Introducing Machine Learning @Google with TensorFlow

Machine Learning for Marketeers
15th May 2018
Rokesh Jankie
rokesh@google.com
Customer Engineer

Google Cloud

Shift in Computing


2 Billion
#io17
#io17

Google's Mission

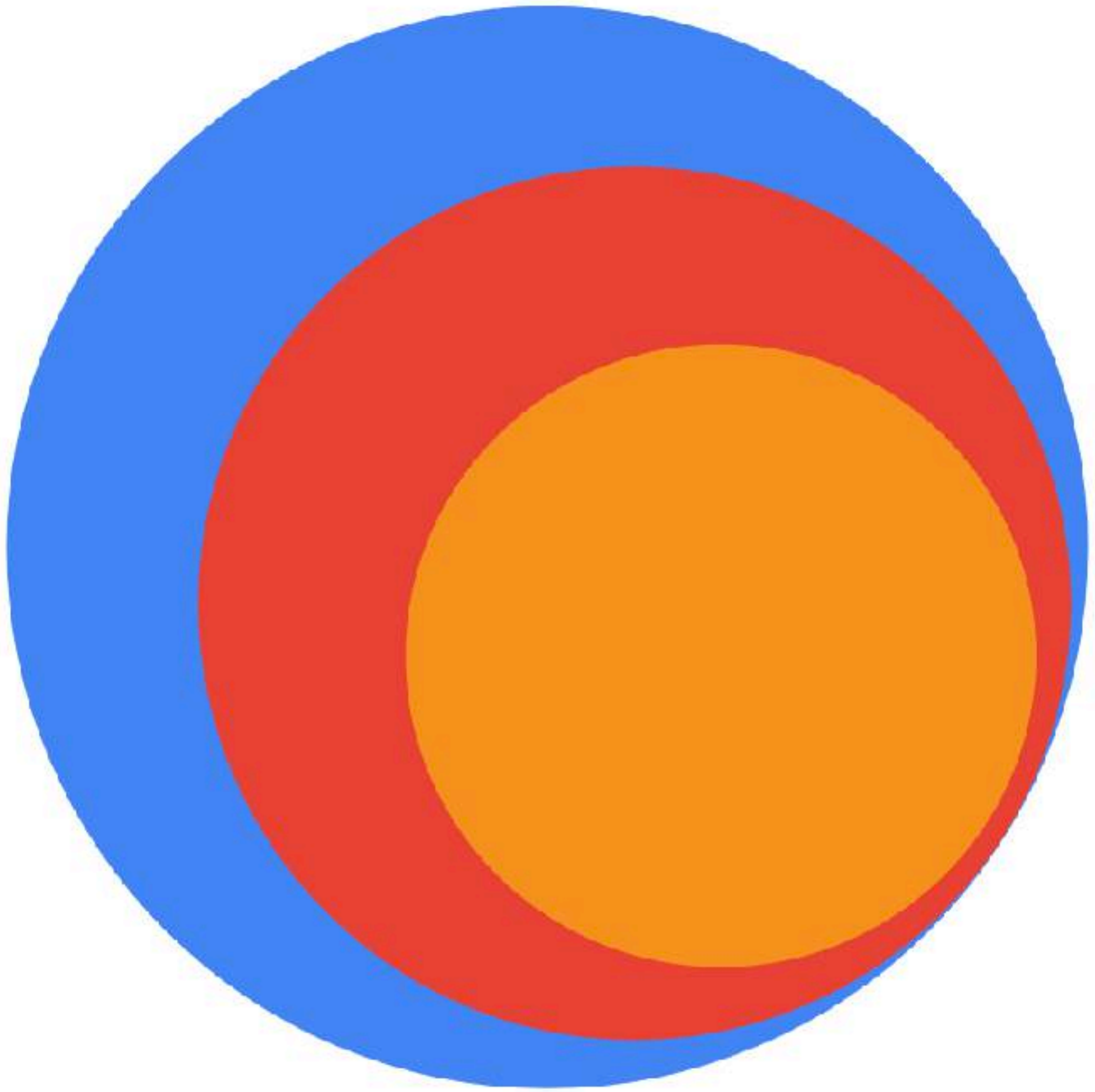
Organize the world's information
and make it universally
accessible and useful

Seven cloud products with

> ONE BILLION

Users

Putting things in perspective



Artificial Intelligence
The science to make things smart

Machine Learning
Building machines that can learn

Neural Network
A type of algorithms in machine learning

Supervised Learning

Labelled data = data + output (target, class, etc) variable. Goals is to learn the relationship between data and output

Semi-Supervised Learning

Partially labelled data

Online/Active Learning

Real-Time incremental learning on data streams

Unsupervised Learning

Unlabelled data.
Goals is to learn associations, groups, etc

Reinforcement Learning

Reinforcement Learning allows the machine or software agent to learn its behaviour based on feedback from the environment.

How Can You Get Started with Machine Learning?

- Three w
  - Use a C
    Natura
  - Use an
    it or fin
  - Develo
    new pr

More flexible, but more effort required

TensorFlow

What is TensorFlow

- Open source Machine Learning library
- Especially useful for Deep Learning
- For research and production
- Apache 2.0 license

TensorFlow History

- DistBelief

Large Scale Distributed Deep Networks

Venue

NIPS (2012)

Publication Year

2012

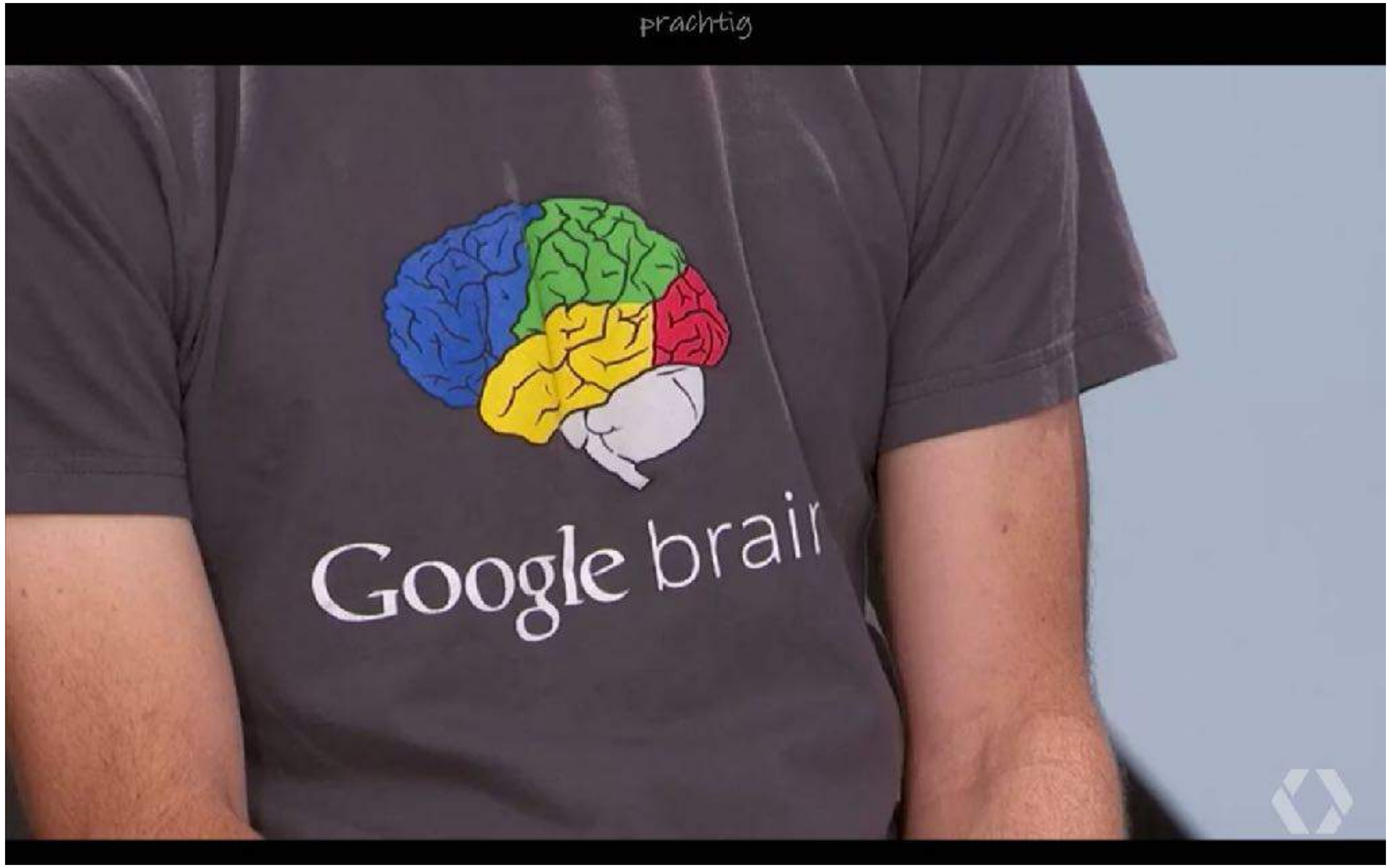
Authors

Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc'Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, Andrew Y. Ng

Abstract

Recent work in unsupervised feature learning and deep learning has shown that being able to train large models can dramatically improve performance. In this paper, we consider the problem of training a deep network with billions of parameters using tens of thousands of CPU cores. We have developed a software framework called DistBelief that can utilize computing clusters with thousands of machines to train large models. Within this framework, we have developed two algorithms for large-scale distributed training: (i) Downpour SGD, an asynchronous stochastic gradient descent procedure supporting a large number of model replicas, and (ii) Sandblaster, a framework that supports a variety of distributed batch optimization procedures, including a distributed implementation of L-BFGS. Downpour SGD and Sandblaster L-BFGS both increase the

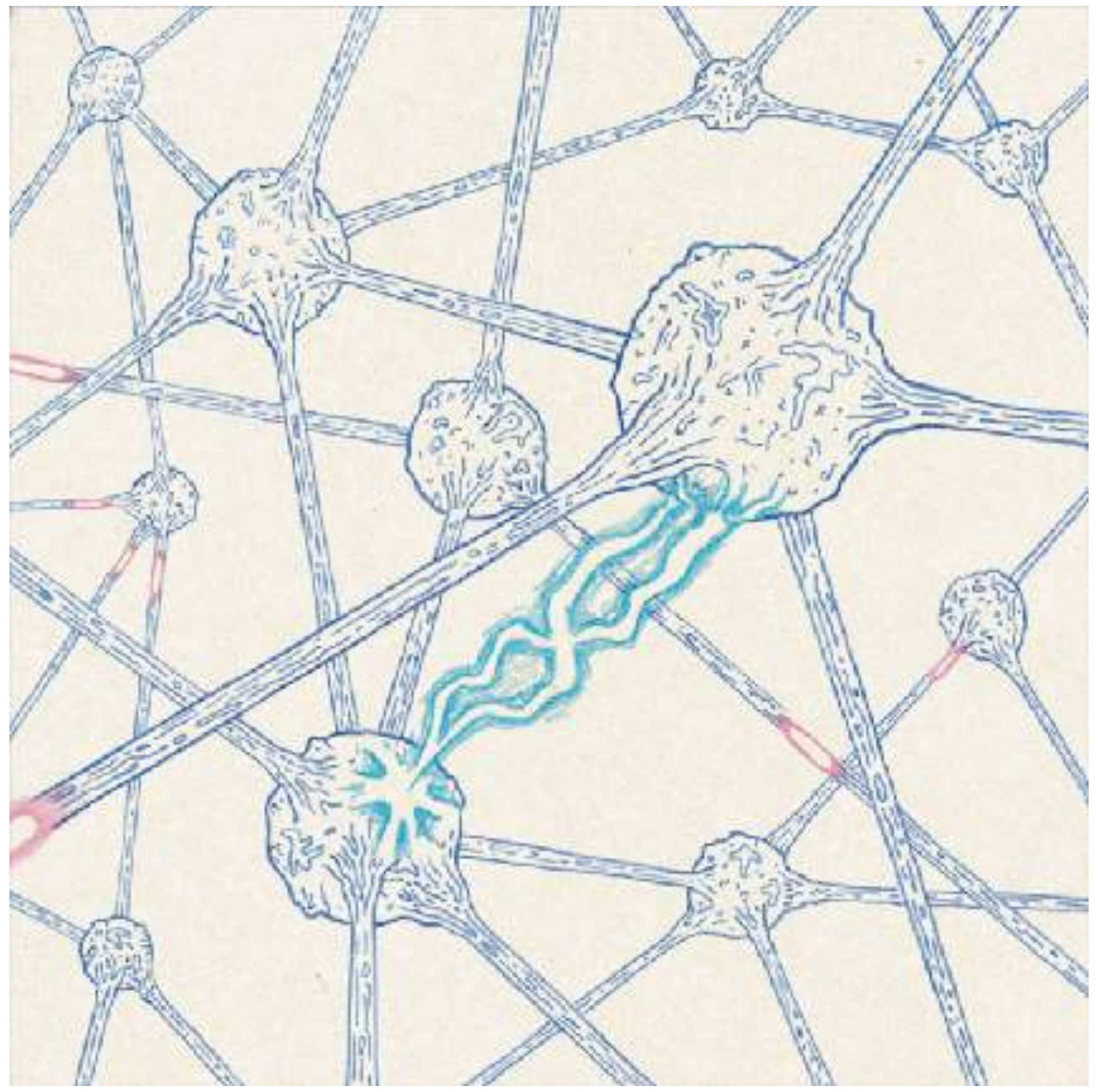
Jeff Dean!

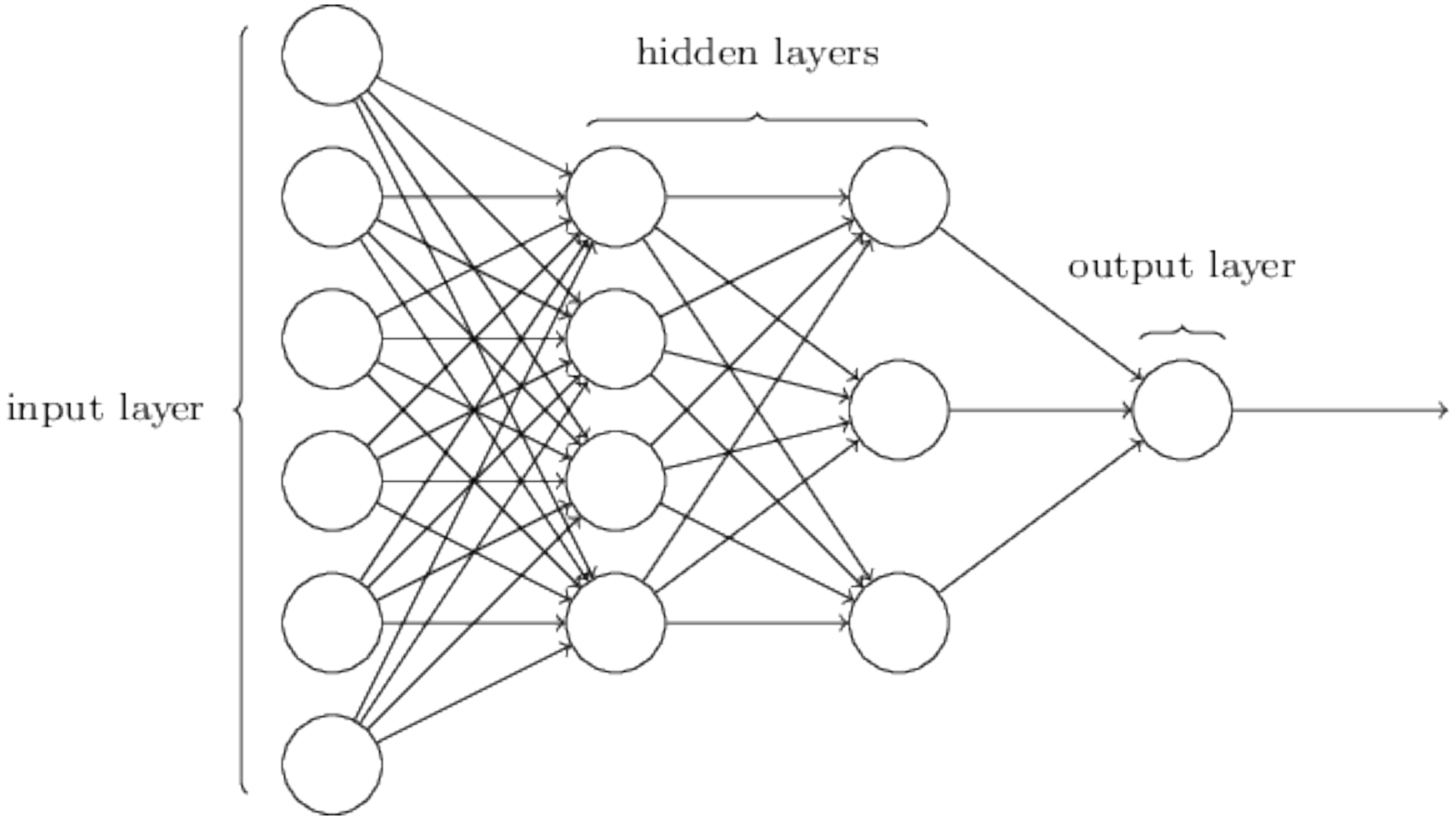


You mentioned Deep Learning? Wazup?

