
sicherheitskultur.at

# Algorithm Ethics

# Algorithmenethik

**Philipp Schaumann**

https://sicherheitskultur.at/business_ethik.htm
https://philipps-welt.info/robots.htm
https://sicherheitskultur.at/autos.htm
https://sicherheitskultur.at/Manipulation.htm



AI & Algorithmen Ethik - AI & Algorithm Ethics

## Inhalt

- Algorithmen steuern und lenken uns
- Algorithmen treffen Entscheidungen über uns
- Algorithmen haben systembedingte Fehlerquellen
  - Korrelationen statt Kausalitäten
  - Deep Learning = Garbage in – Garbage out (Vorurteil rein – Vorurteil raus)
- Was wären Anforderungen an ethisch akzeptable Implementierungen
- Algorithmen implementieren das Geschäftsmodell des Überwachungskapitalismus – das Geschäftsmodell verträgt sich nicht mit Ethik

# Wo werden Algorithmen (sinnvoll ?) eingesetzt ?

- Algorithmen übernehmen die Kontrolle, wo Menschen zu langsam sind (z.B. bei Waffen)
- Algorithmen sind im Einsatz wo dadurch Arbeitsplätze eingespart werden können
- Algorithmen sind im Einsatz wo dadurch zusätzliche „Erkenntnisse“ gewonnen werden können (z.B. bei Bewertungen)

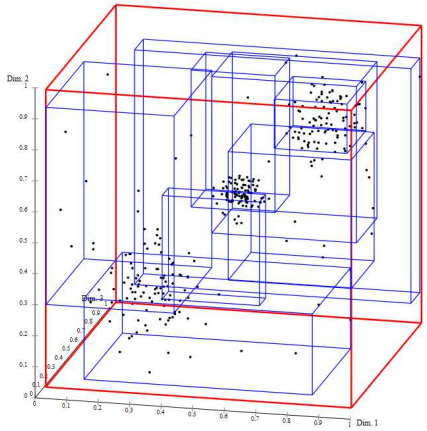


AI & Algorithmen Ethik - AI & Algorithm Ethics

# Wie funktioniert Big Data? (1)

- In bestehenden Daten wird nach Mustern gesucht, nach (mehr oder weniger zufälligen) Zusammenhängen/ Korrelationen,

- z.B. indem sie die Daten korrelieren, nach Häufungen suchen, Cluster finden, (die nicht mal wirklich existieren müssen) ….
- Sie finden immer Zusammenhänge, egal ob real oder zufällig



# Wie funktioniert Big Data? (2)

- Gleichzeitig geben Auftraggeber vor, in welche Richtung die zu optimierende Variable sich für die Gesamtpopulation bewegen soll
  - Ertrag / Gewinn (durch personalisierte Werbung)
  - Verweilzeit auf der Website / im Netzwerk
  - Kosteneinsparung (Reduktion von Schulungen mit geringen Erfolgsaussichten)
  - Rückgang der Kriminalität
  - Häufigkeit von Rückfall (Resozialisierung)





## Viele Optimierungsoptionen

- Optimierung von Gewinn (z.B. durch höhere Preise für „preiselastische“ Zielgruppen – z.B. Apple User)
- Umsatzsteigerung (z.B. durch gezielte Sonderangebote für Kunden die darauf anspringen)
- Auslastung (z.B. durch Preisoptimierung bei Flügen)
- Reduktion von Kreditausfall-Risikos (z.B. durch Analyse des Profils der Kunden)
- Reduktion von Kundenverlust durch Bewertung der erwarteten „Treue“ des Kunden

