CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

INTRODUCTION TO DATA SCIENCE AND AI

DAT405/DIT405, STUDY PERIOD 3, 2021

First of all…

- Very welcome to the course!
- This is a 100% online course that uses the video conference system Zoom and the learning platform Canvas
- We have a record number of students on the course: about 190!
- Presentation

Recordings

All lectures will be recorded as a service to you, the students of the course. The recordings and the lecture slides will be uploaded to the course's Canvas page.

You can choose if you want to have your video and mic turned on or off

You can choose to watch the lectures live (recommended) or offline afterwards

Lecture 1

Introduction to AI

Today

- Info about the course
- Mathematical modeling
- Natural intelligence
- Artificial intelligence
- Applications of rule-based AI

Info about the course

Students from Chalmers

About 171 students from Chalmers

Bachelor's programs

- TKITE 33 — IT
- TKAUT 19 — Automation
- TKDAT 15 — CS
- TKTEM 4 — Mathematics
- …

Master's programs

- MPSOF 25 — Software Engineering
- MPHPC 11 — High Performance Comp.
- MPALG 9 — Algorithms
- MPMEI 8 — Innovation
- exchange 8
- MPCAS 7 — Physics
- MPDES 5 — Design
- …

Students from GU

- 19 from GU?
- Then there are about 190 students on the course in total
- New record for the course!

Student representatives

- Claudio Aguilar Aguilar (laagu@student.chalmers.se), MPSOF
- Felicia Ekener (felicia.ekener@gmail.com), TKAUT
- Frida Grothérus (fridagr@student.chalmers.se), MPDES
- Alexander Jyborn (alexander.jyborn@gmail.com), TKITE
- Marcel Vacante (vacante@student.chalmers.se), MPCSN

Teaching team

13 teachers

- Claes Strannegård (Examiner, Teacher)
- Simon Olsson (Teacher)
- Emilio Jorge (Administrator)

- Divya Grover (TA)
- Arman Rahbar (TA)
- David Bosch (TA)
- Anton Johansson (TA)

- Denitsa Saynova (TA)
- Azadeh Karimisefat (TA)
- Erik Gunnarsson (TA)
- Kaver Hui (TA)
- Panagiotis Moraitis (TA)
- Adnan Fazlinovic (TA)

Formal info

- Name: Introduction to Data Science and AI
- Codes: DAT405/DIT405
- Credits: 7,5
- Duration: 8 weeks (18 Jan – 21 March 2021: Chalmers' Study Period 3)
- Aim: *The course gives a broad introduction to various techniques and theories used in Data Science and AI, with particular focus on their practical applications.*
- *Learning outcomes*: Read about it on Canvas

Learning platform

- We will be using the learning platform Canvas for almost everything
- Here is the homepage of the course on Canvas
- The Home and Syllabus pages are the most important.

Lectures

- All lectures will be given live on Zoom
- Slides and recordings will be uploaded to Canvas after the lectures
- They are for your personal use only. Do NOT distribute!

Because of privacy and copyright laws

- You can ask questions during the lectures using the chat function and I will try to answer
- Technical problems can happen. If we lose connection, we reconnect

Assignments

More details about assignments on Canvas

Not 1, not 3, not 4…

- All assignments are done in pairs (normally same pair all the time)! If you need to make changes, please contact Emilio.

- Once you have found a partner you should both join a pre-existing group on Canvas that is free. Post on the Canvas Discussion forum if you are looking for a partner.

- There will be two organized lab sessions for working on the assignments each week. Then you will be able to get help from a TA.

Modules

There are eight modules: one per week

1. Introduction
2. Regression and classification
3. Clustering
4. Bayesian models
5. Markov models
6. Ethics
7. Neural networks
8. Rule-based AI

Lectures

Each module has two lectures, except for the last, which has only one.

1. Intro to AI
2. Intro to Data Science
3. Regression
4. Classification 1
5. Clustering 1
6. Clustering 2
7. Bayesian statistics
8. Bayesian models
9. Markov models
10. Classification 2
11. Ethics 1
12. Ethics 2
13. Neural networks 1
14. Neural networks 2
15. Rule-based AI

Assignments

Each module has one assignment

1. Data sets
2. Regression and Classification
3. Clustering
4. Naïve Bayes
5. Reinforcement learning
6. Ethics
7. Neural networks
8. Search

Grading

- Chalmers grades
  - Fail
  - 3
  - 4
  - 5
- GU grades
  - U (Fail)
  - G (Pass)
  - VG (Pass with Distiction)
- Assignment grades
  - Each Assignment will be graded with 1-10 points
  - Special rules apply for grades on late submissions
- Course grade
  - At least 5 points on all Assignments needed for a passing grade
  - The course grade will be an aggregation of the assignment grades

More details about grades on Canvas!

Plagiarism

- Plagiarism (Video: stop at 3.00)
  - Education and research are all about using work by others
  - Highly recommended to use work by others but cite when you do
  - Applies to everything: text, images, music, video, code, etc.

- How to avoid plagiarism (Blog post)
  - Cite your sources
  - Include quotations
  - Paraphrase
  - Present your own idea
  - Use a plagiarism checker (to avoid involuntary plagiarism)

- Examples of plagiarism

Plagiarism at Chalmers and GU

Unfortunately, we have had problems with plagiarism on this course before.

- It will be discovered
  - All submissions are automatically scanned by a plagiarism detector. Compares with:
    - Documents on the Internet
    - Other submissions to this course
    - Submissions to other courses
- It will be reported
  - Chalmers and GU are very strict. Teachers are obliged to report.
- It will have consequences
  - The consequences for students involved in plagiarism at Chalmers or GU are very serious (often suspension)

Contact

Contact the teachers or administrator directly only in exceptional cases!

If you want to contact the teaching team, please follow these instructions:

- Questions about the course
  - Look for answers on the Canvas pages
  - Use the Canvas Discussion forum
- Questions about the lectures
  - During lectures: Use the chat room in Zoom
  - Other times: Use the Canvas Discussion forum
- Questions about the assignments
  - During lab sessions: Use the Waglys system to contact a TA
  - Other times: Use the Canvas Discussion forum
  - If you can’t post it, contact the responsible TA directly. Check the Assignments page to see who is responsible.
- Need to change groups?
  - Contact Emilio Jorge

Acknowledgements

- Videos, images and slides come from many different sources

- Thanks to Dag Wedelin, Graham Kemp, Marina Axelsson-Fisk, Fredrik Johansson, Simon Olsson, Alex Berman, and Peter Ljunglöf for letting me use their slides. Thanks also to Emilio Jorge, Devdatt Dubhashi, and Niklas Engsner.

Let’s get started!

Mathematical modeling

What is a model?

