Causal Networks: Learning and Inference

Fabio Stella and Luca Bernardinello
University of Milan-Bicocca
Department of Informatics, Systems and Communication

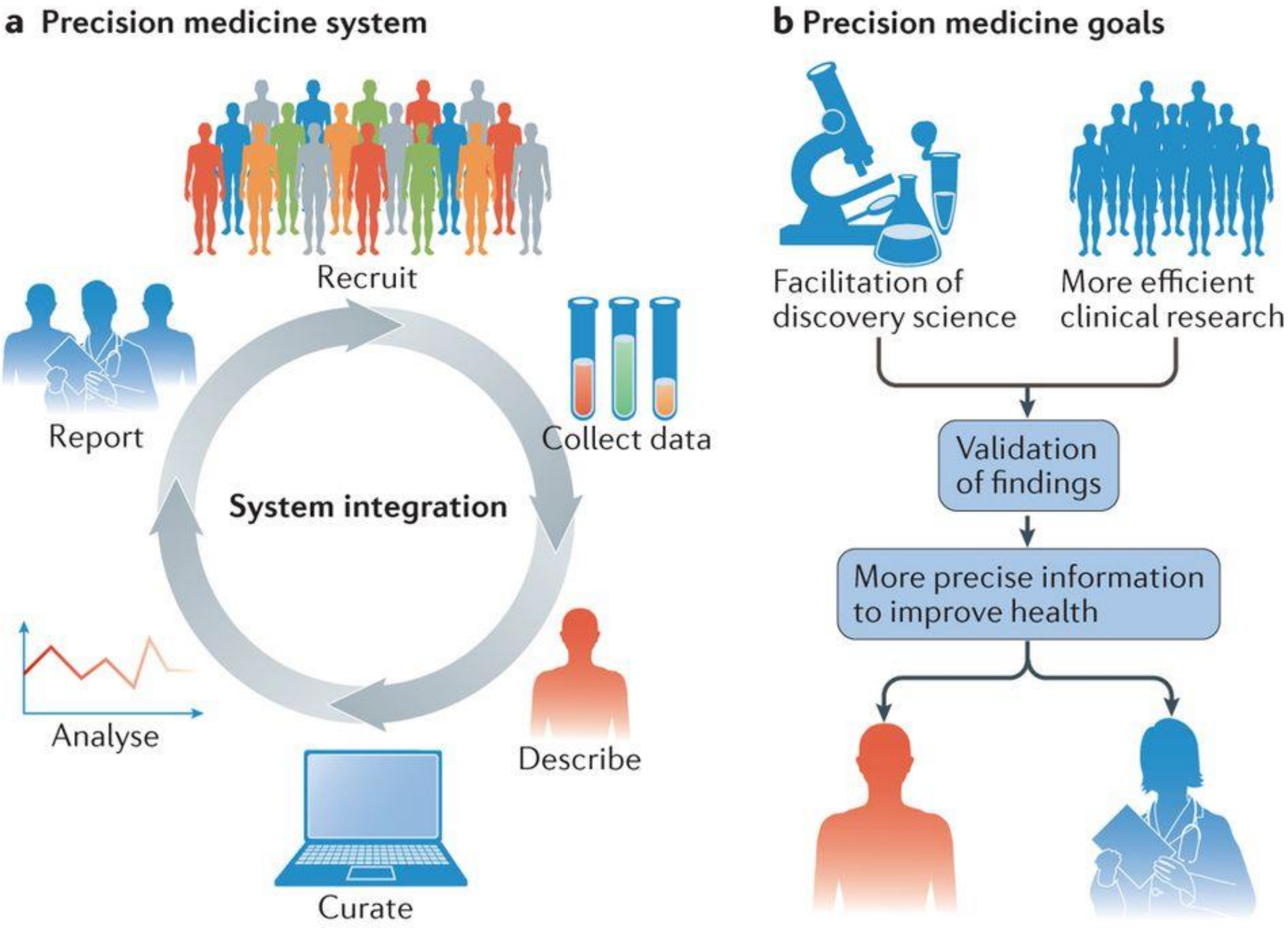
2019 September 23rd – 27th

The 100,000 Genomes Project

- complete sequencing of 70,000 individuals
- 21 PetaBytes of data
- 1 PetaByte of music requires about 2,000 years to be listened

The complexity increases with digital data:

- Electronic Health Record
- Life style
- Nutrition
- Job type
- Geographical
- Social networks

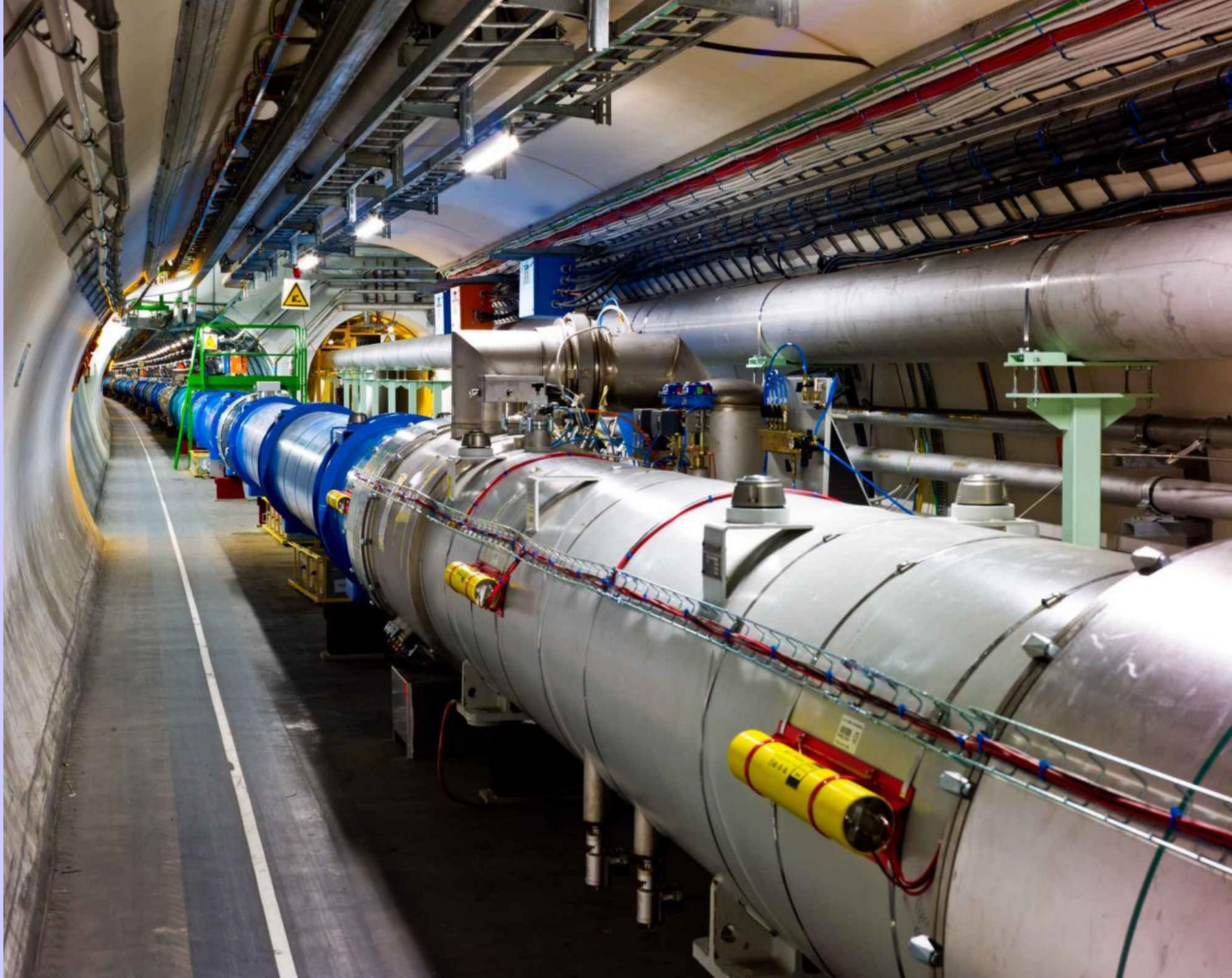


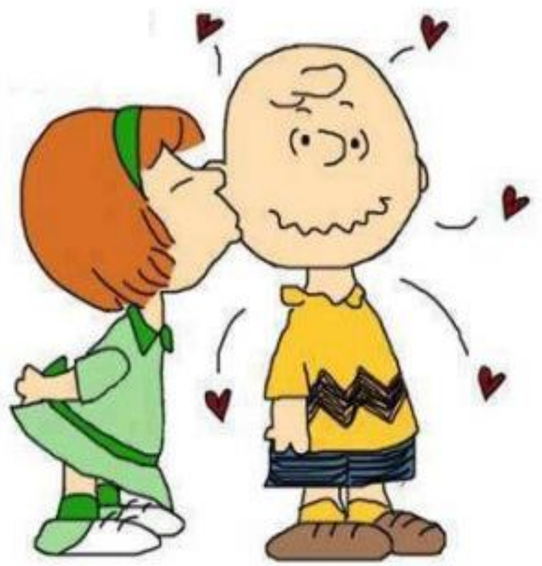
Smart Cities

- Automate
- Control
- See real-time data
- Home automation
- Environment monitoring
- Medical and healthcare
- Smart transportation
- Smart manufacturing
- Energy resource management

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.





"Big data is like teenage sex;
everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is doing it,
so everyone claims they are doing it".

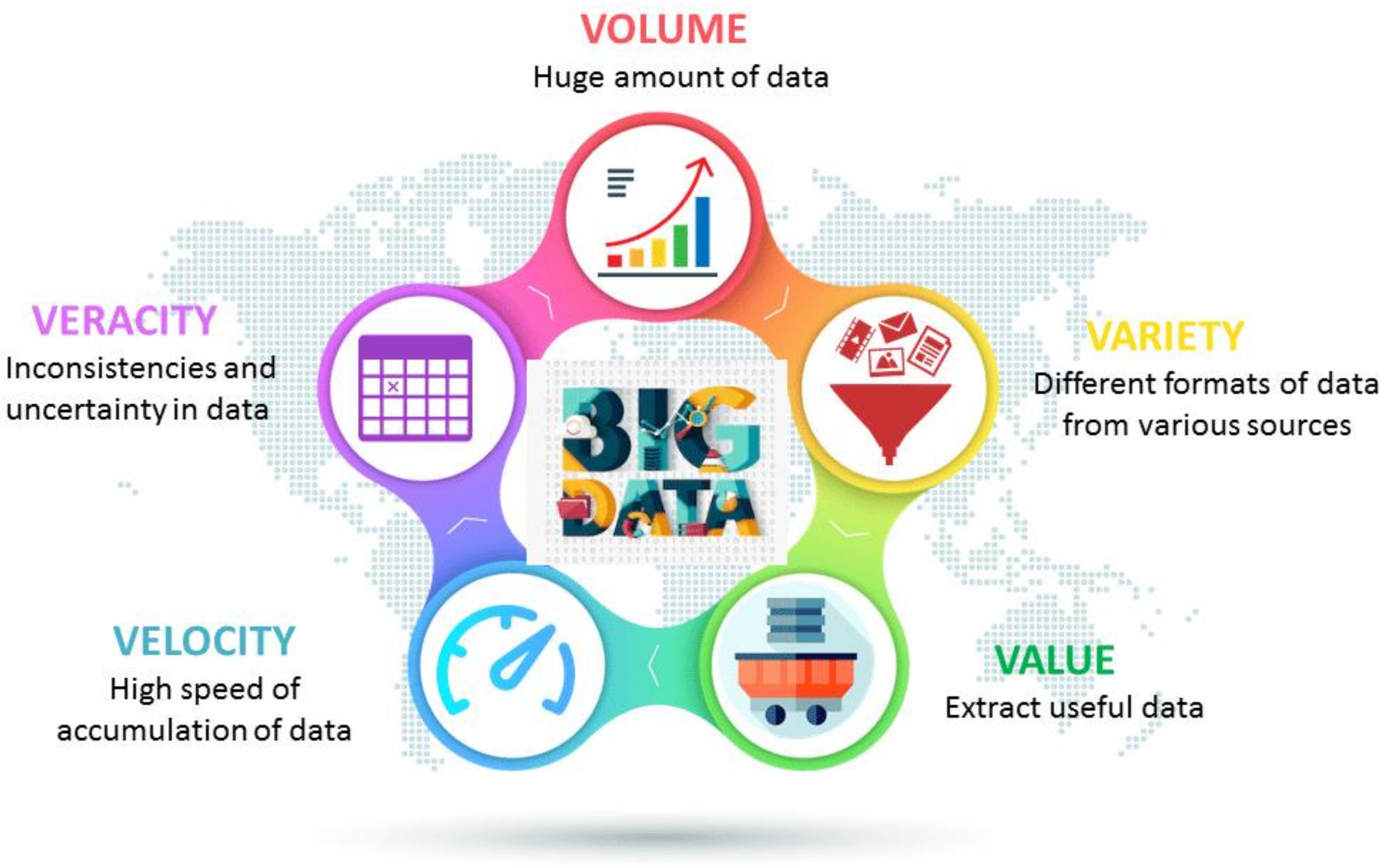
Dan Ariely, Duke University

According to **[1]**,

- **Big Data and Data Science** are being used as buzzwords and are composites of many concepts.
- Big data appears frequently in the press and academic journals.
- In the last five years; birth and growth of many data science programs in academia.
- In 2012, the **White House Office of Science and Technology Policy** announced the **Big Data Research and Development Initiative** that builds upon federal initiatives ranging from
  - ✓ *computer architecture and networking technologies,*
  - ✓ *algorithms,*
  - ✓ *data management,*
  - ✓ *artificial intelligence,*
  - ✓ *machine learning,*
  - ✓ *advanced cyber-infrastructure.*

[1] Brady H. E., *Annual Review of Political Science*, 22 (2019) 297.

