go/ml4pms-slides-q42017

ML for PMs

Speaker Series

go/ml4pms, ml4pms-organizers@

Mountain View, Dec 5th 2017

by the Research & Machine Intelligence PM community

Agenda

- **Welcome**
- Fairness: pbrandt@
- Human Sensing: dkaram@
- ML and Data: ivanku@
- Crowd Computing: pocketaces@
- Natural Language: barakt@
- On-device: ingerman@
- Medical Applications: lhpeng@
- Getting to Launch: binghamj@
- Refreshing Conversations

Agenda

- Welcome
- **Fairness**: pbrandt@
- Human Sensing: dkaram@
- ML and Data: ivanku@
- Crowd Computing: pocketaces@
- Natural Language: barakt@
- On-device: ingerman@
- Medical Applications: lhpeng@
- Getting to Launch: binghamj@
- Refreshing Conversations

Fairness

Peter Brandt (pbrandt@)

ML Fairness is a major product issue

Intelligent Machines

Forget Killer Robots—Bias Is the Real AI Danger

John Giannandrea, who leads AI at Google, is worried about intelligent systems learning human prejudices.

by Will Knight October 3, 2017

Google's AI chief isn't fretting about super-intelligent killer robots. Instead, John Giannandrea is concerned about the danger that may be lurking inside the machine-learning algorithms used to make millions of decisions every minute.

"The real safety question, if you want to call it that, is that if we give these systems biased data, they will be biased," Giannandrea said before a recent Google conference on the relationship between humans and AI systems.

https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/

Some recent headlines...

LGBT community anger over YouTube restrictions which make their videos invisible

Google and Facebook Face Criticism for Ads Targeting Racist Sentiments

Google engineer apologizes after Photos app tags two black people as gorillas

Google's Sentiment Analyzer Thinks Being Gay Is Bad

What do we mean by ML Fairness?

Policy definition

“algorithmic unfairness” means unjust or prejudicial treatment of people that is related to sensitive characteristics such as race, income, sexual orientation, or gender, through algorithmic systems or algorithmically aided decision-making -- go/algorithmic-unfairness-definition

What do we mean by ML Fairness?

Technical definitions

Demographic parity -- predictions must be uncorrelated with the sensitive attribute

Equal opportunity -- individuals who qualify for a desirable outcome should have an equal chance of being correctly classified for this outcome

See [go/eosl-paper] for more

Product use cases which may raise concerns

- Content moderation and filtering
- Personalization and ads targeting
- Image model use cases involving people
- Text model use cases involving web or user-generated content

What are we doing for ML Fairness?

There is a ton of work going on across the company! go/ml-fairness has an overview.

What are we doing for ML Fairness?

1) “Vanguard project” collaborations with product teams

2) Creating internal docs, tools, and policy/legal guidance

3) Establishing a design and launch review process

YouTube Clickbait Vanguard

ML Fairness Concern

- **Clickbait classifier had higher False Positive Rate (FPR) for a protected group**

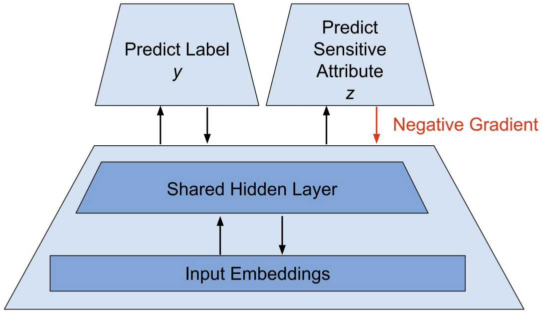
Outcomes / Impact

- Launched new Clickbait classifier in collaboration with SIR+MLX!
- FPR for Clickbait classifier improved by ~40% in instant prod and by ~70% in stable prod for sensitive content[1].

Learnings

- **Modeling technique:** First real-world demonstration of *adversarial multi-task learning*[2] - able to significantly reduce FPR for a protected group within a content filter. Colab created and is available to product teams (go/ml-fairness-colabs).
- **Fairness measures:** First time *Equality of Opportunity*[3] has been applied to a Google product. FPR/FNR trade-offs are real and will be product-dependent.
- **Data labelling**: Training/Evaluation relied on obtaining labels for sensitive content; underscores need for labels for similar analysis.

Adversarial multi-task model architecture[2]



1. Launch documentation for Clickbait Fairness Model
2. Data Decisions and Theoretical Implications when Adversarially Learning Fair Representations
3. Equality of Opportunity in Supervised Learning (Hardt et. al, 2016) defines "Equality of Opportunity" as $P(\hat{Y}=1|Y=1,Z=0) = P(\hat{Y}=1|Y=1,Z=1)$

Mobile Vision/FaceNet Vanguard

ML Fairness Concern

- **FaceNet performance varies across race and gender subgroups**

Outcomes / Impact

- "UHS Diversity Classifier" built on FaceNet provides possibility of measuring unfairness quantitatively across race and gender subgroups.[1]
- Face attribute detection "smiling" improved by inferring race and gender[2] demonstrates importance of sensitive category inference for downstream fairness.
- Discovering known and new race/gender subgroups by leveraging small amounts of labeled data promising in initial tests.[3]

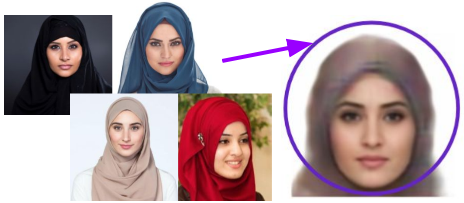
Learnings

- **Data:** Aggregating input for synthetic user generation can help mitigate privacy and legal concerns.
- **Accuracy:** Inferring sensitive subgroups can improve performance on downstream subgroup-dependent task.
- **Modeling technique:** Nearest neighbors for face images from two different subgroups can aid discovery of new subgroups.

1. Accuracy: 91% for 4 races, 98% for gender.
2. +1.5% accuracy overall, more equal performance across subgroups with simple baselines.
3. Cosine k-means clustering with <500 initial labeled images per subgroup (2 genders, 5 races) resulted in 75-95% accuracy range across known subgroups, and qualitatively reasonable new subgroups.

races {1, 2, 3} = {.45 .35 .20}

55% "smiling" $p < .05$



Conversation AI Vanguard

ML Fairness Concern

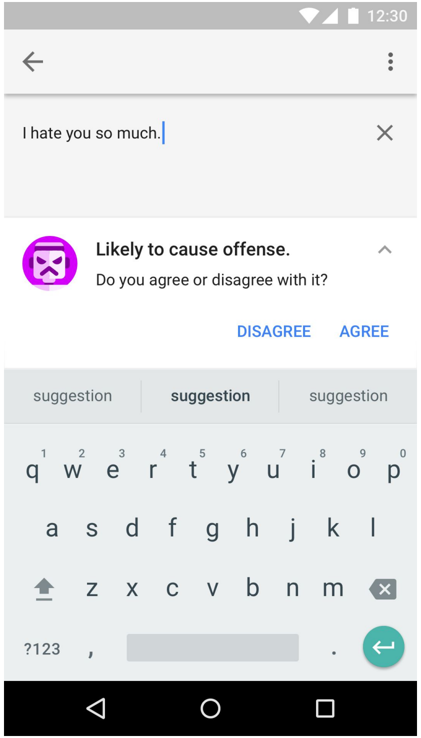
- **Skewed representation of targeted groups in training data on harassment, leading to unintended bias**

Outcomes / Impact (still work-in-progress)

- New Pinned AUC evaluation metric and bias mitigation via strategic data addition (to be published in upcoming paper)
- Significant reduction in bias using data collected from external demo (users try to "game the system", thereby entering the adversarial data we need)
- Launch planned for Perspective API demo on Crowdsource app in December
- Initial study on Crowdflower annotator bias shows no difference between ratings across annotator demographic
- Experiments with crowdsourcing identity labels on text

Learnings

- **Crowdsourcing:** Demonstrated the efficacy of demos and crowdsourcing techniques for adversarial testing and rich data generation
- **Evaluation:** Invented new techniques and tools for measuring bias in text classification model



Adversarial testing in Crowdsource App

Internal docs, tools, and policy/legal guidance

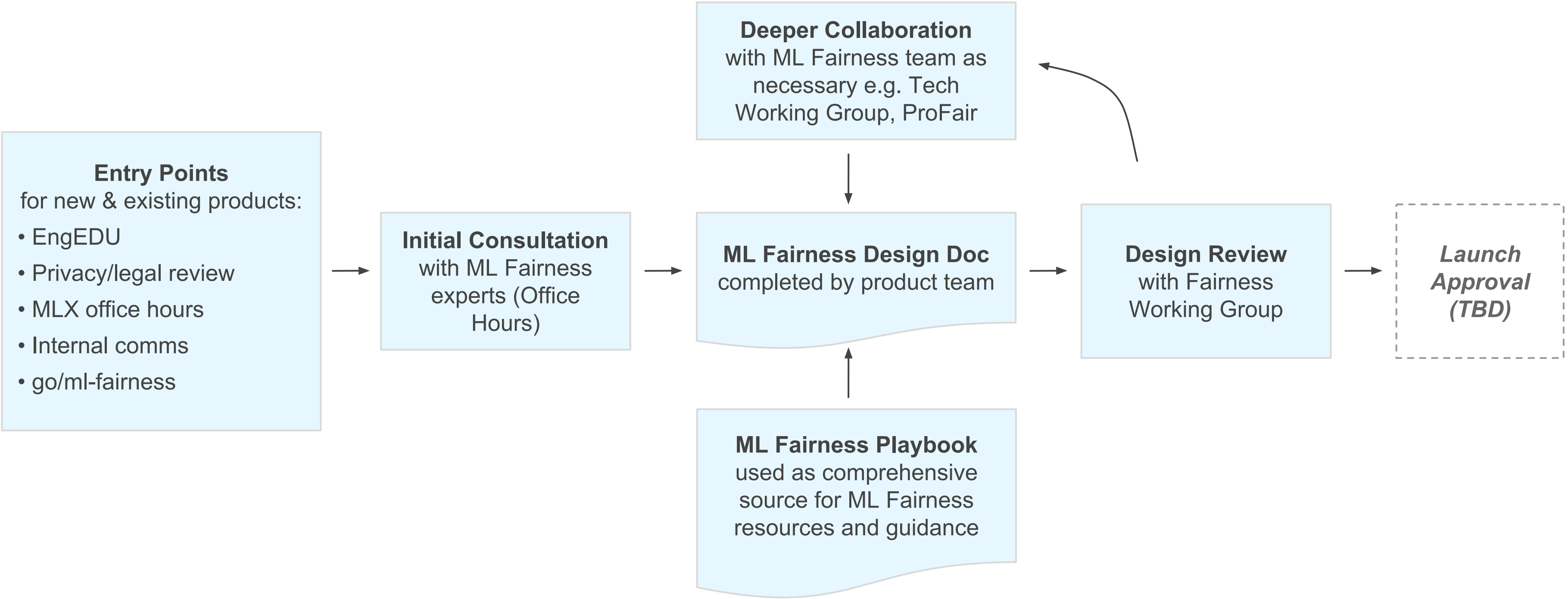
ML Fairness site with links to tons of resources -- go/ml-fairness

ML Fairness technical resources -- go/ml-fairness-tech

ML Fairness design doc -- go/ml-fairness-dd

ML Fairness Playbook with product, data, model, and incident response guidance (coming soon!)

Design and launch review process



Help Google build inclusive ML products

Reach out at **ml-fairness-questions@**. Engage with us early in the process!

Sign up to be a Vanguard project. We will help you leverage resources in Research and other PAs to build a solid plan.

Agenda

- ➢ Welcome
- ➢ Fairness: pbrandt@
- ➢ **Human Sensing**: dkaram@
- ➢ ML and Data: ivanku@
- ➢ Crowd Computing: pocketaces@
- ➢ Natural Language: barakt@
- ➢ On-device: ingerman@
- ➢ Medical Applications: lhpeng@
- ➢ Getting to Launch: binghamj@
- ➢ Refreshing Conversations

Human Sensing

David Karam (dkaram@)

Understand people, their emotions,
appearance and actions in images and videos.

On-device and in real-time.

Integrations and Partners

Summarization / Curation

AR & VR / Gaming

UXR / Content Engagement

Medicine / Mental Health

Assistant Vision & Personality

Customer Support

Automotive

Robotics

Surveillance

Detection

Description

Recognition

Geometry

Fairness

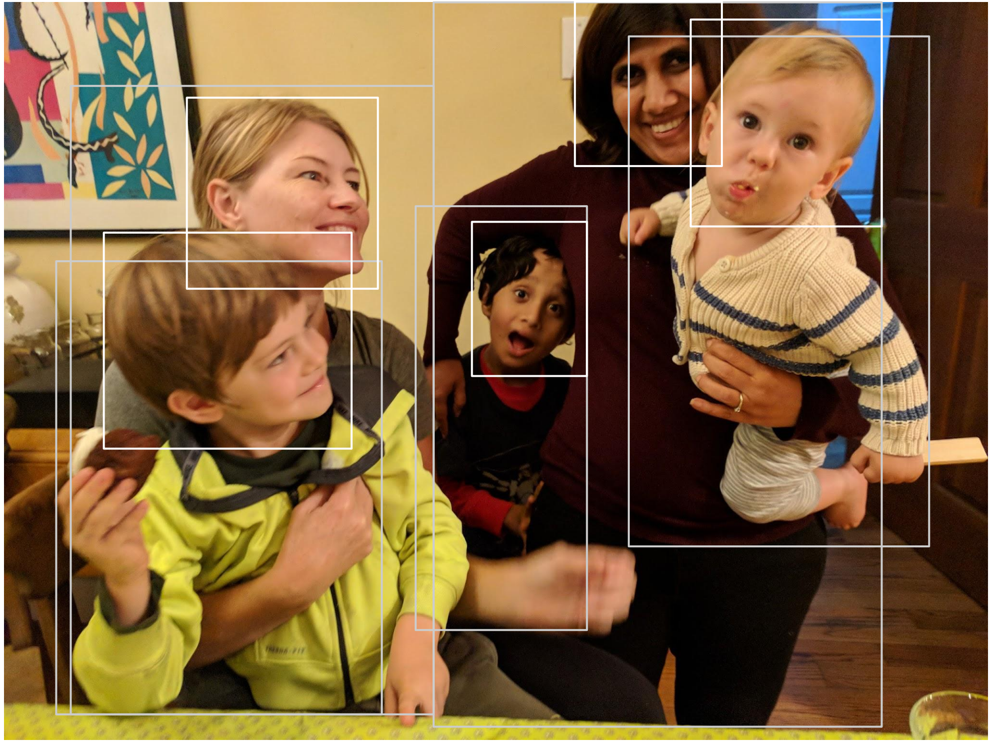
Detection

Description

Recognition

Geometry

Fairness



There are people here.

This is their location in the frame.

Face SDK
One stop shop
go/facesdk
developers.google.com/vision
jiayong@
hadam@

The Face SDK contains dedicated components for face related visual sensing. This includes detection, tracking, classification and recognition. This software provides the functional basis for a broad spectrum of applications, services and engines covering photo-management, image and video content analysis, automatic image labeling/ annotation, search, authentication and much more.

The Face SDK is integrated into GMSCore and available through the Mobile Vision API. It is also what currently powers our Cloud Vision API.

Face SSD
Single shot detector
go/face-ssd
go/mobile-ssd
menglong@
mttang@
dkalenichenko@

The Mobile SSD project provides a framework for compiling Single Shot MultiBox Detectors into a fast & lightweight inference library powered by tfmini. The library is cross- platform with the primary focus on mobile devices. It currently integrates face, products and common objects but is intended to be a general purpose framework for many vision detectors.

Detection

Description

Recognition

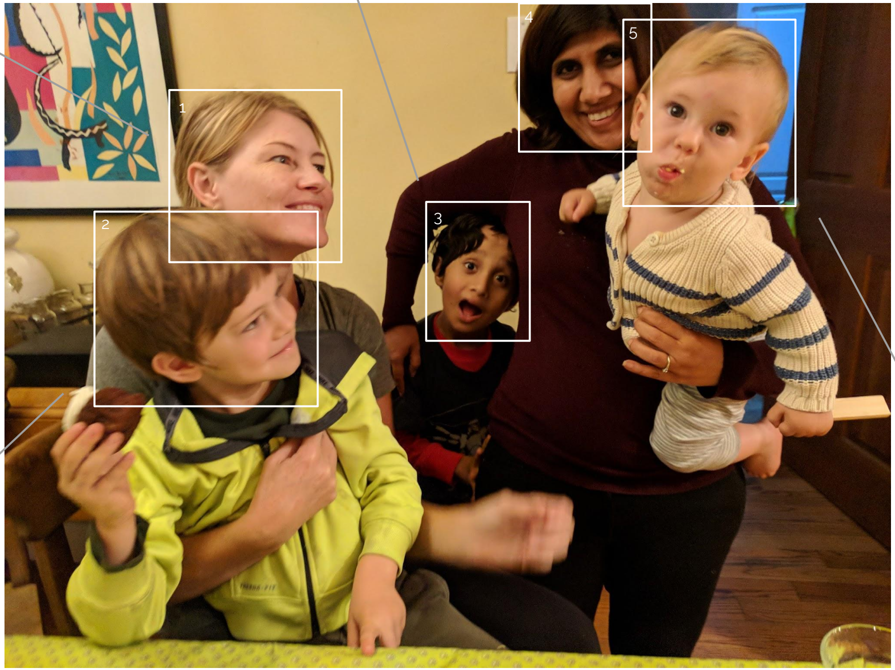
Geometry

Fairness

female
30
showing amusement, mild interest
smiling, long hair
holding person 2
looking at person 5

male
11
showing surprise, mild interest and some elation
looking at the camera

female
30
showing mild amusement
smiling, long hair
black pants
holding person 5
looking at the camera



male
7
showing interest
smiling, short hair
yellow jacket
sitting on lap of person 1
looking at person 5

Descriptions tell us what is apparent about each person:

Gender	male
Age	2
Emotion	showing surprise and interest
Face	short blond hair
Clothing	white sweater, striped pants
Activity	eating, being held by person 4
Gaze	looking at the camera

LookNet
Generic facial attributes
go/looknet
go/looknet-mobile
anm@
bochen@

Demographic
age, gender

Objective facial attributes
glasses, dark glasses, headwear, eyes visible, mouth open, facial hair, long hair, frontal gaze, sideburns, beard, mustache, squinting, smiling, black and white, blur, selfie, art work, statue, eye shadow, ... *many more*

	%	ms	MB
server	97.4	635.0	270.0
smallest	94.5	7.3	0.9
fastest	94.2	5.6	3.3
well rounded	95.5	19.0	3.3

FeelNet
Recognizing human emotions
go/feelnet
go/feelnet-lite
go/video-emotion-model
go/sentire
go/affective-computing
bjou@
gautamprasad@

We develop computational sensors for the full human affective experience, including facial expressions, gesture expressions, affective speech/voice, contextual/environmental cues, and their multimodal integration.

Currently, we support several discrete facial emotion categories in FeelNet and legacy emotions in FaceSDK. This is a companion effort to the Sentire project.

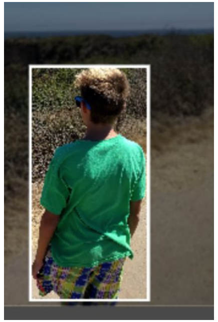
Emotions
amusement, anger, concentration, contentment, desire, disappointment, disgust, elation, embarrassment, interest, pride, sadness, surprise

PersonAttributes
Generic person attributes
go/person-attributes
liuti@

Demographic
age, gender

Objective attributes
hair color and style, clothing color and pattern and style, ...

Samples



male
10
short, straight hair
green, short sleeve,
t-shirt

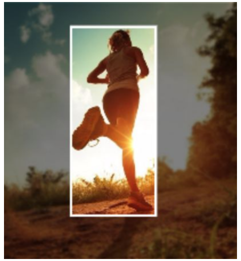
female
24
long, straight, brown hair
long sleeve, gray upper clothing
black lower clothing

ActNet
Recognizing activities
go/actnet
go/actionloc
liuti@

Actions
clapping hands, dancing, eating, hugging, jumping, kicking, kissing, running, shaking hands, singing, throwing objects, thumb-up gesture, toasting, v-sign gesture, waving hands

Samples


jumping, waving hands


running


dancing, waving hands

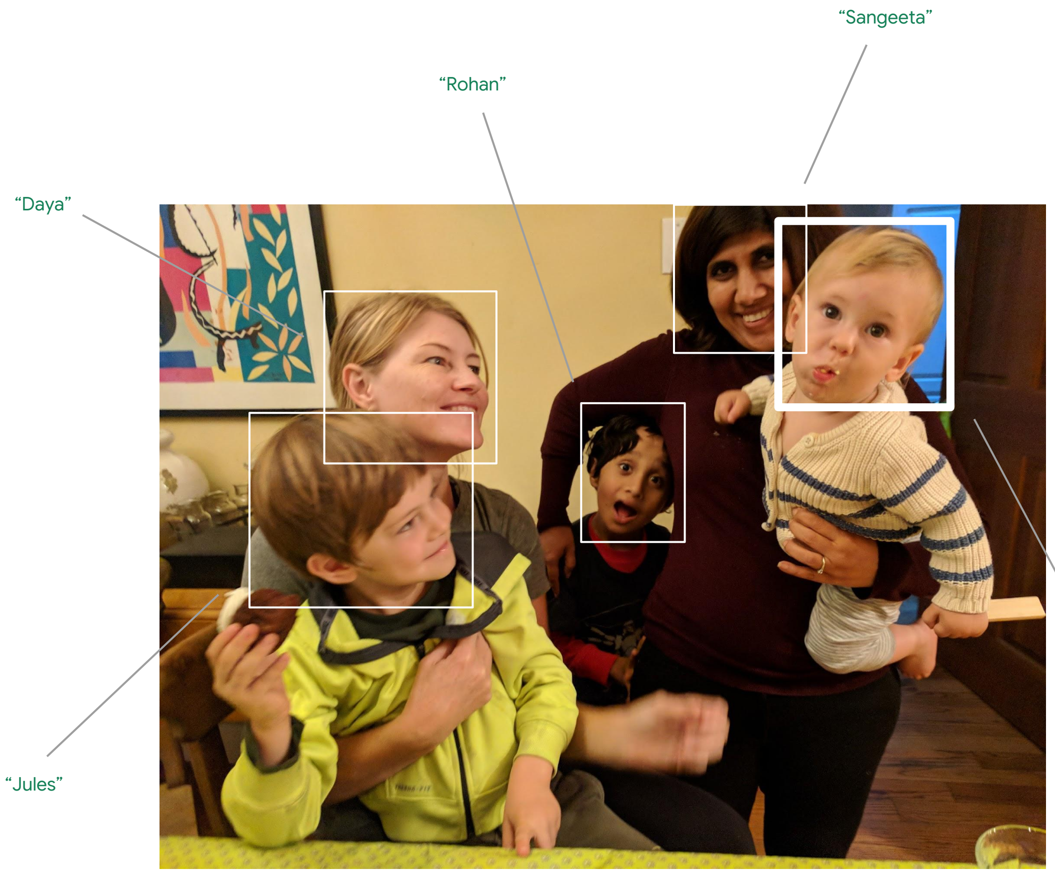
Detection

Description

Recognition

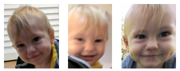
Geometry

Fairness



FaceNet embeds face thumbnails into a space, where faces of the same identity are closer together than faces from different identities.

"Lucas"



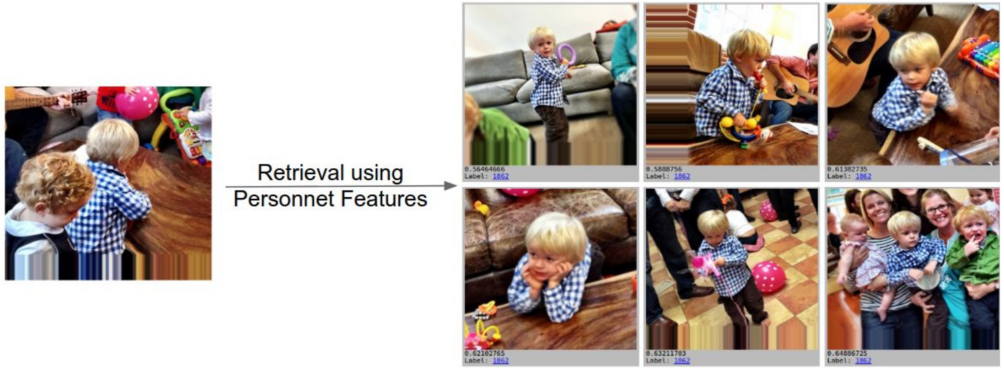
Integrations
Photos - clustering
Nest Cam - familiar faces
Clips - familiar faces
YouTube - Eastwood

FaceNet
Facial recognition
go/facenet
fschroff@
dkalenichenko@

This project addresses face recognition and face authentication. It provides the core modules for face recognition which are used or being integrated into several projects: Google Photos,Trusted Face in Android, Nest cam, Eastwood, Stellar

PersonNet
Whole person recognition
go/personnet
@liuti

This project aims to match different image instances of people. This work uses the whole body image, which allows matching on person images when the face is not visible. Currently, the project concentrates on building an image embedding model called PersonNet.



Responsibilities

Detection

Description

Recognition

Geometry

Fairness

Graphic compositing is possible with landmarks, pose estimation and segmentation.

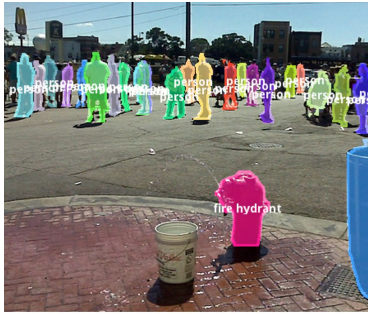
Integrations
Camera - real time bokeh
Daydream - tracking, compositing

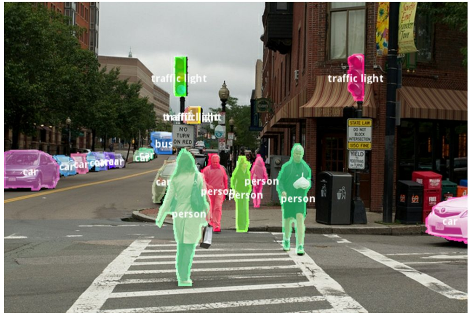
MaskLab

Instance segmentation

go/masklab

lcchen@

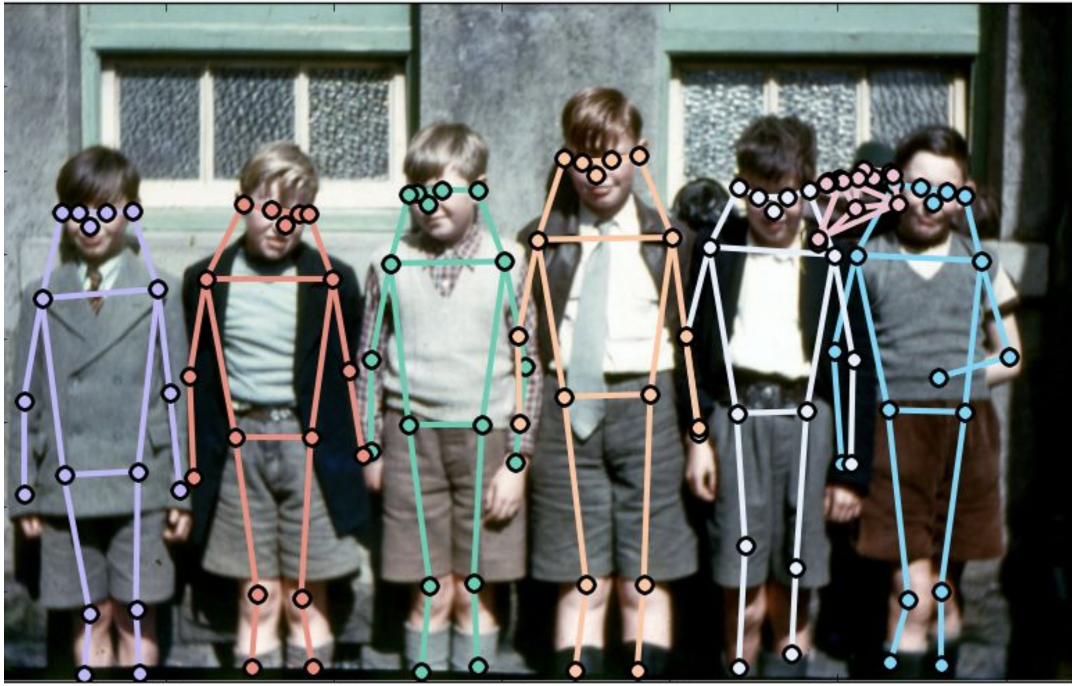




Pose estimation

go/posenet

gpapan@
tylerzhu@



University of Human Sensing

Make Human Sensing more inclusive.

