

Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten*

Prof. Dr. Dr. h.c.mult. *Gunther Teubner*, Frankfurt am Main

Inhaltsübersicht

- I. Digitalisierung: Verantwortungslücken und Dogmatikdefizite*
 - 1. *Das Pfeifen im Walde oder: Die Angst vor dem Homo ex machina*
 - 2. *Konkrete Haftungslücken*
 - 3. *Dogmatische Fehlkonstruktionen*
 - 4. *Volle Rechtssubjektivität?*
- II. Aktanten: Autonomierisiko*
 - 1. *Anthropomorphisierung?*
 - 2. *Kommunikation mit Aktanten*
 - 3. *Gradualisierte digitale Autonomie*
 - 4. *Rechtliche Autonomiekriterien*
 - 5. *Autonome Entscheidungen der Softwareagenten*
- III. Rechtsfragen des Autonomierisikos*
 - 1. *Digitale Verträge*
 - 2. *Vertragliche Haftung*
 - 3. *Deliktische Haftung*
 - 4. *Gefährdungshaftung oder digitale Assistenzhaftung?*
- IV. Hybride: Verbundrisiko*
 - 1. *Mensch-Maschinen-Assoziationen*
 - 2. *Hybride als Rechtssubjekte?*
- V. Multi-Agenten-Systeme: Vernetzungsrisiko*
- VI. Ergebnis*

I. Digitalisierung: Verantwortungslücken und Dogmatikdefizite

1. Das Pfeifen im Walde oder: Die Angst vor dem Homo ex machina

„Verhaltensbiologisch dient das Pfeifen ... als Geräuschmarkierung, wie es bei der Revierabgrenzung üblich ist, und soll Raubtiere, die dem Menschen gefährlich werden könnten, auf Distanz halten. Durch das Signalisieren des Revieranspruchs wird von eigener Unterlegenheitsangst abgelenkt“ (Wikipedia) Die gefährlichen Raubtiere, die in die Reviere des Zivilrechts eindringen, sind nicht-kontrollierbare Algorithmen aus der digitalen Welt, Roboter, allgemeiner Softwareagenten,¹ die über

* Für kritisch-konstruktive Kommentare danke ich Marc Amstutz, Pasquale Femia, Malte Gruber und Dan Wielsch.

¹ Zur Terminologie: „e-person“ dürfte zu nahe an der Bedeutung einer vollen Juristischen Person haften, „Computer“ wiederum zu nahe an der Hardware-Komponente. „Roboter“ ist zu eng, da er regelmäßig Software plus Verkörperung bedeutet. „Algorithmus“ scheint korrekt, aber dies bezeichnet oft nur die mathematische Formel und wird ihrer zugrundeliegenden Materialität und der Dynamik ihrer Interaktion mit der realen Welt nicht gerecht. „Elektronischer Agent“ trifft die Sache recht gut, weil „Agent“ das Element des Handelns betont und Anklänge an Stellvertretung und Hilfsperson enthält. Aber dies berücksichtigt nur elektronische Signale, nicht aber biologische Signale oder Lichtsignale. „Digitaler Agent“ dürfte die Sache am besten treffen. Im Text aber soll der Begriff „autonome Softwareagenten“ verwendet werden, weil er sich inzwischen im Sprachgebrauch am ehesten durchgesetzt hat.

hohe Intelligenz und Lernfähigkeit verfügen und großartige Leistungen vollbringen, zugleich aber auch für die Menschen neuartige ungeahnte Gefahren erzeugen. Doch mit selbstbewussten Geräuschmarkierungen halten viele juristische Autoren die digitalen Raubtiere auf Distanz: Sämtliche Gefährdungen durch die neuartige digitale Spezies, im Vertragsrecht wie im Haftungsrecht, könnten „mit tradierten rechtlichen Instrumenten bewältigt werden“.² Sie sehen keinerlei Anlass, von den bewährten Zurechnungstechniken, die das Computerverhalten stets als Menschenverhalten ausgeben, abzugehen. Softwareagenten könnten rechtlich ohne Probleme als bloße Maschinen, als willige Instrumente in den Händen ihrer menschlichen Herren behandelt werden.³

Informationstechnologierechtsexperten, etwa Spindler, sind dagegen schon vorsichtiger.⁴ Die Rechtsfragen seien nur „einstweilen geklärt“, jedenfalls für durch Programmierung determinierte Softwareagenten. Sobald aber der entscheidende Schritt, dass Softwareagenten über Künstliche Intelligenz verfügten, vollzogen sei, müsse die Rechtsgeschäftslehre, besonders das Erklärungsbewusstsein, und womöglich auch das Haftungsrecht, neu überdacht werden. Das jedoch sei eine Frage, die sich erst in fernen, unbestimmten Zukünften stelle.

Doch ist dies wirklich nur eine Zukunftsfrage? Stellt sich nicht schon heute mit Dringlichkeit das Problem, dass autonome Softwareagenten bisher nicht bekannte Risiken erzeugen, so dass ihnen nur mit neuen zivilrechtlichen Normen begegnet werden kann? Oder gar, dass den Softwareagenten ein eigenständiger privatrechtlicher Status zuerkannt werden muss? Letztlich entscheidend dürfte sein, ob die rasanten digitalen Entwicklungen schon jetzt schwer erträgliche Verantwortungslücken aufgerissen haben. Deshalb lautet die Gegenthese, die in der Fachliteratur, aber auch vom Europäischen Parlament vertreten wird⁵: Unausweichlich verursachen Softwareagenten diese Verantwortungslücken, da ihr autonomes Handeln einen massiven Kontrollverlust menschlicher Akteure mit sich bringt. Zugleich aber nimmt die Abhängigkeit der Gesellschaft von autonomen Softwareagenten im großen Maßstab zu und es ist äußerst unwahrscheinlich, dass die Gesellschaft auf deren Nutzung verzichtet.⁶ Das Zivilrecht steht somit vor der Wahl: *Entweder verleiht es autonomen Softwareagenten einen eigenständigen Rechtsstatus und behandelt sie selbst als verantwortliche Akteure oder es häufen sich immer mehr „Unfälle“, ohne dass ein Verantwortlicher dafür einsteht.* Die

² MüKoBGB/Busche, 7. Aufl. 2015 Vor § 145 Rn. 37.

³ BGHZ 195, 126, Rn. 17; Palandt/Ellenberger, BGB 77. Aufl. 2018, vor § 116 Rn. 2; Staudinger/Singer BGB 2017 Vor §§ 116 ff. Rn. 57; Soergel/Hefermehl BGB 13. Aufl. 1999 vor § 116 Rn. 30; Medicus/Petersen, Bürgerliches Recht, 26. Aufl., 2017 Rn. 256; Cornelius MMR 2002, 353, 355; Sorge, Softwareagenten: Vertragsschluss, Vertragsstrafe, Reugeld, 2006, 36; Kirn/Müller-Hengstenberg KI – Künstliche Intelligenz 2015, 59, 67; Bräutigam/Klindt NJW 2015, 1137, 1138.

⁴ Spindler JZ 2016, 805, 816; Spindler, in: Hilgendorf (Hg.), Robotik im Kontext von Recht und Moral, 2014, 63, 64.

⁵ Europäisches Parlament www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//...582...//DE 2016, 5 ff. Auf das neuartige Autonomierisiko weisen hin: Mayinger Die künstliche Person 2017, 213 ff.; Zech, in: Gless/Seelmann (Hg.), Intelligente Agenten und das Recht, 2016, 163, 175 f.; Janal, in: Gless/Seelmann (Hg.), Intelligente Agenten und das Recht, 2016, 139, 158; Beck AJP 2017, 183, 184. Zwar räumen auch Spindler (Fn. 4, 2016), 816 und Schaub JZ 2017, 342, 346 gewisse Haftungslücken ein, das aber ist für sie kein Anlass, die bewährten rechtlichen Instrumente aufzugeben.

⁶ Matthias, Automaten als Träger von Rechten. 2. Aufl., 2010, 15.

Dynamik der Digitalisierung erzeugt unaufhörlich verantwortungsfreie Räume, die sich in Zukunft ausweiten werden.⁷

2. Konkrete Haftungslücken

Welches sind konkret die Verantwortungslücken? In der Informationswissenschaft werden folgende Sachverhalte genannt: Die Defizienzen entstehen in der Praxis dann, wenn die Software von Teams hergestellt werden, wenn Managemententscheidungen ebenso wichtig sind wie Programmierentscheidungen, wenn Dokumentationen über Anforderungen und Spezifizierungen eine große Rolle im resultierenden Code spielen, wenn trotz der Tests der Code-Genauigkeit viel von „off the shelves“-Komponenten abhängt, deren Herkunft und Genauigkeit unklar ist, wenn die Performanz der Software das Resultat der begleitenden Kontrollen und nicht der Programmerstellung ist, wenn automatisierte Instrumente in der Konstruktion der Software benutzt werden, wenn die Arbeitsweise der Algorithmen von seinen *interfaces* oder sogar vom Systemverkehr beeinflusst wird, wenn die Software in nicht voraussehbarer Weise interagiert, wenn die Software mit Wahrscheinlichkeiten arbeitet oder über Anpassungsfähigkeit verfügt oder selbst das Resultat eines anderen Programmes ist.⁸

Im Recht hat das Auftauchen von Softwareagenten folgende Haftungslücken schon heute erzeugt:

Computervernetzung: Die am schwierigsten zu korrigierende Haftungslücke entsteht im Falle multipler Kausalität bei schädigenden Handlungen mehrerer Computer, wenn diese untereinander vernetzt sind. Mit den Haftungsnormen des geltenden Rechts kommt man hier nicht weiter.⁹ Auch de lege ferenda ist noch kein brauchbarer Vorschlag ersichtlich. Im Falle des *High Frequency Trading* ist dies Risiko offensichtlich geworden.¹⁰ „Wer trägt das Risiko, wenn, wie am 6. Mai 2010, ein Flash-Crash ein gewaltiger Kursverlust des Dow Jones an der Börse an New York dadurch verursacht wurde, dass sich die Algorithmen der Wall Street, welche die Händlersysteme beherrschen, sich für einige Zeit unkontrolliert und unverständlich verhielten und einen Milliardenverlust bewirkten?“¹¹

Big Data: Weitere Haftungslücken entstehen durch Fehleinschätzungen fehlerhafter Big Data-Vorhersagen. Big Data wird dazu eingesetzt, mithilfe außerordentlich großer bestehender Datenmengen zu prognostizieren, wie sich bereits bestehende gesellschaftliche Tendenzen oder Epidemien entwickeln und – notfalls – beeinflussen lassen.¹² Wenn der Grund für die fehlerhafte Berechnung nicht eindeutig festgestellt werden kann, bestehen die Schwierigkeiten, Kausalität und Verschulden zu bestimmen.

⁷ So die zentrale These von *Matthias* (Fn. 6), 111, 113 ff.

⁸ So für viele Analysen *Floridi/Sanders*, in: *Anderson/Anderson* (Hg.), *Machine Ethics*, 2011, 184, 205.

⁹ So deutlich *Spiecker* CR 2016, 698 ff.

¹⁰ Dazu aber *Gruber*, *Soziale Systeme* 2013/14, 327, 330 f.

¹¹ *Kirn/Müller-Hengstenberg* <https://fzid.uni-hohenheim.de/71978.html> 2014, 1, 1.

¹² *Hackenberg*, in: *Hoeren/Sieber/Holznapel* (Hg.), *Handbuch Multimedia-Recht*. 45. Auflage, 39.EL 2014, Teil 16.7 Rn. 58.

Digitaler Vertragsbruch: Immer dann, wenn ein autonomer Softwareagent Vertragspflichten verletzt, dann soll, da er nicht über rechtliche Handlungsfähigkeit verfügt, nicht etwa die Gehilfenhaftung des § 278 eingreifen, sondern eine Haftung ausschließlich dann eintreten, wenn der menschliche Geschäftsherr selbst eine schuldhaftige Pflichtverletzung nach §§ 276, 280 begeht. Hier entsteht eine problematische Haftungslücke: Kann der Betreiber nachweisen, dass der Softwareagent korrekt eingesetzt wurde, ohne dass der Betreiber selbst eine Vertragspflicht verletzt hat, dann ist der Betreiber von jeglicher Haftung befreit.¹³ Soll dann der Kunde den vom Computer der Gegenseite verursachten Schaden tragen?

Außervertragliche Haftung: Im Bereich außervertraglicher Haftung tritt ein ähnliches Problem auf, weil es bei verschuldensabhängiger Haftung nur auf die Pflichtverletzung des Betreibers, Herstellers oder Programmierers ankommen soll. Halten die beteiligten Menschen diese Pflichten ein, dann trifft sie keine Haftung.¹⁴ Selbst wenn die Gerichte kompensatorisch ihre Sorgfaltspflichten überspannen und damit die Grenzen der Zumutbarkeit überschreiten, wird die Haftungslücke nicht geschlossen.¹⁵ Bei Schädigungen durch autonome Roboter führt dies, anders als bei Kraftfahrzeugschädigungen, bei denen zum Glück regelmäßig die Gefährdungshaftung für Betriebsgefahr der Kfz eingreift, dazu, dass der Geschädigte ohne Schutz bleibt.

Gefährdungshaftung: Selbst die zahlreichen rechtspolitischen Vorstöße, die de lege ferenda vorschlagen, Schädigungen durch Softwareagenten mit Gefährdungshaftungsnormen zu begegnen, können erhebliche Haftungslücken nicht vermeiden. Die Leitprinzipien der Gefährdungshaftung können schwerlich als Vorbild dienen, da sie nicht auf das digitale Risiko zugeschnitten sind.

Computererklärungen: Eine fragwürdige Haftungslage ergibt sich schließlich auch im Recht der rechtsgeschäftlichen Computererklärungen, wenn der Softwareagent selbst wie ein „*falsus procurator*“ fälschlich einen Dritten als Vollmachtgeber angibt. In solchen Fällen soll das Risiko vollständig beim Geschäftsherrn des Softwareagenten liegen. Dies sehen viele Autoren als eine übermäßige Belastung an, die sich nicht rechtfertigen lässt, besonders hoch bei distribuierten Aktionen oder bei self-cloning.¹⁶

3. Dogmatische Fehlkonstruktionen

Die Verantwortungsdefizite entstehen, solange die Dogmatik darauf besteht, auf die neuen digitalen Realitäten nur mit dem hergebrachten begrifflichen Instrumentarium zu reagieren. Um dann aber dennoch mit den digitalen Entwicklungen einigermaßen Schritt zu halten, sieht sie sich gezwungen, mit fragwürdigen dogmatischen Hilfskonstruktionen wenigstens manchen der bisher unbekanntenen Risiken der Softwareagenten zu begegnen. Im Recht der Willenserklärungen hält sie unbeirrt daran fest, dass nur natürliche Personen in der

¹³ Z.B. MüKoBGB/*Grundmann*, 7. Aufl. 2016, BGB § 278 Rn. 46; *Kirn/Müller-Hengstenberg* (Fn. 11), 16.

¹⁴ Z.B. *Spindler* (Fn. 4, 2016), 816; ebenso *Bräutigam/Klindt* (Fn. 3), 1140.

¹⁵ Die Tendenz zur Überspannung von Pflichten als Kompensation für dogmatische Defizite kritisieren *Gruber*, *Bioinformatikrecht: Zur Persönlichkeitsentfaltung des Menschen in technisierter Verfassung*, 2015; *Janal* (Fn. 5), 155.

¹⁶ *Sartor* Proceedings of the Workshop on the Law of Electronic Agents http://www.cirfid.unibo.it/*sartor/sartorpapers/g Sartor2002_agenti_software.pdf 2003.

Lage seien, rechtsverbindliche Erklärungen abzugeben. Deshalb muss sie mit problematischen Fiktionen arbeiten. Im Recht der vertraglichen und außervertraglichen Haftung muss sie, wenn Schädigungen auf einen Mensch-Computer-Verbund zurückzuführen sind, das schadensauslösende Ereignis ausschließlich am Handlungsanteil des Menschen festmachen und ist dann nicht mehr in der Lage, im Detail die Haftungsvoraussetzungen zu bestimmen. Und mit der Übernahme der Regeln zur Gefährdungshaftung geht sie einerseits zu weit, andererseits nicht weit genug, weil sie das digitale Risiko wie das bloße Kausalrisiko einer gefährlichen Sache behandelt. Gegenüber den Vernetzungen von Multi-Agenten-Systemen schließlich besteht allgemeine Ratlosigkeit.¹⁷

Wenn in diesen Fällen das Recht auf den Einsatz von autonomen Softwareagenten nur mit den herkömmlichen dogmatischen Mitteln reagiert und damit Verantwortungslücken aufbrechen lässt, wird der Schaden nicht von der Schädigerseite getragen, auch nicht kollektiv über die Gesellschaft verteilt, sondern es gilt ein unbarmherziges *casum sentit dominus*. Hier nun setzt eine massive Kritik ein. Die Schadensfolgen dem Geschädigten aufzubürden, wird zu Recht als rechtspolitisch verfehlt und als fundamental ungerecht kritisiert. Wenn in solchen Situationen das Versagen autonomer Softwareagenten haftungsfrei bleibt, dann setzt dies falsche Anreize für Betreiber, Produzenten und Programmierer.¹⁸ Und es sinkt die gesellschaftliche Bereitschaft, das im Prinzip vielversprechende Potential autonomer Softwareagenten voll zu nutzen. Vor allem aber widerspricht es dem Gerechtigkeitspostulat des notwendigen Zusammenhangs von Entscheidung und Verantwortung.¹⁹

4. Volle Rechtssubjektivität?

Volle Rechtsfähigkeit für Softwareagenten – dies ist die heute vieldiskutierte Antwort mancher Juristen und Politiker auf diese Defizite.²⁰ Das Europäische Parlament hat im Januar 2017 auf der Grundlage des Delvaux-Berichtes beschlossen, „einen speziellen rechtlichen Status für Roboter zu schaffen, damit zumindest für die ausgeklügeltsten autonomen Roboter ein Status als elektronische Personen mit speziellen Rechten und Verpflichtungen festgelegt werden könnte, dazu gehört auch die Wiedergutmachung sämtlicher Schäden, die sie verursachen“.

¹⁷ Ausführlich zu den Einzelheiten der dogmatischen Konstruktionen und zu Gegenvorschlägen unten III-V.

¹⁸ Eidenmüller <https://ssrn.com/abstract=2941001> 2017, 8; ähnlich MüKoBGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46.

¹⁹ So zurecht *Europäisches Parlament* (Fn. 5), 7.

²⁰ Beck (Fn. 5), 187 ff.; Beck, in: Hilgendorf/Günther (Hg.), Robotik und Gesetzgebung, 239, 255 ff.; Beck JR 2009, 225, 229 f.; Günther, Roboter und rechtliche Verantwortung, 2016, 251 ff.; Hilgendorf in: Beck (Hg.), Jenseits von Mensch und Maschine, 2012, 119, 127 f.; Zimmerman SSRN Electronic Journal <https://ssrn.com/abstract=2563965> 2015, ff.; Wettig, Vertragsschluss mittels elektronischer Agenten, 2010; Wettig/Zehendner, Proceedings of the Law and Electronic Agents Workshop 2003, 97 ff.; Matthias (Fn. 6) 83 ff.; Weitzenboeck Int'l J L & Info Tech 2001, 204 ff.; Allen/Widdison Harv. J.L. & Tech. 1996, 25, 34 ff.; Solum N. C. L. 1992, 1231 ff. Unter dem Begriff der „künstlichen Person“, die neben natürliche und juristische Personen stehen soll, vertritt Mayinger (Fn. 5), 166 ff. eine Variante dieser Vorschläge. Zur verfassungsrechtlichen Diskussion Kersten JZ 2015, 1, 7f.; Ingold Der Staat 2014, 193 ff. Strikt ablehnend dagegen Bull, Sinn und Unsinn des Datenschutzes, 2016, 120. Kritisch auch Bryson/Diamantis/Grant J. Artif Intell Law 2017, 273, 289 ff.; Lohmann ZRP 2017, 168 ff.; Schaub (Fn. 5), 345 f.; Solaiman J. Artif Intell Law 2017, 155 ff.; Spindler CR 2015, 766 ff.; Zech (Fn. 5), 202 f.; Gless/Janal JR 2016, 561, 571.

