Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Neues Angriffsziel Maschinenintelligenz – Kann ich schützen was ich nicht verstehe?

Prof. Dr. Reinhard Riedl
Head of BFH-Center for Digital Society

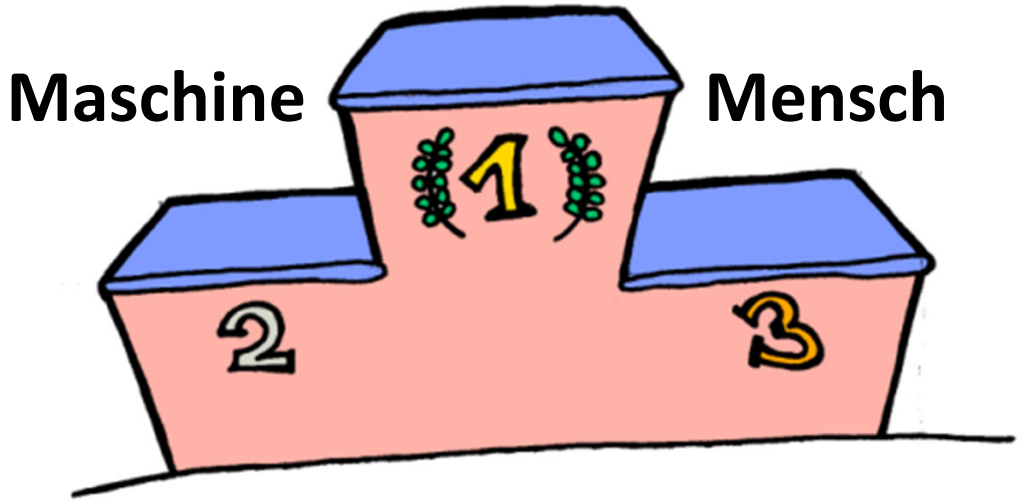
BFH-Zentrum Digital Society

NEIN!

aber

Wettkampf zwischen Mensch und Maschine

Mensch & Maschine

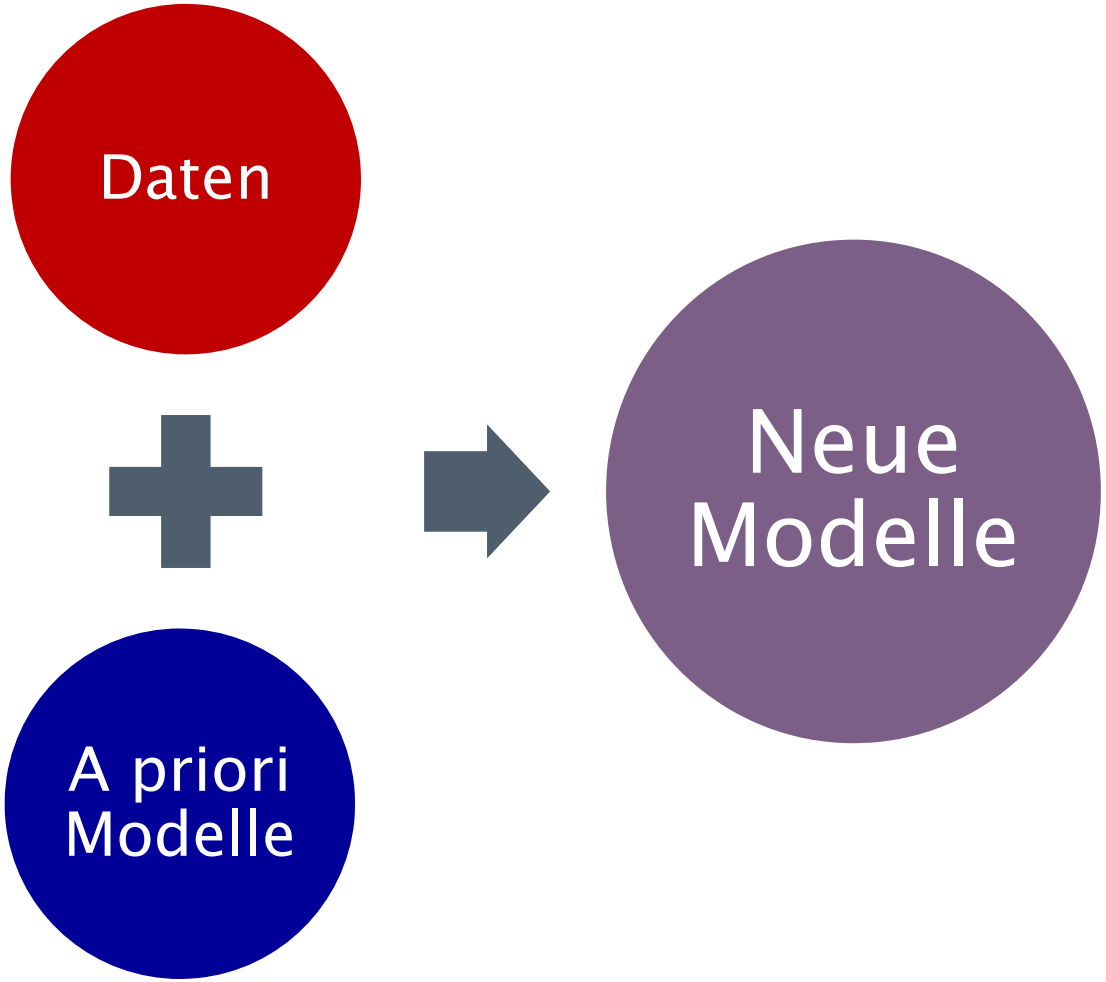


Entscheide, Design und Implementierungen **mit definiertem Kontext**:
die Maschinen schlagen Spezialisten / Führungskräfte zu 80 bis 100 Prozent!

