Phil XX: Ethical Problems and Artificial Intelligence

Instructor: Joshua Kissel
Joshuakissel2014@u.northwestern.edu
Office Hours: XXX

Class Meeting: XXX
Location: XXX

Course Description:

This is an intermediate to advanced level philosophy course in ethics and social-political philosophy for majors and nonmajors. The objective of this course is to introduce you to an aspect of a vibrant and development subfield in philosophy of technology dealing with the ethical issues related to the development of Artificial Intelligence.

This course deals with normative questions related to and arising from present and (likely) future developments in 'Artificial Intelligence' (AI) including 'Artificial General Intelligence' (AGI), the sort of AI that approaches and surpasses human capacities in reasoning and intelligence. The first half of the course explores ethical and political ramifications of AI for human beings as an object of moral reasoning. The second half of the course explores AI as subjects of moral reasoning exploring the possibility that we might owe things to AI.

In particular we will explore how AI might (i) alter the labor market and change the distribution of wealth and work, (ii) ut human existence at risk, perhaps because they become superintelligent and surpassing human beings in rationality, or (iii) be objects of relationships of friendship or sex with human beings. Our attention will then shift to AI as subjects of morality. We will explore whether (i) AI might have sentience that garners them moral standing, or (ii) if they might be moral persons capable of the full gamut of moral life. (iii) We will explore whether AI deserve legal or political rights and (iv) whether they might be able to live a good life at all. (v) We'll also investigate the relationship between morality and rationality and (vi) whether AI might be good moral agents like you or I hope to be.

It is expected but not required that you will have some background in philosophy already. Please contact me directly if you are unsure if you are prepared for this course.

Course Objectives: this course enables students to:

(1) Understand and differentiate AI, AGI, and robotics
(2) Evaluate ethical and social risks posed by AI
(3) Understand and differentiate degrees of moral status (e.g., moral patienthood or personhood)
(4) Consider the details of a particular AI and determine whether it/they should be regarded as bearers of moral status of a particular kind.
(5) Assess the justifiability of a variety of ethical claims related to AI
(6) Be able to apply ethical thinking to a wider class of cases beyond AI (especially facets of environmental and animal ethics, bioethics, aspects of political philosophy)

In addition, students will acquire a background in important areas of philosophy including; critical reasoning, normative ethical theory, political philosophy, philosophy of science and technology, philosophy of mind, the scope of ethics and moral status, and applied questions dealing with many of these areas.

Office Hours:

During my office hours I will be sitting quietly behind a desk, staring at a wall unless students come to meet with me. This time is meant for you to ask questions, discuss philosophy, or just hang out. It is *your* time, and you do not need an excuse or any clarity about what you want to do. I request, but do not require, that you send me an email alerting me to when you want to come, and if you happen to know, what you plan to discuss.

If for whatever reason you cannot make my regular office hours, please send me an email asap with a range of time that you could meet, and we will try to work something out!

Absences:

I trust all of you to make rational decisions with respect to attendance in accord with your own best reasons. You are each **permitted 2 totally unexcused absences** without any requirement to email or in any other way alert me to your absences. You can use these absences to miss class for any reason (e.g., your being sick, tired, wanting to binge a new show or play your favorite video games, to attend some internship or work-related activity, or whatever else.)

Absences beyond this number will amount to a $1/20^{th}$ reduction in your participation grade for this class.

I try not to allow any extra excused absences beyond your freebies. However, I encourage all students to reach out if you run out of freebies but believe you have some special excuse (like a health issue) that might warrant extra accommodation or special exemption without requiring me to disadvantage your peers in this class. These *can and do* sometimes happen.

Screen Policy*:*

This class is a screen-free environment. This means no computers, tablets, phones, or other such devices. This is because I have found students participation and discussion is best when they are undistracted by their own screens or those of their peers. I share all my PowerPoints to make note-taking less laborious. If you violate the policy, you may be marked absent for the day. Special accommodations are exempted. [E.g., medical exemptions.]

Students with Disabilities*:*

Any student needing accommodations should speak directly to AccessibleNU ((847) 467-5530 or accessiblenu@northwestern.edu) and to me as early as possible in the quarter. Be aware that AccessibleNU will help arrange reasonable accommodations for both physical and mental health concerns. Barring unforeseen circumstances, any necessary arrangements should be made within in the first week of class. All discussions will remain confidential.

General Grading Schema:

1. Participation and Attendance 15 % of total. Pass/Fail
2. Ten Reading Responses: 10% of total. Check + (100)/Check (92)/Check -(85)
3. Two Scaffolding Paper Outlines: 30% of total. A-F
    a. First 15%
    b. Second 15%
4. Paper Prospectus and Peer Reviews 15% of total. 'A'-'F'
5. Final Paper (1500-1750 words) 30% of total. 'A'-'F'

A	A-	B+	B	B-	C+	C	C-	D	F
94-100	90-93	87-89	84-86	80-83	77-79	74-76	70-73	60-69	60-0

Assignment Due Dates [Details for Particular Assignments to Come]

Assignment	(Some) Details	Due Date [Examples]
Reading Responses:	Each response is due on Canvas 2 hours before the relevant class takes places. And each must be on a different week. EX: Responses to Srinivasan must be submitted *before* our discussion of her paper on week 5	Through-out
First Paper Outline	This outline must be on a topic from week 1 or 2	Early-Mid
Second Paper Outline	This outline must be on a topic from class 4-6	Mid-Late
Final Paper Proposals and Peer Reviewing	This assignment includes a paper proposal (due after week 9 discussion) and your peer review feedback	Late
Final Paper	Paper of 1500-1750 words on any topic in this course. You may choose to use any paper outline or to start from scratch	End

General Course Outline:

Based on a course with 15; one hour and 50-minute sessions meeting three times a week for 5 weeks. Naturally, the course might be altered to meet less frequently over a longer period.

