

## Herausgeber

Professor Dr. Dr. Eric Hilgendorf, Würzburg  
Professor Dr. Matthias Jestaedt, Freiburg i.Br.  
Professor Dr. Florian Möslin, LL.M. (London), Marburg  
Professor Dr. Astrid Stadler, Konstanz

## Redaktion

Martin Ilder, Tübingen

Mohr Siebeck

19 78. Jahrgang  
6. Oktober 2023  
Seiten 833–884

# Juristen Zeitung

Aufsätze

Professor Dr. Daniel Effer-Uhe, Berlin\*

## Überlegungen zur Automatisierbarkeit der Rechtsanwendung

Die Diskussion darüber, inwiefern die Rechtsanwendung oder Teile der Rechtsanwendung auf Computerprogramme übertragen werden können, beschäftigt die Rechtswissenschaft seit Jahren zunehmend, in den Grundzügen seit Jahrzehnten. Im Folgenden soll es nicht darum gehen, ob eine solche Übertragung sinnvoll und mit verfassungsrechtlichen Vorgaben vereinbar ist, sondern darum, was an Automatisierung – derzeit oder überhaupt – theoretisch möglich ist, denn das Wissen um Möglichkeiten und Grenzen ist Voraussetzung dafür, überhaupt erst sinnvoll darüber nachzudenken, in welchen Bereichen und inwieweit eine Übertragung juristischer Tätigkeiten auf Computer angestrebt oder rechtlich ermöglicht werden sollte.

### I. Digitalisierbarkeit von Regeln

Der juristische Syllogismus geht von Rechtsnormen aus, die nach ihrer Struktur als Regel (oder Konditionalprogramm) konzipiert sind: „Wenn A und B, dann Rechtsfolge R.“ Oder etwas komplizierter: „Wenn A und B und C und D und nicht E, dann Rechtsfolge R.“ Wenn z.B. eine natürliche oder juristische Person eine Willenserklärung abgegeben hat und dabei einem Inhaltsirrtum unterlegen ist und sie innerhalb der Anfechtungsfrist eine Anfechtungserklärung gegenüber dem richtigen Adressaten abgegeben hat und das Rechtsgeschäft nicht von dieser Person bestätigt worden ist, dann ist die Willenserklärung als von Anfang an nichtig zu behandeln. Diese regelbasierten Teile des Rechtssystems werden allgemein als grundsätzlich digitalisierbar angesehen:<sup>1</sup> Regeln lassen sich ohne weiteres in Computercode abbilden,<sup>2</sup> wobei Voraussetzung ist, dass die Tatbestände als Bedingungsgefüge exakt bestimmbar sind, also mit allen Alternativen, Ausnahmen etc.,<sup>3</sup> gegebenenfalls einschließlich negativer oder ungeschriebener Tatbestandsmerkmale, in den Computercode aufgenommen werden. Auch wenn der menschliche Rechts-

anwender sich in einem vertragsrechtlichen Fall normalerweise keine Gedanken über die Geschäftsfähigkeit der Beteiligten macht, sofern es dafür keine konkreten Anhaltspunkte im Sachverhalt gibt, muss der Algorithmus auch die Regeln der §§ 104 ff. BGB abarbeiten – um in den verhältnismäßig wenigen Fällen, in denen es darauf doch einmal ankommt, die Mängel der Geschäftsfähigkeit bei der Falllösung berücksichtigen zu können.

#### 1. Explizite und implizite Programmierung

Es handelt sich insoweit um explizite Programmierung, bei der alle Schritte zur Lösung einer Aufgabe a priori bekannt sind und in Programmcode übersetzt werden.<sup>4</sup> So lassen sich Anspruchsgrundlagen mit ihren Tatbestandsmerkmalen baumartig ausdifferenzieren (z. B. in Form einer graphischen Veranschaulichung aus didaktischen Gründen als „Rulemapping“<sup>5</sup>), die Struktur zeigt dann die zu prüfenden Punkte,<sup>6</sup> und exakt diese Baumstruktur ist es, die auch von einem Programm abgearbeitet wird.<sup>7</sup> Um eine solche explizite Programmierung geht es zum Beispiel, wenn vorgeschlagen wird, dass im Rahmen der Digitalisierung der Justiz Aktenauszüge entlang den Tatbestandsmerkmalen einer Norm automatisch erstellt werden könnten.<sup>8</sup>

Aber genügt derartige implizite Programmierung, um jegliche Rechtsanwendung in Computercode umwandeln zu können? An welchen Punkten der Rechtsanwendung versagt eine solche explizite Programmierung? Und kann dann eine – bei den derzeit verfügbaren Rechtsprüfungsprogrammen kaum genutzte<sup>9</sup> – implizite Programmierung weiterhelfen? Bei der impliziten Programmierung bzw. dem maschinellen Lernen werden nicht bestimmte Arbeitsanweisungen auf einem vom Programmierer vorbestimmten Weg Schritt für Schritt abgearbeitet; vielmehr werden anhand vorliegender Daten Muster – Zusammenhänge und Korrelationen – zwischen Eingangs- und Ausgangsdaten festgestellt und in Sys-

\* Der Autor ist Professor für Bürgerliches Recht, Rechtsgeschichte, Rechtslehre und Rechtspsychologie an der BSP Business and Law School.

<sup>1</sup> Breidenbach/Glatz, in: *Breidenbach/Glatz* (Hrsg.), *Rechtshandbuch Legal Tech*, 2. Aufl. 2021, Kap. 1.1 Rn. 15; Yuan, ebd., Kap. 9.4. Rn. 6.

<sup>2</sup> Breidenbach, in: *Breidenbach/Glatz* (Fn. 1), Kap. 2.2 Rn. 6; Bull/Steffek ZKM 2018, 165, 169 unterscheiden zwischen „Fällen technischer Art“ und „komplexen Fällen“, verstehen unter den Fällen technischer Art aber letztlich Fälle, deren Lösung regelhaft ohne größeren Einfluss von Abwägungsprozessen prädestiniert ist.

<sup>3</sup> Breidenbach, in: *Breidenbach/Glatz* (Fn. 1), Kap. 2.2 Rn. 22.

<sup>4</sup> Von Büna, in: *Breidenbach/Glatz* (Fn. 1), Kap. 3 Rn. 13.

<sup>5</sup> Breidenbach, in: *Breidenbach/Glatz* (Fn. 1), Kap. 2.2 Rn. 9.

<sup>6</sup> Breidenbach/Gaier, in: *Breidenbach/Glatz* (Fn. 1), Kap. 7.2. Rn. 5 f.

<sup>7</sup> Eine entsprechend vorgenommene Visualisierung entspricht also schon einem digitalen Format, kann also bereits als Code begriffen werden (Breidenbach, in: *Breidenbach/Glatz* [Fn. 1], Kap. 2.2 Rn. 12).

<sup>8</sup> So z. B. Gaier, in: *Breidenbach/Glatz* (Fn. 1), Kap. 7.1 Rn. 22.

<sup>9</sup> Fries, in: *Kaulartz/Braegelmann* (Hrsg.), *Rechtshandbuch Artificial Intelligence und Machine Learning*, 2020, Kap. 15.1 Rn. 2.

temverhalten übersetzt.<sup>10</sup> Die Grundidee des maschinellen Lernens liegt also darin, dass das Programm aus einer größeren Anzahl an Beispielen lernt, welche Kombinationen von Eigenschaften eines Objekts eine bestimmte Entscheidung über dieses Objekt wie beeinflussen.<sup>11</sup> So wird beim Deep Learning dem Programm nicht ein abzuarbeitender Algorithmus vorgegeben, sondern ein gewünschtes Ergebnis, und anhand vorliegender Trainingsdaten entwickelt das Programm selbständig einen Algorithmus, der bestmöglich zu den Trainingsdaten passt und so mit hoher Wahrscheinlichkeit auch bei neuen Daten die besten Ergebnisse zeitigt.<sup>12</sup> In einem künstlichen neuronalen Netz nehmen Neuronen Signale auf und geben sie weiter, wenn sie einen bestimmten Schwellenwert erreichen und nicht zugleich ein Sperrsignal eingeht.<sup>13</sup> Den Eingangswerten wird ein Schwellenwert zugeordnet, und wenn die eingehenden Signale diesen Schwellenwert erreichen, erfolgt eine Signalweiterleitung.<sup>14</sup> Die Schwellenwerte kann eine lernende Künstliche Intelligenz (KI) selbständig ändern und neuen Daten anpassen.<sup>15</sup>

Eine erste juristische Herausforderung müsste das Programm allerdings schon vorab leisten: Die Auswahl der relevanten Normen,<sup>16</sup> im Zivilrecht also z. B. die Auswahl einer Anspruchsgrundlage. Das ist durch explizite Programmierung jedenfalls nicht umfassend auch für untypische, selten auftretende Fallgestaltungen möglich. Eine explizite Programmierung stößt überdies bei unbestimmten Rechtsbegriffen, die einer Wertung bedürfen, an ihre Grenzen.<sup>17</sup> So ist z. B. § 826 BGB („Wer in einer gegen die guten Sitten verstoßenden Weise einem anderen vorsätzlich Schaden zufügt, ist dem anderen zum Ersatz des Schadens verpflichtet“) der Form nach eine Regel – man könnte umformulieren „Wenn jemand einem anderen Schaden zufügt (A), dabei vorsätzlich handelt (B) und gegen die guten Sitten verstößt (C), dann ist er dem anderen zum Schadensersatz verpflichtet (R)“ – „Wenn A und B und C, dann Rechtsfolge R.“ Aber das Tatbestandsmerkmal „in einer gegen die guten Sitten verstoßenden Weise“ ist als unbestimmter Rechtsbegriff der wertenden Ausfüllung bedürftig, die ein Computer jedenfalls im Wege expliziter Programmierung nicht leisten kann.

