LINUX DAY MILANO

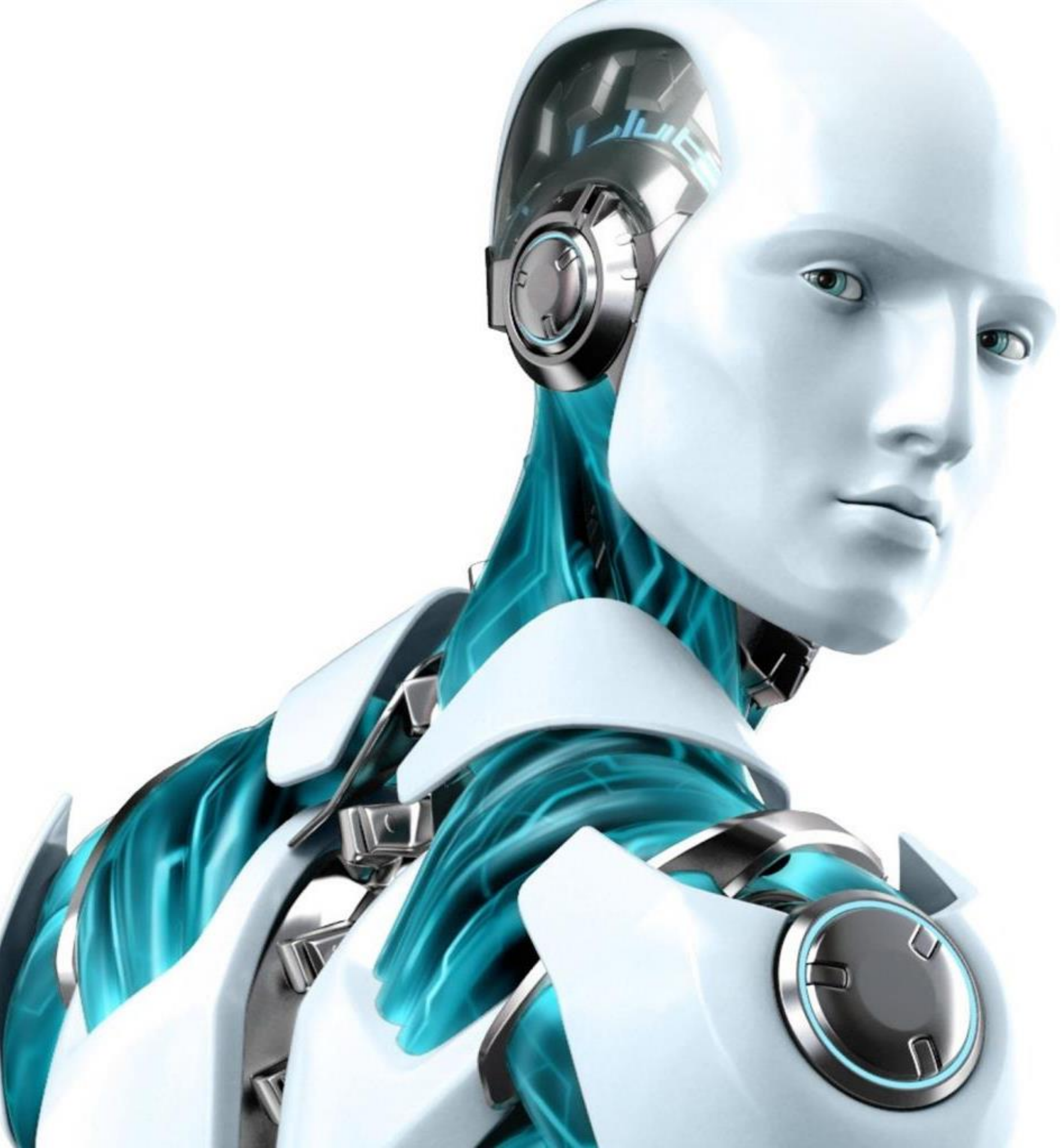
https://linuxdaymilano.org/#sponsors

26 Ottobre 2019

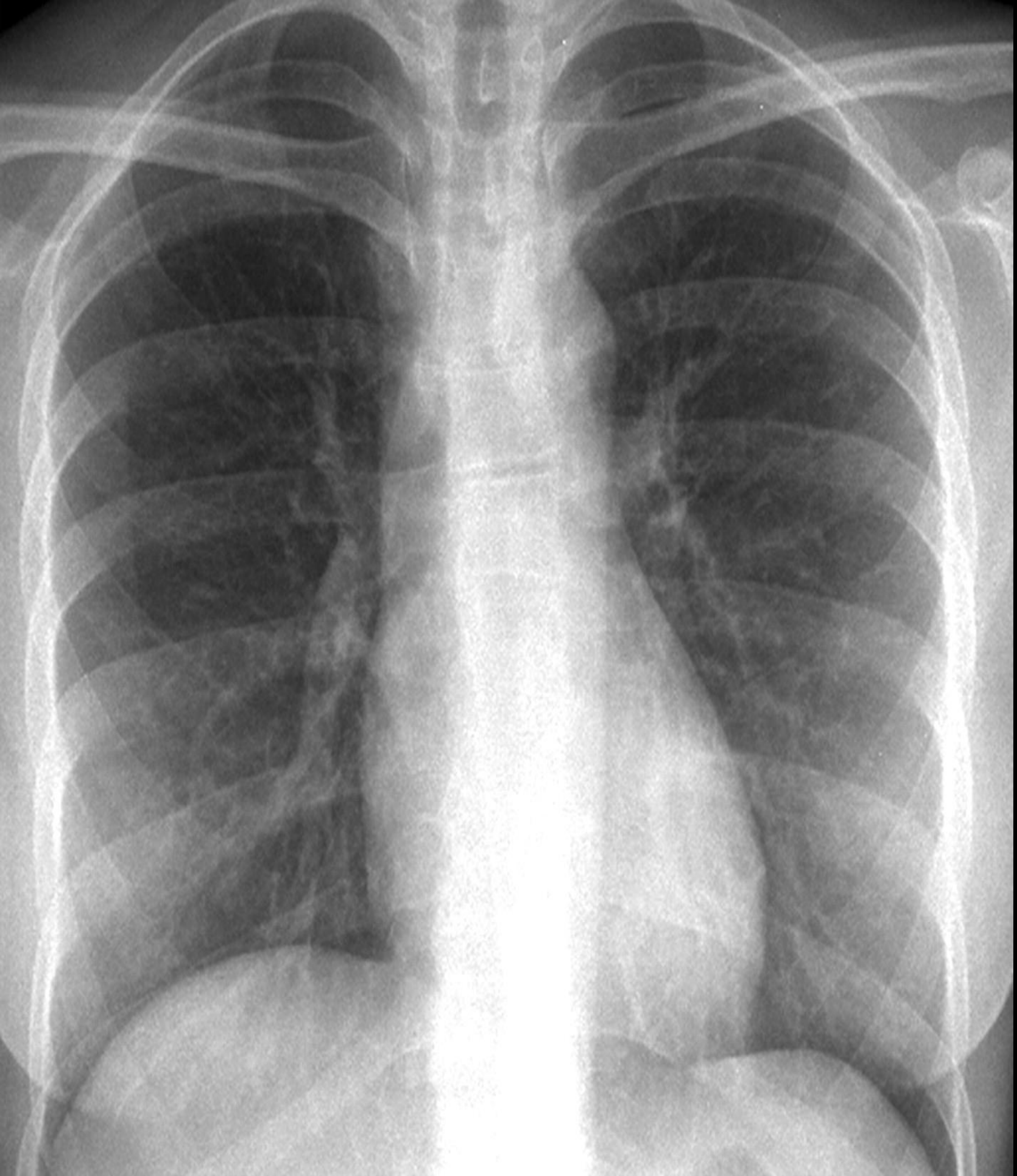
Edificio U7

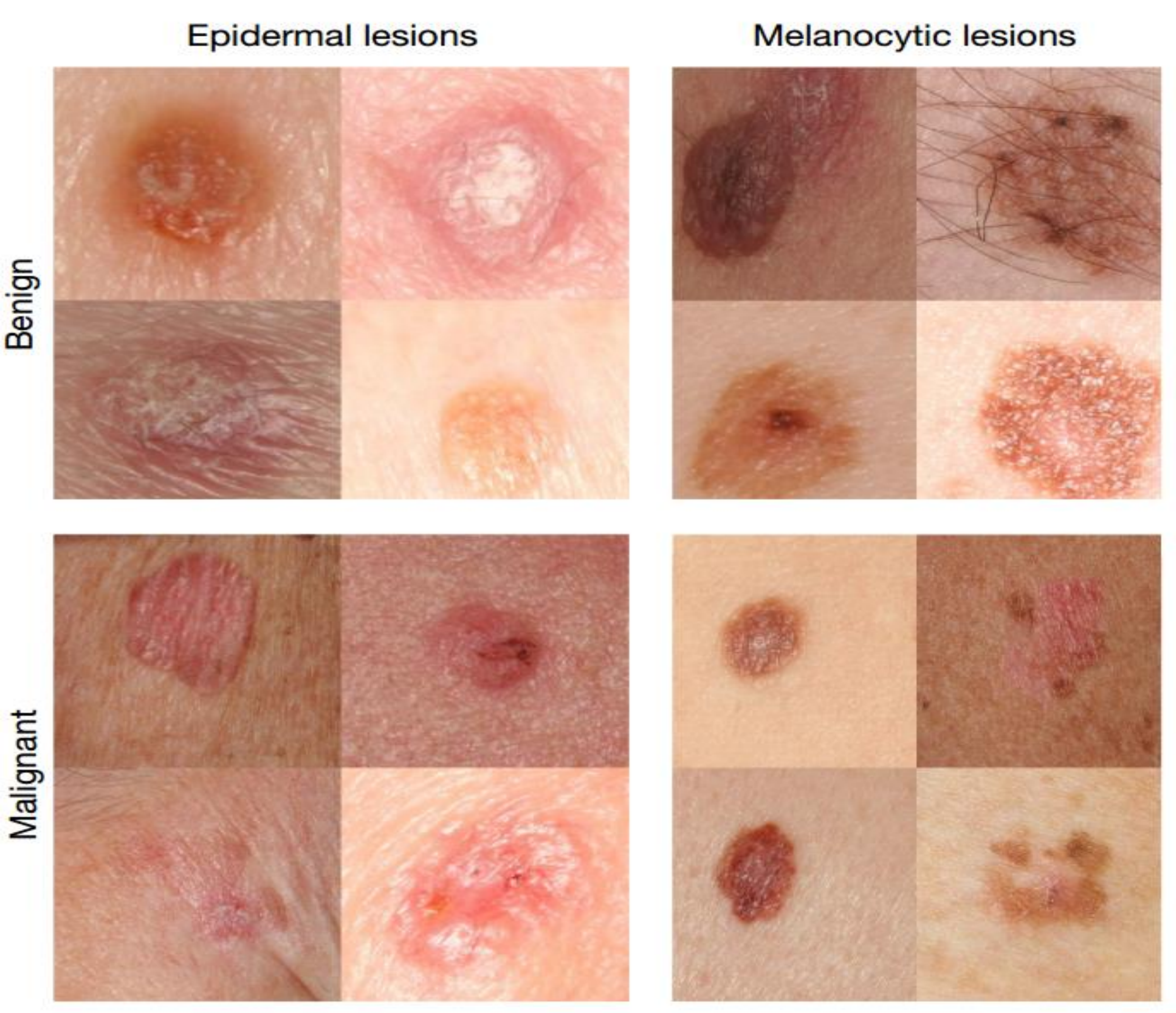
Intelligenza a livello umano o abilità a livello animale?