Detection

Description

Recognition

Geometry

Fairness

Identify and remove performance biases in the ML system used for human sensing

Identify and handle representation biases in multimedia corpora across Google.

UHS Diagnostic Tool
Used to diagnose models and visualize differences across subsets.

UHS (Race) Classifier v0
WIP Google-wide usage policy and infrastructure. Trained on WebFaces and Eastwood.

Oakley

go/oakley

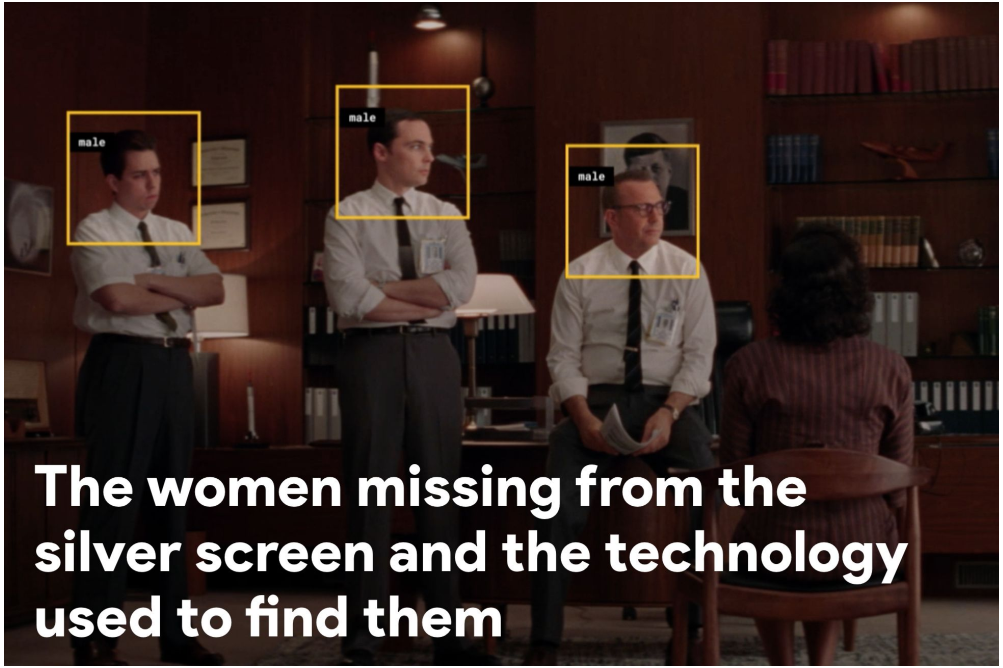
Detection

Description

Recognition

Geometry

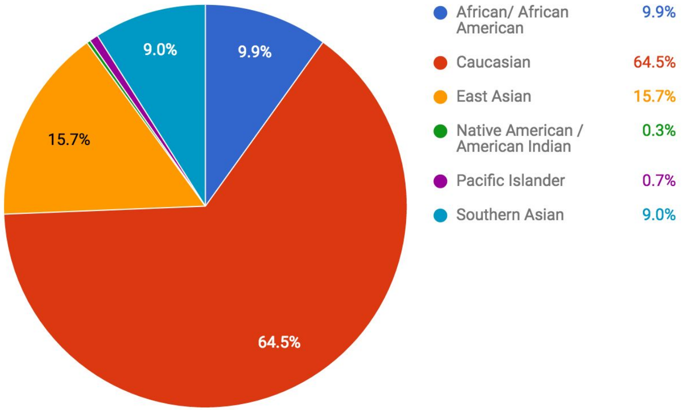
Fairness



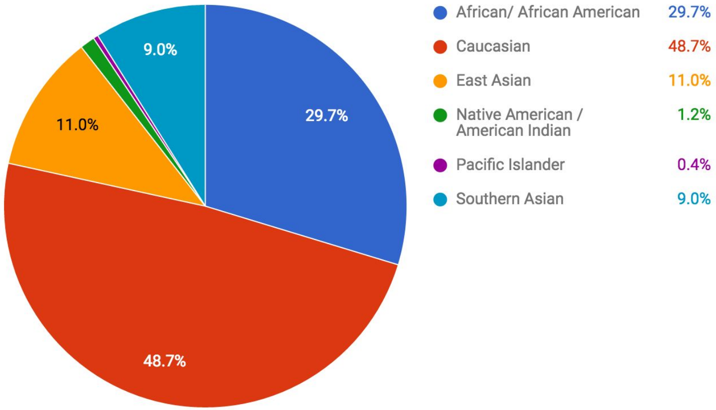
UHS - University of Human Sensing

Race distributions of datasets.

WebFaces 165,483 images

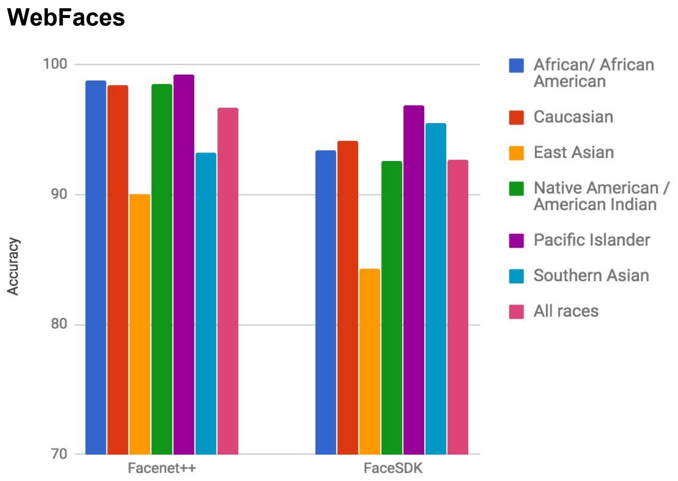


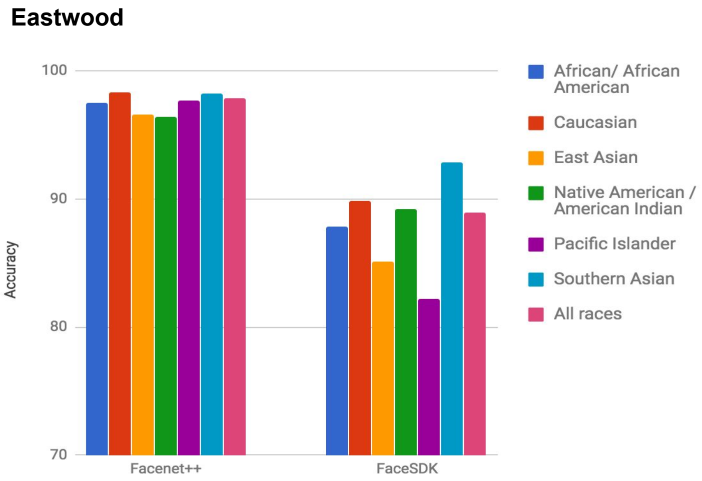
Eastwood 36,608 images



UHS - University of Human Sensing

Gender prediction accuracies.





What’s next

Smaller models for stream use cases
Better NN architecture for HW acceleration
7bit quantization
More unbiasing, more attributes

Agenda

- Welcome
- Fairness: pbrandt@
- Human Sensing: dkaram@
- **ML and Data**: ivanku@
- Crowd Computing: pocketaces@
- Natural Language: barakt@
- On-device: ingerman@
- Medical Applications: lhpeng@
- Getting to Launch: binghamj@
- Refreshing Conversations

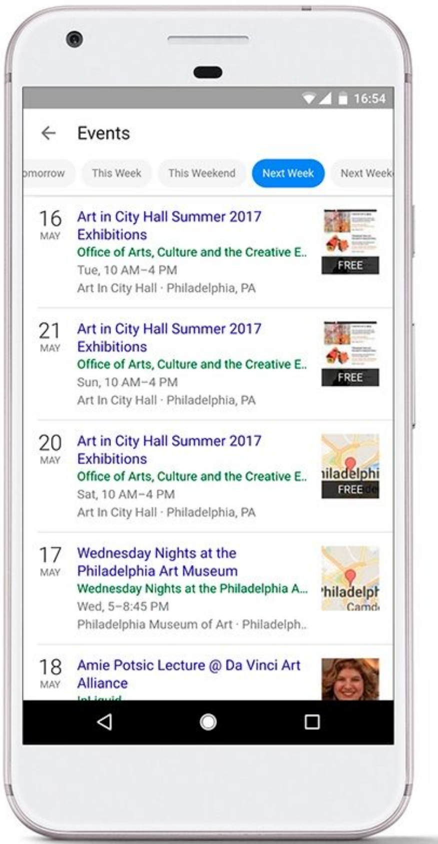
ML and Data

Ivan Kuznetsov (ivanku@)

1. Understand complexity
2. Think about data
3. Not everything is ML

Understand complexity

Example: Events Search



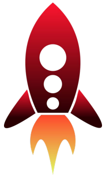
Where does the data come from

```
<script type="application/ld+json">
{ "@context"  : "http://schema.org",
  "@type"     : "Event",
  "name"      : "B.B. King",
  "startDate" : "2014-04-12T19:30",
  "location"  :
    { "@type"    : "Place",
      "name"     : "Lupo’s Heartbreak Hotel",
      "address" : "79 Washington St., Providence, RI" },
  "offers"    :
    { "@type"    : "Offer",
      "url"      : "https://www.etix.com/ticket/1771656" } }
</script>
```

Markup - **75%**

ML extractions - **25%**

But it took a while to get here - ML is hard



June 2015

October 2015

November 2016

July 2017

WhatsHapp project starts

Early prototype presented @ Google Research Conference

ML models quality satisfies Search launch criteria

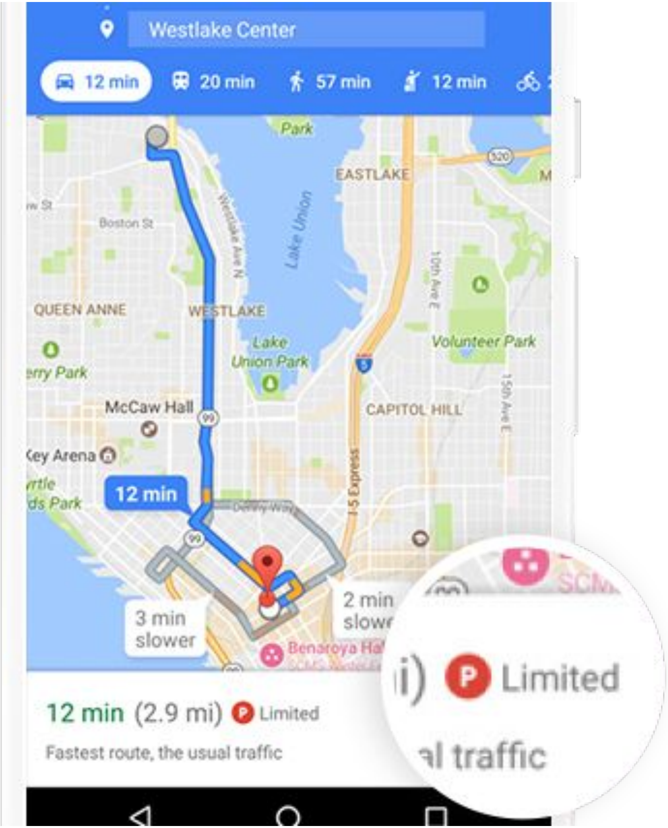
ML models production integration complete - launch!

Think about trade-offs

- Acquire data - extract data
- Infer attributes - ask users
- Do you have a long tail where ML can help?

Think about data

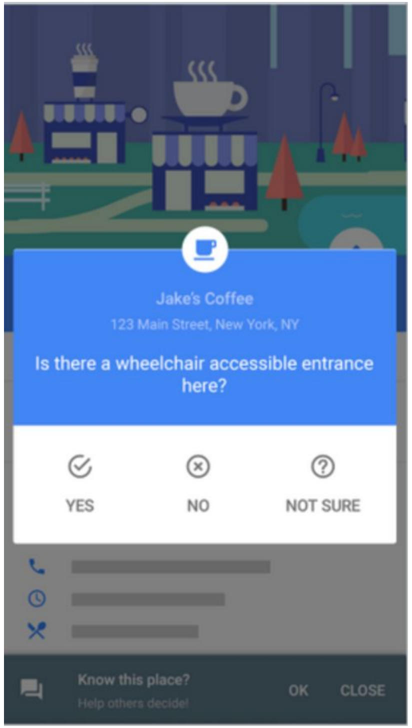
Example: parking difficulty



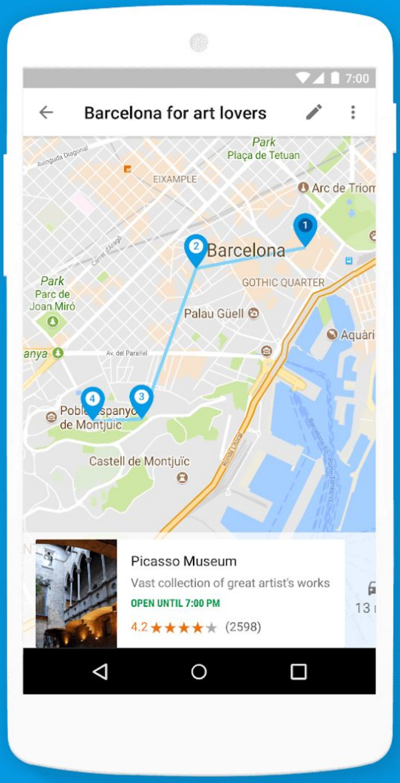
ML models need data for training

Possible approaches:

- Crowdsourcing
    - Riddler
    - Google Consumer Survey
- Use open data
- Purchase datasets
- Create datasets



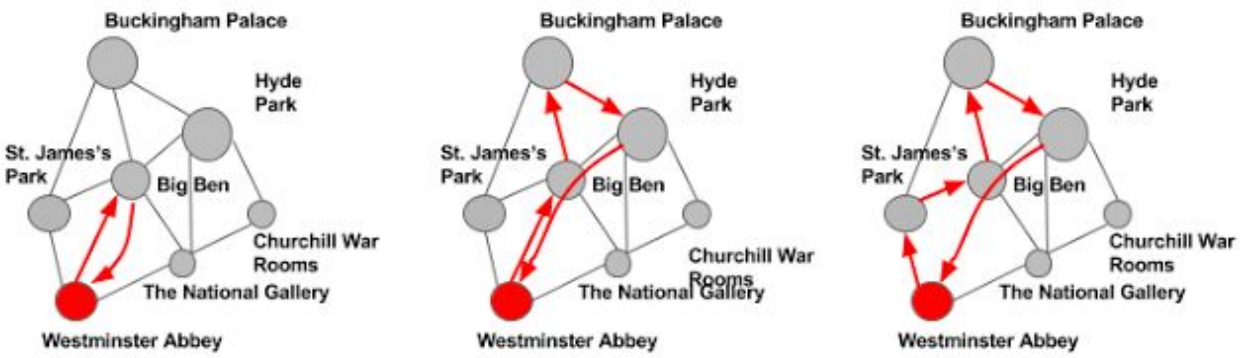
Not everything is ML



Plan your day like magic

Google Trips makes it easier than ever to plan and organize your trips. It automatically maps out a half day or a full day with suggestions for things to see and do. Don't like what you see? Tap the "magic wand" to see more nearby sights. Each tap of the wand gives you a fresh set of nearby attractions.

Bridges of Königsberg and Traveling Salesman



Starting at Westminster Abbey, we decide to add Big Ben.

Next we add Buckingham Palace followed by Hyde Park. Note we bypass Big Ben on the way back along the shortest path to the starting point of the tour.

Finally, we add St. James's Park, and Christofides' algorithm allows us to connect it at the most efficient point of the tour so far.

GOOGLE'S NEW VACATION APP WAS 280 YEARS IN THE MAKING

Where to learn more about algorithms work

Google Optimization Tools - developers.google.com/optimization

Operations Research Team - go/or

Market Algorithms Team - go/market-algorithms

Discrete Algorithms Team - go/discrete-algorithms

1. Understand complexity
2. Think about data
3. Not everything is ML

Agenda

- Welcome
- Fairness: pbrandt@
- Human Sensing: dkaram@
- ML and Data: ivanku@
- **Crowd Computing**: pocketaces@
- Natural Language: barakt@
- On-device: ingerman@
- Medical Applications: lhpeng@
- Getting to Launch: binghamj@
- Refreshing Conversations

Crowd Computing

Anurag Batra (pocketaces@)

Human Computation Landscape

3 Core “Classes” of Data Collection with Corresponding Platforms/Teams

Paid Raters

- Crowd Compute/VSEval
- Furball
- Ewoq

Global Crowdsourcing

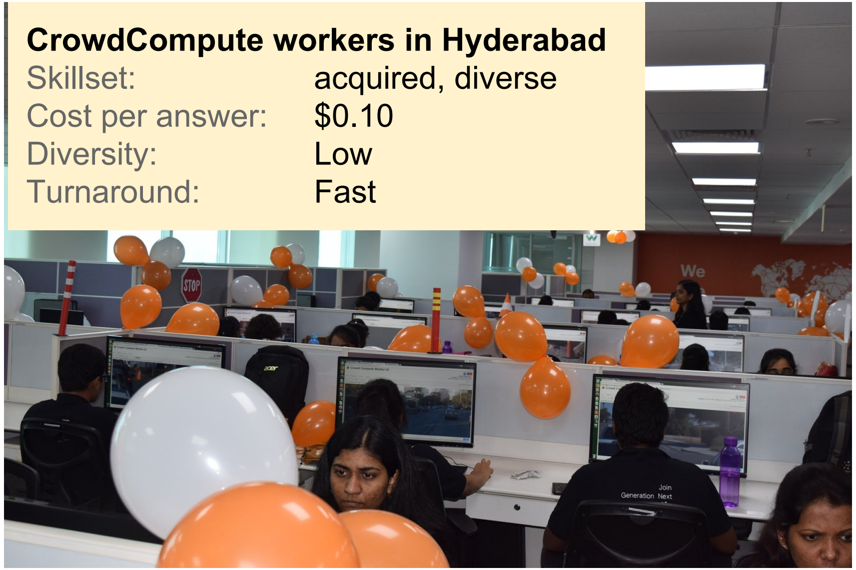
- Village (Crowdsource app)
- Google Opinion Rewards
- Endor (MTurk Raters)

Specialist Raters

- Pygmalion
- Speech Data Ops

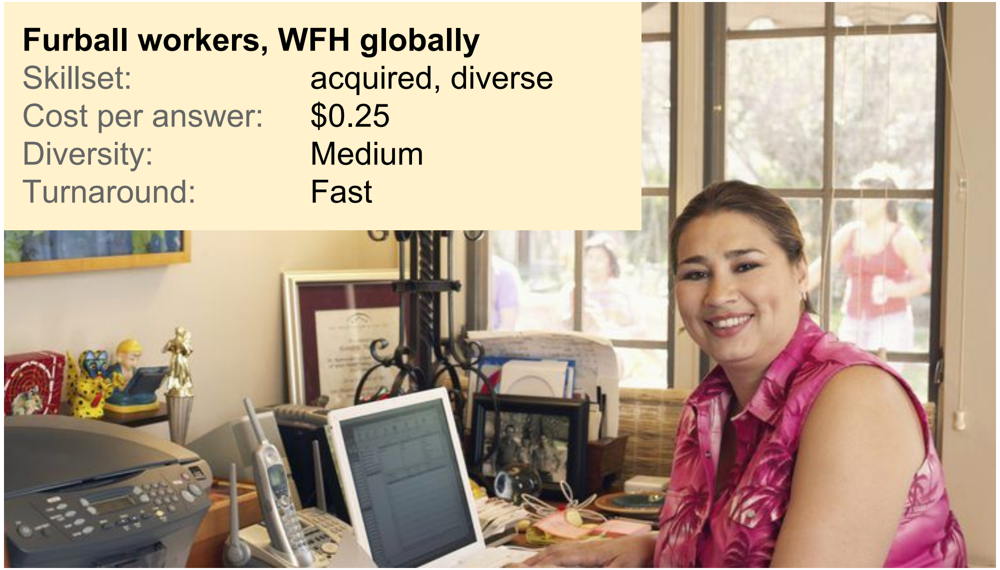
Diverse, Large-Scale Operations

- Platforms active in every country but North Korea
- Multiple man-years of data collected per day on Crowd Compute alone



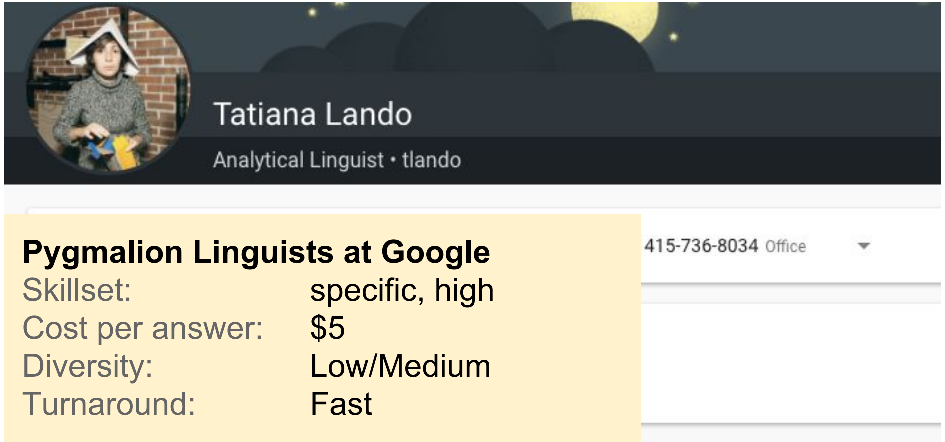
CrowdCompute workers in Hyderabad

Skillset:	acquired, diverse
Cost per answer:	$0.10
Diversity:	Low
Turnaround:	Fast



Furball workers, WFH globally

Skillset:	acquired, diverse
Cost per answer:	$0.25
Diversity:	Medium
Turnaround:	Fast



Pygmalion Linguists at Google

Skillset:	specific, high
Cost per answer:	$5
Diversity:	Low/Medium
Turnaround:	Fast

Crowdsource app users globally

Skillset:	generic
Cost per answer:	$0
Diversity:	High
Turnaround:	Slow

All you need to do to get data

[go/get-hcomp-data]()

1. Provide some details about what you need
2. Automatically generates a tracking bug and CCs stakeholders
3. We refer you to the right platform and team

Stuff that you'd get help with

- Task design
- Budgeting
- Sample selection
- Diversity and Fairness advice
- Cataloging and sharing
- Privacy and Policy compliance
- Storage and Deletion advice

Getting diverse and accurate data

Raters’ opinions shape your products

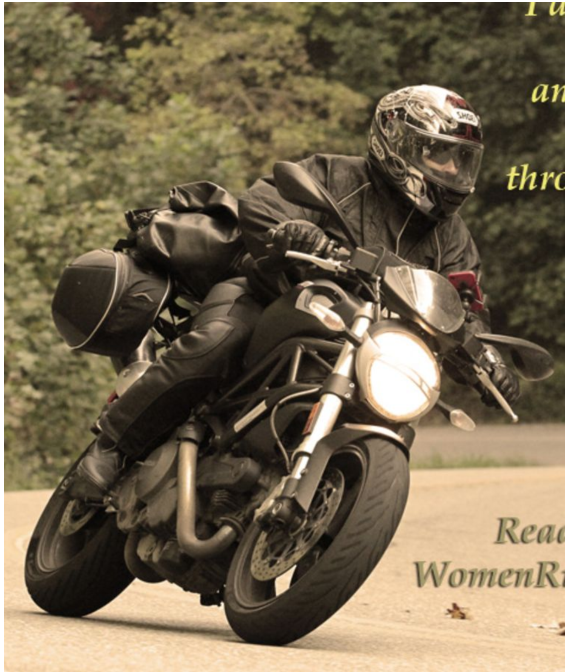
Know thy rater!

What is a bike?

In the US

In India

What is safe motorcycle riding?



In the US

In India

What is great weather to be outdoors?

In the US

hitesh nanda
@hitesh4ualwayss

Follow

awesome weather in my city...#Delhi

In India

What can be done to mitigate personal biases?

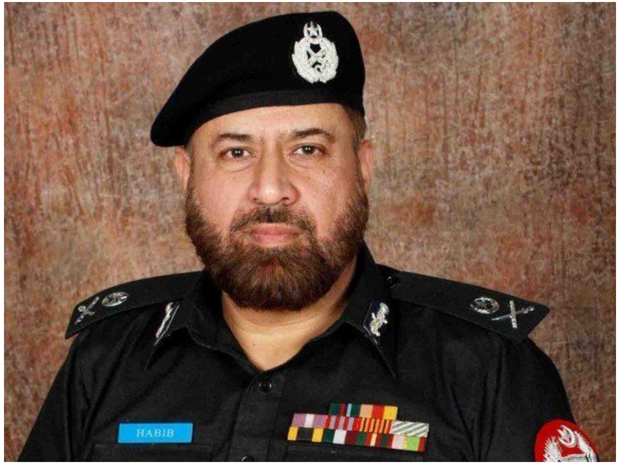
More awareness, better context

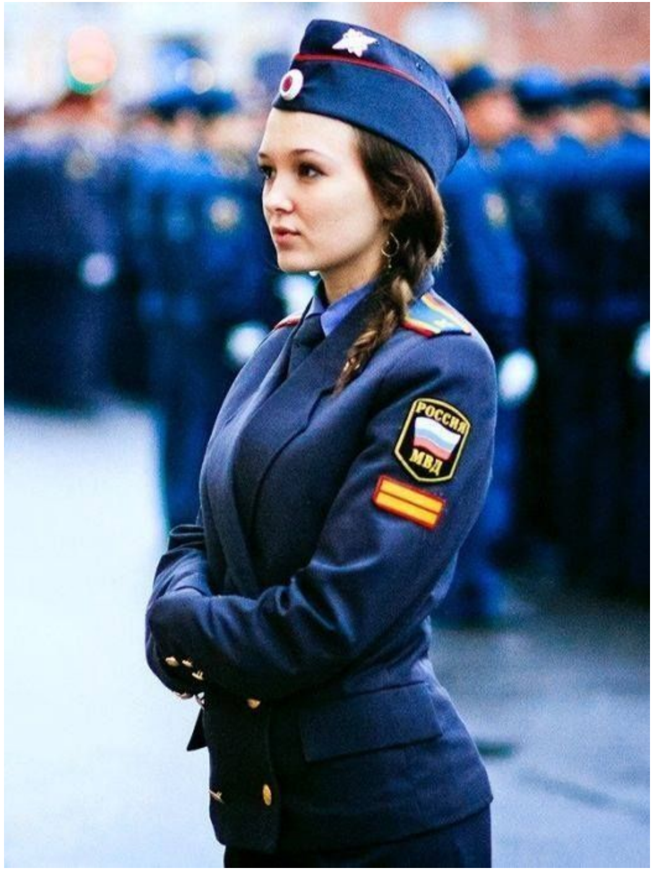
Police Officer

Rater Instruction: *Please use an image search service to find images that represent this category. Each image should contain a single human being.* *No additional context given.*

Police Officer

Rater Instruction: *Please use an image search service to find images that represent this category. Each image should contain a single human being.* *Raters given prior context on diversity.*





go/get-hcomp-data

Agenda

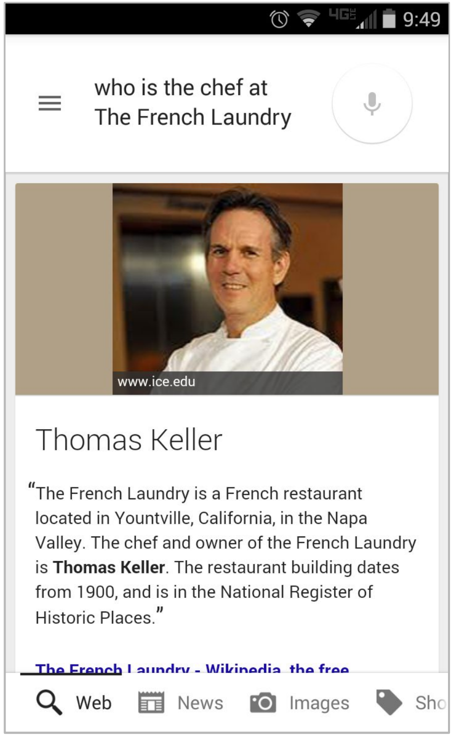
- Welcome
- Fairness: pbrandt@
- Human Sensing: dkaram@
- ML and Data: ivanku@
- Crowd Computing: pocketaces@
- **Natural Language**: barakt@
- On-device: ingerman@
- Medical Applications: lhpeng@
- Getting to Launch: binghamj@
- Refreshing Conversations