Big Data & Machine Learning

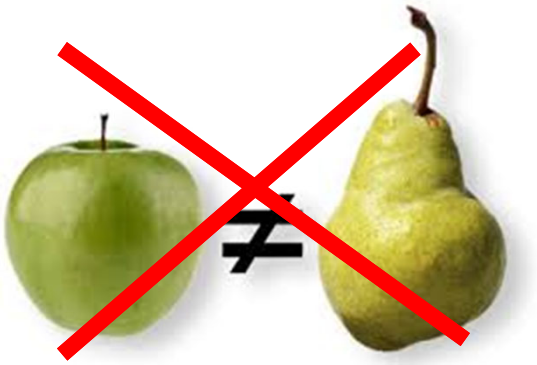
- Offensichtliche Nutzung
  - Entscheidungen treffen oder vorbereiten
- Nicht ganz so offensichtlich
  - Daten-Bias eliminieren
  - Variablen reduzieren
  - Kausale Zusammenhänge identifizieren
  - … aus vielen schlechten genügend viele gute Daten machen
- Der Clou
  - Zusammenhänge ohne menschliche Ratio erkennen

Das Versprechen



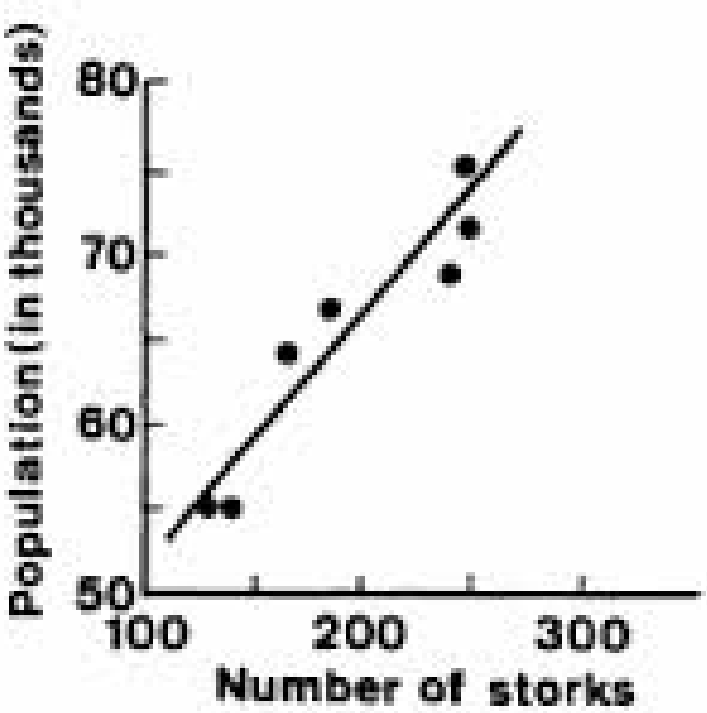
Der Perspektivenwechsel

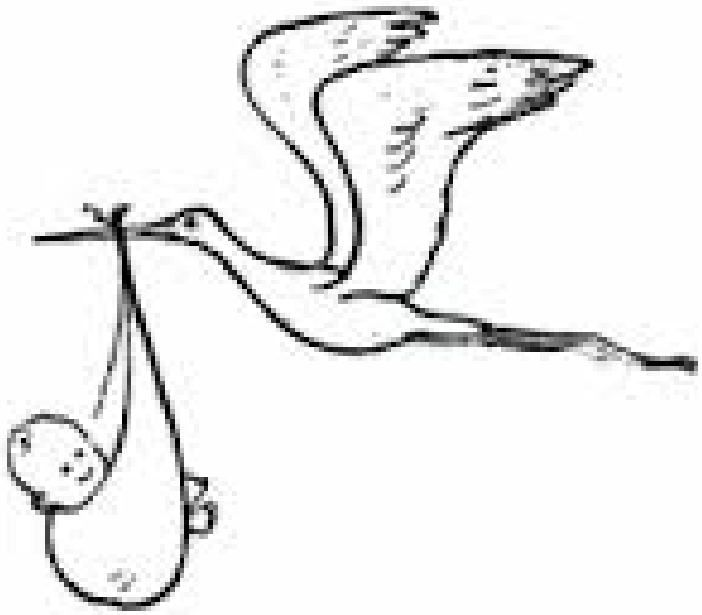
Die vier Paradigmenwechsel



Big Data ist wirklich anders …
… in Bezug auf die Anwendung

Die Mischung ist die neue Datenqualität!