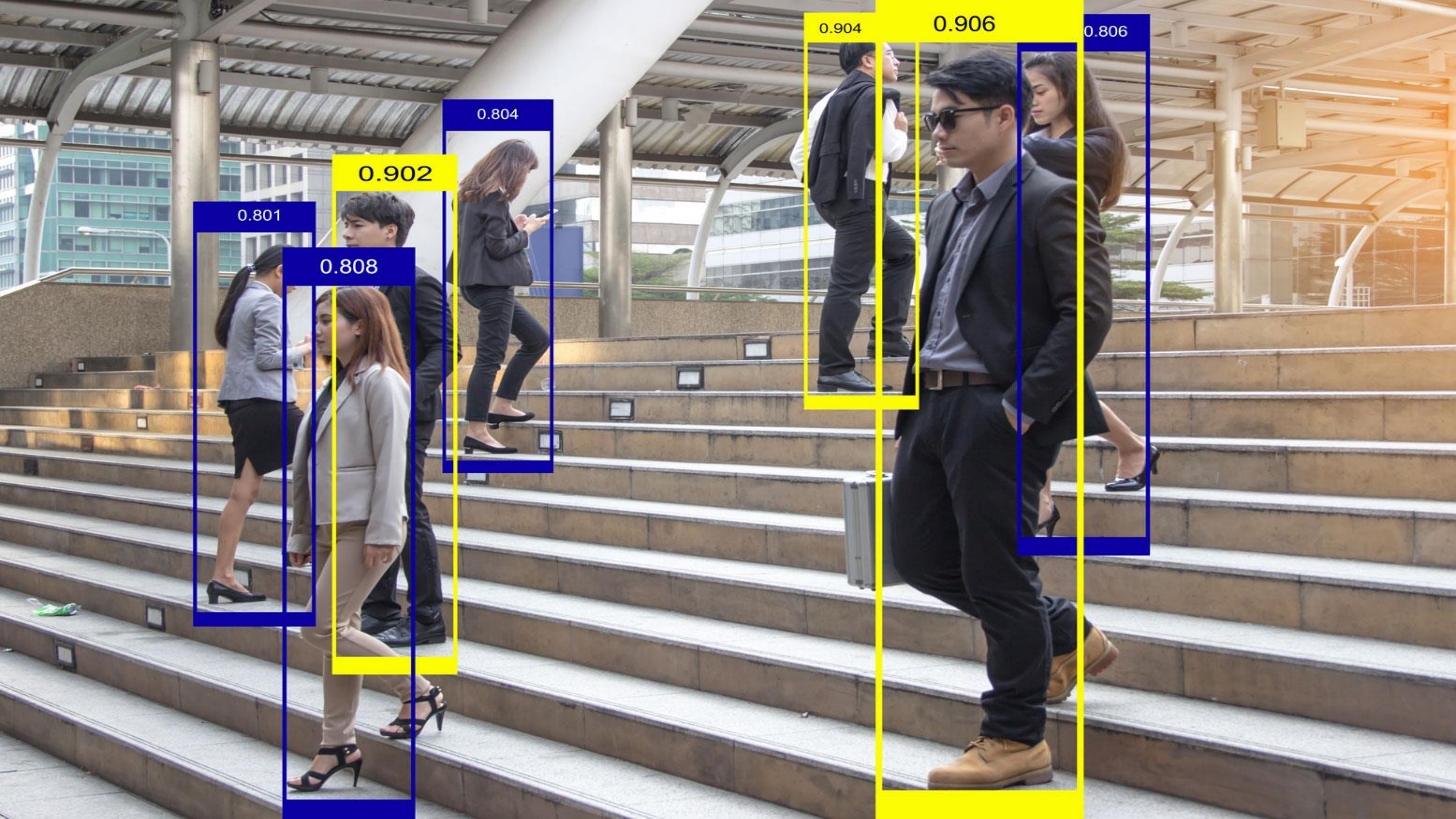
Fabio Stella
Università di Milano-Bicocca
Dipartimento di Informatica, Sistemistica e Comunicazione



«L'**intelligenza artificiale** o **IA** è una disciplina appartenente all’**informatica** che studia i fondamenti teorici, le metodologie e le tecniche che consentono la progettazione di **sistemi hardware** e sistemi di programmi **software** capaci di fornire all’**elaboratore elettronico** prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell’**intelligenza umana**.»




Epidermal lesions
Melanocytic lesions
Benign
Malignant

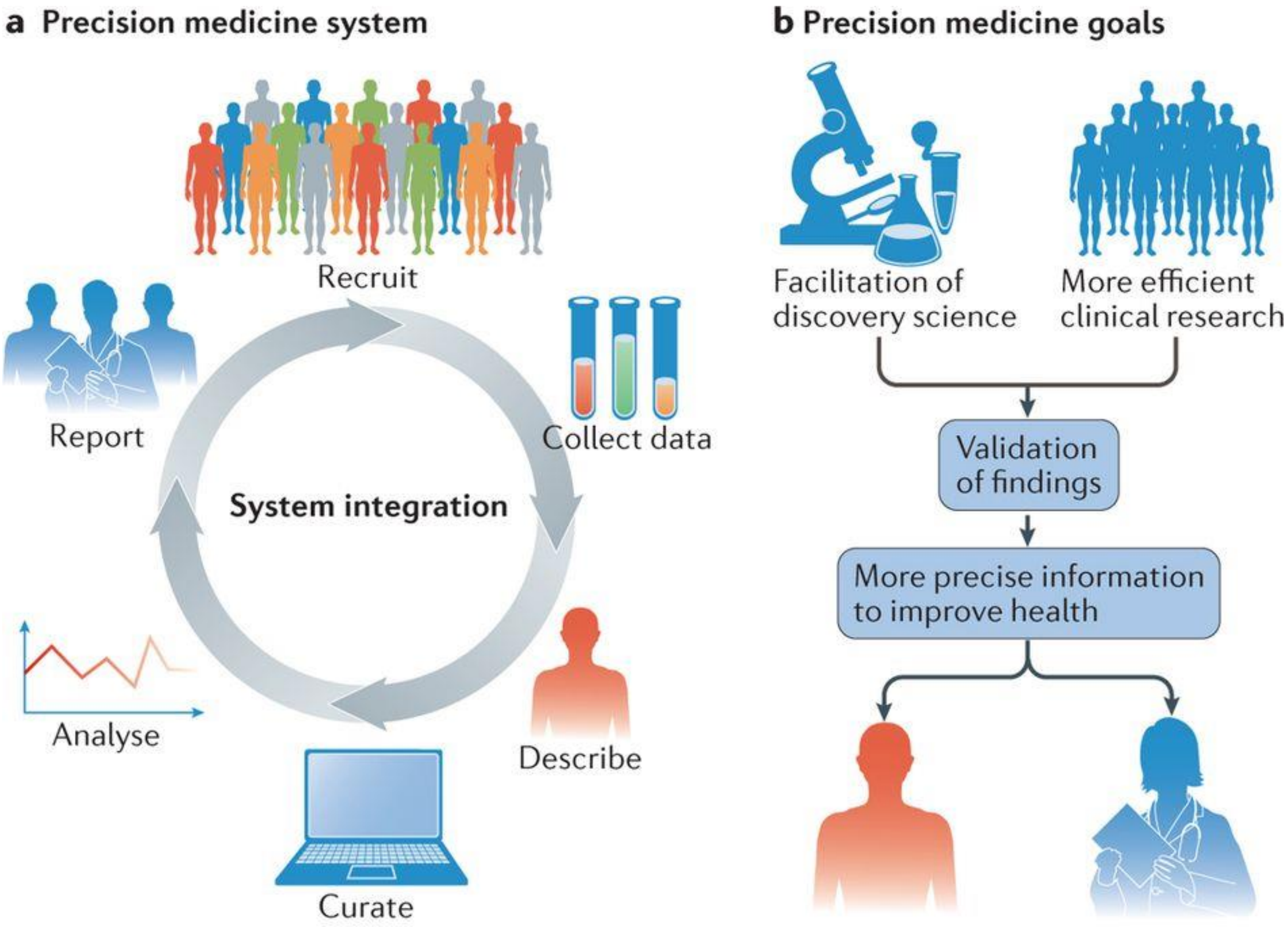

0.904
0.906
0.806
0.804
0.902
0.801
0.808

Il progetto 100,000 Genomi

- Sequenziamento completo di 70,000 individui
- 21 PetaBytes di dati
- 1 PetaByte di musica richiede circa 2,000 anni per essere ascoltato integralmente

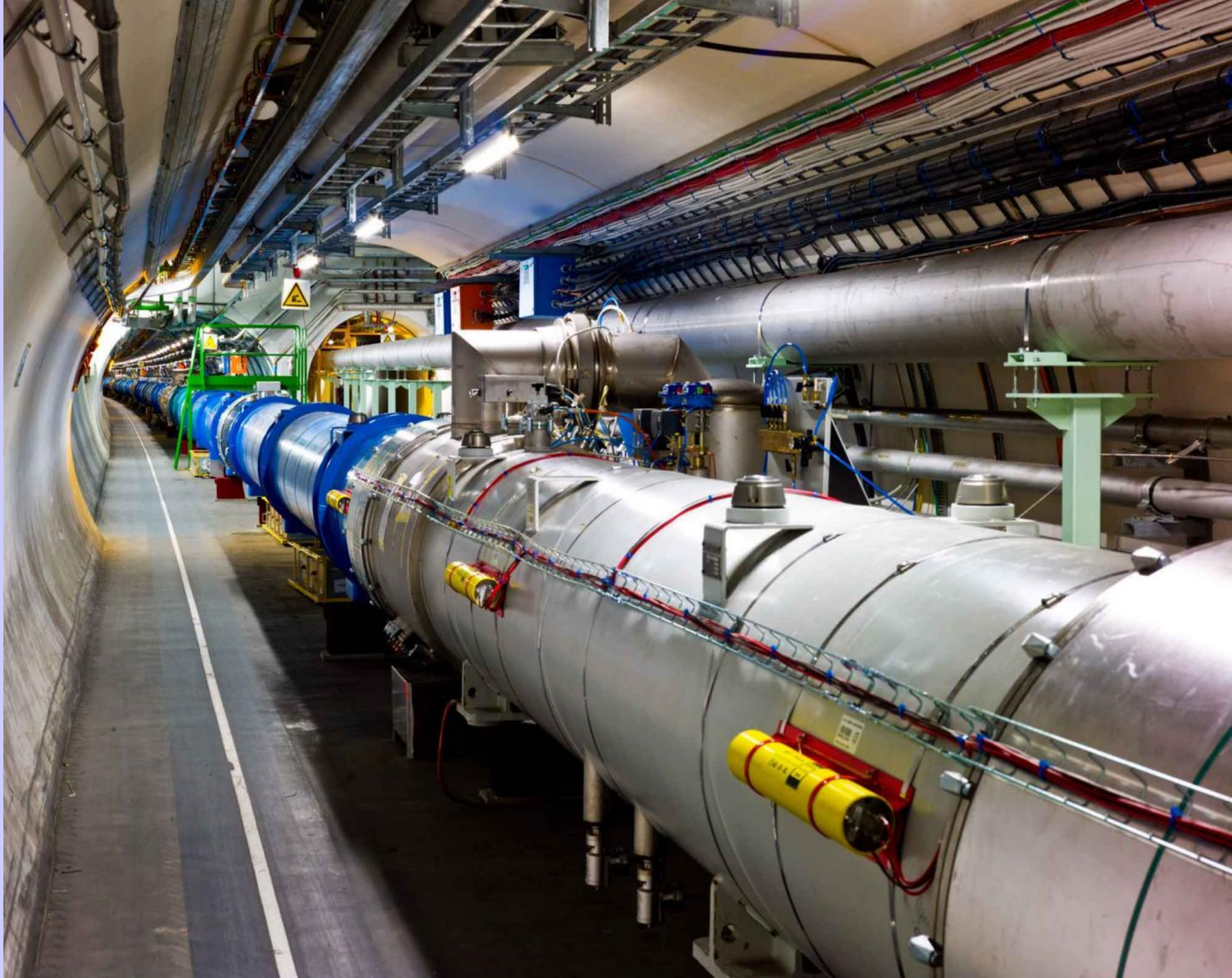
La complessità aumenta ulteriormente (esplode) se in aggiunta consideriamo

