Artificial Intelligence

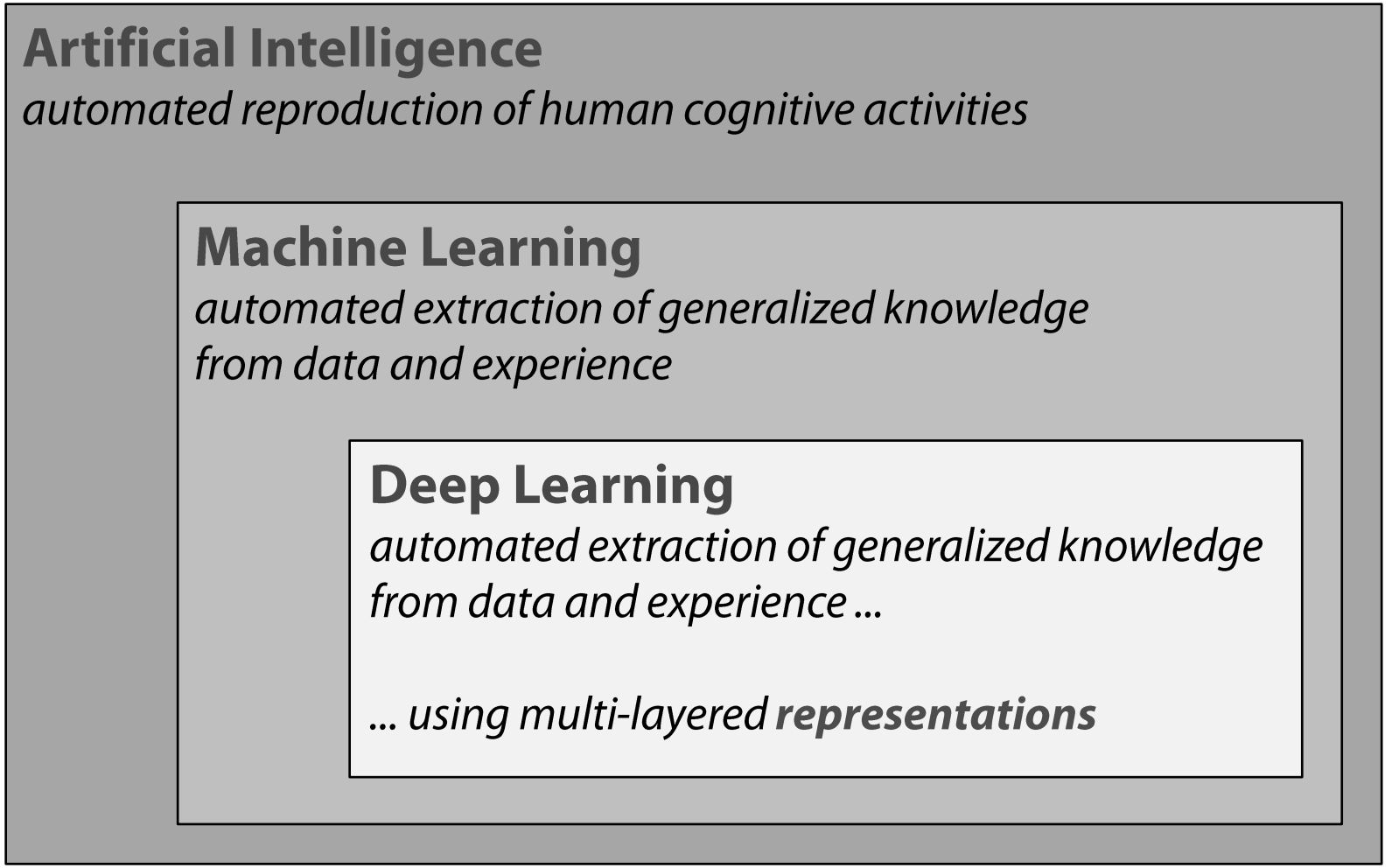
A course about foundations

Introduction

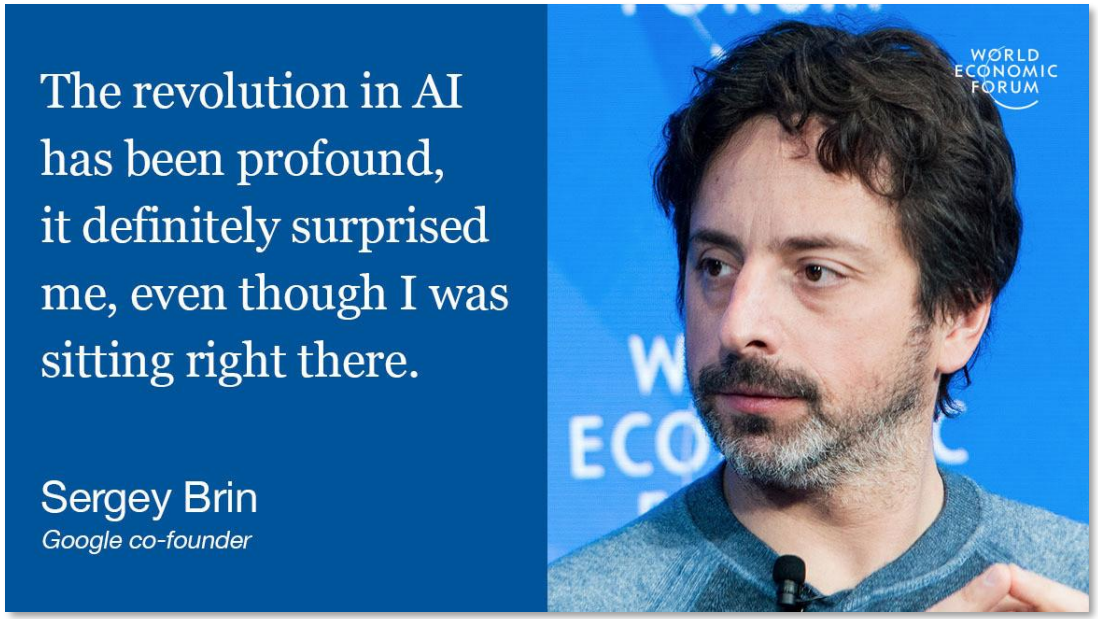
Marco Piastra

Prologue:
what this course is NOT about?

Deep Learning in the Artificial Intelligence Cosmos



AI strikes back?



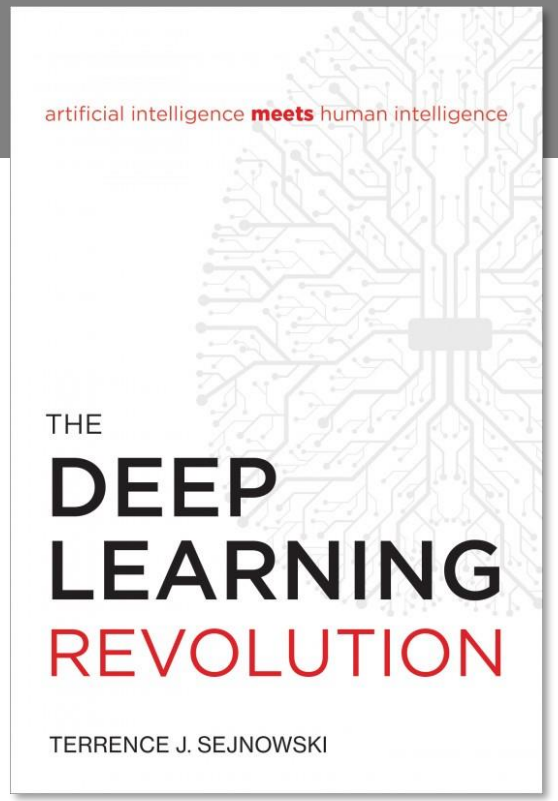
- Sergey Brin [Google Co-Founder, January 2017]

  "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/]

The Deep Learning 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 [...]"

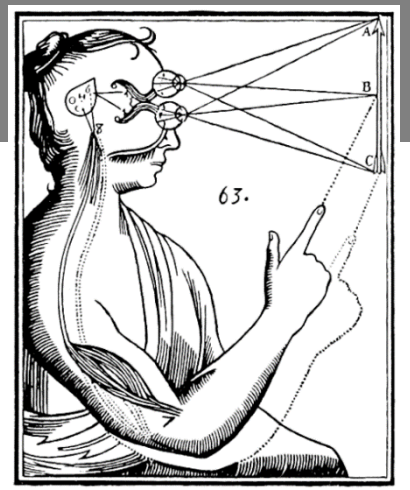
[Quote and image from https://mitpress.mit.edu/books/deep-learning-revolution]

OK, then what this course IS about?

Inference (the very idea)

Artificial Mind?

Mind vs. Brain



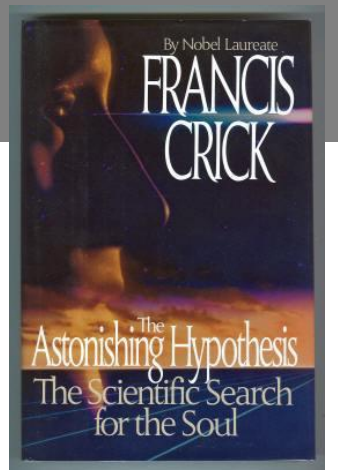
(from Wikipedia)

"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]



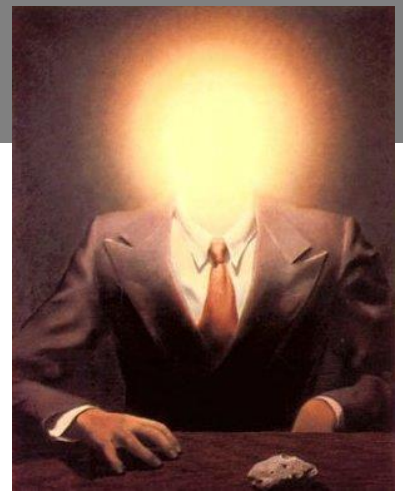
[Image from Wikipedia]

"You, your joys and your sorrows, your memories and your ambitions,
your sense of personal identity and free will,

are in fact no more than the behavior of a vast assembly
of nerve cells and their associated molecules."

