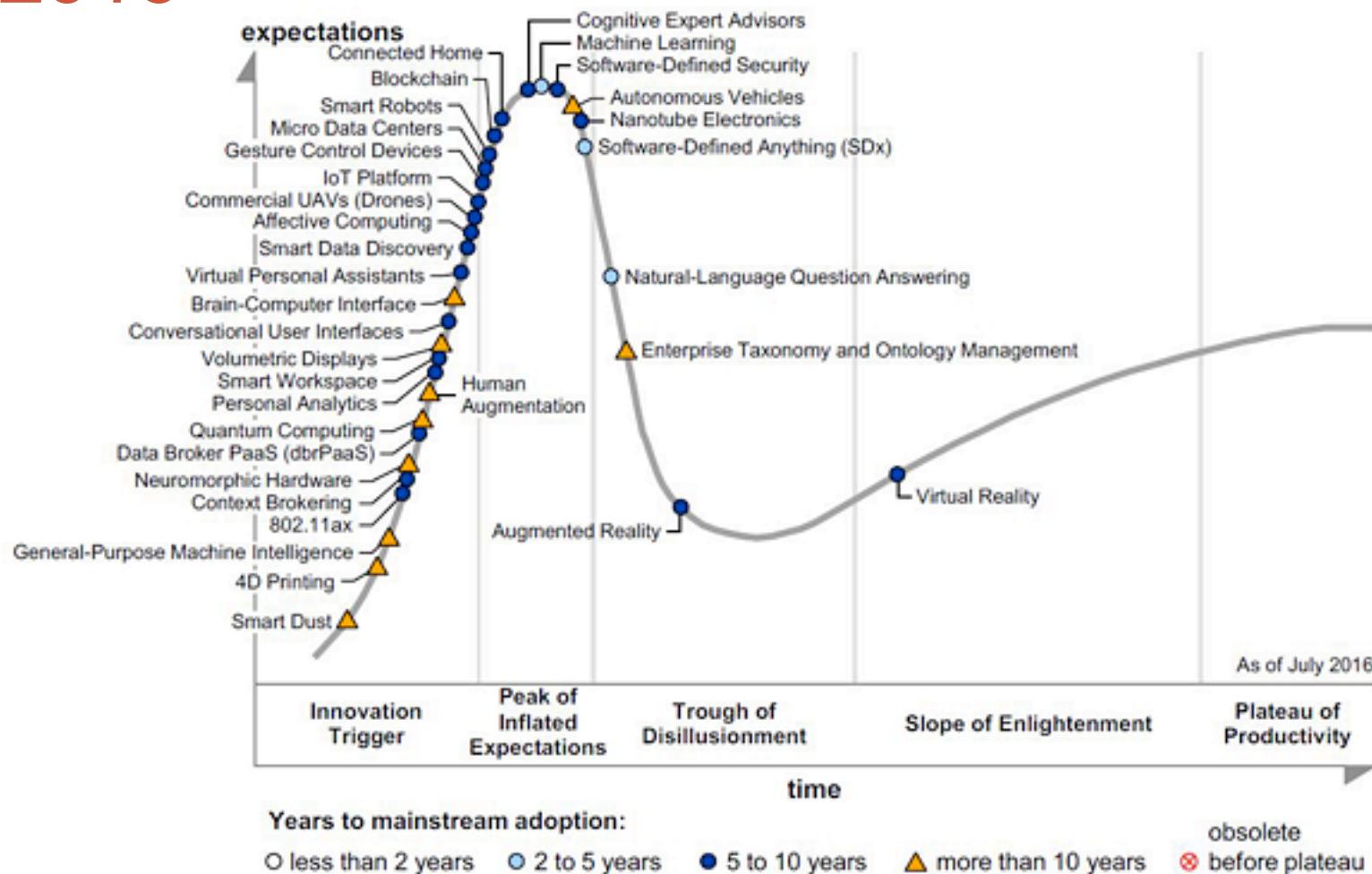


CAPITALIZING THE DATA OPPORTUNITY

Ekapol Chuangsuwanich

Given at Home.tech
November 30, 2017

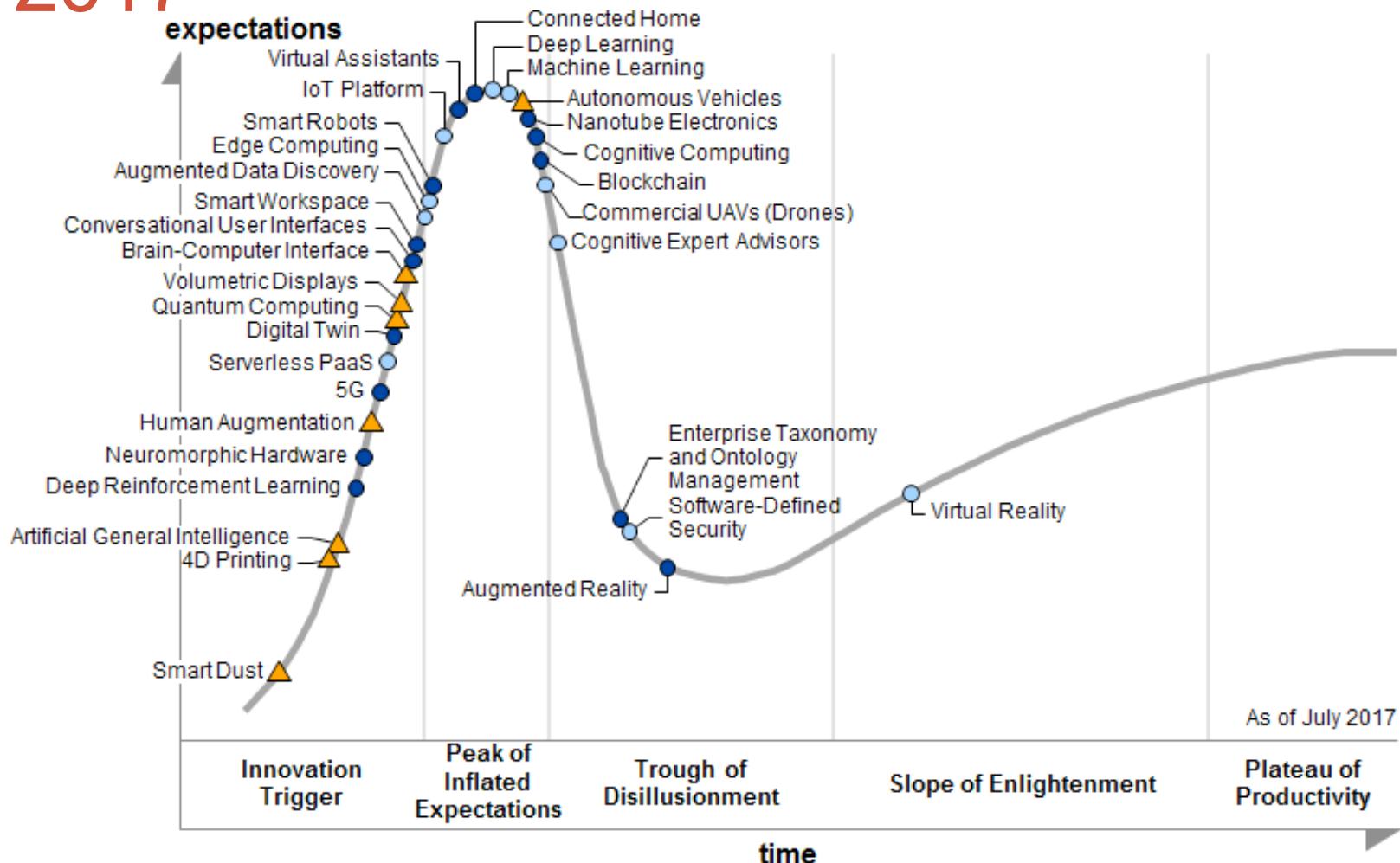
Hype Cycle for Emerging Technologies, 2016



Source: Gartner (July 2016)

<http://www.gartner.com/newsroom/id/3412017>

Hype Cycle for Emerging Technologies, 2017



Years to mainstream adoption:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ✕ before plateau

<https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017>

Machine Intelligence LANDSCAPE

CORE TECHNOLOGIES

ARTIFICIAL INTELLIGENCE	DEEP LEARNING	MACHINE LEARNING	NLP PLATFORMS	PREDICTIVE APIs	IMAGE RECOGNITION	SPEECH RECOGNITION
IBM WATSON MetaMind Numenta ai-one Cycorp Microsoft Research nara Reactor Scaled Inference	vicarious Vision Factory facebook Google SKYMINDD Baidu IDL ersatz SignalSense	rapidminer context relevant Oxdata H2O DATA RAMP Liftoff SPARKBEYOND Azure ML Shat GraphLab Alpine Sense AYASDI	cortical.io LUMINOSO Maluuba wit.ai	AlchemyAPI™ MINDOPS Google bigml indico ALGORITHMIKA PredictionIO Expect Labs	clarifai MADBITS DNNresearch DEXTRO ViSENZE lookflow	GRIDSPACE popUP archive NUANCE

RETHINKING ENTERPRISE

Thousands of machine learning related startups

SALES	SECURITY / AUTHENTICATION	FRAUD DETECTION	HR / RECRUITING	MARKETING	PERSONAL ASSISTANT	INTELLIGENCE TOOLS