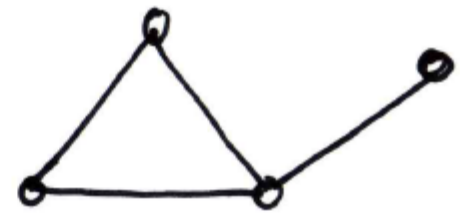
- A description of some kind that enables us to understand and predict some phenomenon

- Examples
  - A 3D model of a house (that exists or not)
  - A 2D sketch of a person
  - A text describing a dog
  - An equation describing a law of nature

- A *mathematical model* is a description that is formulated in mathematical/computational language

Different kinds of models

$$F = G \frac{m_1 m_2}{r^2}$$

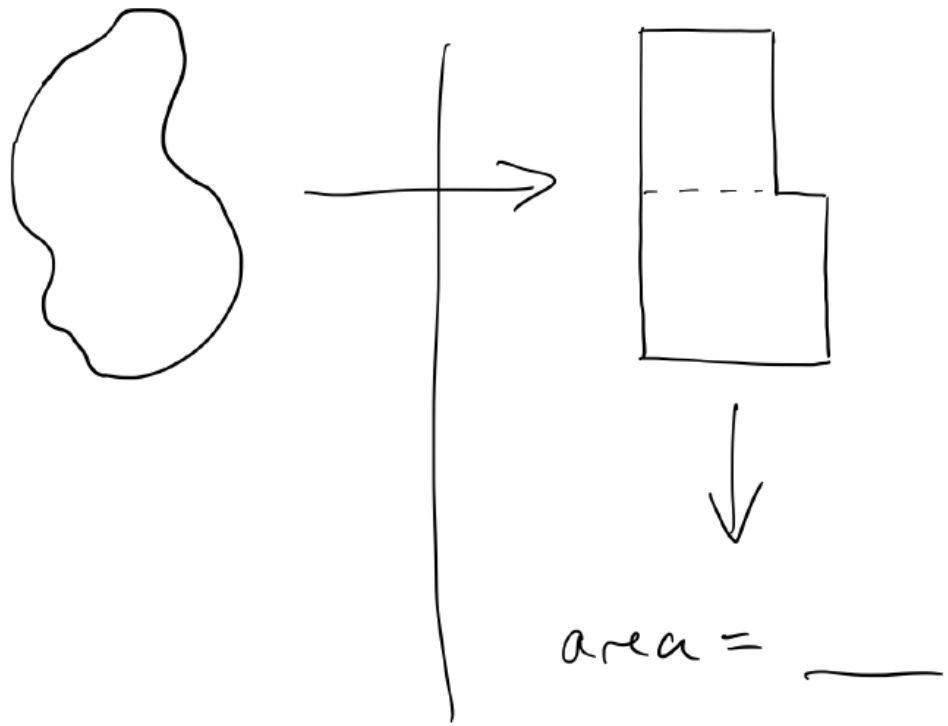


man(Socrates)

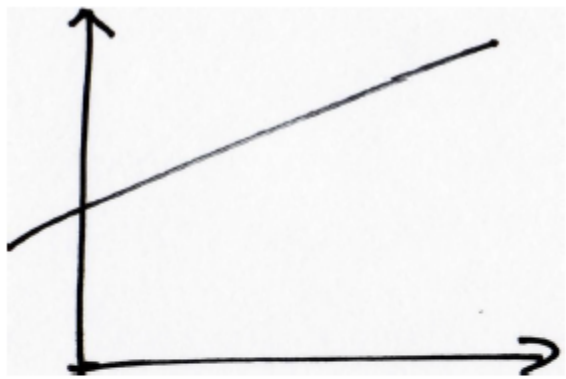
man(X) ⇒ mortal(X)

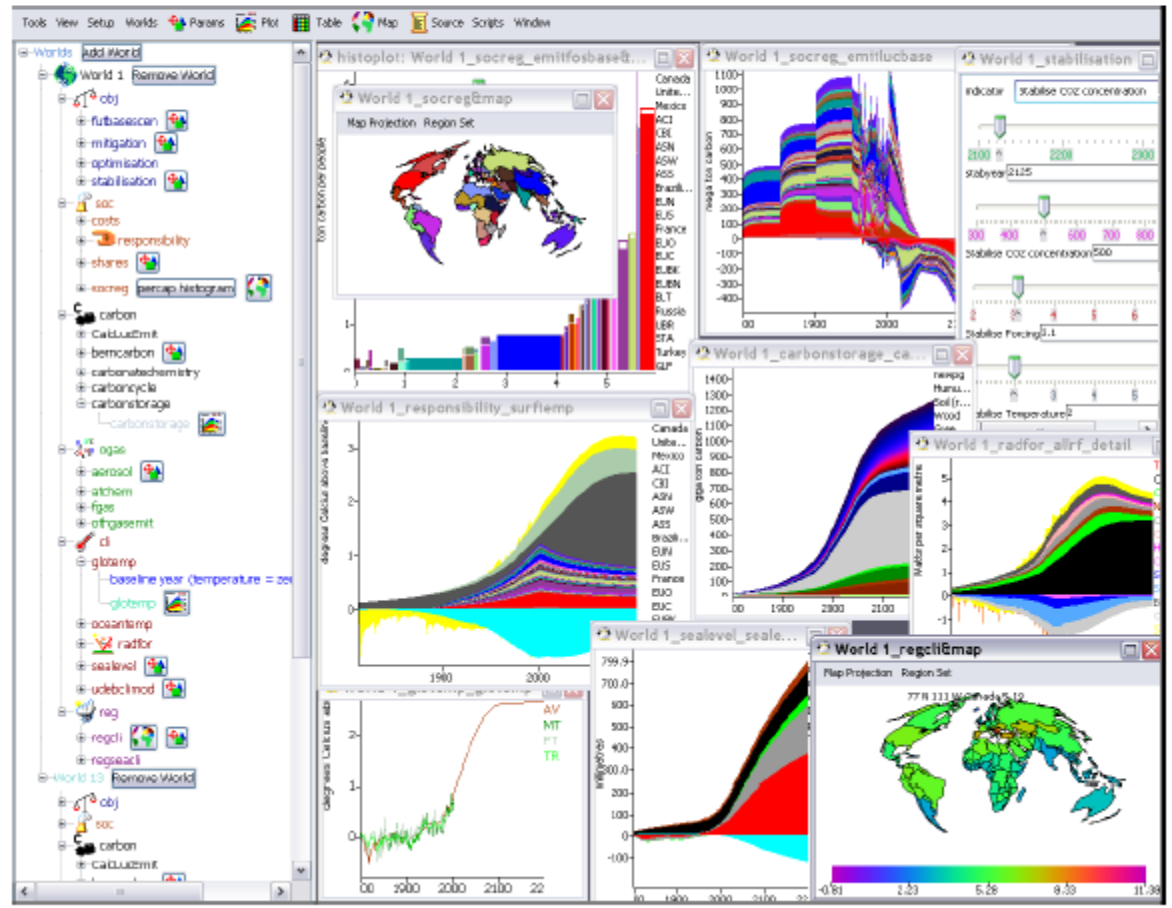
Why models?

