Artificial Intelligence

A course about foundations

UNIVERSITÀ DI PAVIA

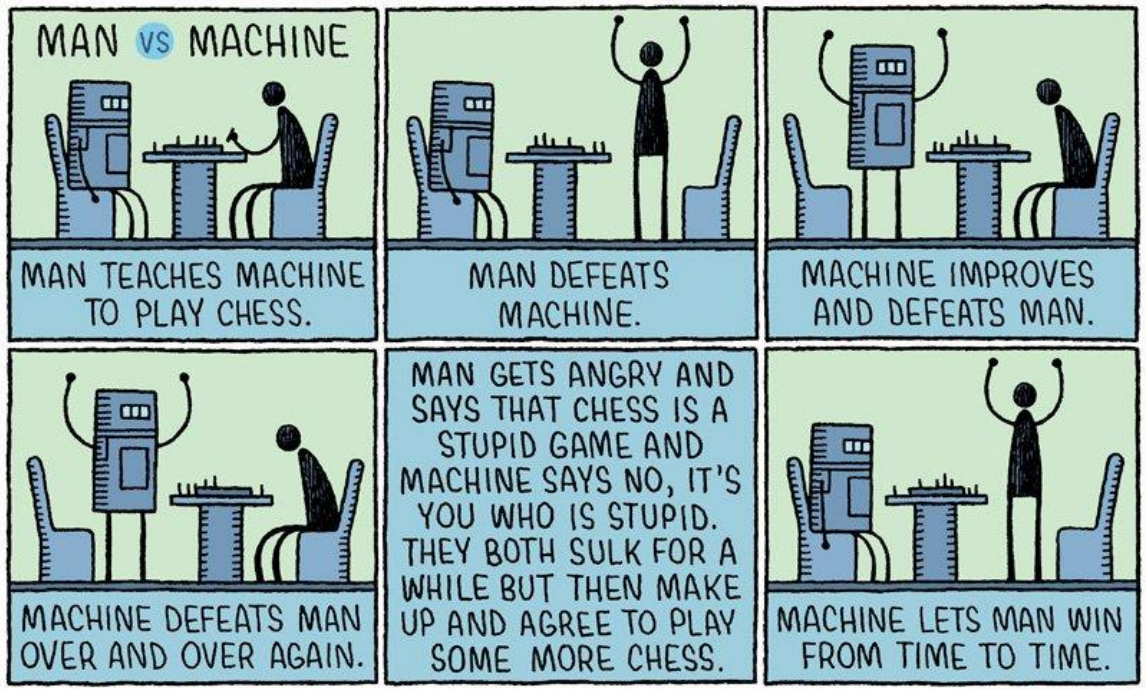
Introduction

Marco Piastra

Artificial Intelligence: *then and now*

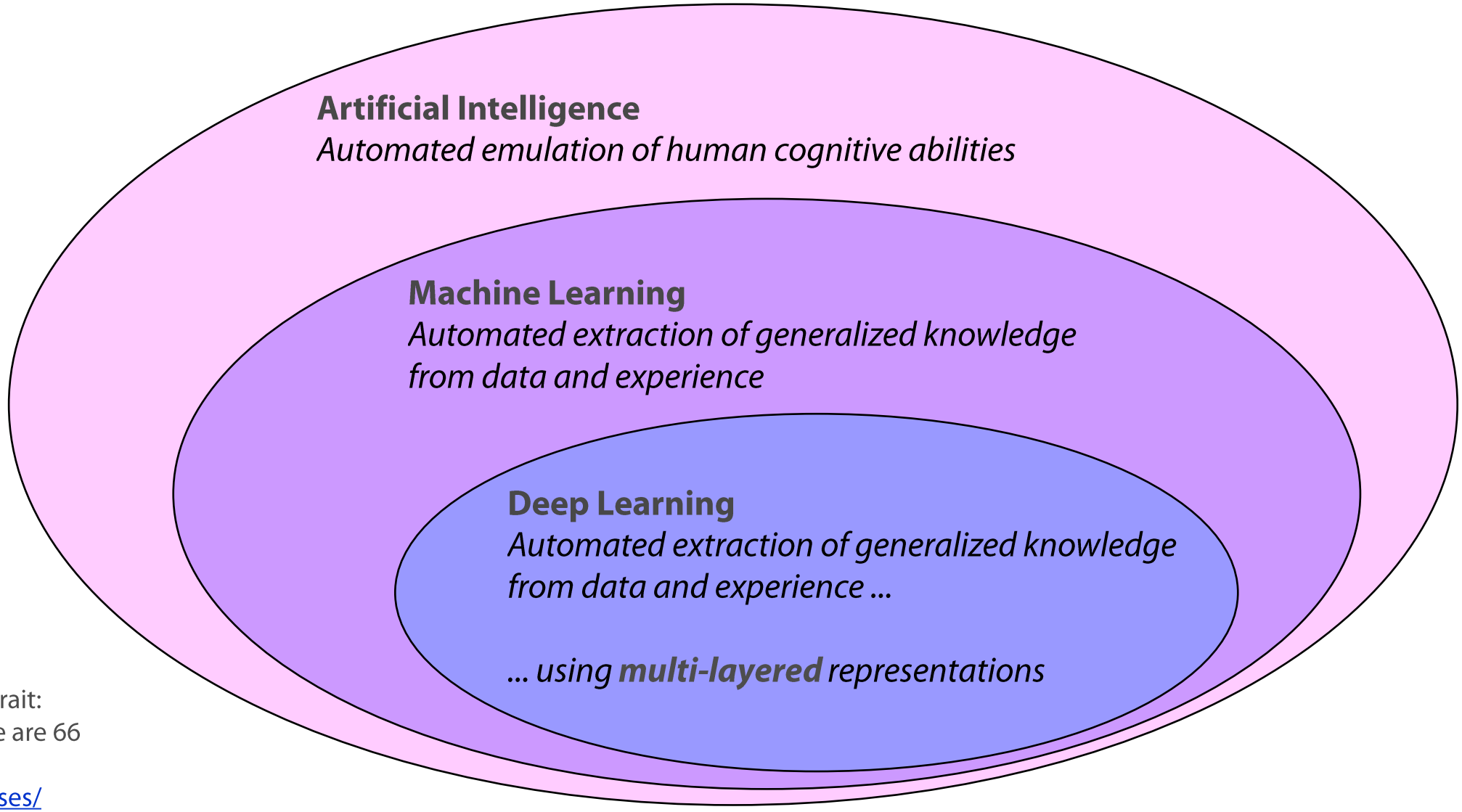
[Image from https://www.tomgauld.com/portfolio]

Computers play chess



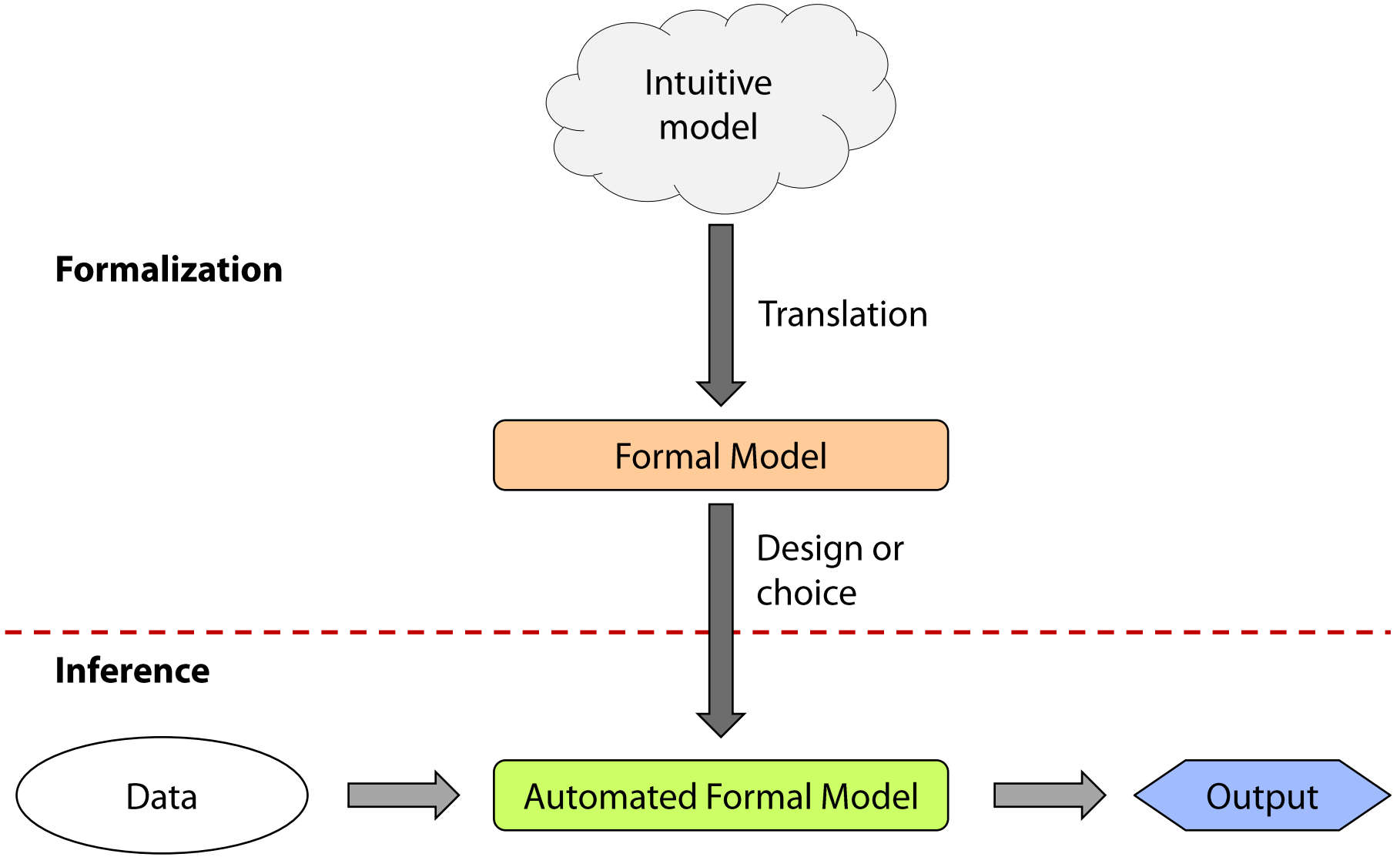
[Image from https://www.tomgauld.com/portfolio]

The Artificial Intelligence Cosmos

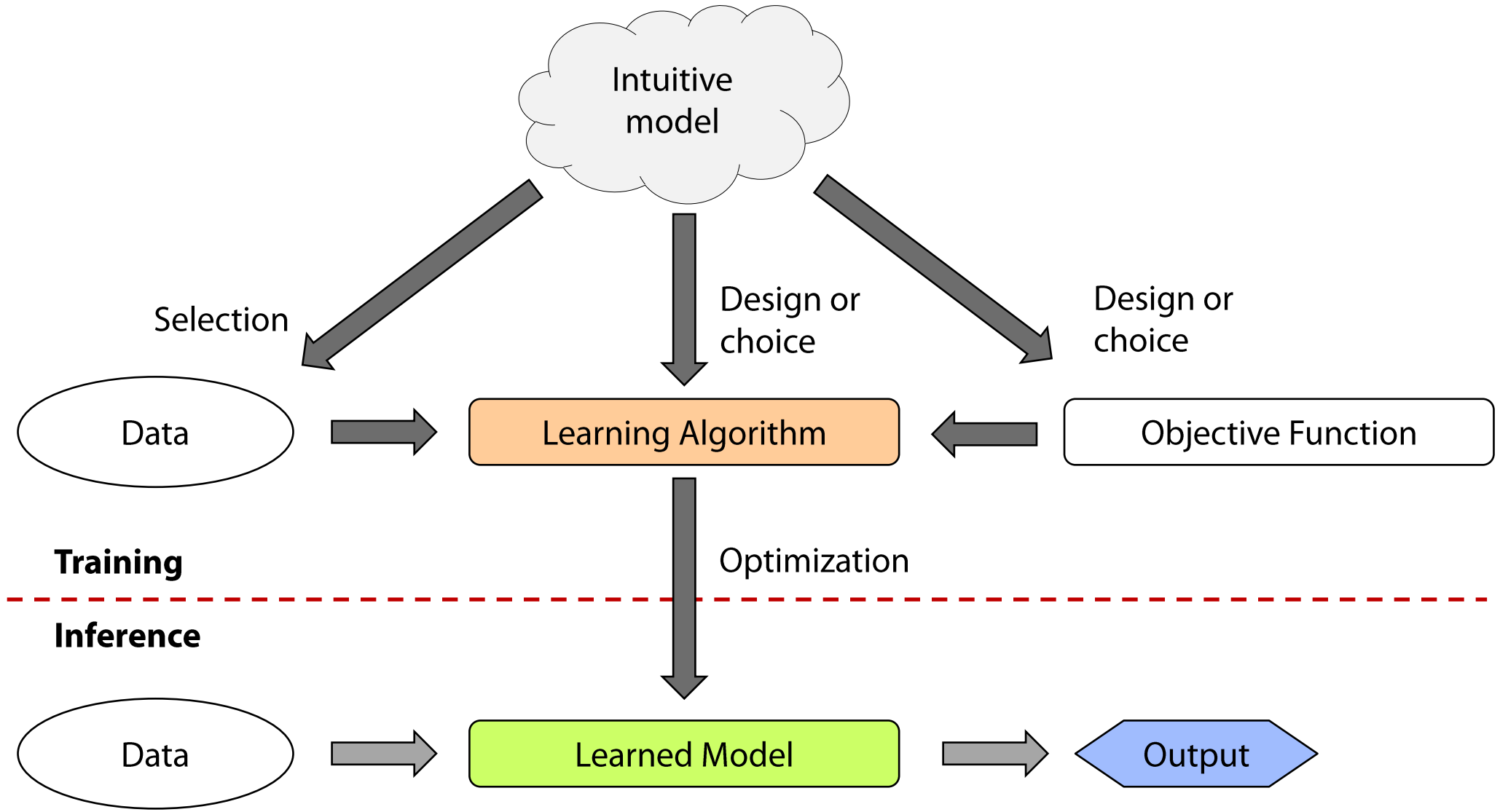


- This is a very simplified portrait: at Stanford, at present, there are 66 AI-related graduate courses https://ai.stanford.edu/courses/

