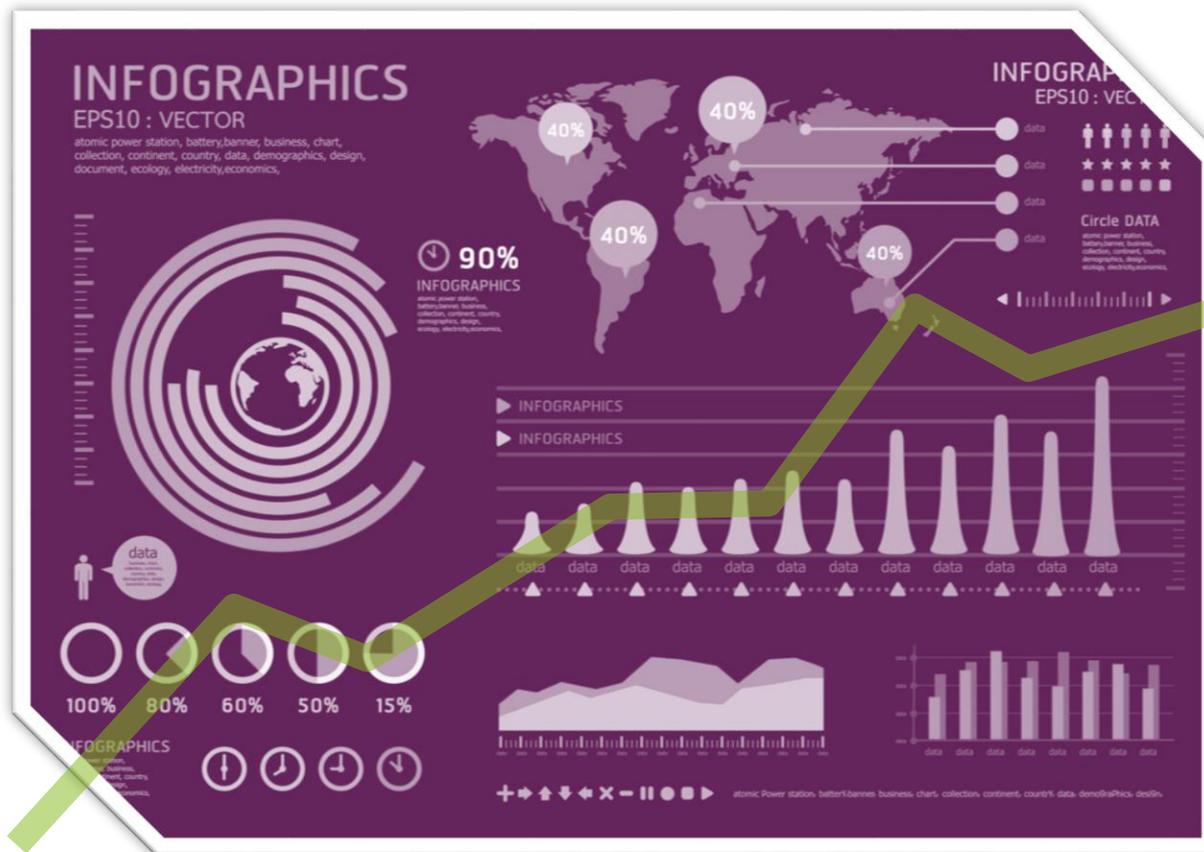


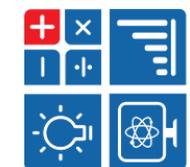
# Der Wert von *Daten* in Zeiten von *Big Data*

Microsoft Vision Days, 21.01.2016

Thilo Stadelmann



data1ab  
[www.zhaw.ch/data1ab](http://www.zhaw.ch/data1ab)



