

Gigerenzer, Gerd; Rebitschek, Felix G.; Wagner, Gert G.

Article

Eine vermessene Gesellschaft braucht Transparenz

Wirtschaftsdienst

Suggested Citation: Gigerenzer, Gerd; Rebitschek, Felix G.; Wagner, Gert G. (2018) : Eine vermessene Gesellschaft braucht Transparenz, Wirtschaftsdienst, ISSN 1613-978X, Springer, Heidelberg, Vol. 98, Iss. 12, pp. 860-868,
<https://doi.org/10.1007/s10273-018-2378-4>

This Version is available at:

<https://hdl.handle.net/10419/196521>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Gerd Gigerenzer, Felix G. Rebitschek, Gert G. Wagner*

Eine vermessene Gesellschaft braucht Transparenz

Ende Oktober 2018 hat der Sachverständigenrat für Verbraucherfragen das Gutachten „Verbrauchergerechtes Scoring“ vorgelegt. Es geht dabei um die sich ausbreitende Praxis, immer mehr Aspekte des Verhaltens der Bürger durch „Scores“ vorherzusagen oder zu steuern. Scoring ist durch eine Reihe von Kernproblemen gekennzeichnet: mangelnde Transparenz und Qualitätskontrolle für Daten und Algorithmen, ungenügende personelle sowie technische Ausstattung der Aufsicht und die zukünftige Gefahr eines kommerziellen Super-Scores ähnlich dem Sozialkredit-System in China. Anbieter von sensitiven Scores sollten verpflichtet werden, den Betroffenen alle Merkmale und deren Gewichtung offenzulegen. Grundsätzlich sind nicht Scoring-Algorithmen, sondern eine nicht hinreichend leistungsstarke Aufsicht das Problem.

Digitalisierung ist in aller Munde, Scoring noch nicht. Was aber ist „Scoring“? Im Gutachten des Sachverständigenrats für Verbraucherfragen (SVRV) ist die Definition eng gezogen: „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 („Computerprogramm“).“¹ Scoring ist nichts Neues; es wird in einer digitalisierten Welt aber in einer bisher nicht gekannten Intensität möglich.² Schul- und Examensnoten sind klassische Scores, die für die Studienzulassung auch zu einem einzigen Score zusammengezogen werden können. Die „Punkte in Flensburg“ des Kraftfahrt-Bundesamtes und der Elo-Score eines Schachmeisters sind weitere Beispiele aus der analogen Welt. Diese Systeme werden akzeptiert da sie völlig transparent sind (was nicht ausschließt, dass man über einzelne Noten und Punkte durchaus auch streiten kann).

Der wahrscheinlich bekannteste Score in Deutschland ist der Schufa-Wert, mit dem die Kreditwürdigkeit erfasst wird. Dieser Wert kann entscheiden, welche Zahlungs-

variante Online-Händler ihrem Kunden anbieten – bei zu niedrigem Wert sieht der Kunde nur Optionen der Vorauszahlung, jedoch keine Option zur Zahlung nach Erhalt der Ware. Der Schufa-Wert kann auch ausschlaggebend dafür sein, ob ein Vermieter einem Bewerber eine Wohnung anbietet und ob eine Bank einem Kunden einen Kredit gibt und zu welchen Konditionen. Weniger bekannt, aber im Gutachten ebenfalls untersucht wurden die Scores anderer Auskunfteien, der Kfz-Haftpflichtversicherungen für sogenannte Telematik-Tarife und jene Scores, die Bonusprogrammen von Krankenversicherungen zugrunde liegen.

Diese sensiblen Lebensbereiche sind zu unterscheiden von Scoring in anderen Bereichen wie etwa Partner-Börsen, derer man sich nicht bedienen muss.³

Der Gesetzgeber sollte kurzfristig mit einer Positivliste regulierungswürdiger Scorings starten, um einer langwierigen Aushandlung einer allgemeingültigen Definition von sensitiven Scorings zuvorzukommen. Im Folgenden treffen wir eine Auswahl der im SVRV-Gutachten gegebenen acht Empfehlungen und konzentrieren uns auf die unserer Ansicht nach vier wichtigsten:

1. Scoring für Verbraucher verständlich machen,
2. Qualität von Score und Daten gewährleisten,
3. Aufsicht durch eine Digitalagentur stärken,
4. Super-Scores verhindern.

* Gerd Gigerenzer und Gert G. Wagner waren federführend für das Gutachten „Verbrauchergerechtes Scoring“ des Sachverständigenrats für Verbraucherfragen verantwortlich, Felix G. Rebitschek hat beide unterstützt.

1 Vgl. Sachverständigenrat für Verbraucherfragen (SVRV): Verbrauchergerechtes Scoring, Gutachten, Berlin 2018, http://www.svr-verb-raucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf (25.11.2018), S. 15.

2 A. Oehler: Grundsätze ordnungsgemäßer Bewertung durch Scoring, in: Wirtschaftsdienst, 97. Jg. (2017), H. 10, S. 748-751; vgl. SVRV, a. a. O., S. 22-23.

3 Vgl. SVRV, a. a. O., S. 140.

Scoring für Verbraucher verständlich machen

Der Schufa-Score macht ein grundlegendes Problem deutlich: das Wechselspiel zwischen dem mangelnden Wissen in der Bevölkerung und der mangelnden Transparenz der Auskunftsteien. Eine repräsentative Erhebung im Auftrag des SVRV zeigt, dass die meisten Deutschen kaum etwas über die Merkmale wissen, die tatsächlich in ihren Score eingehen. Beispielsweise denken zwei von drei Menschen in Deutschland fälschlicherweise, dass ihr Beruf und ihr Vermögen zählen.⁴ Doch das Unwissen ist nicht allein selbstverschuldet, sondern liegt an der mangelnden Transparenz der Anbieter. Nach der derzeitigen Rechtsprechung (sogenanntes „Schufa-Urteil“ des Bundesgerichtshofs 2014) sind Auskunftsteien nicht verpflichtet, den Bürgern zu erklären, wie ihr Score zustande gekommen ist. Bei bestimmten Auskunftsteien wird dies sogar auffällig zu verhindern versucht.⁵ Lediglich die Aufsichtsbehörden, etwa Datenschutzbehörden, erfahren wie der Algorithmus (vgl. Kasten 1) zur Berechnung des Scores arbeitet. Daher können auch jene Bürger das nicht verstehen, die es gerne wissen möchten. Wir fordern dagegen, das Recht der Verbraucher auf Aufklärung über das Recht der Anbieter auf Geheimhaltung zu stellen.