[Crick, F., *The Astonishing Hypothesis*, 1994]

Artificial Brain:
Can machines think?



[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)

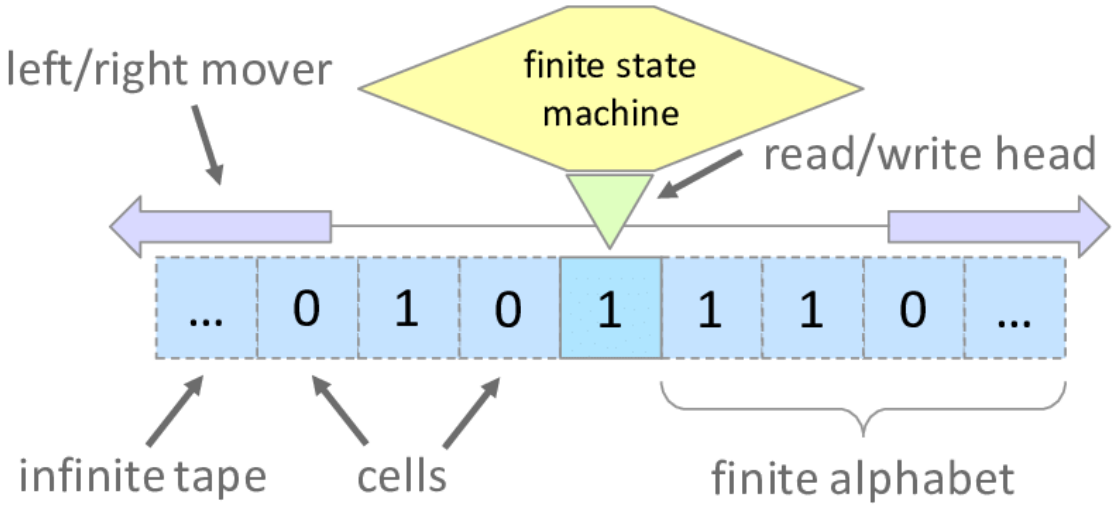
- Informal description *(more to come, later on)*

  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)

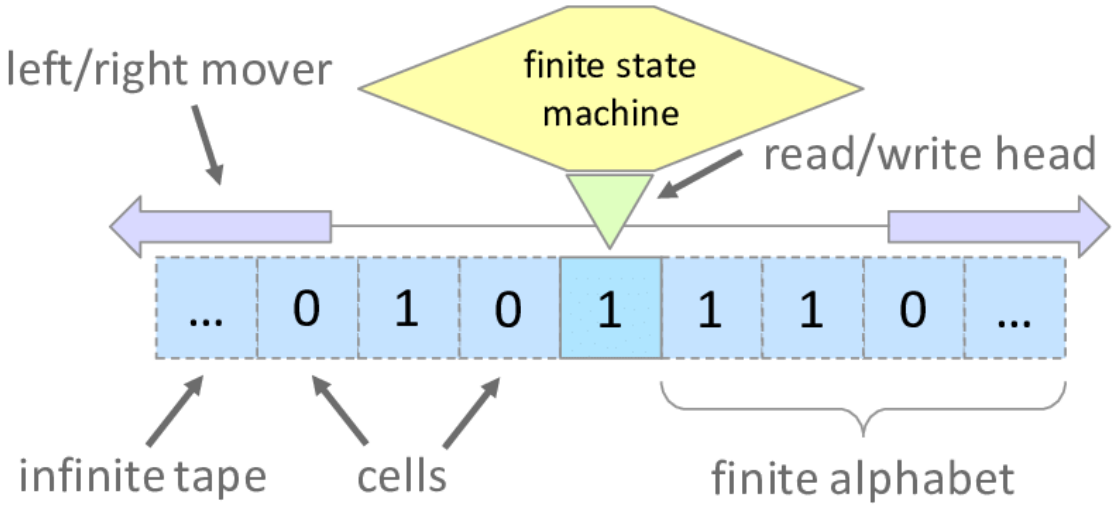
- What is the meaning of this?

  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

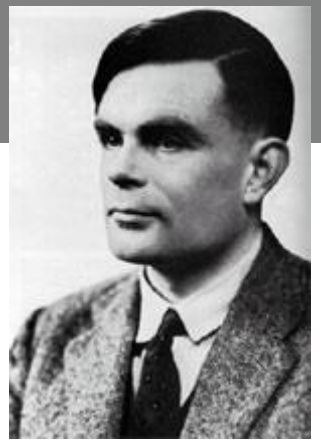
  Is the vice-versa also true?

  (*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)

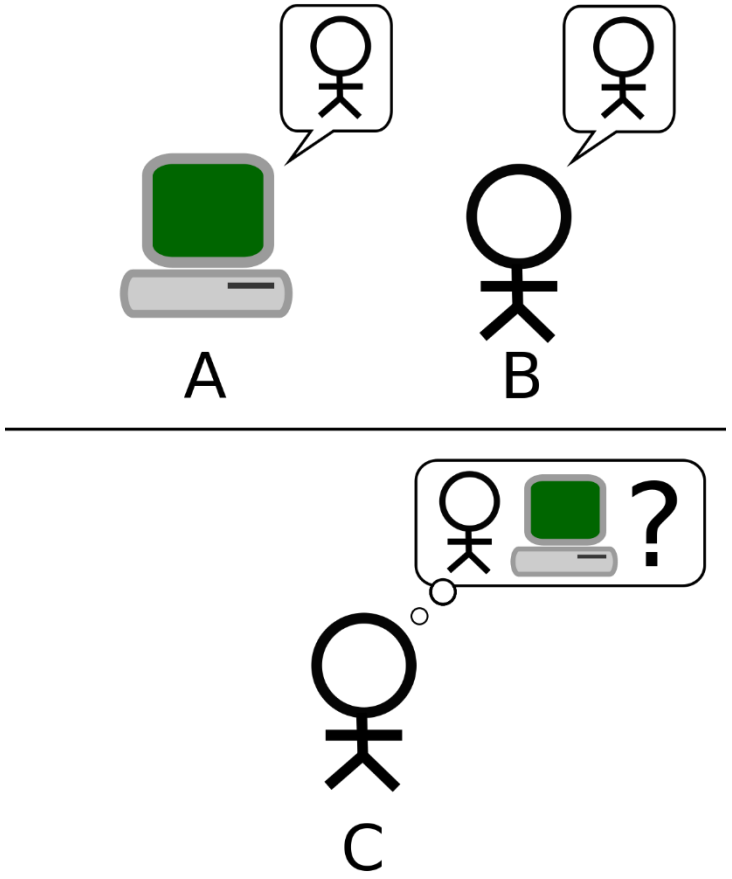
"[The Imitation Game]

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. […]



[Image from Wikipedia]

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

Will the interrogator decide wrongly as often
when the game is played like this as he does
when the game is played between a man and a woman?

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

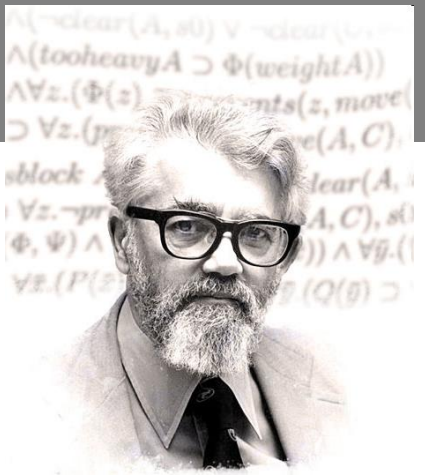
[Turing, A., Computing Machinery and Intelligence, 1950]

An aside question: Are we machines?

