verbraucherzentrale
Bundesverband

Tobias D. Krafft I Prof. Dr. Katharina A. Zweig

TRANSPARENZ UND NACHVOLLZIEHBARKEIT ALGORITHMENBASIERTER ENTSCHEIDUNGSPROZESSE

Ein Regulierungsvorschlag aus sozioinformatischer Perspektive

22. Januar 2019

Gefördert durch:

Bundesministerium der Justiz und für Verbraucherschutz

aufgrund eines Beschlusses des Deutschen Bundestages

Impressum

Verbraucherzentrale
Bundesverband e.V.

Team Digitales und Medien

Rudi-Dutschke-Straße 17
10969 Berlin
digitales@vzbv.de

Bundesverband der Verbraucherzentralen und Verbraucherverbände

Verbraucherzentrale Bundesverband e.V.

INHALT

I. EINLEITUNG

Immer öfter werden Verbraucher automatisch oder teilautomatisch aufgrund von maschinellen Bewertungsmechanismen (algorithmischen Entscheidungssystemen) sortiert oder kategorisiert. Aus der immer größeren Anzahl dieser algorithmischen Entscheidungssysteme (Algorithm Decision Making Systems), kurz ADM-Systeme, erwächst der Wunsch nach mehr Transparenz und Möglichkeiten stärkerer Einsichtnahme, um Fehlurteile zu erkennen und zu vermeiden. Denn nur mit einer angemessenen Kenntnis der Risikopotenziale, können wir die möglichen Gefahren auf ein Minimum beschränken und damit den größtmöglichen gesellschaftlichen Nutzen aus diesen Systemen ziehen.

Natürlich haben Firmen schon immer Entscheidungen über ihre Kunden getroffen, z. B. haben sie Kunden von ihrem Service ausgeschlossen, wenn diese zu oft ihre Bestellungen zurückschickten[1]. Im Zuge der Digitalisierung[2] haben sich aber die Entscheidungsstrukturen in den verschiedenen gesellschaftlichen Bereichen grundlegend geändert, was sich am Beispiel der veränderten Situation des Verbrauchers auf dem digitalen Markt schnell verdeutlichen lässt: Während sich noch in den 1990er Jahren Käufer und Verkäufer auf ähnlichem Informationsniveau begegnet sind, so entstehen aktuell auf dem digitalen Markt völlig neue Verkaufsstrukturen, die ein Ungleichgewicht zwischen den Vertragspartnern begünstigen. Algorithmen übernehmen vermehrt die Aufgaben des menschlichen Entscheiders, wobei den Unternehmen mittlerweile immense Datenbestände (Big Data) zur Verfügung stehen, um die Kaufentscheidungen potenzieller Käufer vorherzusagen und sogar zu beeinflussen. VerbraucherInnen sehen sich dabei mit für sie kaum zu durchschauenden Prozessen konfrontiert und fordern einen transparenten und kontrollierten Umgang mit diesen, um die Ergebnisse nachvollziehen zu können.

Dabei können diese algorithmischen Entscheidungssysteme existenzielle Auswirkungen auf das Leben eines Individuums haben, zum Beispiel wenn sie die Kreditwürdigkeit einer Person bewerten, sie in Versicherungen klassifizieren oder Nachrichten personalisiert zusammenstellen. Demzufolge wird im vorliegenden Gutachten der Fokus auf solche Systeme gelegt, die entweder zur eigentlichen Entscheidungsfindung bei der Beurteilung oder Kategorisierung von Verbraucher herangezogen werden oder vollautomatisiert Entscheidungen über Verbraucher treffen.

Auch die Politik nimmt sich im Bereich der Digitalisierung inzwischen der Problematik algorithmischer Entscheidungssysteme an. Im Rahmen ihres Koalitionsvertrages richteten die Regierungsparteien CDU, CSU und SPD den politischen Fokus auf den Verbraucherschutz im Kontext der digitalen Welt der Algorithmen. Sie setzen sich unter anderem „für Transparenz bei Online-Vergleichs- und Beratungsportalen ein“[3]. Explizit fordern sie

---

[1] Das war auch rechtens – so entschied das Landgericht Hamburg (Az. 416 O 129/03), dass der Versandhandel OTTO Kunden ausschließen darf, die zu oft die bestellten Waren wieder auf Kosten von OTTO zurückschickten. Bericht erschienen in FinanzTest 02/2004.

[2] Als Digitalisierung bezeichnen wir die Übertragung von analogen Prozessen auf den Computer.

[3] CDU/CSU/SPD: Ein neuer Aufbruch für Europa Eine neue Dynamik für Deutschland Ein neuer Zusammenhalt für unser Land Koalitionsvertrag zwischen CDU, CSU und SPD 19. Legislaturperiode, 2018, https://www.bundesregierung.de/resource/blob/975226/847984/5b8bc23590d4cb2892b31c987ad672b7/2018-03-14-koalitionsvertrag-data.pdf?download=1, 01.12.2018, , Zeile 2092 ff.

„Algorithmen- und KI-basierte Entscheidungen, Dienstleistungen und Produkte überprüfbar (zu) machen, insbesondere im Hinblick auf mögliche unzulässige Diskriminierungen, Benachteiligungen und Betrügereien“[4].

Des Weiteren ist der "Strategie Künstliche Intelligenz" der Bundesregierung zu entnehmen, dass sie aktiv „Forschung zu Verfahren zur Kontrolle und Nachvollziehbarkeit algorithmischer Prognose- und Entscheidungssysteme fördern möchten“[5]. Diese Forschungsinteressen werden im Strategiepapier mit einer positiven Bewertung der möglichen Potenziale auch und insbesondere in verschiedenen sensiblen Anwendungsmöglichkeiten begründet. So ist beispielsweise geplant, gezielt die Nutzbarkeit von KI-Systemen im Gesundheitssektor zu fördern[6]. Es wird deutlich, dass die von Gesellschaft und Politik zunehmend aufgegriffene Transparenzdebatte dem wachsenden Bedürfnis der Bevölkerung nach mehr Klarheit über algorithmenbasierte Entscheidungsprozesse Ausdruck verleiht.

Eine unreflektierte Überwachung sämtlicher ADM-Systeme stünde aber in keinem Verhältnis zum dafür erforderlichen Aufwand, was wiederum zu einem Innovationshemmnis führen würde. Ausschlaggebend für den Level an notwendiger Überwachung sollten hierfür die mit dem jeweiligen ADM-System verbundenen Folgen und Risiken für den Einzelnen und die Gesellschaft sein. Wie schon die Verbraucherzentrale Bundesverband (VZBV) in ihrem Thesenpapier „Algorithmenbasierte Entscheidungsprozesse“ festgestellt hat, ist hierbei die Angemessenheit von geforderten Maßnahmen zu überprüfen[7].

Die erforderlichen Transparenzpflichten und Kontrollprozesse unterscheiden sich sowohl in der Tiefe als auch darin, an welcher Stelle sie ansetzen, denn Fehler können sich auf den verschiedensten Ebenen einschleichen: Während der Konzeption des Systems, während seiner Implementierung oder während seiner Nutzung. Ein ADM-System kann beispielsweise auf falschen oder zu speziellen Daten trainiert worden sein, wie das folgende Beispiel des kanadischen Unternehmens Winterlight Labs veranschaulicht. Das Startup vertreibt einen medizinischen Hörtest zur Früherkennung von neurologischen Krankheiten wie Alzheimer, Parkinson und Multiple Sklerose, der auf Grundlage von Untersuchungen mit erkrankten und gesunden Personen beruht. Auf diesen Daten wurde ein statistisches Modell trainiert, das für weitere Personen dann jeweils eine Einschätzung berechnet, ob und wenn ja, an welcher Krankheit diese leiden könnten. Die Firma hatte 2016 erste Forschungsergebnisse in einem wissenschaftlichen Journal publiziert[8]; erst danach wurde folgende Problematik erkannt: Die Software wurde auf englischen Muttersprachlern mit einem speziellen Dialekt trainiert, von denen ein Teil mit Alzheimer diagnostiziert war. Letztere zeigten Wortfindungsschwierigkeiten, so dass die Maschine kleinere Denkpausen bei Antworten auf einfache Fragen mit der Krankheit assoziierte. Während diese Korrelation beim Originaldatenset gute Ergebnisse vorweisen konnte,

---

[4] Ebd.Zeilen 6354 ff.

[5] Bundesministerium für Wirtschaft und Energie, das Bundesministerium für Bildung und Forschung und das Bundesministerium für Arbeit und Soziales: Strategie Künstliche Intelligenz der Bundesregierung, 2018, https://www.bmwi.de/Redaktion/DE/Publikationen/Technologie/strategie-kuenstliche-intelligenz-der-bundesregierung.html, 01.12.2018, Seite 16

[6] Ebd. Seite 37

[7] Verbraucherzentrale Bundesverband: Thesenpapier Algorithmenbasierten Entscheidungsprozesse, 2017, https://www.vzbv.de/sites/default/files/downloads/2018/05/22/dm_17-12-07_vzbv_thesenpapier_algorithmen.pdf, 01.12.2018, Seite 11

[8] Fraser, K. C., Meltzer, J. A., & Rudzicz, F.: Linguistic features identify Alzheimer’s disease in narrative speech. Journal of Alzheimer's Disease 49.2 (2016): 407-422.

kam es später bei Menschen, die mit dem Englischen bzw. dem Dialekt nicht so vertraut waren, aus sprachlichen Gründen zu Verzögerungen bei der Antwort. Diese wurden daraufhin von der Software fälschlicherweise positiv eingestuft, sodass den Personen eine Störung attestiert wurde[9]. Die Auswirkungen solcher Fehldiagnosen auf Betroffene können immens sein – ein eingängiges Beispiel für die Notwendigkeit von Transparenzpflichten und Kontrollprozessen bei der Anwendung solcher ADM-Systeme.

Während bei medizinischen oder ähnlich relevanten Anwendungsgebieten die Überprüfung und Kontrolle algorithmischer Entscheidungssysteme zunehmend in den Fokus der gesellschaftlichen Debatte rücken, gibt es andere, weniger beachtete Bereiche, die zuerst harmloser erscheinen. So hat es bei Verbrauchern zunächst nur Verärgerung ausgelöst, als eine Unregelmäßigkeit der algorithmisch vergebenen Sitzplätze bei Fluggesellschaften aufgedeckt wurde. Eine Umfrage unter Passagieren zeigte, dass zum Beispiel Ryanair 35% der Kunden von anderen Personen ihrer Reisegruppe getrennt saßen, wenn sie nicht bereit waren, dafür extra zu bezahlen. Der allgemeine Durchschnitt liegt dagegen lediglich bei 18%[10].

Wie ist der Schaden einer solchen, algorithmischen Verteilung zu bewerten? Der nahezu doppelt so hohe Anteil bei Ryanair erregte zwar Unmut, allerdings ist der individuelle Schaden für den einzelnen Verbraucher schwer zu bemessen und eher gering. Die isolierte Schadensanalyse bezogen auf die betroffenen Individuen erwies sich jedoch als unzureichend. Erst als eine Studie der Royal Aeronautical Society (RAeS) - zu Deutsch „Königliche Luftfahrtgesellschaft“ – im Rahmen einer sozioinformatischen Analyse das System in seinem Gesamtkontext untersuchte, konnte das tatsächliche Schadenspotenzial getrennt sitzender Familien aufgezeigt werden[11]. So fand man heraus, dass diese bei einer Evakuierung eines Flugzeugs zueinander strömen und somit unter Umständen entgegen der Notfallwege laufen, was zu einer starken Verzögerung bei der Entfluchtung führen kann. Wir nennen eine solche Einbeziehung des Kontextes und die Analyse der Interaktion von algorithmischer Entscheidung und betroffenem sozialen System eine *sozioinformatischen Gesamtanalyse*. Erst damit kann das Schadenspotenzial in einem vorher noch nicht für wesentlich gehaltenen Bereich ausgemacht werden. Diese neue Bewertung begründet die Forderung nach einer dauerhaften Überwachung dieses algorithmenbasierten Entscheidungssystems.

Diesen offensichtlich zu kontrollierenden ADM-Systemen steht eine Vielzahl von Systemen gegenüber, die uns in unserer täglichen Arbeit unterstützen, die Freizeit interessanter gestalten oder uns spezifische Produkte vorschlagen, die genau das treffen, was wir gesucht haben. So kann bei einem ADM-System, welches das bisherige Kaufverhalten analysiert und zu dem Schluss kommt, dass eine Person gerne Kostüme mit auffälligen Mustern trägt, kaum von einem Schadenspotenzial bei einer Fehlkategorisierung gesprochen werden. Wenn eine Kundin sich hier verkannt fühlt, bemerkt sie das nach wenigen fehlgeleiteten Vorschlägen und kann zu einem anderen Anbieter wechseln. Hier genau dieselben Transparenzpflichten zu fordern und die Entscheidungen genauso stark zu

---

[9] Dave Gershgorn: If AI is going to be the world’s doctor, it needs better textbooks, 2018, https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks/, 01.12.2018

[10] Civil Aviation Authority: Civil Aviation Authority launches review of airlines’ allocated seating policies, 2018, https://www.aerosociety.com/media/8534/emergency-evacuation-of-commercial-passenger-aeroplanes-paper.pdf, 01.12.2018

[11] Royal Aeronautical Society: Emergency Evacuation of Commercial Passenger Aeroplanes, 2018, https://www.aerosociety.com/media/8534/emergency-evacuation-of-commercial-passenger-aeroplanes-paper.pdf, 01.12.2018

kontrollieren wie im sensiblen Gesundheitsbereich oder in der Flugsicherheit, ist offensichtlich überflüssig und schädlich aufgrund der dadurch gebundenen Ressourcen im Staat und in den anbietenden Firmen.

Aus den bisherigen Ausführungen lässt sich daher entnehmen, dass eine differenzierte Auseinandersetzung mit der Frage nach den jeweils angemessenen Transparenzpflichten und Kontrollprozessen notwendig ist, um den jeweils erforderlichen Nachvollziehbarkeitslevel zu erreichen. Um die Voraussetzungen hierfür zu schaffen, werden im folgenden Kapitel zunächst die grundlegenden Begriffe *Verbraucher* und *algorithmisches Entscheidungssystem* definiert. Am Schluss des Kapitels wird die *sozioinformatische Gesamtanalysen* erklärt.

Anschließend werden die Forderungen nach „Transparenz“ und „Nachvollziehbarkeit“ algorithmischer Entscheidungssysteme sowohl begrifflich geklärt, als auch in ihren verschiedenen Interpretationen vorgestellt (Kapitel II).

Das Kernstück dieses Gutachtens besteht in der Vorstellung einer neuen Methode, um algorithmischen Entscheidungssysteme in ihrer spezifischen Einbettung in einen sozialen Prozess entsprechend ihrer jeweiligen Risikopotenziale zu bewerten und zu kategorisieren (Kapitel IV).

Darauf aufbauend kann dann die notwendige Regulierungstiefe auf den Ebenen der Transparenz und Nachvollziehbarkeit bestimmt werden. Hierbei werden die verschiedenen Forderungen und Herangehensweisen zur Kontrolle und Überprüfung von ADM-Systemen erläutert und so gezeigt welche Art von Nachvollziehbarkeit und Transparenz für diese jeweils gefordert werden müssen. In dem Zusammenhang schlagen wir vor, was Betreiber von ADM-Systemen bereitstellen müssen, um die jeweils notwendige Transparenz und Nachvollziehbarkeit bezüglich des zuvor bestimmten Risikolevels zu gewährleisten (Kapitel V).

Das letzte Kapitel befasst sich abschließend mit konkreten Handlungsempfehlungen (Kapitel VI).

II. BEGRIFFSKLÄRUNGEN

Dieses Kapitel dient der Klärung zentraler Begriffe und wird der eigentlichen Studie vorangestellt.

Im Mittelpunkt der vorliegenden Studie steht der *Verbraucher*, der auf irgendeine Weise von den hier besprochenen algorithmischen Entscheidungssystemen betroffen ist und den es vor potenziellen negativen Folgen und Auswirkungen zu schützen gilt. Aus dem Grund beginnt das Kapitel mit der Definition dieses Begriffs, wobei anzumerken ist, dass es im Digitalen nicht ganz einfach ist, den *Verbraucher* vom *Nutzer* zu trennen oder diesen von anderen schützenswerten Gruppen wie Arbeitnehmern und Bürgern zu differenzieren.

Es ist danach notwendig, den Begriff des *algorithmisches Entscheidungssystem* zu bestimmen. Ein solches algorithmisches Entscheidungssystem ist jedoch auch immer in den übergeordneten Kontext eines sozioinformatischen Systems integriert, das einen bestimmten Zweck verfolgt. Erst die Identifikation dieser Einbettung ermöglicht es, mögliche Risiken zu bestimmen, die mit der Nutzung eines solchen Systems verbunden sind. Der für diese Studie ebenfalls zentrale Begriffe des *sozioinformatischen Systems* und der *sozioinformatischen Gesamtanalyse* schließt das Kapitel ab.

