

**Anna Beckers  
Gunther Teubner  
Digitale Aktanten,  
Hybride,  
Schwärme**

**Drei Haftungsregime  
für künstliche Intelligenz  
suhrkamp taschenbuch  
wissenschaft**

suhrkamp taschenbuch  
wissenschaft 2444

Wer trägt das Risiko, wenn künstlicher Intelligenz – etwa ChatGPT – schadensträchtige Fehler unterlaufen? Wenn die beteiligten Menschen sorgfältig gehandelt haben, haftet nach geltendem Recht – niemand. Um dieser gravierenden Verantwortungslücke zu begegnen, entwerfen Anna Beckers und Gunther Teubner drei rechtliche Haftungsregime, für die sie Erkenntnisse aus der Soziologie sowie der Moral- und der Technikphilosophie heranziehen: Prinzipal-Agenten-Haftung für Handlungen autonomer Software-Agenten (»Aktanten«), Netzwerkhafung für verdichtete Mensch-KI-Interaktionen (»Hybride«) und fondsbasierte Entschädigung für vernetzte KI-Systeme (»Schwärme«). Ein bahnbrechender Lösungsvorschlag für eine hochaktuelle Problematik.

Anna Beckers ist Professorin für Privatrecht und Gesellschaftstheorie an der Universität Maastricht.

Gunther Teubner ist emeritierter Professor für Privatrecht und Rechtssoziologie an der Goethe-Universität Frankfurt am Main. Zuletzt erschienen: *Regime-Kollisionen. Zur Fragmentierung des globalen Rechts* (stw 1803, zus. mit Andreas Fischer-Lescano) und *Verfassungsfragmente. Gesellschaftlicher Konstitutionalismus in der Globalisierung* (stw 2028).

Anna Beckers  
Gunther Teubner  
Digitale Aktanten,  
Hybride,  
Schwärme

*Drei Haftungsregime  
für künstliche Intelligenz*

Suhrkamp

Die vorliegende Publikation ist eine stark überarbeitete  
und aktualisierte Übersetzung von *Three Liability Regimes for Artificial  
Intelligence* (Hart Publishing, 2022).



Erste Auflage 2024

suhrkamp taschenbuch wissenschaft 2444

Originalausgabe

© Suhrkamp Verlag AG, Berlin, 2024

Alle Rechte vorbehalten. Wir behalten uns auch  
eine Nutzung des Werks für Text und Data Mining

im Sinne von § 44b UrhG vor.

Umschlag nach Entwürfen

von Willy Fleckhaus und Rolf Staudt

Druck und Bindung: C. H. Beck, Nördlingen

Printed in Germany

ISBN 978-3-518-30044-2

[www.suhrkamp.de](http://www.suhrkamp.de)

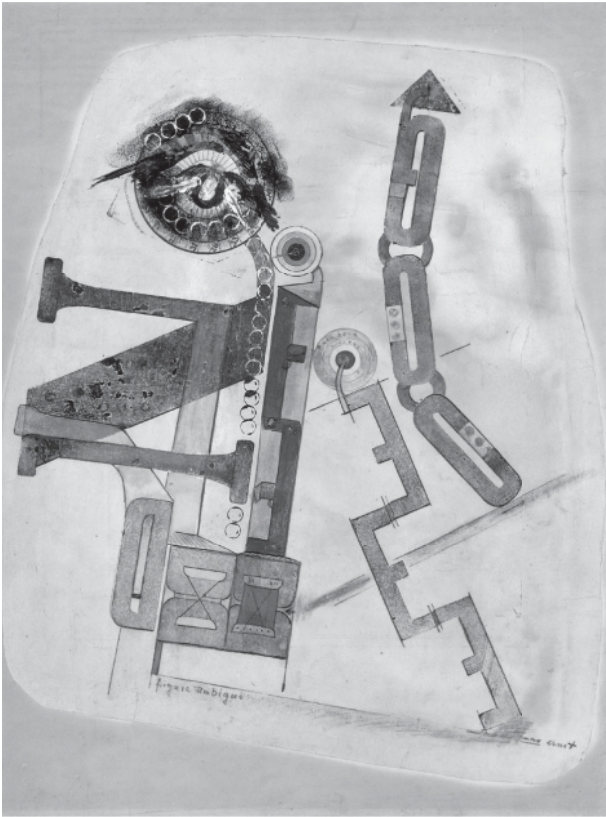


Abbildung 1: *Figure ambiguë*, Bleistiftdurchreibung, Tuschkfeder, Aquarell und Gouache, Collage von Max Ernst, um 1919/1920, 44 x 33 cm, Nationalgalerie, Staatliche Museen zu Berlin, Sammlung Scharf-Gerstenberg, Foto: © bpk/Nationalgalerie, SMB, Sammlung Scharf-Gerstenberg/Jörg P. Anders, © VG Bild-Kunst, Bonn 2023.



