Minds, Machines, and Persons

Mike Deigan

Rutgers University   01:730:329 01

Spring 2021

mike.deigan@rutgers.edu

https://mikedeigan.com/teaching/329

Description

Could we ever create intelligent machines? Should we worry that they will become more intelligent than us? Would they be conscious? Would they have rights? Could we upload our own minds onto machines? Usually questions like these are left to the realm of science fiction, but in this course we will attempt to explore them in a rigorous way, using philosophical tools and ideas from philosophy of mind and ethics.

The course type is "By Arrangement" and so has no single set meeting time, but it will still be primarily discussion-based rather than lecture-based. Each week you will meet on Zoom with a small group of your classmates, and most of the other assignments are designed to ensure that everyone is well prepared for these conversations. The aim is to help you think through the questions of the course together and to develop the important skill of having focused discussions that are both critical and collaborative.

Prerequisites

At least one previous course in philosophy.

Office Hours

I will be in the class Zoom room at preannounced times each week for discussions with whoever wants to drop by. I enjoy talking with people about philosophy, so please feel welcome to come to these, even if you don't have any specific thing you feel you need to ask me about. You don't need to schedule a meeting to come to these.

If you need to meet with me one-on-one, you can either email me to find some time that works for us both, or you can schedule an appointment online for for some time on Tuesdays from (9:30–10:15 EST) here: http://mikedeigan.com/meet
One-on-one meetings will need to be scheduled at least a day in advance.

Assignments

1. *Readings & Perusall Discussion:* there will be readings assigned for each week, typically around 30 pages (I will also post brief videos giving some background and orientation). You will be required to read these carefully and discuss them on Perusall, an online social annotation platform accessed through Canvas.

2. *Group Meeting Participation*: at the beginning of the semester, you will be assigned to a group that will meet weekly over Zoom to discuss one or two issues related to the week's topic. You are required to attend and participate in these meetings. I will join your group for about half of them.

3. *Pre-meet Write-ups:* prior to each meeting, you will write a question you would like answered and a point you would like to make in the meeting. These will be shared with your group to help make a meeting agenda.

4. *Post-meet Reports*: after each meeting, you will write a brief report of the group discussion and what you've taken away from it.

5. *Argument Stress Tests*: over the course of the semester you will write two 1,000 word papers putting pressure on the weakest point in a philosophical argument and determining how best to strengthen the argument.

Grading

Perusall discussion:	25%
Pre-meet write-ups:	10%
Post-meet write-ups:	20%
Meeting participation:	20%
Argument Stress Test #1:	10%
Argument Stress Test #2:	15%

Each week your meeting participation and pre-meet write-ups will be graded for completion, and your Perusall participation and post-meet reports will be be graded on the following scale:

Excellent:	100 points
Good:	90 points
Acceptable:	80 points
Unacceptable:	60 points

One lowest grade in each category will be dropped, and your final grade for each will be the average of the rest. Argument Stress Tests will receive grades on a 100 point scale.

Rubrics for each assignment type will be distributed in advance.

Books

I will post all the readings on the course webpage; no book purchases required.

Policies

Late Policy:

For papers and post-meet write-ups, I'll take off two points for each day after the deadline it is turned in. No late credit for Perusall or pre-meet write-ups.

For all deadlines, I'm willing to be reasonably flexible. If you need an extra couple days because of other commitments or illness, just let me know a few days in advance and I don't mind giving an extension, even for non-emergency situations. If you need a longer extension or a last minute extension, then there needs to be some special and unforeseeable circumstance justifying it.

Schedule

The assigned readings are subject to change. I will notify you of any such changes at least one week in advance and post an updated syllabus on the course webpage.

Can a machine be intelligent? How could we tell?

Week 1   01/19   Turing (1950) "Computing Machinery and Intelligence"
                   Shannon & McCarthy (1956) *Automata Studies*, preface

Week 2   01/25   Shieber (2016) "Principles for Designing and AI Prize, or Why the
                   Turing Test Fails as an Inducement Prize "
                   Legg and Hutter (2007) "Universal Intelligence: A Definition of
                   Machine Intelligence" (selections)

Superintelligence and the Control Problem

Week 3   02/01   Grace et al. (2018) "When Will AI Exceed Human Performance?
                   Evidence from AI Experts"
                   Chalmers (2010) "The Singularity: A Philosophical Analysis" (§§1–4)
                   Chollet (2018) "The implausibility of intelligence explosion"

Week 4   02/08   Chalmers (2010) "The Singularity: A Philosophical Analysis" (§§5–7)
                   Russell (2019) *Human Compatible* (selections)

Week 5   02/15   Russell (2019) *Human Compatible* (more selections)
                   Gabriel (2020) "Artificial Intelligence, Values, and Alignment" (selection)

The moral status of machines

Week 6   02/22   Bryson (2010) "Robots Should Be Slaves"
                   Schwitzgebel and Garza (2020) "Designing AI with Rights, Consciousness, Self-Respect, and Freedom"

Week 7   03/01   Warren (1997) *Moral Status: Obligations to Persons and Other Living
                   Things*, Chs. 1, 5
                   Kagan (2019) *How to Count Animals, More or Less* (selections)

Week 8   03/08   Paper Week (no assigned reading)

          03/13   Argument Stress Test #1 Due!

          03/15   Spring break!

Intentionality

Week 9	03/22	Searle (1980) "Minds, Brains, and Programs" and commentary by Fodor, Lycan, McCarthy, Pylyshyn
Week 10	03/29	Millikan (1993) *White Queen Psychology and Other Essays For Alice* Introduction, Ch. 4
		Dennett (1988) "Evolution, Error and Intentionality"

Consciousness

Week 11	04/05	Chalmers (2003) "Consciousness and Its Place in Nature"
		Churchland (1996) "The Hornswoggle Problem"
Week 12	4/12	Carruthers (2019) *Human and Animal Minds: The Consciousness Questions Laid to Rest* (selections)
Week 13	4/19	Carruthers (2019) *Human and Animal Minds: The Consciousness Questions Laid to Rest* (more selections)
		Schneider (2020) "How to Catch an AI Zombie: Testing for Consciousness in Machines"

Uploading and Personal Identity

Week 14	4/26	Chalmers "The Singularity: A Philosophical Analysis" §§8-11
		Parfit (1986) *Reasons and Persons* (selections)
		Wolf (1986) "Self-Interest and Interst in Selves" (selections)
	05/03	Reading Week
	05/10	**Argument Stress Test #2 due**