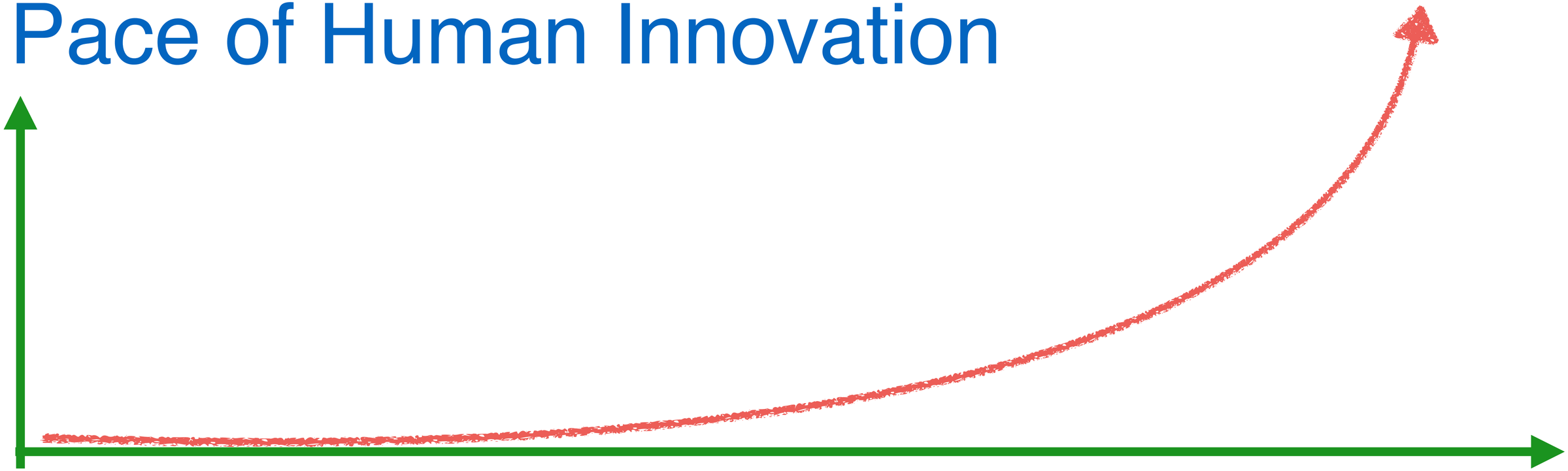
Transatlantic Sync 2019

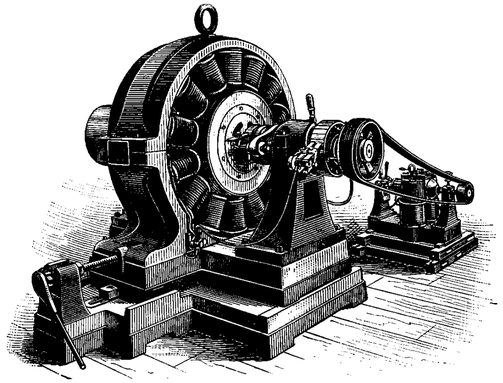
A.I. and Innovation

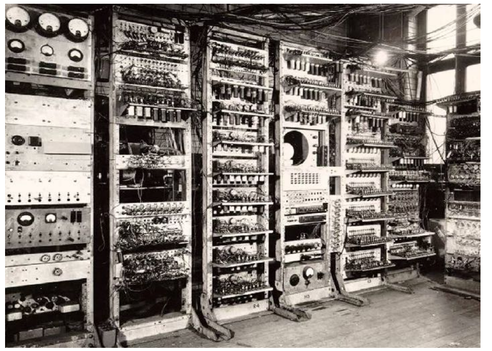
Andreas Bechtolsheim
Arista Networks

Pace of Human Innovation

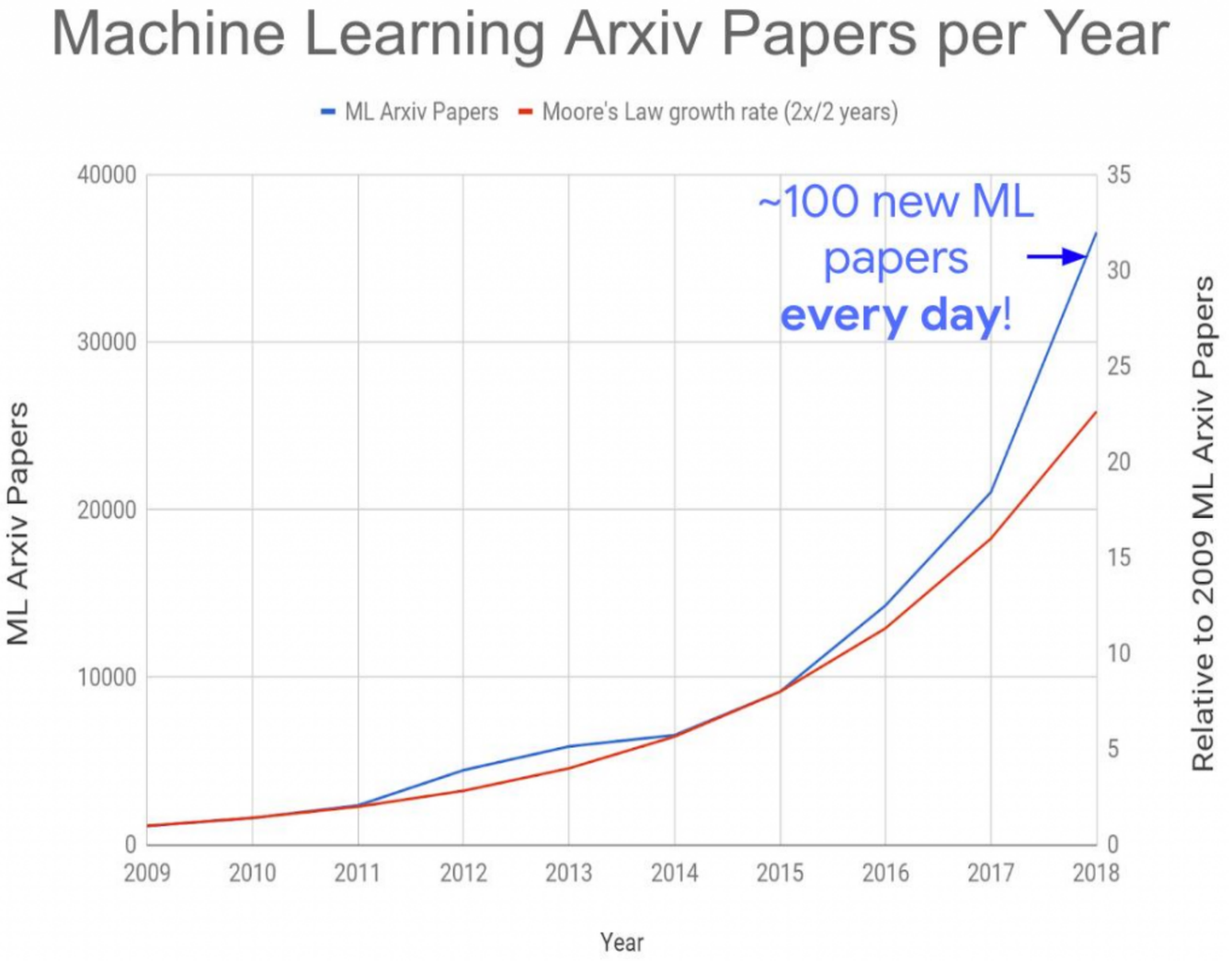


Agricultural Age | Industrial Age | Information Age





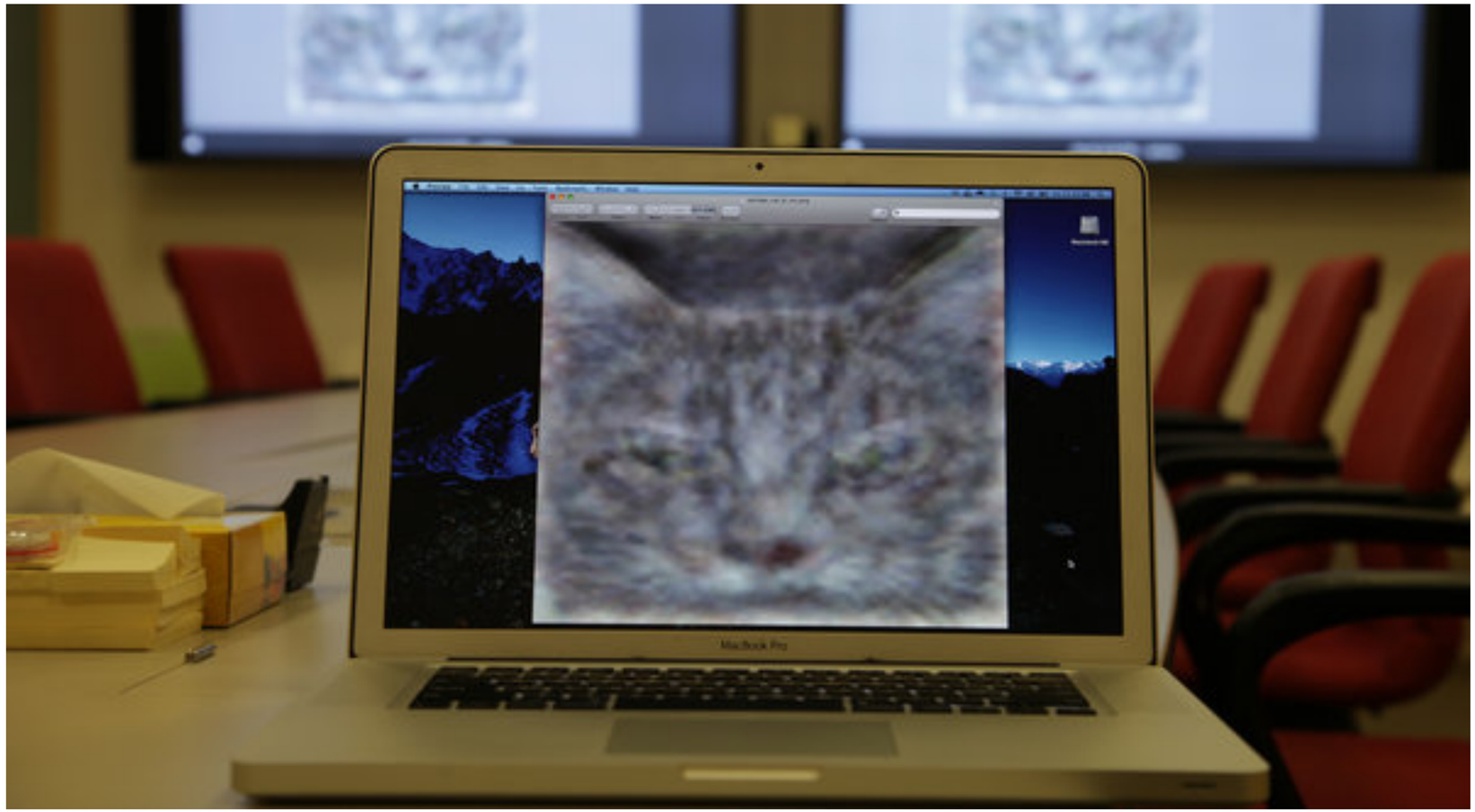
Where are we with Artificial Intelligence?



Machine Learning Papers on Arxiv:

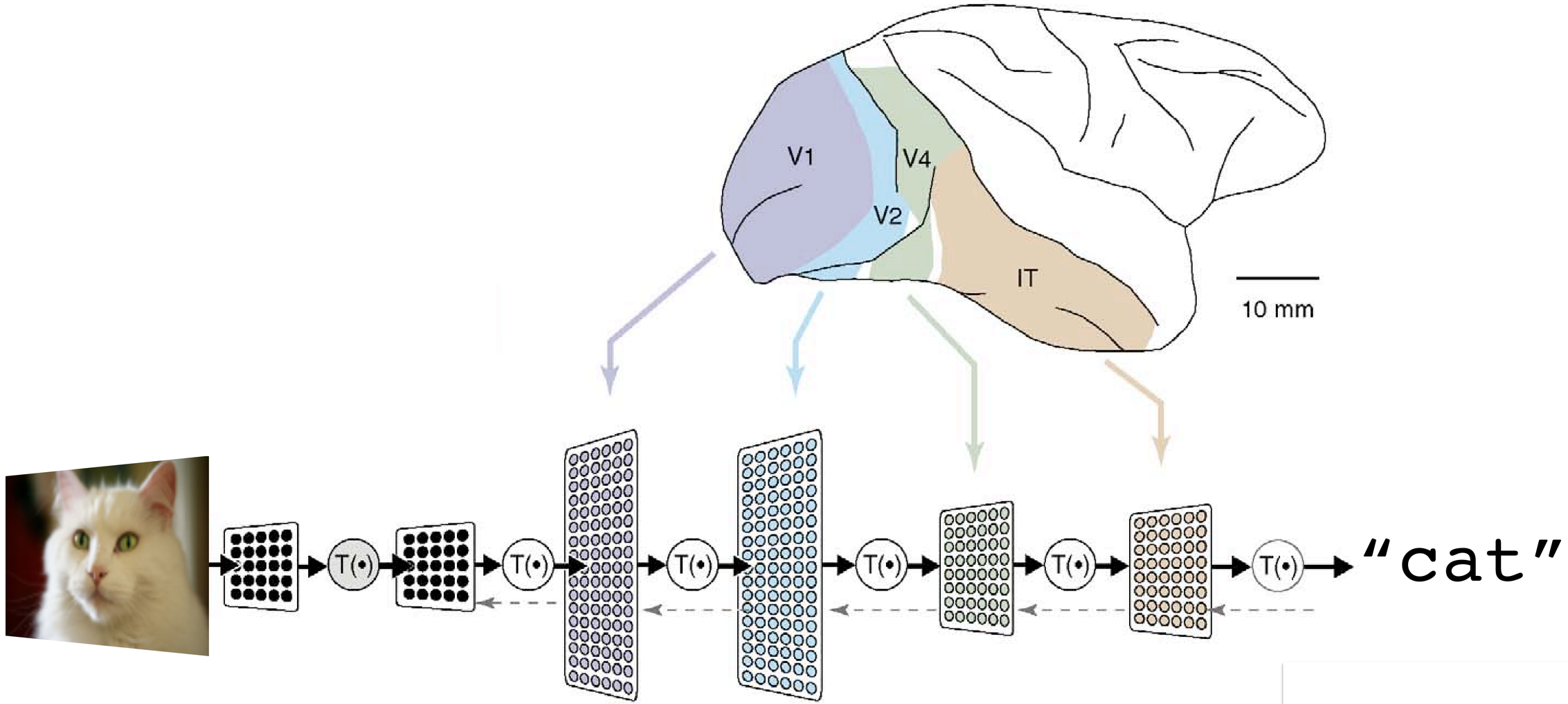
100 in CY 2009
100 per day in 2019

Beginnings of Modern AI: The Google Cat (2012)



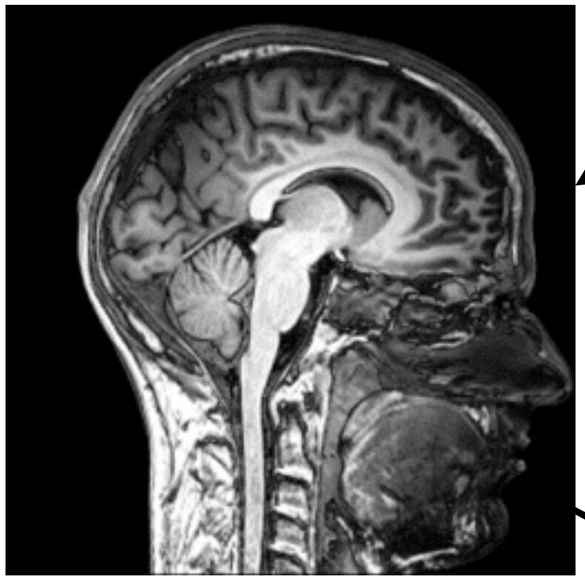
This is what the computer learned a cat looks like

Loosely inspired by the human brain



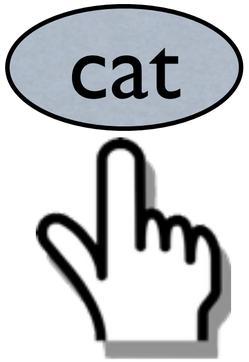
How do Human Neural Networks Work?

