GenEd 1125: Artificial and Natural Intelligence - Syllabus

What does it mean for a machine to be intelligent, how does current artificial intelligence compare with animal intelligence, and should we be worried?

SCIENCE & TECHNOLOGY IN SOCIETY

Instructor: Professor Venkatesh Murthy
Raymond Leo Erikson Life Sciences Professor of Molecular and Cellular Biology
Director of the Center for Brain Science

What is intelligence? An inquiry into the nature of intelligence can take different forms – philosophical, biological, mathematical or technological. In this course, we will use machine intelligence (everything from voice recognizing smartphones to game-playing computers) as a handle to think about natural intelligence (brains and behavior of animals). Although we will start with big, general questions, we will quickly move to concrete queries about brains and computers. This approach, rather than just starting with brains of animals, may be useful in framing more universal questions independent of the specific architecture of brains of animals. As machines increasingly perform tasks that were once thought to be solely in the domain of humans, there is an urgent need for discussions of the moral and societal implications of artificial intelligence. This course targets students interested in brains and computers in equal measure, and who are comfortable discussing ethical concerns.

Teaching Fellows (TFs)

- Lucy Lai (Head TF) | lucylai@g.harvard.edu
- SJ Lee | sukjoon_lee@hms.harvard.edu
- Rockwell Anhoya | ranyoha@g.harvard.edu
- Qianyi Li | qianyi_li@g.harvard.edu
- Kumaresh Krishnan | kumaresh_krishnan@g.harvard.edu

Course Prerequisites

There are no formal prerequisites to this course, and students do not need to be from a technical or science background. Some basic high school math may be required for some problem sets. Problem sets will be a mix of short-form discussion questions and simple technical problems.

Assessment

10% - Lecture attendance / participation*
10% - Section attendance / participation
40% - Problem sets (4 total), 10% each
20% - Mid-term take-home exam
20% - Final paper

*Lectures will be interactive, with discussions throughout. Lectures will also be live-recorded and released after each class.

Summary of Important Dates

All problem sets due on Wednesday at 11:59pm

- 02/02 - PSet 1 handed out (Lectures 1-4)
- **02/09 - PSet 1 due @ 11:59pm**
- 02/16 - PSet 2 handed out (Lectures 5-8)
- 02/22 - Deadline to add / drop a course (fee) and to change from letter-graded to Pass/Fail
- **02/23 - PSet 2 due @ 11:59pm**
- 03/02 - Midterm (take-home) handed out
- **03/09 - Midterm due @ 11:59pm**
- 03/12 to 03/20 - Spring Break!
- 03/30 - PSet 3 handed out (Lectures 14-17)
- **04/06 - PSet 3 due @ 11:59pm**
- 04/13 - PSet 4 handed out (Lectures 18-21)
- **04/20 - PSet 4 due @ 11:59pm**
- 04/27 - Paper proposal due (in Section)
- **05/09 - Final paper due**

Late policy: You are allowed one (1) late pass for problem sets. After that, 1 point will be deducted off your problem set grade for each subsequent day that it is late.

Collaboration and Academic Integrity

Discussion and the exchange of ideas are essential to doing academic work. For assignments in this course, you are encouraged to consult with your classmates as you work on problem sets and exercises. However, after discussions with peers and/or TFs, make sure that you can work through the problem yourself and ensure that any answers you submit for evaluation are the result of your own efforts. If any books, articles, websites, lectures, etc that have significantly helped you with your work, please use appropriate citation practices. Be sure to familiarize yourself with the Harvard honor code, and follow it.

Academic Accommodations

Any student needing academic adjustments or accommodations is requested to present their letter from the Accessible Education Office (AEO) and speak with the professor by the end of the second week of the term. Failure to do so may result in the course head's inability to respond in a timely manner. All discussions will remain confidential, although AEO may be consulted to discuss appropriate implementation.

Course Schedule

Week 1: Intro

Monday, 01/24 — Lecture 1: What is Intelligence?

- Course introduction and overview
- What do we consider intelligent behavior? Can intelligence be reduced to a computation? Does it depend on the hardware performing the computation? What are the differences and commonalities between artificial learning systems and biological brains? We will use the example of visual processing to introduce Marr's levels of understanding: Computation, Algorithm, and Implementation.

