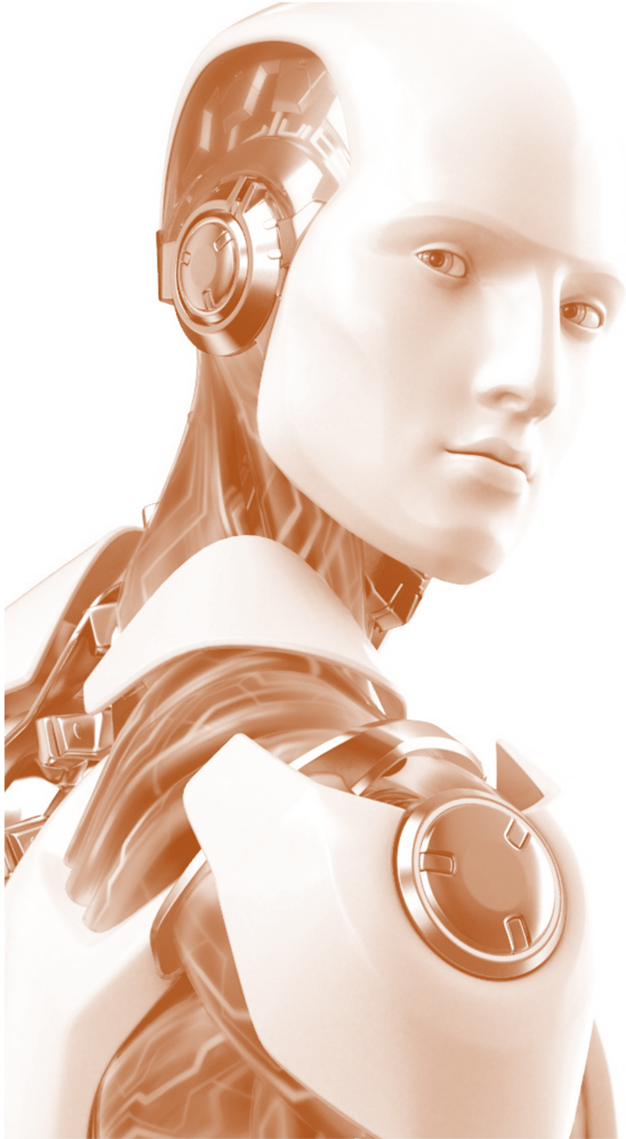


Machine Learning

Lecture 1 - Introduction to Machine Learning

Profa. Dra. Esther Luna Colombini
esther@ic.unicamp.br

Prof. Dr. Alexandre Simoes
alexandre.simoes@unesp.br

LaRoCS – Laboratory of Robotics and Cognitive Systems

UNICAMP

Advanced Institute for Artificial Intelligence

Topics

- Machine Learning History
- Machine Learning Today
- Definitions of Machine Learning
- Machine Learning Approach
- Types of Learning:
  - Unsupervised Learning
  - Supervised Learning
  - Semi-supervised Learning
  - Reinforcement Learning
  - Evolutionary Learning
  - Inductive Learning
  - Deductive Learning
  - Deep Learning
  - Deep Reinforcement Learning

Most valuable resource today

The Economist   Topics   Print edition   More

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

Print edition | Leaders ›

May 6th 2017

Why Now?

BIGGER DATA

BETTER HARDWARE

SMARTER ALGORITHMS

Data

Eevery 60 seconds, 136,000 photos are uploaded, 510,000 comments are posted, and 293,000 status updates are posted.

More than five billion instances of people listening to songs online have been catalogued by Facebook withing 5 months

4 intermediary Facebook friends are usually enough to introduce anyone to a random stranger according to a study analyzing 69 billion friend connections among 721 million people

Your mobile phone apps

Know where you are and your daily habits

Know what you say

Know your preferences

… with your consent

Frameworks

TensorFlow

PYTORCH

scikit learn

colab

Keras

docker

Google Cloud

aws

H2O.ai

Caffe2

ANACONDA

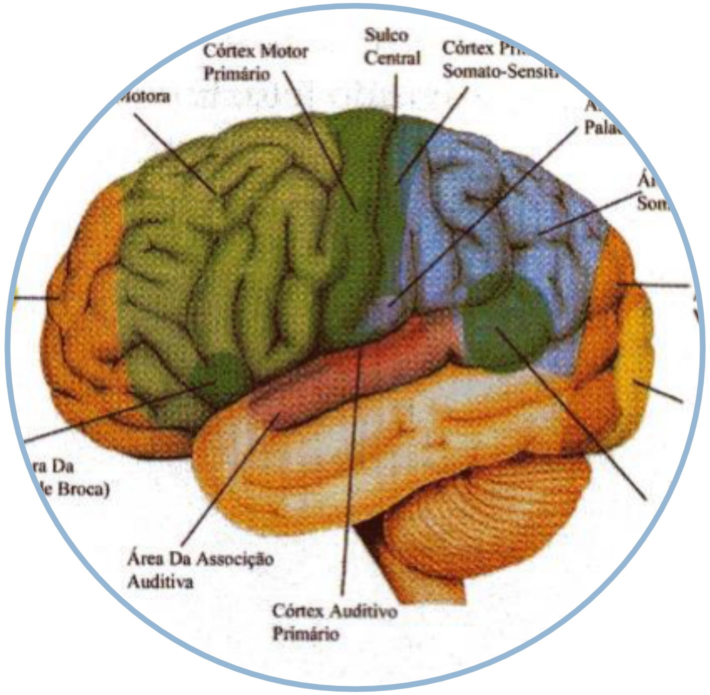
Azure

GPUs (Graphics Processing Unit)

A Glimpse on AI

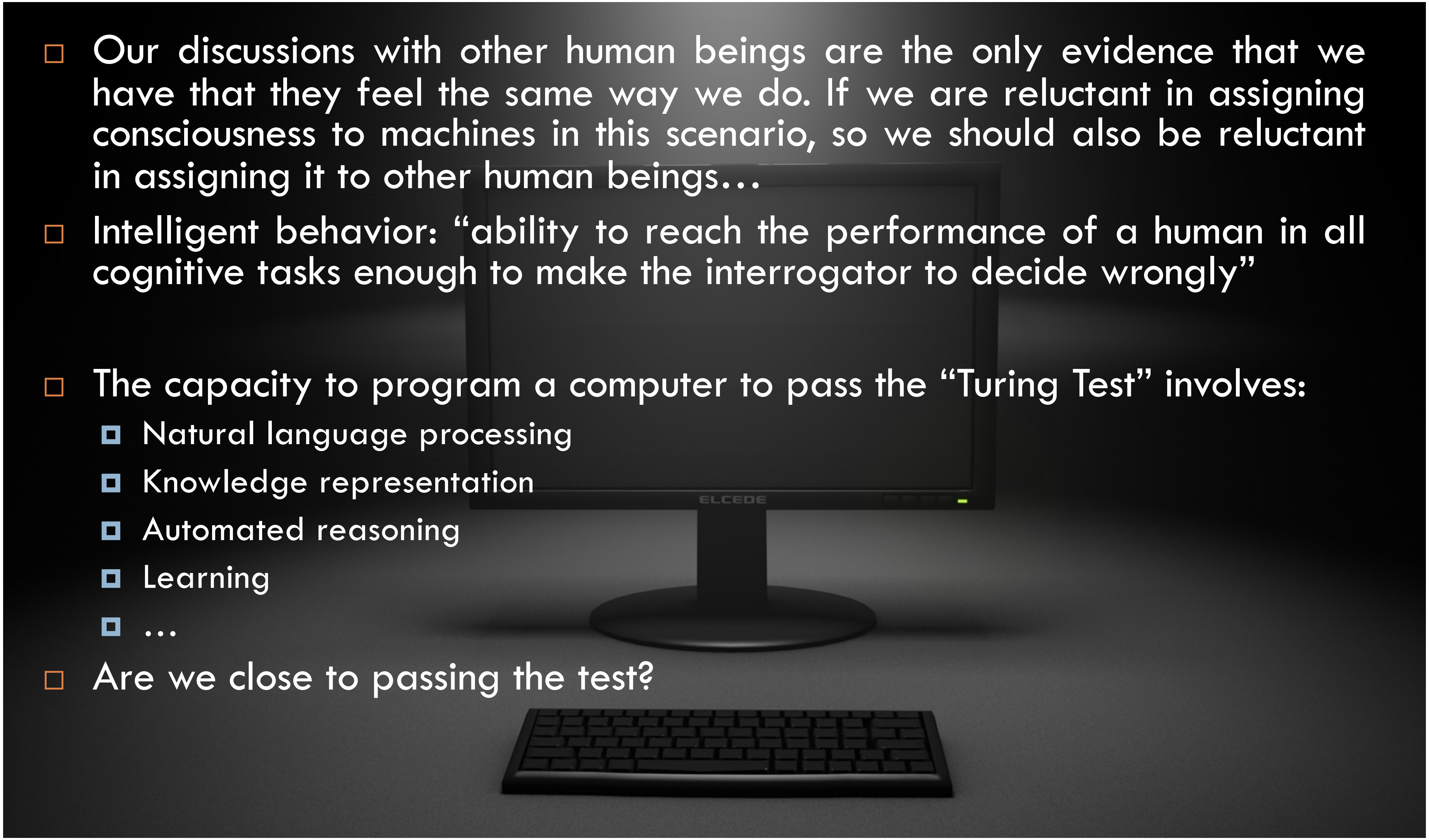
Natural x artificial

Natural x artificial



Turing Test

Turing Test



- Our discussions with other human beings are the only evidence that we have that they feel the same way we do. If we are reluctant in assigning consciousness to machines in this scenario, so we should also be reluctant in assigning it to other human beings…
- Intelligent behavior: “ability to reach the performance of a human in all cognitive tasks enough to make the interrogator to decide wrongly”
- The capacity to program a computer to pass the “Turing Test” involves:
  - Natural language processing
  - Knowledge representation
  - Automated reasoning
  - Learning
  - …
- Are we close to passing the test?

Dartmouth Summer Research Project on Artificial Intelligence (1956)

AI Paradigms

How AI is divided

Symbolic

linguistic metaphor
ex. production systems, agents, planning, learning

Connectionist

brain metaphor
ex. neural networks and reinforcement learning

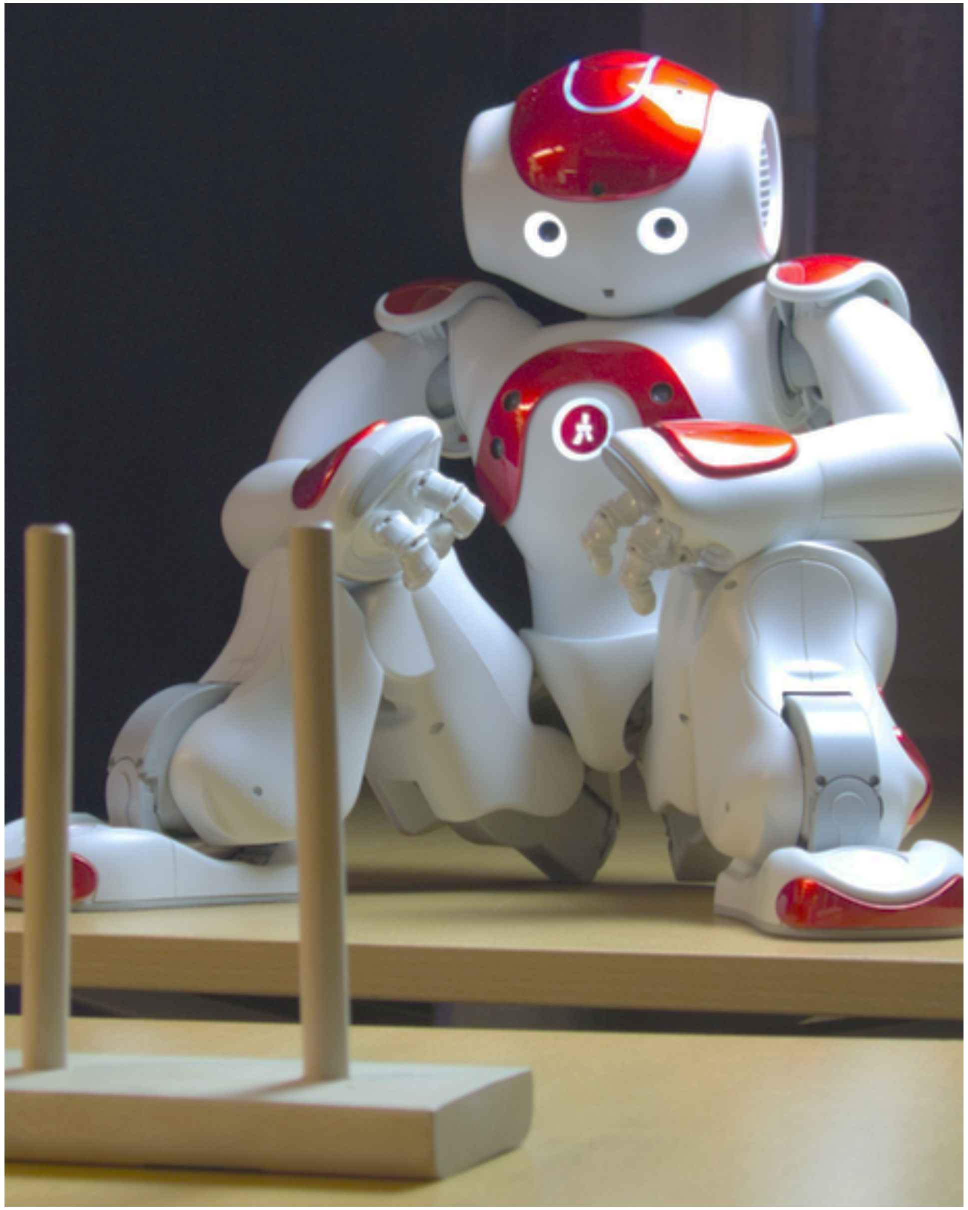
Evolutionist

metaphor of nature
ex. genetic algorithms, artificial life and collective intelligence

