Disentangling deep learning and IP rights

Jean-Marc Deltorn

Centre for International Intellectual Property Studies (CEIPI) – University of Strasbourg, France

Vereniging voor Auteursrecht

Amsterdam 15-06-2018

The views expressed in this presentation are solely those of the author and do not necessarily reflect the position of any other organization or employer.


The Perils of Special Counsels / This Is You
Newswe
THE DOCTOR WILL SEE YOU NOW
HOW AI IS GOING TO CURE OUR SICK HEALTH CARE SYSTEM
PRICE $8.99
THE NEW YORKER
OCT. 23, 2017
INSIDE: A 14-PAGE SPECIAL REPORT ON FINANCIAL
The Economist
How to fix America's
The self-service eco
Time to open up Ind
Inside the anti-brib
Why humans cause
Artificial Intellig
The promise and the peril
TIME
SPECIAL EDITION
Artificial Intelligence
The Future of Humankind

GARDEN LEAF BLOWER SAVE £50 FOR £29.99 SEE PAGE 24

SUNDAY EXPRESS

WITH 5 MAGAZINE, SPORT, FINANCE, REVIEW & TRAVEL NOVEMBER 19, 2017 £1.40

FREE ROYAL MINT COIN FOR EVERY READER

PAGE 49

NOW 40p CHEAPER THAN THE MAIL ON SUNDAY

PLUS The Queen and Prince Philip 70th anniversary photograph: PAGE 3 & 5 MAGAZINE SPECIAL

DRIVERLESS CARS BY 2021

Budget pledge to spend millions on making UK roads fit for the future

By Camilla Tominey and Nick Lester

DRIVERLESS cars will be on the roads within three years as part of a multimillion pound plan to help Britain lead a technological revolution post-Brexit.

The Chancellor will on Wednesday outline Budget plans to "turbo-charge" Britain's tech industries – pumping investment into artificial intelligence, clean fuel technology and next generation 5G phones.

Funding will also go on more computer science teachers and technology skills training.

A source close to Philip Hammond said he wanted to Brexit-proof Britain and ensure the workforce of the future will be able to "enjoy the upside" of the opportunities ahead.

He is expected to announce £75million for developing artificial intelligence, hold reverse-lay-on-road testing to put driverless cars on Britain's highways by 2021 and

TURN TO PAGE 2

Police 'find Gaia's body'

DISCOVERY ENDS TORMENT FOR GIRL'S FAMILY. SEE PAGE 7

EU must give us deal or go bust

THE EU will be "insolvent" if Britain leaves without a trade deal, a leading Brexiteer has predicted, writes Camilla Tominey.

As pressure mounted on the Prime Minister to break the deadlock in Brexit talks by handing billions to Brussels, Jacob Rees-Mogg urged Theresa May to hold firm.

He should not be "blackmailed" into paying a huge divorce bill to the bloc he said.

Mr Rees-Mogg insisted that the UK's negotiating position was much stronger than the EU's.

He said: "If we say that we are not continuing to contribute without a deal to the last 21 months of the multi-annual financial framework then the EU has a huge hole in its budget – it has no legal ability to borrow and it is effectively insolvent for that period.

"So I would say to people who say we're not going to get on to the second stage, well that's fine, but then you haven't got any money for 21 months. Where are you going to get it from? In

TURN TO PAGE 4

DAILY STAR OF SCOTLAND

20p CHEAPER THAN THE SUN...

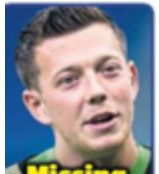
TUESDAY, AUGUST 22, 2017 TOTAL ECLIPSE OF THE SUN JUST 30s

Bake Off back & rude as ever P4

GOT HIM

Van nut shot dead

PAGE 7



Missing McGregor baffles Brendan

BHOYS BOSS CAN'T UNDERSTAND FAILURE TO MAKE SCOTLAND SQUAD

SEE STARSPORT

Sarah in Slutgate CBB storm

PAGES 10-11

WAR FEAR OVER RISE OF THE KILLER ROBOT

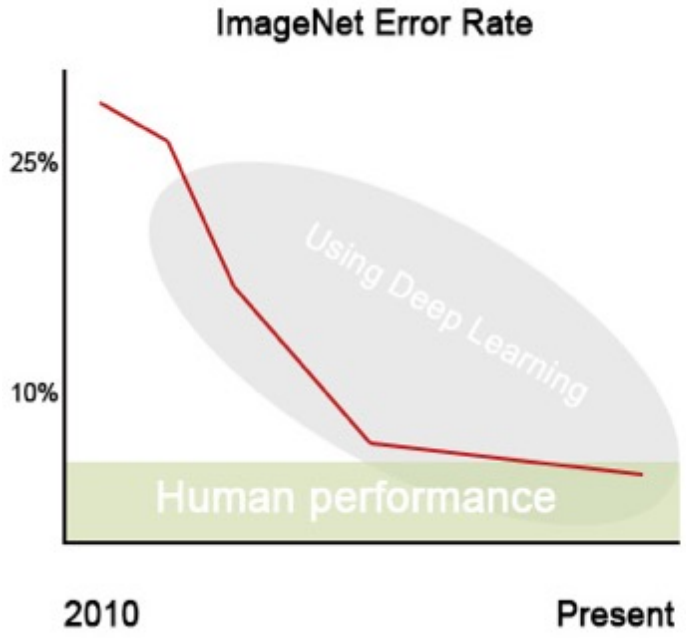
KILLER robots will hunt mankind to extinction in the future unless their development is halted right now, the world's leading experts have warned.

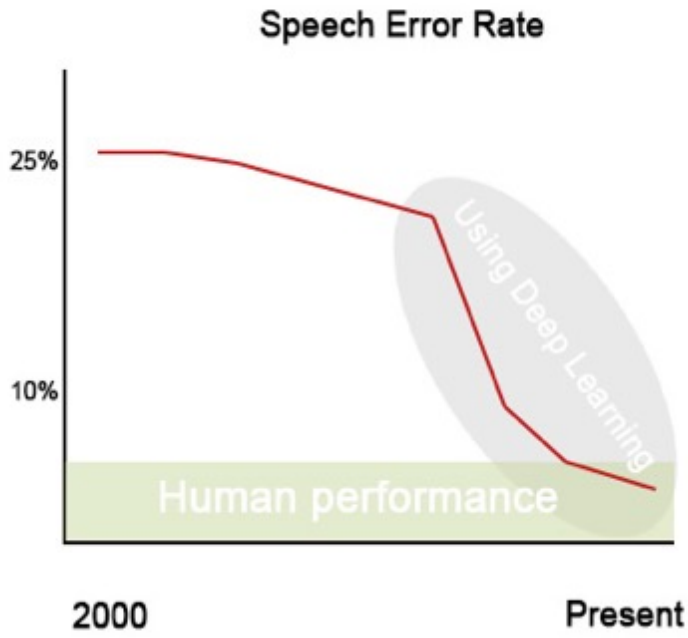
They have called on the United Nations to stop this new arms race, warning: "We do not have long to act."

Full story: Page 5

Experts plead for action

The AI disruption:
Reaching task-specific human performance





“machines behaving in ways that would be called intelligent if a human were so behaving”

John McCarthy, 1955

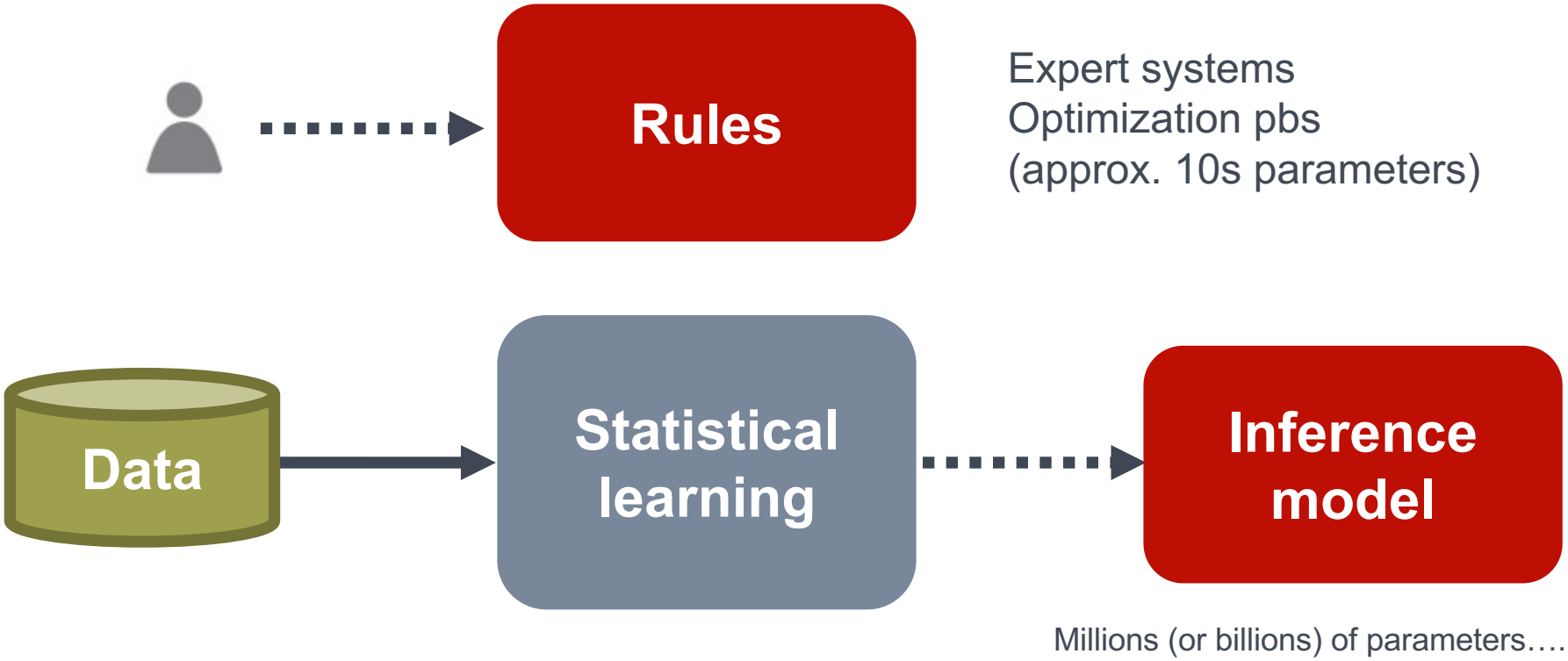
Why now?

