

Was kann Künstliche Intelligenz leisten?

Fach Digitale Transformation, CAS Paralegal

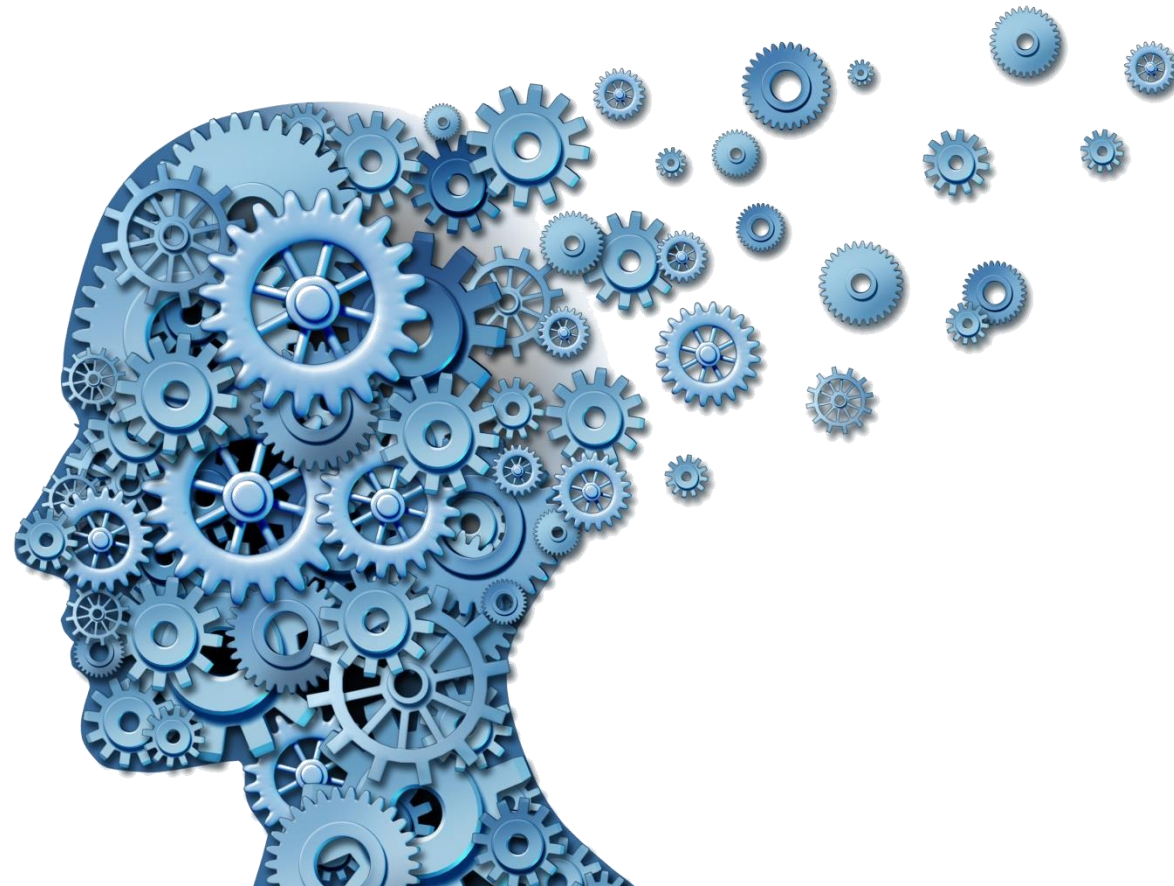
06. September 2019

Thilo Stadelmann

Was ist KI?

Warum ist das jetzt aktuell?

Wie funktioniert das?



Swiss Alliance for
Data-Intensive Services



datalab

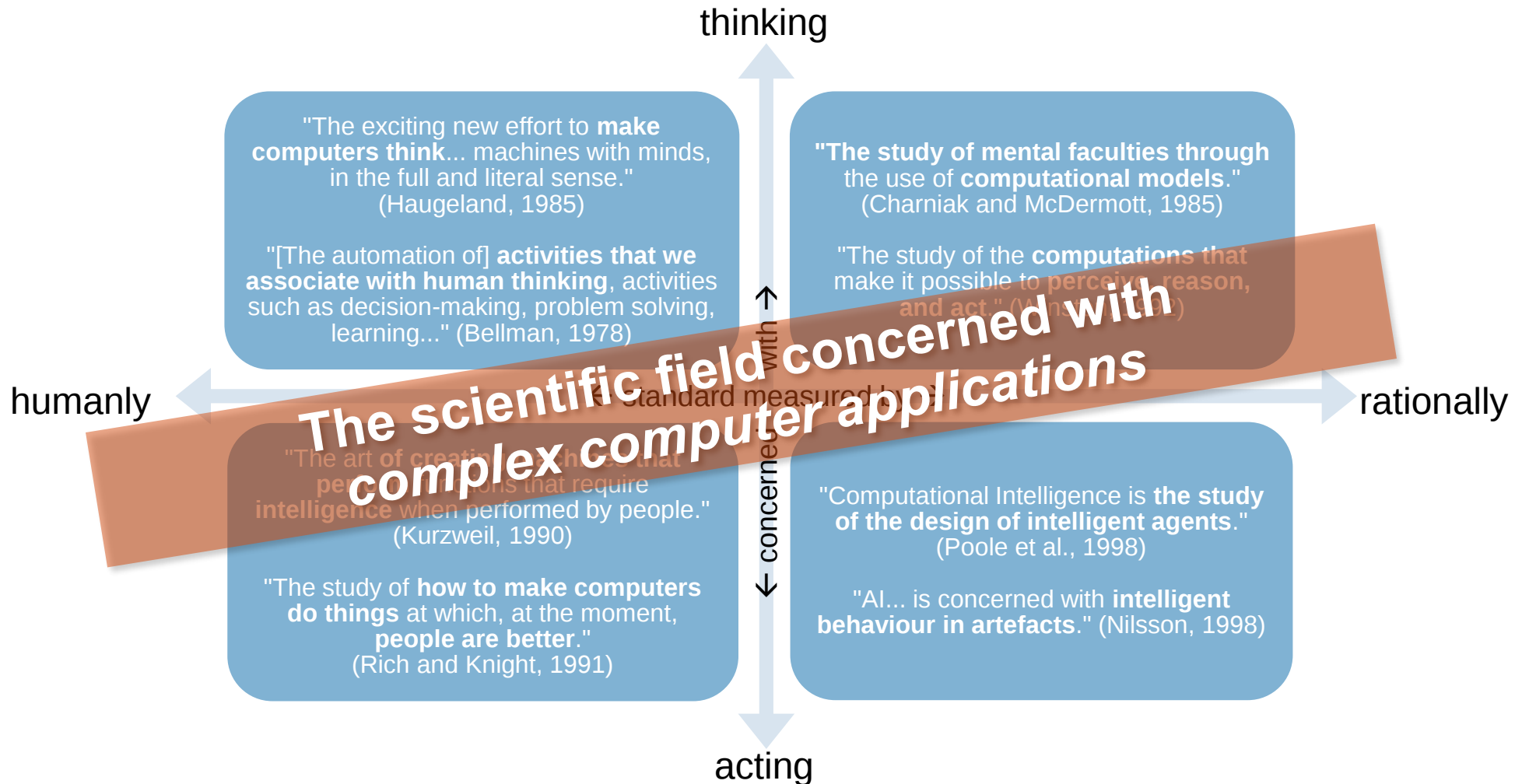
www.zhaw.ch/datalab

Was → Warum? → Wie?

1

Was ist Künstliche Intelligenz?

Was ist künstliche Intelligenz?



