

Spring 2021

HSS 405-002: Artificial Intelligence and Autonomy

Daniel Estrada

Follow this and additional works at: <https://digitalcommons.njit.edu/hum-syllabi>

Recommended Citation

Estrada, Daniel, "HSS 405-002: Artificial Intelligence and Autonomy" (2021). *Humanities Syllabi*. 398.
<https://digitalcommons.njit.edu/hum-syllabi/398>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Humanities Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

HSS 405: Philosophy Seminar

Artificial Intelligence and Autonomy

Syllabus

Instructor: Dr. Daniel Estrada

E-mail: estrada@njit.edu

Discord Server: <https://discord.gg/NxFvdH7>

Classes meet Monday and Wednesday 12:30-1:50pm

Class Meetings: njit.webex.com/meet/estrada

Online Office Hours: By Appointment. Discord is encouraged.

Course Description: Autonomy is a foundational concept in modern western philosophy and is central to many popular moral and political theories. Autonomy is also a term of art in the high tech world, frequently seen in both the regulation and marketing of artificial intelligence and robotics. This course will attempt to weave these two discussions together so that the technical, philosophical, and sociopolitical aspects of autonomy can illuminate each other. The goal of the course is to develop a conception of autonomy that is sensitive to both historical and future uses of the term.

The first unit of the course will look at ancient, modern, and contemporary philosophical perspectives on agency, autonomy, and organizational complexity. The second unit will consider modern perspectives on political autonomy and freedom of the will. The third unit will provide a historical and technical overview of artificial intelligence, addressing issues in philosophy of mind, ethics, and political theory. Students are asked to engage in debates and discussions, and to prepare short writing assignments, a presentation on reading material, and a final exam paper.

Course objectives:

- Develop an appreciation of the historical and philosophical discourse on autonomy, agency, and freedom, and their application to AI and robotics
- Recognize the different positions in the free will debate (determinism, compatibilism, libertarianism) and the reasons supporting each position
- Understand basic concepts in computer science and their relevance for philosophy, ethics, and technology policy
- Engage in peer-led debate and discussion peer on free will, artificial intelligence, machine autonomy (ie, the Tunnel Problem), and robot rights
- Prepare short, peer reviewed, well-researched argumentative essays defending a strong thesis related to the course themes

Books: All readings will be made available on Canvas or in class

Course Overview: Lessons are organized into 3 thematic units, with readings assigned to each week. Video lectures introducing major lesson themes are provided on Canvas to guide students through the readings. During most scheduled class sessions, students will give presentations on the readings, followed by discussion and conversation with the instructor and class. Attendance is required for class debates and peer review sessions. Classes will be recorded and videos made available on Canvas for asynchronous discussion. A Discord server will also be available for asynchronous discussion and other class activities.

Assignments Overview:

Weekly assignments (due **Fridays by midnight**)

One **Weekly Comment** (200-300 word) on Canvas each week

Engage on **Discord** (no word limit, just demonstrate active engagement)

Unit 1 and 2 assignments (see Canvas for due dates)

One **Thought Paper** (700-1000 word) on Canvas each unit

Unit 3 assignments:

Final paper preparation (see due dates on Canvas)

Thesis Proposal

Full Proposal peer review

Full Draft peer review

Exam Paper

Final debate & commentary

Class assignments:

Presentation (solo or in pairs)

Sign up on [Presentation Schedule](#)

Participation in live class debates (including **debate preparation**)

Write **Debate Commentary** (300-500 words) on Canvas after Debates 1 and 2

Required Attendance days for debates and peer review

Extra credit:

Regular attendance

AI & Autonomy memes

Encyclopedia entry

Additional Peer Reviews

Assignments in **bold** are required and graded for credit. Full details on assignment requirements and grading policies are given in the syllabus below.