Statistical / Probabilistic

uncertainty metaphor
ex. Bayesian networks and fuzzy systems

Symbolic paradigm of AI

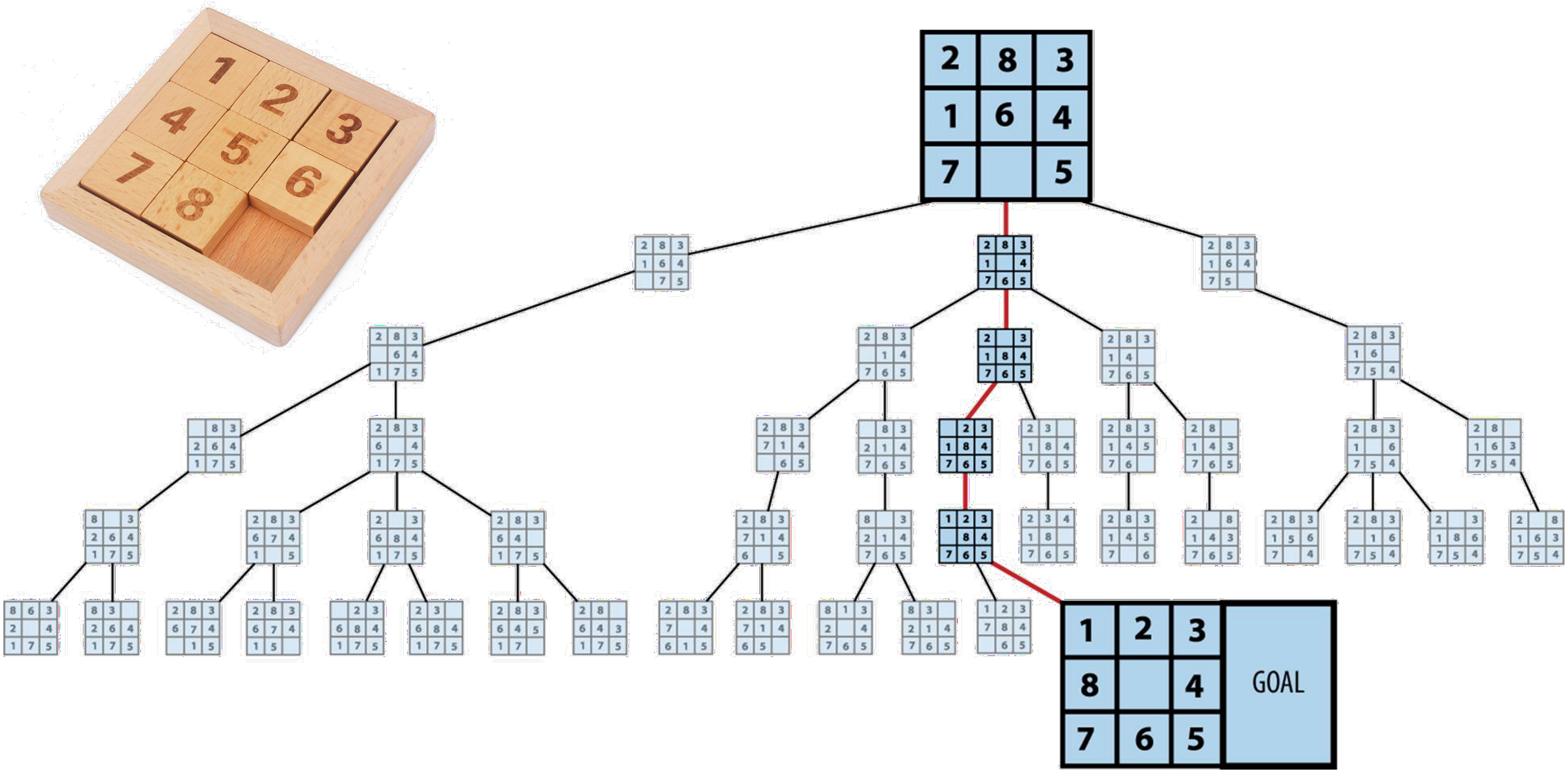


In a system with physical symbols (or formal system) there are symbols that are combined in structures (expressions) and manipulated (through specific processes) to produce new expressions

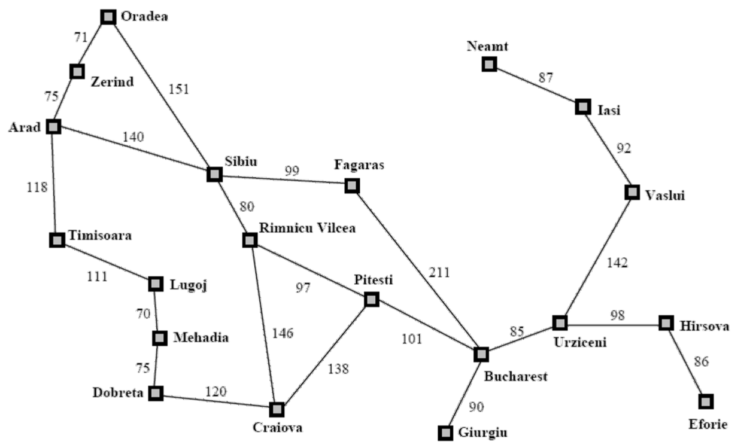
Physical symbol system hypothesis:

“A physical symbols system has the necessary and sufficient means for intelligent action in general”

Symbolic paradigm of AI

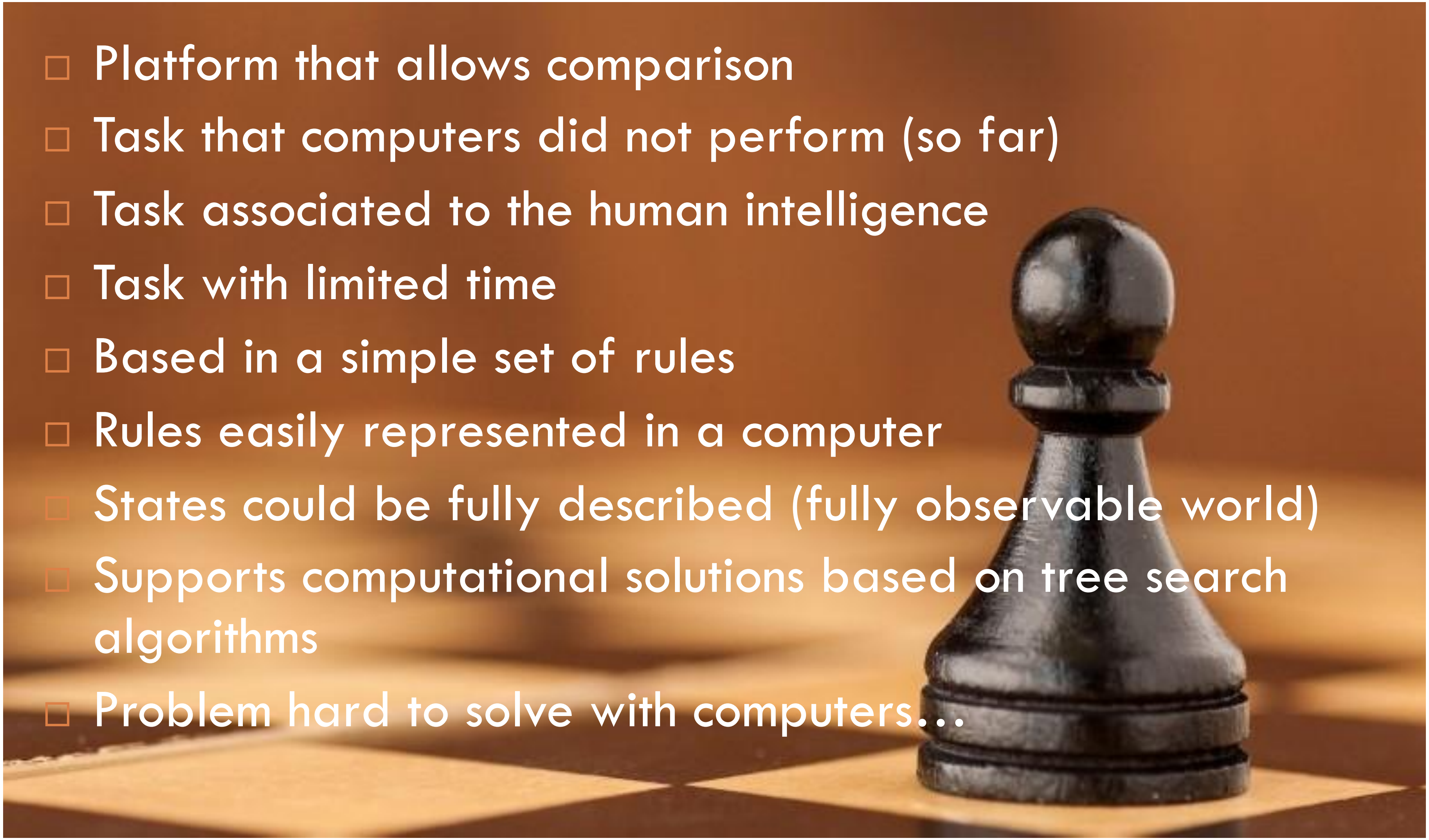


Symbolic paradigm of AI



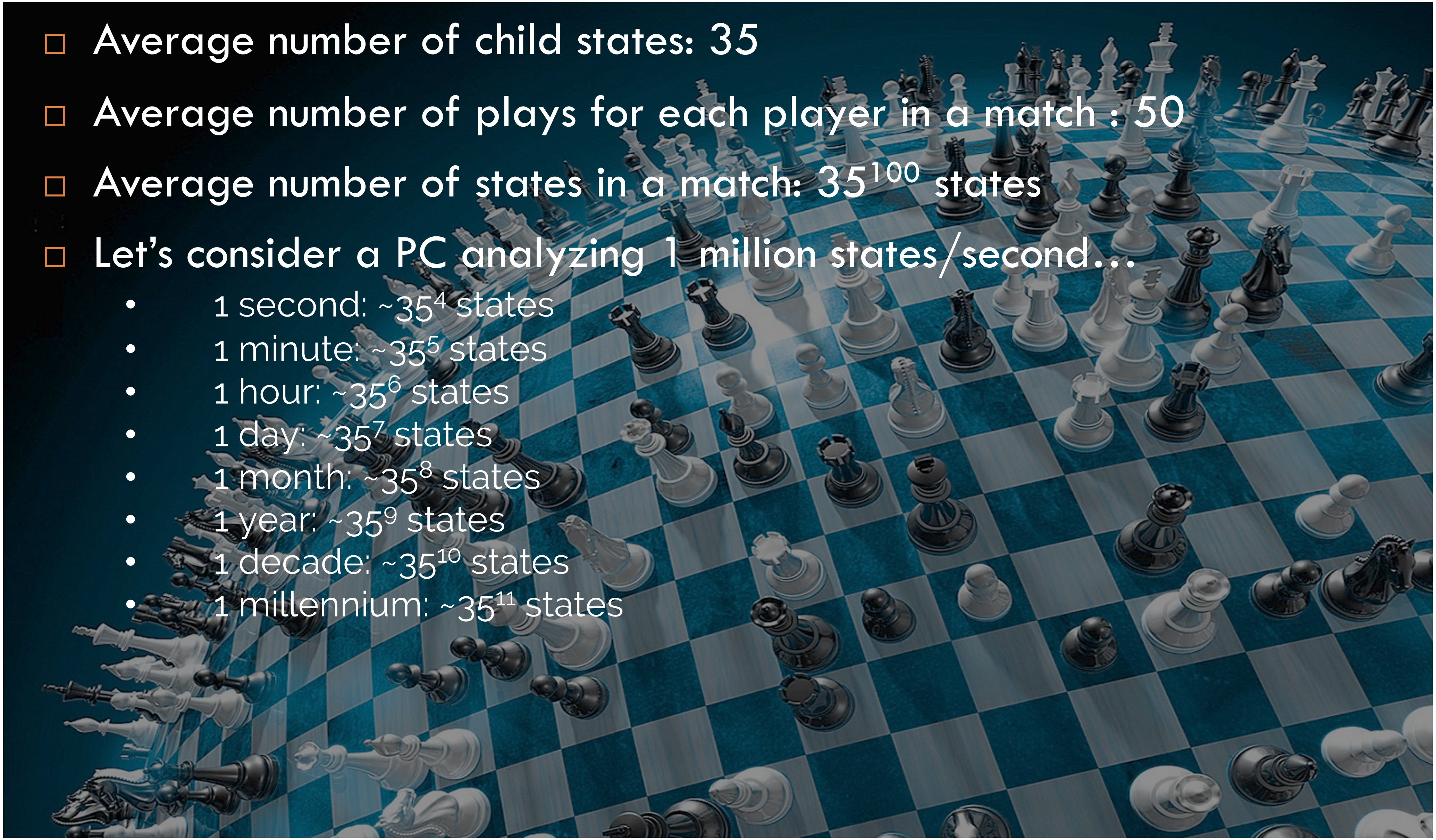

the rematch
GARRY
KASPAROV
DEEP
BLUE

Why chess?



- Platform that allows comparison
- Task that computers did not perform (so far)
- Task associated to the human intelligence
- Task with limited time
- Based in a simple set of rules
- Rules easily represented in a computer
- States could be fully described (fully observable world)
- Supports computational solutions based on tree search algorithms
- Problem hard to solve with computers...

How hard?