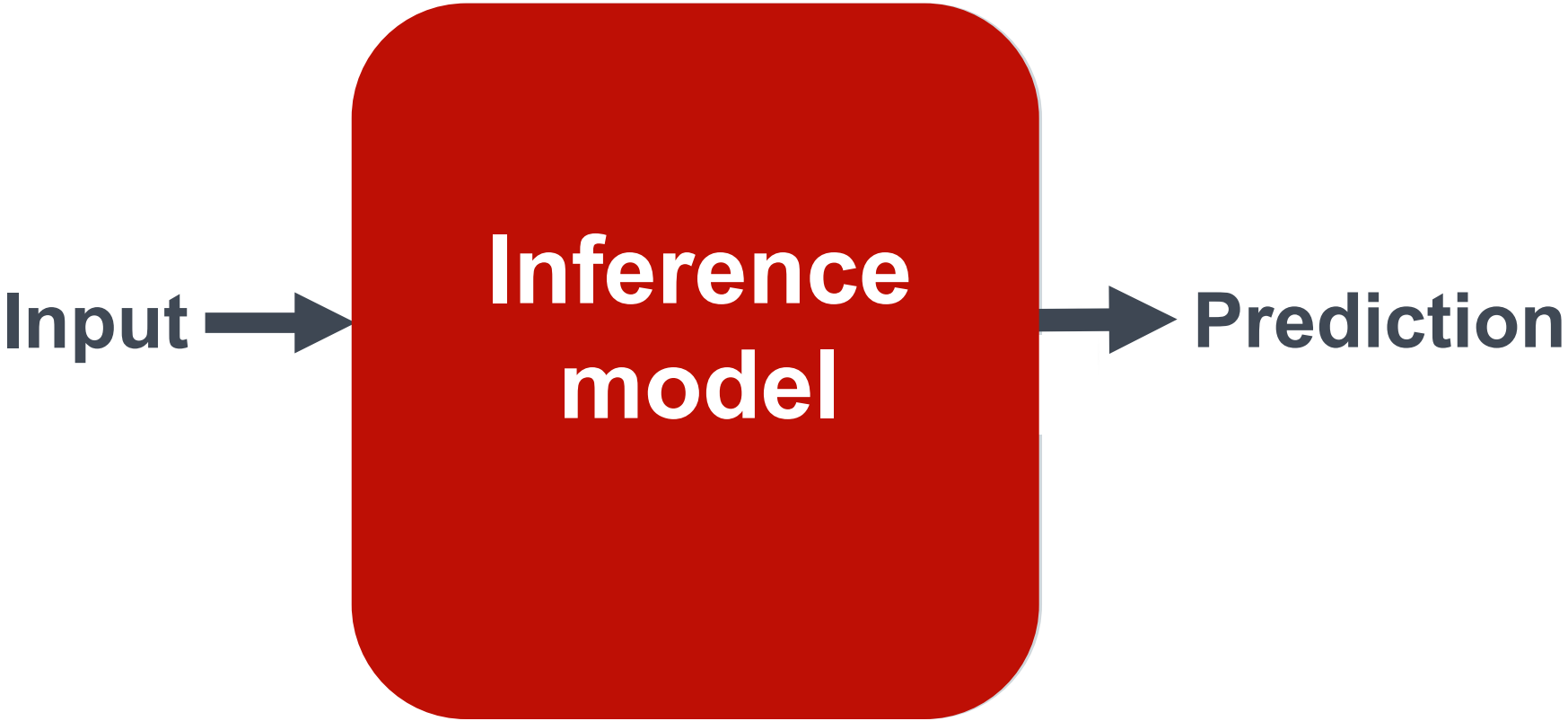
Why now?

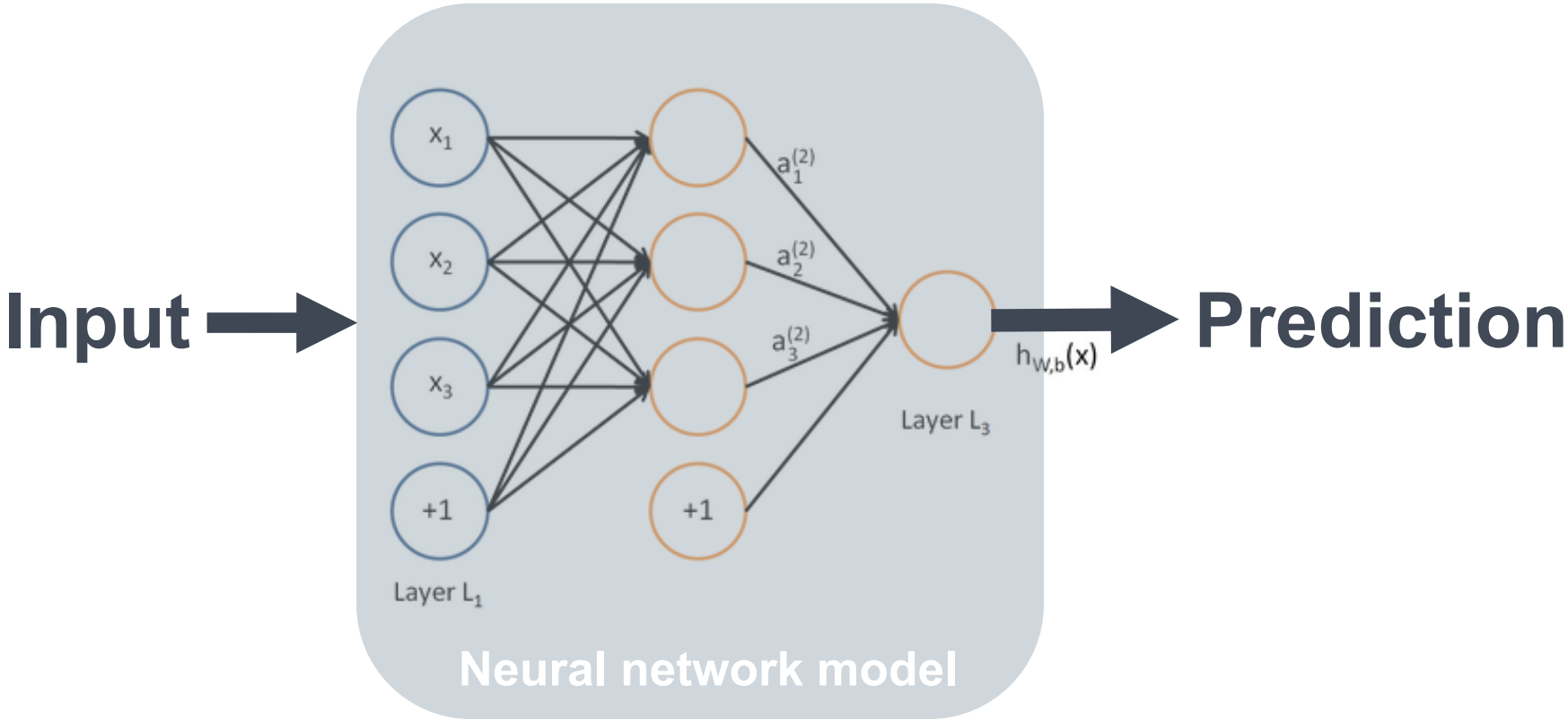
- New learning algorithms

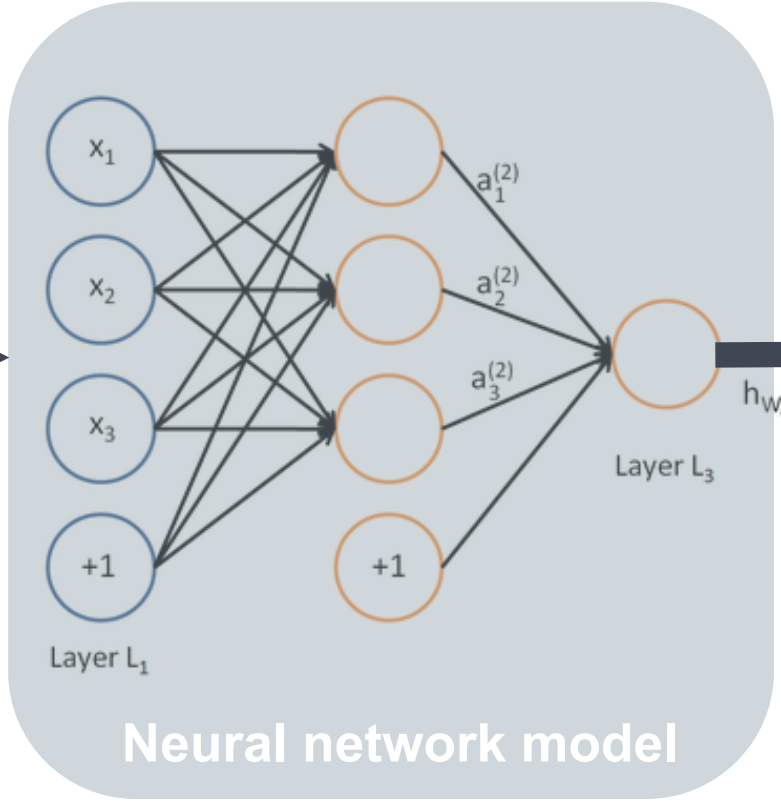

input
AI
prediction

Machine learning: a paradigm shift

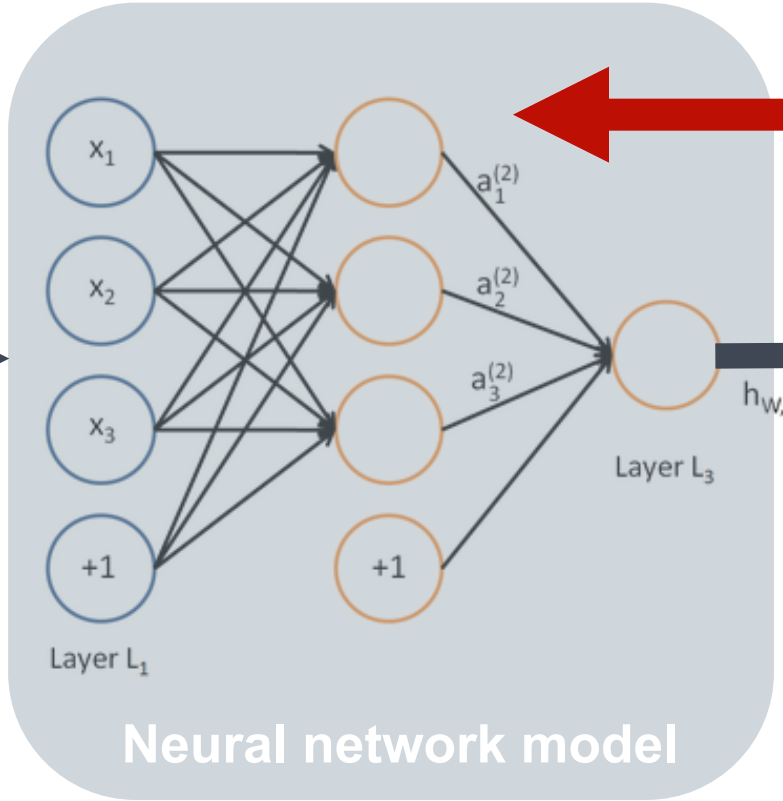



Input
Inference model
Prediction


Input
x_1
x_2
x_3
+1
Layer L_1
$a_1^{(2)}$
$a_2^{(2)}$
$a_3^{(2)}$
+1
Layer L_3
$h_{W,b}(x)$
Prediction
Neural network model


x_1
x_2
x_3
+1
Layer L_1
$a_1^{(2)}$
$a_2^{(2)}$
$a_3^{(2)}$
+1
Layer L_3
$h_{W,b}(x)$
Neural network model

