



BIG DATA

Warum sollte ich auf meine Daten achten?

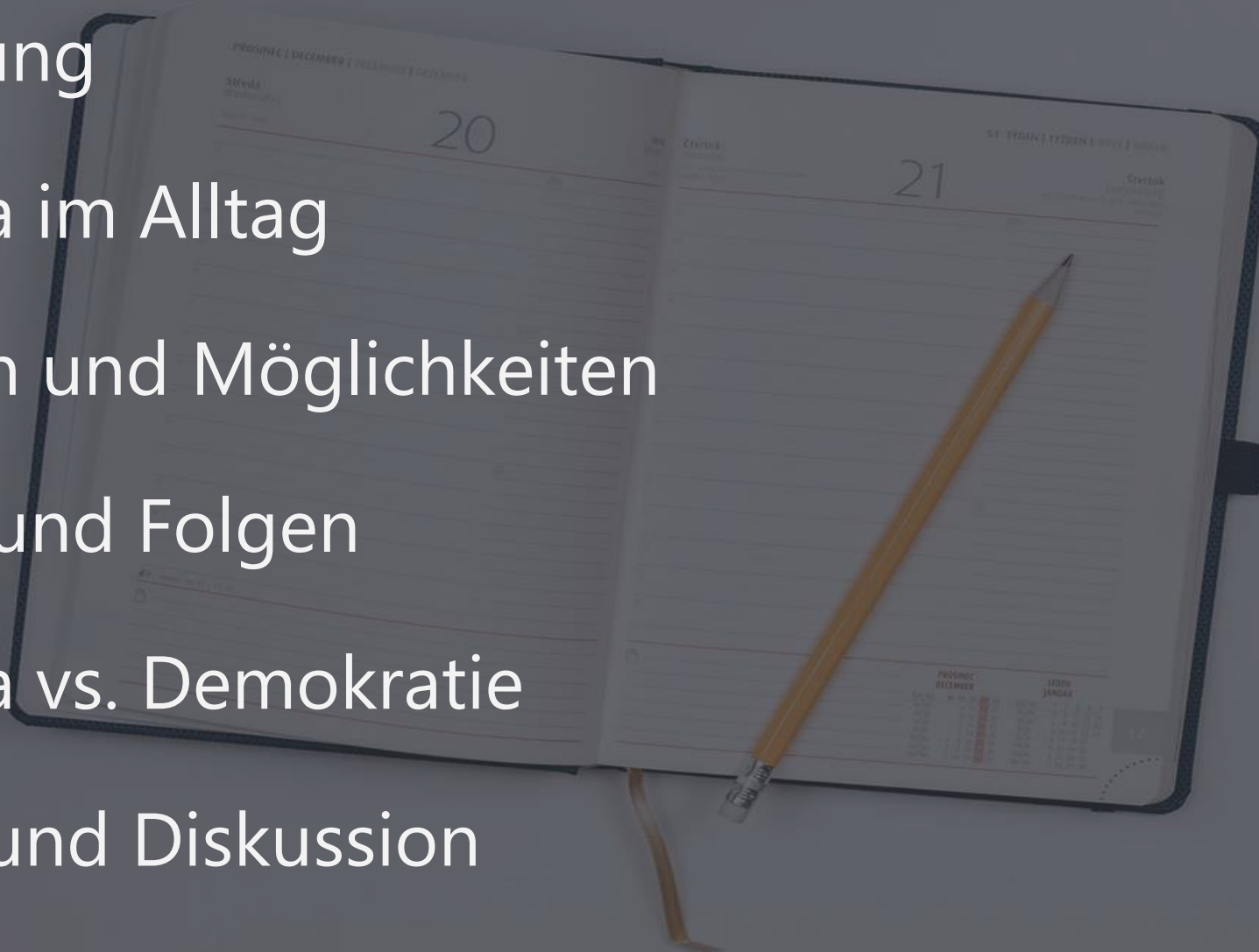
16. Demokratie-Tag Rheinland-Pfalz
28.10.2021

Autor: Jonas Grasediek

© Brett Sayles, pexels.com

Agenda

- Einführung
- Big Data im Alltag
- Chancen und Möglichkeiten
- Risiken und Folgen
- Big Data vs. Demokratie
- Fragen und Diskussion



Einführung

„Die wertvollste Ressource der Welt ist nicht länger das Öl, sondern es sind Daten“

Kiran Bhageshpur, The Economist, 2017

Einführung – Daten sind das neue Öl



Öl-Gewinnung

- Orten
- Erschließen
- Raffinieren
- Speichern
- Der weiteren Verarbeitung zuführen
- Enormer ökonomischer Wert
- Industrielle Revolution



Daten-Gewinnung

- Orten
- Erschließen
- Konsolidieren
- Speichern
- Der weiteren Verarbeitung zuführen
- Enormer ökonomischer Wert
- Digitale Revolution

Quellen:

Dr. Kilian Semmelmann, Daten sind das neue Öl – Herkunft des Zitats und der Stand heute, 2020
<https://datadrivencompany.de/daten-sind-das-neue-oel-herkunft-des-zitats-und-der-stand-heute/>
Maximilian Th. L. Rückert, Die digitale Revolution, 2018
https://www.hss.de/download/publications/PS_479_Digitale_Revolution_02.pdf