Assignment Structure: Grades in the course will be based on the following assignments:

- **Weekly Comments (20%):** Students are expected to leave one **200-300 word comment** on Canvas by the end of each lesson week.
- **Thought Papers (10%):** Students are expected to produce a **700-1000 word (2-3 page) essay** at the end of Units 1 and 2.
- **Exam Paper (15%):** Students are expected to write one **3000-5000 word (7-10 page) essay** at the end of the semester, and to prepare for it with drafts available for peer review in class.
- **Presentation (30%):** Students are expected to prepare one **10-15 minute presentation** to be given during class time at some point during the semester.
- **Debate Commentaries (5%):** Students are expected to participate in class debates, and produce **short (300-500 word) commentaries** on the debate.
- **Participation (20%):** Students are expected to actively participate in debates and discussions on the **Discord** server, and to contribute to **formal class debates and peer review assignments**.

Assignment details:

Weekly comments: A 200-300 word (1-2 paragraphs) comment due at the end of each week, except for debate weeks. Introductions and Final Thoughts are considered part of your Weekly Assignment requirements. Due dates can be found in the schedule below and on Canvas.

Comments are informal opportunities to engage the reading material and peers each week, and to track how your thinking develops as the course progresses. Write about interesting things you notice in the readings, or work out ideas and concepts that interest you. These should be thought of as low-pressure assignments, and exist mostly to keep students engaged in short writing every week. Demonstrate that you're engaged with the readings and lesson themes. Students can submit comments in weekly reading threads, or as responses to thought papers and debate commentary; basically, all comments on Canvas will be considered for weekly comment credit. Students can submit more than one comment each week, but cannot earn more than the maximum points for comments that week. 11 total weekly comments are assigned, but at most 10 will be graded (not including Introductions and Final Thoughts).

Thought Papers: A 700-1000 word (2-3 page) essays in response to material from Units 1 and 2. Essays should be posted in Canvas in the appropriate unit discussion forum. Due dates can be found in the schedule below and on Canvas.

Thought papers can cover any topic relating to that unit's material, core themes, and in-class discussions. Essays can be argumentative, attempting to defend or critique some perspective or position. But thought papers don't need to be argumentative! Students can dive deeper into the reading material with textual analysis, or present additional outside research related to the lesson. Thought papers should be addressed to the class with the purpose of provoking discussion, and to work out ideas in preparation for presentations, debates, and the Exam

paper. Thought papers are semi-formal, should be more polished than weekly comments, and are graded relative to my impression on the effort that went into an essay, including deliberate pre-writing, thoughtful reading and interpretation of readings, evidence of independent research, and sincere engagement with peers.

Presentations: Students are asked to prepare short, 10-15 minute presentations to be given on WebEx during the scheduled class session, followed by leading 10-15 minutes of discussion with the class and Instructor. Students must sign up on the [Presentation Schedule](#) made available on Canvas to select the day for their presentation. Presentations should be short and informative, and should engage one or two readings from that week's lessons. The goal of a presentation should be to inform and educate other students on lesson materials, to explore ideas and arguments the presenter finds interesting, and to prompt the class for further discussion and debate. Slides are strongly encouraged. Presentations are graded on effort and thoughtfulness. This is your chance to lead the class while exploring your interests, so make the best of it!

Exam paper: 7-10 pages (3000-5000 words) typed and double spaced in a standard font and submitted on Canvas. Can address any issue addressed in class, but must take the form of an argumentative essay. Argumentative essays must include a clear thesis statement and a coherent, reasonable defense of the thesis. They should develop an argument in support of the thesis that considers the relevant scholarly literature, and should include some consideration of objections and alternative views to their thesis. The exam paper will include a drafting period in Unit 3, where students are required to prepare thesis statements, proposals, and eventually a full draft of the paper for peer review and feedback. These drafting stages are a requirement for the class, and will be graded for credit. I am looking for essays that demonstrate that students can prepare a well-structured, thoughtful essay that critically develops an argument through direct engagement with the relevant scholarly literature, and that engages sincerely with meaningful objections to their arguments. Details of the grading rubric can be found on Canvas.

Debate Commentaries: After Debates 1 and 2, students are asked to reflect on the debate in a short (300-500 word) commentary. Your primary assignment is to discuss the most interesting thing you heard **someone else** say during the debate (or preparation), and the implications it has for your understanding of the discourse. This is a good opportunity to reflect not just on your own views, but on the structure of the discourse, and whether it's possible to make progress on certain disagreements. The point of debates is not to "win" an argument, but to give the many different perspectives in class an opportunity to be expressed and defended openly in class, and to reflect how these perspectives interact within the discourse.