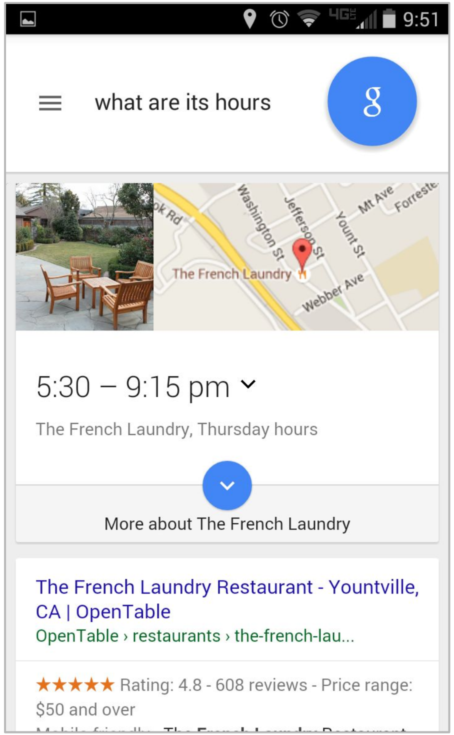
Natural Language Understanding
Barak Turovsky (barakt@)

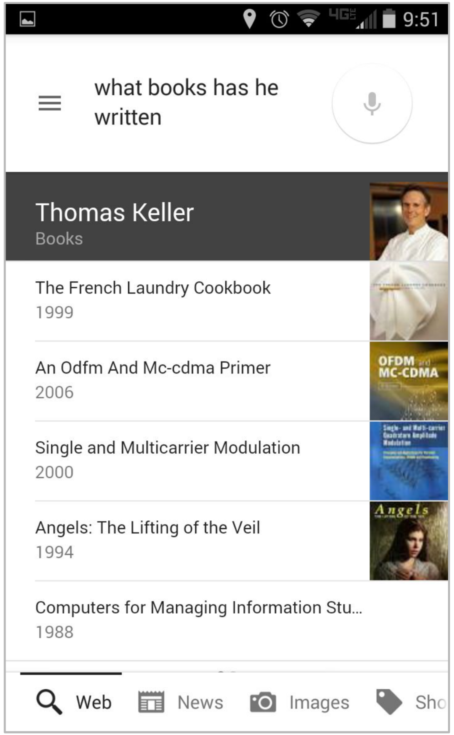
MISSION: USE CUTTING EDGE AI TO UNDERSTAND HUMAN LANGUAGE

MONOLINGUAL

Conversational Search

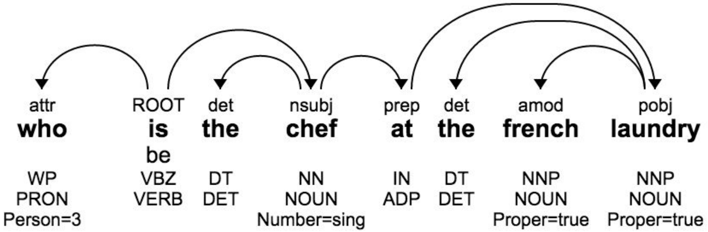


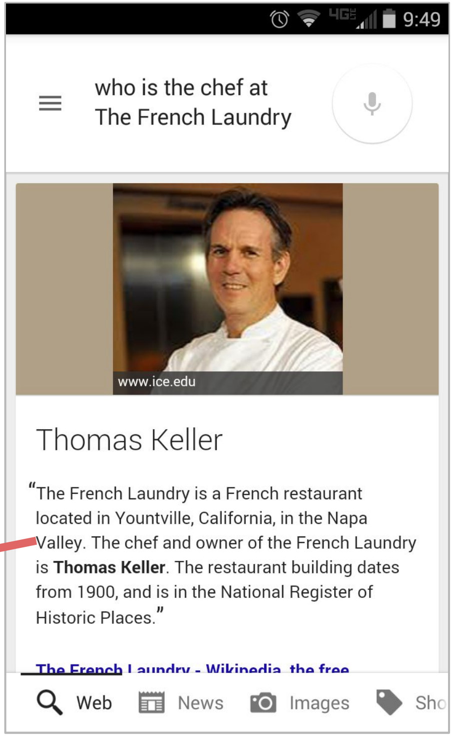




[live on your phone]

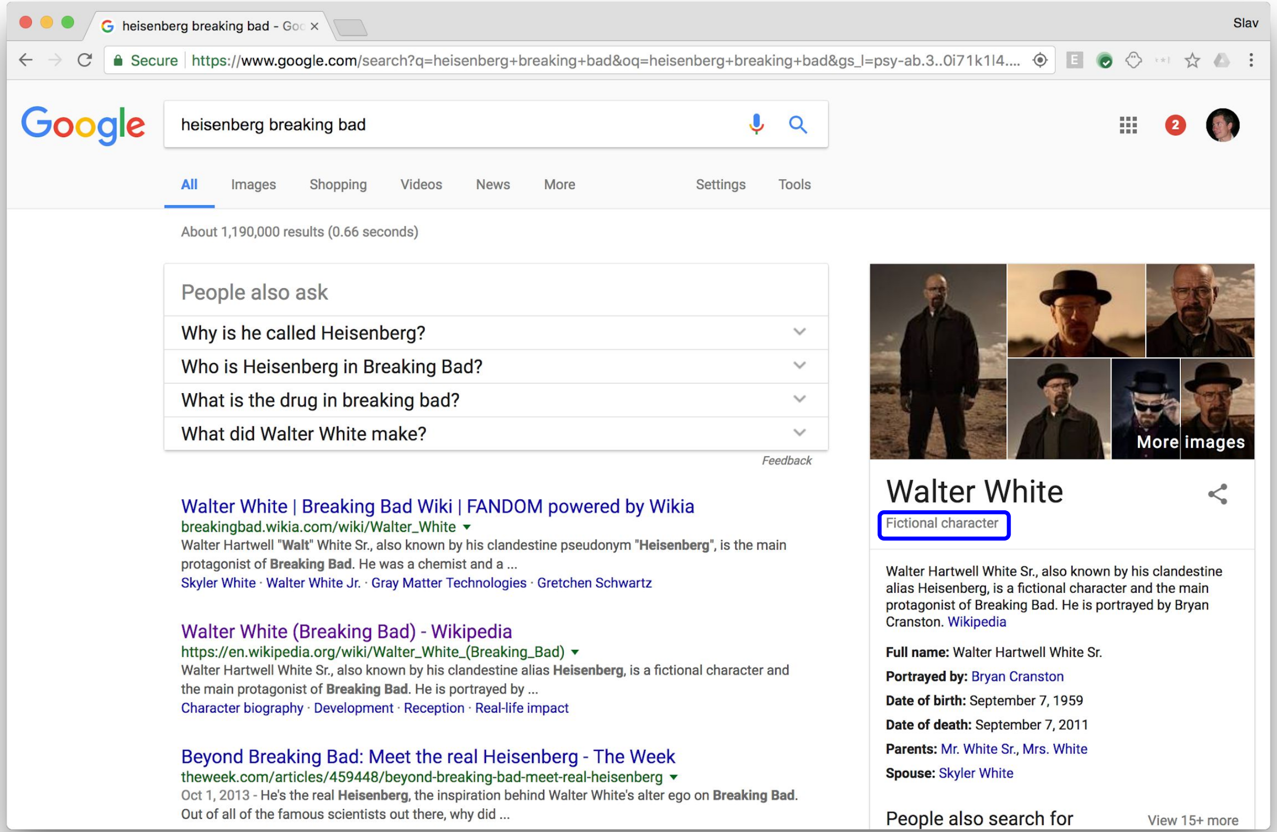
Understanding Queries





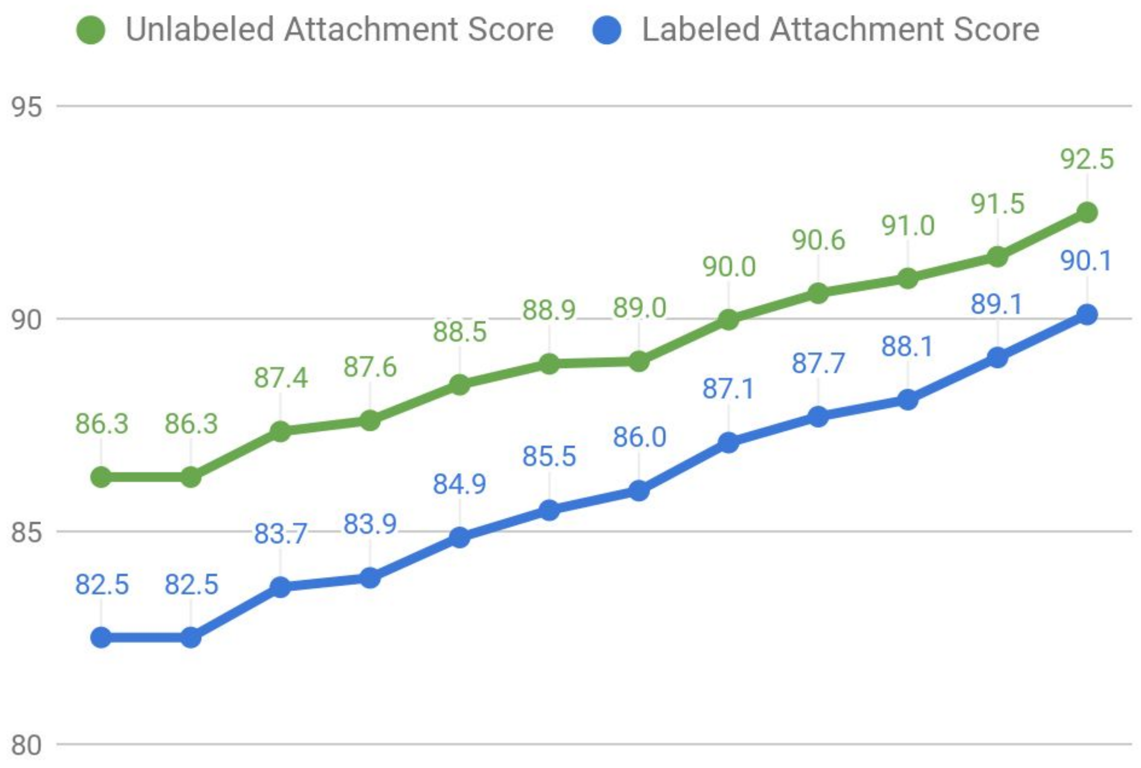
[live on your phone]

Question Answering

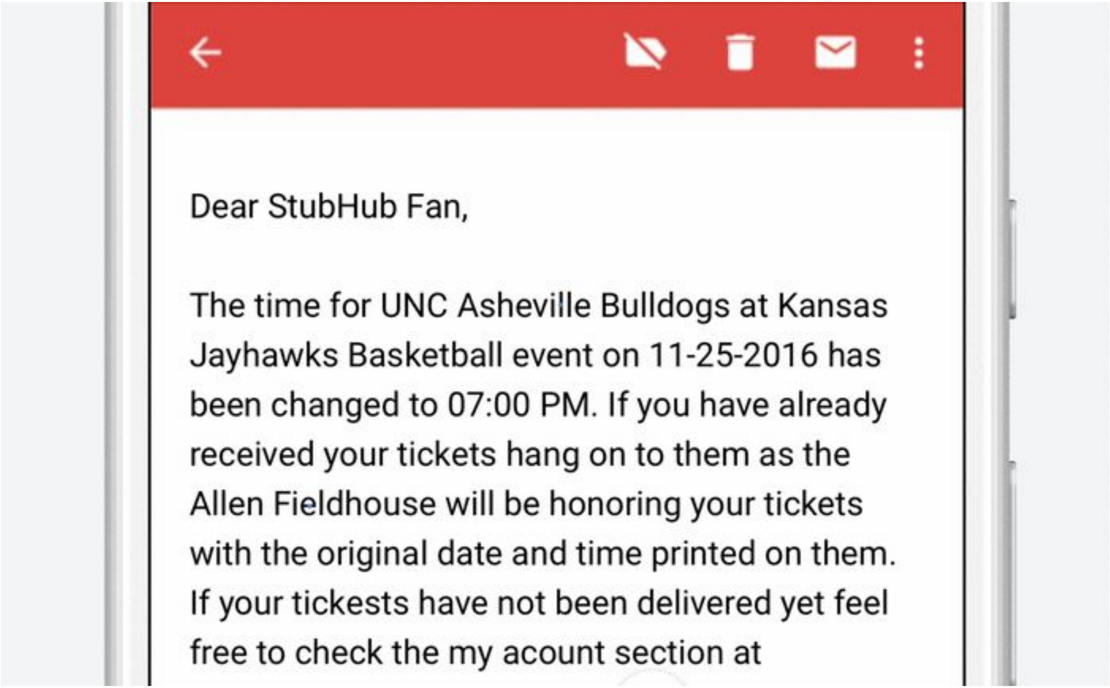


Reading the Web

English Parsing Accuracy Progress (on Web Data)



On-Device NLP



MONOLINGUAL:
GOOGLE TRANSLATE

Why we care about translations

50%

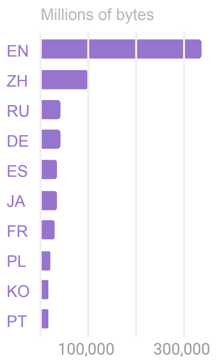
Of the internet content is in English

20%

Of the world population has some English skills

80%

Of the web is in only 10 languages



In order to make the world's information accessible, we need translations

Google Translate is a truly global product

1.2B+

Monthly
active users

95%

Users from
outside US

45%

Of our users
are on mobile

140B

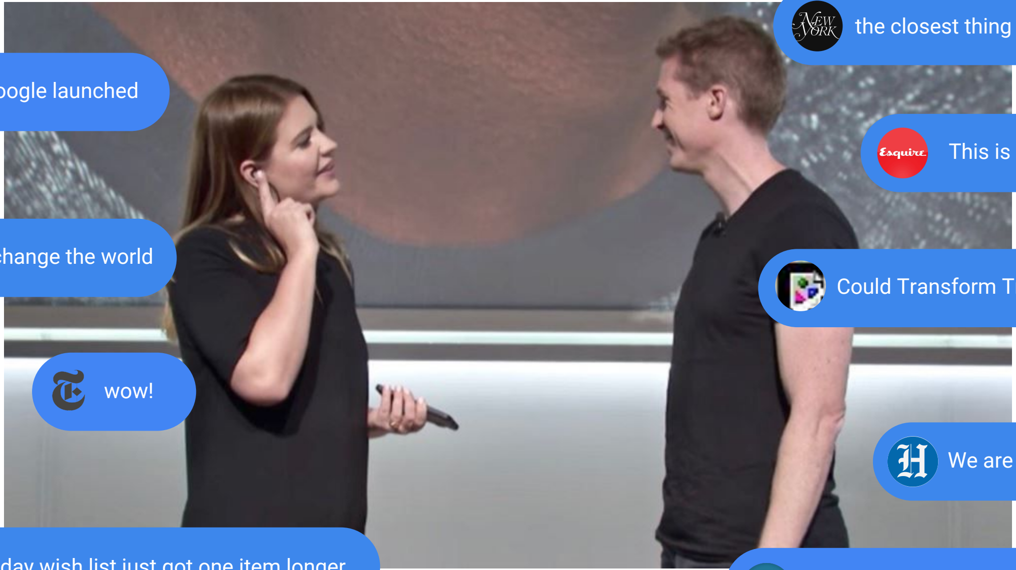
Translated
words/day

550M

Mobile app
downloads

We're about more than just text translations

We help people overcome language barriers



the closest thing we’ve seen to the Babel Fish

The most important gadget Google launched

This is some seriously sci-fi stuff

Google's Pixel Buds translation will change the world

Could Transform Travel With Real-Time Translation

wow!

A powerful differentiator

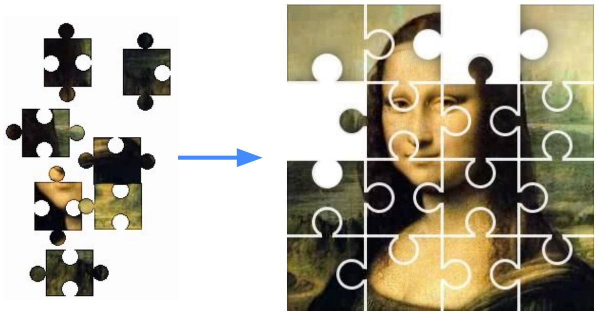
We are living in the future

Your holiday wish list just got one item longer

removal of language barriers is closer than ever

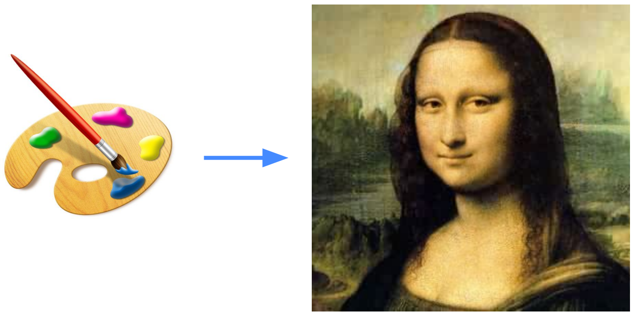
Neural Machine Translation

3rd generation machine translation system



Phrase-based machine translation

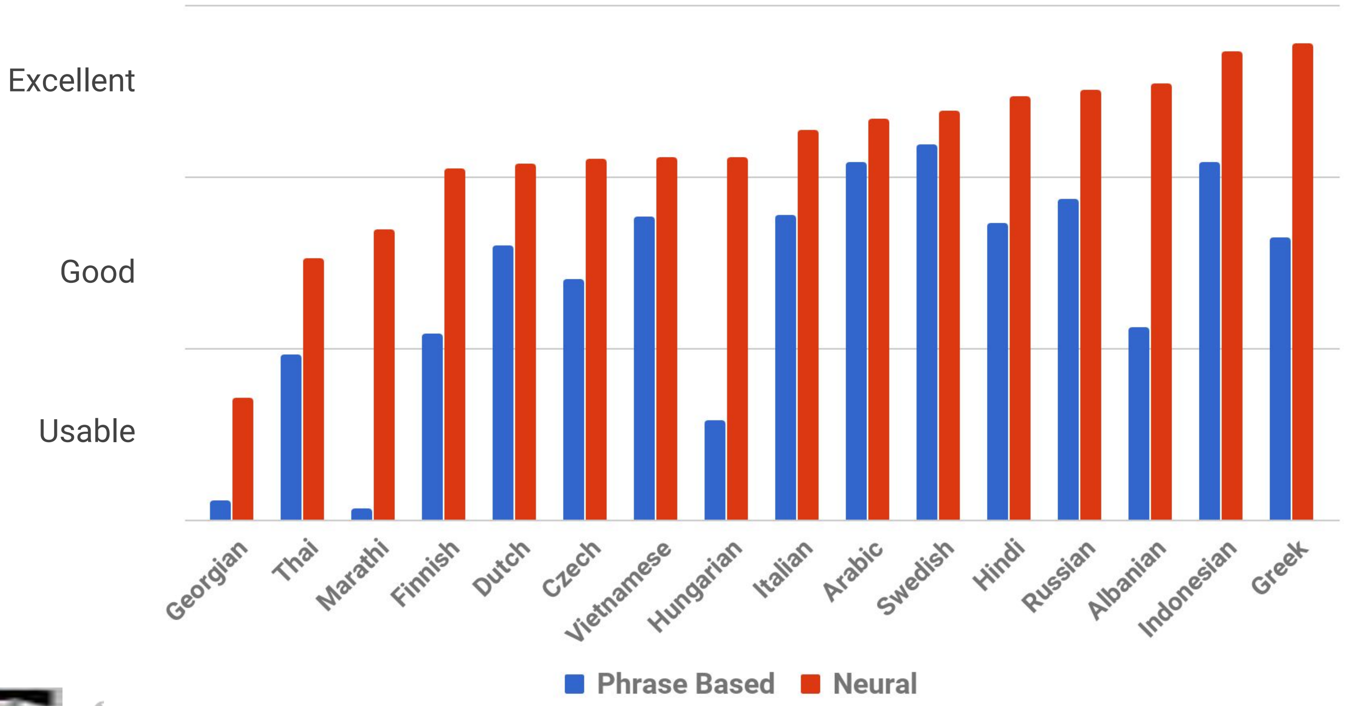
Discrete, local decision



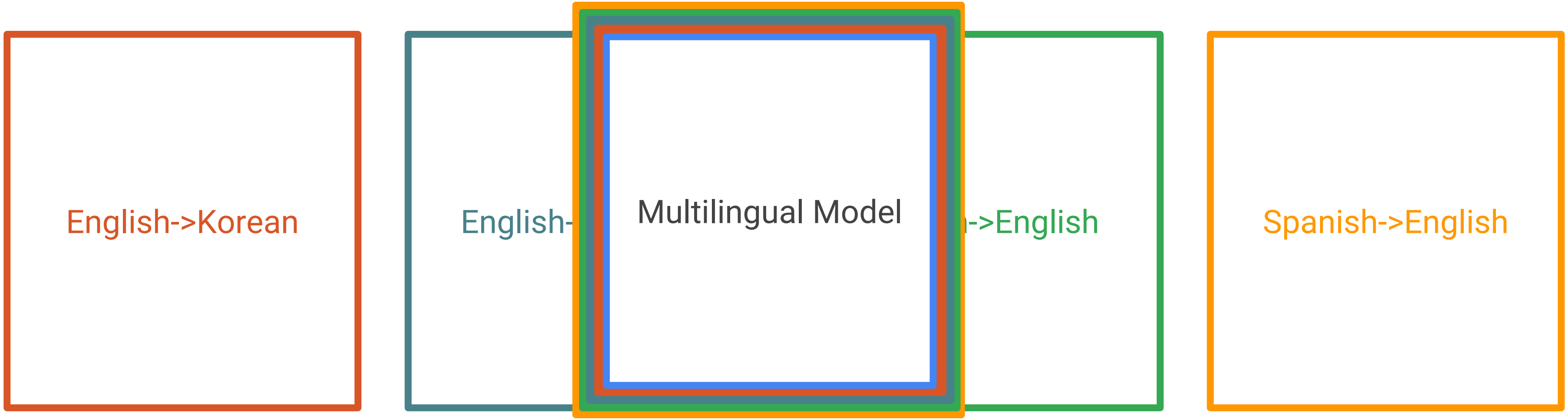
Neural Machine Translation (NMT)

Continuous, global decision

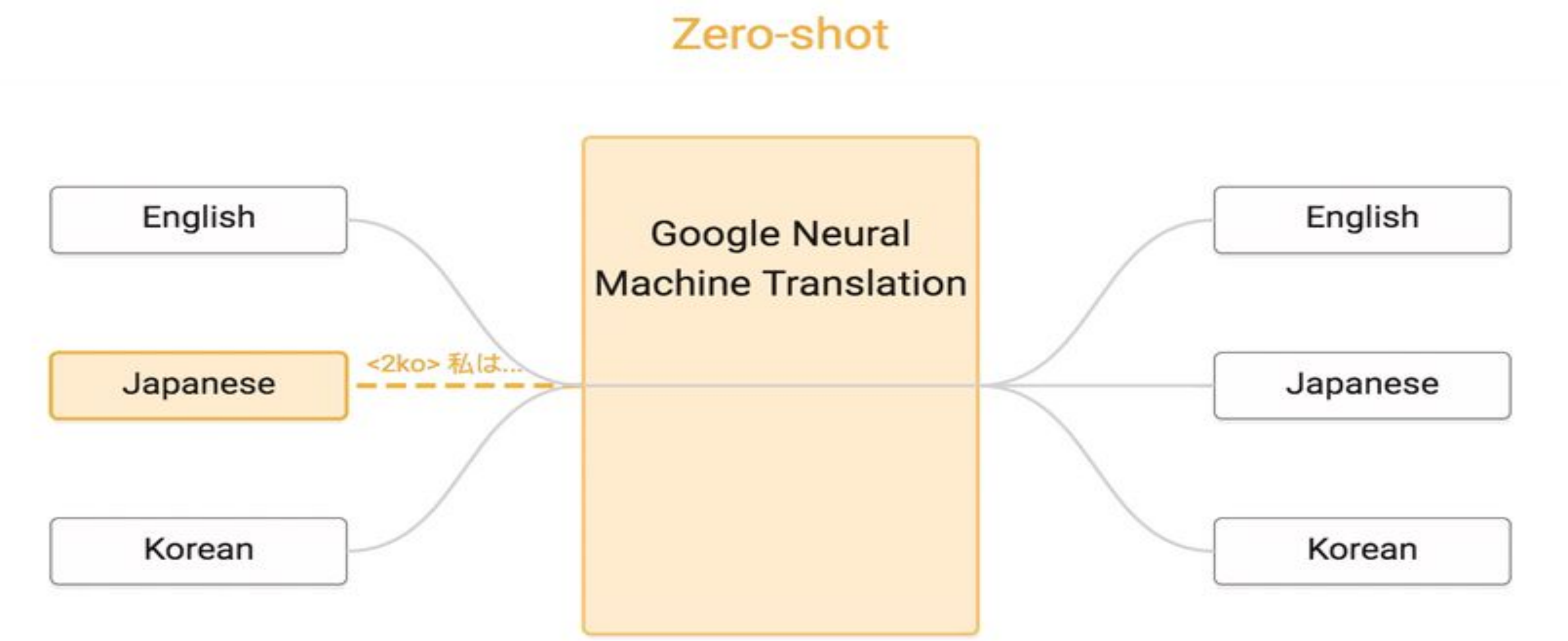
With NMT, we have achieved the most disruptive jump in translation quality in last 10 years!



Multilingual models



Zero-shot translation

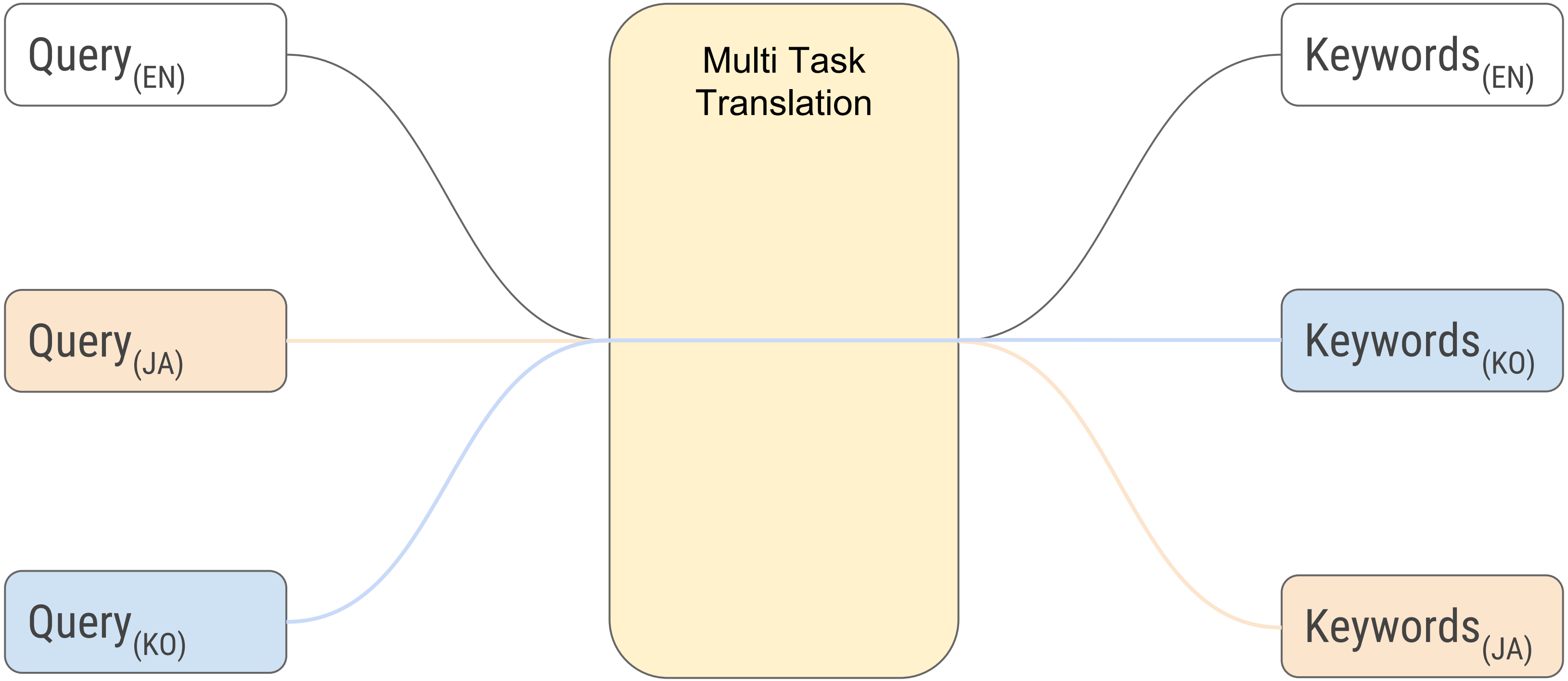


CAN WE DO BOTH?

MULTI TASKS ML

Semantic keyword-based targeting for ads

“Show my ads on queries expressing the ***concept*** conveyed by my keyword.”



