Kai v. Lewinski, Johanna Hähnle*

Was informatisch richtig ist, kann auch juristisch recht sein

Datenvalidität und -qualität als juristische Kategorien

Die „Richtigkeit“ von Daten und Informationen ist für Prozesse von grundlegender Bedeutung, im Computer wie im Recht. Ob und wie die informatischen Konzepte und Kategorien von Validität und Qualität in das Recht übertragen werden, ist Gegenstand des nachstehenden Beitrags.

1 Was ist rechterweise richtig?

Informationen und Daten beeinflussen menschliches Verhalten in nahezu allen Lebensbereichen, und mit fortschreitender Automatisierung ist immer mehr Vertrauen in die bereitgestellten Informationen erforderlich.¹ Manchmal nun wird dieses Vertrauen enttäuscht, wenn Daten und Informationen falsch sind oder jedenfalls zu Fehlern führen. Zwar wirkt sich nicht jede falsche Angabe (Garbage in, Garbage out) aus, umgekehrt können aber auch richtige Daten „falsch“ sein, etwa wenn sie unvollständig, außerhalb eines bestimmten Zusammenhangs² oder (fälschlicherweise) mit einer alternativen Information verknüpft³ verwendet werden.

Was hier die Rechtswissenschaft von der Informatik lernen kann, will der nachstehende Beitrag beleuchten. Zunächst werden

* Besonderer Dank geht an Dr. Armin Gerl (Universität Passau), der stets für Rückfragen zu den informatischen Bezügen zur Verfügung stand.

Der Beitrag ist im Rahmen des BMVI-Teilprojekts „Quo Vadis“ entstanden.

1 Das Phänomen der „Complacency“ beschreibt, dass Personen geneigt sind, einem automatisierten System zu vertrauen, wenn seine Vorschläge in der Regel gute Ergebnisse liefern (vgl. *Parasuraman, Manzey*, HFES 2010, S. 381–410, <https://doi.org/10.1177/0018720810376055>).

2 *Hennemann*, Legal Revolutionary, 2020, S. 77, Rn.4–6.

3 Das Wort „Bank“ kann sowohl mit dem Kontext „Finanzinstitut“ oder „Sitzgelegenheit“ verbunden sein.

Prof. Dr. Kai v. Lewinski

lehrt Öffentliches Recht, Medien- und Informationsrecht an der Universität Passau.

E-Mail: kai.lewinski@uni-passau.de

Johanna Hähnle

ist wissenschaftlicher Mitarbeiterin an der Forschungsstelle für Rechtsfragen der Digitalisierung (FREDI) der Universität Passau.

E-Mail: johanna.haehnle@uni-passau.de

aus einer informatischen Perspektive die Begriffe „Validität“ und „Qualität“ skizziert (2.). Dann werden einige rechtliche Fallkonstellationen vorgestellt, bei denen sich von rechtlicher Warte her Fragen hinsichtlich der „Richtigkeit“ stellen (3.). Anschließend wird überprüft, ob und wie (4.) in der Informatik entwickelte Richtigkeitsbegriffe auch juristisch operabel gemacht werden können.

2 Begriffsbestimmungen

Die Rechtswissenschaft ist zwar allgemein für die Präzision der Begriffsbildung und -verwendung bekannt (und berüchtigt). Diesem Nimbus wird sie aber im Informationsrecht vielfach nicht gerecht:⁴ *Datenschutzrecht*⁵ bezieht sich auf Informationen, *Informationsfreiheitsrecht* dagegen oft auf Daten, § 2 Nr. 2 IFG spricht von „personenbezogene[n] Daten und sonstige[n] Informationen“, § 2 Abs. 3 UIG definiert Umweltinformationen als „umweltbezogene Daten“, und „informationstechnisches System“⁶ meint Datenverarbeitungsanlage.

In der Informatik gibt es zwar (ebenfalls) keine eindeutige Terminologie⁷, aber jedenfalls eine weitgehend konsentiert der Semiotik entlehnte Unterscheidung⁸ zwischen Daten, Information und Wissen:

- Die syntaktische Ebene bildet Zeichen ab, ohne dass es auf deren Bedeutung ankäme (Daten). Auf diese Weise werden Informationen der realen Welt in Zeichenfolgen übersetzt.
- Auf semantischer Ebene stehen die Zeichen(folgen) dann in einem Kontext und werden so zu Information. Im „semantic web“ werden die (weitgehend unstrukturierten) Informationen des Internets mit bestimmten Bedeutungen verknüpft.⁹ Dazu werden Metadaten, seltener (aber nach dem Verständnis dieses Beitrags passender) auch Metainformationen genannt, ge-

4 S.a. *Oster*, JZ 2021, S. 167, 168.

5 Zur Begriffskritik m.w.N. auch v. *Lewinski*, Die Matrix des Datenschutzes, 2014, S. 3 ff.; vgl. schon *Steinmüller*, EDV und Recht – Einführung in die Rechtsinformatik, 1970, S. 87.