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

Artificial Intelligence:
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, emphasis added]

The Physical Symbol System Hypothesis (PSSH)

[Newell, A., Simon, H., *Computer Science as Empirical Inquiry Symbols and Search*, 1976]

A **Physical Symbol System (PSS):**

- a set of entities, called **symbols**
- as components of **expressions** (or symbol structure)
  related in some physical way (such as one token being next to another)
- a collection of **processes** that operate on *expressions* to produce other *expressions*:
  creation, modification, reproduction and destruction

The **Physical Symbol System Hypothesis (PSSH):**

- Any (sufficiently sophisticated) **PSS** is capable of intelligent action

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. "

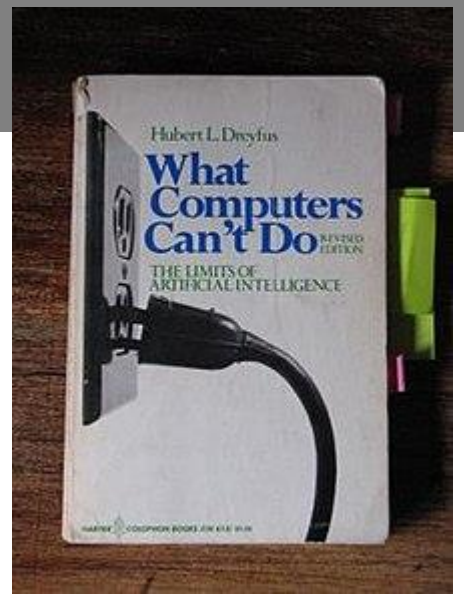
[John McCarthy, PROGRAMS WITH COMMON SENSE, 1959]

- ***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:
Are we *out of control?*

'Out of Control', BBC2 Horizon documentary, 2012

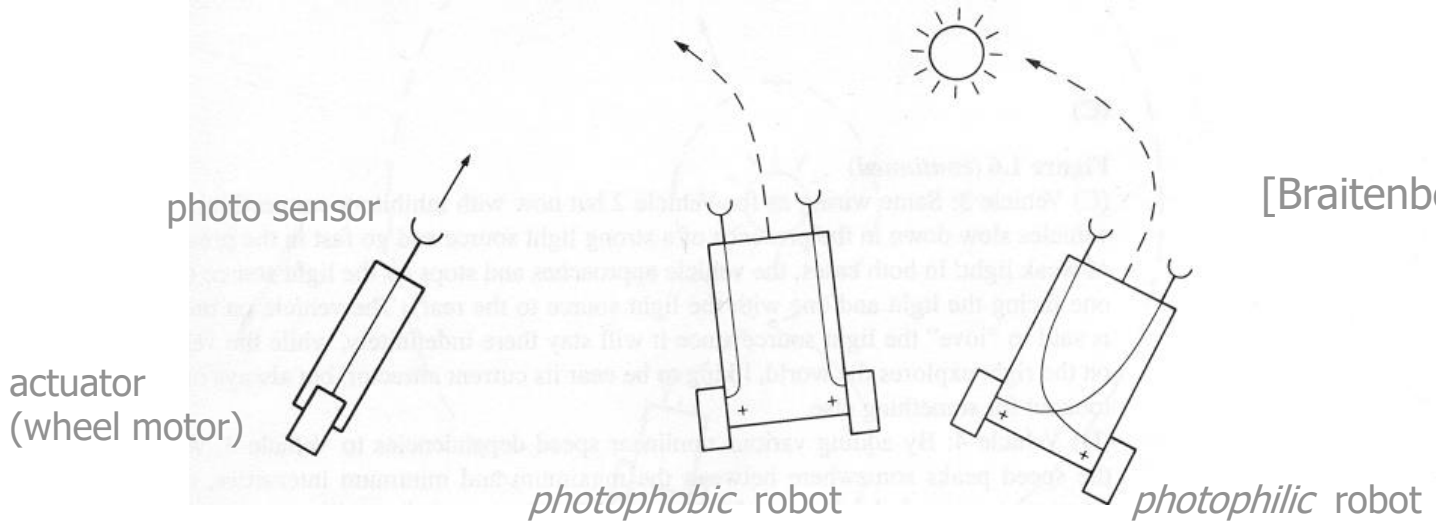
- *How big is the unconscious mind (in humans)?*

  In this BBC documentary, a few senior neuro-scientists were asked to represent on a sheet of paper how much did they think was conscious vs. unconscious activity in the human brain

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

Artificial Intelligence:
the dawn of connectionism

Connections and Behavior



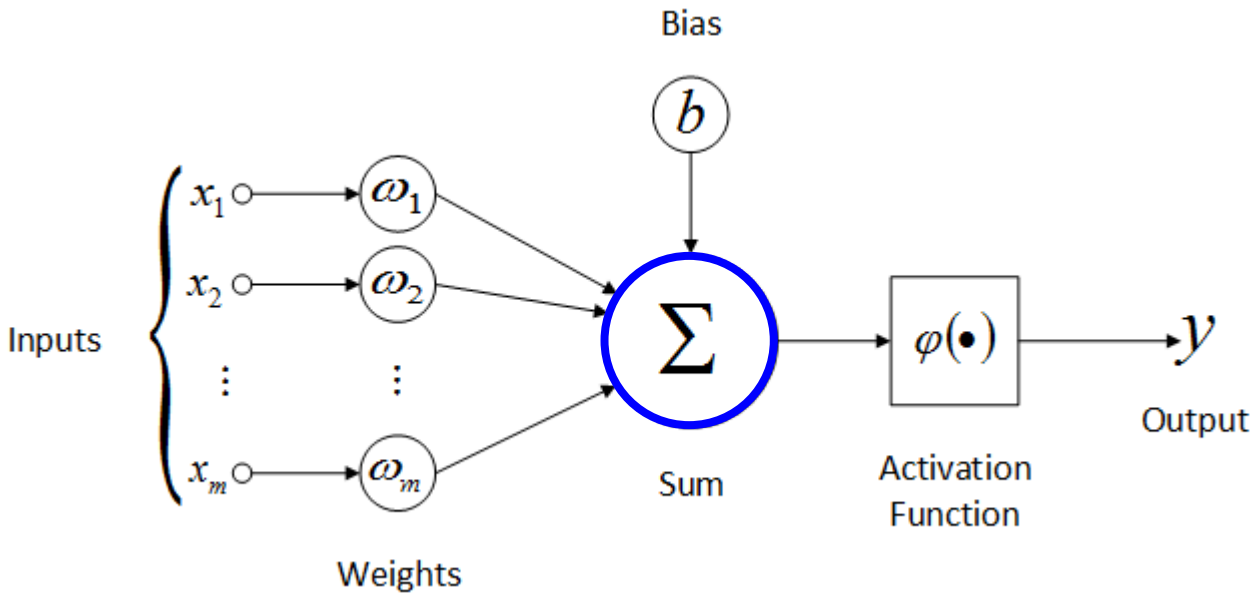
[Braitenberg, V., 1986]

- Direct connection

  These robots by V. Braitenberg have just a *reactive* behavior,
  i.e. no 'thought in between': *since sensors are directly connected to actuators*

  The resulting behavior is remarkable anyway ... ("*intelligence is in the eye of the beholder*")

Emulation or simulation? Connectionism



[Images from Wikipedia]

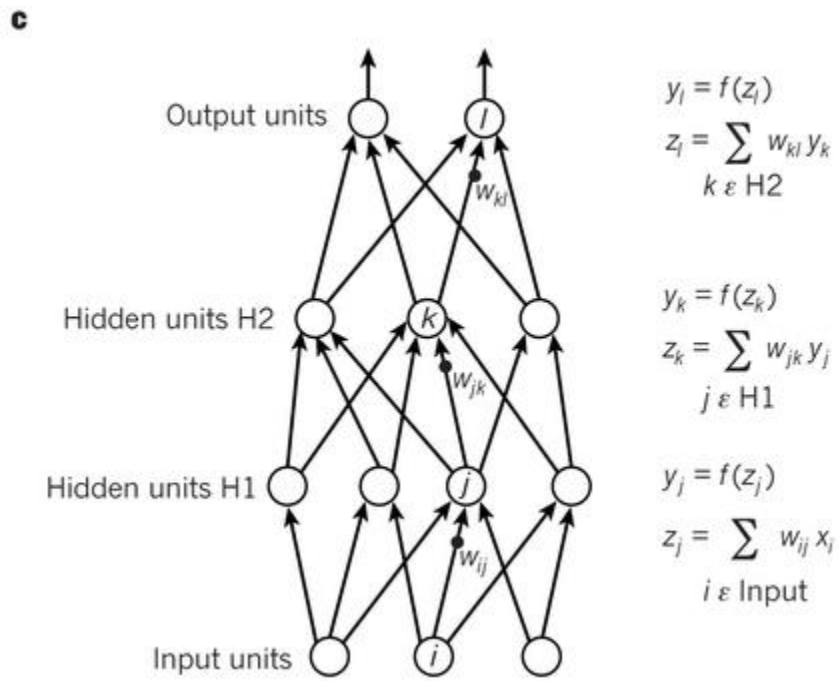
[Rumelhart, D.E., J.L. McClelland 1986]

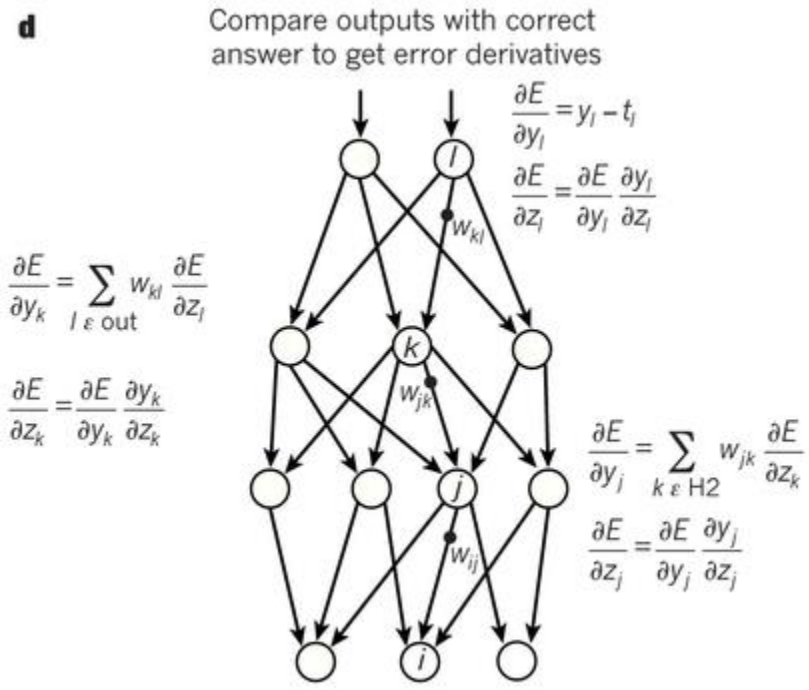
- ***Basic assumption***

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

  "In our view, people are smarter than today's computers because the brain employs a basic computational architecture that is more suited to deal with a central aspect of the natural information processing tasks that people are so good at."
  [Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*]

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:
and the winner was...