Wednesday, 01/26 — Lecture 2: With Great Power Comes Great Responsibility

- Over the past decade, our understanding and ability to measure brain signals has progressed tremendously. Similarly, artificial learning systems are now ubiquitous and can perform tasks previously thought to be reserved to humans. How do we use this knowledge? We will begin to discuss the moral, societal, and ethical implications of these advances and introduce the topics we will cover later in the course.

Section

- TF and class introduction
    - Icebreaker games
    - Purpose of section
    - Grading schemes

Week 2: The Building Blocks of Biological and Artificial Intelligence

Monday, 01/31 — Lecture 3: The building blocks of the brain and neural networks

- We will cover the basics of neurophysiology (single neuron, action potential and synapses). We will begin to discuss how neurons are connected to each other to perform certain functions – simple examples of pattern generators. We will discuss the similarities and differences between neural and AI architecture.

Wednesday, 02/02 — Lecture 4: Guest Lecturer

- What are brains made of and how can we look into them?
- Problem Set 1 handed out (Lectures 1-4)

Section

Week 3: The Building Blocks of Biological and Artificial Intelligence (cont'd)

Monday, 02/07 — Lecture 5: The architecture of the brain

- We will discuss how neurons form brain circuits and brain areas with specific functional roles. We will describe how we can measure the computations performed by the brain. We will discuss new

technologies and how they in turn depend on advances in machine learning. We discuss how the concept of modularity is useful in AI.

Wednesday, 02/09 — Lecture 6: The building blocks of neural networks

- We will introduce the perceptron model and the concept of linear separability. We will discuss the limitations of the perceptron and the problem of credit assignment in large networks. [Demonstration: Perceptrons and simple classifiers]
- **Problem Set 1 due @ 11:59pm**

Section

Week 4: The Building Blocks of Biological and Artificial Intelligence (cont'd)

Monday, 02/14 — Lecture 7: How networks learn

- We will introduce the backpropagation algorithm which is one the essential methods that allow neural networks to learn. We will also discuss the importance of datasets in learning large models that perform complex tasks. [Demonstration: digit recognition, MNIST]. **Compare to how the brain learns complex tasks**

Wednesday, 02/16 — Lecture 8: Guest Lecturer

- Deep Net architectures
- Problem Set 2 handed out (Lectures 5-8)

Section

Week 5: Learning to Represent

Monday, 02/21— President's Day

- No class!
- Tuesday - Deadline to add / drop a course (fee) and to change from letter-graded to Pass/Fail or vice versa

Wednesday, 02/23 — Lecture 9: Primary visual cortex and convolutional neural networks

- Starting with the example of visual recognition, we will introduce the neurobiology of early visual processing (retina/V1). We will then introduce convolutional neural network (CNN) architectures and examples of their success.
- **Problem Set 2 due @ 11:59pm**

Section

Week 6: Learning Complex Representations

Monday, 02/28— Lecture 10: Dimensionality reduction and invariant representations

- We will introduce the notions of invariant representations and dimensionality reduction. Using the example of face representations, we will look at neural representations in inferior temporal cortex and in artificial neural networks. [Demonstration: image recognition in CNNs]

Wednesday, 03/02 — Lecture 11: Guest Lecturer

- Midterm (take-home) handed out

Section

Week 7: Language in Brains and Machines

Monday, 03/07— Lecture 12: Natural Language and Communication

- What forms of communication do animals use with each other? We will look at examples across the animal kingdom: ants, song birds, mouse vocalization, human speech. Is human language qualitatively different? What are the human brain areas dedicated to language? Are they unique to humans?

Wednesday, 03/09 — Lecture 13: Understanding Language

- What does understanding language mean? Is it simply manipulating symbols or is there something else? Recent language models (google translate, GPT-3) work pretty well in real world applications but they make unnatural mistakes. Is "something" still missing?
- **Midterm due @ 11:59pm**

Section

Spring Recess

03/12 to 03/20 — Spring Recess

Week 8: Learning to Act

Monday, 03/21— Lecture 14: Reinforcement Learning: Learning Actions and Values (TF: Lucy Lai)

- How can we learn the value of a new environment? How can we learn which action to take? Using trial and error we can build representations of the world and learn how to optimize rewards. We will introduce and discuss fundamental concepts of reinforcement learning and present some examples. [Demonstration: RL in computer games]

Wednesday, 03/23 — Lecture 15: RL in the brain: the computational role of Dopamine

- We will give a brief overview of the dopaminergic system and see how the activity of dopaminergic neurons can be interpreted within the reinforcement learning framework. This is a great example of

how a detailed quantitative model can make experimentally testable predictions. We will also discuss medical applications of these ideas including deep brain stimulation in Parkinson's disease.

