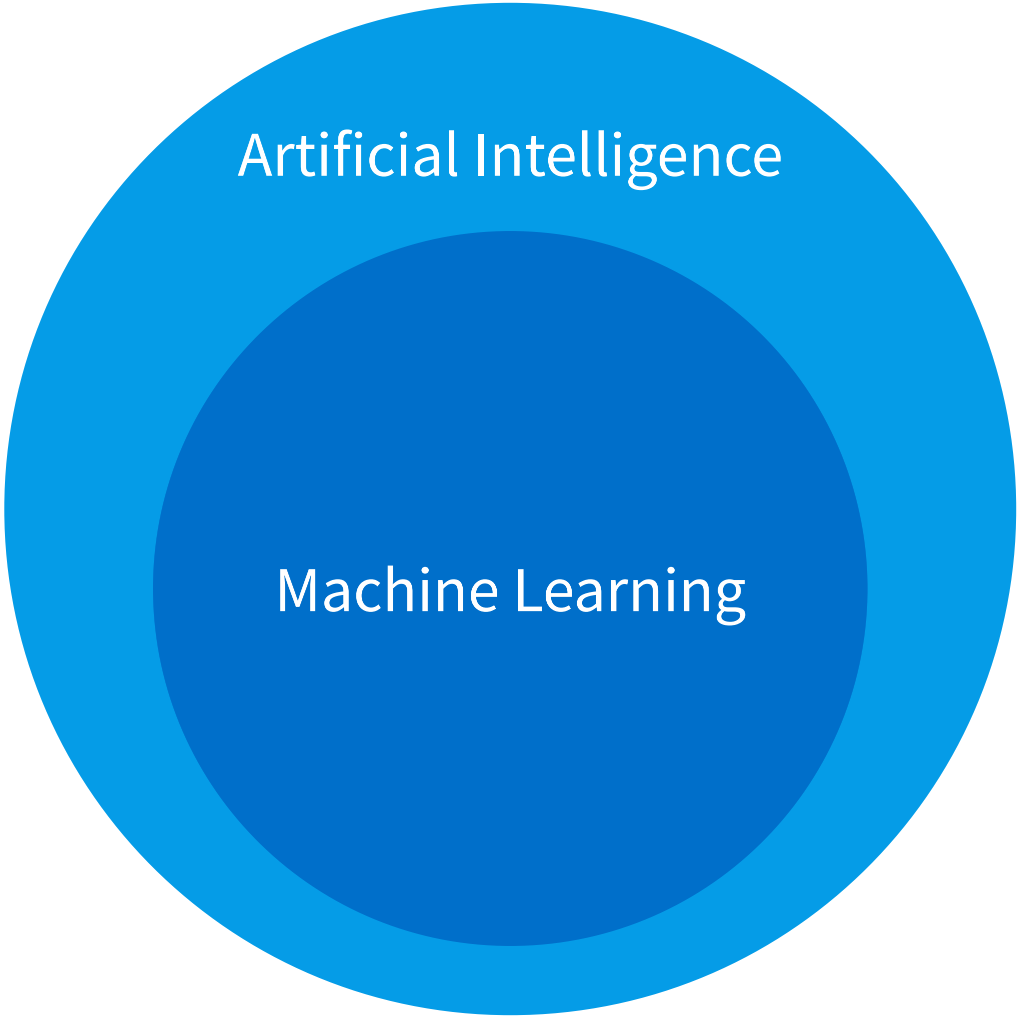
Deep Learning in the Real World

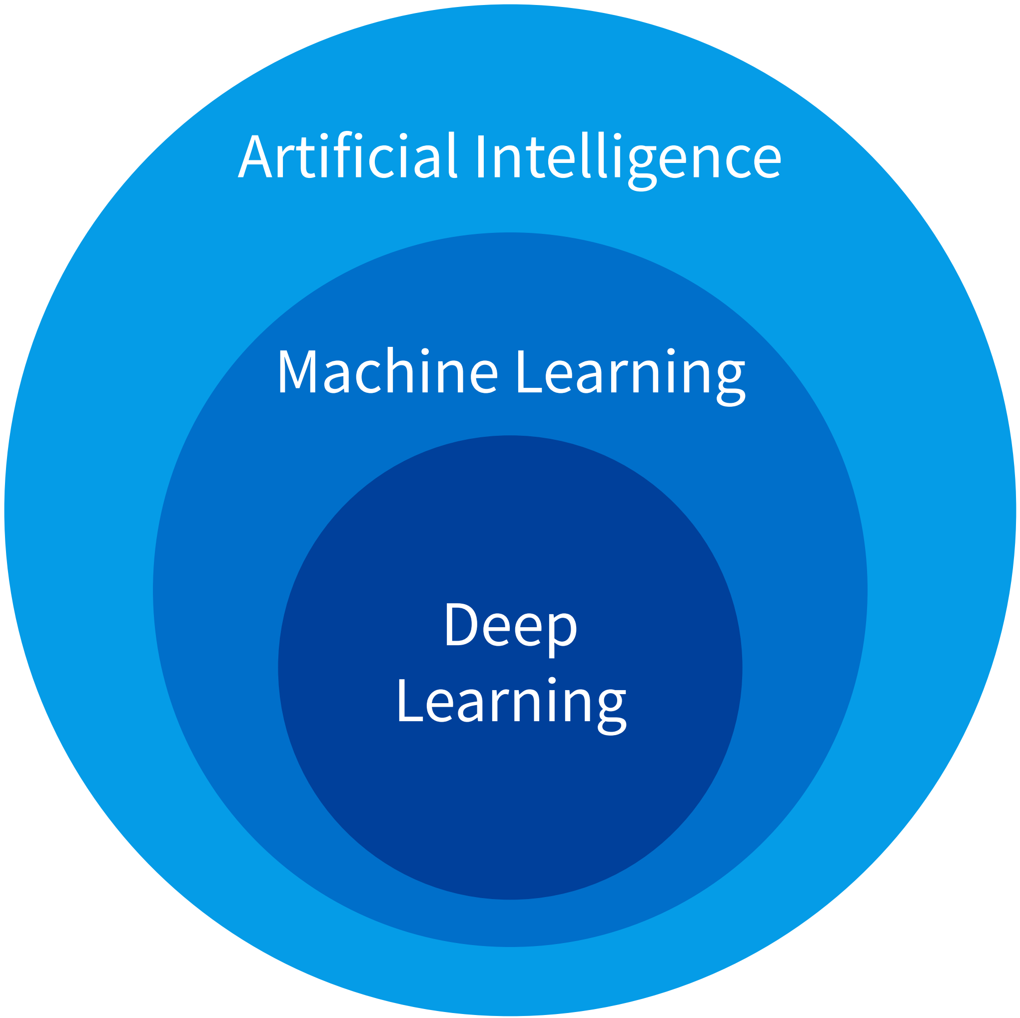
wandb.com

Terminology

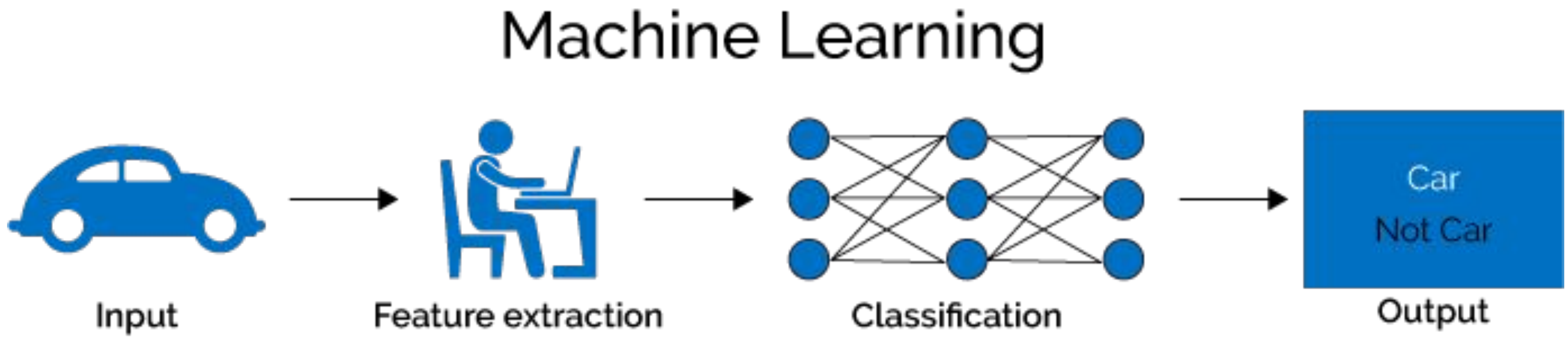
Artificial Intelligence



wandb.com


Artificial Intelligence
Machine Learning
Deep
Learning

Why is deep learning exciting?


Machine Learning
Car
Not Car
Input
Feature extraction
Classification
Output

wandb.com


HXM 950

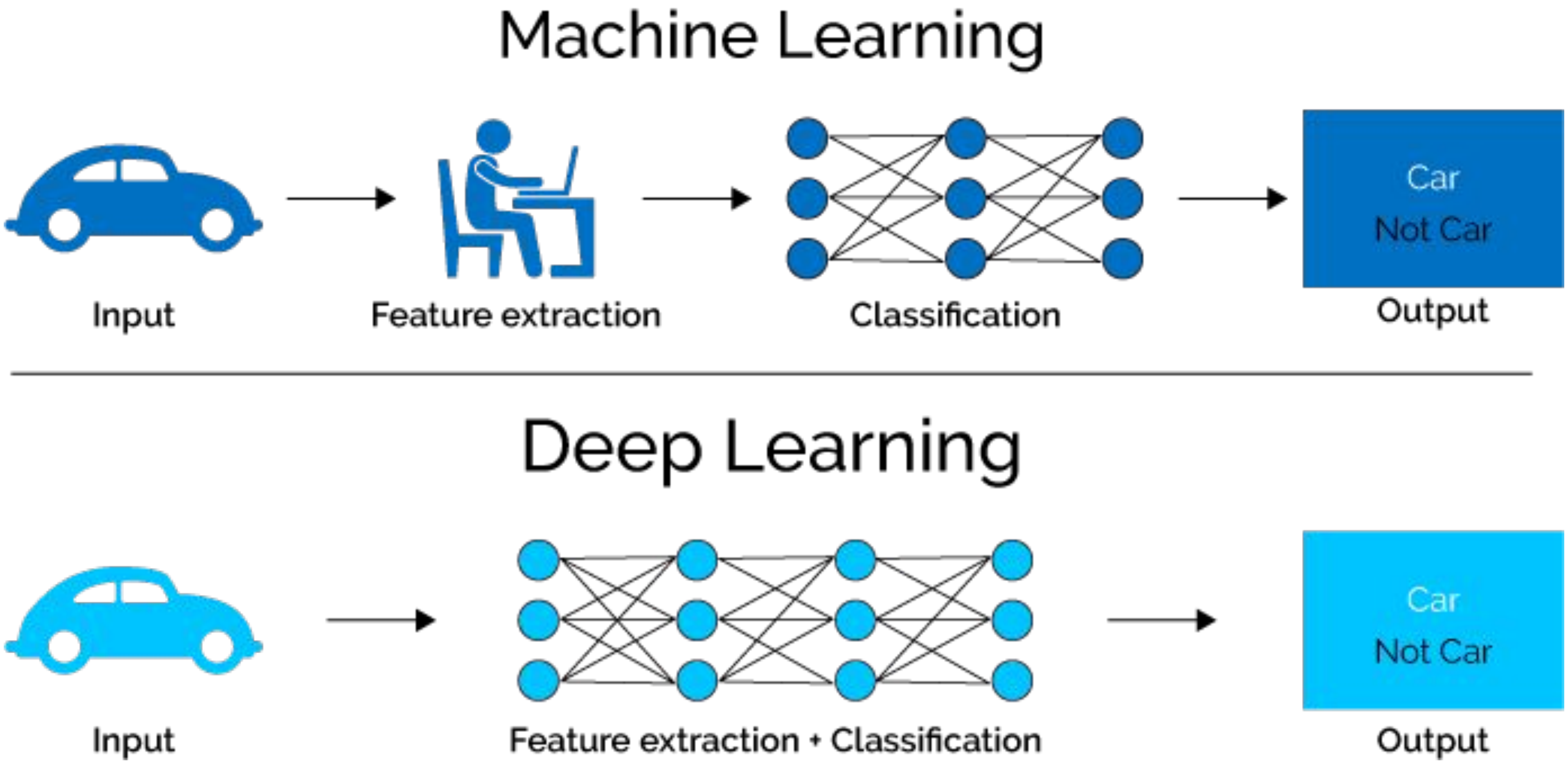

ADVANTAGE

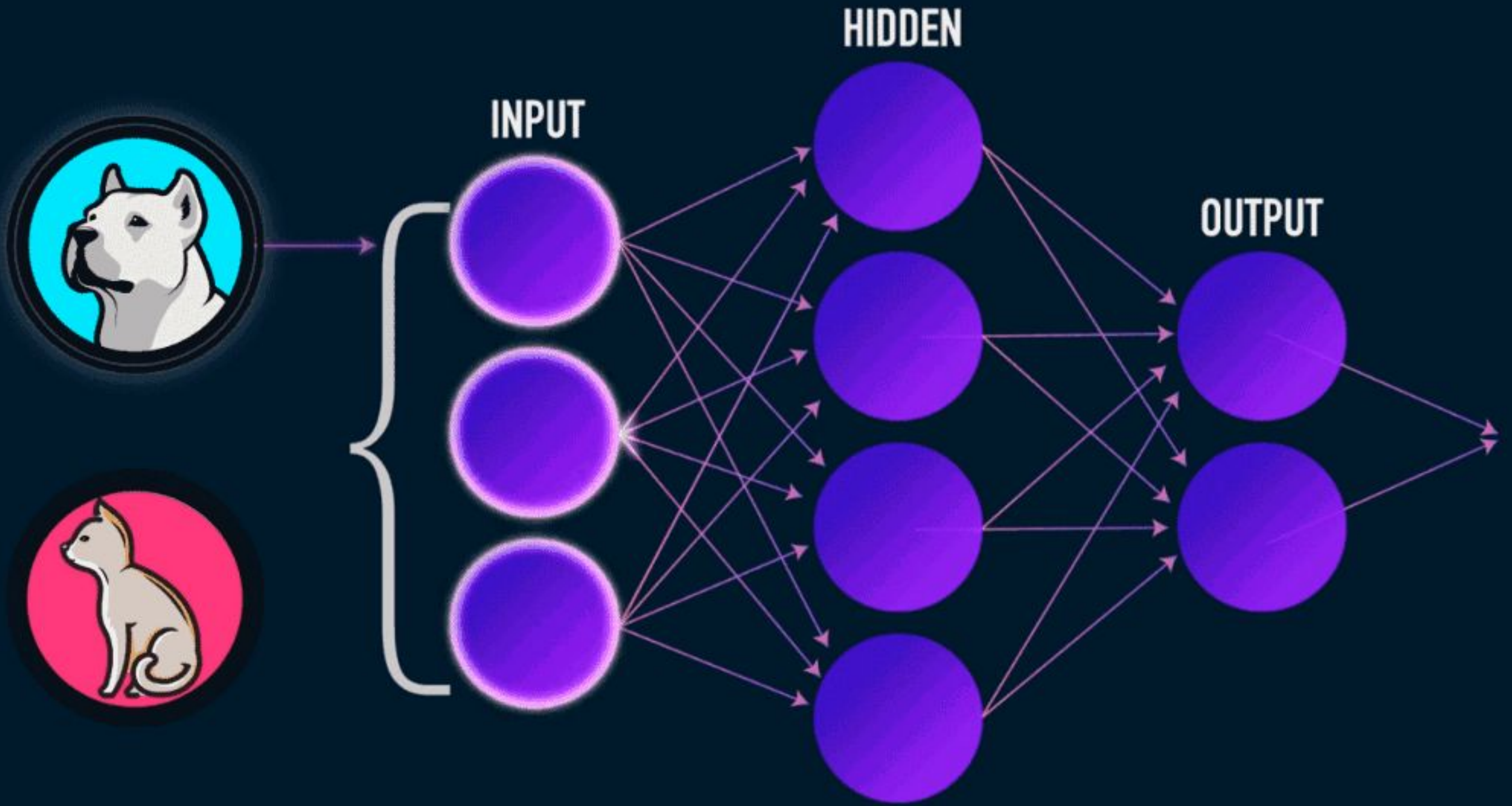
ADVANTAGE

Ferrari


PRAHA
Lamborghini
Lamborghini
Lamborghini PRAHA




Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction + Classification
Car
Not Car
Output


