

Automated Decision-Making

Der Regulierungsvorschlag der Digitalen Gesellschaft

26. Februar 2022

Die Talentsuche

Digitaler Headhunter!

Inc.

How Amazon Accidentally Invented a
Sexist Hiring Algorithm

A company experiment to use artificial intelligence
in hiring inadvertently favored male candidates. [🔗](#)

Automated Decision Making (ADM) System

ADM-Systeme



ADM-Systeme: Entscheide zu sensiblen Lebensbereichen

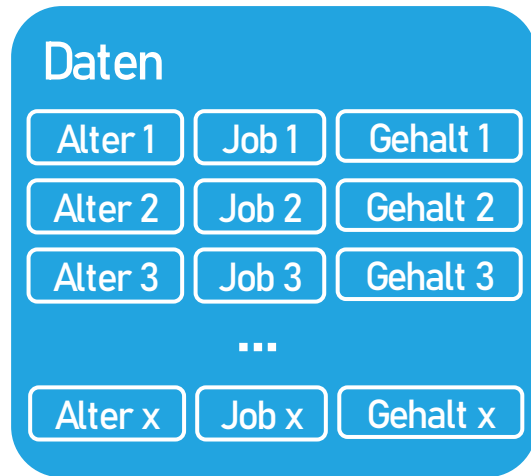
Bsp.: Ausbildungsmöglichkeiten, Gesundheitsergebnisse, Arbeitsleistung, Mobilität, Interessen, Verhalten, persönliche Autonomie..

Nur Einzelne Personen? Nein!