Forderung 1: Der Gesetzgeber soll sicherstellen, dass Scoring-Anbieter den Verbrauchern alle Merkmale, auf deren Basis sie gescoret werden, sowie deren Gewichtung auf verständliche und leicht zugängliche Weise offenlegen.⁶

Ein bekanntes Argument gegen diese Forderung ist die Wahrung des Geschäftsgeheimnisses. Dem steht entgegen, dass die Anbieter zwar alle Merkmale und deren Gewichtung offenlegen sollten, nicht aber den Quellcode oder organisatorische Einzelheiten, die für das Verständnis der Verbraucher nicht wesentlich sind. Die Gefahr, dass Konkurrenten dann das Verfahren der Score-Berechnung übernehmen könnten, ist daher nicht gegeben.

⁴ Vgl. ebenda, S. 94 ff.

⁵ Schufa: „OpenSchufa“-Kampagne irreführend und gegen Sicherheit und Datenschutz in Deutschland, 15.2.2018, <https://www.schufa.de/de/ueber-uns/themenportal/detailseite/themenportal-detailseite.9856.jsp> (18.11.2018). Zu ersten Ergebnissen von OpenSchufa siehe Spiegel Data und BR Data, Blackbox Schufa, <http://www.spiegel.de/wirtschaft/service/schufa-so-funktioniert-deutschlands-einflussreichste-auskunftei-a-1239214.html> (2.12.2018).

⁶ Dies ist die Forderung der Autoren, wohingegen die Mehrheit des SVRV eine schwächere Version vertritt, nämlich, dass nicht alle, sondern nur die „wesentlichen“ Merkmale offengelegt werden sollen. Damit würde man es jedoch wieder den Verbrauchern erschweren, zu verstehen, wie der Score zustande kam. Und wenn es wirklich *alle* „wesentlichen“ Merkmale sind, ist das Geschäftsgeheimnis auch diesbezüglich nicht gewahrt. In der Praxis käme es immer wieder zum Streit darüber, was „wesentlich“ ist. Deswegen kann nur die Pflicht zur vollständigen Offenlegung glaubhafte Transparenz herstellen und Akzeptanz der Scores sicherstellen. Insofern liegt volle Transparenz im Interesse der Scoring-Anbieter.

Prof. Dr. Gerd Gigerenzer ist Direktor des Harding-Zentrums für Risikokompetenz am Max-Planck-Institut für Bildungsforschung (MPIB) in Berlin. Er war von 2013 bis 2018 Mitglied im Sachverständigenrat für Verbraucherfragen (SVRV).

Dr. Felix G. Rebitschek ist wissenschaftlicher Projektleiter am Harding-Zentrum für Risikokompetenz am MPIB in Berlin.

Prof. Dr. Gert G. Wagner ist Fellow am MPIB und Research Associate des Alexander von Humboldt Instituts für Internet und Gesellschaft (HIIG) in Berlin. Er ist Mitglied im SVRV.

Es ist anzumerken, dass Versicherungen, die Telematik-Tarife oder Bonusprogramme für Gesundheitsverhalten anbieten, kein Problem mit voller Transparenz der Merkmale und Gewichtung haben und diese auch dem Verbraucher gegenüber offenlegen.

Ein zweites Argument gegen die Forderung ist, dass dann Verbraucher ihr Verhalten ändern könnten, um ihren Score zu verbessern („gaming“). Dem steht entgegen, dass es sowohl bei Kreditwürdigkeit, Gesundheit als auch Fahrverhalten ja gerade wünschenswert ist, wenn Verbraucher ihr Verhalten zum Besseren ändern. Wenn Personen sich mehr bewegen, um ihren Gesundheitscore zu verbessern oder Rechnungen sofort oder spätestens nach der 1. Mahnung begleichen, ist das im Interesse des Verbrauchers und der Gesellschaft. Sind die Merkmale aber geheim, und haben die Menschen daher – wie die SVRV-Studie gezeigt hat – falsche Vermutungen, ist eine informierte Korrektur des Verhaltens nicht möglich.

Gaming wird nur genau dann ein Problem, wenn sich ein Anbieter im Wesentlichen anstatt auf ursächliche auf nicht-ursächliche Merkmale verlässt. Wenn beispielsweise der bloße Beitritt und nicht auch die aktive Nutzung eines Fitnessstudios den Gesundheitscore erhöht und Krankenversicherungsprämien so indirekt reduziert werden, ist das nicht ausreichend. 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.

Zusammenfassend sind wir der Auffassung, dass Verbraucher das Recht erhalten sollten, zu verstehen, wie ihr

Kasten 1

Algorithmen sind durchschaubar

Dass Algorithmen sehr nützlich sein können, wissen wir schon aus einer Zeit, in der es noch keine Computer und Big Data gab – aber transparent müssen sie sein (wie etwa Schulnoten oder die Punkte in Flensburg). Es gehört zum Kernbestand einer freiheitlichen Demokratie, dass es festgeschriebene Entscheidungsregeln gibt, die für die Betroffenen transparent sind und für die es einen Rechtsweg gibt, der es erlaubt ihre Einhaltung zu prüfen und gegebenenfalls durchzusetzen.

In der Öffentlichkeit greift seit einigen Jahren die Angst vor „selbstlernenden Algorithmen“ um sich, die die Menschheit über kurz oder lang entmachten würden. Zumindest, wenn man einer merkwürdigen Allianz von zwei Gruppen glaubt, die sonst nicht viel miteinander gemein haben: Jene Computerspezialisten, die so gerne von „Big Data“ reden, während sie die Ignoranz von Schöngeltern beklagen, und Feuilletonisten, die gerne betonen, nichts von Computern zu verstehen, aber umso mehr Ängste schüren. Beide Gruppen haben unrecht: Sowohl Entwickler der „künstlichen Intelligenz“ (KI) als auch ihre Kritiker übertreiben bei ihren Thesen, was diese leisten kann und wie undurchschaubar, ja autonom sie werden wird, wenn nicht schon sei.¹ Zunächst muss man festhalten, dass die Qualität der Verfahren, die mit dem Begriff künstliche Intelligenz zusammengefasst werden, nur in bestimmten – wenn auch wichtigen – Bereichen älteren Verfahren der statistischen Analyse überlegen ist: Bild- und Spracherkennung.² Bilder und Sprache sind kurzfristig wenig dynamische Systeme, die kaum auf ihre Umwelt reagieren, sondern statisch genug, dass man ein Computerprogramm, das Bild und Sprache erkennen soll, immer mehr verfeinern kann, wenn man es mit immer mehr Daten füttert. Die KI-Entwickler sprechen gerne von „selbstlernenden“ Systemen, was aber nichts anderes bedeutet, als dass von Menschenhand immer neue Daten hinzugefügt werden, und wie bei jeder statistischen Analyse werden die Ergebnisse besser, wenn ihr mehr Daten – also hier Fotos und Texte – zugrunde liegen. Von „Selbstlernen“ (wie bei Menschen) kann keine Rede sein, denn der Computercode verändert sich nicht von selbst. Alles, was sich ändert, sind die „Gewichte“, und beim Deep Learning gegebenenfalls auch automatisch die Merkmale der analysierten Fotos und Sprache selbst.