6 BVerfGE 120, 274 (Ls. 2 et pass.) – Online-Durchsuchung.

7 *Apel, Behme, Eberlein, Merighi*, Datenqualität erfolgreich steuern, 2015, S. 3; auch *Helfert*, Planung und Messung der Datenqualität in Data-Warehouse-Systemen, 2002, S. 13; *Zech*, Information als Schutzgegenstand, 2012, S. 13 ff.

8 Vgl. zum Semiotischen Dreieck in der Informatik: *Rehbein* in Jannidis/Kohle/Rehbein, Digital Humanities, Eine Einführung, 2017, S. 163; *Helfert* (Fn. 7), S. 14 m.w.N.; *Zech*, (Fn. 7), S. 25 m.w.N.

9 *Breitmann, Casanova, Truszkowski*, Semantic Web, 2007, S. 6.

nutzt. Metainformationen enthalten strukturierte Informationen über die Merkmale anderer Daten.

- Die Verarbeitung der Information wird auf der pragmatischen Ebene abgebildet (Wissen).¹⁰ Auf dieser Ebene zeigt sich, ob die Information für die verarbeitende Stelle von Nutzen ist.¹¹

Diese dritte pragmatische Ebene soll hier jedoch nicht weiter betrachtet werden. Vielmehr liegt der Fokus auf der Daten- und Informationsebene, denn nur wenn es auf diesen beiden Ebenen „richtig“ zugeht, kann eine belastbare Wissenssebene überhaupt entstehen.

Konzepte, um Richtigkeit abzubilden, gibt es in großer Zahl. Im Informationsmanagement als angewandter Informatik haben sich die Begriffe der Validität und Qualität als Maßstäbe für Richtigkeit etabliert. Beide Prinzipien sollen zunächst kurz erläutert werden:

2.1 Validität

Ursprünglich der Statistik und der empirischen Forschung entlehnt,¹² beschreibt Validität die Gültigkeit und Genauigkeit einer Messung: Wurde das erfasst, was erfasst werden sollte? Und wie genau bilden die Messergebnisse den abzubildenden Gegenstand ab?¹³ Hier können je nach Anwendungsgebiet unterschiedliche Überprüfungskriterien greifen.¹⁴ Die Präzision der Messung fragt im Kontext der Validität lediglich danach, ob das Messgerät überhaupt in der Lage ist, den gesuchten Wert zu messen. Die Validität beschreibt, wie genau tatsächliche Werte der realen Welt auf der syntaktischen Ebene, der Datenebene, widerspiegelt werden. Validität kann deshalb als *Datenrichtigkeit* bezeichnet werden.

2.2 Qualität

Qualität beschreibt den Grad von Güte, das Resultat eines Prozesses, ursprünglich von Produkten der Fertigungsindustrie. Wann immer Qualität bestimmt werden soll, werden zwei Ansätze besonders häufig genannt:¹⁵ Einerseits die fünf Ansatzpunkte nach der Definition des Harvard-Wirtschaftswissenschaftlers *David A. Garvin*¹⁶, andererseits die DIN- und DIN ISO EN-Normen. Beiden ist gemein, dass sie einen Kontext definieren, in dem etwas von Qualität ist. Ohne einen solchen Kontext kann keine Qualität gemessen werden. Während DIN EN ISO 9000 Begriffe des Qualitätsmanagements beschreibt, bezieht sich DIN 55350 auf die Qualitätssicherung. Von *Garvins* Ansätzen erweisen sich v.a. Produkt- und

Anwendungsorientierung als Anknüpfungspunkte für Daten- und Informationsqualität sinnvoll. Auch DIN 55350 (Qualität als „Gesamtheit von Eigenschaften und Merkmalen eines Produkts oder einer Tätigkeit, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen“) und DIN ISO EN 9000 („Grad, in dem ein Satz inhärenter Merkmale eines Objekts Anforderungen erfüllt“) setzen an diesen Punkten an. Die vereinbarten oder erwartbaren Eigenschaften variieren von Produkt zu Produkt und bilden sogenannte Qualitätsmerkmale.¹⁷ Mehrere Qualitätsmerkmale können ein Qualitätsmodell bilden.¹⁸

Qualitätsmodelle und -standards wurden ursprünglich für die Fertigungsindustrie entwickelt, lassen sich aber auf die Daten- und Informationsqualität übertragen.¹⁹ Die Qualitätsmerkmale unterscheiden sich allerdings: Für gegenständliche Produkte gibt es vielfach bereits Standards (auf die auch im rechtlichen Kontext Bezug genommen wird²⁰), für Datenqualität hat sich noch kein solcher etabliert, der auch im juristischen Kontext genutzt würde²¹ (vgl. die unterschiedlichen Bewertungen in 3.2. und 3.3.).