1. VERBRAUCHER

Der Begriff *Verbraucher* ist gesetzlich klar geregelt: Jede Person, die einen Vertrag mit einem Unternehmen schließt, um eine überwiegend zu privaten Zwecken genutzte Ware oder Dienstleistung zu erwerben, ist ein/e VerbraucherIn[12] [13]. Nun ist es im digitalen Zeitalter nicht immer so, dass eine solche Dienstleistung im unmittelbaren Tausch gegen Geld erworben wird – viele Dienstleistungen im Internet sind auf den ersten Blick kostenfrei. Dazu gehören die Nutzung von bekannten Suchmaschinen oder sozialen Netzwerkplattformen ebenso wie Produktempfehlungen auf großen digitalen Marktplätzen, sodass man in diesen Fällen auch von *Nutzern* dieser Dienste spricht. Im Rahmen dieser Studie sind sie aber klar auch als Verbraucher und Verbraucherinnen definiert, denn jede Nutzung der genannten Dienstleistungen schließt einen Vertrag mit den jeweiligen Dienstanbietern ein, die in den allermeisten Fällen Unternehmen sind.

Oftmals ist mit der Verwendung dieser Dienste auch die Zustimmung für die Weiterverarbeitung der dabei anfallenden Daten verbunden. Dabei können algorithmische Entscheidungssysteme genutzt werden, um Nutzerinnen und Nutzer zu bewerten. Basierend auf der Bewertung können im Wesentlichen drei verschiedene Fälle auftreten:

1) Es kann damit der Zugang zu einer Dienstleistung gewährt oder verwehrt werden. Dazu gehören beispielsweise die Möglichkeiten, ein Produkt per Rechnung zu bestellen, die Aufnahme als Versicherungsteilnehmer oder die Buchung eines Flugtickets.

---

[12] Bundesministerium für Justiz und Verbraucherschutz: Meine Rechte bei Verbraucherverträgen, 2018, https://www.bmjv.de/DE/Verbraucherportal/KonsumImAlltag/Verbrauchervertraege/Verbrauchervertraege_node.html, 01.12.2018

[13] Bürgerliches Gesetzbuch (BGB) § 13 Verbraucher

2) Es können die Dienstleistungen selbst verändert werden. Dazu gehören alle Formen der Personalisierung von digitalen Diensten, wie beispielsweise die personalisierten Timelines bei Twitter und Facebook und alle Arten von personalisierten Produktempfehlungen.

3) Die Daten können verwendet werden, um den Nutzer oder die Nutzerin für Werbetreibende zu klassifizieren, um damit den Dienst zu finanzieren. Dies ist klassischerweise bei allen "kostenlosen" Webdiensten der Fall, so zum Beispiel bei Suchmaschinen, sozialen Netzwerken oder Fotodatenbanken.

In dieser Studie konzentrieren wir uns auf algorithmische Entscheidungssysteme mit einer lernenden Komponente, wie im folgenden Abschnitt definiert wird.

2. ALGORITHMISCHE ENTSCHEIDUNGSSYSTEME (ADM-SYSTEME)

Algorithmische Entscheidungssysteme (Algorithmic decision making systems, ADM-Systeme) bekommen Informationen über Personen und deren Verhalten und benutzen eine klar definierte Handlungsanweisung (einen *Algorithmus*), um aus dieser Information eine einzige Zahl zu erzeugen. Diese Zahl ist die eigentliche Entscheidung. Dabei kann die Zahl beispielsweise einem Risikowert entsprechen und damit zwischen 0 und 1 liegen. Oder sie kann eine Zahl sein, die einer bestimmten Kategorie entspricht: So würde digital die Antwort auf die Frage, in welche der 39 Schadensfreiheitsklassen eine Autofahrerin eingruppiert wird, als eine Zahl zwischen 1 und 39 repräsentiert werden. Bei letzterer Form ist die Anzahl der möglichen Resultate von vornherein begrenzt – man spricht von einer *Klassifikation*. Wenn prinzipiell jeder Person eine andere Zahl zugewiesen bekommen kann, dann werden Personen nach ihrer Zahl geordnet und man spricht von einem *Scoring.* Ein Beispiel dafür wäre das Schufa-Scoring zur Bewertung der Kreditwürdigkeit einer Person[14].

Die Entscheidungsregeln selbst können von Menschen beispielsweise in einem Entscheidungsbaum explizit festgelegt worden sein, Dann handelt es sich bei dem algorithmischen Entscheidungssystem um ein sogenannte *Expertensystem*. Daneben können aber die Entscheidungsregeln auch aus den Informationen über bisherige Nutzerinnen und Nutzer eines Dienstes und deren Verhalten abgeleitet werden; dann handelt es sich um Algorithmen des *maschinellen Lernens*, die die Entscheidungsregeln aus den Daten abstrahieren.

Ein klassisches Beispiel für ADM-Systeme, die ihre Entscheidungsregeln aus Daten lernen, ist die Gruppe der Produktempfehlungssysteme. Jeder digitale Marktplatz hat zahlreiche Informationen über seine Kunden: Dazu gehören die verwendeten Zahlungsarten, die Information, ob gezahlt wurde oder ob es jemals zu Unregelmäßigkeiten kam, und der Wohnort. Zudem konnte ihr Verhalten beobachtet werden: Welche Produktseiten sie wie oft und wie lange jeweils angesehen haben, ob das Produkt in und wieder aus dem Warenkorb genommen wurde, ob es gekauft und ob es wieder zurückgeschickt wurde. Mit diesen Daten können nun "ähnliche" Kunden bestimmt werden, also solche, die viele gleiche Produkte gekauft haben. Wenn nun Kundin A und Kundin B als ähnlich erkannt

[14] Sachverständigenrat für Verbraucherfragen: Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf, 01.12.2018

wurden, aber B einen Artikel (schon) gekauft hat, den A (noch) nicht hat, kann dieser Artikel Kundin A vorgeschlagen werden.

Häufig benötigt das maschinelle Lernen eine sogenannte *Grundwahrheit*, um wirklich gute Entscheidungsregeln zu identifizieren. Im vorliegenden Beispiel wäre es beispielsweise hilfreich, wenn bekannt wäre, ob das empfohlene Produkt nachher auch gekauft wurde oder nicht. Dann kann basierend auf dieser Information gelernt werden, wann eine Person besonders gut auf Empfehlungen anspricht (vielleicht eher wochentags oder am Wochenende), bei welcher Produktkategorie es gut funktioniert, ob es besser ist, viele Empfehlungen oder wenige auszusprechen und so weiter und so fort. Bei dieser Art des Lernens wird mit Hilfe eines Algorithmus des maschinellen Lernens ein *statistisches Modell* gelernt, das für eine Kombination von Eigenschaften oder Verhalten eine Wahrscheinlichkeit für den Kauf eines Produktes berechnet. Dafür versucht der Algorithmus diejenigen Eigenschaften und Verhalten zu identifizieren, die für die vorliegenden Daten der Vergangenheit am stärksten mit dem Kaufverhalten (der *Grundwahrheit*) korrelieren. Die Annahmen hinter dem Einsatz solcher Methoden sind: Das vergangene Verhalten eines Kunden und/oder das Verhalten von ihm ähnlichen Kunden sagt auch sein zukünftiges Verhalten vorher .

Wenn algorithmische Entscheidungssysteme ihre Entscheidungsregeln aus Daten gelernt haben, gilt es also zwei Algorithmen zu unterscheiden: Den Algorithmus des maschinellen Lernens, der die Entscheidungsregeln aus Daten der Vergangenheit ableitet, und der zweite Algorithmus, der mithilfe der abgeleiteten Regeln die eigentliche Entscheidung trifft. Den ersteren Algorithmus nennen wir im Folgenden auch das *Lernverfahren* – in den allermeisten Fällen ist es dieser Algorithmus und die von ihm identifizierten Regeln, die es zu kontrollieren gilt. Der eigentliche Entscheidungsalgorithmus ist dagegen einfacher strukturiert.

ADM-System:

Algorithmische Entscheidungssysteme (Algorithm Decision Making Systems), kurz ADM-Systeme, enthalten eine algorithmische Komponente, die - basierend auf der Eingabe - eine Entscheidung trifft, also aus Verbraucherinformationen eine einzige Zahl berechnet, die ein *Scoring* oder eine *Klassifikation* darstellt. Wenn der Algorithmus von Experten erarbeitet wurde, spricht man von einem Expertensystem. Daneben gibt es solche, die das Regelsystem mit Hilfe von maschinellem Lernen aus Daten selbstständig ableiten.

Im letzteren Fall besteht das algorithmische Entscheidungssystem aus den Trainingsdaten mit der darin enthaltenen Grundwahrheit, anhand derer gelernt wird, und dem aus dem Lernen entstandenen statistischen Modell, in dem die Entscheidungsregeln festgehalten wurden. Dazu gehört auch das Qualitätsmaß, anhand dessen bestimmt wurde, ob das System schon genügend trainiert ist. Eine Beurteilung des Systems wird leichter, wenn auch die zum Lernen verwendete Methode und die dafür verwendeten Parameter bekannt sind.

Der Prozess, in dem ein algorithmisches Entscheidungssystem bzw. ADM-System geplant, entwickelt, trainiert und genutzt wird, ist gepflastert mit einer Vielzahl von Entscheidungen und Annahmen. Im Arbeitspapier „Wo Maschinen irren können“[15] haben wir die lange Kette der Verantwortlichkeiten vorgestellt (siehe Abbildung 1), anhand derer die grundlegenden Entscheidungen und die jeweiligen Entscheidungsträger verortet werden können.

Die Vielzahl der Entscheider und Entscheidungen im Designprozess erschweren sowohl die Transparenz als auch die Kontrolle und damit die Nachvollziehbarkeit der vom System getroffenen Entscheidungen. Im Folgenden konzentrieren wir uns auf solche Entscheidungssysteme, die über Menschen entscheiden (z. B. durch eine Bewertung oder eine Klassifizierung), und fokussieren uns hauptsächlich auf solche Systeme, die ihre Entscheidungsregeln mit Hilfe von maschinellen Lernverfahren aus Daten lernen. Es ist allerdings auch immer instruktiv, die jeweils geforderten Transparenzpflichten und Kontrollprozesse abzugleichen mit den entsprechenden Forderungen, die an menschliche Entscheider gestellt werden.

Nicht zuletzt ist auch die Einbettung des algorithmischen Entscheidungssystems in den sozialen Prozess, für den es Entscheidungen trifft, wichtig. Daher bedarf es immer auch einer sozioinformatischen Gesamtanalyse, die im Folgenden definiert wird.

3. SOZIOINFORMATISCHE GESAMTANALYSE

Dasselbe algorithmische System, bestehend aus Trainingsdaten (inklusive Grundwahrheit), statistischem Modell und Qualitätsmaß, nach dem optimiert wurde, kann in sehr unterschiedlichen sozialen Situationen eingesetzt werden. Sein Nutzen und die möglichen Gefahren hängen aber vor allen Dingen von dem genauen Kontext seiner Verwendung ab. Um diesen Kontext bestimmen zu können, ist es notwendig, die von einem solchen System Betroffenen zu charakterisieren (das *soziale System*). Dazu gehören immer:

1) Die Verbraucher, die von einem ADM-System bewertet werden.
2) Die Personen oder Institutionen (z. B. Firmen, Staat), die das ADM-System in ihrer Arbeit einsetzen und daraufhin (automatisiert oder manuell) Entscheidungen treffen.
3) Die Personen, die diese Entscheidung vorher getroffen haben und durch das System eventuell ersetzt werden oder deren Arbeitsinhalt sich dadurch stark ändert.
4) Die Nutznießer des Systems, meistens also der Staat oder die Firma, die es einsetzen.

Genau wie menschliche Entscheider können auch Maschinen Fehler bei den durch sie berechneten Entscheidungen machen. Wir stellen die Hypothese auf, dass es vor allen Dingen das Risikopotenzial einer Maschine ist, welches das notwendige Level an Kontrolle bestimmt. Das Risikopotenzial einer Maschine setzt sich auf der einen Seite aus der Summe der Schäden von Fehlurteilen bei Individuen zusammen, die bewertet wurden. Darüber hinaus kann es aber auch gesamtgesellschaftliche Schäden geben, die sich nicht aus dem individuellen Gesamtschaden ableiten lassen. Daher ist es notwendig, in einer sogenannten *sozioinformatischen Gesamtanalyse* verschiedene Szenarien

[15] Zweig, K.A., Fischer, S., Lischka, K.: Wo Maschinen irren können Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung, 2018, https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf, 01.12.2018

durchzuspielen, um den individuellen und gesamtgesellschaftlichen Schaden abschätzen zu können.

So ist beispielsweise der individuelle Schaden, wenn ein soziales Netzwerk in einer Timeline einen eher wenig interessanten Post hoch platziert, eher gering. In wenigen Sekunden können Nutzer abschätzen, ob ein Post sie interessiert oder nicht. Das gilt auch für Fakenews, die in ihrer Timeline platziert werden. Der gesamtgesellschaftliche Schaden – z.B. der Vertrauensverlust in eine Regierung oder die Demokratie als Ganzes – kann aber durch das massenweise Verbreiten von Fakenews auf sozialen Netzwerken enorm sein.

Insgesamt kann die Wucht des Schadenspotenzials abgemildert werden, wenn es einen Markt von verschiedenen ADM-Systemen zu derselben Fragestellung gibt, die für die VerbraucherInnen leicht nutzbar sind. Je höher also der Anteil an Mitbewerbern und je leichter der Wechsel oder die gleichzeitige Nutzung ist, desto weniger externer Kontrolle bedarf es. Daher ist es für die sozioinformatische Gesamtanalyse auch wichtig, den gesamten Markt und die gesetzliche Bestimmung zur Einsichtnahme in den Entscheidungsprozess zu berücksichtigen.

Nicht zuletzt verändert der Einsatz von ADM-Systemen auch das Verhalten der Menschen – sowohl derer, die das System einsetzen als auch derer, die bewertet werden. Im ersteren Fall muss man davon ausgehen, dass es zu einer Delegation von Verantwortung kommt und der menschliche Entscheider vermutlich mehr und mehr dem Resultat der Maschine folgt. Das kann positiv oder negativ sein: Wenn z. B. in einer Bank schon leichte Risikowerte einer Kundin dazu führen, dass ein Kredit nicht vergeben wird, kann eine maschinelle Entscheidung dabei helfen, auch diesen Personen noch einen Kredit zu verleihen, da das persönliche Risiko des Bankangestellten bei einem geplatzten Kredit sinkt.
Es gibt aber auch genügend Fälle, in denen ADM-Systeme Personen in eine Hochrisikokategorie einsortiert haben, die dort nicht hineingehörten. Das im amerikanischen Justizvollzug eingesetzte ADM-System ‚Correctional Offender Management Profile for Alternative Sanctions' kurz COMPAS irrt sich bei 80% der Straftäter, für die es ein hohes gewaltbasiertes Rückfallrisiko prognostiziert[16] und es wird dennoch angewandt.

Eine Analyse des statischen sozioinformatischen Gesamtsystems ist allerdings nicht ausreichend. Gerade wenn Menschen bewertet werden, um damit Zugang zu einer begehrten Ressource zu erlangen (Bestellung auf Rechnung, Kredite, Versicherungen, etc.), kann man davon ausgehen, dass zumindest ein Teil, der zu Bewertenden versuchen wird, die Bewertung zu ihren Gunsten zu manipulieren. Daher ist es auch immer notwendig, eine Forderung nach Transparenz der Entscheidungsregeln für die Allgemeinheit daraufhin zu überprüfen, wie stark dadurch vermutlich die Manipulationsversuche der zu Bewertenden werden könnten. Dies ist insbesondere dann wichtig, wenn die Anzahl der ADM-Systeme auf dem Markt sehr klein und es daher besonders lohnenswert ist, diese Systeme so gut kennenzulernen, dass sie manipulierbar werden.

[16] Tobias Krafft: Qualitätsmaße binärer Klassifikationen im Bereich kriminalprognostischer Instrumente der vierten Generation, 2017, https://arxiv.org/abs/1804.01557, Seite 46

Ein gutes Beispiel für diesen Balanceakt stellen die ausführlichen Handreichungen von Google bezüglich der gewünschten "Search Engine Optimization" von Webseitenbetreibern dar[17], um ein angemessenes Ranking zu erreichen. Diese sind aber nie so detailliert, dass Webseitenbetreiber damit ihr Ranking feingranular steuern können, denn dann würden insbesondere Phishingseiten und andere unerwünschte Webinhalte diese Mechanismen nutzen, um sich zu positionieren.

Eine **sozioinformatische Gesamtanalyse** des möglichen Schadens, der durch ein ADM-System verursacht werden könnte, verlangt daher in einem ersten Schritt die Identifikation aller Akteure und Institutionen, die diesen Schaden erleiden, erzeugen, vermehren oder vermindern können.

In einem zweiten Schritt müssen die dynamischen Veränderungen, die durch den Einsatz des Systems verursacht werden könnten, in die Bewertung des zukünftigen Schadenspotenzials mit einbezogen werden. Dafür gilt es, die möglichen Reaktionen von Bewertenden und Bewerteten zu bedenken.