Autor: Jonas Grasediek

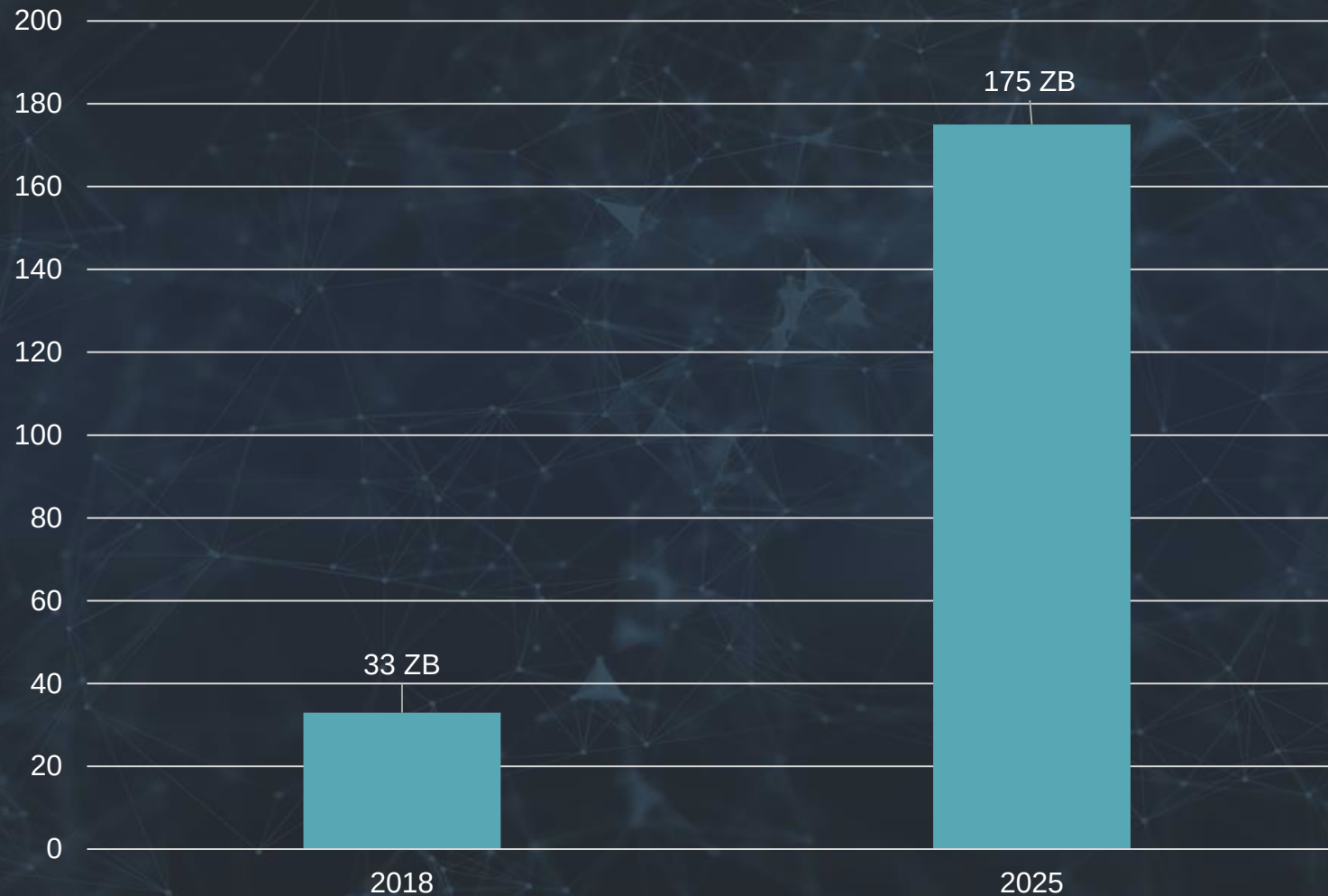
Einführung – Definition

Big Data bezeichnet die Analyse großer Datenmengen aus vielfältigen Quellen in hoher Geschwindigkeit mit dem Ziel, wirtschaftlichen Nutzen zu erzeugen.

Einführung – Definition



Einführung - Big Data



**Prognose zum weltweit
generierten Datenvolumen
bis 2025**

Einführung - Big Data

- 2018 wurden weltweit 33 Zettabyte (ZB) Daten generiert
- Das sind 33 Mrd. Terabyte (TB)
- Das entspricht rund
 - 8,25 Mrd. handelsüblichen 4TB Platten
 - 129 Mrd. iPods mit 256GB Speicher
- Um diese Menge zu veranschaulichen
 - stellen wir uns vor, 1TB würde dem Gewicht einer Büroklammer entsprechen (rund 1 Gramm)
 - Dann würde die weltweite generierte Datenmenge immer noch 33.000 Tonnen entsprechen
 - Das ist in etwa das dreifache Gewicht des Eiffelturms
- Bis 2025 wird ein Anstieg auf 175ZB erwartet

Einführung – Vielfältige Quellen

- Jeder Mensch erzeugt täglich ca. 1,5GB Daten
- Google erweitert seinen Cloud-Speicher täglich um ca. 1.000TB
- In den sozialen Medien werden täglich 4.000TB Daten produziert
- Auf ebay werden täglich 12TB Kundendaten verarbeitet. Die Gesamtgröße der Kundendatenbank beträgt derzeit ca. 4.000TB
- Im Gesundheitswesen existiert derzeit ein Datenbestand von 150.000TB
- Das Internet of Things (IoT) verfügt derzeit über ca. 20,4 Mrd. Geräte, die zusammen 1,6 Mrd. TB Daten pro Jahr erzeugen. Pro Tag werden derzeit ca. 5,5 Mio. neue Sensoren hinzugefügt
- Ein Flugzeug (z.B. A320) besitzt 3.000 Sensoren pro Tragfläche. Pro Flugzeug werden täglich ca. 2,5TB Daten erzeugt
- Im Weltraum (Satelliten, Raumfahrt, etc.) fallen derzeit täglich ca. 1.200TB Daten an
- Autonomes Fahren ist für ca. 4TB Daten pro Tag verantwortlich. In Zukunft werden pro Sekunde und Fahrzeug ca. 1GB Daten erwartet.
- In der Genomforschung fallen jährlich ca. 40.000 TB Daten an.

Quelle: Walter Schadhauser und Rainer Graefen, Wo kommen all die Daten her?, 2017

<https://www.storage-insider.de/wo-kommen-all-die-daten-her-a-608824/>

Michael Kroker, Die Größe des globalen Datenbestands von 33.000 Exabytes anschaulich umgerechnet.

<https://blog.wiwo.de/look-at-it/2019/03/06/die-groesse-des-globalen-datenbestands-von-33-000-exabytes-anschaulich-umgerechnet/>

Autor: Jonas Grasediek

© TheDigitalArtist, pixabay

Einführung – hohe Geschwindigkeiten

- Die derzeitige weltweite Internetbandbreite beträgt 130.487 GB/S
- Das bedeutet pro Sekunde werden ca.
 - 9.961 Tweets auf Twitter abgesetzt
 - 1.107 Fotos auf Instagram hochgeladen
 - 1,2 Stunden Videomaterial auf Youtube hochgeladen
 - 96.116 Suchanfragen bei Google gestartet
 - 3 Mio. E-Mails versendet, wovon ca. 67% (= 2,01 Mio. E-Mails) auf SPAM entfällt

Einführung – Wirtschaftlicher Nutzen erzeugen

- Beispiele für Bereiche in denen wirtschaftlicher Nutzen erzeugt wird, sind:
- Personalisierte Werbung
 - Durch Vermarktung von Daten wie Vorlieben, Kaufverhalten, verwendete Zahlungsmittel, etc.
 - Insbesondere soziale Netzwerke verkaufen diese Daten
- Onlinehandel
 - Erzeugung von besserem Kundenverständnis
 - Erschließung neuer Umsatzquellen
- Prozessoptimierung
 - Mit Hilfe der Bereitstellung aufbereiteter Daten und Ergebnisse (z.B. im Produktionsbereich) können Prozesse effizienter und somit gewinnbringender eingesetzt werden.
- Agrarsektor
 - Durch Verarbeitung von großen Daten in Echtzeit können Bauern ihre Feldnutzung optimieren
- Industrie
 - Durch die Errechnung des optimalen Zeitpunkts für die Beschaffung von Rohmaterialien können große Industrieunternehmen Ausgaben reduzieren
- Börsenhandel
 - Börsenmanager nutzen mathematische Modelle auf Basis großer Datenmengen aus der Wirtschaft, Börse, etc. um den besten Zeitpunkt für bestimmte Investments zu bestimmen