Aber auch hier bestimmt sich der Qualitätsstandard nach dem Kontext der Informationen. Im Zusammenhang mit großen Datenmengen (Big Data/Master Data) wird seit 2009 die ISO-Normserie 8000 entwickelt. Hier wird dann Datenqualität als „Grad, zu dem ein Datensatz Voraussetzungen erfüllt“ („degree to which a set of inherent characteristics of data fulfils requirements“), beschrieben. Die folgenden Dimensionen werden häufig als Qualitätskriterien betrachtet: Availability (Verfügbarkeit), Usability (Nutzbarkeit), Reliability (Verlässlichkeit), Relevance (Relevanz) und Presentation Quality (Lesbarkeit).²² Auch im Forschungsdatenmanagement werden Standards darauf ausgelegt, die Daten anwenderfreundlich aufzubereiten.²³ In diesen Dimensionen wird ebenfalls die Anwendungsorientierung deutlich.

Qualität kann immer nur im Kontext eines Bezugspunktes, eines Standards, bestehen. Sie bildet die *Informationsrichtigkeit* im Kontext ab. Während die Validität die *Datenrichtigkeit*, die Richtigkeit auf der syntaktischen Ebene, meint, sind Qualitätsmerkmale der Standard für die *Informationsrichtigkeit*, also die semantische Ebene. Im World Wide Web werden auf semantischer Ebene Daten um Metadaten ergänzt, um Informationen zu erzeugen.

Nach dieser Identifikation der Kernbegriffe sollen nun Fälle, in denen fehlerhafte Daten und Informationen juristisch relevant wurden, betrachtet werden.

3 Juristische Bedeutung von unrichtigen Daten und Informationen

Unrichtige Daten und Informationen können rechtliche Folgen nach sich ziehen (Schadenersatz, Unterlassung, Strafe usw.). Da-

10 *Apel, Behme, Eberlein, Merighi*, (Fn. 6), S. 4.

11 *Krcmar*, Einführung in das Informationsmanagement, 2015, S. 30.

12 *Soziologie-Lexikon*, (<https://doi.org/10.1515/9783486804362-fm>) S. 695.

13 *Himme* in *Albers/Klapper/Konrad/Walter/Wolf*, Methodik der empirischen Forschung, S. 375–382; *Hilgers, Heussen, Stanzel* in *Gressner/Arndt*, Lexikon der Medizinischen Laboratoriumsdiagnostik, 2019, S. 2425.

14 V.a. in der empirischen Forschung wird häufig die Validität der Messung unbeobachtbarer Werte diskutiert, vgl. *Sedlmeier, Renkewitz*, Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler, S. 85. Hier soll dagegen lediglich die Genauigkeit – dieser Begriff wird häufig austauschbar mit dem der Validität verwendet – von tatsächlich messbaren Werten betrachtet werden (vgl. *Saint-Mont*, Die Macht der Daten, 2014, S. 55).

15 Detaillierte Ausführungen zum Qualitätsbegriff finden sich bei *Helfert*, Planung und Messung der Datenqualität in Data-Warehouse-Systemen, 2002, S. 66 ff.

16 *Garvin*, Sloan Management Review, 1984, S. 40 ff. unterscheidet produktorientierte (objektiv bestimmbare), anwendungsorientierte (also die Geeignetheit für die Verwendung), prozessorientierte (auf die Optimierung des Herstellungsverfahrens gerichtete), wertbezogene (Kosten-Leistungs-Verhältnis) und transzendente (dem philosophischen Konzept der Vortrefflichkeit etc. entsprechende) Ansatzpunkte für Qualität.

17 *Behme* in *Kemper/Mayer*, Business Intelligence in der Praxis.2002, S. 52.

18 *Apel et al.* (Fn. 6).

19 *Wang, Ziad, Lee* 2001, S. 3 f.

20 Im Kaufrecht ist von „mittlerer Art und Güte“, die sich für industrielle Produkte an den Fertigungsnormen orientiert, die Rede; vgl. *Beurskens* in *BeckOGK*, BGB § 243 Rn. 34.

21 https://dl.gi.de/bitstream/handle/20.500.12116/25005/paper4_06.pdf?sequence=1&isAllowed=y, S. 374; in der informatischen Literatur gibt es zahlreiche, wenn auch nicht einheitliche Vorschläge für Metriken und Standards, vgl. beispielsweise *Heinrich/Klier* in *Hildebrand/Gebauer/Mielke*, Daten- und Informationsqualität, 2020, S. 47 ff.

22 *Hoeren*, MMR 2016, S. 8.

23 Vgl. https://www.ruhr-uni-bochum.de/researchdata/de/fair_data.html.

bei werden Fehler der verschiedenen Ebenen (syntaktisch und semantisch) an unterschiedlicher Stelle relevant. In der nachfolgenden kategorisierenden Aufstellung wird skizziert, welche Konstellationen das Recht bereits gut bewältigt und wo die Rechtswissenschaft und erst recht der Gesetzgeber noch auf der Suche nach Kriterien ist.

3.1 Messfehler

Der klassische Messfehler beruht entweder auf einem fehlerhaften Sensor oder einer fehlerhaften Nutzung des Messgeräts. Hier ist bereits die (Daten)Grundlage falsch und damit nicht valide, die Frage der Qualität der hierauf beruhenden Information stellt sich gar nicht (vgl. 2.1.).