Wenn Roboter autonome Entscheidungen fällen, dann sollen sie als „elektronische Personen“, als juristische Personen im Vollsinn, anerkannt werden.²¹

In der Literatur wird ganz in diesem Sinne zur Kompensation der genannten Defizite vorgeschlagen, dass *E-persons* die Fähigkeit haben sollen, als Juristische Vollpersonen Willenserklärungen im fremden, aber auch im eigenen Namen abzugeben, dass sie Eigentum haben, über Geld verfügen, Bankkonten besitzen und Zugang zu Kredit haben sollen.²² Ja, dass sie für ihre Transaktionen Provisionen kassieren und dies selbstverdiente Geld dazu benutzen, um für Schädigungen oder Ordnungswidrigkeiten aufzukommen.²³ Im Haftungsrecht wird eine genuine Eigenhaftung der Softwareagenten gefordert. Entweder soll ihnen dafür ein Fonds eigentumsrechtlich zugeordnet werden, der durch Zahlungen der Beteiligten (Hersteller, Programmierer, Betreiber, Nutzer) aufgefüllt wird, oder eine (Pflicht-)Versicherung soll die Eigenschulden des Agenten abdecken.²⁴ Nicht nur sollen sie zu Trägern subjektiver Rechte werden, sondern sogar in Analogie zu Art. 19 III GG eigene Grundrechte geltend machen können, Rechte auf Entfaltung der Persönlichkeit, auf Nicht-Diskriminierung, auf wirtschaftliche Entfaltungsfreiheit und vor allem das Recht auf Meinungsfreiheit.²⁵

Doch muss dem deutlich widersprochen werden: Forderungen auf volle digitale Rechtssubjektivität gehen an der heutigen Realität vorbei. Wie schon bei allen eingangs geschilderten Verantwortungslücken deutlich wird, geht es bis zurzeit gar nicht um eigennütziges Handeln der Automaten, sondern stets um deren fremdnütziges Handeln für Menschen oder Organisationen, besonders Wirtschaftsunternehmen.²⁶ Ökonomisch gesprochen handelt es sich um eine Principal-Agent-Beziehung, in der der Agent zwar autonom handelt, aber vom Prinzipal abhängig ist.²⁷ Softwareagenten sind digitale Sklaven, aber Sklaven mit übermenschlichen Fähigkeiten.²⁸ Und der Sklavenaufstand muss verhindert werden. Volle Rechtssubjektivität wäre erst dann angebracht, wenn in Wirtschaft und Gesellschaft ihnen selbst Ressourcen zugeordnet würden, mit denen sie eigene Profitinteressen verfolgen. Softwareagenten treten aber – jedenfalls bisher – gar nicht als eigeninteressierte Handlungseinheiten auf, sondern stets in Interaktion mit Menschen, für deren Interessenverfolgung sie autonome Entscheidungen treffen sollen. Zudem ist gegenüber verbreiteten Vorstellungen isolierten Handelns der Computer und dessen unaufhörlicher Zunahme schon heute nach Aussage von Experten eine äußerst enge Verflechtung von digitalen und humanen Handlungen sehr viel häufiger als das Handeln der Algorithmen in Einsamkeit und Freiheit.²⁹ Und

²¹ *Europäisches Parlament* (Fn. 5), 13.

²² Z.B. *Mayinger* (Fn. 5), 227 ff., 230; *Wettig/Zehendner* (Fn. 20), 97 ff; *Koops/Hildebrandt/Jaquet-Chiffelle* Minn. J. L. Sci. & Tech. 2010, 497, 559 ff.

²³ *Mayinger* (Fn. 5), 244 f.; *Koops et al.* (Fn. 22), 554; *Karnow* John Marshall J Comput Inform Law 1994, 1, 12.

²⁴ *Hilgendorf* (Fn. 20), 127 f.; *Matthias* (Fn. 6), 244; *Sartor* (Fn. 16).

²⁵ *Mayinger* (Fn. 5), 233 ff.; *Ingold* (Fn. 20), 205 ff.; *Kersten* (Fn. 20), 2, 8; *Zimmerman* (Fn. 20), 34 ff.; *Karnow* (Fn. 23).

²⁶ So *Schirmer* JZ 2016, 660 ff., 665. Sollten in Zukunft Algorithmen eigennützig handeln, dann wird unter funktionalen Gesichtspunkten an eine Erweiterung ihrer Teilrechtsfähigkeit zu denken sein.

²⁷ *Locus classicus Jensen/Meckling* JFE 1976, 306 ff.

²⁸ Kein Wunder, dass auf die Rechtsstellung der Sklaven im römischen Recht angesichts der parallelen Situation häufig verwiesen wird, dazu *Harke*, in: *Gless/Seelmann* (Hg.), *Intelligente Agenten und das Recht*, 2016, 97 ff.

²⁹ *Karanasiou/Pinotsis* https://www.researchgate.net/publication/316789688_Towards_a_Legal

in Zukunft wird mit zunehmendem Einsatz von künstlicher Intelligenz auch Zahl und Intensität ihrer Interaktionen mit Menschen zunehmen. Der Trend geht also womöglich gar nicht in Richtung isolierter digitaler Agenten, sondern eher zunehmend in Richtung von echten Mensch-Computer-Verbänden, so dass immer weniger das Handeln der einzelnen Akteure selbst, sondern eher das Gesamthandeln solcher Hybride oder ein umfassender Vernetzungszusammenhang das Ziel juristischer Haftungsnormen sein muss.³⁰

In solchen Mensch-Maschinen-Interaktionen ausschließlich den Automaten Rechte und Pflichten zuzurechnen, wie es den Vorschlägen zur vollen Rechtssubjektivität vorschwebt, wird folglich ebensowenig ihrer eigenen Rolle gerecht wie der Rolle der beteiligten Menschen. Es unterschlägt tendenziell den Beitrag der Menschen zum gesamten Handlungszusammenhang und verfehlt deren Vermögen als Haftungspotential. So wie bisher in Wirtschaft und Gesellschaft Softwareagenten eingesetzt werden, ist also weder ihre volle Rechtsfähigkeit noch ihre Beförderung zu Juristischen Personen notwendig. Stattdessen sind eher situationsadäquate Rechtskonstruktionen gefragt. Ihr Rechtsstatus müsste, wie es besonders gründlich Gruber herausgearbeitet hat, in einer funktionalen Sicht genau auf ihre dienende Rolle in Mensch-Maschinen-Interaktionen abgestimmt sein.³¹ Autonome digitale Assistenz – für diese so präzisierte Rolle erscheint eine allgemeine Rechtsfähigkeit nicht notwendig. Vielmehr stellt sich die Frage, ob – und wenn ja wie – dafür eine *partielle Rechtssubjektivität von Softwareagenten* anerkannt werden müsste.³² *Die heute die Diskussion beherrschende Alternative – entweder sind Softwareagenten bloße Instrumente oder sie sind vollwertige Rechtssubjekte – dürfte also falsch gestellt sein.* Verfügt das Recht nicht über subtilere Konstruktionen, um den neuartigen digitalen Gefahren zu begegnen?

II. Aktanten: Autonomierisiko

Die Frage, ob und wie den Softwareagenten ein neuartiger Rechtsstatus einer strikt funktional definierten Rechtssubjektivität zuerkannt werden soll, kann nicht allein dadurch geklärt werden, dass man die eingangs geschilderten Verantwortungslücken aufdeckt. Denn diese sind nur die Symptome von tieferliegenden Risiken, die vom massiven Einsatz autonomer Softwareagenten ausgelöst werden und auf die das Recht womöglich mit einem neuartigen Rechtsstatus reagieren muss.³³ Besonders drei Verantwortungsrisiken hat die

Definition_of_Machine_Intelligence_The_Argument_for_Artificial_Personhood_in_the_Age_of_Deep_Learning, 7 f.; Jones. Vand. J. Ent. & Tech. L., 18, 2015, 77.

³⁰ Karanasiou/Pinotsi (Fn. 29), 8. Zu den Schwierigkeiten, den Vernetzungsaspekt rechtlich zu erfassen Gruber, in: Günther/Hilgendorf (Hg.), Robotik und Gesetzgebung, 2013, 123, 126, 144 f.

³¹ Gruber, in: Gruber/Bung/Ziemann (Hg.), Autonome Automaten: Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft, 2015, 191, 198; Gruber, in: Beck (Hg.), Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, 133 ff. Ebenso Kluge/Müller InTeR 2017, 24 ff. 27; Schirmer (Fn. 26), 660, 663ff.; Gindis JITE 2016, 499 ff. 17.

³² Eine präzise auf Stellvertretung und auf Gehilfenhaftung beschränkte Rechtsfähigkeit elektronischer Agenten in Bezug wurde vorgeschlagen in Teubner ZfRsoz 2006, 5 ff. Dieser Vorschlag soll im Folgenden im Detail ausgearbeitet werden, besonders unter III, 1-4. Für die beschränkte Rechtsfähigkeit von Softwareagenten plädieren auch Loos, in Schulze, Staudenmayer und Lohse (Hg.) Contracts for the Supply of Digital Content 2017, 59, 74 ff.; Gruber (Fn. 15), 227 ff.; Schirmer (Fn. 26), 663 ff.; Koops et al. (Fn. 22), 512 f., 559; implizit Keßler (Fn. 109), 592 f.; Čerkaa/Grigienėa/Sirbikytėb CLSR 2017, 685 ff.

³³ Zech (Fn. 5), 172 ff.

Digitalität aufgeworfen: (1) das Autonomierisiko, das in eigenständigen „Entscheidungen“ der Softwareagenten seinen Ursprung hat, (2) das Verbundrisiko, das auf die enge Kooperation von Mensch und Softwareagent zurückzuführen ist, und (3) das Vernetzungsrisiko, das entsteht, wenn Computer nicht isoliert agieren, sondern in enger Verflechtung mit anderen Computern. Erst eine genauere Analyse dieser Risiken kann Richtungsangaben dafür machen, welcher privatrechtliche Status Softwareagenten zuerkannt werden sollte. Im Folgenden soll eingehend das Autonomierisiko und dessen rechtliche Konsequenzen diskutiert werden, während die anderen beiden Risiken nur eher cursorisch behandelt werden können.

Besonders einschneidend ist das Autonomierisiko, das vom prinzipiell unvorhersehbaren Verhalten selbstlernender Algorithmen erzeugt wird. Es verlangt nach anderen Formen der Risikoabsorption als das schon länger bekannte bloße Automatisierungsrisiko.³⁴ Es verwirklicht sich dann, wenn Softwareagenten tatsächlich, wie in den beteiligten Wissenschaften, besonders der Informationsphilosophie, der Aktor-Netzwerktheorie und der Systemtheorie, vielfach vertreten wird, als handlungsfähige Akteure auftreten.

1. Anthropomorphisierung?

Bedeutet dies, wie oft behauptet, dass Computer kurzschlüssig mit Menschen gleichgesetzt werden? Um nicht falschen Anthropomorphisierungen aufzusitzen, muss man die digitale Handlungsfähigkeit überhaupt erst in ihrer Eigenart verstehen. Dazu ist es notwendig, die Parallele zur Handlungsfähigkeit anderer nicht-menschlicher Akteure, also zu den als juristischen Personen verfassten formalen Organisationen, zu ziehen. Um dazu das Ergebnis einer weitverzweigten Diskussion zum Substrat der juristischen Person kurz zu kennzeichnen: Man muss sich von der vertrauten Vorstellung lösen, das soziale Substrat juristischer Personen sei eine Vielheit konkreter Menschen.³⁵ Als Substrat kommen gar nicht die üblichen Verdächtigen in Betracht, weder Otto von Gierkes berüchtigte «reale Verbandspersönlichkeit», noch Emile Durkheims „Kollektivbewusstsein“, noch James Colemans «Ressourcenpool», ebensowenig Maurice Hauriou «Institutionen».³⁶ Der Kollektivakteur, wie ihn Talcott Parsons, Niklas Luhmann und andere definieren, ist nicht eine Gruppe von Individuen, sondern - eine Kette von Mitteilungen. Organisationen sind nicht Häuser oder Menschengruppen oder Ressourcenpools, sondern Entscheidungsketten. Unter der doppelten Voraussetzung, dass erstens eine solche Kommunikationskette über sich selbst kommuniziert, also eine Selbstbeschreibung herstellt, und dass zweitens gerade dieser Selbstbeschreibung kommunikative Ereignisse als von ihr selbst vollzogenen Handlungen zugeschrieben

³⁴ Zu beiden Risiken Zech (Fn. 5), 175 f.; Janal (Fn. 5), 158.

³⁵ Zum soziologischen Begriff des Kollektivakteurs Luhmann, Soziale Systeme: Grundriß einer allgemeinen Theorie, 1984, 270 ff.; Luhmann, Die Gesellschaft der Gesellschaft, 1997, 834 f.; Luhmann, Die Politik der Gesellschaft, 2000, 241. Zum Verhältnis von Kollektivakteur und Juristischer Person Teubner KritVj 1987, 61 ff. Eine detaillierte Analyse der Personifizierung nicht-menschlicher Entitäten bietet Sprenger Anthropological Theory 2017, 108 ff. 111. Zu durchaus ähnlichen Ergebnissen kommt die Diskursanalyse unter dem Stichwort der Subjektivierung, dazu Marttila Forum: Qualitative Social Research <http://www.qualitative-research.net/> 2015, 4.3. Auch die Kognitionsphilosophie behandelt die Personifizierung von *non-humans* unter ähnlichen Prämissen, Dennett, The Intentional Stance, 1987, 17.

³⁶ Zur Konzeption der „realen Verbandspersönlichkeit“ Gierke, Das Wesen der menschlichen Verbände, 1902; zum „Kollektivbewusstsein“ Durkheim, Über die Teilung der sozialen Arbeit, 2004 [1883] 121 ff.; zum „Ressourcenpool“ Coleman, Foundations of Social Theory, 1990, 325 ff.; zu „Normkomplexen“ Hauriou, Die Theorie der Institution, 1965.

werden, entsteht die soziale Realität eines Kollektivakteurs, das soziale Substrat der juristischen Person.³⁷ In ganz paralleler Weise sind Softwareagenten als Algorithmen, also als mathematisch formalisierte Informationsflüsse zu verstehen, denen in Wirtschaft und Gesellschaft unter bestimmten Bedingungen soziale Identität und Handlungsfähigkeit zugeschrieben wird.³⁸ Die Parallele wird deutlich, wenn man zwei Fehlvorstellungen der Personifizierung nicht-menschlicher Entitäten zurückweist: Falsch ist die Vorstellung, dass im Falle von Organisationen eine Menschenansammlung sich in eine organisatorische reale Verbandspersönlichkeit verwandelt. Und ebenso falsch ist die Vorstellung, dass im Falle von Softwareagenten ein Computer in einen *homo ex machina* transformiert wird. In beiden Fällen geht es exakt um das Gleiche: um die gesellschaftlich wirksame Zuschreibung von eigener Handlungsfähigkeit an - Kommunikationsprozesse.

Den prinzipiellen Unterschied zur menschlichen Handlungsfähigkeit hat Bruno Latour in seiner berühmten Aktor-Netzwerk-Theorie, in der handlungsfähige digitale Prozesse mit dem Neologismus „Aktanten“ belegt werden, herausgearbeitet und zugleich die genaueren Bedingungen nicht-menschlicher Handlungsfähigkeit analysiert.³⁹ Wie schon der Ausdruck „Aktanten“ deutlich macht, wird in Latours Analysen deutlich, dass es keinesfalls um eine Anthropomorphisierung von digitalen Prozessen geht, sondern gerade umgekehrt um eine De-Anthropomorphisierung von Softwareagenten, die – wie Informationswissenschaftler sagen – als „mindless machine(s)“⁴⁰ durch Handlungszurechnung zu nicht-menschlichen Mitgliedern der Gesellschaft geworden sind.

2. Kommunikation mit Aktanten

Doch können wir wirklich davon ausgehen, dass Algorithmen kommunizieren, wie wir das bei Menschen, aber auch bei Organisationen voraussetzen? Steckt in unserer Begegnung mit Algorithmen nur eine Wahrnehmung von Rechenergebnissen oder eine genuine Kommunikation mit gesellschaftlichen Akteuren? Auch hier wieder in aller Kürze: Die Antwort hängt davon ab, ob es in der Kommunikation gelingt, die „Beiträge“ der Softwareagenten als kommunikative Ereignisse im strengen Sinne, also als autonome Einheit von Information, Mitteilung und Verstehen zu aktivieren.⁴¹ Sofern es im kommunikativen Prozess zwischen Menschen und Computern gelingt, Ereignisse zu identifizieren, die als „Mitteilungen“ des Algorithmen, welche eine bestimmte „Information“ enthalten, zu „verstehen“ sind, wird sich in der Tat in der Begegnung von Menschen mit Algorithmen ein genuines Sozialsystem herausbilden. Die „Antworten“ in der Form von Kommunikationen, die wir von Software-Agenten auf unsere Anfragen erhalten, erfüllen alles, was die

³⁷ Dazu eingehend *Teubner* (Fn. 35).

³⁸ Zur Personifizierung von Algorithmen in informationstheoretischer Sicht *Floridi/Sanders* (Fn. 8), 187 ff.; *Thürmel*, in: *Misselhorn* (Hg.), *Collective Agency and Cooperation in Natural and Artificial Systems*, 2015, 45, 52 ff.; in öffentlichrechtlicher Sicht *Ingold* (Fn. 20), 220 ff.; in privatrechtlicher Sicht *Teubner* (Fn. 32).

³⁹ *Latour*, *Das Parlament der Dinge: Für eine politisch Ökologie*, 2001, 93 ff. Latour definiert ihre Handlungsfähigkeit sehr weit als „Widerständigkeit“. Demgegenüber wird im Text sehr viel enger auf ihre Teilnahme an gesellschaftlicher Kommunikation abgestellt.

⁴⁰ *Hildebrandt*, *Smart Technologies and the End(s) of Law*, 2015, 22; *Floridi/Sanders* (Fn. 8), 186.

⁴¹ Dazu grundlegend *Luhmann* (Fn. 35, 1984), 191 ff. In unserem Zusammenhang besonders wichtig ist die Möglichkeit der Kommunikation im strengen Sinne mit nicht-menschlichen Entitäten: *Sprenger* (Fn. 35), 116 ff.; *Luhmann* (Fn. 35, 1984), 116 ff. 119 ff.; speziell zur Kommunikation mit Algorithmen *Esposito ZfSoz* 2017, 249, 254 ff.; *Teubner* (Fn. 32).

Synthese von Information, Mitteilung und Verstehen, die Voraussetzung für Kommunikation im strengen Sinne ist, erfordert.⁴² Und dies gilt auch, wenngleich schwieriger zu begründen in der Gegenrichtung, in der Kommunikation von Mensch zu Computer.

Zwar ist dies eine „asymmetrische“ Interaktion von Menschen und Algorithmen, dennoch bildet sich zwischen ihnen ein genuines selbstproduzierendes Kommunikationssystem heraus. Sie ist asymmetrisch in einem dreifachen Sinne. (1) Den Algorithmen können keineswegs die psychischen Kompetenzen von Menschen zugebilligt werden.⁴³ Ihr Innenleben besteht aus mathematischen Operationen auf der Grundlage elektronischer Signale. Aber dieser Unterschied der Innerlichkeit von Mensch und Computer ist für unsere Frage gar nicht relevant, sofern nur außerhalb der unterschiedlichen Innenleben von Mensch und Computer ein Kommunikationsgeschehen zustande kommt und sofern zugleich das elektronische Innenleben der Algorithmen und das Bewusstsein der Menschen die Kommunikation zwischen ihnen ausreichend irritieren, sodass die Synthese aus Mitteilung, Information, Verstehen gelingen kann. (2) Sie ist in einem anderen Sinne asymmetrisch. In der Kommunikation unter Menschen ist die doppelte Kontingenz symmetrisch beidseitig, weil beide Partner die Wahl ihres Verhaltens von der Verhaltenswahl des anderen abhängig machen.⁴⁴ Dagegen wird zwischen Mensch und Maschine die für Kommunikation notwendige doppelte Kontingenz nur einseitig erlebt, nämlich nur vom Menschen und nicht von der Maschine (jedenfalls im heutigen Entwicklungsstand).⁴⁵ Einseitig erlebte doppelte Kontingenz, wie wir sie in der Mensch-Maschinen-Beziehung antreffen, aber schließt nicht aus, dass dennoch genuine Kommunikation stattfindet. Historisch bekannte Konfigurationen, wie die Kommunikation mit Gott im Gebet ebenso wie animistische Praktiken in traditionellen Gesellschaften, leisten durchaus die Synthese aus Mitteilung, Information und Verstehen.⁴⁶ Aber sie tun dies nur unter der Voraussetzung, dass im Ereignisstrom eine Personifikation des nicht-menschlichen Partners stattfindet, welche eine Zurechnung der Handlung auf den „anderen“ ermöglicht. „Personalität entsteht, wo immer das Verhalten anderer als gewählt vorgestellt wird und durch eigenes Verhalten kommunikativ zu beeinflussen.“⁴⁷ Die Personifizierung nicht-menschlicher Entitäten als ein performatives Ereignis, das die Person als ein semantisches Konstrukt konstituiert, kompensiert also diese Asymmetrie in der Mensch-Maschinenbeziehung. (3) Sie ist schließlich asymmetrisch in Bezug auf das wechselseitige Verstehen in der Kommunikation von Mensch und Maschine. Wenn Verstehen verstanden wird als die Fähigkeit, die Selbstreferenz eines anderen in der eigenen Selbstreferenz nachzukonstruieren, dann können Menschen die Botschaften des Algorithmus verstehen, während dem Algorithmus die Fähigkeit des Verstehens des menschlichen Innenlebens fehlen dürfte. In unserem Zusammenhang kann man jedoch diese Frage offenlassen, weil es in der Kommunikation auf dieses wechselseitige Tiefenverstehen im Innenleben von Mensch und Maschine gar nicht

⁴² Zur Theoriefrage, ob Arbeit mit Computern als Kommunikation zu verstehen ist, auch wenn doppelte Kontingenz womöglich nur einseitig erfahren wird, *Luhmann* (Fn. 35, 1997), 303 f. *Esposito* (Fn. 42), 262 stellt die Frage, ob der Kommunikationsbegriff erweitert oder ein anderer Begriff für die Begegnung Computer-Mensch entwickelt werden muss, lässt die Antwort dann aber letztlich offen.

