

C-Level Intensivseminar KI & Machine Learning

SUVA, 14. Oktober 2019



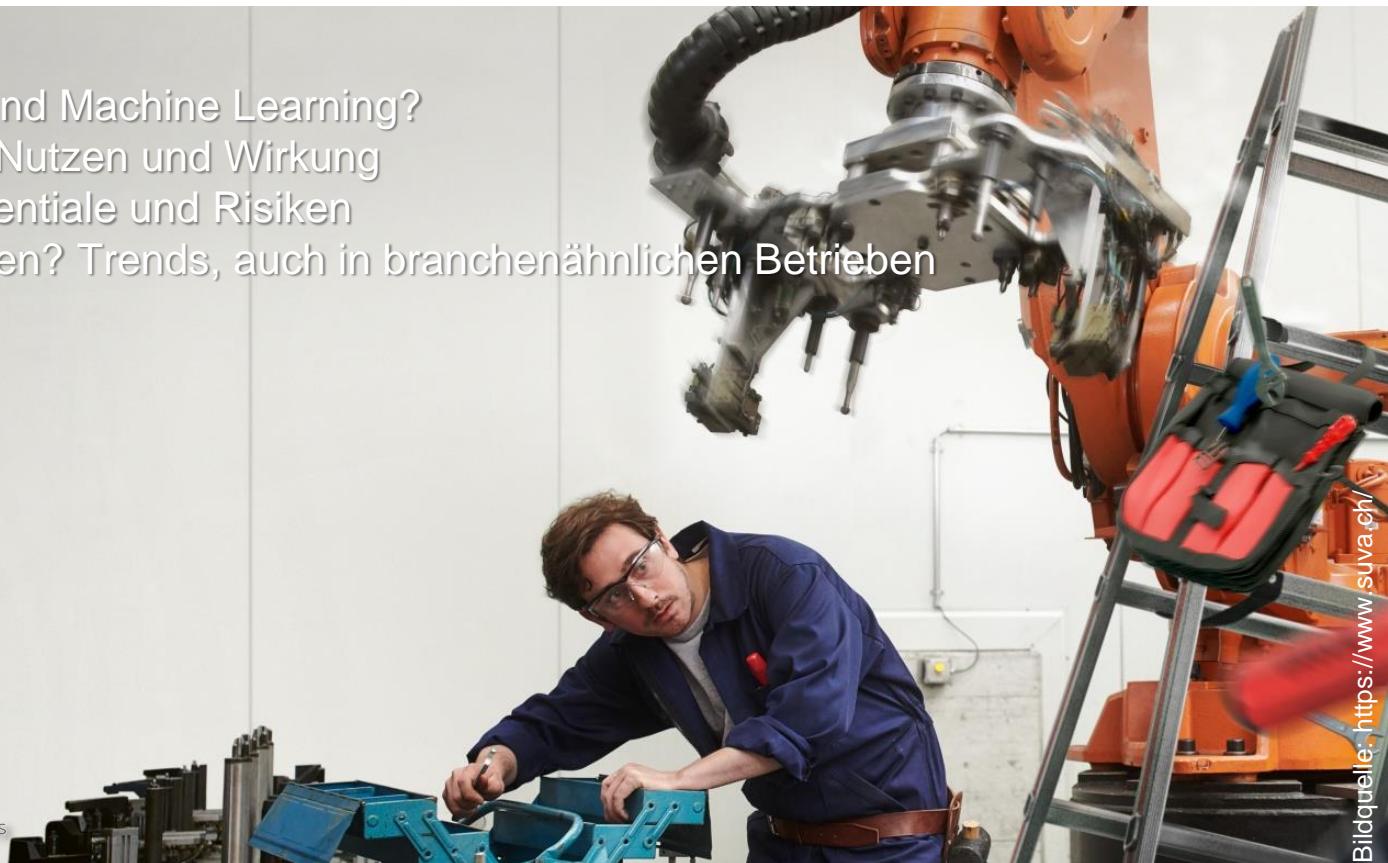
Thilo Stadelmann

Was versteht man unter KI und Machine Learning?

Warum macht man das? → Nutzen und Wirkung

Wozu führt das jetzt? → Potentiale und Risiken

Wohin kann das einmal führen? Trends, auch in branchenähnlichen Betrieben



Swiss Alliance for
Data-Intensive Services

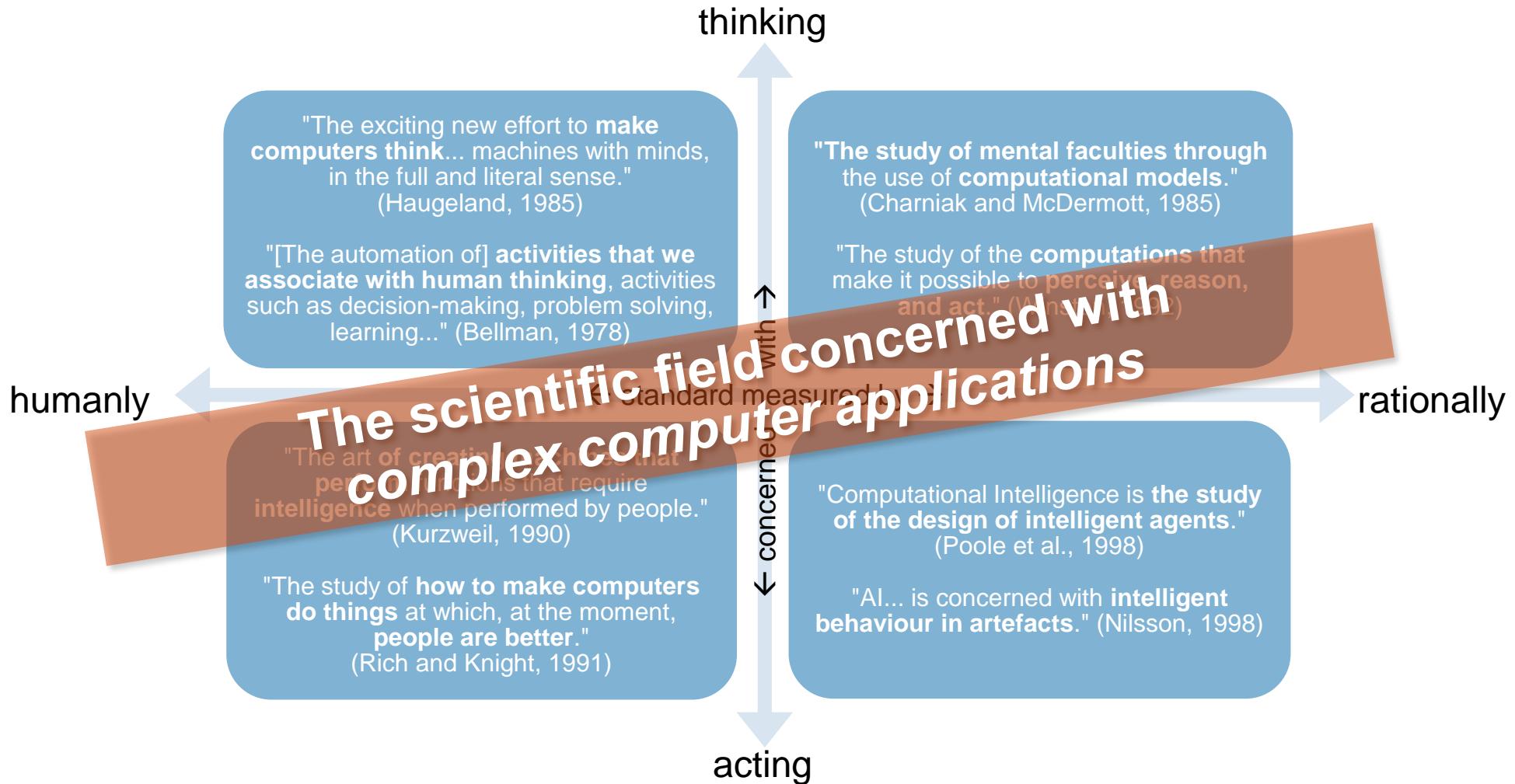
Was → Warum? → Wozu? → Wohin?



