LEGAL REVOLUTIONary
Rechtsmagazin der digitalen Wirtschaft

Trau, schau, wem? Qualitätssicherung durch Datenverwender

Beitragsreihe „Input Control – Datenqualität und Datenvalidität als Grundlage rechtlicher Automatisierungsprozesse“, Abschnitt „Input“

PD Dr. Martin Fries | Privatdozent | Juristische Fakultät, Ludwig-Maximilians-Universität München

6. Mai 2020

LR 2020, Seiten 87 bis 89 (insgesamt 3 Seiten)

Seit Anfang Februar 2020 haben sich über 3 Millionen Menschen bei YouTube angesehen, wie der Berliner Künstler *Simon Weckert* den Online-Dienst *Google Maps* zum Narren hielt, indem er mit einem kleinen Schubwagen voller Smartphones durch Berlin spazierte. Obwohl die Straße menschenleer war, zeigte *Google Maps* einen Stau an. Im Nachhinein stellte sich heraus, dass es tatsächlich sehr schwierig war, den *Google*-Algorithmus zu überlisten. Das Experiment zeigt aber: Fehler sind möglich. Bei datenbasierten Services gilt das Prinzip des *garbage in, garbage out*: Mit der Qualität der verarbeiteten Daten steht und fällt die Qualität jedes Datenverarbeitungsvorgangs. Wer aber wacht über die Datenqualität? Neben dem Datenprovider[1] kommt hier insbesondere diejenige Akteurin in Betracht, die die Daten verwendet. 1

I. Folgen der Verarbeitung unrichtiger Daten

Wer gewerblich Daten verwendet, hat sie manchmal selbst erhoben, regelmäßig aber auch auf Datenmärkten beschafft oder offene Datenquellen angezapft. Kreditinstitute machen gute Konditionen von eingekauften Bonitätsdaten abhängig, Maschinenverleiher kalkulieren ihre Preise auf der Basis zugelieferter Wetterdaten, und Mobilitätsdienstleister richten sich nach offen zugänglichen Bewegungs- und Pünktlichkeitsdaten anderer Anbieter sowie nach Verkehrsstromdaten wie den eingangs genannten Staumessungen, um Zeitpunkt und Bepreisung ihrer Dienstleistung zu steuern. 2

[1] Vgl. *Rossi*, Informationsgewährungspflichten und -möglichkeiten des Staates –Niveau und Sicherung der Qualität staatlicher Daten, LR 2020, 90.

Soweit die verwendeten Daten aus offen zugänglichen Quellen stammen, gibt es niemanden, der Datenverwendern vertraglich für die Datenqualität haftet. In diesen Fällen hängt alles an der Auswahl der Datenquelle durch die Datenverwenderin. Nachlässigkeiten gehen hier regelmäßig zu Lasten ihrer Kunden. Ein Beispiel, das *Thomas Riehm* mir zugerufen hat: Mehrere Kfz-Versicherungen nutzen inzwischen Telematiktarife, die das individuelle Fahrverhalten des Versicherungsnehmers kontrollieren und abhängig davon einen höheren oder niedrigeren Versicherungsbeitrag berechnen. Verlässt sich eine solche Versicherung leichtfertig auf die Richtigkeit allgemein zugänglicher Datenquellen über Geschwindigkeitsbeschränkungen oder die Verkehrsdichte, unterstellt sie ihren Versicherungsnehmern womöglich unberechtigt Fahrfehler und kalkuliert die Versicherungsprämie unrichtig zu deren Nachteil. Dass sie damit eine vertragliche Nebenpflicht zur Rücksichtnahme auf die Interessen des Versicherungsnehmers verletzt, liegt auf der Hand. 3

Das Problem stellt sich in ähnlicher Weise, wo die Datenverwenderin nicht offen zugängliche Daten von einer Lieferantin bezieht. Denn auch wenn sich Letztere vertraglich zur Lieferung „sauberer" Daten verpflichtet, kann es natürlich Fälle geben, in denen ihre Kontrolle versagt. Dann hängt es von den Kontrollmechanismen im Hause der Datenverwenderin ab, ob der Fehler auffliegt. Bleibt er im Verborgenen, erhalten die Kunden wiederum eine mangelhafte Dienstleistung mit unter Umständen manifesten finanziellen Nachteilen. Gerade bei personenbezogenen Daten[2] können aber auch diejenigen, auf die sich die Daten beziehen, von Datenfehlern betroffen sein, etwa wenn sie von der Datenverwenderin oder einem späteren Glied der Datenlieferkette schlechtere Konditionen angeboten bekommen, weil ihre Bonität auf unrichtiger Datengrundlage berechnet wurde. 4

