

# Verhaltens- und Wissenszurechnung beim Einsatz von Künstlicher Intelligenz

Philipp Hacker\*

A. Einführung .....	244	b) Grundsätze für die Zurechnung expliziten KIA-Wissens .....	274
B. Zurechnung im deutschen Privatrecht .....	247	3. Zurechnung impliziten Wissens im rechtsgeschäftlichen Bereich .....	277
C. Verhaltenszurechnung .....	248	a) Implizites KIA-Wissen auf Grundlage der BGH-Rechtsprechung .....	277
I. Vertraglicher Bereich .....	249	b) Grundsätze für die Zurechnung impliziten KIA-Wissens .....	278
1. KI-Agenten als Zurechnungsausgangssubjekt .....	249	aa) Strikte Wissenszurechnung .....	278
a) Menschen und Werkzeuge ...	249	bb) Zurechnung auf Grundlage der Figur des Sich-Verschließens .....	279
b) KI-Agenten als Hybrididentitäten: § 278 BGB analog .....	250	(1) Fallgruppen .....	279
aa) Grade von Autonomie .....	251	(2) Gründe .....	280
(1) Vollständige Autonomie ....	252	(i) Dogmatik .....	280
(2) Schwache Autonomie .....	253	(ii) Korrespondenz von Vor- und Nachteil .....	280
(3) Mittelstarke Autonomie ....	253	(iii) KI-spezifisches Gleichstellungsargument .....	281
bb) Sinn und Zweck von § 278 BGB .....	254	c) Zwischenergebnis .....	282
2. Verschulden bei KI-Agenten ....	256	4. Wissenszurechnung im Deliktsrecht nach dem BGH ....	283
a) Inkompatibilität von KI-Agent und Verschulden .....	256	a) Wissenszurechnung auf Grundlage der BGH-Rechtsprechung .....	283
b) Garantiehaftung .....	258	aa) § 826 BGB .....	283
c) Funktionale Verschuldensäquivalente .....	259	bb) § 831 Abs. 1 BGB .....	284
aa) Pönalisierung .....	260	b) Grundsätze für die KIA-Wissenszurechnung im Deliktsrecht .....	285
bb) Effizienz .....	261	E. Schluss .....	287
cc) Haftungsbegrenzung .....	261		
(1) Anthro-Parallelität .....	261		
(2) KI-Spezifizität .....	263		
II. Deliktischer Bereich .....	265		
III. Zwischenergebnis .....	267		
D. Wissenszurechnung .....	267		
I. Überblick über die Rechtsprechung .....	268		
II. Zurechnung KI-generierten Wissens .....	270		
1. Zurechnungsausgangssubjekt ...	272		
2. Zurechnung expliziten Wissens im rechtsgeschäftlichen Bereich .....	273		
a) Explizites KIA-Wissen auf Grundlage der BGH-Rechtsprechung .....	273		

\* Dr. Philipp Hacker, LL.M. (Yale) ist A.SK Fellow am Wissenschaftszentrum Berlin sowie Habilitand am Lehrstuhl für Bürgerliches Recht, Deutsches, Europäisches und Internationales Privat- und Wirtschaftsrecht der Humboldt-Universität zu Berlin.

*Abstract: Künstliche Intelligenz (KI), basierend vor allem auf Techniken maschinellen Lernens, wird in einer zunehmenden Anzahl von Wirtschaftsbereichen eingesetzt. Dabei stellen sich unausweichlich auch Haftungsfragen. Regelmäßig müssen dabei die Aktivität oder der interne Zustand eines KI-Agenten, zum Beispiel eines Roboters, einem Haftungssubjekt zugerechnet werden. Dieser Beitrag analysiert, ausgehend von der Dogmatik der Verhaltens- und Wissenszurechnung im deutschen Privatrecht, drei zentrale Probleme in diesem Schnittbereich: (i) die mögliche Parallelisierung von KI-Agenten mit menschlichen Akteuren angesichts wachsender maschineller Autonomie; (ii) die Frage des Verschuldens von KI-Agenten; und schließlich (iii) die Zurechenbarkeit von implizitem maschinellem Wissen. Dabei werden aus einer Analyse der jeweils relevanten Risiken Grundsätze für den Umgang mit KI-Agenten entwickelt, die auf Grundlage des geltenden Rechts umgesetzt werden können.*

## **A. Einführung**

Die Übernahme von Verantwortung für, und damit die rechtliche Zurechnung von, Handlungen und Wissen anderer ist ein Kernproblem arbeitsteiliger Gesellschaft, das Rechtsprechung und Literatur seit jeher beschäftigt hat.<sup>1</sup> Dieses Problem verschärft sich in dem Maße, in dem marktwirtschaftliche Entscheidungs- und Handlungsprozesse nicht nur zwischen menschlichen, sondern zunehmend auch unter Einbezug autonomer maschineller Akteure aufgespalten werden. Derartige intelligente, auf Datenverarbeitung basierende Agenten werden zunehmend im Kontext der Künstlichen Intelligenz (KI) entwickelt und bereits in einer Reihe von Sektoren erfolgreich eingesetzt.<sup>2</sup>

KI zeichnet sich dadurch aus, dass ihre Handlungsweise adaptiv ist und nicht regelbasiert prä-programmiert.<sup>3</sup> Sie analysiert und reagiert damit im Rahmen der Situationen, für die sie konfiguriert ist, autonom und kann Aufgaben übernehmen, für die alternativ ein menschlicher Akteur zuständig wäre. Akteure, die eine derartige Selbstständigkeit und quasi-kognitive Kompetenz an den Tag legen, sind aber typischerweise solche, deren Handlungen in arbeitsteiligen Konstellationen – zumindest potentiell – einem Geschäftsherrn zugerechnet werden können. Dieser Ar-

1 Siehe nur monographisch T. Tröger, Arbeitsteilung und Vertrag, Tübingen 2012; P. Buck, Wissen und juristische Person, Tübingen 2001.

2 Siehe etwa den Überblick bei I. Witten et al., Data Mining: Practical Machine Learning Tools and Techniques, 4. Aufl., Morgan Kaufmann, Cambridge, MA 2016, S. 21 ff.; S. Russell/P. Norvig, Artificial Intelligence: A Modern Approach, 3. Aufl., Prentice Hall, Upper Saddle River 2010, S. 28 f.; K. Ashley, Artificial Intelligence and Legal Analytics, Cambridge University Press, Cambridge 2017, S. 3 ff.; Royal Society, Machine learning: the power and promise of computers that learn by example, London 2017, S. 33 ff.

3 Siehe etwa Russell/Norvig (Fn. 2), S. 39; S. Shalev-Shwartz/S. Ben-David, Understanding Machine Learning. From Theory to Algorithms, Cambridge University Press, Cambridge 2014, S. 3 f.; im Einzelnen unten, Fn. 31 und 32 und dortiger Text.

tikel geht der Frage nach, inwiefern dies auf Grundlage der geltenden Zurechnungslehren des deutschen Rechts möglich ist, und inwieweit diese einer Anpassung bedürfen, um die spezifischen Eigenheiten von KI-Agenten<sup>4</sup> (KIA) zu erfassen.

Analytisch lässt sich der Vorgang der Zurechnung auf drei wesentliche Parameter reduzieren: erstens ein Zurechnungsausgangssubjekt, also einen Akteur, dessen Handlungen oder interne Zustände zugerechnet werden; zweitens einen Zurechnungsgegenstand, mithin das Wissen, die Handlung oder das Verschulden, das konkret zugerechnet wird; und drittens ein Zurechnungsendsubjekt,<sup>5</sup> also eine Entität, der etwas zugerechnet wird.

Alle drei Parameter werden sich mit dem Einsatz von KI erheblich wandeln. Erstens geht es beim Zurechnungsausgangssubjekt um eine Änderung des relevanten Akteurstyps, vom Menschen zum KIA. Hier stellt sich die Frage, inwiefern Zurechnungsnormen angesichts wachsender maschineller Autonomie analog auf KIA, und wenn ja auf welche spezifischen Untertypen, angewandt werden können. Zweitens ändert sich der Typ des Wissens oder der Handlung mit dem Einsatz von KI, also der Zurechnungsgegenstand. Im Rahmen der Verhaltenszurechnung ist insbesondere problematisch, inwieweit man KIA einen Verschuldensvorwurf machen kann. Hinsichtlich des Wissens ist gerade bei besonders leistungsstarken KIA (etwa tiefen neuronalen Netzen) ein Wandel von explizit zu lediglich implizit repräsentiertem Wissen zu konstatieren, ähnlich der menschlichen Unterscheidung zwischen bewussten und unbewussten Wissensinhalten. Hier fragt sich, inwiefern in komplexen Strukturen die Marginalisierung unmittelbar erschlossenen Wissens mit weiterreichender Wissensverantwortung korrespondieren sollte. Schließlich, drittens, könnte sich auch das Zurechnungsendsubjekt ändern, dies jedoch nur, wenn KIA mit einer Rechtspersönlichkeit ausgestattet werden.<sup>6</sup> Rechtspolitisch ist dies zwar äußerst fragwürdig;<sup>7</sup> im Grunde würde eine Anerkennung der Rechtspersönlichkeit von KIA für die Zurechnung von Handlungen oder Wissen zu ihnen jedoch keine besonderen Probleme bereiten.

4 Der Begriff Agent wird hier sowohl für KI-Systeme mit (Roboter) als auch ohne physische (nicht-elektronische) Instantiierung (Software) verwendet, siehe *Russell/Norvig* (Fn. 2), S. 34.

5 Begriff bei BGH NJW 1990, S. 975 (1976); *D. Medicus*, Probleme der Wissenszurechnung, Karlsruher Forum 1994, S. 4 (11); auch Zurechnungsobjekt genannt, siehe *Buck* (Fn. 1), S. 108; J. Taupitz, Wissenszurechnung nach englischem und deutschem Recht, Karlsruher Forum 1994, S. 16 (18).

6 Die Zurechnung ist im Übrigen grundsätzlich von der Möglichkeit einer Eigenhaftung des Gehilfen unabhängig, siehe etwa S. *Grundmann*, in: F. J. Sacker et al. (Hrsg.), Münchener Kommentar BGB, Band 2, 7. Aufl., München 2016, § 278 Rn. 1; umstritten für Verkehrssicherungspflichten im Bereich des § 31 BGB, siehe A. *Arnold*, in: F. J. Sacker et al. (Hrsg.), Münchener Kommentar BGB, Band 1, 7. Aufl., München 2015, § 31 Rn. 31; siehe aber BGH NJW 1990, S. 976 (1977 f.). Wo immer relevant wird jedoch hier davon ausgegangen, dass derartige Regeln einer Eigenhaftung von KIA nicht existieren.

7 Siehe etwa J.-E. *Schirmer*, Rechtsfähige Roboter, JZ 2016, S. 660; G. *Teubner*, Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten, *Ancilla Iuris* 2018, S. 106.

Daher lohnt insbesondere die Auseinandersetzung mit den ersten beiden Parametern der Zurechnung, und hier mit drei zentralen Problemen: (i) die mögliche Parallelisierung von KIA mit menschlichen Akteuren im Bereich des Zurechnungsausgangssubjekts angesichts wachsender maschineller Autonomie; (ii) die Frage des Verschuldens von KIA; und schließlich (iii) die Zurechenbarkeit von implizitem Wissen. Damit ist zugleich das Programm des Artikels umschrieben; die Untersuchung wird sich dabei an zwei Hauptbeispiele anlehnen, die hypothetische, aber nicht unrealistische Fälle aus dem Bereich der Verhaltens- bzw. Wissenszurechnung präsentieren.

Fragen der Zurechnung im Bereich der KI werden hier lediglich für das Zivilrecht untersucht, sind aber auch für das Strafrecht von erheblicher Relevanz<sup>8</sup> und können umgekehrt im Zivilrecht von strafrechtlichen Diskussionen um Schuld und Willensfreiheit profitieren.<sup>9</sup> Auch innerhalb des Zivilrechts ist jedoch der Untersuchungsgegenstand des Beitrags denkbar weit. Daher muss sich dieser Artikel auf die genannte Auswahl von Problemen beschränken. So muss etwa die deutsche Dogmatik der Handlungs- und Wissenszurechnung, trotz ihrer Schwachstellen und Widersprüche,<sup>10</sup> im Wesentlichen als Ausgangspunkt der Betrachtung dienen. Diese wird in Teil B kurz allgemein, zu Beginn von Teil D mit Blick auf die spezifische Rechtsprechung zur Wissenszurechnung referiert. Auf dieser Grundlage wird in Teil C die Handlungszurechnung mit den ersten beiden zentralen Probleme diskutiert: Autonomie und Verschulden bei KIA. Die Produkthaftung muss hier außen vor bleiben, nicht nur aus Platzgründen, sondern auch weil dieser im Bereich der KI bereits eine Vielzahl von Beiträgen gewidmet sind.<sup>11</sup> Neben dieser ist jedoch nicht nur in dogmatischer, sondern auch in praktischer Hinsicht die vertragliche und die (sonstige) deliktische Haftung des unmittelbaren Verwenders von KIA rele-

8 Siehe etwa C. Thomale, Wissenszurechnung im Gesellschaftsstrafrecht – deutsche Unternehmen vor französischen Strafgerichten, AG 2015, S. 41; siehe auch BGH NJW 2017, 250 Rn. 29 und dazu A. Reuter, Wissenszurechnung in Unternehmen, ZIP 2017, S. 310 (314).

9 Siehe etwa die Nachweise in Fn. 63 f. und 89 sowie den dortigen Text.

10 Zu einer Kritik der Wissenszurechnung umfassend jüngst etwa H.-C. Grigoleit, Zivilrechtliche Grundlagen der Wissenszurechnung, ZHR 2017, S. 160; G. Wagner, Wissenszurechnung: Rechtsvergleichende und rechtsökonomische Grundlagen, ZHR 2017, S. 203; knapper G. Spindler, Wissenszurechnung in der GmbH, der AG und im Konzern, ZHR 181 (2017), S. 311 (314-319).

11 Siehe etwa G. Wagner, Produkthaftung für autonome Systeme, AcP 217 (2017), S. 707; M. Ebers, Autonomes Fahren: Produkt- und Produzentenhaftung, in: B. Oppermann/J. Stender-Vorwachs (Hrsg.), Autonomes Fahren. Rechtsfolgen, Rechtsprobleme, technische Grundlagen, München 2017, S. 93; 1641 ff.; C. Gomille, Herstellerhaftung für automatisierte Fahrzeuge, JZ 2016, S. 76; G. Spindler, Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien?, CR 2015, S. 766 (769 ff.); M. Geistfeld, A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation, 105 California Law Review 2017, S. 1611; C. Reed et al., Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning, Queen Mary School of Law Legal Studies Research Paper No. 243/2016, <https://ssrn.com/abstract=2853462>; P. Čerka et al., Liability for damages caused by artificial intelligence, 31 Computer Law & Security Review 2015, S. 376; D. Vladeck, Machines without principals: liability rules and artificial intelligence, 89 Washington Law Review 2014, S. 117.

vant. Häufig wird es für den Geschädigten unter Vollstreckungsgesichtspunkten einfacher sein, gegen den Verwender des KIA (typischerweise seinen Vertragspartner) vorzugehen statt gegen einen fernen Hersteller. Teil D widmet sich dann den Fragen der Wissenszurechnung, insbesondere dem Problem des impliziten Wissens. Jeweils werden im Anschluss an die Diskussion der zentralen Zurechnungsfragen Grundsätze für den Umgang mit KIA entwickelt, die im Wesentlichen auf Grundlage des geltenden Rechts umgesetzt werden können. Der Beitrag schließt in Teil E mit einer Zusammenfassung.

## B. Zurechnung im deutschen Privatrecht

Das BGB kennt keine einheitliche Zurechnungslehre. Zurechnungsfragen stellen sich sowohl zwischen natürlichen Personen (zum Beispiel Ehegatten) als auch insbesondere innerhalb von Unternehmen, wo regelmäßig geklärt werden muss, welches Wissen und welche Handlungen von Mitarbeitern der juristischen Person (oder auch Gesamthandsgesellschaft) zugerechnet werden können. Vor allem um die unternehmensinterne Zurechnung wird es in diesem Artikel gehen. Die Väter des BGB haben wenige Zurechnungsnormen eher bruchstückhaft über verschiedene Rechtsmaterien verteilt und insbesondere davon abgesehen, das Prinzip des *respondeat superior* in Gesetzesform zu gießen.<sup>12</sup> Daher stellen Zurechnungsfragen nicht nur ein genuin deutsches Problem,<sup>13</sup> sondern auch eine rechtsprechungsgeprägte Materie dar, die sich von den positiv-rechtlichen Grundlagen nicht selten erheblich entfernt und, in den Worten des BGH, von „wertender Beurteilung“<sup>14</sup> geprägt ist.

Diese Ausgangslage bietet sich eigentlich für eine flexible Eingliederung neuer technologischer Entwicklungen an; die folgenden Teile dieses Aufsatzes werden jedoch verdeutlichen, dass nicht unerhebliche Probleme bei der Bewältigung von Zurechnungsfragen im Bereich künstlicher Intelligenz bestehen. Um diese Friktionen aufzuzeigen, muss zunächst ein kurzer Überblick über die Zurechnung im deutschen Privatrecht gewonnen werden.

Das BGB kennt drei zentrale Zurechnungsnormen (§§ 31, 166, 278 BGB) und mit § 831 BGB eine funktionsähnliche Vorschrift, die freilich eigenes Verschulden des Geschäftsherrn voraussetzt. Bis auf § 166 BGB betreffen diese positivrechtlichen Grundlagen ihrem Wortlaut nach ausschließlich die Zurechnung von Verhalten, nicht von Wissen. § 31 BGB regelt die Zurechnung von Handlungen innerhalb von juristischen Personen und wird analog auf Gesamthandsgesellschaften angewen-

12 Zum Grundsatz des *respondeat superior*, in rechtsvergleichender und rechtsökonomischer Perspektive, siehe etwa Wagner (Fn. 10), S. 214 ff., 260 ff.; umfassend H. Wicke, *Respondeat Superior*, Berlin 2000.

13 Wagner (Fn. 10), S. 205; zu anderen Rechtsordnungen dort ausführlich, S. 210 ff.

14 BGH NJW 1990, S. 975 (976); NJW 1996, S. 1339 (1340); NJW 1997, S. 1917.

det.<sup>15</sup> Danach ist die Gesellschaft verantwortlich für Schäden, die ein Organ einem Dritten durch eine zum Schadensersatz verpflichtende Handlung zugefügt hat. In ähnlicher Weise ordnet § 278 BGB an, dass ein Geschäftsherr innerhalb eines Schuldverhältnisses für das Verschulden von Erfüllungsgehilfen ebenso haftet wie für eigenes; wiederum werden fremde Handlung und fremdes Verschulden einer hierarchisch übergeordneten Instanz zugerechnet. Bei § 831 BGB hingegen macht sich der Geschäftsherr haftbar, wenn einerseits ein Verrichtungsgehilfe objektiv rechtswidrig einen Deliktstatbestand erfüllt und andererseits den Geschäftsherrn selbst ein Auswahl-, Überwachung- oder Instruktionsverschulden trifft. Dies ermöglicht zwar die deliktische Haftung für unterorganschaftliche Mitarbeiter, aber nur um den Preis der Exkulpationsmöglichkeit. Um diese auszuschalten, wurde von der Rechtsprechung etwa § 31 BGB in personeller Hinsicht auf eigenverantwortlich agierende, aber nicht in der Satzung verankerte Repräsentanten der Gesellschaft, wie leitende Angestellte, ausgedehnt.<sup>16</sup>

Demgegenüber regelt lediglich § 166 BGB explizit die Zurechnung von Wissen und dies unmittelbar nur für Fälle der Stellvertretung. Danach ist grundsätzlich ausschließlich das bei dem Vertreter vorhandene Wissen relevant, nur in Weisungsfällen auch das des Vertretenen. Damit ist aber noch nicht gesagt, wie es um das Wissen anderer Hilfspersonen steht. Dies ist denn auch der Anknüpfungspunkt für eine umfangreiche Rechtsprechung, zur analogen Anwendung des § 166 BGB in Fällen der sogenannten Wissensvertretung und zu weitergehenden Pflichten der Wissensorganisation.<sup>17</sup>

Die positivrechtlichen Verankerungen und die Rechtsprechung sind notwendiger Ausgangspunkt für die Zurechnungsprobleme im Bereich der KI, sowohl im Bereich der Verhaltenszurechnung (dazu sogleich) als auch der Wissenszurechnung (Teil D).