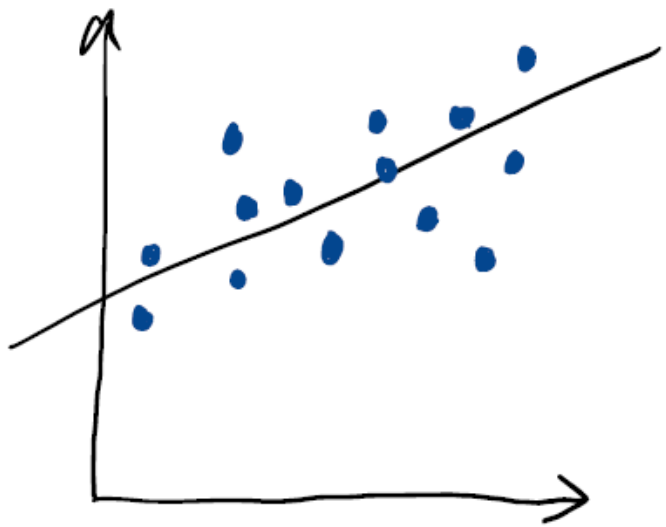
"a convenient way to represent reality so that we more easily can draw conclusions about it"



Simple and complicated models





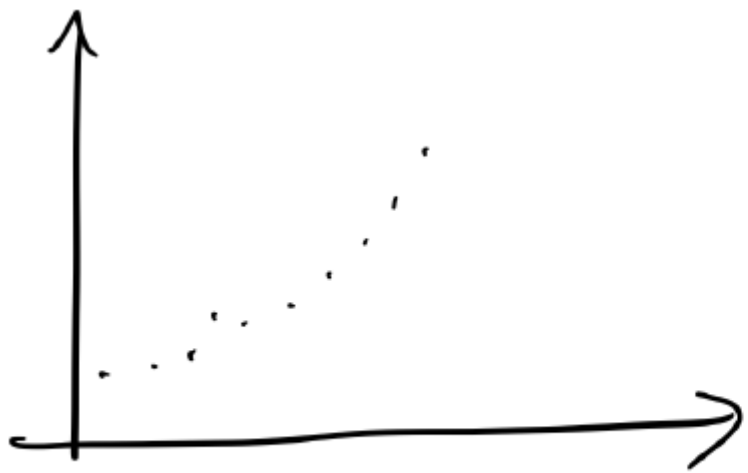


in this case we can hardly expect a perfect fit!

a model is necessary for generalization!

Natural to combine human insight and data

$$p = c \cdot e^{at}$$



Correlation does not imply causation

Interpret the data carefully!

Alzheimer's may be relatively common in Florida, but moving there does not increase your risk of getting it!

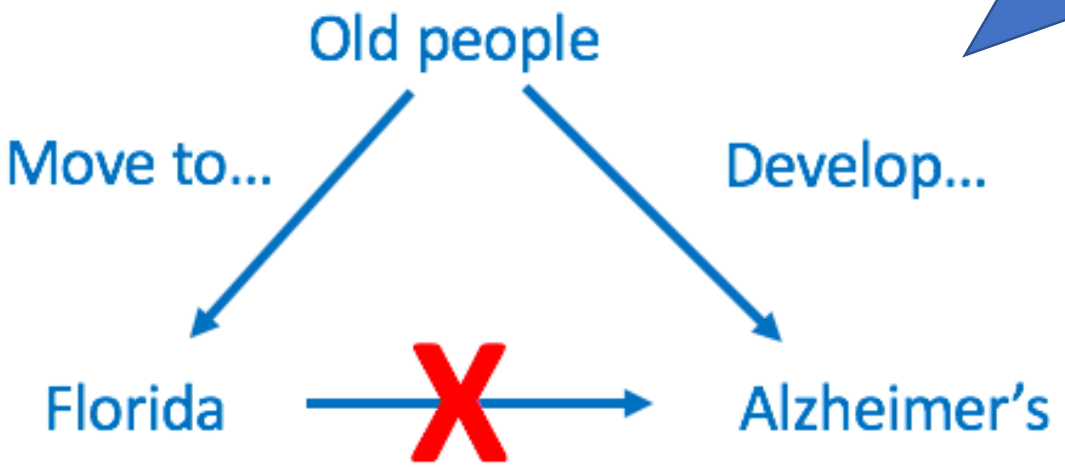
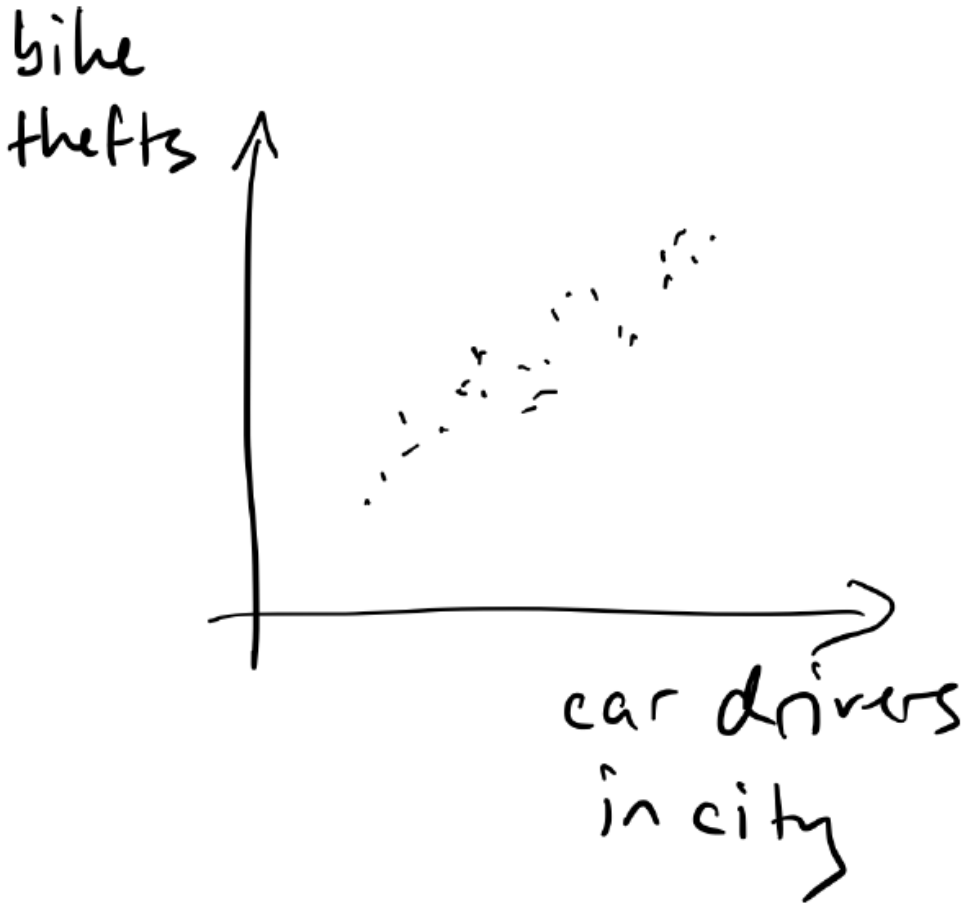