## 2. Subsumtionsschlüsse

Damit aber nicht genug: Die Subsumtion unter einen Rechtsbegriff kann die KI ebenfalls nicht ohne weiteres selbst vornehmen, hier ist der Mensch erforderlich.<sup>18</sup> Heißt es z. B. in

§ 224 Abs. 1 Nr. 1 StGB „durch Beibringung von Gift“, dann kann zwar das Programm dahin leiten, dass das Merkmal „Gift“ bejaht oder verneint werden muss – es kann aber selbst nicht entscheiden, ob eine Substanz Gift in diesem Sinne ist. Das ist zum einen möglich, indem der Benutzer an den „Ast-Enden“ der Baumstruktur in die Sachverhaltsgewinnung einbezogen wird, also die Fakten des Falls vom Programm bei einem Nutzer abgefragt werden:<sup>19</sup> Der Mensch könnte also z. B. gefragt werden, ob dem anderen Gift beigebracht wurde, dabei würde der Mensch die Subsumtion tätigen, ob es sich bei der fraglichen Substanz um Gift handelt und ob sie dem Opfer beigebracht wurde. Dieser Subsumtionsschluss ist eine Form von Klassifikation – es wird ein Begriff, der ein tatsächliches Ereignis oder hier eine tatsächliche Substanz beschreibt, in juristischer Hinsicht klassifiziert<sup>20</sup> (z. B. Arsen als Gift). Natürlich kann bereits in der Programmierung bestimmten Daten, die unmittelbar die Umwelt oder ein Ereignis beschreiben („Arsen“) ein Label („Gift“) beigelegt werden – dann muss die Subsumtion „Arsen ist ein Gift“ nicht mehr von einem Menschen vorgenommen werden. Aber hier tun sich schnell praktische Grenzen auf: Eine vollständig autonome Subsumtion würde nämlich als notwendige Bedingung eine *vollständige* Liste aller Gifte voraussetzen, die anzulegen niemand in der Lage ist – praktisch könnte also nur mit einer unvollständigen Liste gearbeitet werden, die zwar den automatisierten Schluss „Arsen ist ein Gift“ erlaubt, aber nicht den Schluss „jeder in der Liste nicht genannte Stoff ist kein Gift“. Das Programm wäre also zumindest für zweifelhafte oder neuartige Fälle nach wie vor auf menschliche Unterstützung angewiesen (siehe noch unten II. 4. c a. E.).

Der Unterschied zwischen bestimmten und unbestimmten Rechtsbegriffen ist ohnehin nur ein gradueller. Rechtsbegriffe, die vollständig von jeder Wertung frei sind, gibt es – von Ausnahmen wie Zahlen abgesehen – kaum. Selbst ein vermeintlich einfaches Tatbestandsmerkmal mit einem klar bestimmten Kernbereich verfügt meist über Randbereiche, in denen eine Wertung erforderlich bleibt.<sup>21</sup> So ist der Begriff „Stuhl“ auf den ersten Blick klar definiert: Vier Beine, Sitzfläche, Rückenlehne. Aber ist ein Hocker ohne Lehne womöglich auch ein Stuhl im Sinn der Norm? Ist ein kaputter Stuhl, von dessen vier Beinen zwei abgebrochen sind, ein Stuhl? Ist ein Sitzball ein Stuhl? Diese Fragen sind nicht ohne Wertung zu beantworten, und sie sind nicht abschließend für jede Konstellation zu beantworten. Es kann ohne weiteres sein, dass für die eine Rechtsnorm der Sitzball als Stuhl anzusehen ist (weil nach Sinn und Zweck dieser Norm die Eignung als Sitzmöbel entscheidend ist), für die andere Rechtsnorm dagegen nicht (weil Sinn und Zweck der Norm hier die konkrete Art des Sitzmöbels betreffen). Auch bei Normen, die als Regeln gefasst sind, werden also häufig Wertungen, Abwägungsprozesse erforderlich, auf einer untergeordneten Ebene innerhalb der einzelnen Tatbestandsmerkmale. Alle nicht völlig eindeutigen Fälle beinhalten Abwägungsentscheidungen (z. B. hinsichtlich der Gewichtung von Auslegungskanones, die im konkreten Fall in unterschiedliche Richtungen deuten).

<sup>10</sup> Von Bünauf, in: *Breidenbach/Glatz* (Fn. 1), Kap. 3 Rn. 14; *Hähnchen/Schrader/Weiler/Wischmeyer* JuS 2020, 625, 626.

<sup>11</sup> *Niederée/Nejdl*, in: *Ebers/Heinze/Krügell/Steinrötter* (Hrsg.), *Künstliche Intelligenz und Robotik – Rechtshandbuch*, 2020, § 2 Rn. 24.

<sup>12</sup> *Specht/Herold* MMR 2018, 40; *Armour/Eidenmüller* ZHR 183 (2019), 169, 174; *Effer-Uhe* RD 2021, 169, 170.

<sup>13</sup> Vgl. z. B. *Stiemering*, in: *Kaulartz/Braegelmann* (Fn. 9), Kap. 2.1 Rn. 21 ff.

<sup>14</sup> *Matthias*, *Automaten als Träger von Rechten*, 2. Aufl. 2010, S. 23 f.

<sup>15</sup> *Zech*, in: *Gless/Seelmann*, *Intelligente Agenten und das Recht*, 2016, S. 163, 169. Ausführlicher *Niederée/Nejdl*, in: *Ebers/Heinze/Krügell/Steinrötter* (Fn. 11), § 2 Rn. 59 ff.

<sup>16</sup> *Hähnchen/Schrader/Weiler/Wischmeyer* JuS 2020, 625, 627; *Schrader* BRAK-Mitteilungen 2020, 62, 66.

<sup>17</sup> *Krimphove/Niehaus*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.3 Rn. 5; vgl. auch *M. Engel* JZ 2014, 1096, 1097.

<sup>18</sup> *Breidenbach*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.2 Rn. 9; *Kotsoglou* JZ 2014, 451, 456; ähnlich *Grupp*, in: *Hartung/Bues/Halbleib* (Hrsg.), *Legal Tech – Die Digitalisierung des Rechtsmarkts*, 2018, Rn. 1100, 1116, demzufolge sich regelbasierte Prüfungstools überall dort anbieten, wo „Schwierigkeit und Aufwand weniger in der Definition eines einzelnen Tatbestandsmerkmals liegen, sondern wo gerade der Prüfungsprozess als solcher abgebildet werden soll“.

<sup>19</sup> *M. Engel* JZ 2014, 1096, 1099.

<sup>20</sup> *Yuan*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.4 Rn. 7.

<sup>21</sup> Vgl. zu Begriffskern und Begriffshof *Looschelders/Roth*, *Juristische Methodik im Prozeß der Rechtsanwendung*, 1996, S. 134 f.

## II. Digitalisierbarkeit von Abwägungsentscheidungen

Schon auf einer höheren Ebene finden Abwägungsentscheidungen statt, wenn Rechtsnormen nicht als Regeln, sondern als Prinzipien formuliert sind: als Optimierungsgebote. Etwas soll „in einem relativ auf die rechtlichen und tatsächlichen Möglichkeiten möglichst hohen Maße realisiert werden“. <sup>22</sup> Das geläufigste Beispiel für Prinzipien, die ihren „Sinngehalt erst im Zusammenspiel wechselseitiger Ergänzung und Beschränkung entfalten“, <sup>23</sup> sind die Grundrechte des Grundgesetzes. Prinzipienkollisionen – z. B. zwischen Kunstfreiheit und allgemeinem Persönlichkeitsrecht – sind dadurch zu lösen, dass in einer Abwägung nach dem höheren Gewicht der Prinzipien entschieden wird und im Rahmen der praktischen Konkordanz auch Lösungen zum Zug kommen können, in denen beide Prinzipien teilweise verwirklicht werden. <sup>24</sup> Können Prinzipienabwägungen also nicht automatisiert werden, wie es einer verbreiteten Auffassung entspricht? <sup>25</sup>

### 1. Prinzipienabwägungen als Schaffung von Regeln

Zum einen könnten natürlich Prinzipienabwägungen von menschlichen Entscheidern vorgenommen und damit regel-förmige Normen aus den Prinzipien abgeleitet werden, die dann wiederum digitalisierbar sind. Denn das Ergebnis jeder verallgemeinerungsfähigen Prinzipienabwägung ist eine Regel: Wenn das Prinzip A dem Prinzip B unter den näher angegebenen Umständen C vorgeht, und wenn sich unter diesen Umständen C aus der Abwägung der Prinzipien die Rechtsfolge R ergibt, dann gilt eine Regel mit C als Tatbestand und R als Rechtsfolge:  $C \Rightarrow R$ . <sup>26</sup> Alexy formuliert das als sogenanntes „Kollisionsgesetz“ folgendermaßen: „Die Bedingungen, unter denen das eine Prinzip dem anderen vorgeht, bilden den Tatbestand einer Regel, die die Rechtsfolge des vorgehenden Prinzips ausspricht.“ <sup>27</sup> Man kann also jede Prinzipienabwägung in die Form einer Regel überführen, die dann wieder einer expliziten Programmierung zugänglich ist. Das Problem dabei ist allerdings, dass die Anzahl möglicher Prinzipienkollisionen unendlich groß ist. Das zeigt sich schon allein am Bereich der Grundrechte (die natürlich nur einen winzigen Bruchteil der als Prinzip formulierten Rechtssätze unseres Rechtssystems ausmachen): Zwar gibt es nur eine begrenzte Anzahl an einzelnen Grundrechten, so dass die Kombinationsmöglichkeiten abzählbar sind, aber die konkreten tatbestandlichen Konstellationen, unter denen es zu derartigen Kollisionen kommen kann, sind unbegrenzt, wie sich in der Reichhaltigkeit der Rechtsprechung des *BVerfG* zu Grundrechten zumindest andeutet. Praktisch ist die Überführung von Prinzipienkollisionen in Regeln, die dann in Algorithmen umgesetzt werden, kein gangbarer Weg.

<sup>22</sup> Alexy, *Theorie der Grundrechte*, 1985 (hier zitiert nach der Ausgabe Frankfurt am Main 1994), S. 75; Dworkin, *Bürgerrechte ernstgenommen*, 1984, S. 61 f.

<sup>23</sup> Canaris, *Systemdenken und Rechtsbegriff*, 2. Aufl. 1983, S. 53.

<sup>24</sup> Alexy (Fn. 22), S. 79; Effer-Uhe, *Die Bindungswirkung von Präjudizien*, 2008, Rn. 28.

<sup>25</sup> Skeptisch z. B. Grupp, in: *Hartung/Bues/Halbleib* (Fn. 18), Rn. 1100, 1111; Rollberg, *Algorithmen in der Justiz*, 2020, S. 76; Dreyer/Schmees CR 2019, S. 758, 762; zumindest im Hinblick auf den Status quo ähnlich Wischmeyer, in: *Ebers/Heinze/Krügel/Steinrötter* (Fn. 11), § 20 Rn. 39.

