

erschienen in: Maria Reiffenstein und Beate Blaschek (Hrsg.), Konsumentenpolitisches Jahrbuch: Trends, Rechtsentwicklung und Judikatur der letzten zwei Jahre, Bd. 49, Wien 2019, S. 77-111

Verbraucher-Scoring – fair und korrekt?*

von Gerd Gigerenzer, Gert G. Wagner, Johannes Gerberding, Christian Groß, Ariane Keitel, Felix G. Rebitschek und Sarah Sommer

* Der vorliegende Beitrag beruht auf dem Gutachten 'Verbrauchergerechtes Scoring' des Sachverständigenrats für Verbraucherfragen beim deutschen Bundesministerium der Justiz und für Verbraucherschutz (Berlin, 2018; veröffentlicht auch unter www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf). Die Wiedergabe von Ausführungen aus diesem Gutachten im vorliegenden Beitrag ist nicht eigens nachgewiesen.

I. VORWORT

Im Oktober 2018 hat der Sachverständigenrat für Verbraucherfragen (SVRV; www.svr-verbraucherfragen.de) das Gutachten „Verbrauchergerechtes Scoring“ vorgelegt. Auf dieses Gutachten aufbauend stellen dessen federführende Autoren, Gerd Gigerenzer und Gert G. Wagner, gemeinsam mit den Mitarbeiterinnen und Mitarbeitern der Geschäftsstelle des SVRV Johannes Gerberding, Christian Groß, Ariane Keitel und Sarah Sommer, sowie mit Felix G. Rebitschek, ausgewählte Aspekte des Themas Konsumentenscoring vor.¹

In diesem Beitrag werden nach einer kurzen Einführung in die Thematik sowie einem historischen Rückblick, die Wichtigkeit der Transparenz und Verständlichkeit von Scoring für Konsumentinnen und Konsumenten, die Güte von Scoring-Verfahren, die gesellschaftlichen Implikationen sowie die Verknüpfung von Scores aus verschiedenen Lebensbereichen im Fokus stehen. Am Ende eines jeden Unterkapitels finden sich Empfehlungen für konsumentenpolitische Maßnahmen, die darauf abzielen, ein verbrauchergerechtes Scoring zu gewährleisten und damit eine Stärkung der Konsumentinnen und Konsumenten zu erreichen.

II. EINLEITUNG

Im vorliegenden Beitrag geht es um Scoring, basierend auf algorithmischen Verfahren, mit direkter Relevanz für Verbraucherinnen und Verbraucher. Es werden Empfehlungen gegeben, wie Scoring fair, transparent und möglichst diskriminierungsfrei gestaltet werden könnte. Zugleich wird aufgezeigt, welche Herausforderungen, Gefahren und Potenziale mit Scoring im digitalen Zeitalter verbunden sind. Dabei werden insbesondere das etablierte Bonitäts-Scoring sowie neuere Entwicklungen wie das Scoring des Fahrverhaltens bei speziellen Tarifen in der Kfz-Versicherung („pay as you drive“, Telematik-Optionen) und Scoring im Zuge von Bonusprogrammen von deutschen Krankenversicherungen in den Blick genommen. Dies sind Beispiele für Scoring in Lebensbereichen, die für viele Verbraucherinnen und

¹ Für eine knappere Zusammenfassung vgl. *Gigerenzer ua*, Eine vermessene Gesellschaft.

Verbraucher zentral sind. Dem Bonitäts-Scoring etwa kann sich keine Bürgerin und kein Bürger mehr entziehen.

Die Überlegungen und Empfehlungen dieses Beitrags beruhen neben Erkenntnissen aus der aktuellen Fachliteratur und aus Hintergrundgesprächen mit zahlreichen Expertinnen und Experten und Unternehmen auch auf eigenen empirischen Untersuchungen.² Im Rahmen einer Marktstudie wurde die Praxis von Auskunfteien sowie Kfz- und Krankenversicherungen in Bezug auf Scoring und verhaltensbasierte Tarife untersucht. An der schriftlichen Befragung beteiligten sich drei Auskunfteien, zehn Kfz-Versicherer mit Telematik-Angeboten und 62 gesetzliche und private Krankenversicherer. Eine repräsentative Bevölkerungsbefragung hat mittels Telefoninterviews mit über 2.000 Personen in Deutschland ermittelt, was sie über Scoring wissen, wie sie dieses bewerten und ob bzw. unter welchen Umständen sie an scoring-basierten Versicherungsmodellen teilnehmen würden.

A. Definition und Implikationen von Scoring

Scoring ist die Zuordnung eines Zahlenwertes (des Scores) zu einem Menschen zum Zweck der Verhaltensprognose oder Verhaltenssteuerung. Die Bestimmung dieses Zahlenwertes erfolgt in der Regel auf der Grundlage einer breiten Datenbasis durch ein algorithmisches Verfahren, das sehr einfach (zB Addition) oder komplizierter (zB logistische Regression) sein kann. Scoring wird aufgrund der Digitalisierung und der zunehmenden Verfügbarkeit von mehr Daten über einzelne Personen in immer mehr Lebensbereichen relevant, in denen Eigenschaften und Aktivitäten von Menschen analysiert und Prognosen über ihr künftiges Verhalten erstellt werden oder sie zu bestimmten Verhaltensweisen motiviert werden sollen. Der Score kann beispielsweise als Grundlage dafür herangezogen werden, ob und zu welchen Konditionen eine Verbraucherin oder ein Verbraucher einen Immobilienkredit erhält, welche Rückerstattung man bei guter Fahrweise von seiner Kfz-Versicherung erhält und ob jemand sich so gesundheitsbewusst verhält, dass er einen Bonus von seiner Krankenversicherung bekommt. Dies sind Beispiele aus den drei wichtigen Lebens- und Konsumbereichen Finanzen, Mobilität und Gesundheit, in denen Scoring heute eingesetzt wird.

Scoring – vor allem das Bonitäts-Scoring – erfüllt in einer Marktwirtschaft für beide Seiten, Kreditgeber/in und Kreditnehmer/in, eine wichtige Transparenz- und Vertrauensfunktion. Neue Score-basierte Versicherungsprodukte können durchaus Vorteile für Verbraucherinnen und Verbraucher haben, wenn diese sie zum Beispiel zu einer gesundheitsbewussteren

² SVRV (Hrsg), Verbrauchergerechtes Scoring (2018).

Lebensweise motivieren. Neben solchen positiven Effekten kann Scoring aber auch negative Effekte haben: Scoring darf die Betroffenen nicht in die Irre führen, indem es etwa nicht haltbare Versprechen hinsichtlich der Gesundheit macht und Menschen auf Grundlage (reiner) Gruppenzugehörigkeit bewertet (weil sie zB in einer bestimmten Gegend leben; Stichwort Geoscore), sondern es muss zuverlässige Vorhersagen treffen (etwa eine gewisse Bonität eines Verbrauchers oder einer Verbraucherin). Besondere Aufmerksamkeit verdient die Zusammenfassung von Scores aus verschiedenen Lebensbereichen zu einem „Superscore,“ wie es etwa beim chinesischen Sozialkreditsystem (*Social Credit System*) der Fall ist. Superscores eignen sich zur Verhaltenssteuerung mit Hilfe von Privilegien und Sanktionen.

Fair ist Scoring nur dann, wenn eine Reihe von Bedingungen erfüllt ist. So muss beispielsweise das Risiko einer Verwechslung der Person, auf die sich ein Score bezieht, minimiert werden. Ein Widerspruch gegen fehlerhafte Scores, der Bewusstheit und Kenntnis von Fehlern voraussetzt, muss auf einfache Weise möglich sein. Es darf weder direkt noch indirekt auf Basis von geschützten Merkmalen (wie etwa dem Geschlecht) ungerechtfertigt diskriminiert werden. Scoring, das eine Vorhersage treffen will, muss belegen können, wie hoch die Qualität der verwendeten Merkmale für den Zweck der Prognoseerstellung ist und wie genau und realitätsgetreu die Vorhersagen selbst sind. Diese prädiktive Kraft sollte zudem über verschiedene sozio-ökonomische Gruppen hinweg gleich sein. In den folgenden Kapiteln werden nach einem historischen Rückblick die zentralen Herausforderungen von auf Scoring basierenden (neuen) Geschäftsmodellen beleuchtet und es wird umrissen, wie verbrauchergerechtes Scoring gestaltet werden kann, welche Anforderungen dafür erfüllt sein müssen und wie dieses politisch und institutionell unterstützt werden kann.

B. Historischer Rückblick

Die Einschätzung von Personen mit Hilfe einer Zahl (Scoring) hat in unserer Kultur Tradition, man denke nur an Schul- und Examensnoten sowie an den „Schwerbehinderten-Grad“, der wie Bonitäts-Scoring auf einem Wertebereich von null bis hundert definiert ist, oder an die sogenannte „Gliedertaxe“, die zum Ausdruck bringt, welche Bedeutung dem Verlust oder der Funktionsunfähigkeit von Körperteilen und Sinnesorganen beigemessen wird.

In deutschen Schulen werden seit dem 16. Jahrhundert Zensuren vergeben.³ An individuelle Bewertungen von Leistungen und Lernerfolgen sind auch heute noch Konsequenzen, wie die Zulassung zu höheren Klassenstufen oder weiterführenden Schularten sowie das Erreichen von Schulabschlüssen, geknüpft. Als Zulassungskriterium zu vielen europäischen Hochschulen ist neben anderen Kriterien die Abiturnote ein zentraler Wert. In diesem Kontext wird besonders deutlich, dass in den Abschlussnoten nicht nur eine Bewertung vergangener Leistungen gesehen, sondern ihnen auch eine Vorhersagekraft in Bezug auf zukünftige Leistung zugeschrieben wird. Eine gute Abiturnote soll signalisieren, dass auch während des Studiums gute Leistungen zu erwarten sind und ein Abschluss wahrscheinlich ist.

Vor allem im wirtschaftlichen Bereich kommt Scoring eine historische Bedeutung zu, denn hier spielte Risikominimierung eine große Rolle. Für Geschäftsbeziehungen, bei denen Verträge mit bis dahin unbekanntem Partnern eingegangen werden, muss stets ein gewisser Vertrauensvorschub geleistet werden. Aus diesem Bedarf, sich über die Zuverlässigkeit und Zahlungsfähigkeit von Geschäftskundinnen und Geschäftskunden zu informieren, entstanden im 19. Jahrhundert die ersten Auskunfteien. Zu den ersten in Europa gehörten die Auskunfteien *Wys Muller* (gegründet 1861), *Schimmelpfeng* (gegründet 1872) oder *Creditreform* (gegründet 1879). Diese Unternehmen sammelten wirtschaftlich relevante Informationen zu Personen und Firmen und verkauften diese an Unternehmen und Banken. Seither bildeten Auskunfteien eine wesentliche Grundlage für ein funktionierendes Kreditwesen.

Die ersten Versuche, das Risiko eines Zahlungsausfalls von Personen quantitativ zu berechnen und numerisch darzustellen, wurden in den 1940er Jahren unternommen. Bis dahin bestanden rudimentäre Scoring-Systeme (etwa von Versandhäusern) in einem Kriterienkatalog, mit dessen Hilfe vor der Kreditvergabe eine Reihe von Voraussetzungen auf Erfüllung geprüft und zusammengezählt wurden.⁴ Der Mathematiker *Davis Durand* verwendete 1941 in einem Forschungsprojekt zum ersten Mal die Diskriminanzanalyse, um das Ausfallrisiko von Krediten zu bestimmen.⁵ Er analysierte Datensätze über bereits vergebene Kredite daraufhin, welches die entscheidenden Faktoren für eine problemlose Rückzahlung oder für Rückzahlungsschwierigkeiten gewesen waren, und entwickelte einen *Kreditscore*. Das erste Unternehmen, das statistische Modelle für die Kreditvergabe kommerziell entwickelte, war *Fair, Isaac and Company* in Kalifornien (heute bekannt als

³ *Lintorf* in *Lintorf* (Hrsg), Grundschulnoten.

⁴ *Thomas ua*, Credit Scoring.

⁵ *Durand*, Risk Elements.

FICO). Seit den 1950er Jahren vertrieb es Scoring-Produkte an Finanzhäuser, Einzelhändler und Versandhäuser⁶.

In den folgenden Jahrzehnten wurden mathematische Fortschritte von Innovationen im Bereich der elektronischen Datenverarbeitung begleitet, was schließlich weitgehend automatisiertes Bonitäts-Scoring ermöglichte. Die Kombination aus Computern und Algorithmen sowie die Erfahrungen von Unternehmen, dass mit Hilfe von Scoring die Ausfallraten ihrer Kredite sowie Betrugsdelikte deutlich reduziert werden konnten, führten zu den Scoring-Produkten der Auskunfteien wie wir sie heute kennen. Dabei ist zu beachten, dass die Dienstleistung, die Scoring-Firmen wie die Schufa anbieten, nicht nur aus der Berechnung eines individuellen Scores besteht, sondern auch in dessen elektronischen Bereitstellung innerhalb von Sekunden.