Preact AVISO RelateIQ NGDATA CLARABRIDGE FRAMED infer ATTENITY causata	CROSSMATCH EVERVERIFY CYLANCE BITSIGHT AREA 1 SECURITY biomly	sift science SOCURE ThreatMetrix feedzai Brighterion VERAFIN	TalentBin entelo predikt Connectifier gild hiQ CONCEPTNODE	brightfunnel bloomreach CommandIQ AIRPR RADIUS TellApart people pattern Freshplum	Siri Google now Cortana cleversense tempo Reblab KASISTO fusemachines VIV CLARA LABS	ADATAQ Palantir Quid Digital Reasoning

RETHINKING INDUSTRIES

ADTECH	AGRICULTURE	EDUCATION	FINANCE	LEGAL	MANUFACTURING	MEDICAL
METAMARKETS dstillery rocketfuel YieldMo ADBRAIN	BLUEBIRD RIVER ceresimaging THE CLIMATE CORPORATION HONEYCOMB tule X MARKS	Declarative coursera KNEWTON kidaptive	Bloomberg alphasense Dataminr	Lex Machina brightleaf COUNSELYTICS minetta brook BINATIX DiligenceEngine	SIGHT MACHINE MICROSCAN IVISYS BOULDER IMAGING	Parzival transcripcive ZEPHYR HEALTH Genescient grand round table bina TUTE GENOMICS

OIL AND GAS	MEDIA / CONTENT	CONSUMER FINANCE	PHILANTHROPIES	AUTOMOTIVE	DIAGNOSTICS	RETAIL
kaggle biota TACHYUS Flutura	Outbrain newstle SAILTHRU wavii NarrativeScience YSEOP Prismatic ai AUTOMATED INSIGHTS	Affirm inVenture zest finance BILL GUARD LendingClub Kabbage	DataKind thorn the DATA GUILD	Google Continental TENKA MOBILEYE CRUISE	enlitic 3SCAN lumiata ENTAVIS	BAY SENSORS PRISM SKYLABS select euclid