Students who do not attend the debates can expand on the debate theme in their commentaries, and explain what they would have said had they attended. Students who missed attendance should clearly say which days they missed; a Debate Commentary will make up for some points lost as Attendance Credit. Add 200-300 words to your commentary for each absence during the required attendance days of the debate.

Participation: Students are expected to participate in group activities and debates, to attend class regularly, regularly engage the readings, and to put good effort into their written work. Three kinds of engagement will be graded for credit, and are considered requirements for the course:

Weekly Discord activity: The [Discord server](#) will be available 24/7 for class discussion and engagement. It is also the best way to reach me for quick questions, or to schedule a meeting in office hours. Discord is free to use, and can be [downloaded here](#). During class I'll take attendance using the Discord server. Students are also asked to leave questions and comments on Discord during presentations and debates. To stay active on Discord, students are asked to leave one substantive comment of any length (registering attendance credit does NOT count) at least once a week on Discord for small participation credit. This comment **must be different** from your weekly comment on Canvas, although the latter can expand on the ideas in the former. You can think of your Discord activity as weekly informal peer engagement for open discussion, and your Canvas comment as the final written product of that week's work. Weekly participation on Discord is a participation requirement for the course.

Required Attendance: Students are encouraged to attend all class sessions and activities, but some classes are mandatory, and attendance is considered a graded requirement for the course. Attendance is required for all debates (included debate preparation, see the schedule below) and for all live peer review sessions.

Peer review: Canvas will automatically distribute Peer Reviews after certain assignments. Students are asked to grade peer assignments using the same grading rubric I use, with space to leave comments to the author. These comments are *not* anonymous, and should not be confused with your weekly comment assignments. Leave direct content engagement in the discussion forum, and put grading comments in Peer Review. Please don't comment on grades (writing style, grammar, etc) in the discussion forum! The goal of peer review is a) to provide another avenue for feedback and peer engagement over scholarship and writing, b) to give students insight into the grading process, in order to improve their own writing assignments, and c) as a double-check on my own grading practices, and to make sure my grades are fair and in line with class expectations. Peer reviews on Thought Papers and Exam Paper preparation are a requirement for participation credit. Peer reviews on Weekly Comments, Debate Commentaries, and Exam drafts will earn some extra credit.

Extra credit: Students will have four forms of extra credit available:

Regular class attendance: Attendance will be taken every class session on Discord. Students will be given a small amount of extra credit for attendance on days that are not required. All class sessions will be recorded and made available online for students unable to attend.

Encyclopedia Entry: Students will have the option to complete a 3-5 page Encyclopedia entry that gives an introduction and scholarly overview into some important technical concept related to the class themes. Discuss the concept in detail and provide some analysis of the major scholarly concerns and disputes. Students should also prepare a short annotated bibliography of at least 5 scholarly works as relevant background reading on the topic. Entries will eventually be uploaded to [this website](#).

AI & Autonomy memes: Students are encouraged to produce memes on the themes and topics discussed in class. Original memes can earn extra credit at the end of the semester. With permission, some memes will be shared on my Twitter feed. See Canvas for details.

Peer reviews: Some assignments will be available for additional peer review that is not required for class. These peer reviews can earn some extra credit. See Canvas for detail.

Accessibility policy: I want all students to succeed in this class, and I will gladly accommodate the special circumstances and needs of all students to make sure that happens. I understand that life doesn't happen on the semester schedule, and that school work can't always be a top priority. In pandemic conditions we all need to be more flexible with scheduling and difficult work conditions; I understand how medical issues or disability can complicate these challenges. If there is any issue impacting your performance in class, please come talk to me in office hours or send me a message by email or on Canvas! Even if you're behind on assignments, drop me a message letting me know what's up, I'm sure we can figure something out =)

Late policy: Assignments are due at midnight according to the schedule below and on Canvas. The Canvas schedule will be considered the "official" and most updated schedule. I'll allow a short (~30 min) grace period; assignments received at 12:01am will not be marked as late. Assignments will receive a late penalty for one week after the due date. After one week, late work will not be accepted without making arrangements with the instructor. See the rubric on Canvas for more details.