Image from www.acsh.org

Correlation does not imply causation

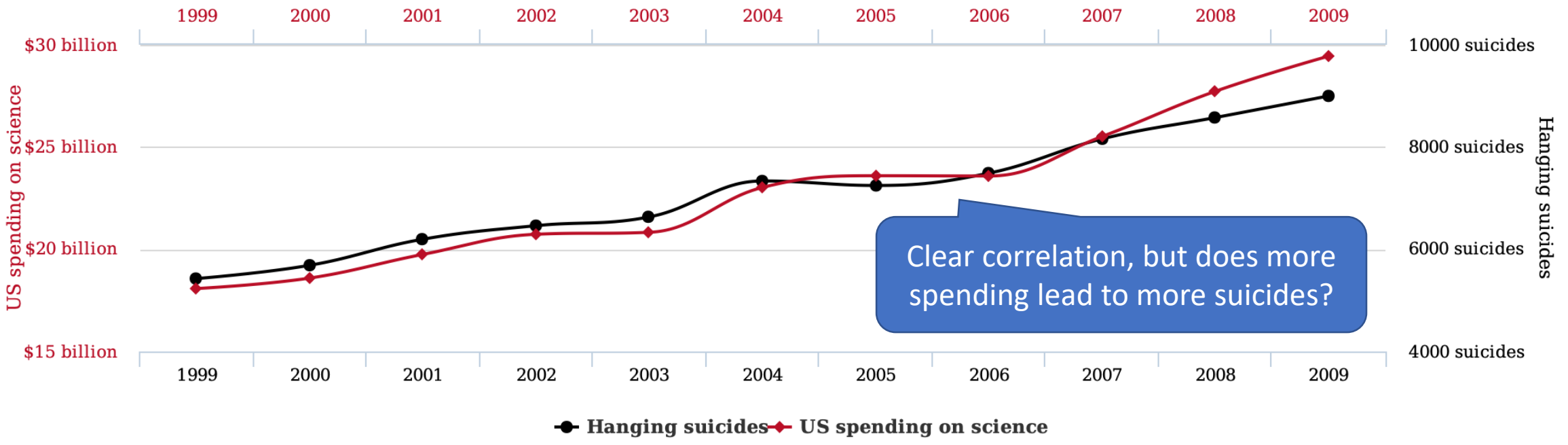


Proper reasoning is cruical!

US spending on science, space, and technology

correlates with

Suicides by hanging, strangulation and suffocation



tylervigen.com

Fundamental rules in all problem solving

understand the problem!

try simple approaches first!

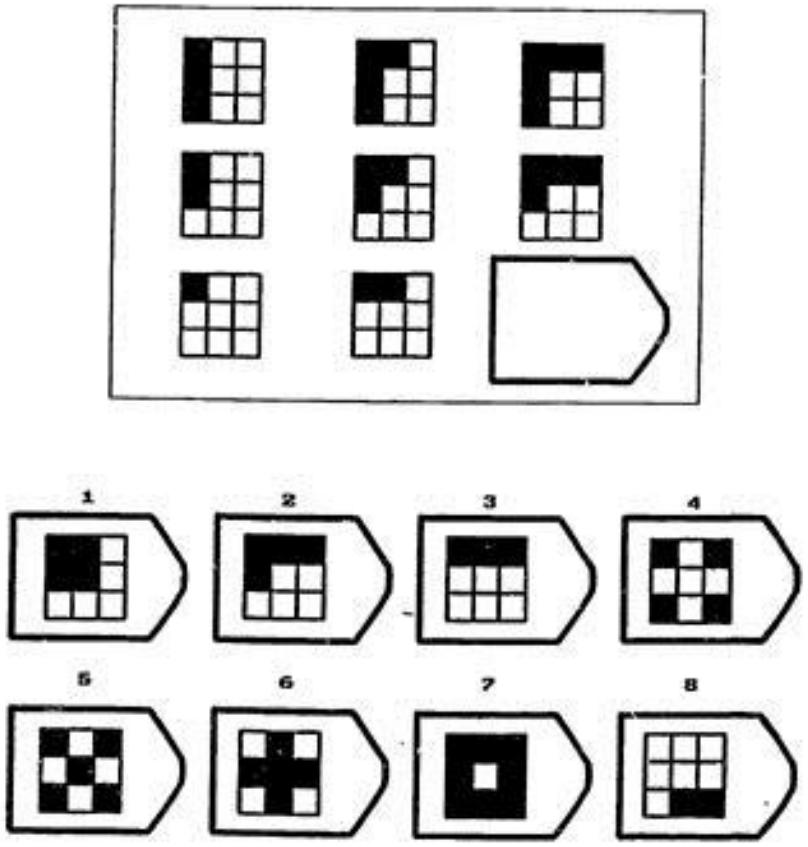
search broadly!

these rules are often broken in the naive application of data science/AI methods

Natural intelligence

What we want to imitate

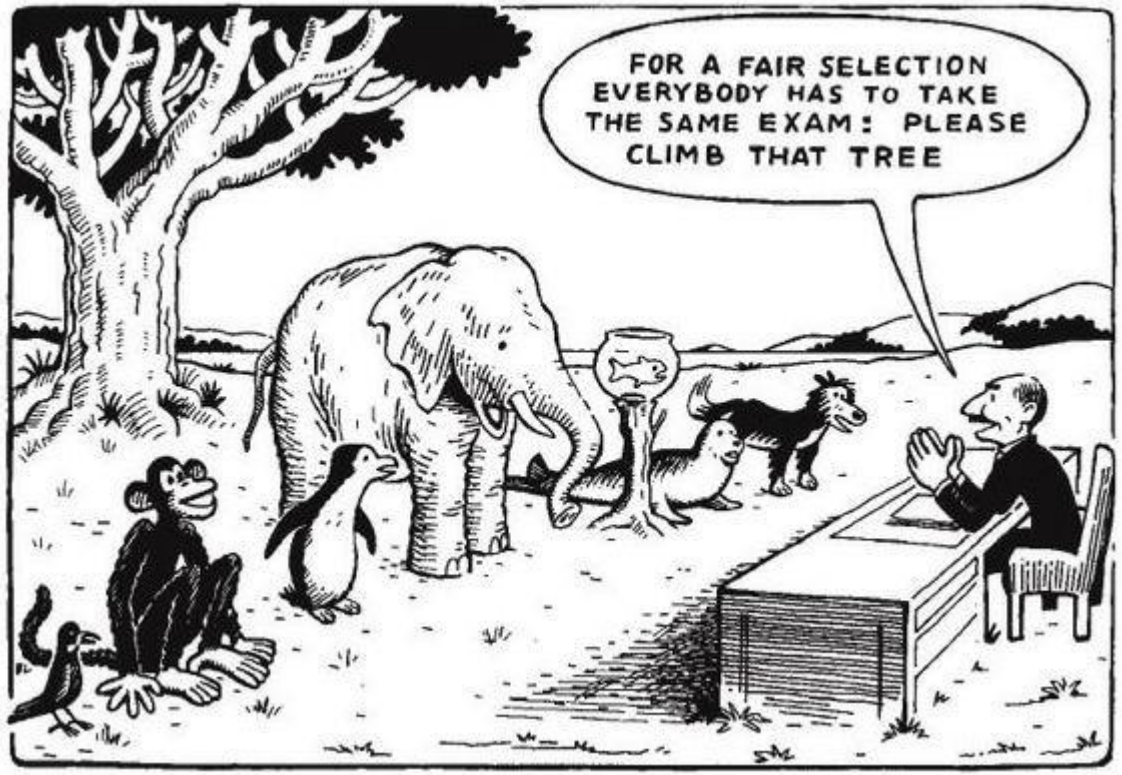
Measuring intelligence



Raven’s progressive matrices

Measuring intelligence

How can we define and measure inteligence?



https://marquetteeducator.files.wordpress.com/

Intelligence

- Intelligence is the ability to solve problems. But what problems? Intelligence is notoriously hard to define and measure. Intelligence tests apply only to humans (with a certain cultural background). They depend on an arbitrary notion of ”task” or ”problem”.

