VISCHER

Einsatz künstlicher Intelligenz.

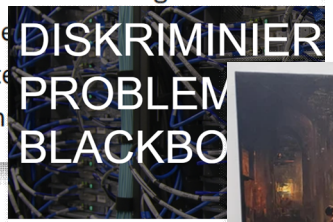
Was ist in der Schweiz (beim Staat) erlaubt?

David Rosenthal, Partner, VISCHER AG
4. Oktober 2022

Noch ein emotionales Thema …

«Technologie ist immer
irgendwas, sie ist aber au
ein Problem» – kün
gezähmt werden

Künstliche Intelligenz nimmt de
Zugl
Grot
Tech



Es ist nicht das erste Mal, dass künst
Menschen mit schwarzer Hautfarbe a
vom KI-unterstützten Einstellungspro
Schräglage, die beim Einsatz von KI de

Künstliche Intelligenz

Google feuert Ingenieur, de

MOTHERBOARD
TECH BY VICE

An AI-Generated Artwork Wor
at a State Fair Fine Arts Compe
Artists Are Pissed

Künstliche Intelligenz. Blic

Eingereicht von:

Grüne Partei der Schweiz

Einreichungsdatum: 17.06.2022

Science

Meet Loab, the AI Art Woman Haunting the Internet

Is she a demon? A Cryptid? Or nothing at all...

Jackson Ryan
Sept. 11, 2022 4:10 p.m. PT

5 min read

Meet Loab
@supercomposite/Twitt

Are there ghosts in our machines? Well, of course not, but a recent viral Twitter thread might have you believing there is something sinister lurking behind your computer screen, just waiting to be unleashed.

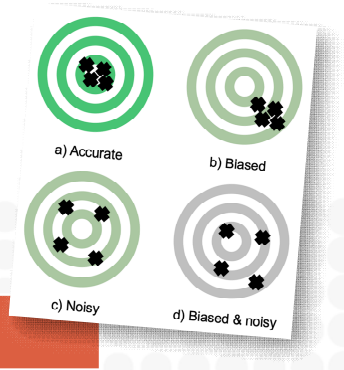
Quellen: NZZ.ch, faz.net, spiegel.de, parlament.ch, greenits.ch, vice.com, cnet.com

Worüber wir reden

- Abgrenzen
  - Systeme, die **programmiert** werden
  - Systeme, die **trainiert** werden
- Wir reden von Systemen, die **Muster erkennen** können (mehr nicht)
  - Problem 1: Korrelation, keine Kausalität
  - Problem 2: "Black Box"
  - Problem 3: "machine bias" / "blind spots"
  - Problem 1 ist statistisch nicht zwingend eins
  - Probleme 2 und 3 sind im Prinzip lösbar
- PS. Es gibt noch **weitere Probleme**

Mensch v. Maschine

Wer trifft die "besseren" Ermessensentscheide?



Basierend auf Noise, Kahneman, Sibony, Sunstein (2021)

Merkmal	Mustererkennung	Mensch
Qualitätsfaktor	Erfahrung (= Training)	Erfahrung, Vorgaben
Bias (Verzerrung)	Ja	Ja
Noise (Verrauschung)	Nein	Ja, bei Ermessen
Skaleneffekt	Ja	Nein, bei Ermessen
Nachvollziehbarkeit	Beschränkt	Beschränkt
Wahrung der Würde	Unklar	Ja

reduzierbar

Unsere Rechtsquellen

Personalrecht

- Rechtliches Gehör
- Begründungspflicht
- Gesetzmässigkeit
- Verhältnismässigkeit
- Diskriminierungsverbot

Personal- und Datenschutzrecht jeweils auf Stufe Bund und Kantone unterschiedlich geregelt

KI-Regulierung

Generelle Regulierung vom Bundesrat bisher (zurecht) abgelehnt. Auch die Schweizer Lehre sieht es offenbar mehrheitlich so.* Anders die Europäische Kommission, die im April 2021 eine Regulierung vorgeschlagen hat.

