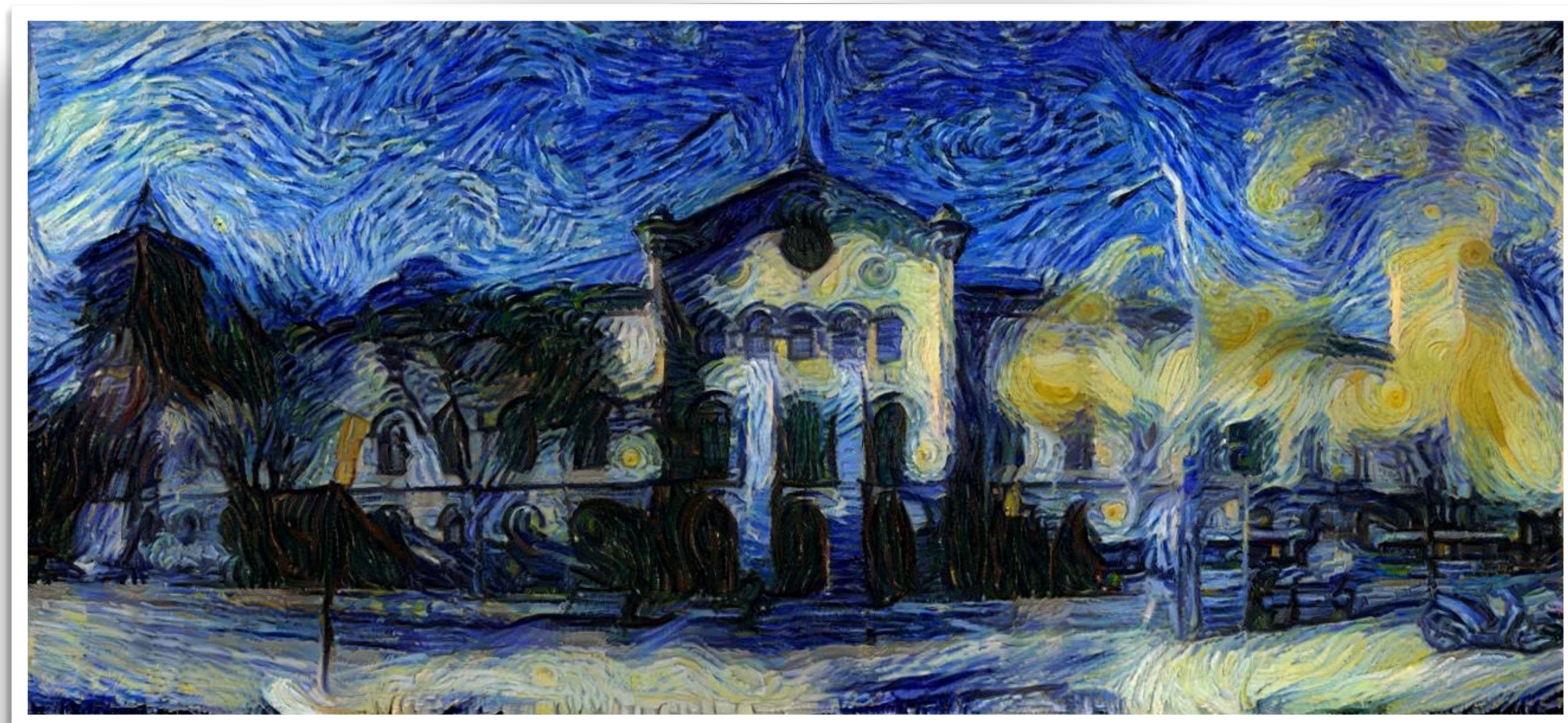


Deep Learning

Jordi Vitrià
jordi.vitria@ub.edu

Course Contents

1. Describe how a (deep) neural network works and combine different types of layers and activation functions. Deep Learning is not magic.
2. Describe how these models can be applied in computer vision, text analytics, time series analysis, etc. Deep Learning is not the final machine learning method.
3. Develop your own models in **Tensorflow**, **Keras** and derivates. You can train (small) deep models in your laptop.

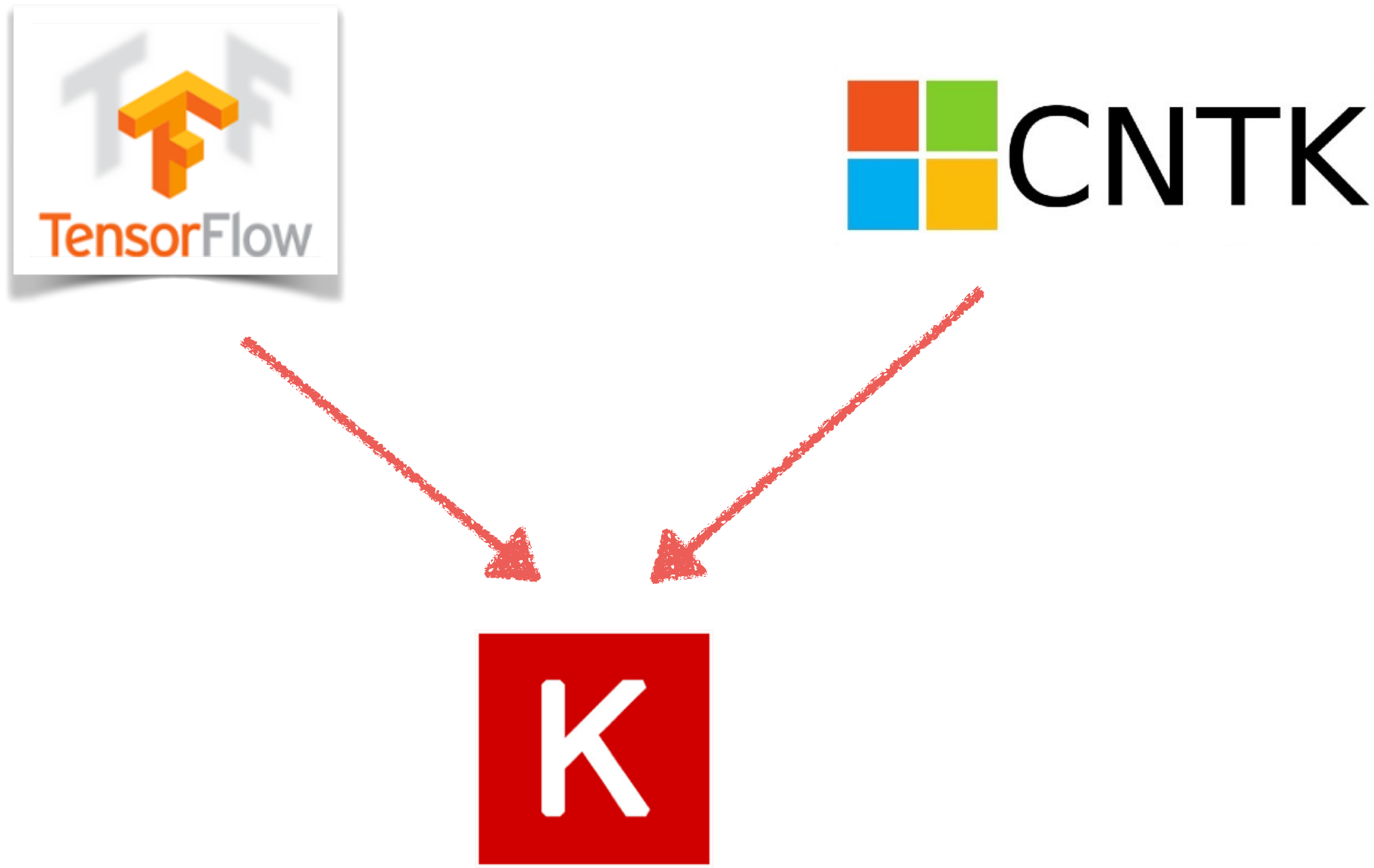
BackEnds

TensorFlow

PyTorch

mxnet Caffe2 Chainer

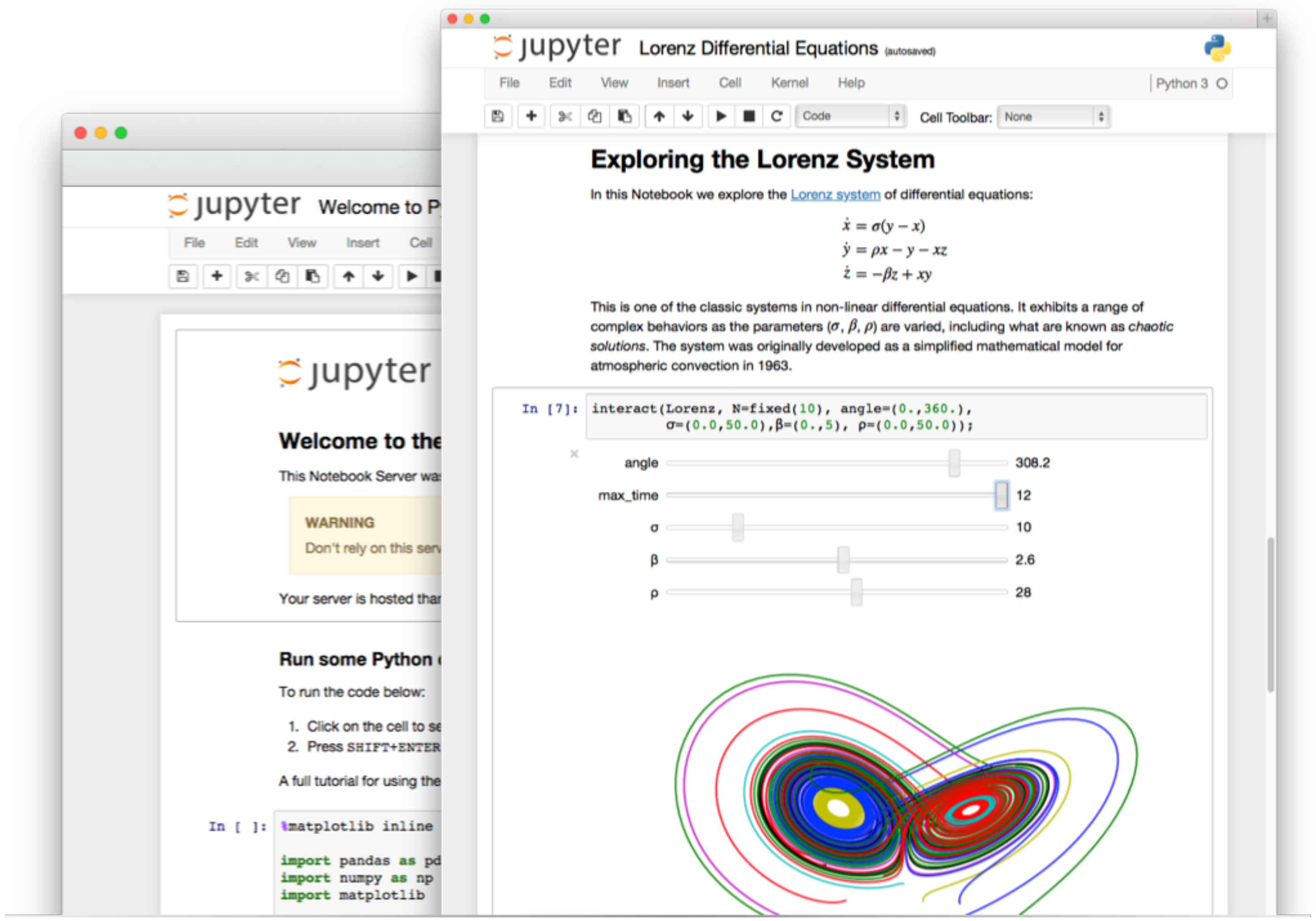
Wrappers



Keras

Approach

We will illustrate all contents with **Jupyter notebooks**, a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text.



How to run a Jupyter Notebook?

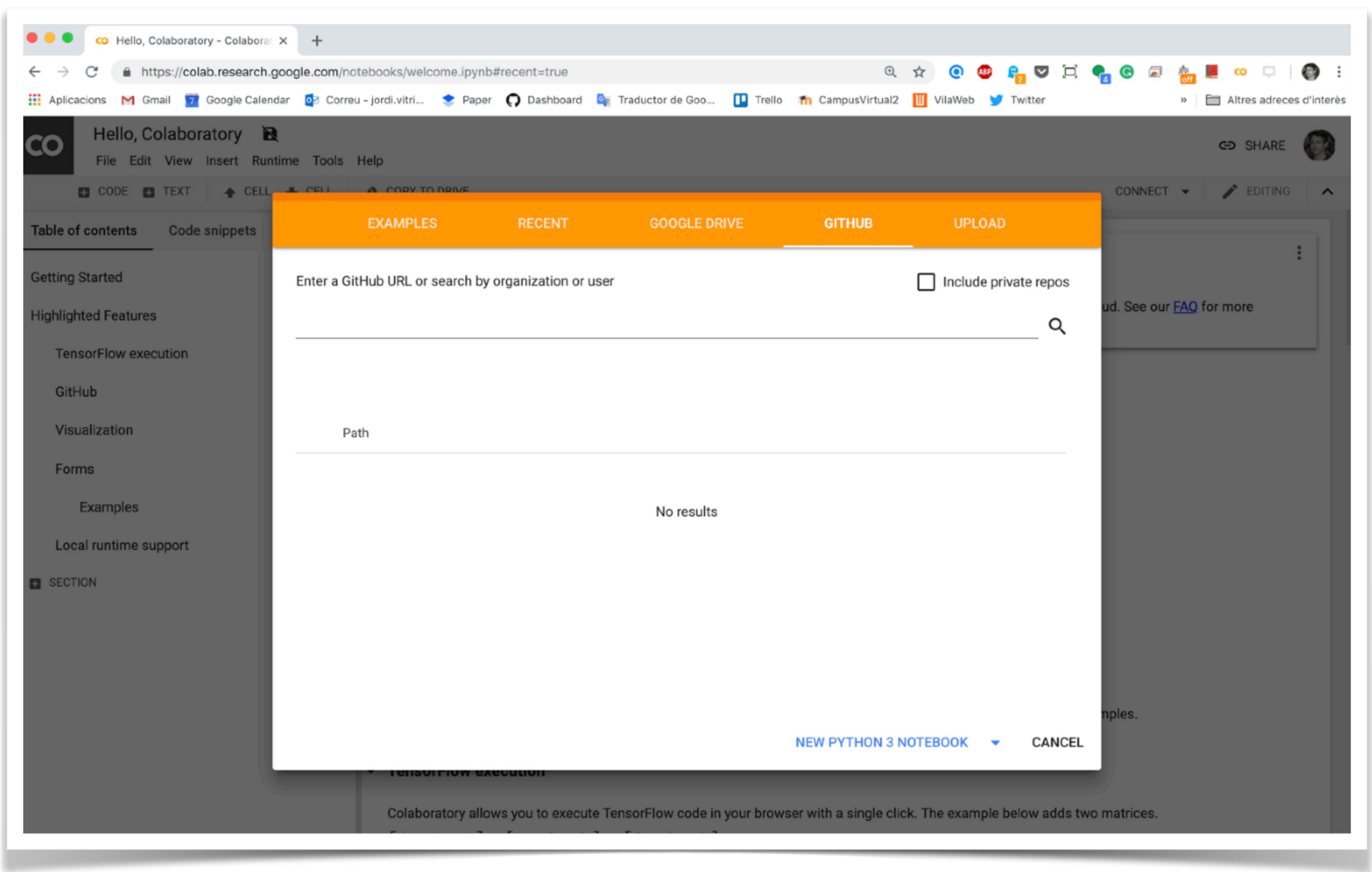
1

The easiest way for a beginner to get started with Jupyter Notebooks is by installing **Anaconda**. Anaconda is the most widely used Python distribution for data science and comes pre-loaded with all the most popular libraries and tools.