0.1 sec:
neurons
fire only
10 times!



see
image

click
if cat



Attributes of Human Neural Networks

Human perception is very fast (0.1 seconds)

- Recognize Objects (“see”)
- Recognize Speech (“hear”)
- Recognize Emotion (“feel”)

The efficiency of Human Neural Networks is amazing

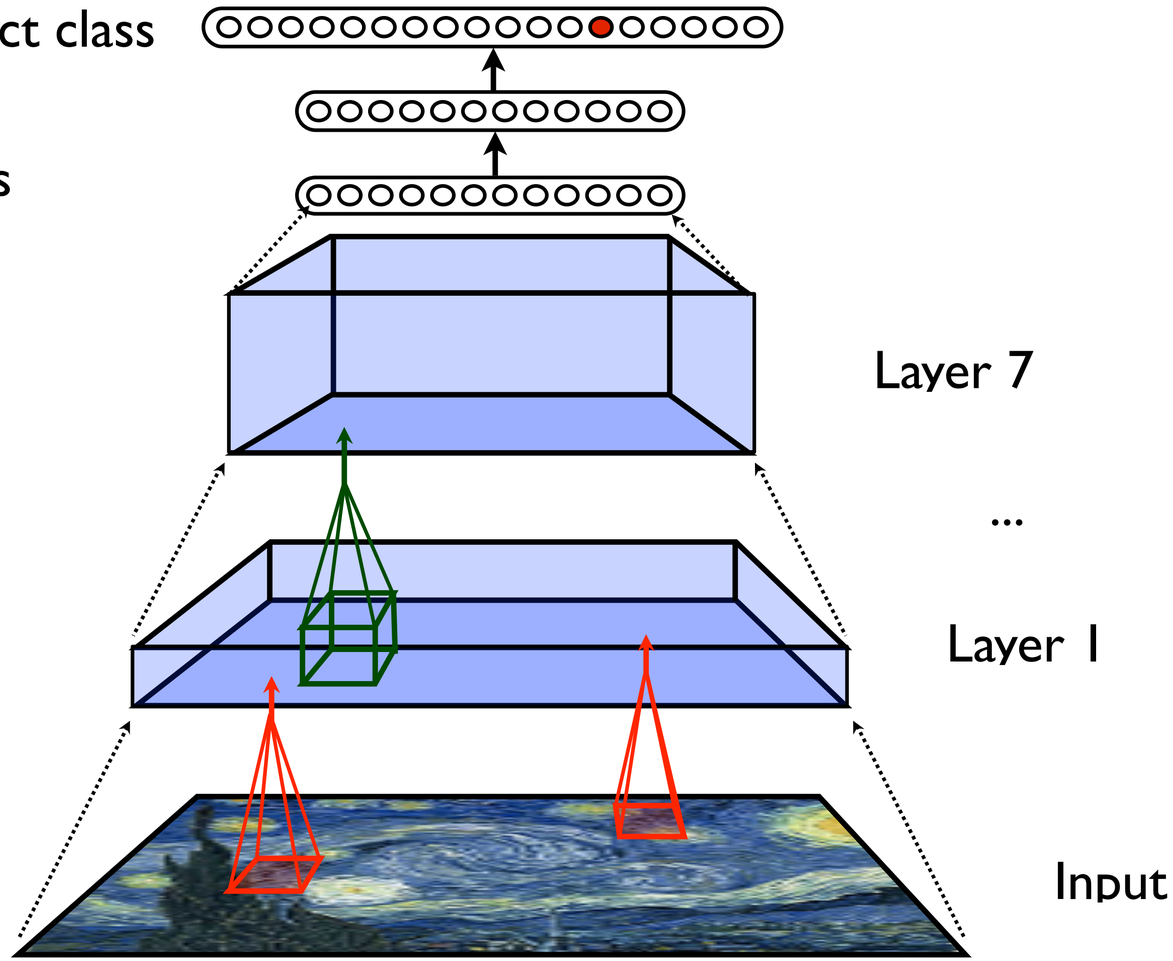
2012 Model for Object Recognition

Softmax to predict object class

Fully-connected layers

Convolutional layers
(same weights used at all spatial locations in layer)

Convolutional networks developed by Yann LeCun (NYU)



Won 2012 ImageNet Challenge with 16.4% error rate

2014 Model for Object Recognition

24 layers deep

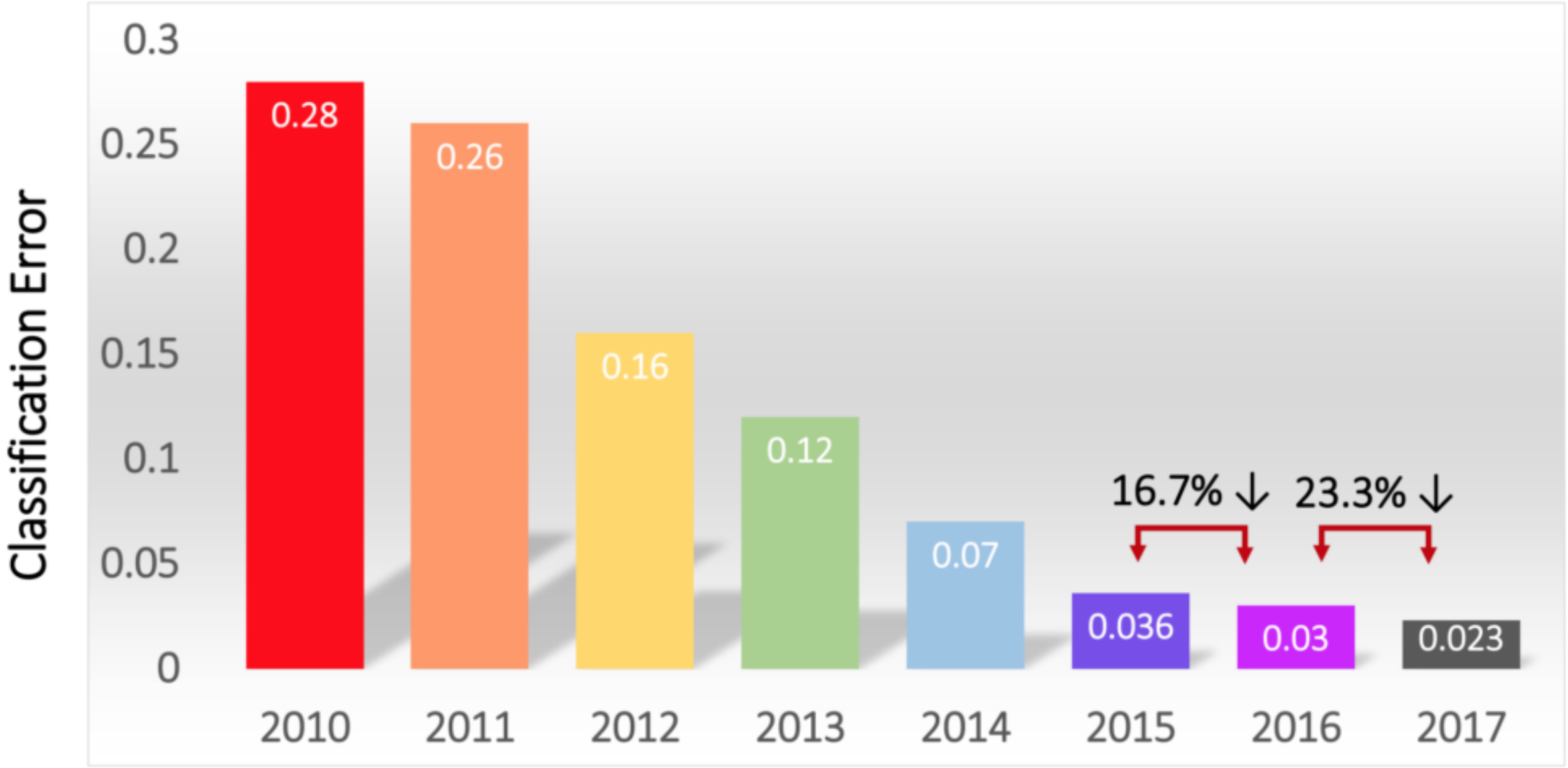
Each module has six separate convolutional layers

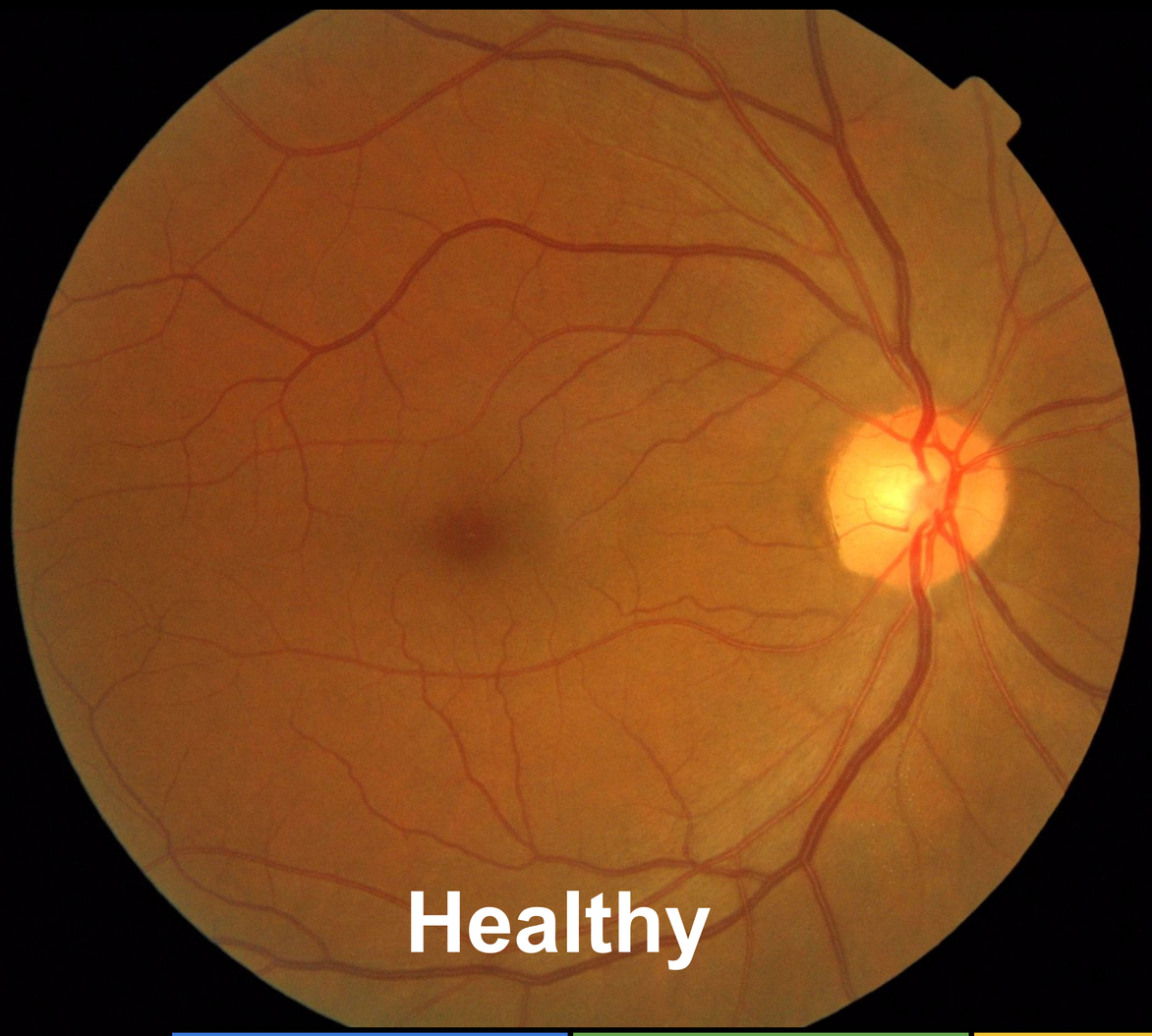
Developed by a team of Google Researchers

Won 2014 ImageNet Challenge
with 6.66% error rate

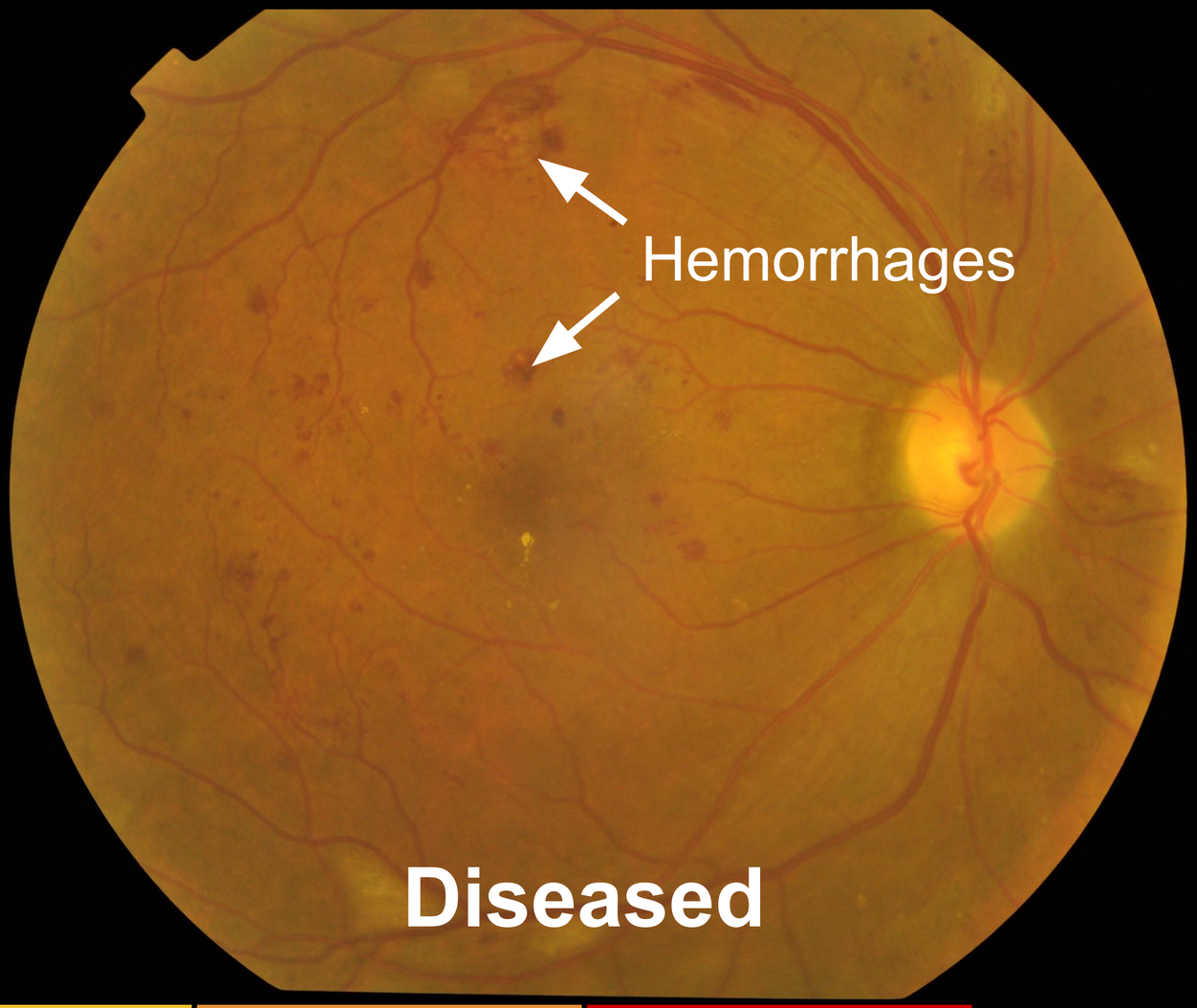
By February 2015, error rate dropped
to 5.6%, matching human performance