Datenschutzrecht

- Gesetzmässigkeit
- Datenrichtigkeit
- Verhältnismässigkeit
- Transparenz
- Zweckbindung
- Diskriminierungsverbot
- Automatisierte Einzelentscheide
- Profiling
- Datensicherheit
- Nachvollziehbarkeit des Tuns

* Florent Thouvenin et al., Ein Rechtsrahmen für Künstliche Intelligenz, Balsthal 2021 (https://bit.ly/3vtHLtn)

Revision IDG ZH

Kanton Zürich
Direktion der Justiz und des Innern

Vorentwurf mit erläuterndem Bericht

vom *Datum*
JI-Nr.: 2019-389

Revision Gesetz über die Information und den Datenschutz (IDG)

wird angeregt (a.a.O. S. 46). Diese Vorschläge zielen damit auf den Schutz von Personendaten bei der Verwendung einer bestimmten Technologie (automatisiertes Verfahren). Da der Schutz von Personendaten (auch bei automatisieren Verfahren) bereits heute durch die geltenden Regelungen des IDG sichergestellt wird, bieten die vorgeschlagenen Ergänzungen keinen ersichtlichen Mehrwert. Eine umfassende Regelung von KI würde gemäss Studie (a.a.O. S. 38) denn auch vor allem Änderungen im Gesetz über die Verwaltungsrechtspflege vom 24. Mai 1959 (VRG; LS 175.2) bedingen. Zu berücksichtigen ist zudem, dass in Bezug auf den Einsatz von KI noch viele Fragen offen sind, insbesondere auch mit Bezug auf den Regelungsbedarf und den Begriff «Künstliche Intelligenz» (vgl. Jonas Fischer/Mathias Fuchs, «Brauchen wir eine Legaldefinition für künstliche Intelligenz?», in: Jusletter 8. November 2021). Der Begriff «Künstliche Intelligenz» stammt aus der Computerwissenschaft und nicht aus der Rechtswissenschaft. Die Künstliche Intelligenz an sich ex[illegible] vielmehr verbergen sich hinter dieser Grundlagentechnologie unterschiedlichste V[illegible] welche je nach Bereich eine abweichende Nutzungstiefe aufweisen [illegible] 7). Eine Klärung der Bedürfnisse und der offenen Fragen und eine umfassende Regelung [illegible] jetzigen Zeitpunkt erscheint damit als verfrüht. Dies kann allenfalls zu einem späteren Zeitpunkt mit einem eigenen Projekt angegangen werden, wenn auch klar ist, welche Anwendungsfälle überhaupt geregelt werden müssen. Denn unter KI-Systeme fallen die unterschiedlichsten Konstruktionen mit uneinheitlichen Risikoprofilen. Dabei können auch die Entwicklungen beim

Neu: Verzeichnis "AES"

III. Das Gesetz über die Organisation des Regierungsrates und der kantonalen Verwaltung vom 6. Juni 2005 wird wie folgt geändert:

Datenkatalog und Verzeichnis algorithmischer Entscheidsysteme

§ 44 a. [1] Die kantonale Verwaltung führt

a. einen gemeinsamen Datenkatalog, in dem Systeme und Register geführt werden, die Identifikatoren und Merkmale enthalten und

b. ein Verzeichnis der verwendeten algorithmischen Entscheidsysteme.

[2] Der Regierungsrat erlässt Ausführungsbestimmungen und ernennt die für die Führung des Datenkatalogs und des Verzeichnisses der algorithmischen Entscheidsysteme zuständige Stelle.

Automatisierte Einzelentscheide

Art. 21 Informationspflicht bei einer automatisierten Einzelentscheidung

[1] Der Verantwortliche informiert die betroffene Person über eine Entscheidung, die ausschliesslich auf einer automatisierten Bearbeitung beruht und die für sie mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt (automatisierte Einzelentscheidung).

