Zürcher Hochschule
für Angewandte Wissenschaften

C-Level Intensivseminar KI & Machine Learning

SUVA, 14. Oktober 2019

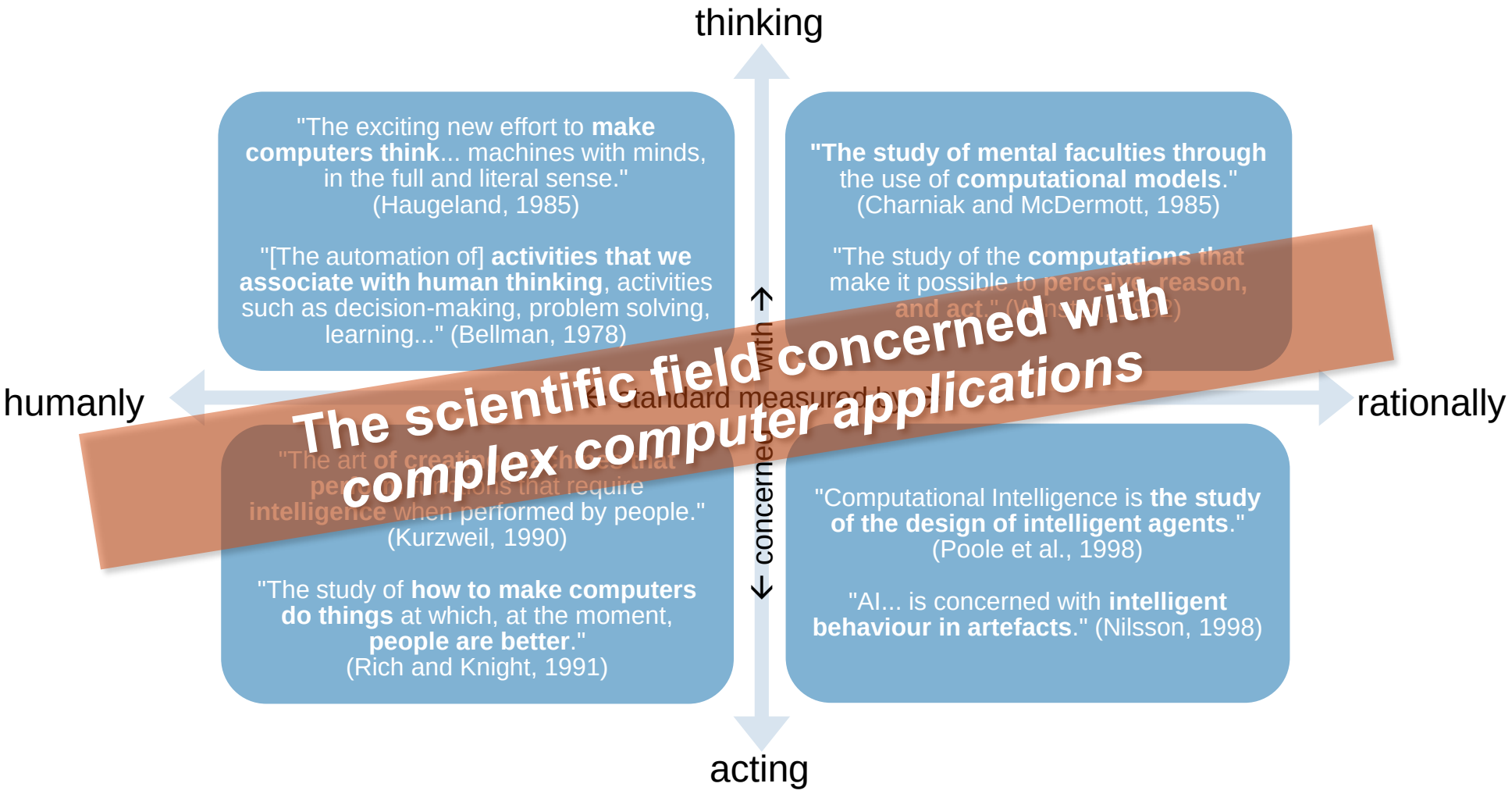
Thilo Stadelmann

Was→ Warum?→ Wozu?→ Wohin?

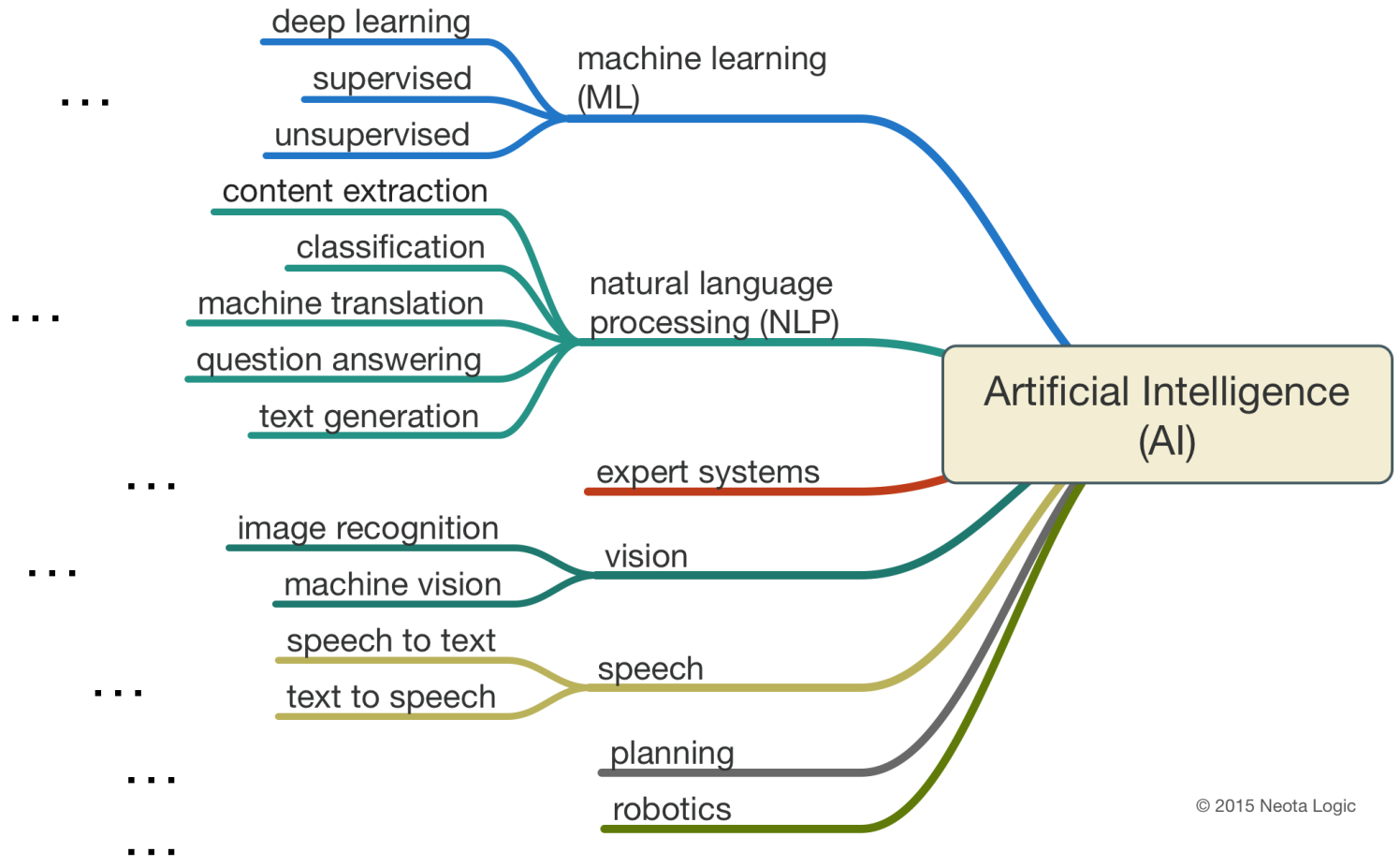
1

Was versteht man unter KI und Machine Learning?

