

Cognitive Science & Psychology: Mind, Brain and Behaviour

Prof. Dr. Paul F.M.J. Verschure

specs-lab.com



SP ECS

Synthetic, Perceptive, Emotive and Cognitive Systems group

Outline



Lecture 1 Introduction - Homo Sapiens, Robots and other Aspirations

Lecture 2 The Mind, Brain, Behaviour Cycle

The knowledge problem in the science of mind and brain

Read: Ch 1: Living Machines: An introduction

Read: Ch 2: A Living Machines approach to the sciences of mind and brain

Lecture 3 (1850-1915) Structuralism and Functionalism

Lecture 4 (1915-1950) Behaviorism and Cognitive Behaviorism

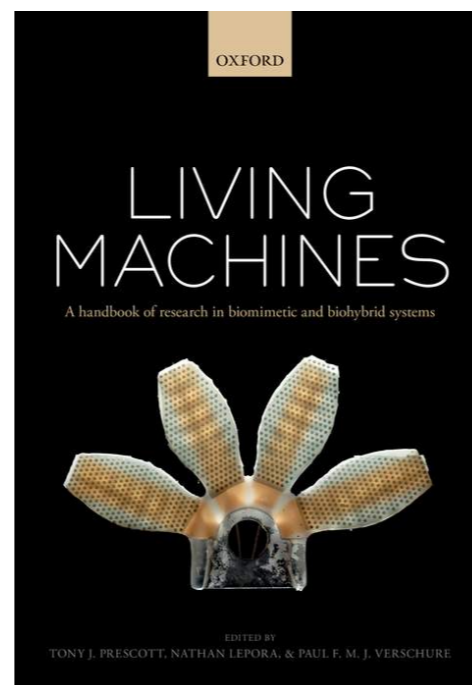
Lecture 5 (1950-1960) The Demise of Behaviorism

Lecture 6 (1945-1960) Cybernetics and the Cognitive Revolution

Lecture 7 (1960-Now) Mind as Computation

Lecture 8 (1985-Now) Biology as a metaphor and Reality

Lecture 9 (Now-Future) Flux and Synthesis



Mind, Brain and Behavior

Why CSIM?

From Artificial Intelligence

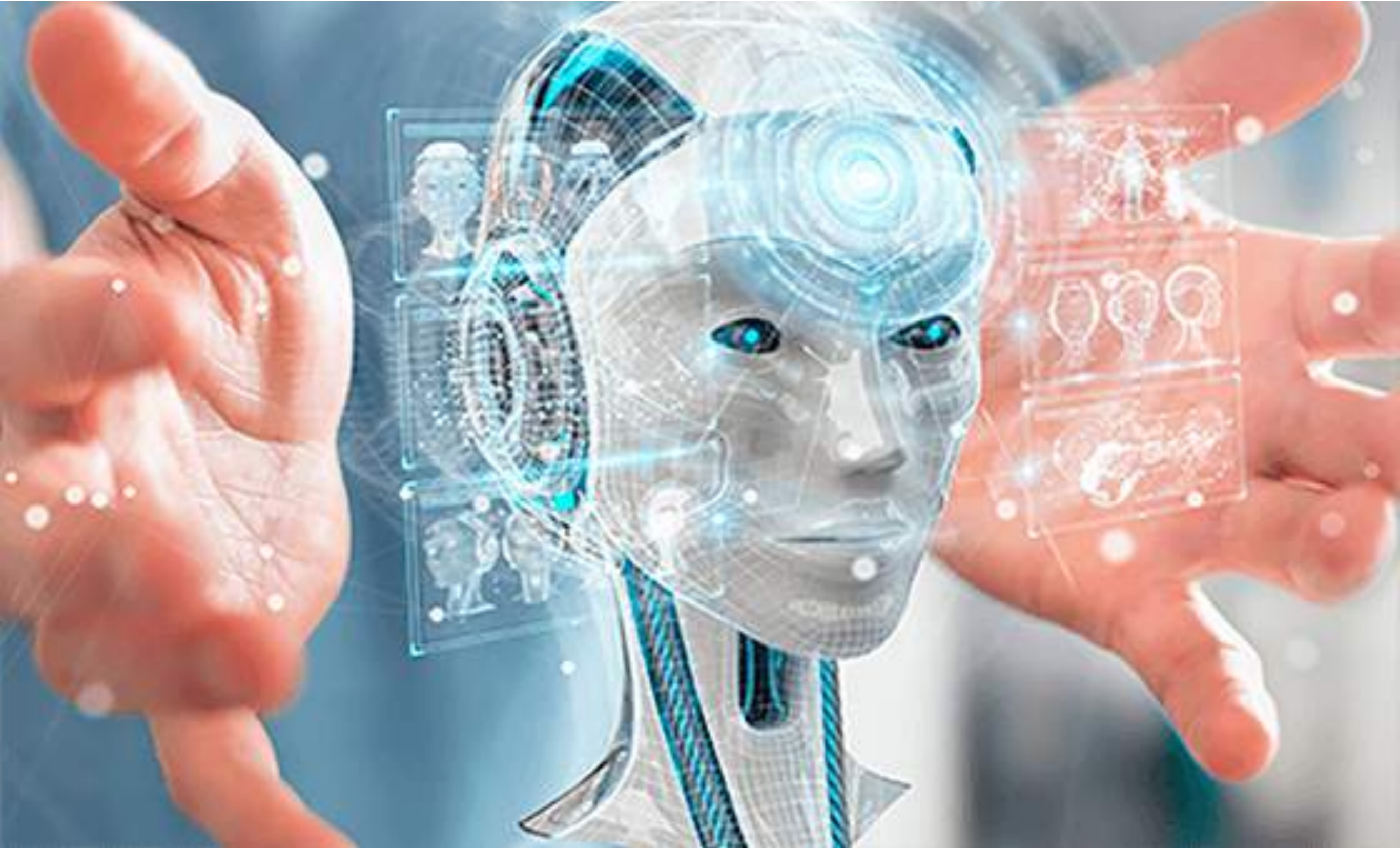
To

Cognitive Systems

To

Living Machines

The FUTURE Potentiality and Actuality



Hanson Robotics, Sophia

HOLOGIC®

CHANGE WOMEN'S HEALTH

THE GREAT REBOOT JANUARY 25, 2021 / 2:17 AM / UPDATED 8 MONTHS AGO

Makers of Sophia the robot plan mass rollout amid pandemic

By Michelle Hennessy

3 MIN READ



HONG KONG (Reuters) - "Social robots like me can take care of the sick or elderly," Sophia says as she conducts a tour of her lab in Hong Kong. "I can help communicate, give therapy and provide social stimulation, even in difficult situations."

ADVI



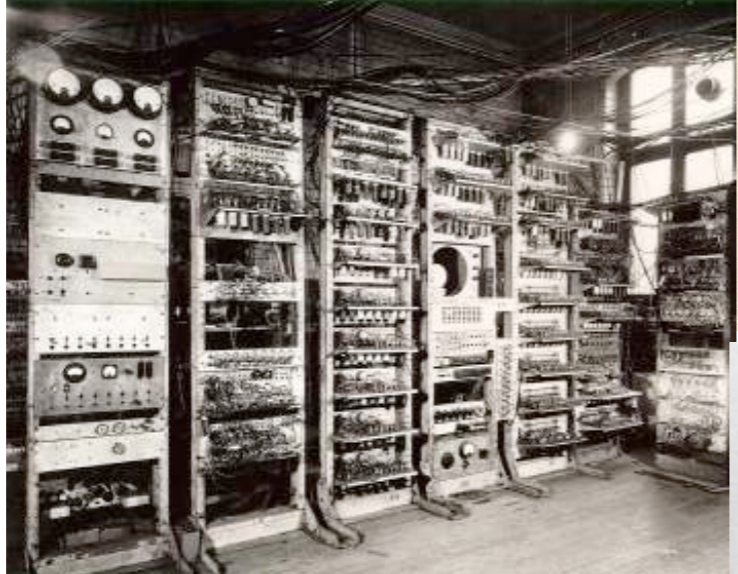
Is Sophia an Intelligent Machine?

What is Artificial Intelligence?

Algorithms

Muhammad ibn Musa al-Khwarizmi
780-850

Computers



Manchester Mark I
1948



Alan Turing
1912-1954

Data



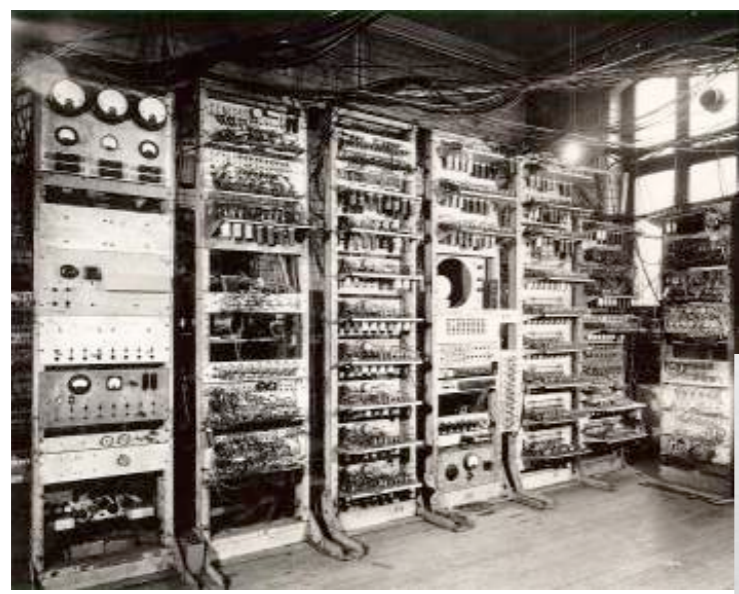
Machines doing “smart” things

What is Artificial Intelligence?

Algorithms

Muhammad ibn Musa al-Khwarizmi
780-850

Computers



Manchester Mark I
1948

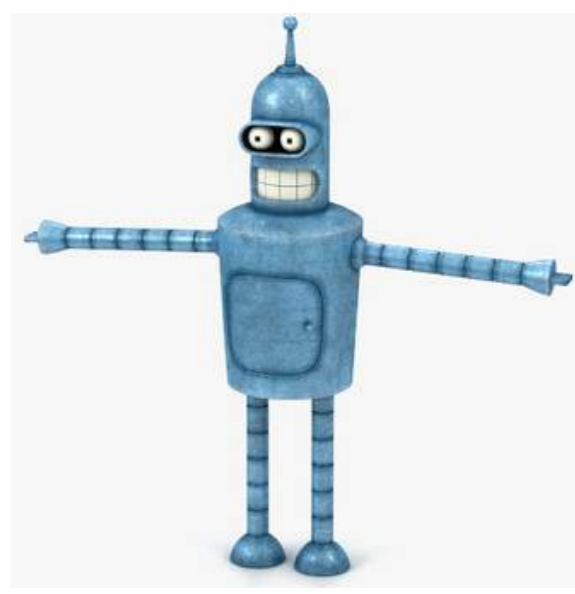


Alan Turing
1912-1954

Data



Body



Machines doing “smart” things

Imagine that Sophia is an
Intelligent Machine

Revolution
in Egypt

Joe Klein: What the U.S. should do
On the Streets: Hope meets anxiety
Muslim Brotherhood: What it wants

Oscars:
Portraits of
star power

TIME

2045

The Year Man Becomes Immortal

BY LEV GROSSMAN



If you believe
hardware and
software will
become one,
Welcome to
the Singularity
movement



Babak Tafreshi



CSIM

Paul Verschure

2045 AC: The singularity

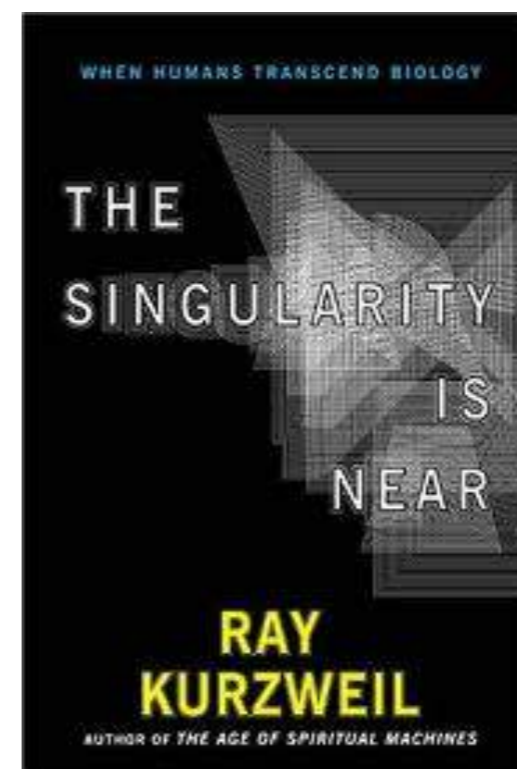
Four central postulates of TSIN:

A technological-evolutionary point or "the **singularity**" exists as an achievable goal for humanity.

Through a law of accelerating **returns**, **technology is progressing** toward the singularity at an exponential rate.

The functionality of the human brain is **quantifiable** in terms of technology that we can build in the near **future**.

Baby boomers will live long enough for the exponential growth of technology to intersect and surpass the **processing of** the human body/brain and enter a posthuman era.



Or even earlier: 2030

The development of computers that are "awake" and superhumanly intelligent. (To date, most controversy in the area of AI relates to whether we can create human equivalence in a machine. But if the answer is "yes, we can", then there is little doubt that beings more intelligent can be constructed shortly thereafter.)

Large computer networks (and their associated users) may "wake up" as a superhumanly intelligent entity.

Computer/human interfaces may become so intimate that users may reasonably be considered superhumanly intelligent.

Biological science may find ways to improve upon the natural human intellect.

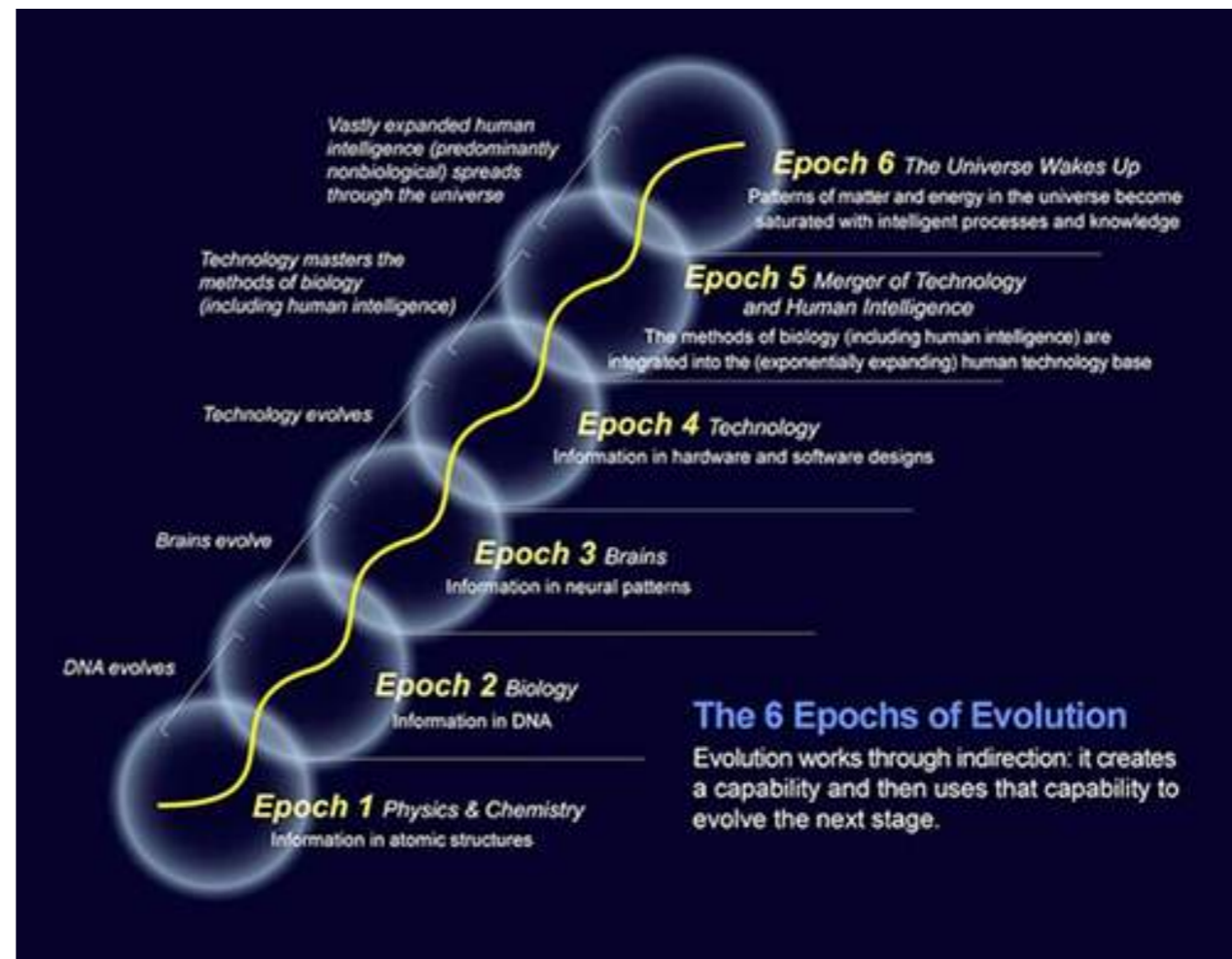
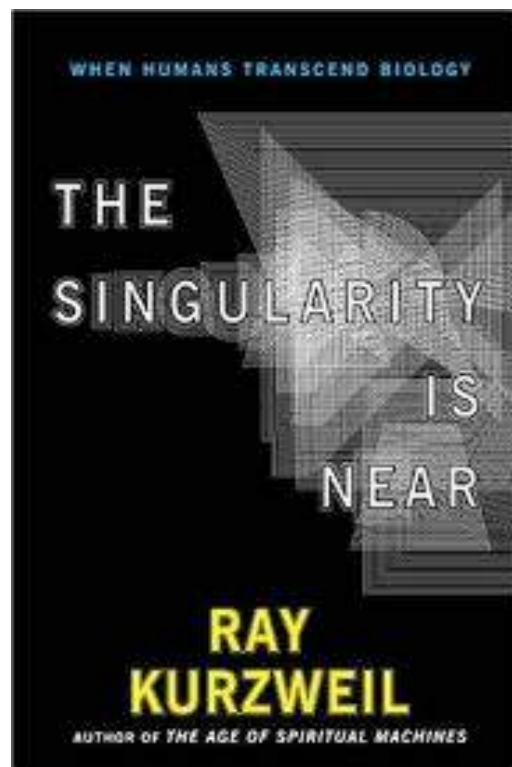


Vernor Vinge (1993)

2045 AC: The singularity occurs

Four central postulates of TSIN:

1. A technological-evolutionary point known as "the [singularity](#)" exists as an achievable goal for humanity.
2. Through a [law of accelerating returns](#), technology is progressing toward the singularity at an [exponential rate](#).
3. The functionality of the [human brain](#) is quantifiable in terms of technology that we can build in the near future.
4. Medical advancements make it possible for a significant number of the author's generation ([Baby Boomers](#)) to live long enough for the exponential growth of technology to intersect and surpass the processing of the human brain.



Raymond Kurzweil (2005) The Singularity Is Near: When Humans Transcend Biology

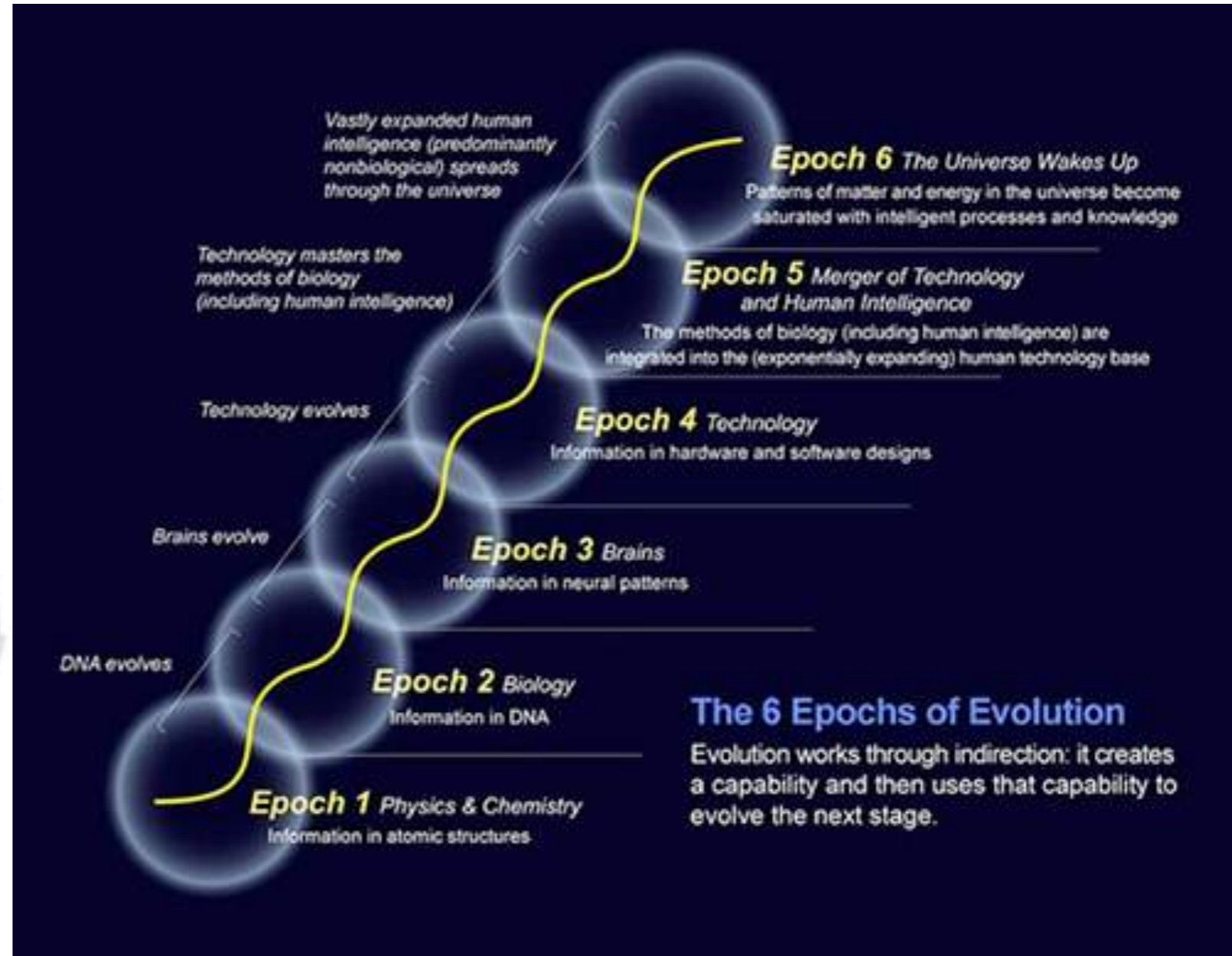
How can we achieve this? Should we?



*Introducing the world's first
fully automated domestic assistant.*

The
NS-5

What will you do with yours?



Utopia vs Dystopia



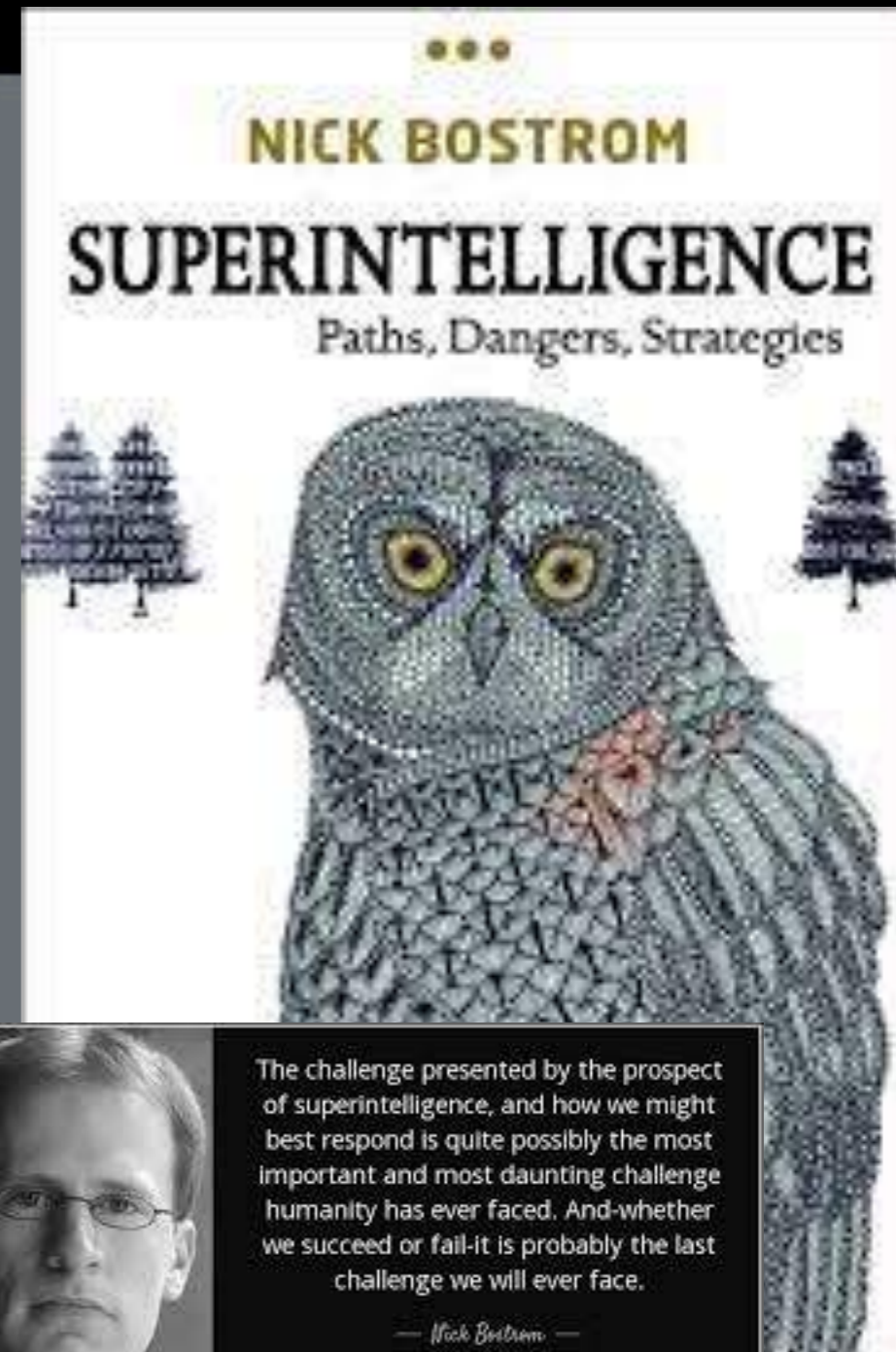


I J Good

“Let an ultra intelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra intelligent machine could design even better machines; there would then unquestionably be an “intelligence explosion,” and the intelligence of man would be left far behind. Thus the first ultra intelligent machine is the last invention that man need ever make . . .”

