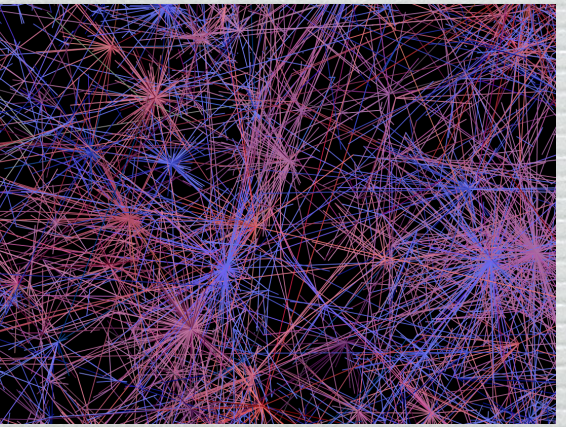
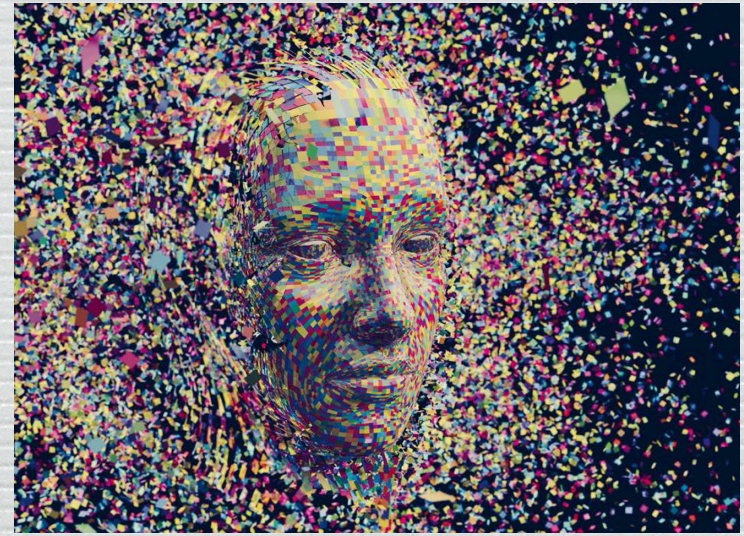
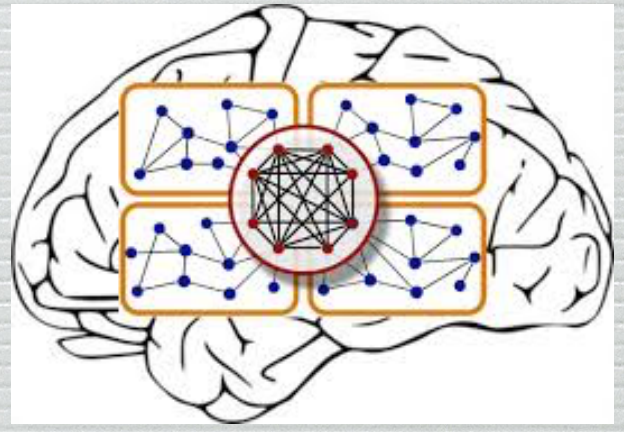
Artificial Intelligence: Success, Limits, Myths and Threats





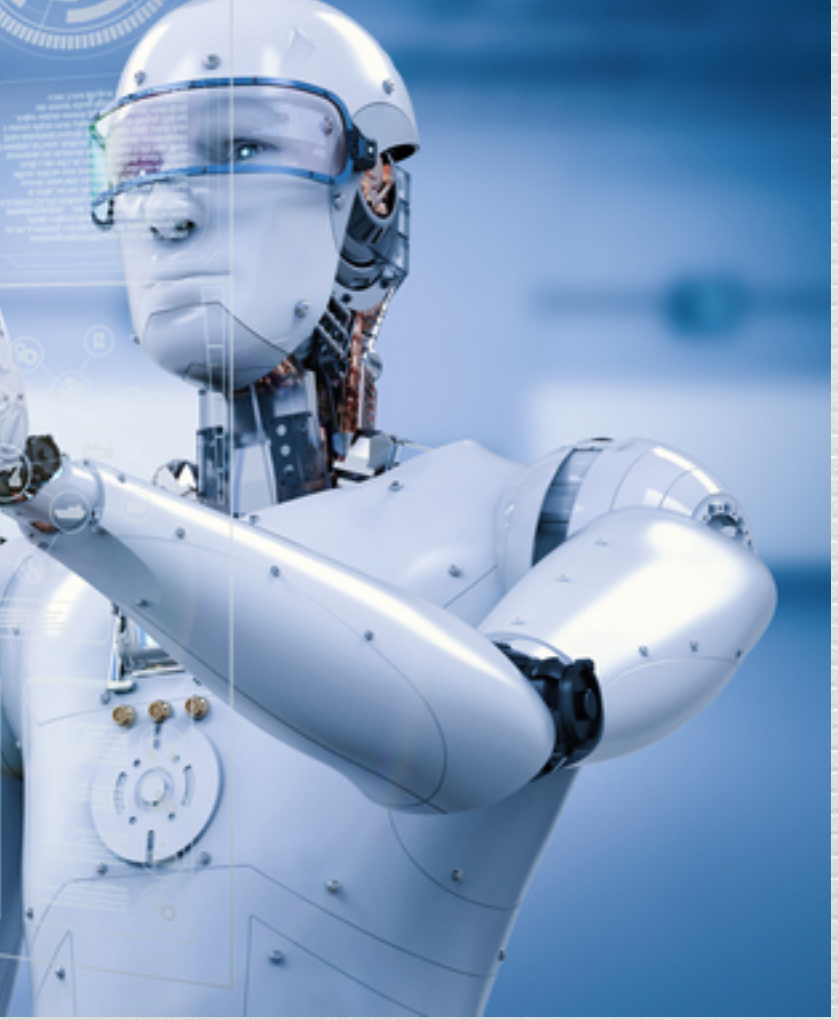


Marc Mézard
Ecole normale supérieure, PSL University

Salam lectures,
ICTP Trieste,
January 27, 2020

Chapter One

Myths and Reality



“What’s actually happening is machines are powering all of us. [.]They may not yet be inside our bodies, but, by the 2030s, we will connect our neocortex, the part of our brain where we do our thinking, to the cloud.”

“We’re going to get more neocortex, we’re going to be funnier, we’re going to be better at music. We’re going to be sexier.”

Ray Kurzweil, Director of Engineering at Google, MIT
Lemelson Prize, National Medal of Technology

“Artificial Intelligence will make you smarter”

Terry Sejnowski, UCSD

AI is "a fundamental existential risk for human civilisation". "In a way that car accidents, airplane crashes, faulty drugs or bad food were not. They were harmful to a set of individuals in society, but they were not harmful to society as a whole".

Elon Musk, Tesla

I fear that AI may replace humans altogether. […]
If people design computer viruses, someone will design AI that improves and replicates itself. This will be a new form of life that outperforms humans.

Stephen Hawking

The new era of AI

Machine learning with « deep neural networks »:

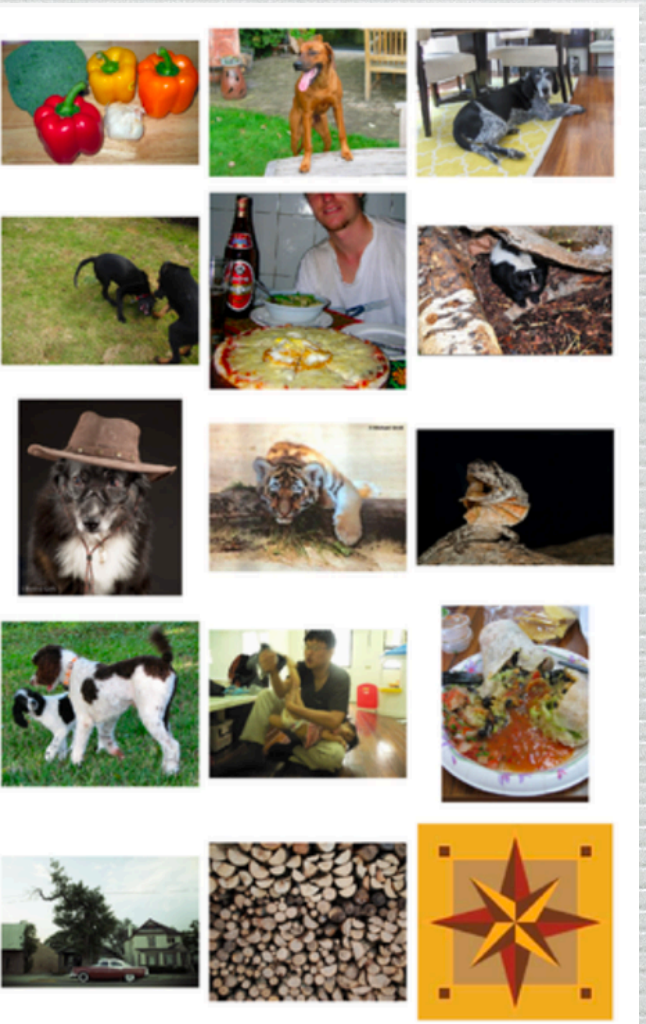
1- Image understanding.

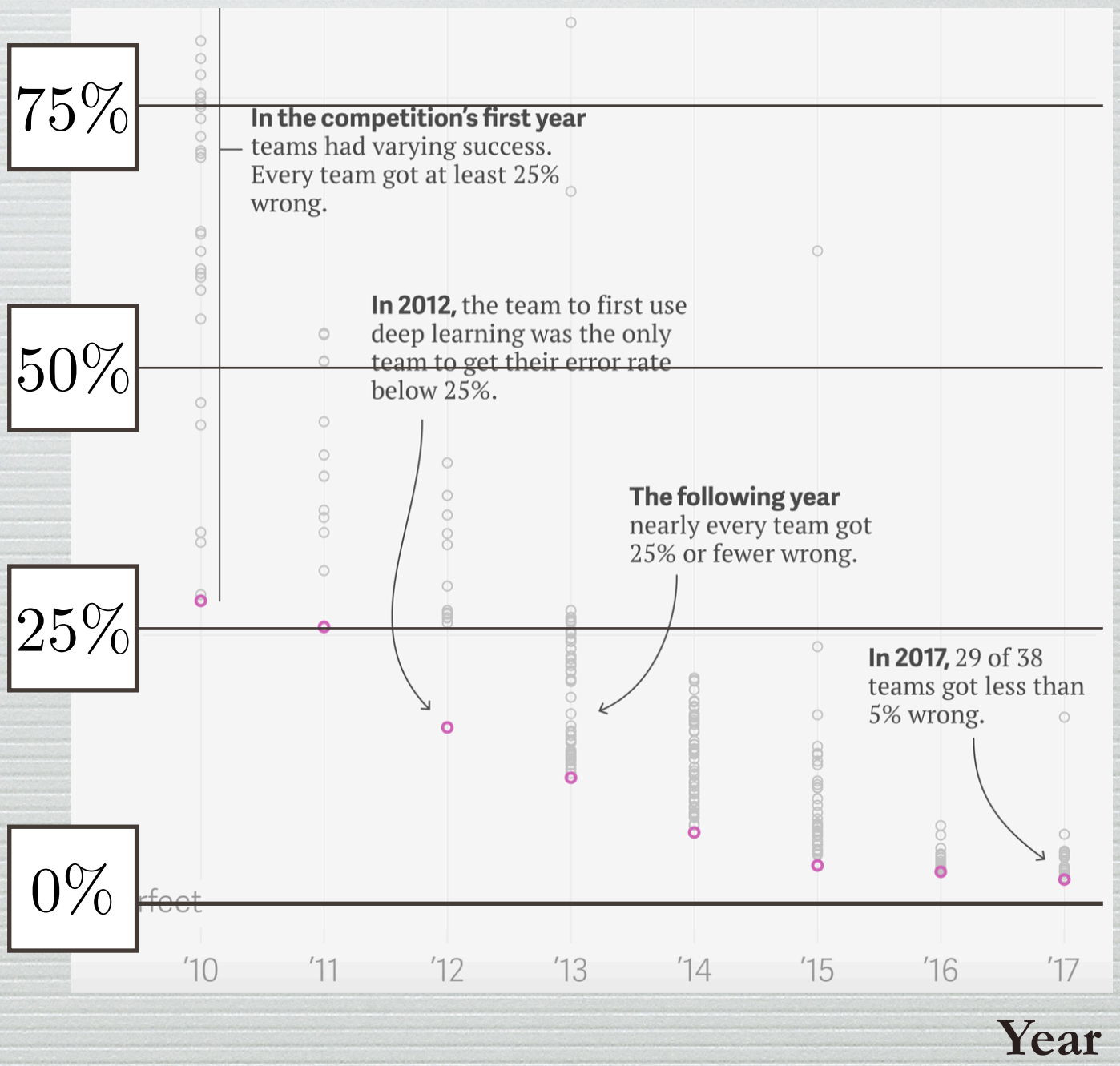
In the last decade, detection, segmentation and recognition of objects and regions in natural images:

- classification,
- detection of faces,
- segmentation
- …

ImageNet Database and challenge

One million images
One thousand categories





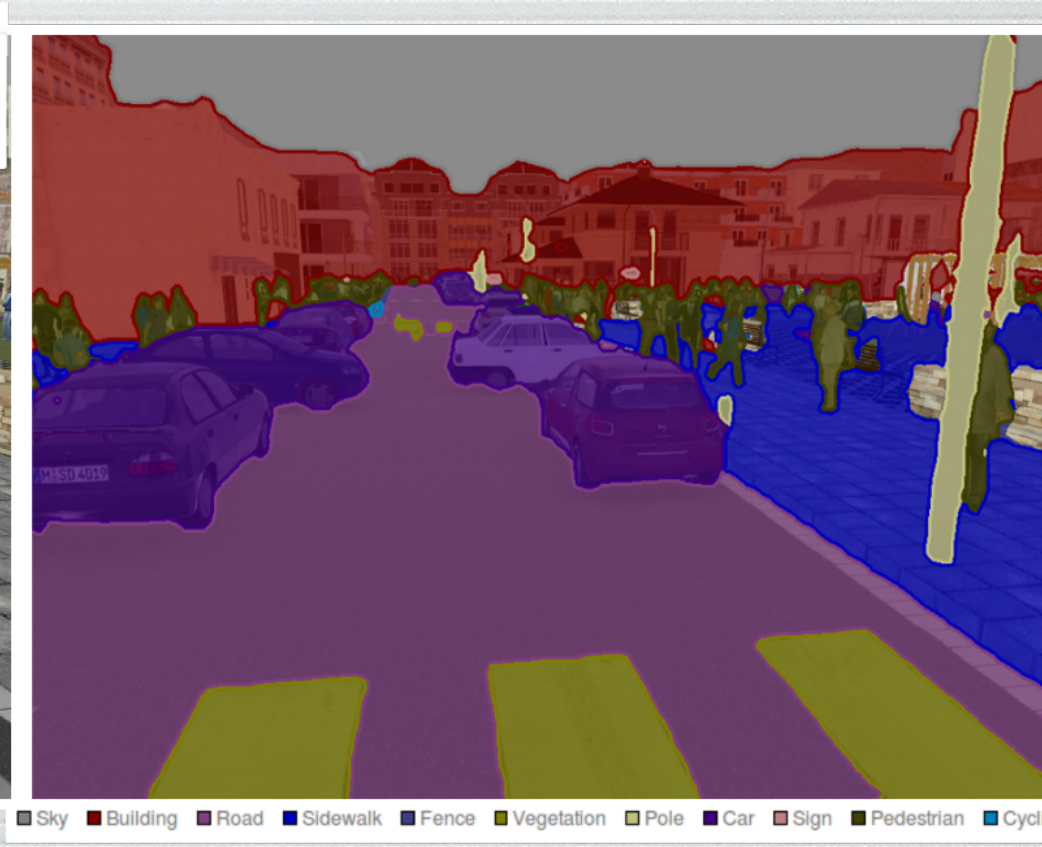
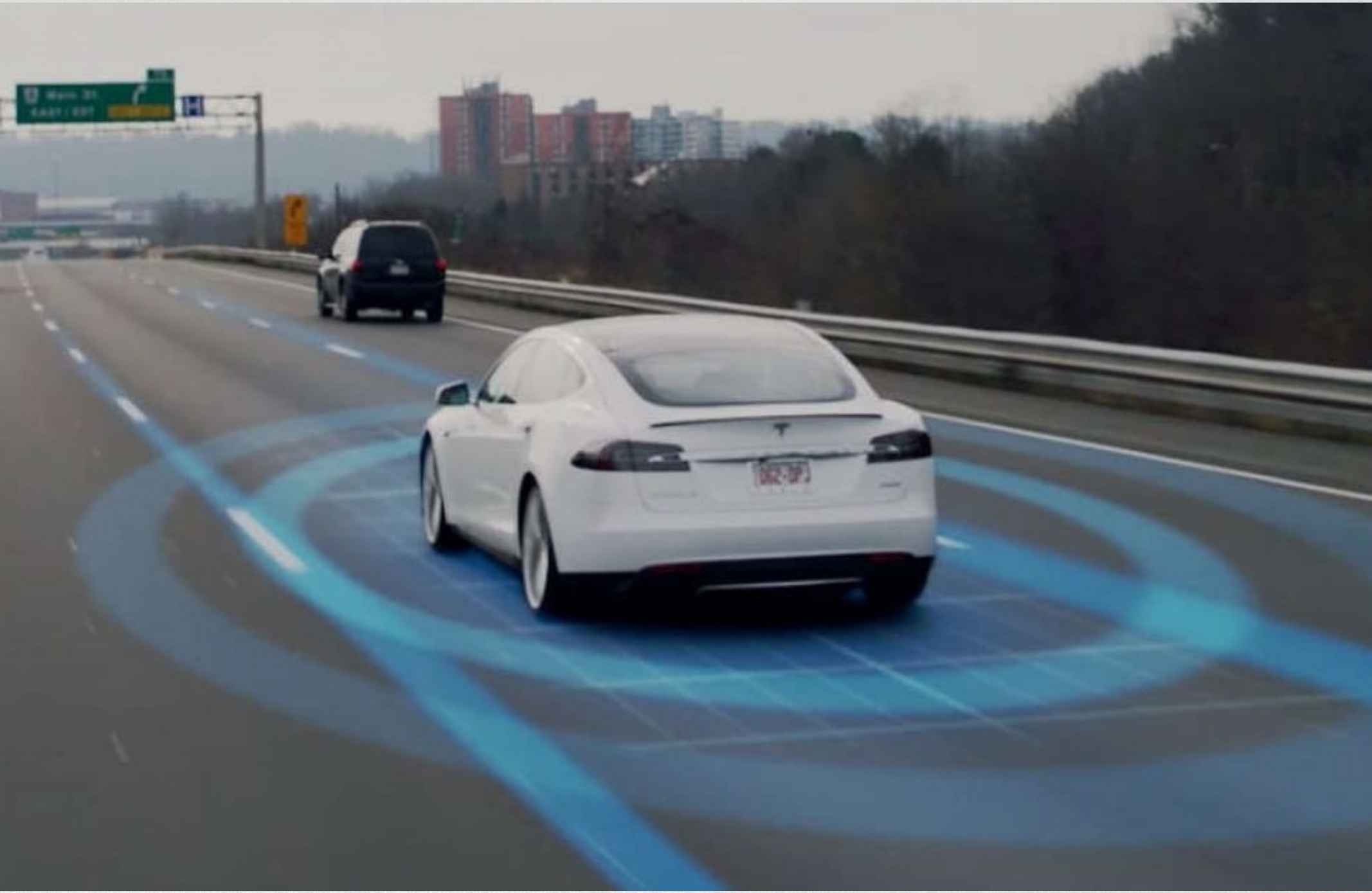


