



Algorithmische Entscheidungen: Transparenz und Kontrolle

Prof. Dr. K.A. Zweig
TU Kaiserslautern
Algorithm
Accountability Lab
@nettwkerin



Die zwei Ängste

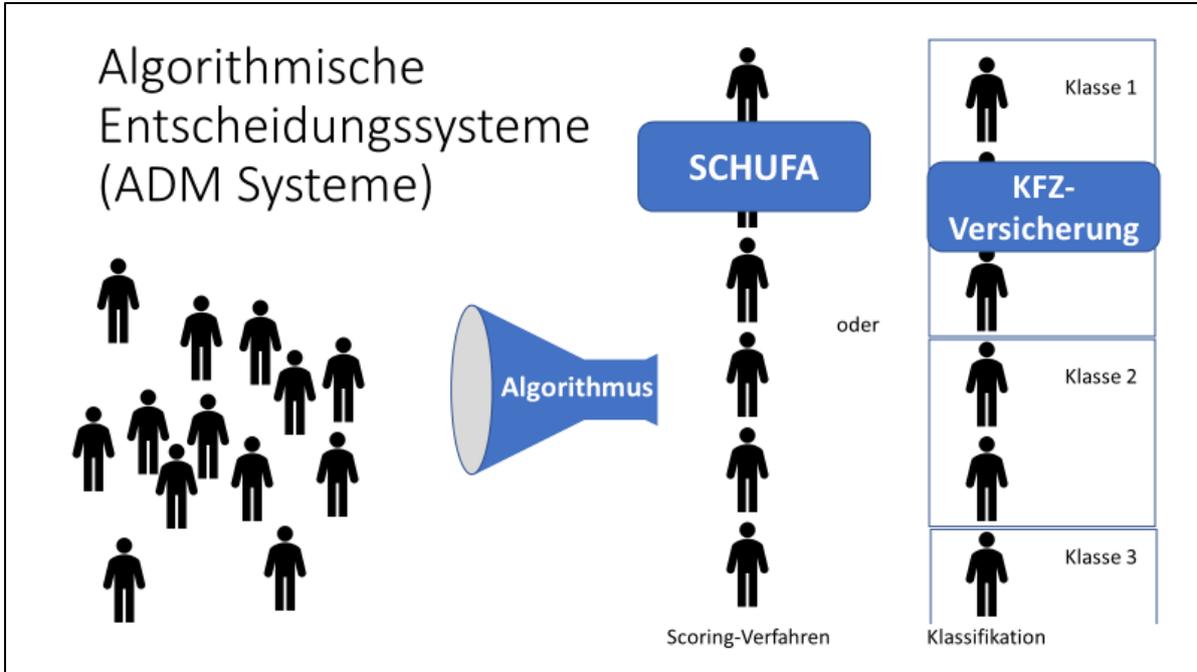
Sie wird dichten

Sie wird richten



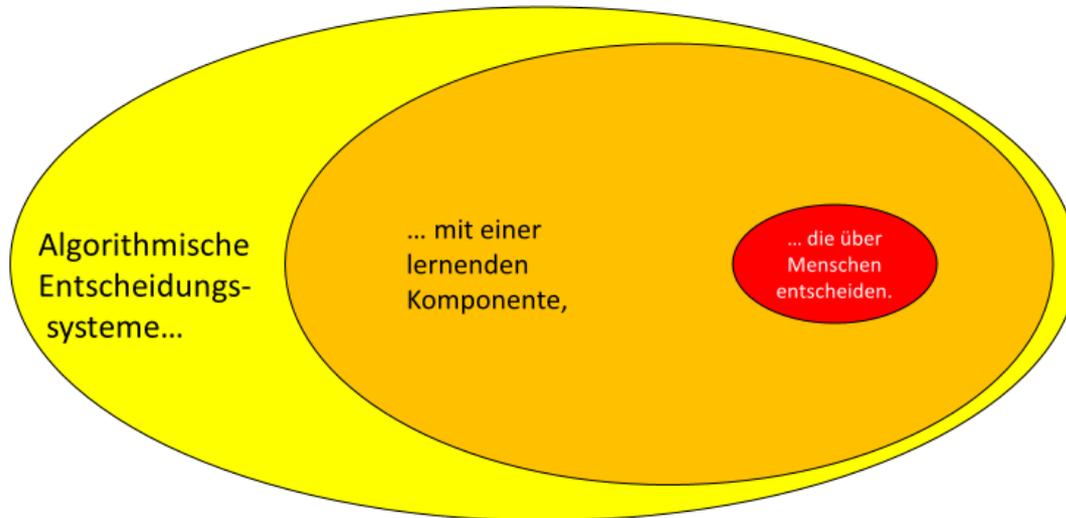
Im Umgang mit KI gibt es zwei Sorgen in der Bevölkerung:

- 1) Dass sie uns ersetzen wird – auch in den menschlichsten aller Tätigkeiten (beispielhaft: "Sie wird dichten")
- 2) Dass sie über uns entscheiden wird (beispielhaft: "Sie wird richten").