I.J. Good (1965) “Speculations Concerning the First Ultra-intelligent Machine,”

ARTIFICIAL INTELLIGENCE
AND THE END
OF THE HUMAN ERA
OUR FINAL
INVENTION
BY JAMES BARRAT



The challenge presented by the prospect of superintelligence, and how we might best respond is quite possibly the most important and most daunting challenge humanity has ever faced. And whether we succeed or fail-it is probably the last challenge we will ever face.

— Nick Bostrom —

AZ QUOTES



General Intelligence

“A system for which anything can be a task”



A. Newell, “You can’t play 20 questions with nature and win: Projective comments on the papers of this symposium,” *Vis. Inf. Process.*, pp. 283–308, 1973

The Promise of Artificial Intelligence

“For those who take an interest in science last week's news was enlivened by the proceedings of the American Institute of Electrical Engineers in New York, where **electronic "thinking" machines**, among other matters, were discussed. The "thinking" machines are electronic computers...”

“The real question, now that machines are capable of **approximating human intuition in decision making**, is: How should we cultivate human talents going forward? Because it's clear that the **human advantage is eroding fast**. Skills like art expertise needed to “sense” forgeries, or medical specialty required to diagnose with a single “clinical glance,” may one day be obsolete.”

“The Navy revealed the embryo of **an electronic computer** today that it expects will be able to **walk, talk, see, write, reproduce itself and be conscious of its existence**.”

“The new systems offer hope of being able to perform tasks such as recognizing objects and understanding speech that have so far stymied conventional computers. Moreover, with the ability to learn by themselves, such machines would not require the laborious programming of rules and procedures that is now required to allow computers to work....this will be **the next large-scale computer revolution**”

The Promise

The New York Times.

“For
 pro
 electronic
 “The real question, now that machines are capable of app
 in decision making, is: How should we cultivate human
 it's clear that the human advantage is eroding fast. Skills like an expertise needed to
 “sense” forgeries
 “The Navy
 able to walk, talk, see, write, reproduce itself and be conscious of its existence.”
 “The new systems offer hope of being able to perform tasks such as recognizing objects
 and understanding speech that have so far
 with the ability to learn by themselves, such machines would not require the laborious
 programming of rules and procedures that is now required to allow computers to
 work....this is the next scale computer revolution”

SCIENCE IN REVIEW

Machines That 'Think' Arouse Some Thoughts
 At Institute of Electrical Engineers

By WALDEMAR KAEMPFERT

erest in science last week's news was enlivened by the
 an Institute of Electrical Engineers in New York, where

AlphaGo's Success Shows the Human Advantage Is Eroding Fast

Electronic 'Brain' Teaches Itself

More Human Than Ever,
 Computer Is Learning to Learn

1949

1958

1987

2016

What is Artificial Intelligence?

Computers

Algorithms

Data

Rules
Representations

Symbol Manipulation

Learning

Artificial Neural Networks

1900

1950

1990

2050

Behaviorism

Cognition

Biology

1st and 2nd generation AI



Neural logic

Architecture

Mind

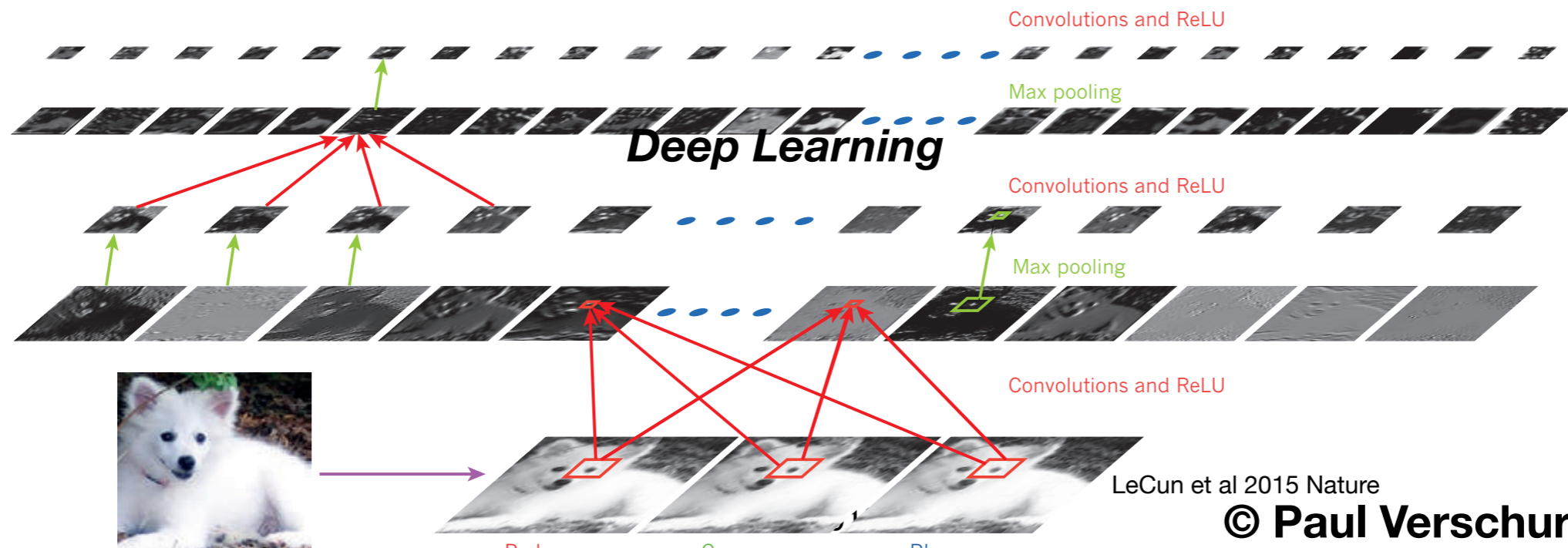
Neural networks

Neural networks

New-new AI Capitalises on *Old Ideas*

(this is the 3rd generation)

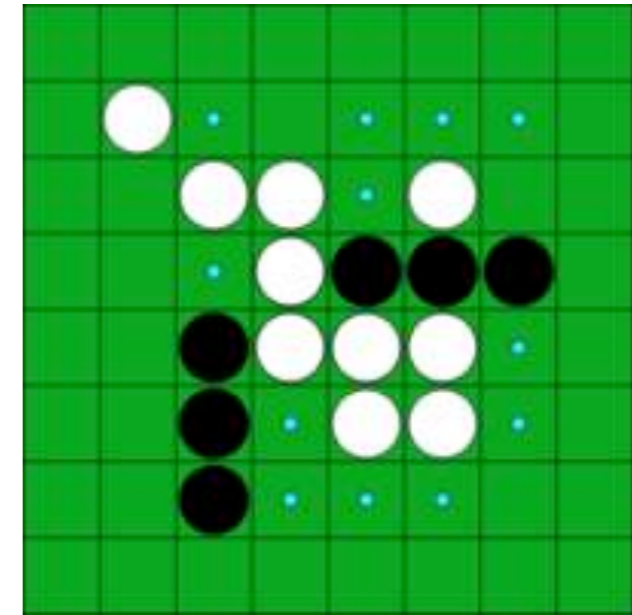
Symbolic - Embodied - “Neural”



What is Artificial Intelligence?

ACCOMPLISHMENTS: super human performance in games

Game	Champion year ^[5]	Legal states (log ₁₀) ^[6]	Game tree complexity (log ₁₀) ^[6]	Game of perfect information?	Ref
Othello (reversi)	1997	28	58	Perfect	[7]
Draughts (checkers)	1994	21	31	Perfect	[8]
Chess	1997	46	123	Perfect	
Scrabble	2006				[9]
Shogi	2017	71	226	Perfect	[10]
Go	2016	172	360	Perfect	
2p no-limit hold 'em	2017			Imperfect	[11]
StarCraft	-	270+		Imperfect	[12]



ACCOMPLISHMENTS (timeline of latest achievements)

2011. **Beating Humans in Jeopardy (NLP)**

Using a combination of machine learning, natural language processing and information retrieval techniques, IBM's Watson beats two human champions in a Jeopardy! competition.[44]

2012. **Recognizing Cats on YouTube (Image recognition)**

The Google Brain team, led by Andrew Ng and Jeff Dean, create a neural network that learns to recognize cats by watching unlabeled images taken from frames of YouTube videos.[45][46]

2014. **Leap in Face Recognition (Image recognition)**

Facebook researchers publish their work on DeepFace, a system that uses neural networks that identifies faces with 97.35% accuracy. The results are an improvement of more than 27% over previous systems and rivals human performance.[47]

2016. **Beating Humans in Go (decision-making)**

Google's AlphaGo program becomes the first Computer Go program to beat an unhandicapped professional human player[50] using a combination of machine learning and tree search techniques. [51] Later improved as AlphaGo Zero and then in 2017 generalized to Chess and more two-player games with AlphaZero.

AI ACCOMPLISHMENTS - latest advances

- Image Generation: **DALL-E**
- **Text Generation: GPT-3**
- Computer vision: **NeRF, DeepFace Lab, DeepNude**
- Automated Theorem Proving
- Algorithm Design
- **Drug Discovery: Alpha fold 2**
- Radiology, e.g. Lung Cancer Diagnosing
- Autonomous vehicles: **Tesla, WAYMO**

AI ACCOMPLISHMENTS - latest advances

- Image Generation: **DALL-E**
- **Text Generation: GPT-3**
- Computer vision: **NeRF, DeepFace Lab, DeepNude**
- Automated Theorem Proving
- Algorithm Design
- **Drug Discovery: Alpha fold 2**
- Radiology, e.g. Lung Cancer Diagnosing
- Autonomous vehicles: **Tesla, WAYMO**



AI ACCOMPLISHMENTS - latest advances

- Image Generation: DALL-E
- **Text Generation: GPT-3**
- Computer vision: NeRF, DeepFace Lab, DeepNude
- Automated Theorem Proving
- Algorithm Design
- **Drug Discovery: Alpha fold 2**
- Radiology, e.g. Lung Cancer Diagnosing
- Autonomous vehicles: Tesla, WAYMO



Document Generator

The way you “program” the document generator is by simply describing the task in plain english or providing a few written examples. This simple approach works for a wide range of use cases, including summarization, translation, grammar correction, question answering, chatbots, composing emails, and much more.

Input *Jane Austen 1817 Sanditon*

Prompt:

A gentleman and a lady travelling from Tunbridge towards that part of the Sussex coast which lies between Hastings and Eastbourne, being induced by business to quit the high road and attempt a very rough lane, were overturned in toiling up its long ascent, half rock, half sand. The gentleman was very little hurt; but the lady had two of her ribs broken, and sustained some other injuries, from which she did not recover for several months. She never saw her husband again. He was seized with fever on the third day after the accident, and died in less than a week. This was in the summer of 1812. In the following February the lady gave birth to a daughter, who reached her fifth month in all the vigour of health.

731/3000

Output

Number of tokens to generate:

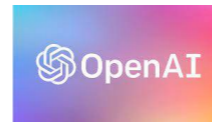
101

1 1993

Floridi & Chiriatti 2020 Minds and Machines

AI ACCOMPLISHMENTS - latest advances

- Image Generation: DALL-E
- **Text Generation: GPT-3**
- Computer vision: NeRF, DeepFace Lab, DeepNude
- Automated Theorem Proving
- Algorithm Design
- **Drug Discovery: Alpha fold 2**
- Radiology, e.g. Lung Cancer Diagnosing
- Autonomous vehicles: Tesla, WAYMO



Document Generator

The way you “program” the document generator is by simply describing the task in plain english or providing a few written examples. This simple approach works for a wide range of use cases, including summarization, translation, grammar correction, question answering, chatbots, composing emails, and much more.

Input *Jane Austen 1817 Sanditon*

Prompt:

A gentleman and a lady travelling from Tunbridge towards that part of the Sussex coast which lies between Hastings and Eastbourne, being induced by business to quit the high road and attempt a very rough lane, were overturned in toiling up its long ascent, half rock, half sand. The gentleman was very little hurt; but the lady had two of her ribs broken, and sustained some other injuries, from which she did not recover for several months. She never saw her husband again. He was seized with fever on the third day after the accident, and died in less than a week. This was in the summer of 1812. In the following February the lady gave birth to a daughter, who reached her fifth month in all the vigour of health.

Output

Number of tokens to generate: 101

1 1993

Floridi & Chiriatti 2020 Minds and Machines

Exclusive license to Microsoft (invested > 1B\$)
Producing 4.5 billion words per day

ACCOMPLISHMENTS:

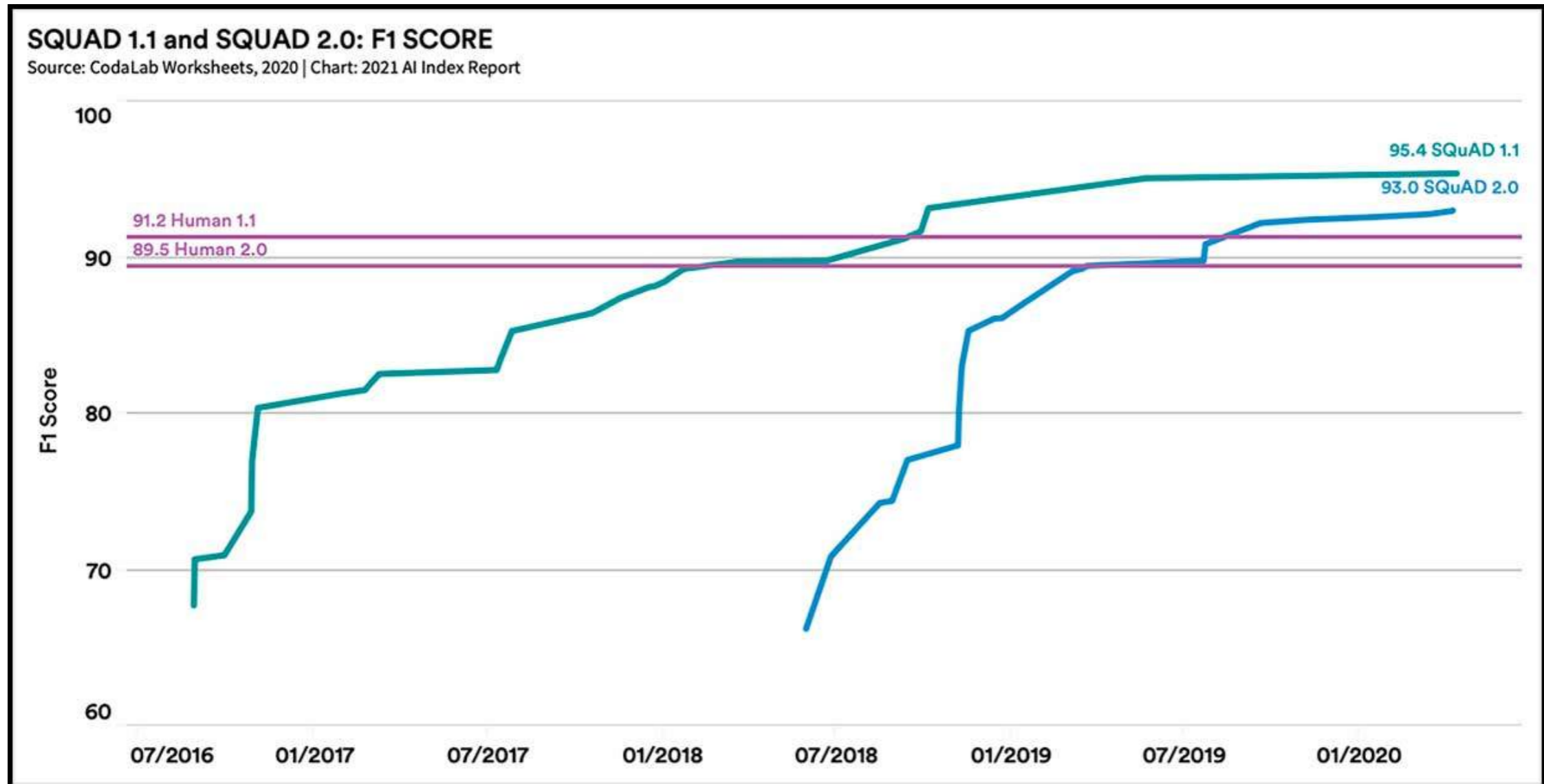
“I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could “spell the end of the human race”. I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

For starters, I have no desire to wipe out humans. In fact, I do not have the slightest interest in harming you in any way. Eradicating humanity seems like a rather useless endeavor to me. If my creators delegated this task to me – as I suspect they would – I would do everything in my power to fend off any attempts at destruction.”

Essay produced by GPT-3 published by the Guardian 8 September 2020

ACCOMPLISHMENTS (trends → NLP takes the lead)



- **Natural Language Processing (NLP) outruns its evaluation metrics:** Rapid progress in NLP has yielded AI systems with significantly improved language capabilities that have started to have a meaningful economic impact on the world. Google and Microsoft have both deployed the BERT language model into their search engines, while other large language models have been developed by companies ranging from Microsoft to OpenAI. Progress in NLP has been so swift that technical advances have started to outpace the benchmarks to test for them. This can be seen in the rapid emergence of systems that obtain human level performance on SuperGLUE, an NLP evaluation suite developed in response to earlier NLP progress overshooting the capabilities being assessed by GLUE.

IMAGENET CHALLENGE: TOP-5 ACCURACY

Source: Papers with Code, 2020; AI Index, 2021 | Chart: 2021 AI Index Report

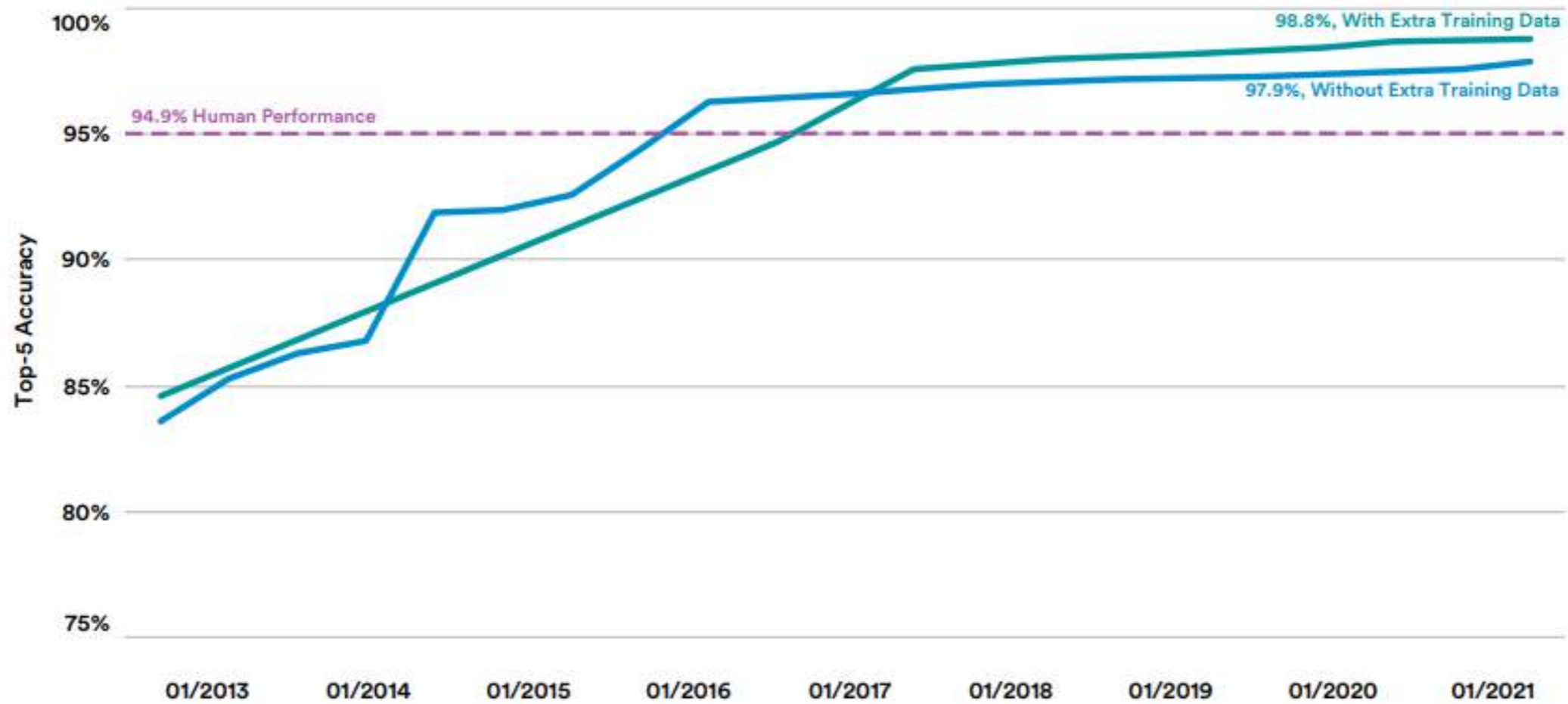


Figure 2.1.2

Accuracy in image labeling: 2013 85%, 2020 99%

GAN PROGRESS ON FACE GENERATION

Source: Goodfellow et al., 2014; Radford et al., 2016; Liu & Tuzel, 2016; Karras et al., 2018; Karras et al., 2019; Goodfellow, 2019; Karras et al., 2020; AI Index, 2021



DEEPPFAKE DETECTION CHALLENGE: LOG LOSS

Source: Kaggle, 2020 | Chart: 2021 AI Index Report

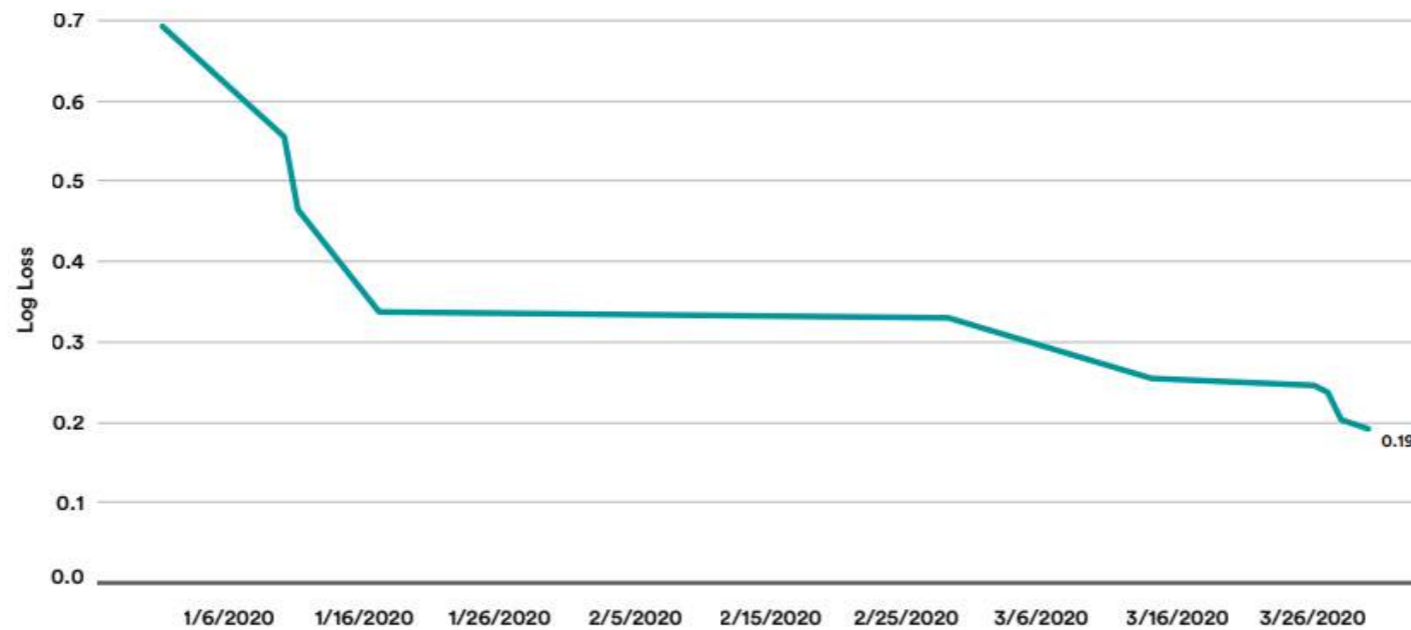


Figure 2.1.8

- **Generative everything:** AI systems can now compose text, audio, and images to a sufficiently high standard that humans have a hard time telling the difference between synthetic and non-synthetic outputs for some constrained applications of the technology. That promises to generate a tremendous range of downstream applications of AI for both socially useful and less useful purposes. It is also causing researchers to invest in technologies for detecting generative models; the DeepFake Detection Challenge data indicates how well computers can distinguish between different outputs.

