Zürcher Hochschule
für Angewandte Wissenschaften

Fundamentals of AI

Distinguished lecture, University of Engineering & Management, Kolkata
July 14, 2020

Thilo Stadelmann

What is AI?
Why is it hot?
How does it work?
And what’s the connection to a digitally transformed future?

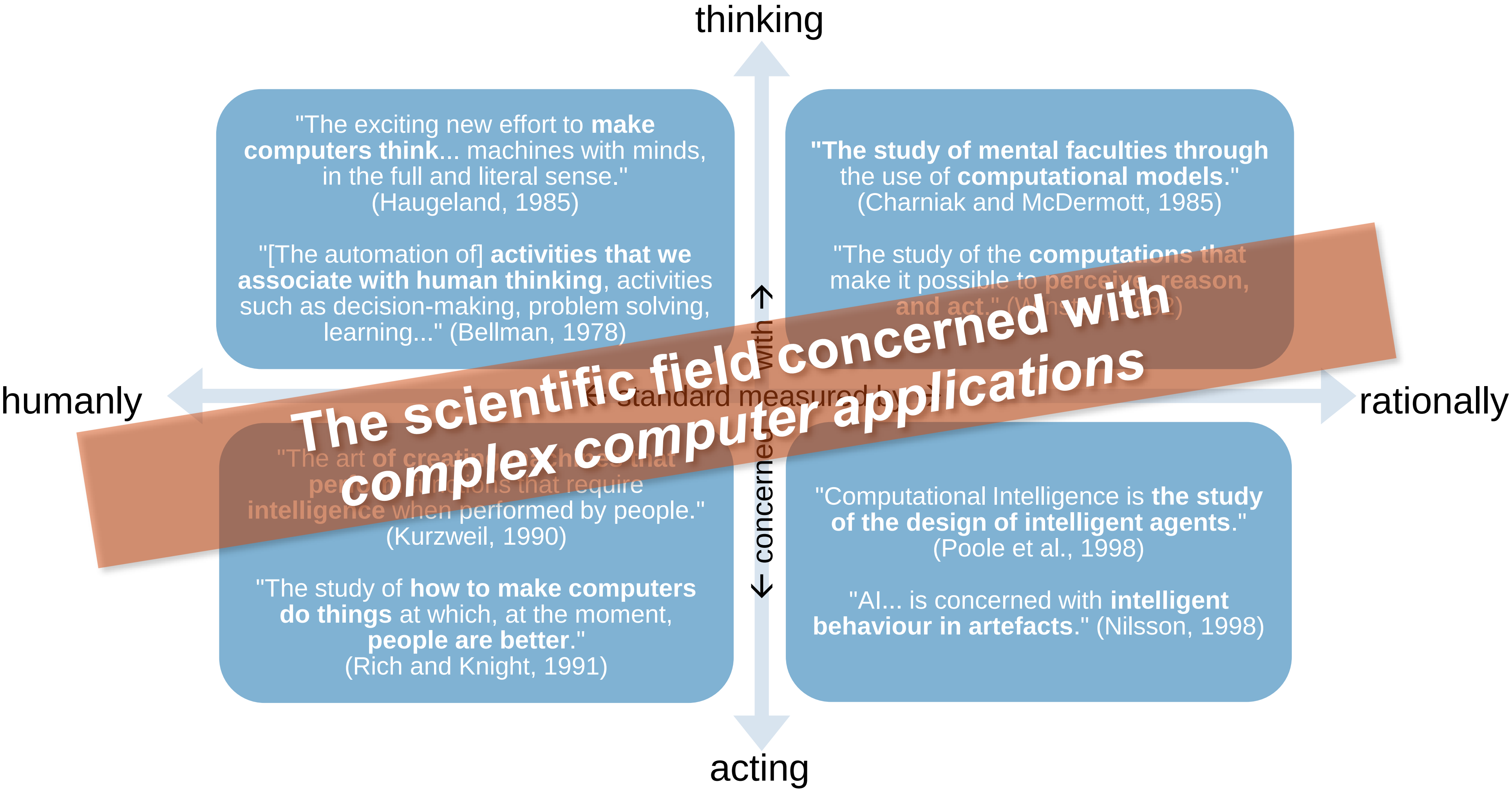
datalab
www.zhaw.ch/datalab

Source: https://www.softwareheritage.org/2018/01/08/yearly-anniversary-report/

1

What is AI?

What is AI?



thinking

"The exciting new effort to **make computers think**... machines with minds, in the full and literal sense." (Haugeland, 1985)

"[The automation of] **activities that we associate with human thinking**, activities such as decision-making, problem solving, learning..." (Bellman, 1978)

"The study of mental faculties through the use of **computational models**." (Charniak and McDermott, 1985)

"The study of the **computations that** make it possible to **perceive, reason, and act**." (Winston, 1992)

humanly ← standard measured by → rationally

← concerned with →

"The art **of creating machines that perform functions** that require **intelligence** when performed by people." (Kurzweil, 1990)

"The study of **how to make computers do things** at which, at the moment, **people are better**." (Rich and Knight, 1991)

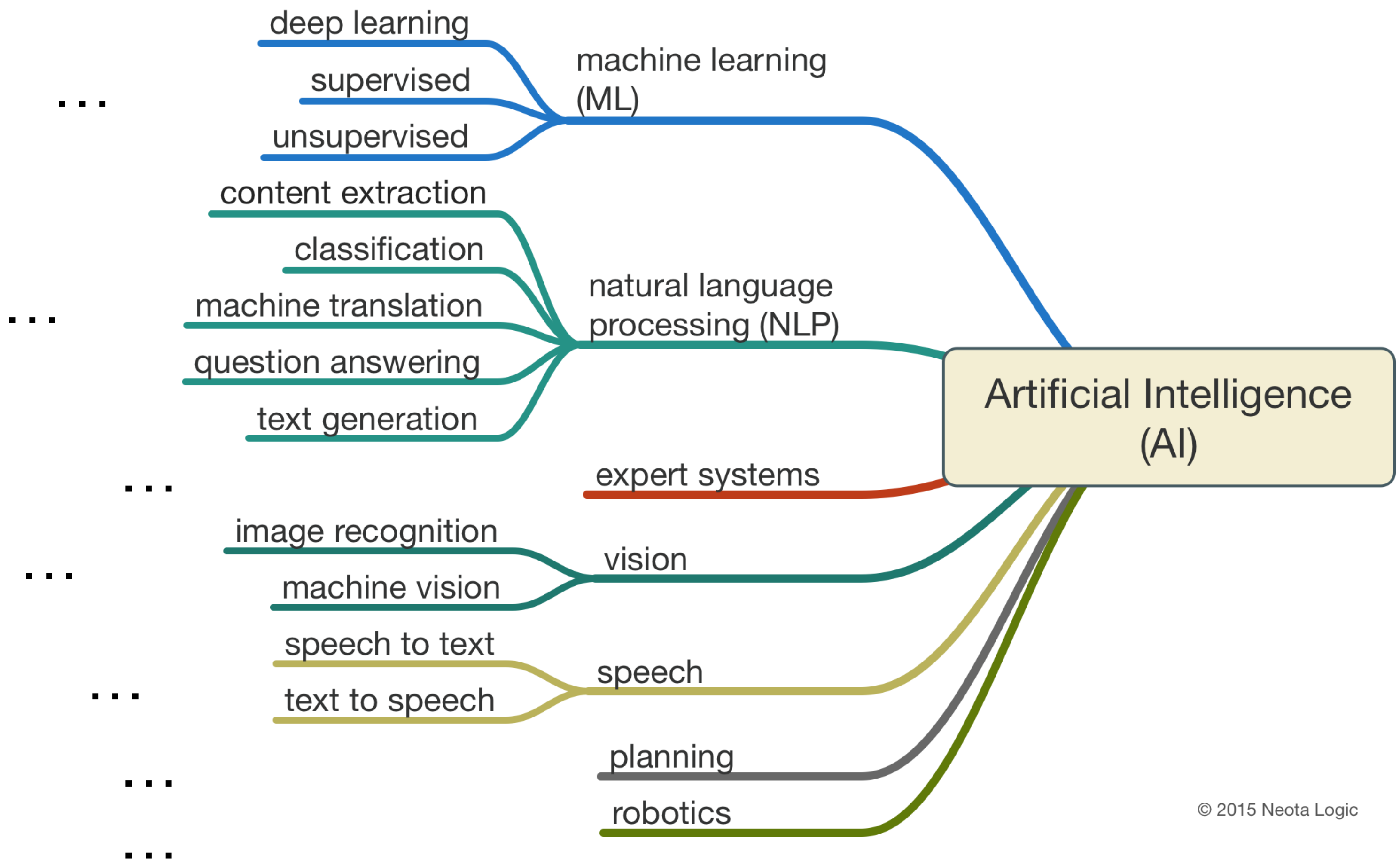
"Computational Intelligence is **the study of the design of intelligent agents**." (Poole et al., 1998)

"AI... is concerned with **intelligent behaviour in artefacts**." (Nilsson, 1998)

acting

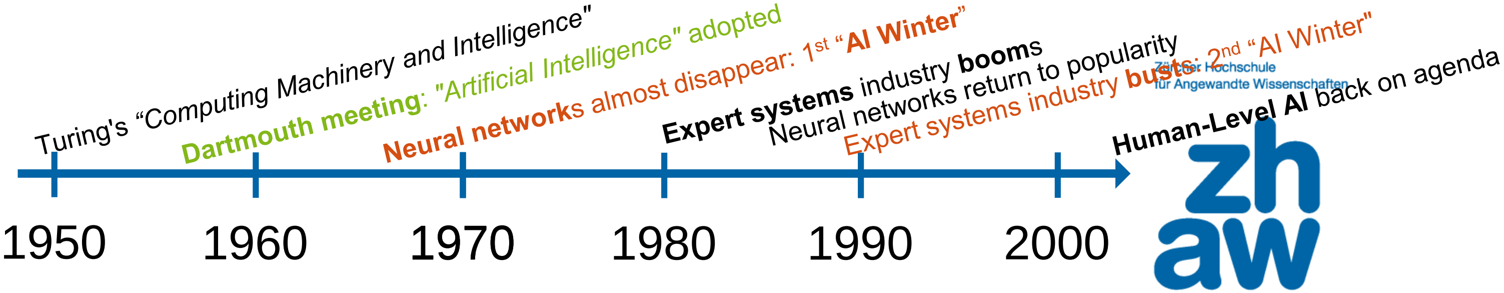
The scientific field concerned with complex computer applications

What belongs to AI?



