

Quo Vadis Big Data: Sicherheit, Schutz und Vertrauen in Big Data

Prof. Dr. Hannes Federrath
Präsident der Gesellschaft für Informatik

Big Data Days 2018, BMBF Berlin, 11. April 2018

Schlagworte im Zusammenhang mit Big Data

- Always Online
- Cloud Computing
- Bring Your Own Device
- Internet of Things
- Data Mining

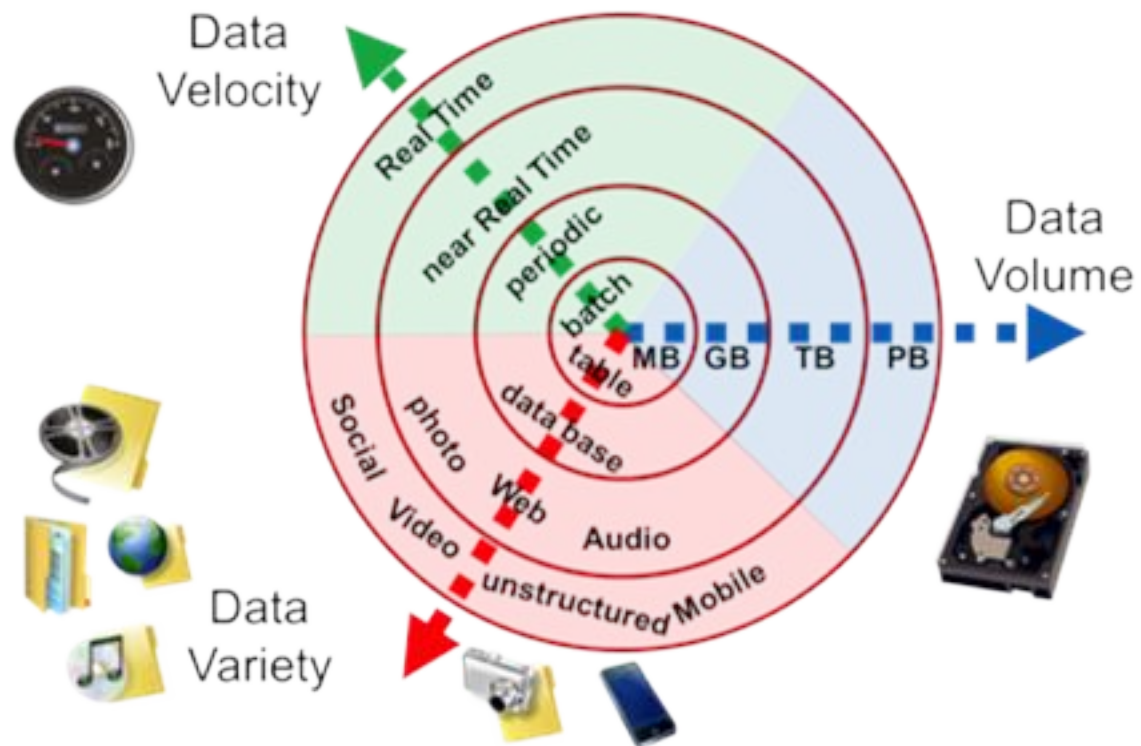


Bild: <https://gi.de/informatiklexikon/big-data/>

Definition von Big Data

Big Data bezeichnet Datenmengen, die

- zu groß sind, oder/und
- zu komplex sind, oder/und
- sich zu schnell ändern und daher

mit herkömmlichen Datenbanksystemen und Datenverarbeitungssystemen nicht mehr effektiv gespeichert und verarbeitet werden können oder durch Anwendung neuer Methoden neue Erkenntnisse aus diesen Datenmengen gewonnen werden können.

in Anlehnung an http://de.wikipedia.org/wiki/Big_Data (2014)



Zukunftsfragen

- Wie wird Big Data und IT in Zukunft...
 - unsere Kommunikation prägen?
 - unsere Arbeit prägen?
 - unser Privatleben prägen?
- Oder eben allgemeiner:
 - unsere Gesellschaft prägen?

Big Data als Instrument für

- Operations Research
- Business Intelligence
- Behavioral Targeting
- Human Resources Management
- Signal Intelligence



Stecknadeln im Heuhaufen



Social networks and personal data – the «Strava Heatmap» case

- Fitness tracker website publishes favourite running paths of their users
- US military personnel was (officially) equipped with fitness tracking devices
- Public heatmap discloses secret locations of military bases



Sources:

<https://twitter.com/Nrg8000/status/957318498102865920>

7 <https://www.theguardian.com/us-news/2018/jan/29/pentagon-strava-fitness-security-us-military>



Anonymized data...