Wie jede statistische Analyse stößt KI schnell an ihre Grenzen, wenn sich ein System ständig verändert – insbesondere wenn auch sogenannte Rückkopplungen eine Rolle spielen. Im Gegensatz zur Einschätzung der Reaktion eines anderen Fahrers durch die KI eines selbstfahrenden Autos sind Rückkopplungen, die es im sozialen Alltag und im Wirtschaftsleben zu interpretieren gilt, ungleich schwieriger. Hier sind einfache und robuste statistische Verfahren gefragt, die auch nicht jeden Fehler in den Daten überinterpretieren. Deswegen wird modernste KI auch nicht zur Berechnung der Bonität eingesetzt – und es ist unwahrscheinlich, dass das in Zukunft der Fall sein wird. Denn KI ist hier einer konventionellen statistischen Analyse, bei der berechnet wird, von welchen Merkmalen das Risiko für einen Kredit-Ausfall abhängt, nicht überlegen.³

Was auch immer noch an digitalen Algorithmen kommen wird, die unser Leben steuern, etwa im Bereich der Schulen und Hochschulen im Zuge sogenannter „Learning Analytics“, das A und O für Akzeptanz ist Transparenz. Das gilt natürlich auch für Algorithmen, die auf KI beruhen. Wenn behauptet wird, sie seien undurchschaubar und dies gefährde unsere Demokratie, dann muss dem entgegengehalten werden, sie sind durchschaubar: man muss nur richtig hinschauen! Nun kann man – zu Recht – argumentieren, dass etliche der heutzutage eingesetzten Algorithmen, auch wenn sie keineswegs „selbstlernend“ sind, bereits geheimnisvoll genug seien: Sie sind nur schwer verständlich und machen manchem dadurch Angst. Das gilt beispielsweise auch für den Schufa-Score. Da das Kredit-Ausfallrisiko nicht mit einfachen Tabellen ausgerechnet wird (was durchaus möglich wäre), sondern in verfeinerter Form mittels „logistischer Regressionen“, sind die Berechnungen für Laien unverständlich und selbst für manche Fachleute nicht immer intuitiv zu verstehen: je nach Spezifikation der Regressionsgleichung sowie der Art und Weise der Ergebnisdarstellung (etwa ohne den Ausweis absoluter Risiken). Trotzdem ist es keineswegs so, dass wir Verfahren, die wie eine Art Black Box aussehen, hilflos ausgeliefert wären, weil sie unverständlich wären und immer unverständlicher werden könnten. Denn auch Laien können sich Beispielfälle anschauen, um zu sehen, was ein Computercode „voraussagt“. Genau dies machen auch konventionelle Statistiker und KI-Spezialisten, um die Plausibilität ihres Programmcodes und ihrer Berechnungen zu prüfen. Wenn man eine ganze Reihe von systematisch ausgewählten Beispielen ausrechnet, kann man gut abschätzen, was ein „Algo-

1 Vgl. auch G. G. Wagner: Transparenz – per Gesetz, in: Freitag vom 31.10.2018, <https://www.freitag.de/autoren/der-freitag/transparenz-per-gesetz> (23.11.2018); und K. Passig: Fünfzig Jahre Black Box, in: Merkur, 71. Jg. (2017), Nr. 823, S. 16-30.

2 Vgl. C. Wick: Deep Learning, in: Informatik-Spektrum, 40. Jg. (2017), H. 1, S. 103-107.

3 Vgl. Sachverständigenrat für Verbraucherfragen (SVRV): Gutachten „Verbrauchergerechtes Scoring“, Berlin 2018, http://www.svr-verbraucherfragen.de/wp-content/uploads/SVRV_Verbrauchergerechtes_Scoring.pdf (12.12.2018), S. 94 ff.

rithmus“ tatsächlich macht. Etwa, ob er bei der Beurteilung der Kreditwürdigkeit Bewohner in bestimmten Gegenden benachteiligt, weil sich dort Nachbarn mit hohen Kredit-Ausfallrisiken ballen.⁴

Die Zuverlässigkeit eines bestimmten Verfahrens anhand von Tests abzuschätzen ist auch die Grundlage für die Stiftung Warentest. Um die Qualität vieler Geräte und deren Software beurteilen zu können, reicht es nicht, die Einzelbestandteile und den Bauplan zu kennen (der ja obendrein oft auch ein Geschäftsgeheimnis ist) und einen „Algorithmen-TÜV“ einzuführen, bei dem wie beim Auto-TÜV darauf geachtet würde, dass nur „sichere“ Software-Bestandteile benutzt werden. Vielmehr muss getestet werden, wie ein Produkt sich im Alltag unter Belastung verhält. Das geht nicht nur mit Kaffeemühlen, sondern auch mit jedem Algorithmus. Dass es Algorithmen gibt, deren Wirkungen schwer für alle Eventualfälle zu testen sind, etwa die Computerprogramme, die Finanzprodukte an- und verkaufen, ist auch richtig. Aber hier ist nicht der Computer-Algorithmus das Problem, sondern das Finanzsystem. Es ist von vielen Rückkopplungen gekennzeichnet und nur schwer zu verstehen. Börsenkrach gab es schon, bevor es Computer gab.

Gegen die Sorge vor dem nicht gänzlich Bekannten hilft auch die Lektüre von Stanisław Lem. Er argumentierte in seiner „Summa Technologiae“, dass jeder Mensch ein hervorragendes Beispiel eines Apparates ist, dessen wir uns bedienen können, ohne seinen Algorithmus im Detail zu kennen.⁵ Als Bürger haben wir zu Recht ein Interesse daran, zu wissen, wie Algorithmen, die für unser Leben eine wesentliche Rolle spielen, funktionieren. Auch wie ein von einem Algorithmus gesteuertes selbstfahrendes Auto funktioniert, müssen wir wissen. Dieses Wissen zu erlangen, ist relativ einfach, der Gesetzgeber muss es nur wollen und durch Gesetze und Verordnungen sicherstellen, dass alle relevanten Algorithmen so getestet werden, wie das beim selbstfahrenden Auto bereits jetzt der Fall ist. Und die Ergebnisse dieser Tests müssen transparent gemacht werden. Das ist z. B. beim Schufa-Score recht einfach zu bewerkstelligen, da man ihn ganz leicht anhand von Beispielfällen berechnen kann. Der Gesetzgeber muss eine solche „Transparenz-Schnittstelle“ nur wollen.⁶ Sie einzurichten ist keine Hexerei, sondern bedarf eines überschaubaren (Forschungs-)Aufwands.⁷

Das Test-Szenario hat bisher keine besondere Aufmerksamkeit genossen.⁸ Denn derzeit muss kein Algorithmen-Entwickler es der kritischen Öffentlichkeit leichtmachen, seine Produkte zu testen. Das Befüllen der Black Box mit Testdaten kann deshalb ein mühsames Unterfangen sein. Neu gemischt würden die Karten, sobald ein Anspruch auf Durchführung von Tests in die Rechtsordnung Einzug hielte. Dann müssten die Algorithmen-Entwickler eine Schnittstelle vorsehen, über die Testdaten eingespeist werden können. Die Arbeit der Algorithmen-Tester würde dies ungemein vereinfachen – so sehr, dass es Fälle geben kann, in denen es keinen Unterschied mehr macht, ob der Algorithmus nun offengelegt oder „nur“ auf Herz und Nieren getestet wurde.

