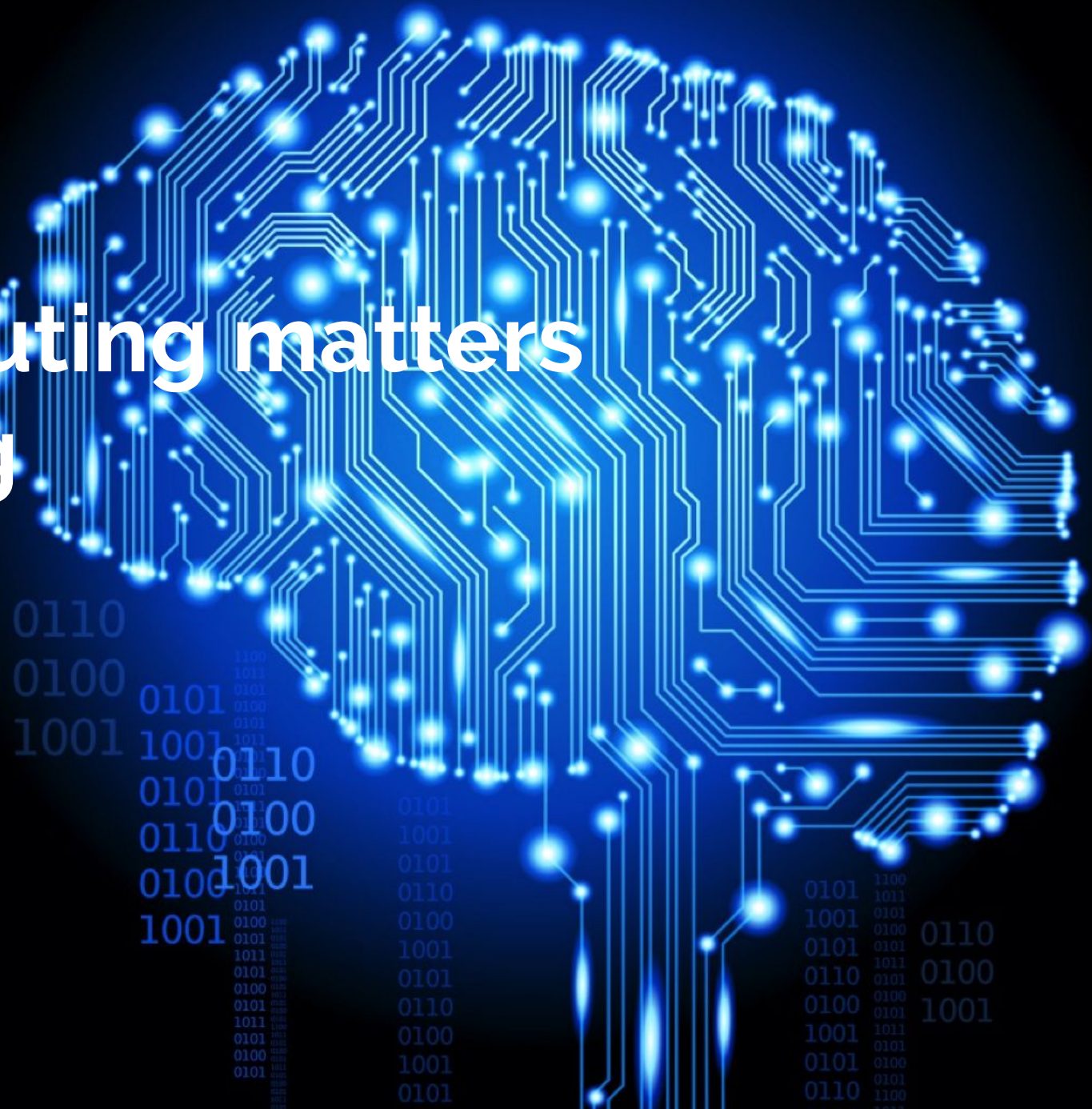
# Why Supercomputing matters to Deep Learning

## ESADE – MIBA (FALL 2017)



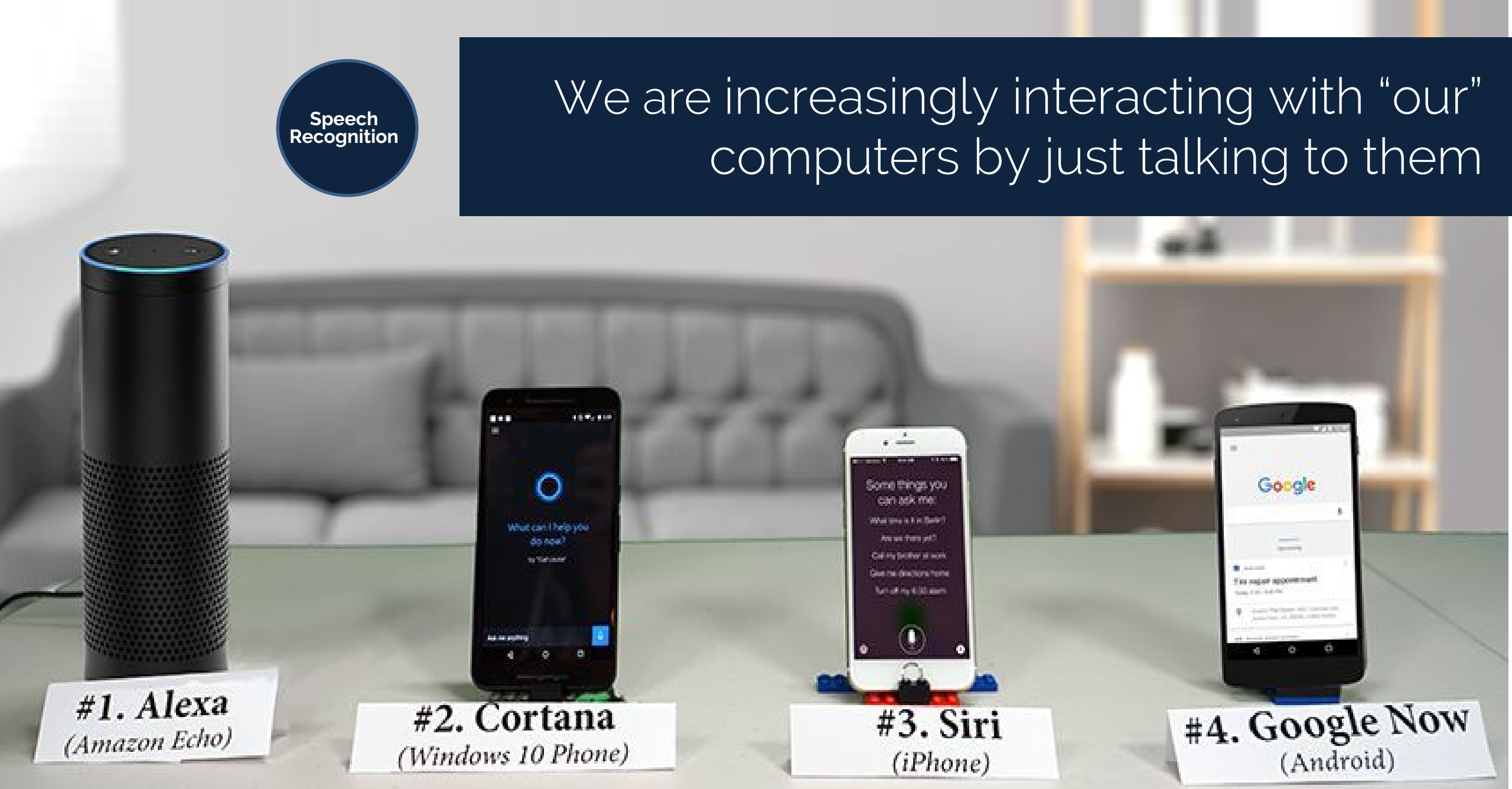
JORDI **TORRES** | FRANCESC **SASTRE**

# Artificial Intelligence is changing our life


# Quantum leaps in the quality of a wide range of everyday technologies thanks to Artificial Intelligence

Speech Recognition

# We are increasingly interacting with "our" computers by just talking to them

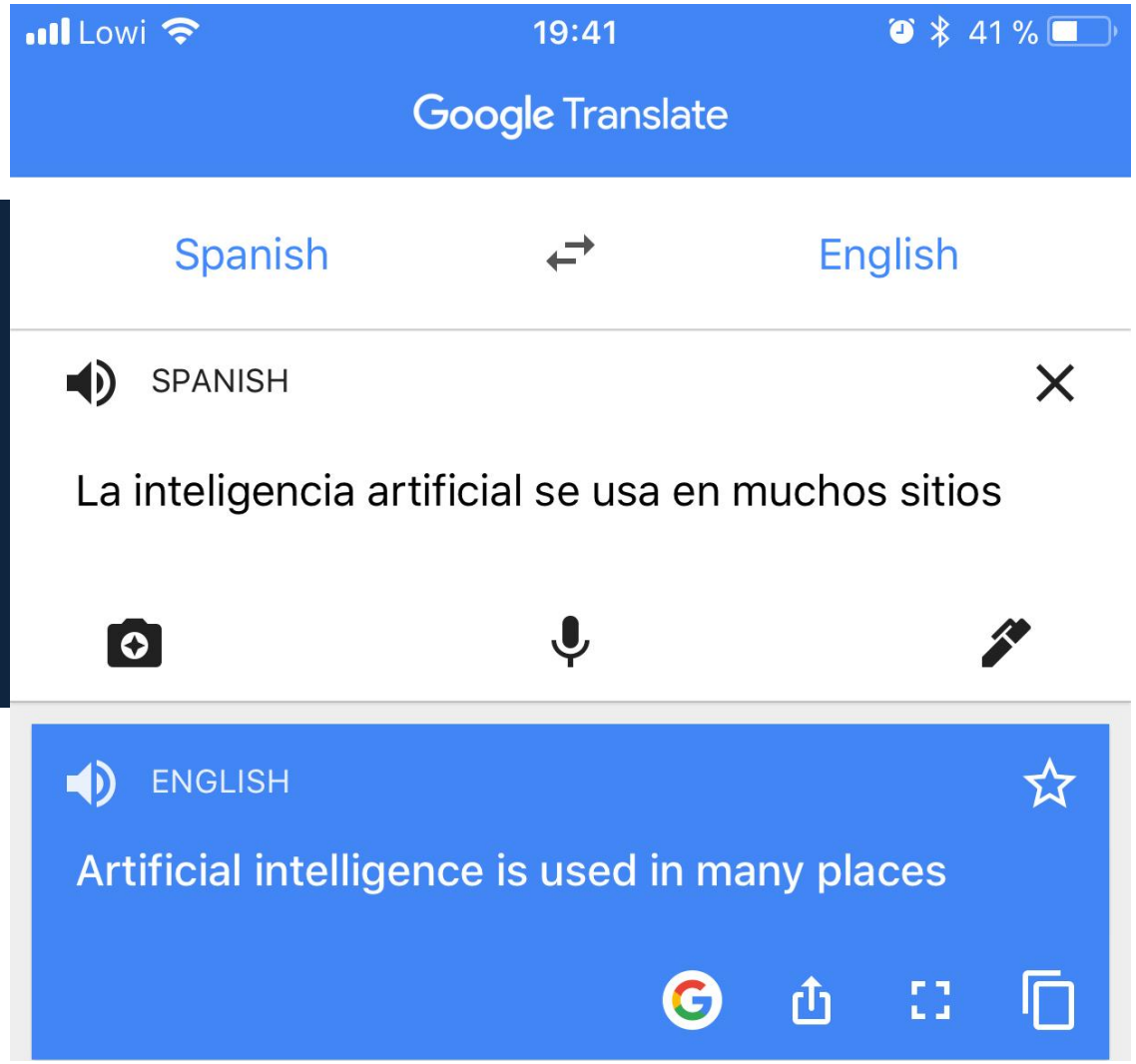





Credits: https://www.yahoo.com/tech/battle-of-the-voice-assistants-siri-cortana-211625975.html