Ein solcher Fehler lag etwa den Abstürzen der Boeing 737 Max²⁴ zugrunde: Der Sensor des MCAS²⁵ hatte fälschlicherweise einen zu hohen Anstellwinkel (und damit einen drohenden Abriss des Luftstroms) gemessen und einen Sinkflug eingeleitet. Die Messwerte des Winkels, auf deren Güte es für die Steuerung eines Flugzeuges ansonsten ankommt, spielten keine Rolle mehr. Ähnliche Fehler (mit weniger drastischen Konsequenzen) lassen sich in den juristischen Fallsammlungen zahlreich finden: defekte Schwimmschalter, die Brände verursachen²⁶, falsch installierte Wasserzähler, die zu hohe Wasserrechnungen nach sich ziehen²⁷, und verkleinerte Gärten aufgrund von Einmessfehlern²⁸.

Diese Fehler der Validität bzw. Datenrichtigkeit unterfallen der juristischen Kategorie der Pflichtverletzungen der Messenden und sind damit rechtsfolgenauslösend.²⁹

3.2 Darstellungs- und Kontextualisierungsprobleme

Auch die Kontextualisierung der Daten oder die Sortierung bzw. Darstellung von Informationen kann fehlerhaft erfolgen. Valide Daten werden durch Umrechnungsfehler, Missverständnisse über Maßeinheiten oder Kommafehler in einen unrichtigen Kontext versetzt. Hier werden richtige Daten nicht als brauchbare (oder als nicht brauchbare) Informationen abgebildet:

- Der sogenannte „Millennium-Bug“³⁰ führte zum Jahreswechsel 1999 auf 2000 zu zahlreichen, teils antizipierten Problemen und rechtlichen Auseinandersetzungen.³¹ Datumsangaben wurden zwar valide eingegeben und auch verarbeitet, jedoch nur die letzten beiden Stellen gespeichert. Dies führte dazu, dass in zahlreichen Anwendungen das Jahr 2000 nicht auf das Jahr 1999 folgte.
- Software, die adlige Namen entgegen der bibliographischen Ordnungsregeln³² nach dem Namensprädikat (also unter „v“) sortiert, wurde in der Rechtsprechung als mangelhaft bewertet.³³

24 Vgl. die Sicherheitswarnung der FFA vom 07.11.2018, AD#: 2018-23-51: [https://rgl.faa.gov/Regulatory_and_Guidance_Library/rgad.nsf/0/83ec7f95f3e5bfbdb8625833e0070a070/\\$FILE/2018-23-51_Emergency.pdf](https://rgl.faa.gov/Regulatory_and_Guidance_Library/rgad.nsf/0/83ec7f95f3e5bfbdb8625833e0070a070/$FILE/2018-23-51_Emergency.pdf)

25 Maneuvering Characteristics Augmentation Systems.

26 BGHZ 67, S. 359 ff. – Schwimmschalter.

27 AG Buxtehude, NJW-RR 2005, S. 1651.

28 OLG München, BeckRS 2017, 157736.

29 Liegt der Fehler im Gerät selbst, so greift in der Regel die Produkthaftung, liegt der Fehler beim Messenden, kommt es auf das Vertretenmüssen an.

30 Auch „Jahr-2000-Problem“ oder „Y2K-Bug“.

31 Wenngleich der teils prophezeite totale Zusammenbruch ausblieb; näher zu Problemen und Lösungen: Bartsch, Software und das Jahr 2000, 1998.

32 DIN 31638; Thieme, Das Deutsche Personenrecht, 2003, S. 120.

33 LG Bonn, Urt. v. 22.7.1997 – 10 O 497/96 (zu einem Adressprogramm); of-fengelassen von OLG Köln, NJW-RR 1999, S. 51, 51.

- Besteht die Möglichkeit, dass ein Geschwindigkeitsmessgerät im Straßenverkehr seine Rohdaten zu unzuverlässigen Geschwindigkeitsergebnissen auswertet,³⁴ muss es die Möglichkeit einer Überprüfung durch die Offenlegung der Rohdaten geben.³⁵ Nur so kann eine angemessene Verteidigung und ein faires Verfahren gewährleistet werden.

In diesen Fällen liegt das Problem in der verwendeten Software und löst unterschiedliche Rechtsfolgen aus: Führt ein Problem während der Kontextualisierung zu einem Vertragsschluss, kann z.B. ein Vertrag nicht mehr angefochten werden,³⁶ es sei denn, der Kontextfehler ist für die andere Vertragspartei erkennbar.³⁷ Entsteht aus einem solchen Fehler ein Schaden (wenn bspw. ein Flugzeug notlanden muss, weil die Treibstoffmenge mit dem Umrechnungsfaktor lb/l statt kg/l berechnet wurde³⁸), so kommt es auf die Vermeidbarkeit des Fehlers und die Zurechnung zum Pflichtenkreis an. Handelt es sich etwa beispielsweise um einen Druckfehler, durch den ein Wert verfälscht wird, bleibt dies medienrechtlich meist folgenlos.³⁹

In all' diesen Konstellationen werden valide Daten so abgebildet, dass sie nicht als „richtige“ Information dienen. Es stellt sich nicht die Frage nach der (vorhandenen) Validität der Daten, sondern dem fehlenden informationellen Wert einer unbrauchbaren Darstellung valider Daten. Die juristische Beurteilung jedoch variiert – um es mit einer Juristenweisheit auszudrücken: „Es kommt darauf an.“ Im nächsten Abschnitt (4.) soll überprüft werden, ob die informatischen Kategorien das juristische „es kommt darauf an“ präzisieren können.