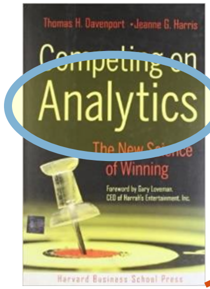
© 2015 Neota Logic

AI in context

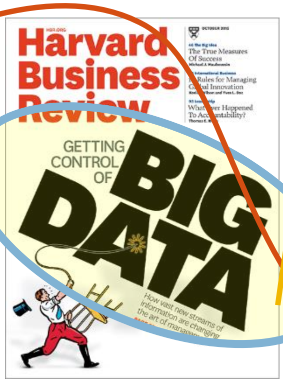


Turing's "Computing Machinery and Intelligence"
Dartmouth meeting: "Artificial Intelligence" adopted
Neural networks almost disappear: 1st "AI Winter"
Expert systems industry booms
Neural networks return to popularity
Expert systems industry busts: 2nd "AI Winter"
Human-Level AI back on agenda
1950
1960
1970
1980
1990
2000

Zürcher Hochschule für Angewandte Wissenschaften



Thomas H. Davenport · Jeanne G. Harris
Competing on Analytics
The New Science of Winning
Harvard Business School Press

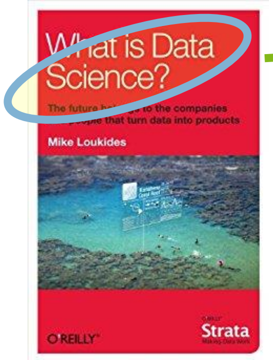


…

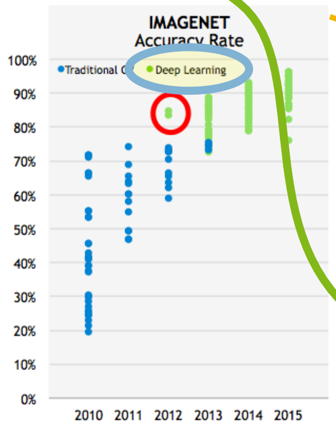
2007

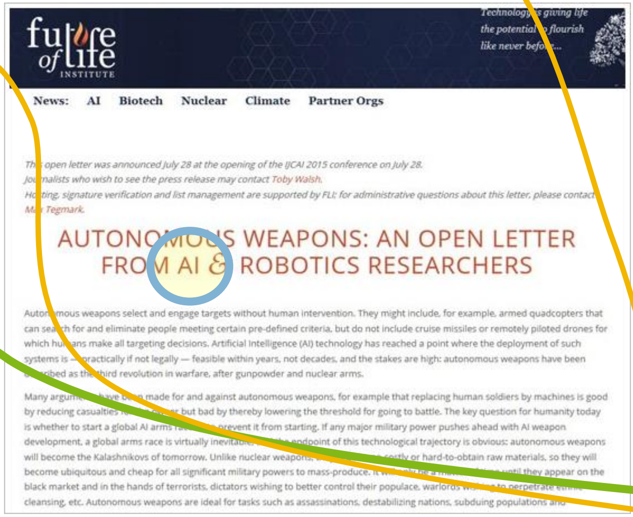
2012

2016



What is Data Science?
The future belongs to the companies and people that turn data into products
Mike Loukides
O'REILLY
Strata





future of life INSTITUTE
Technology is giving life the potential to flourish like never before...

News: AI Biotech Nuclear Climate Partner Orgs

This open letter was announced July 28 at the opening of the IJCAI 2015 conference on July 28.
Journalists who wish to see the press release may contact Toby Walsh.
Hosting, signature verification and list management are supported by FLI; for administrative questions about this letter, please contact Max Tegmark.

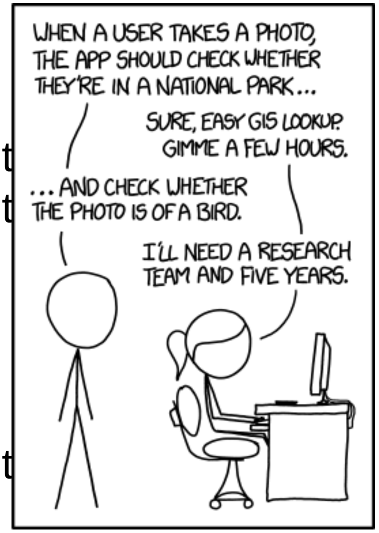
AUTONOMOUS WEAPONS: AN OPEN LETTER FROM AI & ROBOTICS RESEARCHERS

Autonomous weapons select and engage targets without human intervention. They might include, for example, armed quadcopters that can search for and eliminate people meeting certain pre-defined criteria, but do not include cruise missiles or remotely piloted drones for which humans make all targeting decisions. Artificial Intelligence (AI) technology has reached a point where the deployment of such systems is — practically if not legally — feasible within years, not decades, and the stakes are high: autonomous weapons have been described as the third revolution in warfare, after gunpowder and nuclear arms.

Many arguments have been made for and against autonomous weapons, for example that replacing human soldiers by machines is good by reducing casualties for the owner but bad by thereby lowering the threshold for going to battle. The key question for humanity today is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and

What can AI do today?

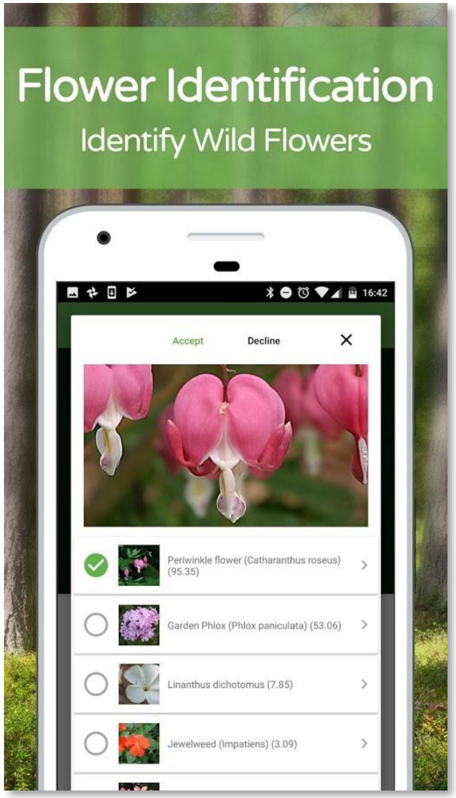
1.	Play a decent game of **table tennis**	ok
2.	**Drive** safely along a curving **mountain road**	ok
3.	Drive safely along **Technikumstrasse** Winterthur	ok (only since recently)
4.	**Buy** a week's worth of **groceries on the web**	ok
5.	Buy a week's worth of groceries **at Migros**	no
6.	**Play** a decent game of **bridge**	ok
7.	**Discover** and prove a new mathematical **theorem**	not complet
8.	**Design** and execute a **research program** in molecular biology	not complet
9.	Write an **intentionally funny** story	no
10.	Give competent **legal advice** in a specialized area of law	ok
11.	**Translate** spoken English **into spoken** Swedish in real time	ok
12.	**Converse** successfully with another person for an hour	no
13.	Perform a complex **surgical operation**	not complet
14.	**Unload** any **dishwasher** and put everything away	no
15.	Compete in the game show **Jeopardy!**	ok
16.	**Write clickbait** articles fully automatized	ok


WHEN A USER TAKES A PHOTO, THE APP SHOULD CHECK WHETHER THEY'RE IN A NATIONAL PARK...
SURE, EASY GIS LOOKUP. GIMME A FEW HOURS.
...AND CHECK WHETHER THE PHOTO IS OF A BIRD.
I'LL NEED A RESEARCH TEAM AND FIVE YEARS.

IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Example: Feasible vs. dangerous

Technology: Computer Vision with Deep Learning

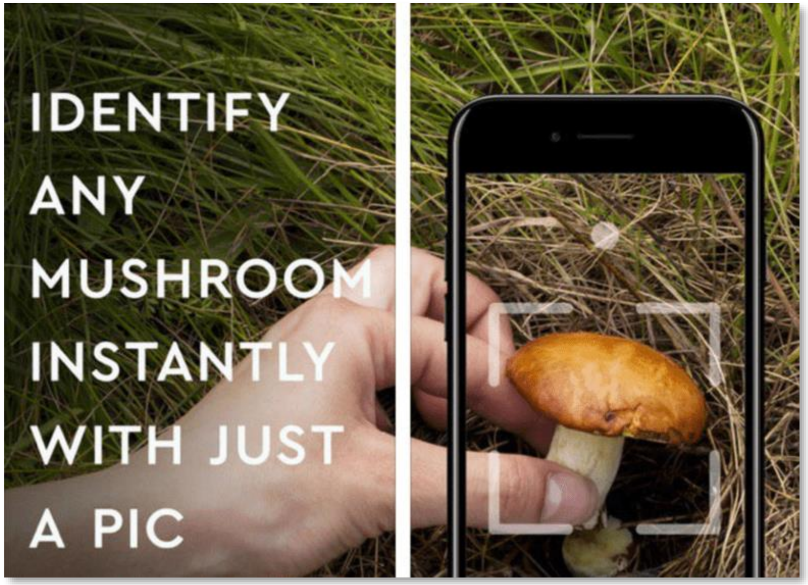


● Periwinkle flower (Catharanthus roseus) (95.35)

○ Garden Phlox (Phlox paniculata) (53.06)

○ Linanthus dichotomus (7.85)

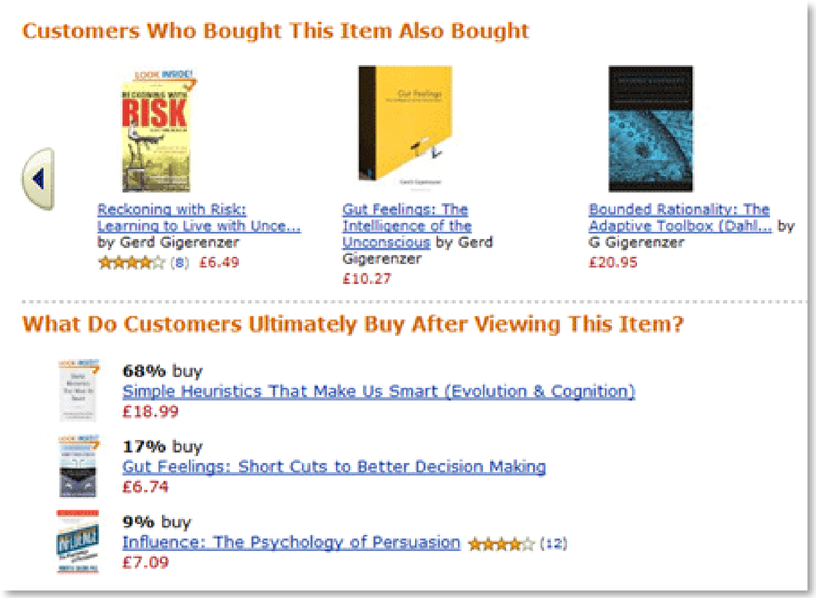
○ Jewelweed (Impatiens) (3.09)



https://www.cultofmac.com/495088/avoid-potentially-deadly-ai-app/

Example: Commercial success vs. regulation

Technology: Recommender Systems



Customers Who Bought This Item Also Bought

Reckoning with Risk: Learning to Live with Unce... by Gerd Gigerenzer ★★★★☆ (8) £6.49

