CS440/ECE448: Artificial Intelligence
(Section R, Spring 2019)

Lecture 1: What is AI?

Julia Hockenmaier

juliahmr@illinois.edu

Welcome to CS440/ECE448 Artificial Intelligence!

Section R (TR, 11am-12:15, DCL1320)
Prof. Julia Hockenmaier (juliahmr@illinois.edu)
TAs: Dhruv Agarwal, Mark Craft, Bryce Kille

Section Q (TR, 12:30-13:45, ECEB1002)
Prof. Mark Hasegawa-Johnson (jhasegaw@illinois.edu)
TAs: Austin Bae, Rahul Kunji, Jialu Li, Jiaxi Nie, Ningkai Wu, Yijia Xu

Email: CS-440-staff@lists.cs.illinois.edu (only from illinois.edu email)

CS440/ECE448 Lecture 1: What is AI?

1. **What will you learn in this class?**
   - What is AI?
   - What will our syllabus cover?

2. **How will we teach this class?**
   - Reading materials
   - Assessment and policies

What will you learn in this class?

What is Artificial Intelligence?

What will our syllabus cover?

What is *Artificial Intelligence*?

- *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 like a human	**2. Acting like a human**
3. Thinking rationally	**4. Acting rationally**

Thinking like a Human: Cognitive Modeling

- **Can we determine how humans think?**
  - Introspection can be misleading, is not scientific
  - Psychological experiments measure behavior
  - Brain imaging measures brain activity (blood flow, electrical activity)
- **Can we reproduce the way humans think in a computer?**
  - Cognitive science and computational neuroscience try to do that
  - But this may be unnecessary for engineering purposes
    (planes don't fly like birds)
  - 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?
  - On some AI tasks, computers outperform humans already (chess)

Acting like a Human: The Turing Test
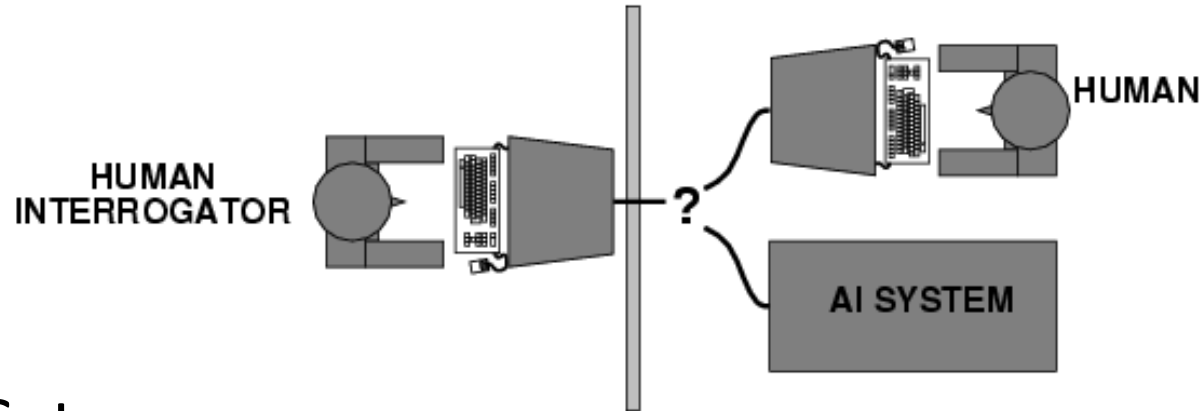
- Setup:
  - A human interrogator poses written questions and reads responses that are either written by a human or by a machine.
  - The interrogator has to tell whether the answers were written by a machine or by a human.
- A machine passes the Turing test if the human interrogator cannot decide whether the answers are written by a machine or human.

Background: 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?"**

Turing predicted that by the year 2000, machines would be able to fool 30% of human judges for five minutes

What capabilities would a computer need to pass the Turing Test?

- Natural language processing
- Knowledge representation
- Automated reasoning
- Machine learning

What's wrong with the Turing Test?

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

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

Winograd schema

- Advantages over standard Turing test
  - Test can be administered and graded by machine
  - Scoring of the test does not depend on human subjectivity
  - Machine does not require ability to generate English sentences
  - Questions cannot be evaded using verbal "tricks"
  - 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

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 your propositional logic classes):

- Given: $p \Rightarrow q$ $q \Rightarrow r$ 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 = database of examples
  - 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 an answer that is often (but not always) correct
- Expert systems = commercial success in the 1970s
  - Radiology, geology, materials science expert systems advised their human users
  - Dating services (match users based on hobbies, etc.)