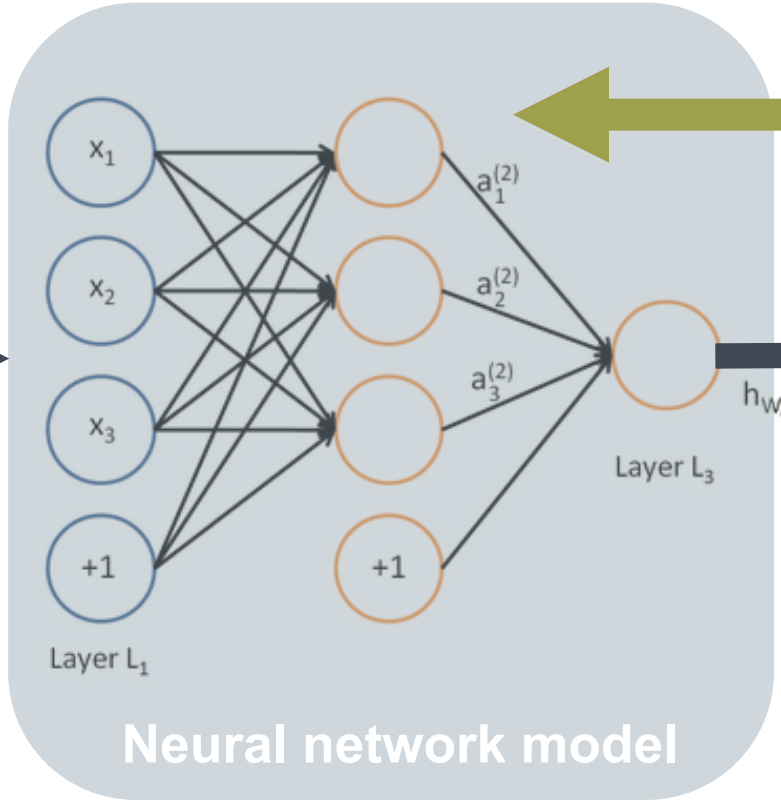
Cat?



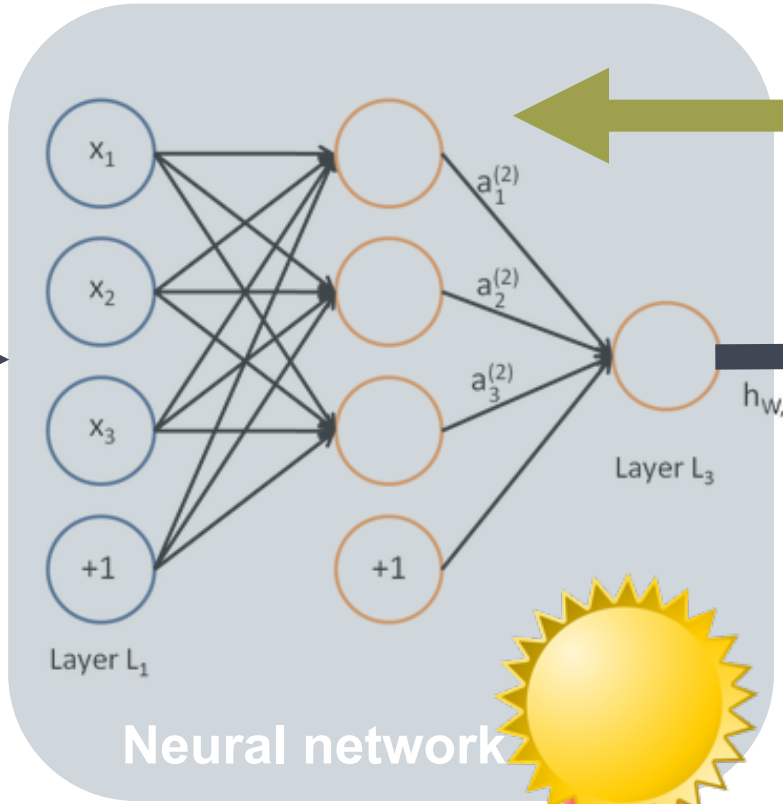
Tweak parameters

Cat?

Tweak parameters

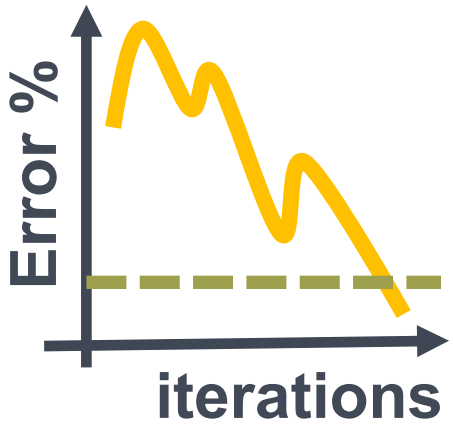


Cyclist ✓



Tweak parameters

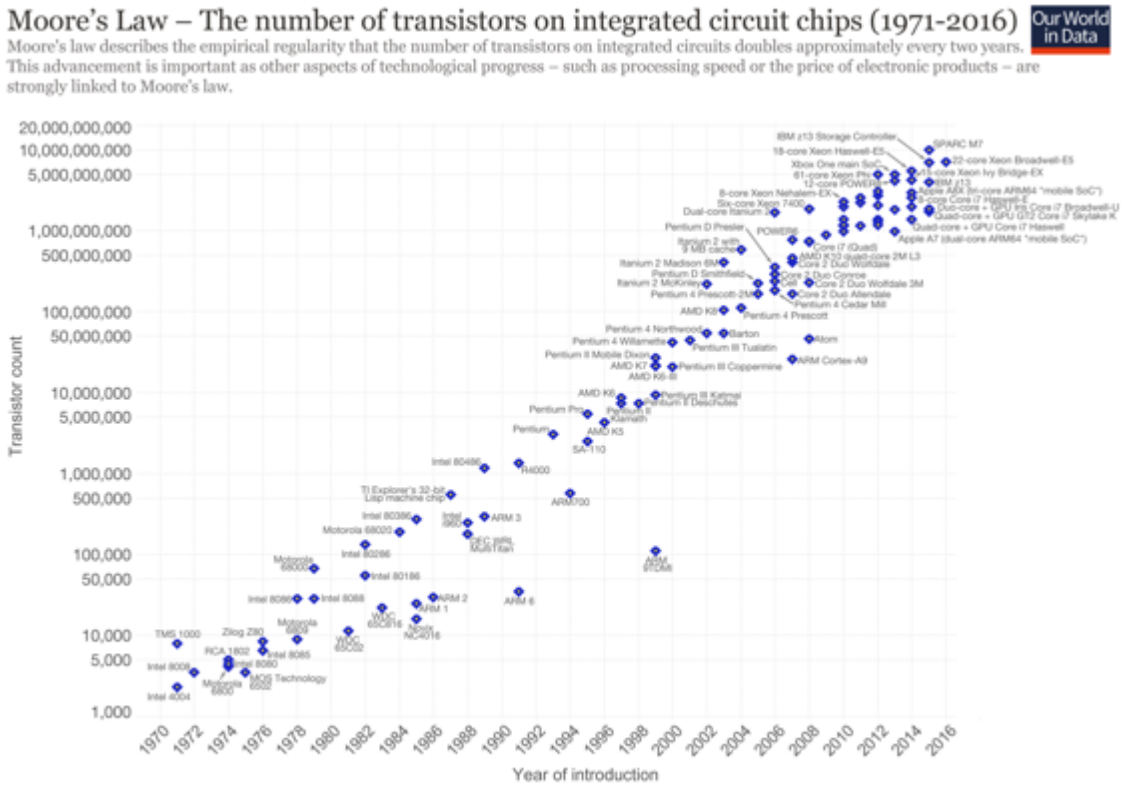
Cyclist ✓



Why now?

- New learning algorithms → Deep learning
- **More computing power**

More CPU

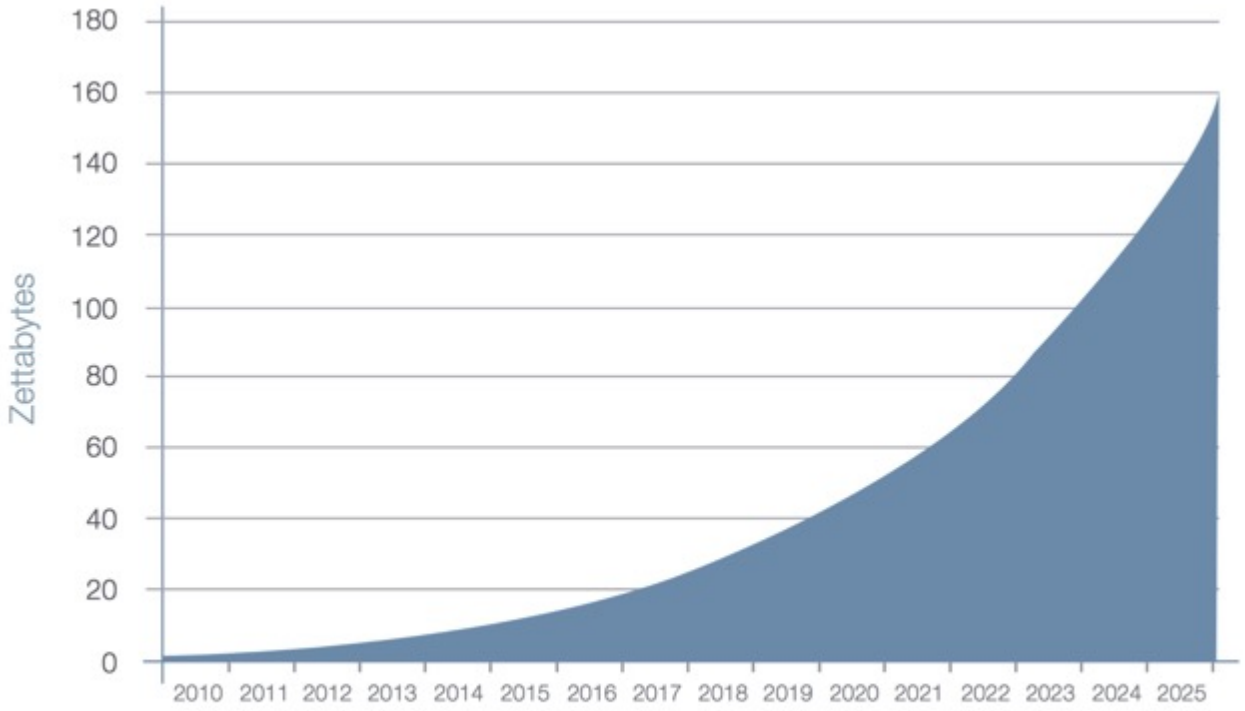


Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldInData.org. There you find more visualizations and research on this topic.