...can reveal comm. relations.



URL: <https://www.politico.com/story/2018/03/05/what-taxi-data-shows-about-the-feds-contact-with-bankers-383751>

What taxi data shows about the Fed's contact with bankers

By VICTORIA GUIDA | 03/05/2018 12:59 PM EST

Share on Facebook | Share on Twitter

Contact between the Federal Reserve Bank of New York and six of the largest U.S. banks seems to increase around the Fed's key interest-rate-setting meetings, according to a new academic study that raises questions about whether this could give top Wall Street institutions an unfair competitive edge.

The study, from the University of Chicago's Booth School of Business, examines granular data on cab rides released by New York's taxi regulator. It singles out lunchtime trips between the New York Fed and the major offices of six banks: Goldman Sachs, Citigroup, JPMorgan Chase, Morgan Stanley, Bank of New York Mellon and Bank of America.

It also includes potential off-site meetups by factoring in "coincidental drop-offs," where someone from the New York Fed and someone from a bank each appeared to be dropped off at the same location around the same time.

Lunchtime coincidental drop-offs happened about 50 percent more often between when an important meeting of the policy-making Federal Open Market Committee started through the following week, according to the study written by David Andrew Finer. That's an average of about 1.2 more taxi rides per meeting.

"I cannot conclusively demonstrate a link between rides and face-to-face meetings, but evidence that individuals are in very close proximity to each other more often around FOMC meetings would complement more indirect evidence of regular informal communication," the study says. It examines taxi data between 2009, when data was first made available, and 2014, before ride shares like Uber and Lyft became popular.

David Andrew Finer: What Insights Do Taxi Rides Offer into Federal Reserve Leakage? Working Paper, Booth School of Business, University of Chicago, March 2018. <https://research.chicagobooth.edu/-/media/research/stigler/pdfs/workingpapers/18whatinsightsdotaxiridesofferintofederalreserveleakage.pdf>

Data Science

- Data Science beschäftigt sich mit einer zweckorientierten Datenanalyse und der systematischen Generierung von Entscheidungshilfen und -grundlagen.
- Interdisziplinäres Forschungsfeld
 - Informatik, Statistik, Mathematik
 - Naturwissenschaften
 - Wirtschaftswissenschaften
- Kompetenzen
 - Programmierung
 - Datentechnik
 - Mustererkennung
 - Prognostik
 - Modellierung
 - Sicherheit
 - Datenlagerung

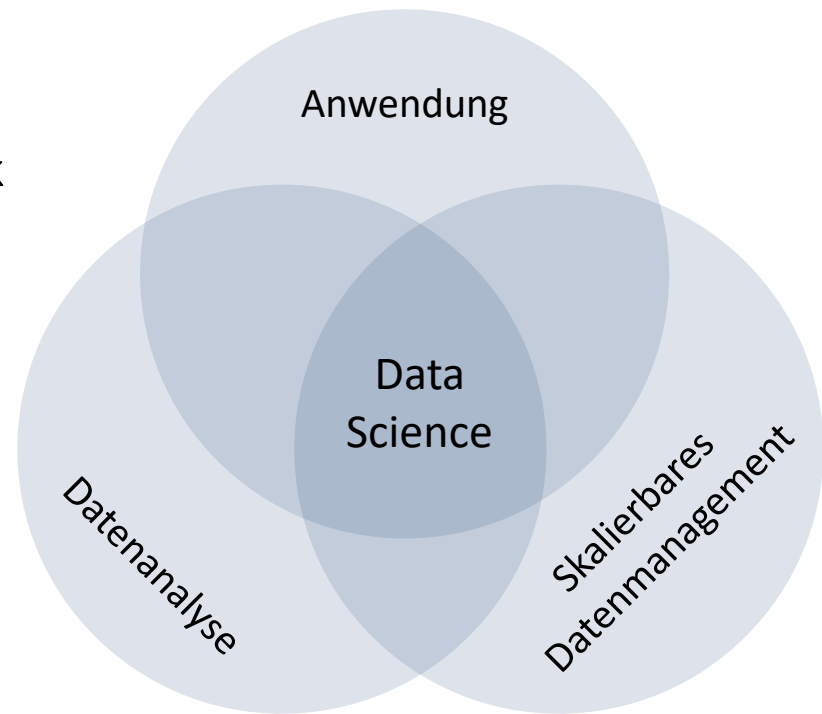


Bild: Markl 2015

Die GI hat eine Task Force «Data Science» zur Präzisierung der Kompetenzen des Data Scientists gegründet.

