

SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

Instructor

Peter R. Murray

OH: Thursday, 10am-12pm
Ladd 108, 518-580-5401

Email: pmurray@skidmore.edu

Course Info

Philosophy 330

Mon., Wed. 2:30pm–4:20pm

Ladd 107

Spring 2017

Course Description

Artificial intelligence (AI) is seemingly everywhere now, and each day brings news of further advances in AI technology and further discussion of its promises and perils. Many of these discussions focus on Terminator-style apocalyptic scenarios in which super-intelligent machines rise up and exterminate humanity. However, the fact that computers will soon be—or already are—piloting our cars, trucks, and planes, helping to make our medical, financial, and legal decisions, tending our children and elderly, and even fighting our wars raises difficult philosophical questions that are closer to hand than such existentially gripping science fiction.

The questions that we will investigate in this seminar include: What ethical principles should we program into AI-piloted vehicles to handle “forced choice” situations in which someone or other will die or be injured and the AI has the ability to influence who that is? Is ethical decision-making even amenable to algorithmic implementation? Who should be morally or legally responsible for AI systems and their actions? The corporations or governments who built them? The AI systems themselves? More generally, what is “intelligence”, and what makes an AI “artificially” intelligence? What is the relation between intelligence and the other aspects of psychology that intelligent beings typically have, e.g., sensations, emotions, moods, beliefs, desires, etc.? Could something be intelligent without having any of those other characteristics? And if not, do we need to create artificial sensations, etc. to create artificial intelligence? And if we are talking about sensations, we are presumably talking about bodies, but now we appear to be talking about creating artificial life in order to create artificial intelligence, which raises a further host of ethical and metaphysical questions. For example, could an AI system come to have ethical duties and commensurate rights? Could the decision to turn an AI system on or off become as ethically laden an issue as whether to bring a child into the world or to end a human being’s life? Or will or should an AI always be just a tool?

Course Objectives

The view of philosophy that will be central to our work in this class is that of philosophy as an activity, as something we do, rather than as a set of doctrines that we learn or memorize. As we will see, the heart of this philosophical activity consists in analyzing, evaluating, and critiquing the argumentation that we present to persuade ourselves and others that certain conclusions are correct. The conclusions that we will be particularly interested in here have to do with important ethical and metaphysical issues related to the development of artificial intelligence. However, the argument- and reasoning-related skills that we will practice and develop in this course are of much broader application and utility not only in your other courses at Skidmore, but elsewhere in your life more generally.



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

Accordingly, the learning objectives in this course are for you to be able to:

- (1) Recognize, charitably reconstruct, and engage in productive ways with arguments in the texts we read and with the arguments of your peers;
- (2) Develop, draft, revise, and polish a sustained, reasoned argument in a paper that engages arguments presented in texts;
- (3) Work collaboratively with your peers to clarify your reasoning and writing by providing and incorporating feedback and suggestions on each other's work;
- (4) Describe the major AI-related ethical, practical, and metaphysical issues that we cover in this course, the arguments that the authors we read have presented for and against various positions on those issues, and the dialectical interrelations among those arguments.

Course Materials

AI: Its Nature and Future (by Margaret Boden, 2016)
(abbreviated "AINF" in course schedule)

Artificial Intelligence: A Philosophical Introduction (by Jack Copeland, 2001)
(abbreviated "AIPI" in course schedule)

The Readings page of the course *Blackboard* site has links to other materials we will work with in this course. (abbreviated "BB" in the course schedule)

Course Requirements

There are 100 possible points in this course. Your final grade in this course is determined by your participation in class, your performance on your Argument Reconstructions and Critiques (ARCs), your papers, and your feedback on your peers' paper drafts:

1. Participation (10 points, i.e., 10% of course grade)

To participate successfully in this class, you must:

- Read the assigned texts before class, think about them, and bring them to class;
- Read your peers' ARCs for a given class meeting before that class meets;
- Come to class consistently and on time (I keep track of attendance). *In addition, students who miss more than four classes, for whatever reason, will fail the course;*
- Participate actively in class discussions (I note down who participated in each class, and we will discuss and practice how to do so productively);