RETHINKING HUMANS / HCI

AUGMENTED REALITY	GESTURAL COMPUTING	ROBOTICS	EMOTIONALrecognition	HARDWARE	DATA PREP	DATA COLLECTION
wearable intelligence ARX APX blippar META layar	THALMICLABS FLUTTER eyeSight™ GestureTek 3Gear Systems nod	Intel LEAP MOTION iRobot Boston Dynamics jibo Anki Evolution robotics	affectiva BEYONDVERBAL EMOTIENT cogito	NVIDIA XILINX Qualcomm NERVANA SYSTEMS TERADEEP Artificial Learning rigetti	TRIFACTA Paxata. tamr Alation	diffbot kimono CrowdFlower Connate WorkFusion import.io

“AI first”



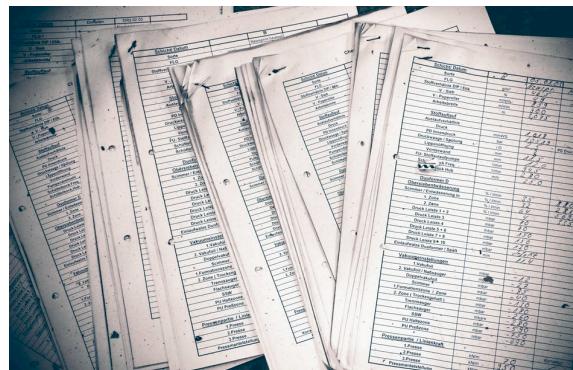
Google I/O – May, 17



Microsoft developer conference – May, 17

Deriving ROI from Machine Learning

- What are needed for Machine Learning in business applications?



Data



Business goal



Team



Business goal

Pick the task



Pick the problem

- The goal can dictate the collection of data and the machine learning methods
- Treat it just like any other business plan
 - Machine Learning tasks are harder to evaluate the potential benefits

Machine learning is not a magic bullet



Example: Recommendation systems

- How can the recommendation system generates value?
- How can it be integrated into the existing UX flow?
- How can this feature differentiate you from competitors?
- What is a minimum requirement for the feature?

Customers who bought this item also bought

Page 1 of 13



Dragonpad USA Pop filter
Studio Microphone Mic
Wind Screen Pop Filter
★★★★★ 3,564
\$8.39 ✓prime



NEEWER Adjustable
Microphone Suspension
Boom Scissor Arm Stand,
Compact Mic Stand...
★★★★★ 3,050

#1 Best Seller in
Microphone Stands
\$12.50 ✓prime



Elgato Chat Link, Party
Chat Adapter for Xbox One
and PlayStation 4
★★★★★ 330
\$9.95 ✓prime



Neewer NW(B-3) 6 inch
Studio Microphone Mic
Round Shape Wind Pop
Filter Mask Shield with...
★★★★★ 882

#1 Best Seller in
Microphone Windscreens &
Pop...
\$6.99



Elgato Game Capture
HD60 S - stream, record
and share your gameplay
in 1080p60, superior low...
★★★★★ 1,544
\$179.94 ✓prime



Elgato Game Capture
HD60, for PlayStation 4,
Xbox One and Xbox 360,
or Wii U gameplay, Full...
★★★★★ 1,544

#1 Best Seller in Internal
TV Tuner & Video...
\$151.74 ✓prime



What tasks are good for machine learning?

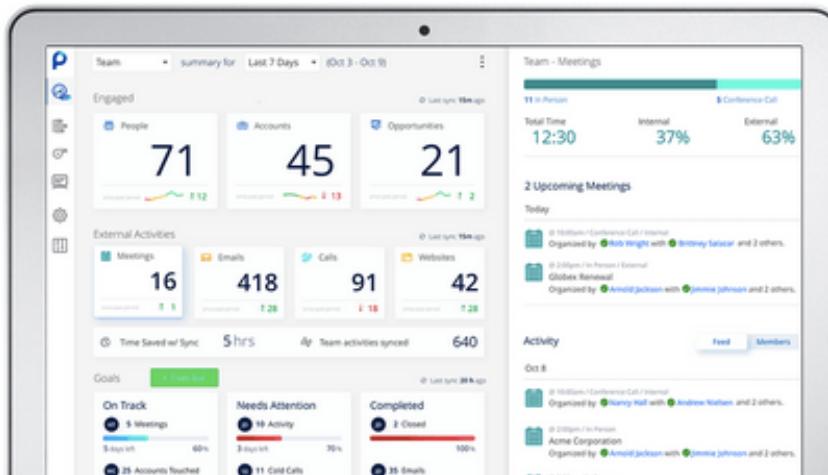
- Tasks that requires less than 1 second to think
 - ASR, Object recognition, Face recognition
- Augmenting human performance
 - Decision support system
 - Healthcare diagnosis
 - Self-driving cars
- Tasks at a large scale





Machine learning in sales

- people.ai



RAMP

Provide new hires with a proven playbook from historical data on your best sales people



BENCHMARK

Create custom metrics and benchmarks to drive sales behavior



ACCOUNT BASED SELLING

Target efforts on the right accounts, with the right activities, at the right time



