

# UNE BRÈVE HISTOIRE DE L'IA ET DE L'APPRENTISSAGE PROFOND

**Marc Duranton**

CEA Fellow

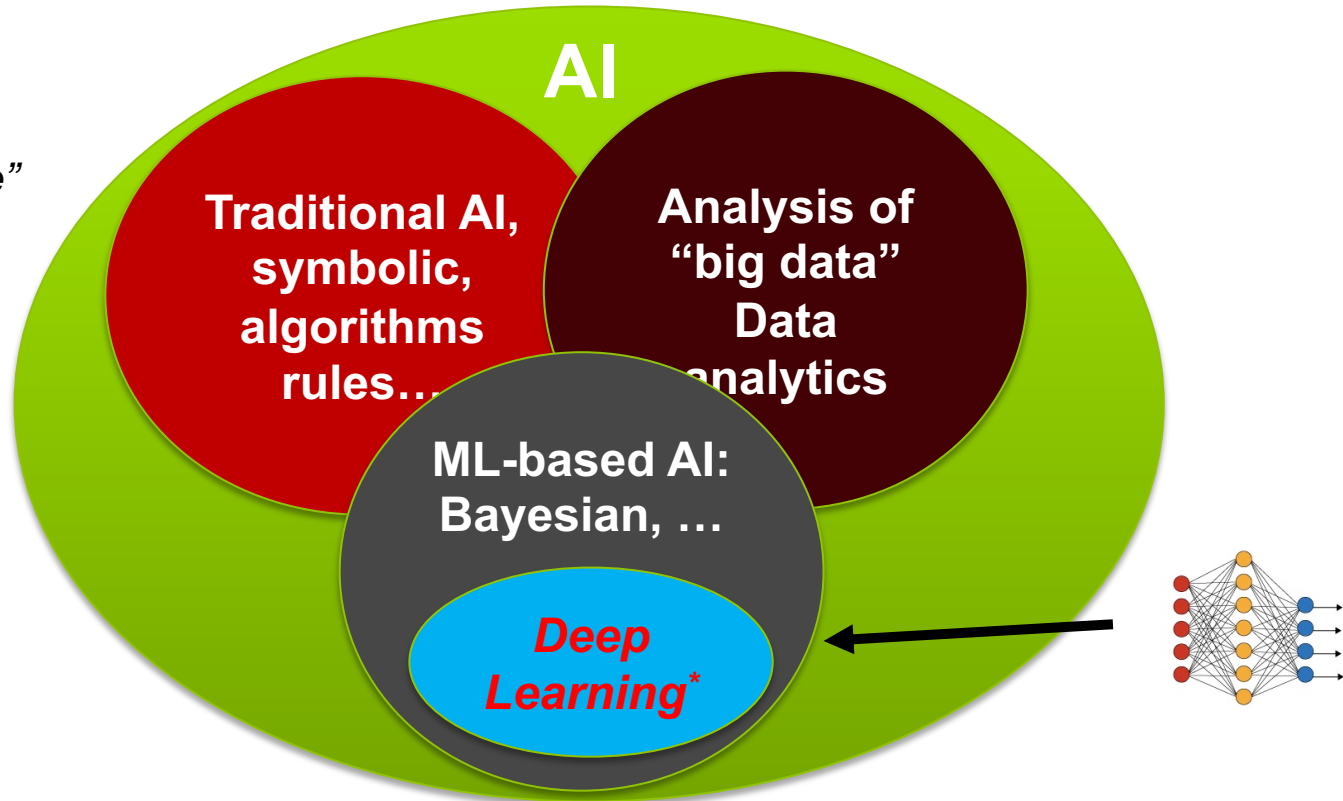
Commissariat à l'énergie atomique et aux énergies alternatives

21 septembre 2020

# KEY ELEMENTS OF ARTIFICIAL INTELLIGENCE

*“...as soon as it works, no one calls it AI anymore.”*

John McCarthy,  
who coined the term  
“Artificial Intelligence”  
in 1956



\* Reinforcement Learning, One-shot Learning,  
Generative Adversarial Networks, etc...

From Greg. S. Corrado, Google brain team co-founder:

- “Traditional AI systems are **programmed** to be clever
- Modern ML-based AI systems **learn** to be clever.

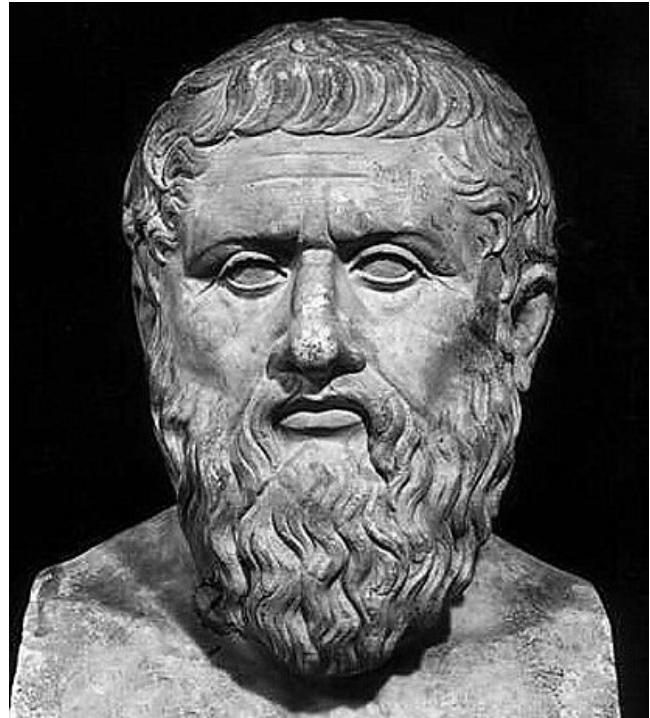


# 387 BC: PLATO

Plato suggested that the brain controls our mental processes

**Human behavior flows from three  
main sources: desire, emotion,  
and knowledge.**

Plato



# 335 BC: ARISTOTLE

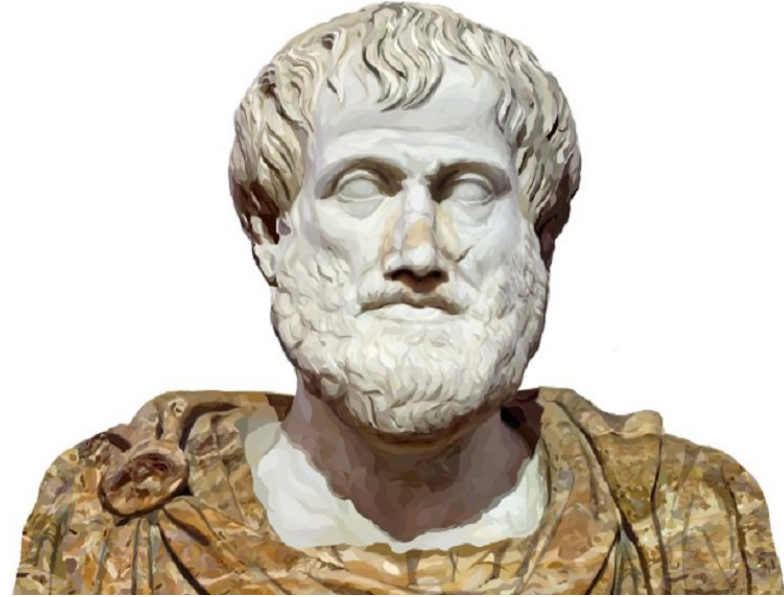
Aristotle invented the inductive reasoning system:

A formal method to represent how human reasons

But...

- He believed that the heart was the seat of behavior,
- He noted the importance of the brain...

...but for cooling the blood !





# NEAR 1644: DESCARTES

## **Descartes' Contributions to Modern Psychology**

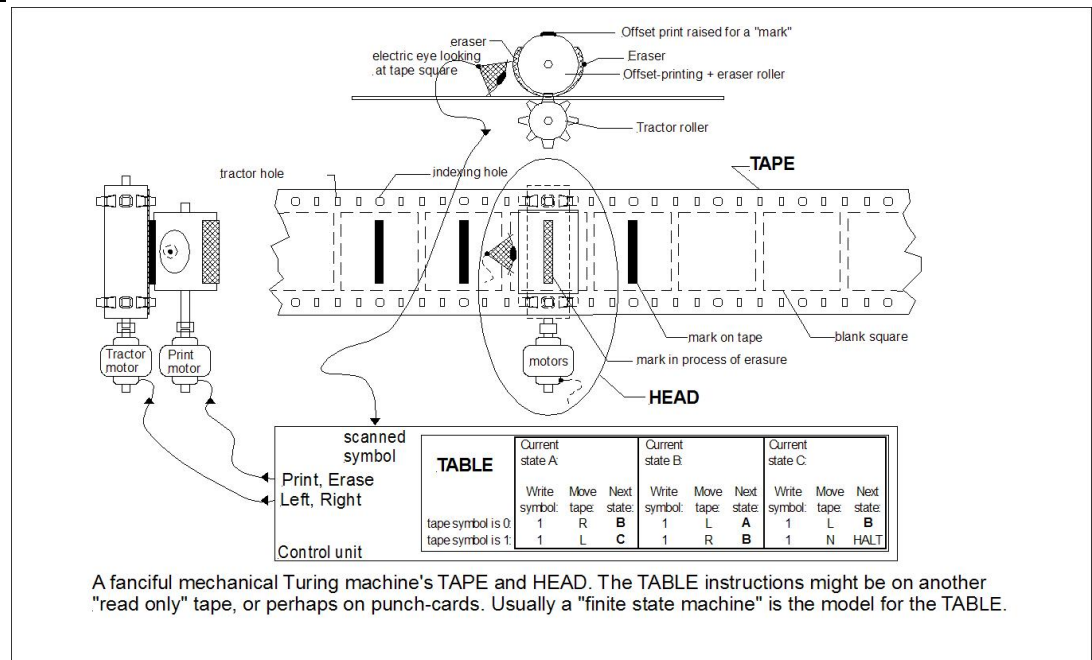
- Laid the foundation for stimulus-response and behavioristic psychology
- Proposed the brain is a mediator of behavior
- Clarified the mind-body relationship
- Arguably the father of physiological and comparative psychology
- Paved the way for the scientific study of consciousness



# 1936: TURING MACHINE

The Turing machine (a-machine - automatic machine) prove properties of computation in general—and in particular, the uncomputability of the Entscheidungsproblem ("decision problem") and prove fundamental limitations on the power of mechanical computation.

A programming language that is Turing complete is theoretically capable of expressing all tasks accomplishable by computers if the limitations of finite memory are ignored.

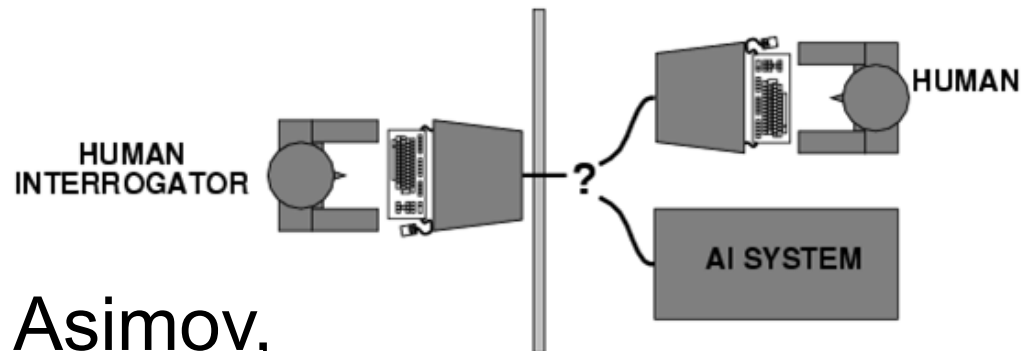
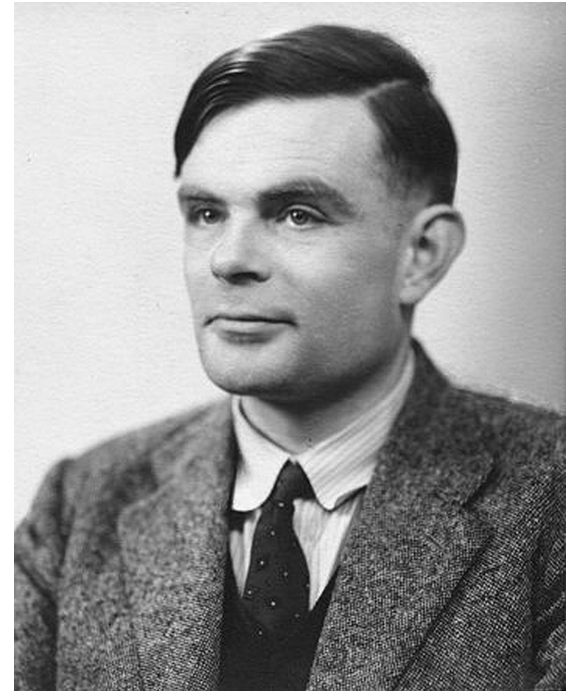




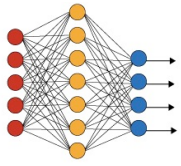
# 1942: ALAN TURING

1942: Any form of mathematical reasoning can be made by a machine.

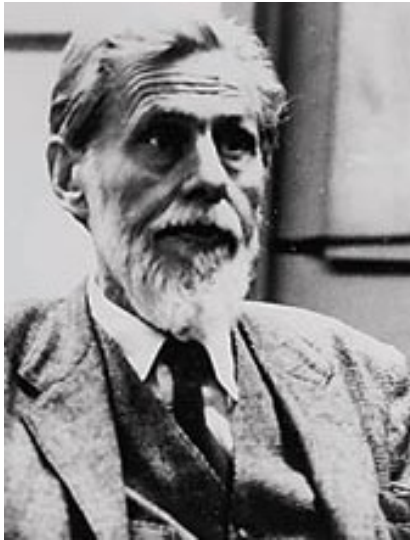
1950: He invented the “Turing test” to check if a system is “intelligent”, i.e. undisguisable from a human



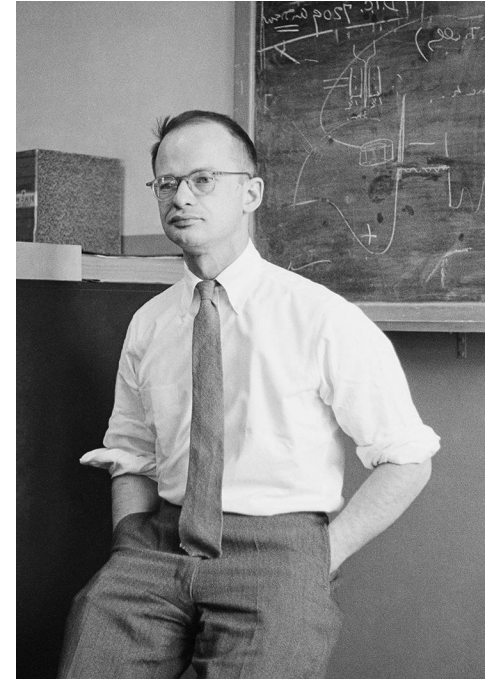
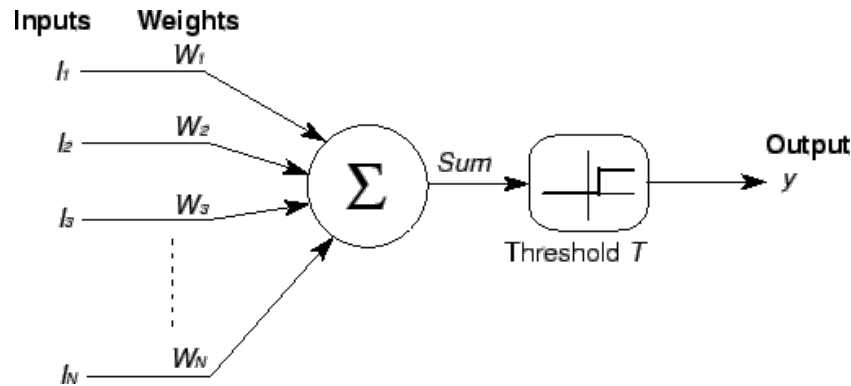
The same year, Isaac Asimov, invented the 3 (4) laws of robotics



# 1943: MCCULLOCH AND PITTS



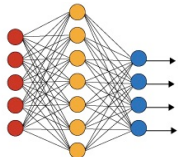
Neurophysiologist and cybernetician



Logician working in the field of computational neuroscience

They laid the foundations of formal Neural Networks





# 1943: MCCULLOCH AND PITTS

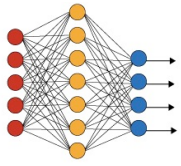
BULLETIN OF  
MATHEMATICAL BIOPHYSICS  
VOLUME 5, 1943

## A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

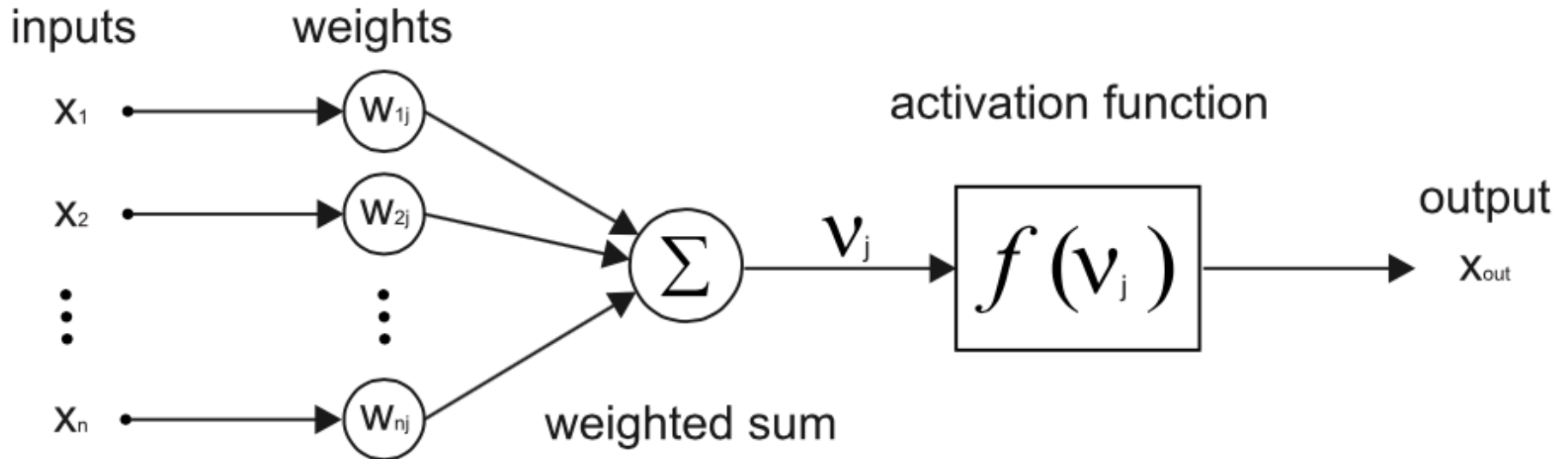
FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,  
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,  
AND THE UNIVERSITY OF CHICAGO

Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

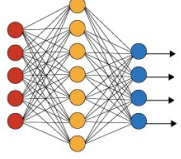


# WHAT IS A NEURAL NETWORK?

**A « formal » neuron:**

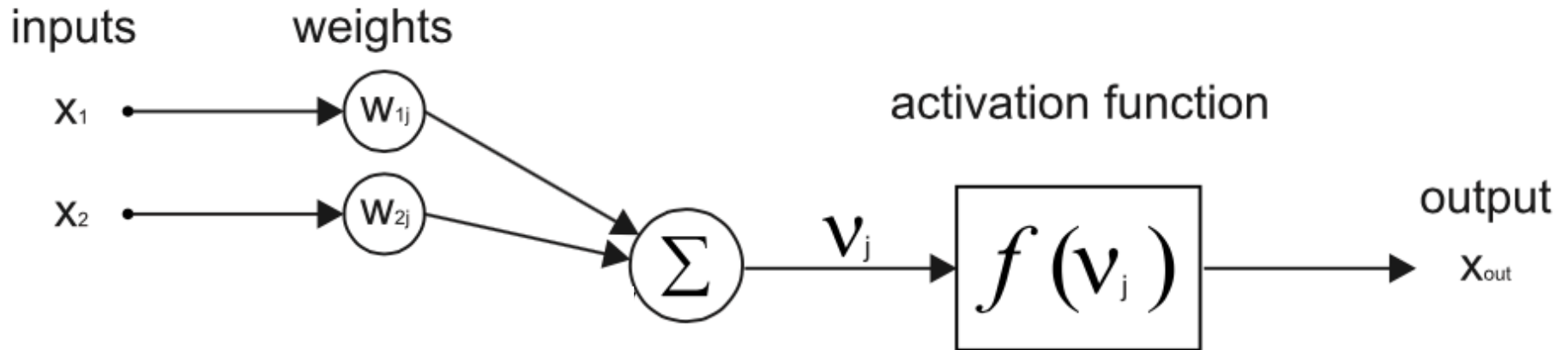






# WHAT IS A NEURAL NETWORK?

The « formal » neuron:

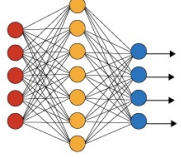


$$V_j = W_{1j} \cdot X_1 + W_{2j} \cdot X_2$$

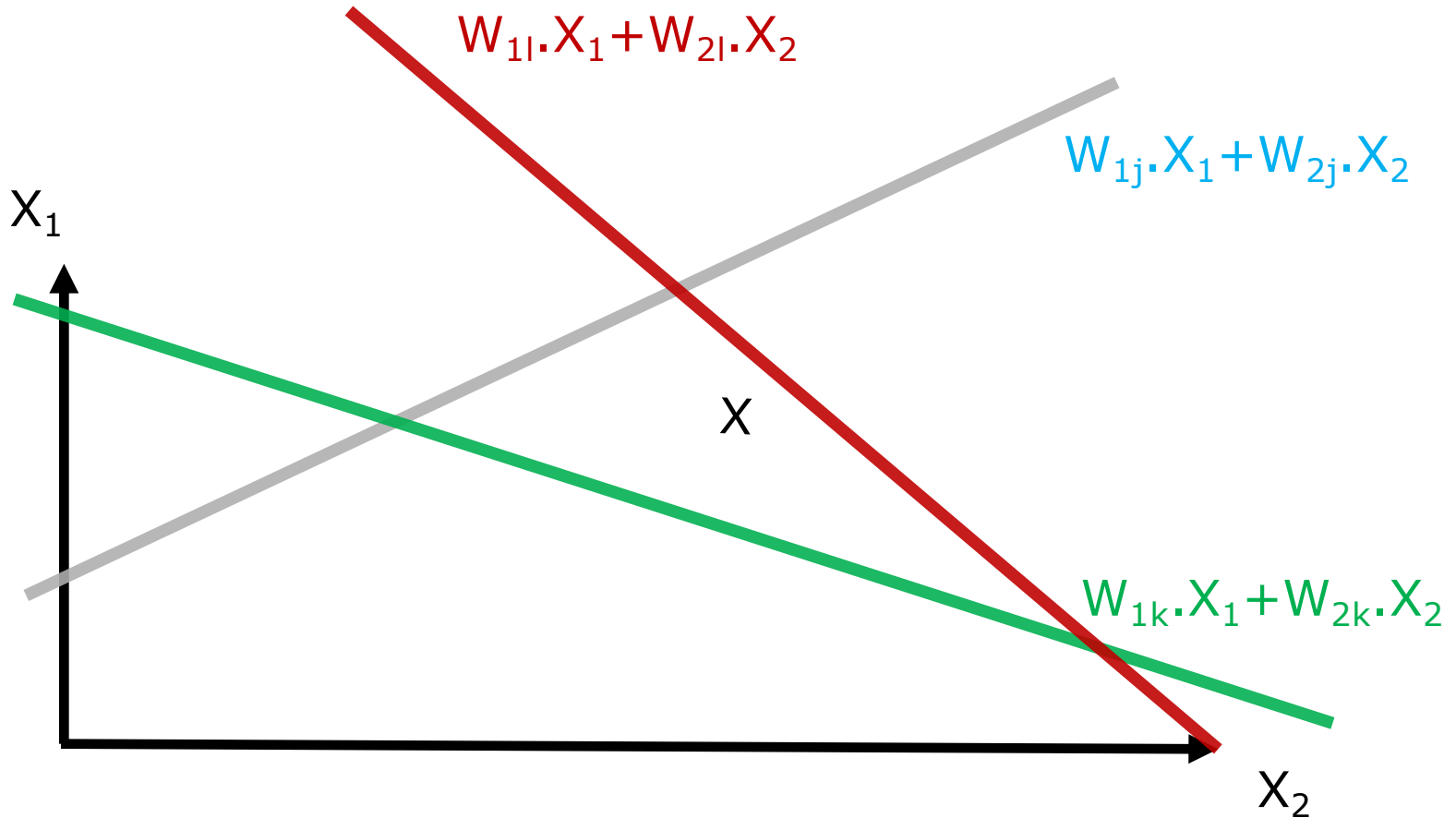
It is the definition of an hyperplane

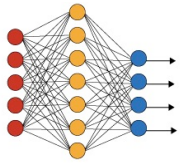
$F(V_j)$  non linear  $\in \{-1, 1\}$  e.g.  $\text{sign}()$  function