Why now?

- New learning algorithms → Deep learning
- More computing power → GPU, TPUs…
- **Availability of vast amounts of (training) data**

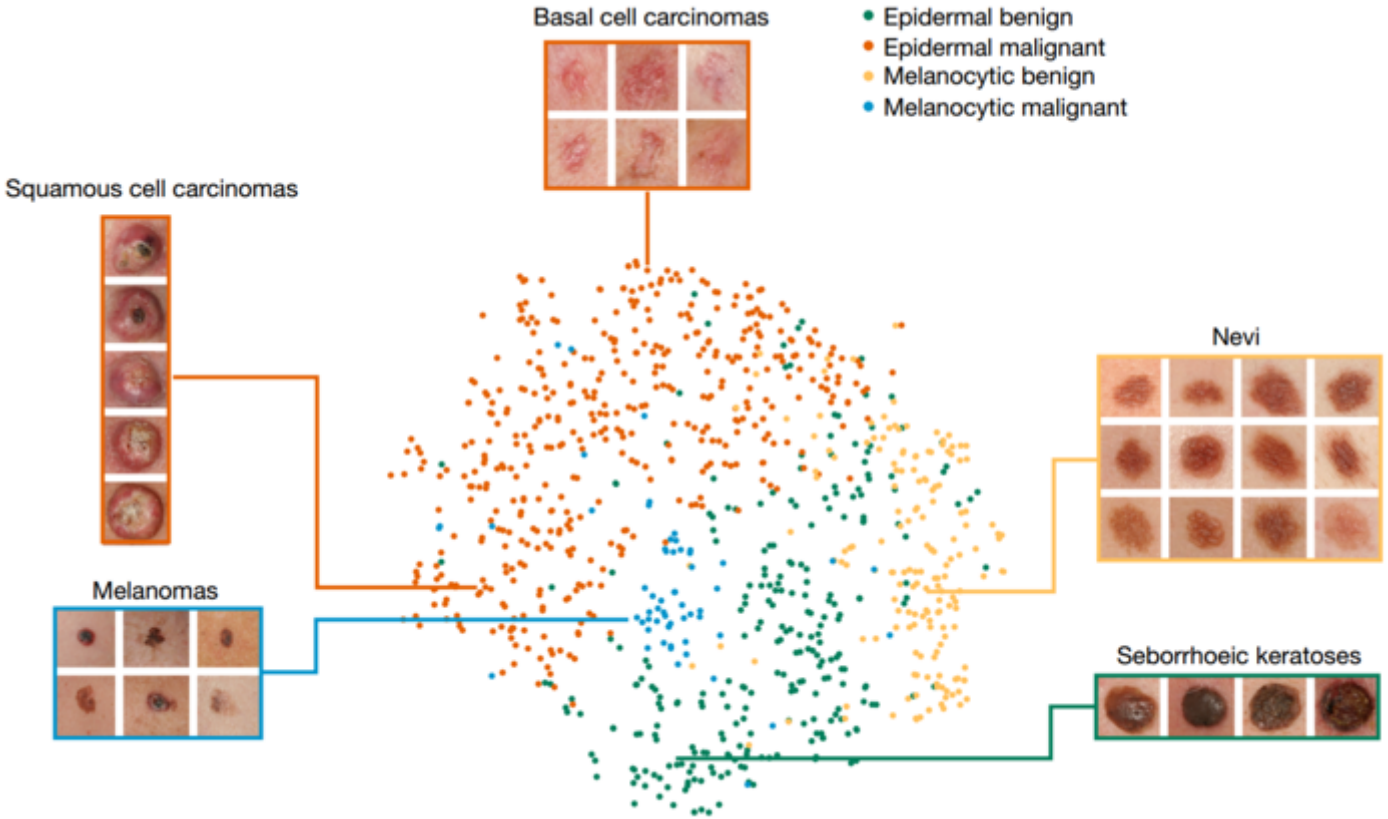
More data



Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

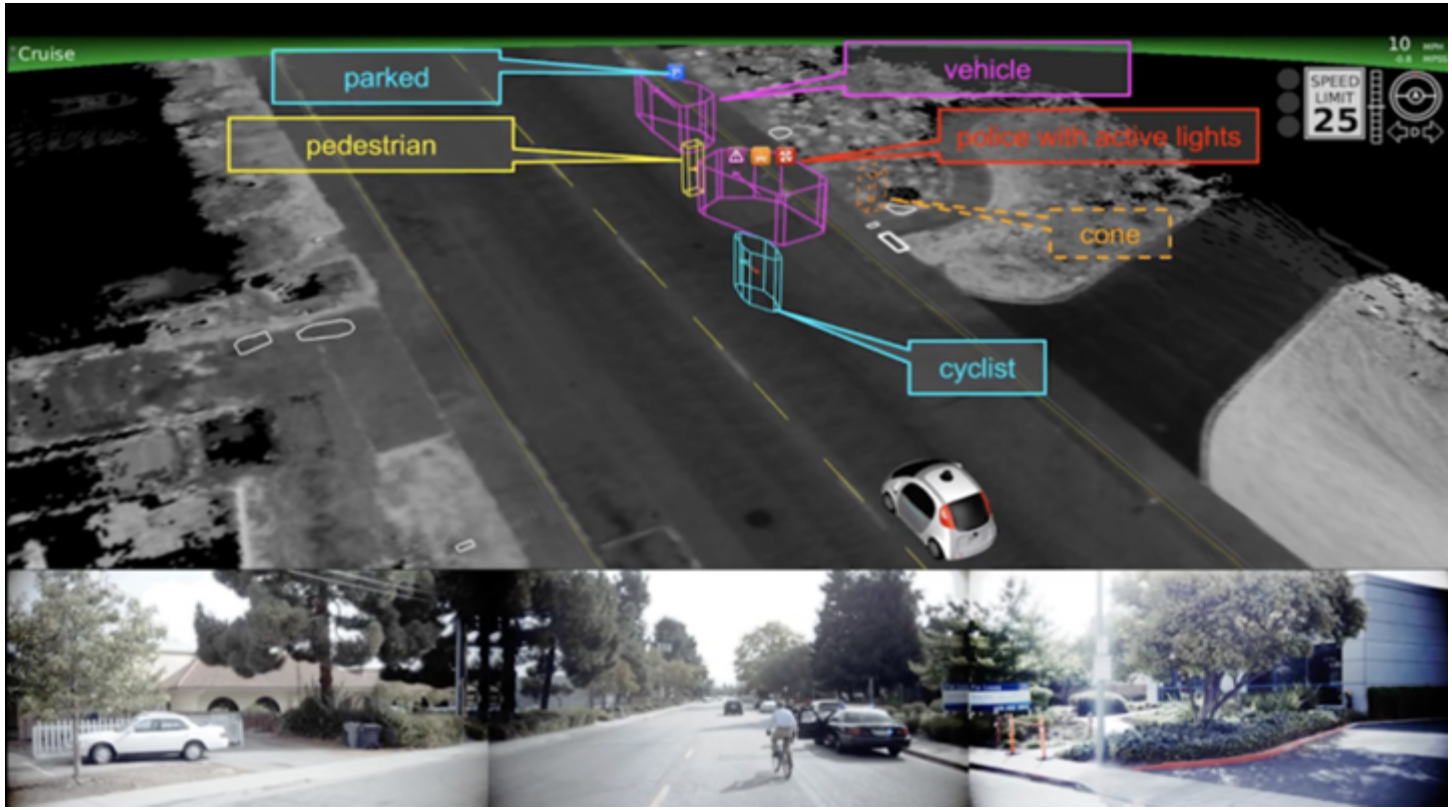
Applications

Human level lesion classification



Esteva et al. 2017 “Dermatologist-level classification of skin cancer with deep neural networks” Nature 542, 115-118

Scene understanding and planning

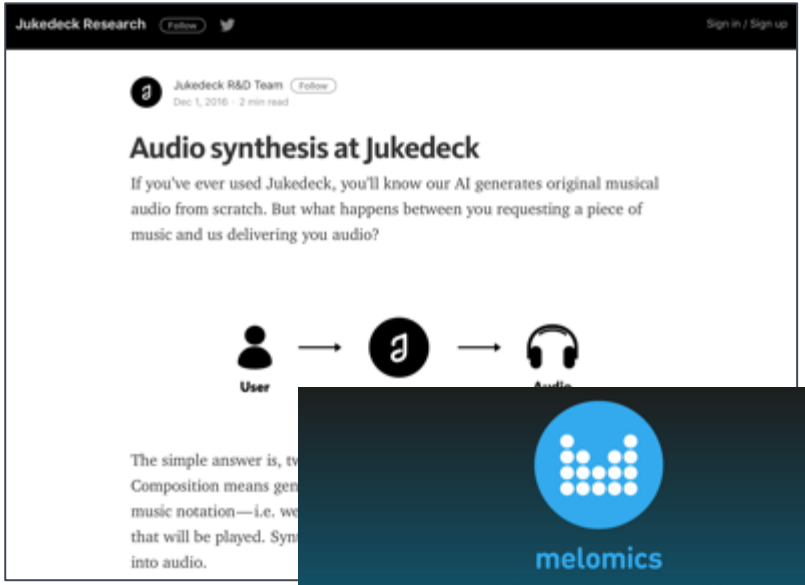


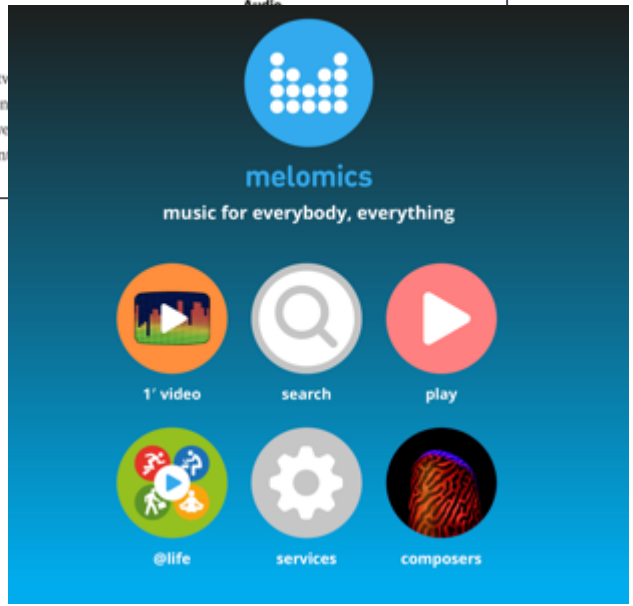
Source: Sacha Arnoud (Waymo) MIT https://selfdrivingcars.mit.edu

