

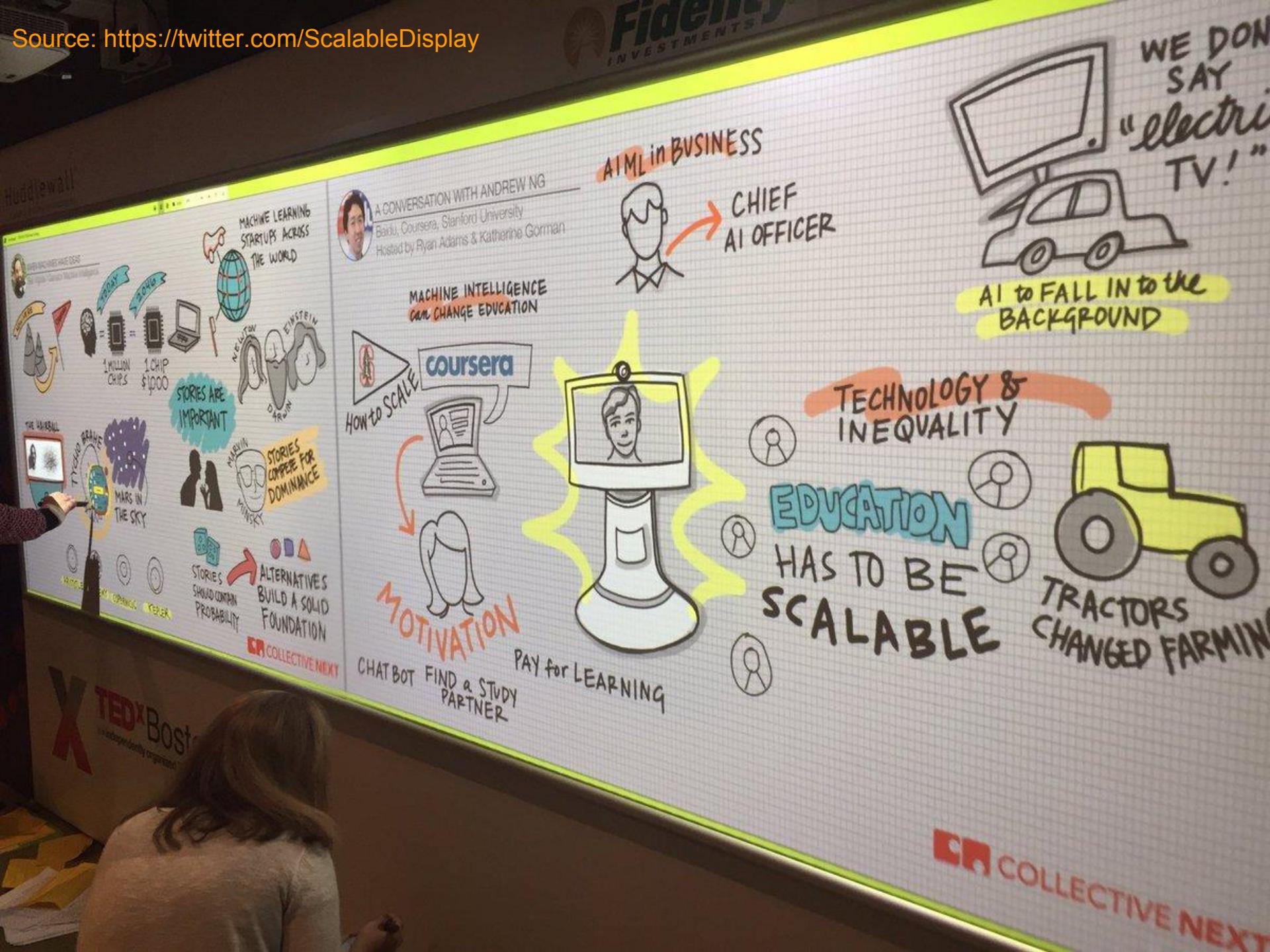


(Big) Data, (Deep) Learning and AI

When big data hits machine learning

Phạm Thành Lâm | Founder @ SaigonApps
HCMC 01.10.2016

Source: <https://twitter.com/ScalableDisplay>



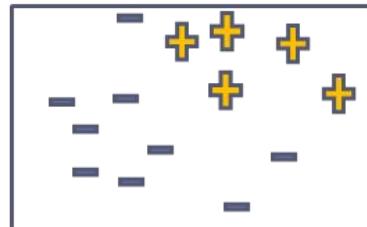
Big Picture: Big Data - Machine Learning/ Data Mining

Machine Learning

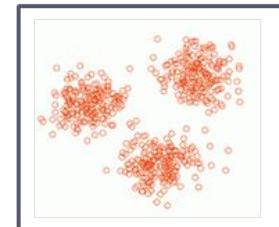
Frequent pattern

TID	Items	TV
1	A, B, C, D, E	2310
2	A, B, C, D, F	2750
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	78
8	C, D, F	455