Ein weiterer Bereich, in dem solche Formen der Risikoeinschätzung schon lange etabliert sind, ist die Welt der Versicherungen. Seit dem 17. Jahrhundert wurden Algorithmen entwickelt, um individuelle Beitragssummen für Lebensversicherungen zu berechnen.⁷ In den 1920er und 1930er Jahren entstand in Deutschland das Interesse, die Beitragsberechnung in der Krankheitskostenversicherung auf mathematisch-statistische Grundlagen zu stellen. Mittels sogenannter Morbiditätstabellen wurde festgestellt, dass je nach Geschlecht, Alter und Beruf der Versicherten unterschiedliche Krankheitskosten zu erwarten waren⁸. Noch heute werden in der privaten Krankenversicherung die Beiträge beim Einstieg individuell berechnet. Auch in der Lebensrisiko- oder Berufsunfähigkeitsversicherung ist dies der Fall. Verbraucherinnen und Verbraucher werden anhand einer Kombination einzelner Merkmale (zB Alter oder Vorerkrankungen) eingestuft, das Risiko für den Versicherer damit abgewogen und die Beitragsätze entsprechend kalkuliert.

Eine besonders komplexe Prämienkalkulation findet in der Regel in der Kfz-Versicherung statt, in der die Tarife aufgrund einer Vielzahl an Merkmalen an die einzelnen Kundinnen und Kunden angepasst werden. Wesentliche Faktoren sind ua die Typklasse des Automodells, die Regionalklasse und die Schadensfreiheitsklasse sowie Eigenschaften wie Zahl der Nutzerinnen und Nutzer, Nutzer- und Fahrzeualter, Kilometerfahrleistung und Stellplatz (GDV, 2016).

⁶ *Dixon/Gellman*, The scoring of America.

⁷ *Daston*, Probabilistic Revolution.

⁸ *Wagner-Braun*, Bedeutung berufsständischer Krankenkassen.

Partner-Vermittlungsagenturen arbeiten ebenfalls – völlig intransparent – mit Algorithmen, die die „Passfähigkeit“ von Menschen berechnen. Da die Scoring-Methoden der Vermittlungsagenturen jedoch Geschäftsgeheimnisse darstellen, sind viele Menschen skeptisch. Dies gilt offenkundig weniger für moderne digitale Partner-Vermittlungsplattformen.⁹ Deren Erfolge sind zwar auch überschaubar¹⁰, aber ihre Attraktivität liegt eher im schnellen Zugang zu potenziellen Partnerinnen und Partnern als in der Fähigkeit der Algorithmen, den richtigen Partner bzw die richtige Partnerin zu finden.

Die hier genannten Beispiele zeigen, dass Scoring durchaus kein neues Phänomen ist. Die Digitalisierung vervielfacht jedoch die Möglichkeiten, Personen zu bewerten, zugleich aber auch die Risiken und Chancen. Dies führt zu einer neuen Dynamik in Wirtschaft und Gesellschaft, die mehr denn je zuvor gezielte politische Maßnahmen verlangt.

III. AUSGEWÄHLTE PROBLEMBEREICHE

In den folgenden Abschnitten werden vier Bereiche in den Blick genommen, in denen das Phänomen Scoring Probleme aufwirft. Neben der Transparenz und Verständlichkeit des Scorings (Abschnitt IV.A.) und der Güte der verwendeten Daten und der algorithmisch-statistischen Verfahren (Abschnitt IV.B.) handelt es sich hierbei um die potenziellen gesellschaftlichen Implikationen von Scoring (Abschnitt IV.C.), insbesondere um die Gefahr der Zusammenlegung von Scores aus verschiedenen Lebensbereichen zu einem Superscore (Abschnitt IV.D.).

A. Transparenz und Verständlichkeit

Durch die Digitalisierung entstanden und entstehen Möglichkeiten, in neuen Lebensbereichen Scoring zu betreiben oder bereits etablierte Scoring-Verfahren datenintensiver, dh differenzierter, auszugestalten. Hier zu nennen sind beispielsweise automatisierte Scoring-Verfahren für die Bewertung von Personal hinsichtlich der Beförderung in Unternehmen bzw

⁹ <https://www.singleboersen-vergleich.de>.

¹⁰ Gigerenzer, Liebestrunken.

für die Personalauswahl (*People Analytics*)¹¹ und Scoring-Verfahren im Polizeibereich, mit denen Einsatz- und Ressourcenplanung vorgenommen werden (*Predictive Policing*).¹² Um die dem Scoring immanenten Chancen, etwa in den Bereichen Personalauswahl, Personalbewertung und Gesundheit künftig voll nutzen zu können, ist es notwendig, dass dafür vom Gesetzgeber Rahmenbedingungen für ein gerechtes Scoring geschaffen und durchgesetzt werden. So könnte digitales Scoring sogar transparenter sein als es menschliche Entscheidungen sind. Einem Scoring-Algorithmus kann man mit wenig technischem Aufwand und ohne dass man den Algorithmus tatsächlich kennen müsste durch systematisches Testen leicht nachweisen, dass er beispielsweise bei der Vergabe von Mietwohnungen bestimmte Gruppen nachteilig bewertet und dadurch ggf ungerechtfertigt diskriminiert.¹³ Der Vermieterin oder dem Vermieter eines Dreifamilienhauses, die bzw der diskriminiert, ohne darüber zu reden, kann man dies nur schwer nachweisen – auch nicht statistisch, da sie bzw er zu selten neu vermietet.

Transparenz bedeutet die Offenlegung von Informationen gegenüber Verbraucherinnen und Verbrauchern durch die Ersteller oder Verwender von Scoring-Verfahren. Diese Forderung geht mit der Herausforderung einher, die transparent gemachten Informationen so aufzubereiten und zu strukturieren, dass deren Verständlichkeit für Verbraucherinnen und Verbraucher gewährleistet wird. Scoring-Verfahren transparent zu gestalten ist nicht zuletzt deshalb wichtig, weil erst ein ausreichendes Transparenzniveau es Verbraucherinnen und Verbrauchern ermöglicht, ihre Rechte, zum Beispiel auf Korrektur eines fehlerhaften Scores, geltend zu machen. Zur Transparenz gehört auch eine rechtlich abgesicherte Möglichkeit gegen Scores und ihre Verwendung zu klagen.

Transparenz ist für die Akzeptanz von Scoring zentral, wie man leicht an traditionellen Verfahren, etwa dem Numerus Clausus für den Hochschulzugang oder den „Punkten in Flensburg“ für die Fahrerlaubnis erkennen kann.

Scoring-Verfahren werden sowohl zur Prognose als auch zur Steuerung von Verhalten eingesetzt. Im Hinblick auf Transparenz muss zwischen diesen beiden Verwendungsabsichten unterschieden werden: Scoring-Verfahren, die auf eine Verhaltensprognose abzielen, wollen üblicherweise das in den Blick genommene Verhalten nicht selbst beeinflussen.

Betrugsscoring versucht etwa kriminelle Geschäfte vorherzusagen und so zu verhindern, ohne

¹¹ Sharp in Kiggins (Hrsg), *The Political Economy of Robots*, 61-80.

¹² Gluba, *Unabhängige Zeitschrift für die kriminalistische Wissenschaft und Praxis* 2014, 347–352.

¹³ vgl Gigerenzer *ua*, *Licht in die Black Box; Zehlike & Wagner*, *Neutrale Algorithmen?*

dass gleichzeitig Kriminelle eine Rückmeldung erhalten und die Chance sich anzupassen. Auch nicht-kriminelle Personen, deren Verhalten Gegenstand einer prognostischen Einschätzung durch ein Scoring-Verfahren sind, werden normalerweise ein Interesse daran haben, zu erfahren, ob ein Scoring stattfindet und zu welchem Zweck. Zweitens werden sie um die Konsequenzen des Scores wissen wollen. Weiterhin können sie daran interessiert sein, zu wissen, auf welcher Datengrundlage die Ermittlung des sie betreffenden Score-Wertes erfolgt, das heißt, welche ihrer Merkmale in die Score-Wert-Ermittlung eingehen. Schließlich können sie daran interessiert sein, in die innere Funktionsweise des Scoring-Algorithmus Einblick zu bekommen, also insbesondere zu erfahren, welches relative Gewicht jedes erfasste persönliche Merkmal für die Ermittlung des Score-Wertes hat.

Diesen Interessen können auf der anderen Seite Geheimhaltungsinteressen des Unternehmens, das ein Scoring durchführt, oder der Allgemeinheit gegenüberstehen. Ein Geheimhaltungsinteresse liegt regelmäßig dann vor, wenn die Prognoseleistung des Scoring-Verfahrens dieses wirtschaftlich wertvoll und damit als Unternehmensgeheimnis schützenswert macht. Wenn bekannt wird, wie ein algorithmisches Verfahren zur Score-Berechnung im Einzelnen ausgestaltet ist, verliert es seinen Charakter als Geschäftsgeheimnis und kann von Konkurrenten übernommen werden.

Scoring kann auch ein Instrument der Verhaltenssteuerung sein (etwa des Fahrverhaltens). Mit diesem Zweck unterlegt, ist Transparenz eine notwendige Bedingung des wirkungsvollen Einsatzes des Scoring-Verfahrens. Denn ein Anreizsystem kann eine zielgerichtete verhaltenssteuernde Wirkung nur dann entfalten, wenn es den Zusammenhang zwischen dem Verhalten und seiner Bewertung offenlegt. Auf das Scoring bezogen: Wenn es dem Verwender des Scoring-Verfahrens darum geht, zur Verbesserung des Score-Wertes zu motivieren, ist zwingend erforderlich, dass er zumindest offenlegt, dass bestimmte Verhaltensweisen „punkteträchtig“ sind und andere wiederum nicht.

Allerdings kann die Verhaltenssteuerung durch Scoring-Verfahren auch „weicher“ erfolgen. Dies lässt sich am Beispiel der Kreditwürdigkeit verdeutlichen. Nicht alle Kriterien für den Score-Wert sind bekannt. Trotzdem kann ein solches Scoring-Verfahren Auswirkungen auf das Verhalten der gescorten Person haben. Sie kann versuchen, ihren Score zu verbessern. Nur ist die Richtung dieser Verhaltensänderung unbestimmter, denn die Verbraucherinnen und Verbraucher können nur vermuten, welche Verhaltensweisen bewertet werden.

Im Zusammenhang mit Scoring herrscht an vielen Stellen Verunsicherung, Skepsis und Unwissen. Vielen Verbraucherinnen und Verbrauchern ist beispielsweise nicht immer klar, dass sie überhaupt gescort werden oder was genau das Ziel des Scorings ist bzw welche persönlichen Merkmale dafür herangezogen werden. In einer im Auftrag des SVRV durchgeführten repräsentativen Bevölkerungsbefragung wurde ein grundlegendes Problem deutlich: Das Wechselspiel zwischen dem mangelnden Wissen in der Bevölkerung einerseits und der verbesserungswürdigen Transparenz der Auskunftsteien andererseits. So weiß ein Großteil der Befragten nicht, welche Merkmale zur Beurteilung der Bonität von Auskunftsteien wie der Schufa verwendet werden. Etwa denken zwei von drei Deutschen fälschlicherweise, dass ihr Beruf und ihr Vermögen relevant für die Berechnung ihres Bonitäts-Scores sind.¹⁴ Die Ursache für dieses Wissensdefizit ist jedoch nicht ausschließlich bei den Verbraucherinnen und Verbrauchern zu suchen, sondern spiegelt auch die mangelnde Transparenz der Auskunftsteien in Bezug auf das Bonitäts-Scoring wider. Auskunftsteien sind nicht verpflichtet, den Bürgerinnen und Bürgern darzulegen, wie ihr Score zustande gekommen ist.¹⁵ Indessen wäre es vorzuziehen, das Recht der Verbraucherinnen und Verbraucher auf Aufklärung über das Recht der Anbieter auf Geheimhaltung zu stellen. Dafür gibt es mehrere Gründe: Als Argument gegen die Offenlegung wird regelmäßig die Wahrung des Geschäftsgeheimnisses herangezogen. Dem steht entgegen, dass die Anbieter zwar alle Merkmale und deren Gewichte offenlegen sollten, nicht jedoch dazu verpflichtet werden sollten, den Quellcode oder andere Einzelheiten, die für das Verständnis der Verbraucherinnen und Verbraucher nicht wesentlich sind. Die Gefahr, dass Konkurrenten dann das Verfahren der Score-Berechnung übernehmen könnten, ist daher auch als extrem gering einzustufen. Zumal zum Wettbewerbsvorteil der Auskunftsteien vor allem auch die elektronische Bereitstellung in Quasi-Echtzeit gehört, deren Techniken und ggf Geheimnisse nicht durch die Offenlegung des Scoring-Verfahrens bekannt werden.