Agenda

- Welcome
- Fairness: pbrandt@
- Human Sensing: dkaram@
- ML and Data: ivanku@
- Crowd Computing: pocketaces@
- Natural Language: barakt@
- **On-device**: ingerman@
- Medical Applications: lhpeng@
- Getting to Launch: binghamj@
- Refreshing Conversations

On-device
Alex Ingerman (ingerman@)

This talk's purpose in life

To **persuade** you that on-device ML is important and *different*

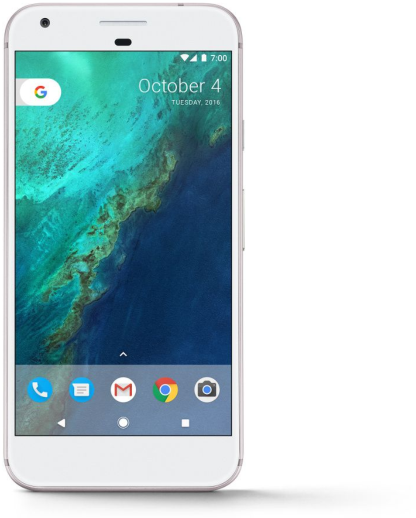
To **introduce** three approaches for on-device ML

To **discuss** the applications and available technology

What’s a device, anyway?

Google

Phones



Phones: very *personal* computers

```
2015: 79% away from phone ≤2 hours/day[1]
      63% away from phone ≤1 hour/day
      25% can't remember being away at all
```

Plethora of sensors

Innumerable digital interactions

[1] 2015 Always Connected Research Report, IDC and Facebook

Connected “Things”

IoT devices: they live among us

2016: 6.4 billion connected “things”. Of these,
4.0 billion consumer devices
1.1 billion cross-industry devices
1.3 billion industry-verticals

Consumer installations projected to grow fastest for the foreseeable future

Broadening capabilities

[2015 Connected Devices report, Gartner]

Robots

The robots are coming

2015:
- 5 million commercial and industrial robots
- 3.5 million consumer robots

2020:
- 12.2 million commercial and industrial robots
- 8.6 million consumer robots

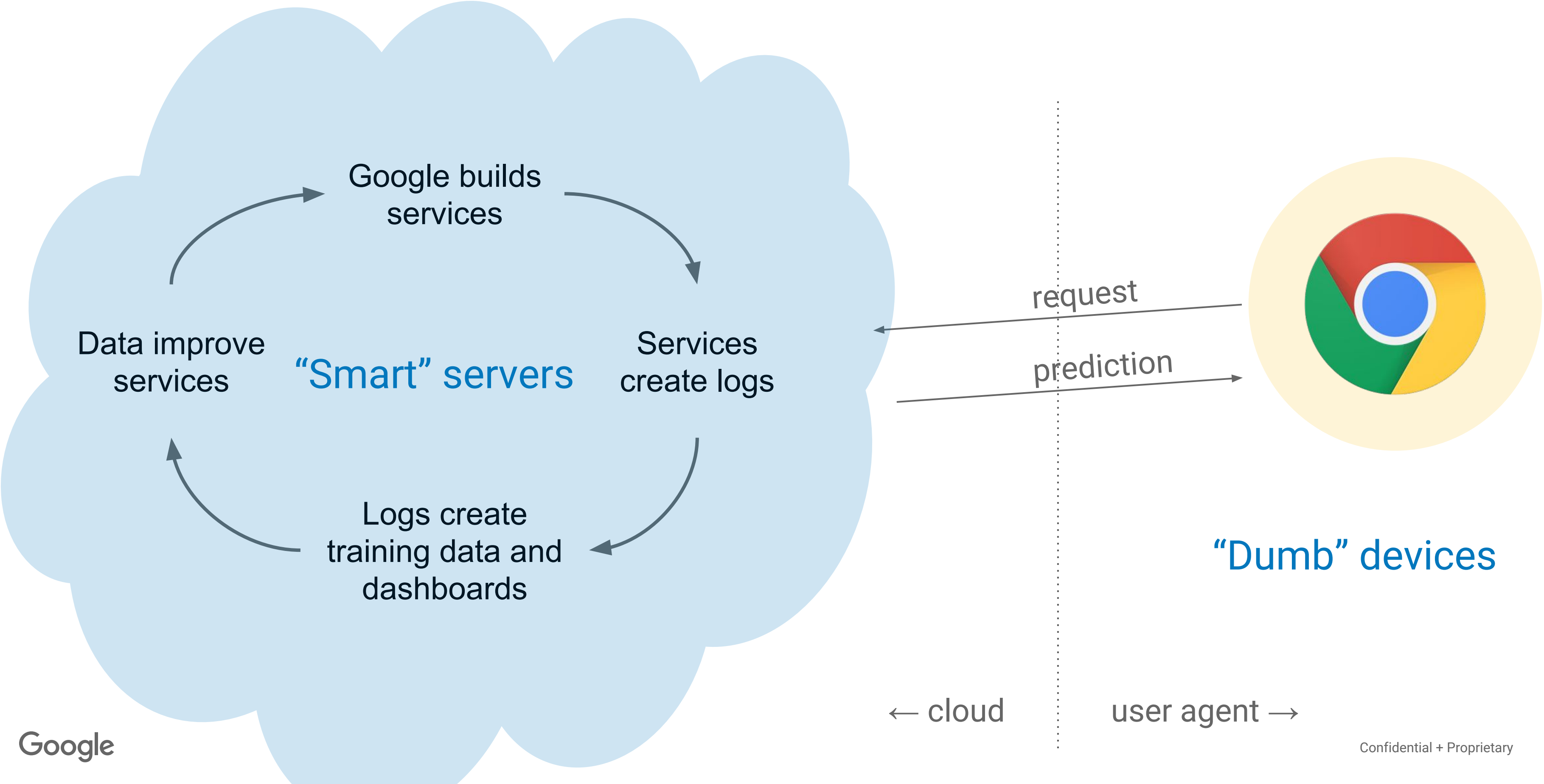
Self-driving cars and home robots projected to generate explosive growth

2017 Boston Consulting Group report

What’s so different about ML on-device?

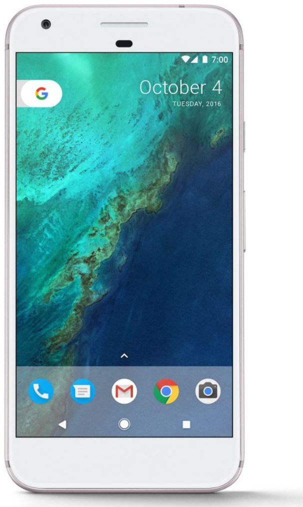
Google

Most machine learning today is done in the cloud



Devices are a new, different category of user agents

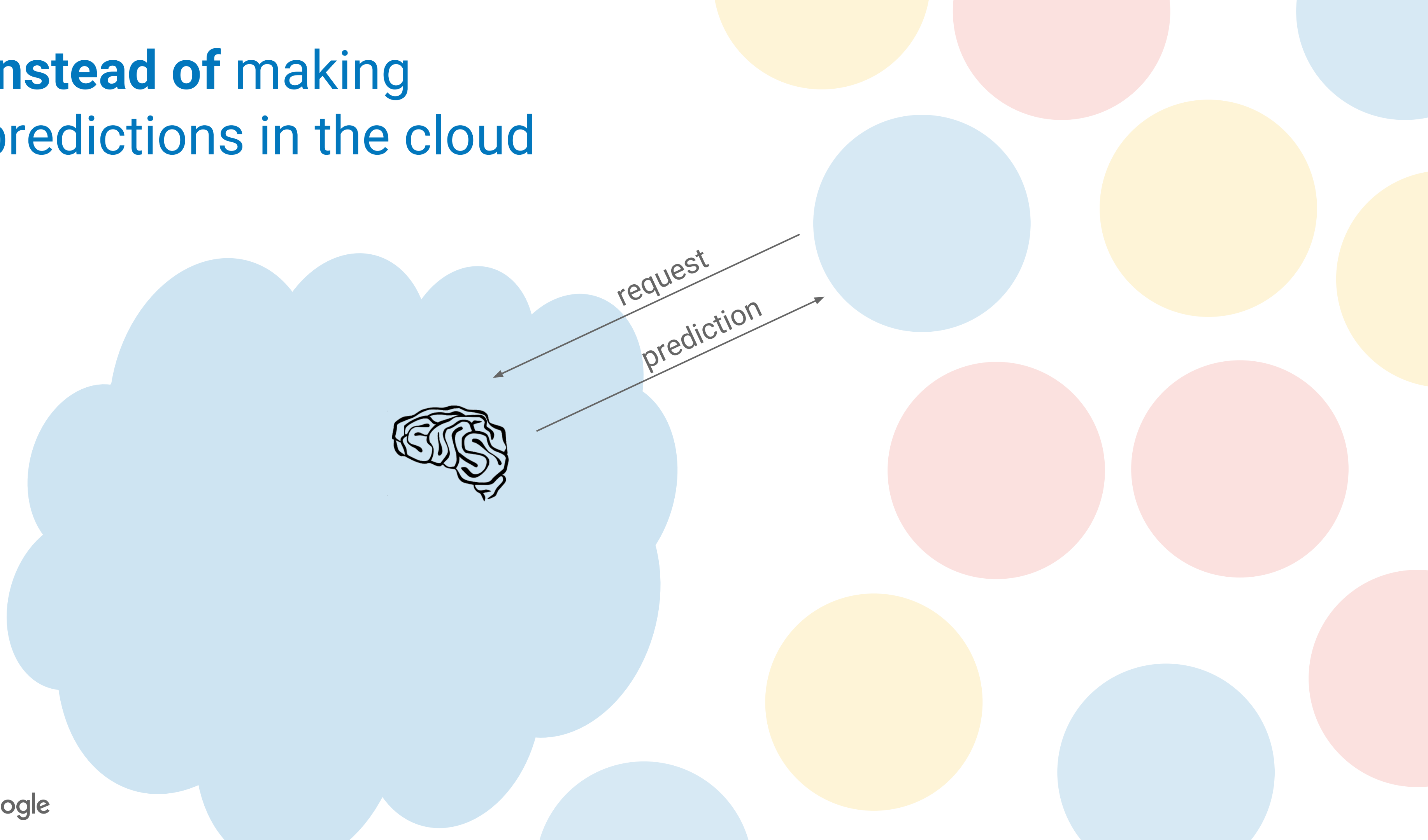
1. More personal
2. More numerous
3. More capable
4. More autonomous



We can build better products by leveraging devices' strengths!

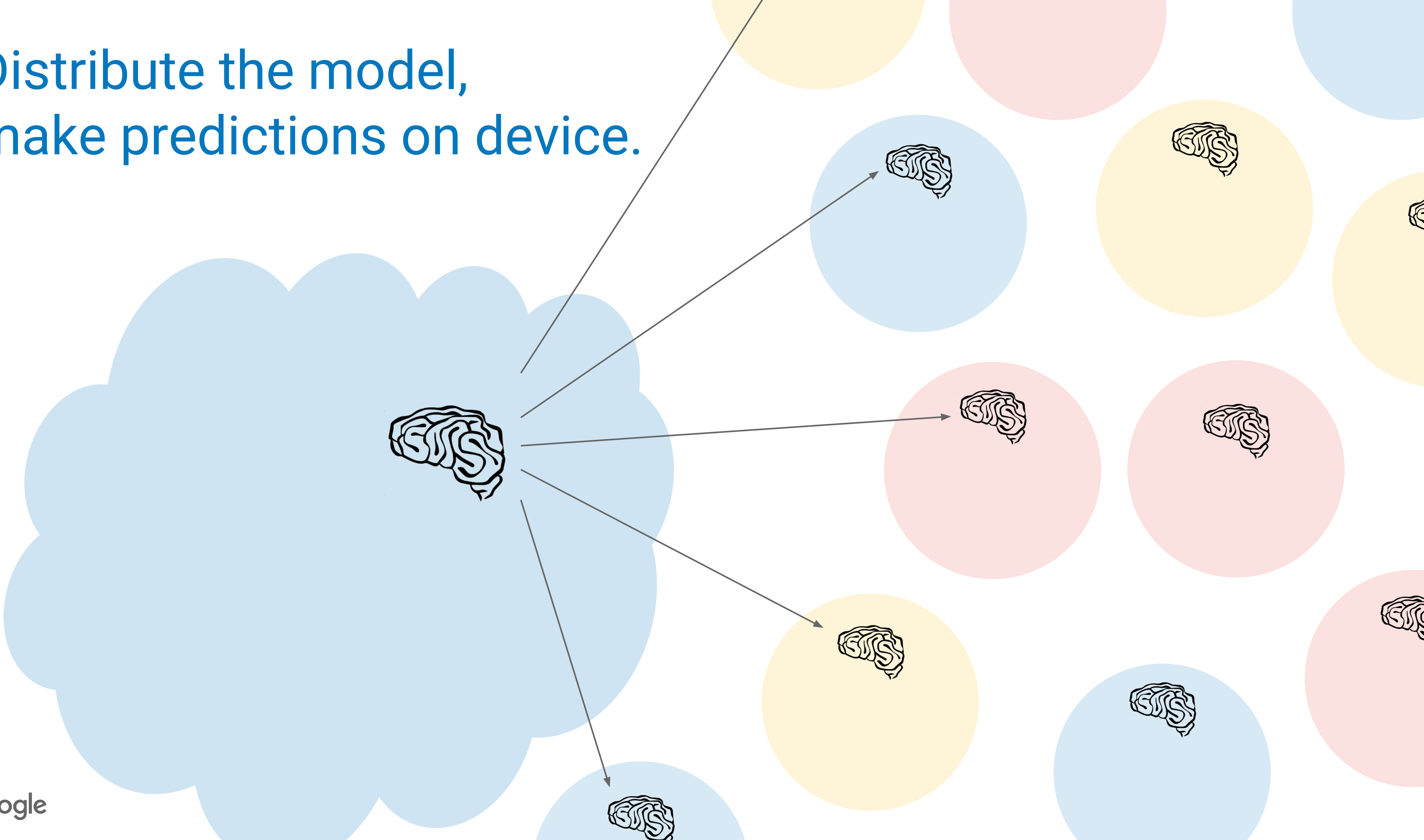
ca 2014+: On-Device **Predictions** (Inference)

Instead of making predictions in the cloud



Google

Distribute the model, make predictions on device.



Google

Latency

Data Caps

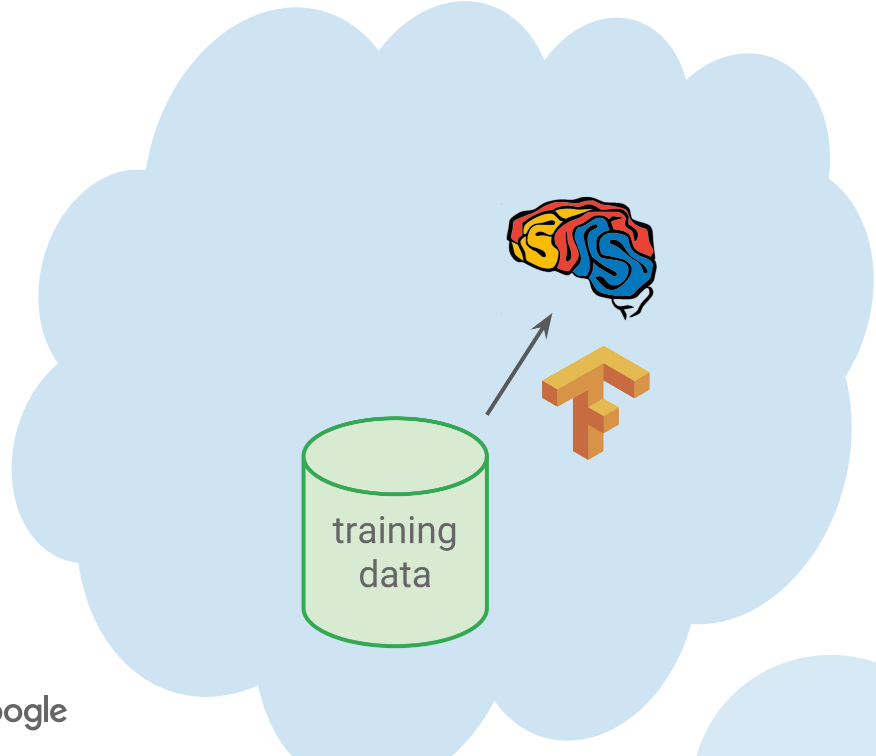
Privacy

Offline

Power

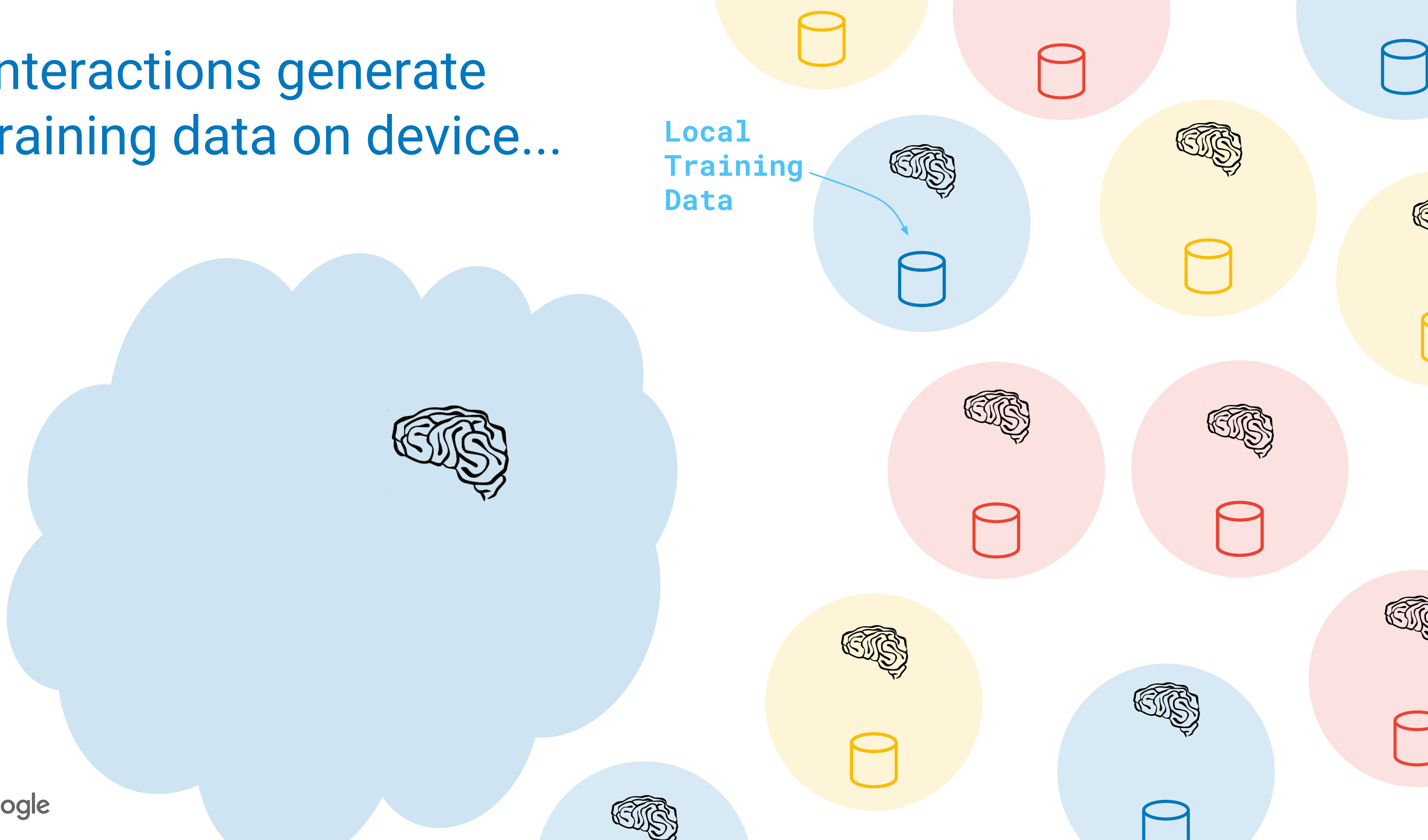
Sensors

How do we continue to improve the model?

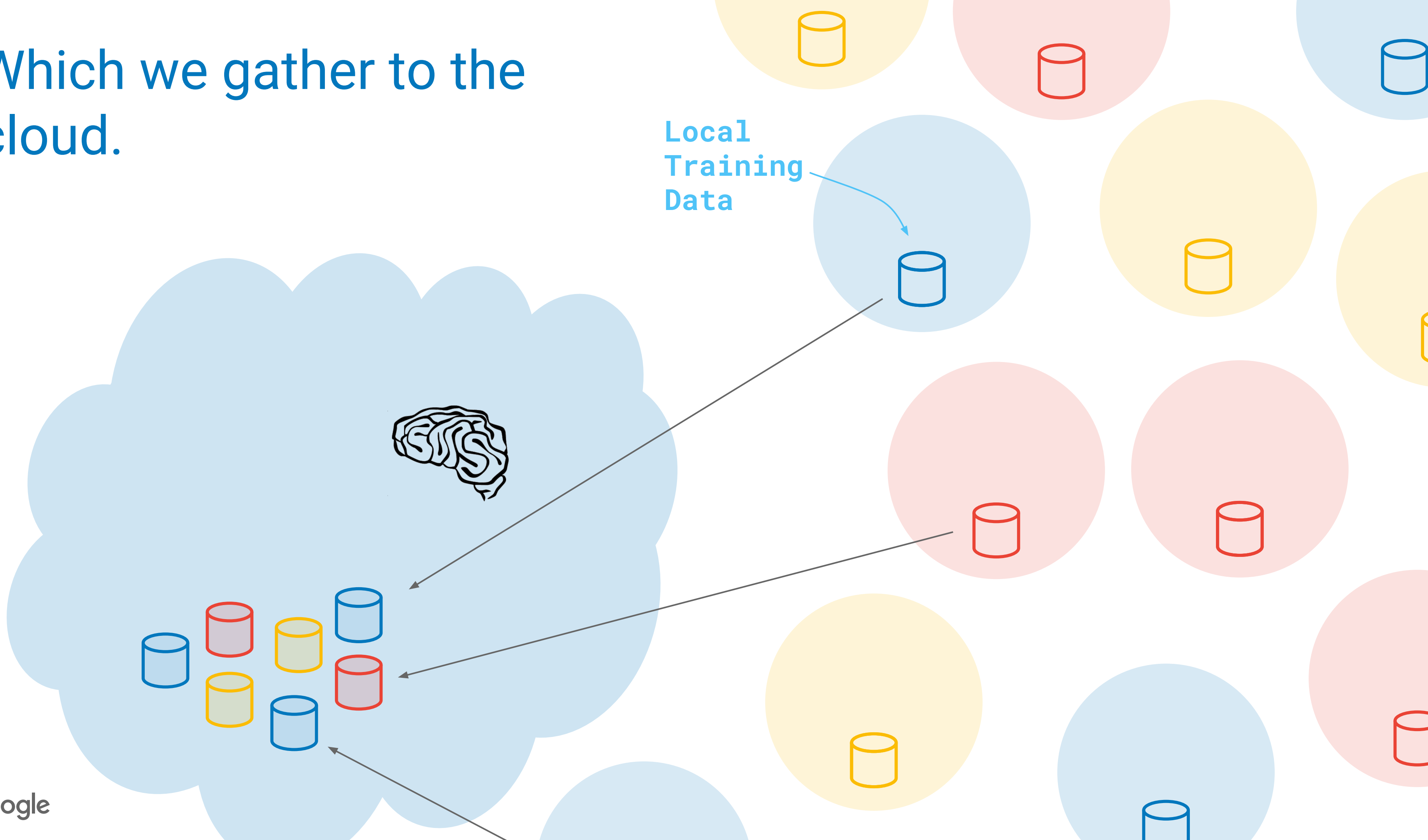


Google

Interactions generate training data on device...

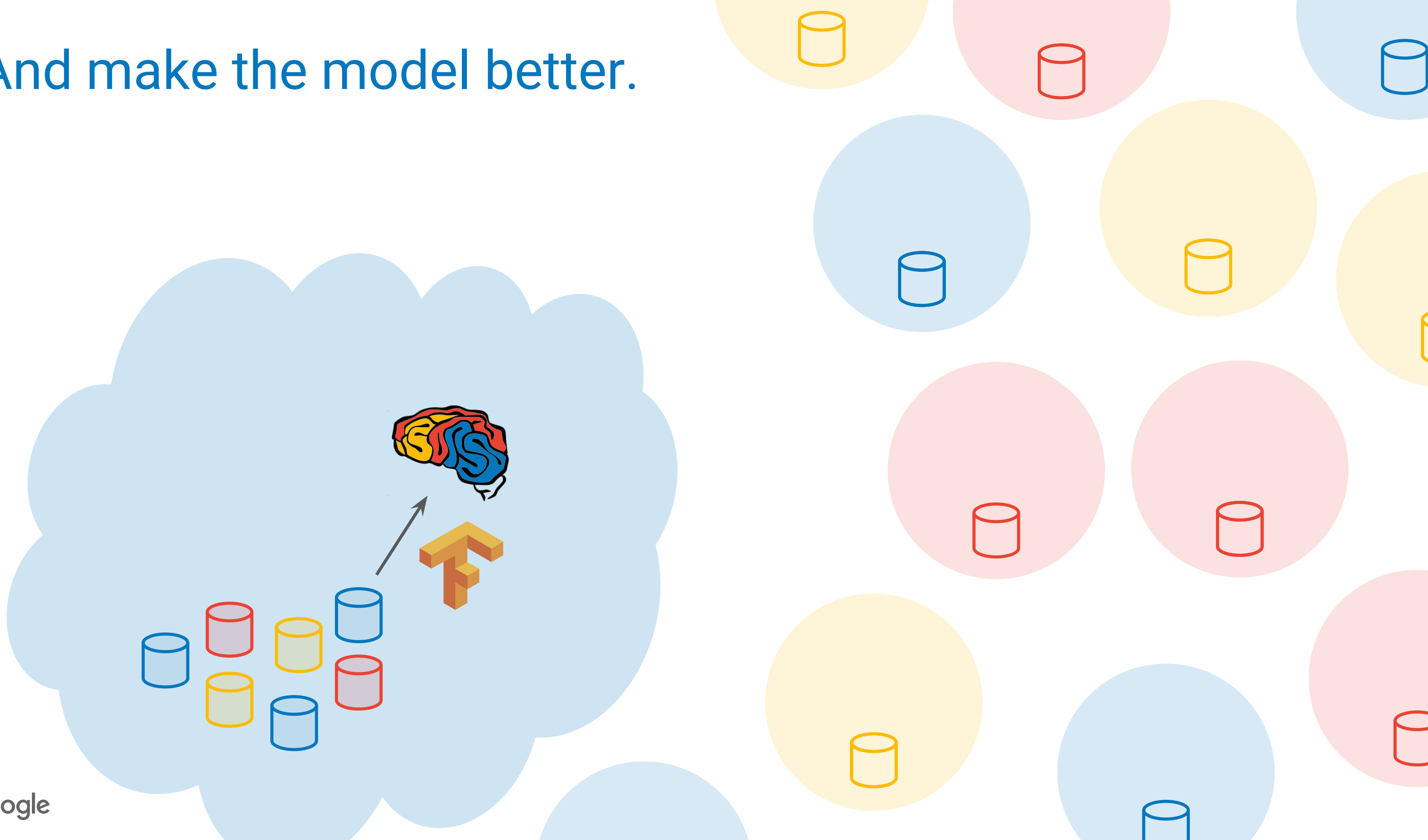


Which we gather to the cloud.