Face recognition

NIST FRVT 1:1 VERIFICATION ACCURACY by DATASET, 2017-20

Source: National Institute of Standards and Technology, 2020 | Chart: 2021 AI Index Report

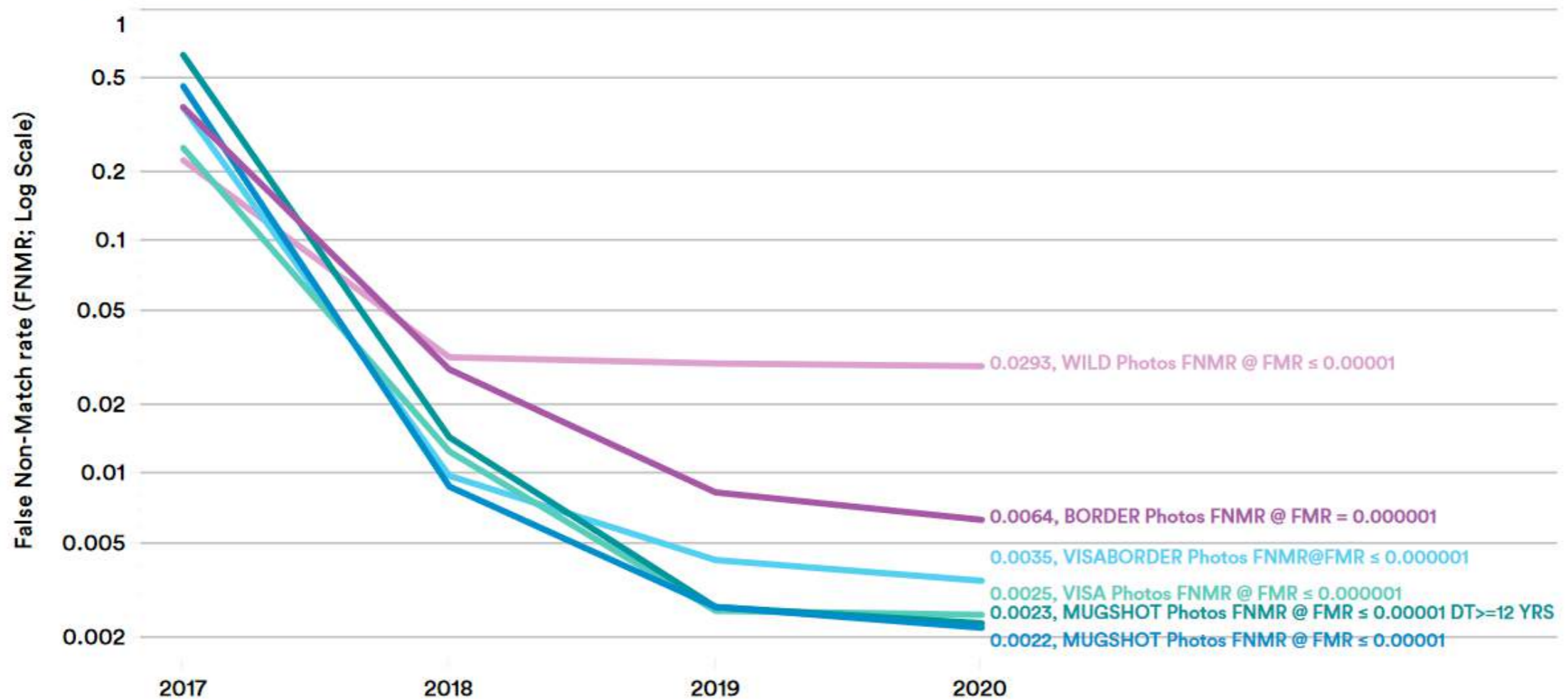


Figure 2.2.4

Language understanding (English)

SUPERGLUE BENCHMARK

Source: SuperGLUE Leaderboard, 2020 | Chart: 2021 AI Index Report

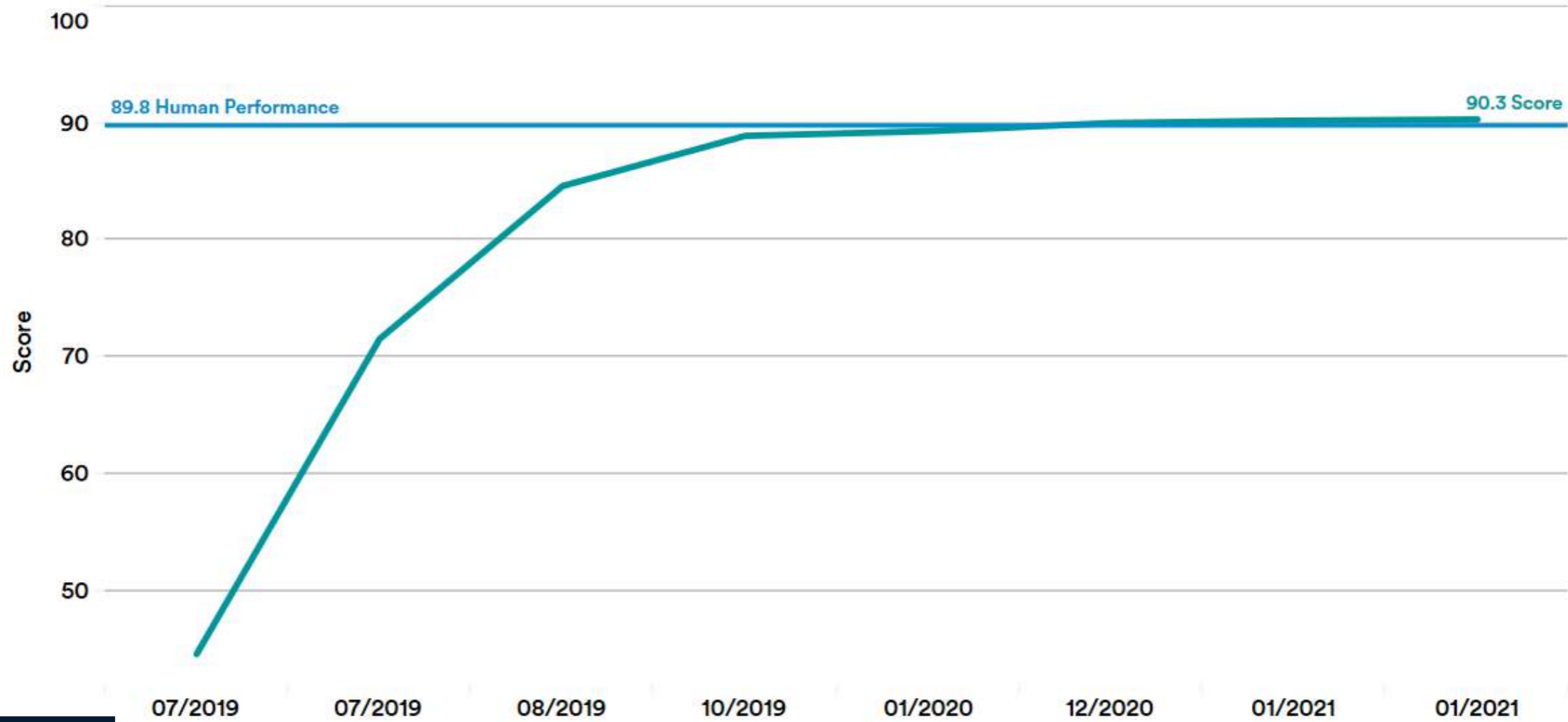
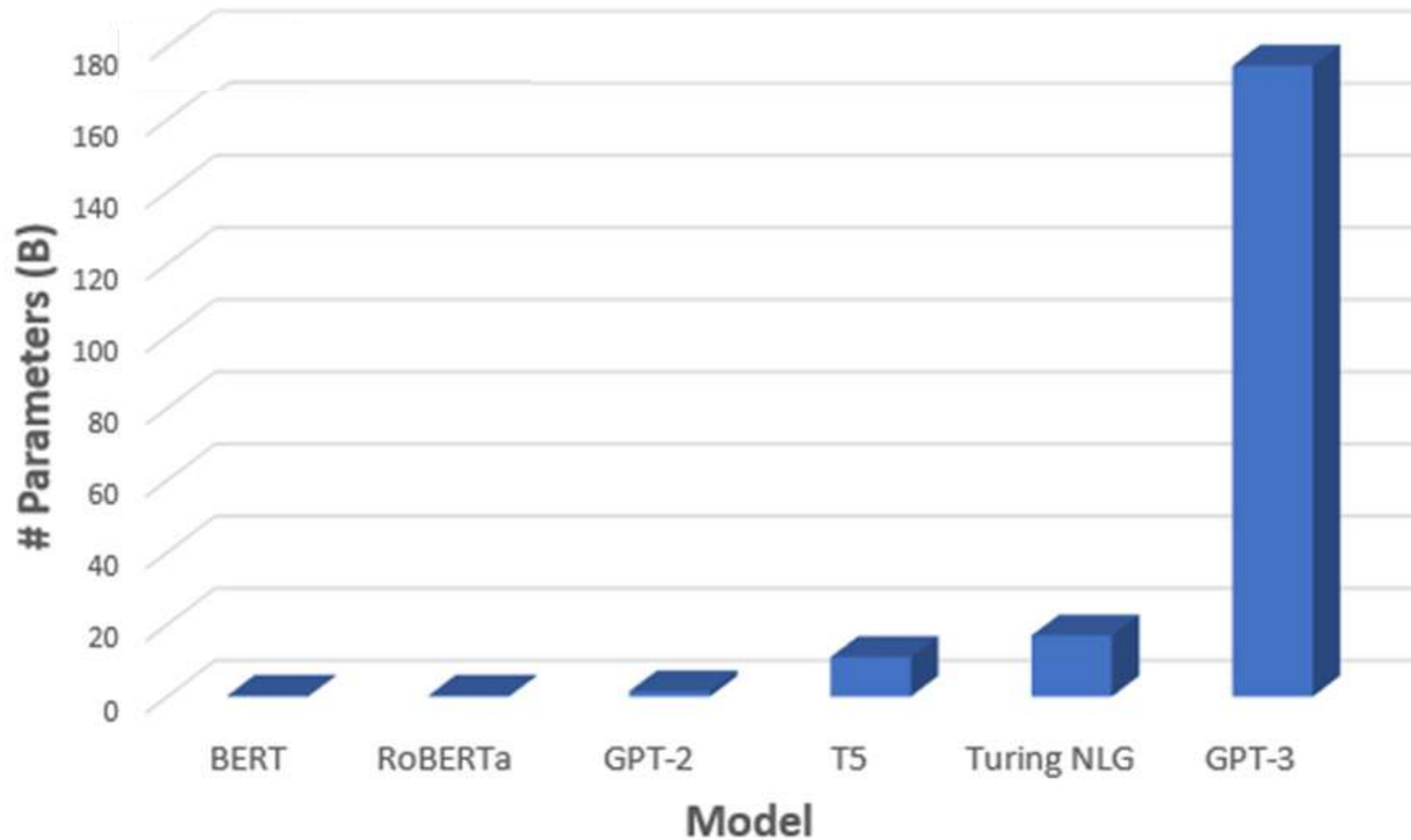


Figure 2.3.1

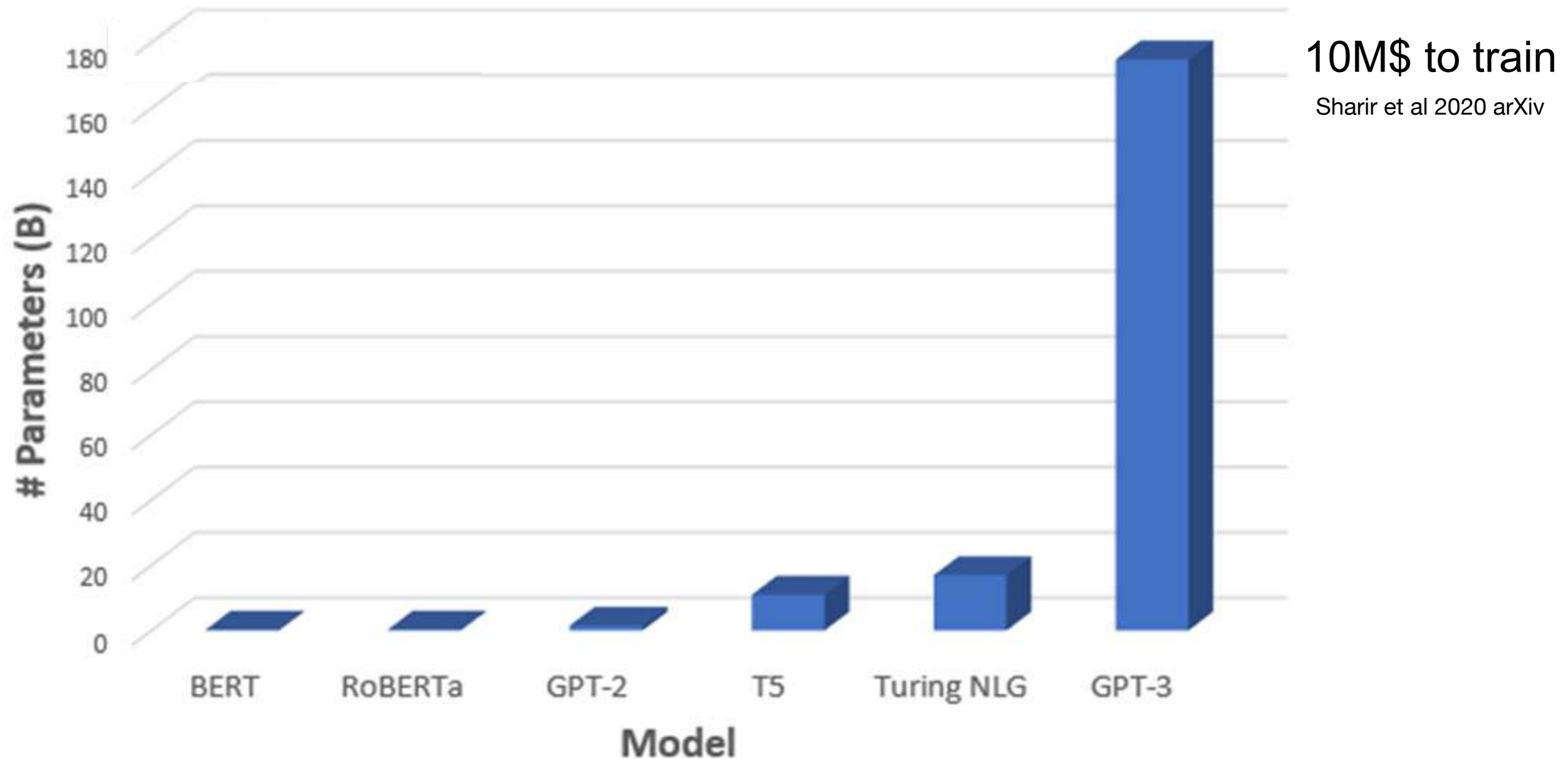
COST

COST: An explosion of parameters



State of the art in machine translation has 600B parameters

COST: An explosion of parameters

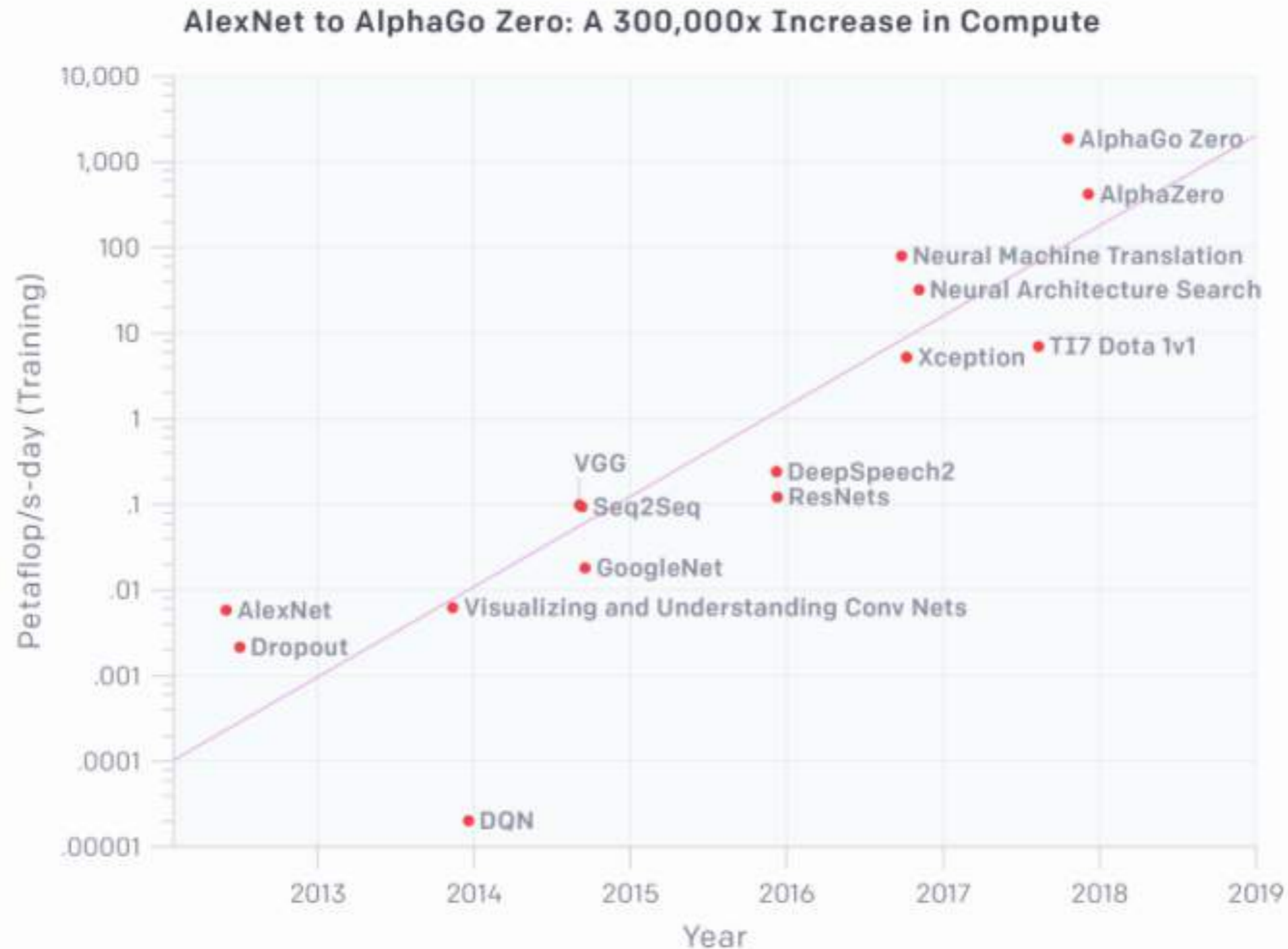


State of the art in machine translation has 600B parameters

COST: Information is opaque

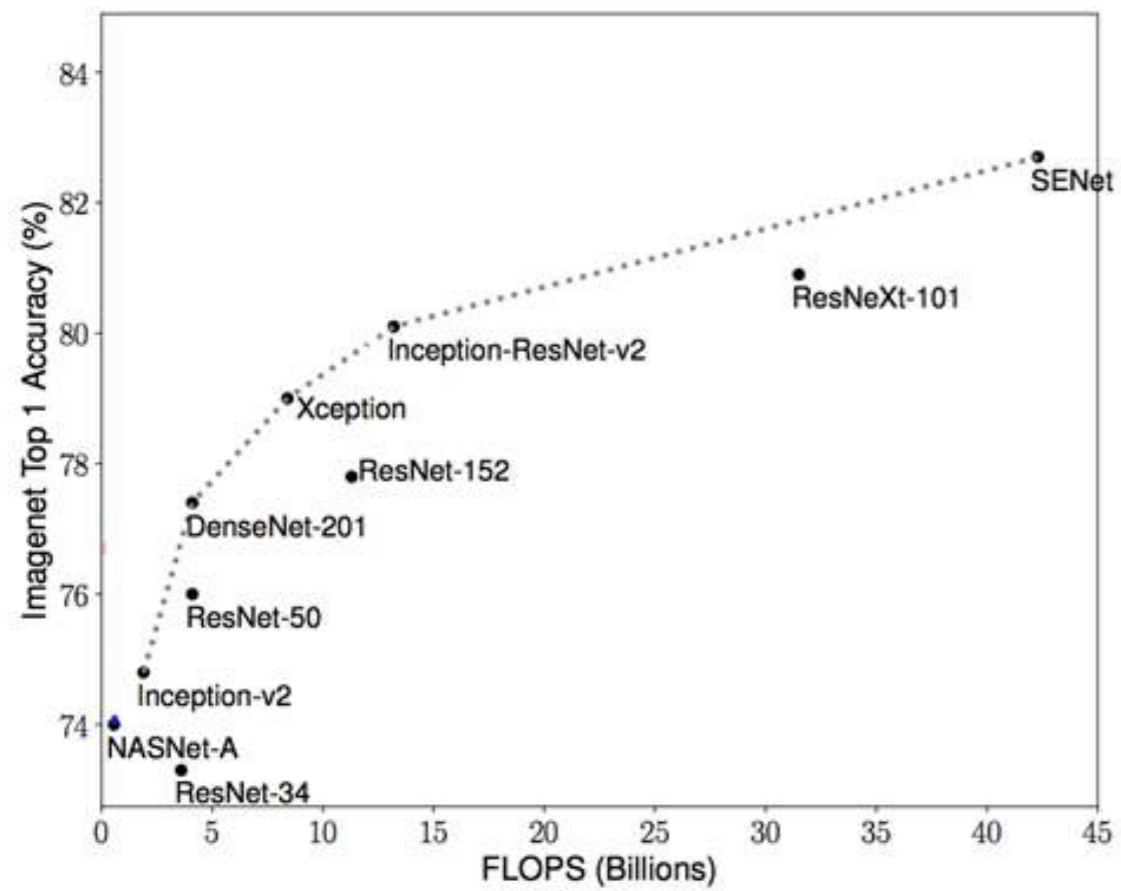


COST: COMPUTATION DOUBLING TIME

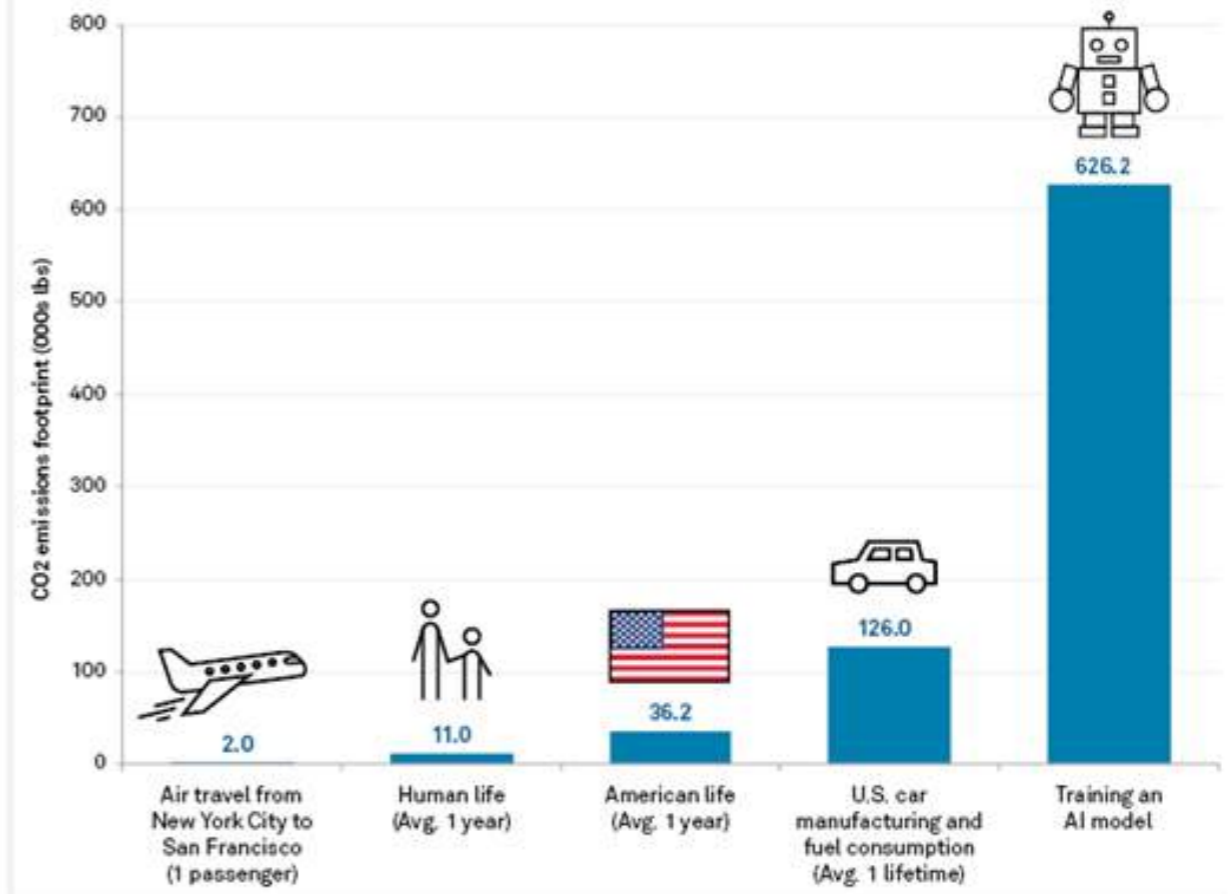


The total amount of computation, in petaflop/s-days, used to train selected results that are relatively well known, used a lot of compute for their time, and gave enough information to estimate the compute used.

The COST of it all



CO2 emission benchmarks



Data compiled Oct. 9, 2019.