- A very different performance measure is *biological fitness*: roughly the number of fertile offspring. Surviving until reproductive age and then reproducing successfully requires solving a series of real-life problems that are posed by the environment: find food, escape predators, reproduce.

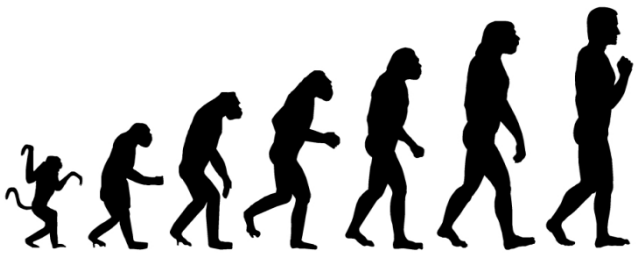
Break?

Evolution

Let’s look at some fundamental forces that have led to the development of intelligence in nature

Nothing in biology makes sense except in the light of Evolution.

T. G. Dobzhansky, 1973



Evolution: development at the population level

Learning

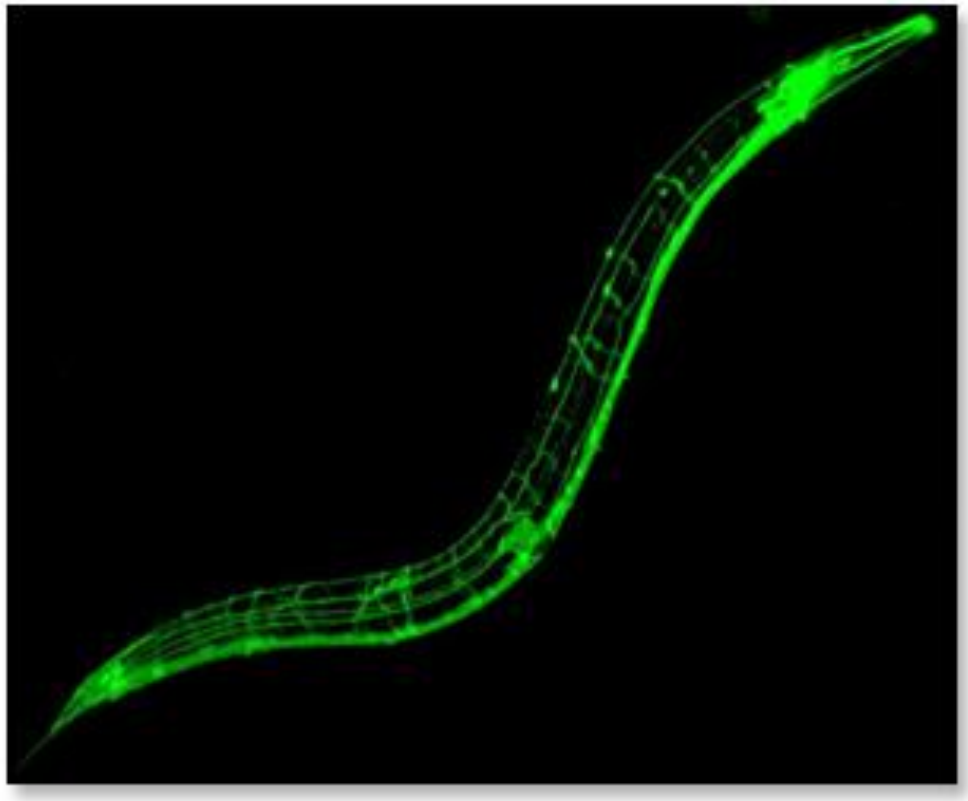
- *Radical constructivism starts from the assumption that knowledge (…) is in the heads of persons, and that the thinking subject has no alternative but to construct what he or she knows on the basis of his or her own experience.*

  von Glasersfeld, 1995

Learning: development at the individual level

Neuroscience

You change physically every time you learn something new!

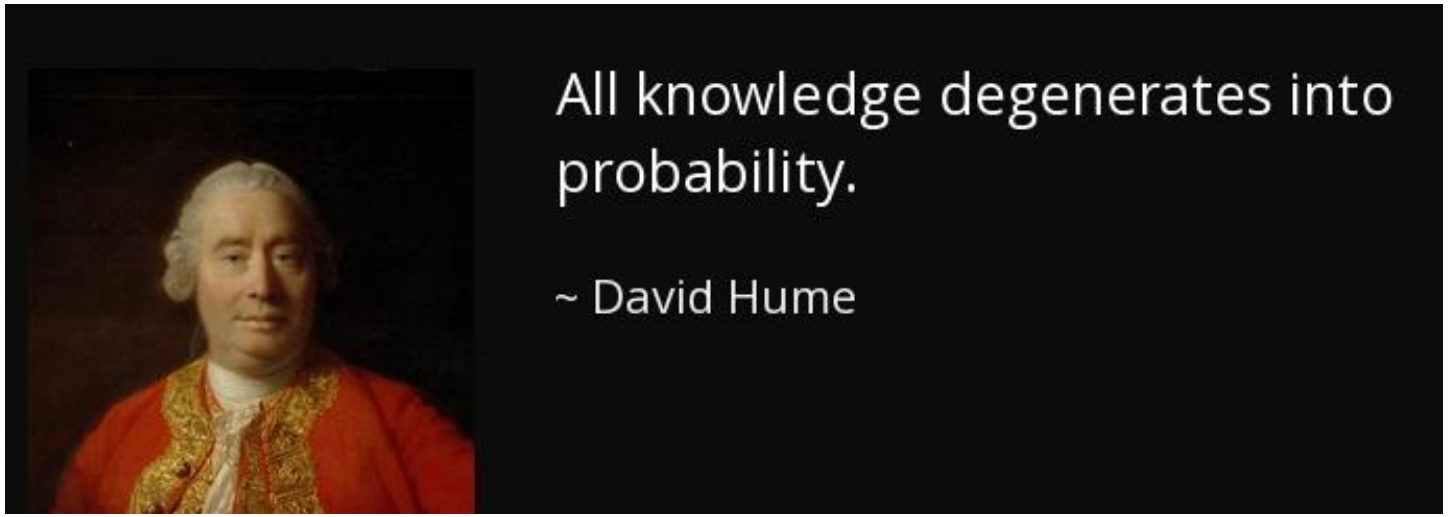


Nervous system of the nematode C. Elegans with its 302 neurons

Almost all animals have nervous systems. Essentially the only exceptions are sponges. Nervous systems typically change physically throughout the entire lifetime of the animal. Neuroplasticity enables learning/adaptation.

Empiricism

Empiricism: All knowledge comes from experience



Animal learning

Everything we do and are is determined by our history of rewards and punishments.

B. F. Skinner

Animals learn from the past and apply that experience to decision-making in the present.
This is only helpful to the extent that the present ressembles the past.