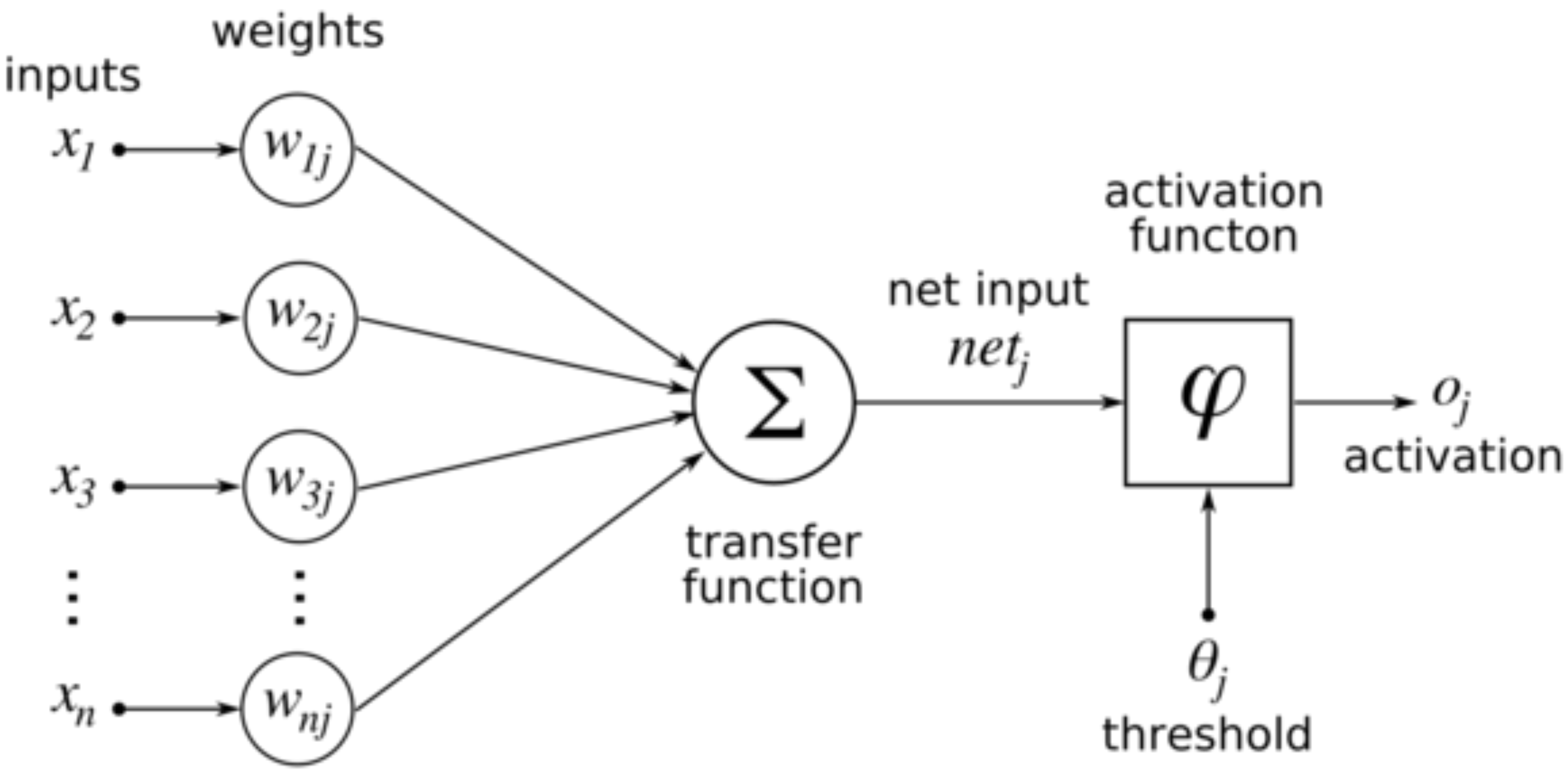
- The first question to answer: What's a Neural Network ?

- Inspired by Biology:





- Two flavours: Supervised an Unsupervised

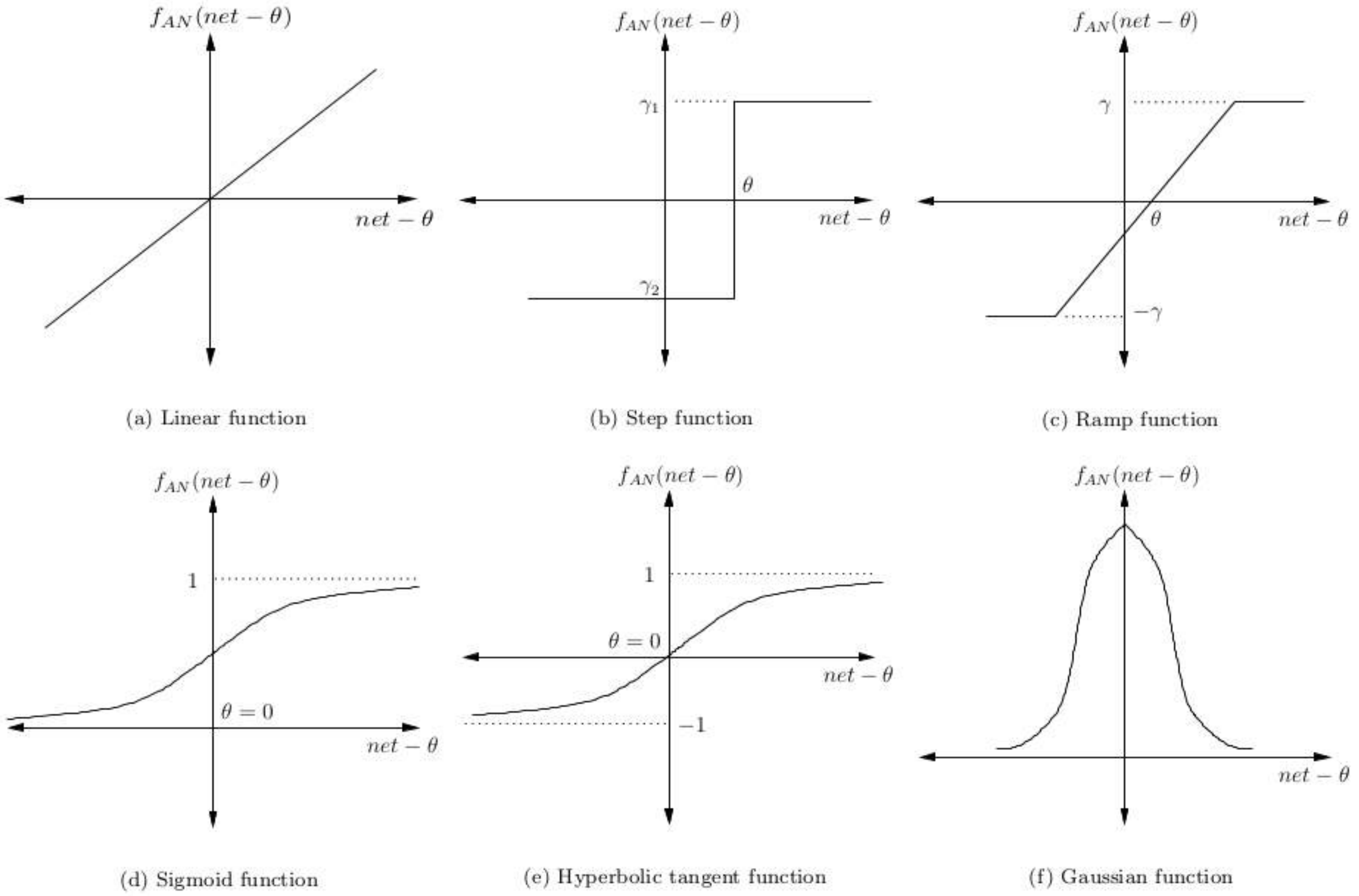


Base Idea of Neural Networks

$$f(x) = y$$

This is what we are trying to solve

Activation Functions



(a) Linear function

(b) Step function

(c) Ramp function

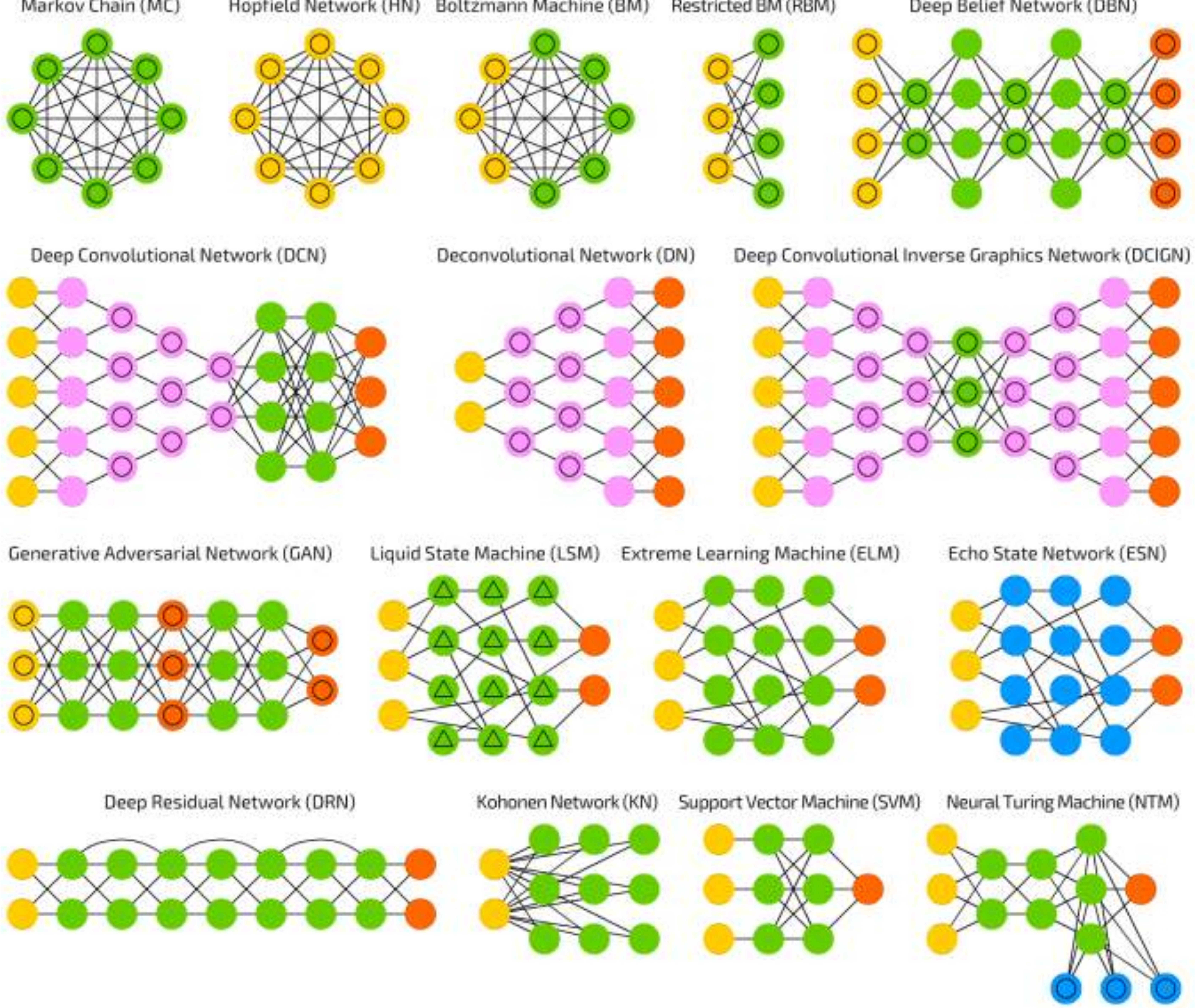
(d) Sigmoid function

(e) Hyperbolic tangent function

(f) Gaussian function

Neura

- Many kinds



What happened in the last decade...

- Algorithms: This area has seen some improvements, but most of the early wins came from fairly old ideas. Now that Deep Learning is showing success we are seeing some good advances as well.

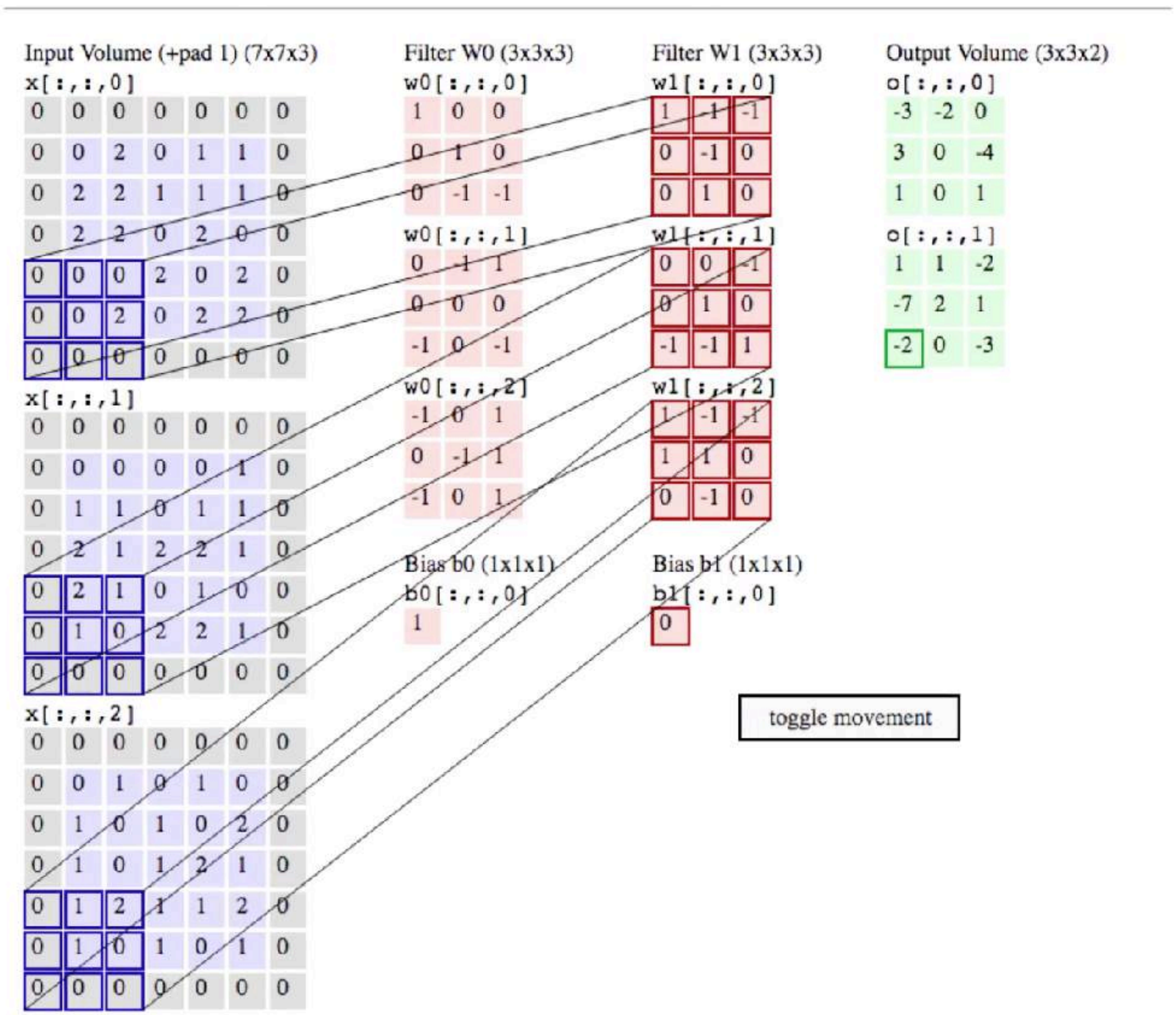
- Datasets: Training large networks is hard without large enough datasets. MNIST can only go so far in pushing the limits. Having datasets like ImageNet has really helped pushed the state of the art in vision.