GI-Beitrag «Turing Bus» zum Wissenschaftsjahr 2018

■ Projekt «Turing Bus»

- benannt nach Alan Turing
- richtet sich an Schülerinnen und Schüler



Bild: Wikipedia

■ Infonauten

- Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler des interdisziplinären Fachs «Informatik und Gesellschaft» (luG)

■ Passagiere

- junge Erwachsene (15-25 Jahre) während der Berufs- und Studienwahl

■ Formate

- Vorträge, Schulungen und Hackathons
- Lehr- und Lernmaterialien veröffentlicht
- Landkarte von luG-Projekten

Die GI ist Partner des Wissenschaftsjahres 2018 «Arbeitswelten der Zukunft».

GI-Beitrag «Turing Bus» zum Wissenschaftsjahr 2018

- Projekt «Turing Bus»
 - benannt nach Alan Turing
 - richtet sich an Schülerinnen und Schüler

Die GI ist Partner des Wissenschaftsjahres 2018 «Arbeitswelten der Zukunft».

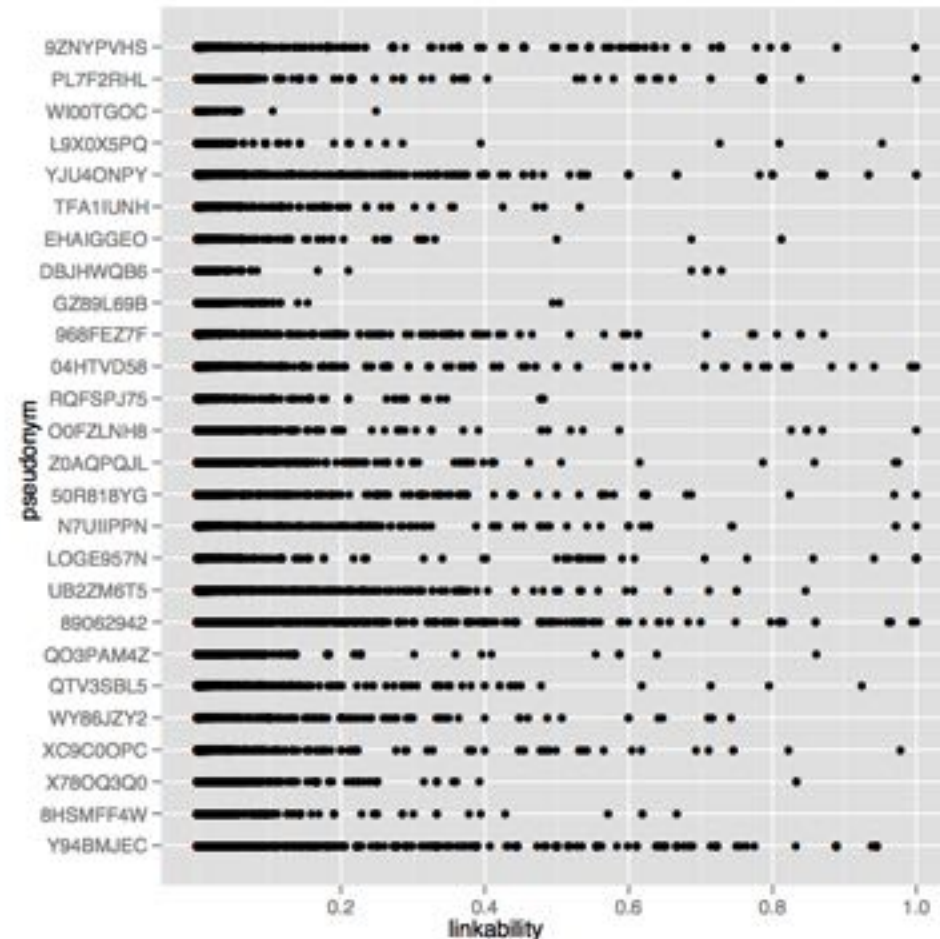


Datenanalyse

- Analyse der Daten auf eine bestimmte Fragestellung hin
- Big Data: Vor Auswertung keine Vorverarbeitung (Aufbereitung) der Daten.
Gründe:

- nicht möglich
- zu ineffizient
- statistisch nicht signifikant
- Analyse von Log-Files,
- Verknüpfung von Datenbanken
- Maschinelles Lernen
- Komplexitätstheorie kann auch bei Big Data nicht ausgeschaltet werden
 - Kryptographie: Knacken von Schlüsseln bleibt exponentiell