HIDDEN
INPUT
OUTPUT

1

Excitement

2

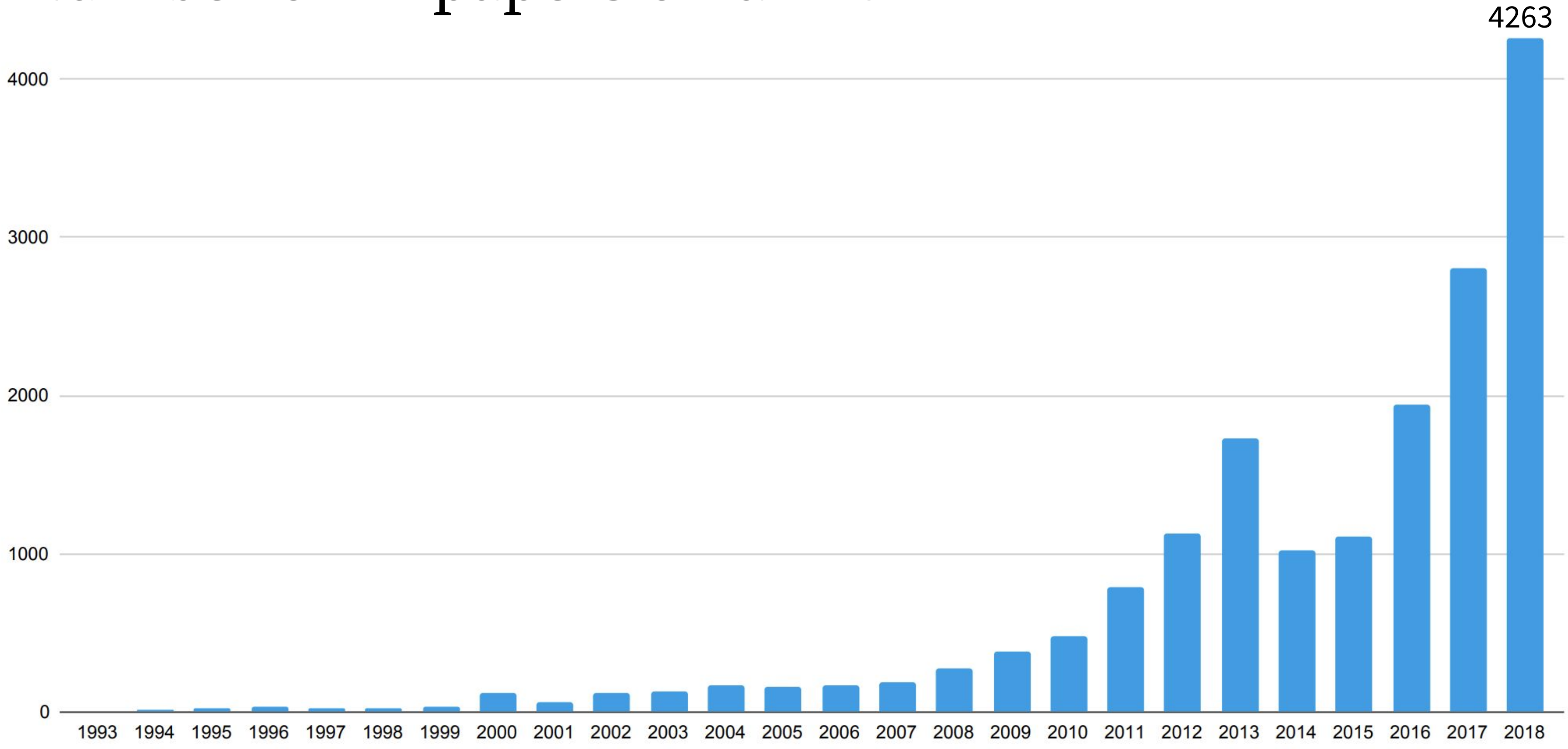
Problems

3

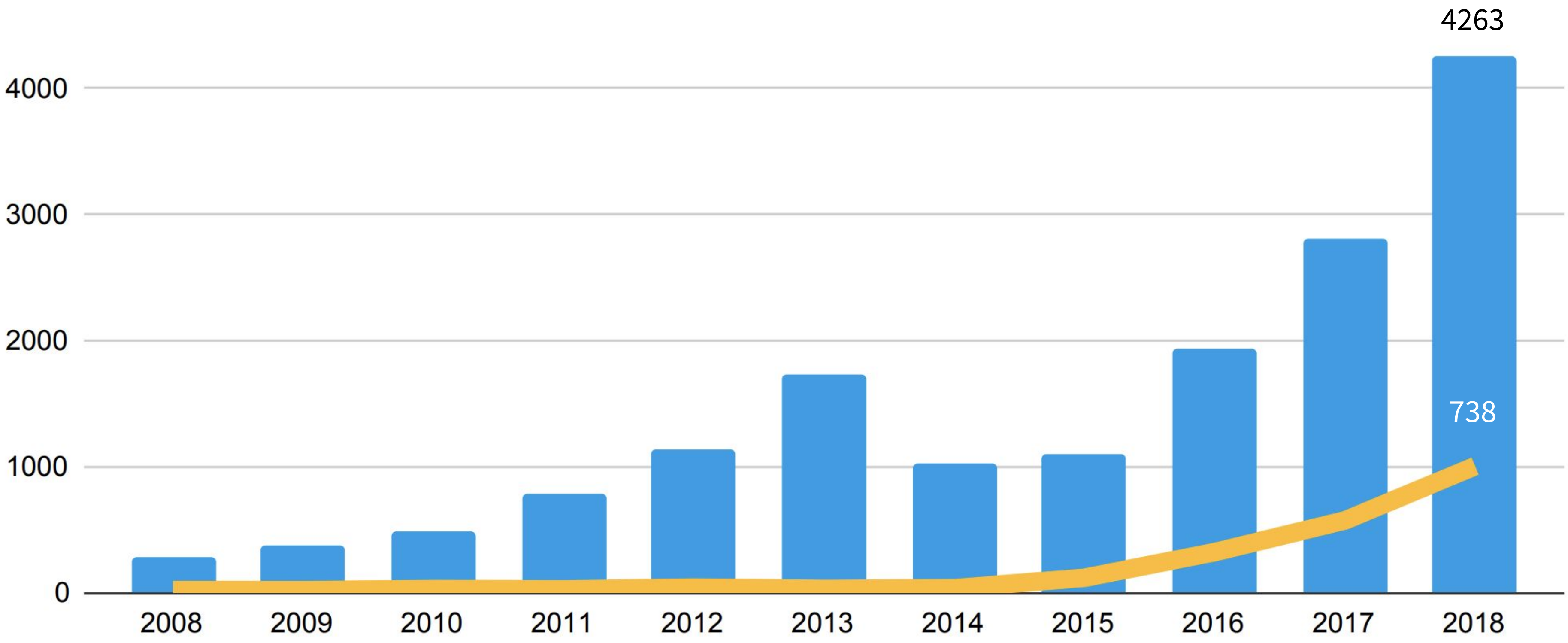
Opportunities

Excitement

Number of AI papers on arXiv

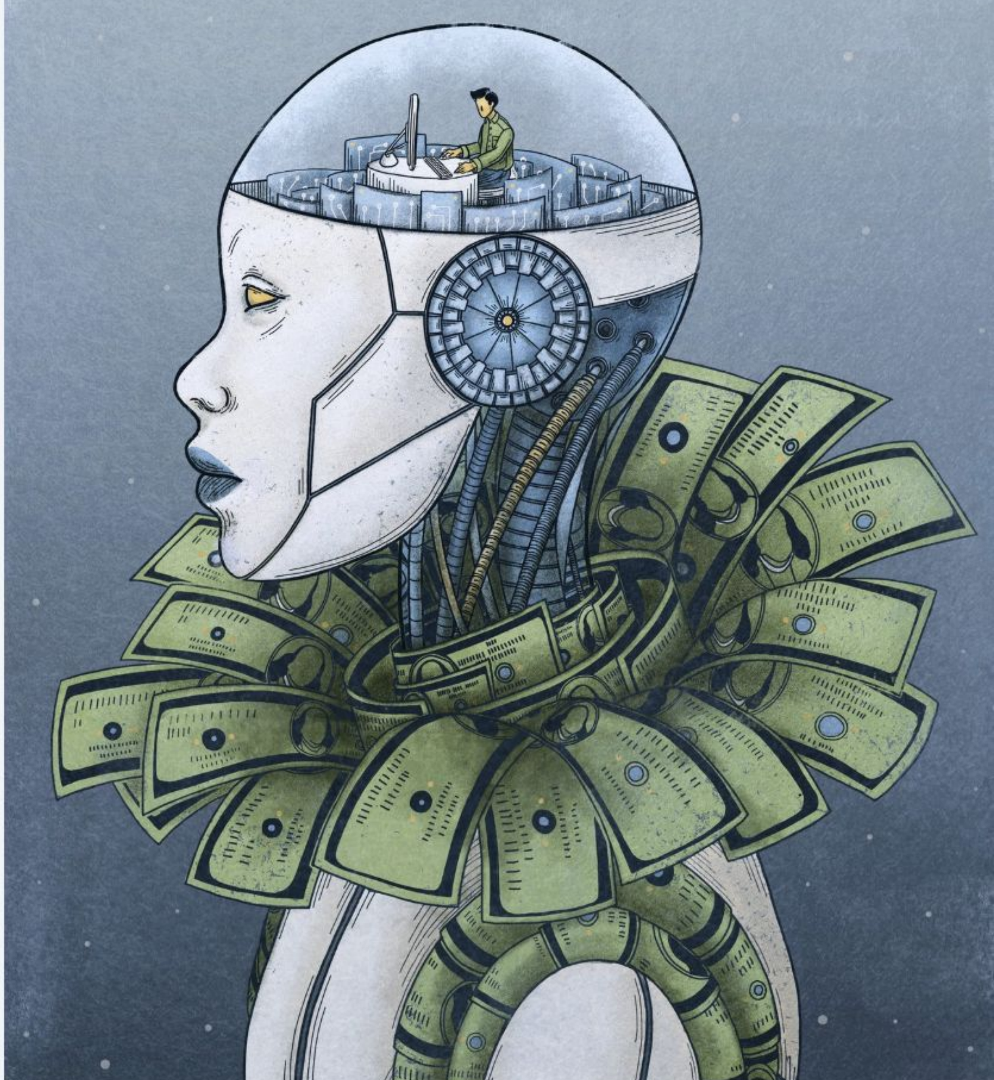


AI papers that include the term "neural network"



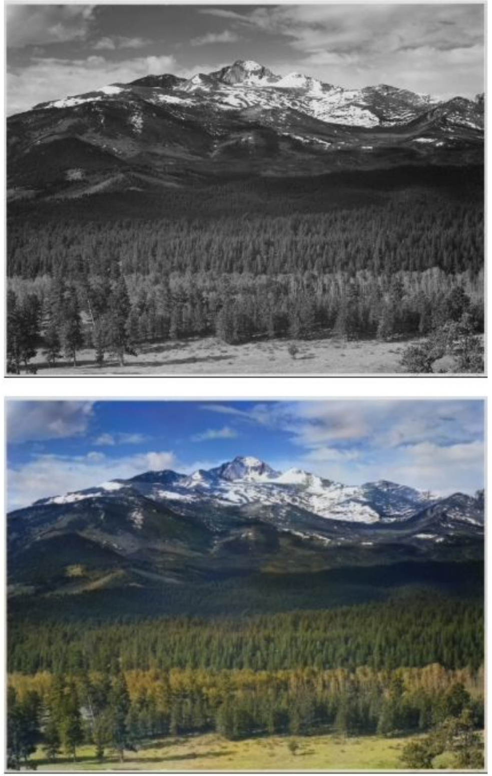
Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent

Nearly all big tech companies have an artificial intelligence project, and they are willing to pay experts millions of dollars to help get it done.



Applications

Colorizing images



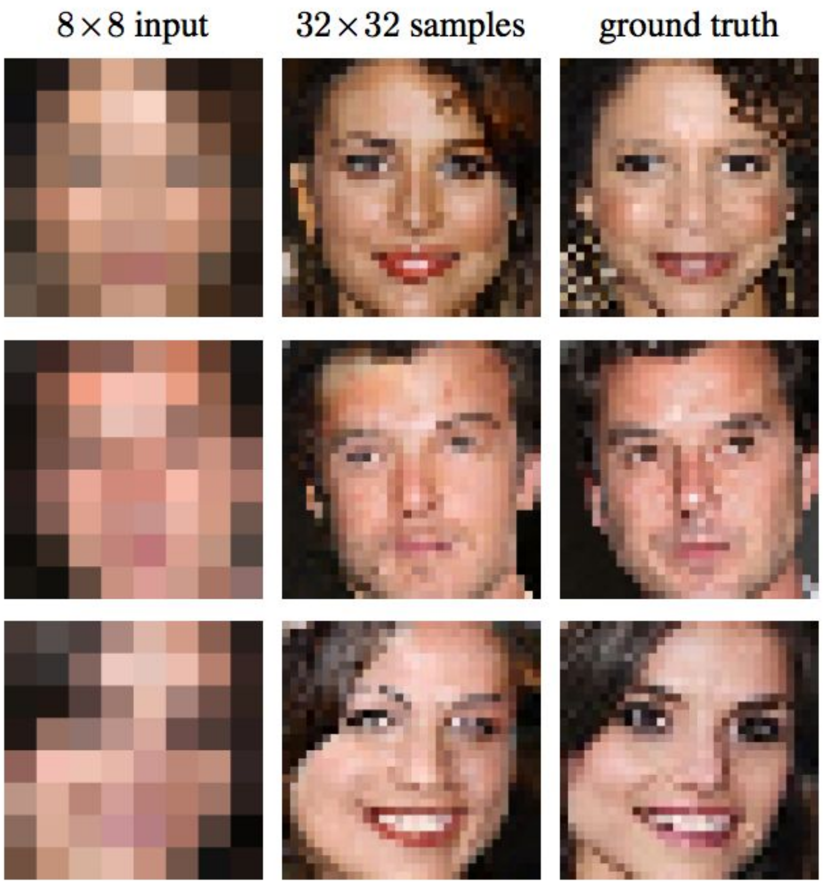
Colorado National Park, 1941

Textile Mill, June 1937

Berry Field, June 1909

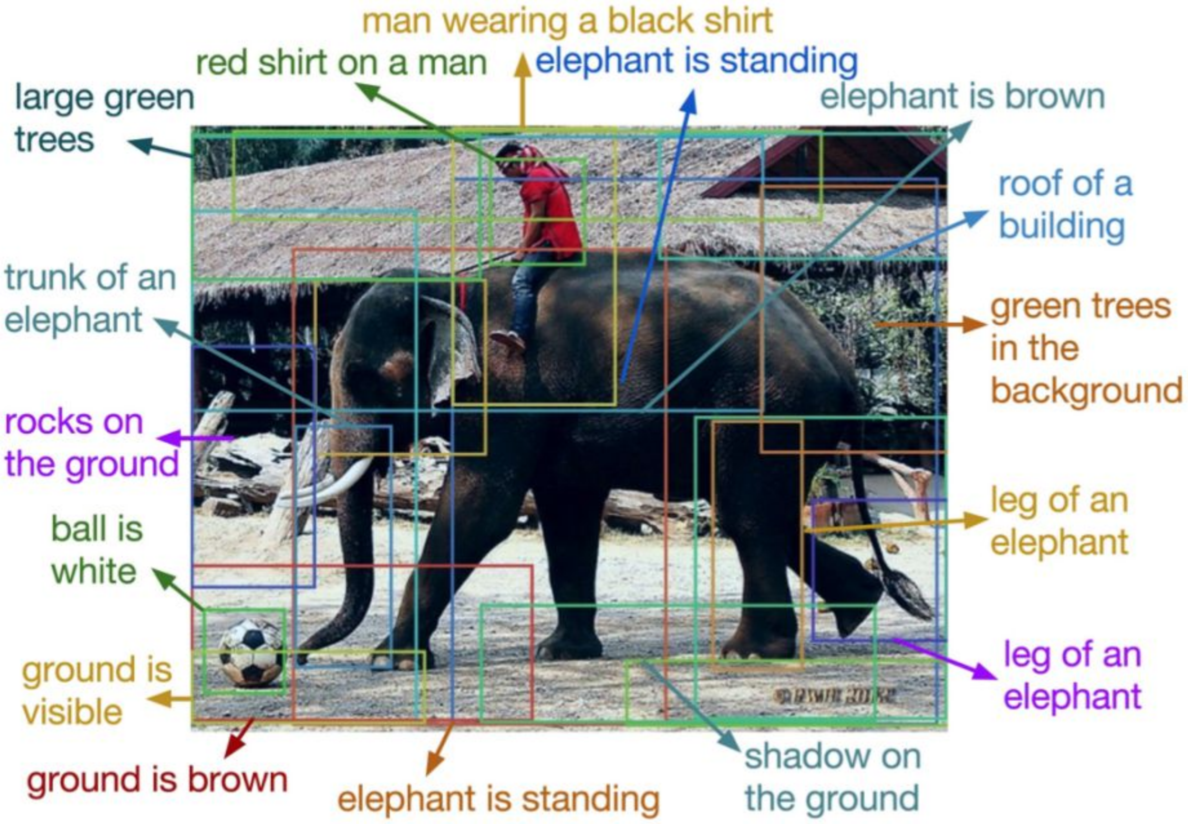
Hamilton, 1936

Restoring pixels



Restoring pixels

Understanding images



Translating text in real time

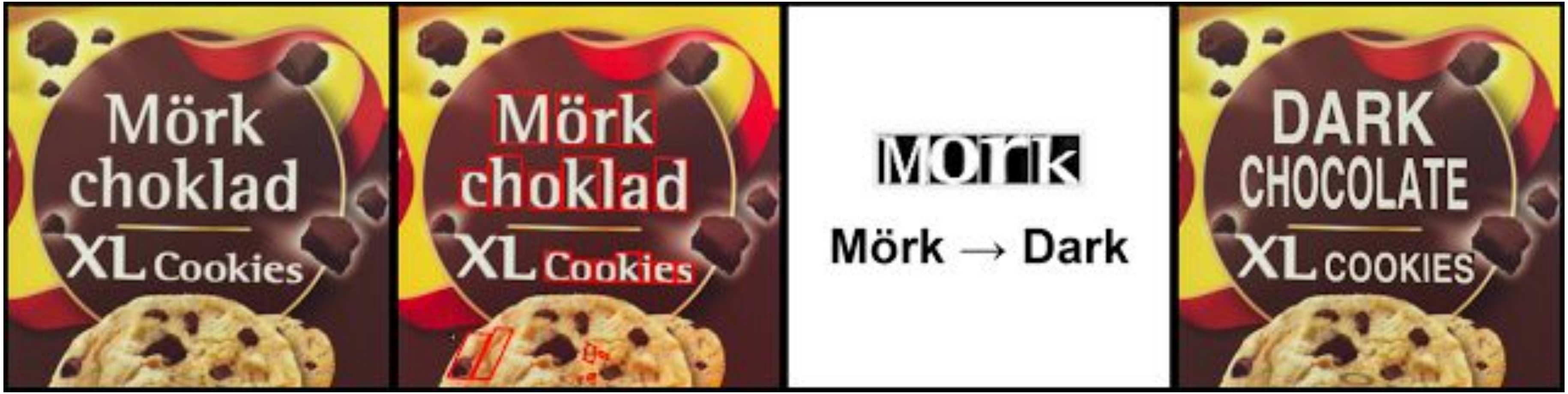


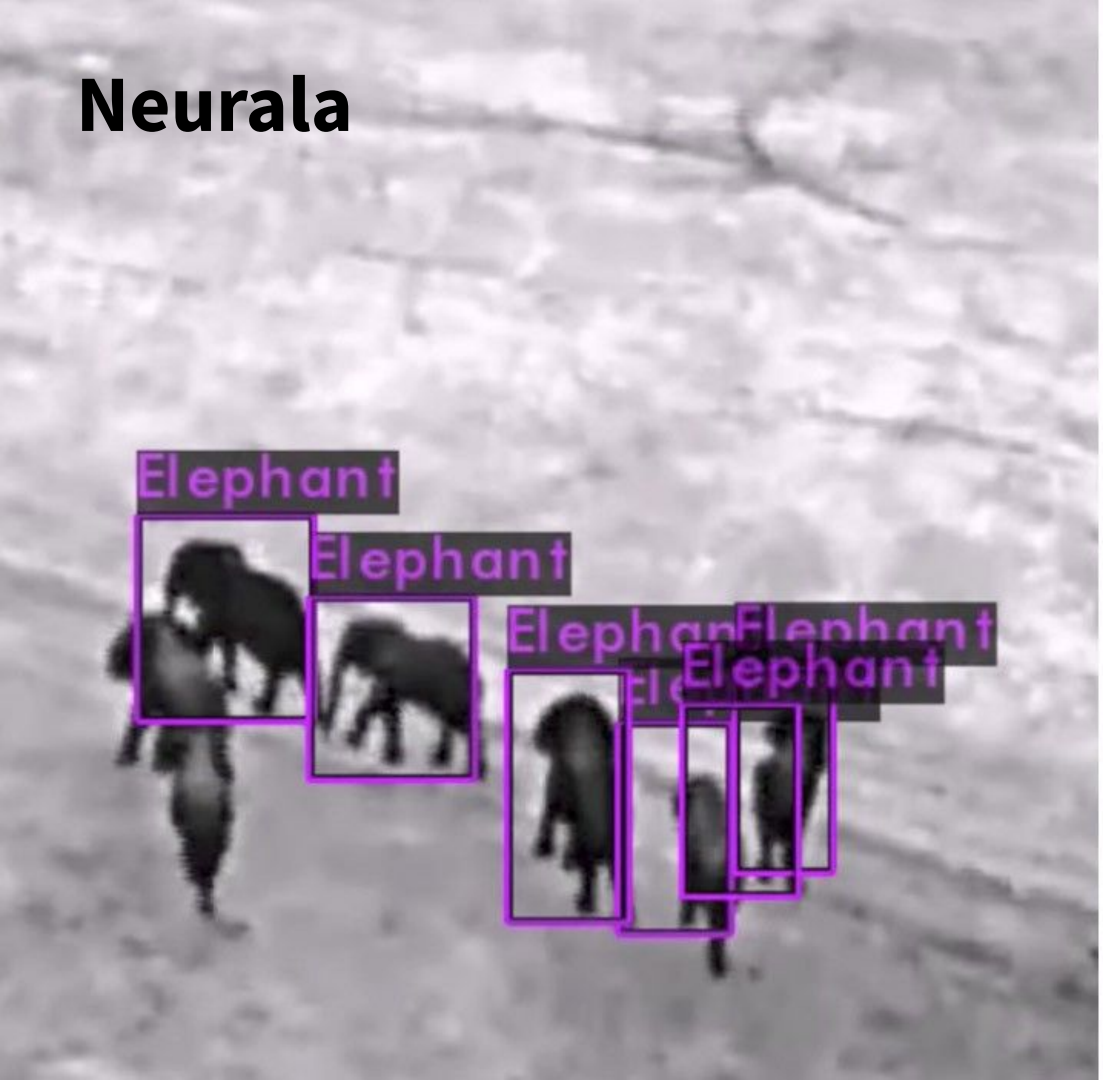
Image translation

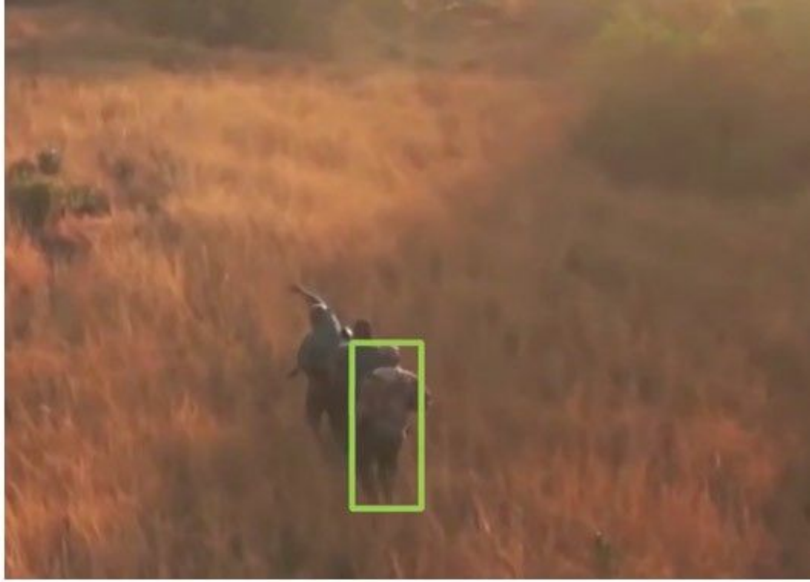
Input Ground Truth Output

wandb.com

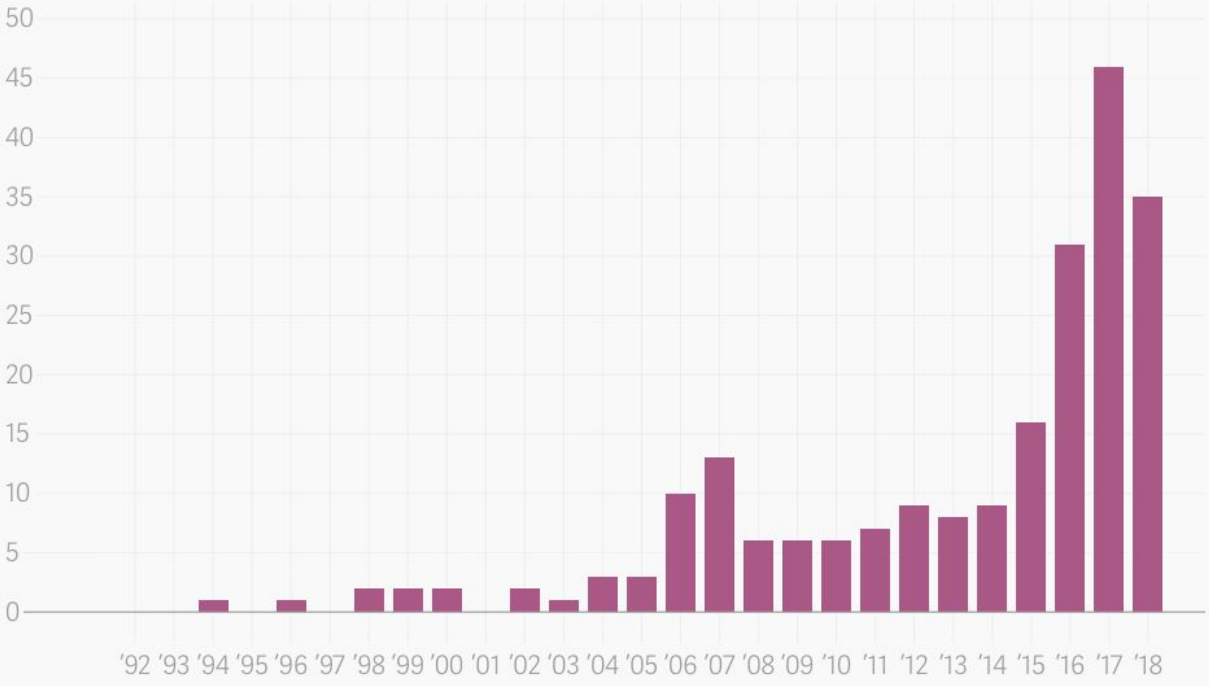
Shelf-Checking Robots To Roam The Aisles At Schnucks Supermarkets