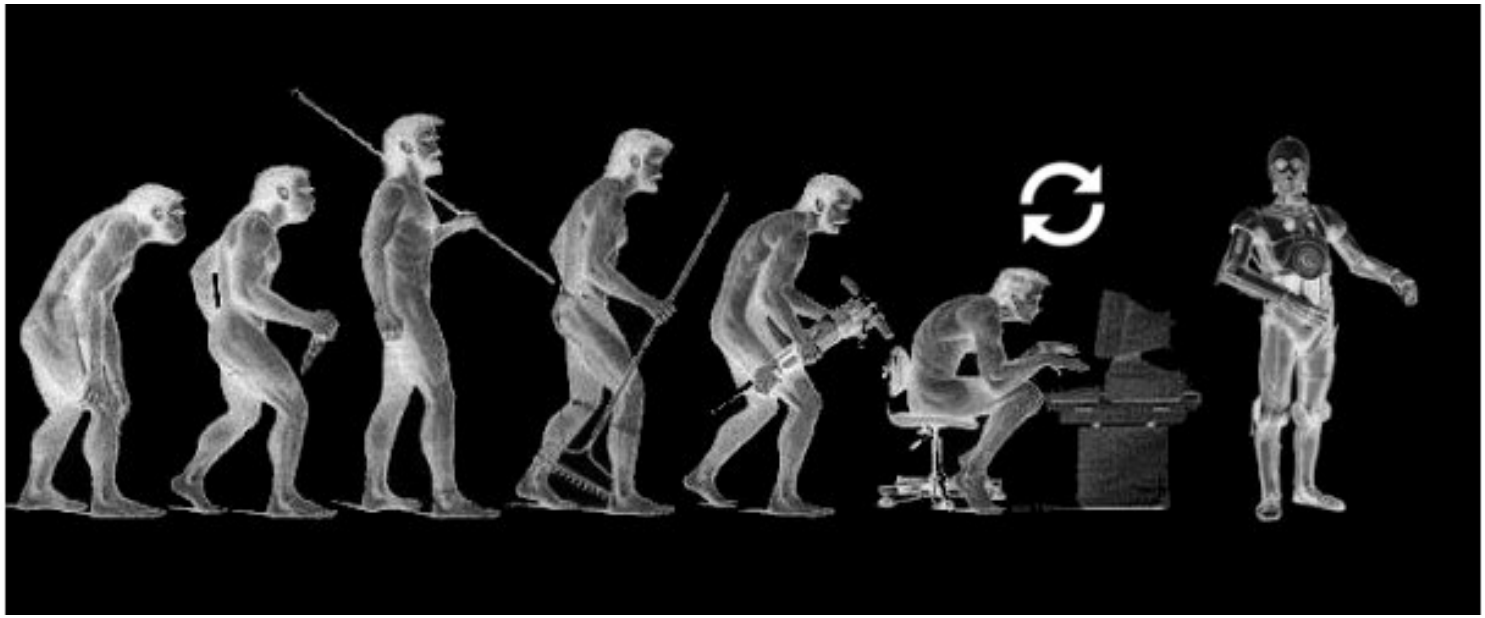




Korrelation ist die neue Kausalität! (oder nicht)