Wert → Anforderungen → Trends

1

Der Wert von Daten

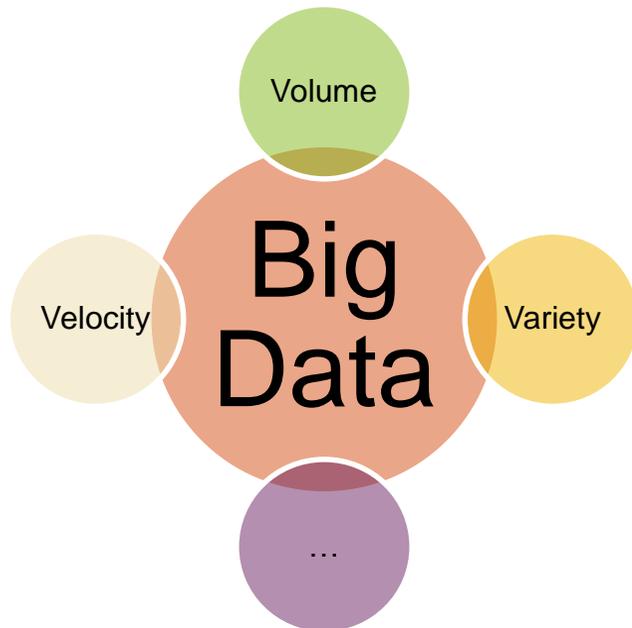
**«Data is the new oil»**

**(Clive Humby, ANA Senior marketer's summit, 2006)**

**Wenn Daten der neue Rohstoff sind...**

- Für wen haben sie dann Wert?
- Worin besteht dieser Wert?

# Der quantifizierbare Wert von Big Data



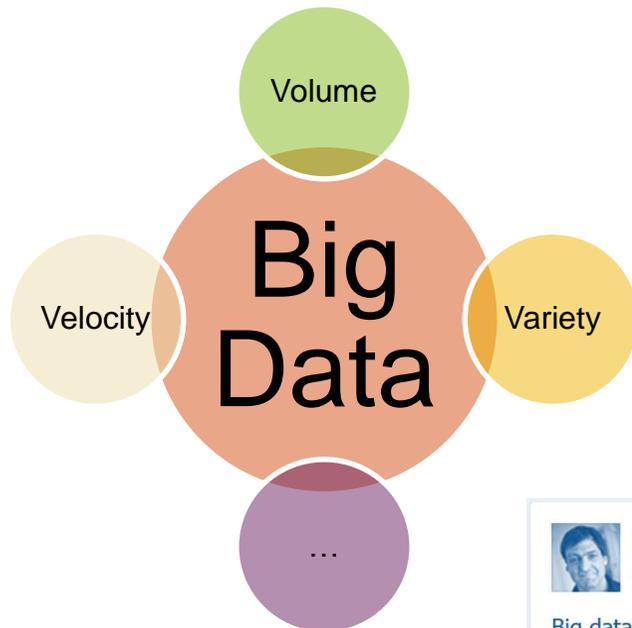
(D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner, 2001.)

# Der quantifizierbare Wert von Big Data



(D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner, 2001.)

# Der quantifizierbare Wert von Big Data



- Big Oil
- Big Tobacco
- Big Pharma
- ...
- **Big Data?**



(D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner, 2001.)

# Big Data – eine grosse *Veränderung*

Es geht nicht um 3/4/... V's.



# Big Data – eine grosse *Veränderung*

Es geht nicht um 3/4/... V's.

Es geht um eine generelle Änderung der Denkweise

- Daten sind *verfügbar* (kostengünstig, für jedes Faktum des Lebens)
- Wir können daraus Lernen (automatisiert, prozessintegriert)



# Big Data – eine grosse *Veränderung*

Es geht nicht um 3/4/... V's.