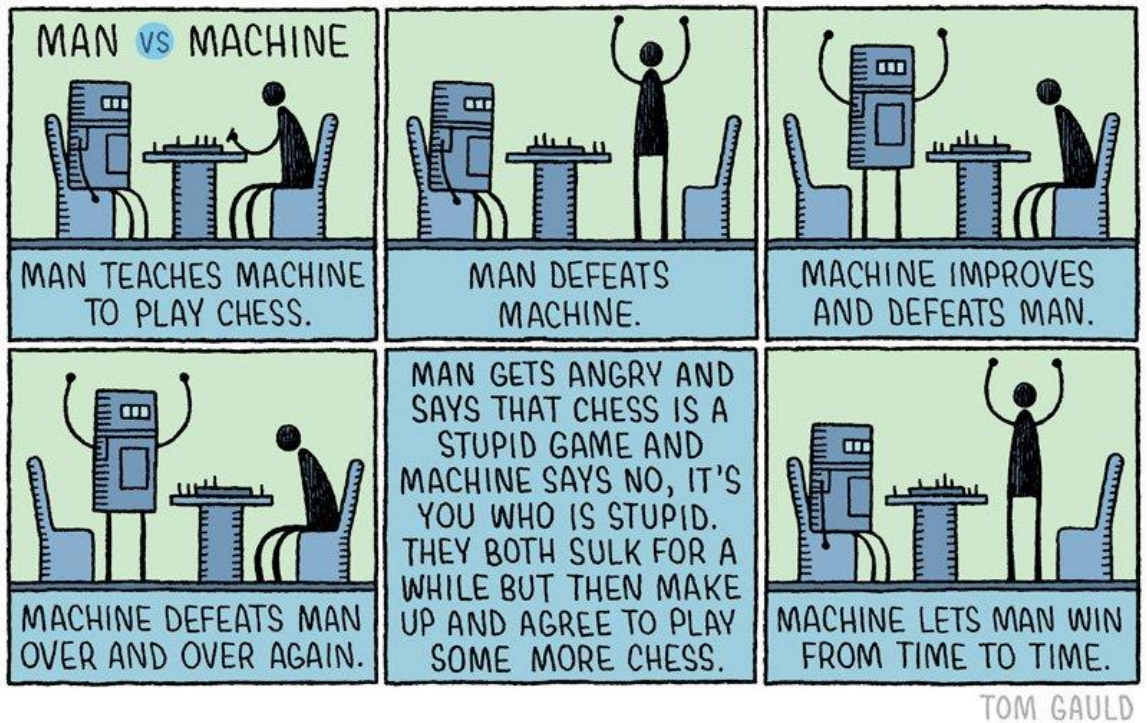
[Image from https://www.tomgauld.com/portfolio

...no one (AI winter)

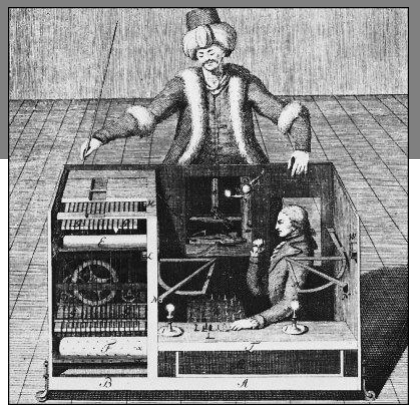
Artificial Brain: can machines think?

(The long and winding road of a question)

Computers play chess



Can machines play chess?



(from Wikipedia)

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*

- 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"

Strategy A or Strategy B?

[Shannon, 1948]

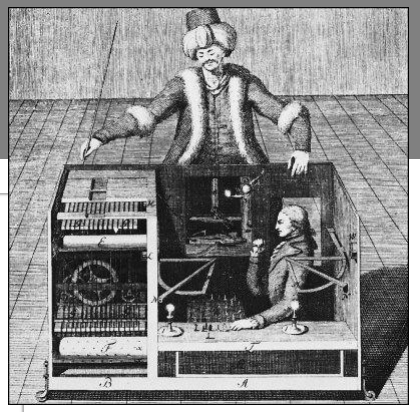
Due to the high computational complexity of Strategy A,
he foresees a progressive development of Strategy B

(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
- *Since then, Strategy A – with significant improvements – dominates the scene*

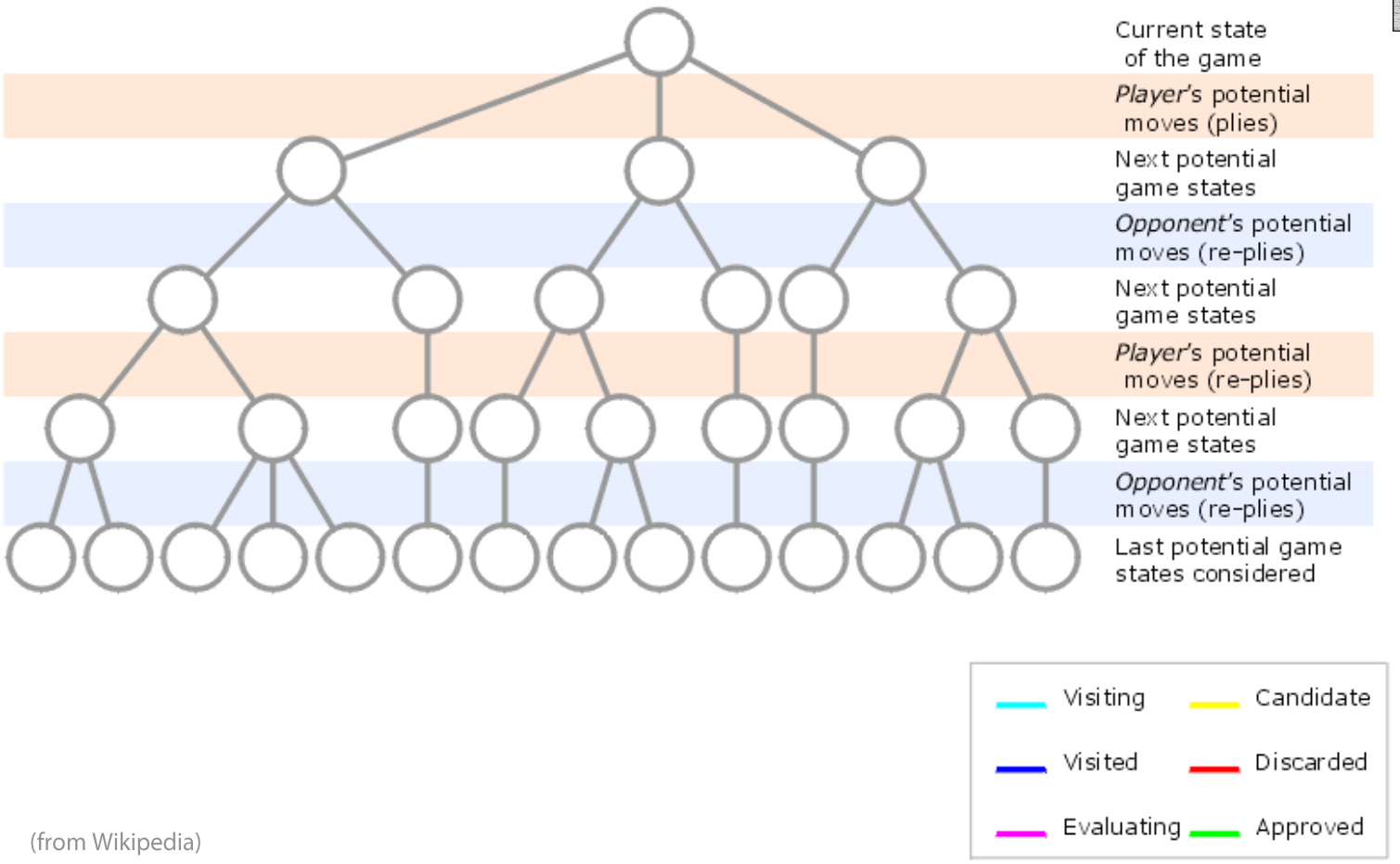
Can machines play chess?



(from Wikipedia)

Strategy A

Minimax on a two-person game tree of 4 plies



(from Wikipedia)

Deep Blue

(from Wikipedia)

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*

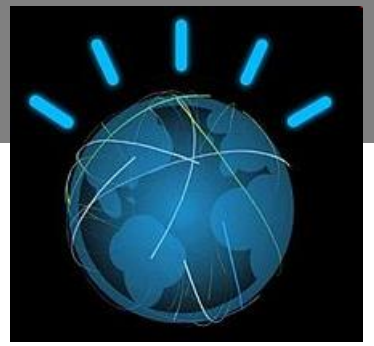
In 1997 the *Deep Blue* system, made by IBM, beats the world chess champion Gary Kasparov

- 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
  - Software 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*)

Computers answer questions

DeepQA (a.k.a. "Watson")



- *Jeopardy!* : a quiz game

(from Wikipedia)

Category: General Science

Clue: When hit by electrons, a phosphor gives off electromagnetic energy in this form.

Answer: Light (or Photons)

Category: Diplomatic Relations

Clue: Of the four countries in the world that the United States does not have diplomatic relations with, the one that's farthest north.