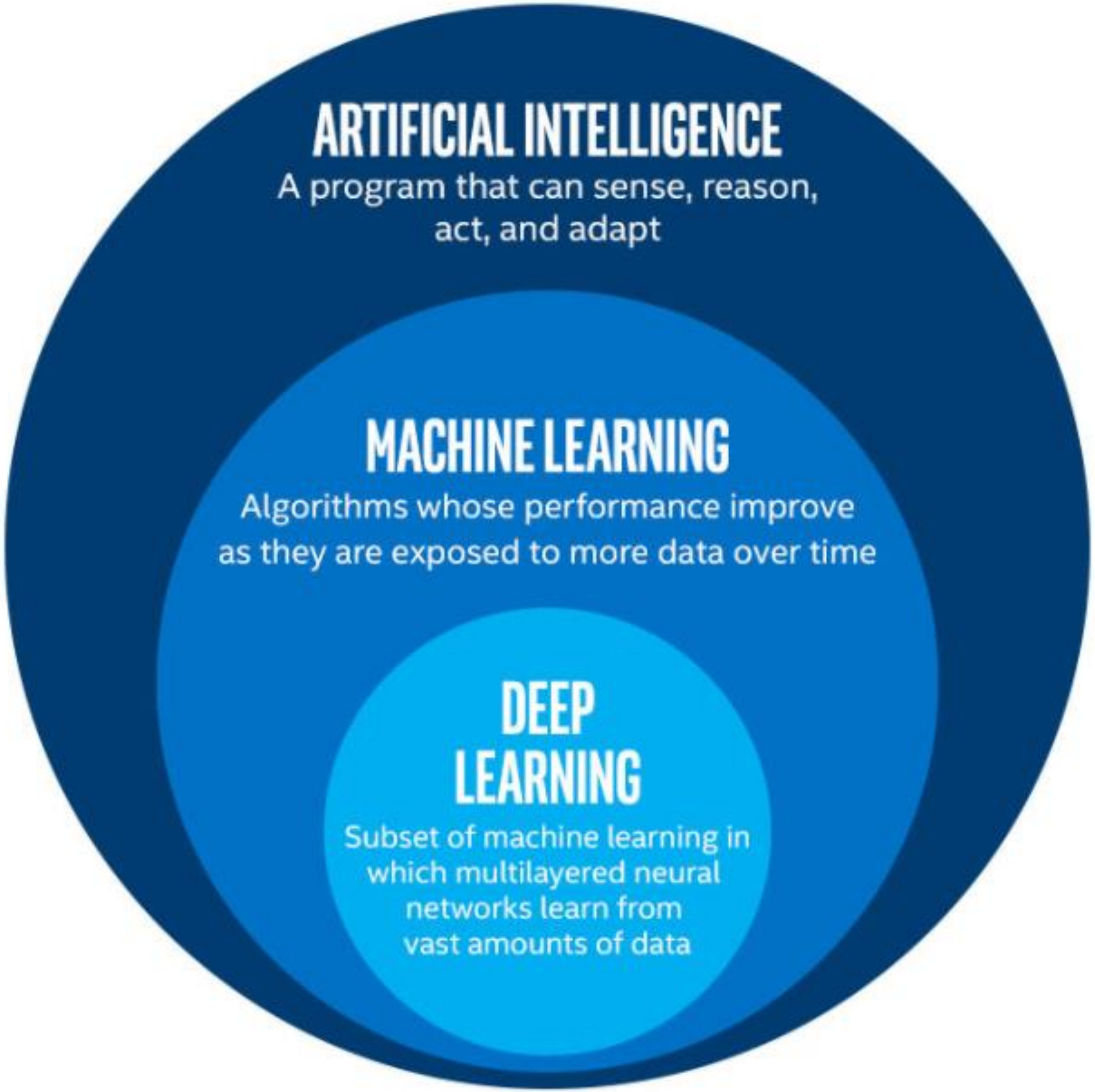
Big Data 5 V's


VOLUME
Huge amount of data
VARIETY
Different formats of data from various sources
VALUE
Extract useful data
VELOCITY
High speed of accumulation of data
VERACITY
Inconsistencies and uncertainty in data
BIG DATA

Artificial Intelligence;

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

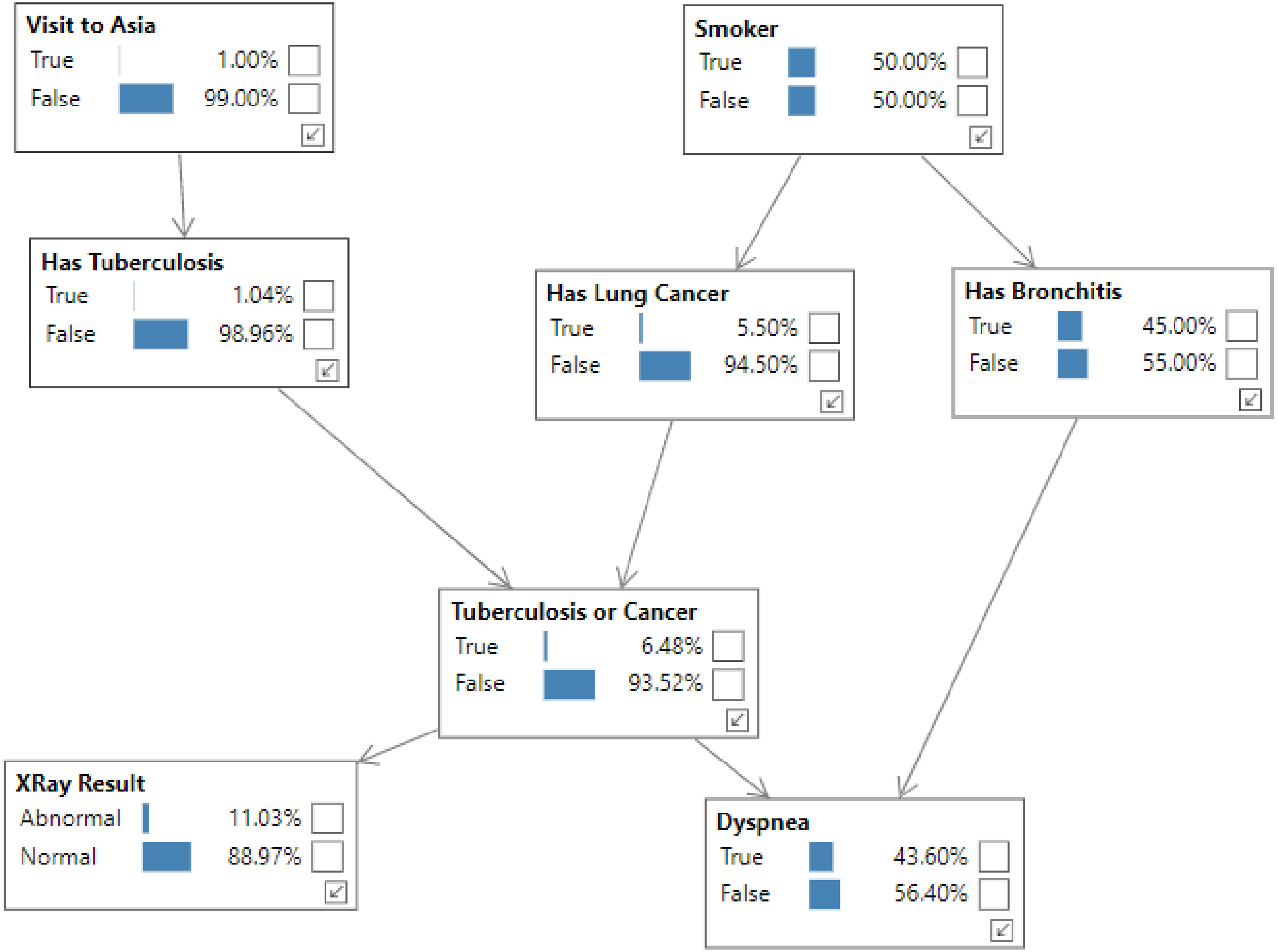
- Colloquially, the term Artificial Intelligence (**AI**) is used to describe machines/computers that mimic "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".
- Two kinds of AI:
  - ✓ Weak
  - ✓ Strong



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.

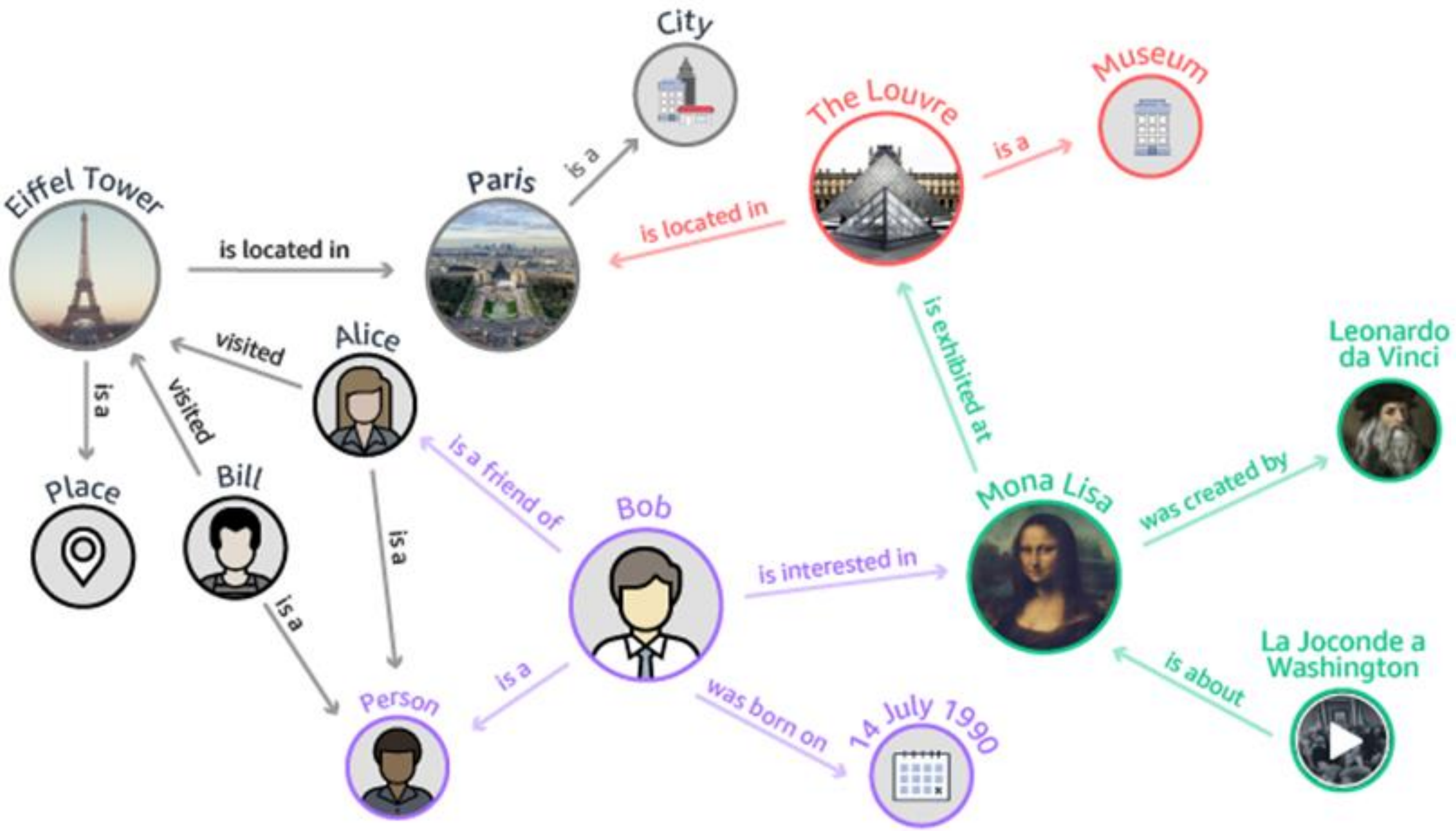


Artificial Intelligence

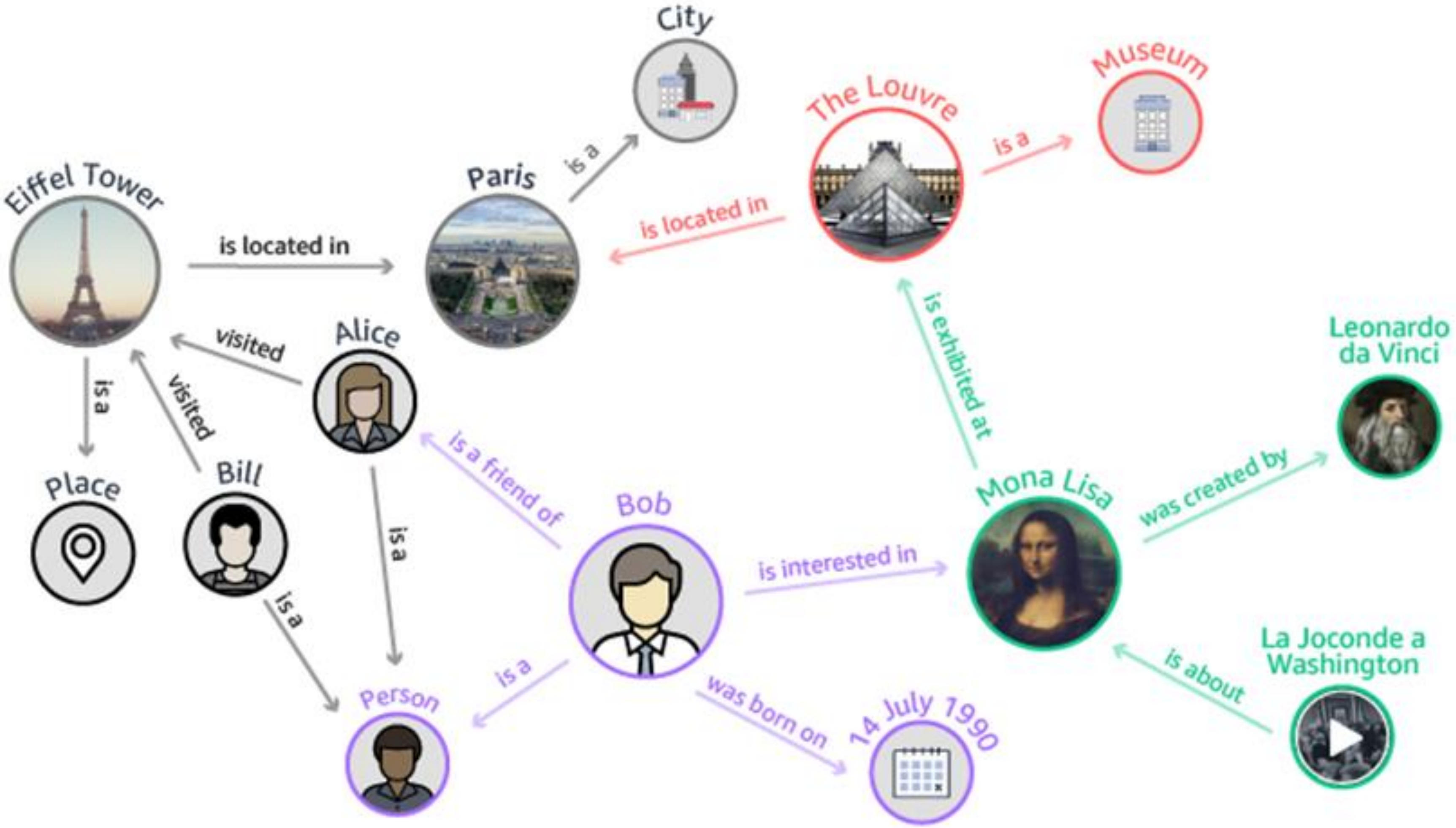
Knowledge Graphs

- A knowledge base used by **Google** and its services to enhance its search engine's results with information gathered from a variety of sources.