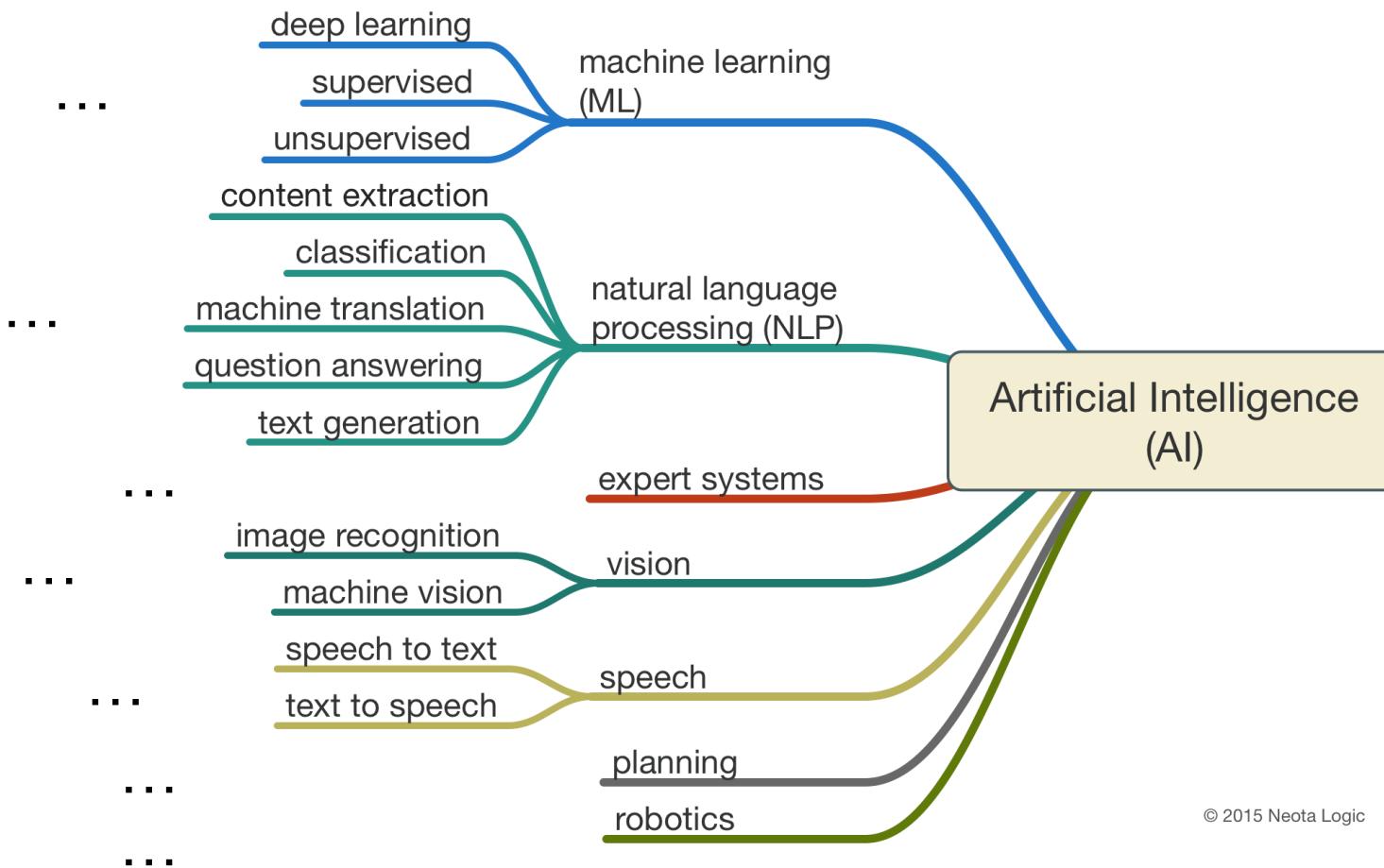
1

Was versteht man unter KI und Machine Learning?

Was ist künstliche Intelligenz?

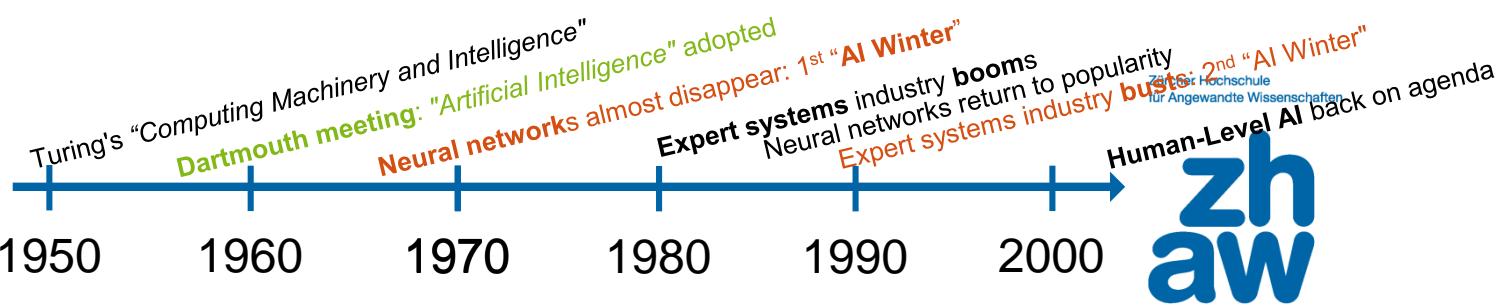


Was gehört zu künstlicher Intelligenz?