Answer: North Korea

Category: Rhyme Time

Clue: It's where Pele stores his ball.

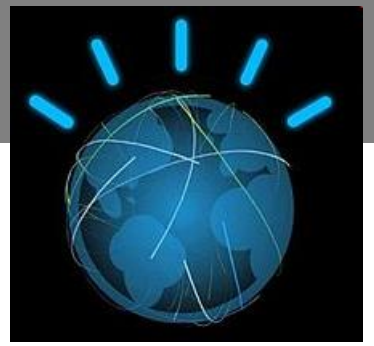
Answer: soccer locker

Category: Lincoln Blogs

Clue: Secretary Chase just submitted this to me for the third time; guess what, pal. This time I'm accepting it.

Answer: his resignation

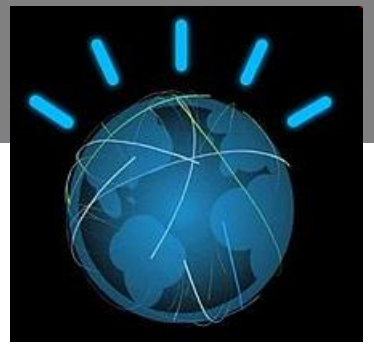
DeepQA (a.k.a. "Watson")



(from Wikipedia)

- DeepQA, 2010 [Ferrucci, D., et al. 2010]
  - The Event (14-18/02/2011)
    - In a sequence of three "*Jeopardy!*" games, Watson beats in a very convincing way the all-times human champions
      - Brad Rutter, winner of the highest amount of money
      - Ken Jennings, winner of the longest string of games
  - *Jeopardy!* : a quiz game
    - In the real game, questions can also be about images, audio or video displays
    - DeepQA can only accept spoken text as input
  - Autonomous search, local memory
    - The rules of the challenge forbid connecting to Internet during the game: DeepQA must use its local memory only
    - It does use Internet during training
  - Conventional hardware, massive parallelism
    - High Performance system, with 2880 standard CPUs (no specialized hardware required)
    - Linux SUSE ES 11, Software in Java and C++, with Apache Hadoop and Apache UIMA
  - (IBM makes business on Watson, now)

Is Watson intelligent?



(from Wikipedia)

- "Does Watson Think?"

  [D. Ferrucci, transcript from video
  http://www.ted.com/webcast/archive/event/ibmwatson]

  "Huh, hmm, what's my favorite response on that?
  (*Do submarines swim?*) […]

  I'd like to look at it as a sort of task-based view:
  when you think of Watson playing Jeopardy!
  it is acting like an intelligent Jeopardy! player,
  if you deconstruct its intelligence
  you're gonna find lots of different algorithms
  no one of them you would look at and say
  "Wow! That's really intelligent! It really understands the question!"
  […]

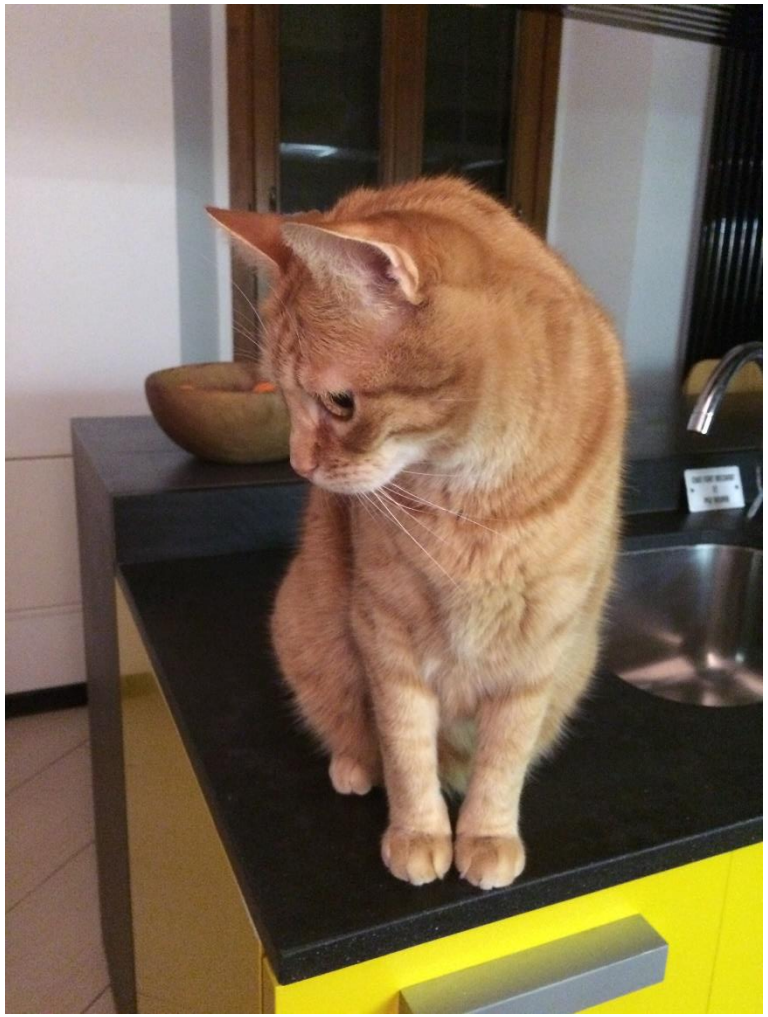
  You have this holistic effect,
  where it's solving a problem that you formally think
  that takes you *think,* to solve that problem, …
  Watson is doing it in a perhaps different way.
  […]

  And I think ultimately of it as a tool, that helps humans solving problems… "

Computers see cats

Artificial Perception

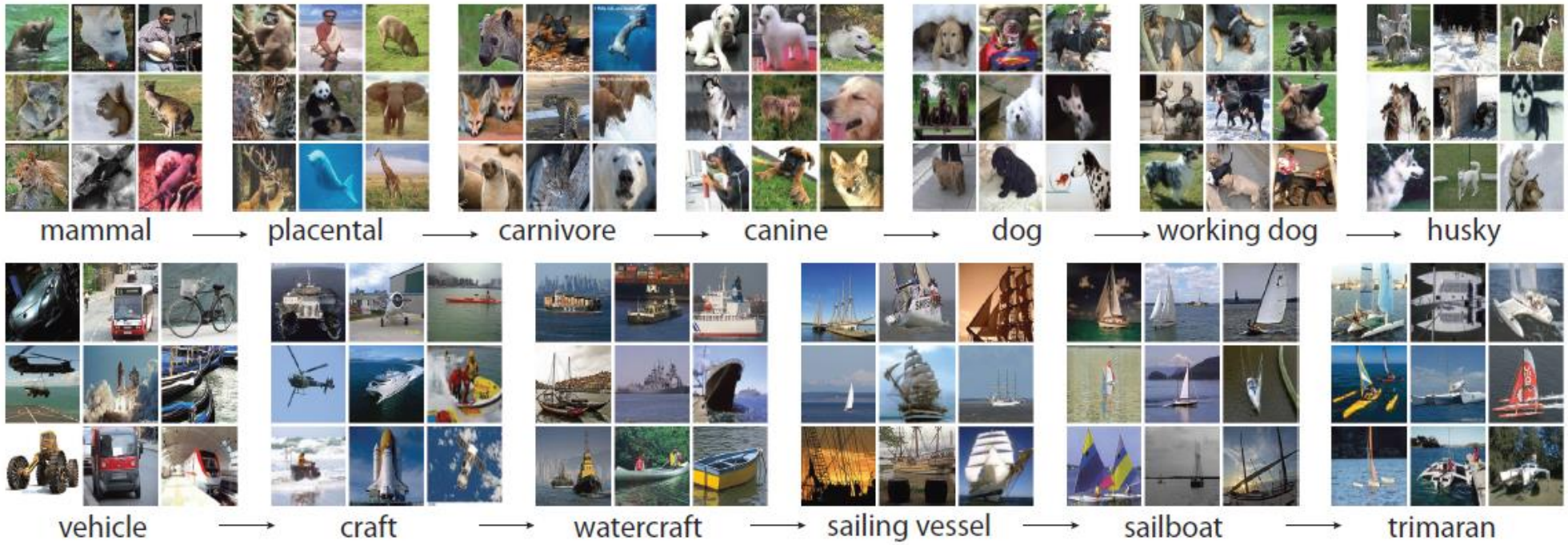
Is there a cat
in this picture?



