PHIL 470/870 Topics in Philosophy of Science — Winter 2021

Instructor: Dr. Catherine Stinson

Topic: Third Wave Artificial Intelligence

Overview
The first wave of AI peaked in the mid-20th century. The second wave began in the late 1980s, with the resurgence of connectionism and rise of machine learning. Some refer to the third wave as a hypothetical near future where artificial general intelligence (AGI) has been achieved. This course focuses on a third wave that is already here. Although we do not have anything close to AGI, a rapid expansion of the scale and ubiquity of the not-so-smart AI we currently have is allowing big tech platforms to surveil, analyze and nudge human behaviour in unprecedented ways.

A loose collection of scholarly and policy work responding to this third wave has been dubbed AI Ethics. The name gives the impression that it is a branch of philosophy, yet philosophers have so far played a negligible role. One of the main goals of the course is to figure out what role (if any) philosophers might usefully play in so-called AI Ethics. A re-invented Philosophy of AI that brings together analytic Philosophy of Science with Critical Studies could play an important role in AI Ethics. Philosophy of Science speaks directly to the concerns of scientists. Critical Studies makes clear what is at stake socially and politically. The trick is to find common ground. The themes below apply this nascent hybrid field to examples of problems facing AI.

Theme A: Democracies are collapsing, and wealth is being drained out of communities into the pockets of surveillance capitalism. Both private companies and police forces are expanding the reach of biometric monitoring from public spaces into the home. These are not inevitable side-effects of tech innovation, and the costs are disproportionally being felt by colonized peoples.

Theme B: Blame for AI's ethical crisis is often assigned to nerd culture, known for being unwelcoming to workplace diversity, scornful of outside expertise, and unconcerned with social issues. Despite popular claims that technology is neutral, AI researchers are beginning to recognize that algorithms can be biased, and that they bear moral responsibility for the effects of their inventions.

Theme C: The "intelligence" central to AI has long gone unquestioned. The tasks that AI is built to do, and idealized descriptions of AGI focus on a narrow range of the intelligent behaviors of humans. It is a rationalist notion of intelligence that stands in contrast to the perceived emotionality and irrationality of women and POC. It is no coincidence that alongside the rise of AI, there has been a resurgence of race science.

Course Delivery
Class will meet for discussion sessions in the **Classroom** channel **on Teams** every **Tuesday at 11:30.** Although the class is scheduled for 3 hours per week, most discussion sessions will last about 90 minutes. **Here** is a list of norms for respectful discussion that I hope we'll all follow.

Where copyright permits, readings will be posted online for collaborative discussion in Perusall. Students are encouraged to use this to ask and answer questions about confusing passages, highlight and discuss interesting issues, raise provocations, or link to additional materials.

Science Fiction pairings are noted on the schedule. Films will be screened occasionally throughout the semester, in the last 90 minutes of class time. This is an optional component of the course.

Reading Schedule

Date	Topic	Readings	Theme
January 12	History of Philosophy of AI	n/a Extra: https://www.coe.int/en/web/artificial-intelligence/history-of-ai Sci Fi: 🎥 Victim of the Brain	all
January 19	Surveillance capitalism	Zuboff, *The Age of Surveillance Capitalism*, Ch 4, 5 Extra: 🎥The Great Hack Abdalla & Abdalla, The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity Pasquale, *The Black Box Society*, Ch 3 Sci Fi: Eggers, *The Circle*	A
January 26	Don't be evil	Douglas, The Moral Responsibilities of Scientists Extra: Heated twitter thread about ethics reviews at NeurIPS Hao, Training a single AI model can emit as much carbon as five cars in their lifetime Sci Fi: 🎥 Frankenstein (1931)	B
February 2	Whose intelligence?	Adam, *Artificial Knowing*, pp 34-47, 99-104, 110-128. Extra: Adam, *Artificial Knowing: Gender and the thinking machine* Leslie et al. Expectations of brilliance underlie gender distributions across academic disciplines Hong & Page, Groups of diverse problem-solvers can outperform groups of high-ability problem solvers Sci Fi: 🎥 Gattaca	C
February 9	Biometric surveillance	🎥 Coded Bias Simone Browne, *Dark Matters*, Ch 3 Extra: Benjamin, *Race After Technology* Sci Fi: 🎥 Minority Report	A

February 23	Biased algorithms	Stinson, Fuck the Algorithm Extra: Liao & Huebner, Oppressive Things Winner, Do Artifacts Have Politics? Crawford, Can an Algorithm be Agonistic? Sci Fi: Asimov, _Franchise_	B
March 2	Race realism in AI	Saini, _Superior: The return of race science_, Ch 9 Extra: https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/ Sci Fi: Okorafor, _Binti_	C
March 9	From surveillance to incarceration	Robertson, Khoo & Song, To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada (excerpt) Extra: https://www.theglobeandmail.com/canada/article-investigation-racial-bias-in-canadian-prison-risk-assessments/ https://washingtoncitypaper.com/article/178161/dc-agency-purchases-ankle-monitors-that-can-record-defendants-every-word-and-motion/ Davis, _Are Prisons Obsolete?_ Sci Fi: 🎥Dirty Computer, 🎥Escape from LA	A
March 15	Diversity	Wylie, Why Standpoint Matters Molten & Rogers, The Actual Science of James Damore's Google Memo Extra: Damore, Google's Ideological Echo Chamber Denton et al., Bringing the People Back In: Contesting Benchmark Machine Learning Datasets Harding (Ed), _The Feminist Standpoint Theory Reader_ 🎥The Social Network	B
March 23	AI's colonial roots	Cave, The Problem with Intelligence Extra: Mbembe, _Necropolitics_ Foucault, _Society Must be Defended_, 17 March, 1976 https://www.wired.com/story/for-superpowers-artificial-intelligence-fuels-new-global-arms-race/ Sci Fi: Leckie, _Ancillary Justice_	C
March 30	Research Talks	https://lizlerman.com/critical-response-process/	all
April 6	Research Talks & Videos	n/a	all