[2] Er gibt der betroffenen Person auf Antrag die Möglichkeit, ihren Standpunkt darzulegen. Die betroffene Person kann verlangen, dass die automatisierte Einzelentscheidung von einer natürlichen Person überprüft wird.

[3] Die Absätze 1 und 2 gelten nicht, wenn:

a. die automatisierte Einzelentscheidung in unmittelbarem Zusammenhang mit dem Abschluss oder der Abwicklung eines Vertrags zwischen dem Verantwortlichen und der betroffenen Person steht und ihrem Begehren stattgegeben wird; oder

b. die betroffene Person ausdrücklich eingewilligt hat, dass die Entscheidung automatisiert erfolgt.

[4] Ergeht die automatisierte Einzelentscheidung durch ein Bundesorgan, so muss es die Entscheidung entsprechend kennzeichnen. Absatz 2 ist nicht anwendbar, wenn die betroffene Person nach Artikel 30 Absatz 2 des Verwaltungsverfahrensgesetzes vom 20. Dezember 1968[6] (VwVG) oder nach einem anderen Bundesgesetz vor dem Entscheid nicht angehört werden muss.

Revidiertes DSG des Bundes

Vollautomatische **Ermessensentscheide** (mit/ohne KI) → **Informationspflicht**

Anspruch auf **"menschliches Gehör"**, was bei Verfügungen das **Verfahren** auch abdeckt

Kantone verzichten meist auf Regelung

Lücke:

Einzelentscheide, die nicht in Verfügungen resultieren, sind aber oft auch ein Profiling

→ nächste Folie

"Keine Regelung nötig, wenn sichergestellt ist, dass bei automatisierten Einzelentscheidungen die betroffene Person informiert wird (z.B. durch Eröffnung der entsprechenden Verfügung) und sie die Möglichkeit hat, sich zur Einzelentscheidung zu äussern (z.B. aufgrund des Anspruchs auf rechtliches Gehör)." (Konferenz der Kantonsregierungen, Leitfaden)

Profiling

Profiling: jede Art der automatisierten Bearbeitung von Personendaten, die darin besteht, dass diese Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten, insbesondere um Aspekte bezüglich Arbeitsleistung, wirtschaftlicher Lage, Gesundheit, persönlicher Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder Ortswechsel dieser natürlichen Person zu analysieren oder vorherzusagen;

Revidiertes DSG des Bundes

Kurz:
Automatisierter Wertentscheid betreffend eine Person

KdK-Leitfaden: Manuell genügt (IMHO falsch)

BJ: "im Wesentlichen automatisiert" genügt

Beispiel: Welcher Bewerber ist für eine bestimmte Funktion am besten **geeignet**?

Kein Profiling ist die **objektive Feststellung** des Sachverhalts (Beispiel: Geleistete Arbeit)

Erhöhte Anforderungen an gesetzliche Grundlage

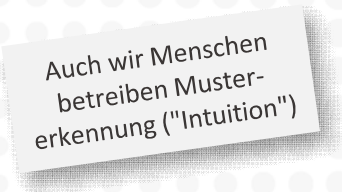
(z.B. Regelung in formellem Gesetz, hinreichend bestimmte Regelung, im Falle bloss indirekter Grundlage nur falls "unentbehrlich" oder mit einer Regelung in einer Verordnung)

Abgrenzen: Der Begriff des "Persönlichkeitsprofils" → meint das Ergebnis einer Datenbearbeitung, ob manuell oder automatisiert (im DSG gestrichen, in den Kantonen wird er zum Teil beibehalten werden)

Fazit: Damit sind viele (einzel-)**personenbezogene KI-Anwendungen erfasst** (andere basierend auf anonymisierten Daten oder zur Prognose des Verhaltens von *Gruppen* von Arbeitnehmern nicht)