- Compute: the <u>biggest game changer</u> in recent years.

163 Zetabytes
(that's 163 trillion gigabytes!)
of data per year by 2025
Google Cloud

Convolution NNs

- Large Networks
- Lots of Input Nodes
- Lots of Layers
- Each Layer with a special task



TensorFlow History: The Cat

It took 16,000 computers to identify a cat!

This was in 2012

“TensorFlow”, the name...

Tensors: multidimensional data arrays

[0.3, 0.2, 0.4, 0.5, 0.9]

But also:

```
[
  [ 0.3, 0.2, 0.4, 0.5, 0.9 ],
  [ 0.6, 0.0, 0.8, 0.5, 0.4 ],
  [ 0.2, 0.8, 0.8, 0.1, 0.3 ]
]
```

“TensorFlow”, the flow ?

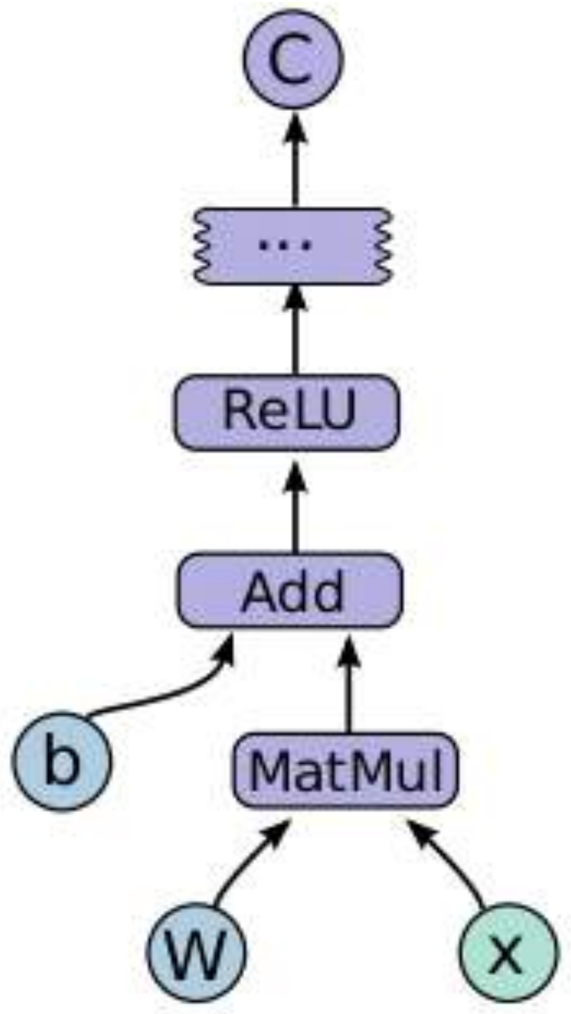
- Math Operations
  - tf.exp   tf.pow tf.tan tf.sign
- Control Flow
  - tf.cond tf.do_while
- Tensor Operations
  - tf.matmul tf.add   tf.reduce_sum tf.cumprod

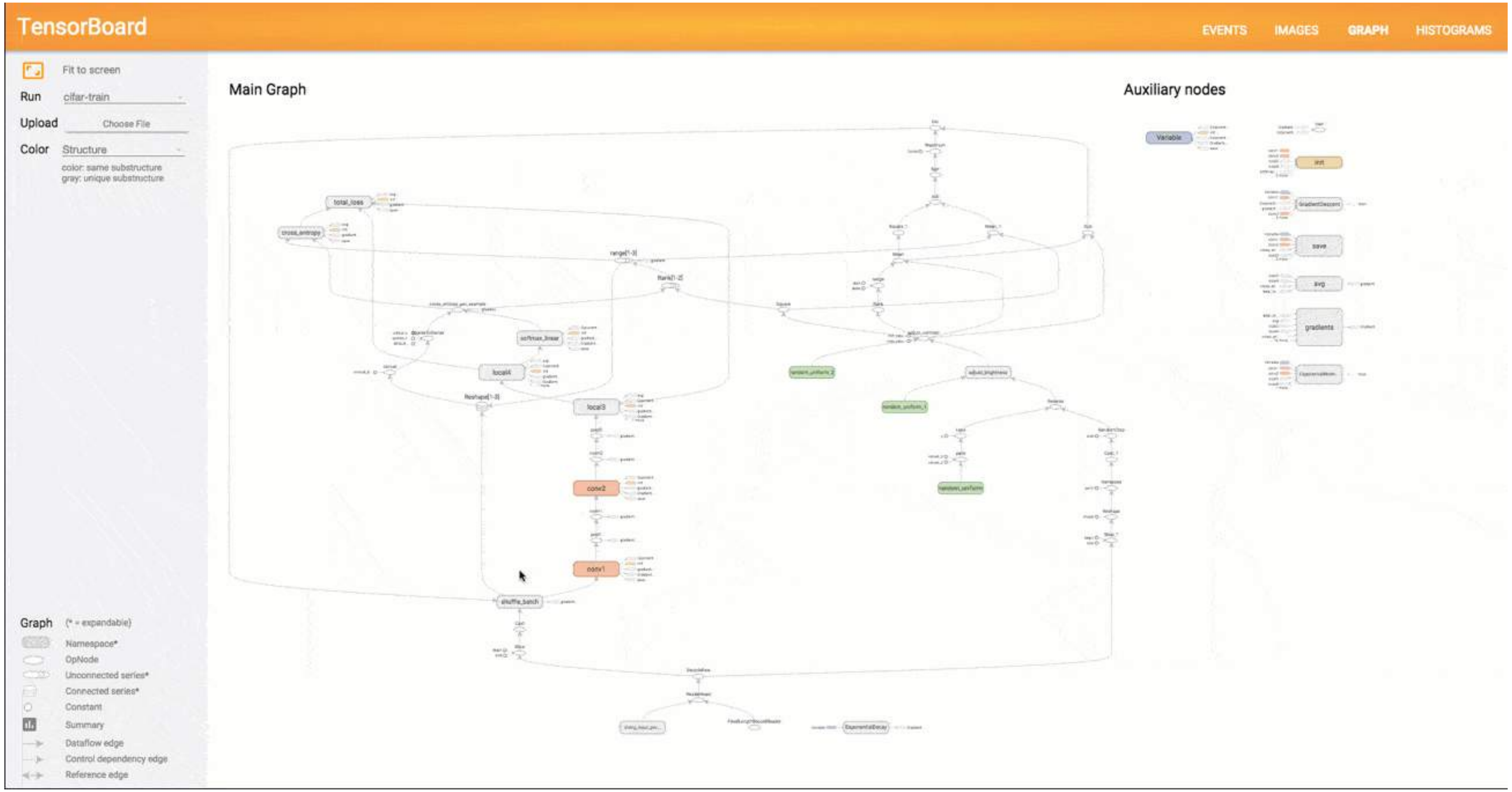
“TensorFlow”, the Graph

Computation is defined as a directed acyclic graph (DAG) to optimize an objective function

- Graph is defined in high-level language (Python)
- Graph is compiled and optimized
- Graph is executed (in parts or fully) on available low level devices (CPU, GPU)
- Data (tensors) flow through the graph
- TensorFlow can compute gradients automatically

“TensorFlow”, the Graph

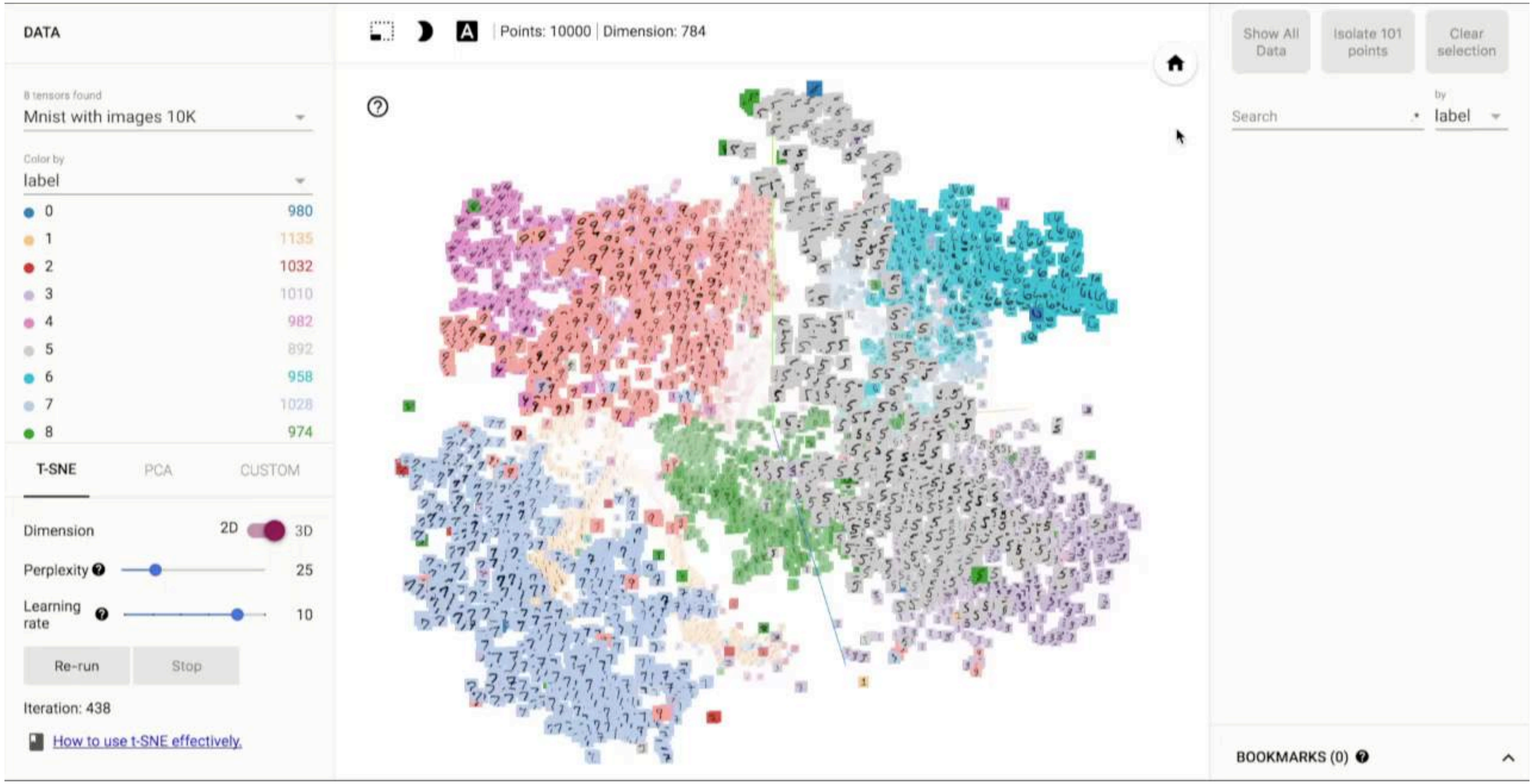




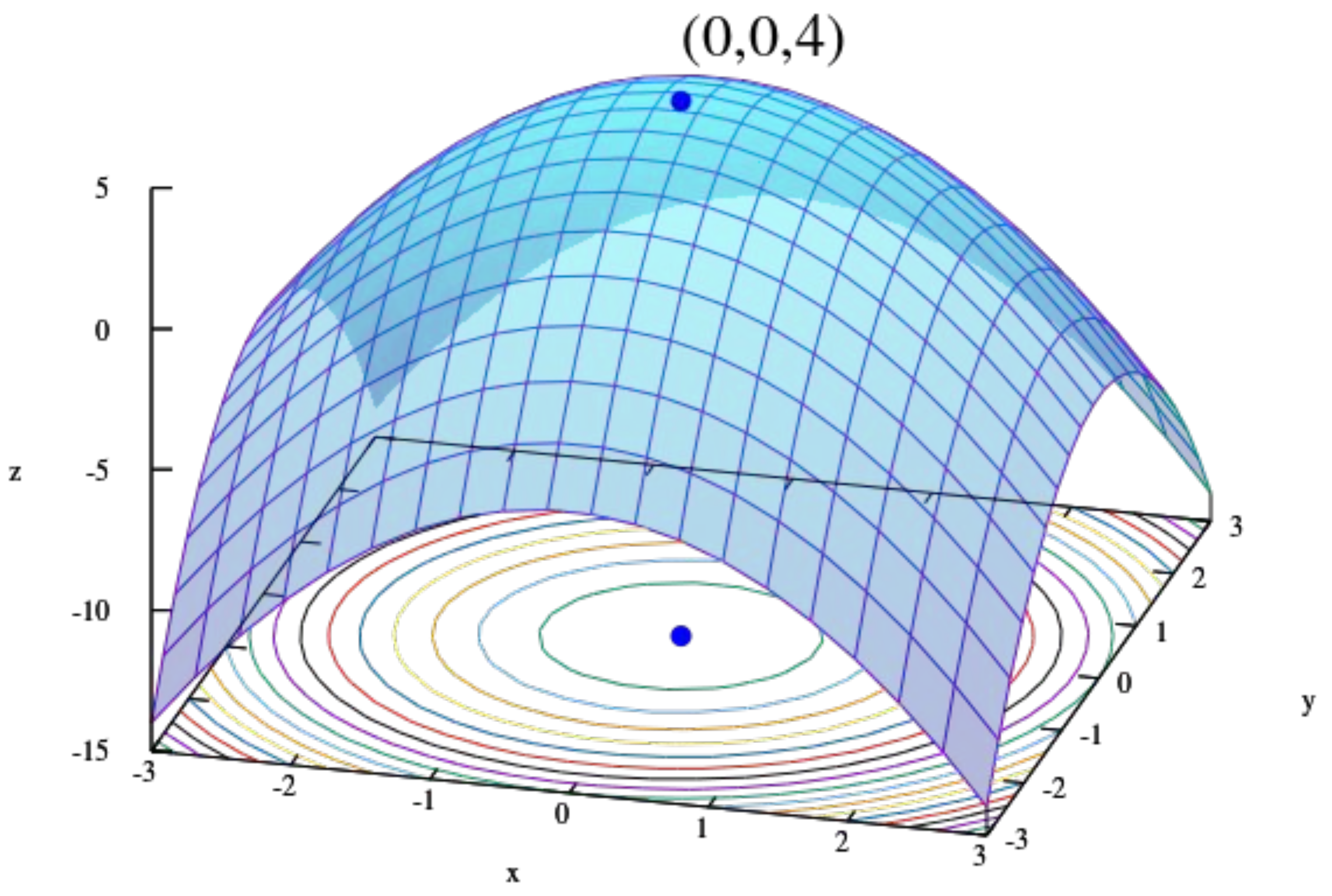
TensorBoard Projector (new)