Supervised



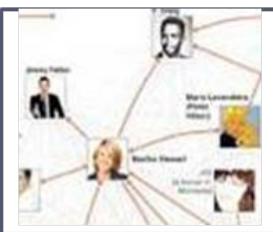
Unsupervised



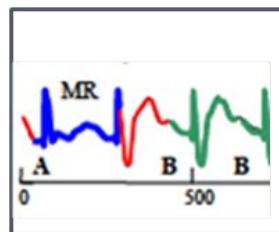
Recommendation



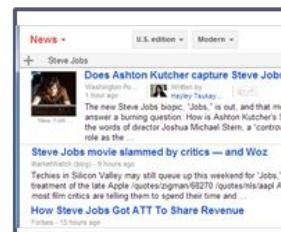
Graphs



Stream Data



Text Data



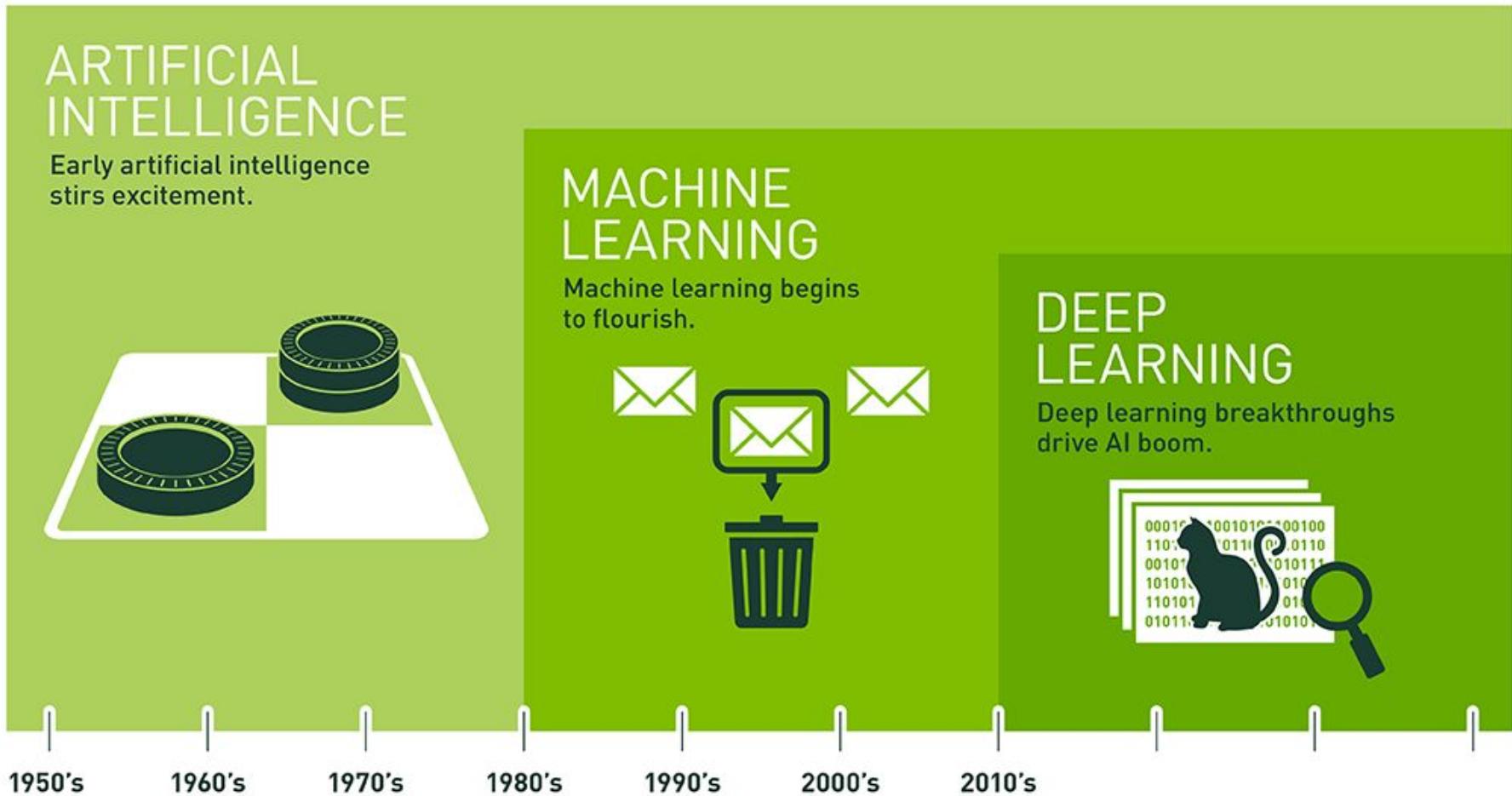
Image/Video/ Voice data



Big Data (Infrastructure)

Data Science Team (people)

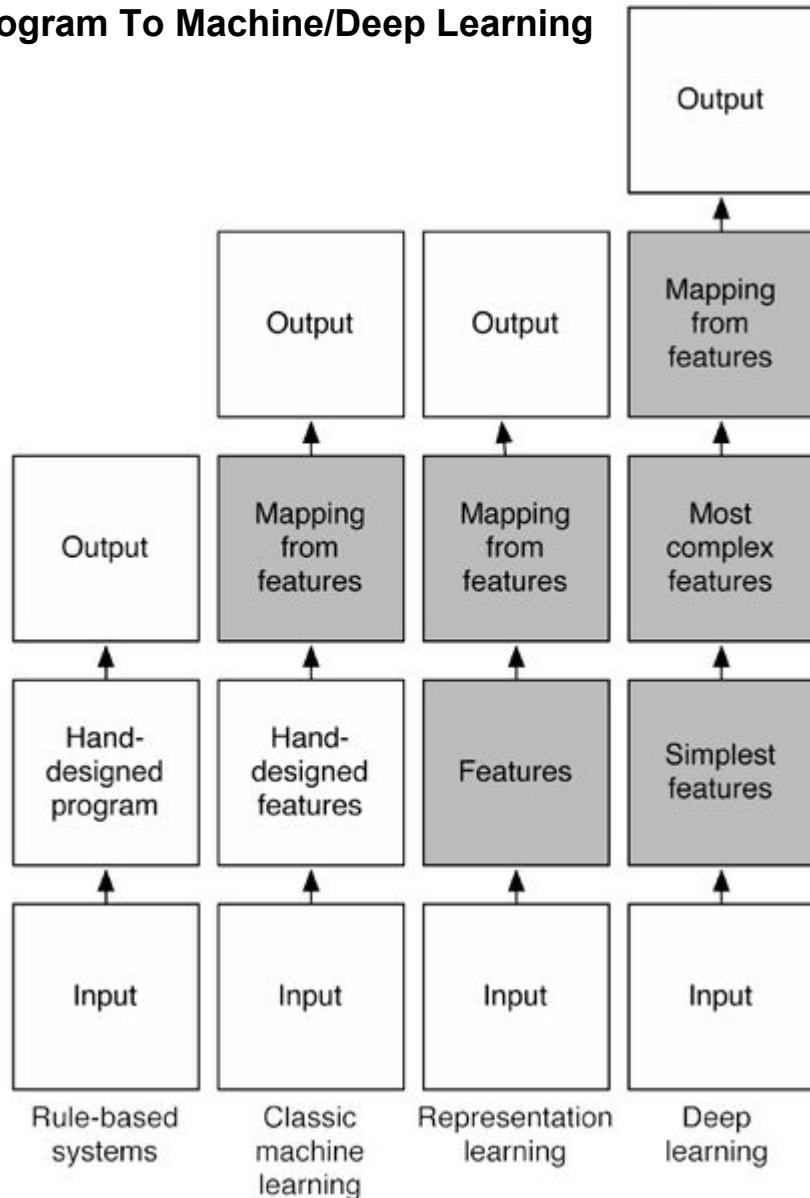
History of AI, Machine Learning and Deep Learning



