# TDT4136 Introduction to Artificial Intelligence

## Lecture 1: Introduction
## (Chapter 1 in the textbook)

Pinar Öztürk

Norwegian University of Science and Technology

# Practical information

All practical Information is in the first part (separate) of Lecture 1 slides.

# Course topic overview

Intelligent agents
Problem solving by Searching
Adversarial Search
Constraint Satisfaction Problems
Logical systems
Knowledge representation
Planning
Game Theory
Ethical issues in AI

# Todays outline

What is AI?
A brief history
The state of the art

# What is AI?

There are no crisp definitions. Here is one from John McCarthy, (Father of the phrase *Artificial Intelligence*)
see `http://www.formal.Stanford.EDU/jmc/whatisai/`

Question: What is artificial intelligence?
Answer: It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

Question: Yes, but what is intelligence?
Answer: Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.

# What is AI? - cont

"It is the science and engineering of making intelligent machines, especially intelligent computer programs" John McCarthy

"AI is the science of making machines to do things that would require intelligence if done by men." Marvin Minsky

"The study of mental faculties through the use of computational models" Eugen Charniak

In short: There is no formal definition covering all aspects of intelligence

# Two aspects of AI

**Science**

The science of understanding intelligent entitites, developing theories which attempt to explain and predict the nature of such entities

Discover ideas about knowledge that help explain various sorts of intelligence

Model functions of the human brain

**Engineering**

Solving real-world problems by employing ideas of how to represent and use knowledge

Engineering of intelligent entities

Produce intelligent behaviour by any means

# Can machines be Intelligent?

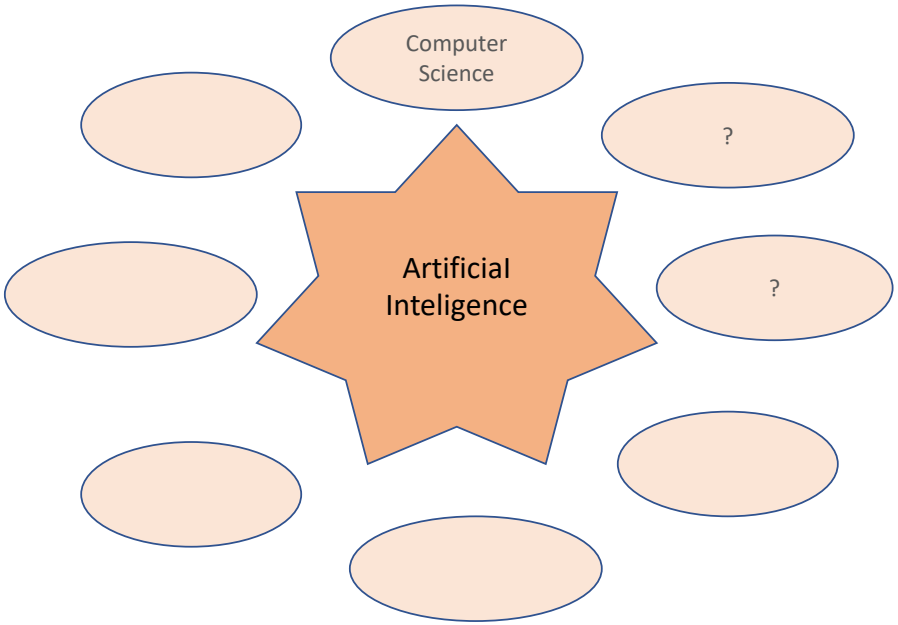
- *Symbolic system hypothesis* (Newell and Simon)
  - Intelligence is substrate neutral
  - A *physical symbol system* has necessary and sufficient means for general intelligent action.

- Biological substrate only (John Searle, philosopher)
  - Intelligence is substrate dependent.
  - The material humans are made of is fundamental for our intelligence.
  - Thinking is possible only in special machines - living ones made of proteins.