$X(X_1, X_2)$  is "above" or "below" the hyperplane



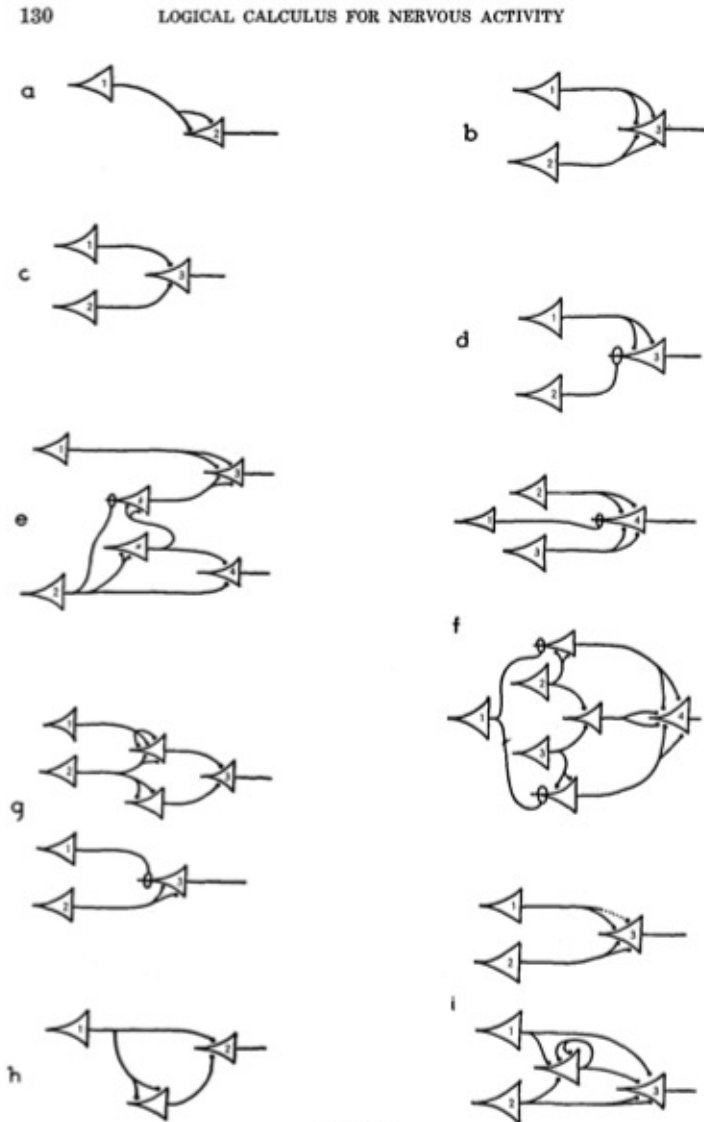
# WHAT IS A NEURAL NETWORK?





# WHAT IS A NEURAL NETWORK?

Association of neurons to make logical functions.  
Example: AND gate



IN 1	IN 2	OUT
0	0	0
0	1	0
1	0	0
1	1	1

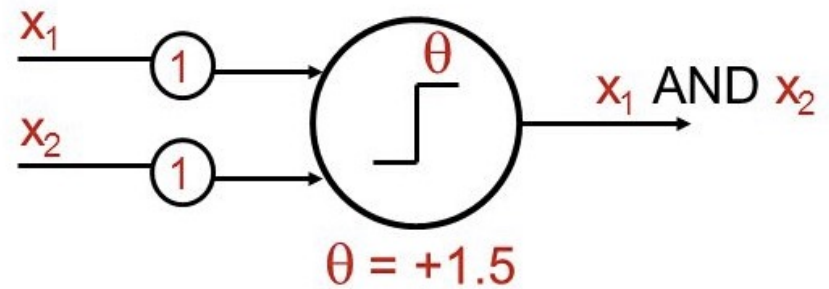
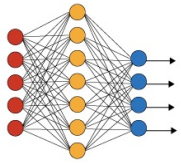
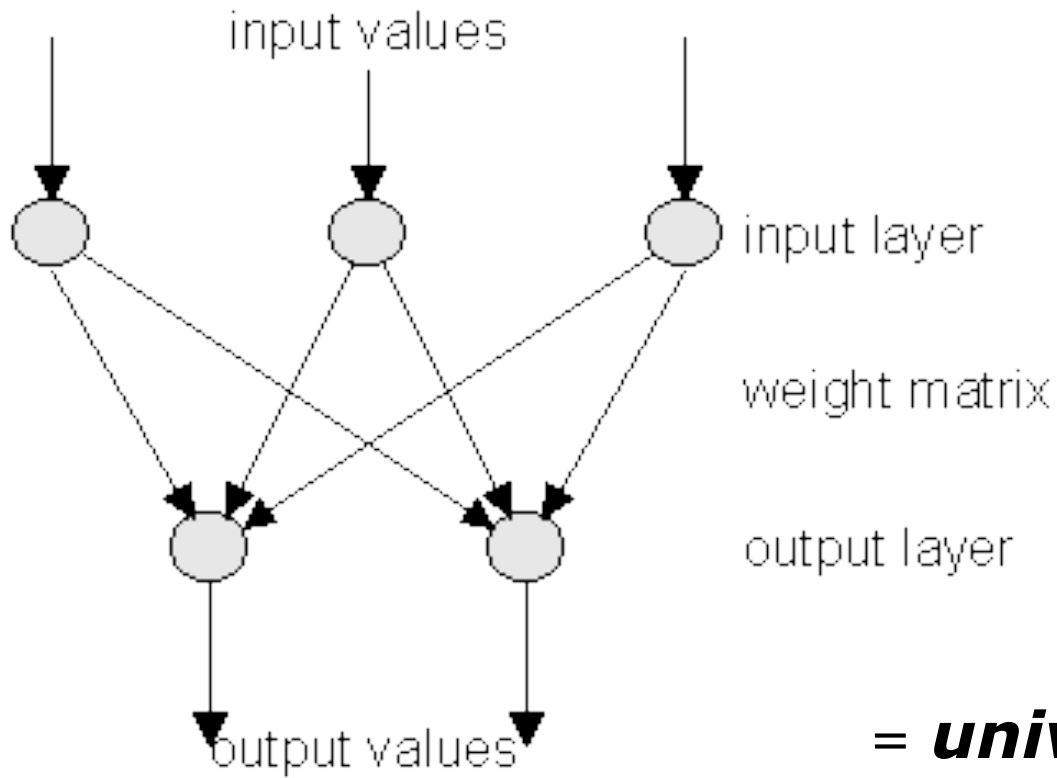


FIGURE 1



# MULTILAYER NETWORK

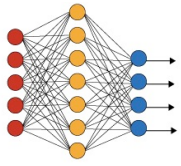


Hyperplane separation

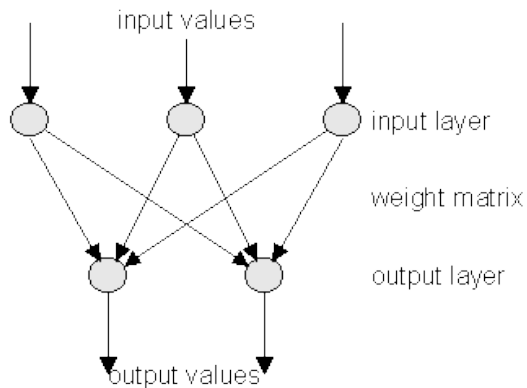
“logic” composition  
Warren McCulloch and  
Walter Pitts, 1943

= ***universal approximator***

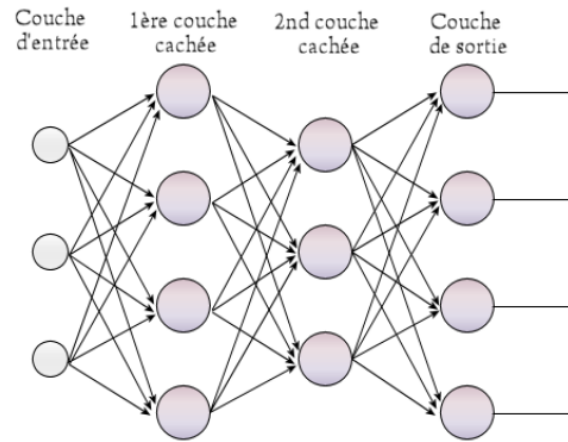




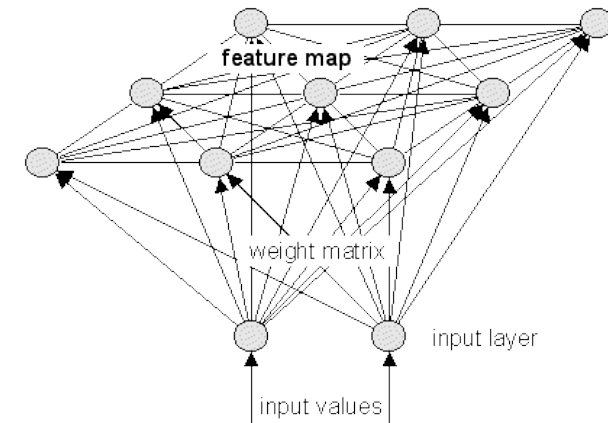
# TOPOLOGY OF NEURAL NETWORKS



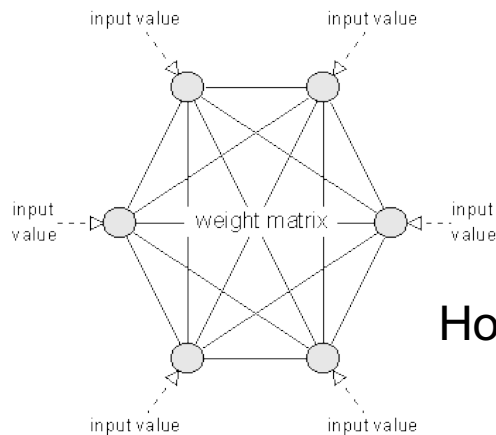
**Perceptron**  
Rosenblatt -- 1957-58



**Multi-layer Perceptron**

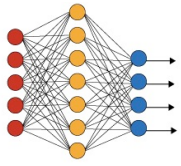


**Kohonen Self-Organizing Maps**

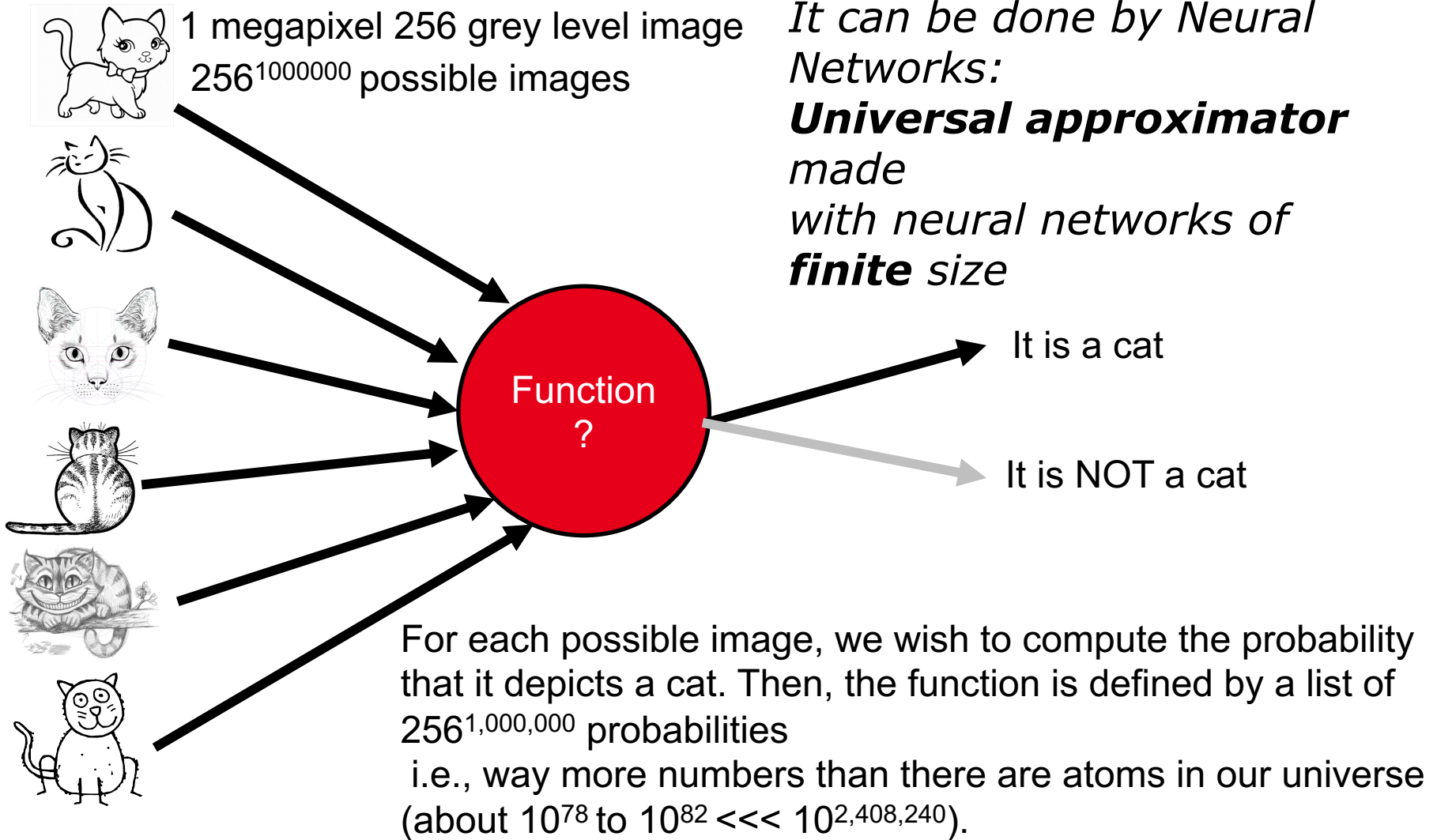


**Hopfield**

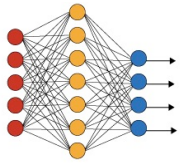
And more...



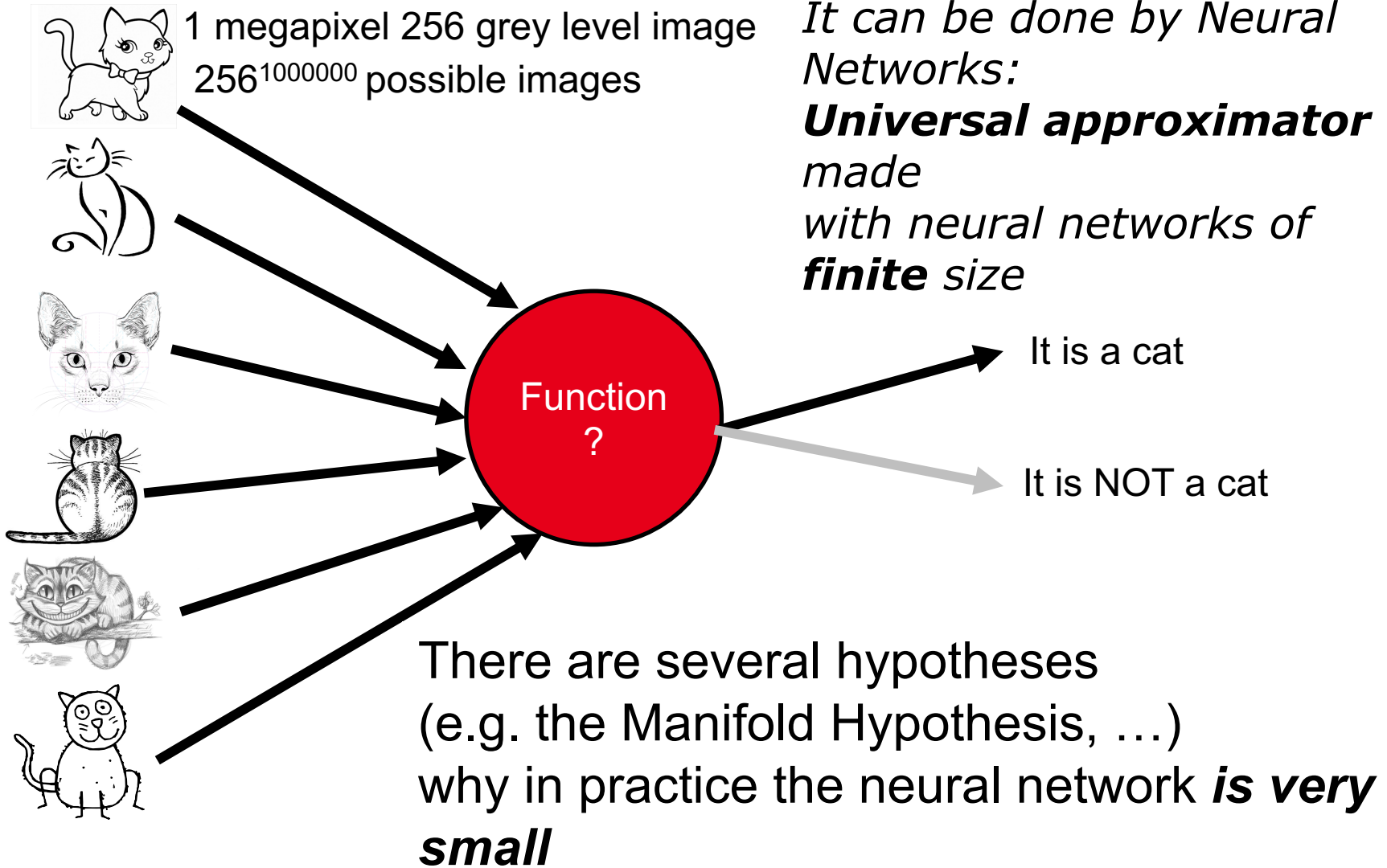
# WHY DOES DEEP LEARNING WORK SO WELL?\*



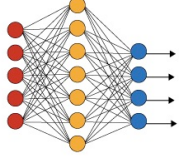
- Work of Henry W. Lin (Harvard) , Max Tegmark (MIT), and David Rolnick (MIT)  
<https://arxiv.org/abs/1608.08225>



# WHY DOES DEEP LEARNING WORK SO WELL?\*



- Work of Henry W. Lin (Harvard) , Max Tegmark (MIT), and David Rolnick (MIT)  
<https://arxiv.org/abs/1608.08225>



# WHY DOES DEEP LEARNING WORK SO WELL?\*

But a picture of a cat is not an arbitrary set of random pixels:

“For reasons that are still not fully understood, our universe can be accurately described by **polynomial Hamiltonians of low order**,”

The laws of physics have other important properties. For example, they are usually **symmetrical** when it comes to **rotation and translation**.

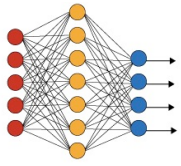
There is another property of the universe that neural networks exploit. This is the **hierarchy of its structure**.

This is why the structure of neural networks is important too: the layers in these networks can approximate each step in the causal sequence.

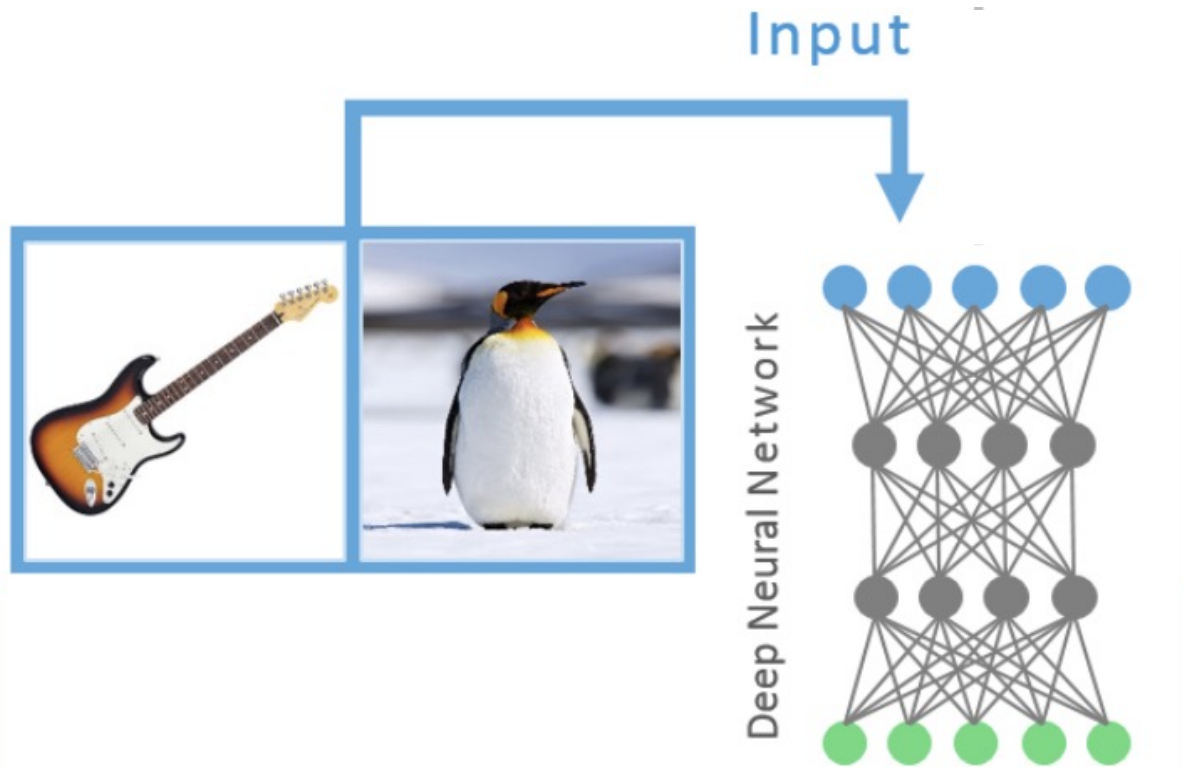
- **These properties mean that neural networks do not need to approximate an infinitude of possible mathematical functions but only a tiny subset of the simplest ones. – *because they are inspired from biological systems that were developed in the context of the real world.***

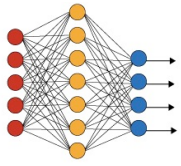
\* Work of Henry W. Lin (Harvard) , Max Tegmark (MIT), and David Rolnick (MIT)