Bild: Gerber 2009

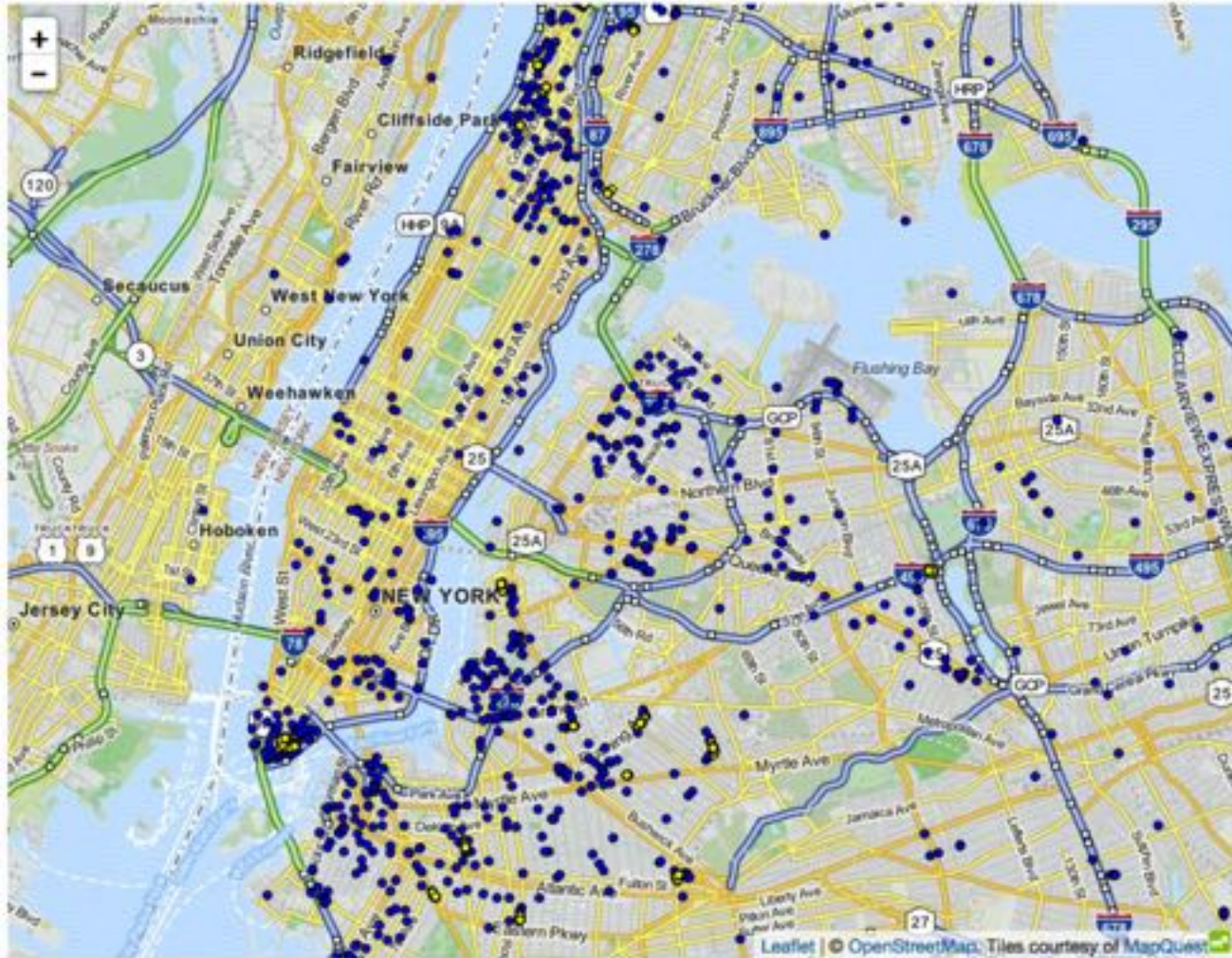


Pseudonymized data...

...can reveal
sensitive location
information.

20 GBytes of
pseudonymized
data

170 million taxi
rides in New York



Drop-off locations for trips starting at Larry Flynt's Hustler Club between midnight and 6 am during 2013.

Source: <http://content.research.neustar.biz/blog/differential-privacy/stripeRaw.html>