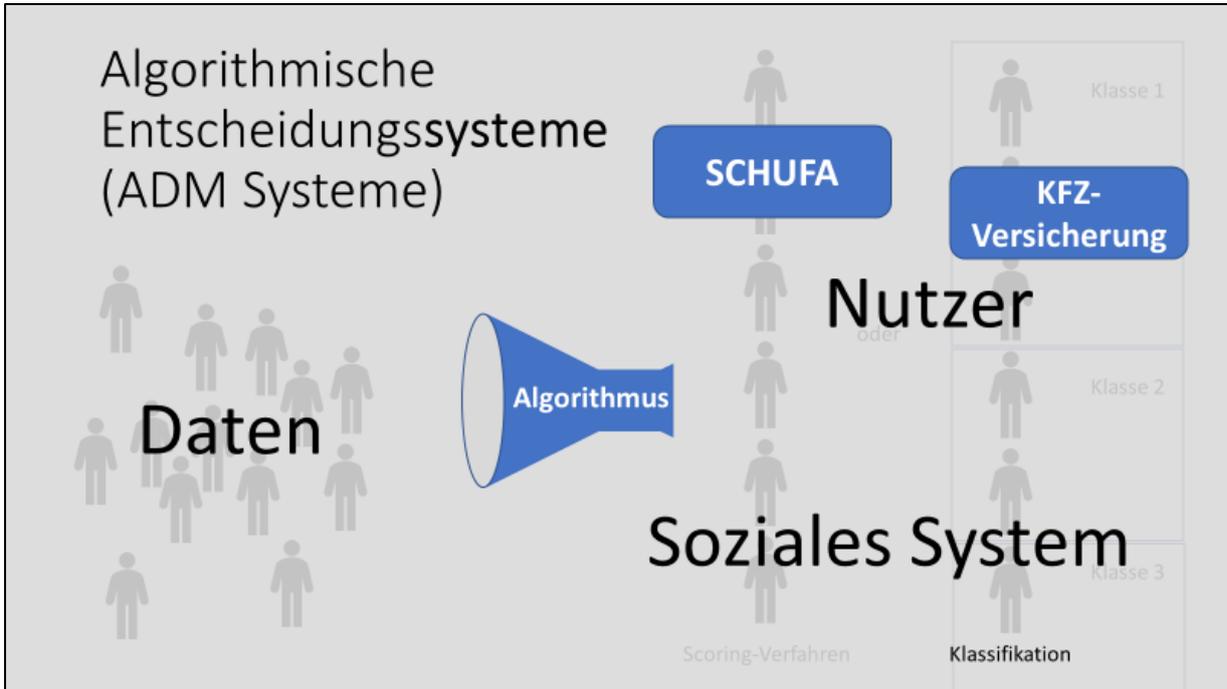


Algorithmische Entscheidungssysteme bekommen einen **Input** – z.B. Daten von Personen. Damit werden die Personen bewertet (bspw. in einem Scoring-Verfahren) oder klassifiziert. Beides passiert heute schon (z.B. bei Schufa und KFZ-Versicherungen), jedoch ist hier sowohl der gesellschaftlich Nutzen klar als auch die Kriterien, nach denen entschieden wird.

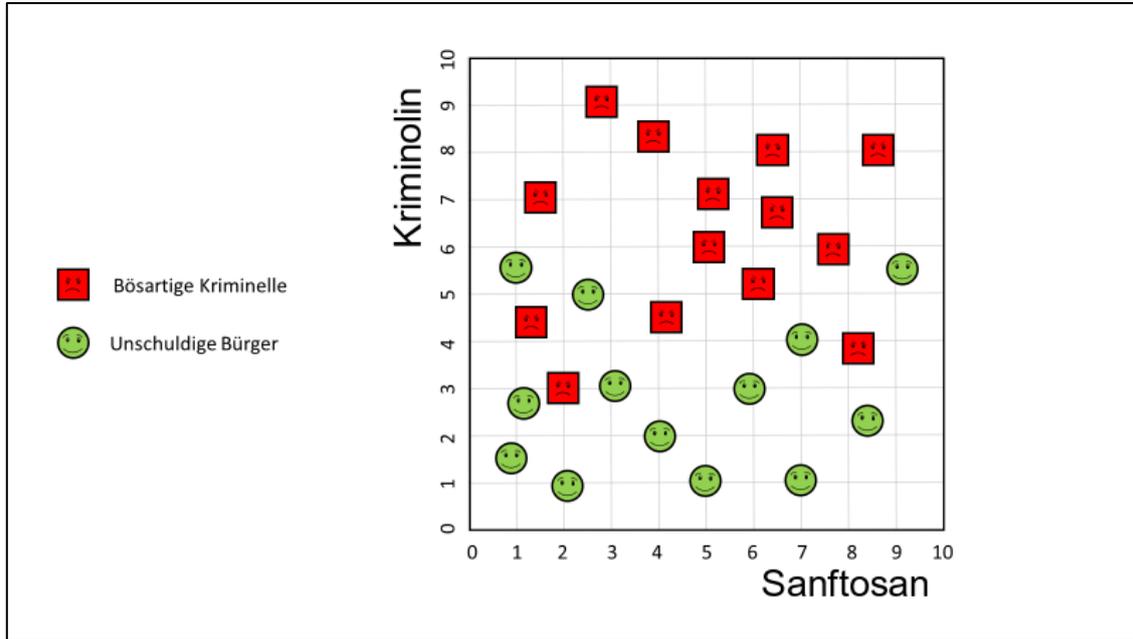
Welche ADM-Systeme sind problematisch?