Es geht um eine generelle Änderung der Denkweise

- Daten sind *verfügbar* (kostengünstig, für jedes Faktum des Lebens)
- Wir können daraus Lernen (automatisiert, prozessintegriert)



(Michael Natusch @ SDS|2014)

➔ Bessere Entscheidungen haben einen messbaren Wert



# Der Wert von Daten ...am Beispiel Twitter Inc.

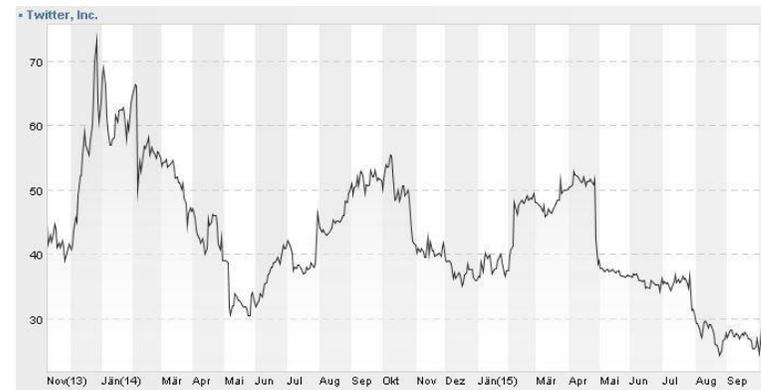


# Der Wert von Daten ...am Beispiel Twitter Inc.



Börsennotiert

**Gesamtbewertung von knapp 21.3 Mrd. \$**



# Der Wert von Daten ...am Beispiel Twitter Inc.

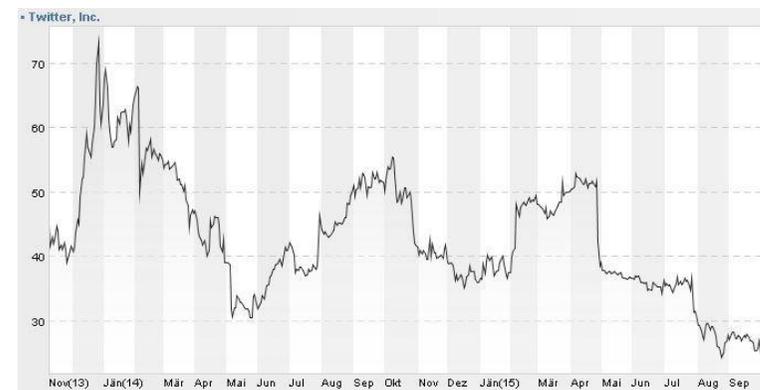
Gesellschaftlich



Börsennotiert

*«Results shown here are impressive, showing that **Twitter data can be used to detect and possibly track Influenza like epidemics in real time.**»*

**Gesamtbewertung von knapp 21.3 Mrd. \$**



(Herland et al., "A review of data mining using big data in health informatics", Journal of Big Data, 2014, 1:2)