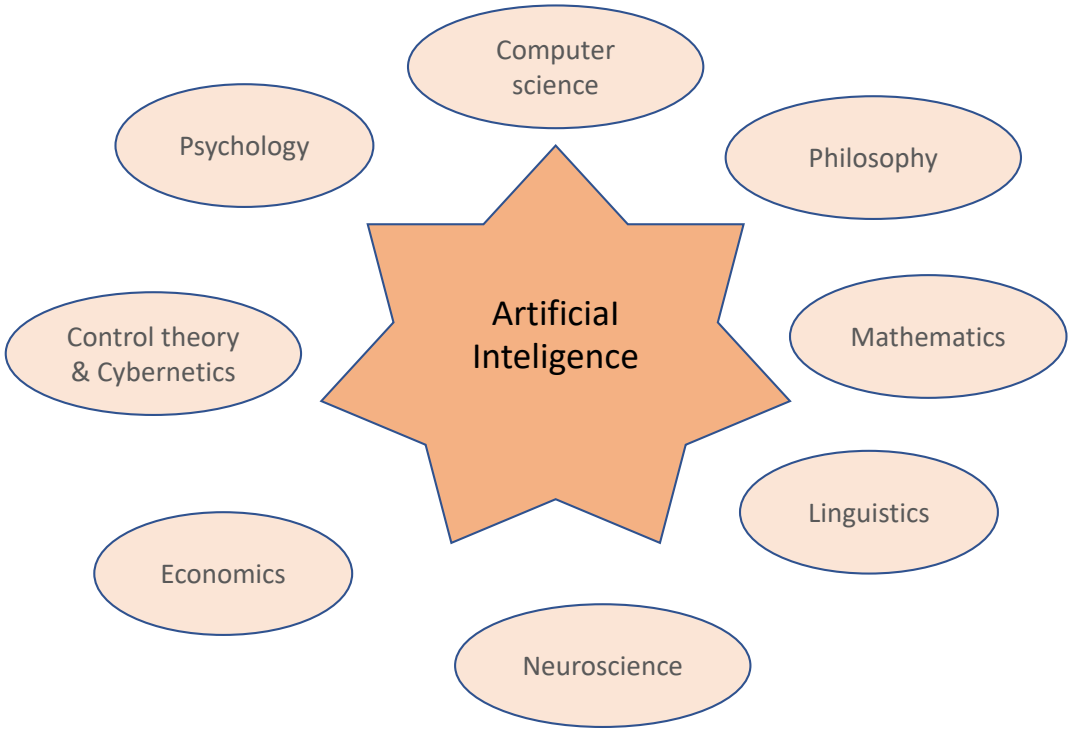
# AI prehistory - grounding disciplines

Computer Science is the main discipline underlying AI

Can we think which other disciplines AI is grounded on?



# AI prehistory - grounding disciplines



# AI prehistory

| | |
|---|---|
| **Philosophy** | logic, methods of reasoning |
| | mind as physical system |
| | foundations of learning, language, rationality |
| **Mathematics** | logic, formal representation |
| | algorithms, computation, probability theory |
| **Statistics** | modeling uncertainty, learning from data |
| **Psychology** | phenomena of perception and motor control |
| | cognitive psychology |

# AI prehistory II

| | |
|---|---|
| **Economics** | formal theory of rational decisions, utility |
| **Linguistics** | knowledge representation |
| | grammar, syntax, semantics |
| **Neuroscience** | neurons as information processing units |
| | synapse as learning mechanism |
| **Control theory** | homeostatic systems, stability |
| | simple optimal agent designs, maximize objective function |
| **Computer science** | engineering, hardware |
| | computational complexity theory |

# Birth of AI field

A Summer Research Project in the Dartmouth College, in 1956 was the birth of the AI research field.