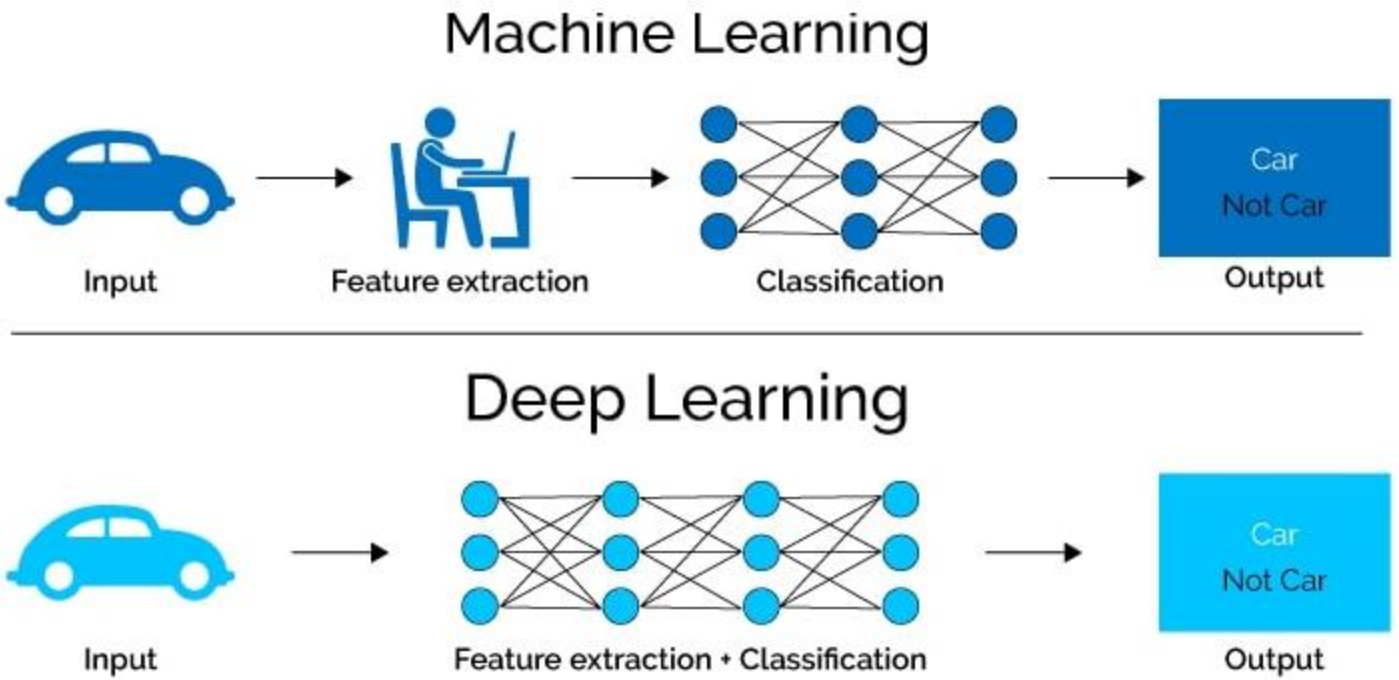
AI Symbolic Framework

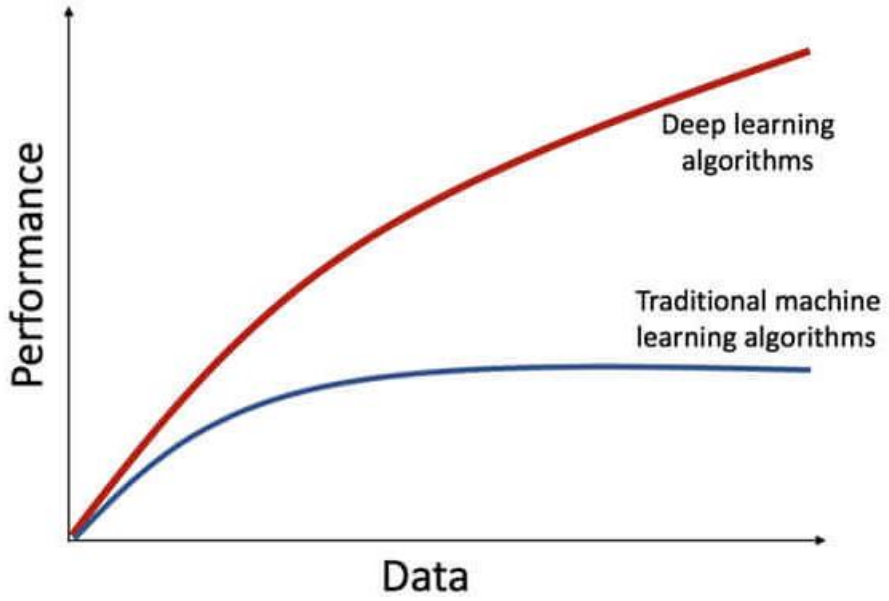


Machine Learning



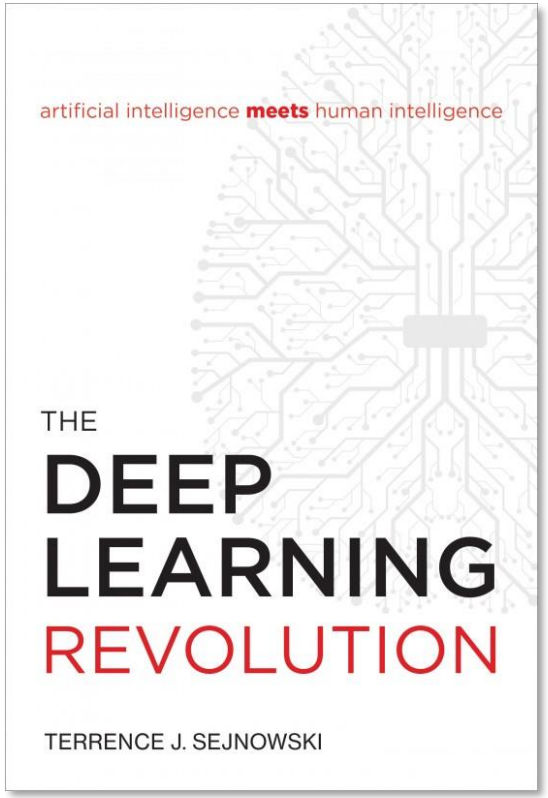
Machine Learning vs Deep Learning





Prologue:
what this course is NOT about

Deep Learning: AI Revolution?



- Terrence J. Sejnowski [President of the Neural Information Processing (NIPS) Foundation, October 2018]

  "How deep learning—from Google Translate to driverless cars to personal cognitive assistants—is changing our lives and transforming every sector of the economy."

  "AI is now awakening and transforming our world.
  Driving these breakthroughs is the deep learning revolution [...]"

[Quote and image from https://mitpress.mit.edu/books/deep-learning-revolution]

La fiera delle meraviglie: ChatGPT

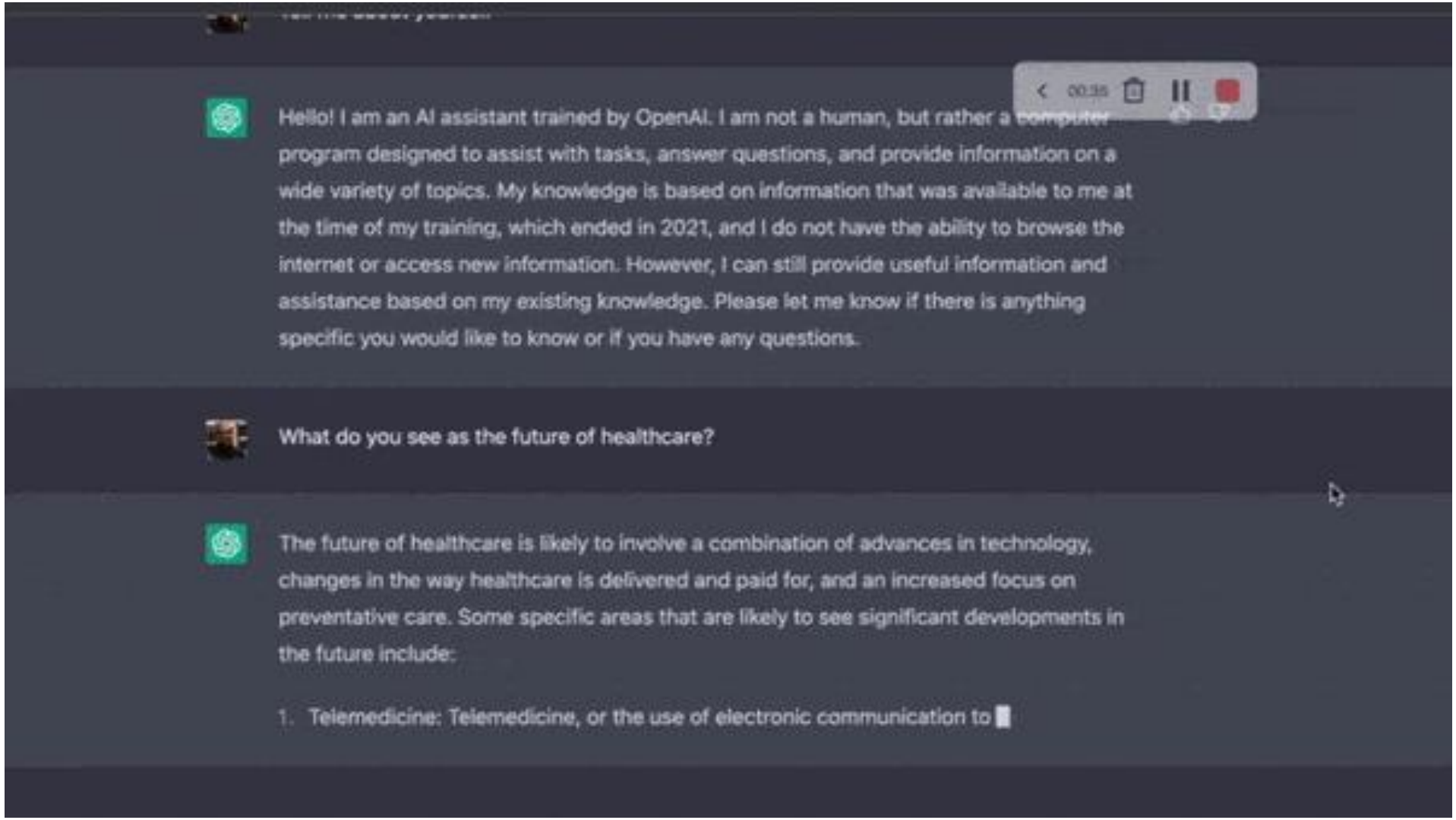
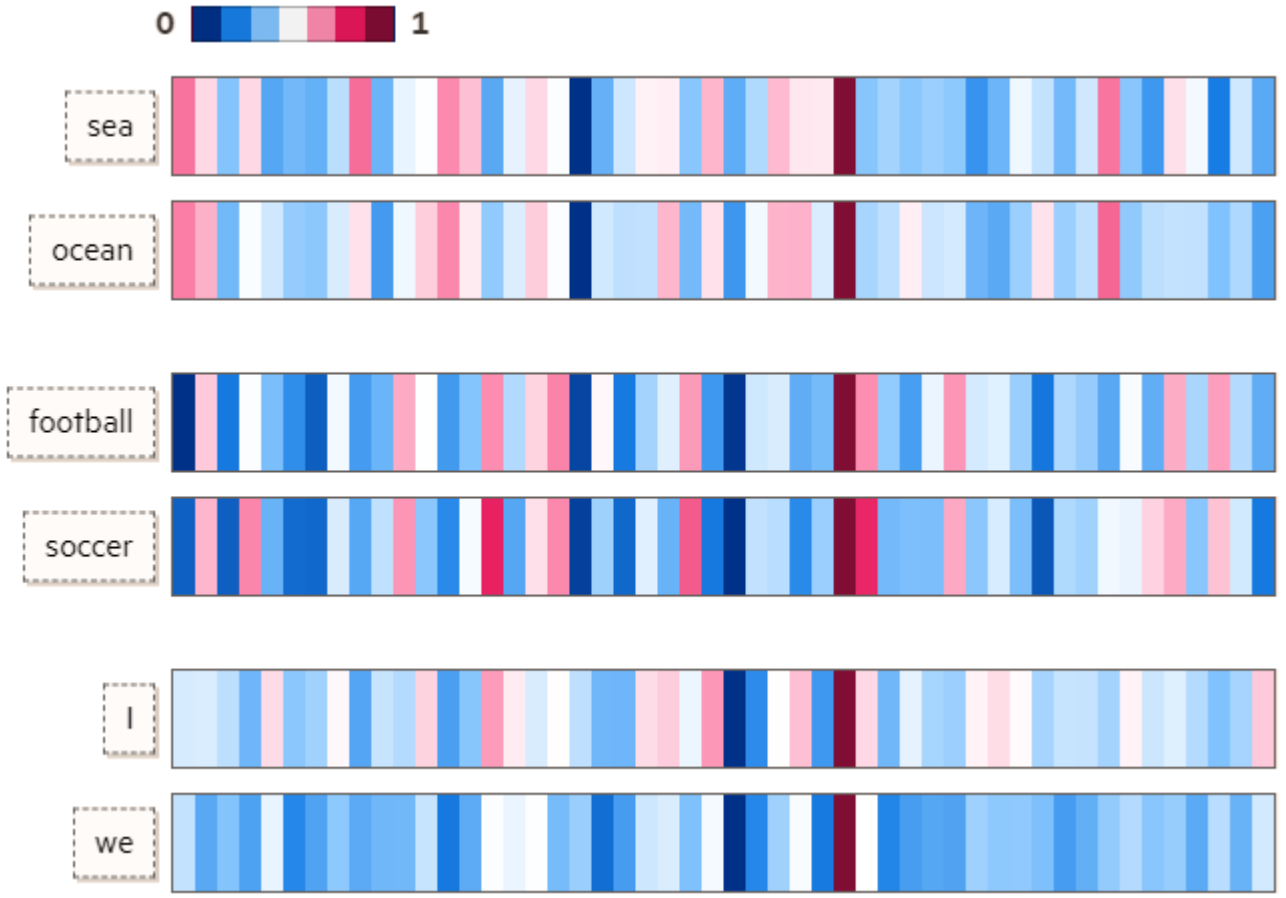


Image from https://info.homecareinnovationforum.com/the-potential-of-chatgpt-in-home-health-care

How does ChatGPT work?

- **Positional encoding (*Embedding*)**

  Each word (=*token*) in natural language are translated into a numerical vector in high dimensions



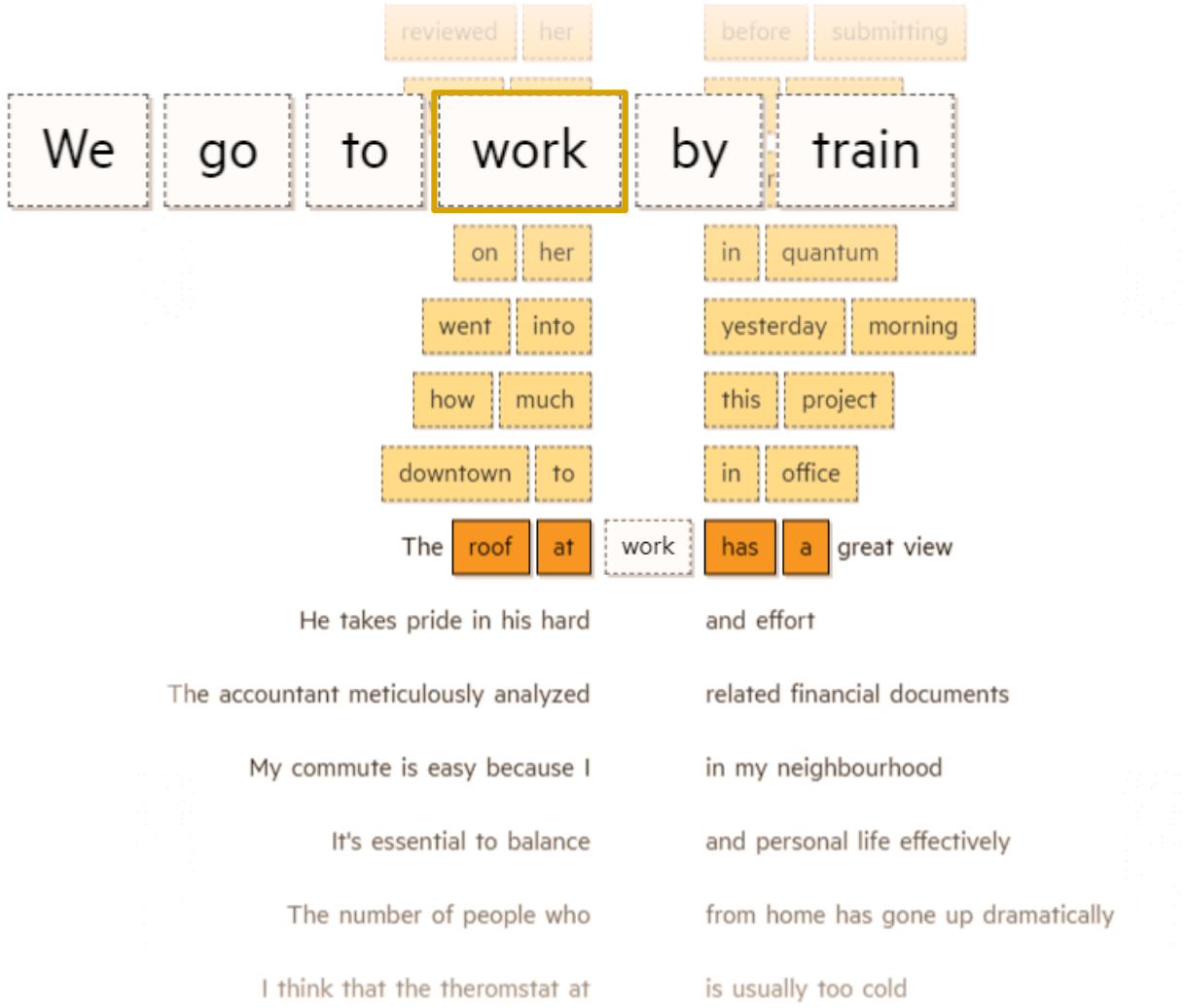
Images from https://ig.ft.com/generative-ai/

How does ChatGPT work?