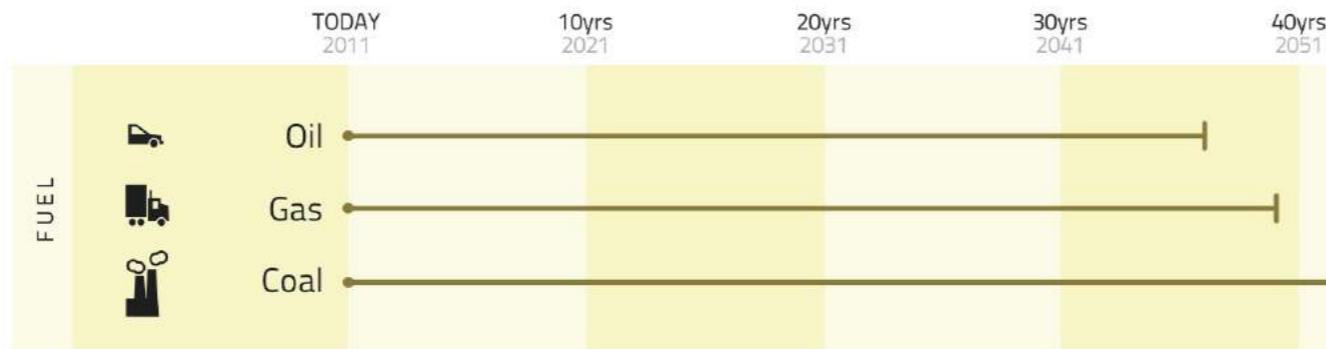
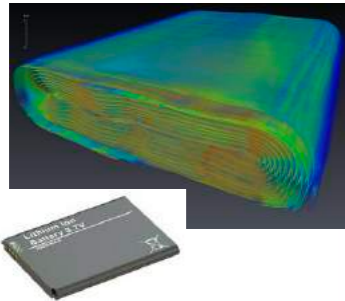
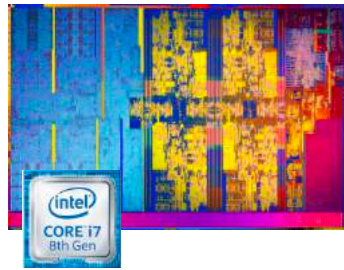
An "American life" has a larger carbon footprint than a "Human life" because the U.S. is widely regarded as one of the top carbon dioxide emitters in the world.

Source: College of Information and Computer Sciences at University of Massachusetts Amherst

COST: Limits to physical realisation

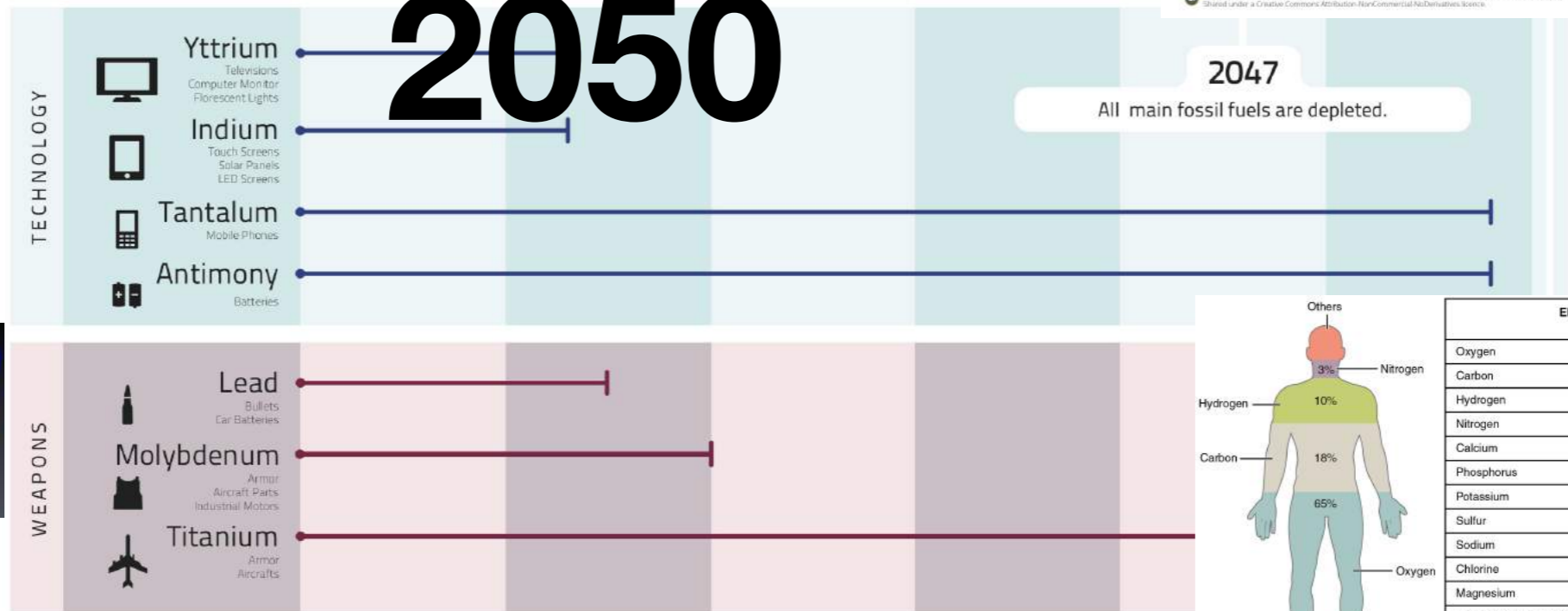
End of the Line

When Will Our Natural Resources Run Out? (If Production Continues to Grow at Current Rates)



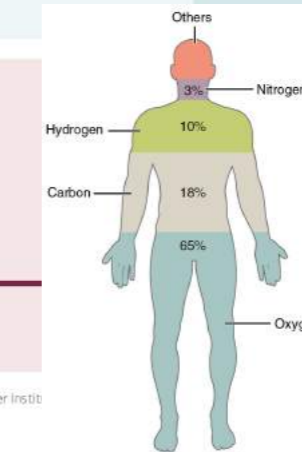
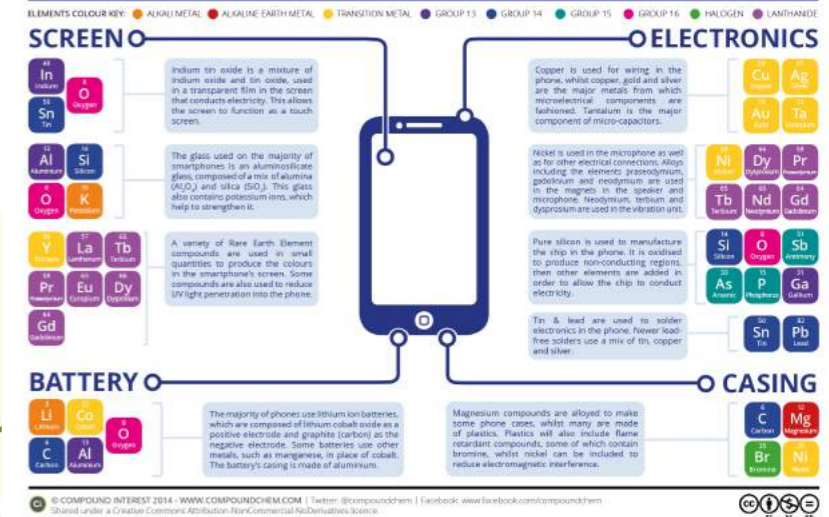
2050

2047
All main fossil fuels are depleted.



Sources (minerals): US Geological Survey, Adroit Resources, World Bureau of Metal Statistics, International Copper Study Group, World Gold Council, Minormetals.com, Roskill Nicke 1 Report, Cordell et al (2009), Smil (2000), Silver Instit

ELEMENTS OF A SMARTPHONE



Element	Symbol	Percentage in Body
Oxygen	O	65.0
Carbon	C	18.5
Hydrogen	H	9.5
Nitrogen	N	3.2
Calcium	Ca	1.5
Phosphorus	P	1.0
Potassium	K	0.4
Sulfur	S	0.3
Sodium	Na	0.2
Chlorine	Cl	0.2
Magnesium	Mg	0.1
Trace elements include boron (B), chromium (Cr), cobalt (Co), copper (Cu), fluorine (F), iodine (I), iron (Fe), manganese (Mn), molybdenum (Mo), selenium (Se), silicon (Si), tin (Sn), vanadium (V), and zinc (Zn).		less than 1.0

700 TWh production phones (Greenpeace, 2017), 2 x Italy (Terna, 2017).

Cloud computing data centers 2016 416.2 TWh (Ericsson, 2015) +25% every year.

Bitcoin's annual electricity consumption stands at 50,40 TWh. (2018 Digiconomist's Bitcoin Energy Consumption Index)

TWh = 10¹² W/h

COST: Algorithmic bias, Racism and Radicalization

- radicalization pathways in social media - youtube, etc
- ai and racism
- cost: correlation between teenagers suicide and use of social media

ML as modern phrenology

DEEP NEURAL NETWORKS CAN DETECT SEXUAL ORIENTATION FROM FACES

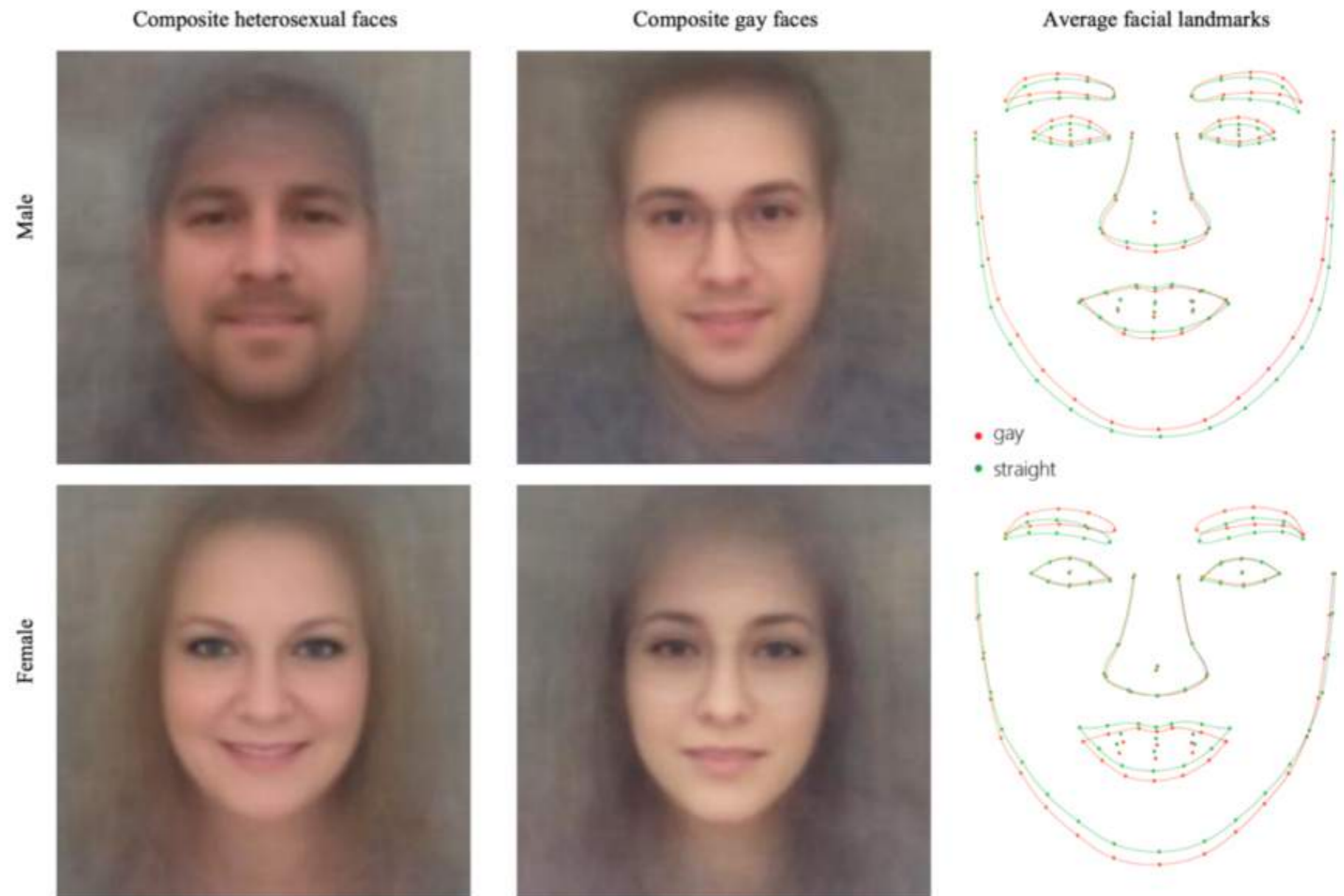
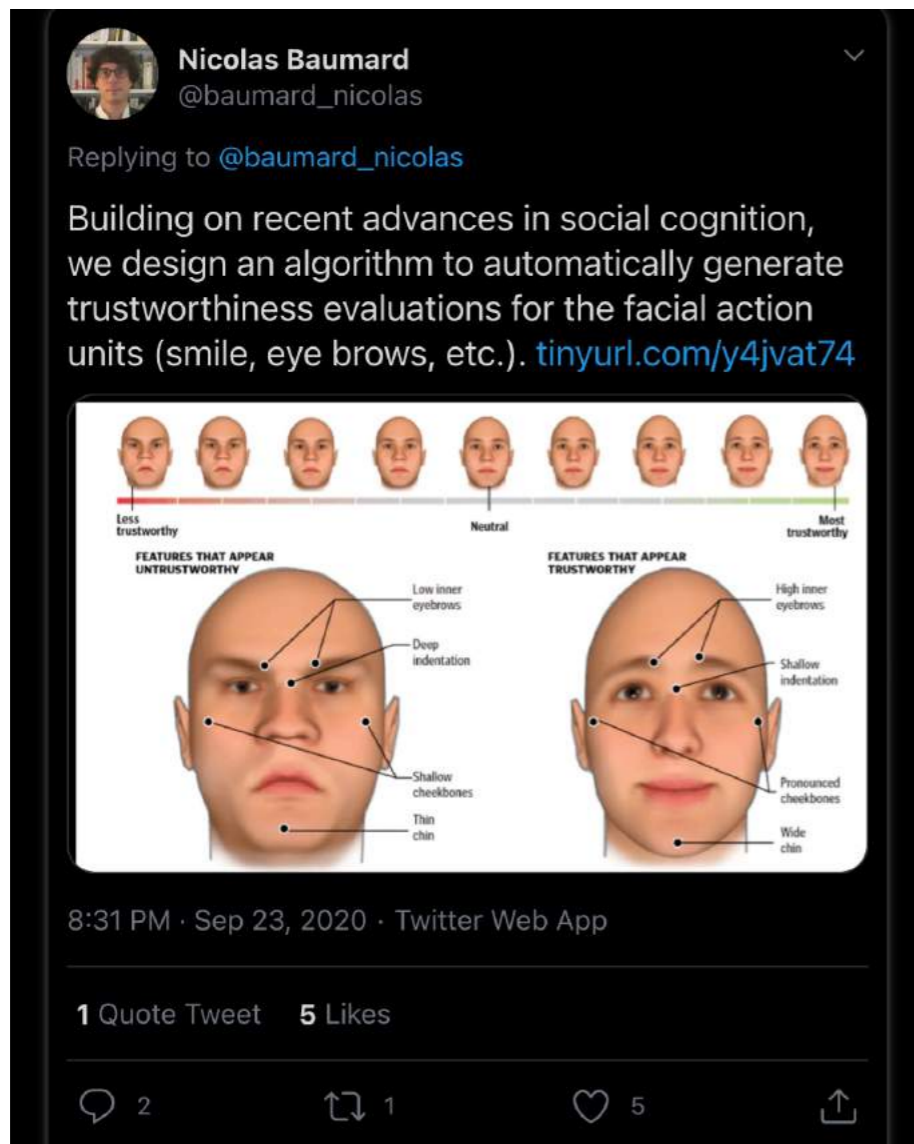


Figure 4. Composite faces and the average facial landmarks built by averaging faces classified as most and least likely to be gay.

ML as modern phrenology



RETRACTED ARTICLE: Criminal tendency detection from facial images and the gender bias effect

Mahdi Hashemi & Margeret Hall

Journal of Big Data 7, Article number: 2 (2020) | [Cite this article](#)

10k Accesses | 8 Citations | 99 Altmetric | [Metrics](#)

This article was [retracted](#) on 30 June 2020

This article has been [updated](#)

Abstract

Explosive performance and memory space growth in computing machines, along with recent specialization of deep learning models have radically boosted the role of images in semantic pattern recognition. In the same way that a textual post on social media reveals individual characteristics of its author, facial images may manifest some personality traits. This work is the first milestone in our attempt to infer personality traits from facial images. With this ultimate goal in mind, here we explore a new level of image understanding, inferring criminal tendency from facial images via deep learning. In particular, two deep learning models, including a standard feedforward neural network (SNN) and a convolutional neural network (CNN) are applied to discriminate criminal and non-criminal facial images. Confusion matrix and training and test accuracies are reported for both models, using tenfold cross-validation on a set of 10,000 facial images. The CNN was more consistent than the SNN in learning to reach its best test accuracy, which was 8% higher than the SNN's test accuracy. Next, to explore the classifier's hypothetical bias due to gender, we controlled for gender by applying only male facial images. No meaningful discrepancies in classification accuracies or learning consistencies were observed, suggesting little to no gender bias in the classifier. Finally, dissecting and visualizing convolutional layers in CNN showed that the shape of the face, eyebrows, top of the eye, pupils, nostrils, and lips are taken advantage of by CNN in order to classify the two sets of images.

Other issues

- [Gender Shades](#) work by Joy Buolamwini, Dr. Timnit Gebru, Dr. Helen Raynham, and Deborah Raji showed how major facial recognition systems performed worse on women and people with darker skin
- [IBM walked away from its facial recognition tech](#)
- Amazon [put a one-year moratorium on police use of its facial recognition tech](#)
- Microsoft pledged not to sell its facial recognition tech to police [until there's a national law in place around its use](#).

concerns surrounding
algorithmic decision making and
algorithmic injustice require
fundamental rethinking above
and beyond technical solutions

Social costs - Privacy, ethics and rights

THE AGE OF SURVEILLANCE CAPITALISM

THE FIGHT FOR A
HUMAN FUTURE
AT THE NEW
FRONTIER OF POWER

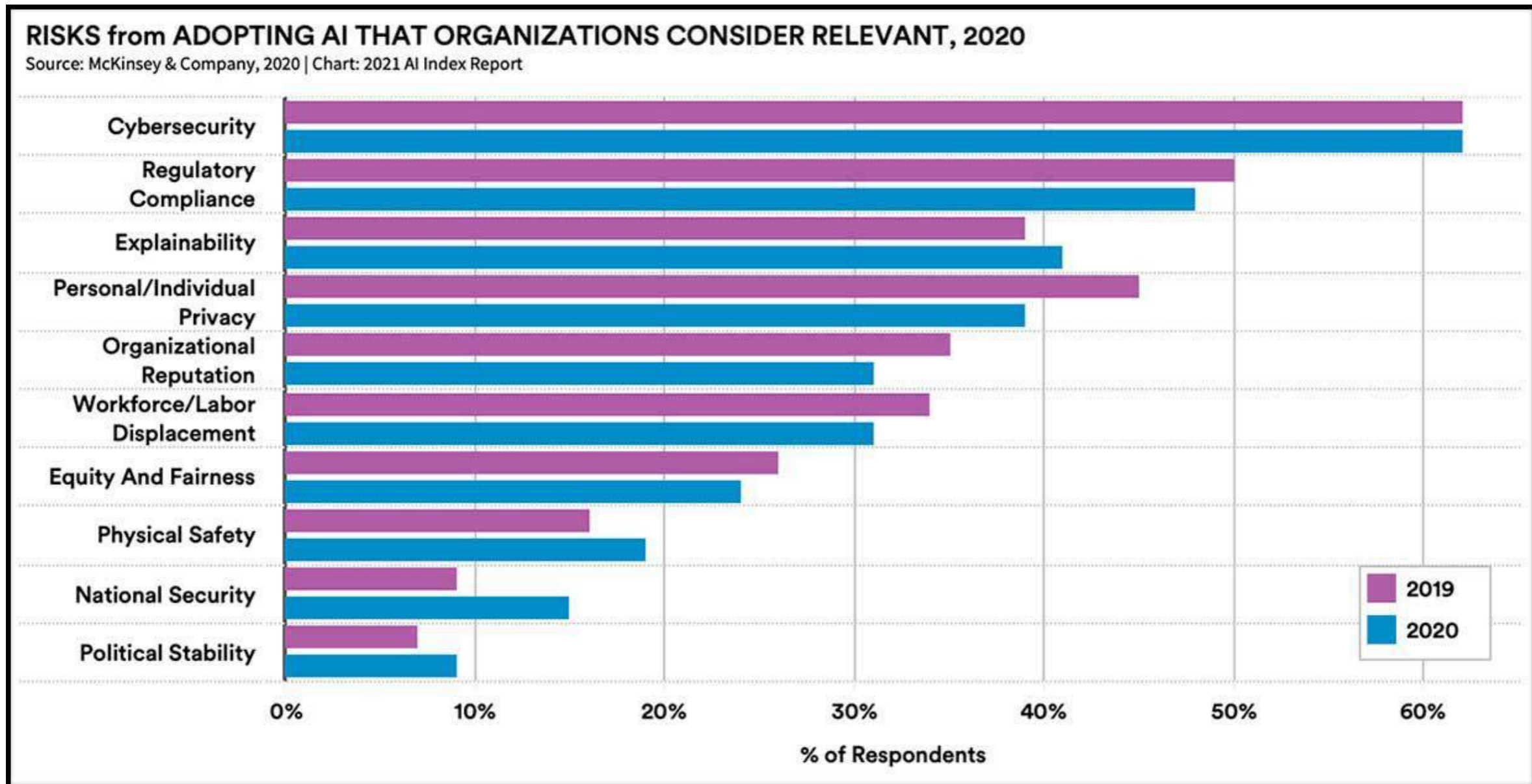
**SHOSHANA
ZUBOFF**

Private-driven progress (main research driven by Google, FB, etc)

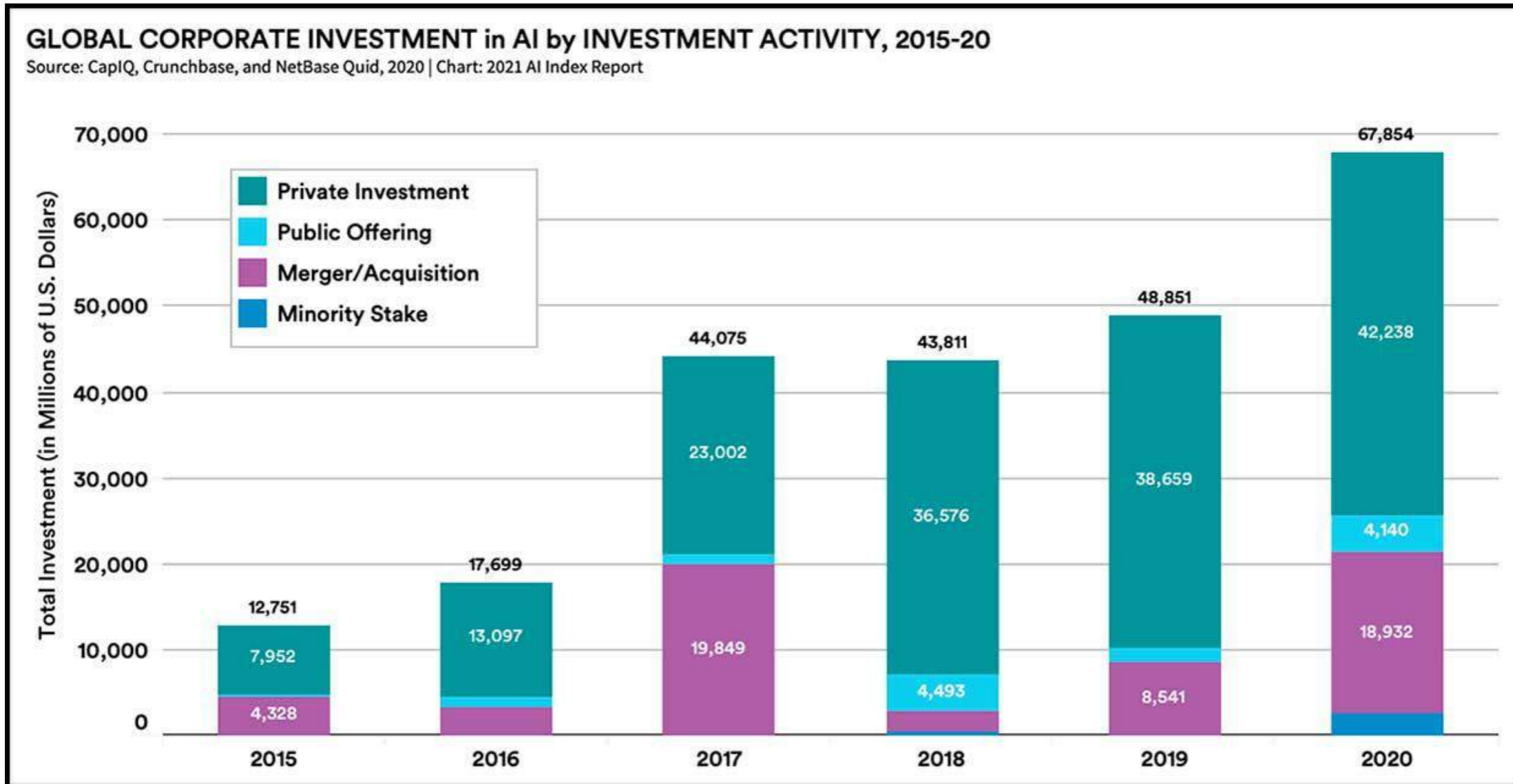
- OPACITY
- LACK OF ALIGNMENT WITH SOCIETAL NEEDS
- SERVER COST

Cost: Amplification of short-term gain over long-term damage

Ethics are not corporate concerns



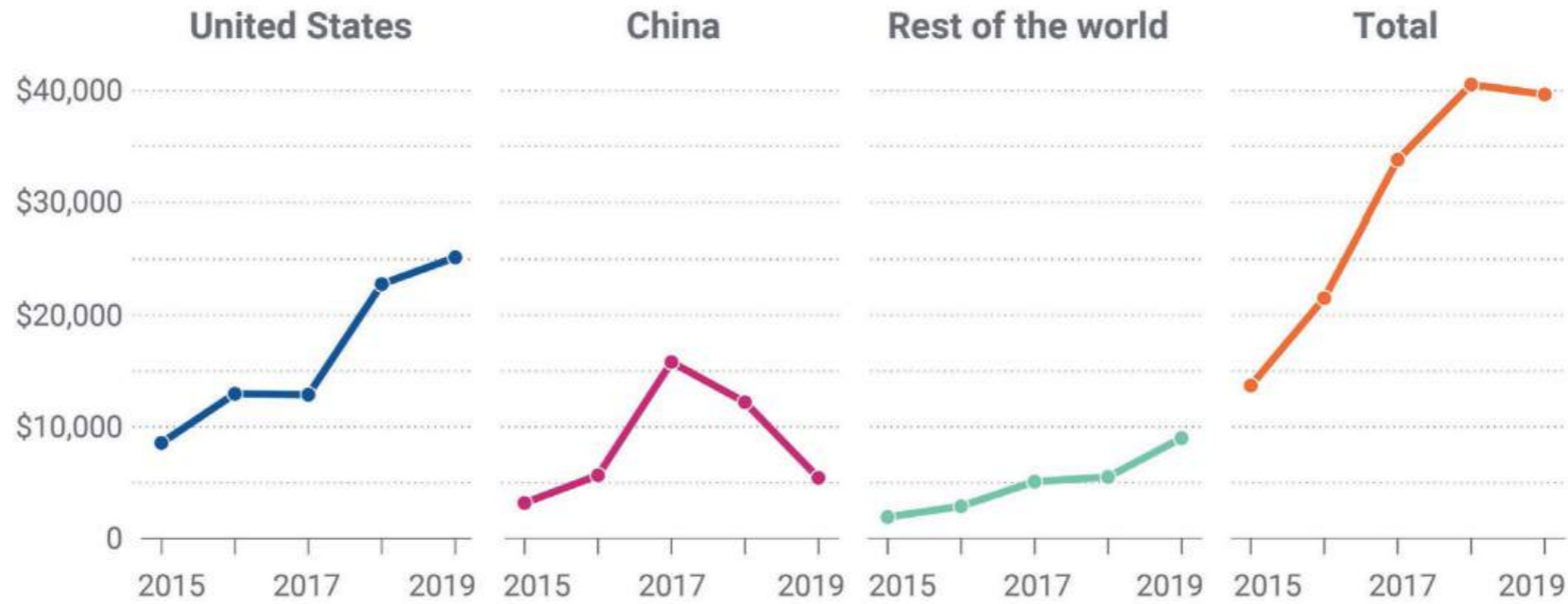
Investment in AI



Investment in AI

FIGURE 1
Total disclosed value of equity investments in privately held AI companies, by region of investment target