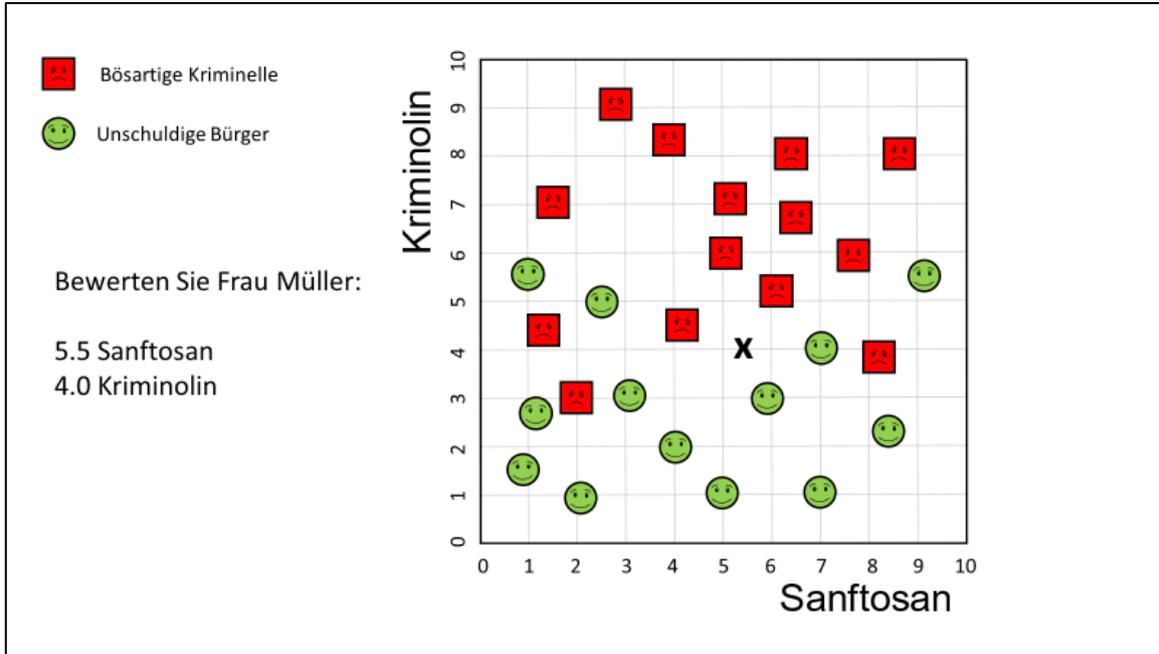
Im Wesentlichen sind nur solche ADM-Systeme problematisch, die mit Hilfe einer aus Daten lernenden Komponente über Menschen entscheiden.



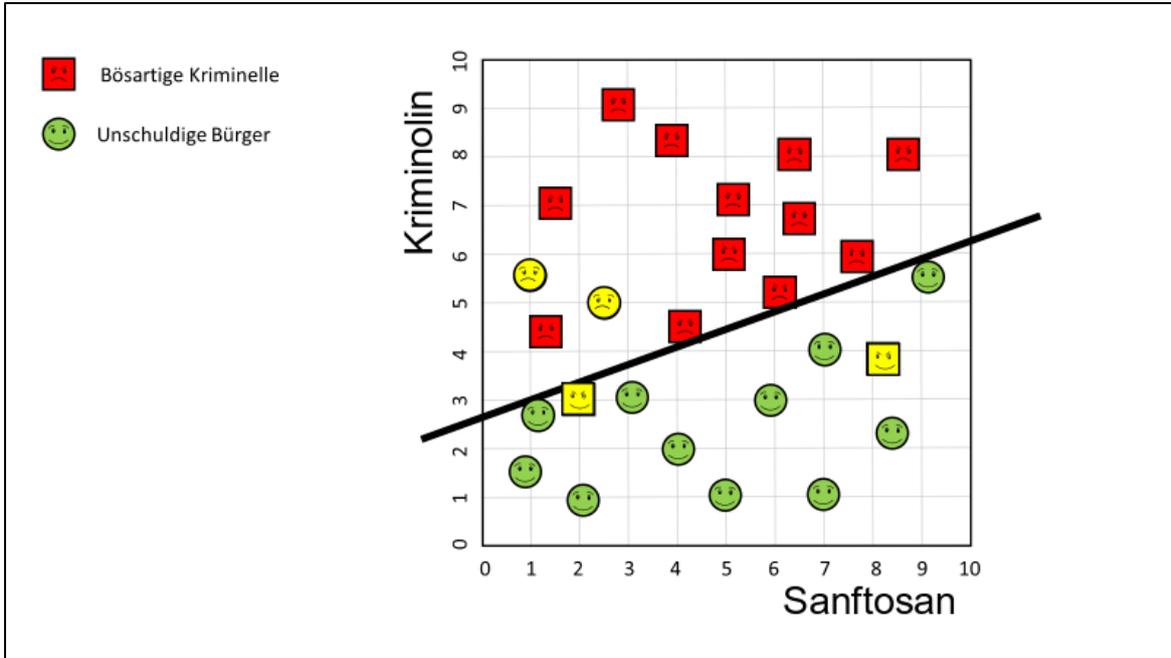
Wir sprechen von algorithmischen Entscheidungssystemen, weil neben dem Algorithmus, der zum Lernen genutzt wird, auch die Daten, die Nutzer und die vom System Bewerteten betrachtet werden. Nicht zuletzt geht es auch um die genaue soziale Fragestellung, die vom System entschieden werden soll. Wir sprechen vom "sozio-informatischen Gesamtsystem".



