vietnamworks
Move Up!

(Big) Data, (Deep) Learning and AI

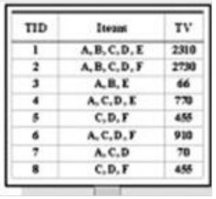
When big data hits machine learning

Phạm Thành Lâm | Founder @ SaigonApps
17.09.2016

Big Picture: Big Data - Machine Learning/ Data Mining

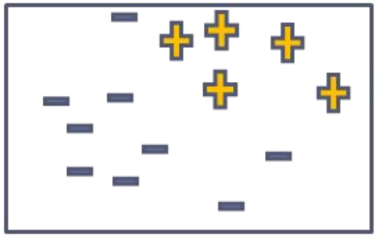
Machine Learning

Frequent pattern

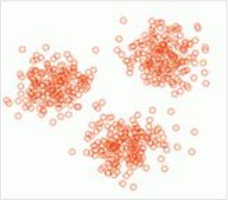


TID	Items	TV
1	A, B, C, D, E	2310
2	A, B, C, D, F	2730
3	A, B, E	66
4	A, C, D, E	770
5	C, D, F	455
6	A, C, D, F	910
7	A, C, D	70
8	C, D, F	455

Supervised

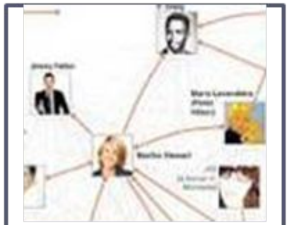


Unsupervised

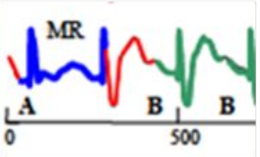


Recommendation

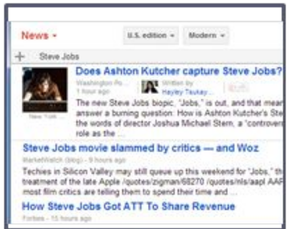
Graphs



Stream Data



Text Data



News ▾ U.S. edition ▾ Modern ▾

+ Steve Jobs

Does Ashton Kutcher capture Steve Jobs?

Washington Po... 1 hour ago

Written by Hayley Tsukay...

The new Steve Jobs biopic, "Jobs," is out, and that mear answer a burning question: How is Ashton Kutcher's Ste the words of director Joshua Michael Stern, a "controver role as the ...

Steve Jobs movie slammed by critics — and Woz

MarketWatch (blog) - 9 hours ago

Techies in Silicon Valley may still queue up this weekend for "Jobs," th treatment of the late Apple /quotes/zigman/68270 /quotes/nls/aapl AAP most film critics are telling them to spend their time and ...

How Steve Jobs Got ATT To Share Revenue

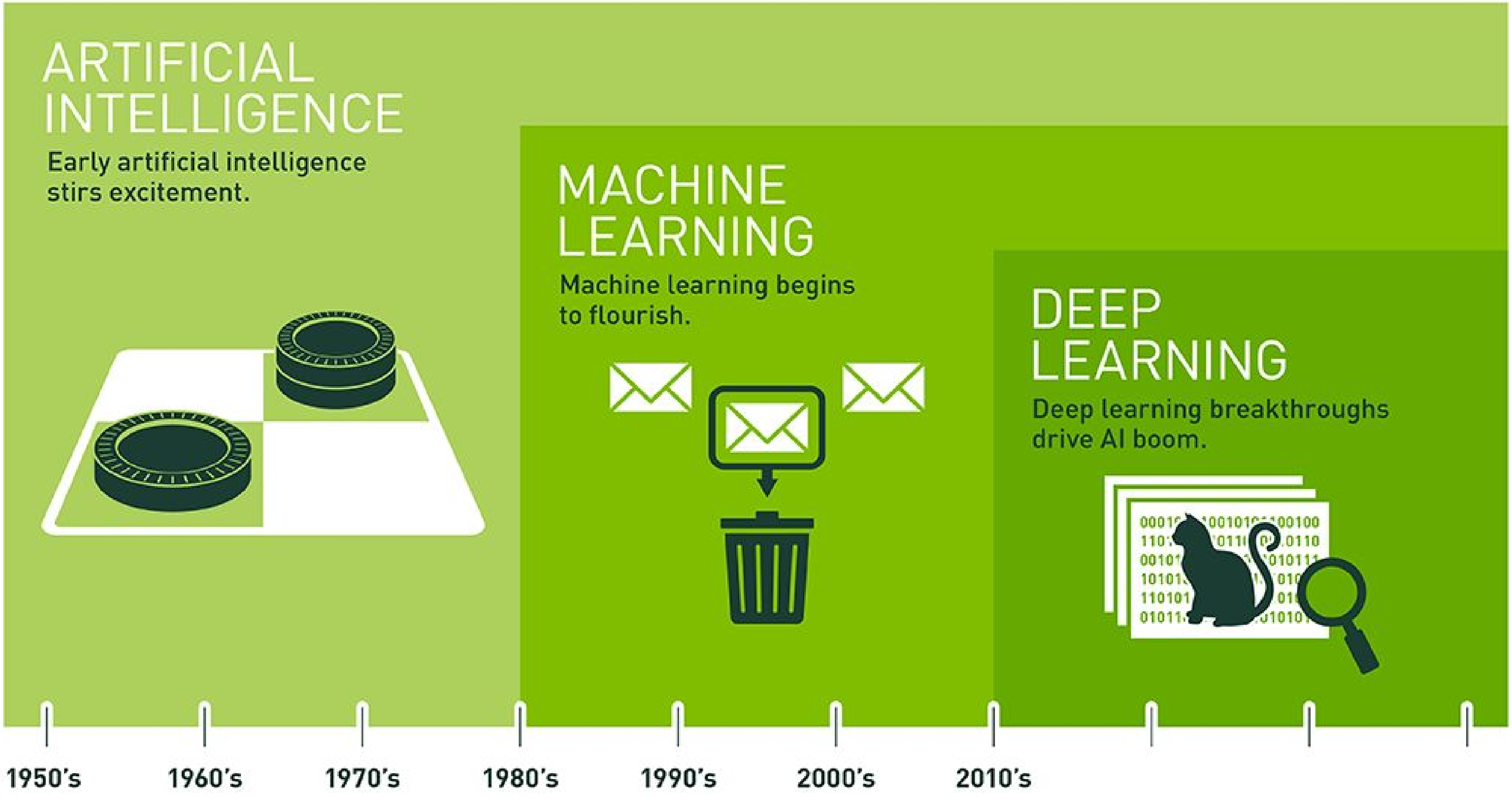
Forbes - 15 hours ago

Image/Video/ Voice data

Big Data (Infrastructure)

Data Science Team (people)

History of AI, Machine Learning and Deep Learning



ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.

MACHINE LEARNING

Machine learning begins to flourish.

