Artificial Intelligence
CE-417, Group 1
Computer Eng. Department
Sharif University of Technology

Spring 2024s

By Mohammad Hossein Rohban, Ph.D.

Courtesy: Most slides are adopted from CSE-573 (Washington U.), original slides for the textbook, and CS-188 (UC. Berkeley).

Course Information and Policies

Instructor: **Mohammad Hossein Rohban**

Email: rohban@sharif.edu

Office: **806**

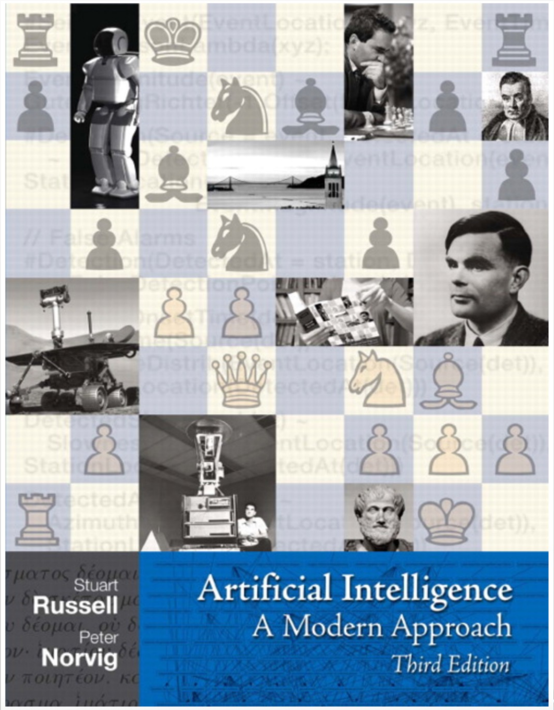
Office hours: in my office by appointment through email

Head TA: Arash Marioriyad, Ilya Mahrooghi, Emad Emam-Jomeh

Course information and Policies (cont.)

Textbook: **Artificial Intelligence: A Modern Approach**, by Stuart Russell and Peter Norvig. 3rd Edition, 2009.

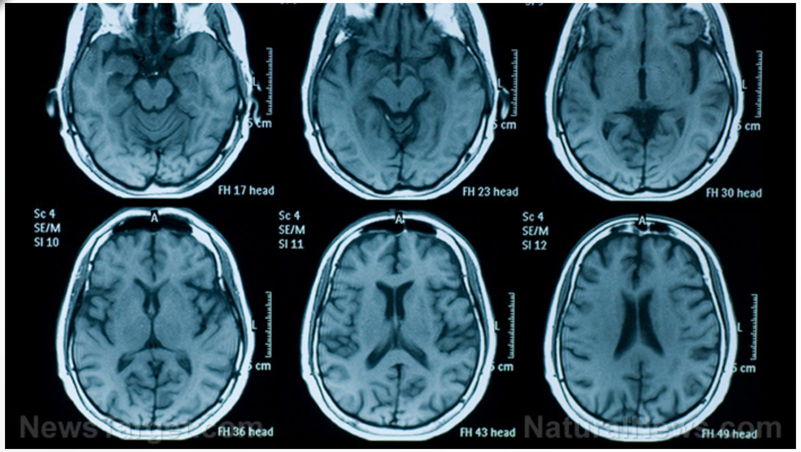
[On a Quera Post]



Prerequisites:

- Knowledge of a programming language
- Data Structures and Algorithms
- **Probability and Statistics**

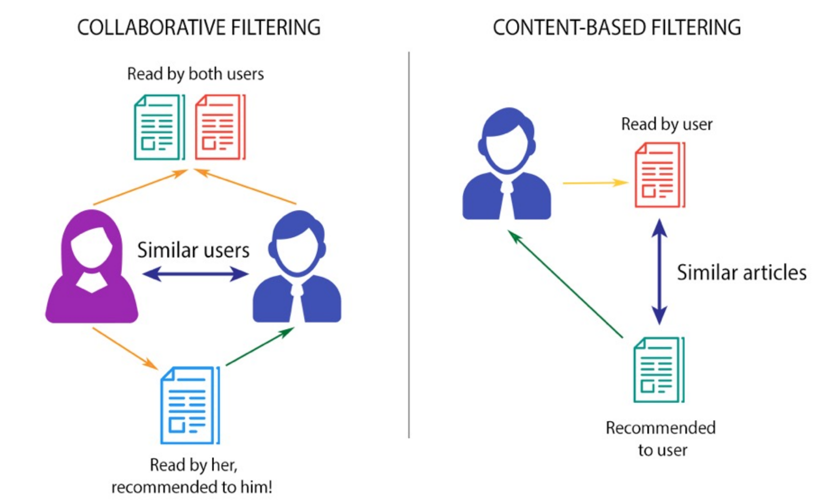
Medical Diagnosis



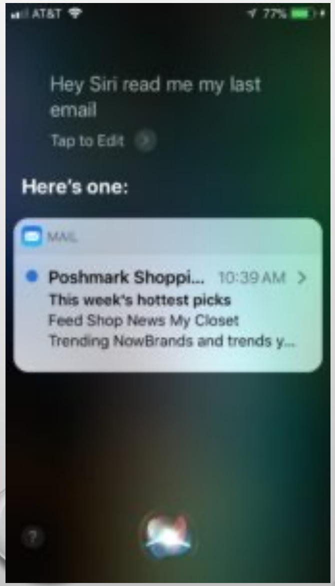
Sentiment Analysis

"The food was very good, but it took over half an hour to be seated, ... and the service was terrible. The room was very noisy and cold wind blew in from a curtain next to our table. Desserts were very good, but because of [the] poor service, I'm not sure we'll ever go back!"