II. Prüfpflichten, Qualitätsstandards oder Gefährdungshaftung?

Um solche Probleme zu vermeiden, kann man Datenverwender in unterschiedlicher Weise in die Pflicht nehmen. Der Gesetzgeber kann beispielsweise Qualitätsprüfpflichten statuieren, wonach Datenverwender zumindest in Stichproben die Plausibilität oder sogar Richtigkeit von Daten kontrollieren müssen. Für den Bereich der personenbezogenen Daten enthält etwa Art. 5 Abs. 1 lit. d) DSGVO eine solche Pflicht. Alternativ sind auch Vorschriften mit Vorgaben zum Verfahren der Datenverarbeitung denkbar, wie sie etwa § 31 BDSG für das Scoring und Bonitätsauskünfte enthält. Je größer der Umfang der verwendeten Daten, desto eher geht die Entwicklung dabei in Richtung automatisierter oder zumindest teilautomatisierter Lösungen. Der europäische Gesetzgeber sieht dies skeptisch im Bereich personenbezogener Daten, vgl. Art. 22 DSGVO, mit Blick auf die Wahrung von Urheberrechten ist ihm dieser Ansatz in Form von Upload-Filtern nach Art. 17 der Urheberrechts-Richtlinie (EU) 2019/790 hingegen durchaus recht. Die im Entstehungsprozess dieser Richtlinie und bis heute geführte Diskussion zeigt freilich auch 5

[2] Vgl. *Hennemann,* Datenrichtigkeit, LR 2020, 77.

die Schwächen einer automatisierten Datenkontrolle: Nicht nur verarbeitete Daten, sondern auch datenkontrollierende Algorithmen können fehlerhaft sein.

Immerhin gibt es Alternativen zu dieser Art von Prüfpflichten. Eine Möglichkeit besteht etwa darin, Datenformate zu zertifizieren oder zu standardisieren (vgl. den Qualitätsstandard ISO 8000), um Transparenz zu fördern und – verpflichtende oder freiwillige – Qualitätskontrollen zu erleichtern. Ein anderer Ansatz wäre die Schaffung einer Gefährdungshaftung zu Lasten von Datenverwendern, wonach diese für Schäden einzustehen hätten, die durch den Einsatz unrichtiger Daten entstehen. Die Frage ist freilich, ob man mit einer scharfen Haftung bei den Datenverwendern an der richtigen Adresse wäre. Aus rechtsökonomischer Warte spricht vieles dafür, eine solche Haftung eher bei derjenigen Akteurin anzusiedeln, die die Daten erhebt, weil sie die Datenqualität am besten steuern kann und damit regelmäßig der *superior risk bearer* sein wird. Zu klären bleibt dann vor allem noch die Frage der Beweislast dafür, dass ein Schaden tatsächlich durch die Bereitstellung oder Verarbeitung unrichtiger Daten verursacht wurde, denn wer sie trägt, wird regelmäßig auch den Schaden zu schultern haben. 6

III. Qualitätssicherung bei Datenverarbeitung durch künstliche Intelligenz

Besonders spannend, aber auch besonders schwierig wird es in Zukunft dort, wo die ursprünglich in die Verarbeitung eingespeisten Rohdaten nicht mehr zu erkennen sind. Wenn nur noch die lernende Maschine weiß, aber selbst sie nicht erklären kann, welche Bedeutung unrichtige Daten für das Ergebnis ihrer Berechnungen haben, wird eine Beweisführung über Details der Datenverarbeitung weitgehend unmöglich. Inwieweit man dieses Klärungsdefizit mit einer statistischen Beschreibung der Datenverarbeitung auflösen kann und möchte, erscheint gegenwärtig noch ungewiss. 7

Gewiss ist demgegenüber, dass die Zahl der datenbasierten Güter- und Dienstleistungsangebote und der datengetriebenen Entscheidungen in naher Zukunft deutlich zunehmen wird. Dem liegt nicht zuletzt ein unternehmerisches Streben nach präziser und möglichst individueller Preisbildung zur Maximierung der Produzentenrente zugrunde, das in Konkurrenz tritt zur einheitlichen Preisbildung auf breit angelegten Märkten. Gleichzeitig gilt es ein Rezept zum Umgang mit manipulierten Daten zu finden, denn mit dem steigenden wirtschaftlichen Gewicht von Daten werden sich auch die Anreize für deren bewusste Verzerrung mehren. Der künstliche Stau in Berlin gab auf seine Weise einen wichtigen Fingerzeig, wie man die Ergebnisse einer Datenverarbeitung beeinflussen kann. Die rechtswissenschaftliche Forschung steht hier noch ganz am Anfang. 8