

## Rechtspersönlichkeit für Künstliche Intelligenz?

CAROLIN KEMPER

### Zitiervorschlag

KEMPER, Rechtspersönlichkeit für KI?, in: cognitio 2018/1.

URL: [cognitio-zeitschrift.ch/2018-1/Kemper](http://cognitio-zeitschrift.ch/2018-1/Kemper)

DOI: <https://doi.org/10.5281/zenodo.1928171>

ISSN: 2624-8417



## Rechtspersönlichkeit für Künstliche Intelligenz?

CAROLIN KEMPER, Ref. jur.<sup>1</sup>

*Seit LAWRENCE B. SOLUMs Aufsatz Legal Personhood for Artificial Intelligences von 1992 wird rege diskutiert, ob bzw. ab wann digitale Agenten Rechtspersönlichkeit erhalten sollen. Die entscheidende Frage ist, wann digitale Agenten zur Teilnahme am Rechtsverkehr fähig sind. In diesem Beitrag soll untersucht werden, ob Agenten, die auf der Technologie der Künstlichen Neuronalen Netzen basieren, "autonom genug" sind, um als Rechtsperson handeln zu können.*

### Inhaltsübersicht

I. Einführung	1
II. Künstliche Neuronale Netze	3
A. Der Aufbau von KNN	3
B. Der Lernprozess	4
C. Deep Learning	5
III. Zur Statusfrage von digitalen Agenten	6
A. Die konventionelle Behandlung von digitalen Agenten	6

B. Volle Rechtsfähigkeit	7
C. Teilrechtsfähigkeit	8
D. Alternative Lösungen	10
IV. Die Konstruktion von rechtlicher Handlungsfähigkeit	10
A. Agenten aus informationstechnischer und philosophischer Sicht	10
B. Handlungs- als Kommunikationsfähigkeit	11
C. Autonome digitale Agenten	11
D. Intentionalität digitaler Agenten	12
V. Ausblick	12

### I. Einführung

Im Zuge der zunehmenden Leistungs- und Lernfähigkeit informationstechnischer Systeme wird diskutiert, ob bzw. ab wann diese sich für einen eigenständigen rechtlichen Status qualifizieren. Bislang geschah dies hauptsächlich in abstrakter Form, indem Kriterien für einen eigenen Rechtsstatus entwickelt wurden.<sup>2</sup> Nunmehr sollen die bereits erarbeiteten Konzepte und Ansätze mit der aktuellen Technologie der Künstlichen Intelligenz abgeglichen werden.

<sup>1</sup> Die Autorin dankt Steve Blum von der Camelot ITLab GmbH für die Unterstützung beim Verständnis der technischen Grundlagen Künstlicher Neuronaler Netze, und für die Ermutigungen, ohne die dieser Beitrag nicht entstanden wäre.

<sup>2</sup> Vgl. statt vieler: TEUBNER GUNTHER, Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten, in: AcP 218 (2018), S. 155 ff. (im Übrigen siehe unten ab III.).

Künstliche Intelligenz ist eines der Schlagworte der Digitalisierung. Einerseits wird es besonders häufig verwendet, andererseits mutet es vage an und erinnert etwas an Science-Fiction. Man denkt an *SkyNet* oder gar an einen Androiden wie *Lieutenant Commander Data* aus *Star Trek: The Next Generation*.

Im Allgemeinen wird von *starker Künstlicher Intelligenz* (KI) oder auch *Artificial General Intelligence* gesprochen, wenn eine menschenähnliche Intelligenz gemeint ist.<sup>3</sup> Bei *schwacher Künstlicher Intelligenz* bzw. *Artificial Narrow Intelligence* werden nur einzelne Aspekte von Intelligenz umgesetzt.<sup>4</sup> Ein Teilgebiet der schwachen Künstlichen Intelligenz ist das maschinelle Lernen (*machine learning*), das sich mit lernfähigen Algorithmen auseinandersetzt. Künstliche Neuronale Netze (KNN) sind eine häufig eingesetzte Methode des maschinellen Lernens<sup>5</sup> und finden häufig in KI-Anwendungen Verwendung.

KNN sind lern- und anpassungsfähig und setzen damit einen Aspekt menschlicher Intelligenz um. Ihrem Namen nach erinnern sie zwar an das menschliche Gehirn mit seinen Neuronen und Synapsen, verwenden jedoch mathematische Annäherungsverfahren, um Fehler zu reduzieren, und «lernen» hierdurch, ihre Aufgaben immer genauer zu erledigen.<sup>6</sup> Sie werden u.a. zur Bilderkennung wie insb. der Schrift- und Gesichtserkennung sowie zur natürlichen Sprachverarbeitung (*Natural Language Processing*) genutzt,

können aber auch Vorhersagen treffen (*Predictive Analytics*). Beispielsweise können KNN aus Daten über Lagerbestände und Verkaufszahlen die künftigen Vorratsmengen vorhersagen, die dann automatisiert bestellt werden können, ohne dass es menschlicher Intervention bedarf.<sup>7</sup> Auch beim *Kreditscoring* ist der Einsatz von KNN zur Vorhersage von Kreditausfällen möglich. Gleichzeitig ist infolge des hohen Abstraktionsgrades ihre Funktionsweise nicht nachvollziehbar, in welchem Masse Entscheidungsfaktoren zu einem Ergebnis beigetragen haben. Aus diesem Grund sind Finanztechnologien (sog. *FinTechs*), insb. der Aktienhandel auf der Grundlage von Kursvorhersagen, Gegenstand vieler Debatten, da diese den Finanzmarkt destabilisieren können.<sup>8</sup> Aber auch bei der Kreditvergabe kann die Intransparenz der zugrundeliegenden informationstechnischen Methoden zur Diskriminierung verschiedener Personengruppen führen.<sup>9</sup>

Aus der oft hohen Genauigkeit und Anpassungsfähigkeit, aber auch der Intransparenz sowie mangelnden Interpretierbarkeit der Ergebnisse von KNN ergeben sich neuartige Probleme und Risiken, die das Recht derzeit noch nicht adressiert. Vor diesem Hintergrund stellt sich die Statusfrage von IT-Systemen, die aufgrund ihrer Lernfähigkeit

<sup>3</sup> RUSSELL STUART/NORVIG PETER, *Künstliche Intelligenz*, 3. Aufl., München 2012, S. 51.

<sup>4</sup> GURKAYNAK GONENC/YILMAZ ILAY/HAKSEVER GUNES, *Stifling artificial intelligence: Human perils*, in: CLSR 32, 2016, S. 749 ff., S. 751 f.

<sup>5</sup> Andere Methoden sind z.B. Entscheidungsbäume, Bayes-Netze, usw., siehe ALPAYDIN ETHEM, *Machine Learning, The New AI*, Cambridge MA, 2016.

<sup>6</sup> Siehe ETZIONI OREN, [Deep Learning Isn't a Dangerous Magic Genie, It's Just Math](#), wired 15.06.16; TRINKWALDER ANDREA, *Die Mathematik neuronaler Netze: einfache Mechanismen, komplexe Konstruktion*, c't 2016, S. 130 ff.

<sup>7</sup> Zum Einsatz von maschinellem Lernen zur besseren Planung im Bereich frischer Nahrungsmittel, vgl. GLATZEL CHRISTOPH et.al.,

[The secret to smarter fresh-food replenishment? Machine learning](#), 2016.

<sup>8</sup> Vgl. zum aktuellen «FinTech-Aktionsplan» der EU-Kommission: MÖSLEIN FLORIAN/OMLOR SEBASTIAN, *Die europäische Agenda für innovative Finanztechnologien (FinTech)*, in: BKR 2018, S. 236 ff.; siehe zur Kritik am Einsatz maschinellen Lernens im Finanzmarkt KNIGHT, [Weltwirtschaftsforum, KI könnte Finanzsystem destabilisieren](#), Technology Review 04.09.2018.

<sup>9</sup> KNIGHT WILL, [Das Problem der diskriminierenden Algorithmen](#), Technology Review 25.07.2017; vor einem Jahr wurde vor diesem Hintergrund sogar ein digitales Antidiskriminierungsgesetz vorgeschlagen: KREMPL STEFAN, [Algorithmen und Scoring: Maas fordert digitales Anti-Diskriminierungsgesetz](#), Heise 03.07.2017. Dem wurde mittlerweile durch Art. 22 der EU-Datenschutz-Grundverordnung Abhilfe geschaffen.