Als positive Beispiele in Bezug auf Score-Transparenz sind Versicherungen zu nennen, welche Telematik-Tarife im Zusammenhang einer Kfz-Haftpflichtversicherung oder Bonusprogramme im Rahmen einer Krankenversicherung anbieten, die die Merkmale und Gewichte für ihre Berechnungen auch der Verbraucherin und dem Verbraucher gegenüber offenlegen. Die Marktstudie des SVRV zeigte, dass die Kundinnen und Kunden von Telematik-Tarifen in der Kfz-Versicherung über die Kriterien, die in die Score-Bildung

¹⁴ infas (Hrsg), Verbraucher-Scoring.

¹⁵ Rothmann ua, Credit Scoring in Österreich; für Deutschland siehe auch das sog „Schufa-Urteil“ des deutschen Bundesgerichtshofs, BGH, Urteil vom 28.01.2014 zum Aktenzeichen VI ZR 156/13.

einfließen, in der Regel gut informiert werden. Üblicherweise bilden drei bis neun Merkmale wie Brems-, Beschleunigungs- und Geschwindigkeitsverhalten die Grundlage für die Scoring-Modelle.¹⁶

Ein häufig herangezogenes weiteres Argument gegen die Transparenz-Forderung ist, dass Verbraucherinnen und Verbraucher bei Offenlegung der Zusammensetzung eines Scores ihr Verhalten ändern könnten, um ihren Score zu verbessern (sogenanntes *gaming*). Hier muss man allerdings differenzieren: Nur wenn die Merkmale, die in das Scoring einfließen, mit dem durch das Scoring ermittelten Umstand in keinem ursächlichen Zusammenhang stehen, ist *gaming* ein Problem. Wenn beispielsweise der bloße Beitritt in ein Fitnessstudio den Gesundheitsscore erhöht und Krankenversicherungsprämien so indirekt reduziert werden, dann reicht eben dieser Beitritt aus, ohne das Fitnessstudio tatsächlich zu nutzen. Die Merkmale in einem guten Scoring-Algorithmus sollten daher ursächlich (wie die aktive Teilnahme am Sport) statt nicht-ursächlich (wie der bloße Eintritt in ein Fitnessstudio) sein. Dann ist das Scoring auch grundsätzlich *fair* und *gaming* irrelevant.

Um bei Verbraucherinnen und Verbrauchern eine *grundsätzliche* Akzeptanz für Scoring zu erreichen, halten wir es für unerlässlich, Scoring transparent und verständlich zu gestalten. Das heißt, es muss sichtbar sein, dass Scoring stattfindet, was das Ziel des Scorings ist und welche Merkmale in welcher Form dafür herangezogen werden. Dabei ist Transparenz kein Selbstzweck, sondern eine notwendige Voraussetzung dafür, dass Fälle von Nicht-Fairness (Diskriminierung) und Fehler besser erkannt werden können. Dies wird deutlich, wenn man sich folgenden Sachverhalt anschaut: Die Tarifierung von Lebensversicherungen und privaten Krankenversicherungen basiert seit jeher auf Scores hinsichtlich gruppenspezifischer Lebenserwartungen und Krankheitskosten. Transparenz in diesem Bereich ermöglicht es, bestimmte Formen des Scorings zu problematisieren und gegebenenfalls auf sie zu verzichten, wenn sie als unfair bewertet werden; was sich im Zeitverlauf durchaus ändern kann. Geschehen ist dies etwa bei geschlechtsspezifischen Tarifen bei privaten Kranken- und Lebensversicherungen. Der Gerichtshof der Europäischen Union¹⁷ hat 2011 den diskriminierenden Gehalt einer geschlechterdifferenzierenden Tarifgestaltung dargelegt.

¹⁶ Aus Verbrauchersicht kritisch einzuschätzen ist allerdings die Einbeziehung von durch Autofahrer/innen nur schwer beeinflussbaren Parametern, wie Tageszeit oder Ort einer Fahrt. Differenzen zwischen Werbeversprechen (Telematik führt zu besserem Fahren) und eigentlichem Zielmerkmal des Scorings (Schadenseintrittswahrscheinlichkeit) sollten transparent gemacht werden.

¹⁷ EuGH, Urteil vom 01. 03. 2011 – C-236/0.

Infolgedessen haben inzwischen „Unisex“-Tarife in der privaten Kranken- und Lebensversicherung Einzug gehalten.¹⁸

1. Scoring-Transparenz als Spezialfall von Algorithmentransparenz

Derzeit wird eine intensive Debatte geführt über ein angemessenes Regulierungsregime für „Algorithmen“.¹⁹ Ein Scoring-Algorithmus ist ein bestimmter Typ Algorithmus.²⁰ Die Diskussion über Scoring-Transparenz lässt sich deshalb als ein Teil der allgemeinen Debatte über die Regulierung von Algorithmen führen.

Mit der Offenlegung des Quellcodes selbst einfacher Computerprogramme ist den Verbraucherinnen und Verbrauchern in aller Regel nicht geholfen. Die Mehrheit der betroffenen Personen sind keine IT-Spezialisten. Und selbst wenn sie es wären: die technische Komplexität der Computersysteme, deren Entscheidungsverhalten „transparent“ gemacht werden soll, macht selbst für die Kontrolle durch Expertinnen und Experten eine andere Form von Transparenz erforderlich als die bloße Offenlegung von Programmcodes.²¹

Transparent ist ein System algorithmischer Entscheidungsfindung nicht dann, wenn es in allen Einzelheiten offen vor dem/der Betrachtenden liegt (was bei neuronalen Netzen durchaus auch schwierig wäre)²², sondern wenn es sich seinen Benutzerinnen und Benutzern zu erklären vermag: „Lässt sich nachträglich begründen oder sichtbar machen, wie der Weg zum Ergebnis aussah?“²³ Deshalb ist es nicht notwendig, dass Transparenz „volles Verständnis“ für Scoring herstellt. Es reicht aus, Möglichkeiten zu schaffen, auch unter den Bedingungen unvollständiger Transparenz und unvollständigen Verständnisses Erkenntnisse über die Funktionsweise eines Algorithmus zu erlangen, das sogenannte *black box tinkering*.²⁴

Der Vorschlag einer Transparenz-Schnittstelle²⁵ knüpft hier ebenso an wie die Vorschläge der Gesellschaft für Informatik, das Testen von Algorithmen als Regulierungsinstrument stark zu

¹⁸ Arentz, Unisex-Tarife.

¹⁹ Ermert, Ethische Grenzen.

²⁰ Just & Latzer, Media, Culture & Society 2016, 238-258.

²¹ Samek, Wiegand & Müller, ITU Journal 2017, 1-10; Selbst/Powles, International Data Privacy Law 2017, 233-242; Montavon, Samek & Müller, Digital Signal Processing 2018, 1-15; Wischmeyer, Archiv des öffentlichen Rechts 2018, 1-66; Gigerenzer, Wagner & Müller, Licht in die Blackbox; Gesellschaft für Informatik (Hrsg), Technische und rechtliche Betrachtungen.

²² vgl etwa Ribeiro, Singh & Guestrin, Model-agnostic interpretability; Burrell; Alber, Lapuschkin & Seegerer, iNNvestigate.

²³ Passig, Merkur 2017, 25.

²⁴ Perel & Elkin-Koren, Florida Law Review 2017, 181-221; Wachter, Mittelstadt & Floridi, International Data Privacy Law 2017, 76-99.

²⁵ Gigerenzer, Wagner & Müller, Licht in die Blackbox.

machen²⁶. Hierzu werden die Eingaben systematisch variiert und die Ausgaben ausgewertet. Dies könnte zB eine datenschutzrechtliche Aufsichtsbehörde im Rahmen von sogenannten Datenschutzüberprüfungen („audits“) nach Art. 58 Abs. 1 Buchstabe b DSGVO verlangen und ggf selbst durchführen. Was im Inneren der Black Box im Detail vor sich geht, erkennt man zwar auf diese Weise nicht notwendigerweise, aber man kann hinreichende Erkenntnisse über die relevanten Funktionsweisen des Algorithmus gewinnen.

Zusammenfassend sind wir der Auffassung, dass Verbraucherinnen und Verbraucher das Recht erhalten sollten, zu verstehen, wie ihre Scores zustande kommen und welche Konsequenzen daran geknüpft sind. Das würde zudem zur schnelleren Entdeckung und Korrektur von Fehlern beitragen. Mangelnde Transparenz im Zusammenhang mit Scoring verhindert darüber hinaus, die Qualität der Scores beurteilen zu können (siehe Abschnitt IV.B.). Außerdem erleichtert eine verbesserte Transparenz des Scorings die Kontrolle ihrer Qualität (dazu Abschnitt IV.B.). Die Kfz-Versicherungen mit ihren Telematiktarifen und Krankenkassen mit ihren Bonusprogrammen erfüllen die Forderung nach Transparenz bereits teilweise, während viele Auskunftfeien in Deutschland sich ihr ohne tragfähige Begründung verschließen.

SVRV-FORDERUNG 1: SCORING FÜR VERBRAUCHERINNEN UND VERBRAUCHER TRANSPARENT UND VERSTÄNDLICH MACHEN

- Die in der DSGVO (siehe Art. 15 Abs. 1 Buchst. h) festgeschriebenen Verständlichkeitsanforderungen für Scoring und score-basierte Geschäftsprozesse sollten durch Datenschutzbehörden operationalisiert werden. Maßstab für Verständlichkeit sollte die Durchschnittsverbraucherin bzw. der Durchschnittsverbraucher sein. Sofern eine nicht mehr ohne weiteres nachvollziehbare Komplexität eines Scoring-Verfahrens gegeben sein sollte, sollte Scoring nicht nur für Aufsichtsbehörden, sondern auch zumindest für Verbraucherorganisationen und Nicht-Regierungsorganisationen nachvollziehbar sein.
- Der Gesetzgeber soll sicherstellen, dass Scoring-Anbieter den Verbraucherinnen und Verbrauchern alle Merkmale, auf deren Basis sie gescort werden, sowie deren Gewichte auf verständliche und leicht zugängliche Weise offenlegen.²⁷

²⁶ Gesellschaft für Informatik (Hrsg), Technische und rechtliche Betrachtungen, mit der Empfehlung zur Schaffung eines „Rechts auf Durchführung von Tests“)

²⁷ SVRV (Hrsg), Verbrauchergerechtes Scoring. Die Forderung, dass *alle* Merkmale (und nicht nur die „wesentlichen“) den Verbrauchern gegenüber offengelegt werden, ist die Auffassung der beiden federführenden Autoren des SVRV-Gutachtens, die nicht von allen anderen Mitgliedern geteilt wurde.

B. Qualität von Scores und Daten

Scoring sollte nicht nur für Verbraucherinnen und Verbraucher verständlich sein, sondern auch bestimmte Qualitätsstandards erfüllen. Scoring, das Verhalten vorhersagen will, wie beispielsweise das Bonitätsscoring, sollte präzise und reliabel die Zielvariable vorhersagen, also im Falle des Bonitätsscorings die Wahrscheinlichkeit, mit der jemand seinen Zahlungsverpflichtungen nachkommt. Bei Scoring, dessen Ziel es ist, Verhalten zu steuern, wie es bei den Bonusprogrammen der Krankenkassen der Fall ist, sollten die gewünschten Effekte (zB Förderung der Gesundheit) durch das Verhalten, zu dem die Gestaltung des Scoring-Verfahrens motiviert, auch tatsächlich erreicht werden können.²⁸

Das heißt im Zusammenhang mit der Qualität von Scores stellen sich ua folgende Fragen: Misst der Score tatsächlich das, was er zu messen vorgibt? Und wie präzise? Werden Verbraucherinnen und Verbraucher durch das Scoring ausreichend, genau und fair klassifiziert? Steuert der Score also das Verhalten bzw sagt das vorher, was er vorgibt? Oder ist das Unternehmensziel von Scoring möglicherweise ein ganz anderes (zB Kundenbindung)?

1. Güte der Scoring-Verfahren

Unabhängig davon, mit welchen statistischen Verfahren Score-Anbieter in der Praxis arbeiten, um Scores zu berechnen, sollte eine hinreichende Score-Güte gewährleistet sein. Wie „gut“ ein Algorithmus, der dem Scoring zugrunde liegt, im Einsatz sein kann, hängt dabei von der Wahl des Algorithmus selbst ab, dem Zielmerkmal, der Wahl der Verbrauchermerkmale sowie weiteren Modellparametern. Die Güte eines Verfahrens hängt also davon ab, wie korrekt es Verbraucherinnen und Verbraucher in Bezug auf ihr Risiko-Verhalten klassifiziert, zB hinsichtlich des Kreditausfallrisikos oder der Unfallwahrscheinlichkeit. Einschränkend sei jedoch angemerkt, dass es keine komplett fehlerfreien Methoden und Entscheidungen gibt. Dies gilt für Algorithmen-basierte Entscheidungen wie für menschliche Klassifizierungen, wie sie zB von Kreditsachbearbeiterinnen und Kreditsachbearbeitern von Banken vorgenommen werden.