⁴³ *Esposito* (Fn. 42), 250.

⁴⁴ Zu doppelter Kontingenz *Parsons/Shils*, *Toward a General Theory of Action: Theoretical Foundations for the Social Sciences*, 1951; *Luhmann* (Fn. 35, 1984), 148 ff.

⁴⁵ Dazu *Luhmann* (Fn. 35, 1997), 304.

⁴⁶ *Sprenger* (Fn. 35), 119 ff.

⁴⁷ *Luhmann* (Fn. 35, 1984), 643.

ankommt. Man muss deutlich zwischen Verstehensprozessen innerhalb des autonomen Kommunikationsgeschehens einerseits und Verstehensprozessen innerhalb der interagierenden Entitäten andererseits unterscheiden.⁴⁸ Für Verstehen innerhalb des verselbständigten Kommunikationsprozesses ist nicht entscheidend, ob der Algorithmus selbst versteht, vielmehr kommt es nur darauf an ob der „Antworttext“ des Algorithmus den „Fragertext“ des Menschen versteht. Sofern der Kommunikationsakt des Algorithmus die Differenz von Mitteilung und Information im Kommunikationsakt des Menschen nachvollzieht und darauf mit einer eigenen Differenz von Mitteilung und Information reagiert, ist ein kommunikativer Verstehensakt vollzogen, ohne dass es – um es nochmals zu betonen – darauf ankäme, ob im Innenleben des Algorithmus die Äußerung des Menschen verstanden wird oder nicht. Auch hier bieten sowohl die Kommunikation mit Gott als auch animistische Praktiken historische Belege für Verstehen im kommunikativen Geschehen, das auch dann zustande kommt, wenn der „andere“ in seinem Inneren nicht verstehen sollte.

Um das bisher Gesagte in einer Kurzformel zu verdichten: *Softwareagenten sind – ebenso wie Unternehmen und andere formale Organisationen - nichts anderes als bloße Informationsströme, die dann zu „Personen“ (oder Teilpersonen) werden, wenn sie im Kommunikationsprozess eine soziale Identität aufbauen und wenn ihnen zusammen mit den notwendigen organisatorischen Vorkehrungen, z.B. Vertretungsregeln, eigene Handlungsfähigkeit effektiv zugerechnet wird.*

3. Gradualisierte digitale Autonomie

Ob nun ein Softwareagent, also ein konkreter digitaler Informationsstrom, darüber hinaus als *autonomer* Akteur zu qualifizieren ist, ist dann die juristisch entscheidende Frage. Dies hängt nicht nur von der ihm zugeschriebenen sozialen Handlungsfähigkeit ab, sondern zusätzlich davon, mit welchen besonderen Qualitäten er in unterschiedlichen Kontexten als eigenständige Person ausgestattet wird. Jeder Sozialkontext schafft sich seine jeweils eigenen Kriterien der Personalität, die Wirtschaft nicht anders als die Politik, die Wissenschaft, die Moralphilosophie - oder das Recht.⁴⁹ Jedes Subsystem schreibt auf eine andere Weise Individuen, Kollektiven oder Algorithmen als seinen „Personen“ Handlungen, Entscheidungen, Vermögen, Verantwortlichkeiten, Rechte und Pflichten zu oder stattet sie mit Kapital, Interessen, Intentionen, Handlungszielen oder Präferenzen aus. Je nach gesellschaftlichem Kontext stößt man auf deutliche Variationen der effektiv geltenden Akteurskonzepte, wie an den unterschiedlichen Definitionen des *homo oeconomicus*, *homo juridicus*, *politicus* etc. deutlich wird. Und auch bei Softwareagenten werden die sozialen Kompetenzen vom jeweiligen Sozialbereich festgelegt, weil er die soziale Identität des Algorithmus konstituiert, wie auch dessen Fähigkeit zu handeln, zu kommunizieren, zu entscheiden, festlegt. Nur ein Beispiel für viele: Soziale Bewegungen sind in der Politik als eigenständige Kollektivakteure anerkannt, in der Wirtschaft und im Recht gelten sie als Unpersonen.⁵⁰

⁴⁸ Luhmann, in: Luhmann/Schorr (Hg.), Zwischen Intransparenz und Verstehen 1986, 72, 93 ff. Luhmann, in: The Institute of Comparative Law in Japan (Hg.), Conflict and Integration 1989, 493 ff.

⁴⁹ Dazu Sprenger (Fn. 35), 118.

⁵⁰ Dazu Tratschin, Protest und Selbstbeschreibung 2016, 79, 221.

In der interdisziplinären Diskussion werden nun ganz unterschiedliche Kriterien angeboten, ab wann einem Softwareagenten Autonomie zuzusprechen ist.⁵¹ In vielen Disziplinen wird die Frage, ob Softwareagenten autonom handeln können, positiv beantwortet. Der Schwellenwert aber, von dem ab Autonomie zugeschrieben werden kann, ist kontrovers. Digitale Autonomie scheint ein gradualisiertes Phänomen zu sein.⁵² Und die Gradualisierung findet nicht nur auf einer einzigen Skala mit unterschiedlichen Autonomiegraden statt, sondern in einem mehrdimensionalen Raum, der mehr oder weniger Autonomie zulässt.⁵³

So legt der zurzeit führende Oxforder Informationsphilosoph Floridi in drei Dimensionen Schwellenwerte für die Zurechnung von Handlungsfähigkeit an nicht-menschliche Entitäten, sowohl an Organisationen als auch an Algorithmen, fest: (1) Interaktion (mit Mitarbeitern und anderen Organisationen), (2) Fähigkeit, aus sich selbst Zustandsänderungen herbeizuführen und (3) Anpassung der Strategien für Entscheidungen. Andere wiederum stellen auf ganz heterogene Eigenschaften ab: auf Denkfähigkeit, Kommunikationsfähigkeit, Verstehen, rationales Handeln, Nicht-Prognostizierbarkeit ihrer Konditionalprogramme⁵⁴, autonome Raumveränderung ohne menschliche Aufsicht⁵⁵, geringe Strukturiertheit des Einsatzbereichs,⁵⁶ eigene Zielverfolgung und Mittelwahl,⁵⁷ Optimierung mehrerer Ziele,⁵⁸ Kontrollfähigkeit, Programmierfähigkeit, Lernfähigkeit,⁵⁹ Selbständerung, Selbstbewusstsein,⁶⁰ Künstliche Intelligenz, moralische Selbstgesetzgebung oder gar Leidensfähigkeit oder schließlich auf ein digitales Gewissen ab.⁶¹

Die Unterschiede gehen nicht notwendig auf wissenschaftliche Kontroversen zurück, die zugunsten der einzig richtigen Lösung zu entscheiden wären, sondern erklären sich meist aus dem je verschiedenen Erkenntnisinteresse der beteiligten Disziplinen sowie aus praktischen Handlungsorientierungen in verschiedenen gesellschaftlichen Bereichen. Während die an Erklärung und Voraussage interessierten Kausalwissenschaften nur dann von Autonomie sprechen, wenn sie eine *black box* modellieren, innerhalb derer sie keine Kausalzusammenhänge mehr analysieren können, wohl aber ihr Außenverhalten beobachten, bauen die sinnverstehenden Sozialwissenschaften konstitutiv auf der Autonomie der Akteure auf, aber auch hier wieder mit deutlichen Unterschieden. So stellt die Ökonomie nutzenorientierte Entscheidungen heraus, definiert also Autonomie als rational choice, während Moral und Ethik Autonomie eher in Formen eines digitalen Selbstbewusstseins suchen.

4. Rechtliche Autonomiekriterien

⁵¹ Einen Überblick bietet Zech (Fn. 5), 170 f., Fn. 16.

⁵² So Hildebrandt (Fn. 40), 21.

⁵³ So besonders die mehrdimensionalen Kriterienkataloge von Lohmann AJP 2017, 152, 155; Zech (Fn. 5), 176; Misselhorn, in: Misselhorn (Hg.), *Collective Agency and Cooperation in Natural and Artificial Systems* 2015, 3, 5 ff.; Thürmel (Fn. 38), 53 ff.; Floridi/Sanders (Fn. 8), 192 f.

⁵⁴ Zech (Fn. 5), 171 f.; Bräutigam/Klindt (Fn. 3), 1137.

⁵⁵ Matthias (Fn. 6), 35.

⁵⁶ Lohmann (Fn. 53), 154.

⁵⁷ Misselhorn (Fn. 53), 6 f.

⁵⁸ Karanasiou/Pinotsi (Fn. 29), 1.

⁵⁹ Matthias (Fn. 6), 17 ff.

⁶⁰ Zimmerman (Fn. 20), 34 ff.

⁶¹ Eine informative Diskussion dieser unterschiedlichen Kriterien bei Misselhorn (Fn. 53), 4 ff.

Das Rechtssystem wiederum muss aus eigenem disziplinären Erkenntnisinteresse und aus eigener Handlungsorientierung die rechtlich relevante Grenze zwischen instrumentalisierten und autonomen Handeln selbst festlegen. Dabei wird es sich aber an der interdisziplinären Diskussion in den Informationswissenschaften orientieren und am Ende ein Autonomiekriterium wählen müssen, das vor dem interdisziplinären Diskussionsstand vertretbar ist.⁶² Ähnlich wie das Recht auch sonst, etwa in Umweltfragen, vorgeht, wenn es angesichts von wissenschaftlichen Befunden einer Gradualisierung selbstverantwortlich Grenzwerte festlegt, so muss es hier aufgrund rechtseigener Kriterien festlegen, ab welchem Grad von wissenschaftlich erforschter Eigenständigkeit digitaler Phänomene im Rechtssinne von deren Autonomie auszugehen ist.⁶³

Denkfähigkeit oder künstliche Intelligenz wird in der Debatte immer wieder als das eigentliche Kriterium vorgeschlagen, das über Rechtssubjektivität entscheidet.⁶⁴ Hier nun ist eine verbreitete Fehlvorstellung zu korrigieren. Für ihre Handlungsfähigkeit geht es gar nicht um die Frage: Welche Art von ontologischen Eigenschaften – Intelligenz, Geist, Seele, reflexive Kapazitäten, Einfühlungsvermögen – muss ein Softwareagent besitzen, um als Akteur gelten zu können?⁶⁵ Auch hier wieder ist das Paradigma von formalen Organisationen als juristischen Personen hilfreich: für die rechtliche Handlungsfähigkeit nicht-menschlicher Agenten sind innere psychische Zustände nicht ausschlaggebend.⁶⁶ „After all, what is interesting in the interaction with algorithms is not what happens in the machine’s artificial brain, but what the machine tells its users and the consequences of this.“⁶⁷ Nicht die inneren Eigenschaften der Agenten, sondern die gesellschaftlichen Interaktionen, an denen die laufenden Operationen des Algorithmus teilnehmen, d.h. eine geschlossene und autonome Abfolge rekursiver Kommunikationen, konstituieren den Algorithmus als Person, als semantisches Artefakt, als kommunikationsfähigen Akteur.⁶⁸ Die Operationen des Algorithmus, also digitale Datenströme, existieren – wie oben schon ausgeführt – als Akteure nicht per se, sondern sie werden erst von Sozialsystemen als adressierfähige Personen konstituiert. Gesellschaftliche Teilsysteme konstruieren sie als semantische Artefakte, indem sie ihnen volle oder eingeschränkte Subjektivität zuschreiben. Zwar basiert dies auf der gesellschaftlichen Unterstellung, dass die kommunizierende Einheit kommunikative Fähigkeiten hat, aber der fiktionale Charakter ist kein Schaden, solange es nur gelingt, mit Hilfe ihrer Beiträge den Kommunikationsfluss mit Erfolg aufrechtzuerhalten. Für den kommunikativen Umgang mit Personen ist ein Name erforderlich, nicht aber eine Entschlüsselung innerer Prozesse ‚in‘ der Person. Das gilt für Organisationen ebenso wie für Algorithmen. Also: Nicht auf die innere Denkfähigkeit der Algorithmen kommt es an, nicht auf „wirkliche“ künstliche

⁶² Allgemein zur autonomen und zugleich abhängigen Rolle des Rechts in interdisziplinären Zusammenhängen *Teubner*, in: *Grundmann/Thiessen* (Hg.), *Recht und Sozialtheorie im Rechtsvergleich: Interdisziplinäres Denken in Rechtswissenschaft und -praxis*, 2015, 145 ff.

⁶³ *Matthias* (Fn. 6), 43 ff.

⁶⁴ *Spindler* (Fn. 4, 2016), 816.

⁶⁵ So entschieden gegen einen Trend in den Ingenieurwissenschaften, die sich auf die „inneren“ Prozesse der Algorithmen konzentrieren und die konstitutiven sozialen Interaktionen vernachlässigen, *Esposito* (Fn. 42), 250; ebenso *Latour* (Fn. 39), 93 ff.; *Teubner* (Fn. 32).

⁶⁶ *Luhmann*, *Organisation und Entscheidung*, 2000, Kap. 13 IV.

⁶⁷ *Esposito* (Fn. 42), 250.

⁶⁸ Für die Konstitution als Person im allgemeinen *Dennett* (Fn. 35), 17; aus systemtheoretischer Sicht: *Luhmann*, *Nachweise* in (Fn. 35) (nur Kollektivakteur) und *Sprenger* (Fn. 35), 114. Speziell für Softwareagenten *Gruber* (Fn. 15), 250 ff.; *Matthias* (Fn. 6), 83 ff.; *Teubner* (Fn. 32); *Solum* (Fn. 20).

Intelligenz, was immer das bedeutet, sondern auf ihre spezielle Beteiligung an gesellschaftlicher Kommunikation. „Artificial communication“ und nicht „artificial intelligence“ ist entscheidend.⁶⁹ Diese Entpsychologisierung, wie sie die Kommunikationstheorie vorschlägt, trifft sich recht genau mit den bekannten Objektivierungstendenzen im Recht der Willenserklärung und im Begriff der Fahrlässigkeit.⁷⁰ Darauf ist zurückzukommen.

Intentionales Handeln dagegen dürfte eine notwendige Voraussetzung für Autonomie sein, sofern darunter kein innerer psychischer Zustand verstanden wird, sondern die Zuschreibung zielgerichteten Handelns durch einen Beobachter gemeint ist – die berühmte „intentional stance“ des Kognitionswissenschaftlers Daniel Dennett: Ob die Agenten in Wahrheit Willensfreiheit besitzen, ist keine wissenschaftlich sinnvolle Frage. Stattdessen: Wenn wegen erhöhter Komplexität eine physikalische Beschreibung nicht möglich ist, kann die Wissenschaft mit Hilfe eines intentionalen Vokabulars das untersuchte System als einen Akteur, der mit Annahmen über die Welt, mit Zielen und Handlungsoptionen agiert, analysieren und dadurch neue Erkenntnisse gewinnen.⁷¹ Die Systemtheorie weitet dies von der Wissenschaft auf andere Beobachter aus.⁷² Nicht nur die Wissenschaft kann Softwareagenten intentional - zum Beispiel in einer Vertragsbeziehung – beschreiben, auch der Partner in einer Interaktion (Ego) kann dieser Beobachter sein, der das Verhalten von Alter nicht mehr kausalistisch, sondern intentionalistisch interpretiert und der dadurch neue Orientierungen für seine eigenen Handlungen findet. Ebenso kann ein ganzes Sozialsystem – in unserem Fall das Recht - dieser Beobachter sein, der Softwareagenten Intentionen zurechnet und daraus Konsequenzen für rechtliche Verbindlichkeit von dessen Erklärungen und Verantwortung für dessen Handlungen zieht.

Jedoch für eine rechtlich relevante Autonomie dürfte die bloße Intentionalität, also Zielgerichtetheit und Mittelwahl eines Agenten, noch nicht ausreichen. Das gleiche gilt für dessen Teilnahme an Kommunikation. Denn auch bloß automatisierte Softwareagenten nehmen schon an Kommunikation teil. Ebenso wie Intentionalität ist die Teilnahme an Kommunikation nur eine notwendige, aber nicht hinreichende Bedingung.

Andere Kriterien wiederum dürften über die juristischen Minimalvoraussetzungen für Autonomie weit hinausgehen. Rationales Handeln etwa, wie es Dennett über die „*intentional stance*“ hinaus als Autonomievoraussetzung nicht-menschlicher Agenten verlangt, mag für *rational choice* von Wirtschaftsakteuren plausibel sein, nicht aber für Rechtsakteure, deren irrationales Handeln bei Rechtsverletzungen gerade im Vordergrund steht.

Ebenso dürften auch andere anspruchsvolle Aktivitäten wie Selbständerung oder selbsttätiges Lernen schon die Minimalvoraussetzungen überschreiten. Gewiss steigern sie noch einmal erheblich die Handlungsautonomie der Agenten und ermöglichen es, Sanktionen und Anreize direkt auf die Agenten auszurichten.⁷³ Aber

⁶⁹ *Esposito* (Fn. 42), 250.

⁷⁰ Dazu *MüKoBGB/Grundmann*, 7. Aufl. 2016, § 276 Rn. 53 ff.; *MüKo/Wagner*, 7. Aufl. 2016, § 823, Rn. 28 ff.

⁷¹ *Dennett* (Fn. 35), 17; *Matthias* (Fn. 6), 41 ff.

⁷² *Luhmann* (Fn. 35, 1984), 155 ff.

⁷³ *Floridi/Sanders* (Fn. 8), 192 ff.

es dürfte unter Gesichtspunkten des Schadensausgleichs nicht angemessen sein, erst dann mit rechtlichen Verantwortungsmechanismen einzusetzen, wenn die Programmkorrekturen nur von ihnen selbst, nicht aber von Programmierern vorgenommen werden.

Schon gar nicht darf, wie schon gesagt, für eine rechtlich relevante Digitalautonomie eine vollausgebildete Künstliche Intelligenz, Empathie, Gefühle, Leidensfähigkeit, Selbstbewusstsein oder gar ein digitales Gewissen verlangt werden. Und erst recht sollte ein haftungsrechtlich relevanter Autonomiebegriff für digitale Phänomene nicht versuchen, Anleihen bei der philosophischen Tradition zu machen, wonach Autonomie als die Möglichkeit des Menschen verstanden wird, sich selbst als freiheits- und vernunftfähiges Wesen zu bestimmen und entsprechend aus Freiheit moralisch zu handeln.⁷⁴ Es ist unhaltbar zu behaupten, dass erst wenn digitale Autonomie zu Selbstbewusstsein erstarkt ist, dann auch Rechtspersönlichkeit indiziert sei. Gewiss sind dies Fragen, die sich der Informationsphilosophie in ihrer Suche nach einer möglichen digitalen Moralität stellen. Aber wenn solche Eigenschaften im Haftungsrecht als Autonomiekriterium verlangt werden, dann leistet dies nur opportunistischem Verhalten der Juristen Vorschub, die es für die Gegenwart getrost bei der traditionellen Dogmatik belassen können, sich aber zugleich ein Hintertürchen offenhalten, für den Fall, dass sie sich mit ihrer exklusiven Menschzurechnung einmal vollständig gesellschaftlich isolieren sollten.⁷⁵

5. Autonome Entscheidungen der Softwareagenten

Entscheidung unter Ungewissheit – dies dürfte das rechtsrelevante Autonomiekriterium sein. Wenn diese Aufgabe tatsächlich an Softwareagenten delegiert wird und sie sich entsprechend verhalten, dann ist das Recht genötigt, ihnen auch Handlungsfähigkeit zuzusprechen. Softwareagenten handeln im Rechtssinne dann autonom, wenn ihr Verhalten nicht mehr ausschließlich einem Reiz-Reaktion-Schema folgt, sondern wenn sie eigene Ziele besitzen, abändern, verfolgen und nicht-prognostizierbare Entscheidungen treffen.⁷⁶

Konkret bedeutet dies: Wenn (1) ein Softwareagent so programmiert ist, dass er zwischen Alternativen zu entscheiden hat, wenn (2) er diese Entscheidung als Optimierung verschiedener Kriterien treffen muss und wenn (3) ein Programmierer das Verhalten des Softwareagenten weder nachträglich erklären noch für die Zukunft voraussagen kann, sondern nur noch ex post korrigieren kann, dann sollte das Recht von Autonomie, d.h. von Entscheidungsfähigkeit und Verantwortungsfähigkeit des Softwareagenten ausgehen. Praktisch impliziert dies die Verpflichtung für den Hersteller, eine *black box* zu installieren, deren Protokollierungsfunktion den Entscheidungsablauf nachvollziehen lässt.⁷⁷ Jedoch - wie stets im Zwischenbereich zwischen technisch-wissenschaftlicher Expertise und Recht - bedeutet dies keine automatische Bindung an die Expertise, sondern es bedarf einer eigenverantwortlichen Rechtsentscheidung. Vergleichbar der Beziehung zwischen

⁷⁴ Kant, Grundlegung zur Metaphysik der Sitten, 2012 (1785), 60.