Class	Topic	Readings	Assignments
1 – Week 1 Introduction to Course	Validity and Soundness Thought Experiments	WATCH: "The Simulation Argument" What Will Future Generations Condemn Us For? – Kwame Anthony Appiah (3 Pages) Class Activity https://www.moralmachine.net/	Icebreaker Discussion

2 – Week 1: Introduction to Course	What Is AI?	Selmer Bringsjord and Naven Sundar Govindarajulu – Artificial Intelligence (41 Pages, Long and Dense Survey Article, Start Early!)	
3 – Week 2 Introduction to Course	Artificial Intelligence vs Artificial General Intelligence	Nick Bostrom and Eliezer Yudkowsky - The Ethics of Artificial Intelligence (20 Pages)	
4 – Week 2 Introduction to Course	Social and Political Wrongs	Iris Marion Young – Five Faces of Oppression (20 Pages)	
5 – Week 3: The Effects of AI on US	AI, Automation, and the Labor Market Film Discussion	Distributive Justice Game (Based on John Rawls) Film: Blade Runner (1982)	
6 – Week 3: The Effects of AI on US	AI and Work	Pegah Moradi and Karen Levy - The Future of Work in the Age of AI (20 Pages) Video: Kurzgesagt In a Nutshell – The Rise of Machines – Why Automation is Different this Time	Early-Term Teaching Evaluations Circulated
7 – Week 4: The Effects of AI on US	Existential Risk	Nick Bostrom – Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards (20 Pages) Video: Sam Harris 'Can We Build AI Without Losing Control Over It'	
8 – Week 4: The Effects of AI on US	Superintelligence as Existential Risk	Eliezer Yudkowsky – Artificial Intelligence as a Positive and Negative Factor in Global Risk (43 pages)	Paper Outline 1 Due End of Week 4
9 – Week 5: The Effects of AI on US	Might AI not be So Bad, or even Good?	Amia Srinivasan – Stop the Robot Apocalypse (10 Pages) Kurzweil – Excerpts from the Book 'The Singularity is Near' (2005)	
10 – Week 5: The Effects of AI on US	Sex Robots	Kathleen Richardson – The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots (6 Pages) Sven Nyholm and Lily Eva Frank – From	

		Sex Robots to Love Robots: Is Mutual Love with a Robot Possible? (20 Pages)	
11- Week 6: The Effects of AI on US	Love and Friendship Film Discussion	John Danaher - The Philosophical Case for Robot Friendship (19 Pages) Movie: Her (2013)	
12 – Week 6: AI as Moral Subjects	Patients and Persons	Peter Singer – 'Equality for Animals' from *Practical Ethics,* 3rd Edition (22 Pages) Mary Anne Warren – On the Moral and Legal Status of Abortion (19 Pages)	Midterm Teaching Evaluations Circulated
13 – Week 7: AI as Moral Subjects	Personhood	Christine M. Korsgaard – Fellow Creatures: Kantian Ethics and Our Duties to Animals (33 Pages)	
14 – Week 7: AI as Moral Subjects	Rights for AI?	Eric Schwitzgebel and Mara Garza – A Defense of the Rights of Artificial Intelligences (19 Pages) Bernard Williams – 'The Human Prejudice' from *Philosophy as a Humanistic Discipline* 132-52 (17 Pages)	
15 – Week 8: AI as Moral Subjects	Can AI Live a Good Life?	Martha Nussbaum – *Women and Human Development: The Capabilities Approach*, Excerpts from I.IV 'Central Human Capabilities' (70-86, 16 Pages)	
16 – Week 8:	Morality and Rationality	Stuart Armstrong – General Purpose Intelligence: Arguing the Orthogonality Thesis (20 Pages)	Paper Outline 2 Due End of Week 8
17 – Week 9: AI as Moral Subjects	Can AI Be Moral Agents?	Regina Rini – Creating Robots Capable of Moral Reasoning is like Parenting (5200 Words)	
18 – Week 9: Wrapping Up	Paper Proposals Film Discussion	Paper Proposal Discussions Film: Ex Machina (2015)	Bring Paper Ideas to Workshop in Class
19 – Week 10: Wrapping Up	Make Up	Make Up Session	End of Term Teaching Evaluations
20 – Week 10: Wrapping Up	Wrap Up	Course Recap and Final Paper Workshops	Final Paper Drafts Due Final Paper Due XX Date