COACHING

Help your team work smarter with AI-powered coaching and personalized feedback



TEAM GOALS

Take action to increase performance and reinforce the right behavior

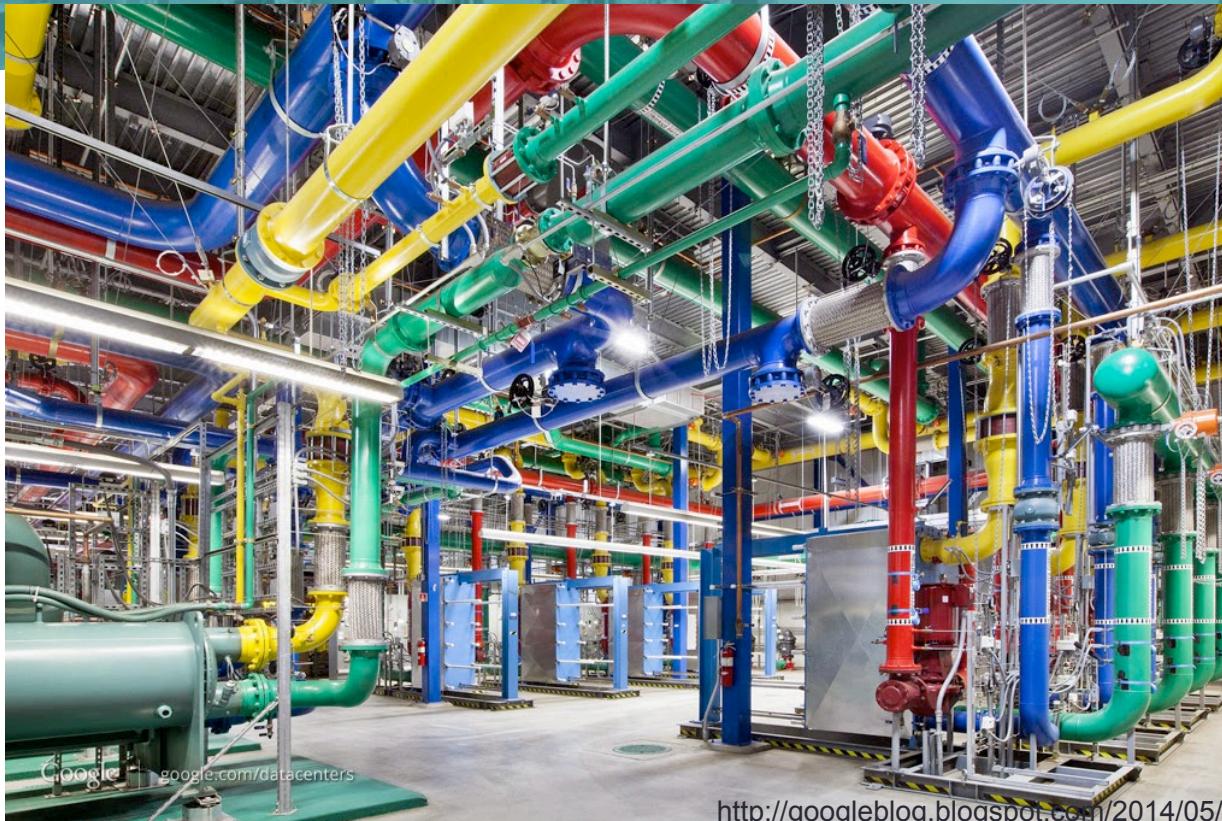


DEAL ANALYTICS

Track multithreading and other key activities needed to win a deal



DeepMind AI Reduces Google Data Centre Cooling Bill by 40%





Aspects to consider

- Metrics
- Timings





Defining a metric

- Machine learning methods are evaluated based on metrics

Speech Recognition - Word Error Rate

Search – Recall rate

Machine Translation - BLEU score

Recommendation system – Click through rate

- Tasks can have multiple metrics
- Understand the relationship between the metric and the business goal



Example: Chatbots

- Task completion rate
- Number of turns to complete a task
- Satisfaction
- Word Error Rate (if speech input)





Getting the human performance baseline

- Find the performance of a human performing the same task



- Guides the development of the machine learning
- Helps in identifying the benefits of machine learning



Understand the trade-offs

- Performance metrics
- Amount of computation power required
- Type of computation required
- Latency