Neurala





Lithium-ion battery incidents of fire or explosion on US airplanes



X-RAY VISION

Deep-learning algorithms are being used to detect lithium-ion batteries in airport luggage

By Akshat Rathi · September 2, 2018


Blue River Technology


Blue River Technology

Solving Rubik's Cube with a Robot Hand

OCTOBER 15, 2019 • 8 MINUTE READ

nature

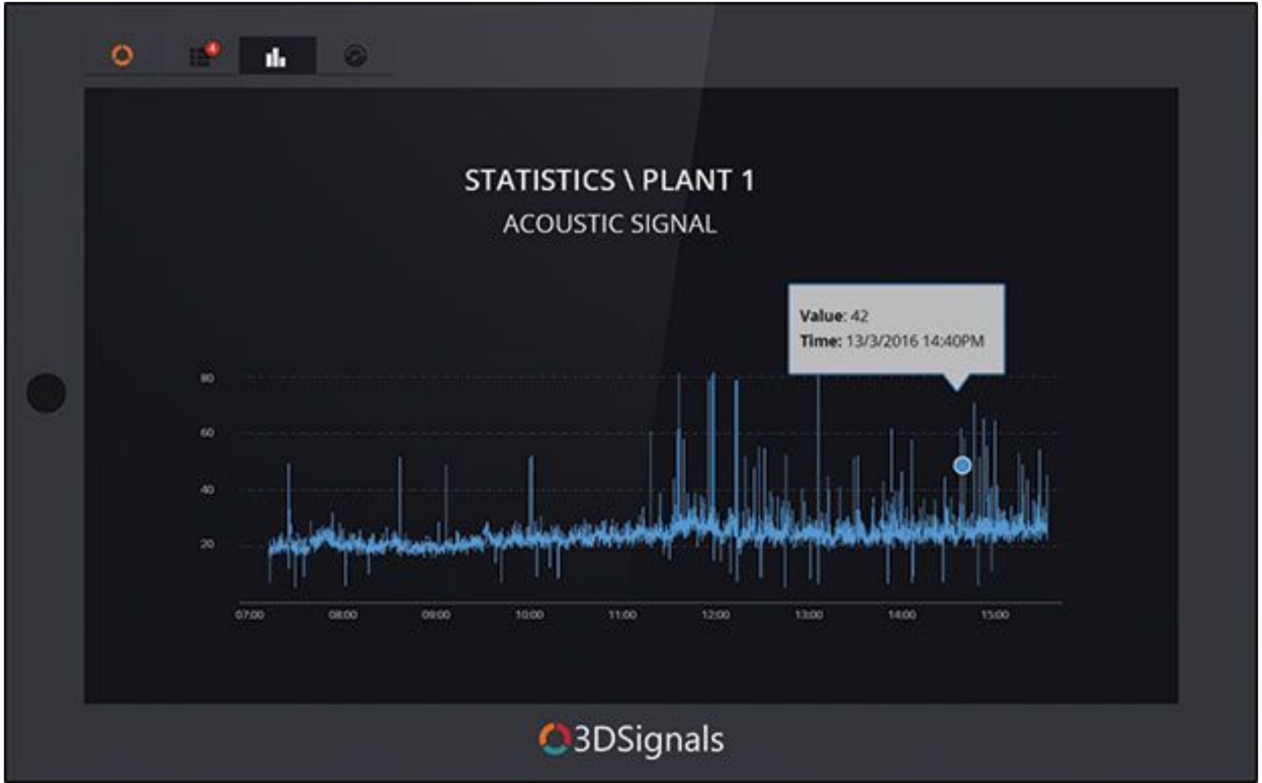
Letter | Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Nature **542**, 115–118 (02 February 2017) | Download Citation

Deep Learning AI Listens to Machines For Signs of Trouble



MACHINE INTELLIGENCE 3.0

ENTERPRISE INTELLIGENCE

VISUAL
Orbital Insight, planet, clarifai, DEEP VISION, cortica, Gigocion, SPACE_KNOW, Captricity, netra, deepomatic

AUDIO
Gridspace, TalkIQ, nexidia, twilio, CAPIO, Expect Labs, Clover, Mobvoi, Qurious.AI, popup archive

SENSOR
PREDIX, C3 IoT, MAANA, Sentenai, PLANET OS, UPTAKE, IMUBIT, Preferred Networks, thingworx, KONUX, Alluvium

INTERNAL DATA
PRIMER, IBM WATSON, Cycorp, Palantir, ARIMO, Alation, sapho, Outlier, Digital Reasoning

MARKET
mattermark, Quid, DataFox, PREMISE, Bottlenose, MOTIVA, enigma, CB INSIGHTS, Tracxn, predata

ENTERPRISE FUNCTIONS

CUSTOMER SUPPORT
DigitalGenius, Kasisto, ELOQUENT, Wiseio, ACTIONIQ, zendesk, Preact, CLARABRIDGE

SALES
collective[i], 6sense, fuse|machines, AVISO, salesforce, INSIDE SALES .COM, clari, Zensight

MARKETING
MINTIGO, Lattice, RADIUS, Liftigniter, [PERSADO], brightfunnel, retention science, COGNICOR, AIRPR, msg ai

SECURITY
CYLANCE, DARKTRACE, ZIMPERIUM, deepinstinct, Sentinel, DEMISTO, graphistry, drawbridge, SignalSense, AppZen

RECRUITING
textio, entelo, Wade & Wendy, hiQ, unitive, SpringRole, GIGSTER, Hire Vue

AUTONOMOUS SYSTEMS

GROUND NAVIGATION
drive.ai, AdasWorks, zoox, MOBILEYE, UBER, Google, TESLA, nuTonomy, Auro Robotics

AERIAL
SKYDIO, SHIELD AI, Airware, dji, LILY, DroneDeploy, pilot ai, SKYCATCH

INDUSTRIAL
JAYBRIDGE, OSARO, CLEARPATH ROBOTICS, fetch robotics, KINDRED, HARVEST AUTOMATION, rethink robotics

AGENTS

PERSONAL
amazon alexa, Cortana, Allo, facebook M, Siri, Replika

PROFESSIONAL
butter.ai, pogo, SKIPFLAG, clara, x.ai, slack, talla, Zoom.ai, sudo

INDUSTRIES

AGRICULTURE
BLUE RIVER, mavrx, tule, TRACE GENOMICS, Pivot Bio, TerrAvion, AGRI-DATA, Descartes Labs, udio, abundant robotics

EDUCATION
KNEWTON, volley, gradescope, CTI, coursera, UDACITY, altschool

INVESTMENT
Bloomberg, sentient, iSENTIUM, KENSHO, alphasense, Dataminr, CEREBELLUM CAPITAL, Quandl

LEGAL
blue J, BEAGLE, Everlaw, RAVEL, seal, ROSS, LEGAL ROBOT

LOGISTICS
NAUTO, Acerta, PRETECKT, clearmetal, Routific, MARBLE, PITSTOP

INDUSTRIES CONT'D

MATERIALS
zymergen, Citrine, Eigen Innovations, SIGHT MACHINE, GINKGO BIOWORKS, nanotronics, CALCULARIO

RETAIL FINANCE
TALA, zest finance, Lendo, earnest, affirm, MIRADOR, wealthfront, Betterment

HEALTHCARE

PATIENT
PULSE, CareSkore, ZEPHYR HEALTH, IBM Watson Health, Oncora, SENTRIAN, Atomwise, Numerate

IMAGE
BUTTERFLY, 3SCAN, ARTERYS, enlitic, BAYLABS, imagia, Google DeepMind

BIOLOGICAL
iCarbonX, color, GRAIL, deep genomics, RECURSION, LUMINIST, Numerate, Atomwise, verily, WHOLE BIOME

shivonzilis.com/MACHINEINTELLIGENCE · Bloomberg BETA

TECHNOLOGY STACK

AGENT ENABLERS
OCTANE.AI, howdy, Maluuba, KITT.AI, OpenAI Gym, Kasisto, AUTOMAT, semanticmachines

DATA SCIENCE
DOMINO, SPARKBEYOND, rapidminer, kaggle, DataRobot, yhat, AYASDI, data iku, seldon, yseop, bigml

MACHINE LEARNING
CognitiveScale, GoogleML, context relevant, Cycorp, HyperScience, naralogics, minds.ai, H2O.ai, SCALED INFERENCE, sparkcognition, loop, GEOMETRIC INTELLIGENCE, deepsense.io, reactive, skymind, bonsai

NATURAL LANGUAGE
agolo, AYLIEN, LEXALYTICS, Narrative Science, loop ai labs, spaCy, LUMINOSO, cortical.io, MonkeyLearn

DEVELOPMENT
SIGOPT, HyperOpt, fuzzy.io, kite, rainforest, lobe, Anodot, Signifai, LAYER 6 AI, bonsai

DATA CAPTURE
CrowdFlower, diffbot, CrowdAI, import.io, Paxata, DATASIFT, amazon mechanical turk, enigma, WorkFusion, DATALOGUE, TRIFACTA, parsehub

OPEN SOURCE LIBRARIES
Keras, Chainer, CNTK, TensorFlow, Caffe, H20, DEEPLEARNING4J, theano, torch, DSSTNE, Scikit-learn, AzureML, neon, MXNet, DMTK, Spark, PaddlePaddle, WEKA

HARDWARE
KNUPATH, TENSTORRENT, Cirrascale, NVIDIA, intel nervana, Movidius, tensilica, GoogleTPU, 10^{26} Labs, Qualcomm, Cerebras, Isosemi

RESEARCH
OpenAI, nnaisense, ELEMENT AI, vicarious, KNOGGIN, Numenta, Kimera Systems, Cogitai

LINKS FROM THIS TALK

bit.ly/ucsd-ml

Elon Musk Promises a Really Truly Self-Driving Tesla in 2020

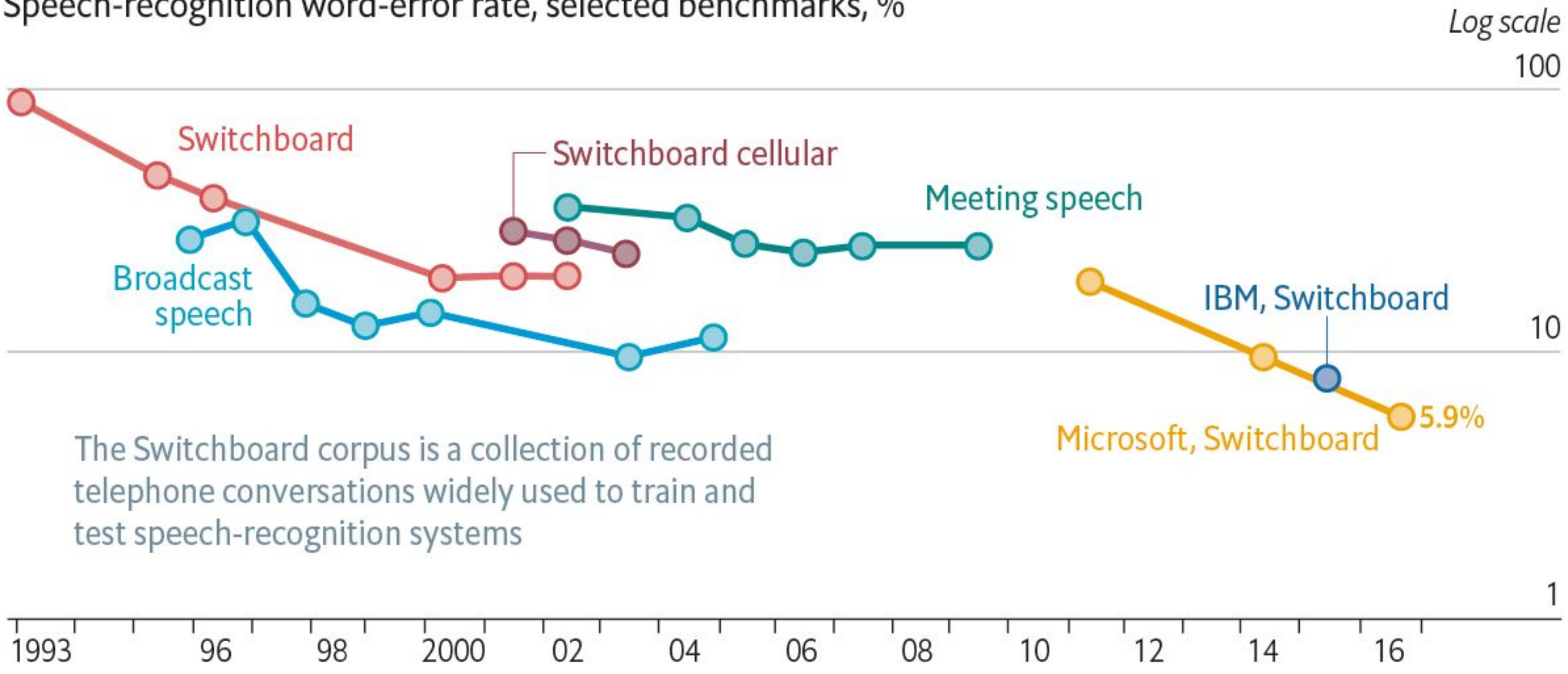
Wired article, February 2019


LEFT REARWARD VEHICLE CAMERA
MEDIUM RANGE VEHICLE CAMERA
RIGHT REARWARD VEHICLE CAMERA

Natural Language

Loud and clear

Speech-recognition word-error rate, selected benchmarks, %

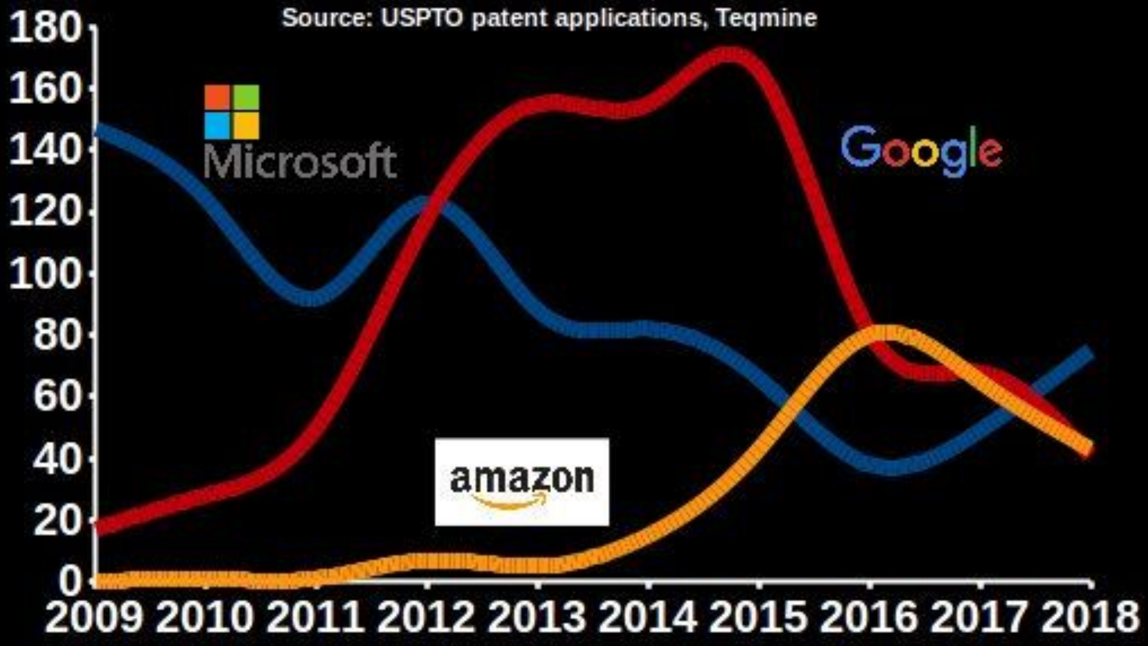


Sources: Microsoft; research papers

MICROSOFT, AMAZON & GOOGLE PATENTING
Natural Language Processing

TOPIC 6:

Topics created by
Teqmine AI
from full-text patents



TEQMINE

GLUE

2018

ML²

NYU

UWNLP

DeepMind

SuperGLUE

1 year later

ML² NYU

UWNLP

facebook Artificial Intelligence

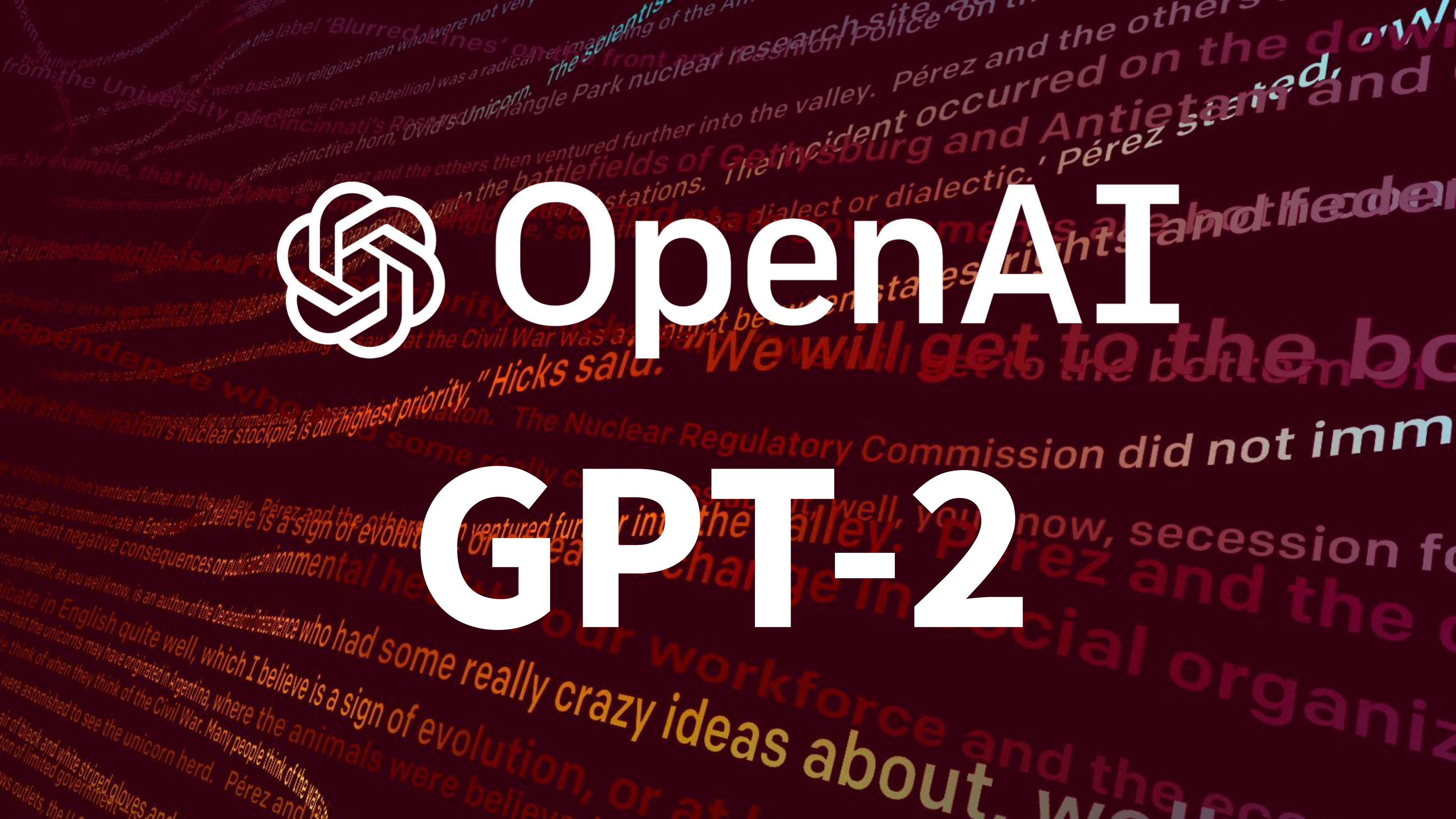
DeepMind

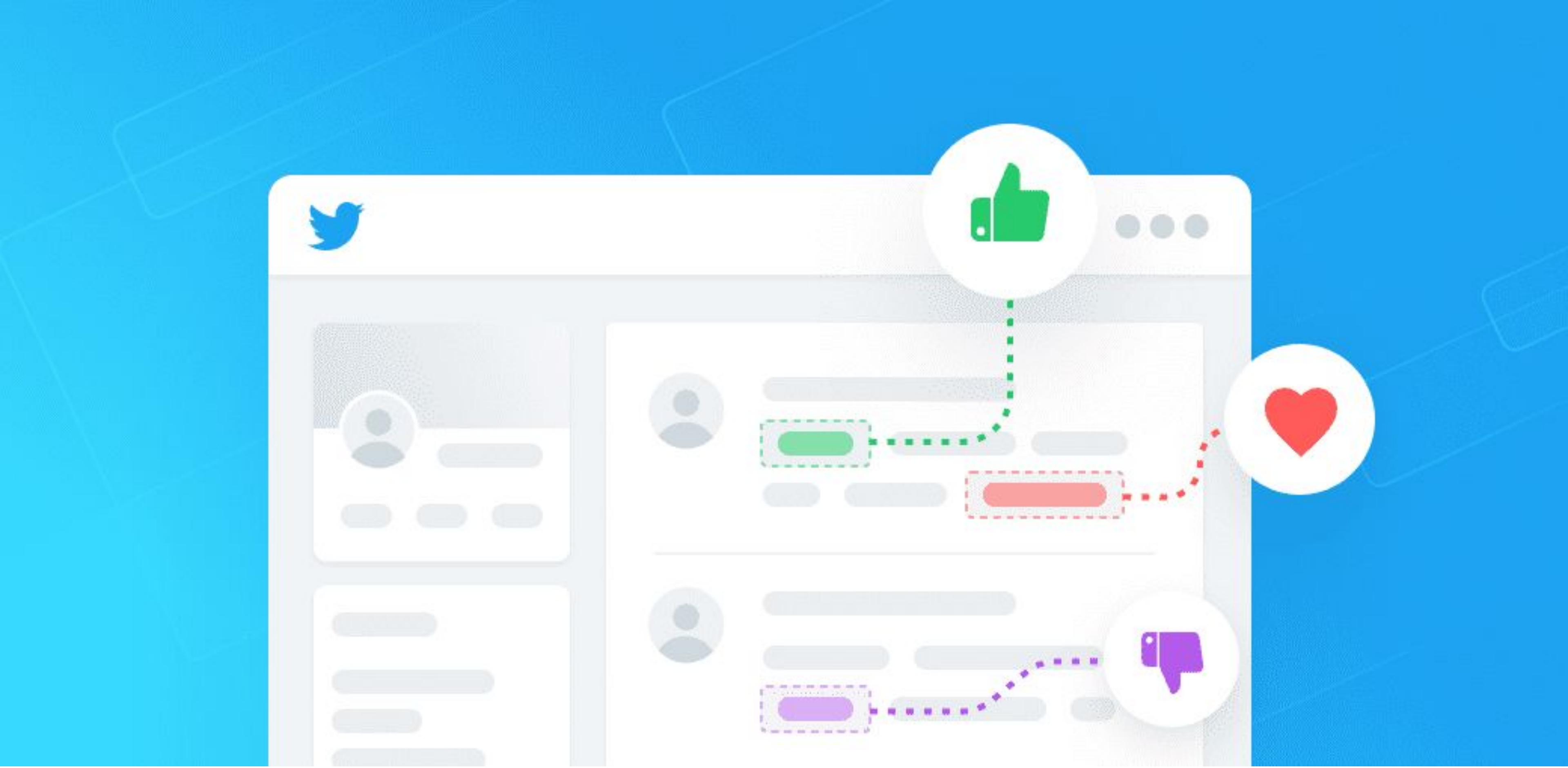
5 months later...

Leaderboard Version: **2.0**

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-b	AX-g
1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	76.6	99.3/99.7
2	T5 Team - Google	T5		88.9	91.0	93.0/96.4	94.8	88.2/62.3	93.3/92.5	92.5	76.1	93.8	65.6	92.7/91.9
3	Facebook AI	RoBERTa		84.6	87.1	90.5/95.2	90.6	84.4/52.5	90.6/90.0	88.2	69.9	89.0	57.9	91.0/78.1
4	IBM Research AI	BERT-mtl		73.5	84.8	89.6/94.0	73.8	73.2/30.5	74.6/74.0	84.1	66.2	61.0	29.6	97.8/57.3
5	SuperGLUE Baselines	BERT++		71.5	79.0	84.8/90.4	73.8	70.0/24.1	72.0/71.3	79.0	69.6	64.4	38.0	99.4/51.4
		BERT		69.0	77.4	75.7/83.6	70.6	70.0/24.1	72.0/71.3	71.7	69.6	64.4	23.0	97.8/51.7
		Most Frequent Class		47.1	62.3	21.7/48.4	50.0	61.1/0.3	33.4/32.5	50.3	50.0	65.1	0.0	100.0/50.0
		CBoW		44.5	62.2	49.0/71.2	51.6	0.0/0.5	14.0/13.6	49.7	53.1	65.1	-0.4	100.0/50.0
		Outside Best		-	80.4	-	84.4	70.4/24.5	74.8/73.0	82.7	-	-	-	-

wandb.com


OpenAI
GPT-2



Twitter sentiment analysis

wandb.com

Cognitive Systems Research

Volume 54, May 2019, Pages 50-61

Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information

Ahmed Sulaiman M. Alharbi [a, b] ... Elise de Doncker [a]

AMERICAN JOURNAL OF
Preventive Medicine

A Practical Approach for Content Mining of Tweets

Sunmoo Yoon, RN, PhD[a,b,*], Noémie Elhadad, PhD[b], Suzanne Bakken, RN, PhD[a,b]

Content mining tweets

Table 1

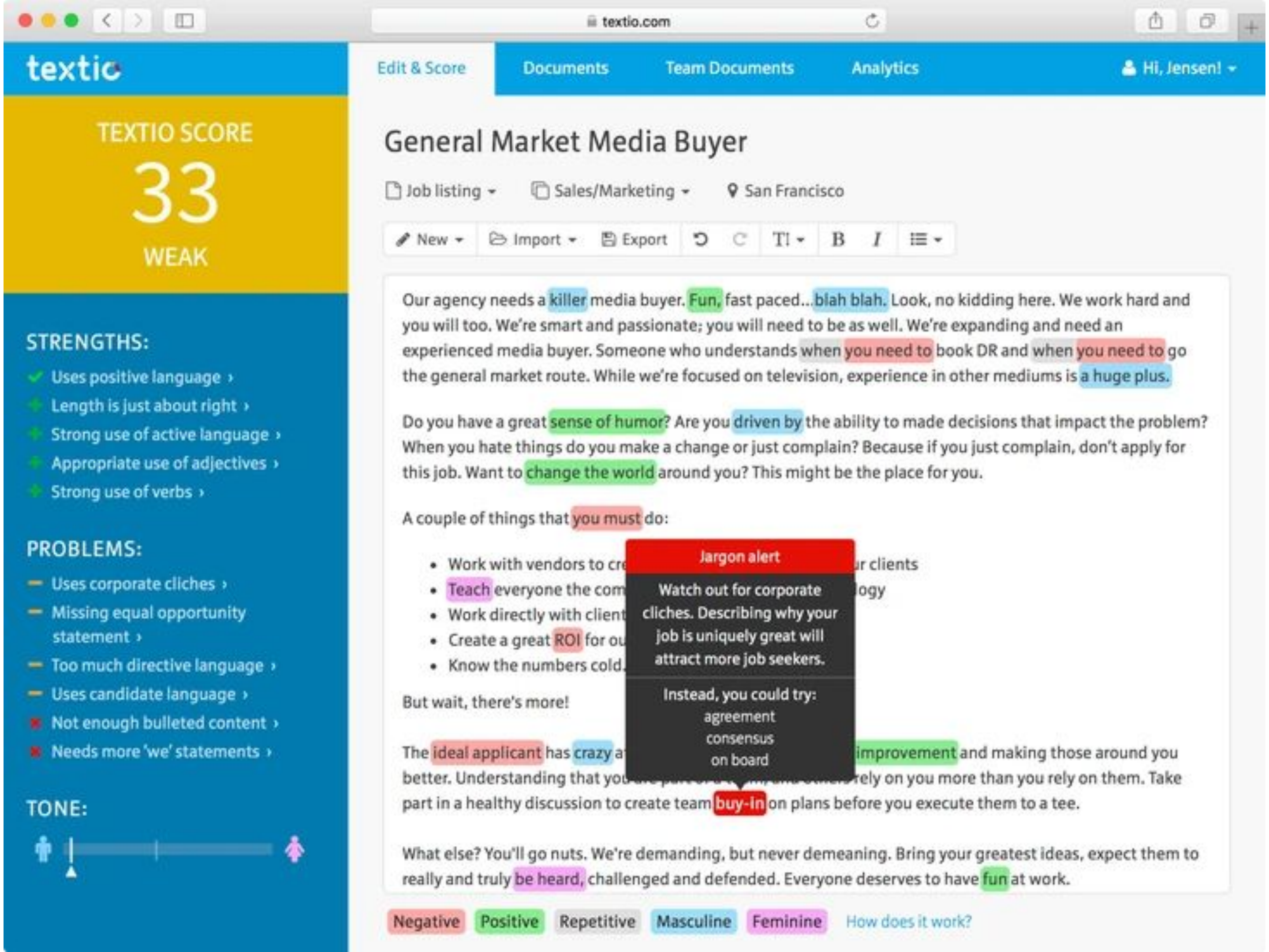
A Sample of physical activity Tweets term-frequency dictionary from the corpus of 174,394 Tweets containing 31,489 terms

Terms	**weight**	**day**	**good**	**time**	**obesity**	**study**
aerobics	0.0229	0.0349	0.0331	0.0335	0	0.0026
bicycling	0.0062	0.0283	0.0182	0.0215	0	0.0024
chop wood	0	0.0705	0.0529	0.0353	0	0

Table 2

Top three unigram, bigram, and trigram for each of the physical activity terms

Term	Unigram	Bigram	Trigram
aerobics	Class	aerobic class	water aerobics class
	Water	water aerobics	minutes doing aerobics
	Step	calories burned	my aerobics class
bicycling	Google	Google maps	bike-friendly cities
	maps	bicycling magazine	Google maps adds
	cities	bike friendly	50 bike friendly
chop wood	fire	i chopped	I chopped wood
	free	of wood	chopped some wood
	today	wood for	chopped wood for


textio.com
textio
Edit & Score
Documents
Team Documents
Analytics
Hi, Jensen!
TEXTIO SCORE
33
WEAK
STRENGTHS:
Uses positive language ›
Length is just about right ›
Strong use of active language ›
Appropriate use of adjectives ›
Strong use of verbs ›
PROBLEMS:
Uses corporate cliches ›
Missing equal opportunity statement ›
Too much directive language ›
Uses candidate language ›
Not enough bulleted content ›
Needs more 'we' statements ›
TONE:
General Market Media Buyer
Job listing
Sales/Marketing
San Francisco
New
Import
Export
Our agency needs a killer media buyer. Fun, fast paced...blah blah. Look, no kidding here. We work hard and you will too. We're smart and passionate; you will need to be as well. We're expanding and need an experienced media buyer. Someone who understands when you need to book DR and when you need to go the general market route. While we're focused on television, experience in other mediums is a huge plus.
Do you have a great sense of humor? Are you driven by the ability to made decisions that impact the problem? When you hate things do you make a change or just complain? Because if you just complain, don't apply for this job. Want to change the world around you? This might be the place for you.
A couple of things that you must do:
Jargon alert
Watch out for corporate cliches. Describing why your job is uniquely great will attract more job seekers.
Instead, you could try:
agreement
consensus
on board
But wait, there's more!
part in a healthy discussion to create team buy-in on plans before you execute them to a tee.
What else? You'll go nuts. We're demanding, but never demeaning. Bring your greatest ideas, expect them to really and truly be heard, challenged and defended. Everyone deserves to have fun at work.
Negative
Positive
Repetitive
Masculine
Feminine
How does it work?

LINKS FROM THIS TALK

bit.ly/ucsd-ml

1

Excitement

2

Problems

3

Opportunities

Problems

Elon Musk Promises a Really Truly Self-Driving Tesla in 2020

Wired article, February 2019

Tesla CEO Elon Musk drops his prediction of full autonomous driving from 3 years to just 2

Fred Lambert - Dec. 21st 2015 3:26 pm ET @FredericLambert

Tesla CEO Elon Musk drops his prediction of full autonomous driving from 3 years to just 2

Fred Lambert - Dec. 21st 2015 3:26 pm ET @FredericLambert

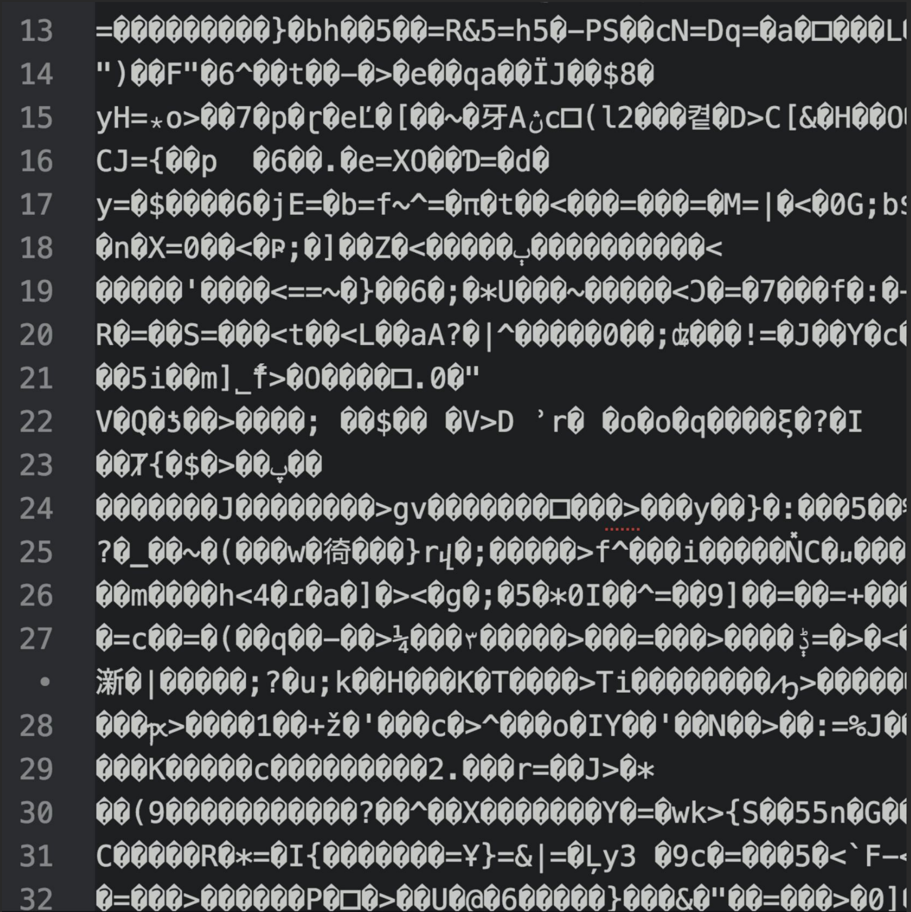
IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close

By CASEY ROSS @caseymross *and* IKE SWETLITZ / SEPTEMBER 5, 2017

Human Code

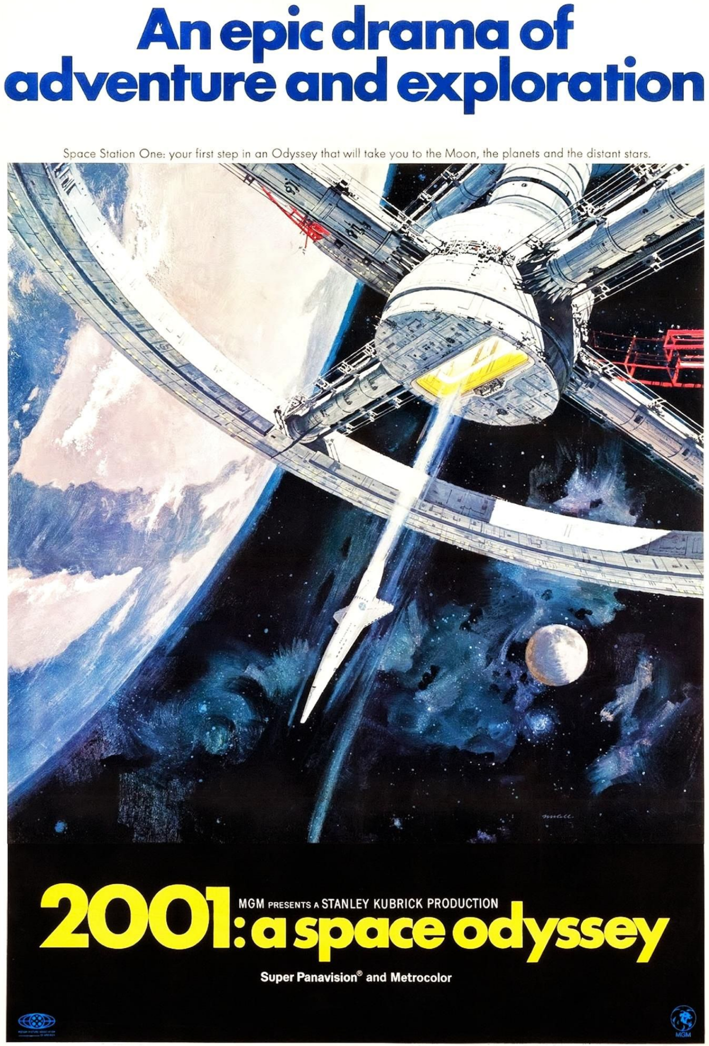
```python
from keras.datasets import mnist
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Dro
from keras.utils import np_utils

(X_train, y_train), (X_test, y_test) = mnist.load_

img_width=28
img_height=28

X_train = X_train.astype('float32')
X_train /= 255.
X_test = X_test.astype('float32')
X_test /= 255.

#reshape input data
X_train = X_train.reshape(X_train.shape[0], img_wi
X_test = X_test.reshape(X_test.shape[0], img_width

# one hot encode outputs
y_train = np_utils.to_categorical(y_train)
```

Machine Code



```diff
@@ -18,29 +17,120 @@
     "metadata": {
      "collapsed": false
     },
+    "outputs": [],
+    "source": [
+     "from bokeh.plotting import figure, output_notebook, show"
+    ]
+   },
+   {
+    "cell_type": "code",
+    "execution_count": 3,
+    "metadata": {
+     "collapsed": false
+    },
     "outputs": [
      {
       "data": {
        "text/html": [
         "\n",
-        "    \n",
-        "<script type=\"text/javascript\" src=\"https://cdn.pydata.org/bokeh/dev/bokeh-0.11.0dev9.min.js\"></script>\n",
-        "<script type=\"text/javascript\" src=\"https://cdn.pydata.org/bokeh/dev/bokeh-widgets-0.11.0dev9.min.js\"></script>\n",
-        "<script type=\"text/javascript\"
+        "    <div class=\"bk-root\">\n",
```

wandb.com


An epic drama of
adventure and exploration
Space Station One: your first step in an Odyssey that will take you to the Moon, the planets and the distant stars.
MGM PRESENTS A STANLEY KUBRICK PRODUCTION
2001: a space odyssey
Super Panavision® and Metrocolor


KASPAROV vs
DEEP BLUE
the rematch
1:57
1:58
AUDEMARS PIGUET
1997

We have bad intuitions about what is hard.