- Cartella clinica elettronica
- Stile di vita
- Dieta
- Tipo di occupazione
- Informazioni geografiche
- Rete di interazione



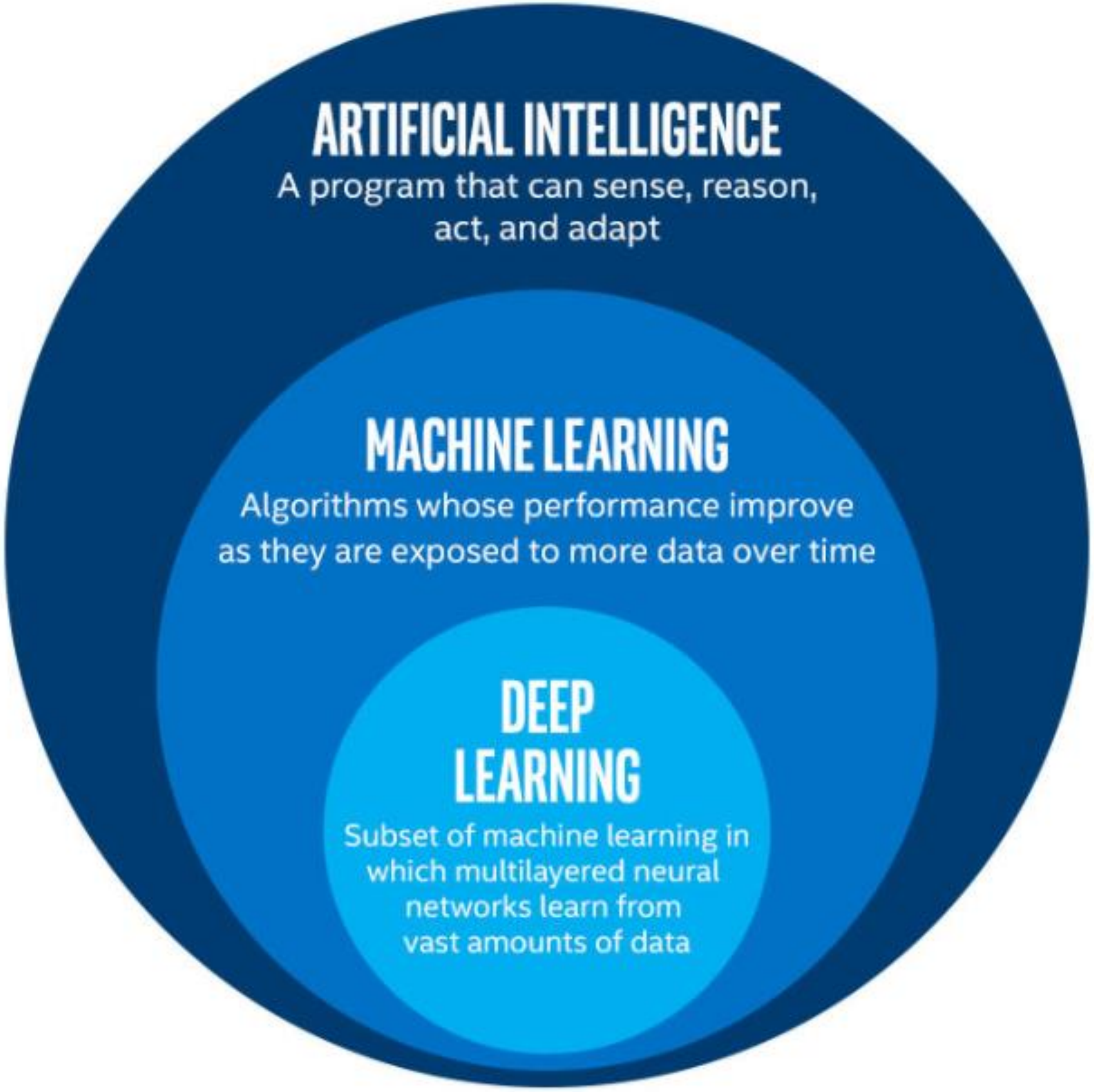
Large Hadron Collider

- 2017 June 29 at CERN, 200 PB of data permanently stored.
- 1 billion collisions per second happening at, ATLAS, CMS, ALICE, and LHCb.



Artificial Intelligence; intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans.

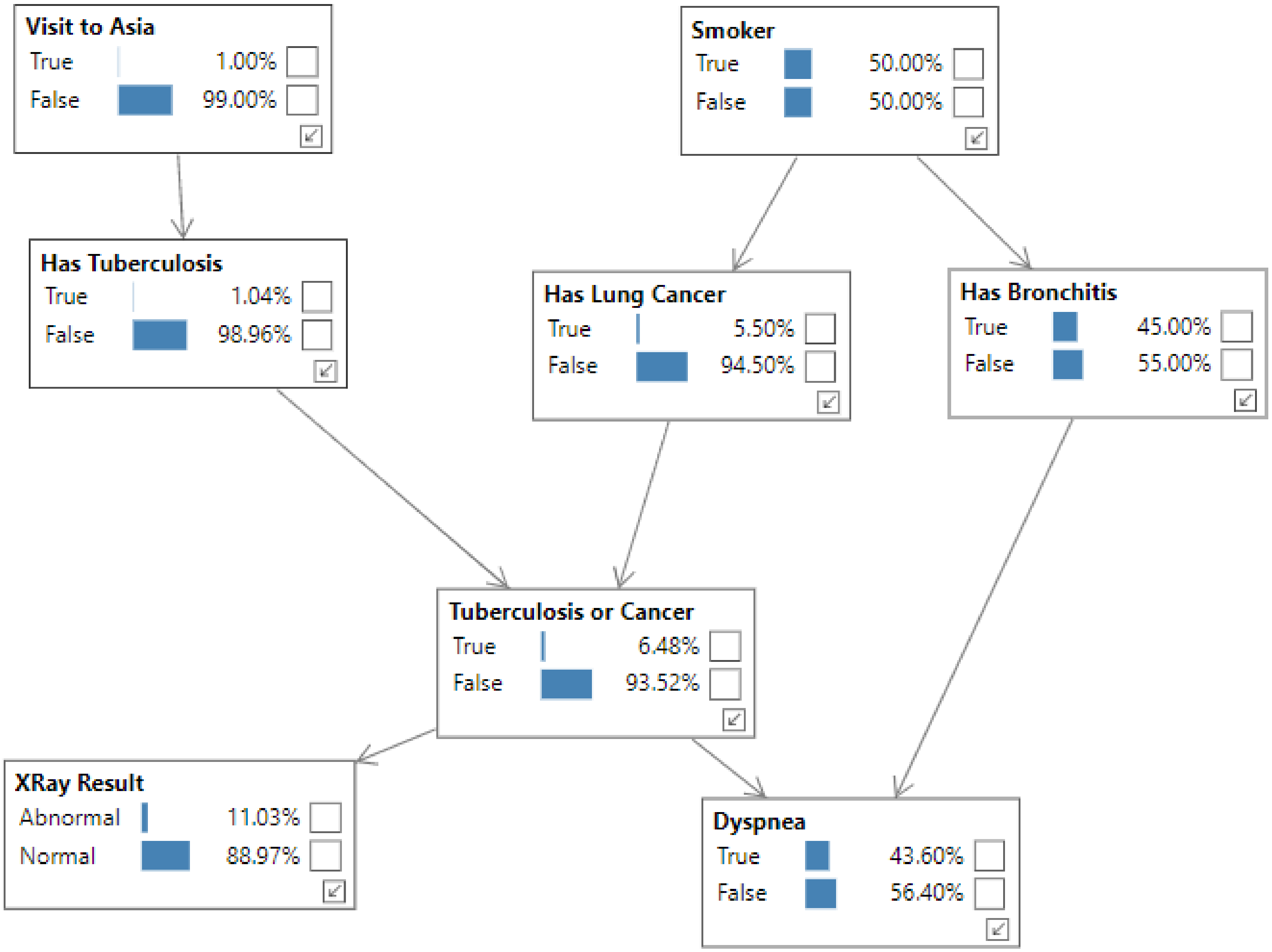
- Colloquially, the term Artificial Intelligence (**AI**) is used to describe machines/computers that mimic “cognitive” functions that humans associate with other human minds, such as "learning" and "problem solving".
- Two kinds of AI:
  - ✓ Weak
  - ✓ Strong



Artificial Intelligence

Bayesian Networks

- A type of statistical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).
- Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes was the contributing factor.
- Structural Causal Models.

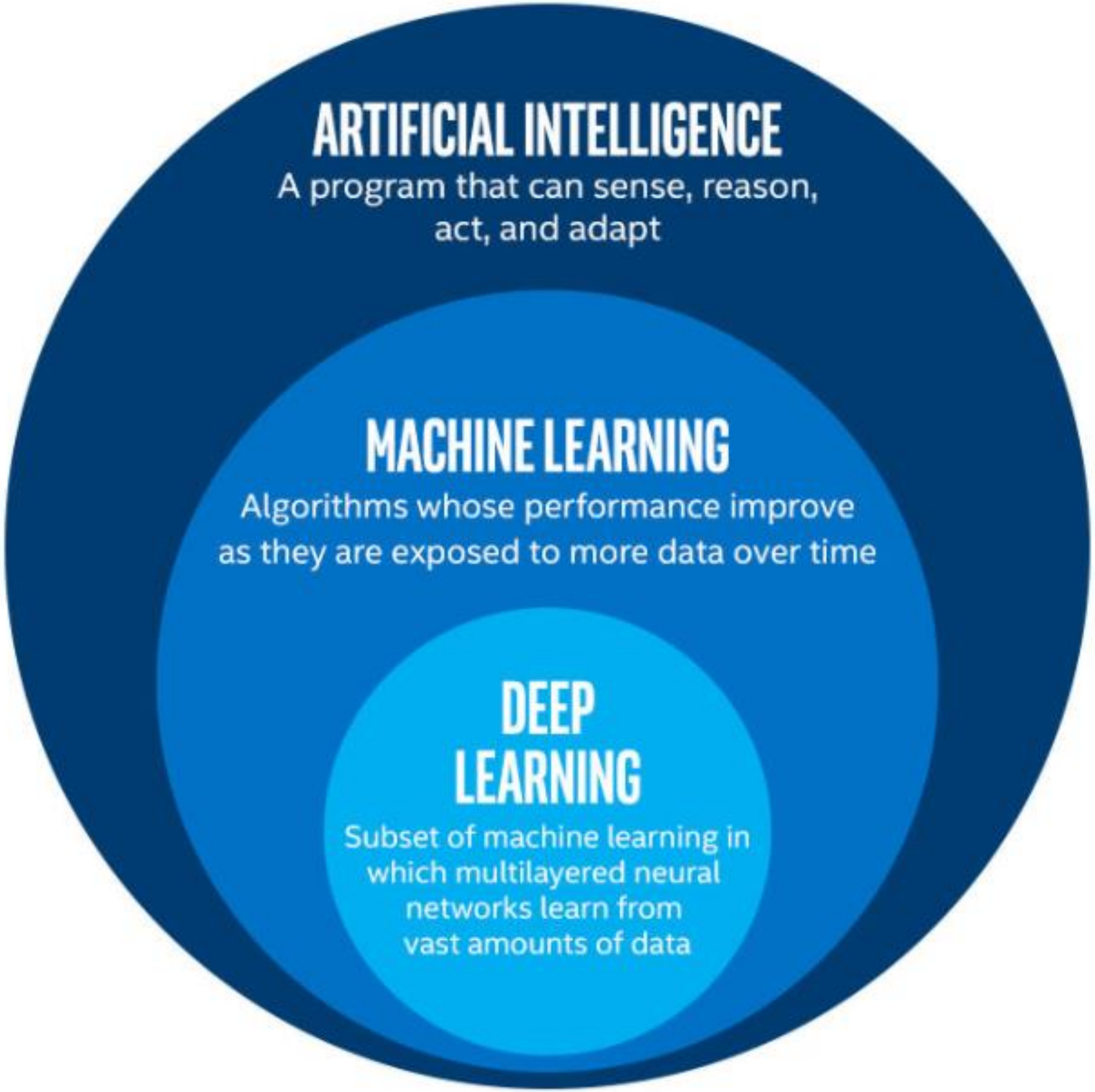


Machine Learning;

algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead.