<sup>26</sup> Effer-Uhe (Fn. 24), Rn. 29.

<sup>27</sup> Alexy (Fn. 22), S. 84.

### 2. Formalisierbarkeit von Abwägungsvorgängen

Sehr wohl theoretisch möglich ist es – trotz verbreiteter Skepsis – aber, Prinzipienabwägungen (und allgemeiner: alle Abwägungsprozesse) in mathematische Formeln zu überführen. Einen Vorschlag dazu habe ich an anderer Stelle <sup>28</sup> gemacht und anhand eines Beispiels veranschaulicht. So kann man sich der Theorie unscharfer Mengen (Fuzzy-Logik) bedienen, um Abwägungen mathematisch zu fassen. <sup>29</sup> Wenn verschiedene Werte (z. B. Verwirklichungsgrade eines Prinzips und eines anderen Prinzips bei Nutzung verschiedener möglicher Auflösungen der Prinzipienkollision) abzuwägen sind, müssen zunächst einmal die Verwirklichungsgrade beziffert werden, was ein gesondertes Problem darstellt, auf das gleich zurückzukommen sein wird. Liegen solche Bezifferungen aber vor, dann kann die Abwägung mit Hilfe der sogenannten „algebraischen Summe“ (im einfachsten Fall bei zwei Summanden:  $x + y - xy$ , bei drei Summanden  $x + y + z - xy - xz - yz + xyz$ ) <sup>30</sup> veranschaulicht werden.

### 3. Problem der Bezifferbarkeit

Praktische Probleme ergeben sich aber bei dem Versuch, Bezifferungen vorzunehmen. So wäre wohl kaum ein Jurist in der Lage, eine abstrakte Ordnung der Relevanz verschiedener Grundrechte vorzunehmen (mit Ausnahme der Menschenwürde als oberstes Prinzip, deren Einordnung durch das Grundgesetz vorgegeben ist), geschweige denn eine solche Rangordnung auch noch intersubjektiv überzeugend vorzunehmen. Wenn eine bloße Ordinalskalierung nicht ausreicht, sondern zur Ermöglichung von Berechnungen eine Intervallskala oder gar eine Verhältnisskala aufgestellt werden soll, <sup>31</sup> werden die Schwierigkeiten noch größer. Hinzu kommt, dass eine bloße Bezifferung der „Wertigkeit“ einzelner Prinzipien nicht ausreicht, sondern es auch und vor allem um Verwirklichungsgrade im Einzelfall geht – so kann z. B. eine Einschränkung des Rechts auf körperliche Unversehrtheit unterschiedlich stark ausfallen (leichte Kratzer vs. Verlust eines Körperteils). Es muss also eine – kaum intersubjektiv überzeugend festzusetzende – Rangstärke eines Prinzips zusätzlich mit einer Verwirklichungsintensität im konkreten Fall multipliziert werden, <sup>32</sup> die ebenfalls durch den Computer bestimmt oder durch den Programmierer im Vorhinein festgesetzt werden müsste. Vor dem Hintergrund dieser Schwierigkeiten können die Einwände, die gegen mit einer Skalierung verbundene mathematische Abwägungskonzepte vorgebracht werden, kaum überraschen. <sup>33</sup> Überdies kann man gegen eine derartige mathematische Berechnung von Abwägungsprozessen zutreffend einwenden, dass sie nicht den tatsächlichen Weg richterlicher Entscheidungsfindung nachzeichnen kann – der Richter beziffert die Abwägungseinflüsse nicht –, sondern allenfalls die Struktur ei-

<sup>28</sup> Effer-Uhe (Fn. 24), Rn. 61 ff., 71 ff.

<sup>29</sup> In ähnliche Richtung deutend, wenn auch auf einer verbalen statt einer formelhaften Ebene verbleibend Krimphove/Niehaus, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.3 Rn. 7.

<sup>30</sup> Vgl. dazu Effer-Uhe (Fn. 24), Rn. 75 f. und Fn. 113.

<sup>31</sup> Vgl. zu den unterschiedlichen Skalenniveaus Döring/Bortz, *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, 5. Aufl. 2016, S. 232 ff.

<sup>32</sup> Vgl. Effer-Uhe (Fn. 24), Rn. 66 mit Fn. 118.

<sup>33</sup> Vgl. z. B. Schlink, *Abwägung im Verfassungsrecht*, 1976, S. 134; Goerlich, *Wertordnung und Grundgesetz*, 1973, S. 136 ff.; Hofmann, *Abwägung im Recht*, 2007, S. 312 f.

ner viel weniger exakten tatsächlichen Abwägungsentscheidung näherungsweise zu illustrieren geeignet ist.<sup>34</sup>

Aber wenn es um die Frage geht, ob eine richterliche Entscheidungsfindung durch eine automatisierte Entscheidungsfindung ersetzt werden könnte, muss das nicht heißen, dass die Entscheidungsfindung des Algorithmus exakt dem richterlichen Weg zur Entscheidung folgt. Möglicherweise könnte die KI sogar – gerade bei Abwägungsentscheidungen – besser abschneiden als ein menschlicher Richter, weil sie einer mathematischen Logik folgt und ihre Ergebnisse berechnet. Entscheidungen menschlicher Entscheider fallen unterschiedlich aus, gerade in Fällen, in denen – wie bei der Strafzumessung oder der Schmerzensgeldbemessung – die Abwägung auf einem Kontinuum erfolgt und es keinen „richtigen“ oder „falschen“ Zielwert gibt. Dabei gibt es einerseits systematische *Verzerrungen*:<sup>35</sup> Richter A verurteilt vielleicht im Durchschnitt zu deutlich höheren Schmerzensgeldern als Richterin B. Daneben fließt aber auch eine *Verrauschung* in die unterschiedlichen Ergebnisse ein:<sup>36</sup> Dass Richter A im Durchschnitt höhere Schmerzensgelder zuspricht, heißt nämlich keineswegs, dass er auch in jedem Einzelfall oberhalb der Entscheidung von Richterin B läge. Zu der grundsätzlich höheren Schmerzensgeldbemessung von Richter A, die einen Teil der Abweichungen zu erklären vermag, kommt also noch eine weitere Beeinflussung der konkreten Ergebnisse durch weitere Faktoren hinzu, die wir als zufällige Schwankung behandeln können, solange wir nichts über die genauen Faktoren wissen, aus denen sich die Abweichung ergibt. Wenn die Automatisierung in der Lage wäre, zumindest einen der beiden Punkte „Verzerrung“ und „Verrauschung“ nennenswert zu reduzieren, wäre das ein denkbare Argument für den KI-Einsatz (dem natürlich andere Argumente entgegengesetzt werden können). Die eben beschriebenen Probleme stehen allerdings einer Lösung anhand expliziter Programmierung entgegen. Es stellt sich also die Frage, ob stattdessen eine implizite Programmierung Abwägungsprozesse jeder Art abbilden und rationaler gestalten könnte.

#### 4. Alternative: Kasuistik und Auswertung von Big Data

Bei ausfüllungsbedürftigen Rechtsbegriffen und General Klauseln wird das Prüfungsschema durch eine ausdifferenzierte Kasuistik strukturiert; es ergibt sich so ein Netz aus Einzelfallentscheidungen, die mal mehr, mal weniger miteinander harmonieren. *Dworkin* stellt sich in einem Gedankenexperiment den fiktiven Richter Herkules mit seinen übermenschlichen Fähigkeiten, übermenschlicher Ausbildung, übermenschlicher Geduld und übermenschlichem Scharfsinn vor.<sup>37</sup> Man könne nun jedem Präzedenzfall (allgemeiner: jeder juristischen Entscheidung) ein bestimmtes Schema von Prinzipien zuordnen, das die Entscheidung des Präzedenzfalls rechtfertigt.<sup>38</sup> Ein übermenschlicher Richter wie Herkules sei theoretisch in der Lage, für neue Fälle die am besten passende Entscheidung zu treffen, die sich am besten in das Gewebe der bisherigen Entscheidungen einfügt.<sup>39</sup> Zwar wird sich dabei praktisch kein völlig fehlerloses

System ergeben – einzelne Entscheidungen werden nicht passen und von Herkules (als Irrtum) außer Acht gelassen werden müssen.<sup>40</sup> Für *Dworkin* ist Herkules nur eine fiktive Figur, ein Gedankenexperiment – ein Richter, den es in der Realität nicht geben kann. Die Möglichkeiten des Deep Learning bringen uns aber vielleicht dem Richter Herkules näher, als es sich *Dworkin* beim Verfassen seines „Taking Rights Seriously“<sup>41</sup> vorgestellt hat. So sind Überlegungen, ob nicht gerade der Computereinsatz Fehler und Verzerrungen, die bei einer teilweise intuitiven Vorgehensweise menschlicher Rechtsanwender vorkommen, reduzieren kann, heute nicht mehr völlig fernliegend.<sup>42</sup> Die Statistik stellt Verfahren zur Verfügung, um innerhalb von Datensätzen Korrelationen festzustellen, insbesondere multiple Regressionsmodelle.<sup>43</sup> Derartige Feststellungen von Korrelationen – zu welchem Anteil ist eine richterliche Entscheidung mit welchem Faktor zu erklären – sind die Stärke von Computerprogrammen. Festgestellt werden können allerdings nur Korrelationen, was nicht gleichbedeutend ist mit Kausalität:<sup>44</sup> Wenn ein Faktor A mit einem Faktor B korreliert, kommen grundsätzlich drei Beziehungen in Betracht – A kann kausal für B sein, B kausal für A, und A und B können auf einer gemeinsamen Drittursache beruhen.<sup>45</sup> Wenn also ein Computerprogramm durch Auswertung einer extrem großen Zahl von Strafurteilen zu dem Ergebnis kommt, dass sich die Strafzumessung mit einem bestimmten (auch vom Zusammenspiel mit anderen Strafzumessungserwägungen abhängigen) Faktor durch ein von Reue getragenes Geständnis erklären lässt, dann bedeutet das nicht, dass tatsächlich naturwissenschaftliche Kausalität vorliegt.<sup>46</sup> Das ergibt sich schon daraus, dass die Erwähnung der Reue im Urteil keineswegs heißt, dass der Täter reuig war, sondern zunächst einmal nur, dass das Gericht das geschrieben hat. Aber selbst, wenn das Gericht daran auch tatsächlich selbst geglaubt hat und mit diesem Glauben vielleicht sogar richtig liegt, ist der tatsächliche Kausalfaktor möglicherweise weniger das Geständnis selbst als das gesamte Verhalten und Auftreten des Angeklagten in der Verhandlung, das sich nicht in eine kompakte sprachliche Formel im schriftlichen Urteil bringen lässt. Der Computer kann also nur eine Korrelation zwischen den Urteilsgründen in seinen Trainingsdaten und dem Tenor seiner Trainingsdaten feststellen, aber keine Kausalbeziehung. Möglicherweise genügt das aber schon, um die Rechtsanwendung durch den Einsatz von KI rationaler zu gestalten.