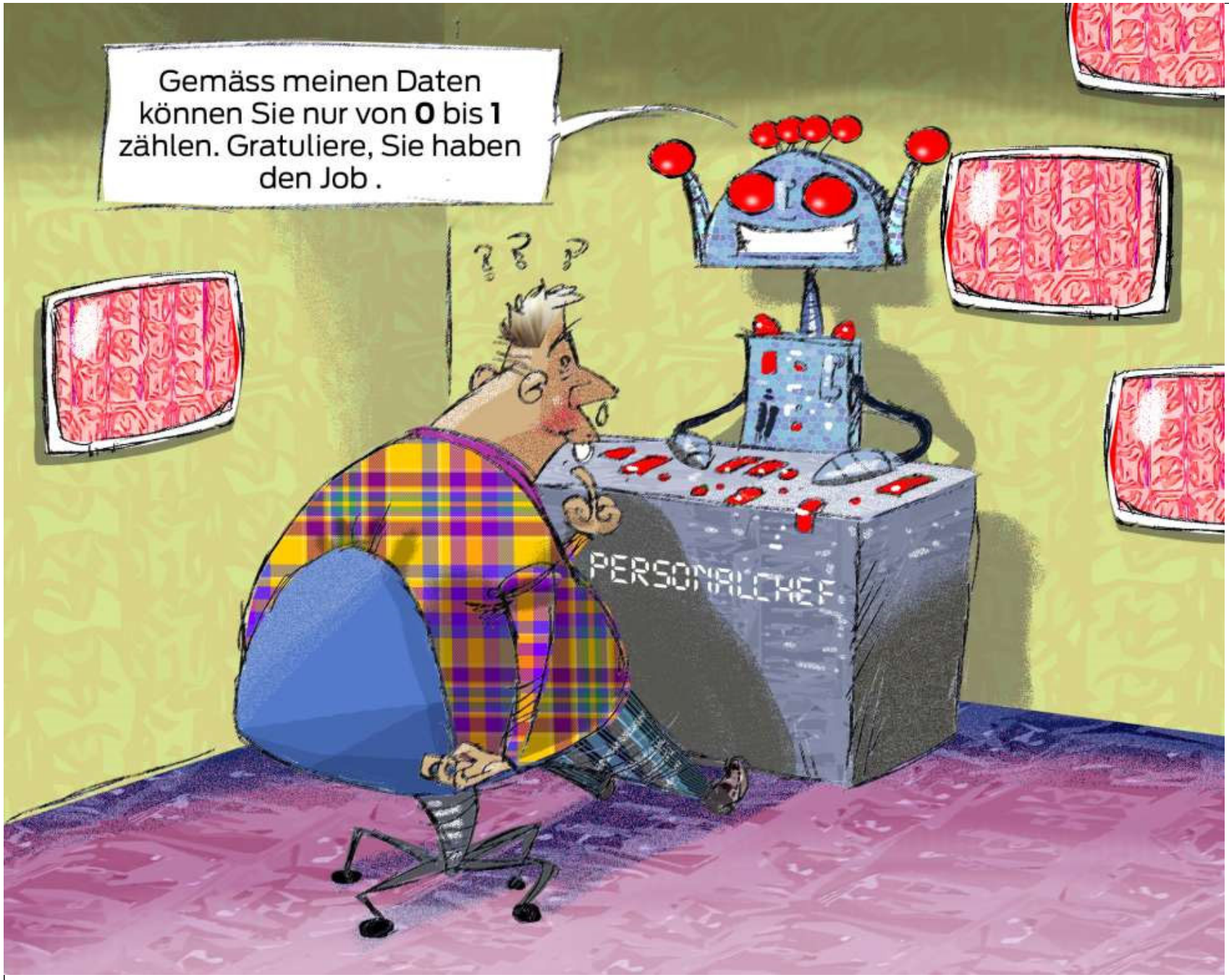
CDC Blogs

Diskriminierung ist die neue Fairness!



Nagarjuna Kadampa Meditation Centre

Maschinen sind die neuen Lernwesen!



https://www.blick.ch/news/wirtschaft/schoene-neue-welt-der-jobsuche-was-chefs-in-zukunft-ueber-uns-wissen-id7251498.html

0
1
1
1
0
1
0
1
0
0
0
1
0
1
1
1
0
0

Konkrete Anwendung im Predictive Policing

Predictive Policing «datafiziert» die Polizeiarbeit

- Chancen: u.a. die Vorhersage von
  - Verbrechen
  - Tätern
  - Täterprofilen
  - Opfern
- Offline und online …

- Typisches Problem:
  - Bias bei der Variablenselektion
  - Bias bei der Datengenerierung

- Unklar:
  - ***Einfügen von Shortcuts bei Datenmangel …***
  - ***Was ist und bedeutet das neue Modell?***

Schlüsselidee 1

Daten **von vielen** darüber, wie sie sich in den Kontexten A und X verhalten

?

Schätzung

bzw. Prognose

Das Verhalten **einer Person** im Kontext A ist bekannt, wir interessieren uns für ihr Verhalten im **Kontext X**

Big Data hilft, unbekannte Eigenschaften zu schätzen, bzw. Prognosen für die Zukunft zu machen