Wert → Anforderungen → Trends

2

Veränderte Anforderungen

# Big Data = grosse Veränderung

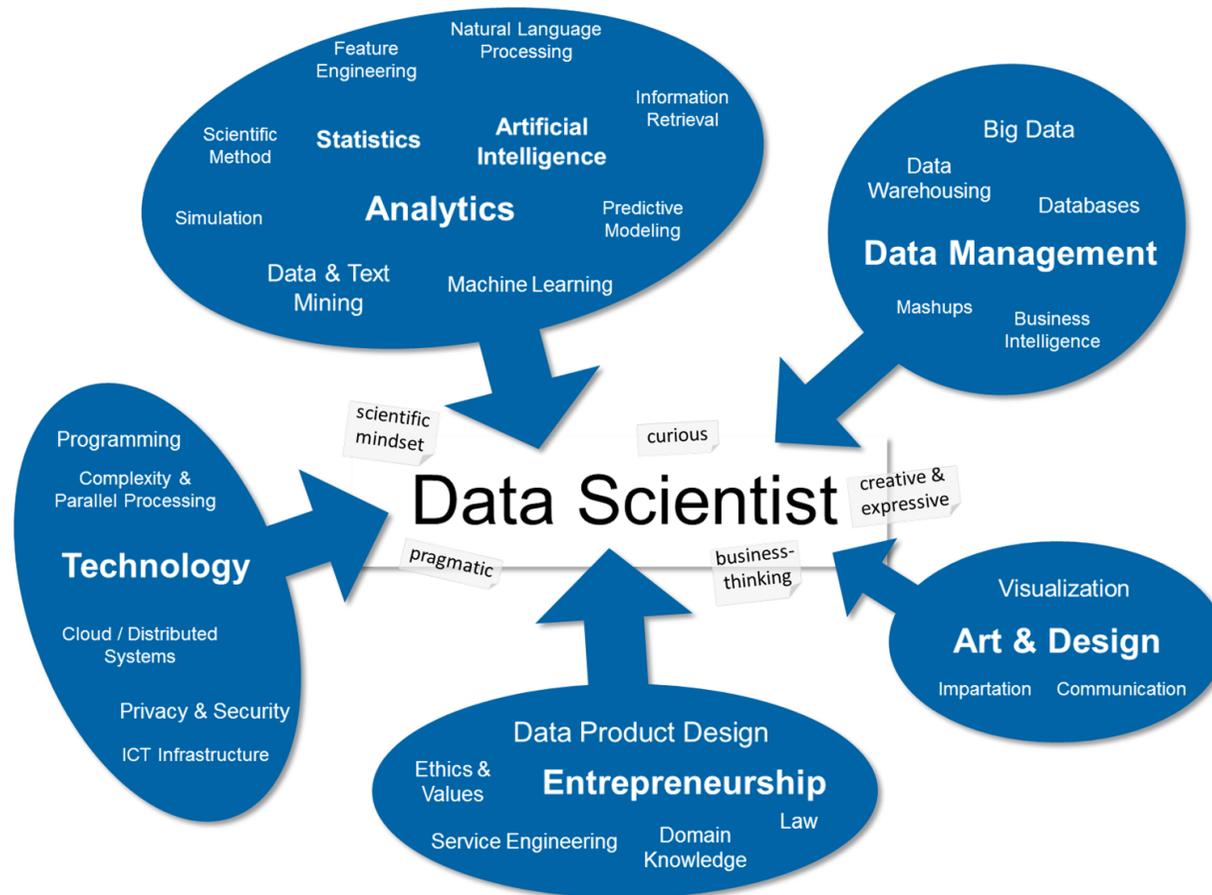
Wenn sich der Denkansatz verändert hat...

- Wer setzt dieses Denken um?
- Wie zieht man geschäftlichen Mehrwert daraus?