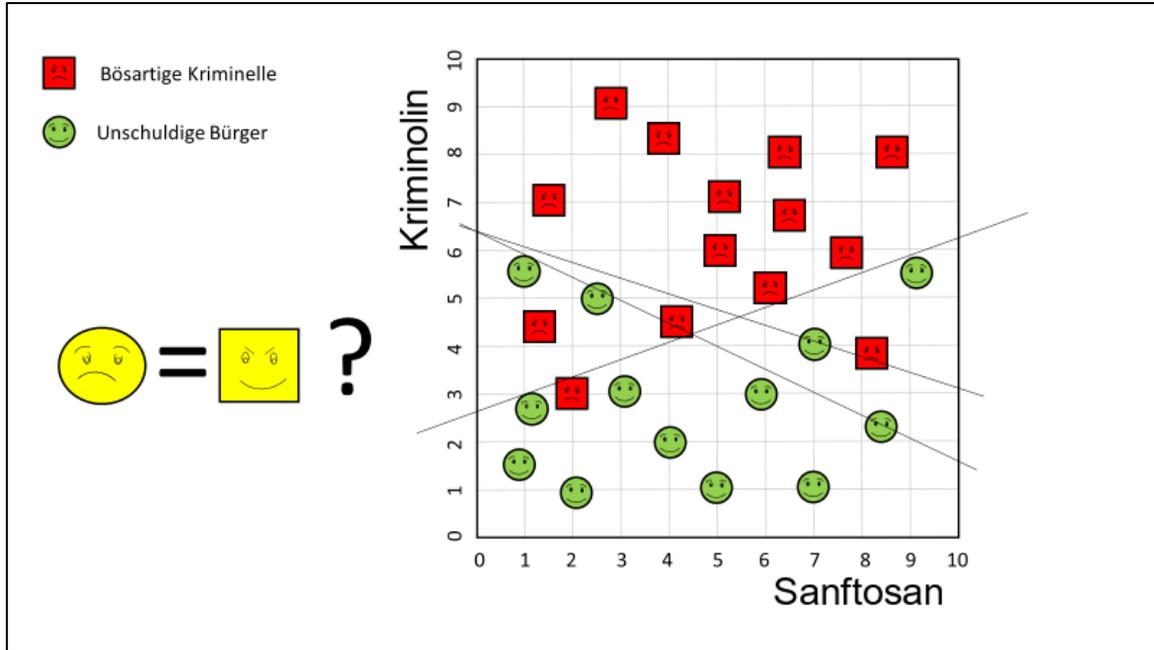
Wie lernen Computer von Daten? Stellen wir uns folgende Situation vor: Eine Biologin entdeckt, dass zwei Hormone im Blut im Wesentlichen vorhersagen, wer kriminell wird: "Sanftosan" und "Kriminolin". Hier sehen wir die Blutgehalte dieser Hormone von verurteilten Kriminellen und unschuldigen Bürgern. Zeichnen Sie eine gerade Linie durch die Datenpunkte, so dass möglichst alle grünen auf der einen Seite und alle roten auf der anderen Seite sind. Ein ähnliches Verfahren wird im maschinellen Lernen "Support Vector Machine" genannt (SVM).



Bewerten Sie nun Frau Müller, mit einem Sanftosan-Wert von 5.5 und einem Kriminolin-Wert von 4.0 (schwarzes Kreuz). Liegt Sie bei Ihnen auf der "grünen" oder der "roten" Seite Ihrer Linie?