Im Kontext der Score-Berechnung werden in der Literatur drei Kategorien von Gütemaßen häufig herangezogen: Maße wie die Falsch-Positiv-Rate und *Hit-Rate* oder die *Area Under*

²⁸ Die Marktstudie des SVRV zeigte beispielsweise, dass manche Krankenkassen ihren Versicherten ermöglichen, mit gemeinnützigen Aktivitäten (zB Blutspenden, Erste-Hilfe-Kurse) Bonuspunkte zu sammeln, die keinen individuellen Gesundheitseffekt haben und somit keinen präventiven Charakter haben. Gleichwohl werden solche Programme gesundheitsförderlich dargestellt (SVRV (Hrsg), Verbrauchergerechtes Scoring).

The Curve (AUC), welche die Trennschärfe (Discriminatory Power) eines Modells beschreiben; Maße wie der Brier-Score, welche die Genauigkeit der Vorhersage eines Modells beschreiben; sowie der *Classification Error*²⁹.

Aus verbraucherpolitischer Perspektive besteht das Problem dabei darin, dass weder Gütemaße von Score-Anbietern – sofern sie denn tatsächlich genutzt werden – berichtet werden, noch gesetzliche Mindestvorgaben für die Güte von Scoring-Verfahren vorliegen. Die Minimal-Anforderung des Gesetzgebers in Deutschland (§ 31 Abs. 1 Nr. 2 des deutschen Bundesdatenschutzgesetzes) ist bereits dann erfüllt, wenn eines der eingangs genannten algorithmischen Entscheidungsverfahren überhaupt angewandt wird – eine völlig unzureichende Qualitätsanforderung. Wünschenswert wäre hier beispielsweise die Verwendung von Kriterien analog zu wissenschaftlichen Standards der empirischen Sozialforschung, die für Veröffentlichungen von Ergebnissen im Rahmen eines Peer-Review-Verfahrens empfohlen werden. Dazu gehören die Offenlegung aller verwendeten Merkmale, deren Kennwerte (Durchschnitt, Standardabweichung etc), des Algorithmus (derzeit meist eine lineare Merkmalskombination), sowie der Güte des Algorithmus in Form der oben genannten Gütemaße. Diese Angaben sollten, wie in der wissenschaftlichen Praxis üblich, einer unabhängigen Begutachtung durch mehrere qualifizierte externe Gutachterinnen und Gutachter unterzogen werden.

Ein weiterer Aspekt mit Blick auf die Güte eines Algorithmus ist die mögliche Verschlechterung der Prognosekraft aufgrund struktureller Veränderungen, sogenannter exogener Effekte. So ist anzunehmen, dass mit wachsenden Zeitabständen aufgrund fortwährender gesellschaftlicher Veränderungen die Stabilität der in Scoring-Modellen ermittelten statistischen Zusammenhänge abnimmt³⁰. Die Prognosekraft eines statistisch berechneten Scorewerts kann sich demnach im Zeitverlauf ändern. Auch können wiederkehrende konjunkturelle Schwankungen Einfluss auf die Verlässlichkeit von Bonitäts-Scores ausüben³¹. Im Gesundheitsbereich könnte eine strukturelle Verschlechterung von Gesundheits-Scores durch den demografischen Wandel und den sich im Durchschnitt verschlechternden Gesundheitszustand der Bevölkerung stattfinden. Scoring-Verfahren müssen also in regelmäßigen Abständen gegen die Einflüsse exogener Effekte und Ausweichreaktionen durch die Gescorten abgesichert werden. Zudem können in alten Daten überkommene soziale Normen kodiert sein, die durch Modelle, die auf nicht aktualisierten

²⁹ zB *Lessmann* ua, *European Journal of Operational Research* 2015, 124-136.

³⁰ *Hand & Henley*, *Journal of the Royal Statistical Society* 1997, 523-541.

³¹ *Schröder* ua in *Schröder/Taeger* (Hrsg), *Scoring im Fokus*.

Daten beruhen, (unwillentlich) perpetuiert werden³² und heutzutage als diskriminierend bewertet werden.

2. Güte der Daten

In der einschlägigen Methoden-Literatur ist ausführlich beschrieben, dass oft weniger die Scoring-Methode für die Güte des Scorings eine Rolle spielt als die (fehlende) Qualität der eingehenden Daten und die Stabilität des gescorten Prozesses.³³

Fehler in der Datengrundlage können sich unmittelbar auf die Güte eines Algorithmus auswirken. Im Extremfall ist der Score, der einer Verbraucherin oder einem Verbraucher zugesprochen wird, schlichtweg falsch, weil er mit einer anderen Person verwechselt wurde (was bei häufigen Namen besonders leicht vorkommen kann, insbesondere dann, wenn Vor- und Nachname gleichermaßen betroffen sind, wie etwa bei „Christine Schäfer“). Bezüglich der Qualität dieser „Entity Recognition“, die immer wieder im Mittelpunkt anekdotischer Evidenz steht, über die in der Presse berichtet wird³⁴, ist wissenschaftlich jedoch bisher wenig bekannt.

Vertreter/innen des Datenschutzes streiten dafür, so wenig personenbezogene Daten wie möglich zu erfassen. Dies fordert aber die eindeutige Zuordnung von Mensch und Datensatz heraus. Beispielsweise kommt es immer wieder bei Umzügen an einen neuen Wohnort zu Problemen bei der Entity Recognition: nämlich dann, wenn ein Score-Anbieter ggf zunächst einen neuen Datensatz anlegen muss, der aufgrund einer möglicherweise fehlenden Datenverknüpfung keine Informationen über die Bonitäts-Historie einer Person enthält. Falls dann die Bonität dieser Person zunächst durch Geo-Scoring approximiert werden muss, kann es sein, dass dieser Person dadurch ein ungünstigerer Score zugewiesen wird als wenn sie erkannt worden wäre.

Mögliche Fehlerquellen über Verwechslungen und Nicht-Erkennung hinaus sind fehlende und veraltete Daten.³⁵ Im Bereich der Kfz-Telematik wird (auch in der Marktbefragung des SVRV) berichtet, dass Daten entweder falsch aufgezeichnet wurden (zB weil es Störungen beim GPS-Signal oder beim Übertragen der Daten gab) bzw sich Fehler in den Daten durch

³² siehe etwa *Lowry & Macpherson*, *British Medical Journal* 1988, 657.

³³ vgl *Hand*, *European Journal of the Operational Research Society* 2005, 1109-1117; *Hand*, *Statistical Science* 2006, 1-14; *Verbeke, Dejaeger, Martens, Hur & Baesens*, *European Journal of the Operational Research Society* 2012, 211-229; *Lessmann ua*, *European Journal of Operational Research* 2015, 124-136(2015)

³⁴ vgl zB Seibt, SCHUFA.

³⁵ zB *Schröder ua* in *Schröder/Taeger* (Hrsg), *Scoring im Fokus*.

eine Verknüpfung mit fehlerhaften Daten ergaben. So verschlechterte sich der Score von Verbraucherinnen und Verbrauchern aufgrund veralteten Kartenmaterials: In diesen Fällen wurde eine Geschwindigkeitsüberschreitung aufgezeichnet, die tatsächlich keine war: Der Fahrer fuhr mit erlaubter Geschwindigkeit – jedoch in einem Bereich, in dem sich einige Zeit zuvor eine Baustelle befand, mit entsprechend reduzierter Geschwindigkeitsbegrenzung.

Verbeke et al. (2012) heben aus Sicht der Entwickler/innen von Algorithmen die Bedeutung der Datenqualität allgemein hervor und messen einer besseren Datenqualität eine höhere Priorität zu als ausgefeilteren Algorithmen zur Berechnung von Scores.

3. Qualität bei der Verhaltensvorhersage: Bonität

Ein Bonitäts-Score soll vorhersagen, mit welcher Wahrscheinlichkeit eine Kundin bzw ein Kunde vertragsgemäß seinen Zahlungsverpflichtungen nachkommen wird. Die Korrektheit und Präzision dieser Vorhersage hängt ganz entscheidend von der Qualität der zugrunde liegenden Daten und des Scoring-Algorithmus ab. Zur Sicherung der Qualität eines Algorithmus ist in Deutschland nach Paragraph 31 BDSG lediglich ein „wissenschaftlich anerkanntes mathematisch-statistisches Verfahren“ gefordert. Dies ist jedoch eine unzureichende Forderung.³⁶

Beim Bonitäts-Scoring werden zwar mathematisch-statistische Verfahren wie die logistische Regression eingesetzt, dieser Umstand sichert jedoch noch nicht automatisch die Qualität des Ergebnisses. Das anhand der logistischen Regression erhaltene Ergebnis kann je nach Auswahl der Merkmale, die für die Berechnung herangezogen werden, und der Qualität der Daten (siehe Abschnitt IV.B.2.) nämlich zu besseren oder schlechteren Prognosen führen. Zudem kann ein solches „mathematisch-statistisches Verfahren“ bei der Vorhersage eine bestimmte Personengruppe ungerechtfertigt diskriminieren oder auch nicht. Die Qualität der Vorhersage (wie auch die mögliche Diskrimination) muss man empirisch prüfen; die Forderung nach einem „wissenschaftlich anerkannten, mathematisch-statistischen Verfahren“ garantiert für sich genommen kein befriedigendes Ergebnis. Die Vorhersagequalität von Bonitäts-Scoring sollte daher mittels konkreter Gütemaße (zB Falsch-Positive-Rate, Hit-Rate) quantitativ gekennzeichnet werden.

Die Festlegung materieller Qualitätskriterien ist zu ergänzen durch eine entsprechende Prüfung von Scoring-Verfahren durch unabhängige Institutionen. Deshalb ist es erforderlich,

³⁶ Gerberding & Wagner, Zeitschrift für Rechtspolitik 2019, 116-119.

Aufsichtsbehörden so auszustatten, dass Scoring-Verfahren kompetent und unabhängig auf ihre Qualität überprüft werden können.

4. Qualität bei Scoring zur Verhaltenssteuerung: Gesundheit und Fahrverhalten

Dem gesetzlichen Auftrag folgend, sollen die Bonusprogramme der Krankenkassen in Deutschland gesundheitsbewusstes Verhalten fördern. Also würde man erwarten, dass Verhalten, welches wissenschaftlich nachgewiesen die Gesundheit stärkt, mit entsprechend mehr Punkten belohnt wird als Verhalten, für das dieser Nachweis nicht vorhanden oder unklar ist. In einer Marktstudie³⁷ wurde jedoch festgestellt, dass viele gesetzliche und private Krankenkassen wenige oder keine Punkte für nachweislich gesundheitsförderliches Verhalten vergeben, aber vergleichsweise viele Punkte für anderes Verhalten, für das in der medizinischen Forschung kaum Evidenz dafür vorliegt, dass der gesundheitliche Nutzen höher als der Schaden ist. Ein Beispiel: Starke Raucherinnen und Raucher verringern ihre Lebenserwartung um bis zu zehn Jahre, und entsprechend ihre Lebensqualität³⁸. Die Teilnahme an Untersuchungen zur Krebsfrüherkennung erhöhen dagegen die Lebenserwartung nicht oder bestenfalls einige Wochen³⁹ und sogenannte „individuelle Gesundheitsleistungen“ (IGeL) wie Glaukom-Früherkennung haben keinerlei erwiesenen Nutzen⁴⁰. Andere wie Eierstockkrebs-Früherkennung haben nicht nur keinen erwiesenen Nutzen, sondern erwiesene, massive Gesundheitsschäden: Jährlich verlieren geschätzt über 10.000 gesunde Frauen in Deutschland Eierstöcke wegen eines falsch-positiven Ultraschalls⁴¹. Trotz dieser wissenschaftlichen Ergebnisse vergeben nur acht der 45 untersuchten Bonusprogramme Punkte für Nicht-Rauchen, aber alle 45 vergeben Punkte für Krebsfrüherkennungsuntersuchungen.

Aus der Marktanalyse und den Interviews mit den Anbietern wird deutlich⁴², dass es bei den Bonusprogrammen nicht immer in erster Linie um die Förderung der Gesundheit geht, sondern um Kundenbindung, Neu-Kundenwerbung und auch den Verkauf anderer Versicherungen. In diesen Fällen wird das angegebene Ziel der Förderung der Gesundheit der Versicherten mit den Geschäftsinteressen der Versicherungen vermischt und im Extremfall dadurch ersetzt.

³⁷ SVRV (Hrsg). Verbrauchergerechtes Scoring 71 ff.