Artificial Intelligence

Pioneers of AI

Leibnitz dreamt of a machine that could be used for answering all questions about mathematics and more. The questions would be expressed in a formal language and processed by a machine called *Calculus ratiocinator.* Kurt Gödel's First Incompleteness theorem shows that no such machine can exist.

Pioneers of AI

Ada Lovelace programmed the first algorithm into a machine (the analytical engine). She also predicted that machines can be used not only for mathematical calculation, but for writing, drawing, and music composition.

Pioneers of AI

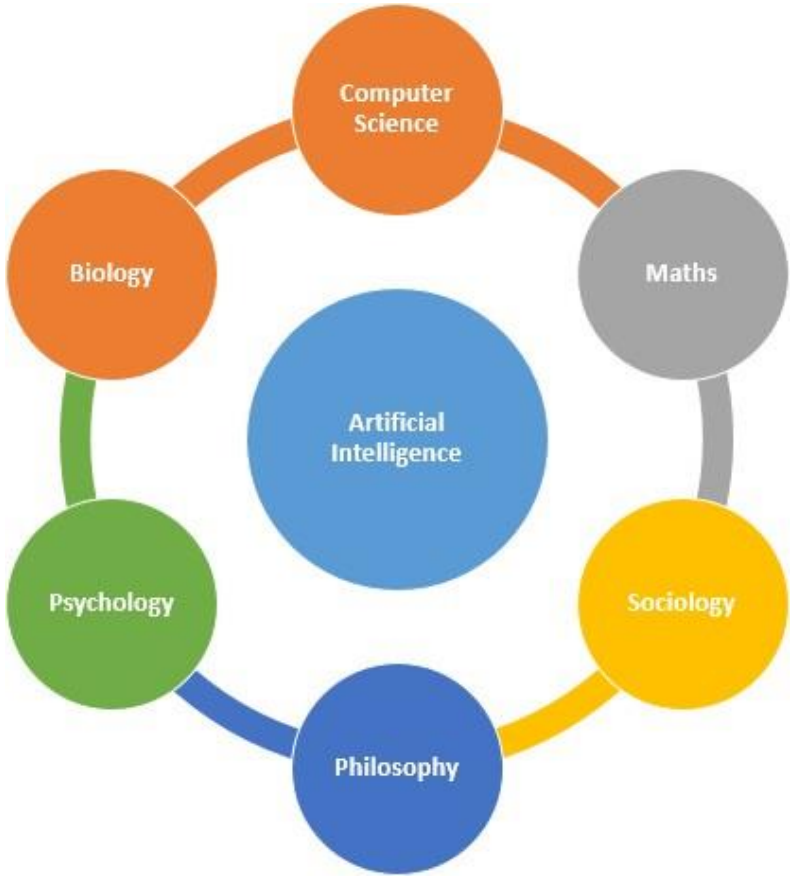
Alan Turing pioneered the study of machine intelligence by making the notion of computation precise (Turing machine), by inventing the programmable computer (universal Turing machine), by making the first chess program, by introducing the idea of machine learning, by introducing the idea of genetic programming, and by suggesting criteria for machine intelligence (Turing test).

Pioneers of AI

Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.
John McCarthy,
Dartmouth Workshop 1956

John McCarthy coined the term “artificial intelligence” with the suggested definition:
“the science and engineering of making intelligent machines”

What is AI?



A multi-disciplinary research field

Or just a part of computer science?

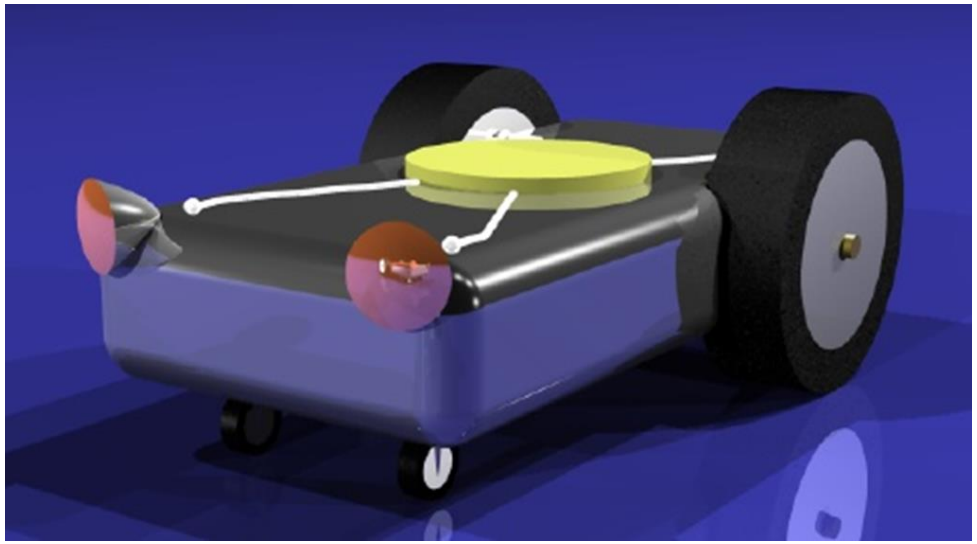
How we imitate flying is not important?

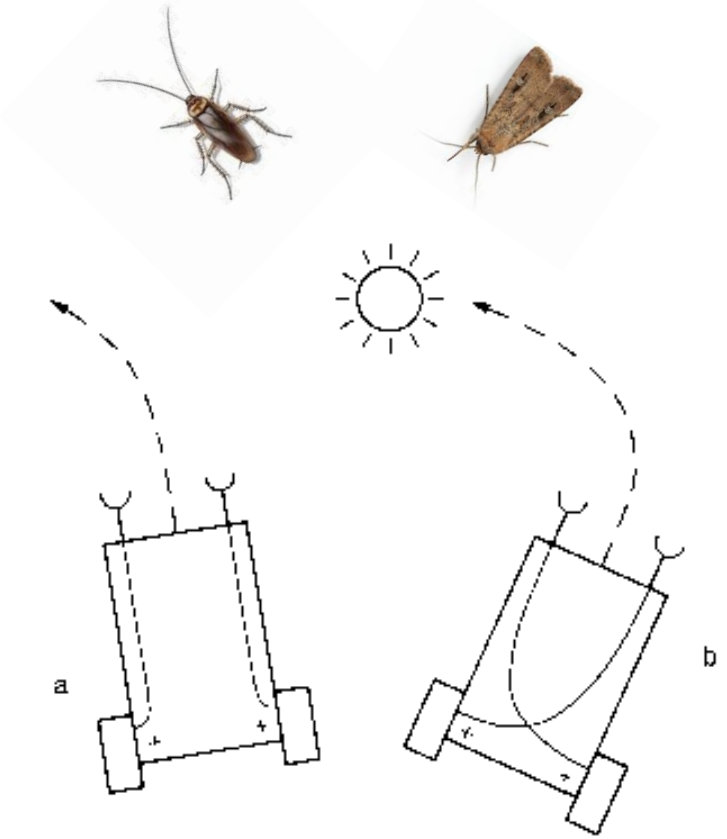
Natural versus artificial flight

We can fly but we still can’t fly like birds Stanford project

Natural versus artificial navigation

How we imitate navigation is not important?