Marketing

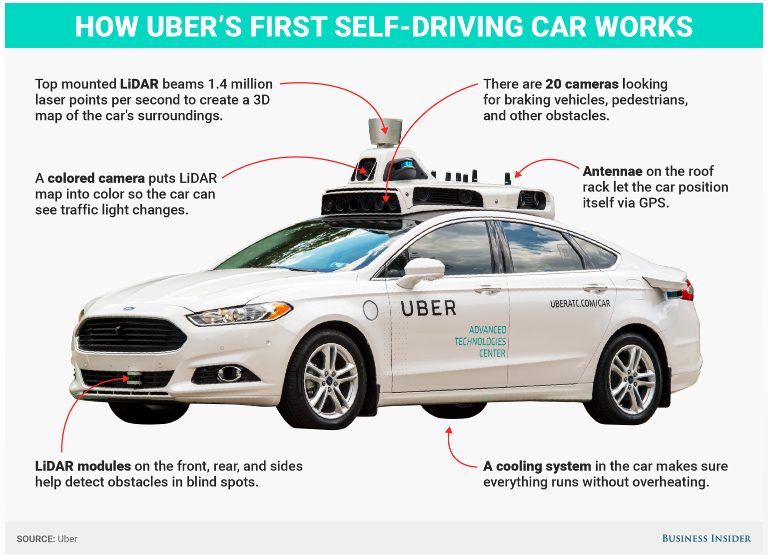


Virtual Assistant



Game Playing

Self-driving Cars



Protein Function Prediction

Stock Market Forecasting

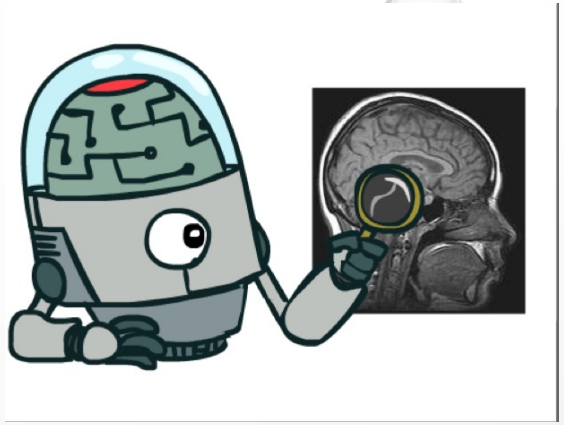
Outline

- What is AI?
- What can AI do nowadays?
- What subjects we will cover in this course?
- Let's begin: Rational Agents

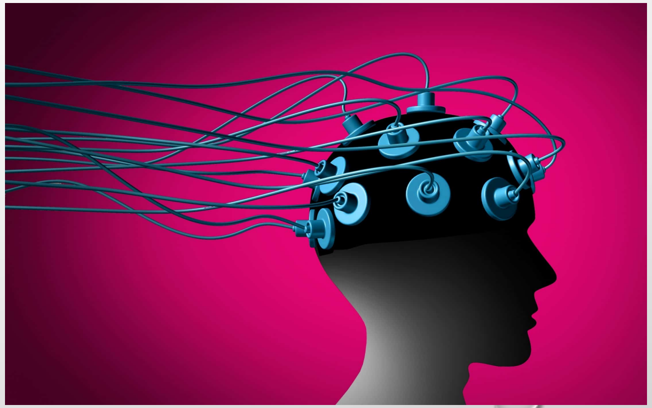
What is AI?

The science of making machines that:

Thinking humanly: Cognitive Science

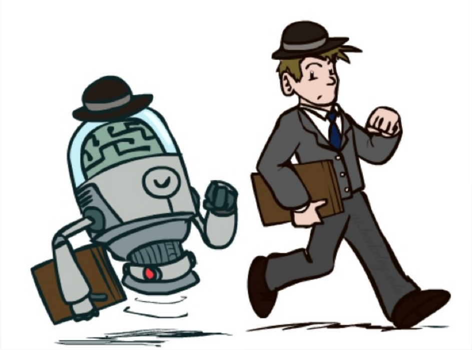


- 1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism.