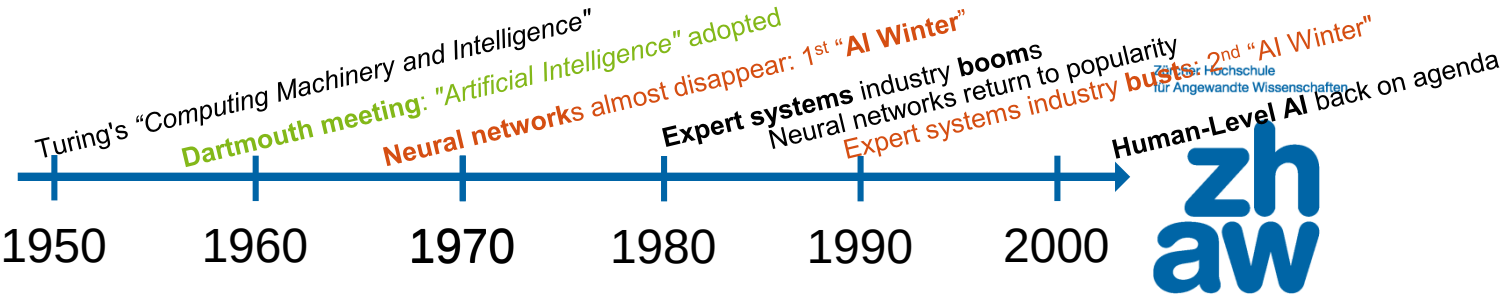
Was ist künstliche Intelligenz?



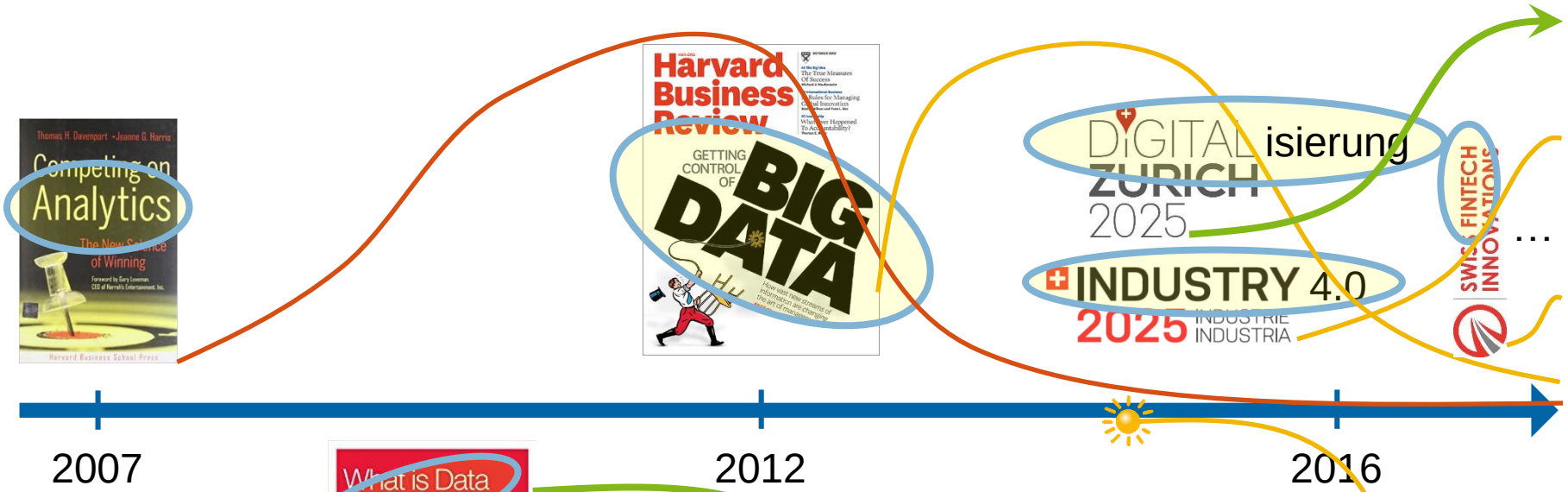
Was gehört zu künstlicher Intelligenz?

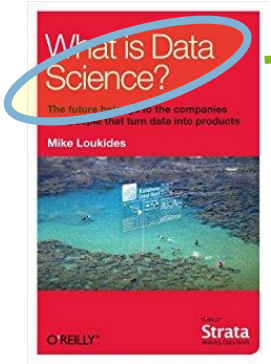


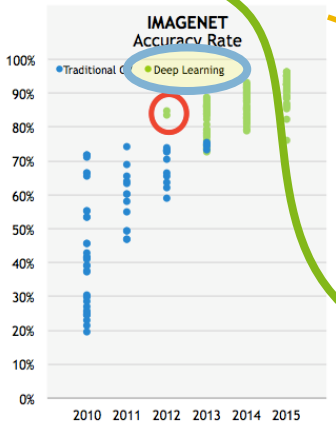
KI im Kontext

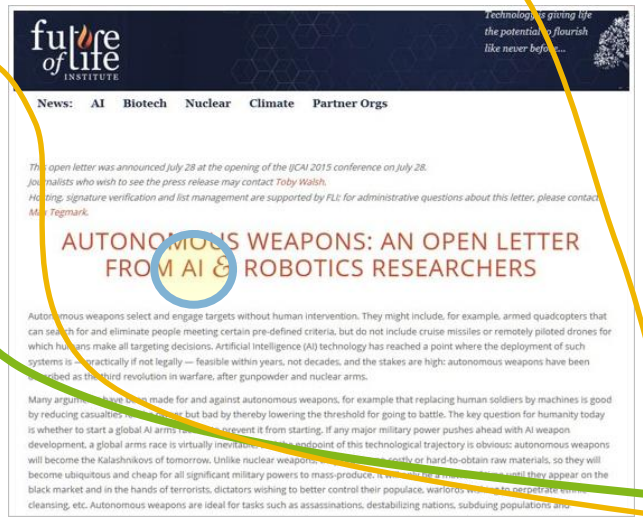


Zürcher Hochschule für Angewandte Wissenschaften









Eine Machine Learning Landkarte

*«…gives computers the ability to learn **without being explicitly programmed**.»*
A. Samuel, 1959

ML

Types of:

- inductive (example-based)
- transductive (example → example)
- deductive (logic-based)

Subtypes of:

- supervised (learn concepts / predict values)
- reinforcement (learn to act)
- unsupervised (find structure)