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

- Bring questions to office hours, post questions or other issues for discussion to the Discussion Forum section of the course *Blackboard* site, and/or respond to questions and issues that other students have posted to the Discussion Forum. *At a minimum, you must meet with me in office hours at least once during the semester, in addition to the Midterm Status Review that we will have to see how the class is going for you.*
- A course Q-and-A notebook will circulate continuously among the students at each class meeting. In it, student may participate by posing a question about material in that class meeting that they feel they are not fully understanding, or by providing a suggested answer to a question raised by a peer. Entries should be identified only by the last four digits of your Skidmore ID, and answers should reference the ID number of the question(s) they are responding to. The notebook is turned in at the end of each class.

2. Argument Reconstructions and Critiques (ARCs) (30 points, i.e., 30% of course grade)

- Beginning in week 3, you will post a 500-word Argument Reconstruction and Critique (ARC) on the readings for one class per week on the ARCs page of the course *Blackboard* site. ARCs for a given class must be posted to the ARCs page by 7 pm the day before the class meets. See the Assignments page of the course *Blackboard* site for guidelines on writing ARCs and anonymized examples of well done student ARCs.
- For one class meeting during the semester, you will each present your ARC to the class, summarize your peers' ARCs and how they compared and contrasted with your own, and offer two questions to launch discussion for that class. See the Assignments page of the course *Blackboard* site for guidelines for ARC presentations.

3. Papers (40 points, i.e., 40% of course grade)

- You will write two ten-page papers in this course. Each paper will count for 20 points (20% of your course grade), with your first draft of the paper counting for 10 of those points, and the final draft of your paper counting for the other 10 points.
- You will have approximately two weeks to complete each paper. The first week, you will complete a draft of your paper, which you will turn in for feedback and suggestions for improvement from a peer in the class, based on guidelines that I will provide. Your peer reviewer will have two days to return your draft to you with their feedback and suggestions, I will also give you feedback on your draft, and you will have the remainder of the two weeks to make revisions to your paper that incorporate all the feedback you've been given before turning in your final draft. Your Skidmore student ID number should be the only identifying information on both your first and final drafts.
- Late papers are accepted up to one week after the deadline, but they lose 1/3 of a letter grade per day. So, if a paper is of A- quality, but it is turned in an hour after the deadline, it will receive a B+. If the same paper is turned in 25 hours after the deadline, it will receive a B, and so on. Papers more than seven days late will not be accepted.



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

4. Peer Feedback on Paper Drafts (20 points, i.e., 20% of course grade)

- For each of the two papers in this class, you will provide one of your peers with feedback and suggestions for improvement on their draft, using guidelines that I provide. Each peer review will be worth 10 points (10% of the course grade).
- Your feedback will be graded on its helpfulness and level of engagement with the paper you review. See the Assignments page of the course *Blackboard* site for guidelines on providing peer feedback and a rubric for assessing the quality of their draft.
- To ensure that your peer has time to incorporate your feedback into her final draft, peer feedback on paper drafts may not be turned in late.

5. Academic Integrity

- Your work in this course is governed by the Skidmore College Honor Code and Code of Conduct (https://www.skidmore.edu/student_handbook/honor-code.php).
- Violations of the Honor Code and/or Code of Conduct will result in your failing this course. If you have any questions regarding whether particular actions constitute violations of the Honor Code or Code of Conduct, please see me to discuss them.

6. Extra Credit

There are two opportunities for extra credit in this class, each of which is worth up to 6 points (i.e., 2/3 of a letter grade):

- After consulting with me and receiving my approval, students may write a three- to four-page paper or do an in-class presentation that explicitly connects a movie, TV show (episode), novel, article, short story, or personal experience to our readings and discussion in this course;
- With my approval, students may give more than one ARC presentation.

Final course grades are determined according to the following scale:

F = below 60 pts	D+ = 67–69 pts D = 63–66 pts D- = 60–62 pts	C+ = 77–79 pts C = 73–76 pts C- = 70–72 pts	B+ = 87–89 pts B = 93–86 pts B- = 80–82 pts	A+ = 97–100 pts A = 93–96 pts A- = 90–92 pts
-------------------------	--	--	--	---

Academic Resources

Skidmore College has a variety of resources available to support your work in this and other courses. Please seek them out and let me know if you have any questions about them. An online listing of available resources is at <https://www.skidmore.edu/advising/support.php>.



