AP/COGS/PHIL 3750 – Philosophy of Artificial Intelligence – Winter 2022

Lecture: Thursdays 2:30-5:30 Curtis Lecture Halls A

Professor: Javier Gomez-Lavin
Office: S438 Ross Building
Office Hours: by appointment only
jglavin@yorku.ca

Course Description:

Artificial intelligence is a deceptively simple term: all you have to do is *build something intelligent*. But when we pause to think about that maxim, it seems obvious that we've done similar things throughout human history. From mesmerizing pressure-plate controlled fountains in royal gardens of Versailles to the 'karakuri' automata built in 17th century Japan, we have many examples of engineered, 'smart' marvels. However, something is still missing in this picture; though clever and entertaining, most wouldn't say that these party-tricks are *intelligent*. It seems that we need to do some philosophical work to figure out *precisely* what we mean by "intelligent" and, perhaps even what counts as "artificial." This class begins with these questions and takes a multifaceted and historically anchored approach to the development of AI to come to a more holistic, socially- and ethically-ensconced account of what it means to be *artificially intelligent*.

We'll begin with a review of what programs and computers are, approaching formal systems and algorithms from the ground-up—eventually hand-coding our own register machines. This will give us an appreciation for the computation-as-symbol-manipulation dogma that lies at the heart of "good old-fashioned AI" that dominated the second half of the 20th century, which we'll review after we consider classic views of intelligence from 20th century psychology and cognitive science. We'll then chart the significant challenges faced by GOFAI and turn towards a plural and embodied taxonomy of intelligences, before dipping our toes into the technical details and philosophical implications of deep learning and neural networks. During the last third of the course we'll review some of the burgeoning social and moral issues raised by the rapid escalation and propagation of AI assistants and algorithms. Specifically, we'll look at the effect of algorithms in incarceration, job recruitment and gatekeeping, the role of feminized-AI assistants in perpetuating unjust stereotypes, proactive steps we can take to mitigate the effects of AI by taking on a decolonialist perspective, and whether we can even design 'ethical' AI.

Organization of the course:

Weekly seminar meetings (3 hours, with breaks). Class meetings will be conducted in-person and on-campus at the designated classroom. Unless otherwise noted

- **NB AS OF 1/7/2022 class through 2/3/2022 will be held on zoom**
- *Note, that this is one "extra" day on zoom as compared to York U guidelines – this choice was made to keep both courses "in-sync" this term.*

Course Learning Objectives:

After completing this course, students should be able to:

1. Hand-code their own register machine, and demonstrate a grasp of the formal foundations of computation.
2. Understand differing accounts of "intelligence" derived from and operating within cognitive science, developmental psychology, computer science, and robotics.

3. Understand the historical roots, trends and challenges faced by paradigms in AI research and development.
4. Understand the social and moral risks posed by the widespread and undertheorized adoption of AI assistants and algorithms.
5. Demonstrate an ability to charitably engage with interdisciplinary sources from a range of media types (including video, technical reports, publications and books).
6. Demonstrate an ability to write in a succinct, clear, and charitable manner either via two essay based tests or a short research paper.

Course Prerequisites:

Prerequisite / Co-requisite: AP/PHIL 2160 3.00 or AP/PHIL 2240 3.00 or permission of the instructor.

Technical Requirements for Taking the Course:

This course is designed to be in-person, as such we expect meetings to take place in-person and on-campus throughout the Winter semester. Students should be prepared for the possibility that the some or all meetings in the course become fully remote, with meetings being hosted on Zoom, as circumstances require. In order to fully participate in Zoom meetings, students should ensure that they have access to a stable, higher-speed internet connection, as well as a computer with a webcam and microphone, and/or a smart device with these features. Students should bring such a device to class as well, in order to complete in-class, online hosted assignments as needed.

All assignments except for in-class exams, the schedule of readings, course announcements, and course policies will be posted to the eClass site for this course. All assignments—including in class quizzes—will either be submitted via eClass, a York U hosted Qualtrics link created by the instructor, or completed within the eClass interface (e.g. posting on discussion forums). The instructor will also use eClass as their primary way of communicating with students.

Students are expected to follow the most up-to-date safety protocols laid out by the Province and by York University, including completing daily screenings. For more information, see: https://www.yorku.ca/bettertogether/students/. Students should be prepared for the possibility that class be held fully remotely, via Zoom, depending on the evolving circumstances of the COVID-19 pandemic.