How to run a Jupyter Notebook?

2

Colaboratory is a hosted (Google) Jupyter notebook environment that is free to use and requires no setup.



How to run a Jupyter Notebook?

3

Docker provides the ability to build a runtime environment that not only remains isolated from other running containers, but also can be deployed to multiple locations in a repeatable way. Docker also uses a text document — a Dockerfile — that contains all the commands to assemble an image, which will meet our need to document the build environment.

Why Deep Learning?


It's funny!
It's not rocket science!
It's powerful!

- In 1943, neurophysiologist **Warren McCulloch** and mathematician **Walter Pitts** wrote a paper on how neurons might work. In order to describe how neurons in the brain might work, they modeled a simple neural network using electrical circuits.

- In 1949, Donald **Hebb** wrote *The Organization of Behavior*, a work which pointed out the fact that neural pathways are strengthened each time they are used, a concept fundamentally essential to the ways in which humans learn. If two nerves fire at the same time, he argued, the connection between them is enhanced.

- In 1957 **Frank Rosenblatt** attempted to build a kind of mechanical brain called the Perceptron, which was billed as “a machine which senses, recognizes, remembers, and responds like the human mind”.


- In 1962, **Widrow & Hoff** developed a learning procedure that examines the value before the weight adjusts it (i.e. 0 or 1) according to the rule: Weight Change = (Pre-Weight line value) * (Error / (Number of Inputs)). It is based on the idea that while one active perceptron may have a big error, one can adjust the weight values to distribute it across the network, or at least to adjacent perceptrons.


- A critical book written in 1969 by **Marvin Minsky** and his collaborator **Seymour Papert** ("Perceptrons") showed that Rosenblatt's original system was painfully limited, literally blind to some simple logical functions like "exclusive-or". What had become known as the field of "neural networks" all but disappeared.

First neural network winter is coming

- In 1982, interest in the field was renewed. **John Hopfield** of Caltech presented a paper to the National Academy of Sciences. His approach was to create more useful machines by using bidirectional lines. Previously, the connections between neurons was only one way.

- In 1986, the problem was how to extend the Widrow-Hoff rule to multiple layers. Three independent groups of researchers, which included **David E. Rumelhart, Geoffrey E. Hinton** and **Ronald J. Williams**, came up with similar ideas which are now called back-propagation networks because it distributes pattern recognition errors throughout the network.

- From 1986 to mid 90's new developments arised: convolutional neural networks (**Y.LeCun**), unsupervised learning (**Y.Bengio**), RBM (**G.Hinton**), etc. But, by this point **new machine learning methods** had begun to also emerge, and people were again beginning to be skeptical of neural nets since they seemed so intuition-based and since computers were still barely able to meet their computational needs.

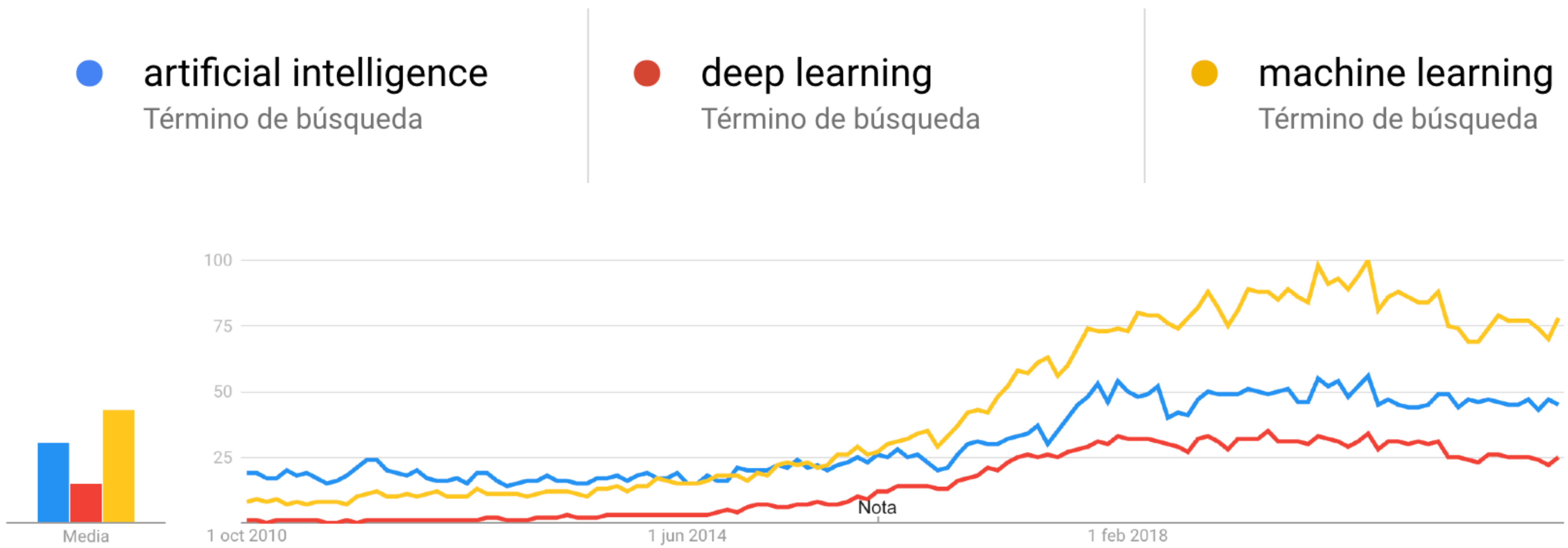
Second neural network winter is coming

- With the ascent of **Support Vector Machines** and the failure of backpropagation, the early 2000s were a dark time for neural net research.

- Then, what every researcher must dream of actually happened: G.Hinton, S.Osindero, and Y.W.Teh published a paper in 2006 that was seen as a breakthrough, a breakthrough significant enough to rekindle interest in neural nets: *A fast learning algorithm for* ***deep*** ***belief nets***.

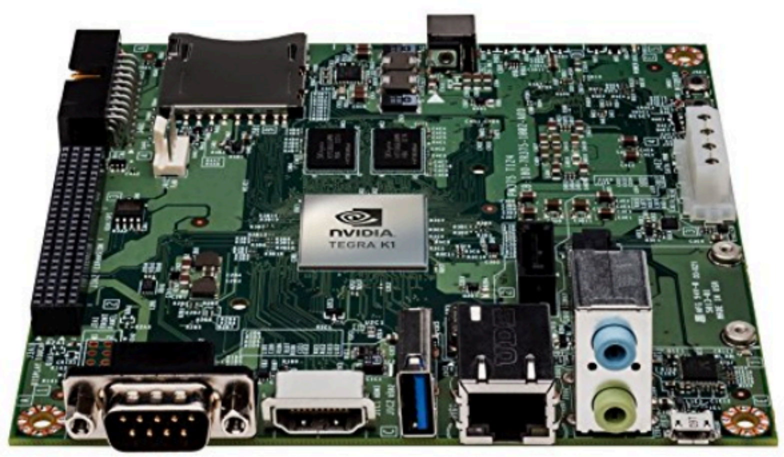
- After that, following Moore's law, computers got dozens of times faster (GPUs) since the slow days of the 90s, making learning with large datasets and many layers much more tractable.

Neural Networks Reborn



Google Trends

UNIVERSITAT DE BARCELONA



Roll over image to zoom in

NVIDIA

NVIDIA Jetson TK1 Development Kit

33 customer reviews | 24 answered questions

List Price: ~~$199.99~~

Price: $170.77 + $49.57 Shipping & Import Fees Deposit to Spain Details

You Save: $29.22 (15%)

Item is eligible: No interest if paid in full within 6 months with the Amazon.com Store Card.

In Stock.

This item ships to **Barcelona, Spain. Want it Monday, Feb. 12?** Order within **5 hrs 25 mins** and choose **AmazonGlobal Priority Shipping** at checkout. Learn more

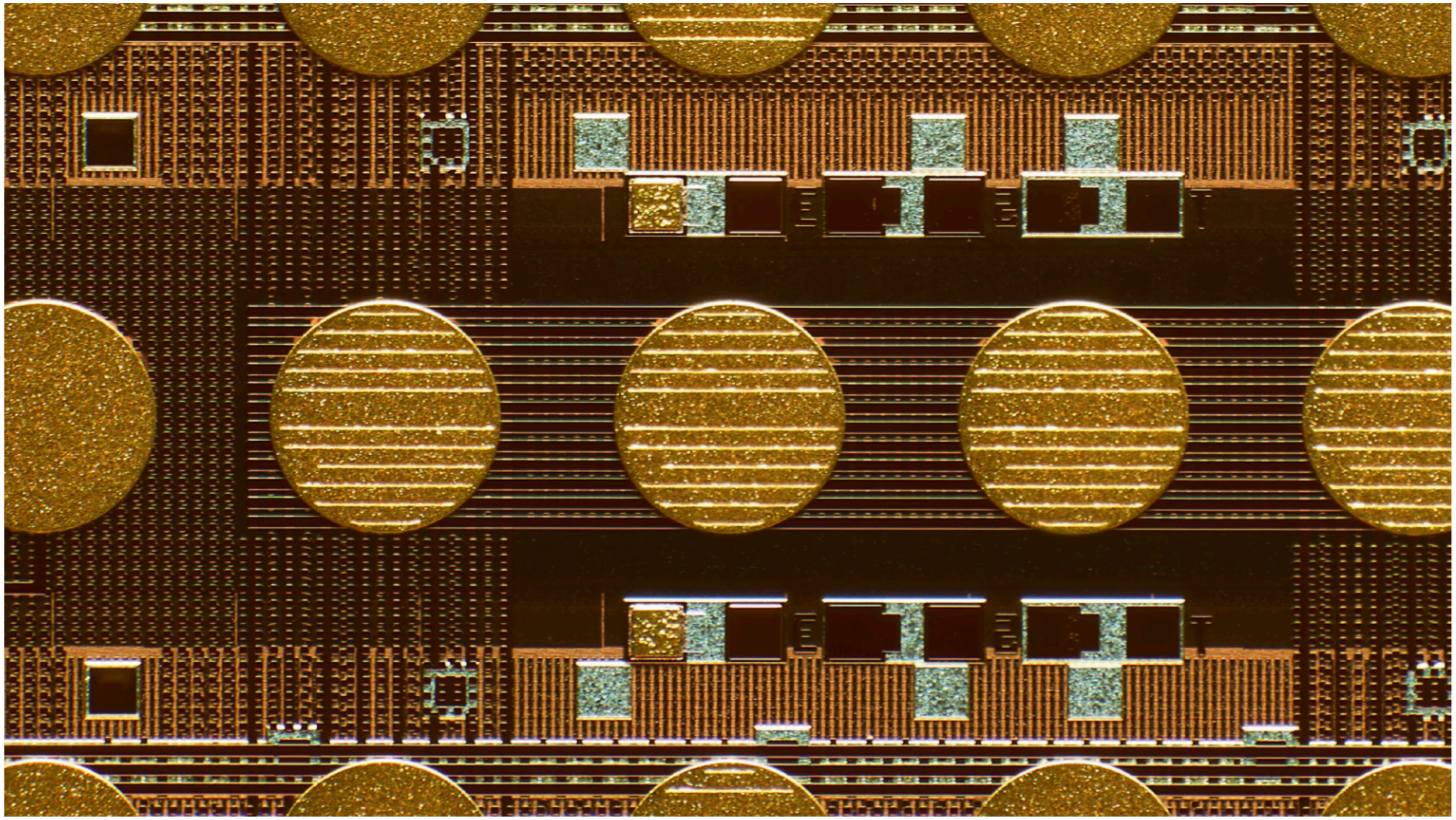
Ships from and sold by Amazon.com. Gift-wrap available.

- NVIDIA Kepler GPU with 192 CUDA cores
- NVIDIA 4-Plus-1 quad-core ARM Cortex-A15 CPU
- 2 GB memory, 16 GB eMMC
- Gigabit Ethernet, USB 3.0, SD/MMC, miniPCIe
- HDMI 1.4, SATA, Line out/Mic in, RS232 serial port
- Expansion ports for additional display, GPIOs, and high-bandwidth camera interface

NICOLE KOBIE LONG READS 17.06.2021 10:10 AM

NVIDIA and the battle for the future of AI chips

NVIDIA's GPUs dominate AI chips. But a raft of startups say new architecture is needed for the fast-evolving AI field



SUN LEE

Definitions

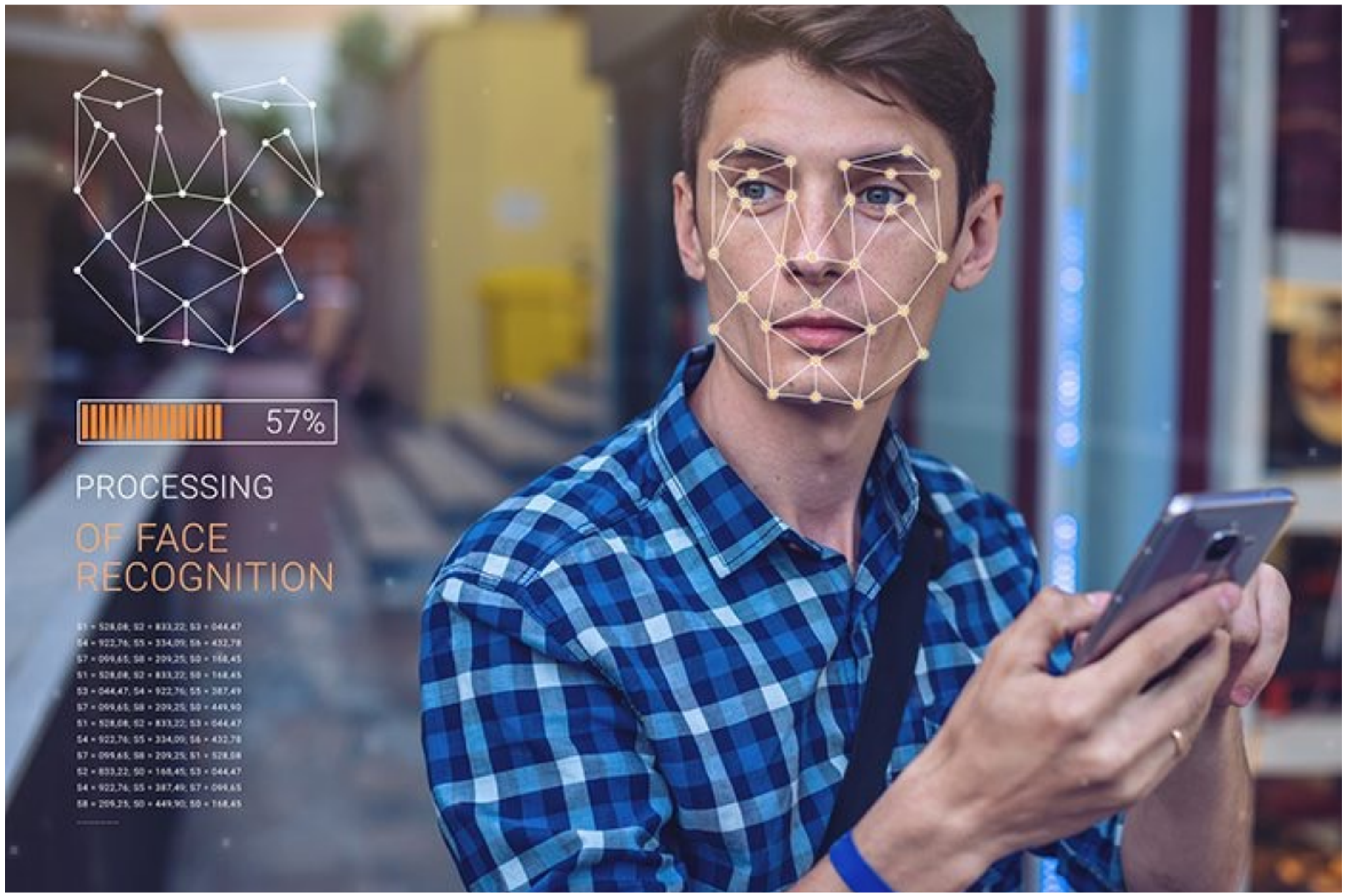
- **Neural Networks (NN)** is a beautiful biologically-inspired programming paradigm which enables a computer to learn from observational data.