Der Fall Cambridge Analytica

<

>

☰

heise.de

↺

⬇

⬆

📄

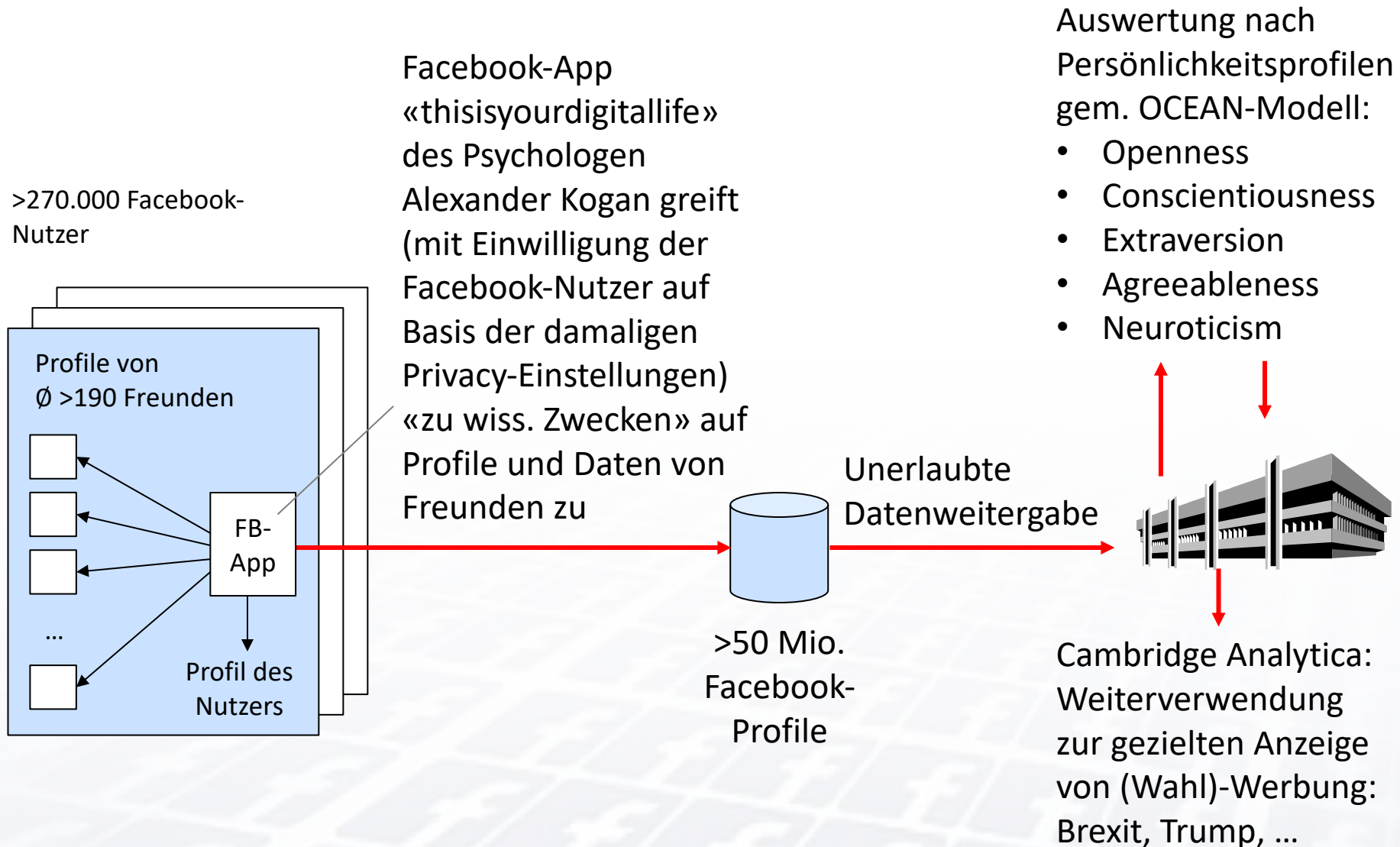
+

Facebook-Datenskandal

Die englische Datenanalyse-Firma Cambridge Analytica hat sich während des US-Wahlkampfes unerlaubt Zugang zu Daten von mehr als 50 Millionen Facebook-Nutzern verschafft. Mit den Informationen soll die Firma geholfen haben, Anhänger des heutigen US-Präsidenten Donald Trump zu mobilisieren und zugleich potenzielle Wähler der Gegenkandidaten Hillary Clinton vom Urnengang abzuhalten. Mithilfe der Nutzerprofile wurden im sozialen Netzwerk gezielt Botschaften als Werbung ausgespielt.