### **C. Verhaltenszurechnung**

Beginnen wir die Untersuchung von Fragen der Verhaltenszurechnung mit einem plastischen Beispiel. Ein typischer Fall arbeitsteiliger Aktivität ist die Erledigung von Reinigungsaufgaben. Bereits jetzt ist jedoch eine Automatisierung dieses Bereichs durch Putzroboter zu beobachten. Daher liegt es nicht fern zu vermuten, dass in Kürze Unternehmen Reinigungsdienste anbieten werden, bei denen nicht menschliche Reinigungsfachkräfte, sondern spezifisch trainierte KIA die eigentliche Arbeit erledigen. Stellen wir uns vor, der Wohnungseigentümer W beauftragt das Reinigungsunternehmen R, wöchentlich mithilfe eines KIA seine Wohnung zu put-

15 Siehe nur BGH NJW 2003, S. 1445 (1446); *Arnold* (Fn. 6), Rn. 15 f.

16 Siehe etwa BGH NJW 1998, S. 1854 (1856); *Arnold* (Fn. 6), Rn. 4 f., 7 und 20 f.

17 BGH NJW 1997, S. 1917; dazu ausführlich in Teil D.I.

zen. Der KIA navigiert selbstständig durch unbekanntes Terrain, indem unter anderem ein tiefes neuronales Netz Sensordaten verarbeitet, Hindernisse erkennt, den Putzbedarf ermittelt und die entsprechenden Reinigungsaktionen auslöst. In der Wohnung des W steht jedoch ein Tisch, dessen dekorativ ausgestatteter Ständer dem Stamm und Astwerk eines Baumes nachempfunden ist. Auf diesem ist eine Vase im Wert von 500 € platziert. Der KIA analysiert bei seinem Einsatz zwar den Tisch, hält ihn aber aufgrund seiner Konstruktion für einen weiter entfernt stehenden Baum<sup>18</sup> und stößt bei seiner Reinigungstätigkeit dagegen. Die Vase geht zu Bruch und W erleidet einen Schaden von 500 €. Dieser Fehler des neuronalen Netzes war allerdings von außen nicht spezifisch vorhersehbar.

Wann immer derartige Fehler von in Unternehmensstrukturen eingebundenen KIA Schäden verursachen, stellt sich die Frage, inwiefern die Handlungen dem Unternehmen als Geschäftsherrn zugerechnet werden können. Aufgrund der bekannten Defizite der deliktischen Haftung aus § 831 BGB liegt der Schwerpunkt der Analyse auf der Möglichkeit und den Grenzen der vertraglichen Haftung.

## I. Vertraglicher Bereich

Dabei macht es Sinn, die Frage der grundsätzlichen Vergleichbarkeit von Mensch und Maschine von derjenigen der Verschuldensfähigkeit der letzteren analytisch zu trennen, was in der bisherigen Literatur nicht hinreichend geschehen ist. Dadurch lassen sich Probleme hinsichtlich des tauglichen Zurechnungsausgangssubjekts abgrenzen von jenen der Erfüllung der weiteren Tatbestandsvoraussetzungen der Zurechnungsnorm durch dieses Subjekt.

### 1. KI-Agenten als Zurechnungsausgangssubjekt

#### a) Menschen und Werkzeuge

Stellen wir uns den geschilderten Fall Reinigungsfall zu Vergleichszwecken außerhalb der KI-Sphäre vor. Dann wird regelmäßig eine Hilfsperson Werkzeuge nutzen (etwa einen Staubsauger), um eine Wohnung zu putzen. Ist diese Hilfsperson bei einem Unternehmen angestellt, mit dem der Wohnungseigentümer einen Reinigungsvertrag abgeschlossen hat, so gilt für die Zurechnung Folgendes: Das Unternehmen haftet nach § 278 S. 1 BGB strikt für Fehler der Hilfsperson (sofern diese, als Verhalten des Geschäftsherrn gedacht, eine schuldhafte Pflichtverletzung darstellen),<sup>19</sup> ferner aus eigenem Verschulden, sofern das Werkzeug nicht sorgfältig ausgewählt, eingesetzt und ausreichend gewartet wurde.<sup>20</sup> Lediglich eine Mindermeinung möchte darüber hinaus § 278 S. 1 BGB analog auf nicht erkennbares

18 Vgl. den Unfall mit einem Tesla im autonomen Fahrmodus: Reuters, Tesla considering two theories that may explain fatal Model S crash, The Guardian (29 July, 2016), <https://www.theguardian.com/technology/2016/jul/29/tesla-two-theories-fatal-model-s-crash-autopilot>.

19 Siehe nur T. Pfeiffer, in: J. Ekkenga (Hrsg.), Soergel. BGB, 13. Aufl., Stuttgart 2014, Band 3/2, § 278 Rn. 25.

technisches Versagen bei (nicht-intelligenten) Maschinen anwenden.<sup>21</sup> Nach der herrschenden Meinung hingegen bleibt es bei der Zurechnung allein der Handlungen des menschlichen Maschinenverwenders.<sup>22</sup> Es besteht grundsätzlich schon keine Nachforschungspflicht dahingehend, bei einem Reinigungsvertrag verwendete Werkzeuge beständig auf Fehler zu untersuchen, sofern diese nicht im Rahmen der allgemeinen Sorgfalt (also durch sorgfältige Auswahl und regelmäßige Wartung) vorhersehbar und vermeidbar sind.<sup>23</sup> Explodiert also etwa der bisher einwandfrei funktionierende Staubsauger während des Reinigungsvorgangs und zerstört eine Vase, ist der der Geschädigte auf die Produkt- und Produzentenhaftung nach § 1 Abs. 1 ProdHG, § 823 Abs. 1 BGB verwiesen. Mit dem Fehler des Werkzeugs wird grundsätzlich der Hersteller, nicht der Geschäftsherr des unmittelbaren Verwenders belastet.

#### b) KI-Agenten als Hybrididentitäten: § 278 BGB analog

Durch den Einsatz von KIA kollabiert nun die kategoriale Unterscheidung zwischen Personal und Maschine: Der KIA ist vielmehr beides in einem. Gegenüber dem Reinigungsunternehmen kann jedoch nur eine der beiden Zurechnungsvarianten greifen. Entweder, man wendet § 278 S. 1 BGB hinsichtlich des Merkmals des Erfüllungsgehilfen analog auf KIA an, oder man betrachtet diese im Wesentlichen als Werkzeuge, für die nicht der Verwender, sondern in erster Linie der Hersteller haftet.

Erfüllungsgehilfe ist nach ständiger Rechtsprechung, wer mit dem Willen des Geschäftsherrn (= Schuldners) bei der Erfüllung der diesem obliegenden Verbindlichkeit als seine Hilfsperson tätig wird.<sup>24</sup> Diese Definition hilft aber bei der Verortung von KIA nicht weiter, da gerade unklar ist, welche Akteurstypen sich hinter dem „wer“ verbergen können. Zugeschnitten ist § 278 S. 1 BGB, und die Definition des Erfüllungsgehilfen, auf menschliche Akteure („Personen“). Die Klassifikation mo-

20 Grundmann (Fn. 6), Rn. 46; S. Lorenz, in: Bamberger/Roth/Hau/Poseck (Hrsg.), Beck'scher Online Kommentar BGB, 46. Aufl., München 2018, § 278 Rn. 17.

21 M. Wolf, Schuldnerhaftung bei Automatenversagen, JuS 1989, S. 899 (901); K. Spiro, Die Haftung für Erfüllungsgehilfen, Stämpfli & Cie, Bern 1984, S. 210 f.; C.-W. Canaris, Bankvertragsrecht, in: Staub, HGB, Berlin 2005, Rn. 367 (für Computerversagen); einschränkend Pfeiffer (Fn. 19) (Analogie nur, wenn nach Vertrag oder Verkehrserwartung Leistungserbringung durch Menschen zu erwarten war).

22 Lieser, Die zivilrechtliche Haftung im automatisierten Geschäftsverkehr, JZ 1971, S. 759 (761); H. Köhler, Die Problematik automatisierter Rechtsgänge, insbesondere von Willenserklärungen, AcP 182 (1982), S. 126 (168); G. Caspers, in: Staudinger, BGB, Berlin 2014, § 278 Rn. 5; D. Medicus/ S. Lorenz, Schuldrecht I: Allgemeiner Teil, 21. Aufl., München 2015, Rn. 386; Lorenz (Fn. 20), Rn. 17; siehe auch BGH NJW 1971, 32 (keine Staatshaftung für eine defekte Verkehrsampelanlage).

23 Vgl. BGH NJW 1978, S. 1157 (1158) (keine Pflicht des Werkunternehmers, von einem Dritten gelieferte und einzubauende Gegenstände zu untersuchen); ebenso BGH, BeckRS 1962, 31189816, unter 3c) (für defekte, eingebaute Heizung).

24 BGH NJW 1968, S. 1569; NJW 1998, S. 3342 (3343).



derner KI stellt daher eine Regelungslücke im Rahmen des § 278 BGB dar,<sup>25</sup> die einen Analogieschluss ermöglichen könnte; hinsichtlich der Vergleichbarkeit der maßgeblichen Wertungskriterien ist jedoch zu differenzieren.<sup>26</sup>

#### aa) Grade von Autonomie

Richtigerweise sollte die Abgrenzung nach dem Grad der Autonomie des KIA erfolgen, da dieser für die Unterscheidung von Personalrisiken einerseits (§ 278 BGB) und Produktrisiken andererseits (Produkt- und Produzentenhaftung) entscheidend ist.<sup>27</sup> Besitzt der KIA eigenständige, quasi-kognitive Fähigkeiten, die ein selbstständiges Navigieren und die Durchführung im Einzelnen nicht vorhersehbarer, adaptiver Aktionen ermöglichen, so ist die Grenze vom bloßen Werkzeug zu einer personenähnlichen Entität überschritten. Sachlich geht es dann nicht mehr (nur) um Produktrisiken, die vornehmlich dem Hersteller zugeordnet sind, sondern um Personalrisiken, die grundsätzlich nach der Anordnung des § 278 BGB der Geschäftsherr zu tragen hat.<sup>28</sup> Der Geschäftsherr verwendet dann den KIA selbst, und nicht den Benutzer eines komplexen Werkzeugs, zur Erfüllung seiner Schuldnerpflicht.<sup>29</sup>

Im Bereich des autonomen Fahrens sind etwa verschiedene Autonomiegrade etabliert worden, die ganz wesentlich davon abhängen, inwieweit der KIA seine Aufgaben ohne die Notwendigkeit des korrektiven Eingreifens von menschlichem Aufsichtspersonal erledigt und auf neuartige Situationen reagiert.<sup>30</sup> Diese Einteilung beruht auf einer allgemeinen Theorie der Autonomie von KI, die neben der Unabhängigkeit der Aktionen von menschlicher Kontrolle<sup>31</sup> auch die adaptive, eigenständige Wissenserweiterung zur Bewältigung neuer Aufgaben, mithin eine quasi-kognitive Lernkompetenz, umfasst.<sup>32</sup> Kurz gefasst geht es also um den Verbund

25 J.-P. Günther, Roboter und rechtliche Verantwortung, München 2016, S. 76 f.; J. Hanisch, Haftung für Automation, Göttingen 2010, S. 191 f.; T. Schulz, Verantwortlichkeit bei autonom agierenden Systemen, Baden-Baden 2015, S. 139; vgl. auch Grundmann (Fn. 6), § 278 Rn. 46.

26 Ohne Differenzierung gegen eine Analogie zu § 278 S. 1 BGB Günther (Fn. 25), S. 84, unter Verweis auf die (eher sachfremde) mangelnde gesellschaftliche Akzeptanz von Robotern.

27 Zu Überschneidungen, siehe unten, Fn. 44 und Text.

28 Siehe zum Kriterium des Personalrisikos bereits BGH NJW 1960, S. 669 (671).

29 Zur Relevanz der Erfüllung der Schuldnerpflicht Lorenz (Fn. 20), Rn. 26.

30 National Highway Traffic Safety Administration, Preliminary Statement of Policy Concerning Automated Vehicles, Washington D.C. 2013, S. 4 f. (mit 5 Stufen); SAE International's Standard J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems, Warrendale 2014 (mit 6 Stufen).

31 Siehe etwa G. Sartor, Contracts in the Infosphere, in: S. Grundmann (Hrsg.), European Contract Law in the Digital Age, Cambridge, Intersentia 2018, S. 263 (267 f.) („sustained independent action“); ähnlich Department of Defense, Defense Science Board, The Role of Autonomy in DoD Systems, Washington D.C., 2012, S. 1; G. Bekey, Current trends in robotics: technology and ethics, in: P. Lin et al. (Hrsg.), Robot Ethics, MIT Press, Cambridge, MA 2012, S. 17 (18); C. Tessier, Robots Autonomy: Some Technical Issues, in: W. Lawless et al. (Hrsg.), Autonomy and Artificial Intelligence: A Threat or Savior?, Springer, Cham 2017, S. 179 (180).

32 Russell/Norvig (Fn. 2), S. 39; Department of Defense, Defense Science Board, Summer Study on Autonomy, Washington D.C. 2016, S. 4; E. Alonso/E. Mondragón, Agency, Learning and Animal-Based Reinforcement Learning, in: M. Nickels (Hrsg.), Agents and Computational Autonomy, Berlin

unabhängiger Handlung mit selbstständigem Lernen. Diese Kriterien etablieren Grade an Autonomie,<sup>33</sup> die sich auch jenseits des autonomen Fahrens verhältnismäßig klar anwenden lassen. So wird man zwischen vollständig autonomen, mittelstark autonomen und schwach autonomen KIA unterscheiden müssen.<sup>34</sup> Damit einher geht eine Differenzierung zwischen einem sich aus weitestgehend selbständiger Handlung ableitenden Personalrisiko und einem einfachen Produktrisiko, das jedes maschinelle Produkt aufweist.

### (1) Vollständige Autonomie

Bei vollständig autonomen KIA muss keine menschliche Aufsicht mehr in der konkreten Situation geleistet werden; der Agent ist in seinen Aktionen von seiner Umgebung vollkommen unabhängig.<sup>35</sup> Jedenfalls hier ist die Personenähnlichkeit mit Blick auf § 278 S. 1 BGB grundsätzlich zu bejahen. Hervorzuheben ist jedoch, dass allein die Abwesenheit menschlicher Aufsicht beim Betrieb einer Maschine noch keine vollständige Autonomie impliziert. Vielmehr besteht diese nach obigen Kriterien erst dann, wenn sich der KIA selbstständig und adaptiv verhält, also die Möglichkeit besteht, dass er in unvorhersehbarer Weise (gute oder ungute) Entscheidungen trifft, die nicht im einzelnen regelbasiert prä-programmiert, sondern die Folge von veränderlichen internen Zuständen sind.<sup>36</sup> Dies unterscheidet ein vollständig autonomes System etwa von solchen, bereits heute gängigen Putzrobotern, die zwar ebenfalls selbstständig durch die Wohnung navigieren, dabei jedoch präskriptiv programmiert sind und keine eigenständigen Entscheidungen fällen.<sup>37</sup> Sie sind automatisiert, nicht autonom.<sup>38</sup> Erst letzteres Kriterium zeichnet das für § 278 S. 1 BGB typische Personalrisiko aus und stellt die Vergleichbarkeit zum menschlichen Erfüllungsgelhilfen her.

Unschädlich ist andererseits, dass technisch vollständig autonome Systeme bislang keine völlig humanäquivalente Autonomiekapazität erreichen, da sie ihre Ziele nicht selbst wählen.<sup>39</sup> Unabhängige Wahl und kritisch-reflektive Veränderung der eigenen Ziele (*second-order preferences*) sind zwar Signum gegenwärtiger philoso-

2005, S. 1 (2 f.); C. Castelfranchi/R. Falcone, Founding Autonomy: The Dialectics Between (Social) Environment and Agent's Architecture and Powers, in: ebd., S. 40 (40 f.); Sartor (Fn. 31), S. 269.

33 Department of Defense (Fn. 31), S. 4; Tessier (Fn. 31), S. 182.

34 Ähnlich Schulz (Fn. 25), S. 44.

35 Castelfranchi/Falcone (Fn. 32), S. 44.

36 Castelfranchi/Falcone (Fn. 32), S. 42; vgl. auch Günther (Fn. 25), S. 37 f.

37 Vgl. Russell/Norvig (Fn. 2), S. 39; Tessier (Fn. 31), S. 181.

38 Department of Defense (Fn. 32), S. 4.

39 Am nächsten kommt dem eine selbst gelernte Belohnungsfunktion (*reward function*) im Bereich *reinforcement learning*, die jedoch bislang auch durch menschliches Feedback (= Zielsetzung) geformt wird, siehe etwa P. Christiano et al., Deep reinforcement learning from human preferences, Advances in Neural Information Processing Systems 2017, S. 4302.

phischer Autonomiekonzepte;<sup>40</sup> relevantes Personalrisiko entsteht freilich bereits durch eigenständige, unvorhersehbare Handlungen im oben genannten Sinne, nicht erst durch eigenverantwortliche Zielsetzungen.

Andererseits sind auch bei vollständig autonomen Systemen nicht alle Aktionen über § 278 S. 1 BGB zurechenbar. Ist etwa ein vollkommen autonomer KIA schlecht verlötet und explodiert daher, so handelt es sich um die Realisierung von Produkt-, nicht Personalrisiko, das allein über das Produkthaftungsrecht zu lösen ist. Der Schutzzweck des § 278 S. 1 BGB ist dann nicht tangiert.

## (2) Schwache Autonomie

Schwach autonome KIA hingegen bedürfen der ständigen menschlichen Anleitung und Überprüfung, da ihr System noch nicht im ausreichenden Maße auf neue Situationen reagieren kann. Dies liegt vor allem an dem Mangel an adäquaten Trainingsdaten, mit denen das Modell auch für seltene oder neuartige Situationen kalibriert werden kann.

Bei schwach autonomen Systemen dürfte man noch nicht von einer Arbeitsteilung dahingehend sprechen, dass signifikante Entscheidungskompetenzen an den KIA delegiert wurden. Vielmehr ist hier der menschliche Überwacher geeigneter Anknüpfungspunkt für Zurechnungsfragen. Primär geht es hier also hinsichtlich des KIA um Produkt-, nicht um Personalrisiken, die aus eigenverantwortlicher Handlung entstehen. Daher ist eine analoge Anwendung von § 278 S. 1 BGB hier abzulehnen.

## (3) Mittelstarke Autonomie

Besonders diffizil ist die Zuordnung im Bereich der mittelstark autonomen KIA. Hier wird man ein Überwiegen des Personal- gegenüber dem Produktrisiko und damit die analoge Anwendbarkeit von § 278 S. 1 BGB auf KIA bejahen können, wenn ein menschliches Eingreifen zwar grundsätzlich noch möglich, aber nur ganz punktuell vorgesehen ist und somit der Großteil der Aufgaben, welche das Risiko einer Rechtsgutsverletzung in sich tragen, faktisch autonom erledigt werden. Dogmatisch lässt sich dies über die Figur des Gehilfengehilfen lösen, da die Nutzung des KIA vom Geschäftsherrn regelmäßig gerade gewollt war.<sup>41</sup> Operiert dahingegen ein KIA in einem Testbetrieb, bei dem zwar die Reaktionskompetenz in vielen Situationen bereits hoch ist, jedoch für eine erhebliche Anzahl von kritischen Konstellationen eine Korrektur durch einen menschlichen Begleiter notwendig ist, so

40 Siehe nur G. Diworkin, *The Theory and Practice of Autonomy*, Cambridge University Press, Cambridge 1988, S. 15–20; dazu P. Hacker, *Verhaltensökonomik und Normativität*, Tübingen 2017, S. 223 ff., besonders S. 229–232.

41 BGH NJW 1983, S. 448; NJW-RR 1988, S. 241 (243); NJW-RR 2012, S. 1316 Rn. 14; Grundmann (Fn. 6), Rn. 44.

dürfte im Zweifel die Ähnlichkeit zu einem komplexen Werkzeug größer, § 278 S. 1 BGB damit nicht einschlägig sein.

#### bb) Sinn und Zweck von § 278 BGB

In teleologischer Hinsicht macht es in den geschilderten Fällen hinreichender Autonomie Sinn, auf die strikte Zurechnung des § 278 S. 1 BGB zu rekurrieren. Ersetzt der KIA in funktionaler Betrachtung eine Person, die als Erfüllungsgehilfe zu qualifizieren wäre, so liegt genau solch ein Fall arbeitsteiligen Vorgehens vor, den § 278 BGB erfassen will.<sup>42</sup> Sofern man nachteilige, eigenständige Entscheidungen des KIA zugleich als Konstruktionsfehler im Sinne des § 3 ProdHG ansieht, ist das Personalrisiko zwar insoweit auch dem Hersteller zugewiesen.<sup>43</sup> Das ändert jedoch nichts daran, dass in vertraglichen Beziehungen nach dem Sinn und Zweck des § 278 BGB das Personalrisiko, also das Risiko unvorhersehbarer, eigenständiger Entscheidungen der Hilfskraft, sei es einer menschlichen oder maschinellen, *jedenfalls auch* vom Geschäftsherrn, als dem ausgewählten Vertragspartner, zu tragen ist.<sup>44</sup> Irrelevant ist auch, ob ein Rechtsverhältnis zwischen Geschäftsherrn und Delegat besteht;<sup>45</sup> insofern ist die mangelnde Rechtspersönlichkeit des KIA unerheblich, entscheidend die tatsächliche Verlagerung des Entscheidungszentrums weg vom Geschäftsherrn.

Weniger überzeugend ist in diesem Kontext allerdings der alleinige Rekurs auf das marktwirtschaftliche Prinzip der Korrespondenz von Vor- und Nachteil als Zurechnungsgrund.<sup>46</sup> Macht sich der Geschäftsherr die autonomen Entscheidungsfähigkeiten der Hilfskraft zur Erweiterung seiner Produktionsmöglichkeiten zunutze, so das Argument, soll er auch für diejenigen Schäden einstehen, die aus einer aus Sicht des Vertragspartners suboptimalen Nutzung dieser Autonomieressourcen resultieren. Allerdings hat die Übertragung von Aufgaben an KIA unter Wettbewerbsbedingungen auch für den Gläubiger (Preis- und weitere Nutzens-)Vorteile,<sup>47</sup> so dass die Anknüpfung allein an Vorteile unscharf ist.