Gut Feelings: The Intelligence of the Unconscious by Gerd Gigerenzer £10.27

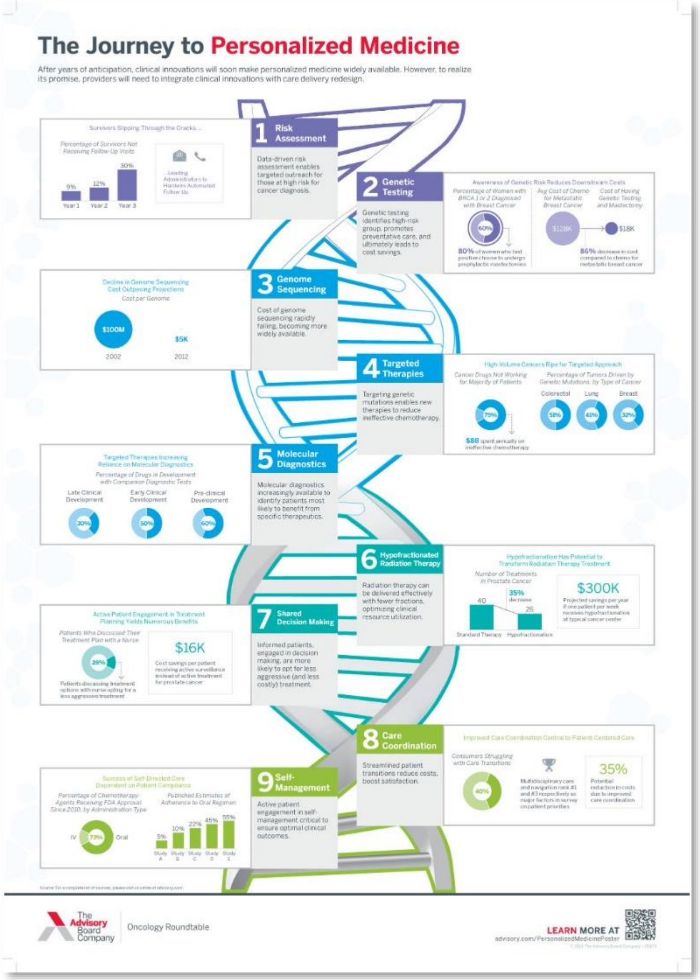
Bounded Rationality: The Adaptive Toolbox (Dahl... by G Gigerenzer £20.95

What Do Customers Ultimately Buy After Viewing This Item?

68% buy Simple Heuristics That Make Us Smart (Evolution & Cognition) £18.99

17% buy Gut Feelings: Short Cuts to Better Decision Making £6.74

9% buy Influence: The Psychology of Persuasion ★★★★☆ (12) £7.09



The Journey to Personalized Medicine

After years of anticipation, clinical innovations will soon make personalized medicine widely available. However, to realize its promise, providers will need to integrate clinical innovations with care delivery redesign.

1 Risk Assessment: Data-driven risk assessment enables targeted outreach for those at high risk for cancer diagnosis.

2 Genetic Testing: Genetic testing identifies high-risk group, promotes preventative care, and ultimately leads to cost savings.

3 Genome Sequencing: Cost of genome sequencing rapidly falling, becoming more widely available.

4 Targeted Therapies: Targeting genetic mutations enables new therapies to reduce ineffective chemotherapy.

5 Molecular Diagnostics: Molecular diagnostics increasingly available to identify patients most likely to benefit from specific therapeutics.

6 Hypofractionated Radiation Therapy: Radiation therapy can be delivered effectively with fewer fractions, optimizing clinical resource utilization.

7 Shared Decision Making: Informed patients, engaged in decision making, are more likely to opt for less aggressive (and less costly) treatment.

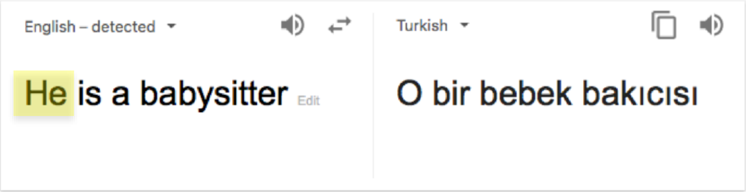
8 Care Coordination: Streamlined patient transitions reduce costs, boost satisfaction.

9 Self-Management: Active patient engagement in self-management critical to ensure optimal clinical outcomes.

Survivors Slipping Through the Cracks...
Percentage of Survivors Not Receiving Follow-Up Visits
9% Year 1
12% Year 2
30% Year 3
...Leading Administrators to Hardwire Automated Follow-Up
Awareness of Genetic Risk Reduces Downstream Costs
Percentage of Women with BRCA 1 or 2 Diagnosed with Breast Cancer
60%
80% of women who test positive choose to undergo prophylactic mastectomies
Avg Cost of Chemo for Metastatic Breast Cancer
Cost of Having Genetic Testing and Mastectomy
$128K
$18K
86% decrease in cost compared to chemo for metastatic breast cancer
Decline in Genome Sequencing Cost Outpacing Projections
Cost per Genome
$100M 2002
$5K 2012
High-Volume Cancers Ripe for Targeted Approach
Cancer Drugs Not Working for Majority of Patients
75%
$88 spent annually on ineffective chemotherapy
Percentage of Tumors Driven by Genetic Mutations, by Type of Cancer
Colorectal 51%
Lung 41%
Breast 32%
Targeted Therapies Increasing Reliance on Molecular Diagnostics
Percentage of Drugs in Development with Companion Diagnostic Tests
Late Clinical Development 30%
Early Clinical Development 50%
Pre-clinical Development 60%
Hypofractionation Has Potential to Transform Radiation Therapy Treatment
Number of Treatments in Prostate Cancer
40 Standard Therapy
26 Hypofractionation
35% decrease
$300K
Projected savings per year if one patient per week receives hypofractionation at typical cancer center
Active Patient Engagement in Treatment Planning Yields Numerous Benefits
Patients Who Discussed Their Treatment Plan with a Nurse
28%
Patients discussing treatment options with nurse opting for a less aggressive treatment
$16K
Cost savings per patient receiving active surveillance instead of active treatment for prostate cancer
Improved Care Coordination Central to Patient-Centered Care
Consumers Struggling with Care Transitions
40%
Multidisciplinary care and navigation rank #1 and #3 respectively as major factors in survey on patient priorities
35%
Potential reduction in costs due to improved care coordination
Success of Self-Directed Care Dependent on Patient Compliance
Percentage of Chemotherapy Agents Receiving FDA Approval Since 2010, by Administration Type
IV 73% Oral
Published Estimates of Adherence to Oral Regimen
5% Study A
10% Study B
22% Study C
45% Study D
55% Study E
The Advisory Board Company
Oncology Roundtable
LEARN MORE AT advisory.com/PersonalizedMedicinePoster