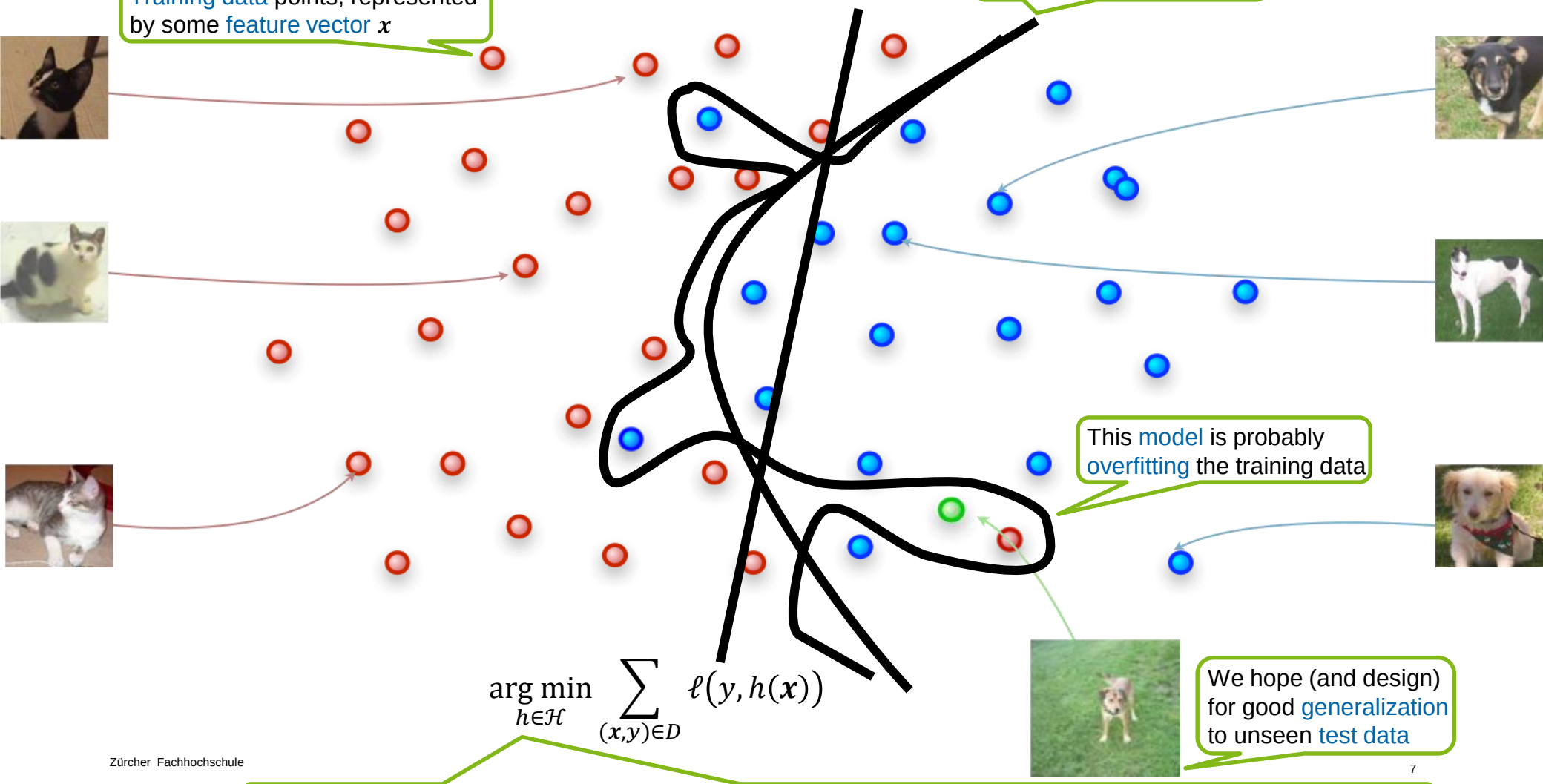
⋮

Search through **hypothesis space**,
driven by **theory** and **empiricism**

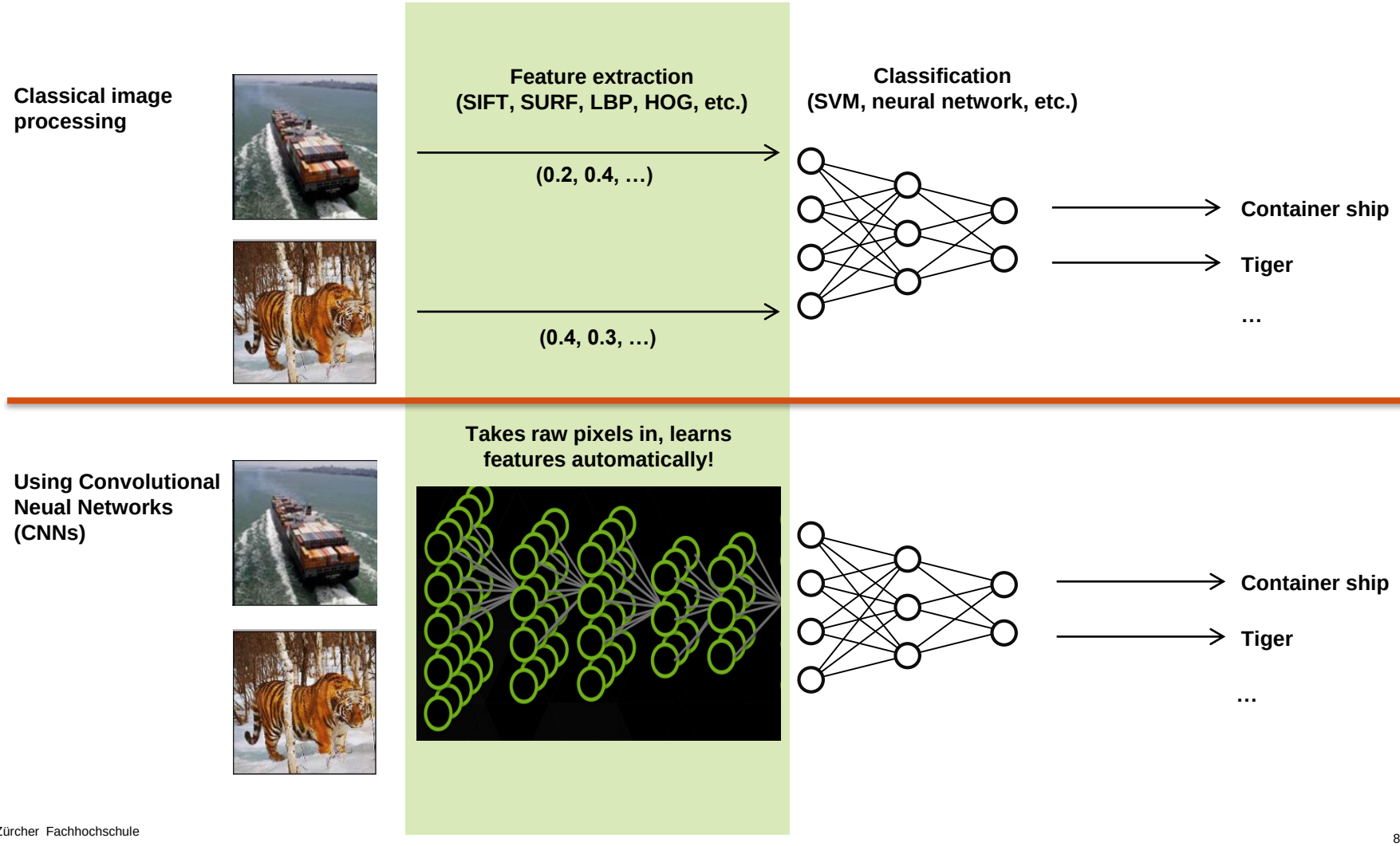
Models: … linear / non-linear | parametric / non-parametric | discriminative / generative | predictive / inferential | deep / shallow …

black box / explanatory

Supervised Machine Learning im Überblick



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



Zürcher Hochschule
für Angewandte Wissenschaften

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

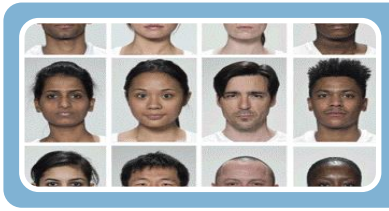
KI: maschinelles Lösen von komplexen (=kann bisher nur der Mensch) Aufgabenstellungen

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)

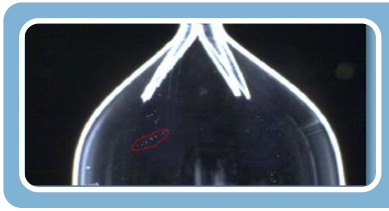


Gesichtserkennung für Stadionzutritt

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

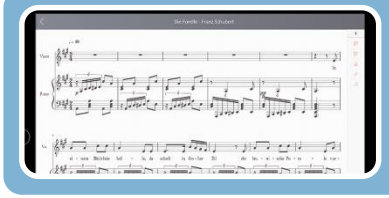
Automatische Artikelsegmentierung

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



Visuelle Qualitätskontrolle in Produktion

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



Digitalisierung von Musikalien

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

Was→ Warum?→ Wozu?→ Wohin?