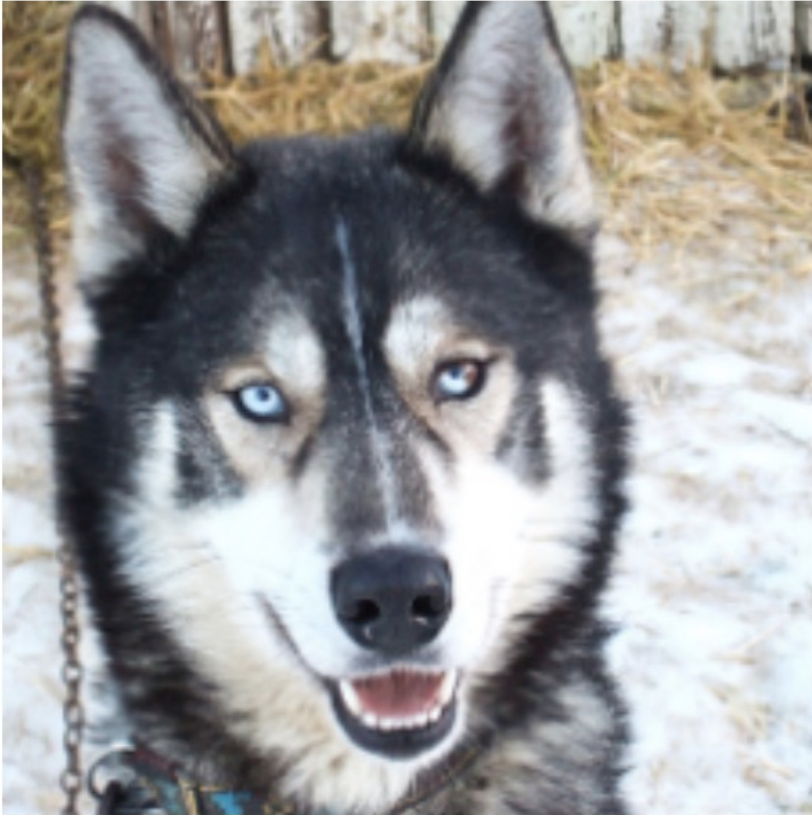


# WHY DOES DEEP LEARNING WORK SO WELL? OR NOT....





## DEEP LEARNING SHOULD LEARN THE RIGHT THINGS



(a) Husky classified as wolf



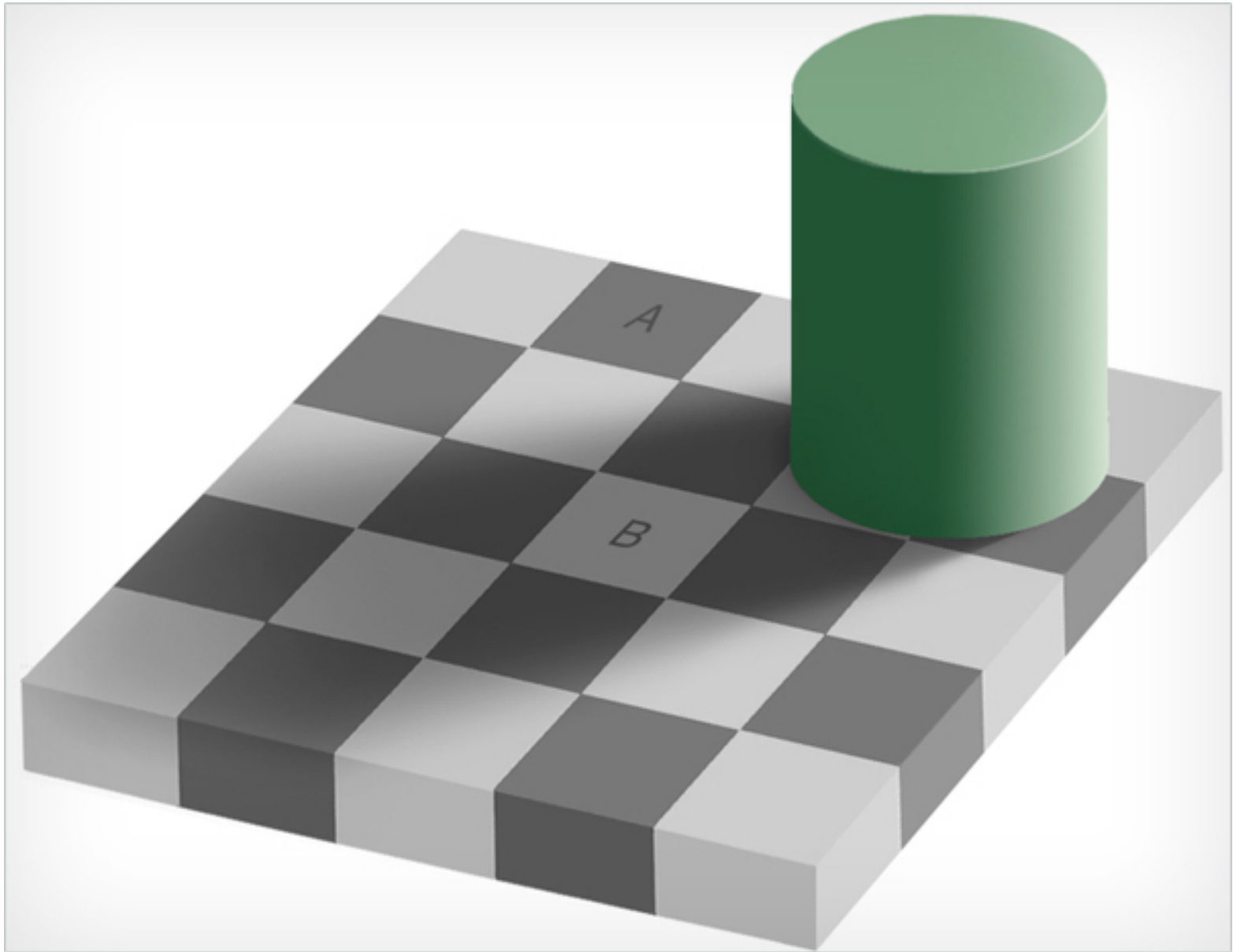
(b) Explanation

From "Why Should I Trust You?": Explaining the Predictions of Any Classifier", Tulio Ribeiro, Marco; Singh, Sameer; Guestrin, Carlos, arXiv:1602.04938, 02/2016.

**BUT OUR BRAIN DOES NOT ALWAYS WORK**

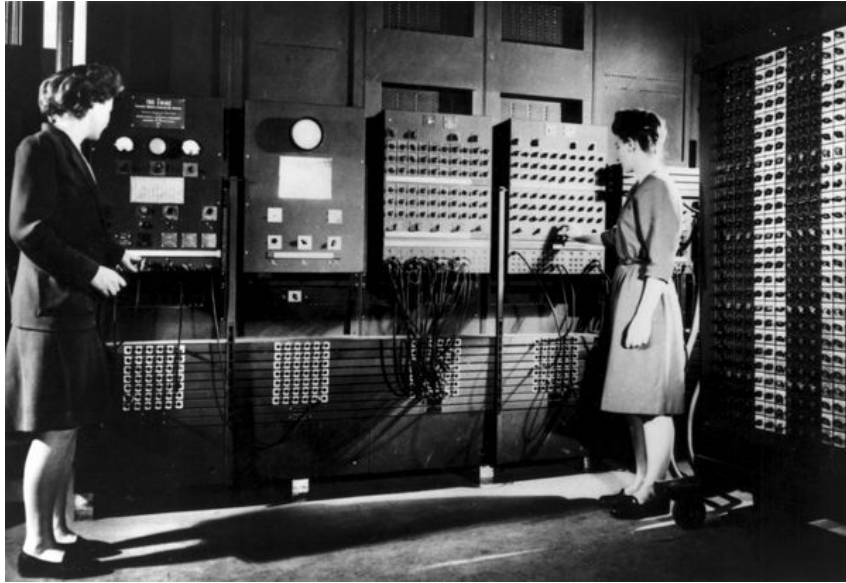


**BUT OUR BRAIN DOES NOT ALWAYS WORK**





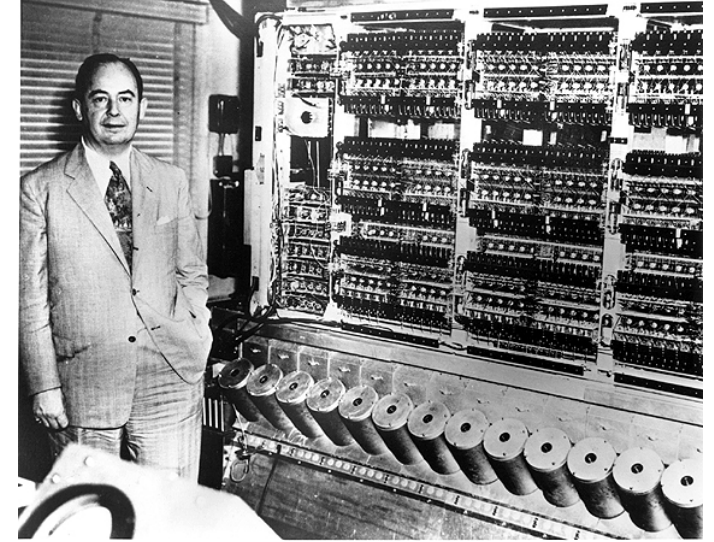
# LOOKING BACK... 1 COMPUTER FOR THE WHOLE PLANET



ENIAC (Electronic Numerical Integrator And Computer), built between 1943 and 1945.

ENIAC contained 20,000 vacuum tubes, 7200 crystal diodes, 1500 relays. It weighed more than 27 t, was roughly 2.4 m × 0.9 m × 30 m in size, occupied 167 m<sup>2</sup> and consumed 150 kW of electricity.

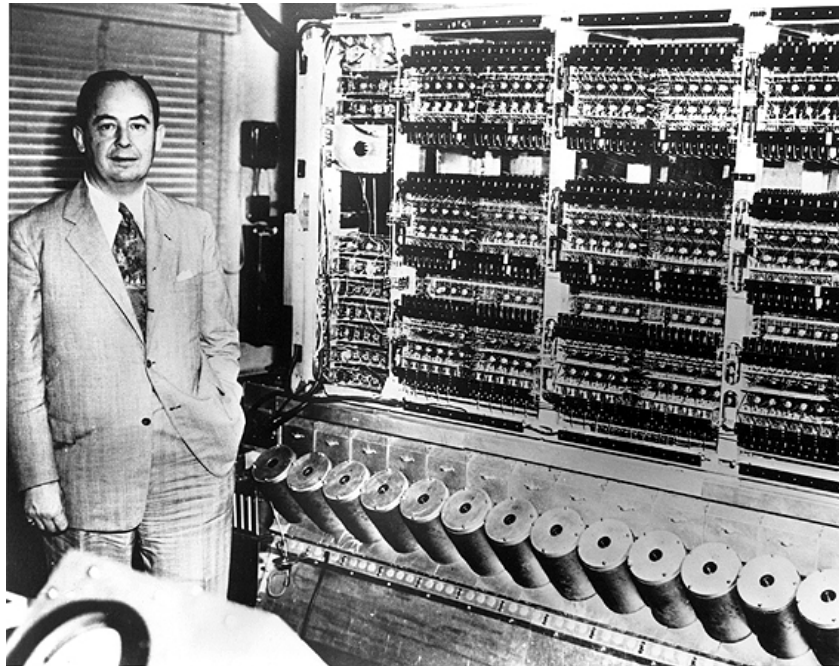
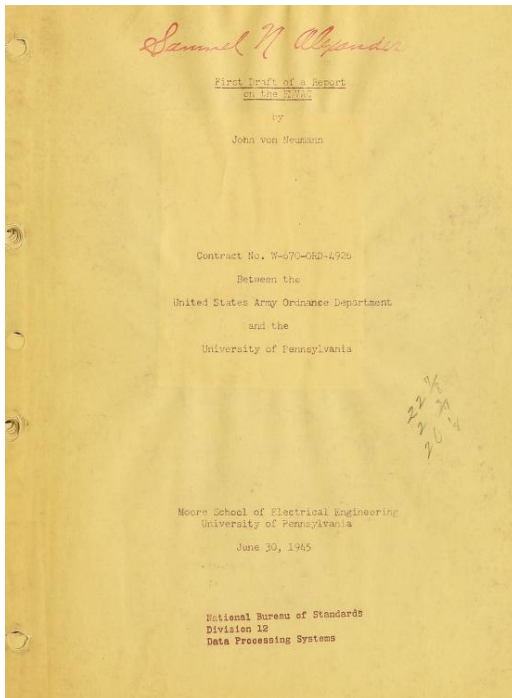
From Wikipedia

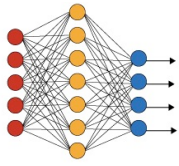


EDVAC was delivered in 1949. Functionally, EDVAC was a binary serial computer with automatic addition, subtraction, multiplication, programmed division and automatic checking with an ultrasonic serial memory capacity of 1,000 44-bit words. EDVAC's average addition time was 864 microseconds and its average multiplication time was 2,900 microseconds.

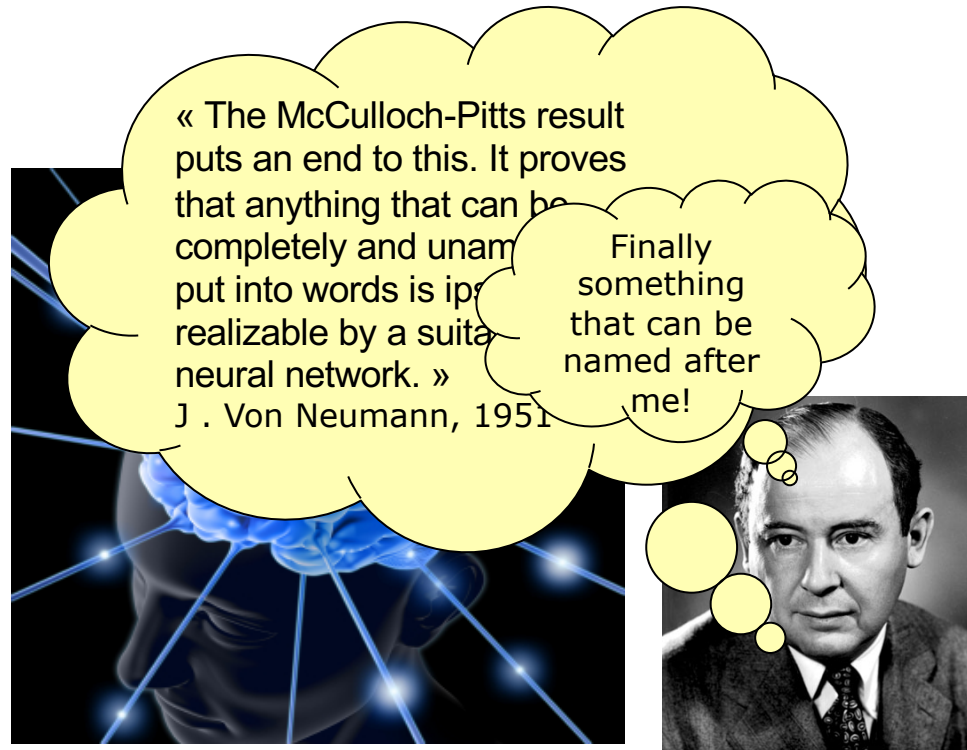
## BACK TO THE ORIGIN: WHAT IS THE TRUE VON NEUMANN ARCHITECTURE?

In “First Draft of a Report on the EDVAC,” the first published description of a stored-program binary computing machine - the modern computer, John von Neumann suggested modelling the computer after Pitts and McCulloch’s neural networks.

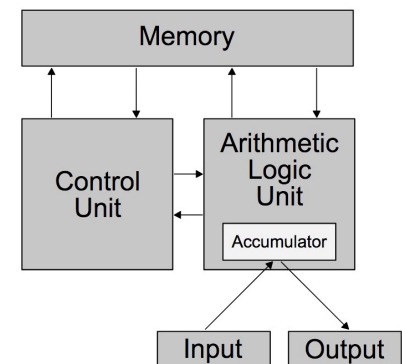




## BACK TO THE ORIGIN: WHAT IS THE TRUE VON NEUMANN ARCHITECTURE?



**But technology was not ready in the 50's,  
leading to realization with sequential processing  
And to the computer architecture we have now...**



## DO YOU KNOW THAT...

John von Neumann

- Founded the field of game theory as a mathematical discipline,
- Created stochastic computing in 1953,
- Constructed the first self-replicating cellular automata,
- Designed a self-reproducing computer program which is considered to be the world's first computer virus in 1949,
- Introduced quantum logic in 1936 with Garrett Birkhoff



# 1948: HOMEOSTAT FROM ROSS ASHBY

The Homeostat is one of the first devices capable of **adapting itself to the environment**, like animals. It exhibited behaviors such as habituation, reinforcement and learning through its ability to maintain homeostasis in a changing environment.

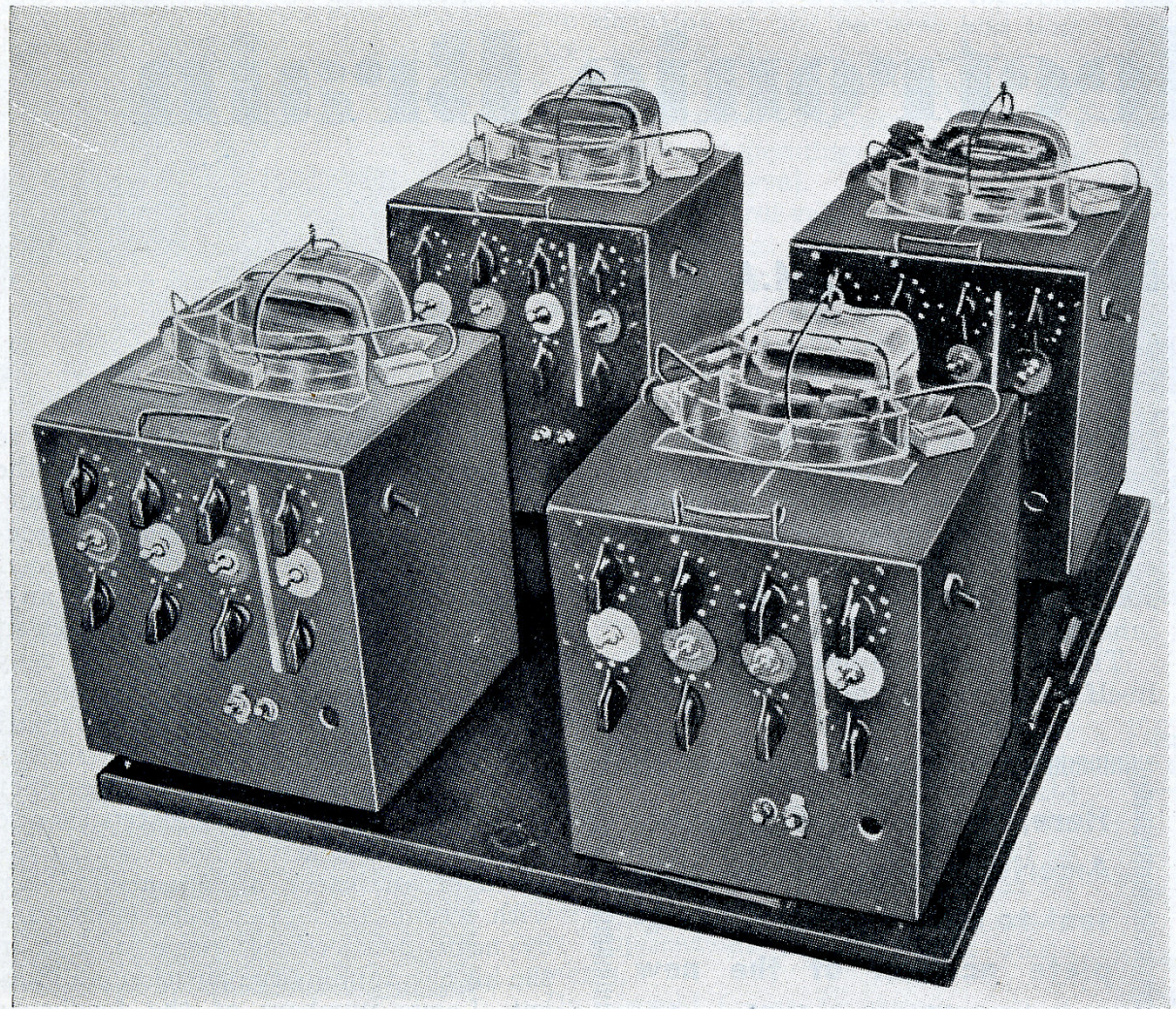
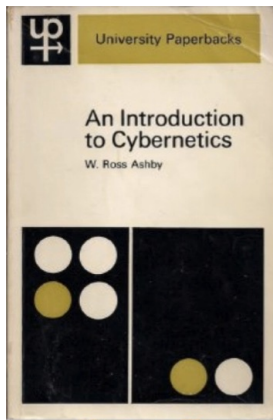


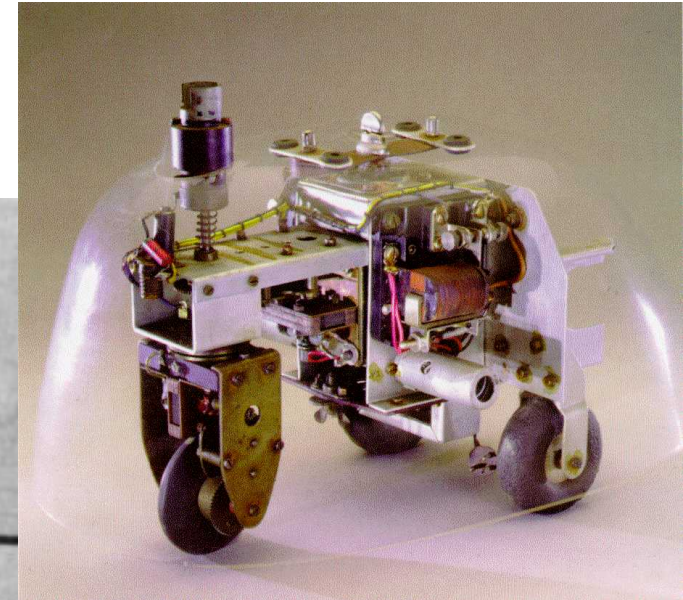
Fig. 1—The homeostat, with its four units, each one of which reacts on all the others.

