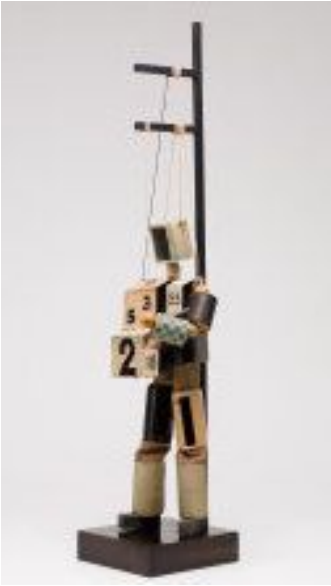
Minds and Machines



The Philosophy of Artificial Intelligence

GESM 120, University of Southern California, Spring 2017

Morning Section
Tuesday and Thursday 11am–12:20pm
Waite Philips Hall (WPH) 206

Afternoon Section
Tuesday and Thursday 2–3:20pm
Grace Ford Salvatori Hal (GFS) 229

Instructor

Jeff Sanford Russell
jeff.russell@usc.edu

Office Hours
Stonier Hall 227
Thursdays 12:30–1:30pm or by appointment

Course Description

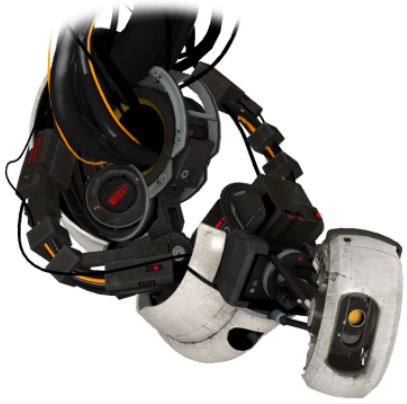
The current world champions at chess, Jeopardy, and go are computers. Computers sort your email, plan air traffic, diagnose diseases, find distant planets, recommend movies, compose music, and drive cars. What else could they do? Will computers ever think, or be creative, or feel sadness or love? How different are human minds from electronic computers? Could we become immortal by uploading ourselves to the cloud? Could robots take all our jobs, or rise up and kill us all? What should we do about these possibilities?

This seminar will be a full-brain workout. You will read philosophy and fiction. You will write an essay, and you will write your own computer program. You will ponder creativity, consciousness, and morality, as well as logic and formal systems.

Objectives

1. To address central questions about being human
2. To evaluate the moral significance of advances in science and technology
3. To critically analyze primary sources of scholarship in philosophy
4. To develop tools for independent critical thinking, through rational and moral reflection, collegial discussion, and clear and precise writing

5. To learn some deep facts about the limits of logic and computation



Course Texts

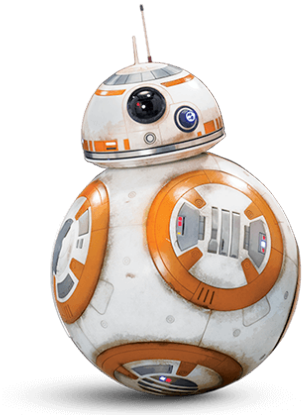
The readings will consist of articles, short stories, short selections from books, and course notes I have written. I will post links to all of the readings on the following webpage:

http://www-bcf.usc.edu/~russ813/ai.readings.html

Requirements

Each assignment is described in more detail below.

Requirement	*Weight*	*Due*
Seminar discussion	10%	every meeting
Check-in assignments	10%	roughly weekly
Analytical essay	15%	February 21
Midterm exam	15%	March 2
Research project		
- **Research question**	10%	April 11
- **Peer feedback**	10%	April 18
- **In-class presentation**	10%	April 25 & 27
- **Final report**	20%	May 5 (TBC)



Seminar Discussion

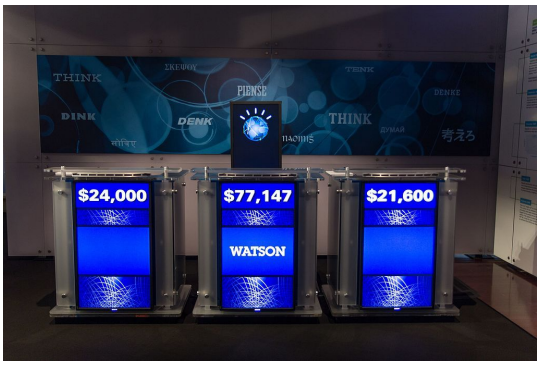
Be present. A seminar is a collective enterprise: how valuable and interesting it is will depend on what you bring to the table. Being present, on time, prepared, and engaged shows respect to the other students in the seminar.

Be prepared. Do the day's reading assignment. Spend time before class thinking about specific questions the reading leaves you with, and specific ideas you would like to discuss.