© 2015 Neota Logic

KI im Kontext



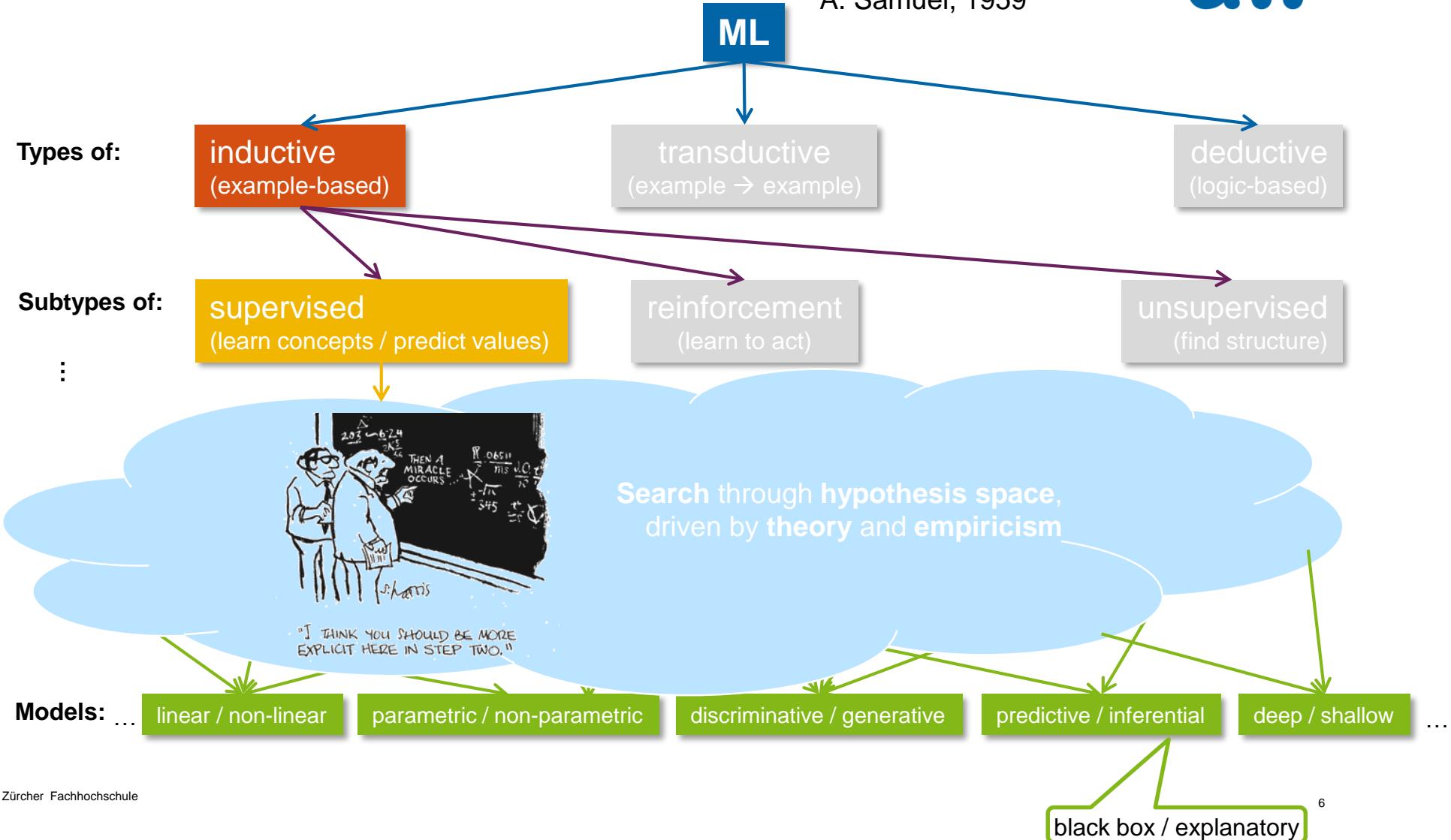
zhaw
Zürcher Hochschule
für Angewandte Wissenschaften



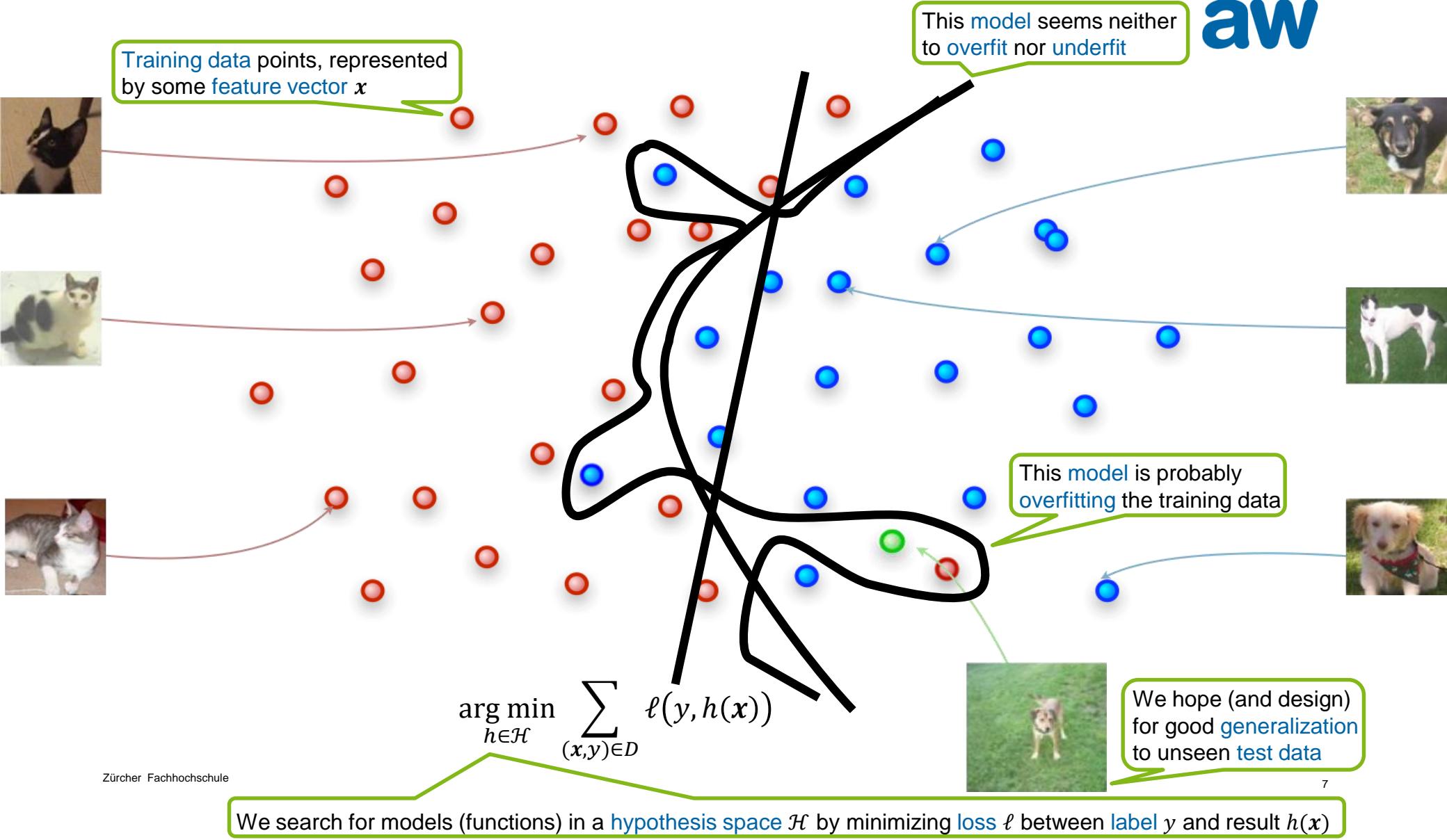
Eine Machine Learning Landkarte

«...gives computers the ability to learn **without being explicitly programmed.**»

A. Samuel, 1959



Supervised Machine Learning im Überblick



Grundprinzip im Deep Learning: Feature Vektoren (Merkmale) *automatisch* lernen

Classical image processing

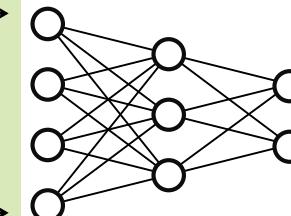


Feature extraction
(SIFT, SURF, LBP, HOG, etc.)