*(From left: Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff)*

McCarth: "...to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

# Brief history of AI

| | |
|---|---|
| 1943 | McCulloch & Pitts: Boolean circuit model of brain |
| 1950 | Turing's "Computing Machinery and Intelligence" |
| 1952-69 | Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine |
| 1956 | Dartmouth meeting: "Artificial Intelligence" adopted |
| 1962 | Rosenblatt's Perceptron for training simple neural networks |
| 1965 | Robinson's complete algorithm for logical reasoning |
| 1972 | The logic programming language PROLOG is created. |
| 1966–74 | Disappointment: AI discovers computational complexity Neural network research almost disappears |

# Brief history of AI -cont.

| | |
|---|---|
| 1969–79 | Early development of knowledge-based systems |
| 1980–88 | Expert systems industry booms |
| 1988–93 | Expert systems industry busts: "AI Winter" |
| 1985–95 | Backpropagation learning returns neural networks to popularity |
| 1988– | Resurgence of probability; general increase in technical depth<br>"Nouvelle AI": ALife, GA |
| 1995– | Agents, agents, everywhere ... |
| 2003– | Human-level AI back on the agenda |
| 2005-2010 | 2005-2010 AI disappoints again, AI is not much appreciated |

# Example from Early History of AI

1943 McCulloch & Pitts: Boolean circuit model of brain

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

## A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

### *I. Introduction*

Theoretical neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a soma and an axon. Their adjunctions, or synapses, are always between the axon of one neuron and the soma of another. At any instant a neuron has some threshold, which excitation must exceed to initiate an impulse. This, except for the fact and the time of its occurrence, is determined by the neuron, not by the excitation. From the point of excitation the impulse is propagated to all parts of the neuron. The velocity along the axon varies directly with its diameter, from less than one meter per second in thin axons, which are usually short, to more than 150 meters per second in thick axons, which are usually long. The time for axonal conduction is consequently of little importance in determining the time of arrival of impulses at points unequally remote from the same source. Excitation across synapses oc-

# McCulloch and Pitts 1943

**Goal:** to understand how the brain produces complex thoughts ("propositions") by using (simple) neurons
- model of a neuron, axon, dendrode - "MCP neuron"
- network of neurons
- transfer of information through on/off mechanism of neurons
- every network of MCP neurons encodes some logical proposition

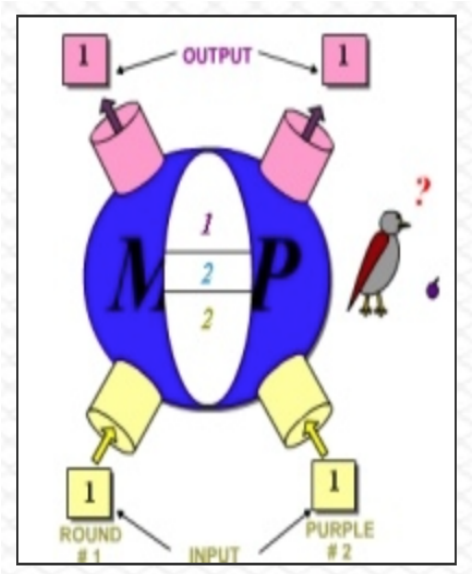
# McCulloch and Pitts 1943 - "Bird example"

Example from:
http://www.mind.ilstu.edu/curriculum/mcp_neurons/mcp_neuron_1.php?modGUI=212&compGUI=1749&itemGUI=3018

Assume a bird will "decide" to eat/not an object (e.g., blue berry, orange, basketball, daisy). It receives/perceives two inputs about the object:
- shape: round or not
- colour: purple or not
- decision: eat only if round purple object (e.g., blue berry)

| Object | Purple? | Round? | Eat? |
|---|---|---|---|
| Blueberry | Yes | Yes | Yes |
| Golf ball | No | Yes | No |
| Violet | Yes | No | No |
| Hot Dog | No | No | No |