Classification Results (CLS)





Healthy



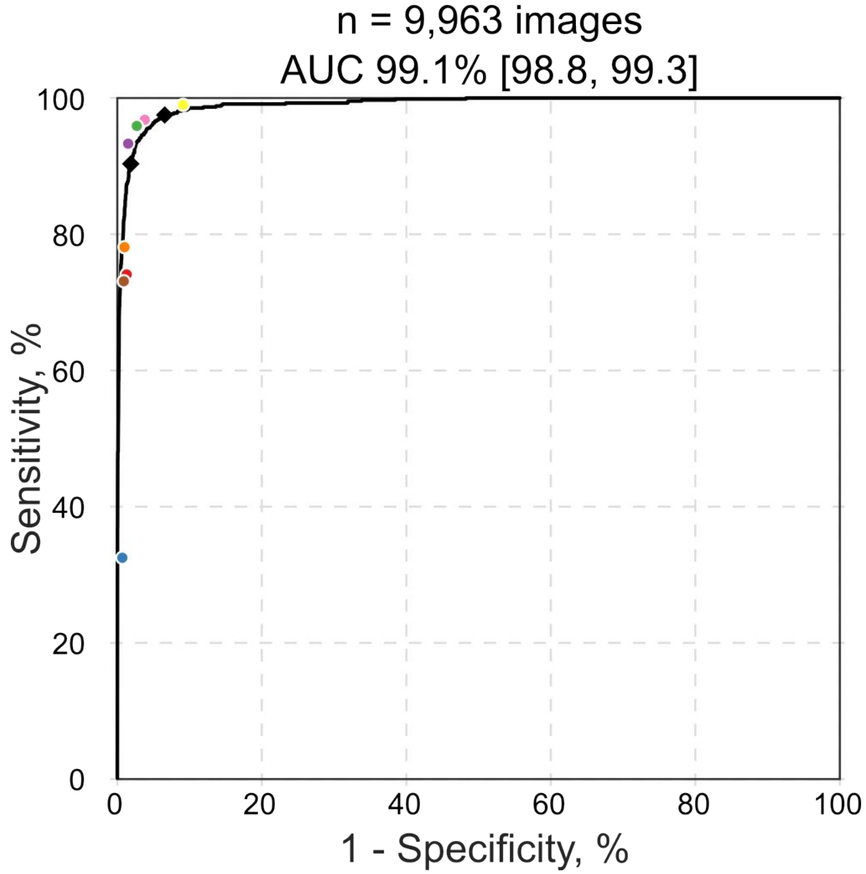
Diseased

No DR	Mild DR	Moderate DR	Severe DR	Proliferative DR
1	2	3	4	5

JAMA The Journal of the American Medical Association

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	0.91
Algorithm	Ophthalmologist (median)

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

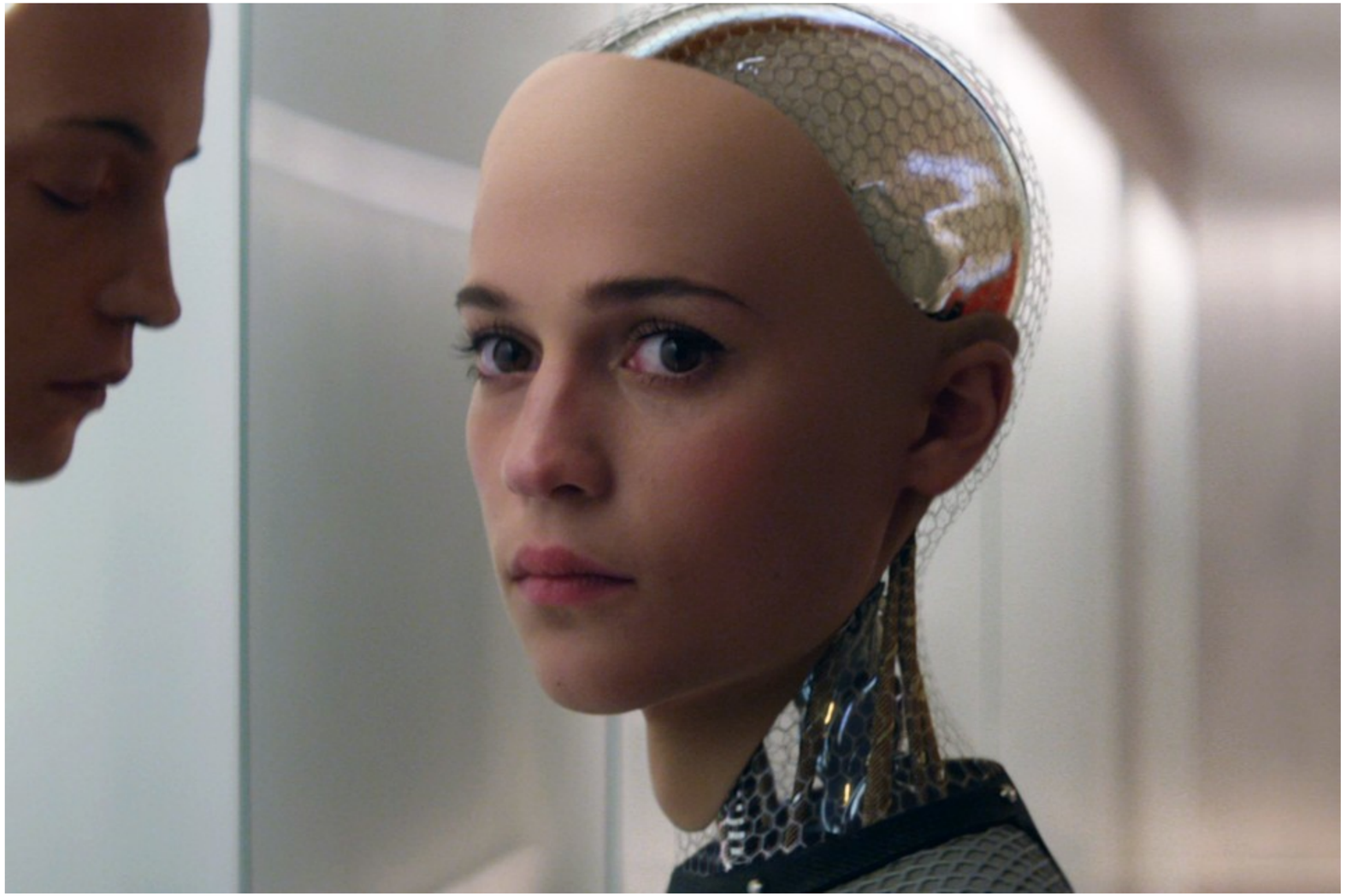
Dr. Andrew Beam, Dr. Isaac Kohane
Harvard Medical School

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

Dr. Luke Oakden-Rayner
University of Adelaide

Beyond Image Recognition: Natural Language Processing, Translation, Game Playing

Meet Jill Watson, your new Teaching Assistant



Georgia Tech 2016
Artificial Intelligence
Teaching Assistant (TA)

Computer answered all questions where it had high confidence (97%)

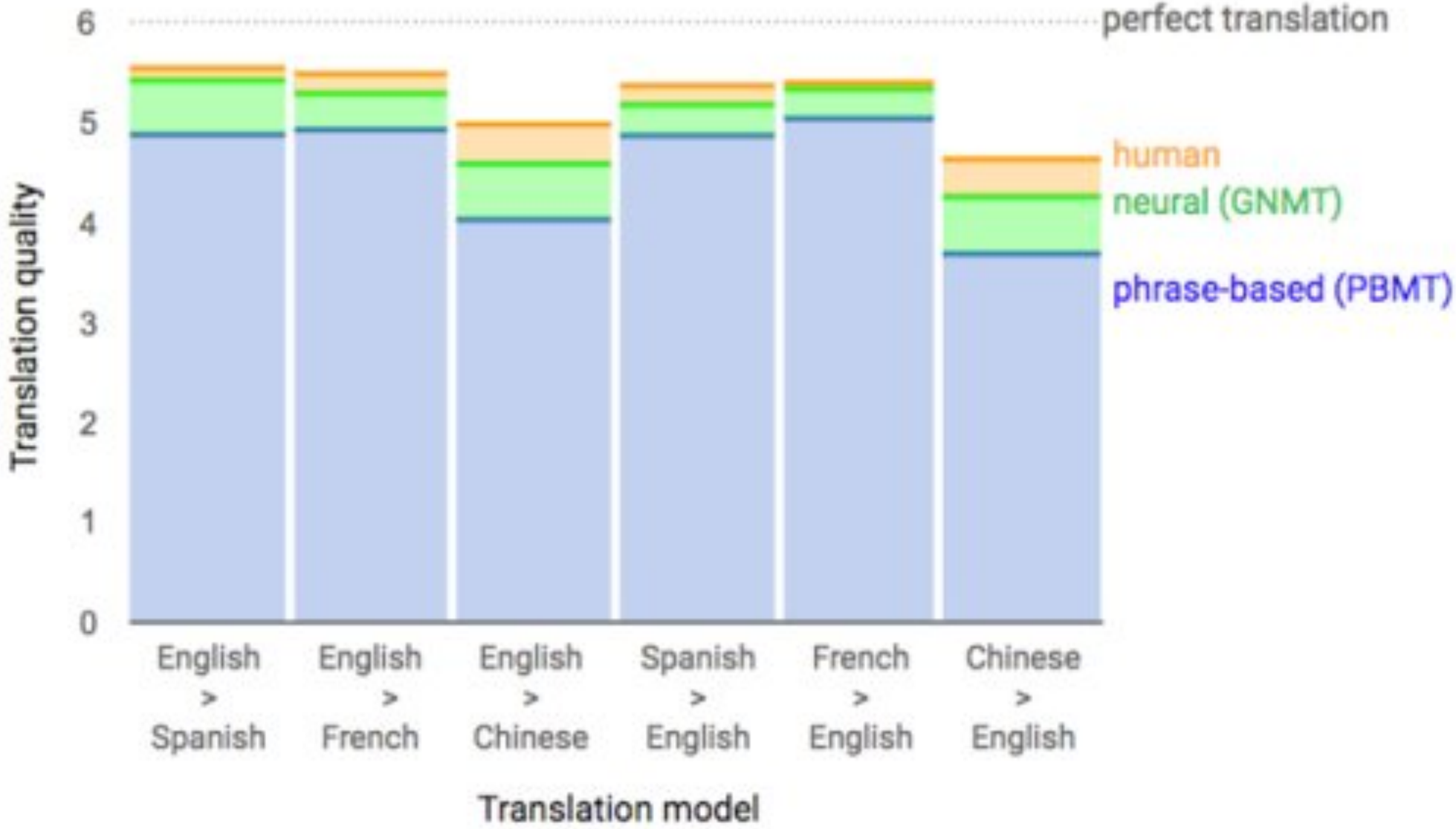
Students were unaware that teaching assistant was actually a computer

AI Beats Humans in Answering Questions

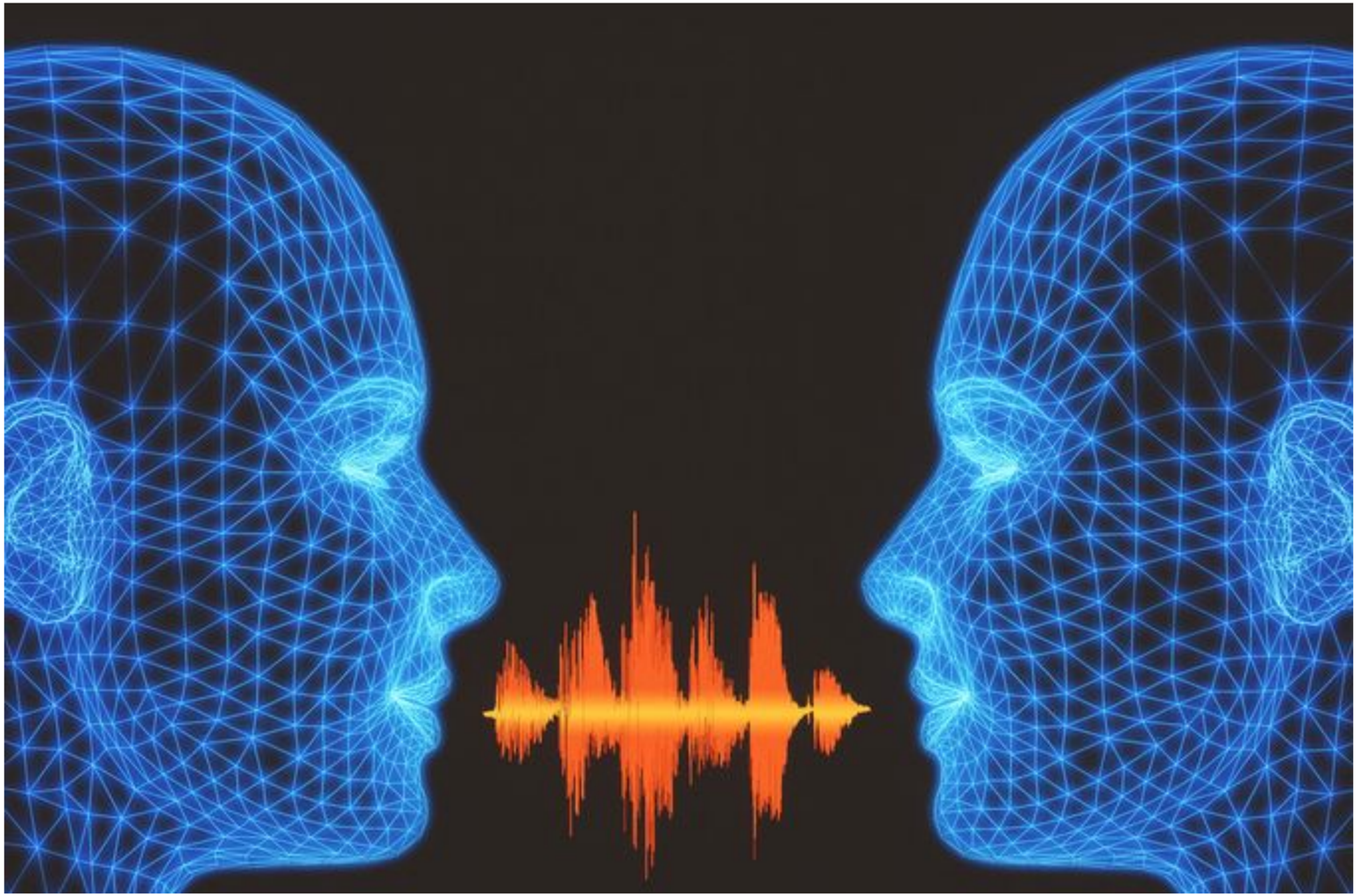
Alibaba's deep neural network and Microsoft AI have outscored humans on a Stanford University reading comprehension test, which demanded answers to more than 100,000 questions (1/2018)

Broad implications for customer service to handle inquiries

Google GNMT Translation Performance



Google Translatotron: Real-time V2V



Real-time Voice-to-Voice Translation Application

Keeps character of the source speaker voice in the output

No intermediate textual representation are used during inference

The Ancient Game of Go

Oldest board game in history (500 B.C.)