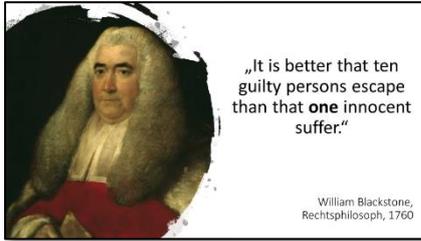


Eine mögliche Linie ist hier gezeigt – wir machen aber insgesamt vier Fehltritte mit dieser Linie: Zwei unschuldige Personen sind oberhalb der Linie und zwei weitere unten (gelbe Smiley).

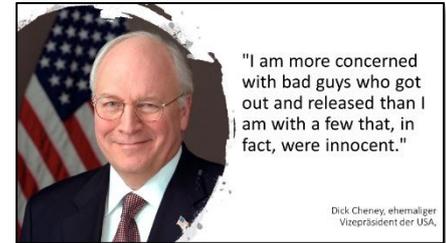


Unter der Maßgabe, dass alle Fehlurteile gleich schlecht sind, gibt es noch mindestens zwei weitere Linien, bei denen insgesamt vier Fehlurteile gemacht werden. Weniger als vier Fehlurteile sind bei diesem Datensatz nicht möglich.

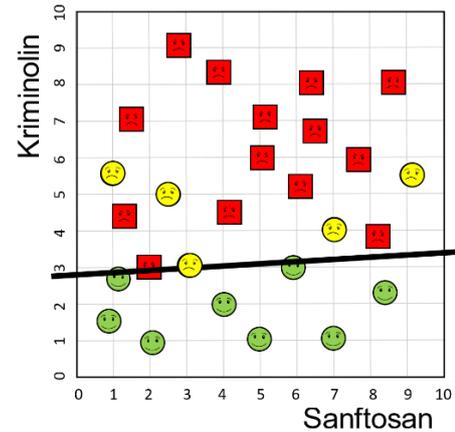
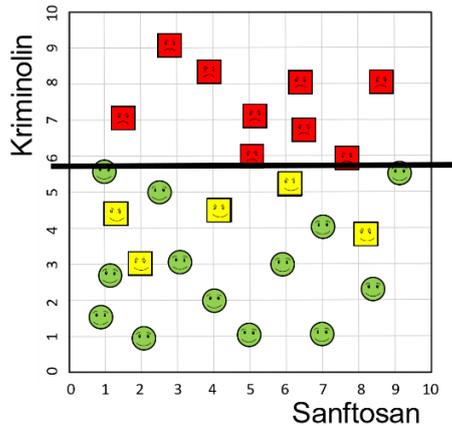
Über das richtige Verhältnis der beiden Fehlurteile lässt sich streiten:

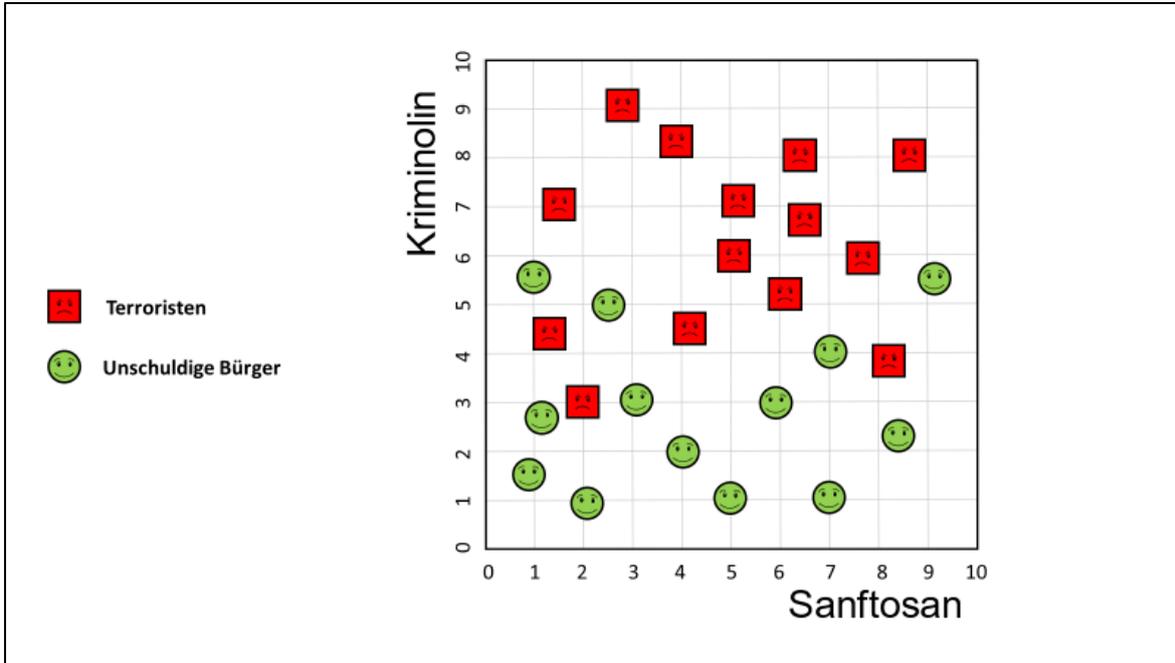


Blackstone will lieber 10 Kriminelle laufen lassen als eine unschuldige Person im Gefängnis zu sehen, Cheney sieht es eher andersherum.