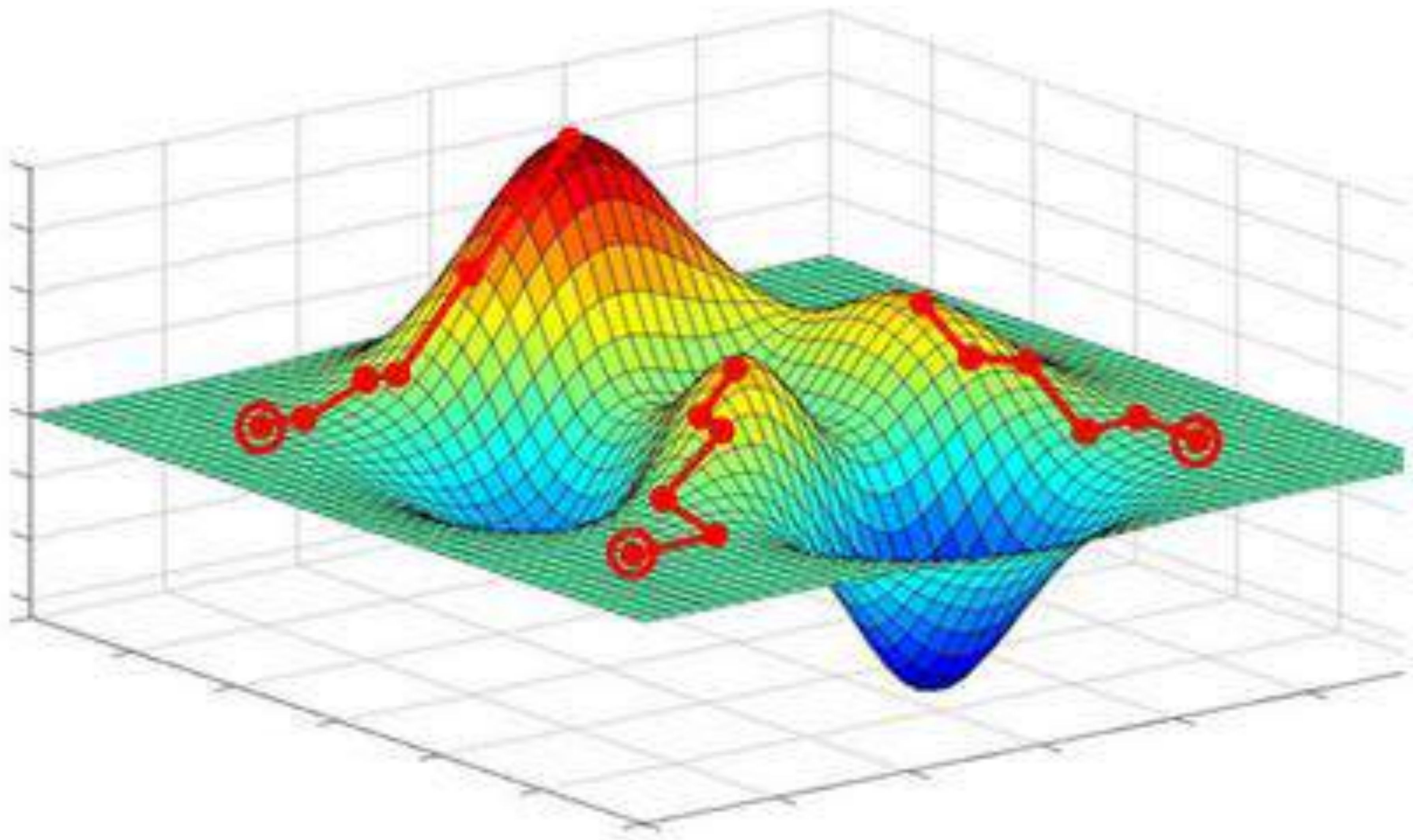
Recently launched

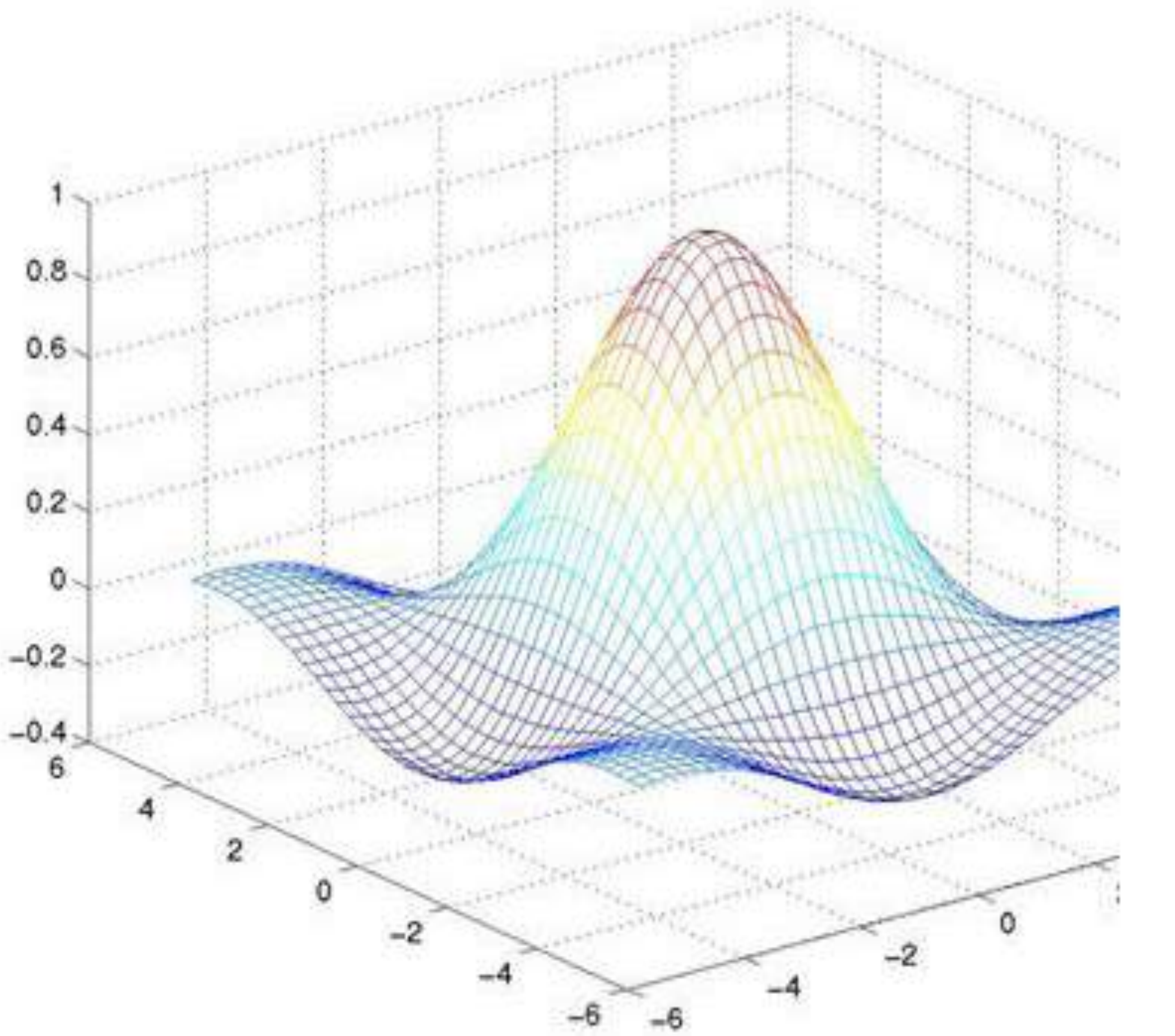
Projector



“TensorFlow”, Gradients?







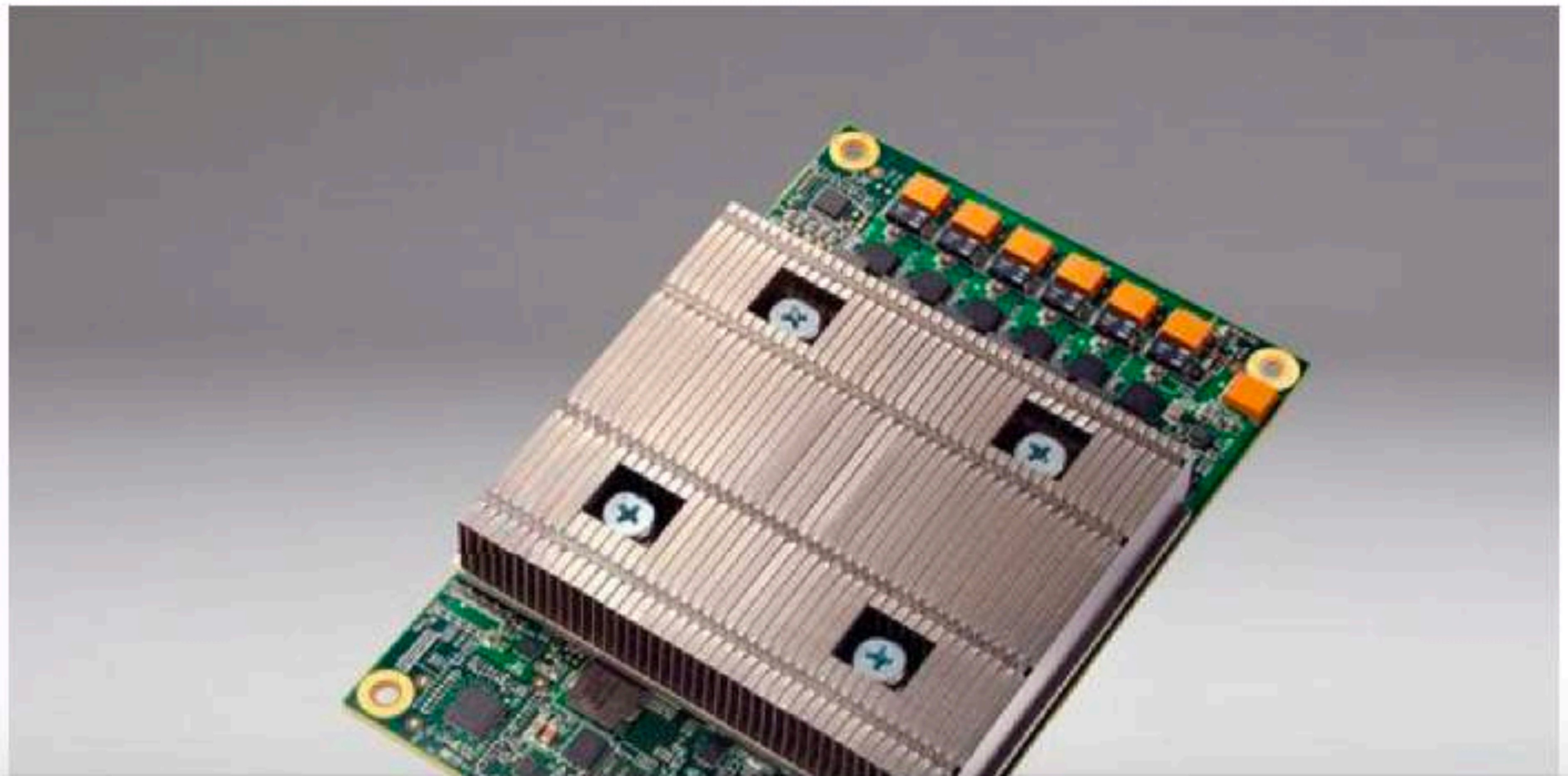
11.5 petaflops per pod

NEWS

Google's Tensor Processing Unit could advance Moore's Law 7 years into the future

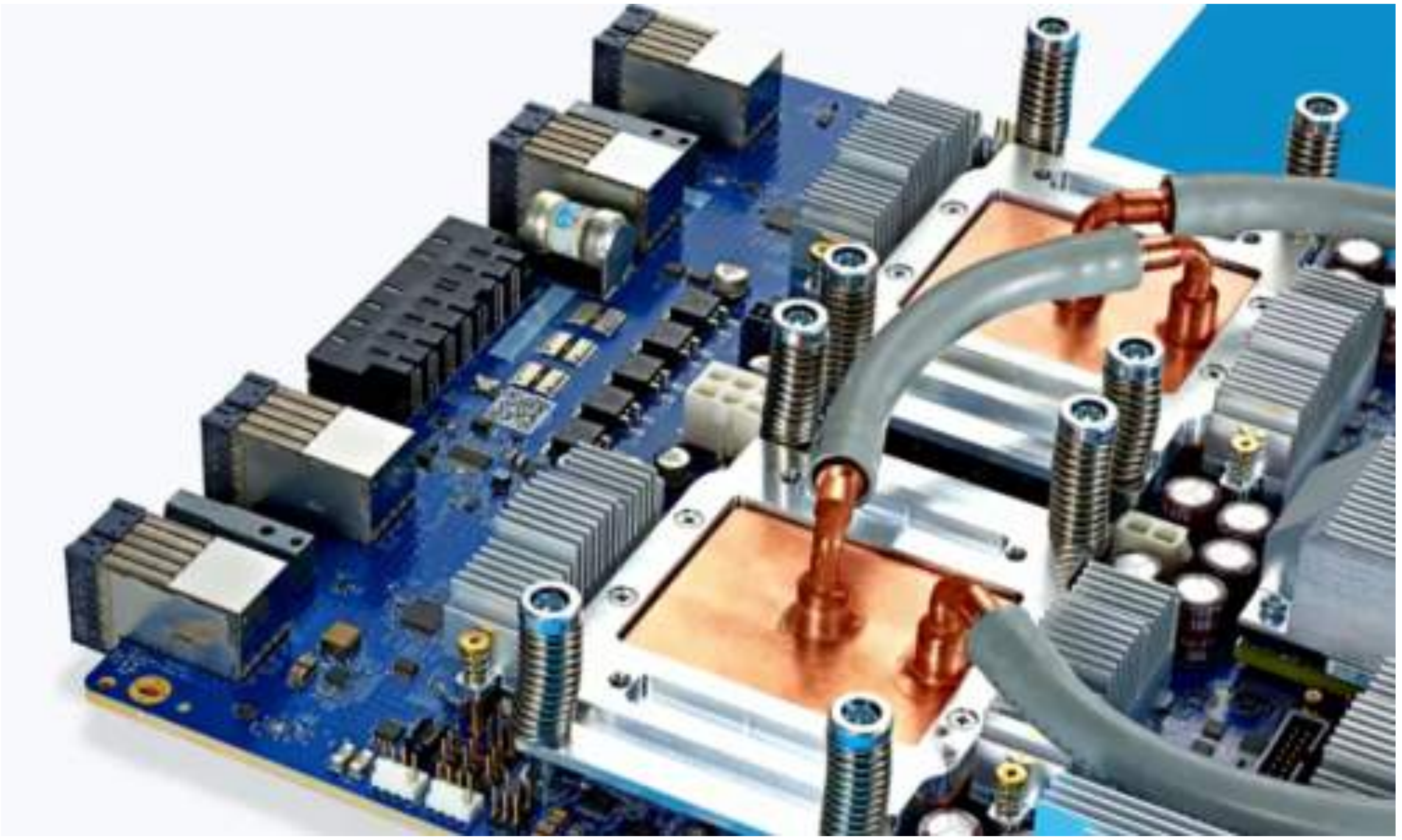
Google unveils a custom chip, which it says advances computing performance by three generations.