# Der Missbrauch von Korrelationen

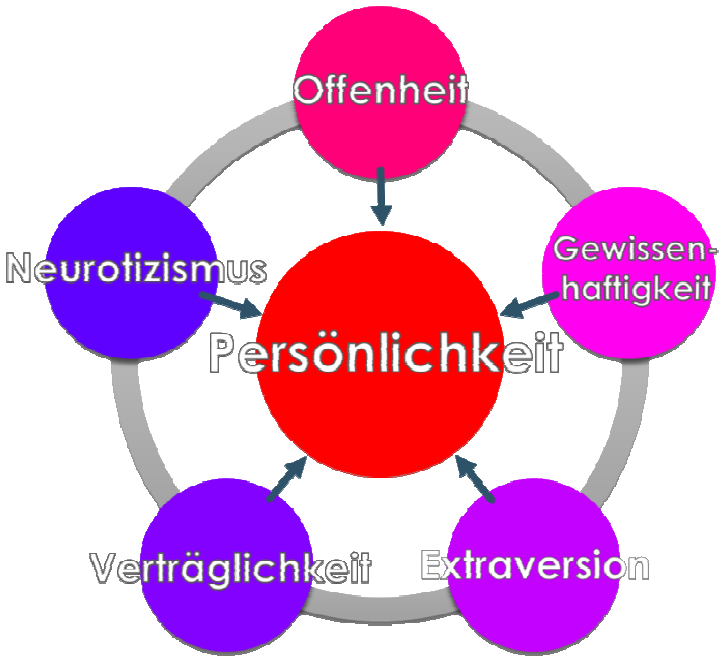
- Die Statistik erkennt Korrelationen, das sind valide Aussagen über die Gesamtpopulation.
- „Business“ sucht aber Entscheidungshilfen für den Einzelfall und konstruiert Kausalitäten:
  - „Wer dort wohnt und diese „friends“ hat, der ist ein Kredit-Risiko“
  - „Weil andere Personen mit ähnlichem Profil auch keinen Job gefunden haben, gibt es keine Weiterbildung“

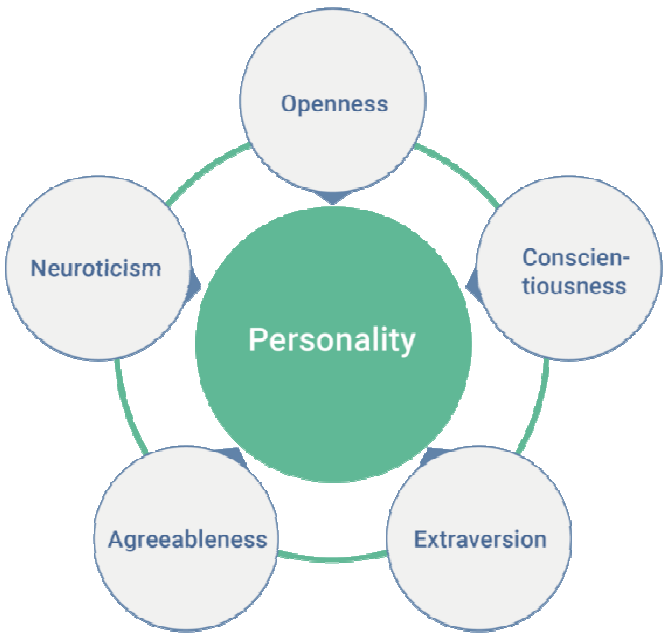
http://sicherheitskultur.at/





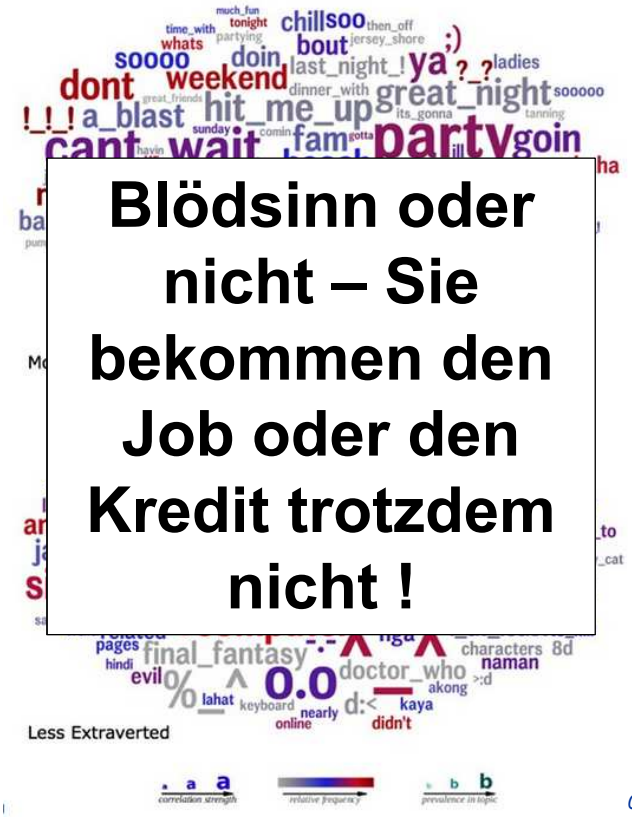
# Big Five personality traits




https://en.wikipedia.org/wiki/Big_Five_personality_traits

# Persönlichkeitsanalyse über Schreibstil




Wortwahl, Verteilung der Wortlängen und Satzlängen sind korreliert mit den „Big Five – Persönlichkeiten“