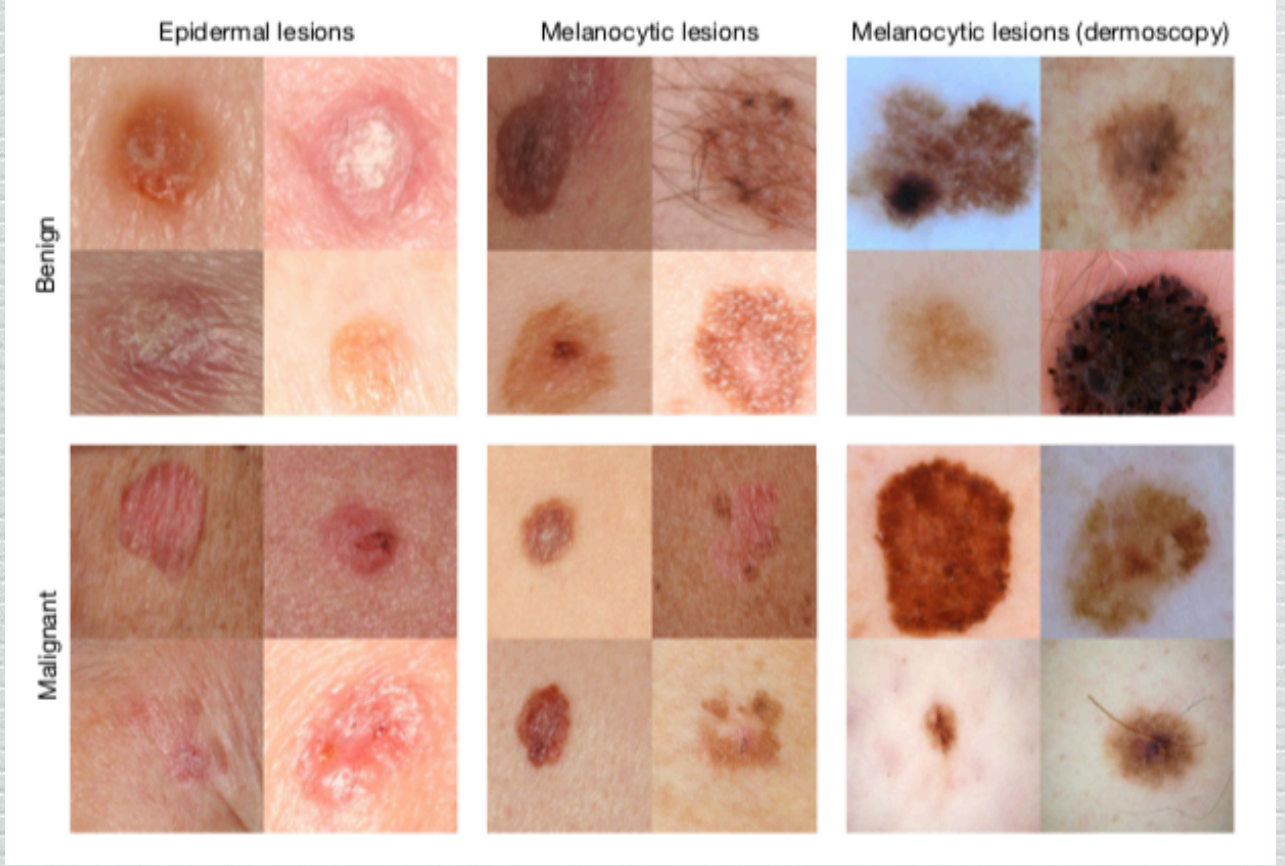
Image segmentation



Convoy of self-driving trucks completes first European cross-border trip

‘Platoon’ of wireless-linked trucks arrives in Netherlands port city of Rotterdam, giving a glimpse of the future of road haulage

▲ Semi-automated trucks are driven on the E19 highway in Vilvoorde on Tuesday as part of the ‘EU truck platooning challenge’. Photograph: Eric Lalmand/AFP/Getty Images

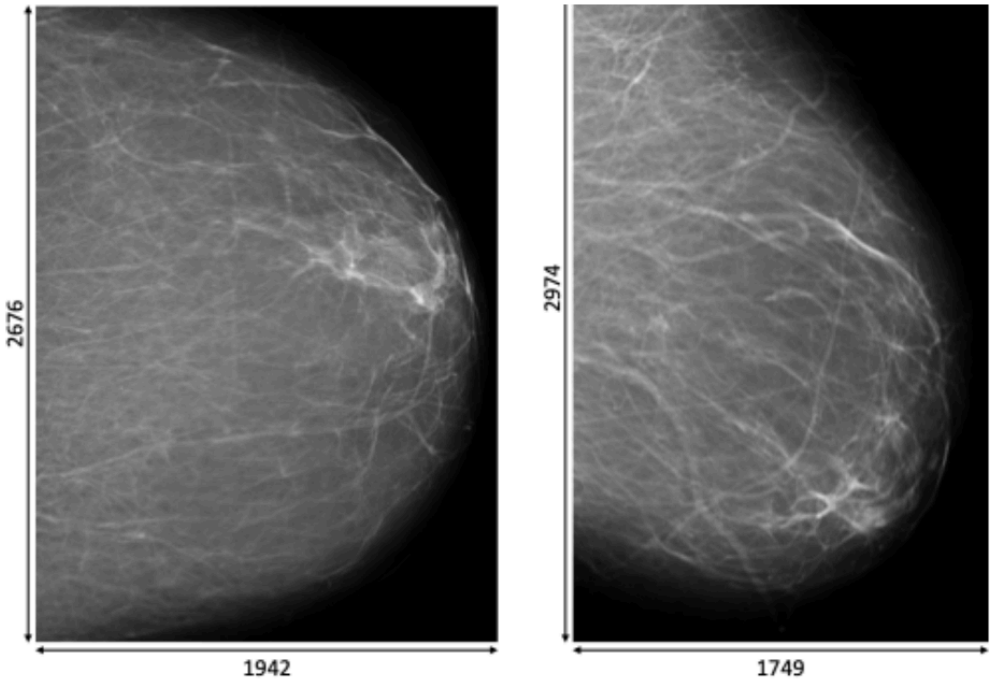


Deep neural network pre-trained on 1.4 millions images (Image-net), then trained on 130,000 clinical images, benign and malignant

Esteva et al. Nature, 2017

« The artificial neural network achieves performance on par with all tested experts, demonstrating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists. »

« Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. »

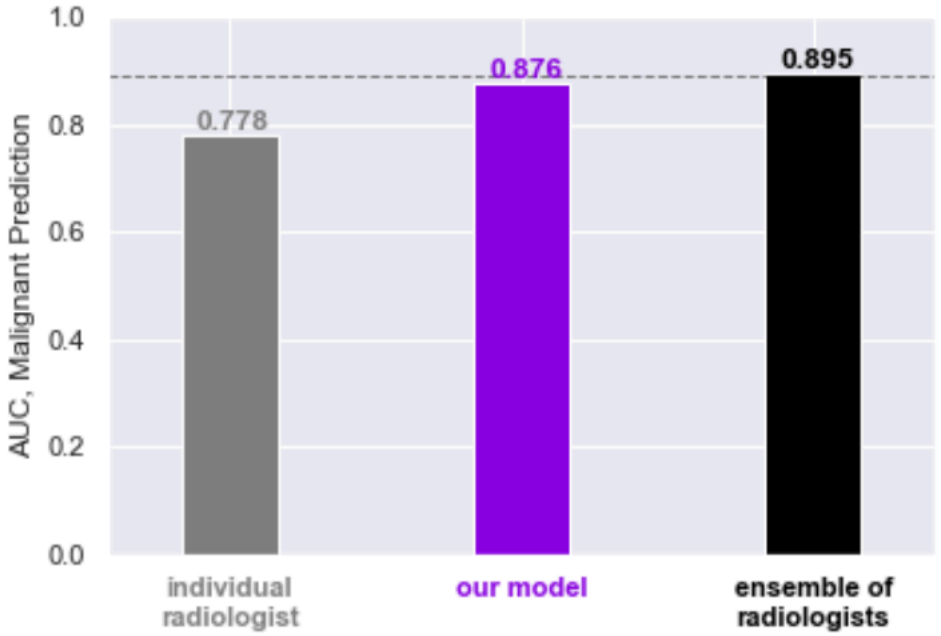


Breast cancer screening: analyzing mammography images

Nan Wu et al., MIDL Conference, April 2019

Giant database of 1 million images !!

Compared performance of the mammography test : individual radiologist, model, and panel of 14 radiologists



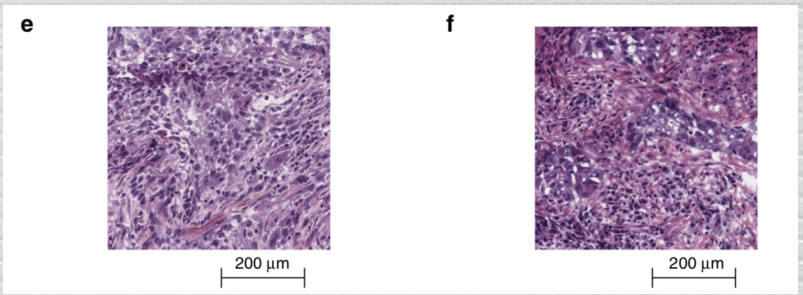
Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features

Kun-Hsing Yu, Ce Zhang, Gerald J. Berry, Russ B. Altman, Christopher Ré, Daniel L. Rubin & Michael Snyder

Nature Communications **7**, Article number: 12474 (2016) | Download Citation

« We aim to improve the prognostic prediction of lung adenocarcinoma and squamous cell carcinoma patients through objective features distilled from histopathology images »

Data : 2200 stained histopathology images of lung adenocarcinoma. Use machine learning to predict short term vs long-term survivors



Kun-Hsing Yu et al. ,
Nature Communications
2016

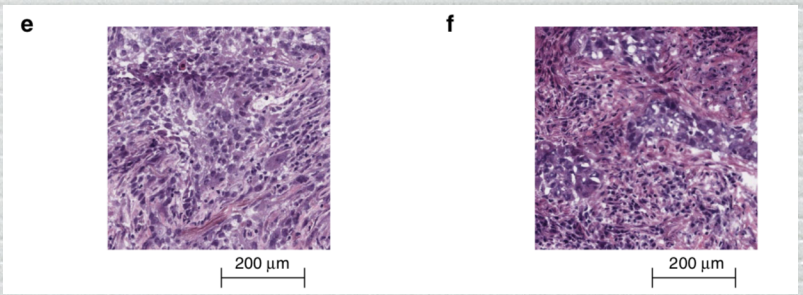
Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features

Kun-Hsing Yu, Ce Zhang, Gerald J. Berry, Russ B. Altman, Christopher Ré, Daniel L. Rubin & Michael Snyder

Nature Communications **7**, Article number: 12474 (2016) | Download Citation

« We aim to improve the prognostic prediction of lung adenocarcinoma and squamous cell carcinoma patients through objective features distilled from histopathology images »

Data : 2200 stained histopathology images of lung adenocarcinoma. Use machine learning to predict short term vs long-term survivors



Kun-Hsing Yu et al. , Nature Communications 2016

« Our results suggest that automatically derived image features can predict the prognosis of lung cancer patients and thereby contribute to precision oncology. Our methods are extensible to histopathology images of other organs. »

The new era of AI

Machine learning with « deep neural networks »:

1- Image understanding.

Since the early 2000s : detection, segmentation and recognition of objects and regions in images. (traffic sign recognition, detection of faces, text, pedestrians and human bodies in natural images, face recognition…)

But also:

2- Language understanding: topic classification, question answering, language translation

3- Lip-reading

4- Predicting the activity of potential drug molecules

5- Analysing particle accelerator data

6- Designing new molecules for biochemistry

7- Playing games (chess, go, poker, video-games,…)

etc.