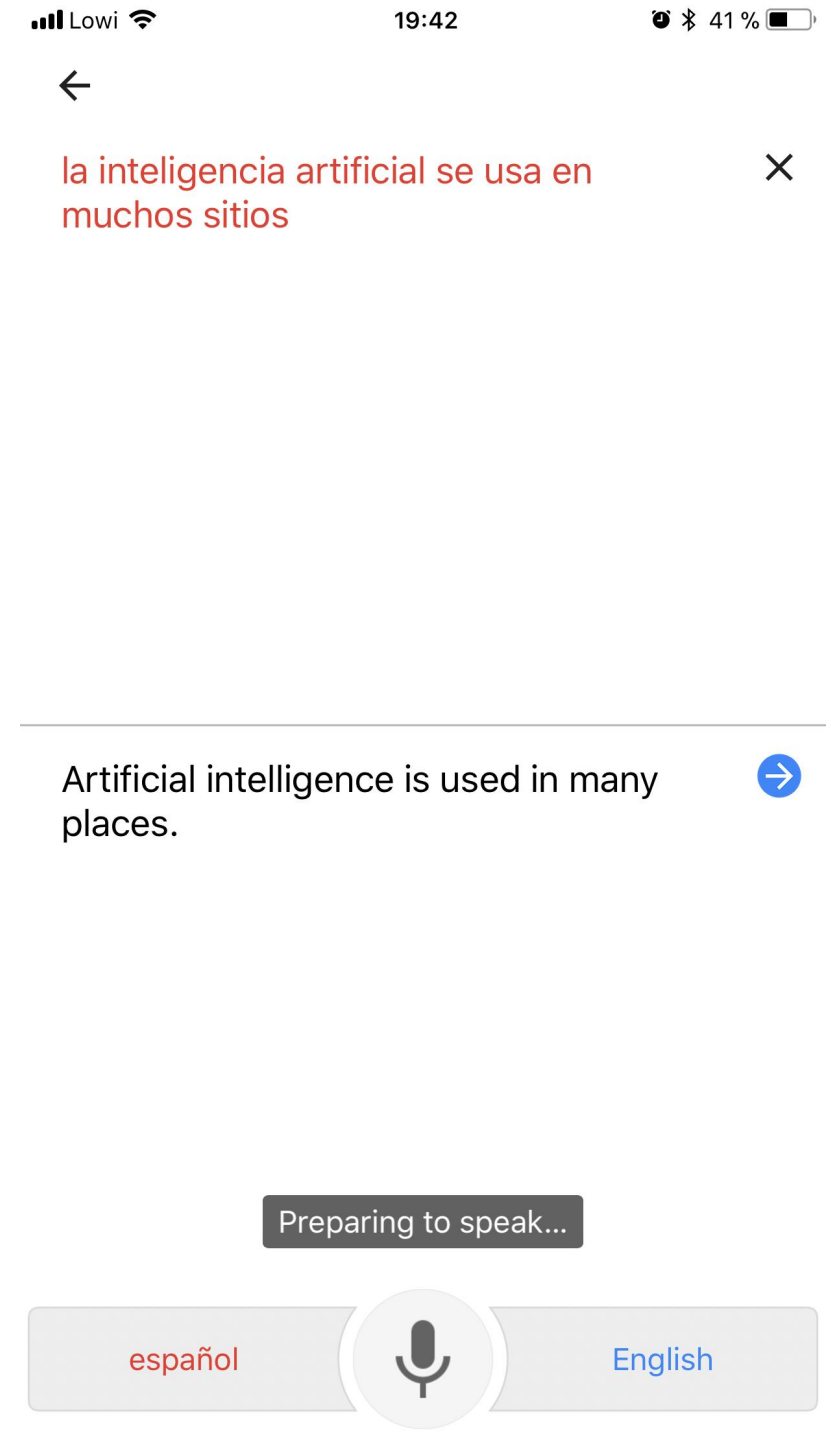
*Google Translate* now renders spoken sentences in one language into spoken sentences in another, for **32 pairs** of languages and offers text translation for **100+ languages**.

Natural Language Processing



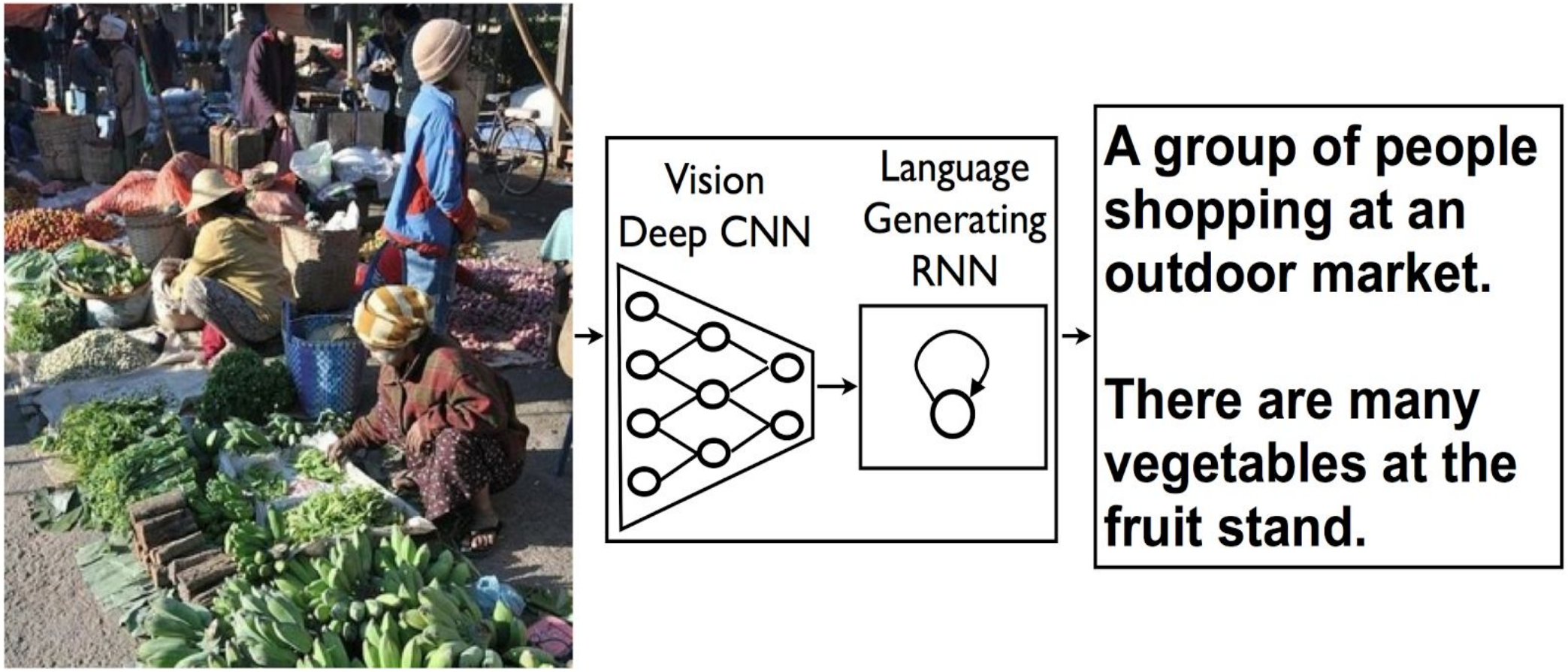
*Google Translate* now renders spoken sentences in one language into spoken sentences in another, for **32 pairs** of languages and offers text translation for **100+ languages**.

**Natural Language Processing**

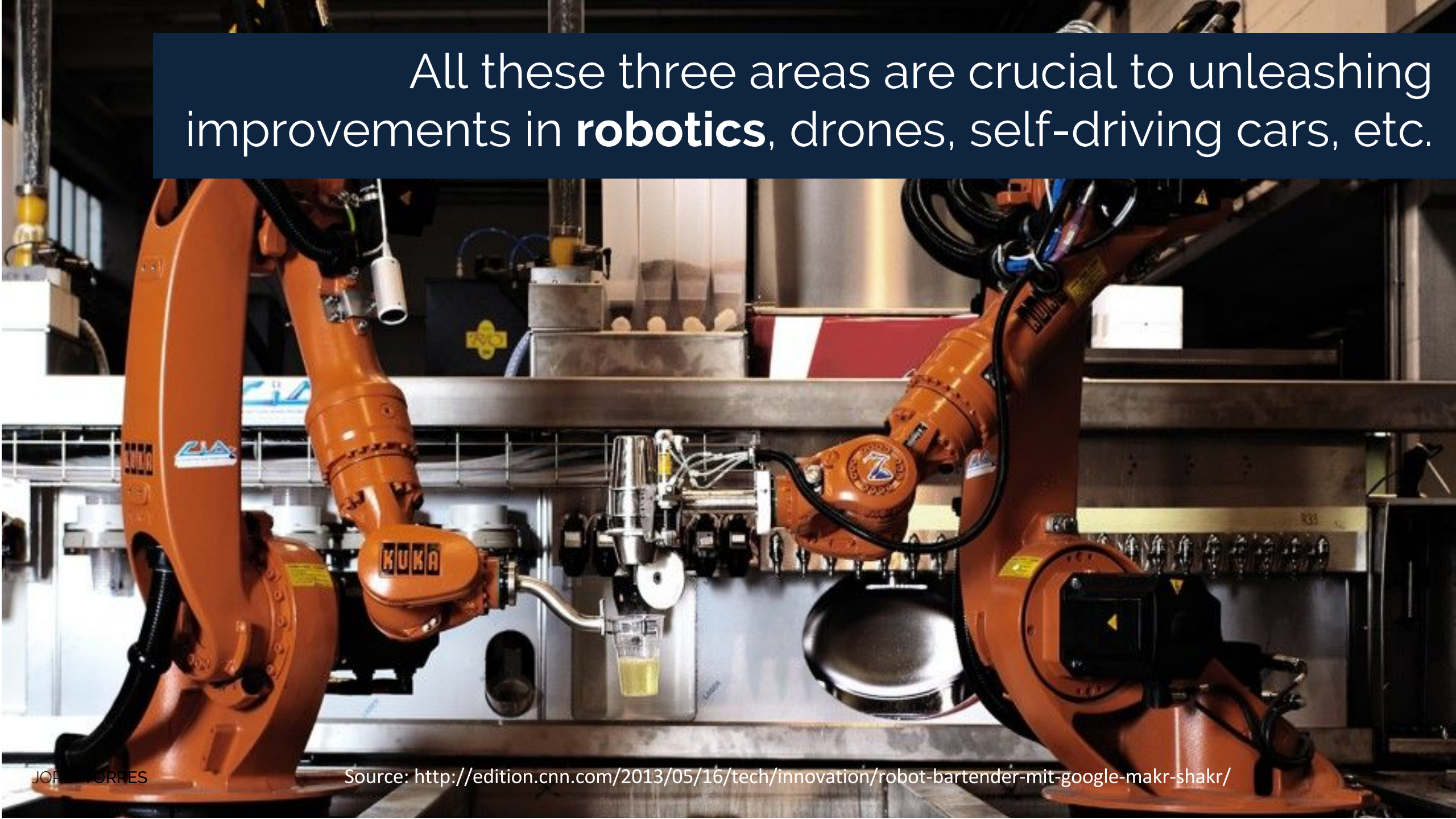


Computer Vision

# Now our computers can recognize images and generate descriptions for photos in seconds.



All these three areas are crucial to unleashing improvements in **robotics**, drones, self-driving cars, etc.