Understand the timings

- Most data are not ready for machine learning
- Data preparation takes time



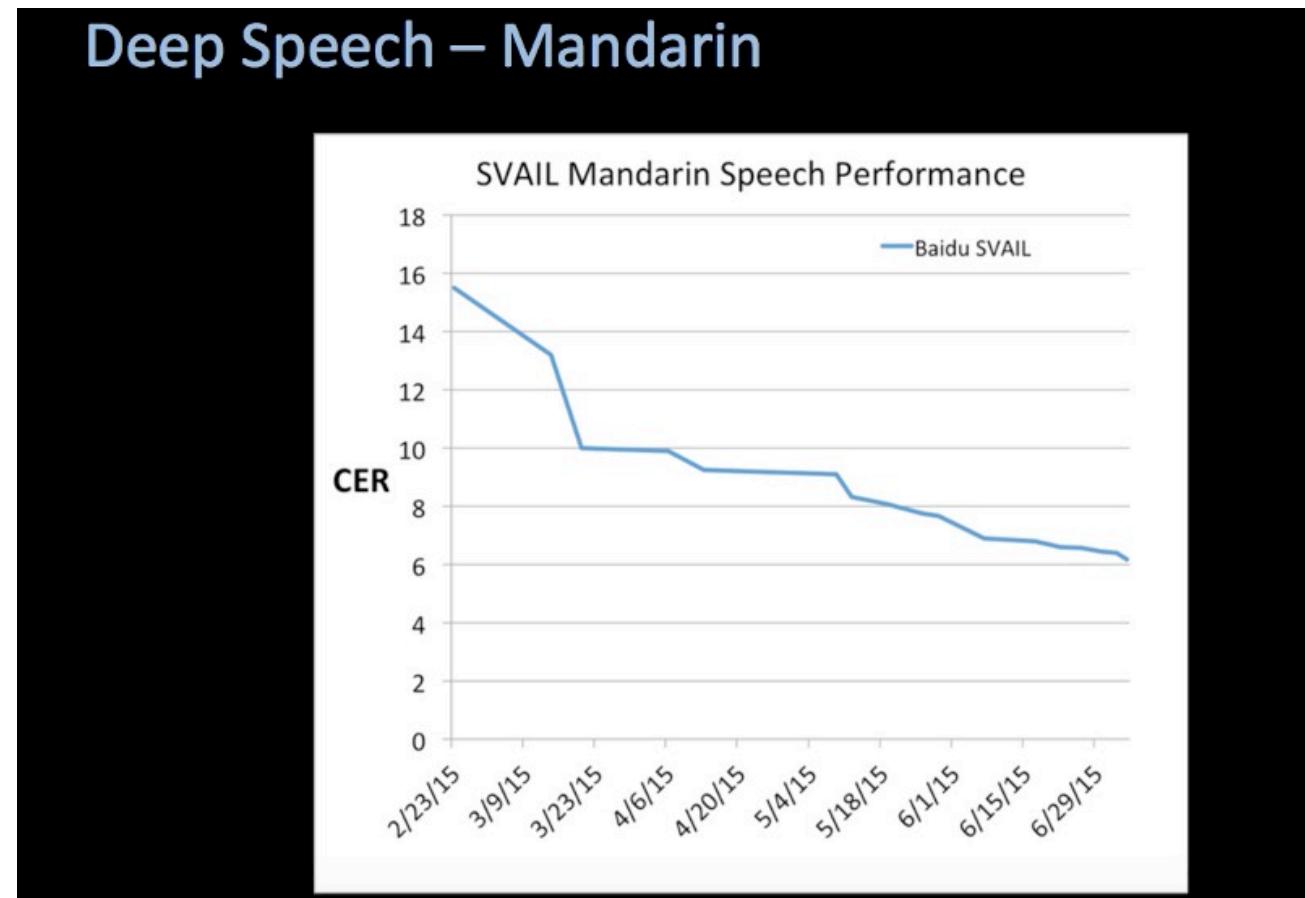
<http://www.nvidia.com/object/drive-px.html>

- Improving performance takes time

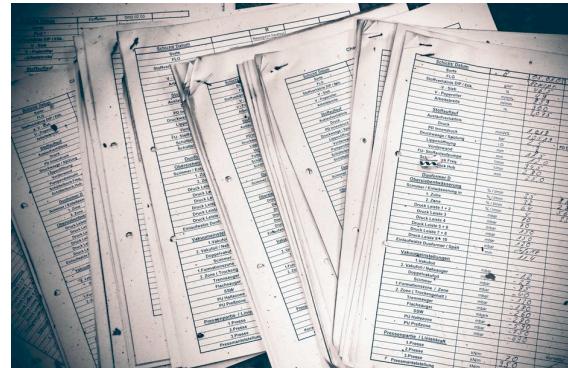


Performance and time spent on development

- 80/20 rule



<https://medium.com/s-c-a-l-e/how-baidu-mastered-mandarin-with-deep-learning-and-lots-of-data-1d94032564a5>



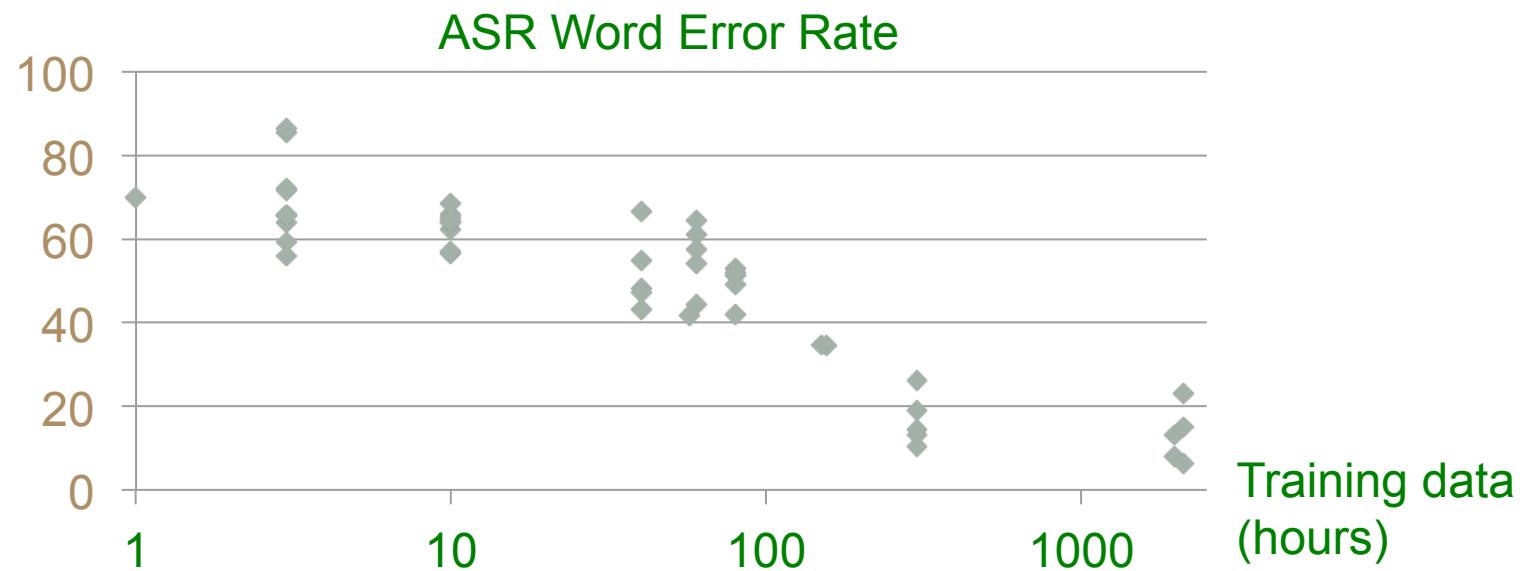
Data

Accumulate data

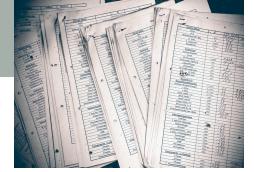


Machine learning uses data

- More data means better models



- You can start a machine learning project without data
 - If the business plan says it's feasible