Excused Late Work: If you have a legitimate excuse that you know about in advance (an academic conference, National Guard duty, expected delivery date, etc.), please make arrangements with me in advance. Extensions for anticipated issues must be arranged at least 48 hours before a deadline to avoid a late penalty. Unexpected emergencies (medical emergencies, deaths in the family, etc.) should be brought to the attention of the Dean of Students with the [Student Concern Reporting Form](#). The Dean's office is equipped to verify your situation confidentially and provide the administrative support you need. The Dean's office can also coordinate with all your instructors for any issues that arise. After an emergency (when you are able to return to school work), let me know what's up (a short note will do). I'll recommend

you contact the Dean with the form linked above, and I'll ask you to come up with a plan for completing your missing work, and we can go from there.

Plagiarism Detection: Students are expected to submit their work to plagiarism detection at the end of every unit. Students should collect their Thought Paper and all comments on Canvas into a single word document (.doc or .docx), to submit to the full document to the plagiarism detection assignment on Canvas. The final exam paper is also required to submit to plagiarism detection. Failure to submit to plagiarism detection will result in zero credit for these assignments.

NJIT Plagiarism Policy

“Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at:

<http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf>.

*Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. **Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university.** If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu”*

Copying and pasting from the web is one form of plagiarism. Failing to provide adequate citations is a form of plagiarism. Copying from your own work counts as plagiarism. Changing a few words in an extensively quoted passage is a form of paraphrase and [may constitute plagiarism](#) (click the link for a detailed explanation of paraphrase and plagiarism). Any work you use should be given adequate citation so I can find and review your sources. Just as in mathematics, you need to show your work! If you use any source in your research, (including Wikipedia and other encyclopedias) *even if you don't quote them directly*, provide a citation.

Suspected cases of plagiarism will be given zero credit for the assignment with a warning about the plagiarism policy. Students found plagiarizing will also forfeit all extra credit opportunities for the semester. Repeated or extreme instances of plagiarism will be reported directly to the Dean of Students as a violation of the [Student Code of Academic Integrity](#).

Grades:

- **200 pts: Weekly comments** (20 pts x 10)
- **300 pts: Presentation**
- **100 pts: Thought papers** (50 pts x 2)
- **150 pts: Exam Paper & Preparation**
 - 100 pts: Exam Paper
 - 10 pts Thesis proposal
 - 15 pts Full Proposal
 - 25 pts Full Draft
- **50 pts: Debate Commentaries** (25 pts x 2)
- **200 pts: Participation**
 - 50 pts Weekly Discord (5 pts x 10)
 - 50 pts Peer Reviews (10 pts x 5)
 - 100 pts Required Attendance: (15 pts * 7)

Total points: 1000 pts

Extra credit:

- **30 pts:** Regular attendance
- **50 pts:** Encyclopedia entry
- **20 pts:** AI & Autonomy memes
- **14 pts:** Additional Peer Reviews

Grade scale:

A : 900+
B+: 850+
B : 800+
C+: 750+
C : 650+
D : 500+
F : below 500

Thematic schedule**Unit 1: 3 Theories of Agency**

Week 1: Introduction to Agency and Autonomy

Week 2: Agents are alive

Week 3: Agents are machines

Week 4: Agents are conscious

Week 5: Class debate

Unit 2: Free Will

Week 6: Classical Determinism

Week 7: Libertarianism

Week 8: Compatibilism

Week 9: Class debate

Unit 3: AI and Autonomy

Week 10: Computation and Machine Learning

Week 11: Can machines think? Turing, Lovelace, and Searle

Week 12: Artifacts and Autonomy

Week 13: AI Ethics

Week 14: Robot Rights

Week 15: Class Debate

Assignment Schedule

Classes meet MW 12:30-1:50pm on WebEx: njit.webex.com/meet/estrada

Attendance will be taken on Discord during class: discord.gg/g2h4S42

Assignments in **bold** are due at **midnight** on the day indicated; for scheduling purposes, the “week” ends Friday at midnight. Saturday comments on Discord count towards the following week. All assignments must be submitted on Canvas in the appropriate folder. “Attendance required” means attendance and participation are factored into your grades. Attendance is encouraged other days and will earn extra credit points.