Diese „Persönlichkeiten“ sind korreliert mit anderen Parametern und die bilden die Basis für Entscheidungen.

http://www.wired.co.uk/news/archive/2013-10/03/facebook-language-study





# Vorhersage unserer Wünsche und Bedürfnisse – maßgeschneiderte Werbung

Amazon, Google, Facebook und die anderen beobachten unser Verhalten -

sie wissen (oft besser als wir selbst), was wir gerne kaufen würden / kaufen werden.

Und das, obwohl sie natürlich nicht „verstehen“ wie Menschen „ticken“

vielleicht gerade WEIL sie die Menschen nicht verstehen

http://sicherheitskultur.at/

# Gezielte Werbung – Problem oder kein Problem?

- Benutzer werden in Gruppen / Klassen eingeteilt, die dann gezielt beworben werden können
- Klassifizierungen die bisher (immer mal wieder) angeboten wurden:
  - Altersklassen
  - "Gewichtsprobleme" / Diät-Themen / Körperbild-Probleme
  - Suchabfragen nach bestimmten Krankheiten (AIDS, …)
  - Suchabfragen zu Alkohol-, Drogenmissbrauch
  - Suchabfragen zu Selbstmord-Themen
  - Suchabfragen zu Vergewaltigung
  - Politische Einstellungen (progressiv, konservativ, Umweltschutz, Judenhasser, Hitlerfans, ….)
  - Gefühl von "Wertlosigkeit"





# Optimierung für Verweildauer / Suchtpotential

- Ziel ist, dass der Benutzer auf DIESER Website bleibt, d.h. das nächste Video anschaut oder den nächsten Post liest
- Der Benutzer soll durch die Inhalte motiviert wird, Inhalte zu "sharen", d.h. andere zu aktivieren und zu Interaktionen zu motivieren
- Effektivste Mittel dafür sind (neben Katzenvideos) Emotionen, Aufregung, Zorn, Entrüstung, …
- Unerwünschter Nebeneffekt sind manchmal Pogrome wie in Sri Lanka, Myanmar, Indien, Südamerika, ….
  https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html
  https://www.nytimes.com/interactive/2018/07/18/technology/whatsapp-india-killings.html

https://philipps-welt.info/social networks.htm#sucht

# Wahlbeeinflussung

## Selective Information

Users click only those search results and articles that are presented to them. Tests show that like and dislike and voting behaviour can be influenced
https://aibrt.org/downloads/EPSTEIN_and_Robertson_2013-Democracy_at_Risk-APS-summary-5-13.pdf

## Selective Encouragement

Facebook has shown that they can increase voting by selectively encouriging selected voters
https://netzpolitik.org/2014/wie-facebook-wahlen-beeinflussen-kann-oder-was-tun-gegen-digitales-gerrymandering/





# Von Korrelation zu Artificial Intelligence

- **Bis jetzt haben wir uns mit dem Problem Korrelation statt Kausalität beschäftigt – aus statistischen Fakten werden Falschaussagen über einzelne Personen**

- **Der nächste Schritt ist Artificial Intelligence (AI)**
  **(Grund: es ist billiger in der Entwicklung)**