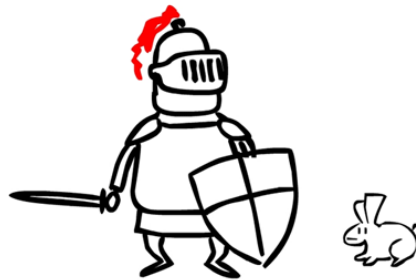
Pragmatisches Designparadigma: Rationale Agenten

Agents

- an **entity that perceives and acts**
- a **function from percept histories to actions** $f: P^* \rightarrow A$

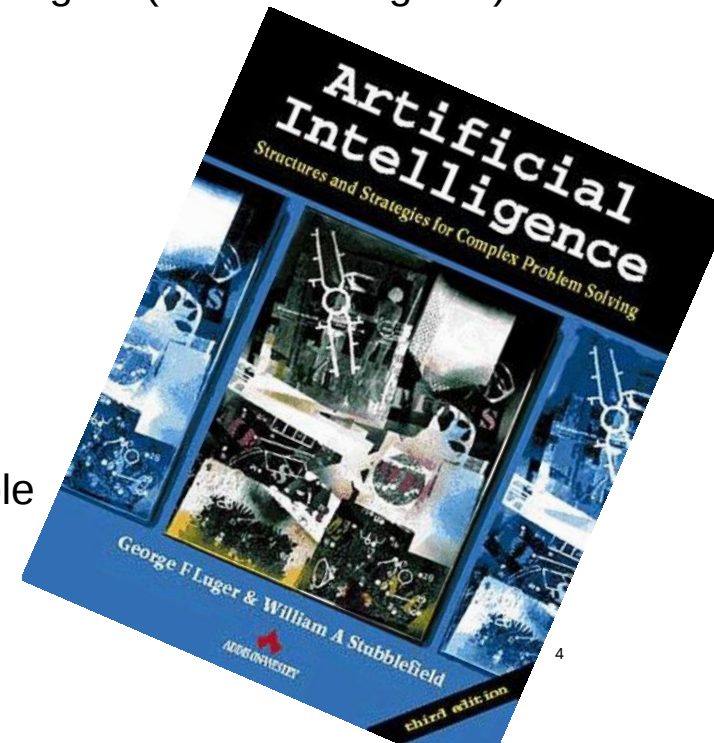
Rational agents

- **For any** given class of **environments** and **tasks**, we **seek** the agent (or class of agents) with the **best performance**

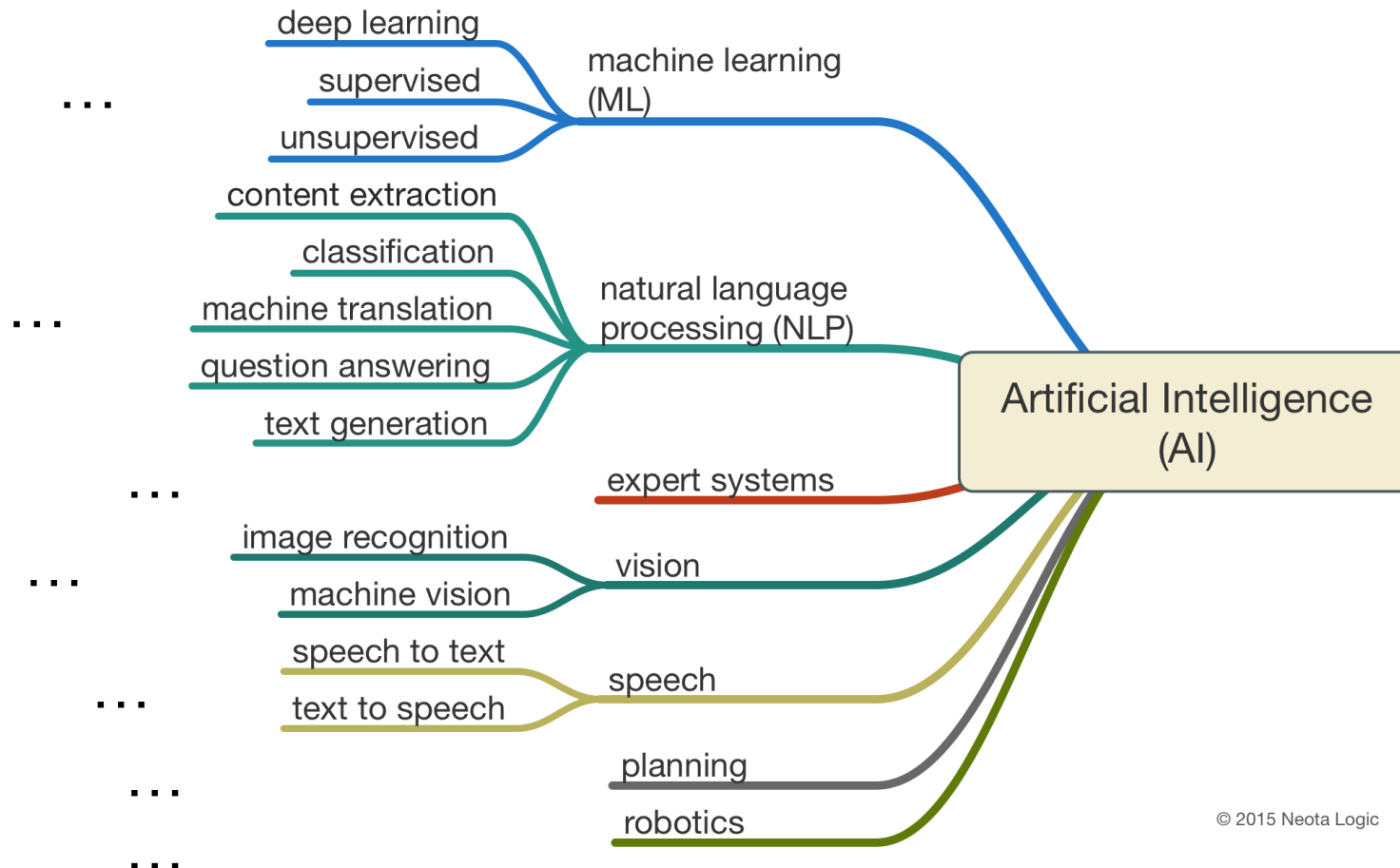


Caveat

- Computational limitations make perfect rationality unachievable
→ **Design best program for given machine resources**

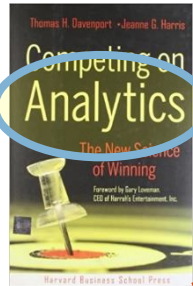
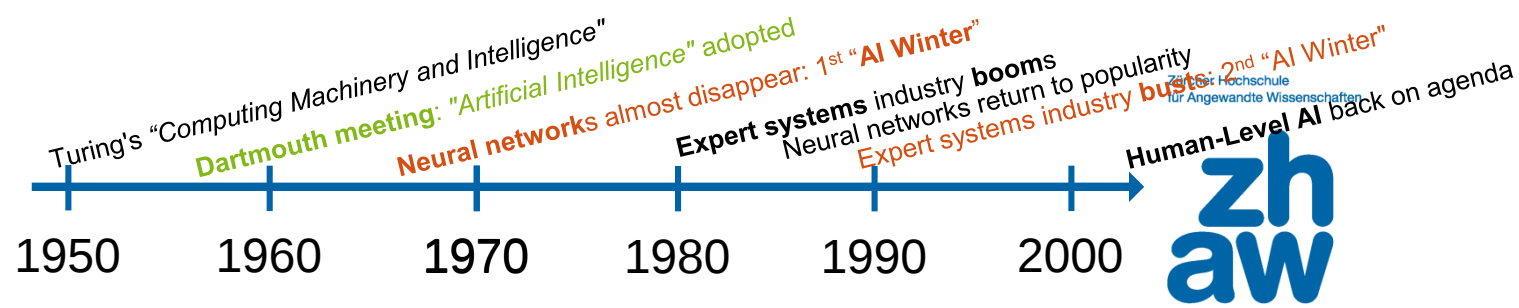


Was gehört zu künstlicher Intelligenz?



© 2015 Neota Logic

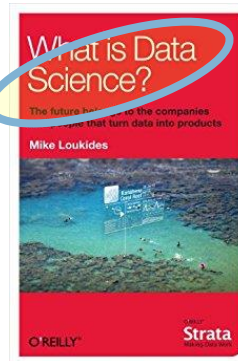
KI im Kontext



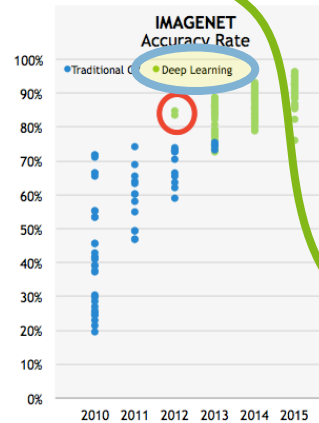
FINTECH INNOVATIONS



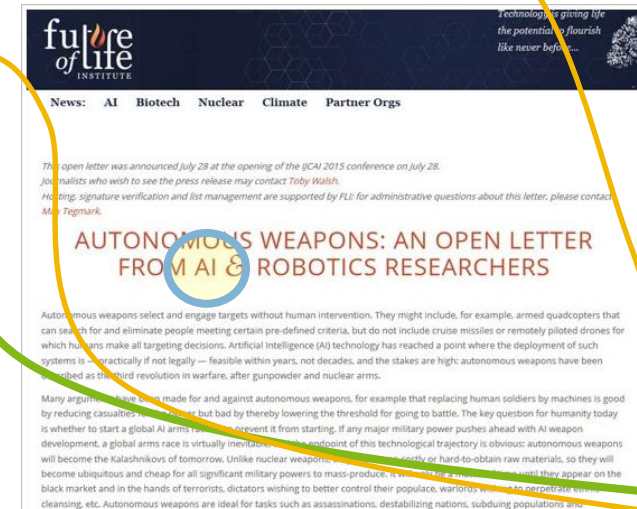
2007



2012



2016



Was kann KI bereits heute?