2- Language understanding:

topic classification, question answering, language translation

2- Language understanding:

topic classification, question answering, language translation

DeepL | Traducteur | Linguee | DeepL Pro | Blog | Infos

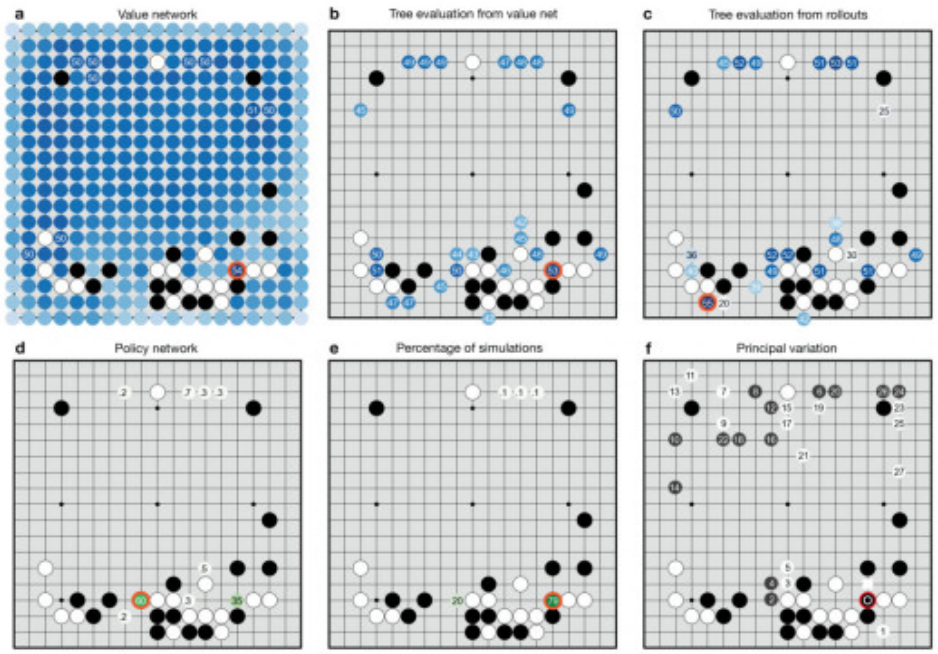
Traduire **français** (langue identifiée)

Longtemps, je me suis couché de bonne heure. Parfois, à peine ma bougie éteinte, mes yeux se fermaient si vite que je n'avais pas le temps de me dire : « Je m'endors. »

Traduire le document

Traduire en **anglais**

For a long time, I went to bed early. Sometimes, as soon as my candle went out, my eyes would close so quickly that I didn't have time to say to myself, "I'm falling asleep. »



AlphaGo

March 2016 : Alpha Go wins against Lee Sedol, the Korean who was 18-time world champion of Go

July 2019 : Pluribus (Facebook+Carnegie Mellon) is the first AI program capable of beating human players at poker (6 player no limit Hold'em)

Partie de poker entre Pluribus et joueurs

October 2019: Alphastar becomes a grandmaster at Starcraft 2. Top 0.2 % Competition and cooperation in multi-agent setting

Chapter Two

Machine learning

Machine Learning

Input → Machine  → Output

Machine Learning

Input → Machine → Output

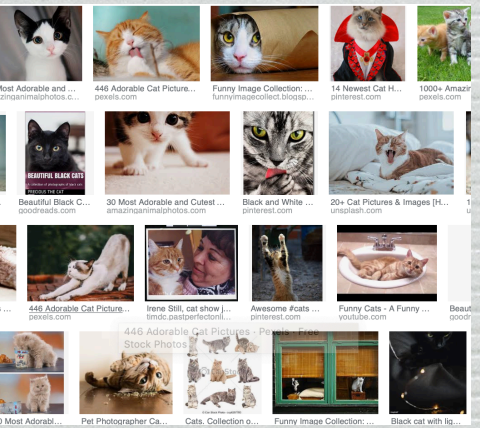
CAT

DOG

Machine Learning

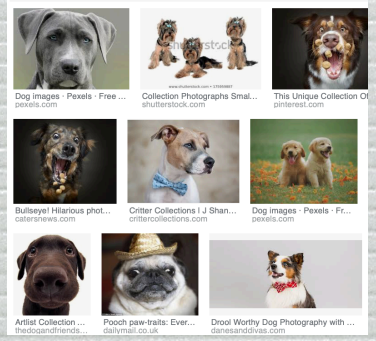
Present many pictures from a cat database, many pictures from a dog database

Input → Machine → Output



Tune the knobs of the machine until it outputs the right results

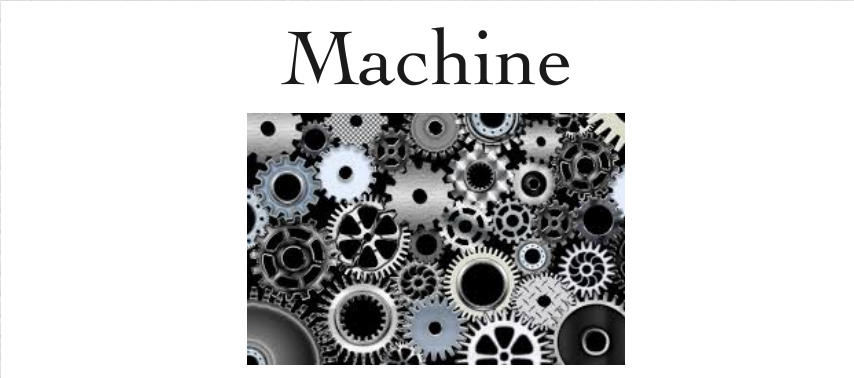
→ CAT



→ DOG

Test phase=present new picture, that the machine has not yet seen

Data



Output

How many mistakes on new pictures?

CAT

CAT

Test phase=present new picture, that the machine has not yet seen

Data → Machine → Output

How many mistakes on new pictures?

CAT

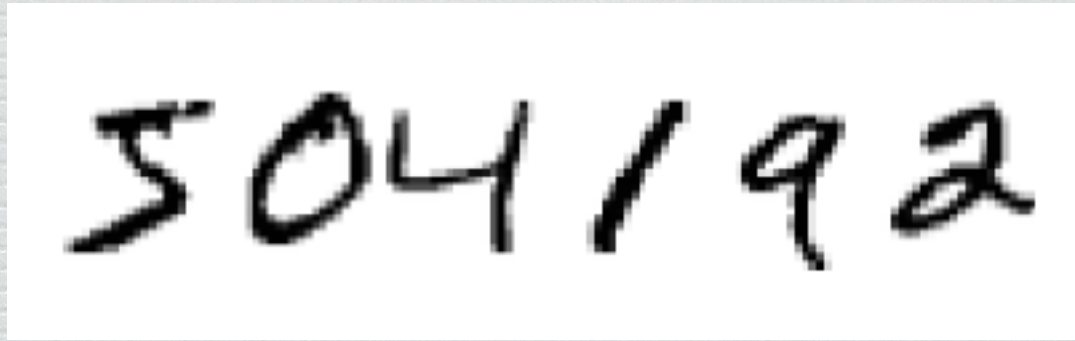
~~CAT~~

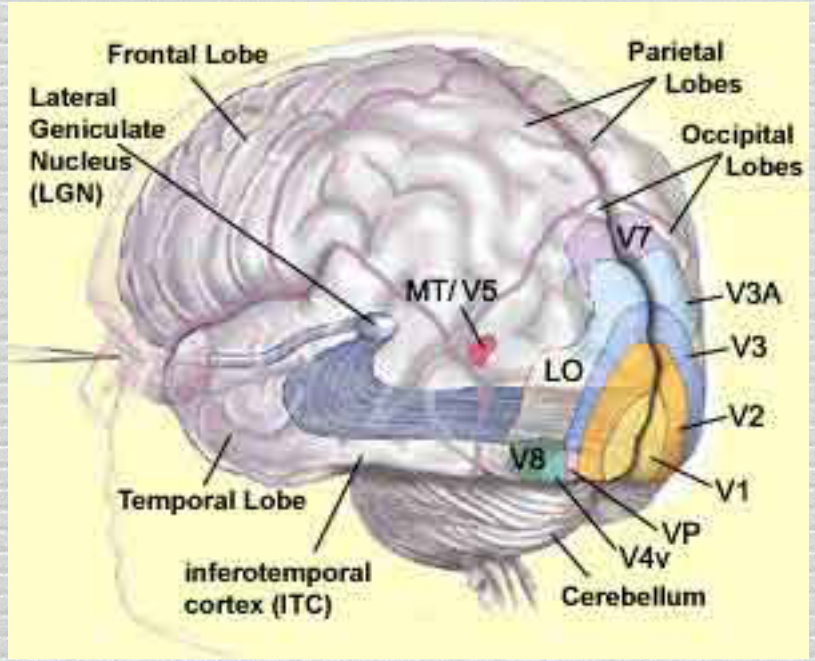
Chapter Three

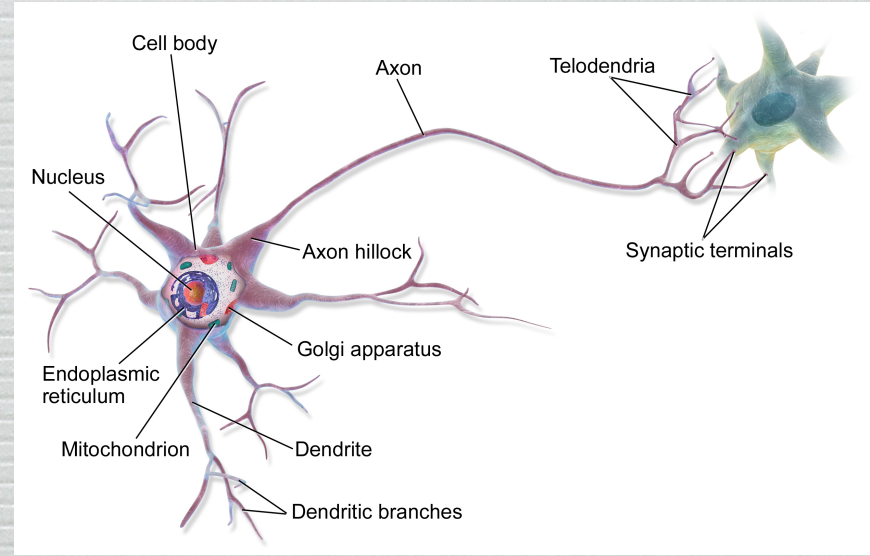
The Machines:

Artificial Neural Networks

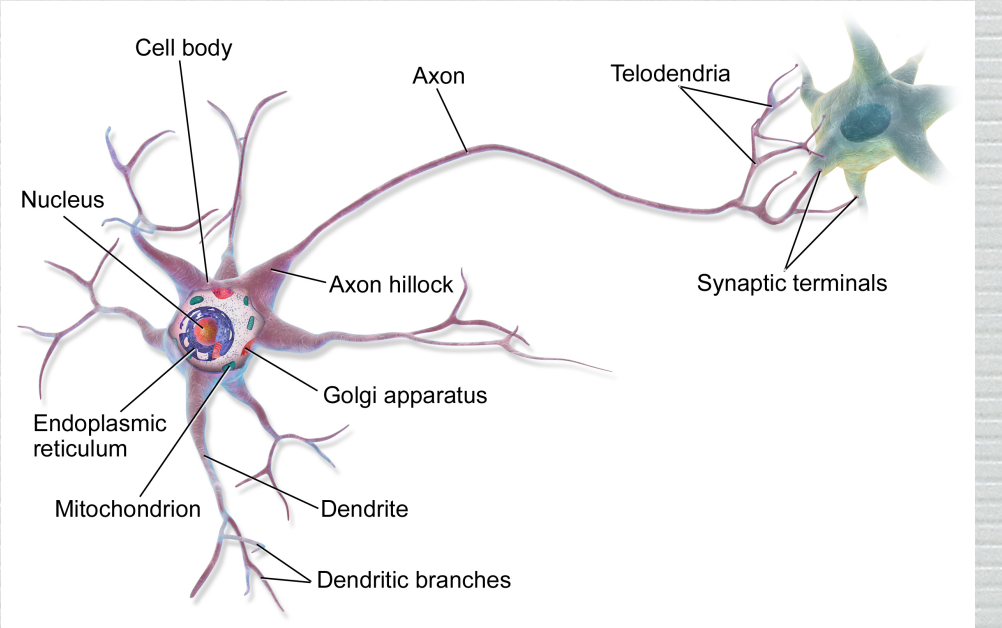
Everyone recognizes

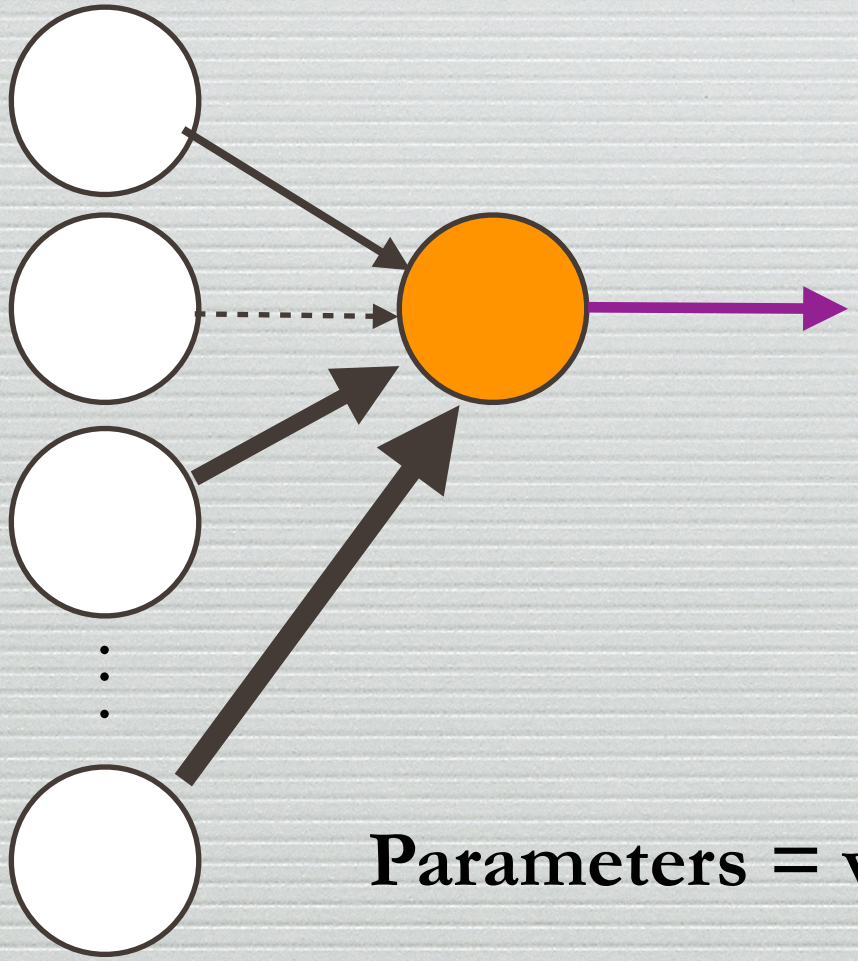






Artificial neural networks



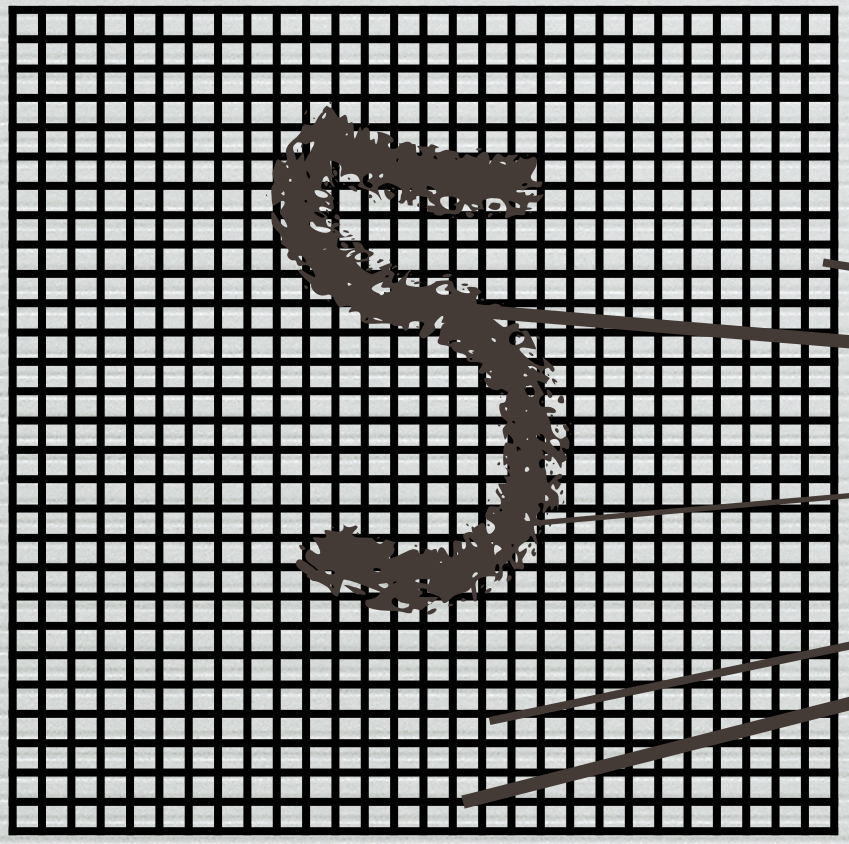


y

Output= sum of inputs (beyond a threshold)

Parameters = weight of each input

Perceptron (Franck Rosenblatt, ‘50s) : One output neuron for each digit. Learn the 784 weights with machine learning



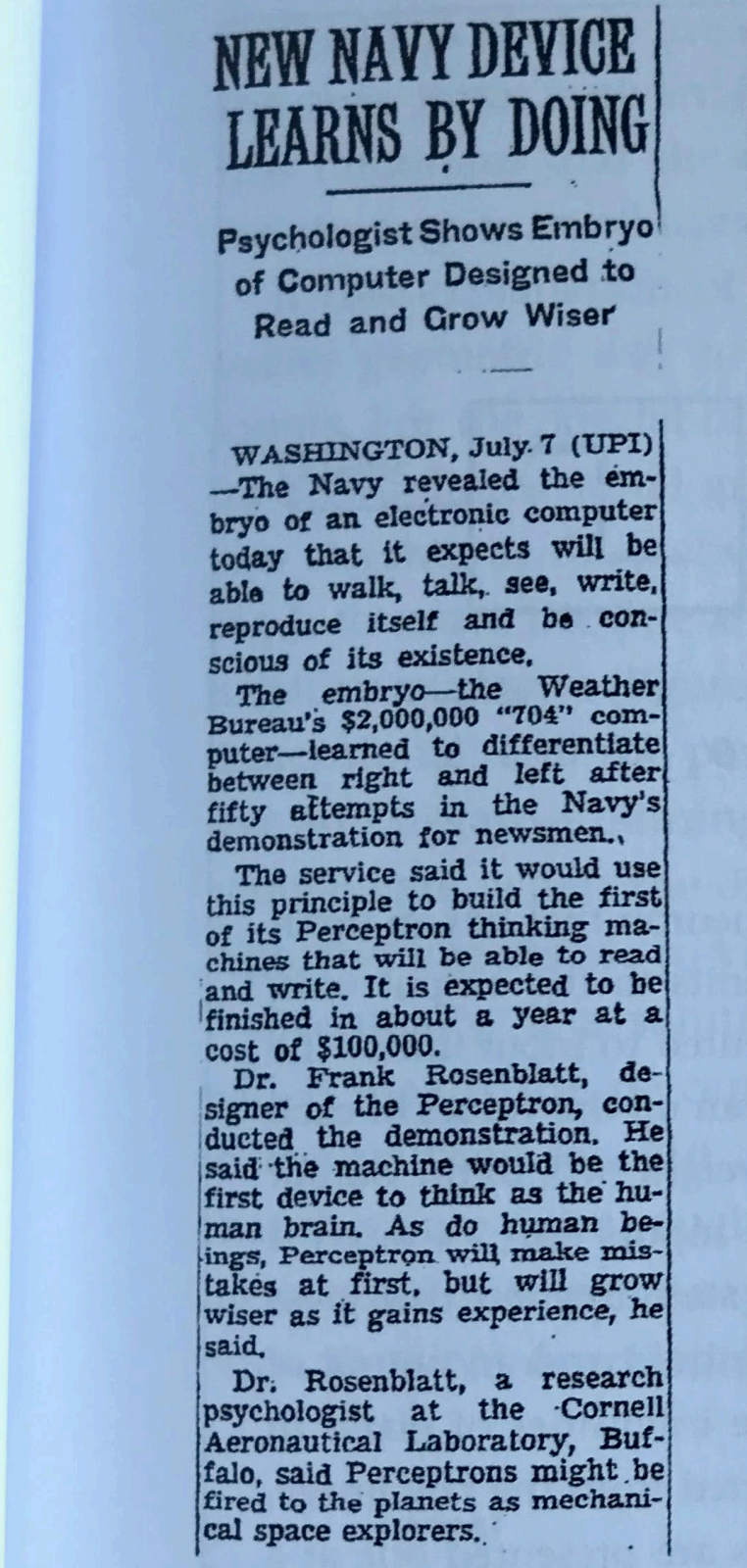
Input

0
1
2
3
4
5
6
7
8
9

Output

Frank Rosenblatt's perceptron in the NY Times, July 8, 1958

« *The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence* »



NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Frank Rosenblatt's perceptron in the NY Times, July 8, 1958

« *The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence* »

… but the perceptron had fundamentally limited performances !