Datenschutzgrundsätze beim KI-Einsatz

- Muss das Organ wissen, **wie die Black Box innen aussieht**?
  - Nein, aber es muss geprüft werden, ob das *Ergebnis* der Bearbeitung richtig und seine Verwendung vertretbar ist
- Grundsatz der **Datenrichtigkeit**
  - Achtung: Es geht i.d.R. um Wertentscheide, nicht Feststellungen
  - Ergebnis muss für den angestrebten Zweck tauglich sein
  - Kommt das Ergebnis in einem vernünftigen Verfahren zustande?
  - Richtigkeit im Einzelfall oder "nur" statistisch betrachtet?
  - Begründbarkeit eines Entscheids als KI-Stolperfalle, da Menschen Begründungen für ihre Entscheide immer nachschieben können
  - Verfahrensqualität auch anders begründbar (z.B. empirisch)



Datenschutzgrundsätze beim KI-Einsatz

- Grundsatz der **Verhältnismässigkeit**
  - Ist das Verfahren nötig und geeignet?
    - Standpunkt 1: Eine maschinelle Bewertung einer Person greift per se stärker in die Persönlichkeit ein
    - Standpunkt 2: Es kommt auf die Qualität des Ergebnisses an
  - Ist das Verfahren zumutbar bzw. verhältnismässig i.e.S.
    - Eine Frage der Menschenwürde, ob ein Mensch alleine von einer Maschine beurteilt werden darf → fallspezifisch
- Grundsatz der **Transparenz**
  - Information, (a) dass eine maschinelle Beurteilung erfolgt, (b) über die Grundzüge der Logik (z.B. trainierte KI), (c) welcher Input benutzt wird, und (d) was mit dem Ergebnis geschieht

Es werden derzeit Methoden entwickelt, um KI-Modelle z.B. durch "reverse engineering" erklärbar zu machen (Stichwort "machine learning explainability")

Datenschutzgrundsätze beim KI-Einsatz

- Verbot der **Diskriminierung** durch den Staat
  - Vor allem eine Frage der Datenrichtigkeit
  - Pro memoria: Menschliche Wertentscheide sind nicht unbedingt besser, aber die Anforderungen an sie sind de facto geringer
  - **Schritt 1:** Anforderung an die Prognosequalität?
    - Höher beim Entlassungs- als beim Einstellungsentscheid?
    - Welche menschliche "Kontrolle" bleibt im Prozess?
  - **Schritt 2:** Entspricht die KI diesen Anforderungen?
    - Red Flag 1: Datenmodell des Inputs ist atypisch
    - Red Flag 2: Output angesichts Datenlage unerklärbar

    } Kausalitätsproblem, zu lösen durch empirischen Nachweis der Richtigkeit

    - Red Flag 3: Training unklar; fehlende Qualitätskontrolle → Qualitätsmassnahmen, Prüfungen, Angaben zur Confidence, Erfahrung

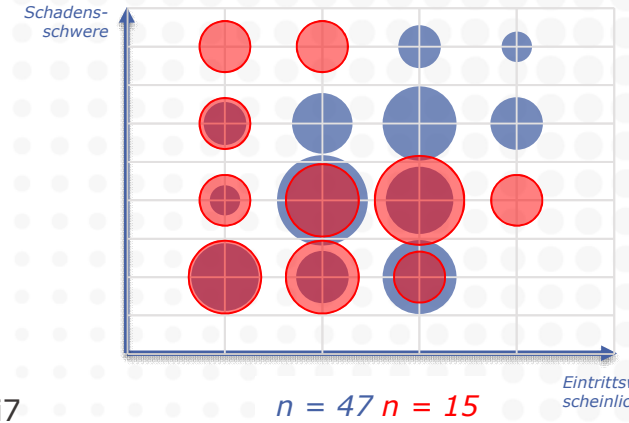
Abzubilden sind diese Risiken in einer sog. **Datenschutz-Folgenabschätzung** (DSFA)

Risiko für betroffene Personen?