Ein juristisches Allheilmittel, um den Interessengegensatz zwischen Offenlegungs- und Geheimhaltungswunsch zu überwinden, ist der Algorithmen-Test nicht. Denn je intensiver man eine Black Box von außen testet, desto mehr gibt sie ihre inneren Geheimnisse preis. Zwar schützt das Grundgesetz Unternehmen davor, ihre Geschäftsgeheimnisse offenlegen zu müssen, letztlich gewinnt der Algorithmen-Tester jedoch ein klares Bild vom geheimen Bauplan des Algorithmus. Wie intensiv dadurch Geschäftsgeheimnisse in Mitleidenschaft gezogen werden, ist freilich eine offene Frage. Denn nicht nur der Score, den ein Algorithmus berechnet, ist z. B. für das Geschäftsmodell einer Auskunft wichtig, sondern die gesamte Online-Abwicklung und Organisation einer Anfrage gehören zu ihrem Geschäftskapital. Einen Algorithmus zu kennen, schafft nicht automatisch auch die Organisation, die für seine geschäftsmäßige Anwendung im Alltag notwendig ist. Überdies ist der Schutz des Geschäftsgeheimnisses nicht absolut. Wie weit er reicht, entscheidet sich erst durch eine Abwägung mit anderen in der Verfassung garantierten Rechtsgütern. Zu diesen gehören allemal das Persönlichkeitsrecht und die informationelle Selbstbestimmung derjenigen, die den sich undurchschaubar gebenden Algorithmen ausgesetzt sind. Die Väter und Mütter des Grundgesetzes haben hier Weitsicht bewiesen. Denn dessen Artikel 14 gewährleistet nicht nur das Eigentum. Im kürzesten Satz des gesamten Grundgesetzes heißt es dort mit maximaler Prägnanz: „Eigentum verpflichtet.“ Eine Ausnahme für das Eigentum in der digitalen Welt ist in der Verfassung nicht vorgesehen.

4 Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, Berlin 2018.

5 S. Lem: Summa Technologiae, (1964), Minneapolis und London 2013, S. 98.

6 G. Gigerenzer, K.-R. Müller, G. G. Wagner: Wie man Licht in die Black Box wirft, in: Frankfurter Allgemeine Zeitung vom 22.6.2018, Nr. 142, S. 15.

7 Gesellschaft für Informatik: Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, Berlin 2018, S. 155 ff.

8 Für das folgende danken wir Johannes Gerberding und Christian Gross für wichtige Hinweise.

Score zustande kommt. Das würde auch helfen, Fehler im Score schnell zu entdecken und zu beseitigen. Die Kfz-Versicherungen und Krankenkassen mit ihren Bonusprogrammen erfüllen diese Forderung bereits jetzt, während Auskunfteien dies nicht tun. Die vorgebrachten Gegenargumente halten wir für Scheinargumente. Die mangelnde Transparenz verhindert darüber hinaus, die Qualität der Scores beurteilen zu können.

Qualität von Scores und Daten gewährleisten

Scoring sollte nicht nur für den Verbraucher verständlich, sondern auch qualitativ hochwertig sein. Das heißt zweierlei: Scoring mit dem Ziel, Verhalten vorherzusagen, etwa die Bonität, sollte dies auch nachweislich leisten. Scoring, dessen Ziel es ist, Verhalten zu steuern, wie bei Bonusprogrammen der Krankenkassen, sollte Verhaltensänderungen wie einen gesünderen Lebensstil auch tatsächlich bewirken.

Scoring zur Verhaltensvorhersage: Bonität

Ein Bonitäts-Score soll vorhersagen, ob ein Kunde vertragsgemäß zahlen wird. Das kann nur gelingen, wenn die Qualität des Scoring-Algorithmus und der dazu verwendeten Daten gesichert ist. Das erste Problem liegt darin, dass zur Sicherung der Qualität des Algorithmus Paragraph 31 Bundesdatenschutzgesetz lediglich ein „wissenschaftlich anerkanntes mathematisch-statistisches Verfahren“ fordert. Dies ist eine unzureichende, da völlig unspezifische Forderung. Beim Bonitäts-Scoring werden anerkannte Verfahren wie eine logistische Regression eingesetzt. Die Methodik sagt jedoch nichts über die Qualität aus, da das Ergebnis, je nach Merkmalsauswahl und der Datenqualität, zu guten oder schlechten Prognosen führen kann.

Als zweites Problem kann ein solches „mathematisch-statistisches Verfahren“ bei der Vorhersage eine bestimmte Personengruppe ungerechtfertigt diskriminieren. Der Begriff „wissenschaftlich anerkanntes“ Verfahren ist nahezu inhaltsleer, vielmehr sollte jeder spezifische Algorithmus auf seine Qualität in der Vorhersage wissenschaftlich geprüft werden. Die Vorhersagequalität von Bonitäts-Scoring ist nicht perfekt, daher soll diese mittels üblicher Gütemaße (z. B. Falsch-Positiv-Rate, Hit-Rate, Gini-Koeffizient) quantitativ gekennzeichnet werden. Die bisherige Praxis ist undurchsichtig. Die Praxis, dass die Gutachten für die Prüfung eines geheimen Algorithmus durch die Datenschutzbeauftragten von den Auskunfteien beauftragt und bezahlt werden, sollte sofort unterbunden werden. Die Aufsicht muss in die Lage versetzt werden, selbst kompetent und unabhängig zu urteilen.

Forderung 2: Die Vorhersagequalität von Bonitäts-Scores soll von einer kompetenten und unabhängigen Stelle (wie einer „Digitalagentur“⁷) überprüft werden. Die Qualität soll quantitativ mittels üblicher Gütemaße für die Gesamtpopulation ausgewiesen werden, aber auch für relevante Teilpopulationen, um ungerechtfertigte Diskriminierung aufdecken zu können.