⁷⁵ Dann kann man schnell die dogmatischen Begriffe ändern und zugleich beteuern, man habe es schon seit langem angekündigt. Zum anderen kann mit dieser Formel die Eigenverantwortlichkeit der Softwareagenten ad infinitum hinausgezögert werden, da dem Menschen vergleichbare Intelligenz und Willensformung auch in Zukunft stets verweigert werden können.

⁷⁶ In dieser Richtung auch *Europäisches Parlament* (Fn. 5), 6; *Mayinger* (Fn. 5), 14; *Kirn/Müller-Hengstenberg* (Fn. 11), 4.

⁷⁷ *Lohmann* (Fn. 53), 158; *Günther* (Fn. 20), 99.

Sachverständigen und Richter bei Fragen der menschlichen Zurechnungsfähigkeit geht es darum, den Umschlagspunkt von Kausalzurechnung zu Entscheidungszurechnung im Einzelnen zu bestimmen. Das Recht zieht bekanntlich in diesem Umschlag noch zusätzliche rechtspolitische und rechtsdogmatische Gesichtspunkte heran.

Warum ist Entscheidung unter Ungewissheit das rechtlich relevante Kriterium? Ungewissheit resultiert aus der Indeterminiertheit der Programmierung und einer geringen Strukturiertheit der Umwelt, mit welcher der Algorithmus konfrontiert wird.⁷⁸ Der Grund für die rechtliche Relevanz ist der fundamentale Zusammenhang von Entscheidung und Verantwortung.⁷⁹ Zwischen der Öffnung von Entscheidungsalternativen in einer ungewissen Umwelt, die an nicht-menschliche Prozessoren delegiert werden, und der dabei entstehenden Verantwortungsproblematik besteht ein unauflöslicher Zusammenhang. Verantwortung im strengen Sinne ist das Einstehenmüssen für Entscheidung unter Ungewissheit, deren Resultat nicht prognostizierbar ist.⁸⁰ Es geht nicht bloß um das Einstehenmüssen für Fehler!⁸¹ Wenn Softwareagenten bei einer voll determinierten Kalkulation Fehler machen, dann bedarf es einer bloßen Fehlerkorrektur. Anders aber bei nicht-determinierten Entscheidungen unter Ungewissheit. Wenn ein solch unkalkulierbares Risiko eingegangen wird, dann kann eine Fehlentscheidung im Vorhinein gar nicht vermieden werden, sondern, wenn sie trotz aller Vorkehrungen eintritt, nur noch bereut werden.⁸² Dies nachträgliche Bereuen von Ungewissheitsentscheidungen aber ist ein klarer Fall von rechtlich gebotener Verantwortung, unter anderem auch von zivilrechtlicher Haftung.

Sollte man dann überhaupt das Risiko eingehen, dass Algorithmen nicht-prognostizierbare Entscheidungen fällen dürfen? Zech vertritt den strengen Standpunkt, dass schon nach geltendem Recht der Einsatz autonomer Algorithmen prinzipiell rechtswidrig ist. Nur der Gesetzgeber könne eine Ausnahme anordnen und dies auch nur dann, wenn er zugleich eine effektive Risikovorsorge trifft.⁸³ Ein solches Verbot mag als eine Extremlösung erscheinen, trifft aber das Problem genau: Das Risiko einer echten Entscheidungsdelegation an nicht-menschliche Akteure ist weder abschätzbar noch beherrschbar.⁸⁴

Ungewissheitsentscheidungen mit Eigenrisiko sind gesellschaftlich sehr viel relevanter als reine Rechenaufgaben mit bloßem Fehlerrisiko. „Der Einsatz vollautonomer beweglicher Maschinen im öffentlichen Raum dürfte am oberen Ende der Risikoskala stehen.“⁸⁵ Auch der häufig angeführte Vorteil, die massive Senkung von Transaktionskosten, kann dies hohe Risiko einzugehen, nicht aufwiegen. Die

⁷⁸ Lohmann (Fn. 53), 154.

⁷⁹ In soziologischer Sicht Luhmann, *Soziologie des Risikos*, 1991, 197 ff.; aus der Sicht des Haftungsrechts Jansen, *Die Struktur des Haftungsrechts*, 2003, 136 ff.

⁸⁰ Zum Zusammenhang von Entscheidungen unter Ungewissheit und Verantwortung Luhmann, *Funktionen und Folgen formaler Organisation*, 1964 172 ff.; Luhmann, *Ökologische Kommunikation: Kann die moderne Gesellschaft sich auf ökologische Gefährdungen einstellen?* 1986, 26.

⁸¹ „Only those questions that are in principle undecidable we can decide“. Zu diesem Unterschied von Entscheidung und Errechnung Foerster, *Cybernetics and Human Knowing* 1992, 9 ff.

⁸² Zum „postdecisional regret“ Luhmann (Fn. 66), 146, 170; Luhmann (Fn. 80), 32 f., 207 ff.

⁸³ Zech (Fn. 5), 191 ff.

⁸⁴ Matthias (Fn. 6), 33 ff.

⁸⁵ Zech (Fn. 5), 176 setzt die Autonomieschwelle relativ hoch an, erst bei Selbständerung und Lernfähigkeit und mangelnde Vorausssehbarkeit.

eigentliche Rechtfertigung liegt im „Entdeckungsverfahren“ des Einsatzes autonomer Algorithmen, in ihrem enormen Innovationspotential. Wenn Computer Entscheidungen unter Ungewissheit treffen, dann ermöglicht dies, etwas gänzlich Neues zu entdecken, etwas, das sich noch keine menschliche Intelligenz ausgedacht hat, und sogar manchmal, das keine menschliche Intelligenz nachvollziehen könnte, herauszufinden, und nicht einfach Errechenbares zu kalkulieren.⁸⁶ Hier steckt das eigentliche Motiv, warum die Gesellschaft es zulässt, dass Menschen Entscheidungen unter Ungewissheit an Algorithmen delegieren. Damit nehmen sie das Risiko katastrophaler Fehlentscheidungen bewusst in Kauf, das Risiko, dass das Entdeckungsverfahren gesellschaftlich höchst unerwünschte Folgen zeitigt. Hier also liegt der tiefere Grund für gesteigerte Anforderung an Verantwortung für Entscheidungen im Unterschied der Verantwortung für bloße Rechenfehler. Ja, Verantwortung gewinnt ihre eigentliche Bedeutung erst hier: als Einstehenmüssen für den „Sprung ins Dunkle“.⁸⁷ Diesen Sprung ins Dunkle als Entscheidung unter Ungewissheit nicht nur Menschen anzuvertrauen, sondern Algorithmen zu überlassen, darin besteht das fundamental Neue. Wenn das Recht das Entdeckungsverfahren autonomer Algorithmen zulässt, wenn es erlaubt, dass Softwareagenten autonome Entscheidungen treffen dürfen, ist es unabdingbar, dass gerade dafür das Recht wirksame Verantwortungsformen für den Enttäuschungsfall bereitstellt. Und dies gilt gerade auch für den Fall, dass den beteiligten Menschen gar keine Pflichtverletzungen unterlaufen sind.

Digitale Ungewissheitsentscheidungen eröffnen ein gänzlich neues gesellschaftliches Labor für Experimente. Nur noch experimentell kann das Neue ausprobiert, nicht mehr vorherberechnet, sondern nur noch nachträglich auf seine Folgen überprüft werden.⁸⁸ Evolutionstheoretisch ausgedrückt: Digitale Ungewissheitsentscheidungen erzeugen vielfältige Variationen, auf die Menschen gar nicht gekommen wären. Verantwortungszuschreibung aber ermöglicht erst ihre gesellschaftlich vertretbare Selektion, die schadensstiftende Variationen wieder ausscheidet. Und in der auf Dauer gestellten Normativierung des richtigen Entscheidens liegt die Retention. Darf man - so muss man die Autoren fragen, welche die eingangs geschilderten Verantwortungslücken hinzunehmen bereit sind - diese neuartigen Risiken den Geschädigten als „Lebensrisiko“ aufbürden, wenn Softwareagenten Ungewissheitsentscheidungen treffen dürfen? Und dies auch noch mit der „humanistischen“ Begründung, dass nur Menschen, nicht Computer im Rechtssinne handeln können?⁸⁹

III. Rechtsfragen des Autonomierisikos

⁸⁶ *Esposito* (Fn. 42), 253.

⁸⁷ Trotz aller Rationalisierung von Entscheidungen bleibt es beim „Mysterium“ der Entscheidung, *Luhmann VerwArch* 1993, 287, 288.

⁸⁸ *Matthias* (Fn. 6), 33 ff.

⁸⁹ *Kirn/Müller-Hengstenberg*, *Rechtliche Risiken autonomer und vernetzter Systeme* 2016, 3.3.7 bei Fn. 799 mwN., sind für diese widersprüchliche Argumentation typisch. Zuerst analysieren sie in aller Härte die zahlreichen Fälle effektiver Delegation von Entscheidungen an Algorithmen, die zum Kontrollverlust der Menschen führen, dann aber verweisen sie auf die heutige Rechtslage, die keine digitalen Rechtssubjekte anerkennt, um dann resigniert bloß noch der Frage nachzugehen, ob nicht die Klärung der Pflichtverletzungen von Menschen wenigstens einige Probleme lösen kann. Die verbleibenden Verantwortungslücken nehmen sie fatalistisch hin. Wie stattdessen eine Rehumanisierung der technischen Welt – auch mit Mitteln des Rechts – eingeleitet werden könnte, dazu eingehend *Gruber* (Fn. 15), 324 ff.

Als Zwischenergebnis lässt sich festhalten. Das Autonomierisiko stellt das Privatrecht vor neue Herausforderungen, freilich nicht in dem Sinne, dass es die volle Personifizierung von Softwareagenten nahelegt. Vielmehr sollte den Softwareagenten ein sorgfältig kalibrierter Rechtsstatus zuerkannt werden. *Die Antwort auf das Autonomierisiko wäre ihr Status als Aktanten, als teilrechtsfähige Akteure, deren autonome Entscheidungen mit Rechtsverbindlichkeit ausgestattet werden und Haftungsfolgen auslösen können.* Dieser neue Rechtsstatus soll nun anhand ausgewählter Einzelproblemen ausgeformt werden. Dabei steht im Vordergrund, ob und wie die eingangs geschilderten Verantwortungsdefizite bewältigt werden können. Aber es wird zugleich darum gehen, wie ein solcher Rechtsstatus von Softwareagenten ermöglicht, die Dogmatik der Rechtsgeschäftslehre und des vertraglichen und außervertraglichen Haftungsrechts in sich stimmig und frei von Fiktionen weiterzuentwickeln.

1. Digitale Verträge

Für die Geschäftspraxis bedeutet es buchstäblich eine revolutionäre Umwälzung, dass Menschen an Algorithmen die verantwortungsvolle Aufgabe delegieren, selbständig Verträge abzuschließen und durchzuführen. Zugleich aber ist die Rechtsgeschäftslehre in ihren dogmatischen Fundamenten berührt, denn deren selbstverständliche Voraussetzung war bisher, dass ausschließlich Menschen – und das gilt auch für die Rechtsakte juristischer Personen - Willenserklärungen abgeben, besonders Verträge abschließen können. Deshalb war es durchaus konsequent, wenn einige Autoren darauf insistierten, dass autonome Computererklärungen nach geltendem Recht unwirksam sind, solange bis die Gesetzgebung sie zulässt und eine eingehende Regelung erlässt, die Verbindlichkeits- und Haftungsfragen regelt.⁹⁰ Genau dies tat die Gesetzgebung in USA und Kanada. Section 14 des Uniform Electronic Transactions Act der USA erklärt Verträge zwischen Algorithmen für wirksam:

„A contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents' actions or the resulting terms and agreements.“⁹¹

Auch das Europäische Parlament erklärt die überkommenen Regeln des Vertragsrechts für inadäquat und spricht von der „Notwendigkeit für neue aktuellere Regeln“.⁹² Das deutsche Recht hingegen sieht hier keinerlei fundamentale Probleme. Es begnügt sich damit, für Computererklärungen in § 312 ein paar Verbraucherschützende Formvorschriften vorzusehen. Als Grund wird angegeben, die Rechtsgeschäftslehre verfüge über einen solchen inneren Reichtum, dass auch Computererklärungen „mit tradierten rechtlichen Instrumenten bewältigt werden“ können.⁹³ Gegen den Protest einzelner Autoren erklären die großen BGB-Kommentare: Die automatisierte Willenserklärung „ist“ (!) die Erklärung dessen, der die EDV-Anlage betreibt, da die Anlage keinen eigenen Willen betätigt, sondern aufgrund ihrer Programmierung (zumindest mittelbar) den Willen des

⁹⁰ Für Unwirksamkeit: *Wettig* (Fn. 20), 162 f.; *Gitter*, Softwareagenten im elektronischen Rechtsverkehr, 2007, 173; *John*, Haftung für künstliche Intelligenz, 2007, 96; Clemens NJW 1985, 1998, 2000 ff.

⁹¹ Section 14 of the Uniform Electronic Transactions Act unter <http://www.law.upenn.edu/bll/ulc/fnact99/1990s/ueta99.pdf>. Ähnlich Sec. 21 of the Canadian Uniform Electronic Commerce Act, unter <http://www.law.ualberta.ca/alri/ulc/current/euecafin.htm>.

⁹² *Europäisches Parlament* (Fn. 5), 7.

⁹³ *MüKoBGB/Busche*, 7. Aufl. 2015, Vor § 145, 37.

Anlagenbetreibers „verlautbart“.⁹⁴ Der BGH bestätigte dies in der Flugbuchungsentscheidung – wenn auch nur in einem *obiter dictum*: „Nicht das Computersystem, sondern die Person (oder das Unternehmen), die es als Kommunikationsmittel nutzt, gibt die Erklärung ab oder ist Empfänger der abgegebenen Erklärung.“⁹⁵

Aber gilt dies auch dann im eigentlich kritischen Fall, wenn der Softwareagent nicht voll determiniert ist, sondern selbst autonom über den Vertrag entscheidet, also eigenmächtig bestimmte Angebote sucht, Verhandlungen mit potentiellen Partnern führt, den Vertragspartner bestimmt, über den Vertragsabschluss entscheidet, die *essentialia* der Willenserklärung konkretisiert, weiterhin den Vertragsablauf bestimmt, Rücktritt und andere Gestaltungsrechte ausübt, bei Vertragsbruch Sanktionen festlegt? Und gilt dies auch dann, wenn auf der Gegenseite ein elektronischer Softwareagent existiert? Besonders im Vertragsgeschehen ist das oben allgemein charakterisierte Autonomierisiko sehr viel höher als das bloße Automatisierungsrisiko. Es sind zwar immer noch die Menschen, die über das Computer-Programm und über dessen generellen Einsatz bestimmen. Aber in der darauffolgenden Vielzahl von Einzelverträgen entzieht sich das konkrete Verhalten des Softwareagenten der Kontrolle des Betreibers. Selbst vom Programmierer ist es nicht mehr im einzelnen determinierbar, nicht mehr prognostizierbar und nicht mehr kontrollierbar.⁹⁶ Doch selbst für diese Situation eines Kontrollverlusts durch Programmierer/Hersteller/Betreiber hält die herrschende Lehre daran fest, dass ausschließlich der menschliche Geschäftsherr selbst die Erklärung abgegeben hat.⁹⁷ Das ist der eigentlich kritische Fall, den die Rechtsgeschäftslehre schlicht mit dem Vertragsschluss durch reaktive Agenten gleichsetzt.⁹⁸ Selbst „lernende“ Software soll keine Ausnahme sein, da sie nicht über Programmgrenzen hinausgehen könne und ihre Funktionalitäten selbst nicht erweitere.⁹⁹

Dies ist eine unhaltbare Fiktion.¹⁰⁰ Und zwar eine Fiktion, die - anders als sonst im Falle von legitimen Rechtsfiktionen - nicht einmal offengelegt wird.

⁹⁴ MüKoBGB/*Busche*, 7. Aufl. 2015, BGB-AT Vor § 145, 37 und Einl. Rn. 185; *Palandt/Ellenberger*, BGB, 77. Aufl. 2018, vor § 116 Rn. 2; *Staudinger/Singer* BGB 2017 Vor §§ 116 ff. Rn. 57; *Soergel/Hefermehl* BGB 13. Aufl. 1999 vor § 116 Rn. 30; *Spindler/Schuster*, Recht der elektronischen Medien. Kommentar 3. Aufl. 2015, vor §§ 116 ff. BGB Rn. 9; *Döpke*, in: *Hoeren/Kolany-Raiser* (Hg.), Big Data in Context. SpringerBriefs in Law, 2018, 1, 15.

⁹⁵ BGHZ 195, 126, Rn. 17. *Kirn/Müller-Hengstenberg* (Fn. 89), 3.3.7 bei Fn. 781 kritisieren den BGH zu Recht deswegen, „weil die Verarbeitung von Abfragedaten in einem Programm kein rein technischer, automatischer und passiver Vorgang ist. Es kommt alleine darauf an, ob auch die Verarbeitung eine Problemlösung einer Aufgabe darstellt, die nun nicht mehr beim Entwickler, sondern stattdessen bei den Agenten autonom, ohne menschliche Mitwirkung, entschieden wird. Für eine technische Bewertung des Suchmaschinenagenten ist insbesondere auf die Komplexität der Aufgabestellung, auf die Autonomie der Entscheidungsfindung, die Unvollständigkeit des erforderlichen Wissens sowie die Schwierigkeit der Entwicklung einer berechenbaren Lösung abzustellen.“

⁹⁶ *Matthias* (Fn. 6), 33 ff.

⁹⁷ Kommentarliteratur siehe Fn. 94. Ebenso *Cornelius* (Fn. 99), 355; *Sorge* (Fn. 3), 36; *Kirn/Müller-Hengstenberg* (Fn. 3), 67. Das soll auch explizit dann gelten, wenn der Softwareagent die *essentialia* des Geschäfts konkretisiert, *Horner/Kaulartz* InTeR 2016, 22 ff. 22; *Spindler* (Fn. 4, 2014), 64; *Medicus/Petersen* (Fn. 3) Rn. 256.; *Bräutigam/Klindt* (Fn. 3), 1137.

⁹⁸ *Spindler* (Fn. 4, 2014), 64.

⁹⁹ *Cornelius* (Fn. 3), 353.

¹⁰⁰ Zur Kritik der Fiktion *Andrade/Novais/Machado/Neves* J. Artif Intell Law 2007, 357, 360; *Allen/Widdison* (Fn. 20), 46 ff. Gewisse Zweifel an der Fiktion plagen auch *Spindler* (Fn. 4, 2016), 816; *Pieper* InTeR 2016, 188 ff. 192; *Köhler/Arndt/Fezer*, Recht des Internet, 2011, 59.

Stattdessen wird es als unbestreitbare Tatsache ausgegeben, als „alternativer Fakt“ sozusagen, dass nur der Geschäftsherr selbst die Erklärungen abgibt. Damit verzichtet die herrschende Lehre auf eine überzeugende dogmatische Lösung. Sie schwankt zwischen verschiedenen wenig plausiblen Konstruktionsversuchen. Entweder arbeitet man mit der blanken Fiktion eines konkreten Geschäftswillens des Prinzipals.¹⁰¹ Oder man begnügt sich mit seinem allgemeinen Erklärungswillen, der mit der Entscheidung, den Softwareagenten einzusetzen, für alle Zukunft wirken soll.¹⁰² Oder es soll schon die Programmierung den Geschäftswillen erzeugen, was jedoch allenfalls bei determiniertem Computerverhalten Sinn macht. Auch die durchaus attraktive dogmatische Konstruktion, die Erklärung des Prinzipals bestehe in einem „Reservoir an bedingt abgegebenen Erklärungen“¹⁰³ überzeugt nur bei reaktiven Softwareagenten, versagt aber offensichtlich im Falle digitaler Entscheidungsautonomie. Oder man leugnet schlicht den Unterschied von determinierten und autonomen Computerverhalten.¹⁰⁴ Oder schließlich ganz ohne Dogmatik: man rechnet einfach dem Geschäftsherrn das Risiko zu.¹⁰⁵ In der Sache gibt die herrschende Lehre das Willenselement beim Prinzipal bis auf die Inbetriebnahme vollständig zugunsten des digitalen Erklärungswertes auf, aber, wie gesagt, ohne dies auch nur zuzugeben.¹⁰⁶

Andere Autoren erkennen wenigstens ihr Dilemma offen an.¹⁰⁷ Einerseits nötigten realistische Analysen der Vertragspraxis zu der Einsicht, dass „mit zunehmender ‚Verselbstständigung‘, also Autonomie des Softwareagenten, der Wille des Nutzers einflusslos wird.“ Andererseits aber seien nach geltendem Recht diese Softwareagenten nun einmal nicht rechtsfähig. Einen Ausweg aus dem Dilemma hoffen sie, mit einem „Vertrauenstatbestand“ zu finden: „dass Erklärungen oder Handlungen eines Softwareagenten, die im Rechtsverkehr einen Vertrauenstatbestand erzeugen, als rechtsverbindlich anzusehen sind“. Doch hat heute kein Mensch Vertrauen darauf, dass bei Internet-Verträgen ausschließlich Menschen die Vertragspartner sind. Computerverträge verdecken nicht eigentlich bestehende Realitäten, worauf mit Vertrauensschutz zu reagieren wäre, sondern der Verbund von Mensch und Computer ist selbst die neue Realität, die ihre eigenen speziellen Regeln benötigt.