Note: There are 11 Weekly Comment assignments listed below. Only 10 are required for “full credit” on this assignment. Same goes for Discord participation.

Unit 1: Six Theories of Agency

Week 1: Introductions

9/2: Class: Introduction to Agency

9/4: **Introductions due**

Presentation schedule due

Week 2: Agents are alive

9/8: Class presentation (note: Tuesday meeting time)

9/9: Class presentation

9/11: **Weekly Comments 1**

Discord 1

Week 3: Agents are machines

9/14: Class presentation

9/16: Class presentation
9/18: **Weekly Comments 1**
Discord 1

Week 4: Agents are conscious
9/21: Class presentation
9/23: Class presentation
9/25: **Weekly Comments 1**
Discord 1

Week 5: Debate
9/28: **Debate preparation (Attendance required)**
9/30: **Debate (Attendance required)**
10/2: **Thought Paper 1**
Debate Commentary 1
Plagiarism detection (TP1 + DC1 + 3xWC1)

Unit 2: Free Will

Week 6: Classical Determinism
10/5: Class presentation
10/7: Class presentation
10/9: **Weekly Comments 2**
Discord 2
Peer Review 1 (on Canvas)

Week 7: Libertarianism
10/12: Class presentation
10/14: Class presentation
10/16: **Weekly Comments 2**
Discord 2

Week 8: Compatibilism
10/19: Class presentation
10/21: Class presentation
10/23: **Weekly Comments 2**
Discord 2

Week 9: Debate
10/26: **Debate preparation (Attendance required)**
10/28: **Debate (Attendance required)**
10/30: **Thought Paper 2**
Debate Commentary 2
Plagiarism detection (TP2 + DC2 + 3xWC2)

Unit 3: AI and Autonomy

Week 10: Agents are computers

11/2: Class presentation

11/4: Class presentation

11/6: **Weekly Comments 3**

Discord 3

Peer Review 2 (on Canvas)

Week 11: Can machines think?

11/9: Class presentation **LAST DAY TO WITHDRAW**

11/11: Class presentation

Thesis proposal

11/13: **Weekly Comments 3**

Discord 3

Thesis Proposal peer review (on Canvas)

Encyclopedia proposals due (extra credit)

Week 12: Artifacts and Autonomy

11/16: Class presentation

11/18: **Full Proposal peer review (in class, Attendance Required)**

11/20: **Full Proposal peer review (on Canvas)**

Weekly Comments 3

Discord 3

Week 13: AI Ethics

11/23: Class presentation

11/25: Class presentation

Weekly Comments 3

Discord 3

THANKSGIVING BREAK

Week 14: Robot Rights

11/30: Class presentation

12/2: Class presentation

12/4: **Weekly Comments 3**

Discord 3

Week 15: Wrap Up & Debate

12/7: **Full Draft peer review (in class, Attendance Required)**

12/9: **Final debate/discussion (in class, Attendance Required)**

12/11: **Final Thoughts**

Final Draft peer review (on Canvas)
Plagiarism Detection (5xWC3 + Final thoughts)
Encyclopedia Entry (extra credit)
AI & Autonomy memes (extra credit)

12/16: Final Exam paper due (on Canvas)
Final Paper Plagiarism Detection

Reading Schedule

See full readings list on Canvas

Unit 1: Six Theories of Agency

Estrada's Agency Lectures & Slides:

- Part 0: What is an agent? [Lecture](#) & [Slides](#)
- Part 1: Agents are alive. [Lecture](#) & [Slides](#)
- Part 2: Agents are machines: [Lecture](#) & [Slides](#)
- Part 3: Agents are conscious: [Lecture](#) & [Slides](#)