Aufgaben selbständig wahrnehmen können (sog. *digitale Agenten*).<sup>10</sup>

## II. Künstliche Neuronale Netze

Die prinzipielle Funktionsweise von KNN wird durch folgendes Beispiel demonstriert: Das lokale Wetter soll vorhergesagt werden. Jeden Tag könnte man hierzu das Wetter beobachten und feststellen, ob es sonnig, regnerisch oder stürmisch ist und welche Temperatur herrscht. Eine zuverlässige Vorhersage für den zweiten Tag der Versuchsreihe kann aber noch nicht getroffen werden; als Entscheidungsgrundlage würde lediglich das Wetter des ersten Tages dienen. Nur über einen langen Zeitraum hinweg ist es möglich, einigermaßen zuverlässig auch Wetterumschwünge vorherzusagen. Der Grund hierfür ist, dass sich viele zufallsbasierte Faktoren auf das Wetter auswirken und es keine klar erkennbaren Bedingungen gibt: wenn am Vortag viele Wolken über den Himmel ziehen, bedeutet dies nicht per se, dass es regnen wird.

Dieses Beispiel verdeutlicht den Anwendungsbereich von KNN: sie erzielen gute Ergebnisse im Umfeld scheinbar zufälliger Einflussfaktoren – also, wenn Konditionalitäten nicht vorhanden oder erkennbar sind. Durch einen Lernprozess nähern sie sich der optimalen Lösung eines Problems immer

weiter an. Je komplexer das Problem ist, d.h. je mehr Faktoren für die Ergebnisfindung einbezogen werden, desto umfangreicher und sorgfältiger muss die Datenbasis ausgewählt sein, damit das Training optimale Ergebnisse liefert.

### A. Der Aufbau von KNN<sup>11</sup>

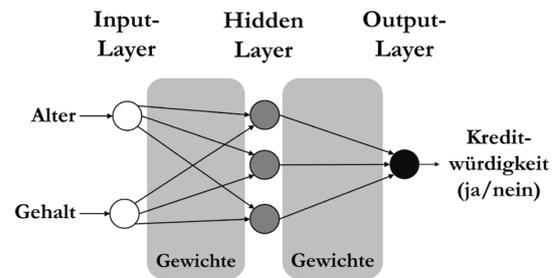


Abb. 1: Vereinfachte Darstellung des Aufbaus und der Funktionsweise eines KNN am Beispiel von *Kreditscoring*.

Ein KNN arbeitet in drei Schichten (*layers*) (siehe Abb. 1): Die erste Schicht besteht aus Input-Neuronen, die Daten, mit denen das KNN arbeiten soll, empfangen. Diese Daten sind in der Regel Faktoren, die das Ergebnis bzw. den Output beeinflussen. Beim *Kreditscoring* wären beispielsweise Alter, Wohnort, Gehalt und Beruf des Kreditsuchenden relevante Faktoren,<sup>12</sup> die in die *Input-Layer*

<sup>10</sup> Im Folgenden beschreibt der Begriff des digitalen Agenten ein Programm, das in rechtserheblicher und selbstständiger Weise Aufgaben wahrnehmen und Entscheidungen treffen kann. Dies umfasst Software-Agenten im engeren Sinne, aber auch Roboter oder Chatbots, sofern Künstliche Neuronale Netze die technische Grundlage ihrer Entscheidungen bilden.

<sup>11</sup> Die folgende – überblicksartig gehaltene – Darstellung von KNN orientiert sich an den Ausführungen von RUSSELL/NORVIG (Fn. 3), S. 845 ff., aber auch ALPAYDIN (Fn. 5), S. 85 ff.; ERTEL WOLFGANG, Grundkurs Künstliche Intelligenz: Eine praxis-orientierte Einführung, 3. Aufl., Wiesbaden 2013, S. 247 ff.; GOODFELLOW IAN/BENGIO YOSHUA/COURVILLE AARON, Deep Learning, Cambridge MA 2016, S. 12 ff., S. 164 ff.; KRUSE RUDOLF et.al., Computational Intelligence, Eine methodische Einführung in Künstliche Neuronale Netze, Evolutionäre Algorithmen, Fuzzy-Systeme und Bayes-Netze,

2. Aufl., Wiesbaden 2015, S. 33 ff.; REY GÜNTHER DANIEL/WENDER KARL F., Neuronale Netze, Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung, 3. Aufl., Bern 2018, S. 15 ff.; SCHMIDHUBER JÜRGEN, Deep Learning in Neural Networks, An Overview, in: Neural Networks 61 (2015), S. 85 ff. In der rechtswissenschaftlichen Literatur wurden KNN u.a. thematisiert von GÜNTHER JAN-PHILIPP, Roboter und rechtliche Verantwortung, Eine Untersuchung der Benutzer- und Herstellerhaftung, München 2016, S. 26 ff.; KARANASIOU ARGYRO/PINOTSIS DIMITIRIS A., [Towards a legal definition of machine intelligence, The argument for artificial personhood in the age of deep learning.](#)

<sup>12</sup> Die ausschliesslich auf Grundlage von personenbezogenen Daten erfolgte Kreditvergabe ist nach Art. 22 der EU-Datenschutz-Grundverordnung unzulässig, bildet aber ein anschauliches Beispiel für die grundsätzliche Funktions- und Verwendungsweise von KNN.

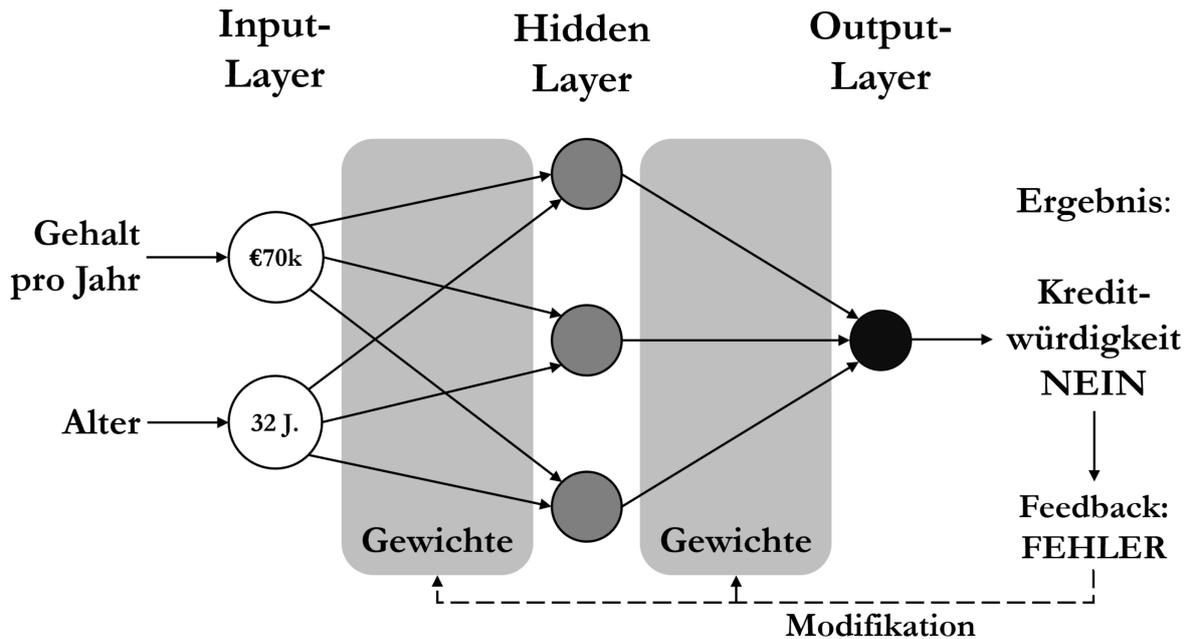


Abb. 2: Schematische Darstellung des Lernprozesses eines KNN am Beispiel von Kreditscoring.

eingeben werden.<sup>13</sup> Output-Neuronen repräsentieren die gewünschte Entscheidung, z.B. ob jemand kreditfähig ist oder den konkreten Kreditscore bzw. den Zinssatz für ein Darlehen. *Hidden-Layers* sind ein- oder mehrstufig miteinander verknüpfte Knotenpunkte, welche die eigentliche Datenverarbeitung repräsentieren. Je mehr Schichten ein KNN hat, desto «tiefer» ist es; je mehr Neuronen (in einer Schicht) ein KNN hat, desto «breiter» ist es.