1. Play a decent game of **table tennis**
2. **Drive** safely along a curving **mountain road**
3. Drive safely along **Technikumstrasse** Winterthur
4. **Buy** a week's worth of **groceries on the web**
5. Buy a week's worth of groceries **at Migros**
6. **Play** a decent game of **bridge**
7. **Discover** and prove a new mathematical **theorem**
8. **Design** and execute a **research program** in molecular biology
9. Write an **intentionally funny** story
10. Give competent **legal advice** in a specialized area of law
11. **Translate** spoken English **into spoken** Swedish in real time
12. **Converse** successfully with another person for an hour
13. Perform a complex **surgical operation**
14. **Unload** any **dishwasher** and put everything away
15. Compete in the game show **Jeopardy!**
16. **Write clickbait** articles fully automatized

ok

ok

ok (only since recently)

ok

no

ok

not complete

not complete

no

ok

ok

no

not complete

no

ok

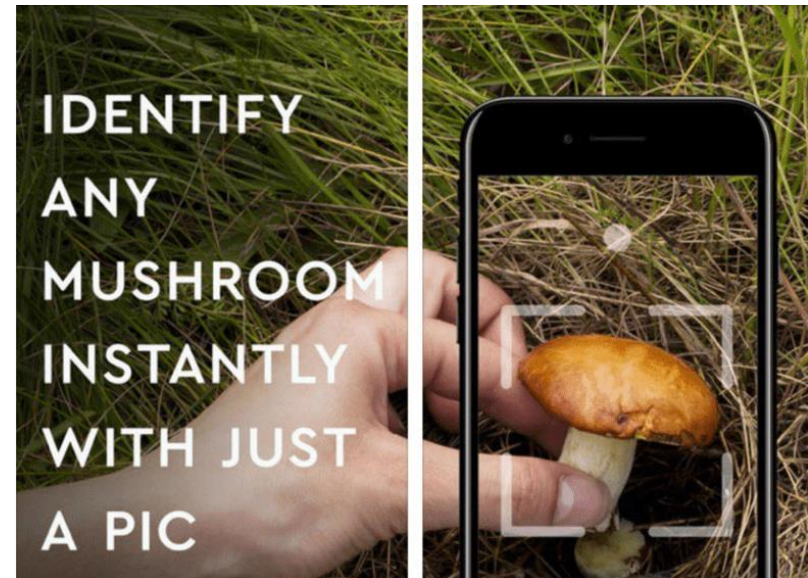
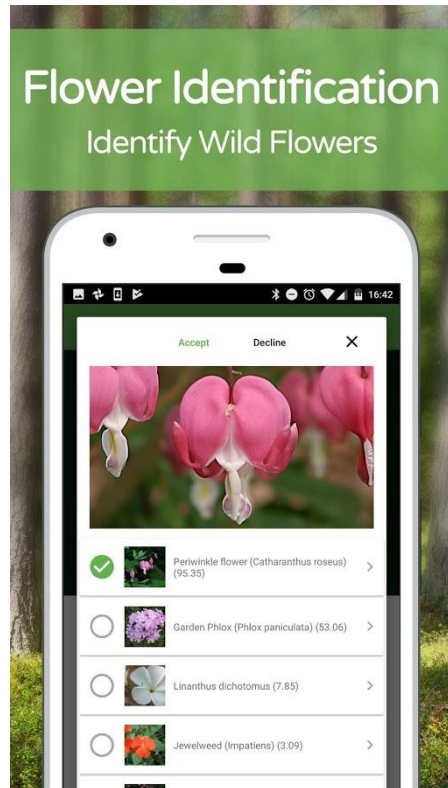
ok



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

Beispiel: Machbar vs. gefährlich

Technologie: Computer Vision mit Deep Learning



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

Beispiel: Markterfolg vs. regulatorische Hürden

Technologie: Recommender Systems

Customers Who Bought This Item Also Bought

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

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

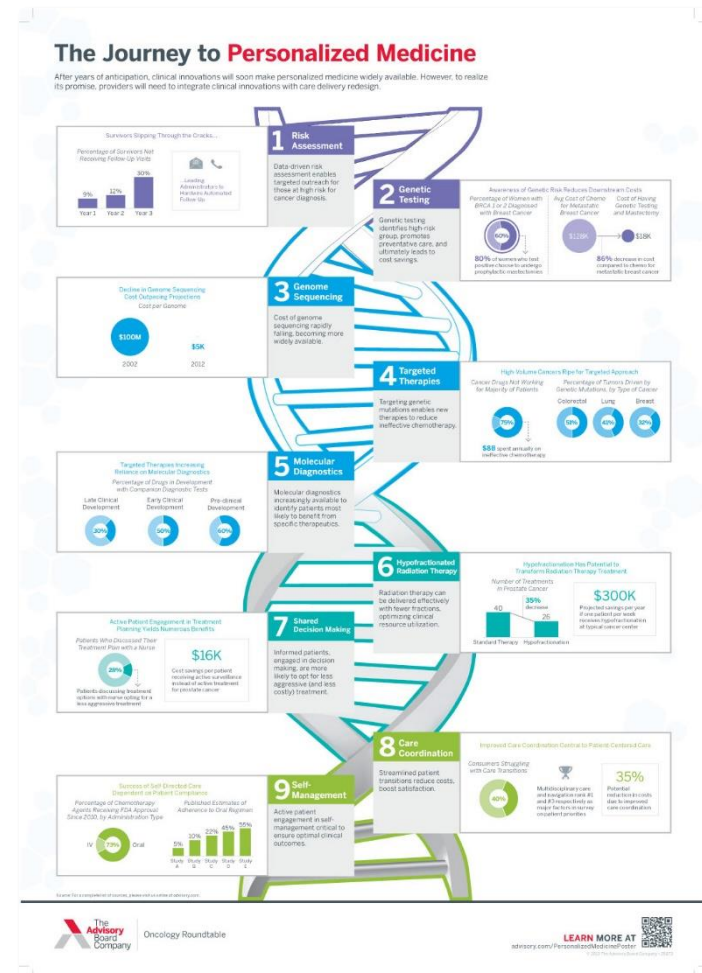
Bounded Rationality: The Adaptive Toolbox (Dahlin) by G Gigerenzer
£20.95

What Do Customers Ultimately Buy After Viewing This Item?

