STS 351 Minds and Machines

Instructor: Dr. Daniel Estrada
E-mail: estrada@njit.edu
Zoom Meeting Link
Office: Cullimore 419
Office Hours: T 2:30-3:30pm and by appt.
Discord: discord.gg/NxFvdH7

Class Meeting: MW 11:30am - 12:50pm CKB 317

Course Description: This course introduces a variety of key concepts, themes, and historical debates in the philosophy of mind. The course is especially concerned with the proposal that the mind is a computing machine. In Unit 1 we will cover the history of computing machinery and intelligence, including classical logic and modern machine learning techniques. In Unit 2 we will look at specific issues in the philosophy of AI, including perception, autonomy, and understanding.
Prerequisites: STS 201 and STS 210, each with a grade of C or better.

Course objectives:

- Develop an understanding of key issues in historical and contemporary debates in the philosophy of AI.
- Understand various arguments for and against a computational theory of mind, and the historical, scientific, and technological context in which they arise.
- Engage in peer-led debate and discussion on critical issues like agency, consciousness, and machine learning.
- Conduct scholarly research through short peer reviewed essays
- Prepare group presentations and conversations exploring the central themes of the course.

Lesson Plan

Unit 1: Thinking machines

Lesson 1: History of AI and computation
Lesson 2: Logic and Computation
Lesson 3: Turing Machines & GOFAI
Lesson 4: Neural networks and machine learning
Lesson 5: Turing and Lovelace

Unit 2: AI Challenges

Lesson 6: Perception and learning
Lesson 7: Sentience and consciousness
Lesson 8: Agency and autonomy
Lesson 9: Understanding

Unit 3: AI Ethics

Lesson 10: Intro to AI Ethics
Lesson 11: Algorithmic bias

Lesson 12: Autonomous weapons and vehicles
Lesson 13: Creativity and copyright
Lesson 14: Robot rights
Lesson 15: Course wrap up

Assignments and expectations

- **Attendance:** Regular classroom attendance is **required**. Students can miss 3 classes without penalty. (10%)
- **Presentation**: One 10-15 min presentation with slides on readings given in class Schedule your presentation on Canvas. (10%)
- **Reading notes**: 300+ words of “reading notes” due for each lesson reflecting on reading assignments in Lesson 2-14. Must be submitted before class on Wednesday for full credit. (24%)
- **Podcasts:** Students are required to participate in a 1 hour recorded group conversation at the end of Unit 1, 2, and 3. (22.5%)
- **Papers:** Students are required to complete two 5-7 page papers, one at the end of Unit 1 and the other at the end of Unit 3. The second paper will require a drafting round and a meeting with the instructor. (25%)
- **Participation:** Introductions and coordinating with podcast groups (8.5%)

Grade break down

Attendance 25 x 4pts = 100
Presentation 1 x 100pts = 100
Notes 12 x 20pts = 240
Podcasts 3 x 75 pts = 225
U1 paper = 100 pts
    U1 paper = 75
    U1 paper proposal = 5
    U1 paper review = 20
U3 paper = 150 pts
    U3 paper = 75
    U3 paper proposal = 5
    U3 paper draft = 20
    U3 paper meeting = 50
Participation = 85
    Intro = 10
    Podcast coordination 3 x 25 pts = 75

Total = 1000 pts

Grade Scale

Final grades are calculated on the following scale:

A: 900+
B+: 830+
B: 770+
C+: 700+
C: 600+
D: 500+
F: < 500

There is a 5 point tolerance for bumping a grade to the next letter when calculating final grades.