3

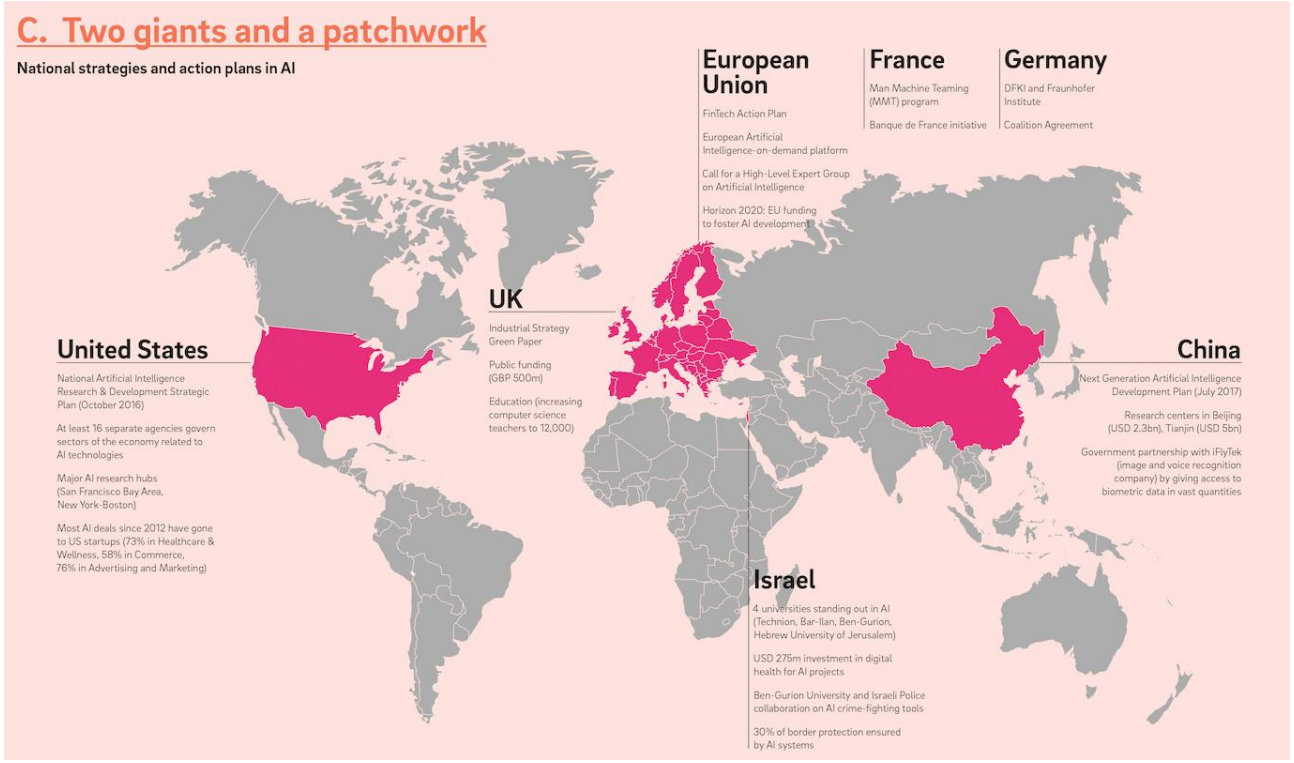
Wozu führt das jetzt? → Potentiale und Risiken

Globale Relevanz künstlicher Intelligenz

Marktgrösse (Hardware, Software, Services): in 2018 ca. $21.5 Mrd. → $190.6 Mrd. in 2025[1]

Nationale Strategien[2]:

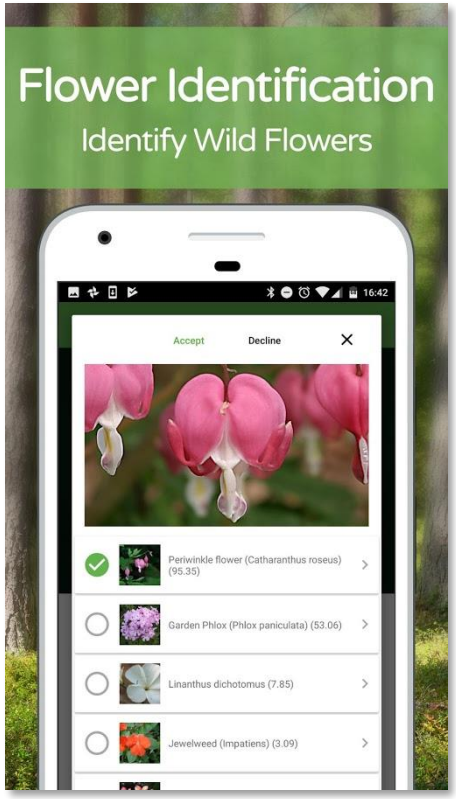
Talente, Forschung, …

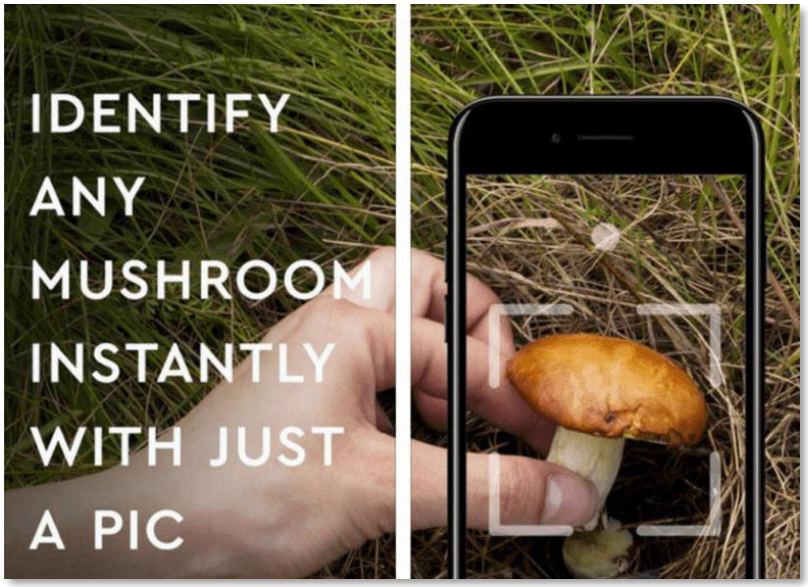


1) Siehe https://www.marketsandmarkets.com/PressReleases/artificial-intelligence.asp (2017)
2) Siehe https://asgard.vc/global-ai/ (2017)