42 Zur Delegation von Arbeit als maßgeblichem Gesichtspunkt siehe auch *Grundmann* (Fn. 6), Rn. 43; siehe auch *Spiro* (Fn. 21), S. 210.

43 Siehe unten Fn. 80 und Text.

44 Vgl., außerhalb von KI, *S. Grundmann* (Fn. 6), Rn. 3; siehe auch unten Fn. 70 und Text; aA wohl *Schulz* (Fn. 25), S. 140 f., der KI-Risiken ausschließlich als Produktrisiken ansieht.

45 BGH NJW 1968, S. 1569; NJW 1998, S. 3342 (3343); *Köhler* (Fn. 22).

46 So aber *Wolf* (Fn. 21), S. 901; *B. Kupisch*, Die Haftung für Erfüllungsgehilfen (§ 278 BGB), JuS 1983, S. 817 (819); *C. Schubert*, in: F. J. Säcker et al. (Hrsg.), Münchener Kommentar BGB, Band 1, 7. Aufl., München 2015, § 166 Rn. 47; *Köhler* (Fn. 22); *Buck* (Fn. 1), S. 314 f.; *M. Baum*, Die Wissenszurechnung, Berlin 1999, S. 64 f.; *E. Schilken*, Wissenszurechnung im Zivilrecht, Bielefeld 1983, S. 224; *M. Brunner*, Zum Risiko von Computerfehlleistungen bei der Abwicklung von Verträgen, München 1970, S. 98 f.

47 *W. Möschel*, Dogmatische Strukturen des bargeldlosen Zahlungsverkehrs, AcP 186 (1986), S. 187 (199 f.); vgl. auch *Buck* (Fn. 1), S. 321 (im parallelen Kontext der Wissenszurechnung); daher eine Analogie zu § 278 BGB ablehnend *Hanisch* (Fn. 25), S. 194; ähnlich *Günther* (Fn. 25), S. 80, jeweils unter Rekurs auf *Spiro* (Fn. 21), S. 60 f.

Maßgeblich ist vielmehr, dass der Geschäftsherr Nutzen aus einem KIA-spezifischen Personalrisiko zieht, das er zwar, wie bei menschlichem Personal auch, nicht *vollständig* kontrollieren, aber regelmäßig *besser* beherrschen kann als der Vertragspartner (*least cost avoider*).<sup>48</sup> Die Zurechnung setzt dann aus ökonomischer Sicht sinnvolle Anreize, diese Kontrolle auch, in ressourcenoptimierender Weise, auszuüben.<sup>49</sup> Die kann nicht nur durch die (ohnehin bereits nach § 276 BGB geschuldete) sorgfältige KIA-Auswahl geschehen, sondern etwa auch durch eine Einpreisung von Haftungsschäden in eine Belohnungsfunktion (*reward function engineering*),<sup>50</sup> welche den KIA adaptiv lenkt,<sup>51</sup> oder auch die regelmäßige technische Überprüfung des situationsadäquaten Funktionierens des KIA (etwa durch algorithmische *audits*).<sup>52</sup>

In denjenigen Fällen hingegen, in denen die Überwachung und korrektive Tätigkeit des menschlichen Aufsichtsakteurs überwiegt, kann und muss (lediglich) dessen etwaige Aufsichtspflichtverletzung nach § 278 S. 1 BGB zugerechnet werden. Das einfache Produktrisiko überwiegt dann das Personalrisiko (in Bezug auf den KIA) derartig, dass neben die Herstellerhaftung nicht noch eine Haftung des Geschäftsherrn treten kann. Daher verfängt in diesem Fall auch eine von § 278 BGB losgelöste Zurechnung zur Risikosphäre des Geschäftsherrn analog § 645 BGB, sofern man die Sphärentheorie überhaupt anerkennen wollte,<sup>53</sup> letztlich nicht.

Im Ergebnis ist daher der KIA selbst genau dann taugliches Zurechnungsausgangssubjekt im Rahmen des § 278 S. 1 BGB, analog einem menschlichen Erfüllungsgehilfen, wenn er dergestalt autonom agiert, dass eine Korrektur durch einen menschlichen Akteur gar nicht oder lediglich ganz punktuell erfolgt.

48 R. Posner, *Economic Analysis of Law*, 9. Aufl. Wolters Kluwer, New York 2014, § 6.8, S. 218; Spiro (Fn. 21), S. 71 f.; Günther (Fn. 25), S. 81 f.; Möschel (Fn. 47), S. 200; A. Wiebe, Die elektronische Willenserklärung, Tübingen 2002, S. 190; im Kontext von KI Spindler (Fn. 11), S. 767.

49 Die Lenkungsmöglichkeiten gehen dabei über das bereits nach §§ 241, 276 BGB *Geschuldete* hinaus, vgl. Spiro (Fn. 21), S. 71 f. Der Geschäftsherr kann diese überobligatorischen Möglichkeiten nutzen und das Haftungsrisiko reduzieren, muss es jedoch nicht (Entscheidung nach Grenzkosten/Grenznutzen).

50 Dies ist wohl das beste Äquivalent zur Instruktion des Erfüllungsgehilfen durch den Geschäftsherrn; siehe etwa P. Christiano, *Thoughts on reward engineering*, AI Alignment, 8. November 2016, <https://ai-alignment.com/thoughts-on-reward-engineering-82b193ec03f6>; A. Agrawal et al., *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, Boston 2018, Kapitel 8; T. White/S. Baum, *Liability for Present and Future Robotics Technology*, in: P. Lin et al., *Robot Ethics 2.0*, Oxford University Press, Oxford 2017, S. 66 (71).

51 Dies gilt für KIA, die, wie die meisten heutzutage besonders erfolgreichen KI-Systeme, auf *deep reinforcement learning* basieren, siehe Y. Li, *Deep Reinforcement Learning: An Overview*, Working Paper (2017), <https://arxiv.org/abs/1701.07274>; zur *reward function* dort, S. 19.

52 Siehe dazu P. Kim, *Auditing Algorithms for Discrimination*, 166 *University of Pennsylvania Law Review Online* 2017, S. 189; P. Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, *Common Market Law Review* (im Erscheinen), <https://ssrn.com/abstract=3164973>, unter 4.2.

53 Dazu ausführlich H. Köhler, *Unmöglichkeit und Geschäftsgrundlage bei Zweckstörungen im Schuldverhältnis*, München 1971, S. 41 ff.

## 2. Verschulden bei KI-Agenten

Vom Status des Erfüllungsgehilfen strikt zu trennen ist nach der gesetzlichen Systematik die Frage von dessen Verschulden. Nach § 278 S. 1 BGB hat der Schuldner ein Verschulden seines Erfüllungsgehilfen in gleichem Umfang zu vertreten wie eigenes Verschulden. Grundsätzlich werden (über den Wortlaut hinaus) sowohl die Handlung als auch das Verschulden des Erfüllungsgehilfen dem Geschäftsherrn zugerechnet. Jedoch ist es nicht trivial, den Begriff des Verschuldens an KIA anzupassen.<sup>54</sup> Grundsätzlich bestehen hier drei Möglichkeiten: das Postulat der Inkompatibilität; der Verzicht auf Verschulden; oder die Konstruktion von funktionalen Verschuldensäquivalenten.

### a) Inkompatibilität von KI-Agent und Verschulden

Erstens kann man mit einer starken Strömung der Literatur die analoge Anwendung des § 278 S. 1 BGB an dieser Stelle scheitern lassen mit dem Argument, dass einem KIA grundsätzlich kein Verschulden zur Last gelegt werden kann.<sup>55</sup> Die zentrale, zumeist unausgesprochene Grundlage dieser Position dürfte darin zu finden sein, dass regelmäßig für den KIA keine gleichwertigen, realistischen Handlungsalternativen bestehen. Auch wenn ein KIA nicht deterministisch programmiert ist, sondern adaptiv auf neue Situationen reagiert, so wird er doch letztlich anhand einer mathematischen Funktion, die während des Trainings erstellt wurde, die aus seiner Perspektive optimale Handlung auswählen.<sup>56</sup> Schädigende Handlungen sind daher für den KIA per definitionem nicht vorhersehbar und vermeidbar,<sup>57</sup> da grundsätzlich die aus seiner Perspektive bestmögliche Aktion durchgeführt wird. Denn der KIA kann nicht in sorgfaltswidriger Weise abgelenkt sein oder sonst eine vermeidbar schlechtere Handlungsalternative wählen; allenfalls könnte man dem Programmierer den Vorwurf machen, bestimmte kritische Situationen nicht ausrei-

54 Je nach Handlungsbegriff stellen sich dort ähnliche Probleme. Handlung ist nach der Rechtsprechung „ein der Steuerung durch Bewusstsein und Willen unterliegendes und insofern grundsätzlich beherrschbares menschliches Verhalten“ (OLG Stuttgart NJW-RR 2017, S. 860 Rn. 15; siehe auch BGH NJW 1987, S. 121). Dass die Aktion nicht menschlichen Ursprungs ist, wird durch den Analogieschluss hinsichtlich des Zurechnungsausgangssubjekts gerade überwunden; zur Frage des Willens und seiner Freiheit sogleich, im nächsten Abschnitt.

55 Caspers (Fn. 22); Köhler (Fn. 22); Medicus/Lorenz (Fn. 22); vgl. auch Lieser (Fn. 22), S. 761, alle jedoch (wohl) mit Bezug auf nicht-intelligente automatisierte Systeme; spezifisch mit KI-Bezug S. Horner/M. Kaulartz, CR 2016, S. 7 (7); F. Möslin, Digitalisierung im Gesellschaftsrecht: Unternehmensleitung durch Algorithmen und künstliche Intelligenz?, ZIP 2018, S. 204 (210); wohl auch ablehnend C. Müller-Hengstenberg/S. Kirn, Intelligente (Software-)Agenten: Eine neue Herausforderung unseres Rechtssystems, MMR 2014, S. 307 (311).

56 Die geschieht Minimierung einer Verlustfunktion (*supervised learning*), siehe etwa Russell/Norvig (Fn. 2), S. 710-712; I. Goodfellow et al., Deep Learning, MIT Press, Cambridge, MA 2016, S. 79 ff.; bzw. durch Maximierung der Belohnungsfunktion (*reinforcement learning*), siehe oben, Fn. 50.

57 Explizit dies (für Maschinen) betonend Köhler (Fn. 22).

chend trainiert zu haben.<sup>58</sup> Mit einer Anwendung von § 278 BGB auf Maschinen wäre demnach auch die Grenze der Analogiefähigkeit überschritten.<sup>59</sup>

Dagegen lässt sich allerdings trefflich einwenden, dass diese Grenzüberschreitung für nicht-intelligente Systeme gelten mag;<sup>60</sup> für hinreichend autonome Systeme hingegen ist eine Analogie nicht a priori ausgeschlossen, wenn funktionale Verschuldensäquivalente bei diesen Systemen etabliert werden können (dazu unter c)). Auch lassen sich augenfällige Parallelen zwischen menschlicher und hinreichend autonomer KIA-Handlung feststellen, anders als im Kontext von nicht-intelligenten Maschinen. Bei solchen KIA, bei denen aufgrund ihrer fortgeschrittenen Autonomie nach dem oben Gesagten ein Analogieschluss überhaupt in Betracht kommt, sind quasi-kognitive Komponenten weit entwickelt. Ähnlich wie beim Menschen ist ihr Handeln aus einer Außenperspektive nicht vorhersehbar.<sup>61</sup> Dies schließt jedoch eine Determination durch interne Zustände nicht aus.<sup>62</sup> Inwiefern nun Handlungen intern frei von neuronaler Vorbestimmung sind, ist hingegen durchaus auch beim Menschen überaus umstritten. Dies ist nicht der Ort, um die Bände füllende Debatte um die Möglichkeiten und Grenzen des freien menschlichen Willens zu eröffnen oder gar zu entscheiden.<sup>63</sup> Es ist jedoch nicht ohne Bedeutung, dass von namhaften Autoren, vor allem im Strafrecht, die Verschuldensfähigkeit des Menschen im Sinne einer alternativen Handlungsmöglichkeit letztlich als unbewiesene normative Setzung betrachtet wird.<sup>64</sup> Dieses Andershandelnkönnen ist auch im Zi-

58 Siehe zu derartigen Trainingsmöglichkeiten etwa S. Greengard, *Gaming Machine Learning*, 60(12) *Communications of the ACM* 2017, S. 14.

59 *Lieser* (Fn. 22), S. 761; aA *Spiro* (Fn. 21), S. 211 mit Fn. 7.

60 Darauf bezogen auch die Wertung von *Lieser* (Fn. 22), S. 761.

61 Zur externen Unvorhersehbarkeit (aufgrund von Komplexität) als Signum menschlichen Handelns etwa W. Singer, Wann und warum erscheinen uns Entscheidungen als frei?, in: T. Stompe/H. Schanda (Hrsg.), *Der freie Wille und die Schuldfähigkeit*, Berlin 2010, S. 15 (16 f.); zu KIA *Castelfranchi/Falcone* (Fn. 32), S. 42.

62 *Ibid.*

63 Monographisch etwa R. Merkel, *Willensfreiheit und rechtliche Schuld*, Baden-Baden 2008; G. Seebaß, *Willensfreiheit und Determinismus*, Berlin 2008; S. Walter, *Illusion freier Wille?*, Stuttgart 2016; überblicksartig K. Smith, *Neuroscience vs philosophy: Taking aim at free will*, 477 *Nature* 2011, S. 23; J. Adam et al., *Nulla poena sine culpa – Was besagt das verfassungsrechtliche Schuldprinzip?*, *NStZ* 2017, S. 7.

64 Siehe etwa C. Roxin *Strafrecht AT I*, 4. Aufl. 2006, § 3 Rn. 55; J. Eisele, in: Schönke/Schröder, *Strafgesetzbuch*, 29. Aufl., München 2014, Vorbem. zu §§ 13 ff. Rn. 110; Merkel (Fn. 63), S. 134; K. Burns/A. Bechara, *Decision making and free will: a neuroscience perspective*, 25 *Behavioral Sciences & the Law* 2007, S. 263 (274); dezidiert gegen die Annahme alternativer Handlungsmöglichkeit beim Menschen auch G. Merkel/G. Roth, *Hirnforschung, Gewalt und Strafe*, in: T. Stompe/H. Schanda (Hrsg.), *Der freie Wille und die Schuldfähigkeit*, Berlin 2010, S. 143 (148 f.; 152-155, 159 f.); Singer (Fn. 61), S. 19-21; Merkel (Fn. 63), S. 121; nuancierter M. Pauen/G. Roth, *Freiheit, Schuld und Verantwortung*, Frankfurt a.M. 2008, S. 46 ff., S. 134 ff., insbesondere S. 167 f. (Selbstbestimmung als determinationsunabhängige Kategorie von Freiheit und Grundlage von Schuld).

vilrecht Kern des Verschuldensvorwurfs.<sup>65</sup> Eine derartige, auf gleich im Einzelnen darzustellenden Sachgründen aufruhende Setzung lässt sich auch bei KIA vornehmen, deren Entscheidungen ebenfalls auf adaptiven, aber determinierten internen Zuständen beruhen.<sup>66</sup> Dies gilt insbesondere dann, wenn dadurch ein ausgewogenes Sanktionssystem zur Erfassung „delegierter Pflichtverletzung“, also der Verwirklichung vertraglich relevanten Personalrisikos, im KI-Bereich erst ermöglicht wird.<sup>67</sup>

In diesem Zusammenhang lässt sich gegen die die Inkompatibilitätsthese schließlich vorbringen, dass die Alternativen zu einer analogen Anwendung des § 278 S. 1 BGB hinsichtlich des Verschuldens nach geltendem Recht unbefriedigend sind. So würde die Haftung des Schuldners auf eine Eigenhaftung für Auswahl-, Instruktions- und Überwachungsverschulden nach § 276 BGB reduziert, welche die gleichen Lücken wie § 831 Abs. 1 BGB aufweist.<sup>68</sup> Der Verweis auf die Produkthaftung dürfte ebenfalls aus Verkehrsschutzgesichtspunkten, gerade im vertraglichen Bereich, nicht ausreichen. Denn hier ist zum einen unklar, wie der Fehlerbegriff im Rahmen von KIA auszulegen ist;<sup>69</sup> zum anderen wird dem Geschädigten ein anderer Schuldner als derjenige, den er sich im Rahmen der Sonderverbindung ausgesucht hat, aufgedrängt, mit nicht nur anders gelagerten Rechtsdurchsetzungskosten, sondern auch differenten Insolvenzrisiken.<sup>70</sup> Genau dies soll aber § 278 S. 1 BGB anerkanntermaßen verhindern.<sup>71</sup>

## b) Garantiehaftung

Ein zweiter, diametral entgegengesetzter Weg bestünde daher darin, auf das Verschuldenserfordernis bei KIA im Rahmen des § 278 S. 1 BGB schlichtweg zu verzichten und damit *de facto* eine Gefährdungshaftung für KIA im vertraglichen Bereich zu etablieren.<sup>72</sup> Dies ginge jedoch über die gesetzliche Systematik des § 278 S. 1 BGB und die darin explizit normierte Verschuldenshaftung für Erfüllungsgehil-

65 OLG Stuttgart NJW-RR 2017, S. 860 Rn. 15 („personale Verantwortung“); G. Wagner, in: F. J. Säcker et al. (Hrsg.), Münchener Kommentar BGB, Band 6, 7. Aufl., München 2017, § 827 Rn. 1; G. Spindler, in: Bamberger/Roth/Hau/Poseck (Hrsg.), Beck'scher Online Kommentar BGB, 46. Aufl., München 2018, § 827 Rn. 1; siehe auch BGH NJW 1968, S. 1132 (1133).

66 Zur Vergleichbarkeit von natürlich und durch Programmierung determinierten Personen, siehe auch Merkel (Fn. 63), S. 104 f.; aA Schulz (Fn. 25), S. 50-52; zur Kritik der normativen Setzung Merkel (Fn. 63), S. 118 und 134.

67 Zu diesem Begründungsstrang von Sanktion angesichts neuronaler Determiniertheit Pauen/Roth (Fn. 64), S. 150 ff.; Merkel (Fn. 63), S. 125 ff.

68 Vgl. Wolf (Fn. 21), S. 900; Günther (Fn. 25), S. 81.

69 Siehe etwa Čerka et al. (Fn. 11), S. 386; Vladeck (Fn. 11), S. 142 ff.; Reed et al. (Fn. 11), S. 6; Wagner (Fn. 11), S. 724 ff.

70 Vgl. R. Schmidt, Rationalisierung und Privatrecht, AcP 166 (1966), S. 1 (24).

71 Spiro (Fn. 21), S. 52-54.

72 So bereits Schmidt (Fn. 72); wohl auch Wiebe (Fn. 47); neuerdings etwa Spindler (Fn. 11), S. 775 f.; Horner/Kaulartz (Fn. 55), S. 13 f.; für den deliktischen Bereich, siehe etwa Vladeck (Fn. 11), S. 146-150.



fen deutlich hinaus. Zur Entscheidung, innerhalb von Sonderbeziehungen den Schuldner nicht mehr nur für (zugerechnetes) Verschulden oder funktionale Äquivalente, sondern strikt haften zu lassen, ist der Gesetzgeber, nicht der Rechtsausleger oder -fortbilder berufen.<sup>73</sup> Eine Diskussion der Vor- und Nachteile einer derartigen Garantiehaftung *de lege ferenda* würde den Rahmen des Beitrags sprengen.<sup>74</sup> Wenn überhaupt, so kommt eine strikte Haftung jedoch wohl nur für den Hersteller, nicht für den Verwender von KIA in Betracht, da jener die insgesamt besten Einwirkungsmöglichkeiten zur Kontrollierung von Risiken in der Phase der Entwicklung und Kalibrierung des algorithmischen Modells hat.<sup>75</sup> Demgegenüber hat der konkrete Verwender (der Geschäftsherr im Rahmen von § 278 BGB) lediglich reduzierte Einwirkungsmöglichkeiten, die auch nur eine (auf bestimmte, verschuldensäquivalente Fälle beschränkte) Haftung auslösen sollten.

### c) Funktionale Verschuldensäquivalente

Vorzugswürdig erscheint daher im Rahmen der Rechtsfortbildung die dritte Möglichkeit, an funktionale Verschuldensäquivalente bei KIA anzuknüpfen. Dieser Weg ist insbesondere einem Rückgriff auf § 242 BGB überlegen, den etwa *Lieser* in Folge der Verschuldensproblematik als Alternative zu einer Analogie von § 278 S. 1 BGB für nicht-intelligente Maschinen propagiert,<sup>76</sup> da er sachnäher, vorhersehbarer und damit auch einfacher versicherbar ist.<sup>77</sup> Als Verschuldensäquivalente kommen letztlich zwei Lösungen in Betracht: eine anthropo-parallele und eine KI-spezifische.