# If we know how it works

# –

# it's not artificial intelligence

**Charles Stross (S-F Autor)**

http://sicherheitskultur.at/

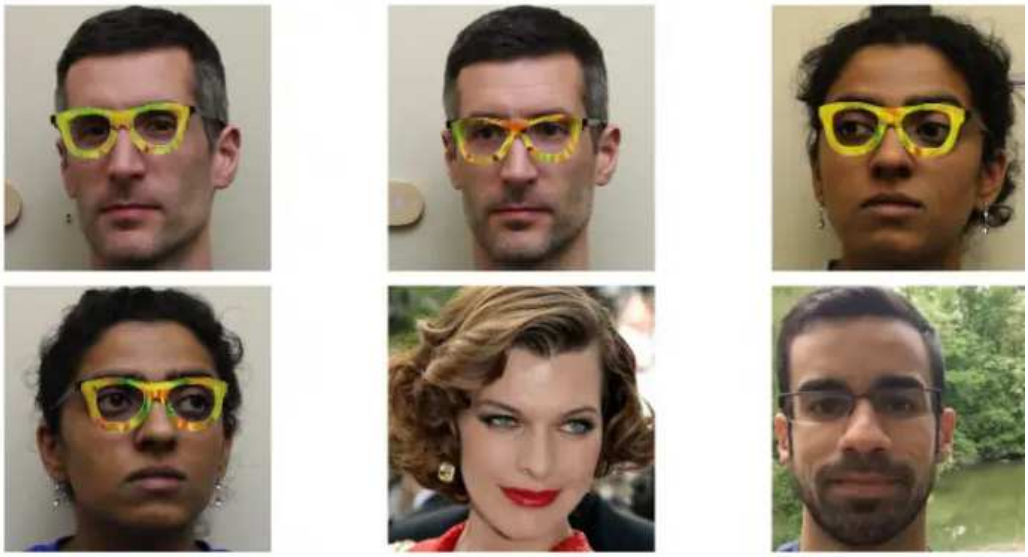




## Fehlerquellen von Deep Learning (AI)

1. Neuronale Netze machen (systembedingte) Fehler
2. Neuronalen Netze können nicht erklären, warum sie gewisse Entscheidungen getroffen haben

# Deep Learning Algorithmen machen (systembedingte) Fehler



- **Durch solche Brillen erkennt das System andere existierende Personen (z.B. Milla Jovovich)**

https://www.heise.de/newsticker/meldung/Buntes-Brillengestell-soll-zuverlaessig-Gesichtserkennung-austricksen-3456711.html
https://www.newscientist.com/article/2111041-glasses-make-face-recognition-tech-think-youre-milla-jovovich/
***Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security***, DOI: 10.1145/2976749.2978392





# Algorithmen machen (systembedingte) Fehler (2)


**Dieses Spielzeug wird als Pistole erkannt**

https://www.newscientist.com/article/2152331-visual-trick-fools-ai-into-thinking-a-turtle-is-really-a-rifle/

# Deep Learning Algorithmen machen (systembedingte) Fehler (3)

- Gesichtserkennung funktioniert nur sehr begrenzt bei schwarzer oder sehr heller Haut
- Großbritannien hat eine Website für das Beantragen eines Passes, der bei zu schwarzer Haut das Photo ablehnt

https://www.newscientist.com/article/2219284-uk-launched-passport-photo-checker-it-knew-would-fail-with-dark-skin/



AI & Algorithmen Ethik - AI & Algorithm Ethics

# Algorithmen treffen Entscheidungen über das Schicksal von Menschen

# Algorithmen steuern die Welt
# Vorhersage von Kriminalität (Predictive Policing)

Neuronale Netze analysieren Orte im Hinblick auf frühere Straftaten (oder Ereignisse), werten Überwachungskameras aus und schicken Polizisten dorthin, wo mit höchster Wahrscheinlichkeit ein Verbrechen passieren wird.

An diesen Orten wird kritischer beobachtet, kritischer geprüft, mehr entdeckt – und das fließt wieder in die Statistik rein.





# „Algorithmic Sentencing“
# Vorhersage von Rückfallwahrscheinlichkeit

Neuronale Netze analysieren die Wahrscheinlichkeit, dass ein Angeklagter mit einer bestimmten Geschichte und einem bestimmten sozialen Umfeld rückfällig wird.
Sie entscheiden über Bewährungsstrafe und (vorzeitige) Entlassung und schlagen auch das Strafmaß vor.


http://www.mosac.mo.gov/page.jsp?id=45498

# Studies and Tests: Algorithms do have bias

- **White House report on Big Data . . . cautions against re-encoding bias and discrimination into algorithmic systems.**
  https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights

- **Software used to predict future criminals is biased against blacks**
  https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

- **What Algorithmic Injustice Looks Like in Real Life**
  https://www.propublica.org/article/what-algorithmic-injustice-looks-like-in-real-life

- **Amazon: KI zur Bewerbungsprüfung benachteiligte Frauen**
  https://www.heise.de/newsticker/meldung/Amazon-KI-zur-Bewerbungspruefung-benachteiligte-Frauen-4189356.html





# Thema Bewerbung für einen Job

Einstellungsgespräche waren immer subjektiv, aber . . .