# Bird example - cont.




- two inputs (each takes value 0 or 1)
- threshold T
- output is a function of inputs and T
- **IF**(sum(inputs))>= T
  **THEN** Output=1

| Object | Purple? | Round? | Eat? |
|---|---|---|---|
| Blueberry | 1 | 1 | 1 |
| Golf ball | 0 | 1 | 0 |
| Violet | 1 | 0 | 0 |
| Hot Dog | 0 | 0 | 0 |

| Object | Purple? | Round? | Total | Greater than or equal to threshold of 1? | Eat? |
|---|---|---|---|---|---|
| Blueberry | 1 | 1 | 2 | Yes | 1 |
| Golf ball | 0 | 1 | 1 | Yes | 1 |
| Violet | 1 | 0 | 1 | Yes | 1 |
| Hot Dog | 0 | 0 | 0 | No | 0 |

# Some killer apps in AI

*1991* During the Gulf War, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people. Saved the US more money than spent on all AI research since 1950
*1997* Deep Blue (IBM) defeated world chess champion Gerry Kasparov

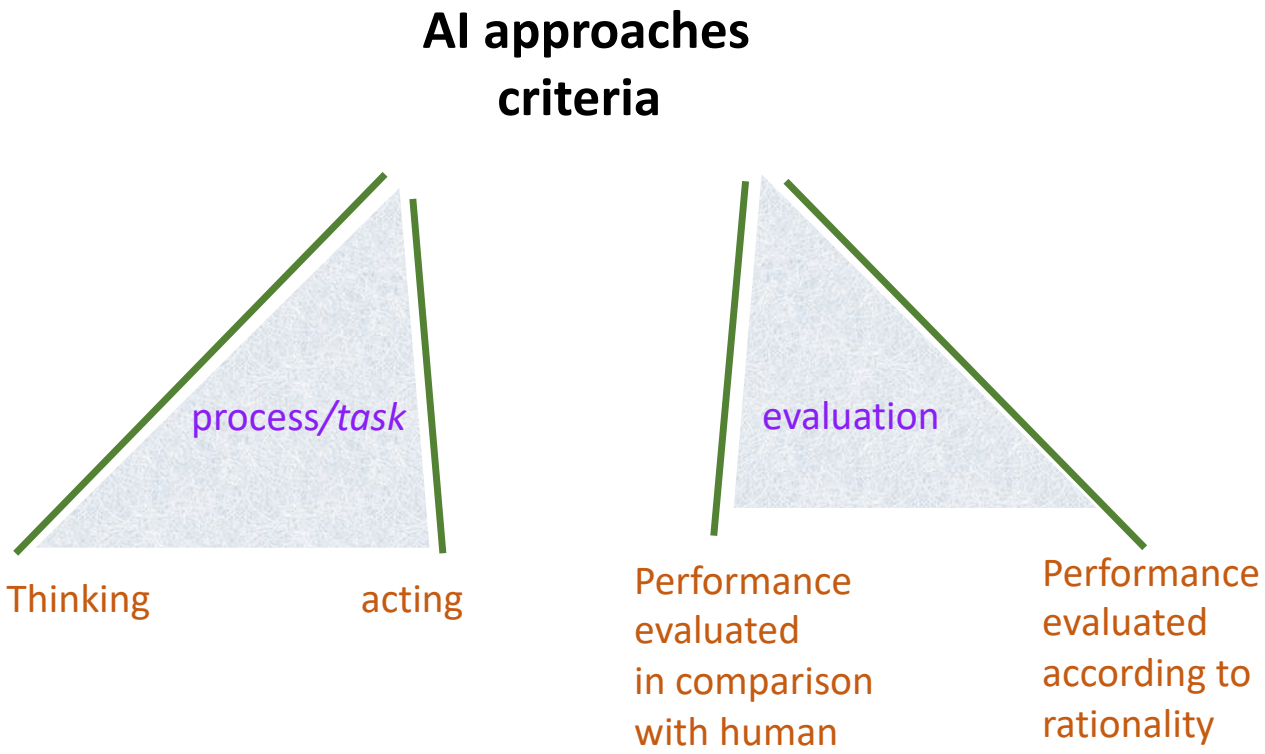
*2011* Watson (IBM) beat human champions on "Jeopardy"

