Artificial Intelligence

Lecture 1

> Seek to be an indispensable complement to something that is getting cheap and plentiful.
>
> *Google chief economist Hal Varian*

IN4010 Artificial Intelligence Techniques

Prof. dr. C.M. Jonker, dr. K.V. Hindriks, dr. M.B. van Riemsdijk

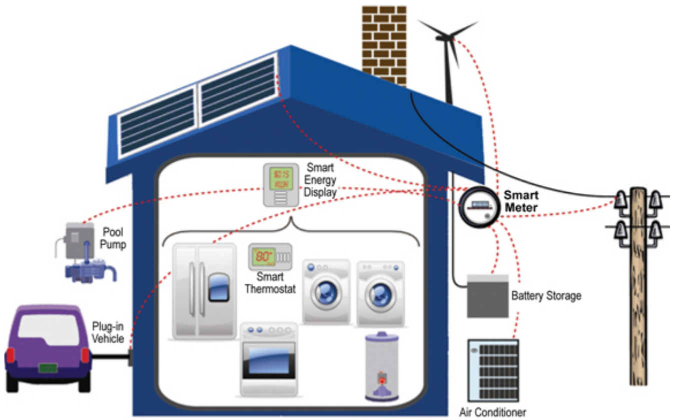
13 September 2017

Interactive Intelligence group ,
Department of Electrical Engineering, Mathematics and Computer Science

Delft University of Technology

Delegating Decision Making to Machines

McKinsey: **Uptake of AI**: by 2025, machines will be able to learn, adjust, exercise judgment, & reprogram themselves



TU Delft

Introducing the Teachers

- Catholijn Jonker

- Koen Hindriks

- Birna van Riemsdijk

Overview

- **Artificial Intelligence definitions**
- **Course outline**
- **What we don’t cover: Machine Learning**

Course Material for this lecture

- Slides

TU Delft

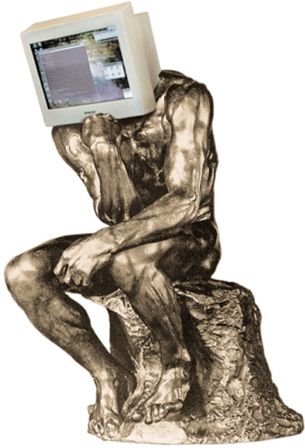
Learning Objectives

In this lecture you will learn to:

- **Discuss various goals** of Artificial Intelligence
- **Design & apply strategies for decision making**
- **Explain** what **cognitive agents** are

ARTIFICIAL INTELLIGENCE

What is Artificial Intelligence (AI)



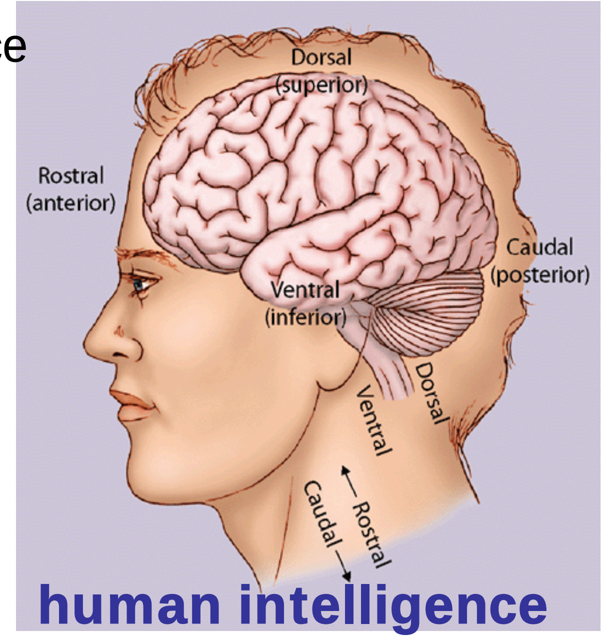
Many definitions have been proposed for Artificial Intelligence

Df 1: Machine ≥ human

- The science that tries to automate processes that humans so far do better than machines