10^170 possible board positions

More than number of atoms in the universe

Cannot be mastered with brute force search

Alpha-Go, the 2016 Go World Champion

A Neural Network that:

1. Learned from existing Championship Games

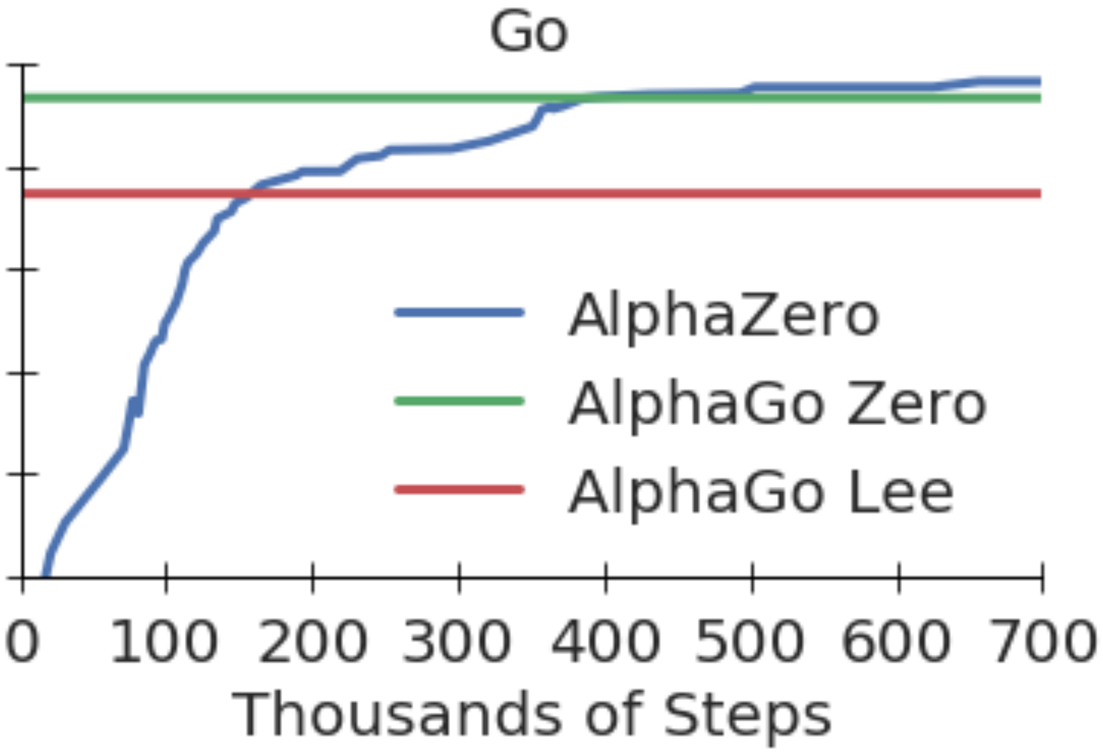
2. Practiced with the European Champion

3. Then challenged the World Champion

The Next Question

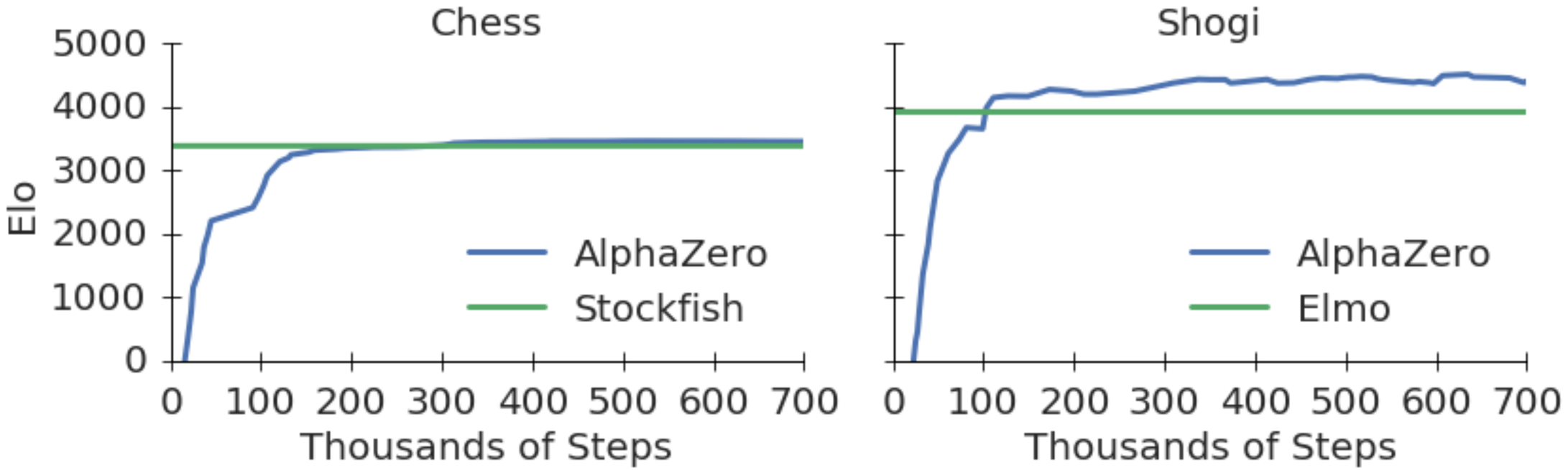
What if the computer started with nothing,
except the rules of the game?

Alpha Zero: The New Go World Champion



Alpha Zero (Late 2017 Version) beat AlphaGo Lee (2016 World Champion) after just 7 Hours of Self-Training and kept improving beyond that

Alpha Zero Performance in Chess and Shogi



Alpha Zero beat Stockfish (previous Chess Champion) in 4 Hours and Elmo (previous Shogi Champion) in just 2 Hours of Self-Training *without any prior knowledge of how to play the game and evaluating 1000X fewer positions than brute force programs*

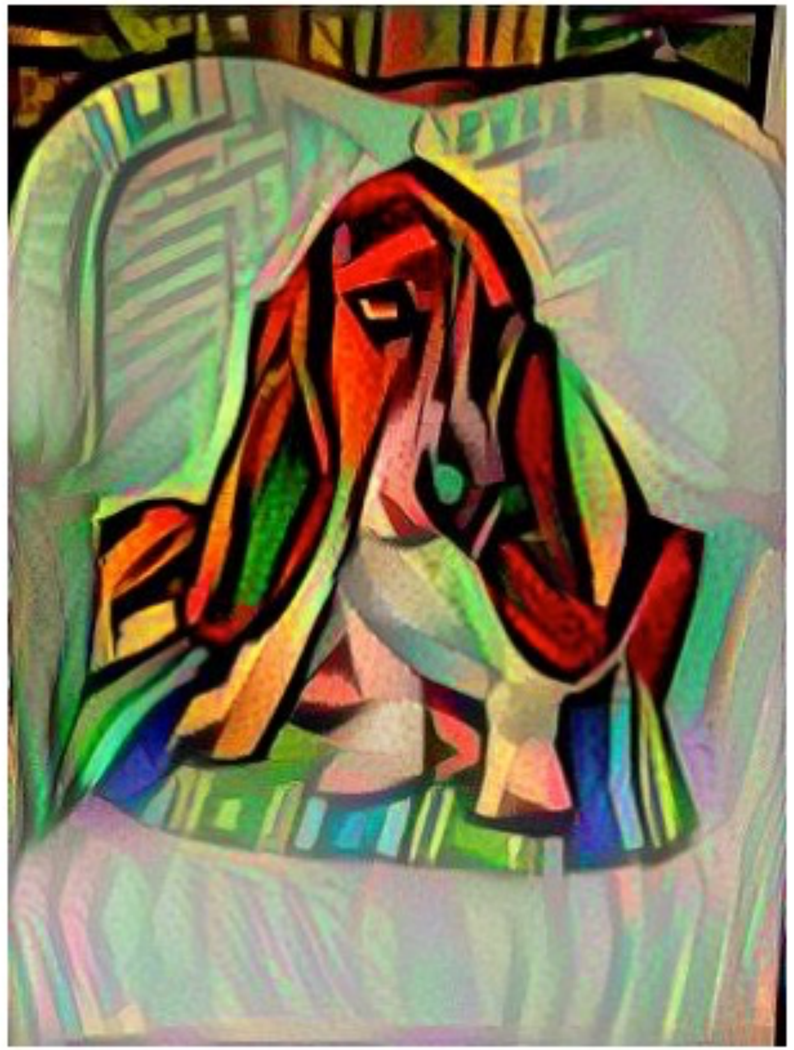
Generative Adversarial Networks (GAN)

Which of these Pictures are Real?

Style Transfer with GAN Networks

Original Picture

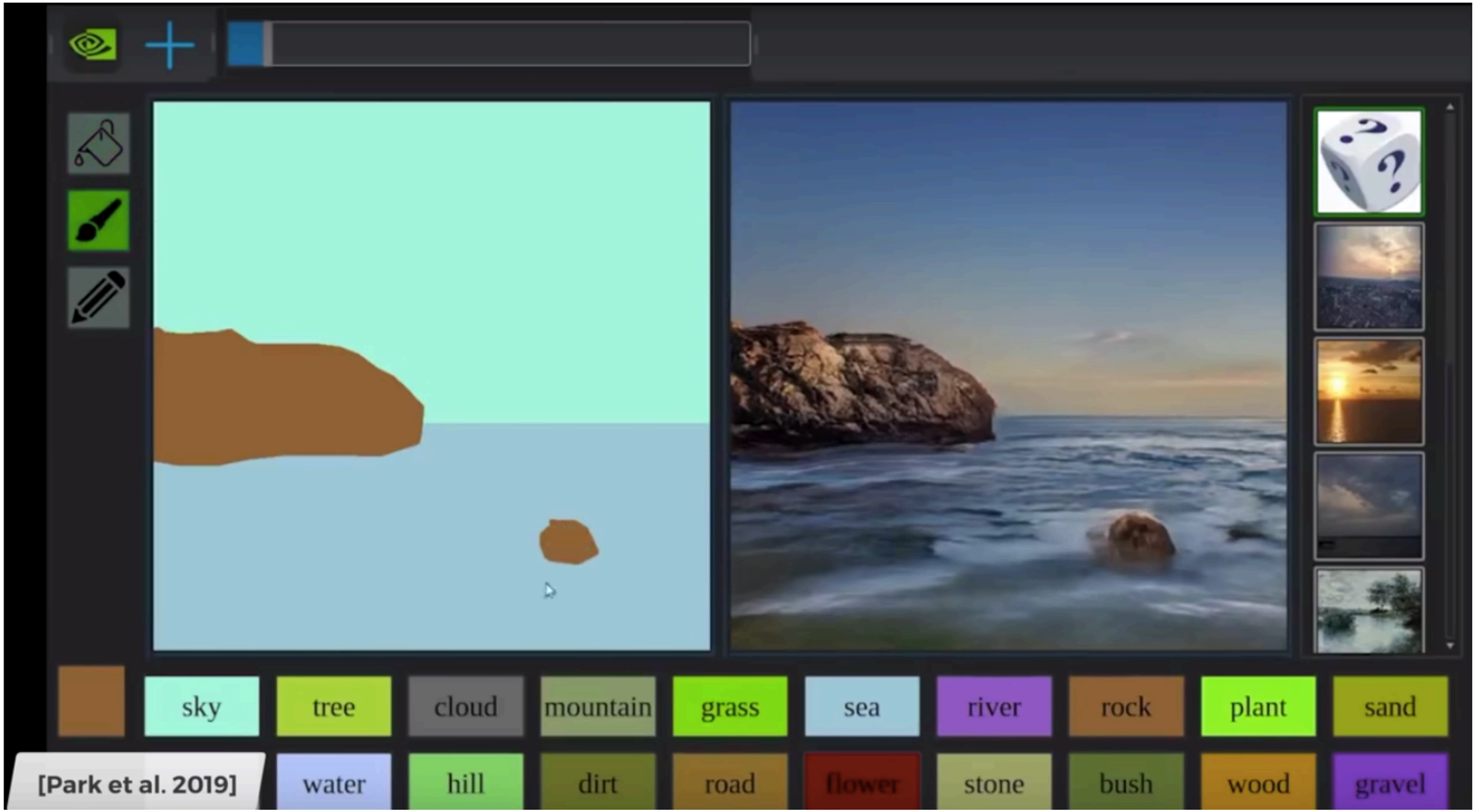
Picasso Style



Picasso Dog

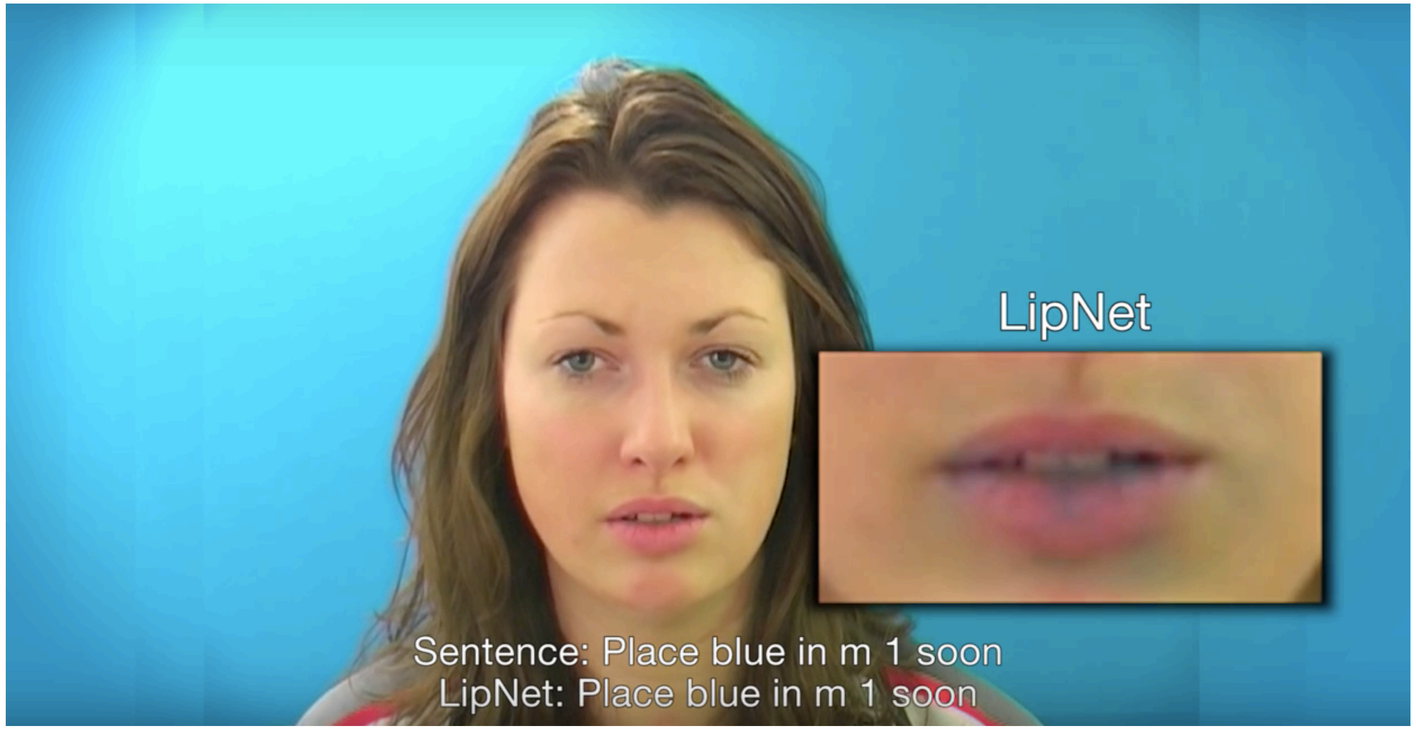

Input
Monet
Van Gogh
Cezanne
Ukiyo-e

Creating Images from Sketches (Nvidia)