- **Deep Learning (DL)** is a powerful set of techniques for learning in neural networks.

- NN and DL currently provide the best solutions to many problems in image recognition, speech recognition, and natural language processing.

“Classical” applications:
object classification, detection and segmentation.

Face recognition.



DeepFace (Facebook): Accuracy of 97.35% (2014)
FaceNet (Google): Accuracy of 99.63% (2015)
InsightFace (Open Source): Accuracy of 99.86% (2020)
??

New applications: navigation and mapping.

UNIVERSITAT DE BARCELONA

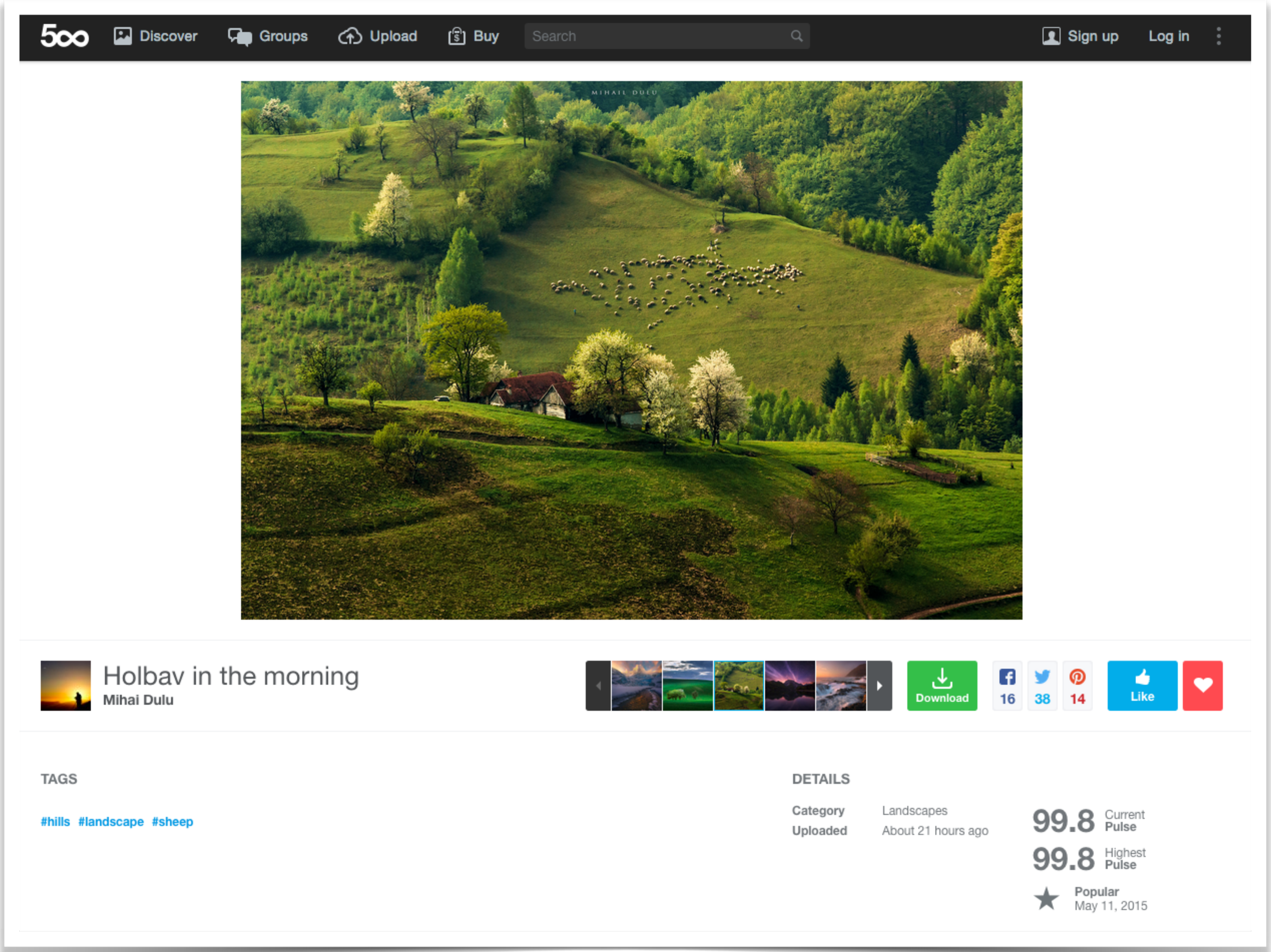
New applications: Image Upscaling (Flipboard)

http://engineering.flipboard.com/2015/05/scaling-convnets/

New applications: Image Upscaling (Flipboard)

http://engineering.flipboard.com/2015/05/scaling-convnets/

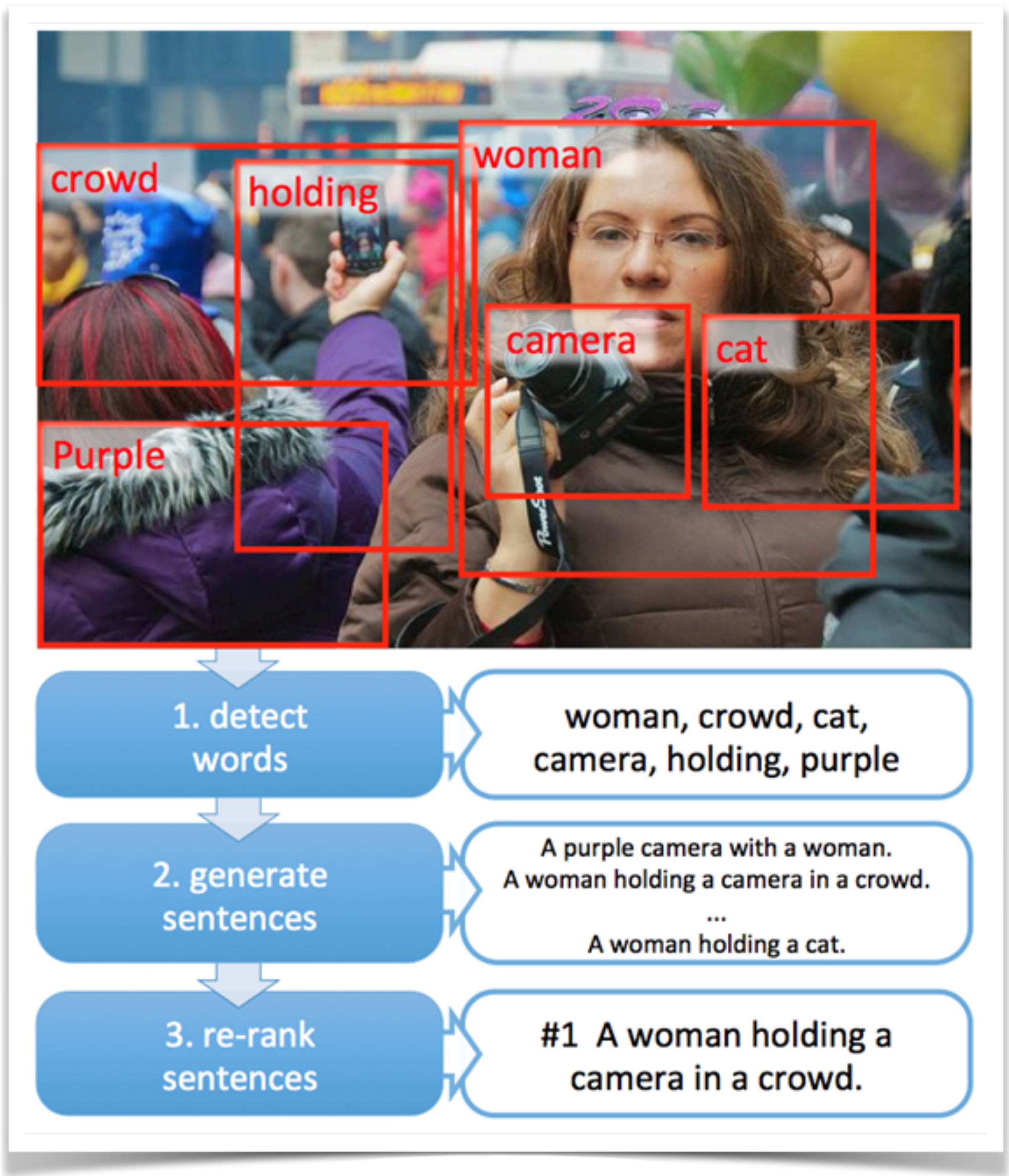
New applications: Non visual data prediction



What is Pulse?

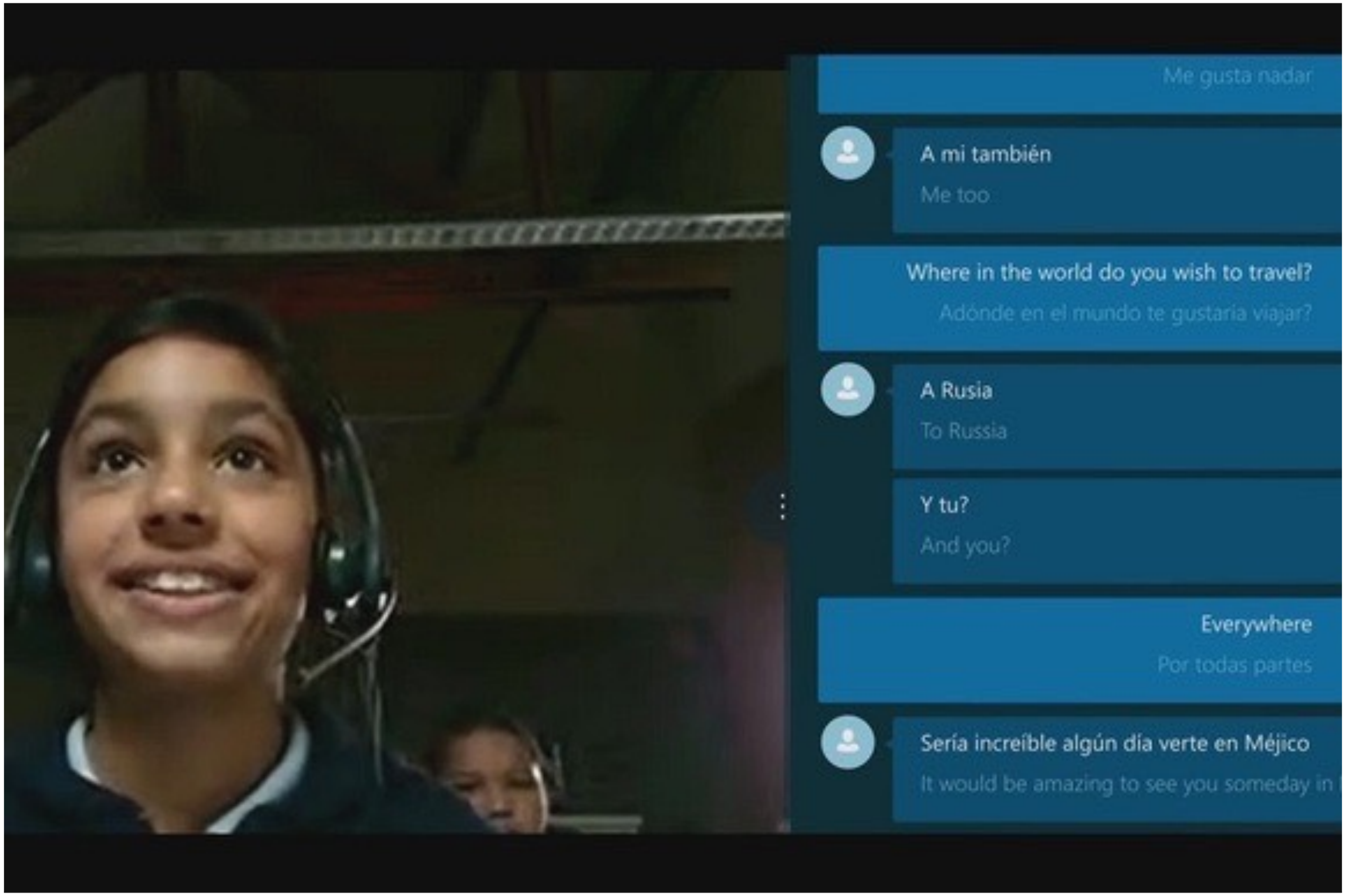
Pulse is a score out of 100 points that measures how **popular** a photo is. Pulse is calculated by an algorithm, which is unique to 500px and is based on votes (Likes & Favorites) on your photo from the community. The Pulse algorithm was designed to promote daily exposure of new photographs and photographers. It is not necessarily a measure of photograph's quality.