[17] Google: Startleitfaden zur Suchmaschinenoptimierung (SEO), 2018, https://support.google.com/webmasters/answer/7451184?hl=de, 01.12.2018

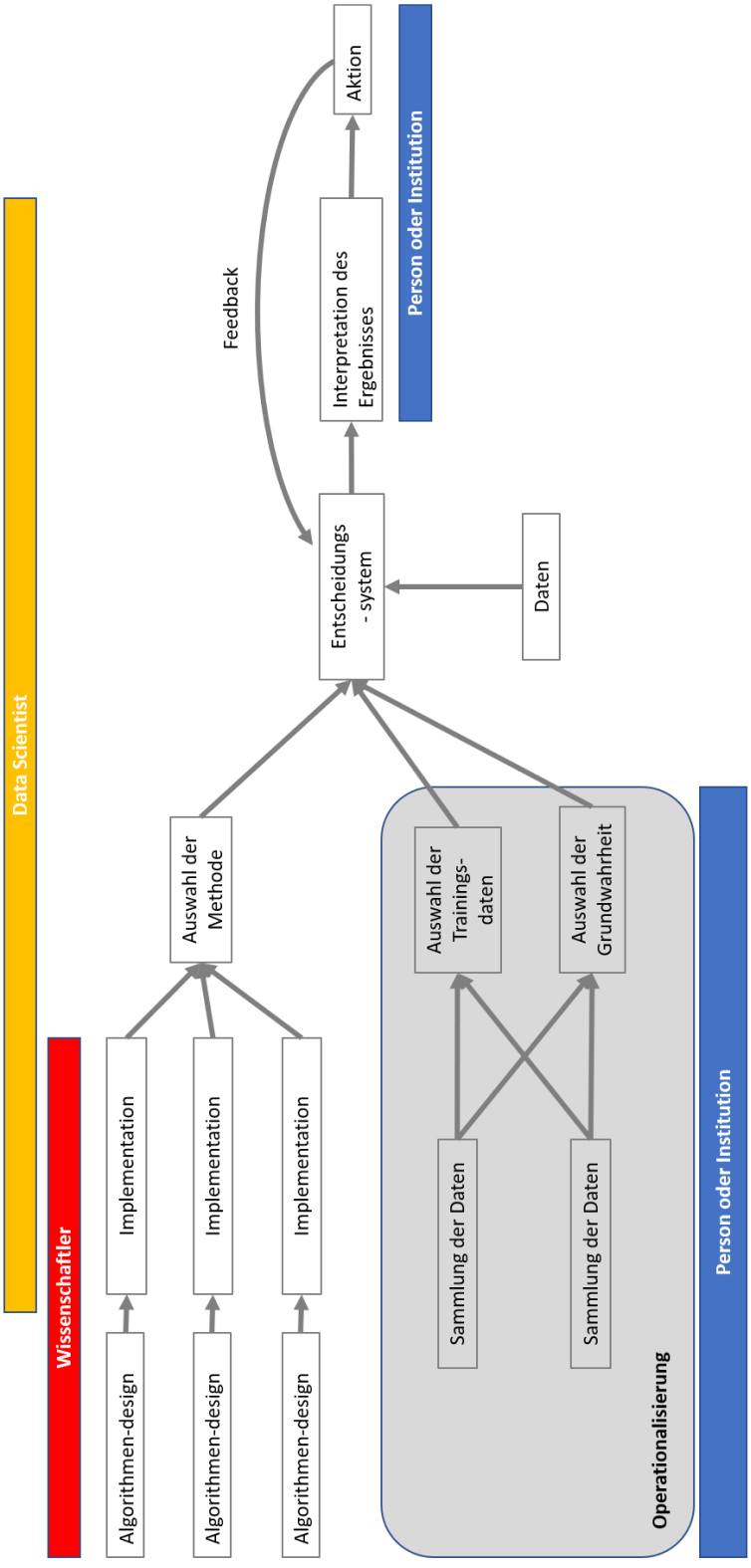

Abbildung 1: Lange Kette der Verantwortlichkeit beim Schaffungsprozess eines algorithmischen Entscheidungssystems (erweiterte Version, ursprünglich aus „Wo Maschinen irren können Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung „[18]). Anhand dieser sollen bestehende Methoden zur Sicherstellung und Gewährleistung der Nachvollziehbarkeit von komplexen ADM-Prozessen verordnet werden.

[18] Zweig, K.A., Fischer, S., Lischka, K.: Wo Maschinen irren können Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung, 2018, https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf, 01.12.2018

III. TRANSPARENZ UND NACHVOLLZIEHBARKEIT

Der öffentlichen Diskussion um die Forderungen nach ‚*Transparenz'* und ‚*Nachvollziehbarkeit'* in der digitalen Welt fehlt es zuweilen an einer klar definierten Begrifflichkeit, was eine Verständigung unnötig erschwert und zu Missverständnissen führen kann, insbesondere wenn die Begriffe für ganz verschiedene Konzepte stehen. Um eine differenzierte Betrachtung zu ermöglichen, soll daher zunächst in diesem Kapitel des Gutachtens näher auf die Begriffe *Transparenz* und *Nachvollziehbarkeit* eingegangen werden.

Der Begriff *Transparenz* leitet sich vom lateinischen ‚*transparens*', zu Deutsch ‚durchscheinend' oder ‚durchsichtig', ab, wobei diese ‚Durchsichtigkeit' sowohl materiell als auch immateriell oder abstrakt im Sinne der ‚Durchschaubarkeit' einer Sache oder eines Sachverhalts verstanden werden kann. Letzteres verdeutlicht das Ziel der Transparenzforderung, das meist nur durch eine zu gewährende Einsichtnahme in den Erschaffungsprozess des in Frage stehenden ADM-Systems erreicht werden kann.

Transparenz wird in der heutigen Gesellschaft zunehmend eingefordert, insbesondere im Bereich des Verbraucherschutzes gibt es verschiedene Transparenzforderungen, beispielsweise zur Erleichterung einer selbstbestimmten Kaufentscheidung. Die Forderungen reichen vom ausreichenden Umfang erforderlicher Informationen bis hin zu einer für den Verbraucher verständlichen Vermittlung. Marina Tamm, Wirtschaftsrechtsprofessorin an der HS Wismar, setzte sich in ihrer Habilitation mit dem Thema Verbraucherschutzrecht auseinander und arbeitete das Transparenzgebot im verbraucherschutzrechtlichen Kontext auf[19]. Sie spricht hier von Transparenz in Verbindung mit notwendigen Vorabinformationen, die unerlässlich sind, damit der Kunde weiß „worauf er sich einlässt, um (auch materiell und nicht nur formell) eine privatautonome Entscheidung treffen zu können"[20]. Die Forderungen beruhen also auf zivil- oder strafrechtlicher Verantwortlichkeit, finden bislang jedoch hauptsächlich im Kontext des Erwerbs von Dienstleistung oder Waren Anwendung.

Durch die im letzten Kapitel angesprochene Digitalisierung verschiebt sich aktuell nicht nur die Position des Verbrauchers vom Kunden zum Nutzer, sondern es bilden sich auch vermehrt Bereiche, in denen Firmen und Institutionen über algorithmische Entscheidungssysteme in Kontakt mit Verbrauchern treten anstatt analog, weshalb versucht wird, diese ähnlichen Transparenzforderungen zu unterwerfen. Die Frage nach Verantwortlichkeit im Falle eines Schadens ist hierbei ein starker Antrieb für die juristische Debatte. Die 2016 verabschiedete und ab Mai 2018 in Kraft getretene EU-Datenschutz-Grundverordnung (kurz DSGVO) enthält neben den Vorgaben zum Thema Datenschutz auch kon-

[19] Tamm, M.: Verbraucherschutzrecht: Europäisierung und Materialisierung des deutschen Zivilrechts und die Herausbildung eines Verbraucherschutzprinzips (Vol. 158). Mohr Siebeck, 2011, Seite 348ff.

[20] Ebd. 347 ff.

krete Vorgaben für Systeme, die auf Grundlage personenbezogener Daten automatisierte Entscheidungen treffen[21]. Hier werden zwar explizit Systeme, welche vollautomatisiert Entscheidungen über Menschen treffen, auf juristischer Ebene verboten, jedoch im selben Atemzug weitreichende Ausnahmen, wie z. B. nationale Gesetze oder explizite Einwilligungen der Verbraucher eingeräumt[22]. Für andere Systeme werden weitere Forderungen in Bezug auf Aufklärungsarbeit für Erhebung, Verarbeitung und Speicherung personenbezogener Daten genannt. Dreyer und Schulz schlussfolgern daraus folgenden Kernpunkt:

„Je besser der Betroffene darüber Bescheid weiß, wer zu welchem Zweck welche personenbezogenen Daten verarbeiten möchte, desto eher kann er sein Recht auf informationelle Selbstbestimmung tatsächlich wahrnehmen.“[23]

An die Art der Aufbereitung und Präsentation dieser Informationen stellt die DSGVO hohe Ansprüche, weshalb sie auch einen Erwägungsgrund (ErwG) mit dem Titel: „Grundsatz der Transparenz“ (ErwG 58 der DSGVO[24]) in der Präambel verankert. Erwägungsgründe werden im rechtlichen Bereich eingesetzt, um aufzuzeigen, welche Überlegungen zum Erlass des Rechtsaktes geführt haben. In diesem Fall gibt der Erwägungsgrund einen Anhaltspunkt, wie im späteren Verlauf der DSGVO Forderungen nach Transparenz auszulegen und zu interpretieren sind:

„Der Grundsatz der Transparenz setzt voraus, dass eine für die Öffentlichkeit oder die betroffene Person bestimmte Information ***präzise****,* ***leicht zugänglich*** *und* ***verständlich*** *sowie in* ***klarer*** *und* ***einfacher Sprache*** *abgefasst ist und gegebenenfalls zusätzlich visuelle Elemente verwendet werden. (…) Wenn sich die Verarbeitung an Kinder richtet, sollten aufgrund der besonderen Schutzwürdigkeit von Kindern Informationen und Hinweise in einer dergestalt klaren und einfachen Sprache erfolgen, dass ein Kind sie verstehen kann.“* (ErwG 58 der DSGVO[25]).

In der bisherigen Datenschutzdiskussion des Verbraucherschutzes bezieht sich das Transparenzgebot auf eine für den Verbraucher verständliche Präsentation notweniger Fakten. Wenn auch diese Akzentsetzung bis jetzt für anstehende Fragestellungen ausreichend war, so muss die Transparenzforderung im aktuellen Kontext von ADM-Systemen insgesamt erweitert werden. Transparent müssen aber nicht nur die relevanten Daten im Hinblick auf einen selbstbestimmten Umgang mit dem System sein, wie z. B., welche Datengrundlage genutzt wird oder wer die eigentliche Entscheidung trifft, ein Mensch oder eine Maschine. Zusätzlich bedarf es der Kenntnis über die zugrundeliegenden Mechanismen, Datenverarbeitung oder etwaiger Informationen über die Konzeption des ADM Systems und seine Einbettung in den sozialen Prozess, in dem eine Entscheidung gefällt wird. Hieraus resultieren verschiedene Transparenzforderungen, auf die in den folgenden Kapiteln noch genauer eingegangen wird.

---

[21] Schulz, W., Dreyer, S.: Was bringt die Datenschutz- Grundverordnung für automatisierte Entscheidungssysteme?, 2018, 1 https://doi.org/10.11586/2018011, 01.12.2018, Seite 18

[22] Ebd. Seite 19ff

[23] Ebd. Seite 23.

[24] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), ABl. EG Nr. L 119 v. 4.5.2016: 1–88

[25] Ebd.

Bei der Forderung nach **Transparenz** im Kontext algorithmischer Entscheidungssysteme stehen Informationspflichten zum Lernverfahren, seiner Qualität, der Datengrundlage und der Einbettung des ADM Systems im Gesamtkontext im Fokus. Generell gilt damit in diesem Kontext, dass Transparenz darauf abzielt, dass eine grobe Evaluation dieser Aspekte ermöglicht wird, aber noch nicht notwendigerweise die Überprüfung der zur Verfügung gestellten Informationen (die eigentliche **Nachvollziehbarkeit**). Je nach Fall muss zudem geklärt werden, wer welche Informationen erhält. In jedem Fall müssen die Informationen präzise, leicht zugänglich und verständlich sowie in klarer und einfacher Sprache vermittelt werden.

Am Beispiel der von Dreyer und Schulz aufgezeigten fehlenden Ansatzpunkte der DSGVO für „gruppen- und gesellschaftsbezogene Ziele wie Nichtdiskriminierung und Teilhabe“[26] wird ein wichtiger Punkt der aktuell entbrannten Debatte um die Kontrolle und Überprüfung von ADM-Systemen aufgezeigt. Eine Überprüfung der Systeme auf diskriminierende Entscheidungen beispielsweise bei unterschiedlichen Personengruppen fällt sowohl aus dem Forderungsrahmen der DSGVO als auch dem bisheriger Transparenzgebote des Verbraucherschutzes.

Unter Umständen kann mit einer reinen Einsichtsforderung einer solchen Problematik kaum entgegengewirkt werden, von der Schwierigkeit diese aufzudecken ganz zu schweigen. Doch im Jahr 2014 beschritt Nicholas Diakopoulos, Assistenz-Professor an der US-amerikanischen University of Maryland College of Journalism, einen gänzlich neuen Weg, indem er vorschlug, ein allgemeines, naturwissenschaftliches Vorgehen auf ein nicht-transparentes ADM-Systems zu transferieren. Er übertrug damit die aus dem Software-Engineering bekannte ‚Blackbox-Analyse‘, in einen öffentlichen Kontext, um damit ein ADM-System ohne direkten Einblick in die interne Entscheidungsstruktur zu analysieren[27].

Hierbei wird das System als ‚Blackbox‘ betrachtet, also als ein System, in das man keinen direkten Einblick erhält. Dennoch kann man ein Verständnis wichtiger Eigenschaften bekommen, indem man dieses mit Daten füttert und aus der Beziehung von Ein- und Ausgabe Rückschlüsse über die innenliegende Mechanik sowie über die Güte des Systems zieht. Wenn es eine öffentliche Schnittstelle für eine solche Analyse gibt, die es erlaubt, möglichst automatisiert viele Anfragen an das ADM-System zu stellen, können damit Qualitätsaussagen des Betreibers über die Entscheidungen seines Systems überprüft werden, was deutlich über den Transparenzbegriff hinausgeht. Die Blackbox-Analyse stellt sich insofern als eine der ersten Methoden dar, welche eine Nachvollziehbarkeit zu schaffen vermag.

---

[26] Schulz, W., Dreyer, S.: Was bringt die Datenschutz- Grundverordnung für automatisierte Entscheidungssysteme?, 2018, 1 https://doi.org/10.11586/2018011, 01.12.2018, Seite 9

[27] Diakopoulos, N.: Algorithmic-Accountability: the investigation of Black Boxes, 2014, https://www.cjr.org/tow_center_reports/algorithmic_accountability_on_the_investigation_of_black_boxes.php, 01.12.2018

Der Begriff des *Nachvollziehens* orientiert sich hier am aktuellen Sprachgebrauch im Kontext von ADM-Systemen und weicht insofern stark von der kontextuellen Verwendung des Begriffs beim Verbraucherschutz ab. Dort gilt eine vollständige Erklärung interner Strukturen als ‚nachvollziehbar', wenn der Verbraucher die erklärte Tatsache für sich nachvollziehen kann – also vollständige Transparenz herrscht. Diese Gleichsetzung gilt aber im Bereich der ADM-Systeme nicht. Transparenzpflichten reichen bei weitem nicht aus, um echte Nachvollziehbarkeit zu erreichen. Daher unterscheiden wir im Folgenden Transparenzpflichten von Kontrollprozessen.

Zusammenfassend lässt sich also feststellen, dass ‚Nachvollziehbarkeit' eines ADM-Systems nicht nur Information erfordert, sondern darüber hinaus eine Überprüfbarkeit von kritischen Resultaten geleistet werden muss. Die kontrollierende Instanz wird hierbei in eine operative, handelnde Rolle versetzt. Um zum Beispiel Entscheidungen bezüglich der Datenbasis nachzuvollziehen, wird nicht nur die Information benötigt, welche Datenquellen herangezogen wurden (Transparenz), sondern es ist zudem notwendig, einen expliziten Zugang zu den verwendeten Trainingsdaten zu gewähren, um die Informationen zu überprüfen.

Bei der Forderung nach **Nachvollziehbarkeit** im Kontext algorithmischer Entscheidungssysteme steht eine Einsichtnahme und Überprüfbarkeit im Vordergrund. Generell gilt, dass Nachvollziehbarkeit darauf abzielt, dass die kontrollierende Instanz in eine operative, handelnde Rolle versetzt wird. Auch hier muss je nach Fall geklärt werden, wer diesen Zugang erhält.

Es sei noch angemerkt, dass ein weiterer Begriff die Sachlage verkompliziert: Unter dem englischen Stichwort *explainable artifical intelligence* werden bestimmte Methoden des maschinellen Lernens zusammengefasst, die eine tatsächliche, von Menschen nachvollziehbare Erklärung zu einer Entscheidung liefern. Es geht also nicht nur um die Entscheidung als solche, ob eine Person beispielsweise einen Kredit erhält oder nicht, sondern zum Beispiel um die zusätzliche Erläuterung, weshalb nicht. Auch dieser Begriff wird häufig im Rahmen der Nachvollziehbarkeit und als eine Möglichkeit von Transparenz diskutiert. Auf Aspekte dieser Thematik wird im Kapitel V näher eingegangen.

Abschließend lässt sich feststellen, dass sich durch den zusätzlichen Schritt von Transparenz zur Nachvollziehbarkeit zwar insgesamt der Arbeitsaufwand erhöht, dennoch ist dies lohnenswert, denn im selben Maß steigt das Vertrauen in die Ergebnisse der ADM-Systeme, da mehr relevante, überprüfbare Fakten geliefert werden und man sich weniger auf externe Aussagen, zum Beispiel der ADM-Betreiber, verlassen muss. Zudem können nur so ADM-Systeme in sensiblen Bereichen, die vielleicht zusätzlich auch rechtlichen Anforderungen genügen müssen, beweisen, dass sie diese erfüllen.