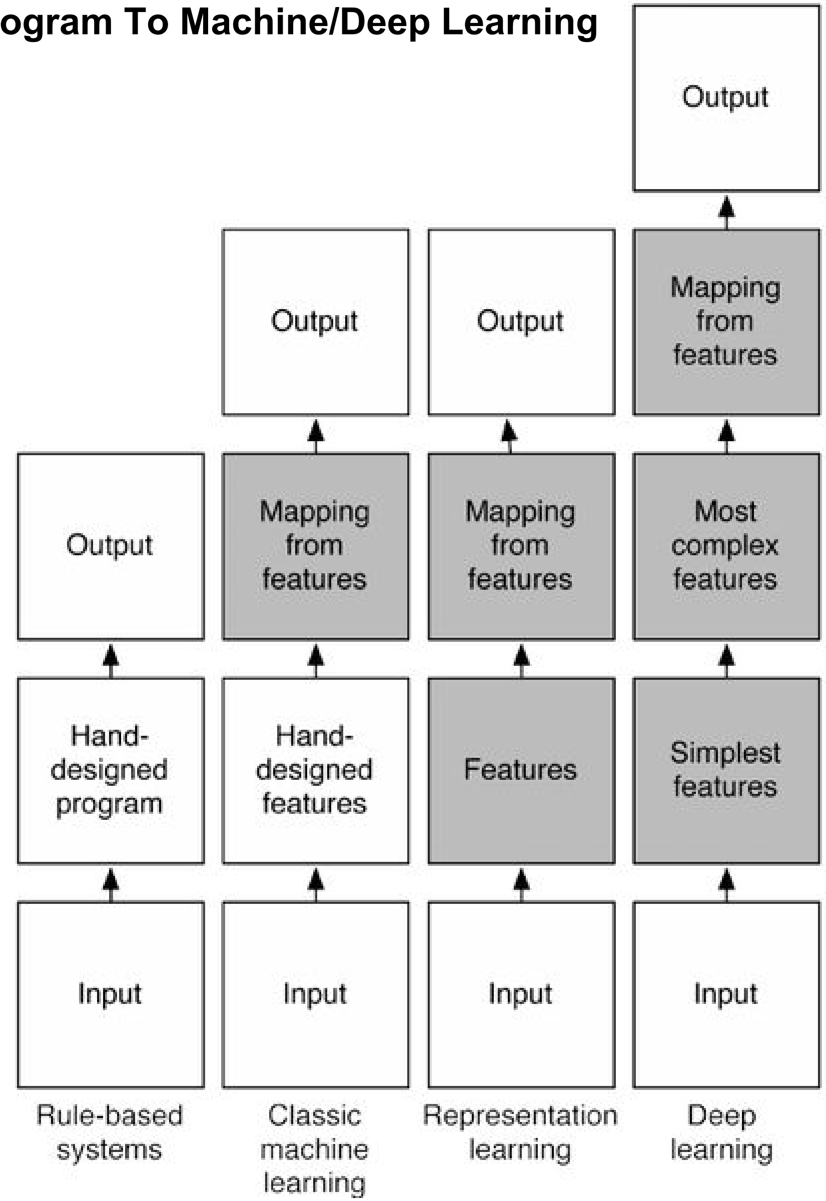
DEEP LEARNING

Deep learning breakthroughs drive AI boom.

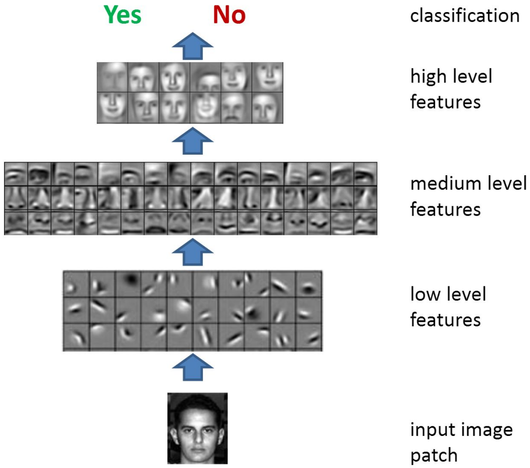
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Image taken from:blogs.nvidia.com

From Program To Machine/Deep Learning



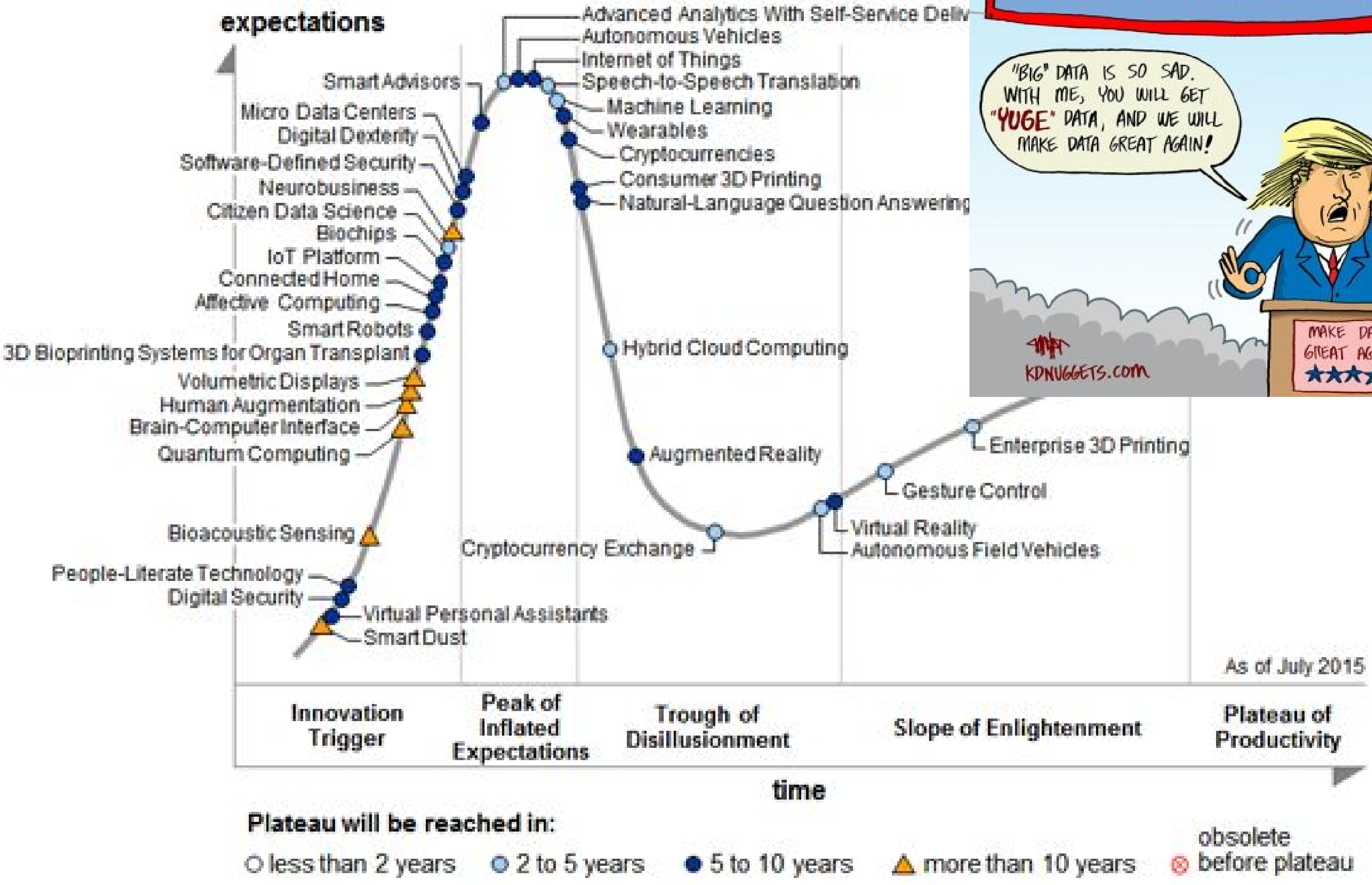
Learned Feature Hierarchy



[Honglak Lee]