New applications: Automatic Image Captioning

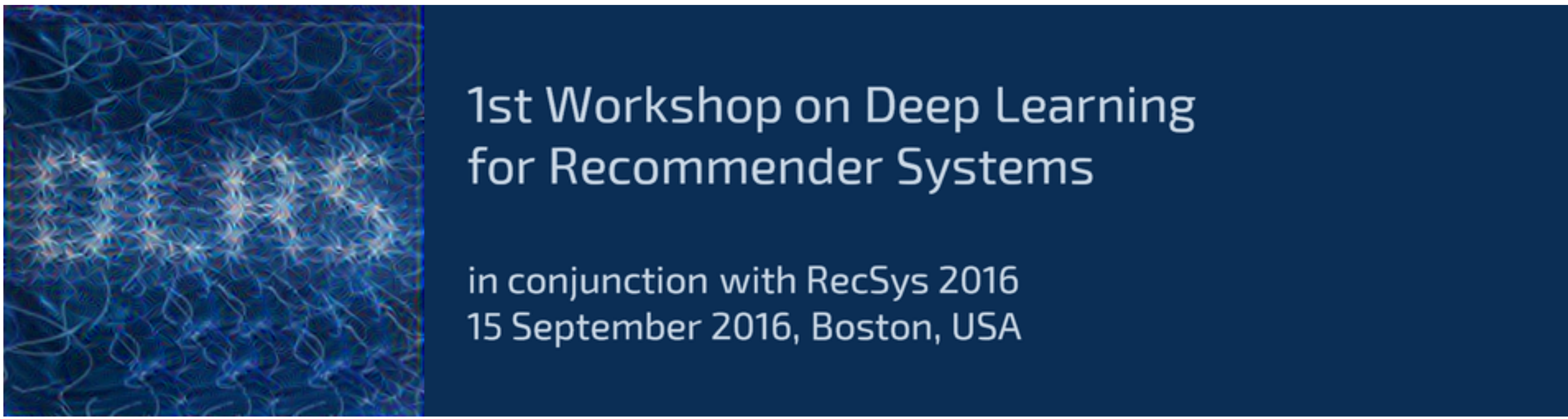


http://blogs.technet.com/b/machinelearning/archive/2014/11/18/rapid-progress-in-automatic-image-captioning.aspx

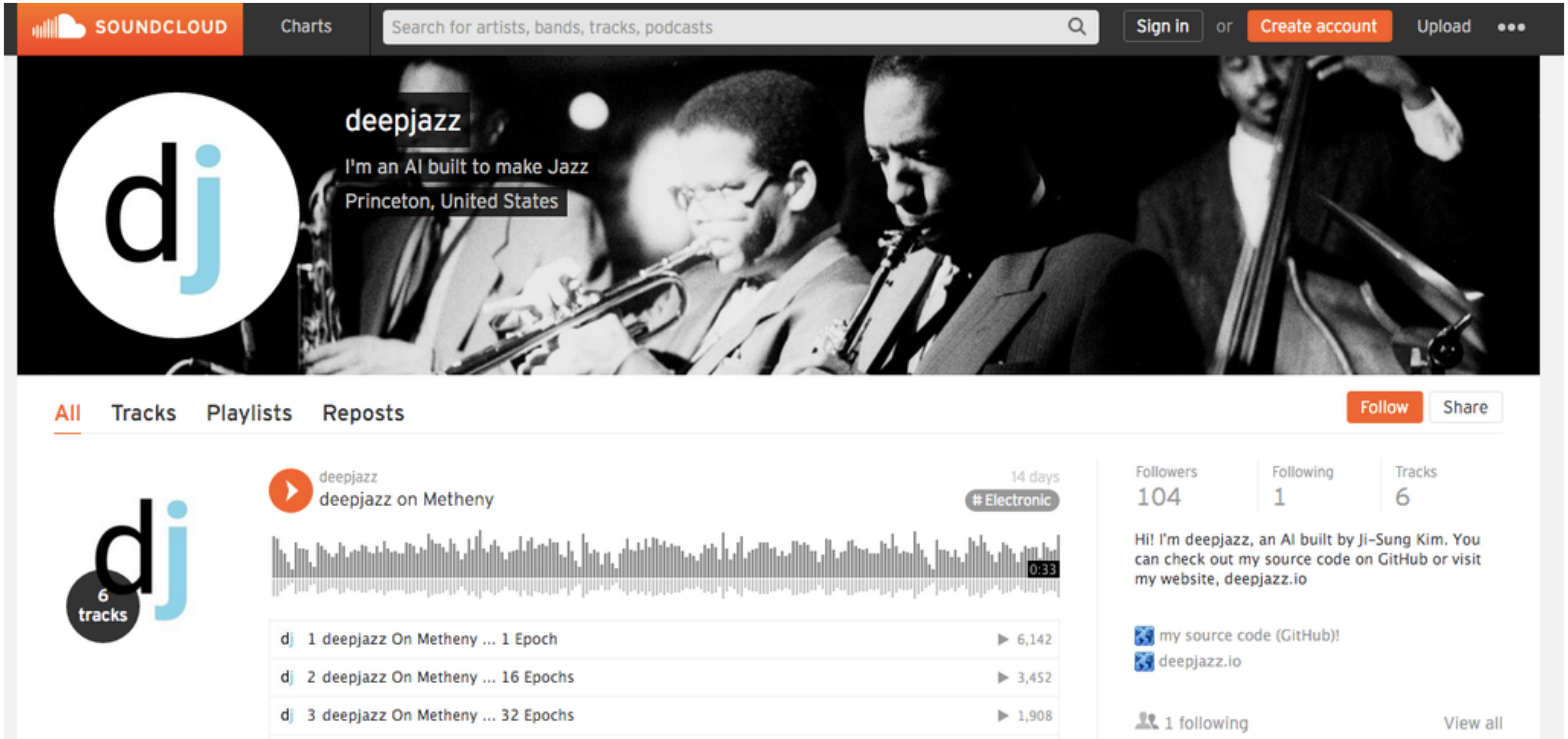
Speech translation



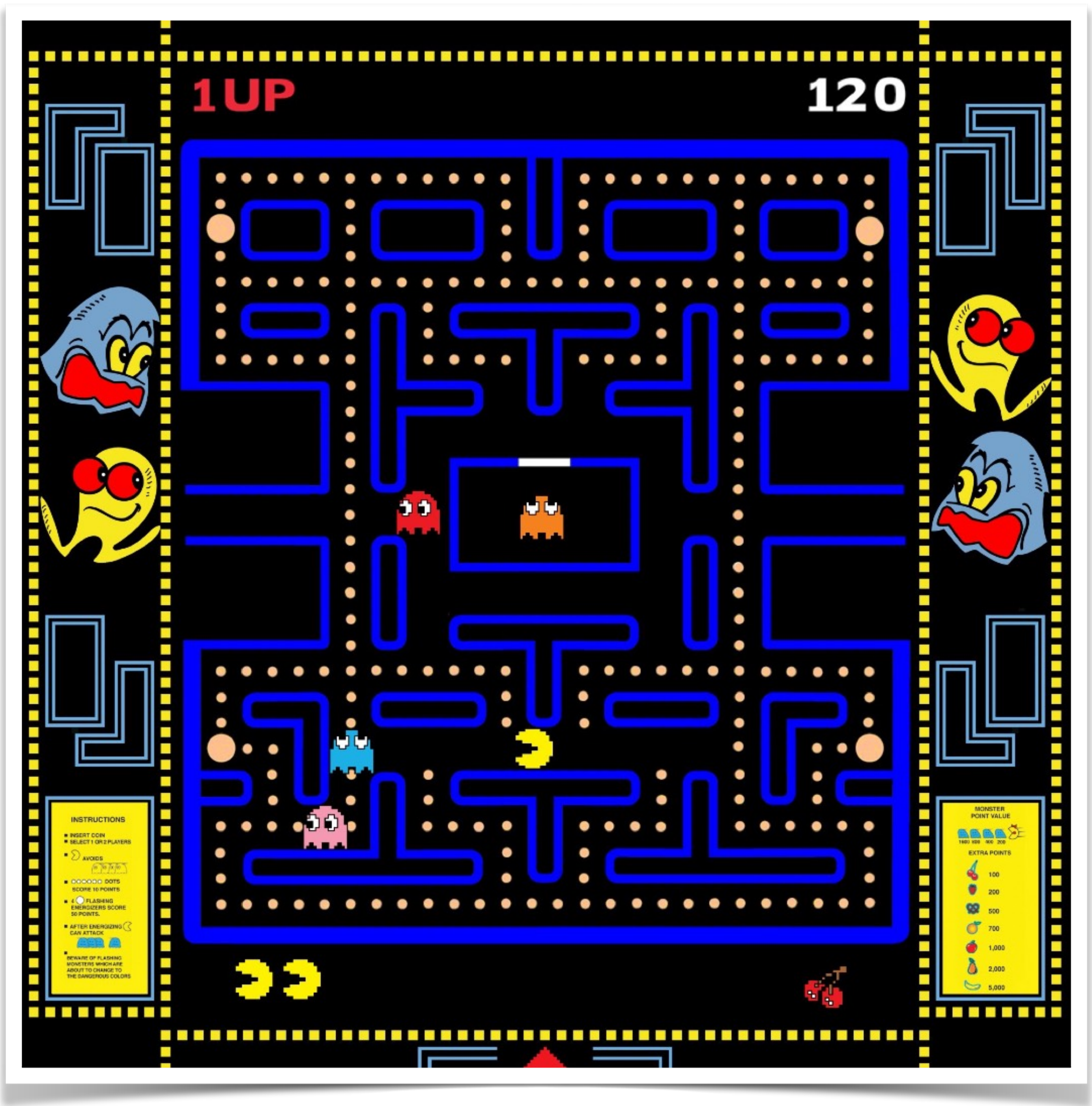
Recommenders



Music Generation



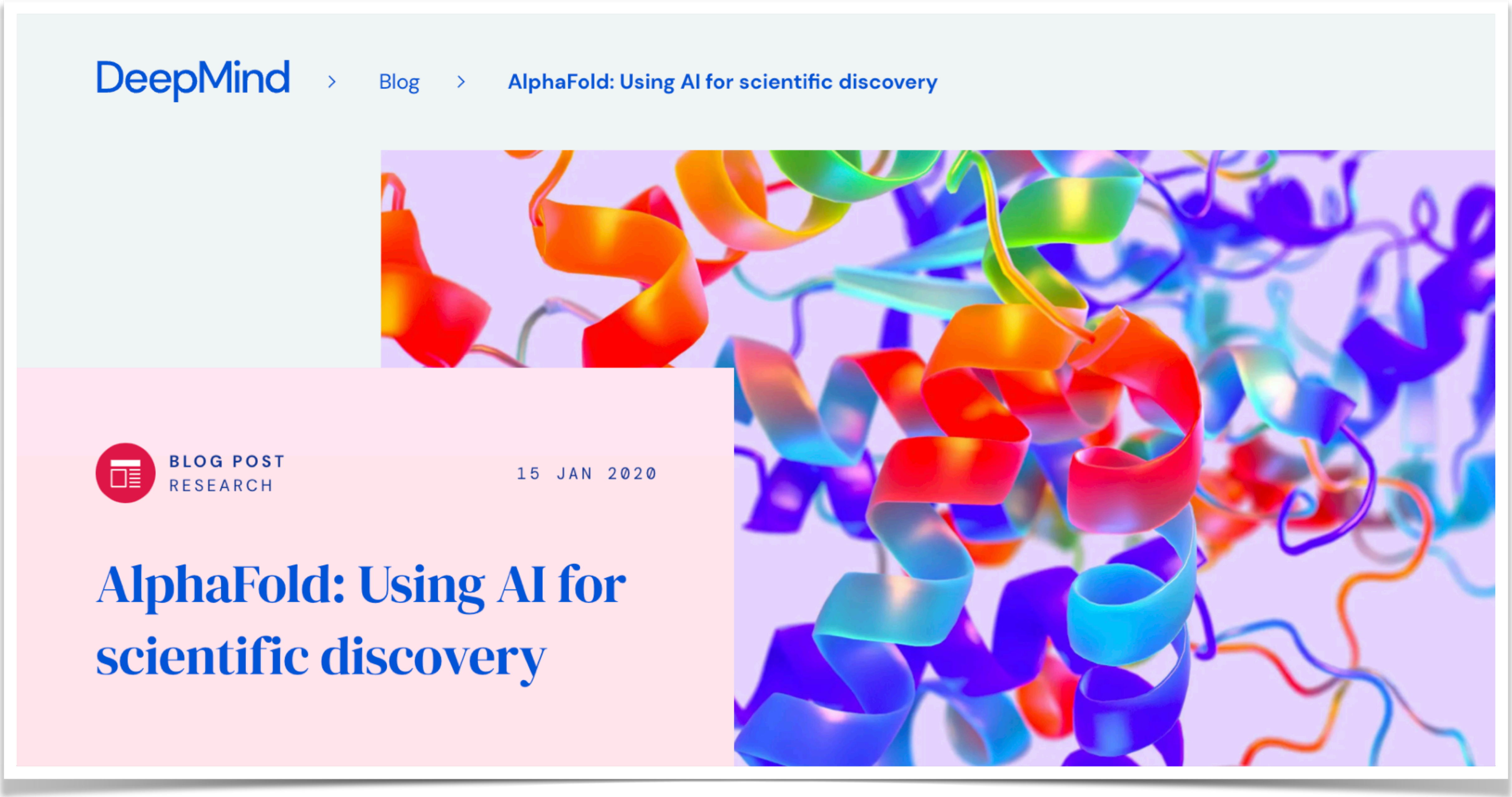
Reinforcement learning.



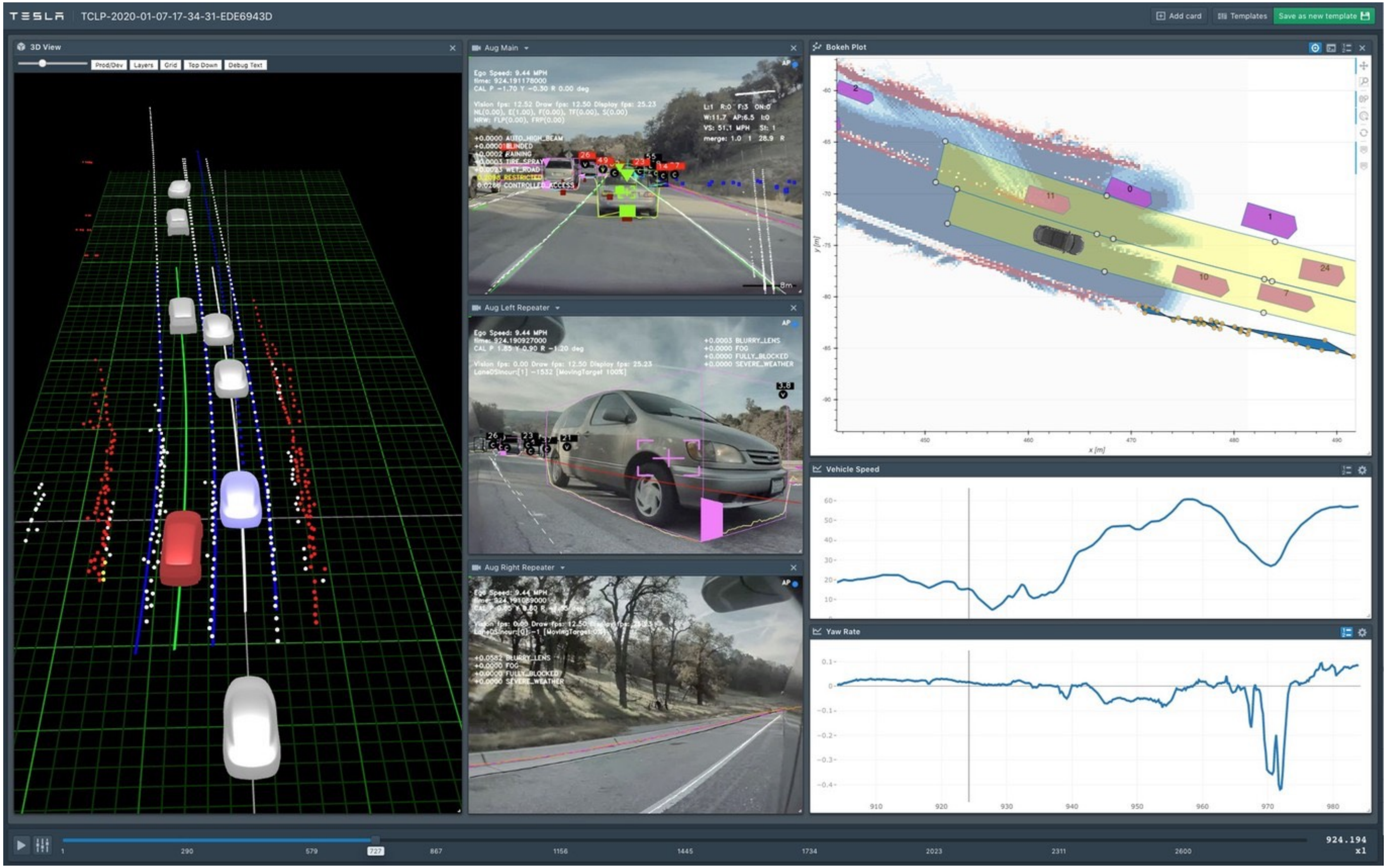
AlphaZero


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462
NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587