Quellen:
Marco Visibelli, Wired, How Big Data Can Mean Big Money — or Big Losses, 2018
<https://www.wired.com/insights/2014/08/big-data-can-mean-big-money-big-losses/>

Philipp Hummel, Heise, Ackern nach Zahlen: Big Data in der Landwirtschaft, 2014
<https://www.heise.de/newsticker/meldung/Ackern-nach-Zahlen-Big-Data-in-der-Landwirtschaft-2240801.html>

BI-Survey, How Do Companies Monetize Their Data?
<https://bi-survey.com/data-monetization>

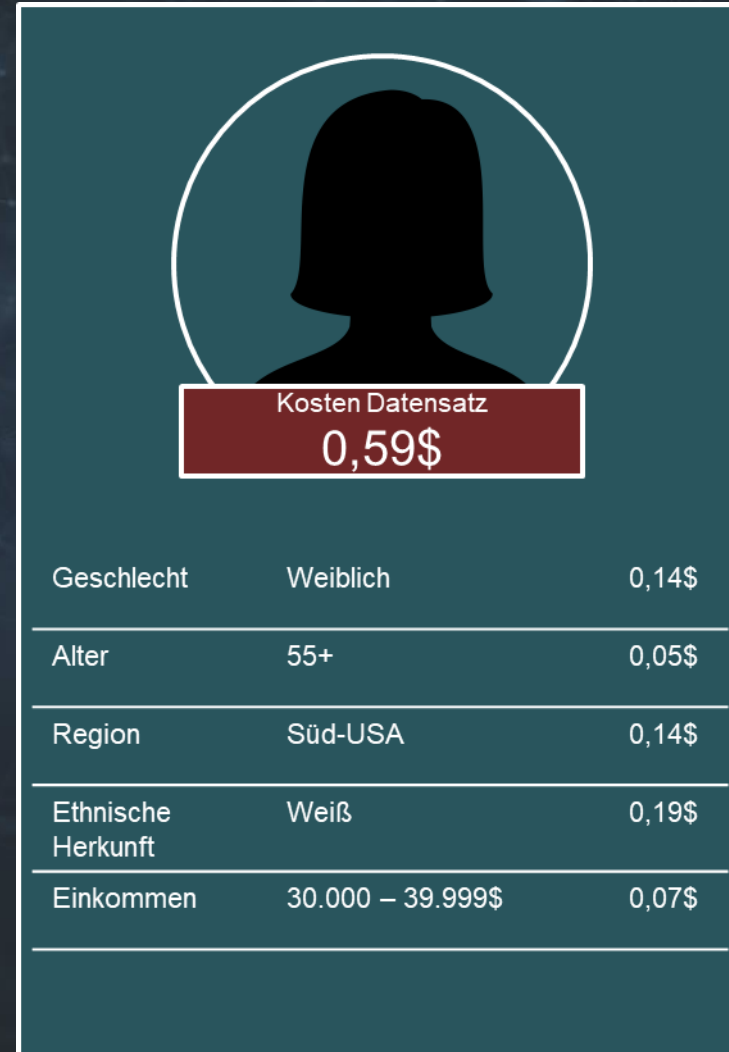
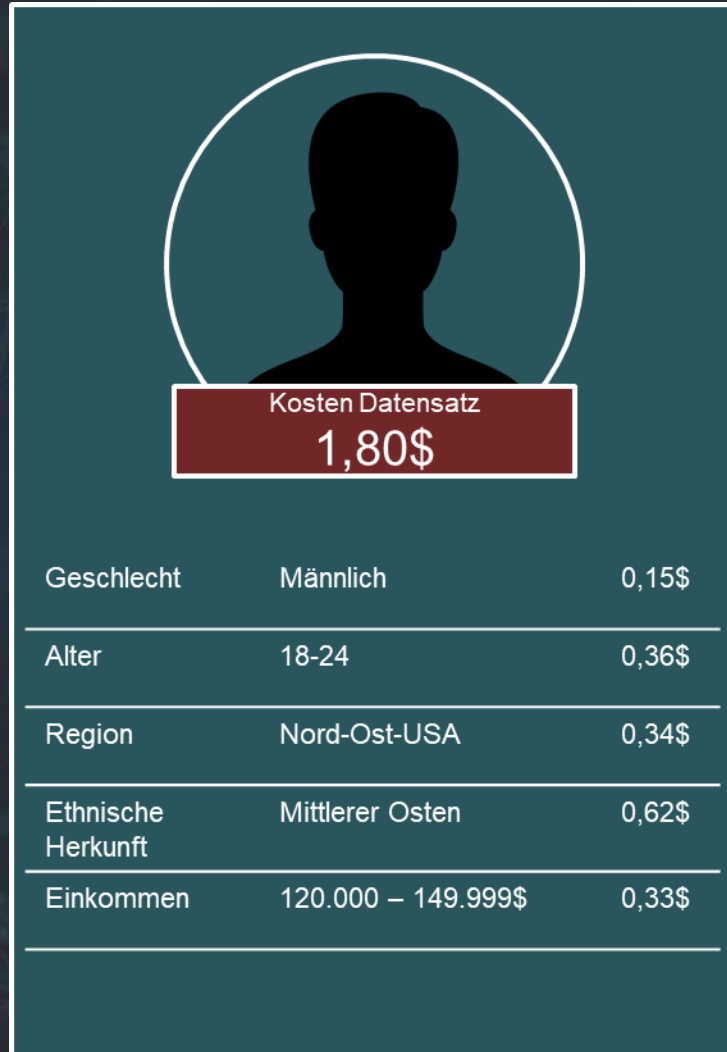
Autor: Jonas Grasediek

Einführung – Was sind meine Daten Wert?

Daten einer amerikanischen Studie zum monetären Wert von Daten

Geschlecht	Männlich	0,15\$
	Weiblich	0,14\$
Alter	18-24	0,36\$
	55+	0,05\$
Ethnische Herkunft	Mittlerer Osten	0,62\$
	Weiß	0,19\$
Jährliches Einkommen	40.000 – 49.999\$	0,02\$
	120,000 – 149.999\$	0,33\$

Einführung – Was sind meine Daten Wert?



Exemplarische
Wertermittlungen
der Studie im
Vergleich

Einführung – Was sind meine Daten Wert?