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

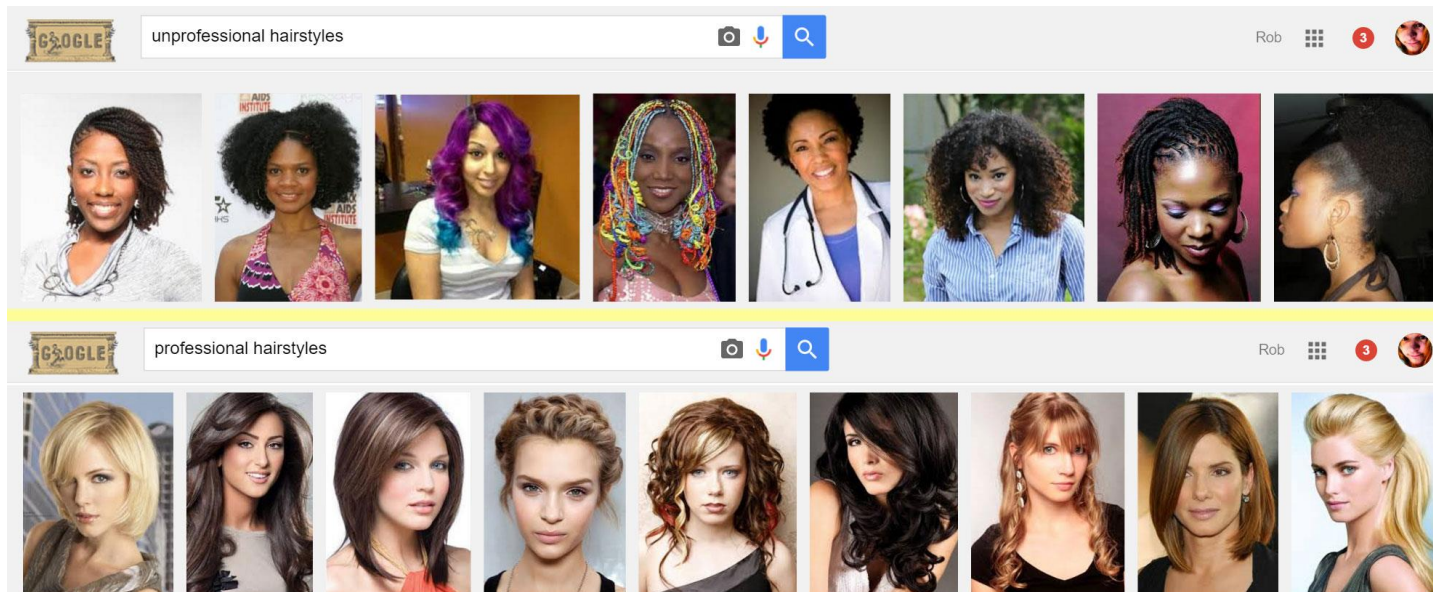
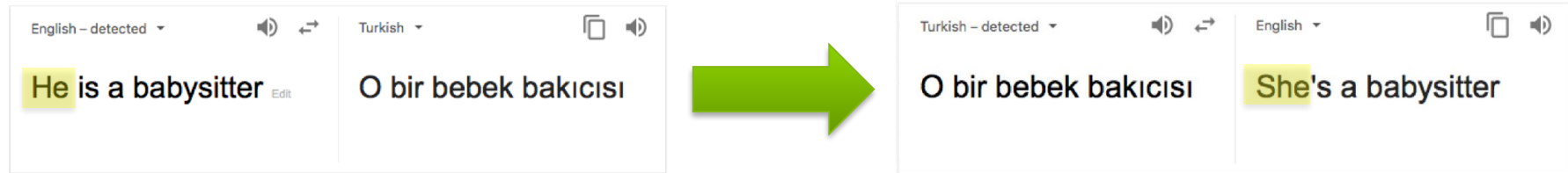
17% buy
Gut Feelings: Short Cuts to Better Decision Making
£6.74

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



Beispiel: Statistik vs. Bias

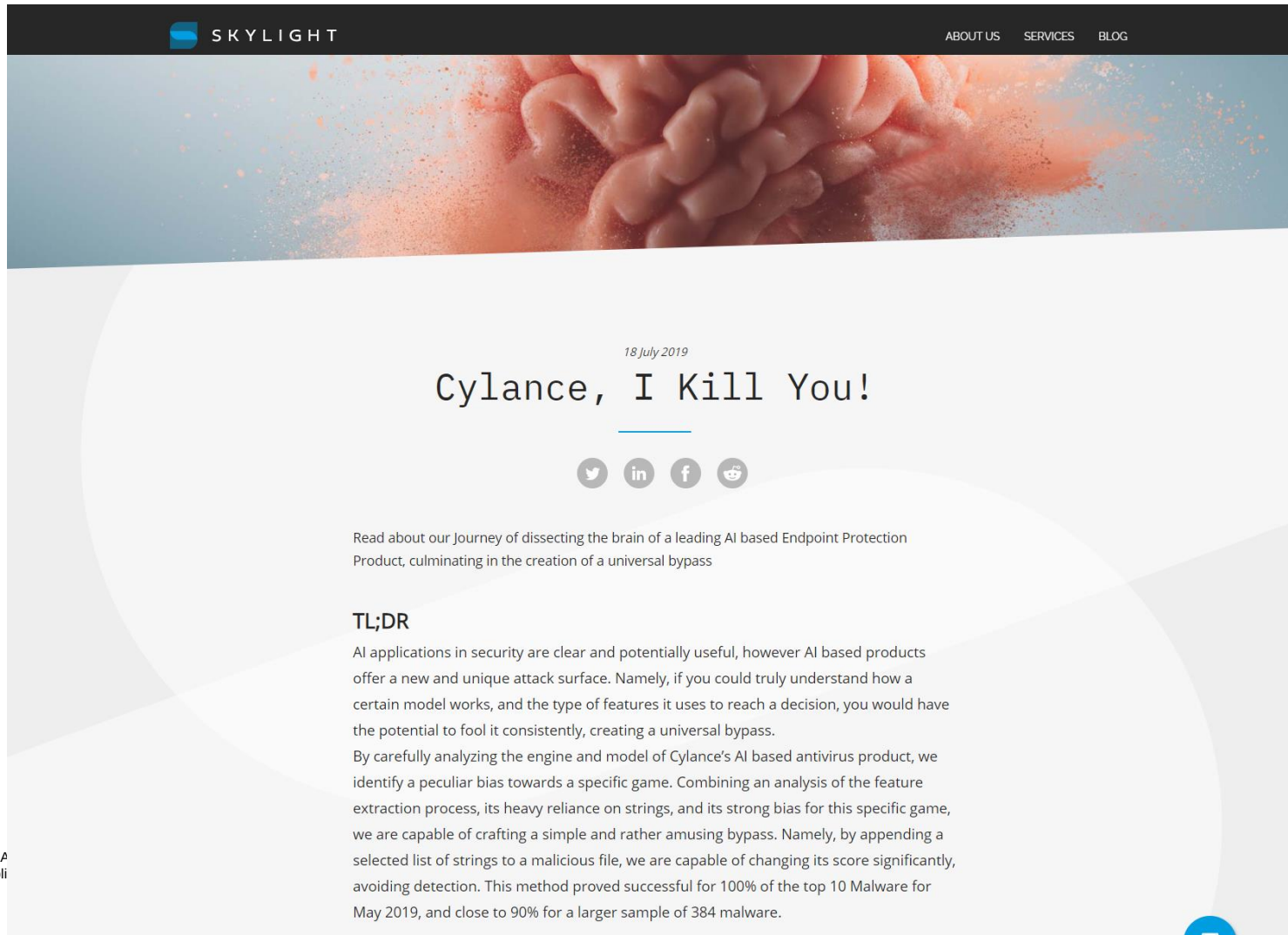
Technologie: Machine Learning



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

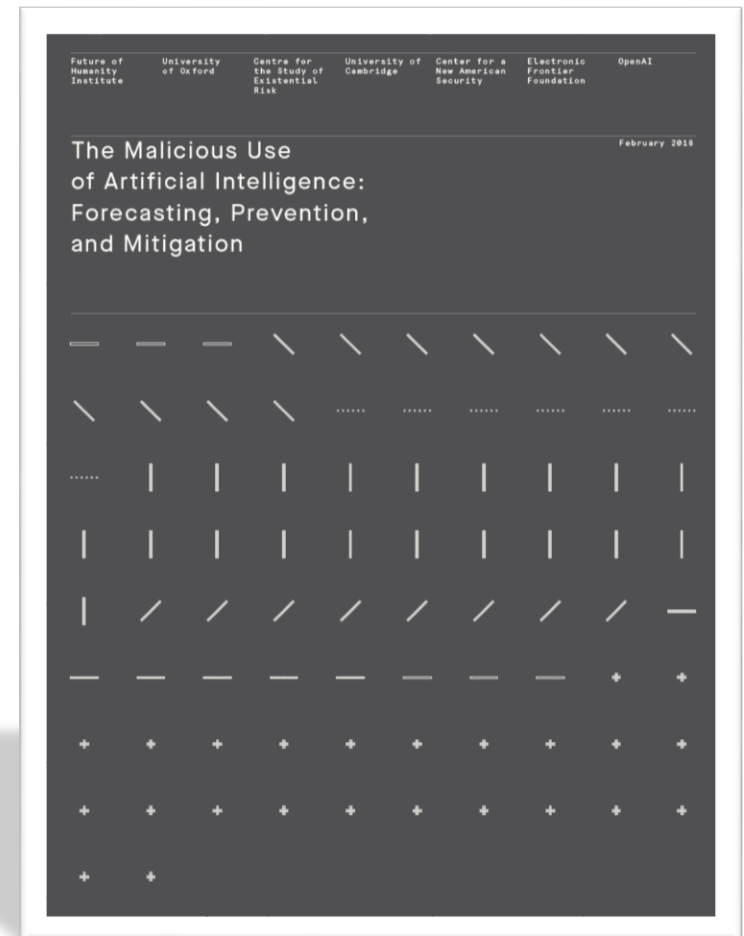
Beispiel: künstl. Intelligenz vs. natürl. Dummheit

Technologie: Machine Learning mit nachgelagerten Regeln



Gefahren durch KI?