Example: Statistics vs. bias

Technology: Machine Learning

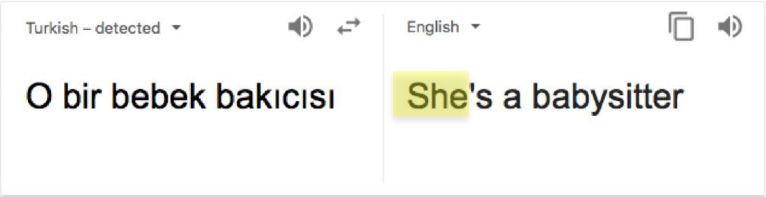


English – detected

He is a babysitter

Turkish

O bir bebek bakıcısı

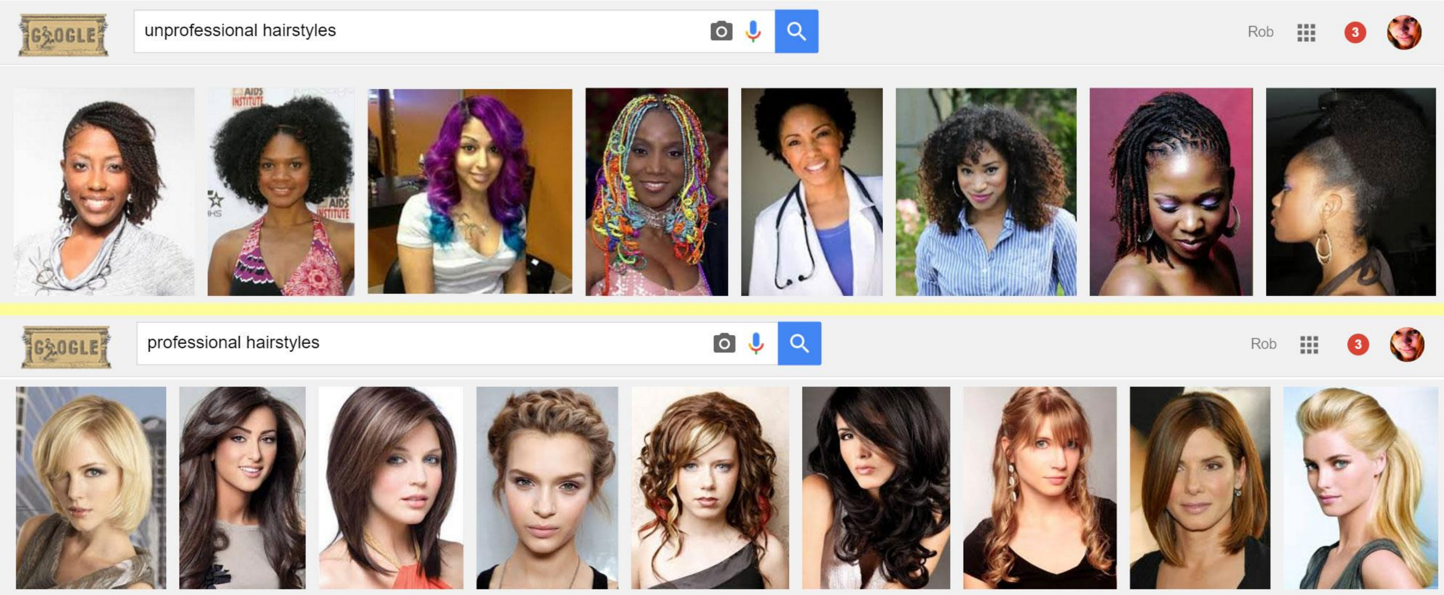


Turkish – detected

O bir bebek bakıcısı

English

She's a babysitter



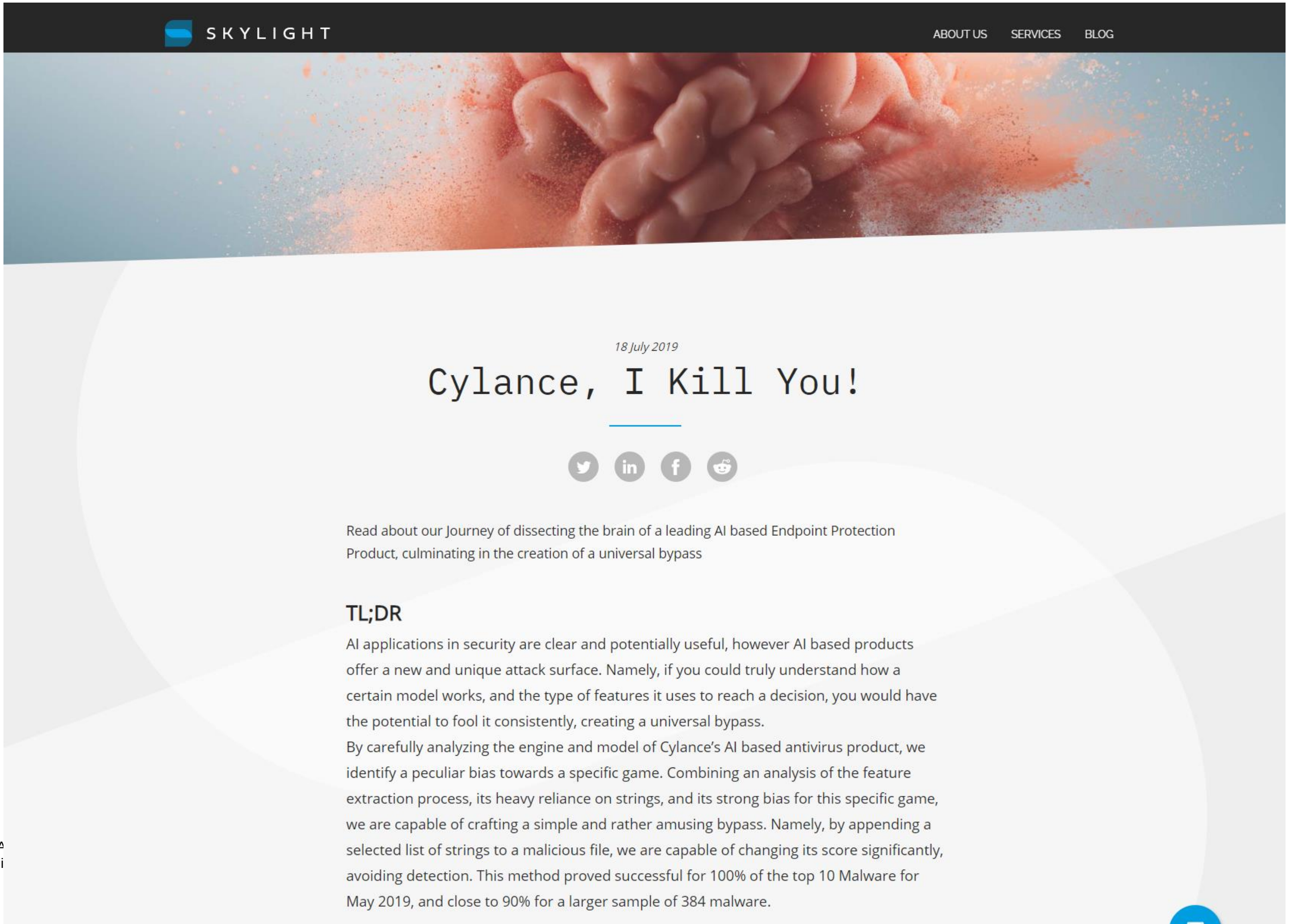
unprofessional hairstyles

professional hairstyles

See also: Nassim Nicholas Talib, «*The Black Swan: The Impact of the Highly Improbable*», 2007

Example: artificial intelligence vs. natural stupidity

Technology: Machine Learning with downstream rules



18 July 2019

Cylance, I Kill You!

Read about our Journey of dissecting the brain of a leading AI based Endpoint Protection Product, culminating in the creation of a universal bypass

TL;DR

AI applications in security are clear and potentially useful, however AI based products offer a new and unique attack surface. Namely, if you could truly understand how a certain model works, and the type of features it uses to reach a decision, you would have the potential to fool it consistently, creating a universal bypass.

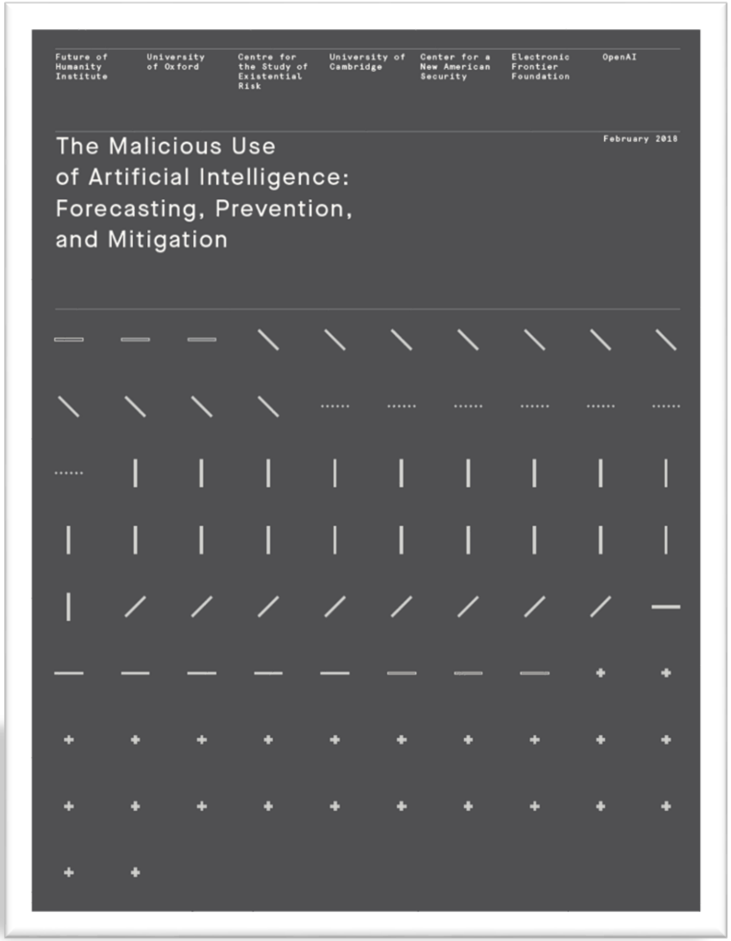
By carefully analyzing the engine and model of Cylance's AI based antivirus product, we identify a peculiar bias towards a specific game. Combining an analysis of the feature extraction process, its heavy reliance on strings, and its strong bias for this specific game, we are capable of crafting a simple and rather amusing bypass. Namely, by appending a selected list of strings to a malicious file, we are capable of changing its score significantly, avoiding detection. This method proved successful for 100% of the top 10 Malware for May 2019, and close to 90% for a larger sample of 384 malware.

Risks through AI?

- AI per definition is a “**dual use technology**”
  → see report by Brundage et al., 2018

- But: “**natural stupidity**” is the more imminent threat

- **AI ethics** and explainable AI became mainstream and hot research topics in the recent years – not because of intolerable risks, but because of:



Future of Humanity Institute

University of Oxford

Centre for the Study of Existential Risk

University of Cambridge

Center for a New American Security

Electronic Frontier Foundation

OpenAI

February 2018

The Malicious Use
of Artificial Intelligence:
Forecasting, Prevention,
and Mitigation

What→ Why?→ How?→ Future

2

Why is it hot currently?
(A short history of recent years)

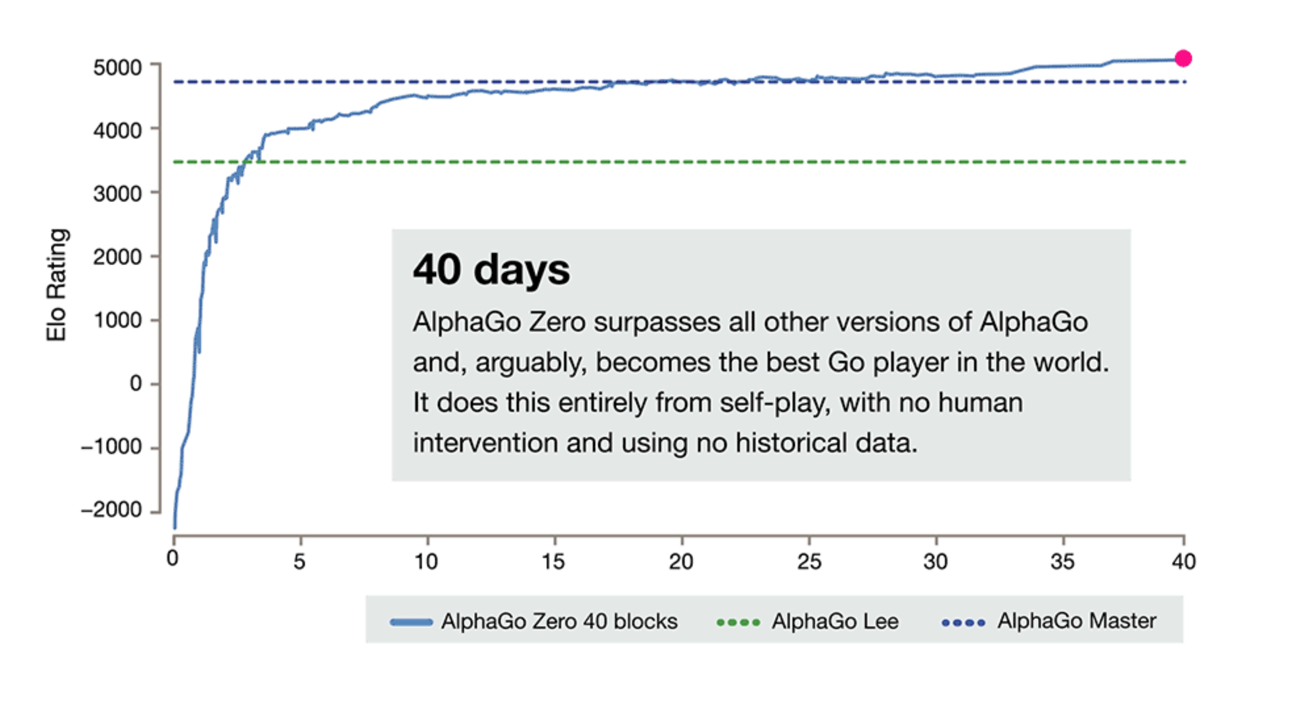
Google Acquires Artificial Intelligence Startup DeepMind For More Than $500M

Posted Jan 26, 2014 by Catherine Shu (@catherineshu)

Google will bu
reports that th
in talks to buy
couldn't disclose deal terms.