Schlüsselidee 2

Daten **von vielen** über unterschiedliche Aspekte

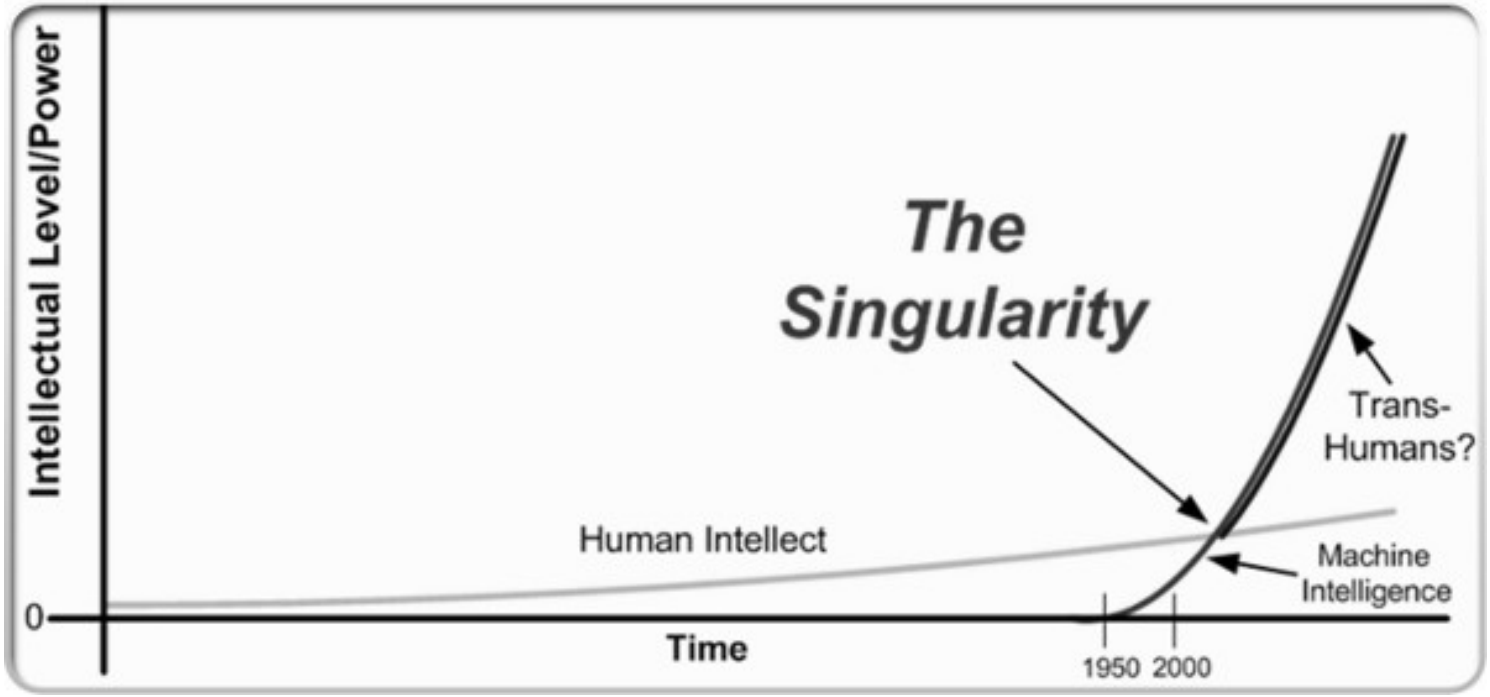
Identifikation eines häufiges Phänomens

Big Data hilft, unbekannte Zusammenhänge
Zu entdecken → neue (Forschungs-)ideen

Kritische Fragen:

Was sind die Kontexte wirklich?

Was beobachten wir bei welchen vielen wirklich?



Saravanan.org

Wozu noch andere Wissenschaften ???!

(2008 – 2015)

Mächtiges & gefährliches und gefährdetes Werkzeug

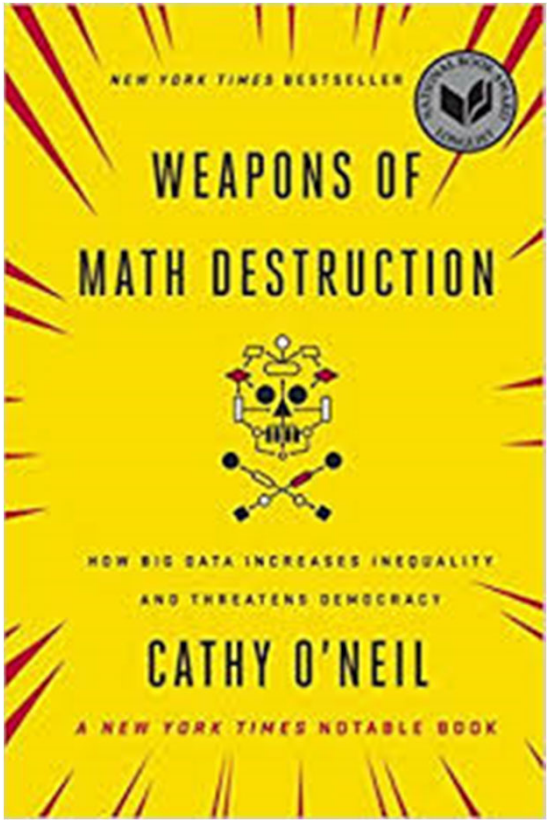
- Grosses Potential
  - Verbesserung der Datenlage
  - Hinterfragen existierender Modell
  - Generierung von Hypothesen
- Paradigmenwechsel
  - Verlangt anderen Umgang mit Ergebnissen
- Probleme
  - In der Praxis fehlende Transparenz
  - Grundsätzliches Problem der transparenten Darstellung
  - Fehlende menschliche Rationalisierung
  - Gutes Angriffsziel

Warum Probleme??

- Praxis: Ich weiss nicht …
  - Woraus ich etwas folgere
  - Womit ich etwas folgere
  - Wie vertrauenswürdig die Folgerung ist
  - Warum ich etwas folgere

- Theorie:
  - Kein klares Klassifikationssystem für Datenqualität
  - Schwierige Beschreibung der Algorithmen
    (das mathematische Begriffssystem passt nicht!!!)
  - Klassisches Problem der «***Inference Uncertainty***»
    - Im Fall von seltenem Feedback
  - Schwierige Ableitung «***menschlicher Rationalisierungen***»
    - Normal, wo «Information Pickup» stattfindet
    - Ein Problem, wo Entscheide gerechtfertigt werden müssen

Ethische Fragezeichen



- Kritisch sind Modelle mit 4 Eigenschaften
  - Fehlende Transparenz
  - ***Ungenügender Feedback-Mechanismus***
  - Hohe Skalierbarkeit
  - Grosse negative Auswirkungen auf Betroffene
- Der Schutz der Privatsphäre ist eine Illusion
  - Pragmatischer ist die Verhinderung der Datennutzung gegen die Betroffenen
- Grundsätzlich zu hinterfragen
  - Gesellschaftlicher Nutzen
  - Gesellschaftliche Auswirkungen
    - ***Jede Messung beeinflusst das Gemessene***

BIG DATA & DEEP LEARNING
bringen viele neue Möglichkeiten,
sind aber NICHT ALLES!