[this is *my* cat, Rabarbaro]

ImageNet Challenge

- The ImageNet Large Scale Visual Recognition Challenge



1,461,406 full resolution images

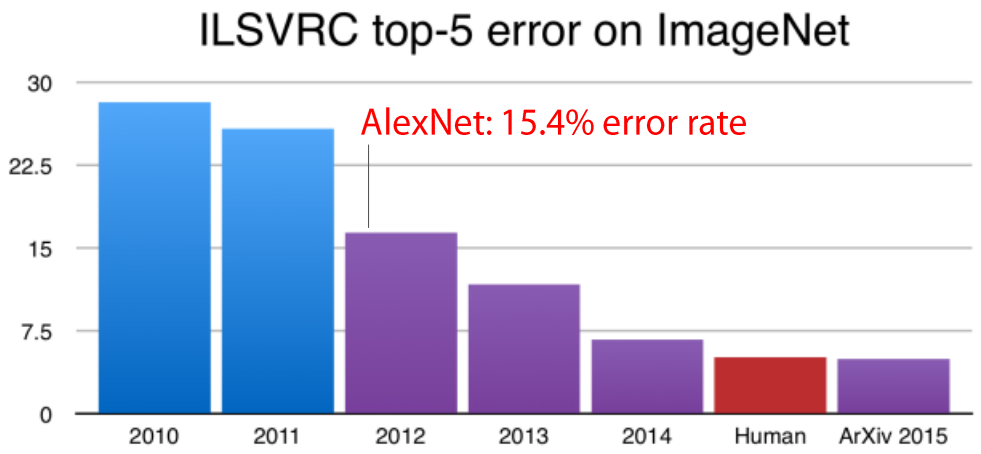
Complex and multiple textual annotation,
hierarchy of 1000 object classes along several dimensions

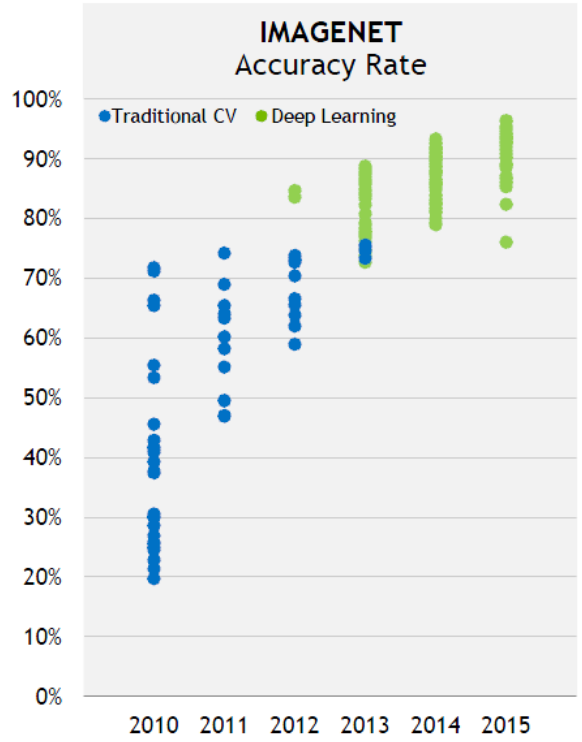
The image classification challenge is run annually since 2010

[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

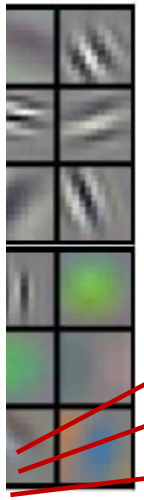
[figures from www.nvidia.com]

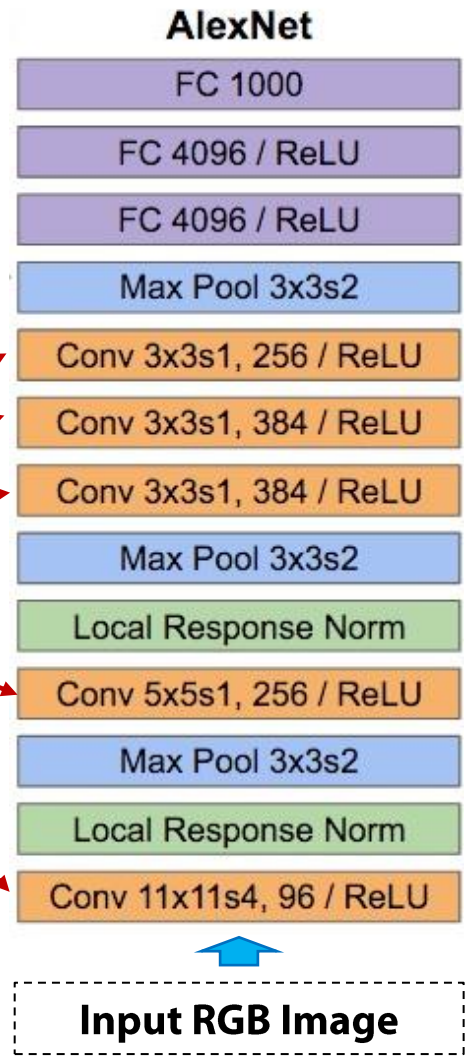
Deep Convolutional Neural Networks (DCNN)

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

Layer 1: 9

Layer 2: 384 deep convc





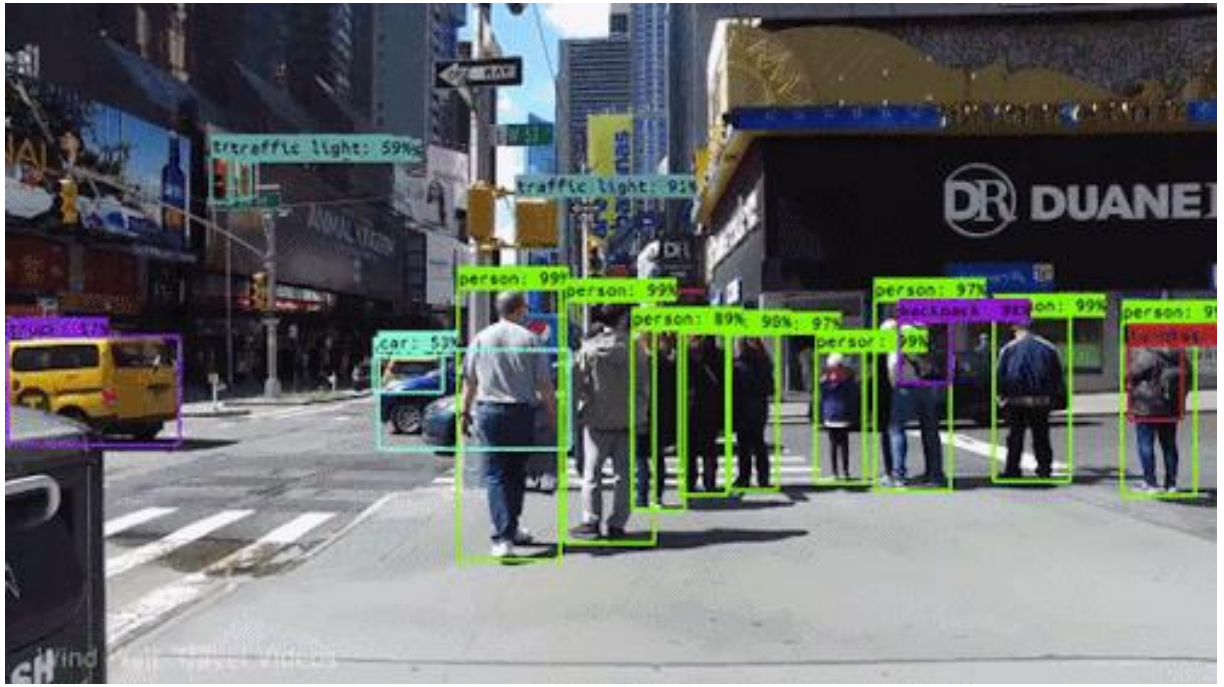
[images from http://cs231n.github.io/convolutional-networks/]

Object (and People) Real-Time Detection

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

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

NOTE:
According to the recent
EU Proposal for a Regulation
about AI, ***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 play other games

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

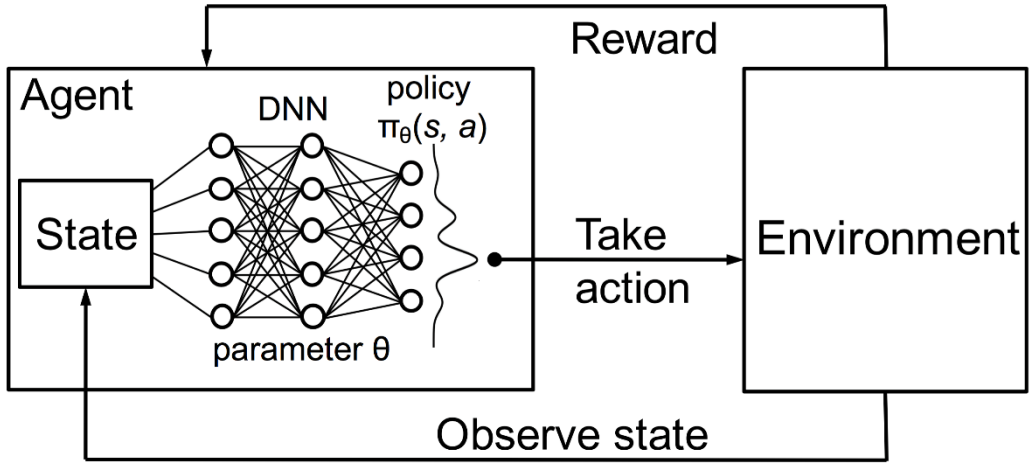
Adversarial self-training:
playing again itself and improving via reinforcement learning

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)

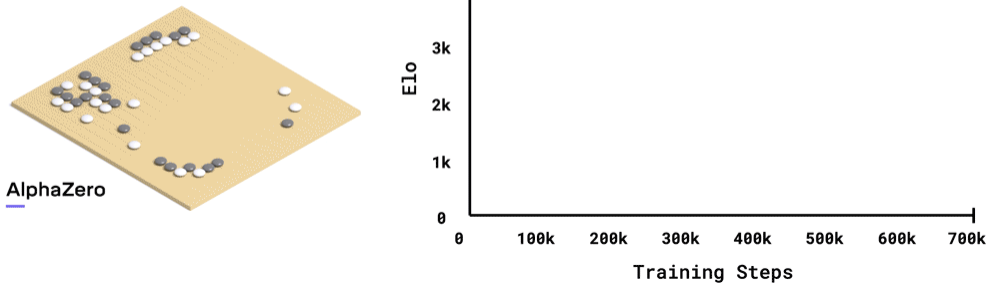
*It selects an **action** in each **state** and receives a **reward** (possibly deferred) as a function of the results obtained*