3.3 Informations- und Datenumfang

Zunächst aber noch wiederum anders stellt sich die Situation bei Datenauswertungskonstellationen dar. Da immer nur eine endliche Datenmenge ausgewertet werden kann, kommt es nicht nur auf die Richtigkeit der Daten an, sondern auch auf deren Auswahl, Zusammenstellung und Gewichtung.

Damit ist nicht so sehr das Problem des „Information Overflow“ gemeint, wie es v.a. im Verbraucherschutz vorkommen kann und besonders sinnfällig in den dokumentierten Informa-

34 Dies lag zuletzt auch bei Geräten vor, die als standardisiertes Messverfahren galten. Vgl. Physikalisch-Technische Bundesanstalt (PTB), Zwischenstand im Zusammenhang mit mutmaßlichen Messwertabweichungen beim Geschwindigkeitsüberwachungsgerät Leivtec XV3, <https://doi.org/10.7795/520.20201215>.

35 BVerfG 2 BvR 1616/18, Rn. 1–72: Grundsätzlich ist bei sog. „standardisierten Messverfahren“ von einer korrekten Messung auszugehen. In der Vergangenheit und auch aktuell gab und gibt es jedoch Fälle, in denen in gewissen Konstellationen zwar korrekte Rohdaten erhoben werden, jedoch durch deren Verwendung bei einem „nachträglichen Schätzwert“ zu einem ungenauen Ergebnis kommen, wenn sich die Messunsicherheit einzelner Werte addiert, statt wie üblich ausgleichend zu wirken (vgl. detailliert hierzu PTB, Antworten auf häufige Fragen zum Laserscanner-Geschwindigkeitsüberwachungsgerät PoliScan der Fa. VITRONIC, <http://dx.doi.org/10.7795/520.20161209B>, S. 3). Durch die Offenlegung der Rohdaten kann ein solcher Fehler nachgewiesen oder ausgeschlossen werden. Vgl. auch Rebler, SVR 2019, S. 401.

36 Armbrüster in MüKoBGB, BGB § 119 Rn. 81.

37 Wenn beispielsweise eine Schulleiterin die Bedeutung der Einheit „Gros“ nicht kennt und statt 25 Rollen Toilettenpapier 3600 Rollen und damit den Bedarf für mehrere Jahre bestellt (LG Hanau NJW 1979, S. 721).

38 <https://aviation-safety.net/database/record.php?id=19830723-0&lang=de>.

39 Meyer, ZUM 1997, S. 26 zu Haftungsaspekten bei Fehlern (u.a. auch Druckfehlern) in wissenschaftlichen Werken. Eine Produkthaftung für bloße, nicht übermäßige Druckfehler wird hier wie auch in BGH NJW 1970, S. 1963 ff., abgelehnt.

tions- und Beratungspflichten bei der Anlageberatung ist.⁴⁰ Hier kann die Anforderung der Vollständigkeit zu einer Dysfunktionalität führen, wenn die Informationen von den Zielpersonen nicht mehr zu Wissen verarbeitet werden können.

Vielmehr ist an Regelungen wie die bei personenbezogenen Score-Werten zu denken, die durch automatisierte Auswertung entstehen. Bei ihnen ist Aktualität und „Richtigkeit“ der Informationen elementar⁴¹ und deshalb gesetzlich geregelt.⁴² Auch die Trainingsdaten⁴³ selbstlernender⁴⁴ KI-Systeme sind auf „Richtigkeit“ i.S.e. Vollständigkeit angewiesen. Andernfalls steigt das Diskriminierungspotential und damit das rechtliche Risiko bei der Verwendung.⁴⁵ KI-Systeme im allgemeinen und deren Trainingsdaten im besonderen sind Gegenstand europäischer Gesetzgebung.⁴⁶ Hier wird ein Regulierungsrahmen angeregt,⁴⁷ konkrete juristische Anforderungen an die Trainingsdaten bestehen aber (noch) nicht.