³⁸ Deutsches Krebsforschungszentrum (Hrsg, Gesundheitliche Folgen des Rauchens.

³⁹ Gigerenzer, British Medical Journal 2016, 352; Prasad, Lenzer & Newman, British Medical Journal 2016, 352.

⁴⁰ IGeL-Monitor (Hrsg), Messung.

⁴¹ Gigerenzer & Kolpatzik, British Medical Journal 2017, 357.

⁴² SVRV (Hrsg), Verbrauchergerechtes Scoring.

Bei der Telematik im Zusammenhang mit Kfz-Versicherungstarifen ist laut Aussagen der Anbieter das Ziel des Scoring-Verfahrens die Verringerung von vermeidbaren Unfällen durch die Verbesserung des Fahrverhaltens. Es liegen keine publizierten Untersuchungen vor, die belegen, dass diese Tarife Unfallrisiken tatsächlich reduzieren und sich auf beide Seiten, Versicherer und Versicherte, positiv auswirken. Auch hier ist es aufschlussreich genauer auf die Merkmale zu sehen, die in den Score eingehen.

Scoring- bzw Tarif-Merkmale wie zu schnelle Beschleunigung und Geschwindigkeitsübertretungen sind unter Kontrolle der Fahrer/innen, andere wie Stadtfahrten versus Landfahrten oder Tag- versus Nachtfahrten sind es nicht oder kaum. Das heißt, ein Fahrer bzw. eine Fahrerin kann erstere beeinflussen, aber letztere weniger. Ein unverzerrter Score über die Fahrfähigkeit der Fahrer/innen ist mit den genannten Faktoren wohl nicht zu erlangen. Zu berücksichtigen ist im Übrigen, dass solche Scoring-Verfahren eine unerwünscht diskriminierende Wirkung haben können. So ergibt sich ein schlechterer Score für Personen, die aus beruflichen Gründen darauf angewiesen sind, meist in der Stadt oder nachts mit dem Auto zu fahren.

Zusammenfassend lässt sich festhalten, dass die Anwendung eines mathematisch-statistischen Verfahrens, wie etwa der logistischen Regression, nicht die tatsächliche Qualität eines Scores bestimmt. Deshalb müssen Anbieter ein hohes Qualitätsniveau für Scoring-Verfahren sicherstellen, und zwar nicht nur für die angewandte Methode, sondern auch mit Blick auf das Ziel des Scorings. Problematisch ist etwa, wenn bei einem „Gesundheits-Score“ massiv gesundheitsschädliches Verhalten wie Rauchen kaum berücksichtigt wird, wohingegen Aktivitäten, deren Gesundheitsnutzen nicht belegt sind, mit hohem Gewicht positiv eingehen.

5. Fazit

Ein Score sollte das messen, was er vorgibt zu messen. Darüber hinaus müssen die Daten, auf deren Basis das Scoring stattfindet, korrekt und angemessen aktuell sein und auch in ausreichendem Umfang vorhanden sein. Insbesondere beim Bonitäts-Scoring ist zudem die korrekte Zuordnung von Informationen zu einer Person (*Entity Recognition*) von zentraler Bedeutung. Denn selbst ein qualitativ sehr hochwertiges Scoring-Verfahren wird bei mangelhafter Qualität der Eingabedaten unpräzise oder sogar falsche Ergebnisse liefern.

SVRV-FORDERUNG 2: SCORE-QUALITÄT GEWÄHRLEISTEN UND DATENQUALITÄT SICHERN

- Die Qualität von Scores soll von einer kompetenten und unabhängigen Stelle (wie einer „Digitalagentur“) überprüft werden. Die Qualität soll quantitativ mittels üblicher Gütemaße (zB Falsch-Positiv-Rate, Hit-Rate) für die Gesamtpopulation ausgewiesen werden, aber auch für relevante Teilpopulationen (zB nach Geschlecht, Alter, Bildung), um ungerechtfertigte Diskriminierung aufdecken zu können.
- Die Datenqualität bei der Score-Entwicklung ist zu gewährleisten und gegenüber den Aufsichtsbehörden zu dokumentieren. In der Anwendung der Verfahren ist die Korrektheit, Vollständigkeit und Aktualität der Daten sicherzustellen. Darüber hinaus ist die nicht-adequate Kennzeichnung des Ziels eines Scoring-Verfahrens aufzudecken und zu verhindern.

C. Gesellschaftliche Implikationen

Verbraucherbezogenes Scoring wirkt sich nicht nur auf individueller Ebene aus (zB über verhaltensbasierte Versicherungsprämien). Vielmehr kann ein solches Scoring vielschichtige Auswirkungen auf die gesamte Gesellschaft haben. Zudem stellen sich ethisch-moralische Fragen mit Blick auf das Spannungsverhältnis zwischen Individualisierung und Solidarität. Schließlich lassen sich auch grundsätzliche Debatten führen, etwa, inwiefern Grundwerte einer demokratischen Gesellschaft, wie die persönliche Freiheit, Autonomie und Solidarität, durch Scoring herausgefordert werden. Zudem wird untersucht werden müssen, ob Scoring zur sozialen Ungleichheit beiträgt oder diese vielmehr abbauen hilft und inwiefern bestimmte Verbrauchergruppen systematisch benachteiligt oder übervorteilt werden. Zum gegenwärtigen Zeitpunkt sind verschiedene beschriebene Implikationen eines zunehmenden Scorings für die Gesellschaft noch spekulativ.

1. Gesellschaftliche Funktionen

Grundsätzlich dienen bestimmte Scoring-Formen dem Funktionieren des wirtschaftlichen Austauschs und sind somit im Interesse der Gesellschaft und der Bürgerinnen und Bürger. Das traditionelle Bonitäts-Scoring schützt Unternehmen vor Zahlungsausfällen und ist somit eine wesentliche Grundlage für die Kreditvergabe an Verbraucherinnen und Verbraucher. Da säumigen Zahlern Kredite verweigert werden oder kostspieliger angeboten werden, bekommen zuverlässige Verbraucherinnen und Verbraucher bessere Konditionen als dies ohne Scoring möglich wäre. Scoringbasierte Entscheidungen können auch dazu beitragen, dass Verbraucherinnen und Verbraucher weniger Willkürentscheidungen, die ggf auf

bewussten oder unbewussten Vorurteilen oder Wahrnehmungsverzerrungen von (ausschließlich) menschlichen Entscheidern beruhen, ausgesetzt sind. Für Verbraucherinnen und Verbraucher sowie Unternehmen wird überdies die Vertragsabwicklung bei Dauerschuldverhältnissen erleichtert (zB der Abschluss von Handyverträgen), da auf ein standardisiertes System der Kredit- bzw Zahlungsfähigkeitsprüfung zurückgegriffen werden kann.⁴³

2. Implikationen bezüglich der Solidarität

Scoring und seine Konsequenzen können auf der Ebene der einzelnen Verbraucherinnen und Verbraucher als fair empfunden werden, wohingegen Scoring sich gesamtgesellschaftlich als unfair und ggf unsolidarisch erweisen kann und umgekehrt: Ist es auf individueller Ebene nicht beispielsweise fair, wenn sich der Beitrag der Versicherungsprämie am freiwilligen Verhalten einer Versicherungsnehmerin bzw eines Versicherungsnehmers bemisst, wenn ein Score also davon abhängt, wie gesundheitsbewusst sich jemand verhält bzw wie umsichtig jemand fährt? Ist es nicht unsolidarisch zu rasen und damit andere Verkehrsteilnehmer/innen zu gefährden? Auf individueller Ebene betrachtet, kann man sich beispielsweise auch fragen, warum man für das Erkrankungsrisiko der rauchenden Nachbarin bzw des rauchenden Nachbarn mitbezahlen soll, wenn man sich selbst gesundheitsbewusst verhält. Umgekehrt ist es ein gesellschaftliches Gut, sich hinsichtlich der Freiheit der Nachbarin bzw des Nachbarn bei seiner Lebensgestaltung solidarisch zu zeigen und ggf die durch das Rauchen anfallenden Gesundheitskosten mitzutragen. Was im Kontext von Scoring als solidarisch und fair erachtet bzw als ethisch-moralisch geboten eingeschätzt wird, muss in einem gesellschaftlichen Diskurs differenziert betrachtet und sorgsam abgewogen werden. Der Schutz Einzelner bzw spezifischer Gruppen reflektiert das gesamtgesellschaftliche Interesse einer auf Solidarität fußenden Gesellschaft. Hierbei müssen also die individuellen oder auch gruppenspezifischen Interessen mit den legitimen Interessen der Gesellschaft in Einklang gebracht werden. Das heißt auch, dass besonders schützenswerte Personen oder Gruppen durch Scoring keine (systematische) Benachteiligung erfahren sollen.

Die Frage, inwiefern nun Scoring die Solidarität innerhalb der Gesellschaft im Allgemeinen und das auf dem Solidaritätsprinzip fußende System der gesetzlichen Krankenversicherung (GKV) im Speziellen unterminieren könnte, wird bereits intensiv diskutiert⁴⁴. Im gesellschaftlichen Sinne lässt sich Solidarität als hohes soziales Gut einer demokratischen

⁴³ Schröder & Taeger (Hrsg), Scoring im Fokus.

⁴⁴ zB Deutscher Ethikrat (Hrsg), Big Data und Gesundheit.

Gesellschaft verstehen, das auf Basis geteilter Wertvorstellungen und Normen funktioniert. So besteht beispielsweise Konsens darüber, dass spezifische Gruppen, wie „verletzliche Verbraucherinnen und Verbraucher“ (zB sozial Schwache, Menschen mit Einschränkungen oder gesundheitlich Beeinträchtigte), in besonderem Maße hilfebedürftig sind, geschützt, gesellschaftlich integriert werden und ihnen soziale und ggf auch finanzielle Unterstützung zuteilwird.⁴⁵

Im Bereich der gesetzlichen Krankenversicherung bedeutet das Solidaritätsprinzip, dass sich der Versicherungsbeitrag nicht an dem Gesundheitsstatus oder am Gesundheitsverhalten orientiert, sondern jeweils am individuellen Einkommen der Versicherungsnehmerin bzw des Versicherungsnehmers ausrichtet. Versicherungsnehmer/innen mit einem guten Gesundheitszustand stehen dabei solidarisch für die Versicherungsnehmer/innen mit einem schlechteren Gesundheitszustand ein (auch wenn der schlechtere Gesundheitszustand auf möglicherweise freiwilligen Entscheidungen, wie zB dem Lebensstil, beruht). Allen Versicherungsnehmerinnen und Versicherungsnehmern stehen hierbei unabhängig von der Höhe der gezahlten Versicherungsbeiträge die gleichen Leistungen zu. Ein zentrales Prinzip der Solidargemeinschaft, das der GKV zugrunde liegt, ist der Risikoausgleich zwischen Gesunden und Kranken. Grundlage hierfür ist eine ausreichend große Anzahl an gesunden Beitragszahlerinnen und Beitragszahlern mit geringer Inanspruchnahme von Leistungen zu relativ gesehen wenigen Beitragszahlerinnen und Beitragszahlern mit Krankheiten und entsprechend höherer Leistungsanspruchnahme.

Da es sich im Kontext von Gesundheit um einen besonders sensiblen und schützenswerten Bereich handelt, ist die gesetzliche Krankenversicherung entsprechend stark reguliert. Die Einführung verhaltensbasierter Versicherungstarife in der GKV ist in Deutschland daher bislang keine Realität. Bei verhaltensbasierten Tarifen würden beispielsweise Personen systematisch benachteiligt, die nicht in der Lage sind, ein bestimmtes Verhalten auszuführen, oder die nicht bereit sind, gesundheitsbezogene und deshalb teils sensible Daten ihrer Versicherung zur Verfügung zu stellen. Würde beispielsweise eine bestimmte Anzahl von täglich zu absolvierenden Schritten für die Tarifierung herangezogen werden, hätten Personen mit Geheinschränkungen einen erheblichen Nachteil, denn sie müssten unter Umständen mehr bezahlen. Es besteht also die Gefahr, dass verhaltensbasierte Tarife das Prinzip der Solidargemeinschaft unterminieren. Insofern erweisen sich die bereits existierenden

⁴⁵ Micklitz ua, Plädoyer für eine differenzierte Strategie.

Bonusprogramme, innerhalb welcher bestimmte Personen benachteiligt werden können, bereits als problematisch.

3. Implikationen bezüglich der Gesellschaftsstruktur

Eine zunehmende freiwillige Selbstüberwachung beispielsweise des Schlafverhaltens, des Bewegungsverhaltens oder des Ernährungsverhaltens durch smarte Geräte und die gegenseitige Bewertung, zB in sozialen Netzwerken, können Druck erzeugen, sich gemäß den Idealvorstellungen einer Leistungsgesellschaft immer stärker selbst zu optimieren. Dies könnte zu einer Aufspaltung in Optimierte und Suboptimierte führen.⁴⁶ So ist denkbar, dass der soziale Status absehbar nicht mehr nur an der formalen Bildung, am Beruf und am Einkommen bemessen wird, sondern auch an einem allgemeinen Gesundheitsscore oder dem durch das Fahrverhalten erlangten Score. Bereits heute wird in einigen Lebensbereichen der Status und Wert einer Person durch Scores beschrieben, erzeugt und festgelegt – kurz: „Zahlen machen Leute“, wie von Steffen Mau⁴⁷ treffend formuliert.