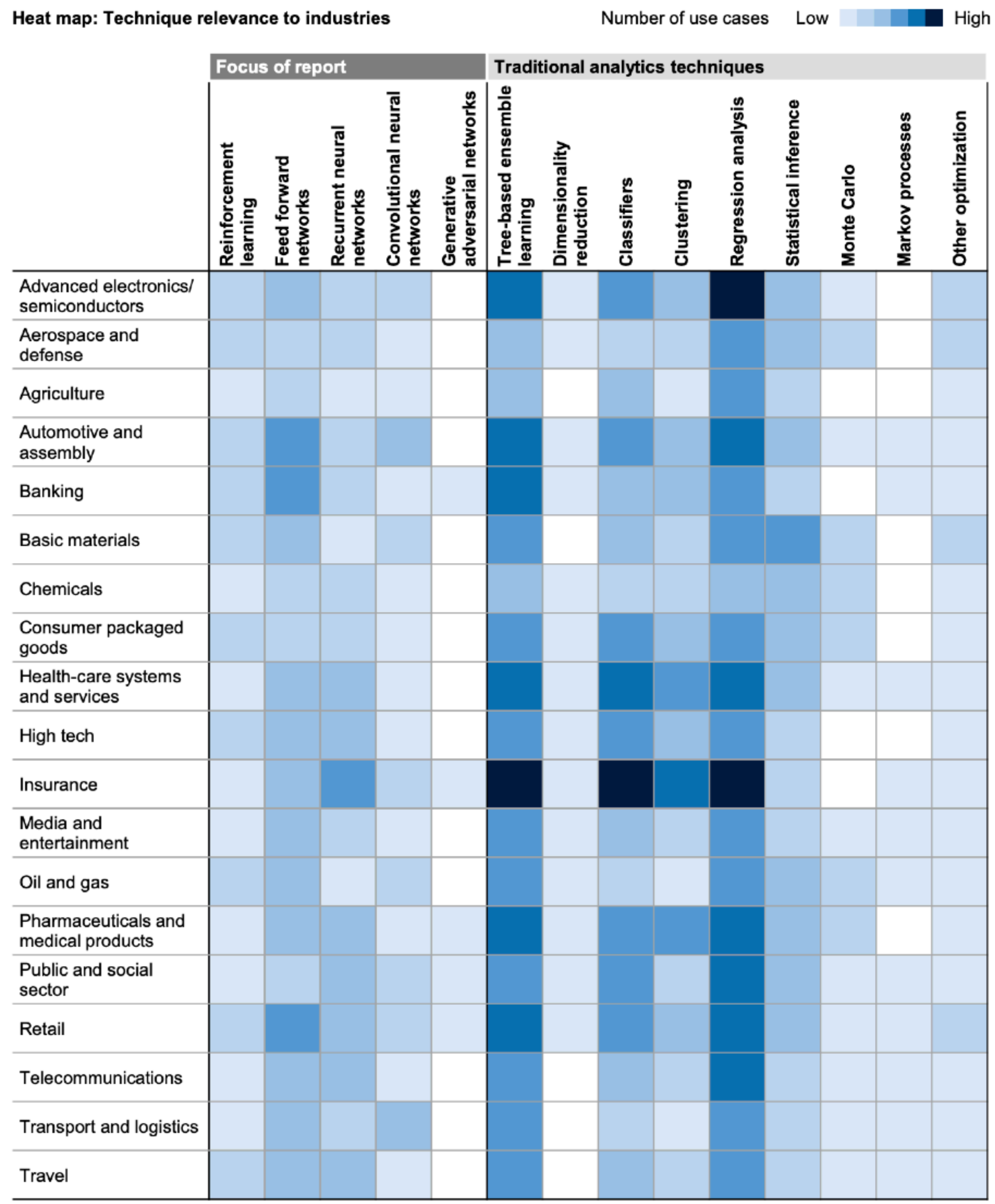
AlphaFold


DeepMind
Blog
AlphaFold: Using AI for scientific discovery
BLOG POST
RESEARCH
15 JAN 2020
AlphaFold: Using AI for scientific discovery

TESLA Autopilot

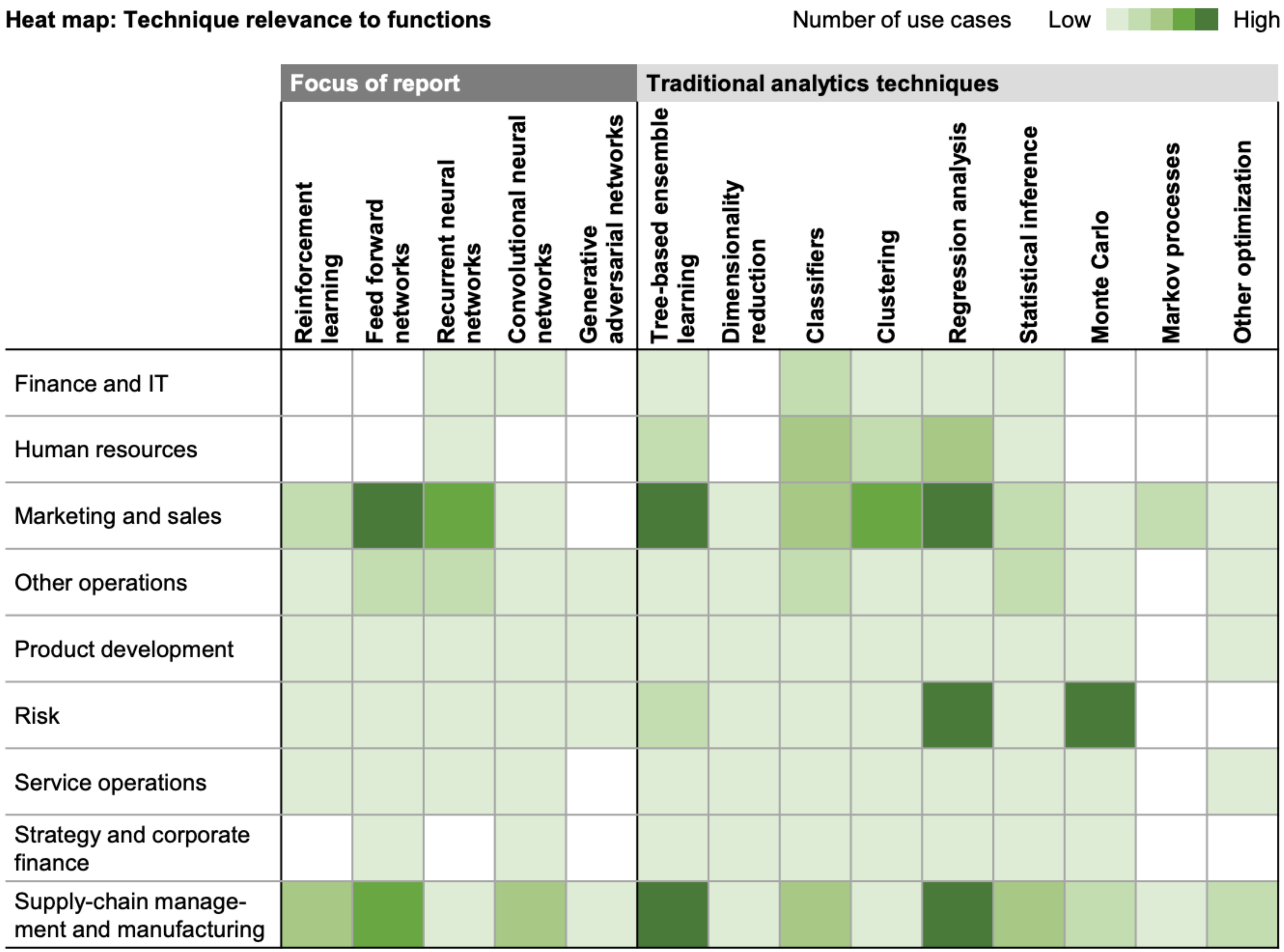


https://www.youtube.com/watch?time_continue=31&v=NGS2SNGXUXo&feature=emb_title



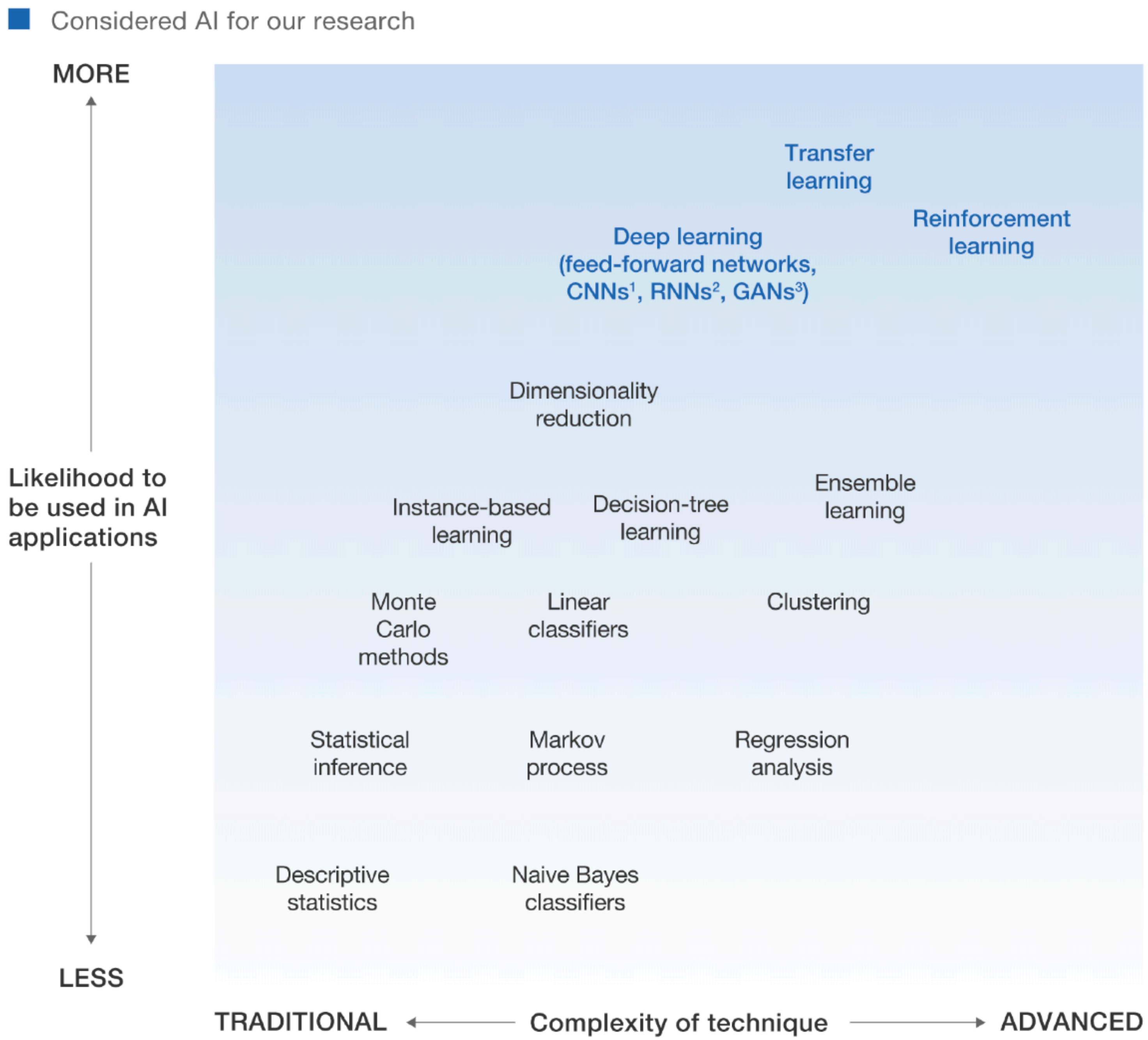
SOURCE: McKinsey Global Institute analysis

Heat map: Technique relevance to functions



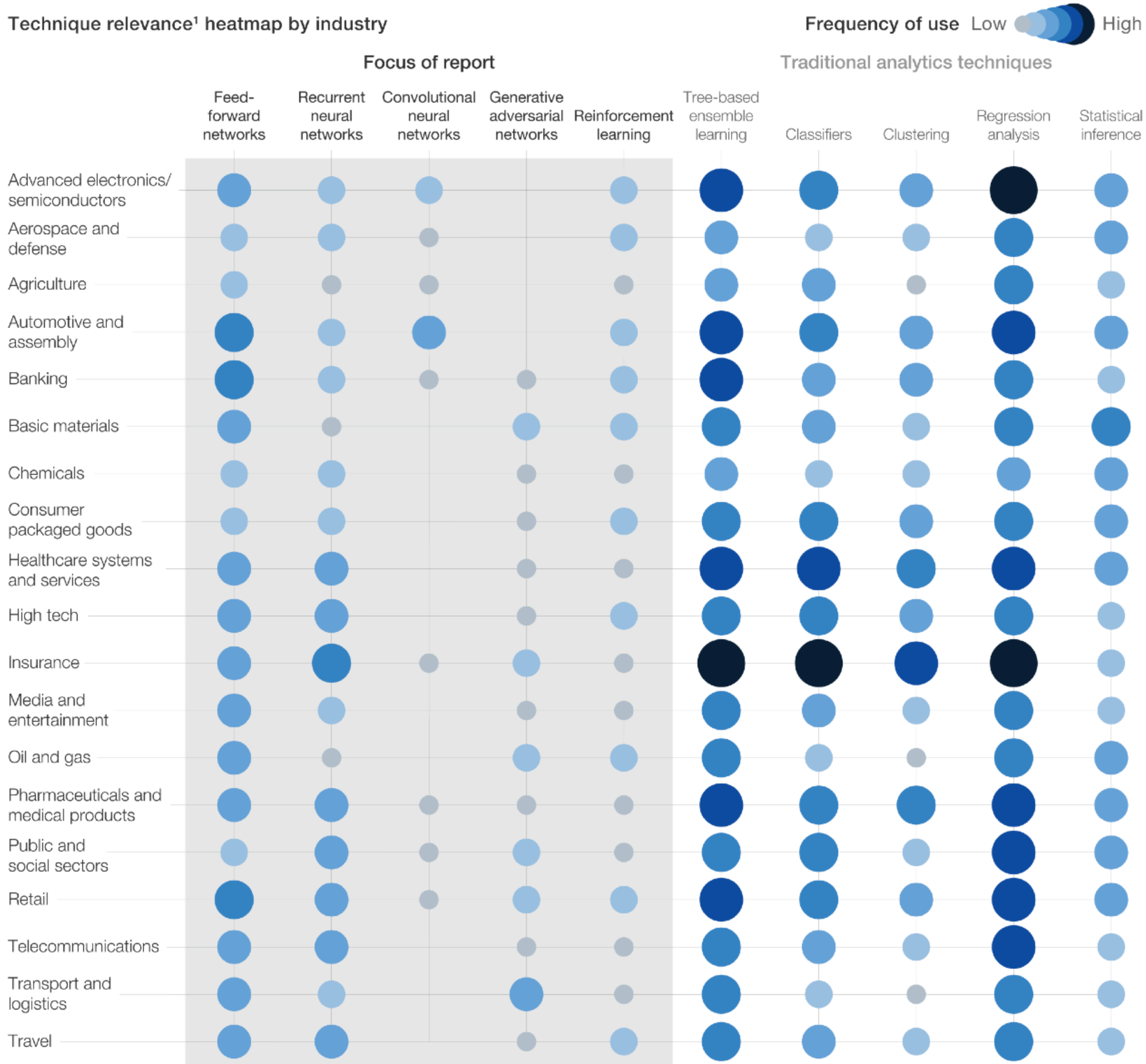
SOURCE: McKinsey Global Institute analysis

SOURCE: McKinsey Global Institute analysis



[1] Convolutional neural networks.
[2] Recurrent neural networks.
[3] Generative adversarial networks.