- Große IT-Unternehmen wie Facebook und Google erzeugen Umsatz indem sie die Daten ihrer Nutzer für Werbezwecke verkaufen
- Unternehmen können dann personalisierte Werbung auf der jeweiligen Plattform schalten
- Im Jahre 2020 wurden somit folgende Werbeeinnahmen erzeugt:
 - LinkedIn – 3 Mrd. \$
 - Facebook – 85 Mrd. \$
 - Google – 150 Mrd. \$

Quellen: Megan Graham , CNBC, Microsoft says LinkedIn topped \$3 billion in ad revenue in the last year, outpacing Snap and Pinterest, 2021

<https://www.cnbc.com/2021/04/27/microsoft-linkedin-topped-3-billion-in-ad-revenue-in-last-year.html>

Martin Coulter, Business Insider, YouTube is a powerhouse for Google, with revenue up 46% to almost \$7 billion, 2021

<https://www.businessinsider.com/youtube-google-revenue-results-covid-19-2021-2>

statista, Facebook's global revenue as of 2nd quarter 2021, by segment, 2021

<https://www.businessinsider.com/youtube-google-revenue-results-covid-19-2021-2>

Megan Graham, CNBC, How Google's \$150 billion advertising business works, 2021

<https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html#:~:text=of%20the%20future,-.But%20Google's%20main%20business%20is%20online%20advertising..the%20company's%202020%20annual%20report.>

Autor: Jonas Grasediek

Big Data im Alltag – Internet of Things

- Jedes physikalische Objekt kann theoretisch mit dem Internet verbunden sein
- 8,4 Mrd. IoT-Geräte in 2020 – 25,4 Mrd. Geräte bis 2030
- ~ 5,5 Mio. neue Sensoren pro Tag
- Daten von
 - Kühlschränken, Waschmaschinen, Kaffeemaschinen
 - Türen, Fenstern, Rollläden
 - Glühbirnen, Uhren, Bilderrahmen
 - Lautsprecher, Fernbedienungen
 - Heizungen, Thermostate
 - Kleidung, Fahrräder, Rucksäcke

Big Data im Alltag – Internet of Things

- Steigerung des persönlichen Komforts
- Steigerung der persönlichen Effizienz
- Aufbereitete Informationen über alle wesentlichen Vorgänge im eigenen Umfeld
- Frühzeitige Erkennung von Gerätedefekten
- Effizienzsteigerung einzelner Geräte
- Sicheres Zuhause durch Überwachung aller Zugänge
- Steigerung der Lebensqualität für die Gesellschaft (z. B. durch Smart Cities)

Quellen:

Marco Visibelli, Wired, How Big Data Can Mean Big Money — or Big Losses, 2018
<https://www.wired.com/insights/2014/08/big-data-can-mean-big-money-big-losses/>
Tas Bindi, ZDNet, Google and Levi's unveil internet-connected jacket, 2014
<https://www.zdnet.com/article/google-and-levis-unveil-internet-connected-jacket/>
Geeta Puri, Thingsphere, 5 Benefits of IOT – for “We, The People”, 2019
<https://thingsphere.com/5-benefits-of-iot-for-we-the-people/>

Autor: Jonas Grasediek

Big Data im Alltag – Streaming

- Zeitangaben
 - Datum und Zeitpunkt
 - Wann und wie lange wurde pausiert?
 - Dauer bis zum Abschluss einer Folge, Staffel, Serie, etc.
 - Zeitpunkt im Abspann, an dem die nächste Folge gestartet wurde
- Geräte
 - Welches Gerät wurde verwendet?
 - Wird zwischendurch das Gerät gewechselt?
 - Anzahl und Gegenstand der Suchen
- Vorlieben
 - Wurde nach einer Pause weitergeschaut?
 - Positiv-Negativ-Bewertungen
 - Wurden alle Inhalte zu Ende gesehen
 - Welche Genres werden bevorzugt? z.B. Horror, Action, Drama
 - Welche Themen werden bevorzugt? z.B. Filme mit Pferden, Filme mit weiblichen Hauptfiguren, etc.

Big Data im Alltag – Streaming

- Persönliche Film- und Serienempfehlungen
- „Nächste Folge ansehen“ entsprechend dem persönlichen Rhythmus
- Möglichst Präzise Vorhersagen über Interessen und Trends (z. B. saisonal)
- Entscheidung über Produktion und Fortführung neuer Filme und Serien
- Mehr Abonnenten = Mehr Geld

Big Data im Alltag – Banken

- Kontobewegungen
- Depottransaktionen
- Nutzungsverhalten
- Genutzte Kanäle (z. B. Online-Banking oder Filiale)
- Zahlungsempfänger
- Regelmäßige Buchungen / Daueraufträge
- Sparpläne
- Immobilienpreise
- Wertpapierkurse
- Anleihen, Fonds, etc.

Big Data im Alltag – Banken

- Individuell angepasste Finanzierungsmodelle
- Individuelle Finanzprodukte
- Optimierung des eigenen Ausgabeverhaltens
- Schnelle Erkennung ungewöhnlicher Kontoaktivitäten
- Besseres Kundenverständnis
- Erkennung illegale Aktivitäten, z.B. Geldwäsche
- Früherkennung künftiger Ereignisse, z.B. Lücken in der Altersvorsorge
- Schnelle Reaktion auf Engpässe

Big Data im Alltag – Wetter

- Atmosphärische Beobachtungen
 - Temperatur, Wind, Luftdruck
 - Wolkenbildung- und Bewegungen
 - Zusammensetzung der Atmosphäre
- Ozeanische Beobachtungen
 - Temperatur, Salzgehalt, Seegang
 - Wassermassenbildung
 - Nährstoffe, Sauerstoffgehalt
- Terrestrische Beobachtungen
 - Schneedecken, Gletscher und Permafrost
 - Grundwasserpegel
 - Kohlenstoffe im Boden

Big Data im Alltag – Wetter

- Prognose von Stromverfügbarkeit von Energieversorgern
- Optimierung der Flugkoordination (z.B. via Sichtdaten)
- Warnung Energieversorger vor Katastrophen
- Unterstützung von Versicherungen und Betroffenen bei Schadensereignissen
- Vorbereitung auf mögliche Katastrophen (Behörden, Energieversorger)
- Vorbereitung auf saisonale Ereignisse (z.B. Pharmaindustrie vor Pollensaison)
- Möglichst genaue Wetterprognosen
- Vorhersage von Pollenflug
- Unwetterwarnungen

Big Data im Alltag – Verkehr

- Verkehrskameras
- Magnetische Sensoren im Straßennetz
- Bewegungsdaten ÖPNV
- Passagierzahlen in Reisezentren
- Anonymisierte Handydaten der Mobilfunkanbieter
- Daten verschiedener Apps (z.B. google maps, waze, etc.)
- GPS-Daten
- Online-Bestellungen, z.B. zur Prognose des erwarteten Lieferaufkommens
- Wetterdaten (Saisonale Einflüsse, Unwetter, Katastrophen)
- Versicherungsdaten (z.B. Unfall-Hot-Spots)

Big Data im Alltag – Verkehr

- Verbesserung der allgemeinen Verkehrssicherheit
- Verbesserung der Verkehrssicherheit an sensiblen Bereichen (z.B. Schulen)
- Prognose
 - von Stoßzeiten (z.B. Berufsverkehr)
 - von Bewegungsströmen bei bestimmten Ereignissen (z.B. Fußballspiele)
 - der optimalen Reisezeit
 - der optimalen Zeiträume für Straßensanierungen
- Optimierte Ampelschaltungen (siehe Los Angeles)
- Skalierbare Angebote im ÖPNV
- Optimierte Navigationsdaten
- Planung und Verbesserung neuer Straßen und Anbindungen
- Kürzere und schneller Lieferwege