IV. KONTEXTABHÄNGIGKEIT DER REGULIERUNGSTIEFE

Gerade im Verbraucherschutz werden bei rechtlichen Fragen in den letzten Jahren zunehmend Forderungen nach Transparenz und Nachvollziehbarkeit im Kontext von ADM-Systemen gestellt. So konstatiert die Landeszentrale für Medien und Kommunikation gemeinsam mit der Verbraucherzentrale Rheinland-Pfalz:

"Um die Nachvollziehbarkeit zu ermöglichen, sollten die Bewertungskriterien und die zugrundeliegenden Daten grundsätzlich offengelegt werden müssen."[28] (LMK & VZRP, 2017).

Allerdings muss realistischerweise berücksichtigt werden, dass der Kostenfaktor für eine erfolgreiche Umsetzung der oben genannten Transparenz- und Nachvollziehbarkeitsforderungen eine entscheidende Rolle spielen wird. Zu hohe Kosten werden sich innovationshemmend auswirken, sodass diese unbedingt zu vermeiden sind. Aus diesem Grund sollten sich die verschiedenen Level an Transparenz und Nachvollziehbarkeit auf die jeweils kleinstmöglichen Teilmengen von ADM-Systemen beschränken.

Wir schlagen eine Kategorisierung der ADM-Systeme anhand zweier Dimensionen vor, und zwar zum einen nach dem möglichen Schadenspotenzial auf individuellem und gesamtgesellschaftlichem Level und zum anderen nach der Möglichkeit der Re-Evaluation bei einer Fehlbeurteilung. Wie wir im Folgenden zeigen werden, ist es zur Evaluation des Schadenspotenzials notwendig, das in Frage stehende ADM-System im gesamten sozioinformatischen Kontext zu betrachten und zu beurteilen, um sowohl die Brisanz der Folgen auf den Verbraucher bzw. Kunden als auch auf die Gesellschaft zu analysieren, die je nach Einsatzbereich des jeweiligen ADM-Systems unterschiedlich zu bewerten sind. Aus diesen beiden Dimensionen – dem Schadenspotenzial und der Leichtigkeit einer Zweitevaluation – ergeben sich die verschiedenen Level an Transparenzpflichten und notwendigen Kontrollprozesse.

Die Klasse der ADM-Systeme mit geringem Schadenspotenzial und hoher Anbietervielfalt bedarf keiner Regulierung. Dazu gehören klassische Produktempfehlungssysteme, beispielsweise im Onlinehandel, die auf jeder Verkaufsplattform zu finden sind und durch ein falsch vorgeschlagenes Produkt weder der Gesellschaft noch dem Verbraucher einen nennenswerten Schaden verursachen. An dieser Stelle wären also ausführliche Transparenzforderungen unangemessen, und es reicht, wenn beispielsweise durch den Verbraucherschutz sporadische Tests oder nur in Verdachtsfällen post-hoc Analysen durchgeführt werden. Einfache Fehler behebt hier der Markt der Anbieter von selbst, da schlechte Vorschläge zu finanziellen Einbußen für die Unternehmen führen, was dort sozusagen ein intrinsisches Bedürfnis nach Verbesserung auslöst.

---

[28] LMK & VZRP: Zwischen Filterblase, Fake News und Demokratiegestaltung – Algorithmen brauchen Transparenz und Kontrolle. Gemeinsames Positionspapier der Landeszentrale für Medien und Kommunikation Rheinland-Pfalz (LMK) und der Verbraucherzentrale Rheinland-Pfalz (VZRP), 2017 https://www.verbraucherzentrale-rlp.de/sites/default/files/migration_files/media247794A.pdf, 01.12.2018

Anders gestaltet sich hingegen der Einsatz eines ADM-Systems im Bereich der KFZ-Versicherungen bei den sogenannten Telematik-Verträgen. Das ADM-System überwacht und beurteilt hier das Fahrverhalten des Kunden, woraus sich die Höhe des zu zahlenden Beitrags ergibt[29]. Hier ergibt sich bei Fehlurteilen auf der Ebene der falsch beurteilten Individuen ein Schadenspotenzial, dessen Ausmaß je nach Einkommen des Kunden unterschiedlich zu bewerten wäre. Ein darüber hinaus gehender gesamtgesellschaftlicher Schaden ist jedoch zunächst nicht erkennbar, da die Systeme momentan noch freiwilliger Natur sind und es auch noch genügend Anbieter ohne Telematik-Angebote gibt[30]. Auch können zum aktuellen Zeitpunkt VerbraucherInnen entscheiden, ob sie einen solchen Tarif und die damit implizierte Überwachung akzeptieren oder einen „klassischen" KFZ-Tarif wählen.

Obwohl es also noch Ausweichmöglichkeiten gibt, sollte hier aus verbraucherschutzrechtlicher Perspektive dennoch ein höheres Maß an Transparenz- und Nachvollziehbarkeitsforderungen gestellt werden. An dieser Stelle ist eine post-hoc-Analyse nicht ausreichend und eine aktive Schnittstelle zur dauerhaften Überprüfung auf diskriminierende Entscheidungen sollte von den Versicherungsgesellschaften eingerichtet werden.

Die erwähnten Beispiele demonstrieren, dass die Regulierungstiefe kontextabhängig vom gesellschaftlichen Bereich unter besonderer Berücksichtigung des möglichen Schadenpotenzials der algorithmischen Entscheidungsfindung auf den Menschen ermittelt werden muss.

Im Folgenden wird der Frage nachgegangen, warum eine isolierte Analyse des ADM-Systems nicht ausreichend ist, sondern das System im gesellschaftlichen Gesamtkontext gesehen und untersucht werden muss. Anschließend wird in Unterkapitel 2 die Bewertung der Gefahr eines sozioinformatischen Systems anhand einer Risikomatrix erarbeitet.

1. KONTEXTABHÄNGIGE ANALYSE DES SCHADENSPOTENZIALS VON ADM SYSTEMEN

Die folgende Vorstellung von drei verschiedenen Szenarien soll die Kontextabhängigkeit als entscheidenden Faktor bei der Überprüfung eines ADM-Systems verdeutlichen, denn jedes ADM-System ist Teil eines sozioinformatischen Systems, in das es eingebettet ist. Da beide Systeme miteinander verflochten sind und sich gegenseitig beeinflussen, können Änderungen sowohl im sozioinformatischen als auch im ADM-System weitgehende Konsequenzen für den gesamten Prozess haben und müssen unbedingt berücksichtigt werden.

Die Beispiele sollen insofern veranschaulichen, wie sich das sozioinformatische System eines ADM-Systems durch den Einfluss unterschiedlicher Anwendungsszenarios konkret verändern kann. Dementsprechend muss auch die erforderliche Regulierungstiefe

[29] Sachverständigenrat für Verbraucherfragen: Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf, 01.12.2018, Seite 18

[30] Wenn sich diese Einbettung des ADM-Systems in den sozialen Prozess (Autoversicherungsmarkt) verändert, wird sich die Einschätzung des Schadenspotenzials drastisch verändern. Die Neubewertung könnte sogar dazu führen, dass man ein Verbot fordern muss.

des ADM-Systems angepasst werden, da dies in unmittelbarem Bezug zu seiner Stellung im gesellschaftlichen Gesamtkontext erfolgen muss. Es bedarf also einer sozioinformatischen Gesamtanalyse, denn eine isolierte Betrachtung eines ADM-Systems ohne Berücksichtigung der jeweiligen Beziehungen innerhalb des betreffenden sozioinformatischen Systems kann zu gravierenden Fehleinschätzungen der zu erwartenden Folgen führen.

1.1 Explizite Änderung des sozioinformatischen Systems durch veränderte Datengrundlage

Der Einfluss des Kontexts des sozioinformatischen Systems auf die notwendige Regulierungstiefe wird bei Betrachtung einer expliziten Änderung des Anwendungsszenarios eines ADM-Systems am einfachsten sichtbar. Viele verbraucherrelevante Szenarien von ADM-Systemen laufen aus informatischer Perspektive auf ein und dasselbe ADM-System und dasselbe dahinterliegende Lernverfahren hinaus, wie zum Beispiel die oben schon genannten Produktempfehlungssysteme (englisch „recommender system“). Ob ein solches System nun auf Grundlage der Vorlieben eines Kunden personalisierte Werbung für unterschiedliche Produkte generiert oder eine individuelle Nachrichtenlandschaft anzeigt, die dahinterliegende Technologie bleibt gleich. Die Maschine bekommt jeweils eine Menge von Objekten (Produkte, Nachrichtenbeiträge usw.), die sie anhand von Informationen über das bisherige Nutzerverhalten und die Eigenschaften der Objekte für jeden neuen Nutzer individuell sortieren und selektieren muss.

In Abbildung 2 wird deutlich, dass hierbei exakt dasselbe ADM-System Anwendung finden kann, wohingegen die Interpretation der Ergebnisse als Werbung für Kleidung oder als Suchergebnisse einer Onlinesuchmaschine diese in völlig andere sozioinformatische Systeme einbettet.

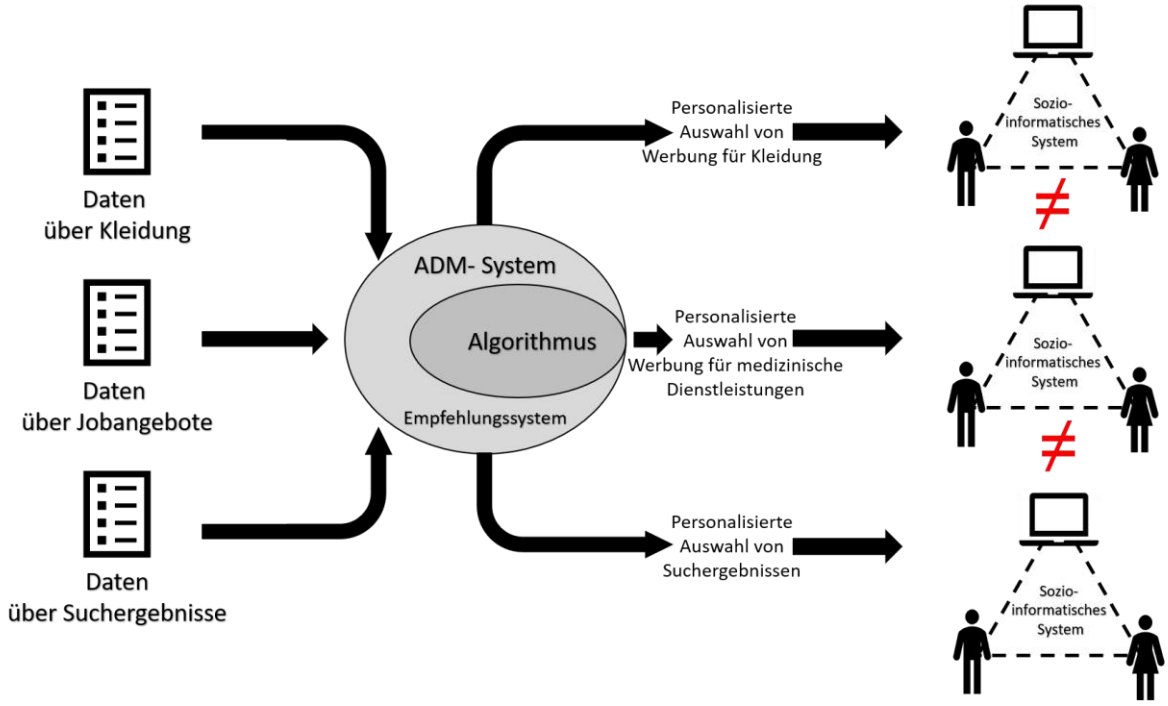

Abbildung 2: Dasselbe ADM-System kann bei expliziter Änderung des sozioinformatischen Systems eingesetzt werden, dies wird verdeutlicht am Beispiel von verschiedenen Anwendungsszenarien von Empfehlungssystemen.

Bei der Betrachtung der verschiedenen Anwendungsszenarien und des jeweiligen Unterschieds hinsichtlich ihres Schadenspotenzials wird der differierende Transparenzbedarf sichtbar, obwohl der Kern des ADM-Systems, das Lernverfahren, sich nicht verändert hat.

Während personalisierte Werbung für Kleidung wie schon erwähnt kaum ein Schadenspotenzial bei einem falschen Vorschlag entfaltet, gibt es Bereiche, in denen sogar Wer-

bung kritisch zu betrachten und zu überwachen ist. So kann eine personalisierte Werbung für medizinische Dienstleistungen, gerade bei Verbrauchern einer eher „labilen“ Personengruppe, bei Empfehlung unorthodoxer oder ungetesteter Verfahren ein hohes individuelles Schadenspotenzial entfalten. In diesem Bereich sollte also vielleicht sogar überlegt werden, auf den Einsatz lernender ADM-Systeme ganz zu verzichten.

In diesem Beispiel wurde gezeigt, dass ein und dieselbe **Art** von ADM-System, angewandt auf verschiedene Daten, zu unterschiedlichen Risikobewertungen führt. Das nächste Beispiel zeigt, dass sogar dasselbe ADM-System auf denselben Daten bei verschiedenen Nutzergruppen unterschiedliche Schadenspotenziale erzeugt.

1.2 Implizite Kontextänderung am Beispiel von automatisierten Videovorschlägen

Das vom Videoportal YouTube[31] verwendete ADM-System zur Erstellung der Vorschläge und für die Autoplayfunktion wird aktuell in den Medien kontrovers diskutiert: Während sich die Vorwürfe von Journalisten bisweilen auf die Auswahl der Videos unter Diversitätsaspekten beschränken (siehe hierzu beispielhaft den Artikel von Michael Moorstedt vom 04.02.2018 in der Süddeutsche Zeitung[32]), so fiel dem Schriftsteller James Bridle das Folgende auf:

Wenn er Videos für Kleinkinder startete, wurden die Titel der vorgeschlagenen Videos immer kryptischer und die Inhalte schienen oftmals wahllos zusammengeschnitten zu sein. Er spricht in einem Blogeintrag auf Medium[33] und einem TED-Talk[34] davon, dass es offenbar mehrere Millionen Aufrufe bei Videos gibt, die weder durch Titel noch durch inhaltliche Güte eine Daseinsberechtigung erringen könnten.

Er untersuchte dies auf journalistischer Basis und stellt folgende Hypothese auf: Es scheint eine Gruppe von YouTube-Accounts zu geben, welche für die Autoplayfunktion optimierte Videos im Kleinkindsektor erstellen, um Kinder, die von ihren Eltern vor den Browser mit eingeschalteter Autoplayfunktion gesetzt werden, als Zuschauer zu gewinnen. Da sich gerade diese Zuschauergruppe nicht aktiv gegen die Vorschläge entscheiden kann – schließlich sind die Kinder hier so jung, dass eine Interaktion mit dem Computer auszuschließen ist – müssen die Videos keinerlei Qualitäten oder Originalität aufweisen. Diese Videos sind vermutlich vom Computer aus erfolgreichen Videos automatisch zusammengeschnitten mit dem bloßen Ziel, dem ADM-System, das die Autoplay-Vorschläge generiert, positiv aufzufallen, um somit von diesem häufig ausgeliefert zu werden.

Die technische Komponente des ADM-Systems ist zwar zum Schutz des Firmengeheimnisses weitgehend unbekannt, jedoch haben Google-Mitarbeiter 2010[35] und 2016[36] die verwendete Methodik zum Generieren der Vorschläge jeweils in einem Paper kurz angerissen. Hierbei handelt es sich um ein ADM-System, welches in ein hochkomplexes sozioinformatisches Gesamtsystem eingebettet ist.

---

[31] https://www.youtube.com

[32] Michael Moorstedt: Youtubes Lügenalgorithmus, 2018, https://www.sueddeutsche.de/digital/netzkolumne-youtubes-luegenalgorithmus-1.3853777, 01.12.2018

[33] James Bridle: Something is wrong on the internet , 2017, https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2, 01.12.2018

[34] James Bridle: The nightmare videos of childrens' YouTube — and what's wrong with the internet today, 2017, https://www.ted.com/talks/james_bridle_the_nightmare_videos_of_childrens_youtube_and_what_s_wrong_with_the_internet_today

[35] Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010). The YouTube video recommendation system. In Proceedings of the fourth ACM conference on Recommender systems (pp. 293-296). ACM.

[36] Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube rec-ommendations. In Proceedings of the 10th ACM Conference on Recommender Sys-tems (pp. 191-198). ACM.

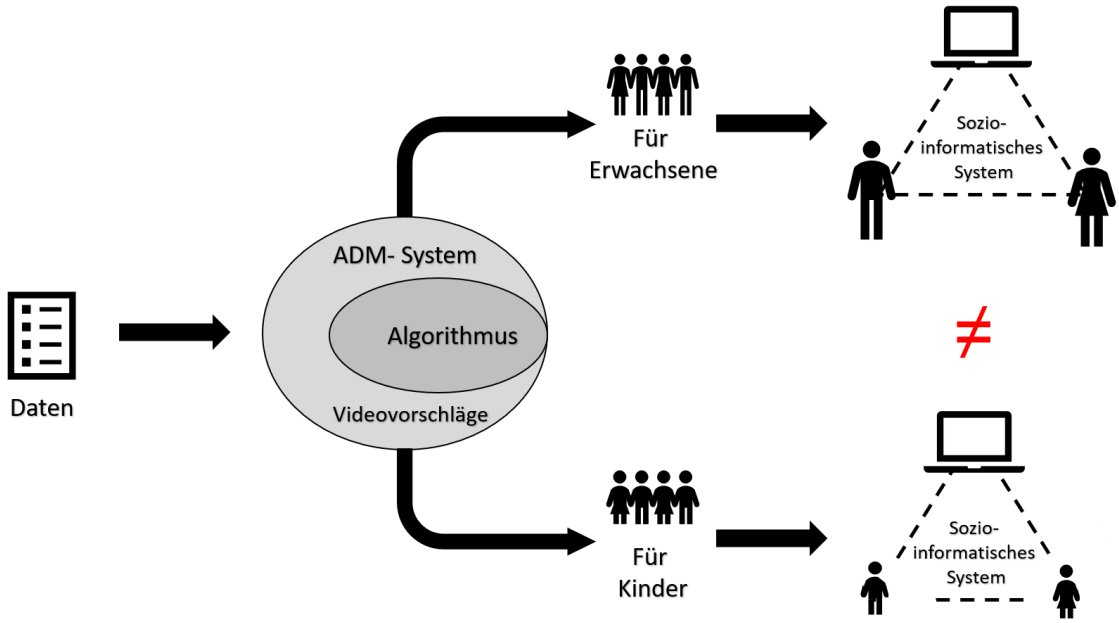

Abbildung 3: Veränderung des sozioinformatischen Systems durch anderes Zielpublikum.