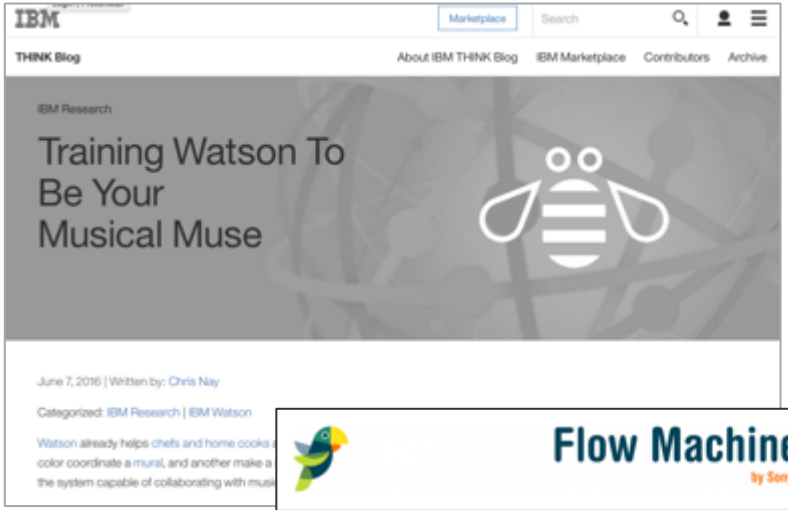
Translating images into text (and speech)

Source: https://code.facebook.com/posts/457605107772545/

(Gatys et al. 2015)


Jukedeck Research
Follow
Sign in / Sign up
Jukedeck R&D Team
Follow
Dec 1, 2016 · 2 min read
Audio synthesis at Jukedeck
If you've ever used Jukedeck, you'll know our AI generates original musical audio from scratch. But what happens between you requesting a piece of music and us delivering you audio?
User
Audio
The simple answer is, tw
Composition means gen
music notation—i.e. we
that will be played. Syn
into audio.


melomics
music for everybody, everything
1′ video
search
play
@life
services
composers


IBM
Marketplace
Search
THINK Blog
About IBM THINK Blog
IBM Marketplace
Contributors
Archive
IBM Research
Training Watson To Be Your Musical Muse
June 7, 2016 | Written by: Chris Nay
Categorized: IBM Research | IBM Watson
Watson already helps chefs and home cooks
color coordinate a mural, and another make a
the system capable of collaborating with musi


Flow Machines
by Sony CSL
MUSICIANS
PROJECTS
EVENTS
PRESS
Flow Machines are cutting-edge algorithms, made to explore new ways to create.
Flow Machines collaborate with musicians to compose the future.
Flow Machines are AI music-making.
FOLLOW US

“All you’re seeing now — all these feats of AI like self-driving cars, interpreting medical images, beating the world champion at Go and so on — these are **very narrow intelligences, and they’re really trained for a particular purpose**. They’re situations where we can collect **a lot of data**.”

Yann LeCun, 2017

“If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future.”

Andrew Ng, 2015

The need for AI protection: global impact

Source: McKinsey Global Institute – Application and value of Deep learning (April 2018)

AI: How to protect?

AI: What ~~How~~ to protect?

AI

- Computer program
- Training corpus / data
- Neural network topology
- Machine learning process
- Hardware
- AI applications
- Inference models

Directive 2009/24/EC

Article 1

Object of protection

1. In accordance with the provisions of this Directive, Member States shall protect computer programs, by copyright, as literary works within the meaning of the Berne Convention for the Protection of Literary and Artistic Works. For the purposes of this Directive, the term 'computer programs' shall include their preparatory design material.

2. Protection in accordance with this Directive shall apply to the expression in any form of a computer program. Ideas and principles which underlie any element of a computer program, including those which underlie its interfaces, are not protected by copyright under this Directive.

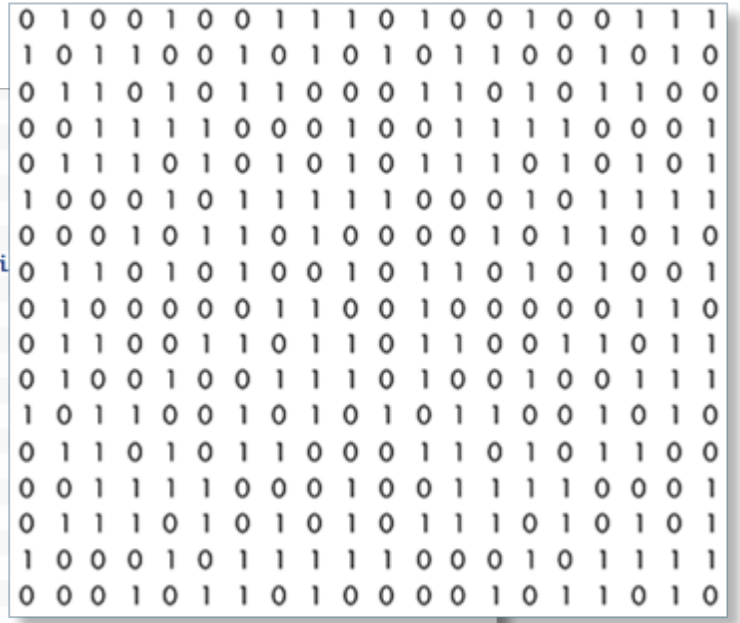
3. A computer program shall be protected if it is original in the sense that it is the author's own intellectual creation. No other criteria shall be applied to determine its eligibility for protection.

Copyrights?

```
# Forward propagate input to a network output
def forward_propagate(network, row):
    inputs = row
    for layer in network:
        new_inputs = []
        for neuron in layer:
            activation = activate(neuron['weights'], i
            neuron['output'] = transfer(activation)
            new_inputs.append(neuron['output'])
        inputs = new_inputs
    return inputs

# Calculate the derivative of an neuron output
def transfer_derivative(output):
    return output * (1.0 - output)

# Backpropagate error and store in neurons
def backward_propagate_error(network, expected):
    for i in reversed(range(len(network))):
        layer = network[i]
        errors = list()
        if i != len(network)-1:
            for j in range(len(layer)):
                error = 0.0
                for neuron in network[i + 1]:
                    error += (neuron['weights'][j] * neuron['delta'])
                errors.append(error)
        else:
            for j in range(len(layer)):
                neuron = layer[j]
                errors.append(expected[j] - neuron['output'])
        for j in range(len(layer)):
```



Only the expression of a computer program is protected

Ideas and principles (including the algorithms) which underlie any element of a program are **not protected** by copyright under this Directive (Art. 1(2)).

Copyrights?

1. **Input a set of training examples**

2. **For each training example** x**:** Set the corresponding input activation $a^{x,1}$, and perform the following steps:

   - **Feedforward:** For each $l = 2, 3, \ldots, L$ compute $z^{x,l} = w^l a^{x,l-1} + b^l$ and $a^{x,l} = \sigma(z^{x,l})$.
   - **Output error** $\delta^{x,L}$**:** Compute the vector $\delta^{x,L} = \nabla_a C_x \odot \sigma'(z^{x,L})$.
   - **Backpropagate the error:** For each $l = L-1, L-2, \ldots, 2$ compute $\delta^{x,l} = ((w^{l+1})^T \delta^{x,l+1}) \odot \sigma'(z^{x,l})$.

3. **Gradient descent:** For each $l = L, L-1, \ldots, 2$ update the weights according to the rule $w^l \to w^l - \frac{\eta}{m} \sum_x \delta^{x,l} (a^{x,l-1})^T$, and the biases according to the rule $b^l \to b^l - \frac{\eta}{m} \sum_x \delta^{x,l}$.

AI

- Computer program → Expression only
- Training corpus / data
- Neural network topology
- Machine learning process
- Hardware
- AI applications
- Inference models

How to protect?

- Copyright?

  → Does not protect functionalities / algorithms

- Trade secrets?

Directive (EU) 2016/943

15.6.2016 EN Official Journal of the European Union L 157/1

DIRECTIVE (EU) 2016/943 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 8 June 2016

on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 114 thereof,

HAVE ADOPTED THIS DIRECTIVE:

CHAPTER I

Subject matter and scope

Article 1

Subject matter and scope

1. This Directive lays down rules on the protection against the unlawful acquisition, use and disclosure of trade secrets.

“Four Elements” of a trade secret:

(1) Information (Art. 2 (1))
(2) **Secrecy** (Art. 2 (1)(a))
(3) Commercial value due to secrecy (Art. 2 (1)(b))
(4) **Measures to keep it secret** (Art. 2 (1)(c))

AI

- Computer program
- Training corpus / data
- Neural network topology
- Machine learning process
- Hardware
- AI applications
- Inference models

Reverse engineering

- **Recital 16:**
  - In the interest of innovation and to foster competition, the provisions of this Directive **should not create any exclusive right** to know-how or information protected as trade secrets.
  - Thus, the independent discovery of the same know-how or information should remain possible.
  - **Reverse engineering of a lawfully acquired product should be considered as a lawful means of acquiring information**

Lawful acquisition, use and disclosure of trade secrets

Article 3(1)(b):

observation, study, disassembly or testing of a product or object that has been made available to the public or that is lawfully in the possession of the acquirer of the information who is free from any legally valid duty to limit the acquisition of the trade secret;

AI

- Computer program
- Training corpus / data → Depends on the data
- Neural network topology → Reverse engineering
- Machine learning process → Reverse engineering
- Hardware
- AI applications
- Inference models

How to protect?

- Copyright?

  → Does not protect functionalities / algorithms

- Trade secrets?

  → Do not protect against reverse engineering

Patents?

→**Protect technical inventions**

- They give owners **the right to prevent third parties from making, using or exploiting an invention** without authorisation.

- They are **valid for up to 20 years**.