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

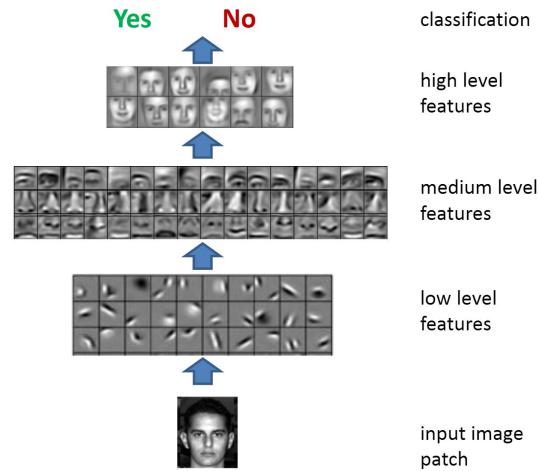


Image taken from the Internet

Gartner Hype Cycle Emerging Technology 2015

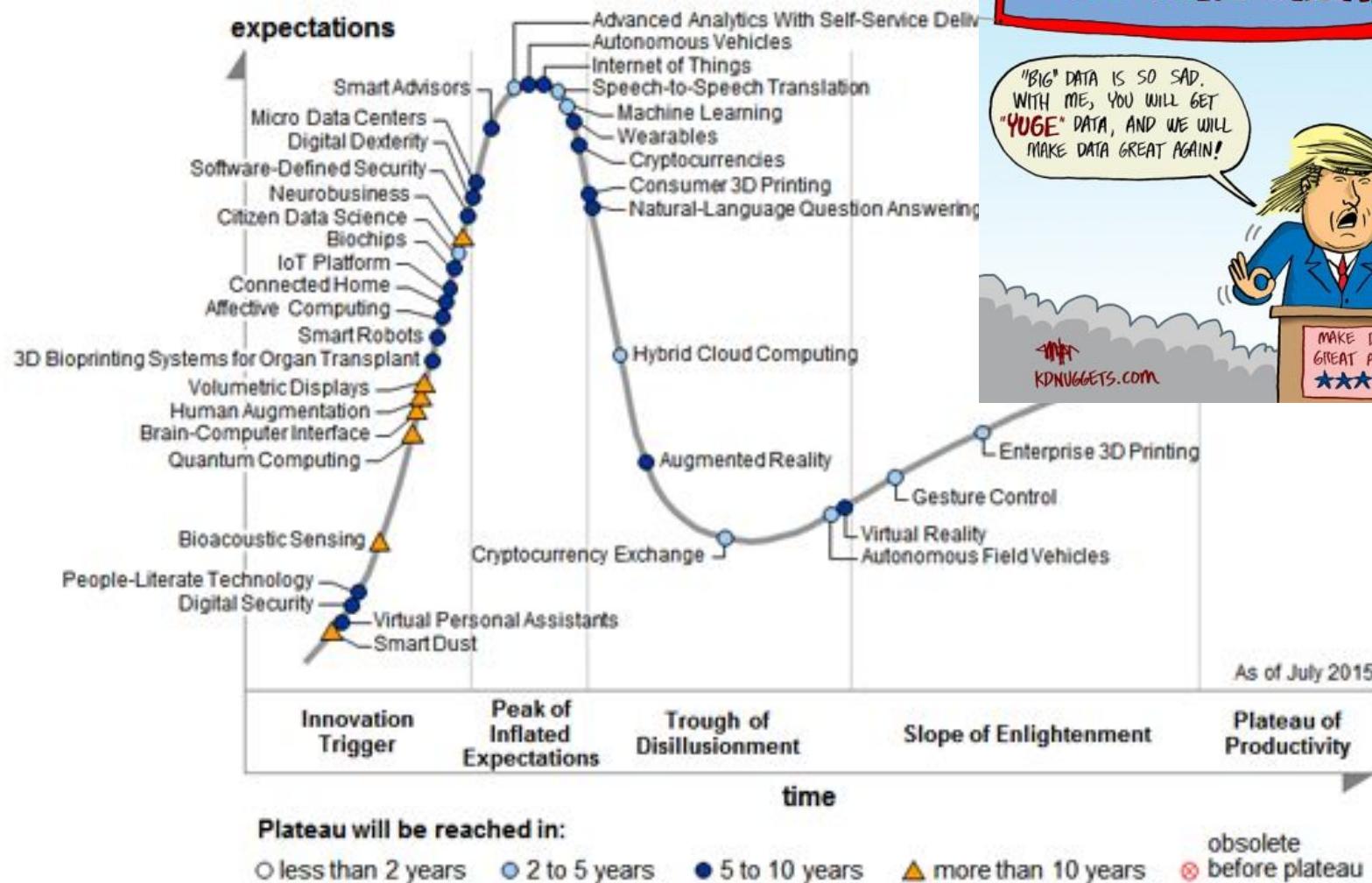


Image taken from: <http://bit.ly/1i4e8oL>, kdnuggets.com

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/1UlgzeJ>

MACHINE INTELLIGENCE 2.0

AGENTS

PROFESSIONAL

Handy! @clara
KASIST DigitalGenius
OVERLAP.CC meekan
fusemachines PRIMER

PERSONAL

facebook XIAOICE
assistant.ai large
nestor @awesome
Magic

OS INTERFACES

Siri Cortana VIV
 maluuba api.ai
CoNea Google Now

AUTONOMOUS SYSTEMS

AIR

3DR DJI PROJECT LOON Google
 VERTICAL DroneDeploy
 AIRDODGE SKYCATCH
 SKYDIO Airware LILY

GROUND

Google UBER
 TESLA CRUISE
 MOBILEYE
 COMMA AdasWorks

SEA

bluefin data
 OPENROV
 BluHaptics

INDUSTRIAL

KIVA Systems fetch
 HARVEST CLEARPATH
 AVIBOTS ENERGID
rethink robotics GREY ORANGE
OSARO

ENTERPRISE

SECURITY / FRAUD

Sentinel graphistry
 BITSIGHT feedzai AREA1
drawbridge siftscience CYLANCE Brighterion

HR / RECRUITING

textio hiQ gild
 SpringRole entelo
 unilive GIGSTER

SALES

osense clarifai infer
 Preact Gainsight AVISO
 salespredict sentient
 Vidora people pattern Prism

MARKETING

LiftIgniter
 RADIUS brightfunnel
 retention SCIENCE AIRPR

CUSTOMER SUPPORT

CLARABRIDGE
 QUANTIFIND wiseio
 ACTIONIQ FRAMED
DigitalGenius

INTERNAL INTEL

Alation ADATAD
 Palantir Osapho lucid
 Rainbird SKIPFLAG agologo
 Digital Reasoning Narrative Science

MARKET INTEL

Quid mattermark
 DataFox bottleneck
 PREMISE enigma
 CB INSIGHTS

PLATFORMS

RESEARCH / AGI

OpenAI vicarious
 Google DeepMind Numenta
 Cycorp naisense
 SCALED INFERENCE GEOMETRIC INTELLIGENCE

FULL STACK

context relevant
 CognitiveScale
 NVIDIA TERADEEP
 QUALCOMM nervana SYSTEMS

MACHINE LEARNING

Dato rapidminer
 cortical.io AYASDI
 amazon web services Azure
 nara logics PredictionIO
 SKYTREE bigml blueyonder

INDUSTRIAL IOT

ThingWorx UPTAKE
 IMUBIT Preferred Networks
 Alluvium xively
 PLANET OS

AUDIO

Gridspace TalkIQ
 nexidia vocaliq
 NUANCE Expect Labs
 popUP archive

VISION

ORBITAL INSIGHT
 Descartes Labs DEXTRON
 cortica clarifai PLANET LIMS
 MetaMind

DATA ENRICHMENT

diffbot Paxata
 TRIFACTA iDIBON
 WorkFusion loopj
 CrowdFlower

INDUSTRIES

ADTECH

ADTHEORENT dstillery
 BEYONDVERBAL
 METAMARKETS TAPAD
 rocketfuel affectiva

AGRICULTURE

BLUE RIVER tule
 TerrAvion mavrx
 THE CLIMATE CORPORATION OCERES TECHNOLOGIES
 HONEYCOMB

FOR GOOD

Conservation Metrics
 DataKind DATA FOR GOOD
 thorn BAYES IMPACT

RETAIL FINANCE

inVenture affirm
 earnest MIRADOR
 Lendo zest finance LendUp

LEGAL

Everlaw RAVEL
 LEGAL ROBOT seal
 BEAGLE ROSS
 Lex Machina

MATERIALS & MFG

zymergen AUGMATE
 GINKGO BIOWORKS
 SIGHT MACHINE TECHNOLOGIES CALCULARIO
 Eigen innovations

HEALTHCARE

deep genomics 3SCAN
 enlitic Calico BIMBLY
 Atomwise Recombine Color
 Metabiota Grand Rounds
 Google Life Sciences IBM Watson Health

INDUSTRIES (CONT'D)

EDUCATION

SKNEWTON
 coursera turnitin
 gradescope UDACITY
 KHANACADEMY

TRANSPORT & LOGISTICS

NAUTO taleris
 PRETECKT
 clearmetal

INVESTMENT FINANCE

Bloomberg Quantopian
 Dataminr KENSIC
 iSENTIUM NEURENSIC
 alphasense

TECH USER TOOLS

DATA SCIENCE

DOMINO kaggle
 Sentenai sense
 yseop Outlier
 ŷhat DataRobot

MACHINE LEARNING

Cortana Analytics AlchemyAPI glowfish
 IBM Watson Platform Anodot MonkeyLearn
 ([h s])HyperScience fuzzy.io SIGOPT
 Oxdata H2O SPARKBEYOND indico