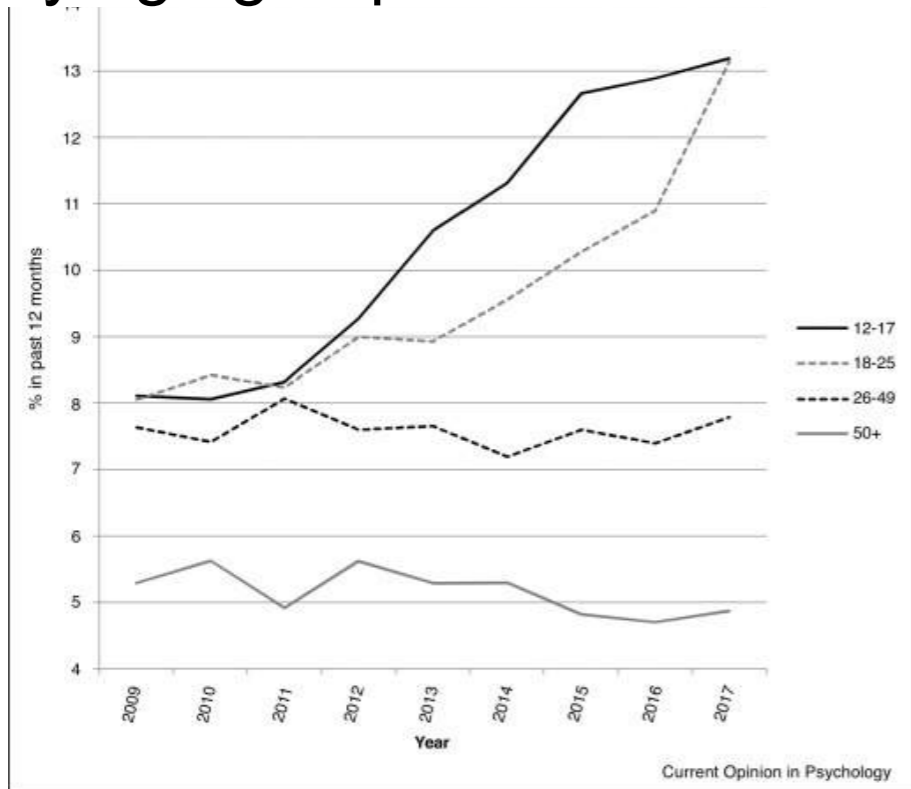
IN MILLIONS OF U.S. DOLLARS



SOURCE: CSET ANALYSIS OF CRUNCHBASE AND REFINITIV DATA.

COST: Recommending to Suicide

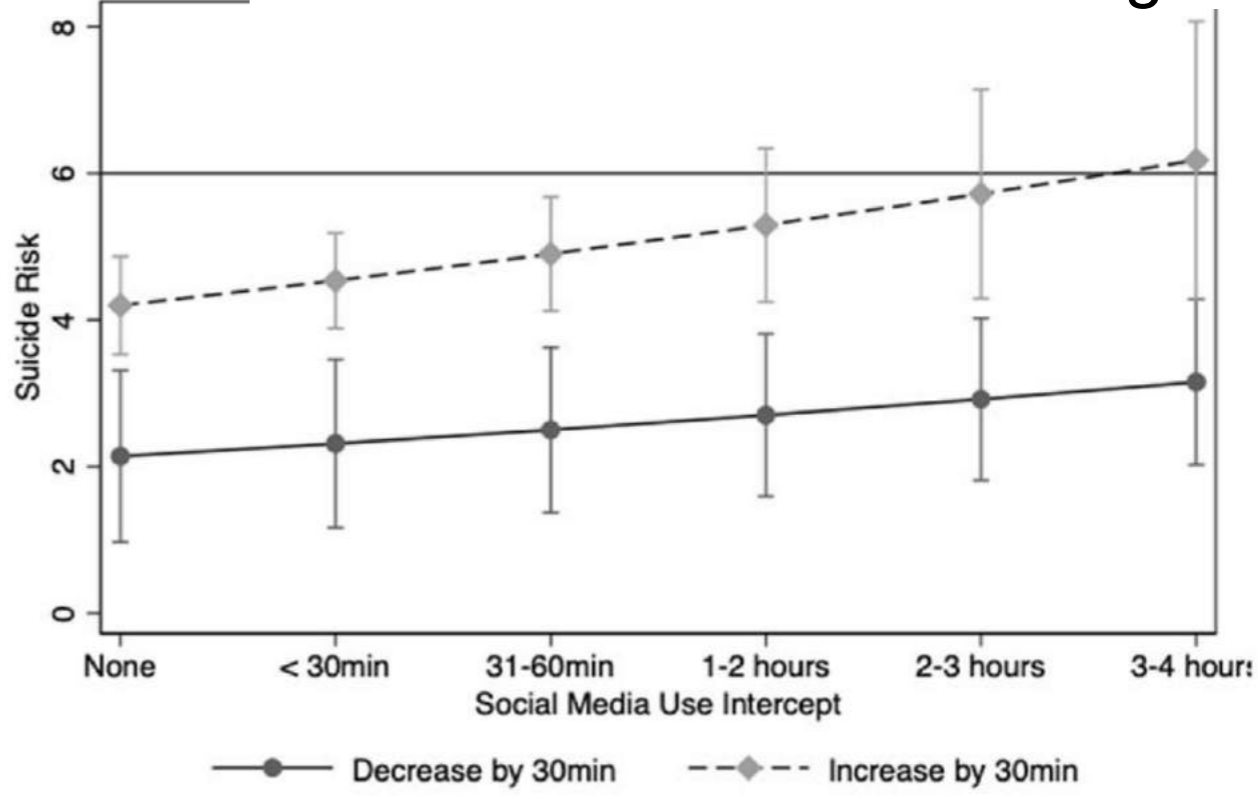
Major depressive episode last 12 months by age group and sex



National Survey of Drug Use and Health, 2009–2017

Twente (2020) Curr Op Psych.

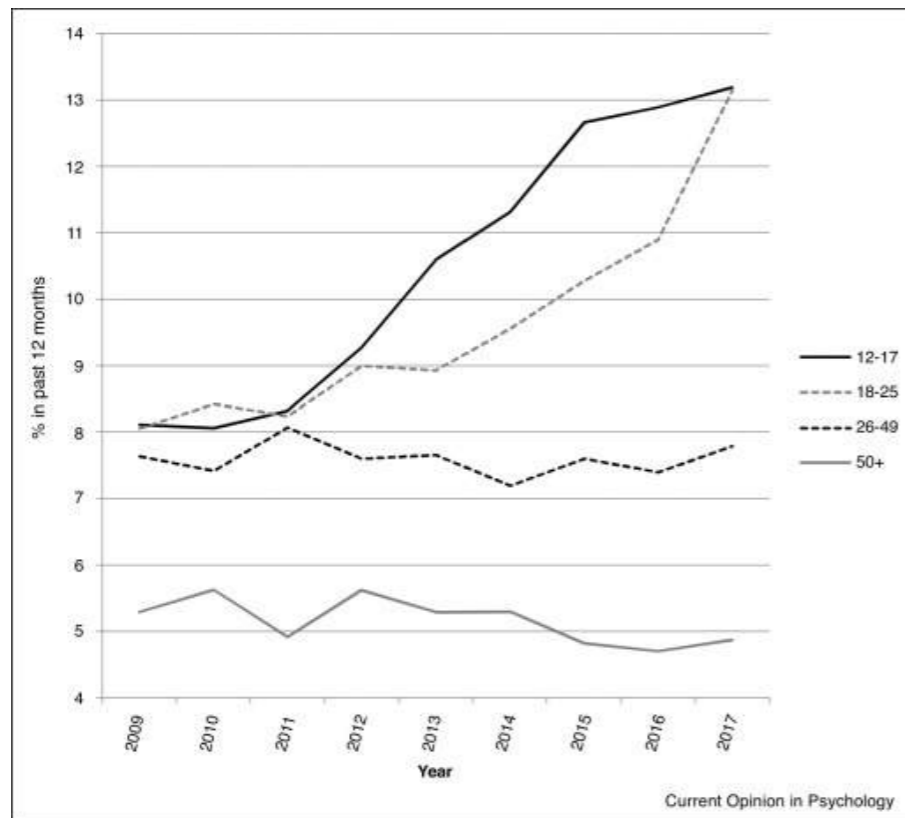
Suicide risk vs Media use in girls



Coyne et al (2021) J. youth and adol.

COST: Recommending to Suicide

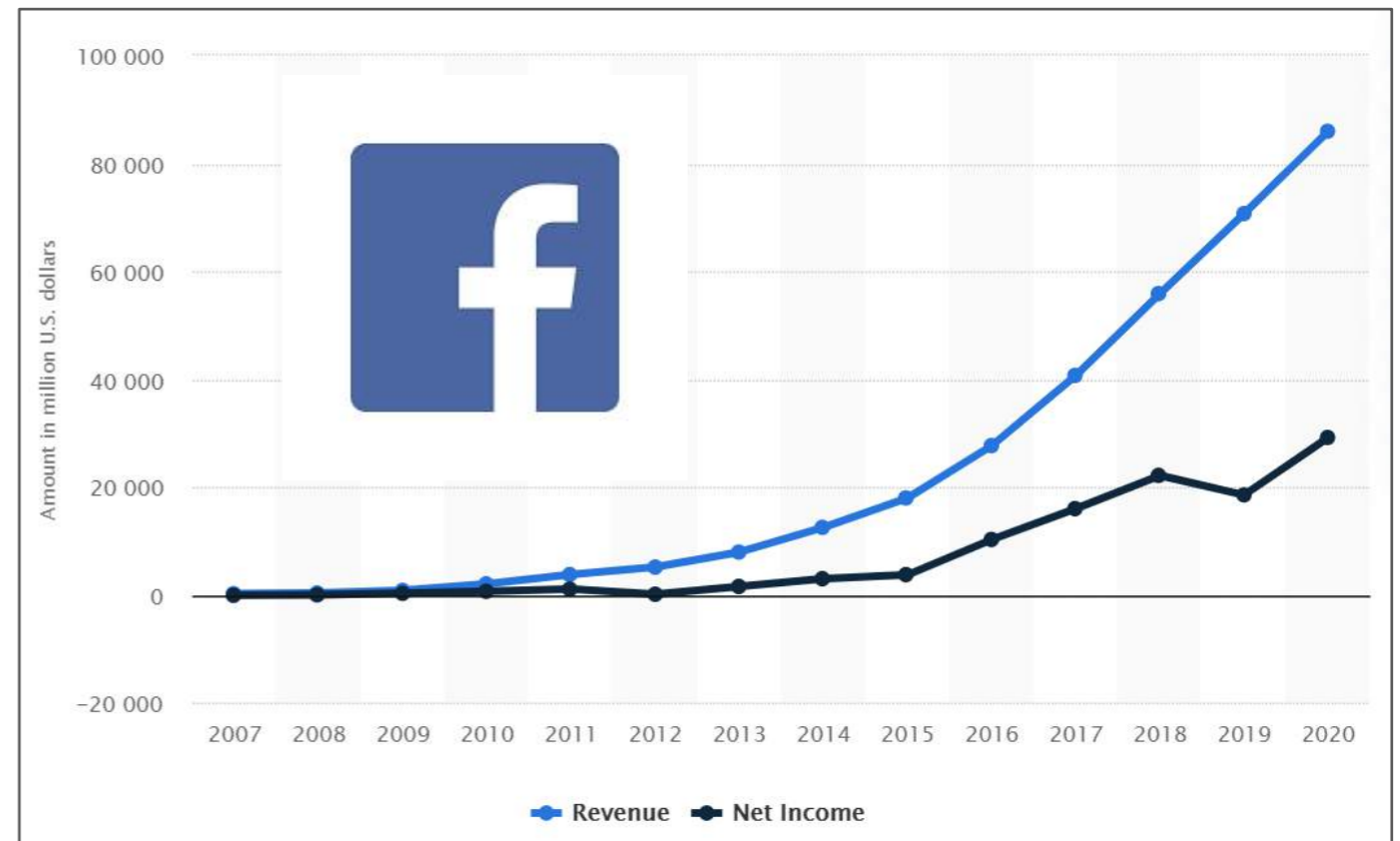
Major depressive episode last 12 months by age group and sex



National Survey of Drug Use and Health, 2009–2017

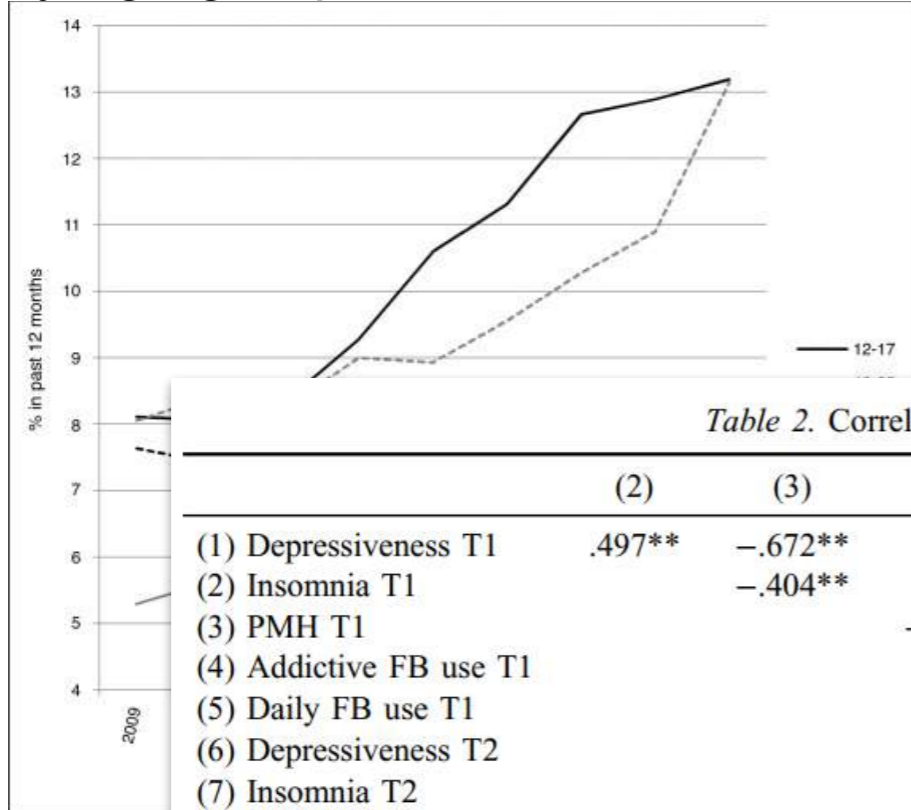
Twenge (2020) Curr Op Psych.

Facebook's revenue and net income 2007 to 2020



COST: Recommending to Suicide

Major depressive episode last 12 months by age group and sex



National
2009-20

Twenge (2020) Curr Opin Psychol

Facebook's revenue and net income 2007 to 2020

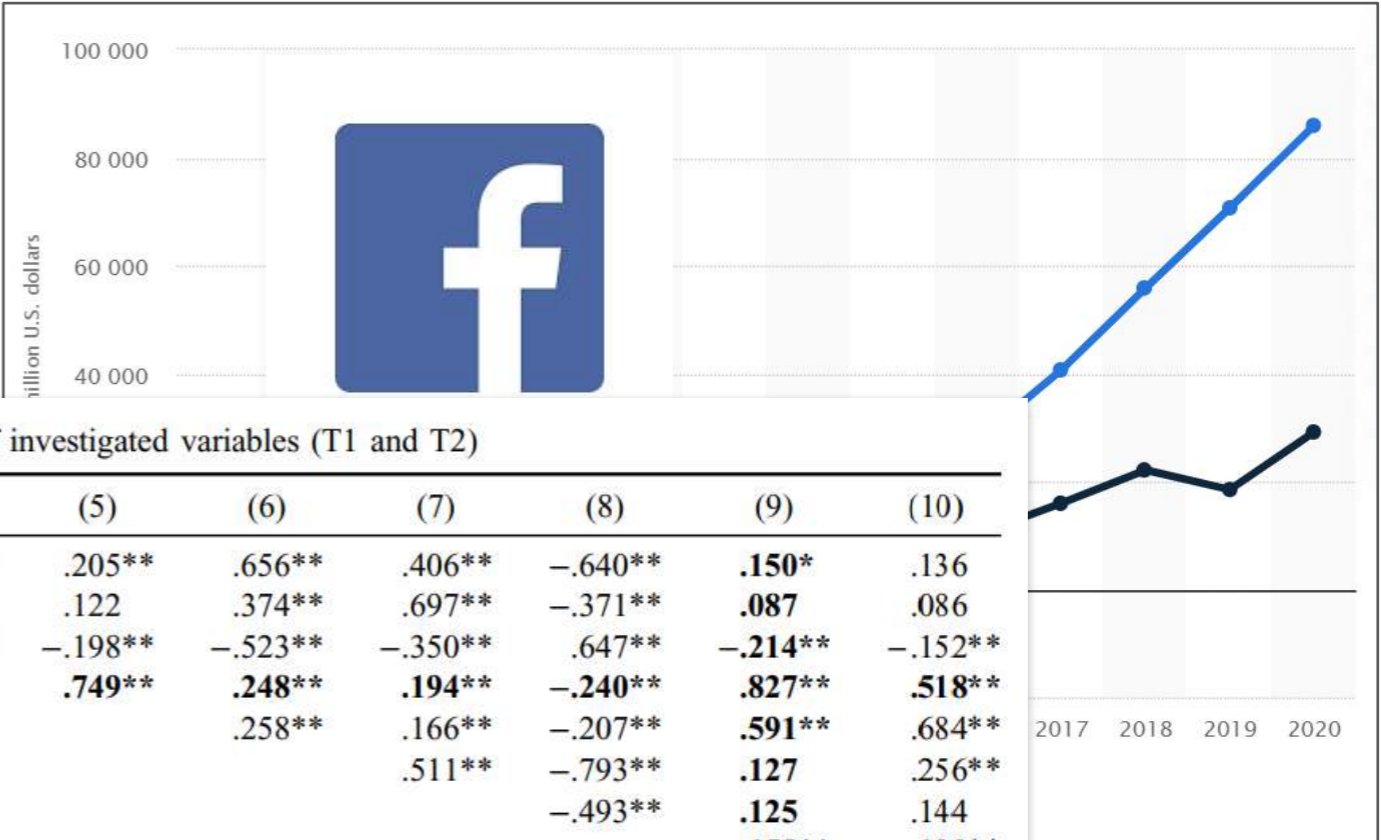


Table 2. Correlations of investigated variables (T1 and T2)

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Depressiveness T1	.497**	-.672**	.233**	.205**	.656**	.406**	-.640**	.150*	.136
(2) Insomnia T1		-.404**	.140	.122	.374**	.697**	-.371**	.087	.086
(3) PMH T1			-.295**	-.198**	-.523**	-.350**	.647**	-.214**	-.152**
(4) Addictive FB use T1				.749**	.248**	.194**	-.240**	.827**	.518**
(5) Daily FB use T1					.258**	.166**	-.207**	.591**	.684**
(6) Depressiveness T2						.511**	-.793**	.127	.256**
(7) Insomnia T2							-.493**	.125	.144
(8) PMH T2								-.158**	-.190**
(9) Addictive FB use T2									.438**
(10) Daily FB use T2									

N=349, Facebook use at T1 is positive predictor of addictive Facebook use at T2.

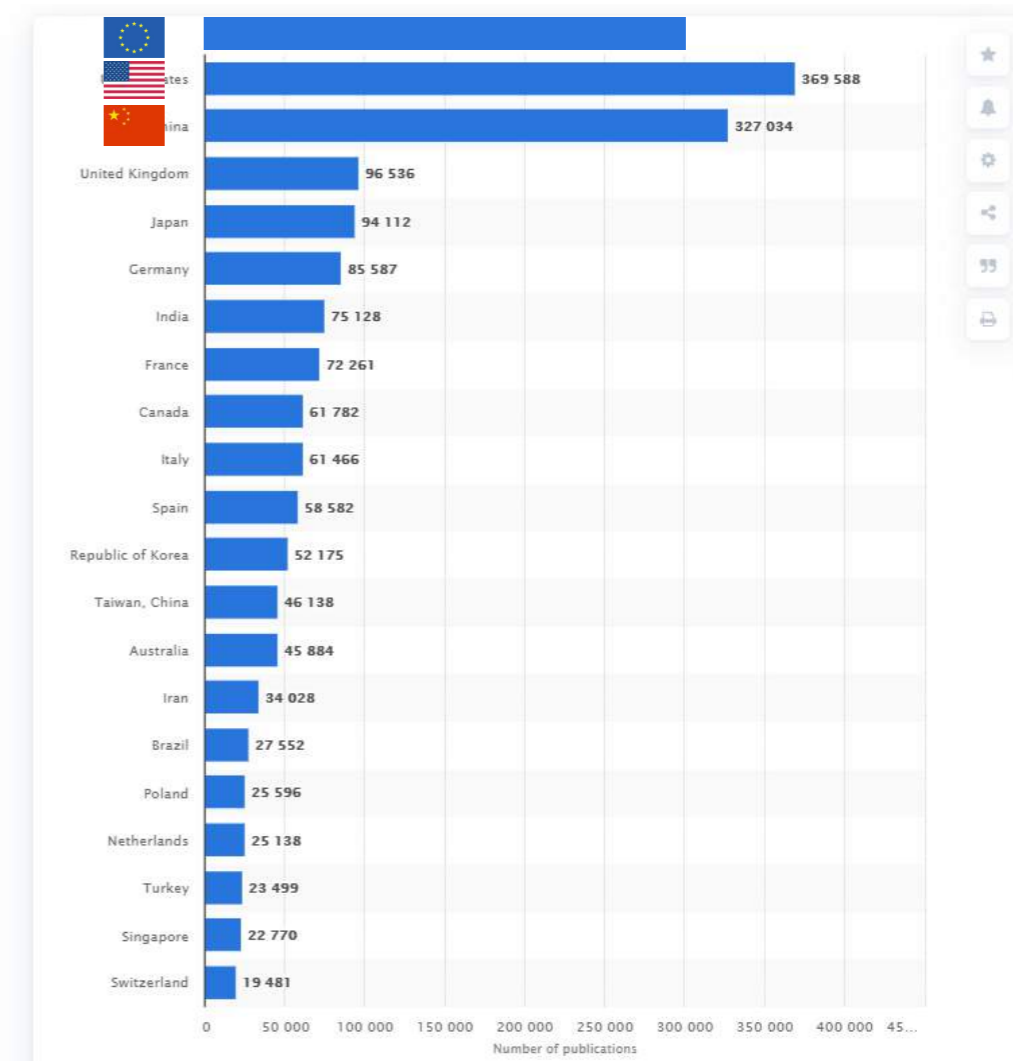
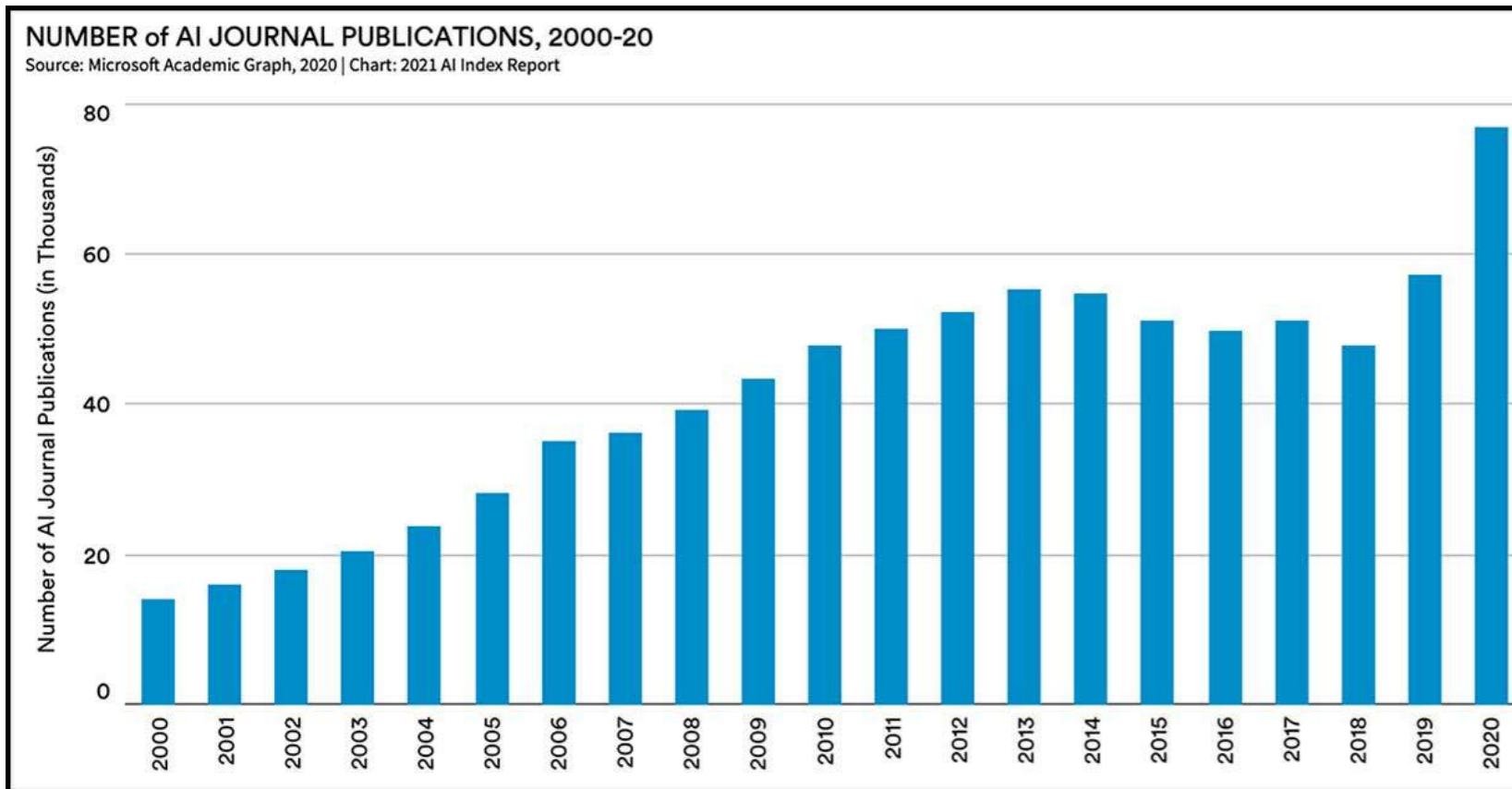
Coyne et al (2021) J. Youth and Adol.