Wie ist es möglich, dass eine Rechtsgeschäftslehre, die in der Vergangenheit erfolgreich auf die Herausforderungen des modernen entpersonalisierten Geschäftsverkehrs mit einer differenzierten Theorie objektiver Willenserklärungen

¹⁰¹ Brauner, Das Erklärungsrisiko beim Einsatz von elektronischen Datenverarbeitungsanlagen, 1988, 41 ff.

¹⁰² Soergel/Hefermehl BGB 13. Aufl. 1999 vor § 116, 8. Wiebe, Die elektronische Willenserklärung, 2002, 204 ff. Wie gequält die dogmatische Konstruktion ist, wird bei Spindler/Schuster (Fn. 94) vor §§ 116 ff. BGB Rn. 9 deutlich: „Der Handlungswille kommt im willentlichen Aktivieren des elektronischen Agenten zum Ausdruck. Erklärungsbewusstsein und Geschäftswille liegen wie bei der Computererklärung im Zeitpunkt der Erzeugung des objektiven Tatbestands der Willenserklärung durch den Agenten nicht vor. Da aber Softwareagenten nur die vom Menschen bestimmten Vorgaben ausführen und vom Anwender willentlich aktiviert werden müssen, muss auch hier eine Zurechnung der Willenserklärung erfolgen.“ Heißt dies, dass Geschäftswillen und Erklärungsbewusstsein beim Computer identifiziert werden müssen, obwohl sie doch gar nicht vorkommen sollen, und anschließend trotzdem auf den Menschen zugerechnet werden?

¹⁰³ Spindler (Fn. 4, 2014), 64.

¹⁰⁴ So ohne jede Begründung MüKoBGB/Busche 7. Aufl. 2015, BGB-AT Vor § 145, 37.

¹⁰⁵ Wiebe (Fn. 102), 204 ff.

¹⁰⁶ Dazu kritisch Schirmer (Fn. 26), 664; Gruber (Fn. 31, 2012), 133, 140, 154 f.; Teubner (Fn. 32).

¹⁰⁷ Kirn/Müller-Hengstenberg (Fn. 89), 2.2.1 bei Fn. 367 und 368.

reagiert hat, auf die digitale Herausforderung nur noch defensiv mit uneingestandenen Fiktionen antwortet? Welch ein Widerspruch, dass das deutsche Privatrecht auf der einen Seite klammheimlich Verträge zwischen Maschinen als rechtsgültig anerkannt hat, auf der anderen Seite aber nicht bereit ist, aus der Delegation von Vertragsentscheidungen an Softwareagenten die notwendigen Konsequenzen für tragende Elemente der Rechtsgeschäftslehre zu ziehen. Stattdessen beruhigt man sich mit der kontrafaktischen Versicherung, dass es auch im Zeitalter massenhafter Internet-Transaktionen immer noch die Menschen seien, die das Vertragsgeschehen beherrschten. Schlimmer noch, man beschwört Willensfreiheit und Menschenwürde, um jede Art von Rechtssubjektivität für Softwareagenten als verfassungswidrig (!) zu blockieren: Es könne das „Grundbekenntnis der Verfassung keine Anerkennung einer Rechtspersönlichkeit mit Rechtsfähigkeit für diese Softwareagenten rechtfertigen.“¹⁰⁸ Zugleich aber – und darin steckt die Ironie – werden angesichts der faktischen Delegation wesentlicher Entscheidungen an seelenlose Maschinen die hohen ethischen Ansprüche an Privatautonomie, wonach menschliche Subjekte das Vertragsgeschehen eigenverantwortlich beherrschen sollen, gerade mit Hilfe dieser Fiktion schlicht aufgegeben.

Deshalb sollte man konsequent auf die Gegenposition übersetzen, wie es eine Reihe von Autoren vertreten:¹⁰⁹ In genauer Entsprechung zu seiner realen Funktion in der Wirtschaftspraxis, ökonomisch gesprochen, in einer *principal-agent*-Beziehung,¹¹⁰ gibt der Softwareagent rechtlich eigenverantwortlich die Willenserklärung ab, handelt aber nicht im eigenen Namen, sondern im Namen des Prinzipals. Die Erklärung des Computers ist als solche rechtsverbindlich und berechtigt und verpflichtet unmittelbar den Prinzipal. *In Analogie zum Recht der Stellvertretung sind autonome Softwareagenten als Repräsentanten ihres menschlichen Prinzipals zu behandeln.*

Dafür braucht man ihnen – wie schon oben gesagt – keine volle Rechtsfähigkeit als juristische Person zuzuschreiben, sondern in funktionaler Sicht genügt die bloße Teilrechtsfähigkeit, die Stellvertretungsfähigkeit.¹¹¹ Dass die Verleihung der Rechtsfähigkeit schon *de lege lata* möglich ist, zeigt die Geschichte des nicht-rechtsfähigen Vereins, besonders der Gewerkschaften, der Vorgesellschaft und jüngst die Entscheidungen zur Rechtsfähigkeit der BGB-Gesellschaft. Dies sind Präjudizien, in denen die Gerichte *praeter legem*, wenn nicht *extra legem*,

¹⁰⁸ Kirn/Müller-Hengstenberg (Fn. 89), 2.2.1 bei Fn. 351.

¹⁰⁹ Beck (Fn. 5), 186; Schirmer (Fn. 26), 664; Kersten (Fn. 20), 7; Gruber (Fn. 31, 2015), 198; Gruber (Fn. 31, 2012), 140 ff.; Wettig (Fn. 20), 179 ff.; Gitter (Fn. 90), 177 ff.; John (Fn. 90), 77 ff.; 154 f.; Teubner (Fn. 32); Sorge (Fn. 3), 118; Schwarz, in: Schweighofer (Hg.), Auf dem Weg zur ePerson, 2001, 67, 67 f.; Fischer Ind. L. J. 1997, 557 ff. Für eine ähnliche Lösung plädieren Günther (Fn. 20), 54 ff.; Gitter/Roßnagel Kommunikation & Recht 2003, 64 ff. 66, da sie eine Analogie zu den Vertretungsregelungen unter dem Gesichtspunkt der Blanketterklärung vorschlagen. Implizit erkennen sie damit die rechtliche Handlungsfähigkeit der Softwareagenten an, obwohl sie sie explizit ausschließen. Vgl. auch John (Fn. 90), 102 ff. Für Stellvertretung, aber nur *de lege ferenda* Loos (Fn. 32), 74 ff.; Keßler Multimedia und Recht 2017, 589, 592; Mayinger (Fn. 5), 227 ff.

Ausdrücklich gegen die Anwendung der Stellvertretung auf Softwareagenten, auch für den Fall, dass sie autonom sind: Kluge/Müller (Fn. 31) 27; Spindler/Schuster (Fn. 94), § 164 BGB Rn. 9-13.

¹¹⁰ Zur ökonomischen Definition der Principal-Agent-Beziehung, die weitgehend der rechtlichen Qualifikation als Stellvertretung entspricht: Jensen/Meckling (Fn. 27).

¹¹¹ Zur durchaus vergleichbaren Konstruktion der Rechtsgeschäfte von Sklaven im römischen Recht Harke (Fn. 28), 97 f.

(beschränkte) Rechtsfähigkeit an zuvor nicht-rechtsfähige Entitäten verliehen haben.¹¹²

Freilich sind die Unterschiede zwischen menschlichen Akteuren und Softwareagenten so groß, dass manche Normen des Stellvertretungsrechts nicht unmittelbar angewendet werden können. Es gilt also, für die Willenserklärungen eines digitalen Stellvertreters in einer sorgfältigen Analogie zum Stellvertretungsrecht adäquate Sonderregeln zu entwickeln.¹¹³ Aus dem traditionell auf menschliche Akteure zugeschnittenen Stellvertretungsrecht ist eine Analogiebasis zu generalisieren und diese dann auf die Eigenheiten digitalen Handelns zu respezifizieren.

Mit Hilfe einer solchen Analogie lässt sich der wohl wichtigste Einwand ausräumen, dass den Softwareagenten die notwendigen Willenselemente fehlen.¹¹⁴ Mit gebotener Sorgfalt ist es durchaus möglich, digitale Äquivalente für die sogenannten subjektiven Voraussetzungen der Willenserklärung, besonders für das Erklärungsbewusstsein, zu finden. An dieser Stelle kommen die bekannten Objektivierungstendenzen in der Rechtsgeschäftslehre der Möglichkeit, dass bewusstseinslose Softwareagenten rechtswirksame Willenserklärungen abgeben können, weit entgegen.¹¹⁵ Sie bestätigen die schon angesprochene allgemeine soziologische These, dass es für die Personifizierung nicht-menschlicher Entitäten nicht auf Bewusstsein, sondern auf Kommunikation ankommt. Bekanntlich hat die Auflösung der Kontroverse Willenstheorie versus Erklärungstheorie in modernen Geltungstheorien und Vertrauenstheorien das Erklärungsbewusstsein entpsychologisiert.¹¹⁶ Auf den subjektiven Willen des Erklärenden kommt es nach der dramatischen Wende in der neueren Rechtsprechung gar nicht mehr an. Der objektive Vertrauensgrundsatz hat die subjektiven Vorstellungen der Parteien verdrängt. Entscheidend ist, dass sich der Erklärende sein falsches äußeres Verhalten im Rechtsverkehr zurechnen lassen muss. Der Erklärende kann sein fehlendes Erklärungsbewusstsein nicht der Geltung der Willenserklärung entgegensetzen, wenn er wie der BGH entschieden hat, „fahrlässig nicht erkannt hat, dass sein Verhalten als Willenserklärung aufgefasst werden könnte und wenn der Empfänger es tatsächlich so verstanden hat“.¹¹⁷

An die Stelle des subjektiven Erklärungsbewusstseins setzt also der BGH eine doppelte objektive Normierung. Damit aber können Softwareagenten, die über elaborierte kognitive Kommunikationsfähigkeiten verfügen, umgehen. Die Normierung besteht aus zwei Normen: erstens, die aus dem Vertrauen entstandene soziale Norm, ob das konkrete Verhalten als eine bindende Willenserklärung

¹¹² Zu dieser Parallele eingehend *Gruber* (Fn. 15), 267 ff. Vgl. BGHZ 146, 341, 344 „beschränkte Rechtssubjektivität der Außen-GbR“; BGHZ 163, 154 – „Teilrechtsfähigkeit der Wohnungseigentümergeinschaft als einer auf Gesetz beruhenden Personengemeinschaft“.

¹¹³ So auch *Gruber* (Fn. 31, 2012), 140 ff.; *Schirmer* (Fn. 26), 664.

¹¹⁴ *Spindler* (Fn. 4, 2014), 816 räumt zwar ein, „dass prima vista die semi-autonome ‚Erklärung‘ des Roboters eher derjenigen eines Stellvertreters ähnelt; doch fehlt es immer noch an einem Erklärungsbewusstsein“.

¹¹⁵ Das sehen auch deutlich *Kirn/Müller-Hengstenberg* (Fn. 89), 2.2. Dann aber machen sie eine Kehrtwendung und erklären apodiktisch, dass Computer selbst keine rechtlich wirksame Erklärung abgeben können, weil sie kein „Ichbewußtsein“ haben.

¹¹⁶ Z.B. *MüKoBGB/Säcker* 7. Auflage 2015 BGB-AT Einleitung, Rn. 158-162 mwN.

¹¹⁷ BGH NJW 1995, 953; BGH NJW 2002, 3629, 2631. *MüKoBGB/Armbrüster*, 7. Aufl. 2015, BGB-AT § 119 Rn. 94.

verstanden werden darf, zweitens, die Pflicht des Erklärenden, diese soziale Norm zu erkennen und nicht gegen diese Pflicht zur Kenntnisnahme fahrlässig zu verstoßen. Mit einer solchen Entpsychologisierung hat die Rechtsprechung selbst schon die gesuchte Generalisierung der Analogiebasis vollzogen, deren Respezifizierung in Sonderregeln für Softwareagenten ohne größere Schwierigkeiten möglich ist. Eine solche Kenntnis sozialer Normen, also wie in einem bestimmten Kontext bestimmte Erklärungen zu verstehen sind, kann durchaus in ein Softwareprogramm übersetzt werden.¹¹⁸ Ja, lernfähige Agenten können diese Information sogar selbsttätig erwerben und sie darüber hinaus bei Änderungen der sozialen Normen oder der Rechtsprechung selbsttätig abändern. Damit ist der Einwand, Softwareagenten als digitale Stellvertreter könnten kein Erklärungsbewusstsein besitzen, weil sie kein Bewusstsein haben, ausgeräumt. Das Äquivalent sind ihre kognitiven Kommunikationsfähigkeiten.

Die konstruktiven Differenzen der beiden Lösungen erzeugen auch haftungsrechtliche Unterschiede. Die herrschende Lehre gerät dann in Schwierigkeiten, wenn der Softwareagent – wie ein *falsus procurator* – sich als berechtigt ausgibt, im Namen eines anderen zu handeln. Da sie die Fiktion aufrechterhält, dass nicht der Softwareagent, sondern der Betreiber gehandelt hat, ist dieser dem hohen Risiko ausgesetzt, dass er nun selbst als vollmachtsloser Vertreter nach §179 haftet.¹¹⁹ Besonders hoch wird dies Risiko bei distribuiertem Aktion¹²⁰ oder *self-cloning*¹²¹ der Softwareagenten eingeschätzt. Viele Autoren sehen dies als unangemessen an, denn der Betreiber kann nicht die Handlungen des Softwareagenten mit genügender Wahrscheinlichkeit voraussehen und er hat wenig Macht, sie zu kontrollieren.¹²² Auch die gelegentlich vorgeschlagene Ausweichlösung, dass der Betreiber eine Gesellschaft gründen könnte, deren einziger Zweck online-trading wäre, ist wenig befriedigend.

Behandelt man dagegen den Softwareagenten als Stellvertreter, dann wäre der Betreiber als Geschäftsherr nicht verpflichtet. Eine Eigenhaftung des Softwareagenten wie ein *falsus procurator* des § 179 ginge mangels Eigenvermögen ins Leere, wie es auch in anderen Fällen, in denen der *falsus procurator* nicht erreichbar ist, geschieht. Nur de lege ferenda wäre eine Eigenhaftung möglich, wenn ihm als juristische Person eigenes Vermögen zugeordnet wäre.¹²³ Solange dies nicht der Fall ist, muss ein interessengerechter Ausgleich zwischen dem Betreiber und Vertragspartner gefunden werden, wonach der Betreiber nicht, wie die herrschende Lehre annehmen muss, automatisch haftet, sondern nur dann, wenn er selbst eine Sorgfaltspflicht verletzt hat.¹²⁴

2. Vertragliche Haftung

¹¹⁸ Zu den rechtlich relevanten kognitiven Fähigkeiten der Softwareagenten *Mayinger* (Fn. 5), 64 ff.

¹¹⁹ *Andrade et al* (Fn. 100), 360; *Sartor* (Fn. 16).

¹²⁰ *Allen/Widdison* (Fn. 20), 42.

¹²¹ *Barfield* Presence 2005, 747 ff.

¹²² *Mayinger* (Fn. 5), 72; *Koops et al.* (Fn. 22), 554.

¹²³ *Mayinger* (Fn. 5), 72, 227, 244 ff. Sofern der Prinzipal eine Versicherung für den Softwareagenten abgeschlossen hat, wäre in Analogie zu § 179 zu erwägen, dem geschädigten Vertragspartner den Zugriff auf die Versicherungssumme zu ermöglichen.

¹²⁴ Hier ist abzuwägen zwischen dem Veranlasser-Risiko des Geschäftsherrn und dem Risiko des Vertragspartners. Das Prinzip sollte die Haftung des Geschäftsherrn sein, eine Ausnahme gälte bei offensichtlichem Fehlverhalten des Computers.

Bei den eben angesprochenen Computer-Erklärungen ging es um die Verbindlichkeit von Verträgen, die Softwareagenten für einen Prinzipal abschließen. Nun geht es um die Situation, in der Menschen oder Unternehmen Verträge abschließen und zur Vertragserfüllung Softwareagenten heranziehen. Wie ist dann eine Schadensverursachung durch den Computer zu beurteilen? Verstößt das Verhalten autonomer Softwareagenten gegen Vertragspflichten, insbesondere schädigt es die Rechtsgüter des Vertragspartners, dann gerät die Lehre, die sich weigert, den Agenten die beschränkten Handlungsfähigkeiten eines Erfüllungsgehilfen zuzusprechen, in große Schwierigkeiten. Nicht nur, dass sie deswegen wie in der Rechtsgeschäftslehre zu merkwürdigen dogmatischen Verrenkungen gezwungen ist, schlimmer ist, dass gerade diese Weigerung erhebliche Haftungslücken aufreißt. Diese sind besonders gravierend, wenn die Schäden beim Einsatz von Pflegerobotern, Medizinrobotern und Fertigungsrobotern entstehen.

Da die Lehre an der Prämisse unbeirrt festhält, dass selbst autonome Softwareagenten nicht handeln können und deshalb auch keine Erfüllungsgehilfen sein können,¹²⁵ kann sie die Vertragspflichtverletzung nicht am eigentlich schädigenden Verhalten des Softwareagenten festmachen, sondern muss sämtliche Einzelelemente der Vertragsverletzung, also die tatbestandlichen Voraussetzungen, die Vertragswidrigkeit des Verhaltens und das Verschulden ausschließlich in der Person des Vertragsschließenden, also des menschlichen Betreibers, suchen. Anspruchsgrundlagen sind dann §§ 276, 280, die auf die Pflichtverletzung durch den Betreiber selbst abstellen¹²⁶. Dies aber ist eine dogmatische Fehlkonstruktion, denn als schuldhaftige Vertragsverletzung kann dann dem Betreiber, wie Kritiker zurecht monieren, nur die Inbetriebnahme (!) des Computers vorgeworfen werden.¹²⁷ Die bloße Inbetriebnahme ist ein viel zu weit gefasster Vorwurf, der zugleich die sonst stets erforderliche detaillierte Prüfung verhindert, ob ein konkretes Schadensgeschehen die einzelnen Voraussetzungen einer Pflichtverletzung erfüllt. Man beruhigt sich damit, dass sich der Betreiber wegen der Verschuldensvermutung des § 280 S. 2 regelmäßig nicht entlasten kann.¹²⁸ Das aber stimmt nicht. Der Prinzipal kann gerade im Fall einer für ihn undurchschaubaren, nicht prognostizierbaren und nicht erklärbaren Schadensverursachung durch den autonomen Softwareagenten relativ einfach die Vermutung widerlegen. Es dürfte geradezu der Regelfall sein, dass ein Betreiber das Verhalten des Softwareagenten weder vorhersehen noch vermeiden kann. Notfalls müsste Beweis über die Prognostizierbarkeit durch Befragung des Programmierers oder Expertengutachten erhoben werden. Gegenüber einer solchen verfehlten Anknüpfung an die Inbetriebnahme als Pflichtverletzung, die letztlich auf eine systemfremde „Veranlasserhaftung“ hinausläuft, bestehen nicht nur rechtsdogmatische Bedenken.

¹²⁵ Palandt/Grüneberg, BGB 77. Aufl. 2018, § 278 Rn. 11; Redeker, IT-Recht, 2012, Rn. 969 f.; Koch, Internet-Recht 2005, § 3, Nr. 1. b, S. 101; Heckmann, Internetrecht 2007, Kap. 4.1 Rn. 18; Cornelius (Fn. 99), 355. Gegen § 278 explizit bei autonomen Softwareagenten: Horner/Kaulartz (Fn. 97), 23; ebenso, wenn auch mit Zweifeln, ob dies nicht falsche Anreize setzt MüKoBGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46.