Interessant ist nun der folgende Aspekt: Während die offensichtliche Nutzergruppe von YouTube, Jugendliche und Erwachsene, eine aktive Interaktionsmöglichkeit hat und sich somit gegen die Autoplayvorschläge entscheiden kann, steigt der mögliche Schaden bei falschen, nicht sinnvollen oder sogar verstörenden Videovorschlägen bei (Klein-) Kindern dramatisch an. Obwohl also alle genutzten Daten, das vollständige ADM-System sowie das eigentliche Ergebnis unverändert bleiben, ändert sich durch das andere Zielpublikum das sozioinformatische System grundlegend (siehe hierzu Abbildung 3). Diese Änderung geht mit einer Steigerung der möglichen Gefahren bei einer Fehlentscheidung einher, sodass deutlich weitreichendere Transparenzforderungen notwendig sind. Die beiden sozioinformatischen Systeme rufen also einzig und allein durch die implizite Kontextänderung einen völlig unterschiedlichen Regulierungsbedarf auf.

Zuletzt zeigen wir noch, dass eine einzige, interne Einstellung eines ADM-Systems zu unterschiedlichen Risikobewertungen führen kann.

1.3. Intransparente Änderung im ADM-System

Als im Jahr 2003 der Suchmaschinenanbieter Google das Start-up Kaltix aufkaufte, welches sich mit personalisierten und kontextsensitiven Suchwerkzeugen beschäftigte[37], begannen die Gerüchte, Google arbeite an einer Möglichkeit, die ausgelieferten Suchergebnisse zu personalisieren[38]. Dies wurde im folgenden Jahr offiziell bestätigt[39] und ab 2009 sprach der Konzern schließlich von einer personalisierten Suche für alle[40] und verankerte dieses Feature klar in den eigenen Datenschutzrichtlinien:

”*We use the information we collect to customize our services for you, including providing recommendations, personalized content, and customized search results*”[41] [42]

Im Jahr 2011 legte der Internetaktivist Eli Pariser durch sein Buch ”The Filter Bubble: What the Internet Is Hiding from You”[43] den öffentlichen Fokus auf sogenannte algorithmisch generierte Filterblasen. Laut Pariser besteht die Gefahr, dass eine individuelle Filterung des Informationsflusses dazu führen könne, dass Gruppen oder Individuen unterschiedliche Fakten präsentiert bekommen, also in einem eigenen Informationsuniversum lebten[44]. Dies sei insbesondere dann besorgniserregend, wenn die Inhalte politisch extremer Natur wären. Weiterhin sieht Pariser durch die einseitige Perspektive die Diskursfähigkeit dieser Bürger geschwächt und befürchtet, dass sie gänzlich verkommen könnte.

Eine Filterblase ist in diesem Sinne eine algorithmische Selektion von Nachrichten, die den eigenen Perspektiven entsprechen, was insbesondere im politischen Bereich zu einer Verhärtung der eigenen politischen Position führen könnte. Nach Pariser benötigt eine solche Filterblase vier Säulen, um entstehen und wirken zu können[45]:

---

[37] Google: Google Acquires Kaltix Corp, 2003, http://googlepress.blogspot.com/2003/09/google-acquires-kaltix-corp.html, 01.12.2018

[38] Matt Hines: Google takes searching personally, 2004, https://www.cnet.com/news/google-takes-searching-personally/, 01.12.2018

[39] Google: Personalized Search Graduates from Google Labs, 2005, http://googlepress.blogspot.com/2005/11/personalized-search-graduates-from_10.html, 01.12.2018

[40] Google: Personalized Search for everyone, 2009, https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html, 01.12.2018

[41] Wir verwenden die von uns erfassten Daten, um unsere Dienstleistungen für Sie anzupassen, einschließlich der Bereitstellung von Empfehlungen, personalisierten Inhalten und benutzerdefinierten Suchergebnissen.

[42] Google: Google privacy policy, 2018, https://policies.google.com/privacy?hl=en, 01.12.2018

[43] Pariser, E.: The filter bubble: What the Internet is hiding from you, 2011, Penguin UK

[44] Ebd. Seite 9

[45] Ebd.

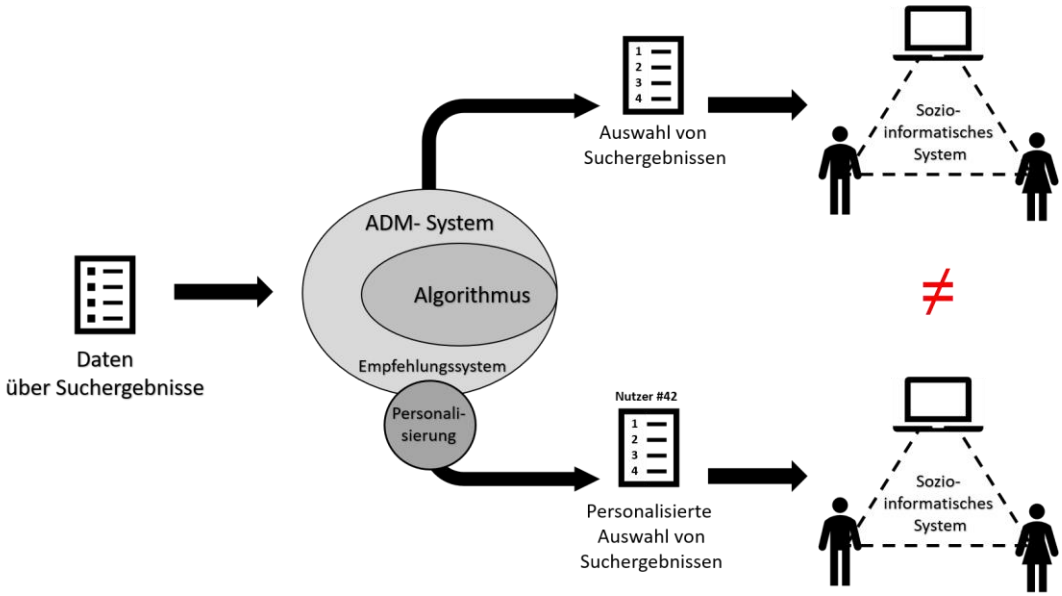

Abbildung 4: Veränderung im ADM-System verändern das Sozioinformatische System.

Zunächst müssen VerbraucherInnen sich lediglich über ein oder mehrere vollständig gleich ausgerichtete Medien informieren (I). Dieses Medium muss die VerbraucherInnen personalisieren (II), also auf Grundlage der individuellen Präferenzen vollständig anders informieren (III). Schließlich muss es sich bei den unterschiedlichen Inhalten noch um kritische Informationen, wie beispielsweise politische Meinungen handeln (IV).

Unser Projekt zur Bundestagswahl 2017[46] hat zur Überprüfung einer solchen personalisierten Suchmaschine eine Methodik vorgestellt, mit deren Hilfe die Gesellschaft einfach und sehr effektiv Punkt II, also den tatsächlichen Personalisierungsgrad einer Suchmaschine, überwachen kann. Bei dieser Studie konnten wir nachweisen, dass der momentane Personalisierungsgrad zu gering ist, um effektiv einen Filterblaseneffekt nach Pariser auszulösen. Solange eine Black-Box-Analyse diesen geringen Personalisierungsgrad fortlaufend für eine repräsentative Nutzermenge und für die unterschiedlichsten Suchanfragen attestiert, ist das Gesamtschadenspotenzial geringer als wenn die Personalisierung erhöht wird. Ein ADM-System, das für Suchanfragen auf Grundlage von Relevanz und anderen Daten für den gesamten Markt eine Auswahl von Suchergebnissen produziert, muss von der Gesellschaft deutlich weniger überprüft werden als ein ADM-System, welches zusätzlich Persönlichkeitsmerkmale und Präferenzen berücksichtigt und somit für jeden Verbraucher eine personalisierte Suchergebnisliste zusammenstellt (siehe Abbildung 4). Hier sind weitreichendere und anhaltende Transparenz- und Nachvollziehbarkeitsforderungen, beispielsweise durch die oben angemerkte Überwachung des Personalisierungsgrades notwendig.

Das Beispiel zeigt zweierlei: Wenn die Bevölkerung die Möglichkeit hat, wesentliche Charakteristika eines ADM-Systems zu überprüfen, kann dies erstens das Vertrauen in das System erhöhen und zweitens dabei helfen, das Gesamtschadenspotenzial zu verringern.

---

[46] Deutsch: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig: Wer sieht was? Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine, 2018 , https://www.blm.de/files/pdf2/bericht-datenspende---wer-sieht-was-auf-google.pdf, 01.12.2018 English: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig: What did you see? Personalization, regionalization and the question of the filter bubble in Google's search engine, 2018 , https://arxiv.org/abs/1812.1094301.12.2018

2. BEWERTUNG DER GEFAHR EINES SOZIOINFORMATISCHEN SYSTEMS ANHAND EINER RISIKOMATRIX

Wie anhand der zuvor genannter Beispiele ausgeführt, sind die Folgen fehlerhafter Entscheidungen je nach sozioinformatischem System, in dem ein ADM-System Anwendung findet, von unterschiedlicher Tragweite. Es spielen hierbei das mögliche Schadenspotenzial auf individuellem und gesamtgesellschaftlichem Level sowie die Möglichkeit der Re-Evaluation bei einer Fehlbeurteilung eine entscheidende Rolle. Das Risiko eines ADM Systems steigt offensichtlich mit dem Schadenspotenzial aber auch mit den fehlenden Einspruchs- und Zweitbewertungsmöglichkeiten.

Wie oben schon skizziert, besteht das Schadenspotenzial dabei aus zwei Anteilen: Dem Gesamtschaden aller Individuen plus einem möglichen, superlinearen gesellschaftlichen Gesamtschaden. Für den Verbraucher ist beispielsweise das individuelle **Schadenspotenzial** eines falschen Kreditscorings existenziell hoch, wenn hierdurch der geplante Eigenheimbau nicht zustande kommen kann, obwohl eine regelmäßige Rückzahlungsmentalität vorliegt. Der gesellschaftliche Schaden durch ein fehlerhaftes Bewertungssystem ist aber ebenfalls beträchtlich: Werden Verbraucher zu gut bewertet, werden die Banken auf die Dauer keine Kredite mehr vergeben; werden Verbraucher zu schlecht bewertet, können ökonomische Chancen nicht realisiert werden.

Bei anderen Fällen ist eines der beiden Schadenspotenziale stärker ausgeprägt: Während Upload-Filter auf individueller Ebene vor allen Dingen Unsicherheit verursachen, könnte ein größerer gesellschaftlicher Schaden durch ein Fehlurteil bei den aktuell diskutierten Upload-Filtern auftreten. Der umstrittene Artikel 13 wurde am 12.09.2018 vom Europäischen Parlament[47] verabschiedet und macht alle *außer* „den kleinsten Internetplattformen haftbar für jegliche Urheberrechtsverletzungen ihrer Nutzer“[48] [49]. Dies führt direkt zur Verpflichtung, jeden Kommentar, jedes hochgeladene Bild durch automatische Filter auf einen solchen urheberechtlichen Verstoß überprüfen zu lassen (Upload-Filter). Da solche Filter aus informatischer Sicht sehr fehleranfällig sind und die Firmen durch die angedrohten Strafen kein Risiko eingehen werden, könnte das Internet in seiner jetzigen Form als offene Informations- und Diskursplattform zunehmend zensiert werden[50]. Obwohl also dem einzelnen Verbraucher lediglich die Partizipationsmöglichkeit durch einen gefilterten Kommentar entzogen wird, entsteht der Gesellschaft als Ganzes dadurch ein möglicherweise großer gesamtgesellschaftlicher Schaden. Es ist zusätzlich zu bedenken das ein weiteres zu bedenkendes Schadenspotential in einer Verhaltensänderung des Verbrauchers durch den Einsatz eines ADM-Systems entstehen kann. Je nach

---

[47] Abstimmung des Europaparlament: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+PV+20180912+RES-VOT+DOC+PDF+V0//DE&language=DE

[48] Julia Reda: Europäisches Parlament befürwortet Uploadfilter und EU-Leistungsschutzrecht, 2018, https://julia-reda.eu/2018/09/ep-pro-uploadfilter-leistungsschutzrecht/, 01.12.2018

[49] Zum Zeitpunkt dieses Gutachtens steht die Trilog-Verhandlungen noch aus, weshalb die zivilgesellschaftlichen und die wissenschaftlichen Bedenken noch Einfluss nehmen können, zumal der Artikel 13 lediglich mit 366 Stimmen dafür, 297 dagegen beschlossen wurde.

[50] Heldt, A.P: Intelligente Upload-Filter: Bedrohung für die Meinungsfreiheit? In: Mohabbat Kar, Resa (Ed.); Thapa, Basanta E. P. (Ed.); Parycek, Peter (Ed.); Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, Kompetenzzentrum Öffentliche IT (ÖFIT) (Ed.): (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft. Berlin, ISBN 978-3-9818892-5-3, 392-416

dem ob dieses Verhalten gewünscht ist, also einer Art Nudging gleich kommt[51] oder beispielsweise Teilhabe verhindert wird.

Ähnlich starken Einfluss auf das Risikopotenzial eines sozioinformatischen Systems hat die **Re-Evaluierungsmöglichkeit** der Systeme. Anzustreben wäre, dass verschiedene Anbieter oder Dienstleister mit ähnlichem Angebot ein unterschiedliches ADM-System nutzen. So könnten Verbraucherinnen und Verbraucher bei einer Fehlentscheidung eines ADM-Systems auf ein anderes ausweichen.

Ein Paradebeispiel für eine solche Diversität bietet der aktuelle Markt an Empfehlungssystemen in der Art „Kunden, die dieses Produkt gekauft haben, kauften auch …". Eine solche Kaufberatung bietet aktuell fast jedes Onlinewarenhaus an und berechnet nach unterschiedlichsten Faktoren eine persönliche Produktempfehlung. Auch wenn die Anbieter natürlich jeweils unterschiedliche Informationsstände zu einer Verbraucherin oder einem Verbraucher haben, steht dem Verbraucher die Möglichkeit zur Verfügung, bei einer offensichtlichen Fehleinschätzung, Vorschläge von einem anderen ADM-System einzuholen und das erste Ergebnis zu evaluieren.

Erhöhte Aufmerksamkeit muss hingegen monopolartigen Strukturen gewidmet werden, wenn Bürgerinnen und Bürger nicht mehr die Auswahl zwischen verschiedenen Anbietern haben, durch die eine Dienstleistung in Anspruch genommen werden kann. In solchen Fällen wird bei einem Fehlurteil die Re-Evaluierung erschwert. Ein starker Monopolisierungsgrad kann jedoch auch durch Maßnahmen der Anfechtbarkeit eines Ergebnisses gemildert werden. So ist zwar die Lohnsteuerjahresausgleichsprüfung in manchen Bundesländern schon automatisiert und natürlich monopolisiert, aber durch Einsprüche kann eine Nachbewertung durch einen menschlichen Sachbearbeiter jederzeit erreicht werden. Die Zweitbewertung muss also nicht notwendigerweise durch andere ADM-Systeme geschehen, sondern darf – und sollte – in wichtigen Fällen auch durch menschliche Re-Evaluation erfolgen.

Mithilfe einer Risikomatrix (siehe Abbildung 5), basierend auf den zwei Merkmalen (Schadenspotenzial durch Fehlurteile, Re-Evaluierungsmöglichkeit), lassen sich die Anwendungsszenarien von ADM-Systemen leicht verorten, sodass man schnell einen ersten Überblick über mögliche Risiken des Systems bekommt.

