Wechselwirkungen von Datenschutz, Systemsicherheit und künstlicher Intelligenz

Prof. Dr. Hannes Federrath
Sicherheit in verteilten Systemen (SVS)
http://svs.informatik.uni-hamburg.de

Algorithmic Decision Making
Predictive Policing
Profiling

Big Data
Data Analytics
Künstliche Intelligenz

Zuverlässigkeit
Verstehbarkeit

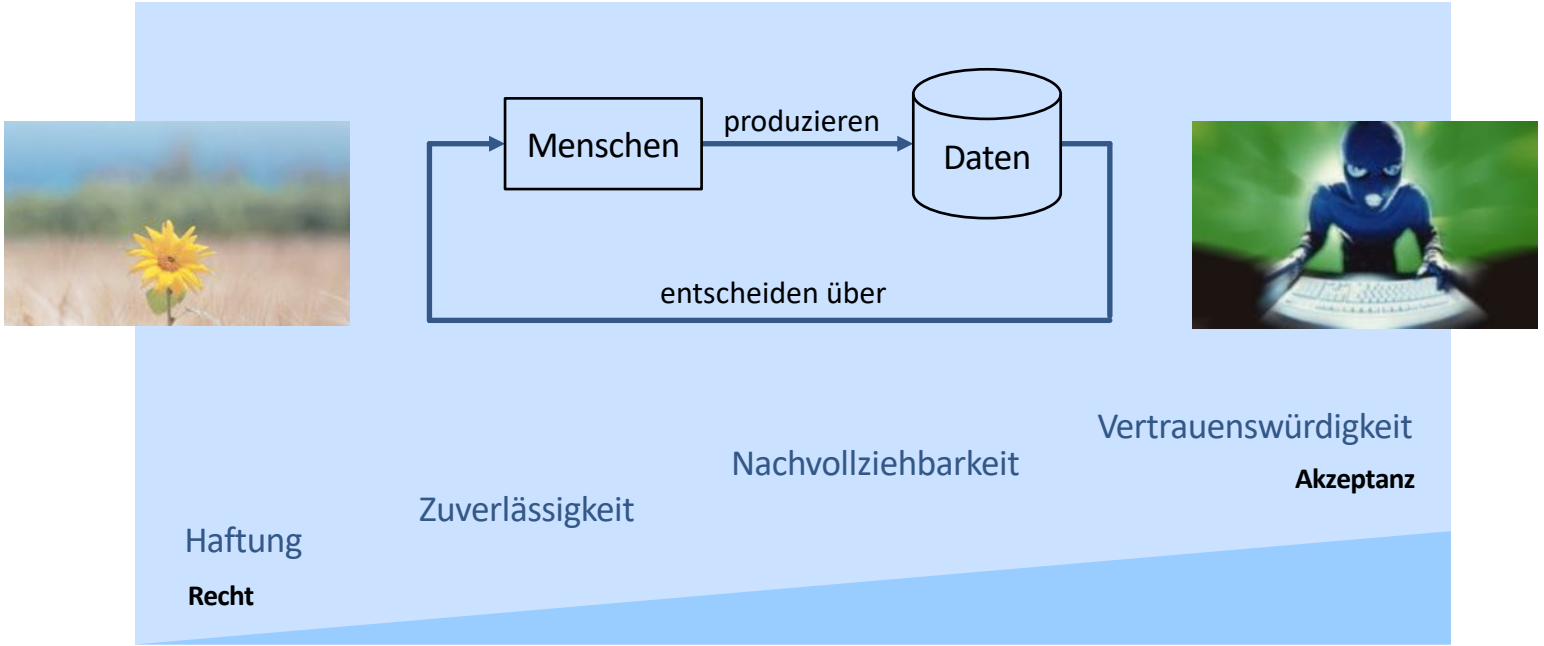
Datenschutz

IT-Sicherheit

Privacy by Design
Security by Design

Künstliche Intelligenz als Blackbox

- Risikoawareness: Spannungsfeld zwischen Haftung und Vertrauenswürdigkeit



Autonome Fahrzeuge

- Haftungsregeln bei Technikversagen
  - juristische Fragen
  - technische Fragen
  - ethische Fragen

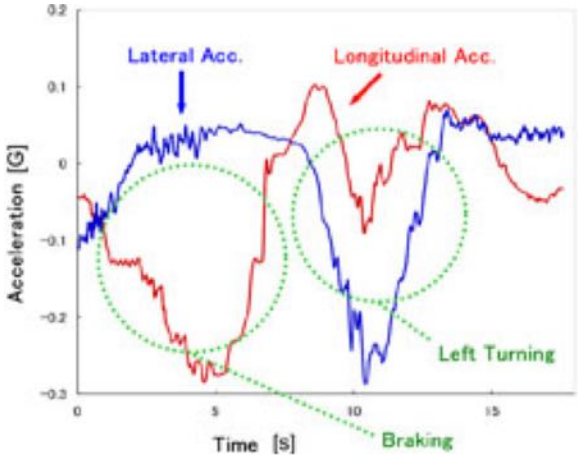
    - ohne Helm: höhere Chance, beim Aufprall zu sterben
    - mit Helm: höhere Chance zu überleben

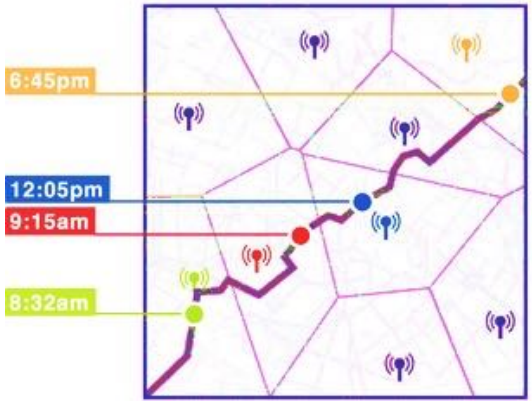
http://www.theguardian.com/environment/green-living-blog/2010/apr/07/david-cameron-cycles-without-helmet

Kontinuierliches Sammeln und Übermitteln von Daten und (ggf. anonymisierte) zentrale Auswertung zur Systemverbesserung, Beweissicherung etc.

Kontinuierliches Sammeln und Übermitteln von Daten

http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html
http://www.npa.go.jp/nrips/en/traffic/section3.html





Klärung der Haftung bei Unfällen:

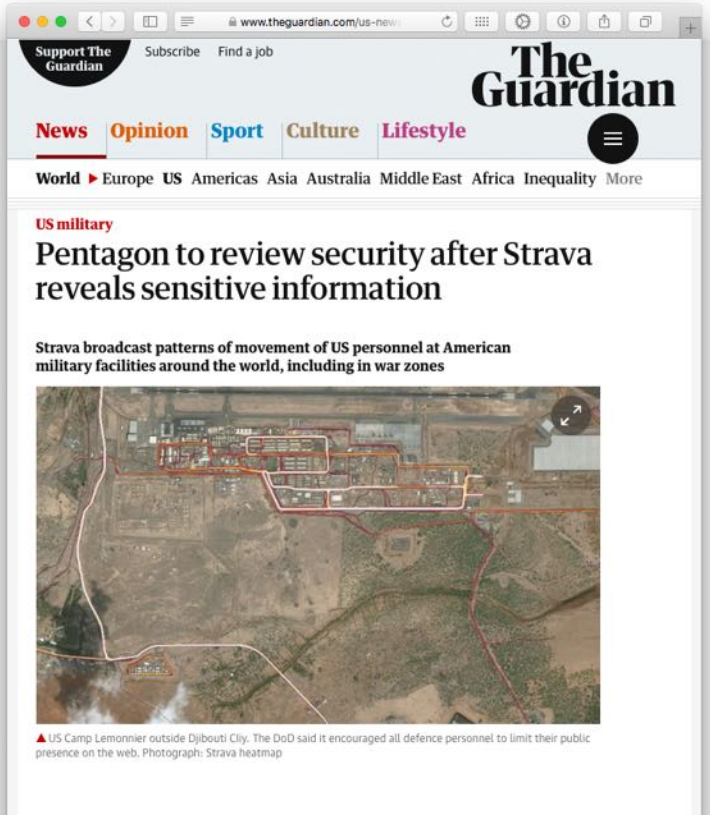
- Rekonstruktion des Fahrverhaltens anhand aufgezeichneter Daten
- Vorteile für den Hersteller, der festlegt, welche Daten gespeichert werden

Gefahr für die Privatsphäre:

- Identifizierung des Fahrers auch anhand anonymisierter Bewegungsdaten möglich; u.U. reichen bereits 4 Ortsangaben aus
- Verräterische Daten wecken Begehrlichkeiten

Soziale Netze und Datenschutz – Der Fall »Strava Heatmap«

- Fitness-Tracker Website veröffentlicht beliebte Laufstrecken
- Soldaten des US-Militärs offiziell ausgestattet mit Fitness-Tracker
- Öffentliche »Heatmap« enthüllt ungewollt geheime US-Bases im Ausland