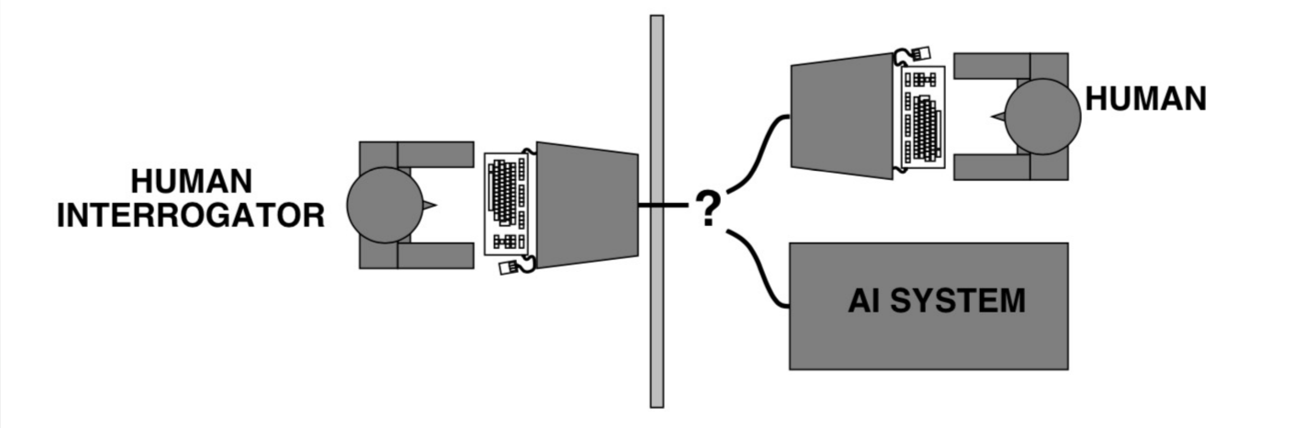
- Requires scientific theories of internal activities of the brain
  - What level of abstraction? "Knowledge" or "circuits"?
  - How to validate? Requires :
    1. Predicting and testing behavior of human subjects (top-down), or
    2. Direct identification from neurological data (bottom-up)
- Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI

Acting Humanly: The Turing Test



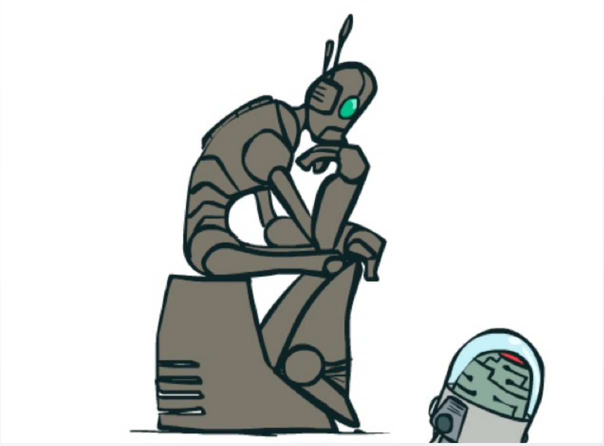
Turing (1950) "computing machinery and intelligence":

- "can machines think?" → "Can machines behave intelligently?"
- Operational test for intelligent behavior: the imitation game



- Caveat: Turing test is not reproducible, constructive, or amenable to mathematical analysis.

Thinking rationally: Laws of Thought

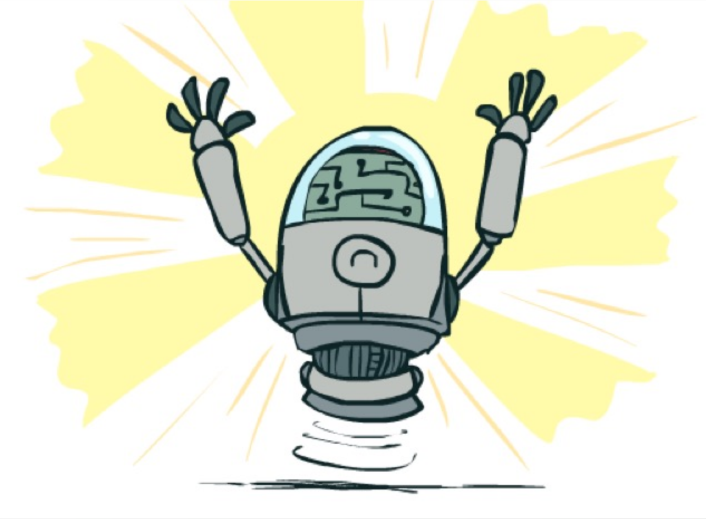


- Aristotle: what are correct arguments/thought processes?
- Several Greek schools developed various forms of logic: notation and rules of derivation for thoughts;
- May or may not have proceeded to the idea of mechanization
- Direct line through mathematics and philosophy to modern AI.
- Caveat: Not all intelligent behavior is mediated by logical deliberation

Rational Decisions

We'll use the term **rational** in a very specific, technical way:

- Rational: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made (not the thought process behind them).
- Goals are expressed in terms of the **utility** of outcomes
- Being rational means **maximizing your expected utility**