#### a) Qualität der Trainingsdaten, insbesondere in den Trainingsdaten enthaltene Verzerrungen

Sofern Abwägungsprozesse mit der Hilfe von Computern vorgenommen werden, benötigen sie Input – Bewertungen, die nicht originär vom Computer vorgenommen werden können, sondern zumindest in den Trainingsdaten andere Quellen (d. h. in der Regel: Menschen) haben müssen, deren Bewertungen dann im Rahmen der Auswertung von Big-Data mit Hilfe von Machine Learning fortgeschrieben werden. Dabei sollte die Qualität der Trainingsdaten möglichst hoch sein,<sup>47</sup> wobei ideal eine möglichst homogene, standar-

34 Vgl. Effer-Uhe (Fn. 24), Rn. 79.

35 Kahneman/Sibony/Sunstein, Noise: Was unsere Entscheidungen verzerrt – und wie wir sie verbessern können, S. 79.

36 Kahneman/Sibony/Sunstein (Fn. 35), S. 83 ff.

37 Dworkin (Fn. 22), S. 182.

38 Dworkin (Fn. 22), S. 198 f.

39 Dworkin (Fn. 22), S. 199 ff.

40 Dworkin (Fn. 22), S. 203.

41 Das Werk wurde im englischen Original im Jahr 1977 publiziert.

42 Vgl. z. B. Adrian Rechtstheorie 48 (2017), 77, 97 im Hinblick auf die intuitive Erfassung von Sprache.

43 Vgl. Mittag/Schüller, Statistik, 6. Aufl. 2020, S. 284 ff.

44 Von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3, Rn. 14, 21.

45 Effer-Uhe/Mohnert, Psychologie für Juristen, 2019, Rn. 11.

46 Von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3 Rn. 21.

47 Bues, in: Hartung/Bues/Halbleib (Fn. 18), Rn. 1156, 1203.

disierte Datenstruktur ist,<sup>48</sup> die beispielsweise bei Gerichtsentscheidungen regelmäßig nicht vorliegt. Zweifelhafte Bewertungen können durch die Auswertung großer Datenmengen grundsätzlich für die Zukunft vermieden werden, wenn sie in den Trainingsdaten bloße Ausreißer darstellen: Qualitativ minderwertige Daten, unter anderem solche Ausreißer-Entscheidungen (die *Dworkins* Richter Herkules als Irrtümer unbeachtet lassen würde), können bei der Analyse großer Datenmengen nämlich „herausgerechnet“ werden, solange die Datenqualität im Großen und Ganzen ausreichend hoch ist.<sup>49</sup> In den Trainingsdaten sehr verbreitete Bewertungen werden allerdings umgekehrt durch den Computer perpetuiert, indem die Trainingsdaten sozusagen für die Zukunft fortgeschrieben werden.<sup>50</sup> Ein Problem stellt das insbesondere dann dar, wenn die Trainingsdaten durch Biases<sup>51</sup> verzerrt sind:<sup>52</sup> Wenn beispielsweise in Trainingsdaten zur Strafzumessung eine systematische Verzerrung zulasten von dunkelhäutigen Angeklagten besteht, würde diese Verzerrung durch die KI perpetuiert werden.<sup>53</sup> Um das zu vermeiden, müssten schon die Trainingsdaten von solchen Verzerrungen bereinigt werden. Man könnte beispielsweise alle direkten Hinweise auf die Hautfarbe in den Trainingsdaten einigermaßen problemlos entfernen. Dazu müsste man nicht einmal die entsprechenden Urteile, die solche Verzerrungen enthalten, komplett eliminieren, vielmehr würde es reichen, die Hinweise in den Urteilen, aufgrund derer sich die Hautfarbe ergibt, zu löschen. Der Grund der Verzerrung würde damit für die KI unsichtbar, die Verzerrung (Bias) wäre für sie nur noch als Streuung (Verrauschung, Noise – also ungerichtete Abweichung unklarer Genese) zu erkennen, und gerade eine solche Streuung würde durch die KI-gestützte Abwägung für die Zukunft reduziert.<sup>54</sup>

Allerdings muss man sich darüber im Klaren sein, dass mehr als eine bloße Reduzierung auf diesem Weg kaum zu erreichen ist: Denn während man direkte Hinweise auf die Hautfarbe unproblematisch entfernen könnte, ist es kaum möglich, auch alle anderen Kriterien zu eliminieren, die indirekt mit einer höheren Wahrscheinlichkeit dunkler Hautfarbe assoziiert sind.<sup>55</sup> Geht es zum Beispiel um jemanden, der als Kind immigrierter Eltern in wechselnden Asylbewerberheimen aufgewachsen ist, dann ist damit eine gegenüber dem Bevölkerungsdurchschnitt erhöhte Wahrscheinlichkeit dunkler Hautfarbe verbunden – die KI würde die Diskriminierung also nicht unmittelbar an die Hautfarbe anknüpfend fortschreiben, wohl aber anknüpfend an die Wohnverhältnisse.<sup>56</sup> Derartige Merkmale wird man aber nicht vollständig

aus den Trainingsdaten eliminieren können; denn es kann bei einer Strafzumessungsentscheidung ja auch völlig berechtigte Gründe geben, aufgrund derer sich Besonderheiten in der Kindheit und Jugend des Angeklagten wie beispielsweise die Wohnsituation im Strafmaß auswirken, ohne dass es sich dabei um eine verbotene Diskriminierung handelt, die auf ungewollten Verzerrungen beruht. Eine bloße Reduktion von Verzerrungen in den Trainingsdaten lässt sich also durch die Elimination der direkt diskriminierenden Anknüpfungspunkte aus den Trainingsdaten erreichen – ein vollständiger Ausschluss derartiger Verzerrungen ist dagegen nicht ohne weiteres möglich.

Andererseits könnte die KI dabei helfen, noch unentdeckte unangebrachte Verzerrungen in den Trainingsdaten überhaupt erst zu identifizieren.<sup>57</sup> Der Machine-Learning-Ansatz zeichnet sich gerade dadurch aus, dass er auch unbekannte Zusammenhänge erkennen kann:<sup>58</sup> Wenn man durch die KI eine Strafzumessungsempfehlung für einen bestimmten Fall vorschlagen lässt, dann einen einzigen Punkt in einer Weise abändert, die eigentlich keine Änderung des Strafmaßes mit sich bringen sollte, und das zu einer relevanten Änderung der Strafmaßempfehlung führt, spricht das dafür, dass man eine Verzerrung entdeckt hat.

#### b) Perpetuierung der in den Trainingsdaten schon enthaltenen Wertentscheidungen

Das Programm kann nicht von sich aus auf einen Wertewandel reagieren, sich nicht auf neue Gegebenheiten einstellen, die in den Trainingsdaten noch keine Berücksichtigung finden, ohne dass neue, nicht vom Computer selbst generierte Daten eingepflegt werden. Vielmehr schreibt das künstliche neuronale Netz nur die bisherige Entscheidungspraxis für die Zukunft fort.<sup>59</sup> Aber auch schon die möglichst „perfekte“ Einpassung neuer Entscheidungen in die bisherige Praxis wäre – als Reduktion von Verrauschung – ein Schritt in die richtige Richtung: Wenn sich z.B. die Schmerzensgeldbemessung tatsächlich an einer endlichen Zahl von Faktoren bemessen würde, könnte der Computer jede neue Entscheidung so treffen, dass sie sich möglichst perfekt in das Gefüge der bisherigen Entscheidungen einpasst, also mit allen zu berücksichtigenden Faktoren und deren Berücksichtigung in der Vergangenheit (ihrem Beitrag zur Erklärung der Höhe vergangener Schmerzensgelder) bestmöglich harmonisiert. Der Computer könnte so bei ausreichend großen Datenmengen – lässt man einmal das Problem der für das Training von Deep-Learning-Programmen erforderlichen hohen Rechenleistung gerade bei umfangreichen Trainingsdaten<sup>60</sup> außer Acht – tatsächlich die Rolle von *Dworkins* Richter Herkules übernehmen! Allerdings ist wohl die Annahme *Dworkins*, dass es immer nur ein einziges richtiges Ergebnis gebe, das Herkules „nur“ erkennen müsse, auch mit Machine-Learning-Methoden nicht ohne weiteres allein durch die KI zu erreichen: Menschliche Entscheidungen im Design der Programme (z. B. bei der Frage, welche Art von Machine Learn-

48 Dreyer/Schmees CR 2019, 758, 761.

49 Wiegerling, in: Breidenbach/Glatz (Fn. 1), Kap. 1.3 Rn. 15.

50 Huber/Giesecke, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 19 Rn. 35; Rühl, in: Kaulartz/Braegelmann (Fn. 9), Kap. 14.1 Rn. 15 f.

51 Vgl. zu verschiedenen Urteilsverzerrungen insbesondere Effer-Uhe/Mohnert (Fn. 45) und Schweizer, Kognitive Täuschungen vor Gericht, 2005.

52 Nink, Justiz und Algorithmen, 2021, S. 173; instruktive Beispiele bei Gössl/Yakar, Handreichung Geschlechterneutrale KI, 2023, S. 55 ff., online zugänglich unter [https://www.schleswig-holstein.de/DE/fachinhalte/G/gleichstellung/Downloads/handreichung\\_geschlechterneutrale\\_ki\\_lang.pdf?\\_\\_blob=publicationFile&v=1](https://www.schleswig-holstein.de/DE/fachinhalte/G/gleichstellung/Downloads/handreichung_geschlechterneutrale_ki_lang.pdf?__blob=publicationFile&v=1) (letzter Abruf 15. 9. 2022).