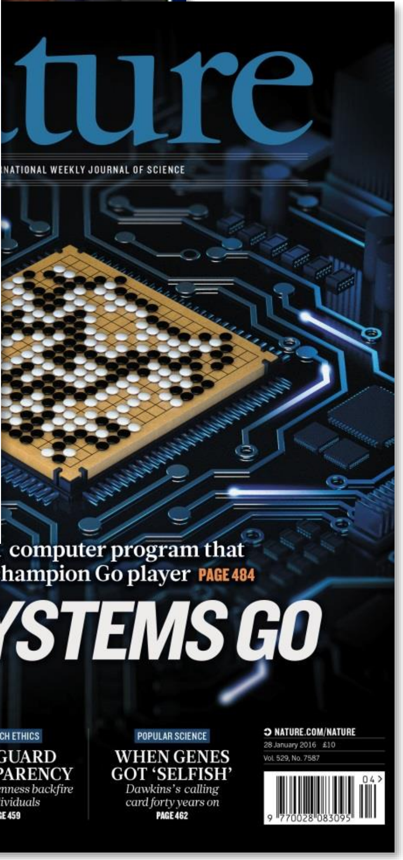
The acquisition was originally confirmed by Google to Re/code.

AlphaGo
Google DeepMind



40 days

AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.



ture

NATIONAL WEEKLY JOURNAL OF SCIENCE

At last — a computer program that can beat a champion Go player PAGE 484

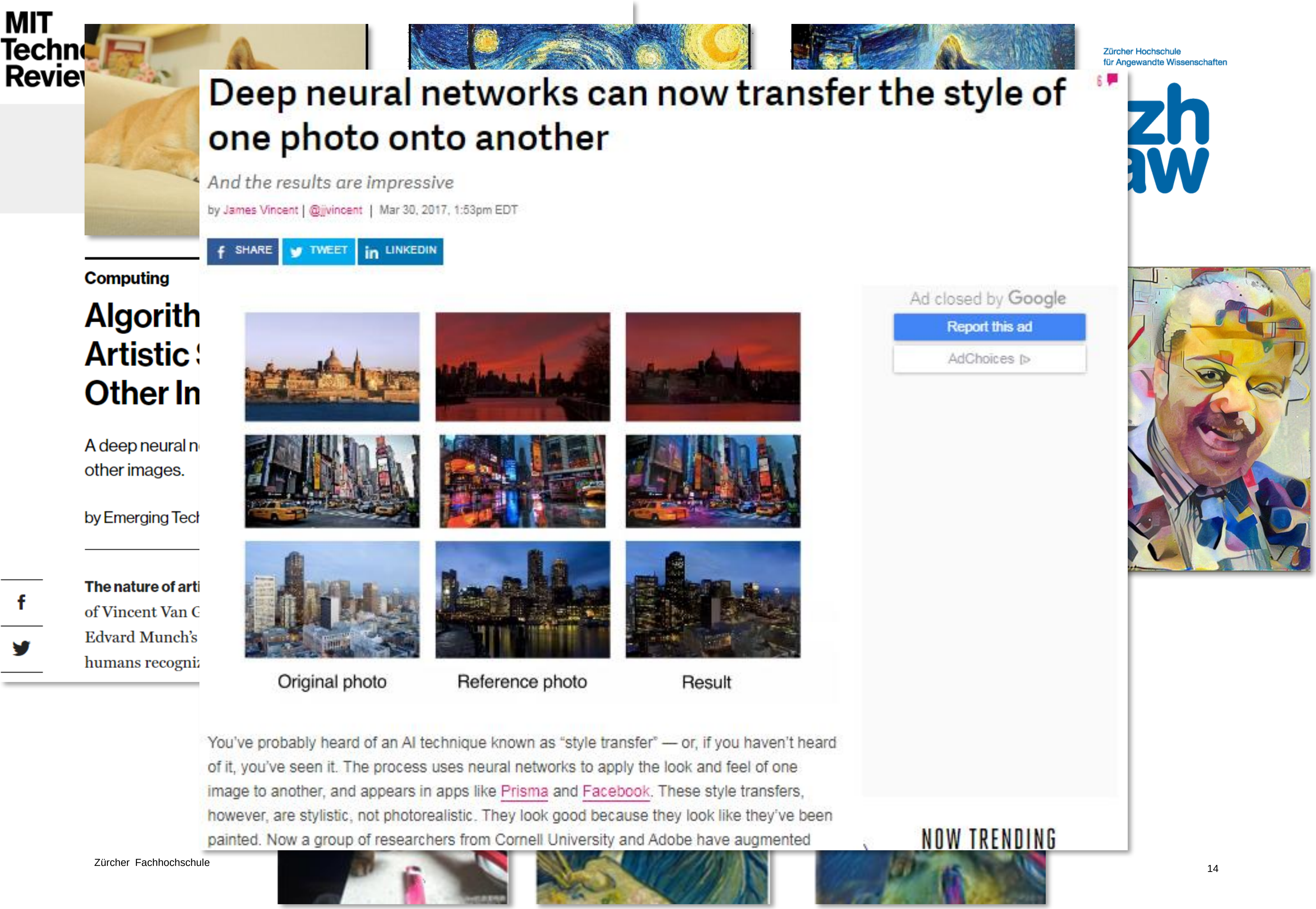
ALL SYSTEMS GO

CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452

RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459

POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462

NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587



MIT Techn Revie

Computing

Algorith Artistic Other In

A deep neural n other images.

by Emerging Tech

The nature of arti of Vincent Van G Edvard Munch's humans recogni

Deep neural networks can now transfer the style of one photo onto another

And the results are impressive

by James Vincent | @jjvincent | Mar 30, 2017, 1:53pm EDT

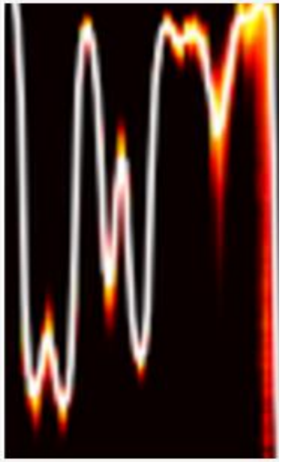
You've probably heard of an AI technique known as "style transfer" — or, if you haven't heard of it, you've seen it. The process uses neural networks to apply the look and feel of one image to another, and appears in apps like Prisma and Facebook. These style transfers, however, are stylistic, not photorealistic. They look good because they look like they've been painted. Now a group of researchers from Cornell University and Adobe have augmented

Künstliche Intelligenz

WaveNet lässt Computersprache natürlich klingen

von Henning Steier / 12.9.201

Die Google-Tochter DeepM
macht auch Musik.

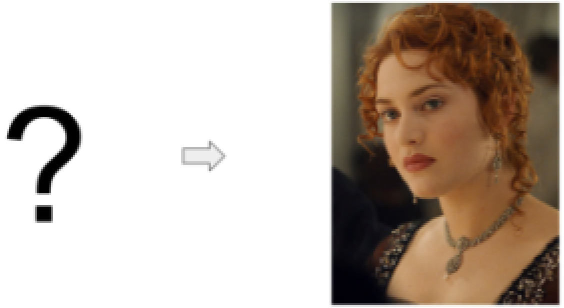


DeepMind lässt WaveNet Spra

Die Google-Tochter Deep
Spiel «Go» Schlagzeilen:
einen der besten mensch
Londoner Unternehmen
erzeugt Sprache, die seh
im Blogeintrag des Unter
Massstab nimmt. Man ha

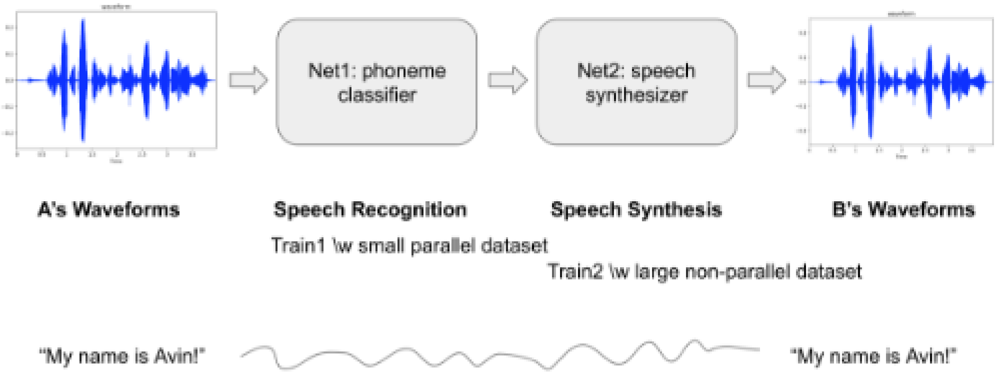
Intro

What if you could imitate a famous celebrity's voice or sing like a famous singer? This project started with a goal to convert someone's voice to a specific target voice. So called, it's voice style transfer. We worked on this project that aims to convert someone's voice to a famous English actress Kate Winslet's voice. We implemented a deep neural networks to achieve that and more than 2 hours of audio book sentences read by Kate Winslet are used as a dataset.



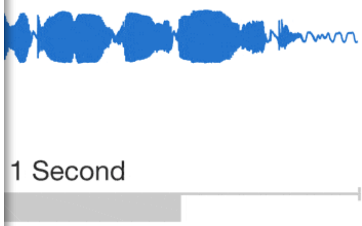
Model Architecture

This is a many-to-one voice conversion system. The main significance of this work is that we could generate a target speaker's utterances without parallel data like <source's wav, target's wav>, <wav, text> or <wav, phone>, but only waveforms of the target speaker. (To make these parallel datasets needs a lot of effort.) All we need in this project is a number of waveforms of the target speaker's utterances and only a small set of <wav, phone> pairs from a number of anonymous speakers.