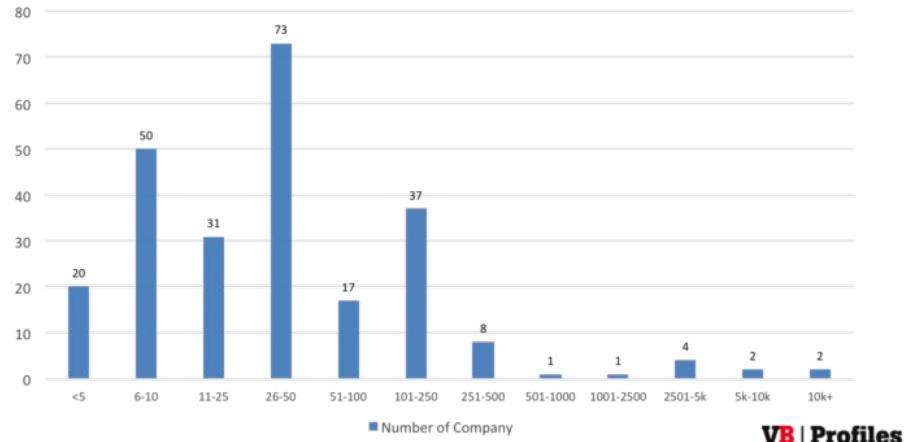
OPEN SOURCE

SKYMLND TensorFlow
 seldon Caffe theano
 Spark MLlib Microsoft DTk spaCy
 DL4J SciKit CGT

VentureBeat reported

Over half of the companies have under 50 employees

- 247 companies
- 50,000 employees
- \$23 billion in funding
- 7 companies with valuations greater than \$1 billion
- A collective valuation of \$107 billion



<http://venturebeat.com/2016/08/16/machine-intelligence-2-0-in-charts-and-graphs>

AI: Artificial Intelligence → Applications and Innovations

Insights we trust.

Top acquirers AI startups MA Timeline

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

Race For AI: Most Active Acquirers In Artificial Intelligence

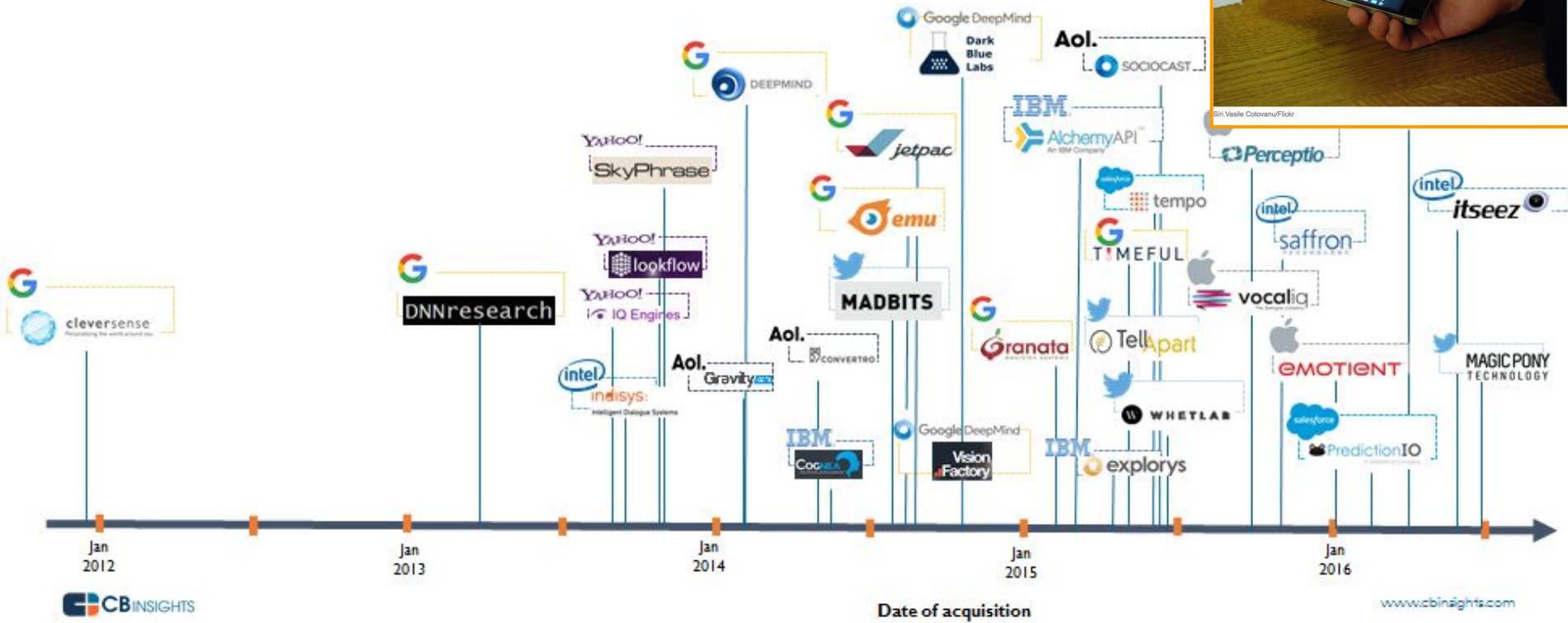


Image taken from: cbinsights.com, econotimes.com

Tech giants(FAGIM) embracing AI

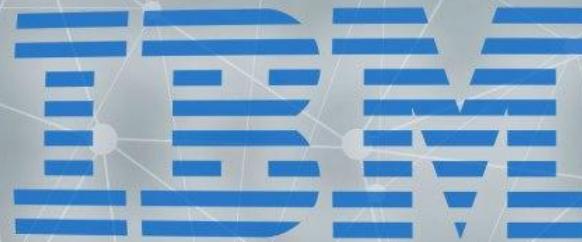
Google	Facebook	Microsoft	Other
<ul style="list-style-type: none">- TensorFlow DL framework and Tensor Processing Unit (TPU), a custom ASIC chip built specifically for machine learning- 100+ different teams working on Google Today, Street View, Inbox Smart Reply, voice search, Google Play, etc.- Magenta to play music- DeepDream for creative pictures- WaveNets: speech synthesis, music creator	<ul style="list-style-type: none">- 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	<ul style="list-style-type: none">- 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	<p>Amazon: unveiling DSSTNE, an open-source AI framework developed to run its recommendation system</p> <p>IBM: Watson/Connie, IBM's AI computer system is able to answer questions posed in natural language, Bluemix apis.</p> <p>Sony: undisclosed investment in Cogitai, a one year old California-based AI startup</p>

Info is curated from: techcitynews.com

The Partnership on AI



Google



Microsoft

Source: Techcrunch

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: The Verge, Luong's Machine Translation slide