Bemerkenswerterweise bestehen aber auch gesetzliche Regelungen oder Konventionen, wonach Angaben *nicht* vollständig sein müssen oder gar nicht dürfen:⁴⁸ Vor Beginn eines Arbeitsverhältnisses darf eine Bewerberin nicht nach einer bestehenden Schwangerschaft gefragt werden;⁴⁹ selbst wenn eine Schwangerschaft vorliegt (und deshalb auch bestimmte Tätigkeiten vorübergehend nicht ausgeführt werden können), gilt die Bewerberin als „nicht schwanger“. In den USA sind bei Bewerbungen anonyme Lebensläufe verbreitet, um eine Diskriminierung aufgrund Alter, Herkunft, Hautfarbe etc. im Auswahlprozess zu vermeiden. Ähnliche „bewusste Unvollständigkeiten“ gibt es im Recht beispielsweise im Erbrecht, wo ein bereits gezeugtes, aber noch nicht geborenes Kind als bereits geboren gilt (§ 1923 Abs. 2 BGB), oder nach § 24 Abs. 2 BZRG, wonach Vorstrafen von Personen, die älter als 90 Jahre sind, zu entfernen sind.

4 Juristische Operabilität informatischer Begriffe?

Es kristallisieren sich drei Ebenen, auf denen die Richtigkeit von Daten bzw. Informationen relevant wird, heraus: Die Ebene der gemessenen, nicht interpretierten Daten (3.1.), der Prozess, der diese (Roh) Daten in ihren Kontext setzt (3.2.) und die der kontextualisierten Daten (Informationen) und ihres Umfangs (3.3.).

Auf der ersten Ebene hat das Recht bereits passende Antworten gefunden (vgl. 3.1. a.E.). Simple Messfehler sind durchaus kein Problem der Digitalisierung – man denke nur an den Mythos des eisenhaltigen Spinats⁵⁰ –, sondern lassen sich mit bestehenden

rechtlichen Kategorien erfassen. Aus informatischer Perspektive ist hier die Validität betroffen (vgl. bei 3.1.). Aus rechtlicher Perspektive haftet der für die fehlerhafte Messung Verantwortliche für die Folgen.⁵¹

Sobald valide Daten als Grundlage vorhanden sind, stellt sich die Frage nach der Informationsrichtigkeit, der Qualität. Hier können die informatischen Dimensionen Availability (Verfügbarkeit), Usability (Nutzbarkeit), Reliability (Verlässlichkeit), Relevance (Relevanz) und Presentation Quality (Lesbarkeit)⁵² (vgl. 2.2.) eine erste Orientierung bieten.

Der Prozess der Kontextualisierung und Darstellung (Ebene zwei, 3.2.) hat valide Daten zum Ausgangspunkt. Hier können sich Probleme ergeben, wenn die Daten nicht in den richtigen Bezug gesetzt beziehungsweise so dargestellt werden, dass sie nicht angemessen genutzt werden können. Die juristische Bewertung variiert in diesem Punkt (vgl. 3.2.): Hier muss geklärt werden, wann Informationen rechtlich „richtig“ sind. Bei „falscher“ Kontextualisierung ist in der Regel insbesondere die Dimension der *Nutzbarkeit* eingeschränkt. Wann diese Einschränkung erkennbar wird und wessen Sphäre sie zuzuordnen ist, variiert und ist rechtlich nicht immer eindeutig greifbar.⁵³

Steht hingegen entweder keine genügende Anzahl valider Daten zur Verfügung, um ausreichende Informationen für den geforderten Kontext zu erreichen, oder ist der Datensatz so gewählt, dass ein algorithmisches System „falsche“ Schlüsse zieht, kann es zu „falschen“ Ergebnissen kommen (Ebene 3, vgl. 3.3.). Informationsqualität steht hier im Fokus der *Verlässlichkeit*.

So wird beim Einsatz von Algorithmen der sogenannte Algorithmic Bias und das Problem der fehlenden Unterscheidung von Korrelation und Kausalität relevant.⁵⁴ Für ein diskriminierungsfreies System muss auf eine angemessene Repräsentation aller Gruppen geachtet werden,⁵⁵ beim Scoring sind bestimmte Anforderungen ebenfalls festgelegt.⁵⁶

Im Umgang mit KI-Trainingsdaten sind aus den gleichen Gründen hohe Anforderungen an die Qualität der Informationen, die sie abbilden sollen, zu stellen.⁵⁷ Für Datenbanken und andere Systeme, die solche Daten aggregieren, gibt es in der Literatur der Informatik bereits verschiedene Metriken und Qualitätsstandards, die solche Daten und Informationen erfüllen müssen.⁵⁸ Juristisch bestehen hier (noch) keine einheitlichen Kriterien, erste Vorschläge hat die EU-Kommission im April 2021 umrissen. Der Fokus der Trainingsdaten soll auf „Relevanz, Repräsentativität, Fehlerfreiheit und Vollständigkeit“ liegen.⁵⁹ Ein

40 BGH NJW 2015, 398.

41 Schantz in Simitis/Hornung/Spiecker, Datenschutzrecht, DSGVO, 2019, Art. 6 Abs. 1 Rn. 138; Art. 5 Abs. 1 lit. d DSGVO.