Es ist bereits gängige Praxis, auf der Basis von Scoring eine Kategorisierung von Personen in spezifische Zielgruppen vorzunehmen, die es ermöglicht, Verbraucherinnen und Verbraucher unterschiedlich zu adressieren und zu behandeln. Eine solche Klassifizierung ist nicht nur deskriptiv und es geht dabei um mehr als die bloße Beschreibung von Unterschieden. Vielmehr kann sie auch eine soziale Selektion nach „wertvollen“ und „weniger wertvollen“ Verbraucherinnen und Verbrauchern implizieren.⁴⁸ Folglich besteht die Gefahr, dass es für Verbraucherinnen und Verbraucher aus den niedrigen Kategorien unter Umständen schwierig sein wird, beispielsweise einen Mobilfunkvertrag, eine Kfz-Versicherung oder eine private Krankenversicherung zu akzeptablen Konditionen abzuschließen. Hierbei stellt sich zudem die Frage, wie einfach oder schwierig es sein wird, von einer schlechten Kategorie in eine bessere wechseln zu können.⁴⁹

Durch die zunehmende Sammlung und Verknüpfung von Daten aus verschiedenen Lebensbereichen (zB Freizeitverhalten, Konsumverhalten, Zahlungsverhalten, Gesundheitsstatus, Bewegungsverhalten, Mitgliedschaften in sozialen Netzwerken, Wohngegend, Arbeitsleben, Einkommen oder Familienstand) zum Zwecke der

⁴⁶ Selke, Lifeloggging.

⁴⁷ Mau, Das metrische Wir.

⁴⁸ Selke, Lifeloggging.

⁴⁹ Mau, Das metrische Wir.

Kategorisierung von Verbraucherinnen und Verbrauchern⁵⁰, erscheint es durchaus denkbar, dass Scoring die Grundlage für die Herausbildung einer digitalen Klassengesellschaft bilden könnte. In dieser möglichen Zukunft würden die Menschen einem permanenten Wettbewerb um den besten Score ausgesetzt sein. Die dadurch potenziell zunehmende Individualisierung der Gesellschaft bei gleichzeitiger Entsolidarisierung könnte zu einer Schwächung des gesellschaftlichen Zusammenhalts führen.

Wenn, wie es in China für den kommerziellen Sesame-Score aufgezeigt wurde, auch „Freunde“ in sozialen Netzwerken nach dem Score ausgewählt werden⁵¹, da niedrige Scores anderer dem eigenen Score schaden können, sind neue Formen sozialer Ausgrenzung denkbar. In diesem Zusammenhang ist der soziale Druck, sich „Score-konform“ zu verhalten aber auch sich überhaupt erst scores zu lassen (falls man eine Wahl hat) nicht zu unterschätzen.

4. Implikationen bezüglich der Wahlfreiheit

Bei hinreichender Verbreitung von Telematiktarifen könnten immer größere Teile der Bevölkerung sich einem starken Anreiz oder auch einem Zwang unterworfen sehen, diese Tarife zu wählen. So ist es denkbar, dass beispielsweise bei Nicht-Teilnahme und Nicht-Preisgabe von persönlichen Daten auf ein potenziell höheres Risiko geschlossen werden könnte und es zu Stigmatisierung und (monetären) Sanktionen kommt. Personen mit Voraussetzungen für einen guten Score sind vermutlich eher geneigt, persönliche Daten preiszugeben. Daher ist es vorstellbar, dass sich auch andere Personen dazu gezwungen sehen, Daten preiszugeben, die zur Teilnahme an einem Telematiktarif notwendig sind, um etwaigen Nachteilen durch eine Nicht-Teilnahme zu entgehen (sog unravelling effect⁵²).

Eine gesellschaftliche Herausforderung wird es daher sein, demokratische Werte wie persönliche (Wahl-)Freiheit, Privatsphäre, Autonomie und Solidarität als hohes Gut zu wahren. Die Wahrung dieser Werte impliziert beispielsweise auch, sich dafür entscheiden zu können, vermeintlich unvernünftig zu handeln, sich nicht sozial erwünscht zu verhalten und sich aktiv dagegen zu entscheiden, gescort zu werden.

Die beschriebenen gesellschaftlichen Implikationen des gegenwärtig zunehmenden Scorings sind in vielerlei Hinsicht noch spekulativ. Sehr wahrscheinlich wird Scoring jedoch breite und

⁵⁰ Saetman, Schneider & Green, The politics of big data.

⁵¹ Kostka, China's Social Credit Systems.

⁵² Peppet, Northwestern University Law Review 2011, 1153-1204.

starke Auswirkungen auf die Gesellschaft haben – in welcher Form bleibt abzuwarten. Daher ist es wichtig, ein Bewusstsein für mögliche gesellschaftliche Konsequenzen von Scoring zu schaffen. In diesem Zusammenhang sollte insbesondere eine Diskussion darüber geführt werden, welche ethischen und moralischen Werte für eine freie demokratische Gesellschaft unerlässlich sind und welche Implikationen im Zusammenhang mit Scoring gesellschaftlich als inakzeptabel und welche als akzeptabel gelten sollen.

SVRV-FORDERUNG 3: TELEMATIKFREIE OPTION SICHERSTELLEN

Telematikfreie Optionen sollten im Versicherungsbereich (insbesondere Kfz-Haftpflichtversicherung, Krankenversicherung) auch zukünftig aufrechterhalten werden. Versicherungsnehmerinnen und -nehmern, die keinen Telematiktarif nutzen, dürfen im Vergleich zu Versicherungsnehmerinnen und -nehmern mit Telematiktarif keine substantiellen Nachteile entstehen. Telematiktarife sind so auszugestalten, dass sie sich wirtschaftlich selbst tragen und dadurch nicht indirekt zu Lasten der Versicherungsnehmerinnen und -nehmer gehen, die keinen Telematiktarif nutzen.

D. Verknüpfung von Scores zu einem Super-Score

Die chinesische Regierung plant, ab 2020 das Verhalten aller Bürgerinnen und Bürger sowie Unternehmen mit einem Sozialkreditsystem (*Social Credit Score*) zu bewerten.⁵³

„Aufrichtiges“ Verhalten im Sinne der Kommunistischen Partei soll belohnt, „unaufrichtiges“ Verhalten negativ sanktioniert werden. Dies kann Auswirkungen auf Ausbildung, Karriere und den gesamten Alltag bedeuten. Beispielsweise hat die Höhe des Scores Auswirkungen darauf, ob Flugzeug- oder Bahn-Tickets erworben werden können, oder ob die eigenen Kinder auf die besten Privatschulen dürfen. Bei einem hohen Score gibt es zB Vorzugsbehandlungen im Krankenhaus und beim Carsharing muss keine Kautions hinterlegt werden.⁵⁴

Die Etablierung eines solchen Sozialkreditsystem ermöglicht die ständige und totale Kontrolle des Verhaltens der Bürgerinnen und Bürger. Ein offizielles Ziel ist es, Vertrauen und Transparenz zwischen Marktteilnehmerinnen und -nehmern herzustellen bzw zu erhöhen. Erste Auswirkungen sind auch auf ausländische Kooperationspartner sowie auf Kundinnen

⁵³ Creemers, Planning outline; State Council of the People's Republic of China (Hrsg), Planning outline.

⁵⁴ Kostka, China's Social Credit System.

und Kunden zu beobachten.⁵⁵ 2014 verkündete der kommerzielle Score-Anbieter Alibaba eine Zusammenarbeit mit dem Luxemburgischen Konsulat: Bei einem entsprechend hohen *Sesame-Score* können Chinesinnen und Chinesen vereinfachte Visaanträge über ein Online-Portal stellen.⁵⁶

Derzeit finden in etwa 40 chinesischen Städten Pilotprojekte statt. Zudem bieten einige kommerzielle Anbieter Sozialkreditsysteme auf freiwilliger Basis an. Die Verfahren basieren, so wird berichtet, auf algorithmischen Big-Data-Analysen. Eine westliche Studie zur öffentlichen Wahrnehmung solcher Systeme zeigte, dass rund 80% der chinesischen Internetnutzer/innen die staatlichen und kommerziellen Sozialkreditsysteme in ihrem Land positiv bewerten.⁵⁷

Konsumenten-Scoring auf Basis riesiger Datenmengen und algorithmischer Berechnungen wird bei Bonusprogrammen, Telematiktarifen und Bonitätsscoring in Deutschland bislang noch nicht angewandt. Ein Trend zu solchen Geschäftsmodellen lässt sich außerhalb Deutschlands allerdings schon in verschiedenen Bereichen beobachten. Diese Entwicklung ist auch vor dem Hintergrund von spezialisierten Datenhändlern zu sehen.

Es stellt sich die Frage, ob Scoring-Systeme, die diese Daten aus verschiedenen Lebensbereichen verknüpfen, auch in Österreich oder Deutschland denkbar wären.⁵⁸ Zwar sind solche Entwicklungen in der westlichen Welt nicht von staatlicher Seite zu erwarten, doch Unternehmen könnten dieses Ziel verfolgen. Durch Verknüpfung personenbezogener Daten aus verschiedenen Lebensbereichen und Quellen ist die Bildung eines Super-Scores, ähnlich dem *Social Credit Score* in China, eine nicht auszuschließende Gefahr. Das Szenario eines solchen Super-Scores, der unter Umständen sogar aufgrund freiwilliger Einwilligungen von Konsumentinnen und Konsumenten, die ihre Daten zur Weitergabe an Dritte zur Verfügung zu stellen, zustande kommen könnte, wird im Folgenden erläutert.

1. Datenhandel und Datenakkumulation

Grundlage eines jeden Verbraucher-Scorings sind Verbraucherdaten, teilweise in erheblichen Mengen.⁵⁹ In vielen Lebensbereichen hinterlassen Verbraucherinnen und Verbraucher Datenspuren. Und die Digitalisierung hat dazu geführt, dass das Sammeln von

⁵⁵ Hoffmann, Social Credit.

⁵⁶ Alibaba Group (Hrsg), Alitrip.

⁵⁷ Kostka, China's Social Credit System.

⁵⁸ siehe dazu auch Al-Ani in Bundesministerium der Verteidigung (Hrsg), China.

⁵⁹ Dixon & Gellmann, The scoring of America.

personenbezogenen Daten so einfach ist wie nie zuvor. Ebenso ist das wirtschaftliche Interesse gewachsen, diese Daten gewinnbringend zu nutzen, zB in den Bereichen Werbung, Onlinehandel, Marktforschung und Politik. Ziel ist meistens, Betrug zu vermeiden, Personen zu identifizieren oder auch Marketingkampagnen und Kundenkommunikation effizienter zu gestalten, indem konkrete Zielgruppen präzise adressiert werden⁶⁰. Daten, die nicht selbst erhoben werden können, werden häufig von Datenhändlern dazugekauft. Deren Geschäftsmodell besteht im Sammeln, Aufbereiten und Verkaufen von individuumbezogenen Daten. In Deutschland soll 2014 das Marktvolumen im Bereich des Handels mit adress- und anderen personenbezogenen Daten bei rund 610 Mio. Euro gelegen haben.⁶¹

Etwa 1.000 Firmen weltweit haben sich darauf spezialisiert, Daten zu sammeln und zu verkaufen, meist ohne Wissen oder bewusst erteilte Zustimmung der Verbraucherinnen und Verbraucher, Anzahl steigend⁶². Einer der größten internationalen Datenfirmen ist die US Firma Acxiom, die auch in Deutschland tätig ist und bereits 2013 Daten über 44 Millionen Deutsche anbot. Viele Menschen sind bereit, ihre Daten gegen kleine Vorteile und Bequemlichkeiten weiter zu geben.⁶³

2. De-Anonymisierung von Daten

Anonymisierte Daten, also solche, die nicht auf die jeweiligen Verbraucher/innen rückführbar sind, werden vom Schutzversprechen des Datenschutzrechts von vornherein nicht erfasst und können daher einschränkungslos erhoben und verwendet sowie von Datenhändlern gehandelt werden. Der konkrete Personenbezug (typischerweise Name, Geburtsdatum, Anschrift) wird also aus den Datensätzen entfernt. Möglich ist auch eine Pseudonymisierung, dh Attribute wie Name oder Anschrift werden durch Pseudonyme oder Kennzeichen ersetzt⁶⁴.