53 Vgl. Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 112; Rühl, in: Kaulartz/Braegelmann (Fn. 9), Kap. 14.1 Rn. 17 f.; Rollberg (Fn. 25), S. 40 ff.; Molavi/Erbguth ITRB 2019, 160, 164.

54 Vgl. Huber/Giesecke, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 19 Rn. 33 zur Reduktion von Ausreißerentscheidungen.

55 Vgl. Greco RW 2020, 29, 39 f.

56 Letztlich tritt das Problem bei allen Merkmalen auf, an die eine Anknüpfung im Antidiskriminierungsrecht als mittelbare Diskriminierung im Sinne des § 3 Abs. 2 AGG angesehen würde – und die insbesondere dann kaum aus den Trainingsdaten eliminiert werden können, wenn es um verdeckte mittelbare Diskriminierung geht, also um Merkmale, die für sich

genommen scheinbar neutral sind, aber eine Personengruppe betreffen, an deren Abgrenzungsmerkmal nach dem AGG eine Ungleichbehandlung nicht ohne rechtfertigenden Grund angeknüpft werden dürfte. Vgl. zu verdeckter mittelbarer Diskriminierung Baumgärtner, in: BeckOGK-AGG, Stand 1. 6. 2023, § 3 Rn. 53, 69 ff.; Thüsing, in: MünchKommBGB, 9. Aufl. 2021, § 3 AGG Rn. 14 ff.

57 Butz u. a. BewHi 2021, 241, 253 f.

58 Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 26, 111.

59 Krimphove/Niehaus, in: Breidenbach/Glatz (Fn. 1), Kap. 9.3 Rn. 52.

60 Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 72.

ing genutzt werden soll<sup>61</sup>) wie auch in der Vorbereitung der Trainingsdaten (z. B. Umgang mit Ausreißer-Ergebnissen, Reduktion der Dimensionalität der Daten mit dem Zweck der Reduktion der erforderlichen Rechenleistung<sup>62</sup>) führen dazu, dass unterschiedliche Programme aufgrund derselben Daten zu – mindestens – leicht unterschiedlichen Ergebnissen führen können: Die Tätigkeit des Programmierers ist ergebnisrelevant.<sup>63</sup> Überdies muss man sich im Klaren sein, dass man zumindest derzeit die Leistungsfähigkeit derartiger Systeme – noch – leicht überschätzen kann: Wenn z. B. eine mit etwa 100 000 Entscheidungen der englischen Ombudsstelle für Finanzdienstleistungen trainierte KI zukünftige Entscheidungen mit einer Sicherheit von 86,6 % voraussagen kann,<sup>64</sup> ist das zwar eine erstaunlich gute Quote – aber eben auch noch weit von 100 % entfernt, und es ist sehr wahrscheinlich, dass das nicht nur an einer Streuung der Ergebnisse der Ombudsstelle liegt,<sup>65</sup> sondern auch an der KI.

Bei neuen Entscheidungen, die gegebenenfalls durch den Computer getroffen würden, stellen sich die verschiedenen Faktoren, aus denen er etwa die „angemessene“ Straf- oder Schmerzensgeldhöhe ableitet, allerdings tatsächlich als Kausalfaktoren dar – Korrelation bedeutet hier tatsächlich Kausalität: Der Computer wirft gerade aufgrund dieser Faktoren eine bestimmte Höhe der Strafe oder des Schmerzensgeldes als „angemessen“ aus. Korrelation in den Trainingsdaten wird in Kausalität übersetzt und so für zukünftige Entscheidungen behandelt, auch wenn sie ursprünglich keineswegs zwingend Kausalität beinhaltete. Bei einfachen neuronalen Netzwerken gibt es also Kausalität, wo ursprünglich in den Trainingsdaten nicht unbedingt Kausalität war. Andererseits ist aber ab einer gewissen Komplexität die Herleitung einer Begründung für eine bestimmte Entscheidung nicht mehr möglich<sup>66</sup> – nachvollzogen werden kann dann zwar, dass bestimmte Faktoren zu einer bestimmten Entscheidung geführt haben, aber kein konkreter Begründungsweg dafür (Black-Box-Verfahren).<sup>67</sup> Mit einem Rechtssystem, das den Anspruch hat, Entscheidungen rational zu begründen, ist das schwerlich vereinbar.<sup>68</sup> Möglicherweise könnten allerdings nicht begründete Entscheidungen durch ein Computerprogramm in einem Stadium vor der Endentscheidung eine Rolle spielen.<sup>69</sup> So könnte ein Richter mit Computerunterstützung eine Plausibilitätsprüfung durchführen, ob die von ihm

ins Auge gefasste Strafhöhe plausibel ist – eine starke Abweichung vom Vorschlag des Computers könnte so für ihn ein Grund sein, die Strafhöhe noch einmal zu überdenken; auch könnte der Computer z. B. den Parteien eines Rechtsstreits ein Ergebnis als Vergleichsvorschlag präsentieren, den sie annehmen können, aber nicht müssen. Auf die damit verbundenen rechtsphilosophischen und verfassungsrechtlichen Probleme soll dieser Beitrag nicht eingehen, weil er sich auf die Frage fokussiert, was überhaupt theoretisch an Automatisierung möglich ist. Und obwohl eine konkrete Entscheidungsbegründung wie durch einen Richter durch die implizit programmierte KI nicht möglich ist, können doch zumindest die Eingabeinformationen anhand ihres Einflusses auf die Entscheidung gewichtet werden.<sup>70</sup>

### c) Quantität der Trainingsdaten

Ein zentrales Problem für die Nutzung künstlicher neuronaler Netze zur Lösung von Rechtsproblemen im Rahmen von Abwägungen liegt darin, dass dafür Input in Form von Daten erforderlich ist – und zwar von ausreichend großen, das heißt sehr großen Datenmengen über Wertungen und unterschiedliche Anlässe/Situationen, denen diese Wertungen entsprechen, in unterschiedlichen Kombinationen.<sup>71</sup> Ist die Menge der Trainingsdaten zu gering, kommt es zu einem „Overfitting“, einer Überanpassung an die Trainingsdaten – das System lernt also konkrete Muster aus den spezifischen Trainingsfällen und weniger verallgemeinerbare Muster.<sup>72</sup> Das stellt in vielen Bereichen ein kaum überwindbares Hindernis dar.<sup>73</sup> Um beim Beispiel der Strafzumessung zu bleiben: Es gibt eine große Anzahl an Urteilen zu Ladendiebstählen, deren Auswertung tatsächlich eine gute Basis dafür darstellen könnte, Korrelationen zwischen einzelnen Strafzumessungserwägungen und der ausgeworfenen Strafe zu erkennen, und jede Woche kommt eine nennenswerte Menge weiterer Urteile hinzu. Das sieht für eher selten von Gerichten angewandte Straftatbestände (z. B. Jagdwilderei) und erst recht im Zuständigkeitsbereich der Schwurgerichte oder allgemeiner bei schwerwiegenden Straftaten schon ganz anders aus – hier wird man nicht auf genügend auszuwertende Urteile kommen, um ähnlich plausibel die „angemessene Strafe“ – z. B. für einen versuchten Totschlag im minder schweren Fall durch einen vermindert Schuldfähigen nach Tatprovokation durch das Opfer – festzulegen (Problem der „unausgewogenen Datensätze“/„imbalanced data sets“ in Bezug auf Fälle, die in der Realität nur selten auftreten<sup>74</sup>). Hinzu kommt, dass nicht nur eine ausreichende Anzahl an Entscheidungen zur Verfügung stehen muss, sondern diese auch in auswertbarer Form, also idealerweise maschinenlesbar und strukturiert, vorliegen müssen.<sup>75</sup> Die Auswahl der anzuwendenden Norm dürfte langfristig auch mit Hilfe impliziter Programmierung daran scheitern, dass für ungewöhnliche Konstellationen und selten von Gerichten angewandte Normen keine ausreichende Datenbasis zur Verfügung stehen wird.

Leistungsfähiger könnte ein KI-System sein, wenn es um die Subsumtion unter bestimmte Tatbestandsmerkmale geht,

61 Vgl. zu Supervised Learning und Unsupervised Learning Kaulartz, in: Kaulartz/Braegelmann (Fn. 9), Kap. 2.2 Rn. 16 ff.

62 Vgl. Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 106, 110.

63 Vgl. Dreyer/Schmees CR 2019, 758, 763; Rollberg (Fn. 25), S. 39.

64 Vgl. Rühl, in: Kaulartz/Braegelmann (Fn. 9), Kap. 14.1 Rn. 10 – die dort genannte Primärquelle [www.case-crunch.com](http://www.case-crunch.com) war zum Zeitpunkt der Vorbereitung dieses Beitrags nicht mehr erreichbar.

65 Vgl. Dreyer/Schmees CR 2019, 758, 760.

66 Von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3 Rn. 21; Stiernerling, in: Kaulartz/Braegelmann (Fn. 9), Kap. 2.1 Rn. 58 ff.; Grupp, in: Hartung/Bues/Halbleib (Fn. 18), Rn. 1100, 1111; Fries, in: Kaulartz/Braegelmann (Fn. 9), Kap. 15.1 Rn. 22 f.; Rollberg ZKM 2020, 208, 209 f.; Nink (Fn. 52), S. 336; vgl. auch Bues, in: in: Hartung/Bues/Halbleib (Fn. 18), Rn. 1156, 1198 mit dem Hinweis auf Versuche, mehr Licht in das Dunkel der Machine-Learning-Algorithmen zu bringen; vgl. zu letzterem Punkt auch Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 128 f. und ausführlich Körner, in: Kaulartz/Braegelmann (Fn. 9), Kap. 2.4.

67 Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 17.

68 Krimphove/Niehaus, in: Breidenbach/Glatz (Fn. 1), Kap. 9.3 Rn. 31; Huber/Giesecke, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 19 Rn. 36 f.; optimistischer Wischmeyer, ebd., § 20 Rn. 56.

69 Ähnlich M. Engel JZ 2014, 1096, 1098; Krimphove/Niehaus, in: Breidenbach/Glatz (Fn. 1), Kap. 9.3 Rn. 50.