Erstere besteht darin, Verschulden anzunehmen, sofern die Aktion des KIA, als Handlung des menschlichen Geschäftsherrn<sup>78</sup> gedacht, sorgfaltswidrig wäre. Letztlich bedingt dies eine Haftung vor allem für solche Fehler, für die dem äußeren Erscheinungsbild nach auch ein Mensch haften würde. Die KI-spezifische Lösung hingegen knüpft an ein Zurückbleiben hinter technischen Standards für die Entwicklung von KIA an.<sup>79</sup> Insofern ähneln die beiden Möglichkeiten jenen, die zur Bestimmung des Fehlers eines autonomen Produkts im Rahmen der Produkthaftung diskutiert werden.<sup>80</sup> Dies überrascht einerseits nicht, da der produkthaftungs-

73 *Wolf* (Fn. 21), S. 901; *Lieser* (Fn. 22), S. 761; *Möschel* (Fn. 47), S. 198.

74 Siehe etwa die Diskussion bei *Wolf* (Fn. 21), S. 900 f.; kritisch *M. Denga*, Deliktische Haftung für künstliche Intelligenz, CR 2018, S. 69 (76 f.); siehe auch Fn. 72.

75 So im Ergebnis auch *Vladeck* (Fn. 11), S. 146-150.

76 *Lieser* (Fn. 22), S. 765.

77 Vgl. auch *Köhler* (Fn. 22).

78 Bei Gesellschaften: ihrer Organe.

79 Vgl. bereits *Wolf* (Fn. 21), S. 902; ferner *J.-P. Günther*, Embodied Robots, in: M.-C. Gruber/J. Bung/S. Ziemann (Hrsg.), *Autonome Automaten*, München 2015, S. 155 (164); zur Bedeutung technischer Standards im Produkthaftungsrecht *C. Förster*, in: *Bamberger/Roth/Hau/Poseck* (Hrsg.), *Beck'scher Online Kommentar BGB*, 46. Aufl., München 2018, § 823 Rn. 683 ff.

80 *Wagner* (Fn. 11), S. 730 ff., besonders S. 733-740; für einen anthropo-parallelen Maßstab *G. Borges*, Haftung für selbstfahrende Autos, CR 2016, S. 272 (275 f.); *Gomille* (Fn. 11), S. 77.

rechtliche Fehlerbegriff letztlich erheblich durch Sorgfaltsanforderungen aufgeladen wurde; andererseits ist jedoch zu beachten, dass, anders als der zumindest im Ausgangspunkt produktbezogene Fehlerbegriff, Verschulden originär auf menschliche Handlung bezogen ist. Zu fragen ist daher, ob die genannten Varianten tatsächlich als funktionale Verschuldensäquivalente angesehen werden können. Dies richtet sich nach dem Sinn und Zweck des Verschuldenserfordernisses, das für den Analogieschluss insoweit ohnehin maßgeblich ist.

#### aa) Pönalisierung

Zunächst könnte man dem Verschuldenserfordernis eine Pönalisierungsfunktion beilegen,<sup>81</sup> gespeist aus der semantischen und ideengeschichtlichen Nähe des Verschuldens zur Schuld (*culpa*).<sup>82</sup> Ein solcher Strafcharakter des Haftungsrechts als Vergeltung für Unrecht wird jedoch mehrheitlich abgelehnt<sup>83</sup> und ist jedenfalls bei der Haftung für fremdes Verschulden schon im Ausgangspunkt verfehlt. Der Geschäftsherr selbst hat sich gar nichts zu Schulden kommen lassen, es bedarf insoweit also schon keiner Bestrafung. Der Erfüllungsgehilfe hingegen muss nach wohl herrschender Meinung nicht notwendig verschuldensfähig, und damit überhaupt dem Grunde nach pönalisierbar, sein.<sup>84</sup> Dies führt insbesondere dazu, dass eine Fremdhafung des Zurechnungsendsubjekts kraft Zurechnung bejaht werden kann, auch wenn eine Eigenhaftung des Zurechnungsausgangssubjekts verneint wird. Im Fall von KIA ist dies *de lege lata* insbesondere deshalb relevant, weil eine Eigenhaftung des KIA mangels Rechtspersönlichkeit ausscheidet.<sup>85</sup> Eine solche gegenüber der Eigenhaftung erweiterte Fremdhafung ist aber auch sachgerecht, da

81 Vgl. *Lieser* (Fn. 22), S. 763 (im Kontext von EDV).

82 Siehe etwa *R. v. Jhering*, Das Schuldmoment im Römischen Privatrecht, Gießen 1867, S. 4 ff., besonders S. 6 f.; *N. Jansen*, Die Struktur des Haftungsrechts, Tübingen 2003, S. 280 ff.

83 Siehe etwa *G. Wagner*, Neue Perspektiven im Schadensersatzrecht-Kommerzialisierung, Strafschadensersatz, Kollektivschaden: Gutachten A für den 66. Deutschen Juristentag, München 2006, S. A 15 und S. A 77; *D. Medicus*, Neue Perspektiven im Schadensersatzrecht – Kommerzialisierung, Strafschadensersatz, Kollektivschaden, JZ 2006, S. 805 (809); *H. Koziol*, Die Objektivierung des Fahrlässigkeitsmaßstabs im Schadensersatzrecht?, AcP 196 (1996), S. 593 (602); nicht zu verwechseln mit dem auf Pflichtverletzungen aufbauenden Sanktionsmodell, siehe dazu *Jansen* (Fn. 82), S. 42, 47 ff.

84 *E. Schmidt*, Zur Dogmatik des § 278, AcP 170 (1970), S. 502 (511 f.); *Canaris* (Fn. 21); *Möschel* (Fn. 47), S. 198–200; *Wolf* (Fn. 21), S. 901 f.; *Spiro* (Fn. 21), S. 211 Fn. 7; *Günther* (Fn. 25), S. 83; *K. Larenz*, Lehrbuch des Schuldrechts: Allgemeiner Teil, Band I, 14. Aufl., München 1987, § 20 VIII, S. 303 f.; *Kupisch* (Fn. 46), S. 821; *Wiebe* (Fn. 47); siehe auch BGH NJW 1960, S. 669 (671) (grundsätzliche Geltung des Sorgfaltsmaßstab des Geschäftsherrn); aA OLG Düsseldorf NJW-RR 1995, S. 1165 (1166) (keine Haftung für verschuldensunfähige Erfüllungsgehilfen); ebenso BayObLG NJW 1970, S. 1550 (1554); *Lorenz* (Fn. 20), Rn. 48; *A. Stadler*, in: R. Stürmer, Jauernig. Bürgerliches Gesetzbuch, 16. Aufl., München 2015, § 278 Rn. 13; *E. Lorenz*, Die Haftung für Erfüllungsgehilfen, in: C.-W. Canaris/A. Heldrich (Hrsg.), 50 Jahre Bundesgerichtshof, Band 1: Bürgerliches Recht, München 2000, S. 329 (377 f.).

85 Siehe oben, Fn. 6 am Ende.

der Gläubiger regelmäßig den Schuldner, nicht aber die Gehilfen (seien sie nun menschliche oder KIA) auswählen kann.<sup>86</sup>

### bb) Effizienz

Auch eine Effizienzfunktion (im Sinne der Steigerung von Kaldor-Hicks-Effizienz)<sup>87</sup> lässt sich am Verschuldenserfordernis kaum belegen. Grundsätzlich wird Effizienz durch eine strikte Haftung besser gewährleistet als durch eine Verschuldenshaftung.<sup>88</sup> Denn erstere setzt im Ergebnis die gleichen Sorgfaltsanreize für den potentiellen Schädiger, kontrolliert aber, anders als die Verschuldenshaftung, zusätzlich noch dessen Aktivitätsniveau, da er sich nicht durch effiziente Sorgfaltsanstrengungen vor Haftung für zusätzliche Aktivitäten schützen kann.

### cc) Haftungsbegrenzung

Letztlich kommt dem Verschuldenserfordernis daher eine Haftungsbegrenzungsfunktion, hinsichtlich der Haftung dem Grunde nach, zu,<sup>89</sup> die gerade bei einer Ausweitung der Haftung auf weitere Zurechnungsausgangssubjekte fundamental ist. Daher wird § 276 BGB von der herrschenden Meinung richtigerweise auch als Zurechnungsnorm verstanden.<sup>90</sup> Eine solche Haftungsbegrenzung können jedoch beide vorgeschlagenen Verschuldensäquivalente durchaus leisten. Es bestehen jedoch Unterschiede und Probleme im Detail.

### (1) Anthro-Parallelität

Die anthro-parallele Lösung hat den Vorzug, sich in puncto Haftungsbegrenzung sehr eng an das Verschuldenserfordernis bei Sachverhalten mit menschlichen Erfüllungsgehilfen anzulehnen. Typischerweise würden ähnliche Sachverhalte haftungsmäßig erfasst, unabhängig davon, ob ein Mensch oder ein (hinreichend autonomer) KIA den Schaden verursacht. Aus dogmatischer Sicht ist dies nur konsequent, da auch bei menschlichen Erfüllungsgehilfen regelmäßig im Rahmen des § 278 S. 1 BGB eine Als-Ob-Betrachtung angestellt werden muss: Erfüllungsgehilfen selbst unterliegen keiner Pflicht aus der Sonderverbindung zwischen Schuldner und Gläubiger und ihre Handlung kann daher nur insoweit, als sie als Handlung

<sup>86</sup> Grundmann (Fn. 6), Rn. 50.

<sup>87</sup> H. Eidenmüller, Effizienz als Rechtsprinzip, Tübingen 1995, S. 51 ff.; Hacker (Fn. 40), S. 307 ff.

<sup>88</sup> Grundlegend S. Shavell, Strict Liability versus Negligence, 9 The Journal of Legal Studies 1980, S. 1 (2 f.); R. Cooter/T. Ulen, Law § Economics, 6. Aufl., Addison-Wesley, Boston 2012, S. 200 ff.; Jansen (Fn. 82), S. 154 f.; S. Grundmann, in: F. J. Sackler et al. (Hrsg.), Münchener Kommentar BGB, Band 2, 7. Aufl., München 2016, § 276 Rn. 32 f.

<sup>89</sup> Für das Strafrecht etwa Roxin (Fn. 64), Rn. 51-53; Merkel/Roth (Fn. 64), S. 155; für das Zivilrecht, Grundmann (Fn. 88), Rn. 8; Spiro (Fn. 21), S. 79; Brunner (Fn. 46); vgl. auch T. Dreier, Kompensation und Prävention, Tübingen 2002, S. 23; G. Caspers, in: Staudinger, BGB, Berlin 2014, § 276 Rn. 7-9.

<sup>90</sup> Grundmann (Fn. 88), Rn. 1; Caspers (Fn. 89), Rn. 3; W. Ernst, in: F. J. Sackler et al. (Hrsg.), Münchener Kommentar BGB, Band 2, 7. Aufl., München 2016, § 280 Rn. 23.

des Geschäftsherrn gedacht wird, eine schuldhafte Pflichtverletzung darstellen.<sup>91</sup> Es entspricht somit der hergebrachten Funktionsweise des § 278 S. 1 BGB, dass der Geschäftsherr grundsätzlich für solche Handlungen haftet, für die er nach §§ 276 f. BGB verantwortlich wäre, hätte er sie selbst ausgeführt.<sup>92</sup> Eine derartige, typisierende Als-Ob-Betrachtung lässt sich ohne Weiteres auch bei KIA anstellen.

Im Rahmen der Frage, ob der Schuldner die konkrete Pflichtverletzung schuldhaft begangen hätte, muss freilich auf einen typisierten Verschuldensvorwurf abgehoben werden, da weder der Schuldner noch sonst ein Mensch *in concreto* gehandelt hat. Dies ist dem Recht jedoch keineswegs fremd: Auch sonst indiziert gerade im Bereich der hier vor allem in Frage stehenden Fahrlässigkeitshaftung die objektive Pflichtwidrigkeit das subjektive Verschuldenselement<sup>93</sup> und ist der Sorgfaltsmaßstab nach § 276 Abs. 2 BGB der „im Verkehr erforderliche“, also ein objektiver.<sup>94</sup> Lediglich auf das im Fahrlässigkeitsbereich ohnehin im deutschen Privatrecht marginalisierte subjektive Verschuldenselement,<sup>95</sup> das auch sonst typischerweise schwer nachzuweisen ist, muss im Rahmen von KIA tatsächlich bei der Als-Ob-Betrachtung verzichtet werden. Dies führt jedoch letztlich nur die von der herrschenden Meinung<sup>96</sup> befürwortete, freilich durch die Einteilung in Verkehrskreise moderierte, Objektivierung des Verschuldensmaßstabs konsequent fort und ist auch insofern verschmerzbar, als die anthropo-parallele Lösung einerseits die sachgerechte Zuweisung von Risiken<sup>97</sup> ermöglicht und andererseits ihrer Begrenzungsfunktion dennoch gerecht wird.

Befindet sich etwa in einer Wohnung eine „Wand“, die in täuschend echter Weise aus dünner Folie eine verputzte Steinwand imitiert, und die durch den KIA beschädigt wird, weil er diese in einer Weise touchiert, die bei einer Steinwand unschädlich gewesen wäre, aber zu einem Einreißen der Folie führt, so ist eine verschuldensäquivalente Pflichtverletzung abzulehnen, wenn die Folieneigenschaft auch für einen sorgfältigen Menschen typischerweise nicht erkennbar gewesen wäre. Anders steht es jedoch, wenn der KIA, wie im Ausgangsfall, die Distanz zu einem Tisch falsch berechnet. Das Umstoßen eines gut sichtbaren Tisches mit der Folge, dass die darauf befindliche Vase zu Bruch geht, ist bei einer menschlichen Reinigungskraft typischerweise sorgfaltswidrig; insofern ist nach der anthropo-parallelen Lösung eine verschuldensäquivalente Pflichtverletzung zu bejahen. Zu fragen ist also

91 Larenz (Fn. 84); Lorenz (Fn. 20), Rn. 50; Spiro (Fn. 21), S. 53 f. und S. 245.

92 Sehr klar Schmidt (Fn. 84), S. 504 und S. 511 f.

93 BGH NJW 1986, S. 2757 (2758); NJW 1994, S. 2232 (2233); G. Wagner, in: F. J. Sackert et al. (Hrsg.), Münchener Kommentar BGB, Band 6, 7. Aufl., München 2017, § 823 Rn. 29 ff. und 82.

94 Siehe etwa BGH NJW 2001, S. 1786 (1787); Schmidt (Fn. 84), S. 512; Grundmann (Fn. 88), Rn. 55; Wagner (Fn. 93), Rn. 38.

95 Siehe nur Koziol (Fn. 83), S. 593.

96 Siehe statt vieler Koziol (Fn. 83), S. 593 f.; Jansen (Fn. 82), S. 425 und S. 440-445.

97 Siehe dazu als Kern der Problematik des Verschuldensmaßstabs, Jansen (Fn. 82), S. 445-447.

jeweils, ob die schadensbegründende Pflichtverletzung von einem Menschen (dem Geschäftsherrn) in der konkreten Situation typischerweise schuldhaft verursacht worden wäre.

Im Ergebnis führt dies zu einer Haftung für solche Schäden, die Menschen typischerweise (bei Beachtung der allgemeinen Sorgfalt) nicht verursacht hätten, für die also Menschen, wenn sie denn eintreten, in der konkreten Situation auch haften würden (Fehler Typ 1: Tisch-Fall). Dagegen haftet der Geschäftsherr nicht für KIA-Schäden, die auch bei einem sorgfältigen Menschen eingetreten wären (Fehler Typ 2: Wand-Fall), etwa auch im Falle höherer Gewalt.<sup>98</sup> Zudem entgeht der Geschäftsherr der Haftung für solche Schäden, die Menschen infolge von Sorgfaltsverletzungen verursachen, die dem KIA aber abtrainiert wurden, so dass schon kein Schaden entsteht (Fehler Typ 3). Dies erscheint vor dem Hintergrund der Beschränkungsfunktion des Verschuldenserfordernisses (Fehler Typ 2), der Ausgleichsfunktion<sup>99</sup> des Haftungsrechts (Fehler Typ 1) und dessen Anreizfunktion zur Schadensvermeidung (Fehler Typ 3) ausgewogen und sachgerecht. Ausgleich ist insbesondere auch dann geschuldet, wenn der KIA durchschnittlich weniger Fehler macht als ein Mensch, da dies an der konkreten Kompensationsbedürftigkeit des einzelnen Schadens nichts ändert.<sup>100</sup> Verbleibende Schwierigkeiten bei der Bestimmung dieses Verschuldensäquivalenzmaßstabs können, wie auch bisher bei Nachweisschwierigkeiten hinsichtlich menschlichem Verschulden, durch Vermutungs- und Beweislastregeln abgefedert werden; § 280 Abs. 1 S. 2 BGB gilt ja etwa auch im Rahmen des § 278 S. 1 BGB.

## (2) KI-Spezifität

Problematisch an dieser Lösung ist jedoch, dass sie gewisse Anreize für Geschäftsherrn setzt, solche KIA zum Einsatz zu bringen, die (sorgfältiges) menschliches Verhalten möglichst gut imitieren, ohne jedoch notwendig darüber hinaus zu gehen. Die Fähigkeit eines KIA, etwa in einer dem Menschen überlegenen Weise die Folieneigenschaft der Wand zu erkennen, bietet zumindest keinen Haftungsvorteil, da für die unterbliebene Erkenntnis auch nicht gehaftet würde (Fehler Typ 2). Dies ist insofern ungünstig, als ein Potenzial von KI darin besteht, Schäden zu vermeiden, die auch sorgfältige Menschen nicht verhindern können. Diese Erwägung spricht im Grundsatz für die KI-spezifische Lösung, nach der nur für solche Schäden gehaftet wird, die dadurch entstehen, dass die gesamte Klasse eines KIA hinter den (im Moment seines Inverkehrbringens oder bei allfälligen Wartungsarbeiten/

98 So auch Brunner (Fn. 46), S. 104 f.

99 Dazu etwa umfassend Jansen (Fn. 82), S. 89 ff.; siehe auch Wagner (Fn. 83), S. A 14.

100 So auch Schulz (Fn. 25), S. 139 f.; Köbler (Fn. 22).

Updates geltenden) technischen Standards zurückbleibt.<sup>101</sup> Dies würde gerade dem Umstand Rechnung tragen, dass KIA häufig systemübergreifend solche Fehler machen, die ein Mensch vermeiden könnte, andererseits aber solche Fehler vermeidet, die ein Mensch machen würde.<sup>102</sup> Wenn in der Summe die erwarteten Schadens- und Schadensvorsorgekosten bei KI-Einsatz geringer sind, wäre ein solcher Zustand durchaus erstrebenswert.

Das zentrale Problem an der KI-spezifischen Lösung ist jedoch, dass allgemein konsentrierte, technische Standards für die Entwicklung von hoch innovativen Produkten wie KIA gegenwärtig nicht erkennbar und, gerade wegen deren Entwicklungsoffenheit und -dynamik, auch nur schwer implementierbar sind. Immerhin hat die US-amerikanische Aufsichtsbehörde freiwillige, allerdings noch sehr allgemeine Standards für autonome Fahrzeuge erarbeitet,<sup>103</sup> die jedoch insbesondere noch keine Regeln für das, für die Schadensvermeidung essentielle, Training des KI-Modells selbst (z.B. Auswahl des Lernalgorithmus und der verwendeten KI-Modellklasse, Validierungszyklen, Trainingsdaten)<sup>104</sup> bieten.<sup>105</sup> Die Europäische Kommission will ebenfalls zwar zu einer Lösung beitragen, belastbare Ergebnisse sind jedoch auch hier keinesfalls gesichert.<sup>106</sup> Auch die seit 2011/12 in Kraft befindliche Norm ISO 26262 bietet für elektronische Komponenten in Fahrzeugen lediglich eine äußerst vage Risikodefinition<sup>107</sup> und ist zudem bereits veraltet,<sup>108</sup> jedenfalls nicht auf maschinelles Lernen zugeschnitten. Solange aber präzise technische Standards für maschinelles Lernen nicht in zuverlässiger Weise bestimmt werden können,<sup>109</sup> ist schon aus diesem Grund auf die anthropo-parallele Lösung zu rekurren.

Dabei sollte zudem nicht verkannt werden, dass auch diese Lösung Anreize für superhumane Fehlervermeidung (Fehler Typ 2 und 3) setzt, die sich allerdings zum Teil (Fehler Typ 2) nicht aus einer Haftungsvermeidung speisen. Vielmehr resultieren sie bei letzteren Fehlern aus dem Umstand, dass dadurch einerseits prozessuale Unsicherheit und Kosten vermieden werden können, andererseits übermenschliche

101 Für eine systembezogene Fehlerdefinition im Produkthaftungsrecht auch explizit *Geistfeld* (Fn. 11), S. 1651 ff.; *Wagner* (Fn. 11), S. 734 ff.

102 *Geistfeld* (Fn. 11), S. 1651 f.

103 *US. Department of Transportation/National Highway Traffic Safety Administration*, Automated Driving Systems 2.0: A Vision for Safety, Washington D.C. 2017, S. 5-15.

104 Siehe etwa *Goodfellow* et al. (Fn. 56), S. 149 f.

105 Vgl. *ibid.*, S. 9.

106 *European Commission*, FinTech Action plan: For a more competitive and innovative European financial sector, COM(2018) 109 final, S. 7-8.