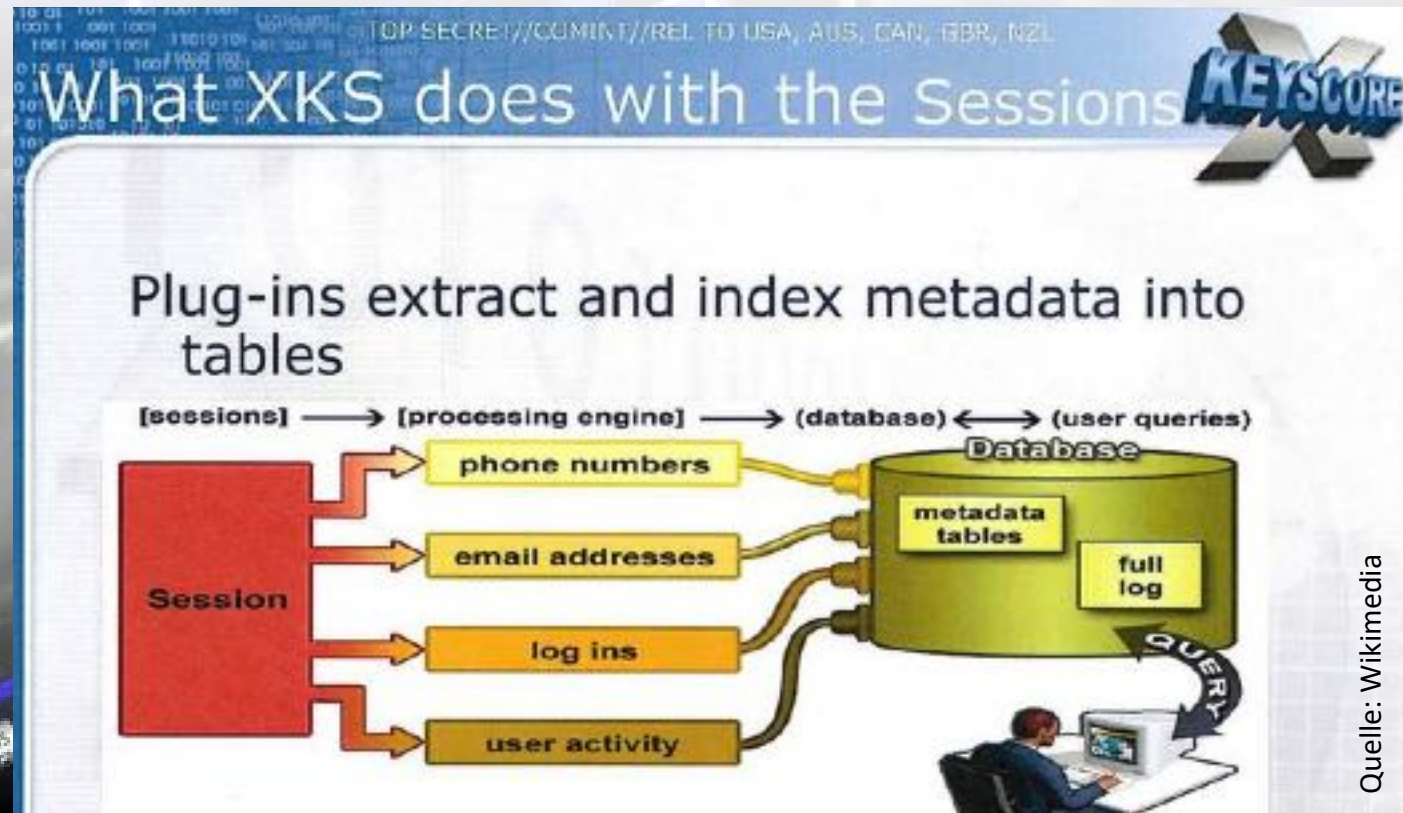


Der im März 2018 bekannt gewordene Datenskandal ist aber offenbar nur die Spitze eines Eisbergs. Laut einem ehemaligen Facebook-Manager habe das Social Network keinerlei Kontrolle über abgeflossene Nutzerdaten.



Big Data – Die Technik im Mittelpunkt!

- unbegrenzter Speicher
- viel Bildschirmfläche
- large-scale computing
- hoher Stromverbrauch
- starke Verkettbarkeit



Big Data – Der Mensch im Mittelpunkt?

- große Transparenz (wünschenswert)
- starke Offenheit (erwartet)
- starke Kontrolle (notwendig)
- Orwells Metapher in 1984
 - Lückenlose Überwachung und Bevormundung von Menschen durch Menschen (Big Brother)
- Neue Metapher in 20xx
 - Lückenlose Überwachung und Profilbildung von Menschen durch Computer (Big Data) ... durch Menschen



Auszug aus Artikel 22 DSGVO

Art. 22 Automatisierte Entscheidungen im Einzelfall einschließlich Profiling

Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.

...