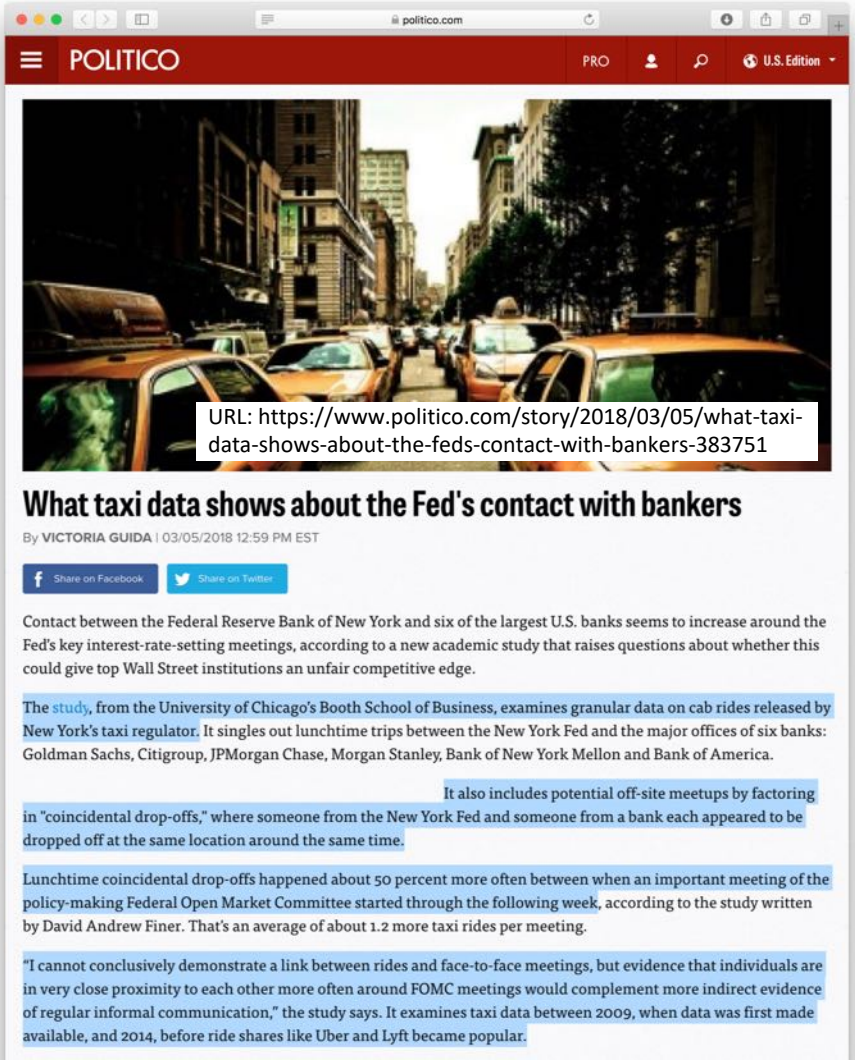


Sources:
https://twitter.com/Nrg8000/status/957318498102865920
https://www.theguardian.com/us-news/2018/jan/29/pentagon-strava-fitness-security-us-military

Anonymisierte Daten…

…können Geheimes verraten.

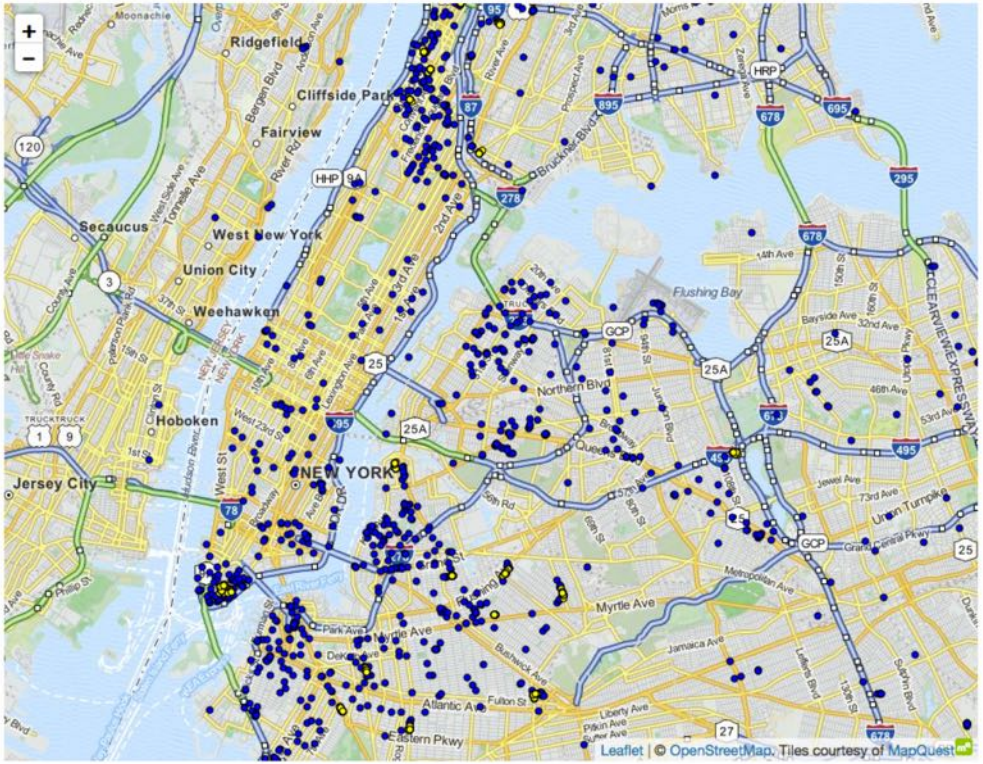
David Andrew Finer: What Insights Do Taxi Rides Offer into Federal Reserve Leakage? Working Paper, Booth School of Business, University of Chicago, March 2018. https://research.chicagobooth.edu/-/media/research/stigler/pdfs/workingpapers/18whatinsightsdotaxiridesofferintofederalreserveleakage.pdf



Pseudonymisierte Daten…

…können Persönlichkeitsrechte verletzen.

20 GByte of pseudonymisierter Daten von 170 Mio. Taxifahrten der New Yorker Taxi-gesellschaft



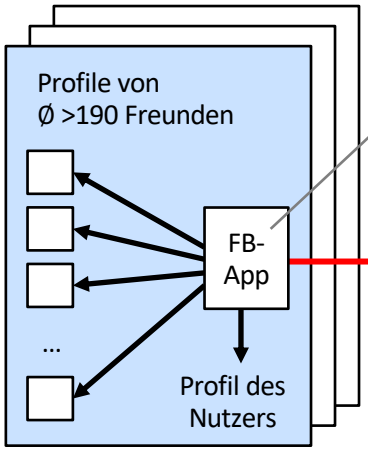
Drop-off locations for trips starting at Larry Flynt's Hustler Club between midnight and 6 am during 2013.

Source: http://content.research.neustar.biz/blog/differential-privacy/stripRaw.html

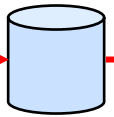
Der Fall Cambridge Analytica

nach: ct 2018, Heft 8, S. 20

>270.000 Facebook-Nutzer

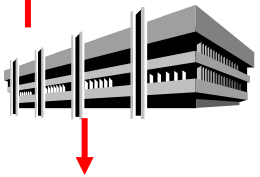


Facebook-App »thisisyourdigitallife« des Psychologen Alexander Kogan greift (mit Einwilligung der Facebook-Nutzer auf Basis der damaligen Privacy-Einstellungen) »zu wiss. Zwecken« auf Profile und Daten von Freunden zu



>50 Mio. Facebook-Profile

Unerlaubte Datenweitergabe



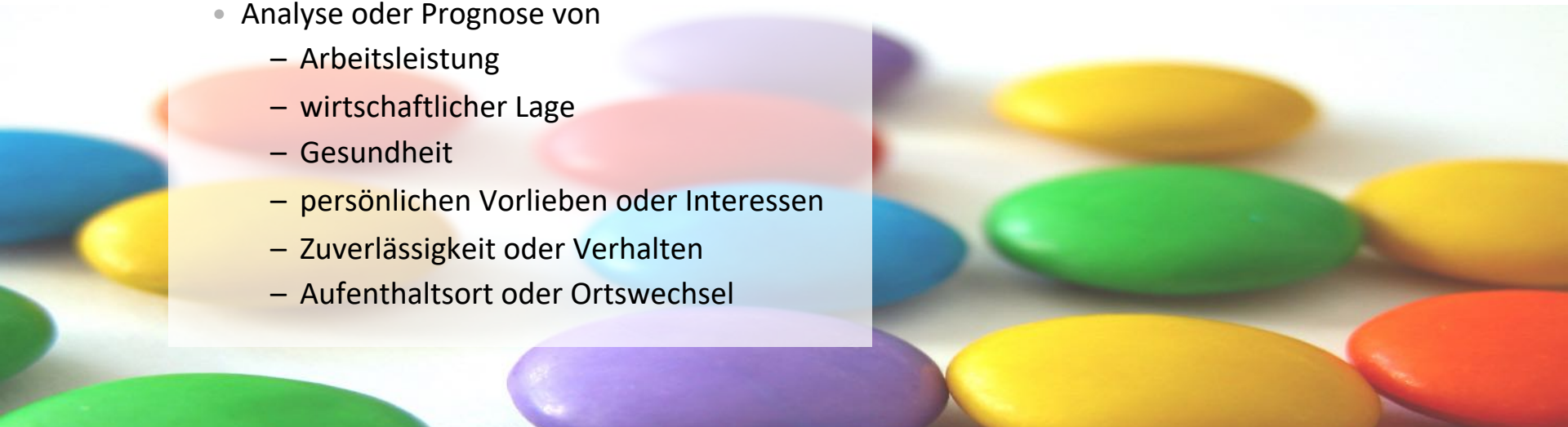
Auswertung nach Persönlichkeitsprofilen gem. OCEAN-Modell:

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

Cambridge Analytica: Weiterverwendung zur gezielten Anzeige von (Wahl)-Werbung: Brexit, Trump, …