Successes of Logicist Approach: Fuzzy Logic

Real numbers (e.g.,room temperature)
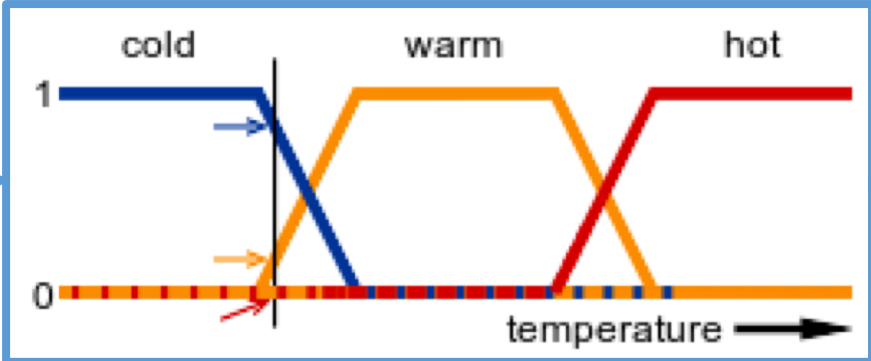
By fullofstars - original (gif): Image:Warm fuzzy logic member function.gif, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2870420
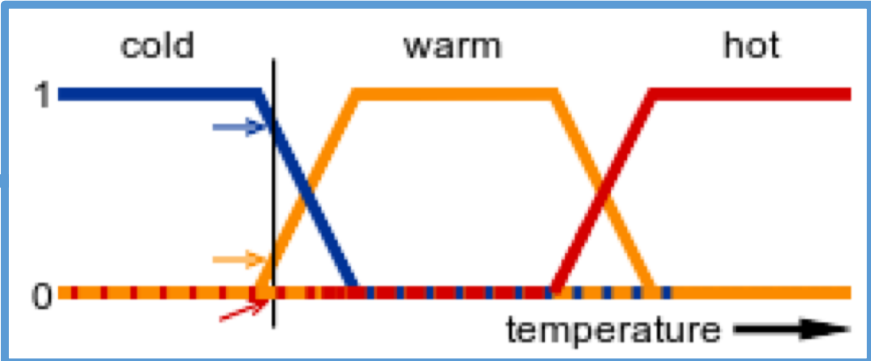
Category Labels (cold, warm, hot)

Logic operations

If cold then turn up the thermostat.

If hot then turn down the thermostat.

Real numbers (e.g., thermostat temperature)

Category Labels (up, down)

# Successes of Logicist Approach: Fuzzy Logic

Example: speed control system of the https://en.wikipedia.org/wiki/Sendai_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: Fragility, and the "AI Winter"

- Expert systems/fuzzy logic work if the number of rules you have to program is small and finite.
- The law of the out-of-vocabulary word: No matter how many words are in your dictionary, there are words you missed.
  - Empirical proof: Hasegawa-Johnson, Elmahdy & Mustafawi, "Arabic Speech and Language Technology," 2017
- Implication: no matter how carefully you design the rules for your expert system, there will be real-world situations that it doesn't know how to handle.
  - This is a well-known problem with expert systems, called "fragility"
  - Corporations and governments reacted to fragility by reducing funding of AI, from about 1966-2009. This was called the "AI Winter."

Failures of Logicist Approach: Humans don't think logically.
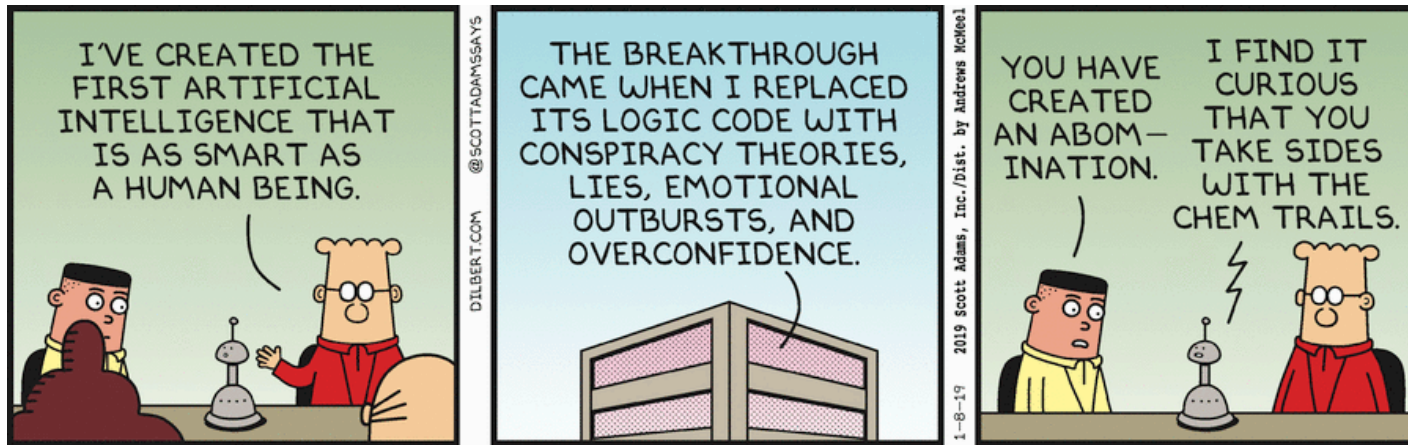
AI definition 4: Acting rationally
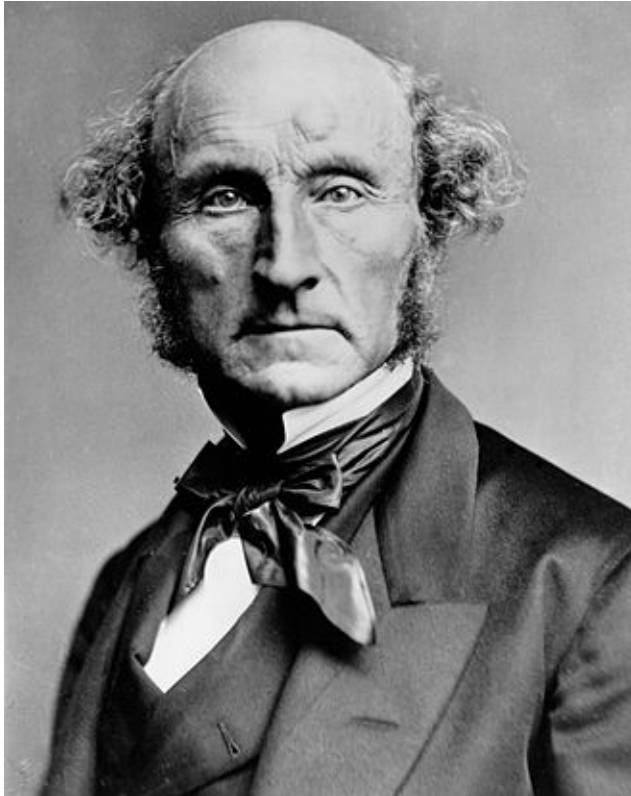
John Stuart Mill, 1806-1873