- The information is presented to users in an **infobox**.




Altre immagini
Enrico Fermi
Fisico
Enrico Fermi è stato un fisico, inventore e accademico italiano naturalizzato statunitense. È noto principalmente per gli studi teorici e sperimentali nell'ambito della meccanica quantistica, e in particolare della fisica nucleare. Wikipedia
Nascita: 29 settembre 1901, Roma
Decesso: 28 novembre 1954, Chicago, Illinois, Stati Uniti
Coniuge: Laura Fermi (s. 1928–1954)
Sepoltura: Oak Woods Cemetery, Chicago, Illinois, Stati Uniti
Figli: Giulio Fermi, Nella Fermi
Libri
Visualizza altri 5 elementi
Notes on Quantum Mechanics
Elementary Particles
1951
Termodin...
1937
Note e memorie
Notes on Thermody... and Statis...
Ricerche correlate
Visualizza altri 15 elementi
Paul Dirac
Erwin Schrödinger
Werner Karl Heisenberg
Niels Bohr
Robert Oppenhei...

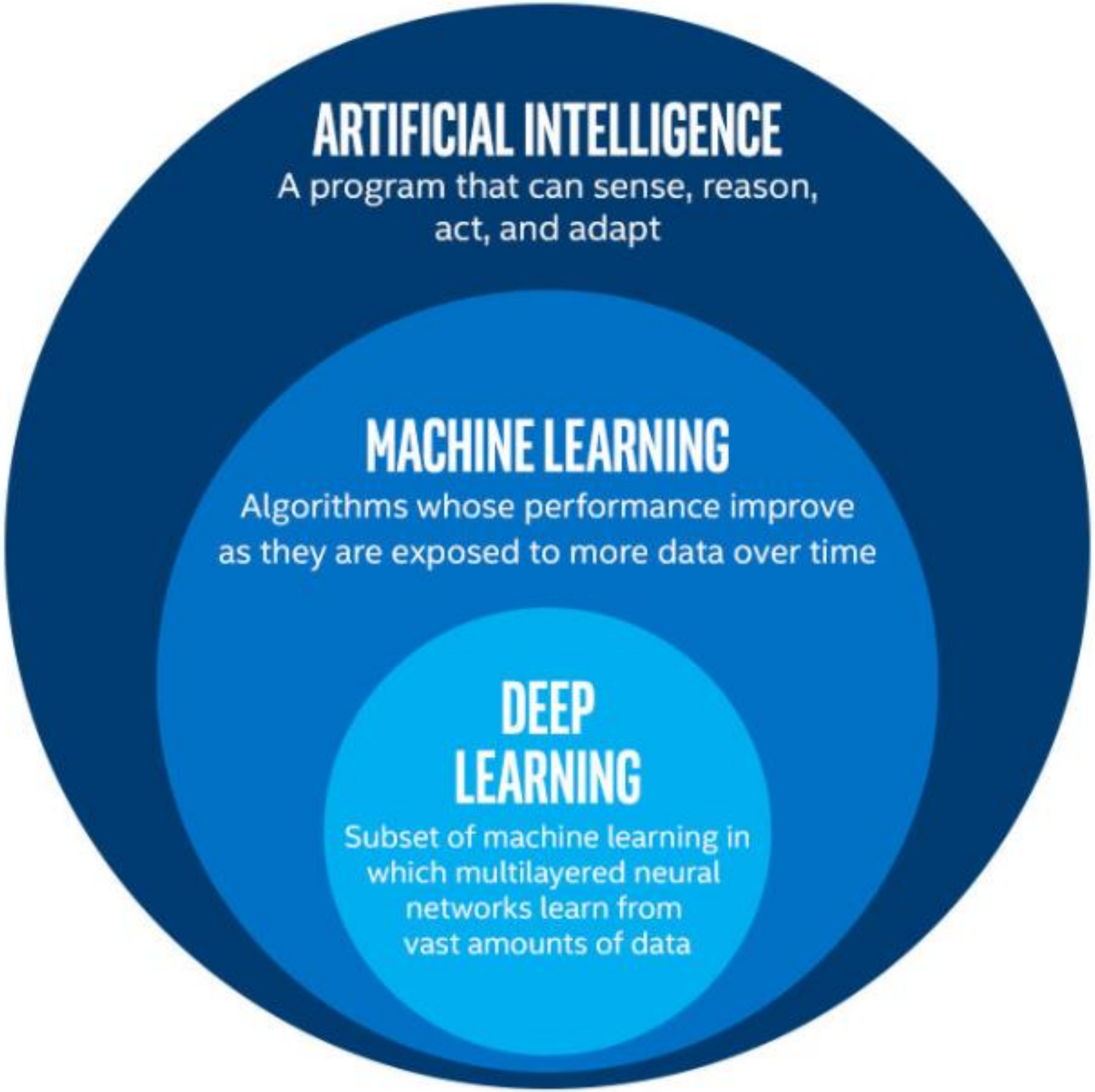

City
Museum
The Louvre
is a
is a
Paris
Eiffel Tower
is located in
is located in
visited
Alice
visited
is a
Place
Bill
is a
is exhibited at
Leonardo da Vinci
is a friend of
Bob
Mona Lisa
was created by
is a
is interested in
La Joconde a Washington
is a
Person
was born on
14 July 1990
is about