¹²⁶ Horner/Kaulartz (Fn. 97), 23; Spindler (Fn. 4, 2014), 63, 66 ff.; Hanisch, in: Hilgendorf (Hg.), Robotik im Kontext von Recht und Moral, 2014, 27 ff.; 32; Hanisch, in: Günther/Hilgendorf (Hg.), Robotik und Gesetzgebung, 2013, 109, 116.

¹²⁷ Dies kritisieren auch Lohmann (Fn. 53), 158; Schirmer (Fn. 26), 664.

¹²⁸ MüKoBGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46; Kirn/Müller-Hengstenberg MultiMedia und Recht 2014, 307, 309 ff.; Spindler (Fn. 4, 2014), 63, 66 ff.

Rechtspolitisch wird dagegen eingewandt, dass damit jede Steuerungsintention der Verschuldenshaftung leerläuft. Zudem sei es innovationsfeindlich, wenn der bloße Betrieb zum Haftungsgrund gemacht wird.¹²⁹

Schlimmer aber ist: Bei dieser dogmatisch und rechtspolitisch fragwürdigen Konstruktion entsteht zugleich eine schwer erträgliche Haftungslücke, die sich in Zukunft noch ausweiten wird, je mehr Aufgaben der Vertragserfüllung an autonome Softwareagenten delegiert werden:¹³⁰ „Kann der Betreiber aber nachweisen, dass der Schadenseintritt nach dem Stand der Technik weder vorhersehbar noch vermeidbar war, so entfällt nach dem Grundgedanken der objektiven Unmöglichkeit der Vermeidung von unerwarteten Problemlösungen bzw. Entscheidungen nach § 275 BGB eine Haftung.“¹³¹

Beide Schwierigkeiten – die Haftungslücke ebenso wie die dogmatische Fehlkonstruktion – lassen sich dagegen problemlos vermeiden, wenn man die Analogie wagt und *die Gehilfenhaftungsregel des § 278 analog auf das Verhalten autonomer Softwareagenten anwendet*. Auch wenn dem Geschäftsherrn selbst kein Fehlverhalten vorzuwerfen ist, haftet er dann für jedes Verhalten des Softwareagenten, sofern dieses die Vertragspflichten verletzt und einen Schaden verursacht.¹³² Die Analogie würde, wie Grundmann aus ökonomischer Sicht darlegt, positiv zur Internalisierung der Kosten von Maschinenversagen beitragen, und da der Entscheidungs- zugleich Kostenträger wäre, hätte er den optimalen Anreiz, Nutzen und Kosten von mehr Maschinensicherheit minimierend abzuwägen.¹³³

Die Analogie beseitigt die Haftungslücke vollständig. Denn der Geschäftsherr kann sich nicht mehr damit entlasten, ihm selbst sei kein Fehlverhalten vorzuwerfen. Nach § 278 kommt es allein auf das Fehlverhalten des Softwareagenten an.¹³⁴ Hier liegt der eigentliche Vorzug der Analogie zur Gehilfenhaftung gegenüber der Haftung des Prinzipals für sein eigenes Fehlverhalten. Denn auch dann, wenn der Prinzipal alle Sorgfaltspflichten beim Einsatz des Computers erfüllt hat, muss er dennoch für ein Entscheidungsversagen des autonomen Softwareagenten haften, genauso wie wenn ein menschlicher Erfüllungsgehilfe gehandelt hätte.

Spindler wendet dagegen ein: Wenn der Betreiber alle Kontrollpflichten im vorgelagerten Bereich eingehalten hat, dann dürfe ihm beim Einsatz von Softwareagenten genauso wenig eine Haftung treffen wie auch sonst beim Einsatz anderer gefährlicher Technologien.¹³⁵ Dieser durchaus bedenkenswerte Einwand

¹²⁹ Hanisch (Fn. 126, 2014), 34.

¹³⁰ Auch MüKoBGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46 spricht von einer Regelungslücke, die durch die technologische Entwicklung bedingt ist.

¹³¹ Kirn/Müller-Hengstenberg (Fn. 11), 16. Ebenso Palandt/Grüneberg BGB 77. Aufl. 2018, § 276 Rn. 20 f.; Erman/Westermann, BGB § 276 Rdnr.13 und § 275 Rdnr.1,5; BGH NJW 2007,762, 763; BGH NJW 2011, 756, 757. Ähnlich Heussen CR 2004,1 ff.

¹³² Für die Analogie Keßler (Fn. 109), 592 f.; Schirmer (Fn. 26), 665; Gruber (Fn. 31, 2015), 198; Teubner (Fn. 32); vorsichtiger nur für die Zukunft befürwortend Günther (Fn. 20), 84. Generell befürwortend auch Soergel/Pfeiffer BGB 13. Aufl. 2014 Rn. 29; Wolf JuS 1989, 899 ff. 901; Spiro, Die Haftung für Erfüllungsgehilfen, 1984 211 f.; Löwenheim BB 1967, 593 ff. 597. Ausführlich hierzu Hanisch, Haftung für Automation, 2010 36 f., 30.

¹³³ MüKoBGB/Grundmann, 7. Aufl. 2016, BGB § 278 Rn. 46, der sich dann aber doch gegen die Analogie ausspricht.

¹³⁴ Schirmer (Fn. 26), 665. Die Noxalhaftung des Sklavenhalters im römischen Recht ist durchaus einer solchen Haftung für elektronische Agenten vergleichbar.

¹³⁵ Spindler (Fn. 4, 2016), 816.

verkennt jedoch die prinzipiell neue Qualität, die beim Entscheidungsversagen autonomer Softwareagenten im Unterschied zu sonstigem Maschinenversagen entsteht.¹³⁶ Auch wenn den beteiligten Menschen keine Pflichtwidrigkeit vorgeworfen werden kann, aber zugleich autonome Softwareagenten vor Entscheidungsalternativen gestellt worden sind, aufgrund ihrer Programme eine schadensverursachende Fehlentscheidung getroffen haben, diese Entscheidung aber von Programmierern weder erklärt noch prognostiziert werden konnte, dann ist die Haftung angezeigt. Also gerade auch dann, wenn in vergleichbaren Fällen für bloßes Maschinenversagen aufgrund fehlgelaufener Kausalverläufe nicht gehaftet wird. Der tiefere Grund liegt, wie gesagt, darin, dass die Gesellschaft, wenn sie neuartige Entscheidungsräume für bis dahin unbekannte autonome Entscheidungsträger zulässt, verpflichtet ist, für wirksame Verantwortungsformen zu sorgen. Entscheidend sind letztlich nicht Gesichtspunkte von Effizienz, von Transaktionskostensparnissen, von utilitaristischen Erwägungen, von soziologischer Jurisprudenz oder von policy-questions, sondern dies ist eine genuine Frage juridischer Gerechtigkeit.¹³⁷ Es ist der Gleichbehandlungsgrundsatz, der hier die Haftung verlangt. Denn würde ein Mensch für die Vertragsdurchführung herangezogen, so haftete der Prinzipal nach § 278 für dessen Pflichtverletzungen, er kann aber nicht von der Haftung befreit sein, wenn für die identische Aufgabe ein Softwareagent herangezogen wird. Diese nicht zu rechtfertigende Privilegierung, die darin besteht, dass die Digitalisierung einem Computerbetreiber einen solchen erheblichen Kostenvorteil zulasten der geschädigten Vertragspartner verschafft, müsste man der Rechtsdogmatik, die unbeirrt an tradierten Rechtsbegriffen festhält, vorwerfen.

Subjektive Voraussetzungen sollten bei der Analogie zur Gehilfenhaftung noch weniger Schwierigkeiten machen als bei der Stellvertretung. Für die Gehilfenhaftung kommt es erst recht nicht auf innere Fähigkeiten an, sondern im Prinzip nur auf Kommunikationsteilnahme. Denn im Rahmen von § 278 ist anerkannt, dass sich der Geschäftsherr sogar das Handeln von unzurechnungsfähigen, also nicht verschuldensfähigen Gehilfen zurechnen lassen muss.¹³⁸ Zudem wird das Problem auch hier wieder durch Objektivierungstendenzen im heutigen Privatrecht, welche die Anforderungen an das Verschulden verändert haben, entschärft. Diese Tendenzen lassen in der „objektiven Fahrlässigkeit“ Rechtswidrigkeit und Schuld beinahe miteinander verschmelzen.¹³⁹ In der Sache wird bei Vertragsverletzungen das Verschulden heute auf einen Spezialfall der Rechtswidrigkeit reduziert. Während die Rechtswidrigkeit die allgemeinen Anforderungen an die Vertragsverletzung definiert, legt das Verschulden die rollen- und situationsspezifische Anforderungen fest, die an das konkrete Verhalten gestellt werden. Das erleichtert die Anforderungen an die Analogie. Denn in der Informationsphilosophie, welche die moralische Verantwortlichkeit von Algorithmen untersucht, argumentiert man, dass die kognitiven Fähigkeiten autonomer Softwareagenten in der Weise entwickelt sind, dass ihnen zwar keine „responsibility“ im Vollsinn zugesprochen werden kann, wohl aber ihre „accountability“.¹⁴⁰ In juristischer Übersetzung hieße dies, dass ihnen Rechtswidrigkeit zugerechnet werden kann, aber kein Verschulden im Sinne von subjektiver Vorwerfbarkeit. Und an dieser Stelle können an Softwareagenten sogar

¹³⁶ So entschieden *Matthias* (Fn. 6), 111 ff.

¹³⁷ So auch *Schirmer* (Fn. 26), 665.

¹³⁸ *MüKoBGB/Grundmann*, 7. Aufl. 2016, BGB § 278 Rn. 50.

¹³⁹ *MüKoBGB/Grundmann*, 7. Aufl. 2016, BGB § 276 Rn. 53 ff.

¹⁴⁰ *Floridi/Sanders* (Fn. 8), 205 ff.

höhere Sorgfaltsanforderungen als an handelnde Menschen gestellt werden, sofern sie wegen ihrer überlegenen Informationsverarbeitungskapazität über höhere kognitive Fähigkeiten verfügen.¹⁴¹ Minimalstandard aber bleiben auch für Softwareagenten die Pflichtanforderungen, die im Verkehr im Allgemeinen gestellt werden.¹⁴²

3. Deliktische Haftung

Im Bereich außervertraglicher Haftung wird das Versagen des geltenden Rechtes gegenüber dem digitalen Autonomierisiko überdeutlich. In aller Härte wird kritisiert, dass hier eine erhebliche Haftungslücke aufgerissen ist. Es wird nachdrücklich das Eingreifen der Legislative gefordert.¹⁴³ Wegen seiner Bedeutung für die wichtigsten Fälle von Computerversagen ist dies der Testfall dafür, ob man den Softwareagenten partielle Rechtsfähigkeit geben soll oder nicht. Nicht nur, weil die Schadenshäufigkeit hoch und die Schadenshöhe beträchtlich ist. Sondern im Gegensatz zu vertraglicher Haftung, bei denen der Vertragspartner sich immerhin „freiwillig“ dem Risiko aussetzt, einen digitalen Vertragspartner zu haben, ist in Fällen außervertraglicher Haftung der Geschädigte unfreiwillig dem Computerrisiko ausgesetzt, trägt also stellvertretend das Risiko der Gesellschaft.

Der eigentliche Grund für das Versagen ist auch hier, dass sich die haftungsrechtlichen Normen ausschließlich darauf konzentrieren, ob Betreibern/Herstellern/Programmierern ein Fehlverhalten vorzuwerfen ist. Dagegen können nach geltendem Recht die Fehlentscheidungen, die Softwareagenten selbst trotz korrekten Verhaltens der menschlichen Beteiligten fällen, nicht sanktioniert werden. Die Haftungslücke entsteht sowohl bei der Produzentenhaftung des § 823 als auch bei der Produkthaftung nach dem Produkthaftungsgesetz, wenn auch in je unterschiedlichem Maße. Der Betreiber ist von der Haftung befreit, wenn er seine Sicherungsvorkehrungen stets an den neuen Stand von Wissenschaft und Technik angepasst hat.¹⁴⁴ Der Produzent haftet nicht, wenn er sämtliche Konstruktionspflichten, Informationspflichten und Produktbeobachtungspflichten erfüllt hat.¹⁴⁵ Besonders fragwürdig erscheint die Haftungsfreistellung für Entwicklungsrisiken angesichts der programmierten (!) Nichtprognostizierbarkeit (!) der Algorithmus-Entscheidungen.¹⁴⁶ Wenn also die menschlichen Beteiligten ihre Sorgfaltspflichten erfüllt haben, dennoch aber der Softwareagent eine rechtswidrige und schadensverursachende Entscheidung trifft, entsteht die vielkritisierte Haftungslücke. Sie wird in Zukunft wachsen, weil mit zunehmender Verselbständigung autonomer Systeme nicht mehr erwartet werden kann, dass die Nutzer das Verhalten durchgängig überprüfen, und entsprechend die Standards der Prüfpflichten sinken werden.¹⁴⁷

Hier de lege lata dennoch eine Haftung zu ermöglichen, dürfte sehr viel schwieriger sein als im Bereich vertraglicher Haftung, wo sich die Analogie zu § 278

¹⁴¹ Jedenfalls dann, wenn die gesteigerten Fähigkeiten implizit in den Vertragsschluss eingingen, dazu MüKoBGB/*Grundmann*, 7. Aufl. 2016, BGB § 278 Rn. 50 mwN.

¹⁴² *Janal* (Fn. 5), 159.

¹⁴³ *Europäisches Parlament* (Fn. 5), 6.

¹⁴⁴ *Döpke* (Fn. 94), 17; *Bräutigam/Klindt* (Fn. 3), 1140; *Kirn/Müller-Hengstenberg* (Fn. 3), 68; *Wendt/Oberländer InTeR* 2016, 58 ff.; *Horner/Kaulartz* (Fn. 97), 24; *Spindler* (Fn. 4, 2014), 70 f.

¹⁴⁵ Dazu eingehend *Wagner AcP* 2017, VI; *Spindler* (Fn. 4, 2014), 72, 78.

¹⁴⁶ Hier sieht auch *Lohmann* (Fn. 53), 158 das eigentliche Problem.

¹⁴⁷ *Horner/Kaulartz* (Fn. 97), 25.

anbietet. Eine effektive Reaktion auf das Autonomierisiko wäre es freilich, wie es in der Literatur auch vorgeschlagen wird,¹⁴⁸ die autonomen Softwareagenten als digitale Verrichtungsgehilfen zu behandeln und die Betreiber stets einstehen zu lassen, wenn sie rechtswidrige Schädigungshandlungen begehen.¹⁴⁹ Doch offensichtlich scheitert eine Anwendung des § 831 am berüchtigten Entlastungsbeweis.¹⁵⁰ Zudem dürfte die Weisungsgebundenheit ein zu enggefasstes Kriterium sein. Die Organhaftung des § 31 wiederum scheitert am Merkmal des verfassungsmäßig berufenen Vertreters. Auf eine - gewiss gewagte - Analogie zu beiden Normen in Richtung einer digitalen Assistenzhaftung *de lege lata* dürfte sich die Rechtsprechung nicht einlassen, nachdem sie nun einmal im Bereich der Unternehmenshaftung ähnliche Lösungen ausgeschlagen hat und stattdessen den Umweg über Organisationspflichten der Unternehmensleitung gegangen ist. Am ehesten noch wird die Rechtsprechung versuchen, das digitale Autonomierisiko ein wenig abzumildern, indem sie die Anforderungen an die Sorgfaltspflichten der menschlichen Beteiligten immer mehr erweitert, sie dann aber voraussichtlich soweit überspannt, dass sie praktisch nicht mehr erfüllbar sind.¹⁵¹

Eine weitere denkbare Lösung wäre eine generelle Unternehmenshaftung für die Handlungen autonomer Softwareagenten, wonach deren Fehlverhalten unmittelbar dem Unternehmen zugerechnet wird.¹⁵² Das wäre im Grunde eine Verbundhaftung, die im umfassenden Rahmen der Unternehmensorganisation dem hybriden Charakter der Mensch-Computer-Interaktion mit einem Mitgliedsstatus des Softwareagenten im Verbund gerecht würde. Aber es wäre unrealistisch, dies beim heutigen Stand des Unternehmenshaftungsrechts *de lege lata* vorzuschlagen.

4. Gefährdungshaftung oder digitale Assistenzhaftung?

Schließlich erschiene eine richterrechtliche Gesamtanalogie zu bestehenden Gefährdungshaftungstatbeständen, wie sie in der Literatur vertreten wird,¹⁵³ aussichtsreich, das digitale Risiko zu bewältigen. Sie dürfte aber an dem bisher unumstößlichen Dogma scheitern, dass es das ausschließliche Privileg des Gesetzgebers ist, neue Gefährdungshaftungen anzuordnen.¹⁵⁴

Aber *de lege ferenda* scheint die Gefährdungshaftung der Königsweg zu sein, auf dem man dem digitalen Autonomierisiko erfolgreich begegnen könnte.¹⁵⁵ Die meisten Autoren fordern auch mit großem Nachdruck eine entsprechende Gesetzgebung, die auf die digitalen Gefahren mit einem neuen Gefährdungshaftungstatbestand reagiert. Vorbilder dafür sind

¹⁴⁸ Eine Analogie zu § 831 erwägen *Kluge/Müller* (Fn. 31), 28.

¹⁴⁹ Diskussion von § 831 für Softwareagenten bei *John* (Fn. 90), 272 ff.; *Günther* (Fn. 2015), 134 f.; *Zech* (Fn. 5), 190.

¹⁵⁰ *Zech* (Fn. 5), 190, 194 f.; *Günther* (Fn. 20), 134 ff.

¹⁵¹ Zu solchen Tendenzen der Rechtsprechung, die Anforderungen an Sorgfaltspflichten zu überspannen, *Lohmann* (Fn. 20), 169; *Lohmann* (Fn. 53), 159; *Gruber* (Fn. 30), 151 ff., 158; *Hanisch* (Fn. 126, 2014), 60.

¹⁵² Zu einer überzeugenden rechtspolitischen Konzeption einer Unternehmenshaftung *MüKoBGB/Wagner*, 7. Aufl. 2017, BGB, § 831, Rn. 3 f.

¹⁵³ Zu der Analogiefrage *MüKoBGB/Wagner*, 7. Aufl. 2017, BGB vor § 823, Rn. 25 ff. mwN.

¹⁵⁴ Dieses Privileg wird auch in der Literatur zur Haftung von Softwareagenten durchgängig respektiert: *Bräutigam/Klindt* (Fn. 3), 1140; *OKBGB/Spindler*, 3. Aufl. 2014, § 823, Rn. 0.21.

¹⁵⁵ *Europäisches Parlament* (Fn. 5), 6; *Günther* (Fn. 20), 237 ff. (mit gewissen Vorbehalten); *Rempe*, *InTeR* 2016, 17, 19; *Schirmer* (Fn. 26), 665; *Bräutigam/Klindt* (Fn. 3), 1140 mwN; *Gruber* (Fn. 30), 198; *Hanisch* (Fn. 126, 2014), 35 f.; *Hanisch* (Fn. 132), 177 ff.; *Spindler* (Fn. 20), 775.

Gefährdungshaftungsnormen für Kraftfahrzeuge, für nukleare Anlagen, für gentechnische Verfahren und für Umweltschäden.¹⁵⁶ Dies führt entweder zu einer Haftung des Software-Betreibers oder des Herstellers, weil dort Veranlassung und wirtschaftlicher Nutzen zusammenfallen?¹⁵⁷

Doch muss hier ein fundamentales Missverständnis ausgeräumt werden. Die Leitprinzipien der Gefährdungshaftung können gar nicht als Vorbild dienen, da sie die Eigenheiten des digitalen Autonomierisikos schlicht verfehlen. Die These heißt: *Nicht eine Haftung für den rechtmäßigen Einsatz gefährlicher Anlagen, sondern Haftung für rechtswidriges Fehlverhalten des autonom entscheidenden Softwareagenten* ist als Grundprinzip einer strikten Haftung für digitales Handeln angemessen. Entsprechend wird hier der Vorschlag unterbreitet, statt Gefährdungshaftung eine „digitale Assistenzhaftung“ d.h. eine strikte deliktische Haftung des Prinzipals für Fehlverhalten des Softwareagenten einzuführen.

Das Risiko digitaler Entscheidungsautonomie ist prinzipiell andersgeartet als die Risiken, auf die es in den bisherigen Fällen der Gefährdungshaftung ankommt.¹⁵⁸ Weder trifft die Unausweichlichkeit des Schadens wie in den Funkenflugfällen, noch die hohe Wahrscheinlichkeit von Unfällen wie im Straßenverkehr oder Flugverkehr die besondere Gefahr digitaler Autonomie. Ausschlaggebend für eine strikte digitale Haftung ist auch nicht die enorme Schadenshöhe wie beim Betrieb von Atomkraftwerkwerken oder bei Starkstromanlagen, ebensowenig der Ausgleich ungleich verteilter Schadenspotentiale wie bei fast allen Konstellationen der Gefährdungshaftung.