- Average number of child states: 35
- Average number of plays for each player in a match : 50
- Average number of states in a match: 35^{100} states
- Let's consider a PC analyzing 1 million states/second…
  - 1 second: ~35^4 states
  - 1 minute: ~35^5 states
  - 1 hour: ~35^6 states
  - 1 day: ~35^7 states
  - 1 month: ~35^8 states
  - 1 year: ~35^9 states
  - 1 decade: ~35^{10} states
  - 1 millennium: ~35^{11} states

How hard?

- Average number of child states: 35
- Average number of plays for each player in a match : 50
- Average number of states in a match: 35^{100} states
- Let's consider a PC analyzing 1 million states/second…
  - 1 second: ~35^4 states
  - 1 minute: ~35^5 states
  - 1 hour: ~35^6 states
  - 1 day: ~35^7 states
  - 1 month: ~35^8 states
  - 1 year: ~35^9 states
  - 1 decade: ~35^{10} states
  - 1 millennium: ~35^{11} states
  - 4 billion years (Andromeda and Milky Way collision): ~35^{14} states

How hard?

- Average number of child states: 35
- Average number of plays for each player in a match : 50
- Average number of states in a match: 35^{100} states
- Let's consider a PC analyzing 1 million states/second…
  - 1 second: ~35^{4} states
  - 1 minute: ~35^{5} states
  - 1 hour: ~35^{6} states
  - 1 day: ~35^{7} states
  - 1 month: ~35^{8} states
  - 1 year: ~35^{9} states
  - 1 decade: ~35^{10} states
  - 1 millennium: ~35^{11} states
  - 4 billion years (Andromeda and Milky Way collision): ~35^{14} states
  - 7 billion years (Sol becomes red giant): ~35^{15} states

How hard?

- Average number of child states: 35
- Average number of plays for each player in a match : 50
- Average number of states in a match: 35^{100} states
- Let's consider a PC analyzing 1 million states/second…

  - 1 second: ~35^4 states
  - 1 minute: ~35^5 states
  - 1 hour: ~35^6 states
  - 1 day: ~35^7 states
  - 1 month: ~35^8 states
  - 1 year: ~35^9 states
  - 1 decade: ~35^{10} states
  - 1 millennium: ~35^{11} states
  - 4 billion years (Andromeda and Milky Way collision): ~35^{14} states
  - 7 billion years (Sol becomes red giant): ~35^{15} states
  - 16 billion years (end of universe?): ~35^{16} states

Deep Blue main innovations

- Software:
  - New function to evaluate states (measures the quality of a state)
- Hardware:
  - 256 processors (200 million positions per second)

Source:
https://www.research.ibm.com/deepblue/

Chess history (50 years)

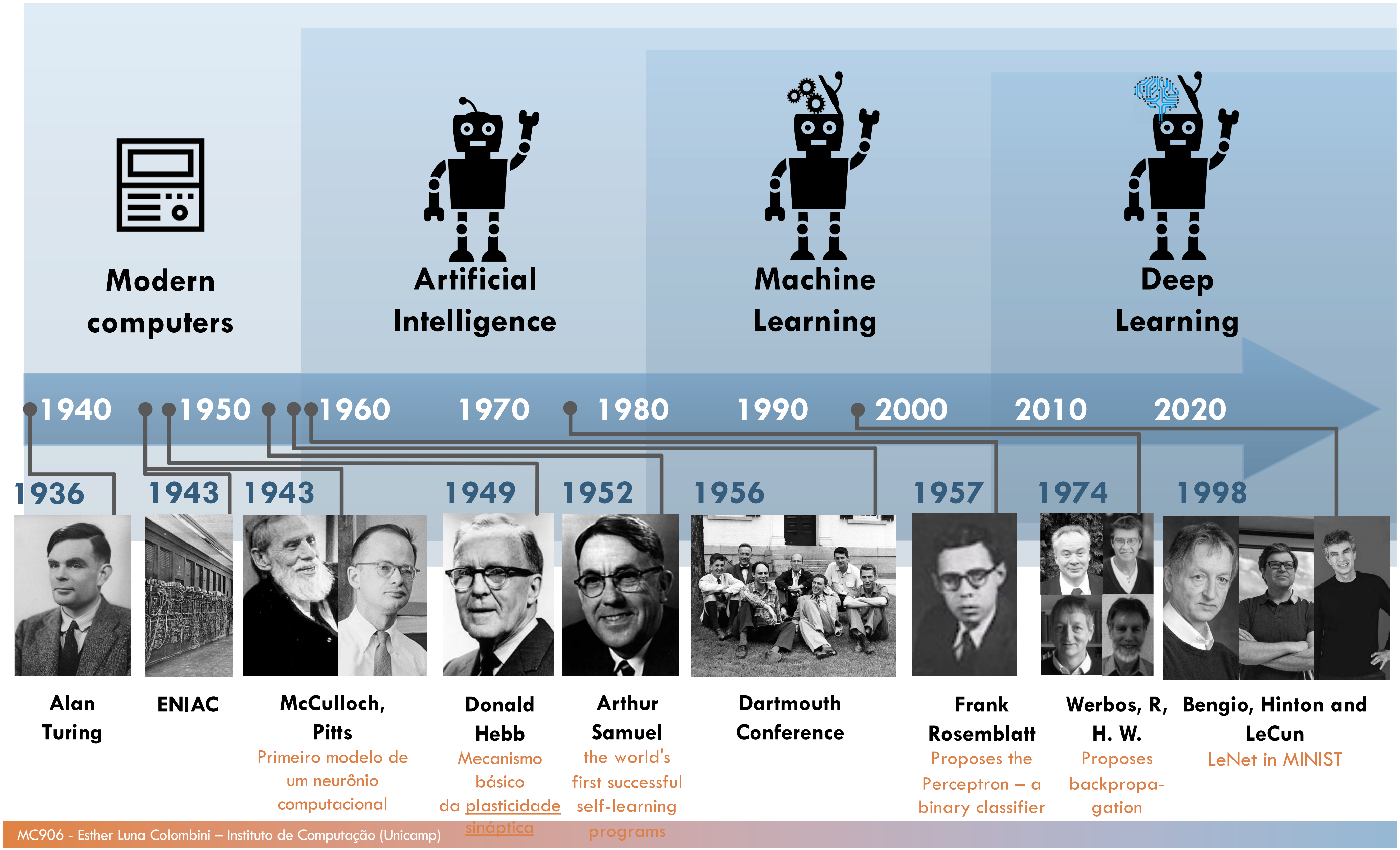
- 1950 – 1st chess program (Claude Shannon and Alan Turing)
- 1956 (MANIAC) – 1st computer to win against a human (amateur) in 23 movements (Los Alamos Scientific Lab)
- 1966-68 (Mac Hack) – computer wins match in MIT (student x professor)
- 1968-78 (Chess x.x) – 1st computer to win a tournament against humans (1.500 positions per second). Master chess player David Levy warns of the fact that, soon, there could be a surprise with the software
- 1981 (Cray Blitz) – 1st computer to win against a chess master (Joe Sentef) in a tornament, and the 1st to receive a master score
- 1989 (CMU-IBM Deep Thought) – It finished a tournament in 1st place (tied with a human), ahead of a former world champion (Mikhail Tal). Deep Thought has lost 2 matches of Kasparov
- 1996 (IBM Deep Blue) – Match of 6 games against Kasparov. Kasparov lost the first, won 3 and conceded 2 draws
- 1997 (IBM Deep Blue) – Improved version of Deep Blue played 6 games against Kasparov. Kasparov won the 1st, lost the 2nd, yielded 3 draws, and then lost the final. 1st computer to beat the world champion in a match.

What is AI?

The automation of activities that we associate with human thinking (decision-making, problem solving, learning…) (Bellman, 1978)	The science that attempts to understand intelligent entities (Russel & Norvig, 1995)	The study of mental faculties through the use of computational models (Charniak and McDermott, 1985)	A field of study that seeks to explain and emulate intelligente behavior in terms of computational processes (Schalkoff, 1990)
The art of creating machines that perform functions that require intelligence when performed by people (Kurzweil, 1990)	The study of how to make computers to things at which, at the moment, people are better (Rich & Knight, 1991)	The branch of computer science that is concearned with the automation of intelligent behavior (Luger & Stubblefield, 1993)	The study of the computations that make it possible to perceive, reason and act (Winston, 1992)

Machine Learning

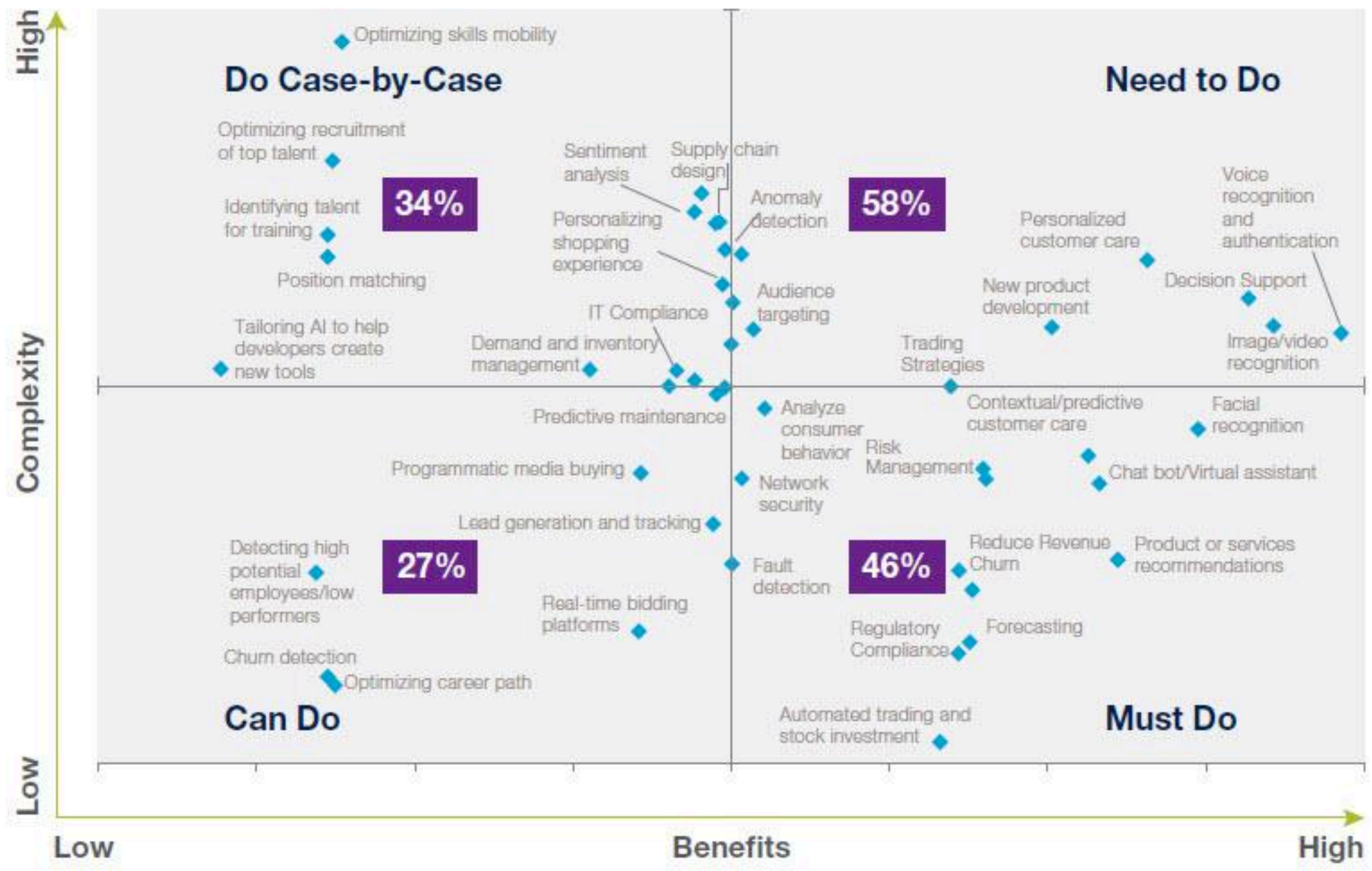
- Why learn?
  - The ability to learn is a fundamental part of the concept of intelligence
  - A learner agent is more flexible
  - Learning allows you to deal with new situations (the world is dynamic)
  - Gives agent autonomy
  - Facilitates the designer task
  - Only the essential is programmed and the system learns the rest
  - To allow for adaptability of the dynamic environment system
  - Reactivity
- How to build programs (agents) that automatically improve with experience?



Modern computers
Artificial Intelligence
Machine Learning
Deep Learning
1940
1950
1960
1970
1980
1990
2000
2010
2020
1936
1943
1943
1949
1952
1956
1957
1974
1998
Alan Turing
ENIAC
McCulloch, Pitts
Primeiro modelo de um neurônio computacional
Donald Hebb
Mecanismo básico da plasticidade sináptica
Arthur Samuel
the world's first successful self-learning programs
Dartmouth Conference
Frank Rosemblatt
Proposes the Perceptron – a binary classifier
Werbos, R, H. W.
Proposes backpropa-gation
Bengio, Hinton and LeCun
LeNet in MINIST

ML Today

Figure 9. Distribution of use cases by benefits and complexity



Font: Forbes, 2018.

Machine Learning is...

- ... the field of study that gives computers the ability to **learn** without being explicitly programmed (Arthur Samuel, 1959)

- … adapting the system such that it can perform the same task or the task in the same population **more efficiently next time** (Simon, 1983)

- ... a set of methods that can automatically detect patterns in data, and then **use the uncovered patterns** to predict future data, or to perform other kinds of decision making under uncertainty (Murphy, 2012)

Machine Learning is...