By Gordon Mah Ung
Executive Editor, PCWorld | MAY 18, 2016 2:08 PM PT



MORE LIKE THIS
Nvidia chief downplays challenge from Google's hyper-specialized AI chip
May's mighty new PC hardware: Mammoth graphics champions and freakishly...
Google says its AI chips smoke CPUs, GPUs in performance tests

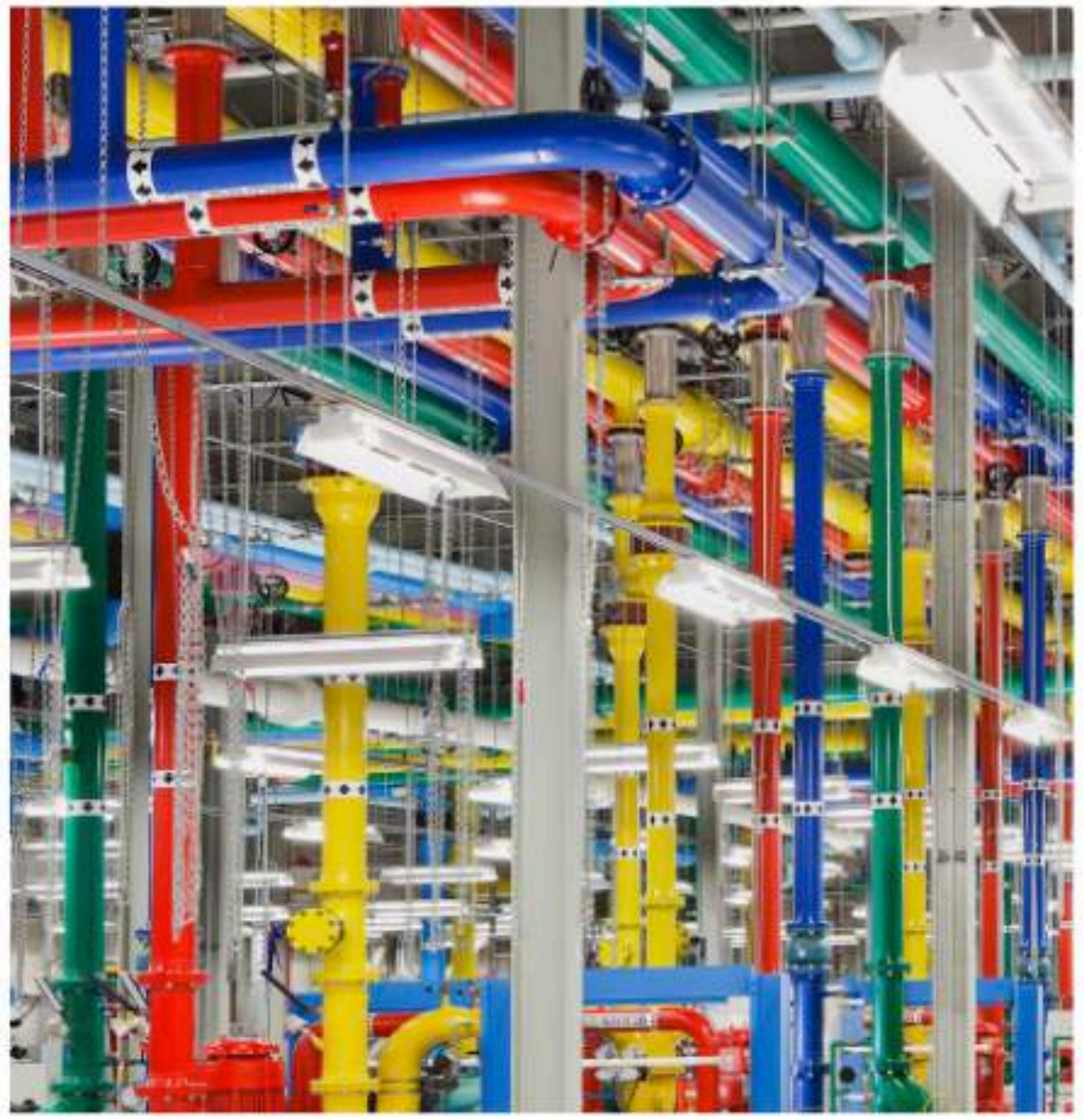
TPU 3.0



8x faster

Portable & Scalable

Your laptop



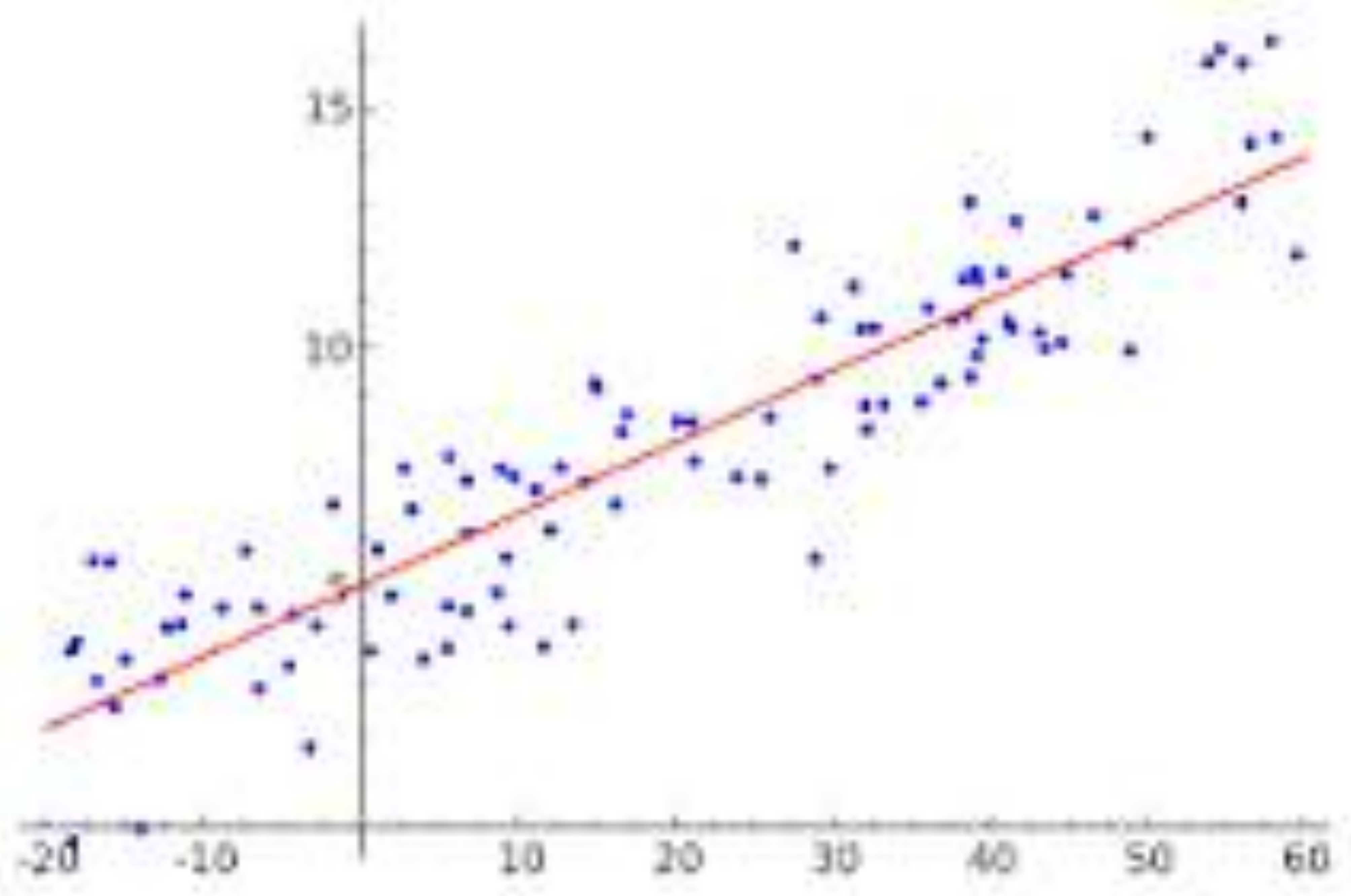
Datacenters

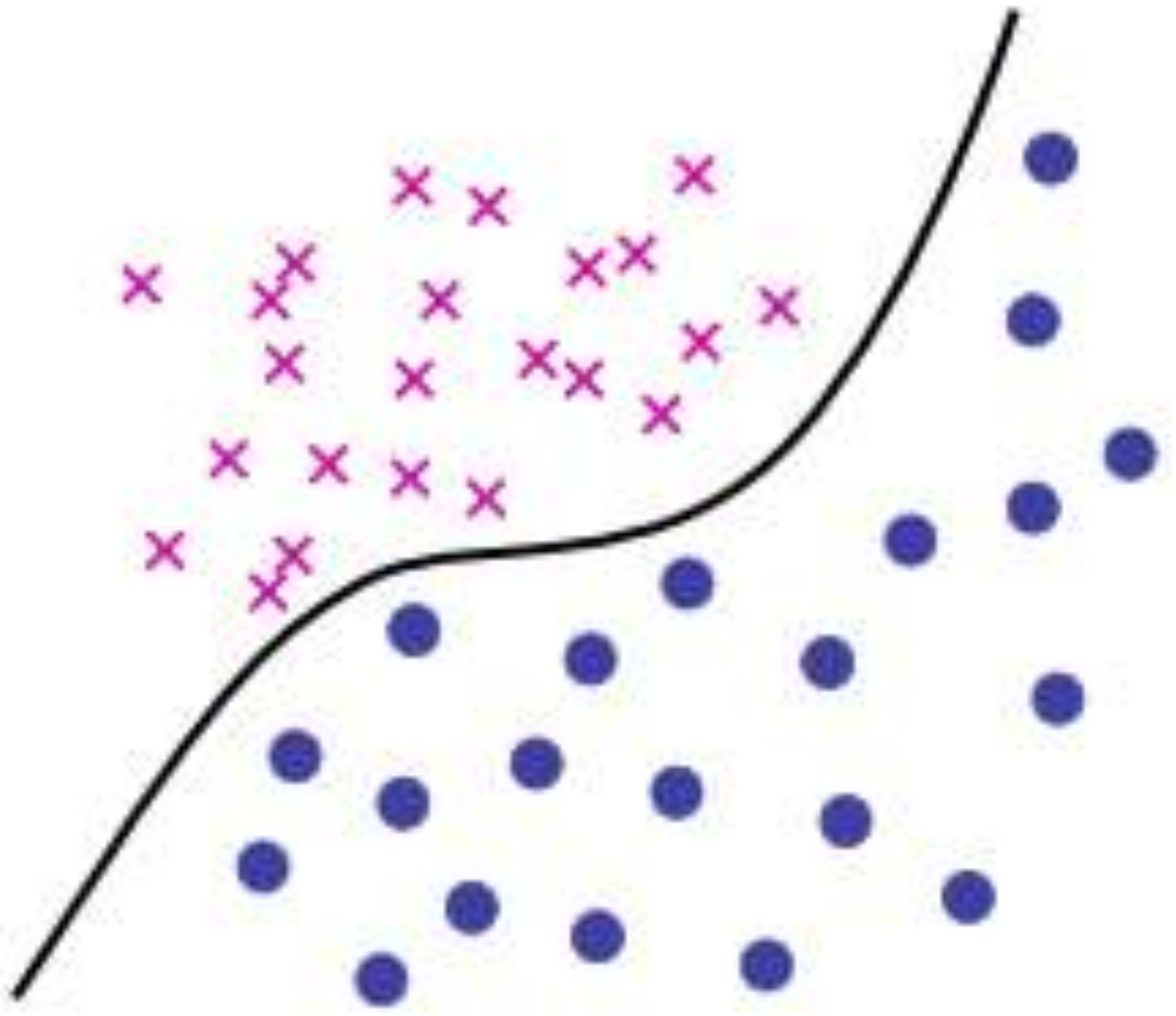
Android

iOS

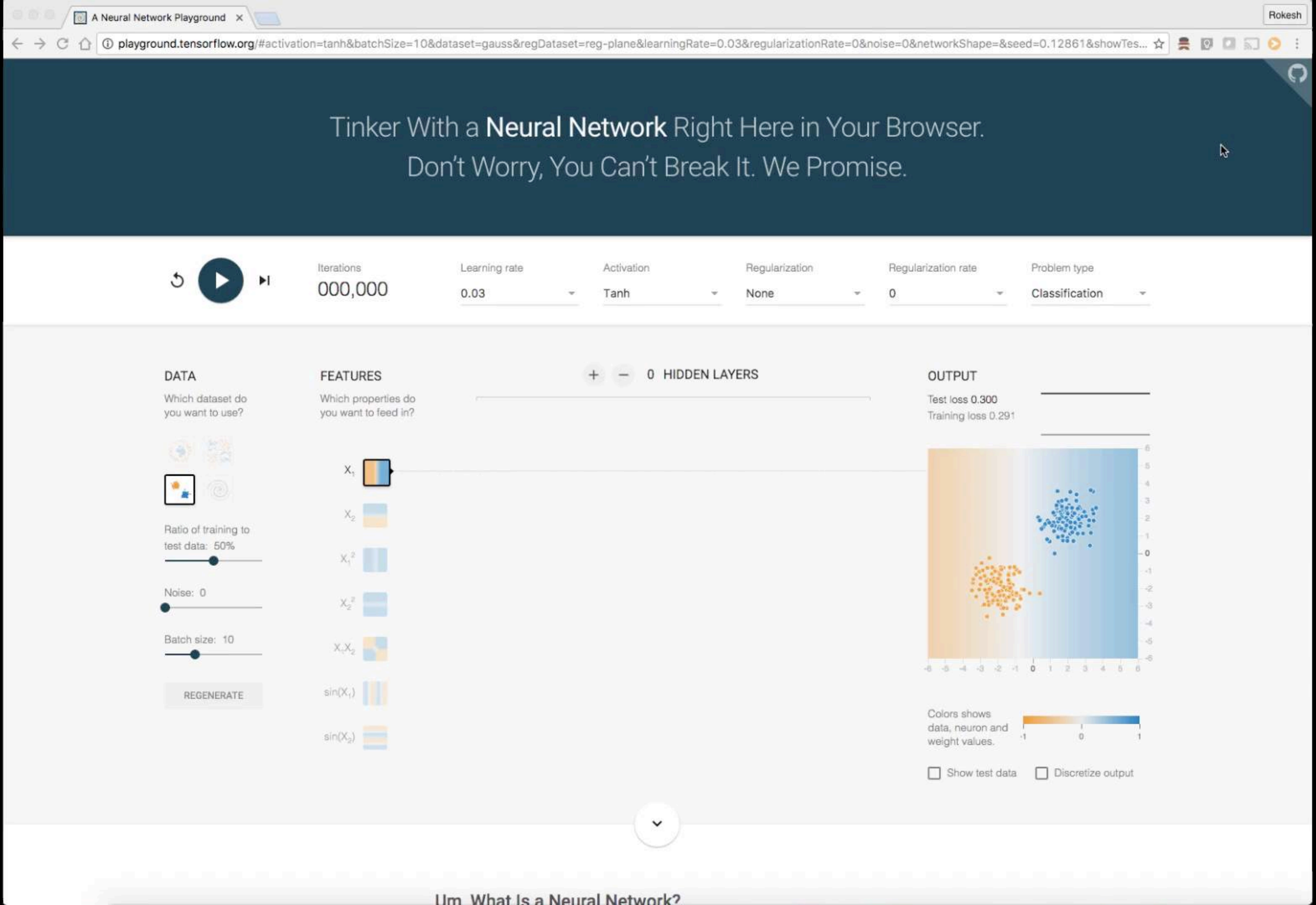
Raspberry Pi

Linear Regression & Classification

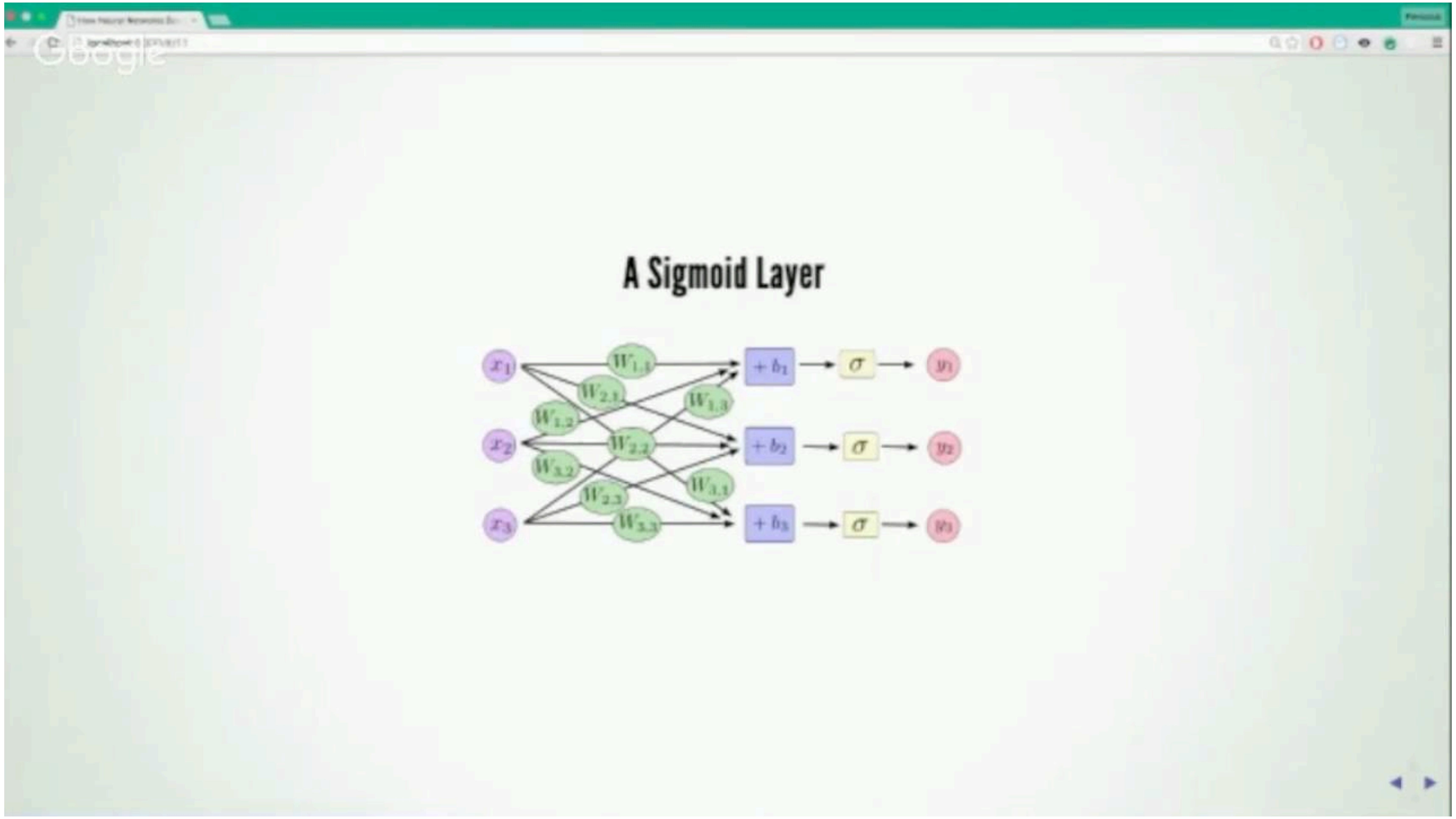




TensorFlow Playground


A Neural Network Playground
Rokesh
playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=gauss&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=&seed=0.12861&showTes...
Tinker With a Neural Network Right Here in Your Browser.
Don't Worry, You Can't Break It. We Promise.
Iterations
000,000
Learning rate
0.03
Activation
Tanh
Regularization
None
Regularization rate
0
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 50%
Noise: 0
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
0 HIDDEN LAYERS
OUTPUT
Test loss 0.300
Training loss 0.291
Colors shows data, neuron and weight values.
Show test data
Discretize output
Um, What Is a Neural Network?

what just happened in Linear Regression ?



Writing your Applications: Where to start ?

- The easiest way to go (nowadays) is Docker image
- When you start the docker image immediately a iPython Notebook is available
- Start playing around with notebooks! (we can do it together!)

Code Sample 1: Simple

```
import tensorflow as tf
import numpy as np

x = tf.constant([[1, 2]])
neg_x = tf.neg(x)

print(neg_x)

with tf.Session() as sess:
    result = sess.run(neg_x)
print(result)
```

Code Sample 2: Which processing unit?

```
import tensorflow as tf

# Creates a graph.
a = tf.constant([1.0, 2.0, 3.0, 4.0, 5.0, 6.0], shape=[2, 3], name='a')
b = tf.constant([1.0, 2.0, 3.0, 4.0, 5.0, 6.0], shape=[3, 2], name='b')
c = tf.matmul(a, b)
# Creates a session with log_device_placement set to True.
sess = tf.Session(config=tf.ConfigProto(log_device_placement=True))
# Runs the op.
print sess.run(c)

```

Code Sample 2 : output

```
root@f5662678fbad:~# python app1.py
Device mapping: no known devices.
I tensorflow/core/common_runtime/direct_session.cc:149] Device mapping:

b: /job:localhost/replica:0/task:0/cpu:0
I tensorflow/core/common_runtime/simple_placer.cc:388] b: /job:localhost/replica:0/task:0/cpu:0
a: /job:localhost/replica:0/task:0/cpu:0
I tensorflow/core/common_runtime/simple_placer.cc:388] a: /job:localhost/replica:0/task:0/cpu:0
MatMul: /job:localhost/replica:0/task:0/cpu:0
I tensorflow/core/common_runtime/simple_placer.cc:388] MatMul: /job:localhost/replica:0/task:0/cpu:0
[[ 22.  28.]
 [ 49.  64.]]