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

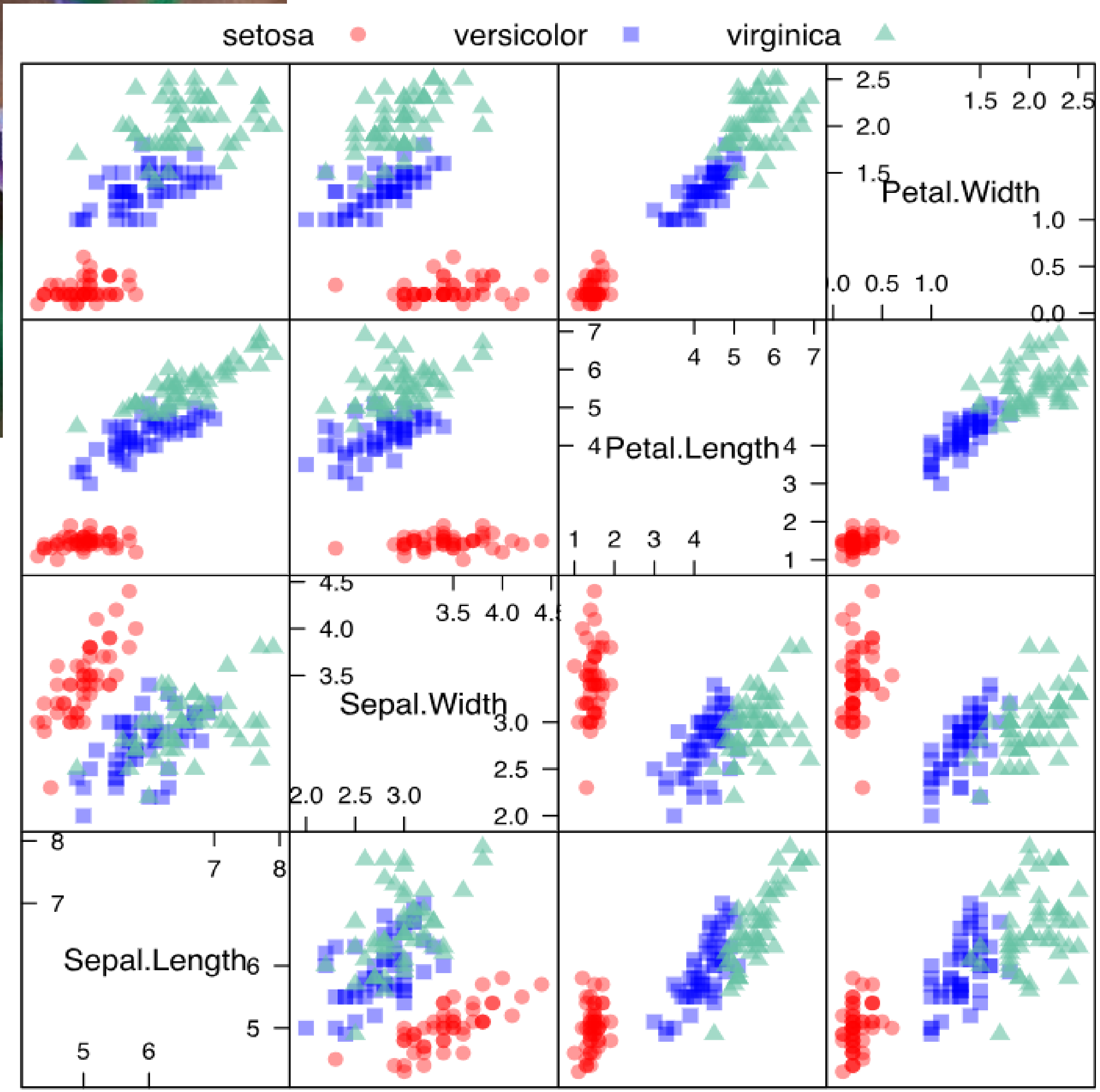
Supervised

- Classification

SETOSA

VERSICOLOR

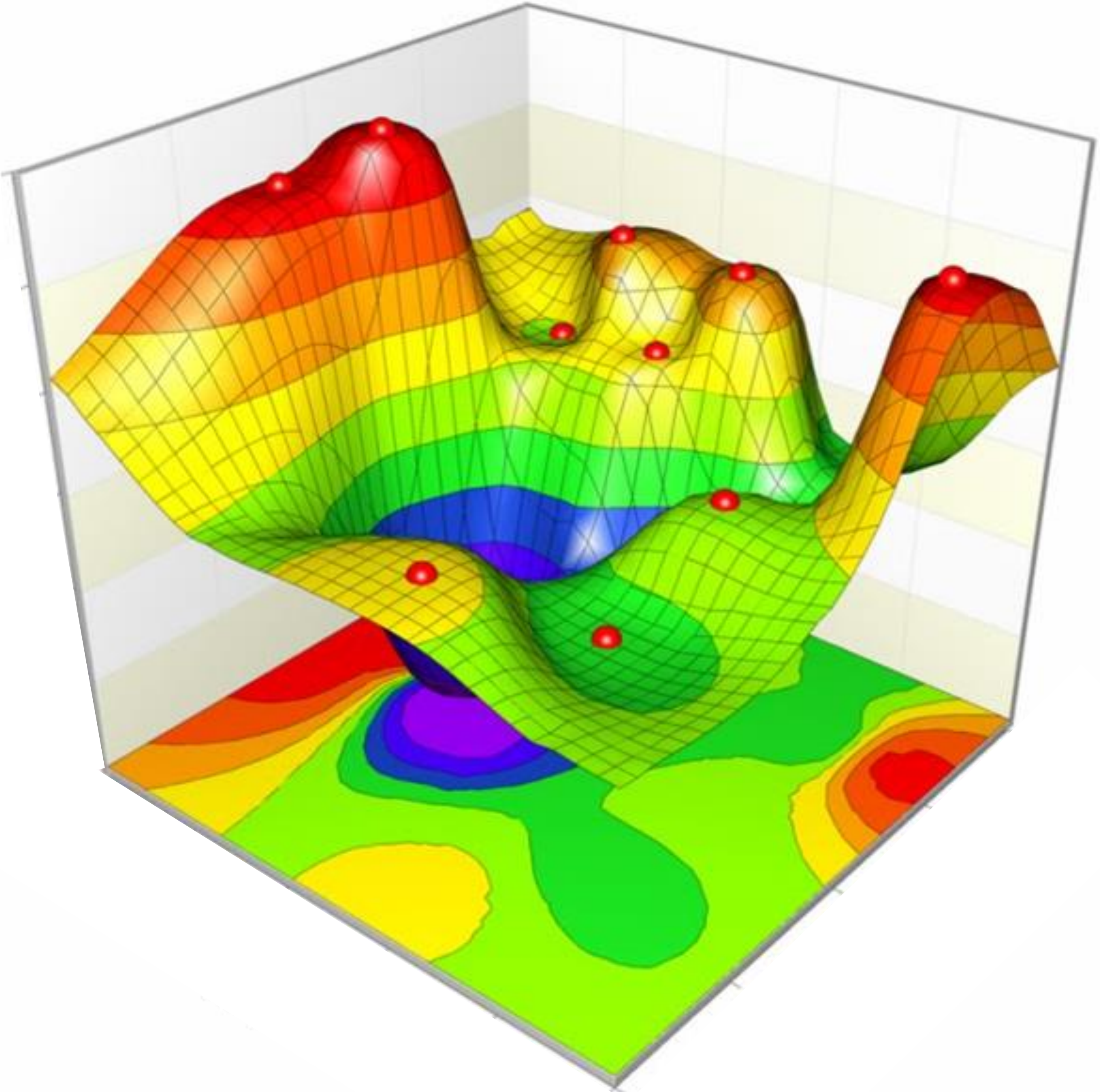
VIRGINICA



Machine Learning

Supervised

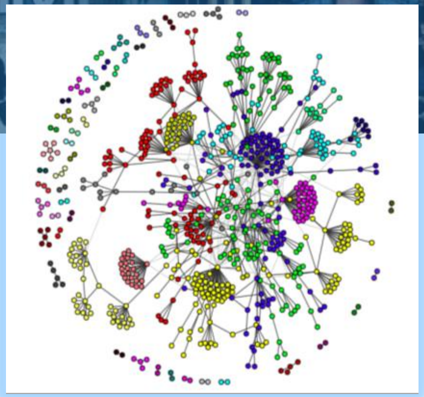
- Curve fitting
- Surface fitting

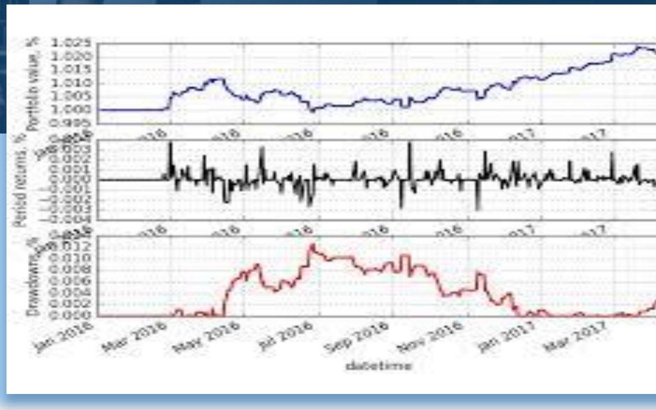


Machine Learning

Self-Supervised

- Recommendation System
- Market Basket Analysis
- Social Network Analysis





Machine Learning

Reinforcement Learning

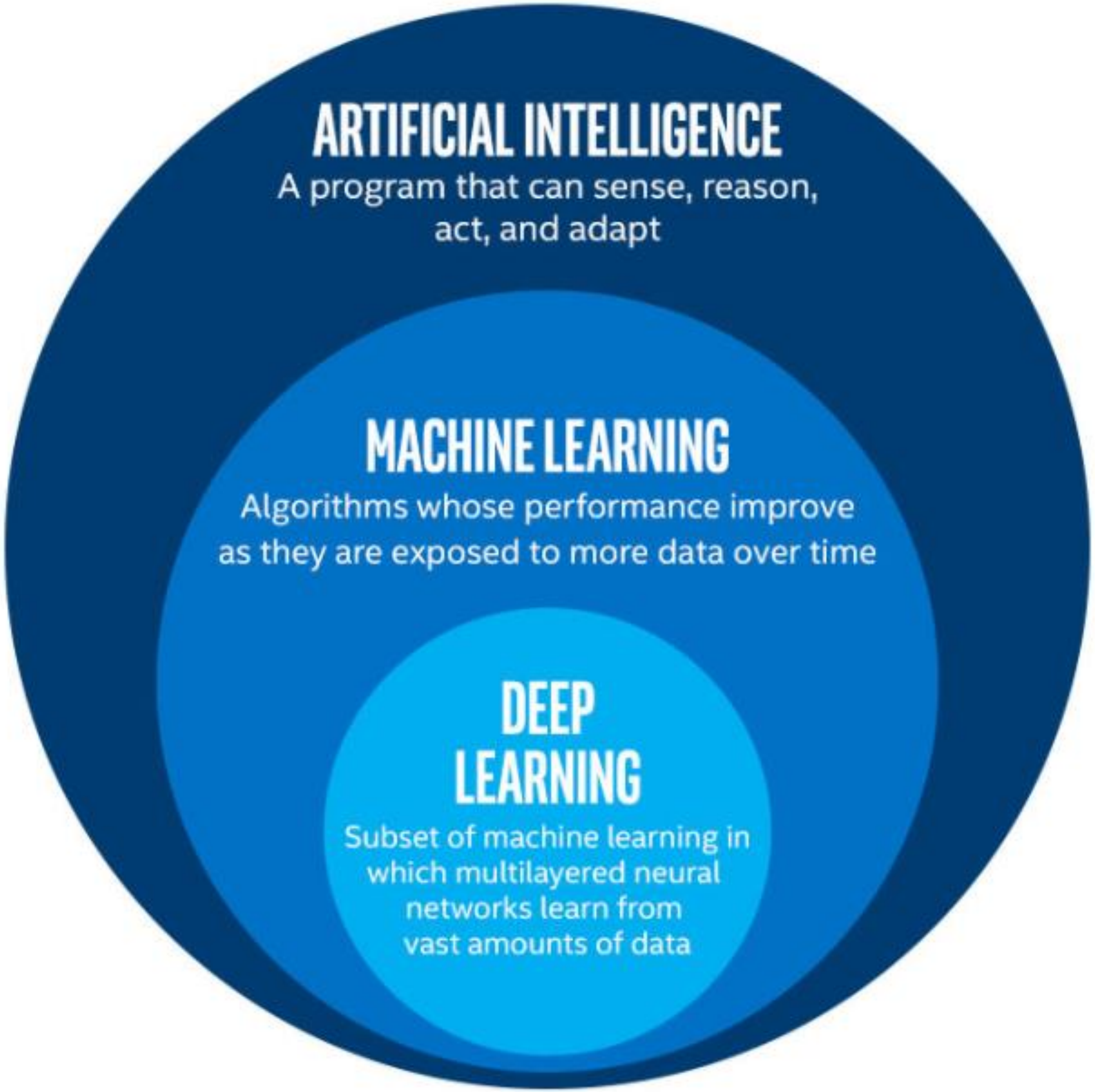
- Learn by interacting with the environment
- The environment reacts to our decisions/actions
- Sequential learning, only at the end of the game we know our performance (reward/punishment)

Deep Learning;

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

Three kinds of DL:

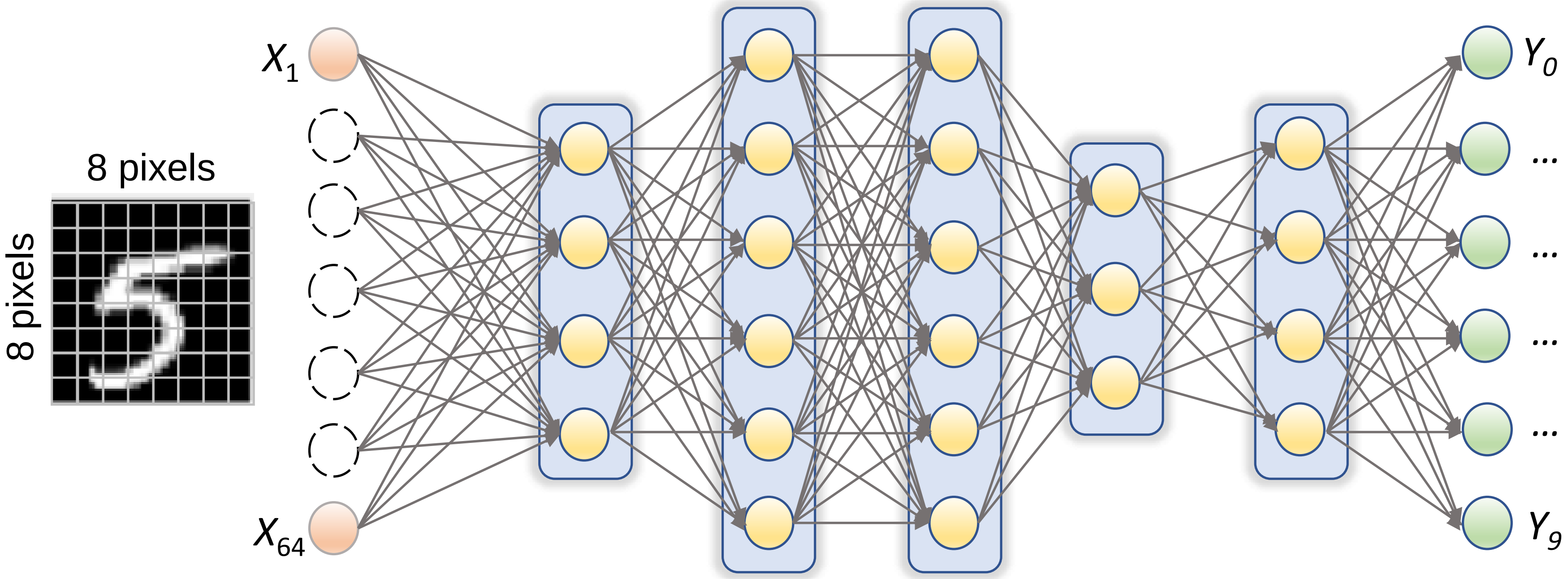
- Supervised
- Self-Supervised
- Reinforcement Learning



Deep Learning

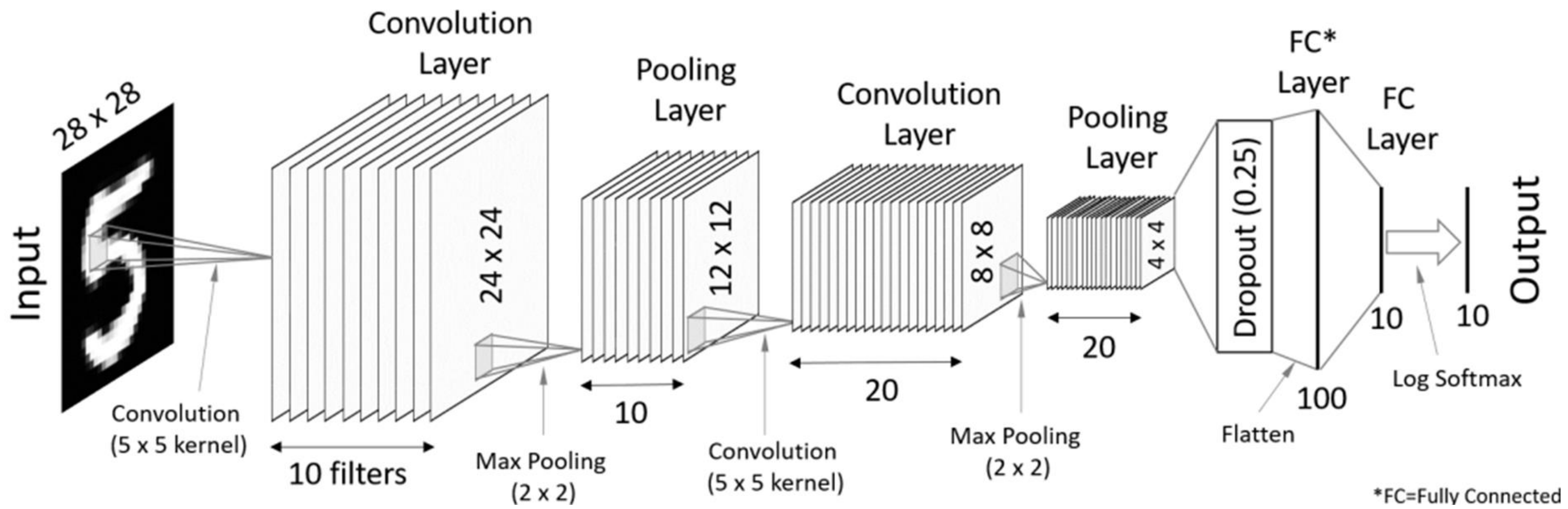
Feedforward Neural Networks

- 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.



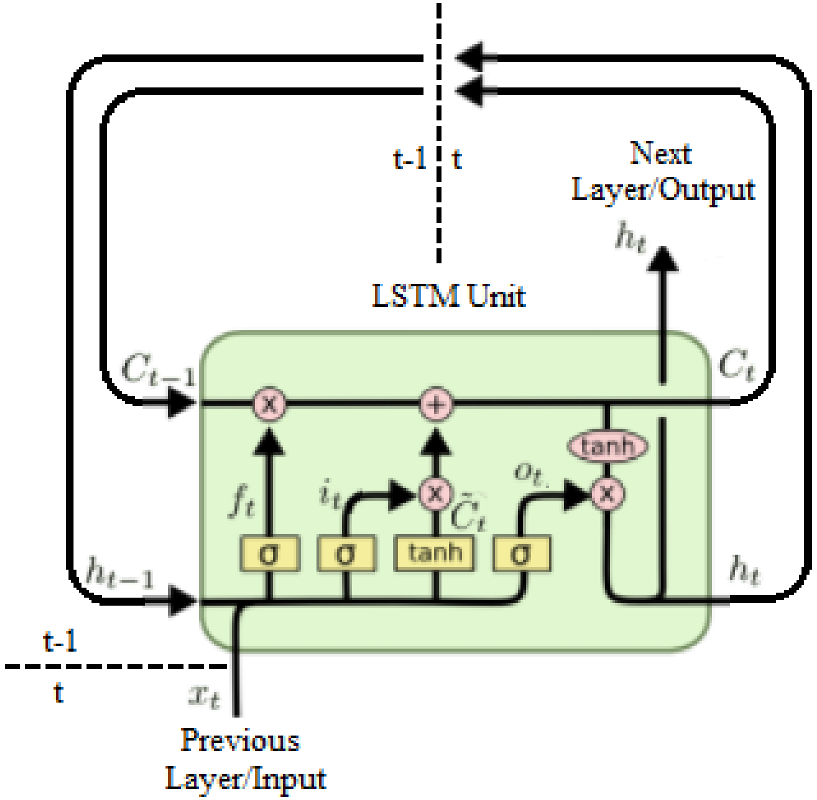
Convolutional Networks

- Regularized versions of multilayer perceptrons which are fully connected and thus prone to overfitting the data.
- Regularization by adding some form of magnitude measurement of weights to the loss function.
- Different approach towards regularization: take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns.