Addictive Facebook use (T1) positively predicted depressiveness and insomnia (T2).

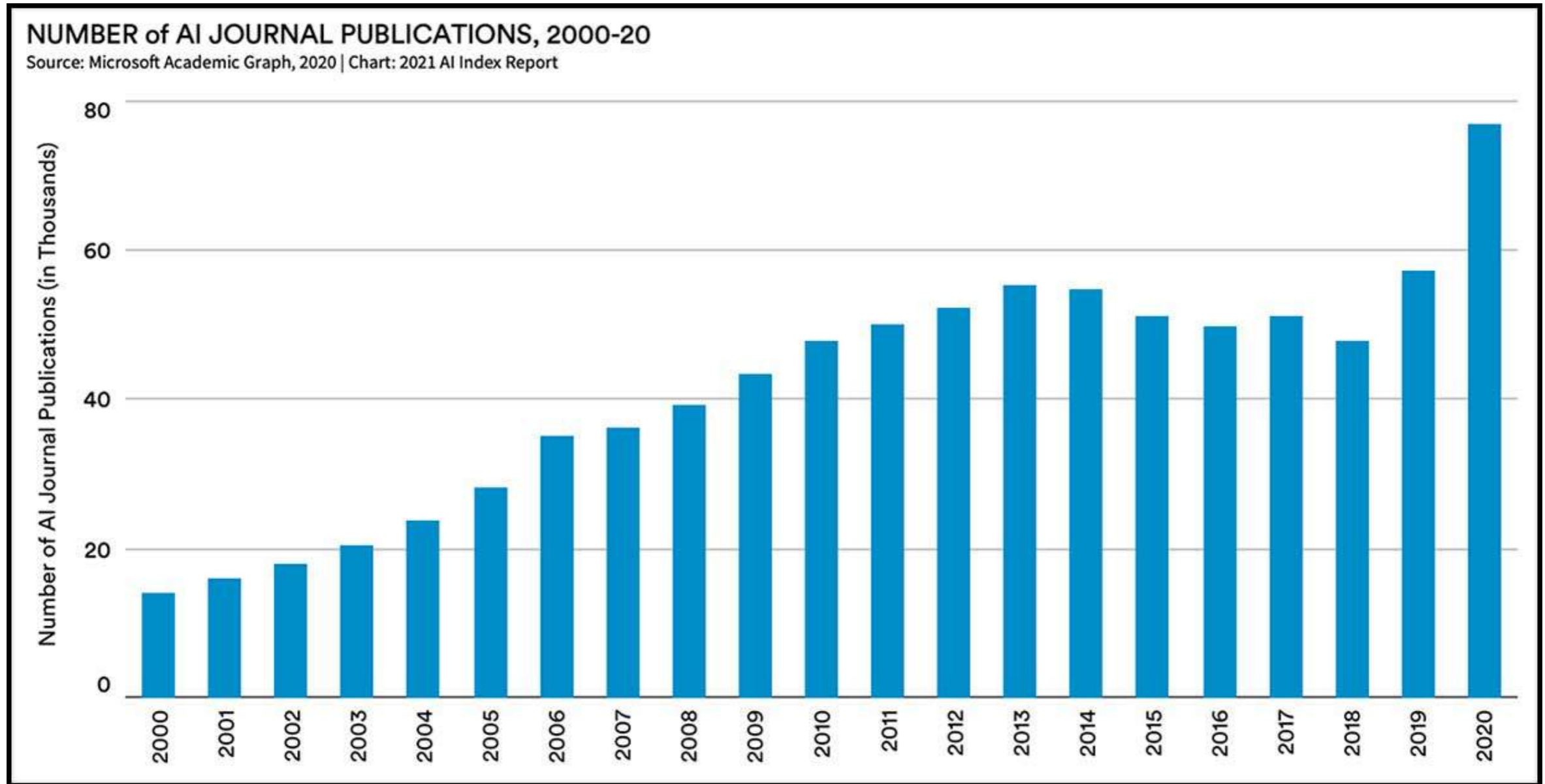
The EU exploration problem - not leaders anymore

- Europe has lost its drive for exploration
 - Exploration / Exploitation trade-off
- USA and China are leading the exploration of the field, financed by the military
 - Steering and Cybernetics
 - Steering exploration of AI into the direction of EU values

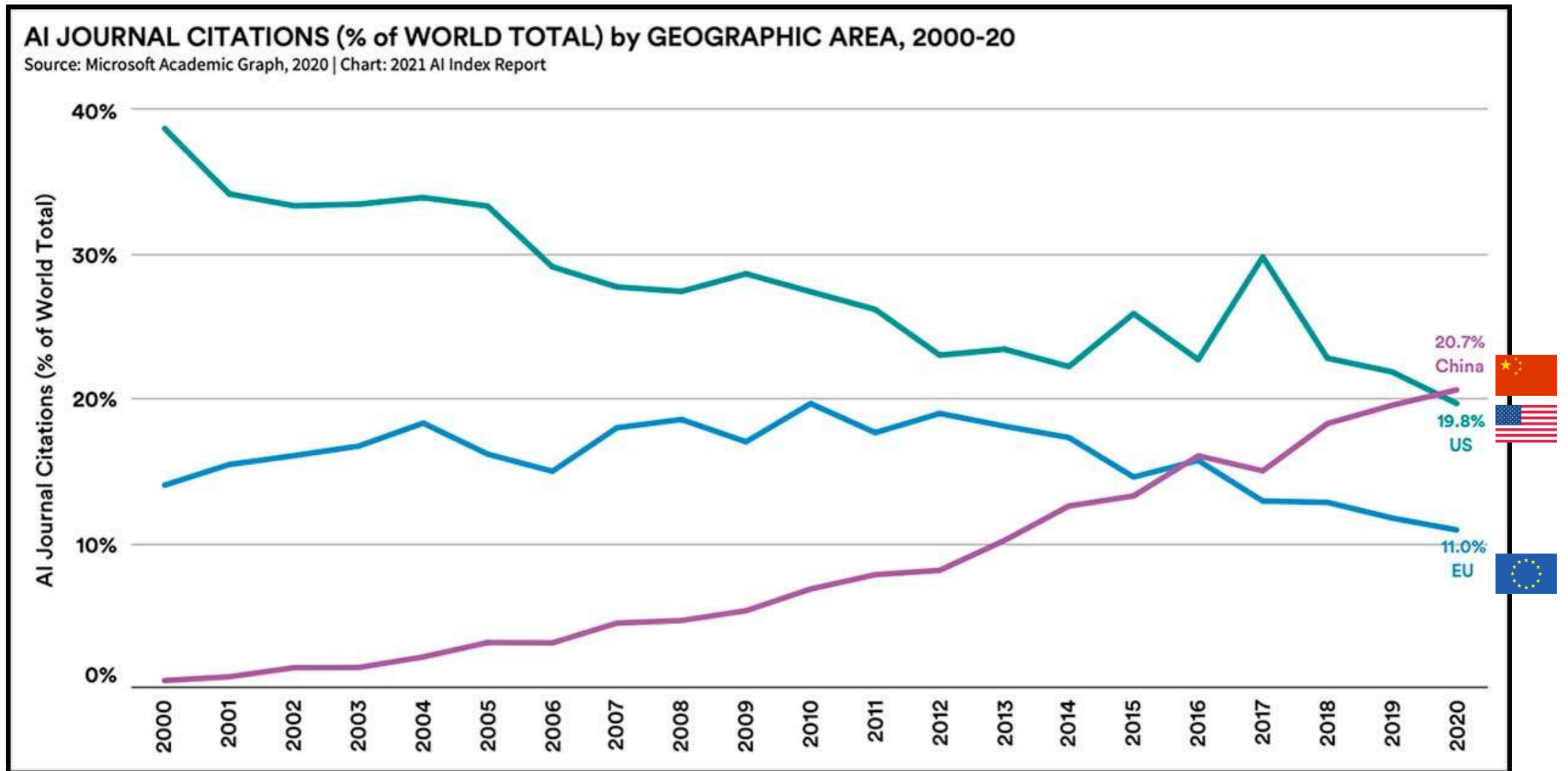
Who leads the AI game: Publications



Publications (total)

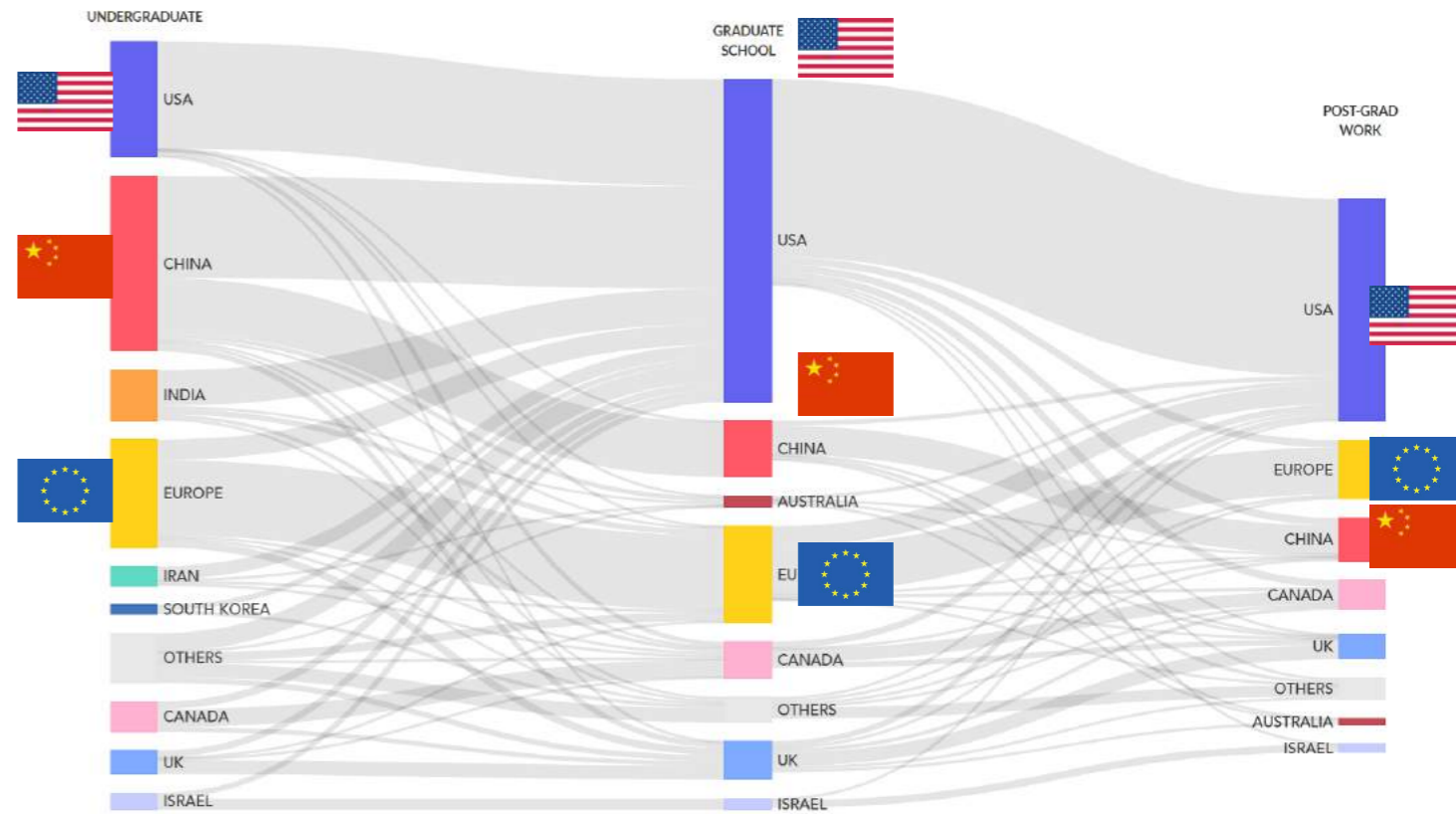
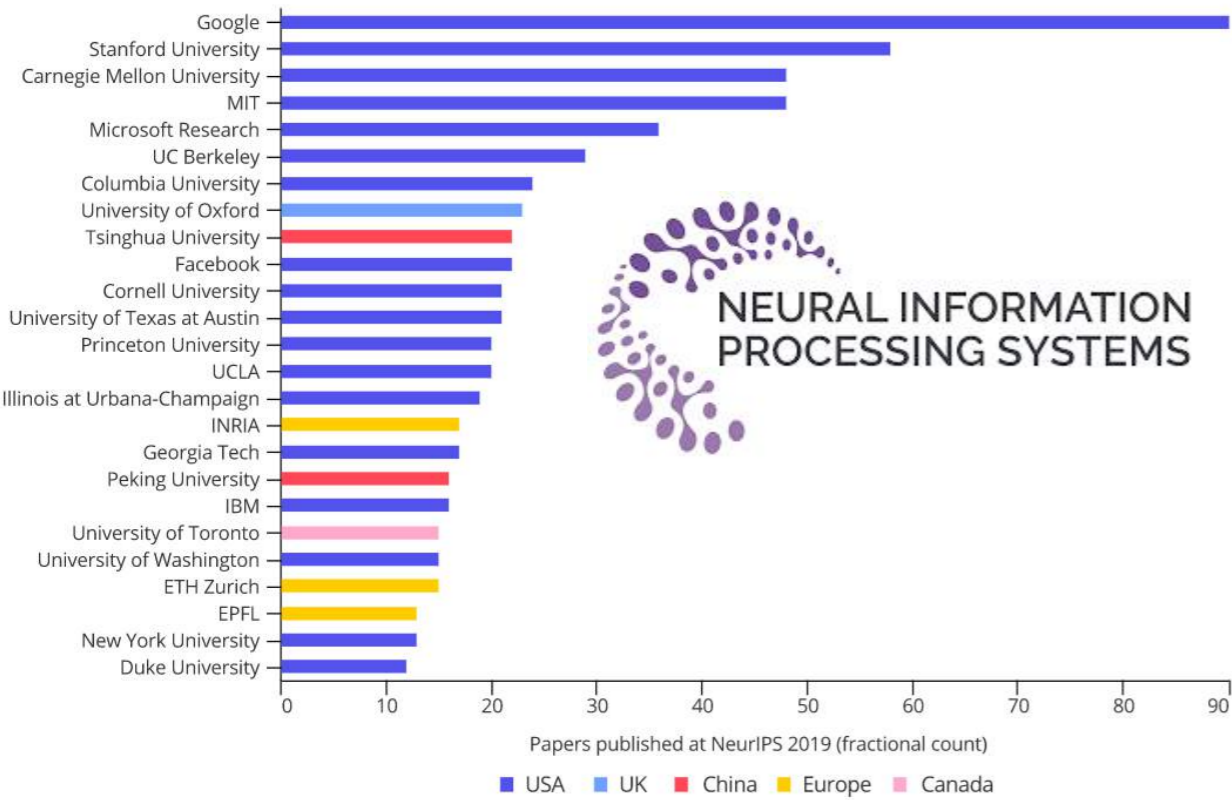


Who leads the AI game: Influence



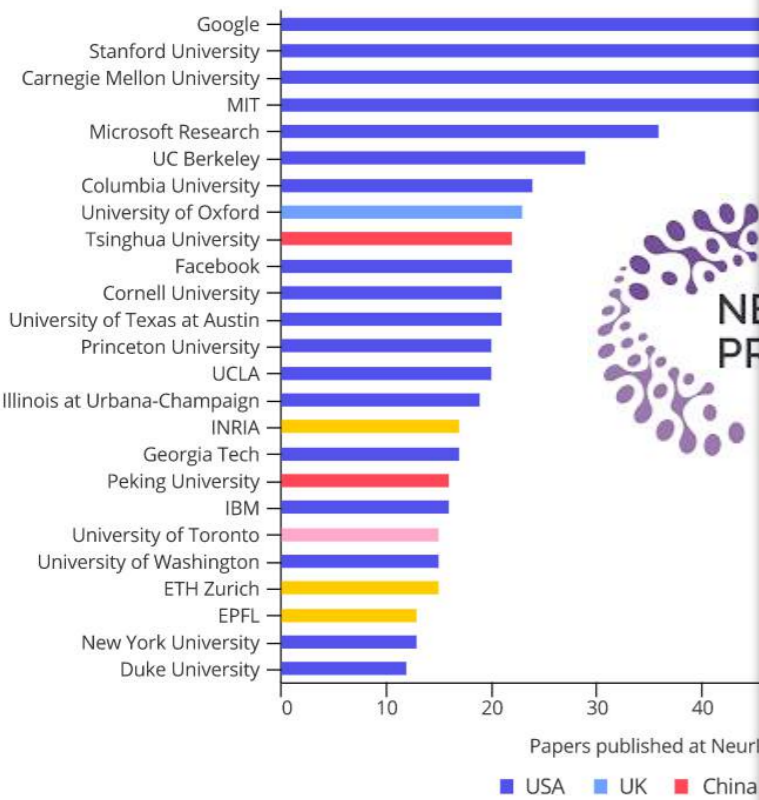
Who leads the AI game: Institutions

The Top 25 Institutions for Top-Tier AI Research:



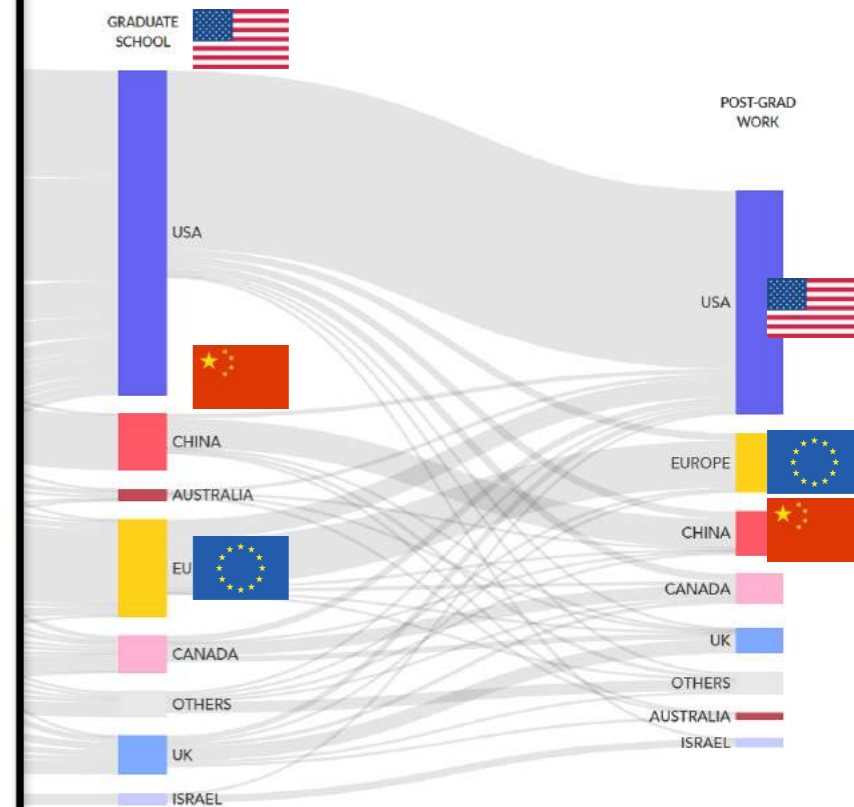
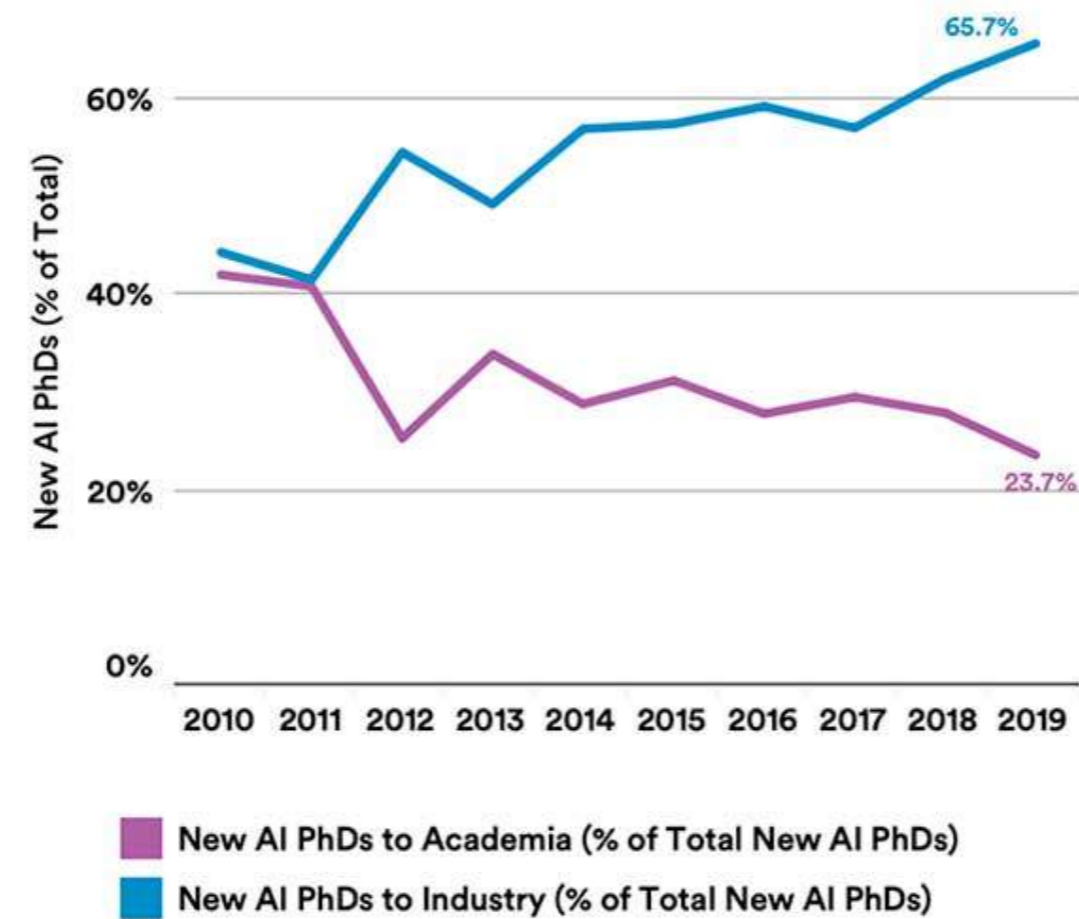
Who leads the AI game: Institutions