```

Code Sample 2 : result

```
import tensorflow as tf

# Creates a graph.
with tf.device('/cpu:0'):
  a = tf.constant([1.0, 2.0, 3.0, 4.0, 5.0, 6.0], shape=[2, 3], name='a')
  b = tf.constant([1.0, 2.0, 3.0, 4.0, 5.0, 6.0], shape=[3, 2], name='b')
  c = tf.matmul(a, b)
# Creates a session with log_device_placement set to True.
sess = tf.Session(config=tf.ConfigProto(log_device_placement=True))
# Runs the op.
print sess.run(c)
```

```
import tensorflow as tf
x = tf.placeholder(tf.float32, [None, 784])

#layer 1
W1 = tf.Variable(tf.zeros([784, 100]))
b1 = tf.Variable(tf.zeros([100]))
y1 = tf.matmul(x, W1) + b1 softmax

#layer 2
W2 = tf.Variable(tf.zeros([100, 10]))
b2 = tf.Variable(tf.zeros([10]))
y2 = tf.nn.softmax(tf.matmul(y1, W2) + b2)

#output
y = y2
y_ = tf.placeholder(tf.float32, [None, 10])
```

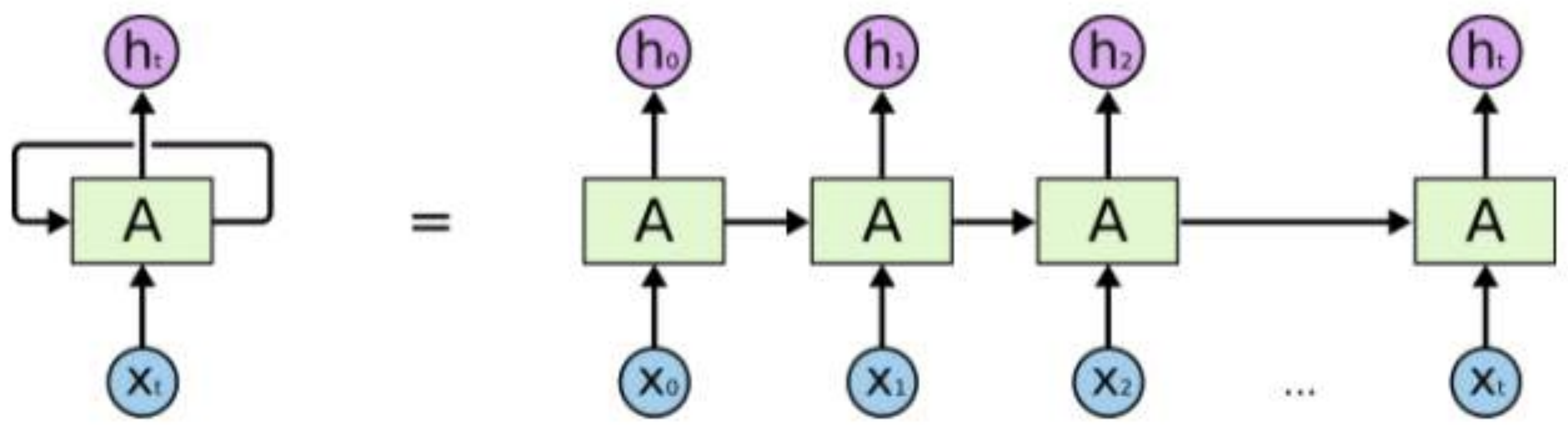
What is the process ?

Examples

- https://github.com/tensorflow/models/tree/master/research/im2txt
- "Lip reading"
- Human-like sound: WaveNet
- Skin cancer detection
- AI Experiments by Google
- Learn how to fly

LSTM Networks

- Special Kind of Recurrent Neural Network: Long Short Term Network



An unrolled recurrent neural network.

- *"Humans don't start their thinking from scratch every second. As you read this essay, you understand each word based on your understanding of previous words. You don't throw everything away and start thinking from scratch again. Your thoughts have persistence"*

- ***Best example: Google Translate***

Companies using TensorFlow

airbnb
AIRBUS DEFENCE & SPACE
ARM
CEVA
CI&T
DeepMind
Dropbox
ebay
Google
IBM Power Systems
intel
JD.COM 京东
mi
Movidius
QUALCOMM
quantiphi
SAP
UBER

Machine Learning Use Cases

Manufacturing

- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Propensity to buy
- Demand forecasting
- Process optimization
- Telematics

Retail

- Predictive inventory planning
- Recommendation engines
- Upsell and cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value

Healthcare and Life Sciences

- Alerts and diagnostics from real-time patient data
- Disease identification and risk satisfaction
- Patient triage optimization
- Proactive health management
- Healthcare provider sentiment analysis

Travel and Hospitality

- Aircraft scheduling
- Dynamic pricing
- Social media – consumer feedback and interaction analysis
- Customer complaint resolution
- Traffic patterns and congestion management

Financial Services

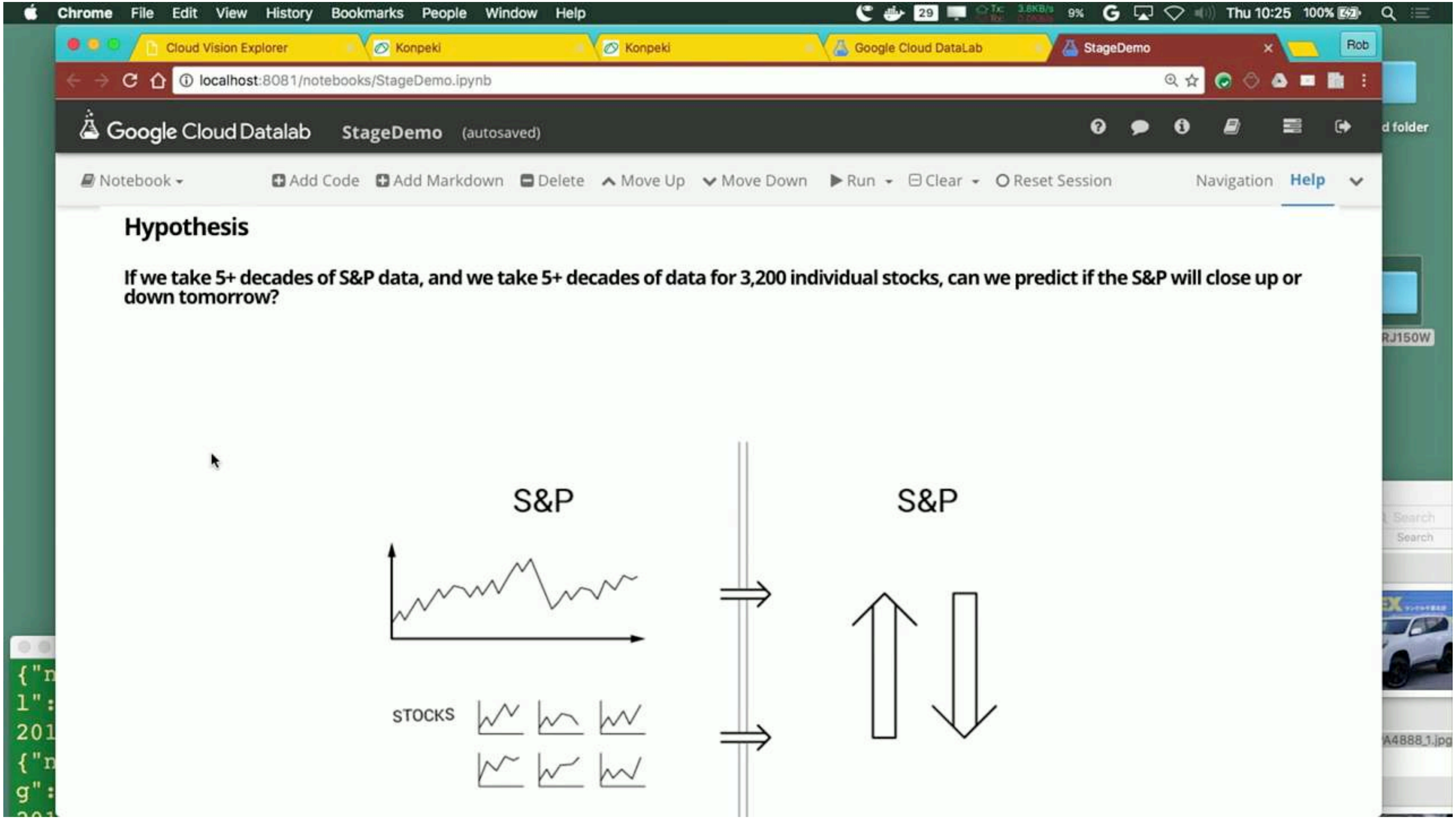
- Risk analytics and regulation
- Customer Segmentation
- Cross-selling and up-selling
- Sales and marketing campaign management
- Credit worthiness evaluation

Energy, Feedstock and Utilities

- Power usage analytics
- Seismic data processing
- Carbon emissions and trading
- Customer-specific pricing
- Smart grid management
- Energy demand and supply optimization