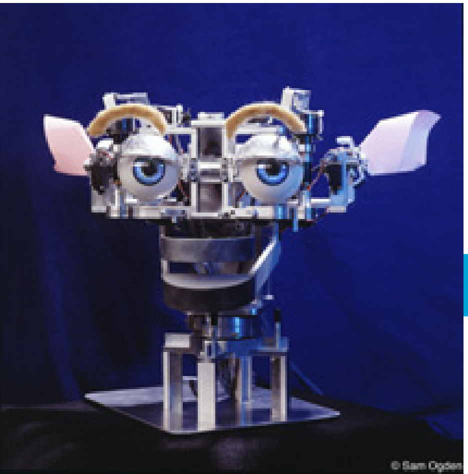
Df 2: Understand and simulate

- The science that aims to understand natural intelligence so well that it can be simulated on a computer.



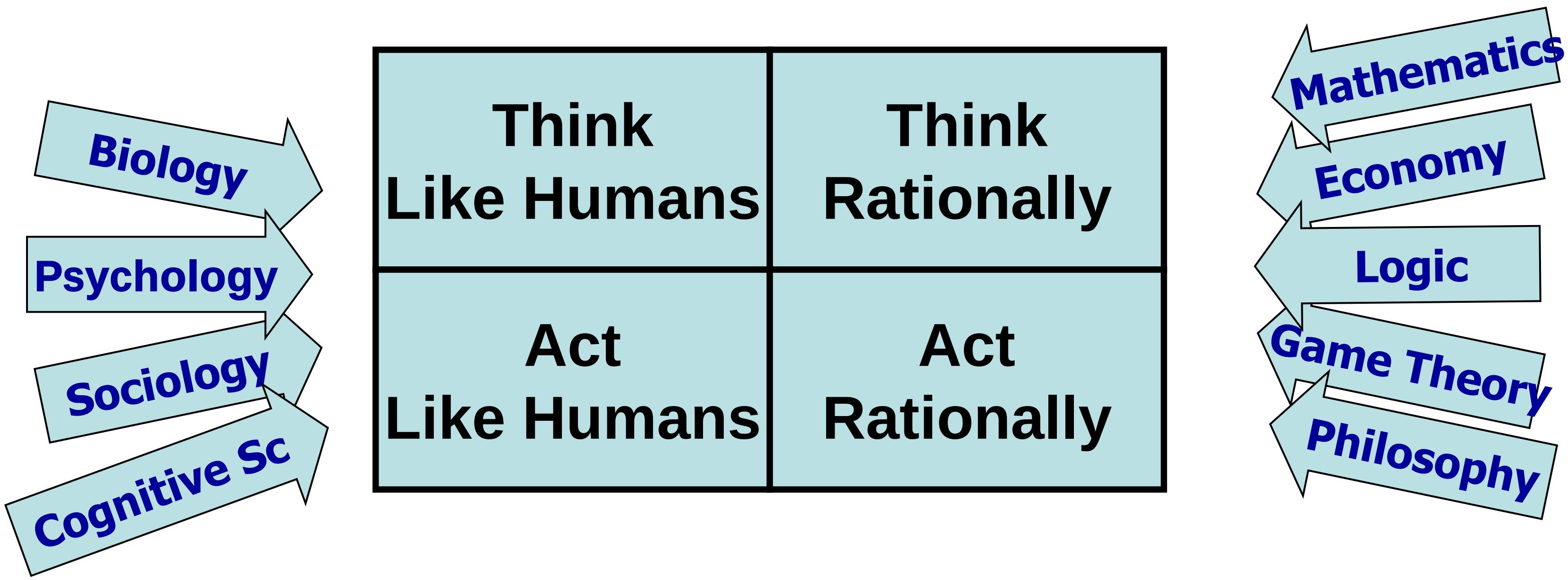
Df 3: Human and computer

- The science of developing intelligent software that supports humans.



What sort of agent system?

What kind of agent system is AI trying to engineer?



TUDelft

Acting humanly: Turing test

Turing (1950) "Computing machinery and intelligence":

- "Can machines think?" → "Can machine behave intelligently?"
- Operational test for intelligent behavior: Imitation Game
- Suggested major components of AI: knowledge, reasoning, language understanding, learning.