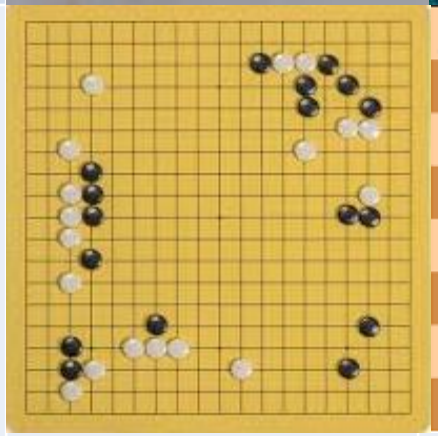




Positive and negative [phototaxis](#) in insects and in [Braitenberg vehicles](#)

Natural versus artificial intelligence

How we imitate intelligence is not important?

	LEARNING	NO LEARNING
ANIMAL		
MACHINE		

Two types of AI

- Rule-based AI
  - A human solution is translated into code
  - Data is not needed
  - No learning takes place

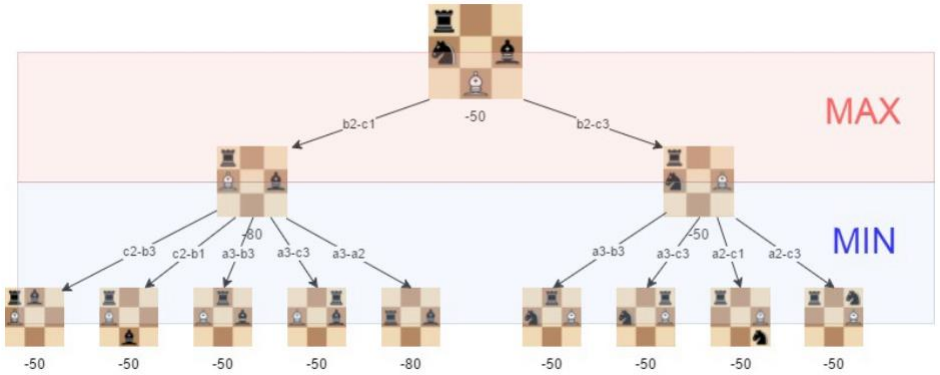
Without learning

This is the traditional way of programming, which is still used on a massive scale in control systems, administrative systems, word processors, spread sheets, etc.

- Machine learning (ML)
  - No human solution is needed
  - Data is needed
  - Learning takes place

With learning

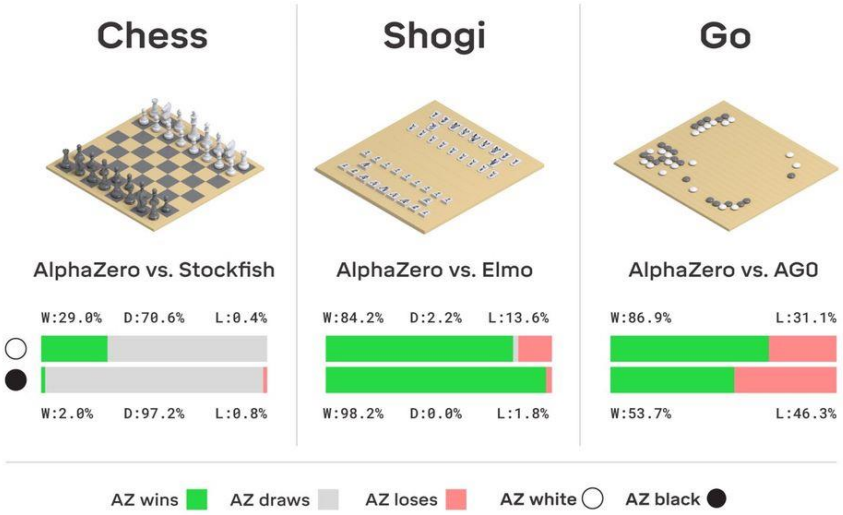
Two types of AI



Rule-based AI: Possible to write rules for chess and then search for the best move

Machine learning: Very difficult to write rules for classifying images of dog and cats!

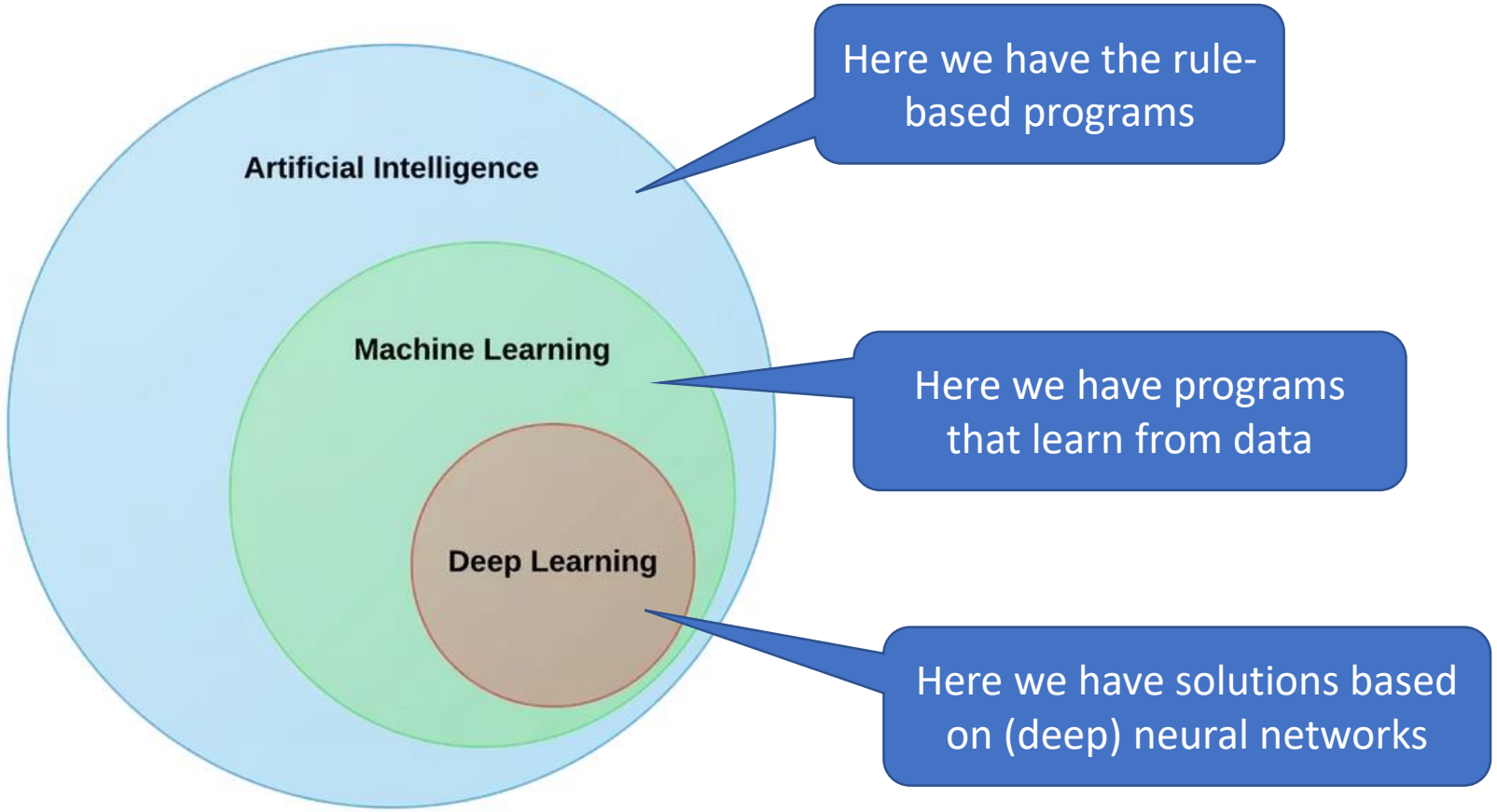
Combining the two types of AI



AlphaZero uses ML for position evaluation and search (Monte Carlo Tree Search) for look-ahead.