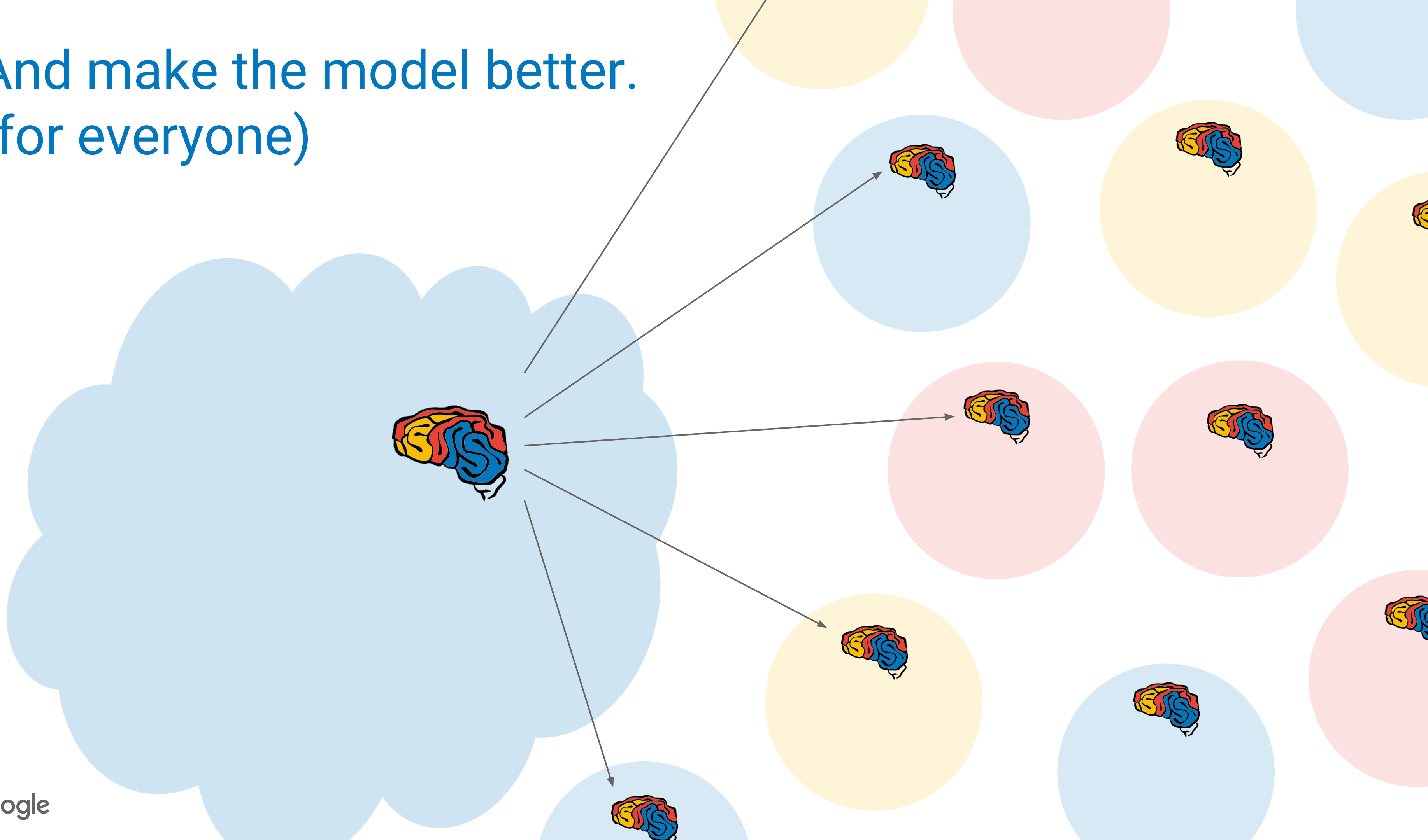


Google

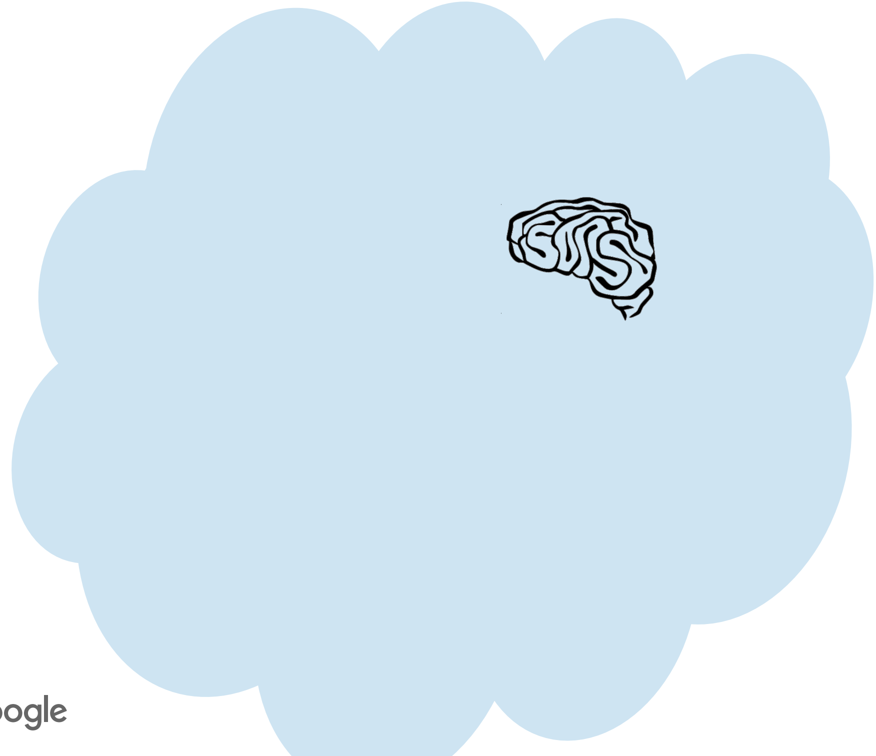
And make the model better.



And make the model better.
(for everyone)



Interactions generate training data on device...



But what about...

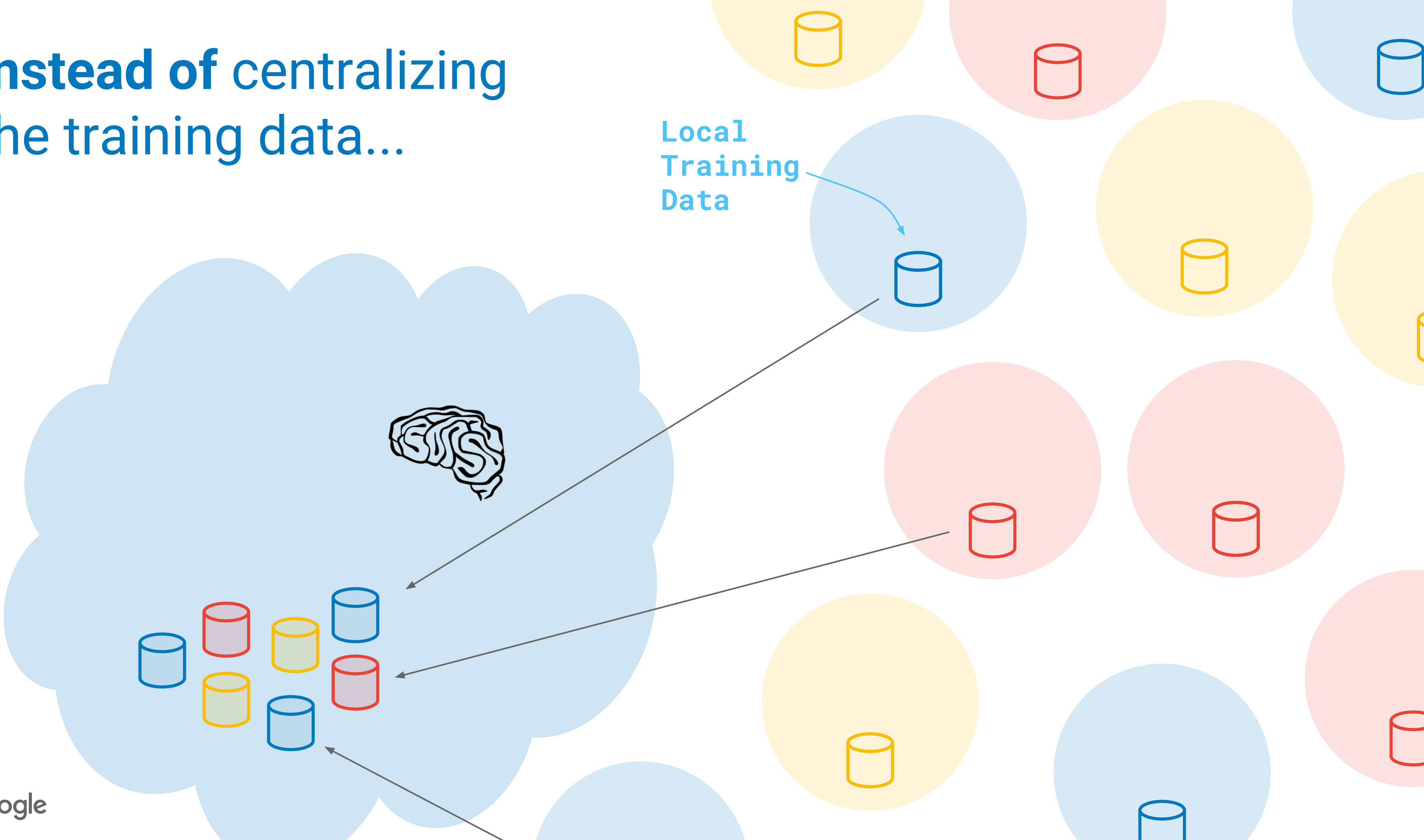
1. Sensitive device data handling?
2. Connectivity constraints?
3. Personalization?

2016+: Device-Local **Learning**

(Personalization)

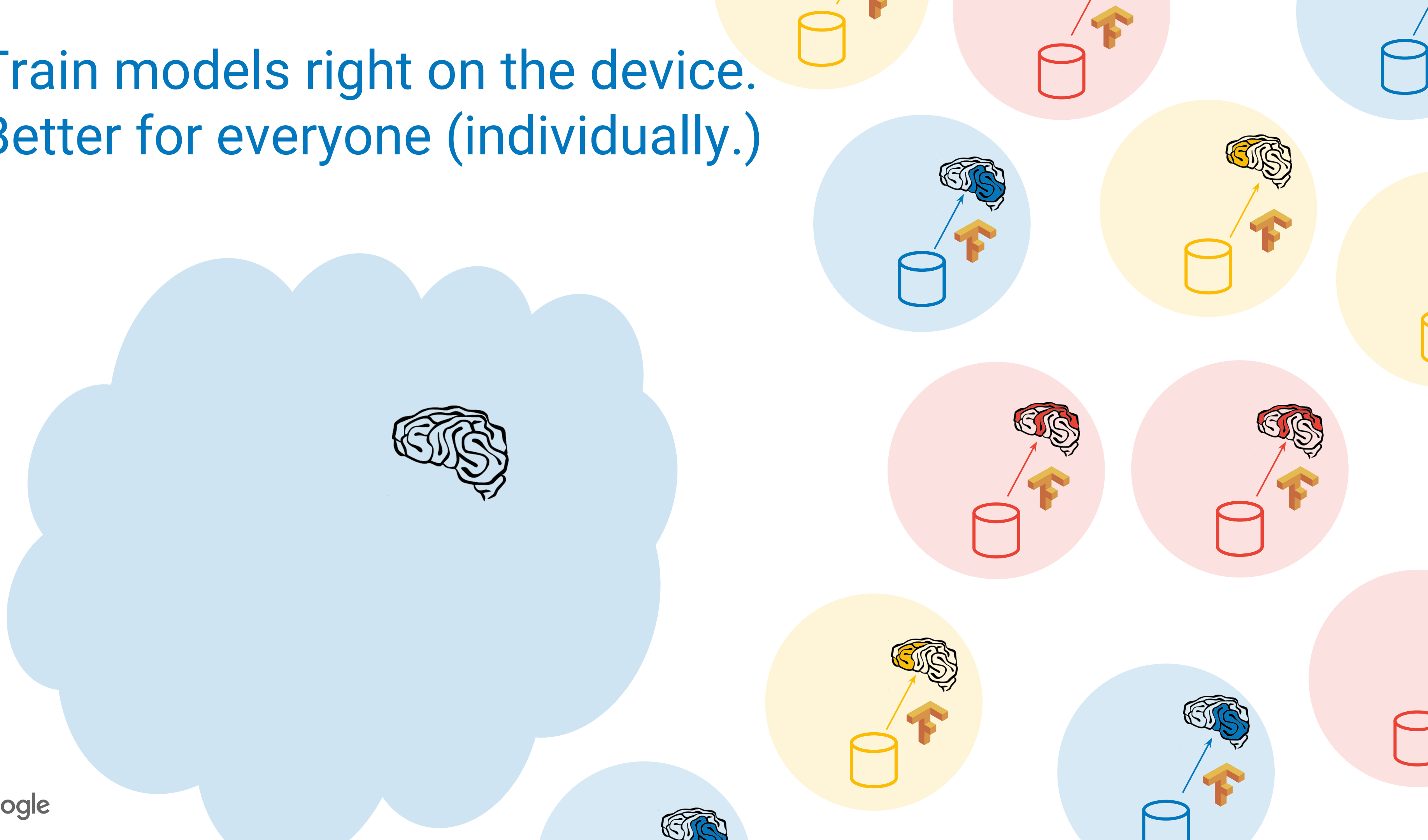
Google

Instead of centralizing the training data...

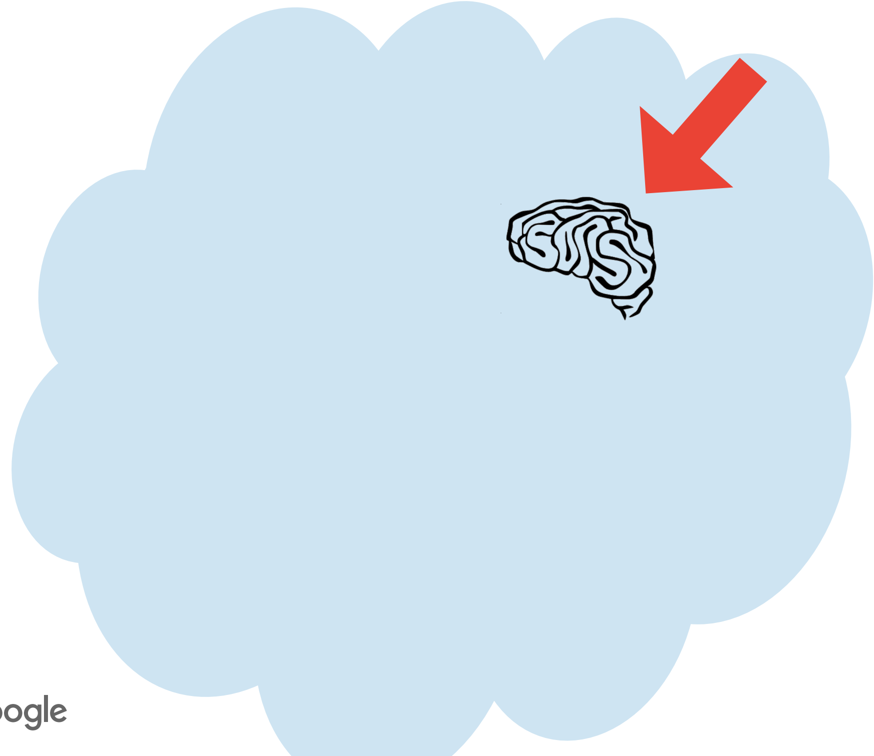


Google

Train models right on the device.
Better for everyone (individually.)



Google



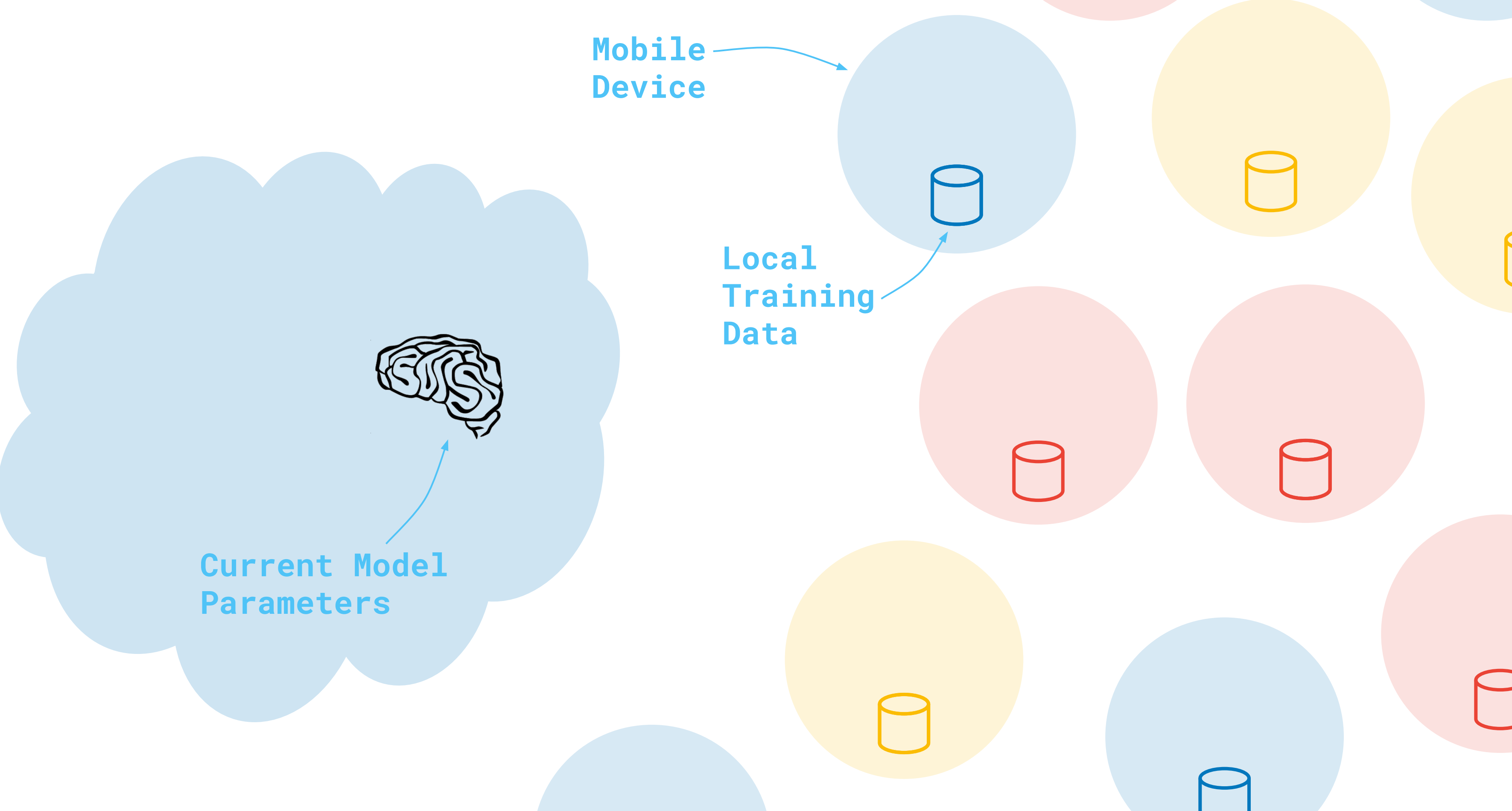
But what about…

1. New User Experience
2. Benefitting from peers' data

2017+: Cross-Device **Learning** (Federation)