[51] Das Gutachten des Sachverständigenrat für Verbraucherfragen geht im Zuge von Scoringverfahren im Gesundheitsbereich detailliert auf diese Problematik ein: Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf, 01.12.2018

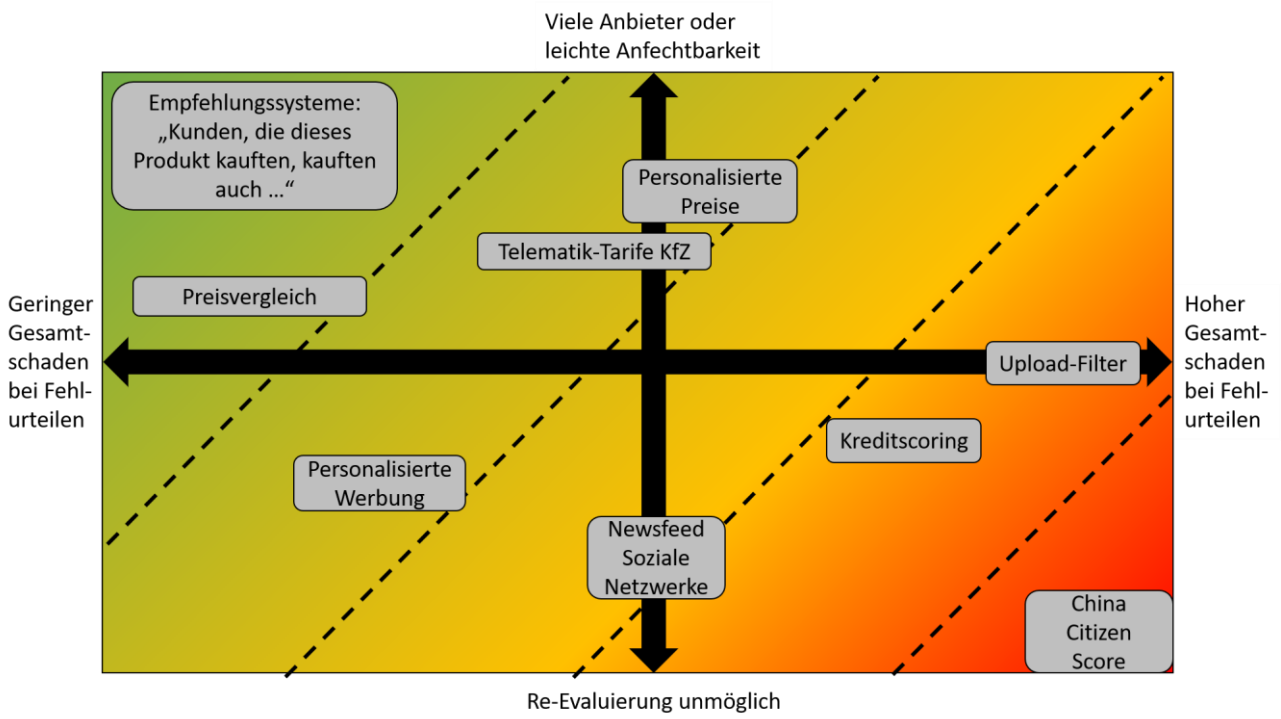

Abbildung 5: Risikomatrix zur Verortung von Anwendungsszenarien von ADM-Systemen, um auf Gefahren und Risiken innerhalb dieser Szenarien zu schließen. Der gewählte Farbverlauf von Grün (oben links, geringes Risiko) über Gelb bis hin zu Rot (unten rechts, hohes Risiko) verdeutlicht das Risikopotenzial der eingeordneten Szenarien.

Hierbei wären beispielsweise Produktempfehlungssysteme in der oberen linken Ecke verortet, da der Verbraucher in allen uns bekannten Fällen problemlos auf eine andere Verkaufs- bzw. Anbieterplattform wechseln kann, die Re-Evaluierungsmöglichkeit also jederzeit durch einen Wechsel des Anbieters gegeben ist. Weiterhin ist von keinen nennenswerten individuellen Schäden oder einem gesamtgesellschaftlichen Schaden durch Fehlurteile von Produktempfehlungssystemen insgesamt auszugehen, so dass diese auf der horizontalen Skala sehr weit links stehen. Die dort grün hinterlegte Farbe signalisiert Entwarnung aufgrund des zu erwartenden minimalen Risikopotenzials – gleichzeitig sinken damit auch die Anforderungen an Transparenz und Nachvollziehbarkeit dieser ADM-Systeme.

Im Gegensatz dazu stellt sich der China Citizen Score, der ab 2020 in der Volksrepublik China Anwendung finden soll[52], in beiden Merkmalen als höchst problematisch dar. Generell werden in diesen Score die verschiedensten personenbezogenen Daten, wie z. B. Wohnort, Alter und Beruf einfließen, genauso aber auch sensible Daten verarbeitet. Dazu gehören das Rückzahlverhalten im Kreditsektor, die Art der sozialen Kontakte oder Straftatbestände. Auch Bewegungsprofile, deren Erstellung eine zunehmende Kameraüberwachung im öffentlichen Sektor ermöglicht, werden eingespeist. Der anhand der gesamten gesammelten Daten algorithmisch ermittelte Score gibt Aufschluss über das staatskonforme Verhalten jedes Bürgers. Ziel ist dabei die umfassende Gesinnungs- und Verhaltenskontrolle der Bevölkerung durch den Staat. Bei ‚erwünschtem‘ Verhalten steigt

[52] Rogier Creemers: Planning Outline for the Construction of a Social Credit System (2014-2020), 2015, https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/, 15.12.2018

der Score und der Bürger wird durch Vergünstigungen belohnt, ein niedriger Score hingegen wird staatlich sanktioniert, sodass ihm Möglichkeiten verwehrt werden, wie z. B. internationale Reisen oder die Erlaubnis, bestimmte Beruf auszuüben[53].

Die Testphase, während der verschiedene Anbieter ihre selbst entwickelten Scores, live mit der Bevölkerung testen können, läuft seit 2014. Sobald diese Testphase beendet ist, wird sich China nach aktuellen Plänen für einen Score entscheiden. Aus demokratietheoretischer Sicht sind mit einem solch weitreichenden Score, der bürgerliches und privates Verhalten bewertet, gesamtgesellschaftliche Schäden zu erwarten, die weit über die Summe der Schäden der Fehlurteile bei den Individuen hinausgeht. Das System ermöglicht einen digitalen Totalitarismus mit bisher unbekanntem Detailgrad. Aus der Perspektive demokratischer Staaten wäre also sowohl das Schadenspotenzial enorm, als auch die Monopolstellung der Bewertung eindeutig und eine Re-Evaluierung ausgeschlossen. Damit würde das System rechts-unten in der Risikomatrix verortet.

Die Beispiele wurden bislang zur Veranschaulichung grundlegender Analyseschritte bewusst plakativ und eindeutig gewählt. Es muss jedoch darauf hingewiesen werden, dass die Verortung eines ADM-Systems in seinem spezifischen sozioinformatischen System auch mithilfe der beiden Skalen mitunter nicht ganz so einfach ist. Abhängig vom subjektiven Standpunkt kann es durchaus zu voneinander abweichenden Positionierungen kommen. So wäre der Citizen Score beispielsweise aus chinesischer Sicht deutlich weiter links einzuordnen, da das chinesische Regime in dem Score natürlich einen großen gesamtgesellschaftlichen Nutzen sehen würde, der den möglichen Schaden bei weitem überwiegt.

Deshalb schlagen wir die Durchführung eines Workshops vor, in dem diese Einordnung mit allen wichtigen Akteuren durchgeführt und im Detail erläutert wird. In diesem Workshop können die Teilnehmer durch eine Vertiefung des Verständnisses um die möglichen Transparenz- und Nachvollziehbarkeitsmethoden, die wir im folgenden Kapitel an die Hand geben, die praktische Umsetzung erarbeiten. Zielführend ist es, den einzelnen Entscheidern den Prozess zu vereinfachen und die persönlich wahrgenommene Ausprägung auf den beiden Skalen bestmöglich einzuordnen. Dabei ist es besonders wichtig, einen Prozess zu definieren, bei dem die Gruppenmitglieder sich am Ende weitestgehend einig sind in der Verortung des ADM Systems.

[53] Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... & Zwitter, A. (2017). Digitale Demokratie statt Datendiktatur. In Unsere digitale Zukunft Springer, Heidelberg, Seite 6

V. TRANSPARENZ- UND NACHVOLLZIEHBARKEITSFORDERUNGEN IN DEN VERSCHIEDENEN REGULIERUNGSKLASSEN

Welche Schlussfolgerungen ergeben sich aus der Verortung eines ADM-Systems in der Risikomatrix? In den Gutachten des Sachverständigenrats für Verbraucherfragen[54] und der Gesellschaft für Informatik[55] wird nur wenig differenziert zwischen den einzelnen Risikopotentialen der Systeme. In den Diskussionen, die wir mit Expertinnen und Experten in den letzten Jahren geführt haben, erwies sich eine Einteilung der Systeme in fünf mögliche Regulierungsklassen als bisher stabil (siehe Abbildung 6). Jede Regulierungsklasse ist assoziiert mit einer Reihe von notwendigen Transparenz- und Nachvollziehbarkeitsforderungen, die den im sozioinformatischen Gesamtsystem ausgelösten Risikopotenzialen entgegenwirken. Insofern sollen die hier aufgeführten Transparenzpflichten und Kontrollprozesse als Anhaltspunkt dienen, was sinnvollerweise im Hinblick auf Transparenz und Nachvollziehbarkeit in Erwägung gezogen werden müsste.

Es ist wichtig festzuhalten, dass sich diese Regulierungsklassen jeweils inkludieren, das heißt, für ein System in Klasse 3 sollten auch alle Forderungen aus den Klassen 0-2 gestellt werden.

Die fünf Regulierungsklassen sind wie folgend gestaffelt:

1) In der Klasse 0 werden keine Transparenzpflichten gefordert und keine Kontrollprozesse dauerhaft installiert. In Verdachtsfällen kommt es zu einer post-hoc-Analyse und die Risikobewertung muss unter Umständen erneut durchgeführt werden.

2) In der Klasse 1 werden erste Transparenzpflichten gefordert. Eine Schnittstelle zur Analyse des Systems als Blackbox muss angeboten werden, sowie die Einbettung des ADM-Systems in den sozialen Entscheidungsprozess beschrieben werden.

3) In der Klasse 2 müssen die Eingangsdaten vollständig (gegenüber einem zu bestimmenden Publikum) beschrieben werden und die Angaben zur Qualität des Entscheidungssystem müssen überprüfbar sein.

4) In der Klasse 3 müssen alle Angaben mindestens für ein Expertengremium in angemessener Zeit nachvollziehbar und überprüfbar sein. Dazu sind verschiedene Interfaces zu den Eingangsdaten und den Resultaten der Maschine notwendig.

---

[54] Sachverständigenrat für Verbraucherfragen: Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf, 01.12.2018

[55] Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf, 14.12.2018

5) Lernende ADM-Systeme, deren Bewertung sie in Klasse 4 sieht, sollten aufgrund eines zu hohen Risikos entweder gar nicht eingesetzt werden oder nur, wenn dem beweisbar ein genügend hoher Gesamtnutzen entgegensteht.

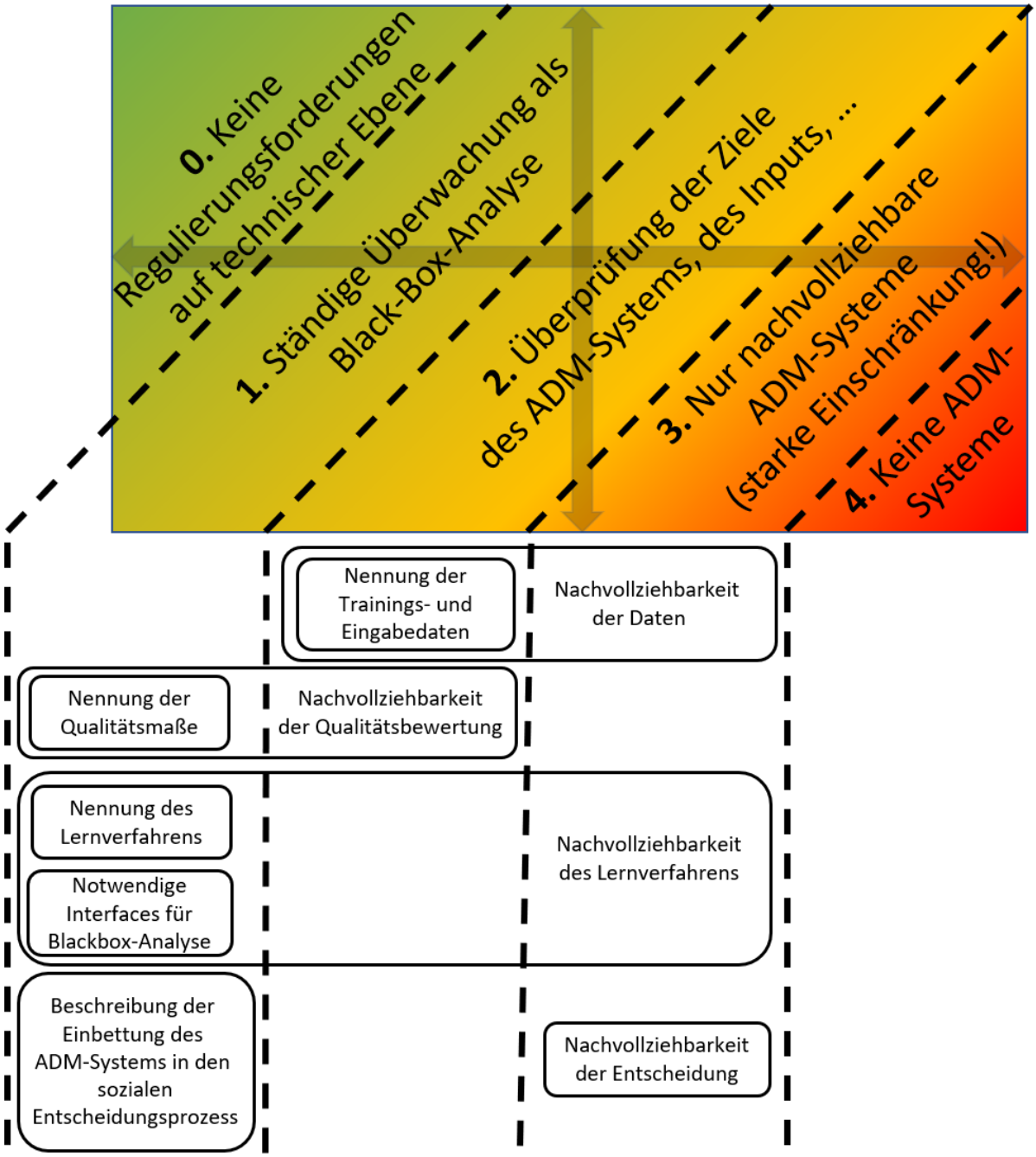

Abbildung 6: Die einzelnen Regulierungsklassen rufen unterschiedliche Transparenz- und Nachvollziehbarkeitsforderungen auf, die in dieser Abbildung zugeordnet werden.

Diese Einteilung ist notwendigerweise zuerst abstrakt. Anhand einer Reihe von Beispielen aus der großen Klasse der *Empfehlungssysteme* (vgl. Abbildung 2) zeigen wir im Folgenden, ab wann ein ADM-System in die nächst risikoreichere Klasse aufsteigt (siehe Abbildung 6).

Hierzu wird in den folgenden Unterkapiteln jeweils ein Empfehlungssystem herausgegriffen, welches in die jeweilige Regulierungsklasse fällt.

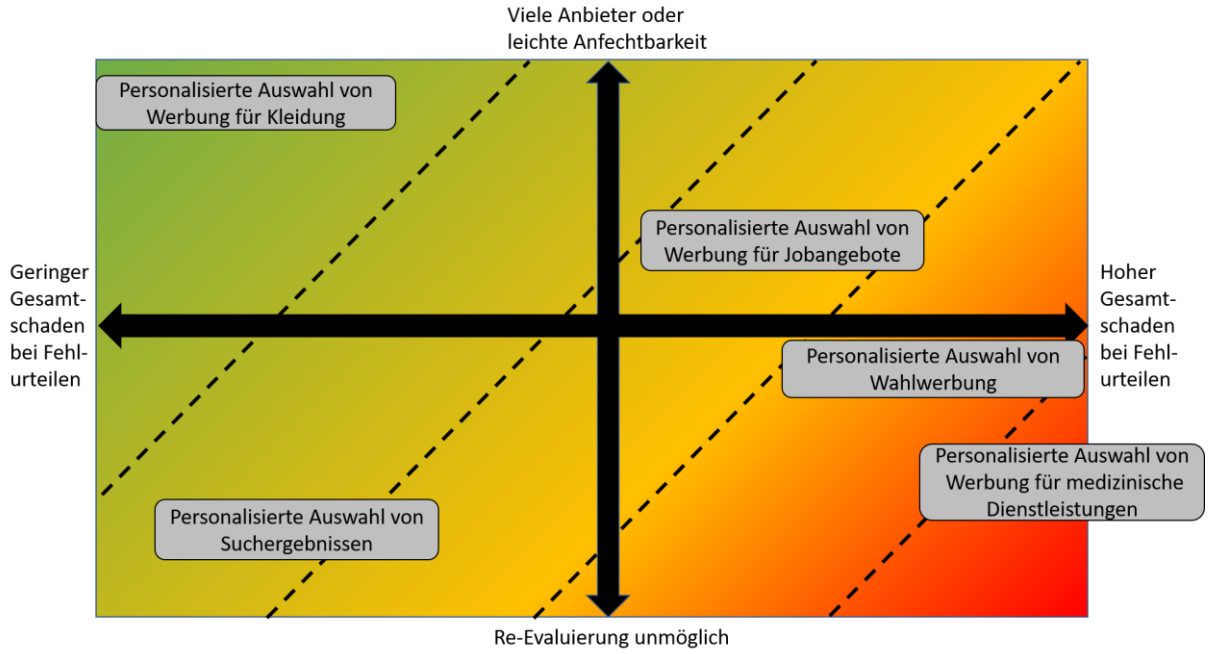

Abbildung 7: Verschiedene Anwendungsgebiete von Empfehlungssystemen. Durch Integration in verschiedene sozioinformatische Systeme verteilen sie sich über die gesamte Risikomatrix und rufen jeweils unterschiedlich starke Transparenz- und Nachvollziehbarkeitsforderungen auf.