Reading a philosophy article is a very different skill from reading a novel, a news article, a textbook, or a scientific paper. It will take practice. You should read an article more than once, and you might need to read some parts many times before they make sense. A good strategy is to read an article quickly once to get the big idea: what is the author arguing for? Where in the article do the important moves of the argument happen? Once you have a mental map of the article as a whole, go back and read it a second time for detail, spending more time on the places you identified as making important moves. What premises is the author relying on? How do they support the author's main argument? After this, if there are important parts you still don't understand, go back and work through them again.

Be engaged. Participate in the conversation creatively, constructively, and reflectively. Listen carefully to others.

Keep your phone on silent and in your bag. Even if you're a great multi-tasker, I find phone-checking very distracting, and many students do as well. Please keep your computer put away during discussion, too, unless you have some special need for it (such as a disability that makes writing by hand difficult).



Check-in Assignments

We will have about eight "check-in" assignments at the beginning of class meetings. These will be very short in-class writing assignments about something related to the assigned reading for that day. These will not be announced ahead of time, and they cannot be made up if you are late or absent. You get one free pass: you can drop one check-in without any penalty. (But please do talk to me if you have a serious problem such as a medical emergency that leads to missing more than one check-in.)

Analytical Essay

Your first writing assignment will be an analytical essay evaluating a philosophical argument about the nature of consciousness and intelligence. This assignment has the following goals:

1. You will analyze the structure of a philosophical argument.
2. You will think critically about both sides of a controversial issue, showing understanding of the strengths and weaknesses of a position.
3. You will show creative independent reflection.
4. You will write clearly and precisely.

It will be 4–5 pages long, and it will include the following three parts.

1. Clearly present **one of the arguments** we have discussed in class for the conclusion that human consciousness cannot be understood as part of the physical, material world:
    a. Descartes' "doubt" argument
    b. Jackson's "knowledge" argument
    c. Chalmers' "zombie" argument

    Make sure to be explicit about the **structure** of the argument. What is a precise statement of its conclusion? What premises does it rely on?

2. Present an **objection** to this argument. You can discuss an idea from one of the readings, or you can present a new idea of your own. In either case, you should be extremely clear about why the objection might show either that one of the argument's premises is false (and which one) or else that the premises do not logically support the conclusion.

3. **Evaluate the objection.** Is it successful, or is there a good way of replying to it? Explain.

I encourage you to meet with me in office hours at least a week before the essay is due to discuss your plan.

Midterm Exam

There will be an in-class midterm covering the first two parts of the seminar (Intelligence and Consciousness, and Formal Systems and Their Limits). The goal is to show knowledge of the texts, understanding of the key philosophical questions, and understanding of the technical limits on formal systems. The exam will include short answer and true–false questions, and it will include an in-class essay about Lucas's argument against "mechanism".

Research Project

You will do a final project on the philosophical significance of an area of current research in AI. For example, your project might be about the ethics of drone warfare, the metaphysics of virtual reality, the politics of self-driving cars, the aesthetics of computer music, or the implications of "intelligent assistants" for philosophy of mind. I will present some example questions and places to start looking in class.

This project will have four different deliverables, which will be due at different points in the semester.

Research Question

The goal of this part is to set a clear direction for your project. This should be two pages long, and have the following parts.

1. **Describe an area of AI research** you will be writing about. (For example: self-driving cars, game-playing, face recognition, medical diagnosis.) Give examples.

2. **State a clear philosophical question** that this research raises, and which you will explore. We will discuss examples of how to come up with good questions in class.

3. **State at least two clear, reasonable answers** to this question. Frame your answers in such a way that it's clear they are not both correct. Briefly give reasons that might support each of these two answers.

Peer Feedback

You will write a 1–2 page response to someone else's proposed research question, helping to refine the question and clarify the reasons for taking one side or another. Your goal is to help your peer to produce a strong presentation and final report. Is the question clear? Does it have at least two interesting, defensible answers? How can the

reasons for one side or the other be strengthened? We will spend some time in class for you to discuss this feedback in person.

Final Presentation

You will give a five-minute in-class presentation about your topic. In your presentation, you should describe the field of AI, giving examples of current and possible future work. You should state your philosophical question, and you should explain the position on this question that you will defend.

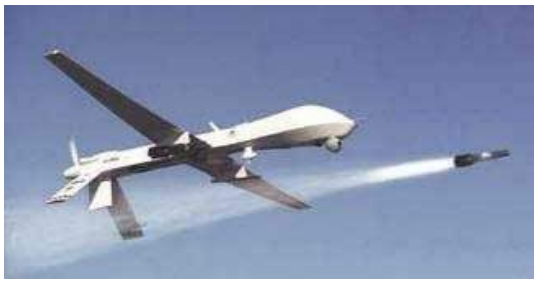
Written Report

Your final written report will be due in class on the day of the final exam. It will be 6–8 pages long, and should do the following.