Big Data im Alltag – Medizin

- Medizinische Vorgeschichte
- Erkrankungen in der Familie
- Krankenakten anderer Patienten mit ähnlichen Symptomen
- Demografische Merkmale
- Lebensstil
- Genetische Mutationen
- DNS-Sequenzen (z.B. bestimmter Tumorzellen)
- Datensätze von smarten Geräten
- Suchanfragen nach bestimmten Krankheiten
- Nennung von Symptomen in sozialen Medien

Big Data im Alltag – Medizin

- Frühe Entdeckung und Eindämmung von
 - Ausbrüchen von Infektionskrankheiten und Seuchenausbrüchen
 - Krebserkrankungen und Erbkrankheiten
 - Diabetes, Herz-Kreislauf-Erkrankungen, Schlaganfällen
 - Psychische Erkrankungen
- Entwicklung und Optimierung von
 - Methoden zur Gesundheitserhaltung
 - Vorsorgeuntersuchungen
 - Behandlungsmethoden
 - Arzneimitteln
- Versorgung von Patienten via Telemedizin
- Verfolgung von Infektionsketten

Big Data im Alltag – weitere Beispiele

- Soziale Medien
- Minutengenaue Sendungsverfolgung
- Rabattportale
- Elektrofahrzeuge
- Schlafoptimierung
- Optimierung des Stromverbrauchs
- Kontinuierliche Kontrolle des Autozustands
- Effizientere Müllentsorgung
- Prognose und Verständnis von Kriminalität
- Präzisere Vorhersage und Unterstützung bei Katastrophen
- Industrie 4.0

Quellen:
Ajoure, Big Data: 11 Beispiele für sinnvolle Anwendungen im Alltag, 2020
<https://www.ajoure.de/lifestyle/technik/big-data-11-beispiele-fuer-sinnvolle-anwendungen-im-alltag/>
Andreas Rabe, Addis Techblog, Infografik: BIG DATA im Alltag, 2015
<https://addis-techblog.de/2015/04/infografik-big-data-im-alltag/>

Autor: Jonas Grasediek

Chancen und Möglichkeiten

- Verbesserung der Gesundheit
- Genauere Wettervorhersagen
- Unterstützung im Straßenverkehr
- Optimierung der eigenen Finanzen
- Individuelle Angebote
- Anpassung der Umwelt an die eigenen Bedürfnisse
- Unterstützung bei Entscheidungsfindung
- Steigerung der gesamten Lebensqualität
- Frühwarnsysteme im Katastrophenschutz
- Verbesserte Sicherheit zu Hause
- Nahezu unbegrenzte Zugriffsmöglichkeiten und Quellen für Nachrichten und Informationen





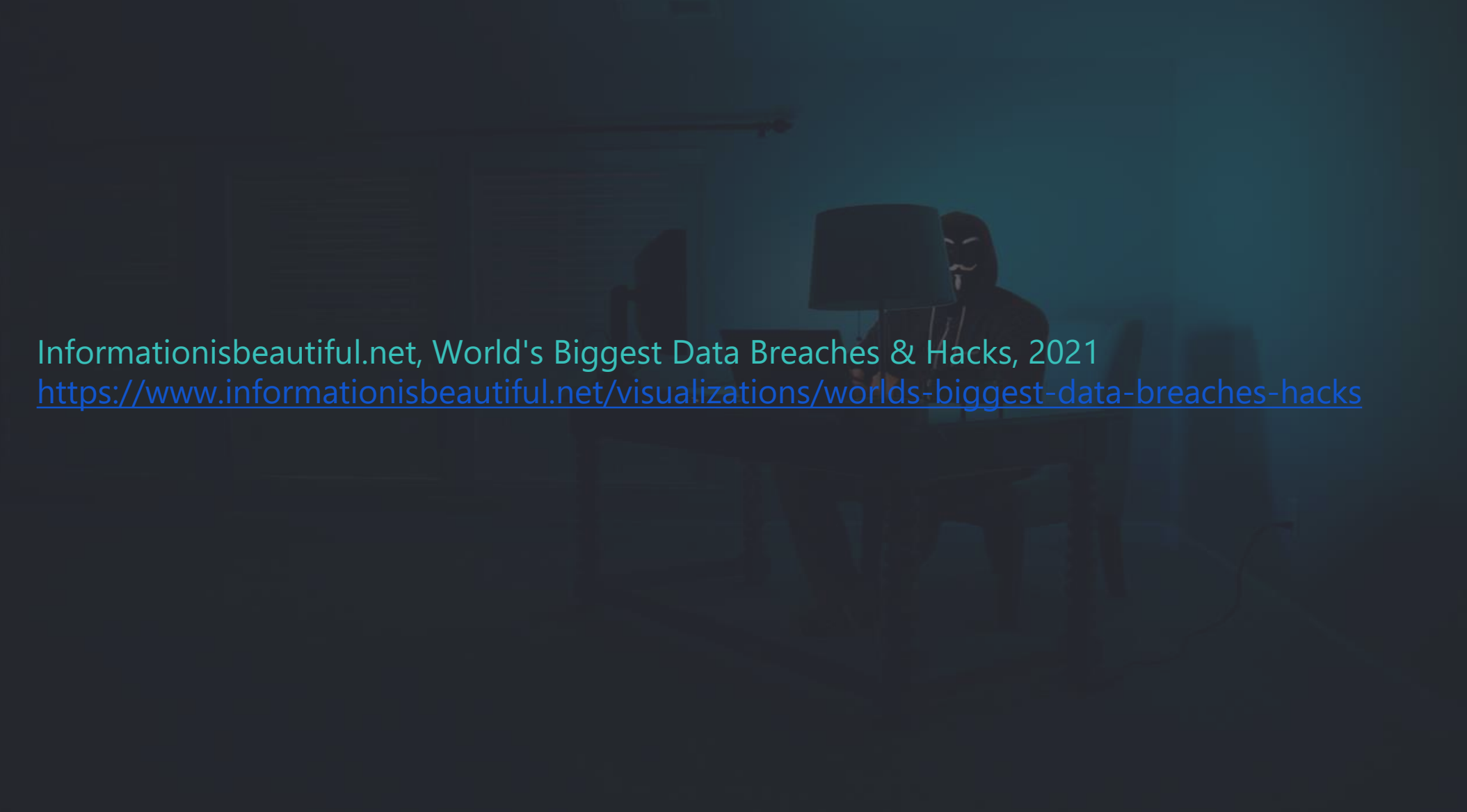
Risiken

Autor: Jonas Grasediek

Risiken – Unsichere Datenspeicherung

- Daten werden unzureichend gesichert
 - Priorität liegt auf dem Erheben, Verarbeiten und Verkaufen von Daten
 - „erstmal Daten erheben, dann schauen, ob und wie man sie gebrauchen kann“
 - Immer wieder erfolgreiche Angriffe auf große Datenspeicher
-
- LinkedIn – 2021 – 700 Mio. Nutzer betroffen
 - Facebook – 2019 – 533 Mio. Nutzer betroffen
 - Alibaba – 2019 – 1,1 Mrd. Nutzer betroffen
 - Yahoo – 2014 – 500 Mio. Nutzer betroffen
 - Adobe – 2013 – 153 Mio. Nutzer betroffen

Risiken – Unsichere Datenspeicherung

A person wearing a black hoodie and a white mask is sitting at a desk in a dimly lit room. They are looking at a laptop. A lamp is on the desk, and there are some papers and a bottle nearby. The background is dark and out of focus.