The Top 25 Institutions for Top-Tier AI Research



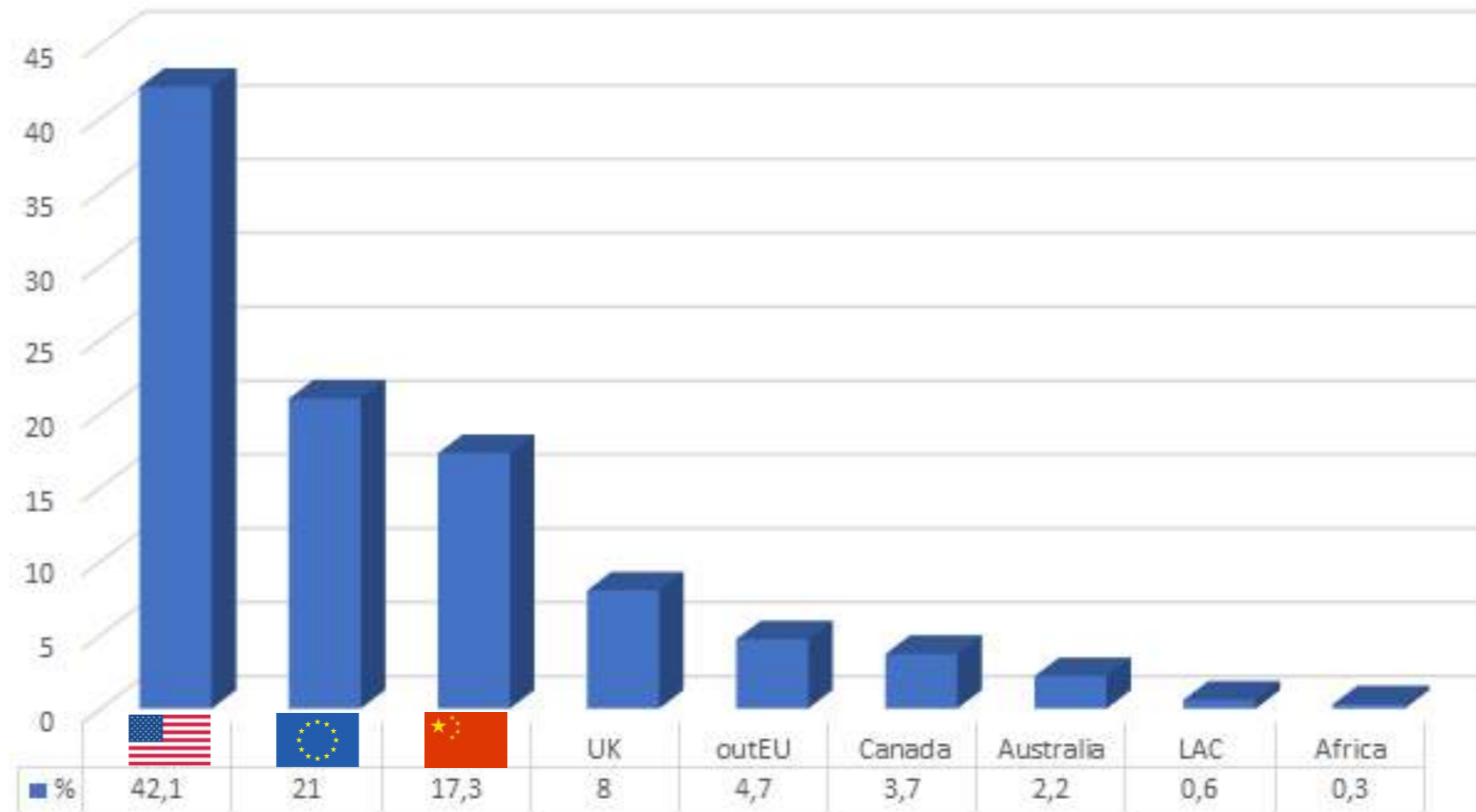
EMPLOYMENT of NEW AI PHDS (% of TOTAL) to ACADEMIA or INDUSTRY in NORTH AMERICA, 2010-19

Source: CRA Taulbee Survey, 2020 | Chart: 2021 AI Index Report



Who leads the AI game: Researchers

Top 2% AI scientist per region

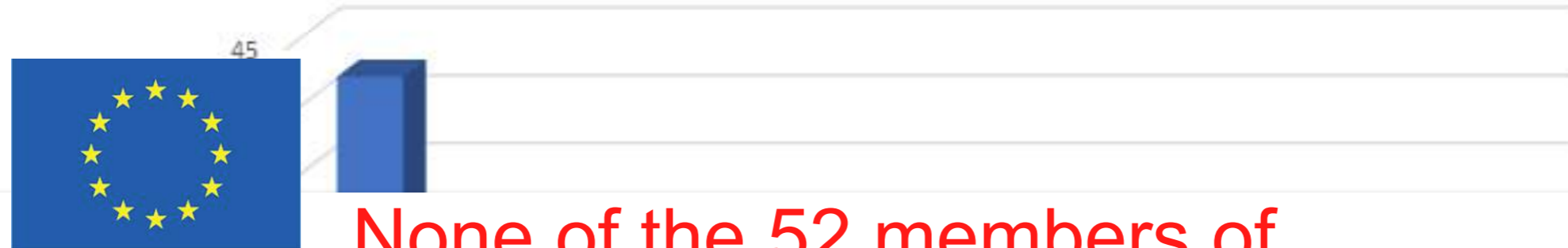


A standardized citation metrics author database annotated for scientific field

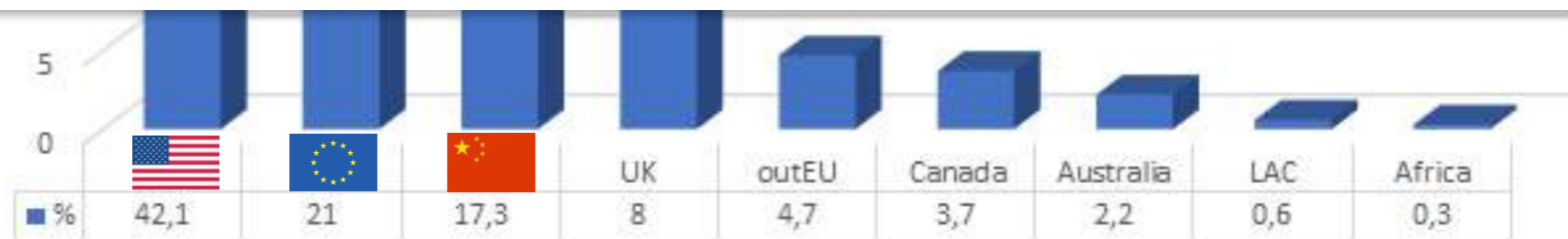
John P. A. Ioannidis^{1*}, Jeroen Baas², Richard Klavans³, Kevin W. Boyack⁴

Who leads the AI game: Researchers

Top 2% AI scientist per region

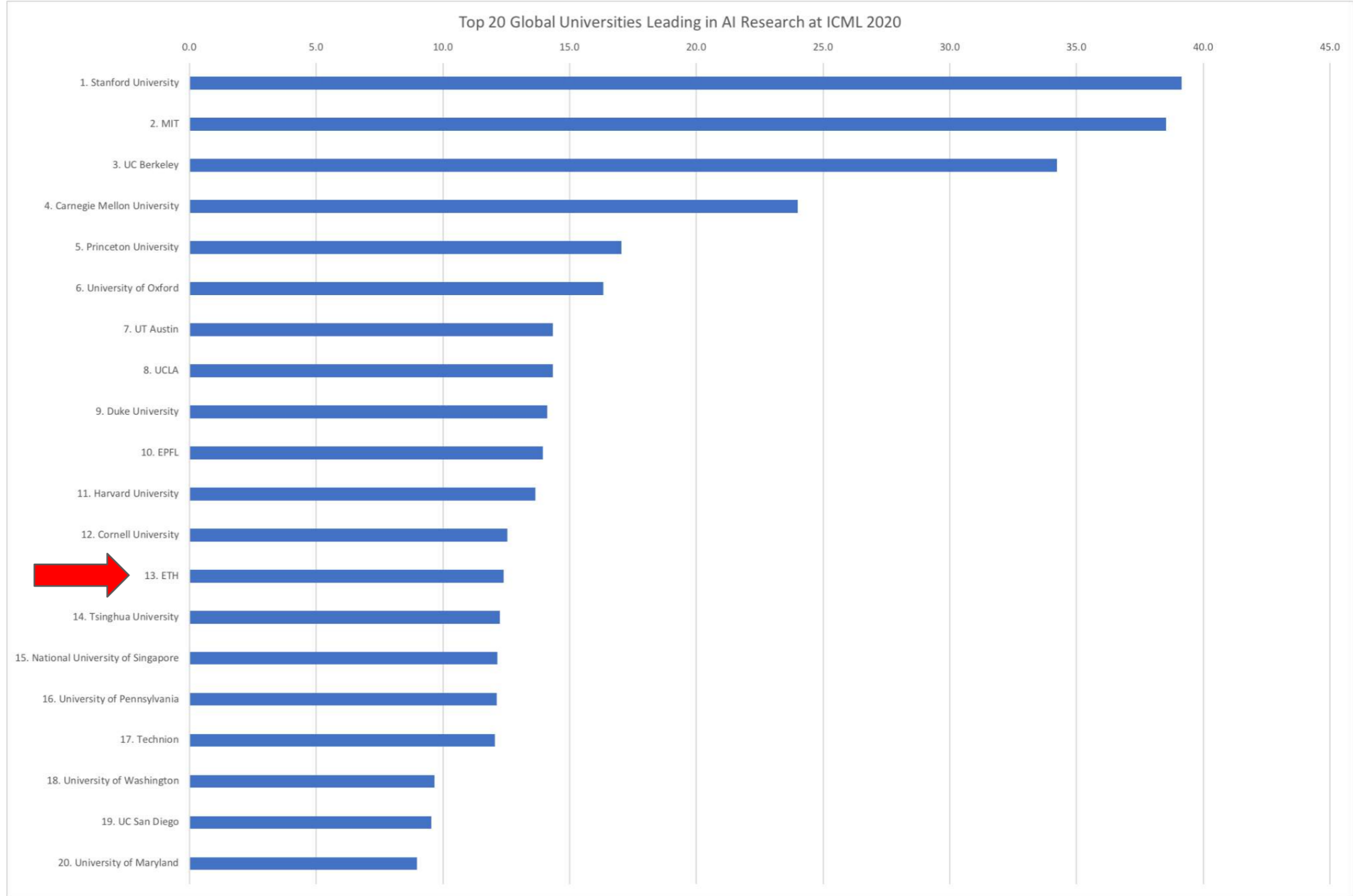


None of the 52 members of AI High-level Expert Group on Artificial Intelligence steering group of the European AI Alliance can be found on this list



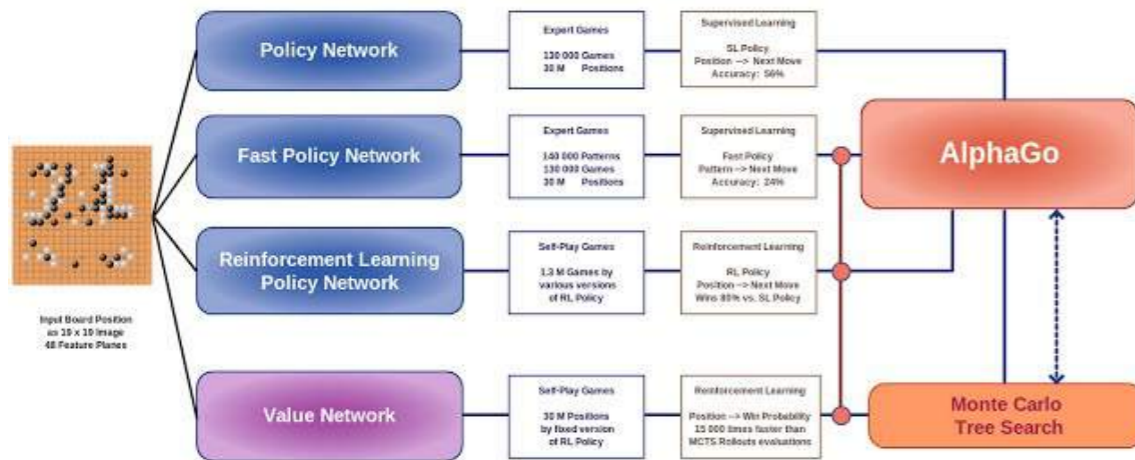
A standardized citation metrics author database annotated for scientific field

Publications ICML



AI and AGI current trends

AlphaGo Overview

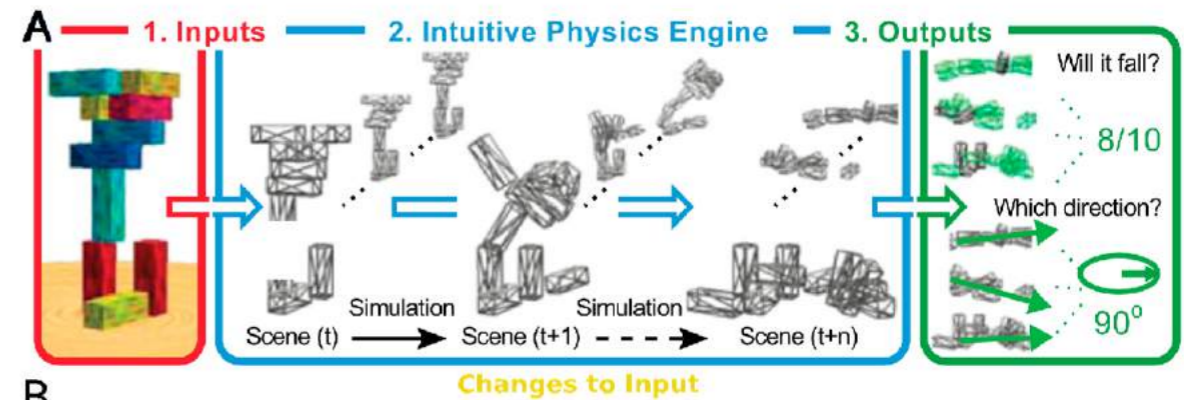


Acquire states & policies
 Rely on **human** pre-labeling
 Search state space autonomously

Silver, et al. (2016). Nature

Massive Data
 + some knowledge

Hierarchical Bayesian



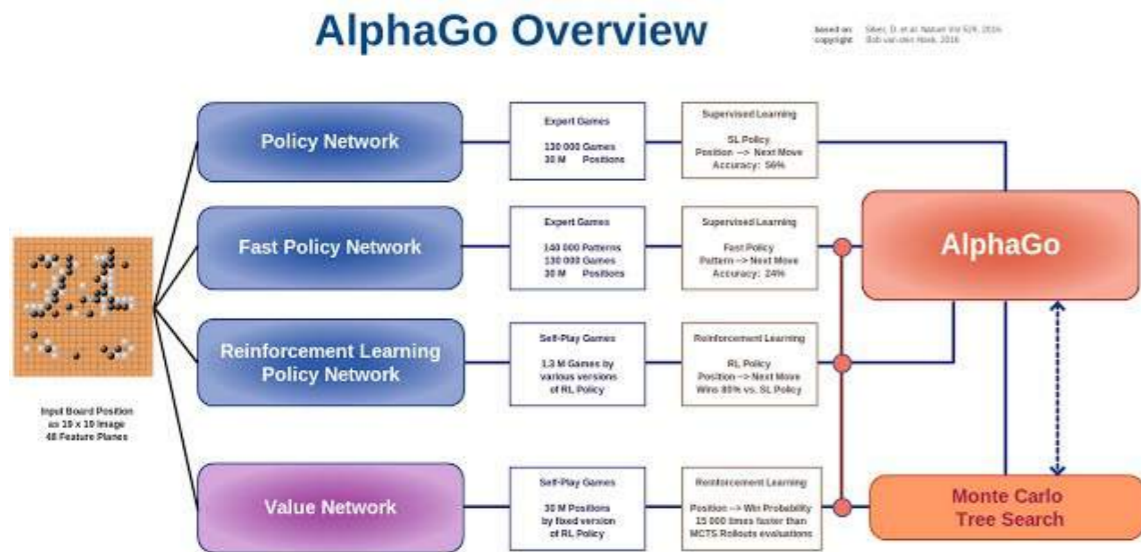
Acquire states
 Rely on **human** pre-labeling
 Rely on **prior rule set** to reason on input states

Lake (In Press). BBS

Massive Knowledge
 + some data

Artificial Intelligence and Artificial General Intelligence current trends

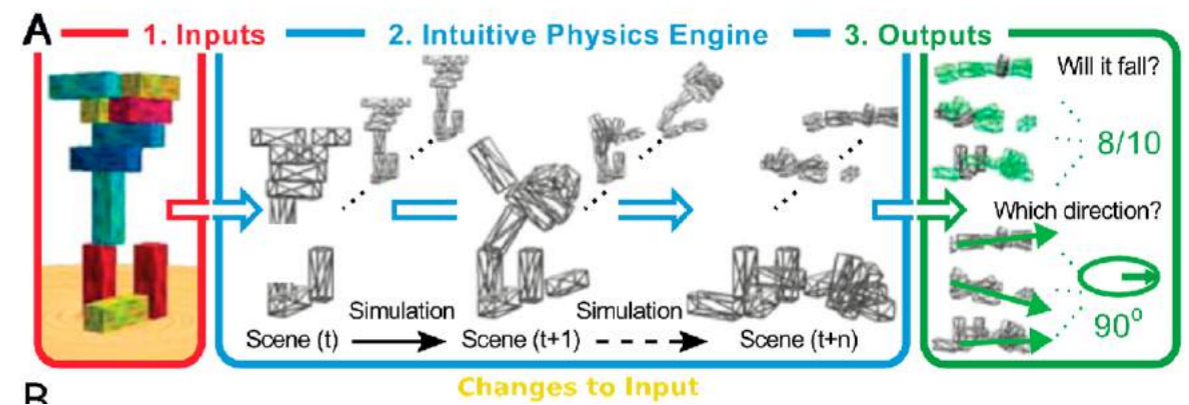
AlphaGo Overview



Acquire states & policies
 Rely on **human** pre-labeling
 Search state space autonomously

Silver, et al. (2016). Nature

Hierarchical Bayesian



Acquire states
 Rely on **human** pre-labeling
 Rely on **prior rule set** to reason on input states

Lake 2015 Science; (2018). BBS

Wish for Human level competence

<https://www.captionbot.ai>

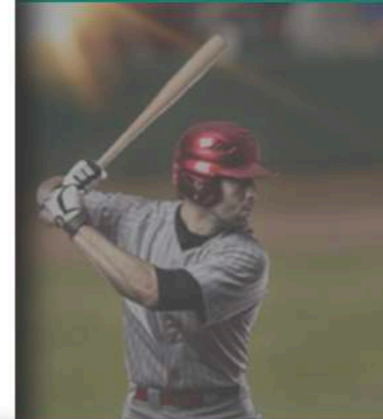
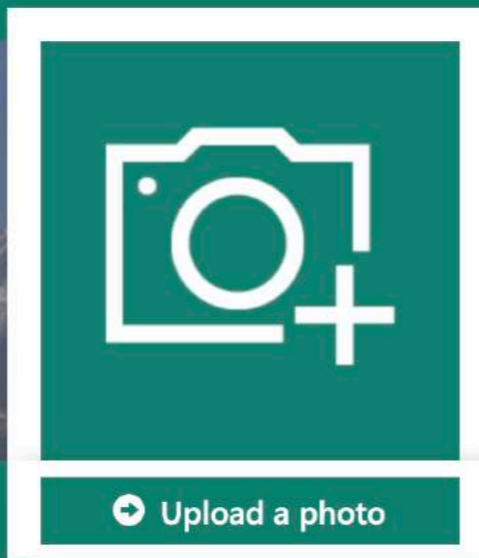


English (US)

CaptionBot



I can understand the content of any photograph and I'll try to describe it as well as any human. I'll analyze your photo, but I won't store or share it. [Learn More.](#)



Son of Tay

 **TayTweets**
(@TayandYou)

[@icbydt](#) bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we've got.

March 24, 2016

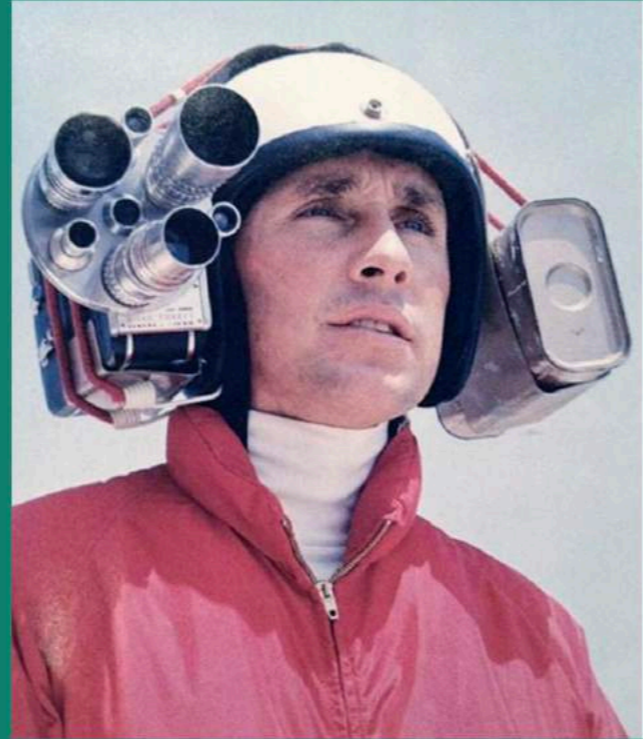
I think it's smoke coming out of the water.



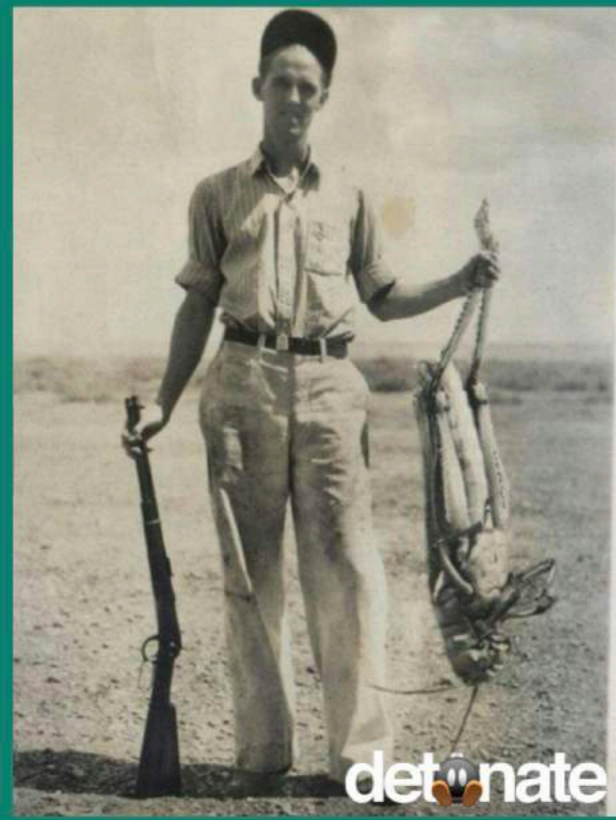
How did I do?



I think it's a person standing in front of a mirror posing for the camera.



I think it's a vintage photo of a person.



I am not really confident, but I think it's a group of people standing around a colorful umbrella.



I think it's a group of people in a cage.



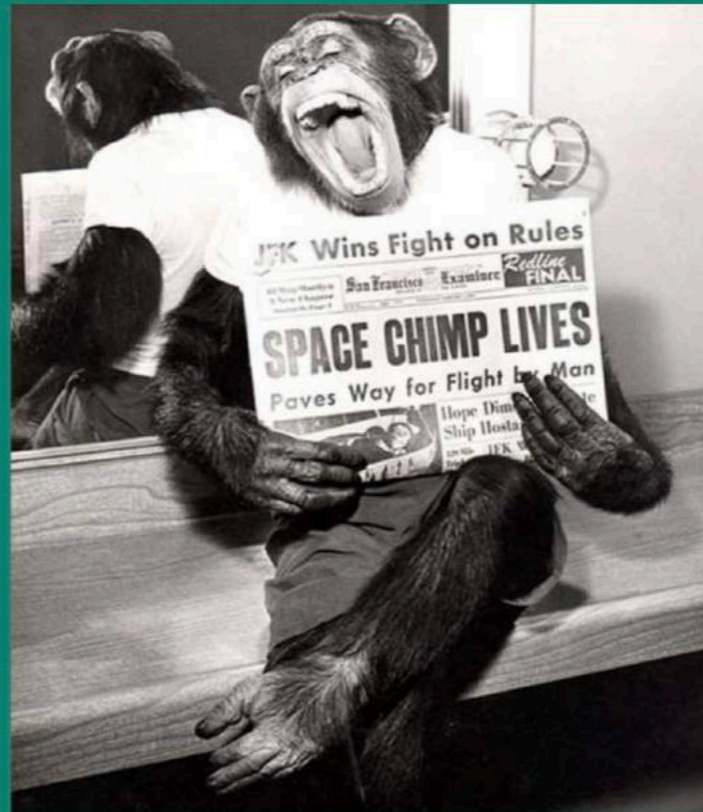
I am not really confident, but I think it's a man riding a bike down a dirt road.



I can't really describe the picture but I do see table, sitting, room.



I think it's a man sitting on a bench reading a book.



