask you to expand on the point/question you have. This isn't meant to a test or to put you on the spot; often we're looking for further clarification or the broader significance of what you say. It certainly doesn't mean what you said was bad, wrong, or inarticulate.

Philosophy isn't called "the great conversation" for nothing. We simply cannot do what we do without lots and lots of discussion. We understand that sometimes speaking in class makes

people nervous, but we expect participation from every student. If you are uncomfortable about speaking in class, please come see one of us and we can discuss ways to mitigate that nervousness.

Please keep in mind that participation can look many different ways. Asking good questions, answering questions about the reading, and adding to class discussion are all ways to participate in class. Dominating every class discussion is not good participation. Please refrain from interrupting us or any of your classmates. Never get the impression that disagreement is unhealthy. Just because a classmate disagrees with your comment it does not entail that she is attacking you personally. Our discussions will always be civil even when they are lively and rife with disagreement.

All of the above forms of participation do require proper preparation. You are expected to not only to complete all of the reading before attending class but also to have critically engaged with the material.

Laptops: Laptops are not permitted because they are bad for the entire classroom environment. Use this course as an opportunity to unplug, be present, use really fun pens and notebooks, and practice your handwriting skills. Exceptions can be made if need be. If you prefer to use a tablet to write your notes, this will be permitted under the requirement that the screen always remains on your note-taking application.

Accomodations: If you believe that you need accommodations for a disability, please contact the Office of Student Disability Services (Parrish 113W) or email studentdisabilityservices@swarthmore.edu to arrange an appointment to discuss your needs. As appropriate, the Office will issue students with documented disabilities a formal Accommodations Letter. Since accommodations require early planning and are not retroactive, please contact the Office of Student Disability Services as soon as possible. For details about the accommodations process, please see http://www.swarthmore.edu/academic-advising-support/welcome-to-student-disability-service.

You are also welcome to contact us privately to discuss your academic needs. However, all disability-related accommodations must be arranged through the Office of Student Disability Services.

Academic Integrity: Plagiarism is a very serious offense. The following is excerpted from the College's official policy on academic misconduct:

"Because plagiarism is considered so serious a transgression, it is the opinion of the faculty that for the first offense failure in the course and immediate suspension for a semester or deprivation of the degree in that year should be the usual punishment; for a second offense the penalty should normally be expulsion."

If you are guilty of academic misconduct for this course, we will pursue official sanctions through the proper channels. Please be aware that this may result in your failure for the course, suspension, or expulsion depending on the circumstances and severity of the case.

Reading Schedule

This syllabus is a living document; please be aware that the schedule elements on this page will change throughout the semester. **It is the student's responsibility to review the Moodle page for the most current schedule.**

Preliminaries
1/23: Course description, policies, and introductions
1/25: How to Write Like a Philosopher: Argument construction
1/28: How to Read Like a Philosopher

Applied Ethical Theory: A Primer
1/30: Relativism, Pluralism, Criticism: Midgley, “Trying Out One’s New Sword”
2/1: Aristotle/Virtue Ethics: Vallor, “Flourishing on Facebook”
2/4: Hume: Wonderly, “A Humean Approach to Violent Video Games”
2/6: Kant/Deontology: Sussman, “What’s Wrong with Torture?”
2/8: Utilitarianism: Singer, “Famine, Affluence, and Morality”
2/11: Feminist ethics: Noddings, “Add Women and Stir”
2/13: Buddhist ethics: McRae, “Equanimity and the Virtue of Open-Mindedness”

What’s Technology?
2/15: Smith, “Franken-Algorithms”
2/18: Jasanoff, Chapter 1, “The Power of Technology”
2/20: Winner, “Do Artifacts Have Politics?”
2/22: Guest Lecture - Ronni Gura Sadovsky
2/25: Jasanoff, Chapter 2, “Risk and Responsibility”
2/27: Noble, Chapter 1, “A Society Searching”

Machine Learning and Algorithm Bias
3/1: Hill, “What an Algorithm Is”
3/4: Noble, Chapter 2, “Searching for Black Girls”
3/6: Noble, Chapters 3 “Searching for People” & 4 “Searching for Protection”
3/8: Bozdag, “Bias in Algorithmic Filtering” (***Midterm Paper Due****)*
3/11-3/15: *Spring Break—No Class*
3/18: Newton "The Trauma Floor", *The Verge*, February 25, 2019
Koebler and Cox, “The Impossible Job: Inside Facebook’s Struggle to Moderate Two Billion People”, *Motherboard*, August 23, 2018

3/20: Barocas & Selbst, “Big Data’s Disparate Impact”
ProPublica Investigation, “Machine Bias”, May 23, 2016
3/22 and 3/25: Caliskan, Bryson, & Narayanan, “Semantics derived automatically from language corpora contain human-like biases”
Google AI, “Text Embedding Models Contain Bias. Here's Why That Matters”
Moritz Hardt, “How big data is unfair”

Surveillance and Privacy
3/27: Thomson, “The Right to Privacy”
3/29: Jasanoff, Chapter 6, “Information’s Wild Frontier”
4/1: Gutting, “Foucault: A Very Short Introduction”
4/3: Chen and Cheung, “The Transparent Self Under Big Data: China’s Social Credit Policy”
4/5: Hull, “Successful Failure: Foucault and Privacy”

Artificial Intelligence
4/8: Turing, “Computing Machinery and Intelligence”
4/10: Thomson, “The Trolley Problem”
4/12: Shaw, “Artificial Intelligence and Ethics”
4/15: Etzioni & Etzioni, “Incorporating Ethics into Artificial Intelligence”
4/17: Martin, “Who Should Decide?
4/19: Basl, “Ethics of Creating Artificial Consciousness”
4/22: ”Bryson, “Robots Should Be Slaves”
4/24: Sparrow, “Killer Robots”

Transhumanism
4/26: Jasanoff, Chapter 5, “Tinkering with Humans”
4/29: Vallor, “Knowing What to Wish For”
5/1: Martin Peterson, “The Ethical Failures Behind the Boeing Disasters”
Whittlestone, Nyrup, Alexandrova, & Stephen Cave, “The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions”
5/3: Current topics - submitted by students

5/16: Final Papers due by **noon** by email (send to both professors) or in person (to Prof. Thomason)