**BIG IDEA**

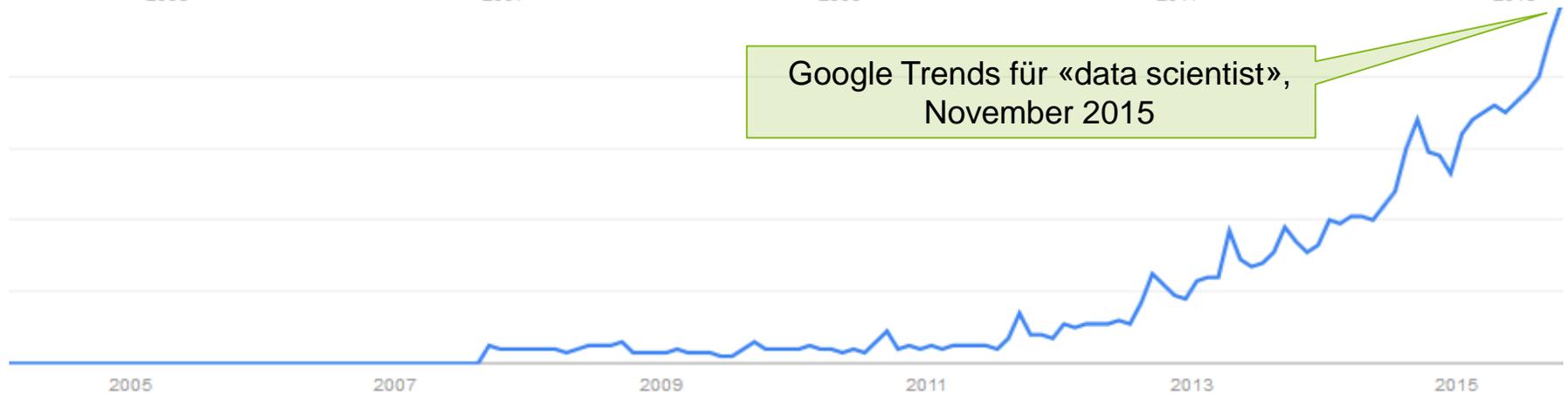
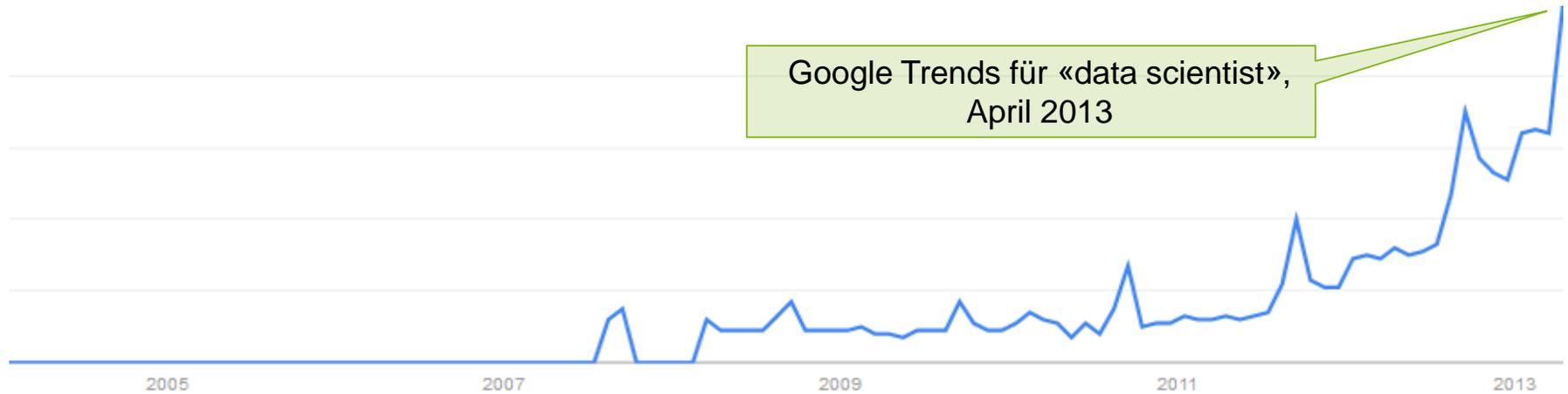
# Ein neuer Typ Mitarbeiter

## Interdisziplinär & praxisbezogen



(Stadelmann et al. (2013). "Applied Data Science in Europe". ECSS 2013, Amsterdam.)

# Nur ein Trend? Kontinuierlich steigend



# Erfahrungen aus der Aus- und Weiterbildung

## ZHAW Diploma of Advanced Studies in Data Science

- Besteht seit Herbst 2014
- Anmeldestand Herbst 2015: Wenige freie Plätze im Herbst 2017

### CAS Data Science Applications

Machine Learning, Big Data Visualization, Design & Development of Data Products, Data Protection & Security

### CAS Information Engineering

Scripting in Python,  
Information Retrieval &  
Text Analytics,  
Databases & SQL,  
Data Warehousing,  
Big Data

### CAS Datenanalyse

Data Description &  
Visualization, Statistical  
Foundations of Analytics,  
Multiple Regression,  
Time Series & Forecasting,  
Clustering & Classification

«But merely using data isn't really what we mean by "data science".  
A data application acquires its value from the data itself...»  
M. Loukides, «What is Data Science?», 2010



# Mehrwert durch «Data Products»

Loukides, «The Evolution of Data Products“, 2011

## Offene Data Products

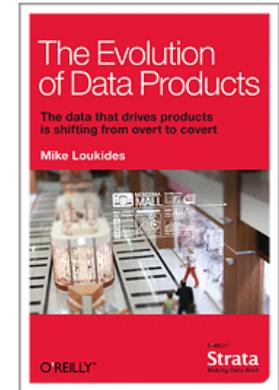
- Basieren auf Daten, liefern Daten
- Beispiel: Produktempfehlung anhand “Likes” → erzeugt Datensätze in einer DB
- ➔ Charakterisiert durch Datentransformation (data in → data out)