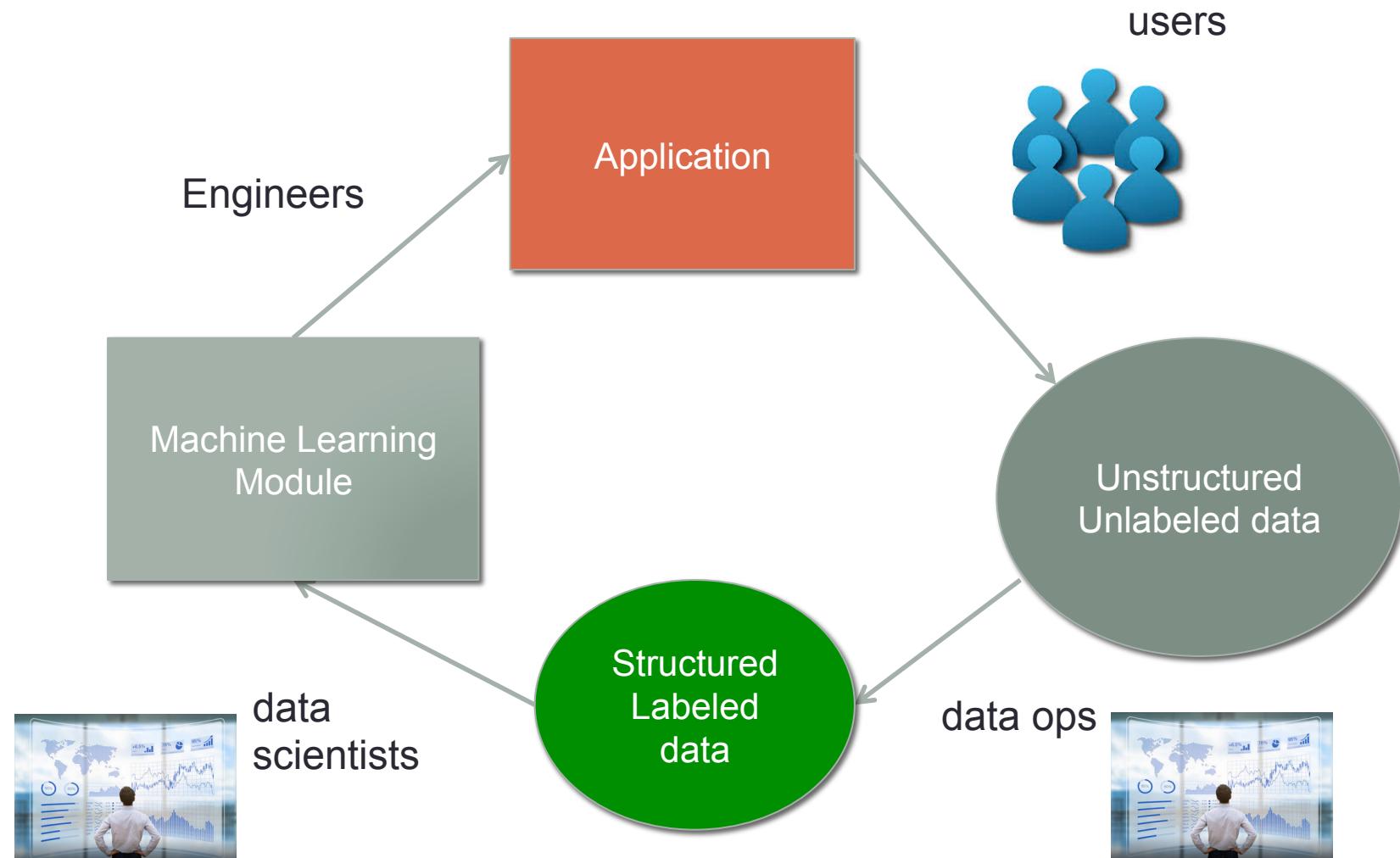


Machine learning uses labeled data



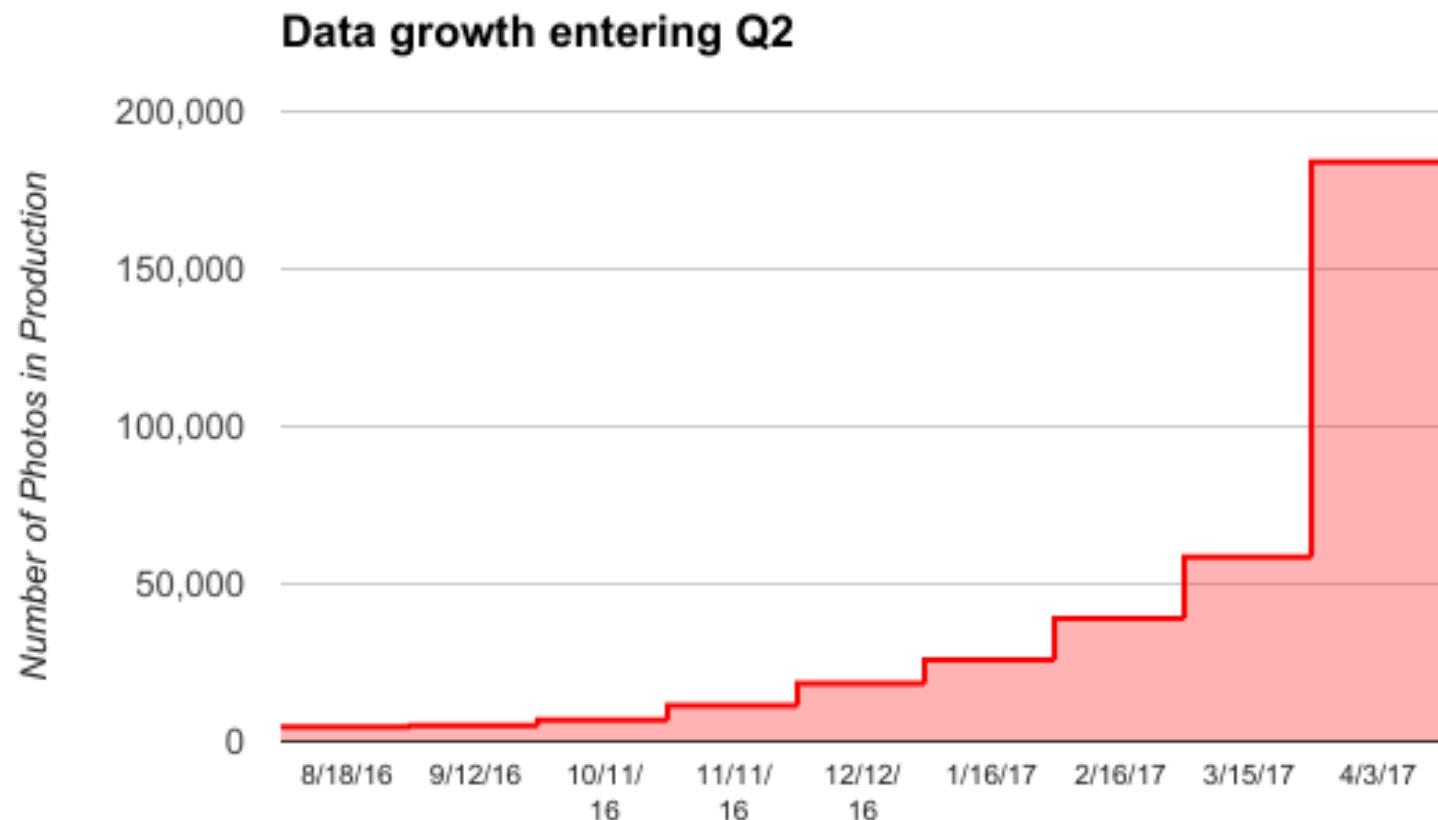


Getting into the data cycle





Getting into the data cycle





Crowdsourcing

- Distribute data labeling to the general public
- Need some incentives

Database
ODOO-ReCaptcha

Email
admin

Password

You have Cross The
Your Account has been
through CAPTCHA!

Log in

Report a problem

Verify

SIGN UP **LOG IN WITH FACEBOOK**





Mturk Example



Is receipt valid?

Yes

Business Name:

Payment Method:

Choose below...

 CASH

 DEBIT

 CREDIT - AMERICAN EXPRESS (AMEX)

 CREDIT - DISCOVER

 CREDIT - MASTERCARD (MCARD)

 CREDIT - VISA (VS)

 CREDIT - NOT SPECIFIED

 FOOD STAMP

 EBT

 SNAP

 WIC

 CHECK CARD

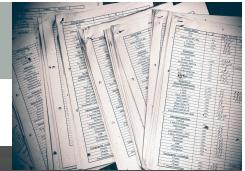
 CHECK

 GIFT CARD

 MONEY CARD

 - Payment method not in above list

 - No payment method is on receipt



Social Listening Labeler x +

localhost:3000/app/project/5257/matchId-string0 5257

Social Listening app project 5257 **matchId-string0%205257**

[Logout of "user1"](#)

Data Info

	Value
dataId	"dataId string"
title	"Title string"
content	... Show all อกตัวก่อนนะจะ เป็นกระซู่แรก รู้สึกทำอะไรไม่ถูกเลย เกี่ยวกับบ้านหลังแรกของเราเลยอยากมากอบรึกษา เรื่มจากเรา กับแฟ้มมองหาบ้านหลังนึง ที่ใกล้บ้านพ่อแม่เราทั้งสองคนค่ะ เลย มาเจอโครงการนี้ค่ะ ตั้งอยู่ที่...
source	"Source string"

Match Info

	Value
matchId	"matchId stringmatchId-string0 5257"
start	"0"
end	"7"

Project Info

	Value
projectId	"projectId string"
 projectName	"projectName string"

Mark as

Valid
Invalid

Marked as

User	Value
user1 (You)	Valid



Team

Build the team



The team

- Data scientists
- Data engineers
- Performance engineers
- Machine learning practitioners
- Labeling team