*2012* Google car obtains driver's license in Nevada, US. By 2014, the cars have driven for 1.1 million km without accident

*2017* DeepMind's Alphago AND Elon Musk's A.I. Destroys Champion Gamer! `https://www.youtube.com/watch?v=XbDmxE0j90Y`

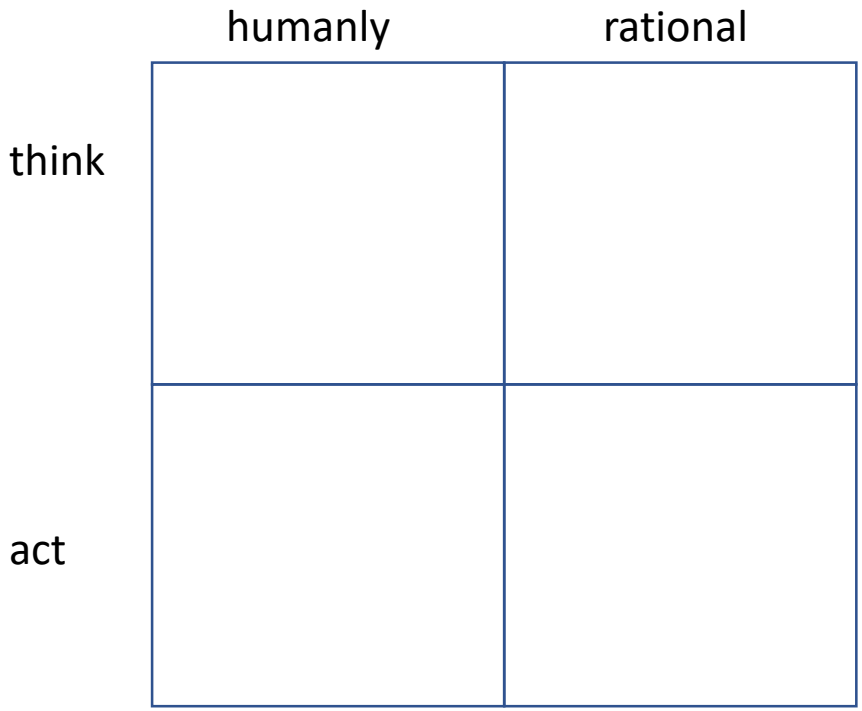
# What is AI?

Russell and Norvig's definition of AI
Two dimensions:



# What is AI?

Types/approaches to AI - as defined by Russel and Norvig:



# What is AI? -cont

## How R&N define AI

| thinking vs. acting \ humanly vs. rationally | | |
|---|---|---|
| | Systems that think like humans | Systems that think rationally |
| | Systems that act like humans | Systems that act rationally<br>Rational Agents |

# Thinking humanly: Cognitive Science

1960s "**cognitive revolution**": information-processing psychology replaced prevailing orthodoxy of **behaviorism**.
Tries to form computational theories of internal activities of the brain.
How to validate? Requires

1. Predicting and testing behavior of human subjects (top-down), or
2. Direct identification from neurological data (bottom-up)

Both approaches (roughly, **Cognitive Science** and**Cognitive Neuroscience**) are now distinct from AI