Die Qualität der Verhaltensvorhersage hängt kritisch von der Qualität der Daten ab. Die Datenqualität beim Bonitäts-Scoring ist weder perfekt (nicht überraschend, aber relevant) noch transparent. Ein Beispiel betrifft den Abgleich der Identitätsmerkmale zum Zweck der „Entity Recognition“ bei der Schufa. Einerseits berichtet die Schufa, dass sich nur wenige Menschen beschwerten, weil sie glauben, dass ihnen ein unzutreffender Score zugeordnet wurde.⁸ Andererseits dürfte nur einem Teil der gescorten Personen bewusst sein, dass er gescort wird, ein kleinerer Teil beantragt die Selbstauskunft, ein noch kleinerer Teil würde Personenverwechslungen bemerken und der kleinste Teil eröffnet ein Ombudsverfahren. Die Zahl der Schiedsverfahren sagt also wahrscheinlich wenig über das Ausmaß dieser Verwechslungen aus.

Daher müssen auch andere Zahlen betrachtet werden. Es ist bekannt, dass es bei Umzügen an einen neuen Wohnort zu Problemen bei der Entity Recognition kommt: Internetanschluss-Anträge vor Umzugsmeldungen beispielsweise führen dazu, dass ein neuer Datensatz für eine Person angelegt wird, für den es definitionsgemäß keine Bonitäts-Geschichte gibt und für die dann mithilfe der Vertragsanschrift (Geo-Scoring) und der Altersverifikation ein potenziell ungünstiger Score gebildet wird. Diese 0,3%⁹ aller 144,1 Mio. Score-Berechnungen pro Jahr¹⁰ entsprechen etwa 430 000 Geo-Score-Berechnungen, die mit der Erwartung aus einer jährlichen Umzugsquote

7 Vgl. diese Empfehlung des SVRV bereits 2017, Sachverständigenrat für Verbraucherfragen: Gutachten „Digitale Souveränität“, Berlin 2017, http://www.svr-verbraucherfragen.de/wp-content/uploads/Gutachten_Digitale_Souver%C3%A4nit%C3%A4t_.pdf (25.11.2018).

8 „Im Jahr 2017 haben 984 Verbraucher schriftlich einen Antrag auf Eröffnung des Schlichtungsverfahrens beim Ombudsmann der Schufa gestellt. Davon waren 336 im Sinne der Verfahrensordnung zulässig. [...] In 42 Fällen ging das Schiedsverfahren zugunsten der Verbraucher aus und dem Verbraucheranliegen wurde stattgegeben. Angesichts der von der Schufa in 2017 jeden Tag erteilten 400 000 Auskünfte ist die Zahl der berechtigten Verbraucheranliegen als gering zu werten“, vgl. H.-J. Papier: Tätigkeitsbericht 2017, 2018, <https://schufa-ombudsmann.de/verfahren/taetigkeitsberichte/taetigkeitsberichte.html> (18.11.2018).

9 Schufa: Wie funktioniert Scoring bei der Schufa?, 2018, <https://www.schufa.de/de/ueber-uns/daten-scoring/scoring/scoring-schufa/> (18.11.2018).

10 Schufa: Schufa in Zahlen, 2018, <https://www.schufa.de/de/ueber-uns/unternehmen/schufa-zahlen/> (18.11.2018).

von etwa 9 % der Bevölkerung¹¹ konsistent und kaum als Ausnahmefälle zu deuten sind.

Dem SVRV ist bewusst, dass eine Verbesserung der Entity Recognition nicht einfach ist, wenn es aus Datenschutzgründen kein zentrales Melderegister geben soll, das für Firmen, die Bonitäts-Scoring anbieten, zugänglich wäre oder jeder in Deutschland lebenden Person eine eindeutige Personen-Nummer zuweist. Deswegen empfiehlt der SVRV, dass die Datenethikkommission der Bundesregierung sich des Problems einer verbesserten Entity Recognition annehmen sollte.

Scoring zur Verhaltenssteuerung: Gesundheit und Fahrverhalten

Dem gesetzlichen Auftrag folgend, sollen die Bonusprogramme der Krankenkassen gesundheitsbewusstes Verhalten fördern. Also würde man erwarten, dass Verhalten, das wissenschaftlich nachgewiesen die Gesundheit stärkt, mit entsprechend mehr Punkten belohnt wird. In einer Marktstudie¹² vom SVRV wurde jedoch festgestellt, dass viele gesetzliche und private Krankenkassen wenige oder keine Punkte für nachweislich gesundheitsförderndes Verhalten vergeben, aber vergleichsweise viele Punkte für ein Verhalten, für dessen Nutzen in der medizinischen Forschung kaum Evidenz vorliegt. Ein Beispiel: Starke Raucher verringern ihre Lebenserwartung um bis zu zehn Jahre und ihre Lebensqualität entsprechend.¹³ An Untersuchungen zur Krebsfrüherkennung teilzunehmen, erhöht dagegen die Lebenserwartung nicht oder bestenfalls einige Tage,¹⁴ und sogenannte individuelle Gesundheitsleistungen (IGeL) wie Glaukom-Früherkennung haben keinerlei erwiesenen Nutzen.¹⁵ 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 Ultraschalls.¹⁶ Trotz dieser wissenschaftlichen Ergebnisse vergeben nur acht der 45 untersuchten Bonusprogramme Punkte für Nicht-Rauchen, aber alle 45 Programme geben Punkte für Krebsfrüherkennung.

Beispielsweise ist die Versicherung Generali Vitality eine der wenigen, die Punkte für den Nicht-Raucher-Status vergibt (eine Selbsterklärung, ein Kreuzchen reicht). Sie gibt aber zugleich Punkte für Verhalten wie die Teilnahme an Prostatakrebsfrüherkennung mittels PSA-Test, dessen Nutzen umstritten, aber dessen Schaden (wie Inkontinenz und Impotenz im Fall unnötiger Behandlung nicht-fortschreitender Krebsformen) nachgewiesen ist.¹⁷ Generali bietet das Bonusprogramm auch nur in Verbindung mit dem Kauf einer Berufsunfähigkeits-Versicherung oder Risikolebensversicherung an und hat angekündigt, mit den privaten Krankenkassen in Deutschland zusammenarbeiten zu wollen.¹⁸ Aus der Marktanalyse und den Interviews mit den Anbietern wird deutlich,¹⁹ dass es bei den Bonusprogrammen nicht immer in erster Linie um die Förderung der Gesundheit geht, sondern um Kundenbindung, Neukundenwerbung und den Verkauf ansonsten schwer verkäuflicher Versicherungen, wie für Berufsunfähigkeit. In diesen Fällen wird das angegebene Ziel der Gesundheitsförderung weitgehend durch Geschäftsinteressen der Versicherungen ersetzt.