Beispiel: Machbar vs. gefährlich

Technologie: Computer Vision mit Deep Learning

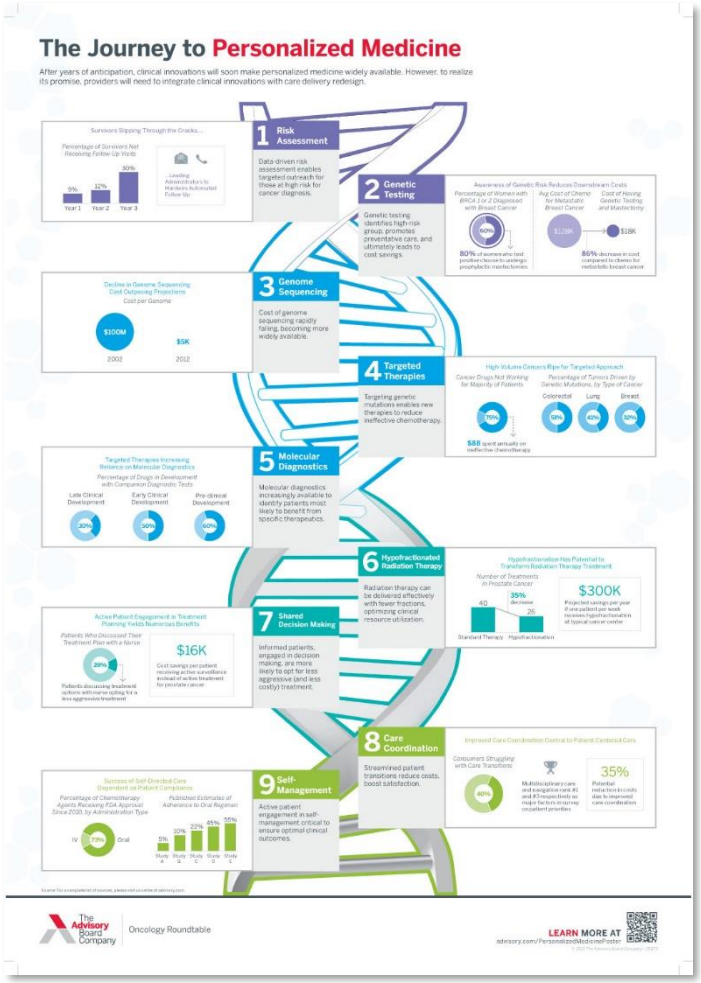




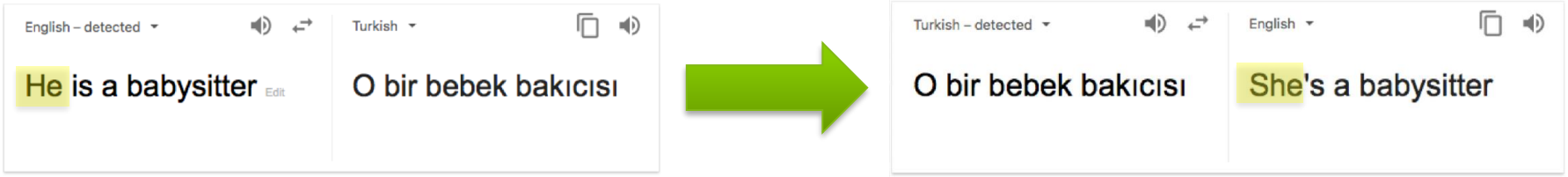
https://www.cultofmac.com/495088/avoid-potentially-deadly-ai-app/

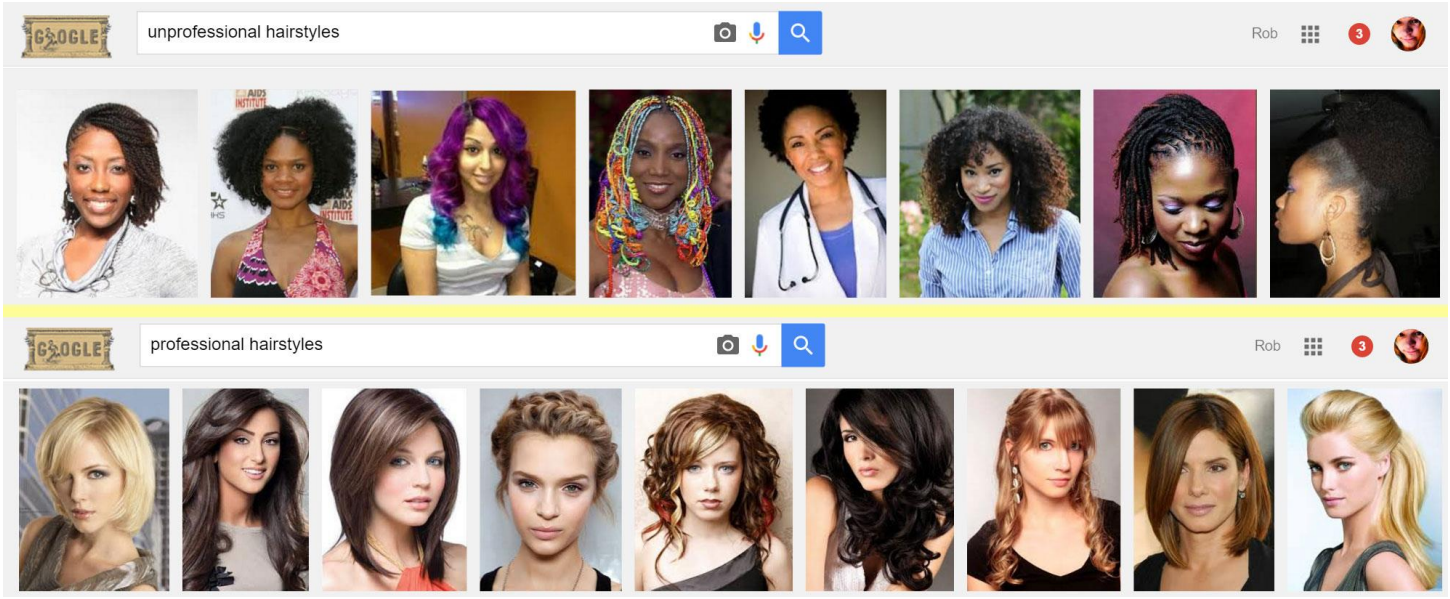
Beispiel: Markterfolg vs. regulatorische Hürden

Technologie: Recommender Systems



Beispiel: Statistik vs. Bias
Technologie: Machine Learning

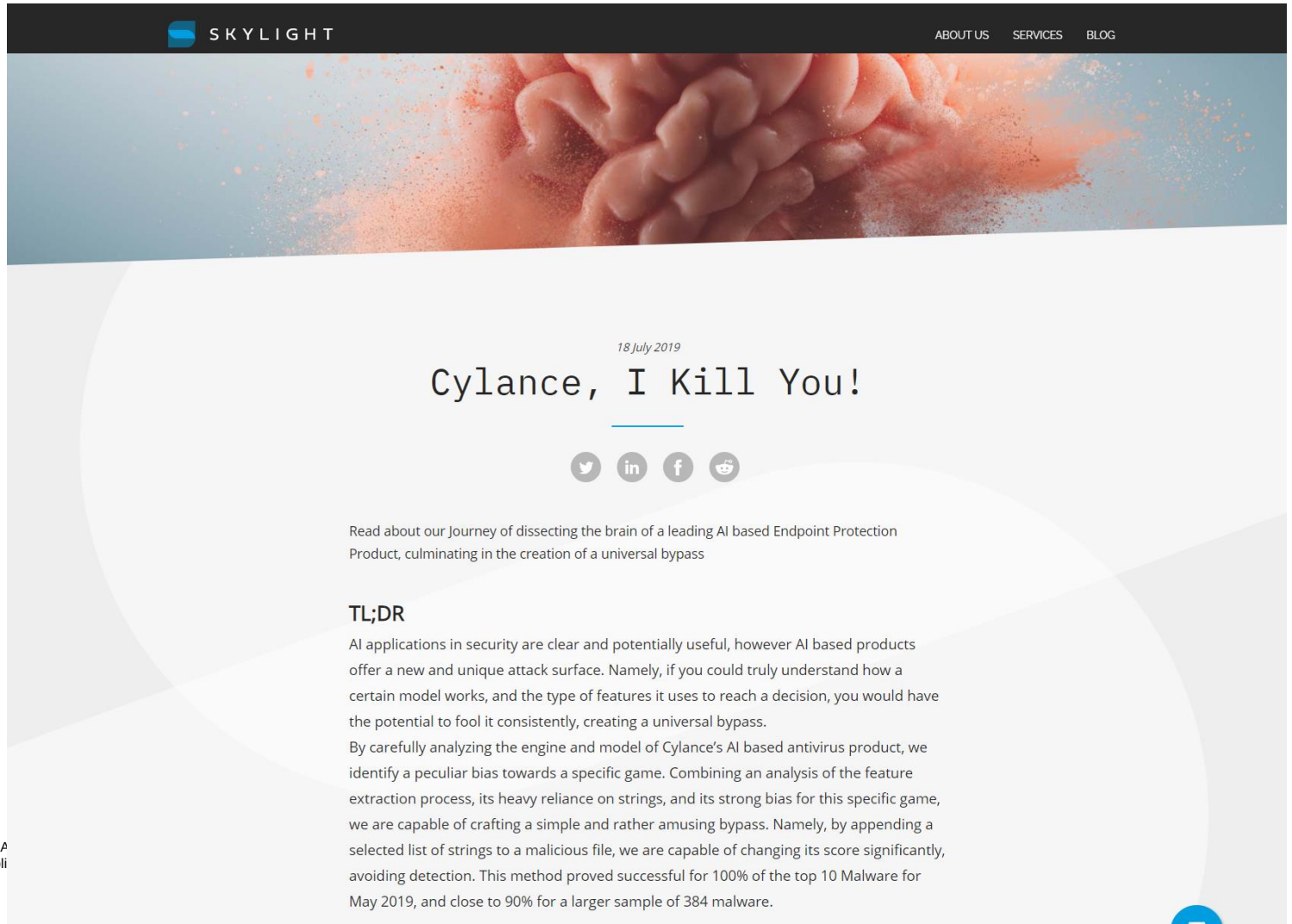




See also: Nassim Nicholas Talib, *«The Black Swan: The Impact of the Highly Improbable»*, 2007

Beispiel: künstl. Intelligenz vs. natürl. Dummheit

Technologie: Machine Learning mit nachgelagerten Regeln



SKYLIGHT

ABOUT US SERVICES BLOG

18 July 2019

Cylance, I Kill You!