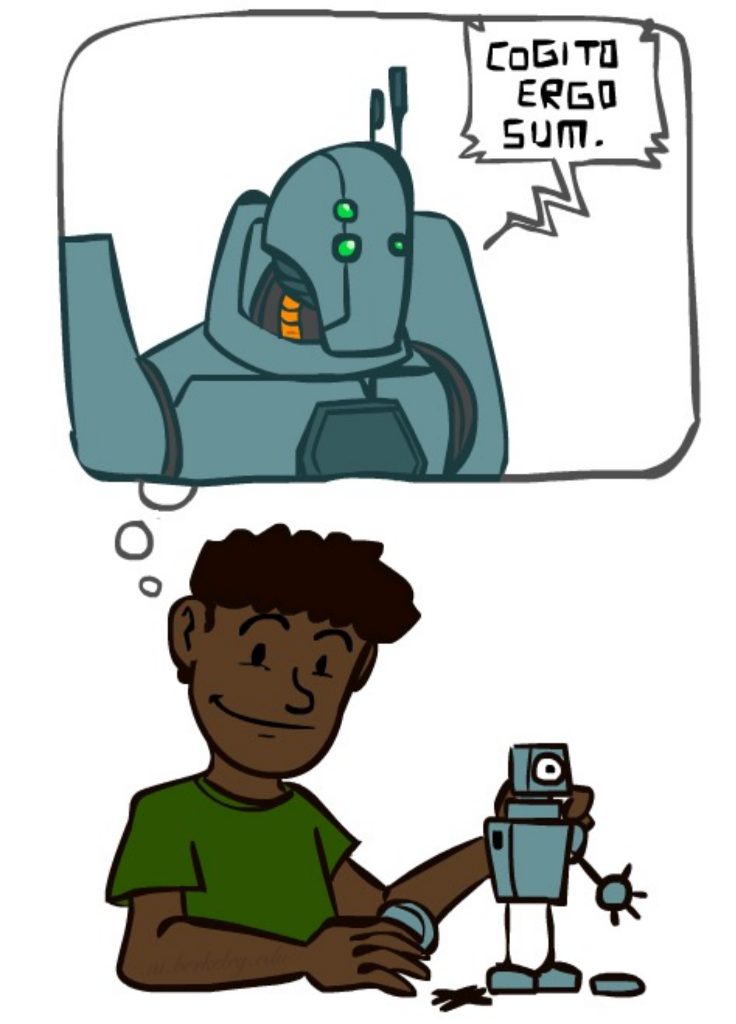


A better title for this course would be:

Computational Rationality

A (Short) History of AI

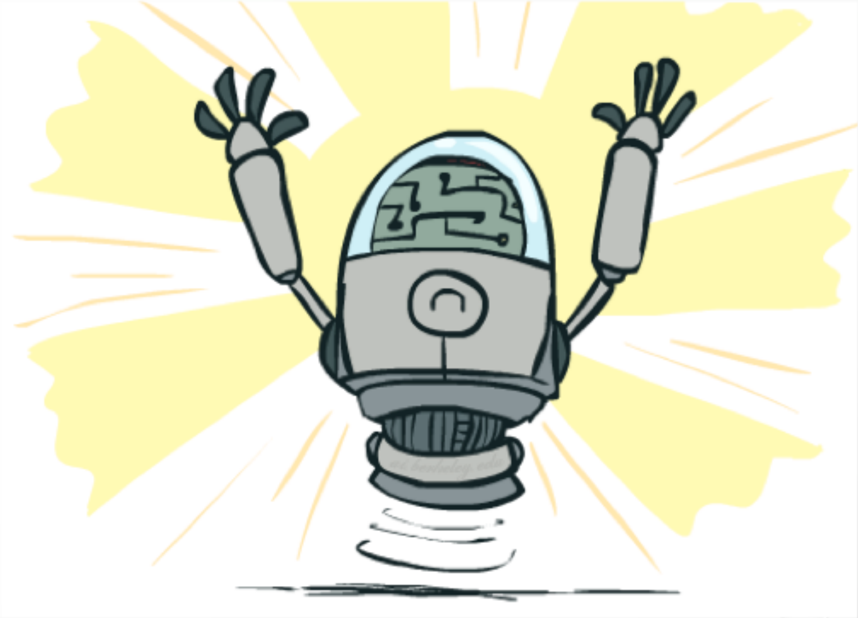
- 1940-1950: Early days
  - 1943: Mcculloch & Pitts: Boolean circuit model of brain
  - 1950: Turing's "computing machinery and intelligence"
- 1950—70: excitement: look, ma, no hands!
  - 1950s: early AI programs, including Samuel's checkers program, Newell & Simon's logic theorist, Gelernter's geometry engine
  - 1956: Dartmouth meeting: "artificial intelligence" adopted
  - 1965: Robinson's complete algorithm for logical reasoning
- 1970—90: knowledge-based approaches
  - 1969—79: early development of knowledge-based systems
  - 1980—88: expert systems industry booms
  - 1988—93: expert systems industry busts: "AI winter"
- 1990—: statistical approaches
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents and learning systems… "AI spring"?
- 2000—: where are we now
- 2012- : Deep Learning
- 2023- : GPTs



What can AI do nowadays?