Three kinds of ML;

- Supervised
- Self-Supervised
- Reinforcement Learning



Machine Learning

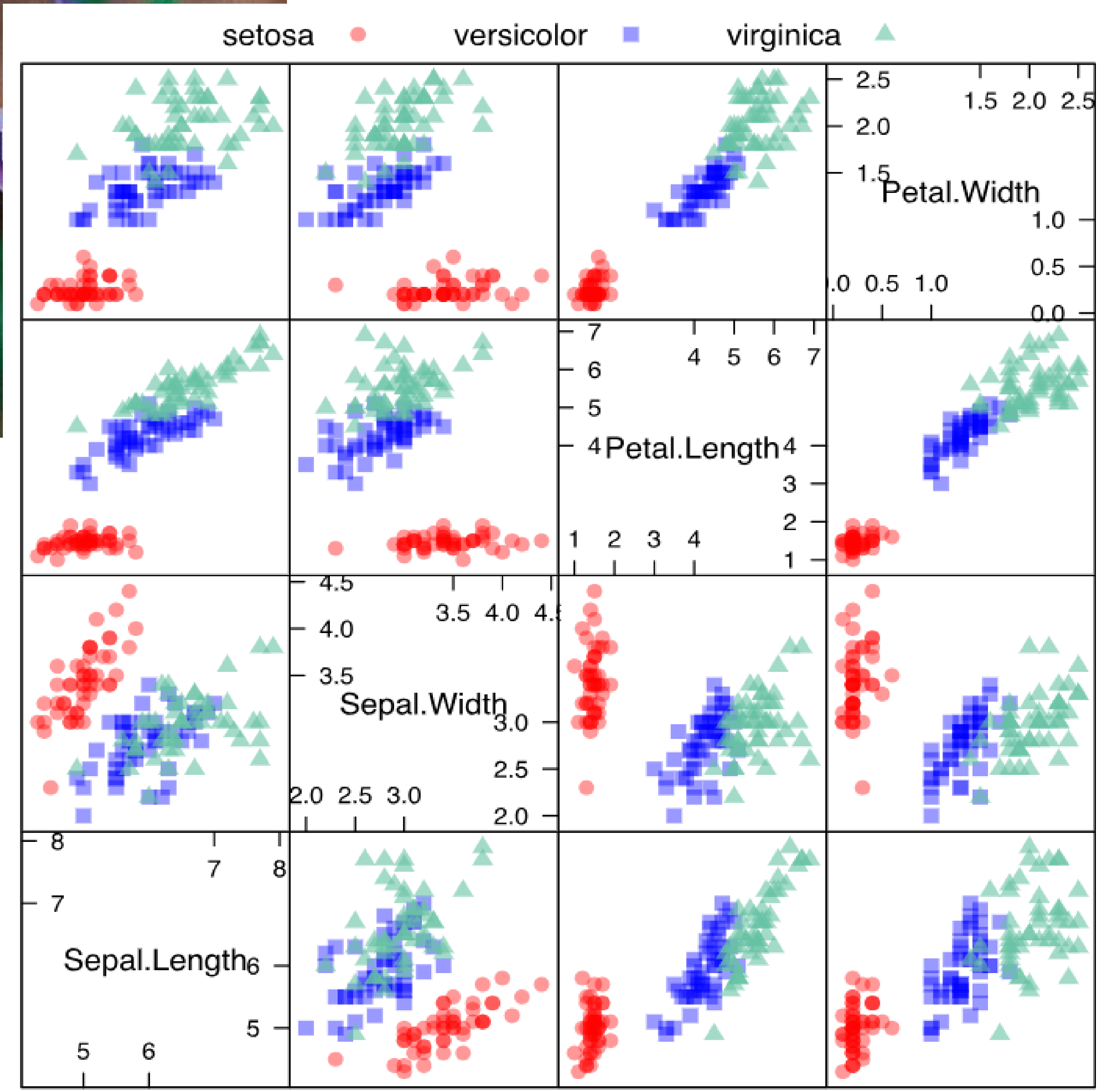
Supervised

- Classification

SETOSA

VERSICOLOR

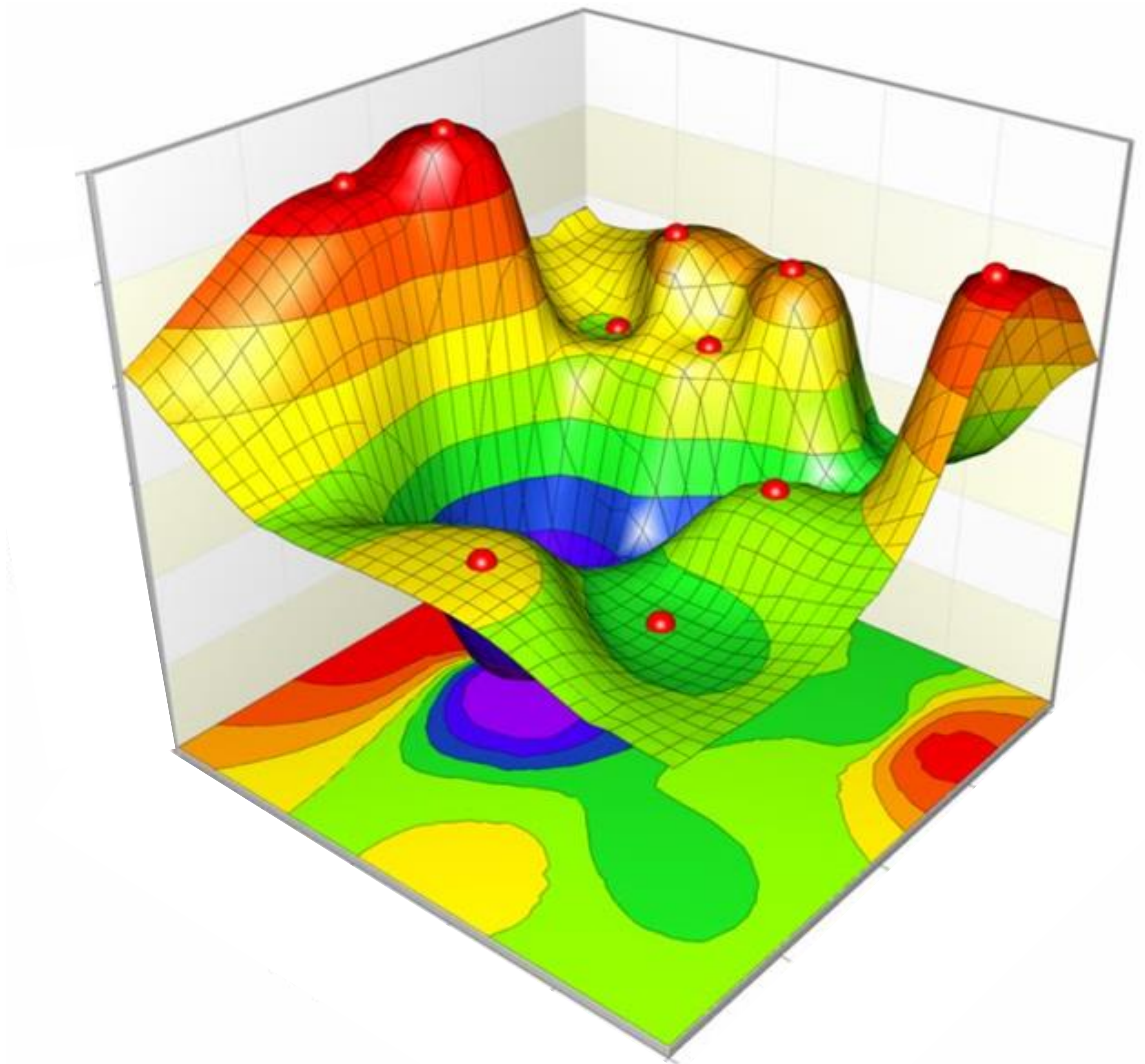
VIRGINICA



Machine Learning

Supervised

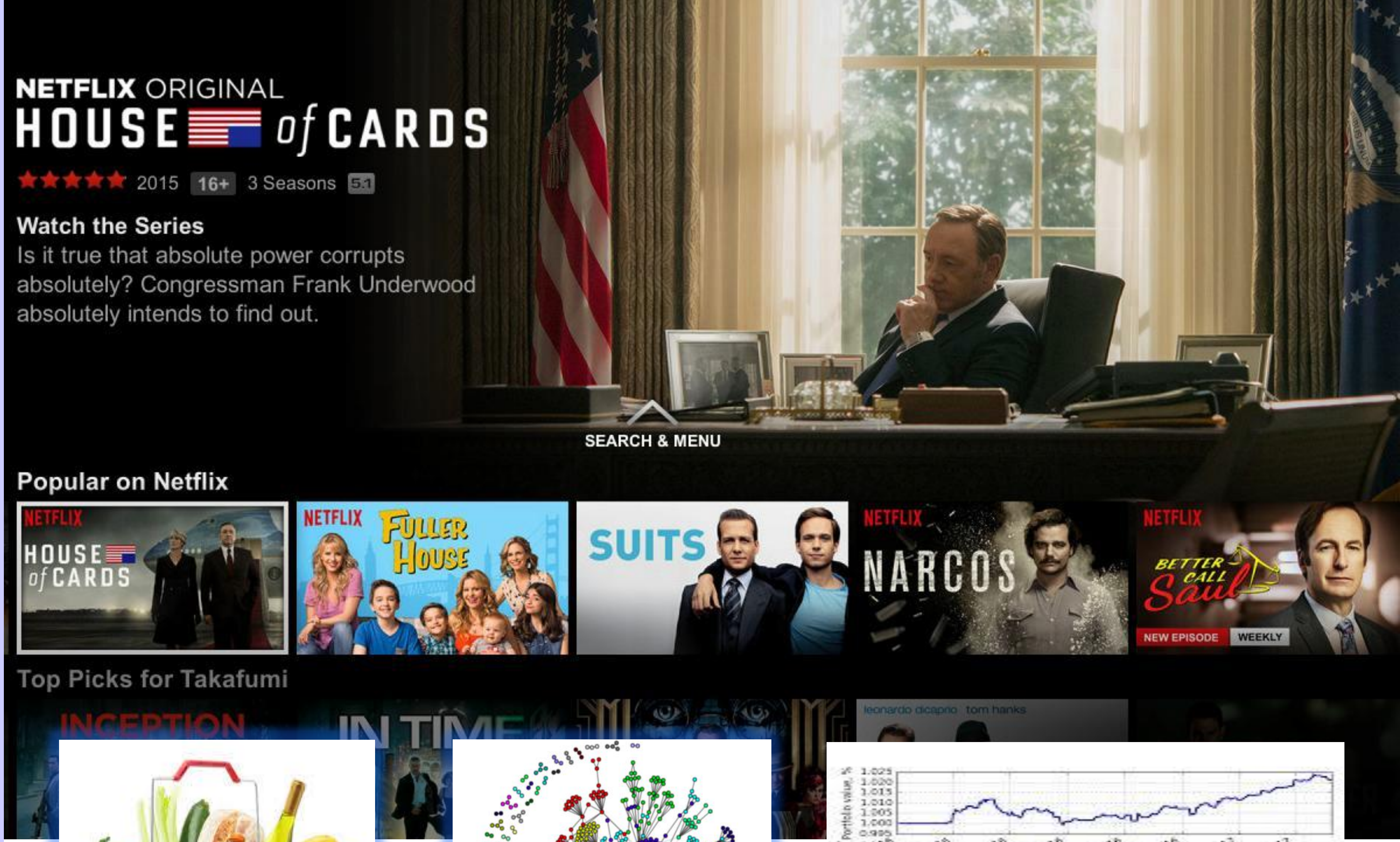
- Curve fitting
- Surface fitting

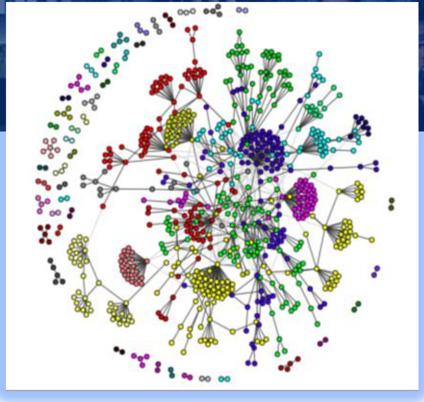


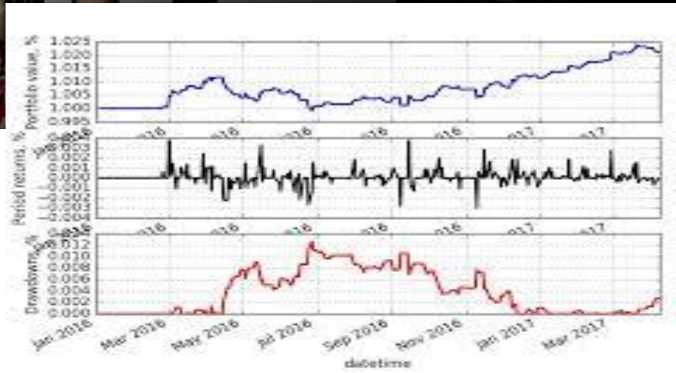
Machine Learning

Self-Supervised

- Recommendation System
- Market Basket Analysis
- Social Network Analysis







Reinforcement Learning: AlphaGo

Mosse Go 10^{761}