1. Give clear exposition of an aspect of current or future technology and its philosophical significance.
2. State a clear and precise philosophical question, and explain in depth the reasons for at least two different answers to it.
3. Defend one of these answers, either by (a) presenting an argument for it and defending that argument against objections, or else (b) presenting an argument against it and showing that an objection to that argument is successful.

Equality, Diversity, and Support

This classroom is a safe environment. Any discrimination on the basis of race, gender, sex, sexuality, socioeconomic status, disability, national origin, religion, or age will not be tolerated. If at any time while at USC you feel you have experienced harassment or discrimination, you can file a complaint: see http://equity.usc.edu for more information. You are also welcome to bring the complaint to any faculty or staff member at USC.



Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. SCampus, the

Student Guidebook, contains the Student Conduct Code in Section 11.00. The recommended sanctions are located in Appendix A. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: http://www.usc.edu/student-affairs/SJACS/.

Statements for Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to us as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m. - 5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

Schedule

Date	Reading	Due
Part 1	**Intelligence and Consciousness**	
Jan 10	First meeting	
Jan 12	Turing, “Computing Machinery and Intelligence” Optional: *The Imitation Game*	
Jan 17	Descartes, *Meditations on First Philosophy*, Meditations 1 and 2	
Jan 19	Valente, “Silently and Very Fast”	
Jan 24	Nagel, "What Is It Like to Be a Bat?"	
Jan 26	**No class** Egan, “Closer”	
Jan 31	Jackson, “Epiphenomenal Qualia”	
Feb 2	Chalmers, “Consciousness and its Place in Nature”	
Feb 7	Dennett, “Quining Qualia”	
Feb 9	No new reading	
Part 2	**The Limits of Computation**	
Feb 14	Course notes (Programming basics, the Church-Turing Thesis)	Analytical essay plan (optional)

Feb 16	Course notes (The Universal Program, Turing's Undecidability Theorem) Pullum, "Scooping the Loop-Snooper"	
Feb 21	Course notes (Gödel's First Incompleteness Theorem)	Analytical essay
Feb 23	Lucas, "Minds, Machines, and Gödel"	
Feb 28	Boolos, "Gödel's Second Incompleteness Theorem Explained in Words of One Syllable." Chiang, "Division by Zero"	
Mar 2	**Midterm**	Midterm (in class)
Part 3	**Identity, Embodiment, and Immortality**	
Mar 7	Locke, *An Essay Concerning Human Understanding*, Book II, Chapter XXVII ("On Identity and Diversity")	
Mar 9	Parfit, *Reasons and Persons*, chapter 10 ("What We Believe Ourselves to Be" and)	
Mar 14 & 16	**Spring break**	
Mar 21	Parfit, chapter 11 ("How We Are Not What We Believe")	
Mar 23	Yang, "Patterns of a Murmuration, in Billions of Data Points"	
Mar 28	Shawl, "Deep End"	
Mar 30	Egan, "A Kidnapping"	

	Nozick, *Anarchy, State, Utopia*, pp. 605–607	
	Optional: *Black Mirror* S3 E4, “San Junipero”	
Part 4	**Robot Politics**	
Apr 4	Bostrom and Yudkowsky, “The Ethics of Artificial Intelligence”	
Apr 6	Continued	
Apr 11	Miller, “The Long-Term Jobs Killer Is Not China. It’s Automation.”	Research question
	The Economist, “Artificial Intelligence: The Impact on Jobs”	
	Manjoo, “A Plan in Case Robots Take the Jobs”	
Apr 13	Danaher, “Will Life Be Worth Living in a World Without Work? Technological Unemployment and the Meaning of Life”	
Apr 18	*Blade Runner*	Peer feedback
Apr 20	Minsky, “Alienable Rights” (TBC)	
Apr 25	Final presentations	Final presentations
Apr 27	Final presentations	Final presentations
May 9	Final report due	Final report

Changes

I may make changes to the schedule or assignments during the semester.

Images

1. *Robot*, Alexandra Exter, 1926. Art Institute Chicago.
2. Siri, Apple.
3. GLaDOS, *Portal 2*. http://theportalwiki.com/wiki/GLaDOS
4. BB-8, *Star Wars: The Force Awakens*.
5. Watson, IBM. Picture by Atomic Taco (CC BY-SA 4.0) via Wikimedia Commons
6. Armed Predator drone firing Hellfire missile. January 20, 2010. Picture by Brigadier Lance Mans, Deputy Director, NATO Special Operations Coordination Centre. http://www.ifpafletcherconference.com/2010/powerpoint/Lance_Mans-Day2-final4web.ppt
7. HAL 9000, *2001: A Space Odyssey*.