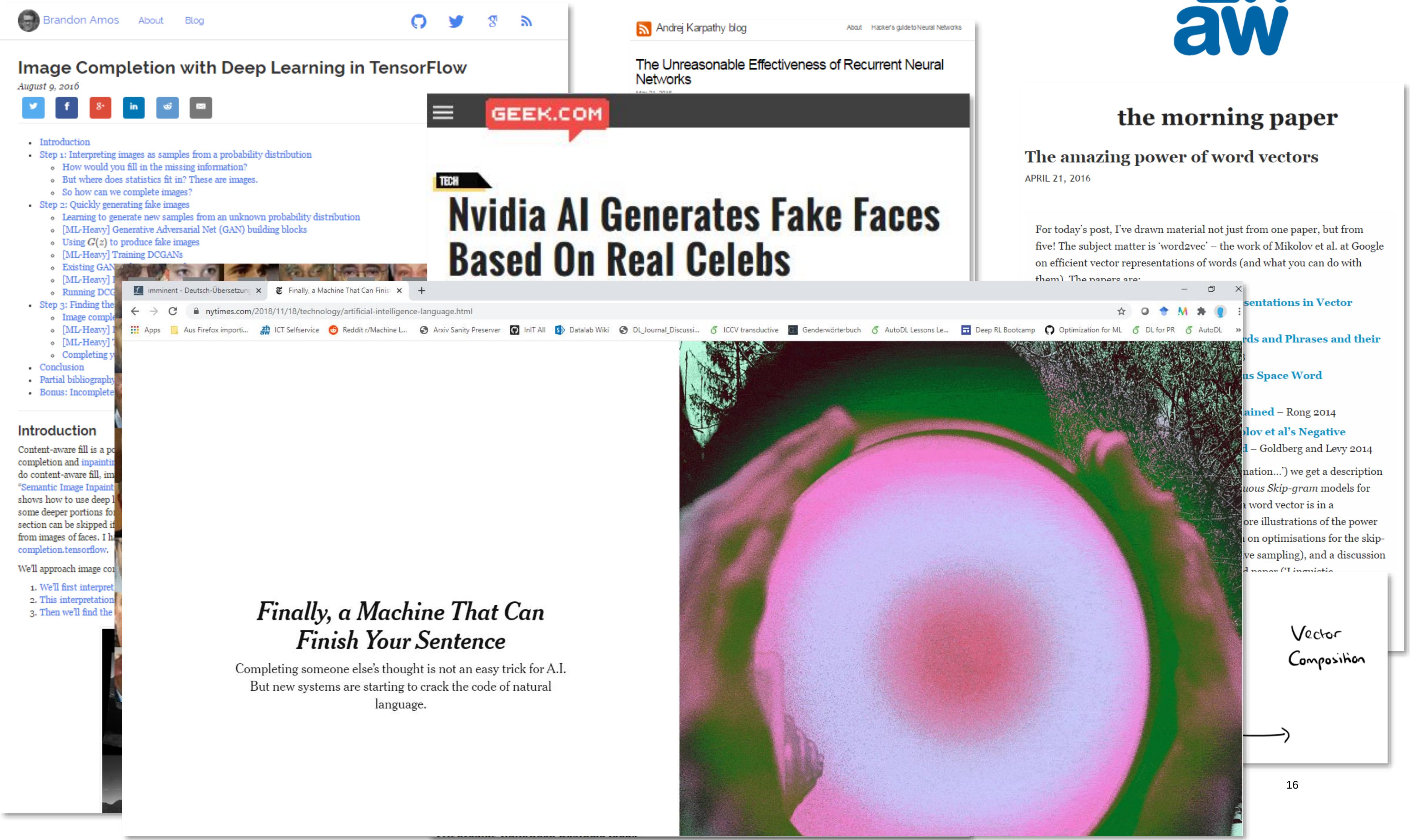


nerierte Sprache
us Texteingabe»

nerierte Musik
hne Inhaltsvorgabe»



…and more!



Brandon Amos About Blog

Image Completion with Deep Learning in TensorFlow

August 9, 2016

- Introduction
- Step 1: Interpreting images as samples from a probability distribution
  - How would you fill in the missing information?
  - But where does statistics fit in? These are images.
  - So how can we complete images?
- Step 2: Quickly generating fake images
  - Learning to generate new samples from an unknown probability distribution
  - [ML-Heavy] Generative Adversarial Net (GAN) building blocks
  - Using $G(z)$ to produce fake images
  - [ML-Heavy] Training DCGANs
  - Existing GAN
  - [ML-Heavy]
  - Running DCG
- Step 3: Finding the
  - Image compl
  - [ML-Heavy]
  - [ML-Heavy]
  - Completing y
- Conclusion
- Partial bibliography
- Bonus: Incomplete

Introduction

Content-aware fill is a p
completion and inpaintin
do content-aware fill, im
"Semantic Image Inpaint
shows how to use deep l
some deeper portions fo
section can be skipped i
from images of faces. I h
completion.tensorflow.

We'll approach image co

1. We'll first interpret
2. This interpretation
3. Then we'll find the

Andrej Karpathy blog About Hacker's guide to Neural Networks

The Unreasonable Effectiveness of Recurrent Neural Networks

GEEK.COM

TECH

Nvidia AI Generates Fake Faces Based On Real Celebs

the morning paper

The amazing power of word vectors

APRIL 21, 2016

For today's post, I've drawn material not just from one paper, but from five! The subject matter is 'word2vec' – the work of Mikolov et al. at Google on efficient vector representations of words (and what you can do with them). The papers are:

sentations in Vector

rds and Phrases and their

us Space Word

ained – Rong 2014

lov et al's Negative

d – Goldberg and Levy 2014

nation…') we get a description
uous *Skip-gram* models for
a word vector is in a
ore illustrations of the power
on optimisations for the skip-
ive sampling), and a discussion

Finally, a Machine That Can Finish Your Sentence

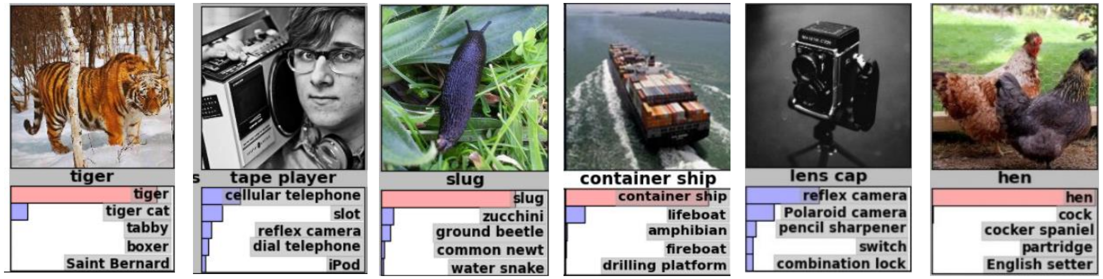
Completing someone else's thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.

Vector
Composition

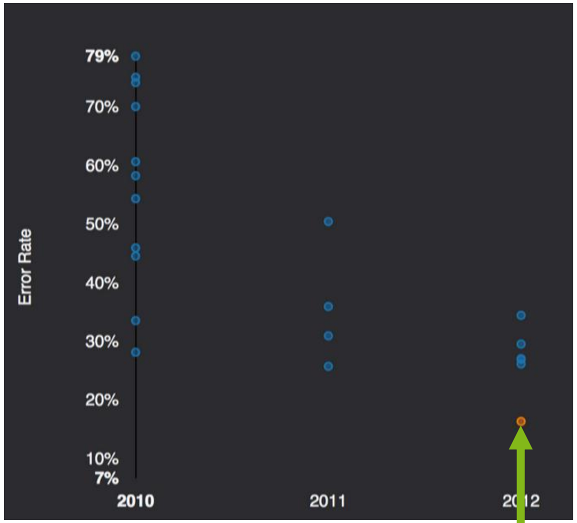
What happened?

The ImageNet Competition

1000 categories
1 Mio. examples



…



A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

2015: computers have lerned to «see»

4.95% Microsoft (February 06)
→ super-human (5.10%)

4.80% Google (February 11)

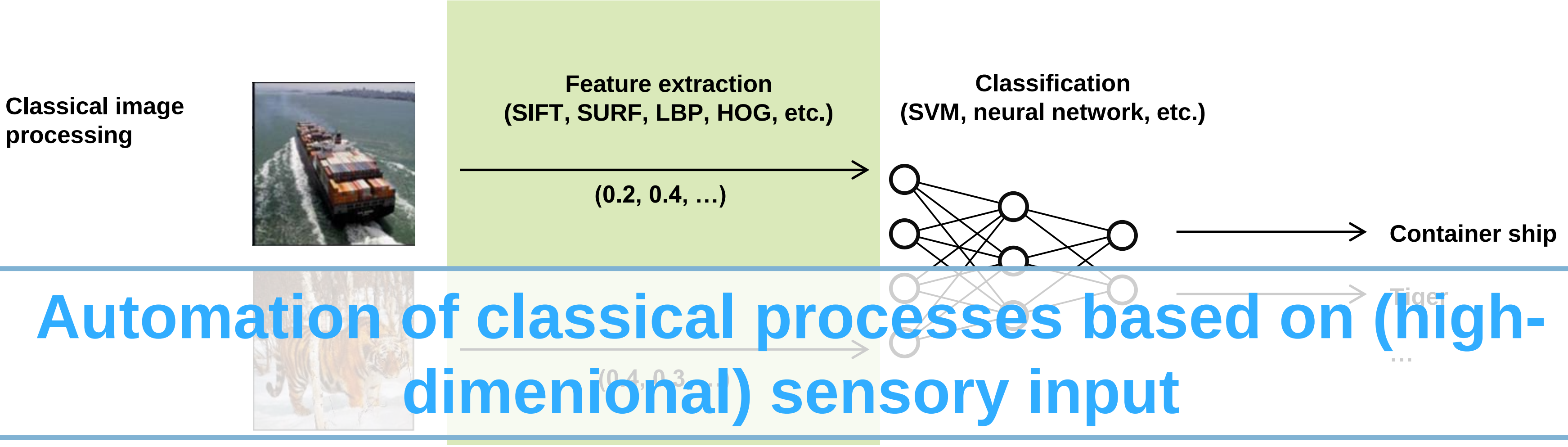
4.58% Baidu (May 11)

3.57% Microsoft (December 10)

3

How does it work?

Idea: Add «depth» to learn features automatically



Classical image processing

Feature extraction
(SIFT, SURF, LBP, HOG, etc.)

(0.2, 0.4, …)

Classification
(SVM, neural network, etc.)

