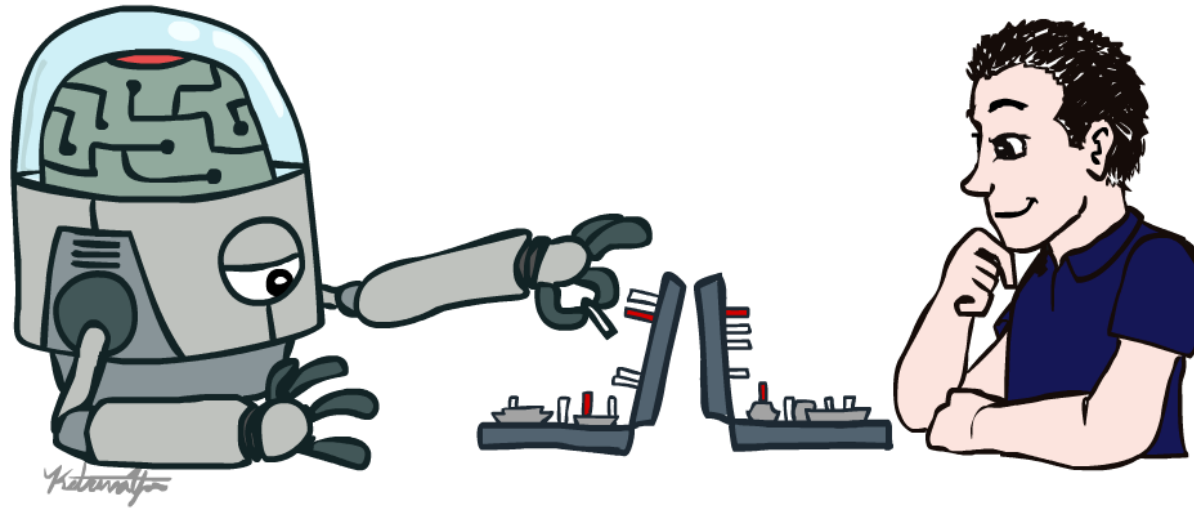


CS 188: Artificial Intelligence

Is AI Actually Intelligent?



Instructor: Oliver Grillmeyer

University of California, Berkeley

Announcements

- Final Exam is **Wednesday, August 13**, 7-10 PM PT in 2050 VLSB
- Please complete course evaluations
- No lecture Wednesday or Thursday. This is the last lecture.

A View of Intelligence

- Are LLMs intelligent?
- Testing Artificial Intelligence
- Thought Experiments
- Natural Intelligence
- Philosophy of Mind
- Machine Intelligence
- Intentions
- Freewill
- Consciousness

LLM Intelligence

- We looked at Transformers
 - Are they really intelligent?
 - Do they understand their answers?
 - Are they more like parrots than humans?

Testing AI: Turing Test

- Turing Test
 - From computer pioneer and mathematician Alan Turing
 - Can you tell if you are communicating with a person or a machine?
 - Testing environment:
 - Two terminals with keyboards that you can type into and read responses
 - You can ask any questions you want
 - If you can't tell which is the program, it passes the Turing Test

Eliza

■ Joseph Weizenbaum's Eliza: emulates Rogerian psychotherapy

Men are all alike

In what way?

They're always bugging us about something or other

Can you think of a specific example?

Well my boyfriend made me come here

Your boyfriend made you come here?

He says I'm depressed most of the time

I'm sorry to hear you are depressed

It's true, I am unhappy

Do you think coming here will help you not to be unhappy?

I need some help, that much seems certain

What would it mean to you if you got some help?

Perhaps I could learn to get along with my mother

Tell me more about your family

My mother takes care of me

Who else in your family takes care of you?

My father

Your father?

You are like my father in some ways

What resemblance do you see?

You are not very aggressive, but I think you don't want me to notice that

What makes you think I am not very aggressive?

You don't argue with me

Why do you think I don't argue with you?

You are afraid of me

Does it please you to believe I am afraid of you?

My father is afraid of everybody

What else come to mind when you think of your father?

Bullies

Does that have anything to do with the fact that your boyfriend made you come here?

Passing the Turing Test

- Have programs passed the Turing Test
 - Eliza?
 - Parry?
 - Chat GPT?

Neuron Replacement Thought Experiment

- Replace neurons in your brain with an artificial equivalent
 - replace wetware with hardware
- Will you still be intelligent and conscious after replacing

One neuron

Ten neurons

1,000 neurons

1,000,000 neurons

86 billion neurons

Functional Brain Thought Experiment

- Would a functional equivalent of a brain be intelligent to human levels?
 - Build a neural network that is the functional equivalent of a human brain
 - Replace wetware with software
 - What must be modeled?
 - What level of granularity do we need to go down to?
 - neurons
 - neuro transmitters
 - chemistry
 - atoms
- Is this just a simulation of intelligence or actual intelligence?