42 Art. 5 Abs. 1 DSGVO regelt die Grundsätze der Datenverarbeitung personenbezogener Daten, EG 71 DSGVO.

43 Eine knappe Übersicht zum Training der KI bei Hacker, GRUR 2020, S. 1025, 1026 f.

44 Zur Unterscheidung selbstlernender und regelbasierter Systeme Unger, ZRP 2020, S. 234, 235 f.

45 Vgl. zum Schadenersatzanspruch nach § 15 AGG v. Lewinski, de Barros Fritz, NZA 2018, S. 620.

46 Europäische Kommission, Weißbuch zur Künstlichen Intelligenz, COM(2020) 65 final.

47 Weißbuch zur Künstlichen Intelligenz (Fn. 46), S. 22.

48 Hennemann, Legal Revolutionary 2020, S. 77, 85.

49 § 3 Abs. 1 S. 2 AGG, LAG Köln, BB 2012, S. 3200.

50 Damals wurde getrockneter Spinat verwendet, der Wassergehalt wurde ignoriert. Die Art der Messung war also nicht geeignet, den Eisengehalt frischen Spinats zu bestimmen. Die Messung war nicht valide.

51 Dies gilt sowohl für Verhältnisse ohne Machtgefälle, wie z.B. einige Vertragstypen (vgl. den oben genannten Einmessfehler), als auch solche mit Machtgefälle (vgl. den Fall des fehlerhaften Zählers, aber auch z.B. fehlerhafte Geschwindigkeits- oder Blutalkoholmessungen).

52 Hoeren, MMR 2016, S. 8.

53 Vgl. die Debatte um die Haftung für falsche Informationen im Fall „Kräuterpfarrer“, EuGH, C-65/20 (Entscheidung gegen eine Haftung am 10.06.2021): Generalanwalt beim EuGH, Schlussantrag v. 15.4.2021 – C-65/20, BeckRS 2021, 7667 Rn. 29 ff.; v. Lewinski, Medienrecht, 2020, § 20 Rn. 61; aber auch die Diskussion über den Irrtum des Täters über Giftmengen: Eser, Bosch in Schönke/Schröder/Eser/Bosch, StGB, 30. Aufl. 2019, § 24 Rn. 22.

54 Freyler, NZA 2020, S. 284, 286.

55 Vgl. Hacker, GRUR 2020, S. 1025, 1027 m.w.N.

56 v. Lewinski, Pohl, ZD 2018, S. 17, 21.

57 Vgl. Hacker, NJW 2020, S. 2142, 2144 m.w.N.

58 Vgl. Fn. 21.

59 Proposal for an Artificial Intelligence Act, COM (2021) 206 final, Art. 10 Abs. 2–5.

möglichst umfangreicher, teils sogar um fiktive Informationen ergänzter⁶⁰ Datensatz ist notwendig.⁶¹

Demgegenüber bestehen auch rechtliche Konstellationen, in denen ein Weniger an Informationen einem höheren Qualitätsstandard entspricht (vgl. 3.3.). Dies mag zunächst paradox klingen, jedoch sind gegenläufige Prinzipien beispielsweise im Datenschutzrecht, das einerseits eine hohe Datenqualität („Richtigkeit“, Art. 5 Abs. 1 lit. d DSGVO) fordert, andererseits aber auch zur Datensparsamkeit (Art. 5 Abs. 1 lit. c DSGVO) mahnt, nicht unbekannt. Wie bereits angedeutet, können solche „Begrenzungserfordernisse“ auch dem Diskriminierungsschutz dienen (vgl. 3.3.). Um dieses Ziel zu erreichen, werden also sehr unterschiedliche Qualitätserfordernisse (möglichst umfangreich, nur so viel wie notwendig, mit Fiktion oder eine Kombination), relevant. Spätestens hier wird deutlich, dass Informationsqualität juristisch betrachtet vom Qualitätsideal der informatischen Perspektive (Verfügbarkeit, Nutzbarkeit, Verlässlichkeit, Relevanz, Lesbarkeit) abweicht.

5 Schlussfolgerungen

Die informatischen Konzepte der Validität und Qualität können den Zugriff für das rechtlich schwer greifbare Konzept der „Richtigkeit“ erleichtern.

Validität als Maßstab für die Richtigkeit der gemessenen/erhobenen Daten erscheint sinnvoll. Bevor die Daten im Kontext zu Informationen werden können, müssen sie ein Realphänomen abbilden. Dabei stellt die Validität nicht das reine Maß der Genauigkeit dar, sondern vielmehr die Eignung der Art der Datenerhebung. Die Validität ist ein Kriterium auf der Ebene der Datenrichtigkeit.

Auf der Ebene der Informationen wird der Kontext der Daten relevant. An dem jeweiligen Zusammenhang (sowohl der Daten als auch der Präsentation der Informationen) orientieren sich die Qualitätsmerkmale⁶², aus denen sich die Informationsrichtigkeit ergibt.