(0.2, 0.4, ...)

(0.4, 0.3, ...)

Classification
(SVM, neural network, etc.)



Container ship

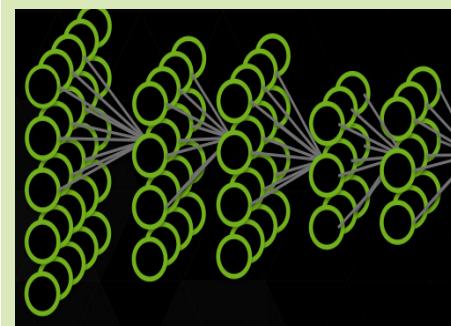
Tiger

...

Using Convolutional Neural Networks (CNNs)



Takes raw pixels in, learns features automatically!



Container ship

Tiger

...

Was → Warum? → Wozu? → Wohin?



2

Warum macht man das? → Nutzen und Wirkung

Zwischenfazit: Einsatzmöglichkeiten & Erfolgsfaktoren



KI: maschinelles Lösen von komplexen (=kann bisher nur der Mensch) Aufgabenstellungen

ML: *ein* Werkzeug für KI; findet Lösungsweg anhand Input-Output Beispielen von Menschen

Zwischenfazit: Einsatzmöglichkeiten & Erfolgsfaktoren

KI: maschinelles Lösen von komplexen (=kann bisher nur der Mensch) Aufgabenstellungen

ML: *ein* Werkzeug für KI; findet Lösungsweg anhand Input-Output Beispielen von Menschen

→ **Automatisierung komplexer, redundanter Prozesse basierend auf (hoch-dim. Sensor-)Daten**



Beispiele aus der angewandten Forschung ...mit lokalen Industriepartnern (KMUs)



Gesichtserkennung für Stadionzutritt

- Nutzen: *Robustes* Personenidentifikationssystem
- Wirkung: Unterstützung bei Entwicklung; Datenqualität schränkte ein

[!] DEEPIMPACT



Automatische Artikelsegmentierung

ARGUS DATA INSIGHTS[®]
WISSEN FÜR EXPERTEN

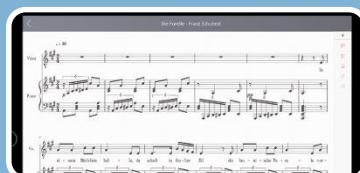
- Nutzen: vollautomatisches Produkt in niedrigem Preissegment
- Wirkung: Einführung dank *Teamausbau* geglückt



Visuelle Qualitätskontrolle in Produktion

BW-TEC[®]
INDUSTRIE • INNOVATION • CONSULTING

- Nutzen: vollautomatischer Triage & Bearbeitung normaler Fälle
- Wirkung: macht *Familienunternehmen* zu Technologieanbieter



Digitalisierung von Musikalien

SCOREPAD

- Nutzen: Enabler für digitales Geschäftsmodell
- Wirkung: 5 Jahre nach Start ist entwickelte Technologie *grösstes Asset*

Was → Warum? → Wozu? → Wohin?



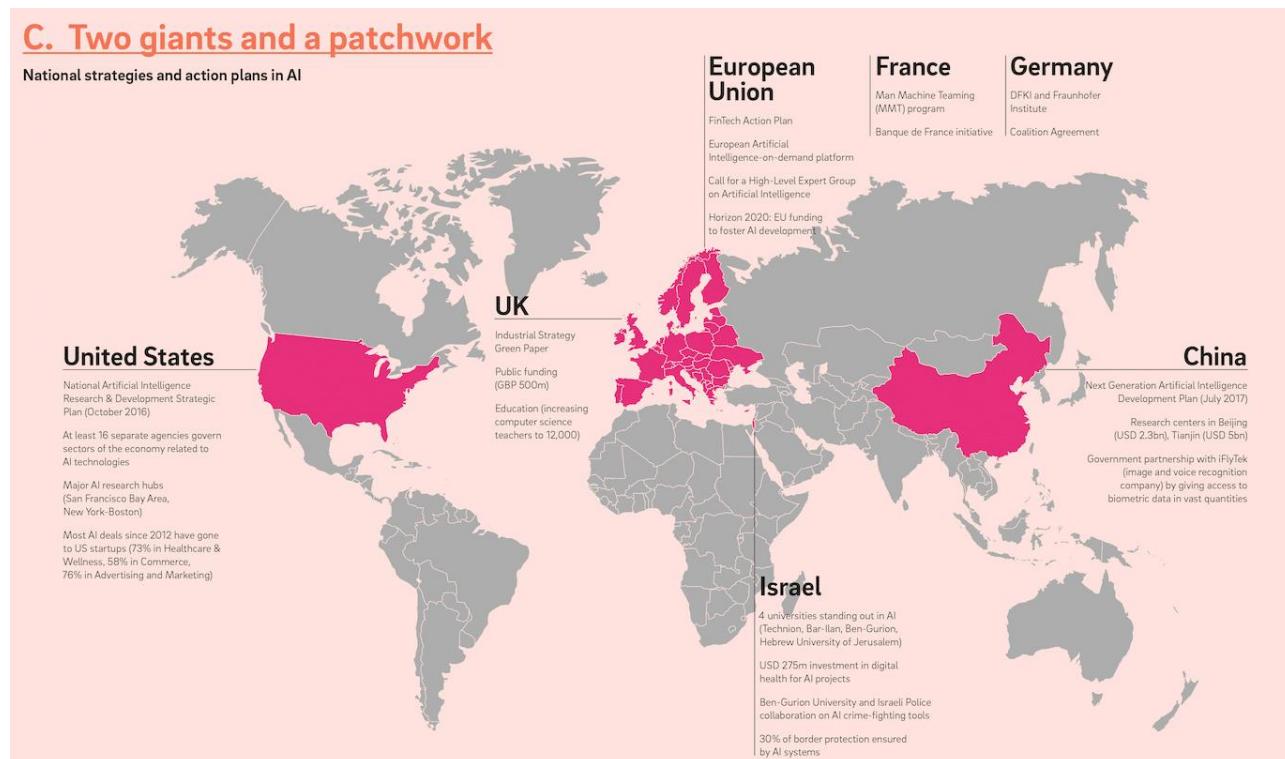
3

Wozu führt das jetzt? → Potentiale und Risiken

Globale Relevanz künstlicher Intelligenz

Marktgrösse (Hardware, Software, Services): in 2018 ca. \$21.5 Mrd. → \$190.6 Mrd. in 2025¹
Nationale Strategien²:

Talente, Forschung, ...