Assignment Schedule

Unit 1: Thinking machines

Lesson 1: History of AI and computation

F 1/19 **Introductions**
Presentation Schedule
L1 Notes (EC)

Lesson 2: Logic and Computation

W 1/24 **L2 Notes**

Lesson 3: Turing Machines & GOFAI

W 1/31 **L3 Notes**

Lesson 4: Neural networks and machine learning

W 2/7 **L4 Notes**

F 2/9 **U1 Paper Proposal**

Lesson 5: Turing and Lovelace

W 2/14 **L5 Notes**

F 2/16 **U1 Podcast**
U1 Plagiarism Check

Unit 2: Challenges for thinking machines

Lesson 6: Perception and learning

W 2/21 **L6 Notes**

F 2/23 **Unit 1 Paper due**

Lesson 7: Sentience and consciousness

W 2/28 **L7 Notes**

Lesson 8: Agency and autonomy

W 3/6 **L8 Notes**

F 3/8 **U1 Paper review**

SPRING BREAK

Lesson 9: Understanding

W 3/20 **L9 Notes**

F 3/22 **U2 Podcast**
U2 Plagiarism check

Unit 3: AI Ethics
Lesson 10: Intro to AI Ethics
W 3/27 **L10 Notes**
Lesson 11: Algorithmic bias
W 4/3 **L11 Notes**
F 4/5 **U3 Paper proposal**
Lesson 12: Autonomous weapons and vehicles
W 4/10 **L12 Notes**
U3 paper meeting
Lesson 13: Creativity and copyright
W 4/17 **L13 Notes**
U3 paper meeting
Lesson 14: Robot rights
W 4/24 **L14 Notes**
U3 paper meeting
Lesson 15: Course wrap up
M **U3 Paper peer review in class**
F 5/3 **U3 Podcast**
U3 Plagiarism Check

W 5/8 **U3 paper due**

Assignment Details

Attendance: Regular class attendance is required, and earns up to 100 points of credit for the semester. Students can miss up to three class sessions before an impact on attendance grade. There are 28 total days of class, so 25 attendance days earn full credit. On time attendance counts for one day. Attendance is considered late if registered more than 5 minutes after class begins and earns 80% credit. Attendance is taken on the class Discord server. Please do not register attendance on Discord until you are actually in your seat in class. Students registering attendance without being in class will lose all attendance credit for the semester. Note that attendance for some classes also earns participation credit! See more info below.

Participation: Participation credit is earned by participating in some classroom activities. This includes the Introduction thread at the start of the semester, see instructions on Canvas. There are also four days of classroom activities where attendance is required and earns participation credit. This includes two class debates and two in class peer review sessions. See the schedule on Canvas. Classroom attendance and participation in these activities will earn participation credit. If you cannot attend class on these days, let me know and I will offer an alternative assignment to make up this participation credit.

Presentations: Students must prepare a 10-15 minute presentation on one of the readings in class. Slides are encouraged but not required. The presentation should offer a close reading of the text, summarizing and explaining (in the student's own words) the main conclusions, concepts, and perspectives discussed in the article. Students are encouraged to develop a critical reading of the text by (for instance) drawing out associations with other class readings and discussions, contributing independent research on the same topic, or offering their own critical analysis and insights on the paper and topic. However, the primary focus of the presentation should be on elucidating the readings for class discussion. Students can work individually or in groups. Presentations must engage the primary readings, but they can also engage with supplemental readings and independent research. Basically, your presentation is your primary opportunity to lead the classroom discussion through the reading list. Take advantage of it!

Reading Notes: Students are expected to complete 300+ words of reading notes each week, posted on Canvas in the appropriate discussion thread. Reading notes document student engagement with the weekly readings. Notes can engage either required or supplemental readings. Notes don't need to be structured as a formal essay. Scattered thoughts and reactions, bullet points, sketches of ideas, etc are fine. However, notes should be primarily in your own words. Notes should not consist mostly of quotes or paraphrasing from the source material. You can include quotes you find important or interesting in your notes, but you should also explicitly explain and react to the quote in your own words. The quote itself doesn't contribute to your notes word count. Notes will be scrutinized for plagiarism, so please be careful to write your notes in your own words! Notes submitted before class Monday earn up to 15/15 points. Notes submitted before class Wednesday earn up to 13/15 points. Notes submitted by Friday earn up to 10/15 points. Notes submitted after one week late earn a maximum of 8/15 points.

Papers are scholarly essays substantively engaging with the debates and ideas found in the readings and lectures for each unit of the course. Papers should demonstrate a clear understanding of the issues presented in the readings and careful critical analysis of the scholarly texts. Papers can be argumentative and defend a particular position or controversial thesis in the debate. Papers can also be clarificatory, seeking to elucidate some complex issue or concept through additional research and reflection. Papers will be developed over several activities at the end of the semester, during which proposals and drafts will be made available for peer review and feedback. Detailed instructions and schedules are on Canvas.