## Geschlossene Data Products

- Daten werden nur intern verwendet; **Ausgabe ist Lebensqualität**
- Beispiel: Navigation durch Karten und Verkehrsinfos → entspanntes Reisen (durch Liste von Wegpunkten)
- Erweiterung: Ergänzung um Fahrverhaltensanalyse → optimierter Kraftstoffverbrauch

## Typische Zutaten

- “Datenausstoss“ (Daten, die nebenbei entstehen)
- “Versteckte Daten“ (schwer zu gebrauchende Daten; keiner weiss, was damit anfangen)
- Kombination verschiedener Datenquellen (auch “open data“)



# Best practices für data product design

Bsp. Empfehlungssystem: Etwas beliebtes vs. **etwas neues** finden

- Die "Filterblase" hält uns ansonsten im Loch landläufiger Meinung und gepflegter Vorurteile
- Gute Treffer sind nicht genug – interessante (d.h. nicht offensichtliche) Treffer sind es
- "Likes" sind das falsche Optimierungsziel
- Vorhersagen sind kein enggültiges Ziel; Entscheidungen sind es

## Benutzer-zentriertes Design

- Menschen können schlecht mit Rohdaten umgehen  
→ gib Ihnen **Optionen zum Entscheiden**
  - Menschen interagieren nicht mit Daten  
→ gib Ihnen **Anwendungen** und
  - Menschen kümmern sich nicht um Daten  
→ gib Ihnen **Mehrwert**
- *Data Products sind kein Job für Nerds, (Value Proposition) Design spielt von Anfang an mit*



Wert → Anforderungen → Trends

3

Trends aus der Forschung

# Technologische Trends

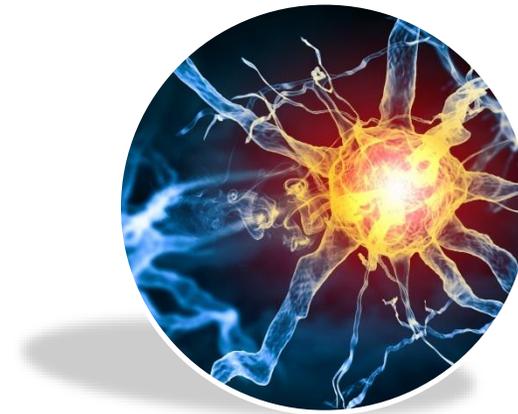
## Datenhaltung

- **Mix** aus klassischen relationalen DBMS...
- ...und Big Data Technologien wie Hadoop Stack