Informationisbeautiful.net, World's Biggest Data Breaches & Hacks, 2021

<https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks>

Risiken – Was sind meine Daten Wert?

Kreditkartendaten	Kanadische Karte inkl. CVV	28\$
	Israelische Karte inkl. CVV	65\$
Soziale Medien	1.000 Instagram-Follower	5\$
	Facebook-Account	65\$
Ausweisdokumente	USA-Selfie mit Ausweis	100\$
	Minnesota Führerschein	20\$
E-Mail / Datenbank Dumps	2,4 Mio. kanadische E-Mails	10\$
	Wählerdatenbank USA (je nach Staat)	100\$

Risiken – Datenqualität

- Falsche Daten
- Ungenauigkeit der Daten
- Mangelnde Integrität
- Manipulation
- Fehlerhafte Datenauswertung
- Echokammereffekt
- Intransparenz gegenüber dem Benutzer

Risiken – Ungenauigkeit der Daten

- Eine Vielzahl der Daten sind ungenau und unterschiedlich zu interpretieren
- Viele Vorhersagen scheitern daran
- Beispiel **Google-Fluence**
 - Projekt zur Vorhersage von Grippeausbrüchen
 - Basierend auf den Suchbegriffen von Nutzern
 - Symptome, Medikamente und Begrifflichkeiten mit Zusammenhang zu Gesundheit / Grippe
 - Projekt wurde 2013 eingestellt da Vorhersagen in 100 von 108 Wochen falsch waren

Risiken – Manipulation

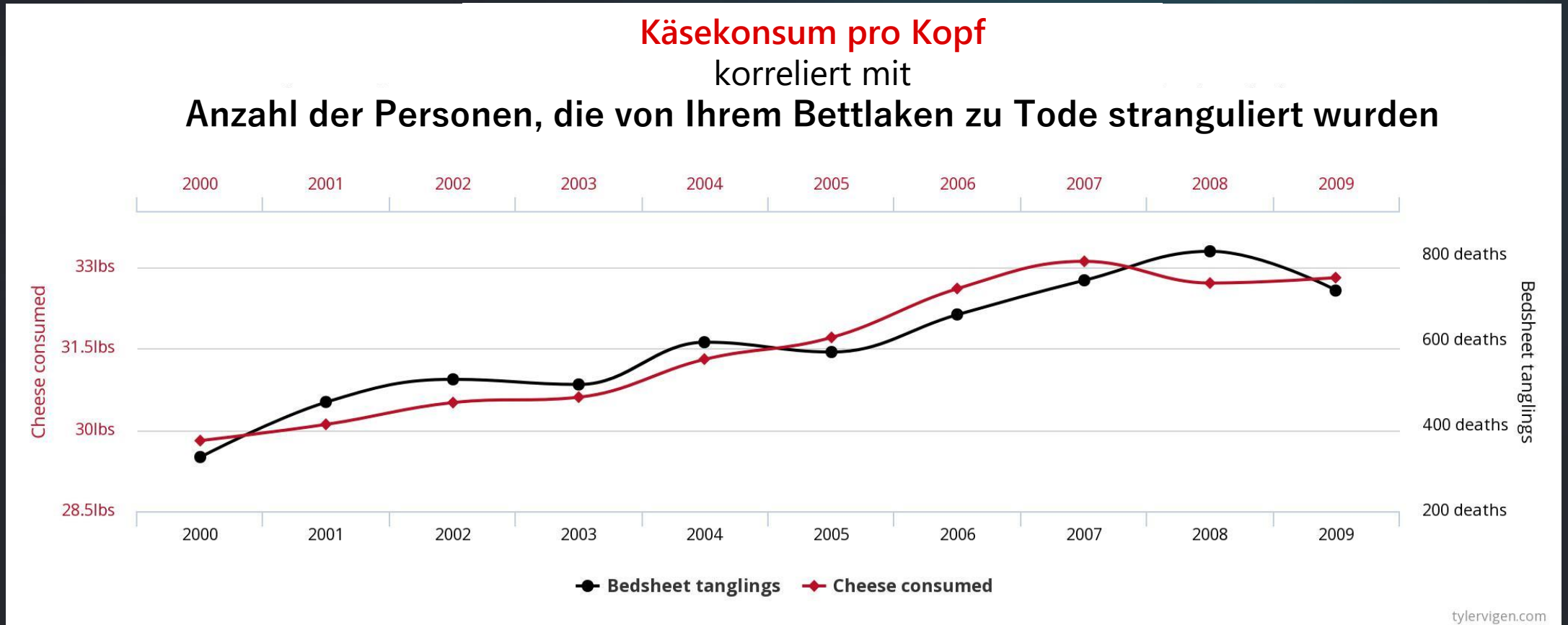
- Simon Weckert (Künstler) machte einen Spaziergang durch Berlin
- Dabei zog er einen Bollerwagen mit 99 ausrangierten Handys hinter sich her
- Bei allen Handys waren die Standortdaten und die Navigation via Google Maps aktiviert
- Aufgrund der mäßigen Geschwindigkeit der 99 Teilnehmer errechnete Google einen Stau auf der Strecke
- Infolgedessen auch Alternativrouten für andere Verkehrsteilnehmer errechnet wurden, sodass diese die Strecke umfuhren

Risiken – Fehlerhafte Datenauswertung

- Datenauswertungen im Big-Data-Bereich sind nicht immer korrekt
- Sie können durch Faktoren wie
 - falsche Daten
 - ungenaue Daten
 - fehlende Datenintegrität
 - gezielte Manipulation
- negativ beeinflusst werden, sodass es zu einer fehlerhaften Datenauswertung kommt

Risiken – Fehlerhafte Datenauswertung

So können bestimmte Datenauswertung beispielsweise eine Korrelation aufzeigen ohne dass diese einen Sinn ergibt, wie folgendes Beispiel aufzeigt:



Risiken – staatliche Überwachung



Hessen-Data

- Identifizierung von Gefährdern
- Führt unverknüpfte Datenbestände zusammen und wertet sie aus
- Kriminalfälle, Telefonüberwachung, Social-Media-Inhalte, Kontaktdaten,
- Einsatz von Palantir-Software

https://www.polizei.hessen.de/File/polizei-hessen-750-tr_18.png

PRISM

- Seit 2007 ein nicht-öffentliches Programm zur Speicherung und Auswertung von Daten
- Telekommunikations- und Internetdaten
- Direkter Zugriff auf Server großer Firmen (z.B. Facebook, Google, Verizon, etc.)
- Kein konkreter Verdacht erforderlich sondern die Annahme der Datenrelevanz
- Ranghohe Politikerinnen und Politiker, Staatsoberhäupter, EU-Kommission, EU-Parlament, EU-Rat, Wirtschaftsunternehmen



https://commons.wikimedia.org/wiki/File:Seal_of_the_U.S._National_Security_Agency.svg