- KI ist per Definition eine **“dual use Technology”**
→ siehe Report von Brundage et al., 2018
- Aber: **“natürliche Dummheit”** ist die grössere Bedrohung
- **Algorithmische Ethik** und **erklärbare KI** sind in den letzten Jahren zu einem top Forschungsfeld geworden – nicht wegen der unkalkulierbaren Risiken per se, sondern:



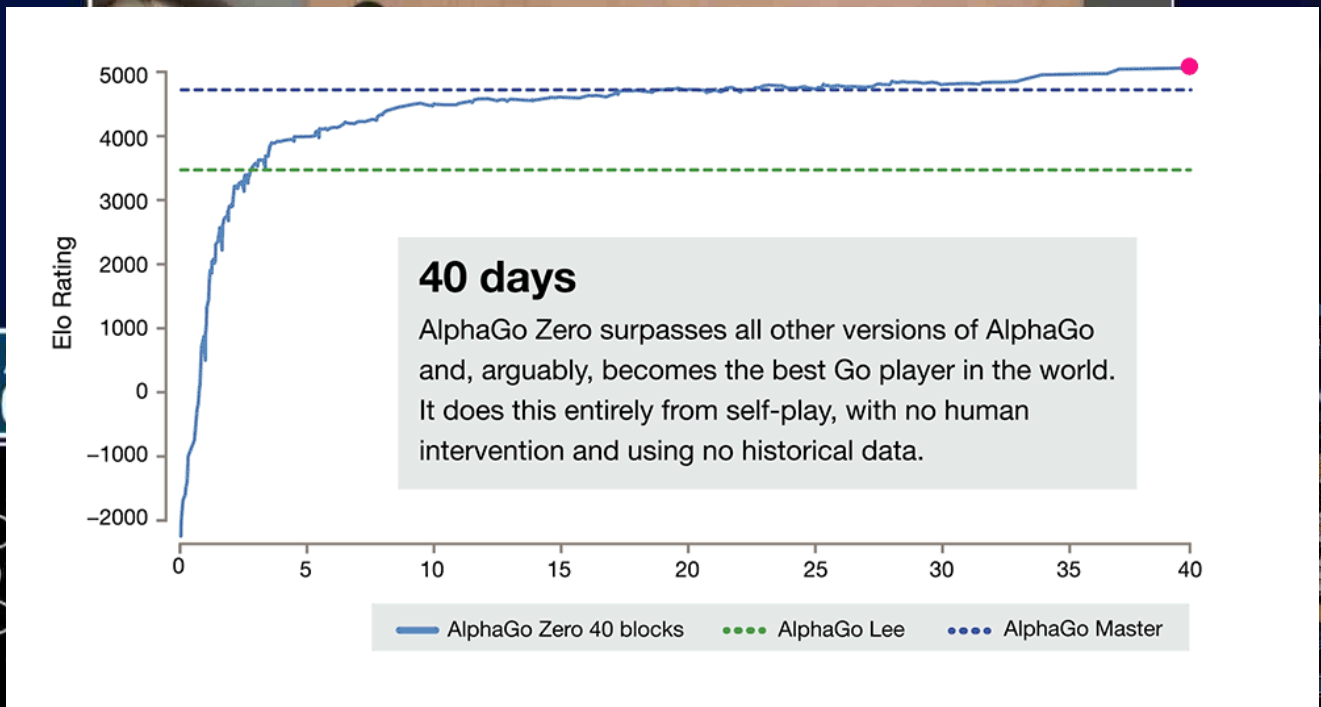
Was → Warum? → Wie?

2

Warum ist das jetzt aktuell?
(Eine kurze Geschichte der letzten Jahre)

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

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



AlphaGo
Google DeepMind

Google will buy DeepMind, reports that the company is in talks to buy the startup. Google couldn't disclose deal terms.

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

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

ALL SYSTEMS GO

CONSERVATION
SONGBIRDS A 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 7557

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

And the results are impressive

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



Computing

Algorithmic Artistic Other In

A deep neural network can transfer the style of one image to another.

by Emerging Tech

The nature of art
of Vincent Van Gogh
Edvard Munch's
humans recogni



Original photo

Reference photo

Result

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

Ad closed by Google

Report this ad

AdChoices

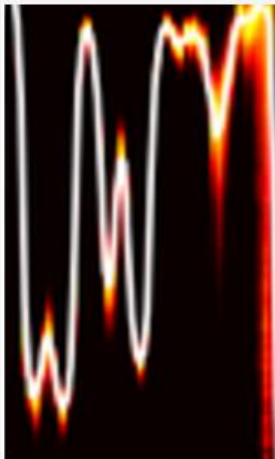


NOW TRENDING

WaveNet lässt Computersprache natürlich klingen

von Henning Steier / 12.9.2017

Die Google-Tochter DeepMind macht auch Musik.



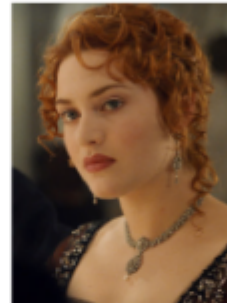
DeepMind lässt WaveNet Spr

Die Google-Tochter DeepMind hat ein Spiel «Go» Schlagzeilen: Das Unternehmen ist eines der besten menschlichen Go-Spieler. Londoner Unternehmen erzeugt Sprache, die sehr natürlich klingt. Im Blogbeitrag des Unternehmens wird erklärt, wie das funktioniert. Der Massstab nimmt. Man hat

Intro

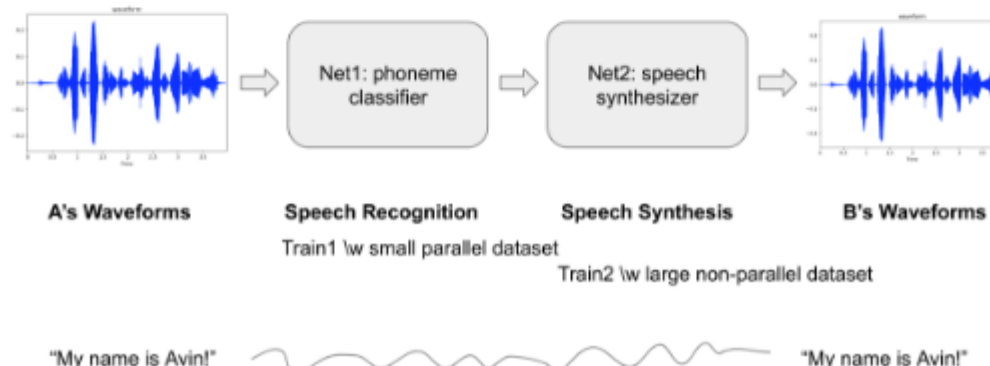
What if you could imitate a famous celebrity's voice or sing like a famous singer? This project started with a goal to convert someone's voice to a specific target voice. So called, it's voice style transfer. We worked on this project that aims to convert someone's voice to a famous English actress [Kate 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
is Texteingabe»

nerierte Musik
ne Inhaltsvorgabe»



1 Second

...und die Liste liesse sich fortsetzen!

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 $C(z)$ to produce fake images
 - [ML-Heavy] Training DCGANs
 - Existing GANs
 - [ML-Heavy] Training DCGANs
 - Running DCGANs
- Step 3: Finding the best completion
 - Image completion
 - [ML-Heavy] Image completion
 - [ML-Heavy] Image completion
 - Completing images
- Conclusion
- Partial bibliography
- Bonus: Incomplete

Introduction

Content-aware fill is a powerful tool for image completion and inpainting. In this post, we'll explore how to use deep learning to complete images of faces. The completion is done using TensorFlow.

We'll approach image completion in three steps:

1. We'll first interpret the image as a probability distribution.
2. This interpretation will be used to generate new samples from an unknown probability distribution.
3. Then we'll find the best completion.



Andrej Karpathy blog

About Hacker's guide to Neural Networks

The Unreasonable Effectiveness of Recurrent Neural Networks

May 22, 2015



TECH

Nvidia AI Generates Fake Faces Based On Real Celebs

BY STEPHANIE MLADT 10.31.2017 :: 10:00AM EST

32 SHARES



I'm getting a distinctly mid-90s "The Rachel" vibe from the woman in the top left corner (via Nvidia)

STAY ON TARGET

AI Shelley Pens Truly Creepy Horror Stories-And You Can Help

Neural Network Serves Up Truly Frightening Halloween Costume Ideas

Celebrity scandals are about to get a lot more complicated.

Nvidia has **developed** a way of producing photo-quality, AI-generated human profiles—by using famous faces.