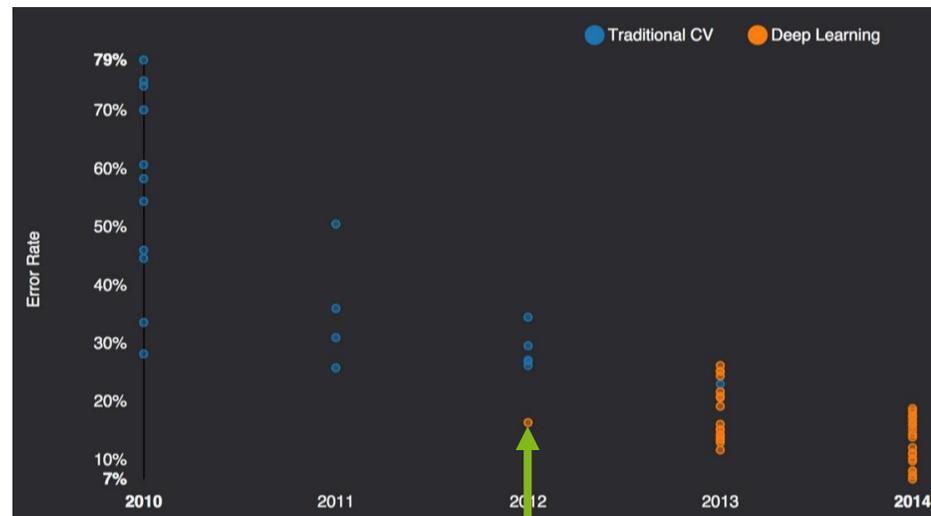
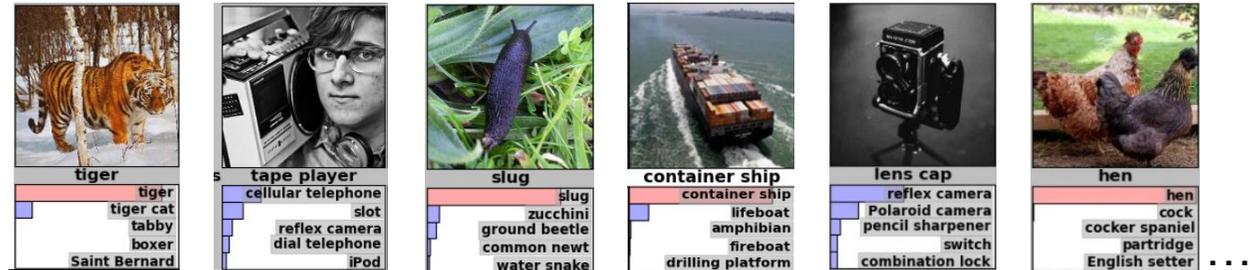
## Datenauswertung

- Erste *übermenschliche* Leistung in Mustererkennungsaufgaben seit 2012
- Auslöser: **Deep Learning** Verfahren



# Hintergrund I: Der ImageNet Wettbewerb

1000 "Klassen"  
1 Mio. Beispielbilder



Fortsetzung 2015:

4.95% Microsoft (Feb 6)

→ lässt menschliche Leistung (5.1%) hinter sich

4.8% Google (Feb 11)

4.58% Baidu (May 11)

...

→ **Computer lernen zu sehen**

# Hintergrund II: Kernidee “feature learning”

Klassischer Ansatz

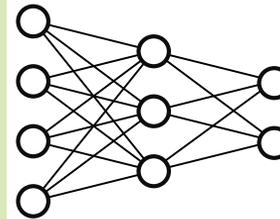


Manuell konfigurierte  
Merkmalsextraktion  
(z.B. LBP, HOG, ...)

(0.2, 0.4, ...)

(0.4, 0.3, ...)

Traditionelle Lernverfahren  
(z.B. SVM, Neuronales Netz, ...)



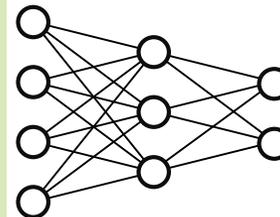
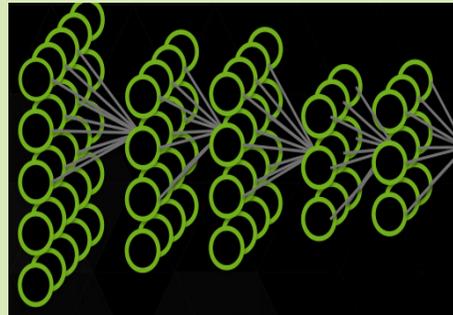
Frachtschiff

Tiger

Neuer Deep Learning  
Ansatz via CNNs  
(Convolutional  
Neural Networks)



Verarbeitet automatisch  
Rohdaten zu visuellen Merkmalen



Frachtschiff

Tiger

# Beispiel: Data-driven Condition Monitoring

## Ein KTI-Projekt des ZHAW Datalab mit mechmine llc

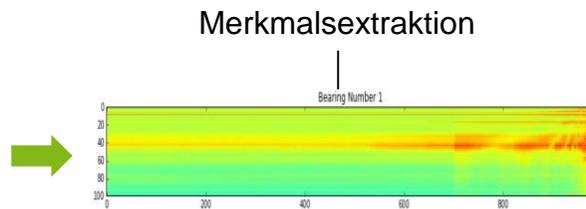
**Situation:** Wartung grosser (Rotations-)Maschinen ist teuer, Defekte sind noch teurer