Die Verbindungen zwischen den Neuronen sind gewichtet. Diese Gewichte werden im Lernprozess verändert und speichern das «Wissen» des Neuronalen Netzes.<sup>14</sup> Das Ergebnis des KNN wird auf Basis der durch Training ermittelten Gewichte zwischen den Neuronen erzielt.

## B. Der Lernprozess

Das Anlernen eines KNN setzt eine grosse Menge Trainingsdaten voraus. Diese müssen das korrekte Ergebnis bereits beinhalten. Der Lernprozess unterteilt sich Trainings-, Test- und Kreuzvalidierungsphase.<sup>15</sup>

In der Trainingsphase versucht das KNN die Ergebnisse der Trainingsdaten möglichst genau vorherzusagen. Dazu werden die Gewichte zunächst zufällig gewählt<sup>16</sup> und dann leicht verändert, sodass sich der Output des KNN an das korrekte Ergebnis annähert (siehe Abb. 2).<sup>17</sup> Dies wird unzählige Male wiederholt, bis die erwünschte Genauigkeit erreicht ist. In der Testphase wird mit einem anderen Teil der Trainingsdaten untersucht, wie genau die Vorhersage des KNN auf bisher unbekanntem Daten ist. In der Kreuz-

<sup>13</sup> Hierbei werden die Daten zunächst normalisiert, d.h. innerhalb einer Skala von 0 bis 1 (als höchster Wert) verortet.

<sup>14</sup> REY/WENDER (Fn. 11), S. 17; STYCZYNSKI ZBIGNIEW A./RUDION KRZYSZTOF/NAUMANN ANDRÉ, Einführung in Expertensysteme Grundlagen, Anwendungen und Beispiele aus der elektrischen Energieversorgung, Berlin/Heidelberg 2017, S. 138.

<sup>15</sup> Klassischerweise wird das Datenset dreigeteilt und je ein Drittel für eine der Phasen verwendet. Vgl. zum Lernprozess REY/WENDER (Fn. 11),

S. 28 f. und zur Kreuzvalidierung Ertel (Fn. 11), S. 218 ff. und KRUSE/BORGELT/BRAUNE (Fn. 11), S. 77 f.

<sup>16</sup> RUSSELL/NORVIG (Fn. 3), S. 66; TRINKWALDER (Fn. 6), S. 133.

<sup>17</sup> Vgl. KARANASIOU/PINOTSIS (Fn. 11), S. 3; zu den verschiedenen Lernregeln mit ihren Vor- und Nachteilen: REY/WENDER (Fn. 11), S. 35 ff. Z.T. wird *random weight guessing* verwendet, d.h. die Gewichte werden zufallsbasiert modifiziert, vgl. hierzu SCHMIDHUBER (Fn. 11), S. 94.

validierungsphase wird die Genauigkeit des KNN mit anderen Algorithmen oder anders aufgebauten KNN verglichen. Unter anderem kann so die geeignete Tiefe des Netzes (d.h. die Anzahl der *Hidden Layers*) ermittelt werden. Erst danach kommt das KNN zur Anwendung.

Soll ein KNN z.B. Kreditscores ermitteln, würden personenbezogene Trainingsdaten wie Alter, Beruf, Wohnort oder Gehalt verwendet werden. Ein Trainingsdatensatz würde aus Informationen über die in der Vergangenheit gewährten Kredite (insb. über Kreditausfälle) zusammengestellt werden. Ein erster Teil dieser Daten würde verwendet werden, um in der Trainingsphase eine geeignete Gewichtung zu ermitteln. Mit einem zweiten Teil der Daten würde in der Testphase gemessen werden, mit welcher Genauigkeit das KNN die Kreditwürdigkeit anhand von personenbezogenen Daten ermitteln kann. Mit einem dritten Teil der Daten würde die erzielte Genauigkeit mit derjenigen von alternativen Modellen sowie mit der Genauigkeit menschlicher Entscheidungen verglichen werden.

Dieses Lernverfahren wird, da bereits manuell im Vorfeld ermittelte Ergebnisse verwendet werden, als *supervised learning* oder überwachtes Lernen bezeichnet. Beim *unsupervised learning* muss dagegen das KNN selbst Datenpunkte anhand ihrer Ähnlichkeit in Cluster einordnen.<sup>18</sup> Daneben existieren

noch weitere Verfahren, wie z.B. das bestärkende Lernen (*reinforcement learning*) bei dem das KNN mehrschrittige Aufgaben lösen muss.<sup>19</sup>

### C. Deep Learning

Hat ein KNN mehr als zwei *Hidden-Layers* und ist es dadurch «tief» und komplex, wird von *Deep Learning* gesprochen.<sup>20</sup> Das Training solcher KNN erfolgt mit besonders vielen Datensätzen (Stichwort *Big Data*). *Deep Learning* wird eingesetzt, wenn der Anwendungsfall die Berücksichtigung vieler Faktoren und damit auch die Verarbeitung der Informationen vieler Input-Neuronen erfordert oder viele *Hidden Layers* mit einer erhöhten Abstraktion komplizierte Aufgaben bewältigen sollen.<sup>21</sup>

Ein Anwendungsgebiet von Deep Learning-KNN ist die Bilderkennung zur Klassifikation oder Segmentierung von abgebildeten Objekten. Ein Beispiel stellt das mit mehr als 20 Schichten ausgestattete *GoogLeNet* dar, das Objekte auf Bildern erkennt. Die Genauigkeit von tiefen KNN ist sehr hoch; so erzielten beispielsweise sogenannte *Convolutional Neural Networks* bei der Erkennung von handgeschriebenen Ziffern eine Genauigkeit von 99,7 %.<sup>22</sup>

Zugleich werden insb. Deep Learning-KNN als undurchschaubar wahrgenommen. Schliesslich kann nicht nachvollzogen werden, was in den Zwischenschichten passiert.<sup>23</sup> Aufgrund der hohen Dimensionalität

<sup>18</sup> Vgl. REY/WENDER (Fn. 11), S. 28.

<sup>19</sup> *Reinforcement learning* wird u.a. thematisiert von MNIH VOLODYMYR ET.AL., Human-level control through deep reinforcement learning, in: *Nature* 518 (2015), S. 529 ff. und ESPOSITO ELENA, Artificial Communication? The Production of Contingency by Algorithms, in: *Zeitschrift für Soziologie* 46 (2017), S. 249 ff., S. 261 f.

<sup>20</sup> Unter Deep Learning fallen auch tiefe *Convolutional Neural Networks* und rekurrente Neuronale Netze. Vgl. zum Deep Learning v.a. WICK CHRISTOPH, Deep Learning, in: *Informatik Spektrum* 2017, S. 103 ff.; GOODFELLOW/BENGIO/COURVILLE (Fn. 11), S. 5 ff., zu *Convolutional Neural Networks* siehe S. 326 ff.

Zu rekurrenten Netzen vgl. REY/WENDER (Fn. 11), S. 66 ff.

<sup>21</sup> Allerdings ist die Komplexität eines KNN abzuwägen mit der benötigten Rechenzeit, die durch jede weitere Schicht zunimmt. Ausserdem muss die *optimale*, nicht die grösstmögliche Tiefe angestrebt werden, um eine Überanpassung (*overfitting*) zu vermeiden.

<sup>22</sup> LECUN YANN, CORTES CORINN, BURGESS CHRISTOPHER J.C., [The MNIST Database of handwritten digits](#). Zu beachten ist, dass auch Menschen Fehler machen und keine Genauigkeit von 100 % erreichen.

<sup>23</sup> Es ist von «opacity» und fehlender «interpretability» die Rede, vgl. BURRELL JENNA, How the machine «thinks», Understanding opacity in machine learning algorithms, in: *Big Data & Society*

sind die Ergebnisse der KNN nicht vorhersehbar und ihre Überprüfung und Kontrolle ist infolge des hohen Abstraktionsgrades schwer.<sup>24</sup> Sie werden deshalb auch als «*black box*» bezeichnet.<sup>25</sup>

### III. Zur Statusfrage von digitalen Agenten

Nunmehr steht folgende Frage im Raum: Können Programme, die auf KNN basieren und selbständig rechtserhebliche Entscheidungen treffen, Rechtspersonen sein?