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

Skidmore is also committed to supporting your mental health and wellbeing. If you are experiencing depression or anxiety, suffering from an eating disorder, struggling with some other psychological difficulty or trauma, or if you just need someone with whom to talk, the Counseling Center (518-580-5555) is an excellent place to get the help you need. More information is available online at <https://www.skidmore.edu/counseling/services.php>. If you need immediate assistance at any time of the day or night, call Campus Safety at 518-580-5567, and they can connect you with the counselor on call.

Skidmore College considers sexual and gender-based misconduct to be one of the most serious violations of the values and standards of the College. Unwelcome sexual contact of any form is a violation of students' personal integrity and their right to a safe environment and therefore violates Skidmore's values. Sexual and gender-based misconduct is also prohibited by federal regulations. Skidmore College faculty are committed to supporting our students and upholding gender equity laws as outlined by Title IX. If a student chooses to confide in a member of Skidmore's faculty or staff regarding an issue of sexual or gender-based misconduct, that faculty or staff member is obligated to tell Skidmore's Title IX Deputy Coordinator. The Title IX Deputy Coordinator will assist the student in connecting with all possible resources for support and reporting both on and off campus. Identities and details will be shared only with those who need to know to support the student and to address the situation through the college's processes. If the student wishes to confide in a confidential resource, The Counseling Center Staff, Health Services, and Victim Advocates are all options available. More information can be found at <https://www.skidmore.edu/sgbm/>.

If you are a student with a disability and believe you will need academic accommodation, you must formally request accommodation from Meg Hegener, Coordinator for Student Access Services. You will need to provide documentation that verifies the existence of a disability and supports your request. For further information, please call 518-580-8150 or stop by the office of Student Academic Services in Starbuck Center.

Strategies for Success

Engaging well in the activity of philosophy is an acquired skill that improves with practice, and the assignments in this class have been organized to provide you with a series of scaffolded opportunities to practice that skill. The readings have been pared down as much as possible to allow you to focus directly on the relevant arguments without losing sight of their context. The ARCs give you an opportunity to reconstruct the argument for a particular view and to begin to think about how that argument might be critiqued and how it is related to other arguments we are addressing in this course. This prepares you for our in-class discussions, in which we will clarify and refine our understanding of the course material, and each of you will have a chance to shape our discussions through the questions you come up with to launch them. Your papers will build on the preliminary work you have done in your ARCs and in-class discussions and allow you to extend and bring further clarity to your thoughts in a sustained piece of reasoning on a single topic. Finally, your peer reviews give you the opportunity to use your own emerging expertise and proficiency with argumentation to help someone else improve their work, to



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

benefit in your own work from the expertise and proficiency of someone else in the class, and to make sure that you are writing for the correct audience in your papers: namely, your peers, rather than me. So, trite as it may sound, my best advice for how to succeed in this class is really for you to do the assigned work.

Though each student is required to do his or her own work, I highly encourage you to work on and discuss the readings in pairs or groups in preparation for writing your ARCs and papers. Often, a breakthrough in your own understanding of the argument for a view is as close as another student working on the same thing. Also, please do come to my office hours or email me to talk through any difficulties you are having with the course material. The week of 3/27–3/31 we will have instructor-student midterm status meetings to make sure we’re all on the same page regarding how you’re doing in the course, and I encourage you at any time to fill out the Instructor Feedback Form that is linked from the course *Blackboard* site to let me know how I can help you better.

