PHIL 470/870 Topics in Philosophy of Science — Winter 2022

Instructor: Dr. Catherine Stinson

Topic: Third Wave Artificial Intelligence

Overview

This course explores recent advances and methods in Artificial Intelligence (AI) from the perspective of philosophy of science. We will focus on epistemic and metaphysical questions about recent AI, but also consider the social context in which the science is being done, and the implications.

Three major themes will be explored:

1. **Understanding Deep Learning:** Do deep learning networks that match human performance on perceptual and linguistic tasks perform these tasks intelligently? What would a demonstration that deep learning is truly intelligent look like? Is deep learning capable of creativity?

2. **The Politics of Data:** Where do the datasets used to train AI come from? How should we interpret improvement on benchmark AI tasks? Can datasets or algorithms be racist?

3. **Interrogating Intelligence:** Whose intelligence is included in and excluded from the aims of AI? Why is eugenics so popular in AI? Should we be worried about an AI singularity?

Course Delivery

This is a seminar course, so most of our class time will be spent in discussion. Mini lectures will often start the class, but the remainder will be semi-structured time where participation is expected. Please come to class having done the required readings, with questions and comments ready.

Here is a list of norms for respectful discussion that I expect we'll all follow.
All usual Queen's rules on grading and academic integrity apply (see https://www.cs.queensu.ca/students/undergraduate/syllabus). You are expected to know and follow those rules.

Texts

Where copyright permits, readings will be available online. Links are provided in the Schedule below.

Absences, Extensions, Emergencies

Minor deviations from the expectations outlined here, like occasional absences or slightly late assignments, do not require any special permission or notice. If you are having major struggles keeping to the schedule, please contact the instructor. Any reasonable requests for accommodations or modifications will be granted. **Presentation times should be strictly observed out of consideration for other presenters.** If you go over the word limits, I may stop reading.

Assessment

15%	Participation
60%	3 Small Assignments OR 1 Research Project
25%	Popular Essay

Participation

Since this is a seminar, you are expected to contribute regularly to class discussions (participating via the chat is acceptable while meeting remotely). The process of mulling over philosophical questions and working out agreement or disagreement in conversation with others is an important learning experience, and skill-building exercise. You will be expected to have your camera on for at least two of the weeks when we are meeting remotely.

Students in PHIL870 will be expected to show deeper engagement with the readings, including familiarity with some of the extra readings.

If being evaluated on the frequency and quality of your contributions to class discussions would cause you undue stress, or you are unable or unwilling to appear on camera, please ask the instructor about alternative modes of participation.

Small Assignments

Each assignment should be related to the assigned material from a different week (or weeks) of class. You are welcome to look at the Extra readings, and to incorporate additional sources, but these are not research assignments. Assignment due dates are **February 1**, **March 1**, and **March 29**.

Default Option — Academic Essay

Write a short paper (2000 words maximum) that explains a philosophical issue raised in one of the readings, and argues for a position.

Alternative Option — Mock Conference Talk

The 3rd assignment may be replaced by a presentation in the style of a conference talk (8-10 minutes) to be delivered in class on **March 29**, and to be followed by a Q&A. The presentation should communicate the argument made in one of your previous Academic Essays or your Popular Essay. Please let the instructor know by March 14 if you plan to give a talk.

Popular Essay

Write a short essay (1000 words maximum) in the style of a Medium post that addresses an issue raised in one of the readings in an accessible way for an educated but non-specialist audience. The topic may be connected to one of your other assignments, but if so must be a distinct piece of writing. This is much harder than it seems, so it will be broken into 3 stages:

- A complete, finished draft of the essay is due **February 15**. (10%)
- A peer review of a classmate's draft is due **March 8**. (5%)
- A final re-write of the essay that takes into account the peer review is due **March 22**. (10%)

Research Project

Option #1

Write a research paper (3000 - 5000 words) exploring a philosophical issue raised in the readings, and arguing for an original position. The extra readings are good starting points.

- An outline of the paper's main thesis and argument structure, (800 - 1600 words), and initial reference list is due is due **March 1**. (10%)
- The final paper is due **March 29**. (30%)
- A presentation in the style of a conference talk (10-15 mins) is to be delivered on **April 5**, followed by a Q&A. (20%)

Option #2

Workshop a possible MA Thesis/Major Research Paper on a topic exploring a philosophical issue raised in the readings. The extra readings are good starting points.

- A proposal in the format specified in the Grad Handbook (extended abstract and bibliography) is due **February 15**. (10%)
- A substantial chunk (3000-5000 words) of well-developed writing (sections or chapters, for example) is due **March 29**. (30%)
- A presentation in the style of a conference talk (10-15 mins) is to be delivered on **April 5**, followed by a Q&A. (20%)

If you choose option 2, you may submit your Popular Essay Draft on **February 21**.

Schedule

Date	**Topic**	**Readings**	**Due**
January 11	Introduction to AI and Deep Learning	Extra: Buckner, Deep Learning: A Philosophical Introduction Mitchell, Why AI is Harder Than we Think	
January 18	Adversarial Examples	https://gradientscience.org/adv/ Buckner, Understanding adversarial examples requires a theory of artefacts for deep learning Extra: https://distill.pub/2019/advex-bugs-discussion/	
January 25	DL and Explanation	Saxe et al., If Deep Learning is the Answer, What is the Question? Extra: Hancox-Li, Robustness in Machine Learning Explanations: Does it Matter? Mittelstadt et al., Explaining Explanations in AI Thompson, Forms of explanation and understanding for neuroscience and artificial intelligence	
February 1	DL and Creativity	Halina, Insightful Artificial Intelligence Extra: watch the AlphaGo documentary https://openai.com/blog/dall-e/ https://openai.com/blog/jukebox/	A1
February 8	Benchmark Datasets	Denton et al., Bringing the People Back In: Contesting Benchmark Machine Learning Datasets Extra: Paullada et al., Data and its (dis)contents: A survey of dataset development and use in machine learning research https://excavating.ai/	

February 15	Large Language Models	https://dailynous.com/2020/07/30/philosophers-gpt-3/ Extra: Bender et al., On the Dangers of Stochastic Parrots: Can Language Models be Too Big?	Popular Essay Draft MRP Proposal
Reading Week			
March 1	Is math racist?	Liao & Huebner, Oppressive Things Extra: Winner, Do Artifacts Have Politics? Crawford, Can an Algorithm be Agonistic?	A2 / Project Outline
March 8	Whose intelligence?	Adam, *Artificial Knowing*, pp 34-47, 99-104, 110-128. Extra: Birhane, The Impossibility of Automating Ambiguity Bostrom, Ethical Issues in Advanced Artificial Intelligence	Popular Essay Reviews
March 14	AI and Eugenics	Stark & Hutson, Physiognomic Artificial Intelligence Extra: read the footnotes	
March 22	AI's colonial roots	Cave, The Problem with Intelligence Extra: Birhane & Guest, Towards decolonizing computational sciences Foucault, *Society Must be Defended*, 17 March, 1976	Popular Essay
March 29	Mock Conference Talks	Q&A sessions will follow a shortened version of the process described here: https://lizlerman.com/critical-response-process/. Please familiarize yourself with the process and come prepared to participate.	A 3 / Project Final Paper
April 5	Project Presentations		