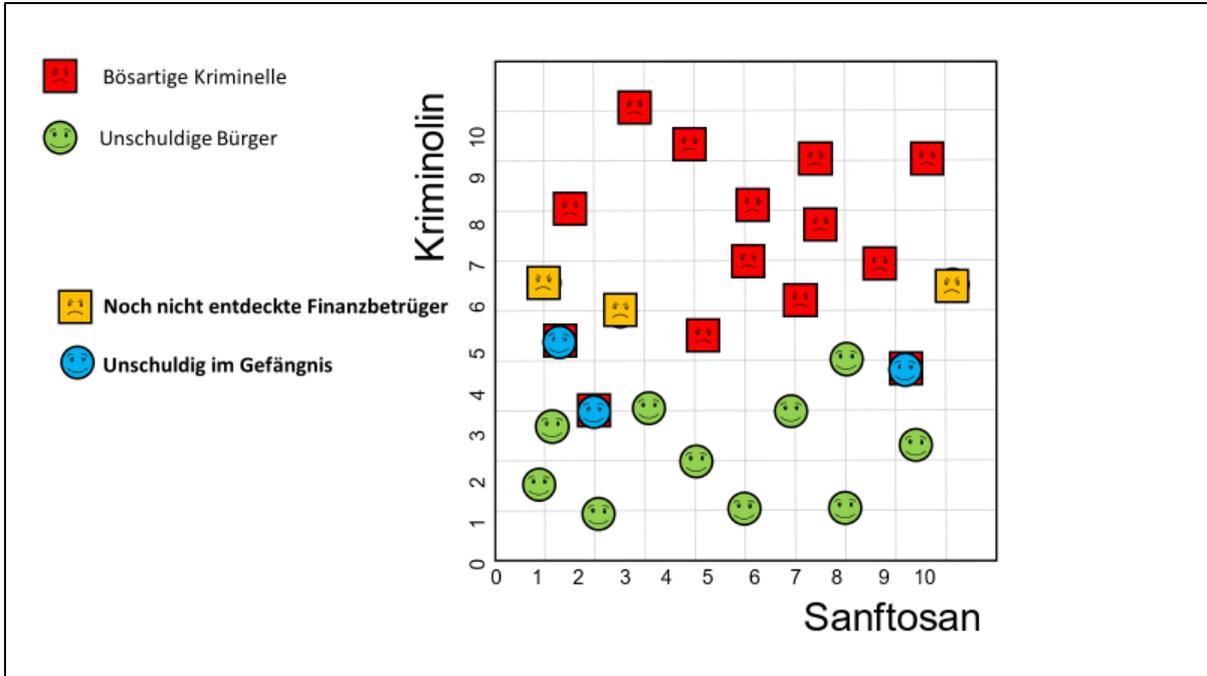


Das führt dann auch zu drastisch unterschiedlichen Linien:





Vielleicht ändert sich ja auch Ihre Meinung zum Verhältnis der Fehlurteile, wenn es nicht nur um irgendwelche Kriminellen sondern um Terroristen geht? Von der Technik her ist es derselbe Algorithmus, aber man sieht nun leicht, dass es individuelle und situative Unterschiede in der Bewertung von "guten Entscheidungen" gibt.



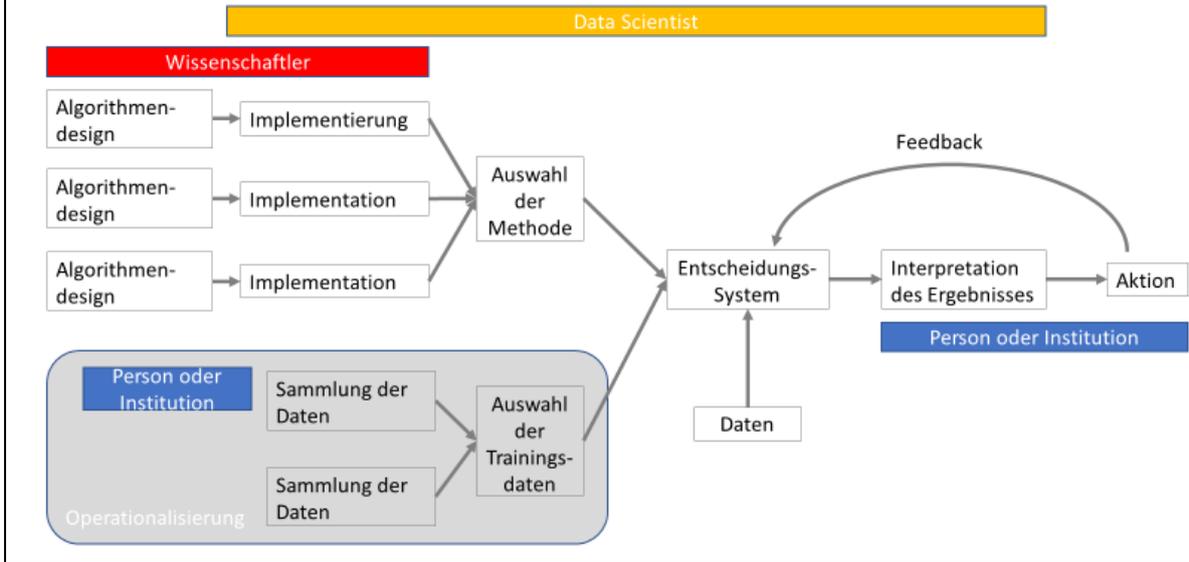
In vielen Fällen sind auch die Datengrundlagen nicht gut genug: Lernt man nur von verurteilten Kriminellen übersieht man solche, die noch nicht entdeckt wurden (orange) und nimmt solche mit auf, die unschuldig im Gefängnis sitzen (blau).