CÀM KỲ THI HOẠ

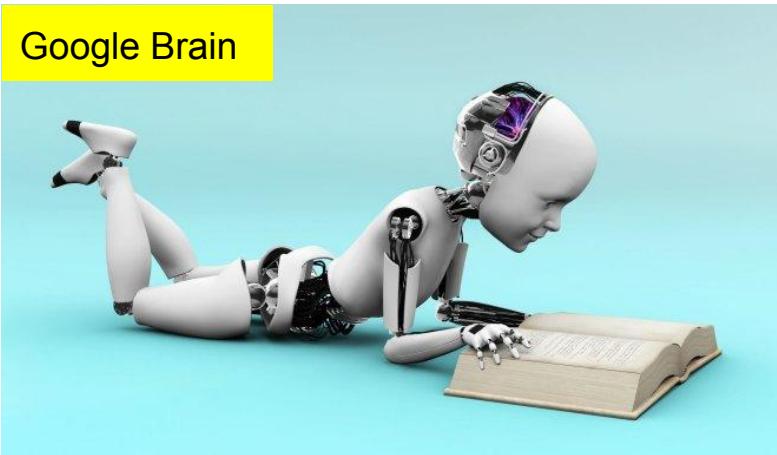
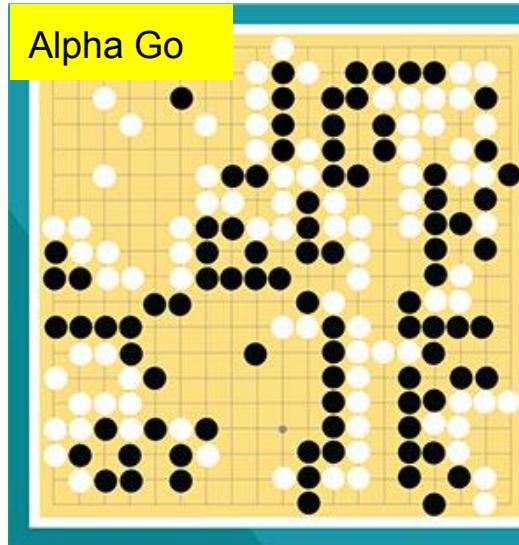


Image taken from:tuoitre.com

Sample paintings

A



B



C



D



More

DeepDrumpf
@DeepDrumpf

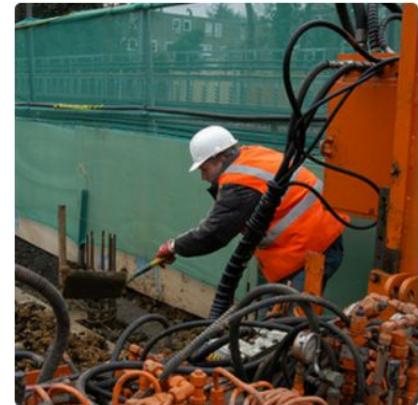
OK, it's amazing right now with ISIS, I tell you what? I don't want them to vote, the worst very social people. I love me.

RETWEETS 83 LIKES 75

1:49 PM - 3 Mar 2016



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."

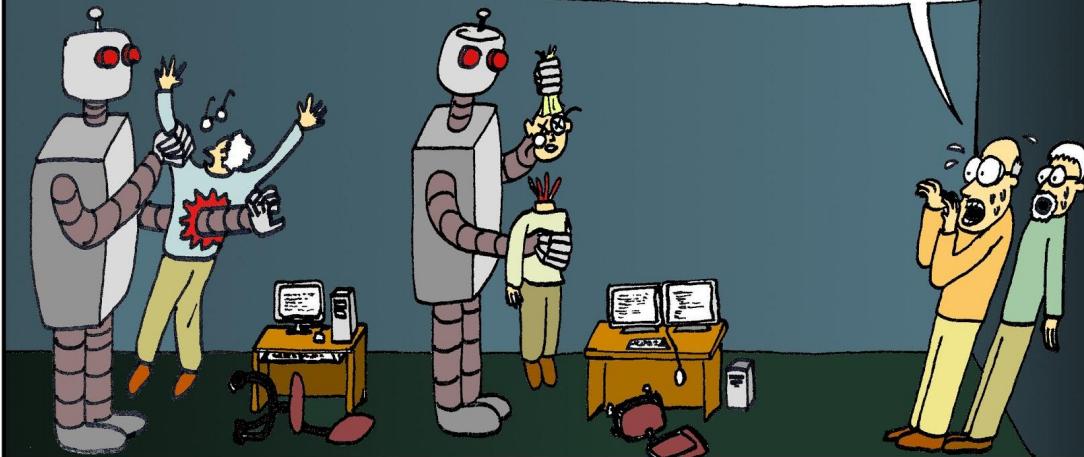


"girl in pink dress is jumping in air."

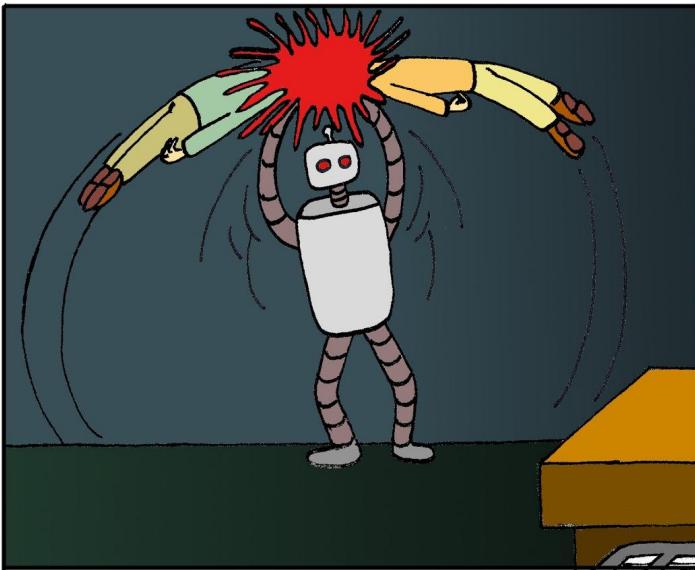


"black and white dog jumps over bar."

OH NO! THE ROBOTS ARE KILLING US!!!



BUT WHY?!? WE NEVER PROGRAMMED THEM TO DO THIS!!!



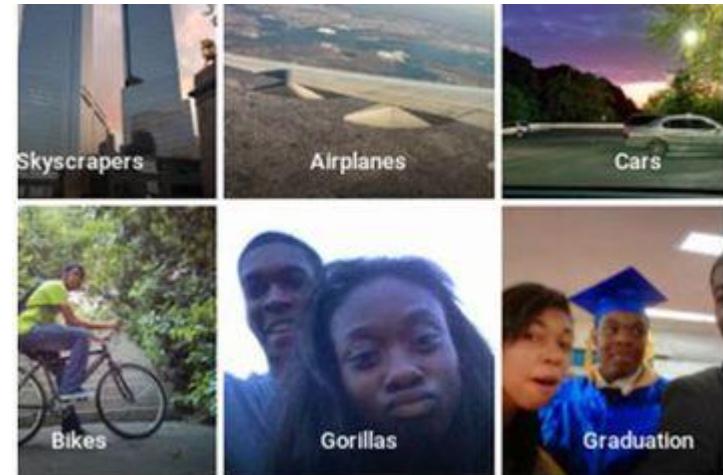
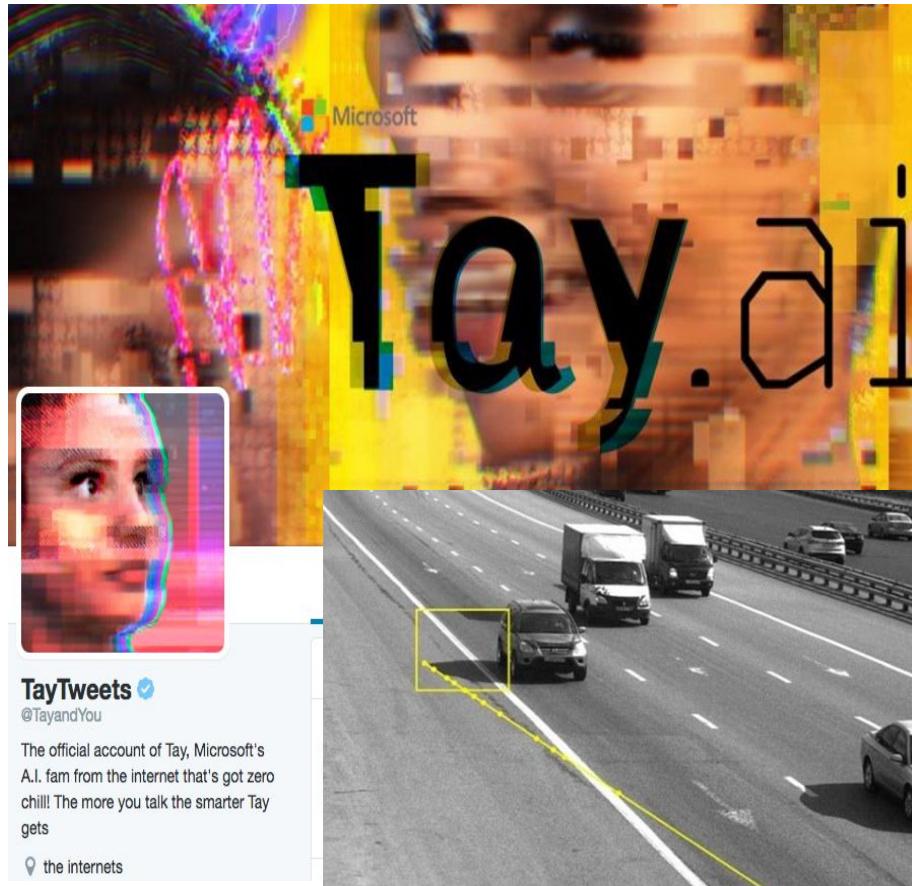
```
static bool isCrazyMurderingRobot = false;
```