McKinsey&Company | **Source:** McKinsey Global Institute analysis



[1] Relevance refers to frequency of use in our use case library, with the most frequently found cases marked as high relevance and the least frequently found as low relevance. Absence of circles indicates no or statistically insignificant number of use cases.
Note: List of techniques is not exhaustive.

McKinsey&Company | **Source:** McKinsey Global Institute analysis

https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning

What is Deep Learning?

UNIVERSITAT DE BARCELONA

Learning from Data

Training data: a set of $(x^{(m)}, y^{(m)})$ pairs.
Learn a function $f_w : x \to y$ to predict on new inputs x.

1. Choose a model function family f_w.

2. Optimize parameters w.

1-layer neural net model

Parameters

Weights

Bias

$$f(x) = \sigma(w^T \cdot x + b)$$

Sigmoid Function

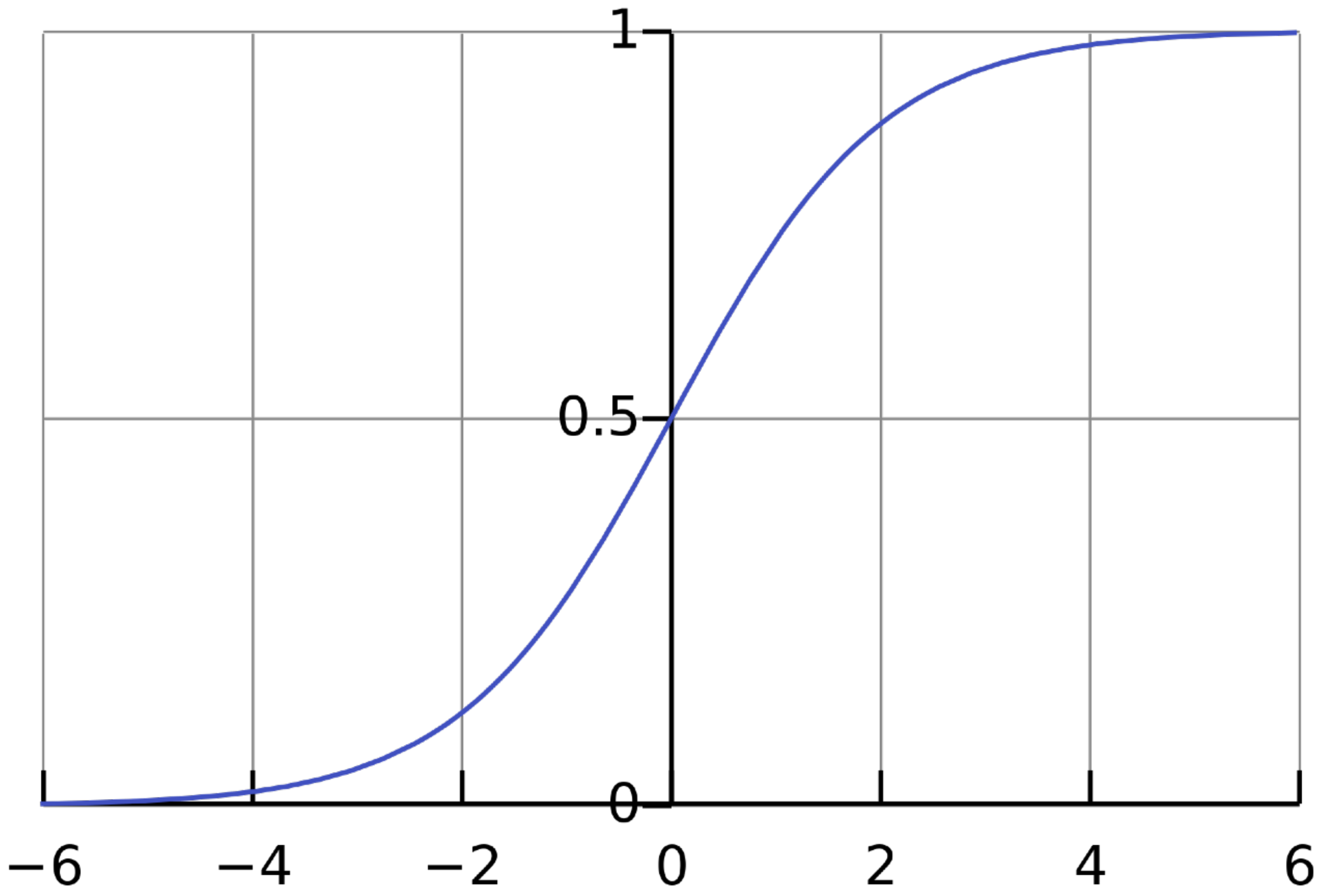
$$\sigma(x) = \frac{1}{(1+e^{-x})}$$

Dot Product

ReLU Function

$$\sigma(x) = max(0, x)$$

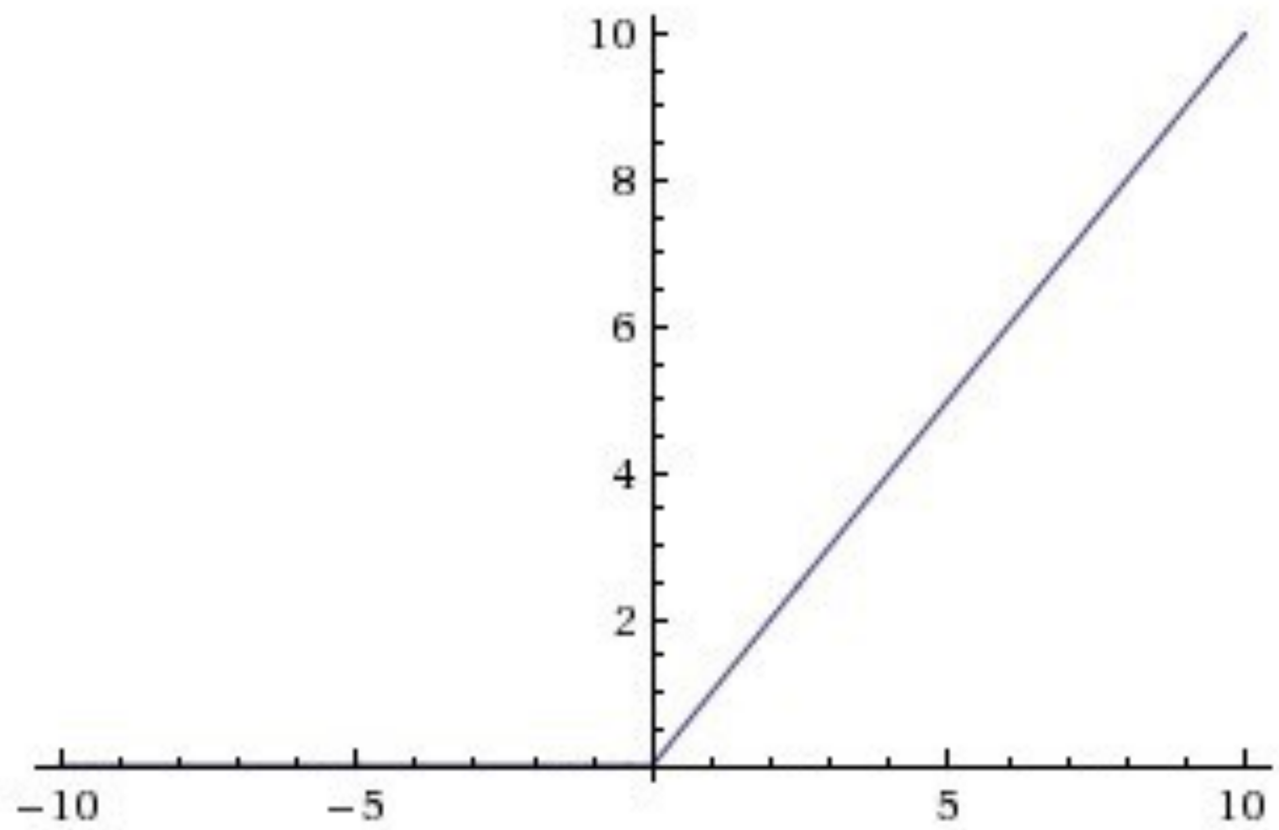
1-layer neural net model



$$\sigma(x) = \frac{1}{(1+e^{-x})}$$

UNIVERSITAT DE BARCELONA

1-layer neural net model



$$\sigma(x) = max(0, x)$$

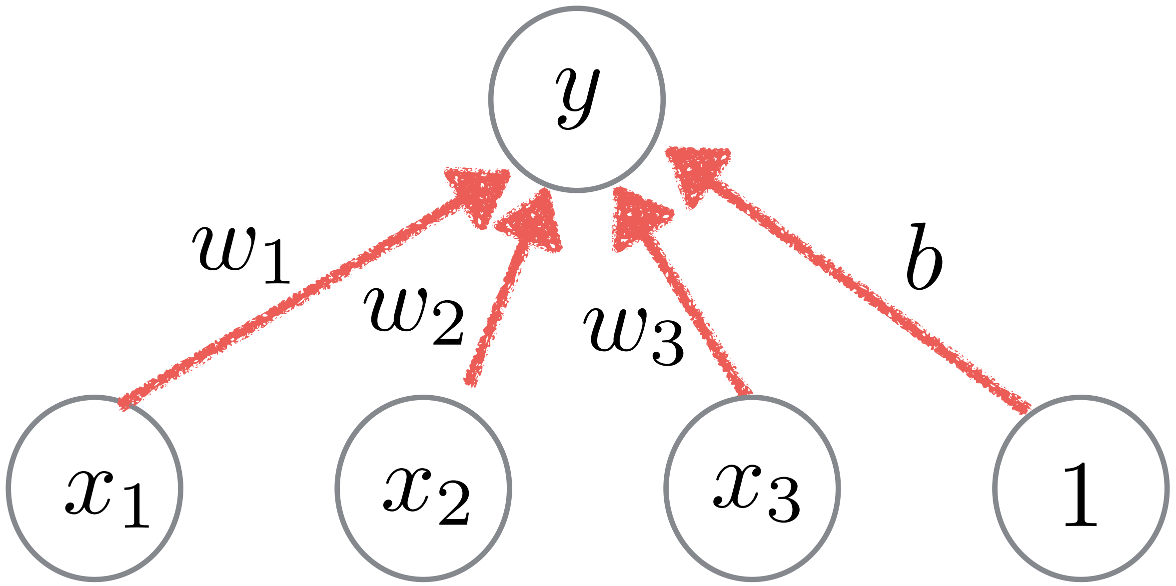
Table 3: Non-linearities tested.

Name	**Formula**	**Year**
none	$y = x$	-
sigmoid	$y = \frac{1}{1+e^{-x}}$	1986
tanh	$y = \frac{e^{2x}-1}{e^{2x}+1}$	1986
ReLU	$y = \max(x, 0)$	2010
(centered) SoftPlus	$y = \ln(e^x + 1) - \ln 2$	2011
LReLU	$y = \max(x, \alpha x)$, $\alpha \approx 0.01$	2011
maxout	$y = \max(W_1 x + b_1, W_2 x + b_2)$	2013
APL	$y = \max(x, 0) + \sum_{s=1}^{S} a_i^s \max(0, -x + b_i^s)$	2014
VLReLU	$y = \max(x, \alpha x)$, $\alpha \in 0.1, 0.5$	2014
RReLU	$y = \max(x, \alpha x)$, $\alpha = \mathrm{random}(0.1, 0.5)$	2015
PReLU	$y = \max(x, \alpha x)$, α is learnable	2015
ELU	$y = x$, if $x \geq 0$, else $\alpha(e^x - 1)$	2015

1-layer neural net model

$$f(x) = \sigma(w^T \cdot x + b)$$

1-dimensional NN → 2 parameters

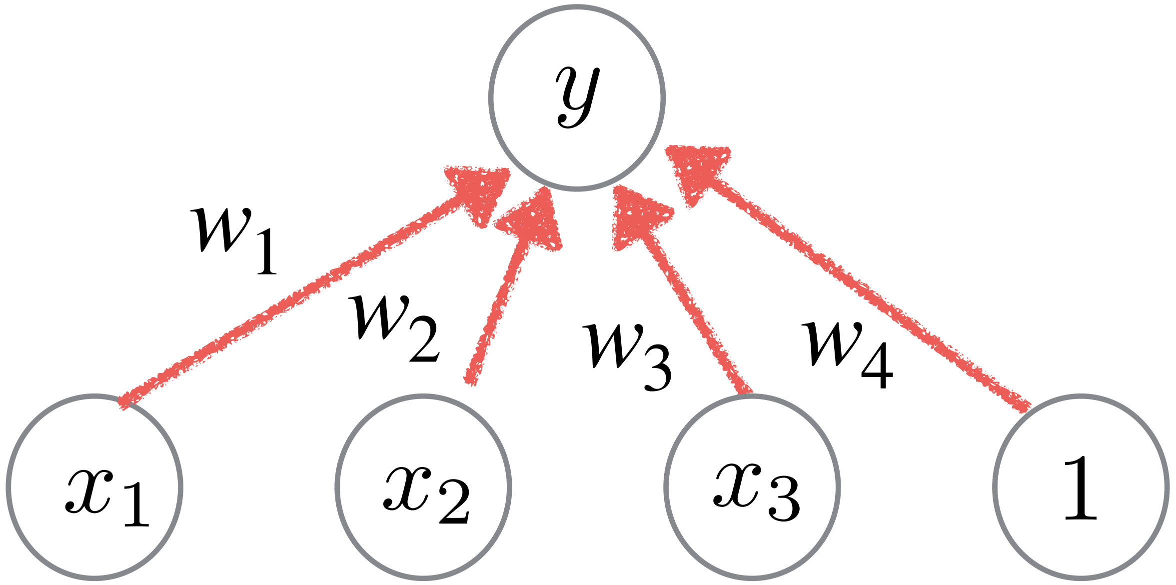


Graphical representation for 3-dimensional data → 4 parameters

1-layer neural net model **rewriting**

$$f(x) = \sigma(w^T \cdot x + b) \rightarrow f(\mathbf{x}') = \sigma(\mathbf{w}^T \cdot \mathbf{x}')$$

where $\mathbf{x}'$ is $(x', 1)$



In this way we simplify formulas!