Mosse scacchi 10^{123}

Atomi universo 10^{81}

Machine Learning

Reinforcement Learning

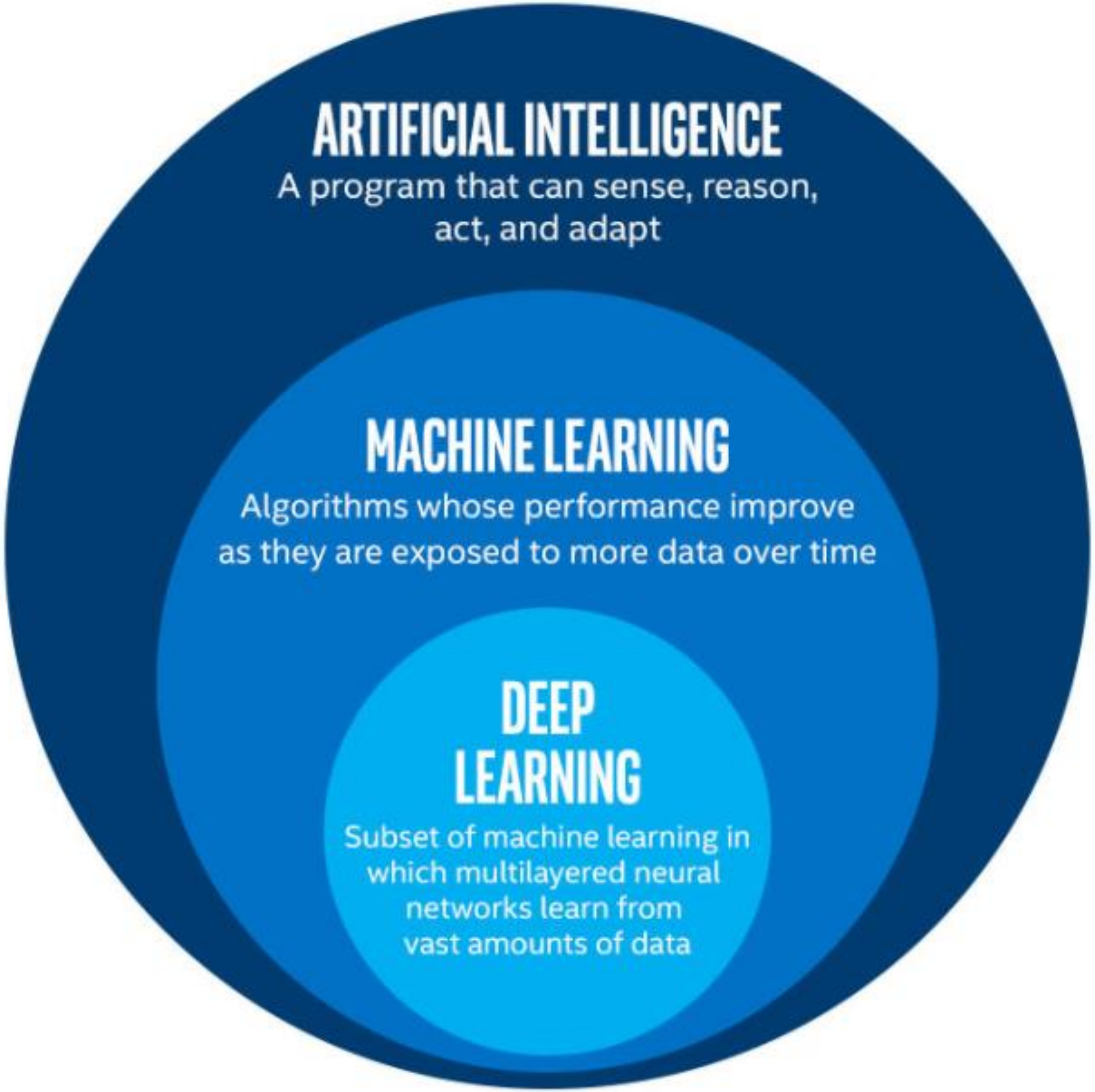
- Learn by interacting with the environment
- The environment reacts to our decisions/actions
- Sequential learning, only at the end of the game we know our performance (reward/punishment)

AlphaGoZero???

Deep Learning; is part of a broader family of machine learning methods based on **Artificial Neural Networks**.

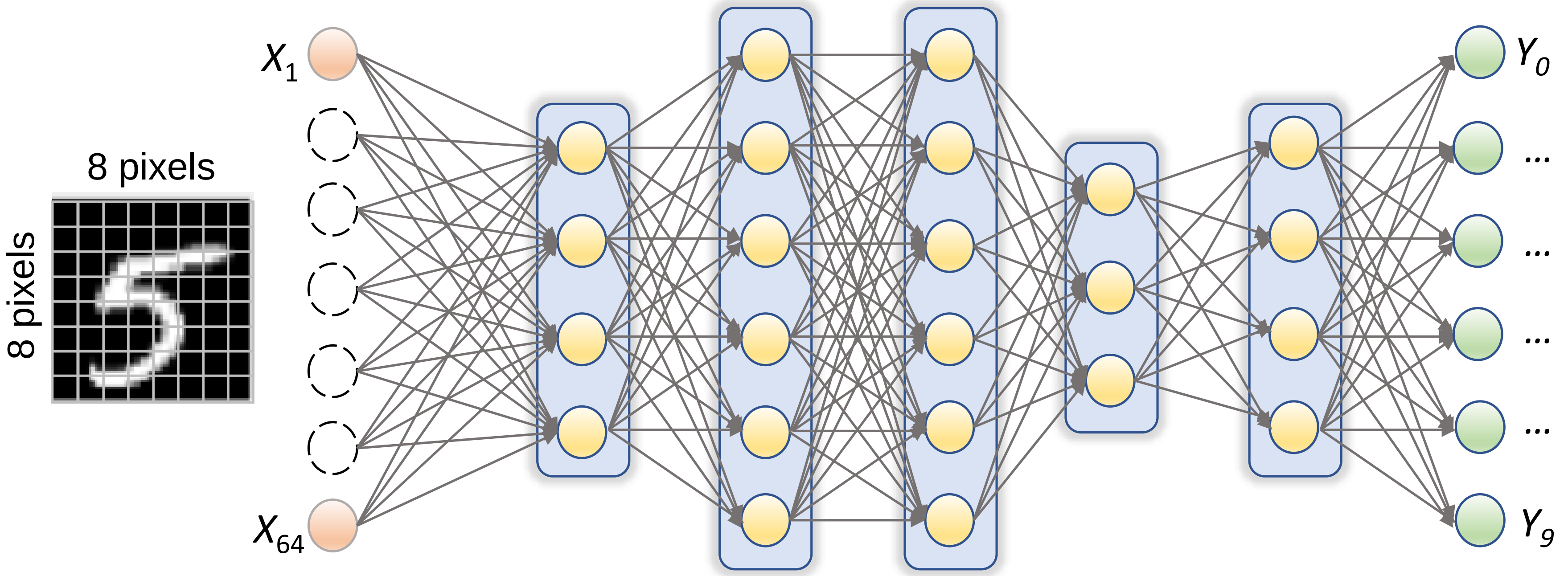
Three kinds of DL;

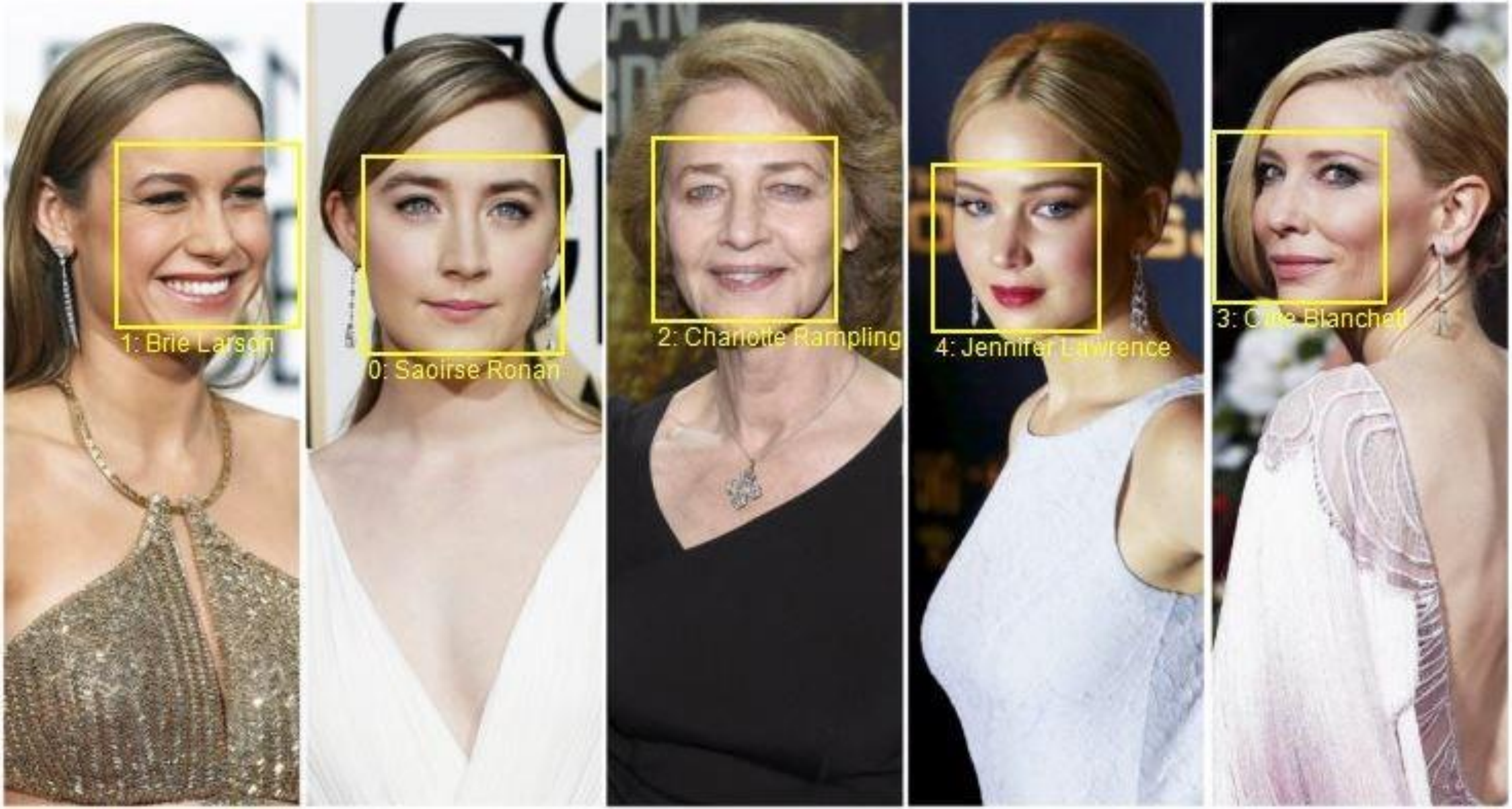
- Supervised
- Self-Supervised
- Reinforcement Learning