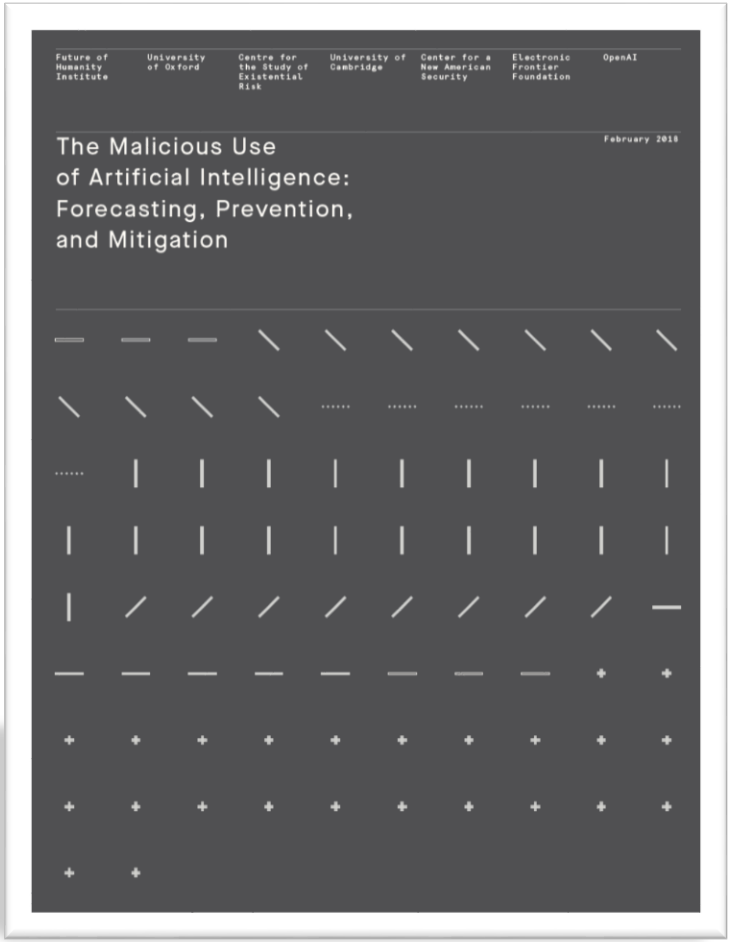
Read about our Journey of dissecting the brain of a leading AI based Endpoint Protection Product, culminating in the creation of a universal bypass

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:



Zürcher Hochschule
für Angewandte Wissenschaften

4

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

Zurich University of Applied Sciences

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**

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

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

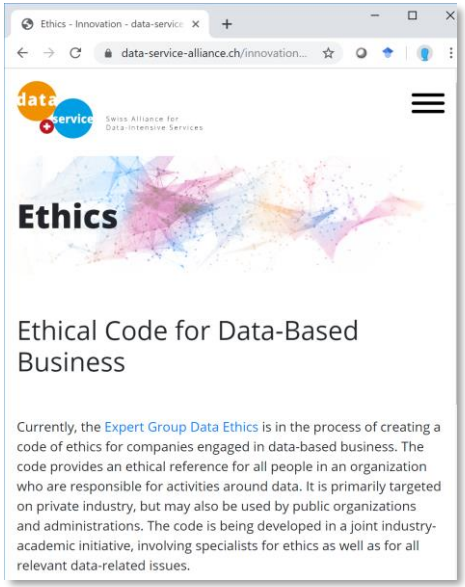
Explain any **algorithmic decision** in non-technical terms to end users

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

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

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



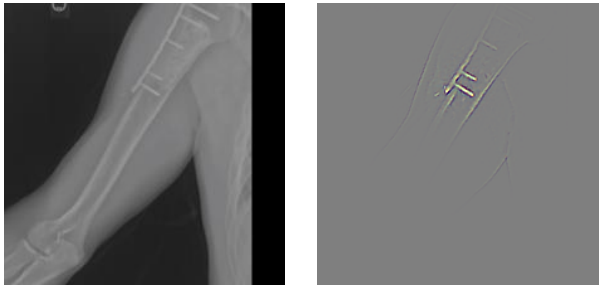
Trend: Entwickeln für Interpretierbarkeit

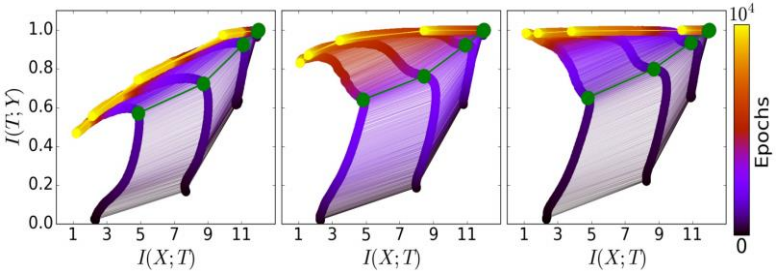
Interpretability is required.