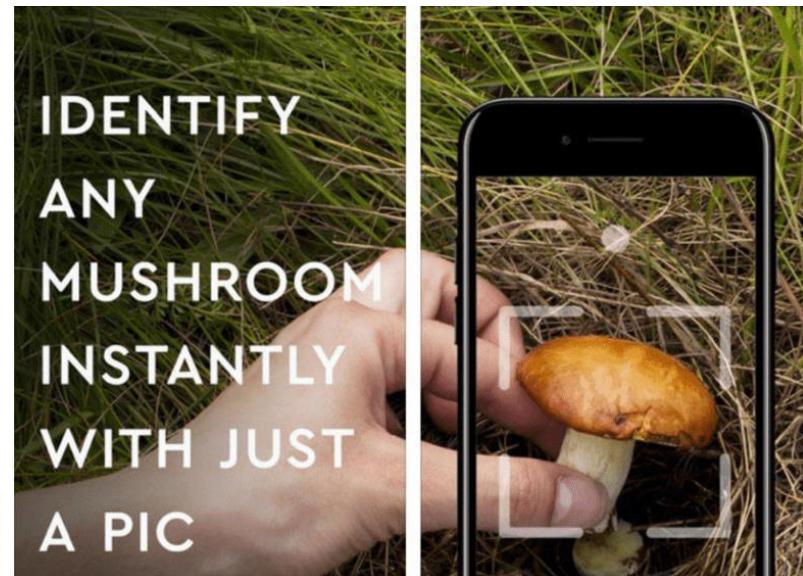
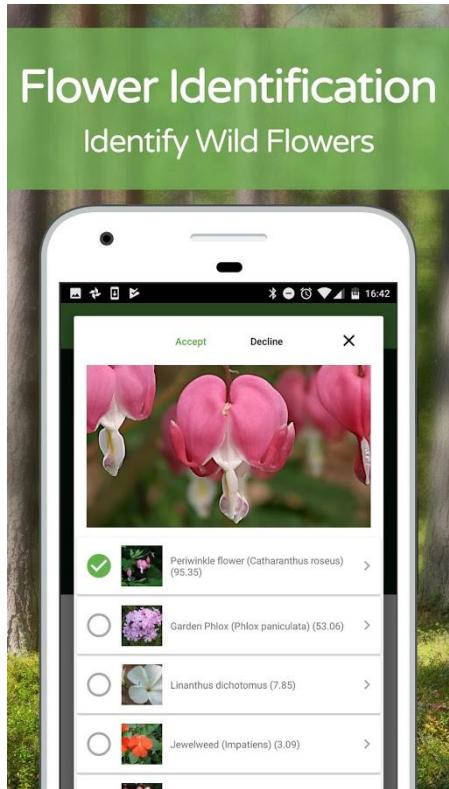


1) Siehe <https://www.marketsandmarkets.com/PressReleases/artificial-intelligence.asp> (2017)

2) Siehe <https://asgard.vc/global-ai/> (2017)

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** by
Gerd Gigerenzer
£20.95

What Do Customers Ultimately Buy After Viewing This Item?



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



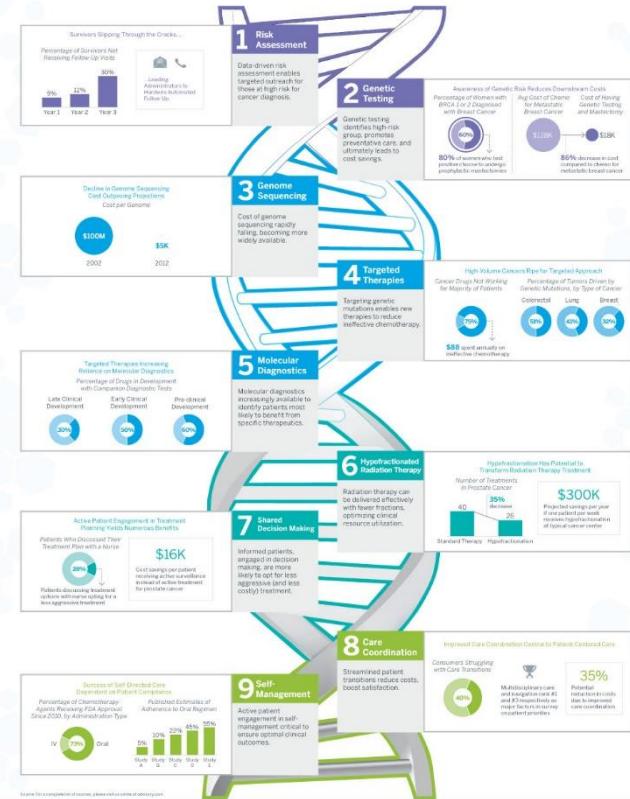
**17% buy
Gut Feelings: Short Cuts to Better Decision Making**
£6.74