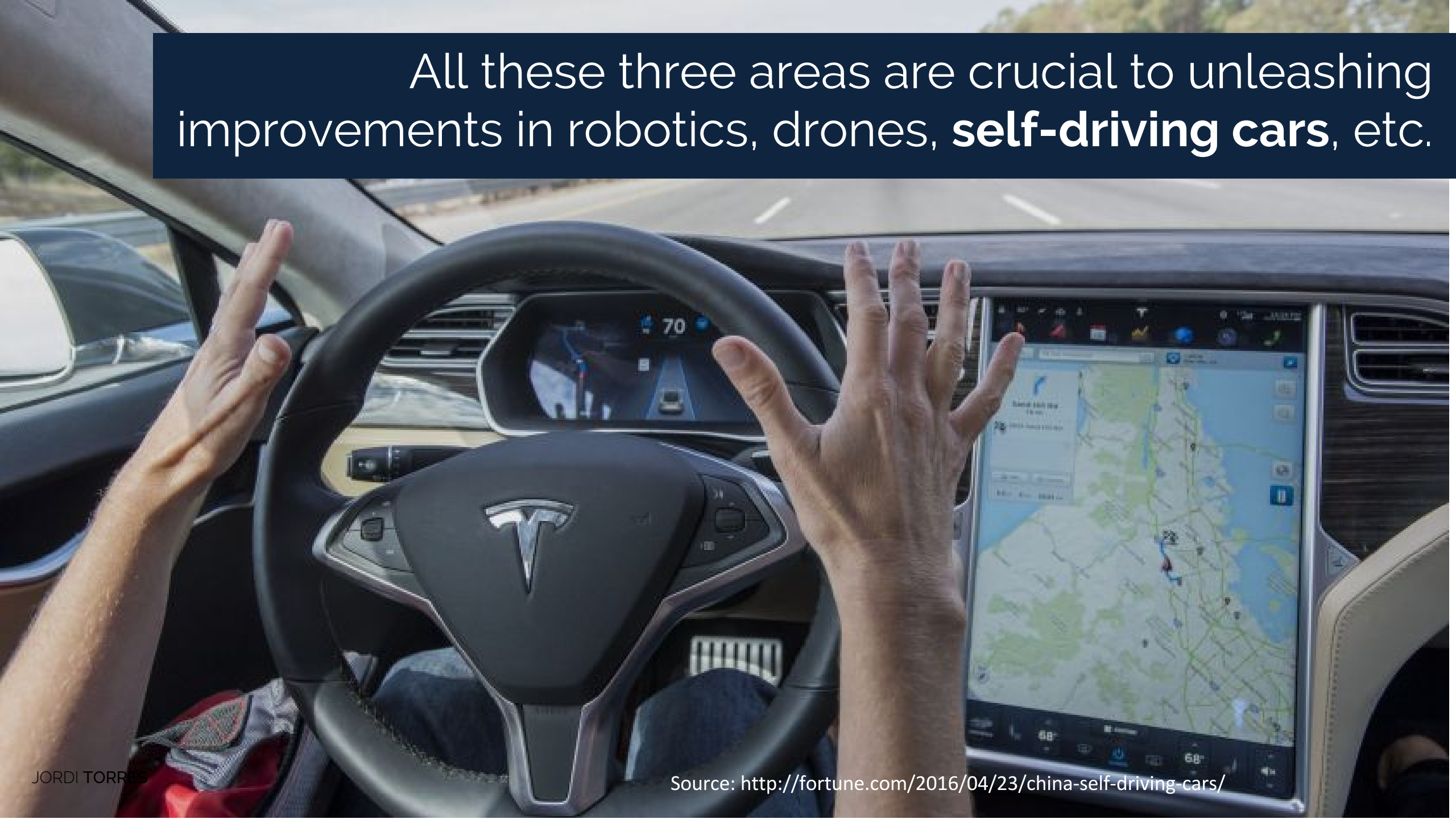

Source: http://edition.cnn.com/2013/05/16/tech/innovation/robot-bartender-mit-google-makr-shakr/

All these three areas are crucial to unleashing improvements in robotics, **drones**, self-driving cars, etc.




Source: http://axisphilly.org/article/military-drones-philadelphia-base-control/

All these three areas are crucial to unleashing improvements in robotics, drones, **self-driving cars**, etc.





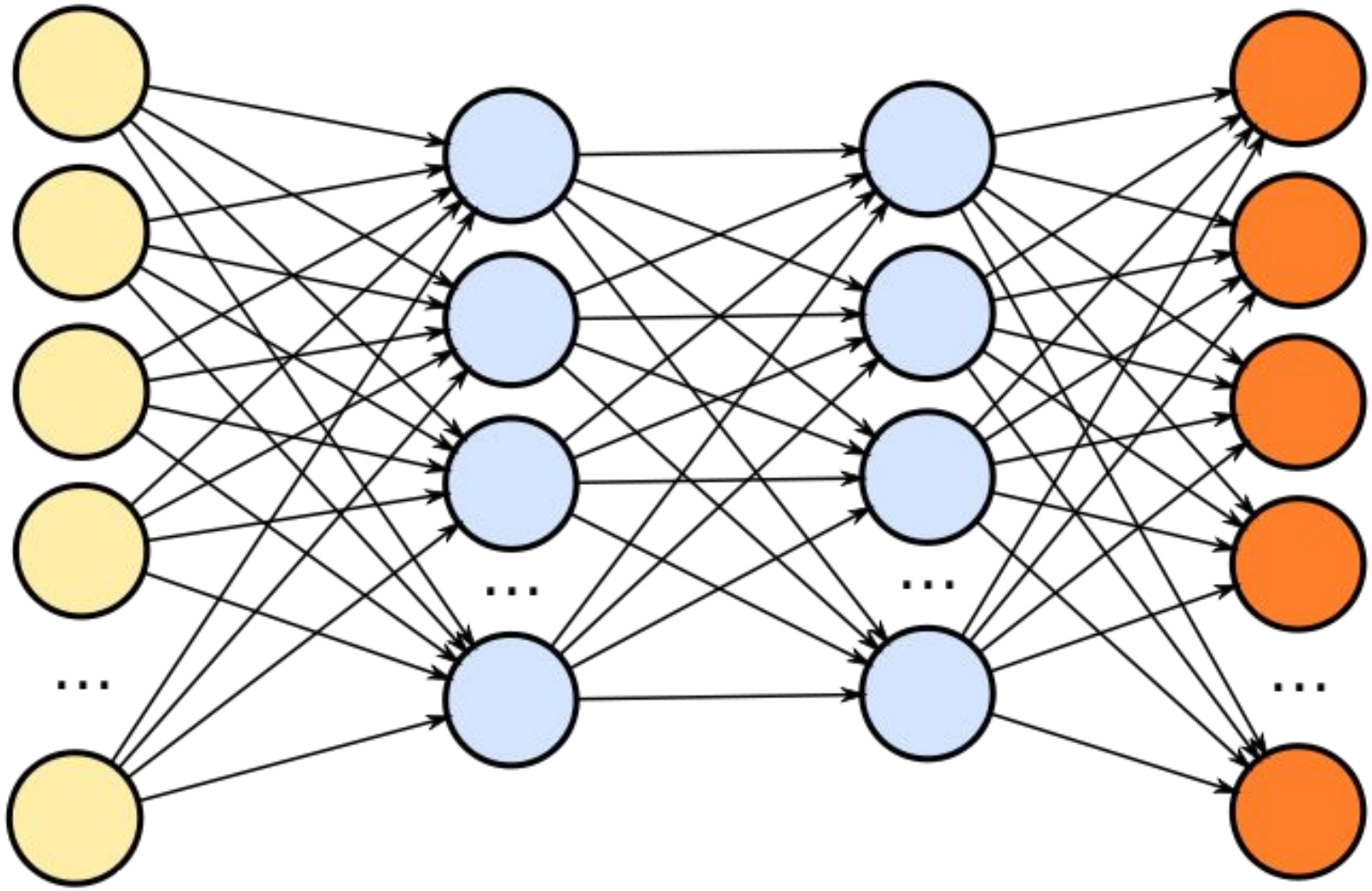
Source: http://fortune.com/2016/04/23/china-self-driving-cars/

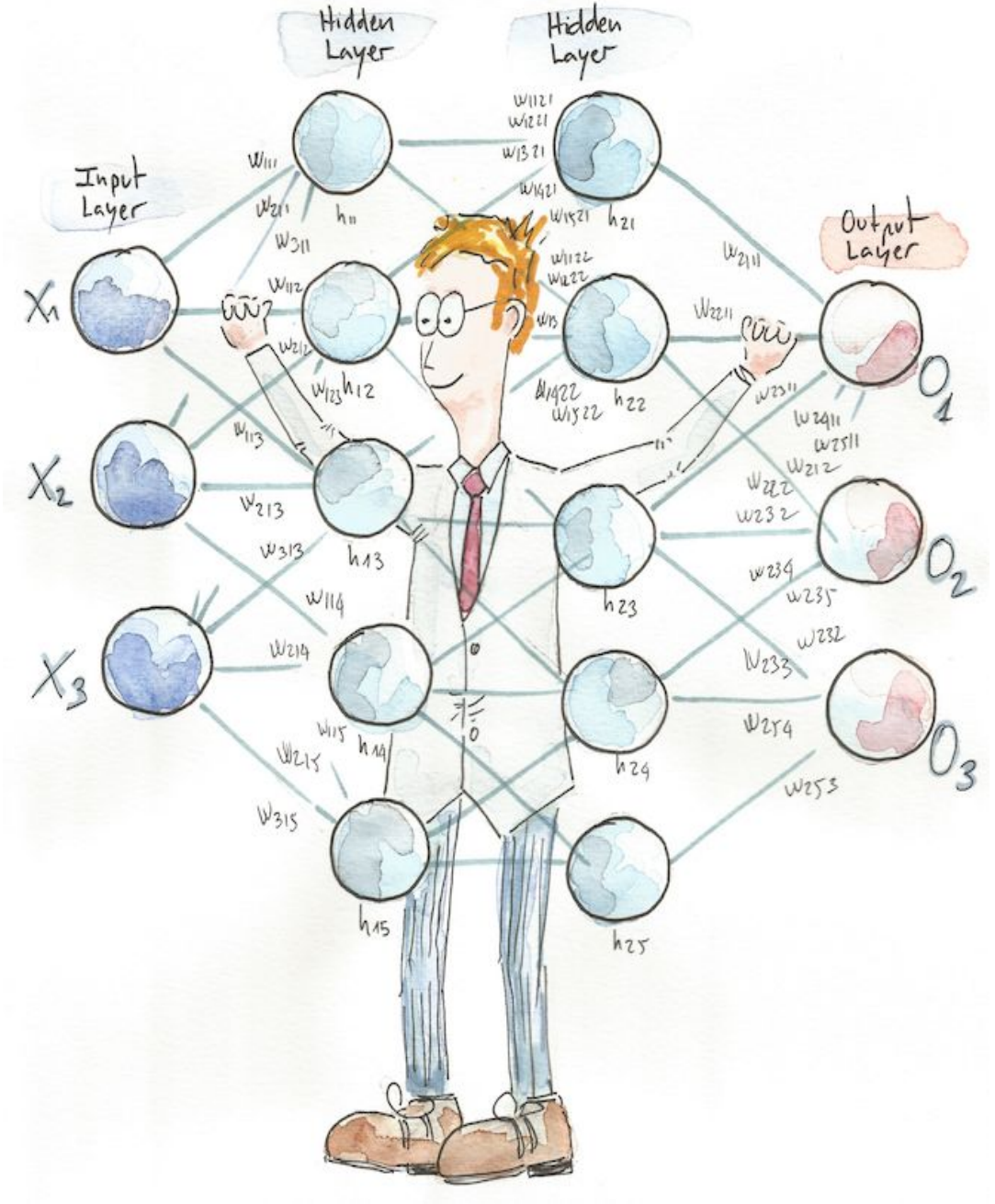
AI is at the heart of today's technological innovation.



image source: http://acuvate.com/blog/22-experts-predict-artificial-intelligence-impact-enterprise-workplace