- ... programming computers to **optimize a performance criterion** using example data or past experience (Alpaydin, 2014)

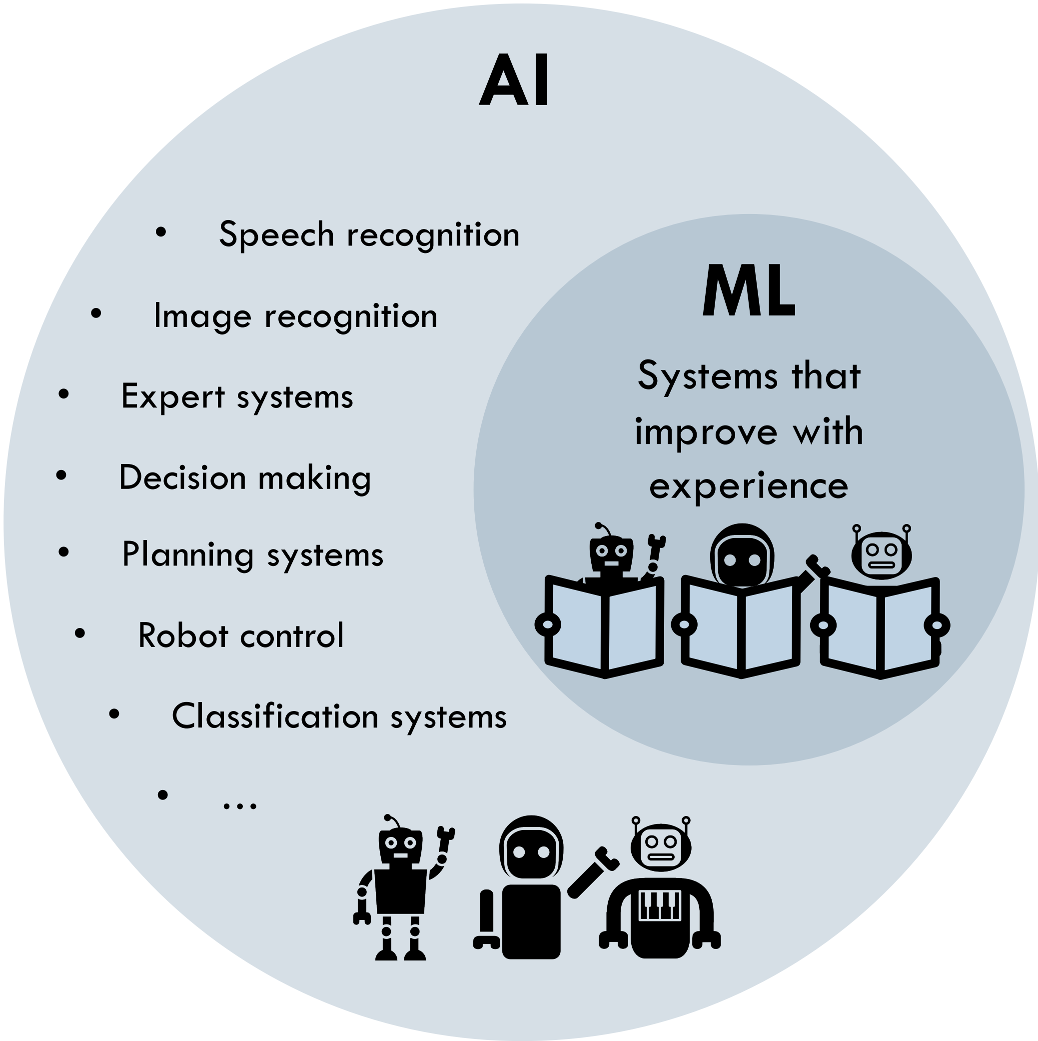
- ... making computers **modify or adapt their actions** (whether these actions are making predictions or controlling a robot) so that these actions get more accurate, where accuracy is measured by how well the chosen actions reflect the correct ones (Marsland, 2015)

- ... the science (and art) of programming computers so they can **learn from data** (Géron, 2017)

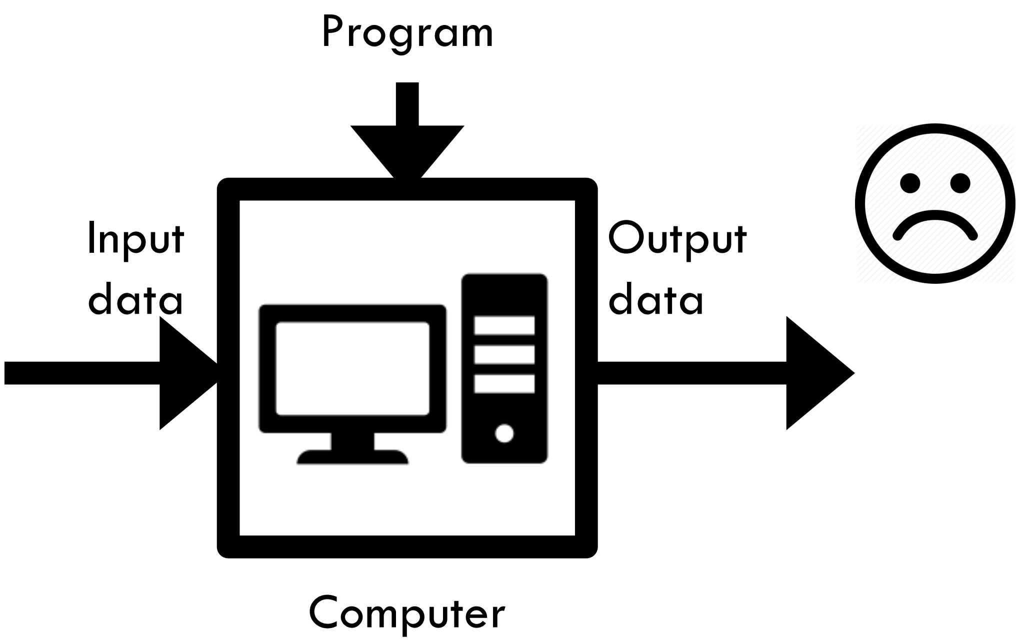
Machine Learning is...

- ... the field concerned with the question of how to construct computer programs that automatically **improve with experience** (Mitchell, 1987)

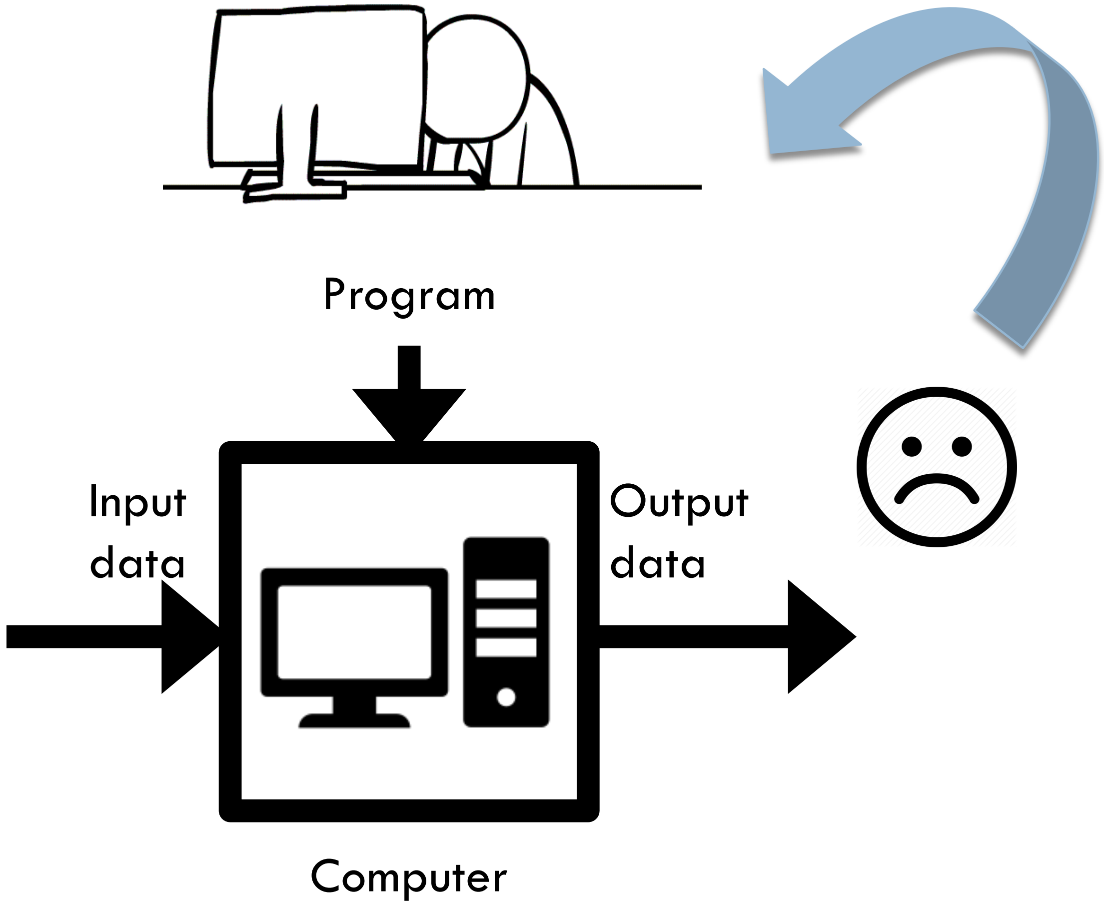
- Formally:
  - A computer program is said to **learn** form experience *E* with respect to some class of task *T* and performance measure *P*, if its performance at tasks in *T*, as measured by P, improves with experience *E*


AI
• Speech recognition
• Image recognition
• Expert systems
• Decision making
• Planning systems
• Robot control
• Classification systems
• ...
ML
Systems that improve with experience