Beim Telematik-Scoring in Kfz-Versicherungstarifen ist das Ziel, vermeidbare Unfälle durch ein verbessertes Fahrverhalten zu verringern. Untersuchungen, die zeigen, dass diese Tarife Unfallrisiken tatsächlich reduzieren und sich auf beide Seiten, Versicherer und Versicherter, positiv auswirken, liegen bislang nicht vor. Auch hier lohnt es sich, genauer auf die Merkmale zu sehen, die in den Score eingehen. Merkmale, wie zu schnelle Beschleunigung und Geschwindigkeitsübertretungen, sind unter Kontrolle des Fahrers, andere wie Stadtfahrten versus Landfahrten sind es nicht oder kaum, d.h., ein Score kann erstere beeinflussen, aber letztere kaum. Dies kann zu einer direkten Diskriminierung von Menschen führen, die aus beruflichen Gründen meist in der Stadt oder nachts fahren.²⁰

Forderung 3: Die nicht-adäquate Kennzeichnung des Ziels eines Scoring-Verfahrens ist aufzudecken und zu verhinder

11 Bundesbaublatt: Wer hat, der bleibt: Laut Techem ziehen immer weniger Mieter um, 2018, http://www.bundesbaublatt.de/news/umzugsquote-wer-hat-der-bleibt_3187016.html (18.11.2018).

12 Vgl. SVRV, a. a. O., S. 71 ff.

13 Deutsches Krebsforschungszentrum: Gesundheitliche Folgen des Rauchens, 2010, https://www.dkfz.de/de/tabakkontrolle/Gesundheitliche_Folgen_des_Rauchens.html (25.11.2018).

14 V. Prasad, J. Lenzer, D. H. Newman: Why cancer screening has never been shown to „save lives“ – and what we can do about it, in: *British Medical Journal*, 352. Jg. (2016), H. h6080; G. Gigerenzer: Full disclosure about cancer screening, in: *British Medical Journal*, 352. Jg. (2016), H. h6967.

15 IGeL-Monitor: Messung des Augeninnendrucks zur Früherkennung eines Glaukoms – Ergebnisbericht, 2012, https://www.igel-monitor.de/fileadmin/user_upload/Messung_des_Augeninnendrucks_Ergebnisbericht.pdf (25.11.2018).

16 G. Gigerenzer, K. Kolpatzik: How new fact boxes are explaining medical risk to millions, in: *British Medical Journal*, 357. Jg. (2017), H. j2460.

17 D. Ilic et al.: Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis, in: *British Medical Journal*, 362. Jg. (2018), H. k3519.

18 Wenn man regelmäßig im Online-Supermarkt „all you need is fresh“ einkauft, kann man bei Vitality mehr Gesundheitspunkte sammeln als beim Nicht-Raucher-Status.

19 Vgl. SVRV, a. a. O., S. 76 ff.

20 Tarife wie BetterDrive oder SecureDrive basieren auf Merkmalen, die viele Fahrer kaum beeinflussen können: Tageszeit (9/10 Anbietern), Straßen-Typ (7/10) und Ort (5/10).

dem. Zur Aufdeckung falsch gekennzeichnete Scores können neben Aufsichtsbehörden auch Verbraucherorganisationen oder die Marktwächter der Verbraucherzentralen beitragen. Irreführende Kennzeichnung ist bereits nach bestehender Rechtslage unzulässig, erhält aber durch Bonusprogramme neue Möglichkeiten.

Aufsicht durch eine Digitalagentur stärken

Die Gewährleistung von Transparenz, Qualität und der Nicht-Diskriminierung sollten zu den Aufgaben von Aufsichtsbehörden und Verbraucherschutzorganisationen zählen. Dazu sind diese jedoch mit ihrer derzeitigen personellen und technischen Ausstattung nicht in der Lage. Aktuell legen z.B. Auskunfteien, wie die Schufa, ihre Verfahren zur Berechnung der Bonität den Datenschutzbehörden vor. Auf der Webseite der Schufa steht: „Das Schufa-Verfahren zur Score-Berechnung ist bereits für Behörden und Aufsichten transparent.“ Für die Schufa ist das der Hessische Datenschutzbeauftragte. Ein Sprecher der Datenschutzbehörde Nordrhein-Westfalen erklärte dagegen, dass den Behörden die erforderliche Fachkunde fehlt, um die Algorithmen wirklich zu bewerten: „Wir können die Gutachten damit nur auf Plausibilität prüfen.“ Nach Recherchen des Bayerischen Rundfunks 2018 verlassen sich die Behörden daher auf Gutachten, die von den Auskunfteien selbst in Auftrag gegeben und bezahlt werden.²¹ Diese Praxis erinnert an jene der Finanzinstitute vor der letzten Finanzkrise, die selbst eine Rating-Agentur auswählten und bezahlten, damit diese ihre Finanzprodukte bewertet. Dadurch wurden inflationär AAA-Bestnoten vergeben, ohne dass diese angemessen gewesen wären, was einer der Hauptanlässe der Krise war.

Die derzeitige geteilte Aufsicht der Bonusprogramme der Krankenkassen zwischen dem Bundesversicherungsamt (für die bundesunmittelbaren gesetzlichen Krankenkassen, die sich über mehr als drei Bundesländer erstrecken) und den Landesbehörden (für Kassen, deren Tätigkeit sich auf maximal drei Bundesländer beschränkt, z. B. die Allgemeinen Ortskrankenkassen) gereicht keiner Seite zum Vorteil. Das Bundesversicherungsamt muss gemäß seinem Auftrag standardisierte Programmevaluationen auf Plausibilität prüfen, wobei die sachgemäße Verwendung der Versichertenbeiträge im Vordergrund steht. Eine methodische oder gesundheitsbezogene Überprüfung der Regeln für bonifizierbare Aktivitäten und ihrer Umsetzung auf ihre Auswirkungen hin findet nicht statt. Dass die Landesbehörden den regionalen Kassen ungleich mehr

²¹ Dies erinnert an die Praxis der Rating-Agenturen wie Standard & Poor's, Moody's und Fitch, deren Scores einer der Hauptanlässe der Krise waren. Diese werden von den Instituten bezahlt, deren Produkte sie bewerten.

Spielraum in der Gestaltung der Bonussysteme einräumen, ermöglicht nur noch mehr Merkmalskombinationen in ihren Programmen, für die es keine Gesundheitsbelege gibt. Zusammengefasst dient diese Wettbewerbslogik zwischen den Versicherern bisher nicht der Gesundheit der teilnehmenden Versicherten.

Forderung 4: Einrichtung einer Digitalagentur, die als Kompetenzzentrum den einzelnen Aufsichtsbehörden hilft, ihre Aufgaben wahrzunehmen. Die Digitalagentur und die Aufsichtsbehörden sind finanziell so auszustatten, dass sie erfolgreich interdisziplinäre Teams von Spezialisten aus Statistik und Informatik zusammenstellen können.