Nicht zu unterschätzen ist hierbei jedoch die Gefahr der „De-Anonymisierung“ oder „Re-Personalisierung“ solcher anonymisierten oder pseudonymisierten Daten. Schon mit dem Vorliegen weniger verknüpfbarer Daten, die identifizierende Merkmale enthalten, sind anonymisierte bzw pseudonymisierte Personendaten wieder mit konkreten Personen

⁶⁰ OECD (Hrsg), economics of personal data.

⁶¹ Goldmedia (Hrsg), Ökonomischer Wert von Verbraucherdaten.

⁶² *McLaughlin*, Die Besserwisser.

⁶³ *Benndorf & Normann*, The willingness to sell personal data.

⁶⁴ Die Frage, inwieweit pseudonyme Daten nicht personenbezogen sind, ist umstritten. Im 26. Erwägungsgrund der DSGVO heißt es: „Einer Pseudonymisierung unterzogene personenbezogene Daten, die durch Heranziehung zusätzlicher Informationen einer natürlichen Person zugeordnet werden könnten, sollten als Informationen über eine identifizierbare natürliche Person betrachtet werden.“

verknüpfbar. So zeigte Sweeney⁶⁵, dass Postleitzahl, Geschlecht und Geburtsdatum ausreichen, um 87 % der US-amerikanischen Bevölkerung eindeutig zu identifizieren. Zur Identifizierung von immerhin 18 % der Bürgerinnen und Bürger reichen Landkreis, Geschlecht und Geburtsdatum. Man kann also argumentieren, dass in Zeiten von Big Data jeder Datenpunkt aufgrund unzähliger Vernetzungsmöglichkeiten mit personenbezogenen Daten potenziell personenbezogen ist⁶⁶. Das Zusammenführen und Abgleichen von mehreren Datensätzen mit personenspezifischen Merkmalen wie Kaufverhalten, Browserverläufen, Suchhistorien etc aus verschiedenen Quellen ermöglicht den Rückschluss auf konkrete individuelle Verbraucherinnen und Verbraucher. Big Data und leistungsfähige Algorithmen lassen dies ohne großen Aufwand zu. Weiterhin ergibt sich daraus die Möglichkeit, durch Datenhändler erworbene Daten auch für Verbraucher-Scoring zu verwenden, da diese in der Regel personenbezogen gebildet werden.

Zusammenfassend lässt sich feststellen, dass Datenakkumulation und -handel in Zeiten von Big Data eine wichtige Rolle spielen und dass mit Algorithmen ein Identifizieren konkreter Verbraucherinnen und Verbraucher auch in großen, anonymisierten Datensätzen möglich ist. Durch rein technische Maßnahmen lässt sich dieses Potenzial nicht kontrollieren. Entsprechend groß sind die juristischen Herausforderungen, gesellschaftlich unerwünschte Personalisierung von Daten durch Gesetze und Verordnungen wirksam zu verbieten.

3. Verknüpfung von Daten zum Super-Score

Mit der steigenden Bedeutung des Datenhandels wächst das Potenzial, Daten aus verschiedensten Lebensbereichen in einer Datenbank und in einem Unternehmen zu vereinen. Potenziell können also viele Daten aus verschiedenen und vermeintlich unabhängigen Lebensbereichen konkreten Verbraucherinnen und Verbrauchern zugeordnet werden und dann auch als Grundlage für Scoring verwendet werden. Ein solcher Super-Score würde bedeuten, dass einzelne Verhaltensweisen von Verbraucherinnen und Verbrauchern lebensbereichsübergreifend weitreichende Konsequenzen nach sich ziehen können.

Hinweise auf eine „Zweckentfremdung“ von Daten und Scores finden sich zum Beispiel in den USA. Bonitätsauskünfte (zB der FICO Score) werden von Autoversicherungsanbietern und Hauseigentümersicherungen zur Beitragskalkulation genutzt⁶⁷. Von Interesse ist hierbei nicht, ob die Kundin oder der Kunde mit einer bestimmten Wahrscheinlichkeit einen

⁶⁵ Sweeney, Simple Demographics.

⁶⁶ Boehme-Neßler, Datenschutz und Datensicherheit 2016, 419-423.

⁶⁷ Consumer Reports, The secret score; Dixon & Gellmann, The scoring of America.

Kredit zurückzahlen wird, sondern die Informationen fließen in die Berechnung einer Wahrscheinlichkeit ein, dass die Kundin oder der Kunde bereit ist, höhere Versicherungsbeiträge zu zahlen, oder dass er einen Schaden melden wird⁶⁸.

Eine internationale Mitgliederbefragung der Society for Human Resource Management⁶⁹ zeigte, dass fast die Hälfte (47 %) der 430 befragten Arbeitgeber einen Bonitätscheck durchführt, bevor sie eine Bewerberin oder einen Bewerber einstellt. Hierbei geht es ebenfalls nicht darum, konkret die Kreditwürdigkeit einer Person zu erfahren, sondern darum, aus Bonitätsinformationen Eigenschaften wie Vertrauenswürdigkeit oder Zuverlässigkeit abzuschätzen⁷⁰.

Die Zweckentfremdung von Daten wird durch Datenhandel befördert. Insbesondere Bonitäts-Scoring ist bereits eng mit dem Geschäftsbereich Datenhandel verknüpft. Alle Auskunftsteien, die an der Befragung des SVRV teilnahmen, gaben an, dass sie den sich bereits im Ausland abzeichnenden Trend der Einbeziehung von Daten aus dem Internet und den Sozialen Medien in das Bonitätsscoring beobachten⁷¹. Zu erwähnen ist an dieser Stelle außerdem, dass ein solches Potenzial auch unabhängig von Datenhändlern vorhanden ist – beispielweise bei großen Versicherungsunternehmen, die verschiedene Versicherungssparten betreiben. Daten könnten hier mit dem Einverständnis der Versicherten zusammengelegt und gemeinsam ausgewertet werden, auch verhaltensbasierte Daten. In der Befragung des SVRV gaben beispielsweise 37 % der befragten Krankenkassen an, dass sie sich vorstellen könnten, Daten aus der elektronischen Patientenakte für den Bereich Bonusprogramme oder verhaltensbasierte Tarife zu nutzen⁷².

Das Szenario eines künftigen Super-Scores, in dem Verhalten in verschiedenen Lebensbereichen dokumentiert, analysiert und für die Score-Berechnung verwendet wird, erscheint im beschriebenen Kontext von innovativen Geschäftsmodellen, Big Data, Datenhandel und De-Anonymisierung durchaus plausibel. Während Entwicklungen wie das Sozialkreditsystem in China von staatlicher Seite in Deutschland sicher nicht zu erwarten sind, stellt sich die Frage, ob Verbraucherinnen und Verbraucher besser vor ähnlichen kommerziellen Entwicklungen geschützt werden müssen – auch durch Aufklärung. Wenn große Datenmengen über Verbraucherverhalten leicht zu erhalten und Analysewerkzeuge

⁶⁸ O'Neil, Weapons of Math Destruction.

⁶⁹ Society for Human Resource Management (Hrsg), Background checking.

⁷⁰ O'Neil, Weapons of Math Destruction.

⁷¹ SVRV (Hrsg), Verbrauchergerechtes Scoring 88.

⁷² SVRV (Hrsg), Verbrauchergerechtes Scoring 88.

profitabel einsetzbar sind, wird sich Datenhandel und Verbraucher-Scoring weiter verbreiten; nicht zuletzt, da Verbraucherinnen und Verbraucher dazu neigen ihre Daten per Klick freizugeben⁷³. Es könnte also sein, dass wir aus Unachtsamkeit freiwillig in ein Super-Score-Szenario „hineinrutschen“. Dass Unternehmen an Super-Scores großes Interesse haben, zeigt zB Douglas Merrill, der Gründer der Auskunftsei Zest Finance, dessen Gegenwartsanalyse auch als Warnung verstanden werden kann: „We feel like all data is credit data, we just don't know how to use it yet“⁷⁴.

Zentral ist: Die im Prinzip sehr wirksamen Sicherungen des europäischen Datenschutzrechts gegen De-Anonymisierungen laufen leer, wenn die Betroffenen freiwillig einer Datenverknüpfung zustimmen. Die repräsentative Bevölkerungsbefragung für das SVRV-Gutachten zeigt⁷⁵, dass ein Drittel der erwachsenen Bevölkerung in Deutschland sich die Zusammenführung von Kfz- und Gesundheits-Scores vorstellen kann: Die Befragten, die sich das vorstellen können, gaben als Argumente dafür an, dies wäre gerecht oder sie könnten davon persönlich profitieren. Damit besteht in der Tat die Gefahr, dass viele Menschen in Deutschland auch ohne einen staatlich verordneten Super-Score (wie in China) einer Verknüpfung ihrer Daten und Scores zustimmen, um nicht nur bessere Versicherungstarife, sondern allerlei Vergünstigungen, beispielsweise beim Mieten einer Wohnung, oder Bequemlichkeiten, etwa bei Tischreservierungen, Zugfahrkarten und Flugtickets, zu bekommen.

Auf die Frage nach einem Sozialkreditsystem in Deutschland wären immerhin 9 % dafür, so ein soziales Punktesystem einzuführen.⁷⁶ China ist derzeit dabei, das Sozialkreditsystem einschließlich Software und Hardware in anderen Staaten zu exportieren; Venezuela ist einer der ersten. Inwieweit die westlichen Demokratien sich in der Zukunft gegenüber effizienten digitalen Autokratien behaupten könnten, welche schnell entscheiden können und mit Hilfe des Super-Scores das Verhalten ihrer Bürgerinnen und Bürger steuern, ist zumindest eine Frage wert.

⁷³ Treiblmaier, Daten- und Informationsqualität 2018, 325-338.

⁷⁴ Hardy, Just the facts.

⁷⁵ SVRV (Hrsg), Verbrauchergerechtes Scoring 93 ff.

⁷⁶ SVRV (Hrsg), Verbrauchergerechtes Scoring 270.

SVRV-FORDERUNG 4: ZUSAMMENFÜHRUNG VON SCORES VERHINDERN

Die De-Anonymisierung und Zusammenführung einzelner Scores durch kommerzielle Konzerne zu einem Super-Score soll frühzeitig beobachtet und verhindert werden. Dazu sollten Gesetzgeber wie Aufsichtsbehörden konkret prüfen, ob bestehende Instrumente der DSGVO (wie Zweckbindungsgrundsatz und Kopplungsverbot) einsetzbar sind oder andere Rechtsgrundlagen geschaffen werden müssen.

E. FAZIT UND AUSBLICK

Der vorliegende Beitrag hat einen Einblick in wesentliche Punkte des Gutachtens „Verbrauchergerechtes Scoring“ des deutschen Sachverständigenrats für Verbraucherfragen (SVRV) gegeben und damit die Herausforderungen, Chancen und Gefahren, wie sie im Kontext von Scoring in zentralen Lebensbereichen von Verbraucherinnen und Verbrauchern auftreten können, umrissen. Darüber hinaus wurden Empfehlungen gegeben, welche (verbraucherpolitischen) Maßnahmen und Voraussetzungen nötig sind, um Scoring fair und verbrauchergerecht zu gestalten.

Hier nur am Rande erwähnte, aber ebenso wichtige Empfehlungen aus dem Gutachten des SVRV reagieren auf das (teils defizitäre) Wissen von Verbraucherinnen und Verbrauchern über Scoring und betreffen die unrechtmäßige Diskriminierung im Zusammenhang mit den eingesetzten Algorithmen sowie die Notwendigkeit einer Stärkung von Aufsichtsbehörden.⁷⁷

Die in das SVRV-Gutachten eingeflossenen Studien und Expertengespräche konnten die hohe Relevanz von Scoring in den Lebensbereichen Bonität, Gesundheit und Kfz-Versicherung verdeutlichen. Insbesondere die Expertengespräche mit Unternehmen und Wissenschaftlerinnen und Wissenschaftlern deuten darauf hin, dass ein zunehmendes Interesse von Unternehmen besteht, Scoring auch auf weitere – nicht ausschließlich verbraucherrelevante – Bereiche auszuweiten, wie etwa bei der Personalauswahl und -beurteilung (Stichwort *people analytics*).

Das Gutachten des SVRV zeigt mit der Ausweitung von algorithmischen Scorings einhergehende Unsicherheiten und Missstände auf: Von Unternehmen werden gegenwärtig häufig Scores verwendet, die wissenschaftlichen Qualitätskriterien nicht genügen, intransparent sind und diskriminierende Ergebnisse liefern. Auf Seiten der Gescorten besteht

⁷⁷ SVRV (Hrsg), Verbrauchergerechtes Scoring 4ff.

Unwissen und Unkenntnis darüber, welche Merkmale zum Scoring herangezogen werden, und welchem Zweck das Scoring dient und es gibt immer weniger Möglichkeiten, sich dem Scoring zu entziehen oder unfaires Scoring anzufechten.