Quellen:

Tobias Singelstein, Netzpolitik.org, Hessen sucht mit Palantir-Software nach Gefährdern, 2019
<https://netzpolitik.org/2019/big-data-bei-der-polizei-hessen-sucht-mit-palantir-software-nach-gefaehrden/#spendenleiste>
Datenschutz.rlp.de, Big Data
<https://www.datenschutz.rlp.de/de/themenfelder-themen/big-data/>
Thilo Weichert, Unabhängiges Landeszentrum für Datenschutz, Prism, Big Data und der Datenschutz – bei uns und in den USA, 2019
<https://www.datenschutzzentrum.de/artikel/98-Prism,-Big-Data-und-der-Datenschutz-bei-uns-und-in-den-USA.html>

Autor: Jonas Grasediek

©Lianhao Qu, unsplash.com

Weitere Risiken

- Umkehrung von Pseudonymisierung
- Unzureichende Pseudonymisierung / Anonymisierung
- Unüberschaubare Komplexität der Daten und Speicher
- Nichterfolgte Datenlöschung
- Verzögerte Reaktion auf Sicherheitsvorfälle
- Unzureichende Sicherheitsvorkehrungen
- Unzureichende Audits der Sicherheitsvorkehrungen
- Unzureichende Protokollierung
- Datenausleitung durch Insider (Leaks, Whistleblower, etc.)
- Staatlich angeordnete Auslieferung von Daten (z. B. Freedom-Act, Cloud-Act)

Folgen – Diskriminierung per Algorithmus

- Trust-Score bei Online-Händlern, zur Bewertung der Bonität von Kunden
 - 16.000 Merkmale werden erfasst, sodass jeder Klick beim Browsen zur Anpassung des Scores beiträgt
 - Browserwechsel beim Online-Shoppen erzeugt beispielsweise einen Abzug
- Online-Screening von Air-BnB
 - KI-System, das Kunden vor der Buchung einem Screening unterzieht
 - Scant Internetseiten und soziale Netzwerke, um negative Persönlichkeitsmerkmale abzuleiten
 - Errechnet einen Vertrauenswürdigkeitsscore (soziale Bonität)
 - Punktabzüge für negative Sprache, Fakeaccounts, etc.

Folgen – Diskriminierung per Algorithmus

- Passagier-Scoring von Airlines
 - Beschwerden führen zu negativem Score
 - Wer nie bei der Kundenhotline anruft erhöht seinen Score
 - Höherer Score = kürzere Wartezeit
 - Höherer Score = höhere Chance auf ein Upgrade in die Business-Class
- Racial-Profiling von Algorithmen
 - Amazon entwickelte 2018 einen Algorithmus, der Gesichter automatisch erkennen und zuordnen sollte
 - Testlauf mit Abgeordneten des US-Kongresses wies eine hohe Fehlerrate auf
 - 28 Mitglieder wurden als vermeintlich verurteilte Straftäter identifiziert
 - Der Großteil der Fehler bei der Gesichtserkennung betraf People-Of-Color

Quellen:

Adrian Lobe, Zeit Online, Big Data - Eine gefährliche Verzauberung der Welt, 2020

<https://www.zeit.de/kultur/2020-01/big-data-algorithmen-zahlen-bedeutung-mystik>

Russell Brandom, The Verge, Amazon's facial recognition matched 28 members of Congress to criminal mugshots, 2018

<https://www.theverge.com/2018/7/26/17615634/amazon-rekognition-aclu-mug-shot-congress-facial-recognition>

Autor: Jonas Grasediek

©Clint Patterson, unsplash.com

Risiken – Social Scoring in China

- China hat ein Sozialkreditsystem eingeführt, dass
- Kopfnote für bestimmtes Verhalten erteilt
- **Pluspunkte** für „gutes“ Verhalten
 - ✓ Pflege von Familienmitgliedern
 - ✓ Blutspende
 - ✓ Bei Grün an der Ampel stehen bleiben
- **Minuspunkte** für „schlechtes“ Verhalten
 - ✗ Zu langes Spielen von Videospielen
 - ✗ Exzessives Online-Shopping
 - ✗ Posten von kritischen Social-Media-Beiträgen
- Gutes Ranking verschafft Vorteile
 - Steuererleichterungen
 - Bevorzugte Behandlung im Krankenhaus!
 - Erleichterte Ausreiseerlaubnis

Quellen:

Adrian Lobe, Zeit Online, Big Data - Eine gefährliche Verzauberung der Welt, 2020

<https://www.zeit.de/kultur/2020-01/big-data-algorithmen-zahlen-bedeutung-mystik>

Ma & Canales, China's 'social credit' system ranks citizens and punishes them with throttled internet speeds and flight bans if the Communist Party deems them untrustworthy

<https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4>

Autor: Jonas Grasediek

Weitere Folgen

- Wegfall der Anonymität im Internet
 - Überall werden jederzeit möglichst viele Daten erhoben, darunter auch personenbezogene
- Filterblasen
 - sind bereits jetzt ein Problem und nur schwer zu durchbrechen
 - Beeinflussen die eigene Identitätsentwicklung, insbesondere bei jugendlichen
- Datengetriebene Verhaltensänderung – Beeinflussung unseres Verhaltens durch Konzerne
 - Telematik und entsprechende Prämien, Social Credit Systeme
- Soziale Diskriminierung
 - Dynamic Pricing (Preise auf iPhone höher als auf Android)
 - Predictive Policing (Die Analyse von Falldaten zur Berechnung der Wahrscheinlichkeit zukünftiger Straftaten zur Steuerung des Einsatzes von Polizeikräften)
 - Versicherungen errechnen Beiträge auf Basis Ihrer Ernährung
- Beeinflussung der politischen Meinungsbildung
 - Einschränkung durch Echokammer
 - Social Bots
 - Filterblasen
- Konformitätsdruck
 - Druck auf ein Individuum zur Anpassung seines Verhaltens oder Denkens, um mit dem Gruppenstandard übereinzustimmen.