Pro: Simple test, empiricist approach.

Con: Not reproducible, constructive, or amenable to mathematical analysis. It does not provide any guidelines for building agents!

Thinking humanly: Cognitive science

- 1960s "cognitive revolution": **information-processing psychology** replace prevailing orthodoxy of behaviorism.
- Requires scientific theories of internal activities of the brain
  - What level of abstraction? "knowledge" or "circuits"?
  - How to **validate**? Requires
    - Predicting and testing behavior of human subjects (top-down)
    - Direct identification from neurological data (bottom-up)
- Both approaches (roughly cognitive science and cognitive neuroscience) are now distinct from AI

Pro: all three fields share one principal direction, share goals; CogSci provides potentially useful theories for AI engineers!

Con: The hard road, and takes into account human specifics such as memory limitations which are(?) irrelevant for machine intelligence

Thinking rationally: Laws of Thought

- Normative (or prescriptive) rather than descriptive **logic**

- Goes back to Aristotle: what are correct arguments / thought processes?

Pro: A huge body of relevant work is available, connecting AI to many other fields such as mathematics and philosophy. Theoretically well-founded.

Con: Not all intelligent behavior is mediated by logical deliberation. What is the purpose of thinking? What thoughts should I have out of all the thoughts (logical or otherwise) that I could have?

Acting rationally

- Rational behavior: doing the right thing
- The right thing: that which is **expected to maximize goal achievement**, given the available information.

Pro: A huge body of relevant work is available, connecting AI to many other fields such as economics and philosophy. Theoretically well-founded.

Con: Much of the theory is quantitative, and it is hard to obtain the “numbers”, or to relate the theory to human performance. But it has been successful in areas humans do not handle well...

Human-like or rational?

Suppose we have an agent that has one of the capabilities below. Should this rather be a human-like or a rational agent?

- Speech recognition
- Negotiating
- Playing a game
    - Educational game
    - Chess
    - Real-time strategy game
- Computer vision
- Planning
- Recommender system

TU Delft

The Ultimate Goal of AI (1/3)

One more definition: Artificial Intelligence and its ultimate goal

The ultimate goal of AI, which we are very far from achieving, is to build a person, or, more humbly, an animal.

Charniak & McDermott 1985

We need an account of **personhood**:

1. To "will", make plans and decisions
2. To have subjective consciousness
3. To be *self*-consciousness
4. To communicate through language
5. To know and believe things, believe things about what others believe, etc
6. To desire changes
7. To reason

This account is in line with capabilities identified in CogSci

The Ultimate Goal of AI (2/3)

We can derive a list of research areas from capabilities of a person

Among others, we need an agent that is able to:

- To **maintain knowledge, beliefs, and desires**
- To **reason** with its knowledge, beliefs, and desires
- **Decide on what to do next**
- **Plan**, taking the future into account (lookahead)
- **Use language to communicate** *meaningfully*

Logic-based AI has made significant contributions to each of these items, and logic in a broad sense seems required to achieve these abilities.

See: Selmer Bringsjord, 2009, The Logicist Manifesto

TU Delft

The Ultimate Goal of AI (3/3)

Q: Why did we not make any progress on providing agents consciousness?

A: We do not have a good theory to help us engineer it!

A similar claim could be made with regards to **learning**, but useful techniques for building learning agents exist.

A classic commonsense problem: Cracking an egg into a bowl

what knowledge do we need to derive plan is sound?

- that the egg is brought into contact with the bowl's edge (and not elsewhere); that this action is done neither too fast nor too slow;
- that the bowl is neither upside down, smaller than the egg, nor made of (say) paper;
- that the egg itself is not hardboiled;
- that an object cannot be two places at once—the contents of the egg cannot be both in the egg and in the bowl;
- that an inanimate object will not move unless the force applied to it is sufficient to counteract any other force;
- gravity will cause the egg contents to fall into the bowl only if the opening made by the crack is facing downwards;
- that the act of cracking an egg does not change the structure of the bowl but does change the structure of the eggshell; ...