Eine Rechtsperson muss Trägerin von Rechten und Pflichten sein können,<sup>26</sup> d.h. befähigt sein, am Rechtsverkehr teilzuhaben.<sup>27</sup> Das «Menschsein» an sich ist keine notwendige Voraussetzung, wie dies bei juristischen Personen ersichtlich wird.<sup>28</sup> Die Verleihung von (Teil-)Rechtsfähigkeit an digitale Agenten ist nur sinnvoll, wenn durch die Eigenarten von KNN Regelungs- oder Haftungslücken entstehen, die durch die Risikoverteilung beim Einsatz von Hilfspersonen besser geschlossen werden können als bei einer konventionellen Haftung des Betreibers von digitalen Agenten.<sup>29</sup>

Ein relativ häufiges Szenario dürfte eine Fehlentscheidung des digitalen Agenten sein, die weder vorhersehbar – und somit

vermeidbar – noch erklärbar ist. Wäre der digitale Agent eine eigene (volle) Rechtsperson, könnte er selbst für entstandene Schäden haften. Alternativ könnte der digitale Agent teilrechtsfähig sein und hierdurch als Stellvertreter oder Erfüllungsgehilfe auftreten. Ein von ihm verursachter Schaden würde dem Geschäftsherrn, also seinem Betreiber, zugerechnet werden. Bei durch Fehlentscheidungen des digitalen Agenten verursachten Schäden hätte nach konventioneller Rechtslage der Betreiber bei ordnungsgemäsem Training und Bedienen des digitalen Agenten keine Sorgfaltspflichten verletzt. Die tatsächliche Schadensverursachung liegt indes beim digitalen Agenten selbst: er hat ein fehlerhaftes Ergebnis geliefert. Der rechtliche Anknüpfungspunkt sollte daher auch beim digitalen Agenten liegen, nicht beim Betreiber. Im Folgenden sollen bereits erarbeitete Ansätze zum rechtlichen Status digitaler Agenten dahingehend überprüft werden, ob sie sinnvolle und pragmatische Lösungen bieten.

#### A. Die konventionelle Behandlung von digitalen Agenten

In der deutschen Zivilrechtsdogmatik wird eine Erklärung, die durch einen digitalen Agenten generiert wurde, als die des Betreibers angesehen.<sup>30</sup> Auch bei schadensverursachenden Handlungen wird nicht an das

2016, S. 1 ff., 9. Bei modularen KNN kann die Ergebnisfindung zumindest etwas transparenter gestaltet werden, siehe MASCHARKA DAVID/TRAN PHILIP/SOKLASKI RYAN/MAJUMDAR ARJUN, Transparency by Design: Closing the Gap Between Performance and Interpretability in Visual Reasoning.

<sup>24</sup> Vgl. zum sog. «*curse of dimensionality*»: DOMINGOS PEDRO, A Few Useful Things to Know About Machine Learning, in: Communications of the ACM 2012, S. 78 ff., 82 f.; sowie ALPAYDIN (Fn. 5), S. 155, 163; KARANASIOU/PINOTSIS (Fn. 11), S. 4.

<sup>25</sup> Vgl. CASTELVECCHI DAVIDE, Can we open the black box of AI?, in: Nature 2016, S. 20 ff.

<sup>26</sup> SCHMITT JOCHEM, in: MüKoBGB, 7. Aufl., München 2015, § 1 Rn. 6; BAMBERGER GEORG, in: BeckOK BGB, Stand: 1.5.2018, § 1 Rn. 10 (wobei Computersysteme ausdrücklich ausgenommen werden).

<sup>27</sup> Dieser Aspekt der Rechtsfähigkeit wird aufgegriffen bei LIPP VOLKER, in: MüKoBGB, 7. Aufl., München 2018, Art. 7 EGBGB, Rn. 4.

<sup>28</sup> Zur Parallele von juristischen und nicht-menschlichen Personen vgl. TEUBNER (Fn. 2), S. 155 ff., S. 164 f., der aus systemtheoretischer Perspektive erläutert, dass das Substrat juristischer Personen gerade nicht Menschen ist, sondern Mitteilungsketten.

<sup>29</sup> Vgl. zu dem Autonomierisiko und den hierdurch entstehenden Haftungslücken: TEUBNER (Fn. 2), S. 155 ff., 163 f., 185 ff.

<sup>30</sup> Vgl. BGHZ 195, 126 Rn. 17; Manche konstruieren dies mithilfe der *offerta ad incertas personas* oder unterstellen eine Blanketterklärung, so z.B. KIRN STEFAN/ MÜLLER-HENGSTENBERG KLAUS D., Intelligente (Software-) Agenten: Eine neue Herausforderung für die Gesellschaft und unser Rechtssystem?, 2014, S. 10; vgl. auch CORNELIUS KAI, Vertragsabschluss durch autonome elektronische Agenten, in: MMR 2002, S. 353 ff., S. 354 f., der von einem «gestreckten Verfahren»

Versagen des Softwareagenten angeknüpft, sondern an eine Sorgfaltspflichtverletzung des Betreibers (§ 276 BGB).<sup>31</sup> Es gibt genügend Stimmen in der rechtswissenschaftlichen Literatur, die (zumindest derzeit) eine eigenständige Rechtspersönlichkeit digitaler Agenten ablehnen. Dies wird zum Teil damit begründet, dass das für die Willensbildung erforderliche Bewusstsein<sup>32</sup> oder die Fähigkeit Pflichten zu erfüllen, fehle.<sup>33</sup> Aus haftungsrechtlichen Gesichtspunkten wird indessen vorgetragen, menschliche Akteure könnten digitale Agenten nutzen, um sich von möglichen Haftungen abzuschirmen, indem diese vom Nutzer auf den digitalen Agenten verlagert wird.<sup>34</sup> Ausserdem müsse erst einmal eine Haftungsmasse geschaffen werden, mit welcher der digitale Agent für Schäden einstehen kann.<sup>35</sup>

## B. Volle Rechtsfähigkeit

Im Jahr 1992 stellte LAWRENCE B. SOLUM als erster die Frage, ob Künstlichen Intelligenzen Rechtspersönlichkeit verliehen werden kann.<sup>36</sup> Damals bezeichnete er seine Darstellungen selbstkritisch als «*imaginative theorizing*» und betonte, dass die Frage erst bei alltäglichen Begegnungen mit Künstlichen Intelligenzen auf pragmatische Weise zu klären sein werde.<sup>37</sup> Zweifelsohne sehen die meisten Vertreter einer künftigen (vollen) Rechtspersönlichkeit für digitale Agenten die Realisierung ihrer Ideen in weiter Zukunft.<sup>38</sup> Zum Teil wird die künftige Organisation eines digitalen Agenten als (Kapital-) Gesellschaft im bereits existierenden Rechtsrahmen befürwortet.<sup>39</sup>

Für auf KNN basierende Agenten scheidet eine volle Rechtspersönlichkeit nach dem

---

spricht; BUSCHE JAN, in: MüKoBGB, 7. Aufl., München 2015, vor § 145 Rn. 37; MEDICUS DIETER/PETERSEN JENS, Bürgerliches Recht, 26. Aufl., München 2017, Rn. 45, 327; SÄCKER JÜRGEN, in: MüKoBGB, 7. Aufl., München 2015, Einl. Rn. 185.

<sup>31</sup> So GRUNDMANN STEFAN, in: MüKoBGB, 7. Aufl., München 2016, § 278 Rn. 46; vgl. aus ethischer Sicht ASARO PETER M., The Liability Problem for Autonomous Artificial Agents, in: Ethical and Moral Considerations in Non-Human Agents (AAAI Spring Symposium Series), Palo Alto 2016, S. 192, der sich für die Ausarbeitung eines an digitale Agenten angepassten Haftungskonzeptes ausspricht.

<sup>32</sup> So z.B. CORNELIUS KAI (Fn. 30), S. 354 f.

<sup>33</sup> SOLAIMAN S. M., Legal personality of robots, corporations, idols and chimpanzees, a quest for legitimacy, in: Artif Intell Law Vol. 25 (2017), S. 155 ff., S. 174 f.; vgl. auch BROŽEK BARTOSZ/JAKUBIEC MAREK, On the legal responsibility of autonomous machines, in: Artif Intell Law 25 (2017), S. 293 ff., S. 301 f., die bezweifeln, dass eine digitale Rechtsperson alltagspsychologisch akzeptiert würde, sodass sie zur rechtlichen Fassade verkommen würde.