Feedforward Neural Networks

- The first and simplest type of artificial neural network devised.
- The information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes.
- There are no cycles or loops in the network.




1: Brie Larson
0: Saoirse Ronan
2: Charlotte Rampling
4: Jennifer Lawrence
3: Cate Blanchett


TrailNet DNN computes pose relative to the trail

DOI:10.1145/3271625

What just happened in artificial intelligence and how it is being misunderstood.

BY ADNAN DARWICHE

Human-Level Intelligence or Animal-Like Abilities?

"The vision systems of the eagle and the snake outperform everything that we can make in the laboratory, but snakes and eagles cannot build an eyeglass or a telescope or a microscope."

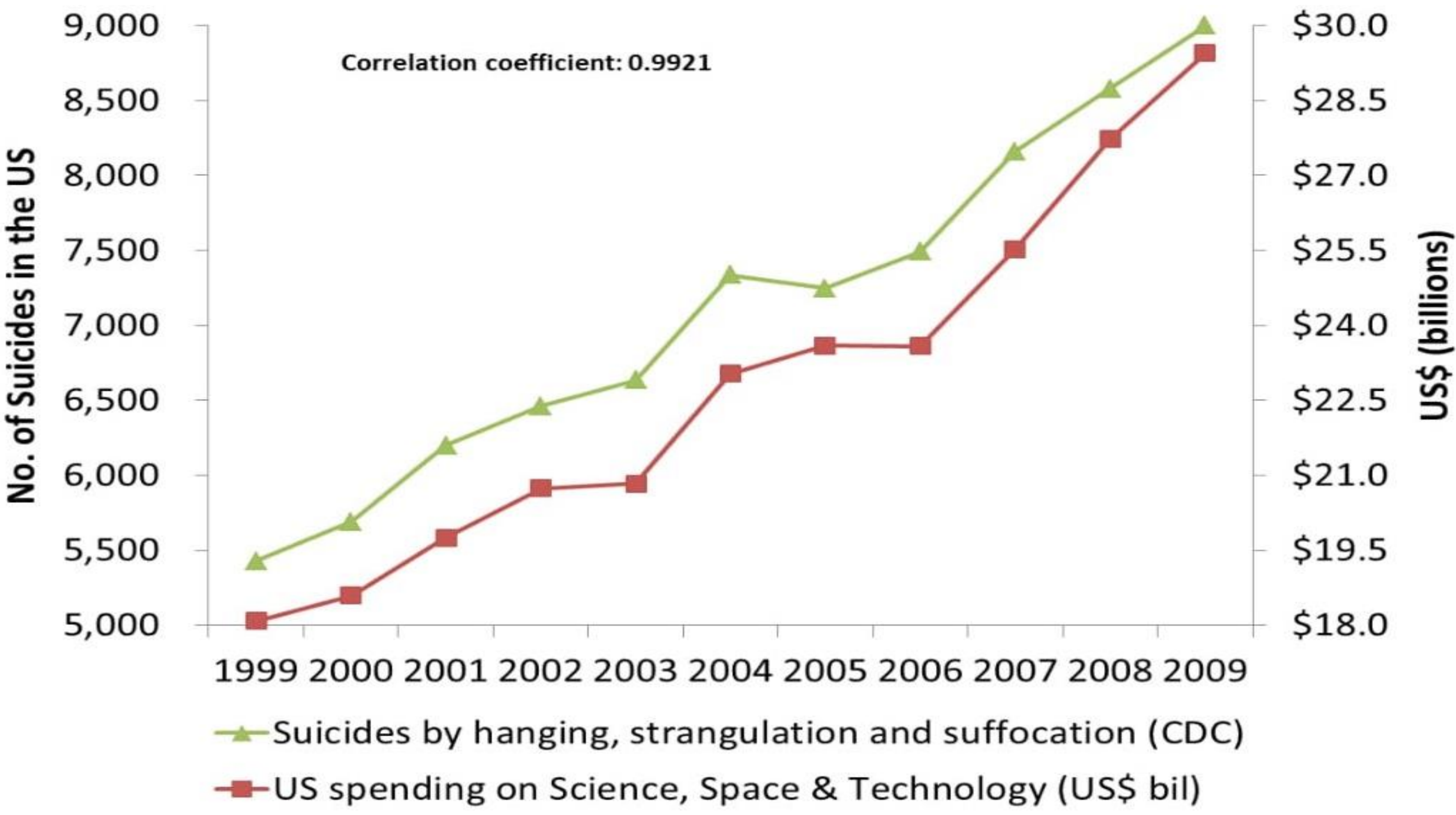
"The vision systems of the eagle and the snake outperform everything that we can make in the laboratory, but snakes and eagles cannot build and eyeglass or a telescope or a microscope."

— *Judea Pearl*

Deep Learning

- Correlazioni spurie

Approssimare può essere fuorviante



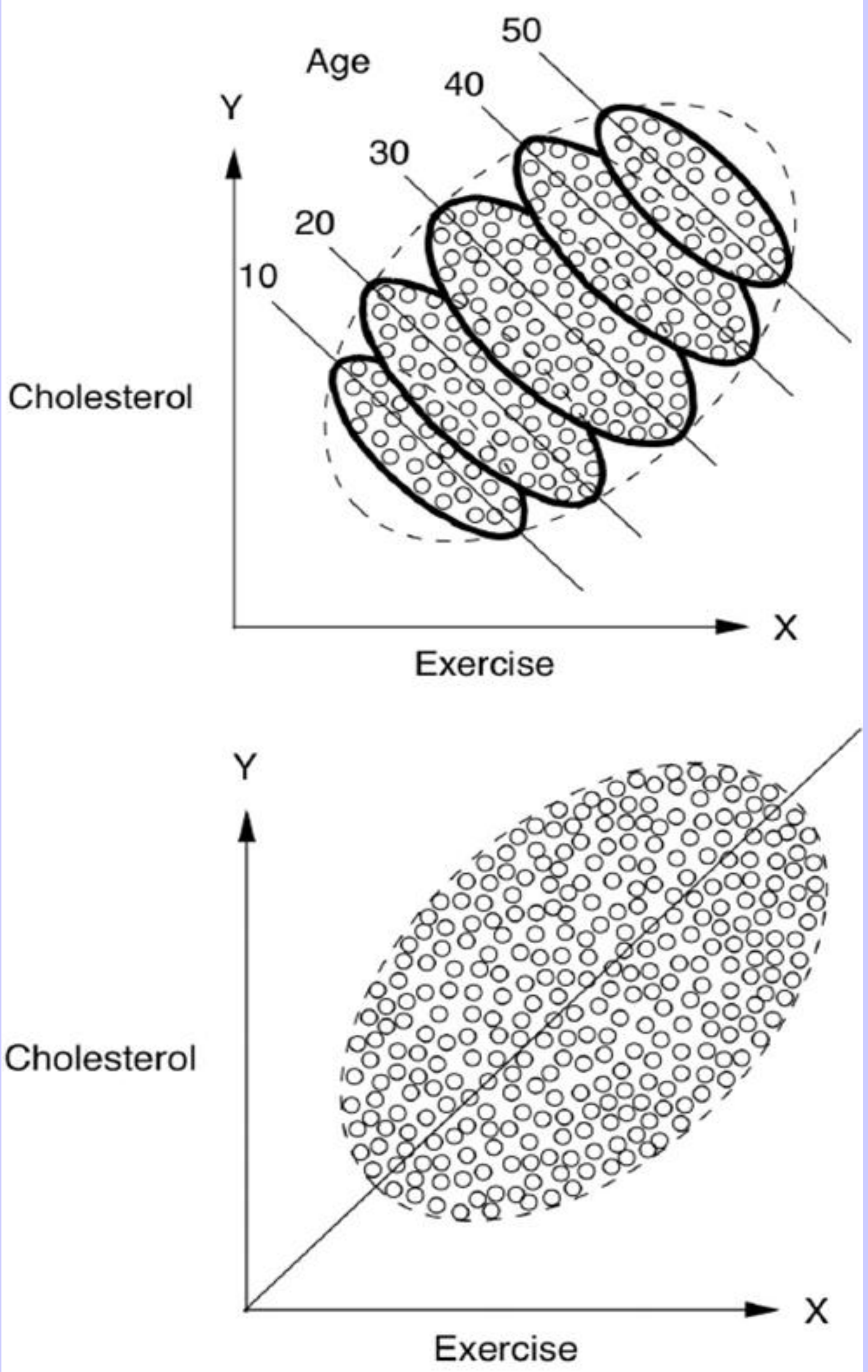
[Spurious Correlations]

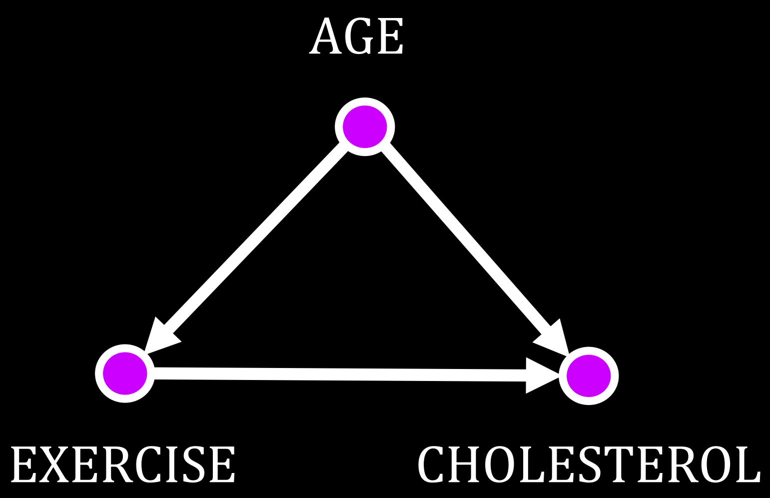
Cosa abbiamo realmente ottenuto?

Exercise impatta il livello del Cholesterol?