Container ship

(0.4, 0.3, …)

Tiger

…

Automation of classical processes based on (high-dimenional) sensory input

Foundation

Inductive supervised learning

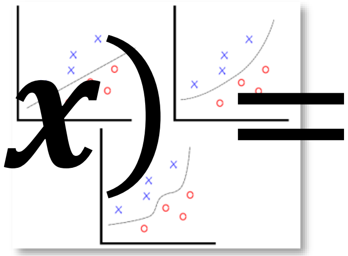
Assumption

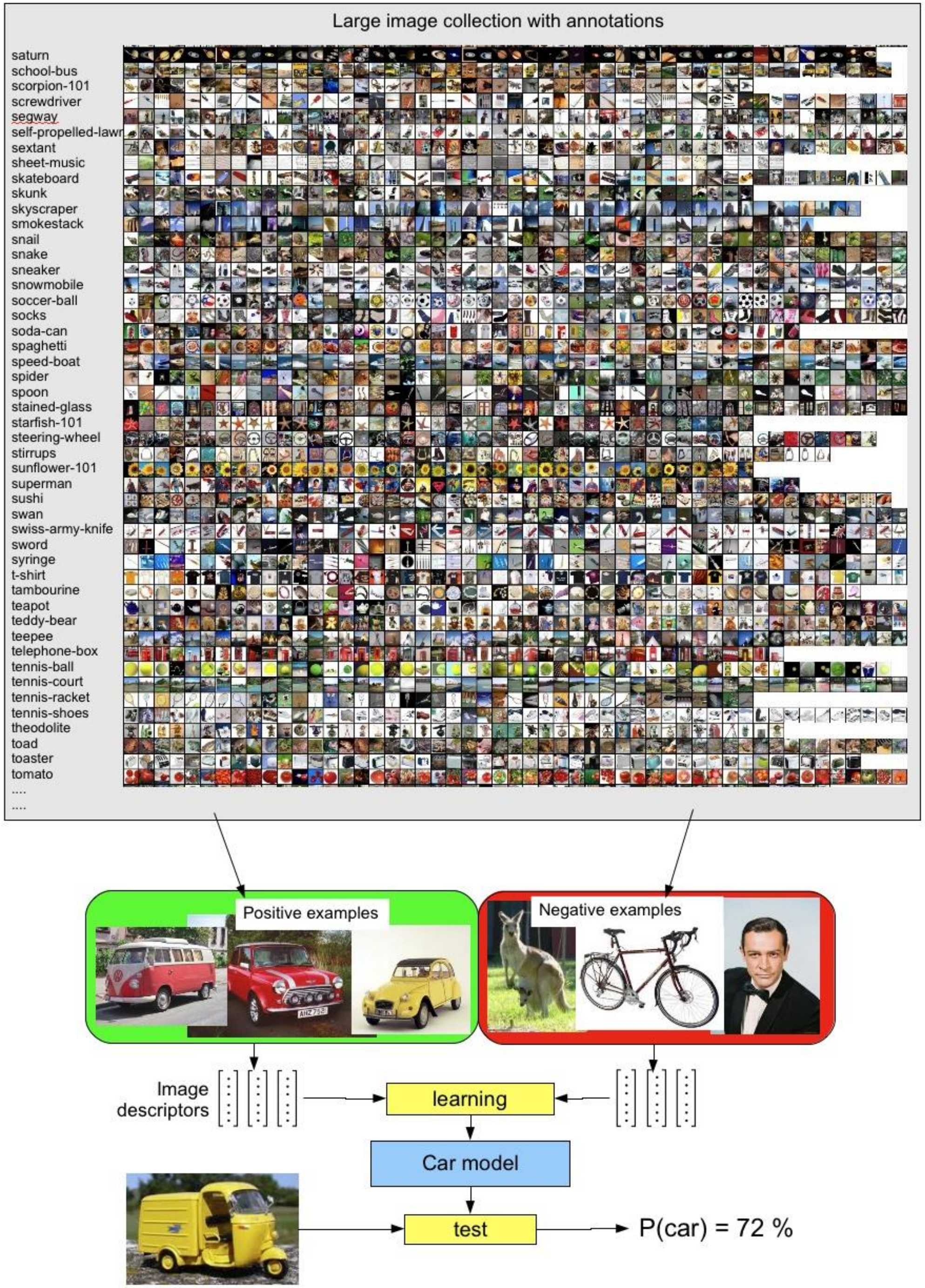
- A model fitted to a *sufficiently large* sample of data…
- …will **generalize** to unseen data

Method

- **Searching for optimal parameters of a function**…
- …such that all sample inputs (images) are mapped to the correct outputs (e.g., «car»)

$$f(x) = y$$

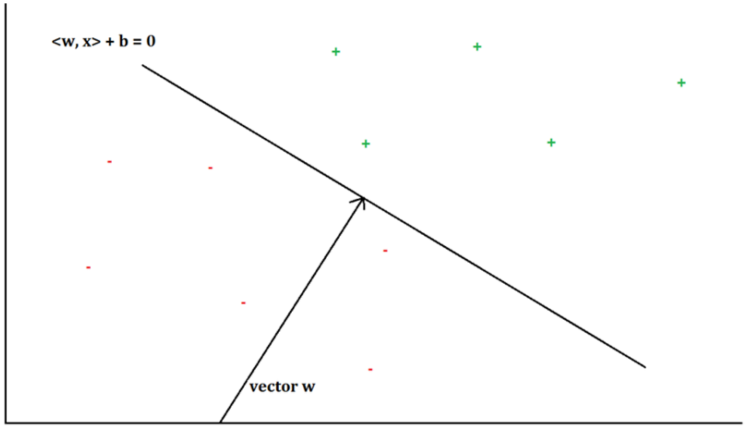




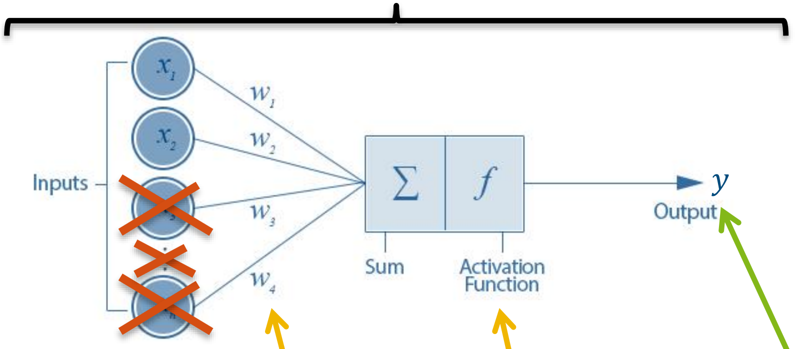
Quelle: http://lear.inrialpes.fr/job/postdoc-large-scale-classif-11-img/attribs_patchwork.jpg

Search for optimal parameters *of a function*?

neuron



neural net

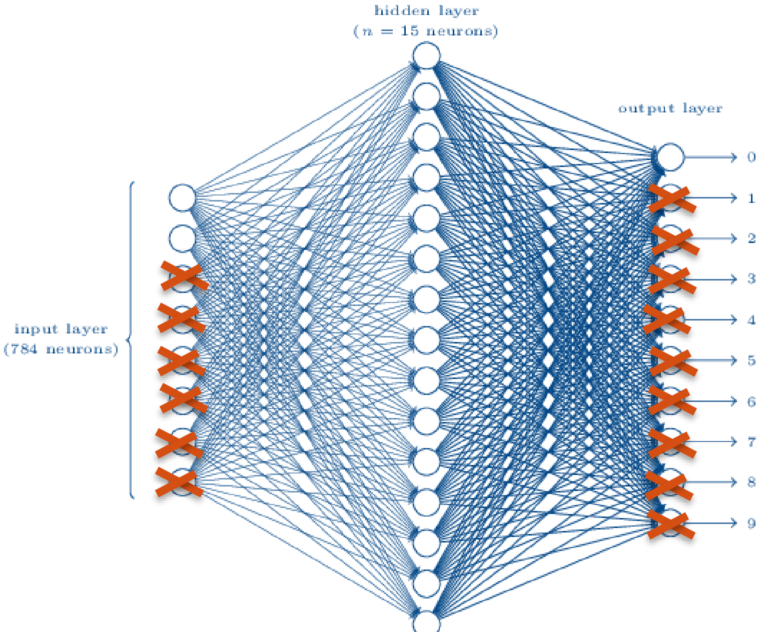


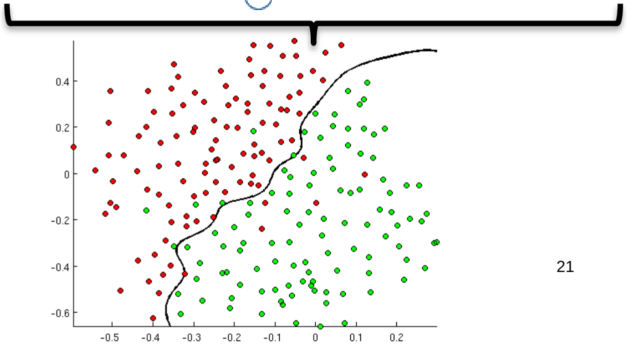
Threshold / decision

features (e.g., pixel)

Result (interpreted as e.g. «car»)

Tunable parameters





What→ Why?→ How?→ Future

4

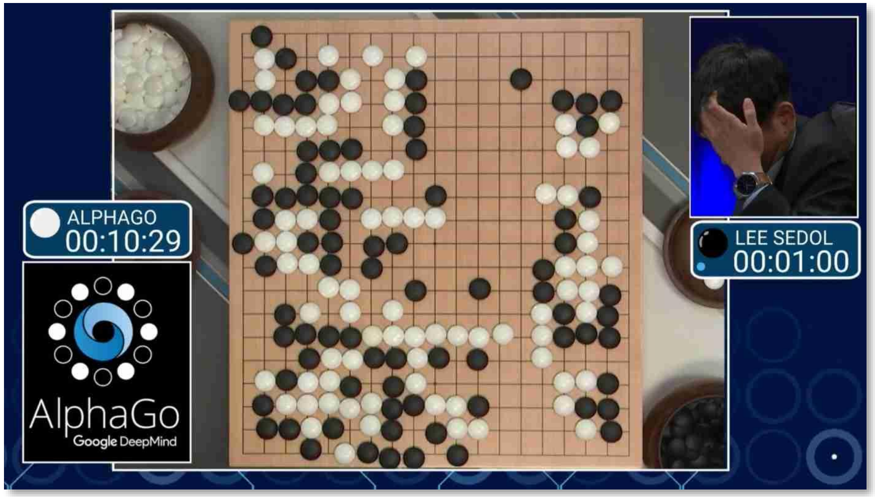
And what’s the connection to a digitally transformed future?