Estrada's intro to ethical theories (from PHIL 334)

- L3: Consequentialism: [Lecture](#) & [Slides](#)
- L5: Deontology: [Lecture](#) & [Slides](#)
- L6: Virtue Ethics: [Lecture](#) & [Slides](#)

Background Readings

Barandiaran et al, [Defining Agency](#)

Shea, [Principles of Autonomy](#)

Taylor, [Personal Autonomy](#)

Schlosser, [Agency](#)

Buss & Westlund, [Personal Autonomy](#)

Stoljar, [Feminist Perspectives on Autonomy](#)

Christman, [Autonomy in Moral and Political Philosophy](#)

Week 2: Agents are alive

Ancient/Classical view:

Aristotle's [De Anima Chapter 2: The soul as bodily organization](#)

Aristotle's [Nichomachean Ethics Ch. 6: The virtues of thought](#)

Perez and Ziemke: [Aristotle, Autonomy, and the Explanation of Behavior](#)

Leunissen and Gotthelf [What's Teleology Got to Do with It?](#)

Lennox [Form, Essence, and Explanation in Aristotle's Biology](#)

Witt and Shapiro: [Feminist history of philosophy](#)

Lange, [Woman is not a rational animal: On Aristotle's reproductive biology](#)

Henry [Optimality and Teleology in Aristotle's Natural Science](#)

Henry [How sexist is Aristotle's developmental biology?](#)

Contemporary view:

Mayr, [What makes biology unique?](#)

Moreno and Mossio, [Biological Autonomy](#)

Rosslenbroich, [On the Origin of Autonomy](#)

Keller: [Organisms, Machines, and Thunderstorms: A history of self-organization, Part Two](#)

Bechtel, [Biological Mechanisms: Organized to Maintain Autonomy](#)

Bechtel, [Thinking Dynamically about Biological Mechanisms](#)

Bechtel, [Why do Biologists use So Many Diagrams?](#)

Smith, [New Theories on the Origin of Life](#)

Week 3: Agents are machines

[Bechtel](#)

[Mental mechanisms, autonomous systems, and moral agency](#)

[Explanation: a mechanist alternative](#)

[Abstraction and the Organization of Mechanisms](#)

Mitchell, [Complexity, A guided tour](#)

Carroll, [The Big Picture](#) (w/ Minute Physics)

[The Laws Underlying The Physics of Everyday Life Are Completely Understood](#)

[The world of everyday experience, in one equation](#)

PBS Space Time: [The First Quantum Field Theory](#)

Demos:

[Conway's Game of Life](#)

Wikipedia: [Cellular Automata](#)

[Demos](#)

[Life in Life](#)

Lenia: [continuous state, continuous space](#)

[The logistic map](#)

[Logistic map widget](#)

[This equation will change how you see the world](#)

Week 4: Agents are conscious

Dennett: [Kinds of Minds](#)

Chalmers (ed), [Philosophy of Mind](#)

Chalmers, [Consciousness and its place in nature](#)

Chalmers, [The meta-problem of consciousness](#)

Descartes' [Meditations on First Philosophy, Meditations 1 & 2](#)

Pitson, [Hume on Primary and secondary qualities](#)

Mercer, [Descartes' debt to Teresa of Ávila](#)

Nagel, [What is it like to be a bat?](#)

Ryle, [Descartes' Myth](#)

Dennett, [The Zombie Hunch: Extinction of an Intuition?](#)

Froese et al, [Embodied social interaction constitute social cognition in pairs of humans](#)

Unit 1 Optional readings:

Agents are Soul

- Aquinas, [Summa Theologica](#)
 - Wiki: [Summa Theologica](#)
 - SEP: [Aquinas](#)
 - SEP: [Divine Illumination](#)
- Kitcher, [Life after Faith](#)
- SEP:
 - [Dualism](#)
 - [Teleological Arguments for God's Existence](#)
 - [Creationism](#)
 - [Religion and Science](#)
 - [Religious Diversity \(Pluralism\)](#)

