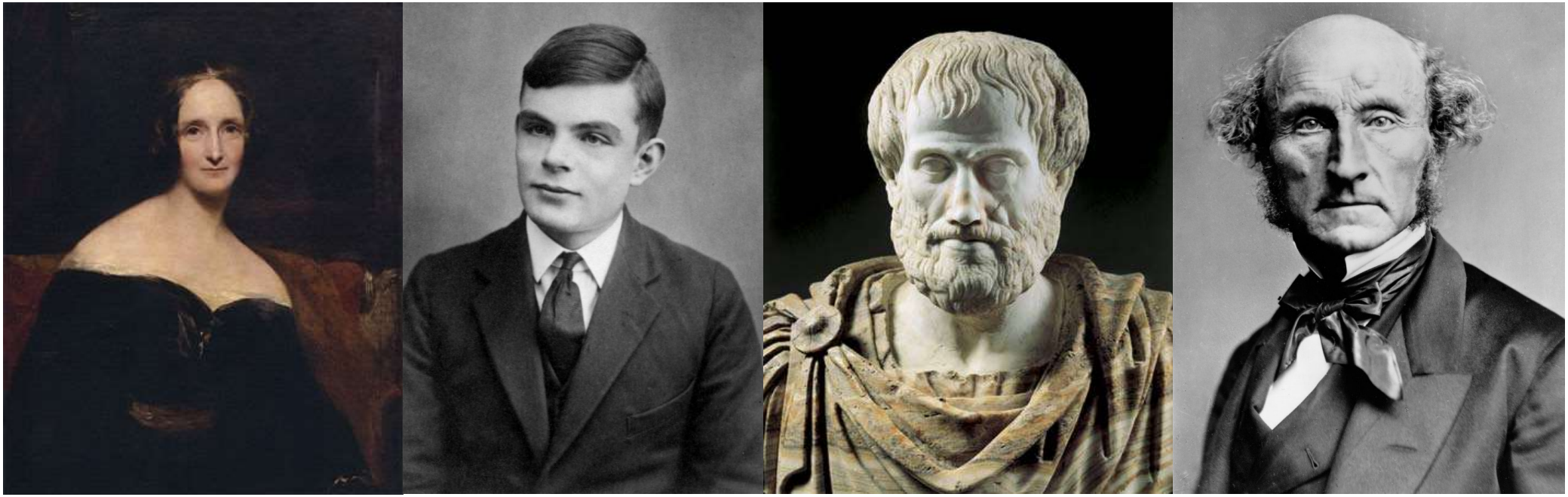
CS440/ECE448: Artificial Intelligence
Lecture 1: What is AI?



CS440/ECE448 Lecture 1: What is AI?

1. Administration: Overview of the Syllabus
2. A two-bit summary of the philosophy of AI
3. Thinking like a Human
4. Acting like a Human
5. Thinking Rationally
6. Acting Rationally

1. Administrative Questions

- Web page
- How is this course graded?
- Policies
- How can I get help?

Web page

Duplicate Websites:

http://courses.engr.Illinois.edu/cs440/

http://courses.engr.Illinois.edu/ece448/

How is this course graded?

- 40%: Exams
  - Mostly from the slides.
  - There will be sample problems included in every deck of slides.
- 60%: MPs (Mini-Projects)
  - Each MP is designed to require about 19 hours of work, including ~14 hours of thinking/ coding/ debugging and ~5 hours of waiting for your computer. Seriously. We really do target 19 hours.
  - (4 credit students: ~25 hours)
  - You can work in teams of up to 3, only if it helps you. Software management exercise.

Policies

- Late MPs:
  - Only if every member of your team has an emergency documented by the emergency dean.
  - If no emergency, penalty is 25% per day.
- Partially late MPs: no such thing
  - We can grade the whole MP as late, or
  - We can grade only the part you submitted on time.
  - Your choice.
- Plagiarism
  - Please DO search online to find good ideas.
  - Please LEARN THE IDEAS, don't COPY THE CODE.
  - Graders will read on-line code repos before grading your MP.

How Can I Get Help?

- Office Hours:
  - Mostly in ECEB 5034. Times listed here: https://courses.engr.illinois.edu/ece448/sp2018/homework.html
- Piazza:
  - https://piazza.com/class/jc8mft43dmb4gu
  - Teaching staff will check piazza at least once/day
  - Fellow students strongly encouraged to give good answers. Extra credit may be given for useful piazza answers.
  - DON'T post code on piazza, either for questions or for answers. You can post pseudo-code if you want.
- Personal e-mails/phone calls/personal conversations:
  - We smile, but we won't be able to help you very much.
- Wikipedia etc: Often very useful. See previous slide.