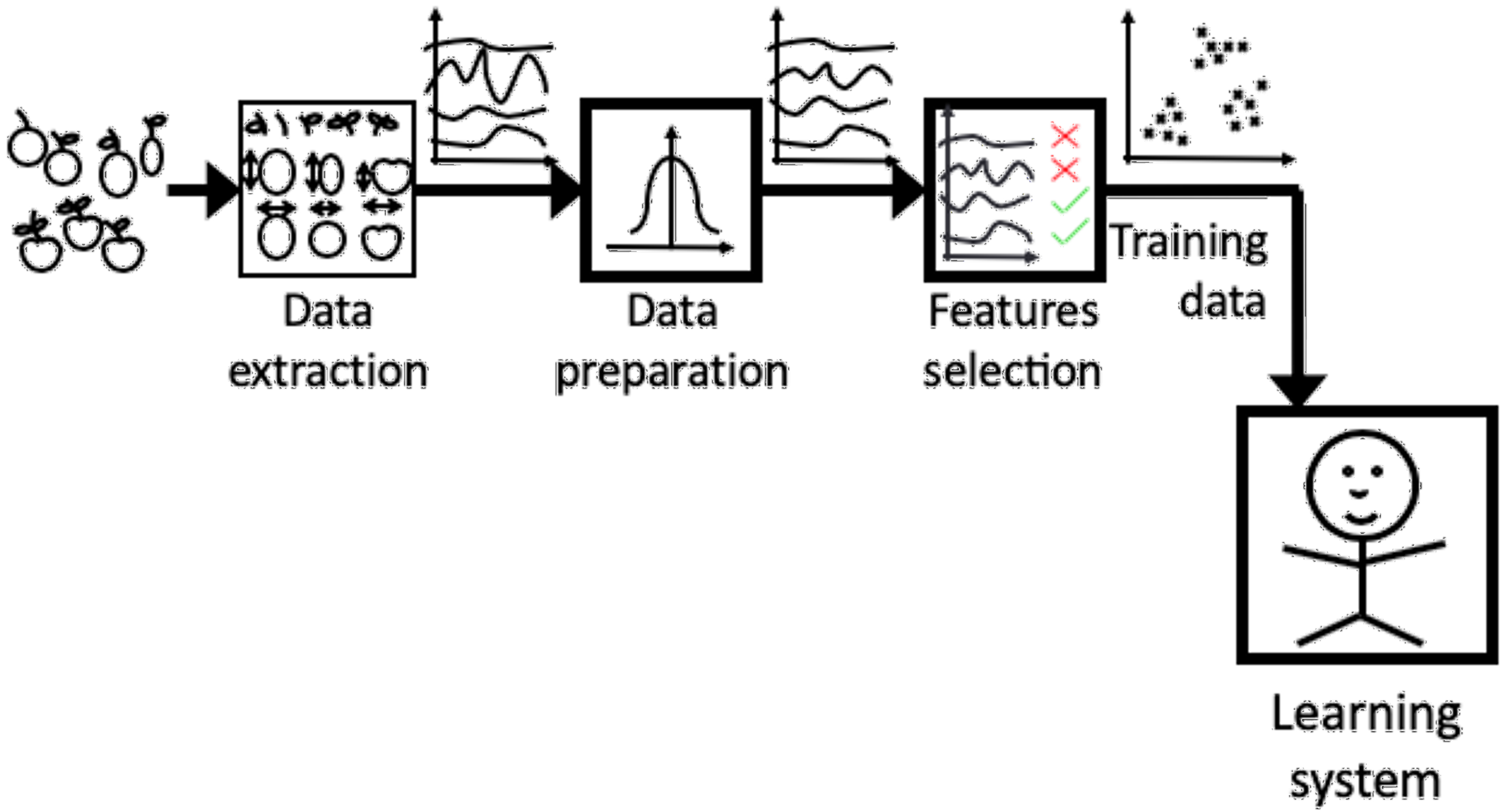
Traditional approach



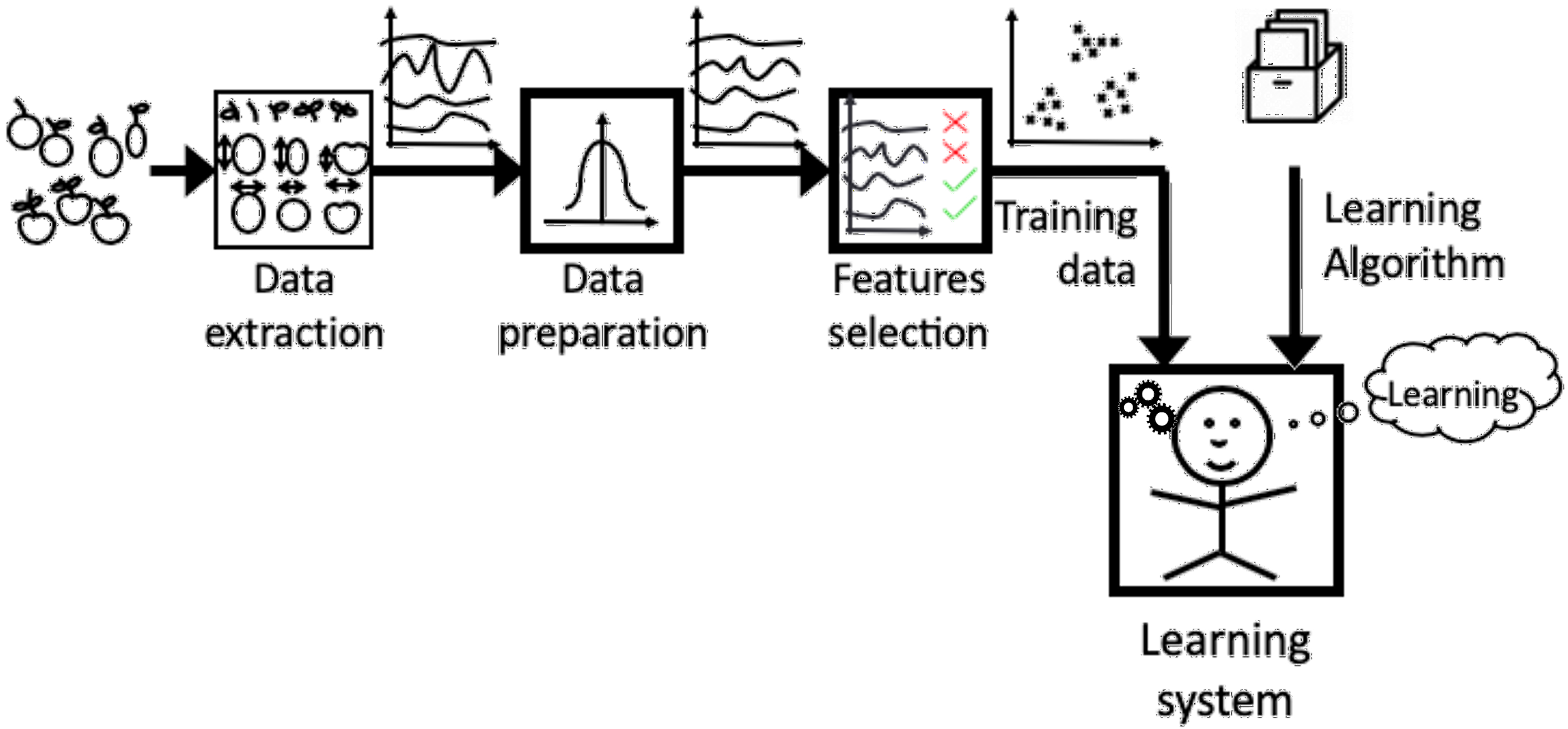
Traditional approach



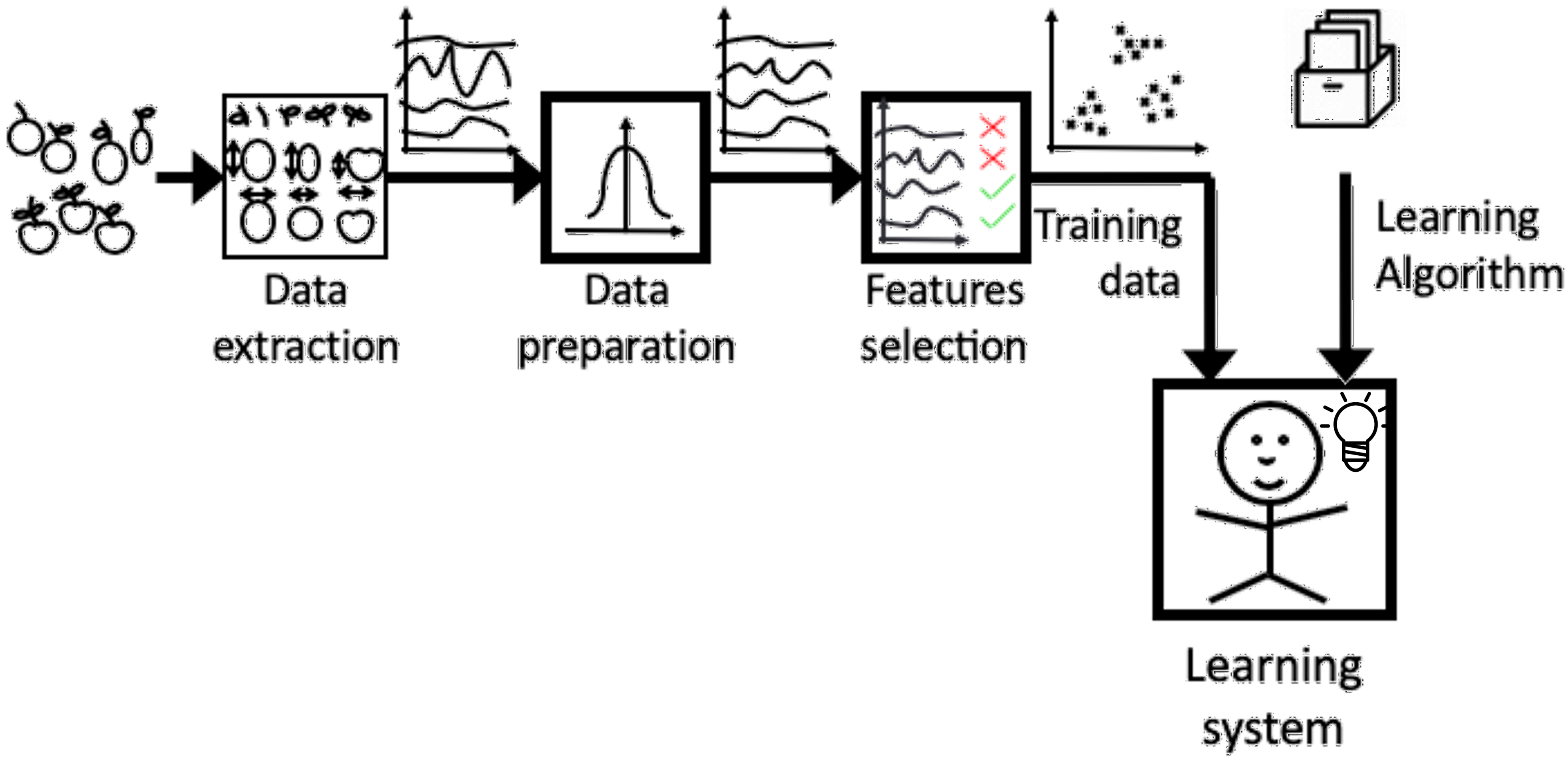
ML approach


Data
extraction
Data
preparation
Features
selection
Training
data
Learning
system

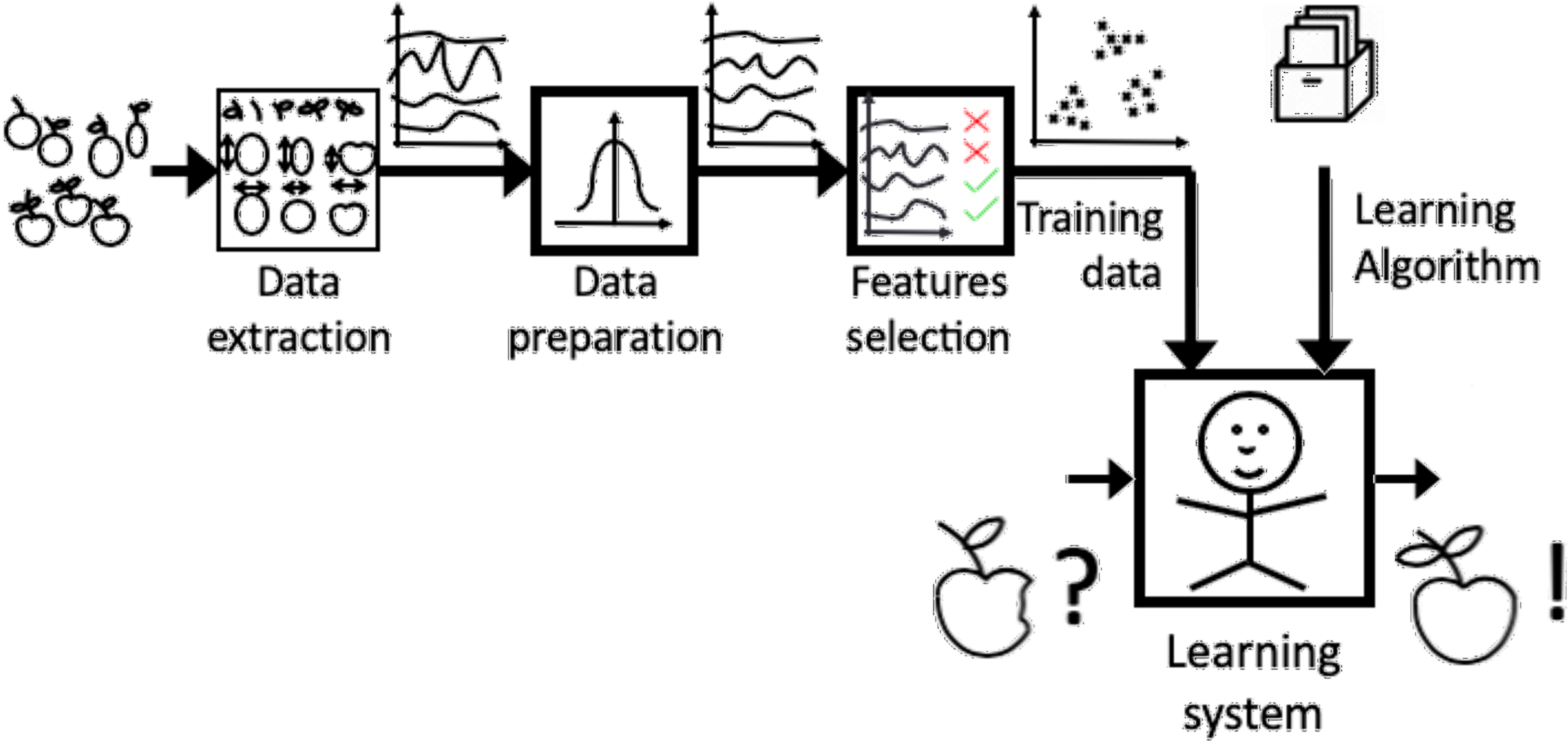
ML approach


Data
extraction
Data
preparation
Features
selection
Training
data
Learning
Algorithm
Learning
Learning
system

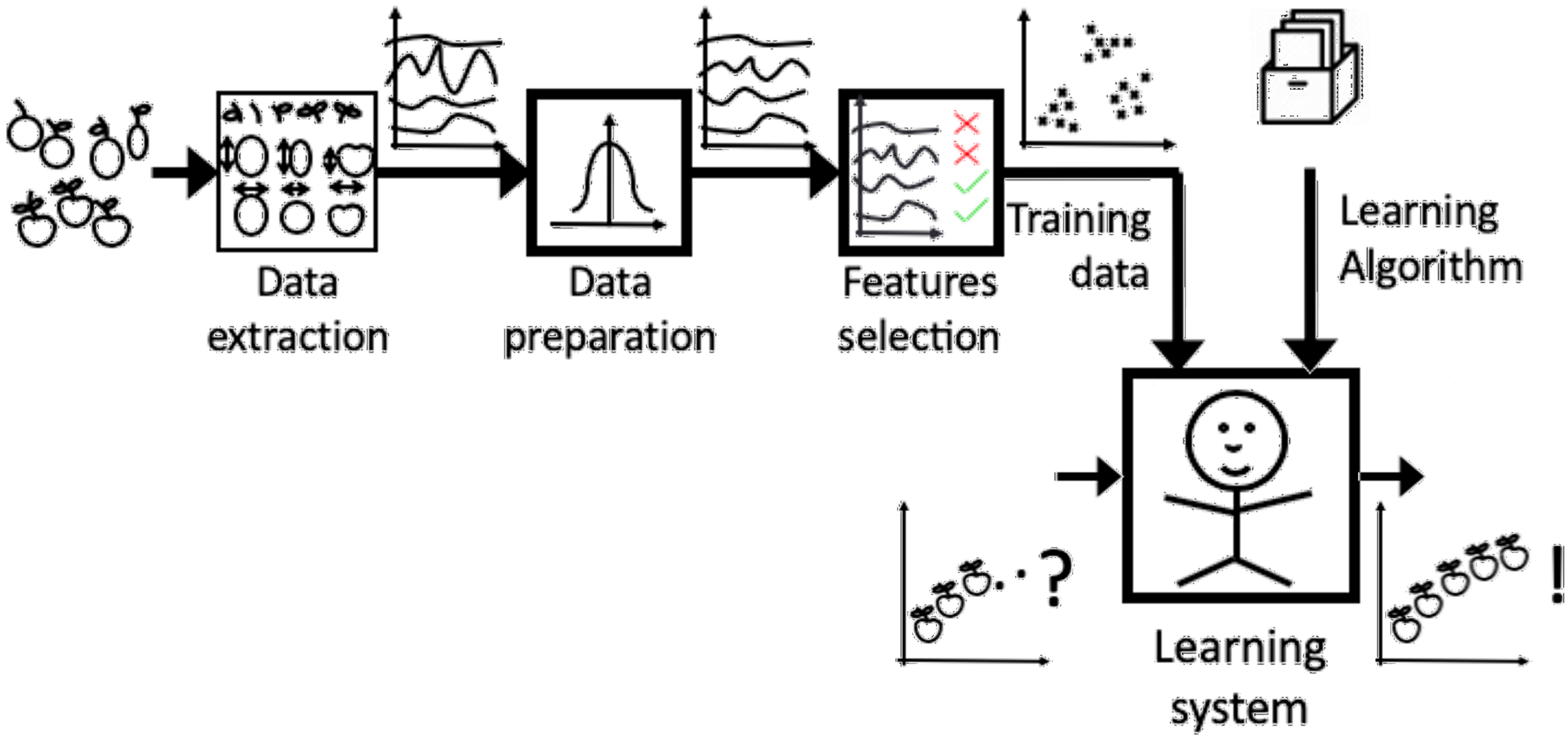
ML approach


Data
extraction
Data
preparation
Features
selection
Training
data
Learning
Algorithm
Learning
system