Big Data vs. Demokratie

- 2016 Haustürwahlkampf von Trump
- Wahlhelfer nutzen eine App, die möglichst genaue Informationen über die besuchten potentiellen Wählerinnen und Wähler enthielten
- Informationen basierten auf 32 verschiedenen Persönlichkeitstypen, die zuvor über soziale Medien erstellt wurden
- Davon abgeleitet waren entsprechende Gesprächsleitfäden, welche die Wahlhelfer nutzten

Cambridge Analytica

- Britisches Datenanalyse-Unternehmen
- Missbrauch einer Drittanbieter-App um an Benutzerprofile zu kommen
- 87 Millionen Datensätze ergattert
- Likes, Aktivitäten, Kontaktdaten, Identität, etc.
- Vordergründig zu wissenschaftlichen Zwecken von GSR verwendet
- Weiterleitung der Daten an Cambridge Analytica <-> ~ 800.000\$ Finanzierung
- Erstellung psychologischer Profile <-> Microtargeting
- Botschaften der Trump-Kampagne auf möglichst kleine Zielgruppen zuzuspitzen
- Komplettes Informationsökosystem (Websites & Blogs), die nicht als Teil der Trump-Kampagne erkennbar waren <-> vermeintlich unabhängige Informationen
- Wähler erhielten immer genau die Forderungen und Versprechen, die laut psychologischem Profil die meiste Wirkung erzielen
- Hilfe bei der Verarbeitung der Daten von Überwachungsfirma Platinir
- Deren Gründer ist Mitglied im Aufsichtsrat von Facebook und Trump-Unterstützer

Cambridge Analytica

- Nachweislich weitere Beteiligung an Wahlkämpfen in der Karibik und Afrika
- Vermutete Beteiligung an der Pro-Brexit-Kampagne – keine eindeutigen Belege
- 2018 – Beschlagnahmung von IT-Geräten, 700TB Daten und 300.000 Dokumenten durch die britische Datenschutzbehörde
- 2018 – Insolvenz
- Bis heute kein eindeutiger Nachweis über die Wirksamkeit von Microtargeting

Quellen:

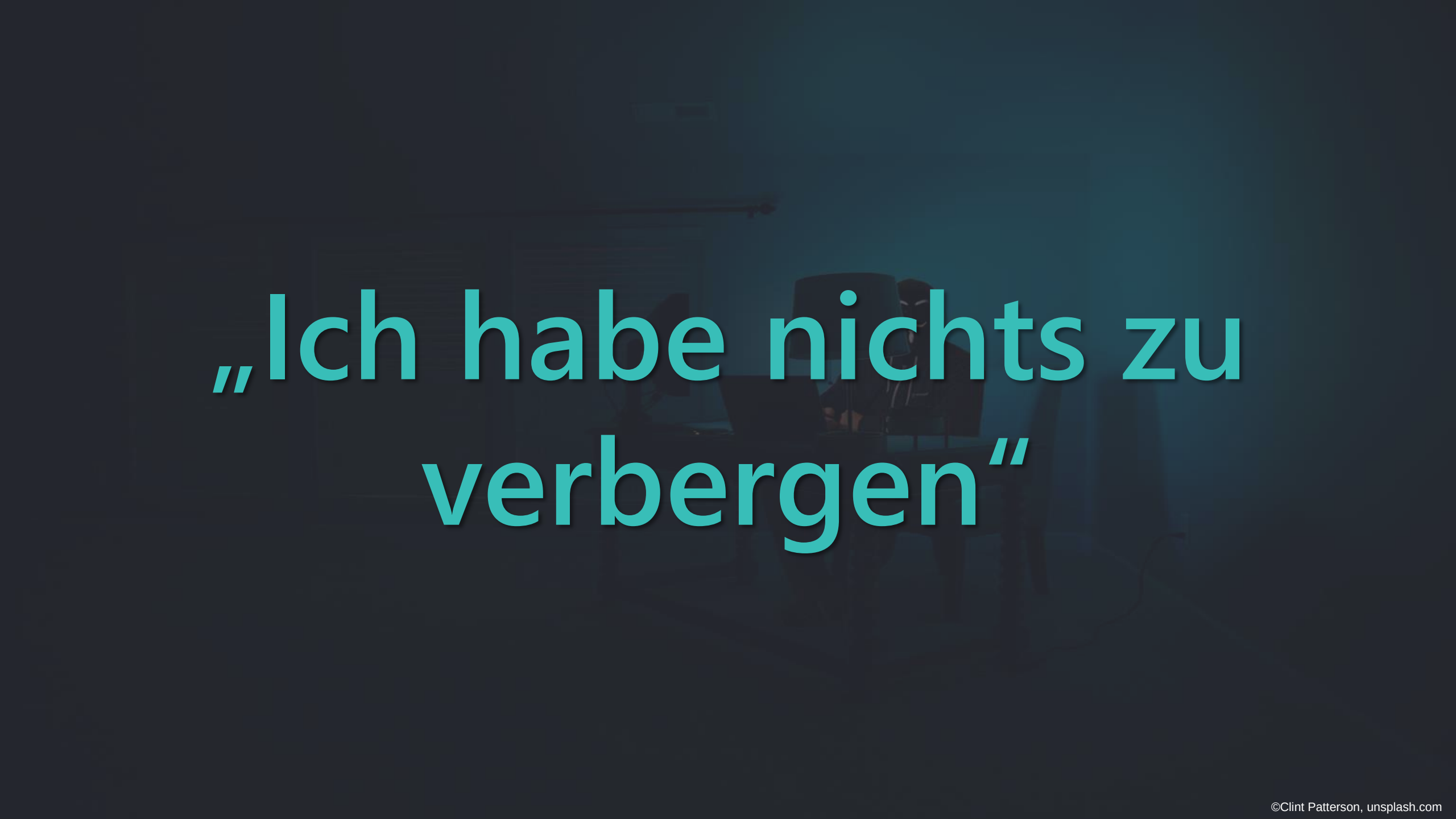
Dachwitz, Thomas & Rebiger, Netzpolitik.org, Was wir über den Skandal um Facebook und Cambridge Analytica wissen, 2018
<https://netzpolitik.org/2018/cambridge-analytica-was-wir-ueber-das-groesste-datenleck-in-der-geschichte-von-facebook-wissen/>
Carole Cadwalladr, The Guardian, The great British Brexit robbery: how our democracy was hijacked, 2017
<https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy>

Autor: Jonas Grasediek

©wenhilario, unsplash.com

Cambridge Analytica – Facebook

- 87 Millionen Datensätze ohne das Wissen der Betroffenen entwendet
- Facebook wusste seit 2015 bescheid
- Aufforderung zur Löschung der Daten erfolgte, wurde aber nicht verifiziert
- Keine Information an Betroffene über Datenabfluss
- Facebook hat lange Zeit mittels Design erschwert, zu kontrollieren wer welche Daten erhält
- Facebook bietet zahlreiche Möglichkeiten Werbung gezielt und auf Basis verschiedener Datenprofile an möglichst genau passende Zielgruppen zu richten
→ Lookalike-Audiences, Custom-Audiences, Audience-Network
- 5 Mrd. Dollar Strafe für Facebook



„Ich habe nichts zu
verbergen“



Zu argumentieren, dass sich Menschen nicht um das Recht auf Privatsphäre kümmern, weil Sie nichts zu verbergen haben, ist nicht anders als zu sagen, dass Sie sich nicht um die freie Meinungsäußerung kümmern, weil Sie nichts zu sagen haben.

-Edward Snowden