2. A two-bit summary of the philosophy of AI

What is Artificial Intelligence?

- Hasegawa-Johnson's Principle of Jargon Understanding:

  If you're trying to understand a piece of jargon, always start by checking the definitions of the component words.

- Artificial (adj., Wiktionary): Man-made, i.e., constructed by means of skill or specialized art.

- Intelligence (noun, Wiktionary): Capacity of mind to understand meaning, acquire knowledge, and apply it to practice.

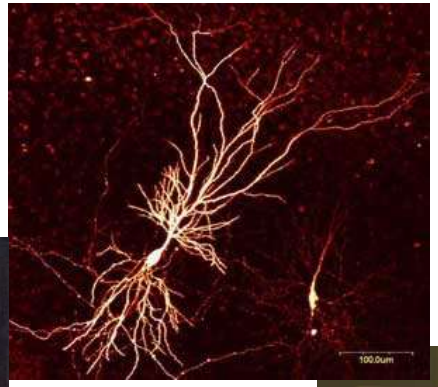
- Artificial Intelligence (implied by above): capacity of a man-made system to understand, acquire, and apply knowledge.

What is Artificial Intelligence?

- Candidate definitions from the textbook:

1. Thinking humanly	**2. Acting humanly**
3. Thinking rationally	**4. Acting rationally**

3. Thinking like a Human



Mary Shelley, author of *Frankenstein: The Modern Prometheus*; Neuron, showing branching of the dendrites; EEG cap; Cortical connectivity map, computed using diffusion tensor MRI

Sample problems

1. Suppose the brain has 100 trillion neurons. How many binary computations per second can the brain perform?

2. The best supercomputers perform far more computations/second than the human brain. If that's true, why have we not yet duplicated a human brain?

How many computations/second?

- Hodgkin-Huxley neuron:
  - Neural computations are binary. Each neuron is either generating an action potential, or not.
  - Action potentials at rates between 1Hz and 1000Hz (1 to 1000 times/second)
  - Each neuron's action potential is communicated to a set of other neurons --- usually 100-1000 other neurons.

Sample problems

1. Suppose the brain has 100 trillion neurons. How many binary computations per second can the brain perform?

Answer: if each neuron performs 1-1000 binary computations/second, then the brain performs up to (100 trillion)X(1000) = 10^17 binary computations/second (100 Peta-ops)

Sample problems

1. Suppose the brain has 100 trillion neurons. How many binary computations per second can the brain perform?

2. The best supercomputers perform far more computations/second than the human brain. If that's true, why have we not yet duplicated a human brain?

Modern neuroimaging techniques

- EEG (electro-encephalography)
  - Good temporal resolution: ~1000 samples/second
  - Poor spatial resolution: ~128 channels for the whole brain. "EEG activity therefore always reflects the summation of the synchronous activity of thousands or millions of neurons that have similar spatial orientation."
- fMRI (functional magnetic resonance imaginge)
  - Better spatial resolution: ~1mm/voxel, ~2000 voxels/brain (vs. 100 trillion neurons)
  - Poor temporal resolution: ~2 seconds/sample
- ECOG (electrocorticography)
  - Spatial resolution of fMRI + temporal resolution of EEG
  - Only for the part of the brain that has been surgically revealed, for a living thinking human.