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

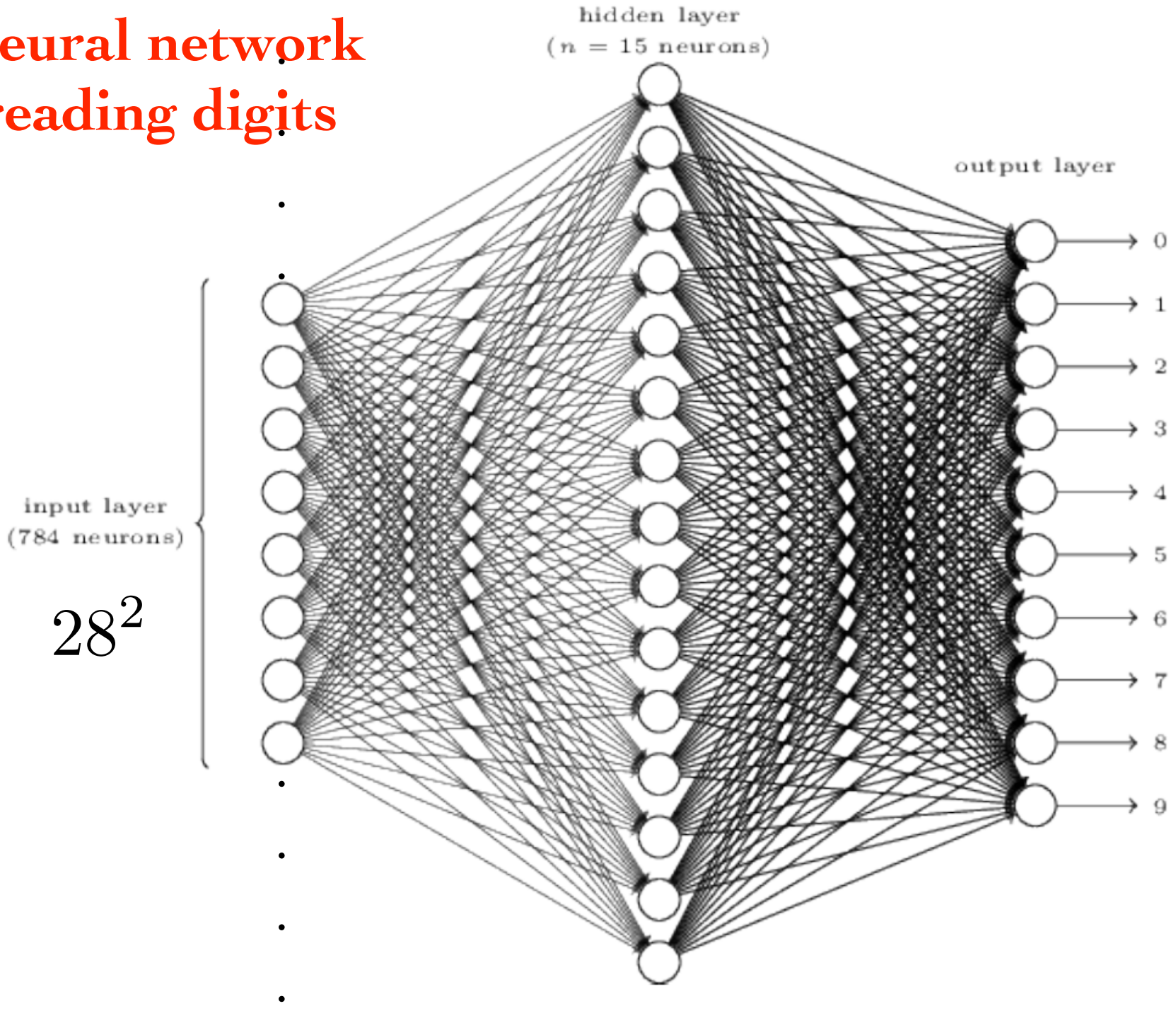
Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

What is new since Rosenblatt's perceptron?

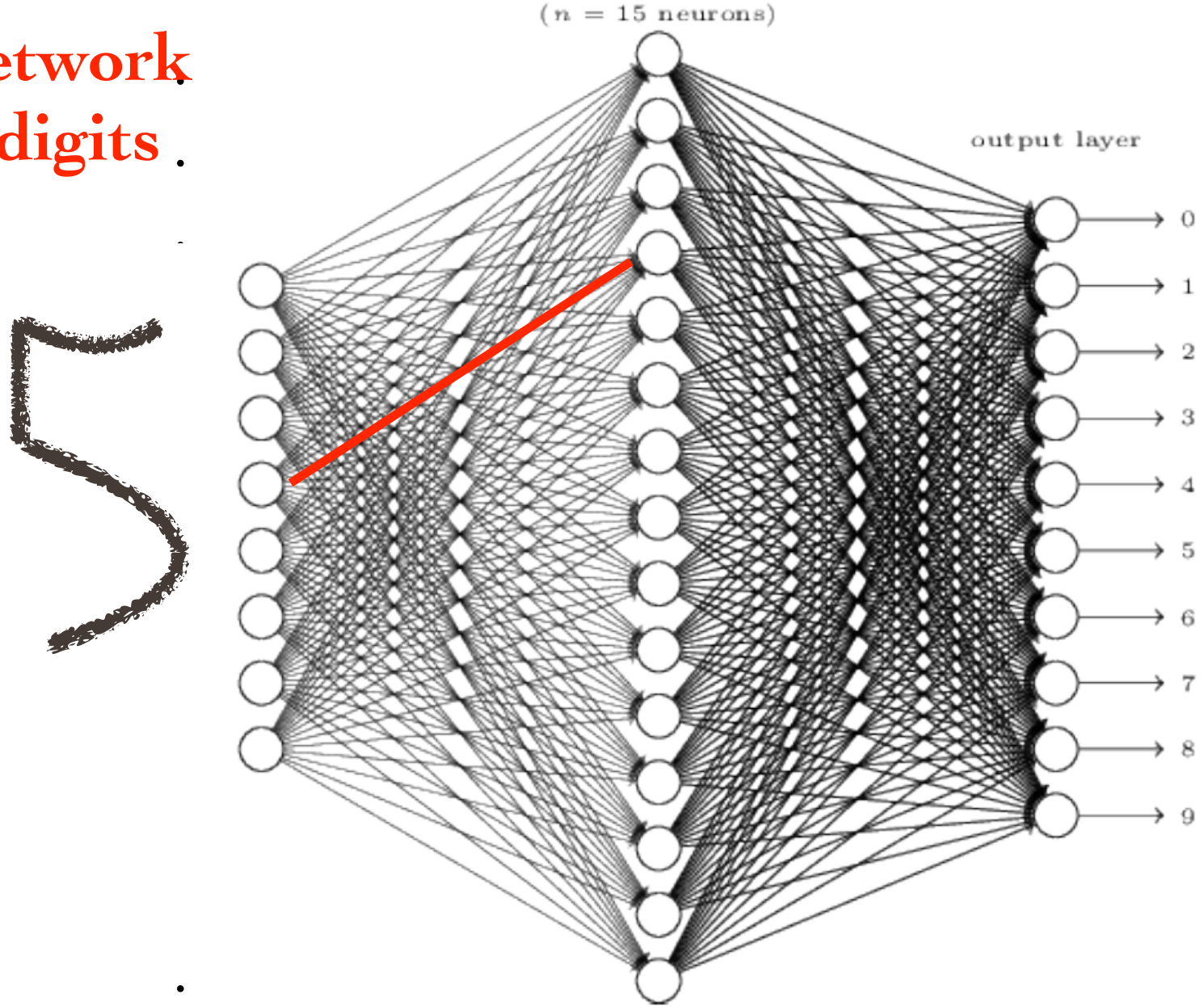
- **Multilayer perceptrons (already in the 80's)**
- **Large databases**
- **Much much more computing power**
- **Several « minor » improvements turned crucial. Details of how each neuron sends its message to other ones**
- **Several layers of neurons : from the perceptron to shallow networks to deep networks**

MNIST database (1998, Le Cun, Cortes, Burges):
70,000 images of handwritten digits

Neural network reading digits

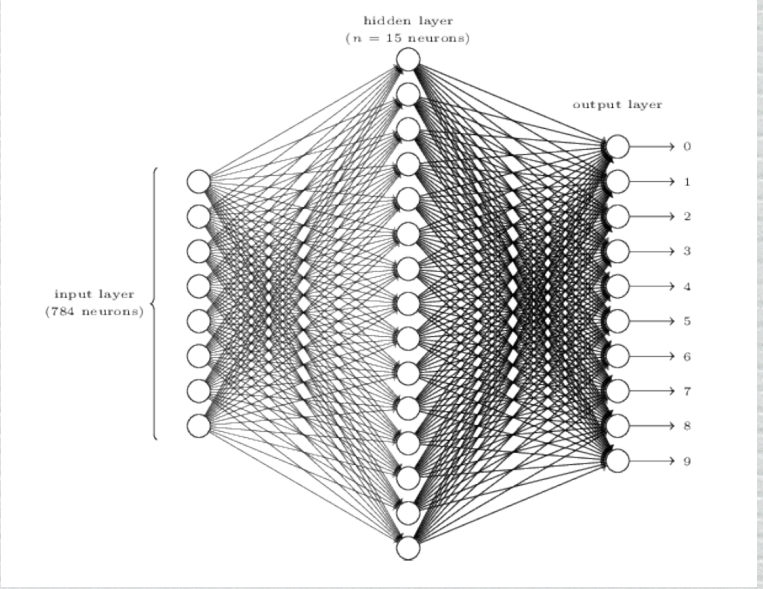


Neural network reading digits



Find the 12000 parameters such that all 60000 digits in the database are well identified !
Stochastic gradient descent

Performance on handwritten digits

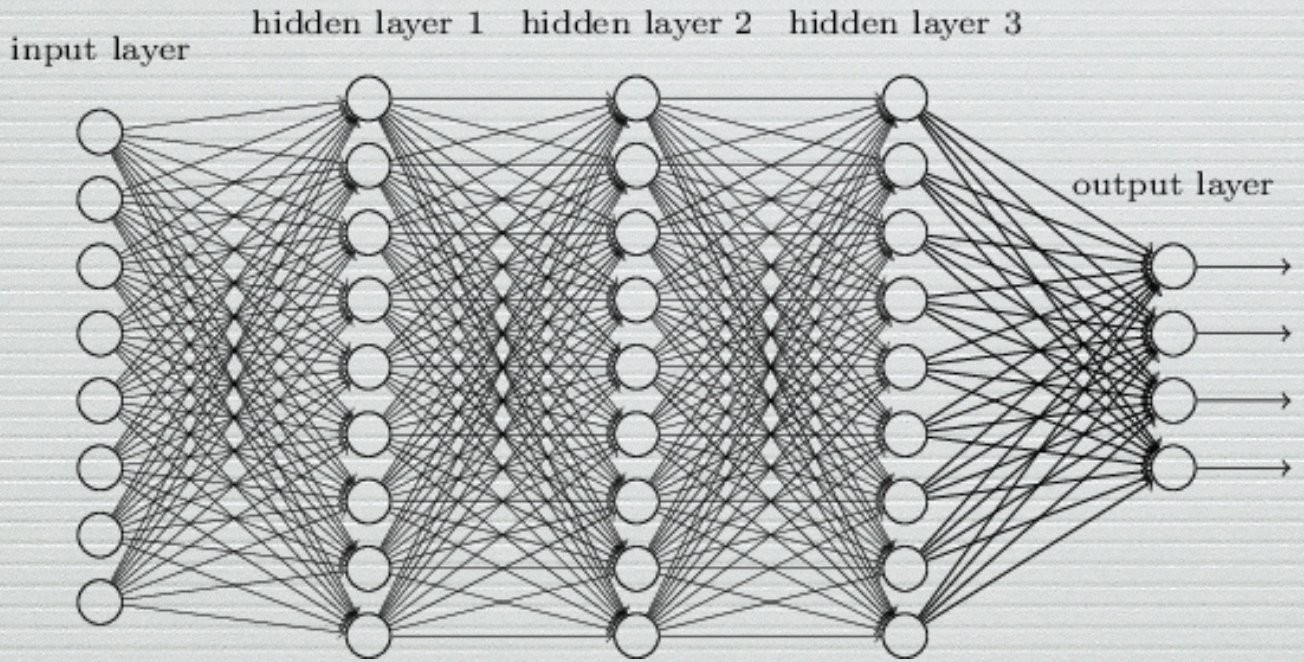


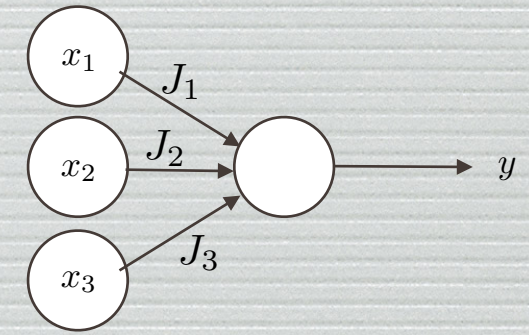
After a (long) training phase on the database: test on remaining 10,000 examples

Simple realization: correct output for 96% of the new images

Better (« deeper ») architecture: correct at 99,8% . 21 « errors » such as:

Deep neural networks

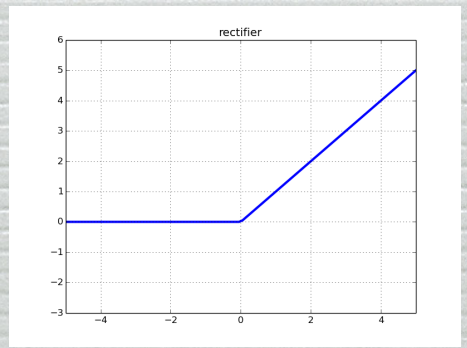




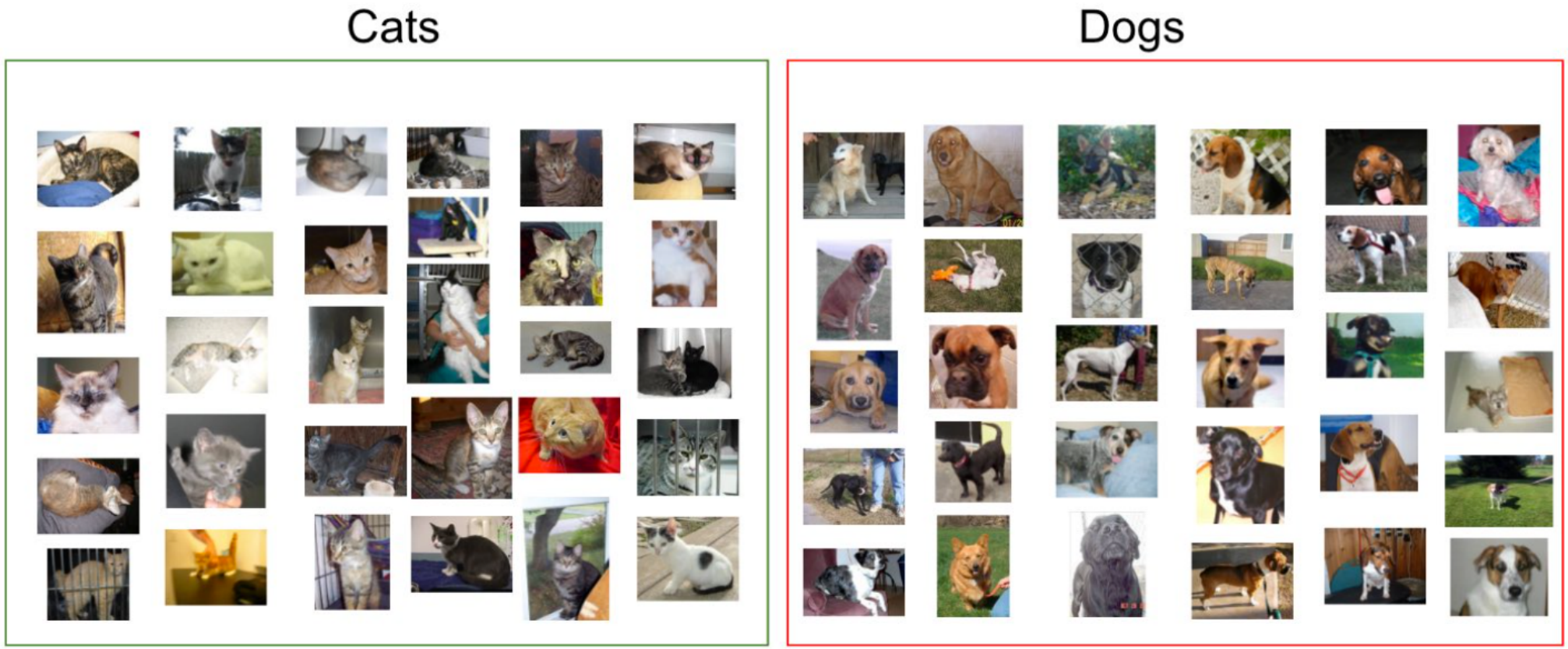
$$y = f(J_0 + J_1 x_1 + J_2 x_2 + J_3 x_3)$$

Can have 100,000 neurons, 100 layers, more than 1,000,000 parameters

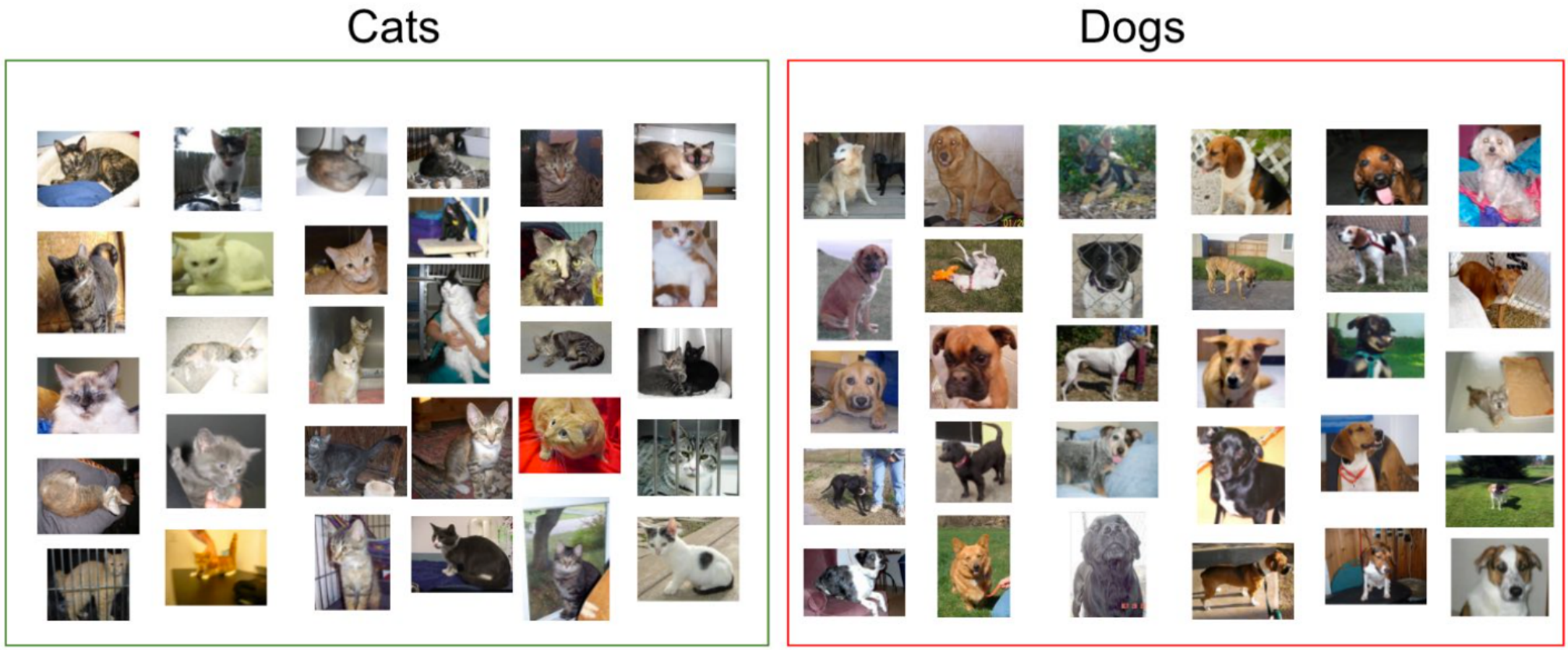
Trained on huge databases



Bigger networks, more parameters. Larger database!

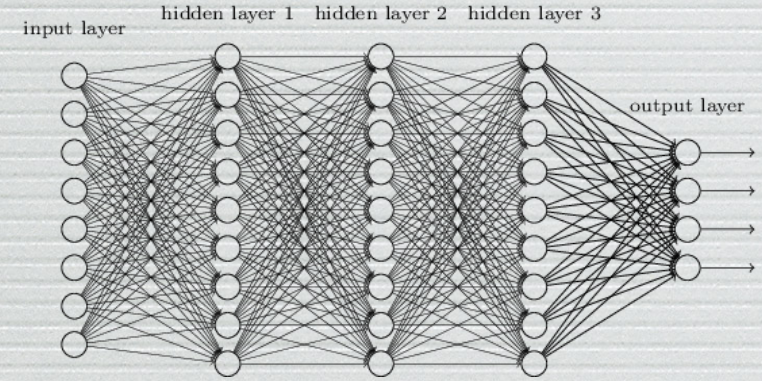


Bigger networks, more parameters. Larger database!



Parenthesis : A couple of puzzles (for the scientists)

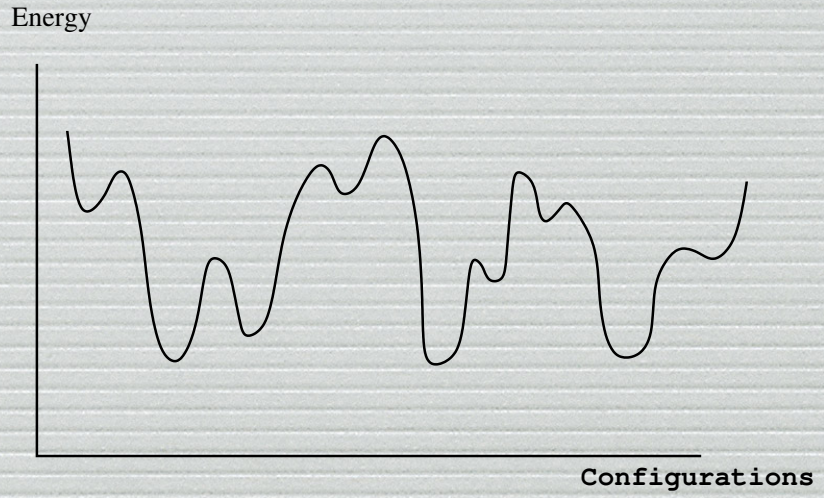
A first puzzle: why does machine learning work?



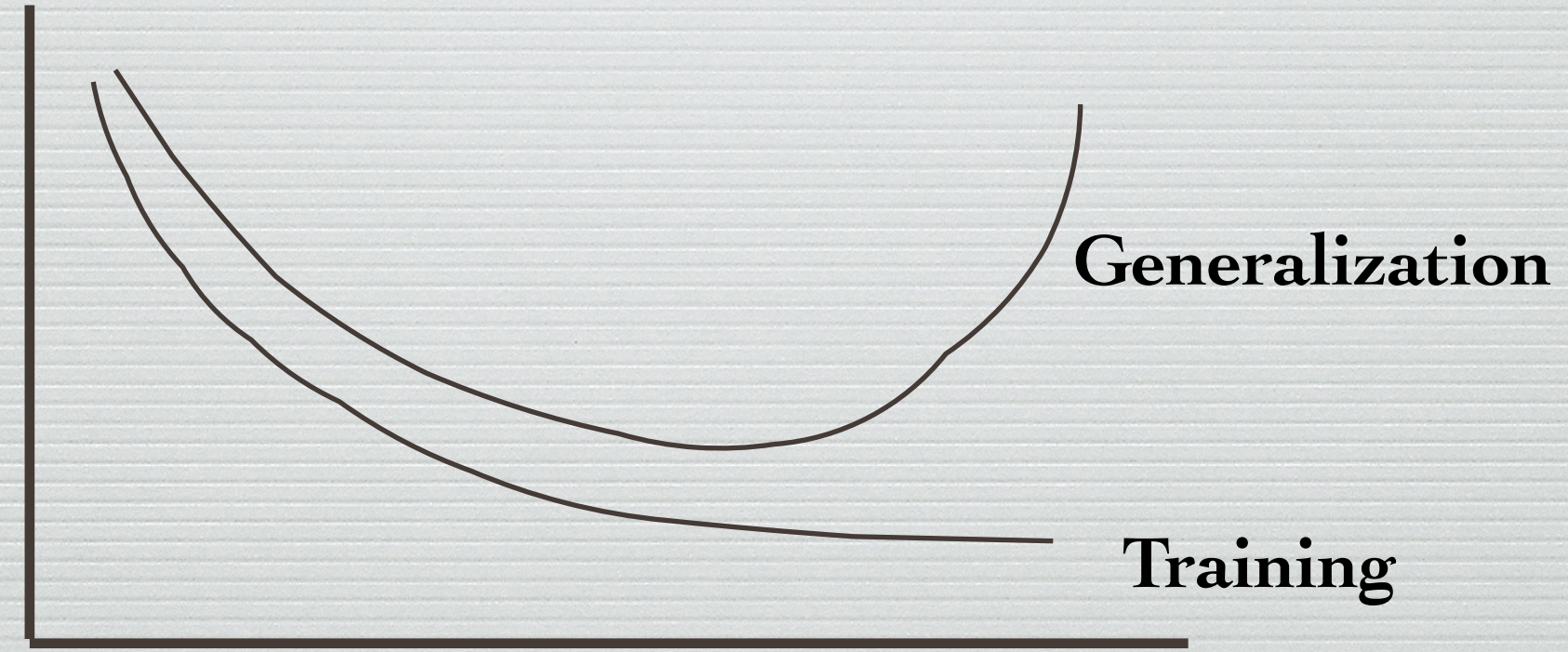
Learning phase: find a set of parameters such that all examples in the training set are correctly classified.
Optimization in a 1,000,000 dimensional space

Landscape?

- **Spin glass ?**
- **Flat regions ?**
- **Many solutions ?**
- **Deep vs shallow networks?**

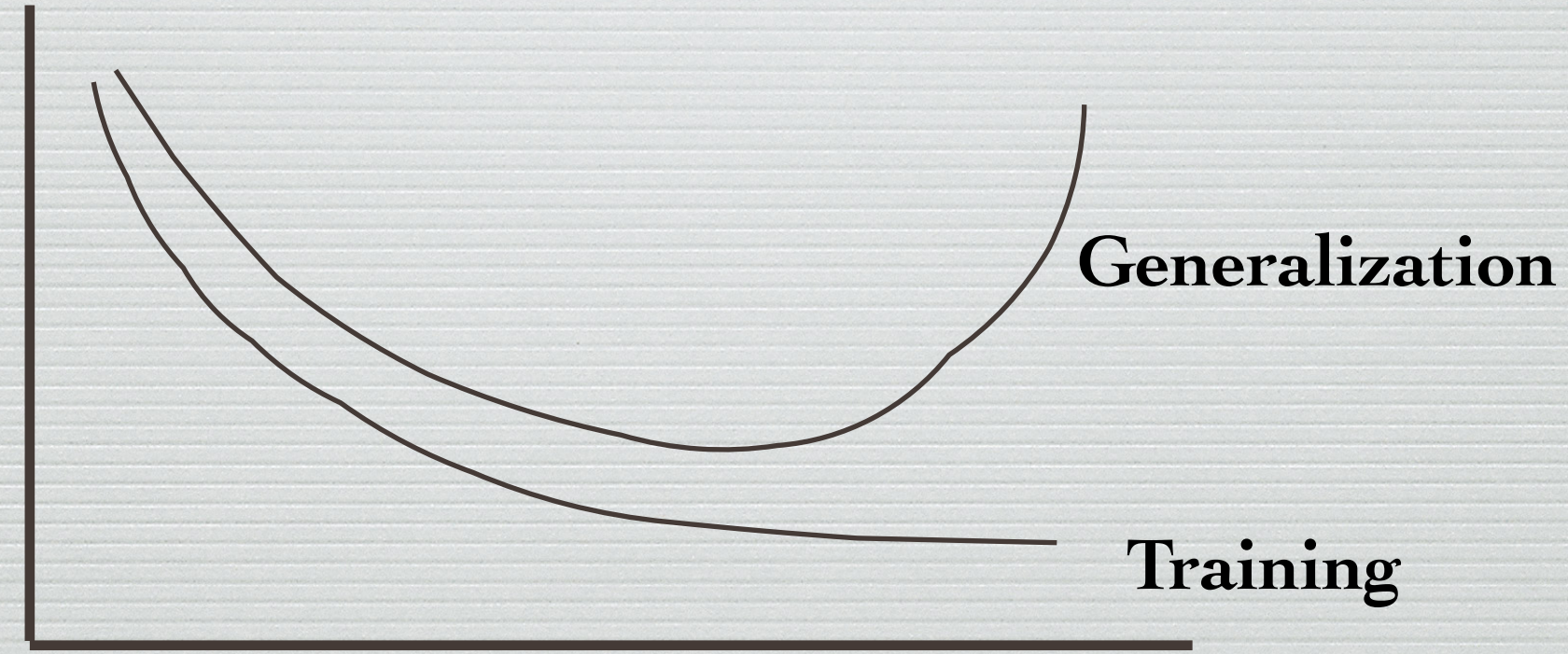


A second puzzle: why doesn't the training « overfit » the data? Why does the network generalize?



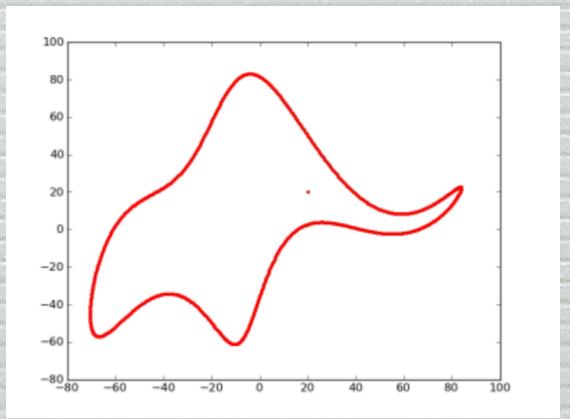
Usual behavior in statistics

A second puzzle: why doesn't the training « overfit » the data? Why does the network generalize?