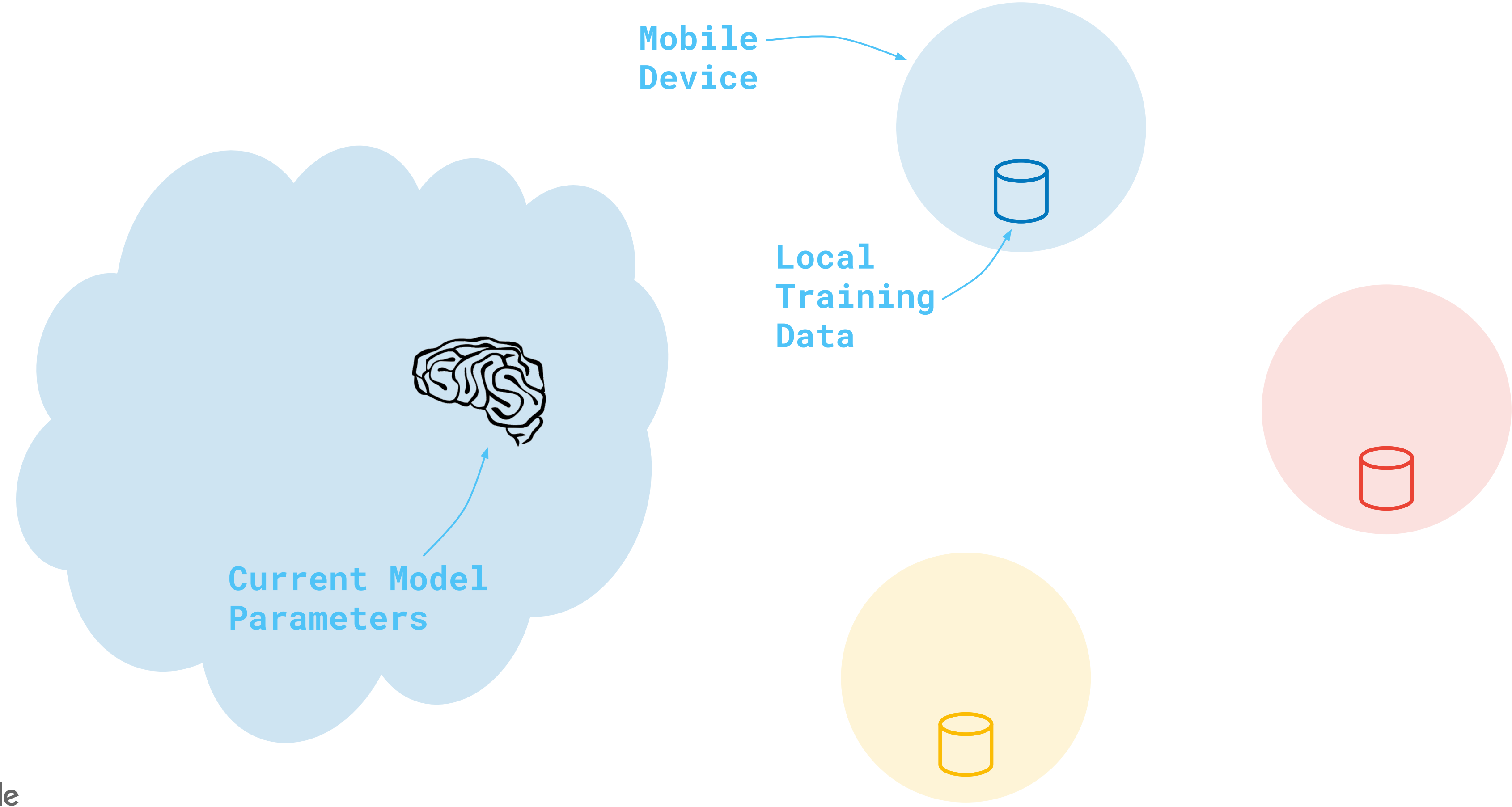
Google

Federated Learning

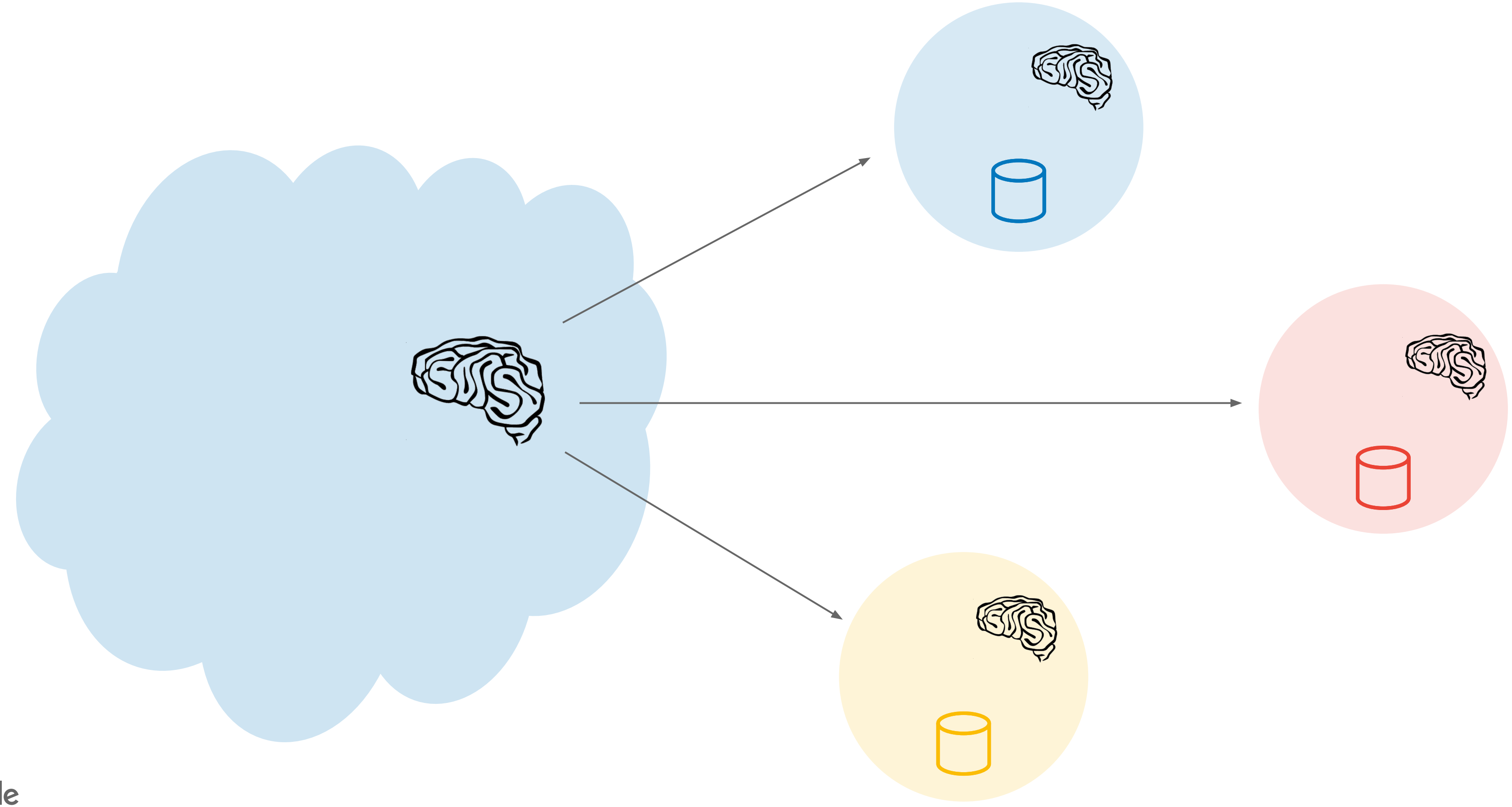


Google

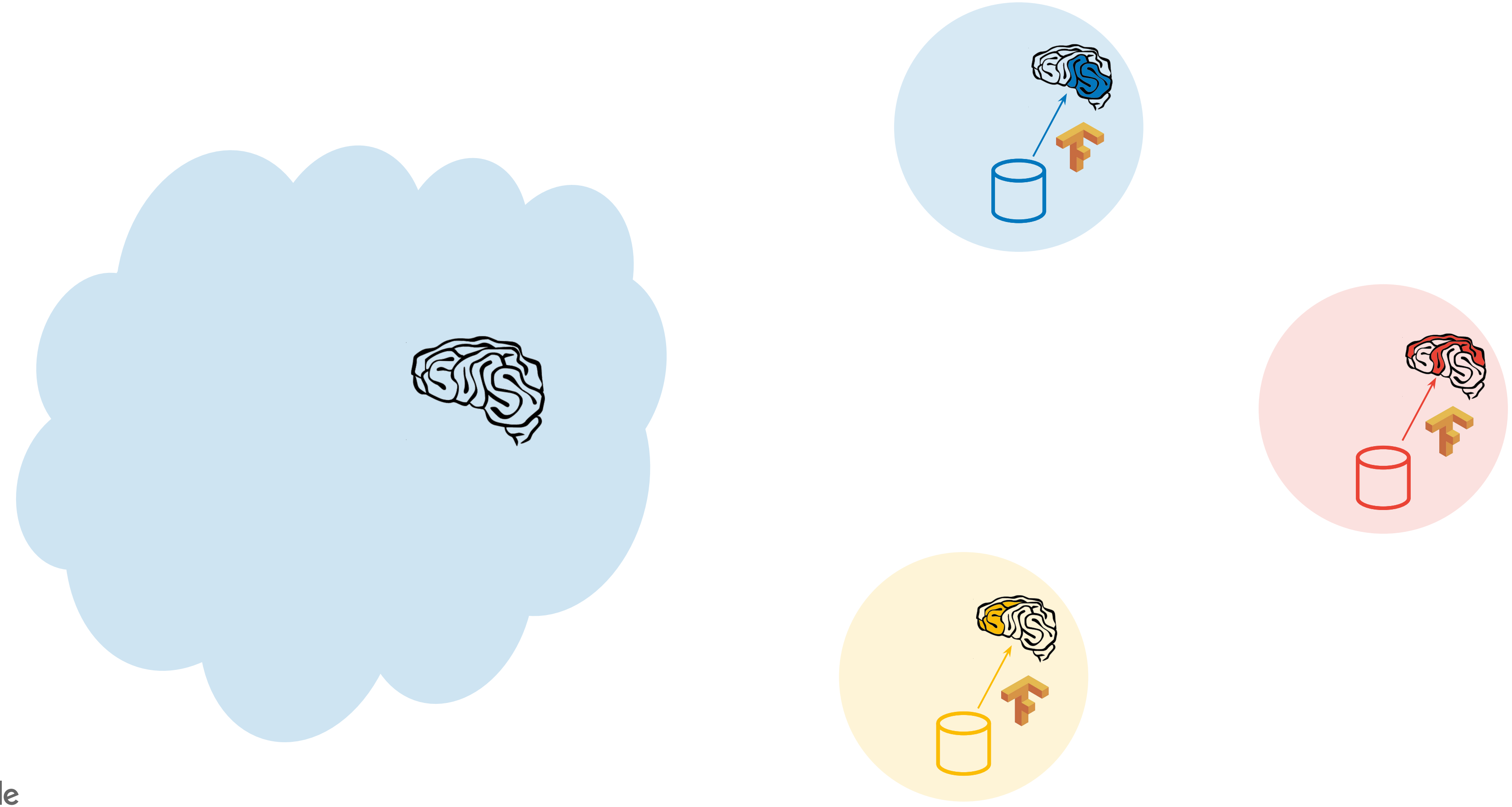
Federated Learning



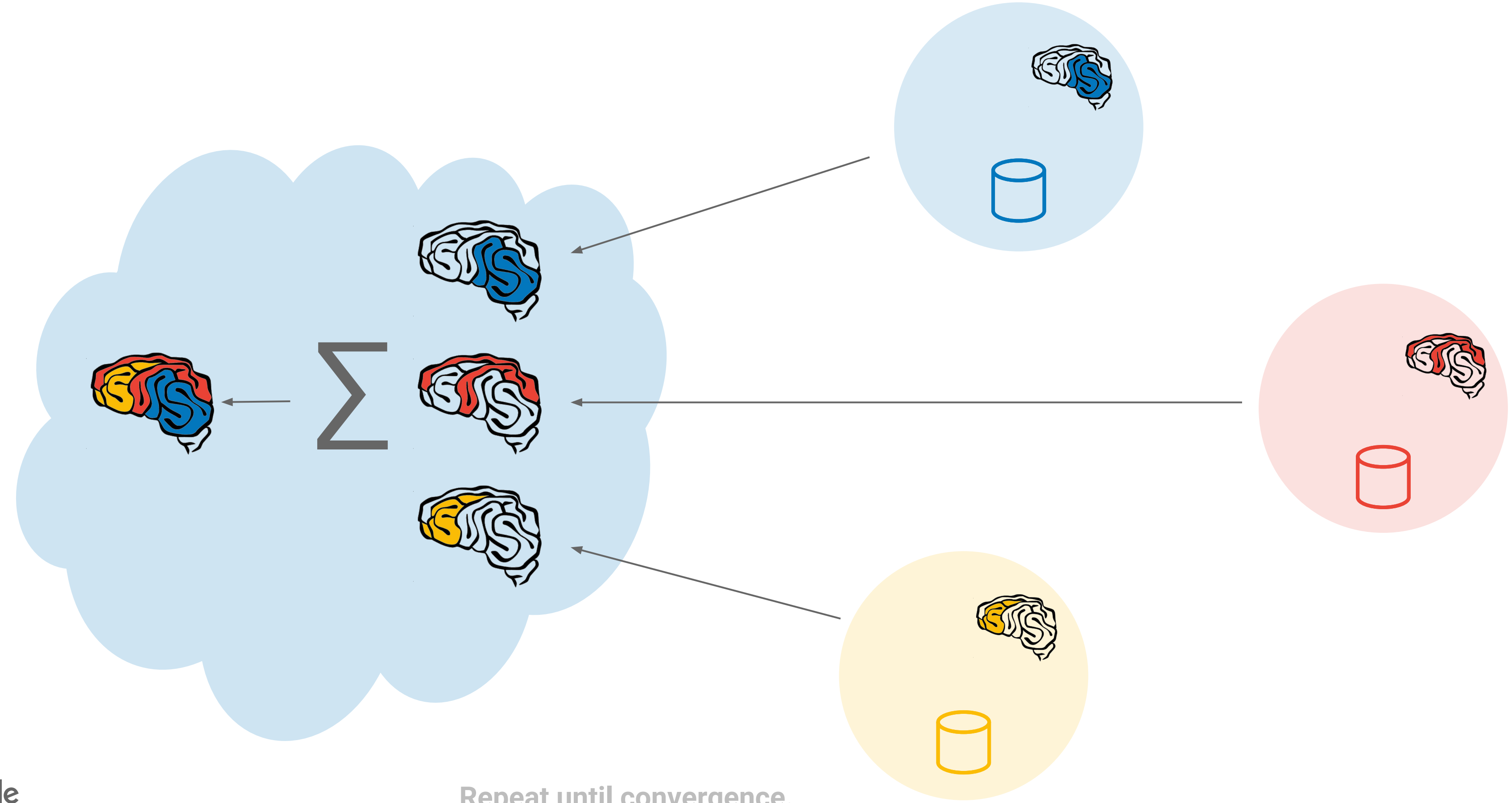
Federated Learning



Federated Learning



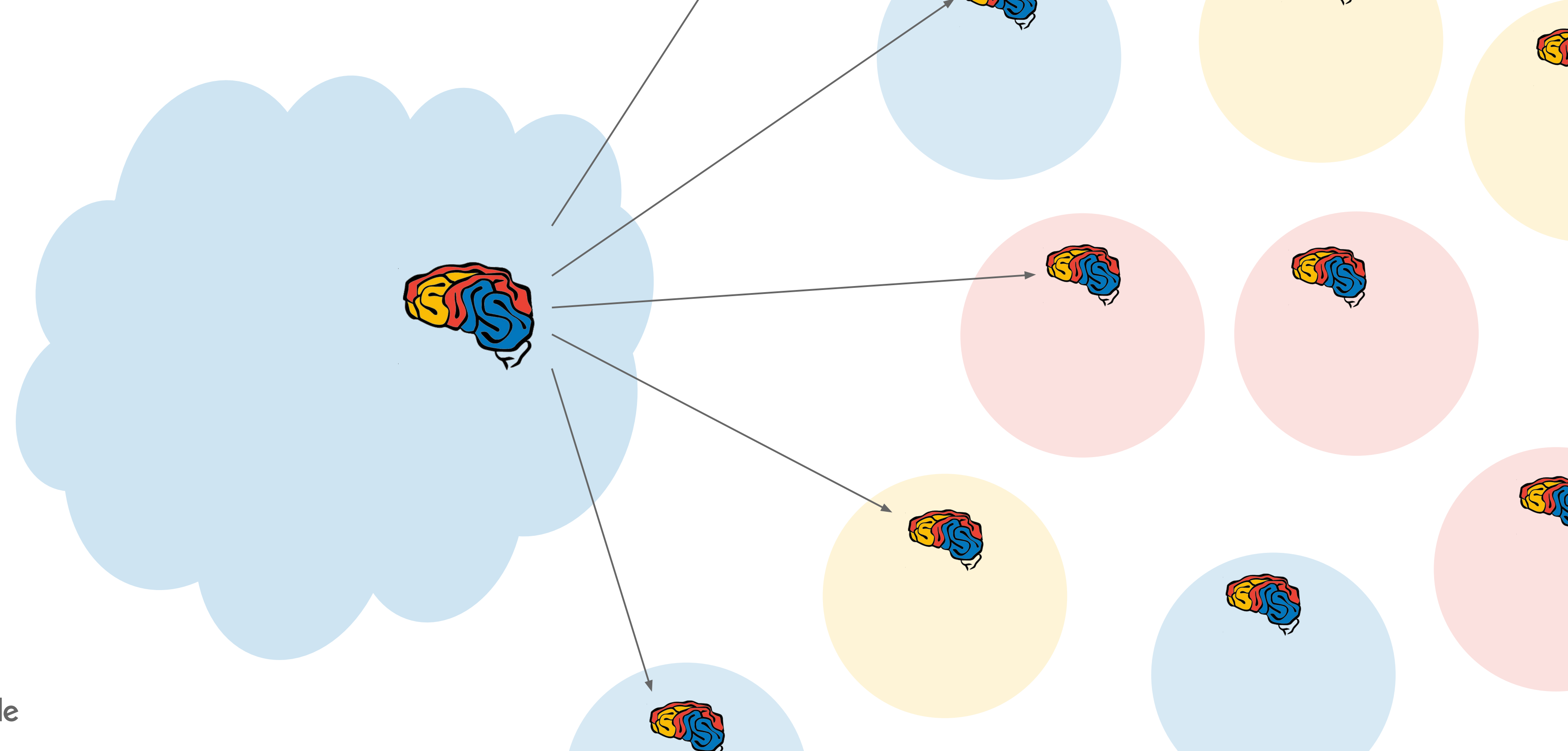
Federated Learning



Google

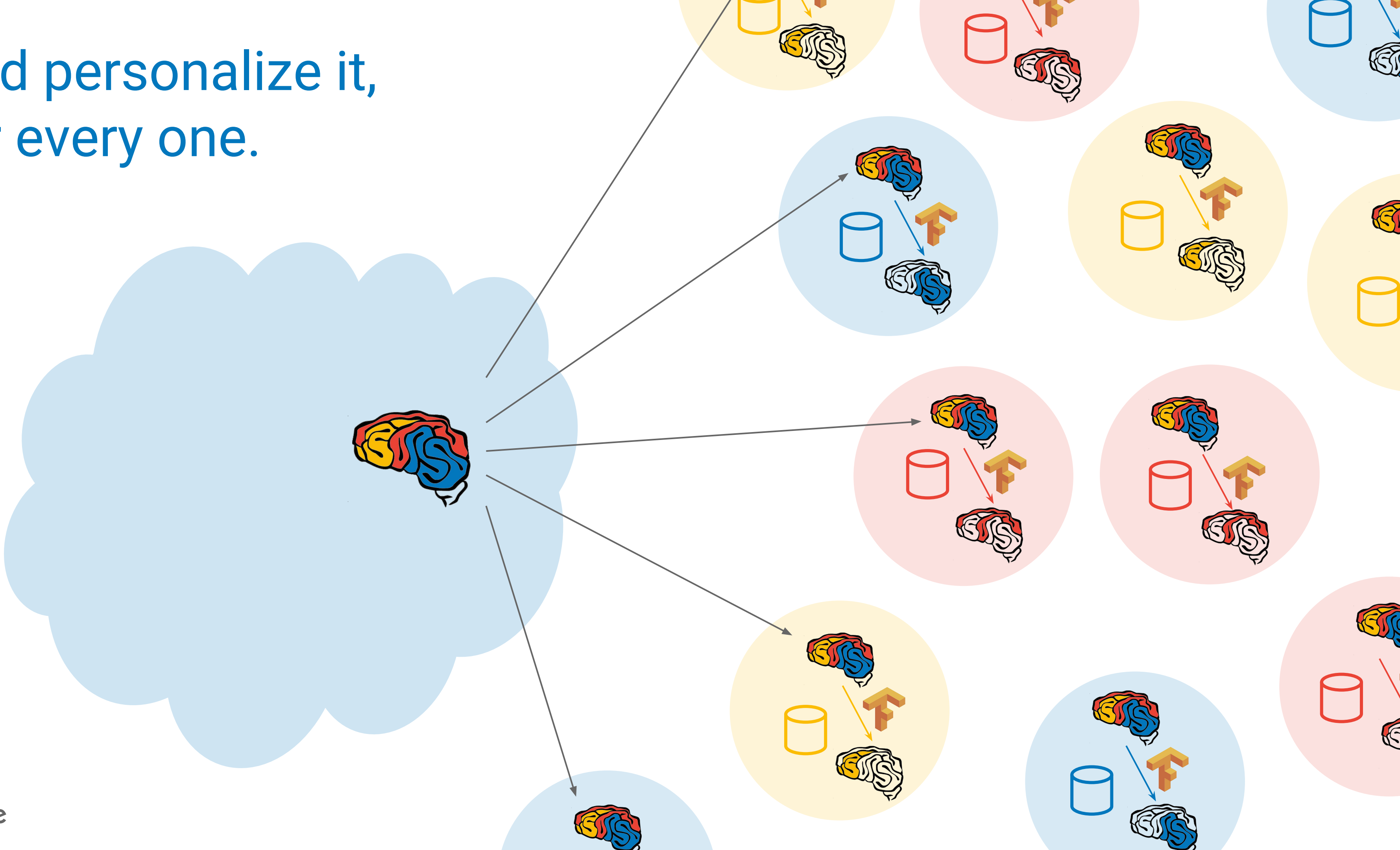
Repeat until convergence.

To make the model better.
(for everyone)



Google

And personalize it, for every one.



Latency

Data Caps

Privacy

Offline

Power

Sensors


Latency
Data Caps
Privacy
SATISFACTION
GENUINE
GUARANTEE
In Vivo
Training & Evaluation
Offline
Power
Sensors
Personalization

Applying on-device learning

What makes a good application?

- On-device data is more relevant than server-side proxy data
- On-device data is privacy sensitive or large
- Labels can be inferred naturally from user interaction
- Want large-scale personalization *and* global model improvements

Example applications

- Language modeling for mobile keyboards and voice recognition
- Image classification for predicting which photos people will share
- Smart reply taking into account all device and user context
- ...

Wrap-up

Devices are crucial to Google’s success

On-device model training enables development of great user experiences while maintaining user trust

On-device ML technology is available today!

Inference: TF Lite (go/tflite-site), Predict-on-device (go/corpsites/pod/home)

Training + inference: Brella (go/brella)

Say hello - ingerman@google.com

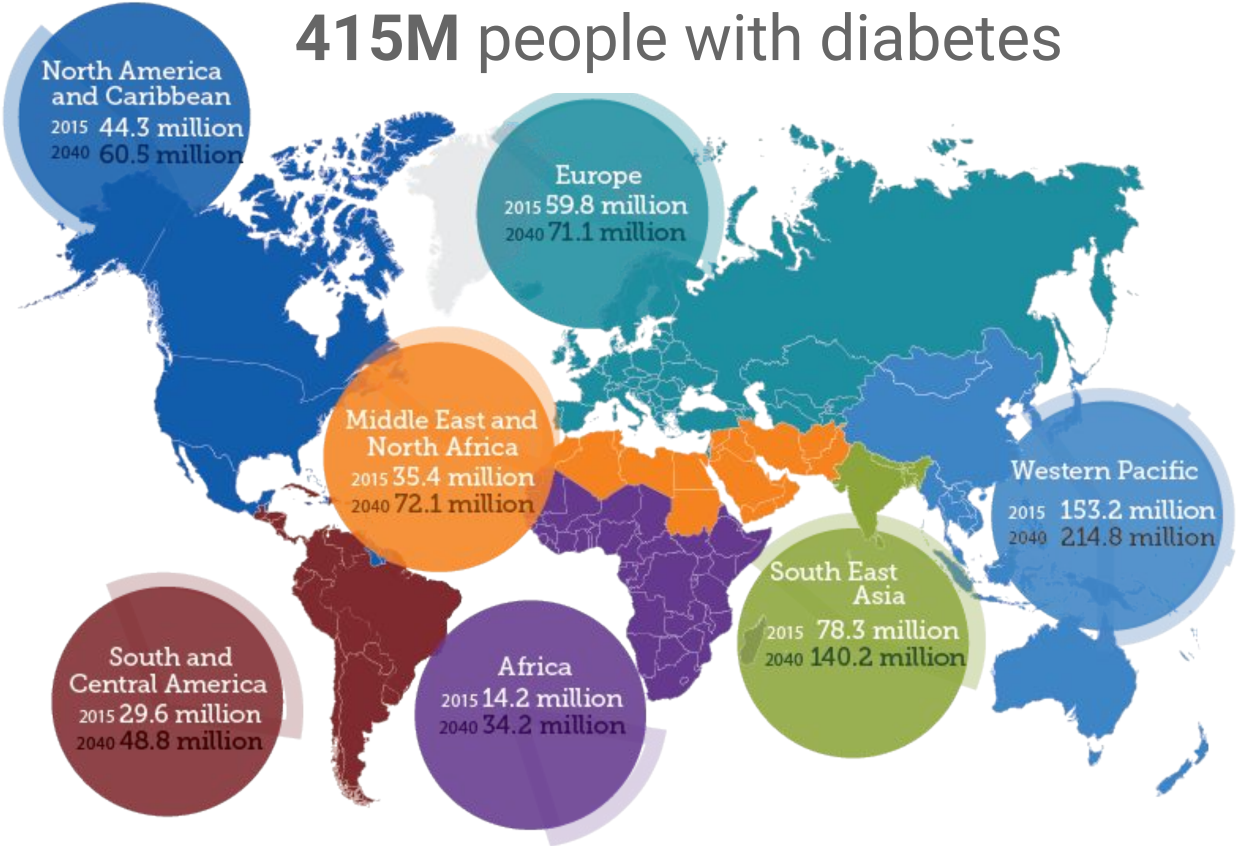
Agenda

- ➢ Welcome
- ➢ Fairness: pbrandt@
- ➢ Human Sensing: dkaram@
- ➢ ML and Data: ivanku@
- ➢ Crowd Computing: pocketaces@
- ➢ Natural Language: barakt@
- ➢ On-device: ingerman@
- ➢ **Medical Applications**: lhpeng@
- ➢ Getting to Launch: binghamj@
- ➢ Refreshing Conversations

Medical Applications

Lily Peng (lhpeng@)

Diabetic retinopathy: fastest growing cause of blindness



Regular screening is key to preventing blindness

=

How DR is Diagnosed: Retinal Fundus Images

No DR	Mild DR	Moderate DR	Severe DR	Proliferative DR


ENQUIRY
INDIA
Shortage of 127,000 eye doctors
45% of patients suffer vision loss before diagnosis

Adapt deep neural network to read fundus images

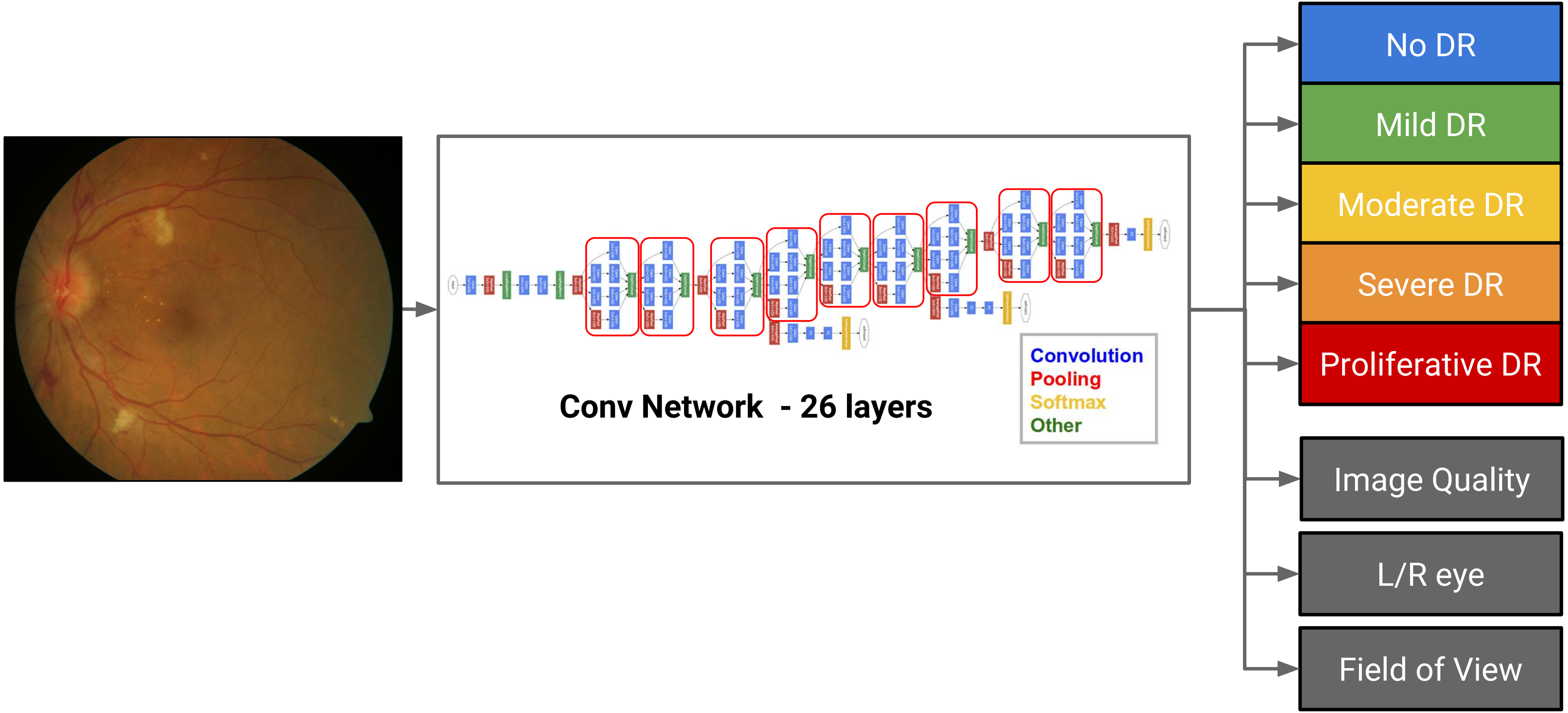
Labeling tool
54 ophthalmologists
130k images

→

880k diagnoses

- No DR
- Mild DR
- Moderate DR
- Severe DR
- Proliferative DR

Adapt deep neural network to read fundus images



ARDA: Automated Retinal Disease Assessment

Drag another image to analyze, or **CHOOSE IMAGE**

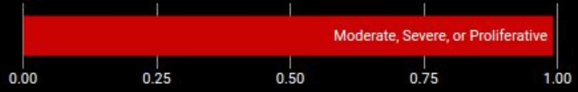
FILENAME (SIZE)

uploaded-retina-image.jpg (2.11M)

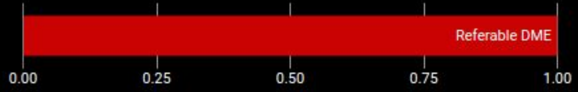
DIAGNOSIS ID

drw-2062

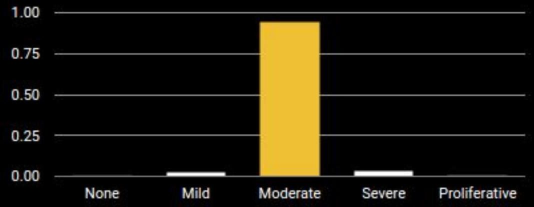
MODERATE+ DIABETIC RETINOPATHY REFERABLE

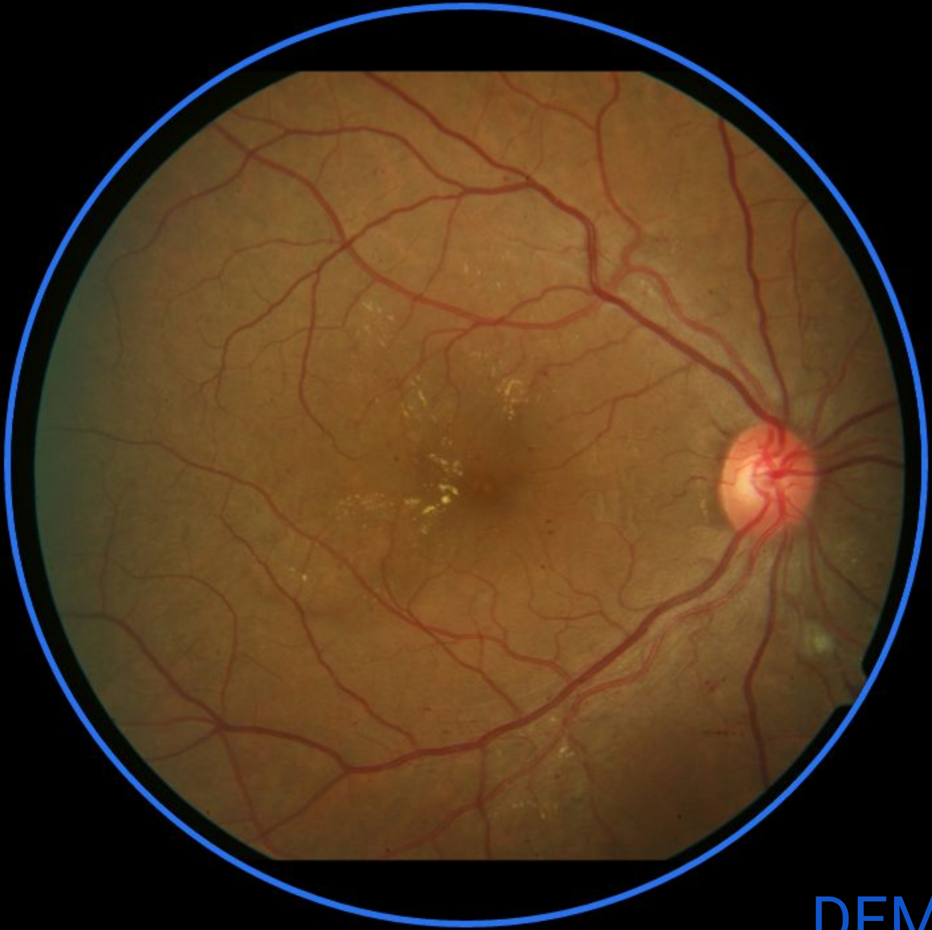


DIABETIC MACULAR EDEMA GRADE



DIABETIC RETINOPATHY GRADE

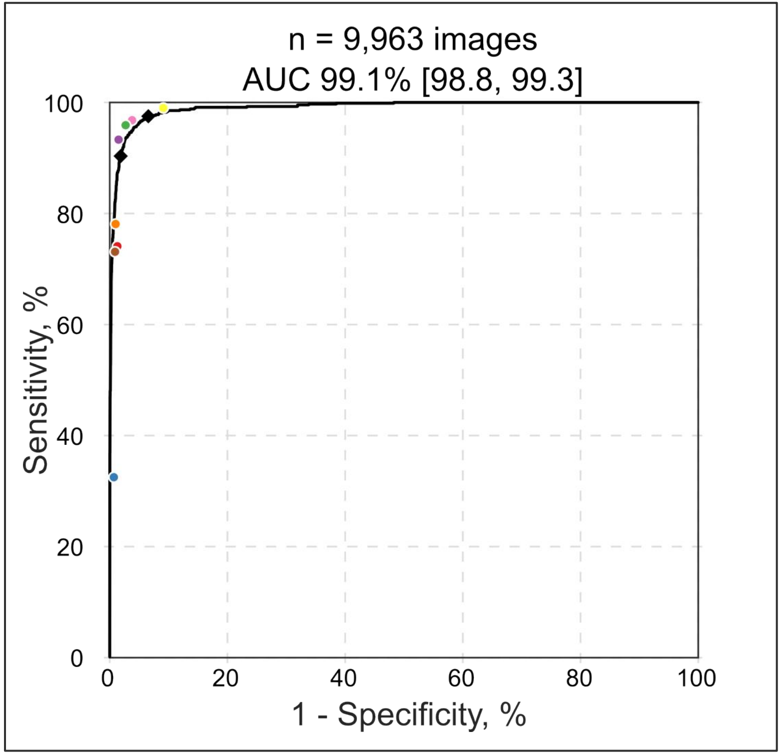




DEMO

JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs



F-score

0.95
Algorithm

0.91
Ophthalmologist (median)

"The study by Gulshan and colleagues **truly represents the brave new world in medicine.**"

Dr. Andrew Beam, Dr. Isaac Kohane
Harvard Medical School

"Google just published this paper in JAMA (impact factor 37) [...] **It actually lives up to the hype.**"

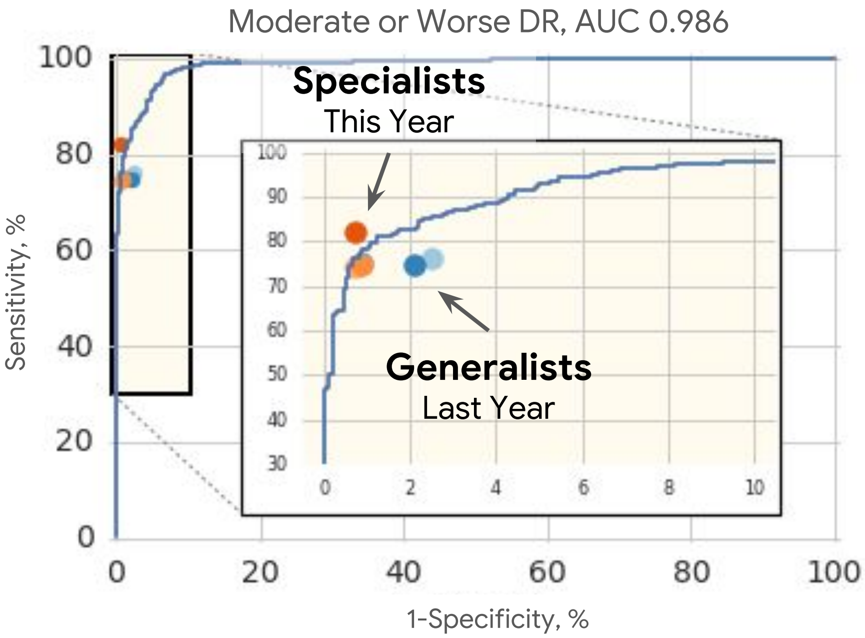
Dr. Luke Oakden-Rayner
University of Adelaide

What's next? Much more to do on path to clinical adoption

CLINICAL TRIALS & REGULATORY

PARTNERSHIPS & HARDWARE

Last Year - On Par with General Ophthalmologists
This Year - On Par with Retinal Specialist Ophthalmologists



	Weighted Kappa
Ophthalmologists Individual	0.80-0.84
Algorithm	0.84
Retinal Specialists Individual	0.82-0.91

Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. J. Krause, et al.