Es braucht auch Small Data
und Rare Data …

… und ein situatives Design + Digital Skills

Big, Small und Rare im Vergleich

BIG DATA

- Data Analytics
- Mathematik in Zentrum
- Entscheidend: viele Daten, passende Algorithmen, Verstehen von Methode & Ergebnis

SMALL DATA

- Neugier, Expertise, Beobachtungsgabe
- Beobachtungen im Zentrum
- Entscheidend: richtig selektieren und interpretieren

RARE DATA

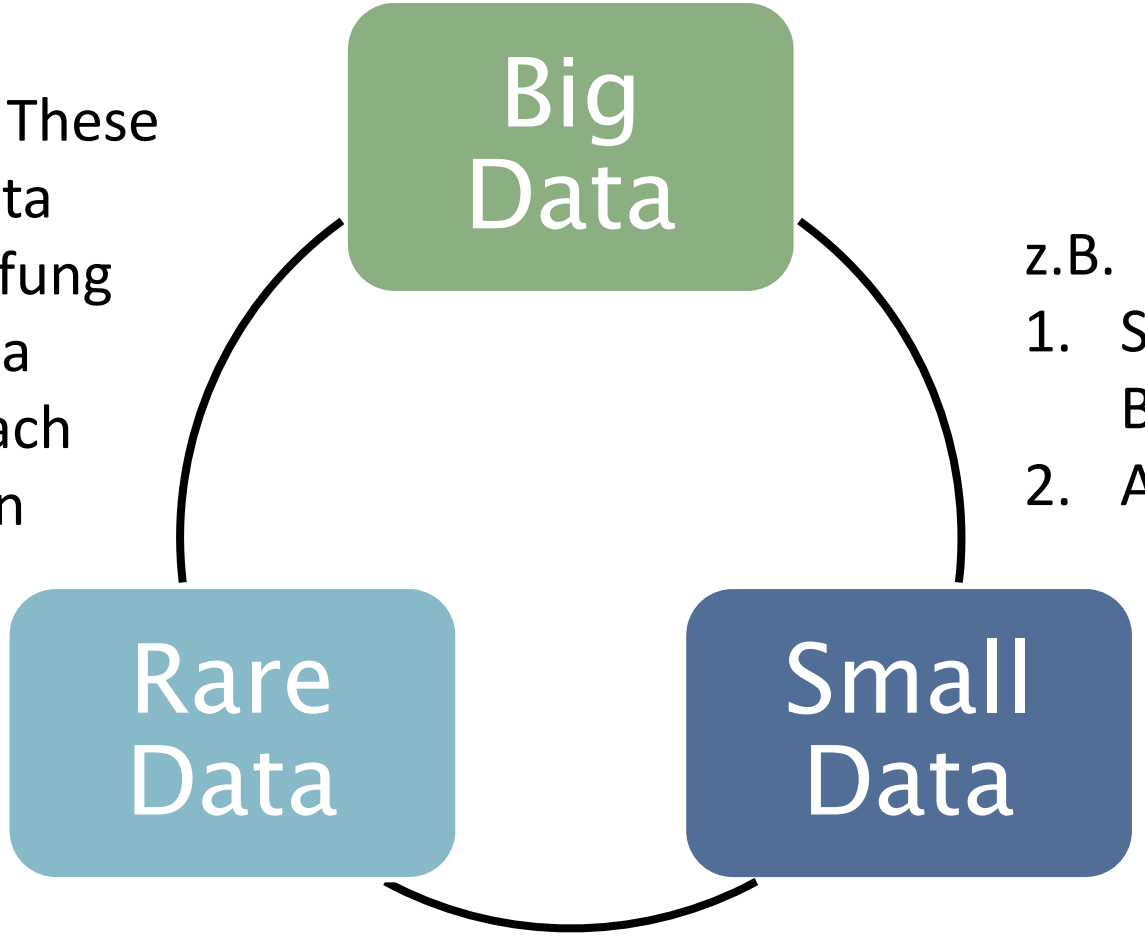
- Analogieschlüsse und Abstraktion
- Modelle und Muster im Zentrum
- Entscheidend: Ideen-Transfer zwischen unterschiedlichen Bereichen

Traditionelles Herangehen

Normalfall: Alle drei sind involviert, irgendeine liefert die These, die anderen machen sie konkreter

z.B.