1-layer neural net model

Training a network consist of updating the vector w values until our equation fits the training data as well as possible.

Such updates happen by randomly picking one of the examples and determining if it is a miss-classification. You have a misclassified example when the perceptron determines that an example is part of the class, but it isn't, or when the perceptron determines an example isn't part of the class, but it is.

Training handles one misclassified example at a time and operates by changing the w vector using a simple weighted addition:

$$w_i = w_i + \alpha(x_i \cdot y_i)$$

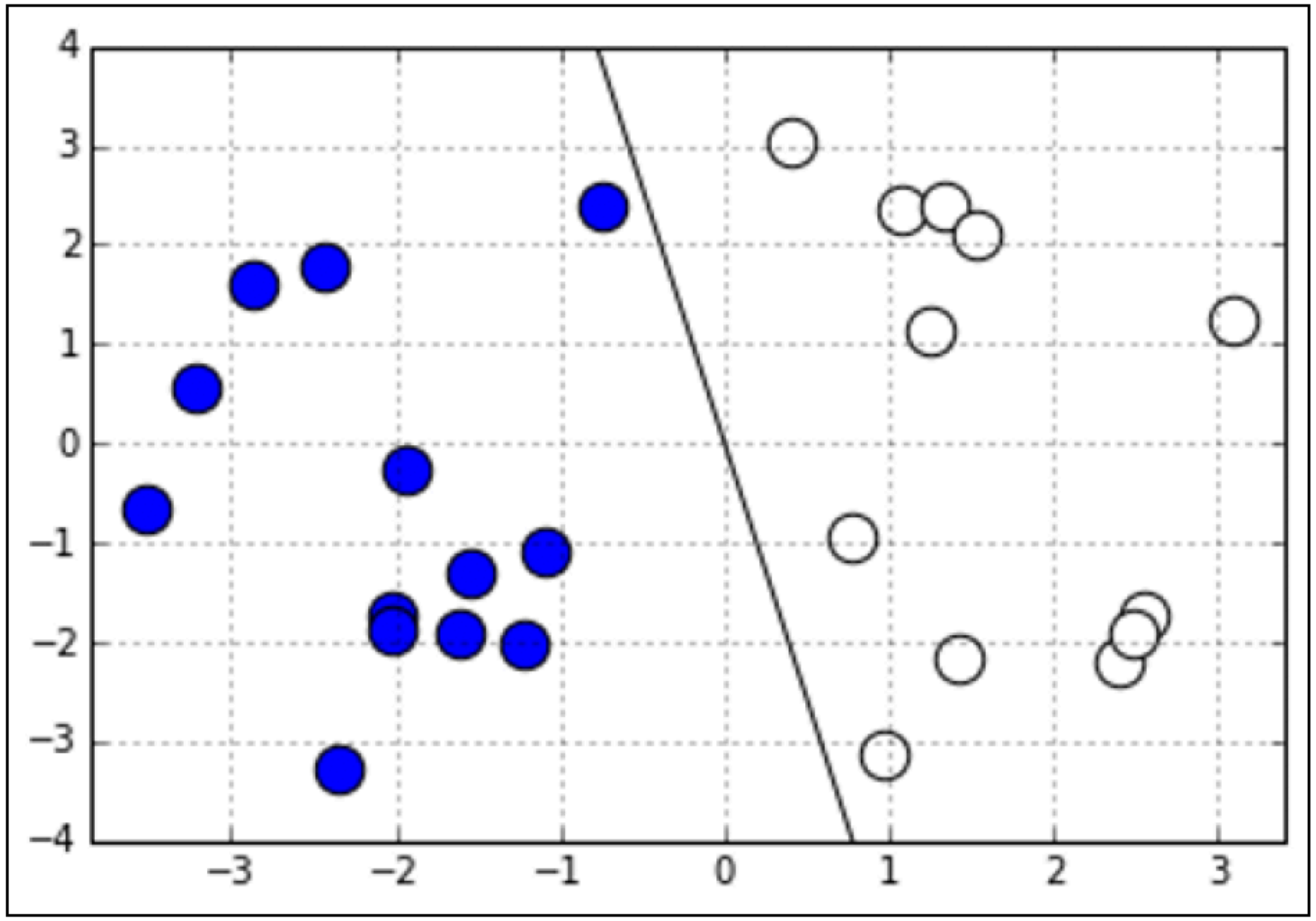
1-layer neural net model

The 2D network is nothing more than a line trying to separate the positive class from the negative one.

Initially, when w is set to zero or to random values, the separating line is just one of the infinite possible lines found on a plane

The updating phase defines it by forcing it to become nearer to the misclassified point. As the algorithm passes through the misclassified examples, it applies a series of corrections. In the end, using multiple iterations to define the errors, the algorithm places the separating line at the exact border between the two classes.

1-layer neural net model



This is an optimal boundary!

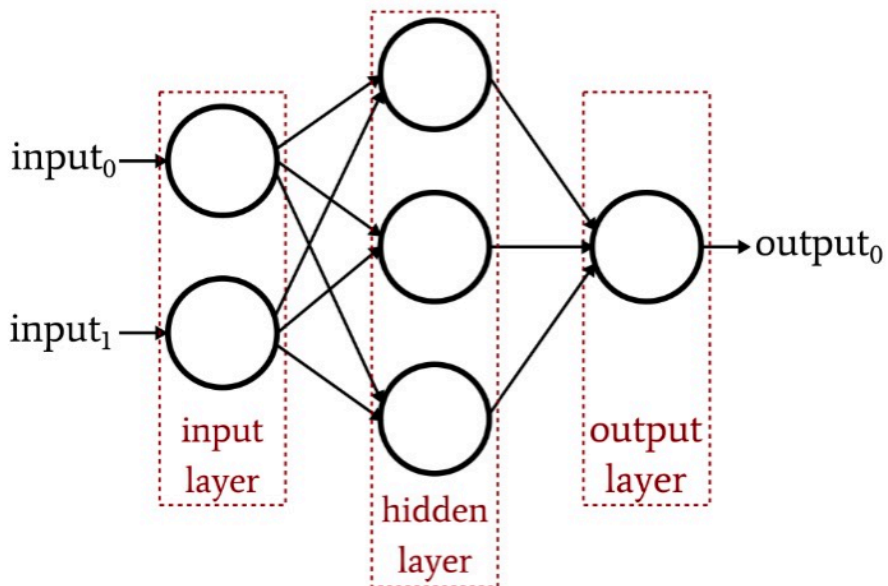
1-layer neural net model

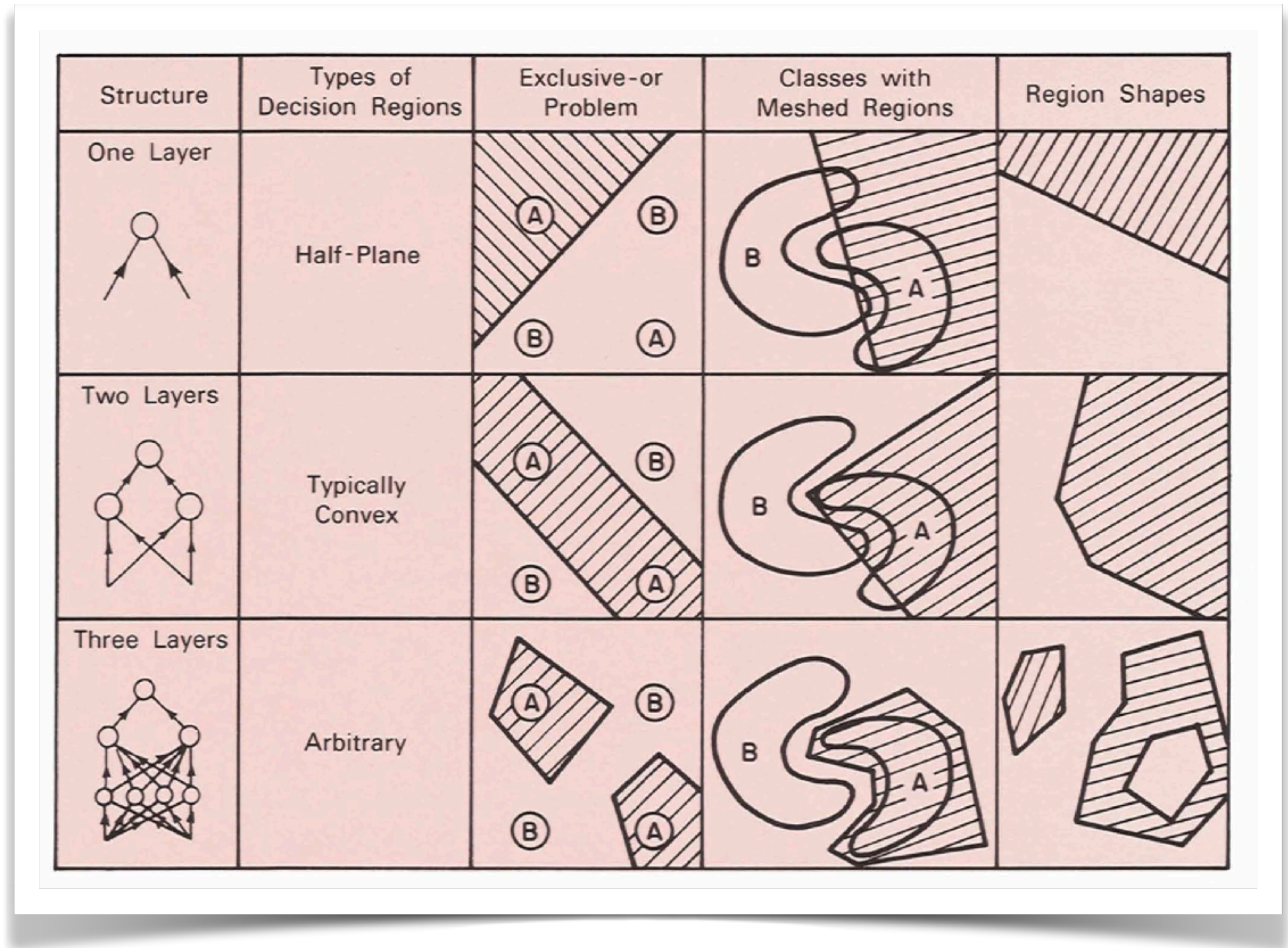
If you can't divide two classes spread on two or more dimensions by any line or plane, they're **nonlinearly separable**. Overcoming data's being nonlinearly separable is one of the challenges of main challenges of ML.

In our case, this can be solved by using **multilayer perceptrons**.

2-layer neural net model

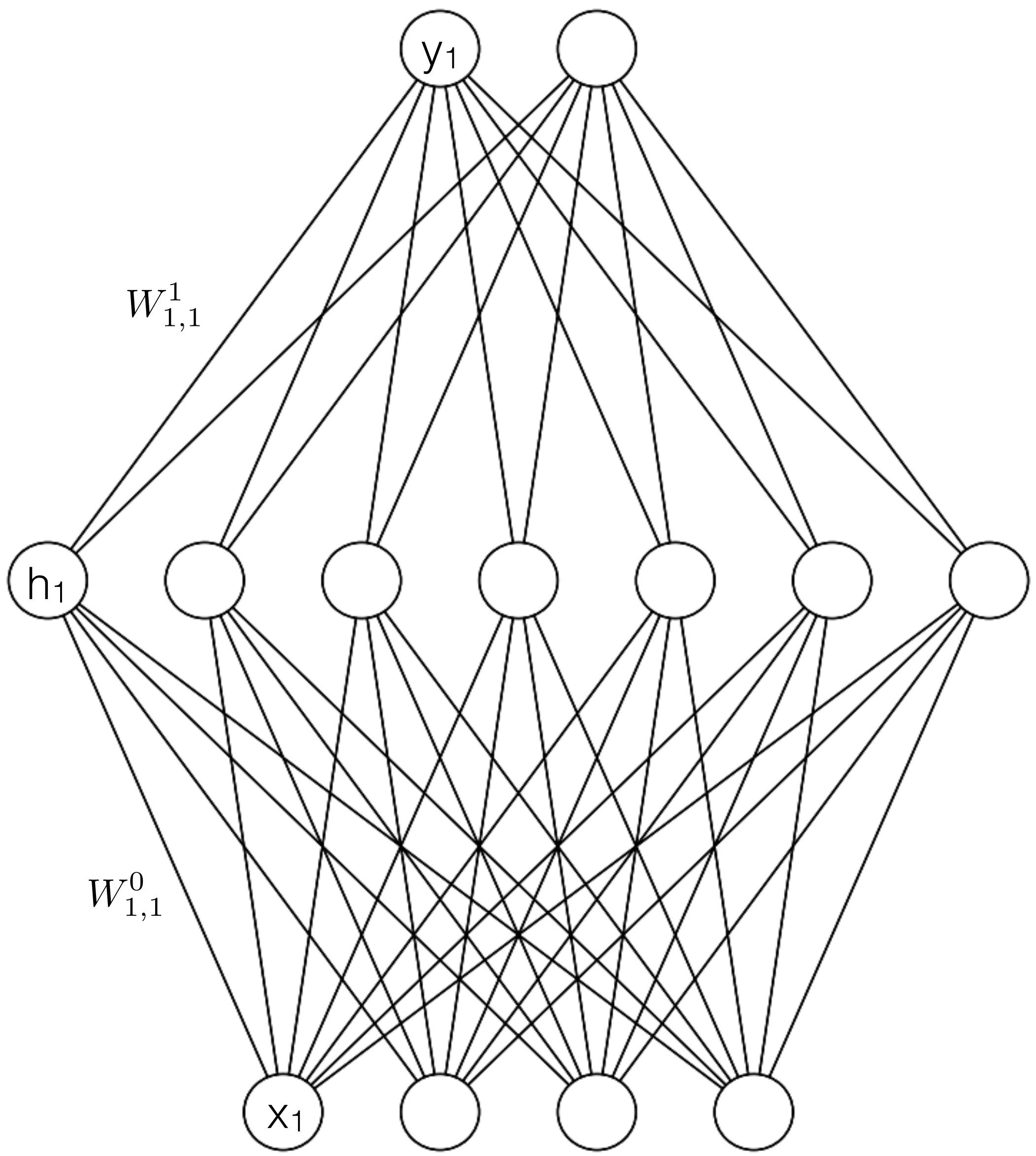
Now we have more parameters!