Honors Requirements: Students registered for the honors section of the course are required to complete all regular assignments, with the following additions:

- Regular attendance is required for honors students
- 16 Reading notes required (up from 12)
- Honors students must prepare 2 presentations (up from 1).
- Honors students must schedule a meeting with the instructor to discuss the U1 paper
- Honors students must complete one of the following:
    - Either write a short (3-5 page) U2 paper (Can be an Encyclopedia write up)

- Or write 3 additional replies to U1 papers
- Or U3 paper must be 10-15 pages (up from 5-7)

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 earn a small late penalty for material submitted after the assignment due dates posted on Canvas. I'll allow a short (~30 min) grace period for assignments due at midnight; assignments received at 12:01am will not be marked as late, but assignments received at 2am will. Late assignments are accepted until the end of the Unit.

Excused Absences: If you have a legitimate excuse that you know about in advance (an academic conference, athletic event, 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 and 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 if you haven't already, and we can discuss a plan for completing your missing work, and go from there.

Plagiarism Policy

Plagiarism Slides

Plagiarism means using work that you did not produce, but presenting it as if it is your own work in assignments. If you did not write the words yourself, you must clearly distinguish that work from your own with quotes and citations. When I am scanning for plagiarism I am looking for long blocks of text that are clearly taken from other sources (possibly with minor modifications) without proper attribution and without distinguishing it from the students own work. Passing off the work of another for credit is plagiarism, and it will not be tolerated in an ethics course.

Copying and pasting from the web is a form of plagiarism. Changing a few words in an extensively quoted passage is a form of plagiarism. Using AI text generators like chatGPT is a form of plagiarism. Failing to provide adequate citations is a form of

plagiarism. Copying from your own work (including work from previous semesters) without acknowledgement counts as plagiarism. *In general, you should never copy large blocks of text from any other source and present it in your own essay as if it were your own words. That includes copying text from online text generators or language translators. Check this link for a detailed explanation of legitimate paraphrase and illegitimate plagiarism. Any work you use should be given adequate citation so your readers 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 dictionaries, Wikipedia and other encyclopedias, and translation tools) even if you don't quote it directly, provide a citation.*

To avoid plagiarism, you must clearly distinguish your work from the work of others. Any work taken from others must be identified with "quotation marks" and explicit citation. Changing a few words in a quote does not make it your work. *If you use online text generators (like chatGPT, Grammarly, or other text sources), you must explicitly identify that text as not being your own work. You must also cite the explicit generator used, including the version and dates it was used. If you use AI text generators at all, you must also supply the full prompt history generating that text as an appendix to your assignment. If you wrote the essay in another language and then used a translator, you should provide the original text in the original language with your submission. If you read a script in any presentation, you must include the text of that script to the plagiarism detection software on Canvas. If you translate your essay from another language, you must include the original untranslated text for comparison. Failure to do so will not earn credit.*

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* ***Note:*** *the research project is a honeypot for cheaters, and typically results in multiple instances of plagiarism in each section. I won't hesitate to fail students who cheat in my ethics course. Consider this your first warning.*

I have substantially reorganized my class around group discussions and presentations to discourage the use of AI text generators. None of the writing assignments in class are "busy work". They all ask you to demonstrate direct engagement with the readings and with the ideas and perspectives of your fellow students. Please take this opportunity to engage your peers in discussions on ethics seriously!

See these Plagiarism Slides with detailed information on the NJIT and course policies on plagiarism, including examples of legitimate and illegitimate paraphrase, to help you understand the plagiarism policy.