**Inspired from biological behavior**



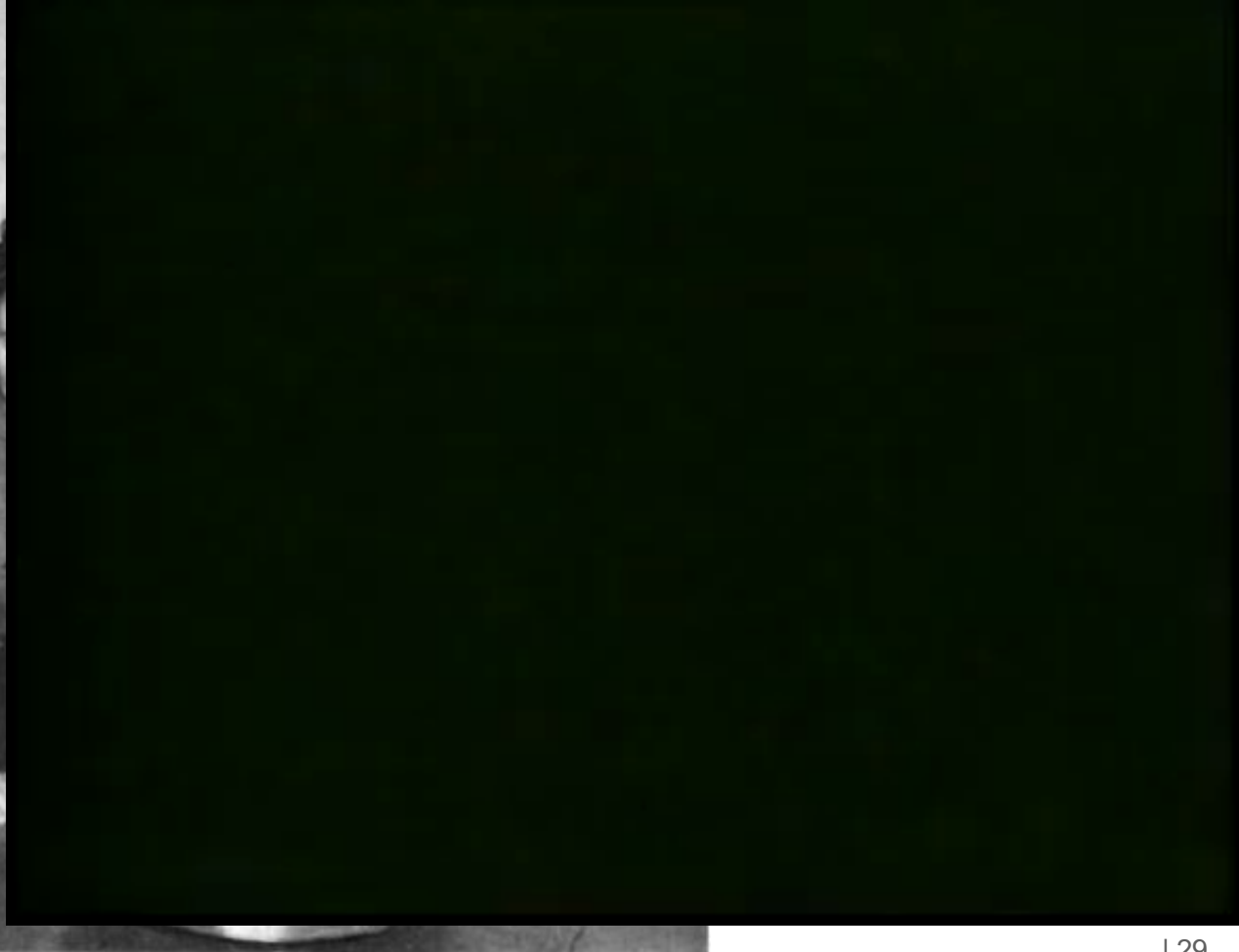
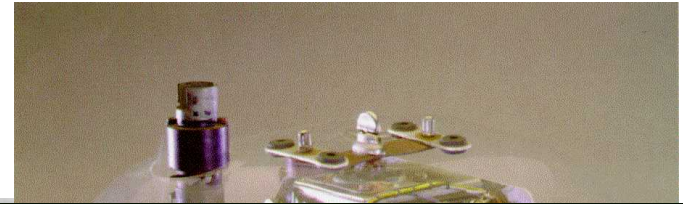
# 1948: GREY WALTER AND HIS TORTOISES

Grey Walter is a neurophysiologist and roboticist. Very simple hardware can have complex behavior as soon as you add feedbacks (from the environment)

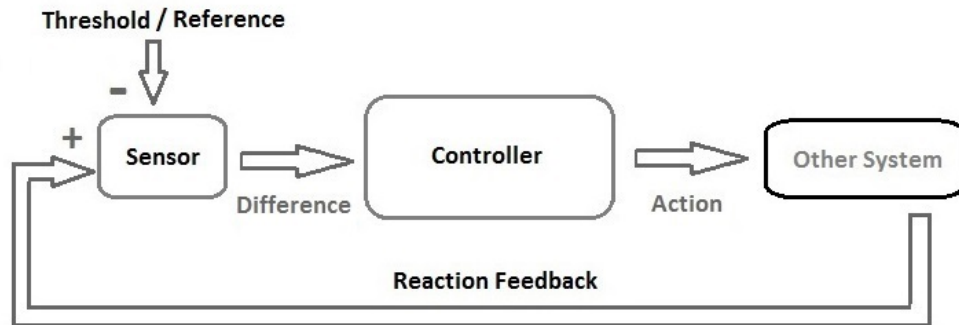
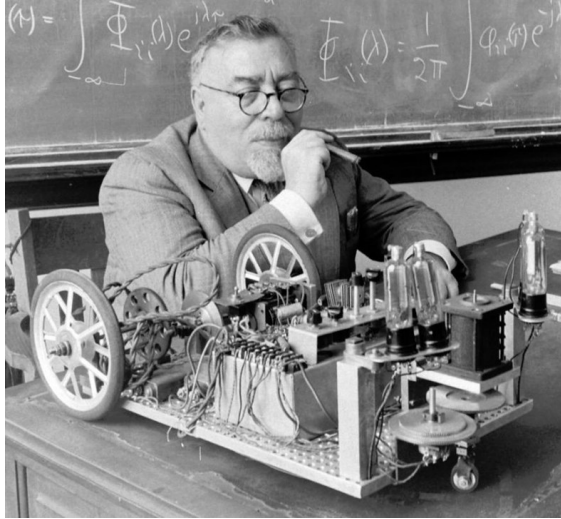


# 1948: GREY WALTER AND HIS TORTOISES

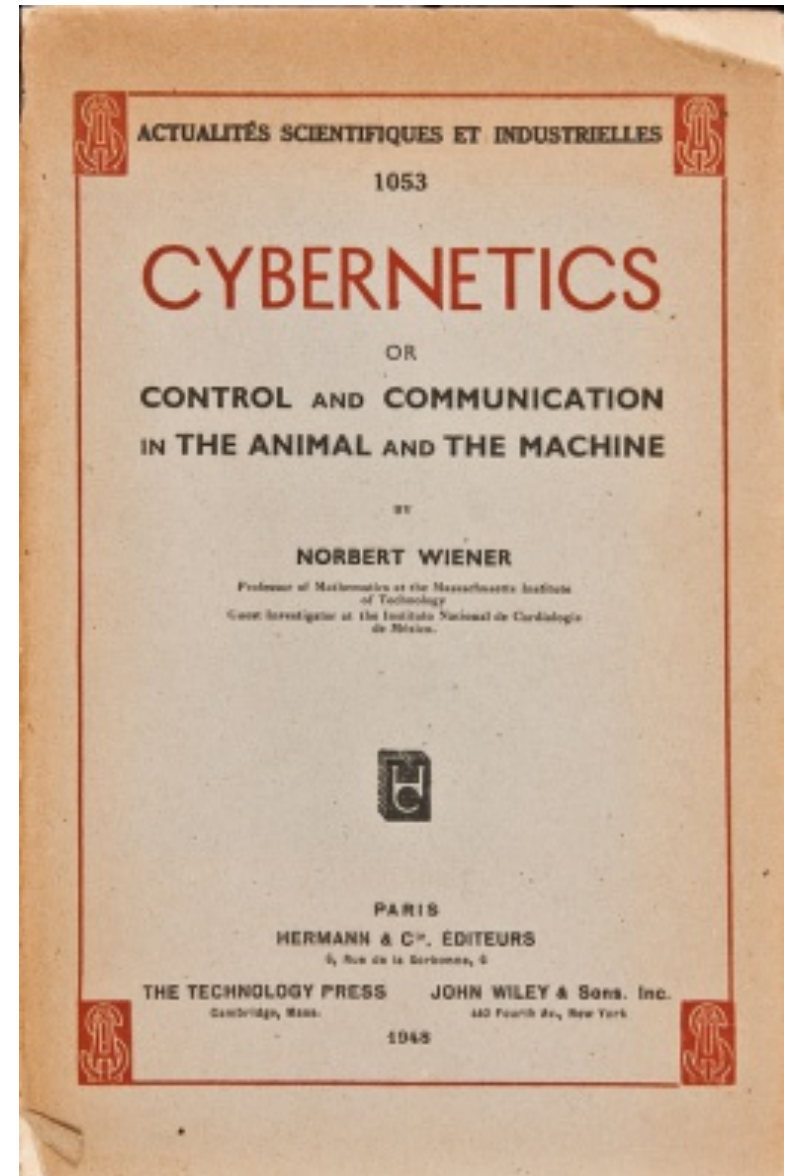
Grey Walter is a neurophysiologist and roboticist. Very simple hardware can have complex behavior as soon as you add feedbacks (from the environment)

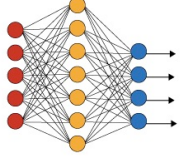


# 1948: NORBERT WIENER



A Cybernetic Loop

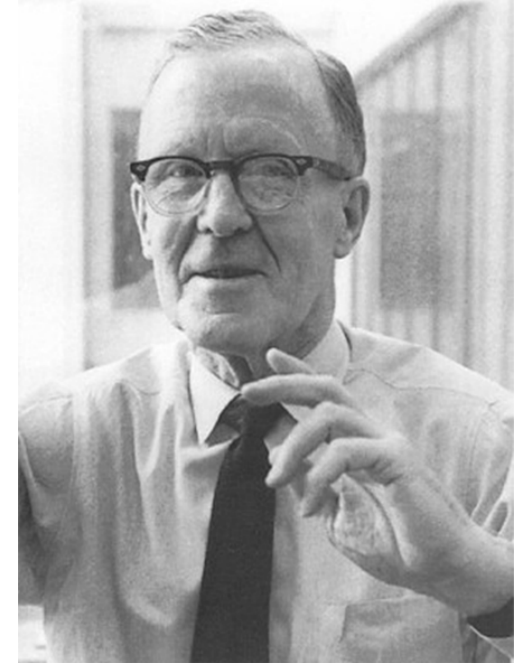




# 1949: DONALD HEBB

Hebb's rule or Hebbian theory: an explanation for the adaptation of neurons in the brain during the learning process

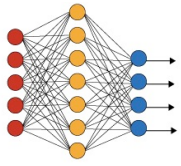
**Basic mechanism for synaptic plasticity:** an increase in synaptic efficacy arises from the presynaptic cell's repeated and persistent stimulation of the postsynaptic cell.



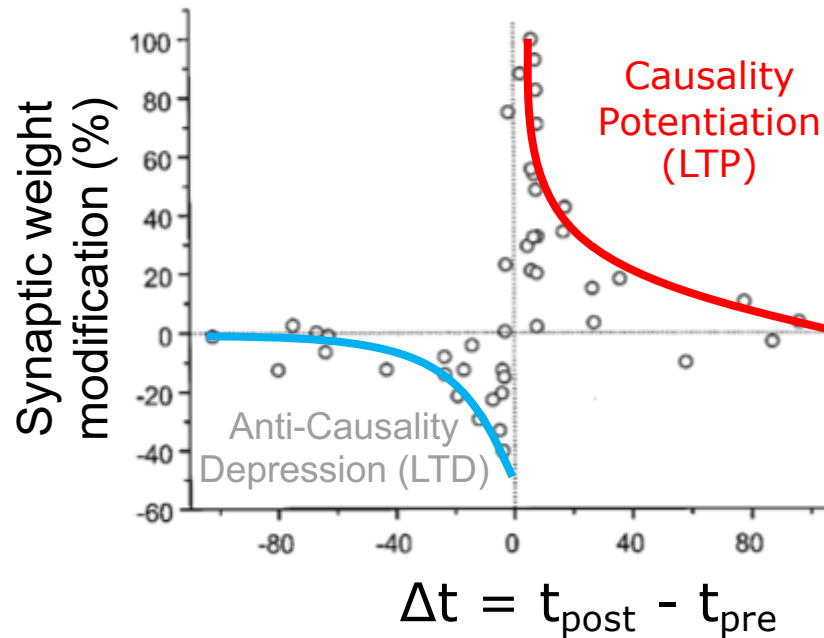
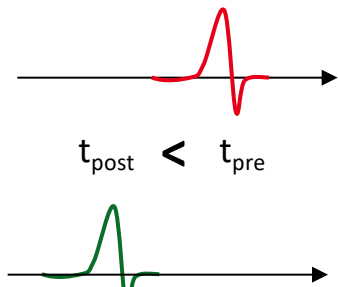
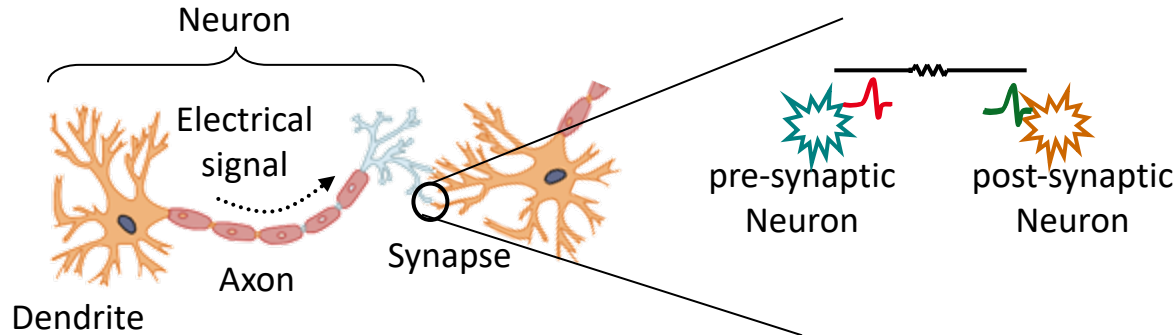
Psychologist, working in the area of neuropsychology

Introduced by Donald Hebb in his 1949 book « *The Organization of Behavior* »

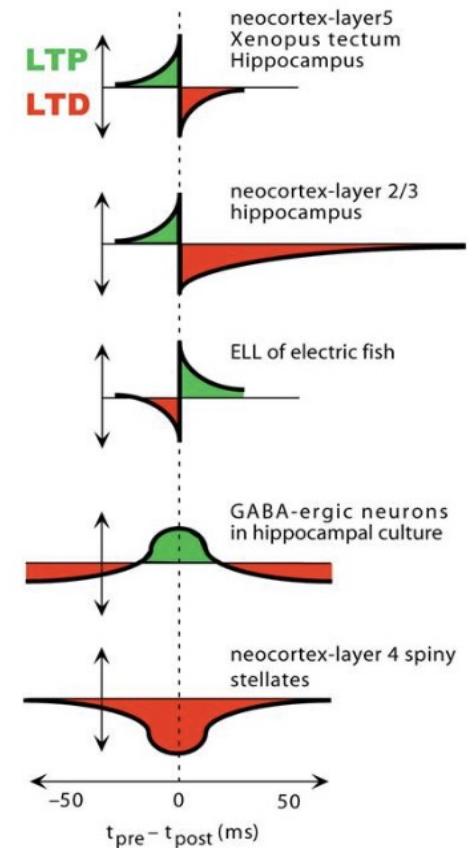


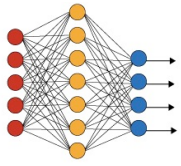


# DERIVED FROM HEBB'S RULE: STDP (SPIKE TIMING DEPENDENT PLASTICITY)



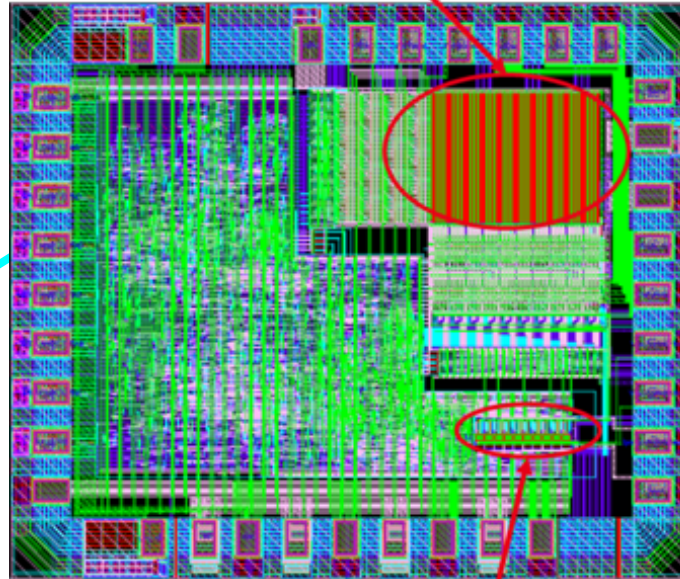
STDP = correlation detector





# SIDE REMARK: INVESTIGATION OF RRAM AS SYNAPSES UNSUPERVISED LEARNING (INFORMATION CODED BY SPIKES)

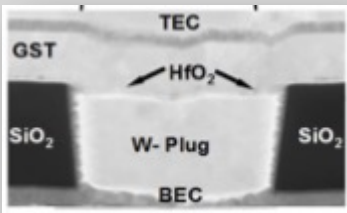
OxRAMs



*Thermal  
effect*

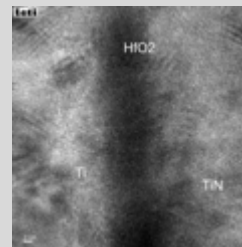
**PCM**

GST  
GeTe  
GST + HfO<sub>2</sub>



*Electronic effect  
oxygen vacancies*

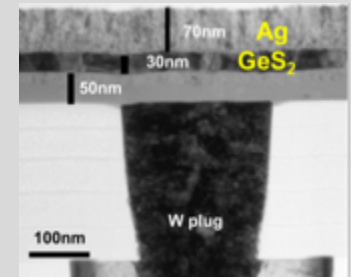
**OxRAM**  
TiN/HfO<sub>2</sub>/Ti/TiN



*Electrochemical  
effect*

**CBRAM**

Ag / GeS<sub>2</sub>

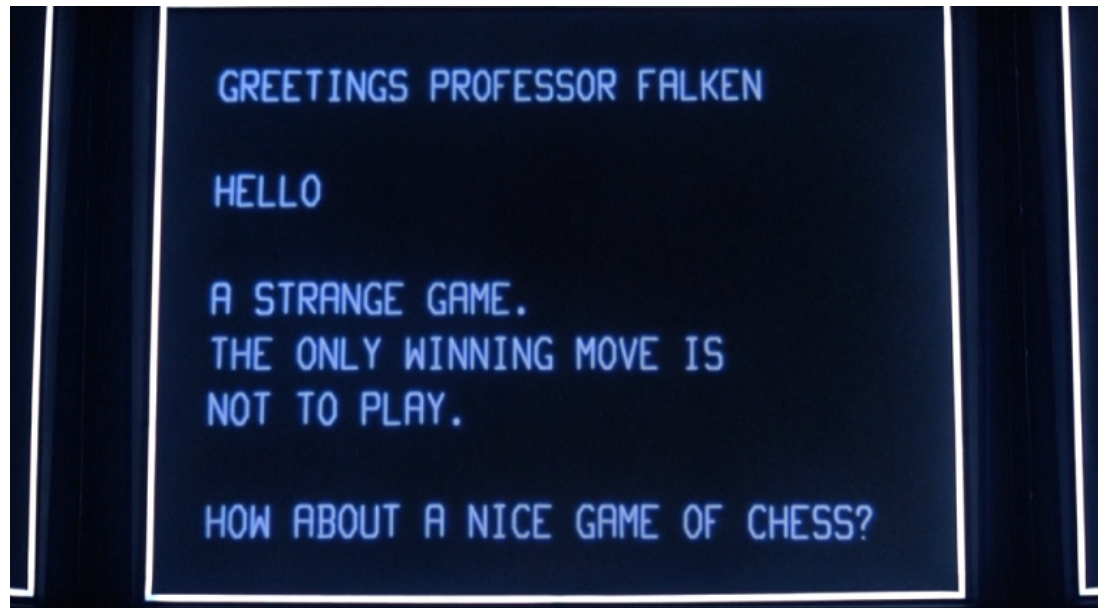
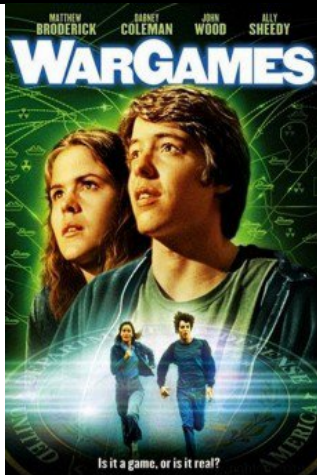
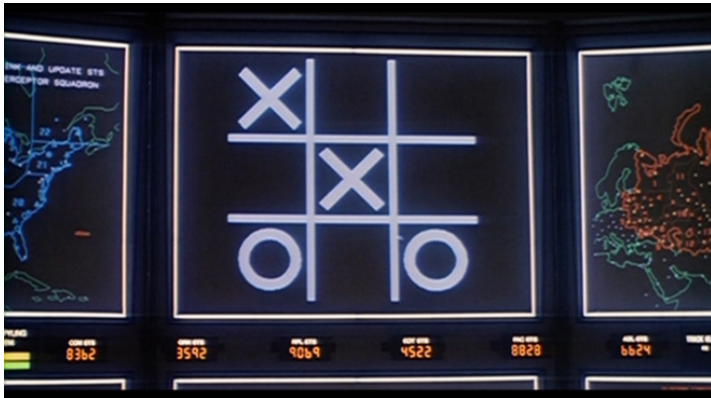


M.Suri, et. al, IEDM 2011  
M.Suri, et. al, IMW 2012 , JAP 2012  
O.Bichler et al. IEEE TED 2012  
M.Suri et al., EPCOS 2013  
D.Garbin et al., IEEE Nano 2013

D.Garbin et al. IEDM 2014  
D.Garbin et al., IEEE TED 2015

# 1952: TIC TAC TOE

In 1952, OXO (or Noughts and Crosses), developed by British computer scientist Alexander S. Douglas for the EDSAC computer at the University of Cambridge, became one of the **first known** video games.



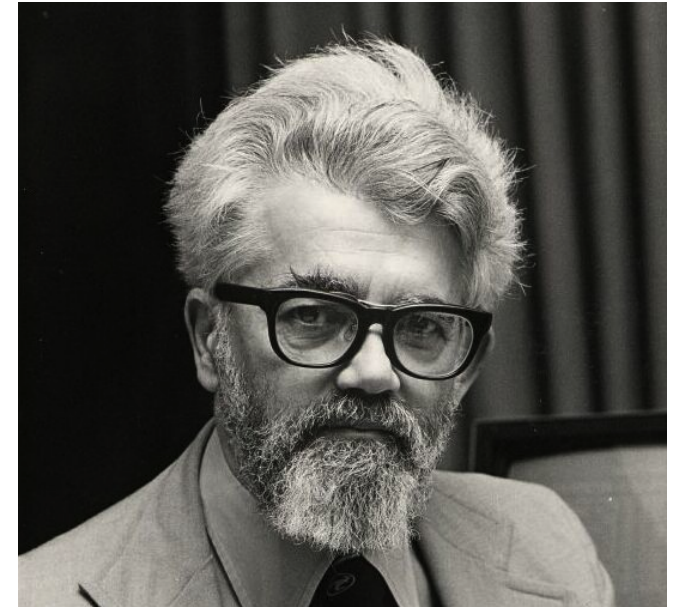
Wargame movie, 1983

# 1955: JOHN MCCARTHY

John McCarthy is one of the "founding fathers" of artificial intelligence, together with Marvin Minsky, Allen Newell and Herbert A. Simon.

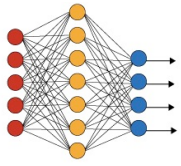
McCarthy coined the term "artificial intelligence" in 1955, and organized the famous **Dartmouth Conference** in Summer 1956. This conference started AI as a science field.

While at MIT, McCarthy developed the programming language **LISP** in 1950, one of the two oldest programming language



```
(defun factorial (n)
  (if (= n 0)
      1
      (* n (factorial (- n 1))) ) )
```

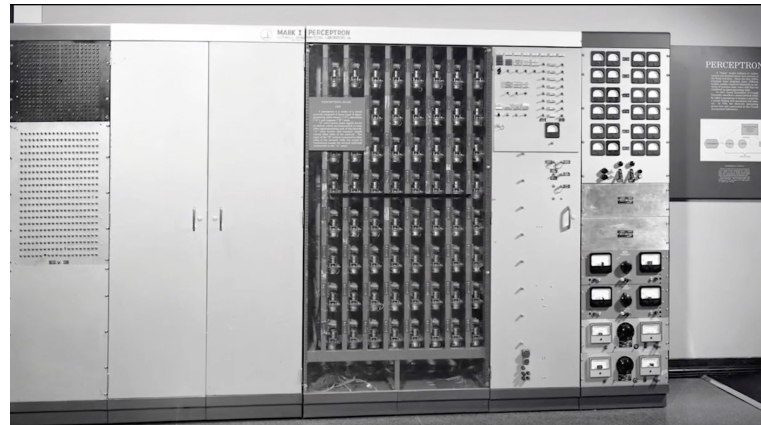
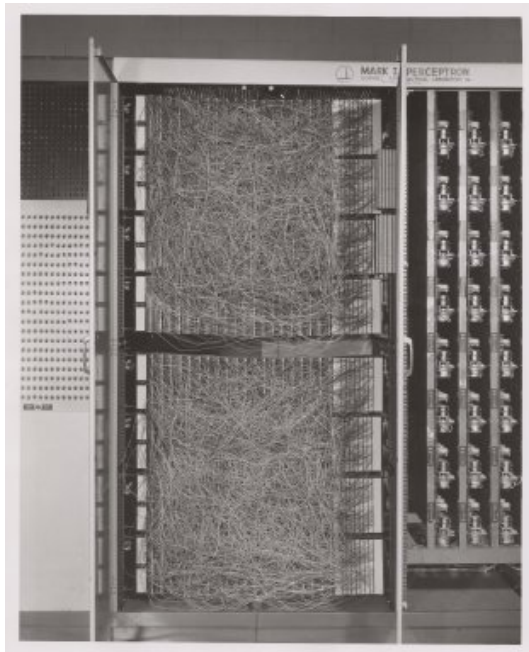




# 1957: THE PERCEPTRON AND F. ROSENBLATT

The perceptron algorithm was invented in 1957 at the Cornell Aeronautical Laboratory by Frank Rosenblatt.