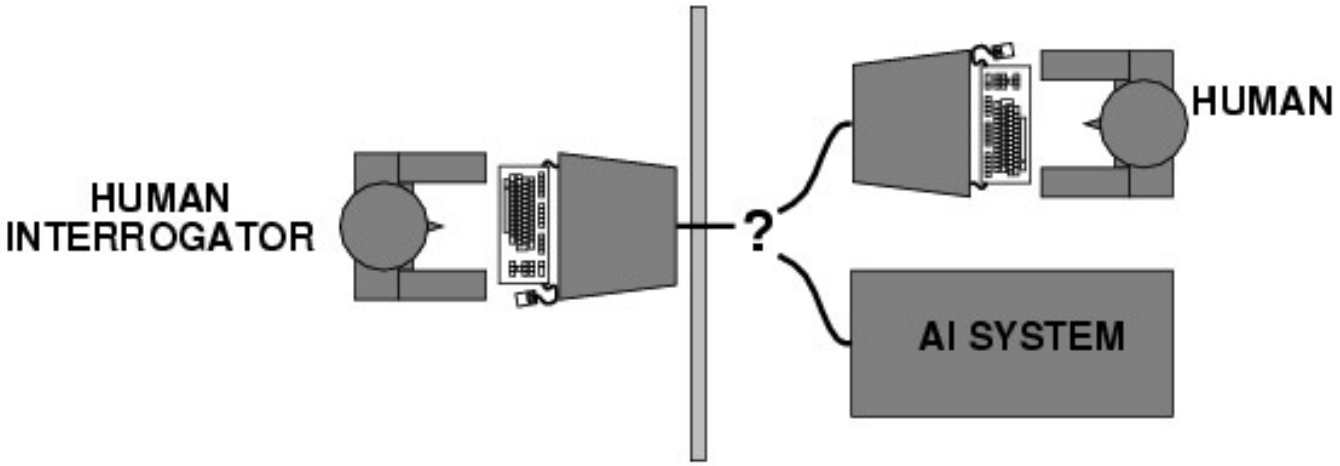
Sample problems

1. Suppose the brain has 100 trillion neurons. How many binary computations per second can the brain perform?

2. The best supercomputers perform far more computations/second than the human brain. If that's true, why have we not yet duplicated a human brain?

Answer: the best available neuro-imaging techniques can measure about 100,000 samples/second, versus 100 quadrillion binary computations being done in the brain. Therefore we can't yet measure what's actually being computed in the brain.

4. Acting like a Human



Schematic of the Turing test; Alan Turing

The Turing Test

- Alan Turing, “Intelligent Machinery,” 1947:

It is not difficult to devise a paper machine which will play a not very bad game of chess. Now get three men as subjects for the experiment. A, B and C. A and C are to be rather poor chess players, B is the operator who works the paper machine. Two rooms are used with some arrangement for communicating moves, and a game is played between C and either A or the paper machine. C may find it quite difficult to tell which he is playing.

We now ask the question, “What will happen when a machine takes the part of A in this game?” Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, “Can machines think?”

- What capabilities would a computer need to have to pass the Turing Test?
  - Natural language processing
  - Knowledge representation
  - Automated reasoning
  - Machine learning
- Turing predicted that by the year 2000, machines would be able to fool 30% of human judges for five minutes

A. Turing, Computing machinery and intelligence, Mind 59, pp. 433-460, 1950

What's wrong with the Turing test?

- Variability in protocols, judges
- Success depends on deception!
- Chatbots can do well using "cheap tricks"
  - First example: ELIZA (1966)
  - Javascript implementation of ELIZA

A better Turing test?

- **Winograd schema:** Multiple choice questions that can be easily answered by people but cannot be answered by computers using “cheap tricks”

•*The trophy would not fit in the brown suitcase because it was so small.*
What was so small?

- *The trophy*
- *The brown suitcase*

H. Levesque, *On our best behaviour*, IJCAI 2013

http://www.newyorker.com/online/blogs/elements/2013/08/why-cant-my-computer-understand-me.html

A better Turing test?

- **Winograd schema:** Multiple choice questions that can be easily answered by people but cannot be answered by computers using "cheap tricks"

•*The trophy would not fit in the brown suitcase because it was so **large**.*
*What was so **large**?*

- *The trophy*
- *The brown suitcase*

H. Levesque, *On our best behaviour*, IJCAI 2013

http://www.newyorker.com/online/blogs/elements/2013/08/why-cant-my-computer-understand-me.html