- **Positional encoding (*Embedding*)**

  Each word (=*token*) in natural language are translated into a numerical vector in high dimensions

  Each vector is computed on the basis of *co-occurrence* of the corresponding word with respect to others, in a large text corpus



Images from https://ig.ft.com/generative-ai/

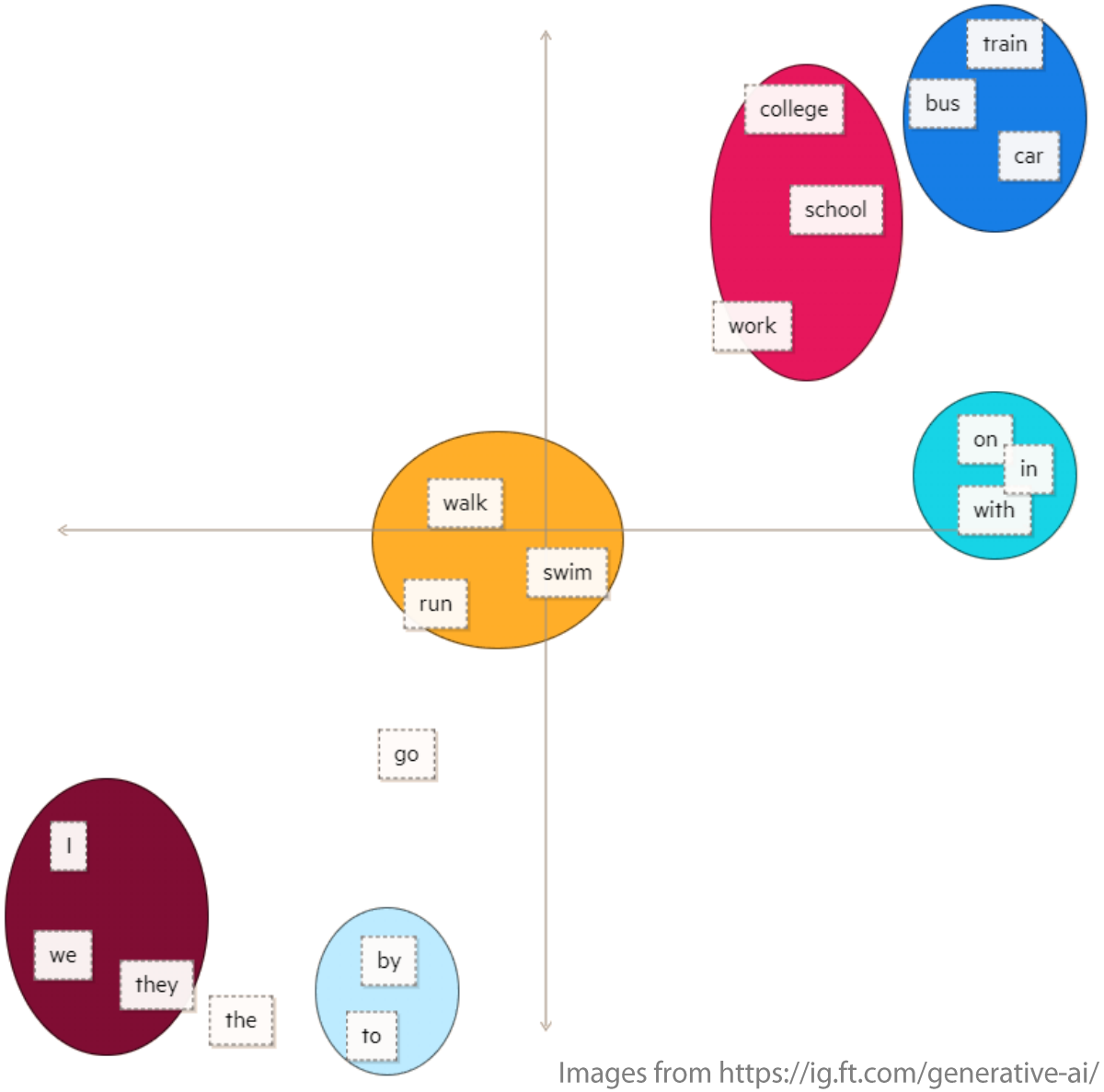
How does ChatGPT work?

- **Positional encoding (*Embedding*)**

  Each word (=*token*) in natural language are translated into a numerical vector in high dimensions

  Each vector is computed on the basis of *co-occurrence* of the corresponding word with respect to others, in a large text corpus

  In this way, numerical closeness among vectors reflects similarity in terms of role or meaning (or both)



Images from https://ig.ft.com/generative-ai/

How does ChatGPT work?

- **Input-output relation**

  Once encoded into vectors, sentences are fed as input to a very sophisticated *artificial neural network*

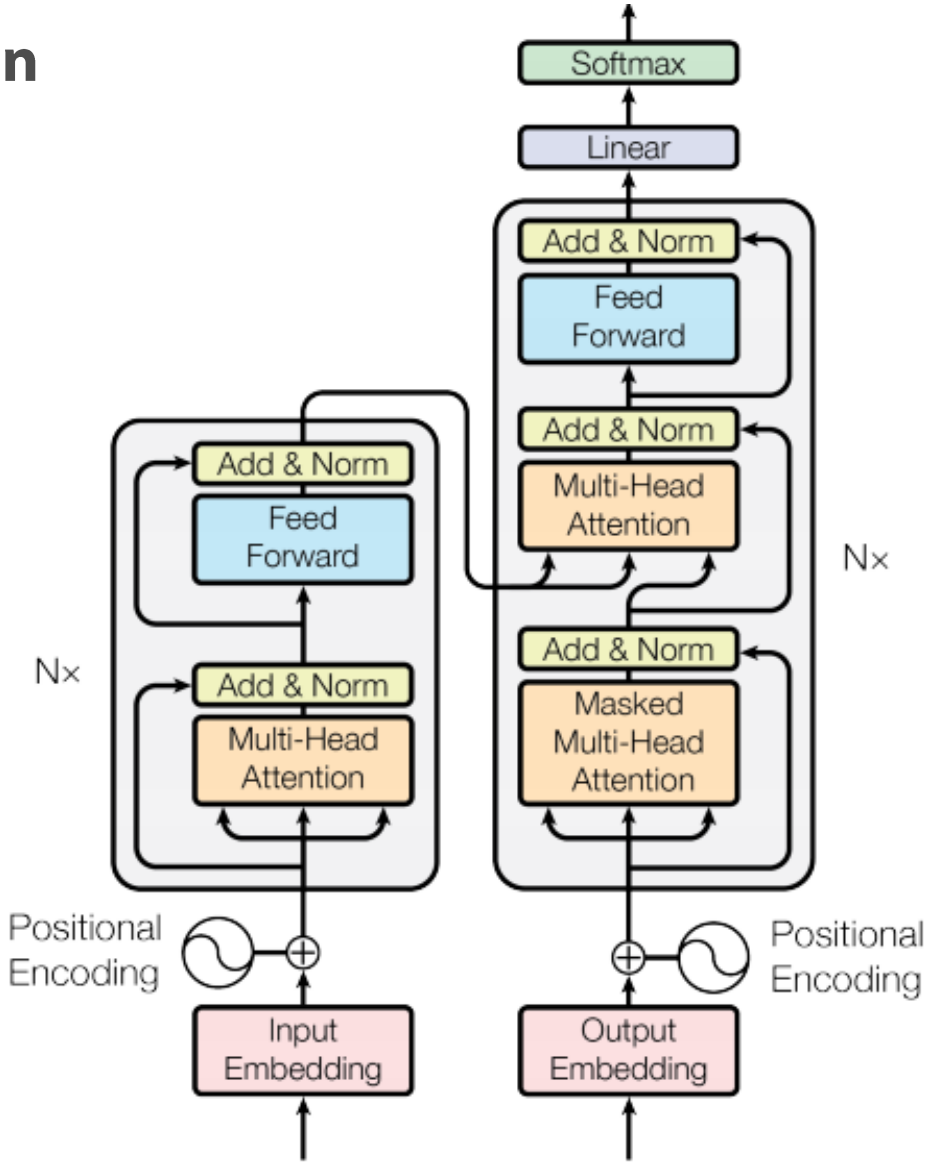


Image from https://arxiv.org/abs/1706.03762

How does ChatGPT work?

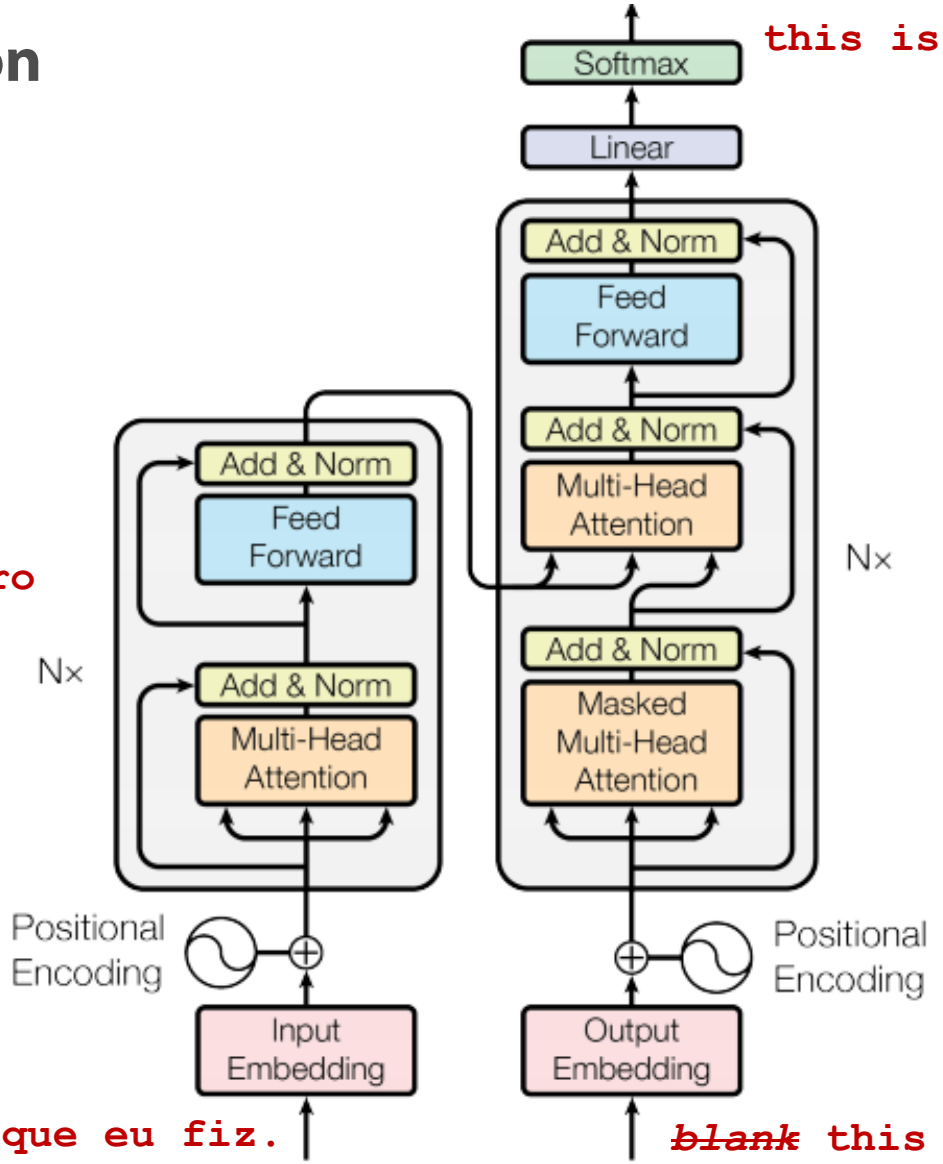
- **Input-output relation**

  Once encoded into vectors, sentences are fed as input to a very sophisticated *artificial neural network*

  Example: translating into English the following sentence in Portuguese:

  este é o primeiro livro que eu fiz.

  Output from the network is fed back as input to the network itself



this is the first book i've ever done.

este é o primeiro livro que eu fiz.

~~blank~~ this is the first book i've ever done

Image from https://arxiv.org/abs/1706.03762

How does ChatGPT work?

- **Input-output relation**

  Once encoded into vectors, sentences are fed as input to a very sophisticated *artificial neural network*

  The *artificial neural network* produces as output a *probability distribution* for either the next word or an entire sequence of words (*beam search*)

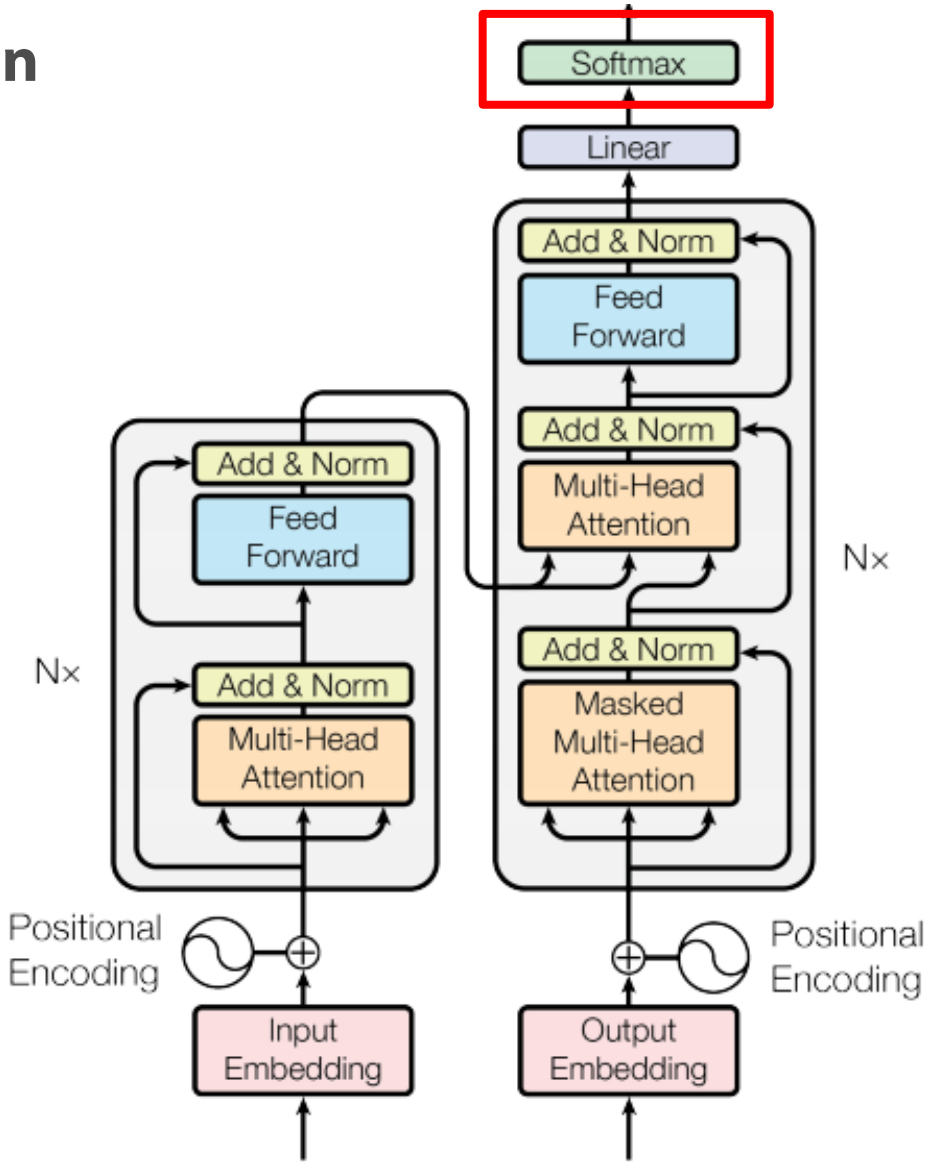
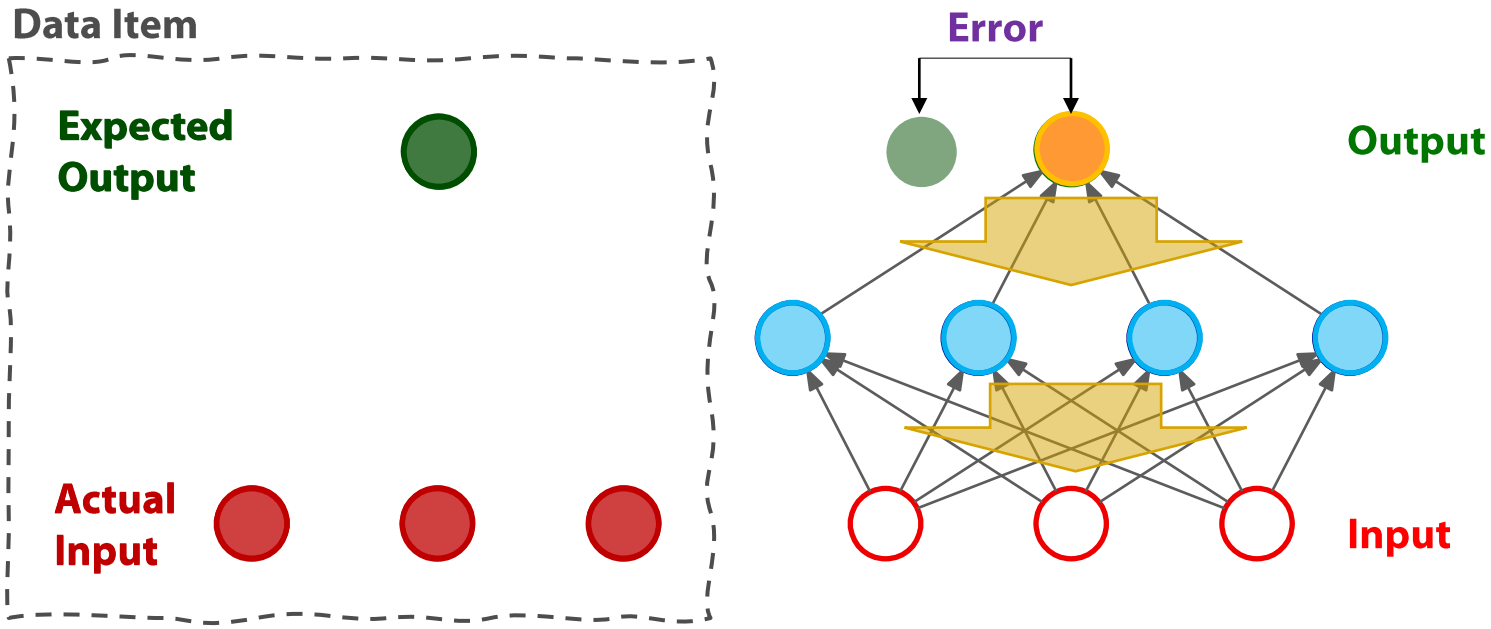