The perceptron was intended to be a machine, rather than a program, and while its first implementation was in software for the IBM 704, it was subsequently implemented in custom-built hardware as the "Mark 1 perceptron". This machine was designed for image recognition: it had an array of 400 photocells, randomly connected to the "neurons". Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors.





# The Perceptron Learning Algorithm

Y LeCun

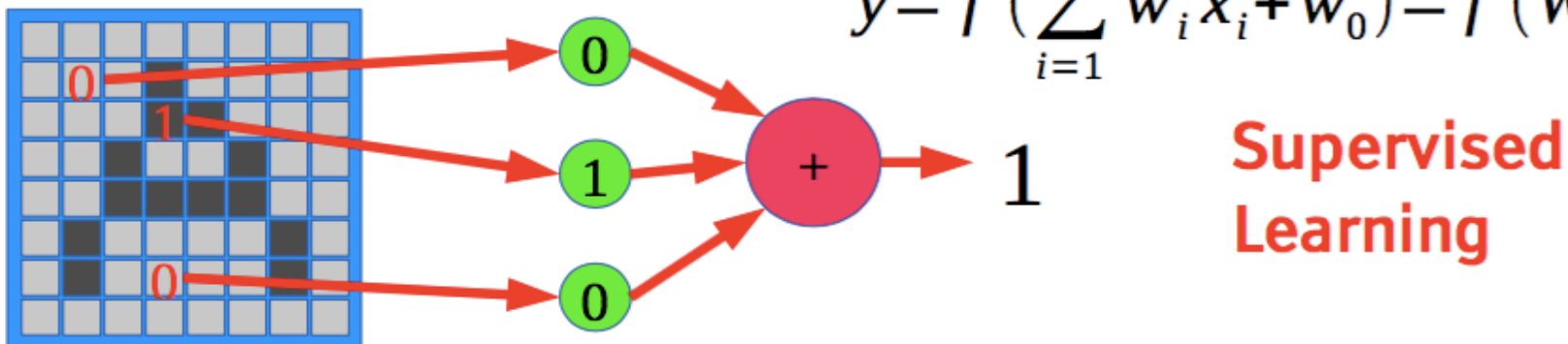
<https://www.college-de-france.fr/site/yann-lecun/course-2015-2016.htm>

- **Training set:**  $(X^1, Y^1), (X^2, Y^2), \dots, (X^P, Y^P)$
- **Take one sample  $(X^k, Y^k)$ , if the desired output is +1 but the actual output is -1**
  - ▶ Increase the weights whose input is positive
  - ▶ Decrease the weights whose input is negative
- **If the desired is -1 and actual is +1, do the converse.**
- **If desired and actual are equal, do nothing**

$$w_i(t+1) = w_i(t) + (y_i^p - f(W'X^p))x_i^p$$

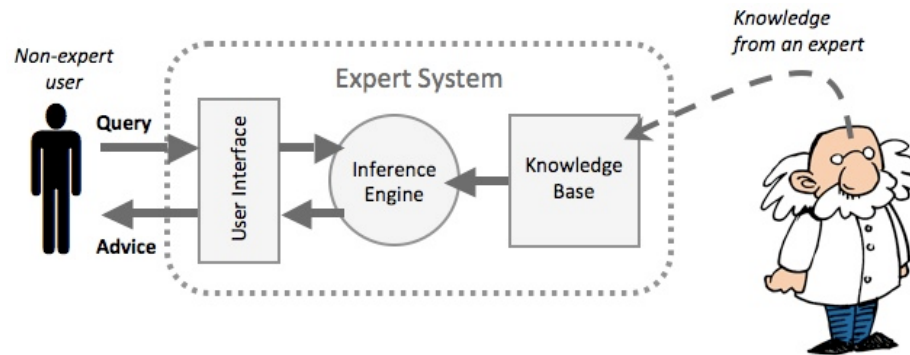
1986: David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams

$$y = f\left(\sum_{i=1} w_i x_i + w_0\right) = f(W'X)$$



# 1965: EXPERT SYSTEMS

Expert systems were introduced by the Stanford Heuristic Programming Project led by Edward Feigenbaum,  
Can also use predicate logic or even Fuzzy Logic



## A) Rulebase Example

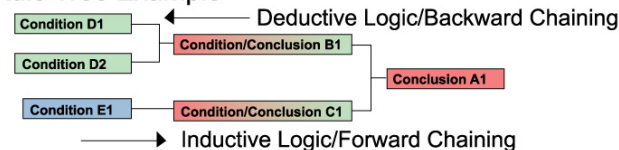
**IF**

- "yes" is equal to `uniform_layer_flow`
- `THETA` is greater than 45.0
- `THETA` is less than or equal to 90.0
- `C4` is greater than  $(Lm / (0.8 * (Hs - H0)))$
- `C6` is greater than  $(Lb / (0.8 * (Hs - H0)))$
- `C9` is less than or equal to  $(Lt / (0.8 * (Hs - H0)))$

**Then**

- `flow_type_ok` is confirmed
- "V2" is assigned to `flow_type`
- "No" is assigned to `wake_attachment`
- `Find coanda_attachment_value`

## B) Rule Tree Example



A Lisp machine

Decoupling:

- The (inference) engine
- The knowledge base

# 1966: ELIZA THE CHAT BOT !

```
;;;
(PRO *****
      ELIZA
ORIGINAL PROGRAM DESCRIBED BY
      JOSEPH WEIZENBAUM

(DEF IN THE COMMUNICATION OF THE ACM JANUARY 1966

(DOC BE SURE THAT THE CAPS LOCK IS ON
(L PLEASE DON'T USE COMMAS OR PERIODS IN YOUR INPUTS

;; *****
;;
;;
;;
;;

*"))

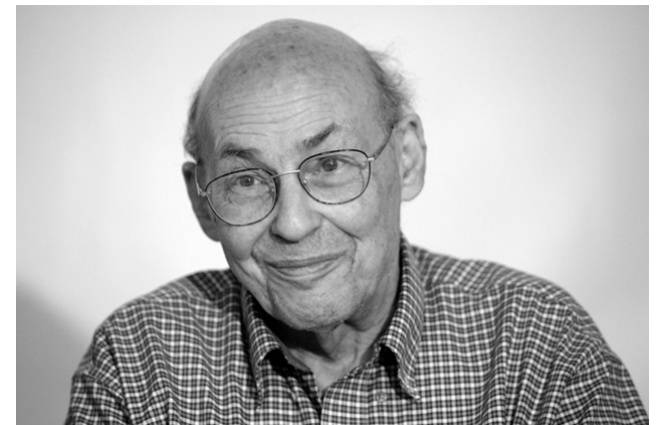
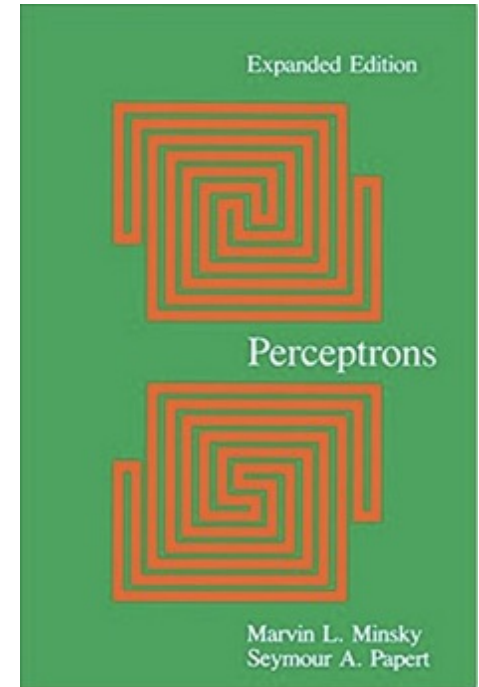
HI! I'M ELIZA. WHAT'S YOUR PROBLEM?
? I AM TIRED
DID YOU COME TO ME BECAUSE YOU ARE TIRED
? NO, I \ I ,ON\
SAY, DO YOU HAVE ANY PSYCHOLOGICAL PROBLEMS?
? NO OF COURSE
ARE YOU SAYING NO JUST TO BE NEGATIVE?
? YES
ARE YOU SURE?
? █
```

# 1969: MARVIN MINSKY

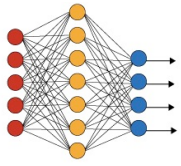
He developed, with Seymour Papert, the first Logo "turtle".

Minsky also built, in 1951, the first randomly wired neural network learning machine, SNARC.

Minsky wrote the book **Perceptrons** (with Seymour Papert), which became the foundational work in the analysis of artificial neural networks. This book is the center of a controversy in the history of AI, as some claim it to have had great importance in discouraging research of neural networks in the 1970s, and contributing to the so-called "**First AI winter**".



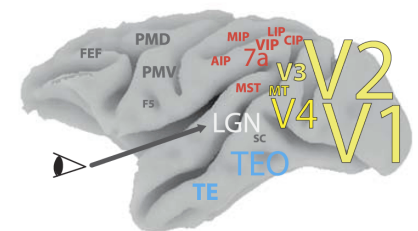
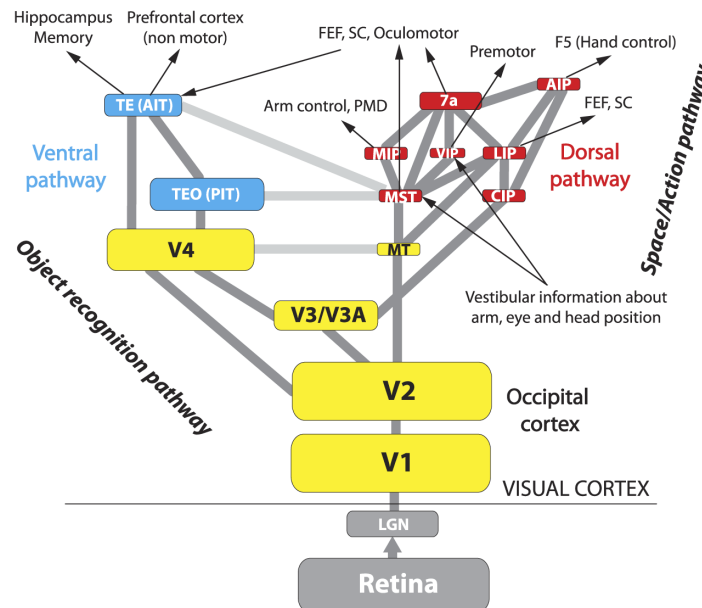
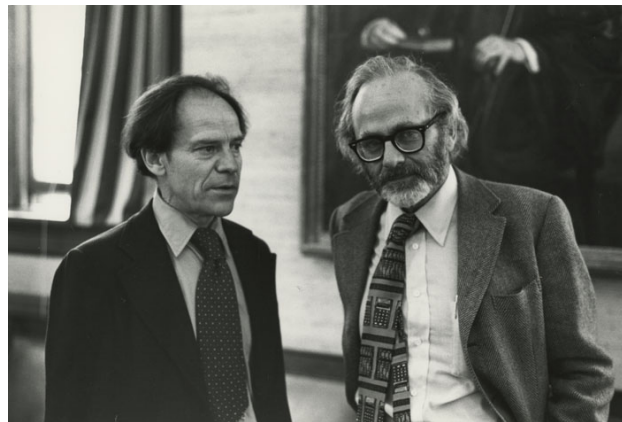




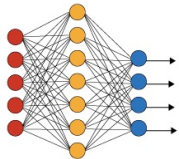
# 1981: DAVID MARR, DAVID HUBEL ET TORSTEN WIESEL

Better understanding how the biological visual system works:

- David Marr: Vision: A computational investigation into the human representation and processing of visual information, which was finished mainly on 1979 summer, was published in 1982 after his death
- Hubel and Wiesel were awarded the Nobel Prize in 1981 for their work on ocular dominance columns in the 1960s and 1970s.







# 1980: KUNIHICO FUKUSHIMA

The first Deep Neural Network, inspired by the visual cortex.



## Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan

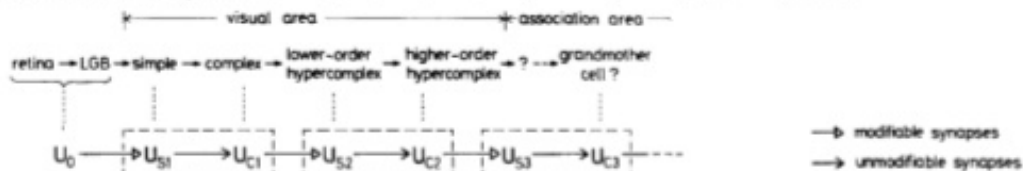


Fig. 1. Correspondence between the hierarchy model by Hubel and Wiesel, and the neural network of the neocognitron

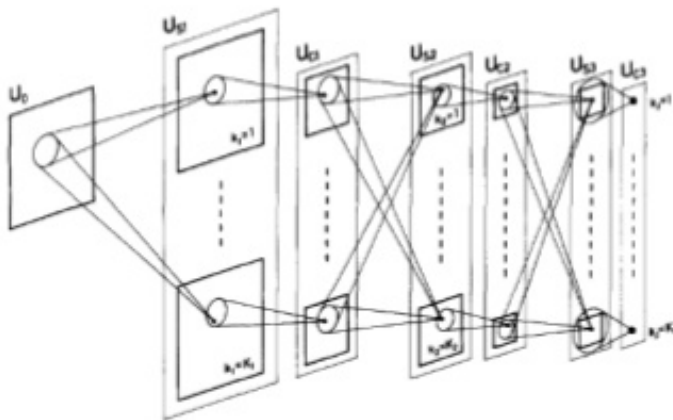
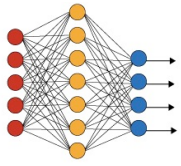


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Biol. Cybernetics 36, 193–202 (1980)



## AROUND 1986: GEOFFREY HINTON

He was one of the first researchers who demonstrated the use of **generalized back-propagation algorithm** for training multi-layer neural networks.

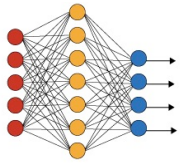
He co-invented **Boltzmann machines** with David Ackley and Terry Sejnowski.

His other contributions to neural network research include distributed representations, time delay neural network, mixtures of experts, Helmholtz machines and Product of Experts

He is now working for Google.



Cognitive psychologist and computer scientist



## AROUND 1985: YANN LE CUN

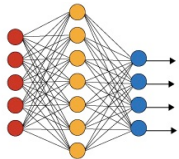
In 1985, he proposed and published (in French), an early version of the learning algorithm known as **error backpropagation**

Near 1989, he developed a number of new machine learning methods, such as a biologically inspired model of image recognition called **Convolutional Neural Networks**, the "Optimal Brain Damage" regularization methods, and the Graph Transformer Networks method which he applied to handwriting recognition and OCR.



The **bank check recognition system** that he helped develop was widely deployed by NCR and other companies, reading over 10% of all the checks in the US in the late 1990s and early 2000s.

In 2013, LeCun became the first director of Facebook AI Research in New York City.



# 1987: NETTALK. A 3 LAYERS PERCEPTION LEARN TO READ

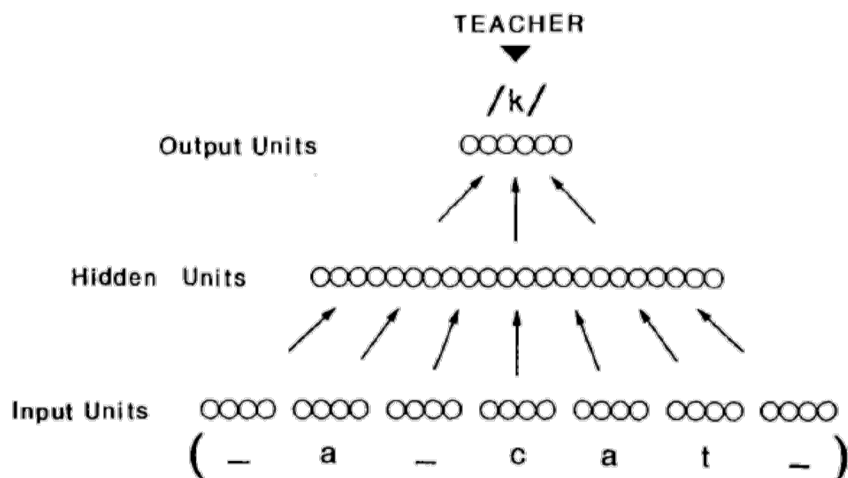
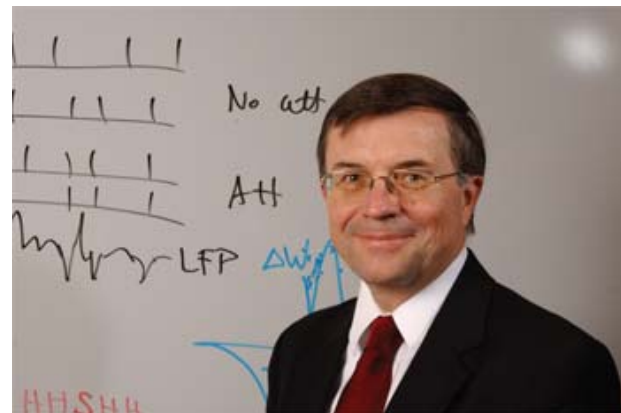
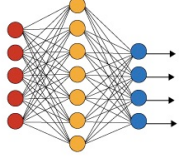


Figure 1: Schematic drawing of the NETtalk network architecture. A window of letters in an English text is fed to an array of 203 input units. Information from these units is transformed by an intermediate layer of 80 “hidden” units to produce patterns of activity in 26 output units. The connections in the network are specified by a total of 18629 weight parameters (including a variable threshold for each unit).

From T. J. Sejnowski and C. R. Rosenberg, “Parallel networks that learn to pronounce English text,” *Complex Systems*, vol. 1, no. 1, pp. 145–168, 1987.

Slide from Christian Gamrat





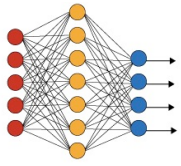
# 1990'S NEUROCOMPUTERS...



## Siemens : MA-16 Chips (SYNAPSE-1 Machine 1994)

- Synapse-1, neurocomputer with 8xMA-16 chips
- Synapse3-PC, PCI board with 2xMA-16 (1.28 Gpcs)
- about 8,000 times as fast as a Sun Workstation (Sparc-2)

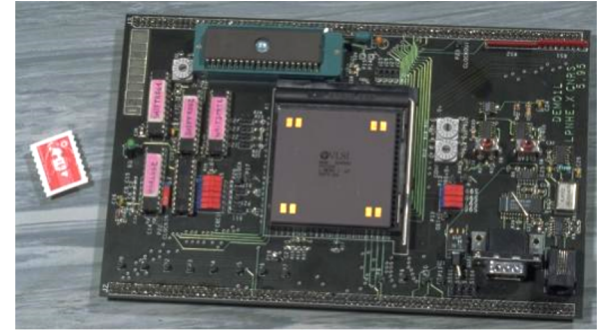




# 1990'S NEUROCOMPUTERS...

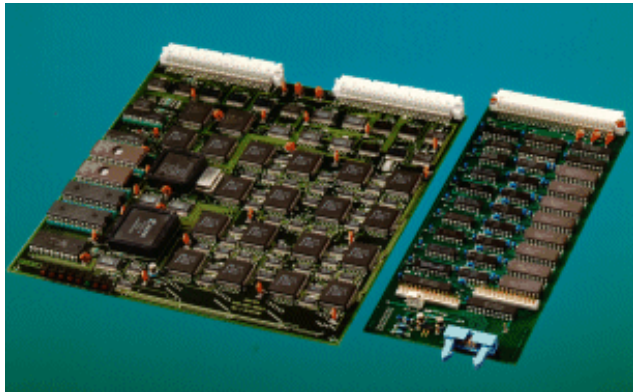
## Philips : L-Neuro

- 1st Gen 16 PEs 26 MCps (1990)
- 2nd Gen 12 PEs 720 MCps (1994)
- Used in satellite, fruit sorting, PCB inspection, sleep analysis, ...

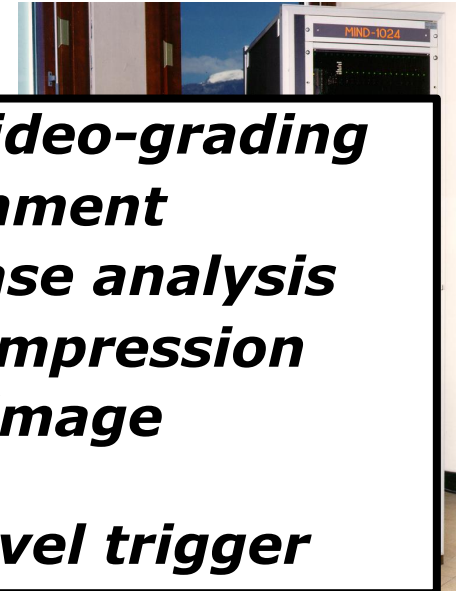


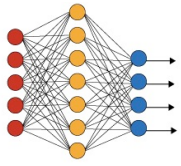
## CEA's MIND machine

- Hybrid analog/digital: MIND-128 (1988)
- Fully digital: MIND-1024 (1991)



- ☐ **Orange video-grading**
- ☐ **Chip alignment**
- ☐ **Sleep phase analysis**
- ☐ **Image compression**
- ☐ **Satellite image analysis**
- ☐ **LHC 1<sup>st</sup> level trigger**

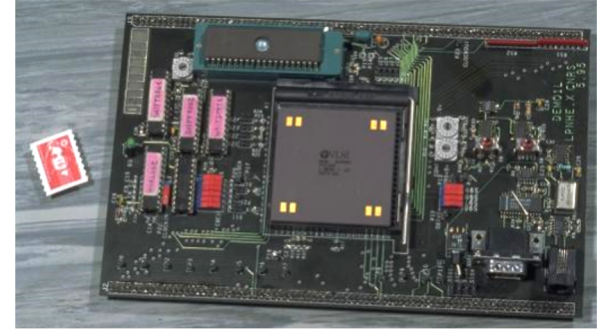




# 1990'S NEUROCOMPUTERS...

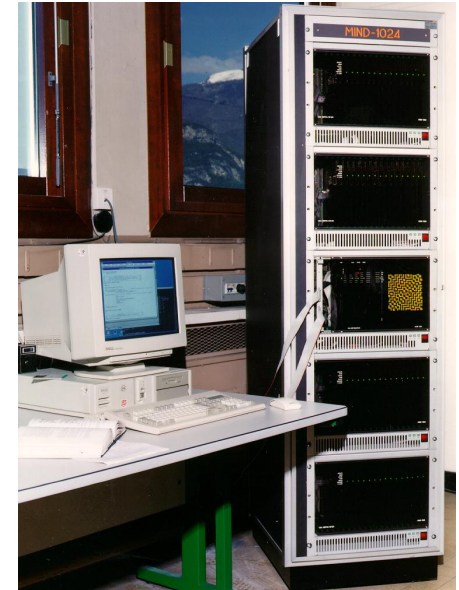
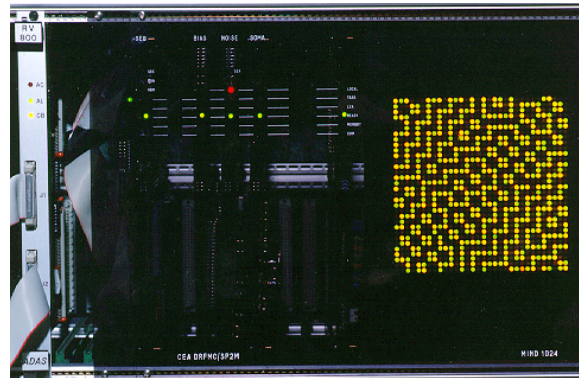
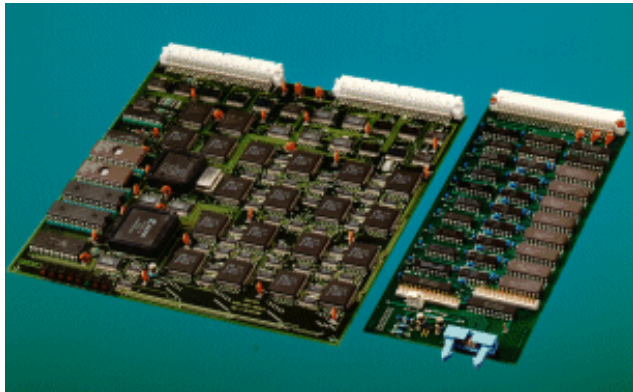
## Philips : L-Neuro

- 1st Gen 16 PEs 26 MCps (1990)
- 2nd Gen 12 PEs 720 MCps (1994)
- Used in satellite, fruit sorting, PCB inspection, sleep analysis, ...