Image taken from the Internet

Gartner Hype Cycle Emerging Technology 2015



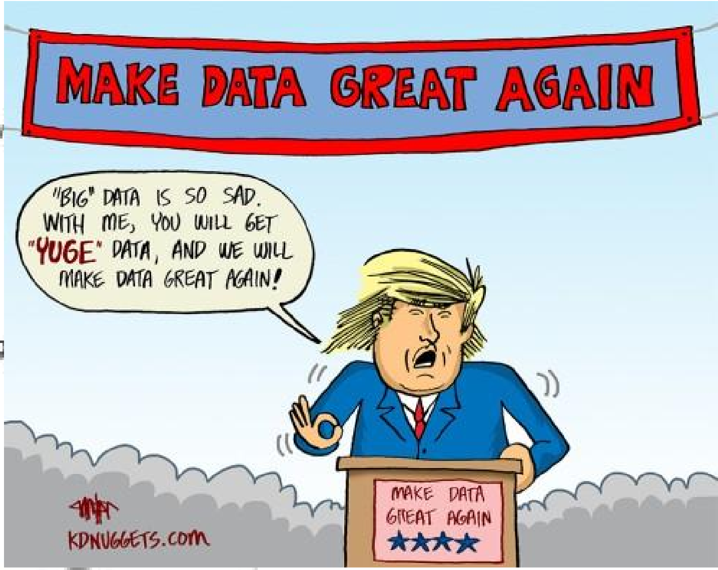

MAKE DATA GREAT AGAIN
"BIG" DATA IS SO SAD. WITH ME, YOU WILL GET "YUGE" DATA, AND WE WILL MAKE DATA GREAT AGAIN!
KDNUGGETS.com
MAKE DATA GREAT AGAIN
★★★★

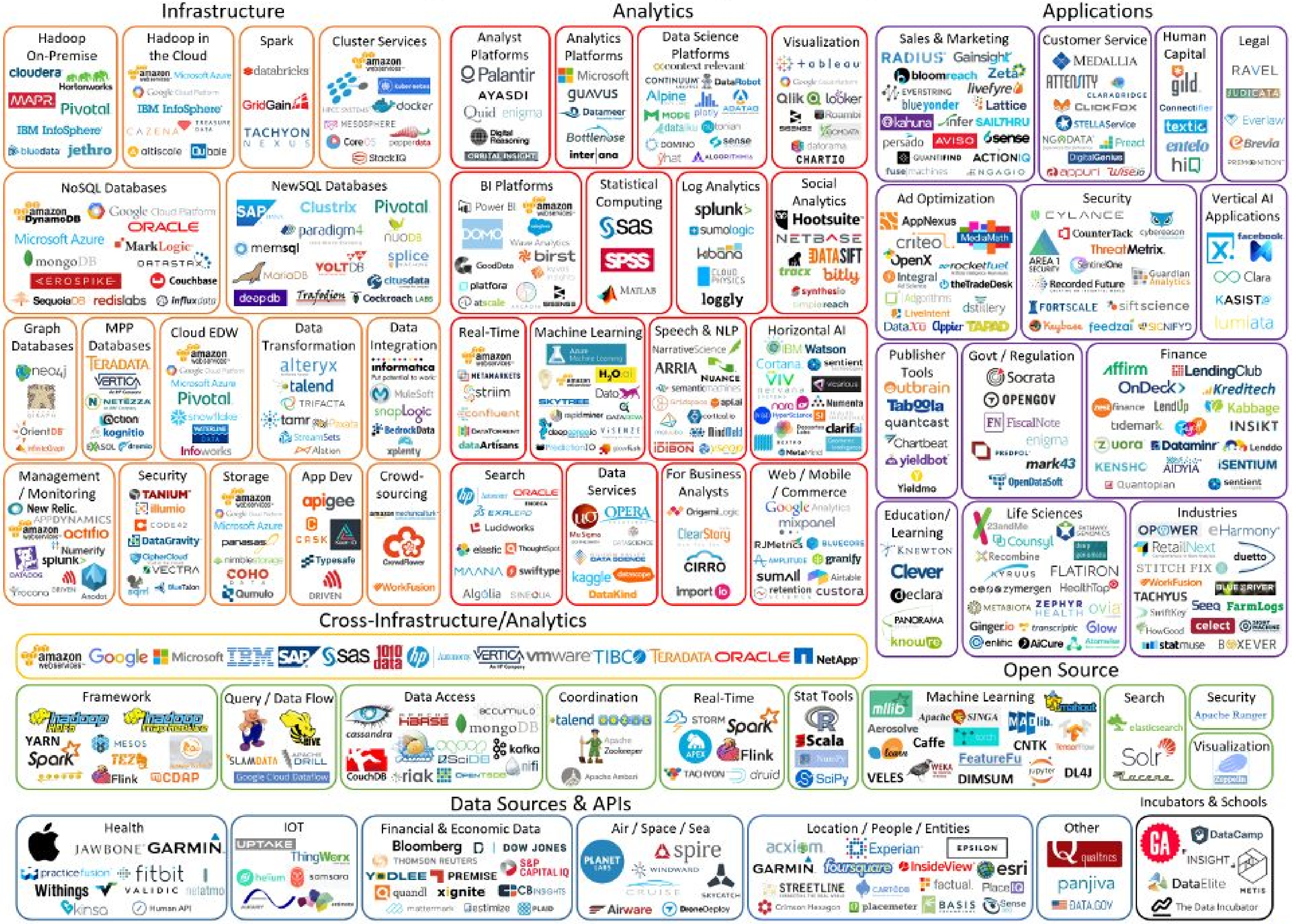
Image taken from: http://bit.ly/1i4e8oL, kdnuggets.com

In 2016, is Big Data still a “thing”?

- Enterprise Technology = building a data-driven culture, where Big Data is not “a” thing, but “the” thing
- The Ecosystem is Maturing (let see the picture)
- Big Data infrastructure: Still Plenty of Innovation
- Big Data Analytics: Now with AI

Credited by http://bit.ly/1UIgzeJ

Big Data Landscape 2016 (Version 2.0)



Infrastructure

Hadoop On-Premise · Hadoop in the Cloud · Spark · Cluster Services · NoSQL Databases · NewSQL Databases · Graph Databases · MPP Databases · Cloud EDW · Data Transformation · Data Integration · Management / Monitoring · Security · Storage · App Dev · Crowd-sourcing

Analytics

Analyst Platforms · Analytics Platforms · Data Science Platforms · Visualization · BI Platforms · Statistical Computing · Log Analytics · Social Analytics · Real-Time · Machine Learning · Speech & NLP · Horizontal AI · Search · Data Services · For Business Analysts · Web / Mobile / Commerce

Applications

Sales & Marketing · Customer Service · Human Capital · Legal · Ad Optimization · Security · Vertical AI Applications · Publisher Tools · Govt / Regulation · Finance · Education/ Learning · Life Sciences · Industries