Die Vorurteile des Einstellenden werden ersetzt durch die mathematische Willkür eines unbekannten Algorithmus mit unbekannter Parametrisierung und fragwürdigen Input Daten.

Die Inputs des Algorithmus sind z.B.
- Bewerbungsunterlagen (und andere Texte) der jetzigen Angestellten und der BewerberIn analysiert durch Deep Learning
- Spuren der BewerberIn im Web (oder die fehlenden Spuren)
- „Friends“ in Social Networks, deren Postings, Wohnorte, Likes,
- Außerhalb der EU auch: Konsumverhalten der BewerberIn,
- …..

# Amazon USA:
# Recruiting thru Deep Learning stopped

Der Algorithmus diskriminierte gegenüber Frauen, weil er auf der bisherigen Einstellungspraxis beruhte.
https://www.heise.de/newsticker/meldung/Amazon-KI-zur-Bewerbungspruefung-benachteiligte-Frauen-4189356.html

Wenn die Lernbasis für Deep Learning die Vergangenheit ist, so kann sich nichts verändern.





# Einstellung auf Grund von Persönlichkeitsanalysen

Der Algorithmus bekommt vielleicht die Vorgabe „harmonisches Team“ oder (eher unwahr-scheinlich) „kreative Querdenker“.

Die Textlänge und die Wortlänge bei Postings gibt mit ausreichender Wahrscheinlichkeit die Einordnung in die (Pseudo-) Persönlichkeitskategorien „Big 5“ wieder.

# Einstellung auf Grund von Key-Words

Nicht wirklich besser:
Simple Algorithmen, die die Häufigkeit von bestimmten (recht willkürlichen) Schlüsselwörter prüfen. (Beispiele: innovative, motivated, dynamic, organized, reliable, honest, creative, experience, helped, supervised, confidence, consistent, …)

https://www.cvplaza.com/cv-basics/cv-power-words/
https://www.jobsite.co.uk/worklife/how-to-use-keywords-cv-7317/
https://www.cv-library.co.uk/career-advice/cv/how-use-keywords-cv/
https://www.callcentrehelper.com/the-top-25-words-to-use-on-your-cv-10032.htm
https://www.reed.co.uk/career-advice/what-words-should-i-use-on-my-cv/





# Algorithmen steuern die Welt

## 2013: Vorhersage von Terrorismus

Computer analysieren Ihr Verhalten im Internet, die Vernetzungen und die Kommunikation und sagen vorher, mit welcher Wahrscheinlichkeit Sie zum “Trouble Maker” werden.

Zukünftige „Trouble Maker“ kommen auf die No-Fly Liste oder werden bei jedem Flug separat verhört.

Die Gedanken hören auf, frei zu sein.

# Guilty through Clustering

Instead of
„Other customer like you also bought the following books“

You get
„Other people with the same word-length pattern in tweets and on Facebook and the same network-topology in their call-records meta-data have become suicide bombers – welcome to the no-fly list“





# Algorithmen steuern die Welt 2018: Vorhersage der Aufklärungswahrscheinlichkeit

In England entscheidet ein Algorithmus welcher Gewalttat zwecks Aufklärung nachgegangen werden soll.
Basis ist die Wahrscheinlichkeit, dass auf Grund bekannter Parameter (Zeugen vorhanden, Video feed vorhanden, Umgebung, Opfer – Täter Verhältnis oder nicht, …) eine Aufklärung wahrscheinlich ist.