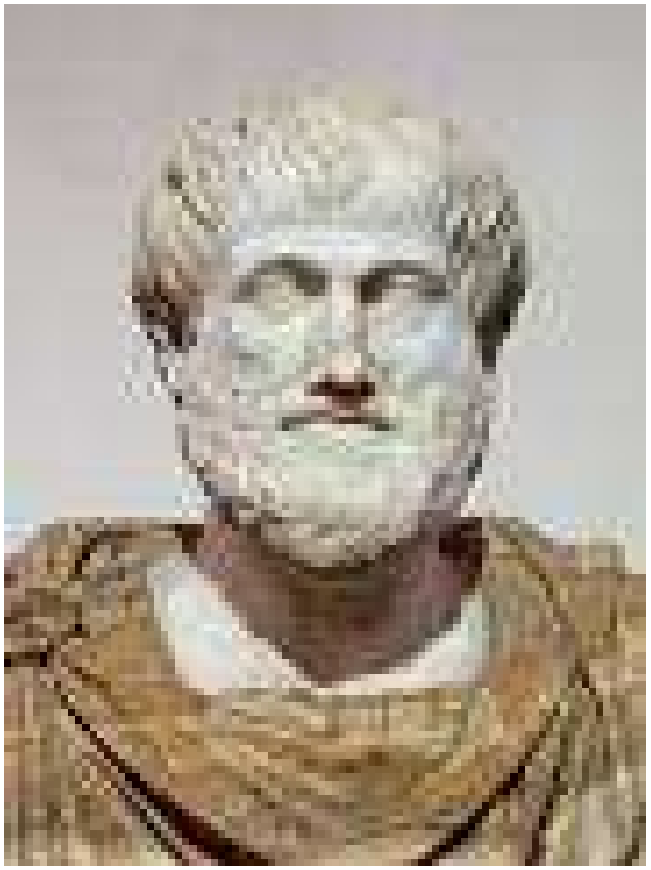
Winograd schema

- Advantages over standard Turing test
  - Test can be administered and graded by machine
  - Does not depend on human subjectivity
  - Does not require ability to generate English sentences
  - Questions cannot be evaded using verbal dodges
  - Questions can be made “Google-proof” (at least for now…)
- Winograd schema challenge
  - Held at IJCAI conference in July 2016
  - Six entries, best system got 58% of 60 questions correct (humans get 90% correct)

Sample questions

- In what way can it be said that a machine that passes the Turing test is intelligent?
- In what way can it be said that a machine that passes the Turing test is _not_ intelligent?
- Give a few reasons why the Winograd schema is a better test of intelligence than the Turing test

AI definition 3: Thinking rationally



Aristotle, 384-322 BC

AI definition 3: Thinking rationally

- Idealized or “right” way of thinking
- **Logic:** patterns of argument that always yield correct conclusions when supplied with correct premises
  - *“Socrates is a man; all men are mortal; therefore Socrates is mortal.”*
- **Logicist approach to AI:** describe problem in formal logical notation and apply general deduction procedures to solve it
- Problems with the logicist approach
  - Computational complexity of finding the solution
  - Describing real-world problems and knowledge in logical notation
  - Dealing with uncertainty
  - A lot of “rational” behavior has nothing to do with logic

Syllogism

- Syllogism = a logical argument that applies deductive reasonining to arrive at a conclusion based on two or more propositions that are asserted to be true.

- Example Problem (you should know this from binary logic classes):

  - Given: $p \Rightarrow q$
  - Given: $q \Rightarrow r$
  - Given: q is false
  - Which of the following are true?