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:

- ★ **Efficient Estimation of Word Representations in Vector Space** – Mikolov et al. 2013
- ★ **Distributed Representations of Words and Phrases and their Compositionality** – Mikolov et al. 2013
- ★ **Linguistic Regularities in Continuous Space Word Representations** – Mikolov et al. 2013
- ★ **word2vec Parameter Learning Explained** – Rong 2014
- ★ **word2vec Explained: Deriving Mikolov et al's Negative Sampling Word-Embedding Method** – Goldberg and Levy 2014

From the first of these papers ('Efficient estimation...') we get a description of the *Continuous Bag-of-Words* and *Continuous Skip-gram* models for learning word vectors (we'll talk about what a word vector is in a moment...). From the second paper we get more illustrations of the power of word vectors, some additional information on optimisations for the skip-gram model (hierarchical softmax and negative sampling), and a discussion of applying word vectors to phrases. The third paper ('Linguistic

R
at
la
ce



hand,

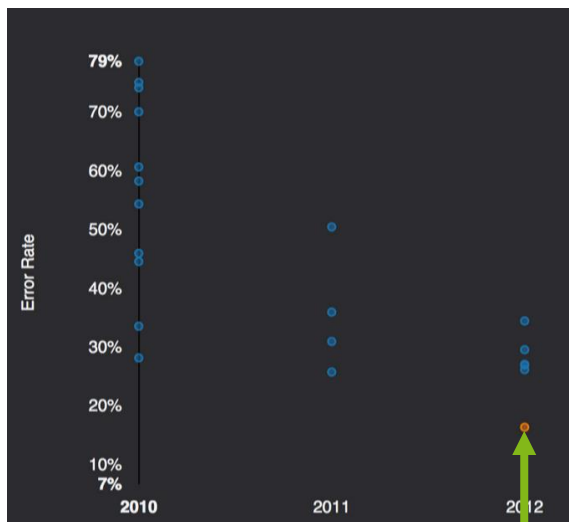
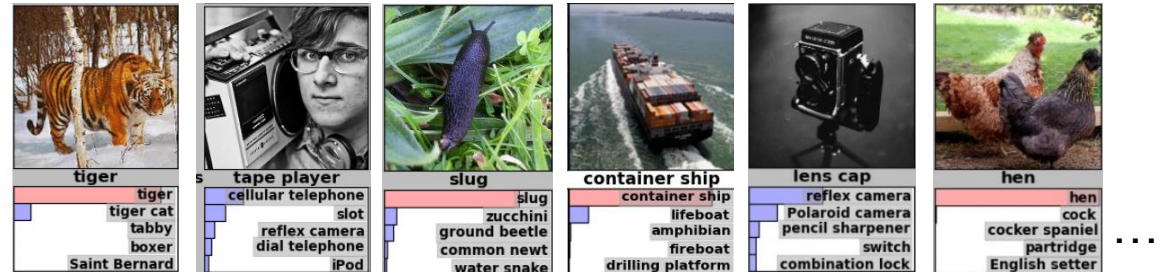
Law,
ds,

Was ist passiert?

Der ImageNet Wettbewerb



1000 Kategorien
1 Mio. Beispiele



2015: Computer *haben* “Sehen” gelernt

4.95% Microsoft (06. Februar)
→ Besser als Menschen (5.10%)

4.80% Google (11. Februar)

4.58% Baidu (11. Mai)

3.57% Microsoft (10. Dezember)

Was → Warum? → Wie?

3

Wie geht das?

Grundlage

Induktives überwachttes Lernen

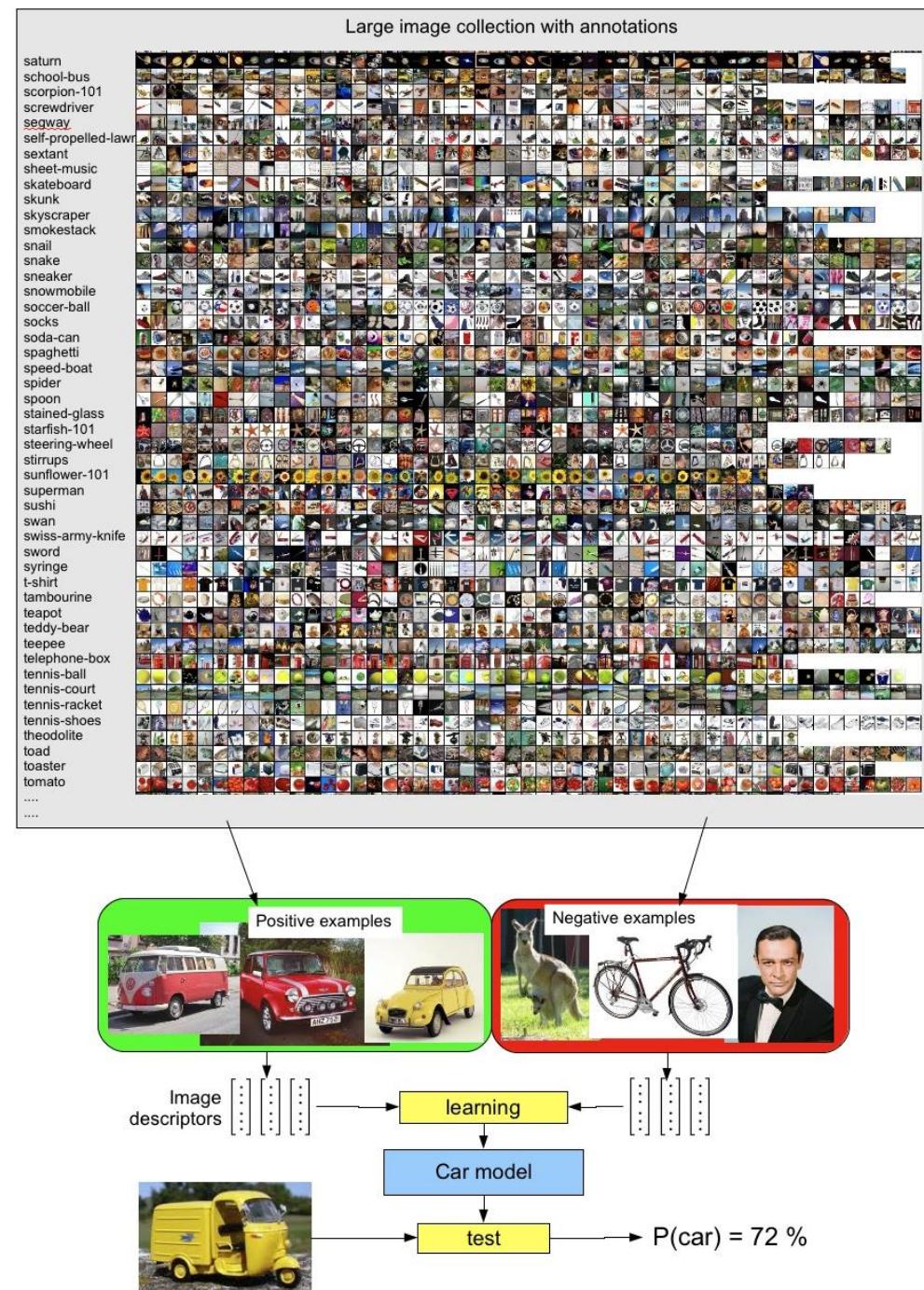
Annahme

- Ein an *genügend viele* Beispiele angepasstes Modell...
- ...wird auch auf unbekannte Daten **generalisieren**