Ordering a meal at a restaurant

what knowledge do we need to derive plan is sound?

- Being able to read menu
- Where to find menu
- How much money
- Recognizing a waiter
- Knowledge allergies
- Food preferences
- Content foods
- How hungry you are
- How many persons

TUDelft

Common-sense Challenge

- Need: Domain specific knowledge and a general understanding of resources, time, and physics.

- But when do we stop writing down all specifications of all conditions that we know to be relevant to a problem?

- And what inference engine do we need?

A “Weak” Notion of Agency

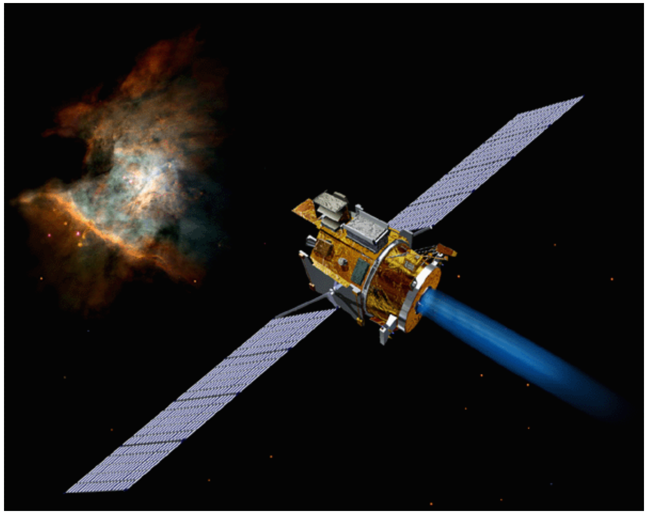
Agents are:

- **autonomous** agents control their own processes
- **reactive** able to receive and respond to information
- **pro-active** able to take initiative to further their goals
- **social** able to communicate and cooperate

Wooldridge & Jennings take a **definitional approach** to characterize agents.

Autonomy in Space

When a space probe makes its long flight from Earth to the outer planets, a ground crew is usually required to continually track its progress, and decide how to deal with unexpected eventualities. This is costly and, if decisions are required *quickly*, it is simply not practicable. For these reasons, organizations like NASA are seriously investigating the possibility of making probes more autonomous — giving them richer decision making capabilities and responsibilities.



"Deep Space 1 launched from Cape Canaveral on October 24, 1998. During a highly successful primary mission, it tested 12 advanced, high-risk technologies in space. In an extremely successful extended mission, it encountered comet Borrelly and returned the best images and other science data ever from a comet. During its fully successful hyperextended mission, it conducted further technology tests. The spacecraft was retired on December 18, 2001." – NASA Web site

http://nmp.jpl.nasa.gov/ds1/

TU Delft

Agents are Reactive

A **reactive** system is one that maintains an ongoing interaction with its environment, and responds to perceived changes that occur in it, in a timely fashion, so that the response is appropriate.

•Reactiveness not relevant in a static environment, e.g. think of compiler.

•Many (most?) interesting environments are dynamic*.* In the real world things change, and information is incomplete.

•Much harder to build agents for dynamic domains: must take into account e.g. possibility of failure – ask itself whether it is worthwhile to perform an action while be responsive to changes!

•Compare Russell & Norvig's Simple Reflex Agent. *Are these reactive?*

Agents are Proactive

A **proactive** system is one that decides on its next action taking its current goals into account, actively monitoring and considering the opportunities for achieving these goals.

- Proactive agents demonstrate **goal directed behavior**, which are ideally in line with our design objectives ("do useful things for us")
- Such agents need to:
  - **generate** and (attempt to) **achieve** goals,
  - should not be driven solely by *current* events, and
  - be able to **recognize opportunities** for furthering their goals.

- Compare Russell & Norvig's Goal-Based Agent. *Are these proactive?*

Combining
Reactive and Proactive Abilities

Combining these properties gives rise to interesting problems.

- Reactive and proactive behavior can be at odds with one another and lead to **potential conflict in choice of action**

- Simple example:

  You can either answer a phone call or finish your diner before you have to leave for a meeting.