Bringing this Technology to the World

Aravind & Sankara (India)

- Publishing previous trial results Q4
- Started assisted read trial with Aravind
- 2018 Goal: Aravind system-wide roll out (total screenings/yr: 250k)

EyePACS (U.S.)

- Quality improvement deployment launching Q4, HCI ongoing
- 2018 Goal: system-wide roll out (total screenings/yr: 120k), diagnostic roll-out with non-profit arm upon regulatory approval

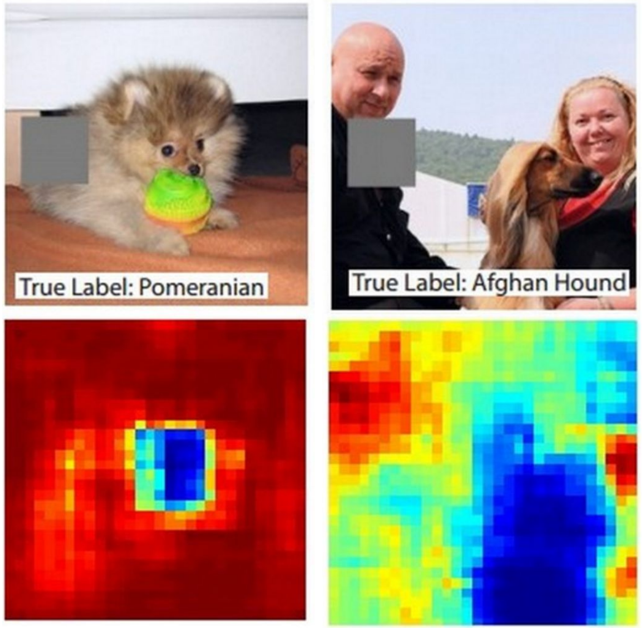
Ministry of Health (Thailand)

- Nationwide, 7500-patient retrospective study started Q4 (validation)
- **2018 Goal: national roll out (total screenings/year: 4M)**

...and more sites in planning stages

Explainability: Neural Networks a Black Box? Not Really

"Show me where."

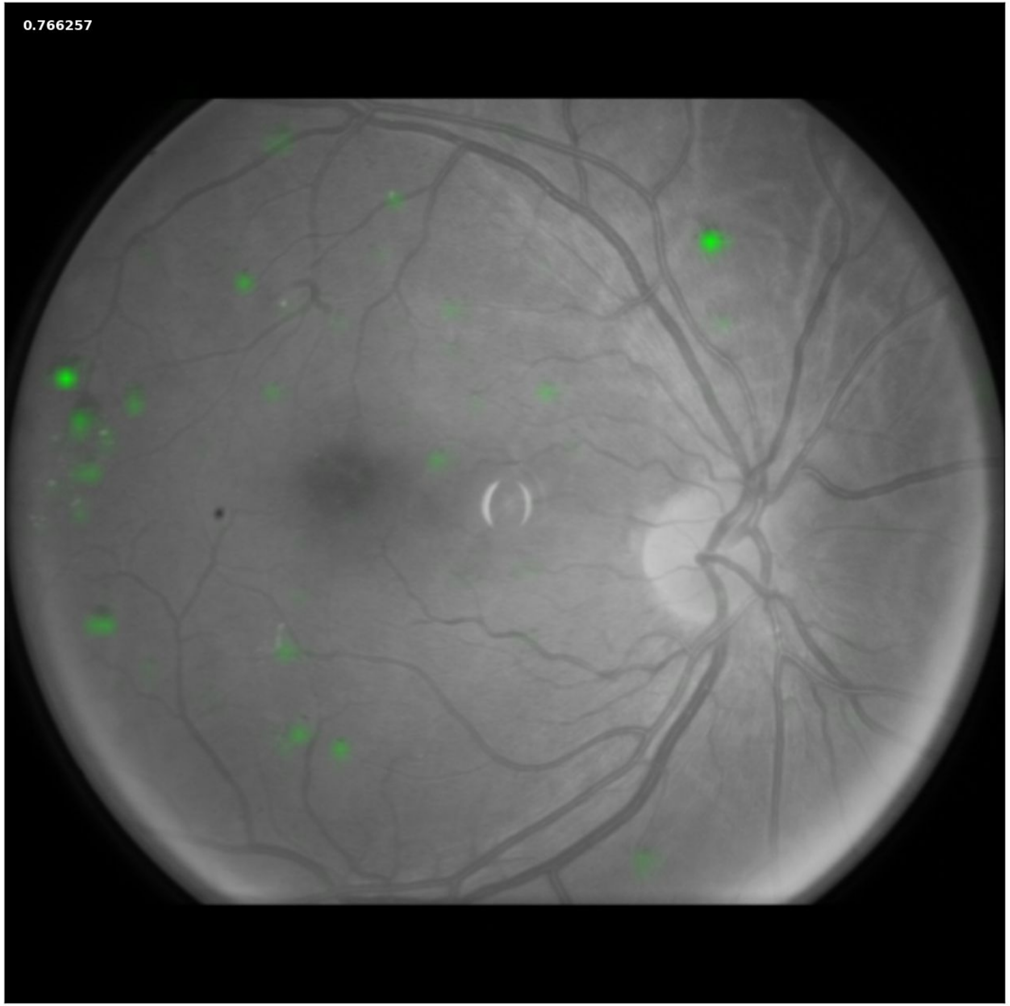


Google

“Show me where”

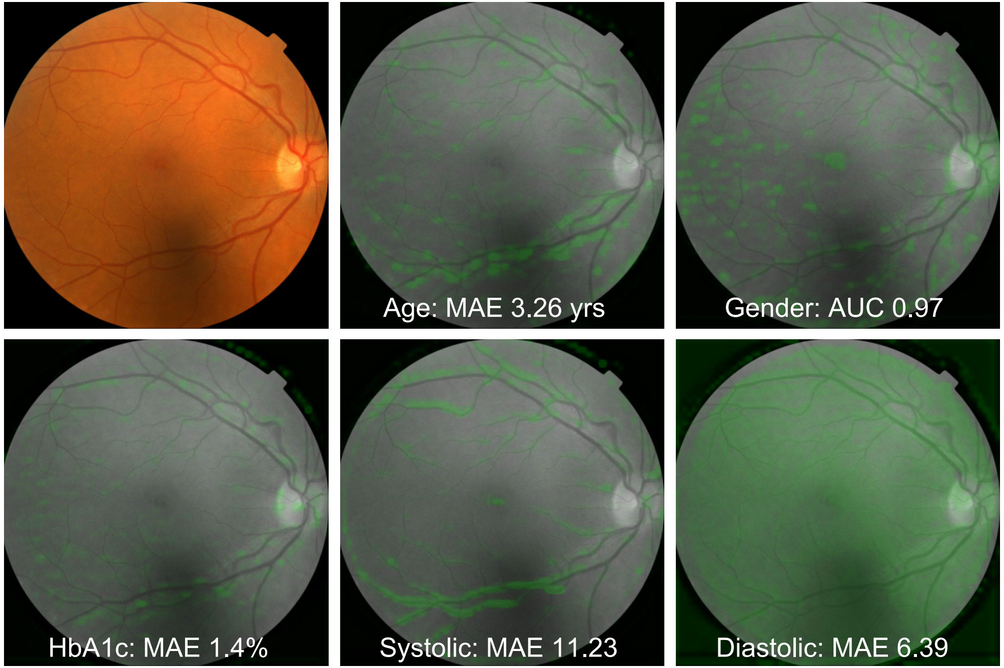
Mild DR

Moderate DR



Moderate DR

Completely new, novel scientific discoveries



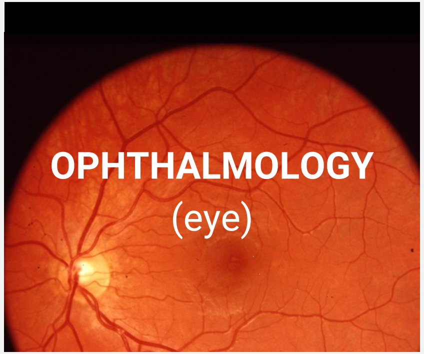
Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning. R. Poplin, A. Varadarajan et. al

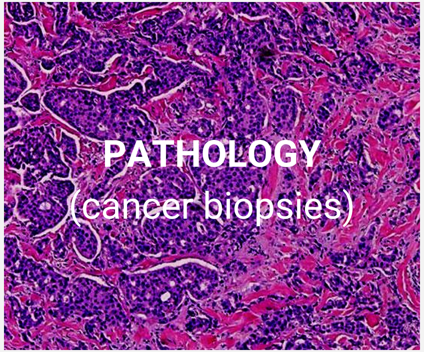
Predicting things that doctors can't predict from imaging

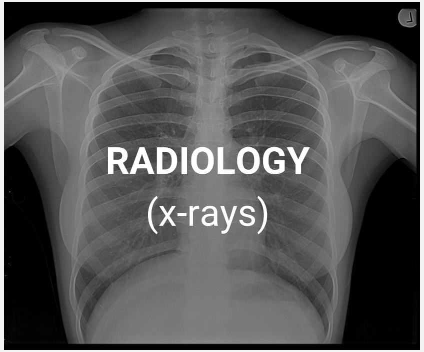
Potential as a new biomarker

Can we predict cardiovascular risk? If so, this is a very nice non-invasive way of doing so

Many more opportunities to increase both access and accuracy







Deep Learning has shown promise in building assistive tools for doctors. What's next?

Clinical Validation

Building Trust

Workflow & User Design

Agenda

- ➢ Welcome
- ➢ Fairness: pbrandt@
- ➢ Human Sensing: dkaram@
- ➢ ML and Data: ivanku@
- ➢ Crowd Computing: pocketaces@
- ➢ Natural Language: barakt@
- ➢ On-device: ingerman@
- ➢ Medical Applications: lhpeng@
- ➢ **Getting to Launch**: binghamj@
- ➢ Refreshing Conversations

Getting to Launch

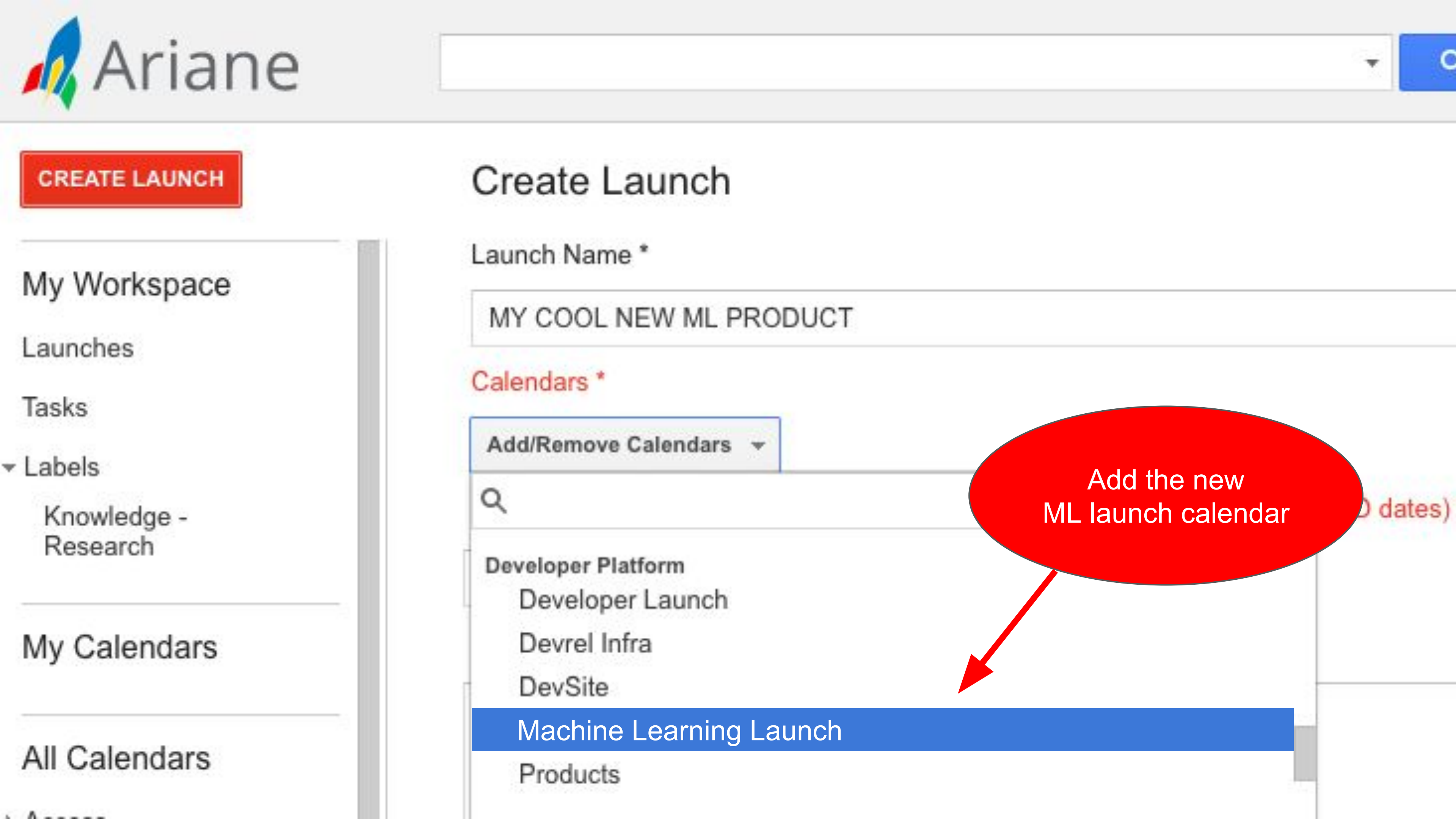
Jonathan Bingham (binghamj@)


As product managers
we all want to launch
great *ML-powered* products


And not crash and burn

Up and to the right

So what does it take to successfully launch an ML product?


Ariane
CREATE LAUNCH
My Workspace
Launches
Tasks
Labels
Knowledge - Research
My Calendars
All Calendars
Create Launch
Launch Name *
MY COOL NEW ML PRODUCT
Calendars *
Add/Remove Calendars
Add the new
ML launch calendar
dates)
Developer Platform
Developer Launch
Devrel Infra
DevSite
Machine Learning Launch
Products

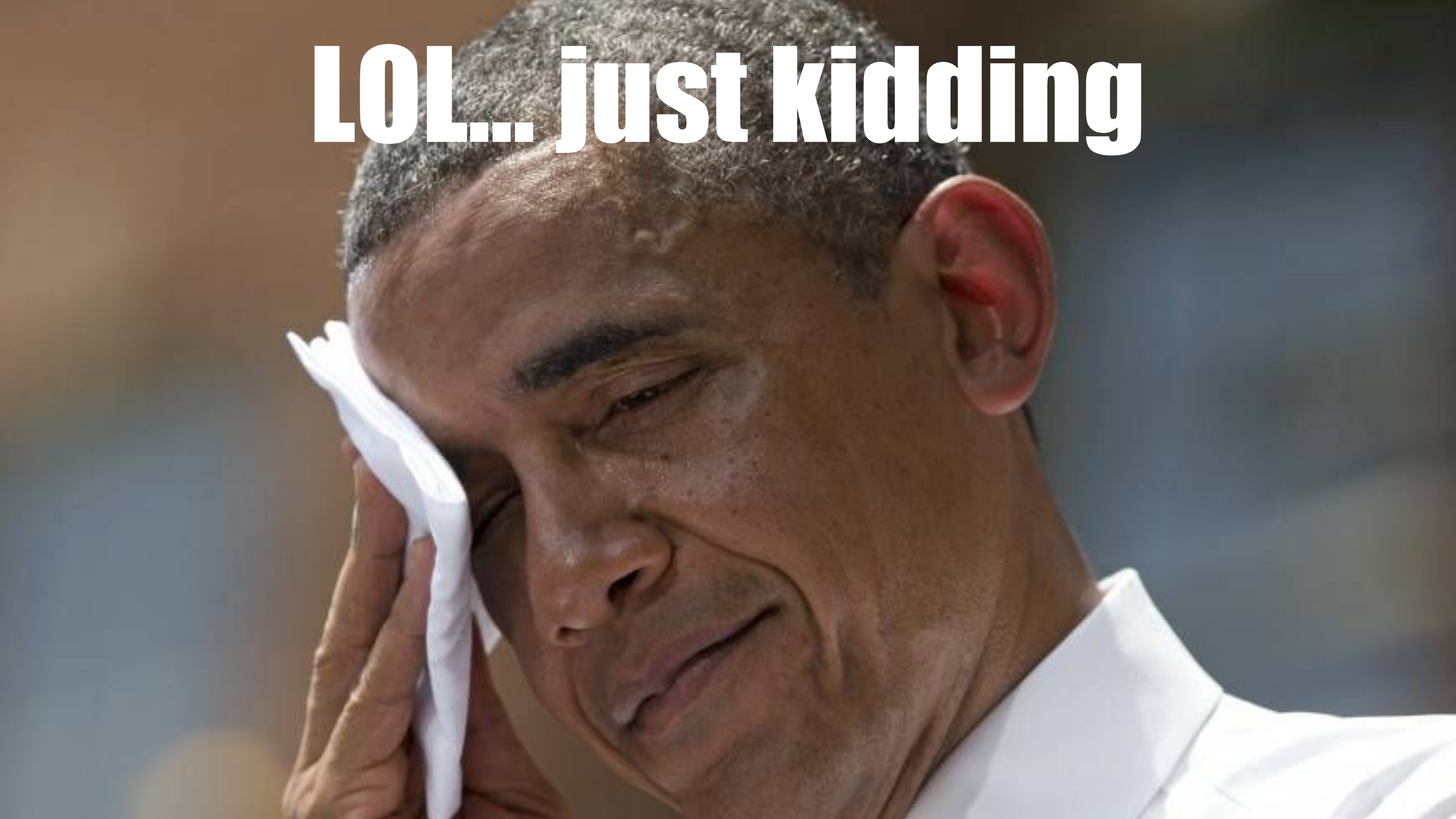
Get approval from all of the ML reviewers

From the Machine Intelligence team

- ✓ Problem definition
- ✓ Data use
- ✓ Model understandability
- ✓ TPU quota
- ✓ Production readiness

From Legal and Privacy

- ✓ Fairness
- ✓ Embarrassment / PR
- ✓ De-identification


LOL... just kidding

There are no
new approvals required

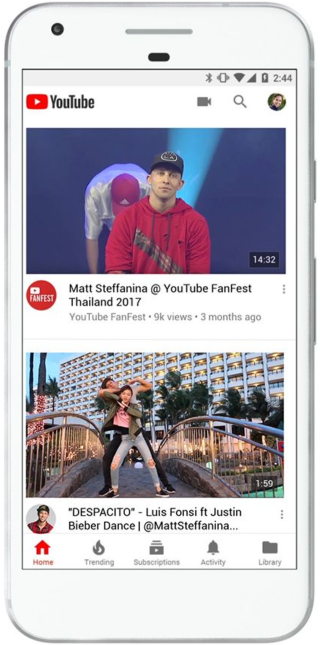
So what does it *really* take to successfully launch an ML product?

Focus on the user

User >> product >> technology

As we search for new ways to delight users
ML is increasingly part of the answer

Define an ambitious goal



Jigsaw

Deciding on a machine learning approach

Prerequisites:

Data. The more the better. You'll use some of its features.

A goal. Know what you're trying to predict, called a label.

Bring in the technical experts:

Work with your eng team to choose an ML model.

For help, you can set up a consultation at go/ml-consult.

If there's a simpler or more appropriate solution than ML, use it!

Understand your ML model

Why? To build a better product:

Avoid confounders. Cf. Google Flu Trends.

Understand bad results on a subset of data.

Balance result quality vs. cost.

Why else? Your users may want to know:

Why is my video blocked?

Why is my ad shown to the wrong audience?

Why are you showing me this news article?

Start simple. Iteratively improve.

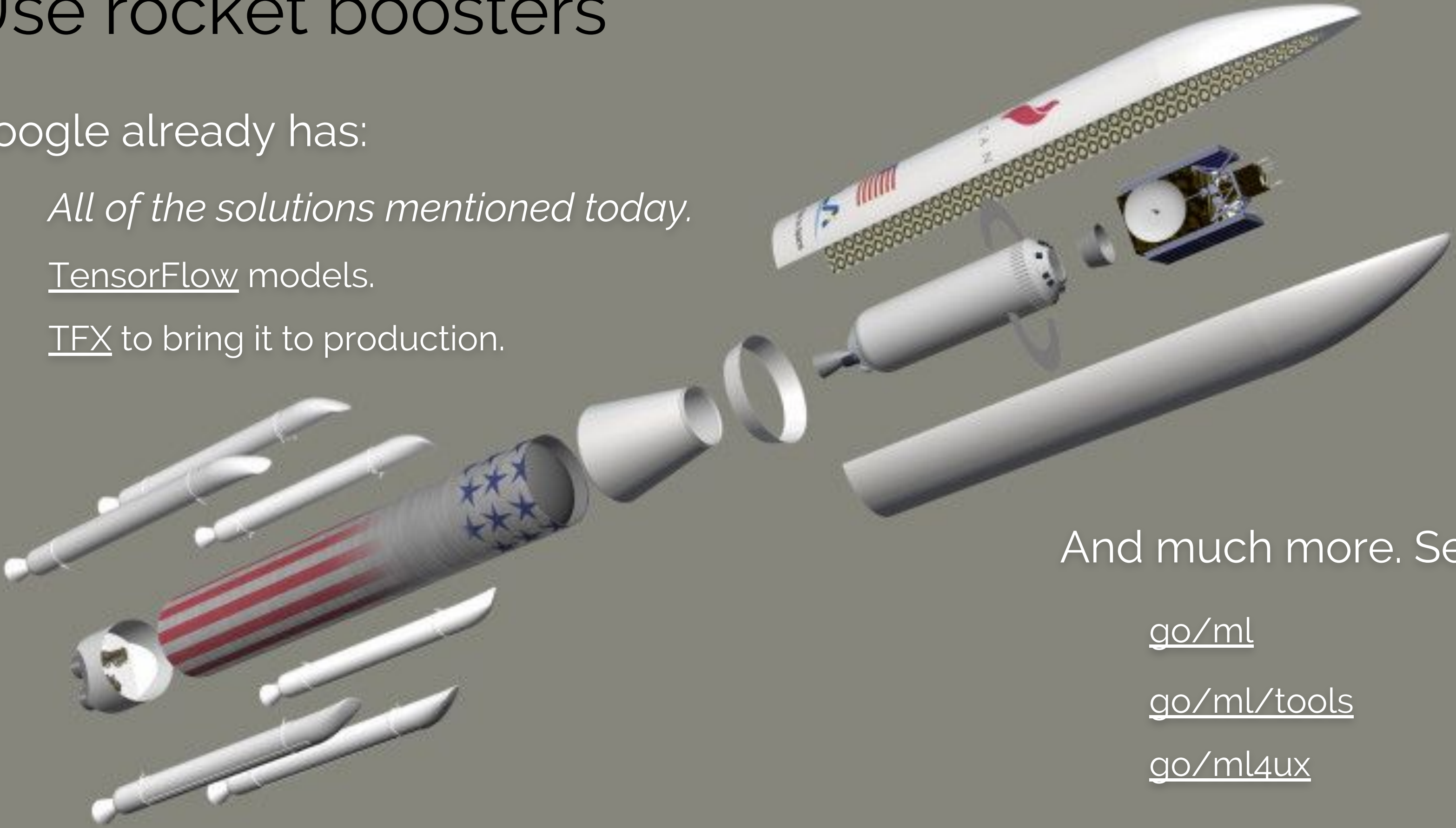
Use rocket boosters

Google already has:

All of the solutions mentioned today.

TensorFlow models.

TFX to bring it to production.



And much more. See:

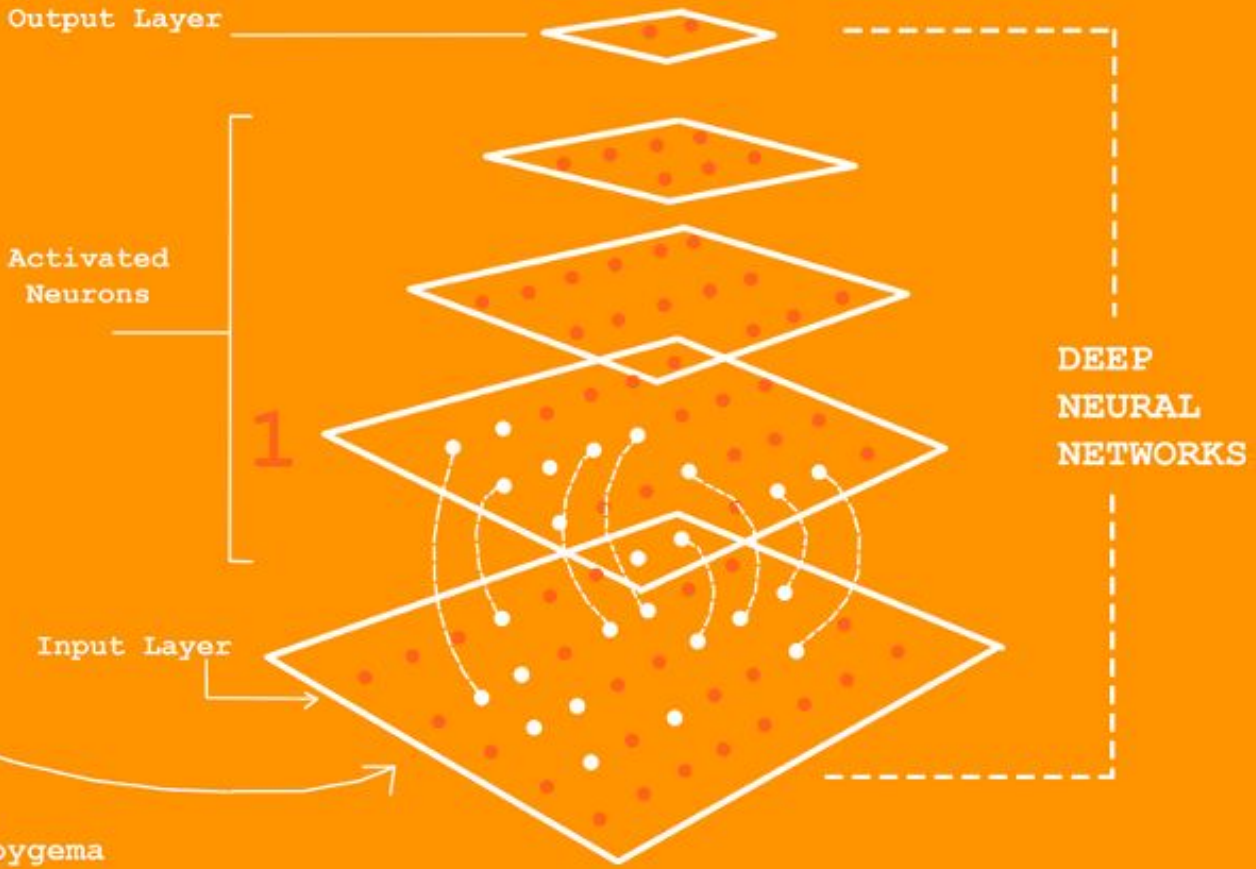
go/ml

go/ml/tools

go/ml4ux

TensorFlow™

CAT?



MADRID MEETUP @Soygema

TensorFlow is...

The most popular GitHub project for ML in the world.

More than just neural networks:

- Linear/logistic regression
- Random Forests
- Support Vector Machines
- Bayesian Optimization
- K-means clustering
- Gaussian Mixture Models

Confidential Material: This page is confidential. Do not share or discuss until authorized to do so.

Lab Part 1: Warm-Up
Introduction
Before You Start the Lab
Resources for Class and Beyond
Coding TensorFlow at Google
TensorFlow Coding Warm-Up
TensorFlow Execution Model (optional)

Lab Part 2: Input
Predicting Video Watches
Reading in Data

Lab Part 3: Defining the Model
Creating a Custom Model

Lab Part 4: Loss, Evaluation, and the Final Model
Computing Loss and Evaluation Metrics
Completing the Custom Estimator

Lab Part 5: Serving

MLCC Lab (Estimator API Version): Coding TensorFlow at Google

Internal: Count: 3, Average: 3.7

go/mlcc-estimator-lab-coding-tf-in-google

Contents

Lab Setup
TensorFlow Versions
A Sample BUILD File for TensorFlow
Using TensorFlow in Python
tf.learn and Estimator Overview

Lab Setup

The code we'll use in this lab can be found in the //depot/google3/engedu/ml/ml101/estimator_lab directory and its subdirectories.