107 Dazu etwa *Wagner* (Fn. 11), S. 730 f.; *Spindler* (Fn. 11), S. 771; ausführlich *E. Helmig*, ISO 26262 – Funktionale Sicherheit in Personenfahrzeugen, InTeR 2013, S. 28 (29).

108 Vgl. *Gomille* (Fn. 11), S. 79.

109 Diesen konzediert implizit auch *Wagner* (Fn. 11), S. 737-740; konstruktiv, aber zugleich skeptisch *Spindler* (Fn. 11), S. 770 f.

Fähigkeiten einen klaren Wettbewerbsvorteil darstellen, da Kunden in jedem Fall an einer Fehlervermeidung interessiert sein dürften, zumal dann, wenn es um Fehler geht, für die der Vertragspartner nicht haftet.

Hinzu kommt, dass das Kriterium der Verletzung technischer Standards tendenziell zu einer Vermengung von Personal- und Produktrisiko führt, da technische Standards sowohl für die eine als auch die andere Risikokategorie gelten können – häufig wohl sogar primär für das Produktrisiko. Diese Standards müssen im Bereich von KI, wie erwähnt, vor allem die sorgfältige Kalibrierung des algorithmischen Modells erfassen,<sup>110</sup> die bei maschinellem Lernen zentral für die Schadensvermeidung ist.<sup>111</sup> Solche maschinenlernbezogenen Standards allerdings richten sich primär an den Hersteller, nicht den Verwender; insofern sind sie im Produkthaftungsrecht, nicht aber im Vertragsrecht richtig aufgehoben. Demgegenüber erfasst die anthropo-parallele Lösung präziser das KI-Äquivalent des Personalrisikos – sie verbürgt gerade jene schulderspezifische Als-Ob-Betrachtung, die zentraler Bestandteil der Funktionsweise der Zurechnung im Vertragsrecht ist. Auch dies spricht letztlich für diese Variante.

## II. Deliktischer Bereich

Mit denselben Argumenten hinsichtlich der Unterscheidung des Produktrisikos vom Personalrisiko<sup>112</sup> lässt sich bei hinreichend autonomen KIA auch die analoge Anwendung des § 831 BGB hinsichtlich der Tatbestandsvoraussetzungen des Verrichtungsgehilfen begründen.<sup>113</sup> Dies ist in der Literatur nicht unumstritten,<sup>114</sup> es sprechen dafür jedoch gute Gründe.

Verrichtungsgehilfe ist, wer mit Wissen und Wollen des Geschäftsherrn weisungsgebunden in dessen Interessenbereich tätig wird.<sup>115</sup> Eine Weisungsgebundenheit wird sich auch bei hinreichend autonomen KIA regelmäßig annehmen lassen, da diese zwar, ähnlich Menschen, zu selbstständigen und adaptiven Handlungen befähigt sind, jedoch in einem durch den Geschäftsherrn determinierten Kontext, zu im Voraus bestimmten Zeiten und Zwecken an einem konkreten Ort eingesetzt werden. Der Geschäftsherr (die Gesellschaft) kann den KIA jederzeit faktisch von sei-

110 Siehe oben, Fn. 104 f.

111 *Geistfeld* (Fn. 11), S. 1644 ff.; *Wagner* (Fn. 11), S. 734 ff.

112 Bzw., in der bei § 831 BGB mitunter gebräuchlichen Terminologie: Betriebsrisiko: BGH NJW 1957, S. 785 (787).

113 Zu weiteren deliktischen Anspruchsgrundlagen, die hier aus Platzgründen übergangen werden müssen, siehe etwa *Günther* (Fn. 25), S. 124 ff.; *Hanisch* (Fn. 25), S. 41 ff.

114 Gegen einen Analogieschluss *Günther* (Fn. 25), S. 135 (unter Verweis auf mangelnde Handlungsfähigkeit); *Schulz* (Fn. 25), S. 147 f. (wegen mangelnder Verschuldensfähigkeit); für einen Analogieschluss *Hanisch* (Fn. 25), S. 196; *Denga* (Fn. 74), S. 74.

115 Ständige Rechtsprechung, siehe bereits RGZ 92, 345 (346 f.); heute etwa BGH NJW 2009, S. 1740 Rn. 11.

ner Aufgabe abziehen;<sup>116</sup> der KIA ist zudem vom Geschäftsherrn, mehr noch als ein menschlicher Gehilfe, organisatorisch abhängig.<sup>117</sup> Eine Grenze wäre erst dann erreicht, wenn in einer (noch nicht absehbaren) Zukunft KIA als gänzlich selbstständige Unternehmer (*artificial general intelligence*) auftreten würden.<sup>118</sup> Die objektiv rechtswidrige Erfüllung des Deliktstatbestands kann aus denselben Gründen wie die Frage des Status als Zurechnungsausgangsobjekt *für die Zwecke* der analogen Anwendung von § 831 Abs. 1 BGB bei hinreichend autonomen KIA bejaht werden: Auch im Rahmen von § 831 Abs. 1 BGB geht es um durch den Geschäftsherrn besser kontrollierbare Personalrisiken. Eigenes Verschulden des Verrichtungsgehilfen ist, anders als bei § 278 BGB, nach herrschender Meinung grundsätzlich nicht erforderlich,<sup>119</sup> so dass dieses Problem entfällt.

Da es bei § 831 Abs. 1 BGB jedoch zusätzlich auf eigenes Auswahl-, Instruktions- oder Überwachungsverschulden des Geschäftsherrn ankommt, ist das Schwert der deliktischen Gehilfenhaftung ungleich weniger scharf als das der vertraglichen, wie allgemein bekannt ist.<sup>120</sup> Der Geschäftsherr darf zwar nicht blind auf das fehlerfreie Funktionieren der vermeintlich objektiven Technik vertrauen. Insbesondere mit Blick auf KIA wird er sich jedoch regelmäßig nach § 831 Abs. 1 S. 2 BGB exkulpieren können, wenn er (oder der Hersteller) den entsprechenden KIA-Typ regelmäßig unter Realbedingungen getestet hat und auf bekannt gewordene Fehlerquellen angemessen reagiert wurde.<sup>121</sup> Die notwendige Testfrequenz bestimmt sich dabei richtigerweise nach der Wahrscheinlichkeit und Schwere einer möglichen Rechtsgutsverletzung.<sup>122</sup> Dieselben Resultate ergeben sich hinsichtlich Auswahl, Instruktion, Wartung und Fehlerreaktion bei den ohnehin zur vertraglichen Pflicht aus §§ 241 Abs. 2, 276 BGB<sup>123</sup> und zu § 831 Abs. 1 BGB im Wesentlichen parallelen<sup>124</sup> unternehmerischen Organisationspflichten im Rahmen des § 823 Abs. 1

116 Vgl. G. Wagner, in: F. J. Säcker et al. (Hrsg.), Münchener Kommentar BGB, Band 6, 7. Aufl., München 2017, § 831 Rn. 14.

117 Zu diesem Kriterium siehe BGH NJW 2009, S. 1740 Rn. 11; die Abhängigkeit ergibt sich schon daraus, dass KIA heutzutage, und auch in absehbarer Zukunft, spezifische Probleme autonom lösen, jedoch nicht allgemein alle menschlichen Tätigkeiten übernehmen können (*artificial narrow intelligence*, nicht *artificial general intelligence*, siehe *Executive Office of the [US] President*, Preparing for the Future of Artificial Intelligence, Washington D.C. 2016, S. 7; P. Wang/B. Goertzel (Hrsg.), *Theoretical Foundations of Artificial General Intelligence*, Atlantis Press, Amsterdam 2012).

118 Zu *artificial general intelligence*, siehe die Nachweise in Fn. 117.

119 Siehe etwa BGH NJW 1971, 31 (32); Wagner (Fn. 116), Rn. 29; zum Sonderproblem bei § 826 BGB unten, Teil D.II.4.a)ii.

120 Siehe nur Wagner (Fn. 116), Rn. 1 f.

121 Vgl. zum Pflichtenprogramm des Geschäftsherrn Wagner (Fn. 116), Rn. 35-38; BGH NJW-RR 1996, S. 867 (868); spezifisch zu KIA Dengä (Fn. 74), S. 75 f.; Horner/Kaulartz (Fn. 55), S. 8 f.

122 G. Spindler, Haftung der Geschäftsführung für IT-Sachverhalte, CR 2017, S. 715 (723).

123 Dazu oben Fn. 20 und Text.

124 Wagner (Fn. 116), Rn. 2 und 11. Als Unterschied bleibt freilich die jedenfalls im Grundsatz fehlende Verschuldensvermutung bei § 823 Abs. 1 BGB.



BGB.<sup>125</sup> Wenn man jedoch der hier vorgeschlagenen Auslegung des § 278 S. 1 BGB folgt, ist eine deliktische Haftung von ungleich geringerer Relevanz.

### III. Zwischenergebnis

Die vorangehende Untersuchung hat, anders als die bisherige Literatur, klar zwischen zwei Fragen unterschieden: einerseits derjenigen, ob KIA taugliche Zurechnungsausgangssubjekte und damit Erfüllungsgehilfen sein können; und andererseits derjenigen, ob der Verschuldensbegriff sinnvoll auf sie angewandt werden kann. Beide Male geht es insofern vor allem um eine analoge Anwendung von § 278 S. 1 BGB; die Antworten sind jedoch auch von Relevanz für eine analoge Anwendung von § 831 Abs. 1 BGB.

Im Rahmen der ersten Frage ist richtigerweise nach dem Grad der Autonomie des KIA zu differenzieren. Vollständig autonome und signifikant mittelstark autonome KIA, die adaptiv und im Wesentlichen ohne die Notwendigkeit menschlichen Eingreifens agieren, können analog § 278 S. 1 BGB als Erfüllungsgehilfen qualifiziert werden. Hier überwiegt das Personalrisiko das Produktrisiko. Gleiches gilt für die Eigenschaft als Verrichtungsgehilfe analog § 831 Abs. 1 BGB.

Für die zweite Frage ist entscheidend, worin man die zentrale Funktion des zivilrechtlichen Verschuldenserfordernisses sieht. Erkennt man dieses als Vehikel der Haftungsbegrenzung, so lassen sich funktionale Verschuldensäquivalente für KIA definieren. Vorzugswürdig ist letztlich eine anthropo-parallele Lösung, bei der für solche durch KIA verursachte Schäden gehaftet wird, die dem Geschäftsherrn bei situationsgleicher Verursachung durch einen menschlichen Erfüllungsgehilfen als schuldhafter Pflichtverletzung zugerechnet worden wären. Dies gründet auf einer Verschränkung einer Als-Ob-Betrachtung mit einem objektivierten Verschuldensmaßstab, die auch bei menschlichen Erfüllungsgehilfen letztlich unerlässlich ist.

### D. Wissenszurechnung

Das dritte zentrale Problem im Bereich der Zurechnung beim Einsatz von KIA betrifft deren Wissen, also einerseits den explizit festgehaltenen Output und andererseits bestimmte interne Zustände. Dieses Problem ist, soweit ersichtlich, in der Literatur bislang noch nicht behandelt worden. Relevant wird es vor allem bei selbstlernenden Vorhersagesystemen (*prediction systems*), die nicht notwendig in einen physischen Roboter eingebaut sind, aber aus einem vorhandenen Input einen Output (die Vorhersage) errechnen. Auch hier muss zunächst geklärt werden, wie *de lege lata* die Zurechnung von KI-generiertem Wissen gelöst wird. Im Bereich der Wissenszurechnung hat sich aufgrund der äußerst punktuellen positivrechtlichen

125 Zu dieser Pflicht der Gefahrensicherung im Organisationsbereich etwa BGH NJW-RR 1996, S. 867 (868); Wagner (Fn. 93), Rn. 97 ff.; konkret zu § 823 Abs. 1 BGB und KI-Risiken Spindler (Fn. 11), S. 767 ff.; Dengä (Fn. 74), S. 71 ff., besonders S. 74; Schulz (Fn. 25), S. 142 ff.

Regelung (§ 166 BGB) eine besonders eigenständige Rechtsprechung entwickelt, weitgehend losgelöst von positivrechtlichen Grundlagen.

## I. Überblick über die Rechtsprechung

Zu unterscheiden sind Grundsätze der Wissenszurechnung und Wissenszusammenrechnung für den rechtsgeschäftlichen Verkehr, insbesondere für den Umgang mit juristischen Personen und (teil-)rechtsfähigen Gesamthandsgesellschaften,<sup>126</sup> von jenen des Deliktsrechts. Im rechtsgeschäftlichen Bereich werden schon nach § 31 BGB (analog) Handlungen *und* Wissen von verschiedenen vertretungsberechtigten Organwaltern der Gesellschaft „kumulativ, gleichsam mosaikartig“<sup>127</sup> zugerechnet, um den Vertragspartner so zu stellen, als habe er nicht mit einer in verschiedene Wissens- und Verantwortungsbereiche aufgespaltenen Entität, sondern mit einer natürlichen Person kontrahiert.

Darüber hinaus lassen sich zwei Tendenzen erkennen. Erstens muss sich ein Unternehmen, unterhalb der Schwelle des verfassungsmäßig berufenen Vertreters, analog § 166 Abs. 1 BGB, das Wissen eines „Wissensvertreters“ zurechnen lassen.<sup>128</sup> Dies ist, ähnlich wie bei der Repräsentantenhaftung des § 31 BGB, jeder, „der nach der Arbeitsorganisation des Geschäftsherrn dazu berufen ist, im Rechtsverkehr als dessen Repräsentant bestimmte Aufgaben in eigener Verantwortung zu erledigen und die dabei angefallenen Informationen zur Kenntnis zu nehmen sowie gegebenenfalls weiterzuleiten.“<sup>129</sup>

Zweitens hat jedoch die Figur des Wissensvertreters in der jüngeren Rechtsprechung des V. Zivilsenats des BGH zunehmend einer allgemeinen Pflicht (präziser eigentlich: Obliegenheit) des Unternehmens zur Organisation eines effektiven Wissenstransfers zwischen Mitarbeitern und Abteilungen Platz gemacht.<sup>130</sup> Dabei ist die konkrete Stellung des Mitarbeiters nicht mehr bedeutsam.<sup>131</sup> Die Grenze der Zurechenbarkeit ist jedoch spätestens dort erreicht, wo mit einer Speicherung und

126 Für die Übertragbarkeit der Zurechnungslehren auf letztere ausdrücklich BGH NJW 1996, S. 1339 (1341); NJW 2001, S. 359 (360); stärker differenzierend noch BGH NJW 1995, S. 2159 (2160) (zur KG).

127 BGH NJW 1990, S. 975 (976) (unter Rekurs lediglich auf §§ 31, 89 BGB, nicht auf § 166 BGB); BGH NJW-RR 2006, S. 771 (772).

128 Die Figur der Wissensvertretung kommt ursprünglich aus dem Versicherungsvertragsrecht, siehe bereits RGZ 101, 402 (403); grundlegend *Richardi*, Die Wissensvertretung, AcP 169 (1969), S. 385; diese Figur betonend *Spindler* (Fn. 10), S. 319.

129 BGH NJW 1992, S. 1099 (1100); siehe bereits BGH NJW 1971, S. 1702 (1703); NJW 1982, S. 1585 (1586) („allgemeiner Rechtsgedanke“); so bereits *Richardi* (Fn. 128), S. 397.

130 So wird der Wissensvertreter in einem der letzten Urteile des BGH zur Materie gar nicht mehr erwähnt: BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 16; siehe auch die Zusammenfassung bei BGH NJW 2001, S. 2535 (2536); zur Differenz auch *Grigoleit* (Fn. 10), S. 162 f., S. 193; siehe aber, beide Stränge vermischend, die Entscheidung des I. Zivilsenats, BGH NJW 2016, S. 3445 (3450).

131 BGH NJW-RR 2006, S. 771 (772).

Weitergabe des Wissens an die letztlich entscheidungszuständigen Stellen, bzw. mit der Abfrage des Wissens durch diese, aus Verkehrsschutzgesichtspunkten nicht mehr gerechnet werden kann.<sup>132</sup> Dies kann verschiedene Gründe haben. Erstens ist ohnehin nur das „typischerweise aktenmäßig festgehaltene Wissen“ tauglicher Zurechnungsgegenstand.<sup>133</sup> Die Speicherungsnotwendigkeit beurteilt sich nach der ex ante Erheblichkeit der Tatsache.<sup>134</sup> Zweitens muss der Informationsaustausch zwischen dem Wissenden und der konkret entscheidungsberufenen Stelle „möglich und naheliegend“<sup>135</sup> sein, also ein hinreichender Anlass für die Weitergabe des Wissens vorliegen.<sup>136</sup> Ist drittens der Anlass nicht bedeutend genug, oder sind die Suchkosten zu hoch, muss einmal gespeichertes Wissen nicht mehr im konkreten Entscheidungsfall abgefragt werden.<sup>137</sup> Entscheidend ist für alle drei Dimensionen (Speicherung; Weitergabe; Abfrage) letztlich eine Zumutbarkeit im konkreten Fall.<sup>138</sup> Damit will der BGH unter dem Verweis auf die Gleichstellung mit natürlichen Personen eine überproportionale Belastung von Unternehmen verhindern.<sup>139</sup> Diese Maßstäbe der Wissenszurechnung wirken unmittelbar für die Frage des Verschuldens (z.B. im Rahmen der Arglist) fort: Die Kenntnis wird unterstellt und deren Nichtaufdeckung zum Anknüpfungspunkt des Verschuldens, auch wenn dem konkreten Entscheidungsträger eine Aufdeckung faktisch wegen Nichtkenntnis unmöglich war.<sup>140</sup>

Die Frage der unternehmensinternen Zurechnung wird damit zu einer der adäquaten Organisation des Wissenstransfers (Wissensverantwortung),<sup>141</sup> die in Verkehrsschutzerwägungen wurzelt.<sup>142</sup> Dabei geht es dem BGH selbst im Bereich der Arglist ganz explizit nicht um ein moralisches Werturteil, sondern um eine angemessene Risikoverteilung zwischen den Parteien.<sup>143</sup> Allerdings wird an der Lehre von der

132 Vgl. BGH NJW 1992, S. 1099 (1100); BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 16, 21 f.

133 BGH NJW 1990, S. 975 (976) im Anschluss an M. Bohrer, Anmerkung, DNotZ 1991, S. 124 (129); zu Folgen fehlerhaften Protokollierens des Wissens (keine Zurechnung des tatsächlichen Wissens), siehe BGH NJW 1996, S. 1205 (1206).

134 BGH NJW 1996, S. 1339 (1341).

135 BGH NJW 1989, S. 2879 (2881); NJW 1992, S. 1099 (1100).

136 BGH NJW 1996, S. 1339 (1341); NJW 1997, S. 1917 (1918).

137 BGH NJW 1996, S. 1339 (1341); BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 16.

138 Ibid.

139 BGH NJW 1996, S. 1339 (1341).

140 BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 24.

141 Vgl. Bohrer (Fn. 133), S. 125; Taupitz (Fn. 5), S. 25 ff.; Medicus (Fn. 5), S. 10 f.; aufgenommen von BGH NJW 1996, S. 1339 (1341).

142 BGH NJW 1996, S. 1339 (1341).

143 BGH NJW 1990, S. 975 (976).

Wissensverantwortung bemängelt, sie verwische die Grenzen zwischen positivem Wissen und Wissenmüssen.<sup>144</sup>

Im Deliktsrecht wiederum stellt sich die Lage diametral anders dar.<sup>145</sup> Das KG hat kürzlich im *Fondsprospektfall* einen Vorstoß unternommen, die genannten Grundsätze der Wissenszurechnung und -zusammenrechnung auch dort fruchtbar zu machen.<sup>146</sup> Der für das Deliktsrecht zuständige VI. Zivilsenat hat sich in der Revision zu einem derartigen Bekenntnis nicht durchringen können<sup>147</sup> und ihnen insbesondere im Rahmen des § 826 BGB der Sache nach eine deutliche Absage erteilt.<sup>148</sup> Dies liegt nach dem Dafürhalten des Senats darin begründet, dass § 826 BGB auf dem Vorwurf eines moralischen Unwerturteils aufruhe, der nur den handelnden und nach § 31 BGB maßgeblichen Personen gemacht werden könne und nicht zurechenbar sei; der nach § 31 BGB konkret relevante Vertreter muss daher selbst die Kenntnisse erworben haben, aus denen sich der moralische Vorwurf und damit die Sittenwidrigkeit speist.<sup>149</sup> Gleiches gilt für das Wollenselement im Rahmen des Schädigungsvorsatzes.<sup>150</sup> § 826 BGB hat nach dem BGH „personalen Charakter“,<sup>151</sup> wesentliche Tatbestandselemente sind damit einer Zurechenbarkeit entzogen. In der Literatur hat die Entscheidung ein geteiltes Echo gefunden.<sup>152</sup> Sie ist gleichwohl geltendes (Richter-)Recht und insofern auch Maßstab für die Zurechnung im Bereich künstlicher Intelligenz.