<sup>34</sup> Vgl. BRYSON JOANNA J./DIAMANTIS MIHAILIS E./GRANT THOMAS D., Of, for, and by the people, the legal lacuna of synthetic persons, in: Artif Intell Law 25 (2017), S. 273 ff., S. 285 ff., 288; SPINDLER GERALD, Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz, Braucht das Recht neue Haftungskategorien?, in: CR 2015, S. 766 ff., S. 774 f.

<sup>35</sup> Hierauf verweisen GLESS SABINE/JANAL RUTH, Hochautomatisiertes und autonomes Autofahren, Risiko und rechtliche Verantwortung, in: JR 2016, S. 561 ff., S. 571 und SCHAUB, Interaktion von Mensch und Maschine, in: JZ 2017, S. 342 ff., 345 f.

<sup>36</sup> SOLUM LAWRENCE B., Legal Personhood for Artificial Intelligences, in: N.C.L.Rev 70 (1992), S. 1231 ff.

<sup>37</sup> Vgl. SOLUM (Fn. 36), S. 1286 f.

<sup>38</sup> Vgl. KOOPS BERT-JAAP/HILDEBRANDT MIREILLE/JAQUET-CHIFFELLE DAVID-OLIVER, Bridging the Accountability Gap, Rights for New Entities in the Information Society?, in: Minnesota Journal of Law, Science & Technology 11 (2010), S. 497 ff., S. 557 ff.; ALLEN TOM/WIDDISON ROBIN, Can Computers Make Contracts?, Harv. J.L. & Tech. 9 (1996), S. 25 ff., S. 40; ZIMMERMAN, Machine Minds: Frontiers in Legal Personhood, 2015, S. 41; CHOPRA SAMIR/WHITE LAURENCE F., A Legal Theory for Autonomous Artificial Agents, Ann Arbor 2011), S. 186 ff. (die davon ausgehen, dass digitalen Agenten Rechtspersönlichkeit dann zugesprochen wird, wenn die Notwendigkeit hierzu besteht).

<sup>39</sup> BAYERN SHAWN ET.AL., Gesellschaftsrecht und autonome Systeme im Rechtsvergleich, in: AJP 2017, S. 192 ff.; HÄUSERMANN DANIEL M., Autonome Systeme im Rechtskleid der Kapitalgesellschaft, in: AJP 2017, S. 204 ff.; kritisch: LOPUCKI LYNN M., Algorithmic Entities, in: Washington University Law Review 95 (2018), S. 1 ff., S. 67: «Entity law is not only incapable of regulating AEs, it is incapable of regulating much of anything.»

derzeitigen Stand der Technik aus. Diese sind nur in Bezug auf eine spezielle Aufgabe einsetzbar. Sind sie beispielsweise auf Vorhersagen von Lagerbeständen und dem automatischen Nachbestellen auf Grundlage der Vorhersage ausgelegt, heisst das nicht, dass sie sich selbst im Falle eines Rechtsstreits vertreten oder auch nur einen Rechtsanwalt damit betrauen könnten.<sup>40</sup> Ausserhalb ihres konkreten Einsatzbereichs sind sie nicht handlungsfähig, weshalb Menschen als für den Agenten Handelnde involviert werden müssten, was die Sinnfrage einer vollen Rechtspersönlichkeit aufwirft.

Insbesondere ein eigener Status in Bezug auf Grund- bzw. Menschenrechte (auch *Posthuman Rights*, wie sie die *Robot Liberation Front* verlangt) erfordert wohl «Menschsein», mindestens aber eine artifizielle Form von freiem Willen oder Bewusstsein.<sup>41</sup>

Zu guter Letzt ist zu berücksichtigen, dass digitale Agenten weder eigene Interessen verfolgen noch einen Selbstzweck erfüllen: sie dienen Menschen bzw. Unternehmen, oder werden zu Forschungszwecken eingesetzt. Sie agieren ausschliesslich fremdnützig,

sodass sie im Rechtsverkehr die Rolle von Hilfspersonen einnehmen. Im Folgenden sind daher die Konzepte zu einer Teilrechtsfähigkeit zu beleuchten, die genau diesem Umstand Rechnung tragen.

### C. Teilrechtsfähigkeit

Ein eigenständiger Status im Sinne einer Teilrechtsfähigkeit wäre nur sinnvoll, wenn er der Funktionsweise und den Einsatzszenarien von digitalen Agenten gerecht wird und losgelöst von Analogien zum Menschen und zu juristischen Personen ist.<sup>42</sup>

CURTIS E. A. KARNOW schuf das Konzept der «*electronic persona*», um den Herausforderungen des Digitalisierungsprozesses gerecht zu werden.<sup>43</sup> Eine solche Konzeption würde das Auftreten als Stellvertreter bei Vertragsschlüssen sowie die Haftung bei Fehlverhalten ermöglichen,<sup>44</sup> vorausgesetzt es besteht eine Haftungsmasse. Eine Registereintragung kann Sicherheit bzgl. des Status und der Solvenz des digitalen Agenten bringen.<sup>45</sup> Gleichzeitig wäre ein solches Register kostenintensiv und schwer aktuell zu halten.<sup>46</sup>

<sup>40</sup> Dieses Problem wird u.a. angesprochen von: BRYSON/DIAMANTIS/GRANT (Fn. 34), S. 288; SPINDLER (Fn. 34), S. 775.

<sup>41</sup> Vgl. KOOPS/HILDEBRANDT/JAQUET-CHIFFELLE (Fn. 38), S. 557 ff.; dagegen spricht sich KERSTEN JENS, Menschen und Maschinen, Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen, in: JZ 2015, S. 1 ff., S. 6 ff., für die Gewährung einer Rechtspersönlichkeit sowie für die grundrechtliche Gleichstellung mit juristischen Personen i.R.v. Art. 19 Abs. 3 GG aus (insb. bzgl. der Berufsfreiheit).

<sup>42</sup> Vgl. BECK SUSANNE, Der rechtliche Status autonomer Maschinen, AJP 2017, S. 183 ff., S. 186 ff.; ČERKA PAULIUS/GRIGIENĖ JURGITA/SIRBIKYTĖ GINTARĖ, Is it possible to grant legal personality to artificial intelligence software systems?, in: CLSR 33 (2017), S. 685 ff., S. 695 f.

<sup>43</sup> KARNOW CURTIS E. A., The Encrypted Self, Fleshing Out the Rights of Electronic Personalities, J. Marshall J. Computer & Info. L. 13 (1994), S. 1 ff., S. 4 ff.; vgl. auch DERS., Liability for Distributed Artificial Intelligences, in: Berkeley Tech. L.J. Vol. 11 (1996), S. 147 ff., S. 193 ff., zu dem von ihm entworfenen «*Turing-Register*».

<sup>44</sup> WETTIG STEFFEN/ZEHENDNER EBERHARD, A legal analysis of human and electronic agents, in: Artif Intell Law 12 (2004), S. 111 ff., S. 127 f.; ALLEN/WIDDISON (Fn.38), S. 41 ff., wobei sie von «agents» für die Unternehmen reden; KOOPS/HILDEBRANDT/JAQUET-CHIFFELLE (Fn. 38), S. 555 ff.; BECK SUSANNE, Über Sinn und Unsinn von Statusfragen, zu Vor- und Nachteilen der Einführung einer elektronischen Person, in: Günther Jan-Philipp/Hilgendorf Eric (Hrsg.), Robotik und Gesetzgebung, Baden-Baden 2013, S. 239 ff., S. 256 f.; vgl. BECK SUSANNE, Der rechtliche Status autonomer Maschinen, in: AJP 2017, S. 183 ff., S. 189 f., die eine Regelung des Status gegenüber kleinteiligen Einzelfallregelungen vorzieht.

Vgl. auch PAGALLO UGO, The Laws of Robots. Crimes, Contracts, and Torts, Dordrecht 2013, S. 103 ff., 132 ff., der das Konzept eines der e-Person ähnlichen Digital Peculium vorschlägt.

<sup>45</sup> ALLEN/WIDDISON (Fn. 38), S. 42 f.; WETTIG ZEHENDNER (Fn. 44), S. 128; KARNOW, 1996 (Fn. 43), S. 193 ff.; BECK, 2013 (Fn. 44), S. 256.; KOOPS/HILDEBRANDT/JAQUET-CHIFFELLE (Fn. 38), S. 555 f.