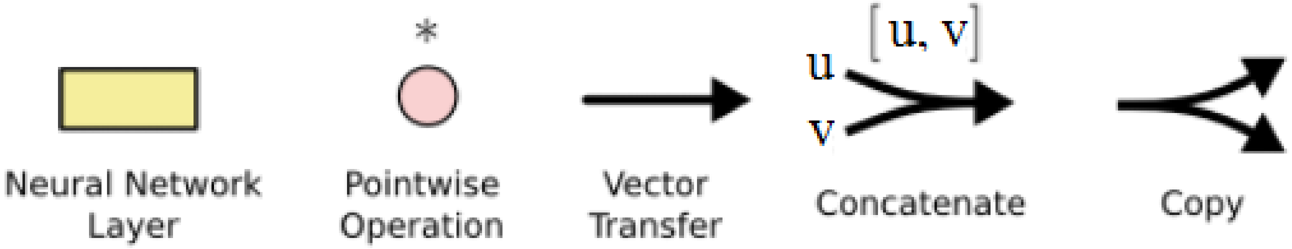


LSTM Networks

- An artificial Recurrent Neural Network architecture.
- Unlike standard feedforward neural networks, LSTM has feedback connections that make it a "general purpose computer" (it can compute anything that a Turing machine can).
- LSTM started to revolutionize speech recognition, outperforming traditional models in certain speech applications.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$



Data Science

- Applied Mathematics
- Computer Science
- Neuroscience
- Engineering
- Statistics
- Biology
- Physics

Data science is the application of computational and statistical techniques to address or gain insight into some problem in the real world (*J. Zico Kolter, Carnegie Mellon University, 2018*)

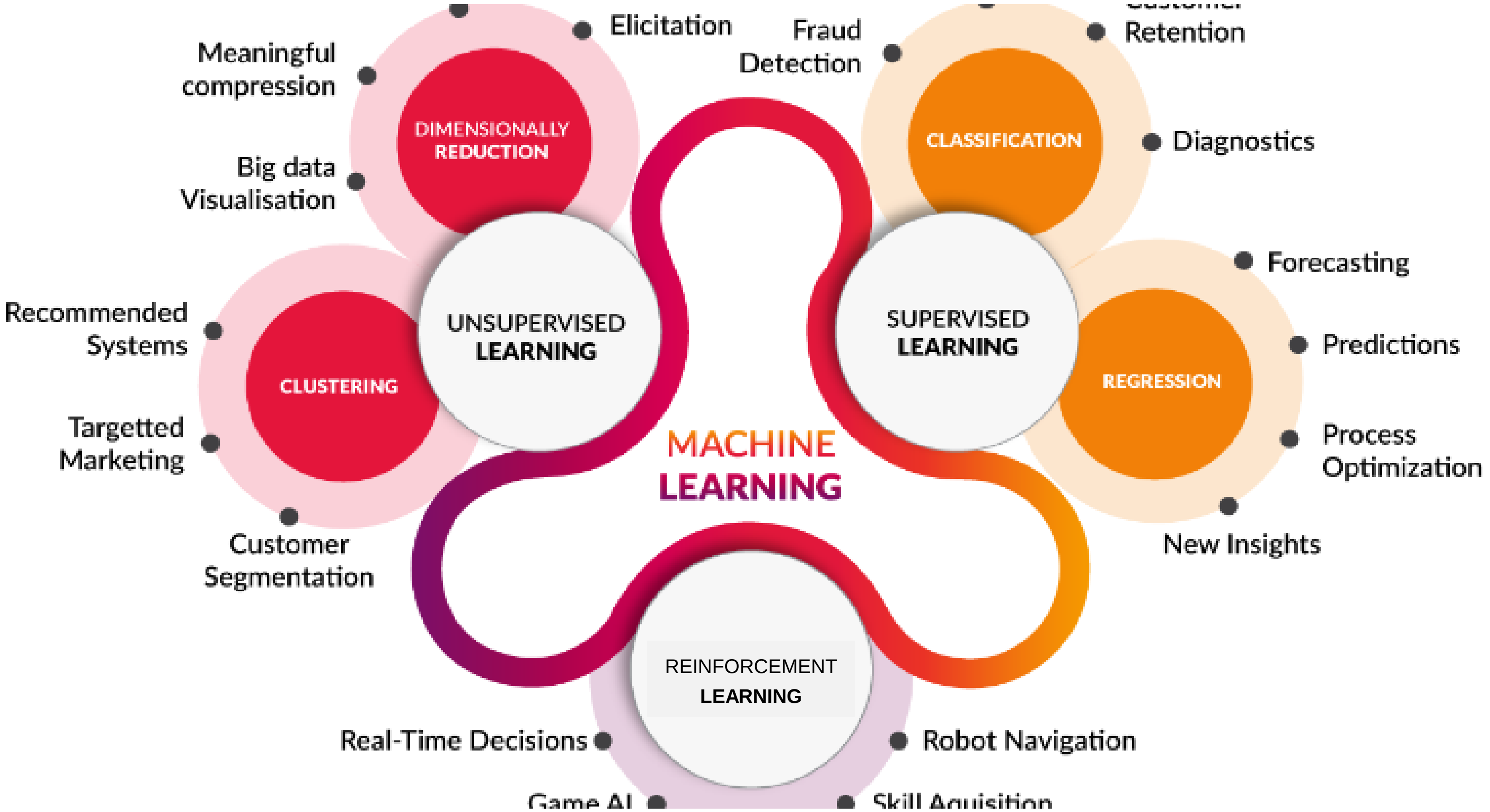
Data science is an extraordinary multidisciplinary and interdisciplinary challenge to apply the scientific method empowered by

- data,
- models,
- computational resources,
- open source software, and

… the most sophisticated technology we have today, i.e. the human being.