70 Krimphove/Niehaus, in: Breidenbach/Glatz (Fn. 1), Kap. 9.3 Rn. 31 f.

71 Krimphove/Niehaus, in: Breidenbach/Glatz (Fn. 1), Kap. 9.3 Rn. 28; Grupp, in: Hartung/Bues/Halbleib (Fn. 18), Rn. 1100, 1110.

72 Dreyer/Schmees CR 2019, 758, 759 f.

73 Vgl. von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3, Rn. 37 f.

74 Niederée/Nejdl, in: Ebers/Heinze/Krügel/Steinrötter (Fn. 11), § 2 Rn. 86.

75 Rühl, in: Kaulartz/Braegelmann (Fn. 9), Kap. 14.1 Rn. 13; Rollberg (Fn. 25), S. 81.

z. B. um die oben genannte Klassifikation von Arsen als Gift: Durch Training anhand einer Vielzahl gelabelter Daten, also von verschiedenen Stoffen, die in den Trainingsdaten mit einer Vielzahl ihrer Eigenschaften enthalten und als Gift oder Nicht-Gift klassifiziert sind, könnte die KI trainiert werden, zu neuen Daten vorherzusagen, ob ein Stoff als Gift eingeordnet werden sollte.<sup>76</sup>

### III. Datafizierung

Das Problem, dass Input in Form von Daten erforderlich ist, ist nicht auf die Auswertung von Big Data beschränkt. Auch für Subsumtionsschlüsse, die keine Abwägungsprozesse erfordern, sind Daten erforderlich.

#### 1. Erforderlichkeit von Ereignisdaten

Relevant für die Rechtsanwendung sind insbesondere sogenannte Ereignisdaten, die ein Ereignis beschreiben, das stattgefunden hat oder möglicherweise stattfinden wird.<sup>77</sup> Mit einem Beispiel von *Yuan*<sup>78</sup>: Für eine Subsumtion unter § 223 StGB könnte einerseits das historische Ereignis „Person A schlägt der Person B mit der Handfläche ins Gesicht“ als Input für den Computer verwendet werden – das wäre das Ereignisdatum, das bloß das Lebensereignis beschreibt; andererseits könnte der Input auch lauten „Person A gibt der Person B eine Ohrfeige“. Die Beschreibung als Ohrfeige setzt eine bereits erfolgte Klassifikation der Ohrfeige als Schlagbewegung voraus, die bei der Datenerfassung geleistet wurde und nicht erst bei der Datenauswertung vom Computer vorgenommen wird.<sup>79</sup> Allerdings wird man sagen müssen, dass zwischen den beiden Formulierungen des Lebensereignisses nur ein gradueller Unterschied besteht. Schon die Formulierung „Person A schlägt der Person B mit der Handfläche ins Gesicht“ beinhaltet nämlich eine Klassifikation der Armbewegung als Schlag (in Abgrenzung zum Beispiel zum Streicheln). Eine vollkommen klassifikationsfreie Übersetzung von Lebensereignissen in natürliche Sprache ist nicht möglich. Sehr wohl kann aber eine Datenerfassung ohne Übersetzung in natürliche Sprache erfolgen. Das ist z. B. bei autonomen Fahrzeugen der Fall, die ihre Umgebung durch Sensoren erfassen und daraus unmittelbar (rechtliche) Schlüsse im Hinblick auf Vorschriften der StVO ziehen.<sup>80</sup> Die autonomen Fahrzeuge generieren mit anderen Worten ihre Ereignisdaten selbst, ein Umweg über die natürliche Sprache ist hier nicht erforderlich. Auch die Verkehrsvorschriften sind bei ihnen aus natürlicher Sprache bereits in Computercode übersetzt, ohne dass der Computer noch sprachliche Äußerungen selbständig inhaltlich erfassen müsste. Allerdings bleibt das Problem, dass nicht ausnahmslos alles in Algorithmen übersetzt ist. Häufige Rechtfertigungsgründe für prima-facie-Verkehrsverstöße (z. B. rechtfertigende Pflichtenkollision, wenn eine Handlung – Lenken – und eine alternative Handlung oder Unterlassung – Weiterfahren oder Bremsen – jeweils Schäden herbeiführen würden) lassen sich erfassen, nicht jedoch ausnahmslos jede denkbare Situation.

#### 2. Keine vollständige Datafizierung

Reine Ereignisdaten stehen in der Rechtsanwendung abgesehen von Spezialfällen wie dem autonomen Auto so gut wie nie zur Verfügung. Das ergibt sich schon daraus, dass eine totale Datafizierung – die Skalier- und Kalkulierbarkeit aller Lebens- und Naturvorgänge – nicht möglich ist und mit jedem erfassten Datum andere, nicht erfasste Daten einhergehen.<sup>81</sup> Es ist in der Regel weder sinnvoll noch auch nur möglich, *sämtliche* Eigenschaften eines Objekts in Form von Daten abzubilden.<sup>82</sup> Zwar liegen enorme „Datenberge“ vor, die noch vor wenigen Jahrzehnten völlig undenkbar gewesen sind und von KI-Systemen ausgewertet werden könnten.<sup>83</sup> Durch die Digitalisierung, den zunehmenden Einsatz von Computertechnologien in der gesamten Gesellschaft, die eine Datafizierung mit sich bringt, vergrößert sich überdies der Bestand an Daten täglich – Informationen über reale Prozesse und Gegebenheiten materialisieren sich in Daten.<sup>84</sup> Diese Datafizierung ist notwendige Bedingung für den Einsatz von Algorithmen, da nur der datafizierte Ausschnitt der realen Welt für Algorithmen zugänglich ist.<sup>85</sup> Aber Daten sind Ergebnisse von Erfassungs- und Sammelprozessen; welche Daten vorliegen und welche nicht, ist damit auch von der Intention beeinflusst, die mit der Messung oder Sammlung der Daten verfolgt wird.<sup>86</sup> Eine Datenerfassung setzt insbesondere eine Festlegung voraus, was es wert ist, als Datum erfasst zu werden.<sup>87</sup> Daten liegen also nicht einfach so in Reinform vor, sondern haben sich aus bestimmten Zwecksetzungen der Datensammlung ergeben.<sup>88</sup> Damit ist Vollständigkeit der relevanten Faktoren in den Daten illusorisch. In den Strafzumessungserwägungen in einem Urteil finden sich z. B. verschiedene Faktoren, die das Gericht bei der Festlegung der Strafe geleitet haben. Das heißt aber noch lange nicht, dass derselbe Fall vom selben Gericht zwingend mit derselben Strafhöhe bedacht werden würde, wenn die im Urteil genannten Strafzumessungserwägungen allesamt identisch sind. So kann z. B. Sympathie oder Antipathie für den jeweiligen Angeklagten auf Seiten des Gerichts die Strafzumessung zumindest geringfügig – und manchmal auch mehr als nur geringfügig – beeinflussen, ohne dass sie im Urteil Erwähnung findet. Wenn eine KI mit Daten aus Urteilen trainiert wird, sind diese Daten also ebenso unvollständig wie die Daten, die sich auf Lebensereignisse beziehen und anhand derer die KI subsumieren soll. Damit kann aber auch die Feststellung von Korrelationen durch die KI negativ beeinflusst werden; je vollständiger zumindest die stärkeren Einflussfaktoren in den Daten erfasst sind, desto geringfügiger werden die Einschränkungen ausfallen.

### IV. Natürliche Sprache und Computersprache

Das autonome Auto umschifft im Vergleich zu einem Computerprogramm, das ganz allgemein – etwa im Rahmen der Vorbereitung gerichtlicher Entscheidungen (oder auch aus

<sup>76</sup> Vgl. *Yuan*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.4 Rn. 10.

<sup>77</sup> *Yuan*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.4 Rn. 12.

<sup>78</sup> *Yuan*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.4 Rn. 12.

<sup>79</sup> *Yuan*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.4 Rn. 12.

<sup>80</sup> *Yuan*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 9.4 Rn. 15.

<sup>81</sup> *Wiegerling*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 1.3 Rn. 15.

<sup>82</sup> *Niederée/Nejdl*, in: *Ebers/Heinze/Krügel/Steinrötter* (Fn. 11), § 2 Rn. 80.

<sup>83</sup> *Von Bünau*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 3 Rn. 8.

<sup>84</sup> *Von Bünau*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 3 Rn. 8.

<sup>85</sup> *Von Bünau*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 3 Rn. 11.

<sup>86</sup> *Wiegerling*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 1.3 Rn. 14.

<sup>87</sup> *Wiegerling*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 1.3 Rn. 14.

<sup>88</sup> *Wiegerling*, in: *Breidenbach/Glatz* (Fn. 1), Kap. 1.3 Rn. 14.

Anwaltssicht: deren Prädiktion) – Recht anwenden soll, zwei wichtige Problemkreise: Einerseits ist für eine umfassende Rechtsanwendung durch eine KI eine Übersetzung von Normen aus natürlicher Sprache in Computercode erforderlich, andererseits regelmäßig die Verarbeitung von Daten, die nicht als Ereignisdaten vorliegen, sondern zumindest zum Teil ebenfalls in natürlicher Sprache vorliegen (z. B. in Form des Tatsachenvortrags einer Partei).