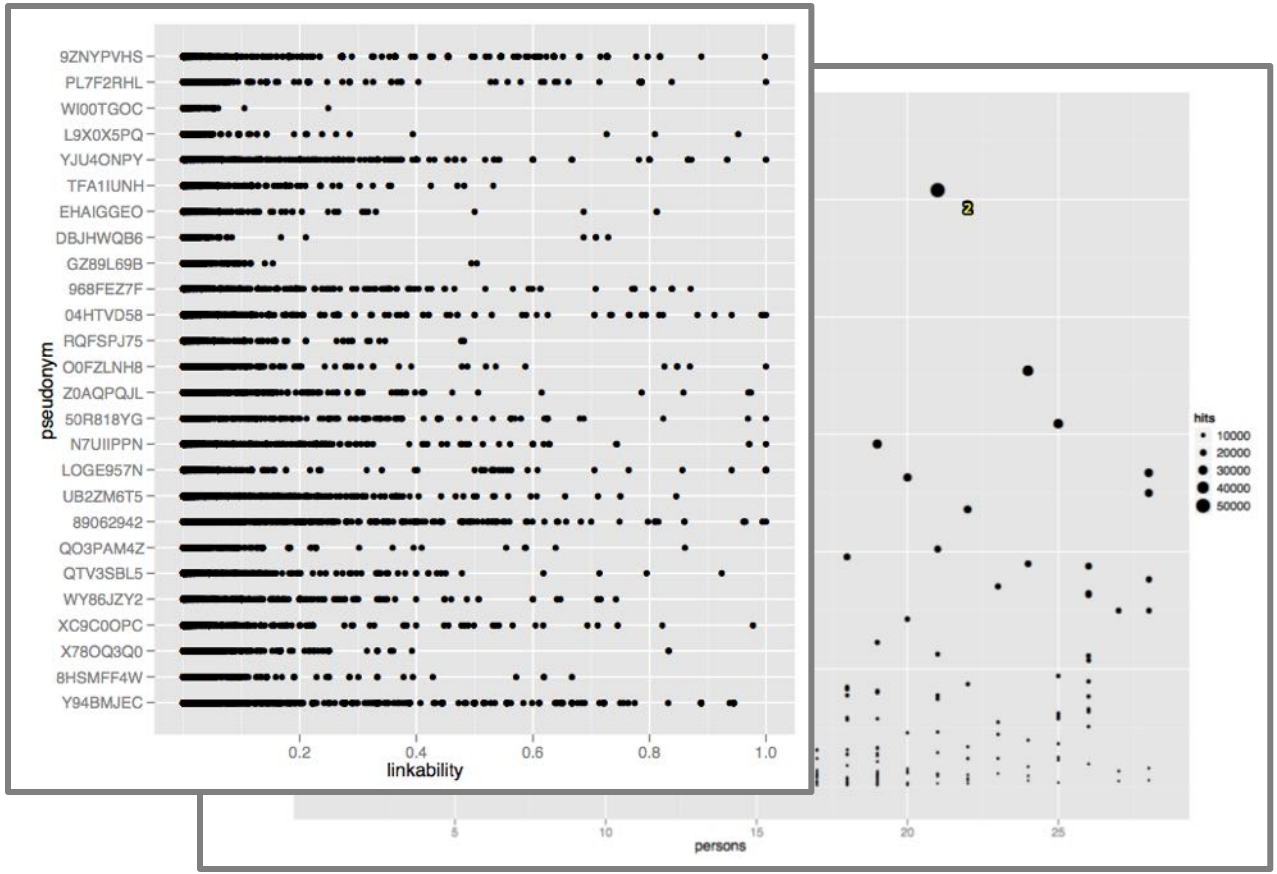
Automatisierte Entscheidungen im Einzelfall einschl. Profiling