Die vorgeschlagene Digitalagentur, die etwa in Form eines „Bundesinstituts“ eingerichtet werden könnte, bündelt Kompetenzen, um die Arbeit der einzelnen Aufsichtsbehörden zu unterstützen. Interdisziplinäre Teams sind unabdingbar, um Transparenz und Qualität kompetent zu prüfen und auf dieser Basis ungerechtfertigte Diskriminierung von einzelnen Personengruppen zu entdecken. Hierbei ist ein besonderes Augenmerk auf Personal mit Methodenkenntnissen zu legen, die auch mögliche zukünftige technologische Innovationen beim Scoring, wie etwa Big-Data-Technologien, verstehen und kritisch bewerten können, die über die heute ausschließlich im Einsatz befindlichen logistischen Regressionen hinausgehen.²² Diese Forderung bedeutet aber nicht einfach mehr Personal, sondern mehr und besser bezahlte Spezialisten aus Informationstechnik und Statistik.

Vorbild für eine verbesserte Aufsichtspraxis könnte die Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) sein. Eine rein datenschutzrechtliche und oberflächliche Methodenprüfung, wie sie der Gesetzgeber bislang vorschreibt, reicht nicht aus. Vielmehr sind materielle Prüfungen – analog der BaFin-Bankenaufsicht – hinsichtlich Transparenz, Qualität und ungerechtfertigter Diskriminierung notwendig. Die Telematiktarife der Kfz-Versicherer und die privaten Krankenversicherungen unterliegen bereits der Prüfung durch die BaFin.

Super-Scores verhindern

Die Zusammenführung aller Scores eines Menschen ist derzeit ein methodisches Problem, dessen technische Lösung in naher Zukunft möglich sein wird. Diese beruht auf einer De-Anonymisierung von Daten und deren Verknüpfung zu einem Super-Score.

²² N. Jentzsch: Kreditwürdigkeitsanalysen im Zeitalter von Big Data: Innovation oder Revolution?, in: Wirtschaftsdienst, 96. Jg. (2016), H. 9, S. 644-647, <https://archiv.wirtschaftsdienst.eu/jahr/2016/9/neue-finanztechnologien-bankenmarkt-in-bewegung/>.

Die chinesische Regierung arbeitet derzeit gemeinsam mit Sesame Credit (Alibaba), Wechat (Tencent) und anderen Konzernen an einem Sozialkredit-System, das bis 2020 jedem Bürger zugeordnet werden soll. Der entsprechende Wert umfasst nicht nur die Bonität, sondern Daten aus allen Lebensbereichen, einschließlich dem sozialen und politischen Verhalten. In Kombination mit einem landesweiten Videoüberwachungssystem mit Gesichtserkennungssoftware („Himmelsnetz“) soll das moralische Verhalten der Bürger erhöht, Korruption verringert und „Harmonie“ erzeugt werden.²³ „Aufrichtige Bürger“ mit hohem Super-Score erhalten Vergünstigungen (ähnlich einer Vielflieger-Karte). „Unaufrichtige Bürger“ mit niedrigem Score wird Freiheit entzogen, diese dürfen etwa nicht mehr mit Schnellzügen fahren oder ihre Kinder nicht auf bestimmte Schulen gehen. Der Super-Score ermöglicht eine ständige und totale Kontrolle des Verhaltens der Bürger.

In Gesprächen des SVRV mit Experten zum Sozialkredit-System zeigte sich auch, dass die Entwicklung in China für Deutschland, wenn auch mit Verzögerung, direkt relevant werden könnte. Derzeit ist das System in China in einem experimentellen Stadium, an dem etwa 40 Regionen teilnehmen, die ihre eigenen Scoring-Systeme entwickeln. Die Kommunistische Partei wird das effektivste davon auswählen, landesweit umsetzen, um das moralische Verhalten – wie Harmonie und „Aufrichtigkeit“ – der Gesellschaft verbessern: was zu einer weiteren Stärkung der Volkswirtschaft führen soll.

Umfragen in China zeigen, dass etwa 80 % der Bürger den Super-Score für gut halten und dass die Fürsprache umso größer ist, je höher das Bildungsniveau ist.²⁴ Ein solches System könnte dann zusammen mit Software und Hardware an andere autokratisch geführte Staaten verkauft werden – Kandidaten wären Thailand, Singapur, aber auch Ungarn und die Türkei – die dann mit Hilfe dieser digitalen Technologie ein effektives System der Überwachung und Belohnung ihrer Bürger einrichten können, von dem die Stasi nur hätte träumen können. Ein solches System

kann schneller Entscheidungsgrundlagen schaffen als unsere oft langsamen Wege der Demokratie. Inwieweit die westlichen Demokratien dabei unter Druck geraten und sich dieser neuen Form einer digitalen Autokratie entgegenstellen könnten, ist zumindest eine Frage wert.

In Deutschland besteht derzeit keinerlei Gefahr einer ähnlichen totalen Überwachung durch den Staat. Doch ist zu befürchten, dass internationale Konzerne Super-Scores ohne staatliche Vorgabe aus kommerziellen Gründen entwickeln und anbieten. Etliche Firmen weltweit haben sich darauf spezialisiert, Daten zu sammeln und zu verkaufen, meist ohne Wissen oder Zustimmung der Verbraucher. Auch in Deutschland sammeln Firmen wie Arvato (Bertelsmann), Otto (Tochter: Eos Gruppe) und Deutsche Post Daten.²⁵ Einer der größten internationalen Datenfirmen ist die US-Firma Acxiom, die auch in Deutschland tätig ist und nach eigenen Angaben bereits im Jahr 2013 Daten über 44 Mio. Deutsche anbot.²⁶ Und viele Menschen sind bereit, ihre Daten gegen kleine Vorteile und Bequemlichkeiten weiterzugeben.

Die kommerzielle Überwachungsgesellschaft ist nicht nur Möglichkeit, sondern schon Realität geworden. Sobald das technische Problem der De-Anonymisierung und der Zusammenführung aller persönlichen Daten zu einem Wert gelöst ist, wird Druck entstehen, diesen auch in Deutschland anbieten zu können. In Europa ist die De-Anonymisierung von Menschen verboten – aber dieses Verbot verliert seine Wirkung, wenn Menschen freiwillig einer Datenverknüpfung zustimmen. Eine derartige Möglichkeit der Zusammenführung von Scores beginnt bei Unternehmen, bei denen derselbe Kunde mehrere Versicherungen, etwa Kranken- und Kfz-Versicherung, abgeschlossen hat. Die repräsentative Bevölkerungsbefragung für das SVRV-Gutachten zeigt,²⁷ dass sich ein Drittel der erwachsenen Bevölkerung in Deutschland die Zusammenführung von Kfz- und Gesundheits-Scores vorstellen kann: Die Befragten gaben an, dies wäre gerecht oder sie könnten davon persönlich profitieren. Damit besteht 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 günstigere Versicherungstarife, sondern allerlei Vergünstigungen, beispielsweise beim Mieten einer Wohnung, oder Bequemlichkeiten, etwa bei Tischreservierung

23 Vgl. SVRV, a. a. O., S. 61 ff.; R. 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/> (2.3.2018); R. Creemers: Opinions concerning accelerating the construction of credit supervision, warning and punishment mechanisms for persons subject to enforcement for trust-breaking (inoffizielle Übersetzung der Direktive der Zentralen Volksregierung), 2016, <https://chinacopyrightandmedia.wordpress.com/2016/09/25/opinionsconcerning-accelerating-the-construction-of-credit-supervisionwarning-and-punishment-mechanisms-for-persons-subject-toenforcement-for-trust-breaking> (2.3.2018).