Image from https://arxiv.org/abs/1706.03762

Artificial Neural Networks

- ***Learning* is an *optimization* process of numerical parameters**

  Using a very large dataset, made of data plus *annotations* (*data items*)



Input data
are fed as input

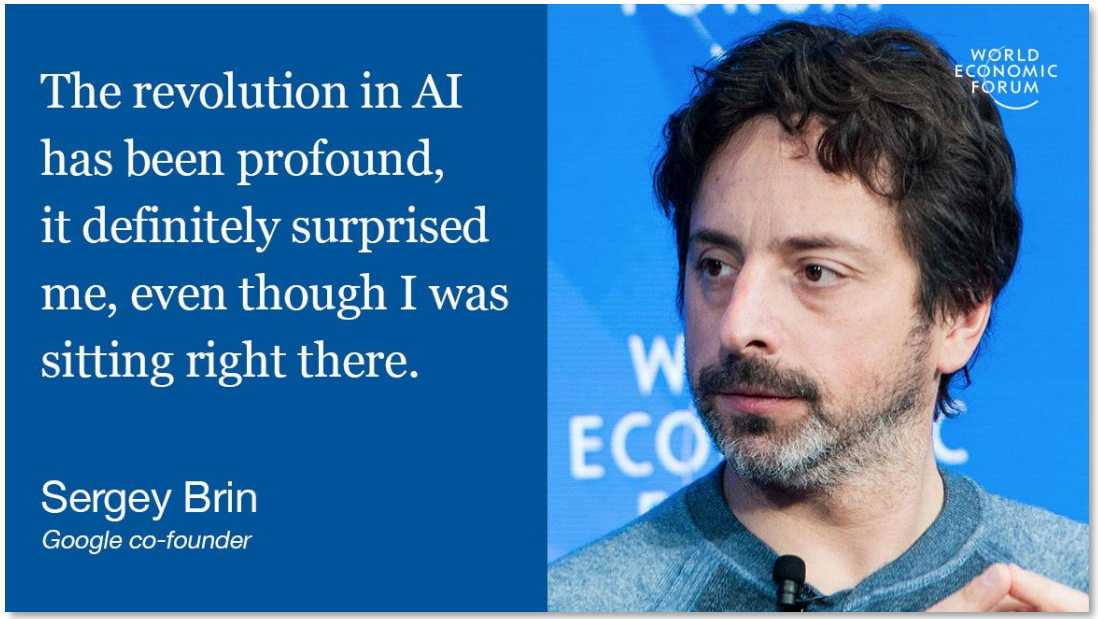
Input is propagated forward
to compute output

Error is propagated backward
to improve *parameters*

Overall method:

- *show a data item*
- *improve*
- *repeat*
- *several million times ...*

AI revolution?



- Sergey Brin [Google Co-Founder, January 2017]

  "I didn't pay attention to it [i.e. Artificial Intelligence] at all, to be perfectly honest."

  "Having been trained as a computer scientist in the 90s,
  everybody knew that AI didn't work.
  People tried it, they tried neural nets and none of it worked."

[Quote and image from https://www.weforum.org/agenda/2017/01/google-sergey-brin-i-didn-t-see-ai-coming/]

AI revolution?

An artificial neural network is capable to abstract an *associative function* (input-output) from a very large quantity of annotated data

- ***Why they work now (and did not before)***

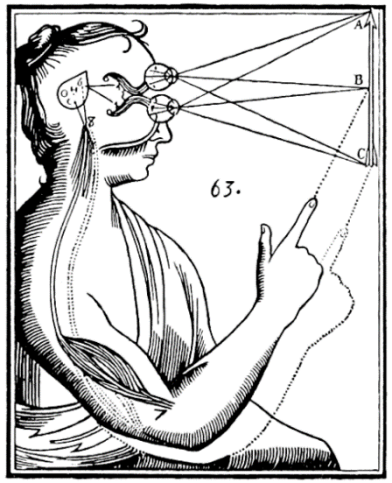
  Recent progresses are due:
  - better mathematical assessement, for both representation and optimization
  - substantial improvement in parallel computational power
  - introduction of newer and more sophisticated, layered network architectures
  - vast number of numerical parameters
    (ChatGPT has 1.5 billions)

OK, what this course IS about, then?

"Inference"
(the very idea)

Bits of AI History: Can machines think?

Mind vs. Brain



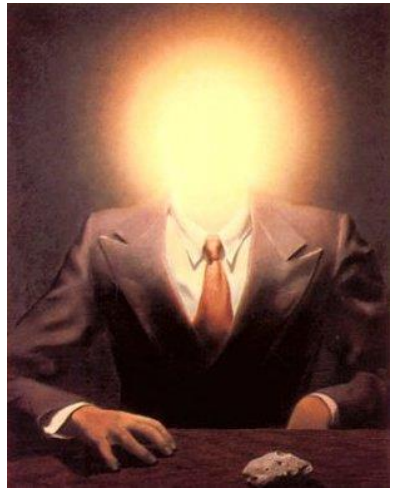
(from Wikipedia)

"I had after this described the **reasonable soul**, and shown that *it could by no means be educed from the power of matter*, as the other things of which I had spoken, but that it must be expressly created;

and that it is not sufficient that it be lodged in the human body exactly like a pilot in a ship, unless perhaps to move its members,

but that it is necessary for it to be joined and united more closely to the body, in order to have sensations and appetites similar to ours, and thus constitute a real man"

[Descartes, R., Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences, 1637, - English version from Project Gutenberg]



[Image from Wikipedia]

"Because we do not understand the brain very well
we are constantly tempted to use the latest technology
as a model for trying to understand it.

In my childhood we were always assured that the brain
was a telephone switchboard ('*What else could it be?*').

I was amused to see that Sherrington, the great British neuroscientist, thought that the brain worked like a telegraph system. Freud often compared the brain to hydraulic and electro-magnetic systems. Leibniz compared it to a mill, and I am told some of the ancient Greeks thought the brain functions like a catapult.

At present, obviously, the metaphor is the digital computer."

[Searle, J. R., *Minds, Brain and Science*, 1986]

Turing Machine (A. Turing, 1937)

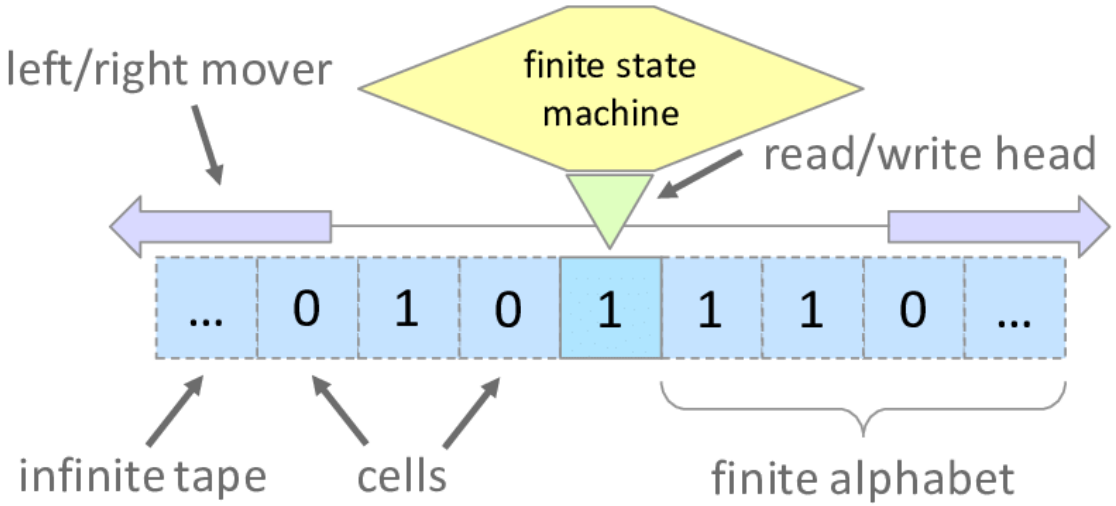
- Informal description *(more to come, later on)*

  An infinite **tape**, made up of individual **cells**
  Each cell contains a **symbol**, from a finite **alphabet**

  A **read/write head**, which can move in each direction (one cell at time)

  A **finite state machine**:

  - The machine starts in an *initial state*
  - Each *state transition* is governed by the input symbol and the current state
  - The *next state* is stored into a register
  - The *output* is written to the cell
  - Then the head moves (i.e. *Left*, *None*, *Right*)



[https://www.researchgate.net/publication/341817215_Quantum_Accelerated_Estimation_of_Algorithmic_Information/figures?lo=1]

Turing Machine (A. Turing, 1937)

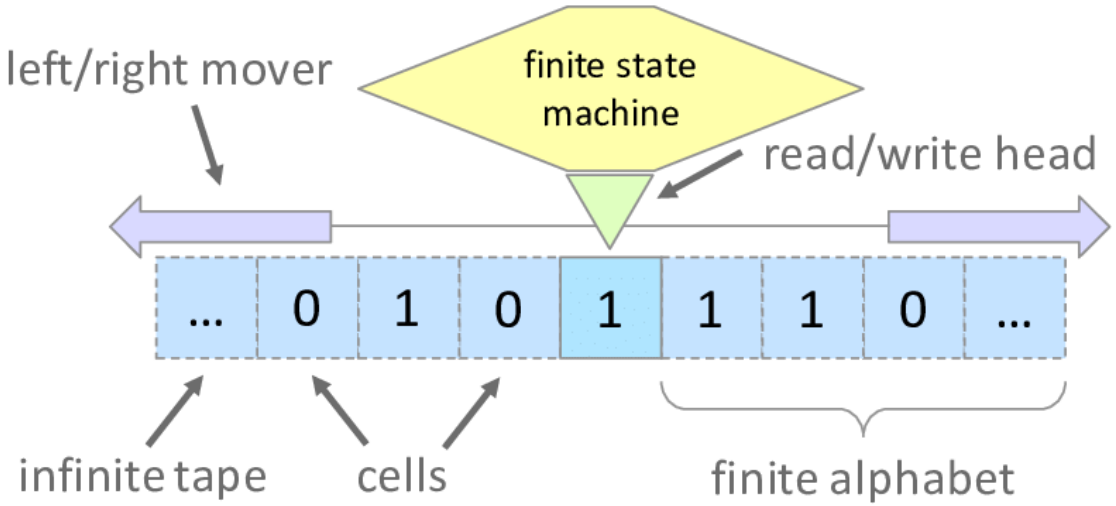
- What is the meaning of this?

  The Turing Machine is a mathematical model
  of a physical computing device

  Any given problem for which there is a Turing Machine
  that computes the solution is clearly computable by a physical machine

  Is the vice-versa also true?

  (*Whenever a problem is computable by a physical does it exist a Turing Machine for it?*)



[https://www.researchgate.net/publication/341817215_Quantum_Accelerated_Estimation_of_Algorithmic_Information/figures?lo=1]

Church-Turing Thesis

- A possible formulation* (from Wikipedia):

> *"Every 'function which would naturally be regarded as computable' can be computed by a Turing machine."*

The vagueness in the above sentence gives raise to different interpretations.
One of these (though not entirely equivalent) is (from Wikipedia)*:*

> *"Every 'function that could be physically computed' can be computed by a Turing machine."* **

Searle: "... At present, obviously, the metaphor is the digital computer."

** Caution: there is no such a thesis in the original writings of either author.*
Its formulation could be extrapolated from both, hence the attribution (made by others)

*** Quantum computation shatters complexity theory, but is (almost) innocuous to computability theory*

Can machines think? (the Turing Test)

- **The Imitation Game** *(textual interactions only)*

"A man (A), a woman (B),
and an interrogator (C) who may be of either sex.

The interrogator stays in a room apart from the other two.

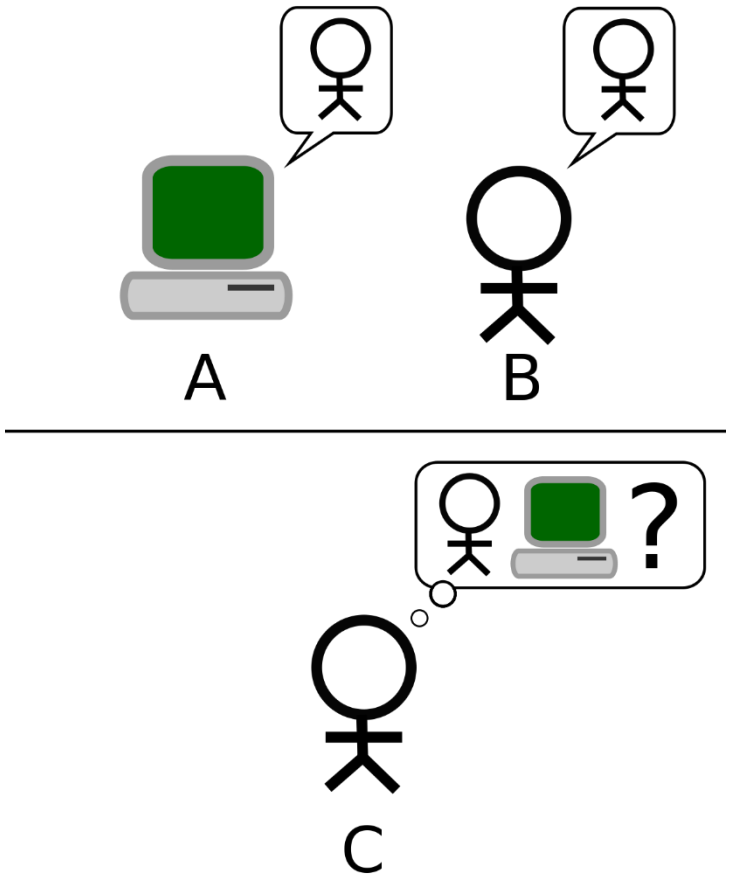
The object for the interrogator is to determine
which of the other two is the man and which is the woman.

The interrogator is allowed to put **questions** to A and B.
[…]

We now ask the question,
'What will happen when a machine takes the part of A in this game?'

[...]