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

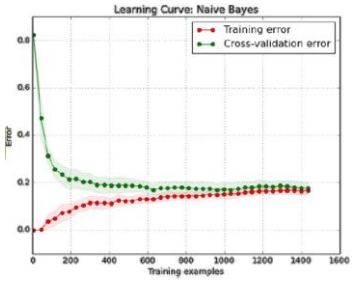
negative X-ray

positive X-ray

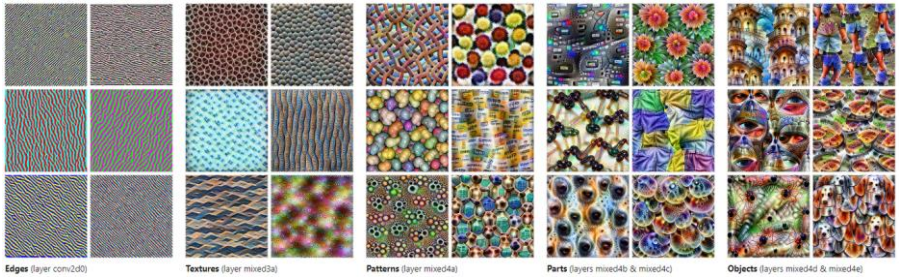




DNN training on the Information Plane



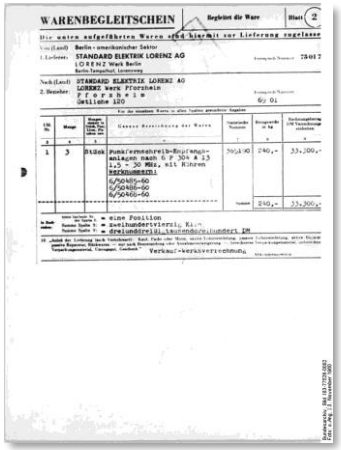
a learning curve

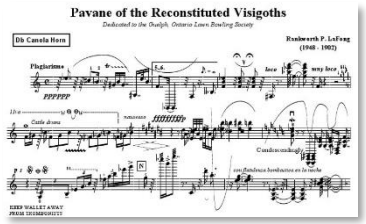


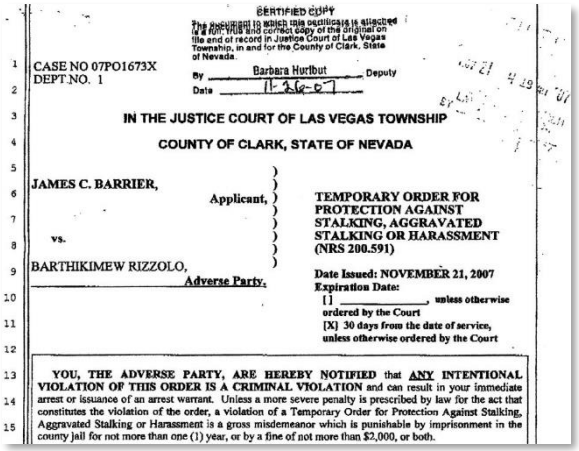
feature visualization

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







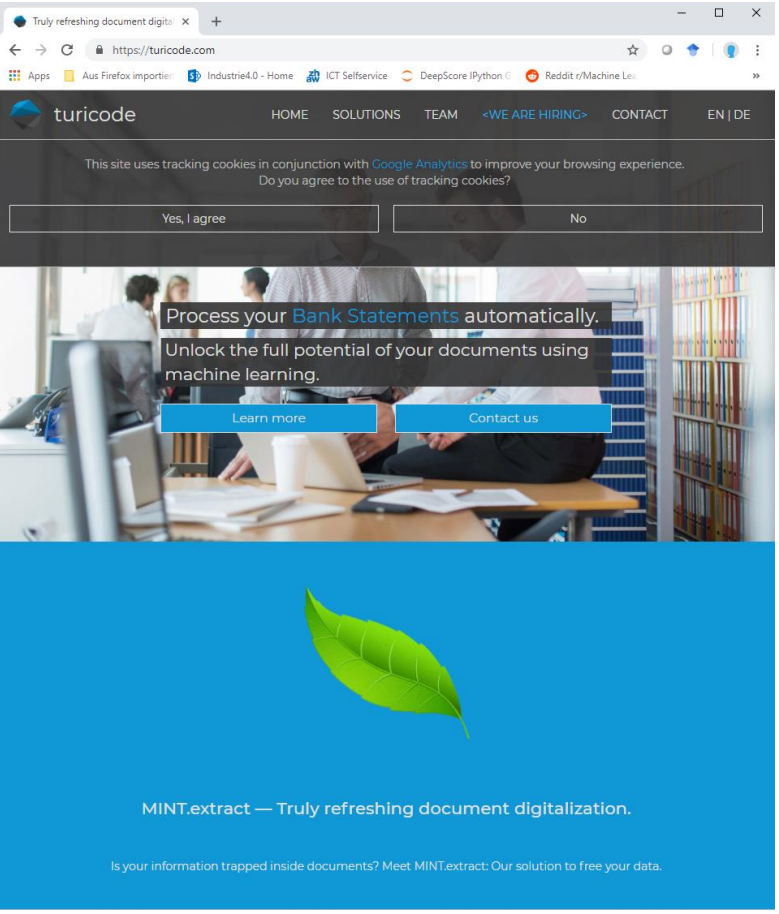
Documents

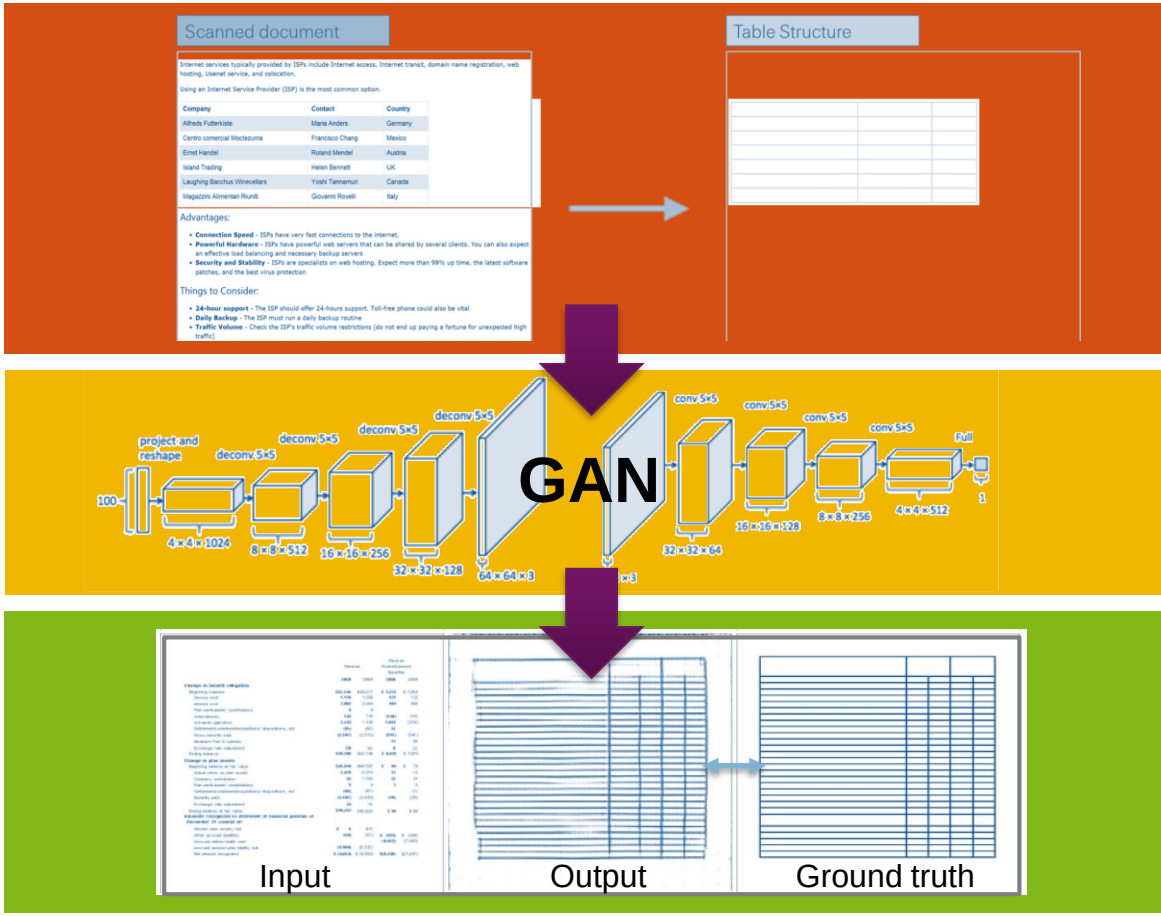
- **Ubiquitous** in human communication and every scenario involving an office
- Somewhat structured for human expert; **unstructured** w.r.t machines
- **Great use case for** various **AI** techniques, including computer vision techniques

Own scientific community

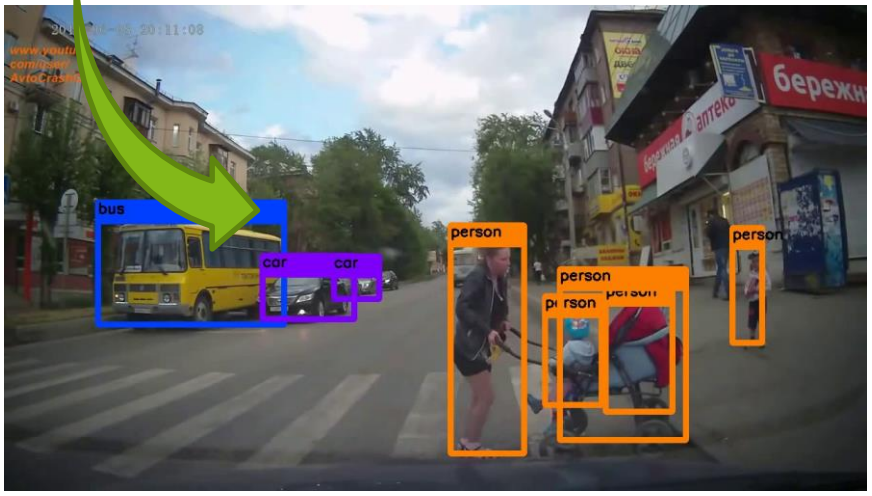
- IAPR’s biannual Intl. Conference on Document Analysis & Recognition (ICDAR):
  character & symbol recognition, printed/handwritten text recognition, graphics analysis & recognition, document analysis & understanding, historical documents & digital libraries, document based forensics, camera & video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

