Lecture 1

Introduction to AI

1.1 What is Artificial Intelligence?

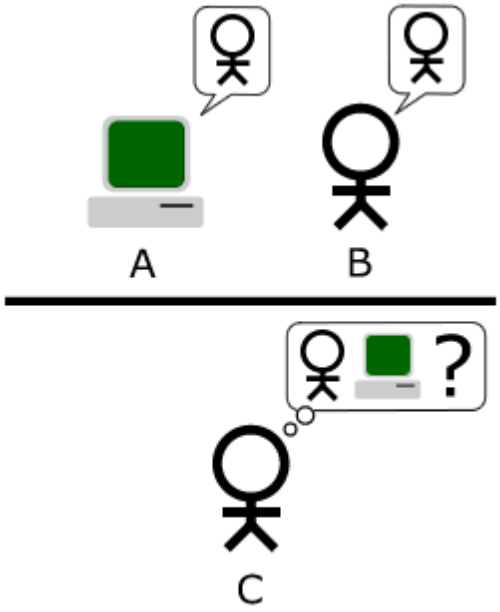
Artificial Intelligence (AI) is a field of study that focuses on understanding our **intelligence**, as well as building **intelligent entities**. AI applies ideas, techniques and algorithms from various related fields. (1)

- Philosophy; Logic
- Mathematics; Statistics; Neuroscience
- Computer Science and Engineering
- Linguistics

AI composes of several subfields, such as playing chess, chatter bots, proving mathematical theorems, music composition, etc. AI definition can be categorized into 4 categories:

	Thinking	*Acting*
Humanly	Thinking humanly	Acting humanly
Rationally	Thinking rationally	**Acting rationally**

Acting humanly: This approach was proposed by Alan Turing in 1950 based on the idea that an intelligent system should behave like a human, and it should be able to fool a human interrogator.



Source: http://commons.wikimedia.org/wiki/File%3ATuring_Test_version_3.png

Thinking humanly: This approach is focusing on how humans think. Human thought can be caught through (1) catching our own thoughts as they go by, (2) psychological experiments e.g. observing a human in action. Then, we can implement a program based on the obtained theories.

Cognitive science is a field of study linking AI and psychological experiment techniques to build theories of the workings of the human mind.

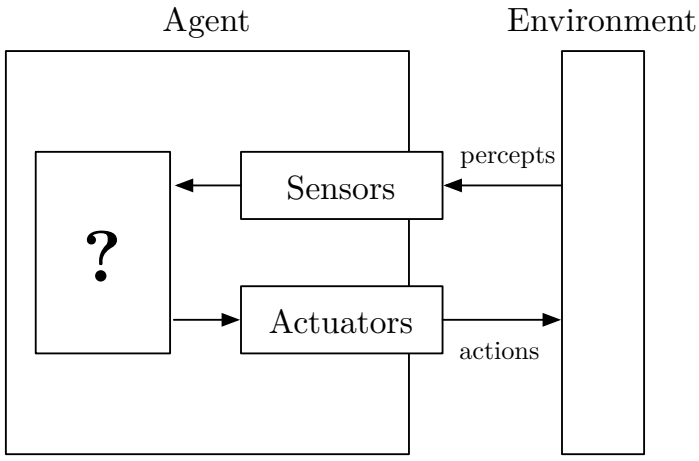
Thinking rationally: This approach aims to codify thinking. *Logic* is a field of study that focuses on the law of thought, or how correct conclusions can be obtained from facts.

$$\frac{\forall x\,(human(x) \Rightarrow mortal(x)) \qquad human(Socrates)}{mortal(Socrates)}$$

Acting rationally: This is the most recent AI approach. It focuses on building a **rational agent** which acts so as to accomplish the best expected outcome by using various techniques.

1.2 Agents

An **agent** is anything that possesses **sensors** to accept inputs from the **environment**, and responses to the environment through its **actuators**. Each agent has an **agent function** to control its bahavior. (2.1)



1.3 Rationality

(2.2.1)

> *For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has.*
>
> — *Russell and Norvig (AIMA p.37)*

In short, a **rational agent** is an agent that does the **right thing**. Each rational agent has a **performance measure** defining the criteria of success. It evaluates sequences of environment states. Rationality is not **omniscience**. A rational agent focuses only on its expected performance measures. It does not need to maximize the real performance.

$$\begin{aligned} \textit{Rational Agent} = (\textit{Performance Measure}) + (\textit{Environment}) \\ + (\textit{Actuators}) + (\textit{Sensors}) \end{aligned}$$

For example, an *automated taxi* do not need to drive (badly) like a real taxi driver, but it should maximize its performance i.e. safety, following traffic rules, profit, and customer convenience, etc.

This automated taxi does not need to be omniscience. It may not be able to avoid some objects falling from the sky.