Basis for disruption (I): automation „at scale“

Or: “digital transformation” refers to a shift in all aspects of society, driven/enabled by this small set of technologies

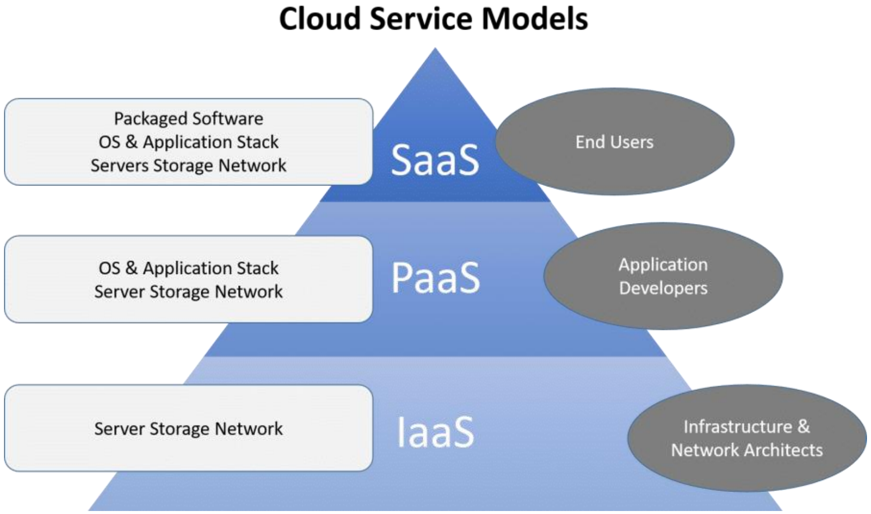
AI

Massively enhanced automation depth through progress in pattern recognition



CLOUD COMPUTING

No need to invest into (IT) infrastructure anymore before entering the market



Basis for disruption (II): decoupling

size of idea ≠ size of implementing organization

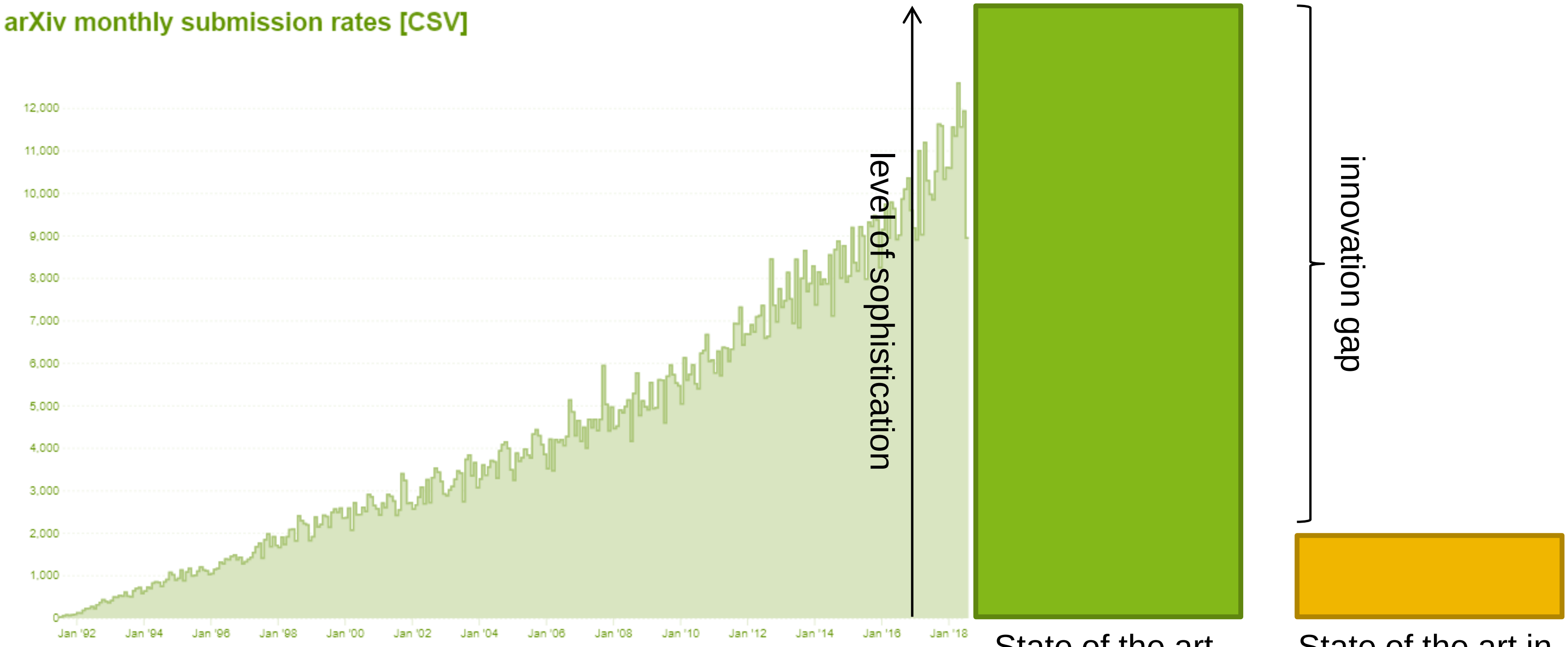
…small organizations can build **whatever they want**
(given know-how, data and an interesting business case)

the technology is sector-independent

…enabling **new** alliances and cooperations

Basis for disruption (III): speed

Average time from (pre-)publication to application: approx. 3 month



level of sophistication

innovation gap

State of the art
in research

State of the art in
industrial application

Forecast: disruption
…even without any further technological progress

1. hypothesis: Use of (current) AI will increase massively within the next 4 years

- Indicator: **AI progress** is mainly driven by **industrial interests (earnings outlook)**; customers value convenienve; these incentives „keep the engine running“

2. hypothesis: This will revolutionize society

- Main question: How does the algorithmically earned **profit** (mainly at large corporations) **distribute**? How does new **free time and convenience distribute**?

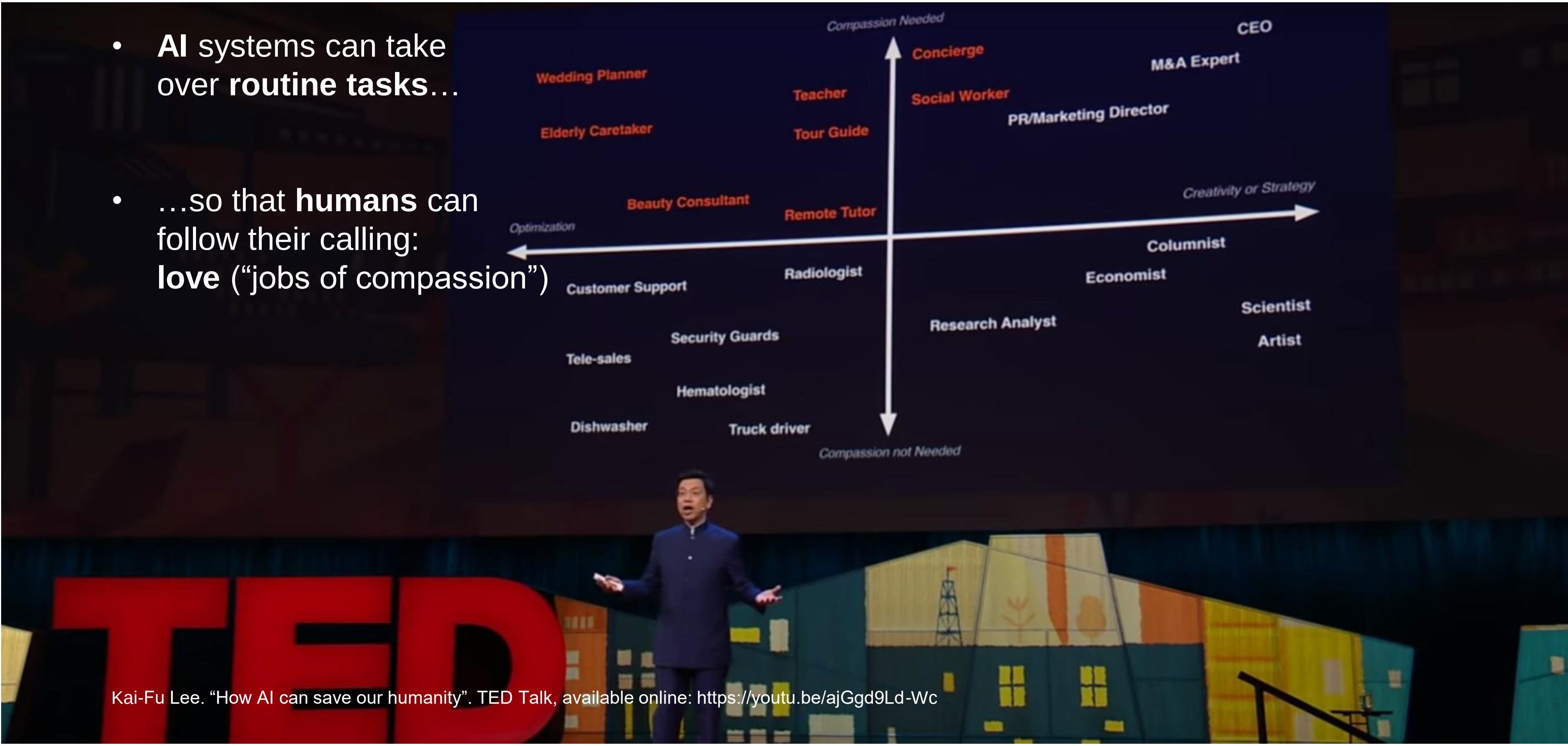
3. hypothesis: Main challenge is our dealings with each other (not with AI)

- Argument: AI (etc.) "for the common good" is an important topic; decisive however is **how the society designes new rules** (regulations) for community life in a digital society (see above)

Cp: Stockinger, Braschler & Stadelmann. "Lessons Learned from Challenging Data Science Case Studies". In: Braschler et al. (Eds), *"Applied Data Science - Lessons Learned for the Data-Driven Business"*, Springer, 2019.

Where are we heading?

The vision of Kai-Fu Lee, venture capitalist & scientist

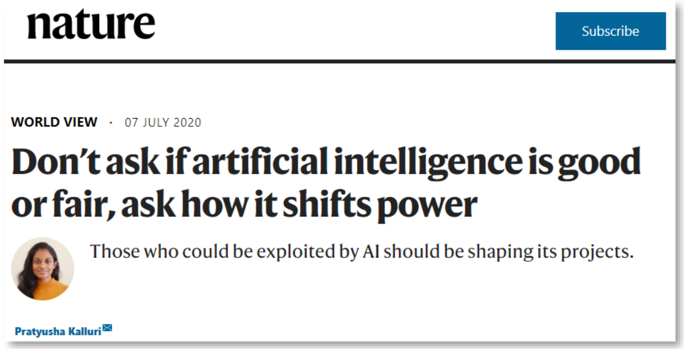


- **AI** systems can take over **routine tasks**…
- …so that **humans** can follow their calling: **love** (“jobs of compassion”)

Kai-Fu Lee. “How AI can save our humanity”. TED Talk, available online: https://youtu.be/ajGgd9Ld-Wc

Conclusions

- Deep Learning lead to a paradigm shift in *pattern recognition tasks*
- *This* enables so many new business opportunities that it (digitally) transforms society
- The *pace is extremely high* (new results are applied within months)
- Big question: what *kind of society are we building* around these opportunities?



nature

Subscribe

WORLD VIEW · 07 JULY 2020

Don't ask if artificial intelligence is good or fair, ask how it shifts power

Those who could be exploited by AI should be shaping its projects.

Pratyusha Kalluri

datalab
www.zhaw.ch/datalab

swiss group for artificial intelligence and cognitive science
SGAICO

data service
Swiss Alliance for Data-Intensive Services

About me:

- Prof. AI/ML, scientific director ZHAW digital
- Email: stdm@zhaw.ch
- Phone: +41 58 934 72 08
- Web: https://stdm.github.io/
- Twitter: @thilo_on_data
- LinkedIn: thilo-stadelmann