The Ethics of Artificial Intelligence

Summer 2022

Professor

Eric Sampson
Email: sampsone@rhodes.edu
Office: Clough 401
Office hours: T, TH (2 - 3pm, or appt./Zoom)

Meeting Details

Days: T, TH
Time: 12:30 – 1:45pm
Place: Buckman Hall 325
Course: PHIL 250-01

Course Description

AI technology has the potential to dwarf the impact of other revolutionary technologies such as the printing press, electricity, antibiotics, and the internet. But AI is developing so quickly there has been little time to reflect on the nature, scope, and (dis)value of that impact. This has given rise to a host of pressing moral questions that we are only beginning to consider (let alone answer). Among the many such questions we'll consider in this class are these: How might AI transform the world for unimaginable good? How might it pose an existential threat, and what can we do to mitigate it? Should governments attempt to regulate the development of AI, and if so, how? Can AI make moral judgments? If so, what moral judgments should we program them to make? And how do we avoid programming our own biases and moral failings into them? Could AI become conscious, and if so, what (if any) moral obligations might this impose on humans? For instance, can AI have rights, interests, or welfare? Could I merge with a superintelligent AI, becoming superintelligent myself? What would that even mean, and would it be morally OK for me to do it? Could I befriend, fall in love with, or even have sex with, an AI? If so, *should* I? Will AI lead to mass unemployment, and if so, what should be done for those who are left jobless? How might AI be used by militaries, governments, employers, and others with interests in surveillance and what (if any) moral obligations might this impose on those with the technology? Finally, how can AI be used to capture our attention and engagement and what obligations (if any) do we have to resist such attempts?

Texts

There is no textbook for this course. All readings are available on the course website.

Course Objectives

- Develop ability to recognize and describe important positions and arguments in the history of ethics
- Develop and exercise a capacity to represent sympathetically, and critically evaluate, arguments for and against your own treasured moral and political views
- Develop and exercise the ability to articulate *your own* ideas in writing and speech

Course Requirements

• Participation	10%	
• Reading Quizzes	15%	(In-class throughout semester)
• Midterm Take Home Exam	20%	Due: Sun., Oct. 17, 11:59p (Canvas)
• Argument Analysis Paper	20%	Due: Tue, Nov. 23, 11:59pm (Canvas)
• Final Exam	20%	Time: Mon., Dec. 13, 1pm (Canvas)

Participation

Participation begins by reading the assigned readings carefully *before class*. You'll then need to contribute to class discussion, at some point, by asking questions or making comments. Some people are shy. I get that. Shy people can either rack up their participation points on the back half of the semester once they become more comfortable with the class setting, or by coming to office hours, or by chatting with me about course material over email or after class. Making lots of comments in class is not the only (or even best) way to receive a good participation grade. Quality matters too. The best thing to do is strike a nice balance between quality and quantity. Sometimes you'll say stuff that doesn't quite make sense. That's fine. Philosophy is hard and you're allowed to struggle. Feel free to contribute even if you're not 100% clear about what's going on. There's no penalty for making a good-faith effort but not quite getting it right. That's how you learn literally anything—trying and failing a bunch until you get it.

Attendance

Attendance is expected. You can miss three (3) course meetings without any notice and without penalty. Each absence beyond those three will result in a 2-point deduction from your participation grade. (Obviously, if you get COVID or something, and can't make it to class for a long time because of illness, I'm not going to tank your grade.)

Grading Scale

A: 94 – 100	A-: 90 – 93	B+: 87 – 89	B: 84 – 86	B-: 80 – 83	C+: 77 – 79
C: 74 – 76	C-: 70 – 73	D+: 67 – 69	D: 64 – 66	D-: 60 – 63	F: < 60

Office Hours & Accessibility

I'm happy to meet with you at any time to discuss assignments or simply to talk more about the topics of the class. Come to my office hours, or if those times don't work, email me to set up an appointment. Zoom works too.