ML approach


Data
extraction
Data
preparation
Features
selection
Training
data
Learning
Algorithm
Learning
system

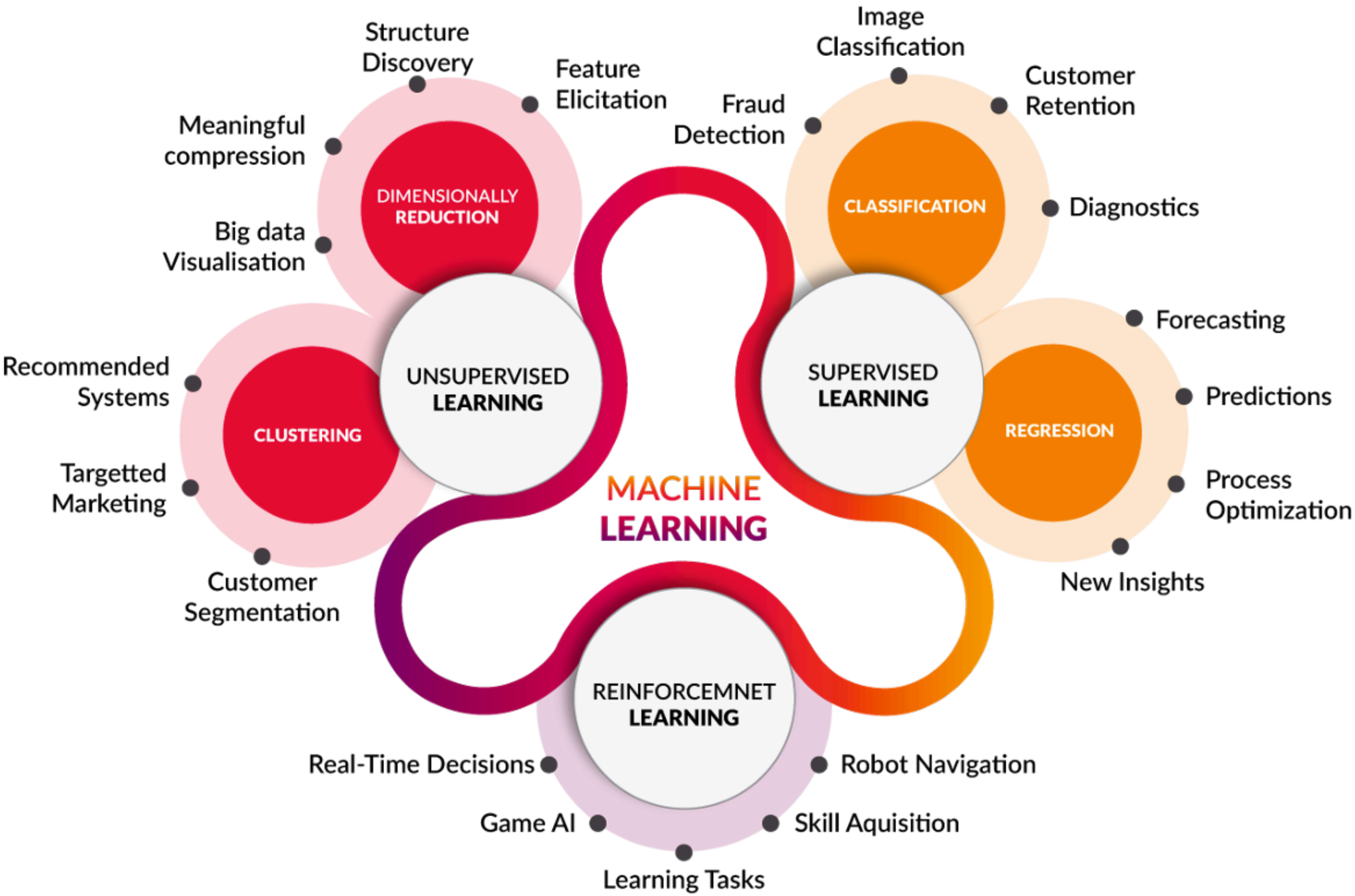
ML approach


Data
extraction
Data
preparation
Features
selection
Training
data
Learning
Algorithm
Learning
system
?
!

Learning types

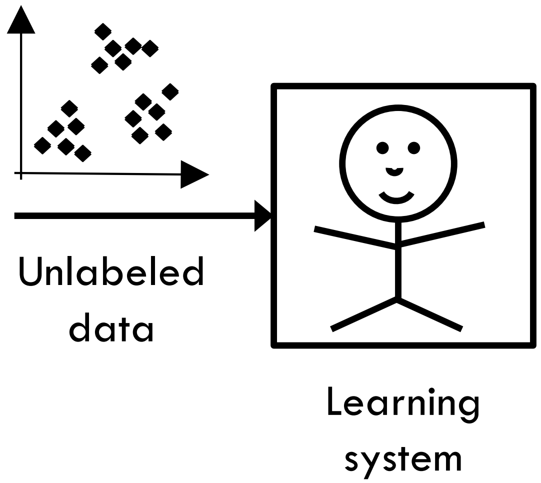
- According to the input/output relation:
  - Supervised learning
  - Unsupervised learning
  - Semi-supervised learning
  - Reinforcement learning
- According to the data/model relation:
  - Inductive learning
  - Deductive learning
- According to the nature of the algorithms:
  - Evolutionary Learning
  - Deep learning
  - Deep reinforcement learning
  - ...

Learning paradigms



Unsupervised Learning

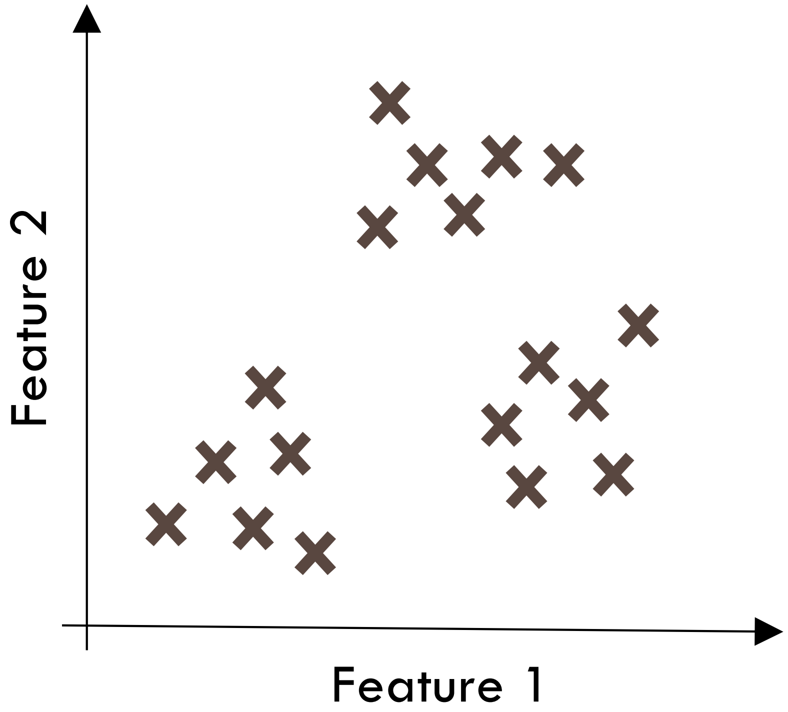
- The learning has **no teacher.** Learning algorithms try to identify relations between data with no additional information
- Examples in **training set** have no labels/targets



Unsupervised Learning

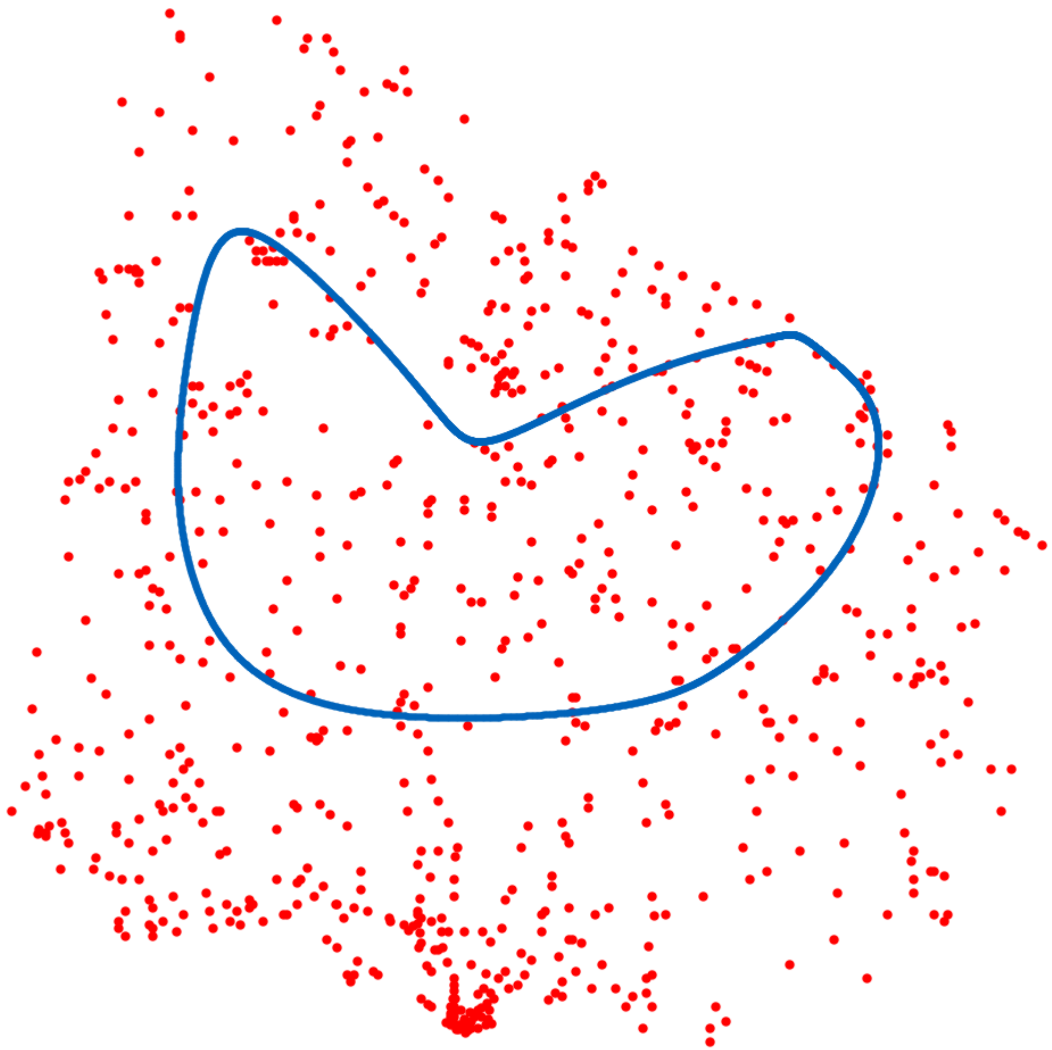
$$X = \left\{ x_i \right\}_{i=1}^{N} = \left\{ x_{i1}, \ldots, x_{id} \right\}_{i=1}^{N}$$

- *X*: training set
- *x*: training example
- *N*: number of training examples of the training set
- *d*: number of dimensions (attributes or features) of the examples