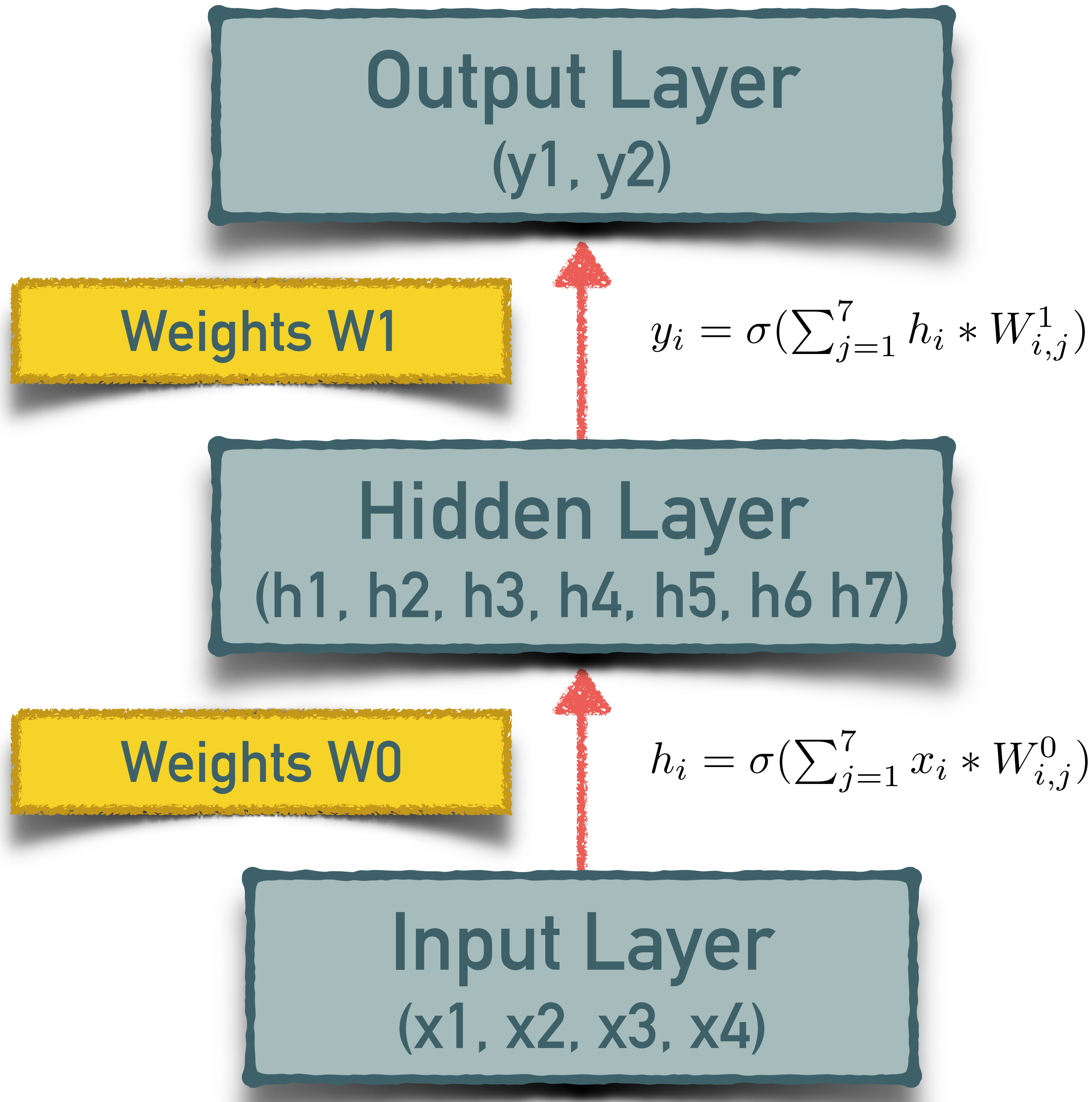




https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.semanticscholar.org%2Fpaper%2FNeural-Network-Classifiers-for-Speech-Recognition-P..%2F7e73bffddc02223a8037e07da898cb194a64a619%2Ffigure%2F4&psig=AOvVaw04Iny0XJdW373JtlK_6TCD&ust=1601564371382000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCMiWwrGSkewCFQAAAAAdAAAAABAV

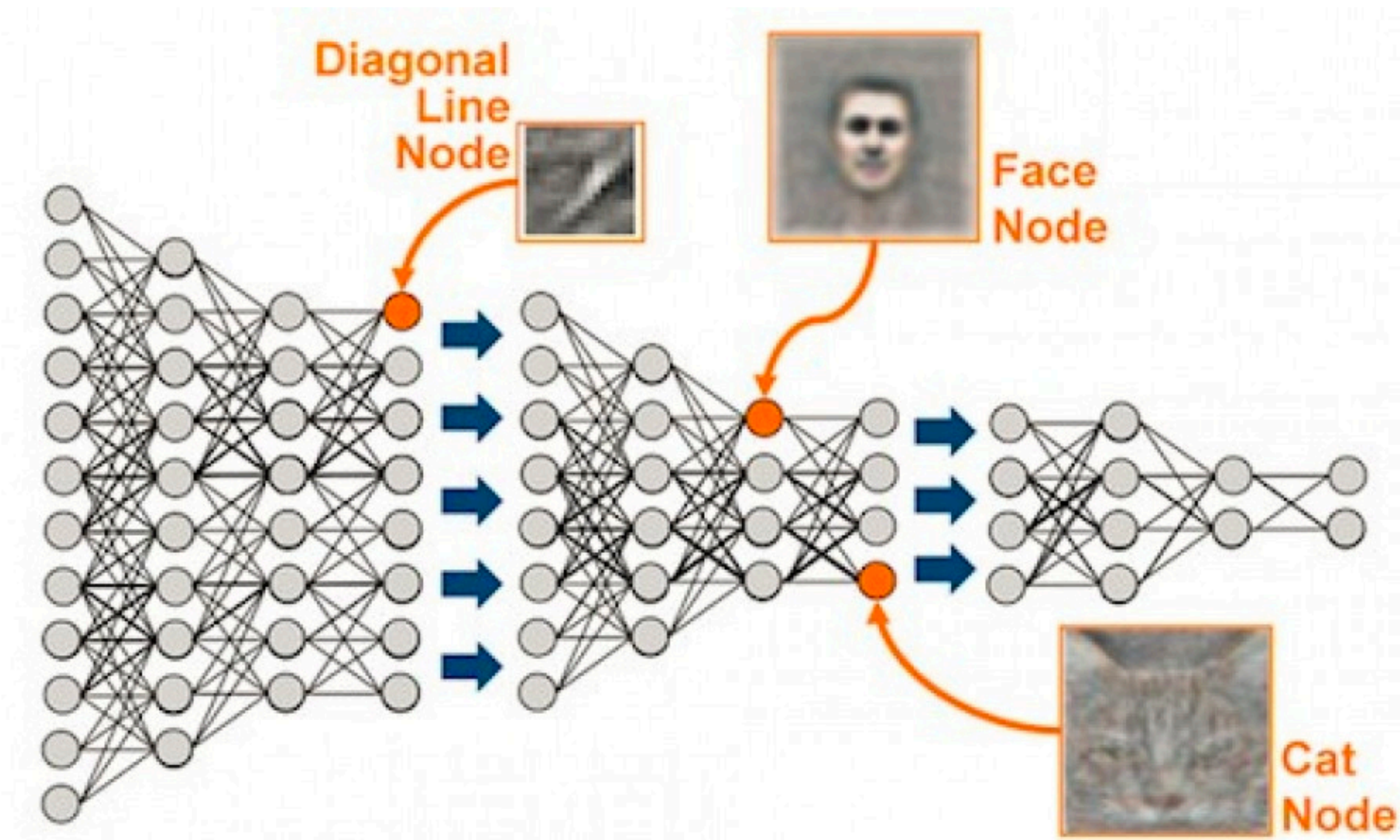
2-layer neural net model



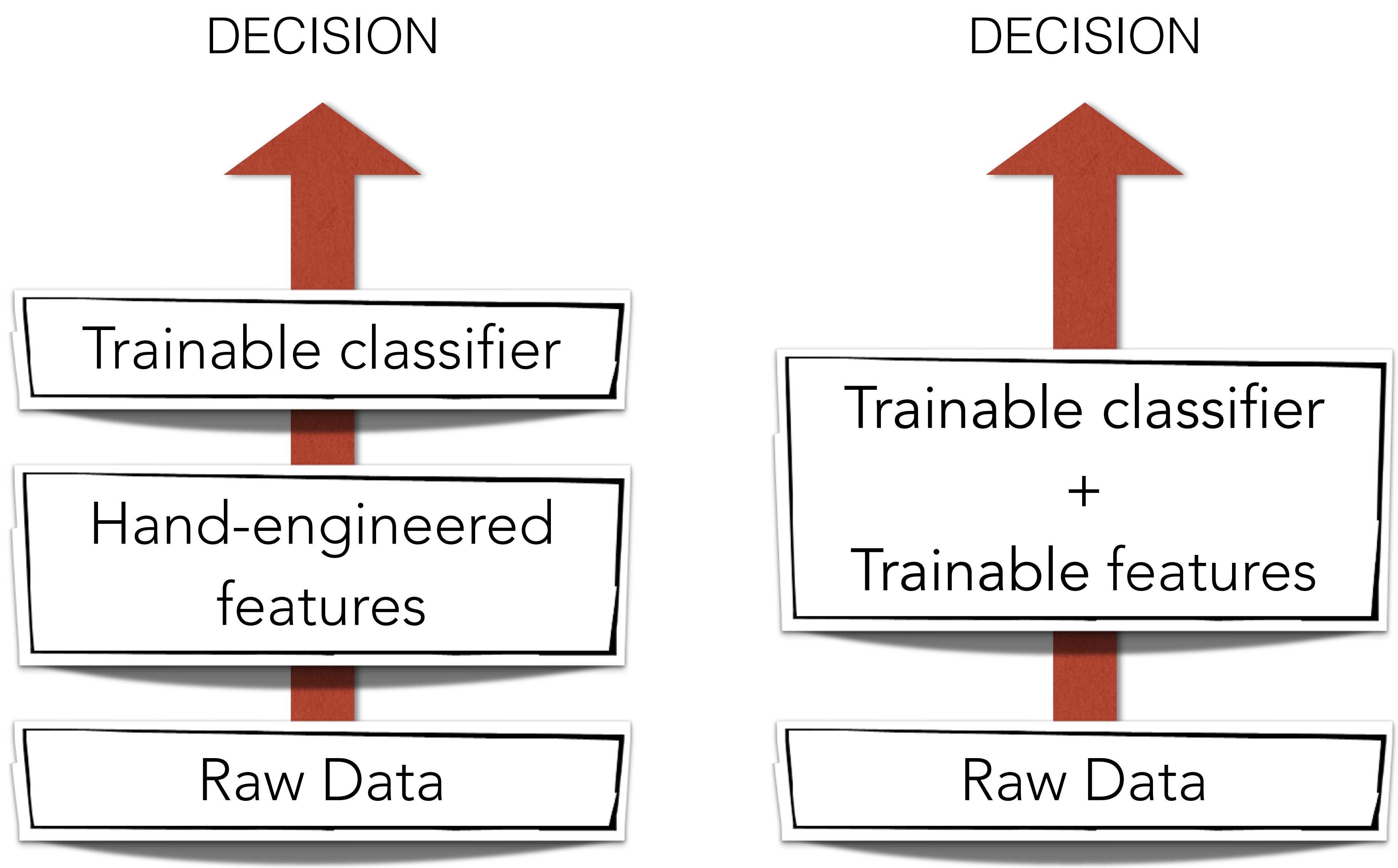


Computing the activation of one layer from the previous one can be written as a matrix-vector multiplication!

Deeper is better!

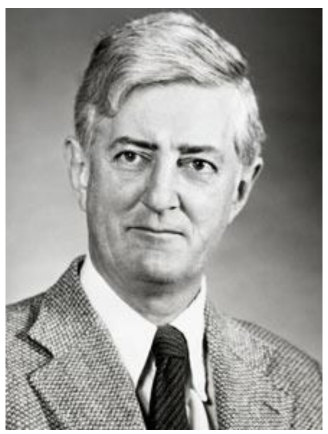


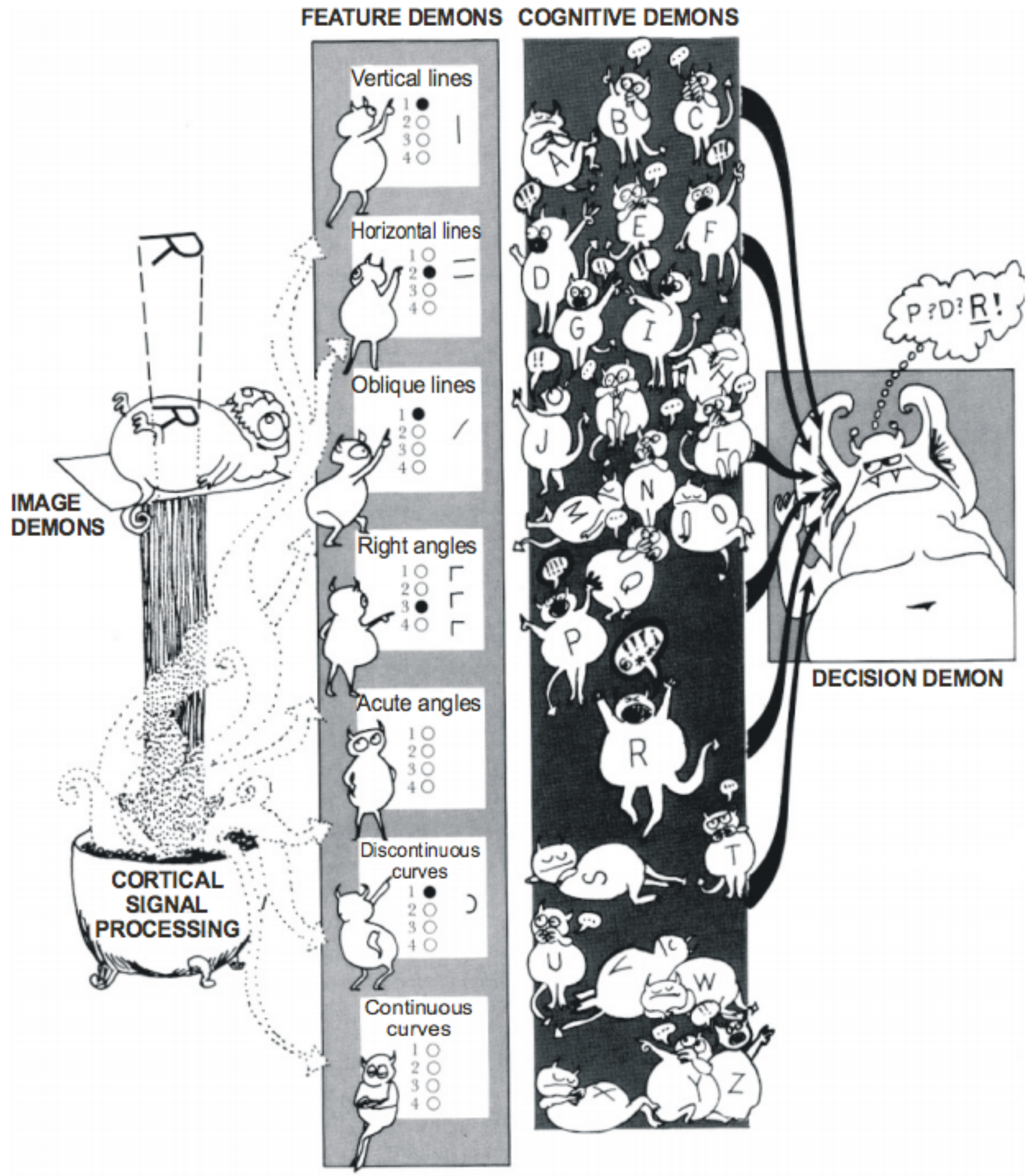
http://sdat.ir/en/sdat-blog/item/1-deep-learning



STANDARD MACHINE LEARNING

DEEP LEARNING





Pandemonium

An early architecture of vision, proposed by Oliver Selfridge in 1959.

Hype is not new

1958 New York Times…

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.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.