High competition for recruitment even in USA

- Compute

Common misconceptions business people have

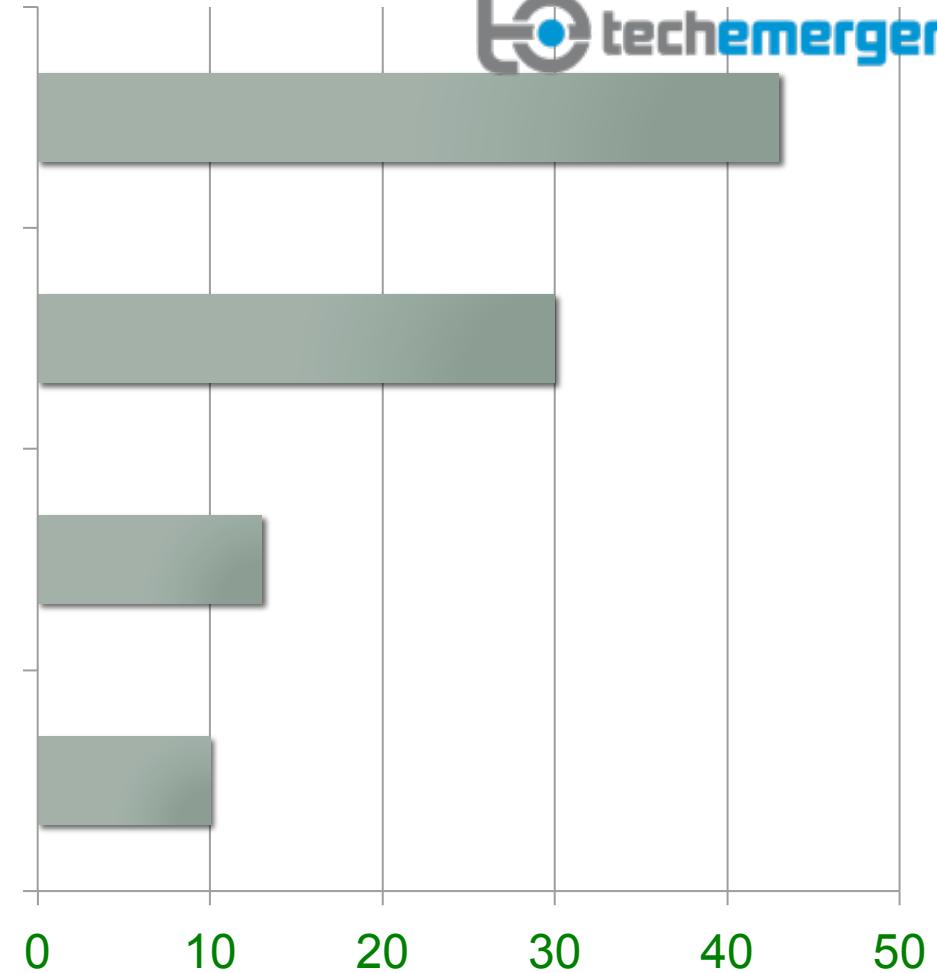


Wrong expectation of capabilities

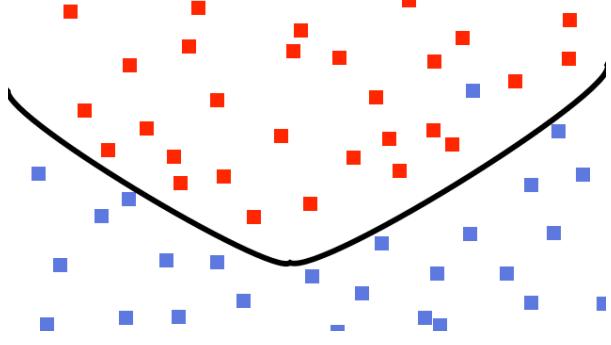
Underestimating resources needed

Technical misunderstanding

Not understanding what AI is



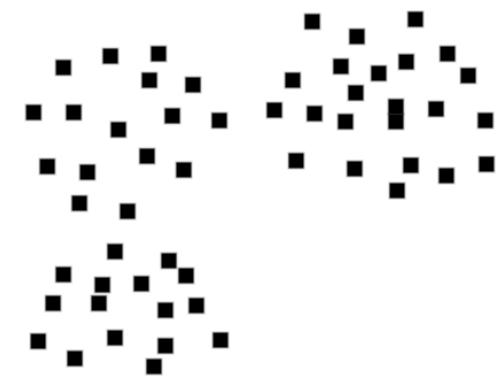
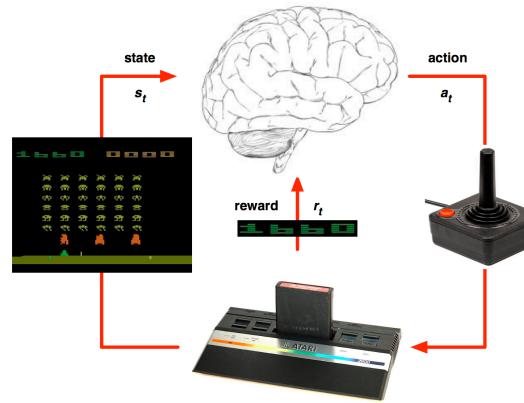
3 Modes of Learning



Supervised Learning



Reinforcement Learning



Unsupervised Learning





Yann LeCun

March 14, 2016 ·

Follow

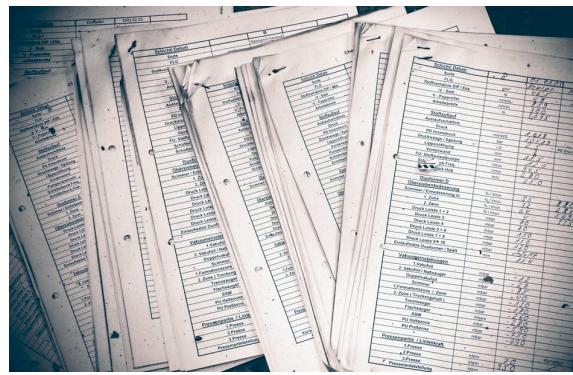
Statement from a Slashdot post about the AlphaGo victory: "We know now that we don't need any big new breakthroughs to get to true AI"

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.



Conclusion



Data



Business goal



Team