Simpson's Paradox

- Non importa quanti dati abbiamo a disposizione, non siamo in grado di conoscere quale alternativa sia corretta





Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug
- Among those who took the drug, a lower percentage recovered than among those who did not
- However, when we partition by gender, we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug
- Among those who took the drug, a lower percentage recovered than among those who did not
- However, when we partition by gender, we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 93% vs 87% male

Simpson’s Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug
- Among those who took the drug, a lower percentage recovered than among those who did not
- However, when we partition by gender, we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 73% vs 69% female

Simpson’s Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug
- Among those who took the drug, a lower percentage recovered than among those who did not
- However, when we partition by gender, we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 78% vs 83% general population!

Simpson's Paradox

- Named after Edward Simpson (born 1922), statistician
- A group of sick patients are given the option to try a new drug
- Among those who took the drug, a lower percentage recovered than among those who did not
- However, when we partition by gender, we see that:
  - *more* men taking the drug recover than do men are not taking the drug, and
  - *more* women taking the drug recover than do women are not taking the drug!

Example 1.2.1 We record the recovery rates of 700 patients who were given access to the drug. A total of 350 patients chose to take the drug and 350 patients did not. The results of the study are shown in **Table 1.1**.

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

Drug vs non-drug takers recovery rates:

- 93% vs 87% male
- 73% vs 69% female
- 78% vs 83% general population!

Should a doctor prescribe the drug; to whom?

Should a policy maker approve the drug for use?

Understand the causal story behind the data

- What mechanism generated the data?
- Suppose: estrogen has a negative effect on recovery
  - Women less likely to recover than men, regardless of the drug

From the data:

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			**No Drug**		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

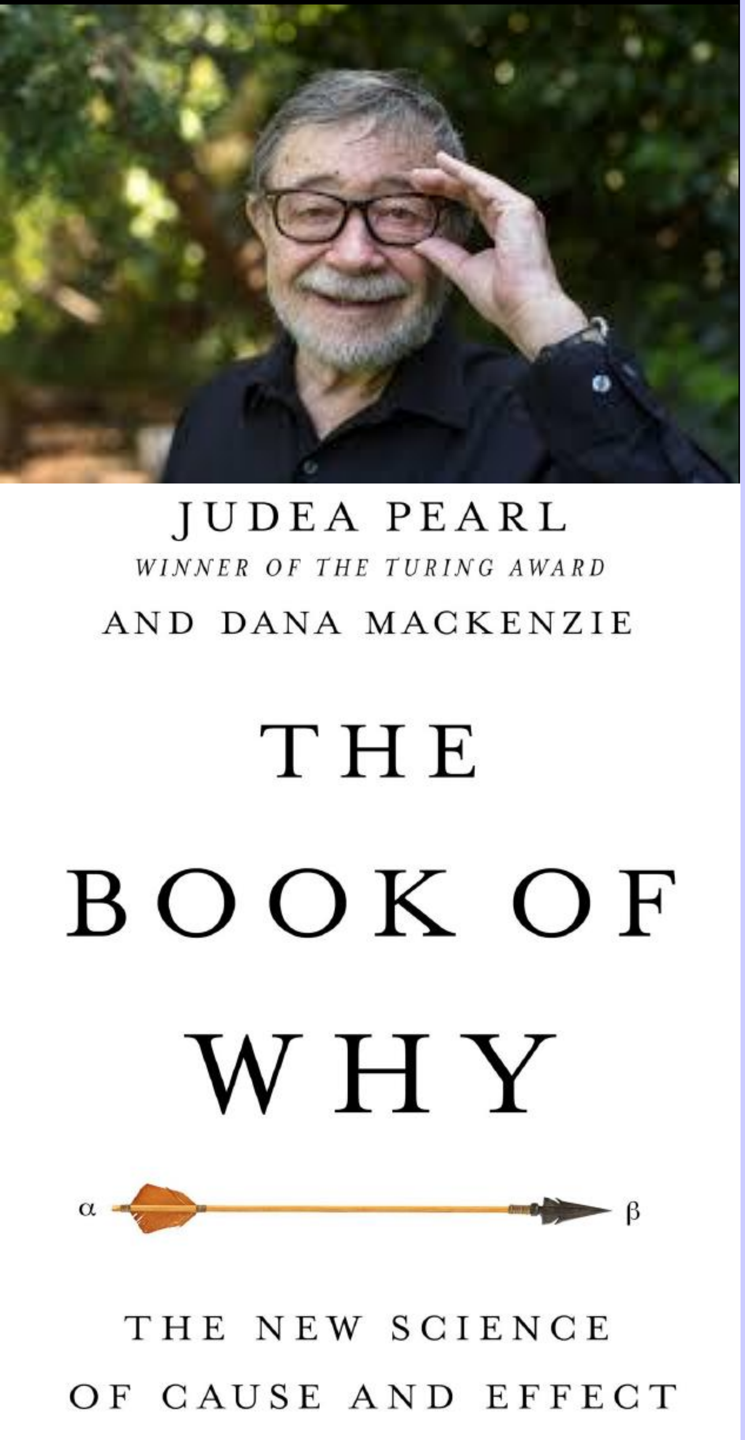
Conclusion: the drug appears to be harmful but it is not

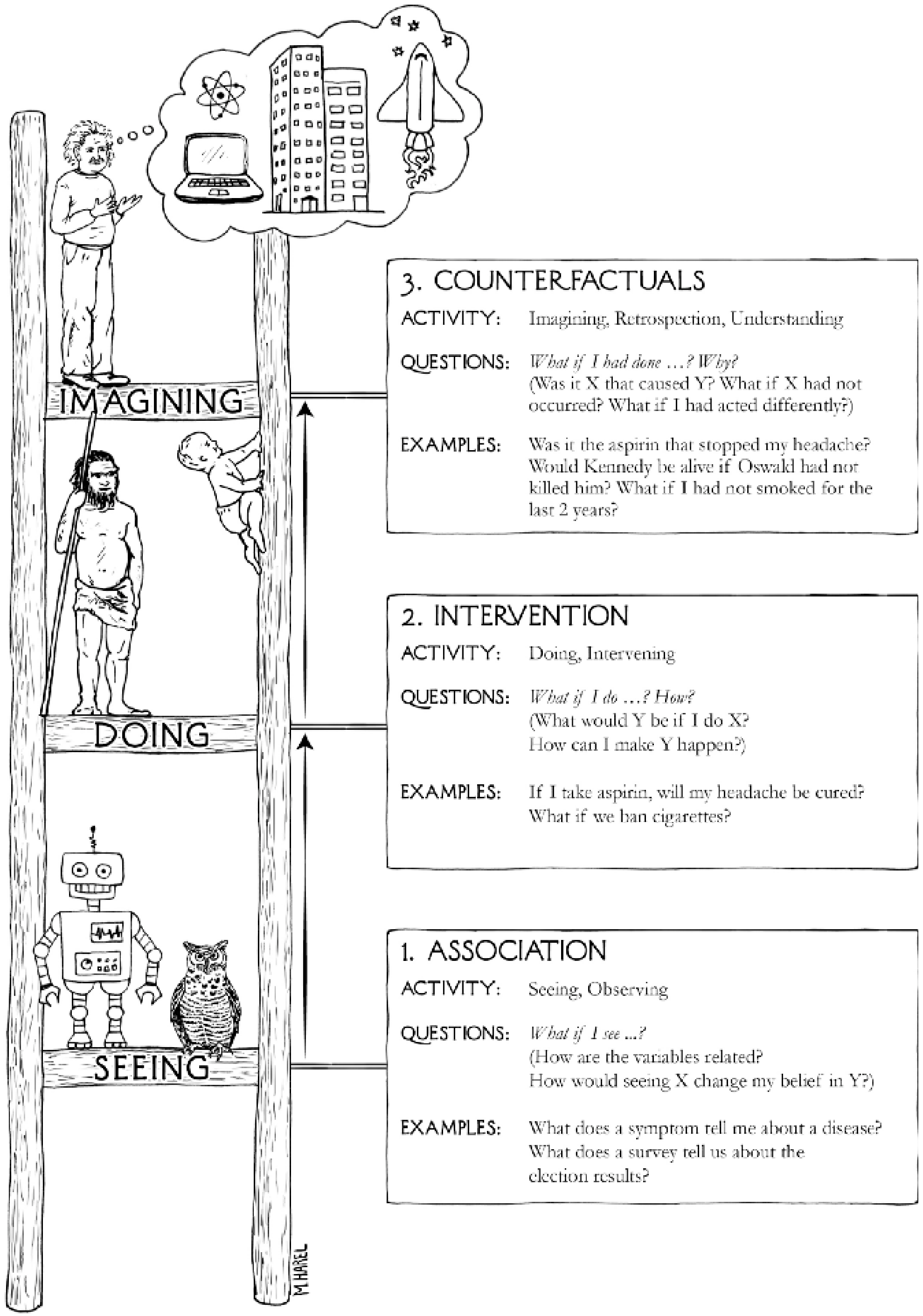
- If we select a drug taker at random, that person is more likely to be a woman
- Hence less likely to recover than a random person who doesn't take the drug

Causal Story

- Being a woman is a common cause of both drug taking and failure to recover.
- To assess the effectiveness we need to compare subjects of the same gender.

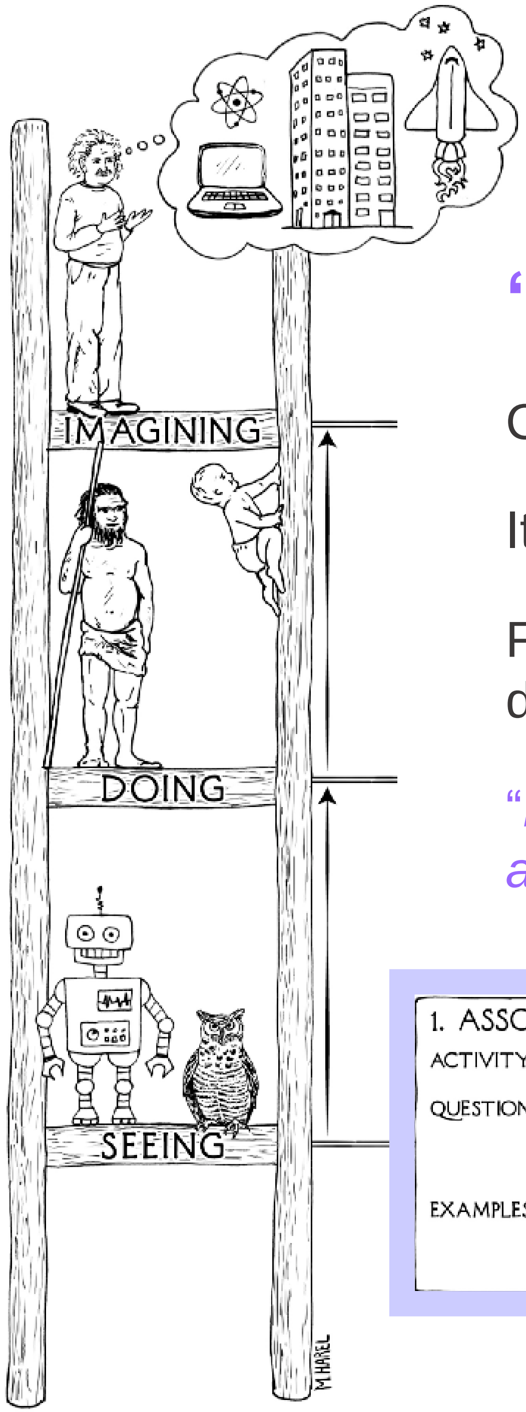
(Ensures that any difference in recovery rates is not ascribable to estrogen)





The Ladder of Causation

Seeing; we are looking for regularities in observations.



"What if I see …?"

Calls for predictions based on passive observations.

It is characterized by the question "*What if I see …?*"

For instance, imagine a marketing director at a department store who asks,

"How likely is a customer who bought toothpaste to also buy dental floss?"