1. Big Data These
2. Small Data Überprüfung
3. Rare Data Suche nach Lösungen

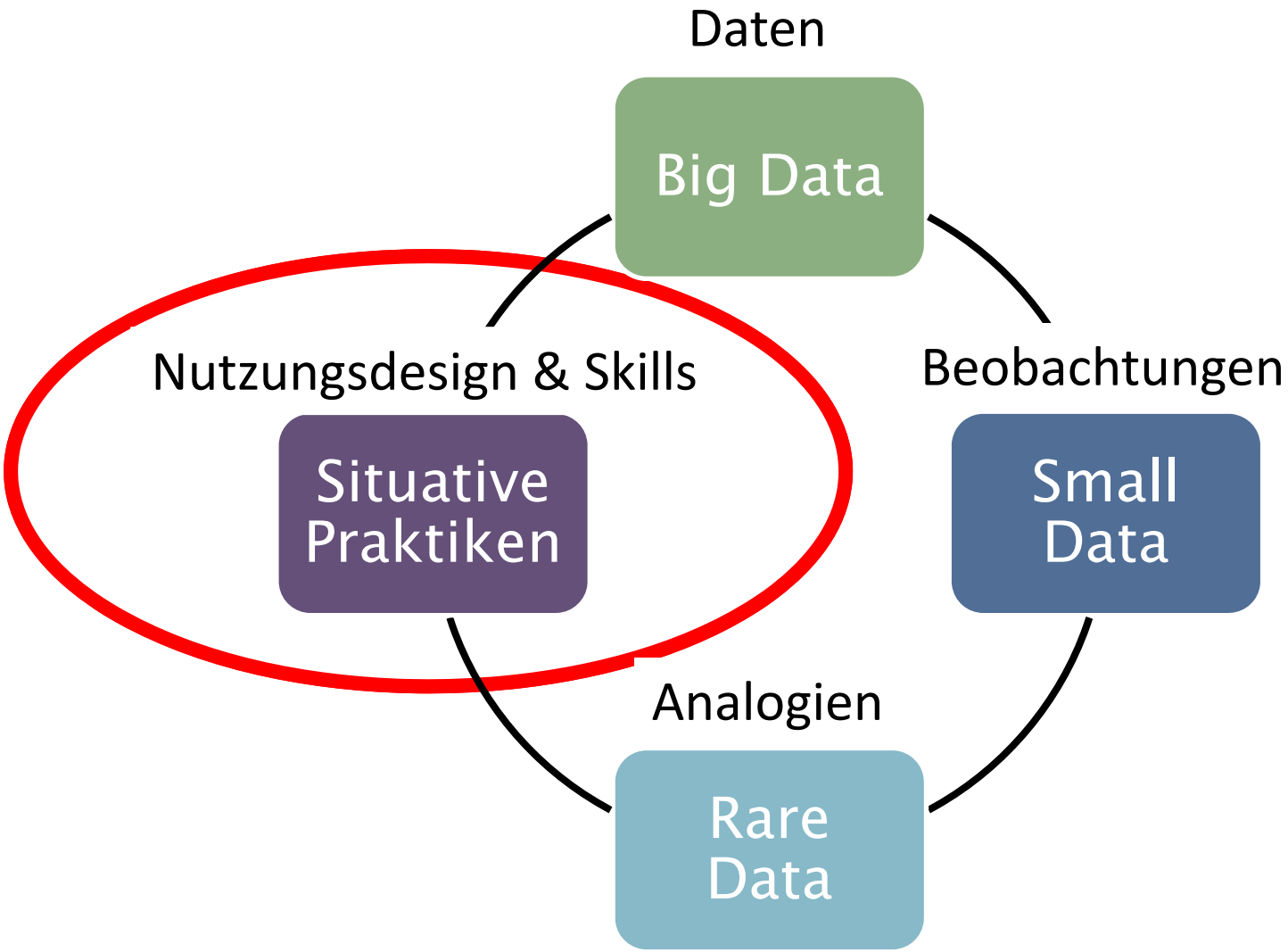


z.B.

1. Small Data Beobachtung
2. A/B Tests

z.B. 1. Rare Data Idee für neue Dienste
2. Big Data Detailanalysen

3 Erkenntnismethoden + Nutzungsprinzipien



Natürliches Angriffsziel

- Deep Learning
  - Konfiguration der neuronalen Netze
  - Einspeisen zusätzlicher Lerndaten
  - Blockieren von relevantem Feedback

- Nutzung des Deep Learnings
  - Lernbasierte Datenlagen und Entscheide
  - D.h. ***das Handeln ohne menschliche Rationalisierung und natürliche menschliche Validierung***

- Auch ohne Angriffe
  - Nachvollziehen von Deep Learning & Evidenz (→ theoret. Probl.)
  - Situatives Design des Einsatzes
  - Instandhaltung von Deep Learning