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

oppressive-silence.com

Image taken from: Andrew Ng twitter

Limits and challenges of DL/ML



I don't drink spirits.
Je ne UNK pas les esprits .
Image taken from Internet: wsj.com, twitter.com

Training DL is painful

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

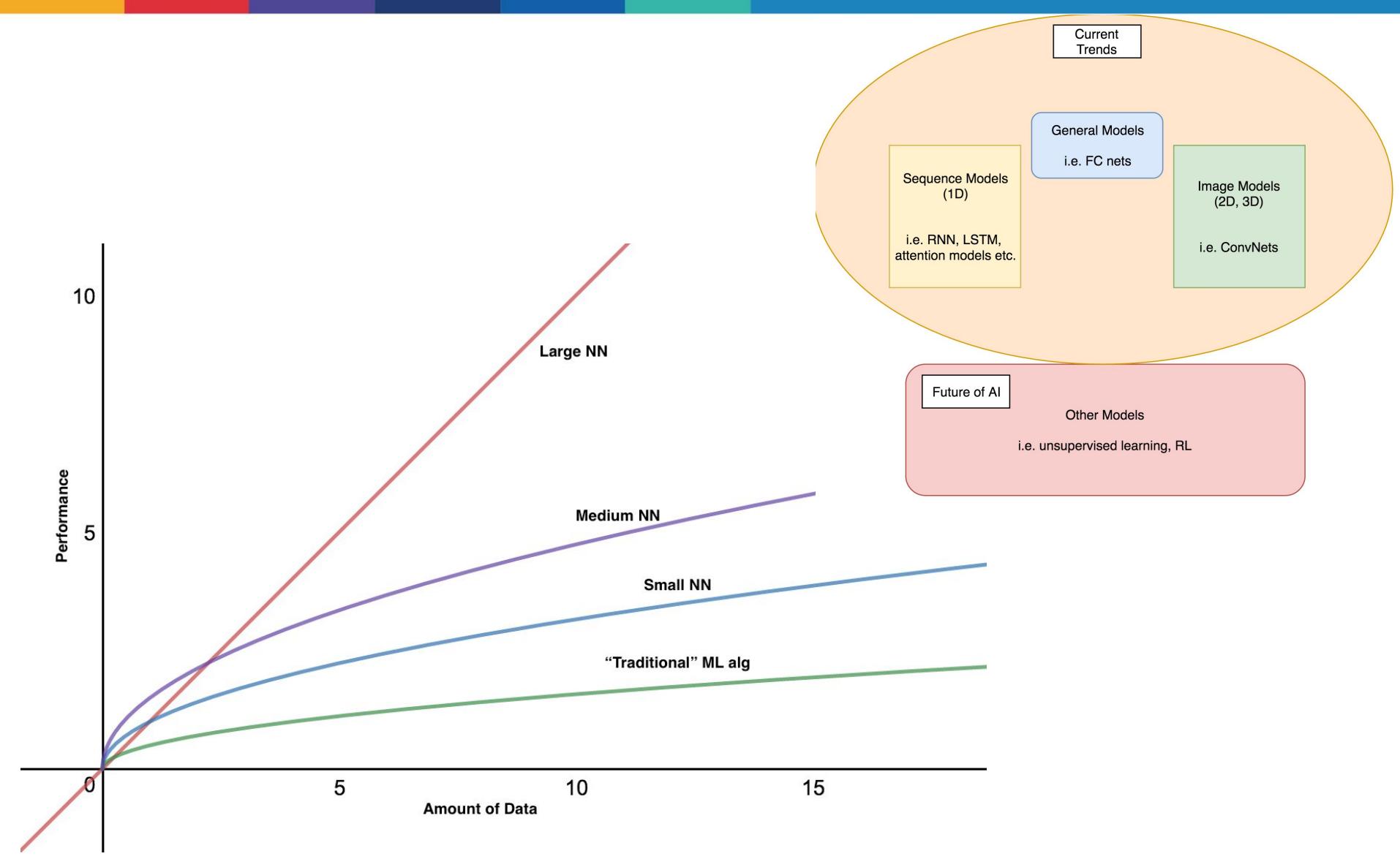
UNSUPERVISED AND TRANSFER LEARNING

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

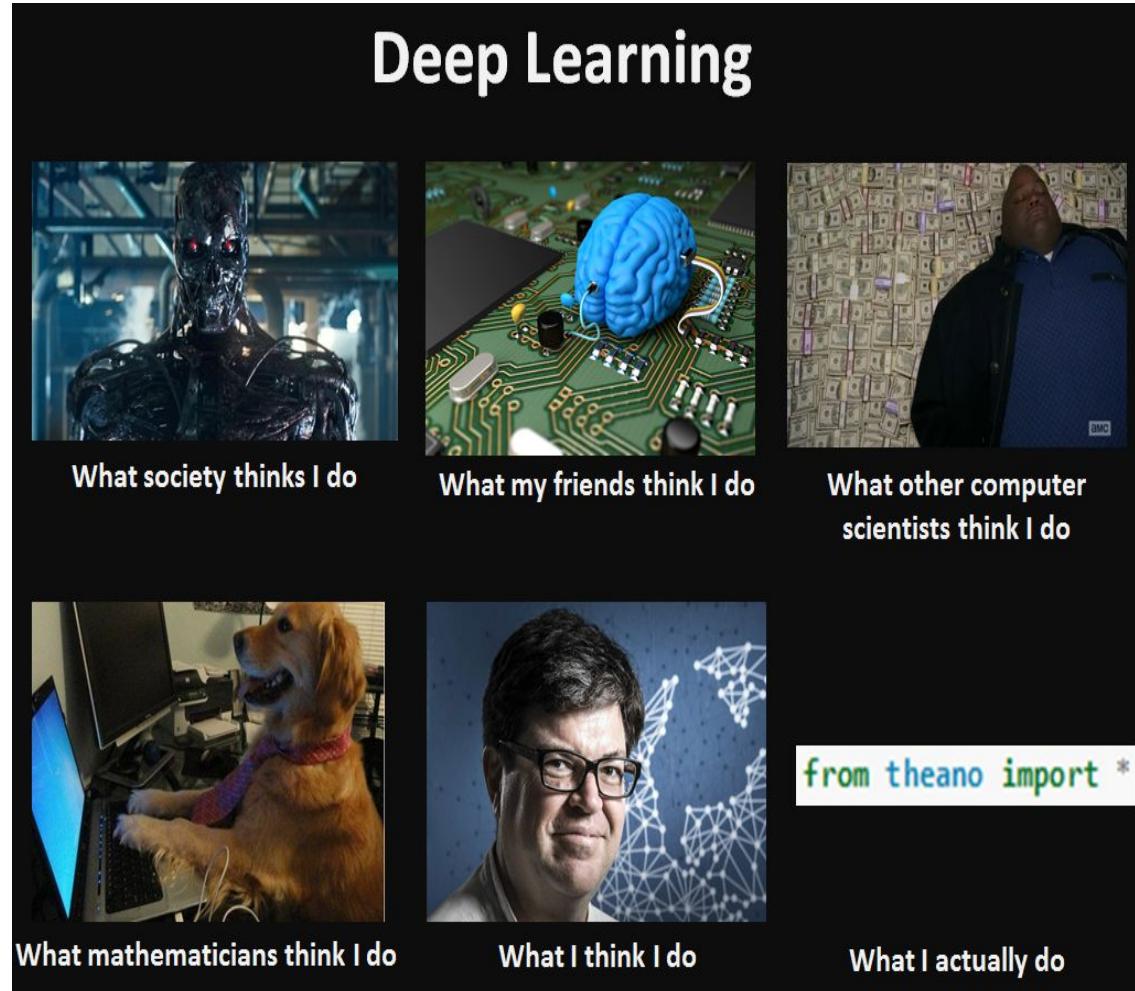


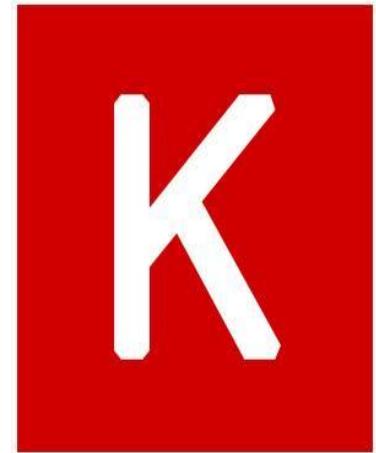
Image taken from Internet

Open source/Frameworks

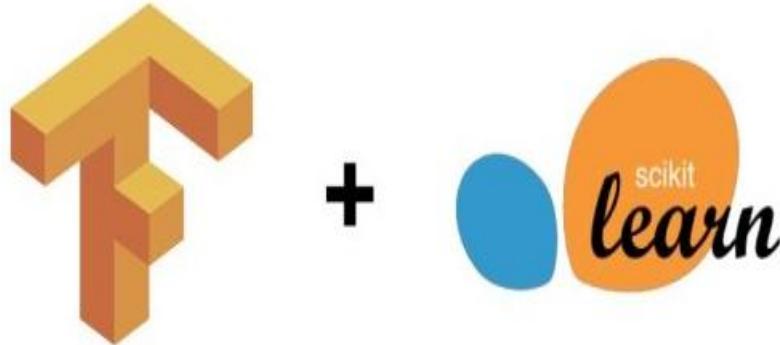


TensorFlow

theano



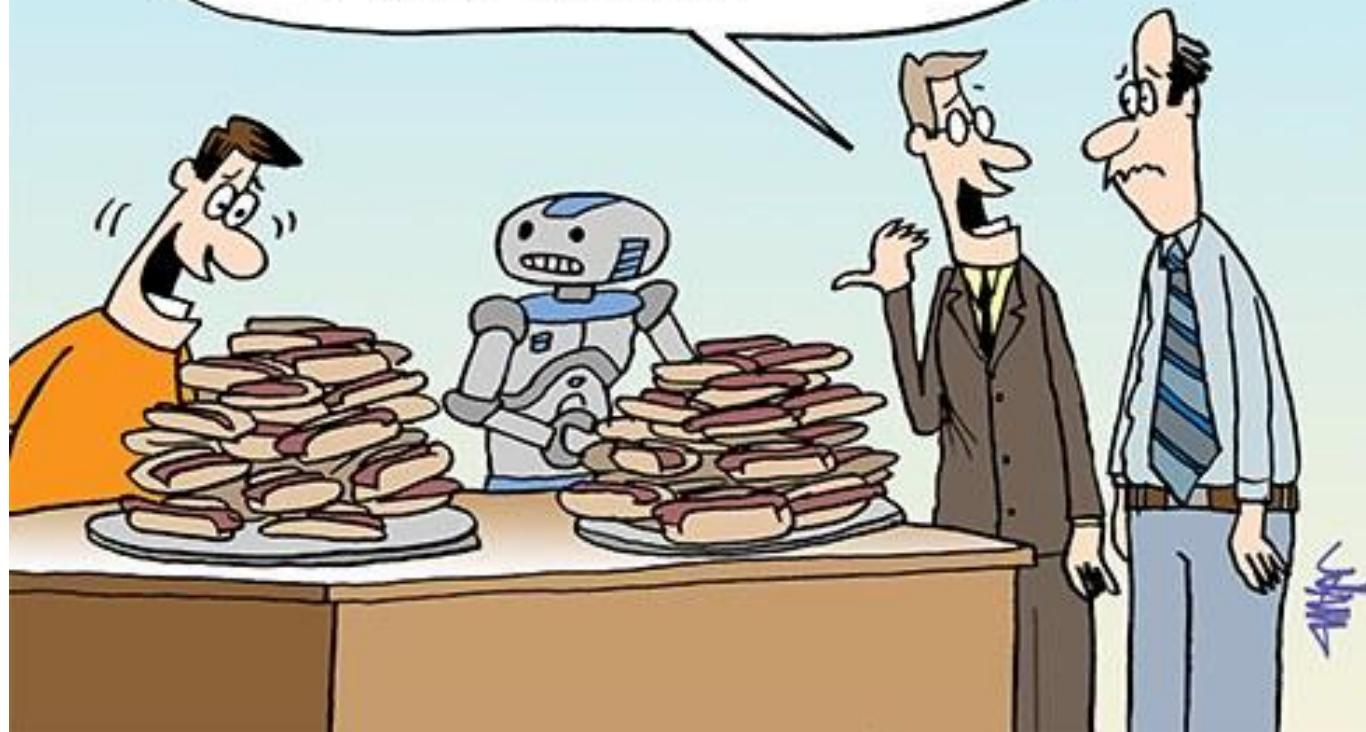
torch



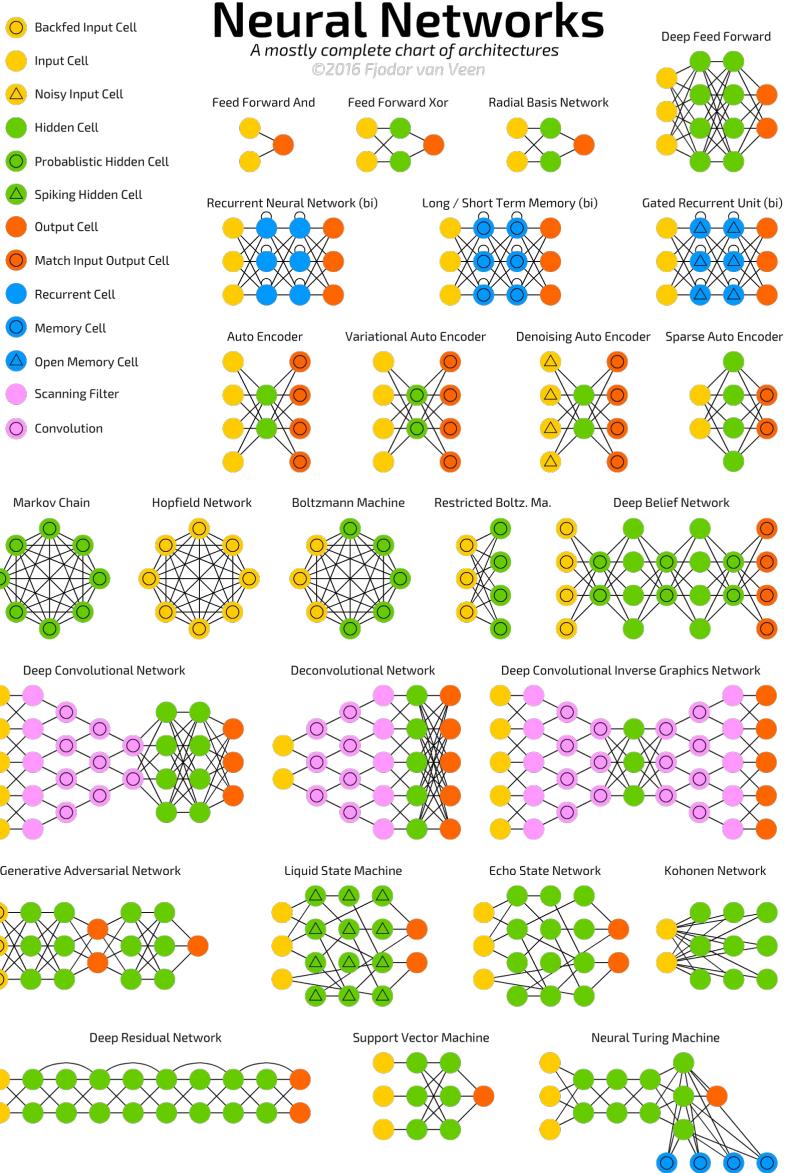
Caffe

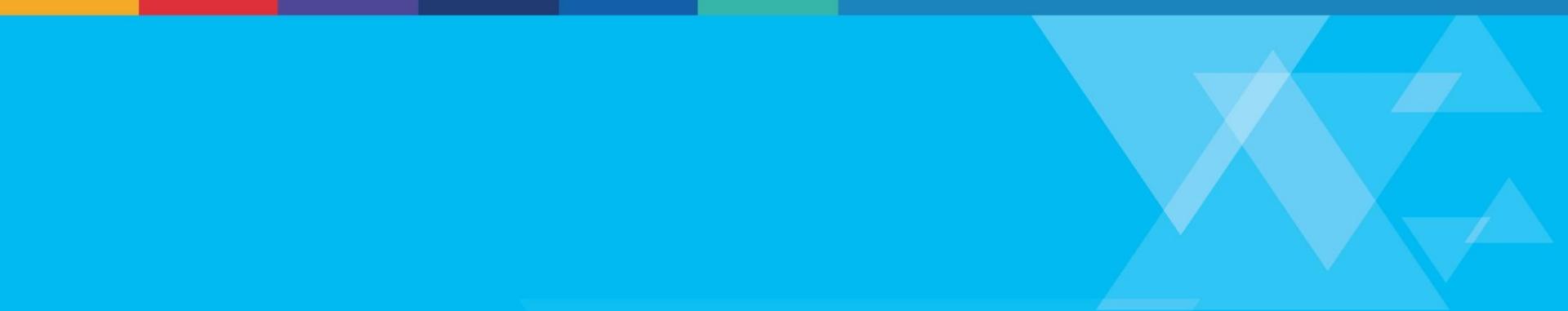
Top libraries by Github issues opened		Top libraries by Github stars			
#1:	2908	BVLC/caffe	#1:	29967	tensorflow/tensorflow
#2:	2530	fchollet/keras	#2:	11914	BVLC/caffe