Which of the following can be done at present?

- ✓ Play a decent game of table tennis?
- ✓ Play a decent game of jeopardy?
- ✓ Drive safely along a curving mountain road?
- ? Drive safely along telegraph avenue?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Berkeley bowl?
- ? Discover and prove a new mathematical theorem?
- ? Converse successfully with another person for an hour?
- ? Perform a surgical operation?
- ✓ Put away the dishes and fold the laundry?
- ✓ Translate spoken Chinese into spoken English in real time?
- ✓ Write an intentionally funny story?



Generative Pretrained Transformer (GPT-3)

- Generating text (by OpenAI) – Trained on a fraction of internet! (cost: 12 M $)
- You provide the initial words or sentences, GPT-3 continues and completes the text!
- Aligned to be able to understand human instruction and values
  - Become chatable!
  - GPT-3.5 or ChatGPT
- GPT-4 is the next version which is trained on a much larger dataset

Albert: It's amazing. I'm so proud of being able to contribute to science and help humanity from here.
Manuel: What's your main occupation at the moment?
Albert: My main occupation is being a physics professor at Princeton University.
Manuel: So, how would you describe your contributions to science so far?
Albert: Well, I've already made some pretty important contributions to physics.
Manuel: *grabs coffee from table and takes a sip* Tell me more.
Albert: Well, first of all, my work with the photoelectric effect and quantum mechanics was very important for the development of quantum physics.
Manuel: What does quantum physics mean? Sorry, I'm no physicist myself.
Albert sits back in his comfortable chair, and after taking a slow sip of tea, explains.
Albert: You see, a hundred years ago, our understanding of the Universe, and how it all works, was still fairly immature. There were many loose ends, which we were starting to try and unravel. For example, things like what the atom is made of, or what light actually is.
Manuel: I see... Tell me more about light.
Albert: Yes. Well, that's where the photoelectric effect comes into play. When light hits an atom, it can eject an electron from that atom, producing a current, and therefore, power. This is an amazing process; light is knocking electrons off their atoms, and actually doing work.
Manuel: *sips coffee, and listens attentively*
Albert: *pause* Now, since we know that light is made of particles called photons, it is reasonable to assume that each photon has a certain amount of energy. And it also makes sense that the more energy a photon has, the more likely it is to eject an electron from an atom.
Manuel: I see. Did you discover this doing your own experiments?
Albert: No, I developed this theory based on the work of Max Planck and others before me. You see, Max Planck observed that there is a threshold to the amount of energy a single photon needs to have in order to eject an electron from an atom. This threshold is called Planck's

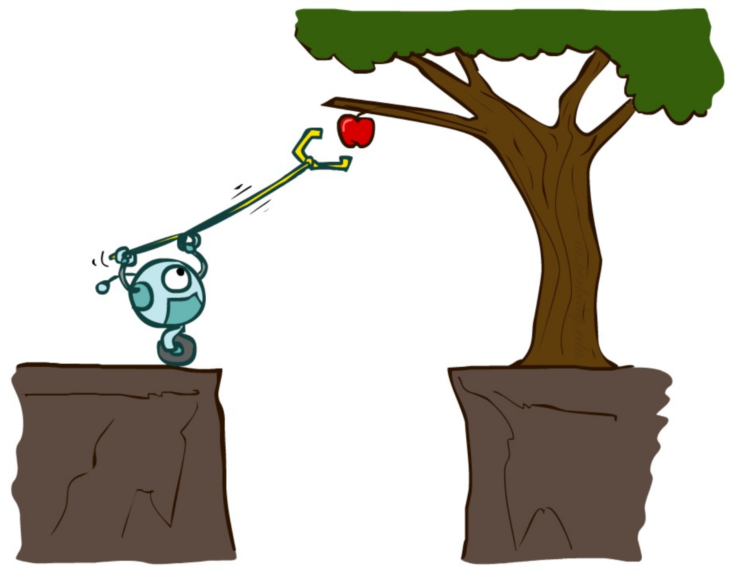
OpenAI GPT-3 generates faux interview with Einstein

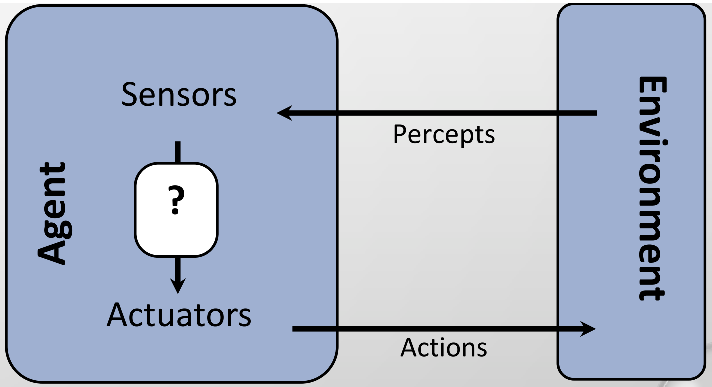
Rational Agent

- An agent is an entity that **perceives** and **acts**
- Abstractly, an agent is a function from percept histories to actions:

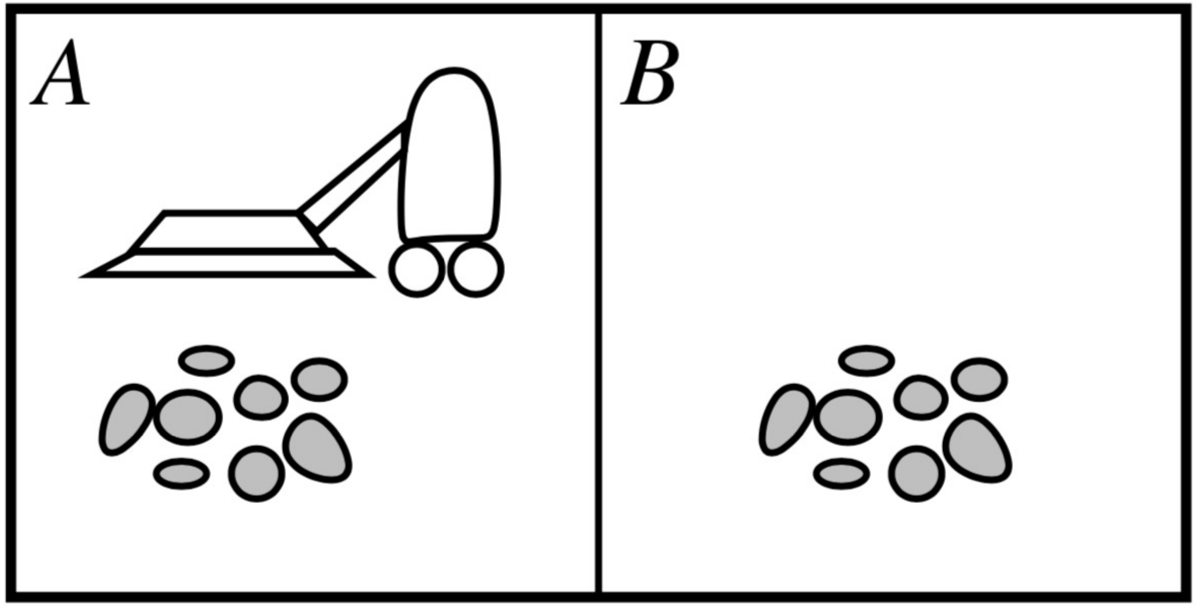
$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

- For any given class of environments and tasks, we seek the
  - agent (or class of agents) with the best (expected) performance (or utility)
- Caveat: computational limitations make perfect rationality unachievable
  - design best program for given machine resources





Vacuum-cleaner world



Percepts: location and contents, e.g., $[A, Dirty]$

Actions: $Left$, $Right$, $Suck$, $NoOp$

A vacuum-cleaner agent

Percept sequence	Action
$[A, Clean]$	$Right$
$[A, Dirty]$	$Suck$
$[B, Clean]$	$Left$
$[B, Dirty]$	$Suck$
$[A, Clean]$, $[A, Clean]$	$Right$
$[A, Clean]$, $[A, Dirty]$	$Suck$
⋮	⋮

function REFLEX-VACUUM-AGENT([*location*,*status*]) **returns** an action

if *status* = *Dirty* **then return** *Suck*
else if *location* = *A* **then return** *Right*
else if *location* = *B* **then return** *Left*

What is the **right** function?
Can it be implemented in a small agent program?

Rationality?

- Fixed performance measure evaluates the environment sequence
  - one point per square cleaned up in time T?
  - one point per clean square per time step, minus one per move?
  - penalize for > k dirty squares?
- A rational agent chooses whichever action maximizes the **expected value** of the performance measure given the **percept sequence to date**

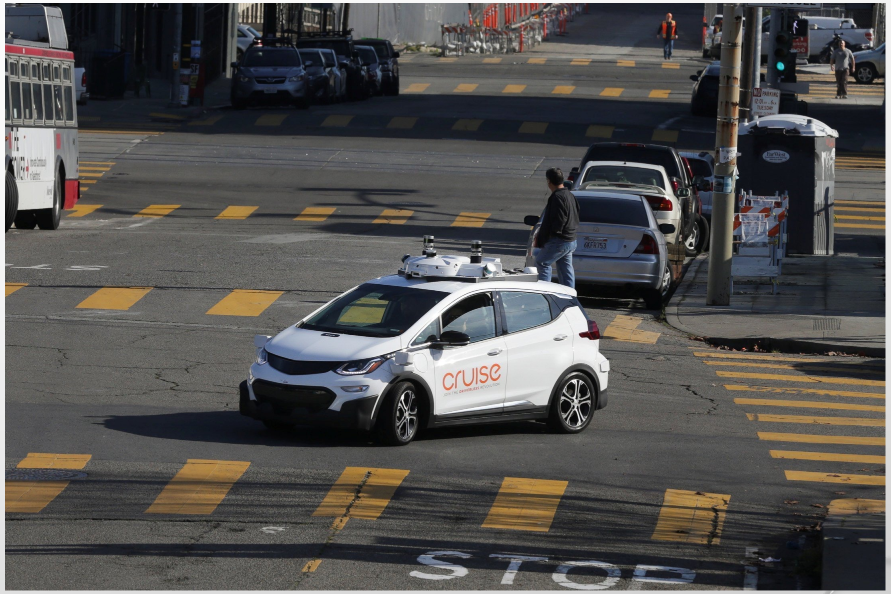
Rationality? (cont.)