Wie gut sind die Robo-Richter?

- Ganz schön schlecht: COMPAS
 - Hochrisiko-Kategorie:
 - Gewöhnliche Kriminaltaten: nur zu 50% richtig!
 - Schwere Straftaten: nur zu 20% richtig!
- Ein amerikanisches Terroristenidentifikationssystem tönt:
 - „Nur 0.008% falsch Positive!“
 - Bei 55 Millionen Einwohner sind das 4.400 Unschuldige, um wenige Hundert zu identifizieren.
 - Von den „Hochrisikopersonen“ also vermutlich unter 20%!
- Im medizinischen Bereich teilweise besser als Doktoren!



Lange Kette der Verantwortlichkeiten



An der Entwicklung eines ADM-Systems sind so viele Personen beteiligt und es werden so viele Entscheidungen getroffen, dass es schwierig (aber nicht unmöglich) ist, den Überblick zu behalten.

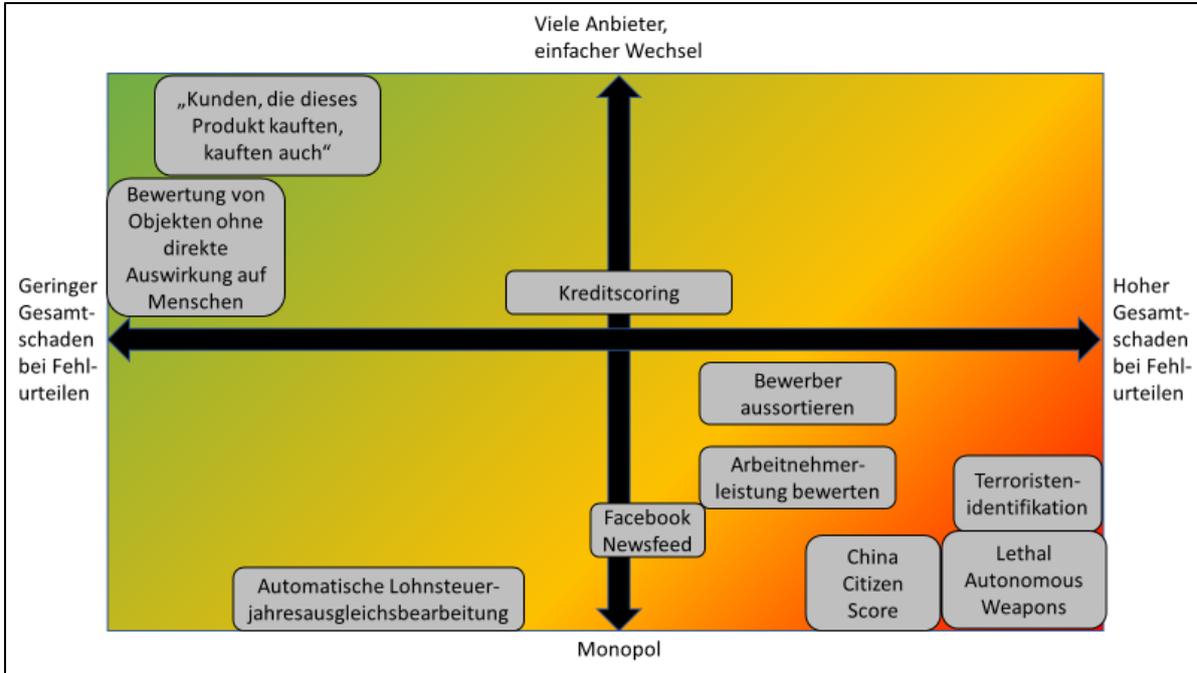
Wie bewerten bezüglich der Regulierungsnotwendigkeit?