# Many of these breakthroughs have been made possible by a family of AI known as Neural Networks



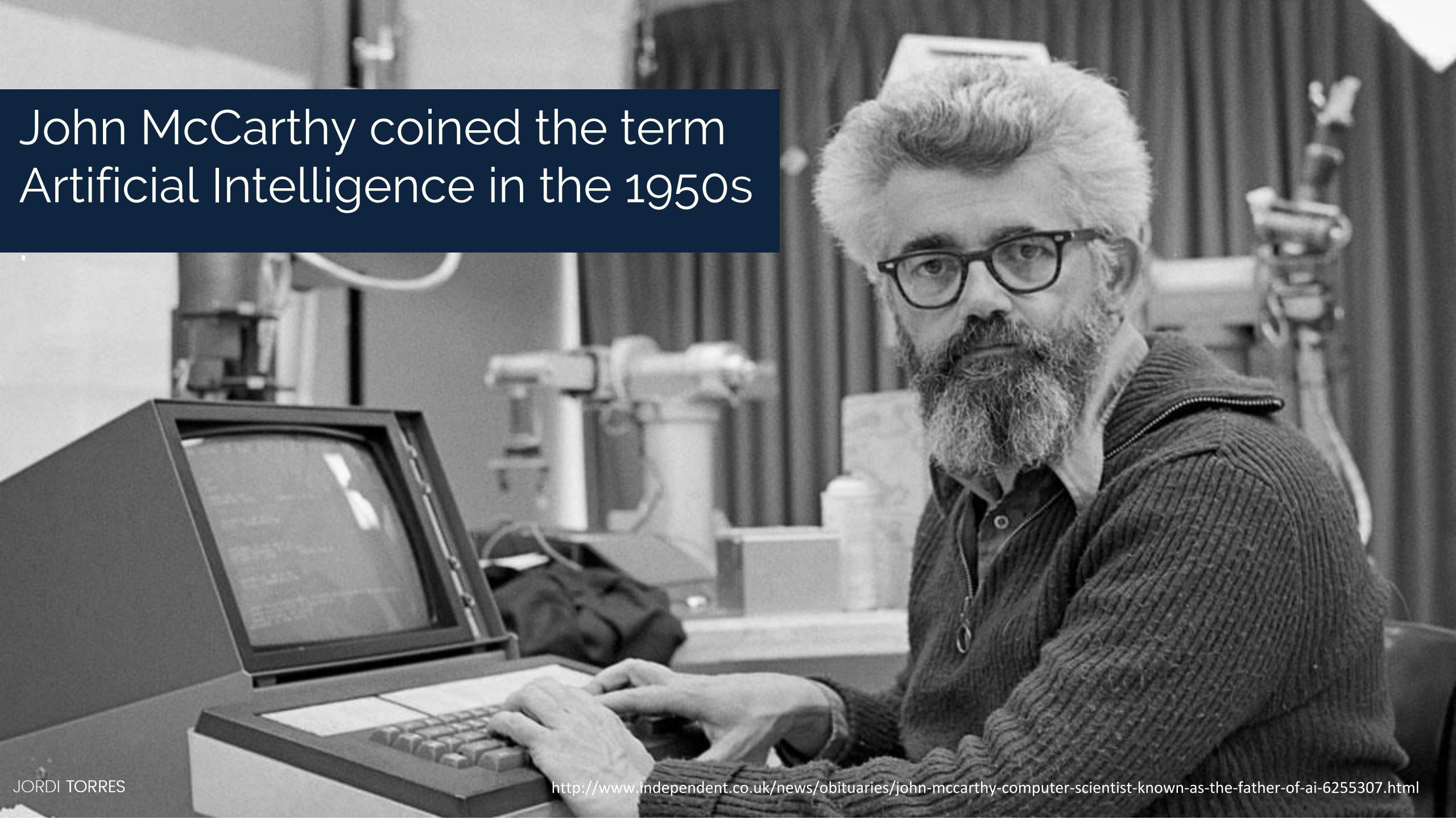



Neural networks, also known as a **Deep Learning,** enables a computer to learn from observational data

Although the greatest impacts of deep learning may be obtained whe it is integrated into the whole toolbo of other AI techniques


Universitat Politècnica de Barcelona
FACULTAT D'INFORMÀTICA
GUIA DOCENT
Curs 1982/83

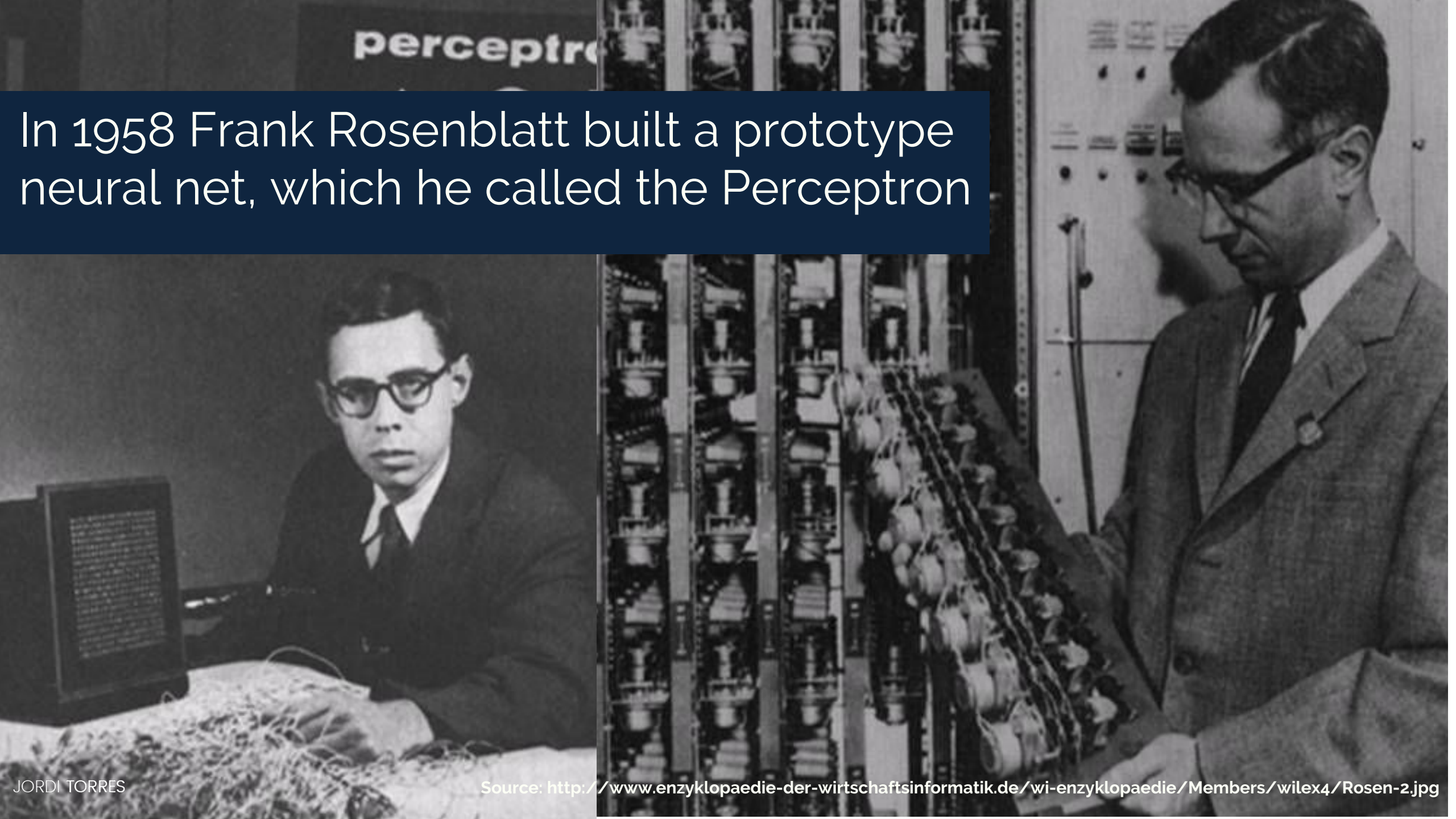
# John McCarthy coined the term Artificial Intelligence in the 1950s





http://www.independent.co.uk/news/obituaries/john-mccarthy-computer-scientist-known-as-the-father-of-ai-6255307.html

# In 1958 Frank Rosenblatt built a prototype neural net, which he called the Perceptron





Source: http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/Members/wilex4/Rosen-2.jpg

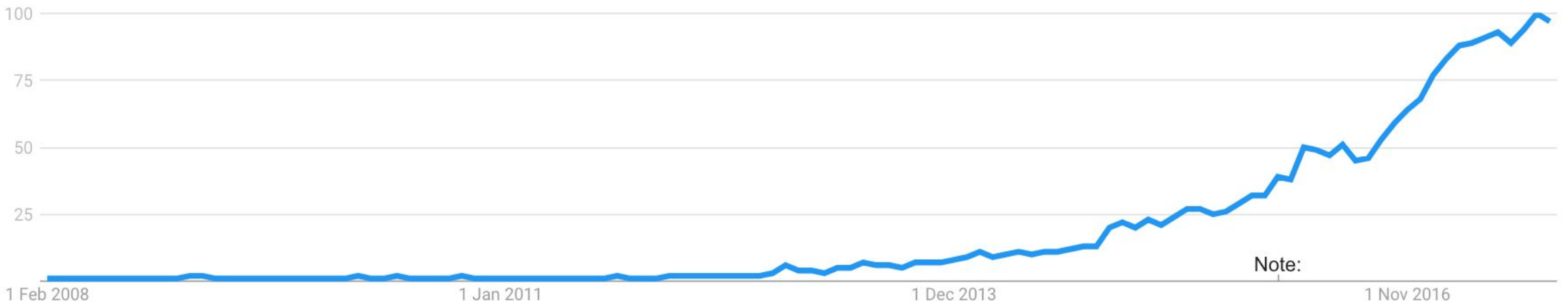
Google Trends | Explore

"deep learning"
Search term

So why did Deep Learning only take off few years ago?