These questions replace our original, 'Can machines think?' "



[Image from Wikipedia]

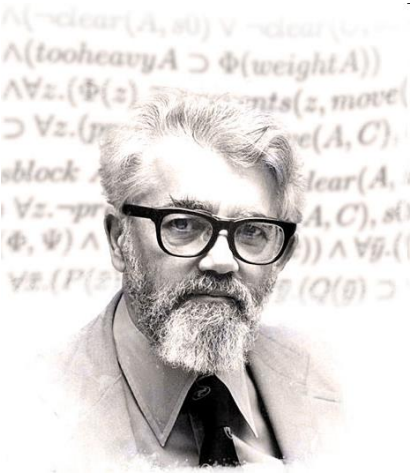
[Turing, A., Computing Machinery and Intelligence, 1950]

An aside question: Are we machines?

[Image from https://xkcd.com/329/]

Bits of AI History:
The beginning, a symbolic approach

"Artificial Intelligence" (first appearance of the term)



[Image from Wikipedia]

"We propose that a two-month, ten-man study
of **artificial intelligence** carried out during the summer of 1956 [...]

The study is to proceed on the basis of the conjecture
that 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. […]

It may be speculated that a large part of human thought
consists of manipulating **words** according to **rules** of **reasoning**
and **rules** of **conjecture**."

[John McCarthy et al., 1955 – *all emphases added*]

Automated Symbolic Logic

"The only way we know of expressing abstractions [...] is in language.
That is why we have decided to program a system which reasons verbally. "

[John McCarthy, PROGRAMS WITH COMMON SENSE, 1959]

- **Formal, Symbolic Logic *as a viable candidate***

  In those days, Formal Logic seemed an obvious choice

  - its formalism is derived from natural language
    (since G. Frege, 1879)
  - it has a clear semantics
    (since 'The Semantic Theory of Truth', A. Tarski, 1930)
  - it is *compositional*
    (the meaning of a complex expression is determined by its constituent and the rules used to combine them)
  - it can be turned into a *computing system*
    (see Herbrand's Theorem, 1930)

Automated Symbolic Logic

"The only way we know of expressing abstractions [...] is in language.
That is why we have decided to program a system which reasons verbally. "

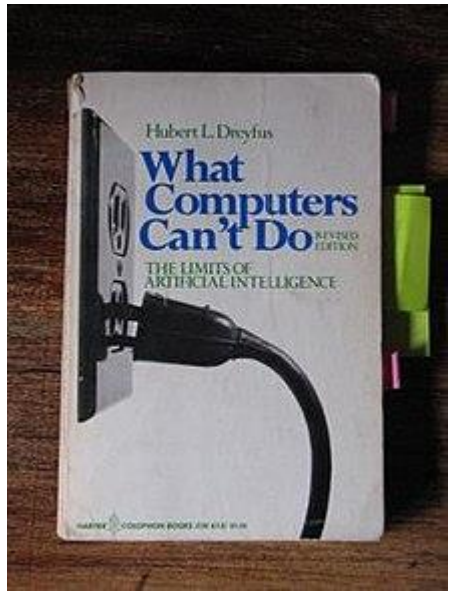
[John McCarthy, PROGRAMS WITH COMMON SENSE, 1959]

- ***Logicism* dominated the early period of AI**

  The logical approach in three theses:
  (from 'Logic and artificial intelligence', N. J. Nilsson, 1991 – *emphasis added*)

  1. Intelligent machines will have ***knowledge*** of their ***environments***.
  2. The most versatile intelligent machines will represent much of their knowledge about their environments ***declaratively***.
  3. For the most versatile machines, the language in which declarative knowledge is represented must be at least as expressive as ***first-order predicate calculus***.

Criticism of the symbolic approach



[Image from Wikipedia]

[H. Dreyfus, 1972]

Human intelligence and expertise depend primarily
on ***unconscious*** processes
rather than *conscious* symbolic manipulation,
and that these *unconscious* skills can never be fully captured in formal rules.

*(These ideas are now embedded in the so-called **sub-symbolic** approach)*

An Aside Question: Are we *out of control?*

'Out of Control', BBC2 Horizon documentary, 2012

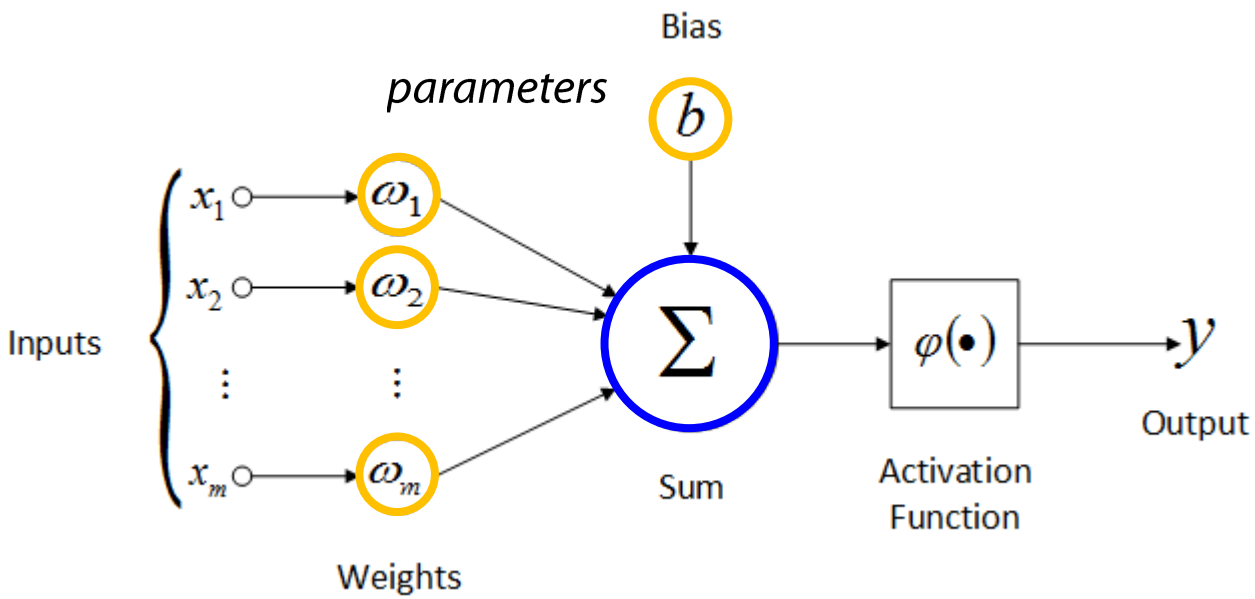
- *How big is the unconscious mind (in humans)?*

  In this BBC documentary, several senior neuro-scientists were asked to represent on a sheet of paper the extent of conscious activity (vs. unconscious) in the human brain

[Image from https://www.bbc.co.uk/programmes/p00pyhx2]

Bits of AI History :
The Dawn of Connectionism

Artificial Neural Networks



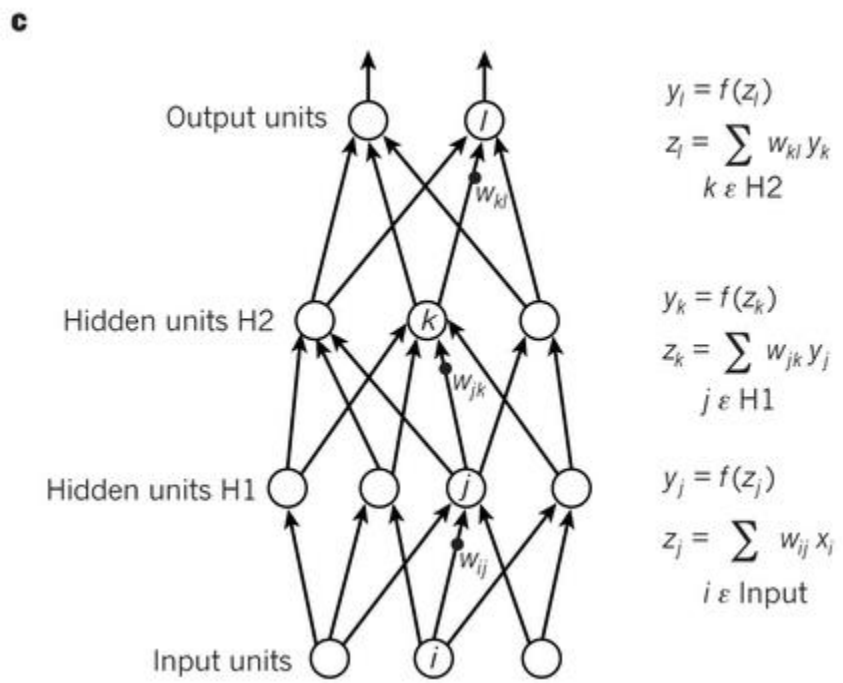
[Images from Wikipedia]

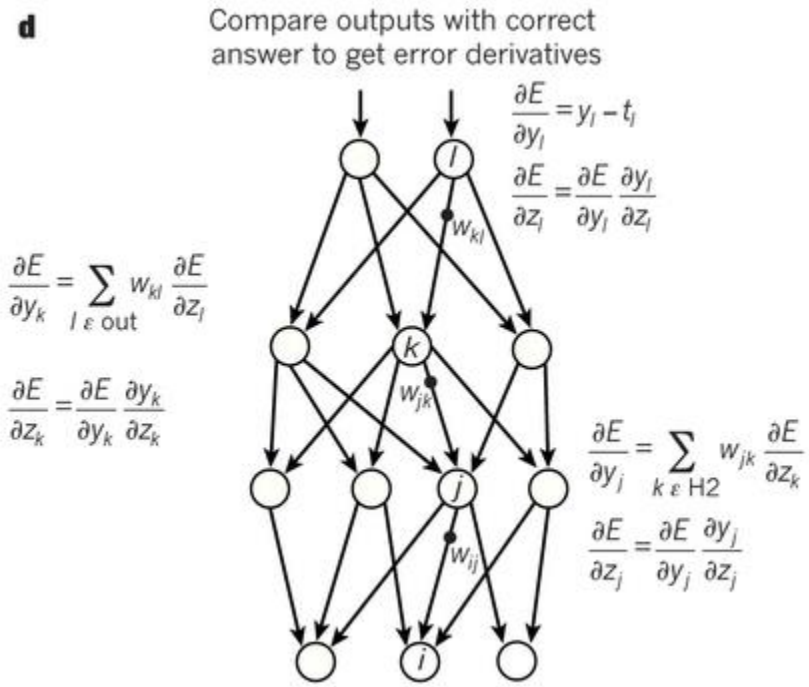
[Rumelhart, D.E., J.L. McClelland 1986]

- ***Basic assumption***

  Mental phenomena can be described
  by interconnected networks of simple and often uniform units

Artificial Neural Network





[figure from LeCun, Bengio, Hinton, *Deep Learning*, Nature 521, 2015]

Function approximation

This is what an artificial neural network does

Supervised learning

The parameters (i.e. *weights*) are "learnt" from a (large) set of *data items*: pairs of input and expected output

Incremental optimization

a.k.a. "*backward propagation*"

Weights are progressively corrected to minimize a *loss function*, namely the difference between actual and expected outputs

Artificial Intelligence:
"Emulation of human cognitive abilities" ...
In which sense?
And to what extent?

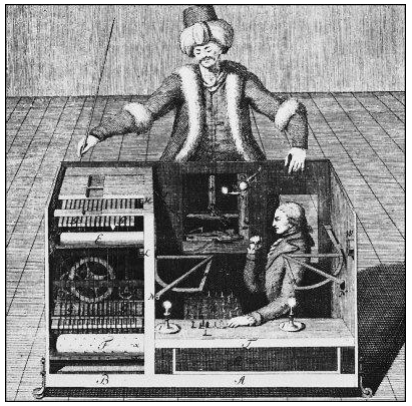
Computers play games

Can Machines Play Chess?

In 1945 A. Turing mentions playing chess as an example of intelligent human activity that some days machines could perform

In 1946 A. Turing defines the first *algorithm* for playing chess

In 1948 C. Shannon wrote a famous article on the possible strategies for playing chess *automatically*



(from Wikipedia)

- **Programming a Computer for Playing Chess** [Shannon, 1948]

  More than 10^{43} different legitimate chessboard configurations

  More than 10^{120} possible games

  Strategy A

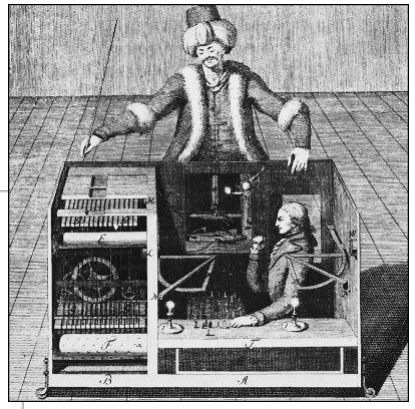
  Starting from the current position, the machine *looks forward* by exploring <u>all</u> possible positions in the game not farther away than k moves

  *The computer chooses its move by **backward induction** using a <u>value function</u> (MINIMAX method)*

  Strategy B

  "A good human player examines only ***a few selected variations*** and carries these out to a reasonable stopping point"

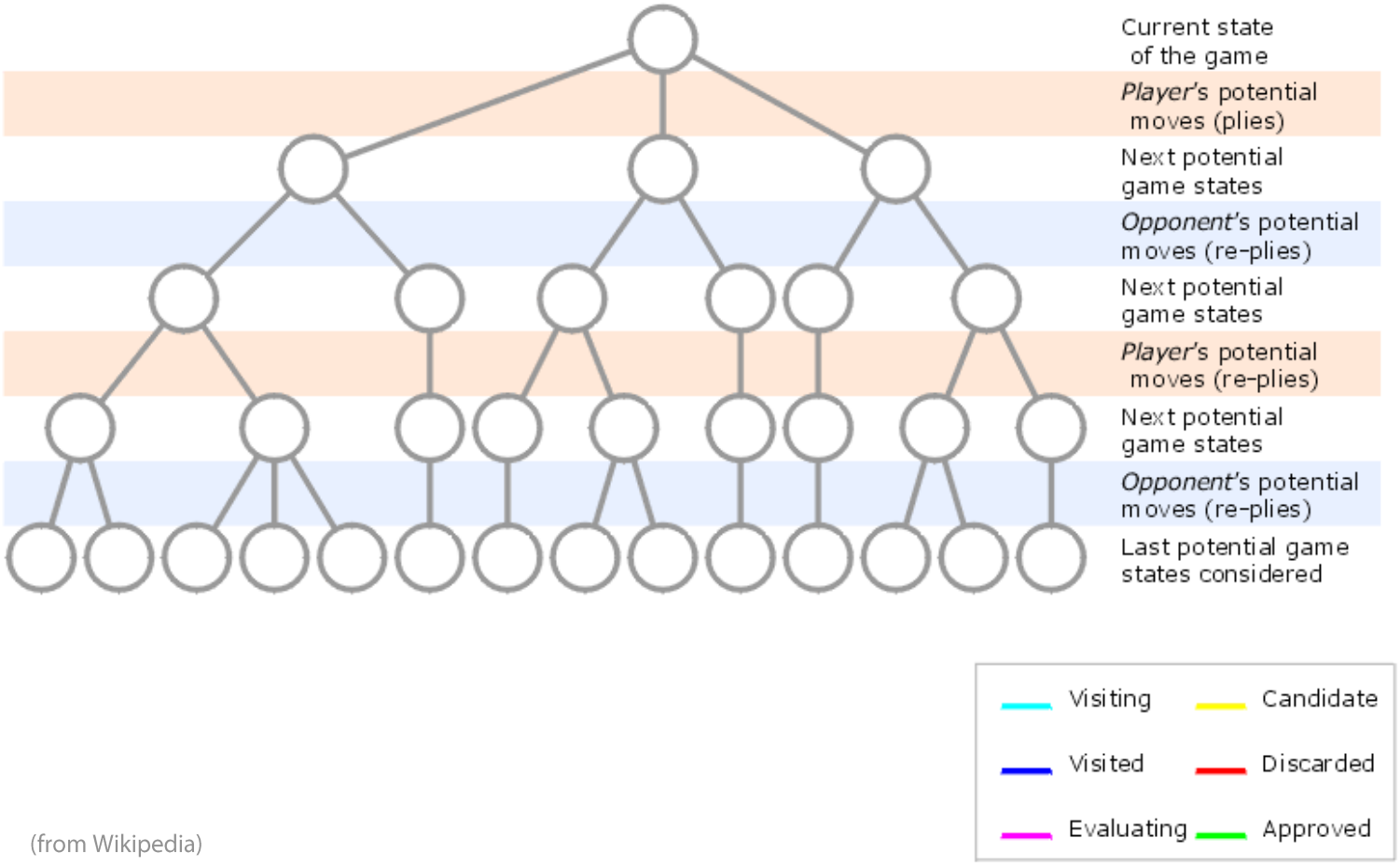
Can Machines Play Chess?