Während also Validität als Konkretisierung des juristischen Richtigkeitsbegriffs einen eindeutigen Anhaltspunkt darstellt, ist das Ausmaß der Qualität von der Erfüllung von Qualitätsmerkmalen abhängig. Hier haben insbesondere in Bezug auf die digitale Entwicklung juristische Standards noch Entwicklungsbedarf.⁶³

Die für Datenqualität im Big Data-Kontext entwickelten Dimensionen Verfügbarkeit, Nutzbarkeit, Verlässlichkeit, Relevanz und Lesbarkeit⁶⁴ können rechtlich operable Leitplanken zur juristischen Bewertung der Richtigkeit darstellen.

Den genauen „Fahrstil“ hingegen muss das Recht selbst herausarbeiten:

- Im staatlichen Umgang mit Informationen hat sich für die passive Bereitstellung ein Fokus auf Verfügbarkeit und Lesbarkeit⁶⁵, bei der aktiven Bereitstellung eine Hervorhebung der

Verlässlichkeit herausgebildet.⁶⁶ Den Rahmen dieser Standards geben die Grundrechte vor, die Präzisierung erfolgt durch die verfassungsgerichtliche Rechtsprechung.

- Im Zivilrecht finden sich ebenfalls zahlreiche Wertungsmaßstäbe für Qualität, die aber vornehmlich von den Vertragsparteien gesetzt werden können. Hier steht oftmals das Kriterium der „Nutzbarkeit“ im Vordergrund. So stellt die „vereinbarte Beschaffenheit“ (vgl. § 434 BGB) als subjektives Kriterium auch objektiver Maßstäbe (vgl. z.B. § 3 Abs. 1 ProdHaftG) Qualitätsstandards dar. In Vertragskonstellationen, die durch Dysbalancen der Parteien keine individuellen Abreden zulassen, gewinnen objektive Kriterien an Einfluss, insb. im digitalen Raum nehmen diese Verträge zu.⁶⁷

Um der Dimension der „Relevanz“ Rechnung zu tragen, kennt das Recht mit dem Prinzip der Fiktion auch die Situation, dass unabhängig von realen Phänomenen eine Information als richtig festgestellt wird.⁶⁸

Qualitätsmerkmale bzw. die Schwerpunkte innerhalb der großen Dimensionen können gesetzlich abgebildet und durch Rechtsprechung präzisiert oder entwickelt werden, wenn eine Normierung nicht vorliegt.

Gesetzliche Vorgaben können auch einen Einfluss auf den Umfang der Qualitätsmerkmale, die für einen Qualitätsstandard herangezogen werden dürfen, haben. Beispiele hierfür finden sich etwa im Allgemeinen Gleichbehandlungsgesetz und der Datenschutz-Grundverordnung.

Im Recht wie in der Informatik ist Richtigkeit nicht binär – bei der Frage, worauf es ankommt, kann die Informatik wertvolle Ansätze bieten, auf die das Recht Bezug nehmen kann und auch sollte.

Open Access

Open Access funding enabled and organized by Projekt DEAL. Copyright: v. Lewinski, Hähnle 2021

Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 (CC BY) International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/ die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- [1] *Apel, Detlef/Behme, Wolfgang/Eberlein, Rüdiger/Merighi, Christian*, Datenqualität erfolgreich steuern: Praxislösungen für Business-Intelligence-Projekte, 3. Aufl., Heidelberg, 2015.
- [2] *Helfert, Markus*, Planung und Messung der Datenqualität in Data-Warehouse-Systemen, Bamberg, 2002.
- [3] *Hoeren, Thomas*, Thesen zum Verhältnis von Big Data und Datenqualität – Erstes Raster zum Erstellen juristischer Standards, MMR 2016, 8-11.
- [4] *Zech, Herbert*, Information als Schutzgegenstand, Tübingen 2012.

66 Aufgrund des Vertrauens in staatliche Informationen ist insb. bei Warnungen etc. großer Wert auf ein tatsächliches Zutreffen der Informationen zu legen, vgl. BVerfG, NJW 2002, S. 2621 – Glykol; BVerfG, NJW 2002, S. 2626 – Osho.

67 *Staudenmayer*, ZEuP 2019, S. 663, 679.

68 *Groh*, Creifelds, Rechtswörterbuch, Art. „Fiktion“; für Beispiele vgl. auch zu 3.

60 Vgl. Fn 55.

61 *Apel, Kaulartz*, RD 2020, S. 24.

62 Vgl. zur Kontextbezogenheit dieser Standards *Rossi*, Legal Revolutionary 2020, S. 90, 91.

63 Ausführlich zu den Herausforderungen: *Zech*, Information als Schutzgegenstand, 2012, S. 183 ff.; zu Big Data und den offenen Fragen *Hoeren*, MMR 2016, S. 8.

64 *Hoeren*, MMR 2016, S. 8.

65 Vgl. Vorschriften, die Anforderungen an Lesbarkeit/Verfügbarkeit (vgl. § 19 GeoIDG, § 12a Abs. 5 EGVG) stellen, nicht aber an inhaltliche „Richtigkeit“ (vgl. § 18 Abs. 1 S. 2 GeoIDG, § 12a Abs. 8 EGVG, § 7 Abs. 3 S. 2 IFG).