24 Vgl. G. Kostka: China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval, 23.7.2018, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3215138 (25.11.2018).

25 A. Gruber: Woher all die vielen E-Mails kommen, Spiegel online vom 1.7.2016, <http://www.spiegel.de/netzwelt/web/adressshaendler-was-die-branche-ueber-sie-weiss-wie-sie-sich-wehren-a-1098938.html> (25.11.2018).

26 C. McLaughlin: Die Besserwisser, in: Zeit vom 4.7.2013, <https://www.zeit.de/2013/28/acxiom> (25.11.2018).

27 Vgl. SVRV, a. a. O., S. 93 ff.

gen, Zugfahrkarten und Flugtickets, zu bekommen.²⁸ Auf die Frage nach einem Sozialkredit-System in Deutschland sind immerhin fast 10 % dafür, so ein soziales Punktesystem einzuführen.²⁹

Forderung 5: Die De-Anonymisierung und Zusammenführung einzelner Scores zu einem Super-Score durch kommerzielle Konzerne soll frühzeitig beobachtet und verhindert werden. Dazu sollten Gesetzgeber wie Aufsichtsbehörden konkret prüfen, ob bestehende Instrumente der Datenschutz-Grundverordnung (wie Zweckbindungsgrundsatz und Kopplungsverbot) einsetzbar sind oder andere Rechtsgrundlagen geschaffen werden müssen.

Ausblick

Wir haben einen kurzen Einblick in wesentliche Punkte des Gutachtens „Verbrauchergerechtes Scoring“ des Sachverständigenrats für Verbraucherfragen gegeben. Weitere Themen und Empfehlungen, auf die wir hier nur am Rande eingegangen sind, betreffen notwendiges Wissen und Kompetenzen zu Scoring, ungerechtfertigte Diskriminierung und sogenannte telematikfreie Optionen.³⁰

Wie die repräsentative Studie des SVRV zeigt, ist das Wissen der Bevölkerung darüber, ob sie gescort wird, und wenn ja, aufgrund welcher Merkmale, sehr gering bis nicht vorhanden.³¹ Die informierte Teilhabe an einem Scoring, dem man sich kaum entziehen kann, sollte sichergestellt werden. Diskriminierung ist ungerechtfertigt, wenn diese real existierende Unterschiede zwischen Personengruppen übertreibt. Damit Diskriminierung erkannt wird, müssen Scoring-Algorithmen transparent sein und die Aufsicht muss verbessert werden.

28 Auch bei uns wächst die Kultur des sozialen Scorings: ob Likes auf Facebook, Anzahl der Follower auf Twitter oder Sterne auf Airbnb.

29 Vgl. SVRV, a. a. O., S. 270.

30 Vgl. SVRV, a. a. O., S. 140 ff.

31 Vgl. SVRV, a. a. O., S. 94 ff.

Schließlich befürworten viele Menschen – wie die repräsentative Erhebung für den SVRV gezeigt hat³² – eine Option von Scoring im Bereich der Telematik und des Gesundheitsverhaltens. Doch sollte sichergestellt werden, dass anderen Verbrauchern, die das nicht möchten, auch in Zukunft telematikfreie Optionen offenstehen, ohne dass diese dadurch Nachteile erfahren.

Wir möchten abschließend ausdrücklich darauf hinweisen, dass transparentes und qualitätsgesichertes Scoring, das tatsächlich zu als positiv bewerteten Verhaltensänderungen führt, grundsätzlich sinnvoll ist. Jedoch gibt es auch hier eine Grenze, die durch den Super-Score markiert ist.

„Digital first – Bedenken second“, so lautete der Slogan, mit dem die FDP bei der Bundestagswahl 2017 um Wählerstimmen warb.³³ Politik und Industrie setzen auf die flächendeckende Digitalisierung, vom Smart TV zum Smart Home, von der Schule³⁴ bis zur Telemedizin. Neue Bereiche des Scoring, denen sich die Betroffenen faktisch nicht entziehen können, zeichnen sich konkret in Form von „People Analytics“ (für Einstellungs- und Beförderungentscheidungen) und künftig in Form einer elektronischen Patientenakte im Gesundheitsbereich ab.

Die Industrie setzt auf Scoring von immer mehr Lebensbereichen, aber es gibt wenig Wissen, wenig Transparenz und so gut wie keine Qualitätskontrolle. Die Antwort darauf lautet Transparenz, Qualitätskontrolle, kompetente Aufsichten und kein Super-Score.³⁵

32 Vgl. SVRV, a. a. O., S. 97 ff.

33 G. Gigerenzer et al., a. a. O.

34 O.V.: Weg frei für den Digitalpakt Schule, tagesschau.de, 2018, <https://www.tagesschau.de/inland/schulen-digitalisierung-grundgesetz-101.html> (25.11.2018).

35 Vgl. inzwischen auch: Deutsche Wissenschaftsakademien: Privatheit in Zeiten der Digitalisierung, Stellungnahme, 2018, https://www.leopoldina.org/uploads/tx_leopublication/2018_Stellungnahme_BigData.pdf (12.12.2018).

Title: A Scored Society Needs Transparency

Abstract: Scores are numerical ratings used to predict or steer people's behaviour. Traditionally, people have not only received scores on their educational achievements (grades) but also on their creditworthiness. The current spread of digital technologies is broadening the possibilities for assessing and rating individuals. This entails both novel risks and opportunities for scored individuals. These opportunities, however, can only be fully realized when society's various legitimate expectations are met. It is then that consumers will accept and benefit from scoring systems. These expectations will include simple and effective ways for individuals to appeal their score. Protected characteristics (such as gender) may not be used as a basis for unwarranted discrimination, be it directly or indirectly. When scores are calculated for the purpose of making predictions, the quality of the criteria applied and the reliability of the predictions must be demonstrated. Predictions need to adhere to the objective of the scoring system concerned, and scores should not be applied to different objectives. Above all, scoring systems must be comprehensible to those who are scored.

JEL Classification: G18, I10, I20, K24, K38, Z18