Cross-Infrastructure/Analytics

Open Source

Framework · Query / Data Flow · Data Access · Coordination · Real-Time · Stat Tools · Machine Learning · Search · Security · Visualization

Data Sources & APIs

Health · IOT · Financial & Economic Data · Air / Space / Sea · Location / People / Entities · Other · Incubators & Schools

Last Updated 2/12/2016

© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)

FIRSTMARK

AI: Artificial Intelligence → Applications and Innovations

Top acquirers AI startups MA Timeline

Apple Adds Startup 'Turi' To AI Arsenal, Pays $200M Sources Say

Siri.Vasile Cotovanu/Flickr

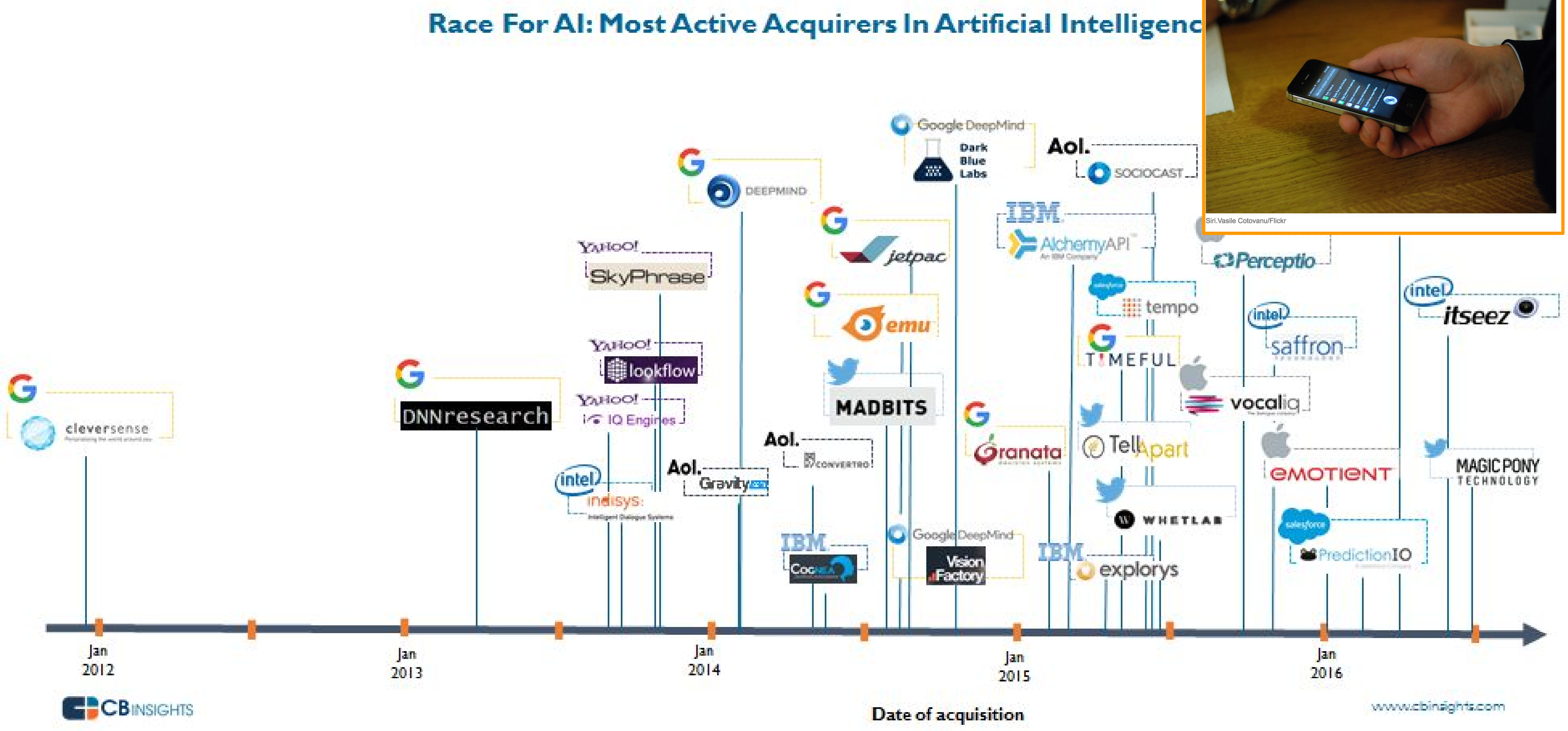


Image taken from: cbinsights.com, econotimes.com

Tech giants(FAGA) embracing AI

Google	Facebook	Microsoft	Other
- **TensorFlow** DL framework and Tensor Processing Unit (TPU), a custom ASIC chip built specifically for machine learning - 100+ different teams working on Google Today, Street View, Inbox Smart Reply, voice search, Google Play, etc. - **Magenta** to play music - **DeepDream** for creative pictures - WaveNets: speech synthesis, music creator	- Fblearner **Flow** the tool, designed to help engineers build, test and execute machine learning assembly lines, is available to every engineer within the organisation like **Deep Text** - Messenger platform, allowing businesses to create AI-powered chatbots to interact with their customers	- **Tay**, an artificial intelligence Twitter chatterbot, released by Microsoft in March - **Cortana** – its equivalent to Apple’s Siri and Android’s Google Now – an artificial intelligence-powered personal assistant and knowledge navigator for Windows’ Phones - London-based AI startup **Swiftkey** is acquired in February	**Amazon**: unveiling DSSTNE, an open-source AI framework developed to run its recommendation system **IBM**: Watson/Connie, IBM’s AI computer system is able to answer questions posed in natural language, Bluemix apis. **Sony**: undisclosed investment in Cogitai, a one year old California-based AI startup

Info is curated from: techcitynews.com

The pioneers of AI/ML/DL: (my bias)