The conditions of patentability

Patents are granted **for inventions** in all fields of technology

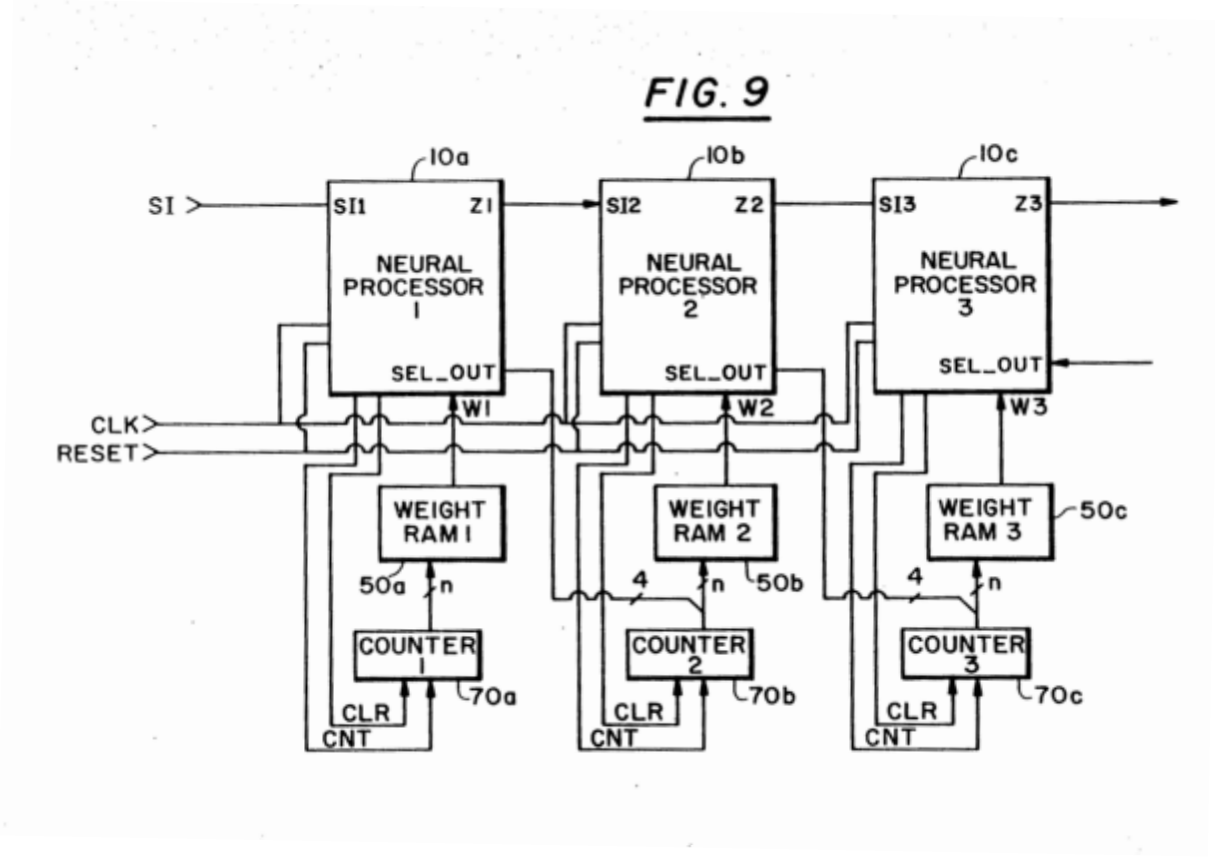
To be patentable, inventions must:

- be **new** (art. 54 EPC)
- involve an **inventive step** (art. 56 EPC)
- be **industrially applicable** (art. 57 EPC)

AI

- Computer program
- Training corpus / data
- Neural network topology
- Machine learning process
- Hardware
- AI applications
- Inference models

W0 9314461: Neural processor apparatus



EP2591443

Method for assisting vehicle guidance over terrain

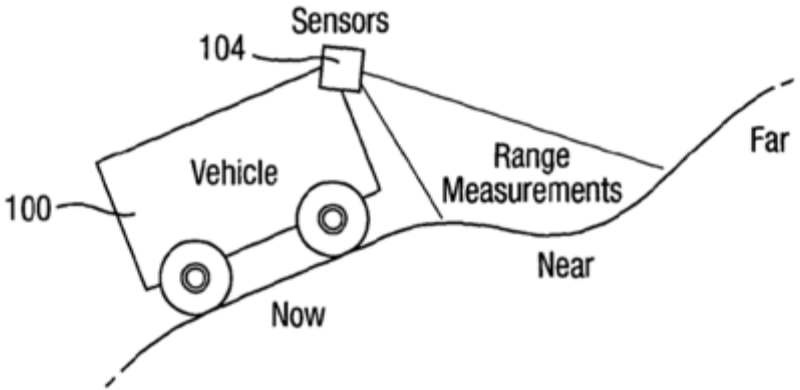
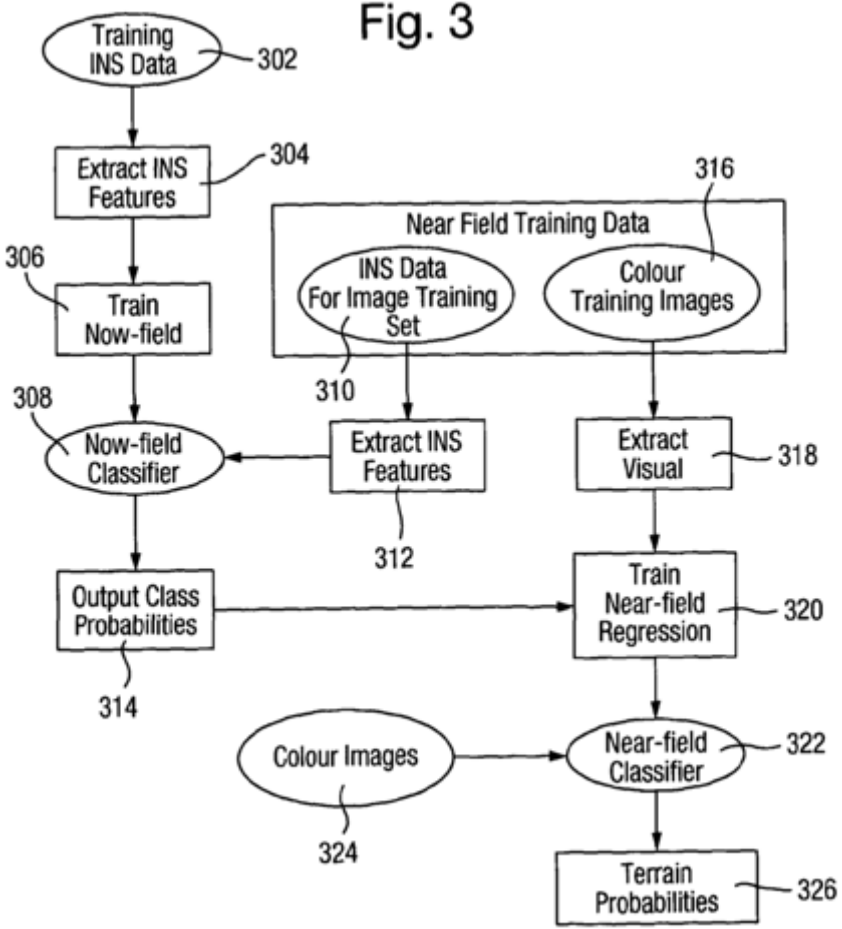
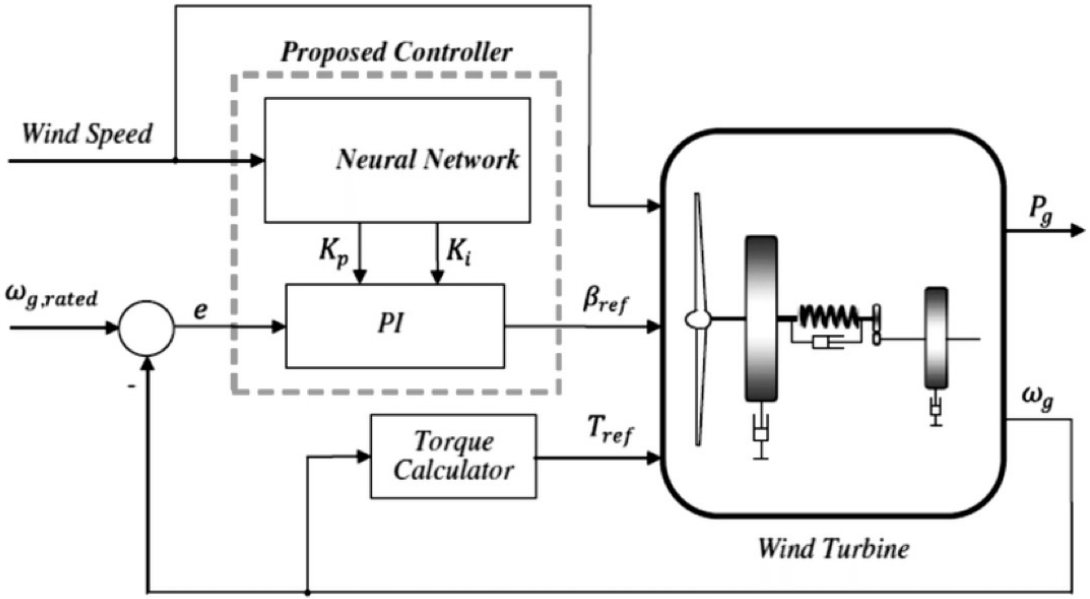