Worldwide | 09/01/2008 - 01/10/2017 | All categories

Interest over time

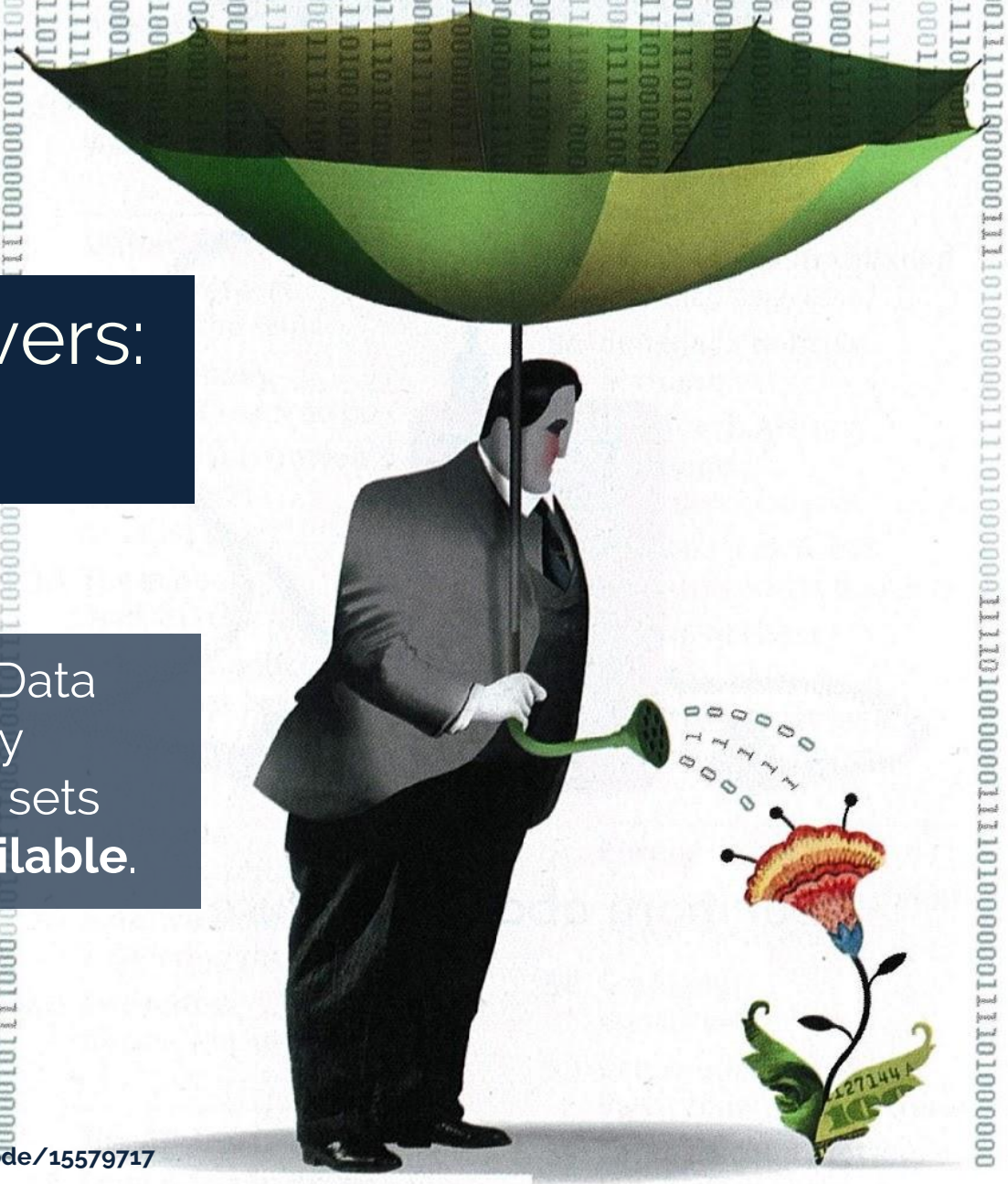


# One of the key drivers: The data deluge

Source:http://www.economist.com/node/15579717

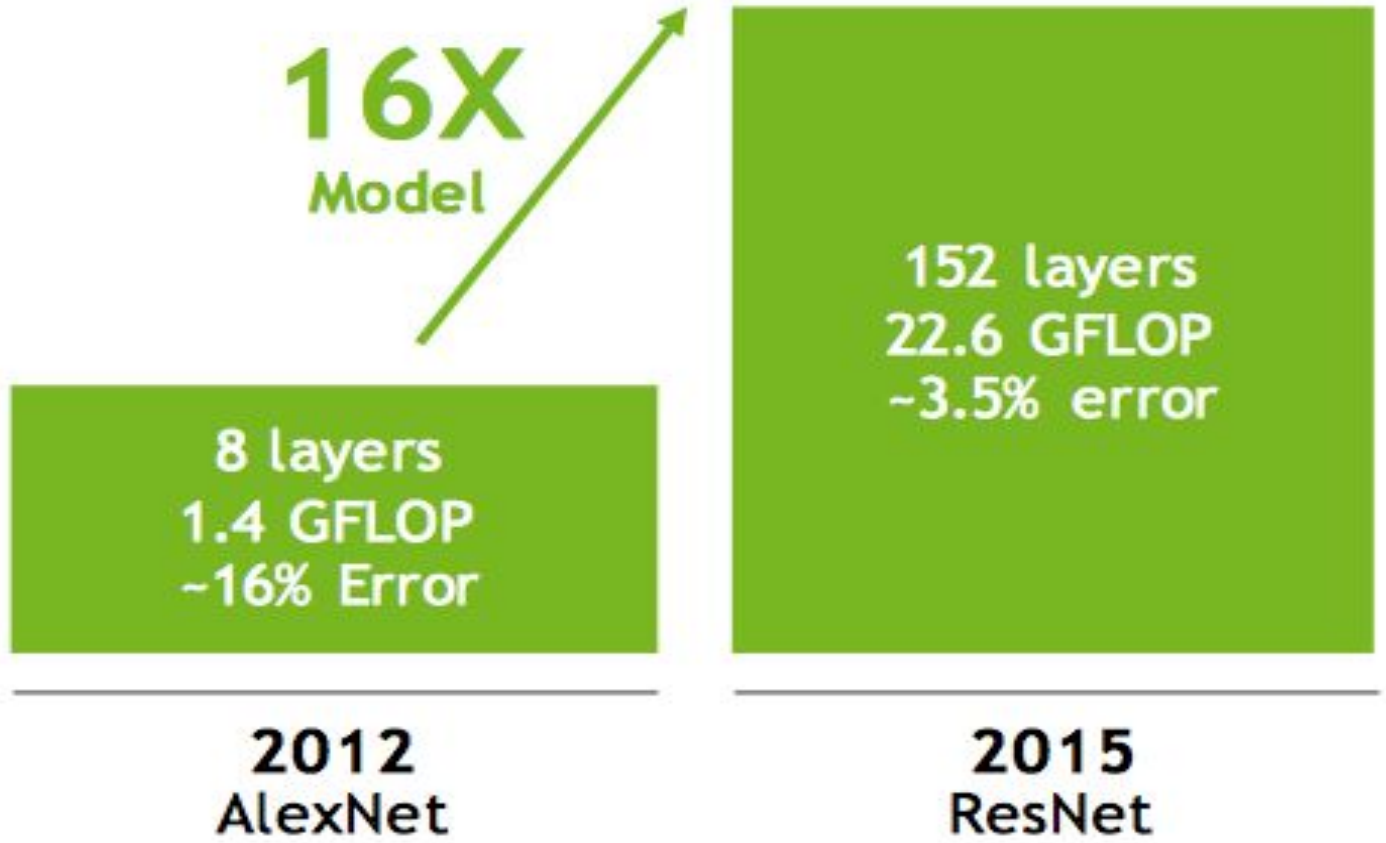
# One of the key drivers: The data deluge

Thanks to the advent of Big Data AI models can be “trained” by exposing them to large data sets **that were previously unavailable**.

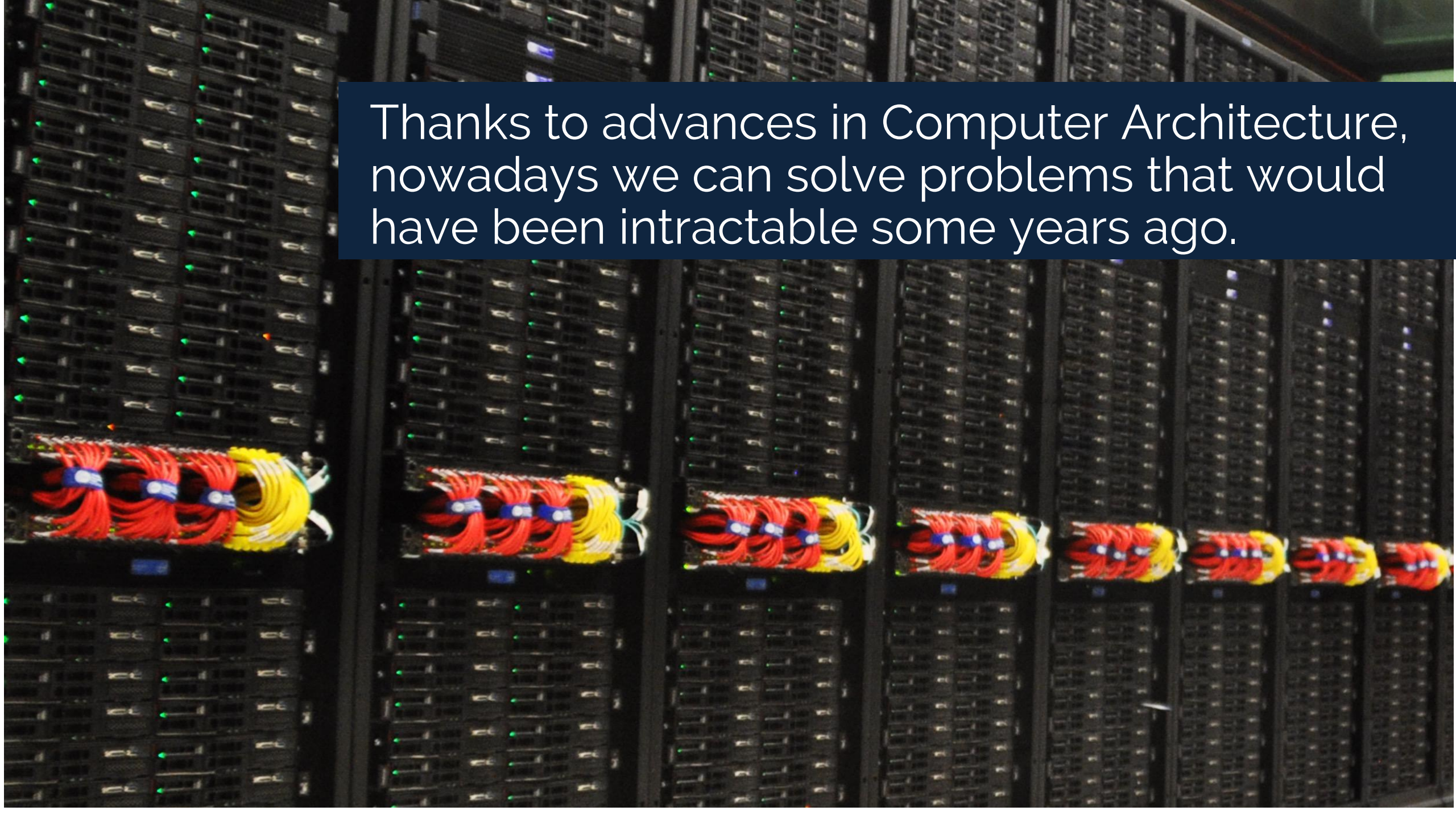


Source:http://www.economist.com/node/15579717

# Training DL neural nets has an insatiable demand for Computing




source: cs231n.stanford.edu/slides/2017/cs231n_2017_lecture15.pdf


Thanks to advances in Computer Architecture, nowadays we can solve problems that would have been intractable some years ago.

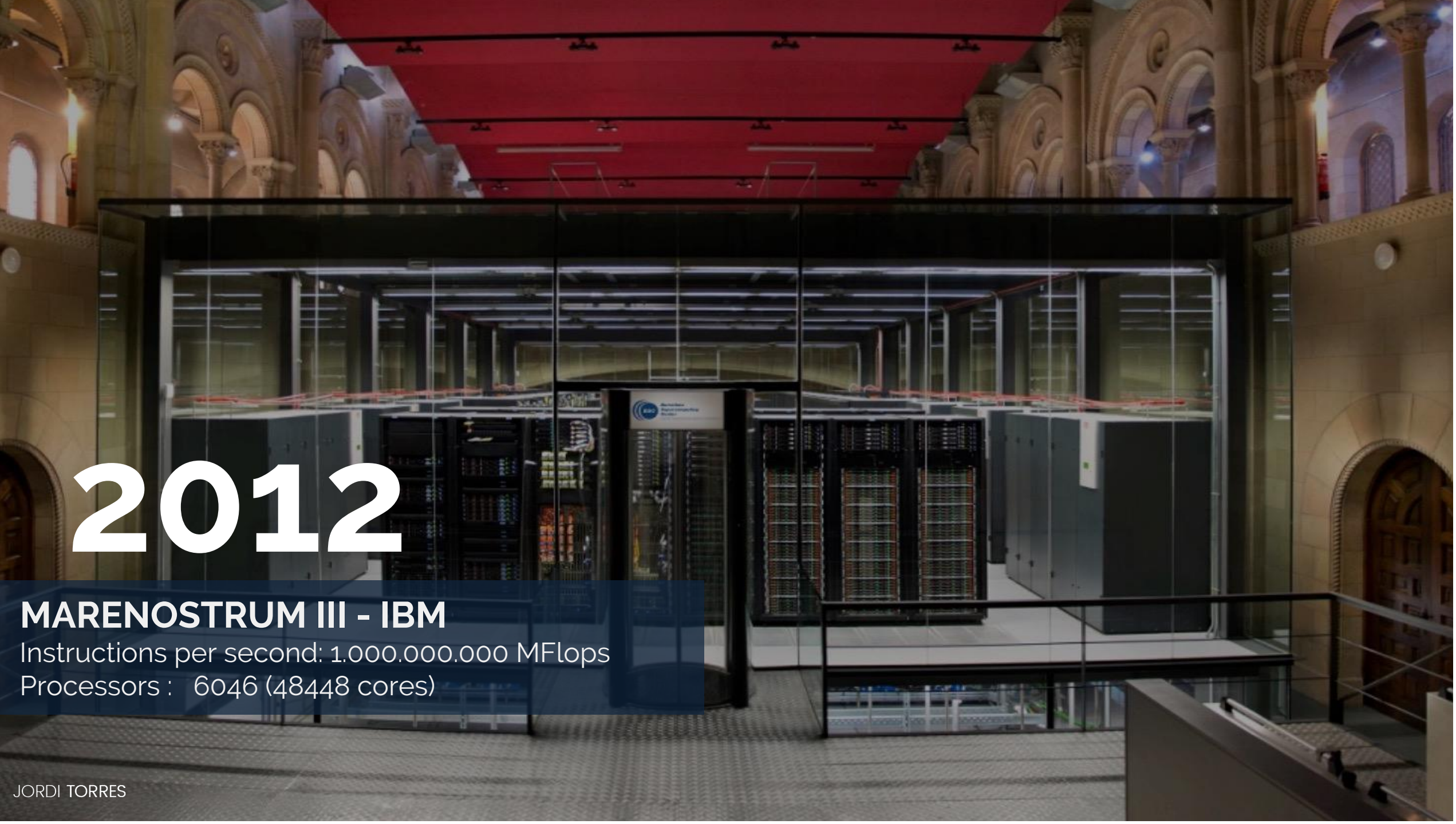
# 1982

## FACOM 230 – Fujitsu

Instructions per second: few Mips * (M = 1.000.000)