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

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

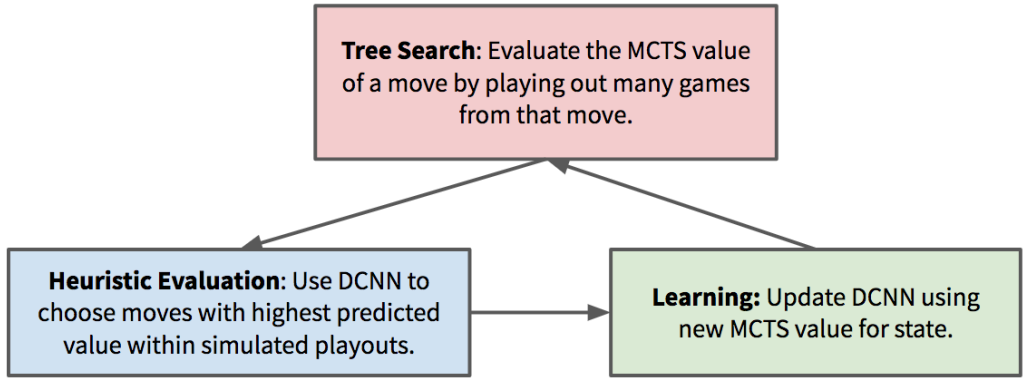
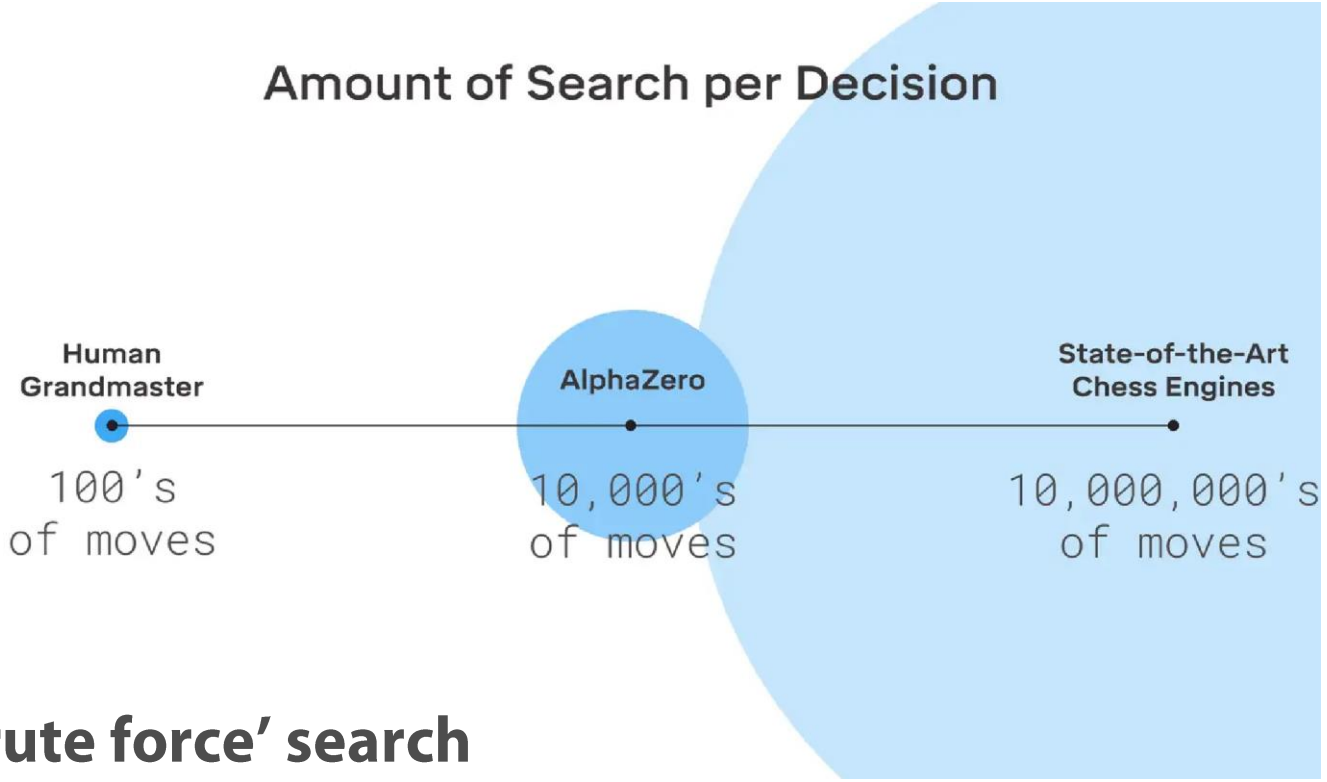


Image from: https://nikcheerla.github.io/deeplearningschool/2018/01/01/AlphaZero-Explained/

Beyond Emulating Humans: AlphaZero (2018)

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



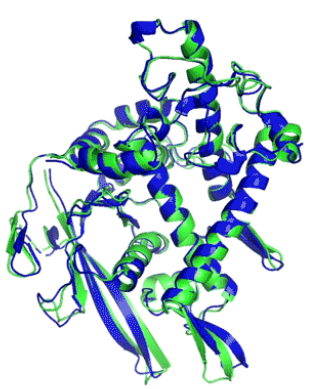
- **AlphaZero uses much less 'brute force' search**
  - When playing, the search process is driven by its neural network
    - *It acts like a memory of past experiences*
  - While training, it learns through a huge amount of self-playing
    - *But it is a faster learner than Alpha Go*

Computers foresee how proteins fold

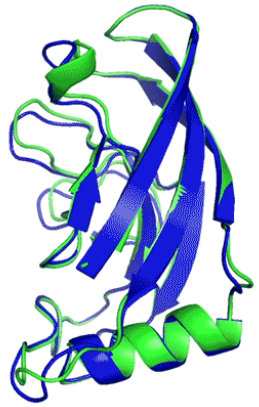
AlphaFold (2020)

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

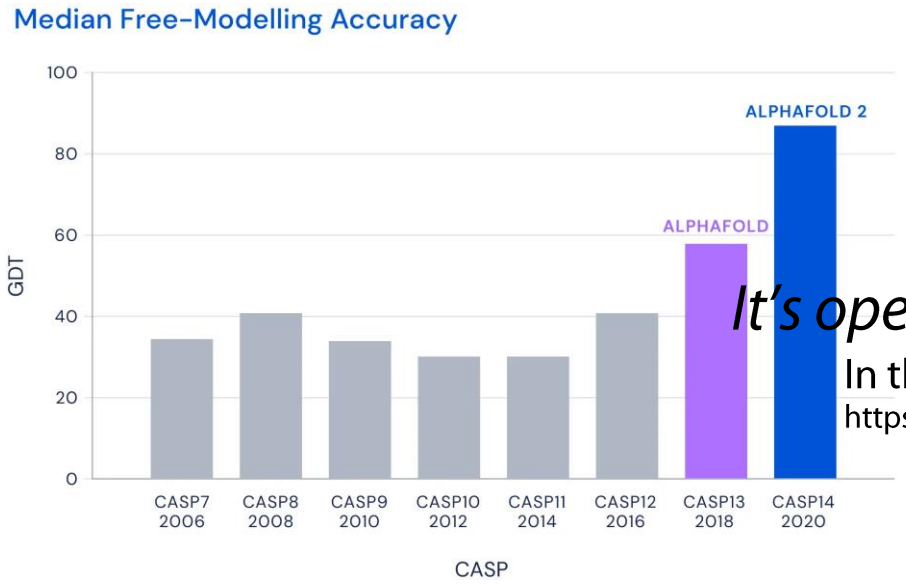
- **Predicting how proteins fold**
  - Several Neural Networks
    - *Trained separately, working together*
  - A big step ahead
    - *Substantial improvement over the best actual systems yet*