Section

Week 9: Applications

Monday, 03/28— Lecture 16: Learning to play complex games

- Tic-tac-toe, Chess, Go, video games: humans have long designed games requiring planning. Artificial systems can now beat humans in all of these. We will ask what is behind these algorithms. Can we solve them by looking at all possible moves? How much is built in and how much is learned?

Wednesday, 03/30 — Lecture 17: Learning how to drive a car / Robotics

- Automated driving has long been a goal of artificial intelligence. We will give examples of early systems and ask what the components needed to perform in a "natural" environment" are.
- Problem Set 3 handed out (Lectures 14-17)

Section

Week 10: What does the Black Box do?

Monday, 04/04— Lecture 18: Mind-Reading

- Recent advances in non-invasive neurotechnology allow us to decode brain activity with increasing precision. This offers exciting clinical opportunities through neural prostheses to recover lost abilities such as speech or walking. But commercial applications also raise questions about the ownership and privacy of thoughts.
- Last day upon which students may withdraw from a year-long course.

Wednesday, 04/06 — Lecture 19: Interpreting Machines

- Many algorithms learn directly from data using defined learning rules. They can solve complex tasks, but does that mean we understand how they solve it? This is important to understand how they will behave when presented with new data and whether those outcomes can be trusted.
- What do we understand and what do we not understand about current algorithms? How do we assess performance and who is ultimately making the decision?
- **Problem Set 3 due @ 11:59pm**

Section

Week 11: The Limits of Intelligence

Monday, 04/11— Lecture 20: Human Lapses & Biases

- Human (animal) perception and decision making is rife with illusions and biases – some examples are size illusions, implicit biases etc. Why and how do these arise? How do priors (assumptions) influence these? How do we reconcile these lapses and biases with "intelligent" behavior?
- We will start with simple examples of search trees to highlight the interaction between learning from experience and prior knowledge. We will introduce Bayes' rule which allows us to combine probabilistic quantities. We will explore the limits of exhaustive search in a complex tree and why we need algorithms in practice.

Wednesday, 04/13 — Lecture 21: Algorithmic bias (or bias in neural networks)

- Any algorithm is an assumption of an implicit probabilistic model. Here we will explore how these assumptions change the type of representations we learn. We will discuss different forms of bias in neural networks.
- We will start with biases due to non-representative training datasets and then discuss more subtle (but equally impactful) forms. Can fairness be defined mathematically or will there always be a subjective component in complex scenarios? We will give examples of algorithmic decision approaches and biases they can generate (financial decisions, judicial system, etc).
- Problem Set 4 handed out (Lectures 18-21)

Section

Week 12: Reality?

Monday, 04/18— Lecture 22: Is this the real life? Is this just fantasy?

- Neural networks make assumptions and carefully designed stimuli can exploit these. We will discuss how visual illusions can give us insight into visual processing in biological systems and how adversarial attacks can highlight flaws in artificial systems.
- We will also discuss how approximations can be exploited (eg. color monitors, audio compression in MP3, deepfakes, etc). Demonstration: GANs, deepfakes

Wednesday, 04/20 — Lecture 23: What is Consciousness, and Can Machines Have It? (TF: Lucy Lai)

- How can we define consciousness in brains and machines? Can AI ever become conscious?
- **Problem Set 4 due @ 11:59pm**

Section

Week 13: Towards General AI & What is Missing?

Monday, 04/25— Lecture 24: Building Machines that Think Like Humans

- How do we build machines that think like humans? There are many specific tasks and problems in which current AI does extremely well. What are some things they are not good at? We will discuss how future machines should be designed to think and learn more like humans.

- What about energy budgets? While the brain consumes only 20W of power, the latest AI models run on thousands of GPUs. Why is there such a difference? What can be done about it?

Wednesday, 04/27 — Lecture 25: Course Summary

Feedback session

Section

- Final paper topic idea due to section leader this week
- Present 1 slide of your paper topic / rough outline

Final due 05/09 @ 11:59pm