Processors : 1

Credits: http://www.ithistory.org/sites/default/files/hardware/facom230-50.jpg

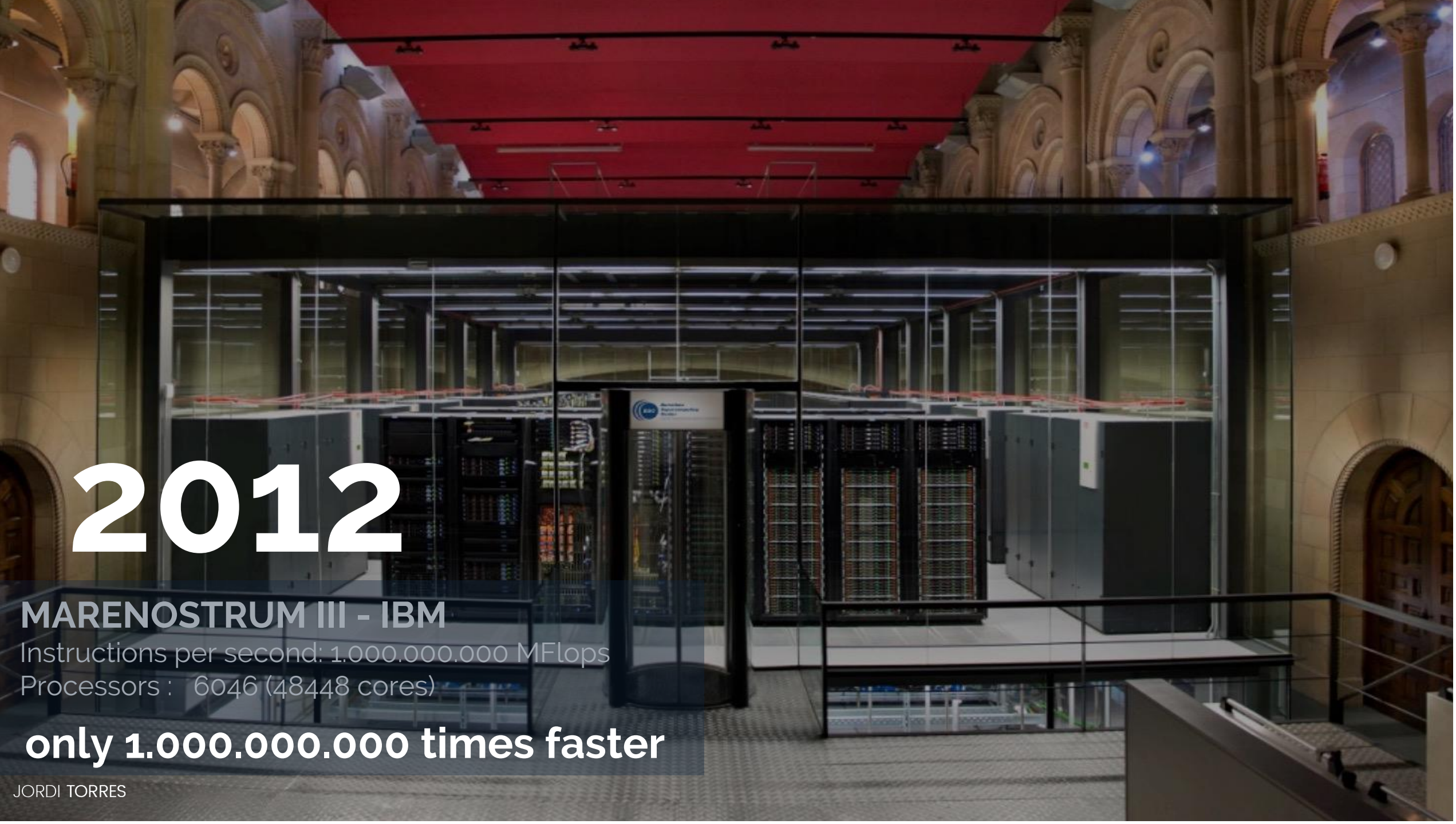


# 2012

## MARENOSTRUM III - IBM

Instructions per second: 1.000.000.000 MFlops

Processors : 6046 (48448 cores)



# 2012

## MARENOSTRUM III - IBM

Instructions per second: 1.000.000.000 MFlops

Processors : 6046 (48448 cores)

**only 1.000.000.000 times faster**

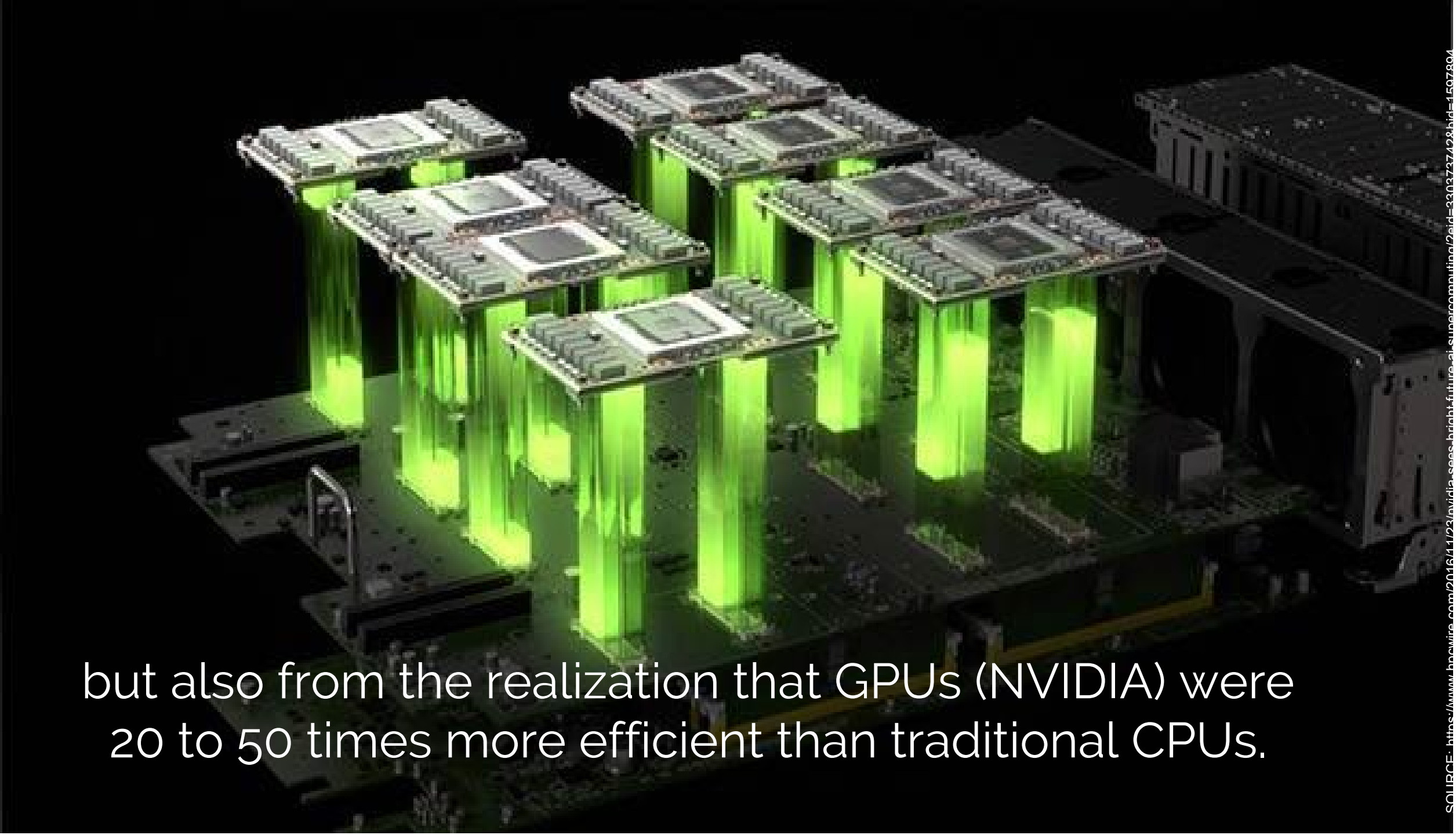
# CPU improvements!

Until then, the increase in computational power every decade of “my” computer, was mainly thanks to CPU

# CPU improvements!

Until then, the increase in computational power every decade of “my” computer, was mainly thanks to CPU

Since then, the increase in computational power for Deep Learning has not only been from CPU improvements . . .

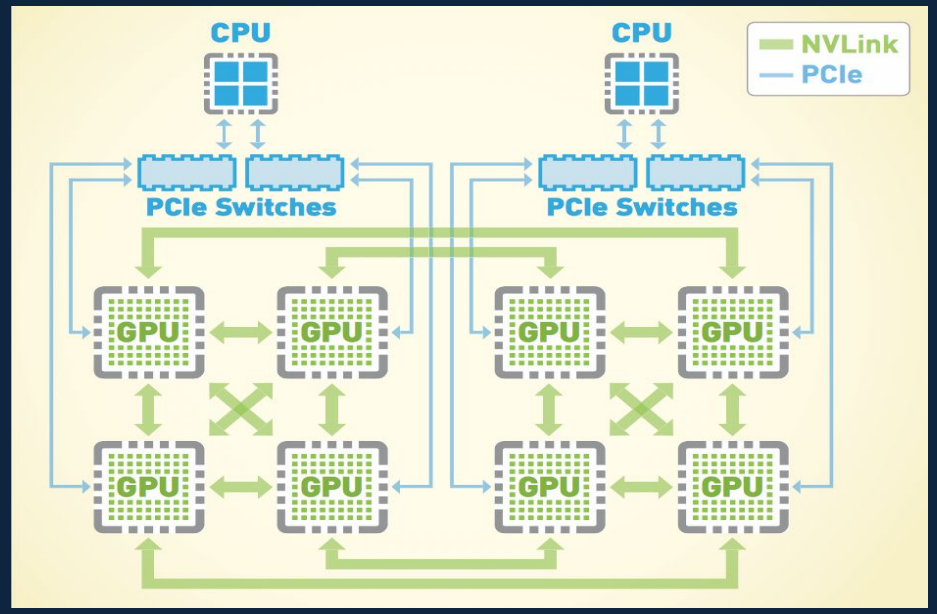


but also from the realization that GPUs (NVIDIA) were 20 to 50 times more efficient than traditional CPUs.

SOURCE: https://www.hpcwire.com/2016/11/23/nvidia-sees-bright-future-ai-supercomputing/?eid=3303737428&bid=1507804

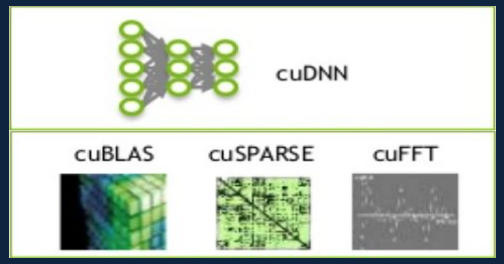
# Deep Learning requires computer architecture advancements

Fast tightly coupled network interfaces




Dense computer hardware

AI specific processors


Optimized libraries and kernels

COMPUTING POWER
is the real enabler!

# What if I do not have this hardware?

Now we are entering into an era of computation democratization for companies !

And what is “my/your” computer like now?

# And what is “my/your” computer like now?


Source: http://www.google.com/about/datacenters/gallery/images

And what is “my/your” computer like now?