## II. Zurechnung KI-generierten Wissens

Diese Differenzierungen bereiten bei der Anwendung auf KI-generiertes Wissen erhebliche Probleme. Ein Fall mag dies genauer veranschaulichen: Stellen wir uns vor, ein Unternehmen bedient sich zur Verwaltung seines erheblichen Grundstücksbestands eines KIA, der entscheiden soll, welche Grundstücke von dem Unterneh-

144 Siehe nur *Grigoleit* (Fn. 10), S. 195; *W. Flume*, Die Haftung für Fehler kraft Wissenszurechnung bei Kauf- und Werkvertrag, AcP 197 (1997), S. 441 (443, 445); *B. Dauner-Lieb*, Wissenszurechnung im Gewährleistungsrecht – Ethische Neutralisierung der Arglist? –, in: Hönn/Konzen/Ereutz (Hrsg.), FS Kraft, München 1998, S. 54–56; *T. Gasteyer/C.-U. Goldschmidt*, Wissenszurechnung bei juristischen Personen und im Konzern, AG 2016, S. 116 (118); dagegen etwa *Schubert* (Fn. 46), Rn. 48.

145 Zur deliktischen Verjährung BGH NJW 2001, S. 2535 (2536 f.); siehe ferner Urteil vom 28.2.2012 – VI ZR 9/11, BeckRS 2012, 08778, Rn. 14.

146 KG Urt. v. 27.8.2015 – 2 U 42/09, BeckRS 2016, 17728, Rn. 52–54.

147 BGH NJW 2017, S. 250 Rn. 23.

148 Ibid. Rn. 23–26.

149 Ibid.

150 Ibid. Rn. 26.

151 Ibid. Rn. 23.

152 Zustimmend etwa *B. Handke*, BB-Kommentar. „Keine Wissenszurechnung und -zusammenrechnung innerhalb einer juristischen Person bei Haftung nach § 826 BGB“, BB 2016, S. 2636; *S. Frisch*, Anmerkung zum Urteil des BGH vom 28.6.2016, Az. VI ZR 536/15 – Zur Haftung einer juristischen Person aus § 826 BGB, EWiR 2016, S. 761 (762); *M. Schwab*, BGB AT und Schuldrecht BT: Keine Wissenszusammenrechnung im Deliktsrecht, JuS 2017, S. 354 (356); wohl auch *Reuter* (Fn. 8), S. 310; kritisch *G. Wagner*, Anmerkung, JZ 2017, S. 522.

men in eigener Regie baulich entwickelt und welche verkauft werden sollen. Trainiert wird der Beispiels-KIA auf historischen Daten hinsichtlich bisheriger Nutzungs- und Verkaufsentscheidungen von Grundstücken. Ein Faktor, der erheblich den Wert der Grundstücke und die Entscheidung über deren Handhabung beeinflusst, sind etwaig vorhandene Altlasten. Deren Vorliegen ist jedoch kein direkter Inputfaktor, da ihr Vorhandensein zum maßgeblichen Entscheidungszeitpunkt nicht bekannt ist; häufig ist jedoch das Vorhandensein von Altlasten redundant in anderen Faktoren encodiert ist. Der KIA „erkennt“ diese Korrelationen von Altlastenbefall mit gewissen Risikofaktoren und empfiehlt nun gerade die Grundstücke zum Verkauf, die tatsächlich mit Altlasten belegt sind, ohne jedoch bei seiner Empfehlung diese Korrelationen ausdrücklich offenzulegen. Vielmehr sind sie lediglich als interne Zustände, als implizites Wissen, im algorithmischen Modell repräsentiert. Erfolgt ein Verkauf unter Ausschluss der Gewährleistung, so stellt sich die Frage, ob der Käufer nach Entdecken der Altlasten geltend machen kann, der Verkäufer habe deren Vorliegen arglistig verschwiegen, § 444 BGB, oder ihn gar vorsätzlich sittenwidrig geschädigt, § 826 BGB. Dies wiederum bemisst sich danach, inwiefern das Wissen des KIA dem Unternehmen zugerechnet werden kann.

Man wird hier schon im Ausgangspunkt zwischen explizitem und implizitem Wissen des KIA unterscheiden müssen. Erwägungen dazu fehlen in der rechtswissenschaftlichen Literatur bislang. Explizites Wissen umfasst sicherlich den ausdrücklichen Output (die Vorhersage) des algorithmischen Modells, in unserem Fall also die Klassifikation des Grundstücks als Eigennutzung oder Verkauf. Zudem sollten unter explizites Wissen solche Faktoren für die Gewinnung des Output fallen, die unmittelbar aus dem Modell ersichtlich sind. Inwiefern solche Faktoren existieren, ist von Modellklasse zu Modellklasse unterschiedlich. Bei einigen Verfahren (etwa Entscheidungsbäumen<sup>153</sup> oder linearer Regression)<sup>154</sup> können die wichtigsten Faktoren für den Output mitunter unmittelbar, d.h. mit sehr geringem Aufwand, aus dem Modell abgelesen werden.

Dies ist bei anderen Modellen (etwa tiefen neuronalen Netzen) nicht der Fall. Hier beginnt vielmehr das implizite Wissen.<sup>155</sup> Informationen über die Gründe für den

153 A. Groll et al., Prediction of the FIFA World Cup 2018 – A random forest approach with an emphasis on estimated team ability parameters, Working Paper (2018), <https://arxiv.org/abs/1806.03208>, S. 9 f.

154 Goodfellow et al. (Fn. 56), S. 104.

155 Siehe etwa M. Jordan/T. Mitchell, Machine learning: Trends, perspectives and prospects, 349 Science 2015, S. 255 (255 f. und 258). Letztlich ist der Übergang zwischen expliziten und impliziten Wissensformen zwar fließend: Wenn die Unterscheidung auch bei einigen KI-Modellen klar durchführbar ist, so existieren doch aufgrund der Vielzahl von KI-Modellen zahllose Graustufen (vgl. etwa den Unterschied zwischen *decision trees* und *random forests* hinsichtlich der Interpretierbarkeit: Groll (Fn. 153), S. 8-10; grundsätzlich Z. Lipton, The Mythos of Model Interpretability, ICML Workshop on Human Interpretability in Machine Learning, 2016, <https://arxiv.org/abs/1606.03490>). Nichtsdestoweniger soll daran zu Analysezwecken festgehalten werden; die genaue Zuordnung ist dann eine Frage des Einzelfalls.

expliziten Output sind in einem tiefen neuronalen Netz in einer impliziten Funktion, und den hochdimensionalen internen Verbindungen zwischen Neuronen in den tiefen Netzwerklagen, repräsentiert.<sup>156</sup> Sie sind also nicht einfach von außen zugänglich. Entscheidend für implizites Wissen dürfte sein, dass die relevante Tatsache (z.B. Altlastenbefund) zwar nicht direkt aus dem Modell ablesbar ist (sonst explizites Wissen), aber intern so encodiert ist, dass sie durch spezifische, aufwändigere Verfahren interpretierbaren maschinellen Lernens im Nachhinein eruiert werden könnte. In solchen Verfahren bestimmt ein zweiter Algorithmus, der das ursprünglich verwendete algorithmische Modell untersucht, welche Faktoren für die ursprüngliche algorithmische Prognose relevant waren.<sup>157</sup> Wenn sich aus diesen Faktoren plausiblerweise die relevante Tatsache ergibt, kann man von implizitem Wissen sprechen. Wenn selbst dieser Schluss nicht möglich ist, wird man statt von implizitem Wissen, jedenfalls in rechtlicher Hinsicht, von Nichtwissen ausgehen müssen.

## 1. Zurechnungsausgangssubjekt

Im (vor-)vertraglichen Bereich stellt sich insbesondere die Frage, inwiefern der Altlastenbefall arglistig verschwiegen wurde. Dies setzt das Bestehen einer Aufklärungspflicht und zumindest bedingten Vorsatz hinsichtlich der Kausalität von deren Verletzung für den Vertragsschluss voraus.<sup>158</sup> Nach der genannten Rechtsprechung zur Wissensverantwortung ist eine Zurechnung von Wissen im (vor-)vertraglichen Bereich immer dann möglich, wenn es sich um Tatsachen handelt, die typischerweise aktenmäßig festgehalten, weitergegeben und bei der konkreten Entscheidung abgerufen werden.

Dass die relevante Tatsachenkenntnis nicht durch einen menschlichen Mitarbeiter, sondern einen KIA erlangt wurde, spielt lediglich dann eine Rolle, wenn man an der Figur des Wissensvertreters analog § 166 BGB festhält, die für menschliche Wissenspersonen entwickelt wurde. Selbst in diesem Rahmen jedoch dürfte der maschinelle Ursprung des Wissens dann keinen Unterschied machen, wenn explizites Wissen lediglich eine Empfehlung darstellt, die Grundlage einer menschlichen Entscheidung ist, diese jedoch nicht vollständig ersetzt. Konkret kann dann jedenfalls für explizites Wissen auf denjenigen Mitarbeiter abgestellt werden, dem dieses Wissen unmittelbar zur Verfügung steht und von dem es genutzt wird.

156 *Goodfellow et al.* (Fn. 56), S. 164 f. und S. 405.

157 Siehe etwa *B. Kim et al.*, Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV), Working Paper, 2018, <https://arxiv.org/abs/1711.11279>; *M. Ribeiro et al.*, "Why Should I Trust You?" Explaining the Predictions of Any Classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016, S. 1135; *Lipton* (Fn. 155).

158 Siehe nur BGH NJW 1990, S. 975 (976).

Anders liegen die Dinge im Fall von implizitem Wissen. Dieses Wissen steht dem Mitarbeiter gerade nicht unmittelbar zur Verfügung, so dass auf dem Boden der Lehre vom Wissensvertreter nur der KIA selbst als Zurechnungsausgangssubjekt infrage kommt. Hier lassen sich dieselben Differenzierungen wie im Rahmen der Verhaltenszurechnung fruchtbar machen:<sup>159</sup> Bei hinreichender Autonomie des KIA wird man davon ausgehen können, dass dieser funktionales Äquivalent eines entsprechenden menschlichen Mitarbeiters ist. Ob er zudem Aufgaben „in eigener Verantwortung“ wahrnimmt, ist dann auf Basis des konkreten Einsatzes anthropoparallel zu beurteilen.

Noch klarer wird man auf Grundlage der Lehre von der Wissensverantwortung den maschinellen Ursprung des Wissens für unproblematisch halten können, da im Rahmen dieser Lehre der einzelne Mitarbeiter gerade zu Gunsten eines Fokus auf den Sachgehalt des Wissens und die ordnungsgemäße Organisation seines Transfers im Unternehmen zurücktritt. Dahinter verbirgt sich auch die eigentliche Sachfrage: Welche Typen von KIA-generiertem Wissen können dem Unternehmen, allgemeiner: dem Verwender des KIA, zugerechnet werden?

## **2. Zurechnung expliziten Wissens im rechtsgeschäftlichen Bereich**

An dieser Stelle wird nun die Unterscheidung zwischen explizitem und implizitem Wissen maßgeblich. Für beide Wissenstypen muss daher separat geklärt werden, wie sie auf Grundlage der Rechtsprechung des BGH im rechtsgeschäftlichen Bereich zu behandeln sind, damit darauf aufbauend in kritischer Auseinandersetzung mit Rechtsprechung und Literatur jeweils allgemeine Grundsätze für die Zurechnung spezifischer Formen von KIA-Wissen entwickelt werden können.

### **a) Explizites KIA-Wissen auf Grundlage der BGH-Rechtsprechung**

Explizites KIA-Wissen kann unproblematisch aktenmäßig festgehalten werden. Es entspricht im Wesentlichen menschlichen Bewusstseinsinhalten, die Mitarbeitern gewissermaßen präsent vor Augen stehen und daher notationsfähig sind. Ob das konkrete explizite KIA-Wissen tatsächlich gespeichert, weitergegeben und abgerufen werden muss, bemisst sich, wie dargestellt, nach der erwarteten Erheblichkeit der Tatsachen. Da der Einsatz von KIA mit nicht zu vernachlässigenden (Anschaffungs-, Personal- etc.) Kosten einhergeht und die Kalibrierung eines algorithmischen Modells aufwändig ist, darf man annehmen, dass zumindest der Output regelmäßig diese Erheblichkeitsschwelle überschreiten wird.

Wenn also das Modell, in Abwandlung des Ausgangsfalles, keine Klassifikation der Grundstücke nach Eigennutzung und Verkauf vornehmen, sondern unmittelbar die Wahrscheinlichkeit von Altlastenbefall bestimmen soll, so wäre das Ergebnis (der konkrete Altlastenscore eines Grundstücks) sicherlich zu speichern, weiterzugeben

159 Siehe oben, Teil C.I.1.b.aa.

und im Verkaufsfall abzufragen. Die elektronische Datenverarbeitung reduziert hier die Organisations- und Suchkosten ganz erheblich, so dass eine Zurechnung in den allermeisten Fällen zumutbar und zu bejahen sein sollte.<sup>160</sup> Anders kann es lediglich bei solchen Faktoren liegen, die zwar unmittelbar aus dem Modell ablesbar sind, deren Relevanz sich jedoch ex ante nicht unmittelbar erschließt. Beispielsweise ist denkbar, dass der wichtigste Faktor für eine Verkaufsentscheidung die Anzahl der Bäume auf einem Grundstück ist. Bei solchen Faktoren, bei denen lediglich eine enge Korrelation mit dem jeweiligen Output (hier: Verkaufswürdigkeit) besteht, nicht aber eine Kausalität für diese Klassifizierung naheliegt, und die auch nicht eindeutig auf einen Altlastenbefall (oder eine sonst wichtige Tatsache) hindeuten, dürfte bereits die Notwendigkeit der Speicherung, jedenfalls aber die Notwendigkeit der Weiterleitung und Abfrage zu verneinen sein.

#### b) Grundsätze für die Zurechnung expliziten KIA-Wissens

Explizites KIA-Wissen kann demnach sowohl auf Grundlage der Figur des Wissensvertreters analog § 166 BGB als auch auf Grundlage der Lehre von der Wissensverantwortung zugerechnet werden. Maßgeblicher Unterschied ist, dass erstere Variante an die Stellung des KIA (oder desjenigen, der sein explizites Wissen zur Kenntnis nimmt) anknüpft,<sup>161</sup> zweite an die Erheblichkeit der zuzurechnenden Tatsache. Da es letztlich unter Risiko- wie auch Anreizgesichtspunkten nicht auf die organisatorische Stellung (Eigenverantwortlichkeit),<sup>162</sup> sondern auf die Bedeutung der jeweiligen Tatsache ankommt, sollte diese den Ausschlag geben. Dies begründet eine an dem Sachgewicht des jeweiligen Wissens, nicht an der formalen Stellung des KIA orientierte Zurechnung.

Ähnlich wie bei der Verhaltenszurechnung geht es auch bei der Wissenszurechnung letztlich um eine sachgerechte Risikoverteilung<sup>163</sup> und eine dadurch bewirkte Verhaltenssteuerung.<sup>164</sup> Dies ist richtigerweise auch der Kern des vom BGH in ständiger Rechtsprechung vorgebrachten,<sup>165</sup> von Teilen der Literatur aber zurecht als unscharf kritisierten Gleichstellungsarguments.<sup>166</sup> Allerdings stehen hier drei andere

160 So auch *Spindler* (Fn. 122); *ders.* (Fn. 10), S. 321.

161 Vgl. *Richardi* (Fn. 128), S. 387 f.; detailliert *Grigoleit* (Fn. 10), S. 183 f.

162 Gegen die Relevanz der Eigenverantwortung auch R. *Waltermann*, Arglistiges Verschweigen eines Fehlers bei der Einschaltung von Hilfskräften, NJW 1993, S. 889 (891 f.); *Buck* (Fn. 1), S. 163 f.; M. *Schultz*, Zur Vertretung im Wissen, NJW 1990, S. 477 (479); *Grigoleit* (Fn. 10), S. 184; H.-C. *Grigoleit*, LM § 166 Nr. 43, Bl. 3; *Taupitz* (Fn. 5), S. 25; *Medicus* (Fn. 5), S. 10 und S. 14.

163 So auch BGH NJW 1990, S. 975 (976); *Buck* (Fn. 1), S. 314-316 und S. 323; *Waltermann* (Fn. 162), S. 893; *Baum* (Fn. 46), S. 225 ff, S. 360 f. und S. 366 ff.; *Taupitz* (Fn. 5), S. 27 ff.; *Schubert* (Fn. 46).

164 Vgl. *Buck* (Fn. 1), S. 329.

165 BGH, NJW 1992, S. 1099 (1100); NJW 1996, S. 1339 (1340 f.); Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 16; zustimmend etwa *Waltermann* (Fn. 162), S. 894; *Medicus* (Fn. 5), S. 15 f.

166 Dazu noch eingehender unten, Fn. 204 und Text.



Risiken als bei der Verhaltenszurechnung im Fokus, die jedoch ebenso durch Arbeitsteilung geschaffen werden: das strategische Nachweisrisiko, das strategische Delegationsrisiko und das zufällige Aufspaltungsrisiko.

Erstens ist es denkbar, dass eine Tatsache innerhalb des Unternehmens handlungsleitend wird, sich dies später aber nicht mehr nachweisen lässt. Daraus erwächst das Risiko einer strategisch relevanten, mit einer Aufklärungspflicht belegten Informationsasymmetrie, deren Verletzung jedoch sanktionslos bleibt, wenn positives Wissen des relevanten Entscheidungsträgers nicht nachgewiesen werden kann. Durchaus möglich ist ja etwa, dass einem nach § 31 BGB relevanten Entscheidungsträger Wissen mündlich kommuniziert wurde, dieser das aber ex post erfolgreich abstreitet; der beweisbelastete Vertragspartner wird dann regelmäßig, trotz möglicher Beweiserleichterungen, das Nachsehen haben,<sup>167</sup> wenn es auf den Nachweis positiven Wissens des Entscheidungsträgers, nicht hypothetischen Wissen infolge Wissensorganisation ankommt. Die Lehre von der Wissensverantwortung hat damit eine nicht zu unterschätzende Rechtsdurchsetzungsfunktion.<sup>168</sup> Zweitens besteht das Risiko, dass bei einer an der formalen Stellung orientierten Anknüpfung inkriminierendes Wissen strategisch an die Unternehmensperipherie delegiert und damit nicht sozial nützlich eingesetzt wird.<sup>169</sup> Schließlich mag die Nutzung von für den Vertragspartner vorteilhaftem Wissen ungeplanter Weise aus Gründen der Wissensaufspaltung unterbleiben, so dass die arbeitsteilig verfahrenende Organisation daraus zwar (anders als im ersten Fall) keinen unmittelbaren strategischen Vorteil erlangt, ihr Gegenüber aber mangels Aufklärung dennoch einen Nachteil erleidet.<sup>170</sup>

All diese Risiken bestehen auch und gerade bei der Vermittlung expliziten Wissens durch KIA, das bei Vorhersagesystemen regelmäßig noch von einem menschlichen Letztentscheider zur Kenntnis genommen werden, mithin weitergeleitet oder abgefragt werden muss. Zwar wird man diese Risiken jeweils noch im Einzelfall bewerten und mit Blick auf die konkrete Wissensnorm entsprechend mit zumutbaren Organisationsanforderungen abstimmen müssen;<sup>171</sup> die Gefahr strategischen Handelns wie auch die Anreize zur Vermeidung zufällig unterlassener Wissensnutzung sprechen jedoch grundsätzlich für die flexiblere, von formalen Gesichtspunkten abstrahierende Lehre von der Wissensverantwortung, die hinsichtlich des Zurech-

167 Vgl. etwa BGH NJW 2003, S. 2380; NJW 2001, S. 359 (360).

168 Vgl. D. DeMott, When is a Principal Charged with an Agent's Knowledge? 13 Duke Journal of Comparative and International Law 2003, S. 291 (300).

169 Siehe Wagner (Fn. 152), S. 525; DeMott (Fn. 168), S. 299 f.; C.-U. Goldschmidt, Die Wissenszurechnung, Berlin 2001, S. 136.

170 Baum (Fn. 46), S. 226.

171 Vgl. Goldschmidt (Fn. 169), S. 35.

nungsgrunds, wie schon die Verhaltenszurechnung,<sup>172</sup> an die Schaffung beherrschbarer Risiken anknüpfen kann.<sup>173</sup>

Dabei dürfte dem zurechnungsbeschränkenden Aspekt der Speicher- und Suchkosten für die ordnungsgemäße Wissensorganisation eine zunehmend geringere Bedeutung zukommen in dem Maße, in dem explizit repräsentierte Informationen digital aufbereitet, erfasst und abgerufen werden (können).<sup>174</sup> Umgekehrt wird jedoch selten eine Person den gesamten Geschehensablauf (Wissen und Handlung) in der Hand halten, wie dies etwa von Grigoleit in seiner Kritik an der Lehre von der Wissensverantwortung für die Wissenszurechnung gefordert wird.<sup>175</sup> Dieses Modell personaler Handlungsverantwortung entgegen besseren Wissens mag zwar der ursprünglichen Konzeption der Zurechnungsnormen (im Bereich positiven Wissens) besser entsprechen; es kommt jedoch in der in spezifische Aufgaben fragmentierten KI-Welt schnell an seine Grenzen. Zumindest für den KI-Bereich scheint das Kriterium der unipersonalen Wissens- und Handlungsverantwortung nur schwer operationalisierbar: Vielmehr wird zunehmend ein Ineinandergreifen verschiedener KIA zu gewärtigen sein, die spezifische Komponenten eines komplexen Vorgangs bearbeiten. Dies entspricht dem Umstand, dass erhebliche Fortschritte bei spezialisierter KI erzielt wurden, weniger jedoch bei genereller KI.<sup>176</sup> Zudem werden die KIA, wenn sie wie häufig als Beratungssysteme (*human-machine teaming*), nicht als Letztentscheidungsinstanz konzipiert sind,<sup>177</sup> in vielfältiger Weise mit menschlichen Akteuren innerhalb des Unternehmens interagieren, welche die Prognosen auswerten und umsetzen. Auch dies erfolgt voraussichtlich eher dezentral als in einer Hand, zumal wenn man die Durchbrüche bei dezentralen Konsensalgorithmen im Bereich Blockchain berücksichtigt, deren Verbindung mit KI in absehbarer Zukunft zu erwarten ist.<sup>178</sup> Will man daher Wissenszurechnung im KI-Bereich funktionsfähig halten, bietet sich vor allem die Lehre von der Wissensverantwortung an. Dafür bestehen, wie gesehen, jedenfalls in diesem Bereich auch gute Gründe.