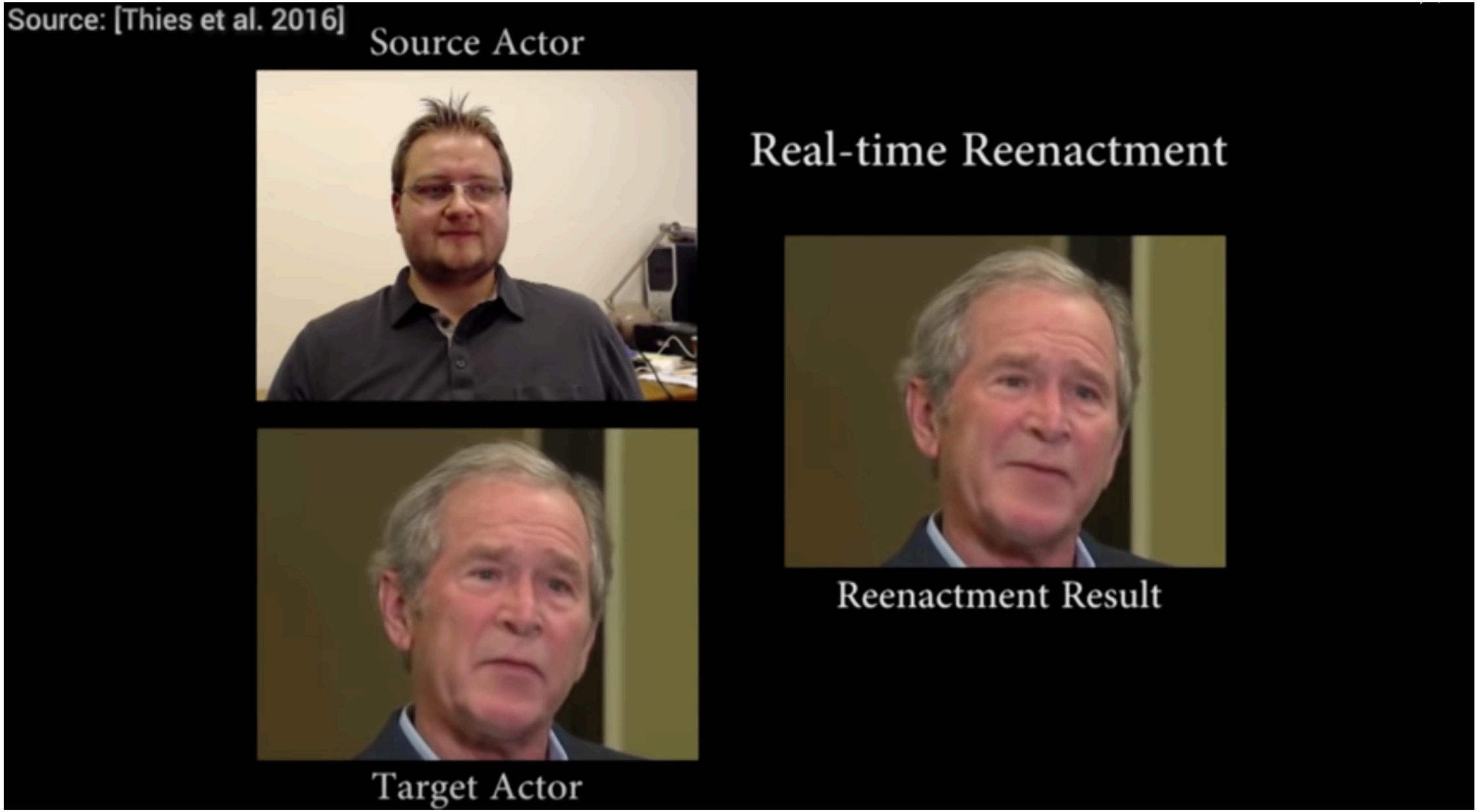
One Hour of Imaginary Celebrities (NVidia)

LipReading SuperHuman Performance

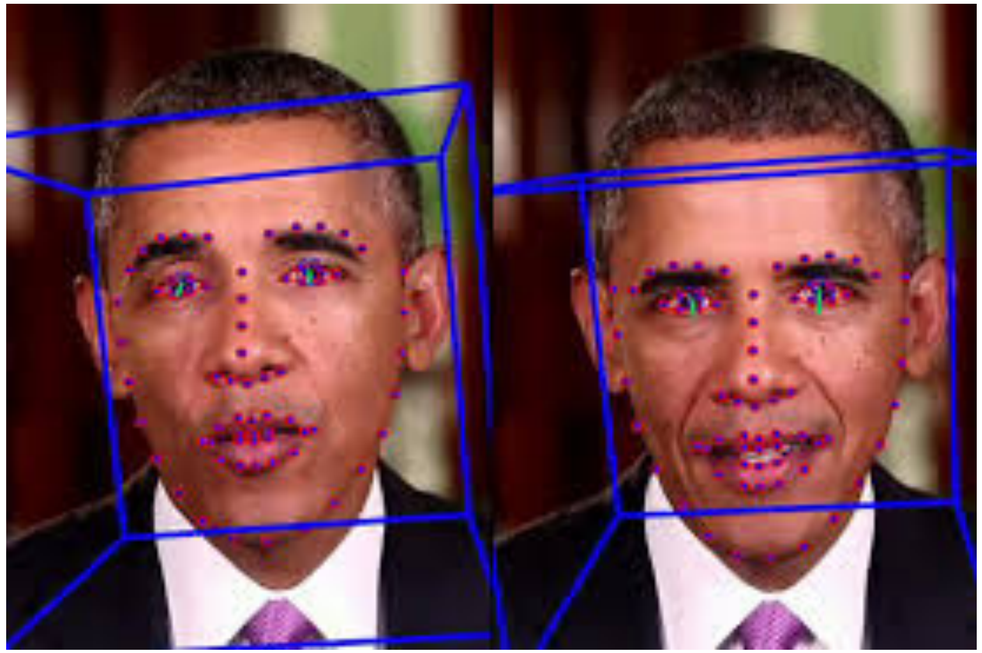


Oxford University LipNet Program: 93% Accuracy
Experienced Human Lipreader: 52% Accuracy

Real-Time Reenactment



Unmasking Deep Fake Videos



Using A.I. to unmask deep fake Videos
(Research by UC Berkeley)

FaceForensics++: Learning to Detect Manipulated Facial Images

Andreas Rössler[1] Davide Cozzolino[2] Luisa Verdoliva[2] Christian Riess[3]
Justus Thies[1] Matthias Nießner[1]

[1]Technical University of Munich [2]University Federico II of Naples [3]University of Erlangen-Nuremberg

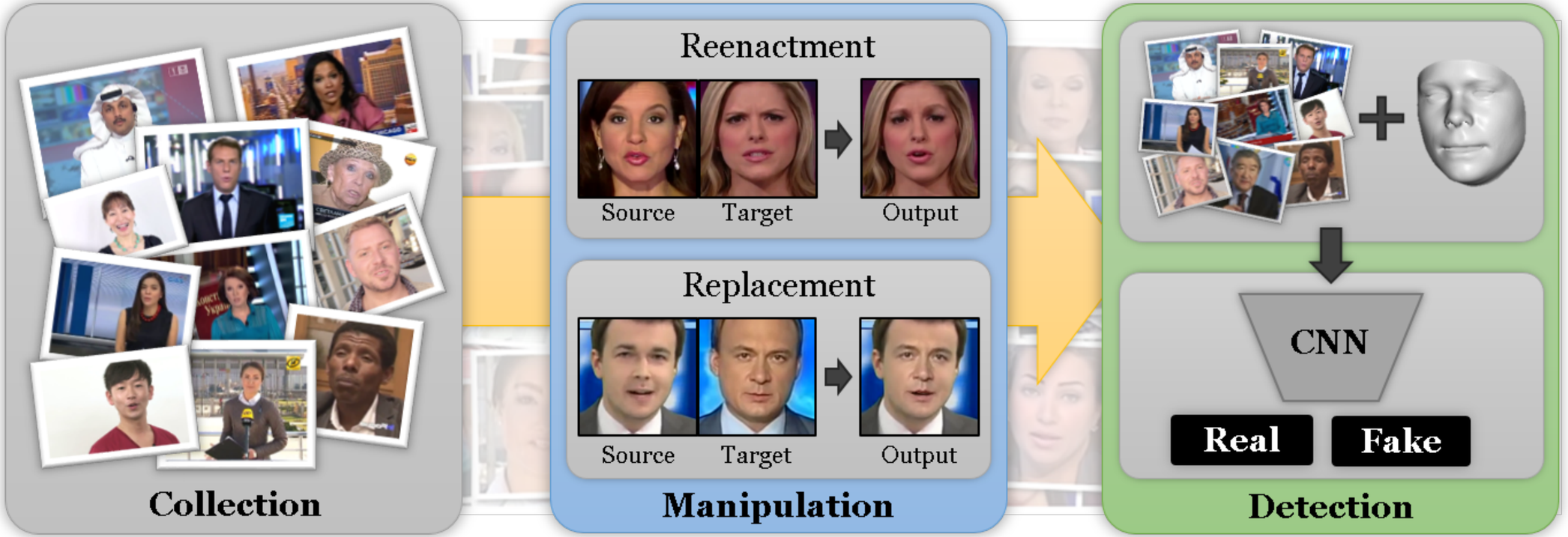


Figure 1: *FaceForensics++* is a dataset of facial forgeries that enables researchers to train deep-learning-based approaches in a supervised fashion. The dataset contains manipulations created with four state-of-the-art methods, namely, *Face2Face*, *FaceSwap*, *DeepFakes*, and *NeuralTextures*.

Fake Video Detection Accuracy

Compression	Raw	HQ	LQ
[14] XceptionNet Full Image	82.01	74.78	70.52
[27] Steg. Features + SVM	97.63	70.97	55.98
[17] Cozzolino *et al.*	98.57	78.45	58.69
[10] Bayar and Stamm	98.74	82.97	66.84
[51] Rahmouni *et al.*	97.03	79.08	61.18
[5] MesoNet	95.23	83.10	70.47
[14] XceptionNet	**99.26**	**95.73**	**81.00**

Table 1: Binary detection accuracy of our baselines when trained on all four manipulation methods. Besides the naïve full image XceptionNet, all methods are trained on a conservative crop (enlarged by a factor of 1.3) around the center of the tracked face.

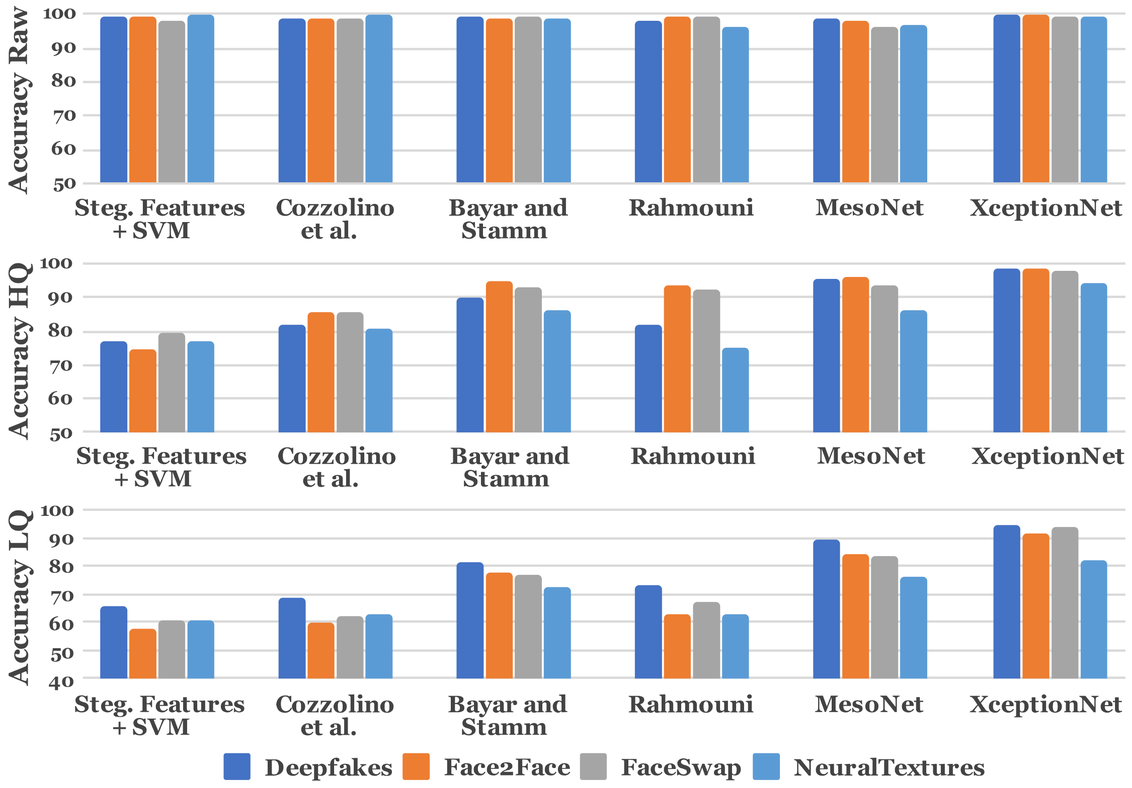


Figure 6: Binary detection accuracy of all evaluated architectures on the different manipulation methods using face tracking when trained on our different manipulation methods separately.

Infrastructure for A.I.