Participation Options
Breadth Stream
- attend all classes, and provide feedback on research talks
- do main readings before each class
- sometimes talk/chat in class and write in Perusall
- complete 3 small assignments

Depth Stream
- choose 1 of the 3 themes to focus on
- attend all classes for your theme, and provide feedback on research talks
- do main and some extra readings (from the reading schedule, or of your own choosing) before each class covering your theme
- often talk/chat in class and write in Perusall when your theme is covered
- complete 1 research project and give a research talk

Assessment
Breadth Stream:
10%   Participation
30% x 3   Small Assignments

Depth Stream:
10%   Participation
60%   Research Project
30%   Research Talk

Absences, Extensions, Emergencies, Global Pandemics…
Minor deviations from the expectations outlined here, like occasional absences or slightly late assignments, do not require any special permission or notice. Do what you can, when you can. If you are having major struggles, please contact the instructor. Any reasonable requests for accommodations or modifications will be granted.
Word counts and presentation times, however, should be strictly observed.

Small Assignment Options
Small Assignments must cover at least **2 different themes**. Due dates are **January 28**, **February 25**, and **March 25**. Submit your assignments in OnQ. Choose from the following options:

Small Assignment Option #1 — Academic Paper
Write a short paper (2000 words maximum) that explains a philosophical issue raised in one of the readings, and argues for a position. (This assignment may be done up to 3 times.)

Small Assignment Option #2 — Video Explainer
Create a 3-5 minute video that explains an issue or concept raised in one of the readings in an accessible way, and considers at least 2 positions one could take on the issue. Post the video (or a link to it) in Teams AND submit in OnQ. (This assignment may only be done once.)

Small Assignment Option #3 — Class Presentation + Study Notes
Give an 8-10 minute presentation in class that summarizes the main reading(s), and raises questions for discussion. Presentations must be scheduled with the instructor at least 1 week in advance. Dates are limited. Submit study notes (approx. 800 words) covering your summary of the reading(s) and

the discussion that followed, suitable for a student who missed that class. Post the notes in Teams AND submit in OnQ. (This assignment may only be done once.)

Small Assignment Option #4 — Sci Fi Paper
Write a short paper (2000 words maximum) that expands upon a philosophical issue raised in one of the science fiction pairings, and argues for a (real world) position on the issue, informed by the reading(s). (This assignment may be done up to 2 times.)

Small Assignment Option #5 — Popular Essay
Write a short essay (1200 words maximum) in the style of a blog post, magazine article, or newspaper Op-Ed that explains an issue raised in one of the readings in an accessible way, and argues for a position. Also write a pitch for the article, and submit evidence that you have sent the pitch to an appropriate venue for your essay (for example, a copy of the email). Your essay should follow the style guide of your chosen venue. (This assignment may be done up to 2 times.)

Research Project Options
Submit your assignments in OnQ

Research Project Option #1
Write a research paper (6000 words maximum) exploring a philosophical issue raised in the readings for your chosen theme, and arguing for an original position.
- A brief statement of the paper's main topic, partial reading list, and a choice of journal where the paper would fit is due **January 28**. (Later changes to topic and journal are permitted.) (5%)
- An outline, partial draft, or short excerpt (maximum 1000 words) is due **February 25**. (15%)
- A final paper written in a format accepted by the chosen journal is due **March 25**. (40%)

Research Project Option #2
Workshop a possible Thesis/Major Research Paper on a topic related to your chosen theme (with no strings attached).
- A brief statement of the paper's main topic, partial reading list, and your intention to workshop a Thesis/MRP is due **January 28**. (5%)
- A proposal in the format specified in the Grad Handbook (abstract and bibliography) is due **February 16**. (15%)
- A substantial chunk (6000 words maximum) of well-developed writing (chapter drafts, for example) is due **March 25**. (40%)

Research Talk Options
Your Research Talk option should be communicated to the instructor by **March 23**.

Research Talk Option #1
Prepare a 13-15 minute talk based on your Research Project aimed at an audience with some familiarity but no expertise in your research topic. Prepare and practice what you'll say, and make a slide deck (Powerpoint or similar) of A/V aids that support your talk. Presentations will be delivered during class time on **March 30** and **April 6**. A short Q&A with the class will follow.

Research Talk Option #2
Create a 5-8 minute video based on your Research Project that introduces your topic in an accessible way and expands on one or two main points. Post the video or a link to Teams AND submit in OnQ by **April 5**. Your video will be shown in class on **April 6**, followed by a short Q&A.