172 Siehe oben, Fn. 48 und Text.

173 So auch Taupitz (Fn. 5), S. 27; Medicus (Fn. 5), S. 11 f.; Baum (Fn. 46), S. 225 ff.; BGH NJW 1996, S. 1339 (1341); J. Drexler, Wissenszurechnung im Konzern, ZHR 161 (1997), S. 491 (511 ff.).

174 Siehe oben, Fn. 160.

175 Grigoleit (Fn. 10), S. 176 f.

176 Siehe oben, Fn. 117.

177 Dies ergibt sich vor allem aus der unterschiedlichen Fehleranfälligkeit von Mensch und Maschine, siehe Executive Office of the [US] President (Fn. 117), S. 10 f.; Jordan/Mitchell (Fn. 155), S. 260.

178 Siehe etwa P. de Filippi/A. Wright, Blockchain and the Law, Harvard University Press, Cambridge, MA 2018, Kapitel 9; Jordan/Mitchell (Fn. 155), S. 259.

### 3. Zurechnung impliziten Wissens im rechtsgeschäftlichen Bereich

Im Fall von explizit repräsentiertem Wissen lassen sich daher die Fragen der Zurechnung auf Grundlage der BGH-Rechtsprechung im Wesentlichen zufriedenstellend lösen; dies ändert sich mit dem Übergang zu implizitem Wissen.

#### a) Implizites KIA-Wissen auf Grundlage der BGH-Rechtsprechung

Implizites Wissen findet beim Menschen seine Entsprechung in unbewussten Kenntnissen: jenen Entscheidungsvariablen, die trotz ihrer Entscheidungsbeeinflussung nicht mit Aufmerksamkeit belegt wurden und daher nicht ins deklarative Gedächtnis Eingang finden, sondern lediglich unbewusst wirksam werden.<sup>179</sup> Trotz ihrer neurobiologisch unbestrittenen Entscheidungsrelevanz handelt es sich dabei gerade nicht um solche Tatsachen, die aktenmäßig festgehalten werden müssen, da dies mangels Eingang in das Bewusstsein des Mitarbeiters schon gar nicht möglich ist.<sup>180</sup> In ähnlicher Weise wird man bei implizitem KIA-Wissen sagen müssen, dass dieses schon der Speicherbarkeit ermangelt. Für die Zwecke der Wissenszurechnung wäre es dann nicht vorhanden.

Dagegen lässt sich allenfalls einwenden, dass implizites Wissen definitionsgemäß mit technischen Verfahren „ins Bewusstsein gehoben“ werden, also explizit gemacht werden kann. Medicus hat überzeugend ausgeführt, dass auch bei Menschen nicht nur präsent (bewusstes) Wissen zurechenbar ist, sondern in normativer Überformung des schwer explorierbaren subjektiv-zerebralen „Präsenzbestands“ auch eine gewisse Erinnerungsanstrengung verlangt werden kann.<sup>181</sup> Dies ist zwar umstritten;<sup>182</sup> in ähnlicher Weise könnte man jedoch argumentieren, auch der KIA-Verwender müsste die technischen Möglichkeiten der Explizierung von implizitem Wissen nutzen. Dies würde jedoch als allgemeine Pflicht zu weit führen. Erstens sind die Methoden der post-hoc Interpretierung intransparenter KI-Modelle noch äußerst experimentell und, anders als die Modelle selbst, kaum in der Breite kommerziell einsetzbar.<sup>183</sup> Hier stets die neuesten Methoden zur Gänze ausschöpfen zu müssen, würde Unternehmen gegenüber Menschen, die „nur“ ihr Gedächtnis befragen müssen, in der Tat überproportional belasten und damit gegen das Gleichstellungsargument des BGH verstoßen. Zweitens würde eine allgemeine Pflicht zur expliziten Wissensgewinnung den Boden der Pflichtentrias des BGH

179 Siehe *Singer* (Fn. 61), S. 17 f.

180 Vgl. zur grundsätzlichen nicht möglichen Zurechnung unbewussten Wissens auch *Grigoleit* (Fn. 10), S. 174; *Buck* (Fn. 1), S. 246.

181 *Medicus* (Fn. 5), S. 5 f.

182 Ausdrücklich für eine Beschränkung auf (wie auch immer zu definierendes) präsent (bewusstes) Wissen *I. Koller*, Wissenszurechnung, Kosten und Risiken, JZ 1998, S. 75 (78 mit Fn. 48); siehe auch Fn. 180.

183 Siehe die Forschungsnachweise in Fn. 157; ferner *A. Selbst/S. Barocas*, The Intuitive Appeal of Explainable Machines, Fordham Law Review (im Erscheinen), <https://ssrn.com/abstract=3126971>, S. 37.

(Speicherung, Weitergabe, Abfrage) verlassen. Auf eine *eingeschränkte* diesbezügliche Pflicht wird jedoch gleich noch zurückzukommen sein (unter b)bb).

#### b) Grundsätze für die Zurechnung impliziten KIA-Wissens

Die bisherige Rechtsprechung und Literatur hält daher nur wenige Ressourcen für die Zurechnung impliziten Wissens bereit. Die Figur des Wissensvertreters greift nicht, da es, in anthropomorpher Diktion, um unbewusstes Wissen geht. Die Lehre von den Organisationspflichten versagt ebenfalls, da implizites Wissen nicht (ohne Weiteres) aktenmäßig festgehalten werden kann. Zwar wahrt dessen grundsätzliche Nicht-Zurechnung die Gleichstellung zwischen Mensch und Maschine insofern, als auch ein einzelner menschlicher Verkäufer einen Altlastenfall, der ihm lediglich unbewusst bekannt ist, nicht arglistig verschwiegen hätte. Dies führt jedoch in der Konsequenz dazu, dass die Wissenszurechnung gerade bei besonders potenten Verfahren maschinellen Lernens, wie tiefen neuronalen Netzwerken, regelmäßig versagt. Man kann darin letztlich doch eine Schlechterstellung des Vertragspartners gegenüber dem Einsatz von Menschen erkennen: Wo ein menschlicher Akteur ausdrücklich nach bestimmten Tatsachen hätte suchen müssen, die bei entsprechender Relevanz auch festzuhalten und weiterzugeben wären, operiert der KIA mit impliziter, nicht zurechenbarer Repräsentation von Sachverhalten. Dies ist nicht nur aus Sicht des Geschädigten misslich, sondern auch, weil es innerbetriebliche Substitutionsentscheidungen (zwischen Mensch und KIA) zugunsten des KIA verzerrt,<sup>184</sup> was die bereits jetzt spürbaren distributiven Folgen des Einsatzes von KI noch verschärfen dürfte.<sup>185</sup> Insgesamt überantwortet diese Lösung zu einseitig die genannten Risiken des Einsatzes von (nicht interpretierbarer) KI dem Vertragspartner des Verwenders.

#### aa) Strikte Wissenszurechnung

Um dem abzuhelfen, könnte man analog zum Vorschlag der strikten Haftung für schädigende Handlungen eines KIA eine strikte Zurechnung impliziten KIA-Wissens implementieren. Dies würde jedoch weit über das Ziel hinausschießen.<sup>186</sup> Eine derartige strikte Zurechnung wäre völlig konturlos und würde den Vertragspartner insofern ungerechtfertigt besserstellen, als auch jene impliziten Inhalte erfasst wären, die ein menschlicher Mitarbeiter (in der gegebenen Zeit) niemals hätte erkennen und festhalten können. Sie käme damit, in menschlicher Übertragung, einer Zurechnung aller unbewussten mentalen Zustände von relevanten Unternehmens-

184 Vgl. Wagner (Fn. 10), S. 258; Posner (Fn. 48).

185 Vgl. L. Nedelkoska/G. Quintini, Automation, skills use and training, OECD Social, Employment and Migration Working Papers, No. 202, OECD Publishing, Paris 2018, <http://dx.doi.org/10.1787/2e2f4eea-en>.

186 Im Ergebnis ebenso gegen eine strikte Wissenszurechnung, außerhalb des KI-Bereichs, Spindler (Fn. 10), S. 322; D. Verse, Doppelmandate und Wissenszurechnung im Konzern, AG 2015, S. 413 (416).

mitarbeitern gleich, was niemand ernsthaft in Erwägung zieht. Zudem wird eine strikte Zurechnung impliziten Wissens auch nicht von der Erwägung der Korrespondenz von Nutzen und Risiko getragen. Denn beileibe nicht alle impliziten Wissensbestände fließen signifikant in die KIA-Prognose ein und bieten damit konkreten Nutzen; vielmehr gilt dies nur für solche, die eine entsprechend hohe Gewichtung durch das algorithmische Modell erfahren. Schließlich wäre eine solch weitgehende Zurechnung kaum operationalisierbar, da jeweils unklar wäre, welche Tatsachen implizit in dem Modell repräsentiert sind – sie lassen sich eben nicht direkt aus diesem ablesen.

### bb) Zurechnung auf Grundlage der Figur des Sich-Verschließens

Demgegenüber erscheint es vorzugswürdig, eine Zurechnung impliziten Wissens von KIA nur ausnahmsweise und unter besonderen Voraussetzungen vorzunehmen: Wenn sich die Substitution expliziten Wissens durch implizites Wissen als bewusstes Sich-Verschließen darstellt. Damit lässt sich eine angemessene Risiko- und Interessenbalance herstellen, die richtigerweise hinter den Fragen der Wissenszurechnung steht.<sup>187</sup> Dies würde in der Tat bedeuten, der bisherigen organisatorischen Pflichtentrias der Speicherung, Weitergabe und Abfrage von Wissen eine vierte Dimension hinzuzufügen: diejenige der Gewinnung expliziten Wissens, allerdings nur unter eng eingegrenzten Voraussetzungen. Dies muss dann konsequenterweise für alle Formen arbeitsteiligen Vorgehens, nicht nur für Gesellschaften, gelten.<sup>188</sup>

### (1) Fallgruppen

Zwei Fallgruppen kommen in Betracht. Erstens die „Flucht in implizites Wissen“, von der dann gesprochen werden kann, wenn bereits vor dem Einsatz des KIA ein Verdacht hinsichtlich der inkriminierenden Tatsachen bestand und anstelle einer Sachverhaltsaufklärung ein nicht interpretierbares Modell eingesetzt wurde. Dies ist insbesondere dann der Fall, wenn sich aus den Inputdaten die relevante Tatsache plausibler Weise ableiten ließe. Im Fall des Altlastenbestands mag dies der Typ und die Dauer der vorherigen Nutzung sein, etwa bei längerer Nutzung durch einen toxische Substanzen verarbeitenden Betrieb.

Zweitens wird es Fälle geben, in denen der Output Verdacht erregen müsste. Denn auch wenn die Modelle komplex sind, so ist der Abgleich von Input und Output Daten verhältnismäßig einfach zu bewerkstelligen und kann routinemäßig in den Einsatz von KI integriert werden (*input-output testing*).<sup>189</sup> So ist etwa denkbar,

187 Siehe etwa *Spindler* (Fn. 10), S. 319 f.; *Drexler* (Fn. 173); siehe auch bereits oben, Fn. 163.

188 *Bohrer* (Fn. 133), S. 130; *Medicus* (Fn. 5), S. 12; vgl. auch *Spindler* (Fn. 10), S. 316; *Dauner-Lieb* (Fn. 144), S. 53.

189 *Kim* (Fn. 52), S. 190 f.; *Hacker* (Fn. 52), unter 4.2.

dass KIA eingesetzt werden, um zu bestimmen, welchem potentiellen Vertragspartner welche Konditionen für eine Kreditkarte angeboten werden sollen. Ergibt sich dabei, dass für Kunden aus sozialen Brennpunkten (Input) besonders komplexe Verträge mit versteckten Kosten (Output) vom KIA vorgeschlagen werden, so liegt der Verdacht nahe, dass letztlich mangelnde finanzielle Sophistikation und erwartbare Fehler der Kunden durch die Vertragskonstruktion ausgenutzt würden (sog. *exploitative contracts*).<sup>190</sup> Für derartige, aus Datenverarbeitung resultierende Angebotskampagnen gibt es empirische Belege;<sup>191</sup> auch in solchen Fällen sollte man von einem bewussten Sich-Verschließen, hier vor der Kenntnis der beschränkten Sophistikation und Rationalität der Vertragspartner, absehen.

## (2) Gründe

Diese Lösung gründet im Wesentlichen auf drei Argumenten: einem dogmatischen und zwei darüber hinausgehenden Sachgründen.

### (i) Dogmatik

Erstens kann sich eine Zurechnung von Wissen, das sich aufdrängte, im konkreten Fall aber nicht explizit erschlossen wurde, an eine dogmatische Figur anlehnen: Weitgehend konsentiert ist, dass Nicht-Wissen dem positiven Wissen dann gleichgestellt (bzw. eine Berufung darauf nach § 242 BGB versagt) ist, wenn der potentielle Wissensträger sich bewusst vor der positiven Kenntnis verschlossen hat.<sup>192</sup> Denn es handelt derjenige rechtsmissbräuchlich, der sich von positiver Kenntnis sehenden Auges abwendet, um deren negative Konsequenzen zu vermeiden.<sup>193</sup> Selbst ein pointierter Kritiker der Lehre der Wissensverantwortung wie *Grigoleit* erkennt diese Kategorie als notwendige Extension und Operationalisierung des Falls positiven Wissens an.<sup>194</sup>

### (ii) Korrespondenz von Vor- und Nachteil

Zweitens würde die Verneinung der Zurechnung in den genannten Fallgruppen dazu führen, dass der Anwender nicht interpretierbarer KI die Vorteile dieser potentiellen Methoden durch bewusste Ersetzung inkriminierenden expliziten Wissens durch implizites Wissen internalisieren, ihre Nachteile jedoch externalisieren könnte. Der Altlastenfall belegt sehr deutlich, dass, anders als im Bereich der Verhal-

190 P. Hacker, Personal data, exploitative contracts, and algorithmic fairness, 7 International Data Privacy Law 2017, S. 266 (269-272).

191 H. Ru, Hong/A. Schoar, Do Credit Card Companies Screen for Behavioral Biases?, NBER Working Paper 22360, 2016.

192 Buck (Fn. 1), S. 448 ff.; Taupitz (Fn. 5), S. 29; Goldschmidt (Fn. 169), S. 138 ff.; Gasteyer/Goldschmidt (Fn. 144), S. 117 f.

193 Ausführlich Buck (Fn. 1), S. 448 ff.

194 Grigoleit (Fn. 10), S. 175.

tenszurechnung,<sup>195</sup> mit dem Vorteil für den KIA-Verwender durchaus nicht typischerweise ein Vorteil auch für den Vertragspartner einhergeht. Vielmehr entsteht eine neue Form von Informationsasymmetrie, die für den informationell Unterlegenen nachteilig ist.

Eine völlige Freizeichnung entspräche daher nicht der marktwirtschaftlichen Korrespondenz von Nutzen und Risiko.<sup>196</sup> Vielmehr droht eine zunehmende Aushöhlung von wissensbasierten Haftungsnormen und eine Unterminierung des damit einhergehenden Anreizsystems zu sozial erwünschtem Verhalten. Eine Nichtzurechnung sich aufdrängenden Wissens setzt missliche Anreize dazu, sozial nützliche, aber für den jeweiligen Geschäftsherrn nachteilige Tatsachen von dem relevanten Wissensträger organisatorisch fernzuhalten bzw. den Wissenstransfer so zu organisieren, dass ein Nachweis des Wissens nicht möglich ist. Akteure könnten also bewusst Aufgaben an KIA zu delegieren, um sich damit Wissensressourcen zu erschließen, ohne für deren Inhalt geradestehen zu müssen. Dies löst nicht nur den Nutzen vom Risiko neuer Technologien; es führt tendenziell auch zu einer Unterproduktion von (sozial nützlichem) explizitem Wissen,<sup>197</sup> wie etwa der Fall der Altlasten zeigt. Die Figur des bewussten Sich-Verschließens ist daher auch rechtsvergleichend verbreitet und ökonomisch grundsätzlich effizient.<sup>198</sup>

Besteht jedoch kein derartiger Anlass, davon auszugehen, dass der KIA-Verwender sich konkretem Wissen bewusst verschlossen hat, kann auch dann eine Wissenszurechnung nicht vorgenommen werden, wenn implizite Wissensbestände Vorteile gebracht haben und grundsätzlich technisch zu heben gewesen wären. Dies würde vielmehr die Organisationspflichten des Unternehmens, die auch vom BGH immer mit Blick auf die erwartete Erheblichkeit und damit den konkreten Anlass interpretiert werden, in unzumutbarer Weise überspannen.

### (iii) KI-spezifisches Gleichstellungsargument

Drittens fordert auch das Gleichstellungsargument eine, wenn auch mit Augenmaß betriebene, anlassbezogene Zurechnung implizit repräsentierten Wissens. Dabei ist jedoch zu präzisieren, was mit Gleichstellung gemeint ist. In traditionellen Fällen der Wissenszurechnung betrifft dies die Gleichstellung von Fällen der Interaktion mit einem einzigen Individuum mit jenen der Interaktion mit einer Mehrzahl an arbeitsteilig verfahrenen Individuen, also die Aufspaltung von Wissen in personaler und zeitlicher Hinsicht.<sup>199</sup> Bei Fällen mit KI-Bezug tritt das Desiderat der Gleichbe-

195 Dazu oben, Fn. 47.

196 Dazu oben, Fn. 46.

197 Zur Unterproduktion von Wissen infolge von dessen Eigenschaft als öffentliches Gut, siehe *Hacker* (Fn. 40), S. 411 f.; vgl. auch *DeMott* (Fn. 168), S. 297-300.

198 *Wagner* (Fn. 10), S. 266 f.

199 BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 16.

handlung von Fällen der Aufgabenerledigung durch KIA mit jenen der Leistungserbringung durch Menschen hinzu. Natürliche Personen können typischerweise gerade nicht einzelne Aspekte einer Faktensammlung gezielt vergessen oder gezielt im unbewussten statt im bewussten Gedächtnis ablegen. Genau diese Möglichkeit würde die strikte Nichtzurechnung von implizitem KI-Wissen jedoch ermöglichen. Dies bedeutete unter anderem, dass diejenigen Unternehmen, die statt auf explizites Mitarbeiterwissen auf (mit implizitem Wissen operierende) KIA setzen, geringere erwartete Schadenskosten hätten und damit verzerrt niedrige Preise anbieten könnten, was wiederum alloкатive Ineffizienz impliziert.<sup>200</sup>

Zwar hat die Gleichstellungsthese auch beachtliche Kritik erfahren,<sup>201</sup> die vor allem darin gründet, dass eine erweiterte Wissenszurechnung arbeitsteilig verfahren- de Entitäten im Ergebnis schlechter stellt, da sie typischerweise deutlich mehr Wissensbestände zu verwalten haben als Einzelperson.<sup>202</sup> Dies kann jedoch aufgrund der gesteigerten technischen Möglichkeiten des Wissensaustauschs und -abrufs jedenfalls im KI-Bereich nur bedingt überzeugen. Typischerweise wächst zugleich mit der Quantität der Faktenbestände auch die Qualität der technischen Wissensverarbeitung. Größeres Wissensrisiko korrespondiert so mit besserer Möglichkeit zu (Wissens-)Risikomanagement.<sup>203</sup> Dies deutet auf den zutreffenden Kern der Gleichstellungsthese: die Ausbalancierung des Risikos von Informations- und Haftungsvorteilen durch Wissensaufspaltung mit Zumutbarkeitsgesichtspunkten.<sup>204</sup> Die Anlassbezogenheit vermittelt der Pflicht zur Gewinnung expliziten Wissens die nötige Beschränkung, so dass Möglichkeit und Notwendigkeit der Wissensproduktion in einem angemessenen Verhältnis stehen. Dass damit eine „wertende Öffnung“<sup>205</sup> des Tatbestandsmerkmals der „Kenntnis“ einhergeht, ist ebenso unbestreitbar wie, in den gezeichneten Konturen, letztlich sachgerecht.