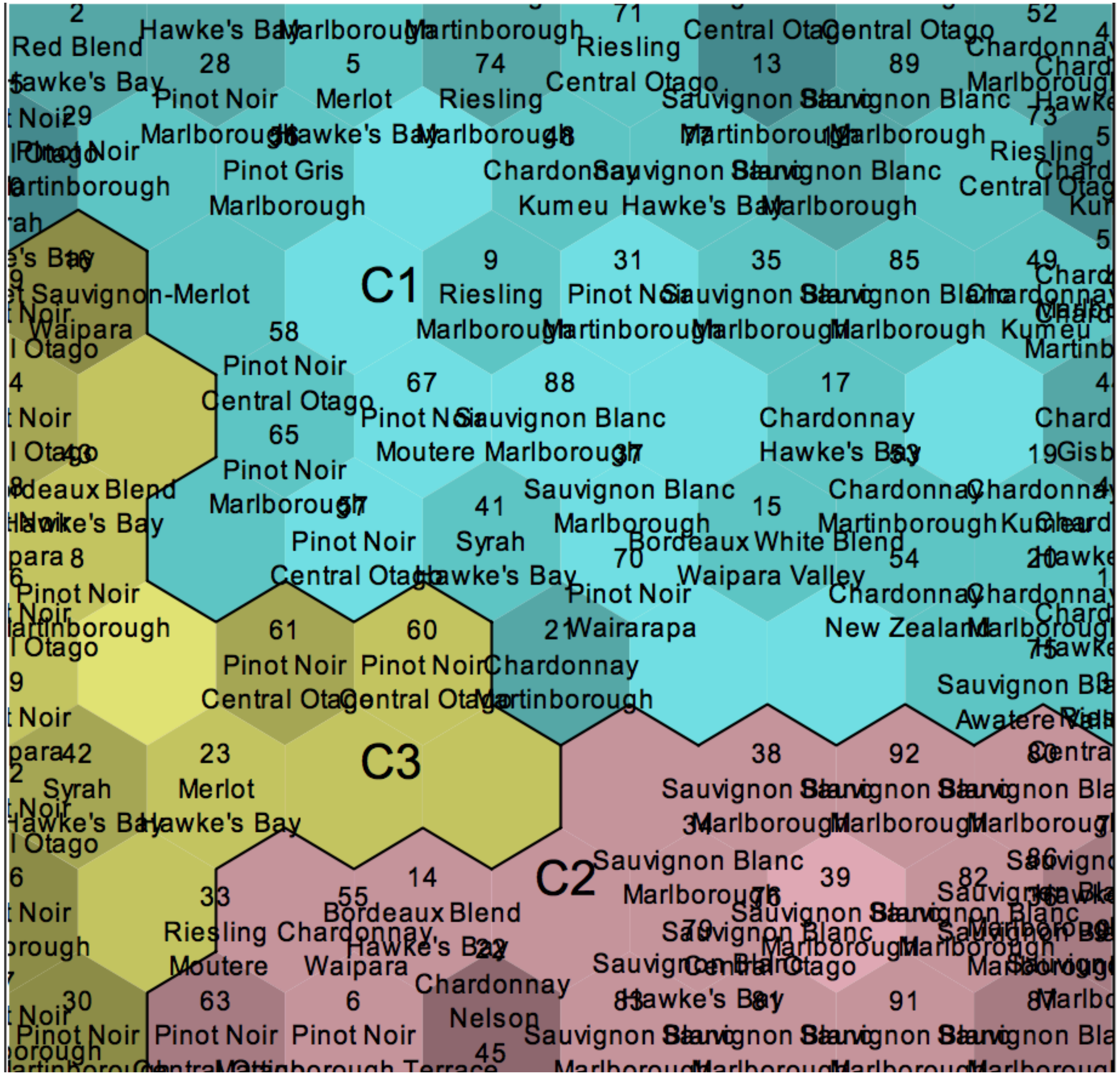
Unsupervised Learning

SOM to solve TSP

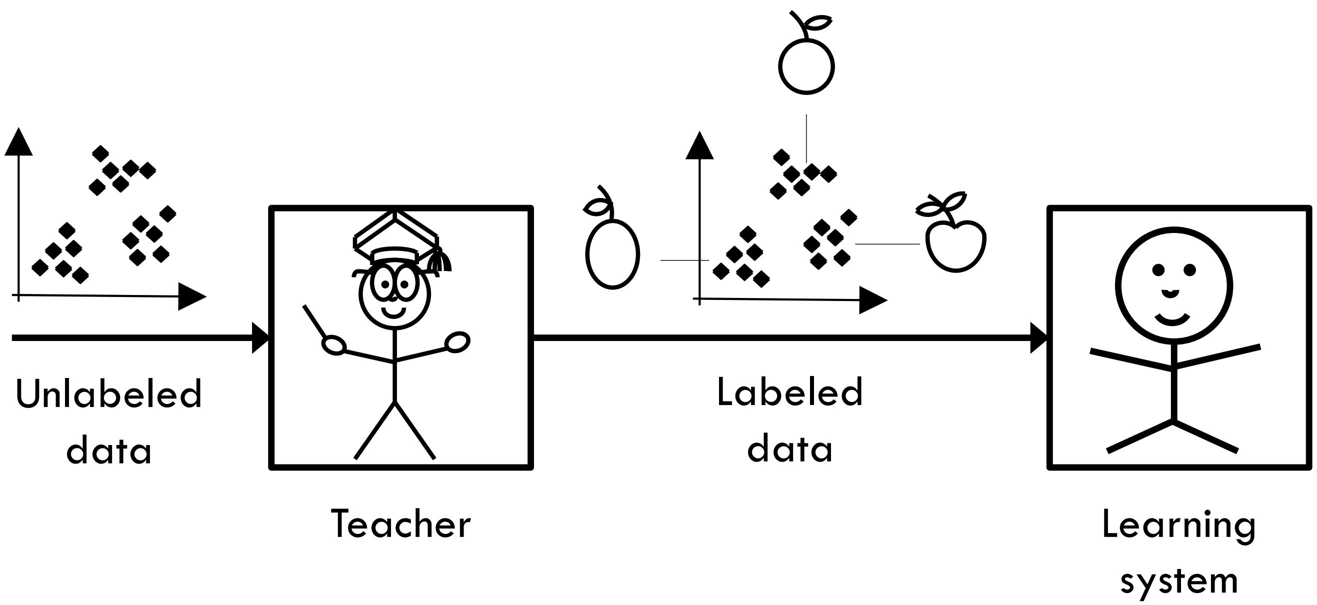


Unsupervised Learning

Kohonen self-organising maps in the data mining of wine taster comments. P. Sallis1 , S. Shanmuganathan1 , L. Pavesi & M. C. J. Muñoz. Auckland University of Technology, Australia 2 Universidad Catolica del Maule, Chile

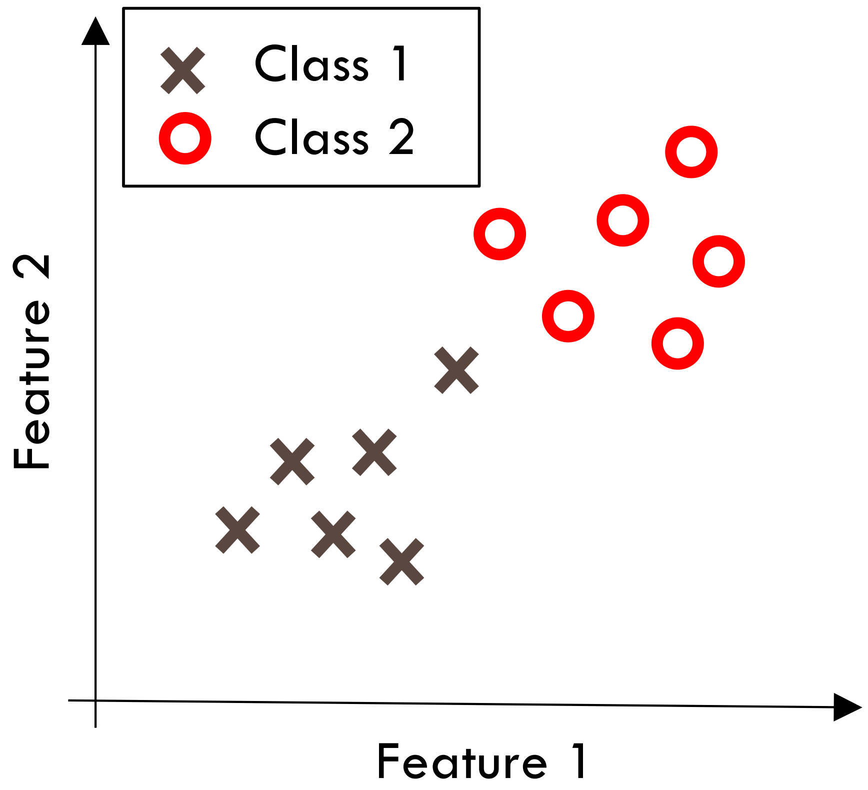


- The learning has a **teacher**, that is, someone that assigns the correct labels to the training examples
- The examples in **training set** have correct responses (targets or labels)



$$X = \{x_i, t_i\}_{i=1}^{N} = \{x_{i1}, \ldots, x_{id}, t_i\}_{i=1}^{N}$$

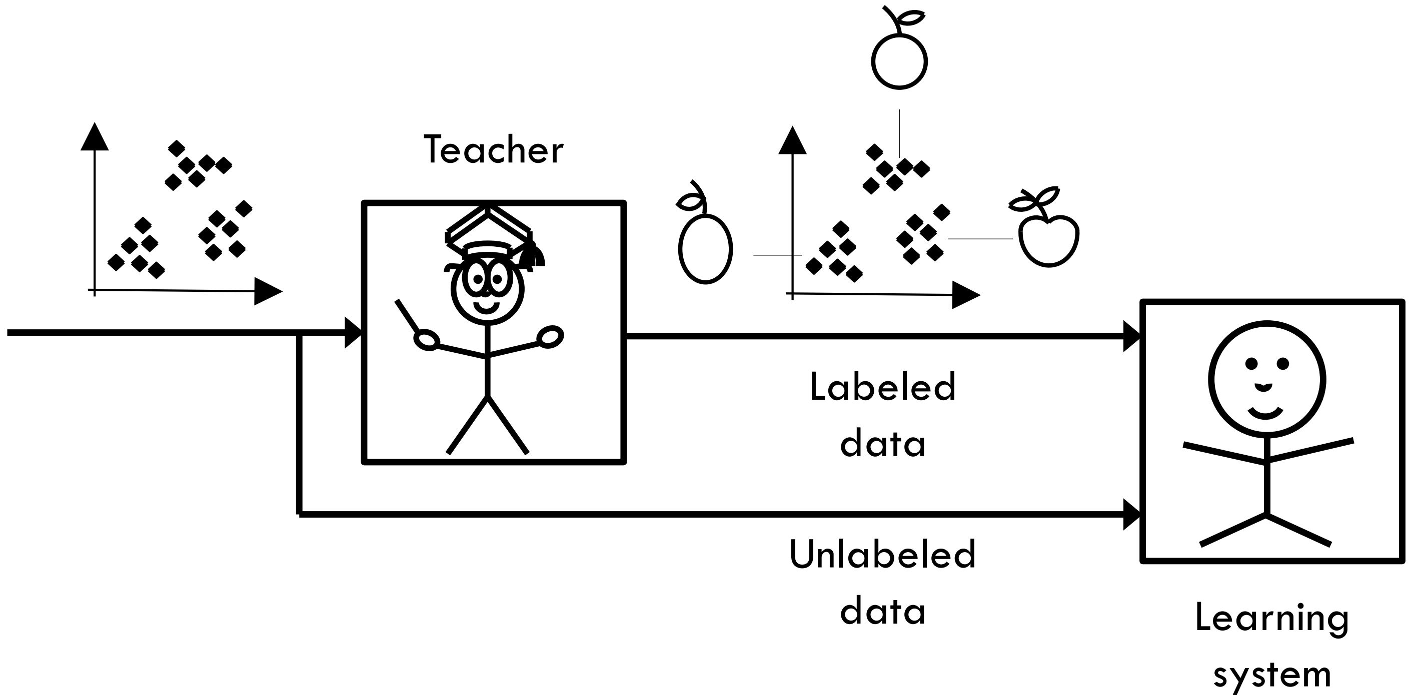
- *X*: training set
- x: training example
- *N*: number of training examples of the training set
- *d*: number of dimensions (attributes or features) of the training examples
- *t*: target or label, that relates the training example to one of the classes available




WOMBO.AI

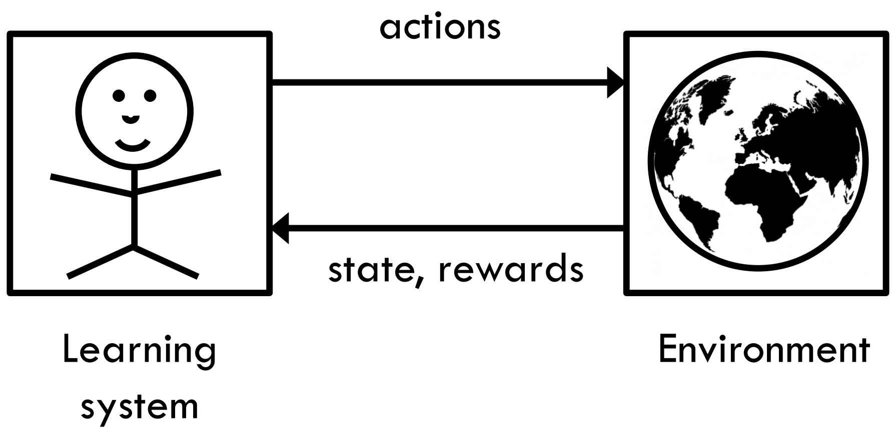
Semi-supervised Learning

- It is a framework of algorithms proposed to improve the performance of supervised algorithms through the use of both **labeled** and **unlabeled** data



Reinforcement Learning (RL)

- The learning process is based on **trial and error**
- The learning system chooses **actions** and receives **feedbacks** from the environment (rewards/punishments) that tells it how good is the system performance
- There is **no training set**

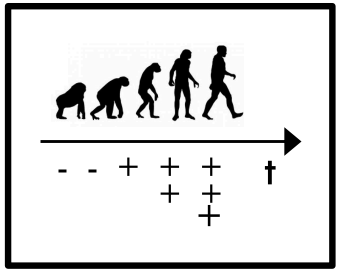


Reinforcement Learning Example

Chelsea Finn, Sergey Levine, Pieter Abbeel

Evolutionary Learning

- The learning process is inspired on the **evolution** of biological organisms
- The learning system has internal models of a **population** that successively generates descendants with some **mutations** in order to stimulate the development of individuals with better chance to achieve a successful **performance** in the environment
- There is **no training set**