<sup>46</sup> Vgl. ALLEN/WIDDISON (Fn. 38), S. 43; SANTORO SERENA/ONIDA TESSA/ROMANO FRANCESCO, Electronic Agents And The Law

Partiell wird auch eine Versicherung für digitale Agenten gefordert, die Schäden kompensieren kann.<sup>47</sup>

Alternativ hierzu kann digitalen Agenten eine Teilrechtsfähigkeit zukommen, durch die sie als Hilfspersonen (v.a. §§ 164 ff., 278 BGB) handeln können.<sup>48</sup> Beispielsweise kann ein digitaler Agent mangels Rechtsfähigkeit kein Stellvertreter sein und ohne freien Willen könnte auch die Abgabe einer Willenserklärung zweifelhaft sein. Die Abgabe einer eigenen Willenserklärung, - in Abgrenzung zur Botenschaft, d.h. der Übermittlung einer fremden Erklärung - wird allerdings aus Gründen des Verkehrsschutzes nach dem äusseren Erscheinungsbild bestimmt.<sup>49</sup> Die durch einen digitalen Agenten abgegebene Erklärung beruht auf den eigenständig ermittelten Ergebnissen des KNN, ohne dass der Ergebnisfindungsprozess nachvollziehbar ist. Damit wird der Inhalt der Erklärung vom digitalen Agenten selbst geschaffen.<sup>50</sup>

Diskutiert wird, wie der Fall eines *falsus procurator* (lat. falschen Vertreters) zu behandeln ist, d.h. wenn der digitale Agent den ihm zugewiesenen Entscheidungsspielraum verlässt. Um jegliche Haftungslücken zu schliessen, könnte man digitale Agenten mit Eigenvermögen,<sup>51</sup> beispielsweise im Rahmen einer Versicherungslösung,<sup>52</sup> konzipieren. Alternativ könnte man die Haftung durch entsprechende Anwendung von § 179 Abs. 3

Satz 2 BGB aufgrund der Vermögenslosigkeit des digitalen Agenten ausschliessen.<sup>53</sup>

Dieses Problem stellt sich bei KNN nicht, da Restriktionen des Entscheidungsspielraums in die Software implementiert werden können. Werden solche Vorkehrungen nicht getroffen, wird regelmässig eine Sorgfaltpflichtverletzung des Betreibers vorliegen.

Ein digitaler Agent könnte ebenso ein Erfüllungsgehilfe im Sinne von § 278 Satz 1 BGB sein. Die Einordnung als blosses Werkzeug oder Hilfsmittel erscheint angesichts der Fähigkeit zu selbständigen Entscheidungen unangemessen.<sup>54</sup> Eine Verschuldenshaftung des Betreibers würden zu Haftungslücken führen, schliesslich werden Schäden in der Regel durch die dem digitalen Agenten immanente Funktionsweise hervorgerufen. Der einzige Anknüpfungspunkt für ein Fehlverhalten wäre demnach die Inbetriebnahme des Systems.<sup>55</sup> Würde man hieran anknüpfen, wäre dies – mangels Verbots eines Einsatzes von KNN – eine Gefährdungshaftung. Die Anknüpfung an ein fehlerhaftes Verhalten des digitalen Agenten wäre demgegenüber vorzuziehen, schliesslich erfüllt er die gleiche Rolle wie ein menschlicher Erfüllungsgehilfe.<sup>56</sup> Freilich haftet der Betreiber dennoch, wenn der Fehler aus einem unsorgfältig betriebenen Trainingsprozess herrührt, während der Software-Hersteller für eine fehlerhafte Architektur des KNN haftet. Gerade letzteres sollte aber aufgrund der

Of Agency, An analysis of how the use of electronic agents could be integrated into law, in: CRi 2004, S. 104 ff., S. 108.

<sup>47</sup> KOOPS/HILDEBRANDT/JAQUET-CHIFFELLE (Fn. 38), S. 555 f.

<sup>48</sup> GRUBER MALTE C., Was spricht gegen Maschinenrechte?, in: Gruber Malte C./Bung Jochen/Ziemann Sascha (Hrsg.), Autonome Automaten. Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft, Berlin 2014, S. 191 ff., S. 198 f.; TEUBNER (Fn. 2), S. 177 ff.; SANTORO/ONIDA/ROMANO (Fn. 46), S. 109, die von einer «minor» oder «quasi personality» sprechen.

<sup>49</sup> Vgl. KESSLER OLIVER, Intelligente Roboter, neue Technologien im Einsatz, in: MMR 2017, S. 589 ff., S. 592, wonach die Erklärung als Ergebnis eines «menschlichen Denkvorgang[s]» erscheinen muss; TEUBNER (Fn. 2), S. 183, der von

«digitalen Äquivalenten» für subjektive Voraussetzungen spricht.

<sup>50</sup> Vgl. GRUBER (Fn. 48), S. 198.

<sup>51</sup> GLESS/JANAL (Fn. 35), S. 571, kritisieren eine eigene Haftungsmasse, die als totes Kapital ökonomisch wenig sinnvoll wäre.

<sup>52</sup> Der digitale Agent würde dann über «Versicherungsvermögen» verfügen, so TEUBNER (Fn. 2), S. 195; siehe zur Versicherungslösung auch (aber im Kontext einer Gefährdungshaftung): BRÄUTIGAM PETER/KLINDT THOMAS, Industrie 4.0, das Internet der Dinge und das Recht, in: NJW 2015, S. 1137 ff., S. 1138.

<sup>53</sup> SPECHT LOUISA/HEROLD Sophie, Roboter als Vertragspartner?, in: MMR 2018, S. 40 ff., S. 43.

<sup>54</sup> Vgl. KESSLER (Fn. 49), S. 592., vgl. zum Verrichtungsgehilfen S. 593 f.

<sup>55</sup> Vgl. TEUBNER (Fn. 2), S. 185 f.

<sup>56</sup> TEUBNER (Fn. 2), S. 188.

Open-Source-Verfügbarkeit sehr selten vorkommen.<sup>57</sup>

Um auch im deliktrechtlichen Bereich eine Gefährdungs- wie auch Verschuldenshaftung des Betreibers zu vermeiden, schlägt GUNTHER TEUBNER eine *digitale Assistenzhaftung* als strikte deliktische Haftung für das Entscheidungsrisiko digitaler Agenten vor.<sup>58</sup> Diese soll an das rechtswidrige Verhalten von digitalen Agenten anknüpfen und neben Sach- und Personenschäden auch Vermögensschäden abdecken.

#### D. Alternative Lösungen

Statt den Herausforderungen durch digitale Agenten mit einer Veränderung des rechtlichen Status zu begegnen, könnte auch ein speziell auf lernfähige Software ausgelegtes Haftungskonzept implementiert werden.<sup>59</sup> Diskutiert wird auch die Rechtsfähigkeit von sog. *Hybriden*, die aus Menschen und digitalen Agenten bestehen.<sup>60</sup>

### IV. Die Konstruktion von rechtlicher Handlungsfähigkeit

Um als Stellvertreter, Erfüllungs- oder Verrichtungsgehilfe auftreten zu können, muss ein digitaler Agent handlungsfähig sein und (im Gegensatz zu einem blossen Werkzeug) autonom handeln können. Ein Stellvertreter muss eine *eigene* Willenserklärung

abgeben und ein Erfüllungs- oder Verrichtungsgehilfe muss einen Schaden selbst verursacht haben. Diese Voraussetzungen müssen von einem digitalen Agenten erfüllt werden.

#### A. Agenten aus informationstechnischer und philosophischer Sicht

In der Informatik gilt ein System als Agent, wenn dieses in intelligenter oder rationaler Weise handelt.<sup>61</sup> Wann ein Agent oder ein Programm «intelligent» handelt, ist aber aufgrund der Mehrdeutigkeit des Begriffs «Intelligenz» nicht genau zu bestimmen – bedeutet Intelligenz Denken, Bewusstsein, oder reicht intelligentes Handeln?<sup>62</sup>

Die Frage nach der Denkfähigkeit von Maschinen hat ALAN TURING durch das *imitation game* ersetzt: eine Person stellt einem Menschen und einer Maschine über einen Bildschirm und eine Tastatur Fragen, ohne zu wissen, welcher Gesprächspartner Maschine und welcher Mensch ist. Wenn die fragende Person den Menschen nicht mehr eindeutig von der Maschine unterscheiden kann, könne man der Maschine zumindest intellektuelle Fähigkeiten, die mit dem Denkvermögen des Menschen vergleichbar sind, zuschreiben, ohne «Intelligenz» oder «Denkvermögen» definieren zu müssen.<sup>63</sup>

<sup>57</sup> Z.B. ist TensorFlow ein Open-Source Framework für maschinelles Lernen und wird von vielen Anwendungen verwendet. Fehler im Code werden daher schnell entdeckt und behoben.