KLASSE 0: KEINE REGULIERUNGSFORDERUNGEN AUF TECHNISCHER EBENE

Ein Empfehlungssystem, das zur personalisierten Auswahl von Werbung für Kleidung genutzt wird, ist ein eingängiges Beispiel für ein Klasse-0 ADM-System. Ein solches ADM-System bedarf durch das geringe Schadenspotenzial keiner dauerhaften Überwachung oder genauen Analysen. Auch wenn das Gutachten der Gesellschaft für Informatik[56] eine generelle Forderung zur Anmeldung von jeglichen ADM-Systemen bei einer noch genauer zu spezifizierenden Kontrollinstitution fordert, sehen wir bei Klasse-0-Systemen dafür im Allgemeinen noch keinen Bedarf.

Natürlich kann es aufgrund unbekannter Schadenspotenziale z. B. durch nicht vorhergesehene Nebenwirkungen eines ADM-Systems doch zu Schäden kommen. Bei Verdachtsmomenten, wenn beispielsweise das ADM-System bei einer bestimmten Personengruppe signifikant teurere oder überteuerte Artikel bewirbt als bei anderen Gruppen, sollte eine sogenannte post-hoc-Analyse durchgeführt werden. Der Begriff „post-hoc-Tests“ bezeichnet Tests, die nicht zur Designzeit schon angedacht waren. In diesem Fall sind es also Überprüfungen im Nachhinein, die die Öffentlichkeit oder dafür geeignete Kontrollinstitutionen nicht von Anfang in Erwägung gezogen haben, für die es jetzt aber einen Verdachtsfall gibt.

[56] Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf, 14.12.2018

Hierbei würde man zum Beispiel die durchschnittlichen Preise für die beiden beworbenen Personengruppen berechnen und auswerten, ob es sich bei dem Unterschied um eine statistische Abweichung oder um eine tatsächliche Ungleichbehandlung handelt. Sollten durch einen post-hoc-Test Schäden für die Verbraucher identifiziert werden können, kann dann entsprechend anderer Verbrauchertäuschungen verfahren werden. Das Schadenspotenzial des ADM-Systems sollte nach einer positiven post-hoc-Analyse neu evaluiert werden. Dadurch könnte es in die nächsthöhere Klasse kommen. Dementsprechend müssten die Betreiber des Empfehlungssystems darauf reagieren und zusätzlich Schnittstellen bereitstellen, um dieses ADM-System beispielsweise als Blackbox zu analysieren.

KLASSE 1: STÄNDIGE ÜBERWACHUNG ALS BLACKBOX-ANALYSE

Die von Nicholas Diakopoulos 2014 übertragende Methode der Blackbox-Analyse[57] stellt eine Möglichkeit dar, ein algorithmisches Entscheidungssystem ohne direkten Zugang zur eigentlichen Code-Basis, also ohne tiefergehendes Verständnis, was innerhalb des Systems passiert, zu analysieren. So kann das System dahingehend überprüft werden, ob auffällige Effekte, wie vermeintliche Diskriminierung von einzelnen Personengruppen tatsächlich auf fehlerhafte Entscheidungen des ADM-Systems zurückgeführt werden können. Als Faustregel für Systeme, bei denen eine solche Analyse notwendig ist, gilt:

ADM-Systeme, die in einem sozialen Entscheidungsprozess eingesetzt werden, der rechtlichen Anforderungen unterliegt, müssen eine Überwachung in Form von Blackbox-Analysen ermöglichen und liegen damit mindestens in Klasse 1.

Eine Online-Suchmaschine, die für jeden Nutzer zu einer Suchanfrage eine personalisierte Suchergebnisreihenfolge erstellt, basiert auf einem Empfehlungssystem. Wie bereits in Kapitel IV.1.3 beschrieben, stellt Googles Suchmaschine ein solches ADM-System zur Verfügung. Im Sommer 2017 wurde von AlgorithmWatch und dem Algorithm Accountability Lab der TU Kaiserslautern eine Studie mit dem Namen "Datenspende #BTW2017" durchgeführt. Für diese Projekt spendeten über 4000 Bürgerinnen und Bürger ihre Suchergebnisse zu einer Reihe von Suchanfragen. Das Ergebnis zeigte, dass Google aktuell kaum bis keine personalisierten Suchergebnisse ausliefert[58]. Solange dies so ist, schätzen wir das Schadenspotenzial der Suchmaschine als relativ gering ein. Zudem gibt es mit Bing, Yahoo und anderen Suchmaschinen einen Markt, auch wenn dieser nicht allzu groß ist. Damit verorten wir Googles Suchmaschine – unter dem Aspekt der Filterblasenbildung und bei geringem Personalisierungsgrad – in Klasse 1.

---

[57] Diakopoulos, N.: Algorithmic-Accountability: the investigation of Black Boxes, 2014, https://www.cjr.org/tow_center_reports/algorithmic_accountability_on_the_investigation_of_black_boxes.php, 01.12.2018

[58] Deutsch: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig: Wer sieht was? Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine, 2018 , https://www.blm.de/files/pdf2/bericht-datenspende---wer-sieht-was-auf-google.pdf, 01.12.2018 English: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig: What did you see? Personalization, regionalization and the ques-tion of the filter bubble in Google's search engine, 2018 , https://arxiv.org/abs/1812.1094301.12.2018

Klasse 1-Systeme müssen die folgenden Transparenzpflichten erfüllen:

1.1 Nennung der Qualitätsmaße

Bei der Bewertung der Güte eines algorithmisches Entscheidungssystems, also der Frage, wie gut die getroffenen Entscheidungen sind, handelt es sich meist um eine statistische Auswertung, wie gut das System „gelernt“ hat, aus den ihm zur Verfügung stehenden Daten die richtige Prognose zu treffen. Je nach Fall stehen über 20 verschiedene Verhältnisse und Kennzahlen zur Verfügung, die unterschiedliche Fokusse bei der Optimierung setzen können. Einerseits gibt die Wahl des Optimierungskriteriums also Einblicke in die vom Betreiber gewählten Ziele des ADM-Systems aber erst die genaue Angabe des Qualitätsmaßes und des Verfahrens dahinter ermöglicht es, die Angemessenheit der Qualitätsgüte, also auch der erreichten Güte für den Anwendungsfall zu beurteilen.

Transparenz in diesem Kontext gibt also Auskunft über das für die Systemevaluation verwendeten Qualitätsmaß und wie gut das System dieses erfüllt.

1.2 Nennung des Lernverfahrens

Algorithmische Entscheidungssysteme können die verschiedensten Verfahren aus dem maschinellen Lernen nutzen, um aus der zur Verfügung gestellten Datenmenge ein statistisches Modell zu entwickeln. Dieses Modell beinhaltet die tatsächlichen Entscheidungskriterien, die später Anwendung finden. Da jede Methode, wie beispielsweise logistische Regression, Entscheidungsbaum bzw. -bäume oder neuronale Netze weitreichende Annahmen über Zusammensetzung der Daten, Fehler in den Daten und die Güte des gelieferten Modells stellt, kann nur bei Nennung des verwendeten Lernverfahrens überprüft werden, ob diese Annahmen im spezifischen sozioinformatischen Gesamtsystem gelten oder sichergestellt sind.

1.3 Beschreibung der Einbettung des ADM-Systems in den sozialen Entscheidungsprozess

Auf der Frage in welchen sozialen Entscheidungsprozess das algorithmische Entscheidungssystem wie eingebettet ist, liegt der Fokus dieser Forderung. Es geht also um Transparenz hinter der Entscheidung: Wer trifft die eigentliche Entscheidung? Wird das ADM-System entscheidungsunterstützend eingesetzt oder liefert es für jedes zu bewertende Objekt oder Individuum eine binäre Zuweisung zu einer Klasse und trifft somit quasi die finale Entscheidung? Zusätzlich ist zu erklären, ob das System eine Entscheidung mit einer Erklärung, die über die reinen Daten des Individuums hinausgeht, liefert oder nicht.

Zudem müssen sie für die notwendigen Kontrollen die folgenden Interfaces bereitstellen:

1.4 Notwendige Interfaces für Blackbox-Analysen

Für die Untersuchung eines algorithmischen Entscheidungssystems als Blackbox muss es möglich sein, das System mit variablen Daten zu füttern und diese mit den gelieferten Ergebnissen des ADM-Systems auswerten zu können. Aus der Relation von Ein- und Ausgabe lassen sich Aussagen über das System überprüfen. Die Studie der Gesellschaft für Informatik zum Thema „Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren“[59] differenziert hierbei verschiedene Auditformen, also Herangehensweisen, um ein solches ADM-System mit oder ohne Beteiligung der Nutzer untersuchen zu können und ohne direkten Einblick in das Innenleben des ADM-Systems zu benötigen. Wir verweisen deshalb auf die dortige Beschreibung der Auditformen „Nichtinvasive“, „Scraping“, „Sock Puppet“ und „Crowdsourced“[60] und erwähnen nur ergänzend, dass das von uns erarbeitete Verfahren zur Datenspende 2017[61] von diesen Methoden abweicht.

Die Liste möglicher Herangehensweisen ist also keineswegs abschließend, sondern lediglich richtungsweisend, was wir unter einer Blackbox-Analyse verstehen. Eine solche muss jedoch immer auf das ADM-System in seinem konkreten sozioinformatischen System angepasst werden.

Praxisanwendung

Es ist instruktiv, sich anzusehen, ob Google als Betreiber schon heute die Anforderungen der Klasse 1 erfüllen würde. Dies diskutieren wir im Folgenden:

Da es in diesem Fall unmöglich ist, zu ermitteln, was die tatsächliche „richtige“ Ergebnisauswahl und Sortierung einer personalisierten Suchergebnisliste ist, ist positiv zu notieren das Google relativ offen kommuniziert, wie das Bewertungsverfahren aussieht. Google selbst stellt dafür sogenannte “Search Quality Evaluators” ein, welche Suchergebnislisten anhand eines selbst erarbeiteten „Search Quality Rating“ (Bewertung der Suchqualität) bewerten sollen. Hier arbeitet Google sehr transparent und geht über die für diese Regulierungsklasse notwendige Forderung einer reinen Nennung des Qualitätsmaßes (siehe Abschnitt 1.1) hinaus, indem sowohl in einem ausführlichen Dokument[62] (über 150 Seiten) beschrieben wird, wie diese Bewertung funktioniert und zusätzlich Einblicke in die genutzte Bewertungsplattform und den Evaluierungsprozess gewährt werden. Einzig das Ergebnis des Bewertungsprozesses hält Google unter Verschluss.

---

[59] Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf, 14.12.2018

[60] Ebd. Seite 66ff.

[61] Deutsch: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig: Wer sieht was? Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine, 2018 , https://www.blm.de/files/pdf2/bericht-datenspende---wer-sieht-was-auf-google.pdf, 01.12.2018 English: Tobias D. Krafft, Michael Gamer, Katharina A. Zweig: What did you see? Personalization, regionalization and the question of the filter bubble in Google's search engine, 2018 , https://arxiv.org/abs/1812.1094301.12.2018

[62] Google: Search Quality Guidelines, 2018, https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf, 01.12.2018

Über das aktuell verwendete Lernverfahren allerdings (siehe Abschnitt 1.2) und zur Frage, wie das verwendete statistische Modell zur Erstellung solcher individueller Ergebnislisten entsteht und funktioniert, gibt es von Google kaum eine Aussage.

Das oben genannte Datenspende-Projekt wäre deutlich einfacher gewesen, wenn Google Schnittstellen, sogenannte Interfaces angeboten hätte, um das ADM-System wissenschaftlich überprüfen zu können, aber es war auch so mit relativ geringem Aufwand möglich (siehe Abschnitt 1.3). Denn Google äußert sich zur Einbettung des algorithmischen Entscheidungssystems in den Entscheidungsprozess (siehe Abschnitt 1.4) sehr offen, was die Überprüfung vereinfacht hat. Das Unternehmen klärt auf, dass der gesamte Prozess vom Crawling, also der Abtastung des Internets, über die Indizierung der gefundenen Webseiten bis hin zur Bereitstellung der auf den Nutzer zugeschnittenen Suchergebnisliste vollständig automatisiert abläuft[63].

KLASSE 2: ÜBERPRÜFUNG DER ZIELE DES ADM-SYSTEMS

Bei ADM-Systemen mit gravierenden Schadenspotenzialen bei Fehlurteilen sowie eingeschränkten Möglichkeiten einer Re-Evaluation reicht eine reine Blackbox-Analyse nicht aus, es müssen aktiv die Ziele des ADM-Systems überprüft werden und klar kommuniziert werden, welche Trainings- und Eingabedaten verwendet werden, damit eine Nachvollziehbarkeit der Ergebnisse ermöglicht wird.

Personalisierte Auswahl von Werbung für Jobangebote, wie sie aktuell durch Onlineplattformen wie Stepstone[64] angeboten werden, nutzen Empfehlungssysteme, welche aus unserer Sicht unter diese Regulierungsklasse fallen. Dies begründen wir vor allen Dingen mit möglichen Diskriminierungen, die darauf beruhen könnten, dass gewisse Jobs beispielsweise nur jungen Personen angeboten werden – ein möglicher Verstoß gegen die Altersdiskriminierung, wenn dafür nicht gute Gründe vorliegen.

Klasse 2-Systeme müssen der folgenden Transparenzpflicht nachkommen:

2.1 Nennung der Trainings- und Eingabedaten

Da die verwendeten Lernverfahren das statistische Modell, das die eigentlichen Entscheidungsregeln des späteren ADM-Systems repräsentiert, aus den zur Verfügung gestellten Trainingsdaten lernen, ist es wichtig, darüber Auskunft zu erteilen. Trainingsdaten bestehen immer aus einer Menge an Eigenschaften einer zu bewertenden Person oder eines zu bewertenden Objekts verknüpft mit der tatsächlichen Eigenschaft oder dem tatsächlichen Verhalten, das durch das ADM-System vorhergesagt werden soll. In dieser Regulierungsklasse reicht es, die verwendeten Eigenschaften eines zu bewertenden Objekts zu nennen und zu kommunizieren woher die tatsächlichen Instanzen dieser Daten kommen. Werden diese selbst gesammelt (wie beispielsweise bei Google) oder wurden andere Datenquelle herangezogen. Durch das Offenlegen der verwendeten Ei-

[63] Google: So funktioniert die Google-Suche, 2018, https://support.google.com/webmasters/answer/70897?hl=de, 01.12.2018

[64] https://www.stepstone.de/

genschaften kann sichergestellt werden das keine rechtlich fragwürdigen Variablen Einfluss auf die Entscheidung nehmen, wie beispielsweise das Geschlecht oder das Alter. Zusätzlich kann auf offensichtliche Proxy-Variablen hin überprüft werden, die eine fragwürde Eigenschaft, nach der sich die Entscheidung nicht richten sollte, abbilden[65].

Außerdem müssen sie für die notwendigen Kontrollen folgenden Zugang bereitstellen:

2.2 Nachvollziehbarkeit der Qualitätsbewertung

Um die Qualitätsbewertung nachzuvollziehen und sich nicht auf die vom Betreiber kommunizierten Werte zu verlassen, sind die Betreiber des ADM-Systems dazu aufgefordert, die Ergebnisse des ADM-Systems in einer Form bereitzustellen, sodass die überwachenden Instanzen die angegebenen Qualitätsmaße berechnen, also nachvollziehen können. Sollte es sich hierbei nicht um ein allgemein bekanntes Qualitätsmaß handeln, ist zusätzlich sicherzustellen, dass der gesamte Berechnungs- und Bestimmungsweg aufgedeckt wird, um diese nachzuvollziehen und zu überprüfen.

Praxisanwendung

Auch hier ist es wieder erkenntnisreich sich anzusehen, wie weit das Klasse 2-System Stepstone unsere Forderungen erfüllt. Bezüglich der Transparenzpflicht von Trainings- und Eingabedaten (siehe Abschnitt 2.1), sind uns keine Veröffentlichungen bekannt, die diese Informationen beinhalten und somit gibt es aktuell keine Möglichkeit, die verwendeten Daten auf Rechtmäßigkeit, beispielsweise im Rahmen der DSGVO hin, zu überprüfen. Stepstone sagt zwar in ihren FAQs, dass es für die Auffindbarkeit ihrer Anzeige bei der Stichwortsuche entscheidend sei, „im Anzeigentext alle wichtigen Schlüsselwörter für die Beschreibung der Stelle zu verwenden“[66], aber welche Informationen darüber hinaus einfließen oder von externen Quellen herangezogen werden, ist nicht bekannt.

In der 3. Regulierungsklasse wird erstmals zusätzlich zu Transparenz-Aspekten Nachvollziehbarkeit gefordert. Die überwachenden Institutionen müssen durch das Nachvollziehen der Qualitätsbewertung (siehe Abschnitt 2.2) eine aktive Rolle einnehmen, um die tatsächlichen Ziele des ADM-Systems zu überprüfen. Auch hierzu geben die Betreiber von Stepstone keine Möglichkeit, da weder das verwendete Qualitätsmaß noch wie gut das genutzte System in diesem abschneidet veröffentlicht ist. Auf die Überprüfung der Forderungen der niedrigeren Klassen haben wir an dieser Stelle verzichtet.

---

[65] Beispielsweise lässt sich aus der Bestellhistorie im Drogeriebereich schnell auf das Geschlecht schließen, Männer kaufen hier wahrscheinlich deutlich weniger Produkte zur Monatshygiene und dafür mehr Bartpflegeprodukte.