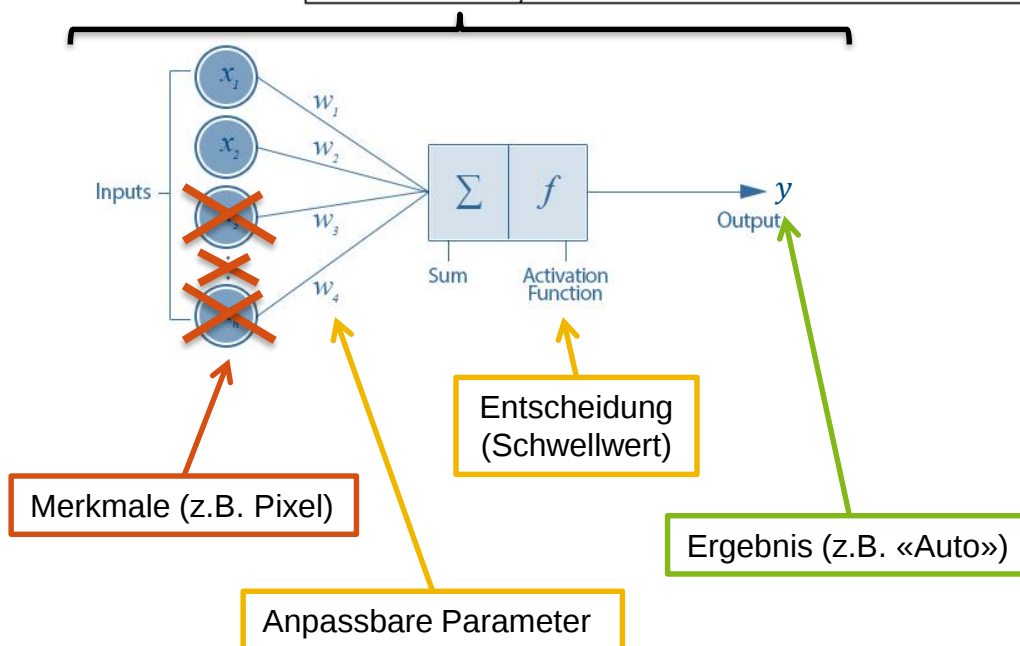
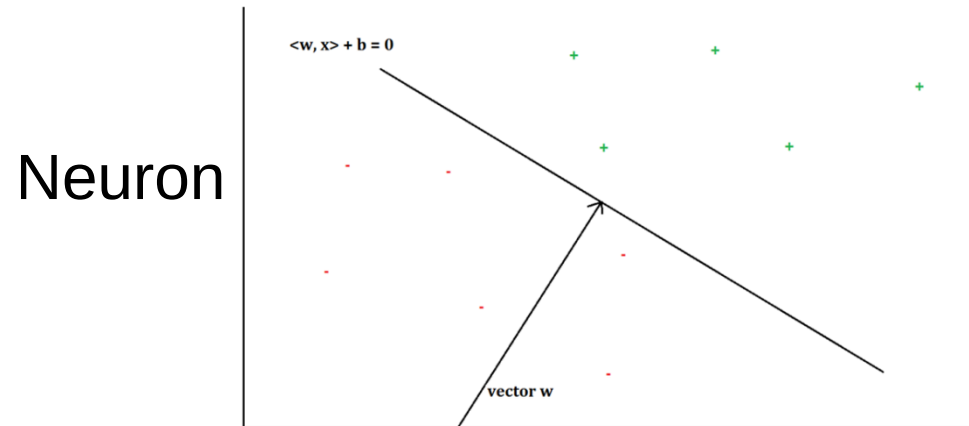
Methode

- **Suchen der Parameter einer gegebenen Funktion...**
- ...so dass für alle Beispiele Eingabe (Bild) auf Ausgabe («Auto») abgebildet wird

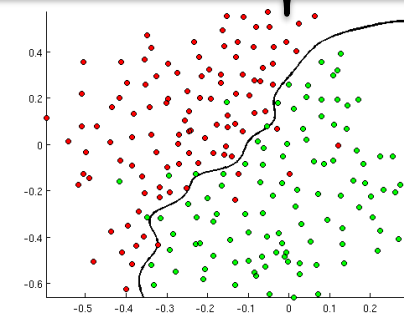
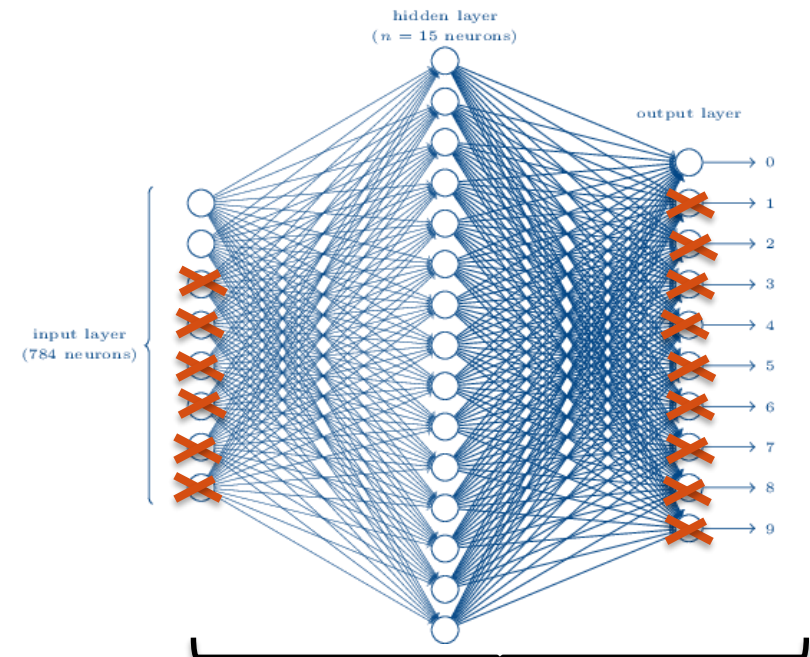
$$f(x) = y$$



Suche der Parameter *einer Funktion*?



Neuronales Netz



Schlussfolgerungen

- Deep Learning hat zu Paradigmenwechsel in *Mustererkennungsaufgaben* geführt
- Die Zeit vom Grundlagenresultat zur praktischer Anwendung beträgt wenige Monate
- Es gibt Methoden zum Hineinschauen in neuronale Black Boxes (siehe Anhang)
- Spezifische Aufgaben lassen sich sehr gut automatisieren (z.B. Ähnlichkeitssuche)



Zu mir:

- Prof. KI/ML, Scientific Director ZHAW digital
- Email: stdm@zhaw.ch
- Telefon: 058 934 72 08
- Web: <https://stdm.github.io/>
- Twitter: @thilo_on_data
- LinkedIn: thilo-stadelmann



Mehr zum Thema:

- Data+Service Alliance: www.data-service-alliance.ch
- KI: <https://sgaico.swissinformatics.org/>
- Zusammenarbeit: datalab@zhaw.ch

ANHANG

Developing for algorithmic fairness

FAT / ML

The FAT ML code of conduct

See <http://www.fatml.org/resources/principles-for-accountable-algorithms>

Purpose

- Help developers to **build algorithmic systems in publicly accountable ways**
- Accountability: the **obligation to report, explain, or justify** algorithmic decision-making & **mitigate** any **negative social impacts** or potential harms

Premise

- *A **human ultimately responsible** for decisions made/informed by an algorithm*

Principles

- **Responsibility, Explainability, Accuracy, Auditability, Fairness**

Make available somebody who will take care of adverse individual / societal effects

Explain any **algorithmic decision** in non-technical terms to end users

Report all sources of **uncertainty / error** in algorithms & data

Enable 3rd parties to **probe & understand** system **behavior**

Ensure algorithmic **decisions are not discriminatory** w.r.t. to people groups

Making it actionable

- **Publish a Social Impact Statement**
- ...use above **principles as a guiding structure**
- ...**revisit three times** during development process: design stage, pre-launch, post-launch

Suche der Parameter einer Funktion?

Wahrscheinlichkeit [%] für bestimmtes Ergebnis

- Unser Neuronales Netz: $f_W(x) = y$
mit Bild x , echtem Resultat y und Parametern W
($W = \{w_1, w_2, \dots\}$ anfangs zufällig gewählt)

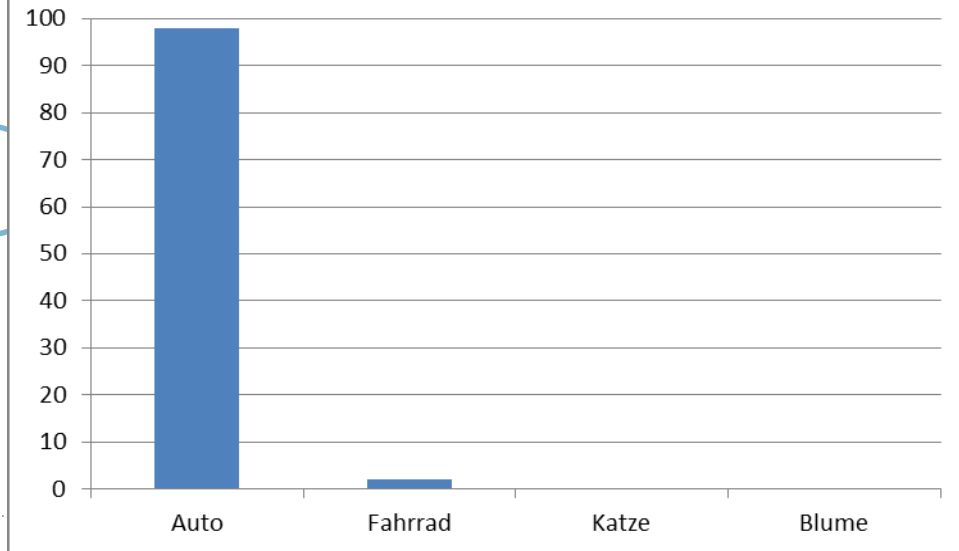
- Fehlermass: $l(W) = \frac{1}{N} \sum_{i=1}^N (f_W(x_i) - y_i)^2$
Durchschnitt der quadratischen Abweichungen
über alle Bilder (Loss)

$$l(W) = \frac{1}{N} \sum_{i=1}^N (f_W(x_i) - y_i)^2$$

Durchschnitt (über
alle Beispiele)

Differenz IST – SOLL
(Fehler)

Bestraft grosse Fehler
überproportional
stärker



← Fehlerlandschaft

Methode: Anpassung der Gewichte
von f in Richtung der steilsten
Steigung (abwärts) von J

Was «sieht» das Neuronale Netz?