### c) Zwischenergebnis

Die Differenzierung zwischen implizitem und explizitem KIA-Wissen ist zu Analysezwecken nützlich und für die Wissenszurechnung relevant. Die Einteilung richtet sich danach, mit welchem Aufwand die Tatsache aus dem Modell ausgelesen werden kann. Bei hohem Aufwand (implizites Wissen) wird man von einem tauglichen

200 Vgl. *Wagner* (Fn. 10), S. 259 und 263.

201 *Koller* (Fn. 182), S. 77 ff.; *Grigoleit* (Fn. 10), S. 190-192; *Spindler* (Fn. 10), S. 315-317; *Buck* (Fn. 1), S. 321-323; *Flume* (Fn. 144), S. 443, 445; differenzierend nach Geschäftsbereichen *Baum* (Fn. 46), S. 186-190.

202 So etwa *Grigoleit* (Fn. 10), S. 190 f.; *Koller* (Fn. 182), S. 78; *Buck* (Fn. 1), S. 318; *Baum* (Fn. 46), S. 207; *Goldschmidt* (Fn. 169), S. 33.

203 Siehe etwa die Vorgaben in § 26 f. VAG und § 25a KWG für Banken und Versicherungen, sowie BaFin, Rundschreiben 09/2017 (BA) – Mindestanforderungen an das Risikomanagement – MaRisk, besonders AT 7.2.

204 Vgl. *Spindler* (Fn. 10), S. 319 f. und S. 322.

205 Formulierung bei *Bohrer* (Fn. 133), S. 129; kritisch etwa *Flume* (Fn. 144); Rechtfertigung mangelnder gesetzlicher Grundlage bei *Medicus* (Fn. 5), S. 13.



Zurechnungsgegenstand nur bei einem konkreten Anlass ausgehen können, der ein bewusste Sich-Verschließen nahelegt. Explizites Wissen, das mit vergleichsweise geringem Aufwand aus dem Modell ersichtlich ist, stellt hingegen grundsätzlich einen relevanten Zurechnungsgegenstand dar, so wie man auch eine gewisse Gedächtnisanstrengung von Menschen normativ erwarten muss. Der konkret zu betreibende Aufwand sollte, wie auch die Pflichten der Speicherung, Weitergabe und Abfrage, durch Zumutbarkeitserwägungen beschränkt werden. Letztlich bietet die Lehre von der Wissensverantwortung, mit ihrer von einer Abwägung zwischen Organisationsaufwand und Erheblichkeit für konkrete Risiken und Rechtsgutsgefährdungen getragenen Zumutbarkeitsskala,<sup>206</sup> ein solides Fundament für einen angemessenen Interessenausgleich im Einzelfall, auch und gerade im Bereich der KI.

#### 4. Wissenszurechnung im Deliktsrecht nach dem BGH

Neben dem Vertragsrecht stellen sich ähnlich Fragen der Wissenszurechnung auch im Deliktsrecht. Dies kann eigenständige Relevanz erlangen, wenn keine (quasi-)vertragliche Sonderverbindung zwischen dem KIA-Verwender und dem Geschädigten besteht, etwa wenn der KIA durch ein selbständiges, vom Vertragspartner unabhängiges Unternehmen eingesetzt wird,<sup>207</sup> oder wenn vertragliche Ansprüche bereits verjährt sind.<sup>208</sup>

##### a) Wissenszurechnung auf Grundlage der BGH-Rechtsprechung

Maßgeblich ist hier insbesondere § 826 BGB, der auf einem besonderen Wissensvorsprung beruhende primäre Vermögensschäden sanktioniert und in diesem Bereich § 831 Abs. 1 BGB zunehmend zurückdrängt.<sup>209</sup> So wird auch in dem *Fondsprospektfall* des BGH neben § 826 BGB eine Haftung nach § 831 BGB mit keinem Wort erwähnt.

##### aa) § 826 BGB

Im Rahmen des § 826 BGB interessiert besonders, inwiefern die Kenntnis der die Sittenwidrigkeit begründenden Tatsachen zugerechnet werden kann, wenn diese Tatsachen allein dem KIA explizit oder implizit bekannt sind. Bei einer Zurechnung analog § 31 BGB kommen zwei Möglichkeiten in Betracht. Erstens könnte der KIA ein verfassungsmäßig berufener Vertreter oder Repräsentant des Unternehmens sein, der selbst *in toto* § 826 BGB verwirklicht. Damit stellt sich auf

206 Für eine ähnliche, am Maßstab der Verhältnismäßigkeit des Aufwands zur Verbesserung eines KI-Modells orientierte Lösung im Bereich der algorithmischen Diskriminierung siehe *Hacker* (Fn. 52), S. 19 f.

207 Bei KI-Beratung kann jedoch § 311 Abs. 3 BGB relevant werden, was hier nicht weiter zu verfolgen ist.

208 Vgl. G. Wagner, F. J. Säcker et al. (Hrsg.), Münchener Kommentar BGB, Band 6, 7. Aufl., München 2017, Vor § 823 Rn. 79.

209 *Reuter* (Fn. 8), S. 311; grundsätzlich zur sinkenden Bedeutung von § 831 BGB G. Wagner, Sinn und Unsinn der Unternehmensstrafe, ZGR 2016, S. 112 (129-131).

Grundlage der vom BGH vertretenen Theorie des § 826 BGB inhärenten, persönlichen moralischen Vorwurfs die schwierige Frage, inwiefern dieser gegenüber einem KIA erhoben und KI allgemein als moralische Maschine (*moral machine*) verstanden werden kann. Dies wird zwar in der Maschinenethik lebhaft diskutiert,<sup>210</sup> braucht jedoch hier nicht im Einzelnen ausgeführt werden, da die Einordnung eines KIA als nach § 31 BGB relevanter Vertreter wohl Komponenten von genereller KI voraussetzt, deren Entwicklung gegenwärtig nicht absehbar ist.

Ist demnach die direkte Erfüllung von § 31 BGB durch den KIA bis auf weiteres unwahrscheinlich, so verbleibt als zweite Möglichkeit, das Wissen des KIA nach den Grundsätzen der Wissenszurechnung, letztlich der Lehre von der Wissensverantwortung, direkt der juristischen Person<sup>211</sup> selbst bzw. deren nach § 31 BGB relevanten Organen<sup>212</sup> zuzurechnen. Diese Möglichkeit wurde nun durch den BGH im *Fondsprospektfall* im Ergebnis zunichte gemacht unter Berufung auf den personalen Charakter des Unrechtsgehalts von § 826 BGB: Jedenfalls dieser entzieht sich nach dem *Fondsprospekt*-Urteil des BGH einer Zurechnung.<sup>213</sup> Gleiches gilt für das voluntative Element des Vorsatzes.<sup>214</sup>

#### bb) § 831 Abs. 1 BGB

Daher verbleibt letztlich nur eine analoge Anwendung von § 831 Abs. 1 BGB, die ähnlich wie im Bereich der Handlungszurechnung gestaltet werden kann.<sup>215</sup> Unproblematisch ist die Handlung eines menschlichen Verrichtungsgehilfen auf Grundlage des KIA-Wissens. Wenn hingegen Wissen und Handlung im KIA zusammenfallen, wird man den hinreichend autonomen KIA selbst analog einem Verrichtungsgehilfen qualifizieren können, der objektiv rechtswidrig einen Deliktstatbestand verwirklicht, in dessen Rahmen Wissen relevant ist (mit besonderen Problemen allerdings bei § 826 BGB).<sup>216</sup> Dabei wird freilich § 831 Abs. 1 BGB nicht

210 W. Wallach/C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, Oxford 2008; K. Abney, *Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed*, in: P. Lin et al. (Hrsg.), *Robot Ethics*, MIT Press, Cambridge, MA 2012, S. 35; P. Bello/S. Bringsjord, *On how to build a moral machine*, 32 *Topoi* 2013, S. 251.

211 Wagner (Fn. 152), S. 523; ders. (Fn. 10), S. 210 (jeweils unter Verweis auf BGH NJW 2001, S. 359 (360); diese Entscheidung ließe sich jedoch auch anders deuten, siehe Grigoleit (Fn. 162), Bl. 3).

212 So Reuter (Fn. 8), S. 311 f.; so wohl auch BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 21 (Zurechnung zum Leiter des Liegenschaftsamts).

213 BGH NJW 2017, S. 250 Rn. 16; zudem würde sich auch in diesem Kontext die Problematik der Zurechnung impliziten Wissens stellen.

214 BGH NJW 2017, S. 250 Rn. 26.

215 Oben, Teil C.II.

216 Siehe, außerhalb der KI-Problematik, zu dieser Konstellation Gasteyer/Goldschmidt (Fn. 144), S. 119; Wagner (Fn. 152), S. 525; allerdings wird man, wenn der BGH bei seiner Interpretation der persönlichen moralischen Vorwerfbarkeit bleibt, § 826 BGB nur schwer auf KIA anwenden können; siehe oben, Fn. 210. Zudem müsste der KIA dafür auch im Rahmen von § 831 BGB mit Vorsatz (oder äquivalent dazu) gehandelt haben, siehe BGH NJW 2014, S. 1380, Rn. 11.

als Norm der Wissenszurechnung genutzt,<sup>217</sup> sondern als eigenständige Anspruchsgrundlage, welche die Haftung erstreckt auf fremde Handlungen, für deren Deliktsqualität wiederum Wissen relevant sein kann. Jedoch bleibt es auch hier bei der bereits beschriebenen Exkulpationsmöglichkeit, so dass die praktische Relevanz nicht zu hoch sein dürfte.

## b) Grundsätze für die KIA-Wissenszurechnung im Deliktsrecht

Die Probleme der deliktischen Wissenszurechnung im KI-Bereich kristallisieren sich daher an der Erfüllung von § 826 BGB. Für eine detaillierte Untersuchung ist hier kein Raum; drei zentrale Punkte können aber festgehalten werden.

Erstens sollte, in Nachvollzug einer beim Arglistenfordernis schon vollzogenen Entwicklung,<sup>218</sup> auch bei § 826 BGB auf einen persönlichen moralischen Vorwurf verzichtet werden,<sup>219</sup> die Garantie eines rechtlich rekonstruierten, ethischen Minimalstandards<sup>220</sup> als lediglich *ein* Element in dem breiten Anwendungsbereich von § 826 BGB verstanden werden. Die Verletzung dieses rechtlich qualifizierten Minimalstandards kann als ein hinreichendes, nicht aber notwendiges Kriterium der Sittenwidrigkeit gelten. Damit öffnet sich § 826 BGB nicht nur einer breiteren Anzahl von Fallkonstellationen (Ergänzungsfunktion),<sup>221</sup> sondern nur ein solches Verständnis kann die bisherige Kasuistik plausibel erklären, insbesondere wenn man an kapitalmarktrechtliche Aufklärungspflichtverletzungen denkt, deren moralischer Unwertgehalt doch eher zweifelhaft erscheint.<sup>222</sup> Ein Beharren auf der persönlichen moralischen Vorwerfbarkeit als *notwendige* Bedingung hingegen führt, wie der vom BGH entschiedene *Fondsprospektfall* eindringlich vor Augen führt, zu erheblichen Problemen in der Anwendung von § 826 BGB auf juristische Personen; bei KIA verschärfen sich diese Probleme noch. Ein individualistisch konzipiertes Haftungsrecht kommt an seine Grenzen, wenn zunehmend Aufgaben an KIA ausgelagert werden.<sup>223</sup> Zur Wahrung der Filterungsfunktion der Sittenwidrigkeit kann dann auf die besondere Schwere einer Pflichtverletzung oder rechtswidrigen Handlung abgestellt werden.<sup>224</sup> Damit ist einerseits eine Grenze zu bloß objektiv rechts- oder vertragswidrigem Verhalten markiert,<sup>225</sup> andererseits sind solche Pflichtverletzungen erfasst, die eine gewisse Erheblichkeit für den Geschädigten entfalten und

217 Dazu etwa *Buck* (Fn. 1), S. 183 ff., besonders S. 186 f.

218 BGH NJW 1990, S. 975 (976); dazu kritisch *Dauner-Lieb* (Fn. 144), S. 55.

219 *J. Oechsler*, in: Staudinger, BGB, Berlin 2018, § 826 Rn. 26; *Wagner* (Fn. 152), S. 524.

220 Vgl. *Grigoleit* (Fn. 10), S. 171.

221 *C. Förster*, in: Bamberger/Roth/Hau/Poseck (Hrsg.), Beck'scher Online Kommentar BGB, 46. Aufl., München 2018, § 826 Rn. 2.

222 *Wagner* (Fn. 152), S. 524.

223 Vgl., für juristische Personen, nicht KI, *Wagner* (Fn. 152), S. 523.

224 *K. Larenz/C.-W. Canaris*, Lehrbuch des Schuldrechts: Besonderer Teil, Band II/2, 13. Aufl., München 1994, § 78 II 1. b), S. 451.

225 Vgl. BGH NJW 2017, S. 250 Rn. 16; *Wagner* (Fn. 208), Rn. 13 f.

bei denen die Abweichung vom Verkehrsüblichen „unerträglich“ ist.<sup>226</sup> Bei KIA kann hier wiederum, in Ermangelung technischer Standards, auf eine anthropo-parallele Sichtweise zur Bestimmung der Schwere abgestellt werden. Damit kann der vom Bundesverfassungsgericht betonten Funktion des § 826 BGB, flexibel neue soziale Umstände zu erfassen,<sup>227</sup> Rechnung getragen werden.

Zweitens sollten im Rahmen des so verstandenen § 826 BGB die oben beschriebenen Grundsätze der Wissenszurechnung aus dem Vertragsrecht übernommen werden.<sup>228</sup> Zwar hat sich der Gläubiger seinen deliktischen Schuldner, anders als im Vertragsrecht, nicht ausgesucht. Für eine Übertragung spricht jedoch, dass, wie das KG zu Recht betont hat, die Zurechnung auch auf *allgemeinen* Verkehrsschutzgesichtspunkten, nicht (nur) auf spezifischen Auswahlentscheidungen im rechtsgeschäftlichen Verkehr aufruft.<sup>229</sup> Gerade die Kriterien der Risikoschaffung und -beherrschung sind anerkannte Grundlagen deliktischer Pflichten;<sup>230</sup> auch im deliktischen Verkehr beherrscht derjenige, der arbeitsteilig vorgeht, die genannten drei Risiken der Wissensaufspaltung besser als die Allgemeinheit. So setzt etwa nur eine Übertragung der Maßstäbe der Wissenszurechnung auch im deliktischen Bereich hinreichende Anreize zu einer Wissensorganisation, die nachteilige Information nicht bewusst delegiert, sondern Wissensverantwortung praktiziert.

Kommt es mithin für die Schwere der Pflichtverletzung im Rahmen der Sittenwidrigkeit auf Wissensbestände an, so sollten diese auf Basis der Lehre von der Wissensverantwortung zurechenbar sein. Errechnet etwa der KIA explizit einen hohen Altlastenscore, der verschwiegen wird, könnte man, entgegen dem BGH, von einer Sittenwidrigkeit qua Wissensverantwortung ausgehen.<sup>231</sup> Gleiches gilt für solche besonders schwerwiegenden Umstände, die implizit encodiert sind, vor denen sich der Geschäftsherr jedoch bewusst verschlossen hat.

Es bleibt dann drittens noch das Problem der Zurechnung der voluntativen Elemente des Vorsatzes im Deliktsrecht, die der BGH bei § 826 BGB klar verneint hat;<sup>232</sup> eine ausführliche Erörterung würde auch hier zu weit führen. Schon im humanen Kontext zeigen sich hier in besonderer Schärfe die Gegensätze der beiden

226 Zur Sittenwidrigkeit als Unerträglichkeit bei BGH NJW 1970, S. 657 (658); siehe auch Förster (Fn. 221), Rn. 21; hier lassen sich auch die vom BGH angeführten Kriterien der Ziele, Mittel, Gesinnung und Folgen (BGH NJW 2017, S. 250 Rn. 16) relevant machen.

227 BVerfG NJW 2015, S. 3083 Rn. 15.

228 So auch G. Wagner, in: F. J. Säcker et al. (Hrsg.), Münchener Kommentar BGB, Band 6, 7. Aufl., München 2017, § 826 Rn. 38 f.; aA Frisch (Fn. 152).

229 KG Urt. v. 27.8.2015 – 2 U 42/09, BeckRS 2016, 17728, Rn. 52; so auch Bohrer (Fn. 133), S. 129; einschränkend Buck (Fn. 1), S. 318; aA Spindler (Fn. 10), S. 318; Frisch (Fn. 152).

230 Siehe oben, Fn. 125; BGH NJW-RR 2018, S. 726 (727).

231 Vgl. auch BGH, Urteil vom 10.12.2010 – V ZR 203/09, BeckRS 2011, 01685, Rn. 21: Schwammbefall von „überragende[r] rechtliche[r] Bedeutung“; für Altlastenfälle Wagner (Fn. 152), S. 525.

232 BGH NJW 2017, S. 250 Rn. 26.

Perspektiven auf die Wissenszurechnung: Einerseits lässt sich Vorsatzhaftung durch Wissenszurechnung bei Auseinanderfallen von Wissen und Handlung dogmatisch aufgrund des konstitutiven subjektiv-voluntativen Elements, anders als Fahrlässigkeitshaftung, dogmatisch letztlich nur durch eine Fiktion begründen<sup>233</sup> (wenn nicht ein bewusstes Sich-Verschließen vorliegt).<sup>234</sup> Eine sachgerechte Risiko- und Verhaltenssteuerung benötigt andererseits wertende Elemente der Zurechnung auch im Vorsatzbereich,<sup>235</sup> wie die Arglistrechtsprechung zeigt.

Legt man Letzteres zugrunde, so ließe sich das Problem der voluntativen Elemente grundsätzlich, unter Anwendung von objektiven Erfahrungssätzen auf Grundlage zugerechneten Wissens,<sup>236</sup> bei KIA auch in anthropo-paralleler Weise lösen. Wenn sich der BGH des Spannungsverhältnisses seiner Arglistrechtsprechung zum *Fondsprospektfall* noch stärker bewusst wird,<sup>237</sup> mag die Frage der Vorsatzhaftung bei Einsatz von KIA dereinst Anlass für eine Entscheidung des Großen Senates für Zivilsachen bieten. Bis dahin erscheint eine Kehrtwende des für Deliktsrecht zuständigen VI. Zivilsenats jedoch heute weniger wahrscheinlich als eine Bejahung der skizzierten Anwendbarkeit der Zurechnungskriterien auf KIA in den übrigen besprochenen Kontexten.

## E. Schluss

Der Beitrag entwickelt auf Grundlage des bestehenden Rechts und seiner Dogmatik Lösungen für Zurechnungsprobleme, die durch den Einsatz von mit künstlicher Intelligenz ausgestatteten Agenten anstelle von menschlichen Mitarbeitern entstehen. Dabei kann als maßgeblicher Zurechnungsgrund sowohl in den Fällen der Verhaltenszurechnung als auch der Wissenszurechnung die Schaffung von Risiken gelten, die durch den Geschäftsherrn besser als durch den Vertragspartner oder deliktisch Geschädigten beherrscht werden können. Die spezifischen Risiken differieren jedoch. Geht es im Bereich der Verhaltenszurechnung vor allem um die Zuweisung des Personalrisikos und dessen im KI-Kontext schwierige Abgrenzung zu einem Produktrisiko, so sind im Bereich der Wissenszurechnung das strategische Nachweisrisiko, das strategische Delegationsrisiko und das zufällige Aufspaltungsrisiko relevant.

233 Gegen eine auf Zurechnung basierende Vorsatzhaftung daher *Grigoleit* (Fn. 10), S. 178; *Flume* (Fn. 144); kritisch auch *Buck* (Fn. 1), S. 334 f.

234 Vgl. auch *Buck* (Fn. 1), S. 84 und S. 419 sowie die Nachweise oben, Fn. 192–198.

235 Für eine Zurechnung etwa *Baum* (Fn. 46), S. 153 und S. 301 f.; *Wagner* (Fn. 228), Rn. 38 f.; *ders.* (Fn. 152), S. 524 f.; nuancierter *ders.* (Fn. 10), S. 265 f. (Ablehnung unter der Prämisse der Effizienz der Haftungsbeschränkung auf Vorsatz).

236 So explizit BGH NJW 1992, S. 975 (976); dazu auch *Buck* (Fn. 1), S. 333 und S. 423; *K. Reischl*, „Wissenszusammenrechnung“ auch bei Personengesellschaften?, JuS 1997, S. 783 (788); *Förster* (Fn. 221), Rn. 39 f.

237 Dazu *Grigoleit* (Fn. 10), S. 165 f.

Auf dieser Grundlage lassen sich die drei spezifischen, hier behandelten Probleme der Zurechnung beim Einsatz von KI lösen. Das erste, die Frage der grundsätzlich parallelen Behandlung von menschlichen und KI-Entscheidungsträgern, stellt sich sowohl im Bereich der Verhaltens- als auch der Wissenszurechnung. Sie lässt sich für hinreichend autonome KIA bejahen. Das Problem einer adäquaten Konzeption des Verschuldens von KIA hingegen ist spezifisch im Bereich der Verhaltenszurechnung angesiedelt und kann insbesondere durch eine analoge Anwendung von § 278 S. 1 BGB unter Berücksichtigung von funktionalen Verschuldensäquivalenzen gelöst werden. Die Frage des Umgangs mit implizitem Wissen hingegen fällt allein in den Kontext der Wissenszurechnung und verlangt nach einer mit Augenmaß betriebenen Anpassung der Lehre von der Wissensverantwortung.