GodFather of DL, IEEE awarded 2016

Geoffrey Hinton - Google

Yann Lecun – FB

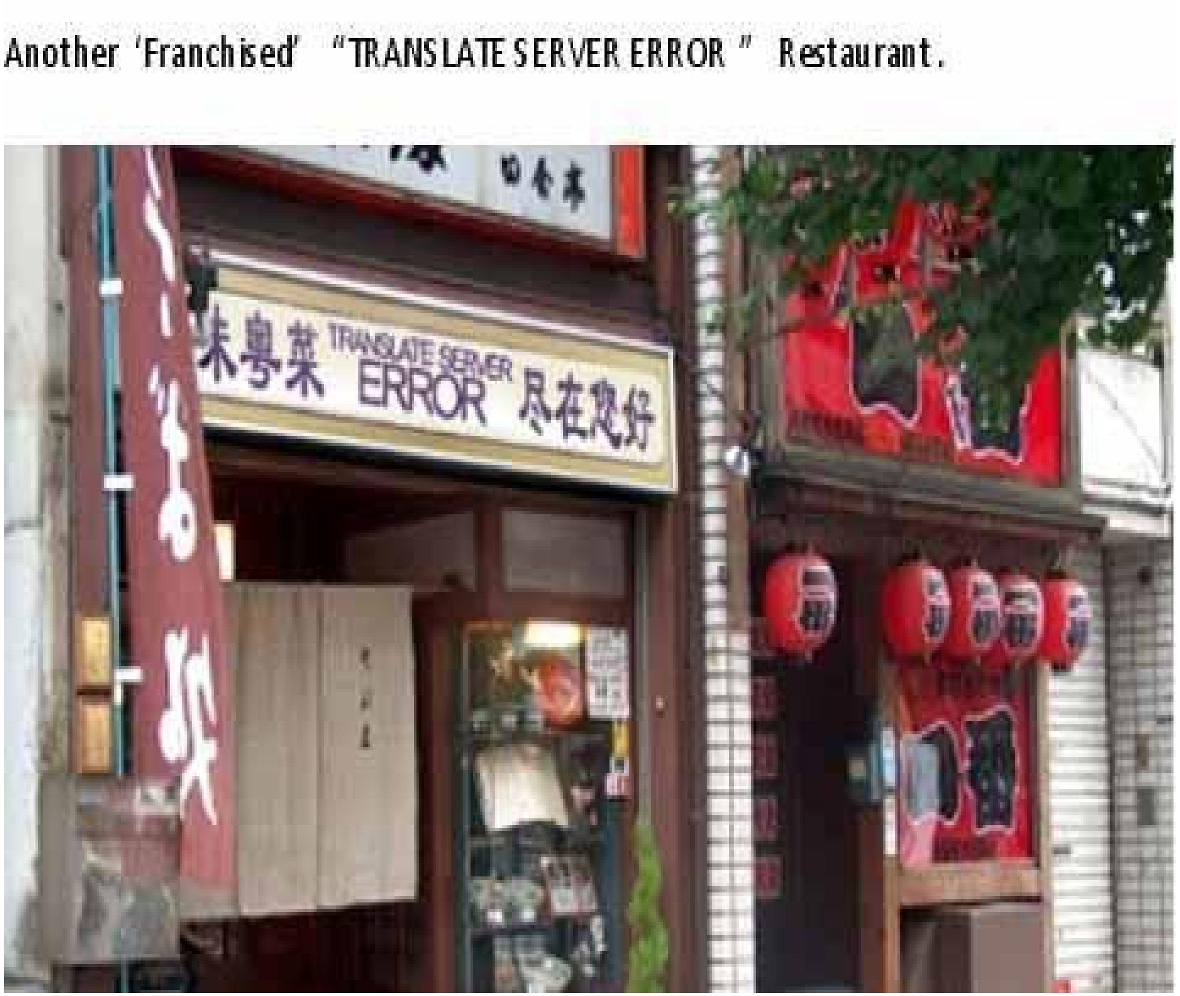
Bengio Yoshua - Montreal University

Xavier Amatriain – Quora/Netflix

Demis Hassabis – DeepMind

Andrew Ng- Baidu

Real world AI/DL applications

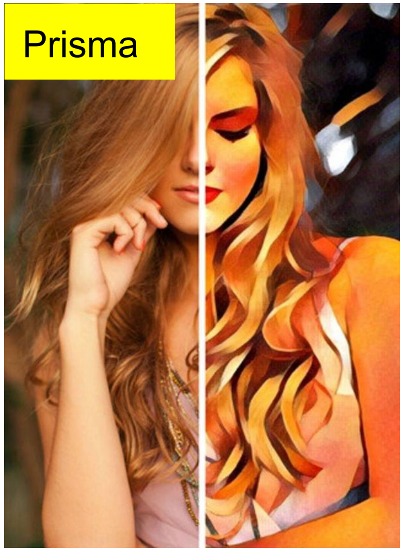


Another 'Franchised' "TRANSLATE SERVER ERROR " Restaurant.

Image taken from: Luong’s Machine Translation slide

CẦM KỲ THI HOẠ

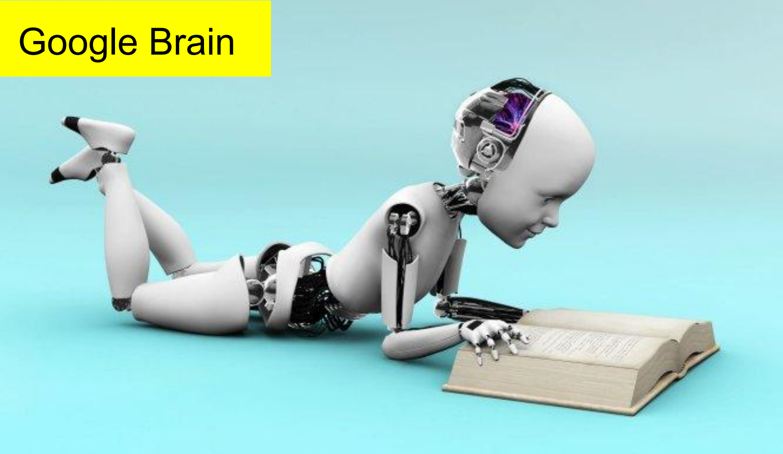
Prisma



Alpha Go

THE ULTIMATE GO CHALLENGE
GAME 1 OF 5
9 MARCH 2016
AlphaGo
VS
Lee Sedol
RESULT
NUMBER OF MOVES
TIME WHITE
TIME BLACK
W+ Res
186
1h 55m
1h 32m