Auszug aus Erwägungsgrund 71:

Die betroffene Person sollte das Recht haben, keiner Entscheidung [...] unterworfen zu werden, die ausschließlich auf einer automatisierten Verarbeitung beruht und die rechtliche Wirkung für die betroffene Person entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt, wie die automatische Ablehnung eines Online-Kreditanspruchs oder Online-Einstellungsverfahrens ohne jegliches menschliche Eingreifen. Zu einer derartigen Verarbeitung zählt auch das „Profiling“, das in jeglicher Form automatisierter Verarbeitung personenbezogener Daten unter Bewertung der persönlichen Aspekte in Bezug auf eine natürliche Person besteht, insbesondere zur Analyse oder Prognose von Aspekten bezüglich Arbeitsleistung, wirtschaftliche Lage, Gesundheit, persönliche Vorlieben oder Interessen, Zuverlässigkeit oder Verhalten, Aufenthaltsort oder Ortswechsel der betroffenen Person, soweit dies rechtliche Wirkung für die betroffene Person entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.

Automatisierte Entscheidungen im Einzelfall einschl. Profiling

■ DSGVO Art. 22 (1) Automatisierte Entscheidungen

- automatisierte Entscheidungen verletzen das Persönlichkeitsrecht
- Profiling verletzt das Persönlichkeitsrecht
- Erwägungsgrund 71 nennt Beispiele:
 - Online-Kreditantrag
 - Online-Einstellungsverfahren
 - Analyse oder Prognose von
 - Arbeitsleistung
 - wirtschaftlicher Lage
 - Gesundheit
 - persönlichen Vorlieben oder Interessen
 - Zuverlässigkeit oder Verhalten
 - Aufenthaltsort oder Ortswechsel

Ethische Dimensionen

- Dreiteilung nach Simon, 2016:

- Ethik des Berufs
- Ethik des Designs
- Ethik der Nutzung



vorangegangene
Beispiele

Ethische Dimensionen

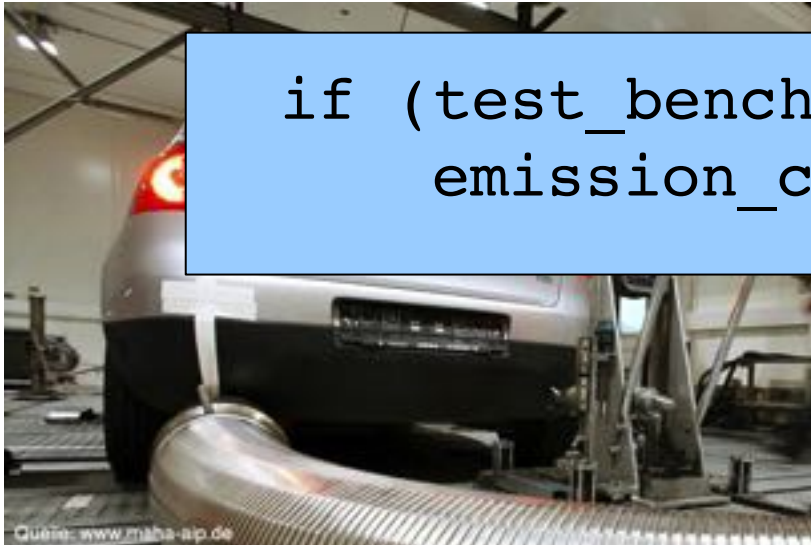
■ Dreiteilung nach Simon, 2016:

- Ethik des Berufs
- Ethik des Designs
- Ethik der Nutzung



```
if (test_bench_detected())  
    emission_clean(FULL)  
else  
    <ignore_laws>
```

Abgasmessungen auf dem Prüfstand und mobil



```
if (test_bench_detected())  
    emission_clean(FULL)
```