Hierarchien komplexer werdender Merkmale

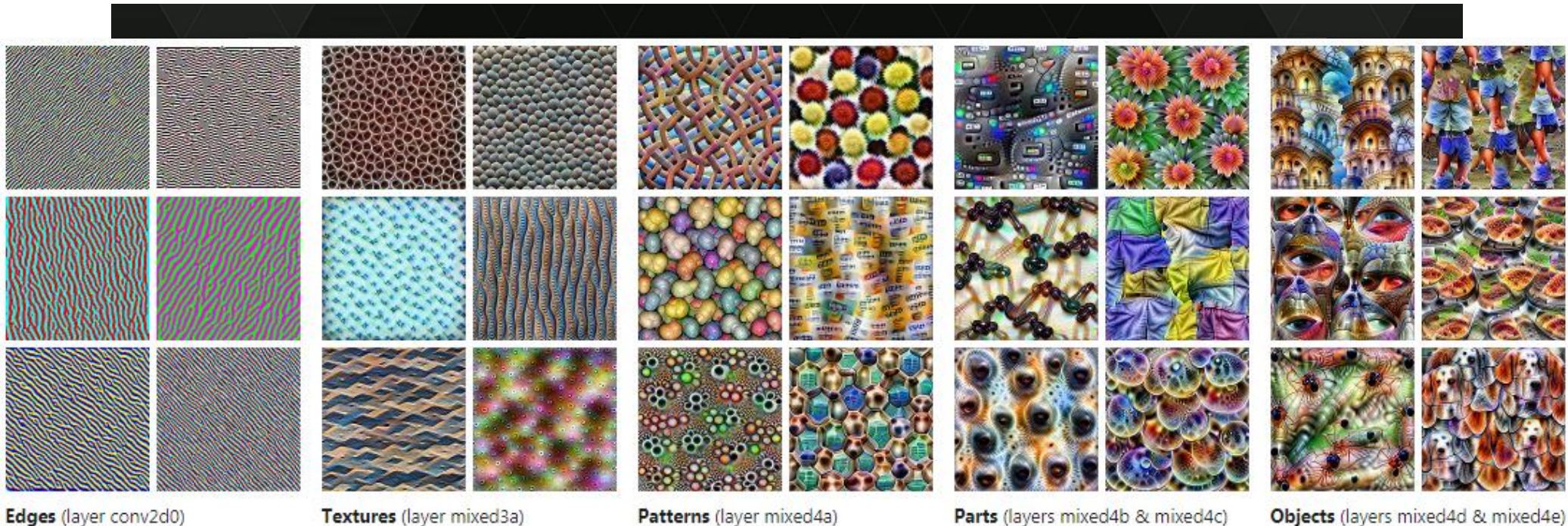


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011.
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

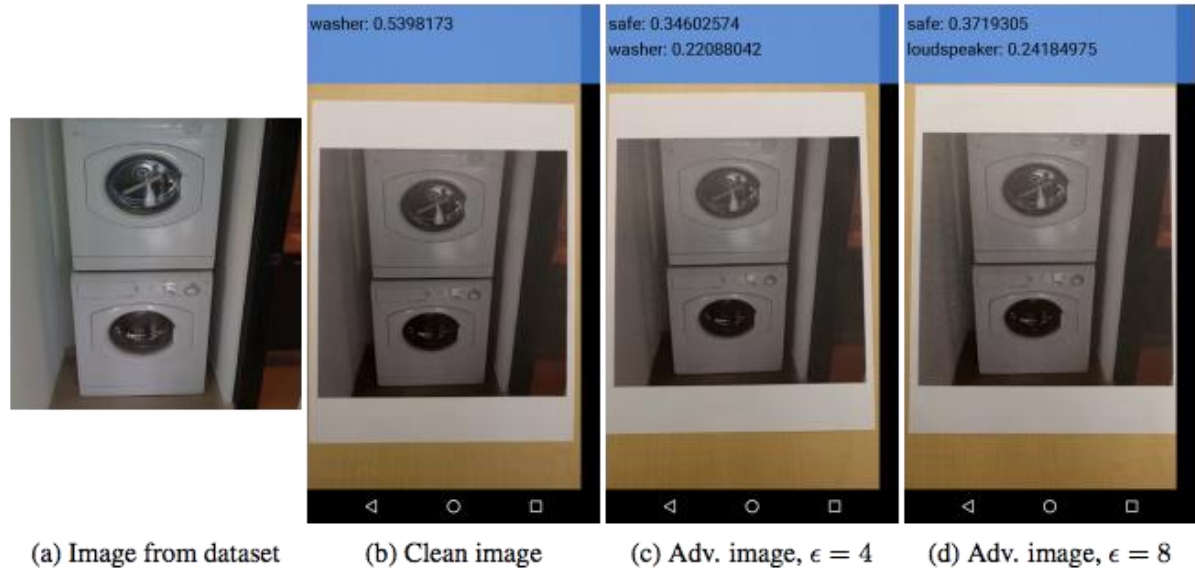
Quellen: <https://www.pinterest.com/explore/artificial-neural-network/>
Olah, et al., "Feature Visualization", Distill, 2017, <https://distill.pub/2017/feature-visualization/>.

Wie schlussfolgert die Maschine?

«Debugging» für Einblicke in die vermeintliche «Black Box»

Verdeutlichen ein Problem:

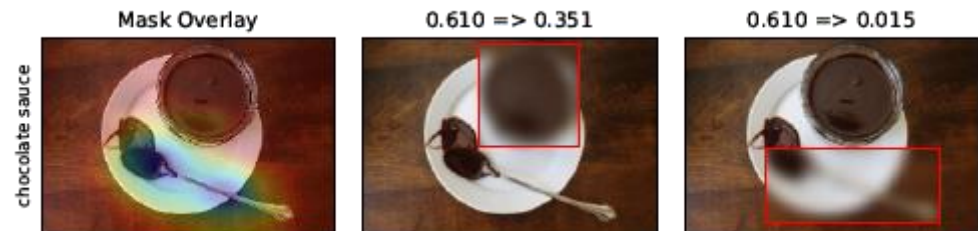
- Adversarial Examples



<https://blog.openai.com/adversarial-example-research/>

Bieten eine Lösung:

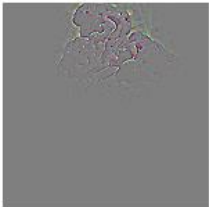
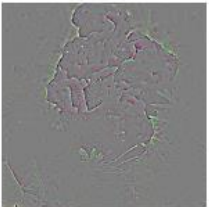
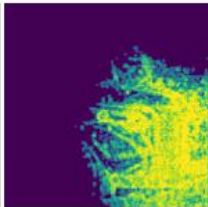
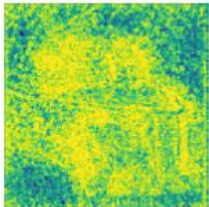
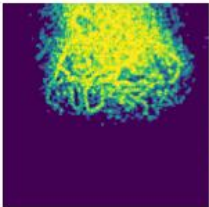
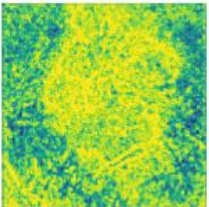
- Saliency Maps



Ruth C. Fong & Andrea Vedaldi, «Interpretable Explanations of Black Boxes by Meaningful Perturbation», 2017

Trace & detect adversarial attacks

...using average local spatial entropy of feature response maps

	Original	Adversarial	Original	Adversarial
Image:				
Feature response:				
Local spatial entropy:				

Amirian, Schwenker & Stadelmann (2018). «Trace and Detect Adversarial Attacks on CNNs using Feature Response Maps». ANNPR'2018.