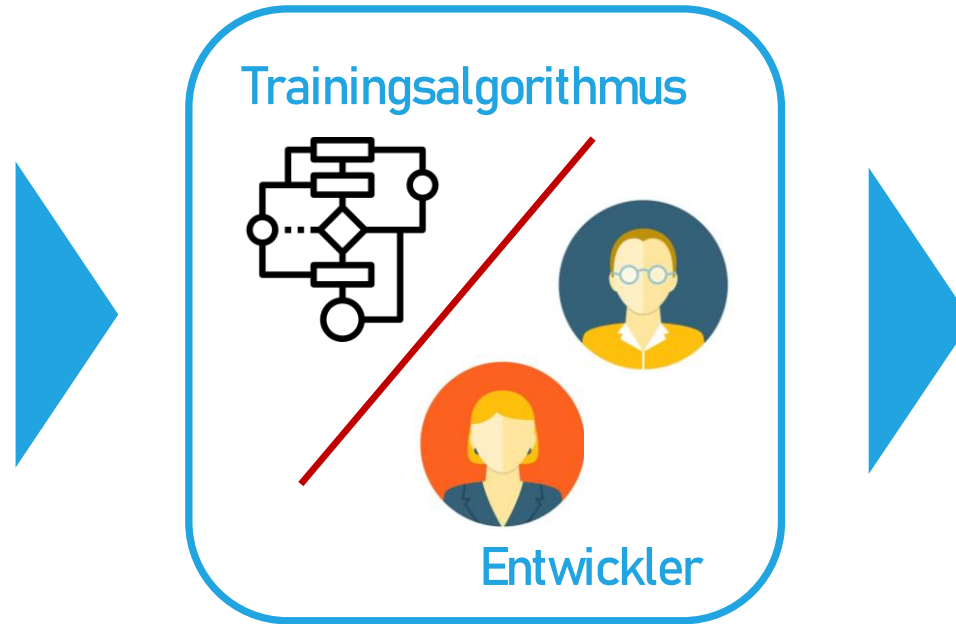
Bsp.: optimale Schülerzuordnung



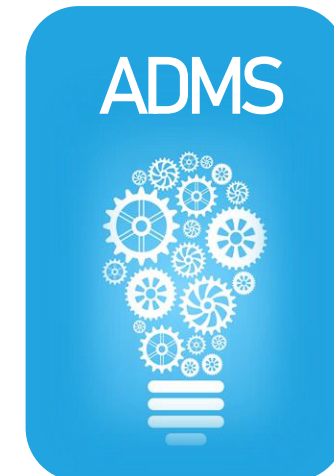
ADM-Systeme erschaffen



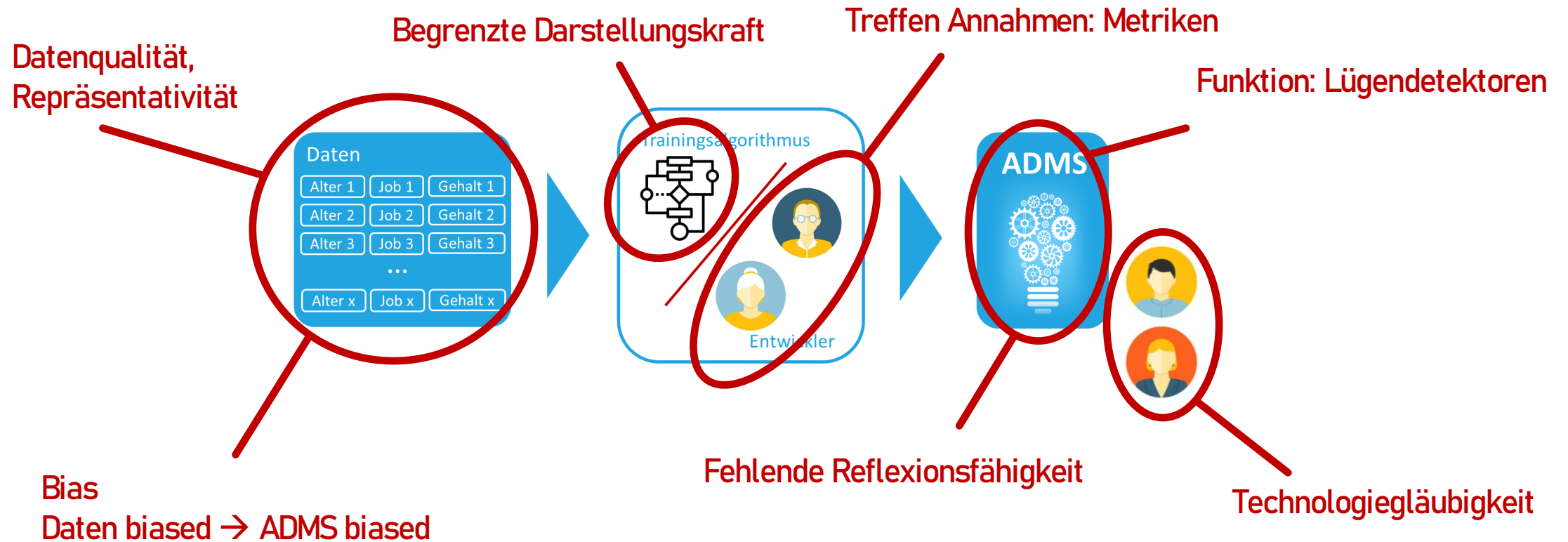
Trainings-Daten



Konzeption / Training



Diskriminierung und Co



ADM-Systeme reproduzieren gesellschaftliche Normen

Werte?

Suchbegriff... **SUCHEN**

Blick

News Sport People

Schweiz Regionen Ausland **Politik** Wirtschaft Leserreport

«In gegenseitigem Einvernehmen»

Bund feuert Porno-Sekretärin

Publiziert: 17.08 Uhr, Aktualisiert: 17.34 Uhr



Zuerst hat die Sekretärin in Bern... Nun ist die f... Job ganz los... Einvernehm...

Nachdem ihre N... Bundeshaus un... geworden sind, Sekretärin A.* a... geschickt. Freig...

Nun ist sie ihren Parlementsdiens... Mitarbeiterin hät... Arbeitsverhältnis... Einvernehmen a... Parlementsdiens... zum Schluss ge... Interesse beider...