Source: http://www.google.com/about/datacenters/gallery/images

# Huge data centers!

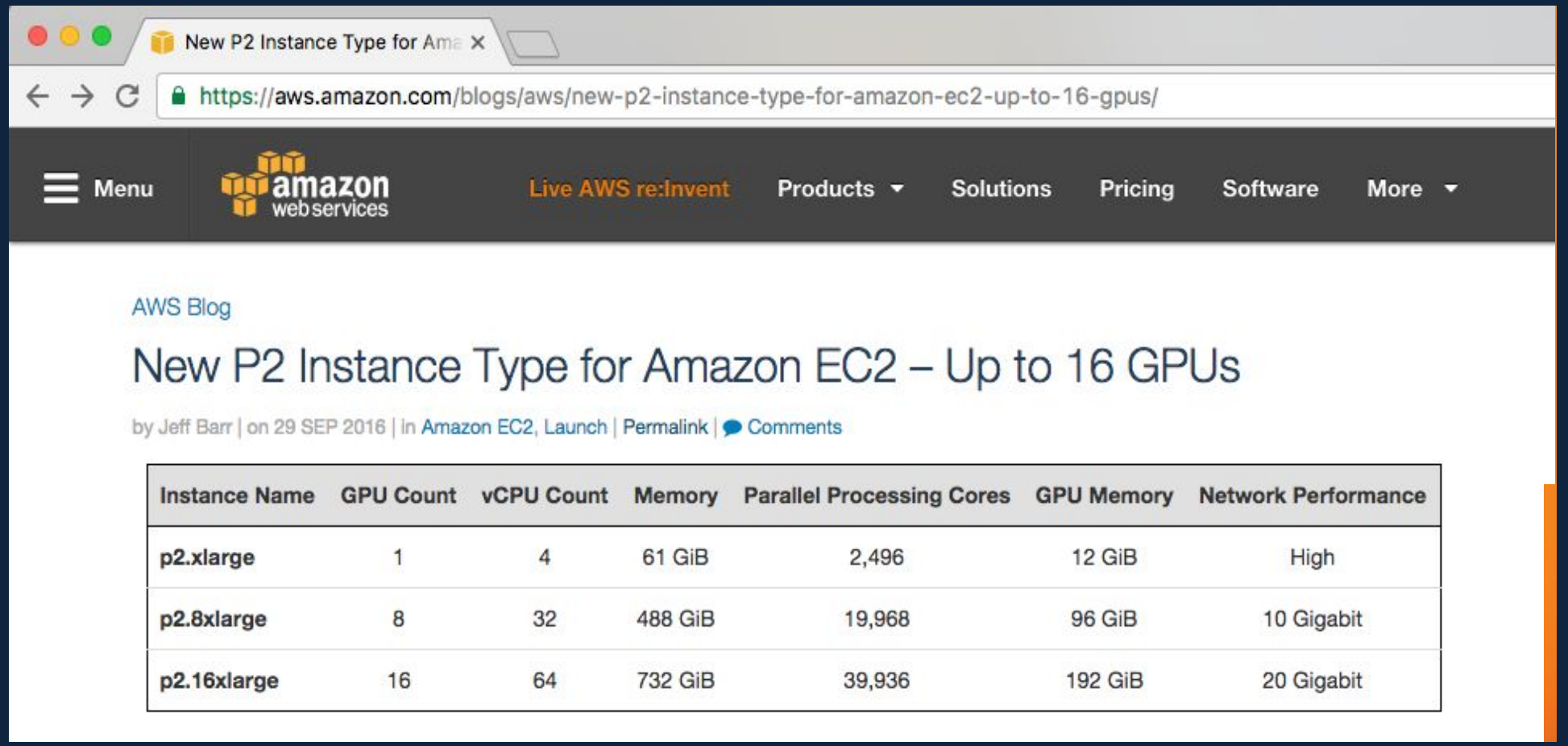

Credits: http://datacenterfrontier.com/server-farms-writ-large-super-sizing-the-cloud-campus/

28.000 m2
Foto: Google


28.000 m2
Foto: Google


28.000 m2
Foto: Google

For those (experts) who want to develop their own software, cloud services like Amazon Web Services provide GPU-driven deep-learning computation services




| Instance Name | GPU Count | vCPU Count | Memory | Parallel Processing Cores | GPU Memory | Network Performance |
|---|---|---|---|---|---|---|
| **p2.xlarge** | 1 | 4 | 61 GiB | 2,496 | 12 GiB | High |
| **p2.8xlarge** | 8 | 32 | 488 GiB | 19,968 | 96 GiB | 10 Gigabit |
| **p2.16xlarge** | 16 | 64 | 732 GiB | 39,936 | 192 GiB | 20 Gigabit |

# And Google ...

Google Cloud Platform

Search

Why Google | Products | Solutions | Launcher | Pricing | Customers | Documentation | Support >

## GRAPHICS PROCESSING UNIT (GPU) BETA

Leverage GPUs on Google Cloud for machine learning and scientific computing

VIEW MY CONSOLE

VIEW DOCUMENTATION

## Accelerated Cloud Computing

Scientists, artists, and engineers need access to massively parallel computational power. Google Cloud now offers virtual machines with GPUs capable of tens of teraflops of performance. Deep learning, physical simulation, and molecular modeling take hours instead of days on NVIDIA® Tesla® K80 GPUs. Regardless of the size of your workload, GCP provides the perfect GPU for your job. AMD FirePro and NVIDIA® Tesla® P100s are coming soon.

# And Google ...

Google Cloud Platform

Search CONSOLE

Why Google | Products | Solutions | Launcher | Pricing | Customers | Documentation | Support >

## CLOUD TPU ALPHA

Train and run machine learning models faster than ever before

SIGN UP TO LEARN MORE

## Accelerated Machine Learning

Machine learning (ML) has the power to greatly simplify our lives. Improvements in speech recognition and language understanding help all of us interact more naturally with technology. Businesses rely on ML to strengthen network security and reduce fraud. Advances in medical imaging enabled by ML can increase the accuracy of medical diagnoses and expand access to care, ultimately saving lives.

And all major cloud platforms...

Microsoft Azure
IBM Cloud
Aliyun
Cirrascale
NIMBIX
Outscale
. . .

Cogeco Peer 1
Penguin Computing
RapidSwitch
Rescale
SkyScale
SoftLayer
. . .

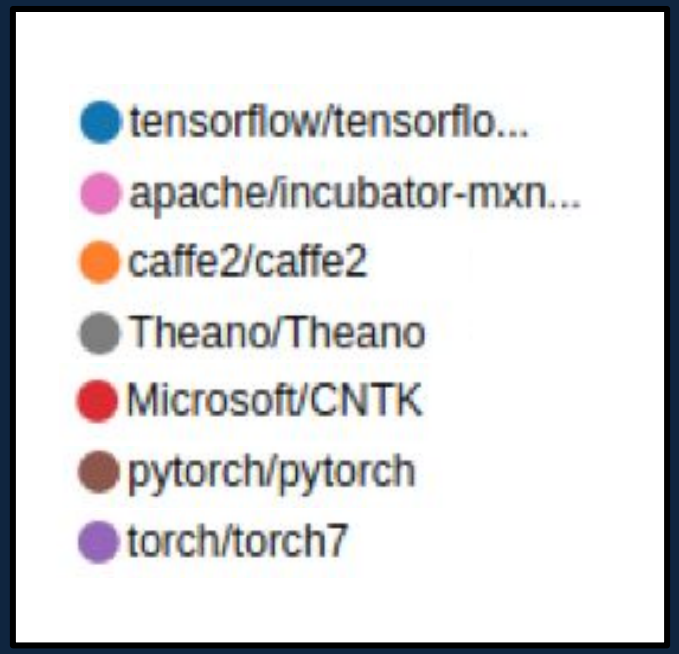
And for “less expert” people, various companies are providing a working scalable implementation of ML/AI algorithms as a Service (**AI-as-a-Service**)