- 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

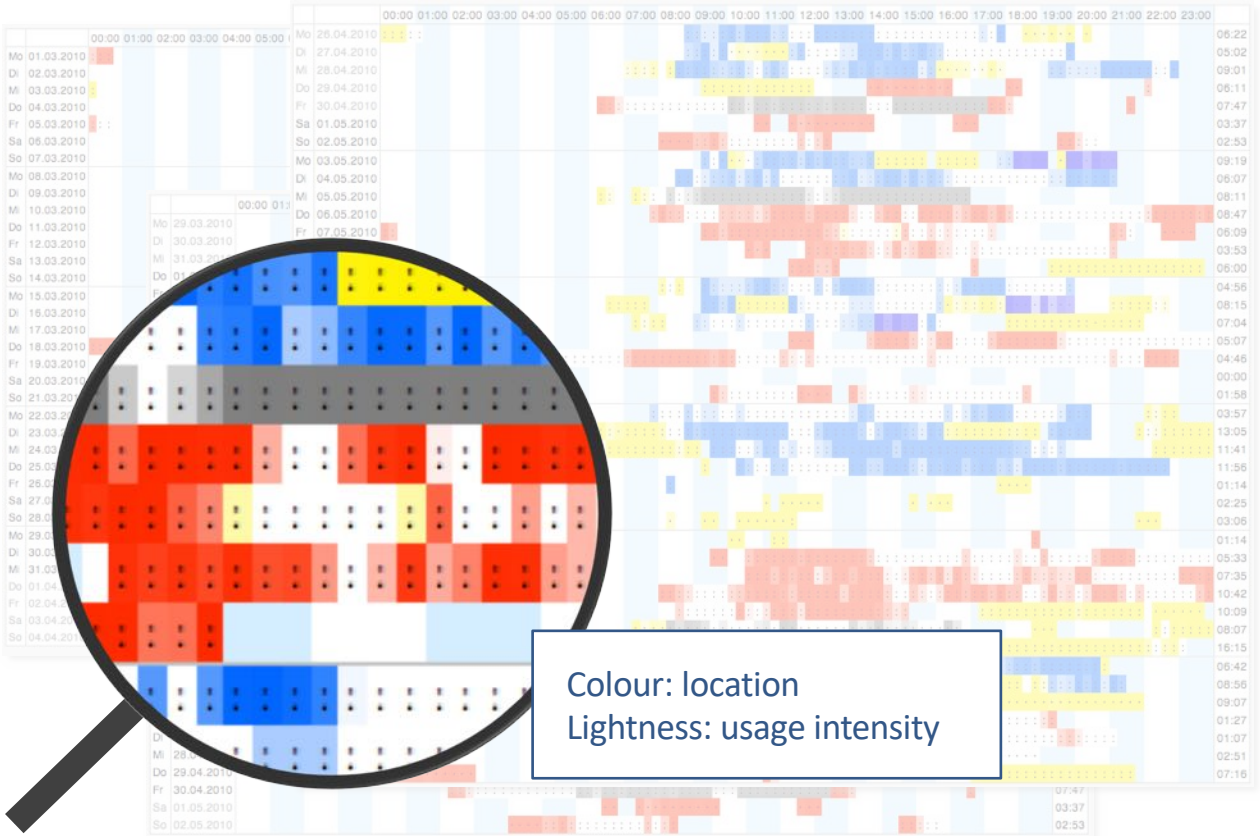


Datenanalyse des Surfverhaltens

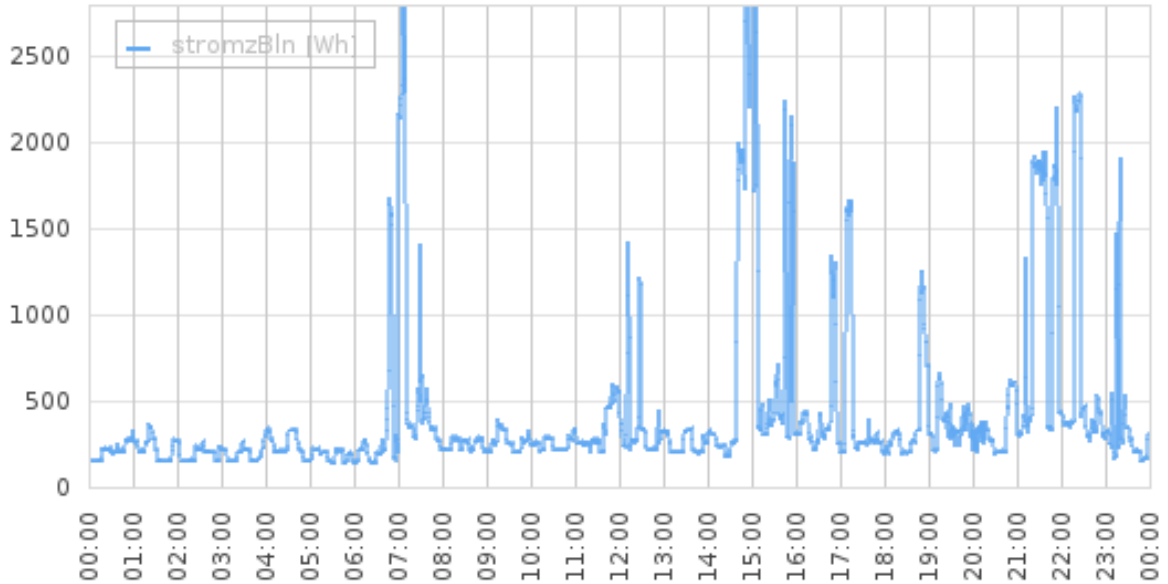
Gerber, 2009

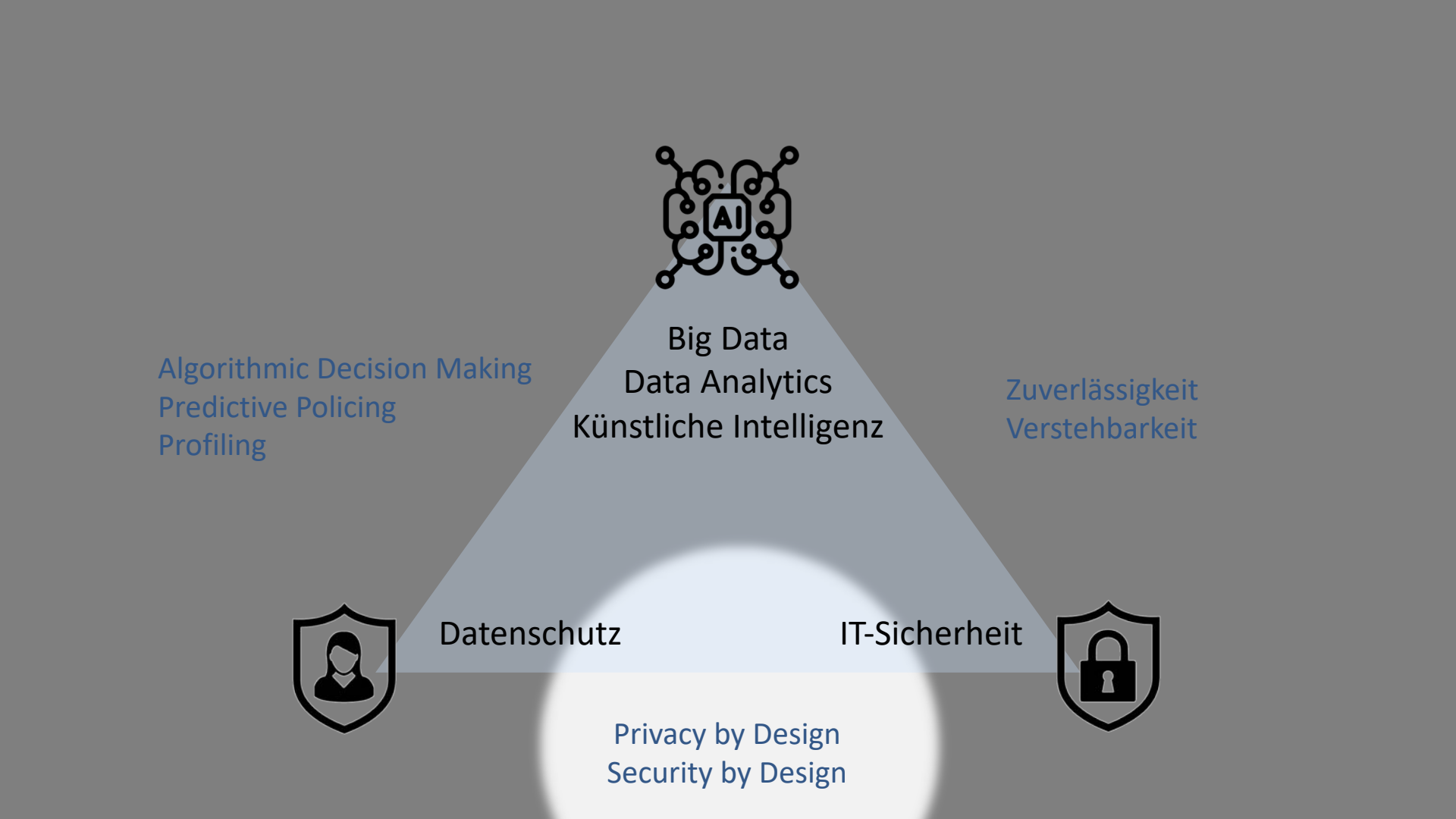


Usage timelines and ip-geo-tagging



Stromverbrauch verrät Infos über persönliche Lebensverhältnisse

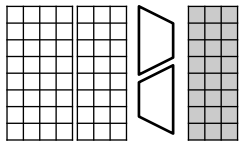



Big Data
Data Analytics
Künstliche Intelligenz
Algorithmic Decision Making
Predictive Policing
Profiling
Zuverlässigkeit
Verstehbarkeit
Datenschutz
IT-Sicherheit
Privacy by Design
Security by Design

Privacy by Design

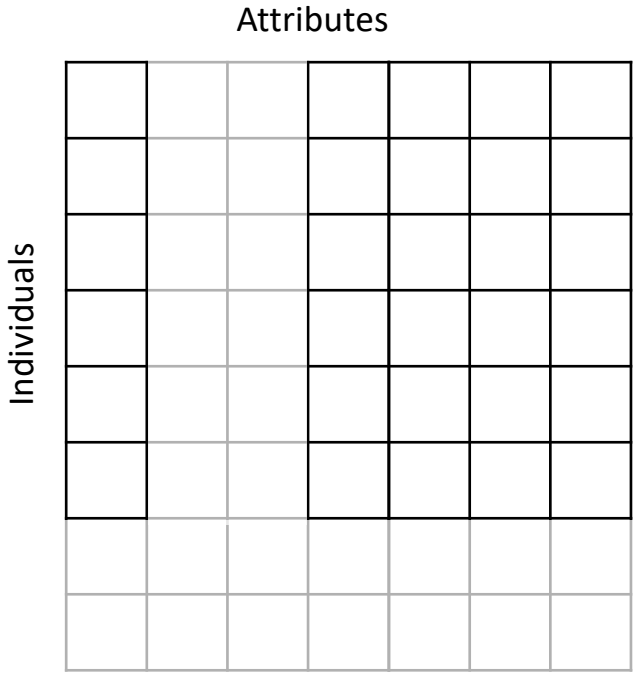
- Stand der Technik
- Kosten
  - ROSI
- Art, Umfang, Umstände, Zwecke
  - Verbot mit Erlaubnisvorbehalt
  - Einwilligung
  - Erforderlichkeit
  - Direkterhebung
- Privacy by Default
  - Beschränkung auf Verarbeitungszweck
  - Beschränkung des Umfangs
  - Beschränkung auf Speicherfristen
  - Beschränkung der Zugänglichkeit
- Risiko-management
- **Technisch-organisatorische Maßnahmen**
  - Datenschutz-grundsätze
    - Transparenz
      - Policy
      - Bereitschaft und gute Dokumentation
      - Big Data
      - Algorithmen
    - Zweckbindung
    - **Daten-sparsamkeit**
      - Anonymität
      - Unbeobachtbarkeit
      - Minimise
      - Separate
      - Aggregate
      - Perturbate
      - Hide
    - Datensicherheit
      - Pseudonymisierung
      - Verschlüsselung
      - Vertraulichkeit
      - Integrität
      - Verfügbarkeit
      - Belastbarkeit
      - Wiederherstellbarkeit
      - Notfallmanagement
      - Überprüfbarkeit
      - Bewertung
      - Evaluierung
  - Betroffenen-rechte
    - Transparenz
    - Benachrichtigung
    - Auskunft
    - Berichtigung
    - Löschung
    - Widerspruchsrecht
    - Recht auf Vergessenwerden
    - Recht auf Einschränkung der Verarbeitung
    - Recht auf Datenübertragbarkeit

Enforce
- Inform
- Control
- Demonstrate

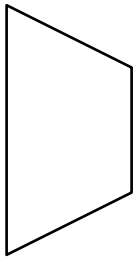


Privacy design strategies

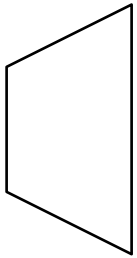
in Anlehnung an Hoepmann, 2013



minimise

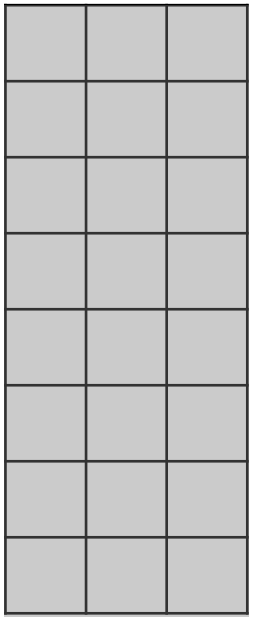


aggregate



perturbate

separate



hide

Privacy design strategies

in Anlehnung an Hoepmann, 2013

- Technisch
  - Minimise: Nur notwendige Daten speichern und verarbeiten
  - Separate: Daten verteilt verarbeiten und speichern
  - Aggregate: Daten auf das notwendige Maß zusammenfassen
  - Perturbate: Daten durch zufällige Störungen ungenau machen
  - Hide: Daten nicht in offener Form speichern

- Organisatorisch
  - Enforce: Durchsetzung einer Datenschutz-Policy (access control)
  - Inform: Betroffene über Datenverwendung informieren (P3P)
  - Control: Eingriffsmöglichkeit der Betroffenen (informed consent)
  - Demonstrate: Überprüfbarkeit (privacy management, logging)

Neubestimmung des Begriffs personenbezogene Daten

nach: Hoffmann-Riem 2018, S.55

- Personenbezogene Daten sind auch Daten, die als Ergebnis einer Big-Data-Analyse entstehen.
  - allgemein und ohne Herleitung aus Daten speziell der konkret betroffenen Person
  - Beispiele: Person wohnt in einem bestimmten Stadtteil; daraus Ableitung von Finanzkraft, Herkunft, sexueller Orientierung, Gesundheit
- Personenbezogene Daten sind auch Daten, deren Personenbezug durch Anonymisierung entfällt.
  - Möglichkeiten der Deanonymisierung und Ableitung von Eigenschaften dürften nicht unterschätzt werden
  - Beispiele: New York Taxi Data Analytics, Strava Heatmap
- ebenso kritisch pseudonymisierte, aggregierte, perturbierte, verschlüsselte Daten betrachten

Datenschutz-Folgenabschätzung

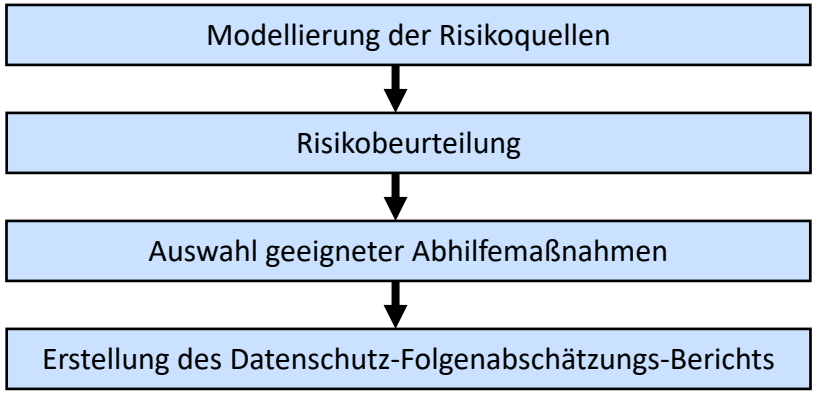
- Auszug aus Art. 35 DSGVO:

(1) Hat eine Form der Verarbeitung, insbesondere bei Verwendung neuer Technologien, aufgrund der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge, so führt der Verantwortliche vorab eine Abschätzung der Folgen der vorgesehenen Verarbeitungsvorgänge für den Schutz personenbezogener Daten durch. Für die Untersuchung mehrerer ähnlicher Verarbeitungsvorgänge mit ähnlich hohen Risiken kann eine einzige Abschätzung vorgenommen werden.