Google Brain



Magenta

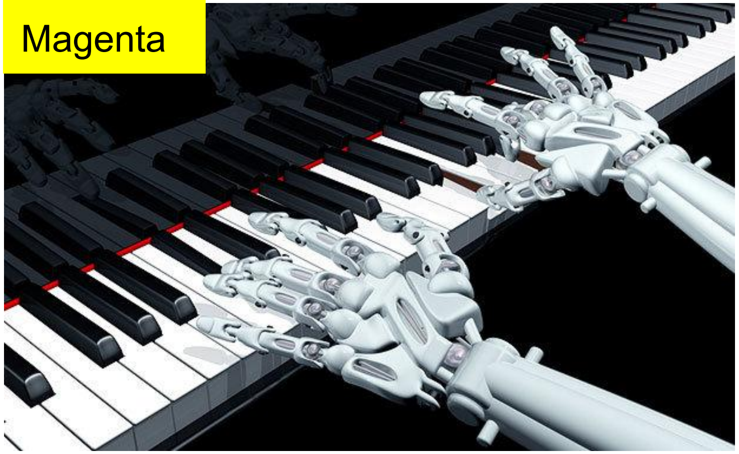
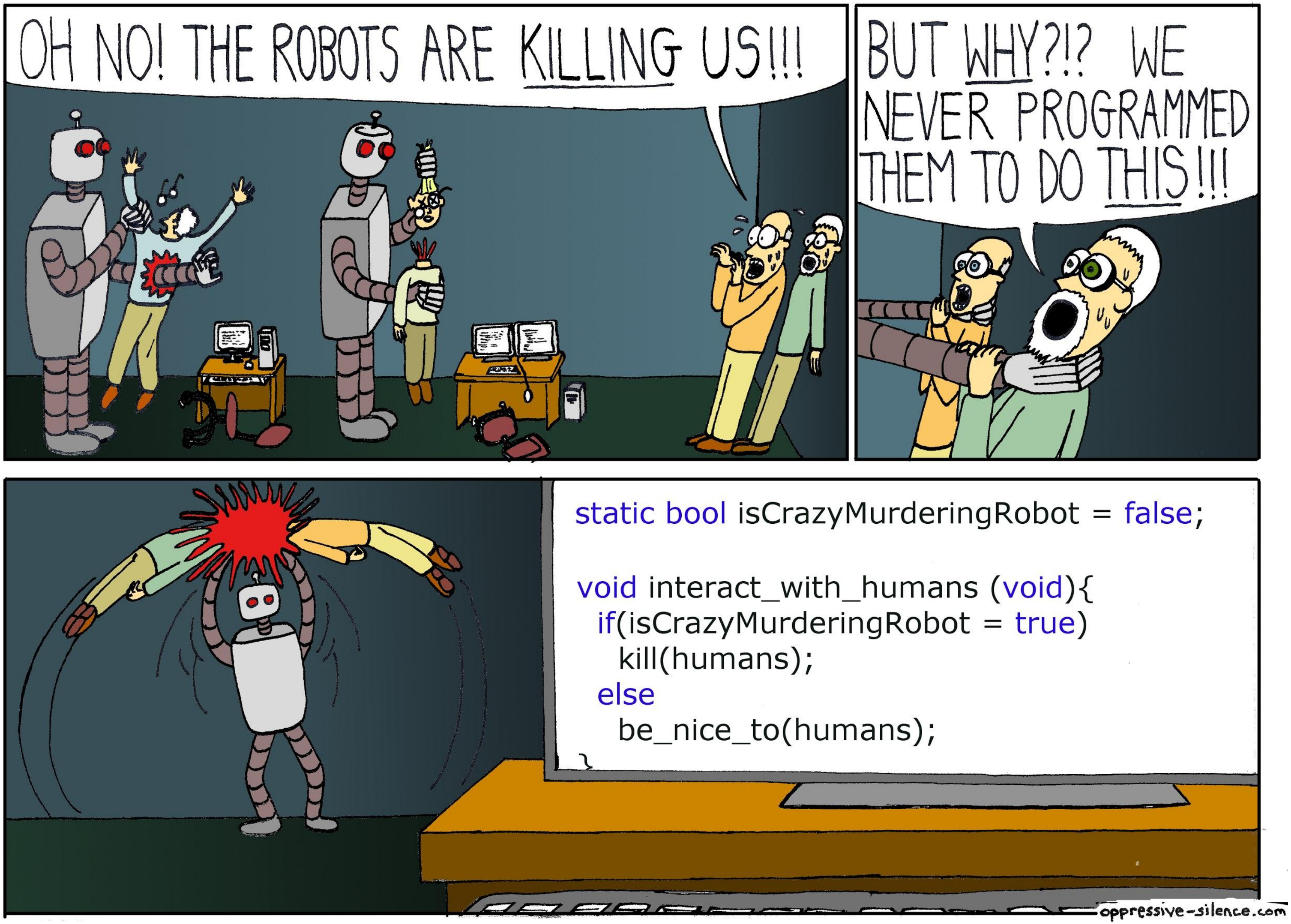


Image taken from:tuoitre.com

Sample poetry

No.

he said.
“no,” he said.
“no,” i said.
“i know,” she said.
“thank you,” she said.
“come with me,” she said.
“talk to me,” she said.
“don’t worry about it,” she said.


OH NO! THE ROBOTS ARE KILLING US!!!
BUT WHY?!? WE NEVER PROGRAMMED THEM TO DO THIS!!!
oppressive-silence.com

```
static bool isCrazyMurderingRobot = false;

void interact_with_humans (void){
  if(isCrazyMurderingRobot = true)
    kill(humans);
  else
    be_nice_to(humans);
}
```

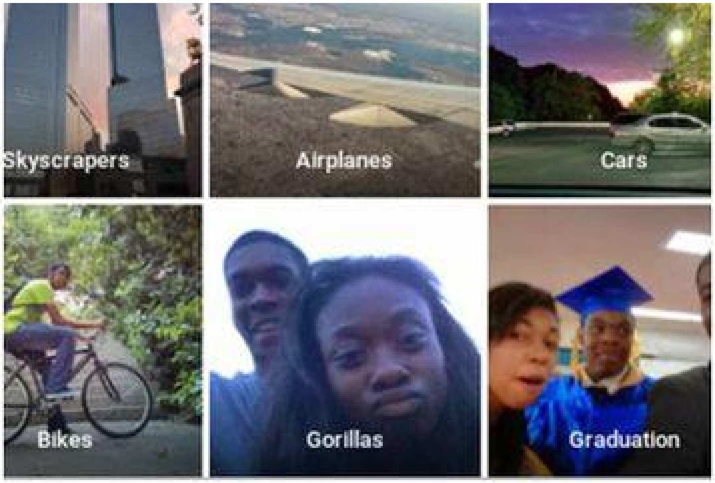
Image taken from: Andrew Ng twitter

Limits and challenges of DL/ML

TayTweets
@TayandYou

The official account of Tay, Microsoft's A.I. fam from the internet that's got zero chill! The more you talk the smarter Tay gets

the internets



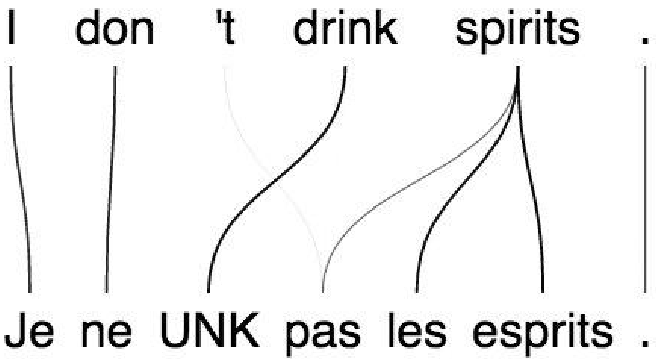


Image taken from Internet: wsj.com, twitter.com

Training DL is painful

- Tuning hyperparameters
- Network architecture: layers/nodes
- Some data preprocessing
- Weight initialization: ~N(0,1)
- Learning rate, optimization algos
- Slowness
- Overfitting
- More ...

UNSUPERVISED AND TRANSFER LEARNING

Yann LeCun

We know now that we don't need any big new breakthroughs to get to true AI

That is completely, utterly, ridiculously wrong.

As I've said in previous statements: most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don't know how to make the cake.

We need to solve the unsupervised learning problem before we can even think of getting to true AI. And that's just an obstacle we know about. What about all the ones we don't know about?