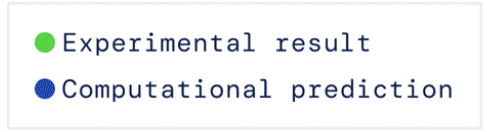


T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



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

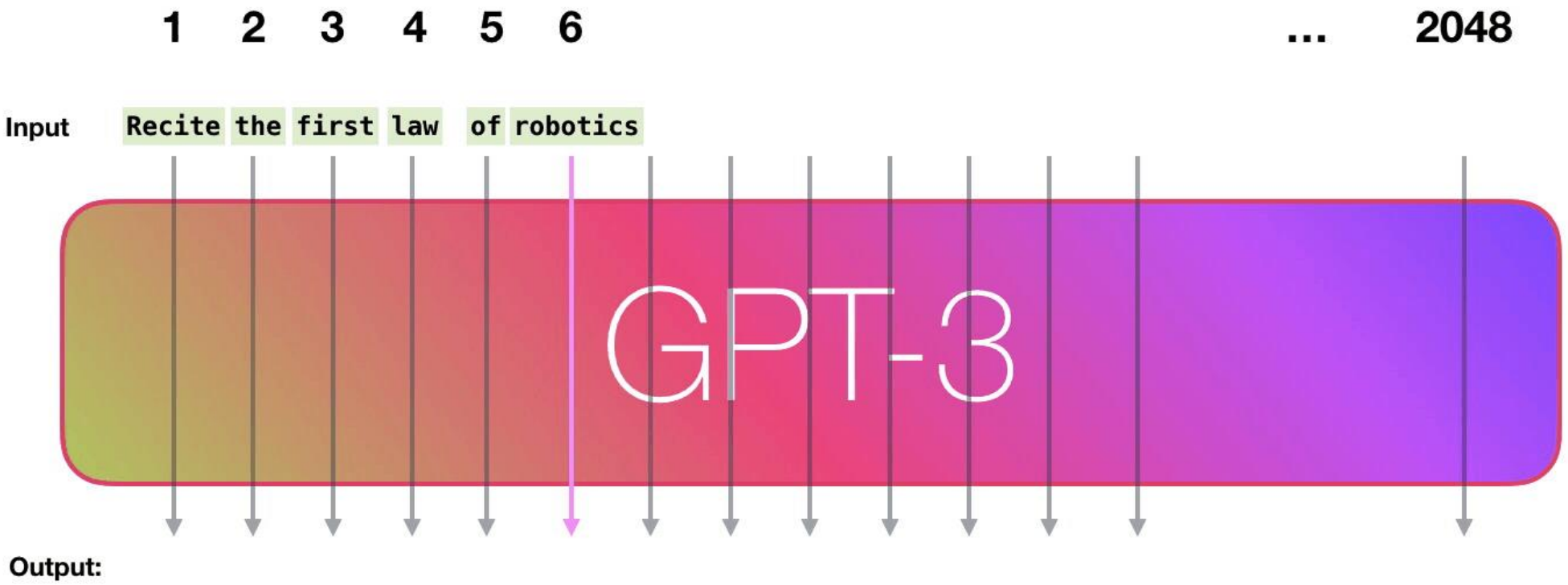
Computers write articles

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)

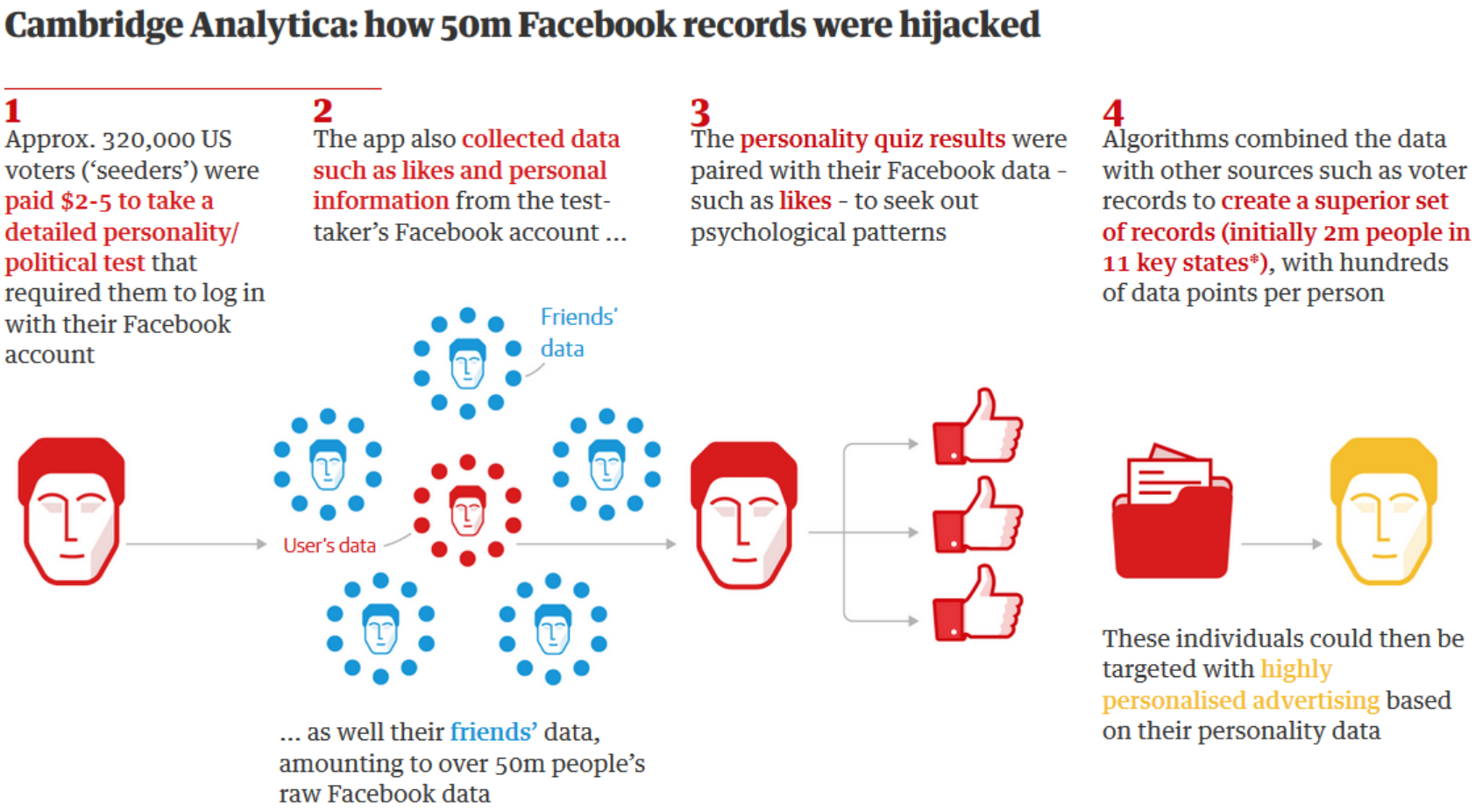
Computer-based personality judgments are more accurate than those made by humans

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)

- **The Cambridge Analytica scandal** [2018]

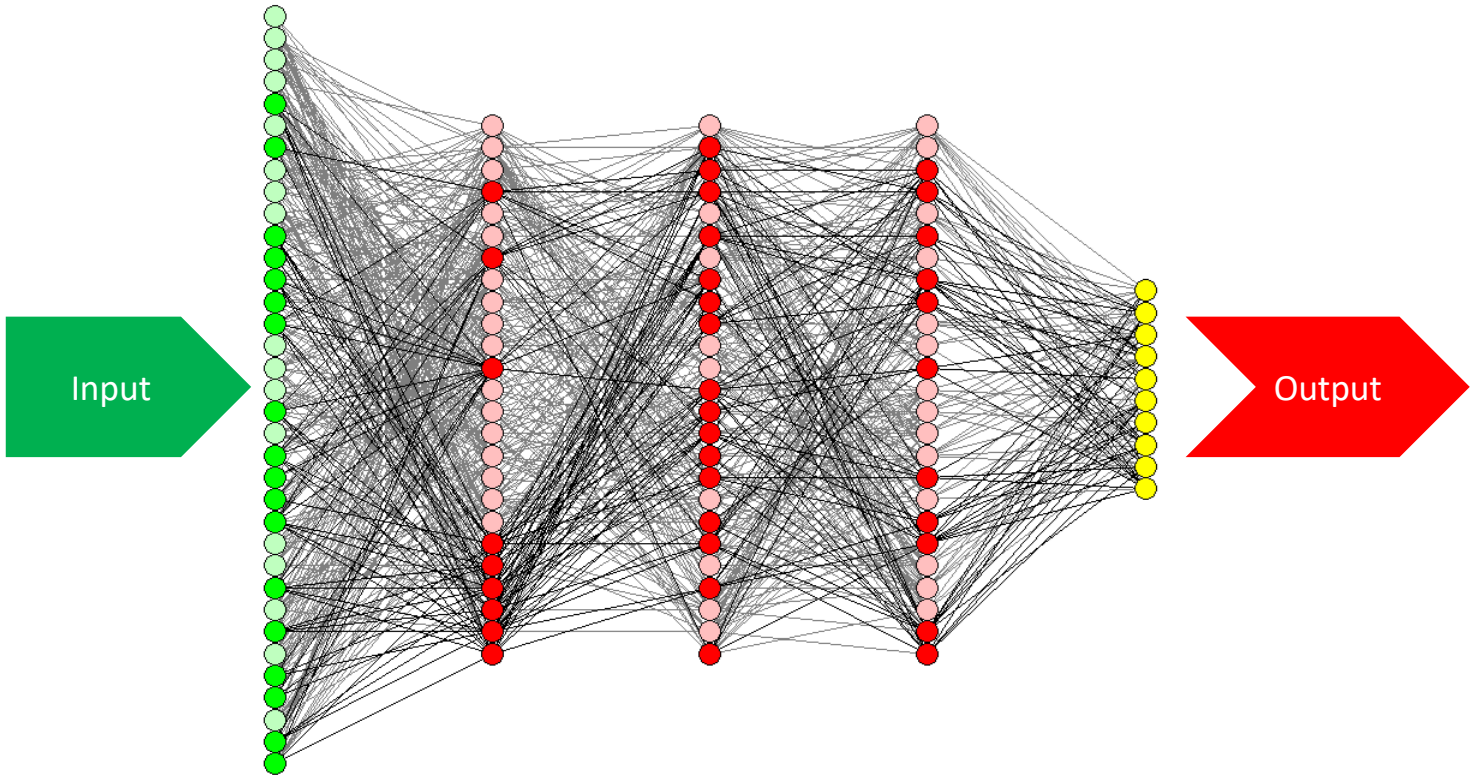


[Graphics from https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm]

The Social Dilemma, 2020 - NETFLIX

Personality and The Machine

- *Opacity*
- *Complexity*
- *Unpredictability*
- *Autonomy*
- *Data*



An EU approach to the regulation of AI

A risk-based approach to regulation

*Not mutually exclusive

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

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]

- pause -

(This page intentionally left blank)

Artificial Intelligence:
a course about foundations

- *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*)