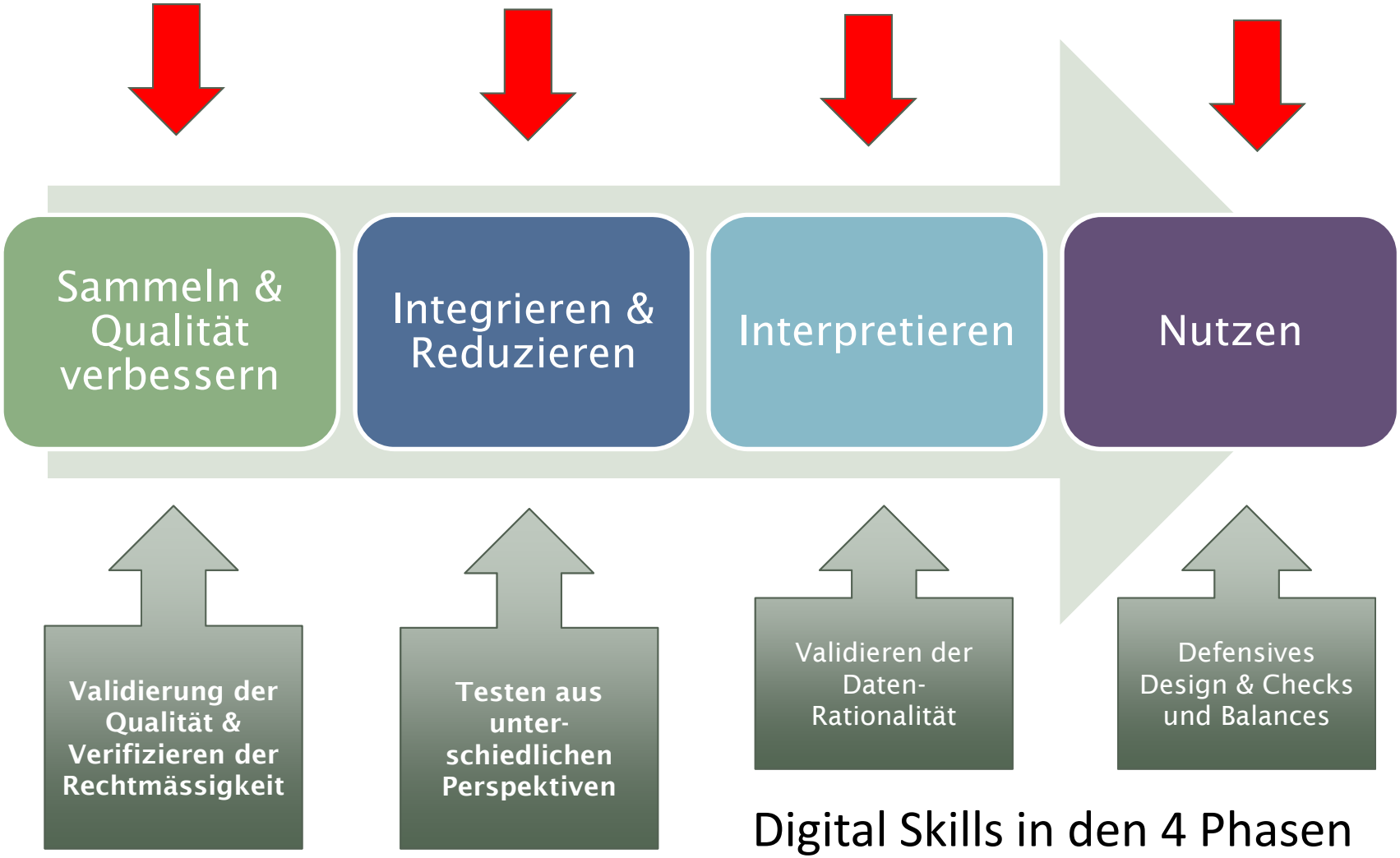
Problem: Wir kennen das schwache Glied in der künstlich intelligenten Argumentation nicht …

Schutz gegen Angriffe

- Deep Learning
  - Bester Schutz
    - Vollständige («kausale») Aufzeichnung
    - Billigere Lösungen sind problematisch
  - Ethisch wichtig
    - Einbau von häufigem Feedback unter menschlicher Aufsicht
- Nutzung von Big Data und Maschinenlernen
  - Unabhängiges System für Qualitätskontrolle
  - Defensive Nutzungspraktiken
  - Plus Forschung zu «Rationalisierungsalgorithmen» und zu Inference Uncertainty
- Systemische Herangehensweise reduziert den Angriffsraum!


SKILLS
education
business
efficiency
experience
capabilities
knowledge
teamwork
career
analyzing
training
talent
awareness
clever
skills

Angriffe & byzantine Fehler …



Digital Skills in den 4 Phasen

Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Zusammenfassung

Chancen und Risiken
Die drei Grand Challenges

BFH-Zentrum Digital Society

Chancen

PAL Verlag

- Instrument zur Forensik
  - IT-Forensik
  - Predictive Policing
- Umfassendere Perspektive
- Weniger menschlicher Bias
- Schnellere Erkenntnisse

Risiken

- Verletzung der Privatsphäre
- Fehlendes Verständnis
- Inkompetente Auswertungen & Interpretation
- Inkompetente Nutzung
- Diskriminierung & Förderung der Ungleichheit
- Neues Angriffsziel

Die 3 Grand Challenges

- Verfügbarkeit von Daten
  - Soziales Miteinander???
- Big Data Skills
  - Weiterbildung!!!
- Transdisziplinäre Zusammenarbeit
  - Auch in der Forensik.

Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Danke! Fragen?

reinhard.riedl@bfh.ch

Forschung am BFH-Zentrum Digital Society

- 6 Fakultäten (Departemente)
  - Wirtschaft; Technik & Informatik; Hochschule der Künste Bern; Gesundheit; Soziale Arbeit; Architektur, Holz und Bau;
- 6 Themen
  - 2 Methodenthemen; Design 4 Future Fitness; Big & Open Data;
  - 2 Fachthemen: Identität & Privatsphäre; Cybersecurity & IT-Forensik;
  - 2 Anwendungsthemen: Gebäude & Städte; Gesundheitsversorgung & E-Health;
- Mehr als 12 Disziplinen
- Ein Thema
  - Nutzung der Digitalisierung / digitalen Transformation als Enabler (Menschen, Organisationen, Staat, Fachdisziplinen)
- Eine Online-Zeitschrift
  - www.societybyte.swiss