Amazon scraps secret AI recruiting tool that showed bias against women

Wie Datenschutzfachleute es einschätzen (spontan v. strukturiert in der Gruppe)

"Eine Beraterfirma will ein Computerprogramm verwenden, das von jeder Stellenbewerbung eines Beraters einen "Erfolgsscore" aufgrund seines bisherigen Werdegangs und seiner Noten berechnet. Der Score basiert auf den bisherigen Erfahrungen des Unternehmens mit vergangenen Einstellungen. Der Score dient nur der Information des Hiring-Partners. Es wird niemand automatisch aussortiert."



Aufsatz zu Risikobeurteilungen im Datenschutz: https://bit.ly/3f1fNi7

Sonderprobleme beim Training einer KI

- Problem 1: **Eigeninteressen** der KI-Provider
  - Provider wollen Daten für das Training ihrer eigenen KI-Modelle
  - Fehlende gesetzliche Grundlage und Zweckentfremdung, wenn staatliche Daten zum Training privater Modelle benutzt werden
  - Lösung 1: Anonymisierung (aber oft nicht praktikabel)
  - Lösung 2: Organ erstellt eigene KI-Modelle (Statistikprivileg)
- Problem 2: **Datensicherheit**
  - Sind die Datentöpfe für das Training wirklich sicher?
- Problem 3: **Re-Identifikation** der **Trainingsdaten** der Modelle
  - Fehlen Vorkehrungen, sind KI-Modelle nicht unbedingt anonym

KI-Anwendung im öffentlichen Bereich?

- Es geht um Wertentscheide, nicht Sachverhaltsfeststellungen
- Bedingung der **Richtigkeit** und **Eignung** im Prinzip erfüllbar
  - Faktisch höhere Anforderungen an ein KI als an den Menschen
  - Qualität mancher KI-Lösungen noch nicht hinreichend belegt
- Bedingung der **Würde**, **Begründ- und Nachvollziehbarkeit**
  - Lösung durch tw. Verzicht auf vollautomatisierte Entscheide
- Bedingung der **gesetzlichen Grundlage** für ein Profiling
  - Betroffen ist jede computergestützte Bewertung, nicht nur KI
  - Wo ist ein Profiling z.B. aufgrund der Masse an Input-Daten wirklich nötig (Beispiele IT-Sicherheit, Missbrauchserkennung)?

Zielvorgaben?
Qualität Algorithmen?
Qualität Trainingsdaten?
Qualitätssicherung?

Einstellung zum Einsatz von KI wird sich **über die Jahre ändern** (eine Frage der Gewohnheit)

VISCHER

Danke für Ihre Aufmerksamkeit!

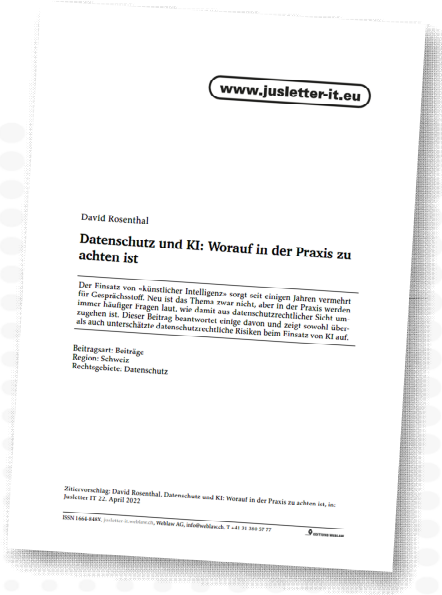
Fragen: drosenthal@vischer.com

Zürich
Schützengasse 1
Postfach
8021 Zürich, Schweiz
T +41 58 211 34 00

www.vischer.com

Basel
Aeschenvorstadt 4
Postfach
4010 Basel, Schweiz
T +41 58 211 33 00

Genf
Rue du Cloître 2-4
Postfach
1211 Genf 3, Schweiz
T +41 58 211 35 00



Mit diversen Quellennachweisen

https://www.rosenthal.ch/downloads/Rosenthal-KI-Datenschutz.pdf