<sup>58</sup> TEUBNER (Fn. 2), S. 192.

<sup>59</sup> Vgl. ASARO (Fn. 31), S. 192 f.; SPINDLER (Fn. 34), S. 775 f., spricht sich für eine Gefährdungshaftung und eine Haftpflichtversicherung aus.

<sup>60</sup> Vgl. v.a. GRUBER MALTE-C., Zumutung und Zumutbarkeit von Verantwortung in Mensch-Maschine-Assoziationen, Ein Beitrag zur zivilrechtlichen Entwicklung der Roboterhaftung, in: Hilgendorf Eric/Günther Jan-Philipp (Hrsg.), Robotik und Gesetzgebung, Baden-Baden 2013, S. 123 ff.; TEUBNER (Fn. 2), S. 196 ff.; aber auch ALLEN/WIDDISON (Fn. 38), S. 40. Eine Auseinandersetzung mit Hybriden kann an dieser Stelle aber leider nicht erfolgen. Vgl. ferner zu

Multi-Agenten-Systemen, TEUBNER (Fn. 2), S. 201 ff.

<sup>61</sup> RUSSELL/NORVIG (Fn. 3), S. 60, 64; vgl. auch FRANKLIN STAN/GRAESSER ART, Is It an agent, or just a program?, A taxonomy for autonomous agents, in: Müller Jörg/Wooldridge Michael J./Jennings Nicholas R. (Hrsg.), Intelligent Agents III, Agent Theories, Architectures, and Languages, Berlin/Heidelberg 1997, S. 21 ff.

<sup>62</sup> RUSSELL/NORVIG (Fn. 3), S. 22 ff.

<sup>63</sup> TURING ALAN, Computing Machinery and Intelligence, in: Mind 1950, S. 433 ff., S. 434 f. Vgl. in der Diskussion um die Rechtssubjektivität digitaler Agenten auch TEUBNER GUNTHER, Elektronische Agenten und grosse Menschenaffen, Zur Ausweitung des Akteursstatus in Recht und Politik, in: Zeitschrift für Rechtssoziologie 2006, S. 5 ff., S. 9, der die Debatte weg von ontologischen Eigenschaften führt.

Auch im Rechtsverkehr ist demnach weniger das Denkvermögen oder Bewusstsein eines Agenten entscheidend. Stattdessen müssen Agenten in der Lage sein, in rechtserheblicher Weise miteinander zu kommunizieren. Rechtliche Handlungsfähigkeit setzt folglich Kommunikationsfähigkeit voraus.<sup>64</sup>

## B. Handlungs- als Kommunikationsfähigkeit

In der Systemtheorie sind weniger die Agenten selbst Gegenstand, sondern ihre Kommunikation miteinander.<sup>65</sup> Im Rahmen der Kommunikation des den Agenten umgebenden sozialen Systems werden Akteure konstruiert, indem ihnen Subjektivität *zugeschrieben* wird.<sup>66</sup> Die Zuschreibung von Subjektivität wird dabei vor allem als Strategie zum Umgang mit Ungewissheit genutzt.<sup>67</sup> Diese Ungewissheit stammt daher, dass die Akteure intransparent und unberechenbar sind; wie eine *black box*.<sup>68</sup> Hierdurch entsteht Kontingenz, d.h. die «Ausschließung von Notwendigkeit und Unmöglichkeit [...] etwas, was weder notwendig ist noch unmöglich ist; was also so, wie es ist (war, sein wird), sein kann, aber auch anders möglich ist»<sup>69</sup>. Die Kontingenz ist doppelt; denn jeder der Kommunikationspartner erfährt die Ungewissheit sowie die vom anderen erfahrene Ungewissheit und beide machen ihre Handlung abhängig voneinander.<sup>70</sup>

Wenn uns digitale Agenten ihre Ergebnisse mitteilen, besteht eine «autonome Einheit

von Information, Mitteilung und Verstehen»,<sup>71</sup> – das Ergebnis, als Information mitgeteilt, kann *verstanden* werden.<sup>72</sup> Eine doppelte Kontingenz (also auch auf der Seite des digitalen Agenten) ist aber nur schwer vorstellbar. ELENA ESPOSITO hat für die Kommunikation mit Algorithmen den Begriff der «virtuellen doppelten Kontingenz» geschaffen.<sup>73</sup> Kontingenz wird zwar nur vom Menschen erlebt, doch schließt dies eine (wenn auch nur einseitig erlebte und damit asymmetrische) Kommunikation nicht aus.<sup>74</sup> Ein innerer Akt des Verstehens ist gerade nicht erforderlich für Kommunikation; entscheidend ist, dass sich ein kommunikativer Akt nach aussen manifestiert.<sup>75</sup>

Selbstlernende digitale Agenten, also vor allem solche, die auf KNN basieren, generieren neue Informationen,<sup>76</sup> (etwa Vorhersagen) und damit auch neue Perspektiven und Kontingenz. Der Inhalt ihrer Mitteilungen ist unvorhersehbar und schlichtweg nicht interpretierbar. Aufgrund der kommunikativen Fähigkeiten digitaler Agenten können diese als adressierfähige Personen in Sozialsystemen konstituiert werden (*artificial communication* statt *artificial intelligence*).<sup>77</sup> Sie *erscheinen* als handlungs- und kommunikationsfähig, sodass ihnen in beschränkter Weise Subjektivität zugeschrieben werden kann.

## C. Autonome digitale Agenten

Agenten sind autonom, wenn sie ihren Zustand nicht nur als Reaktion auf Stimuli

<sup>64</sup> TEUBNER (Fn. 2), S. 165; a.M. z.B. SPINDLER GERALD, *Digitale Wirtschaft – analoges Recht, Braucht das BGB ein Update?*, in: JZ 2016, S. 805 ff., S. 816, der «tatsächlich Eigenschaften der Intelligenz und selbständiger Zielsetzung» fordert.

<sup>65</sup> TEUBNER (Fn. 2), S. 165 mit Verweis auf LUHMANN NIKLAS, *Soziale Systeme, Grundriss einer allgemeinen Theorie*, Nachdruck Frankfurt/Main 2001, S. 191 ff. (zur Kommunikation und dem Begriff der Mitteilung, v.a. ab S. 193), vgl. zur kollektiven Handlungsfähigkeit S. 270 ff. So TEUBNER (Fn. 63), S. 9.

<sup>67</sup> Vgl. TEUBNER (Fn. 63), S. 11 ff.; TEUBNER (Fn. 2), S. 174 ff.; GRUBER (Fn. 48), S. 197 f.

<sup>68</sup> Vgl. hierzu LUHMANN (Fn. 65), S.156 f.; TEUBNER (Fn. 63), S. 12.

<sup>69</sup> LUHMANN (Fn. 65), S. 152.

<sup>70</sup> Vgl. ESPOSITO (Fn. 19), S. 255.

<sup>71</sup> TEUBNER (Fn. 2), S. 166.

<sup>72</sup> Vgl. ESPOSITO (Fn. 19), S. 258 f.

<sup>73</sup> ESPOSITO (Fn. 19), S. 260 (*virtual double contingency*).

<sup>74</sup> So ESPOSITO (Fn. 19), S. 255 ff.; TEUBNER (Fn. 2), S. 167 f., die beide Beispiele wie *Luhmanns Zettelkasten* oder *Beten zu Gott* anführen.

<sup>75</sup> So TEUBNER (Fn. 2), S. 168.

<sup>76</sup> Vgl. ESPOSITO (Fn. 19), S. 261 f. sowie auf S. 258 f.: «The machine in this case is not only behaving in such a way as to allow users to think that it communicates, it actually produces information from a different perspective».