(from Wikipedia)

Strategy A

Minimax on a two-person game tree of 4 plies



(from Wikipedia)

Strategy A or Strategy B?

[Shannon, 1948]

Due to the high computational complexity of Strategy A,
he foresees a progressive development of Strategy B

(i.e. "Computer can improve by emulating humans")

How did it go, in reality?

- At the early stages of computer chess technology, Strategy B was preferred
- During the period 1959-1962 a first 'credible' player was developed [Kotok-McCarthy]
  (at the *beginner* level)
- In 1973 the developers of the soon-to-be world champion in computer chess players abandoned Strategy B in favor of Strategy A

From then on, for quite some time and through significant improvements, Strategy A dominated the scene

Deep Blue Beats the World Champion

In 1945 A. Turing mentions playing chess as an example of intelligent human activity that some days machines could perform

In 1946 A. Turing defines the first *algorithm* for playing chess

In 1948 C. Shannon wrote a famous article on the possible strategies for playing chess *automatically*

In 1997 the *Deep Blue* system, made by IBM, beats the world chess champion Gary Kasparov

(from Wikipedia)

- **Deep Blue, 1997** [Campbell, Hoane, Hsu, F., 2001]
  - 30 standard CPUs (120Mhz) + 480 special-purpose CPUs ('chess search engines', each evaluating >2.5M moves per second)
  - Three-layered hardware architecture, 30 GB of RAM
  - *S*oftware written in C
  - Dedicated team of software and hardware engineers, 10 year of development
  - Wide usage of a large database of recorded games played by grand masters

(A *supercomputer* for those times - *It was turned off at the end of the match*)

Games of strategy: AlphaGo (2016)

Image from: https://nikcheerla.github.io/deeplearningschool/2018/01/01/AlphaZero-Explained/

- **Mastering the game of Go with deep neural networks and tree search**
  [2016, D. Silver, et al. (22 authors), http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html]

There are more possible positions in Go than there are atoms in the universe

Sophisticated machine-learning techniques

Strategy selection via Monte Carlo Tree Search (MCTS)

Deep neural networks (trained on human matches) for both guidance and learning

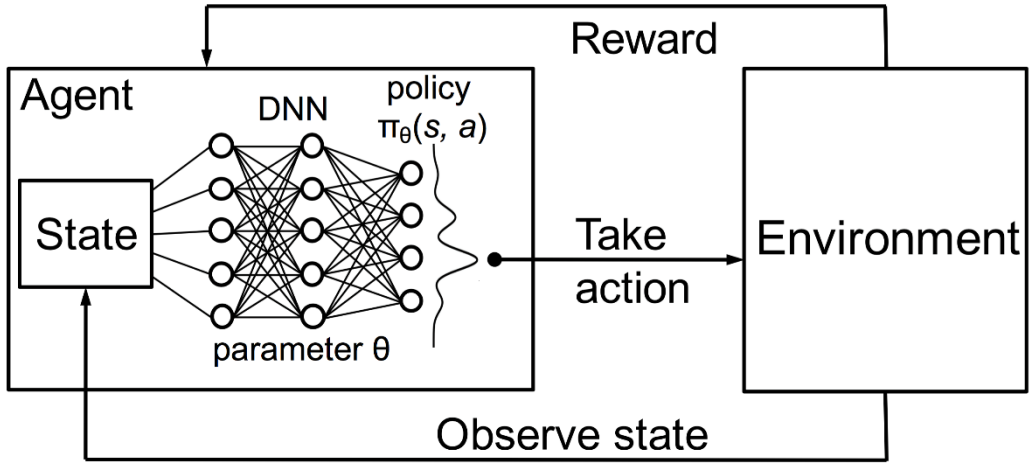
Adversarial self-training:
playing again itself and improving via reinforcement learning

Super-human player?

On March 2016, AlphaGo won 4-1 against the legendary Lee Sedol,
the top Go player in the world over the past decade

Deep Reinforcement Learning (DRL)

- ***A Deep Neural Network learns a policy***



The agent interacts with an environment (it could be a copy of itself)

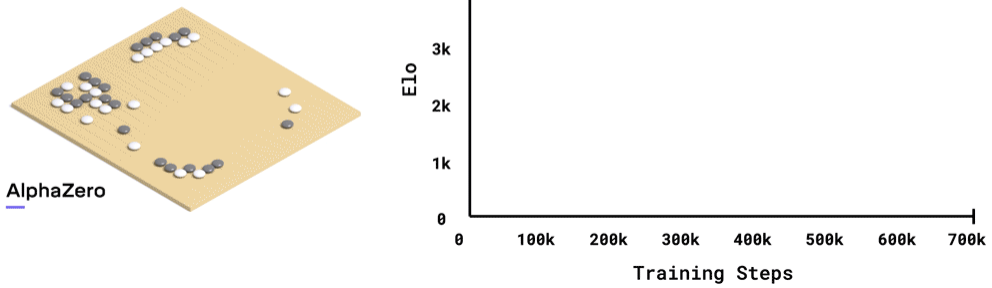
*It selects an **action** in each **state** and receives a **reward** (typically deferred) as a function of the new state obtained*

The DRL system optimizes its policy

Beyond Emulating Humans: AlphaZero (2018)

Image from: https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go

Its predecessor AlphaGo is heavily reliant on the experience of human players



- **AlphaZero learns by itself**
  [2018, D. Silver, et al. (13 authors), https://science.sciencemag.org/content/362/6419/1140.full]

  Basic Knowledge Only

  It just knows the basic rules of the games

  Learning via Self-Play

  It plays against a (frozen) copy of itself

  MCTS and DCNN in a closed loop

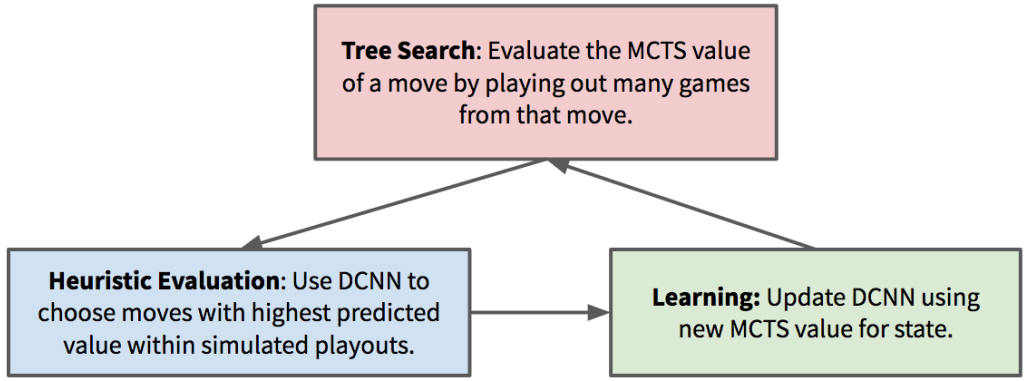
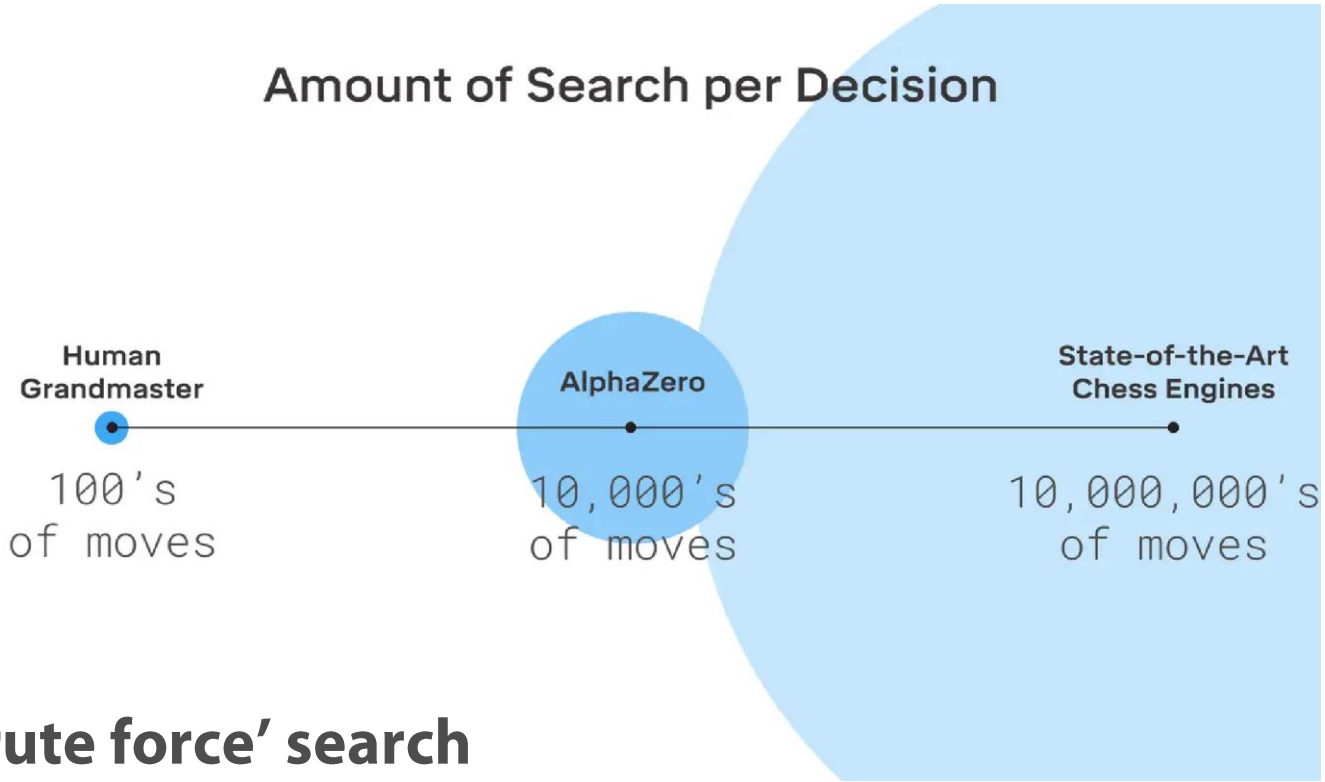


Image from: https://nikcheerla.github.io/deeplearningschool/2018/01/01/AlphaZero-Explained/

Beyond Emulating Humans: AlphaZero (2018)

Image from: https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go

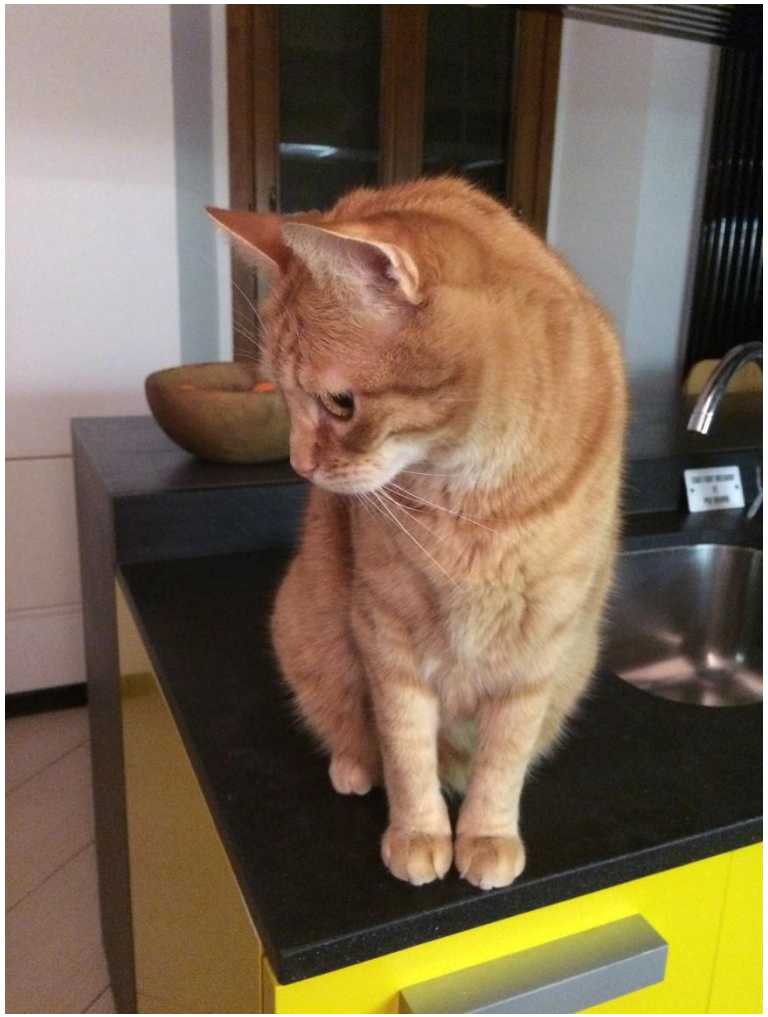


- **AlphaZero uses much less 'brute force' search**
  - When playing, the search process is driven by its neural network
    - *It acts like a memory of past experiences*
  - While training, it learns through a huge amount of self-playing
    - *But it is a faster learner than Alpha Go*

Computers can see cats

Artificial Perception

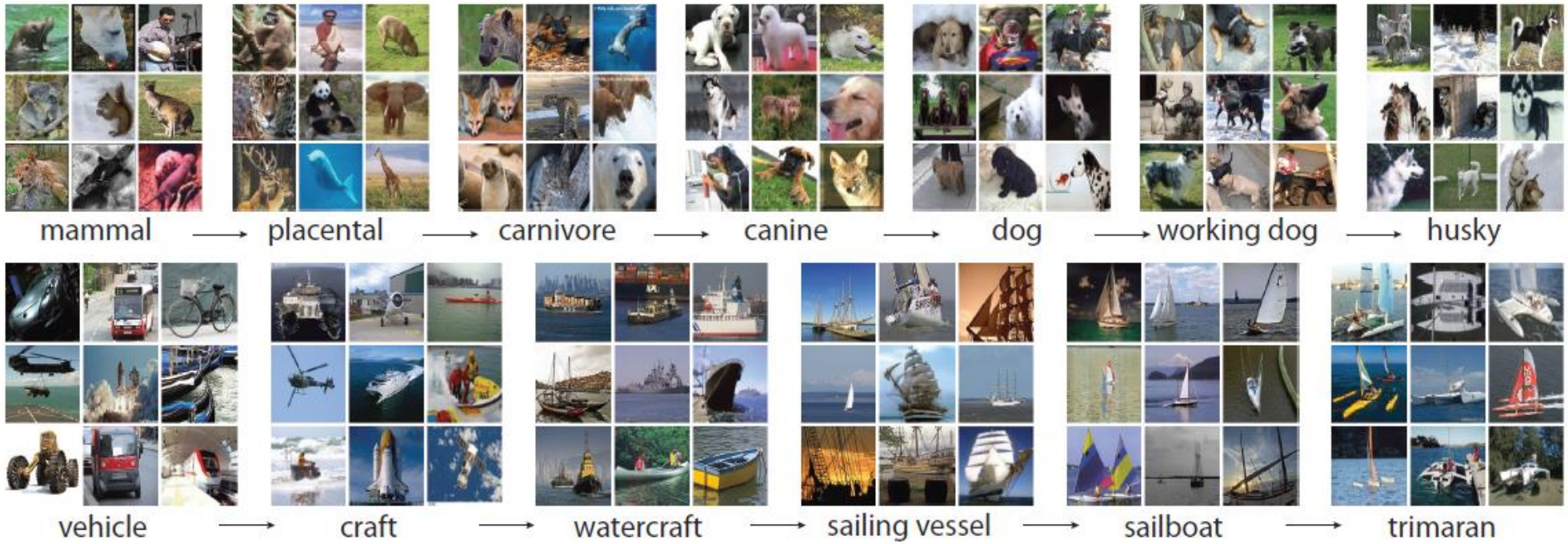
Is there a cat
in this picture?