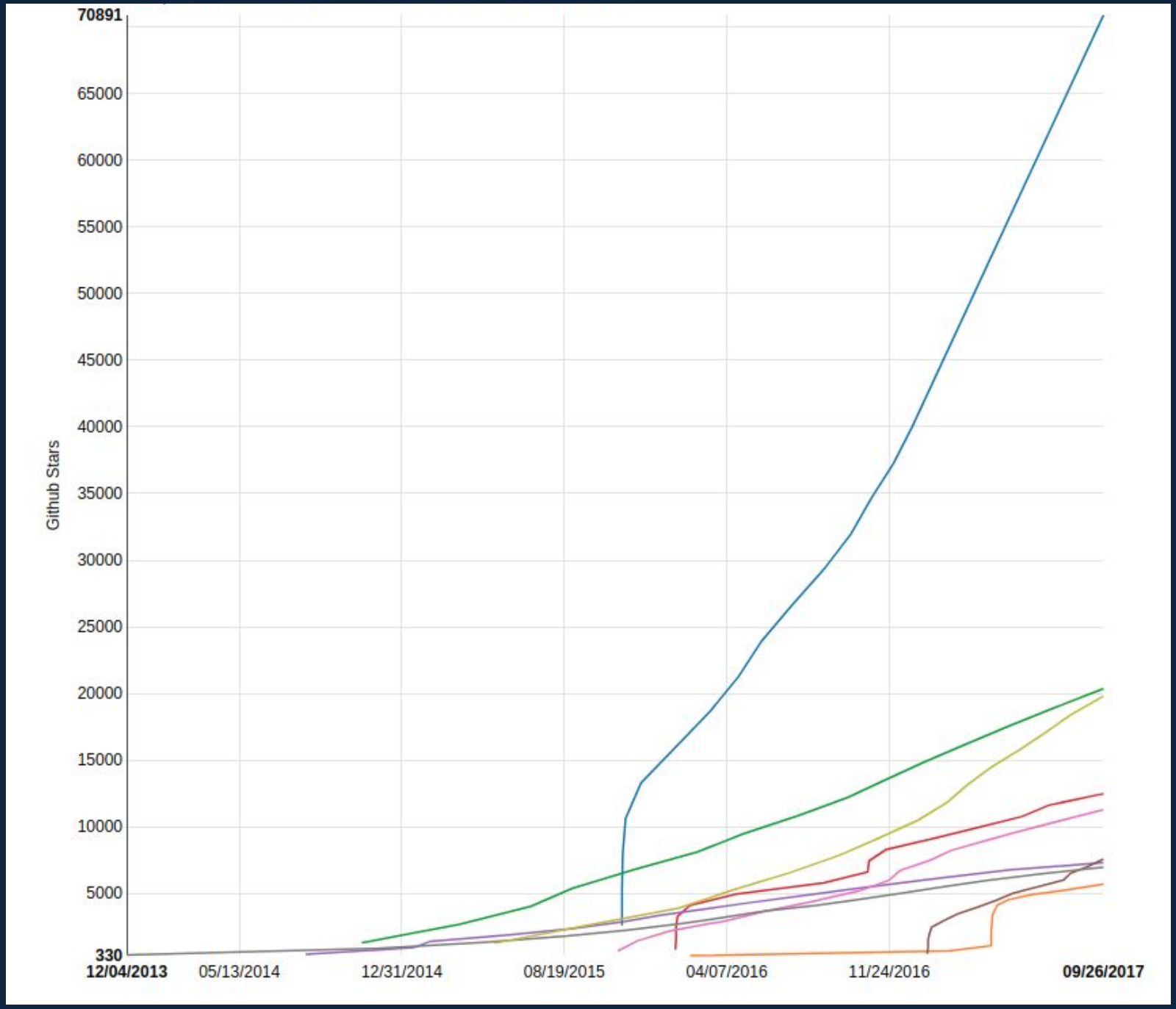

Source: http://www.kdnuggets.com/2015/11/machine-learning-apis-data-science.html

# An open-source world for the Deep Learning community

Many **open-source DL software**
have greased the innovation process

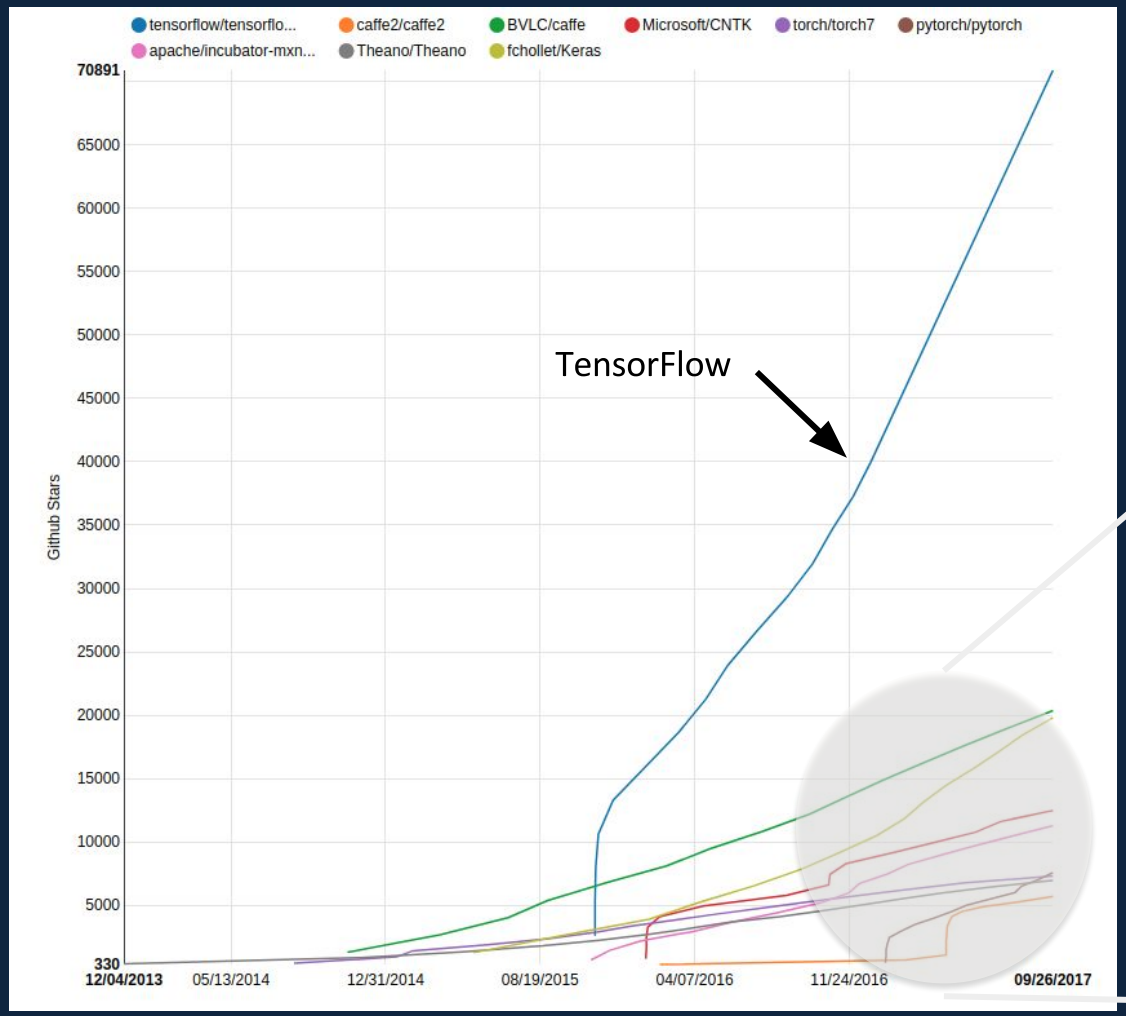
# Github Stars

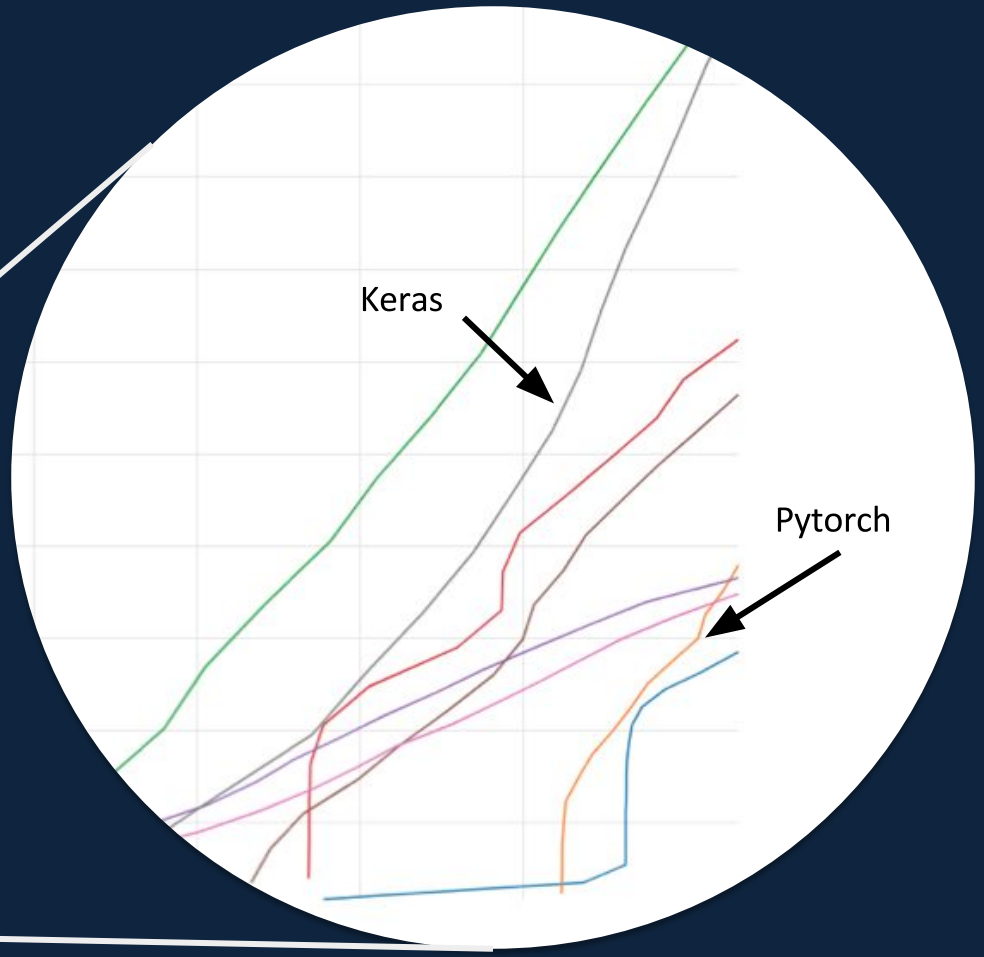



source: Francesc Sastre

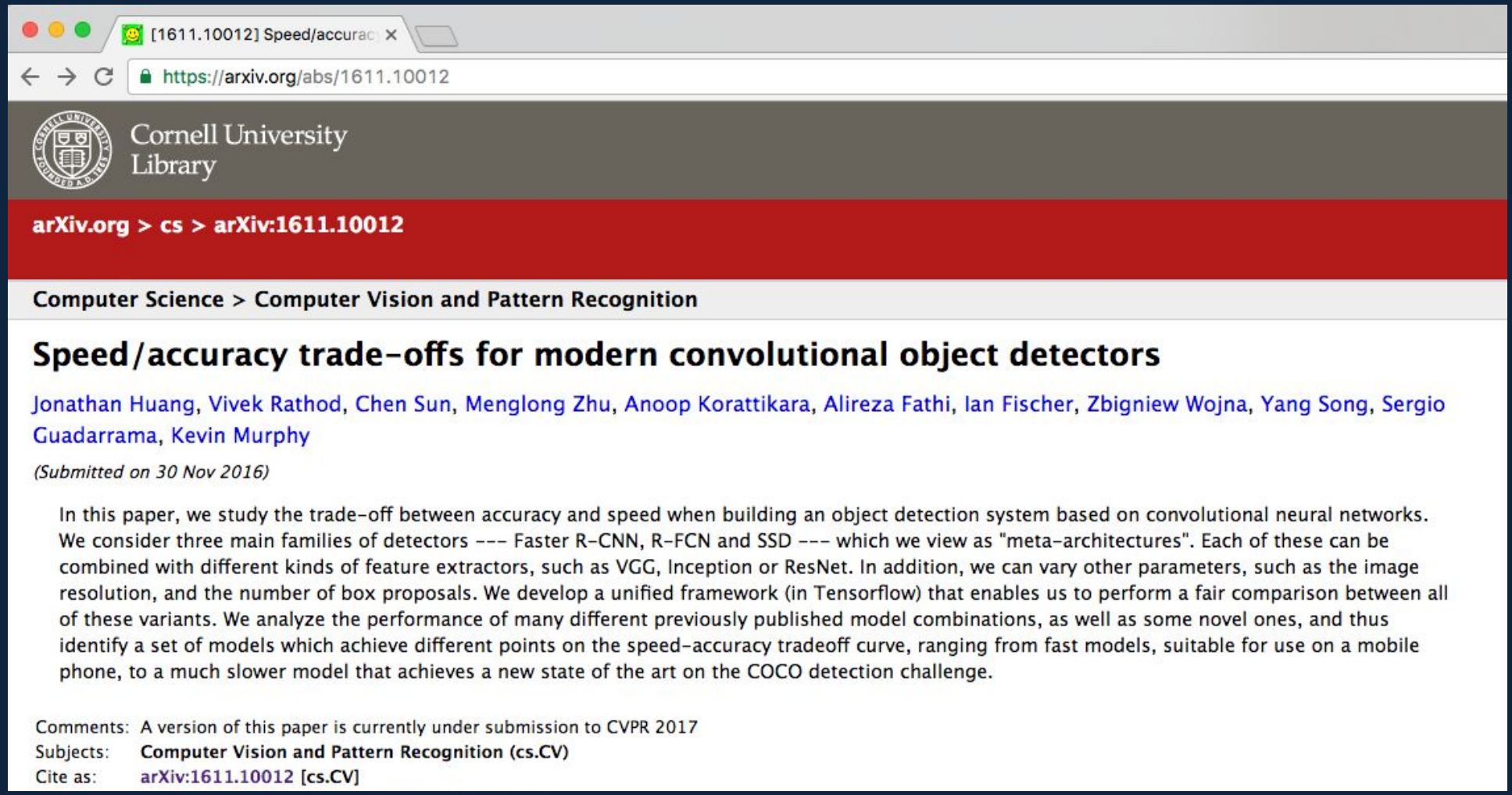
# In this course: **we will consider the 3 frameworks with steepest gradient**

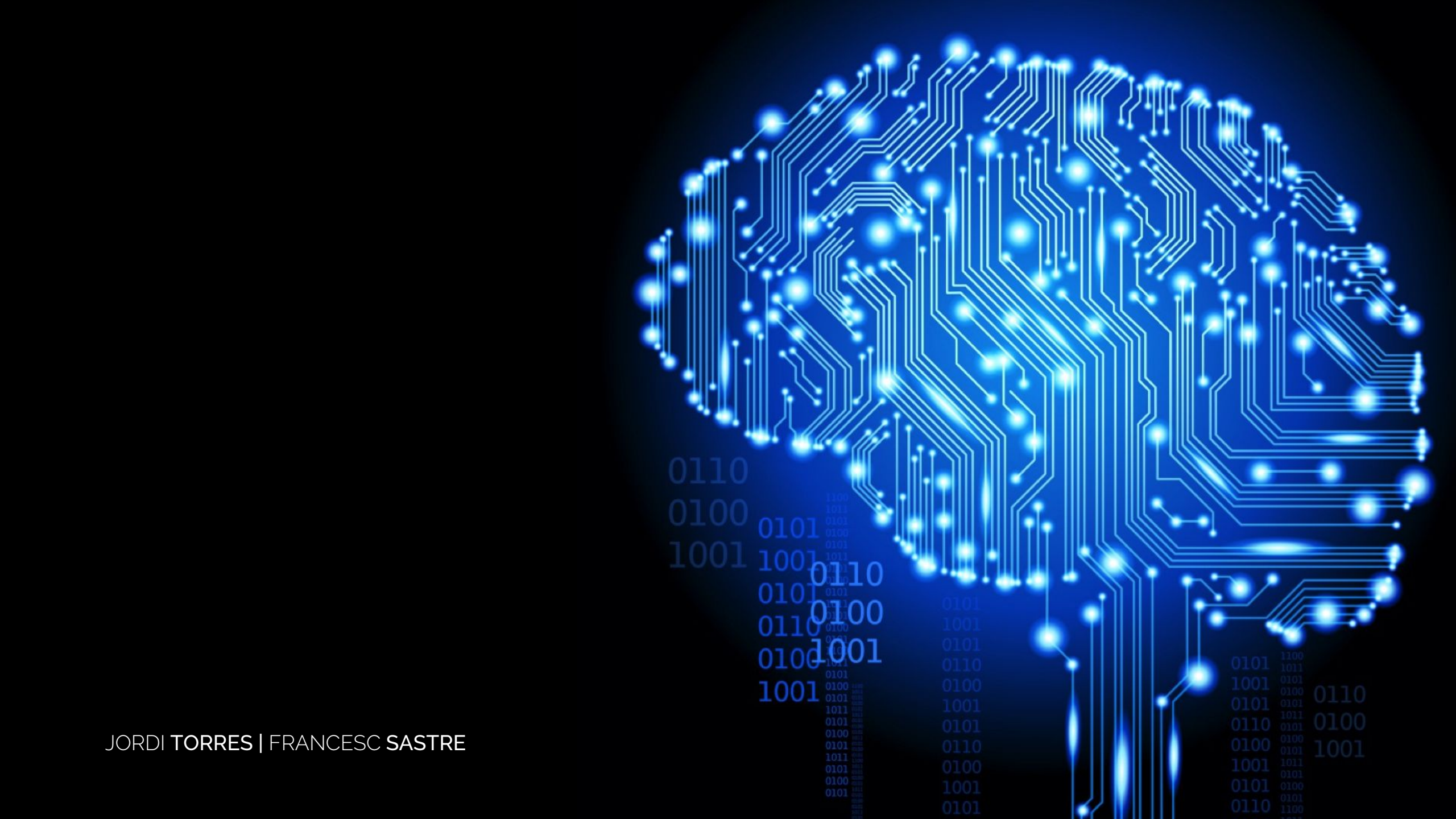



frameworks with more slope

and no less important, **an open-publication ethic**, whereby many researchers publish their results immediately on a database without awaiting peer-review approval.





JORDI **TORRES** | FRANCESC **SASTRE**