### 1. Erschließung des semantischen Gehalts

Zur Verarbeitung von natürlicher Sprache als dominantem Transportmedium für Informationen im Recht wäre daher grundsätzlich eine Erschließung und Kontextualisierung des semantischen Gehalts von Texten durch den Computer erforderlich.<sup>89</sup> Dabei geht es heutzutage nicht mehr so sehr um den Versuch eines echten „Verstehens“ durch den Computer – die Frage ist vielmehr, inwieweit eine Textanalyse durch künstliche neuronale Netze im Wege der Auswertung großer Datenmengen erfolgversprechend ist.<sup>90</sup> KI-Systeme replizieren im Wesentlichen das Verhalten von eingespeisten Datenmustern,<sup>91</sup> und das ist theoretisch auch mit Daten in Form von Texten in natürlicher Sprache möglich. Allerdings ist das Problem der Überführung einer in einem Text kodierten Bedeutung in eine „maschinell verarbeitbare Repräsentation“ momentan noch weitgehend ungelöst, was für Bereiche, in denen – wie im Recht – Informationen in Textform vorliegen, aber präzise erfasst werden müssen, Probleme bereitet.<sup>92</sup> Im Rahmen des sogenannten Natural Language Processing (NLP) ist eine Textkategorisierung noch weitgehend unproblematisch möglich, eine strukturierte Informationsextraktion dagegen derzeit nur auf dem Weg maschinellen Lernens und mit umfangreichen Trainingsdaten für jedes Aussageschema denkbar.<sup>93</sup> Eine vollständige Formalisierung natürlicher Sprache ist dagegen unmöglich.<sup>94</sup> Das ergibt sich schon daraus, dass formale Sprachen vollständig definiert und in ihrer Interpretation eindeutig sein müssen – also insbesondere die Bedeutung des sprachlichen Ausdrucks nicht kontextabhängig sein darf –, während die natürliche und damit auch die juristische Sprache nicht eindeutig und präzise, sondern kontextabhängig sind, sich derselbe Ausdruck also abhängig vom Kontext auf Unterschiedliches beziehen kann.<sup>95</sup>

### 2. Simulation von juristischem Denken

Eine semantische Entschlüsselung natürlicher Sprache ist aber möglicherweise gar nicht nötig, um zu einer Rechtsanwendung durch Computer zu gelangen. Das wurde besonders klar von Axel Adrian herausgearbeitet. Sein Gedankengang ist der Folgende: Ausgehend von der Frage, ob eine Maschine juristisches Denken oder juristische Sprache so umfassend simulieren kann, dass sie einen Rechtsfall insgesamt entscheiden kann,<sup>96</sup> stellt er zunächst fest, dass die

natürliche Sprache assoziativ und damit individuell erlernt wird und nicht über logische Strukturen mit Phänomenen der Realität verbunden ist.<sup>97</sup> Die Idee einer semantischen Bedeutung der natürlichen Sprache sei also nur eine Illusion.<sup>98</sup> Ebenso werde das Recht assoziativ erlernt: Professoren führten ihren Studierenden komplexe Ideen und Strukturen vor und ermöglichten ihnen so, juristisches Denken nachzuvollziehen – dadurch habe aber jeder Student sein eigenes einzigartiges Bild davon, wie juristisches Denken funktioniert.<sup>99</sup> Damit könne es aber nicht mehr darum gehen, eine zutreffende semantische Bedeutung der natürlichen Sprache zu ermitteln, vielmehr werde nur der Austausch semantischer Bedeutung *simuliert* – das aber könnten auch Maschinen tun.<sup>100</sup> Dass Austausch semantischer Bedeutung nur eine Illusion sei, müsse aber nicht bedeuten, dass keinerlei Bedeutung ausgetauscht wird; denn – so nimmt Adrian plausibel an – durch eine gigantische Zahl an Sprachpraktiken und Reaktionen entstünden Muster holistischer Netze der an der Kommunikation Beteiligten, die sich ausreichend ähneln, so dass zumindest eine Verständigung darüber möglich ist, ob man sich versteht oder nicht.<sup>101</sup>

Der praktische Weg führt über das oben geschilderte maschinelle Lernen künstlicher neuronaler Netze.<sup>102</sup> Denn zu einer spezifischen Rechtsfrage ließen sich neben dem Wortlaut des Gesetzes auch Texte aus Urteilen, Fachliteratur und weiteren Quellen erschließen, die diese Rechtsfrage betreffen, daher ließe sich auch überprüfen, welche Worte und Textbausteine in gleicher oder ähnlicher Form häufig oder selten verwendet werden.<sup>103</sup> Dementsprechend könne man auch ohne Kenntnis einer semantischen Bedeutung wissen, wer welches Zeichen verwendet und wie häufig welches Zeichen in Bezug auf andere Zeichen als Reaktion verwendet wurde.<sup>104</sup> Mit anderen Worten: Es ist theoretisch möglich, auf diese Art und Weise Entscheidungen (nicht: Entscheidungsbegründungen!) maschinell zu erstellen, die in ihren Ergebnissen hinreichend nah an menschlichen juristischen Entscheidungen liegen, so dass man von einer geglückten Simulation einer juristischen Entscheidung sprechen kann. Da dieses Modell nur auf Sprachpraktiken beruht und die bestehenden Praktiken weiterführt, umgeht es jegliche Festlegung darüber, wie Auslegung, Rechtsfortbildung, Subsumtion etc. funktionieren – das Modell ist gegenüber jeglicher methodischen Festlegung indifferent.<sup>105</sup> Das Problem, dass grundsätzlich eine vollständige Formalisierung juristischer Normen, eine Überführung in Computersprache erforderlich ist, um eine automatische Subsumtion zu ermöglichen,<sup>106</sup> kann Adrian so umschiffen. Er erkaufte sich das, indem er sich damit begnügt, den Computer nicht denselben oder auch nur einen ähnlichen Entscheidungsweg verfolgen

89 Breidenbach/Glatz, in: Breidenbach/Glatz (Fn. 1), Kap. 1.1 Rn. 17.

90 Breidenbach/Glatz, in: Breidenbach/Glatz (Fn. 1), Kap. 1.1 Rn. 17.

91 Von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3 Rn. 6.

92 Von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3 Rn. 18, 25; Rollberg (Fn. 25), S. 56 ff.

93 Von Bünau, in: Breidenbach/Glatz (Fn. 1), Kap. 3 Rn. 26 ff.

94 Kotsoglou JZ 2014, 451 (452).

95 Kotsoglou JZ 2014, 451 (453 f.); Rollberg (Fn. 25), S. 58; Nink (Fn. 52), S. 211.

96 Adrian Rechtsstheorie 48 (2017), 77, 80; ähnlich Grupp, in: Hartung/Bues/Halbleib (Fn. 18), Rn. 1100, 1115, der ausführt, dass die Automation menschliche Denkprozesse gar nicht nachbilden müsse, solange das Ergebnis juristischer Überprüfung standhält.

97 Adrian Rechtsstheorie 48 (2017), 77, 85.

98 Adrian Rechtsstheorie 48 (2017), 77, 91.

99 Adrian Rechtsstheorie 48 (2017), 77, 88.

100 Adrian Rechtsstheorie 48 (2017), 77, 91.

101 Adrian Rechtsstheorie 48 (2017), 77, 92.

102 Adrian Rechtsstheorie 48 (2017), 77, 111 f. verweist insoweit auf das Instrumentarium der Fuzzy-Logik, also der Theorie unscharfer Mengen, vgl. dazu bereits oben II. sowie Krimphove/Niebaus, in: Breidenbach/Glatz (Fn. 1), Kap. 9.3 Rn. 7 ff. und allgemein Joerden, Logik im Recht, 3. Aufl. 2018, S. 325 ff.

103 Adrian Rechtsstheorie 48 (2017), 77, 101 f.

104 Adrian Rechtsstheorie 48 (2017), 77, 102.

105 Vgl. Adrian Rechtsstheorie 48 (2017), 77, 110; vgl. zu den Problemen, die eine Überführung der Methodenlehre in Computercode im Hinblick auf die Auslegungskanonens mit sich bringt (z. B. nach h. M. kein eindeutiges Rangverhältnis der verschiedenen Auslegungskanonens) exemplarisch Yuan, in: Breidenbach/Glatz (Fn. 1), Kap. 9.4 Rn. 5.

106 M. Engel JZ 2014, 1096, 1097.

zu lassen, den auch ein menschlicher Rechtsanwender geht, sondern nur möglichst ähnliche Ergebnisse wie bei menschlicher Rechtsanwendung zu erwarten. In der Tat wäre das vollkommen ausreichend, wenn es nur um die Frage der Prädiktion juristischer Entscheidungen ginge.<sup>107</sup> Auch wäre eine Gesellschaft zumindest denkbar, die sich sogar für das Erzeugen verbindlicher Entscheidungen mit einer möglichst perfekten Einpassung in die bisherige Entscheidungspraxis zufrieden gibt. Eine solche Gesellschaft könnte theoretisch die Rechtsanwendung auf Computer verlagern. Das würde sogar für Fragen der Beweiswürdigung gelten:<sup>108</sup> Auch im Hinblick auf Beweismittel könnte eine KI analysieren, in welchen Fällen (aufgrund welcher Zeichenketten wie z. B. „wird bestritten“ im Beklagtenschriftsatz) es zu einer Beweisaufnahme kommt, auf welche Aussagen von Zeugen, auf welche Inhalte von Urkunden und Sachverständigengutachten etc. in welchem Anteil der Fälle wie reagiert wurde, und das könnte Basis für eine Entscheidung der KI sein. Nicht zugänglich ist einem solchen Modell freilich der Bereich der Rechtsfortbildung – Analogieschlüsse beispielsweise, die in den Trainingsdaten nicht schon auffindbar sind, wird die KI nicht ziehen können.<sup>109</sup> Analogieschlüsse, bei denen der menschliche Gedankengang nachvollzogen wird, sind ohne semantisches Verständnis nicht denkbar, und eine Simulation menschlicher Analogieschlüsse kann durch eine lernende KI nur gelingen, wenn der konkrete Analogieschluss bereits in den Trainingsdaten zu finden ist.

Dass es sich bei der Simulation menschlicher Entscheidungen durch eine KI bislang und auch auf absehbare Zeit nur um ein theoretisches Modell und nicht schon um einen praktisch gangbaren Weg handelt, ergibt sich bereits daraus, dass eine ausreichend Datenbasis erforderlich ist, aus der sich hinreichend genaue Gesetzmäßigkeiten ableiten lassen. Mit den Worten der Statistik gesprochen: Die vom Computer ermittelten Folgen von Zeichenketten in den Ausgangsdaten und den Ergebnissen dürfen nicht auf Zufall beruhen, die Abweichungen zwischen Fällen mit und Fällen ohne bestimmte Zeichenketten müssen „statistisch signifikant“ sein – in der Statistik geht man meist von statistischer Signifikanz aus, wenn die Wahrscheinlichkeit eines zufälligen Ergebnisses unter 5 % liegt – diese Hürde kann aber durchaus auch höher oder niedriger angesetzt werden (z. B. auf 1 % oder auf 10 %).<sup>110</sup> Die KI muss sich nicht mit abstrakten Signifikanzschwellen befassen, sondern kann sich damit zufrieden geben, dass das künstliche neuronale Netz die Ergebnisse bestmöglich erklärt, abbildet und für die Zukunft fort schreibt – das ist durchaus schon jenseits der in der Statistik üblicherweise genannten Schwellenwerte möglich. Statistische Signifikanz bedeutet nicht zwingend einen starken Effekt: Auch ein sehr geringer Effekt kann statistisch relevant werden, wenn die Datengrundlage ausreichend groß ist, um die Wahrscheinlichkeit eines zufälligen Ergebnisses unter die Signifikanzschwelle abzusenken. Letztlich geht es in der Sache um sogenannte multiple Regressionsanalysen. Belastbare Ergebnisse werden dabei nur bei ausreichend großen Datenmengen erzielt,<sup>111</sup> und daran fehlt es bislang in vielen Berei-

chen – und wird es in manchen Bereichen auch dauerhaft fehlen. So gibt es in unserem Rechtssystem Normen, die so gut wie nie forensisch relevant werden (man denke nur an die Regelungen des BGB zu Bienenschwärmen). Aber auch bei Normen, die mehr praktische Anwendung finden, gibt es so viele Einflussfaktoren, dass es keine Selbstverständlichkeit ist, dass genügend für das Training der KI geeignete Daten produziert werden.