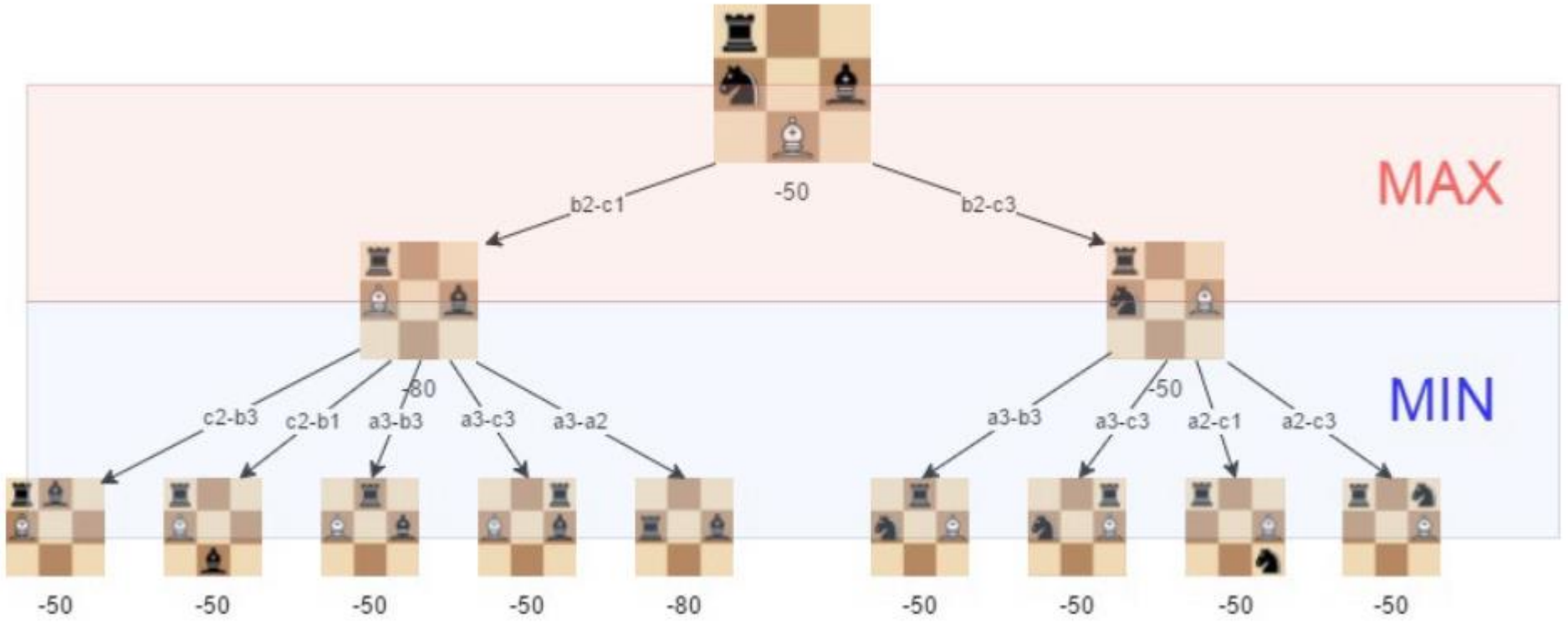
Autonomus vehicles use ML for object classification and rule-based systems for driving.

Terminology



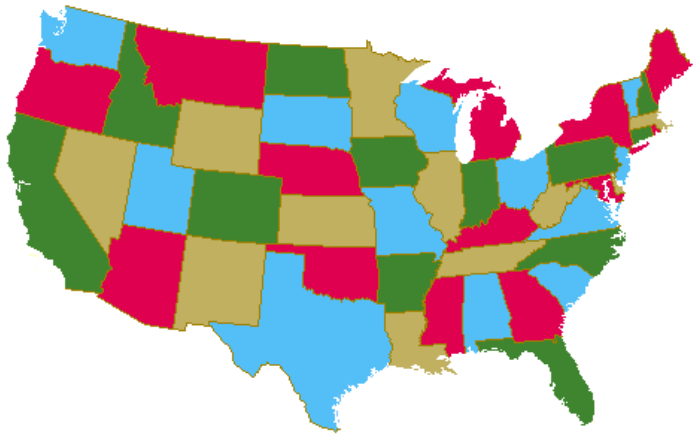
Applications of rule-based AI

Games


MAX
MIN
-50
b2-c1
b2-c3
-80
-50
c2-b3
c2-b1
a3-b3
a3-c3
a3-a2
a3-b3
a3-c3
a2-c1
a2-c3
-50
-50
-50
-50
-80
-50
-50
-50
-50

Theorem proving

Mathematics is formulated in a language (like English or logic). Theorems are sentences and proofs are lists of sentences. No data is involved, so ML is powerless here. But rule-based AI can help prove new theorems!



Four-color theorem, 1976

Kepler conjecture, 1998

[Computer-aided mathematics]

Proof verification

Data plays no role here. Theorems often quantify over the set of all natural numbers, etc!

- The Fundamental Theorem of Algebra
- The Fundamental Theorem of Arithmetic
- Ramsey's Theorem
- The Central Limit Theorem
- Gödel’s First Incompleteness Theorem
- The Law of Large Numbers

http://us.metamath.org/mm_100.html#5

System verification

- Prove properties of
  - Hardware applications
  - Financial applications
  - Medical applications

Testing can only cover a tiny fraction of a large state space, so it is not suitable for large function-critical systems. Theorem-proving (a form of rule-based AI) gives a much higher level of reliability (when it can be used).

System verification

Rule-based AI is used for verifying traffic control systems, e.g. the Paris Metro. One can prove that there can never be two green lights in opposite directions, etc. With 1000+ switches there are 2^{1000}+ states, so exhaustive testing is impossible.

Control systems

No learning here. A fixed algorithm that takes a map and two locations as input.

No learning here. The Chalmers minibus moves on a virtual track and breaks if any of the 8 LIDARs detects motion near by.

Next time

- We will see several examples of machine learning applications
- We will also introduce the field of data science

Until next time

- Please install Anaconda (not just Miniconda)
- That will give you Python 3.8 and Jupyter Notebook
- You will also get several packages:
  - Pandas (data science)
  - NumPy (math)
  - Matplotlib (plots)
- Also please install Tensorflow (for neural networks)