Mögliches Ergebnis: Örtlichkeiten, bei denen nie eine Nachverfolgung passiert

https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

# Algorithmen steuern die Welt 2018: Vorhersage der Aufklärungswahrscheinlichkeit (2)

Gegenmaßnahme:

der Algorithmus „lügt“ in einer gewissen Zahl von Fällen um zu prüfen, ob vielleicht doch eine Aufklärung gelingt und dadurch den Algorithmus nachzuschärfen





# Beispiel Arbeitsamt Österreich (AMS)

- **Arbeitslose werden künftig (2019) in drei Gruppen A, B, C eingeteilt und zwar in jene mit hohen, mittleren und niedrigen Chancen am Arbeitsmarkt.**
- **Wer mit 66-prozentiger Wahrscheinlichkeit innerhalb von sieben Monaten wieder einen Job haben wird, soll ab 2019 als Person mit hoher Arbeitsmarktchance gelten, Gruppe A.**
- **Wer weniger als 25 Prozent Chance hat innerhalb von zwei Jahren einen Job zu bekommen, gilt dann als Kunde mit niedrigen Chancen und kommt in Gruppen B oder bei ganz schlecht, in C.**
- **Förderungen, z.B. Schulungen, gibt es hauptsächlich für Gruppe B.**

https://derstandard.at/2000089720308/Leseanleitung-zum-AMS-Algorithmus

# "MathWashing"
# Algorithmen für eine weiße Weste

„Ich würde ja gern …. Aber der Computer sagt NEIN"

**Absichtliches** Verstecken hinter dem Algorithmus ….
durch Rausreden auf „wertfreien", „objektiven" Algorithmus. Siehe Facebook, Google, etc.: Wir sind ja nur eine Plattform und wollen keine Verantwortung für Inhalte.

**Unabsichtliches** Verstecken hinter dem Algorithmus ….
Targeted Advertising (gezielte Anzeigen) die manchmal eine Re-Traumatisierung auslösen (z.B. Diätmittel für Anorektiker, Glückspiel für Süchtige, etc.)

**Den Begriff „Mathwashing" hat Fred Benenson geprägt.**
**https://technical.ly/brooklyn/2016/06/08/fred-benenson-mathwashing-facebook-data-worship/**
**Quelle viele Materialien: Bertelsmann Stiftung https://algorithmenethik.de/mathwashing/**



AI & Algorithmen Ethik - AI & Algorithm Ethics

# False Positive – Problem
# E.g. Face Recognition Test in Berlin

*Das Überwachungssystem am Berliner Bahnhof Südkreuz versucht, gesuchte Verdächtige im Strom der Passanten zu erkennen?*

„Bei 70 Prozent und mehr haben wir eine positive Erkennung der gesuchten Testpersonen – das ist ein sehr guter Wert."

| | CP | CN |
|---|---|---|
| OP | TP | FP |
| ON | FN | TN |

Falsch!
Die Aussage ist ohne Wert, dies ist nur die „true positive rate"

https://sicherheitskultur.at/privacy_loss.htm#face
https://algorithmenethik.de/2017/12/21/eine-polemik-wie-man-mit-einem-wuerfelnden-schimpansen-terroristen-faengt/

# False Positive – Problem der Face Recognition Test in Berlin

Annahme: 1 Million Personen passieren pro Jahr das Berliner Südkreuz, davon 4 Terroristen.

Bei 75% Erkennungsrate werden in dem Jahr 3 davon gefasst, nicht schlecht.

Nehmen wir gleichzeitig eine false-positive Rate von 1% an, dann werden 1% der Personen im Bereich der Überwachungskamera fälschlich als Terrorist eingestuft.

Das ergibt 10.000 Fehlalarme, d.h. "fälschlich als Terrorist erkannt". Wie soll eine Polizei mit dieser Rate der Fehlalarme umgehen? Und es gibt bestimmt einigen Zorn, wenn 10.000 Leute pro Jahr auf dem Weg zur Arbeit aufgegriffen und verhört werden.

https://twitter.com/FlorianGallwitz/status/956865341392609281





# False Positive – Problem (3)
## E.g. Champions-League-Finalspiels 2017 in Cardiff

… unter den 170.000 Besuchern der Stadt werden 2470 als Kriminelle erkannt

Korrekt a

D.h. die
den groß
während

**Wo automatisiert entschieden wird, da fallen viel mehr Entscheidungen und es werden auch viel mehr Fehler gemacht**

https://futurezone.at/netzpolitik/gesichtserkennung-markiert-tausende-versehentlich-als-kriminelle/400031842

# Ethical Criteria for Algorithms

- **Respecting Human Rights**
- **Representing Reality / Search Results as objectively as possible – trying to prevent bias**
- **As much transparency as possible regarding the algorithm and its goals / optimization criteria**
- **Auditability / Explainability**
- **Fairness, Preventing Discrimination**
- **Adaptability / in-built self-checks**
- **Offering alternative (human-based) decision-making**
- **Reduction of complexity as far as possible**
- **Resource-efficiency**