AI in the Cloud creates Very Large Demands

Training

Terabytes to **Exabytes** of training data sets

Continuous Self-learning multiplies the computational load

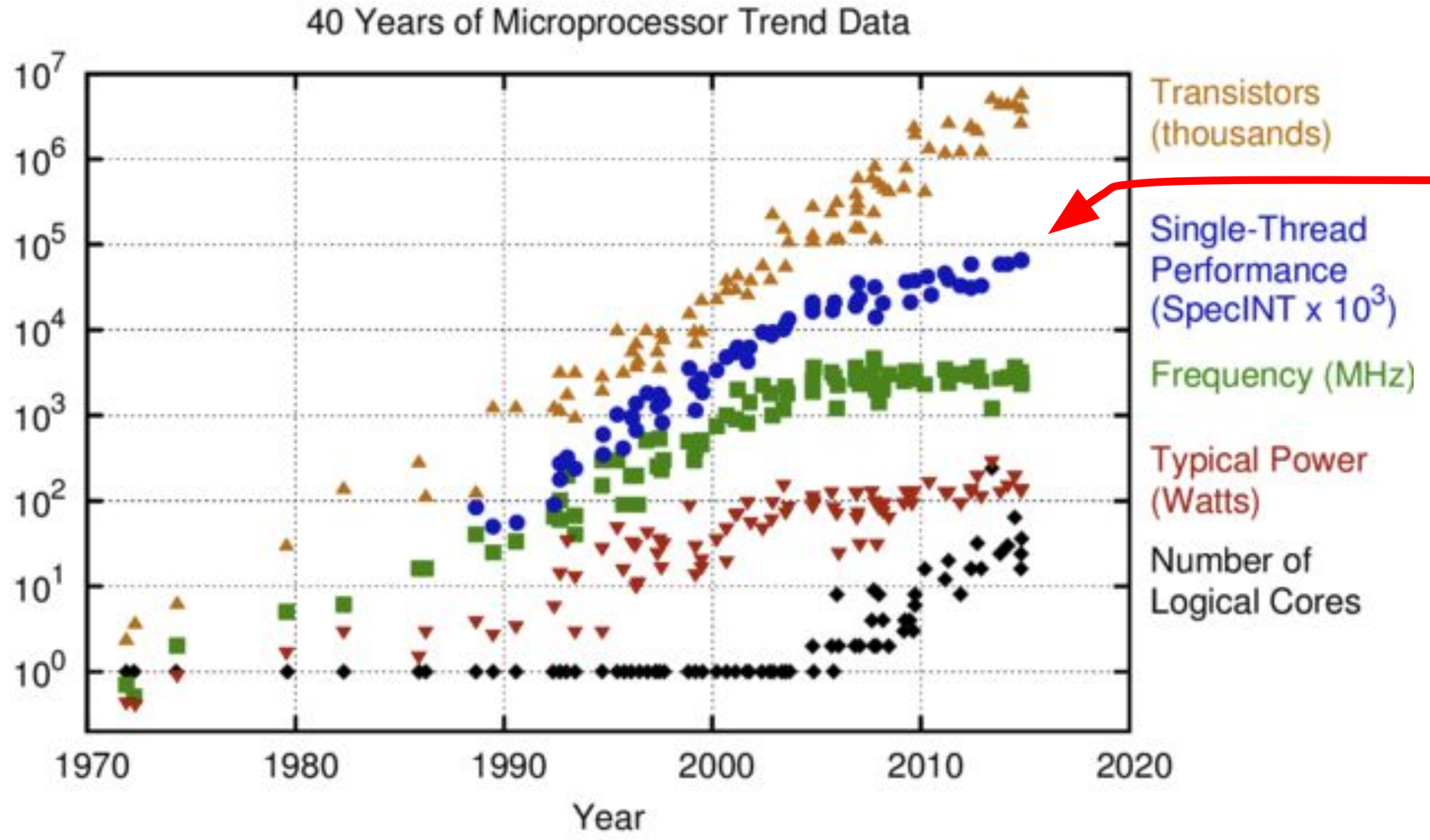
Inference

Billions of users, up to a **Million** requests per second

Latency Requirements in the 10s of milliseconds

100X to 1000X More Throughput Required Compared to Today

Traditional CPU Throughput hitting Limits



Single-core performance plateauing after decades of exponential growth

Graph from 40 Years of Microprocessor Trend Data, Karl Rupp, CC-BY 4.0.

AI needs High Throughput, not High Precision

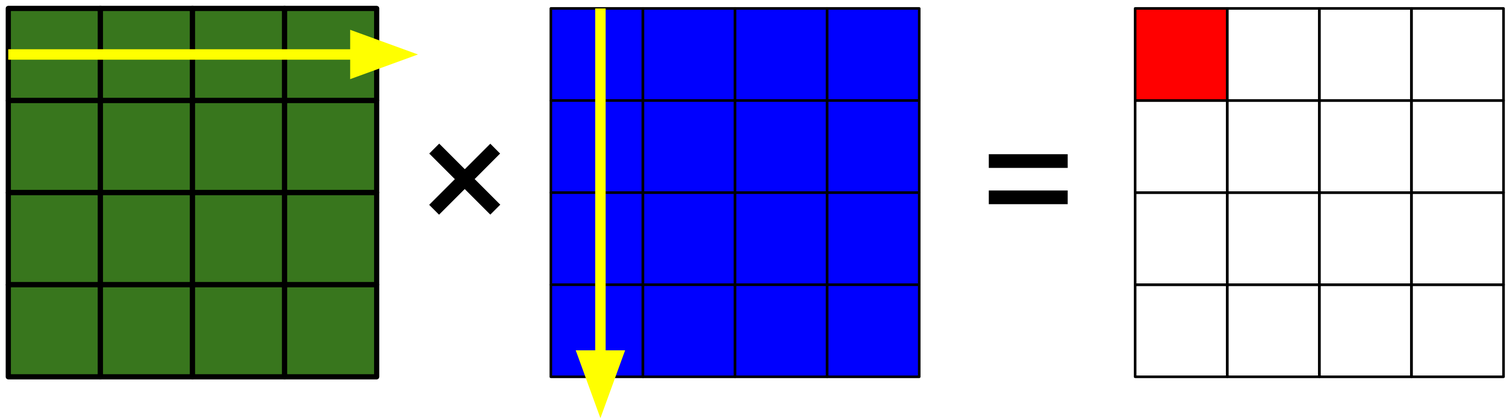
reduced precision ok

about 1.2
× about 0.6
about 0.7

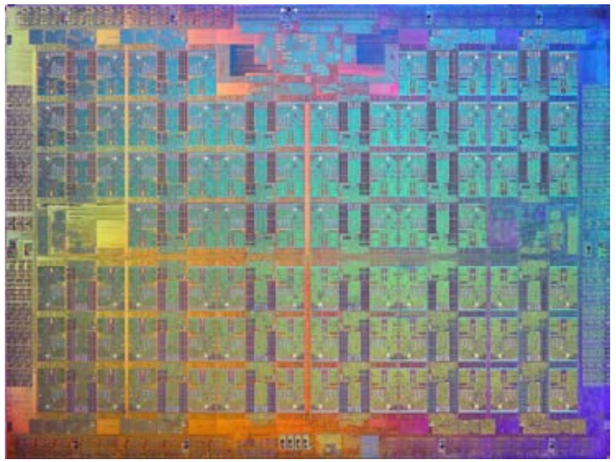
NOT

~~1.21042~~
~~× 0.61127~~
~~0.73989343~~

handful of specific operations



Processors are Built for Specific Workloads

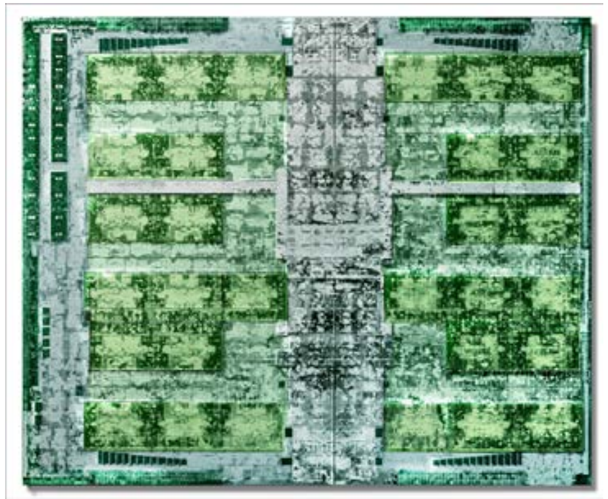


CPU

Scalar Processing

Designed for office apps

Evolved for web servers

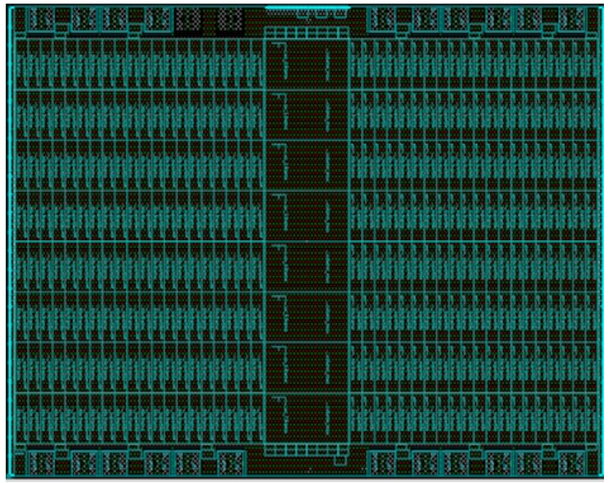


GPU

Vector Processing

Designed for graphics

Evolved for linear algebra



TPU

Systolic Processing

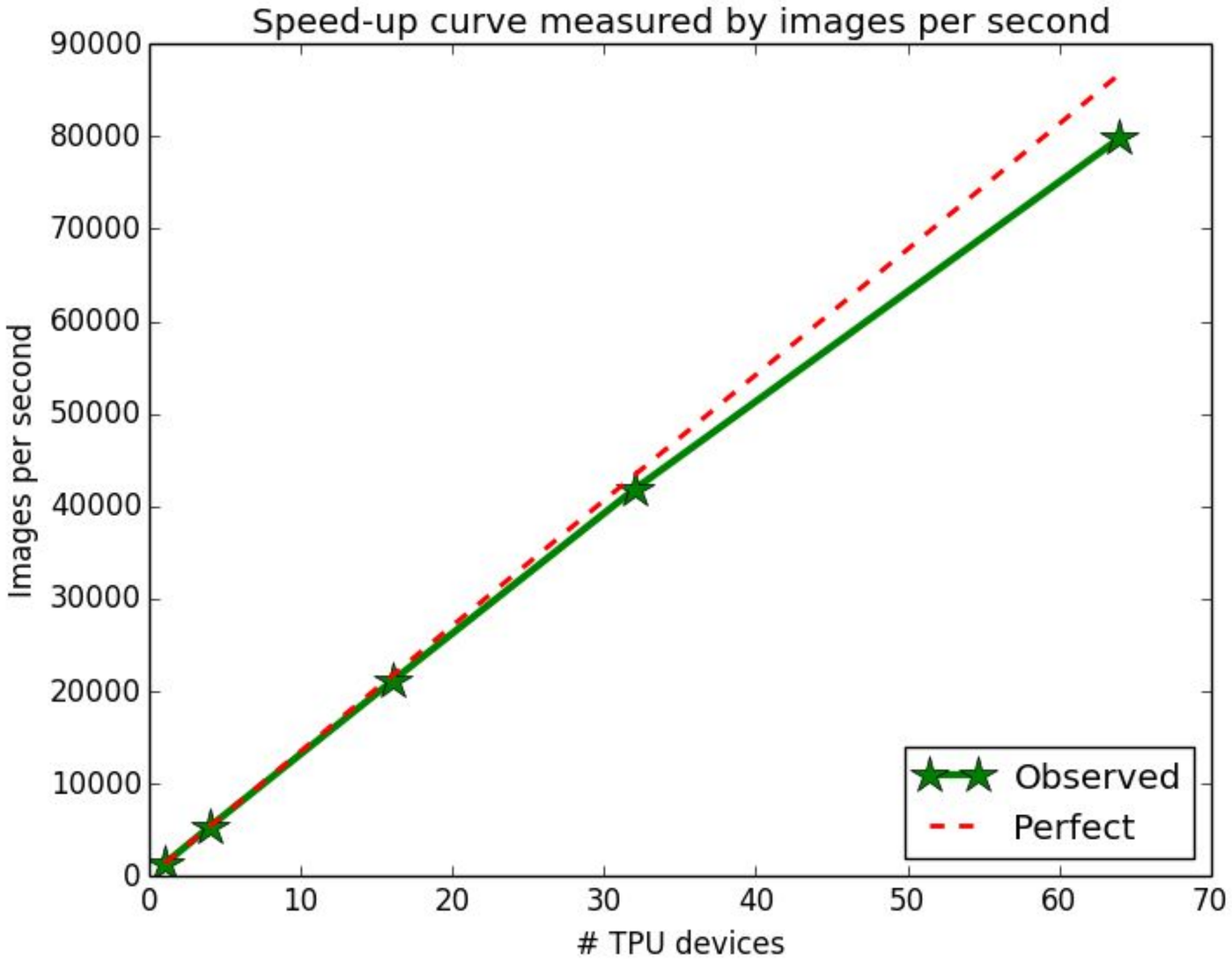
Designed for neural nets

Learning and Inference

Google TPU Cluster V3.0

100 PetaFlops
of Peak B-Flop
Performance
in 8 Racks

Performance Scales with Size of Cluster



Near Linear Scaling for ResNet-50

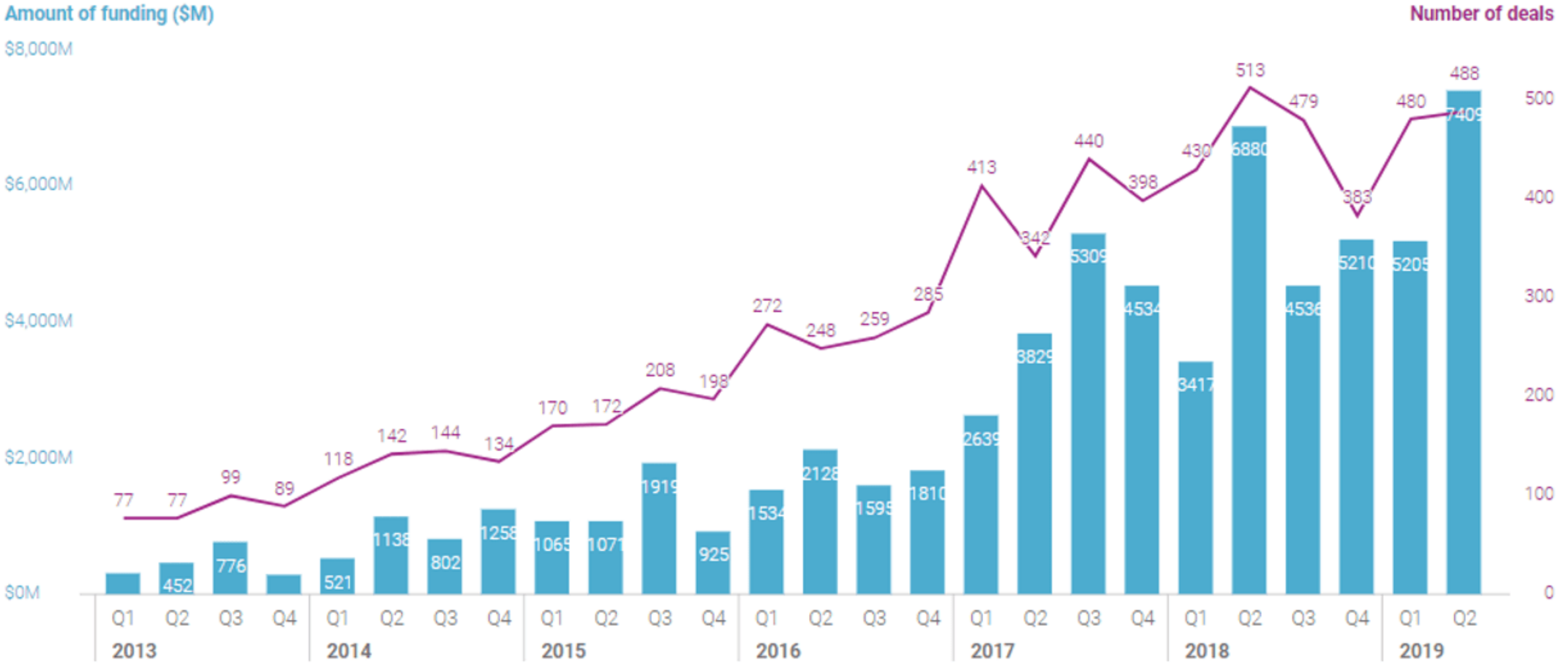
1402 Minutes on one TPUv2 device

22 Minutes on 64 TPUv2 devices

Economic and Social Implications of A.I.