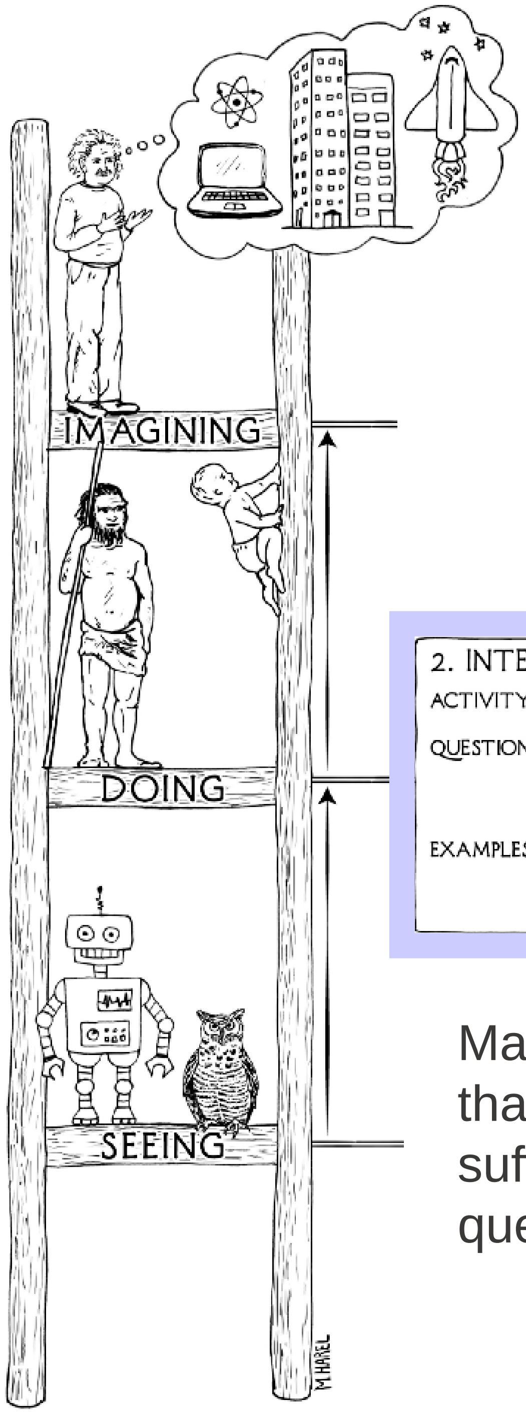
1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

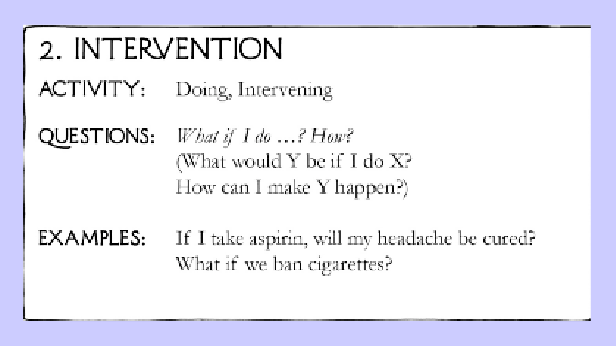
Intervention; ranks higher than association because it involves not just seeing but changing what is.



"What if I do …?" & "How?"

We step up to the next level of causal queries when we begin to change the world. A typical question for this level is

"What will happen to our floss sales if we double the price of toothpaste?"



2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

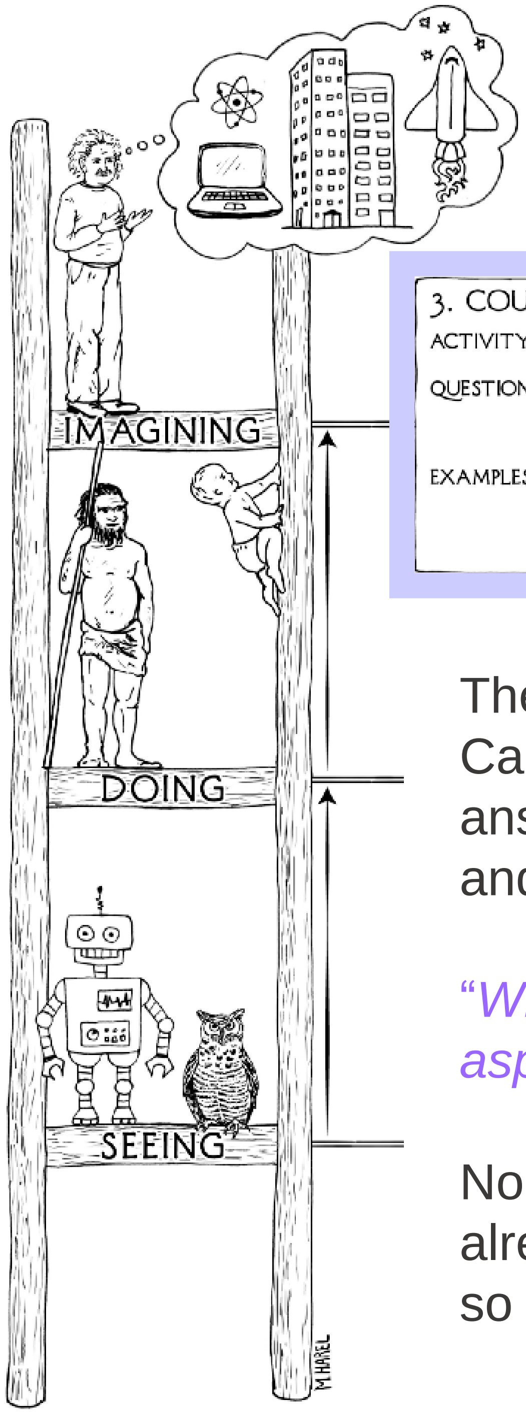
This already calls for a new kind of knowledge, absent from the data, which we find at rung two of the Ladder of Causation, **Intervention**.

Many scientists have been quite traumatized to learn that none of the methods they learned in statistics is sufficient even to articulate, let alone answer, a simple question like

"What happens if we double the price?"

Counterfactuals; ranks higher than intervention because it involves **imagining**, **retrospection** and **understanding**.

"What if I had done …?" & "Why?"



3. COUNTERFACTUALS

ACTIVITY:	Imagining, Retrospection, Understanding
QUESTIONS:	*What if I had done …? Why?* (Was it X that caused Y? What if X had not occurred? What if I had acted differently?)
EXAMPLES:	Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

We might wonder, My headache is gone now, but

- *Why?*
- *Was it the aspirin I took?*
- *The food I ate?*
- *The good news I heard?*

These queries take us to the top rung of the Ladder of Causation, the level of **Counterfactuals**, because to answer them we must go back in time, change history, and ask,

"What would have happened if I had not taken the aspirin?"

No experiment in the world can deny treatment to an already treated person and compare the two outcomes, so we must import a whole new kind of knowledge.

eduopen

Catalogo ▾ Istituzioni Appuntamenti Certificazioni Approfondisci

eduopen

Catalogo ▾ Istituzioni Appuntamenti Certificazioni Approfondisci

UNIVERSITA' DI MILANO
BICOCCA

Data Mining - Classification

Parte di

Percorso in **Introduction to Data Min**

About the Course

Learn how to formulate and solve classification problems for
Mining and Business Intelligence applications such as; fraud
customer churning, network intrusion detection, ... You will lea
develop, validate and apply a data mining workflow to solve bi
non-binary classification problems. The course is self-contain
does not require any programming skills. Hands-on lectures a
the KNIME open source software platform.

Data Mining - Classification

Il corso

This course introduces basic concepts and methods of Data M
Classification. In particular, the course will provide the followi
type, summarization, missing data replacement, and data pre-
introduced together with the following concepts; explanatory
set, classifier (learner and inducer), performance measures (a
folds cross validation, overfitting and underfitting, curse of dir
operating characteristic curve, lift and cumulative gains chart

eduopen

Catalogo ▾ Istituzioni Appuntamenti Certificazioni Approfondisci

UNIVERSITA' DI MILANO
BICOCCA

Text Mininig

Parte di

Percorso in **Introduction to Data Mining**

About the Course

Natural language text is everywhere, social networks, business, finance, medicine and biology are just few of the many sources of natural language text. However, computers are not fit to process natural language text. Indeed, Data Mining methods and algorithms, which operate on structured data, can not be directly applied to unstructured data for knowledge extraction. The Text Mining course, the last one of the Introduction to Data Mining pathway, introduces methods and tools for knowledge extraction from natural language text. The course assumes you are familiar with methods and models presented in the previous two courses, namely Data Mining: Classification and Data Mining: Clustering and Association. The course shows that Text Mining allows to formulate and solve problems in Business Intelligence, Finance, Recommendation, Medicine, Biomedicine, Social Networks, and Intelligence Gathering to mention just a few. In particular, the course introduces methods, models and algorithms for; natural language text preprocessing, text categorization, text clustering, topic modeling and information extraction.

Text Mininig

Il corso

This course introduces basic concepts and methods of Text Mining with specific reference to natural

Autoapprendimento
Iscriviti adesso!

Università degli Studi di Milano-Bicocca

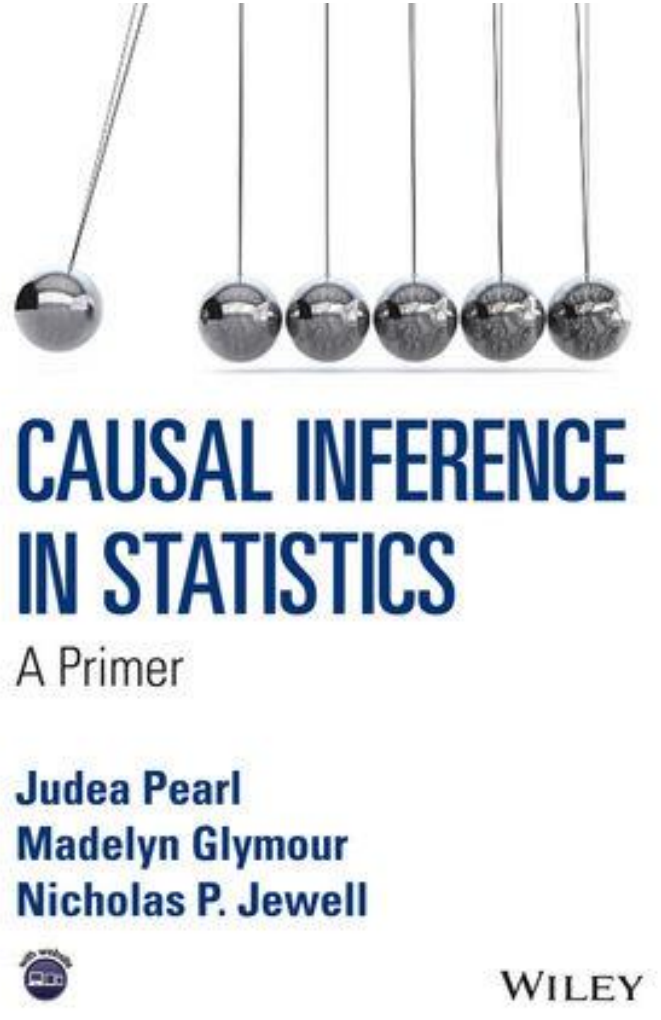
Introduzione al Data Mining

In corso

ATO

Pedagogia ed | Imparare a imparare: i | Introduzione a Stata | Counselling e

References




DOI:10.1145/3271625

What just happened in artificial intelligence and how it is being misunderstood.

BY ADNAN DARWICHE

Human-Level Intelligence or Animal-Like Abilities?

"The vision systems of the eagle and the snake outperform everything that we can make in the laboratory, but snakes and eagles cannot build an eyeglass or a telescope or a microscope."