Course Requirements:

Each week will feature at least two primary readings. These are often difficult texts and will reward multiple reviews. Expect to spend at least two hours outside of class on these texts per week.

Weekly quizzes: 33% Ten short answer quizzes (and one homework assignment) will be administered at the beginning of class – via an eClass or Qualtrics link. Each assignment is worth 3% of your total grade. If you have a valid excuse (e.g., illness, religious observance etc.) you will be provided a secondary link to a similar quiz to be completed within an agreed upon timeframe. You are still responsible for knowing the material covered in class. Each quiz will have a series of short and essay questions that you'll be asked to answer based on that week's reading. You must be in class to complete the quiz, unless otherwise noted.

Midterm Exam: 30%
The midterm will be held March 3rd. The exam will feature a series of short-answer and lengthier essay questions. No aids (e.g., notes, books, electronic devices etc.) are permitted unless otherwise specified by the University's Student Accessibility Services ("SAS"). Any student seeking accommodations (e.g., extra-

time, alternative test-location etc.) must have communicated with SAS and completed their required procedures *prior* to the test date.

Final Paper xor Final Exam: 30%
In this class you will have a choice, to be made March 17th, as to whether you will complete a final, IBAC styled, paper (2000-3000 words) or complete a final, in class essay based exam. The paper will be broken up into a series of graded components, late papers may be docked 5% of the total paper grade each day they're late. You cannot complete both for a grade.

Paper Option:
I will pass around topics and a rubric on March 10th. Topics and abstracts will be due March 24th (5%). Your background section (1000 words or so) will be due March 31st (5%), your analysis section will be due the following week (April 7th) (5%). Your final paper will be due on or around the same day as the final exam (TBD) (15%).

Exam Option:
The exam will be in-class, no-notes or books, on the final exam date (TBD). It will be similar to the midterm exam (2hrs).

Participation and Attendance: 7%
Students are expected to attend each class and participate in the discussion sections. Participation will be noted.

Grading:

Weekly quizzes will be graded as Exemplarily, Pass, or Fail, with a corresponding point reduction. The chart below details how to convert your overall weekly quiz grade to letter grades for that portion of the course:

"A+"	"A"	"B+"	"B"	"C+"	"C"	"D+"	"D"	"E"	"F"
31-33	27-30	23-26	20-22	16-19	13-15	10-12	7-9	4-6	< 4

Exams and paper components will be graded letter-wise (i.e., A+, A, B+, B etc.). For the purposes of estimation, when converting from a ordinal (letter) scale to a continuous (percentage) scale, *I* use a truncated measure that corresponds more closely to US American standards (e.g. A+ = 95% or greater, A = 90-94.9%, and so forth) rather than York's bizarre percentage-conversion guidelines (where an A can be anywhere from an 80-89 but a B only covers half the range [i.e. 70-74]). Bear in mind that percentages are estimates, *letter grades* are what are entered into York's system at the end of the semester.

If you wish to estimate your final grade, convert respective letter grades into a ballpark percentage score, add the scores together and divide by three, then reconvert the result into a letter grade using the same scale as before. This is, of course, only a rough estimate as it would not include your participation grade.

Here's an example below that uses York's typical percentage-conversion guidelines:

Weekly Quiz Score	**Midterm**	**Final XOR Paper**	**Estimated Grade**
25 → B+	A	B	
77	85	73	77+85+73 = 78.3rep

Conversion: B+ ish
Depending on participation grade it could stay as a B+ or become a B or an A

"Wait, why can't our work just be graded along a percentage scale?" This is good question, but ultimately, as a third-year philosophy course, our task is not to memorize a series of facts or formulae, but rather to identify, critique and build written arguments that are often more than the sum of their parts. As such it's not feasible or even pedagogically sound to assign a *quantitative* score to most work in this course. What makes one argument or essay a 90% instead of a 85% or "5" instead of a "6"? Why not a "5.5"? Why not a "5.52"? and so on. Why is one component (e.g., have a thesis statement) worth 1 point and not 2, or three?

Rather, if it helps, you can think of the letter grades as follows: "A+" Excellent, really stand-out work. "A" Very good, no obvious flaws or errors. "B+" Entirely satisfactory, but missing a few details. "B" satisfactory, but missing some key details or elements. "C+" Not entirely satisfactory, though perhaps shows genuine effort. "C" borderline satisfactory, barely meets the criteria set out. "D" incomplete or otherwise marginally acceptable etc.