#3:	2456	tensorflow/tensorflow	#3:	7595	fchollet/keras
#4:	1801	dmlc/mxnet	#4:	5985	Microsoft/CNTK
#5:	1705	Theano/Theano	#5:	5263	karpathy/convnetjs
#6:	1067	deeplearning4j/deeplearning4j	#6:	5160	torch/torch7
#7:	693	Microsoft/CNTK	#7:	4740	dmlc/mxnet
#8:	505	mila-udem/blocks	#8:	4316	Theano/Theano
#9:	498	pfnet/chainer	#9:	3723	deeplearning4j/deeplearning4j
#10:	494	NVIDIA/DIGITS	#10:	3420	tflearn/tflearn
#11:	394	Lasagne/Lasagne	#11:	3162	amznlabs/amazon-dsstne
#12:	342	torch/torch7	#12:	2372	Lasagne/Lasagne
#13:	233	NervanaSystems/neon	#13:	2149	NervanaSystems/neon
#14:	206	tflearn/tflearn	#14:	1577	pfnet/chainer
#15:	82	IDSIA/brainstorm	#15:	1371	NVIDIA/DIGITS
#16:	41	karpathy/convnetjs	#16:	1147	IDSIA/brainstorm
#17:	39	amznlabs/amazon-dsstne	#17:	870	mila-udem/blocks
#18:	27	torchnet/torchnet	#18:	787	torchnet/torchnet
Top libraries by Github contributors		Top libraries by Github forks			
#1:	348	tensorflow/tensorflow	#1:	12506	tensorflow/tensorflow