1. Schadenstiefe

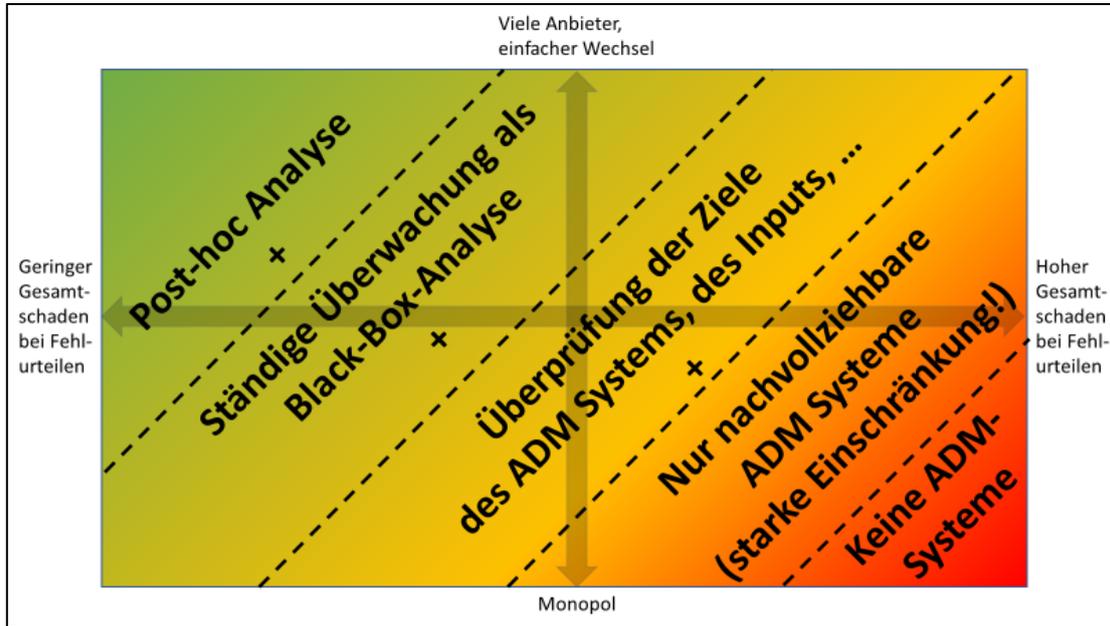
$$\sum \text{Schaden für Individuum(Fehlurteil)} \\ + \text{Schaden für Gesellschaft(Fehlurteil)}$$

2. Anbietervielzahl und Wechselmöglichkeit

Wir glauben, dass zwei Dimensionen im Wesentlichen determinieren, mit welcher Regulierungstiefe vorgegangen werden sollte: Wie hoch der mögliche Gesamtschaden von Fehlurteilen ist und die Möglichkeit des Wechsels zu einem anderen ADM-System.



Hier sind einige ADM-Systeme exemplarisch eingezeichnet.



In Abhängigkeit von der Platzierung eignen sich unterschiedliche Tiefen der Überprüfung. Wo es oben links ausreicht, nach Verdachtsfällen im Nachhinein zu analysieren, brauchen andere Systeme eine ständige Überwachung (z.B. auf unerlaubte Diskriminierung). Die im gelben Bereich sollten zudem inhaltlich von Experten auf ihr Design überprüft werden. Im orangefarbenen Bereich sollten nur solche Systeme eingesetzt werden, die nachvollziehbar sind – dies stellt eine sehr starke Einschränkung dar. Im roten Bereich finden ADM-Systeme ihre Grenze.

Weiterführende Referenzen

1. Studie für die Bertelsmann-Stiftung:
Zweig, Fischer & Lischka: „[Wo Maschinen irren können](#)“ (Serie AlgoEthik, No. 4, 2018)
2. [Zwei Kapitel im Band \(Un\)Berechenbar?](#) des Fraunhofer FOKUS, Kompetenzzentrum ÖFIT, 2018
 1. Zweig & Krafft: „[Fairness und Qualität algorithmischer Entscheidungen](#)“
 2. Krafft & Zweig: „[Wie Gesellschaft algorithmischen Entscheidungen auf den Zahn fühlen kann](#)“
3. Studie für die Konrad-Adenauer-Stiftung „Algorithmische Entscheidungen: Transparenz und Kontrolle“ (Zweig, erscheint 2018)
4. Studie vom Fraunhofer FOKUS, Kompetenzzentrum Öffentliche IT (ÖFIT): Opiela, Mohabbat Kar, Thapa & Weber: [Exekutive KI 2030 – Vier Zukunftsszenarien für Künstliche Intelligenz in der öffentlichen Verwaltung](#), 2018)

Alle Bilder gemeinfrei von der Wikipedia oder von Pixabay unter [CC0-Lizenz](#) (mit Dank!) oder eigene Werke .

Kontakt:

Prof. Dr. Katharina A. Zweig
zweig@cs.uni-kl.de
twitter: @nettwwerkerin