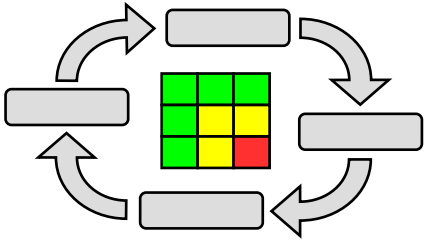
- Beispiele:
  - Scoring
  - Videoüberwachung
  - medizinische Daten
  - KI-Technologien

Methodik einer Datenschutz-Folgenabschätzung

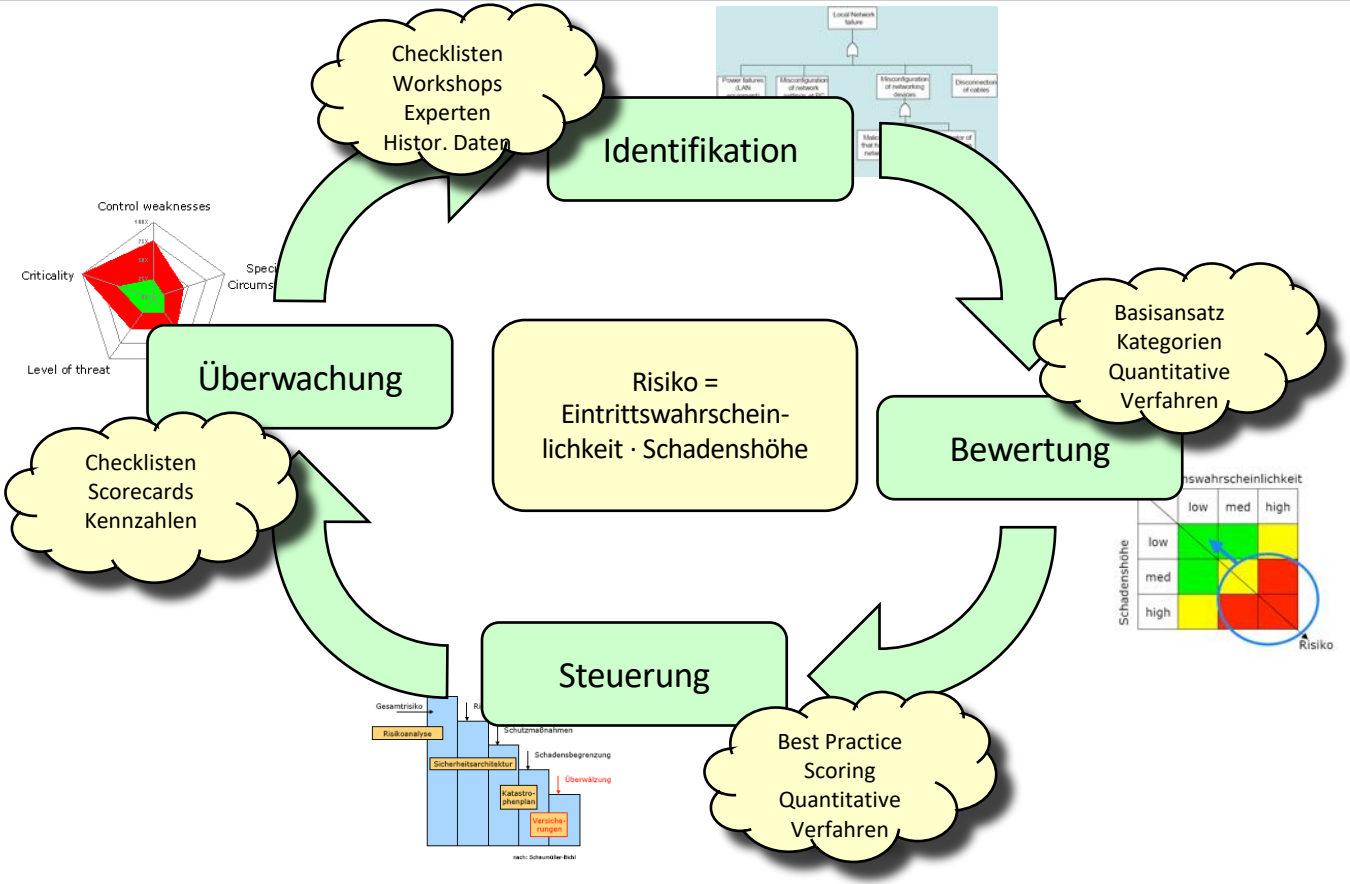
- Vorgehen gemäß Kurzpapier Nr. 5 »Datenschutz-Folgenabschätzung nach Art. 35 DSGVO« der Datenschutzkonferenz (DSK)



- marginale Unterschiede zum Risikomanagement aus der IT-Sicherheit
  - entweder Risikomanagementkreislauf
  - oder BSI-Standard zur Risikoanalyse verwenden



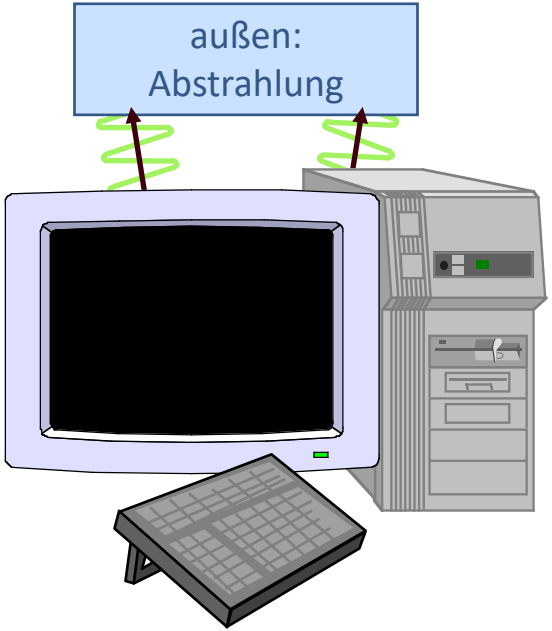
Risikomanagement Kreislauf



Privacy by Design

- Stand der Technik
- Kosten
  - ROSI
- Art, Umfang, Umstände, Zwecke
- Privacy by Default
  - Verbot mit Erlaubnisvorbehalt
    - Einwilligung
    - Erforderlichkeit
    - Direkterhebung
  - Beschränkung auf Verarbeitungszweck
    - Beschränkung des Umfangs
    - Beschränkung auf Speicherfristen
    - Beschränkung der Zugänglichkeit
- Risiko-management
- Technisch-organisatorische Maßnahmen
  - Datenschutz-grundsätze
    - Transparenz
      - Policy
      - Big Data
      - Bereitschaft und gute Dokumentation
      - Algorithmen
        - Enforce
        - Inform
        - Control
        - Demonstrate
    - Zweckbindung
    - Daten-sparsamkeit
      - Anonymität
      - Unbeobachtbarkeit
      - Minimise
      - Separate
      - Aggregate
      - Perturbate
      - Hide
    - Datensicherheit
      - Pseudonymisierung
      - Verschlüsselung
      - Vertraulichkeit
      - Integrität
      - Verfügbarkeit
      - Belastbarkeit
      - Wiederherstellbarkeit
      - Notfallmanagement
      - Überprüfbarkeit
      - Bewertung
      - Evaluierung
  - Betroffenen-rechte
    - Transparenz
    - Benachrichtigung
    - Auskunft
    - Berichtigung
    - Löschung
    - Widerspruchsrecht
    - Recht auf Vergessenwerden
    - Recht auf Einschränkung der Verarbeitung
    - Recht auf Datenübertragbarkeit
    - Bereitschaft

Abschirmen



… und Vibrationen!

(sp)iPhone: Decoding Vibrations From Nearby Keyboards Using Mobile Phone Accelerometers

Philip Marquardt*
MIT Lincoln Laboratory
244 Wood Street, Lexington, MA USA
philip.marquardt@ll.mit.edu

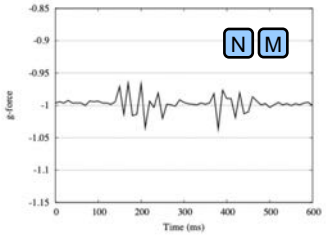
Arunabh Verma, Henry Carter and Patrick Traynor
Georgia Institute of Technology
{arunabh.verma@, carterh@, traynor@cc.}gatech.edu

ABSTRACT

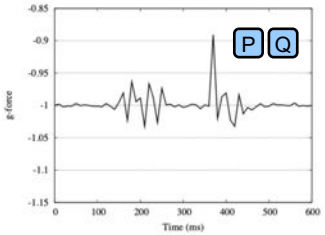
Mobile phones are increasingly equipped with a range of highly responsive sensors. From cameras and GPS receivers to three-axis accelerometers, applications running on these devices are able to experience rich interactions with their environment. Unfortunately, some applications may be able to use such sensors to monitor their surroundings in unintended ways. In this paper, we demonstrate that an application with access to accelerometer readings on a modern mobile phone can use such information to recover text entered on a nearby keyboard. Note that unlike previous emanation recovery papers, the accelerometers on such devices sample at near the Nyquist rate, making previous techniques unworkable.

to web browsers, a progressively sophisticated set of sensors are enabling these devices to more actively interface with the world around them. From gestures captured by accelerometers for games to augmented reality applications displaying metadata tags on video from the camera in real-time, mobile phones are becoming adept at capturing and harnessing rich features and data from their surroundings.