Machine Learning

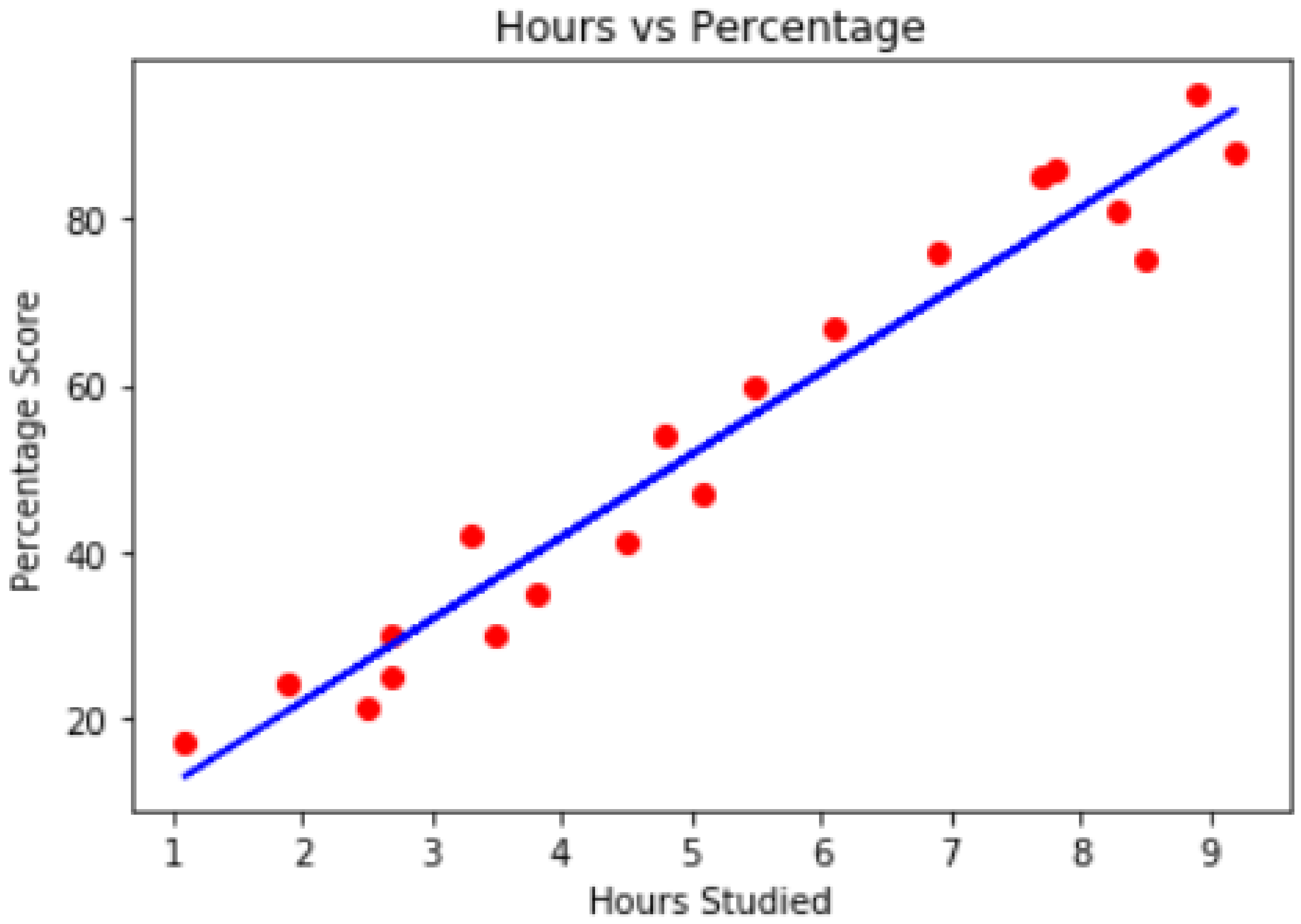
- Many different names for learning



But most of machine learning nowadays
is just **curve fitting**

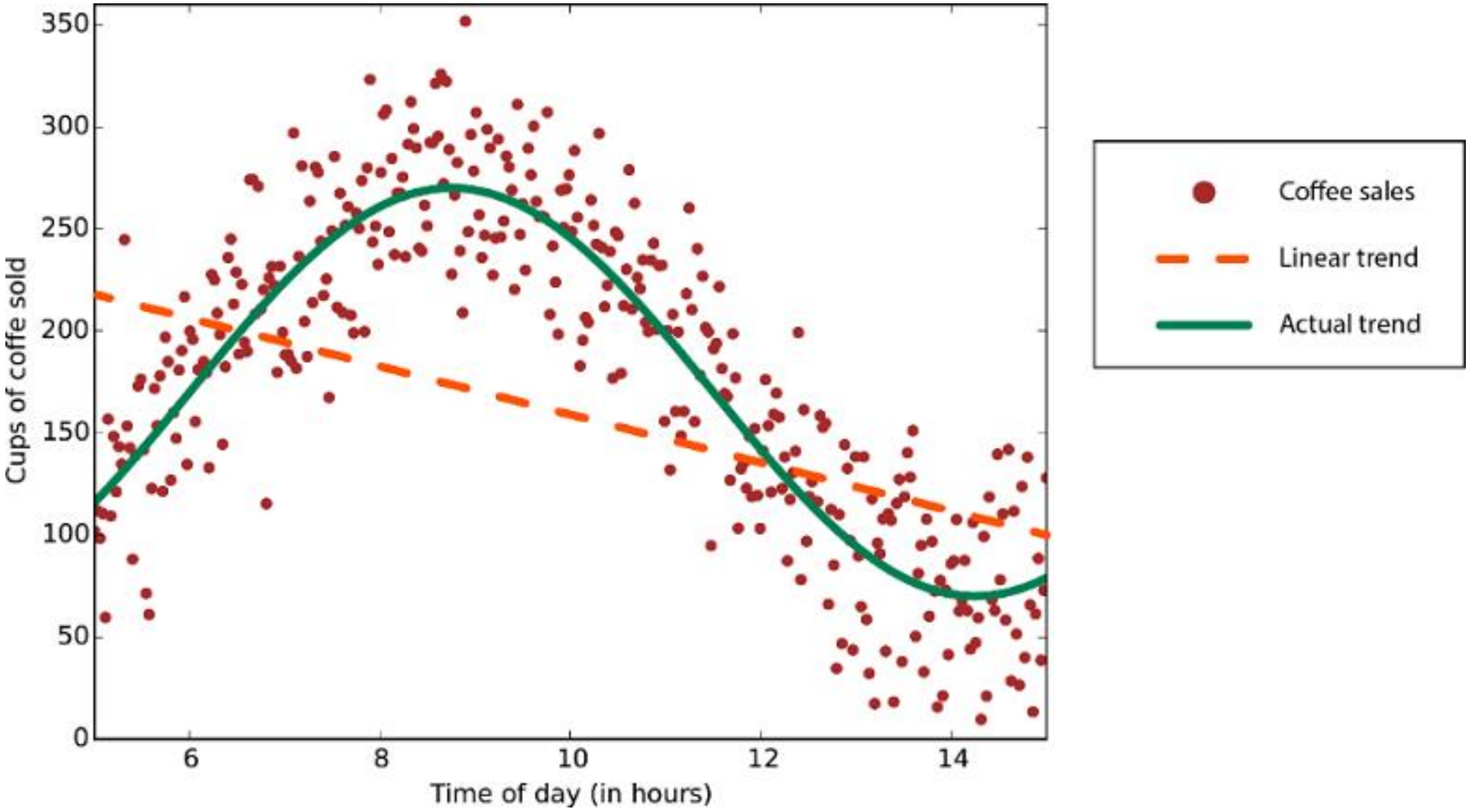
Machine Learning

- Curve fitting (correlations) - linear



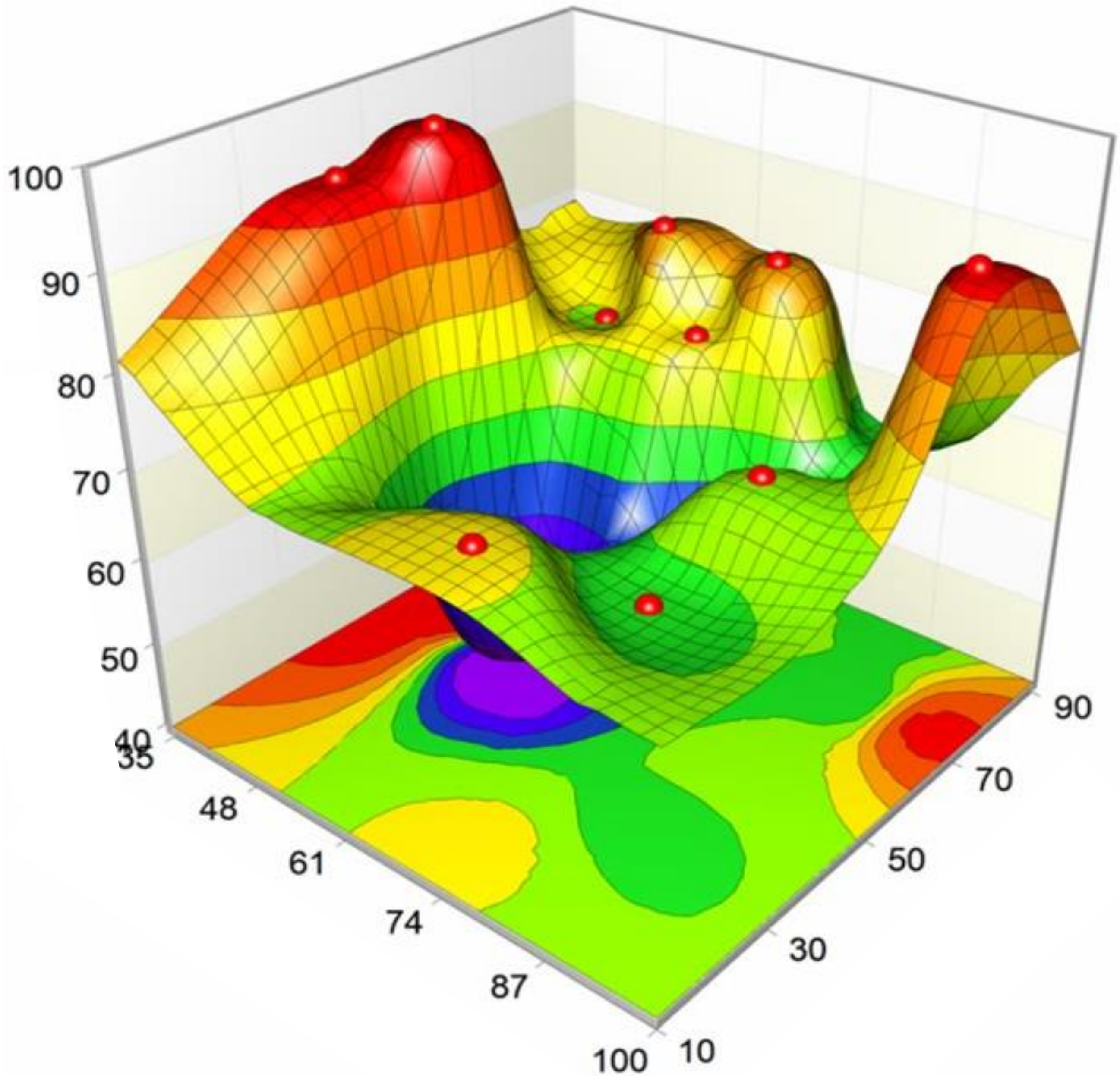
Machine Learning

- Curve fitting - nonlinear



Machine Learning

- Curve fitting - multidimensional



Machine Learning

- Deep Neural Networks

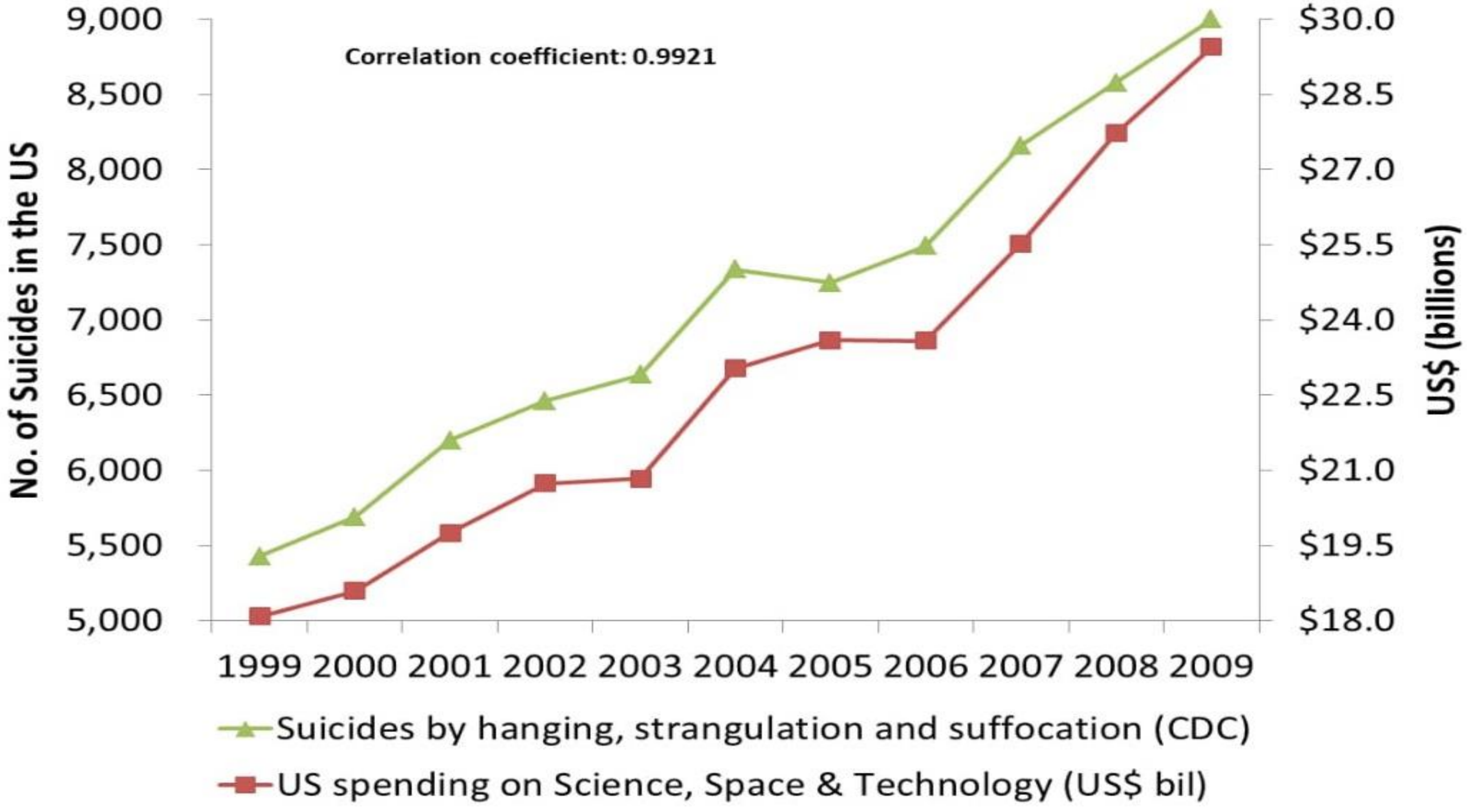
Highly dimensional, highly nonlinear

curve fitting

Machine Learning

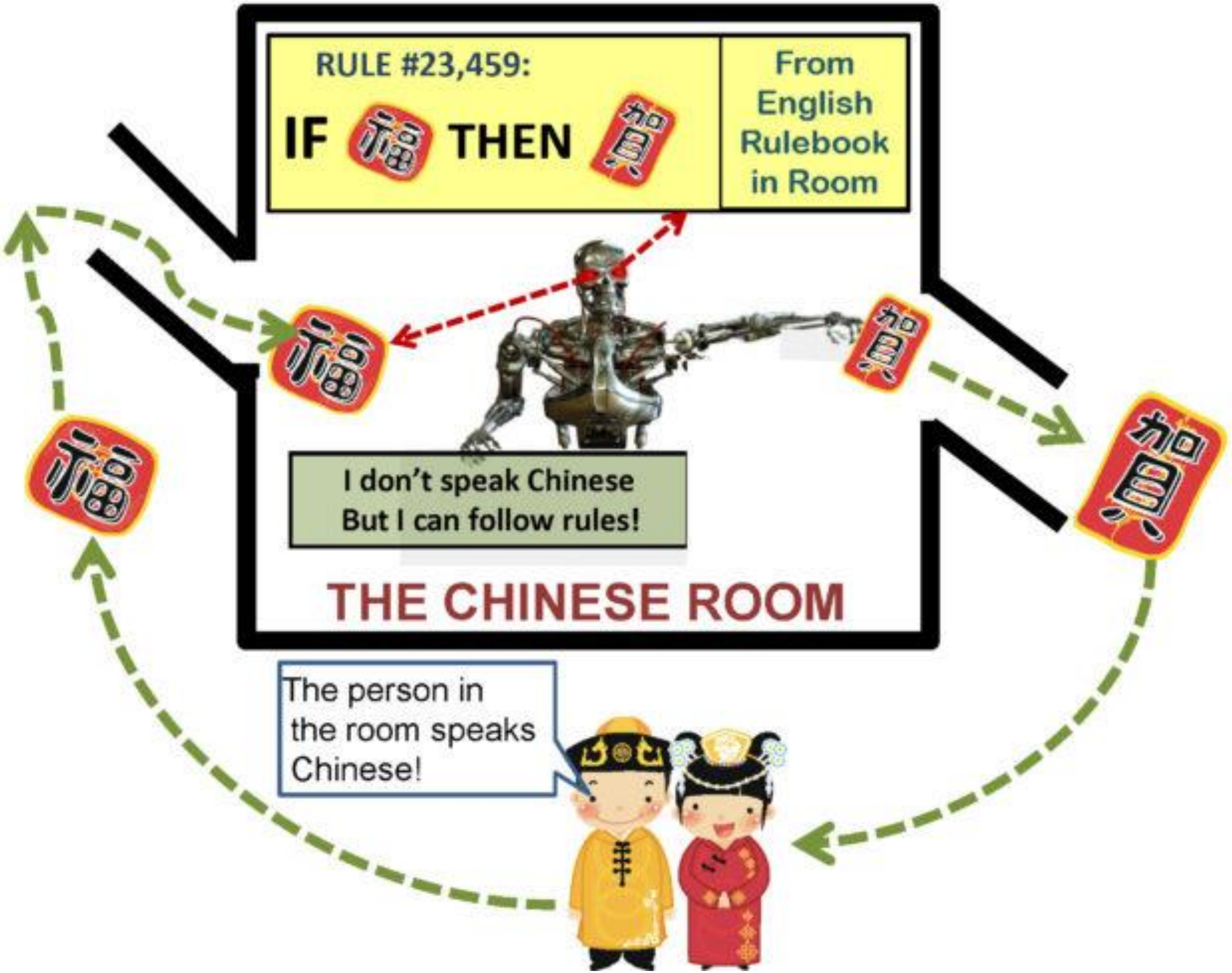
- Spurious Correlations

Fitting can be highly misleading



[Spurious Correlations]

Fitting does not give us any understanding



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 an eyeglass or a telescope or a microscope."

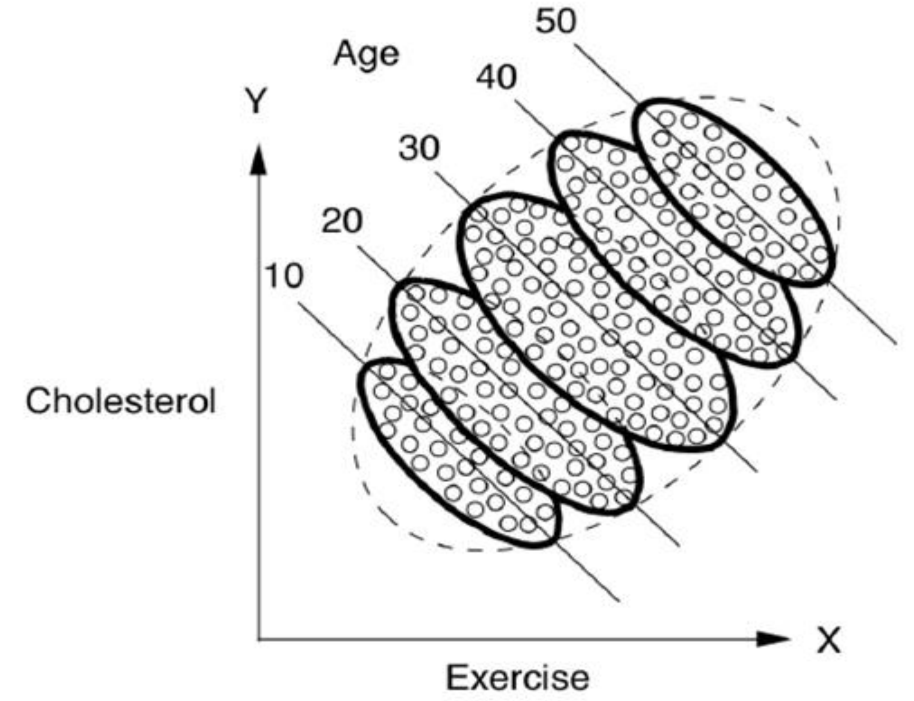
— *Judea Pearl*

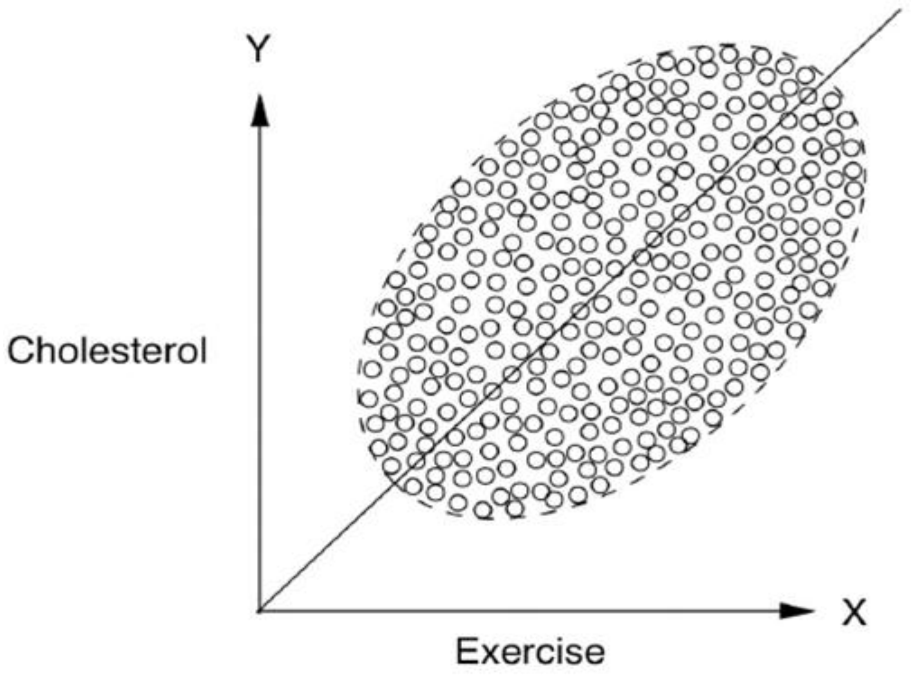
What can truly be achieved?

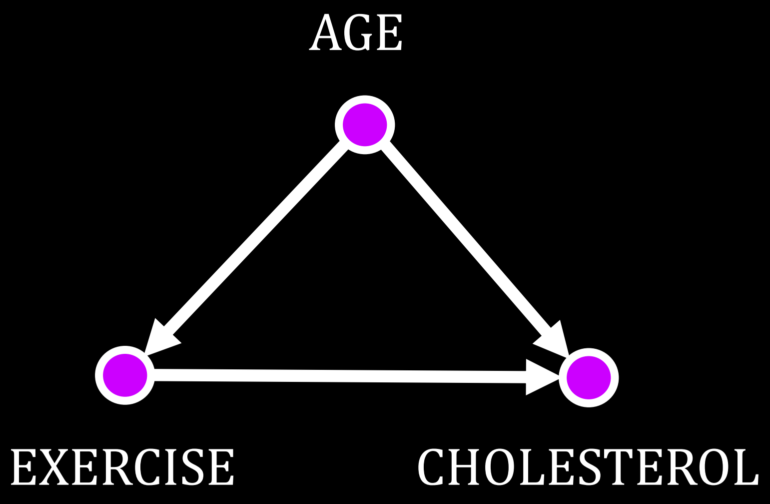
Does Exercise affect Cholesterol?