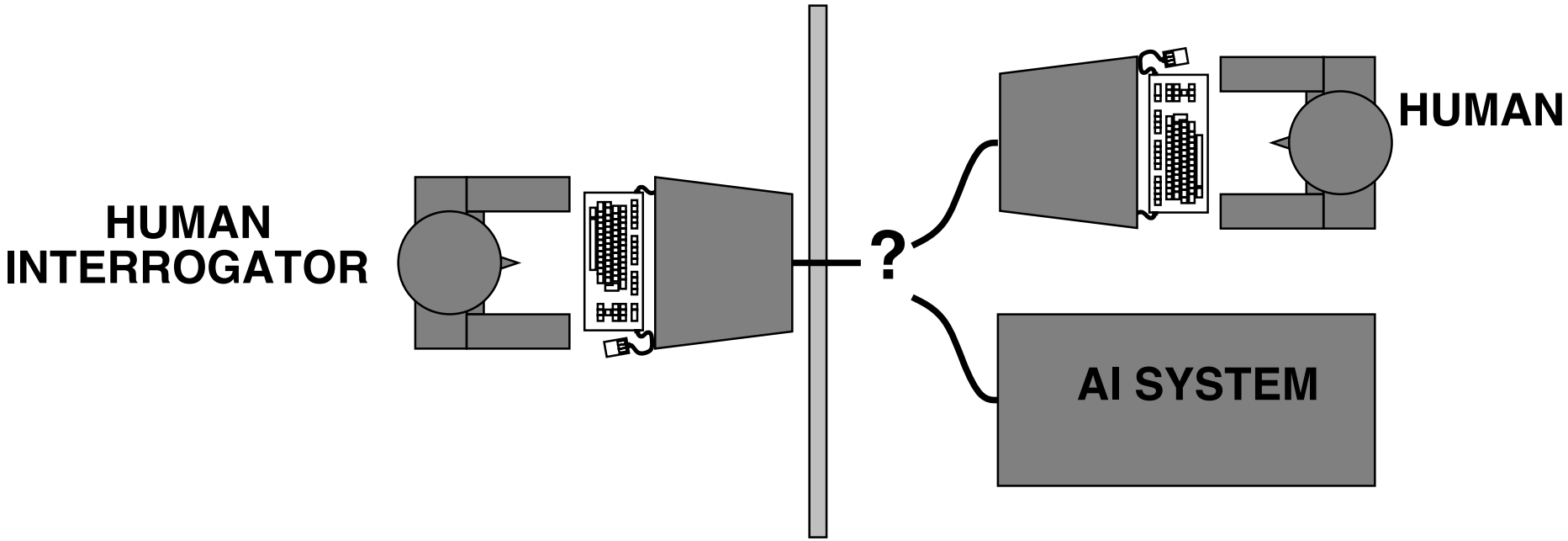
Both share with AI the following characteristic:
the available theories do not explain (or engender)
anything resembling human-level general intelligence

# John McCarthy talks about AI

https://www.youtube.com/watch?v=0zipf13jRr4

# Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":
Can machines think?" $\longrightarrow$ "**Can machines behave intelligently**?"
Operational test for intelligent behavior: the **Imitation Game**



# Acting Humanly - Eliza

- One of the most famous early AI programs: Eliza, the computer psychotherapist, created by Joseph Weizenbaum in 1966 at MIT.
- Eliza functions by "twisting the statements of her 'patients' back at them in the classic manner of a non-directive psychotherapist."
- The fact that it understands the subject's statements is an illusion.
- Surprisingly, many users were taking its performance quite seriously.

# Acting Humanly - Eliza 2

**young woman:** Men are all alike.
**eliza:** In what way?
**young woman:** They're always bugging us about something specific or other.
**eliza:** Can you think of a specific example?
**young woman:** Well, my boyfriend made me come here.
**eliza:** Your boyfriend made you come here?
**young woman:** He says I'm depressed much of the time.
**eliza:** I am sorry to hear you are depressed.
**young woman:** It's true. I'm unhappy.
**eliza:** Do you think coming here will help you not to be unhappy?
**young woman:** I need some help, that much seems certain.
**eliza:** What would it mean to you if you got some help?
**young woman:** Perhaps I could learn to get along with my mother.
**eliza:** Tell me more about your family.
**young woman:** My mother takes care of me.
**eliza:** Who else in your family takes care of you?
**young woman:** My father.
**eliza:** Your father.
**young woman:** You are like my father in some ways.
**eliza:** What resemblance do you see?
**young woman:** You are not very aggressive but I think you don't want me to notice that.
**eliza:** What makes you think I am not very aggressive?
**young woman:** You don't argue with me.
**eliza:** Why do you think I don't argue with you?
**young woman:** You are afraid of me.
**eliza:** Does it please you to believe I am afraid of you?
**young woman:** My father is afraid of everybody.
**eliza:** What else comes to mind when you think of your father?
**young woman:** Bullies.
**eliza:** Does that have anything to do with the fact that your