## CEA's MIND machine

- Hybrid analog/digital: MIND-128 (1986)
- Fully digital: MIND-1024 (1991)

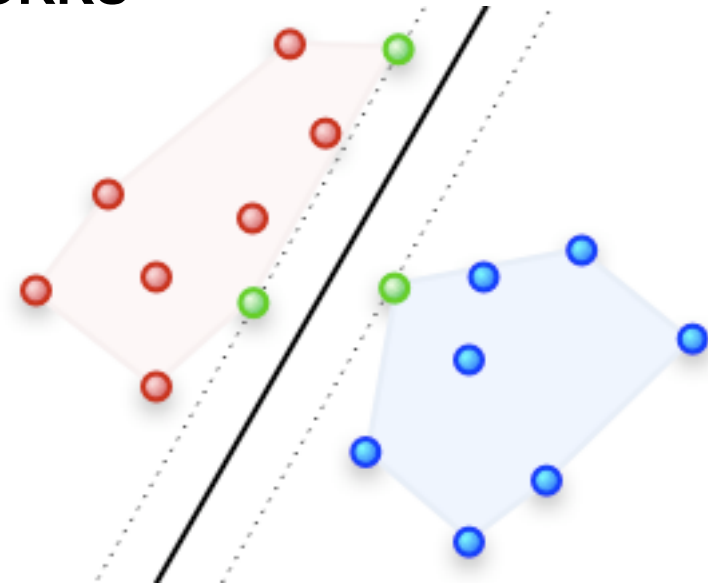


## 1995: SVM OR THE 2<sup>ND</sup> WINTER OF NEURAL NETWORKS

### Support Vector Machines (SVMs)

The original SVM algorithm was invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963.

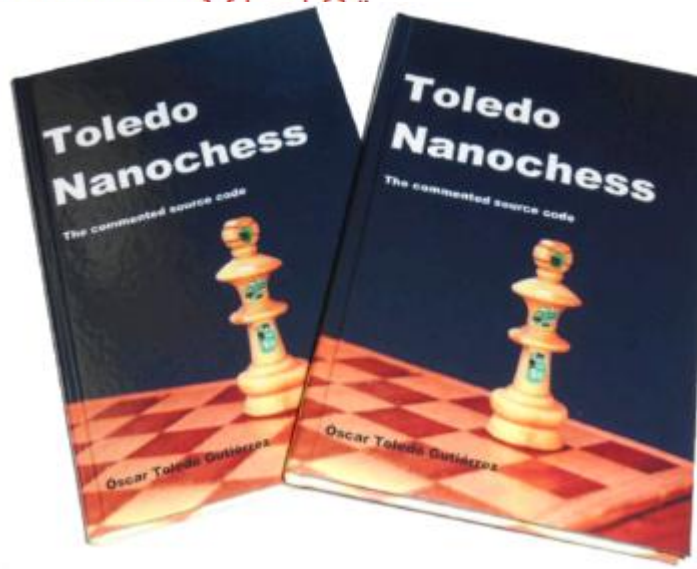
In 1992, Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes. The current standard incarnation (soft margin) was proposed by Corinna Cortes and Vapnik in 1993 and published in 1995.





# WHAT IS THIS PROGRAM DOING?

```
char*l="ustvrtuqqqc
" 76Lsabcddcba .pk
#define F getchar()8
#define v X(0,0,0,21
#define Z while(
#define _ ;if(
#define P return--G,
B,i,y,u,b,I[411],*G=
=78-h<<x,p,*g,n,*m,A
p=0)){q=o&z^y _ q<7)
==w){g=q|p+a-S?0:I+S
K;J=n=o&z;E=I[p-a]&z
+=(1-q?1[p/x+5]-1[0/
!(I[p+1]^n)+1[n&7]*5
```



```
L,E,d,O=e,N=-M*M,K
v 0,0)>M;do{ _ o=I[
I[p+=C[1]-64]_!w|p
!(r-2&7))P G[1]=0,
9-189-h-q:0 _ s)L
+(q?0:!(I[p-1]^n)+
h&&L>z|d){p[I]=n,O
```

A book of 170 pages was written to explain how it works.  
the Toledo Nanochess source code is extremely complicated.  
For each source line there are multiple reasons for some coding decisions.

```
x-(B=F)){1=I[B+=(X-F)*X]&z;b=F;b+=(X-F)*X;Z x-(*G=F))1=*G^8^y;};else v u,5);v u,
1);}}
```

```
rn bqkbnr
pppppppp
.....
.....
.....
.....
PPPPPPPP
RNBQKBNR D2D4

rn bqkbnr
pppppppp
.....
.....
.....
.....
P.....
PPP.PPPP
RNBQKBNR

rn bqkbnr
ppp.pppp
.....
.....
p.....
P.....
PPP.PPPP
RNBQKBNR
```



# 1997: CHESS AND DEEP BLUE

As far back as the mid-60s, chess was called the "Drosophila of artificial intelligence" – a reference to the fruit flies biologists used to uncover the secrets of genetics –  
1997 – Deep Blue wins a six-game match against Garry Kasparov.



# HISTORY OF COMPUTING AND CHESS

1948 – **Norbert Wiener**'s book *Cybernetics* describes how a chess program could be developed using a depth-limited minimax search with an evaluation function.

1950 – **Claude Shannon** publishes "Programming a Computer for Playing Chess", one of the first papers on the problem of computer chess.

1951 – **Alan Turing** is first to publish a program, developed on paper, that was capable of playing a full game of chess.

1956 – **John McCarthy** invents the alpha-beta search algorithm.

1962 – The first program to play credibly, *Kotok-McCarthy*, is published at MIT

1967 – *Mac Hack Six*, by Richard Greenblatt et al. introduces transposition tables and becomes the first program to defeat a person in tournament play

1968 – David Levy makes a bet with AI researchers that no computer program would win a chess match against him within 10 years.

1977 – The first microcomputer chess playing machines, *CHESS CHALLENGER* and *BORIS*, were created

1978 – David Levy wins the bet made 10 years earlier, defeating *Chess 4.7* in a six-game match by a score of 4½–1½. The computer's victory in game four is the first defeat of a human master in a tournament.[11]

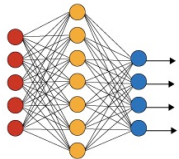
1994 – First time a chess program (*ChessGenius*) defeated a World Champion (*Garry Kasparov*) at a non blitz time limit.

1997 – **Deep Blue** wins a six-game match against *Garry Kasparov*.

2006 – The undisputed world champion, *Vladimir Kramnik*, is defeated 4–2 by *Deep Fritz*.

2015 – *Super Micro* is now the smallest computer implementation of chess at a size of only 455 bytes

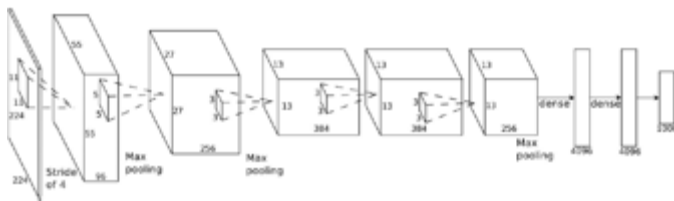
2017 – **AlphaZero** learn to play superhuman chess in less than 4 hours



# 2012: DEEP NEURAL NETWORKS RISE AGAIN

They give the *state-of-the-art performance* e.g. in image classification

- **ImageNet classification (Hinton's team, hired by Google)**
  - 14,197,122 images, 1,000 different classes
  - Top-5 17% error rate (huge improvement) in 2012 (now ~ 3.5%)



**"Supervision"** network

Year: 2012

650,000 neurons

60,000,000 parameters

630,000,000 synapses

- **Facebook's 'DeepFace' Program (labs headed by Y. LeCun)**

The 2018 Turing Award recipients are Google VP Geoffrey Hinton, Facebook's Yann LeCun and Yoshua Bengio, Scientific Director of AI research center Mila.

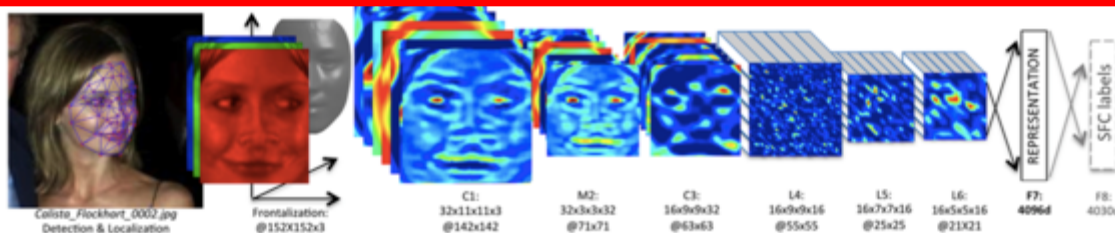


Figure 2. Outline of the *DeepFace* architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

From: Y. Taigman, M. Yang, M.A. Ranzato, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification"

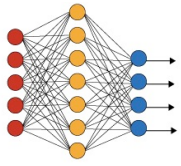
# ImageNet: Classification

Y LeCun

- Give the name of the dominant object in the image
- Top-5 error rates: if correct class is not in top 5, count as error
- Black: ConvNet, Purple: no ConvNet

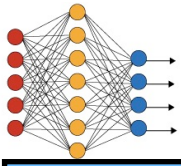
2012 Teams	%error	2013 Teams	%error	2014 Teams	%error
Supervision (Toronto)	15.3	Clarifai (NYU spinoff)	11.7	GoogLeNet	6.6
ISI (Tokyo)	26.1	NUS (singapore)	12.9	VGG (Oxford)	7.3
VGG (Oxford)	26.9	Zeiler-Fergus (NYU)	13.5	MSRA	8.0
XRCE/INRIA	27.0	A. Howard	13.5	A. Howard	8.1
UvA (Amsterdam)	29.6	OverFeat (NYU)	14.1	DeeperVision	9.5
INRIA/LEAR	33.4	UvA (Amsterdam)	14.2	NUS-BST	9.7
		Adobe	15.2	TTIC-ECP	10.2
		VGG (Oxford)	15.2	XYZ	11.2
		VGG (Oxford)	23.0	UvA	12.1





## COMPETITION ON IMAGENET !

Name of the algorithm	Date	Error on test set
Supervision	2012	15.3%
Clarifai	2013	11.7%
GoogLeNet	2014	6.66%
<b>Humain level</b> (Adrej Karpathy)		<b>5%</b>
Microsoft	05/02/2015	4.94%
Google	02/03/2015	4.82%
Baidu/ Deep Image	10/05/2015	4.58%
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences	10/12/2015 (le CNN a 152 couches!)	3.57%
Google Inception-v3 (Arxiv)	2015	3.5%
WMW (Momenta)	2017	2.2%
	Now	?



# EXAMPLES OF RESULTS (IMAGENET)



**sea slug**

sea slug  
flatworm  
coral reef  
sea cucumber  
coral



**brown bear**

brown bear  
otter  
lion  
ice bear  
golden retriever



**jellyfish**

jellyfish  
coral  
polyp  
isopod  
sea anemone



**barracouta**

barracouta  
rainbow trout  
gar  
sturgeon  
coho



**basenji**

basenji  
boxer  
corgi  
Saint Bernard  
Chihuahua



**polyp**

polyp  
sea anemone  
coral  
sea slug  
flatworm



**howler monkey**

howler monkey  
spider monkey  
raccoon  
bullfrog  
indri



**leopard**

leopard  
jaguar  
cheetah  
snow leopard  
Egyptian cat



**American lobster**

American lobster  
tick  
crayfish  
king crab  
barn spider



**mosquito**

mosquito  
harvestman  
cricket  
walking stick  
grasshopper



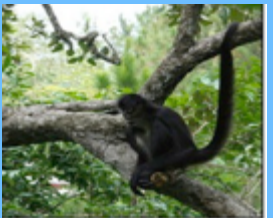
**wolf spider**

wolf spider  
weevil  
grasshopper  
tarantula  
common iguana



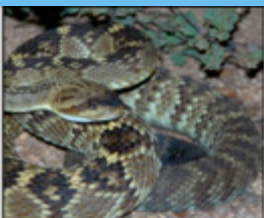
**mite**

mite  
black widow  
cockroach  
tick  
starfish



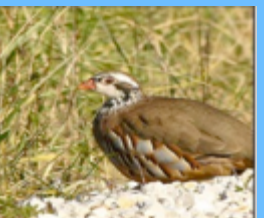
**spider monkey**

howler monkey  
spider monkey  
gorilla  
siamang  
American beech



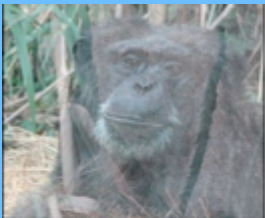
**night snake**

hognoose snake  
night snake  
horned viper  
spiny lobster  
loggerhead



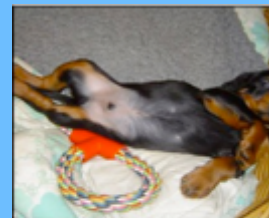
**ruffed grouse**

partridge  
ruffed grouse  
pheasant  
quail  
mink



**chimpanzee**

gorilla  
cougar  
chimpanzee  
baboon  
lion



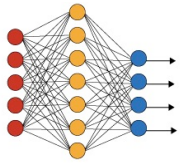
**Gordon setter**

Chihuahua  
Doberman  
basenji  
corgi  
ffordshire bullterrier

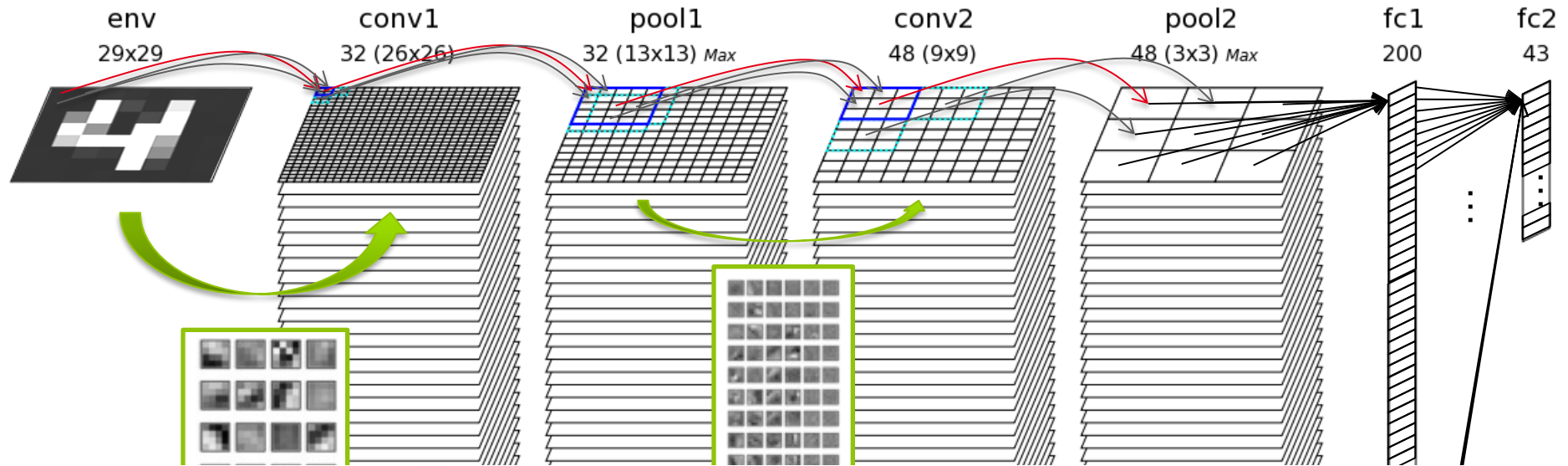


**cherry**

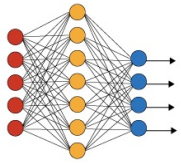
dalmatian  
grape  
elderberry  
ffordshire bullterrier  
currant



# WHAT IS A CNN?

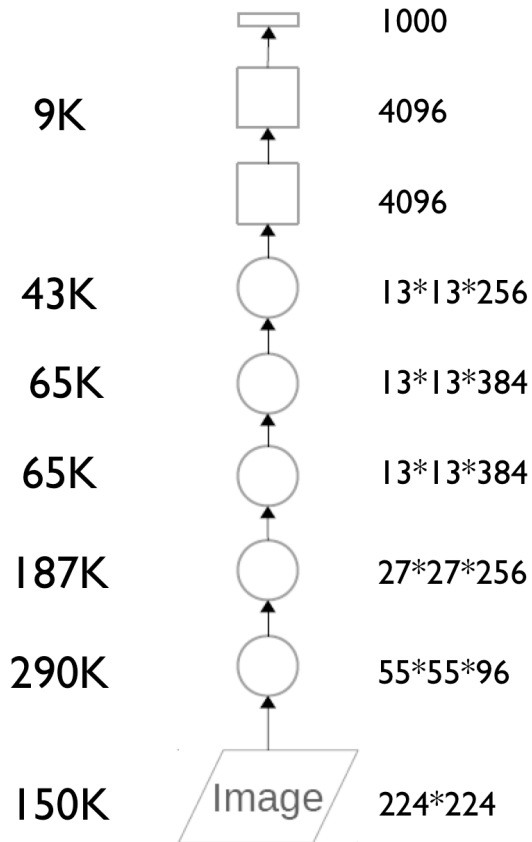


input layer      hidden layer 1      hidden layer 2      hidden layer 3



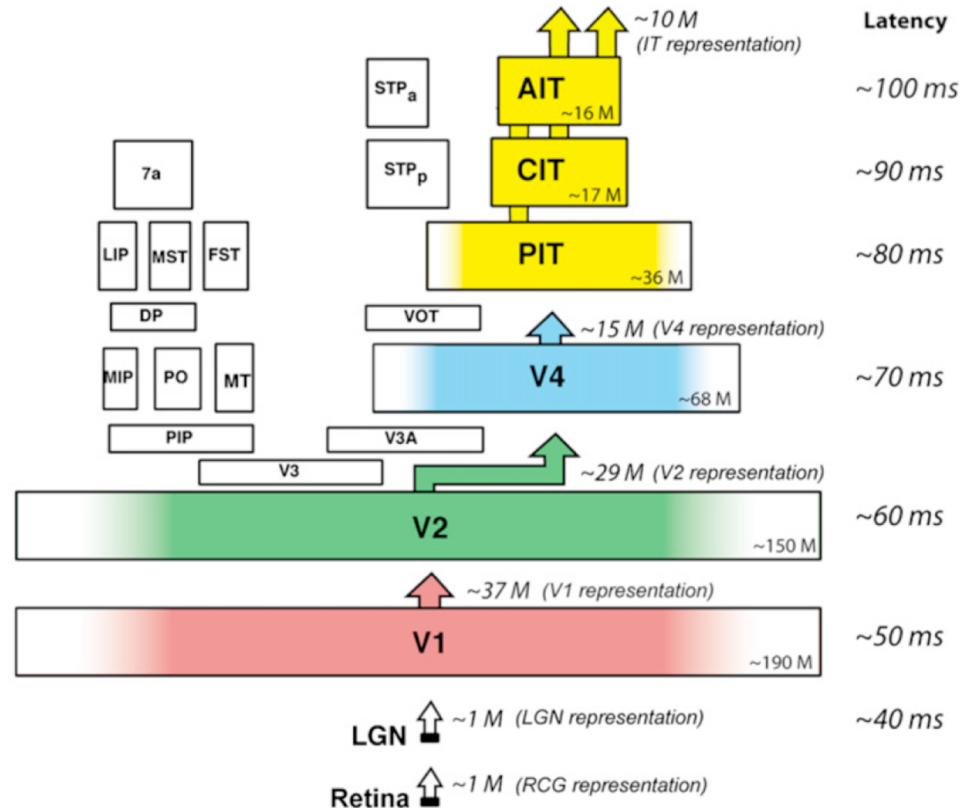
# SUPERVISION VS PRIMATE VISION

## Supervision



Total 650 K neurons

## Primate Visual System



Total 478 M neurons

From Simon Thorpe



# Deep Learning is Everywhere (ConvNets are Everywhere)

## Lots of applications at Facebook, Google, Microsoft, Baidu, Twitter, IBM...

- ▶ Image recognition for photo collection search
- ▶ Image/Video Content filtering: spam, nudity, violence.
- ▶ Search, Newsfeed ranking

## People upload 800 million photos on Facebook every day

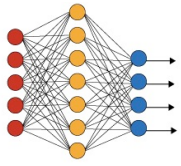
- ▶ (2 billion photos per day if we count Instagram, Messenger and Whatsapp)

## Each photo on Facebook goes through two ConvNets within 2 seconds

- ▶ One for image recognition/tagging
- ▶ One for face recognition (not activated in Europe).

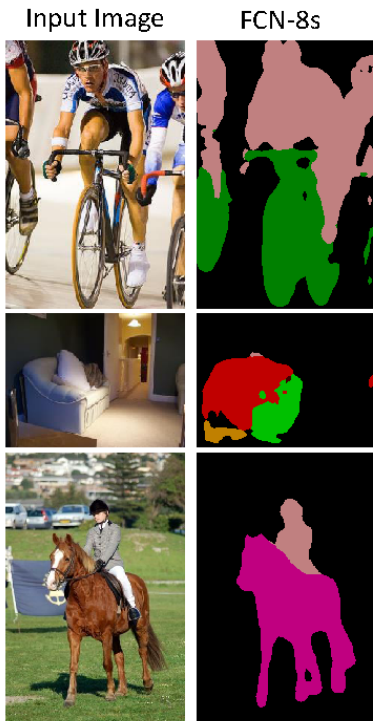
## Soon ConvNets will really be everywhere:

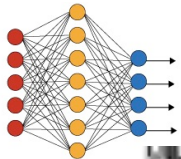
- ▶ self-driving cars, medical imaging, augmented reality, mobile devices, smart cameras, robots, toys.....