Course Schedule (subject to revision)

WEEK 1	Introduction: Course overview and philosophical background	1/23: Introduction to the course, syllabus	NO READINGS
		1/25: Ryle on the Official Doctrine	<ul style="list-style-type: none"> • Concepción, “How to Read Philosophy” (BB) • Graff and Berkenstein, “I Take Your Point” (BB) • Ryle, “Descartes’ Myth” (BB) • [recommended] Rosen et al, “A Brief Guide to Logic and Argumentation”
WEEK 2	The philosophical background of AI	1/30: Mind-Brain Identity Theory	<ul style="list-style-type: none"> • Smart, “Sensations and Brain Processes” (BB) • [optional] Kim, “Mind as the Brain”
		2/1: Machine Functionalism and the Computational Model of the Mind	<ul style="list-style-type: none"> • Putnam, “The Nature of Mental States” • Boden, pp. 132-4 (AINF) • [optional] Kim, “Mind as Computer: Machine Functionalism”

SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE

METAPHYSICS OF MIND AND ETHICAL ISSUES

WEEK 3	Is symbol processing thinking? Is thinking symbol processing?	2/6: Symbol Systems Hypothesis (SSH): symbol processing is thinking	<ul style="list-style-type: none"> • Newell & Simon, “Computer Science as Empirical Inquiry” (BB) • Copeland, Ch. 4 (AIPI) • Copeland Chs. 2, 5 (AIPI)
		2/8: Strong Symbol Systems Hypothesis (SSSH): thinking is symbol processing	<ul style="list-style-type: none"> • [review] Newell & Simon, “Computer Science as Empirical Inquiry” • Fodor, “Fodor’s Guide to Mental Representation” (BB) • Copeland, Ch. 9 (AIPI)
WEEK 4	What would show that a symbol-processing computer is thinking?	2/13: Turing Test	<ul style="list-style-type: none"> • Screening: “Ex Machina” • Turing, “Computing Machinery and Intelligence” (BB) • Copeland, Ch. 3 (AINI) • Boden, 119-21 (AINF)
		2/15: Searle’s “Chinese Room” argument	<ul style="list-style-type: none"> • Searle, “Minds, Brains, and Programs” (BB) • [optional] Sloman, “Did Searle Attack Strong Strong AI or Weak Strong AI?” (BB)
WEEK 5	What would show that a symbol-processing computer is thinking? (cont.)	2/20: The Systems Reply to Searle’s “Chinese Room” argument	<ul style="list-style-type: none"> • Copeland, Ch. 6 (AIPI) • Boden, pp. 134-7 (AINF)
	Further issues for symbol-processing AI: Intentionality	2/22: Can computers give a damn?	<ul style="list-style-type: none"> • Haugeland, “Understanding Natural Language” (BB)
WEEK 6	Further issues for symbol-processing AI: Consciousness	2/27: Does understanding require consciousness? Topics for Paper 1 distributed	<ul style="list-style-type: none"> • Block, “Troubles with Functionalism” • Boden, pp. 119-22, 128-32 (AINF)



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE

METAPHYSICS OF MIND AND ETHICAL ISSUES

WEEK 6	Further issues for symbol-processing AI: Emotion	3/1: Is emotion and affect necessary for cognition?	<ul style="list-style-type: none"> • Simon, “Motivational and Affective Controls of Cognition” • Sloman and Croucher, “Why Robots Will Have Emotion” • [optional] Sloman, “Review of Picard’s <i>Affective Computing</i>” • [optional] Picard, “Response to Sloman’s Review”
WEEK 7	Is symbol-processing hopeless or unnecessary as an approach to AI?	3/6: The Frame Problem DUE: Drafts of Paper 1	<ul style="list-style-type: none"> • Dennett, “Cognitive Wheels” • Copeland, p. 91 (AIPI) • Boden, pp. 43-4 (AINF)
		3/8: Anti-Representationalism DUE: Reviews of Paper 1 drafts	<ul style="list-style-type: none"> • Brooks, “Elephants Don’t Play Chess” • [optional] Brooks, “Intelligence without Representations” • Kirsch, “Today Earwig, Tomorrow Man”
WEEK 8	<p>SPRING BREAK: NO CLASS 3/13, 3/15 Reading: Copeland, Ch. 10 (AIPI); Boden, Ch. 4 (AINF)</p>		
WEEK 9	Rediscovery of an old approach: Parallel Distributed Processing and Neural Networks	3/20: Factual Background for PDP and Neural Nets Sign up for Midterm Status Meetings for week of 3/27–3/31	<ul style="list-style-type: none"> • [review] Copeland, Ch. 10 • [review] Boden, Ch. 4 • Lecun et al (2015), “Deep Learning” (BB) • Geitgey, “Machine Learning is Fun!” (parts 1-6) (BB) <p>NO ARCs or ARC PRESENTATIONS</p>
		3/22: Neural Net Successes DUE: Final Drafts of Paper 1	<ul style="list-style-type: none"> • In-class screening: “Watson: The Science behind an Answer” <p align="center">NO READINGS</p>