## V. Ausblick: Primary Rules, Secondary Rules und die Rule of Recognition

Der Beitrag hat sich im Wesentlichen mit der Frage befasst, inwiefern bestehende Regeln (Gesetze oder untergesetzliches Recht) in Computercode überführt werden und Abwägungsprozesse durch Operationen von KI-Systemen ersetzt werden können. Es wurde insoweit also von bestehenden Regeln ausgegangen, die vom zuständigen Normgeber bereits gesetzt sind. Aber selbstverständlich lassen sich die Ausführungen auf einer Metaebene auch auf die Prüfung übertragen, ob eine Norm Teil des geltenden Rechts ist. So könnte eine entsprechend programmierte KI aus Ereignisdaten aus dem Gesetzgebungsverfahren (zur Einbringung des Gesetzesentwurfs, zu den Lesungen, zu Änderungsanträgen, zu Abstimmungen in Bundestag und Bundesrat etc.) die Prüfung vornehmen, ob eine Norm überhaupt geltendes Recht geworden ist.<sup>112</sup> *H. L. A. Hart* unterscheidet zwischen „Primary Rules“ als Regeln, die Pflichten auferlegen, und „Secondary Rules“ als Regeln, die Kompetenzen übertragen.<sup>113</sup> So wäre z. B. § 242 StGB als verhaltensregelnde Norm eine Primary Rule, gleichzeitig überträgt sie allerdings auch – im Zusammenwirken mit weiteren Normen beispielsweise der StPO – Kompetenzen an den Rechtsstab, der ermächtigt wird, unter bestimmten Voraussetzungen eine Strafe festzulegen und zu vollstrecken, ist insoweit also auch Secondary Rule. Die Regelungen des Grundgesetzes zur Gesetzgebung sind Secondary Rules, die Kompetenzen zur Gesetzgebung vermitteln, ebenso aber z. B. auch die Regeln des BGB zum Kaufvertrag, die den Vertragsparteien die Kompetenz vermitteln, innerhalb eines gewissen Spielraums (z. B. §§ 134, 138 BGB) die Rechtsfolgen „Verpflichtung zur Übergabe und Übereignung einer bestimmten Kaufsache und zur Zahlung eines bestimmten Kaufpreises“ privatautonom zu setzen. Neben diesen Arten von Normen sei aber auch noch eine „Rule of Recognition“ erforderlich, anhand derer überprüft werden kann, ob eine Norm zum geltenden Recht gehört oder nicht.<sup>114</sup> Auch die Übertragung dieser Prüfung auf eine KI könnte theoretisch erwogen werden, jedenfalls, wenn man sich erst einmal einig ist, wie diese Rule of Recognition eigentlich aussehen soll – man denke nur an die vielfältigen Überlegungen zur Letztbegründung von Normen auf naturrechtlicher<sup>115</sup> oder positivistischer<sup>116</sup> Grundlage, an *Kelsen*

<sup>107</sup> *Rühl*, in: *Kaulartz/Braegelmann* (Fn. 9), Kap. 14.1 Rn. 6; kritisch (allerdings auch nicht auf die bloße Prädiktion bezogen) *Nink* (Fn. 52), S. 205 ff.

<sup>108</sup> Grundsätzlich zweifelnd *M. Engel* JZ 2014, 1096, 1100.

<sup>109</sup> Ähnlich auch für eine explizite Programmierung *M. Engel* JZ 2014, 1096, 1098; in der Tendenz anders *Greco* RW 2020, 29, 37 f.

<sup>110</sup> *Blasius/Thiessen*, Argumentieren mit Statistik, 2021, S. 18, 118 ff.

<sup>111</sup> *Adrian* Rechtslehre 48 (2017), 77, 111 f.

<sup>112</sup> Die Spezialfälle des Gewohnheitsrechts und des Richterrechts, die ja auch rechtstheoretisch in den Einzelheiten nicht unumstritten sind (vgl. *Wank*, Juristische Methodenlehre, 2020, § 5 Rn. 28 ff. m. w. N.), sollen hier außen vor bleiben. Vgl. dazu z. B. *Nink* (Fn. 52), S. 229 ff.

<sup>113</sup> *Hart*, The Concept of Law, 2. Aufl. 1994 (hier zitiert nach der Ausgabe 1997), S. 81.

<sup>114</sup> *Hart* (Fn. 113), S. 105 f.

<sup>115</sup> Vgl. dazu *Ellscheid*, in: *Hassemer/Neumann/Saliger* (Hrsg.), Einführung in die Rechtsphilosophie und Rechtstheorie der Gegenwart, 9. Aufl. 2016, S. 143 ff.

<sup>116</sup> Dazu *Horn*, Einführung in die Rechtswissenschaft und Rechtsphilosophie, 6. Aufl. 2016, Rn. 102 ff.

Grundnorm<sup>117</sup> und das Münchhausen-Trilemma<sup>118</sup>. Die sich bei der Übertragung der Prüfung einer Rule of Recognition an eine KI stellenden Herausforderungen sind im Grundsatz mit denjenigen bei der Anwendung von Primary und Secondary Rules vergleichbar – praktisch stellen sich dahingehende Fragen aber schon deshalb nicht, weil dazu erst einmal Einigkeit über den Inhalt einer Rule of Recognition erreicht werden müsste.

## VI. Fazit

Die intensiv diskutierte Frage, ob alles, was in technischer Hinsicht automatisierbar ist, auch automatisiert werden sollte<sup>119</sup> und inwieweit diese Automatisierung mit rechtlichen Vorgaben z. B. des Grundgesetzes vereinbar ist,<sup>120</sup> wurde in diesem Beitrag nicht thematisiert. Im Fokus stand die Frage, inwieweit eine Automatisierung der Rechtsanwendung überhaupt möglich ist. Regeln können grundsätzlich unproblematisch im Wege einer expliziten Programmierung in Computercode überführt werden. Das allein ermöglicht aber noch keine automatisierte Rechtsanwendung. Bei der Frage, ob ein Merkmal tatsächlich gegeben ist, ist nämlich grundsätzlich der Mensch erforderlich, wenn sich nicht schon aus

einer dem Programm zur Verfügung stehenden Datenbank ergibt, ob das Merkmal erfüllt ist oder nicht. Erforderliche Abwägungen, die nicht nur im Rahmen von Prinzipienabwägungen und bei unbestimmten, Rechtsbegriffen auftauchen, können zwar grundsätzlich mit dem Instrumentarium der Fuzzy-Logik formalisiert werden. Voraussetzung ist aber stets das Vorliegen von Bezifferungen für die Wertungen, an denen sich die Abwägung ausrichten soll – und die Festlegung intersubjektiv überzeugender Skalen ist kaum möglich.

Können selbstlernende Programme durch die Auswertung großer Datenmengen Abhilfe schaffen? Im Grundsatz ist das möglich, wenn man bereit ist, auf eine dem menschlichen Richter vergleichbare Entscheidungsfindung zu verzichten und sich mit anderweitig erreichten, aber vergleichbaren Ergebnissen zufrieden zu geben. Allerdings sind die dafür erforderlichen Datenmengen für eine Vielzahl seltener angewandter Rechtsnormen weder qualitativ noch quantitativ in Sicht.

Zumindest ein Teil der erforderlichen Daten wird zudem nur in Form natürlicher Sprache vorliegen – so voraussichtlich auch in Zukunft die Sachverhaltsschilderung. Von einem echten Textverständnis ist der Computer noch weit entfernt; allenfalls kann eine mehr oder weniger überzeugende Simulation menschlichen Denkens durch maschinelles Lernen gelingen. Die Probleme, die für eine umfassende Automatisierbarkeit der Rechtsanwendung – wenn sie denn überhaupt gewollt wäre – zu überwinden sind, sind nach wie vor enorm. Vieles ist zwar theoretisch automatisierbar, allerdings muss man bereit sein, dafür erhebliche Abstriche in der Vergleichbarkeit zwischen menschlicher und maschineller Entscheidungsfindung hinzunehmen. Bei anderen Punkten ist – beispielsweise wegen der nicht ausreichenden Datafizierung – noch nicht einmal eine theoretische Automatisierbarkeit absehbar.

<sup>117</sup> Kelsen, *Reine Rechtslehre*, 1. Aufl. 1934 (hier zitiert nach dem Nachdruck 1994), S. 62 ff.

<sup>118</sup> Einführend zum Münchhausen-Trilemma in der Erkenntnistheorie Schmidt-Salomon, in: *Aufklärung und Kritik*, Sonderheft 5/2001 zu Hans Albert, S. 42 ff., online zugänglich unter <http://www.schmidt-salomon.de/muench.htm> (letzter Abruf 15.9.2022). Das Problem, dass eine Letztbegründung immer nur auf einem von drei unbefriedigenden Wegen (infiniter Regress, Zirkelschluss, Regressabbruch) versucht werden kann, ist für die Rechtswissenschaft bei der Letztbegründung von Normen virulent.

<sup>119</sup> Dazu statt vieler Greco RW 2020, 29, 46 ff. m. w. N.

<sup>120</sup> Dazu z. B. Nink (Fn. 52), S. 260 ff.; Rühl, in: *Kaulartz/Braegelmann* (Fn. 9), Kap. 14.1 Rn. 21 ff.