Unfortunately, the array of sensors in these devices can also be used in unintended ways. As many have suggested, malware could potentially gain access to a mobile phone's camera to take photos or video of the owner and their surroundings [16, 37]. In cases where the camera may not be pointed at an interesting target, a malicious

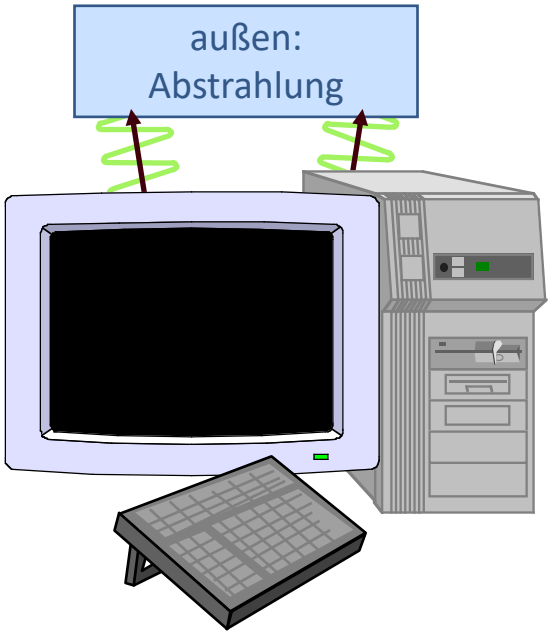


Character pair "nm"



Character pair "pq"

Rechner



… auch Tastaturklappern?

Attacks by keyboard acoustic emanations

- 10-minute sound recording of a user typing English text using a keyboard
  - recovers <= 96% of typed chars
  - makes use of machine learning
  - no need for a labeled training
- can even recognize random text
  - < 20 attempts for 90% recognition of 5-character random passwords
  - < 75 attempts for 80% recognition of 10-character random passwords

Source: Li Zhuang, Feng Zhou, J. D. Tygar: Keyboard Acoustic Emanations Revisited. Proc. 12th ACM Conference on Computer and Communications Security, November 2005, pp. 373-382.

Big Data
Data Analytics
Künstliche Intelligenz

Algorithmic Decision Making
Predictive Policing
Profiling

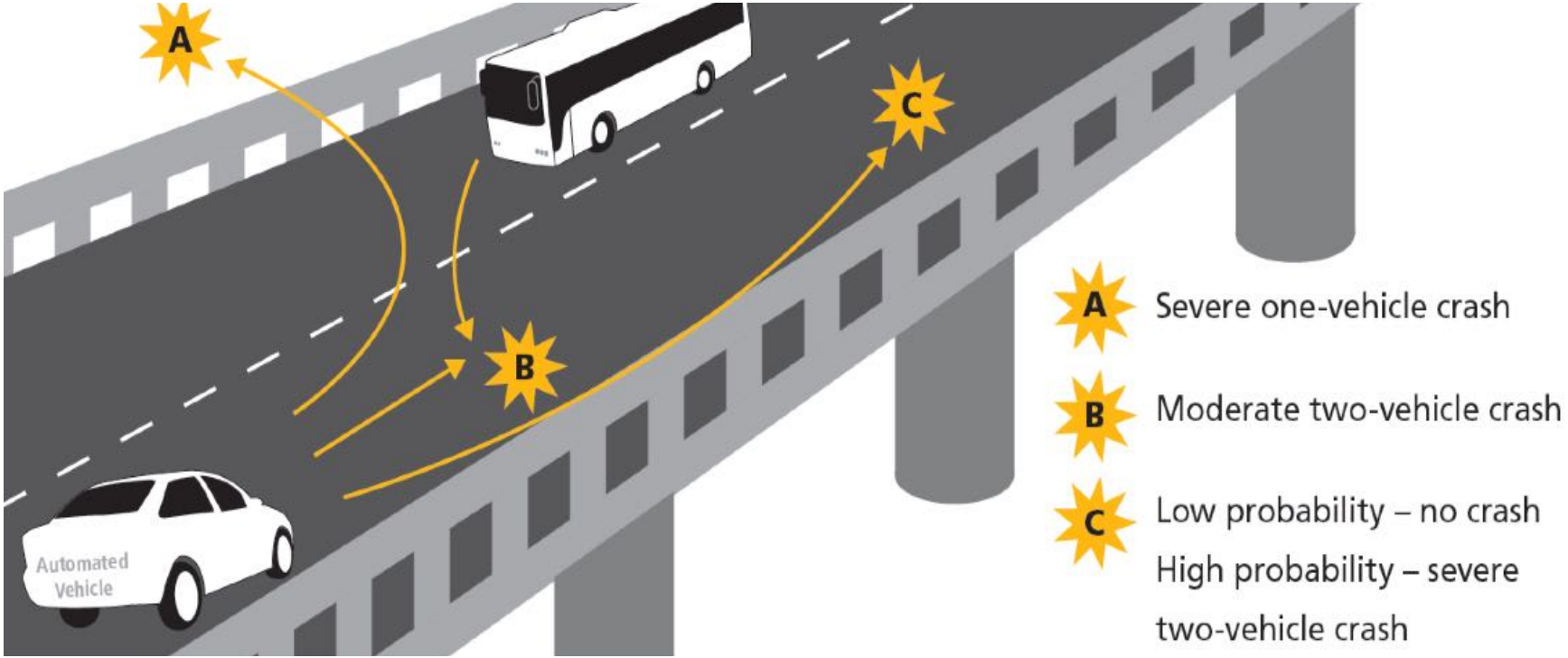
Zuverlässigkeit
Verstehbarkeit

Datenschutz

IT-Sicherheit

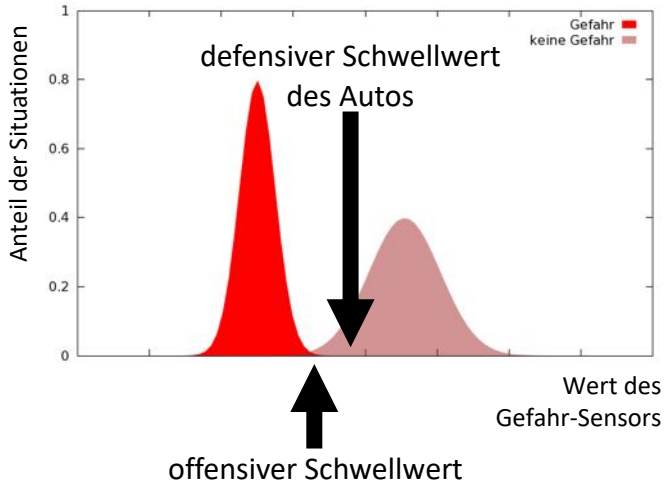
Privacy by Design
Security by Design

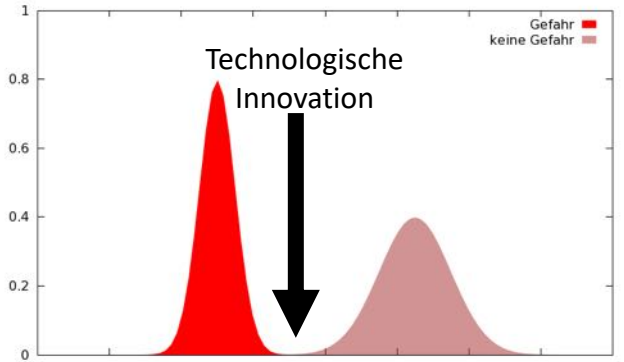
http://people.virginia.edu/~njg2q/ethics.pdf



Erkennung von Gefahren ist eine Frage von Wahrscheinlichkeiten

REALITÄT	ERKENNUNG: Gefahr	ERKENNUNG: »normal«
Gefahr	richtig positiv	falsch negativ
»normal«	falsch positiv	richtig negativ