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

WEEK 10	Machine Ethics	3/27: Why is it important to make “friendly” AIs?	<ul style="list-style-type: none"> • Moor, “The Nature, Importance, and Difficulty of Machine Ethics” (BB) • Allen et al, “Why Machine Ethics” (BB) • Yudokowsky, “Friendly AI” (BB)
		3/29: Is an algorithm for ethics even possible?	<ul style="list-style-type: none"> • Purves et al, “Autonomous Machines, Moral Judgment, and Acting for the Right Reasons” • Burri, “What is the Moral Problem with Killer Robots?” (BB) • [optional] McDowell, “Virtue and Reason” (BB)
WEEK 11	Ethical Issues for Neural Nets	4/3: Machine Bias: Racism	<ul style="list-style-type: none"> • Screening: “Big Data, Inequality, and the Law” • Brennan, “Can Computers Be Racist?” • Sweeney, “Discrimination in Online Ad Delivery”
		4/5: Machine Bias: Sexism	<ul style="list-style-type: none"> • Bolukbasi et al, “Man is to Programmer as Woman is to Homemaker” • Datta et al, “Automated Experiments on Ad Privacy Settings” (BB)
WEEK 12	Artificial Life	4/10: What makes something alive?	<ul style="list-style-type: none"> • Screening: <i>Her</i> • Benner, “Defining Life” (BB) • Weber, “What is Life?” (BB) • Boden, Ch. 5 (AINF) • [optional] Putnam, “Robots: Machines or Artificially Created Life?” (BB)



SYLLABUS FOR PHILOSOPHY 330: ARTIFICIAL INTELLIGENCE
METAPHYSICS OF MIND AND ETHICAL ISSUES

WEEK 12	Artificial Life (cont.)	<p>4/12: Could a piece of code be alive?</p>	<ul style="list-style-type: none"> • Screening: <i>Black Mirror</i>, “White Christmas” • Turkle, “Artificial Life as the New Frontier” (BB) • Pargellis, “Self-Organizing Genetic Codes and the Emergence of Digital Life” (BB) • [optional] The OpenWorm project and virtual <i>C. elegans</i>
WEEK 13	Moral Standing of Artificial Life (cont.)	<p>4/17: Personhood and Moral Standing</p>	<ul style="list-style-type: none"> • Dennett, “Conditions of Personhood” (BB) • Beauchamp, “The Failure of Theories of Personhood” (BB) • [optional] Goodwin, “Experimental Approaches to Moral Standing” (BB)
		<p>4/19: Moral Standing and Legal Rights for Machines</p> <p>Topics for Paper 2 distributed</p>	<ul style="list-style-type: none"> • Screening: “Interview with Shelly Kagan” • Heller, “If Animals Have Rights, Should Robots?” (BB) • Calverly, “Legal Rights for Machines” (BB)
WEEK 14	Robot Rights	<p>4/24: Arguments <i>for</i> extending rights to robots</p>	<ul style="list-style-type: none"> • Screening: <i>TNG</i>, “The Measure of a Man” • Darling, “Extending Legal Protections to Social Robots” (BB) • Gunkel, “A Vindication of the Rights of Machines” (BB)
		<p>4/26: Arguments <i>against</i> extending rights to robots</p> <p>DUE: Drafts of Paper 2</p>	<ul style="list-style-type: none"> • Bryson, “Robots should be Slaves” (BB) • Bryson, “Proposal for the Humanoid Agent-builders League (HAL)” (BB)
WEEK 15	Conclusion	<p>5/1: Course Recap</p> <p>DUE: Reviews of Paper 2 Drafts</p>	NO READINGS

Paper 2 Due: Friday, May 10, 12pm (final draft to Safe Assign, first draft and comments to Prof. Murray’s box (Ladd 108))