  Problem: avoid something bad, achieve something good of more or less equal value, in short window of opportunity. People make different choices, how should agents choose?

- Designing an agent that can balance the two remains an open research problem.

Social Ability

A **social** system is a system that is able to communicate and coordinate with other agents in order to share the knowledge it has and to achieve goals where agents depend on each other.

- Many environments consist of cooperating or possibly competing agents (compare strategic environments).
- In such environments, an agent cannot or should not attempt to achieve its goals without taking those of others into account.
- The ability to coordinate and communicate assumes:
  - the availability of an agent-communication language.
  - the ability to recognize other agent's intentions ("mind reading")

For the interested reader:

➔ Relevant theories: Speech act theory (Linguistics), Theory of Mind (Psychology)

Outline

Concluding:

Whatever else, AI is an **engineering** discipline:

- It aims at developing techniques to engineer systems.
- But it needs **theories** to guide this development that allow us engineers to construct the **toolbox** we need!

Course setup:

- **First period** of this course: focus on decision-theoretic AI
- **Second period** of this course: focus on logic-based AI

This lecture:

- explore various attempts to make the notion of an **agent** more concrete, and identify some areas that need to be studied.

TU Delft

Course Overview – Details

First Quarter

Artificial Intelligence

Negotiation

Automated Negotiation

Game Playing

Rational Decisions

Complex Decisions

Game Theory

Second Quarter

Artificial Intelligence & Cognitive Agents

Automated Reasoning & KR

Logic Programming & Mental States

Actions & Planning & MAS

Reinforcement Learning in Agents

Coordination & Serious Gaming

What is not covered in this course

Computational Intelligence - B.Sc. Q1 - 150 students

- general intro to learning
- classification, regression, search
- MLP, backprop, perceptron, bit of deep learning
- competitive, Hebbian, associative learning, SOM
- Bayesian networks
- genetic algorithms, evolutionary strategies
- reinforcement learning
- swarms, particle swarm optimization

Not covered ...

Neural Networks [IN4015] - M.Sc. Q3+4 - 60/70 stud.

- Q3 lectures
- - perceptron, backprop
- - competitive learning, bit of associative
- - deep learning, CNNs, autoencoders, RBMs
- - overfitting, why and what
- - RNNs, extreme learning machines
- - genetic algorithms, evolutionary strategies
- - reinforcement learning

Not covered ...

- **Neural Networks Q4 seminar MSc level**
- - applications of NNs
- - applications to computer vision and robotics
- - learning from imitation
- - affective computing, emotion recognition
- - neuro-evolution, genetic algorithm training of NNs
- - swarm computing, advanced topic

Not covered ...

Machine Learning [IN4320] - M.Sc. Q3+4, 25/50 stud.

- - regularization and sparsity
- - learning theory, PAC learning, AdaBoost
- - stochastic gradient descent, CNNs
- - semi-supervised learning
- - multiple-instance learning
- - reinforcement learning

TU Delft

Not covered …

- Computer vision
- Data mining
- Image processing
- Pattern recognition
- Robotics
- Human-agent-robot teamwork
- Social Signal Processing

TU Delft

Organizational (check Brightspace)

Reading

Russell and Norvig and additional material as indicated on Brightspace

Tests

On Thursdays (voluntary basis, check it with your fellow students)

Tutorials

Tutorial on various topics to help you prepare for tests

Practical Assignment

Register by the end of the week via email (pa.ai@ii.tudelft.nl).

Organizational – this week

Reading:

- Slides

Prepare for Tutorial

- Tutorial on GENIUS (platform for automated negotiation)

Practical Assignment: **Register** by the end of the week via email (pa.ai-EWI@lists.tudelft.nl). Download and read the first assignment before tomorrow's meeting

Next Tutorial (tomorrow): bring your **laptop**

Practical assignment

- Form groups of 4 people
  - right now & here, or at the tutorial,
  - send an e-mail to pa.ai-EWI@lists.tudelft.nl with your names and contact info (one e-mail per group)

- Assignment
  - instructions will be published on Brightspace
  - at your own time and place