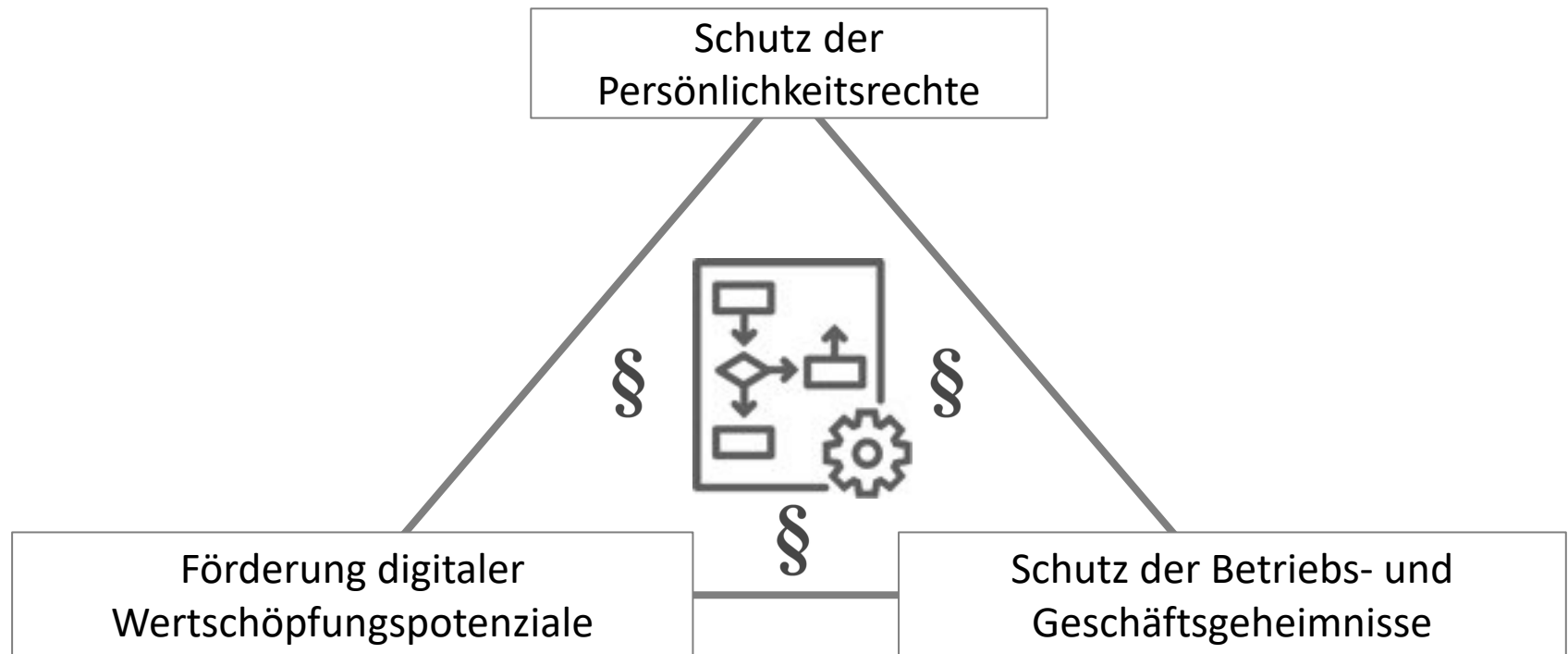
(Pseudocodedarstellung)



- Adaptive Motorsteuerung erkennt anhand von Lenkradbewegungen, Umgebungsluftdruck, Raddrehzahlen und Motorlaufzeit die Messbedingungen

	Abgasreinigung	Kraftstoffverbrauch
Prüfstand	verbessert	hoch
Straße	reduziert	verringert

Spannungsfeld der Regulierung von Algorithmen



nach: Martini 2017

GI vom BMJV mit einer Machbarkeitsstudie für ein Algorithmengesetz beauftragt.

Ansatzpunkte für die Regulierung von Algorithmen



- Art. 22 DSGVO
- Transparenz
 - Kennzeichnungspflicht
 - Begründungspflicht
 - Transparenzanforderungen
- Ex-ante-Kontrolle
- AGG

- Kontrollalgorithmen & Standardisierung
- Risikomanagement
- Protokollierung der Programmabläufe

- Algorithmic Responsibility Kodex mit Erklärungspflicht

- Haftung
 - Beweislastverteilung
 - Gefährdungshaftung
- Erweiterung prozess. Handlungsräume
 - Abmahnbefugnisse
 - Verbandsklagerecht / Schiedsstelle
 - Nebenfolgenkompetenz

nach: Martini 2017

GI vom BMJV mit einer Machbarkeitsstudie für ein Algorithmengesetz beauftragt.

Ethische Dimensionen

■ Dreiteilung nach Simon, 2016:

- Ethik des Berufs
- Ethik des Designs
- Ethik der Nutzung



Association for
Computing Machinery



IEEE
*Advancing Technology
for Humanity*

Ethische Leitlinien der Gesellschaft
für Informatik

ACM Code of Ethics and Professional
Conduct

IEEE Code of Ethics

Die ethischen Leitlinien der Gesellschaft für Informatik



Präambel

Art. 1 Fachkompetenz

Art. 2 Sachkompetenz und kommunikative Kompetenz

Art. 3 Juristische Kompetenz

Art. 4 Urteilsfähigkeit

Art. 5 Arbeitsbedingungen

Art. 6 Organisationsstrukturen

Art. 7 Lehre und Lernen

Art. 8 Forschung

Art. 9 Zivilcourage

Art. 10 Soziale Verantwortung

<https://gi.de/ueber-uns/organisation/unsere-ethischen-leitlinien/>



INFORMATIK BEWEGT,
INFORMATIK IST ZUKUNFT,
WIR SIND INFORMATIK!