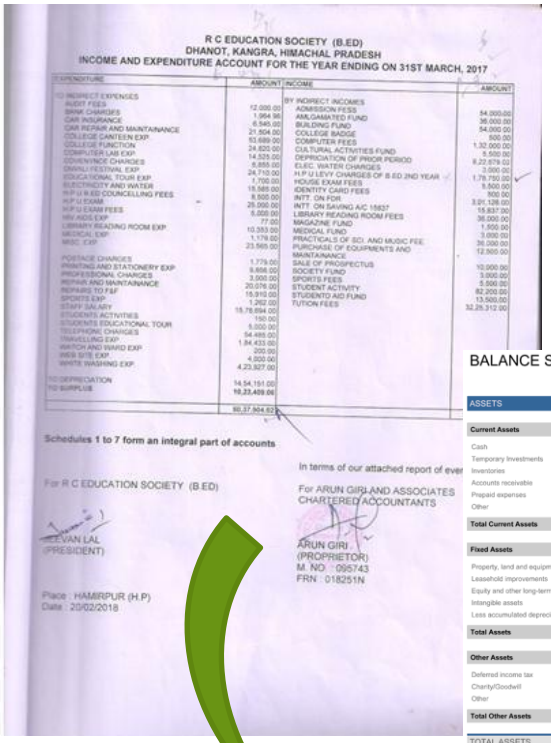
Beispiele aus der (lokalen) Wirtschaft





Beispiele aus der (lokalen) Wirtschaft (contd.)





R C EDUCATION SOCIETY (B.ED)
DHANOT, KANGRA, HIMACHAL PRADESH
INCOME AND EXPENDITURE ACCOUNT FOR THE YEAR ENDING ON 31ST MARCH, 2017

Schedules 1 to 7 form an integral part of accounts

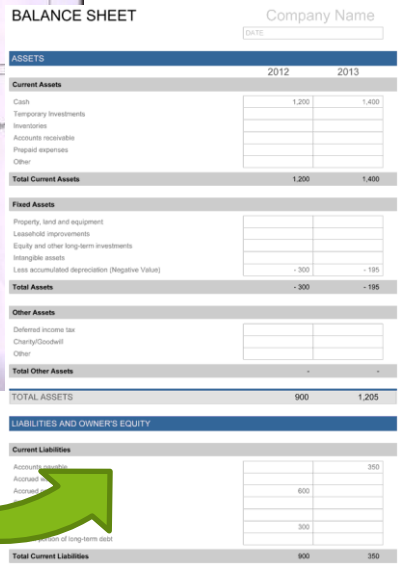
For R C EDUCATION SOCIETY (B.ED)

In terms of our attached report of even

For ARUN GIRI AND ASSOCIATES
CHARTERED ACCOUNTANTS

ARUN GIRI
(PROPRIETOR)
M. NO. : 096743
FRN : 018251N

Place : HAMIRPUR (H.P)
Date : 20/02/2018



BALANCE SHEET

Company Name

DATE

ASSETS	2012	2013
Current Assets		
Cash	1,200	1,400
Temporary Investments		
Inventories		
Accounts receivable		
Prepaid expenses		
Other		
Total Current Assets	1,200	1,400
Fixed Assets		
Property, land and equipment		
Leasehold improvements		
Equity and other long-term investments		
Intangible assets		
Less accumulated depreciation (Negative Value)	- 300	- 195
Total Assets	- 300	- 195
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 [illegible]		
Accrued [illegible]	600	
	300	
[illegible] of long-term debt		
Total Current Liabilities	900	350

Zwischenfazit: Entkopplung

Grösse der Idee ≠ Grösse des Unternehmens

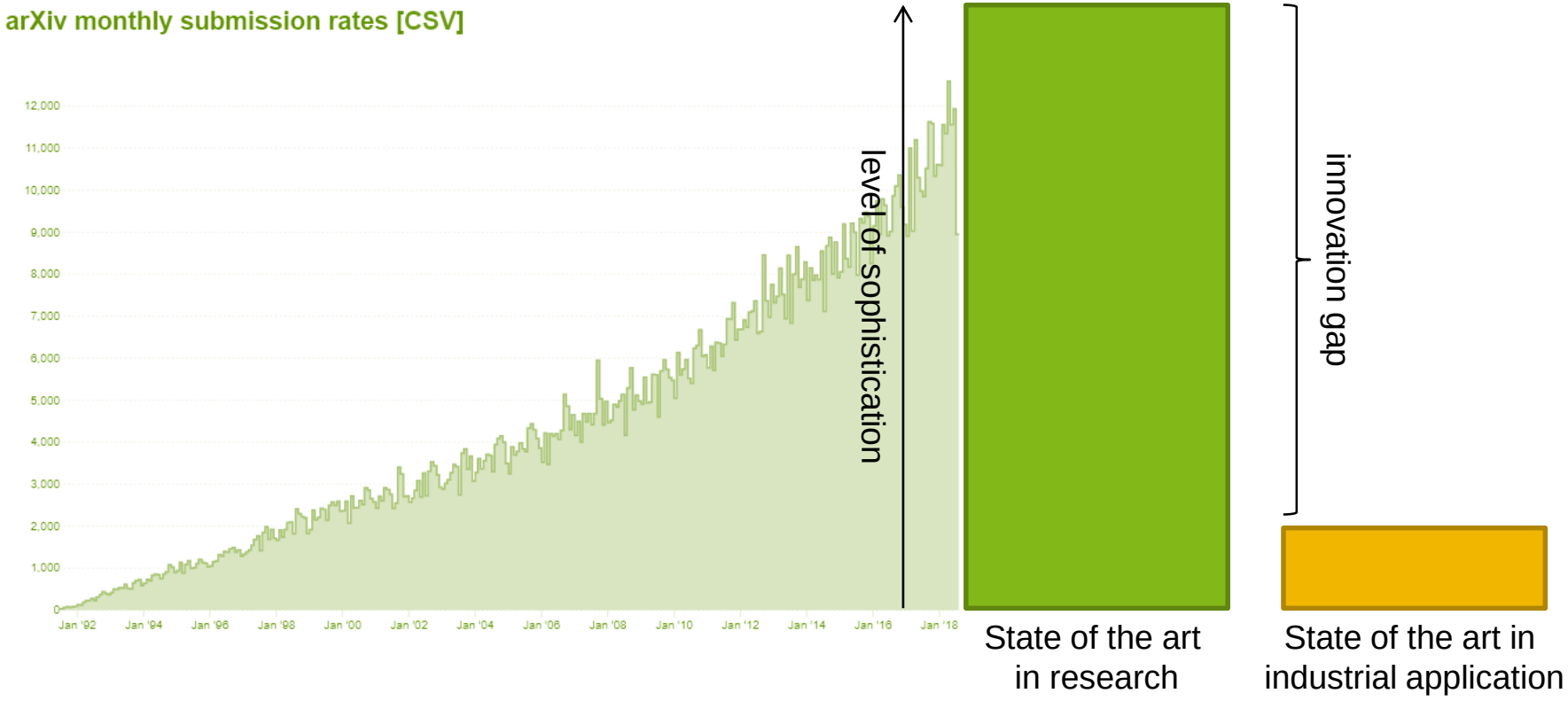
…KMUs können **bauen was auch immer sie mögen**
(gegeben Know-how, Daten und einen interessanten Business Case)

Technologie ist branchenunabhängig

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

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

datalab
www.zhaw.ch/datalab

data service
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