Simpson's Paradox

- No matter how much data you collect, the question can not be answered when using the data alone

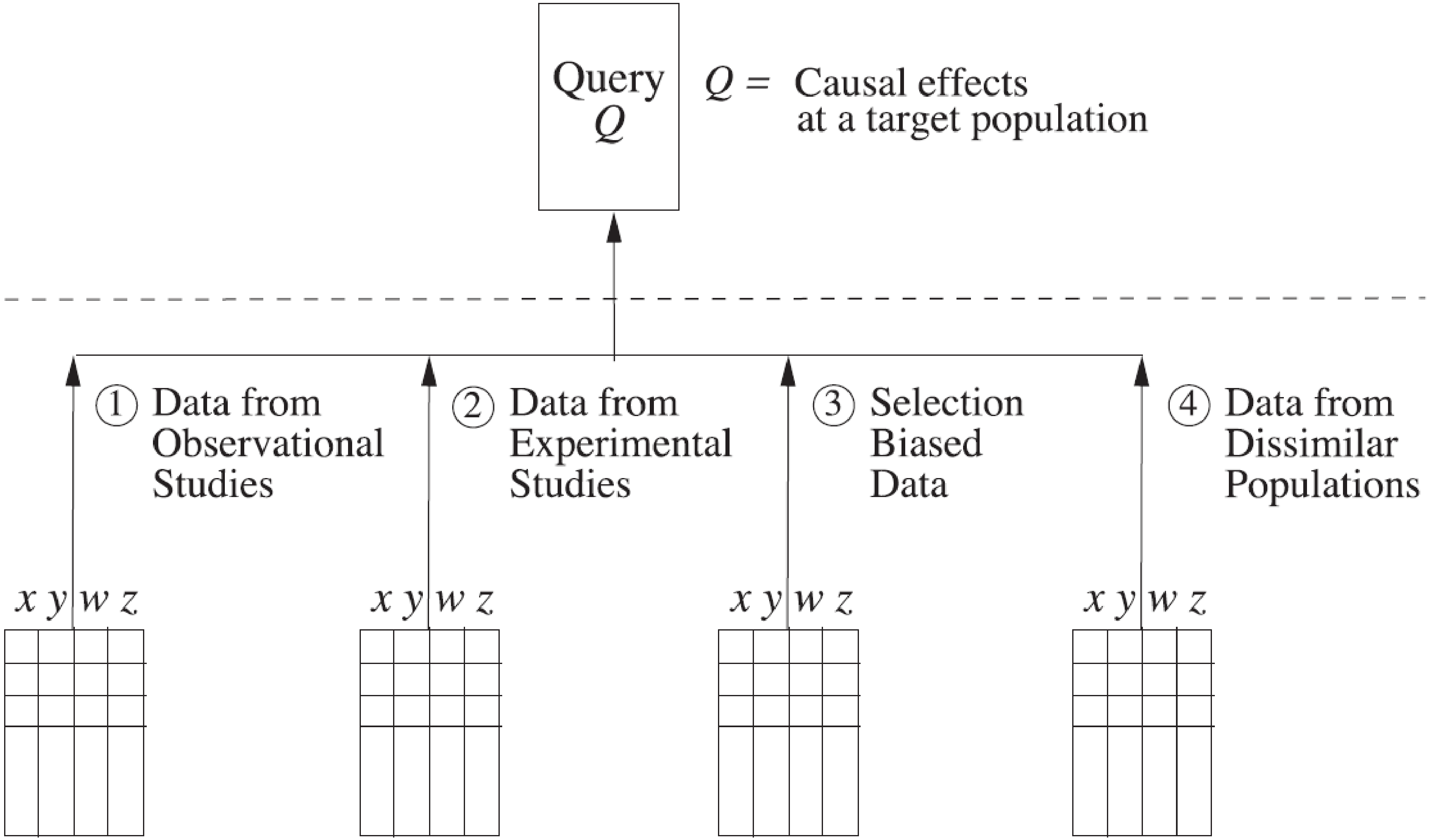






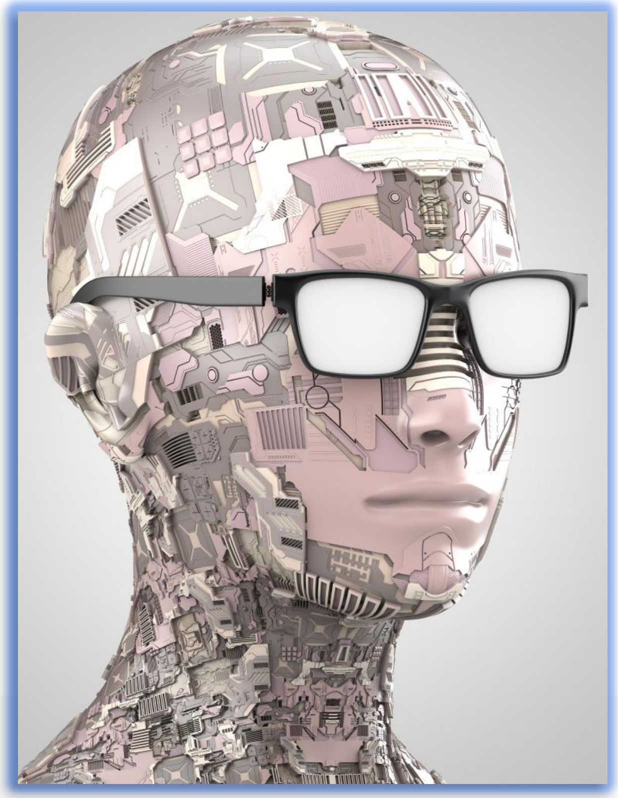
Big Data and the data-fusion problem

- Piecing together multiple datasets collected under heterogeneous conditions (i.e., different populations, regimes, and sampling methods) to obtain valid answers to queries of interest.
- The availability of multiple heterogeneous datasets presents new opportunities to big data analysts, because the knowledge that can be acquired from combined data would not be possible from any individual source alone.



Causation is the key

- What we (computers so far) miss is causal knowledge
- For predicting the consequences of actions
- Causation for us is a synonym of understanding
- Actual intelligence needs causal knowledge
- But causal knowledge is not in the data!



Introduction to Causal Inference

Why study causation?

- To make sense of data
  - effect of smoking on lung cancer?
  - effect of education on salaries?
  - effect of carbon emissions on the climate?
- To understand how we have an effect
  - malaria caused by mosquitos or by mal-air?

Why study causation?

- To make sense of data
  - effect of smoking on lung cancer?
  - effect of education on salaries?
  - effect of carbon emissions on the climate?
- To understand how we have an effect
  - malaria caused by mosquitos or by mal-air?
- To guide actions and policies
  - pack mosquito nets or use breathing masks?
  - reduce CO_2 emissions?
  - have a degree?
  - stop smoking?

Why study causation separately from statistics?

Why study causation separately from statistics?

- What can causation tell us that ML doesn't?
  - Causation is not an aspect of ML
  - It is an addition to ML
- Causation uncovers facts that ML cannot
  - None of the previous problems can be articulated in the language of ML
- Let us introduce these aspects with a famous statistical puzzle …

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!

The drug appears to help

men and women,

but hurt the 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%

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)

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

Should a doctor prescribe this drug or not?

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

- In the general population, the drug might improve recovery rates because of its effect on blood pressure.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

- In the general population, the drug might improve recovery rates because of its effect on blood pressure.
- in subpopulations—the group of people whose posttreatment BP is high and the group whose posttreatment BP is low—we, of course, would not see that effect; we would only see the drug's toxic effect.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

Should a doctor prescribe this drug or not?

Once again, the answer follows from the way the data were generated.

- In the general population, the drug might improve recovery rates because of its effect on blood pressure.
- in subpopulations—the group of people whose posttreatment BP is high and the group whose posttreatment BP is low—we, of course, would not see that effect; we would only see the drug's toxic effect.

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

Should a doctor prescribe this drug or not? **YES**

- Only by BP-segregating the data we can see the toxic effect
- It makes no sense to segregate the data; we should use the combined data

Data Segregation

- We have solved the problem using gender-segregated data
- Then let's just segregate the data whenever possible, right?

WRONG!!!

- Consider a drug affecting recovery by lowering blood pressure (BP)
- Unfortunately, it has also a toxic effect

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%

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

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

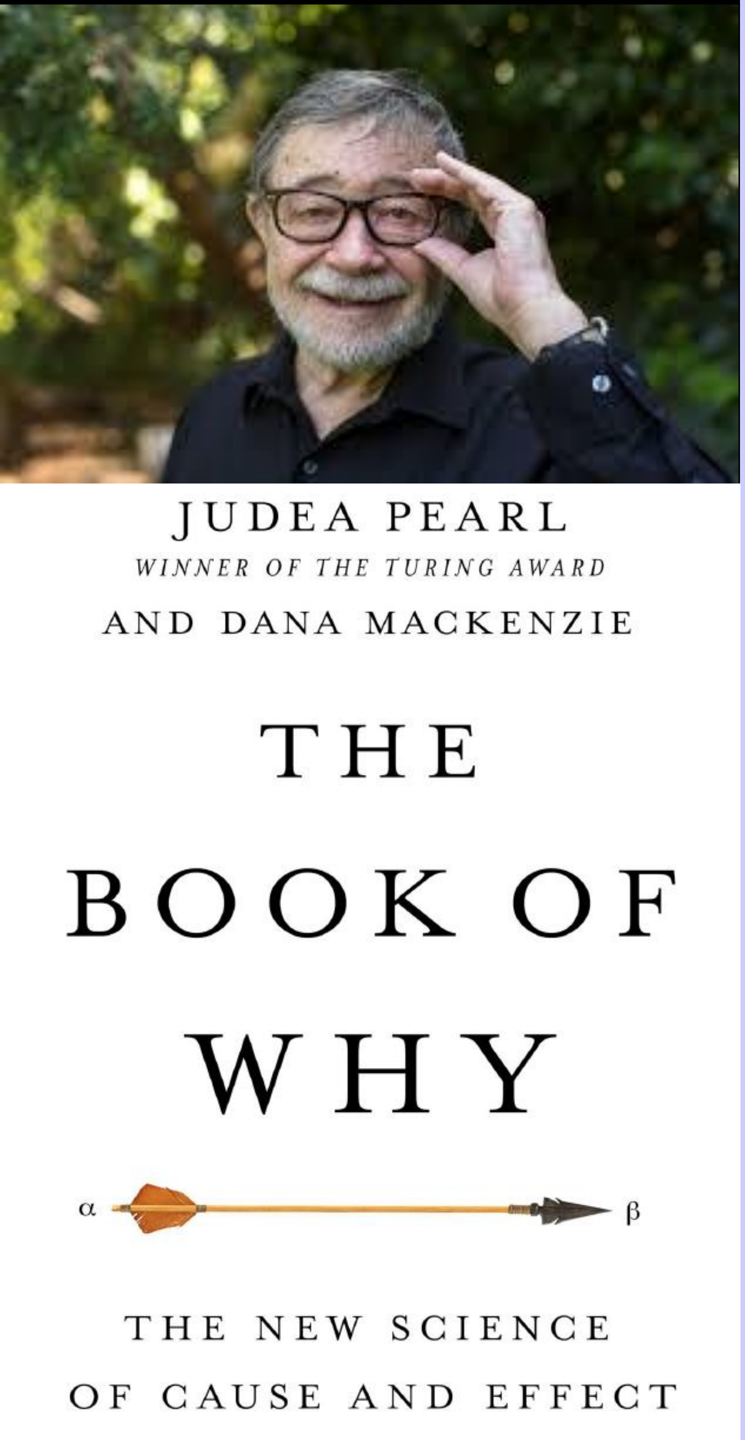
Should a doctor prescribe this drug or not? **YES**

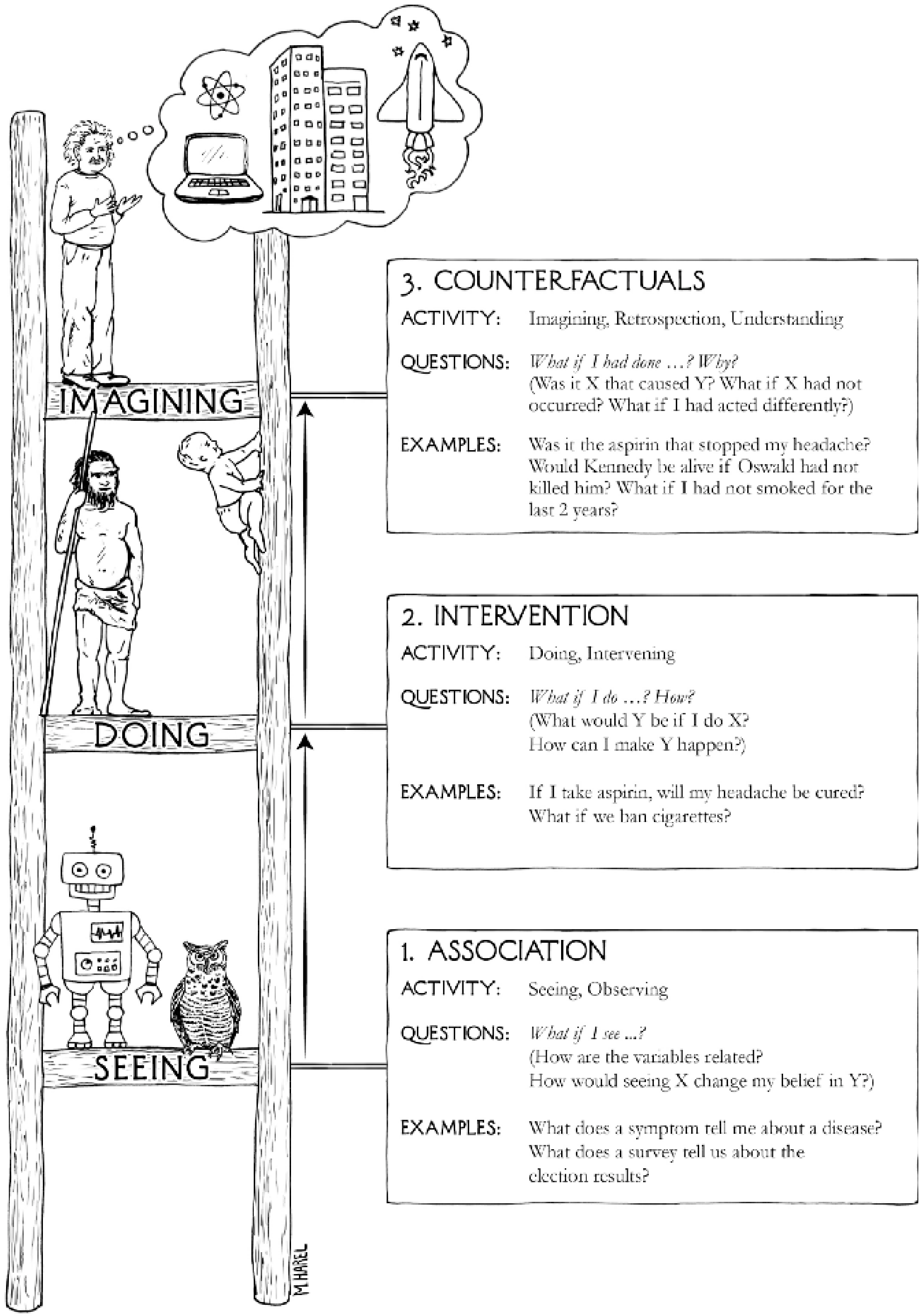
- Only by BP-segregating the data we can see the toxic effect
- It makes no sense to segregate the data; we should use the combined data

Note that the data are the same of Simpson's paradox.