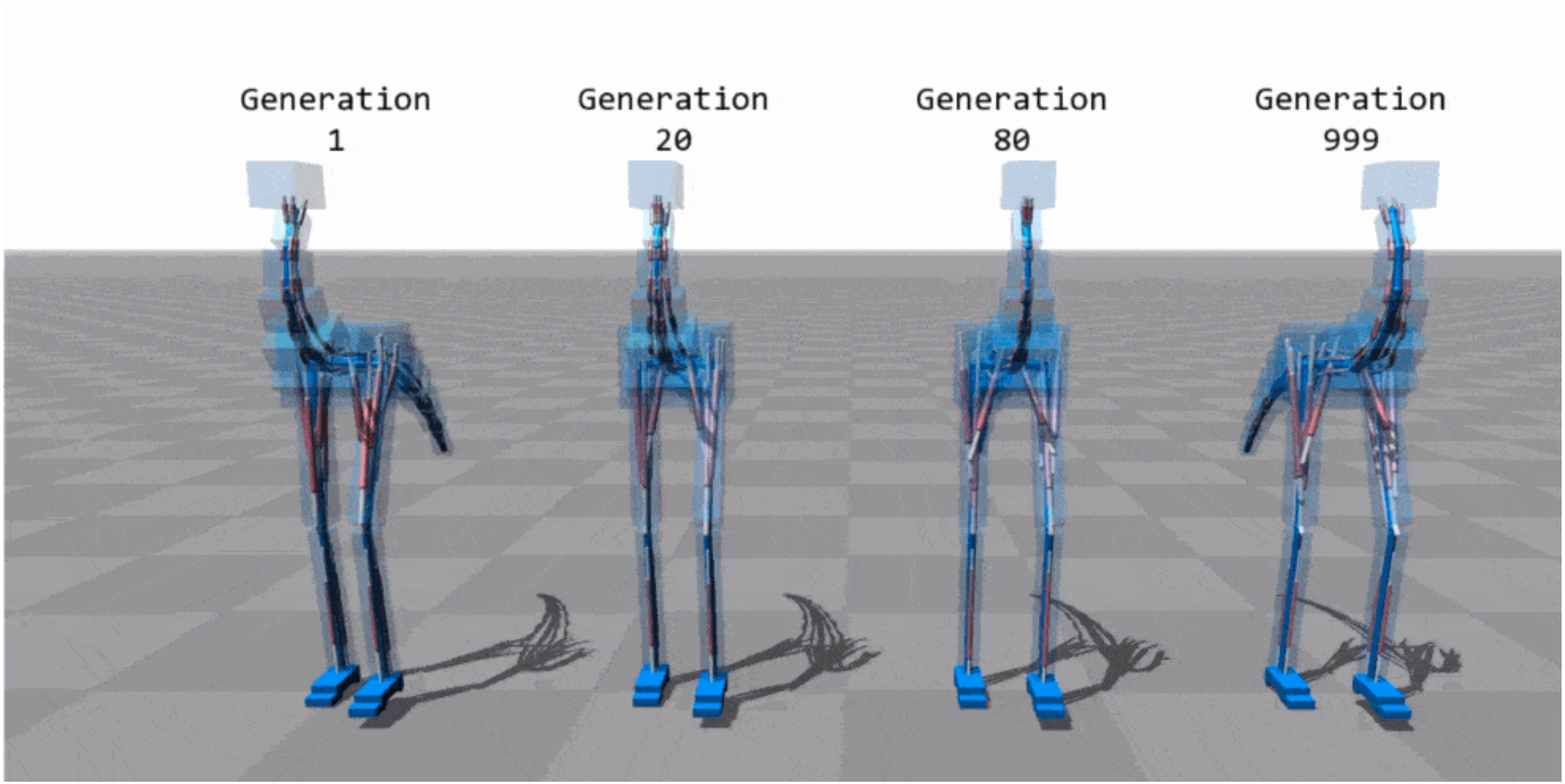


Learning system

Evolutionary Learning Example

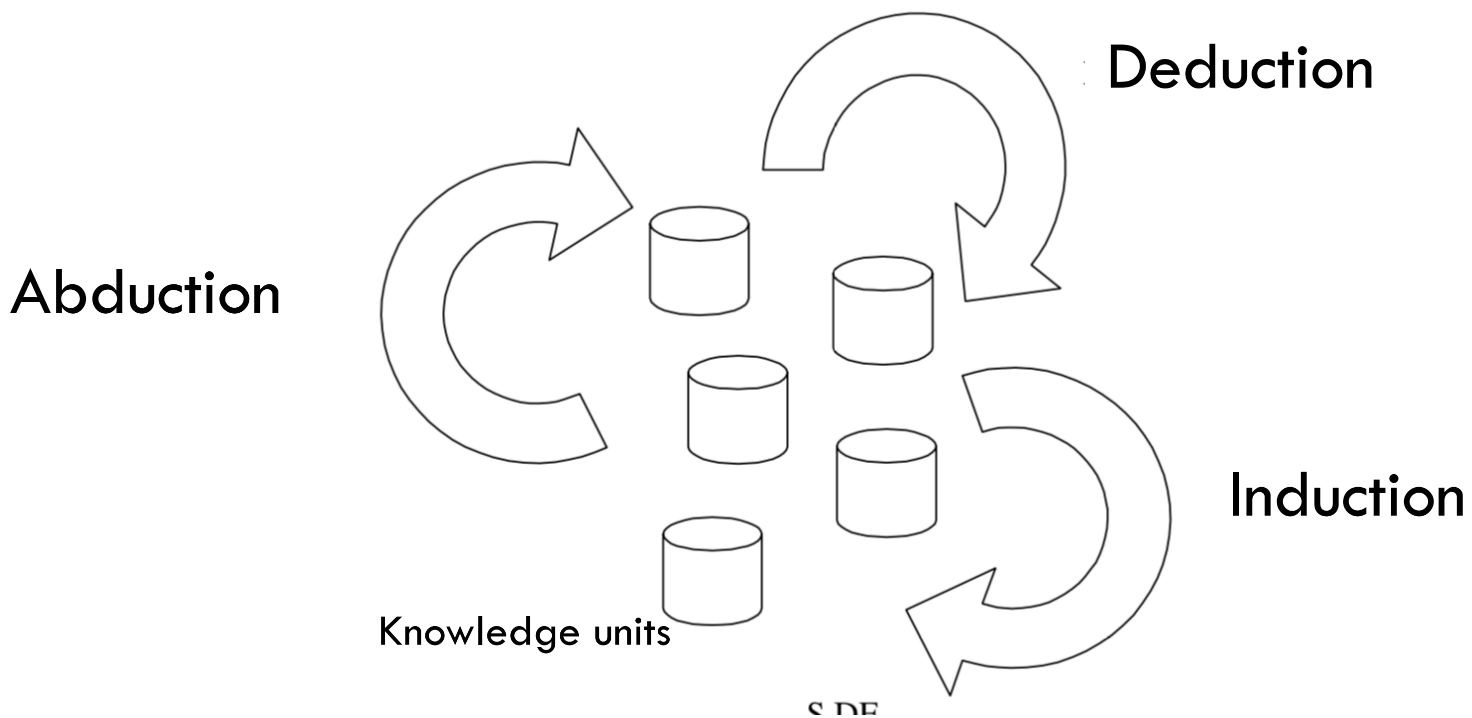
https://dzone.com/articles/feature-selection-using-genetic-algorithms-in-r

- Feature Selection using GA



Inductive x Deductive Learning

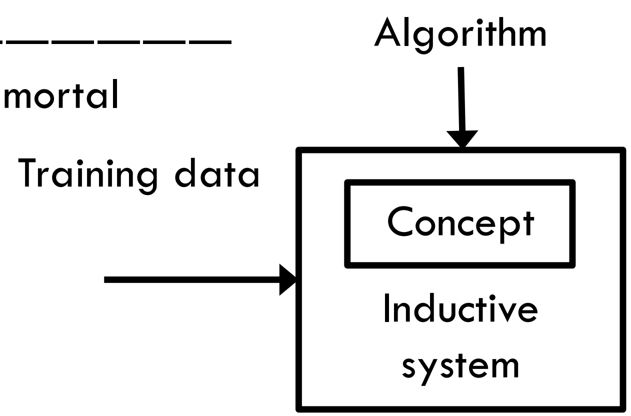
- Reasoning is usually performed in three ways:
  - **Abductive reasoning deals with guesswork,** involves reasoning toward possible conclusions based on guesswork (a best guess)
    - It is a type of reasoning that is used in formulating a hypothesis for further testing

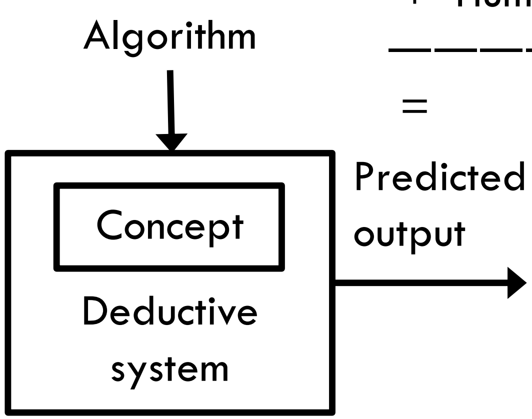


Inductive x Deductive Learning

- **Inductive learning:**
  - System obtains new knowledge (refines its model/concept) from specific information/data
  - New data can change the knowledge
  - Is the reasoning which, after considering a sufficient number of particular cases, concludes a general truth
  - **Inductive reasoning deals with probability** and involves reasoning toward likely conclusions based on data
- **Deductive learning:**
  - System obtains new knowledge (refines its model/concept) typically using logic
  - Uses assumptions arguments to obtain a conclusion. The conclusion makes explicit an already existing knowledge in the premises
  - The knowledge is not new (it is implicit in initial knowledge)
  - **Deductive reasoning deals with certainty** and involves reasoning toward certain conclusions

Socrates is human
\+ ?
———————————————
= Socrates is mortal





Socrates is human
\+ Humans are mortal
———————————————
= ?

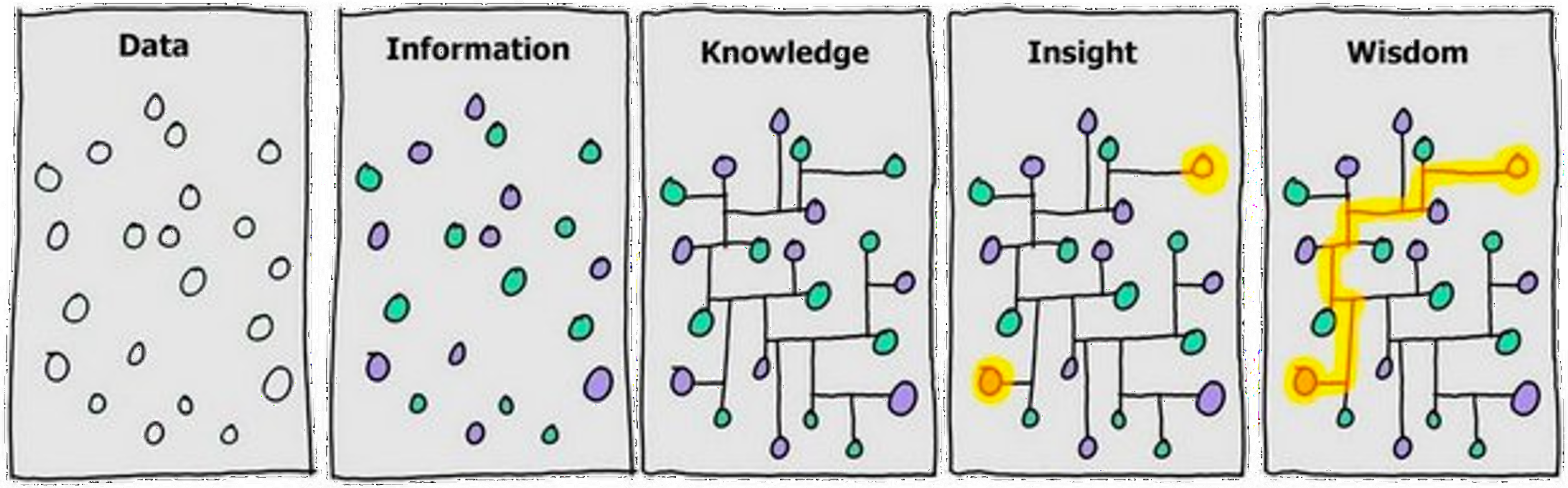
Inductive x Deductive Learning

- In Summary:
  - **Abduction** is forming a hypothesis
    - **taking your best shot**
  - **Induction** is like analyzing the data from testing a hypothesis
    - **conclusion merely likely**
  - **Deduction** would be used in drawing certain logical conclusions from the data gathered
    - **conclusion guaranteed**

- A learning approach that enables the computer to learn **complicated concepts** by organizing them **hierarchically,** in a way that, if we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers (Goodfellow et al, 2016)

- **Deep Reinforcement Learning (DRL)** is a method based on deep learning technique that extends the traditional Reinforcement Learning (RL) to the entire process from sensors to motors

Path from data to wisdom…



About (human) Learning

Learning is a constructive process carried out by the subject and by no one else through **physical actions** (acting on the objects) and **mental actions** (reorganizing mental structures)

To learn, each subject will **mentally construct the objects of his environment** and this will happen through **his own action** and not through the action of the teacher or any other person

It can not be "deposited", "absorbed" or "awakened"

The **action of the subject** is the starting point

ML in real world applications

Extra-class activities

Lecture 1

- Activities
  - Tutorial 1
    - What is ML?
  - Reading:
    - What can machine learning do? Workforce implications. By Erik Brynjolfsson and Tom Mitchell
    - Other reference papers available in the website

References

Lecture 1

- MARSLAND, S. Machine Learning: an algorithm perspective. CRC Press, 2nd edition, 2015.
- ALPAYDIN, E. Introduction to Machine Learning. MIT Press, 3rd edition, 2014.
- SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. Redes Neurais Artificiais para engenharia e ciências aplicadas. Curso prático. Artliber Editora, 2010.
- SUTTON, R. S.; BARTO, A. G. Reinforcement learning: an introduction. MIT Press, Cambridge, MA, 2nd edition in progress, 2017.
- Notas de aula, MC886, Sandra Ávila.

This material is part of the Machine Learning Course

By Esther Colombini and Alexandre Simões