    a. p is true

    b. p is false

    c. r is true

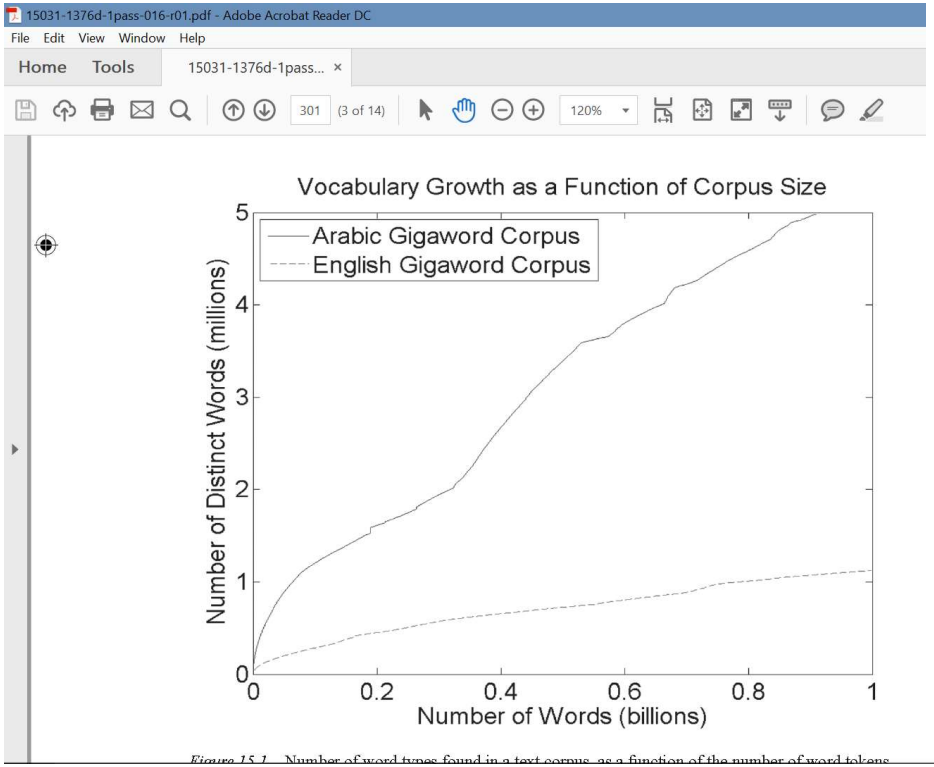
    d. r is false

Successes of Logicist Approach: Expert Systems

- Expert system = (knowledge base) + (logical rules)
  - Knowledge base = easy to collect from human judges and/or encyclopedia
  - Logical rules = easy to deduce from examples, and easy to verify by asking human judges
  - Combination of the two: able to analyze never-before-seen examples of complicated problems, and generate the correct answer

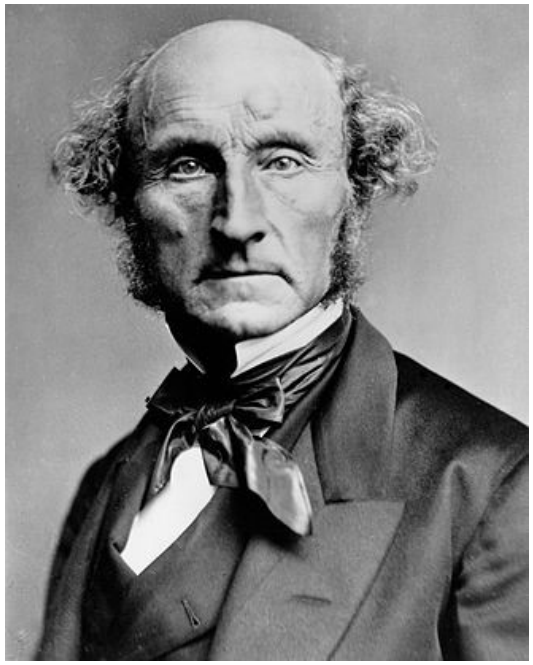
- Example: speed control system of the https://en.wikipedia.org/wiki/Sendai_Subway_Namboku_Line. "This system (developed by Hitachi) accounts for the relative smoothness of the starts and stops when compared to other trains, and is 10% more energy efficient than human-controlled acceleration."

Failures of Logicist Approach: Robust AI



- Humans commonly believe that there are a finite number of facts that must be entered into a knowledge base. Evidence suggests that this is incorrect.
- Example (Hasegawa-Johnson, Elmahdy & Mustafawi, “Arabic Speech and Language Technology,” 2017): the number of distinct words in any corpus of text is linearly proportional to the number of words. In English, a never-before-seen word occurs ~once/1000 words; in Arabic, ~once/180 words.

AI definition 4: Acting rationally



John Stuart Mill, 1806-1873

AI definition 4: Acting rationally

- A **rational agent** acts to optimally achieve its goals
  - Goals are application-dependent and are expressed in terms of the **utility of outcomes**
  - Being rational means **maximizing your (expected) utility**
- This definition of rationality only concerns the decisions/actions that are made, not the cognitive process behind them
- An unexpected step: rational agent theory was originally developed in the field of economics
  - Norvik and Russell: "most people think Economists study money. Economists think that what they study is the behavior of rational actors seeking to maximize their own happiness."
- By standard economic definitions: a rational agent is any agent that acts rationally, e.g., human 3-year-olds:

•https://www.youtube.com/watch?v=EjzadtXR-zs

Utility maximization formulation

- Advantages
  - Generality: goes beyond explicit reasoning, and even human cognition altogether
  - Practicality: can be adapted to many real-world problems
  - Naturally accommodates uncertainty
  - Amenable to good scientific and engineering methodology
  - Avoids philosophy and psychology
- Disadvantages
  - Practical disadvantage: In practice, utility optimization is subject to the agent's computational constraints (**bounded rationality** or **bounded optimality**)
  - Theoretical disadvantage: does being human involve anything other than rationality? Are the other aspects of being human important? Would it be useful if a machine had them, whatever they are?
  - What is utility?

What is AI?

1. What is the advantage of defining AI in terms of action, rather than thought? What is its disadvantage?
2. What is the advantage of defining AI in terms of being rational, instead of being human-like? What is its disadvantage?
3. What is the Turing test? What are some problems with the Turing test? How might they be solved? What is the Winograd schema? Why is it better than the Turing test? Why might it fail?
4. What is the logicist approach to AI?
5. What is a rational agent? What is utility?