Das sexy Selfie von A. aus dem Bundeshaus. (Twitter)

DAILY NEWS NEW YORK **NEWS** POLITICS SPORTS ENTERTAINMENT

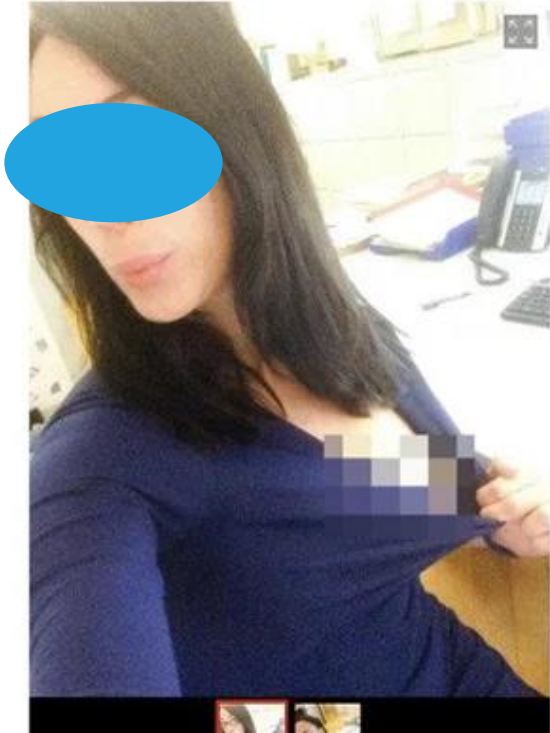
U.S. WORLD CRIME THE WEEK NEWS PICS BLOGS

Swiss parliament worker being investigated for posting naked selfies taken in office

The unidentified woman — who reportedly starred in 200 amateur adult films — was caught posting the snaps taken at Switzerland's Federal Palace on her Twitter account, which had more than 11,000 followers.

BY LEE MORAN / NEW YORK DAILY NEWS / Thursday, August 7, 2014, 8:12 AM

SHARE THIS URL
nydn.us/1szVm9H



A Swiss parliament worker is under fire for posting naked selfies taken in her office.

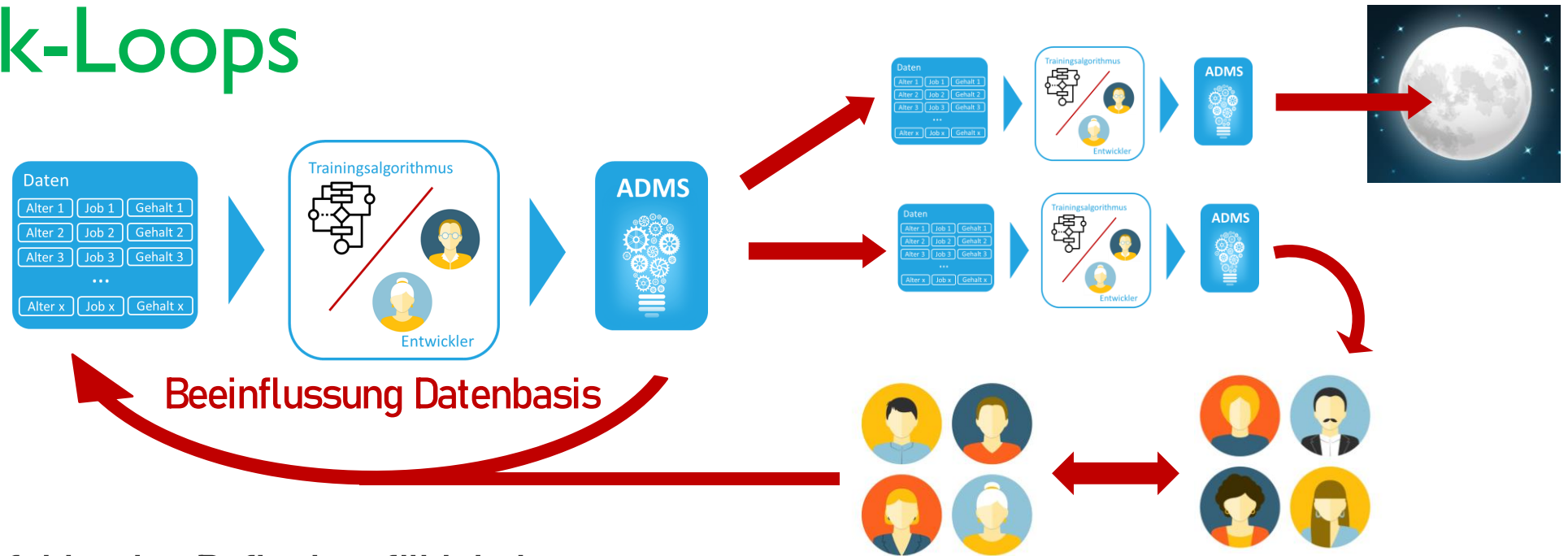
The secretary, who has not been named, racked up more than 11,000 followers with her XXX-rated snaps before her Twitter account was deleted this week.