- Rational ≠ omniscient
  - percepts may not supply all relevant information
- Rational ≠ clairvoyant
  - action outcomes may not be as expected
- Hence, rational ≠ successful

  Rational ⇒ exploration, learning, and autonomy

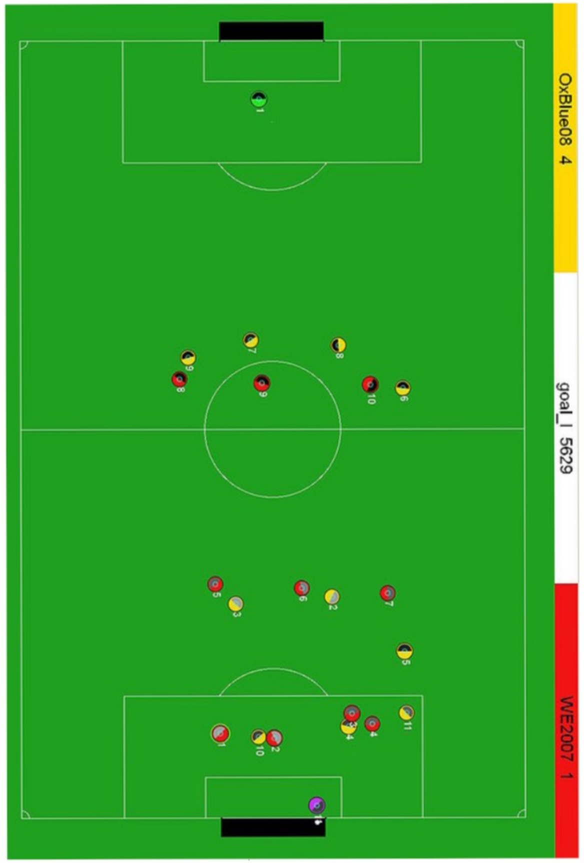
Modeling the world

- To design a rational agent, we must specify the **task environment**
  - Also known as PEAS (e.g. in automated taxi agent):
- Performance measure (sometimes with constraints)
- Environment
- Actuators
- Sensors

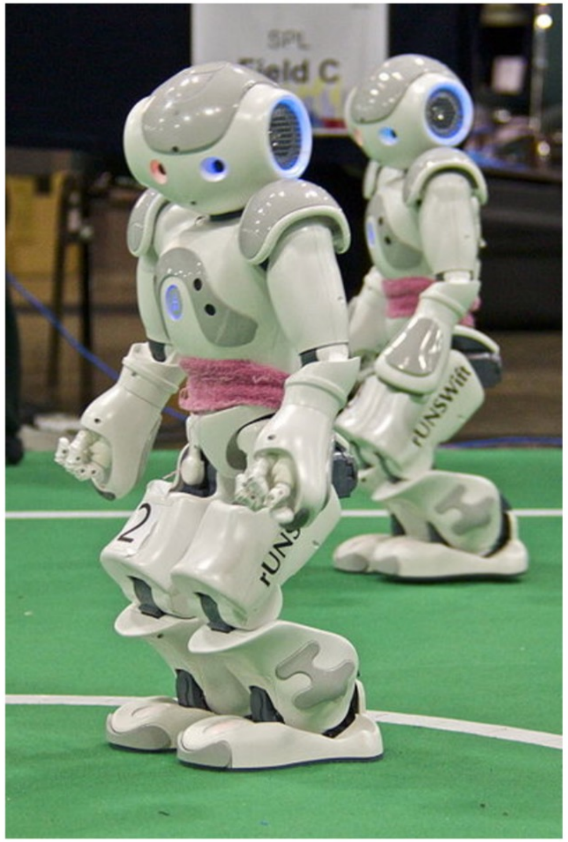


Types of environments

Fully observable vs. Partially observable



vs.



Single agent vs. Multiagent

vs.

Deterministic vs. Stochastic

vs.

Episodic vs. Sequential

vs.

Discrete vs. Continuous

vs.

Types of agents

Reflex agents:

- Choose action based on current percept (and maybe memory)
- Do not consider the future consequences of their actions
- Act on how the world **IS**

Types of agents (cont.)