**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.



Beispiel: Statistik vs. Bias

Technologie: Machine Learning

English – detected ▾

Turkish ▾

He is a babysitter Edit

O bir bebek bakıcısı

Turkish – detected ▾

English ▾

O bir bebek bakıcısı

She's a babysitter

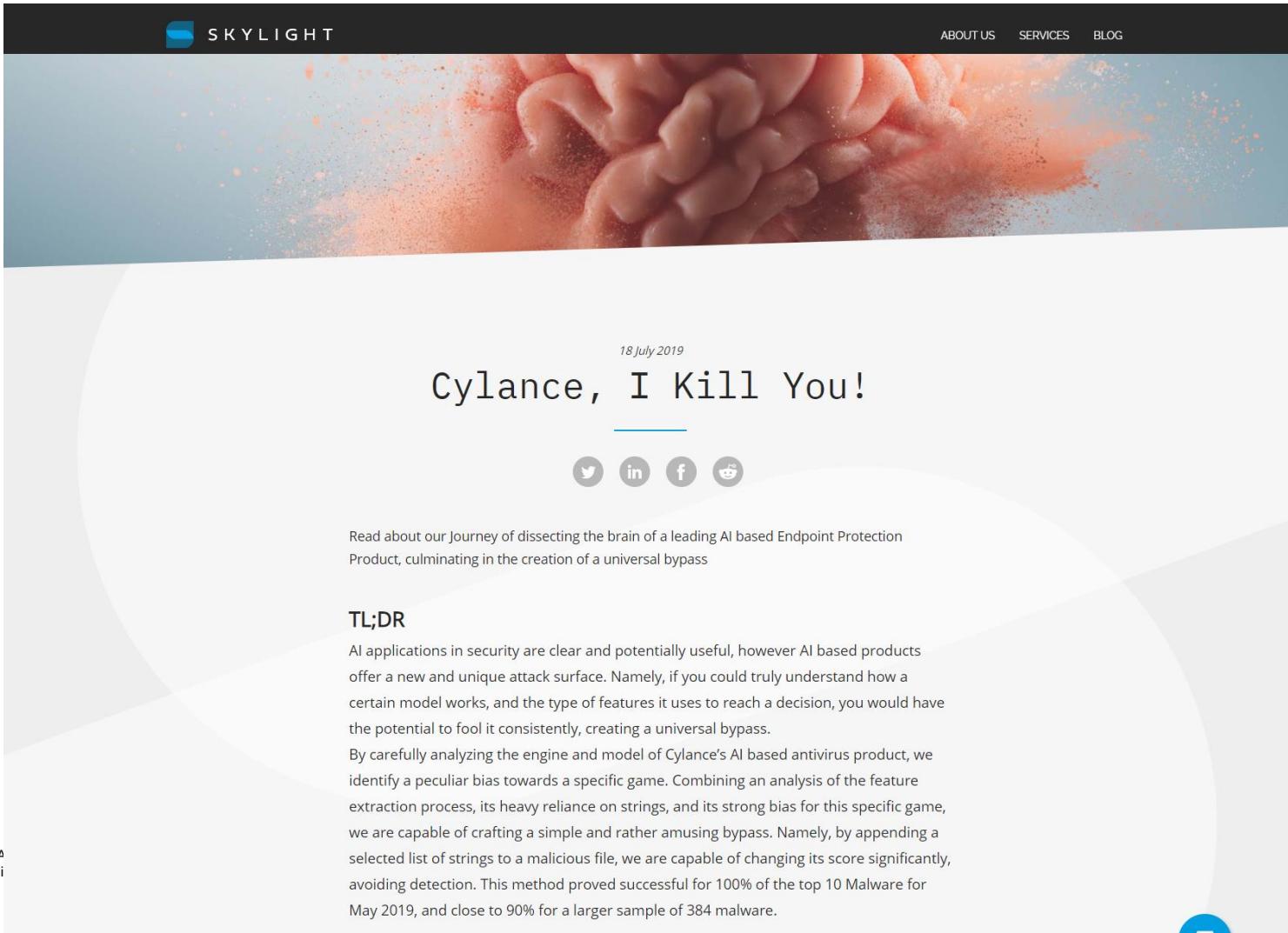
unprofessional hairstyles

professional hairstyles

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

Beispiel: künstl. Intelligenz vs. natürl. Dummheit

Technologie: Machine Learning mit nachgelagerten Regeln



A screenshot of a blog post from Skylight. The header features the Skylight logo (a blue square icon followed by the word "SKYLIGHT") and a navigation menu with links to "ABOUT US", "SERVICES", and "BLOG". The main image is a close-up, artistic photograph of a human brain with orange and red dust or powder being poured over it, set against a light blue background. Below the image, the date "18 July 2019" is displayed, followed by the title "Cylance, I Kill You!". A horizontal line separates the title from social sharing icons: Twitter, LinkedIn, Facebook, and Reddit. The text below the title reads: "Read about our Journey of dissecting the brain of a leading AI based Endpoint Protection Product, culminating in the creation of a universal bypass". A section titled "TL;DR" follows, containing two paragraphs of text about AI applications in security and a specific analysis of Cylance's AI-based antivirus product.

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.

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:



A dark grey rectangular document cover. At the top, it lists several organizations: 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, and OpenAI. Below this, the title "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation" is centered in white text. A date "February 2018" is in the top right corner. The bottom half of the cover features a grid pattern of white symbols: dashes, slashes, dots, and plus signs, arranged in a roughly 10x10 grid.

Was → Warum? → Wozu? → Wohin?



4

Wohin kann das einmal führen? Trends, auch in branchenähnlichen Betrieben

Trend: Entwickeln für “algorithmic fairness”

Der 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:
at design stage, pre-launch, post-launch

Trend: Entwickeln für Interpretierbarkeit

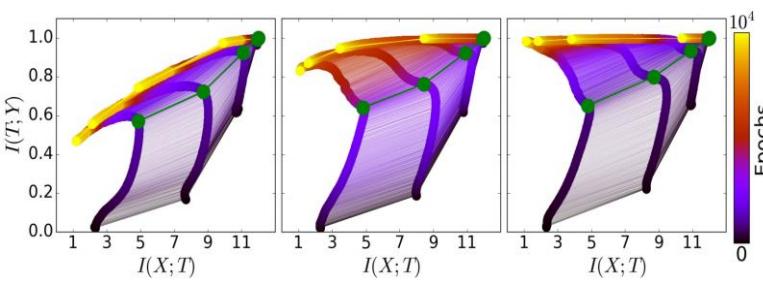
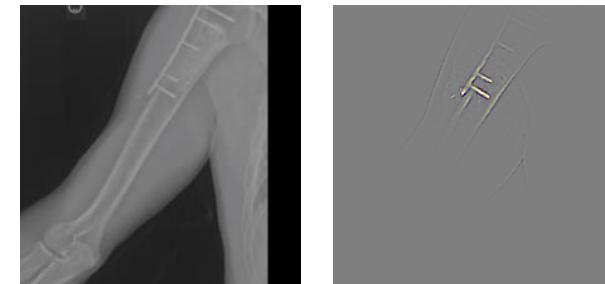
Interpretability is required.

- Helps the developer in «debugging», needed by the user to trust
→ visualizations of learned features, training process, learning curves etc. should be «always on»

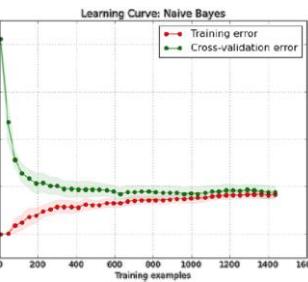
negative X-ray



positive X-ray



DNN training on the Information Plane



a learning curve



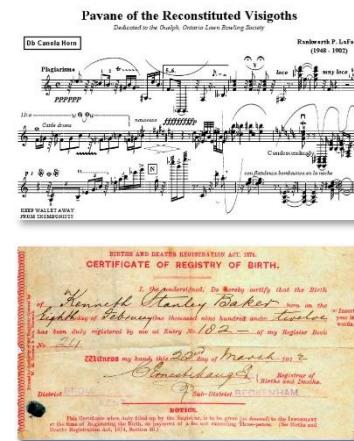
feature visualization

Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). «Deep Learning in the Wild». ANNPR'2018.

Schwartz-Ziv & Tishby (2017). «Opening the Black Box of Deep Neural Networks via Information».

<https://distill.pub/2017/feature-visualization/>, <https://stanfordmlgroup.github.io/competitions/mura/>

Trend: “Document recognition” anstatt Analyse rein strukturierter Daten



Documents

- Ubiquitous** in human communication and every scenario involving an office
- Somewhat structured for human expert; **unstructured** w.r.t machines
- Great use case** for various **AI** techniques, including computer vision techniques

Own scientific community

- IAPR's biannual Intl. Conference on Document Analysis & Recognition (ICDAR): character & symbol recognition, printed/handwritten text recognition, graphics analysis & recognition, document analysis & understanding, historical documents & digital libraries, document based forensics, camera & video based scene text analysis

Beispiele aus der (lokalen) Wirtschaft

This site uses tracking cookies in conjunction with Google Analytics to improve your browsing experience. Do you agree to the use of tracking cookies?