Fig. 3



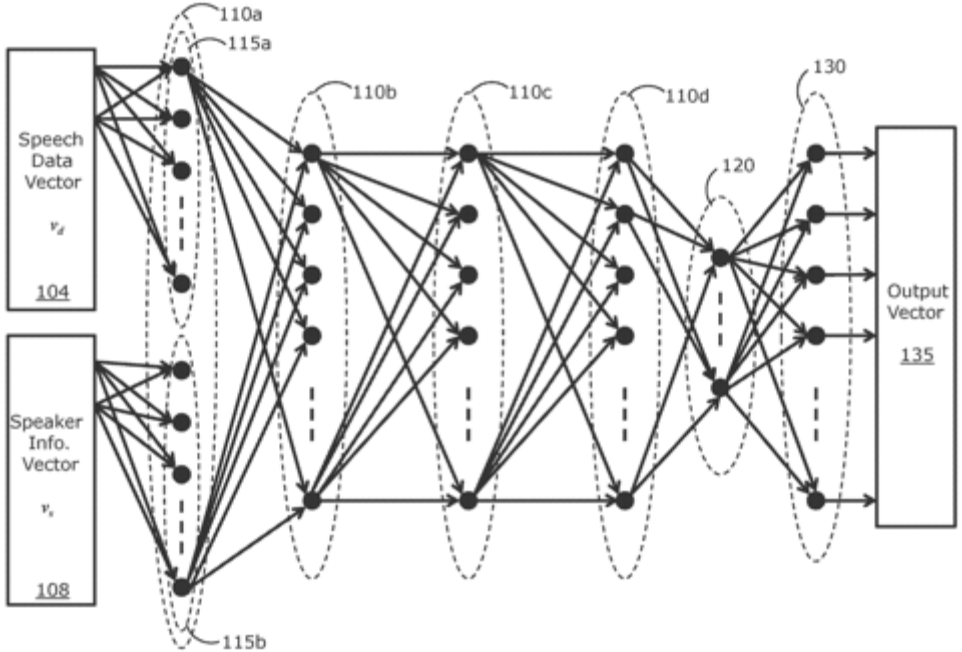
EP2801000

- **Method for controlling a turbine using a recurrent neural network**



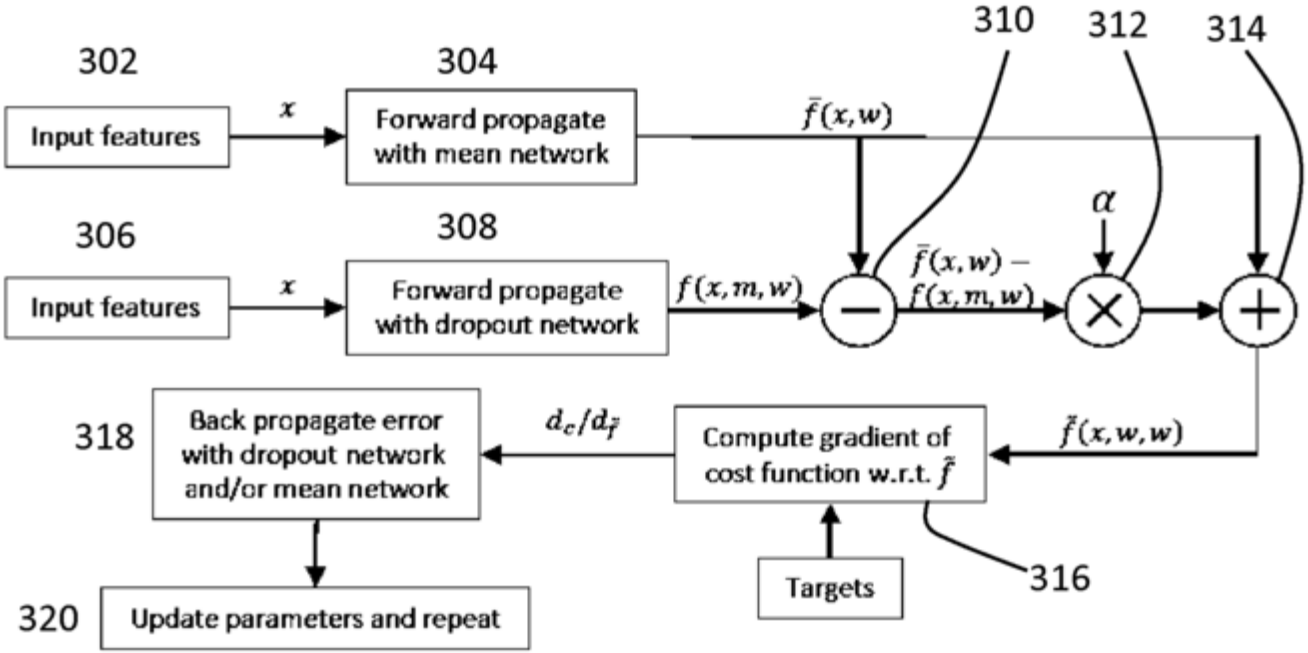
EP2887055

- **Method and apparatus for speech recognition using neural networks with speaker adaptation**



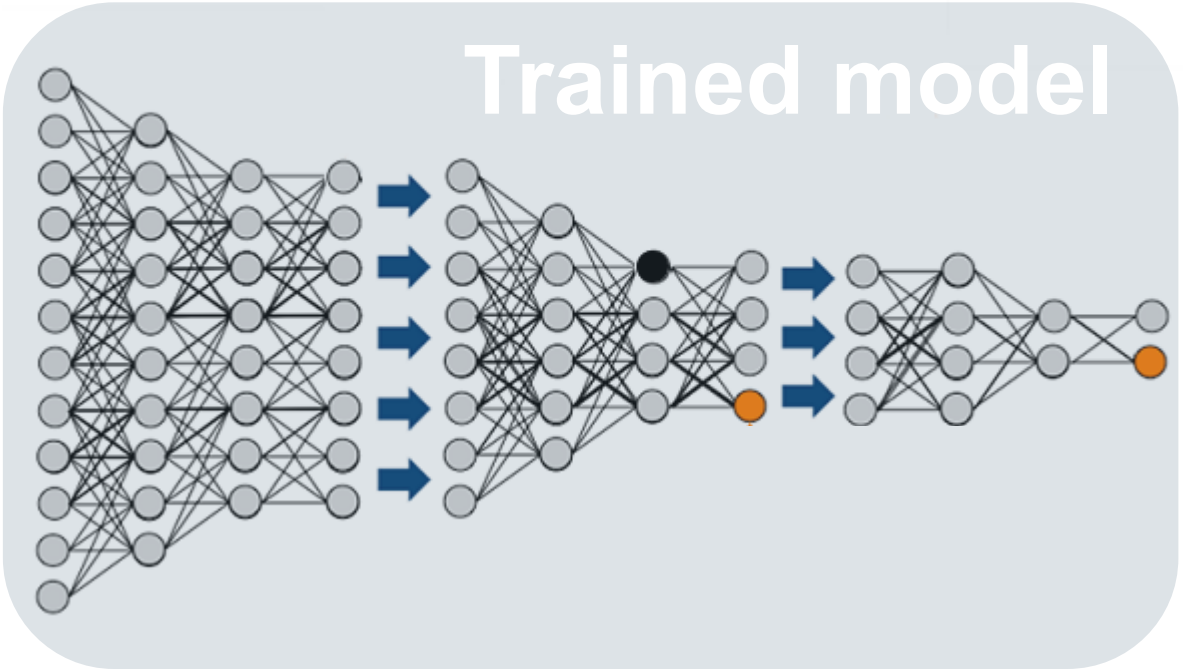
WO 2016145516 (A1)

- System and method for training a deep neural network



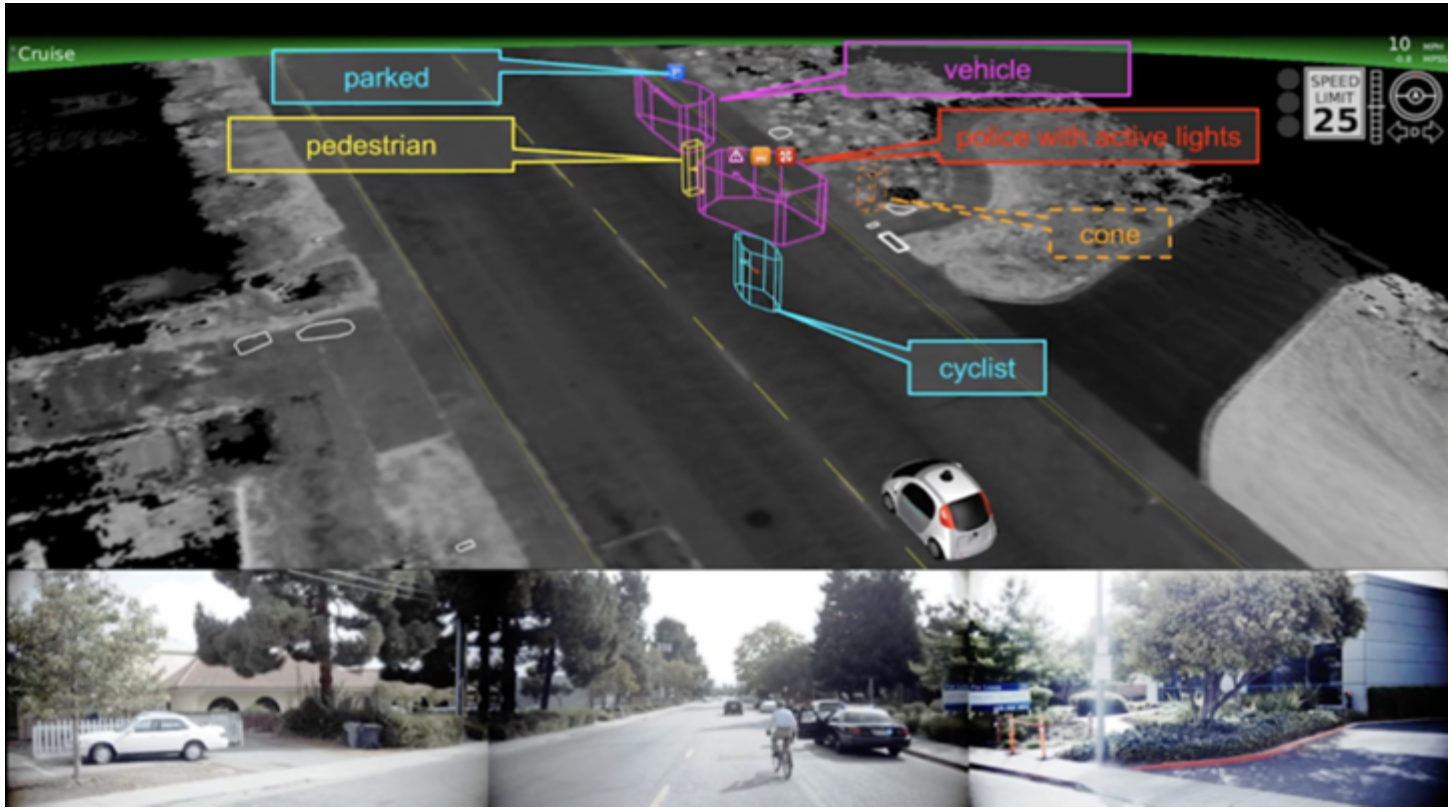
AI

- Computer program
- Training corpus / data
- Neural network topology
- Machine learning process
- Hardware
- AI applications
- Inference models ?...



No pedestrian

Pedestrian



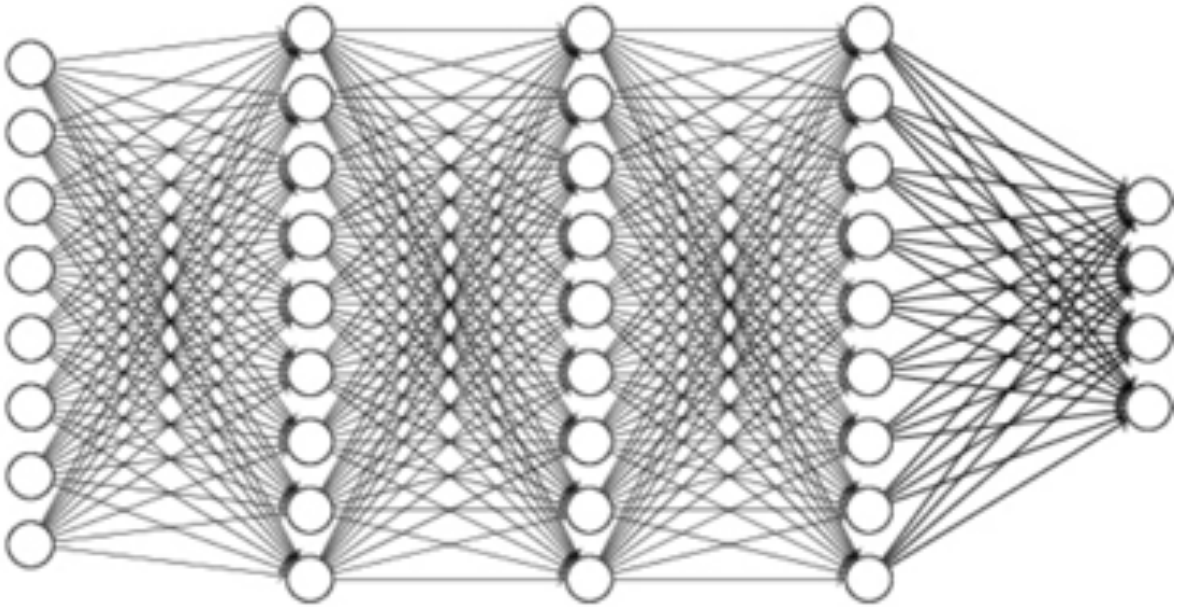
Source: Sacha Arnoud (Waymo) MIT https://selfdrivingcars.mit.edu

Protecting deep learning models

- ML models are the results of significant investments.
- They are the operative element of the AI pipeline

→ **How to protect such entities ?**

Trade secrets?