Google Cloud ML Engine

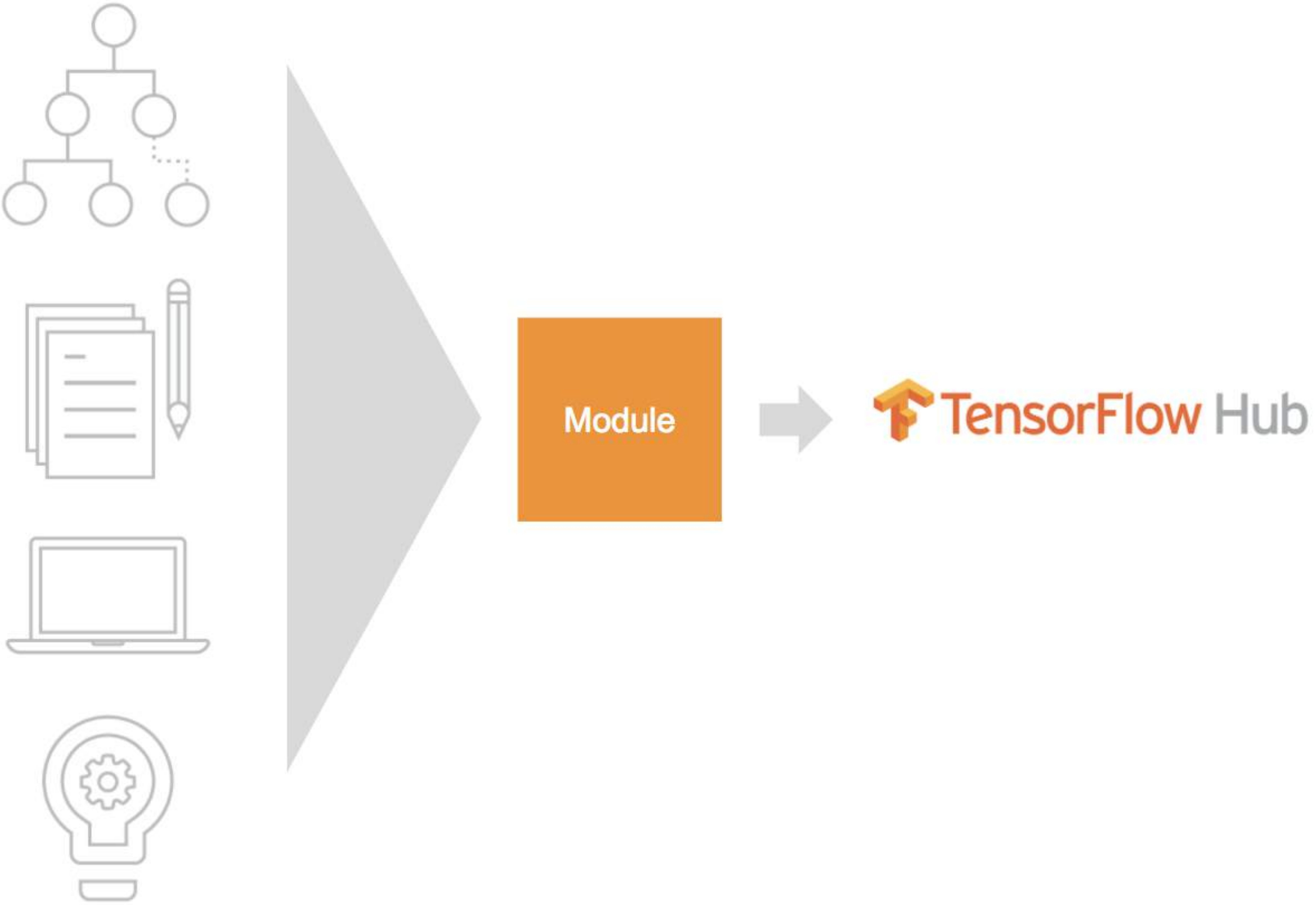
Announced in October 2016 at Google Horizon Event

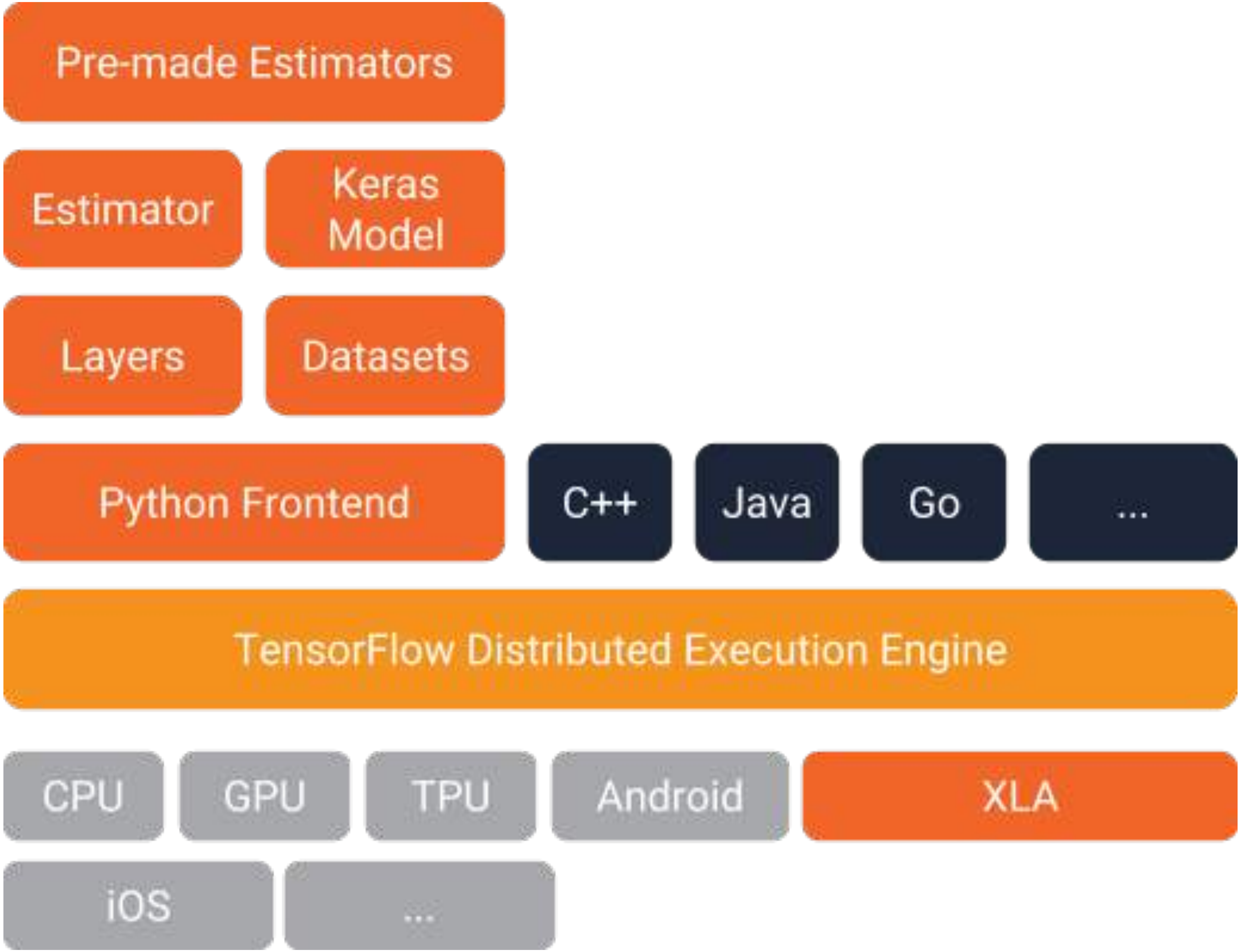

Chrome File Edit View History Bookmarks People Window Help
Thu 10:25 100%
Cloud Vision Explorer
Konpeki
Konpeki
Google Cloud DataLab
StageDemo
Rob
localhost:8081/notebooks/StageDemo.ipynb
Google Cloud Datalab
StageDemo (autosaved)
Notebook
Add Code
Add Markdown
Delete
Move Up
Move Down
Run
Clear
Reset Session
Navigation
Help
Hypothesis
If we take 5+ decades of S&P data, and we take 5+ decades of data for 3,200 individual stocks, can we predict if the S&P will close up or down tomorrow?
S&P
S&P
STOCKS

Recent developments

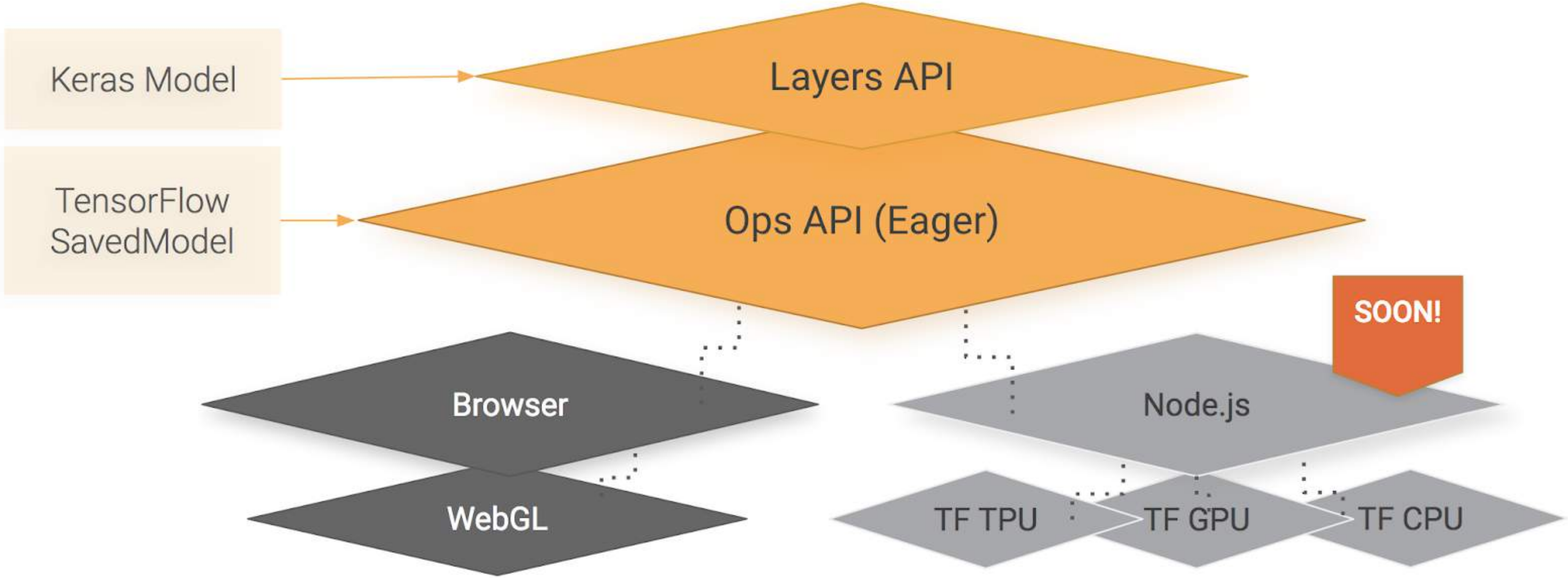
TensorFlow :Keras & Estimators

- A more abstract layer on top of TensorFlow.
  - Makes code more readable.
- https://blog.keras.io/keras-as-a-simplified-interface-to-tensorflow-tutorial.html
- TensorFlow High-level APIs


Module
TensorFlow Hub


Pre-made Estimators
Estimator
Keras Model
Layers
Datasets
Python Frontend
C++
Java
Go
...
TensorFlow Distributed Execution Engine
CPU
GPU
TPU
Android
XLA
iOS
...

TensorFlow.js



Transfer Learning

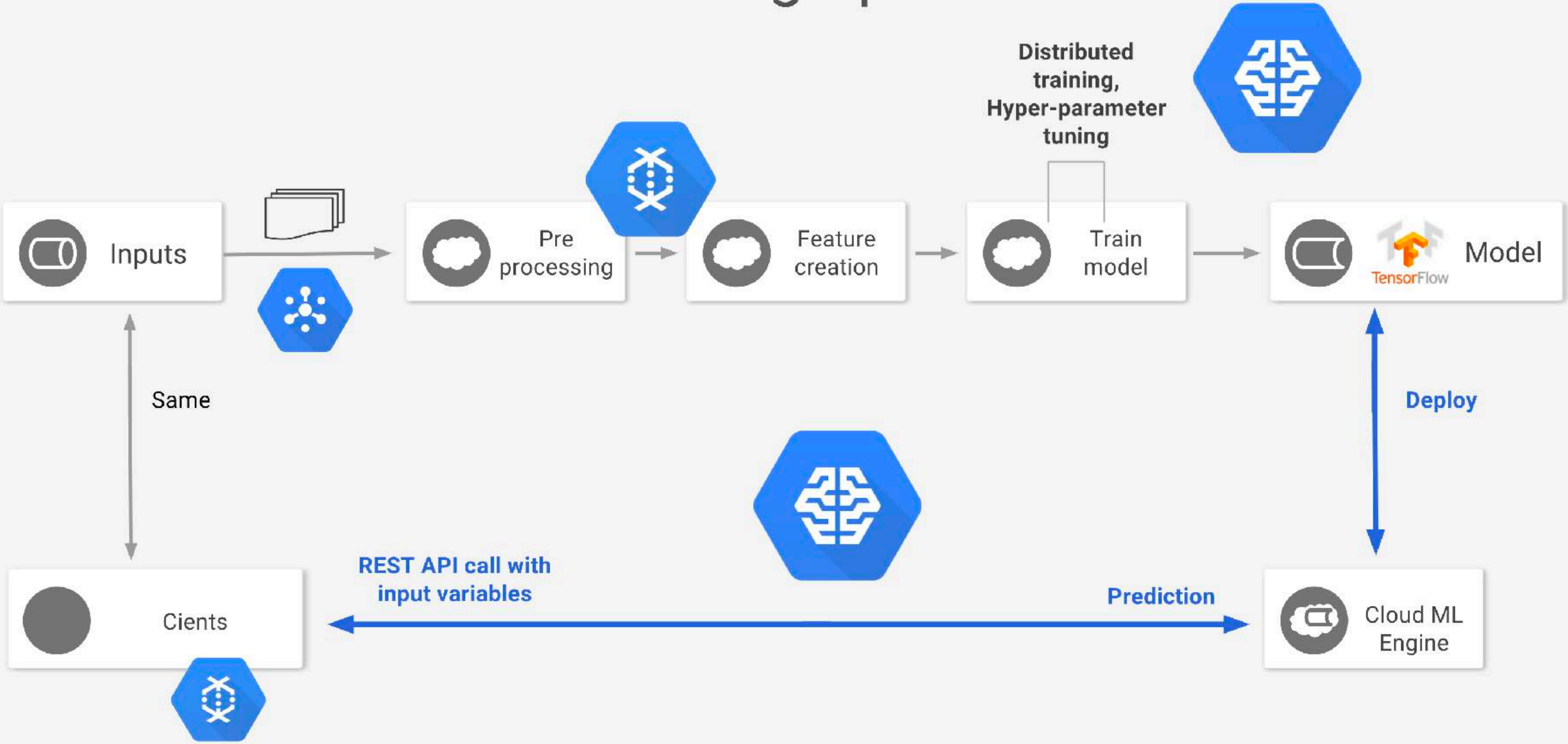
- *Transfer learning is a machine learning method which utilizes a pre-trained neural network. For example, the image recognition model called Inception-v3 consists of two parts:*
    - *Feature extraction part with a convolutional neural network.*
    - *Classification part with fully-connected and softmax layers.*
- *In transfer learning, when you build a new model to classify your original dataset, you reuse the feature extraction part and re-train the classification part with your dataset. Since you don't have to train the feature extraction part (which is the most complex part of the model), you can train the model with less computational resources and training time.*