Agents are Sociopolitical Subjects

- Jenkins, [Social Identity](#)
- SEP [Marx](#)
 - Brooks, [An introduction to Marx's Labor Theory of Value](#)
- SEP [Foucault](#)
 - [Discipline and Punish](#)
- SEP [Arendt](#)
 - [The Human Condition](#)

Unit 2: Free Will

Week 6: Classical Determinism

Fischer et al., [Four views on Free Will](#)

Craver, [Mechanisms in Science](#)

Week 7: Libertarianism

Rose, [Sartre and the problem of universal human nature revisited](#)

Welchman, [Locke on Slavery and Inalienable Rights](#)

Week 8: Compatibilism

Frankfurt, [Freedom of the Will and the Concept of a Person](#)

Vargas, [Compatibilism evolves?](#)

Unit 3: AI and Autonomy

Background:

Estrada Lecture Series: What is AI?

- [Part 1: History of AI Lecture & Slides](#)
- [Part 2: The Day Computers took over the World Lecture & Slides](#)
- [Part 3: Machine Learning and Baby Computers Lecture & Slides](#)
- [Part 4: Algorithmic Injustice Lecture & Slides](#)
- [Part 5: Robot Rights Now Lecture & Slides](#)
- EE L13: [Autonomy in weapons and vehicles Lecture & Slides](#)
- EE L14: [Technological unemployment Lecture & Slides](#)

Week 10: Agents are computers

Aaronson, [Why philosophers should care about computational complexity](#)

Haugeland, [Semantic Engines: Introduction to Mind Design](#)

Haselager, [Robotics, philosophy, and the problems of autonomy](#)

Week 11: Can machines think?

Turing, [Computing Machinery and Intelligence](#)

Turing, [Lecture on the Automatic Computing Engine](#)

Saygin, [Turing's test, 50 years later](#)

Searle, [Minds Brains and Programs](#)

Hayles: [Can computers create meanings? A cyber/bio/semiotic Perspective](#)

Week 12: Artifacts and Autonomy

Baker, [The shrinking difference between artifacts and natural objects](#)

Winner, [Do artifacts have politics?](#)

Haraway, [A Cyborg Manifesto](#)

Liao & Huebner [Oppressive Things](#)

Mills, [Ideal theory as ideology](#)

Week 13: AI Ethics

Asaro [What should we want from a robot ethic?](#)

Boulamwini [How I'm fighting bias in algorithms](#)

Boulamwini [AI, aint I a woman?](#)

HPE [The Ethics of AI](#)

Autonomy law and policy

Lin, [Why Ethics Matters for Autonomous Cars](#)

Propublica, [Machine bias in sentencing](#)

[MIT: Moral Machine](#)

Jacques (2019) [Why The Moral Machine is a Monster](#)

Drones and Weapons

Roff and Moyes (2016) [Meaningful Human Control, Artificial Intelligence, and Autonomous Weapons](#)

[Rise of the Drones](#)

[Brief animated history of drones](#)

The Intercept: [Obama's Drone Wars](#)

[Survey of Automated Weapons Systems](#)

[UN CCW Research and Reports](#)

Asaro: [Autonomous weapons](#)

[Campaign to Stop Killer Robots](#)

[Petition the UN](#)

FLI [Open letter on Autonomous Weapons](#)

Week 14: Robot Rights

Danaher: [Should Robots have Rights? Four perspectives](#)

Salvini [How safe are robots in urban environments? Bullying a service robot](#)

Bryson, [Robots should be slaves](#)

[Of by and for the people: the legal lacuna of synthetic persons](#)

Birhane and van Dijk, [Robot rights? Let's talk about Human Welfare instead](#)

Gunkel: [Can and should robots have rights?](#)

Estrada: [Human supremacy as posthuman risk](#)

[Robot rights: cheap yo!](#)

[Alignment, fair play, and the rights of service robots](#)

[Sophia and her critics](#)

Robots in film and other media:

Jenkins, [Convergence Culture](#)

Big Joel: [What's with AI?](#)

Lindsay Ellis: [The Problem of Lady Robots: Transformers and Feminist Theory](#)

Estrada: [Real Robot Movies](#)

Greg Egan, [Orphanogenesis](#)