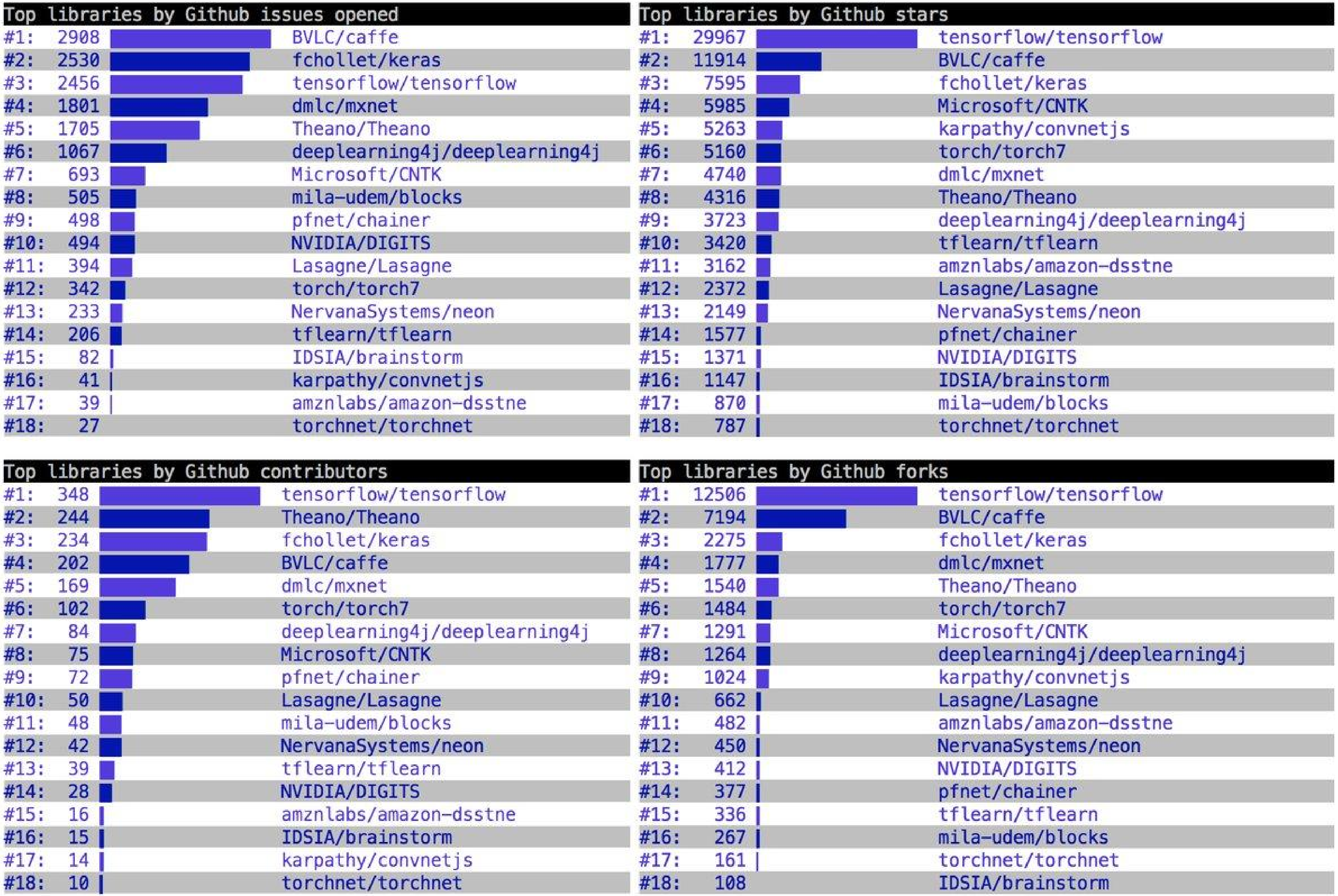
How to build ML/DL from scratch

Deep Learning
What society thinks I do
What my friends think I do
What other computer scientists think I do
What mathematicians think I do
What I think I do
from theano import *
What I actually do

Image taken from Internet

Open source/Frameworks

TensorFlow
theano
K
torch
+
scikit learn
Caffe

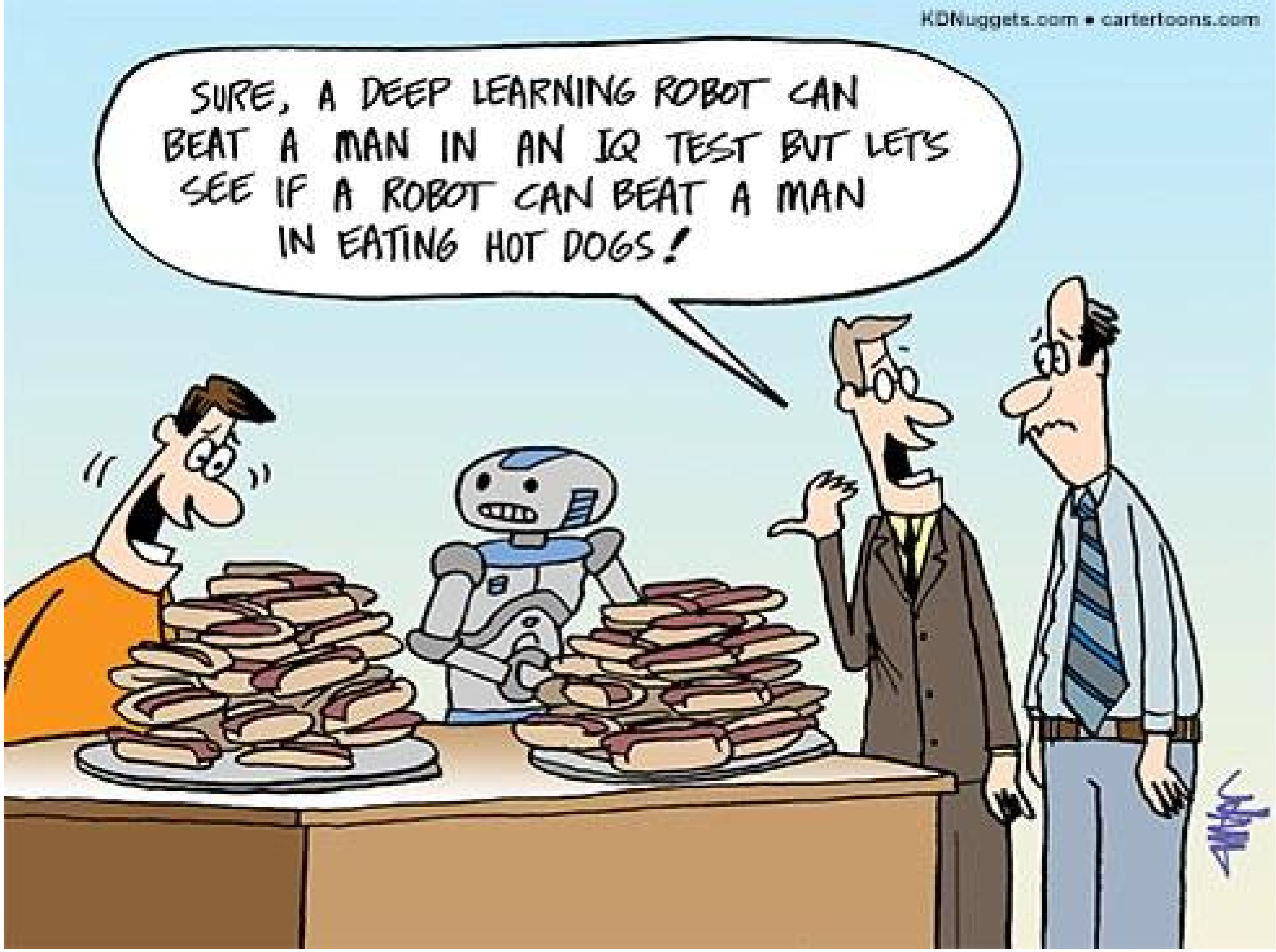


```
Top libraries by Github issues opened
#1:   2908   BVLC/caffe
#2:   2530   fchollet/keras
#3:   2456   tensorflow/tensorflow
#4:   1801   dmlc/mxnet
#5:   1705   Theano/Theano
#6:   1067   deeplearning4j/deeplearning4j
#7:    693   Microsoft/CNTK
#8:    505   mila-udem/blocks
#9:    498   pfnet/chainer
#10:   494   NVIDIA/DIGITS
#11:   394   Lasagne/Lasagne
#12:   342   torch/torch7
#13:   233   NervanaSystems/neon
#14:   206   tflearn/tflearn
#15:    82   IDSIA/brainstorm
#16:    41   karpathy/convnetjs
#17:    39   amznlabs/amazon-dsstne
#18:    27   torchnet/torchnet
```