Abschließend sei aber auch noch einmal deutlich hervorgehoben, dass Scoring grundsätzlich in vielen Bereichen, wie beispielsweise bei der Beurteilung der Bonität, wichtige positive Funktionen sowohl für Verbraucher/innen als auch für Unternehmen erfüllt. Transparentes und qualitätsgesichertes Scoring, das tatsächlich zu Verhaltensänderungen führen würde, wie zum Beispiel eine für sich und andere weniger gefährdende Fahrweise, ist grundsätzlich sinnvoll und begrüßenswert, da es dem Einzelnen und auch der Gesellschaft zu Gute käme.

Welchen Einfluss Scoring zukünftig auf die verschiedenen Lebensbereiche von Verbraucherinnen und Verbrauchern haben wird, kann zum gegenwärtigen Zeitpunkt nur spekuliert werden. Damit Scoring jedoch in Zukunft gerecht, nicht unzulässig diskriminierend und transparent abläuft, ist die Politik schon heute gefragt, Maßnahmen zu ergreifen, die dies gewährleisten. Unser Beitrag bietet dafür grundlegende Hinweise und Vorschläge.

Über gesetzliche Maßnahmen hinaus ist jedoch auch ein gesellschaftlicher Diskurs darüber notwendig, unter welchen Umständen Scoring akzeptabel bzw. inakzeptabel ist. Auch ist es unerlässlich, in zentralen Lebensbereichen, wie bei der Kranken- und Kfz-Versicherung den Bürgerinnen und Bürgern stets die Wahl zu überlassen, ob sie an einem Scoring teilnehmen wollen oder nicht. Denn echte Wahlfreiheit impliziert auch, sich frei zu entscheiden, wie man leben und welche Daten man teilen will, und auch die Möglichkeit, sich sozial unerwünscht, unangepasst oder „unvernünftig“ zu verhalten. Dies ist ein hohes Gut freiheitlicher demokratischer Gesellschaften, das es zu bewahren gilt.

LITERATURVERZEICHNIS

Al-Ani, Das chinesische Social Credit System – ein digitales Gesellschaftsmodell mit globalen Auswirkungen, in Bundesministerium der Verteidigung (Hrsg), China – Perspektiven und Herausforderungen. Bericht zur 3. Sitzung des Netzwerks „Strategie und Vorausschau“ (2018).

Alber/Lapuschkina/Seegerer, iNNvestigate neural networks!, <https://arxiv.org/pdf/1808.04260.pdf>

Alibaba Group (Hrsg), Alitrip Introduces Credit-based Visa Application Service for Qualified Chinese Travelers (2014), https://www.alibabagroup.com/en/news/press_pdf/p150604.pdf

- Arentz*, Unisex-Tarife in der Krankenversicherung: Folgen für die deutsche private Krankenversicherung und das Reformkonzept der Bürgerversicherung, Otto-Wolff-Institut Discussion Paper No. 02 2013.
- Benndorf/Normann*, The willingness to sell personal data, *The Scandinavian Journal of Economics* 2018, 1260-1278.
- Boehme-Neßler*, Das Ende der Anonymität. Wie Big Data das Datenschutzrecht verändert, *Datenschutz und Datensicherheit* 2016, 419-423.
- Burrell*, How the machine ‘thinks’: Understanding opacity in machine learning algorithms, *Big Data & Society* 2016.
- Consumer Reports*, The secret score behind your rates (2015), <https://www.consumerreports.org/cro/car-insurance/credit-scores-affect-auto-insurance-rates/index.htm>
- Creemers*, Planning outline for the construction of a social credit system (2014-2020) [Inoffizielle Übersetzung der Direktive der Zentralen Volksregierung] (2014), <https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/>
- Daston*, The domestication of risk: Mathematical probability and insurance 1650-1830, in Krüger/Daston/Heidelberger (Hrsg), *The probabilistic revolution. Volume 1: Ideas in History* (1987).
- Deutscher Ethikrat (Hrsg), *Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme* (2017).
- Deutsches Krebsforschungszentrum (Hrsg), *Gesundheitliche Folgen des Rauchens* (2010), https://www.dkfz.de/de/tabakkontrolle/Gesundheitliche_Folgen_des_Rauchens.html
- Dixon/Gellmann*, The scoring of america: how secret consumer scores threaten your privacy and your future (2014), http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf
- Durand*, Risk elements in consumer instalment financing (1941).
- Ermert*, Ethische Grenzen: Streit in der EU um Regeln für künftige KI-Systeme (2019), <https://www.heise.de/ct/artikel/Ethische-Grenzen-Streit-in-der-EU-um-Regeln-fuer-kuenftige-KI-Systeme-4405247.html>.
- Gesamtverband der Deutschen Versicherungswirtschaft e.V. (Hrsg), So setzt sich der Versicherungsbeitrag für einen Pkw zusammen (2016), <https://www.gdv.de/de/themen/news/so-setzt-sich-der-versicherungsbeitrag-fuer-einen-pkw-zusammen-11804>
- Gerberding/Wagner*, Qualitätssicherung für „Predictive Analytics“ durch digitale Algorithmen, *Zeitschrift für Rechtspolitik* 2019, 116-119.
- Gesellschaft für Informatik (Hrsg), *Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen* (2018).
- Gigerenzer*, Liebestrunken – Vermittlungsbörse schießt statistisches Eigentor (2015), <http://www.rwi-essen.de/unstatistik/50>
- Gigerenzer*, Full disclosure about cancer screening, *British Medical Journal* 2016, 352.
- Gigerenzer/Rebitschek/Wagner*, Eine vermessene Gesellschaft braucht Transparenz, *Wirtschaftsdienst* 2018, 860-868.

- Gigerenzer/Kolpatzik*, How new fact boxes are explaining medical risk to millions, *British Medical Journal* 2017, 357.
- Gigerenzer/Wagner/Müller*, Wie man Licht in die Black Box wirft (2018), <http://www.faz.net/aktuell/feuilleton/debatten/wie-man-algorithmen-transparent-machen-kann-15652267.html>
- Gluba*, Predictive Policing – eine Bestandsaufnahme, *Kriminalistik - Unabhängige Zeitschrift für die kriminalistische Wissenschaft und Praxis* 2014, 347–352.
- Goldmedia (Hrsg), *Ökonomischer Wert von Verbraucherdaten für Adress- und Datenhändler. Studie im Auftrag des Bundesministeriums der Justiz und für Verbraucherschutz* 2017, https://www.goldmedia.com/fileadmin/goldmedia/2015/Studien/2017/Verbraucherdaten_BMJV/Studie_Wert_Daten_Adresshaendler_Goldmedia_BMJV_2017.pdf
- Hand/Henley*, Statistical classification methods in consumer credit scoring: a review, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 1997, 523-541.
- Hand*, Good practice in retail credit scorecard assessment, *European Journal of the Operational Research Society* 2005, 1109-1117.
- Hand*, Classifier technology and the illusion of progress, *Statistical Science* 2006, 1-14.
- Hardy*, Just the facts. Yes, all of them (2012), <https://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html>
- Hoffmann*, Social credit: technology-enhanced authoritarian control with global consequences, *ASPI International Cyber Policy Centre Policy Brief* 2018.
- IGeL-Monitor (Hrsg), *Messung des Augeninnendrucks zur Früherkennung eines Glaukoms – Ergebnisbericht* (2012), https://www.igelmonitor.de/fileadmin/user_upload/Messung_des_Augeninnendrucks_Ergebnisbericht.pdf
- infas (Hrsg), *Verbraucher-Scoring: Repräsentativbefragung zur Akzeptanz und Kenntnis über (neuartige) Scoring-Methoden. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen* (2018).
- Just/Latzer*, Governance by algorithms: Reality construction by algorithmic selection on the internet, *Media, Culture & Society* 2016, 238-258.
- Kostka*, China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval (2018), <https://ssrn.com/abstract=3215138>
- Lessmann/Baesens/Seow/Thomas*, Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, *European Journal of Operational Research* 2015, 124-136.
- Lintorf*, Zur individuellen und gesellschaftlichen Bedeutung von Schulnoten, in Lintorf (Hrsg), *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale* (2012).
- Lowry/Macpherson*, A blot on the profession, *British Medical Journal (Clinical Research ed.)* 1988, 657.
- Mau*, *Das metrische Wir. Über die Quantifizierung des Sozialen* (2017).
- McLaughlin*, *Die Besserwisser* (2013), <https://www.zeit.de/2013/28/acxiom>
- Micklitz/Oehler/Piorkowsky/Reisch/Strünck*, Der vertrauende, der verletzte oder der verantwortungsvolle Verbraucher? Plädoyer für eine differenzierte Strategie in der Verbraucherpolitik. Stellungnahme des wissenschaftlichen Beirats Verbraucher- und Ernährungspolitik beim BMELV (2010).

Montavon/Samek/Müller, Methods for interpreting and understanding deep neural networks, *Digital Signal Processing* 2018, 1-15.

OECD (Hrsg), Exploring the economics of personal data: a survey of methodologies for measuring monetary value, *OECD Digital Economy Papers* No. 220 2013.

O'Neil, Weapons of math destruction: how big data increases inequality and threatens democracy (2016).

Passig, Fünfzig Jahre Black Box, *Merkur - Deutsche Zeitschrift für Europäisches Denken* 2017, 16-30.

Peppet, Unraveling privacy: The personal prospectus and the threat of a full-disclosure future, *Northwestern University Law Review* 2011, 1153-1204.

Perel/Elkin-Koren, Black box tinkering: Beyond disclosure in algorithmic enforcement, *Florida Law Review* 2017, 181-221.

Prasad/Lenzer/Newman, Why cancer screening has never been shown to “save lives” – and what we can do about it, *British Medical Journal* 2016, 352.

Ribeiro/Singh/Guestrin, Model-agnostic interpretability of machine learning (2016), arXiv:1606.05386

Rothmann/Sterbik-Lamina/Peissl, Credit Scoring in Österreich (2014).

Saetman/Schneider/Green, The politics of big data. Big data, big brother? (2018).

Samek/Wiegand/Müller, Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models, *ITU Journal: ICT Discoveries* 2017, 1-10.

Schröder/Taeger (Hrsg), Scoring im Fokus: ökonomische Bedeutung und rechtliche Rahmenbedingungen im internationalen Vergleich (2014).

Schröder/Lang/ Lerbs/Radev, Ökonomische Bedeutung und Funktionsweise von Credit Scoring, in Schröder/Taeger (Hrsg), Scoring im Fokus: ökonomische Bedeutung und rechtliche Rahmenbedingungen im internationalen Vergleich (2014).

Seibt, Wie ich bei der SCHUFA zum "deutlich erhöhten Risiko" wurde (2018), <http://www.spiegel.de/wirtschaft/service/SCHUFA-wie-ich-zum-deutlich-erhoehten-risiko-wurde-a-1193506.html>

Selbst/Powles, Meaningful information and the right to explanation, *International Data Privacy Law* 2017, 233-242.

Selke, Lifelogging. Wie die digitale Selbstvermessung unsere Gesellschaft verändert (2014).

Sharp, Policy Implications of People Analytics and the Automated Workplace, in Kiggins (Hrsg), *The Political Economy of Robots* (2018), 61-80.

Society for Human Resource Management (Hrsg), Background checking – the use of credit background checks in hiring decisions (2012), <https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/creditbackgroundchecks.aspx>

State Council of the People's Republic of China (Hrsg), Planning Outline for the Construction of a Social Credit System (2014-2020). Report from the National Development and Reform Commission (2014), http://www.ndrc.gov.cn/gzdt/201501/t20150105_659408.html

SVRV (Hrsg), Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen (2018).

Sweeney, Simple demographics often identify people uniquely (Data Privacy Working Paper 3) (2000).

Thomas/Crook/Edelman, Credit scoring and its applications, second edition (2017).

- Treiblmaier*, Ein Entscheidungsmodell zur Weitergabe persönlicher Daten im Internet, Daten-und Informationsqualität 2018, 325-338.
- Verbeke/Dejaeger/Martens/Hur/Baesens*, New insights into churn prediction in the telecommunication sector: A profit driven data mining approach, European Journal of Operational Research 2012, 211-229.
- Wachter/Mittelstadt/Floridi*, Why a right to explanation of automated decision-making does not exist in the general data protection regulation, International Data Privacy Law 2017, 76-99.
- Wagner-Braun*, Zur Bedeutung berufsständischer Krankenkassen innerhalb der privaten Krankenversicherung in Deutschland bis zum Zweiten Weltkrieg (2002).
- Wischmeyer*, Regulierung intelligenter Systeme, Archiv des öffentlichen Rechts 2018, 1-66.
- Zehlike/Wagner*, Neutrale Algorithmen? Von wegen! (2019), <https://www.sueddeutsche.de/kultur/kuenstliche-intelligenz-algorithmen-diskriminierung-1.4432946?reduced=true>.