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"*

Appendix A: Reading Schedule

Unit 1: Thinking machines

Lesson 1: History of Automation

Estrada (2023) History of AI audio lecture and slides
W 1/17

- Mullaney et al (2021) Your Computer on Fire (Intro, Ch 1, 2, 6, 7, 8, 9)
- Benjamin (2019) Race after technology Intro, Ch 2, Ch 3
- BobbyBroccoli (2022, YouTube) The image you can't submit to journals anymore

Lesson 2: Logic and Computation

Estrada slides: Arguments, logic, and computation
M 1/22

- Schotch (2006) Introduction to logic and its philosophy
- Shapiro and Kissel (2022 SEP) Classical logic
- Parsons (2017, SEP) The traditional square of opposition
- Crash Course (2016, YouTube) Philosophical Reasoning

W 1/24

- Aaronson (2011) Why philosophers should care about computational complexity
- Chalmers (1994) On implementing a computation
- Piccinini (2021, SEP) Computation in physical systems
- Horst (1999) Symbols and computation
- Haugeland (1981) Semantic Engines: Introduction to Mind Design

Lesson 3: Turing Machines & GOFAI

M 1/29

- Haugeland (1981) Semantic Engines: Introduction to Mind Design

    - Turing (1947) Lecture on the Automatic Computing Engine
    - Turing (1950) Computing Machinery and Intelligence
    - Saygin (2000) Turing's test, 50 years later
    - Searle (1980) Minds Brains and Programs

W 1/31

- van Rooij et al (2023) Reclaiming AI as a theoretical tool for cognitive science
    - Haugeland (1994) On the nature and plausibility of cognitivism
    - Piccinini (2010) Mind as neural software?
    - Chalmers (2012) Computational foundations of cognitive science
    - Gallistel (2017) The neurobiological bases for the computational theory of mind
    - Rescorla (2020, SEP) Computational theory of mind

Lesson 4: Neural networks and machine learning

M 2/5

- Bender et al (2021) On the dangers of stochastic parrots
    - Hinton (2007) The next generation of neural networks
    - LeCun, Bengio, & Hinton (2015) Deep Learning
    - Vaswani et al (2017) Attention is all you need
    - Marcus (2018) Deep Learning: A critical appraisal
    - Kriegeskorte & Golan (2019) Neural Network Models and Deep Learning

W 2/7

- Art of the Problem (2023): How neural networks learned to talk,
    - How NNs learn
    - How NNs learn concepts
- Tensorflow Playground (demo)
- Tensorflow Embedding Projector (demo)
- 3blue1brown: Neural Networks. video series (S3 E1-4)
- Computerphile: Neural Networks video series
    - How AI image generators work
    - Stable Diffusion in code
    - How GPT3 works
    - AI Language models and transformers

Lesson 5: Turing and Lovelace

M 2/12

- Turing (1950) Computing Machinery and Intelligence
    - Saygin (2000) Turing's test, 50 years later
    - Searle (1980) Minds Brains and Programs

W 2/14

- Bringsjord (2000) Creativity, the Turing Test, and the (Better) Lovelace Test
    - Abramson (2008) Turing's responses to two objections
    - Whittaker (2023) Plantations, Computers, and Industrial Control

L5 Notes

Unit 1 Podcast

Unit 1 Plagiarism Check

Unit 2: Challenges for thinking machines

Lesson 6: Perception and learning

M 2/19

W 2/21 L6 Notes

Unit 1 Paper due

Lesson 7: Organization and autonomy

M 2/26

W 2/28 L7 Notes

Lesson 8: Autonomy and agency

M 3/4

W 3/6 L8 Notes

Unit 1 Paper review

SPRING BREAK

Lesson 9: LLMs and language

M 3/18

W 3/20 L9 Notes

Unit 2 Podcast

Unit 2 paper pre-write

Unit 3: AI Ethics

Lesson 10: Intro to AI Ethics

M 3/25

W 3/27 L10 Notes

Unit 2 Paper due

Lesson 11: Algorithmic bias

M 4/1

W 4/3 L11 Notes

Lesson 12: Autonomous weapons and vehicles

M 4/8

W 4/10 L12 Notes

Unit 3 paper meeting

Lesson 13: Creativity and copyright

M 4/15

W 4/17 L13 Notes

Unit 3 paper meeting

Lesson 14: Robot rights

M 4/22

W 4/24 L14 Notes

Unit 3 paper meeting

Lesson 15: Course wrap up

M 4/29

Unit 3 Podcast

Unit 3 Plagiarism Check

5/8 Final Paper due