# PIXEL WISE IMAGE SEGMENTATION

- DNN technique: Fully-CNN + Unpooling (for high resolution segmentation)





# IMAGE ROI EXTRACTION AND CLASSIFICATION

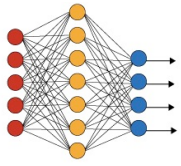
■ D











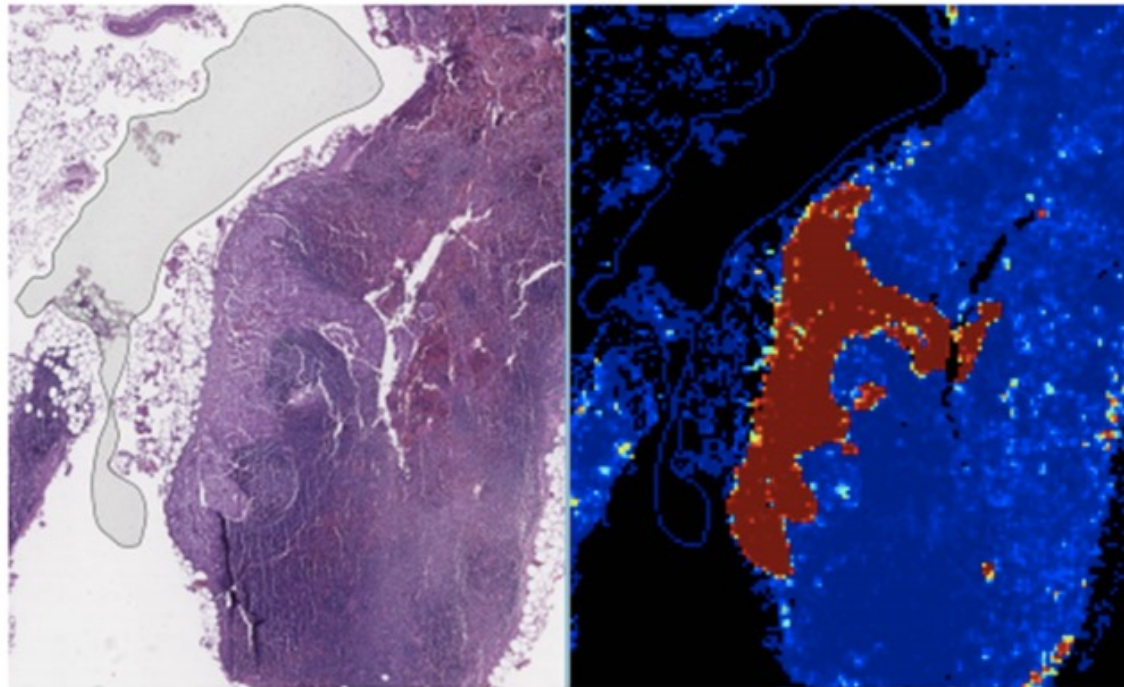
## IMAGE ANALYSIS

# Detecting Cancer Metastases

Tumor localization score  
(FROC):

Pathologist: 0.73

AI model: **0.89**  
**(better)**



*Detecting Cancer  
Metastases on Gigapixel  
Pathology Images (2017)*

# DEEP MANTA

## MANY-TASK DEEP NEURAL NETWORK FOR VISUAL OBJECT RECOGNITION

### Applications

Driving assistance, autonomous driving  
Smart city  
Video-protection  
Advanced Manufacturing



### Technology

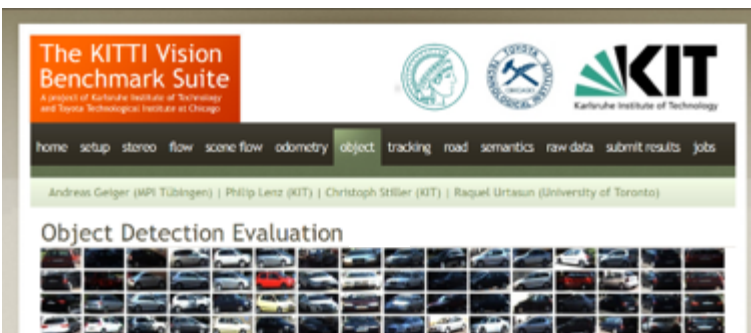
- 1 Object detection
- 2 Fine-grained recognition
- 3 Accurate pose estimation
- 4 2D/3D localisation
- 5 Part localisation
- 6 Part visibility characterization

### Performance

KITTI Benchmark:

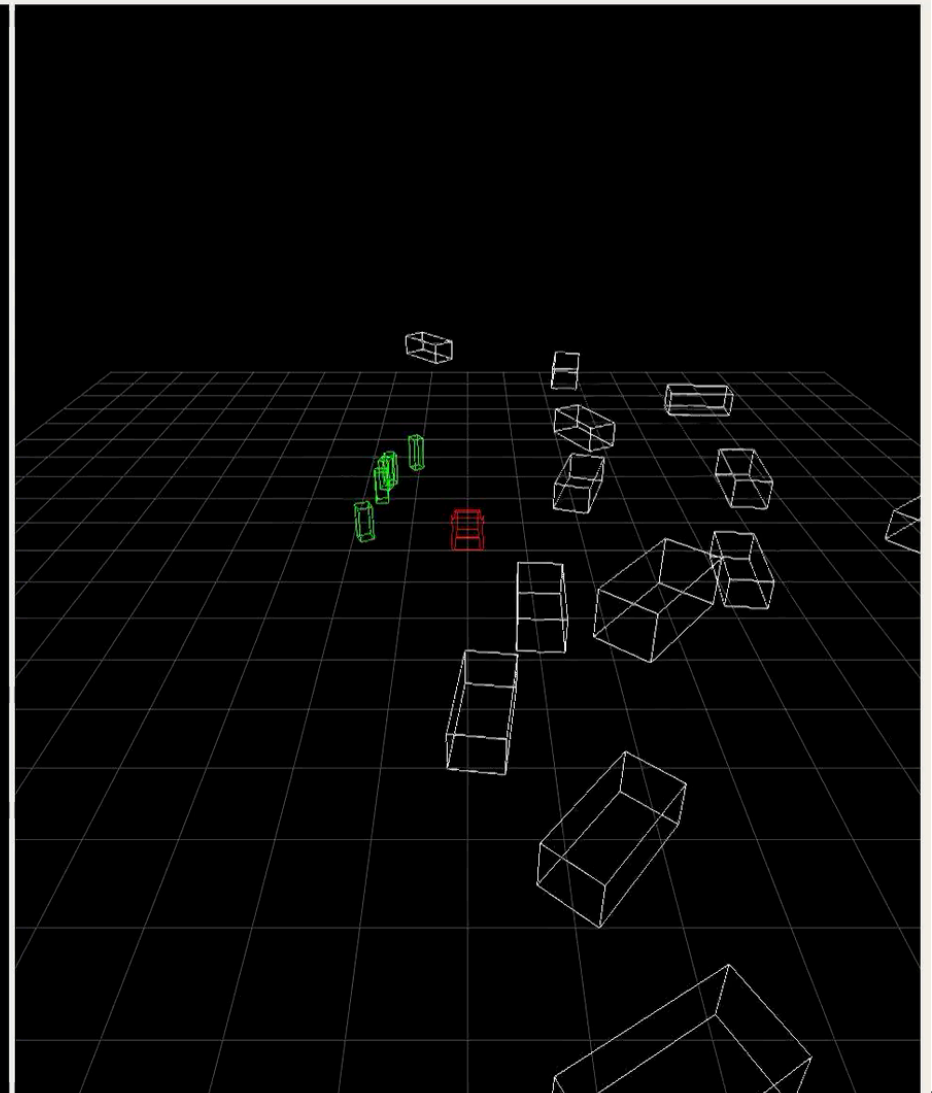
- 1st rank in vehicle orientation estimation
- Top-10 in object detection

Runs at 10 Hz on Nvidia Gtx 1080



CVPR 2017 : F. Chabot, M. Chaouch, J. Rabarisoa, C. Teulière and T. Château  
Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image.

# MANY-TASK DEEP NEURAL NETWORK FOR VISUAL OBJECT RECOGNITION





# Exponential increase of performances in 33 years



Production car of 1985  
Lamborghini Countach 5000QV  
Max speed 300 Km/h



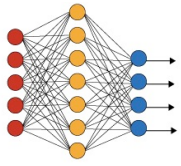
X 100 000 000



27 times the speed of light  
Warp 3 ?  
Star Trek Enterprise  
(Year: about 2290)

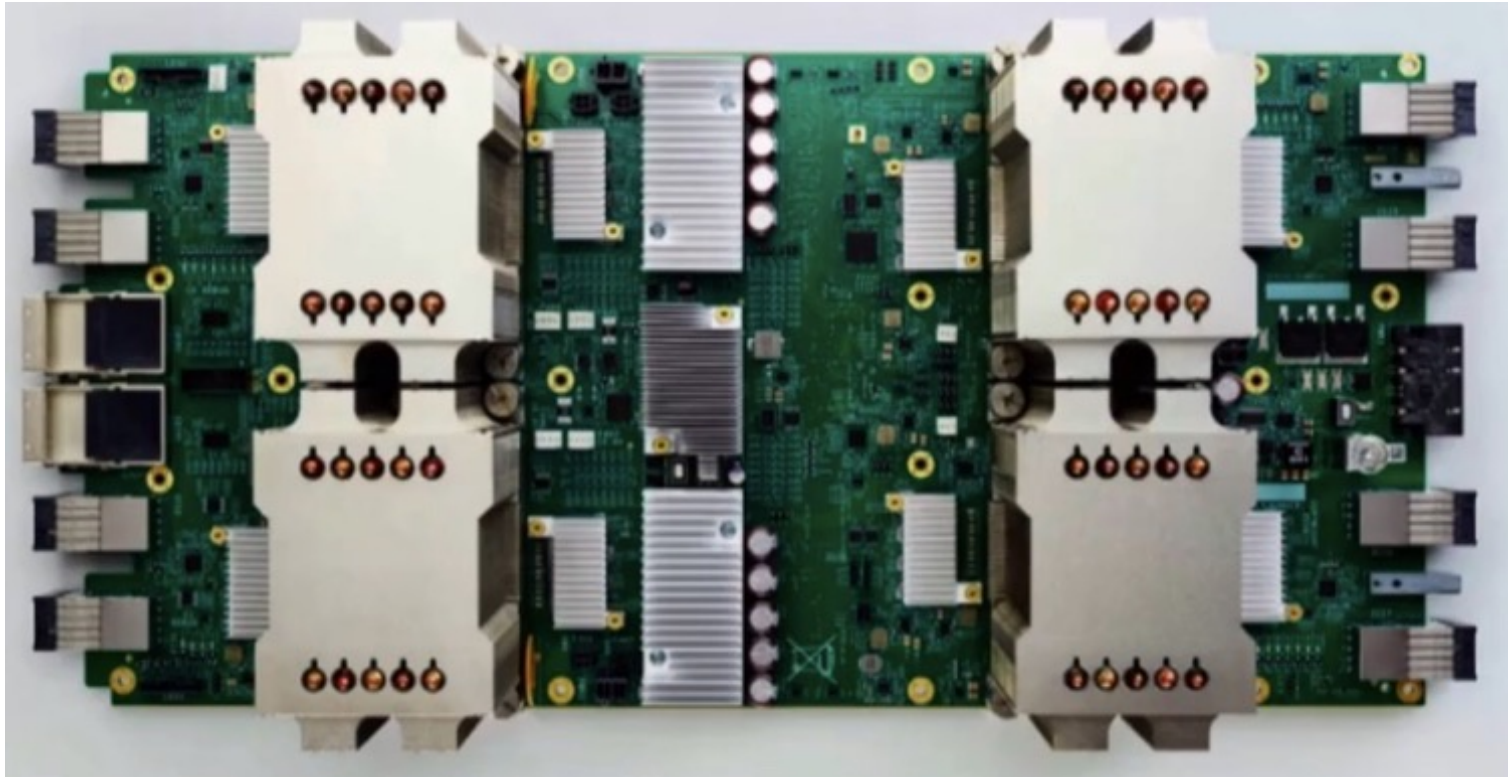
Peta =  $10^{15}$  = million of milliard



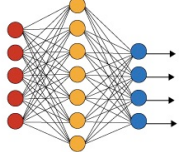


## 2017: GOOGLE'S CUSTOMIZED HARDWARE...

... required to increase energy efficiency  
with **accuracy adapted to the use (e.g. float 16)**



Google's TPU2 : training and inference in a **180 teraflops<sub>16</sub>** board  
(over 200W per TPU2 chip according to the size of the heat sink)

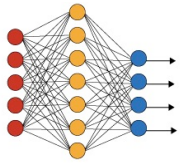


## 2017: GOOGLE'S CUSTOMIZED TPU HARDWARE...

... required to increase energy efficiency  
with accuracy adapted to the use (e.g. float 16)



Google's TPU2 : **11.5 petaflops<sub>16</sub>** of machine learning number crunching  
(and guessing about 400+ KW..., 100+ GFlops<sub>16</sub>/W)

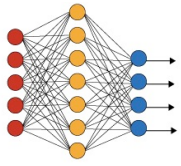


# 2017: THE GAME OF GO

Ke Jie (human world champion in the “Go” game), after being defeated by AlphaGo on May 27th 2017, will work with Deepmind to make a tool from AlphaGo to further help Go players to enhance their game.







# ALPHA ZERO: SELF-PLAYING TO LEARN

The program started from random play given no domain knowledge except the game rules according to an [arXiv paper](#) by DeepMind researchers published Dec. 3.

"I always wondered how it would be if a superior species landed on Earth and played Go for himself. "It plays

Nielsen told BBC.

"Now I know."



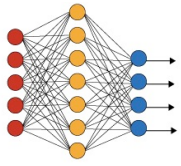
max tegmark  
@tegmark

"What we're seeing here is a model free from human bias and presuppositions. It can learn whatever it determines is optimal, which may indeed be more nuanced than our own conceptions of the same," MIT computer scientist Nick Hynes told Gizmodo following the October victory.

"AlphaZero was not 'taught' by endgame tables, and side pawns. This would be like building a combustion engine, then it experiments numerous times with every combination possible until it builds a Ferrari. ... The program had four hours to play itself many, many times, thereby becoming its own teacher."

"It's like an alien civilisation inventing its own mathematics."



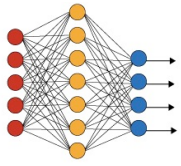


# ALPHAGO ZERO: SELF-PLAYING TO LEARN

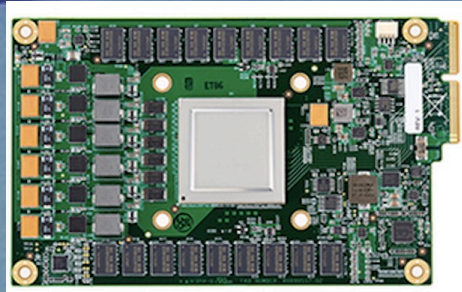
## AlphaGo Zero

Starting from scratch





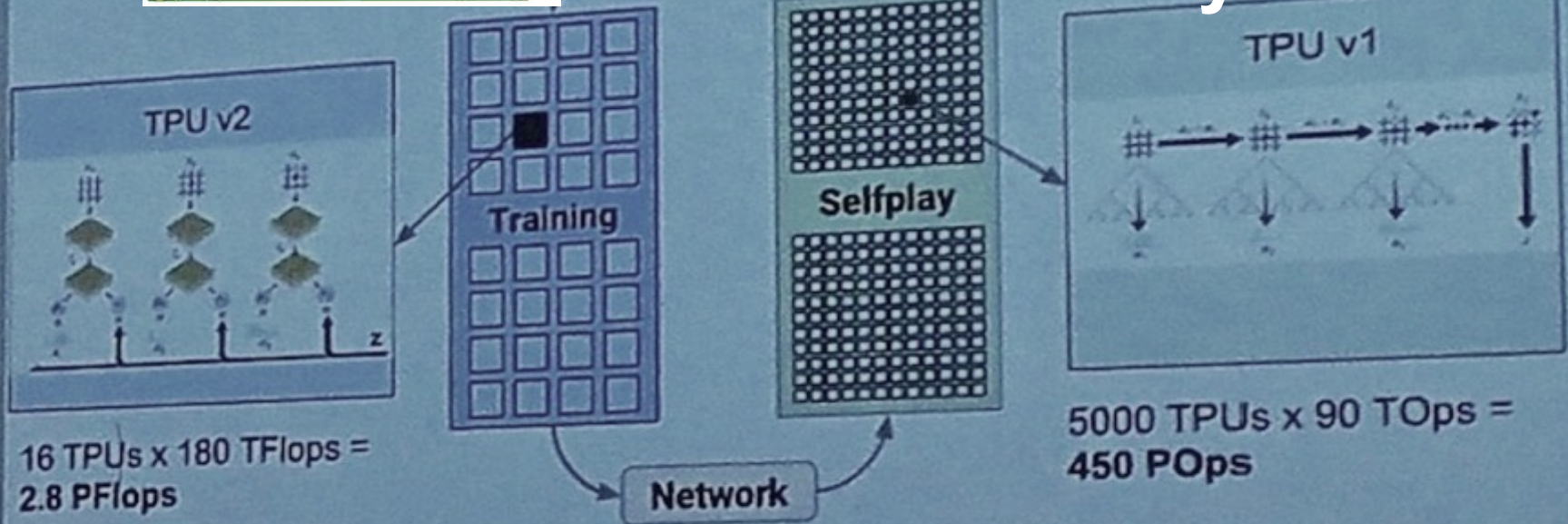
## ALPHAZERO: COMPUTING RESOURCES



AlphaZero algorithm

X 5000 = 200 KW\*

X 40 days...



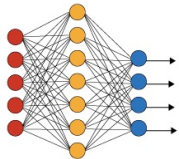
Peta =  $10^{15}$  = million of milliard

From Google Deepmind

AI Research at Google — DOMINIK GREWE & MARCO CORN

\* <https://cloud.google.com/blog/products/gcp/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu>





# 2019: OPENAI WINS DOTA 2 ESPORT GAME

## OpenAI Five Defeats Dota 2 World Champions

OpenAI Five is the first AI to beat the world champions in an esports game, having won two back-to-back games versus the world champion Dota 2 team.

It is also the first time an AI has beaten esports pros on livestream.



Human View

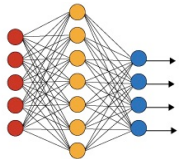


AI View

3.006	-1.386	-0.4695	0.883	1	0.84
-0.3154	-0.5425	-0.5	0.866	0	0.82
3.11	-1.36	-0.9336	0.3584	1	0.78
-2.324	2.863	0.9746	0.225	0	0.86
3.037	-1.361	-0.7773	0.6294	1	0.82
-1.387	2.951	0.988	0.1565	0	0.74
3.023	-0.9395	0.05234	-0.9985	0	0.66
2.951	-0.5747	0.01746	1	0	0.72
2.963	-1.303	0.3906	0.9204	0	0.68
2.834	-3.164	0.01746	-1	0	0.68
3.127	-1.368	0.6562	0.755	1	0.55
3.088	-1.366	0.4695	0.883	0	0.55
2.984	-1.398	-0.225	0.9746	1	0.55
3.037	-1.391	0.788	0.6157	0	0.55
3.076	-1.438	0.883	0.4695	0	0.55
-2.412	2.846	0.996	0.08716	1	0.3

## Cooperative mode

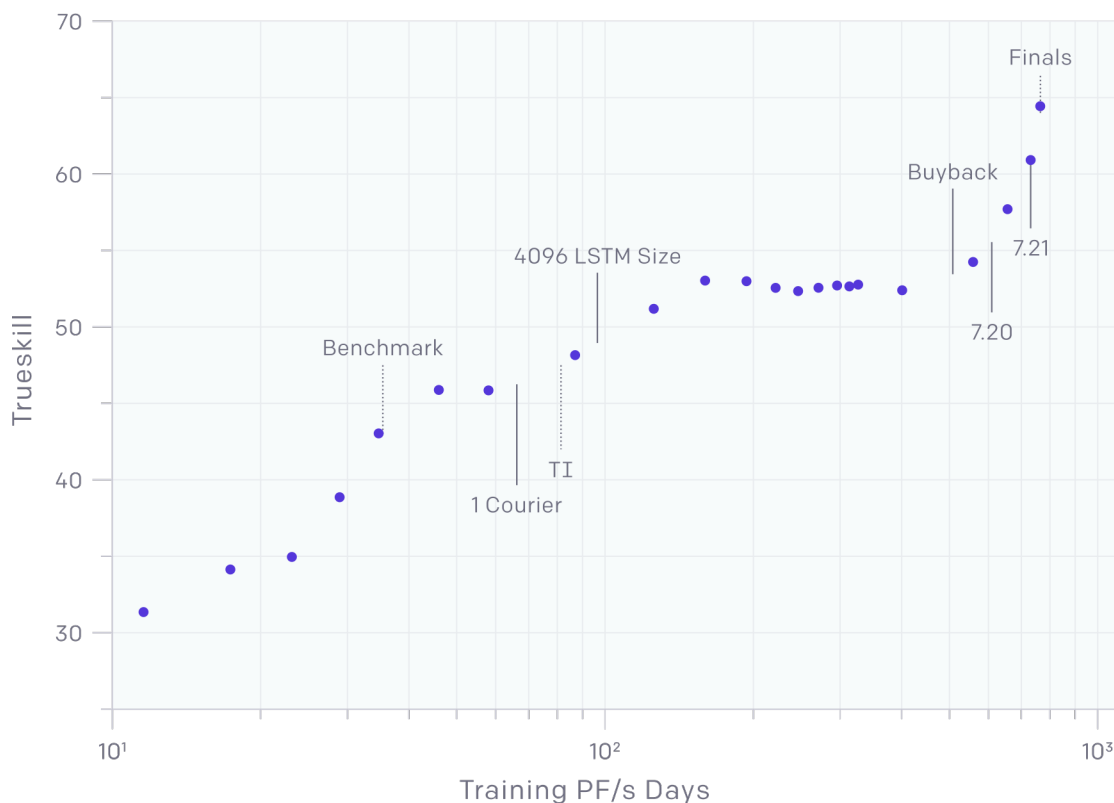
OpenAI Five's ability to play with humans presents a compelling vision for the future of human-AI interaction, one where AI systems collaborate and enhance the human experience. Our testers reported feeling supported by their bot teammates, that they learned from playing alongside these advanced systems, and that it was generally a fun experience overall.



# 2019: OPENAI WINS DOTA 2 ESPORT GAME

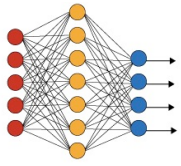
## OpenAI Five Defeats Dota 2 World Champions

In total, the current version of OpenAI Five has consumed **800 petaflop/s-days** and experienced about 45,000 years of Dota self-play over 10 realtime months (up from about 10,000 years over 1.5 realtime months as of The International), for an average of 250 years of simulated experience per day. The Finals version of OpenAI Five has a 99.9% winrate versus the TI (tournament) version.