**Ziel:** Wartung dann einplanen, wenn Defekt vorausgesagt wird, nicht einfach regelmässig

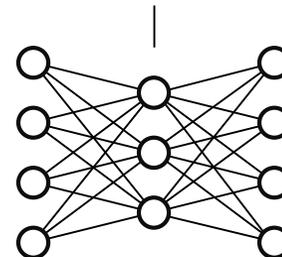
**Herausforderung:** Ansatz finden, der sich automatisch an neue Maschinen anpasst

**Lösung:** Anomalieerkennung aus dem Machine Learning anwenden, um den Normalzustand jeder Maschine sowie Abweichungen davon rein aus Sensordaten herauslesen zu können; Kombination klassischer Ansätze laut Industriestandard sowie u.a. Deep Learning Autoencoders

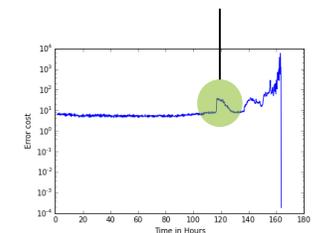
Vibrationssensor



z.B. RNN Autoencoder



Früherkennung von Defekten



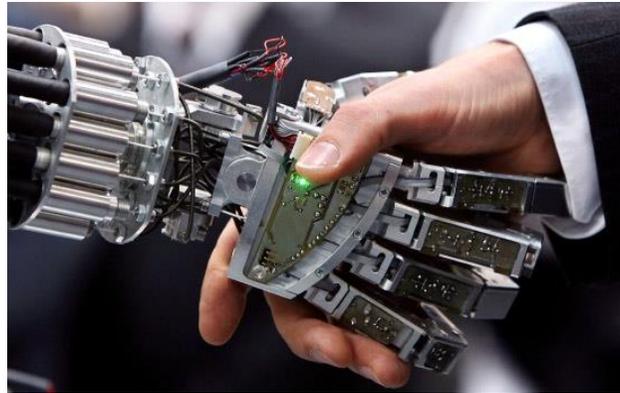
# Wohin führt die Reise? Eine Prognose.

## Maschinen

- ...machen, worin sie gut sind: «Number Crunching», nicht müde werden, Muster erkennen

## Menschen

- ...machen, worin sie unersetzbar sind: Entscheidungen treffen



➔ Datenanalyse in Zeiten von Big Data hat folgenden Wert:  
Ein Navigationsgerät – es zeigt Routen auf, wir dürfen selber fahren

# Zusammenfassung

- Es braucht *nicht viele* Daten, um im Big Data Zeitalter zu profitieren
- Es braucht gute Leute, gut ausgebildet als *Data Scientists*
- Mehrwert kommt aus Data Products, die Daten verwenden, aber *Lebensqualität liefern*
- Analytics- und Storage Technologien entwickeln sich *rasant* weiter
- Es werden immer *Kombinationen von Ansätzen* eingesetzt



## Mehr zu mir:

- Leiter ZHAW Datalab
- [thilo.stadelmann@zhaw.ch](mailto:thilo.stadelmann@zhaw.ch)
- 058 934 72 08
- [www.zhaw.ch/~stdm](http://www.zhaw.ch/~stdm)

## Mehr Data Science?

- Weiterbildung [www.zhaw.ch/datalab](http://www.zhaw.ch/datalab)
  - Konferenz: [www.zhaw.ch/datalab/sds2016](http://www.zhaw.ch/datalab/sds2016)
  - Projekte: [datalab@zhaw.ch](mailto:datalab@zhaw.ch)
  - Verband data+service: In Gründung, offen für Mitglieder.
- ➔ Fragen Sie gerne an.