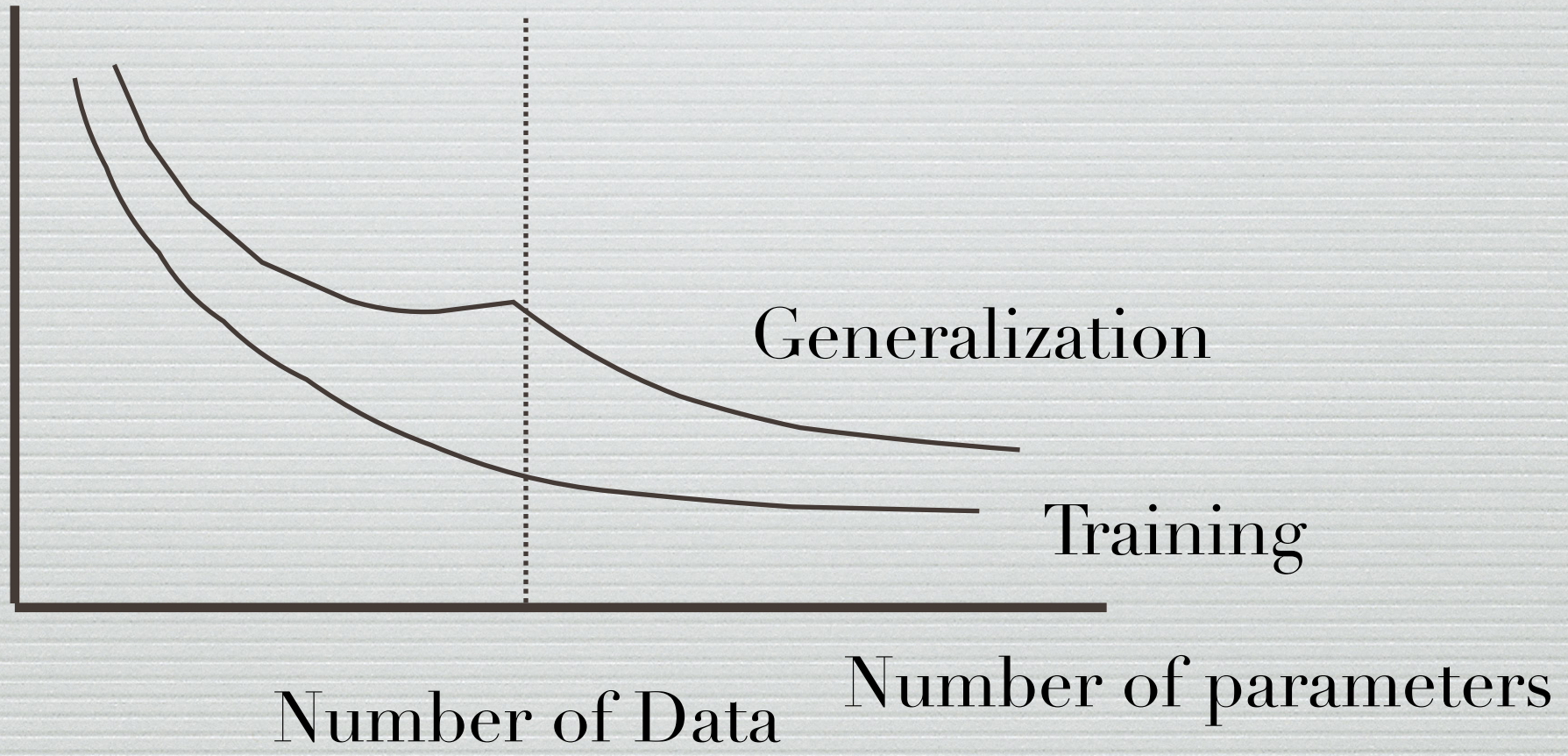


Usual behavior in statistics

With 4 parameters I can fit an elephant (J. von Neumann)

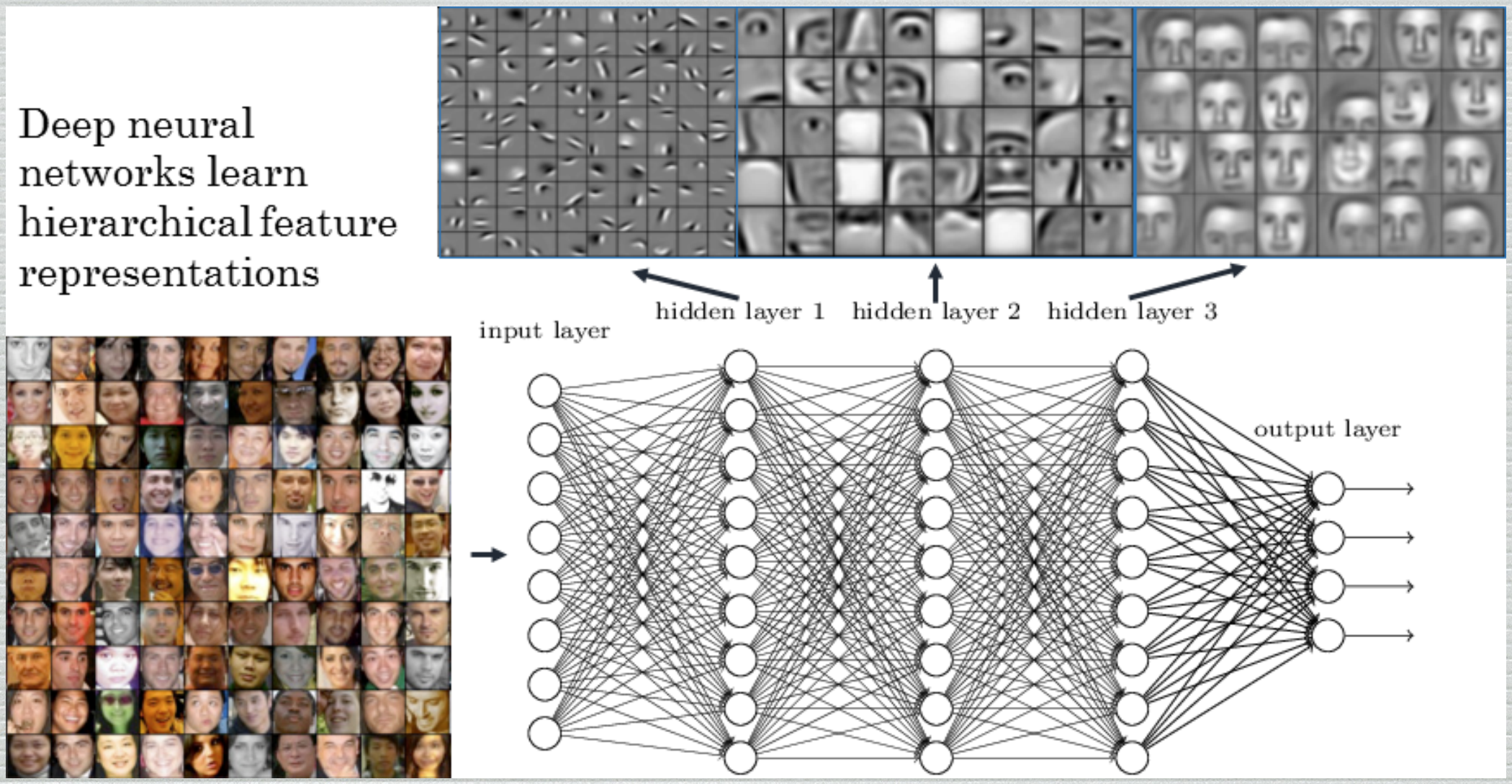


Deep networks



New computing paradigm. Collective representation of information, going to larger scales. Robust.

Deep neural networks learn hierarchical feature representations



Chapter Four

Why deep networks are not (yet?) a panacea

Three main problems:

- **1- Needs very large amount of data, preprocessed, labelled**
- **2- No understanding, no guarantee of results**
- **3- No « general » intelligence, no reasoning, no representation of the world, no consciousness, no attention**

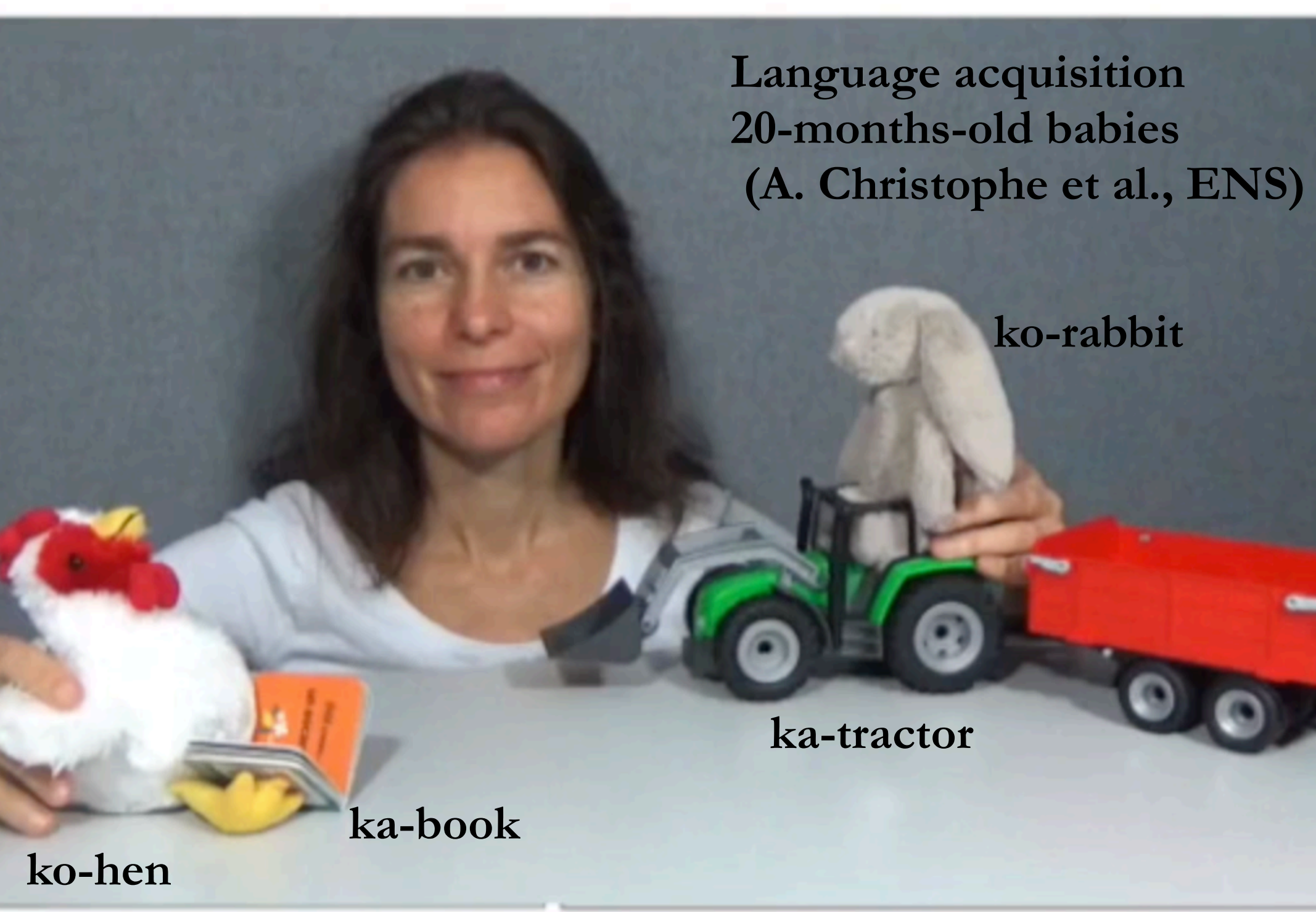
1- Huge amount of labelled data is necessary for learning in deep networks

- This is unpractical from a technological point of view
- This gives a proof that deep networks are still very far from mimicking the brain

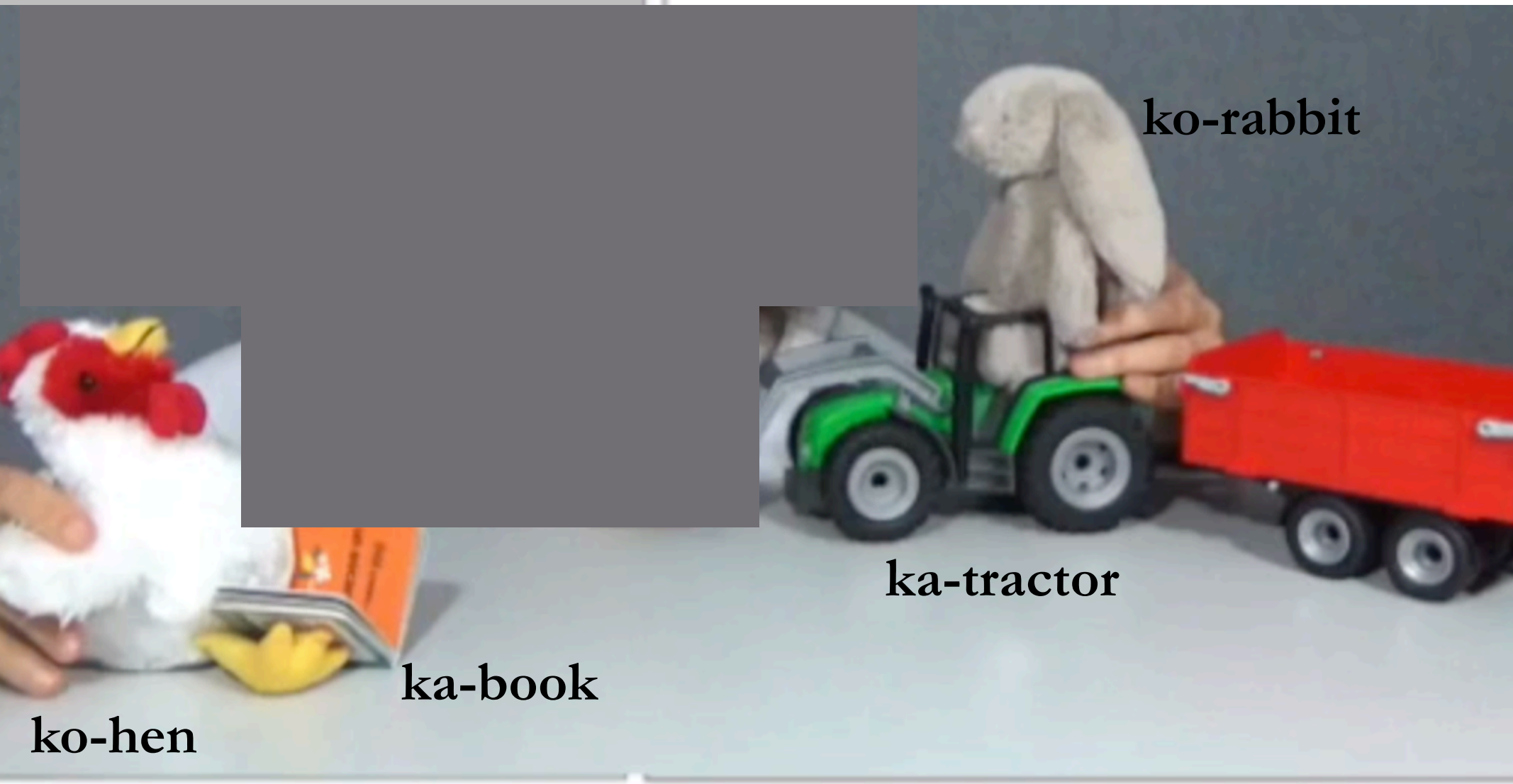
1- Huge amount of labelled data is necessary for learning in deep networks

- This is unpractical from a technological point of view
- This gives a proof that deep networks are still very far from mimicking the brain

How many examples of a cat does a baby see before he builds the mental representation of the concept « cat » ?

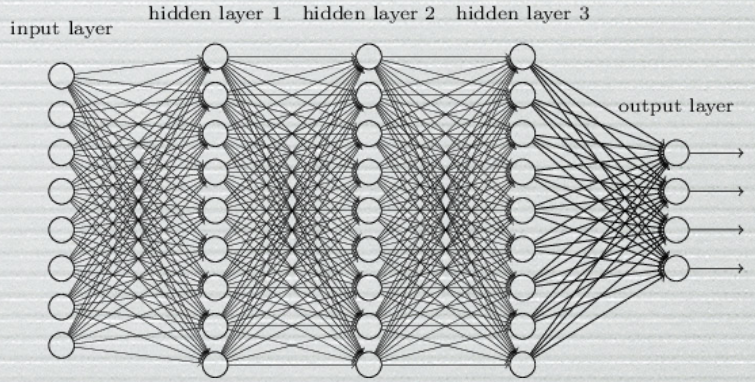

Language acquisition
20-months-old babies
(A. Christophe et al., ENS)
ko-rabbit
ka-tractor
ka-book
ko-hen

Oh, look at **ko bamoule**!
Do you see **ko bamoule**?



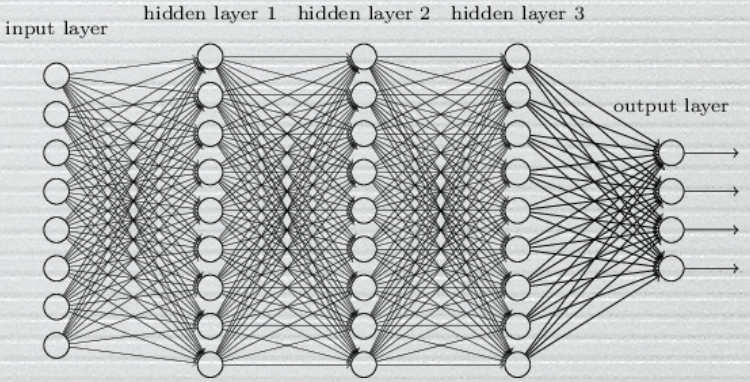
Challenge: build a machine (a neural network?) that learns a language on the basis of what has been heard by a baby in his first year of life

2- What do we know about the deep network, what do we understand?