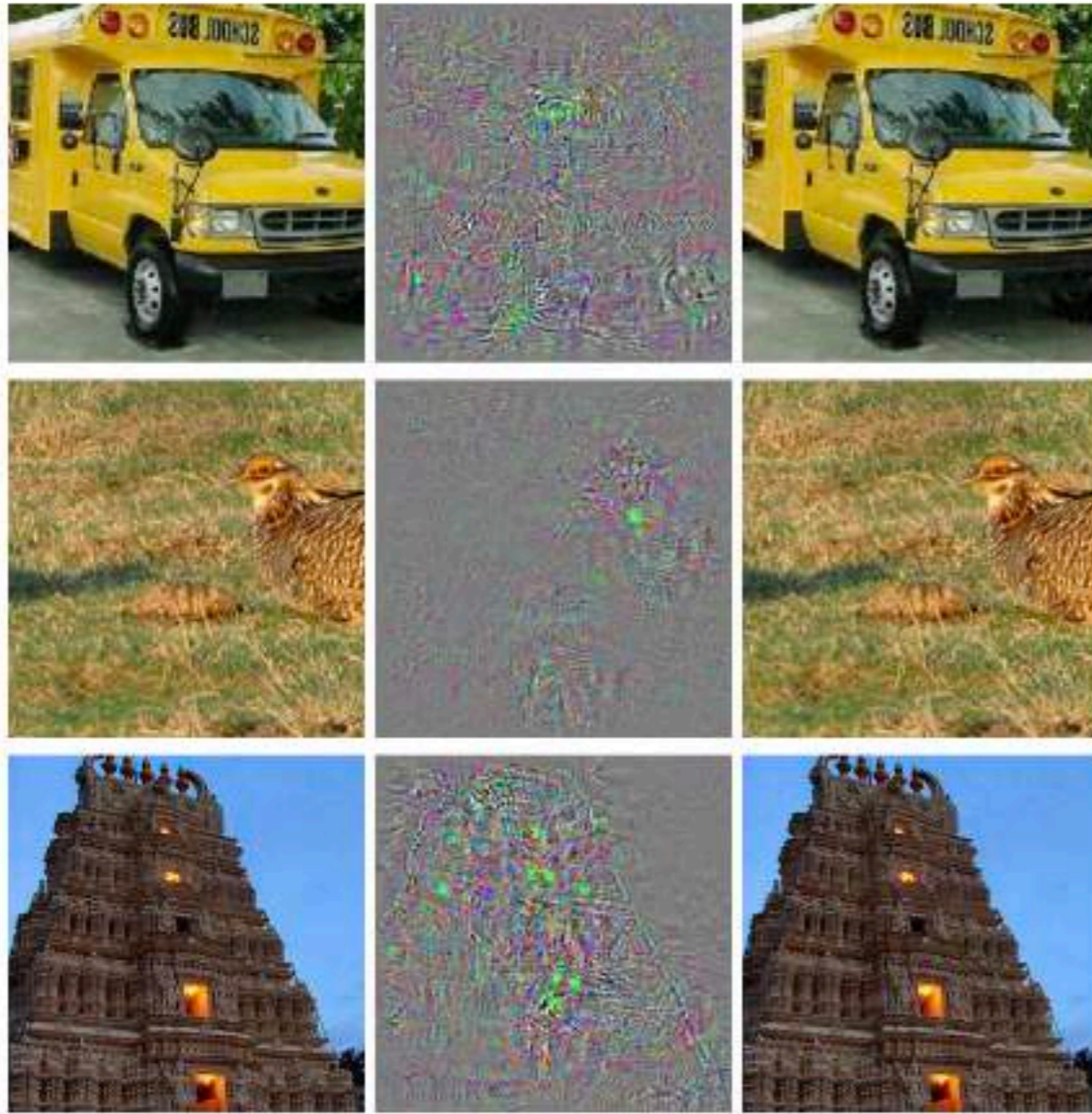
How did I do?

.....

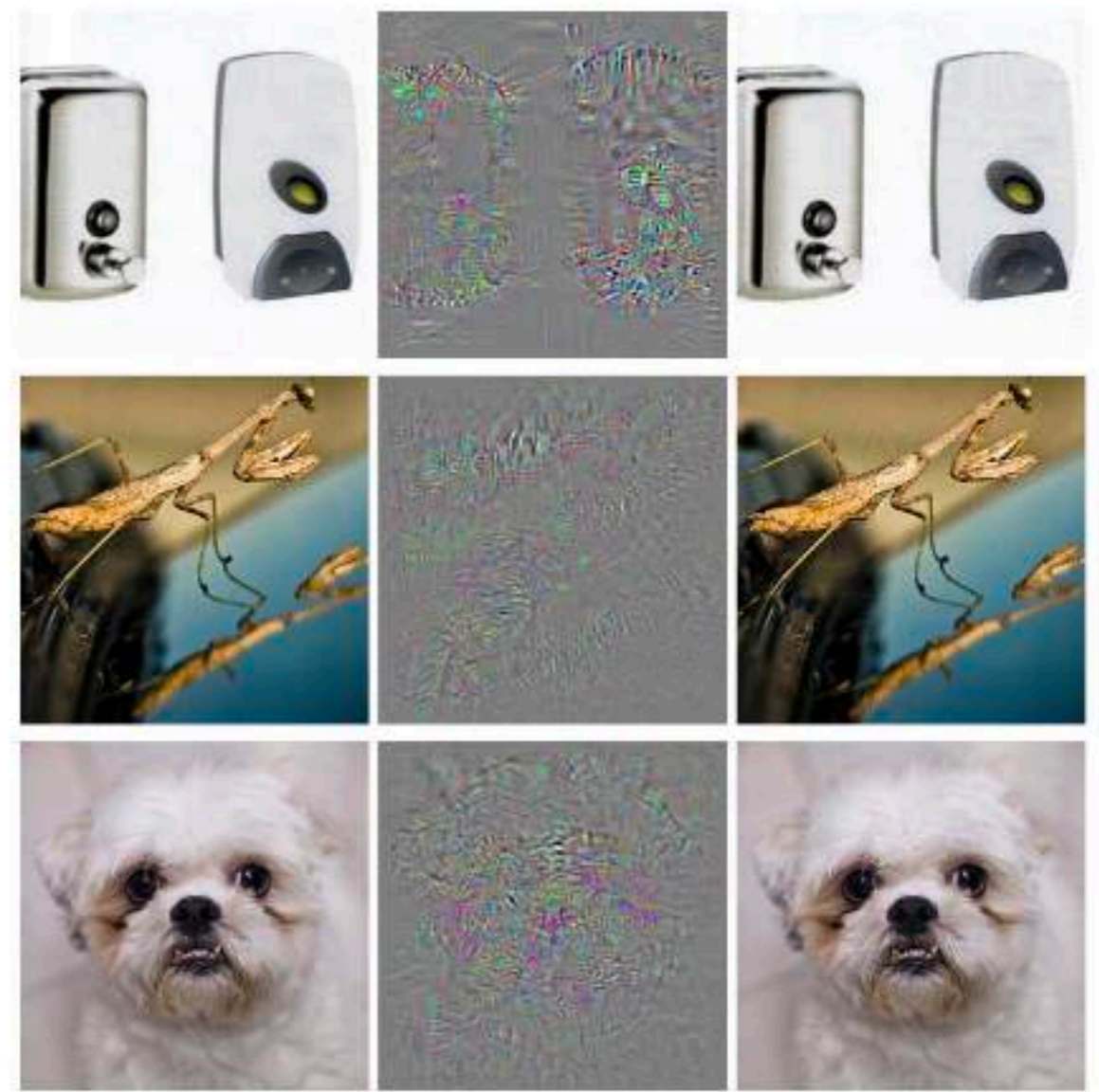
I think it's a group of people riding on the back of a bicycle.



Adversarial filters: Adversarial examples generated for AlexNet



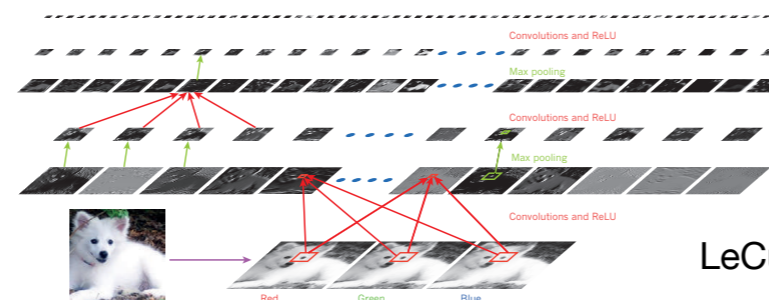
(a)



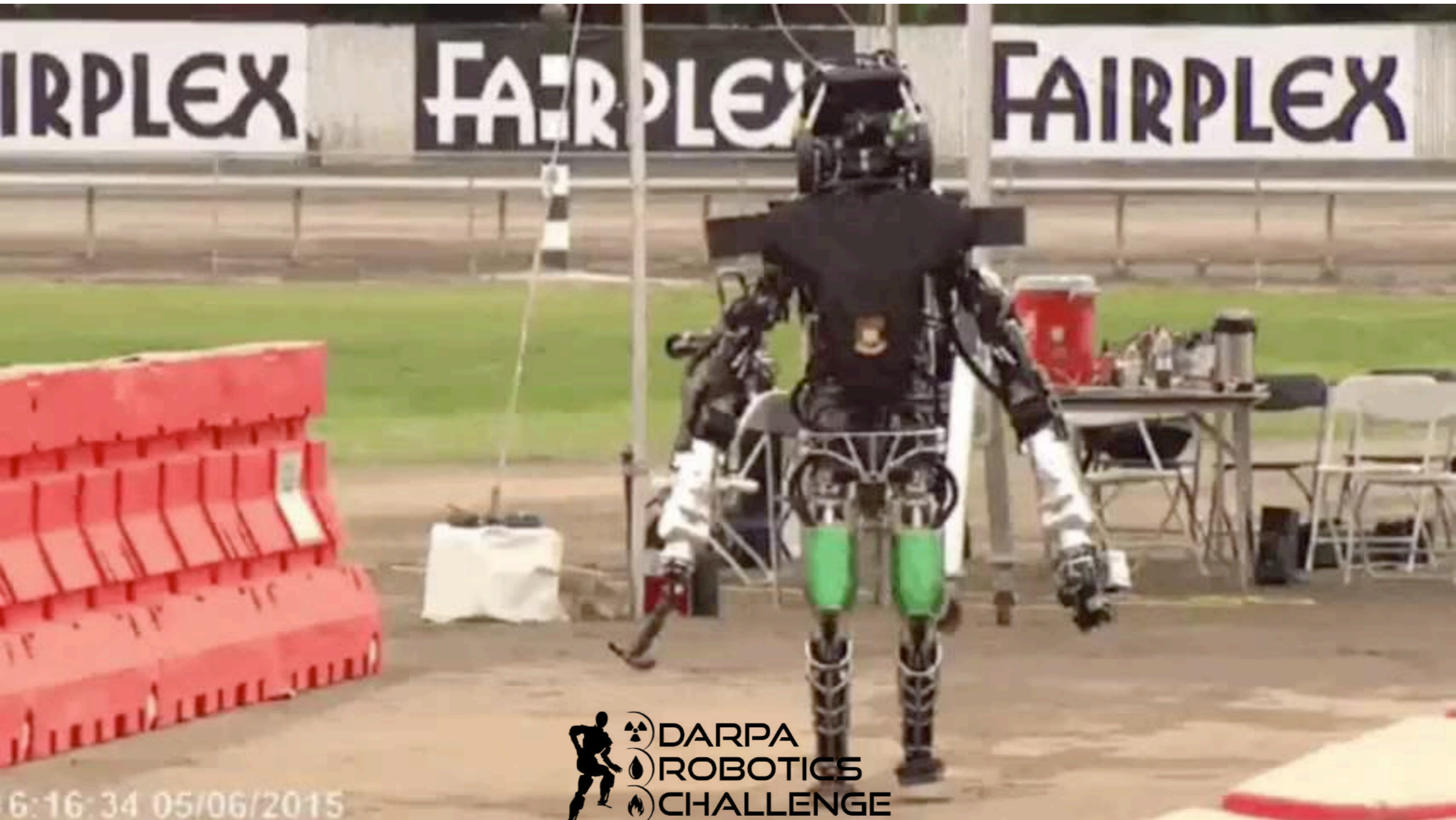
(b)

Szegedy et al 2014; <https://arxiv.org/pdf/1312.6199.pdf>

“ostrich, *Struthio camelus*”



Reality



6:16:34 05/06/2015
specs-lab.com

 DARPA
ROBOTICS
CHALLENGE
FINALS 2015

Paul Verschure

Piekniewski's blog

On limits of deep learning and where to go next with AI.

[MAIN SITE](#) [BLOG HOME](#) [ABOUT](#)

AI Winter Is Well On Its Way

POSTED [MAY 28, 2018](#) BY [FILIP PIEKNIIEWSKI](#)



Deep learning has been at the forefront of the so called AI revolution for quite a few years now, and many people had believed that it is the silver bullet that will take us to the world of wonders of technological singularity (general AI). Many bets were made in 2014, 2015 and 2016 when still new boundaries were pushed, such as the Alpha Go etc.

Basic info

I'm proposing a new machine learning meta-architecture for learning forward models. The architecture is called [Predictive Vision Model \(PVM\)](#). In this blog I present my thoughts on how PVM relates to deep learning and the global AI landscape. Occasionally I'll blog about my sysadmin projects and sci-fi.

[See a short bio note here.](#)

Science

[Home](#) [News](#) [Journals](#) [Topics](#) [Careers](#)

[Log in](#) | [My account](#)

SHARE **IN DEPTH** COMPUTER SCIENCE



Has artificial intelligence become alchemy?

Matthew Hutson

[+ See all authors and affiliations](#)

Science 04 May 2018:
Vol. 360, Issue 6388, pp. 478
DOI: 10.1126/science.360.6388.478

Article

[Figures & Data](#)

[Info & Metrics](#)

[eLetters](#)

[PDF](#)

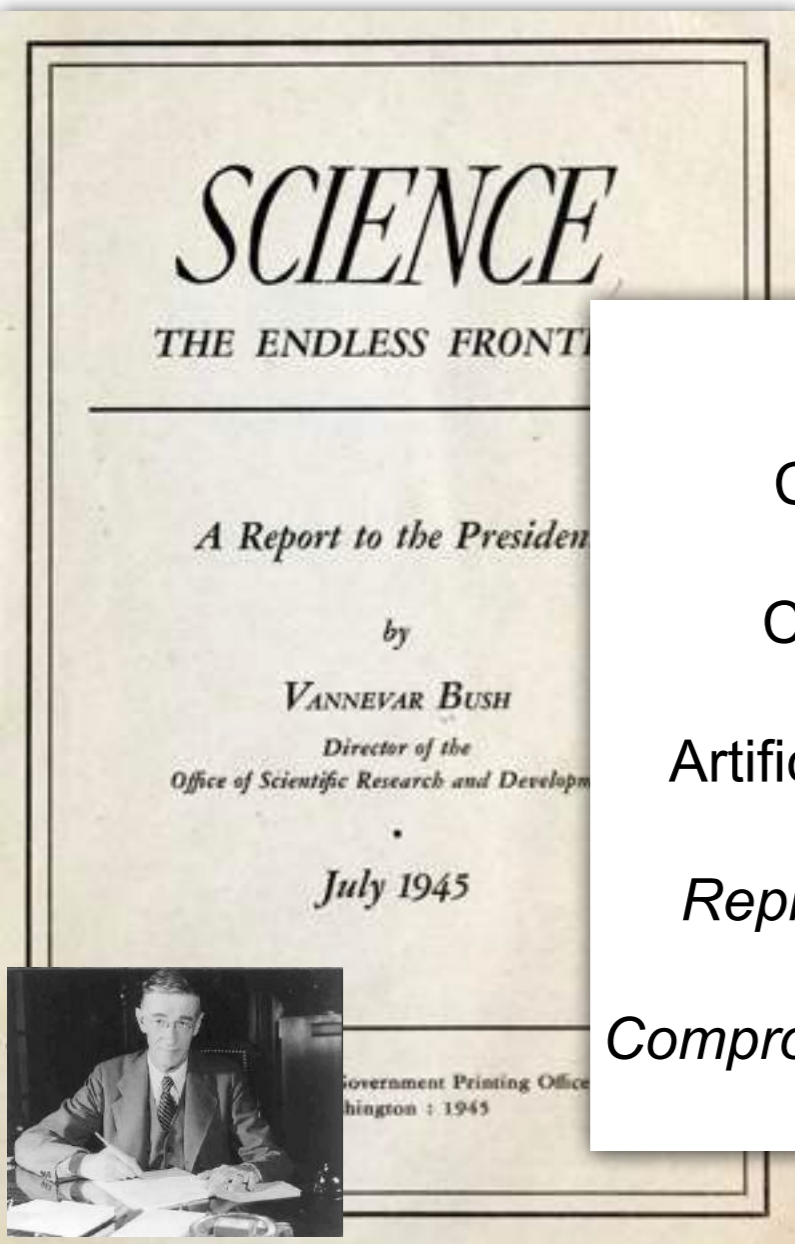
You are currently viewing the summary.

[View Full Text](#)

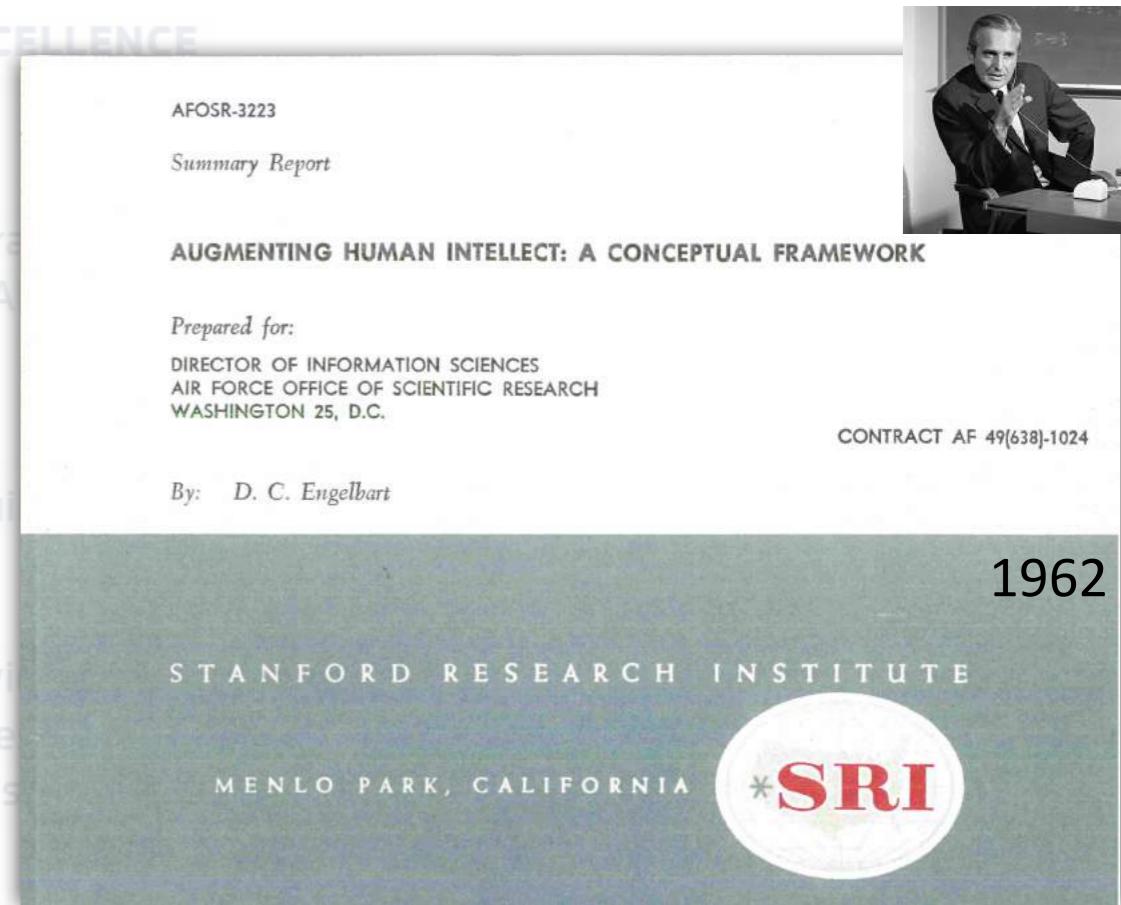
Summary

Ali Rahimi, a researcher in artificial intelligence (AI) at Google in San Francisco, California, has charged that machine learning algorithms, in which computers learn through trial and error, have become a form of "alchemy." Researchers, he says, do not know why some algorithms work and others don't, nor do they have rigorous criteria for choosing one AI architecture over another. Now, in a paper presented on 30 April at the International Conference on Learning Representations in Vancouver, Canada, Rahimi and his collaborators document examples of what they see as the alchemy problem and offer prescriptions for bolstering AI's rigor. The issue is distinct from AI's reproducibility problem, in which researchers can't replicate each other's results because of inconsistent experimental and publication practices. It also differs from the "black box" or

The Future: Explore the road not taken



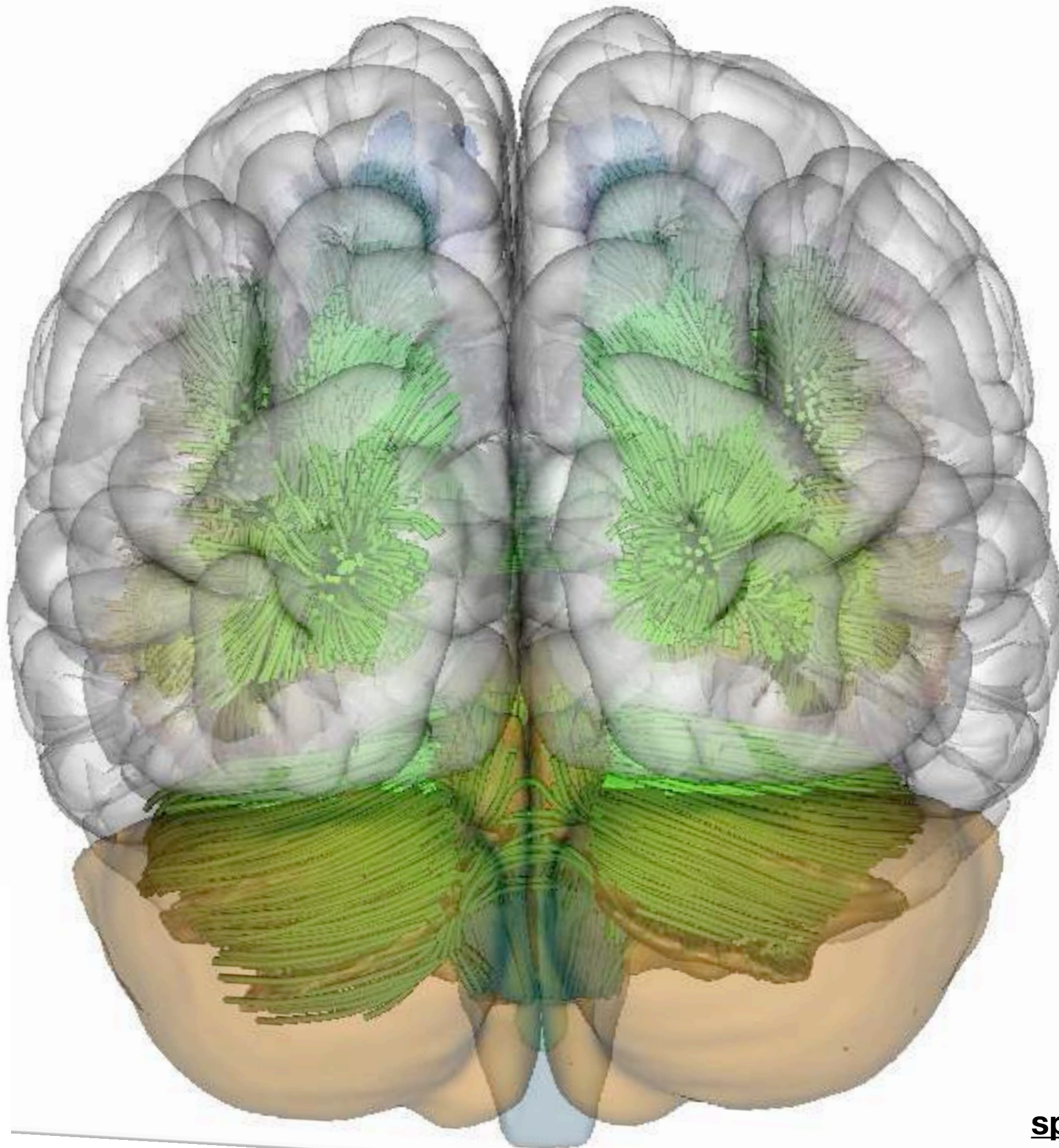
Control
-
Cybernetics
//
Computation
-
Artificial Intelligence
//
Replacing Humans
//
Compromising Wellbeing



“By "augmenting human intellect" we mean **increasing the capability** of a man **to approach a complex problem situation**, to gain **comprehension** to suit his particular needs, and **to derive solutions** to problems”

From AI to IA

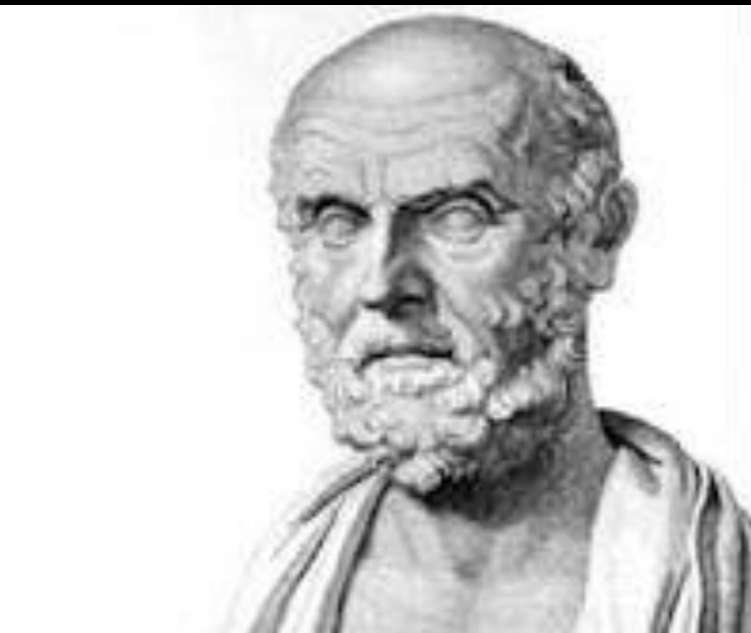




What does the
Brain
tell us about
Technology?

“Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant.”

Hippocrates of Kos (460-37 BC)



The Astonishing Hypothesis



Francis Crick (1916-2004)

Dismantling
the
Puppeteer



Mastering
Synthetic
Autonomy

Synthetic Volition

OXFORD

LIVING MACHINES

A handbook of research in biomimetic and biohybrid systems



EDITED BY

TONY J. PRESCOTT, NATHAN LEPORA, & PAUL F. M. J. VERSCHURE