[this is *my* cat, Rabarbaro]

ImageNet Challenge

- The ImageNet Large Scale Visual Recognition Challenge



1,461,406 full resolution images

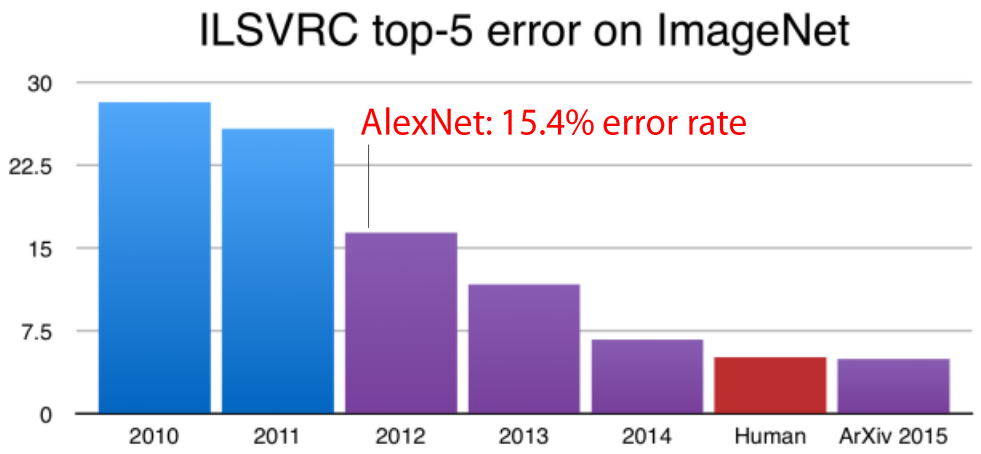
Complex and multiple textual annotation,
hierarchy of 1000 object classes along several dimensions

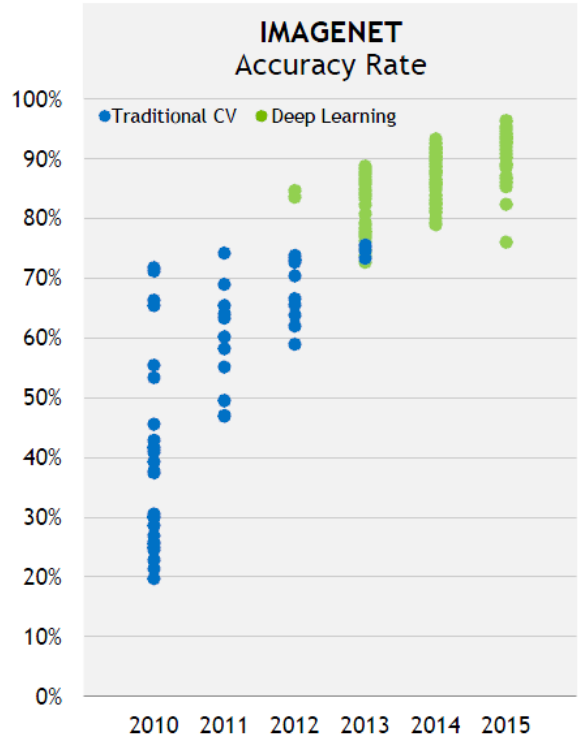
The image classification challenge is run annually since 2010

[figures from www.nvidia.com]

ImageNet Challenge

- The ImageNet Large Scale Visual Recognition Challenge





1,461,406 full resolution images

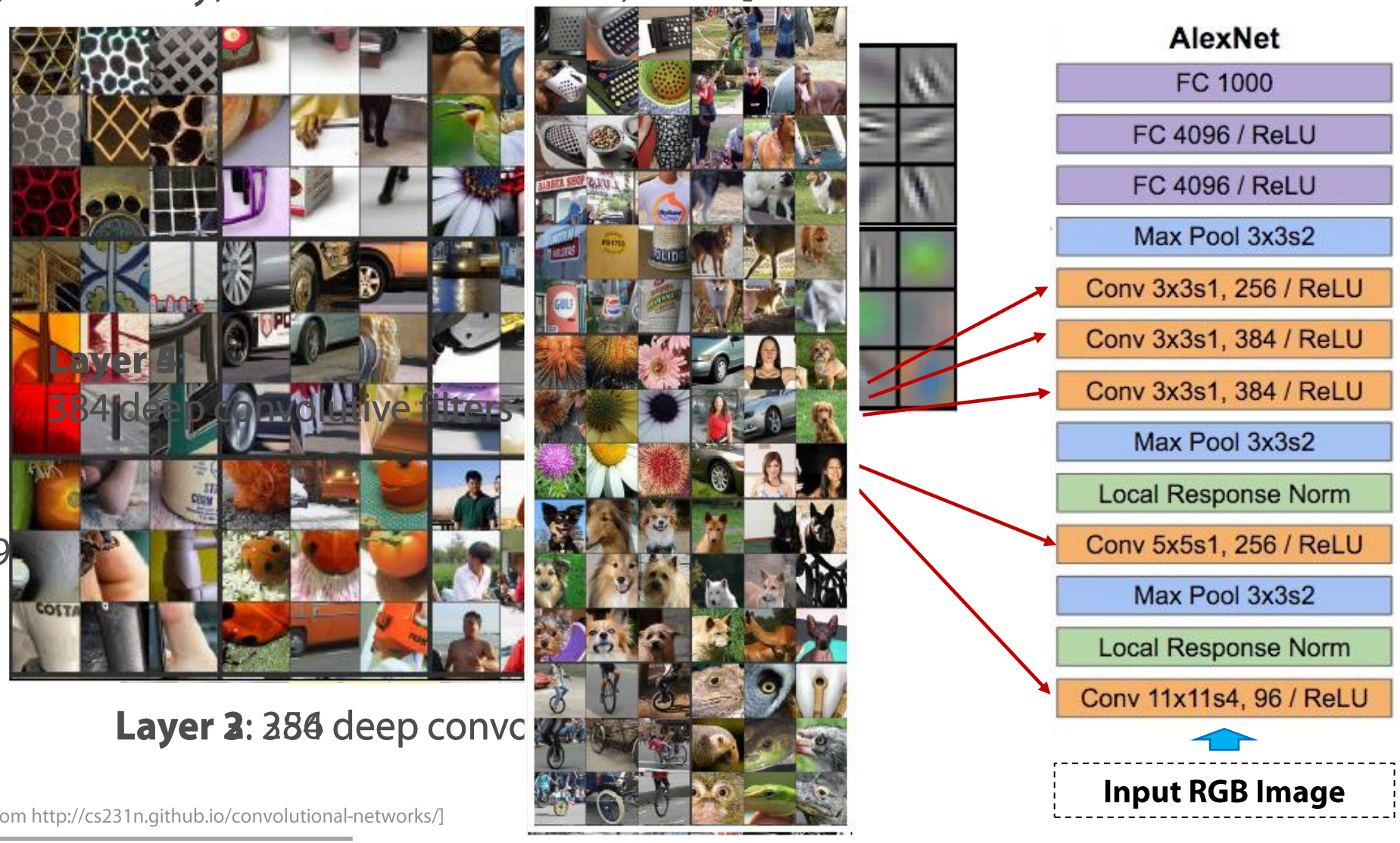
Complex and multiple textual annotation,
hierarchy of 1000 object classes along several dimensions

The image classification challenge was run annually from 2010 to 2017

[figures from www.nvidia.com]

Deep Convolutional Neural Networks (DCNN)

- ***AlexNet*** [Krizhevsky, Sutskever & Hinton, 2012]



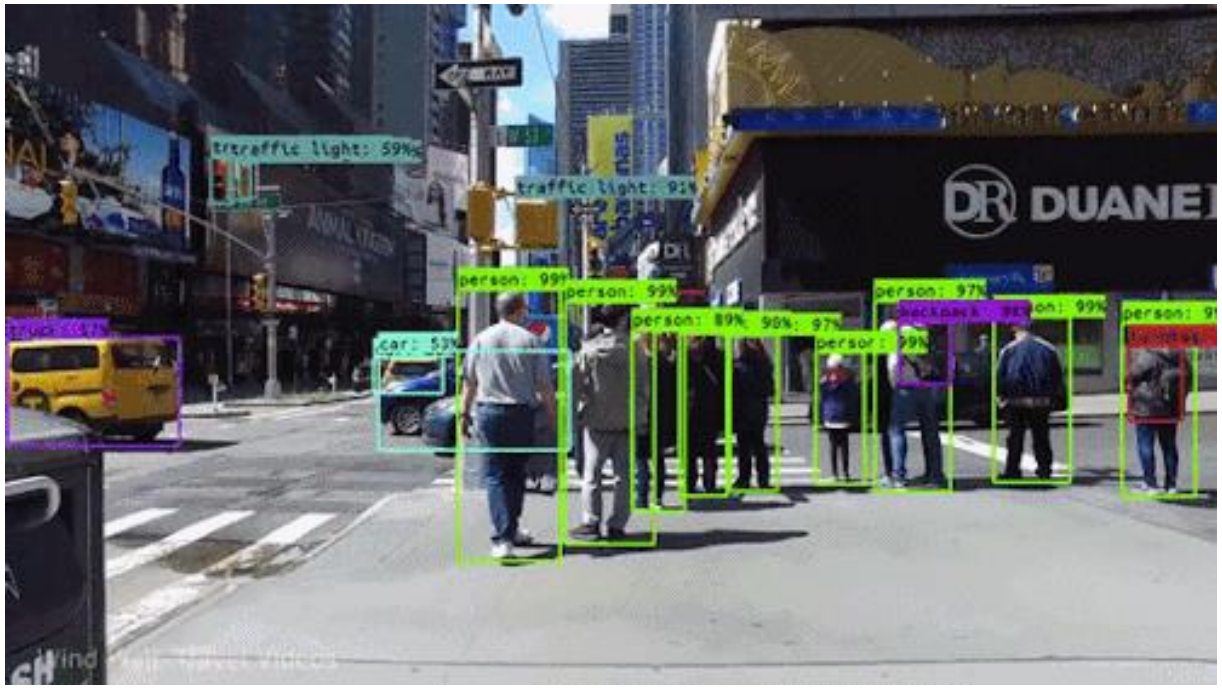
[images from http://cs231n.github.io/convolutional-networks/]

Object (and People) Real-Time Detection

- *Deep Convolutional Networks have evolved since then ...*

Now these system can identify objects
and people in real-time from videos

NOTE:
According to the recent
EU Artificial Intelligence Act,
***remote biometric identification** (RBI)*
in public places will require
a special authorization
(more on this subject, later on)



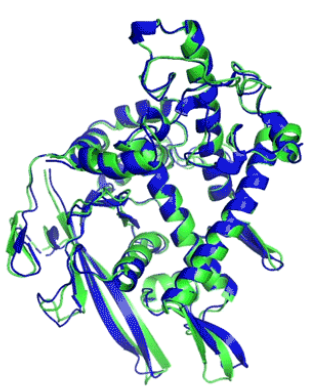
[Image from: https://sgu.ac.id/id/computer-vision-artificial-intelligence-why-is-it-important/]

Computers foresee how proteins fold

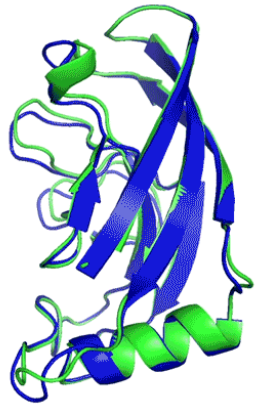
AlphaFold (2020)

Images from https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

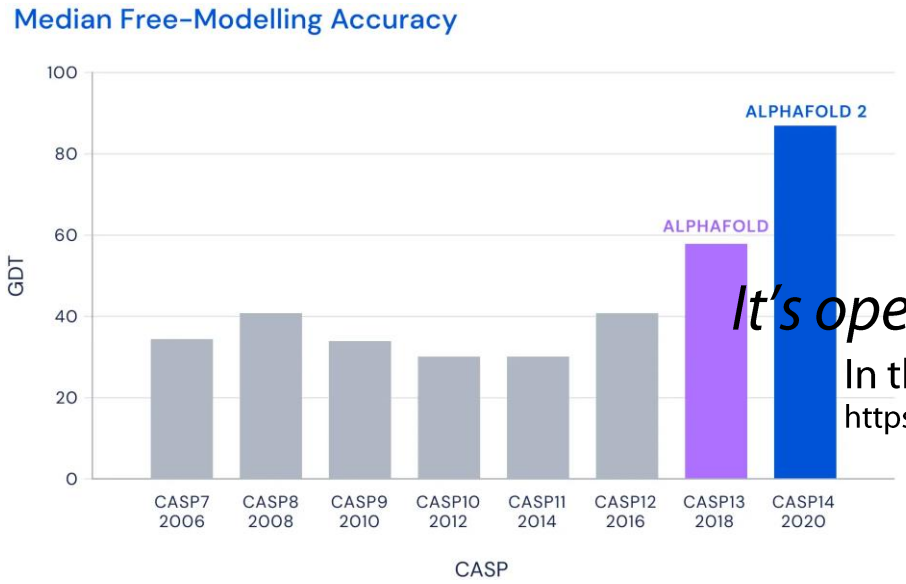
- **Predicting how proteins fold**
  - Several Neural Networks
    - *Trained separately, working together*
  - A big step ahead
    - *Substantial improvement over the best actual systems yet*

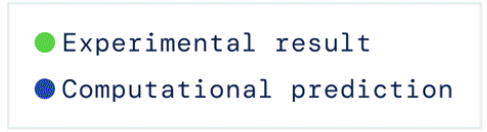


T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



T1049 / 6y4f
93.3 GDT
(adhesin tip)





It's open source

In the version used for the CASP13 challenge
https://github.com/deepmind/deepmind-research/tree/master/alphafold_casp13

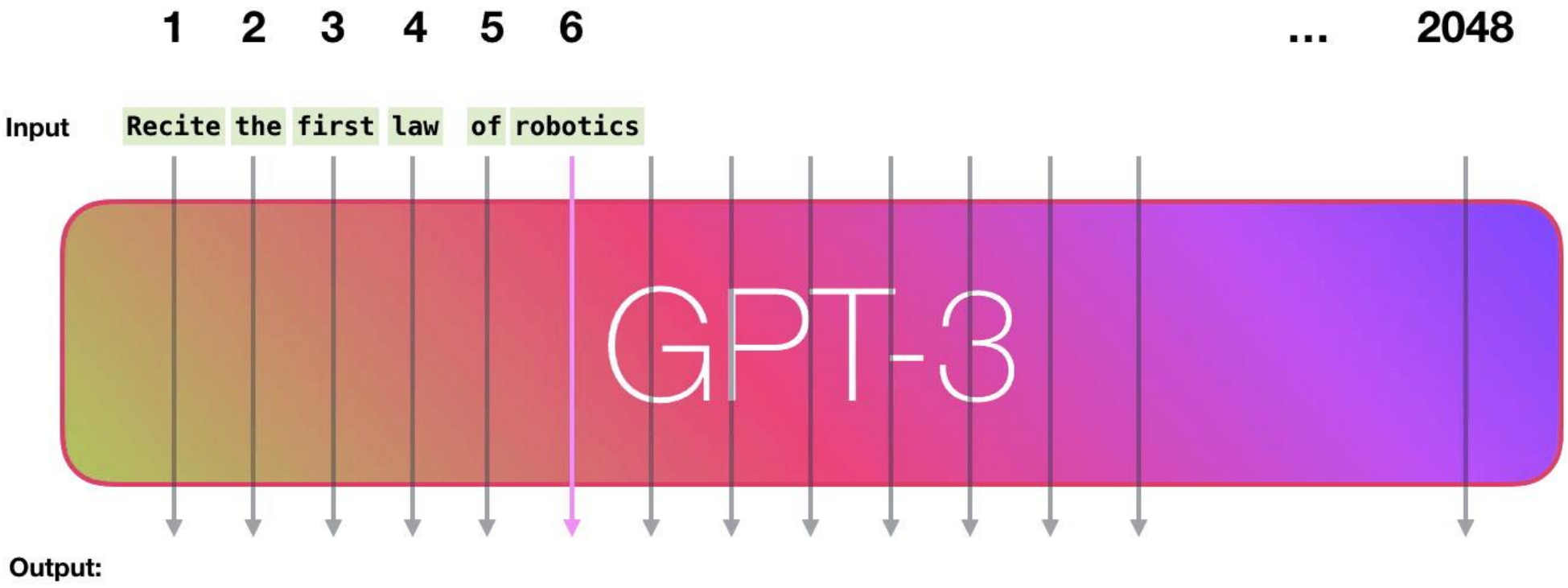
Computers write text

GPT-3 (2020)

Image from https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

GPT-3 (2020)

Image from http://jalammar.github.io/how-gpt3-works-visualizations-animations/



Computers can be misleading (and manipulative)