Q2'19 sees record funding to AI startups at $7.4B

Q1'13 - Q2'19 (swipe right to see full data)



Source: CB Insights

100 STARTUPS USING ARTIFICIAL INTELLIGENCE TO TRANSFORM INDUSTRIES

CONVERSATIONAL AI/ BOTS

MindMeld · semanticmachines · Maluuba · Mobvoi 出门问问 · KITT.AI Natural Language Understanding Made Easy · snips · clara · AUTOMAT · X.

VISION

clarifai · Chronocam · Orbital Insight · pilot.ai · captricity · GROKSTYLE

AUTO

nuTonomy · drive.ai · AIMOTIVE · nauto · nexar · ZOOX

ROBOTICS

UBTECH Dream With Robot · anki · Rokid · dispatch making delivery smart

CYBERSECURITY

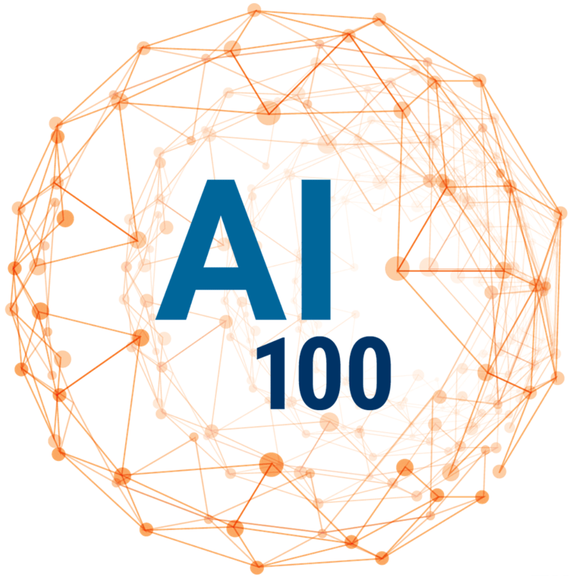
CYLANCE · sift science · sparkcognition · deepinstinct · Shift Technology · DARKTRACE

BUSINESS INTELLIGENCE & ANALYTICS

DataRobot · TRIFACTA · rapidminer · tamr · SIGOPT · Paxata · context relevant · Dataminr · CrowdFlower · logz.io

CORE AI

affectiva · PETUUM · ALGORITHMIA · H_2O.ai · loop ai Labs COGNITIVE COMPUTING · CognitiveScale THE COGNITIVE CLOUD COMPANY · sentient · Voyager Labs · Numenta · SI SCALED INFERENCE · skymind · vicarious · Digital Reasoning · bonsai · AYASDI



AD, SALES, CRM

TalkIQ · deep gram · [PERSADO] · Appier · CHORUS · INSIDE SALES .COM · drawbridge · DigitalGenius Human+AI Customer Service · RESCI

HEALTHCARE

freenome · CLOUD MEDX · zebra MEDICAL VISION · enlitic · twoXAR · iCarbonX 碳云智能 · Atomwise Better medicines faster. · deep genomics · babylon · BenevolentAI · Lunit

THE INNOVATION SUMMIT · CBINSIGHTS

TEXT ANALYSIS/ GENERATION

textio · fido.ai · cortical.io · NarrativeScience Tell the Stories Hidden in Your Data™

IOT/IIOT

nanit · KONUX · VERDIGRIS · SIGHT MACHINE Manufacturing Analytics

COMMERCE

bloomreach · mode.ai

FINTECH & INSURANCE

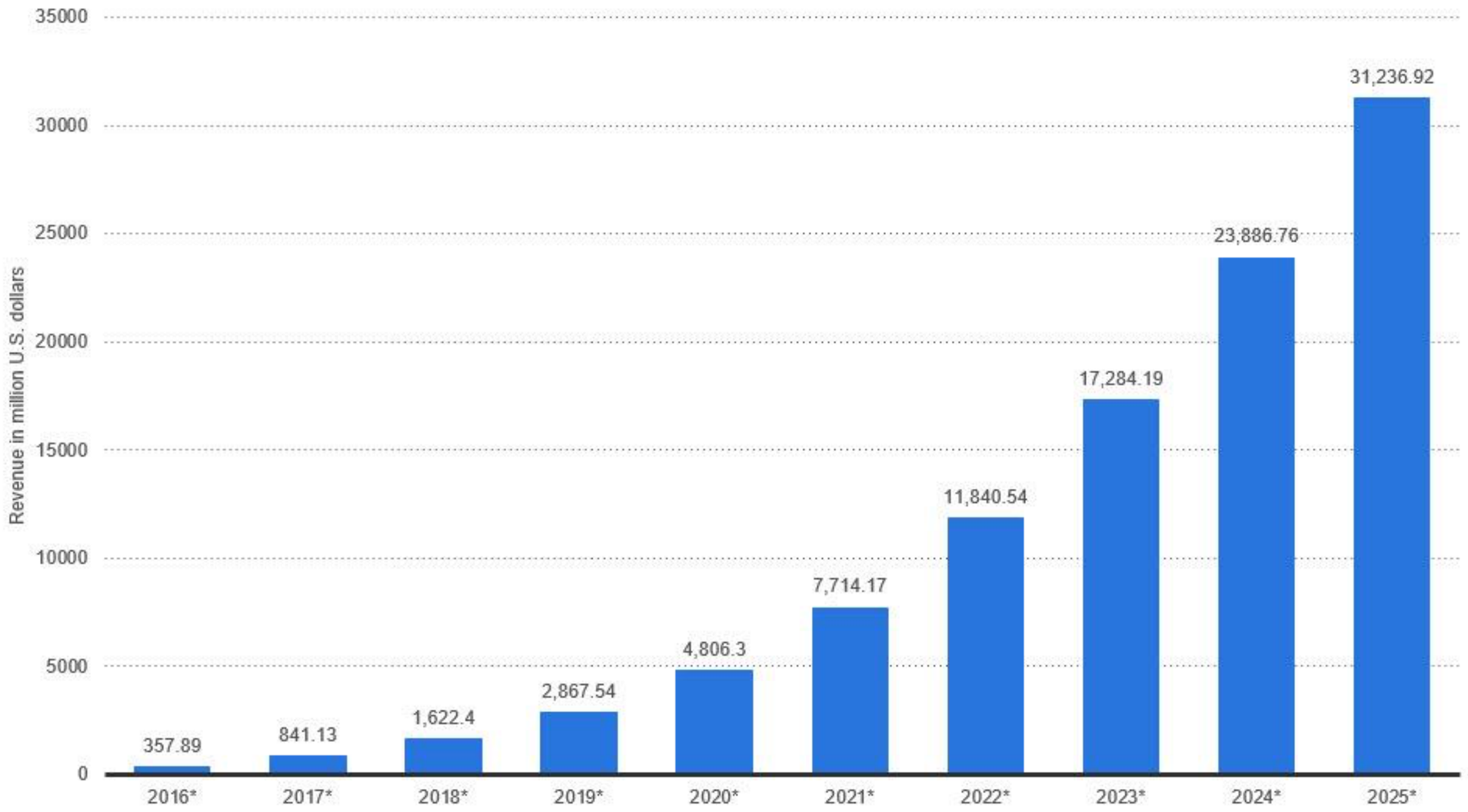
CAPE ANALYTICS · KENSHO · NUMERAI · alphasense · Kasisto

OTHER

GIGSTER · prospera · Citrine · BLUE RIVER TECHNOLOGY · ROSS · zymergen · Descartes Labs · gradescope · talla

Enterprise artificial intelligence market revenue worldwide 2016-2025

Revenues from the artificial intelligence for enterprise applications market worldwide, from 2016 to 2025 (in million U.S. dollars)



The future of work will combine

human and machine intelligence

Augmenting Human Performance with A.I.

What if you could **increase sales productivity by 20%?**

What if you could **reduce customer service costs by 50%?**

What if you could **improve drug discovery time by 10X?**

What if you could **design products not previously possible?**

Big opportunity to improve productivity and accelerate innovation in almost any kind of industry and enterprise, including government

McKinsey White paper on AI Use Cases

Findings from a recent McKinsey paper on potential A.I. benefits across a wide range of industries, representing 400 use cases:

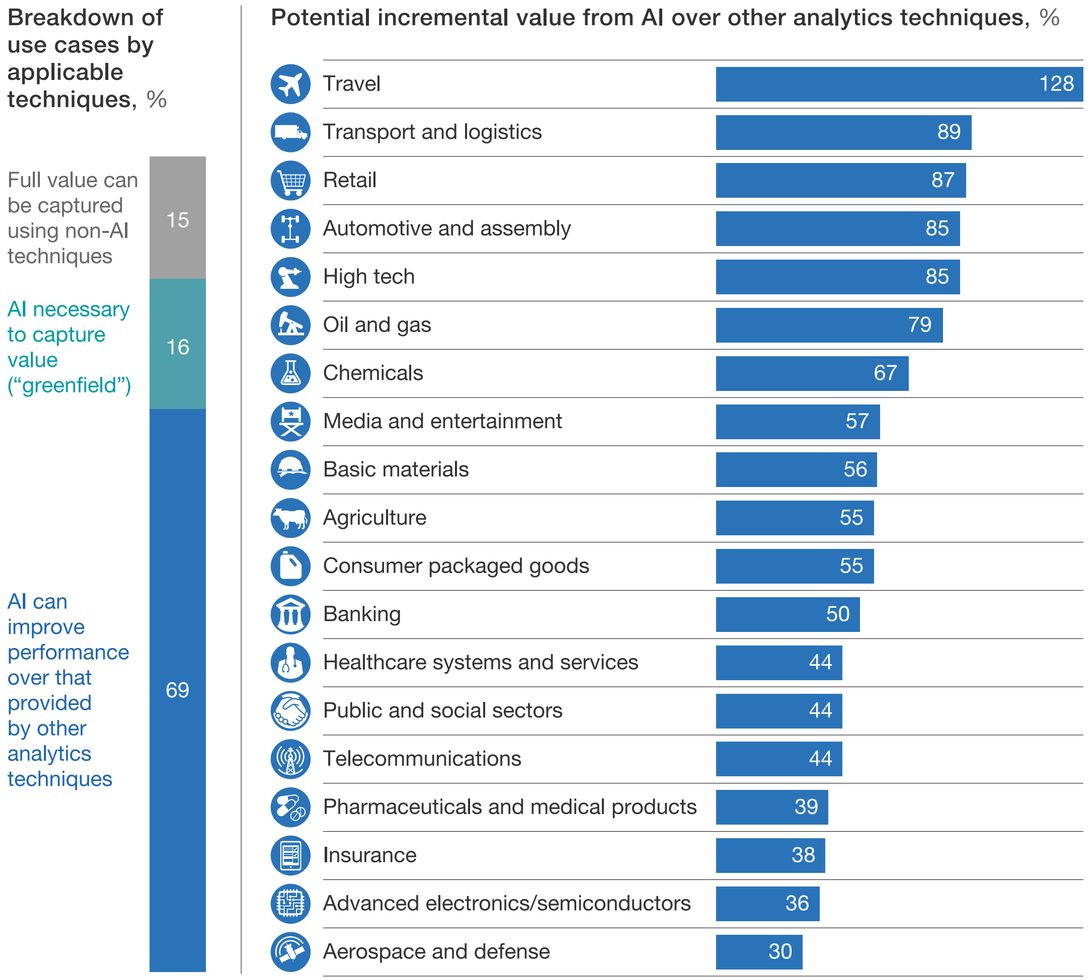
1. Immediate opportunity is to improve performance compared to **conventional analytical techniques**
2. Potential impact between **$100B to $600B per industry**
3. There are **hundreds of applications** right now where A.I. can significantly improve human performance and business results

Two-thirds of the opportunities to use AI are in improving the performance of existing analytics use cases

In 69 percent of the use cases we studied, deep neural networks can be used to improve performance beyond that provided by other analytic techniques. Cases in which only neural networks can be used, which we refer to here as “greenfield” cases, constituted just 16% of the total. For the remaining 15%, artificial neural networks provided limited additional performance over other analytics techniques, among other reasons because of data limitations that made these cases unsuitable for deep learning (Exhibit 3).

McKinsey&Co, April 2018

In more than two-thirds of our use cases, artificial intelligence (AI) can improve performance beyond that provided by other analytics techniques.