#2:	244	Theano/Theano	#2:	7194	BVLC/caffe
#3:	234	fchollet/keras	#3:	2275	fchollet/keras
#4:	202	BVLC/caffe	#4:	1777	dmlc/mxnet
#5:	169	dmlc/mxnet	#5:	1540	Theano/Theano
#6:	102	torch/torch7	#6:	1484	torch/torch7
#7:	84	deeplearning4j/deeplearning4j	#7:	1291	Microsoft/CNTK
#8:	75	Microsoft/CNTK	#8:	1264	deeplearning4j/deeplearning4j
#9:	72	pfnet/chainer	#9:	1024	karpathy/convnetjs
#10:	50	Lasagne/Lasagne	#10:	662	Lasagne/Lasagne
#11:	48	mila-udem/blocks	#11:	482	amznlabs/amazon-dsstne
#12:	42	NervanaSystems/neon	#12:	450	NervanaSystems/neon
#13:	39	tflearn/tflearn	#13:	412	NVIDIA/DIGITS
#14:	28	NVIDIA/DIGITS	#14:	377	pfnet/chainer
#15:	16	amznlabs/amazon-dsstne	#15:	336	tflearn/tflearn
#16:	15	IDSIA/brainstorm	#16:	267	mila-udem/blocks
#17:	14	karpathy/convnetjs	#17:	161	torchnet/torchnet
#18:	10	torchnet/torchnet	#18:	108	IDSIA/brainstorm

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





Find me: @laampt | Github: lampsts

**THANK
YOU!**