[66] Stepstone: FAQ, 2018, https://www.stepstone.de/stellenanbieter/kontakt/faq/ 01.12.2018

KLASSE 3: NUR NACHVOLLZIEHBARE ADM-SYSTEME

Wenn ein ADM-System in seinem sozioinformatischen Gesamtsystem bei Fehlentscheidungen ein von der Gesellschaft gesetztes Risikopotential überschreitet, müssen die Betreiber einer Reihe von Nachvollziehbarkeitsforderungen nachkommen.

Wahlwerbung ist beispielsweise ein Fall, bei dem der Zugang von Parteien durch Rechtsnormen und Gesetze wie dem öffentlich-rechtlichen Rundfunkstaatsvertrag (RStV), geregelt ist[67]. Dieser gleiche Zugang aller Parteien für die „mindestens eine Landesliste (…) zugelassen wurde“ (RStV, § 42 (2)) zu Slots im Fernsehen und Radio sollte sich auch auf Wahlwerbung im Internet transferieren lassen. Ein algorithmisches Entscheidungssystem, das also entscheidet, welche Wahlwerbespots an wen ausgeliefert werden, eröffnet ein immenses gesamtgesellschaftliches Schadenspotenzial, beispielsweise wenn (gewollt oder nicht) Fehlentscheidungen getroffen werden, indem der ausgespielte Wahlwerbespotmix nicht divers genug ist, also einzelne Parteien bevorzugt oder vernachlässigt werden würden. Eine Einordnung dieses ADM-Systems in Klasse 3 ist also gerechtfertigt.

Von allen Klasse 3-Systeme müssen vollständige Nachvollziehbarkeit gefordert werden, dies bezieht sich vor allem auf diese drei Bereiche:

3.1 Nachvollziehbarkeit der Daten

Um zu überprüfen, ob die verwendeten Datensätze zum Trainieren des ADM-System tatsächlich um diskriminierende Eigenschaften bereinigt wurden, muss ein Betreiber eines ADM-Systems in dieser Regulierungsklasse gegenüber einer überwachenden Instanz die verwendeten Trainingsdaten zur Verfügung stellen. Erst wenn die tatsächlich verwendeten Datensätze vorliegen kann sichergestellt werden ob die Qualität der verwendeten Daten der getroffenen Entscheidung gegenüber ausreichend ist. So muss überprüft werden, ob die beim Sammeln großer Datenmengen auftretenden Messfehler[68] ausreichend Bedacht wurden, also entweder herausgefiltert wurden oder zumindest innerhalb des Verarbeitungs- oder Entscheidungsprozess adäquat Beachtung geschenkt wurde.

3.2 Nachvollziehbarkeit des Lernverfahrens

Bei ADM-Systemen mit einem solchen Schadenspotenzial müssen Fehlerquellen und unerwünschte Entscheidungsmuster konsequent und verlässlich ausgeschlossen werden. Hierzu ist ein tatsächlicher Einblick in den Lernprozess des ADM-Systems und den darin wirkenden Code notwendig, da durch Blackbox-Analysen beispielsweise keine 100-prozentigen Aussagen getroffen werden können. Mit einer solchen Analyse ist lediglich möglich, aus den Relationen von Eingabedaten und Ergebnis des ADM-Systems statistische Prognosen über das wahrscheinliche Verhalten zu treffen.

Um eine Aussage über ein Systemverhalten zu treffen, ist es entweder nötig zusätzlich zum verwendeten Lernverfahren die dort angewandten Hyperparameter zu kommunizie-

[67] Detaillierte Informationen sind beispielsweise dem Merkblatt „Informationen für rheinland-pfälzische Rundfunkveranstalter bezüglich der Ausstrahlung von Wahlwerbung“ der Landeszentrale für Medien und Kommunikation (LMK) zu entnehmen was unter folgendem Link zu erreichen ist: https://lmk-online.de/fileadmin/user_upload/Bilder/01_Die_LMK/09_Weitere_Infos/Merkblaetter_und_Formulare/Merkblatt_Wahlwerbung_2014.pdf

[68] Viele dieser Fehler sind unter dem Begriff „Named-entity recognition“ zusammengefasst.

ren, also beispielsweise wie viele Ebenen das neuronale Netz hat welches man trainieren möchte, oder welche maximalen Tiefe der Entscheidungsbaum erreichen darf. Durch diese zusätzlichen Parameter ist es der kontrollierenden Instanz möglich mit den zur Verfügung stehenden Trainingsdaten den Lernprozesses tatsächlich zu wiederholen, ihn also nachzuvollziehen.

Eine andere Möglichkeit ist der im Gutachten der Gesellschaft für Informatik beschriebene tatsächliche Code Audit[69]. Hierbei wird der kontrollierenden Instanz eine beglaubigte Kopie des Programmcodes zu Prüfzwecken übergeben. Es ist darauf zu achten, dass nicht nur das fertig trainierte statistische Modell übergeben wird, sondern tatsächlich auch der Programmcode, welcher genutzt wurde um dieses zu trainieren. Bei der Formulierung einer solchen Forderung ist darauf zu achten, dass es einem Betreiber sehr leichtfällt, Programmcode in eine Form zu bringen, dass dieser zwar noch vollständig funktionstüchtig ist, jedoch selbst für einen erfahrenen Programmierer nicht mehr zu lesen oder bewerten ist. In Frankreich wurde 2017 eine Software zur Berechnung der Steuer veröffentlicht[70], die über 17 000 Variablen und etwa 1 000 Funktionen enthält. Als weiteres Hindernis wurde nur der Programmcode in einer selbst entwickelten Sprache namens "m" veröffentlicht, ohne einen Compiler oder Interpreter. So konnten die Wissenschaftler zunächst nicht mit einem Computer die Funktionalität des Programms überprüfen. Der gewünschte Compiler wurde erst von einem bundesweiten Hackathon[71] geliefert. Es ist daher klar, dass das Konzept der Transparenz mit Sensibilität behandelt werden muss und die Offenlegung des Algorithmus ohne Anforderung an die Verständlichkeit nicht unbedingt zu einer tieferen Einsicht führt.

3.3 Nachvollziehbarkeit der Entscheidung

In dieser Klasse sind die möglichen Schäden so hoch, dass hier einsortierte ADM-Systeme mit einer lernenden Komponente nur erklärbare Methoden des maschinellen Lernens einsetzen dürfen. Die Einteilung der Methoden in "erklärbare" und „nicht-erklärbare“ ist momentan noch Gegenstand der wissenschaftlichen Debatte. Einige Methoden sind aber so einfach, dass die meisten Wissenschaftler im Bereich der künstlichen Intelligenz sich darauf einigen würden, dass sie nachvollziehbare Entscheidungen schaffen. Hierzu gehören lineare und logistische Regressionen sowie einzelne Entscheidungsbäume, solange sie nicht auf zu hochdimensionalen Daten trainiert wurden. Einig ist man sich auch, dass nach aktuellem Forschungsstand neuronale Netze keine Erklärungen zu getroffenen Entscheidungen liefern können. Es ist aber ebenfalls Gegenstand aktueller Forschung, trainierte neuronale Netze im Nachhinein erklärbar zu machen.

---

[69] Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf, 14.12.2018, Seite 65f.

[70] https://github.com/etalab/calculette-impots-m-source-cod , 01.12.2018

[71] Direction Générale des Finances Publiques, Etalab, Hackathon “#CodeImpôt”, https://forum.openfisca.fr/t/code-source-de-la-calculette-impots-et-outils-connexes/37 , 29.09.17

Praxisanwendung

Ein System zur automatisierten Auswahl, welche Wahlwerbespots an wen ausgeliefert werden, muss eine Gleichbehandlung sicherzustellen und diese gegenüber einer staatlichen Institution belegen, also nachweisen. Diese überwachende Instanz müsste nicht nur Zugang zu den verwendeten Trainingsdaten haben (siehe Abschnitt 4.1), sondern die Möglichkeit eingeräumt bekommen, Einblick in das tatsächliche Entscheidungssystem zu erhalten (siehe Abschnitt 4.2) um eine rechtliche Überprüfung durchzuführen.

Es ist an dieser Stelle sogar darüber nachzudenken, den Forderungen der Gesellschaft für Informatik (GI,2018, Seite 173) ab dieser Regulierungsklasse zuzustimmen und eine Meldepflicht bei einer staatlichen Institution zu fordern, da es sich hier um ADM-Systeme handelt, welche weitreichendes gesellschaftliches Schadenspotenzial hat.

KLASSE 4: KEINE ADM-SYSTEME

Aus unserer Sicht gibt es sozioinformatische Gesamtsysteme mit so einem hohem Risiko, dass hier tatsächlich lernende ADM-Systeme gar nicht verwendet werden sollten. Dies ist ebenfalls Teil der aktuellen Diskussion, die nicht nur wissenschaftlich, sondern insbesondere gesamtgesellschaftlich geführt werden muss. Wir nehmen an, dass für die Frage danach, was für ADM-Systeme eine Gesellschaft für wünschenswert hält, auch neue, demokratietheoretische Prozesse entwickelt werden müssen: Hierfür ist zum Beispiel zu eruieren, wer diese Entscheidung trifft, welche Ausbildung diese Personen haben oder erhalten müssen und welcher Informationen sie für die Entscheidung bedürfen.

VI. HANDLUNGSEMPFEHLUNGEN

Die oben vorgestellte Risikomatrix hat sich aus unseren Forschungen in den letzten drei Jahren als ein erster, belastbarer Versuch ergeben, gleichzeitig die Risiken von ADM-Systemen zu kontrollieren und dabei neue, innovative digitale Unterstützungen nicht unnötig durch eine Überregulierung zu hemmen. Die dringendste Handlungsempfehlung besteht somit darin, die Nützlichkeit der Risikomatrix zu testen und in einem zweiten Schritt die zur Umsetzung der darin jeweils geforderten Transparenzpflichten und Kontrollmaßnahmen nötigen rechtlichen Grundlagen zu schaffen.

Wir halten es für sinnvoll, in einer Reihe von Workshops mit Beteiligten aus Verbraucherschutz, Social Media Firmen, Startups, Politikern und Rechtswissenschaftlern weitere Beispiele für ADM-Systeme zusammenzutragen, diese gemeinsam in der Risikomatrix zu verorten und für die Beispiele die entsprechenden, detaillierteren Maßnahmen zu bestimmen. Dazu gehört zum Beispiel die jeweilige Benennung des Zielpublikums, gegenüber dem Transparenz hergestellt werden muss und die Fragen nach der Gruppe, die Zugang zu eventuellen Interfaces bekommen sollte. In einem letzten Schritt sollte dann überprüft werden, ob die Anzahl der Risikoklassen sich auch in diesen Workshops als stabil erweist. Nicht zuletzt sollte in solchen Workshops auch eruiert werden, wo in der Risikomatrix entsprechende "analoge" Entscheidungsmechanismen zu verorten sind, welchen Regulierungen diese unterliegen, um damit auch in dieser Richtung einen Abgleich zu schaffen.

Aus unserer Forschung, die in dieser Studie zusammengefasst und um das Konzept der Risikomatrix und die Einteilung aller ADM-Systeme in fünf Risikoklassen ergänzt wurde, ergibt sich in jedem Fall eindeutig, dass ein Algorithmen-TÜV nicht zielführend ist. Im Gegensatz zu anderen technischen Artefakten können ADM-Systeme nicht ohne den sozioinformatischen Kontext, in dem sie eingebettet sind, evaluiert und freigegeben werden. Eine Agentur für ADM-Verfahren, wie von der Gesellschaft für Informatik[72] oder Digitalagentur, wie vom Sachverständigenrat für Verbraucherfragen [73] vorgeschlagen, die vor allen Dingen die Erfüllung der Transparenzpflichten übernimmt und auch einen Teil der notwendigen Kontrollen durchführt, ist dagegen auch aus unserer Sicht zu unterstützen. Daneben gibt es aber auch viele schon etablierte Institutionen, die in die Kontrolle mit eingebunden werden könnten, z.B: Verbraucherschutzorganisationen und spezifische NGOs in diesem Bereich.

Die Handlungsempfehlung an betroffene ADM-System-Anbieter ist eindeutig, möglichst viele der gelisteten Transparenzpflichten freiwillig zu erfüllen. Das erhöht das Vertrauen in die Nutzung der Dienste und ermöglicht eine genauere Schadenspotenzialbewertung. Solange wenig Informationen vorherrschen, muss bei dieser Bewertung immer der schlimmstmögliche Fall betrachtet werden, was zu einer erhöhten Risikoklasse führen kann. Das Beispiel Suchmaschine hat gezeigt, dass nur die Untersuchbarkeit des Personalisierungsgrades dafür sorgt, dass unsere momentane Risikoklasseneinschätzung

[72] Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf, 14.12.2018, Seite 174

[73] Sachverständigenrat für Verbraucherfragen: Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen, 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf, 01.12.2018, Seite 146

für Googles Suchmaschine in der Klasse 1 liegt. Da z. B. eine solche Untersuchung für den Personalisierungsgrad des NewsFeeds auf Facebook aus technischen und datenschutzrechtlichen Gründen nahezu unmöglich ist, liegt aus unserer Sicht der mögliche Schaden durch eine einseitige Auswahl von Nachrichten im Newsfeed deutlich höher, was eine Einstufung in Risikoklasse 2 erfordert.

Nicht zuletzt wollen wir darauf hinweisen, dass alle getroffenen Aussagen nicht nur für ADM-Systeme im engen Verbraucherschutzkontext gelten. Es ist im digitalen Bereich erstens sowieso sehr schwierig, eine klare Grenze zwischen "Nutzern" und "Verbrauchern" zu ziehen. Aber auch dort, wo Menschen in einem Kontext algorithmisch bewertet werden, wo wir die Beurteilten klar anders benennen, z. B. am Arbeitsplatz, im Sozial- oder Finanzsystem, in den Schulen und Universitäten, gilt das Schema der Risikobewertung genauso. Es ist daher dringend erforderlich, dass die Regulierungen der technischen Grundlagen von ADM-Systemen nicht auf der einen Seite im Verbraucherschutzrecht, im Arbeitsschutzrecht, in den Hochschulverordnungen und im Finanz- oder Steuerrecht parallel etabliert wird. Stattdessen bedarf es eines generellen rechtlichen Ansatzes, der für die jeweiligen Anwendungsgebiete detailliert wird. Auch dieser Aspekt bedarf der weiteren, interdisziplinären Zusammenarbeit.

In diesen anderen Bereichen – am Arbeitsplatz, in der Ausbildung, im Steuer- und Justizsystem – wird auch klar, dass eine Regulierung von ADM-Systemen nicht nur auf technischer Ebene notwendig ist. Es ergeben sich beispielweise viele weitere Fragen im Bereich Datenschutz und Arbeitsschutz, wenn es um den Einsatz von KI in diesen Bereichen geht. Weiterhin müssen proaktiv in der Bildungspolitik Maßnahmen ergriffen werden, um den Personen, deren Arbeitsplatz sich durch ADM-Systeme verändert dabei zu helfen, sich diesen Änderungen anzupassen oder sie für andere Tätigkeiten weiterzubilden. Es werden weitgehende sozialpolitische Maßnahmen nötig sein für diejenigen, die andere Tätigkeiten nicht ausüben können.

Trotzdem ist der Fokus auf die verbraucherseitigen Aspekte von ADM-Systemen besonders zu begrüßen, da er es erlaubt, die Bevölkerung auf die große Tragweite des Einsatzes von ADM-Systemen mit Hilfe von Beispielen aus dem Alltag aufmerksam zu machen. Wir empfehlen daher dringend, diese Beispiele didaktisch aufzubereiten und einer breiteren Öffentlichkeit vorzustellen. Insbesondere im Jahr der Wissenschaft mit dem Thema "Künstliche Intelligenz" wäre es wichtig, die Verbraucherinnen und Verbraucher hier in die anstehenden Diskussionen und Entscheidungen einzuführen.

- Ein ADM-System muss also zumindest diese Minimalforderungen erfüllen, kann jedoch durch zusätzliche Transparenz in Bereichen, die eigentlich erst in höheren Regulierungsklassen zum Tragen kommen, Vertrauen schaffen.
- Durch verschiedene Regulierungsklassen, mit „minimaler“ Regulierung minimieren wir den gesamtgesellschaftlichen Aufwand einer ADM-Regulierung

Die vorliegende Studie fokussiert sich zwar auf solche ADM-Systeme, die speziell Verbraucherinnen und Verbraucher betreffen, dennoch lassen sich die hier gezogenen regulatorischen Schlussfolgerungen auf alle ADM-Systeme beziehen, die zur Bewertung oder Klassifizierung von Menschen genutzt werden.

Dies gilt insbesondere bei der Diskussion über die noch ausstehende Klärung diverser rechtlicher Fragen. Es würde die Durchsetzungskraft und Verständlichkeit der Regulierung von ADM-Systemen vermindern, wenn Auswirkungen auf Verbraucher im Verbraucherrecht behandelt würden, Folgen für Kriminelle im Strafrecht und allgemeine Konsequenzen für Bürger im Zivilrecht behandelt würden. Die Richtlinien für den Einsatz von Algorithmen des maschinellen Lernens sollten in Gesetzen aus einem Guss und an einer Stelle zementiert werden, damit in Deutschland und Europa die Chancen von künstlicher Intelligenz genutzt und die Risiken kontrolliert werden können.

VII. ABBILDUNGSVERZEICHNIS