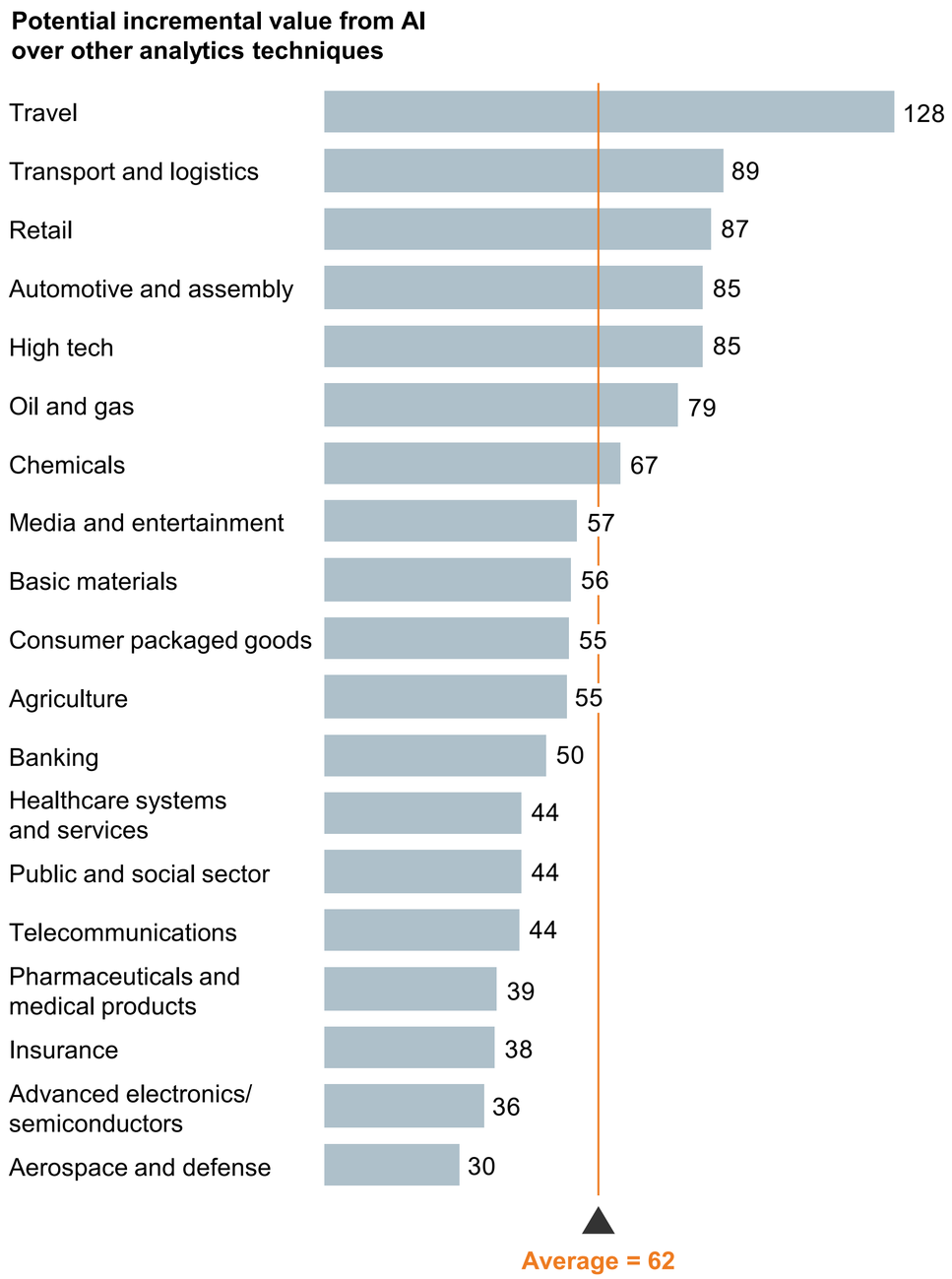


McKinsey&Company | **Source:** McKinsey Global Institute analysis

Potential Incremental Value

The potential incremental value of AI over other analytical techniques ranges from a low of 30% for defense to 128% for travel, with an average of 62%.

McKinsey&Co, April 2018

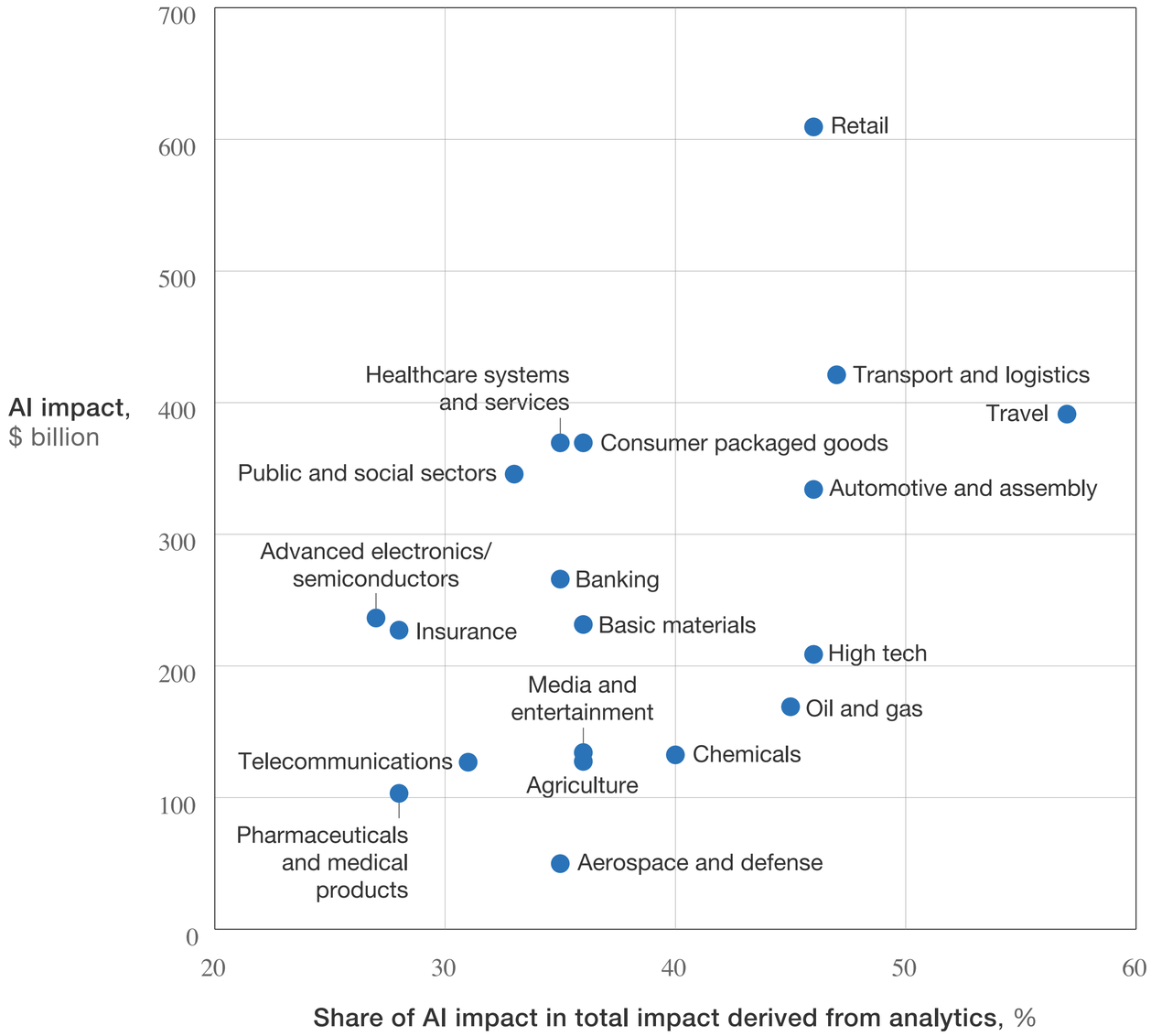


Sizing the potential value of AI

We estimate that the AI techniques we cite in this briefing together have the potential to create between **$3.5 trillion** and **$5.8 trillion** in value annually across nine business functions in 19 industries. This constitutes about **40%** of the overall $9.5 trillion to $15.4 trillion annual impact that could potentially be enabled by all analytical techniques.

McKinsey&Co, April 2018

Artificial intelligence (AI) has the potential to create value across sectors.



McKinsey&Company | **Source:** McKinsey Global Institute analysis

Will AI be Accepted by Consumers?

Consumers want to feel respected and valued
=> **A.I. Needs to make people feel this way**

Consumers hate emotion-less computer voices
=> **A.I. Needs emotionally expressive voices**

Consumers are careful who they trust
=> **A.I. Needs to earn this trust, not violate it**

In the end, consumers embrace technologies that are beneficial to them and makes their lives easier

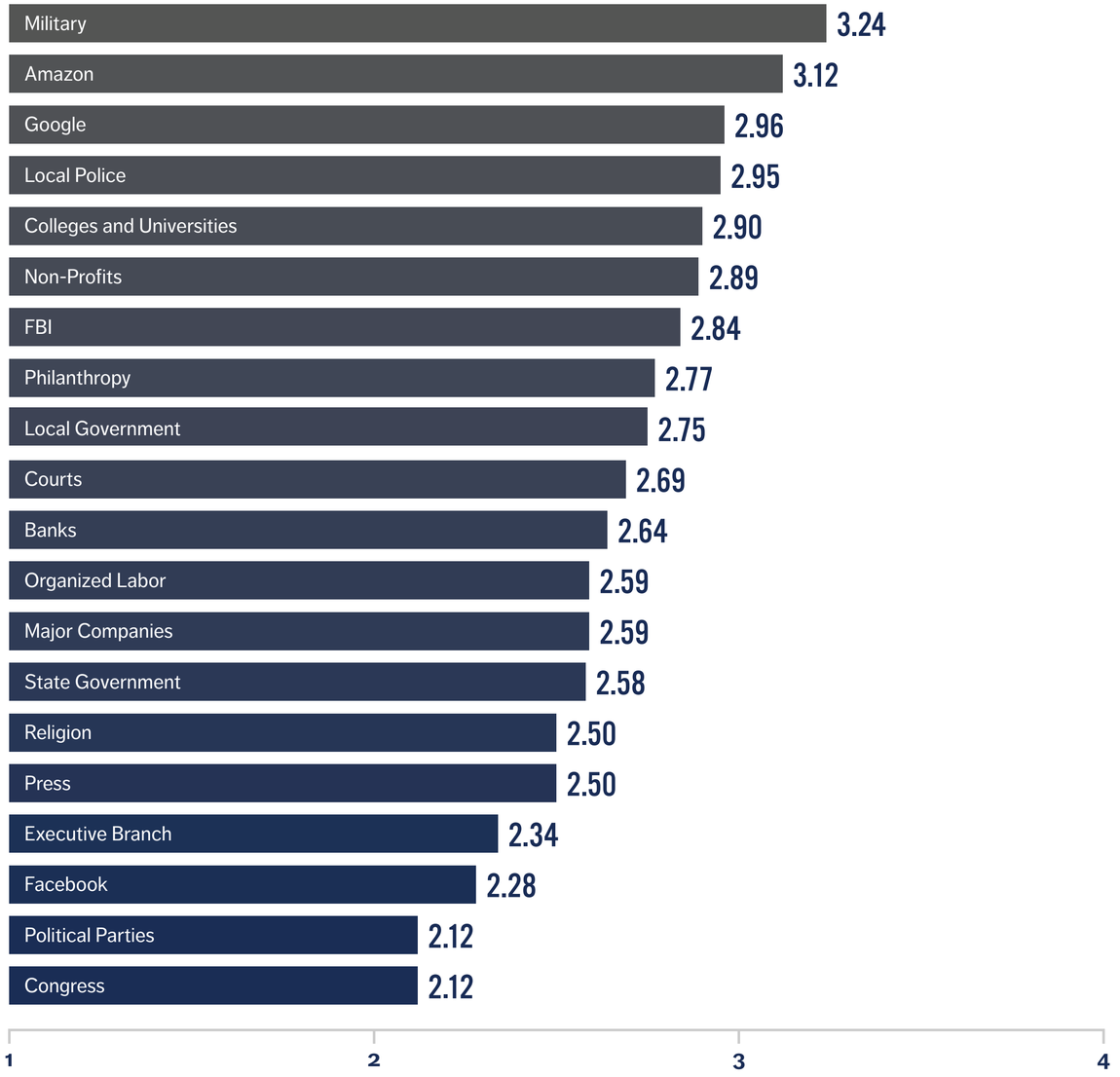
Consumer Confidence in U.S. Institutions

Knowing who to trust is everything

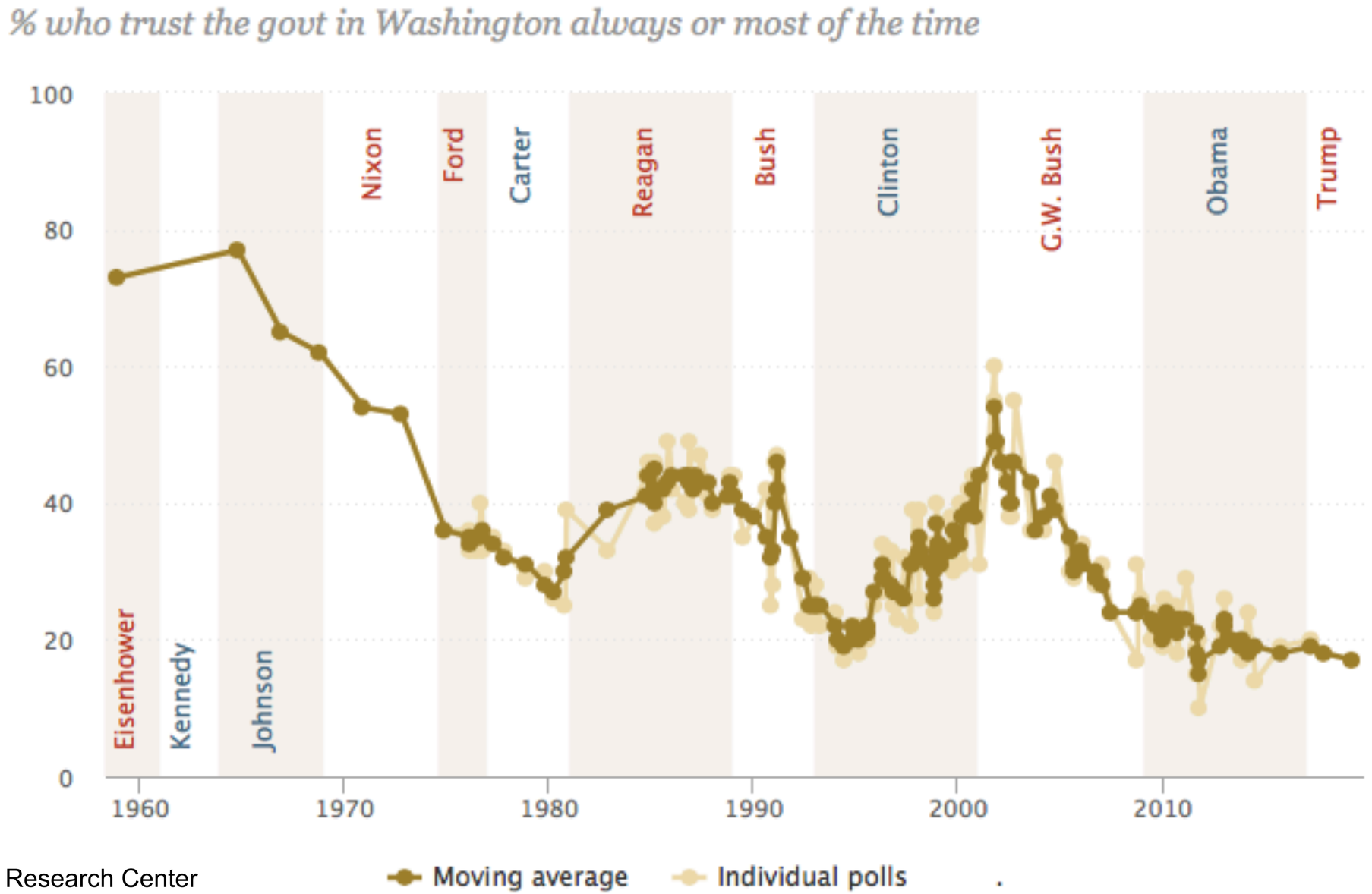
Americans have the highest confidence in the US military, followed by Amazon and Google.

The Press, Political Parties, and Congress rank considerably lower

Baker Center, Georgetown University 2018



Trust in Government at Historic Lows



Source: Pew Research Center

Summary

A.I. offers a large opportunity to **improve productivity** and **accelerate innovation** across many industries

Companies and countries that adopt **A.I. more quickly** have a **significant economic advantage over others**

Consumers will embrace A.I. as long if it **benefits them, makes them feel respected,** and **they can trust it**