Allgemeiner ausgedrückt, besteht zur Gefährdungshaftung der folgende gravierende Unterschied. Gefährdungshaftung setzt ein beim Einsatz gefährlicher Sachen, der aber wegen seines gesellschaftlichen Nutzens erlaubt, also rechtmäßig ist.¹⁵⁹ Gefährdungshaftung ist der seltene Fall einer Schadenshaftung trotz rechtmäßigen Verhaltens der Eigentümer/Nutzer/Betreiber. Gehaftet wird, wenn sich die typische Betriebsgefahr verwirklicht, also wenn Kausalabläufe fehlgelaufen sind. Bei Softwareagenten kommt es aber gerade nicht auf die Sachgefahr eines falsch funktionierenden Computers, also auf das Kausalrisiko, an, sondern auf das Entscheidungsrisiko, auf die andersgeartete Gefahr, dass sich dessen autonome Entscheidungen als Fehlentscheidungen herausstellen. Zurechnungsgrund ist nicht der Einsatz eines Objekts erhöhter Gefahr, sondern das rechtswidrige Verhalten des Algorithmus, den der Prinzipal rechtmäßig zu eigenen Nutzen eingesetzt hat.¹⁶⁰

Weil die „Autonomiegefahr“¹⁶¹ digitaler Entscheidungen schlechterdings nicht mit der Betriebsgefahr der Gefährdungshaftung gleichgesetzt werden kann, müssen hier auch unterschiedliche Verantwortungsprinzipien und unterschiedliche Normen eingreifen. Haftungsnormen für Fehlentscheidungen autonomer Agenten können

¹⁵⁶ § 7 StVG, § 26 AtomG, §§ 1 ff. HPfIG oder §§ 1, 4 ProdHaftG, § 8 BDSG, § 2 UmweltHG

¹⁵⁷ Palandt/Sprau, BGB 77. Aufl. 2018, vor § 823 Rn. 5.

¹⁵⁸ Zu den bei der Gefährdungshaftung bisher aufgetauchten Risiken eingehende Analysen bei Jansen (Fn. 79), 621 ff.

¹⁵⁹ *Locus classicus* Esser, Grundlagen und Entwicklung der Gefährdungshaftung 1941. Zur heutigen juristischen Interpretation Palandt/Sprau, BGB 77. Aufl. 2018, § 823 Einl 11. Zu einer soziologischen Interpretation Luhmann (Fn. 79), 69 f.

¹⁶⁰ Zu dem entscheidenden Unterschied von Gefährdungshaftung und strikter Rechtswidrigkeitshaftung MüKoBGB/Wagner, 7. Aufl. 2017, BGB, § 831, Rn. 4.

¹⁶¹ Janal (Fn. 5), 158.

nicht auf das Kausalrisiko von Sachen gestützt werden, sondern sind auf das Entscheidungsrisiko von Akteuren zuzuschneiden. Auch an dieser Stelle realisiert sich die schon oben angesprochene Funktion der Personifizierung nicht-menschlicher Entitäten, dass sie die Kausalzurechnung durch eine Handlungszurechnung ersetzt.¹⁶² Wenn das Privatrecht, wie hier vorgeschlagen, Softwareagenten als Stellvertreter und als Erfüllungsgehilfen, also als rechtlich handlungsfähig behandelt, dann ist es schlechthin ausgeschlossen, bei außervertraglichen Schädigungen mit einer bloßen Kausalhaftung zu operieren.

Kurz: De lege ferenda sollte in der Tat eine „digitale Assistenzhaftung“, d.h. eine strikte außervertragliche Haftung für Schädigungen durch autonome Softwareagenten, eingeführt werden, aber keinesfalls als Gefährdungshaftung für den rechtmäßigen Einsatz einer gefährlichen Anlage, sondern als *Haftung eines Prinzipals für rechtswidrige Fehlentscheidungen seines Softwareagenten*.¹⁶³

Praktisch relevant wird dieser grundsätzliche Unterschied für zwei Haftungsvoraussetzungen: (1) Anders als bei der Gefährdungshaftung wird für rechtswidriges Verhalten und zwar des Softwareagenten gehaftet. (2) Die Haftung erstreckt sich auf mehr als bloß auf Personen- und Sachschäden. Bisherige Regeln der Gefährdungshaftung anzuwenden, ginge also einerseits zu weit, weil sie auch bei rechtmäßigem Verhalten, das letztlich die Schädigungen verursacht, Haftung auslöst. Andererseits ginge sie nicht weit genug, weil sie regelmäßig nur für Personen- und Sachschäden Ersatz gewährt.

Bei der Gefährdungshaftung wird bekanntlich nicht nach Rechtswidrigkeit gefragt, bei Softwareagenten dagegen wird sie zum Dreh- und Angelpunkt der Haftung.¹⁶⁴ Es kommt entscheidend darauf an, ob die Handlung des Softwareagenten rechtswidrig war. Während es bei der Gefährdungshaftung ausreicht, wenn bloße Kausalität zwischen Betriebsgefahr und Schaden besteht, muss beim Einsatz von Softwareagenten geprüft werden, ob dessen Verhalten rechtswidrig war oder nicht.¹⁶⁵ Besonders bei den sogenannten „Rahmenrechten“, also etwa beim Persönlichkeitsrecht oder beim Recht am Unternehmen, aber auch bei sittenwidrigen oder wettbewerbswidrigen Handeln des Softwareagenten wird der Unterschied relevant, da die Rechtswidrigkeit bekanntlich erst nach einer eingehenden Interessenabwägung festgelegt werden kann. Und hier werden zusätzlich Besonderheiten der Softwareagenten zu berücksichtigen sein. Anfangs wird sich das Rechtswidrigkeitsurteil an den Fähigkeiten menschlicher Akteure orientieren. Das aber ist nur der Minimalstandard, der stets einzuhalten ist. Darüber hinaus können höhere Anforderung an Softwareagenten gestellt werden, wenn sie

¹⁶² Dazu Teubner (Fn. 32).

¹⁶³ Ähnlich Hanisch (Fn. 126, 2014), 46, der statt einer Gefährdungshaftung eine Haftung für „maschinelles Fehlverhalten“ fordert. Ähnlich auch mit dem rechtspolitischen Vorschlag einer Hilfspersonenhaftung für Roboter Müller AJP 2014, 595, 601.

¹⁶⁴ Auch der Entwurf des Europäischen Parlaments verkennt die Besonderheit einer Softwarehaftung im Unterschied zu Gefährdungshaftung, da diese nur Kausalität zwischen schädlichen Verhalten des Computers und dem Schaden verlangt, *Europäisches Parlament* (Fn. 5), Rn. 27. Richtig dagegen Hanisch (Fn. 126, 2014), 46, der maschinelles Fehlverhalten zur Haftungsvoraussetzung macht. Auch in der rechtspolitischen Diskussion, ob „Fahrsysteme“ Rechtssubjektivität verliehen werden soll, werden den Fahrsystemen Verhaltenspflichten auferlegt, also eine Handlungshaftung und keine bloße Kausalhaftung etabliert. Dazu Bodungen/Hoffmann NZV 2015, 521,

¹⁶⁵ So zurecht Janal (Fn. 5), 159; Hanisch (Fn. 126, 2014), 46.

über höhere kognitive Fähigkeiten verfügen.¹⁶⁶ Es wird Aufgabe der Rechtsprechung sein, im Einzelfall konkrete Pflichtenstandards für die Agenten zu definieren.

Der zweite wichtige Unterschied zur Gefährdungshaftung betrifft den Umfang der Haftung. Während Gefährdungshaftungsnormen regelmäßig die Haftung für Vermögensschäden ausschließen oder die Haftung auf nur wenige Rechtsgutsverletzungen beschränken und dies angesichts des jeweils typischen Risikos auch gerechtfertigt ist, sind bei Softwareagenten solche Haftungsbeschränkungen nicht akzeptabel. Denn ihre typischen Risiken verwirklichen sich nicht bloß in den Kontexten, in denen gefährliche Anlagen Unfälle verursachen, sondern in all den Kontexten, in auch Menschen rechtswidrige Entscheidungen fällen. Wie bei schädigenden Handlungen natürlicher Personen müssen auch hier Vermögensschäden ausgeglichen werden, sofern das Handeln der Agenten ein Rechtsgut verletzt oder gegen ein Schutzgesetz verstoßen hat oder sittenwidrig war. Und das macht einen großen Unterschied.

Bei selbstfahrenden Kfz treten sowohl das traditionelle Sachrisiko wie das neuartige Autonomierisiko auf. Entsprechend sind hier Regeln der traditionellen Gefährdungshaftung für die Betriebsgefahr immer anwendbar.¹⁶⁷ Das gilt auch nach der Novelle zum StVG 2017. Doch ist die Problematik dort weniger dringlich, weil die traditionelle Halterhaftung für die Betriebsgefahr des Kfz schon die wichtigsten Risiken abdeckt und die Besonderheiten des Algorithmen-Versagens erst im Regress zwischen Halter und Hersteller virulent werden. Dort scheint aber eine spezielle Haftungsnorm, die das Fehlverhalten des Algorithmus sanktioniert, erforderlich. In anderen Fällen digitaler Autonomie gibt es keine traditionelle Gefährdungshaftung, so dass hier ausschließlich die neuen Regeln anzuwenden sind.

Auf welche Vermögensträger eine solche strikte Haftung für Computer-Fehlentscheidungen dann zuzurechnen sind, dafür sollte nach dem Vorbild der „Zurechnungs- und Haftungseinheit“, in der die Rechtsprechung bei Kfz-Gefährdungshaftung Fahrer, Halter und Versicherung zusammengefasst hat.¹⁶⁸ Betreiber, Hersteller und Programmierer des Softwareagenten sollten zu einem ähnlichen Haftungsverbund zusammengeschlossen werden. Auch der Vorschlag von Hanisch einer abgestuften Haftung zwischen Betreiber und Hersteller, wonach primär der Betreiber haftet und sekundär der Hersteller, ist in diesem Zusammenhang überlegenswert.¹⁶⁹ Eine die Hersteller betreffende Pflichtversicherung für die digitalen Risiken würde Härten für den Betreiber, besonders wenn sie nicht gewerbsmäßig handeln, abfedern.¹⁷⁰ Die Einführung von Höchstsummen entspräche der Versicherungslogik.

De lege ferenda sinnvoll wäre durchaus eine Generalklausel, die eine deliktische Veranlasserhaftung für autonome Algorithmen vorschreibt. *Die Generalklausel müsste nur anordnen, dass für schadensverursachendes Handeln*

¹⁶⁶ Janal (Fn. 5), 159; allgemein für Sorgfaltsstandards: MüKoBGB/Wagner, 7. Aufl. 2017, BGB § 823 Rn. 40: „wenn ein konkreter Akteur über besondere, überdurchschnittliche Kenntnisse und Fähigkeiten verfügt, muss er sie im Interesse der Schadensvermeidung auch einsetzen, um der Haftung zu entgehen.“

¹⁶⁷ Zech (Fn. 5), 185 mwN.

¹⁶⁸ Borges CR 2016, 273, 274, Fn. 23 mwN.; Zech (Fn. 5), 204.

¹⁶⁹ Hanisch (Fn. 126, 2014), 55 ff.

¹⁷⁰ Europäisches Parlament (Fn. 5), Rn. 29.; Horner/Kaulartz (Fn. 97), 26.

*des Softwareagenten der Veranlasser eintreten muss, sofern das Handeln des Softwareagenten rechtswidrig war.*¹⁷¹

Versicherungslösungen, speziell Zwangsversicherungen, stellen ihrerseits eine Änderung des Rechtsstatus der Softwareagenten dar. Sie ordnen eine Haftungsmasse ausschließlich den Handlungen der Softwareagenten zu, der damit zwar nicht zu einer juristischen Person mit Eigenvermögen wird, aber den Rechtsstatus eines Haftungssubjekts mit „Versicherungsvermögen“ erhält. Nur eben dass dieses Vermögen nicht eigentümlich zugeordnet wird, sondern sozialisiert wird, entweder auf die Versichertengemeinschaft oder gar auf die Gesamtgesellschaft.¹⁷² In diesem Sinne verleihen Versicherungslösungen funktionale Äquivalente einer vollen Rechtssubjektivität. Praktisch wichtiger ist: Auch Versicherungslösungen implizieren die Handlungsfähigkeit der Softwareagenten. Wenn Versicherungen für die sogenannte Gefährdungshaftung eintreten, dann wäre es für sie ein nicht kalkulierbares Risiko, wenn sie für jeden kausal verursachten Schaden aufkommen müssten, ohne dass ein digitales Fehlverhalten vorliegt. Sie werden nur im Falle rechtswidrigen Handelns des Agenten Versicherungsschutz zu gewähren bereit sein, besonders dann, wenn die Ersatzpflicht über Personen- und Sachschäden hinausgeht.

Die vielfach geforderte Gefährdungshaftung für Softwareagenten ist also in Wahrheit nichts anderes als eine Variante einer „Gehilfenhaftung“, die notwendig rechtswidriges Verhalten des Agenten voraussetzt und deren Umfang nicht auf Personen- und Sachschäden begrenzt ist. Sie würde - implizit oder explizit – eine beschränkte Rechtsfähigkeit der Softwareagenten anerkennen. Diese Haftung für autonome Entscheidungen der Softwareagenten hätte gewisse Ähnlichkeiten mit dem Haftungsmodell des § 831, jedoch mit dem gravierenden Unterschied, dass dem Geschäftsherrn kein Entlastungsbeweis offensteht und es auf Weisungsgebundenheit nicht ankommt. Ähnlichkeiten bestehen auch mit dem Haftungsmodell des § 31, jedoch entfällt die Beschränkung auf verfassungsmäßig berufene Vertreter.

IV. Hybride: Verbundrisiko

1. Mensch-Maschinen Assoziationen

Als Antwort auf das Autonomierisiko wurden bis hierher dogmatische Lösungen vorgeschlagen, in denen Algorithmen mit beschränkter Rechtsfähigkeit ausgestattet und als eigenständig entscheidende Stellvertreter oder Hilfspersonen behandelt werden. Damit bewegt man sich noch auf einigermaßen sicherem Boden der Rechtsdogmatik. Sehr viel unsicheres Gelände aber betritt man, wenn man den Verbund von Menschen und Algorithmen selbst als Regelungsgegenstand ins Visier nimmt. Der Anlass dafür ist, dass die bisher diskutierte Lösung, den Softwareagenten beschränkte Rechtsfähigkeit zuzuerkennen, ebenso wie die Lösung der traditionellen Dogmatik, das Softwareverhalten einfach als Menschenverhalten zu fingieren, auf eine problematische Alternative hinführt. Einerseits: Ausschließlich den beteiligten Menschen Rechte und Pflichten zuzuschreiben - wie es die traditionelle Dogmatik tut - erfasst die digitalen Risiken nicht ausreichend. Denn es ignoriert die autonome Rolle der Automaten, überspannt beinahe mit Notwendigkeit die Pflichten der

¹⁷¹ Hanisch (Fn. 126, 2014), 46, 54.

¹⁷² Zur sozialversicherungsrechtlichen Lösung Janal (Fn. 5), 157.

beteiligten Menschen und reduziert in Bezug auf Rechte den Schutz der Automaten. Andererseits: Die Handlungen - wie hier vorgeschlagen - ausschließlich den Automaten zuzurechnen, ist auch nicht problemlos, denn man behandelt dann den Beitrag der Menschen im Mensch-Maschinen-Verbund nur in individualrechtlichen Kategorien von Stellvertretung und Vollmacht, was den Eigenheiten des Verbunds selbst nicht gerecht wird. Auch hier bietet wieder Bruno Latours Netzwerk-Aktor-Theorie einen Ausweg. Sie realisiert die Mensch-Maschinen-Interaktion realitätsgerecht als Kollektivphänomen eigener Art: Nach Latour besitzen „Aktanten“ nicht nur eine Sprache und einen widerständigen Körper, sondern auch das Vermögen, „Assoziationen“ zu bilden.¹⁷³ Die Fähigkeit zum gesellschaftlichen Handeln verdanken sie nicht nur ihrem individuellen Status als „Aktanten“, sondern darüber hinaus ihrer Mitgliedschaft in „Hybriden“, d.h. in Assoziationen von menschlichen Akteuren und nicht-menschlichen Aktanten. Damit entsteht – ebenso wie in jeder Assoziation – ein Ressourcenpool, der im Falle der Hybriden die beschränkten Handlungskapazitäten der Aktanten mit den kommunikativen Fertigkeiten realer Menschen kombiniert. „Die psychosystemischen Kompetenzdefizite der Non-Humans werden durch die verteilte Intelligenz der Sozialsysteme zuverlässig kompensiert.“¹⁷⁴ Und es ist gerade die Verbindung von menschlichen und nicht-menschlichen Handlungskapazitäten innerhalb der Hybriden, die es nicht-menschlichen Entitäten erlaubt, auch an komplexeren politischen Verhandlungen, wirtschaftlichen Transaktionen und juristischen Beziehungen teilzunehmen.¹⁷⁵

Ähnlich wie eine formale Organisation entwickelt der Mensch-Maschinen-Verbund selbst eine phänomenale Innenperspektive, eine Selbstwahrnehmung als ein „lebendiger“ Prozess, eine eigene Hierarchie von Präferenzen, eigene soziale Bedürfnisse und politische Interessen, die weder auf die Eigenschaften der beteiligten Menschen noch auf die der Algorithmen reduziert werden können. Die Interaktionen innerhalb des Verbundes machen die Hybride mit körperschaftlichen Akteuren vergleichbar. Die beteiligten Akteure handeln nicht für sich selbst, sondern „für“ den Hybriden als eine emergente Einheit, als Assoziation von Menschen und Nicht-Menschen. Sie tun es in der gleichen Weise wie auch Manager im Unternehmen, die nicht im eigenen Namen handeln, sondern als „Agenten“ in Vertretung für ihren „Prinzipal“, also für das Unternehmen selbst als soziales System.¹⁷⁶ Und dort gibt es Interessen- und Orientierungskonflikte zwischen den Mitgliedern und zwischen ihnen und dem Verbund, wie sie angesichts des bekannten *Agency-Problems* bei korporativen Akteuren diskutiert werde, für deren Lösung zahlreiche Rechtsnormen entwickelt worden sind. Solche Konflikte existieren auf ähnliche Weise auch in den Assoziationen von Menschen und Nicht-Menschen. Und vergleichbare institutionelle Regulierungen – z.B. die Formalisierung der Pflichten und Verantwortlichkeiten des Geschäftsführers, die Ultra-Vires-Lehre, die Prüfung der Repräsentativität bei der Sammelklage – dienen dazu, das *Agency-Problem* einzudämmen, welches auch in den Mensch-Maschinen-Assoziationen auftaucht. In der Tat werden in der Rechtsdogmatik zum Electronic Contracting schon ähnliche

¹⁷³ Latour (Fn. 39), 104.

¹⁷⁴ So Lorentzen, in: Rammert/Schulz-Schaeffer (Hg.), Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik, 2002, 101 ff. 110.

¹⁷⁵ Eingehend zur Kooperation von Menschen und Algorithmen Misselhorn (Fn. 53); Fiebich/Nguyen/Schwarzkopf, in: Misselhorn (Hg.), Collective Agency and Cooperation in Natural and Artificial Systems 2005, 25 ff..

¹⁷⁶ Sie handeln weder für die *shareholders* noch für die *stakeholders*, wie viele Autoren annehmen möchten, sondern ihr Prinzipal ist ausschließlich das Unternehmen, dazu Teubner (Fn. 35).

Lösungsmöglichkeiten bezüglich der „Prinzipal-Agenten-Beziehung“ zwischen dem vertragsschließenden Computer und den menschlichen Vertragsparteien diskutiert.¹⁷⁷

Dies ist eine kollektivrechtliche Sicht, die das Recht aus der nicht unproblematischen individualrechtlichen Alternative befreit, wonach die fragliche Handlung entweder ausschließlich dem Menschen oder ausschließlich dem Algorithmus zurechnen ist. Im Gegensatz zur Individualzurechnung ist sie in der Lage, das emergente Phänomen des Mensch-Maschinen-Verbundes auch rechtlich einzufangen.¹⁷⁸ Denn sie berücksichtigt zum einen die Emergenz in den Innenbeziehungen, die jenseits der Existenz der Binnenakteure die überpersonale Einheit ihres Verbundes erzeugt. Zum anderen wird sie der Emergenz in den Außenbeziehungen gerecht: der Mensch-Algorithmus-Verbund als eigenständiges Handlungssystem tritt selbst mit Dritten in kommunikative Beziehungen und nicht die beteiligten Menschen oder Algorithmen. Den hier auftretenden Risiken, die aufgrund der fast nicht mehr auflösbaren Verflechtung der Einzelhandlungen von Menschen und Algorithmen entstehen, lässt sich besser begegnen, wenn man die Mensch-Algorithmus-Assoziation, den Hybrid als solchen, als gemeinsamen Zurechnungspunkt für Handlungen, Rechte und Pflichten identifiziert.