I'm committed to making class fully accessible regardless of disabilities. If I can do anything to help make the class more accessible to you, let me know, or (if you would prefer) have the Accessibility Office contact me on your behalf.

Plagiarism

Do us all a favor and don't plagiarize. Plagiarism is the representation of another's words, thoughts, or ideas as one's own without attribution in connection with submission of

academic work, whether graded or otherwise. If you quote something, put it in quotes and cite it using whichever citation convention you like. If you use someone's ideas, cite them and put the idea in your own words. If you have questions about what constitutes plagiarism, talk to me (by email or whatever) and I'll be happy to help.

Course Schedule

The Ethical Toolbox

Week 1 — What's AI? What's Ethics? What are we doing here?
Shafer-Landau, "Consequentialism"

Week 2 — Shafer-Landau, "Kantian Ethics"
Shafer-Landau, "The Ethic of Prima Facie Duties"

The Basics of AI Ethics

Week 3 — Bostrom & Yudkowsky, "The Ethics of Artificial Intelligence"
Future of Life Institute, "Benefits and Risks of Artificial Intelligence"
World Economic Forum, "Top 9 ethical issues in Artificial Intelligence"

The Singularity

Week 4 — YouTube: Can We Build AI Without Losing Control of It?
Bostrom, *Superintelligence*, Chs. 2-6
Chalmers, "The Singularity: A Philosophical Analysis"

The Value Alignment Problem

Week 5 — YouTube: What happens when our computers get smarter than we are?
Bostrom, *Superintelligence*, Chs. 7-8, 12
Yudkowsky, "Alignment for Advanced Machine Learning Systems"
Wallach & Vallor, "Moral Machines: From Value Alignment to Virtue"

Racist AI (and other biases)

Week 6 — YouTube: The era of blind faith in big data must end
Hudson, "Technology is Biased Too. How do We Fix It?"
Castro, What's Wrong with Machine Bias?

Crime Prediction, Prevention, and Punishment

Week 7 — YouTube: How Cops are Using Algorithms to Predict Crimes
Surden, "Ethics of AI in Law: Basic Questions"
Barabas, "Beyond Bias: Ethical AI in Criminal Law"

Autonomous Weapons

Week 8 YouTube: Killer Robots in War and Civil Society
Sparrow, "Killer Robots"
Scholz & Galliott, "The Case for Ethical AI in the Military"
Asaro, "Autonomous Weapons and the Ethics of Artificial Intelligence"

Autonomous Cars

Week 9 YouTube: Are We Ready for Driverless Cars?
Nyholm, "The Ethics of Crashes with Self-driving Cars: A Roadmap, I"
Nyholm, "The Ethics of Crashes with Self-driving Cars: A Roadmap, II"

Robot Sentience and Personhood

Week 10 Kingwell, "Are Sentient AIs Persons?"
Schneider, "How to Catch an AI Zombie: Testing for Consciousness"
Schneider, "Could You Merge with AI?"

Robot Rights

Week 11 YouTube: A.I. Ethics: Should We Grant Them Personhood?
Basl and Bowen, "AI as Moral Right-Holder"
Liao, "The Moral Status and Rights of Artificial Intelligence"
Schwitzgebel & Garza, "Designing AI with Rights, Consciousness, Self-Respect, and Freedom"

Falling in Love with AI

Week 12 YouTube: Sex Robots
Devlin, "The Ethics of the Artificial Lover"
Danaher, "Sexuality"

Automation, Mechanization, and the Future of Work

Week 13 YouTube: Humans Need Not Apply
YouTube: Will Automation Take all our Jobs Away?
Moradi & Levy, "Future of Work in the Age of AI: Displacement or Risk-Shifting?"
James, "Planning for Mass Unemployment: Precautionary Basic Income"

AI and the Attention Economy

Week 14 Castro & Pham, "Is the Attention Economy Noxious?"
Castro & Aylsworth, "Is There a Duty to Be a Digital Minimalist?"