Policies:

- I only allow incompletes (DEF grades) for genuinely extraordinary reasons.
- I do not provide extra-credit opportunities.
- Keep in mind that I may not respond to your email outside of business hours (e.g., in the evening, over holidays, or during weekends).
- **Plagiarism and Cheating**: Under no circumstances will plagiarism or cheating be tolerated.

Attendance: Attendance is mandatory, with exceptions due to illness, permitted excuses as required by university policy, and failure to pass the YU screen daily screening. In the event you cannot attend in person class, you must still complete the daily quiz during the start of our class time, as it will be made unavailable shortly afterwards. If you continuously miss class, I may contact you to discuss options for your future in the course. Attendance and active engagement are the keys to successfully completing the course.

Exam makeups: Alternative arrangements and makeups will not be allowed, except as required by university policy: e.g., if the course is required to move online, alternative arrangements (e.g., an essay exam delivered via eClass or Qualtrics) will be devised.

Assignment Submissions: Assignments will be submitted via eClass or qualtrics. Additionally, I may request that you print out and hand in hardcopies of your papers and their subcomponents at the relevant point in the semester.

Papers: Late papers may be accepted at the instructor's discretion, if there are circumstances calling for an extension. Papers turned in late without reasonable mitigating circumstances may incur a half letter grade reduction in the overall grade. It best that you contact me as soon as possible if you're going to be late (preferably before the due date).

Contacting the instructor: You should contact the instructor via email or the eClass direct message function. Please note that we will be responding to messages within 24 hours during working days (i.e. not on weekends or holidays).

Academic honesty and integrity: In this course, we strive to maintain academic integrity to the highest extent possible. All submitted coursework must be an expression of the student's own understanding and ideas. Please familiarize yourself with the meaning of academic integrity by completing SPARK's Academic Integrity module at the beginning of the course. Breaches of academic integrity range from cheating to plagiarism (i.e., the improper crediting of another's work, the representation of another's ideas

as your own, etc.). All instances of academic dishonesty in this course will be reported to the appropriate university authorities, and can be punishable according to the Senate Policy on Academic Honesty.

Turnitin: To promote academic integrity in this course, students will be normally required to submit their written assignments to Turnitin (via the course eClass) for a review of textual similarity and the detection of possible plagiarism. In so doing, students will allow their material to be included as source documents in the Turnitin.com reference database, where they will be used only for the purpose of detecting plagiarism. The terms that apply to the University's use of the Turnitin service are described on the Turnitin.com website.

"Closed-book" policy: It is expected that students will complete in-person quizzes and exams on their own without help from any other person, and without access to notes, articles covered in class, or other materials that have been provided over eClass. These quizzes and exams are designed so that you can complete them without referring to minutia in the text, and doing so with consume valuable time that would be better used in thinking through the essay and answer prompts. In the event that the course is moved online, quizzes and exams may be "open book" at the instructor's discretion. In this event, more information will be provided. Intellectual property: All course material (this syllabus, power points, assignments, paper rubrics, etc.), except the outside assigned articles, is the intellectual property of the course instructor and cannot be reproduced in any way without my permission. Assigned articles are the intellectual property of their respective copyright holders and usually cannot be reproduced or posted publicly.

Course recordings: Any recordings for this course should be used for educational purposes only and as a means for enhancing accessibility. Students do not have permission to duplicate, copy and/or distribute the recordings outside of the class (these acts can violate not only copyright laws but also FIPPA).

Student conduct: All students are expected to treat their fellow students and the instructor with respect and charity, both in class in person and on any of our online platforms. Especially through mediums like Zoom and the course eClass, no form of harassment, trolling, or disrespect will be tolerated.

Student Accommodations: We are committed to fairly accommodating students with disabilities. Please contact the instructors and Student Accessibility Services (https://accessibility.students.yorku.ca/) as soon as possible, and we will all work together to find a fair accommodation. Note that in addition to sending the letter, accommodations for individual assignments must specifically be requested well ahead of the assignment's deadline.

When we zoom:

Zoom sessions: For those attending the live Zoom classes, you are not allowed to take any screenshots or recordings of any kind. This is to respect the privacy of your fellow students. Recordings of the lecture portions (which only involve myself) and my slides will be made available on eClass. You also do not have permission to reproduce any lecture recordings on any platforms or websites outside of eClass.