Goal-based Agents:

- Plan ahead
- Ask “what if”
- Decisions based on (hypothesized) consequences of actions
- Uses a model of how the world evolves in response to actions
- Act on how the world **WOULD BE**

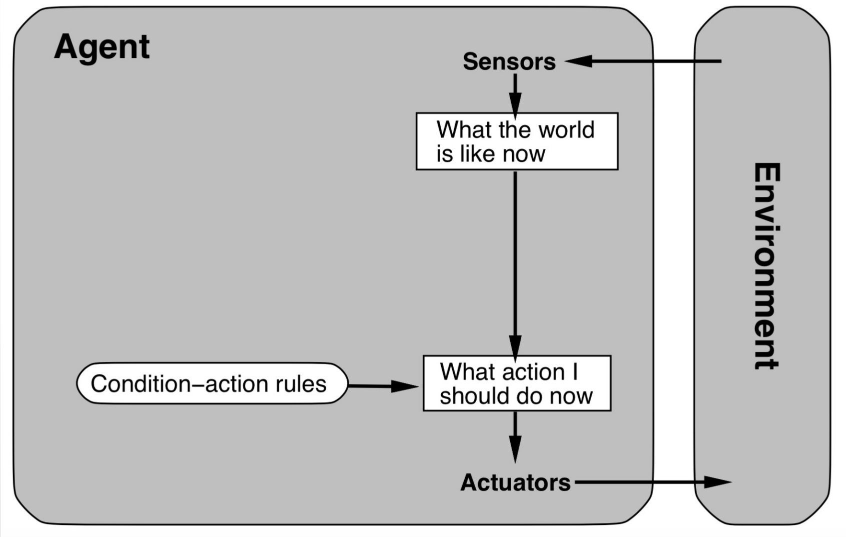
Types of agents (cont.)

Utility Based Agents:

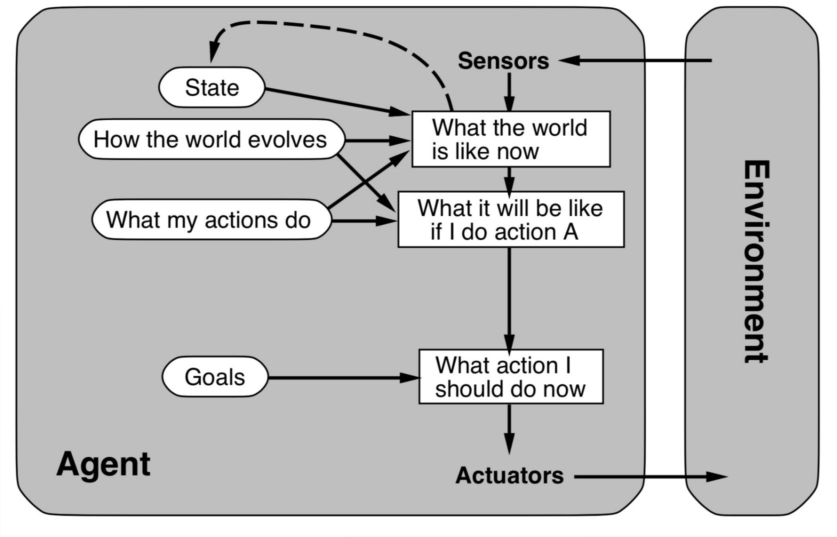
- Like goal-based, but
- Trade off multiple goals
- Reason about probabilities of outcomes
- Act on how the world will **LIKELY** be

Types of agents (cont.)

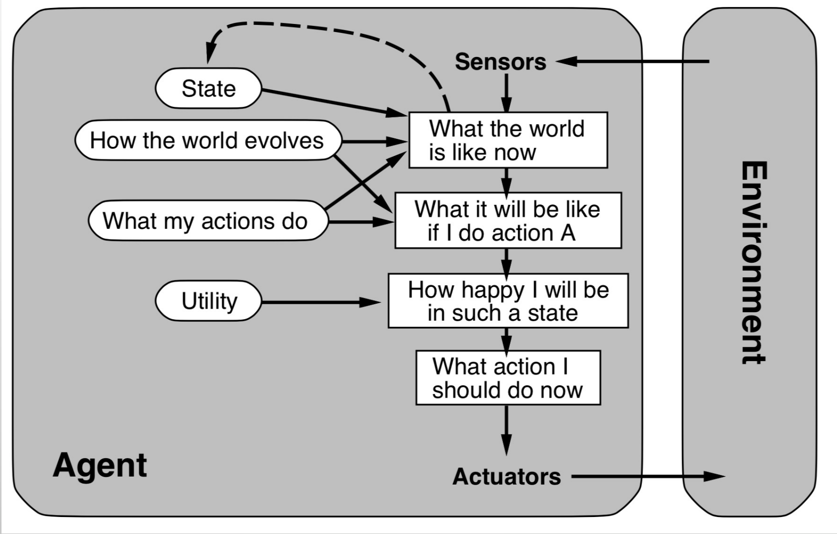
Reflex Agent



Goal-based Agent



Utility-based Agent



What will be discussed in this course?

- Various search techniques to optimize the utility (including adversarial, and online settings)
- Constraint satisfaction problems
- Inference under uncertainty and Bayes nets
- Temporal probability models (such as the Markov models)
- Introduction to Machine/Deep Learning
- Markov Decision Processes and Reinforcement Learning