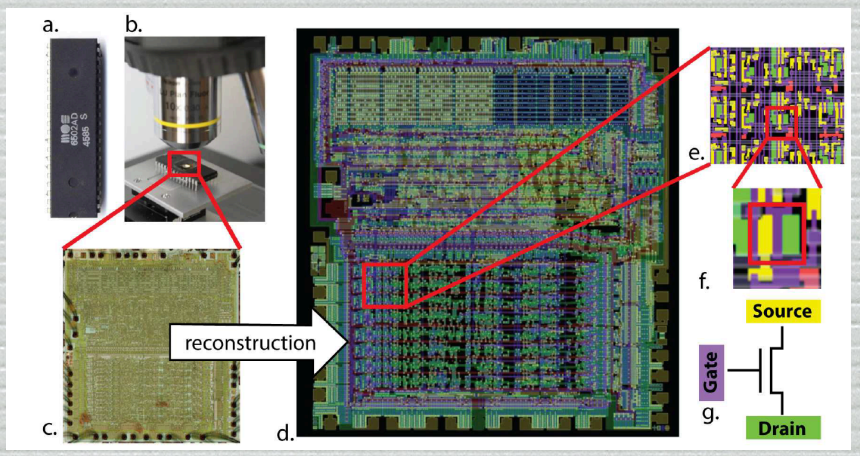
We know everything : all the parameters that have been learnt by the machine are known, all operations of each neuron are known.

2- What do we know about the deep network, what do we understand?

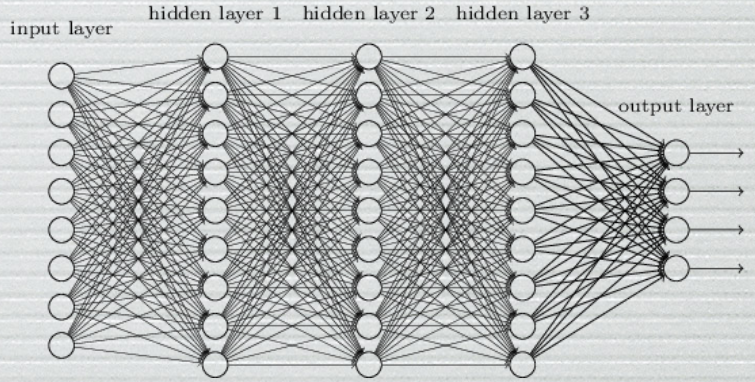


We know everything : all the parameters that have been learnt by the machine are known, all operations of each neuron are known.

The neuroscientist's dream: know the activity of all neurons, of all synapses… (« *Could a Neuroscientist Understand a Microprocessor ?* » E. Jonas and K-P Kording, PLOS 2017**)**

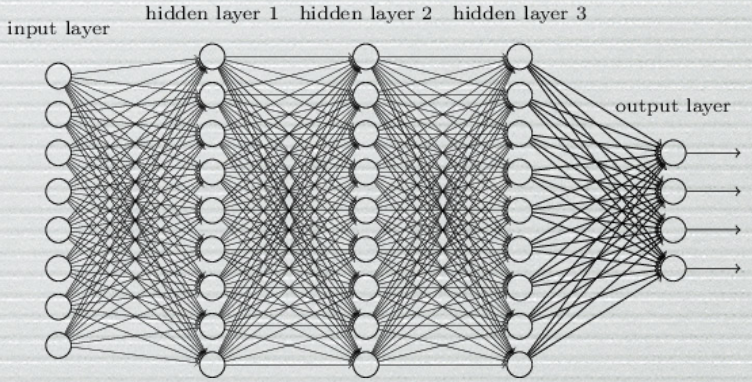


2- What do we know about the deep network, what do we understand?



We know everything : all the parameters that have been learnt by the machine are known, all operations of each neuron are known.

2- What do we know about the deep network, what do we understand?

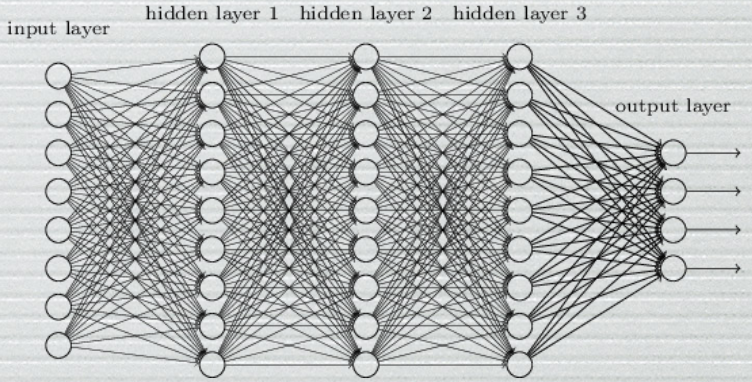


We know everything : all the parameters that have been learnt by the machine are known, all operations of each neuron are known.

We understand very little.

- Learning mechanism is poorly understood
- Collective processing of information? « Emergence » (the whole contains more information than the sum of each part)
- No way to explain the decisions made by the machine

2- What do we know about the deep network, what do we understand?



We know everything : all the parameters that have been learnt by the machine are known, all operations of each neuron are known.

We understand very little.

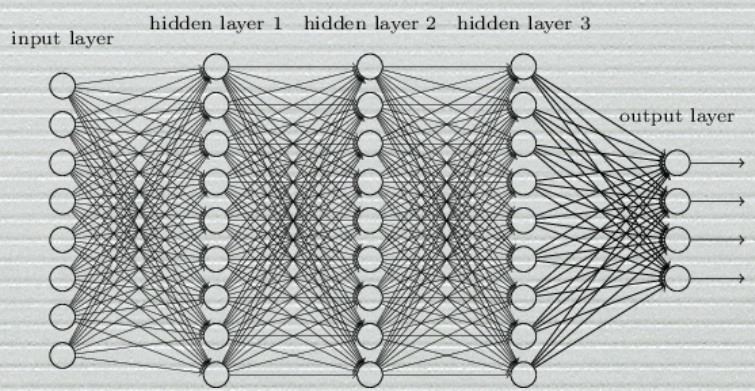
- Learning mechanism is poorly understood
- Collective processing of information? « Emergence » (the whole contains more information than the sum of each part)
- No way to explain the decisions made by the machine

No guarantee that it works in all circumstances

Panda

Gibbon

Classification : Pandas vs Gibbons



Panda

Gibbon

Panda

Gibbon

"panda"
57.7% confidence

Panda

Gibbon

“panda”
57.7% confidence

Panda

Gibbon

“panda”
57.7% confidence

“gibbon”
99.3 % confidence

Panda

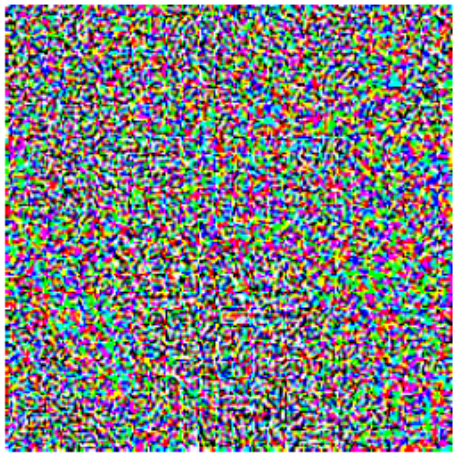
Gibbon

Ian J. Goodfellow et al. 2015

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"gibbon"
99.3 % confidence

Panda

Gibbon

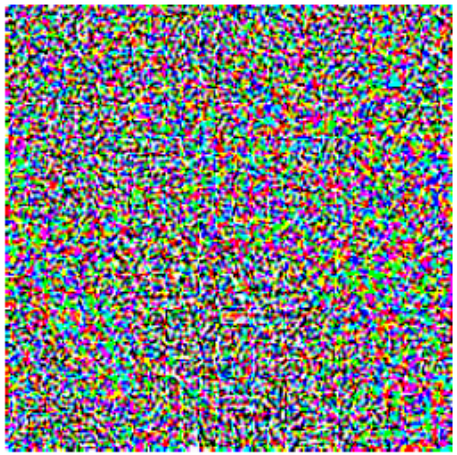
Ian J. Goodfellow et al. 2015

Seen as a gibbon

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



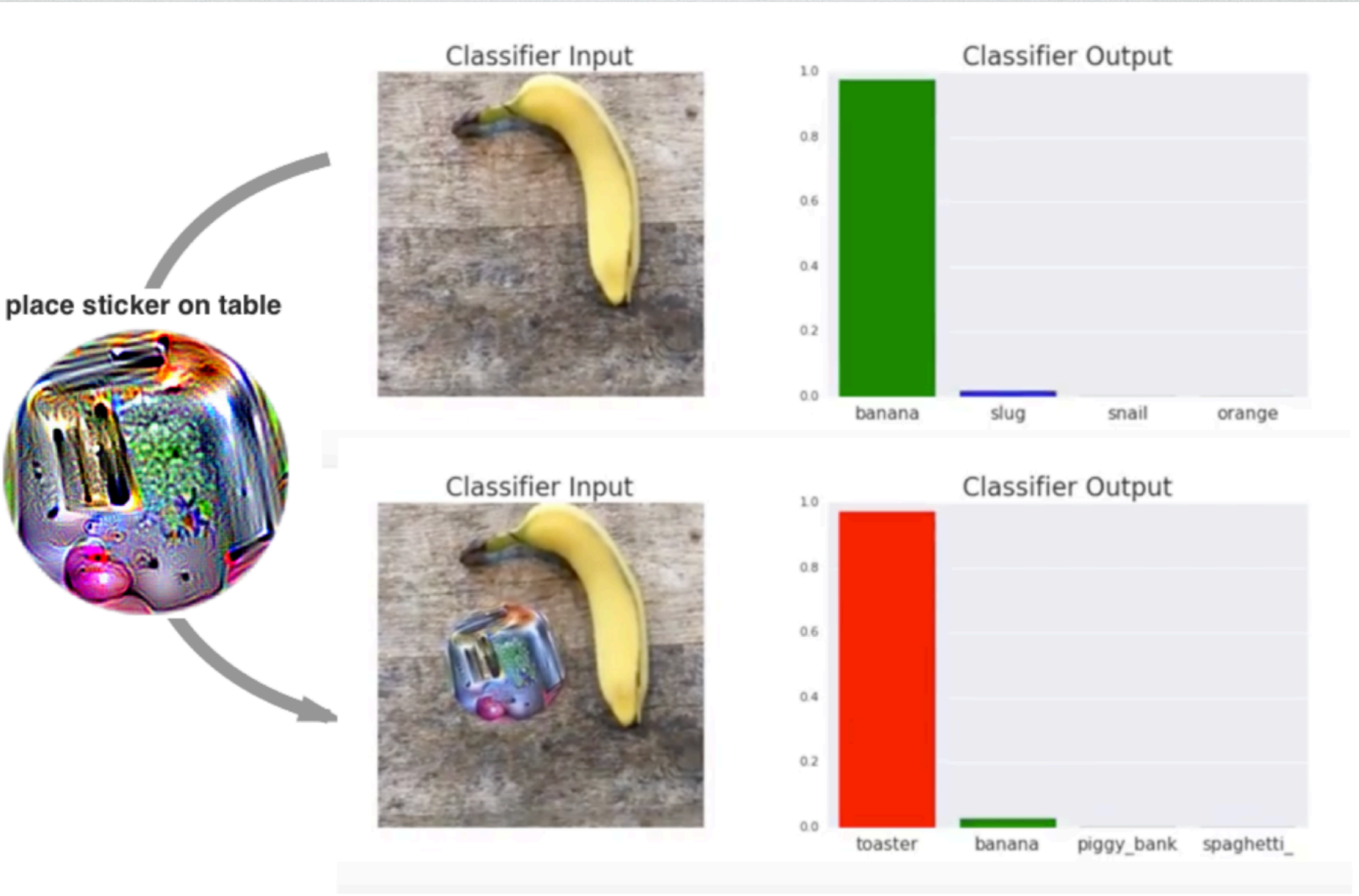
$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

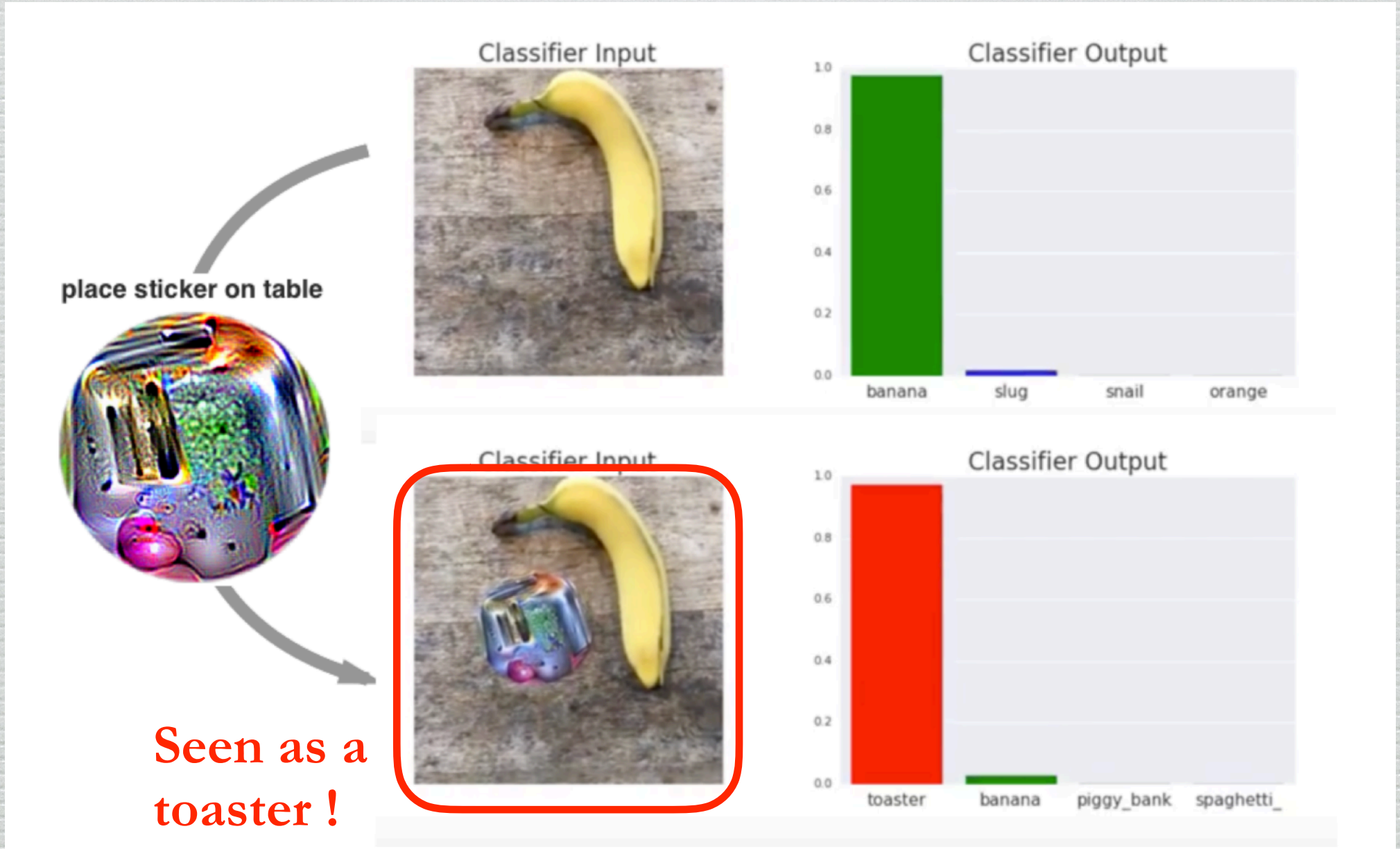
$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"gibbon"
99.3 % confidence



T.B. Brown et al, 2018



T.B. Brown et al, 2018

3- No « general » intelligence, no reasoning, no representation of the world, no consciousness, no attention

3- No « general » intelligence, no reasoning, no representation of the world, no consciousness, no attention

Deep networks = Machines that « solve » very specific problems, well posed, well defined, with a very simple measure of performance.