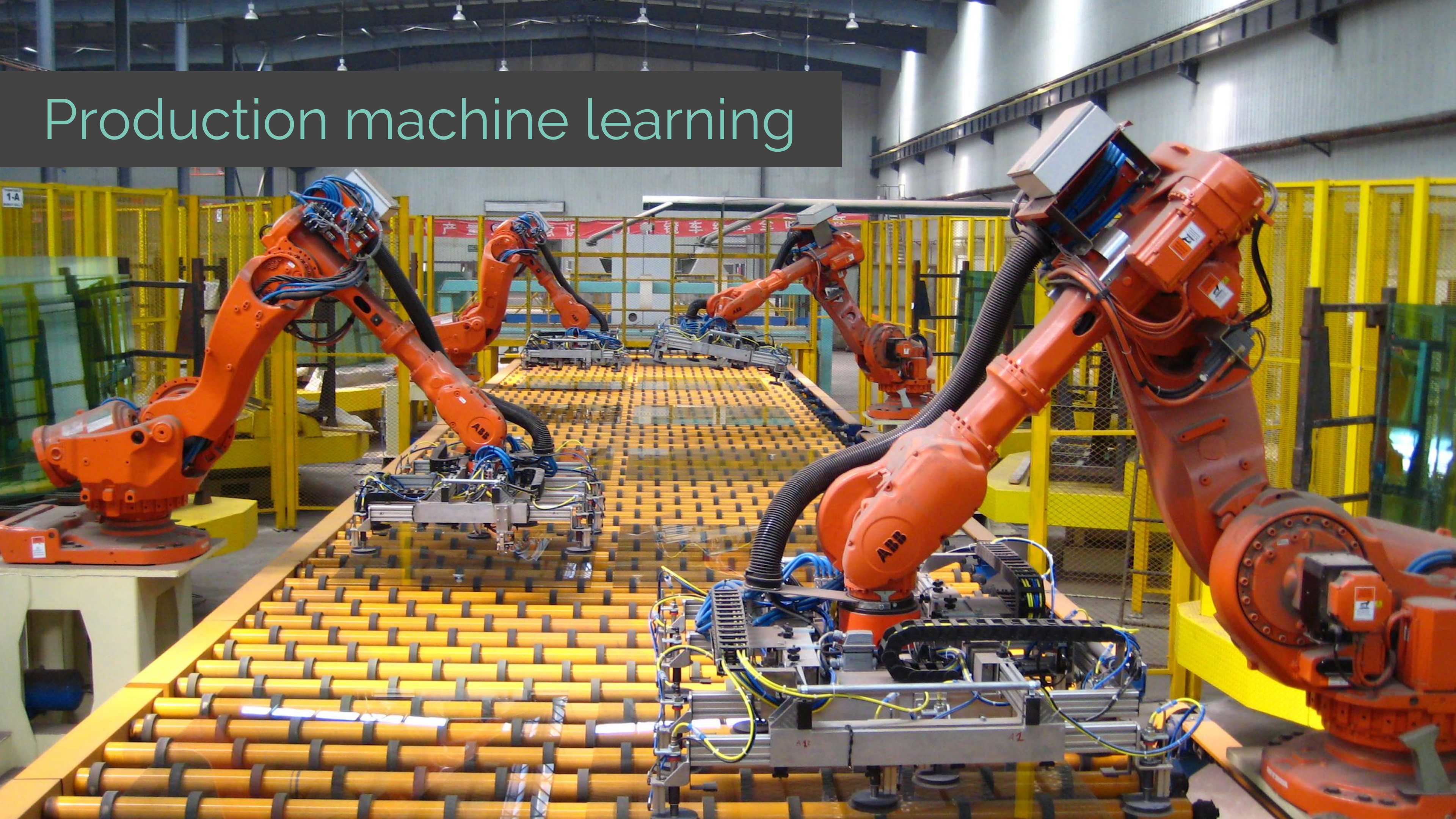
If you haven't already, start a build now. Create a new CITC client from a terminal with `prodaccess` and start a build in the background by using the following commands:

```
g4d -f mlcc_lab
blaze build //engedu/ml/ml101/estimator_lab:all
```

You will be developing code throughout the lab, but doing this build now will save time later. (And, yes, in Google3, you do have to blaze-build Python code before running it).

TensorFlow Versions

Production machine learning



Production ML is 1% inspiration

Production ML is 1% inspiration, 99% plumbing

Configuration

Data visualization

Job orchestration

Monitoring

Workflow tools

Data ingestion
↓
Feature engineering
↓
Data validation
↓
Model training
↓
Model evaluation & validation
↓
Skew detection
↓
Serving

TensorFlow Extended (TFX)

A production ML platform available to all.

Let's you focus on creating
the best ML model for your product,
not wrangling infrastructure.

Over 250 ML pipelines
checked-in across Google PAs.

Google Express

Google Play

Google Sheets

Respect Engine

Google Plus

Google Analytics

Google Maps

Google Shopping

Google AdWords

YouTube

GDN

Google Chrome

Google Knowledge Graph

Google Cloud Platform

DeepMind

Google Live Cases

pay

Android Pay

Kellogg's FLAKES

EngProd Flakes

Google Search

Google Apps for Work

Google Docs

Google Flights

Corp Eng

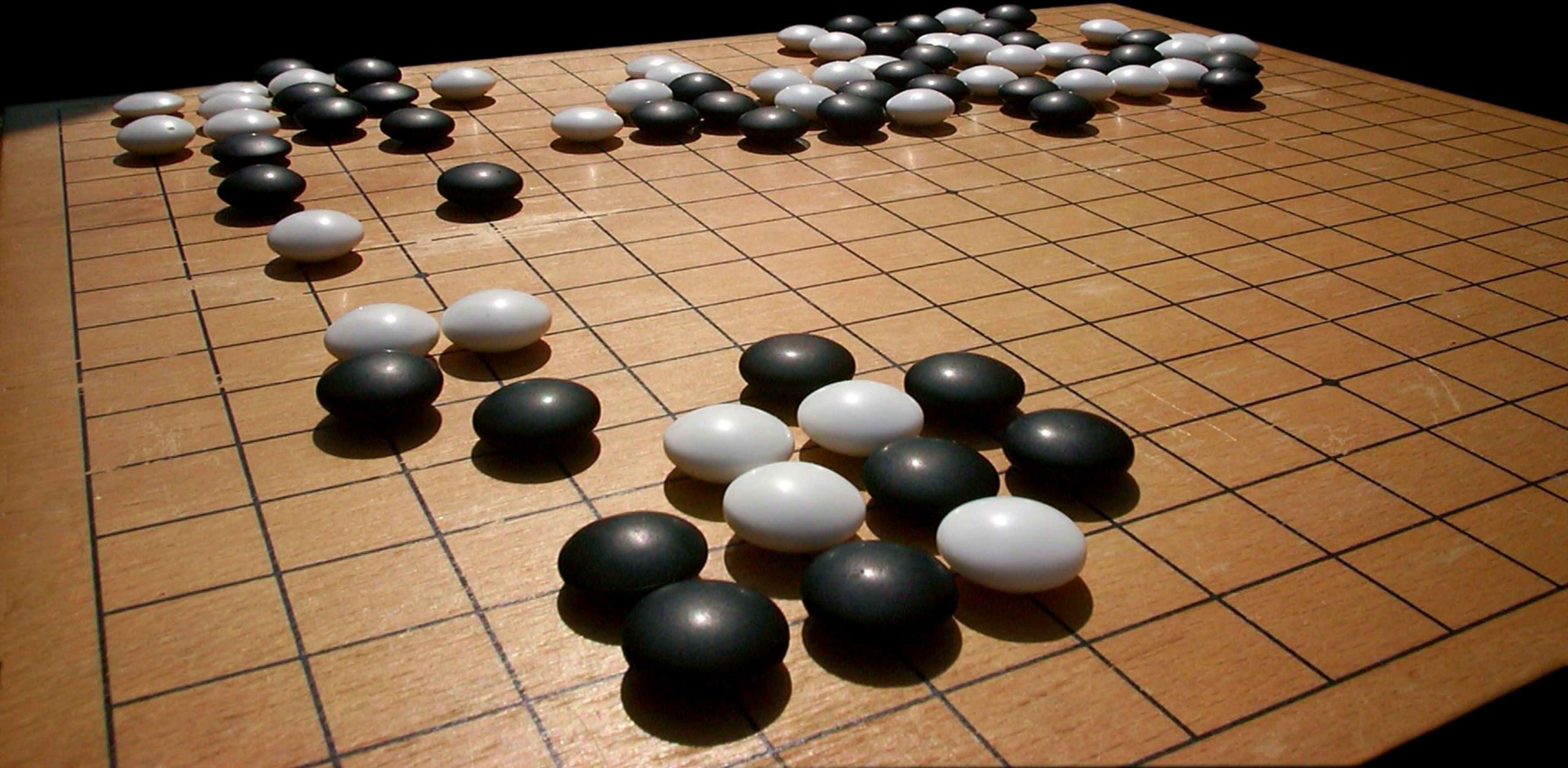
Jigsaw

+ more

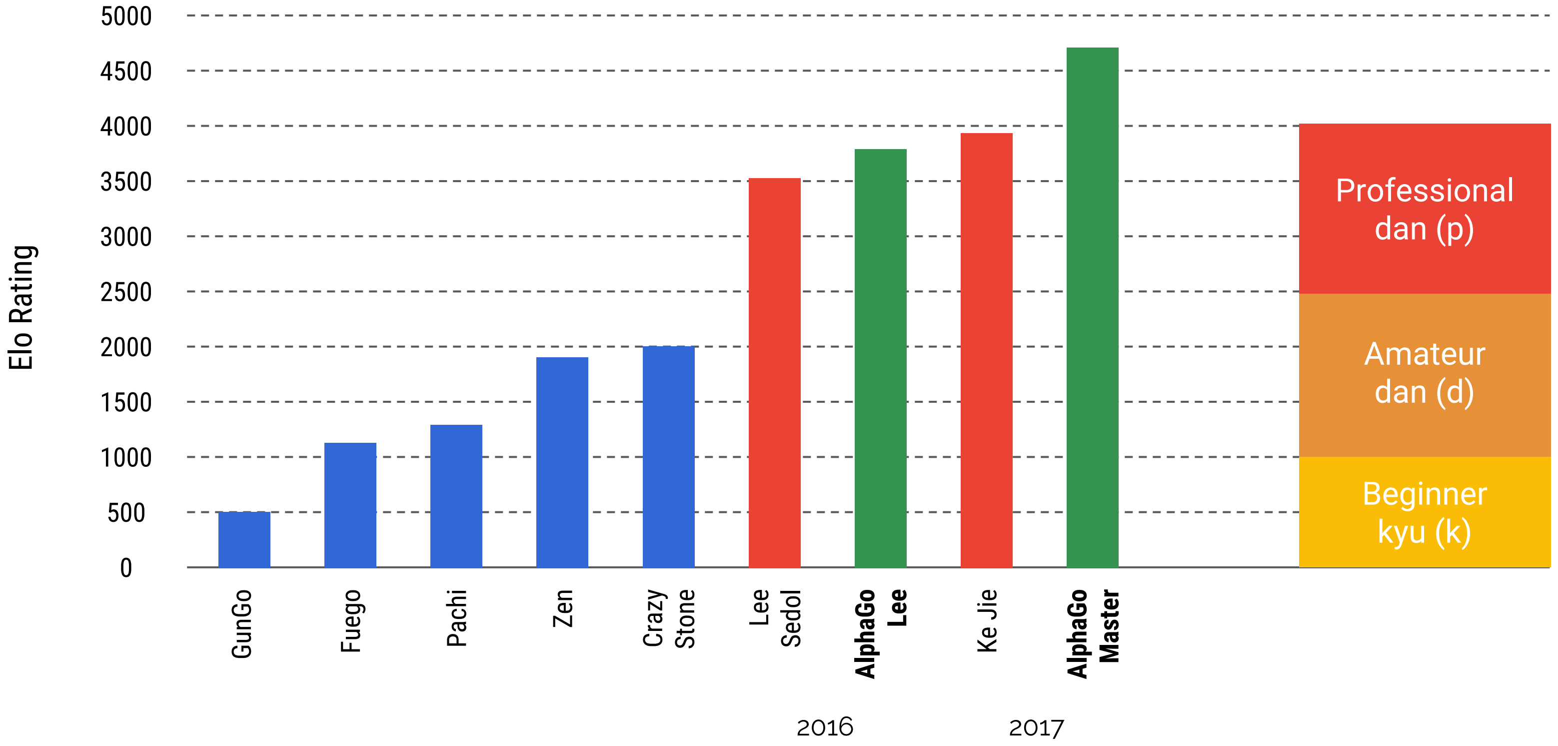
RESEARCH & MACHINE INTELLIGENCE

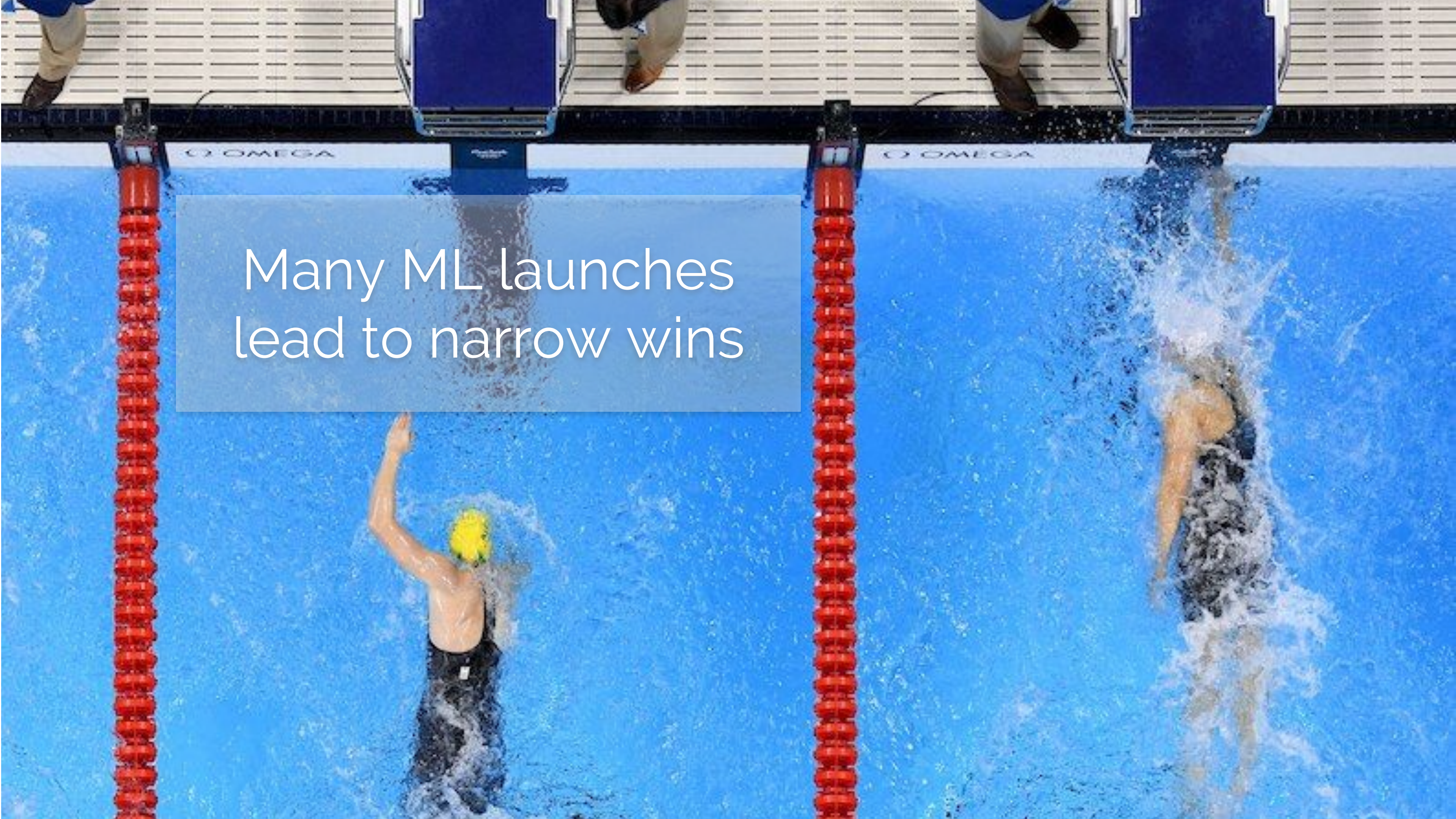
Have realistic expectations

Sometimes realistic means groundbreaking



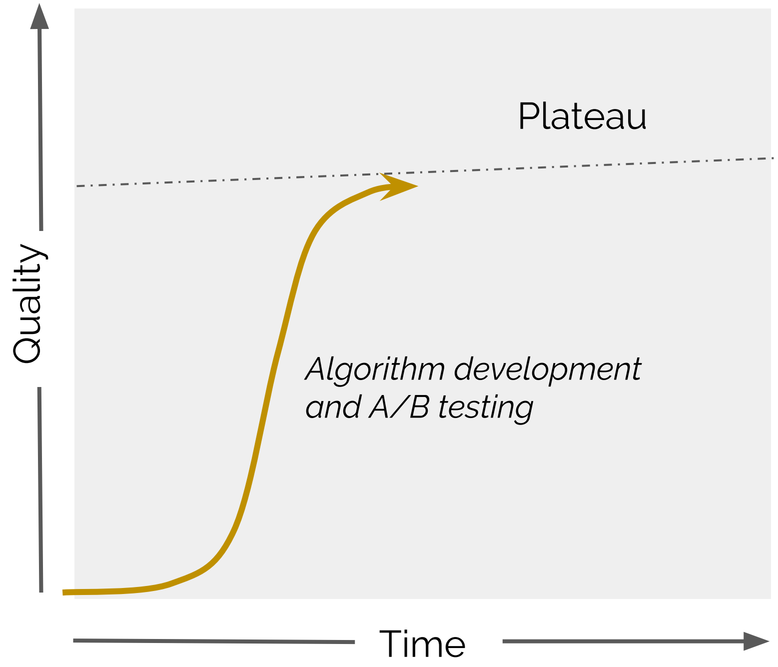
AlphaGo exceeds human ability




Many ML launches
lead to narrow wins
OMEGA
OMEGA

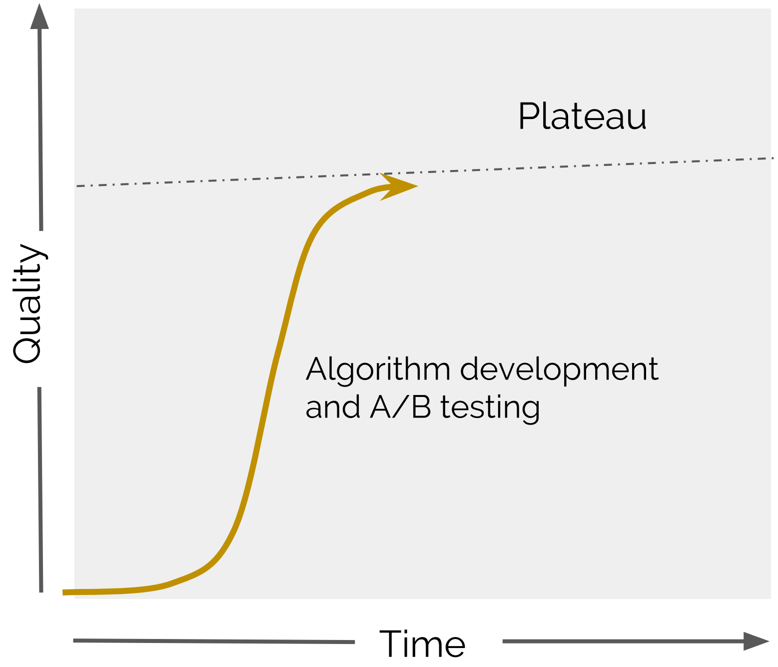
More often realistic means …

Without machine learning

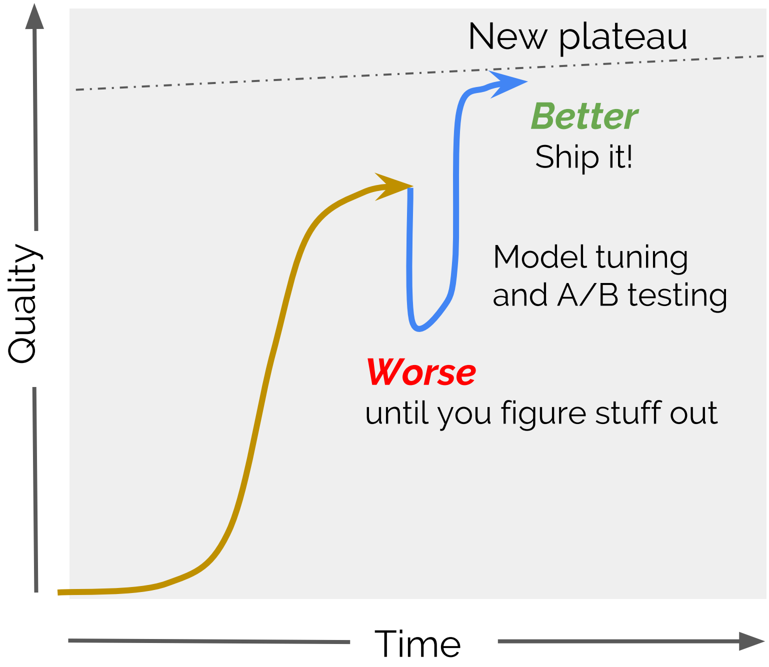


More often realistic means … modest

Without machine learning



With machine learning

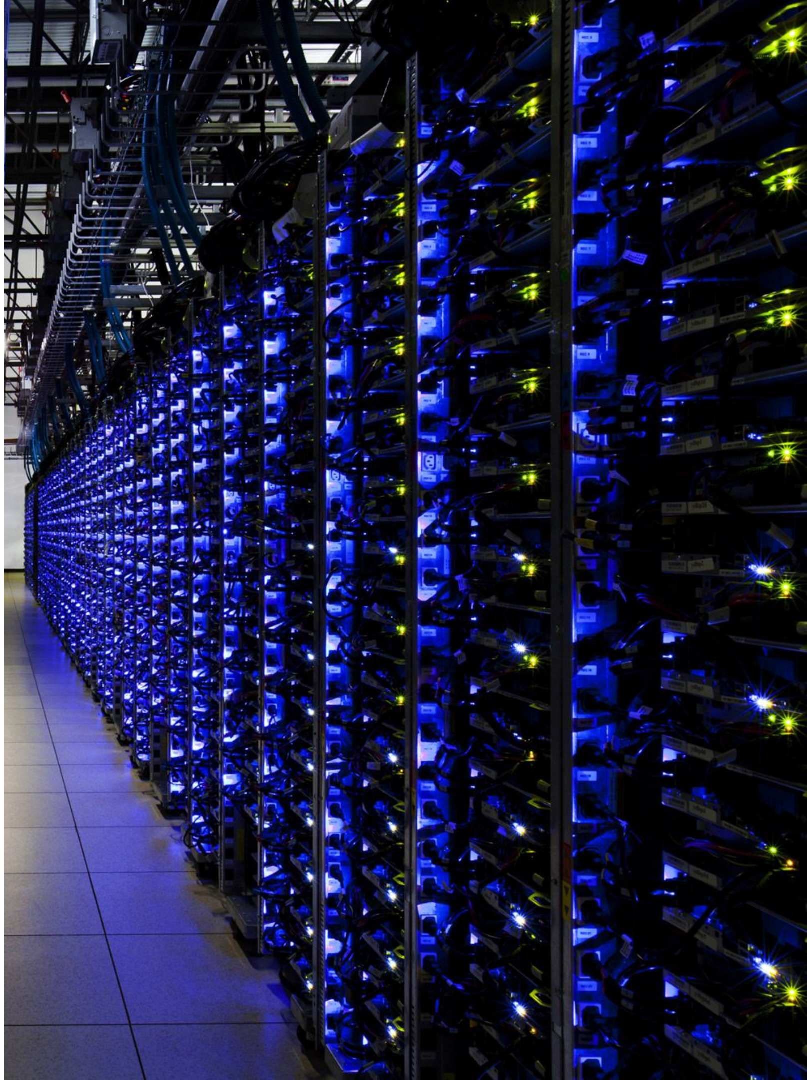


Develop. Measure. Repeat.

Google pioneered the use of data and measurement to improve products.

ML is an evolution in the same direction made possible with more data + compute.

With ML, rather than hand-tune algorithms, let computers find the best solution.

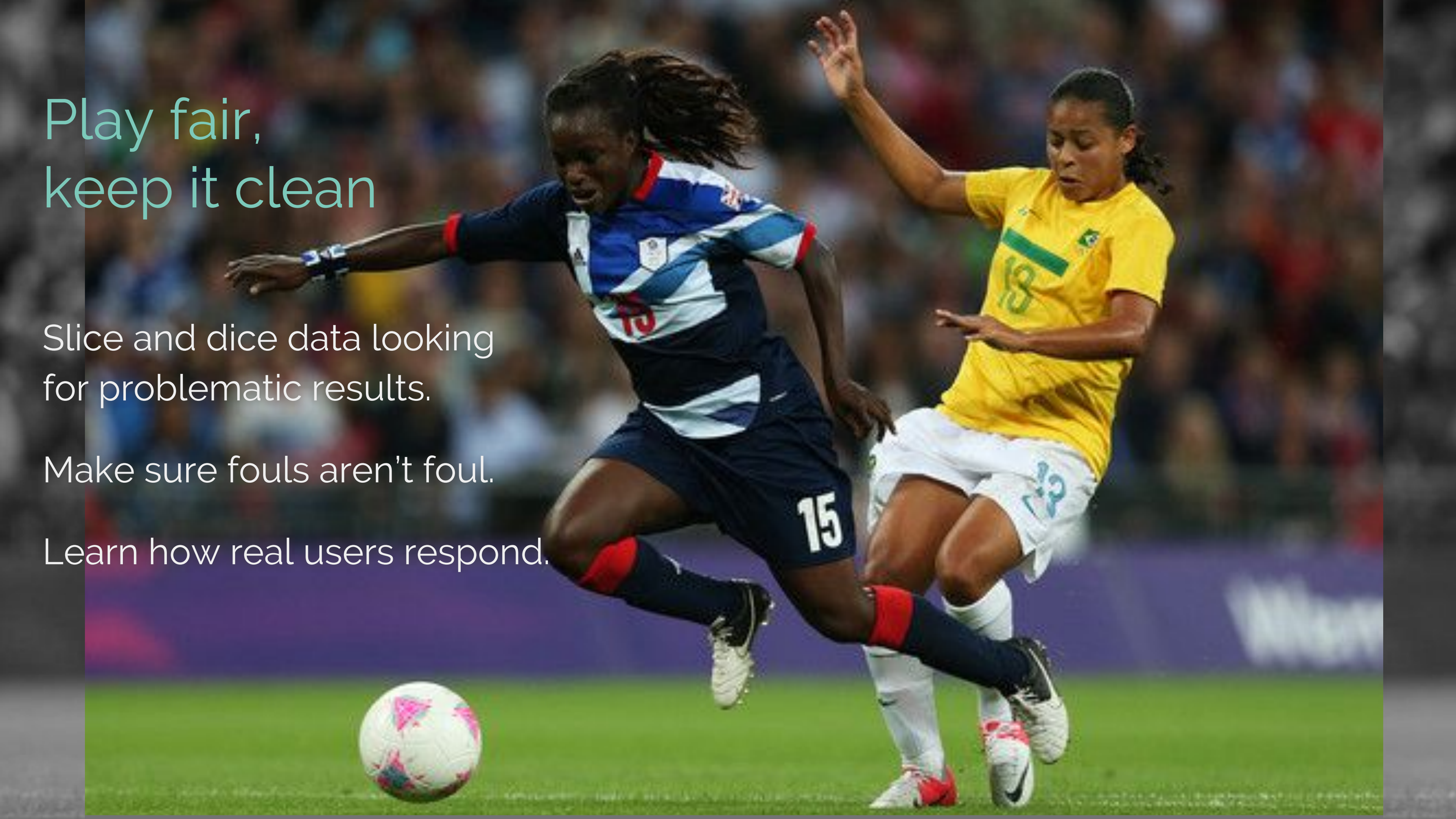


Play fair,
keep it clean

Slice and dice data looking for problematic results.

Make sure fouls aren't foul.

Learn how real users respond.



Build great products

You, too, can get an ML badge

on your Teams page.

Let's launch!

What next?

Talk to ML experts:

- go/ml-consult

Learn more:

- go/ai-first-pm
- go/ml
- go/mlcc

Thanks for listening!

binghamj@

Agenda

- ➢ Welcome
- ➢ Fairness: pbrandt@
- ➢ ML and Data: ivanku@
- ➢ Crowd Computing: pocketaces@
- ➢ Natural Language: barakt@
- ➢ On-device: ingerman@
- ➢ Human Sensing: dkaram@
- ➢ Medical Applications: lhpeng@
- ➢ Getting to Launch: binghamj@
- ➢ **Refreshing Conversations**