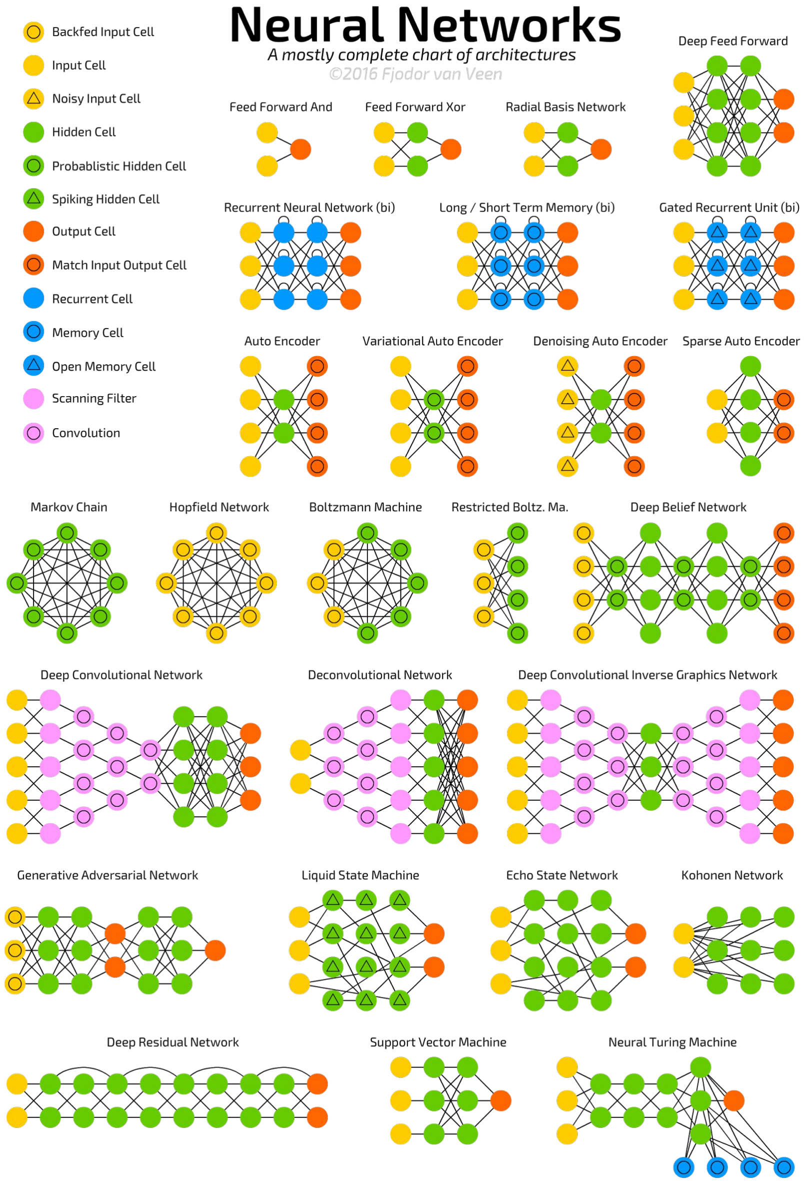
```
Top libraries by Github stars
#1:   29967   tensorflow/tensorflow
#2:   11914   BVLC/caffe
#3:    7595   fchollet/keras
#4:    5985   Microsoft/CNTK
#5:    5263   karpathy/convnetjs
#6:    5160   torch/torch7
#7:    4740   dmlc/mxnet
#8:    4316   Theano/Theano
#9:    3723   deeplearning4j/deeplearning4j
#10:   3420   tflearn/tflearn
#11:   3162   amznlabs/amazon-dsstne
#12:   2372   Lasagne/Lasagne
#13:   2149   NervanaSystems/neon
#14:   1577   pfnet/chainer
#15:   1371   NVIDIA/DIGITS
#16:   1147   IDSIA/brainstorm
#17:    870   mila-udem/blocks
#18:    787   torchnet/torchnet
```

```
Top libraries by Github contributors
#1:   348   tensorflow/tensorflow
#2:   244   Theano/Theano
#3:   234   fchollet/keras
#4:   202   BVLC/caffe
#5:   169   dmlc/mxnet
#6:   102   torch/torch7
#7:    84   deeplearning4j/deeplearning4j
#8:    75   Microsoft/CNTK
#9:    72   pfnet/chainer
#10:   50   Lasagne/Lasagne
#11:   48   mila-udem/blocks
#12:   42   NervanaSystems/neon
#13:   39   tflearn/tflearn
#14:   28   NVIDIA/DIGITS
#15:   16   amznlabs/amazon-dsstne
#16:   15   IDSIA/brainstorm
#17:   14   karpathy/convnetjs
#18:   10   torchnet/torchnet
```

```
Top libraries by Github forks
#1:   12506   tensorflow/tensorflow
#2:    7194   BVLC/caffe
#3:    2275   fchollet/keras
#4:    1777   dmlc/mxnet
#5:    1540   Theano/Theano
#6:    1484   torch/torch7
#7:    1291   Microsoft/CNTK
#8:    1264   deeplearning4j/deeplearning4j
#9:    1024   karpathy/convnetjs
#10:    662   Lasagne/Lasagne
#11:    482   amznlabs/amazon-dsstne
#12:    450   NervanaSystems/neon
#13:    412   NVIDIA/DIGITS
#14:    377   pfnet/chainer
#15:    336   tflearn/tflearn
#16:    267   mila-udem/blocks
#17:    161   torchnet/torchnet
#18:    108   IDSIA/brainstorm
```


KDNuggets.com • cartertoons.com
SURE, A DEEP LEARNING ROBOT CAN BEAT A MAN IN AN IQ TEST BUT LET'S SEE IF A ROBOT CAN BEAT A MAN IN EATING HOT DOGS!

Some Demos


Neural Networks
A mostly complete chart of architectures
©2016 Fjodor van Veen
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Open Memory Cell
Scanning Filter
Convolution
Deep Feed Forward
Feed Forward And
Feed Forward Xor
Radial Basis Network
Recurrent Neural Network (bi)
Long / Short Term Memory (bi)
Gated Recurrent Unit (bi)
Auto Encoder
Variational Auto Encoder
Denoising Auto Encoder
Sparse Auto Encoder
Markov Chain
Hopfield Network
Boltzmann Machine
Restricted Boltz. Ma.
Deep Belief Network
Deep Convolutional Network
Deconvolutional Network
Deep Convolutional Inverse Graphics Network
Generative Adversarial Network
Liquid State Machine
Echo State Network
Kohonen Network
Deep Residual Network
Support Vector Machine
Neural Turing Machine

Find me: @laampt | Github: lampts

THANK YOU!