$20,000 • 530 teams

Search Results Relevance

Enter/Merge by

Mon 11 May 2015 Mon 6 Jul 2015 (37 days to go)

Dashboard

- Home
  - Data
  - Make a submission
- Information
  - Description
  - Evaluation
  - Rules
  - Prizes
  - Timeline
- Forum
- Scripts
- Leaderboard
- My Team

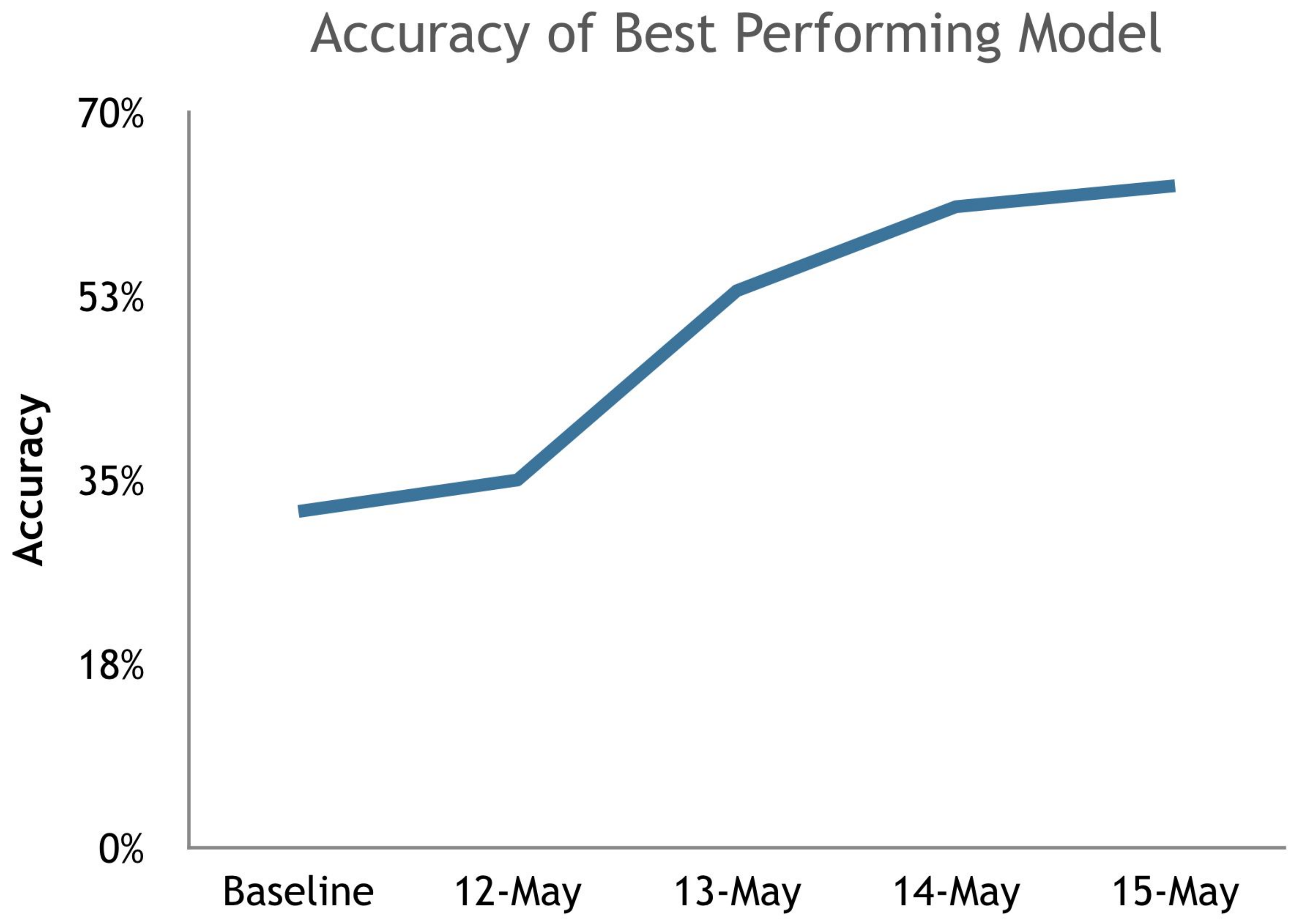
Competition Details » Get the Data » Make a submission

Data Files

File Name	Available Formats
sampleSubmission.csv	.zip (51.25 kb)
test.csv	.zip (4.51 mb)
train.csv	.zip (1.87 mb)

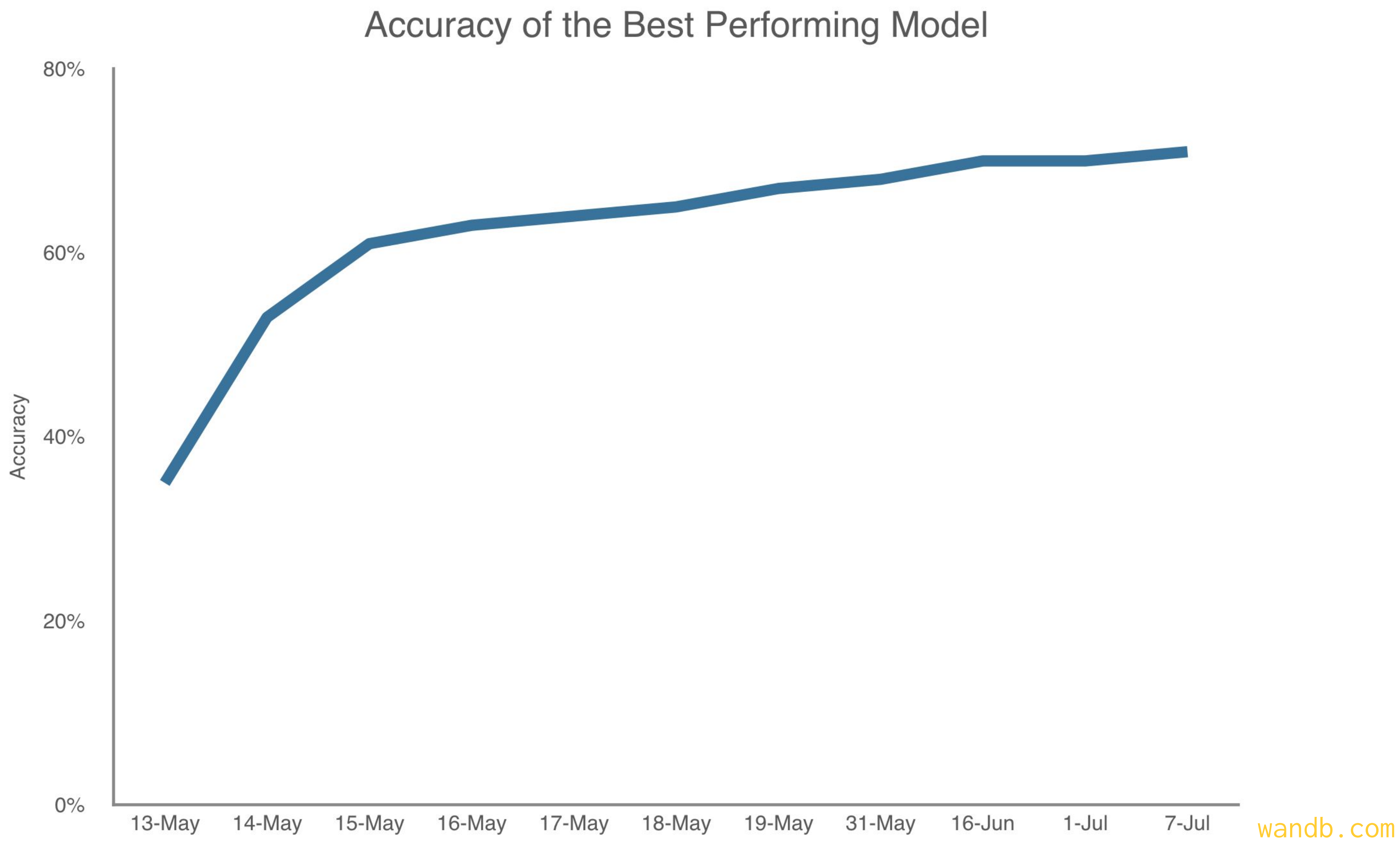
See this script for a quick exploration of the data

To evaluate search relevancy, CrowdFlower has had their crowd evaluate searches from a handful of eCommerce websites. A total of 261 search terms were generated,



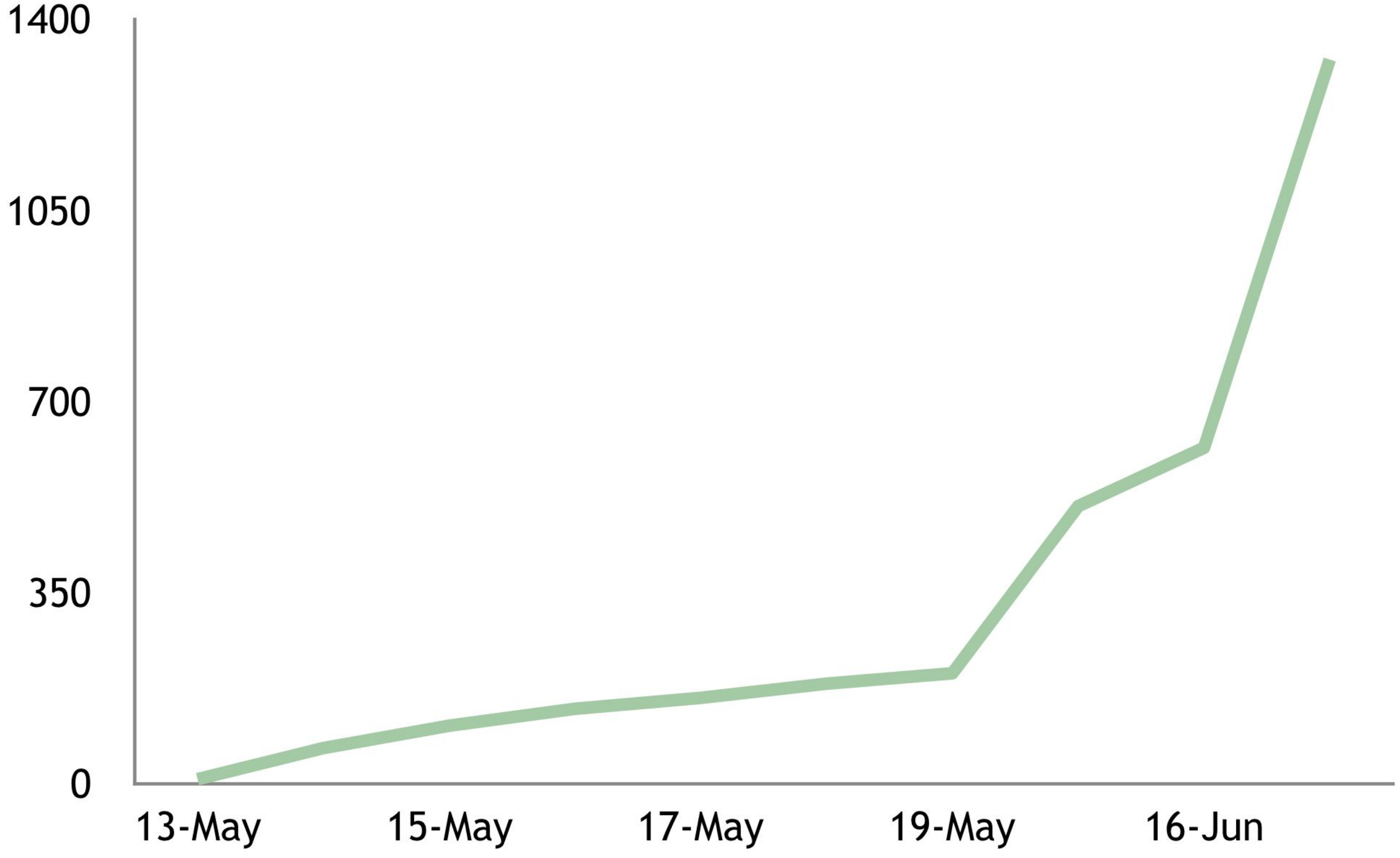
Accuracy of Best Performing Model
70%
53%
35%
18%
0%
Accuracy
Baseline
12-May
13-May
14-May
15-May

ındb.com



Accuracy of the Best Performing Model
80%
60%
40%
20%
0%
Accuracy
13-May
14-May
15-May
16-May
17-May
18-May
19-May
31-May
16-Jun
1-Jul
7-Jul
wandb.com

Number of Participating Teams



Exponentially more effort for vanishing returns

Netflix Prize

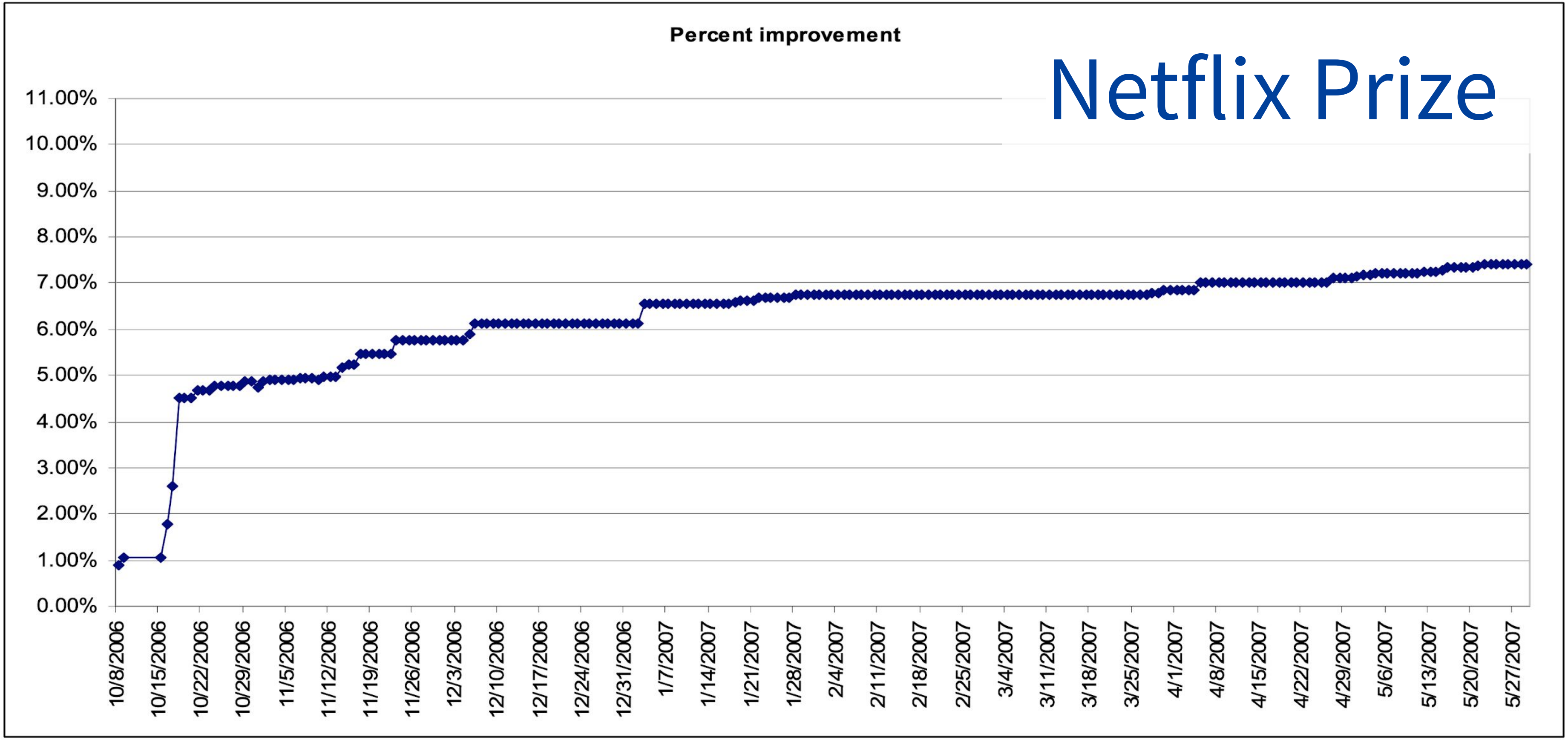
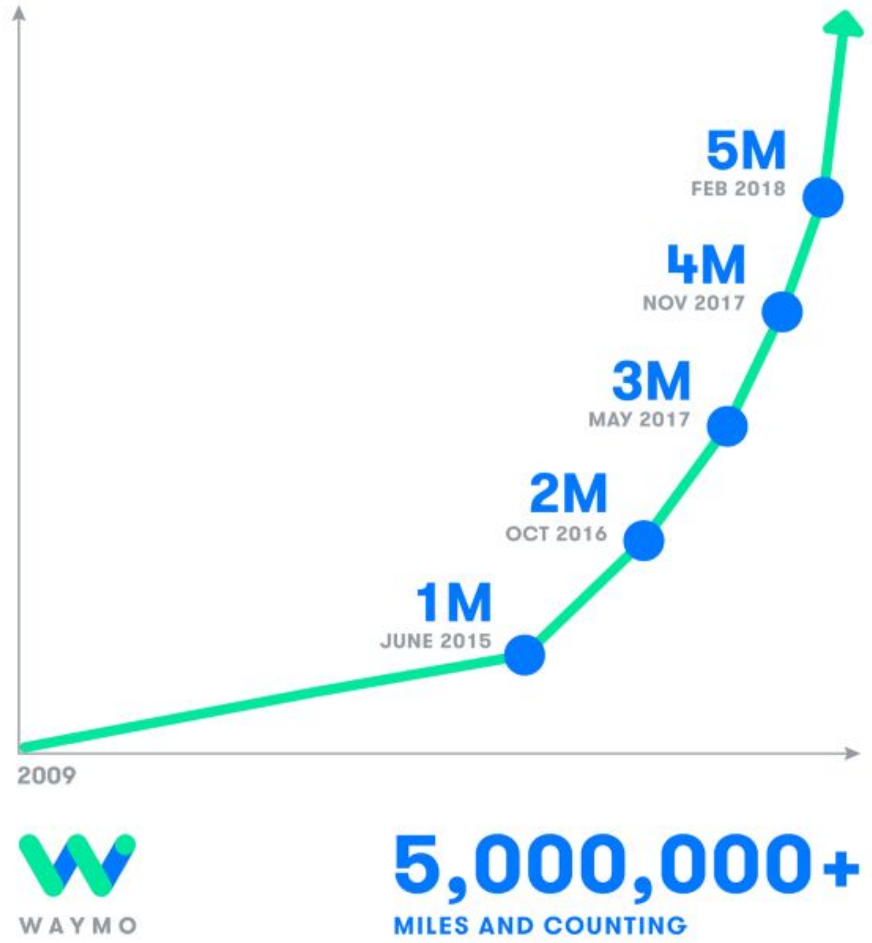
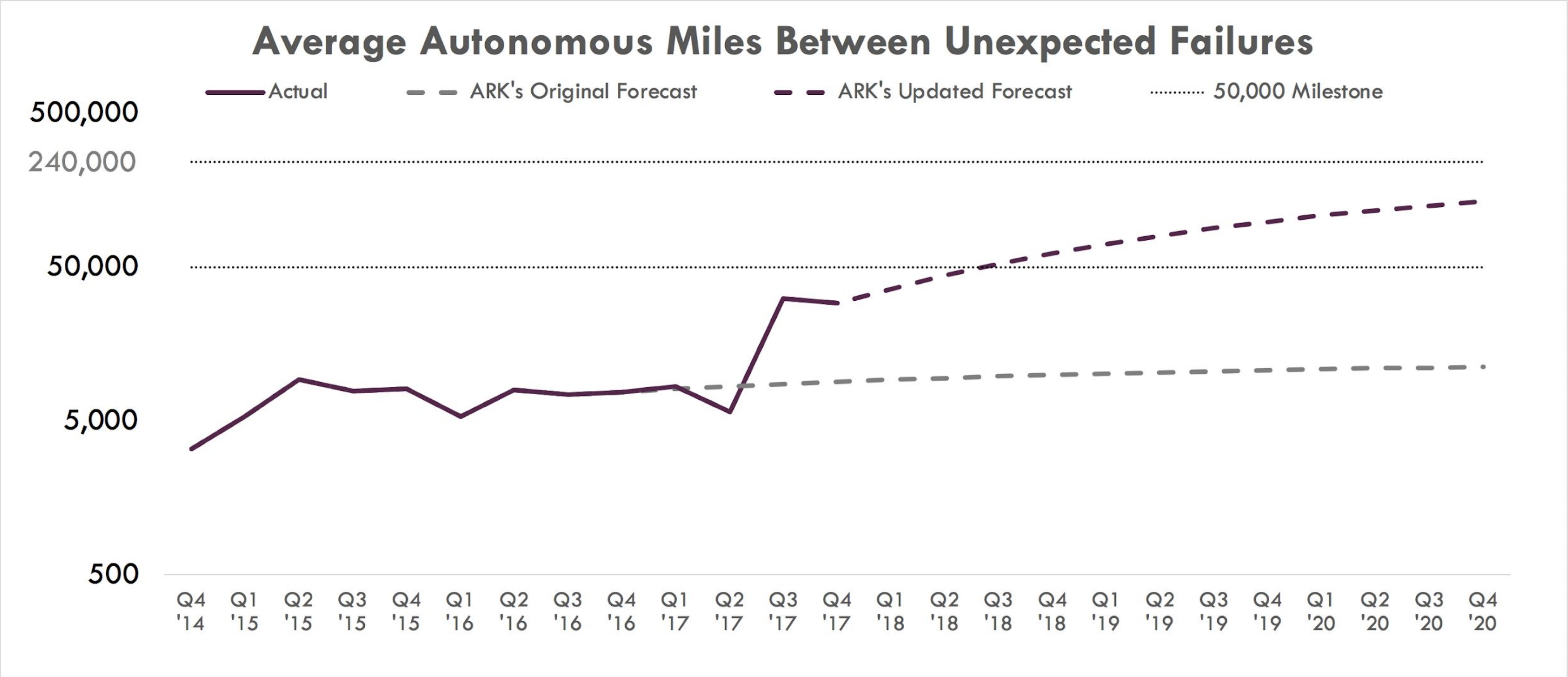