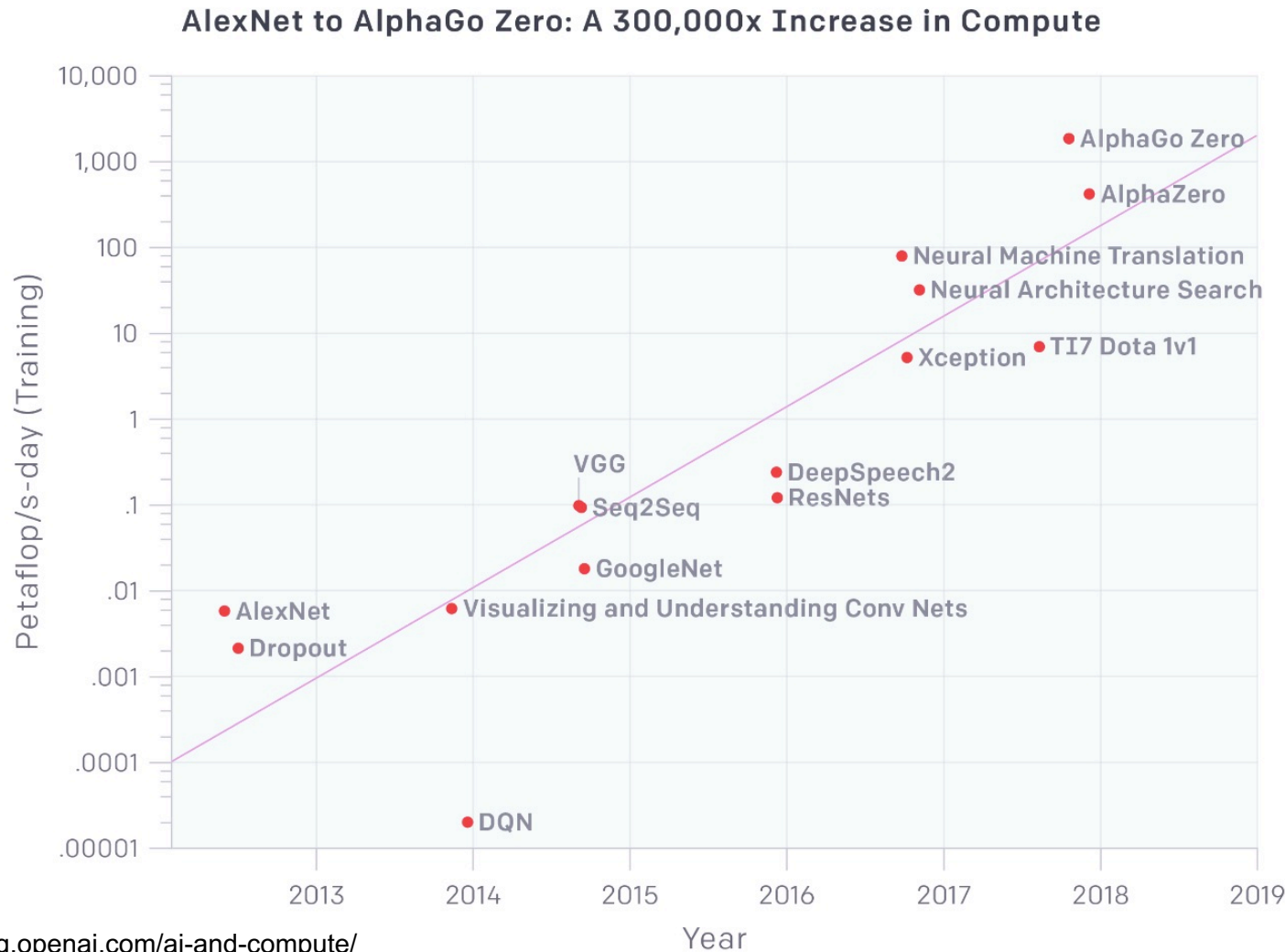
Peta =  $10^{15}$  = million of milliard



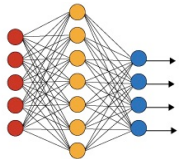


# EXPONENTIAL INCREASE OF COMPUTING POWER FOR AI TRAINING

***“Since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a 3.5 month-doubling time  
(by comparison, Moore’s Law had an 18-month doubling period)”***

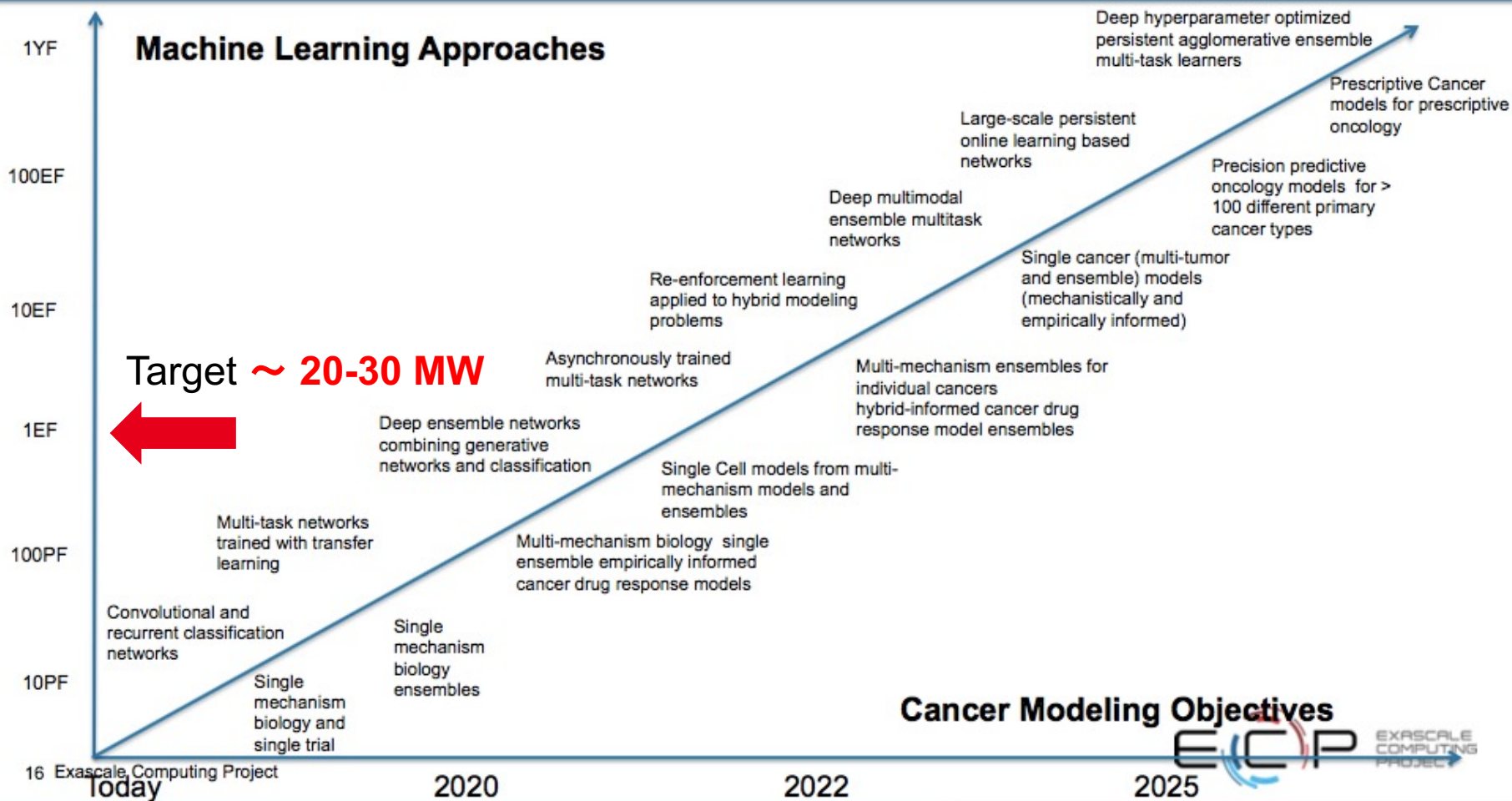


\* <https://blog.openai.com/ai-and-compute/>



# ALWAYS MORE COMPUTING RESSOURCES

## Roadmap for Integration of Deep Learning and Simulation for Predictive Oncology



From Paul Messina, Argonne National Laboratory

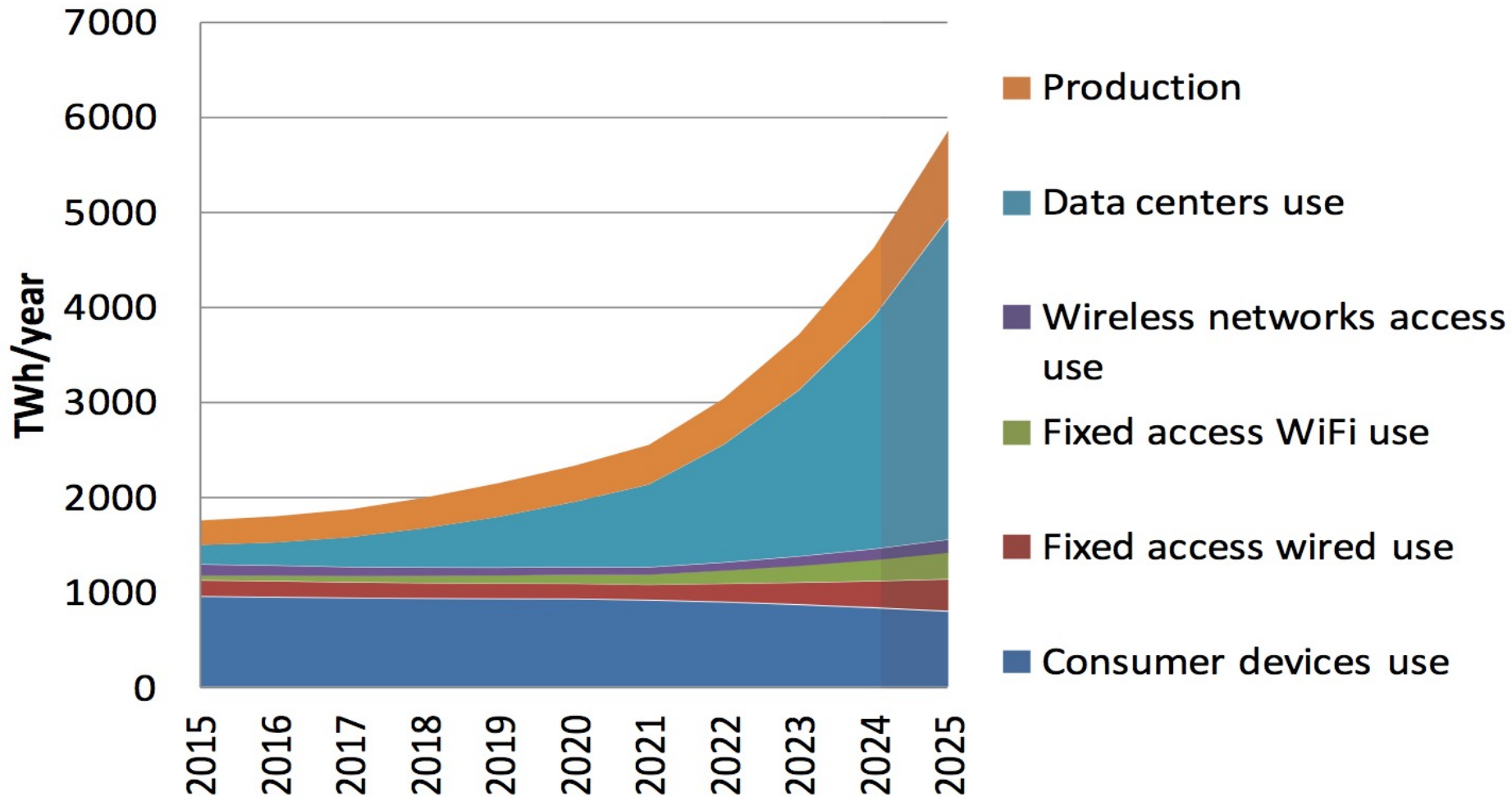


**HOUSTON**

**WE HAVE A PROBLEM...**

# The problem:

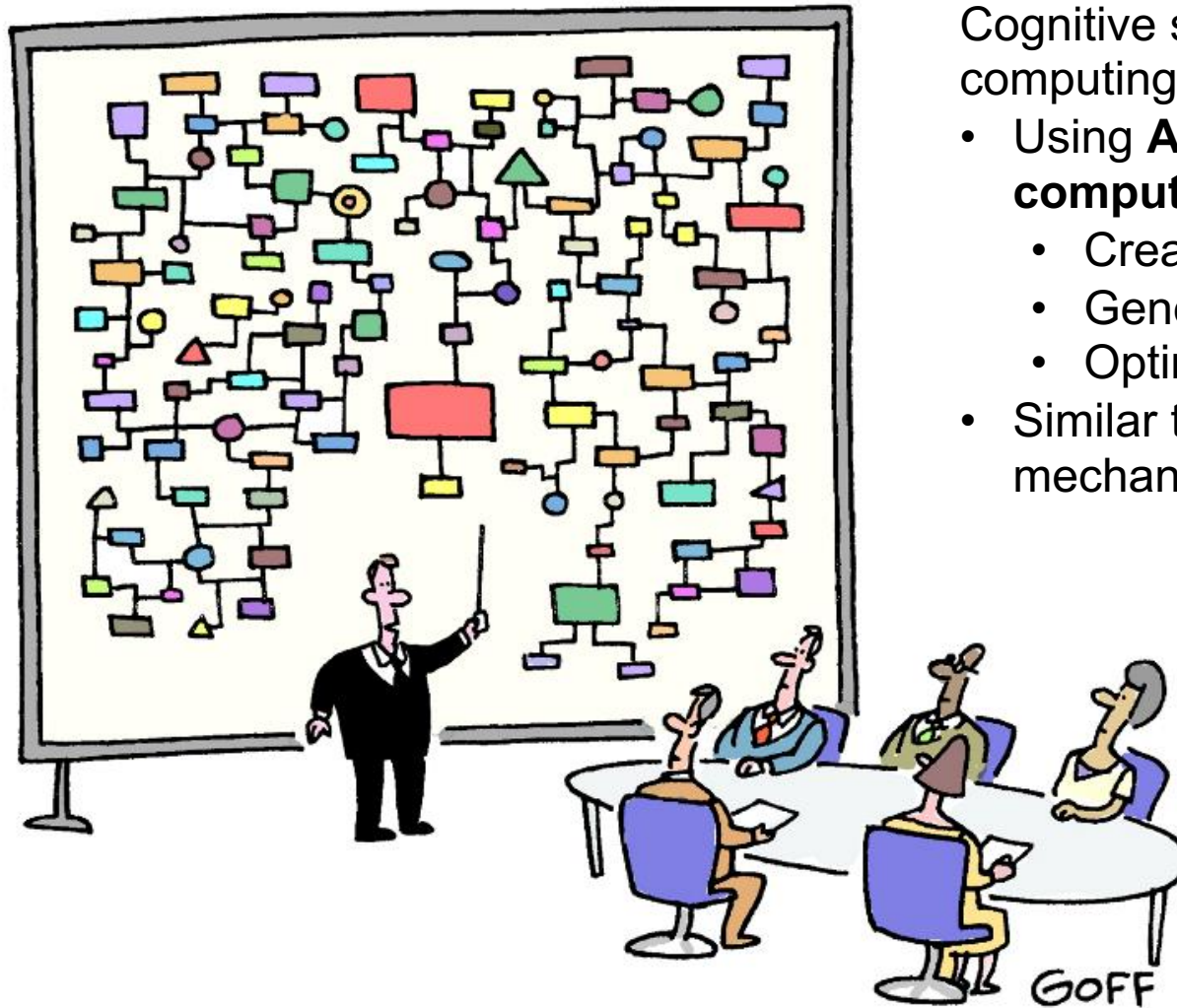
## Expected case scenario



From "Total Consumer Power Consumption Forecast", Anders S.G. Andrae, October 2017



# Managing complexity



Cognitive solutions for complex computing systems:

- Using **AI techniques for computing systems**
  - Creating new hardware
  - Generating code
  - Optimizing systems
- Similar to ***Generative design*** for mechanical engineering

**"And that's why we need a computer."**

# AI FOR MAKING COMPUTING SYSTEMS: “GENERATIVE DESIGN” APPROACH

The user *only states desired goals and constraints*

-> The *complexity wall* might *prevent explaining* the solution

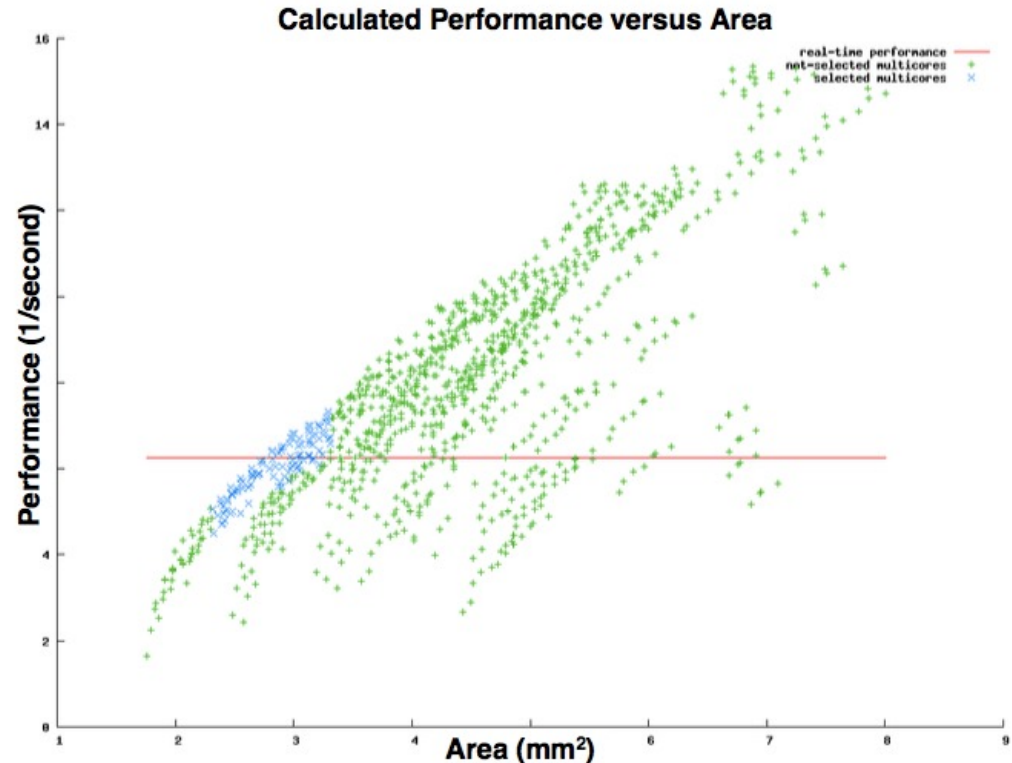


Motorcycle swingarm: the piece that hinges the rear wheel to the bike's frame

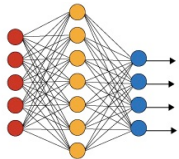
# EXAMPLE: DESIGN SPACE EXPLORATION FOR DESIGN MULTI-CORE PROCESSORS<sup>1</sup> (2010)

- Ne-XVP project – Follow-up of the TriMedia VLIW (<https://en.wikipedia.org/wiki/Ne-XVP>)
- 1,105,747,200 heterogeneous multicores in the design space
- 2 millions years to evaluate all design points
- AI inspired techniques allowed to reduce the induction time to only few days

**=> x16 performance increase**

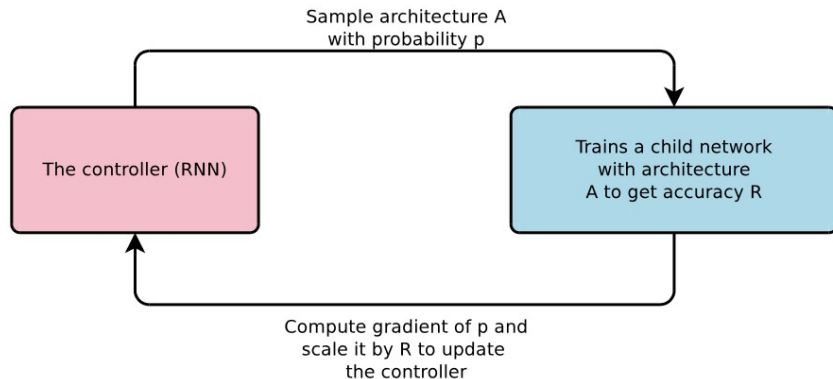


<sup>1</sup> M. Duranton et al., "Rapid Technology-Aware Design Space Exploration for Embedded Heterogeneous Multiprocessors" in Processor and System-on-Chip Simulation, Ed. R. Leupers, 2010



# 2017: GOOGLE; USING DEEP LEARNING TO DESIGN DEEP LEARNING

**“Neural Architecture Search”, using a recurrent neural network to compose neural network architectures using reinforcement learning on CIFAR-10 (character recognition)**



From arXiv:1611.01578v2, Barret Zoph, Quoc V. Le  
Google Brain

Model	Depth	Parameters	Error rate (%)
Network in Network (Lin et al., 2013)	-	-	8.81
All-CNN (Springenberg et al., 2014)	-	-	7.25
Deeply Supervised Net (Lee et al., 2015)	-	-	7.97
Highway Network (Srivastava et al., 2015)	-	-	7.72
Scalable Bayesian Optimization (Snoek et al., 2015)	-	-	6.37
FractalNet (Larsson et al., 2016)	21	38.6M	5.22
with Dropout/Drop-path	21	38.6M	4.60
ResNet (He et al., 2016a)	110	1.7M	6.61
ResNet (reported by Huang et al. (2016c))	110	1.7M	6.41
ResNet with Stochastic Depth (Huang et al., 2016c)	110	1.7M	5.23
	1202	10.2M	4.91
Wide ResNet (Zagoruyko & Komodakis, 2016)	16	11.0M	4.81
	28	36.5M	4.17
ResNet (pre-activation) (He et al., 2016b)	164	1.7M	5.46
	1001	10.2M	4.62
DenseNet ( $L = 40, k = 12$ ) (Huang et al. (2016a))	40	1.0M	5.24
DenseNet ( $L = 100, k = 12$ ) (Huang et al. (2016a))	100	7.0M	4.10
DenseNet ( $L = 100, k = 24$ ) (Huang et al. (2016a))	100	27.2M	3.74
Neural Architecture Search v1 no stride or pooling	15	4.2M	5.50
Neural Architecture Search v2 predicting strides	20	2.5M	6.01
Neural Architecture Search v3 max pooling	39	7.1M	4.47
Neural Architecture Search v3 max pooling + more filters	39	37.4M	3.65

Several other interesting **“Auto-ML”** research projects are on-going



# 2017: A NEW BEATLES SONG ?



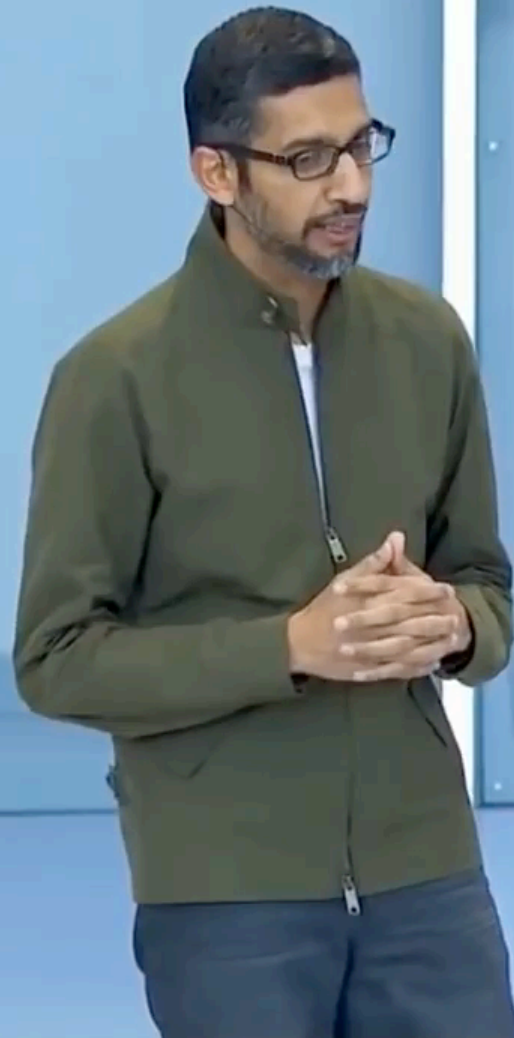
Sony Computer  
Science Laboratories,  
Paris



**No, this song is created by an IA inspired by the Beatles**

# OTHER APPLICATIONS OF DEEP LEARNING

## 2018: GOOGLE DUPLEX: RESTAURANT

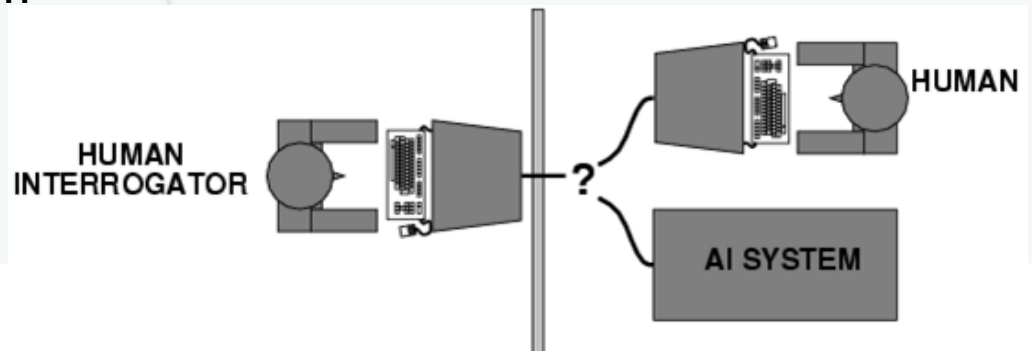


# 2018: GOOGLE DUPLEX: RESTAURANT



*“For when? Tomorrow? Or weekend?”*

Remember the Turing test of 1950...

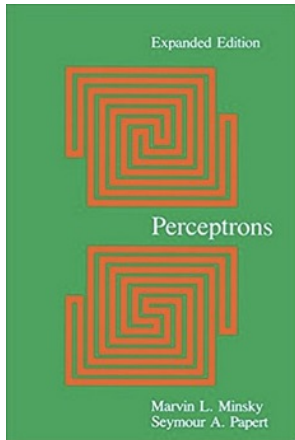




# WHAT'S NEXT FOR DEEP LEARNING AND AI?

**1<sup>st</sup> Winter: 1970s**  
*The Lighthill report*  
(published in 1973)

**2<sup>nd</sup> Winter: 1987**  
*Perceptrons (2<sup>nd</sup> ed)*  
Minsky & Papert



**3<sup>rd</sup> Winter: 1993**  
SVM  
Vapnik & Cortes (1963)

**4<sup>th</sup> Winter or**  
**Plateau of Productivity?**

## Gartner Hype Cycle for Emerging Technologies, 2019



## Gartner Hype cycle - 2019

Paas: Platform as a Service

## CONCLUSION: WE LIVE AN EXCITING TIME!

***“The best way to predict the future is to invent it.”***

*Alan Kay*













# Thank you for your attention

[marc.duranton@cea.fr](mailto:marc.duranton@cea.fr)



**leti**

Centre de Grenoble  
17 rue des Martyrs  
38054 Grenoble Cedex

**list**

Centre de Saclay  
Nano-Innov PC 172  
91191 Gif sur Yvette Cedex