Video policy: I understand that some of you might have privacy concerns about using your video during Zoom classes, and that others might have limited internet bandwidth that might make live-streaming difficult. Therefore, it is not required that students have their videos on if meetings must be moved online. That being said, <u>having video cameras on can make a big difference in terms of creating a sense of community within our class</u>, so that we feel more like a normal, in-person classroom. For this reason, I do very sincerely encourage you to use your video as much as possible if privacy or tech issues are not preventing you from doing otherwise. If you do need to keep your video off, please be sure to put a profile photo on your Zoom account, so that your classmates feel like they are talking to a person, rather than a black box with a name in it.

Zoom Chat policy: If meetings are held on Zoom, the option to privately message other students in the chat will be disabled. Messages in the public chat should be respectful and stay on point.

Texts

Papers and readings will be posted on our eClass site. It is your responsibility to check the eClass site for new readings as they are made available. Please read readings ahead of the schedule class time.

READING LIST DRAFT 1 [1/7/2022]

	Date	Topic	Reading	A. 1	A. 2
1	13-Jan	syllabus day: *what does it mean to be artificial* **No class – Prerecorded video to be Watched!**	**Corbyn**, Z. (2021) "Microsoft's Kate Crawford: 'AI is neither artificial nor intelligent" ***Watch prerecorded video (to drop ~1/11/22) and complete Quiz on Qualtrics.***	Q1	-
2	20-Jan	**Defining programs and computation.** **On Zoom** (links on eClass)	**Carter, M.** Chapter 7 "Formal Systems" & Chapter 8 "Computability" *optional:* Chapter 9 "Universal Machines"	-	*Homework assignment -- counts as Q2*
3	27-Jan	**What counts as intelligence?** **On Zoom** (links on eClass)	**James, W.** (1890/1918). "The scope of psychology" **Krakauer, J.** (2019) "The intelligent reflex" **Fodor, Jerry** (1985). "Precis of *Modularity of Mind*" 1-5.	Q3	
4	3-Feb	**Turing Machines and GOFAI** **On Zoom** (links on eClass)	**Haugeland, J.** (1989) "Computer Architecture" (selections) *Video: "Algorithms: The Secret Rules of Modern Living" **Crane, T.** (2003)"Computers and Thought"	Q4	
5	10-Feb	**Wither (gof)AI?**	**Dryefus, H.** (1992) Introduction to "What Computers Still Can't Do" (selections) **Brooks, R**. (1991)"Intelligence without Representation" **Gopnik, A.** (2019) "What babies tell us about AI"	Q5	
6	17-Feb	**Plural approaches to intelligence(s)**	**Dennett, D**. (1991) "Real Patterns" **Gopnik, A.** (2017) "An AI that knows the world like children do" **Clark, A**. (1997) "The situated infant"	Q6	
	25-Feb	no meeting	*Spring Break*		
7	3-Mar	*Midterm*	*Midterm*	Q7 Midterm evaluation	
8	10-Mar	**Connectionism and Machine Learning**	**Buckner, C.** (2019) "Deep learning: A Philosophical Introduction"	Q8	*Intro to paper*

			Sinton, C. (2020) "From Implausible Artificial Neurons to Idealized Cognitive Models"		*writing & topics*
9	17-Mar	**Algorithmic Biases**	**Sinton, C.** (2021) "The Dark Past of algorithms that associate appearance and criminality" **O'Neil, C.** (2016) Weapons of Math Destruction (selections) **McGlotten, S.** (2014) "Black Data"	Q9	*Decide between Paper or Final Exam*
10	24-Mar	**Feminized AI**	**Read, H. et al.** (*ms*) "A please for integrated empirical and philosophical research on the impact of feminized AI workers" **Gomez-Lavin, J et al** .(*ms*) "Measured interactions of feminized AI and task allocation". **Alesich, S. & Rigby, M.** (2017) "Gendered Robots: Implications for our humanoid future"	Q10	*Topic / Abstract Due*
11	31-Mar	*no class*	*No class – Instructor at SSPP Conference, background section still due if completing final paper*	-	*Back-ground section due*
12	7-April	**AIs and Knowledge Production**	**First Nations Information Governance Centre** (2005)"Ownership, Control, Access, and Possession (OCAP) or Self-Determination Applied to Research" (*selections*) **Mohamed, S. et al.** (2020) "Decolonial AI: Decolonial Theory as Sociotechnical foresight in AI" **Maitra, S.** (2020) "AI and Indigenous perspectives: Protecting and empowering *Intelligent* Human beings"	Q11	*Analysis Section Due*
-	TBD		*Final exam (TBD)*		*Paper due*