GAN: Generative Adversarial Networks

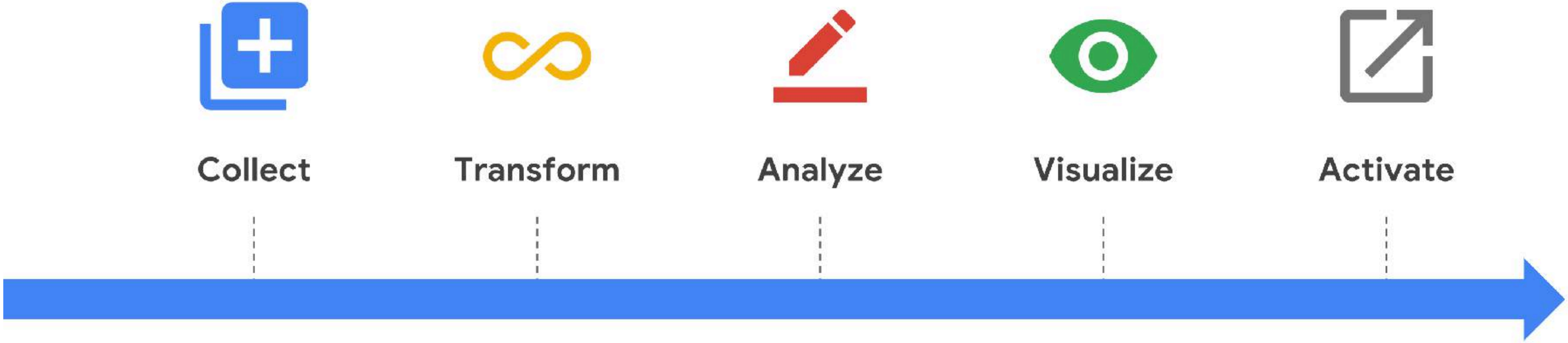
- http://blog.aylien.com/introduction-generative-adversarial-networks-code-tensorflow/
- http://blog.evjang.com/2016/06/generative-adversarial-nets-in.html
- https://bamos.github.io/2016/08/09/deep-completion/
- https://arxiv.org/pdf/1701.00160.pdf
- A new area of research where “generation of output” is done (here’s a popular one)

Using in Production on ML Engine

A useful and common design pattern

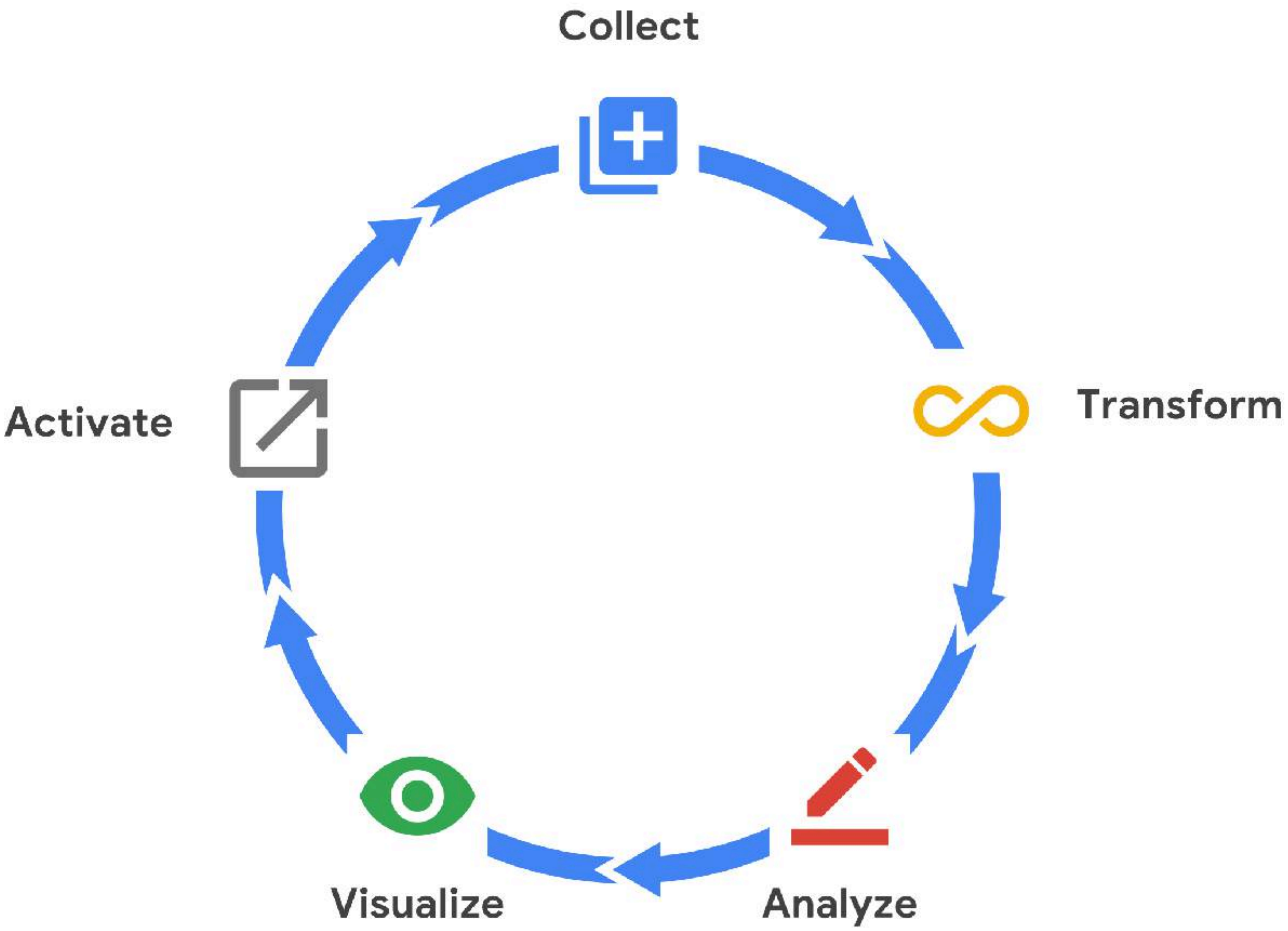


Manage your data journey seamlessly



Our vision

- **More** data sets
- **More** ML models
- **More** activation connectors



Google Cloud

Possibilities…many, but I want to highlight one specifically

AlphaGo

BBC | News | Sport | Weather | iPlayer | TV | Radio | More | Search

NEWS

Find local news

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertainment & Arts | More

Technology

Google achieves AI 'breakthrough' at Go

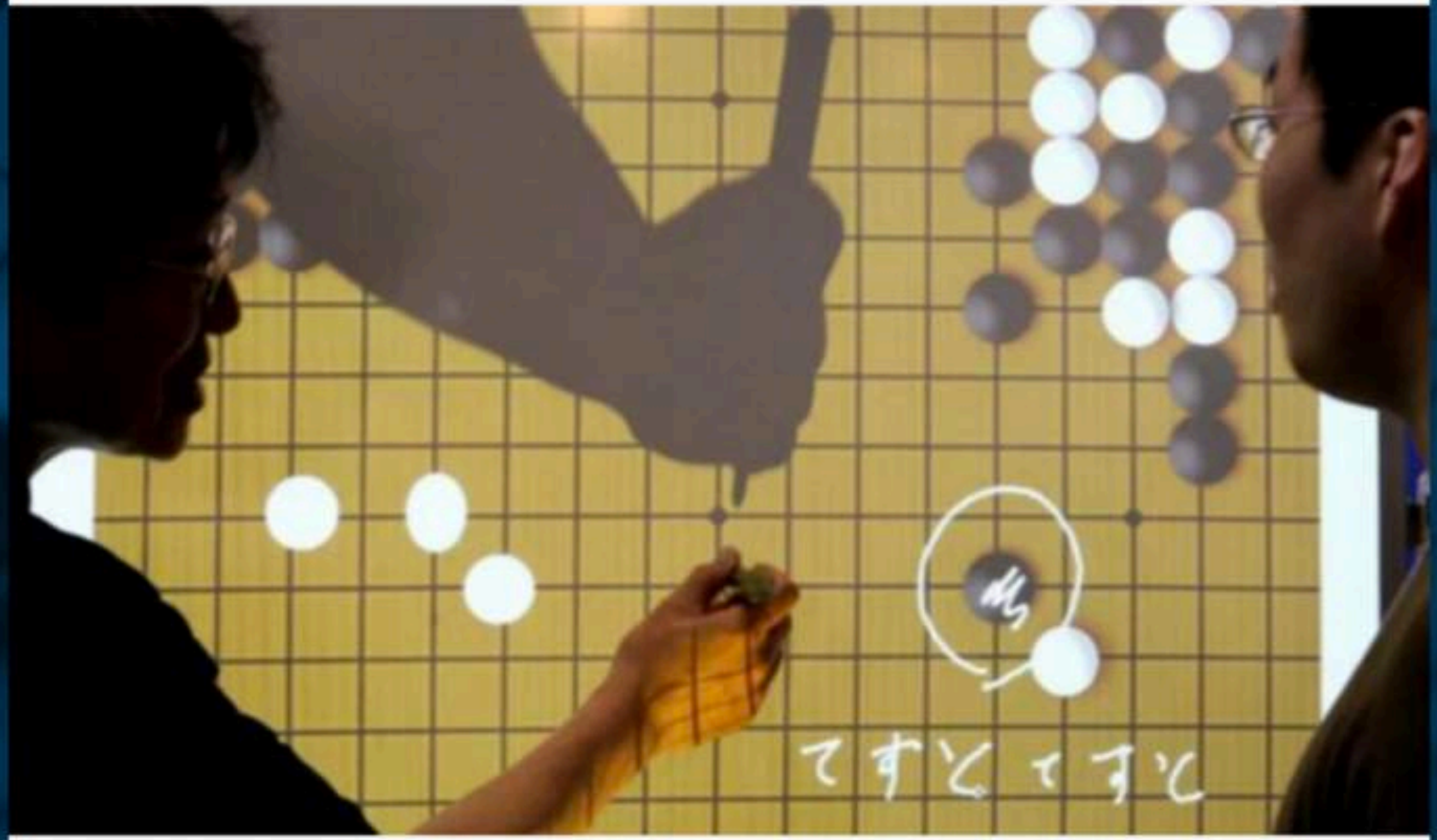
An artificial intelligence program developed by Google beats Europe's top player at the ancient Chinese game of Go, about a decade earlier than expected.

27 January 2016 | Technology

- How did they do it?
- What is the game Go?
- Facebook trains AI to beat humans at Go

Google's AI just cracked the game that supposedly no computer could beat

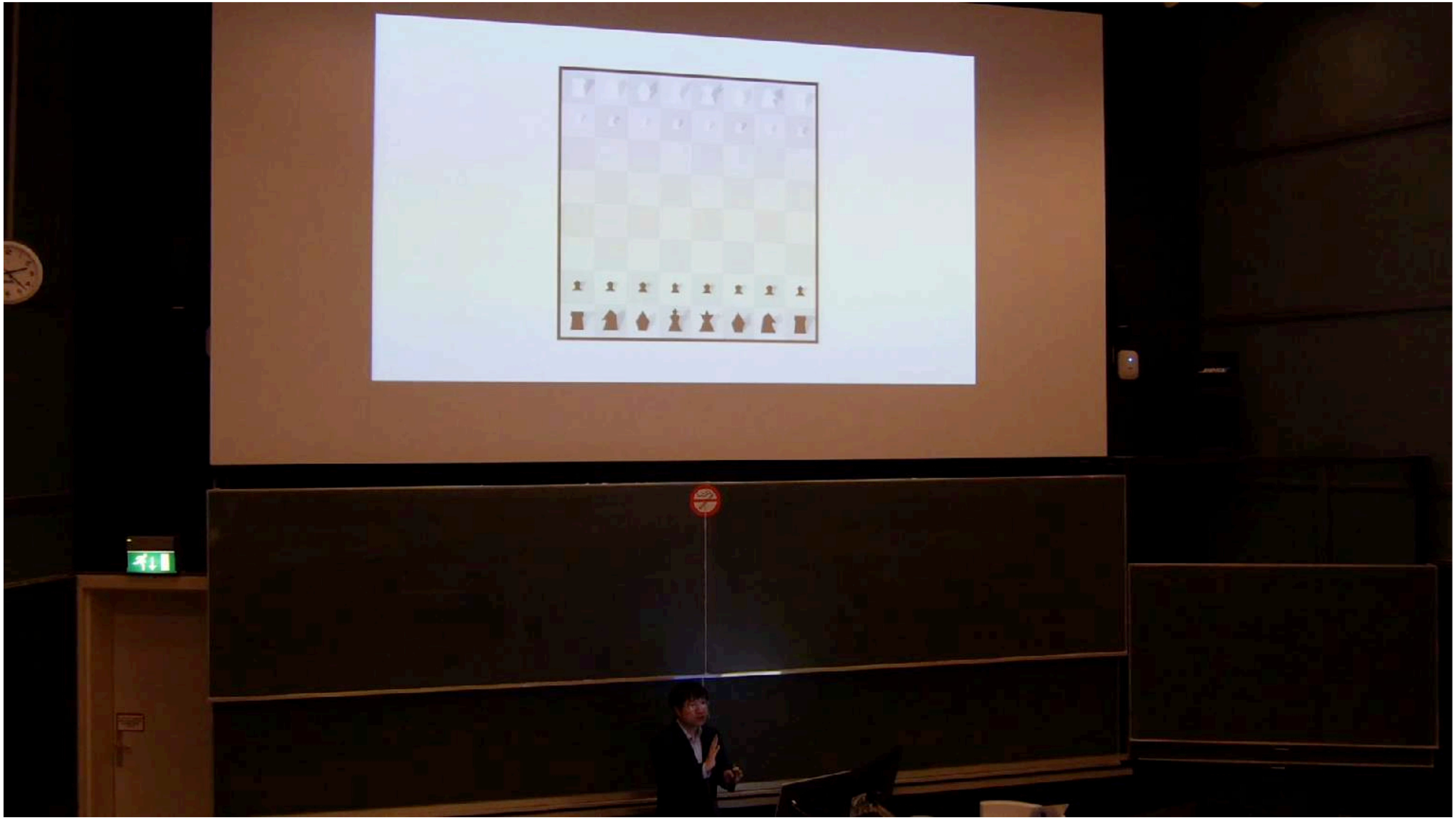
By Mike Murphy | January 27, 2016



Going up. (Reuters/Kiyoshi Ota)

Computers have slowly started to encroach on activities we previously believed only the brilliantly sophisticated human brain could handle. IBM's Deep Blue supercomputer beat Grand Master Garry Kasparov at chess in 1997, and in 2011 IBM's Watson beat former human winners at the quiz game *Jeopardy*. But the ancient board game Go has long been one of the major goals of artificial intelligence research. It's understood to be one of the most difficult games for computers to handle due to the sheer number of possible moves a player can make at any given point. Until now, that is.

Solution space



“The possibilities and power are innumerable”

– Aja Huang

After beating the world's elite Go players, Google's AlphaGo AI is retiring

Posted May 27, 2017 by Jon Russell (@jonrussell)

Next Story

Google's AlphaGo — the AI developed to tackle the world's most demanding strategy game — is stepping down from competitive matches after defeating the world's best talent. The latest to succumb is Go's top-ranked player, Ke Jie, who lost 3-0 in a series hosted in China


NETFLIX
WATCH FREE FOR A MONTH
GET STARTED
SIGN IN
Google DeepMind
ALPHAGO
AlphaGo
2017 | undefined | 1h 30m
Seemingly simple but deceptively complex, the game of "Go" serves as the backdrop for this battle between artificial intelligence and man.
Genres: Documentaries, Science & Technology Documentaries
Director: Greg Kohs


AlphaGo Zero
0:02 / 2:13
YouTube

What's next ?

- tensorflow.org
- github.com/tensorflow
- Want to learn more?
- Udacity class on Deep Learning, goo.gl/iHssII
- https://www.tensorflow.org/tutorials/

- Guides, codelabs, videos

- MNIST for Beginners, goo.gl/tx8R2b
- TF Learn Quickstart, goo.gl/uiefRn
- TensorFlow for Poets, goo.gl/bVjFIL
- ML Recipes, goo.gl/KewA03
- TensorFlow and Deep Learning without a PhD, goo.gl/pHeXe7

Google.ai

Research

Tools

Applied AI

Final Note, by the man himself!

What will the Machine Learning landscape look like in five to ten years?

Thanks!

www.tensorflow.org

https://cloud.google.com/free/