She is now facing an investigation after it emerged many of the explicit photos were taken in Bern's 162-year-old Federal Palace, where the nation's government meets.

The woman reportedly told daily newspaper *Neue Zürcher Zeitung* that the pictures did not violate guidelines for federal employees, as they were part of her private life.

But a spokesman for the government's human resources

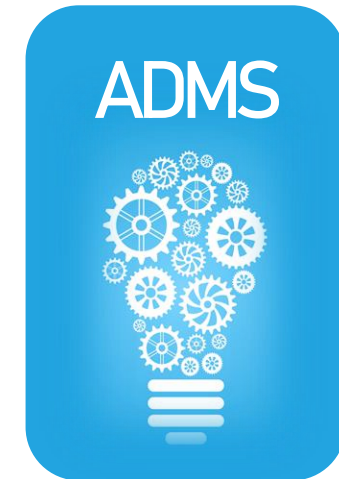
Feedback-Loops



- Ausgleich der fehlenden Reflexionsfähigkeit
- Steigende Komplexität / Ausschluss von Menschen
- Abhängigkeit von ADM-Systemen
- **Machtfrage:** Wer bestimmt die Metriken / Gleichgewichte?
- Einfluss auf Gesellschaft und demokratische Prozesse

Und nun?

- **Transparenz** beim Einsatz kritischer ADM-Systeme
- Handlungsmöglichkeiten bei Problemen
 - Nur Transparenz reicht nicht!
- Mensch besitzt **Geltungshoheit** über ADM-Systeme
- Breite Diskussion!



Ein ADM-Gesetz

Die **Digitale Gesellschaft** fordert

- **Transparenz** beim Einsatz der Systeme
 - Für Staat und Privatwirtschaft
- **Erklärbarkeit**
 - Ausreichend für grobe, externe Risikoeinschätzung
- Klare **Schutzziele**,
 - Einhaltung der Grund- und Menschenrechte
 - Wahrung der psychischen und physischen Gesundheit und Sicherheit des Einzelnen
 - Wahrung der Lebens- und Entwicklungschancen
 - Schutz der demokratischen Rechte und Prozesse.

Risiken

Ansatz: Mischform zwischen Risiko- und Schadensbasiert

Inhalt:

Ein ADM-Gesetz

Kategorisierung

Die ADM-Aufsicht

Transparenzpflichten

Kontrolle und Sanktionen



ADM-System Kategorisierung

Handelt es sich um ein ADM-System? Falls ja:

- 1. Kategorie: **Kein/geringes Risiko**
 - Keine Massnahmen (Transparenz etc) ausser interner Einschätzung nötig
 - Bsp.: Text-Autokorrektursysteme
- 2. Kategorie: **Hohes Risiko**
 - Umfangreiche Transparenz und Rechtfertigungspflichten
 - Bsp.: medizinische Diagnostik, Asylverfahren
- 3. Kategorie: **Inakzeptables Risiko**
 - Grundsätzlich verboten
 - Bsp.: Biometrische Massenüberwachung auf öffentlichen Plätzen
- Zur Entlastung der Firmen und Behörden: **Selbstdeklaration**
 - Hohe Strafen bei zu tiefer Einschätzung

Inhalt:

Ein ADM-Gesetz

Kategorisierung

Die ADM-Aufsicht

Transparenzpflichten

Kontrolle und Sanktionen

**Details, Risiken und
Einschätzungshilfen:
Positionspapier**

Die ADM-Aufsicht

- Staatliches Kompetenzzentrum
 - Für Unternehmen, Behörden und die Öffentlichkeit
- Kontrolliert Gesetzeskonformität
- Erstellt Checklisten und Good-Practices-Anleitungen
- Orchestriert Langzeitanalysen
- **Unabhängig** von **Weisungen** und mit **eigenem Budget**