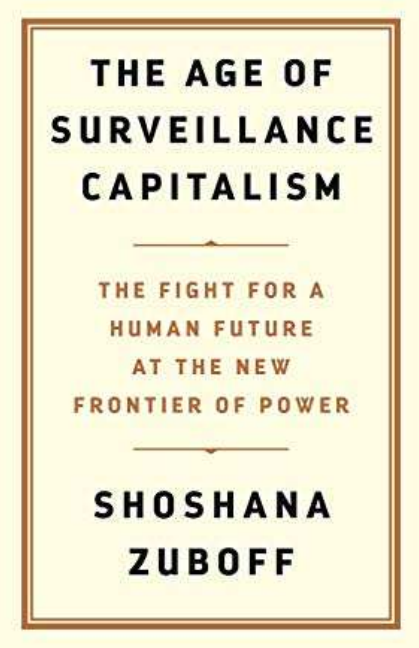


# Probleme mit dem (derzeitigen) Geschäftsmodell des Internets

- Das dominierende Geschäftsmodell im Internet ist der "Überwachungskapitalismus"
- So viel Daten sammeln wie möglich und daraus Entscheidungen berechnen die eine Optimierung des Geschäftserfolgs erreichen
- "Da ist kein Platz für ethische Überlegungen" (Ex-Facebook-Investor Roger McNamee)

# More Information

- **AI Now Institute, New York University, https://ainowinstitute.org/AI_Now_2018_Report.pdf**
- **https://algorithmenethik.de/**
  **sehr lesenswerter wöchentlicher Newsletter**
- **Critical voices: Women in AI**
  **https://www.womeninai.co/**
  **data 4 good Vienna**
  **https://data4good.viennadatasciencegroup.at/**
- **Shoshana Zuboff**
  **The Age of Surveillance Capitalism →**

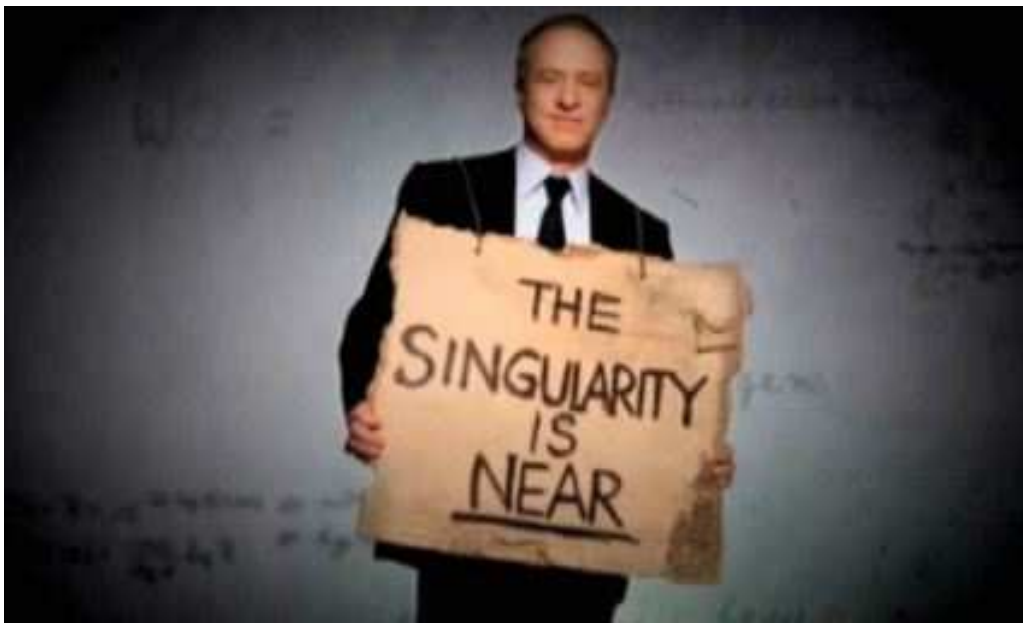




AI & Algorithmen Ethik - AI & Algorithm Ethics

# Danke




**Ray Kurzweil**

**Mehr dazu:**
**http://sicherheitskultur.at/Manipulation.htm**
**http://philipps-welt.info/robots.htm#asimovlaws**