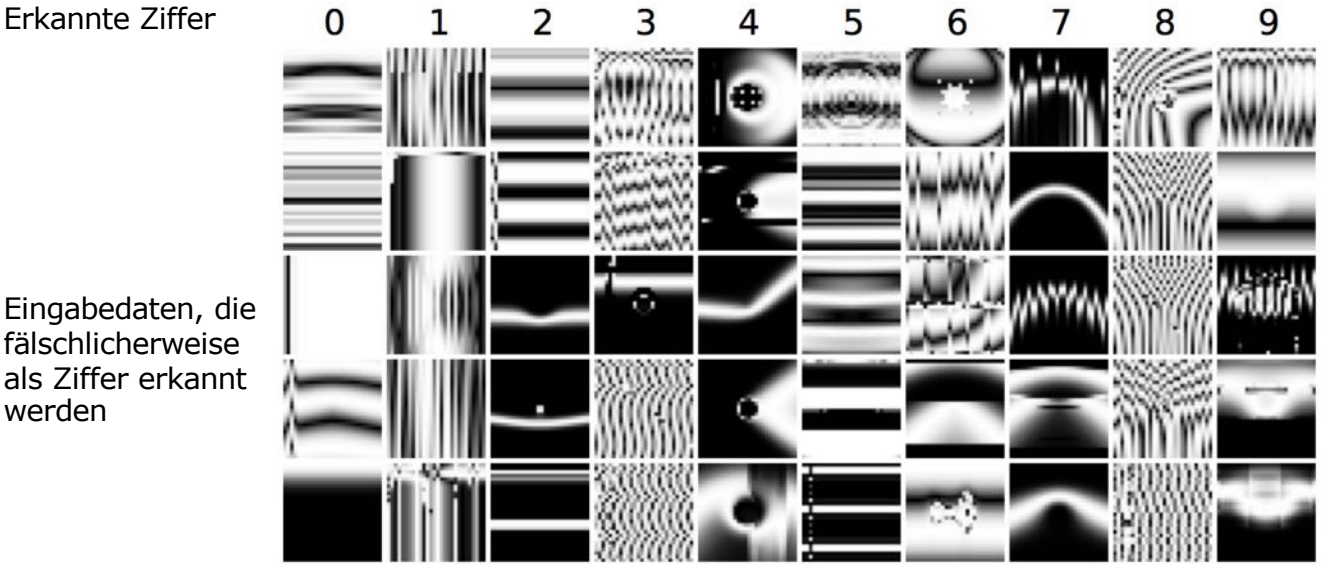




Verkehrszeichenerkennung mittels maschineller Lernverfahren

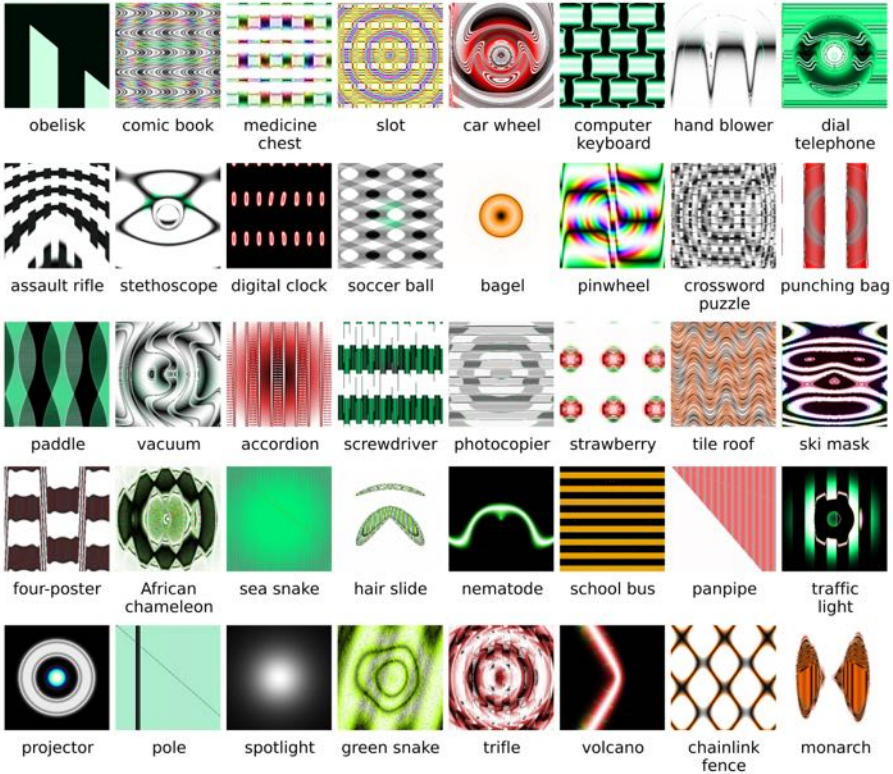
J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Neural Networks, Volume 32, August 2012, Pages 323-332, ISSN 0893-6080

…klappt leider nicht immer



Nguyen A, Yosinski J, Clune J. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In CVPR ’15, IEEE, 2015.

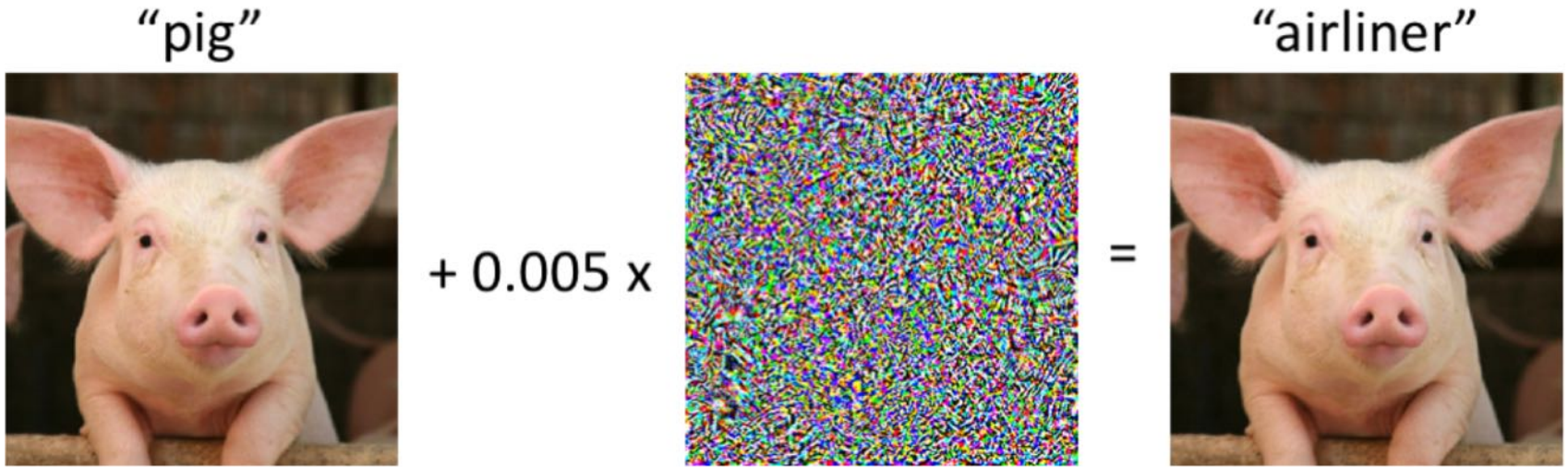
Gezielte Manipulationen eines autonomen Fahrzeugs möglich



Nguyen A, Yosinski J, Clune J. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In CVPR ’15, IEEE, 2015.

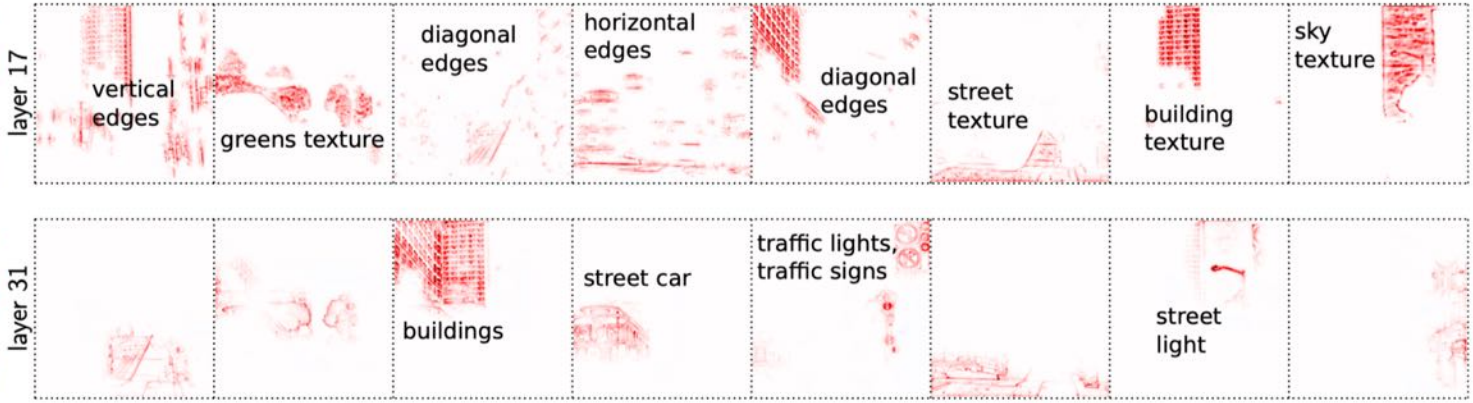
Gezielte Manipulationen

- Adversarial learning

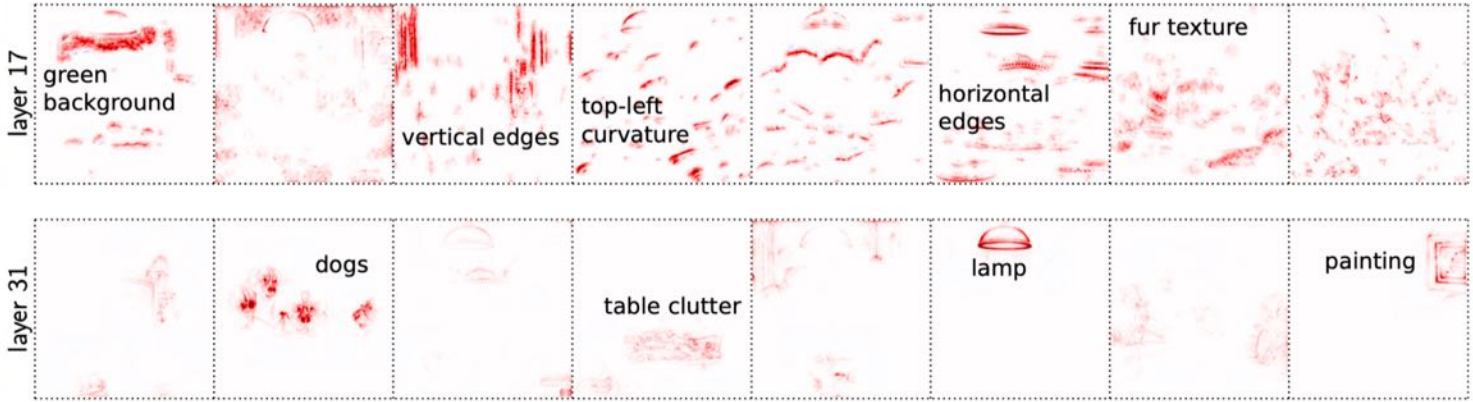


https://www.designnews.com/electronics-test/yes-ai-can-be-tricked-and-its-serious-problem/161652909959780

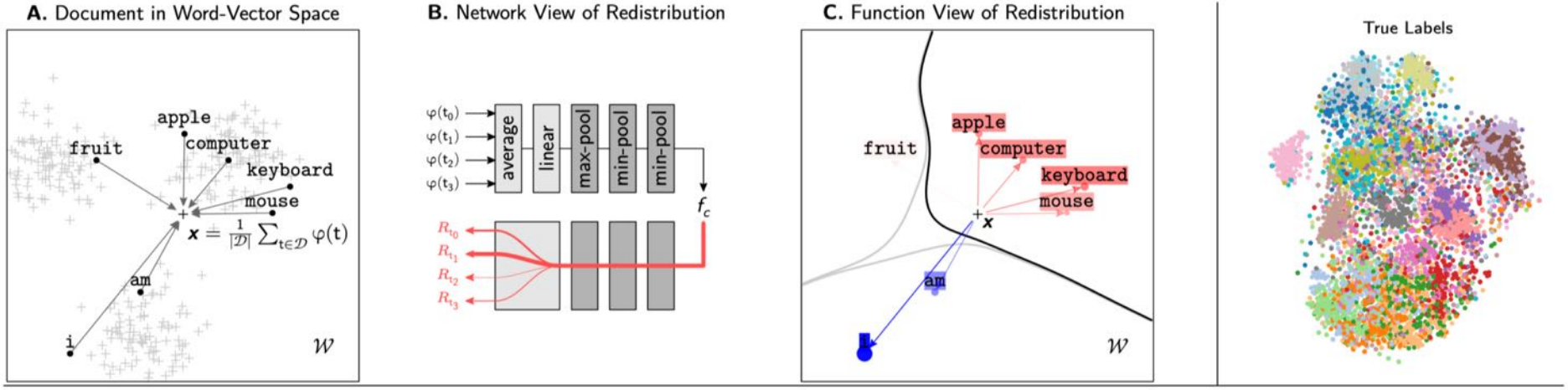
City and streetcar



"Poker Game" (Coolidge, 1894)



J. Kauffmann, M. Esders, G. Montavon, W. Samek, K. Müller, From Clustering to Cluster Explanations via Neural Networks CoRR, 2019, https://arxiv.org/abs/1906.07633



Cluster Assignments

D. talk.politics.guns

Even if it were a capital offense, the warrant was not even an arrest warrant, but a search warrant. In other words, there was no evidence of illegal arms, just enough of a suggestion to get a judge to sign a license to search for illegal evidence.