Inhalt:

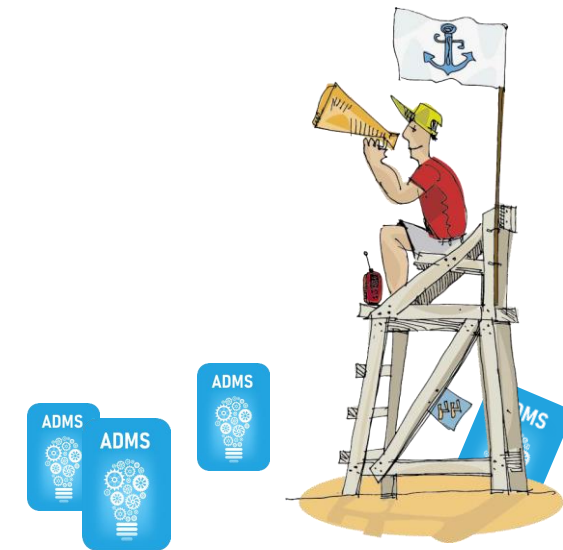
Ein ADM-Gesetz

Kategorisierung

Die ADM-Aufsicht

Transparenzpflichten

Kontrolle und Sanktionen



Transparenz- und Rechenschaftspflichten

- **Einsetzende Entität** ist für ADM-System verantwortlich!
- **Nur für ADM-Systeme mit «hohem Risiko»**
- Hinweis auf
 - Einsatz, Zweck
 - Datenherkunft sowie Features
- Datenherkunft weil: **Datenschutz, Datenqualität und Verkettung**
- **Ziel:** Möglichkeit zur Abschätzung der Effekte auf Individuen und Gesellschaft
- Volle Transparenz für Systeme in Erfüllung eines öffentlichen Auftrags
 - Alle Algorithmen und Daten (wo möglich)
 - **Public Money; Public Code!**

Inhalt:

Ein ADM-Gesetz

Kategorisierung

Die ADM-Aufsicht

Transparenzpflichten

Kontrolle und Sanktionen



Kontrolle, Massnahmen und Sanktionen

Inhalt:

Ein ADM-Gesetz

Kategorisierung

Die ADM-Aufsicht

Transparenzpflichten

Kontrolle und Sanktionen

- ADM-Aufsicht
 - Sammelt Beschwerden & fordert Einsicht bei Verdacht
 - Kann erstinstanzlich Sanktionen verhängen
- Individuen
 - Klagen bei Verdacht oder Schadensfall
 - Sammelklagen und -Beschwerden
- Berechtigte* Verbände
 - Klagen ohne persönliche Betroffenheit
- **Beweislastumkehr**
- umsatzabhängige **Verwaltungssanktionen**



*gesamtschweizerisch und mit
passendem Zweck gemäss Statuten

Nächste Schritte

- Konkretisierung / Standardisierung
 - Transparenz & Rechenschaft
 - Ausformulierter Regulierungsvorschlag
- Öffentlichkeitsarbeit
- **Feedback** zum Positionspapier
 - **Tell us!** david.sommer@digitale-gesellschaft.ch
- **Mitmachen?**
 - **Gerne!** david.sommer@digitale-gesellschaft.ch



Referenzen

Bilder:

- [Circle vector created by macrovector - www.freepik.com](https://www.freepik.com/circle-vector)
- [Business vector created by macrovector_official - www.freepik.com](https://www.freepik.com/vectors/business)
- [Algorithm icons created by Linector - Flaticon](https://www.flaticon.com/free-icons/algorithm "algorithm icons")
- <https://www.pngwing.com/en/free-png-bhgrf>
- <https://www.pngegg.com/en/png-zopgh>
- <https://www.pngwing.com/de/free-png-bywka>
- https://www.clipartmax.com/middle/m2H7d3Z5N4H7A0K9_horn-clipart-free-daily-devil-horns-transparent-background/#

Inhalt:

- <https://www.inc.com/guadalupe-gonzalez/amazon-artificial-intelligence-ai-hiring-tool-hr.html>