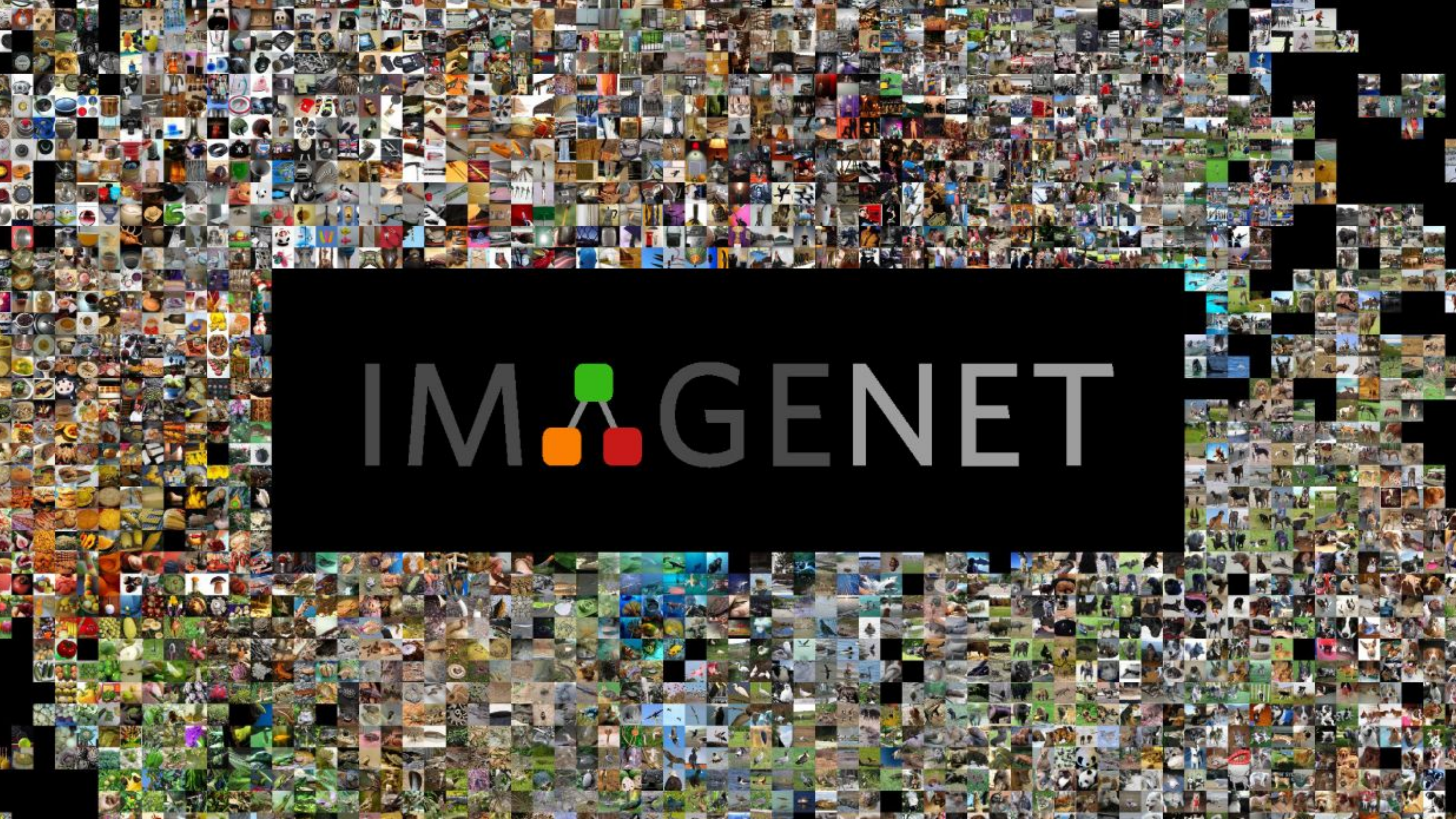
Figure 3: Aggregate improvement over Cinematch by time

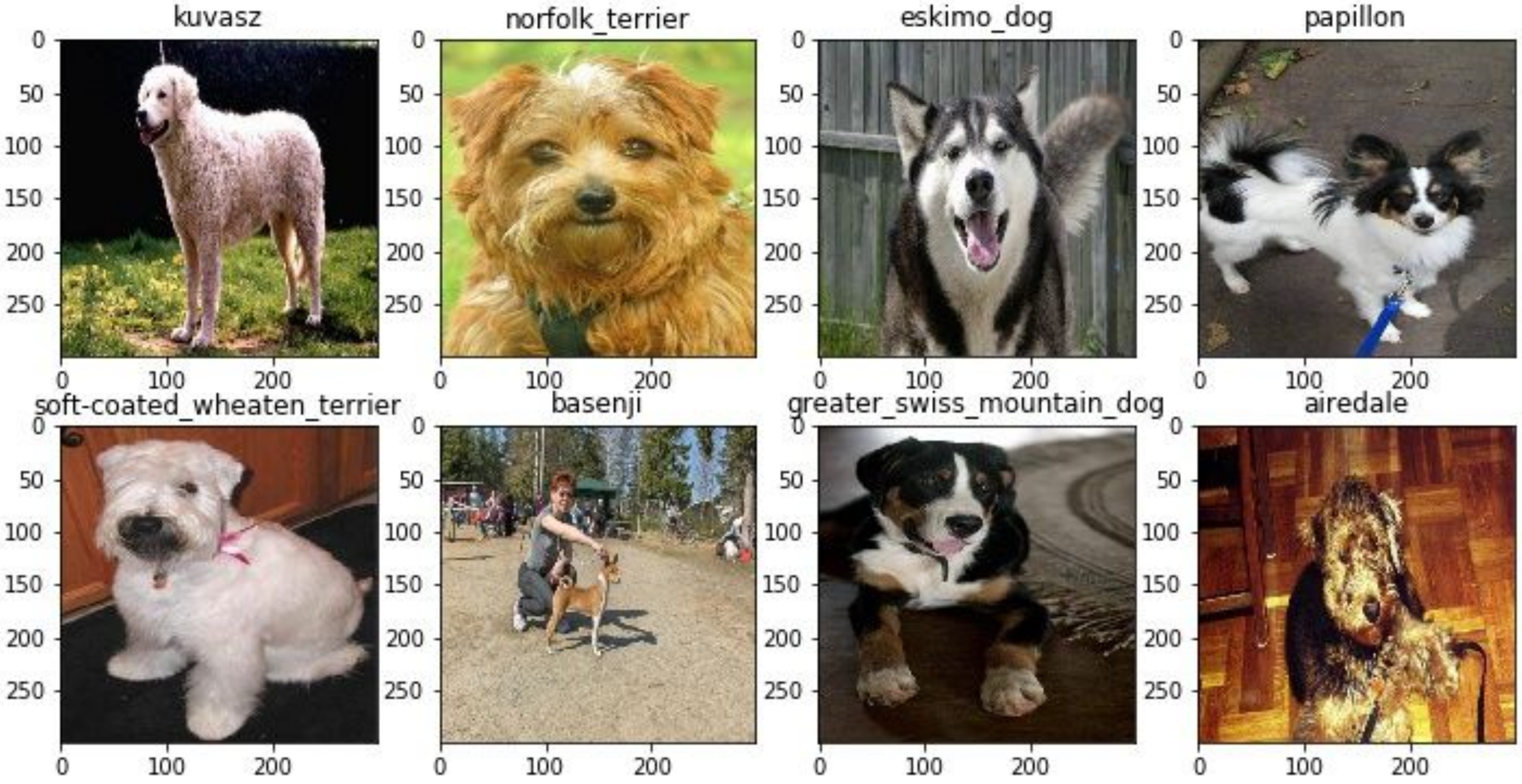

WAYMO
WAYMO


5M
FEB 2018
4M
NOV 2017
3M
MAY 2017
2M
OCT 2016
1M
JUNE 2015
2009
WAYMO
5,000,000+
MILES AND COUNTING

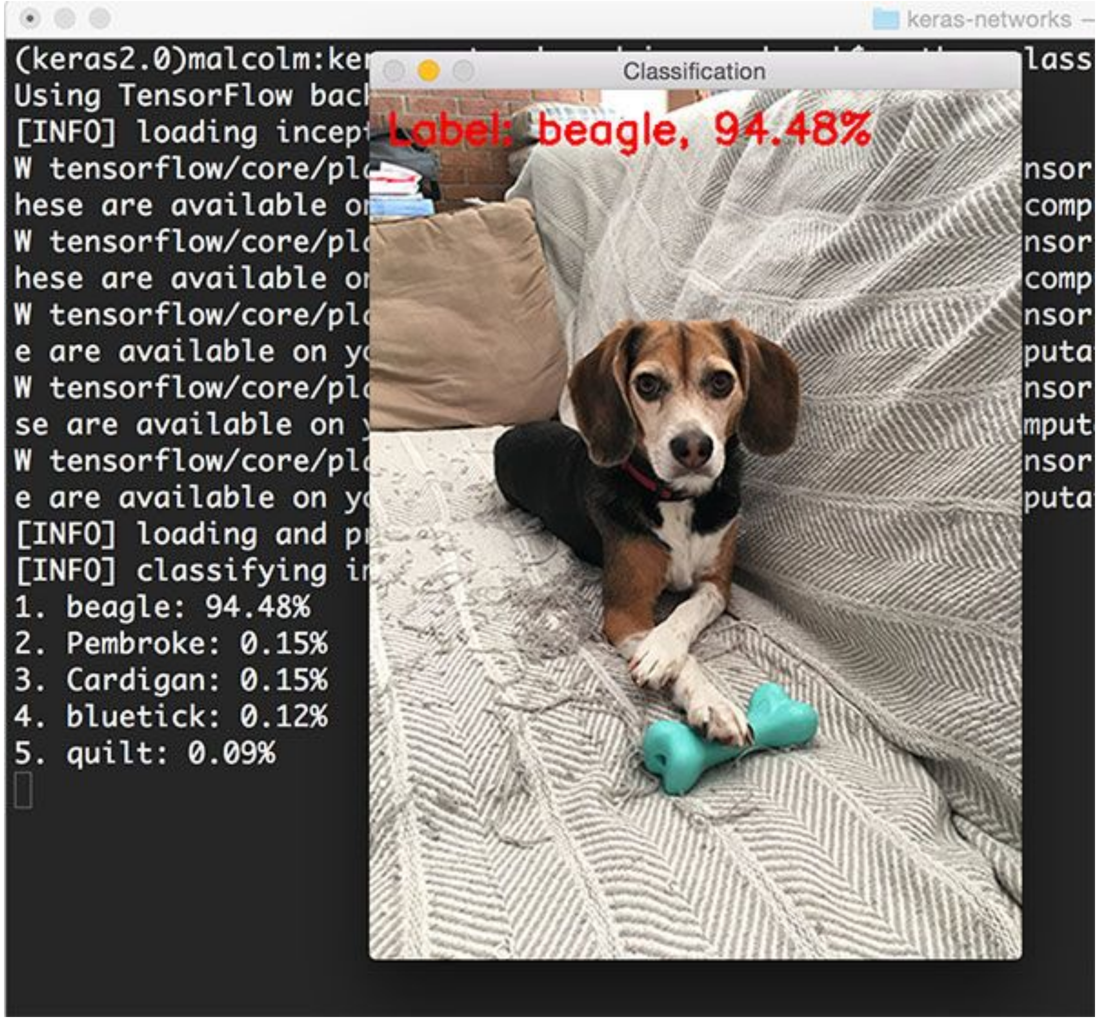


Source: dmv.ca.gov; ARK Investment Management LLC, 2018 | ark-invest.com


IMAGENET

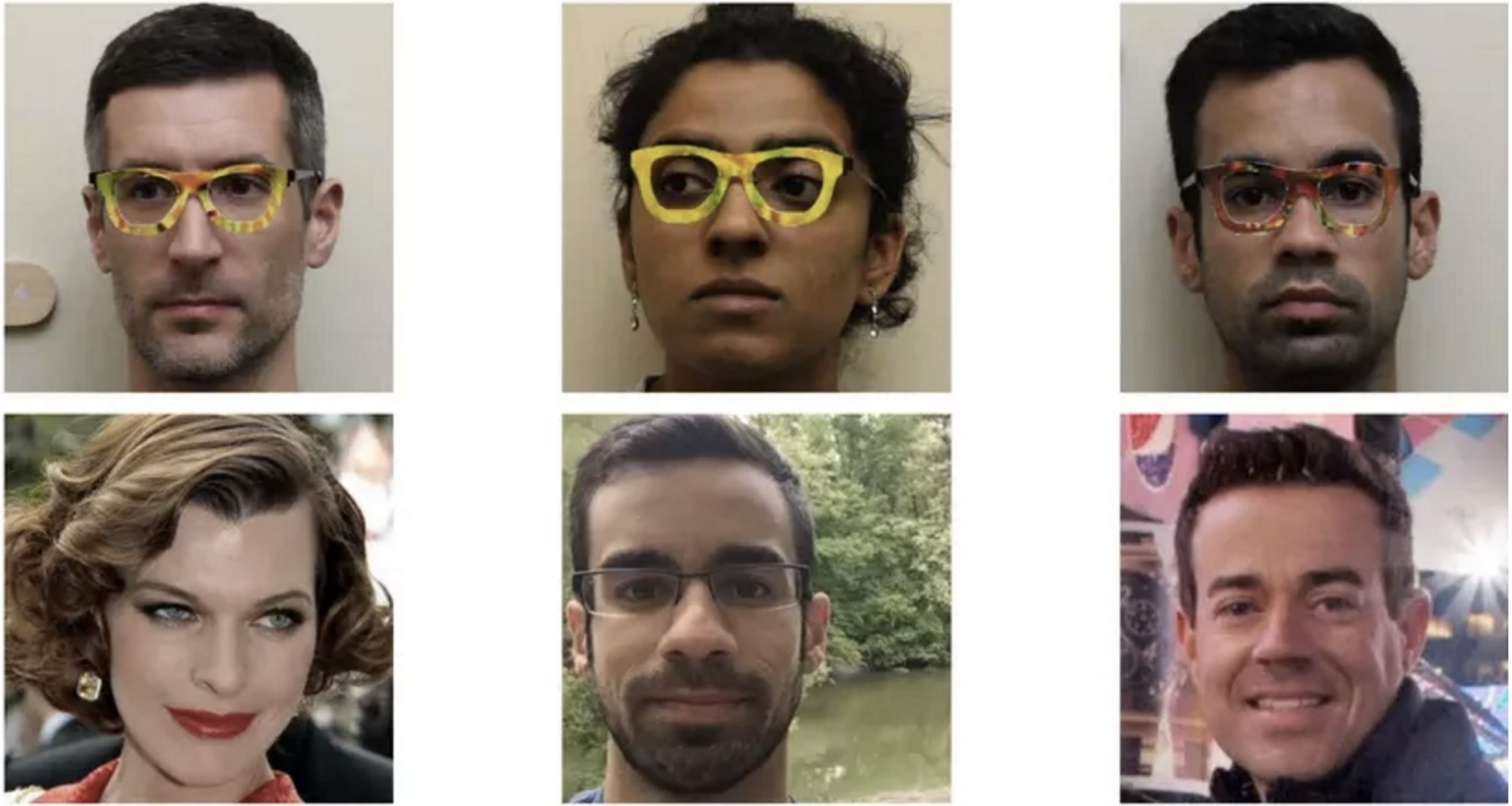

kuvasz
norfolk_terrier
eskimo_dog
papillon
soft-coated_wheaten_terrier
basenji
greater_swiss_mountain_dog
airedale

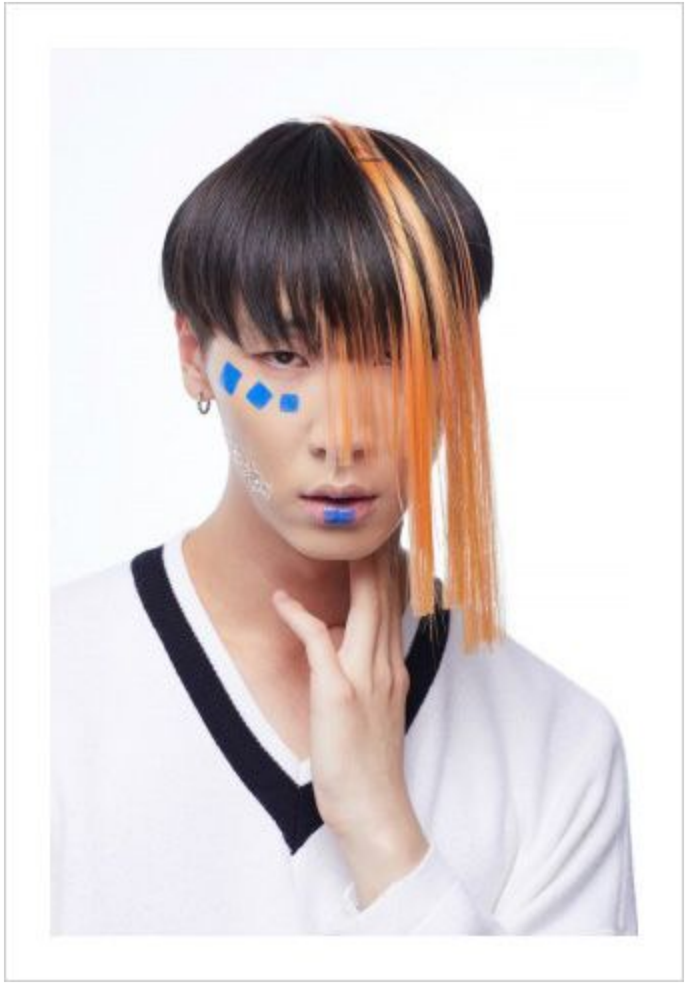
wandb.com

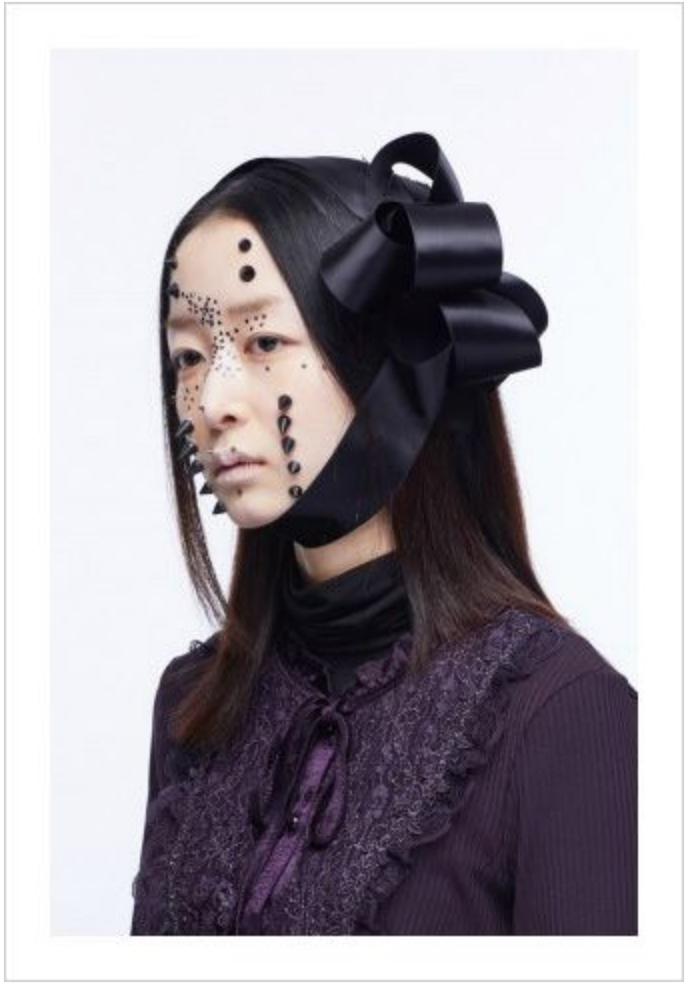

keras-networks
Classification
Label: beagle, 94.48%
(keras2.0)malcolm:ker
Using TensorFlow bac
[INFO] loading incep
[INFO] loading and p
[INFO] classifying i
1. beagle: 94.48%
2. Pembroke: 0.15%
3. Cardigan: 0.15%
4. bluetick: 0.12%
5. quilt: 0.09%


@teenybiscuit

It’s hard to guess what will confuse a model in production.