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
  - Norvig 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."

Utility maximization formulation: Advantages

- **<u>Generality</u>**: goes beyond explicit reasoning, and even human cognition altogether

- **<u>Practicality</u>**: can be adapted to many real-world problems. Avoids philosophy and psychology.

- **<u>Solvability:</u>** Amenable to good scientific and engineering methodology

For all of these reasons, this course will **<u>usually</u>** adopt this definition:

An "artificial intelligence" is a machine that acts rationally (reasons out a plan of action) in order to maximize some measure of utility (a measure of how good is the resulting situation)

Utility maximization formulation: Disadvantages

- **Practical disadvantages**: *can* a machine act rationally in order to achieve a desirable outcome? Why or why not?

- **Theoretical disadvantages**: *should* a machine act rationally in order to achieve a desirable outcome? Why or why not?

So, what is Artificial Intelligence?

- Machine Learning
- Knowledge Representation
- Reasoning and Planning
- Natural Language Processing
- Computer Vision
- Robotics
- …

What will you learn in this class?

- Part 1: Background, History, Terminology
- Part 2: Search, Constraint Satisfaction, Planning, Game Theory
- Part 3: Learning and Probabilistic Models
- Part 4: Impact and Applications of AI

How will we teach this class?

Course Admin

Website, Compass, Piazza, Office Hours, Book

- **Class Website:** For lecture slides, videos, policies etc.
http://courses.engr.illinois.edu/cs440

- **Compass:** To submit assignments, etc.
http://compass2g.illinois.edu

- **Piazza:** For discussions, announcements, etc.
https://piazza.com/class/jp4a2vscy5r76m

- **Office Hours:** to be announced (TAs and professors)

- **Textbook:** S. Russell and P. Norvig, Artificial Intelligence, 3rd Ed.

Piazza policies

- 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.

How is this course graded? (**3 Credit** section)

- **40%: Exams**
  - Mostly from the slides. The page http://courses.engr.Illinois.edu/cs440/lectures.html includes sample problems from the textbook.
- **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.
  - You can work in teams of up to 3, only if it helps you. Software management exercise.

Exam policies

- Exam 1 will be an in-class exam
- Exam 2 will be during the assigned finals slot.
- Exam 2 will only cover material after Exam 1.
- Each exam counts 20% towards the three-hour credit
- No books, cheat sheets, calculators etc. allowed.

- Contact us asap if you need DRES accommodations

MP Policies

- Late MPs:
  - Only if every member of your team has an emergency documented by the emergency dean.
  - If no emergency, penalty is 10% per day.
  - No homework accepted more than 7 days late.
  - DO THE HOMEWORK. Even partly, even 6 days late. If you miss ONE MP, you will probably not pass.
- 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 is this course graded? (4 Credit section)

- 75% determined as for the 3 Credit section
- **25% literature survey**
  - Task: write a survey of a research area/task/problem in AI that covers a number of original research papers (current and/or past)
  - Purpose: Read, explain and critique original research
  - **We will provide more instructions next week**