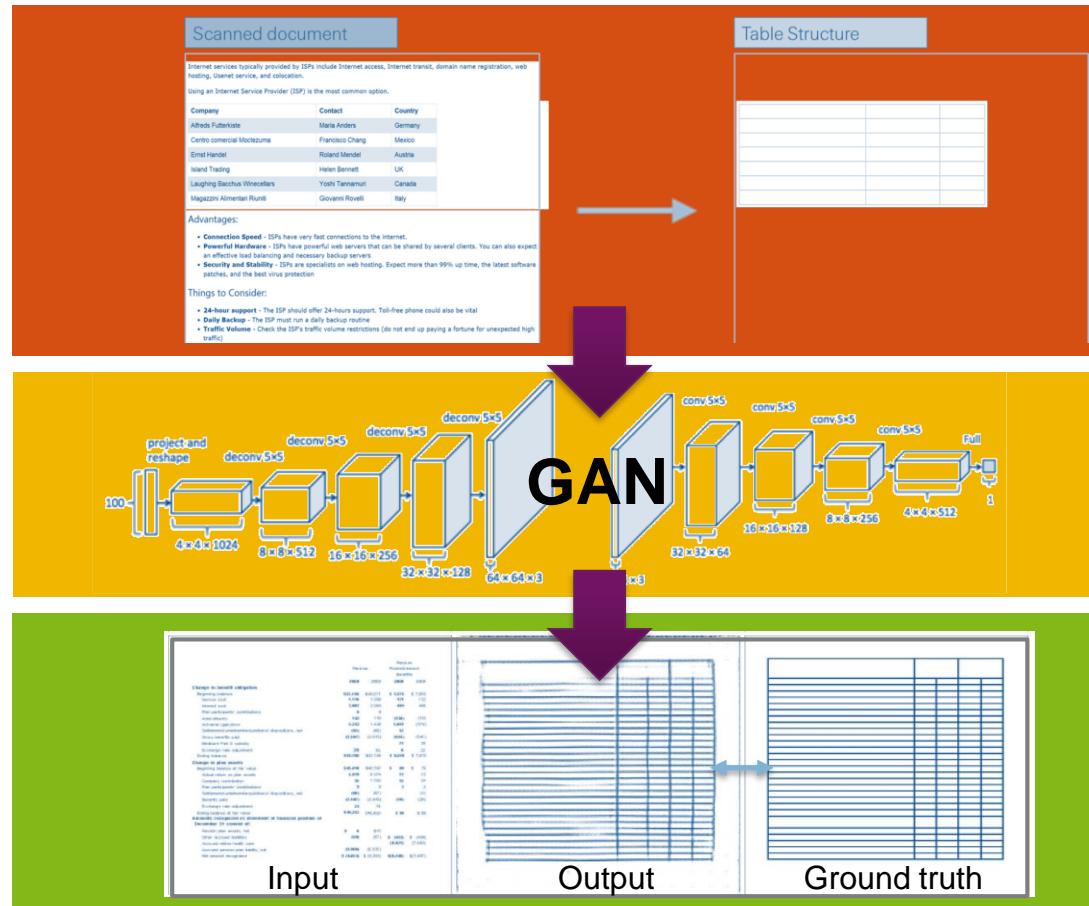
Yes, I agree No

Process your Bank Statements automatically.
Unlock the full potential of your documents using machine learning.

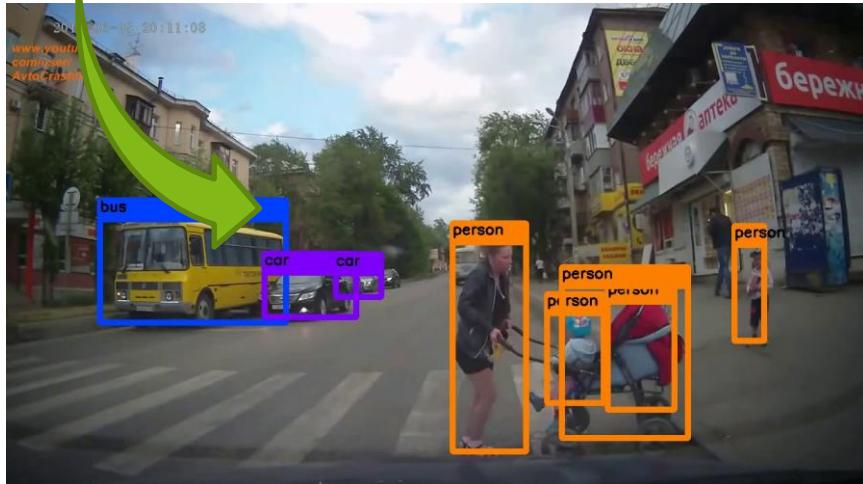
Learn more Contact us

MINT.extract — Truly refreshing document digitalization.

Is your information trapped inside documents? Meet MINT.extract: Our solution to free your data.



Beispiele aus der (lokalen) Wirtschaft (contd.)



R C EDUCATION SOCIETY (B.E.D) DHANOT, KANGRA, HIMACHAL PRADESH INCOME AND EXPENDITURE ACCOUNT FOR THE YEAR ENDING ON 31ST MARCH, 2017		
EXPENDITURE	AMOUNT	INCOME
10 INDIRECT EXPENSES		BY INCOME
AUDIT FEES	12,000.00	ADMISSION FEES
BRANDS/DRUGS	1,994.00	AMMAGAMATED FUND
CAR INSURANCE	4,930.00	BOOK PURCHASE
CATERING AND MAINTAINANCE	1,000.00	COLLEGE BADGE
COLLEGE CANTINEN EXP	21,304.00	COFFEE DRINKS
COMPUTER EXP	5,089.00	COURSES
CONTRIBUTION EXP	34,820.00	CULTURAL ACTIVITIES FUND
DEPARTMENTAL EXP	14,232.00	DEPRICATION OF PRIOR PERIOD
DEVONIMENT CHARGE	8,855.00	DEPRECIATION OF PROPERTY
EDUCATIONAL TOUR EXP	24,710.00	H.P.U LEVY CHARGES OF B.E.D 2ND YEAR
FESTIVAL EXP	1,700.00	INTERNSHIP EXP
FIRE AND WATER EXP	18,568.00	IDENTITY CARD FEES
H.P.U R.E.C.C. MAINTENANCE FEES	9,380.00	INDUSTRY EXP
H.P.U EXAM	26,000.00	INTT. ON SAVING AC 1657
H.P.U LIBRARY EXP	1,080.00	LIBRARY READING ROOM FEE
H.R.V ADD EXP	77.00	MANAGEMENT EXP
H.R.V RECORDING ROOM EXP	10,330.00	MEDICAL FUND
MEDICAL EXP	1,120.00	MEMBERSHIP FEE
MREC EXP	23,560.00	PURCHASE OF EQUIPMENTS AND MATERIALS
POSTAGE CHARGES	1,770.00	SALE OF PROSPECTUS
PROFESSIONAL STATIONERY EXP	8,640.00	SOCIETY FUND
PROFESSIONAL CHARGES	8,400.00	STAFF EXP
REPAIRS AND MAINTAINANCE	3,000.00	STUDENT ACTIVITY
REPAIRS TO FAF	20,076.00	STUDENTS SCHOLARSHIP FUND
STAFF SALARY	18,129.00	TUTION FEES
STUDENTS ACTIVITIES	15,170.00	
TELEPHONE CHARGES	150.00	
TRAVEL AND LOCAL TOUR	54,480.00	
WARCH AND VARDI EXP	1,84,432.00	
W.E.S.T. EXP	200.00	
WEBS ITINERARY EXP	4,800.00	
10 DEPRIFICATION	16,14,11.00	
10 SURPLUS	18,21,459.00	
	86,87,344.00	

Schedules 1 to 7 form an integral part of accounts.