You Can Trick Self-Driving Cars by Defacing Street Signs

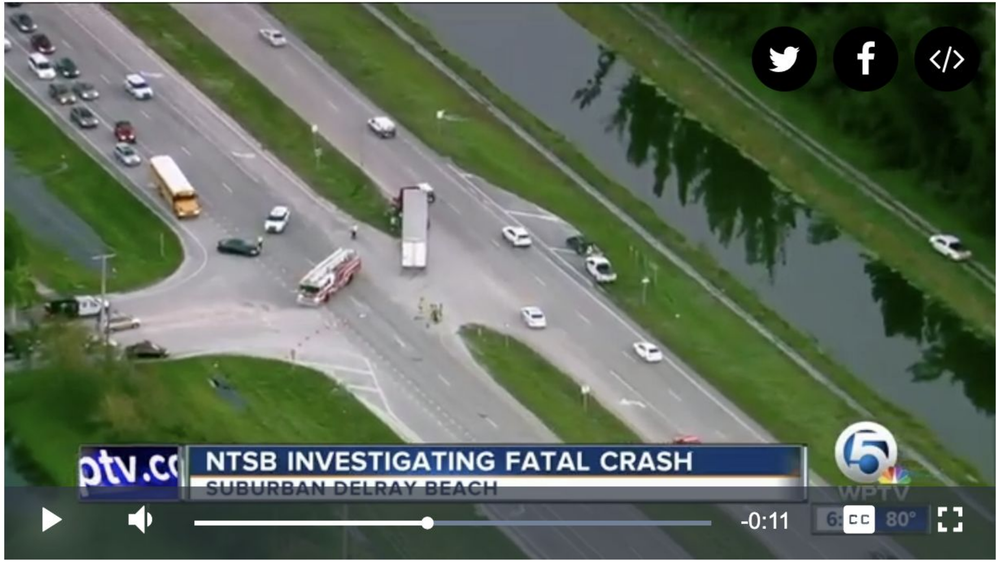
By Catalin Cimpanu

August 7, 2017

Tesla driver killed after crash involving semi in western Delray Beach

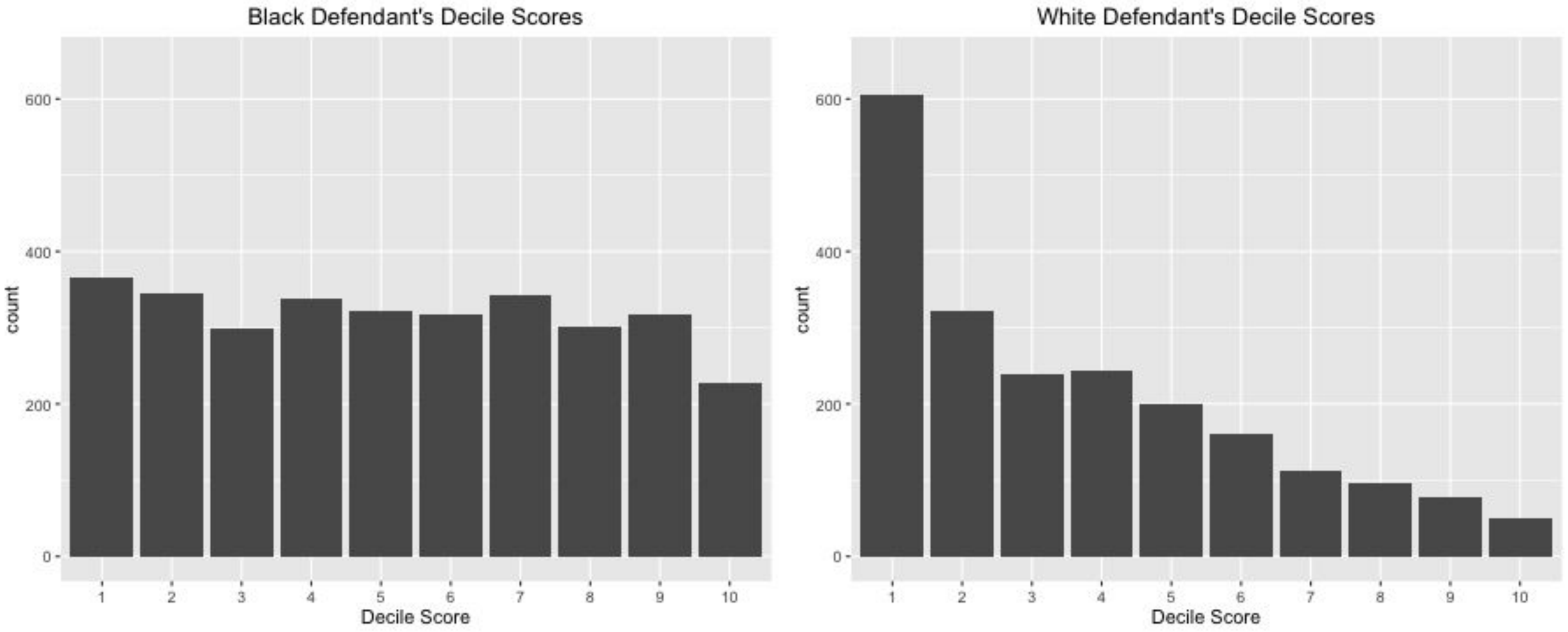
Posted: 7:10 AM, Mar 01, 2019 **Updated:** 4:52 PM, Mar 02, 2019

By: Scott Sutton



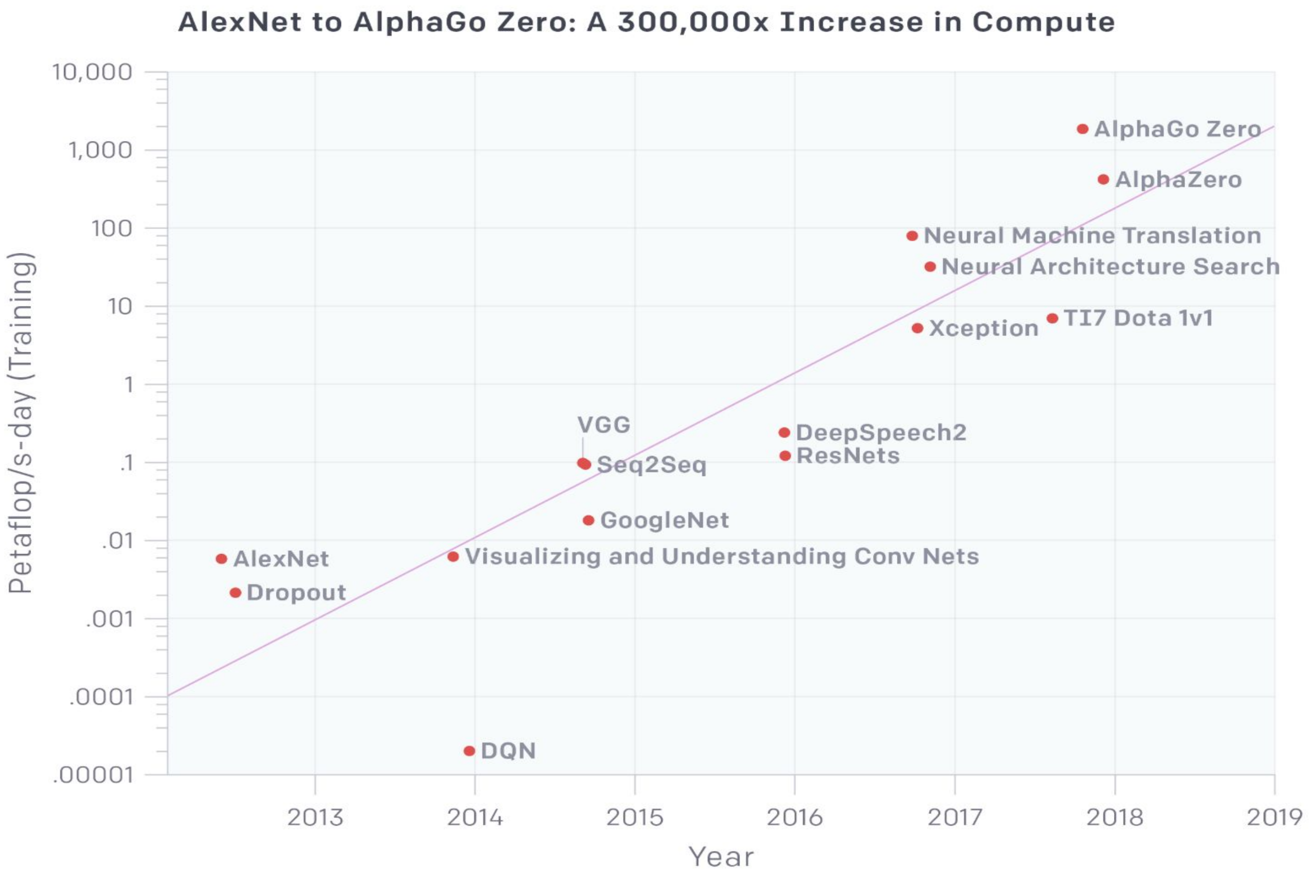
COMPAS Recidivism scores

COMPAS



Black Defendant's Decile Scores
White Defendant's Decile Scores
count
Decile Score
0
200
400
600
1
2
3
4
5
6
7
8
9
10

How do we get a model to explain why it made predictions?



https://openai.com/blog/ai-and-compute/

Many of the major advancements are driven by massive compute

1

Excitement

2

Problems

3

Opportunities

Opportunities

Explainability

“Why Should I Trust You?” Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

ABSTRACT

Despite widespread adoption, machine learning models remain mostly black boxes. Understanding the reasons behind predictions is, however, quite important in assessing *trust*, which is fundamental if one plans to take action based on a prediction, or when choosing whether to deploy a new model. Such understanding also provides insights into the model, which can be used to transform an untrustworthy model or prediction into a trustworthy one.

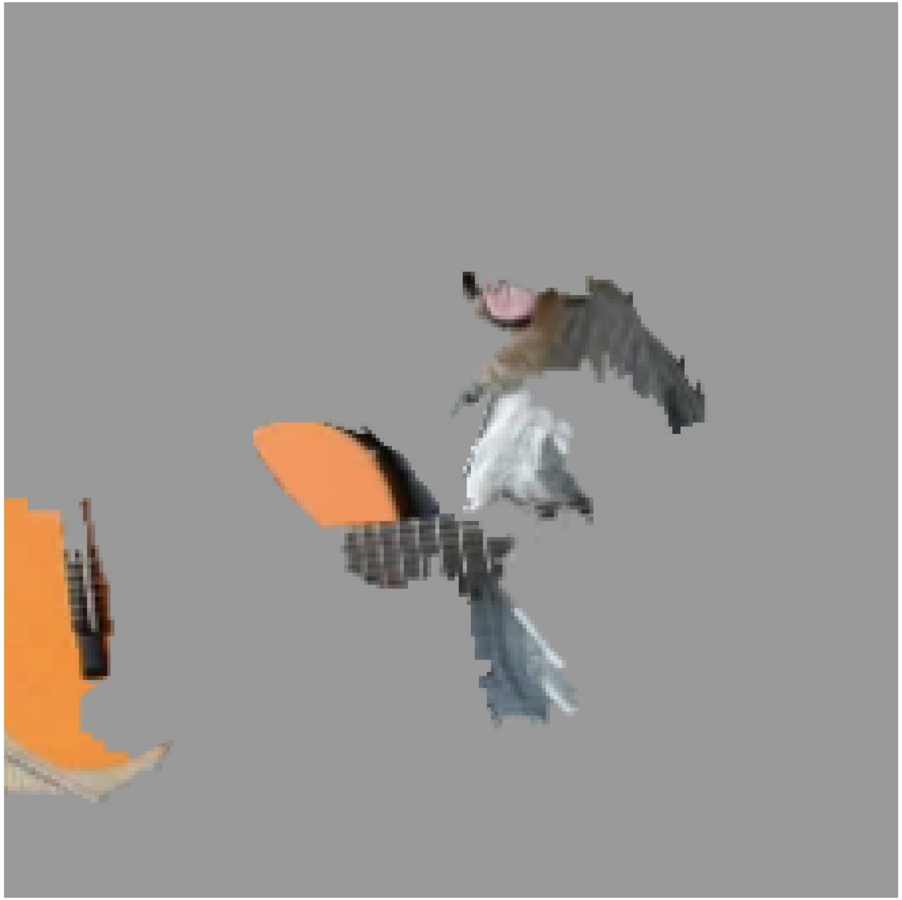
how much the human understands a model's behaviour, as opposed to seeing it as a black box.

Determining trust in individual predictions is an important problem when the model is used for decision making. When using machine learning for medical diagnosis [6] or terrorism detection, for example, predictions cannot be acted upon on blind faith, as the consequences may be catastrophic.

Apart from trusting individual predictions, there is also a need to evaluate the model as a whole before deploying it “in the wild”. To make this decision, users need to be confident

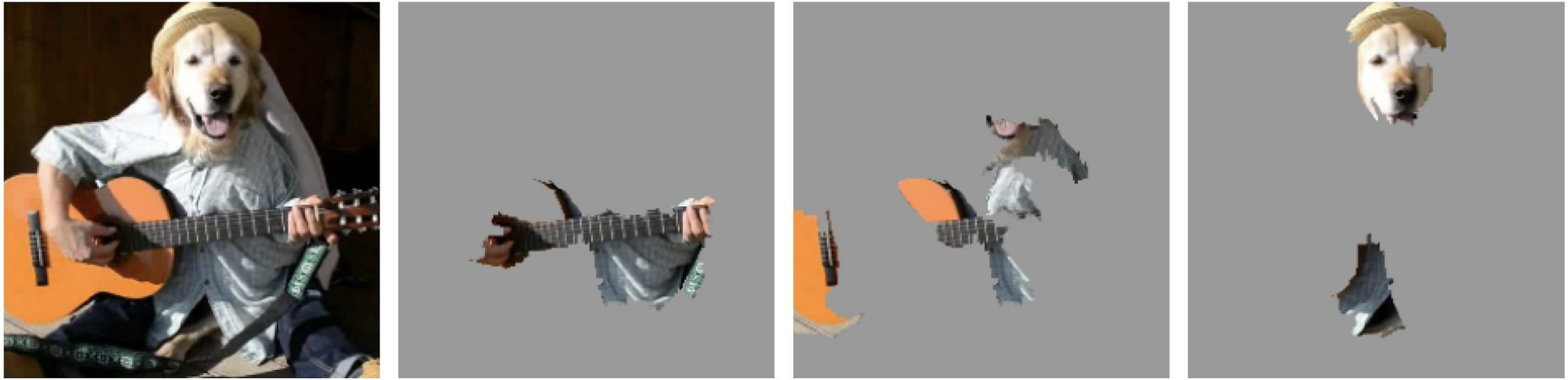
(a) Original Image

(b) Explaining *Electric guitar*



(c) Explaining *Acoustic guitar*

(d) Explaining *Labrador*



(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

Explainable Deep Neural Networks for Multivariate Time Series Predictions

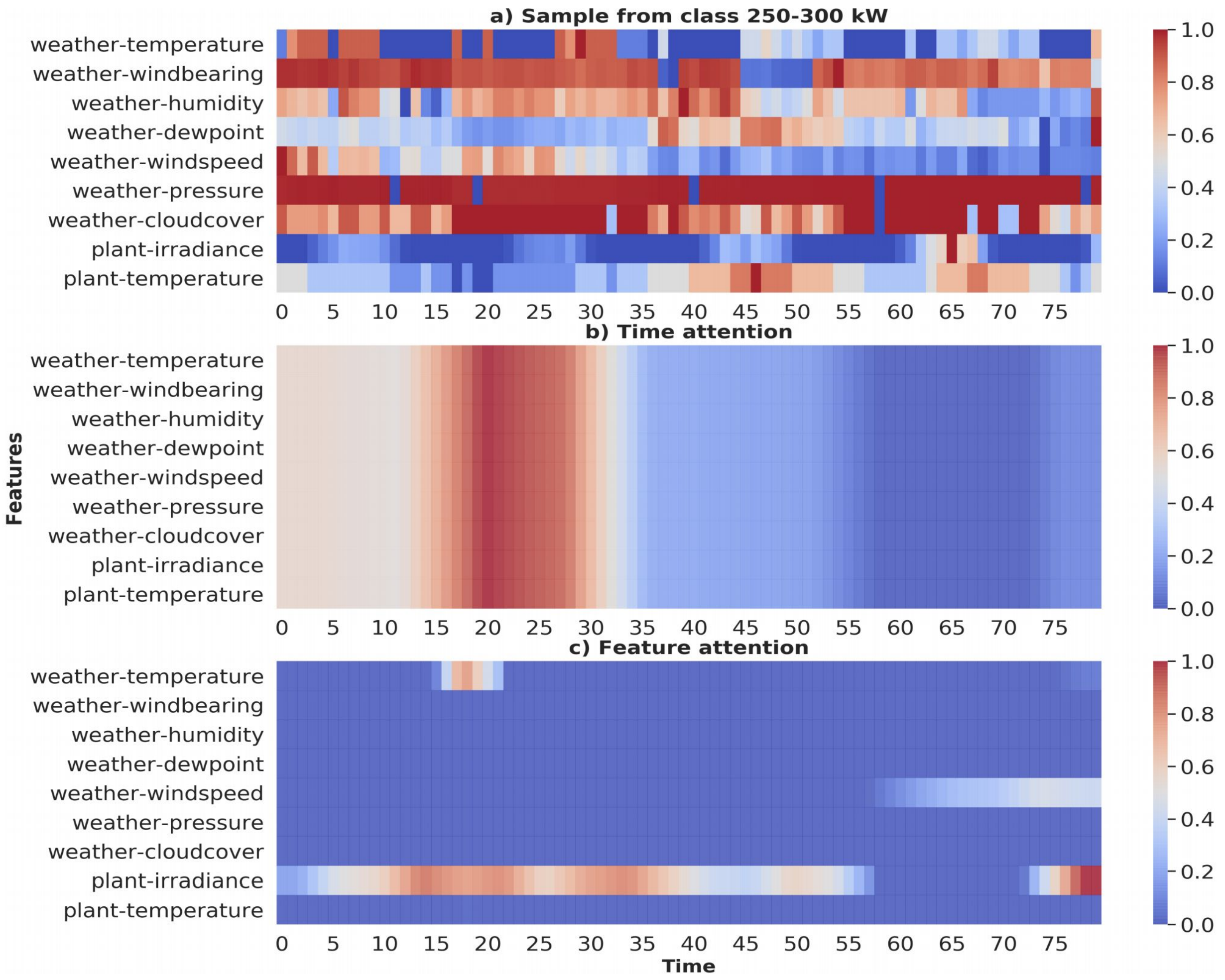
Roy Assaf and **Anika Schumann**
IBM Research, Zurich
{roa, ikh}@zurich.ibm.com

Abstract

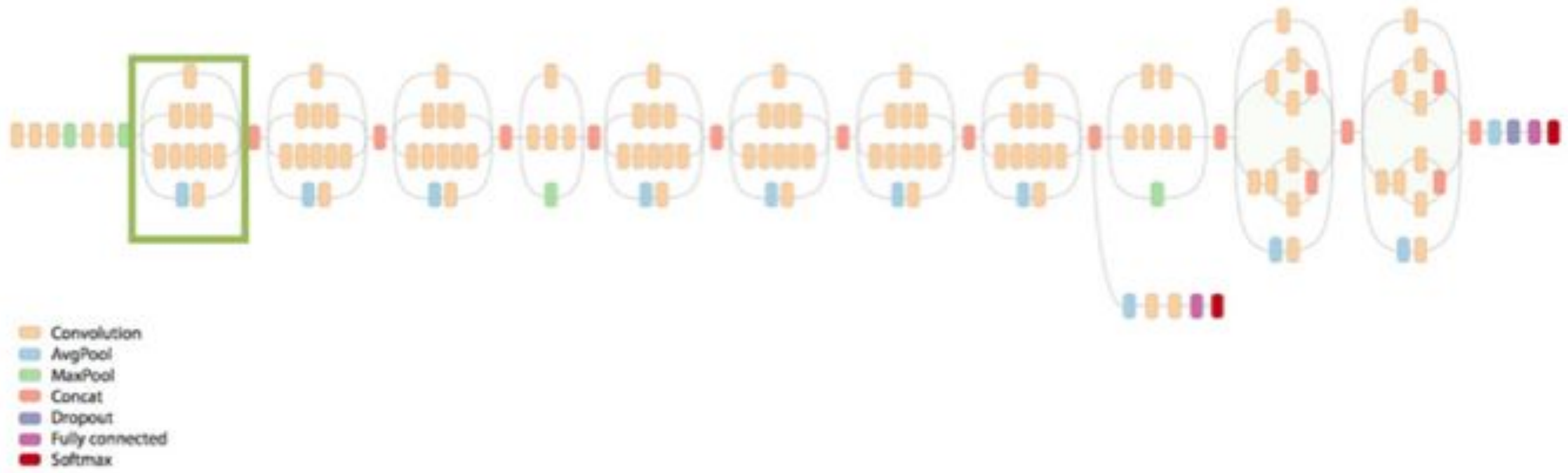
We demonstrate that CNN deep neural networks can not only be used for making predictions based on multivariate time series data, but also for explaining these predictions. This is important for a number of applications where predictions are the basis for decisions and actions. Hence, confidence in the prediction result is crucial. We design a two stage convolutional neural network architecture which uses particular kernel sizes. This allows