Personality and The Machine

Two well-known articles by Kosinski et al. [2012 and 2014]

Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

[a]Free School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and [b]Microsoft Research, Cambridge CB1 2FB, United Kingdom

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved February 12, 2013 (received for review October 29, 2012)

Computer-based personality judgments are more accurate than those made by humans

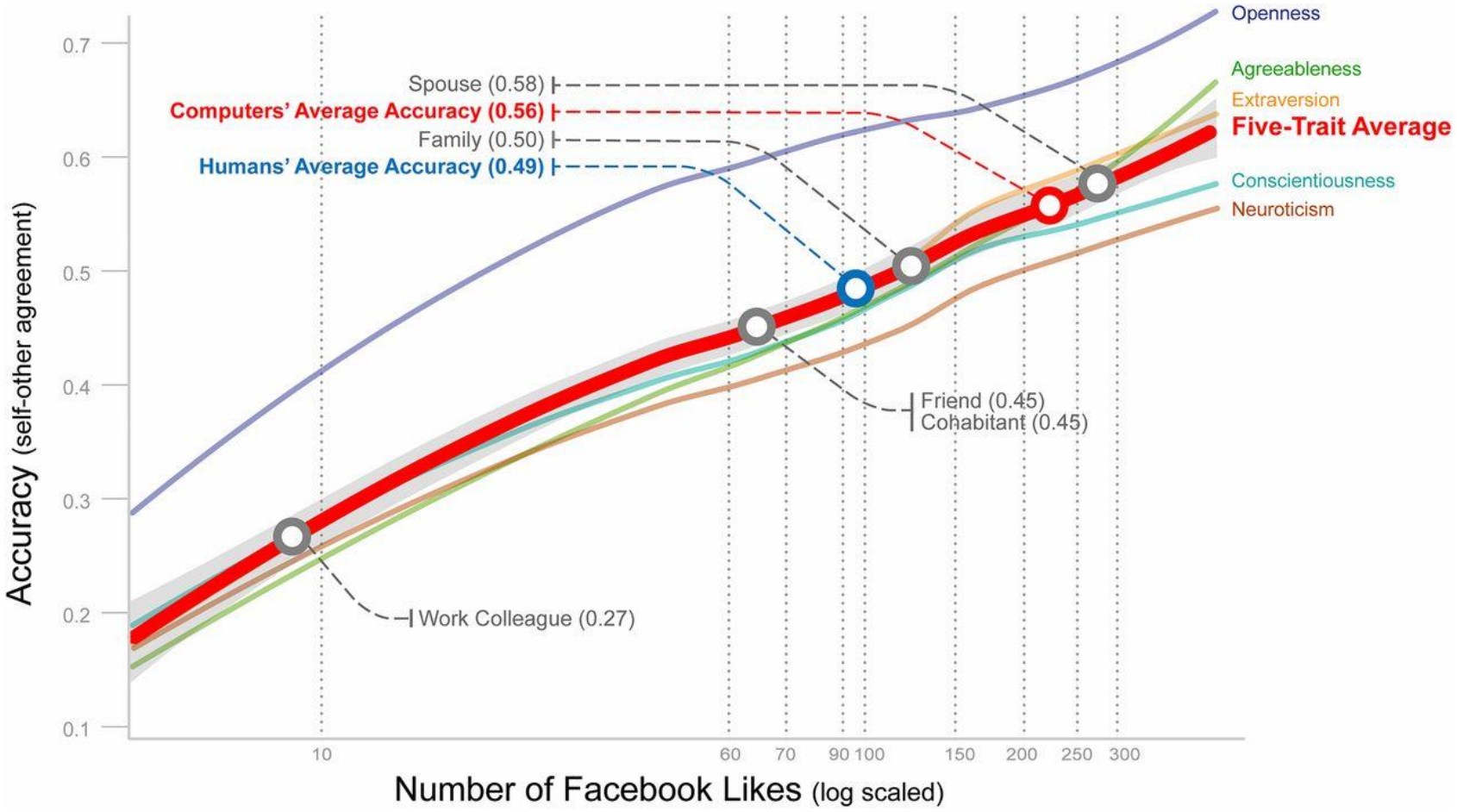
Wu Youyou[a,1,2], Michal Kosinski[b,1], and David Stillwell[a]

[a]Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom; and [b]Department of Computer Science, Stanford University, Stanford, CA 94305

Edited by David Funder, University of California, Riverside, CA, and accepted by the Editorial Board December 2, 2014 (received for review September 28, 2014)

We All Leave Traces Behind Us

- **The "Big Five" personality traits are predictable from Facebook likes**



Wu Youyou et al. PNAS 2015;112:4:1036-1040

The Cambridge Analytica Scandal (2015)

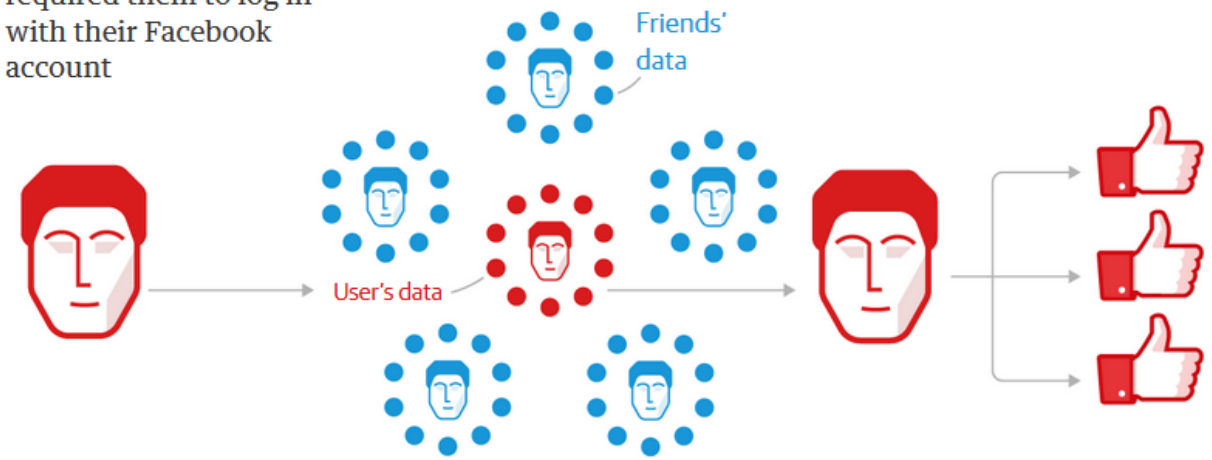
Cambridge Analytica: how 50m Facebook records were hijacked

1
Approx. 320,000 US voters ('seeders') were **paid $2-5 to take a detailed personality/ political test** that required them to log in with their Facebook account

2
The app also **collected data such as likes and personal information** from the test-taker's Facebook account ...

3
The **personality quiz results** were paired with their Facebook data - such as **likes** - to seek out psychological patterns

4
Algorithms combined the data with other sources such as voter records to **create a superior set of records (initially 2m people in 11 key states*)**, with hundreds of data points per person



... as well their **friends'** data, amounting to over 50m people's raw Facebook data

These individuals could then be targeted with **highly personalised advertising** based on their personality data

Guardian graphic. *Arkansas, Colorado, Florida, Iowa, Louisiana, Nevada, New Hampshire, North Carolina, Oregon, South Carolina, West Virginia

[Graphics from https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm]

Computers Can Learn From You

(and you may not be aware of that)

The Social Dilemma, 2020 - NETFLIX

Computers must obey the law

The Artificial Intelligence Act

14 June 2023 – The European Parliament adopted its negotiating position on the AI Act, with 499 votes in favour, 28 against, and 93 abstentions

What is the EU AI Act?

The AI Act is a proposed European law on artificial intelligence (AI) – the first law on AI by a major regulator anywhere. The law assigns applications of AI to three risk categories. First, applications and systems that create an **unacceptable risk**, such as government-run social scoring of the type used in China, are banned. Second, **high-risk applications**, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements. Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated.

https://artificialintelligenceact.eu/

A risk-based approach to regulation

Unacceptable risk
e.g. social scoring
— **Prohibited**

High risk
e.g. recruitment, medical devices
— **Permitted** subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

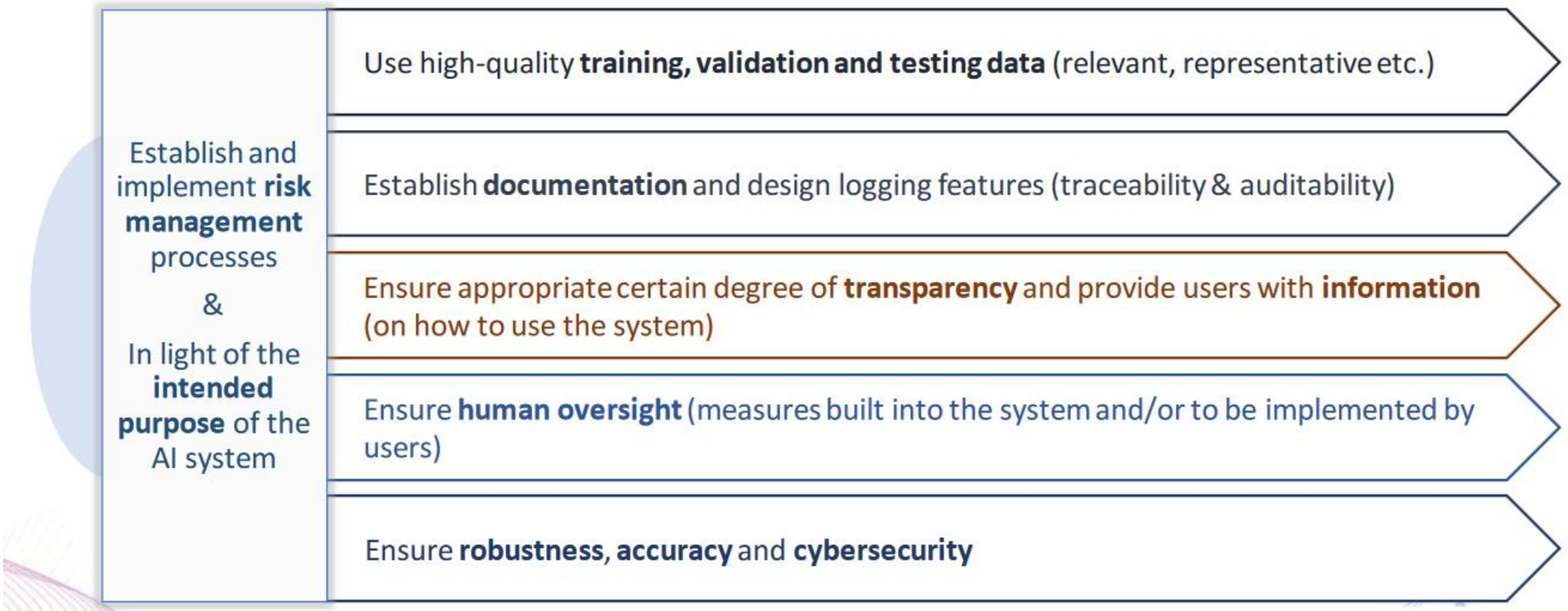
AI with specific transparency obligations
'Impersonation' (bots)
— **Permitted** but subject to information/transparency Obligations

Minimal or no risk
— **Permitted** with no restrictions

European Commission

https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

Requirements for high-risk AI (Title III, chapter 2)



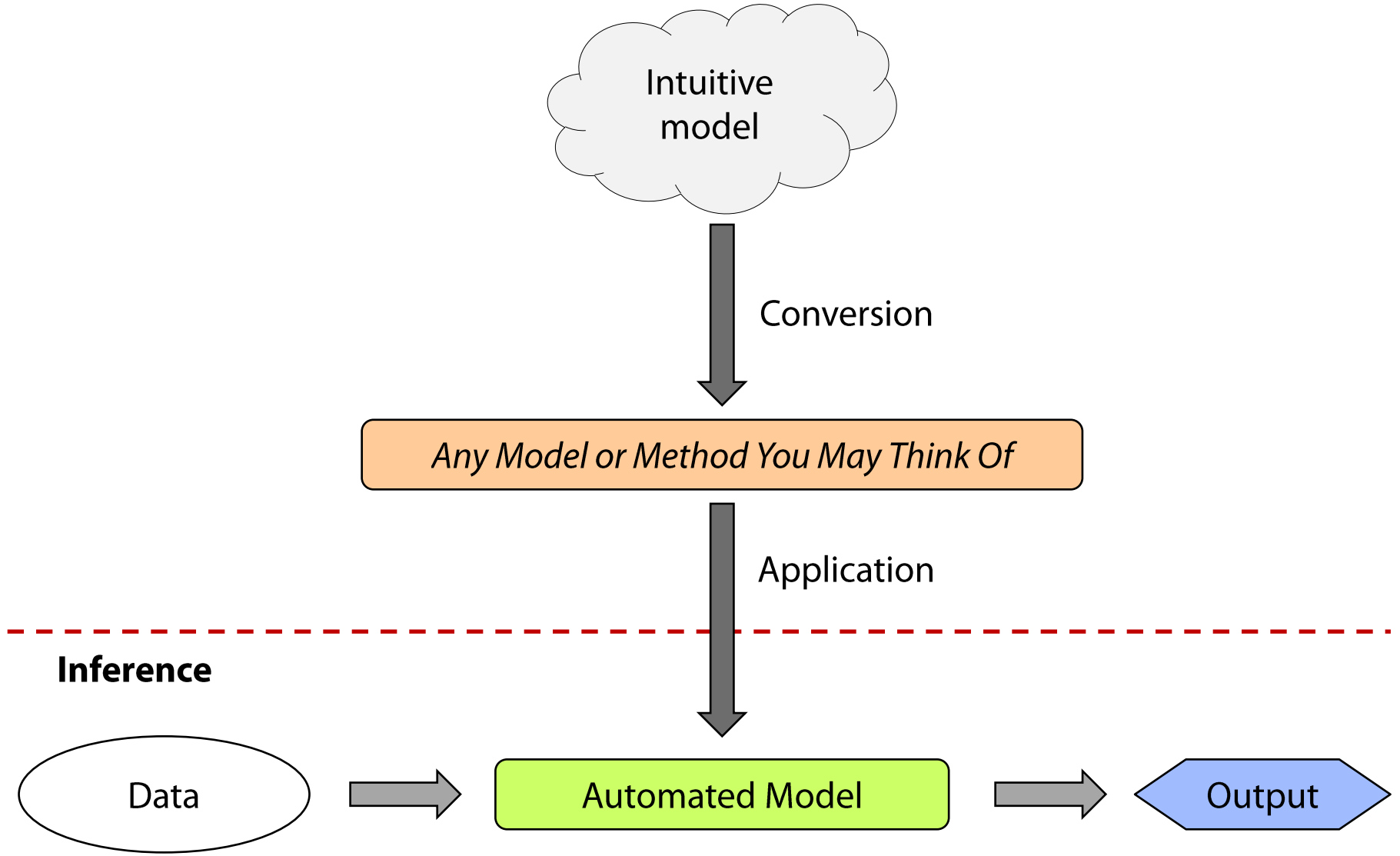
https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

- pause -

(This page intentionally left blank)

Artificial Intelligence:
a course about foundations

Inference as a Fundamental Trait



What Is Inference?

- **From a theoretical standpoint**

  What is the connection between input and output, in theory?
  (Can it be 'so precisely described' – see John McCarthy?)

  What are the justifications, strength and limitations of such connection?

  Can such connection be computed, at all?

- **From the point of view of computer engineering**

  To what extent, if any, can inference be translated into a computer algorithm?

  Are there practical limitations (computational complexity)?

  If so, what kind of compromises need we adopt to make it computationally feasible?

Artificial Intelligence: *Part 1*

- *Reasoning with symbols*

  Propositional logic, first-order logic, logic programming (hints)

  Representation: language and semantics *(logic formulae and their meaning)*

  Inference: entailment, a semantic relation among formulae

  Automation: can machines compute entailment?

  Plausible reasoning: beyond the scope of logical entailment

Artificial Intelligence: *Part 2*

- *Reasoning with numbers*

  Machine learning *(the probabilistic way)*

  Representation: probability, random variables, graphical models
  Inference: answers to queries from joint probability distributions
  Causality: can graphical models be interpreted as *causal* models, as well?

  Supervised learning: learning from *completely observed* and well-formed data items
  Unsupervised learning: when some parts of data items are either *missing* or *hidden*
  Reinforcement learning: learning from experience (even *online*)