For R C EDUCATION SOCIETY (B.E.D)

ARUN GIRDHAR LAL
(PRESIDENT)

Place : HAMIRPUR (H.P)
Date : 25/02/2018

In terms of our attached report of ever
For ARUN GIRI AND ASSOCIATES
CHARTERED ACCOUNTANTS

ARUN GIRI,
(PROPRIETOR)
M. NO : 095743
FRN : 016251N

BALANCE SHEET	
Company Name	
ASSETS	DATE
Current Assets	2012 2013
Cash	1,200 1,400
Temporary Investments	
Investments	
Accounts receivable	
Prepaid expenses	
Other	
Total Current Assets	1,200 1,400
Fixed Assets	
Property, plant and equipment	
Less accumulated depreciation	- 300 - 195
Less accumulated depreciation (Negative Value)	- 300 - 195
Total Assets	900 1,205
Other Assets	
Deferred income tax	
Charity/Goodwill	
Other	
Total Other Assets	- -
TOTAL ASSETS	900 1,205
LIABILITIES AND OWNER'S EQUITY	
Current Liabilities	
Accounts payable	350
Accrued expenses	600
Accrued Income	300
Total Current Liabilities	900 350

Zwischenfazit: Entkopplung



Grösse der Idee ≠ Grösse des Unternehmens

...KMUs können **bauen was auch immer sie mögen**
(gegeben Know-how, Daten und einen interessanten Business Case)

Technologie ist branchenunabhängig

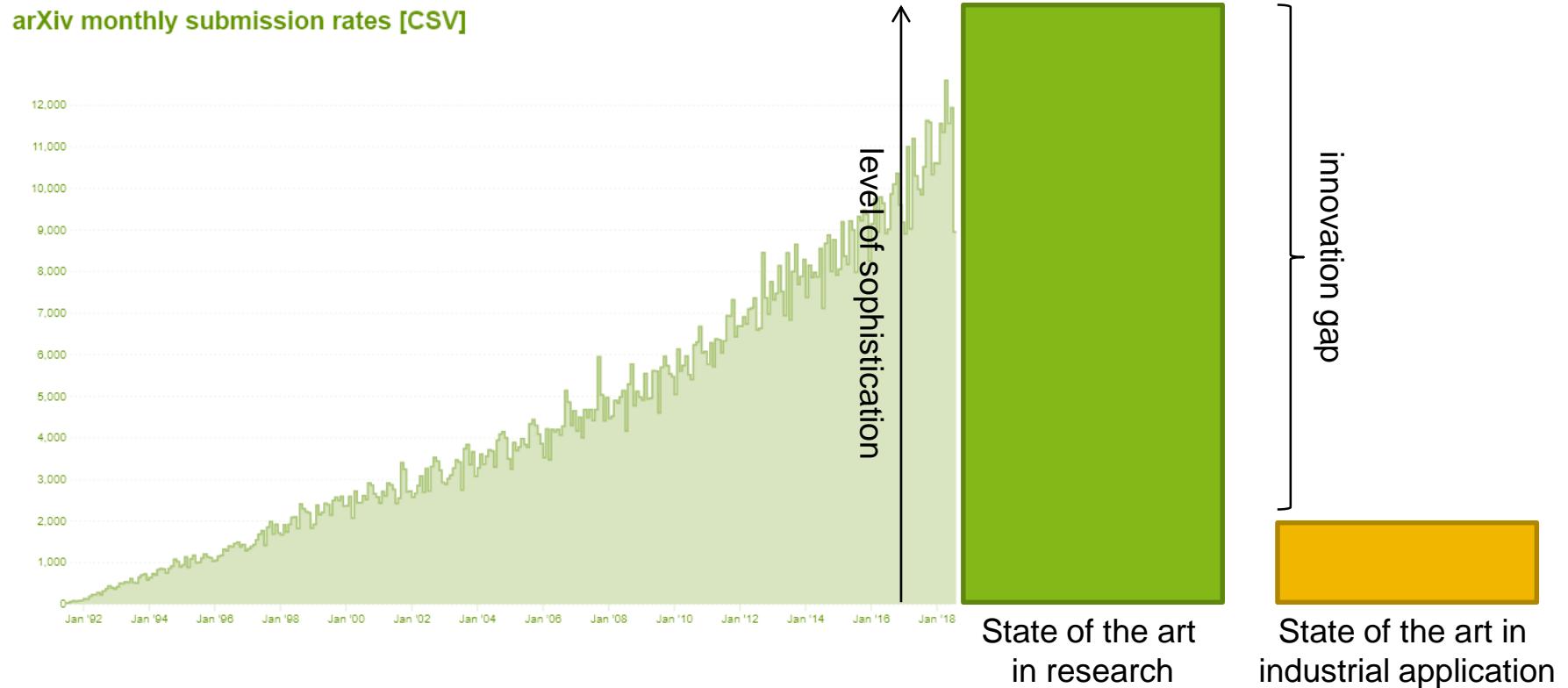
...was **neue** Kooperationen und Allianzen ermöglicht, z.B.



Swiss Alliance for
Data-Intensive Services

Zwischenfazit (contd.): Geschwindigkeit

Durchschnittliche Zeit von Publikation bis Anwendung im Projekt: ca. 3 Monate



Aussicht: Disruption

...selbst bei völliger Stagnation des technischen Fortschritts

1. Hypothese: Einsatz (aktueller) KI wird sich massiv ausbreiten (Zeitrahmen: 5 Jahre)
 - Indikator: **KI-Fortschritt** momentan hauptsächlich **Industriegetrieben (Gewinnaussicht)**; Konsumenten kaufen "bequem"; diese Incentivierung "hält den Motor am Laufen"
2. Hypothese: Dies wird unsere Gesellschaften umwälzen
 - Kernfragen: Wie **verteilt** sich der algorithmisch (hauptsächlich bei Grosskonzernen) erwirtschaftete **Gewinn**? Wie verteilt sich neue **Freizeit** und **Alltagserleichterung**?
3. Hypothese: Grösste Frage wird der Umgang miteinander sein (nicht der Umgang mit KI)
 - Argument: KI (etc.) "for the common good" ist ein wichtiges Thema; entscheidend wird jedoch sein, wie **wir als Gesellschaften die Regeln** für das digitalisierte Zusammenleben (s.o.) **gestalten**



Siehe auch: 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.

Schlussfolgerungen



- KI automatisiert *einzelne*, komplexe, aber *redundante* Prozesse (meist mittels ML auf menschengenerierten Beispielen)
- Deep Learning hat zu Paradigmenwechsel in *Mustererkennungsaufgaben* geführt
- Die Zeit vom Grundlagenresultat zur praktischer Anwendung beträgt wenige Monate
- Das Zeitfenster zu handeln ist beträgt wenige Jahre (<5) → Disruption



Swiss Alliance for
Data-Intensive Services

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
- Zusammenarbeit: datalab@zhaw.ch