E. sci.crypt

You can find the salient difference in any number of 5th amendment related Supreme Court opinions. The Court limits 5th amendment protections to what they call "testimonial" evidence, as opposed to physical evidence.

The whole question would hinge on whether a crypto key would be considered "testimonial" evidence. I suppose arguments could be made either way, though obviously I would hope it would be considered testimonial.

F. rec.motorcycles

I'm not sure on the older bikes, but the Yamaha Virago 535 has spec'd seat height of 27.6 in. and the Honda Shadow 27.2 in.

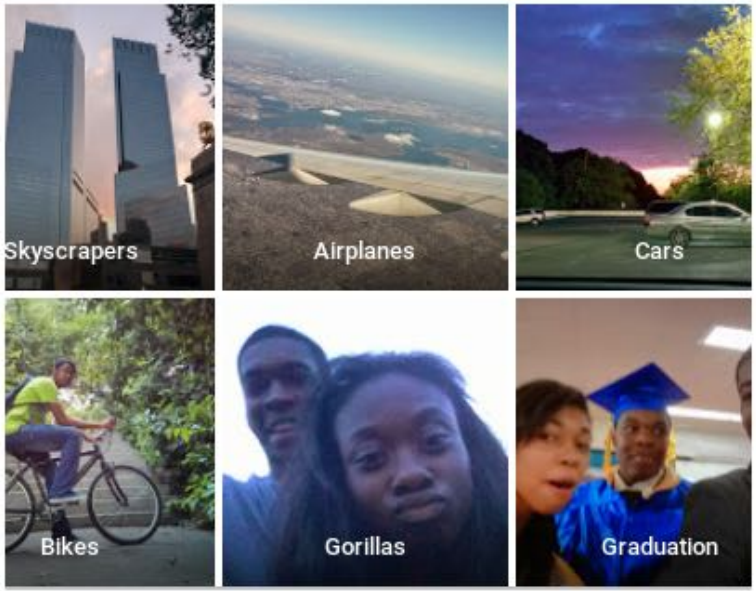
G. misc.forsale

For Sale: A Thule Car rack with 2 bike holder accessories.
Comes with Nissan Pathfinder brackets but you can buy the appropriate ones for your car cheap.
Looking for $100.00 for everything. I live in the Bethesda area.
Thanks for your interest.

J. Kauffmann, M. Esders, G. Montavon, W. Samek, K. Müller, From Clustering to Cluster Explanations via Neural Networks CoRR, 2019, https://arxiv.org/abs/1906.07633

Algorithmen sind nicht neutral, sondern enthalten immer Wertungen

- Machine Learning:
  - Auswahl des Trainingssets erzeugt stereotype Muster



c|net BEST PRODUCTS REVIEWS NEWS VIDEO HOW TO SMART HOME

Google apologizes for algorithm mistakenly calling black people 'gorillas'

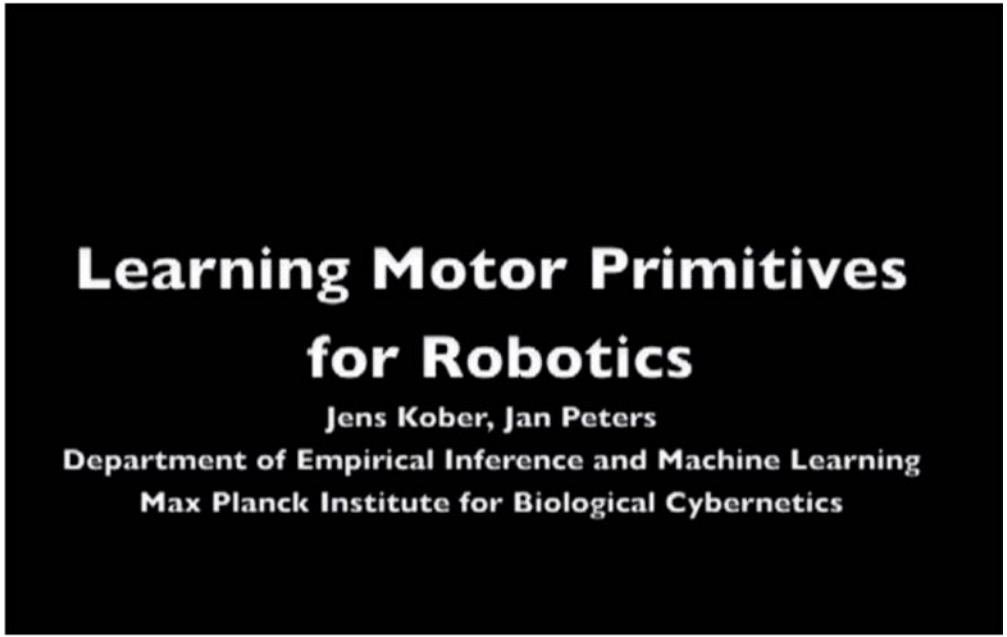
The search giant is under fire after its Photo app offensively mislabeled user's photos. It points to another challenge Silicon Valley companies have to face when developing next-gen tech: sensitivity.

Richard Nieva July 1, 2015 5:32 PM PDT ES 99+

https://www.cnet.com/news/google-apologizes-for-algorithm-mistakenly-calling-black-people-gorillas/

Algorithmen sind nicht neutral, sondern enthalten immer Wertungen

- Machine Learning:
  - Auswahl des Trainingssets erzeugt stereotype Muster
  - verfestigt solche Muster (algorithmische Normalisierung)
  - Indeterminismus erzeugt Unschärfe
  - macht Algorithmen anpassungsfähiger

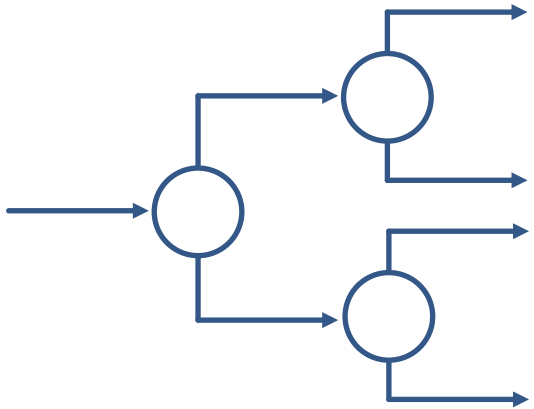


https://www.ias.informatik.tu-darmstadt.de/Research/LearningMotorPrimitives

Algorithmen sind nicht neutral, sondern enthalten immer Wertungen

- Machine Learning:
  - Auswahl des Trainingssets erzeugt stereotype Muster
  - verfestigt solche Muster (algorithmische Normalisierung)
  - Indeterminismus erzeugt Unschärfe
  - macht Algorithmen anpassungsfähiger

- Einfache Entscheidungslogiken
  - Entwickler hat expliziten Entscheidungsbaum hinterlegt
  - Deterministische Verhaltensweise



Algorithmen sind nicht neutral, sondern enthalten immer Wertungen

- Machine Learning:
  - Auswahl des Trainingssets erzeugt stereotype Muster
  - verfestigt solche Muster (algorithmische Normalisierung)
  - Indeterminismus erzeugt Unschärfe
  - macht Algorithmen anpassungsfähiger
- Einfache Entscheidungslogiken
  - Entwickler hat expliziten Entscheidungsbaum hinterlegt
  - Deterministische Verhaltensweise

inf.uni-hamburg.de

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DEPARTMENT OF INFORMATICS
SECURITY AND PRIVACY

HOME COURSES THESES RESEARCH PEOPLE SERVICE

UHH → MIN-Fakultät → Fachbereich Informatik → Einrichtungen → Arbeitsbereiche → Security and Privacy → Home

WORKING GROUP ON «SECURITY AND PRIVACY»

Security and Privacy

Information systems become more and more important in critical infrastructures, while the Internet has evolved to a critical infrastructure itself. The secure operation of these infrastructures is vital and their failure can have severe impacts up to the loss of human lives.
Security refers to the fact that protection goals are achieved in the presence of malicious attacks and system failures. Typical security goals can be confidentiality, integrity, accountability, and availability. Security and privacy in information systems addresses both technical and organizational aspects, such as building and establishing security concepts and security infrastructures as well as risk analysis and risk management.
Privacy can be a conflicting goal to security, but they can also benefit from each other. Hence, it is necessary to balance both when developing secure information systems.

Prof. Dr. Hannes Federrath
Fachbereich Informatik
Universität Hamburg
Vogt-Kölln-Straße 30
D-22527 Hamburg

Telefon +49 40 42883 2358

federrath@informatik.uni-hamburg.de

https://svs.informatik.uni-hamburg.de