2 Method for Explainable Deep Network

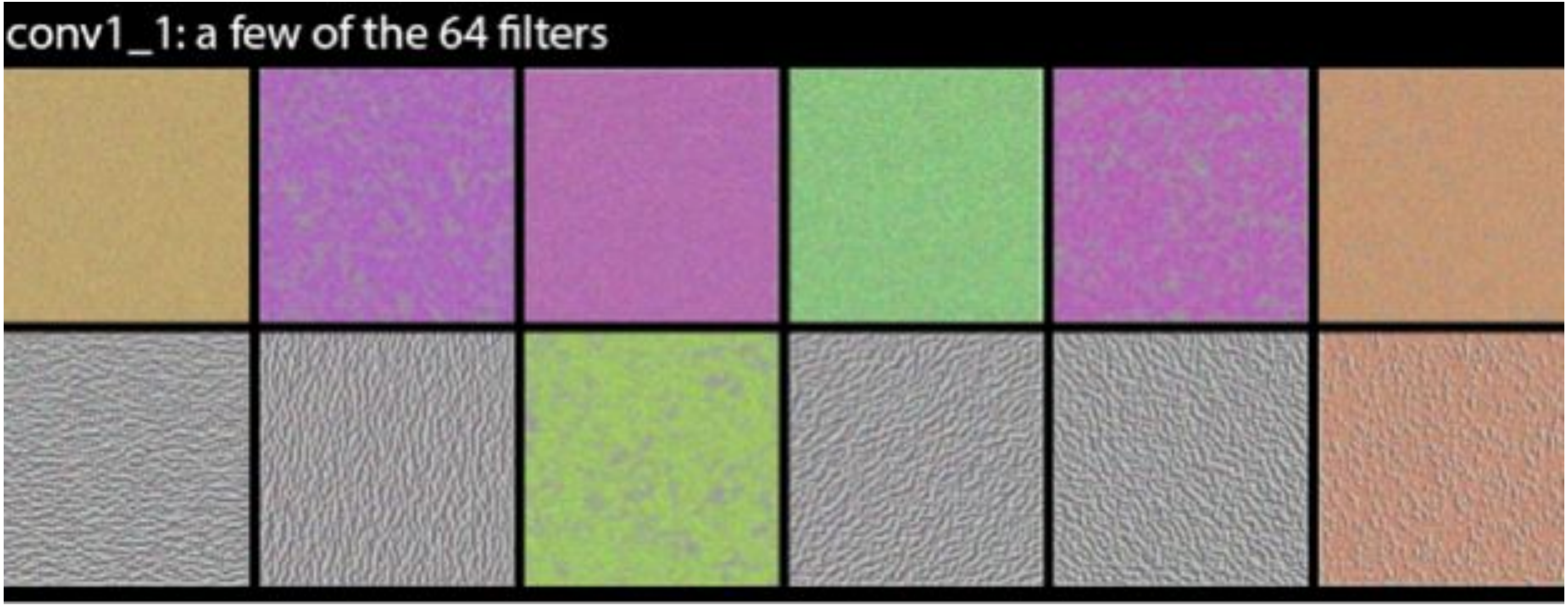
In order to achieve explainable predictions for both the time dimension and the features of the data, we develop a two stage CNN architecture. The first stage consists of a convolutional layer and utilises a 2D convolution with filter size $k \times 1$ which considers k time steps with 1 feature at a the time. This allows us to learn filters which are able to recognise important patterns that occur separately in the different features. This stage is followed by a 1×1 convolution [Lin *et al.*, 2013] and is used in state-of-the-art networks such as in the incep-


a) Sample from class 250-300 kW
b) Time attention
c) Feature attention
Features
Time
weather-temperature
weather-windbearing
weather-humidity
weather-dewpoint
weather-windspeed
weather-pressure
weather-cloudcover
plant-irradiance
plant-temperature
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
0.0 0.2 0.4 0.6 0.8 1.0

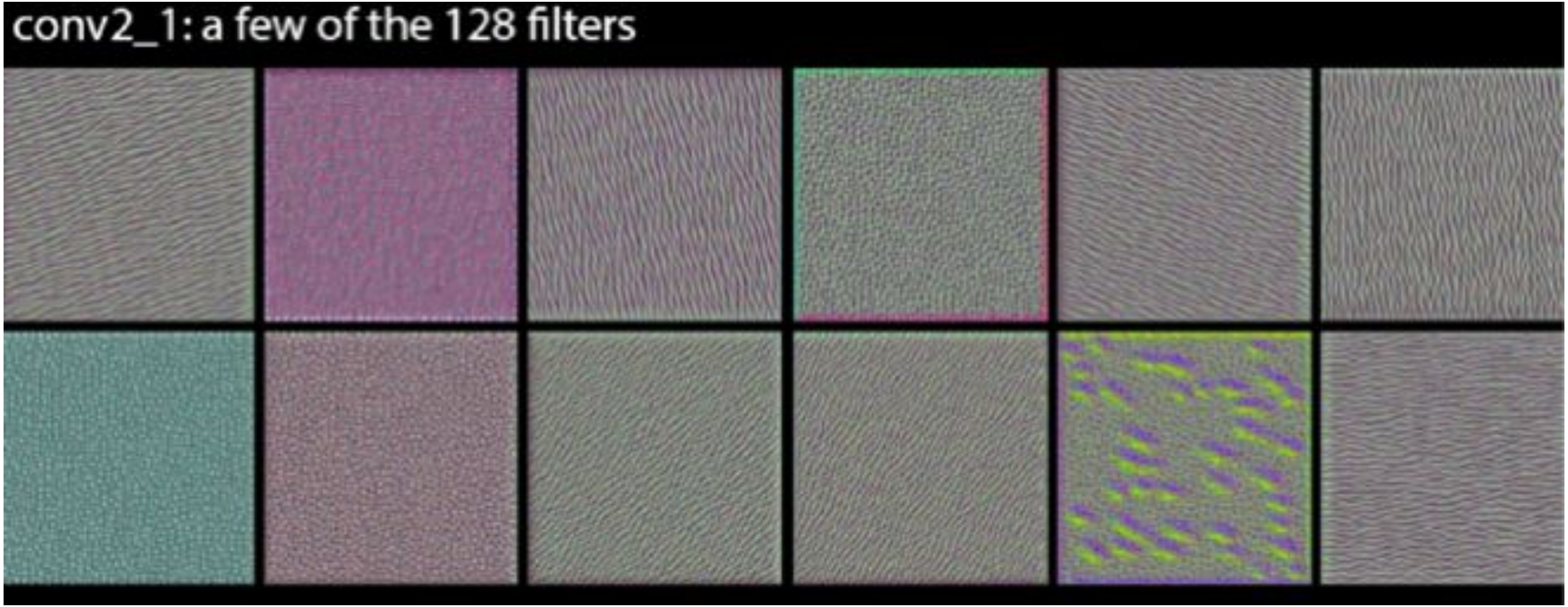
Inception

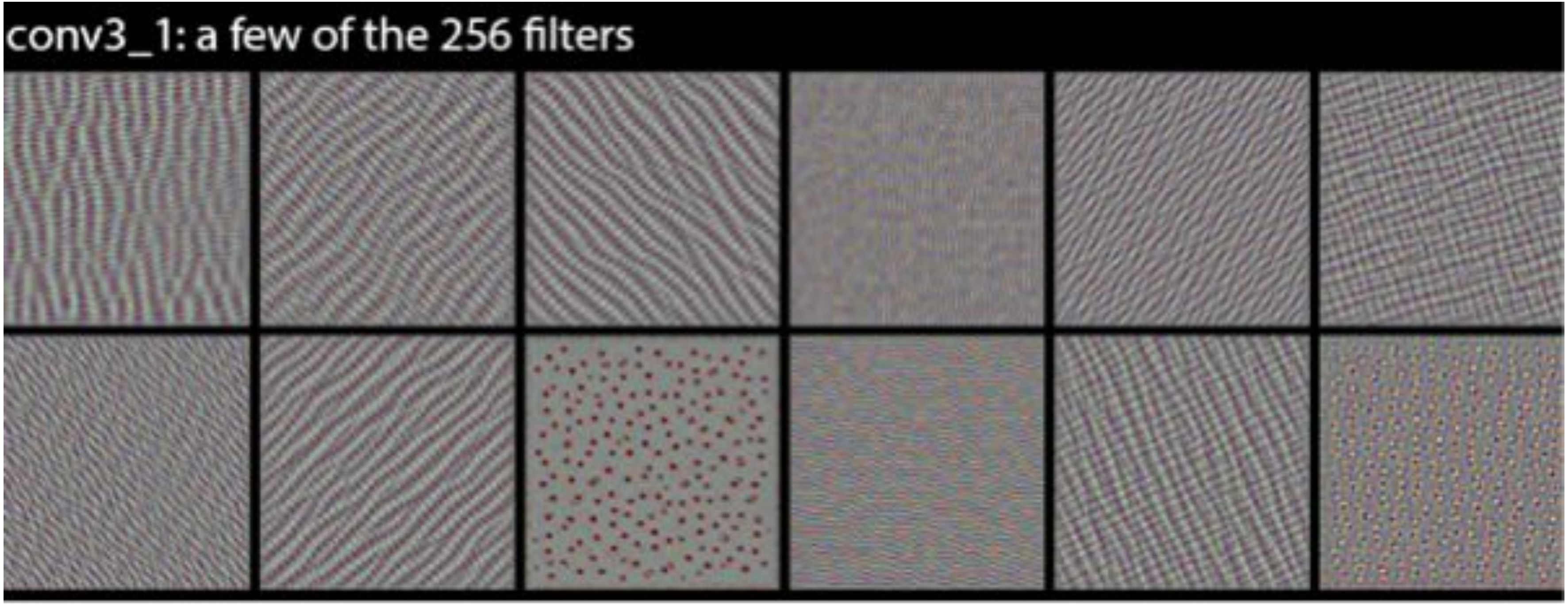


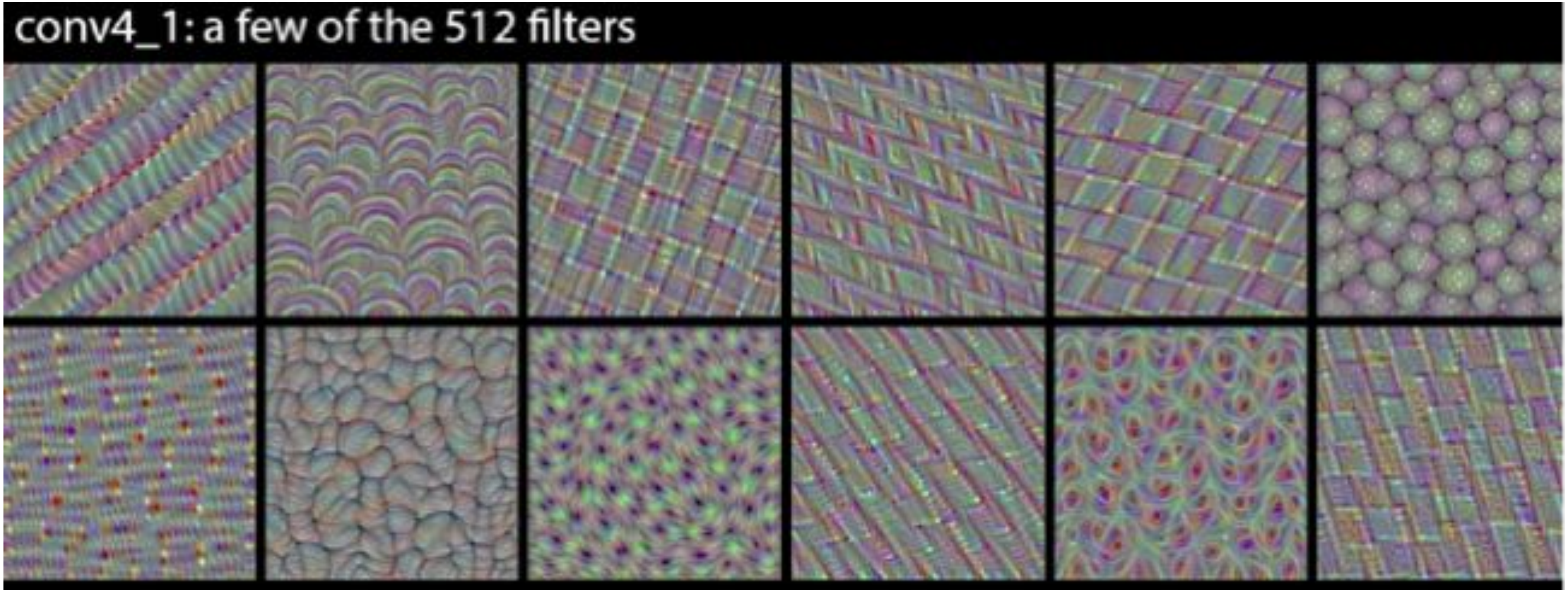
wandb.com

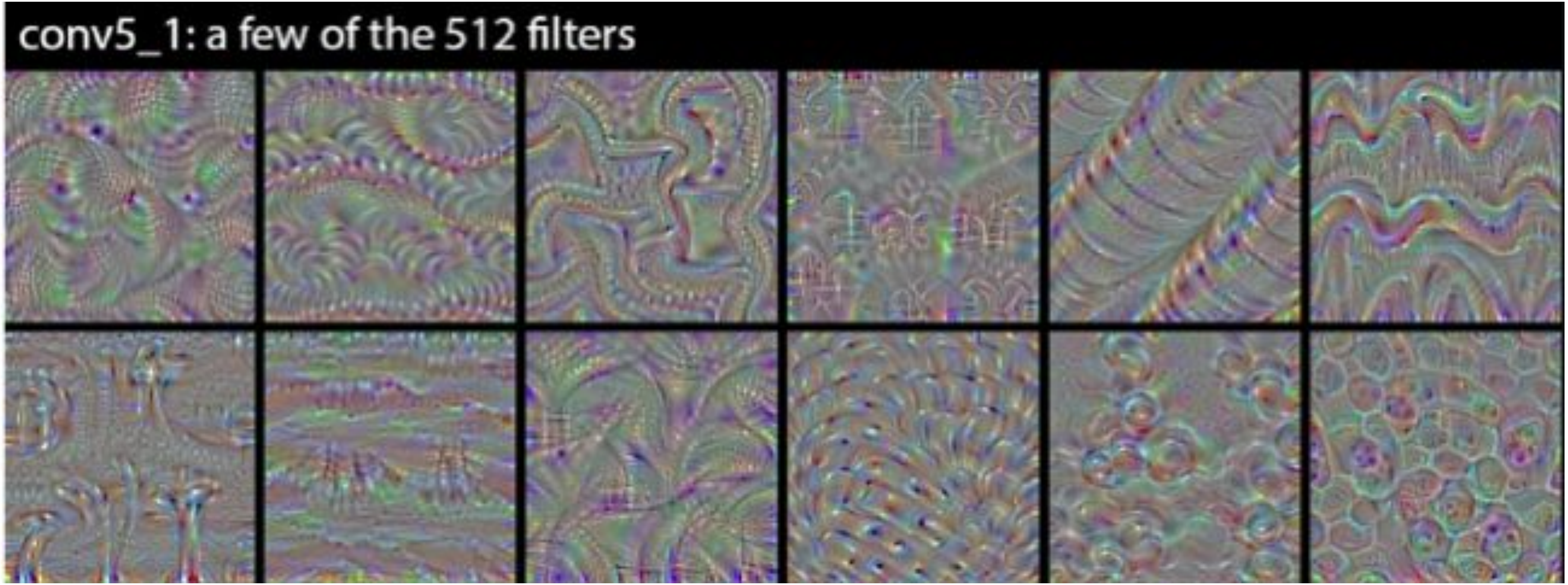


https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html


conv2_1: a few of the 128 filters


conv3_1: a few of the 256 filters


conv4_1: a few of the 512 filters


conv5_1: a few of the 512 filters

Reproducibility

Science

IN DEPTH | COMPUTER SCIENCE

Artificial intelligence faces reproducibility crisis

Matthew Hutson

MIT Technology Review

Artificial Intelligence Feb 18

Machine learning is contributing to a "reproducibility crisis" within science

wandb.com

Artificial Intelligence Confronts a 'Reproducibility' Crisis

Machine-learning systems are black boxes even to the researchers that build them. That makes it hard for others to assess the results.

Runs (295)

1 Filter Group Sort Tag Move

Name (122 visualized)	Tags	Runtime	Train loss:	best_epo...	best_epo...	best_epo...	best_eva...	best_eva...
f8gce7l2	2 layer, 3evalEpochs, both, cacheFeats, ctxtGRU, diff 0, graph, no bbo	17h 12...	0.1271	24	49	49	0.7741	0.7743
tmkrp4qh	2 layer, 3evalEpochs, both, cacheFeats, diff 0, driver, graph, no bbox	1d 19h ...	0.1714	78	67	67	0.7726	0.7653
3hede45x	2 layer, 3evalEpochs, both, cacheFeats, diff 0, driver, graph, no bbox	1d 19h ...	0.1736	34	63	56	0.772	0.7683
0owdyuep	2 layer, 3evalEpochs, both, cacheFeats, diff 0, driver, graph, no bbox	1d 20h ...	0.1522	34	45	18	0.7715	0.7698
y2f6jbke	2 layer, 3evalEpochs, both, cacheFeats, ctxtGRU, diff 0, graph, no bbo	18h 44...	0.1626	67	37	37	0.7738	0.7743
w11tf9dy	2 layer, 3evalEpochs, both, cacheFeats, ctxtGRU, diff 0, graph, no bbo	18h 49...	0.1132	63	26	26	0.7785	0.7862
8ft1iyok	2 layer, 3evalEpochs, both, cacheFeats, diff 0, graph, no bbox, pedGR	17h 41...	0.1165	65	16	65	0.7723	0.7638
vmhulexq	2 layer, 3evalEpochs, both, cacheFeats, diff 0, graph, no bbox, pred30	13h 6m...	0.2152	24	63	63	0.7113	0.6996
q409s2mj	2 layer, 3evalEpochs, both, cacheFeats, diff 0, graph, no bbox, pedGR	23h 27...	0.2292	55	45	45	0.7567	0.7414
rkke81wd	2 layer, 3evalEpochs, both, cacheFeats, diff 1, graph, no bbox, pedGR	1d 7m 41s	0.1191	11	25	8	0.7682	0.7683
im9teeq6	2 layer, 3evalEpochs, both, cacheFeats, diff 1, graph, no bbox, pedGR	13h 50...	0.113	8	25	25	0.7676	0.7683
t25gqvmd	2 layer, 3evalEpochs, both, cacheFeats, diff 1, graph, no bbox, pedGR	13h 57...	0.1123	11	25	25	0.7671	0.7698
68go88t4	2 layer, 3evalEpochs, both, cacheFeats, graph, no bbox, pedGRU, prec	17h 18...	0.2117	32	42	43	0.7735	0.7653
hy1eq7ua	2 layer, 3evalEpochs, both, cacheFeats, graph, no bbox, pedGRU, prec	29m 34s	0.1188	2	3	3	0.7524	0.7444
6cx2xwqk	2 layer, 3evalEpochs, both, cacheFeats, graph, no bbox, pedGRU, prec	16h 49...	0.187	41	42	43	0.7681	0.7698

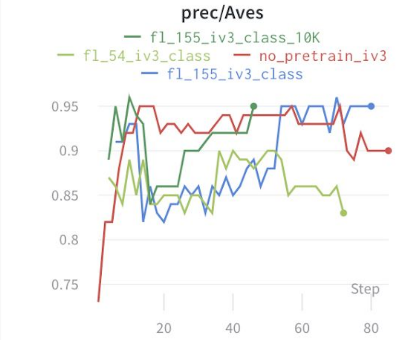

Runs (295)
Name (122 visualized)
f8gce7l2
tmkrp4qh
3hede45x
0owdyuep
y2f6jbke
w11tf9dy
8ft1iyok
vmhulexq
q409s2mj
rkke81wd
im9teeq6
t25gqvmd
68go88t4
1-50 of 122
VISUALIZATIONS
Add a visualization
eval_loss
pvjhtfdm
5ojka19l
out6k74e
le4jp9gx
vci4z47p
k9s2pf5e
0gculszi
tfwpustj
9vx787tv
yc5w9w1g
3
2.5
2
1.5
1
0.5
0
Step
500
1k
1.5k
best_eval_acc
pvjhtfdm
5ojka19l
out6k74e
le4jp9gx
vci4z47p
k9s2pf5e
0gculszi
tfwpustj
9vx787tv
yc5w9w1g
0.81
0.8

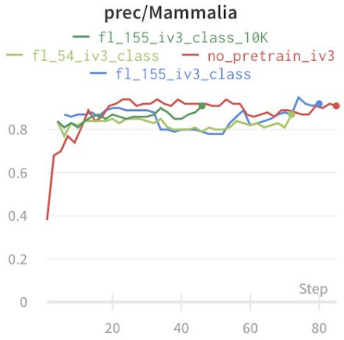
Per-class precision with InceptionV3

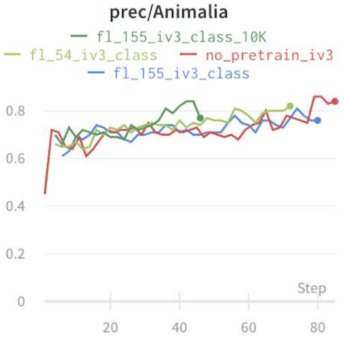
More data for kingdoms

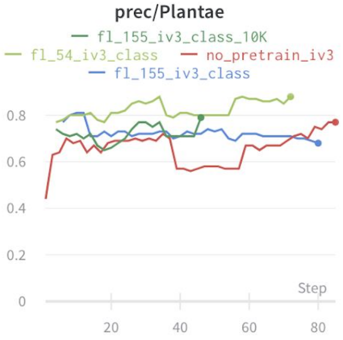
How well do we perform on each of the 10 classes? Look at per-class precision for different models (5K examples, hence noisy/jagged plots)

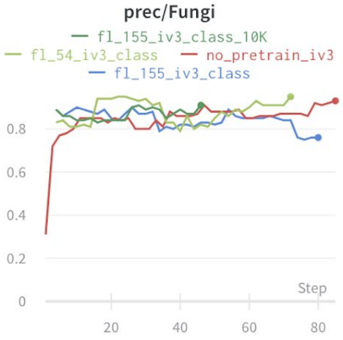
- **per-class precision**: Animalia and Plantae seem to the be worst (70-80). Birds are best (up to 95). Molluscs and Reptiles slightly worse than Amphibia, Arachnida, Fungi, Insects, Mammals (all 80-90). Birds have most species represented, may

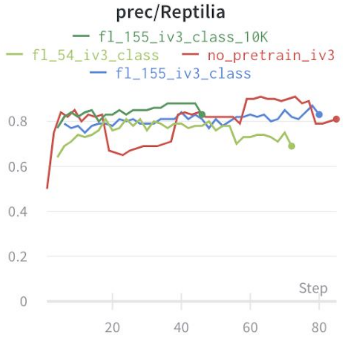












LINKS FROM THIS TALK

bit.ly/ucsd-ml