# How humany is Sofia (Hong Kong Kong firm Hanson Robotics)?

https://www.youtube.com/watch?v=suRuQbDXcrc

# Thinking rationally: Laws of Thought-1

- Aristotle: what are correct arguments/thought processes?
  Formalize "correct" reasoning using a mathematical model
- Several Greek schools developed various forms of **logic**:
  *notation* and *rules of derivation* for thoughts;
- Direct line through mathematics and philosophy to modern AI

# Thinking rationally: Laws of Thought-2

# Thinking rationally: Laws of Thought-3

# Thinking rationally: Laws of Thought -4

Problems:
1) Not all intelligent behavior is mediated by logical deliberation
2) What is the purpose of thinking? What thoughts *should* I have out of all the thoughts (logical or otherwise) that I *could* have?
3) Formalizing (informal) common sense knowledge is difficult
4) General deductive inference is computationally intractable

# Acting rationally: Rational Agents approach

- **Rational** behavior: doing the right thing
- The right thing: that which is expected to maximize goal achievement, given the available information and computational abilities
- Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

# Two main AI paradigms

**Good old fashioned AI** (GOFAI)
**Situated embodied AI** (SEAI)

# Good old fashioned AI

- Most of the successes of AI in the 1970s and 1980s were due to research based on Newell's and Simon's *Physical Symbol System Hypothesis*
- A physical symbol system has the necessary and sufficient means for general intelligent action.
- Newell and Simon viewed intelligence as *symbol manipulation*, and hypothesized that it didn't make difference what physical medium – brain, paper, or computer – was used to do the symbol manipulation
- Hence a special emphasis on *symbolic representations*, which can be interpreted as representing situations in the real world, e.g. "a block world"

# Critiques of GOFAI

- **Searle:** a program (or any physical symbol system) could not be said to understand the symbols that it uses; the symbols have no meaning for the machine
- **Brooks:** our most basic skills of motion, survival, perception, balance etc. do not seem to require high level symbols at all; the use of high level symbols was more complicated and less successful
- **Harnad:** the **symbol grounding** problem: an agent does not perceive symbols, instead the brain converts sensory inputs into higher level abstractions, e.g. symbols

# Emergence of SEAI

- The GOFAI approaches turned out to be brittle and very little robust when deployed on real-world problems
- Trying to define a model of the world turned out to be quite hard - this led to Brooks' statement that "the world is its own best model"
- Situated and embodied AI focuses on having a body (i.e. motor skills) in a physical environment
- Swarm intelligence, subsymbolic AI, genetic algorithms, neural networks - however, this is not a big part of this course, this is covered in more detail in other AI courses.

# Course topic overview

**Next Lecture:** Intelligent Agents (chapter 2 in R&N book)

# More Videos

**Computers with common sense**( Doug Lenat):
https://www.youtube.com/watch?v=2w_ekB08ohU

**Artificial Intelligence: The Common Sense Problem** (Hubert Dreyfus)
https://www.youtube.com/watch?v=SUZUbYCBtGI

**AI Boom** (Rodney Brooks):
https://www.youtube.com/watch?v=ovbzjLobiDo

**IBM Watson :** https://www.youtube.com/watch?v=_Xcmh1LQB9I

**Jeopardy :** https://www.youtube.com/watch?v=WFR3lOm_xhE

**State of the art AI** (Minsky):
https://www.youtube.com/watch?v=aODnFdU_hds