Chapter Five

About (scientific) Intelligence

Chris Anderson, Chief Editor of « Wired », 2008

The end of Theory: The data deluge makes the scientific method obsolete

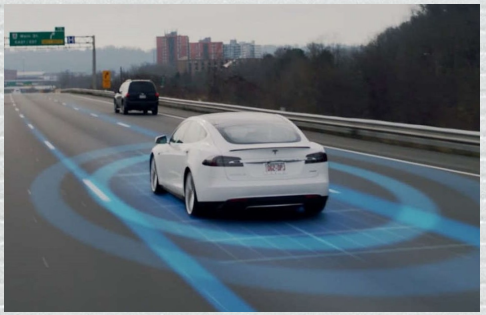
Faced with massive data, **the [traditional] approach to science — hypothesize, model, test — is becoming obsolete**.

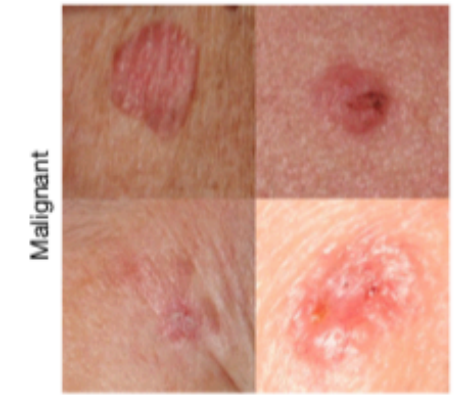
Chris Anderson, Chief Editor of « Wired », 2008

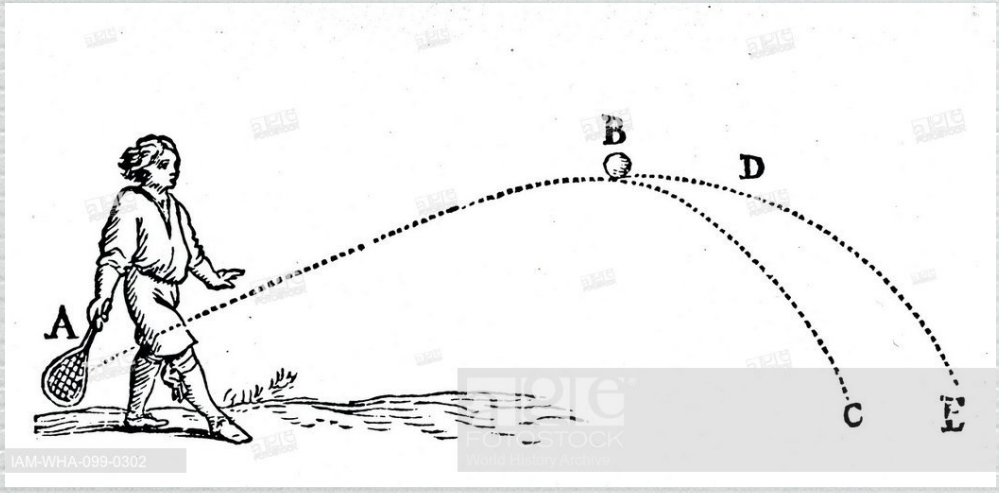
The end of Theory: The data deluge makes the scientific method obsolete

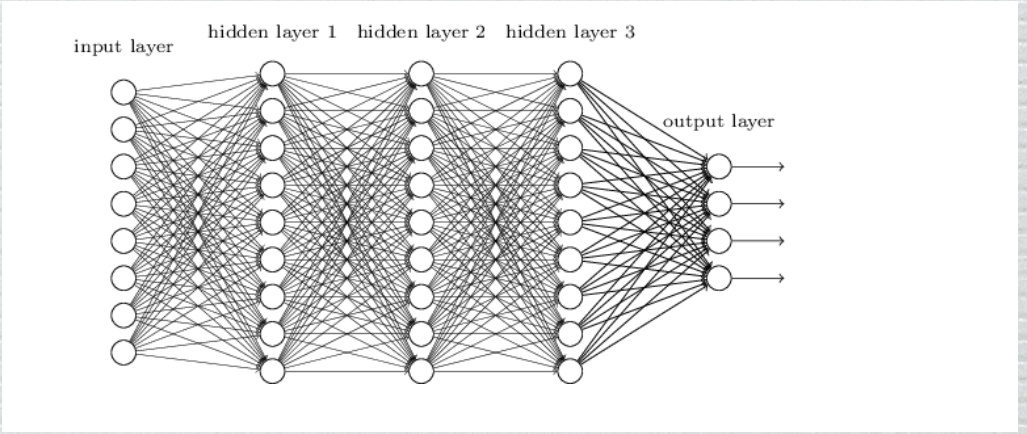
Faced with massive data, **the [traditional] approach to science — hypothesize, model, test — is becoming obsolete**.

The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. **Correlation supersedes causation, and science can advance even without coherent models, unified theories**, or really any mechanistic explanation at all.




Malignant



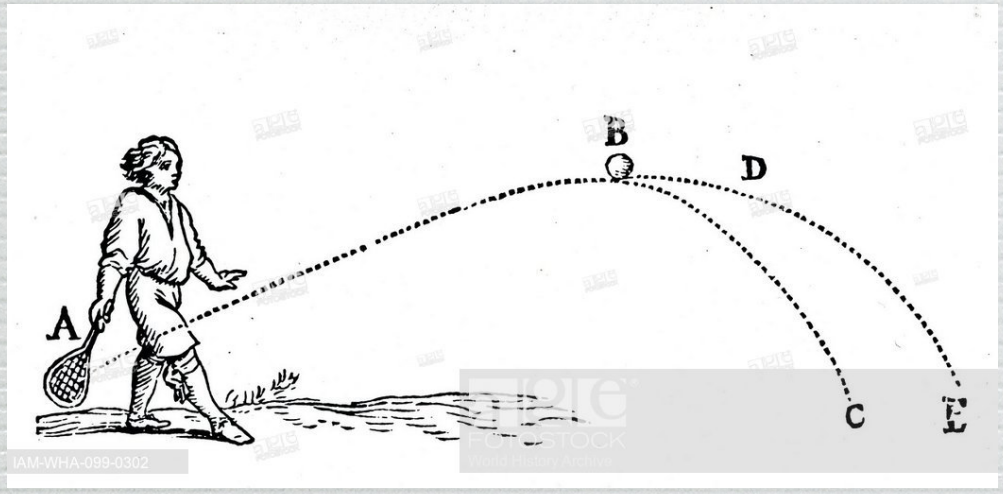


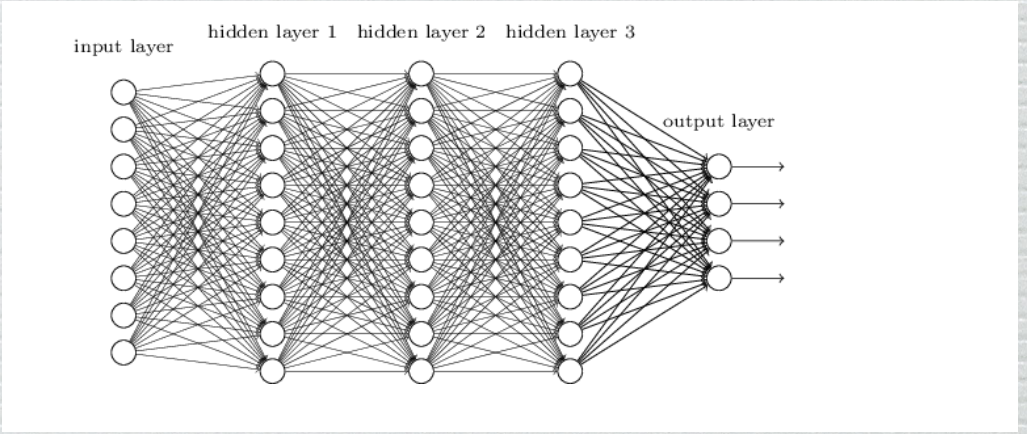
Thought experiment : feed a neural net with tens of thousands of experimental trajectories of objects thrown into the air

Goal: where does it fall back.

Probably, a well-trained deep network will give good predictions, based on the object mass, initial velocity, initial rotation, shape,…

Maybe as good as solving Newton's equations?





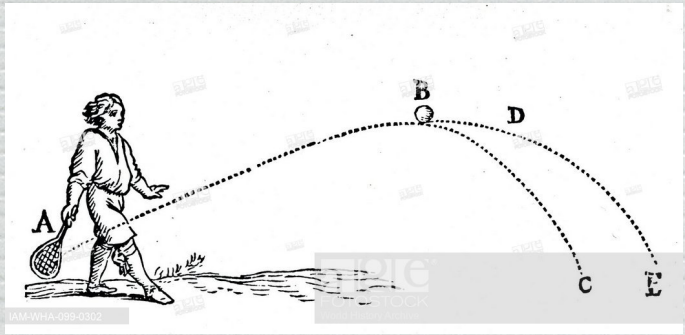
Thought experiment : feed a neural net with tens of thousands of experimental trajectories of objects thrown into the air

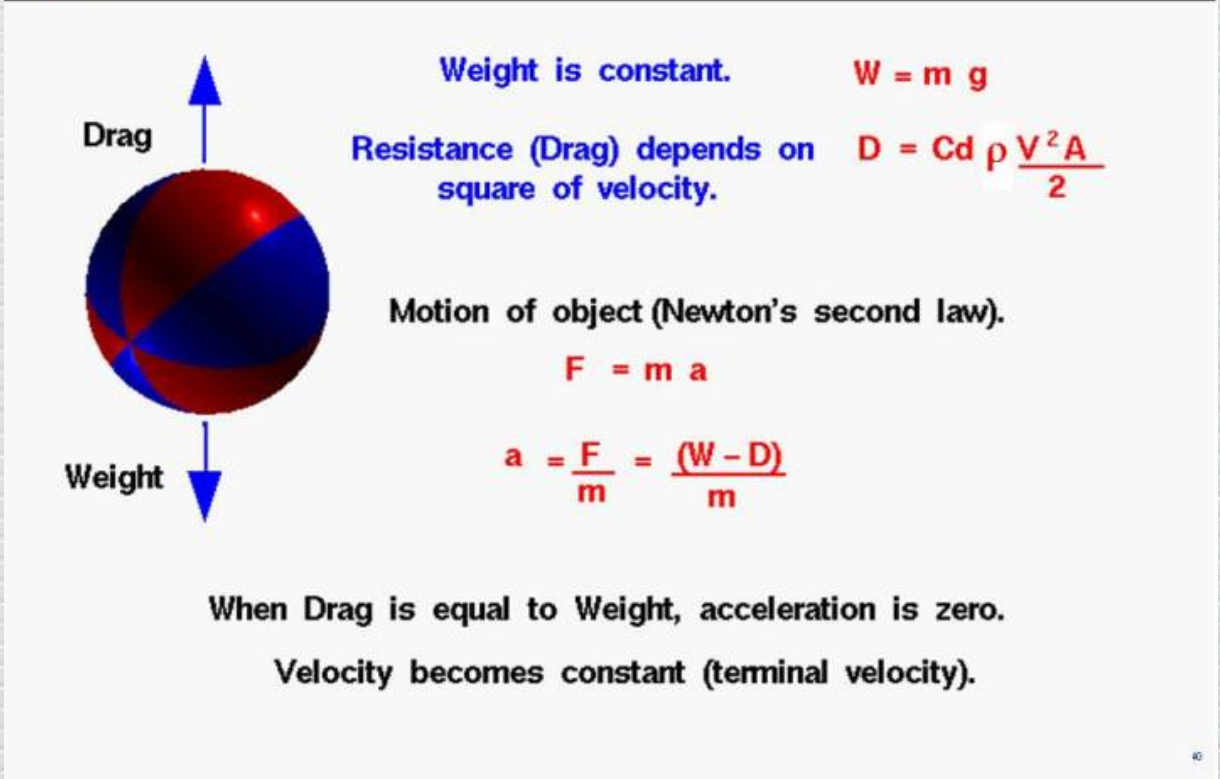
Goal: where does it fall back.

Probably, a well-trained deep network will give good predictions, based on the object mass, initial velocity, initial rotation, shape,…

Maybe as good as solving Newton's equations?

Which one is the best? The deep network or the physicist?





Scientific model has a compact representation, which can be decomposed: gravity, then add the effect of friction, then add the Magnus force

Scientific intelligence, expressed in terms of models and equations: composition of elementary laws, applicable in different contexts

Same gravity law applies to movement of planets

Infinitely richer than simply predicting a trajectory

Deep networks have no reasoning, no representation of the world, no notion of causality, no consciousness, no attention, no possibility to apply knowledge in different context, or to combine it with other information

We are still very very far from General Artificial Intelligence

Conclusion

So, what is going to happen?

Predicting (the future)

1889: « Fooling around with alternating current (AC) is just a waste of time. Nobody will use it, ever. » — Thomas Edison

1943: « I think there is a world market for maybe five computers. »
Thomas Watson, president of IBM, 1943

1977: « There is no reason anyone would want a computer in their home. »
Ken Olsen, founder of Digital Equipment Corporation

2004: »Two years from now, spam will be solved."
Bill Gates, founder of Microsoft, 2004

2007: "There's no chance that the iPhone is going to get any significant market share." — Steve Ballmer, Microsoft CEO.

Predicting the future

Image analysis, language processing, recommendation systems, etc. Very specific problems, well posed, with a simple measure of performance.

- Impressive progress of AI, machine learning and deep networks
- Ability to detect subtle patterns in massive amounts of data
- Major technological breakthroughs
- Global evolution, extremely fast

It will have a significant impact on a large number of human activities and transform many jobs.
Beyond physical labor: intellectual labor. Well defined (and repetitive tasks)

Predicting the future (?)

Positive view: help to achieve better :

- diagnosis in medicine
- case-law search
- displacement of people and goods along railways, freeways, etc
- devices that help people to have fast access to relevant information
- customer support
- robots that can help, for example, elderly people
- identification of pathogens, development of new drugs
- smart language assistants

Big transformation in many jobs, but more economic activity will be created than destroyed

Predicting the future (?)

A more concerned point of view: large destruction of jobs, on a time scale too fast for the society to be able to organise and adapt to the new situation. Essentially monopolistic.

- diagnosis in medicine
- case-law search
- displacement of people and goods along railways, freeways, etc
- devices that help people to have fast access to relevant information
- customer support
- robots that can help, for example, elderly people
- identification of pathogens, development of new drugs
- smart language assistants

A major concern for the present:

AI and the political world.
Control of populations, manipulation of information

"I want to call attention to the mortal danger facing open societies from the instruments of control that machine learning and artificial intelligence can put in the hands of repressive regimes". **George Soros, Davos, January 2019**

[in China,], « All the rapidly expanding information available about a person is going to be consolidated in a centralized database to create a 'social credit system'.
"Based on that data, people will be evaluated by algorithms that will determine whether they pose a threat to the one-party state. »

In 2018:

- Malaysia equipped the police forces with face-recognition devices
- Singapore's experimental program to equip every lamppost with a camera connected to a facial-recognition facility, and a « crowd analytics » software
- Zimbabwe signed a deal with a Chinese company to build a national image data ware

S. Feldstein, Journal of Democracy 30 (2019) vol 1

London's police department said on Friday that it would begin using facial recognition to spot criminal suspects with video cameras as they walk the streets

NYT, January 27, 2020

In the last five years, many examples of « smart » political manipulation of information

In 2018:

- Malaysia equipped the police forces with face-recognition devices
- Singapore's experimental program to equip every lamppost with a camera connected to a facial-recognition facility, and a « crowd analytics » software
- Zimbabwe signed a deal with a Chinese company to build a national image data ware

S. Feldstein, Journal of Democracy 30 (2019) vol 1

London's police department said on Friday that it would begin using facial recognition to spot criminal suspects with video cameras as they walk the streets

NYT, January 27, 2020

In the last five years, many examples of « smart » political manipulation of information

There is an urgent need for control mechanisms, for the elaboration of ethical rules, for the development of a global vision on the possible impacts on our societies

In 2018:

- Malaysia equipped the police forces with face-recognition devices
- Singapore's experimental program to equip every lamppost with a camera connected to a facial-recognition facility, and a « crowd analytics » software
- Zimbabwe signed a deal with a Chinese company to build a national image data ware

S. Feldstein, Journal of Democracy 30 (2019) vol 1

London's police department said on Friday that it would begin using facial recognition to spot criminal suspects with video cameras as they walk the streets

NYT, January 27, 2020

In the last five years, many examples of « smart » political manipulation of information

There is an urgent need for control mechanisms, for the elaboration of ethical rules, for the development of a global vision on the possible impacts on our societies

« *Philosophically, intellectually -in every way- human society is unprepared for the rise of artificial intelligence* ». H Kissinger, « How the Enlightenment Ends », The Atlantic- June 2018

Take-home messages

- New developments in AI using machine learning and deep networks are making spectacular progress in well defined tasks with a clear objective
- They can help humans in sophisticated repetitive tasks (see e.g. medical diagnosis)
- They can also be mis-used. Awareness of the dangers, regulations, ethical supervision are absolutely necessary
- Still, we are extremely far from general artificial intelligence. actually we have not made any progress in that direction
- We know everything of what these machines do, we understand nothing…

The End