Lessons Learned

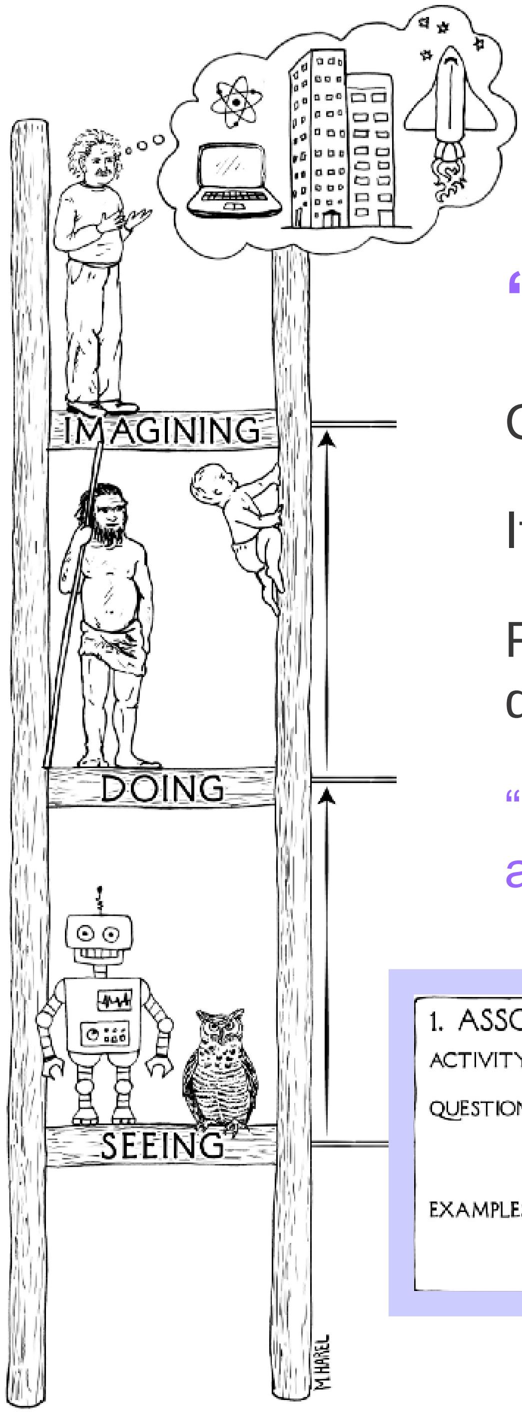
- **Information that allowed us to make a correct decision**
  - the timing of the measurements
  - that the treatment affects blood pressure
  - that blood pressure affects recovery
- **All this information was not in the data**
  - as statisticians rightly say, correlation is not causation
  - hence there is no method that can determine the causal story from data alone
  - whence no ML method can aid in our decision
- **The same holds for Simpson’s paradox**
  - the paradox arises out of our conviction that treatment cannot affect sex
  - if it could, we could explain it as in our blood pressure case
  - but we cannot test the assumption using the data





The Ladder of Causation

Seeing; we are looking for regularities in observations.



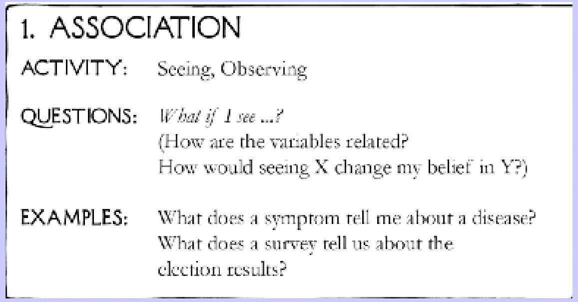
"What if I see …?"

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?"



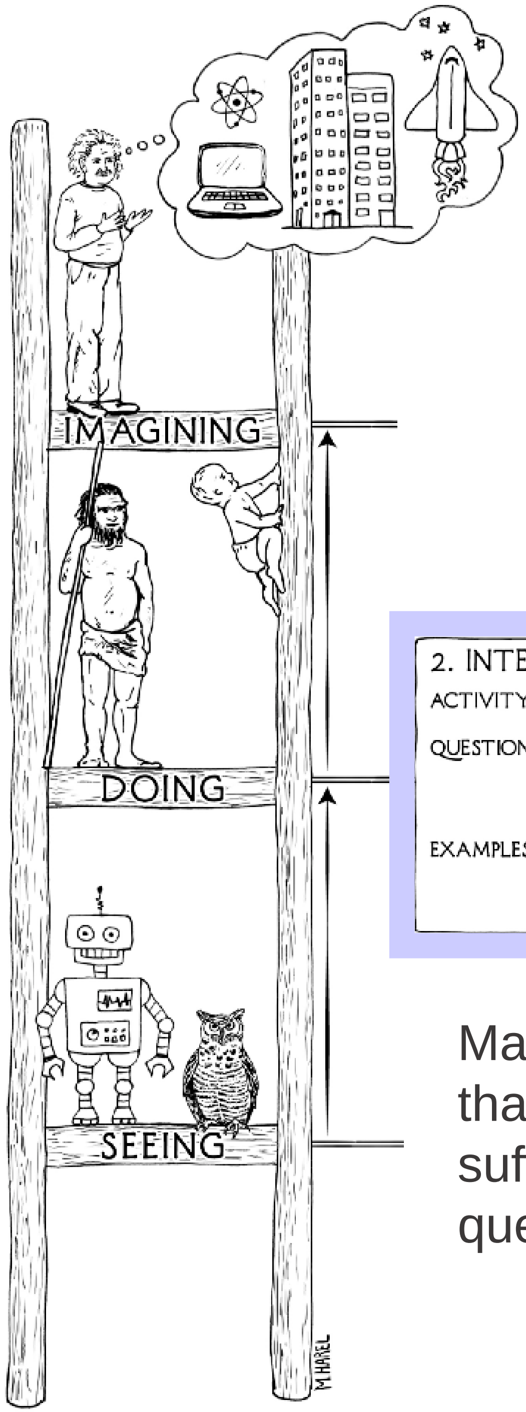
1. ASSOCIATION

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?

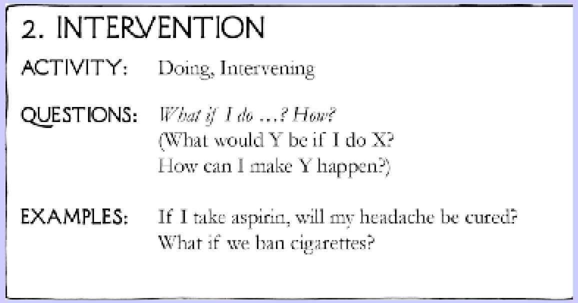
Intervention; ranks higher than association because it involves not just seeing but changing what is.



"What if 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?)

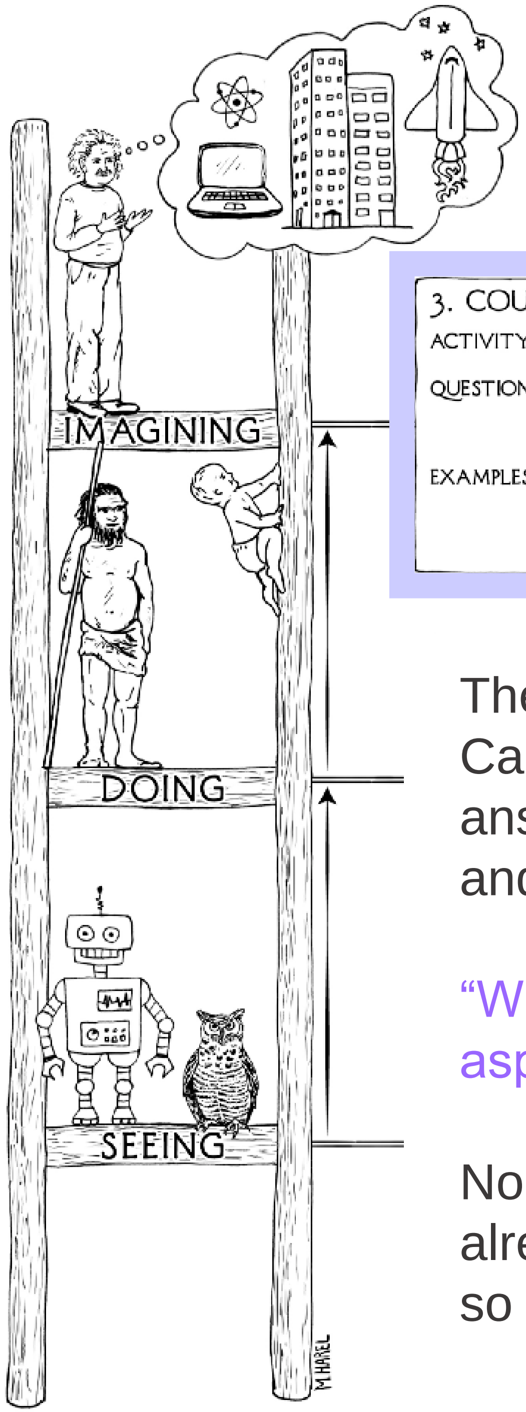
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

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.

Extra-Statistical Methods

- Methods to express and interpret causal assumptions
  - To describe problems of any complexity, where intuition no longer helps
  - To solve them mechanically as we solve algebraic equations
- In particular we need:
  1. A working definition of “causation”
  2. A method to articulate causal assumptions (i.e., a model)
  3. A method to link causal models to data
  4. A method to draw conclusions from model and data

Extra-Statistical Methods

- Methods to express and interpret causal assumptions
  - To describe problems of any complexity, where intuition no longer helps
  - To solve them mechanically as we solve algebraic equations

- In particular we need:
  1. **A working definition of “causation”**
  2. A method to articulate causal assumptions (i.e., a model)
  3. A method to link causal models to data
  4. A method to draw conclusions from model and data

- We start by 1
  - As simple as it can be, it has eluded philosophers for centuries
  - We use an operational definition:

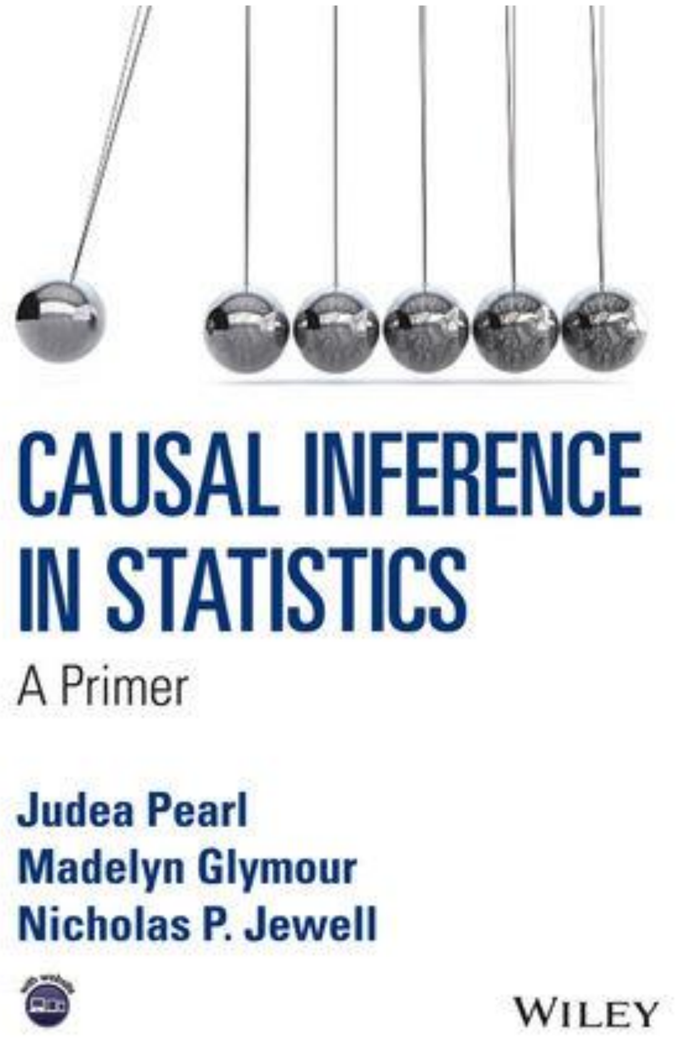
A variable *X* is a *cause* of a variable *Y* if *Y* “listens” to *X* and decides its value in response to what it hears.

Next Steps

Before moving on to actual causal methods we need:

- Some elementary concepts from probability and statistics
  - Most causal statements are uncertain (e.g., “careless driving causes accidents”)
  - Uncertainty is expressed by probability
  - Probability is at the heart of statistics (i.e., learning from data)
- Some graph theory
  - Causal stories will be represented by graphs
  - Graphs will be the basis of solutions methods

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."