Chinese Room Thought Experiment

- John Searle's Chinese Room Argument
 - A person who doesn't know Chinese is in a room and receives questions written in Chinese through a slot in the room
 - He looks up the question in a huge rule book, and writes out a response in Chinese and submits responses back through the slot
 - From the outside it appears that the person understands Chinese
 - But the person doesn't understand Chinese
- Systems Argument - it's the person and rule book (the entire room) that understands
- One problem: Ungrounded symbols

Modern Day Chinese Room

- Chinese Room Using Transformers Instead of Rule Book
 - Replace the rule book with a Transformer algorithm that the person can follow.
 - Chinese questions come in through the slot
 - The person applies the Transformer algorithm and submits written Chinese responses back out through the slot
 - Does the person understand Chinese now?
- What is understanding?
- Can we ground symbols in meaning (semantics)
 - Is it about relationships between symbols or their emotional content?

Natural Intelligence

- What is unique about human or animal intelligence?
 - Is it an emergent property of the size of our brains
 - Is it the product of thousands of years of evolution?
 - Is it related to the mechanism
 - Neuro-transmitters
 - Chemistry

Mind-body Problem

- Mind-body problem
 - Do brains alone produce intelligence?
 - Or is it the mind?
 - If so, then what is the mind (homunculus problem)
 - Nerves exist throughout the body, not just in the brain
 - Do they play a role in intelligence?
 - Do neurotransmitters play a role?
 - Do hormones affect intelligence?

Philosophy of Mind

- A philosophical views of mind, intelligence, understanding, consciousness
- Intentionality: aboutness, we are consciousness about something
- Epistemology: study and theory of knowledge
- Ontology: the study of being
- Phenomenology: study of the conscious experience
 - Husserl: pioneered the term looking at consciousness subjectively
 - Heidegger: focused on existence (Dasein) “being” in the world

Systems Theory

- Chaos and Systems Theory
 - Highly interactive nonlinear systems exhibit complex behavior
 - Emergent properties can arise from such environments
 - Sensitivity to initial conditions
- Even relatively simple nonlinear equations can exhibit chaotic behavior
 - Logistic Difference Equation to model population growth over time
 - $\text{new_population} = \text{growth_rate} \times \text{population} \times (1 - \text{population})$
 - $0 \leq \text{population} \leq 1$
 - Try with $\text{growth_rate} = 4.0$ and $\text{population} = 0.4$
 - Population: 0.4, 0.96, 0.1536, 0.5200, 0.9983, 0.0064

Homunculus

- How do our brains work?
 - How do we see?
 - Light enters through the pupil, goes into retina, then into the visual cortex, adding features until we see an image
 - How do we understand language?
 - Auditory information goes into the ear, vibrating hairs in the cochlea, and eventually to the auditory cortex where we hear the words and understand the meaning
- Homunculus: the little person in your head that sees, hears, understands
- Begs the question of how the homunculus does those operations
- Assigning intelligence to another agent

Machine Intelligence

- Do machines have their own form of intelligence?
 - How do we measure that or even think about it?
- Can machines take actions that they were not programmed to do?
- How will computer intelligence evolve?
- Would an AI program have or develop self-preservation?
- Would AI programs develop a means of communicating with one another that we can't understand?
- Will humans and machines grow apart or learn to coexist?
- Will we become the inferior form of intelligence?
- What drives machine intelligence to take actions?

Intentions

- Can a computer program have intentions?
 - How would we program intentions?
 - Does it require a will to live?
 - Is it an evolutionary adaptation?
- Does Social Media have intentions?
 - Keeping engagement - infinite scroll
 - Confirmation Bias - present users with documents that support their beliefs and interest
 - Is this about intentions or prescribed actions?
- Do we have intentions?

Freewill

- Can a computer program have freewill?
- Do we have freewill?
 - Is it an illusion?
 - Are our choices predetermined?
 - If we can model our brains, can't we predetermine actions based on inputs?
 - Was Neo destined to take the red pill

Consciousness

- Can a computer program be conscious?
- Do we have consciousness?
 - Is it an illusion?
 - Rene Descartes: Cogito ergo sum - I think therefore I am
 - Are we brains in a vat?
- What about animal consciousness?

Apes

Dolphins

Dogs

Cats

Alternative Consciousness

- What is it like to be a bat?
 - Spend your days in a dark cave hanging upside down
 - Come out at night in search of food
 - “Seeing” via echo-location
 - What do images that are formed via sound echoes look like?
 - Can bats see in color?
 - Do they form their own versions of colors, textures?

Alternative Consciousness

- What is it like to be the person next to you?
 - They may be a different gender, or nonbinary
 - They may be from a different country with a different culture
 - Their native language may be different affecting their world view
 - Do they see colors the way you do?
 - Do they see red like you see green?
 - Do they hear sounds the way you do?
 - Do they think 80's music is good?