Billions of parameters!...

Reverse engineering the ML model… with ML

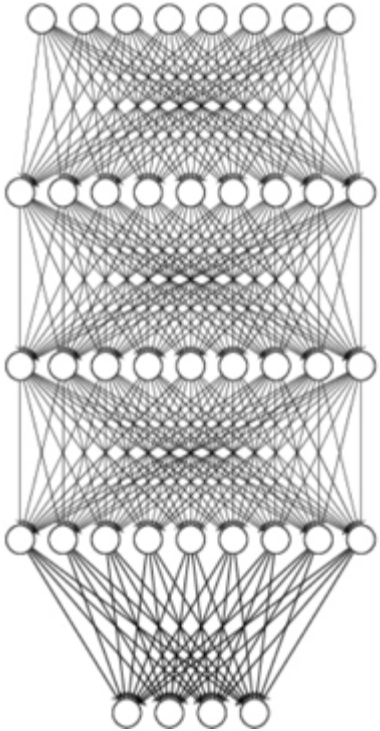
« *With the proposed black-box attack approach, an adversary can* ***use deep learning to reliably infer the necessary information by using labels previously obtained from the classifier under attack, and build a functionally equivalent machine learning classifier*** *without knowing the type, structure or underlying parameters of the original classifier.»*

Shi, Sagduyu & Grushin, How to steal a machine learning classifier with deep learning, IEEE, 2017)

→ Relying on trade secrecy only may be risky!...

Patents?

Claiming Deep learning models?



$w_{1,1}$=0.10201, $w_{1,2}$=0.00783,
$w_{1;3}$=0.23998, $w_{1,4}$=0.55410,
$w_{1,5}$=0.00341, $w_{1,6}$=0.10201,
$w_{1,7}$=0.00681, $w_{1;8}$=0.13389,
$w_{1,9}$=0.65453, $w_{1,10}$=0.01981,
$w_{1,11}$=0.00341, $w_{1,12}$=0.1021,
$w_{1,13}$=0.04681, $w_{1;14}$=0.93110,
$w_{1,15}$=0.7853, $w_{1,16}$=0.02901, …
…
…
…
$w_{1,9901}$=0.07421, $w_{1,9902}$=0.40201, ……

Databases?

Sui generis database?

CHAPTER III

SUI GENERIS RIGHT

Article 7

Object of protection

1. Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively **a substantial investment in either the obtaining, verification or presentation of the contents** to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.

Sui generis database?

The European Court of Justice ruled that **database contents, which evidence a substantial investment in the creation of data rather than in its obtainment, are excluded from protection**

- *British Horseracing Board Ltd and Others v. William Hill Organization Ltd*, ECJ case C-203/02, 9 Nov. 2004 (from England);
- *Fixtures Marketing Ltd v. Oy Veikkaus Ab*, ECJ case C-45/02 (from Finland);
- *Fixtures Marketing Ltd v. Organ- ismos prognostikon agonon podosfairou AE*, ECJ case C-444/02 (from Greece);
- *Fixtures Marketing Ltd. v. Svenska Spel AB*, ECJ case C-338/02 (from Sweden).

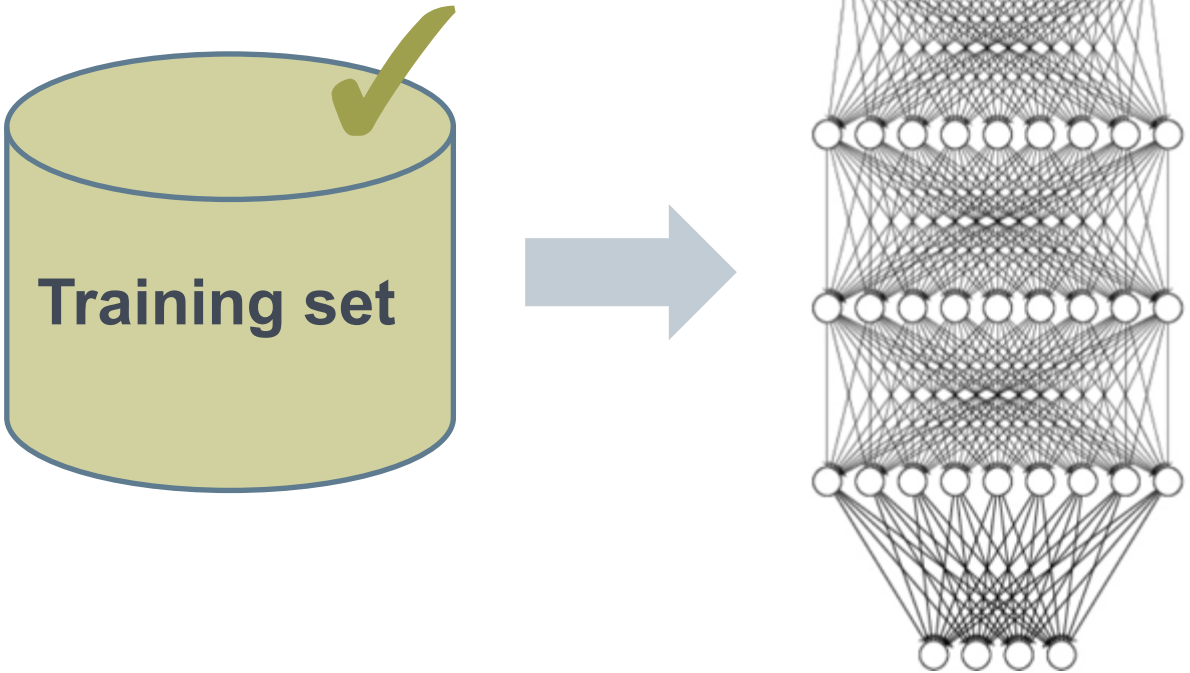
Sui generis database?

- Investment in a database must **refer to the finding and collecting of *existing* data**.

- **It cannot refer solely to data creation**.

- Training process of AI (inference) model consists in defining (i.e. “creating”, through the process of machine learning) the values of the parameters that constitute the model…

Sui generis database → Training data?



AI

- Computer program
- Training corpus / data
- Neural network topology
- Machine learning process
- Hardware
- AI applications
- Inference models → Patents? Trade Secrets? ~~Database~~

AI

- Computer program → Copyright (source code)
- Training corpus / data → Sui generis DB, Trade secret…
- Neural network topology → Patents, Trade secret…
- Machine learning process → Patents, Trade secret…
- Hardware → Patents
- AI applications → Patents
- Inference models → Patents? Trade Secrets? ~~Database~~

Article 64(2) EPC

"If the subject-matter of the European patent is a process, the protection conferred by the patent shall extend to the products **directly** obtained by such process."

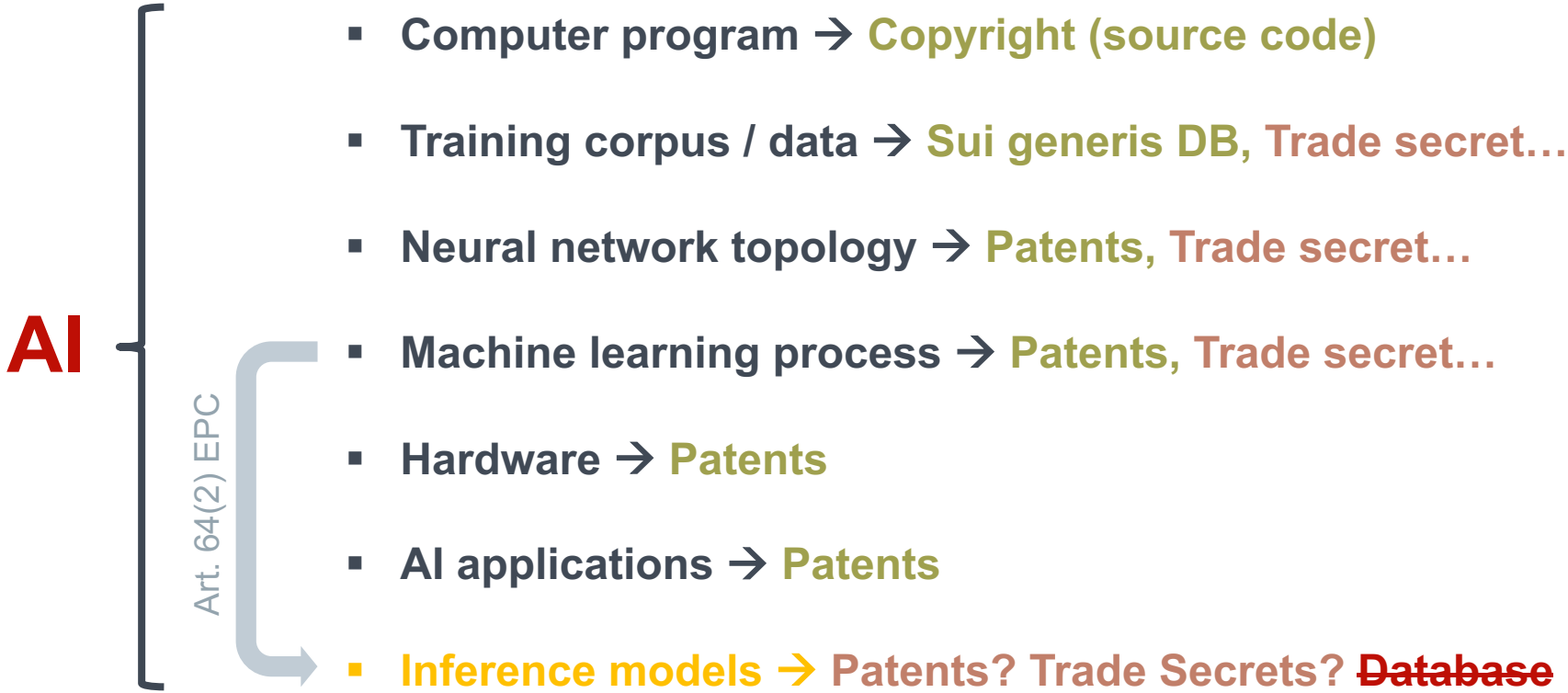
If the training process is patentable

→ The parameters are the direct product of the training

→ As such they are protected under Art. 64(2) EPC

However:

Modifying (ever slightly) the weights ends the protection…


AI
Art. 64(2) EPC
Computer program → Copyright (source code)
Training corpus / data → Sui generis DB, Trade secret…
Neural network topology → Patents, Trade secret…
Machine learning process → Patents, Trade secret…
Hardware → Patents
AI applications → Patents
Inference models → Patents? Trade Secrets? ~~Database~~

The protection of deep learning

→ Multiple (sub-)compontents
→ Requires a combination of IPRs

Open questions:
→ Training corpus…
→ Inference model…