<sup>77</sup> ESPOSITO (Fn. 19), S. 253 f.; TEUBNER (Fn. 2), S. 171 f.

aus der Umwelt, sondern nach eigenen (internen) Regeln verändern können.<sup>78</sup> Aufgrund der Entscheidung des digitalen Agenten unter Ungewissheit wird dessen Autonomie angenommen: der digitale Agent hat zwischen Alternativen zu entscheiden, und soll dabei ein Optimum an Genauigkeit erreichen, ohne dass die Entscheidung vorhersehbar oder interpretierbar ist.<sup>79</sup> Diese aus Ungewissheit resultierende, kaum überprüfbare Entscheidung öffnet überhaupt erst einen Raum für Verantwortung im Sinne eines (rechtlichen) Einstehenmüssens für diese Entscheidung.<sup>80</sup> Insofern lässt sich die *Fähigkeit* der digitalen Agenten autonom zu handeln *konstruieren*, ohne Rückgriff auf menschliche Eigenschaften wie Wille oder Bewusstsein nehmen zu müssen. Für nichtmenschliche, aber handlungsfähige Akteure hat BRUNO LATOUR den Begriff der «*Aktanten*» geschaffen,<sup>81</sup> um eine Anthropomorphisierung von digitalen Agenten zu vermeiden.

#### D. Intentionalität digitaler Agenten

Vielfach sind subjektive Voraussetzungen bzw. innere Tatsachen wie die Fahrlässigkeit im Haftungsrecht und der Wille zur Abgabe einer Willenserklärung notwendig. Man kann aber mit Fug bezweifeln, dass digitale Agenten einen Willen haben. Allerdings kann man annehmen, dass der digitale Agent zielgerichtet handelt. Dem digitalen Agenten könnte man einen bestimmten Willen

*zuschreiben*, der als die hinter der Handlung stehende Intention angenommen wird.<sup>82</sup>

Um Agenten intentionale Zustände unabhängig von Bewusstsein oder Willensfähigkeit zuzuschreiben, hat DANIEL DENNETT die *intentional stance* (die Betrachtung eines Verhaltens anhand der vermuteten Intentionen) entwickelt.<sup>83</sup>

Will man das Verhalten eines Agenten erklären oder vorhersagen, wäre dies nach der *physical stance* (der physischen Bauweise des Systems) oder nach dem *design* (die Konzeption bzw. Zweckrichtung des Systems) denkbar. Aufgrund der Komplexität selbstlernender digitaler Agenten ist diese Perspektive nur beschränkt anwendbar. Stattdessen kann man die *intentional stance* zugrunde legen, indem wir die Rationalität des Systems annehmen und auf dessen Intentionen und Ziele schließen. Diese Betrachtungsweise fügt sich in die Objektivierungstendenzen in Bezug auf die Sorgfaltsmassstäbe im Haftungsrecht ein<sup>84</sup> und erlaubt die Zuschreibung innerer Tatsachen in einer dem Zivilrechtssystem entsprechenden Weise.

#### V. Ausblick

Wie genau der Rechtsstatus digitaler Agenten ausgestaltet wird, bleibt vorerst offen: sie könnten weiterhin lediglich als Werkzeuge angesehen werden oder aber Teilrechtsfähigkeit bzw. in weiter Zukunft auch volle Rechtspersönlichkeit erhalten. Sie

<sup>78</sup> FLORIDI LUCIANO/SANDERS J. W., On the Morality of Artificial Agents, in: *Minds and Machine* 14 (2004), S. 349 ff., S. 357 bzgl. Agenten allgemein und S. 363 in Bezug auf moralische Agenten.

<sup>79</sup> Vgl. TEUBNER (Fn. 2), S. 174, mit Verweis auf BORGES, Rechtliche Rahmenbedingungen für autonome Systeme, in: *NJW* 2018, S. 977 ff., S. 978.

<sup>80</sup> TEUBNER (Fn. 2), 174 ff.

<sup>81</sup> LATOUR BRUNO, Das Parlament der Dinge: für eine politische Ökologie, Frankfurt/Main 2010, S. 108, 285.

<sup>82</sup> Vgl. TEUBNER (Fn. 2), 172.

<sup>83</sup> Siehe zur *intentional stance*: DENNETT DANIEL C., Mechanism and Responsibility, in: Watson Gary (Hrsg.), *Free Will*, London 1982, S. 150 ff., S.

154 ff.; DERS., *The intentional stance*, Cambridge MA 1987, S. 13 ff.; vgl. CHOPRA/WHITE (Fn. 38), S. 12 ff.; GRUBER (Fn. 48), S. 197 f.; SARTOR GIOVANNI, Cognitive automata and the law, electronic contracting and the intentionality of software agents, in: *Artif Intell Law* 17 (2009), S. 253 ff., S. 256 ff., zur *intentional stance* insb. S. 260 ff.

<sup>84</sup> So TEUBNER (Fn. 2), S. 172; vgl. GRUNDMANN STEFAN, in: *MüKoBGB*, 7. Aufl., München 2016, § 276 Rn. 54 sowie WAGNER GERHARD, in: *MüKoBGB*, 7. Aufl., München 2017, § 823 Rn. 38 ff. zum objektiven Sorgfaltsmassstab bei der Fahrlässigkeit; auch beim Vertragsschluss ist der Rechtsbindungswille «aus Sicht eines verständigen Adressaten» vorliegen, vgl. BUSCHÉ JAN, in: *MüKoBGB*, 7. Aufl., 2015, § 145 Rn. 7.

könnten als digitale Assistenten, als Hilfspersonen oder im Rahmen einer eigens für sie entwickelten Haftungsordnung für Schäden eintreten. Hierfür könnte ein verpflichtendes Eigenvermögen oder eine Versicherungspflicht eingeführt werden.

Wichtig ist, die Sorgfaltspflichten des Betreibers eines digitalen Agenten genauer auszugestalten. Verwendet er ein KNN, das beispielsweise in einer *Public Cloud* (d.h. unter Berücksichtigung der Daten anderer Nutzer) fortlaufend trainiert wird, könnten die Trainingsdaten nicht mehr zum Anwendungsprofil des Betreibers passen und zu fehlerhaften Ergebnissen führen. Diesen Risiken kann der Betreiber, der die spezifischen Bedürfnisse des Anwendungsbereichs des digitalen Agenten kennt, am besten begegnen.

Für andere Risiken ist die Zurechnung von Verantwortung deutlich diffuser. So sollte z.B. *Tay*, ein Chatbot von Microsoft, auf Twitter durch die Interaktion mit anderen Nutzern lernen.<sup>85</sup> Nach kurzer Zeit gab sie aufgrund der gezielten Beeinflussung durch andere Twitter-Nutzer rassistische, gar nationalsozialistische Tweets von sich. Hätte sie auf Twitter andere Nutzer beleidigt und würden diese wegen Ehrverletzung Schmerzensgeld verlangen, wäre unklar, wer zur Verantwortung zu ziehen wäre. Microsoft als Unternehmen hat *Tay* zwar verwendet, hatte aber die Entwicklung hin zu rassistischen Tweets weder vorhergesehen noch beabsichtigt. Die einzelnen Twitter-Nutzer, von denen *Tay* den Rassismus «gelernt» hat, haben aber so diffus zur Rechtsverletzung beigetragen, dass eine Verantwortlichkeit kaum zu konstruieren wäre.

Der Status einer *vollen* Rechtspersönlichkeit für Künstliche Intelligenz ist angesichts des derzeitigen Stands der Technik zwar wenig sinnvoll. Fraglich ist aber, wie die bereits jetzt bestehenden Risiken rechtlich zu lösen sind, die einerseits in der zunehmenden Autonomie, andererseits in der fehlenden Nachvollziehbarkeit der Ergebnisse digitaler

Agenten begründet liegen. Die Verleihung von *Teilrechtsfähigkeit* ermöglicht, Entscheidungen und Rechtsverletzungen von digitalen Agenten deren Betreibern zuzurechnen. Für eine Eigenhaftung des digitalen Agenten besteht dagegen kein Bedürfnis: entweder muss sich der Betreiber die Entscheidungen des digitalen Agenten zurechnen lassen oder, wenn der vorgesehene Bereich möglicher Entscheidungen verlassen wurde, haftet derjenige, der die ordnungsgemäße Programmierung des digitalen Agenten zu verantworten hat. Dieser Lösungsweg scheint sich am besten in die Zivilrechtsdogmatik einzufügen und vermeidet eine pauschale Gefährdungshaftung des Betreibers. Stattdessen kann bei der Haftung zwischen der sorgfältigen Programmierung eines digitalen Agenten und dessen Nutzung im Rechtsverkehr differenziert werden.

Eine Alternative bietet die digitale Assistenzhaftung als neues Regelungsregime, wobei hierfür die Frage der Haftungssumme zu klären wäre.

<sup>85</sup> Siehe hierzu u.a. SICKERT TERESA, [Vom Hipster-Mädchen zum Hitler-Bot](#), Spiegel Online vom 24.03.2016.