2. *Hybride als Rechtssubjekte?*

Sollte man dann auch rechtsdogmatisch echte Hybrid-Lösungen erwägen? Sollte man die Rechtskonstruktion eines Mensch-Maschinen-Verbundes einführen? Dies ist eine durchaus ernst zu nehmende, wenn auch bisher noch nicht ausprobierte Alternative. Die einzelnen rechtsgeschäftlichen Handlungen der Softwareagenten und die der beteiligten Menschen würden, ohne beide im Einzelnen trennen zu müssen, zu einer einheitlichen Handlung des Mensch-Maschinen-Verbundes zusammengezogen und würden sowohl Rechtsbindungen als auch Haftungsansprüche erzeugen. Im Gegensatz zum individualistischen Recht der Stellvertretung, das die Einzelhandlungen von Prinzipal und Agent klar voneinander trennt und den Prinzipal zum Vertragspartner erklärt, würde hier der Mensch-Maschinen-Verbund selbst zum eigentlichen Vertragspartner. Für vertragliche und außervertragliche Haftung würden die einzelnen Haftungsvoraussetzungen zusammengesetzt aus ihrer beider Verhalten, ohne dass ihre Einzelbeiträge mühsam und häufig willkürlich auseinandergerechnet werden müssten, wie das in der Haftung für Hilfspersonen notwendig wäre. Eine Wissenszurechnung schließlich fände weder ausschließlich auf den Computer noch ausschließlich auf den Menschen statt, sondern stets auf den Verbund. Auf Vermögen würde haftungsrechtlich zugegriffen, wo immer es sich im Hybriden befindet.

Ja, in einer etwas gewagten Interpretation könnte man sogar die BGH-Rechtsprechung und einen Teil der Lehre, die - wie oben beschrieben - kontrafaktisch behaupten, das Computer-Verhalten sei in Wahrheit nichts anderes als das Verhalten des Menschen,¹⁷⁹ als eine implizite Verbundlösung verstehen. Man muss nur dazu bereit sein, die Formulierung „Mensch“, der als Zurechnungspunkt fungiert, als Chiffre für den Verbund des Hybriden, als Träger der Gesamtheit der menschlichen und digitalen Handlungen, zu interpretieren. Auch der Rechtsschutz

¹⁷⁷ Daniel Santa Clara Computer & High Tech. L. J. 2004, 319, 344 ff.

¹⁷⁸ Kersten (Fn. 20), 4 ff.; Ingold (Fn. 20), 220 ff.

¹⁷⁹ BGHZ 195, 126, Rn. 17.

von Computern gegenüber schädigenden Eingriffen kann implizit als Schutz des Hybriden verstanden werden. Das „Computergrundrecht“ des BVerfG kann man durchaus, wie es Gruber tut, in dieser Richtung interpretieren, da es weder den Computer allein noch das Persönlichkeitsrecht des Nutzers allein, sondern den gesamten Interaktionszusammenhang von Mensch und Computer in seiner Integrität schützt.¹⁸⁰ Man könnte in den durchaus ambivalenten Formulierungen der Rechtsprechung eine Art juristischer Übergangsemantik sehen, die gleitende Übergänge zwischen der Vorstellung traditioneller Individualakteure und der neuartiger Kollektivakteure, eben der Mensch-Maschinen-Verbände, erlaubt. Sie würde es darüber hinaus ermöglichen, den zurzeit beobachtbaren Prozess der Transformation der Algorithmen - von „*inanimate objects*“, über „*life-forces*“, „*animate beings*“, „*graded persons*“ bis zu „*full persons*“¹⁸¹ - auch in juristischer Begrifflichkeit abzubilden. Sie würde übertriebene Kontrastierungen und überflüssige Kontroversen zu vermeiden helfen. Solche Übergangsemantiken hat es im Recht schon immer gegeben. Sie erleichterten stets mehr oder weniger apokryphe Rechtsinnovationen.¹⁸²

Jedoch: die Hybride tatsächlich als rechtliche Einheit zu konstruieren, würde bedeuten, dass das Recht eine völlig neuartige gesellschaftsrechtliche Konstruktion, die Mensch-Maschinen-Assoziation, schaffen müsste.¹⁸³ Dies wäre in der Tat ein wagemutiger - um nicht zu sagen: waghalsiger - Schritt, auf den die heutige Privatrechtsdogmatik kaum vorbereitet ist. Algorithmen als vollwertige Mitglieder einer neuartigen Assoziation? Dieser Schritt wäre noch in einem viel stärkeren Maße eine Rechtsfortbildung extra legem als es der ohnehin schon recht kühne Schritt ist, den Softwareagenten als Repräsentanten ihres menschlichen Prinzipals Teilrechtsfähigkeit zuzuerkennen. Obwohl die kollektivrechtliche Verbundlösung im Vergleich mit der individualrechtlichen Repräsentantenlösung der Realität der Mensch-Maschinen-Interaktion letztlich eher angemessen ist, dürften Rechtsprechung und Lehre es wohl vorziehen, statt dem Verbund sämtliches Verhalten einheitlich zuzurechnen, das Verhalten der einzelnen Partizipanten, also der Einzelmenschen und der Softwareagenten, rechtlich als eine Art Repräsentation individuell Handelnder zu konstruieren.

Wie dem auch sei, selbst wenn es bei der individualrechtlichen Lösung verbleibt, ist auch dann die Einheit des Mensch-Algorithmus-Verbundes juristisch nicht irrelevant; sie wird nun als rechtlich bedeutsame Hintergrundannahme wirksam. Ein interessantes Vorbild, mit einer solchen *unitas duplex* umzugehen, bietet schon das geltende Vertragsrecht mit dem „Vertragszweck“, dem gemeinsamen Zweck der Individualparteien in der Austauschbeziehung des Schuldvertrages. Bekanntlich wird über den Vertragszweck der Vertrag nicht selbst als Kollektiveinheit konstruiert, sondern weiterhin streng individualistisch als

¹⁸⁰ BVerfGE 120, 274, 302. Dazu die subtilen Interpretationen von Gruber (Fn. 15), 131 ff.

¹⁸¹ Eine solche Steigerung von Personifizierungsvorgängen nicht-menschlicher Entitäten beschreibt Sprenger (Fn. 35), 111.

¹⁸² Prominentes historisches Beispiel für eine solche Übergangsemantik bietet das subjektive Recht in seinen Transformationen: von *jus*, über *facultas*, Pflichtrecht, Freiheitsrecht, Rechtsmacht zur Interessenverfolgung bis hin zum sozialstaatlichen Anspruch, dazu Luhmann, Gesellschaftsstruktur und Semantik: Studien zur Wissenssoziologie der modernen Gesellschaft Bd. 2, 1981, 45 ff.

¹⁸³ So die tastenden Überlegungen von Gruber (Fn. 30), 158; Allen/Widdison (Fn. 20), 40 mit Verweis auf die soziale Realität der Hybride. Für eine moralische Verantwortung von Mensch-Maschinen-Verbänden als „*responsibility networks*“ spricht sich Neuhäuser, in: Misselhorn (Hg.), Collective Agency and Cooperation in Natural and Artificial Systems, 2015, 131 ff. 143 f. aus.

Beziehung zwischen den einzelnen Vertragspartnern verstanden. Aber der Vertragszweck, verstanden als der gemeinsame Austauschzweck der Parteien, bringt die übergreifende Einheit der Vertragsbeziehung anders zur Geltung. Er erzeugt konkrete rechtliche Wirkungen – für die Auslegung der Parteierklärungen, für die Normierung von Treupflichten und für die Entscheidung in Streitfällen, ob und in welcher Intensität ein Vertragsbruch vorliegt. Ja, in Dauerschuldverhältnissen, relationalen Verträgen und ganz besonders in Vertragsverbänden, z.B. in Zulieferer- und Vertriebsnetzen, schiebt sich dieser Vertragszweck immer stärker in den Vordergrund – nun unter den Titeln Verbundzweck, Finalnexus oder Netzzweck –, und dies stets, ohne dass gesellschaftsrechtliche Konstruktionen gewählt werden müssen.¹⁸⁴ In ähnlicher Weise müssten die Mensch-Computer-Verbände nicht selbst mit rechtssubjektiven Qualitäten ausgestattet werden, aber ein „Verbundzweck“ könnte ihrem hybriden Charakter gerecht werden, eine rein individualistische Konstruktion korrigieren und die Einheit des Mensch-Maschinen-Verbundes auch rechtlich zur Geltung bringen.

Letztlich sind die individualrechtliche und die kollektivrechtliche Lösung gar nicht als einander ausschließende Alternativen zu verstehen. Sie könnten durchaus sinnvoll nebeneinander bestehen. Die Agenten-Prinzipal-Lösung wäre dann angebracht, wenn die Algorithmen im Sozialleben als klargeschnittene Akteure auftreten. Wenn sie dagegen in einem dichten Interaktionszusammenhang mit menschlichen Akteuren eingebettet sind, dürfte die Verbundlösung vorzuziehen sein. Beide Lösungen hätten jedoch gemeinsam zur notwendigen Voraussetzung, dass sie beschränkte Rechtsfähigkeit an unterschiedliche nicht-menschliche Akteure – an Aktanten oder an Hybride – verleihen müssten.

V. Multi-Agenten-Systeme: Vernetzungsrisiko

Die Rechtstechnik der Personifizierung von Aktanten oder Hybriden gerät jedoch dann an ihre Grenzen, wenn autonome Algorithmen in einem Multi-Agenten-System vernetzt sind. Personifizierung hat ein bestimmbares Substrat noch bei klar identifizierbaren Software-Agenten oder eindeutig abgrenzbaren Mensch-Maschinen-Assoziationen, nicht aber mehr bei komplexen Computervernetzungen.¹⁸⁵ Das hier auftretende Vernetzungsrisiko zerstört Annahmen über Individualität von Akteuren, die für die Zurechnung von Handlung und Verantwortung konstitutiv sind.¹⁸⁶ Sowohl der Handlungsträger als auch die Kausalzusammenhänge sind dann schwierig, wenn nicht unmöglich, aufzuklären.¹⁸⁷

Spiecker zeigt in aller Deutlichkeit die Verantwortungslücke des geltenden Rechts. Ihr durchaus konsequenter Vorschlag einer Pro-Rata-Haftung aller an der Vernetzung Beteiligten, die für das rechtswidrige Verhalten der vernetzten Algorithmen einstehen müssten,¹⁸⁸ aber ist kaum praktikabel. Angesichts der beinahe unendlichen Vernetzung von Algorithmen im digitalen Raum ist eine

¹⁸⁴ Zur neuartigen Kategorie des Netzzwecks: *Grundmann* AcP 2007, 718, 720, 729 ff.; *Teubner*, Netzwerk als Vertragsverbund: Virtuelle Unternehmen, Franchising, Just in Time in sozialwissenschaftlicher und juristischer Sicht, 2004, 148 ff.; *Rohe*, Netzverträge 1998, 78 ff.

¹⁸⁵ *Spiecker* (Fn. 9), 701.

¹⁸⁶ *Beck*, in: *Battaglia/Mukerji/Nida-Rümelin* (Hg.), Rethinking Responsibility in Science and Technology, 2014, 167 ff.; *Floridi/Sanders* (Fn. 8), 205 ff.

¹⁸⁷ *Zech* (Fn. 5), 170.

¹⁸⁸ *Spiecker* (Fn. 9), 703.

Abgrenzung der Haftungssubjekte praktisch unmöglich. Hinzukommen kaum vertretbare Ermittlungskosten: „Linking abstract persons' actions in the information society to their principals may require considerable effort, perhaps at a higher cost than the damage at issue.“¹⁸⁹

Ein Ausweg wäre, rechtliche Verantwortung nicht mehr „Personen“ als Handlungsträgern, sondern nur noch identifizierbaren Handlungen, also einer „anonymen Matrix“ von sozialen und digitalen Prozessen selbst (und nicht auf Strukturen), zuzurechnen. Die eigentlichen Zurechnungspunkte für Verantwortung wären dann autonome Entscheidungen und nicht mehr die Entscheidungsträger. Letztlich wären nicht Menschen, nicht Organisationen, nicht Netzwerke, nicht Softwareagenten, nicht Algorithmenvernetzungen, sondern die Entscheidungen selbst wären in die Verantwortung zu nehmen.

In einer solchen Situation bliebe nichts anderes übrig, als dass das Recht von sich aus autonom abgegrenzte Risiko-Pools zur Eingrenzung dieser Handlungszusammenhänge konstruiert und zwar gerade auch dann, wenn im digitalen Raum keine faktischen Anhaltspunkte für Personifizierung identifizierbar sind.¹⁹⁰ Hier verließ das Recht endgültig die Akteursperspektive, weil es nicht einmal mehr nach Einzel- oder Kollektivakteuren sucht, sondern sich auf die riskanten Entscheidungen als solche konzentriert. Es macht Handlungsketten selbst verantwortlich, ohne sich um deren Eigenschaft als organisierte Willensbildungseinheiten zu kümmern. Der entscheidende Unterschied einer solchen Risikohaftung zu bekannten Formen der Organisationshaftung besteht darin, dass das Haftungsrecht sich nicht mehr auf in der Realität bestehende organisatorische Arrangements oder wenigstens auf Kooperationszusammenhänge bezieht, sondern selbst neuartige Risikonetzwerke definiert, um nicht zu sagen: dekretiert. Und sobald die Handlungen individueller, kollektiver Akteure oder die Kalkulationen digitaler Akteure in einen solchen Raum geraten, würden sie alle zu „Zwangsmitgliedern“ eines solchen Risikopools - also nicht kraft privatautonomer Entscheidung, sondern kraft autoritativer Anordnung des staatlichen Rechts. Freilich gibt es noch einen privatautonomen Anteil am Geschehen, er ist aber auf die bloße Eintrittsentscheidung beschränkt. Sie befinden sich dann in der Kollektivhaftung, ohne Rücksicht auf ihren Willen und ohne Rücksicht auf den Kausalzusammenhang zwischen den von ihnen individuell verantwortbaren Handlungen und dem eingetretenen Schaden.¹⁹¹

Der Risikopool würde nicht mehr durch kooperative, organisatorische oder technische Strukturen bestimmt. Er sollte vielmehr als „digitaler Problembereich“ definiert werden, dessen Grenzen durch die Eignung für kollektives Risikomanagement zu bestimmen sind. Letztlich entscheidend sind weder Kausalzusammenhänge noch vorgegebene kooperative Strukturen - so wichtig beides im Einzelfall auch sein mag -, sondern zentrales Kriterium sollte die Fähigkeit des Pools zum Risikomanagement

¹⁸⁹ *Koops et al.* (Fn. 22), 553.

¹⁹⁰ In diese Richtung, das „*behaviour*“, also die Handlungen und nicht die Handlungsträger, selbst als Zurechnungspunkt zu definieren, denkt anscheinend, wenn auch etwas vorsichtiger, *Hildebrandt* (Fn. 40), 26: „Because the agents may be distributed on and possibly mobile between different hardware applications and because as a multi-agent system it is capable of changing shape (polymorphous), it is not always easy to identify where the emerging agent is located and what is and is not a part of it at any point in time. However, in so far as the emergent behaviours of the system allow its identification as a unity of action, it can be qualified as an agent, whatever the underlying embodiment“.

¹⁹¹ Zu parallelen Überlegungen im Umwelthaftungsrecht *Wagner*, Kollektives Umwelthaftungsrecht auf genossenschaftlicher Grundlage, 1990.

sein. Zugegeben, das wäre "opportunistische" Zurechnung,¹⁹² in diesem Fall aber nicht kausaler, sondern kollektiver Art. Für den Zweck der kollektiven Haftung identifiziert kraft autoritativer Anordnung das Recht konkrete digitale Risikozusammenhänge der Offline- und der Online-Welt mit dem Hintergedanken, ein soziales Gebilde zu schaffen, das diese Risiken einigermaßen präventiv beherrschen oder wenigstens zur Schadensabwicklung herangezogen werden kann.

Dies Risikomanagement betrifft erstens die Abwicklung bereits entstandener Schäden. Das Recht gleicht in Fällen mehrfacher Kausalität die Schäden so aus, dass es einen ausreichenden finanziellen Pool schafft, der die Verluste abdeckt und das Risiko verteilt ("deep pocket", "risk spreading"). Zweitens - und das ist womöglich wichtiger - bedeutet Risikomanagement die kollektive Steuerung zukünftigen Verhaltens, also besonders die Programmierung, aber auch die Einsatzbedingungen der Algorithmen. Das Recht zieht die Grenzen des Risikopools derart, daß eine realistische Basis für eine aktive und gemeinsame Prävention von Risiken in Gebieten geschaffen wird, wo Schädigungsprobleme konzentriert sind. Unter beiden Gesichtspunkten isoliert das Recht den sozialen Bereich kollektiver Verantwortlichkeit so, daß sich eine funktionsfähige digitale Technologie zur Bewältigung der digitalen Vernetzungsrisiken entwickeln kann.

VI. Ergebnis

Die drei geschilderten digitalen Risiken - Autonomierisiko, Verbundrisiko, Vernetzungsrisiko - stellen das Privatrecht vor die Herausforderung, für autonome digitale Informationssysteme einen neuen Rechtsstatus zu bestimmen. Freilich nicht wie es oft gefordert wird, dass sie die volle Personifizierung von Softwareagenten, Mensch-Computer-Assoziationen oder Multi-Agenten-Systemen nahelegen. Als Antwort auf die drei Risiken sollte vielmehr den Algorithmen jeweils ein auf ihre konkrete Rolle sorgfältig kalibrierter Rechtsstatus zuerkannt werden. Für das Autonomierisiko ist es eine adäquate Antwort, den Softwareagenten den Status als teilrechtsfähige Akteure zuzuerkennen. Ihre autonomen Entscheidungen sind mit Rechtsverbindlichkeit auszustatten und falls rechtswidrig sollten sie Haftungsfolgen auslösen. Sie erhalten damit die beschränkte Rechtssubjektivität, als Stellvertreter bindende Verträge für andere abzuschließen. Zugleich sind sie in Fällen vertraglicher und außervertraglicher Haftung als rechtlich handlungsfähige Hilfspersonen anzuerkennen, so dass das maschinelle Fehlverhalten selbst - und nicht bloß das Verhalten der dahinterstehenden Unternehmen - einen Pflichtverstoß darstellt, für den die Unternehmen einstehen müssen.

Während also eine klargeschnittene rechtsdogmatische Antwort auf das Autonomierisiko möglich ist, sind Lösungen für das Verbundrisiko und das Vernetzungsrisiko bisher nur in Konturen erkennbar. Eine denkbare Antwort auf das Verbundrisiko wäre ihr Rechtsstatus als Mitglied einer Mensch-Maschinen-Assoziation. Eine Maximallösung *de lege ferenda* würde den Verbund selbst als rechtliches Zurechnungsobjekt für Handlungen, Rechte und Pflichten vorsehen - eine Lösung, mit der man eine rechtsdogmatische *terra incognita* betreten würde. Eine Minimallösung *de lege lata* dagegen würde den Rechtsbegriff des Verbundzweckes einführen, der rechtsgeschäftliche Erklärungen der Softwareagenten ebenso wie ihre Rechte und Pflichten orientieren könnte. Die Antwort auf das Vernetzungsrisiko

¹⁹² Luhmann (Fn. 79), 129.

schließlich wäre der Status eines vom Haftungsrecht selbst autonom zu definierenden Risikopools, der die Rechtsstellung der Algorithmen im Kontext eines umfassenden digitalen Informationsflusses festlegen und die Haftung des Pools ausschließlich am Ergebnis eines rechtswidrigen Verhaltens des Pools festmachen würde.

Die drei neuen Formen eines digitalen Rechtsstatus für autonome Softwareagenten sind (1) Akteur mit beschränkter Rechtssubjektivität, (2) Mitglied eines Mensch-Maschinen-Verbunds, (3) Teilelement eines Risikopools. Ihre Konkretisierung ist darauf auszurichten, ob und wie die eingangs geschilderten Verantwortungsdefizite bewältigt und Anreize zur Schadensprävention gesetzt werden können. Kernstück des digitalen Rechtsstatus ist, Algorithmen als handlungsfähige Akteure mit beschränkter Rechtsfähigkeit anzuerkennen. Aber zugleich muss ein solcher Rechtsstatus von Softwareagenten es ermöglichen, die Dogmatik der Rechtsgeschäftslehre und des Haftungsrechts konsistent und ohne unhaltbare Fiktionen weiterzuentwickeln, dies freilich nicht als Selbstzweck, sondern um im Falle von Schadensfällen und anderen sozialen Konflikten, die vor das Recht gebracht werden, die Forderung, Gleiches gleich zu behandeln und Ungleiches ungleich, auch im digitalen Raum zu erfüllen.