# Inhalt

Vorwort .....	II
Kapitel 1 – Digitalisierung: Emergente Risiken .....	15
I. Verantwortungslücken im geltenden Recht .....	15
1. Szenarien .....	21
2. Realitätsverweigerung in der Privatrechtsdogmatik ..	25
II. Kritik der wichtigsten Reformvorschläge .....	28
1. Verschuldenshaftung .....	28
2. Gefährdungshaftung und Produkthaftung .....	30
3. Eigenhaftung einer E-Person .....	34
4. Falsche Einheitslösungen .....	40
5. (Rechts-)Form folgt (Sozial-)Funktion .....	44
III. Sozio-digitale Institutionen .....	46
1. Intermediäre zwischen Technik und Recht .....	46
2. Typologie des Maschinenverhaltens .....	54
3. Typologie sozio-digitaler Institutionen .....	55
4. Typologie der Rechtsrisiken .....	59
Kapitel 2 – Aktanten: Autonomierisiko .....	64
I. Sozio-digitale Institution: Digitale Assistenz .....	64
1. Künstliche Intelligenz als Akteur? .....	64
2. Algorithmen als Aktanten .....	69
3. Kommunikation mit Algorithmen .....	72
4. Gradualisierte digitale Autonomie .....	82
5. Rechtliche Autonomiekriterien .....	83
6. Autonome Entscheidungen der Softwareagenten ...	88
7. Sozialer Akteurstatus und juristische Personifizierung	91
II. Risiken digitaler Assistenz .....	94
1. Das allgemeine Risiko algorithmischer Autonomie ..	94
2. Spezielle Risiken digitaler Assistenz .....	98
III. Das Recht des algorithmischen Vertrags .....	103
1. Ungültigkeit algorithmischer Verträge? .....	104
2. Algorithmen als bloße Werkzeuge? .....	106
3. Unsere Lösung: Elektronische Agenten als Stellvertreter .....	112
4. Teilrechtsfähigkeit – Konstellation I .....	116
5. Äquivalente für subjektive Tatbestandsmerkmale ...	118



6. Äquivalente zur Vollmacht .....	122
7. Haftungsprobleme bei unbefugtem Handeln des Agenten? .....	123
IV. Recht der vertraglichen Haftung .....	127
1. Das Dilemma der Werkzeuglösung .....	127
2. Unsere Lösung: Haftung für algorithmische Erfüllungsgehilfen .....	132
3. Teilrechtsfähigkeit – Konstellation II .....	135
V. Recht der außervertraglichen Haftung .....	141
1. Verschuldenshaftung? .....	142
2. Produkthaftung? .....	145
3. Gefährdungshaftung? .....	151
4. Unsere Lösung: Digitale Assistenzhaftung .....	154
5. Teilrechtsfähigkeit – Konstellation III .....	164
6. Der <i>reasonable algorithm</i> .....	164
7. Haftende Akteure .....	167
Kapitel 3 – Hybride: Verbundrisiko .....	170
I. Sozio-digitale Institution: Mensch-Algorithmus-Assoziation .....	171
1. <i>Collective moral autonomy</i> .....	171
2. Emergenz von Hybridität .....	176
3. Die Organisationsanalogie .....	180
II. Verbundrisiko der Hybriden .....	182
III. Unser Vorschlag <i>de lege ferenda</i> : vollwertige Rechtssubjekte .....	184
IV. Unser Vorschlag <i>de lege lata</i> : Netzwerkhafung .....	188
1. Mensch-Algorithmus-Interaktionen als Netzwerke ..	188
2. Netzwerkhafung .....	191
3. Handlungs- und Haftungszurechnung bei Hybriden ..	196
4. Haftende Akteure .....	197
5. Anteilige Netzwerkhafung .....	200
6. Externe Haftungskonzentration: Der <i>One-stop-shop</i> -Ansatz .....	204
7. Interne Haftungsverteilung: Pro-Rata-Netzanteil ...	206
8. Gesamtstruktur der Netzwerkhafung .....	207
Kapitel 4 – Schwärme: Vernetzungsrisiko .....	208
I. Sozio-digitale Institution: Digitale Vernetzung .....	208
1. Nichtkommunikative Kontakte .....	210
2. Distribuierte kognitive Prozesse .....	214

II. Vernetzungsrisiko der digitalen Schwärme .....	218
III. Assistenten- oder Kollektivhaftung für Vernetzungsrisiken?	223
1. Anwendung bestehender Haftungsregime .....	223
2. Assistenten- oder Produkthaftung? .....	225
3. Kollektivhaftung? .....	227
IV. Unsere Lösung:	
Vergesellschaftung des Vernetzungsrisikos .....	232
1. Vollständige oder teilweise Vergesellschaftung? .....	232
2. Rechtlich definierte Risikopools .....	234
3. Verwaltung des Risikopools: Die Fondslösung .....	236
4. Finanzierung des Fonds: <i>Ex-ante</i> - und <i>Ex-post</i> -Komponenten .....	238
5. Beteiligung und Verwaltung .....	240
6. Entschädigung und Wiedergutmachung .....	244
7. Globale Vernetzung und nationale Regulierung .....	248
Kapitel 5 – Drei Haftungsregime: Anwendungsbereiche, Konkurrenzen, exemplarische Fälle, Ausblick .....	251
I. Synopsis haftungsrechtlicher Normen .....	251
II. Maschinenverhalten, sozio-digitale Institutionen und Haftungsrecht .....	253
1. Unterschiede sozio-digitaler Institutionen .....	253
2. Unterschiedliche Behandlung der haftenden Akteure	262
3. Unterschiedlicher Rechtsstatus von Algorithmen ...	269
III. Konkurrenzen zwischen den Haftungsregimen .....	270
1. Abgrenzungen .....	270
2. Vorrangregeln? .....	273
IV. Exemplarische Fälle .....	276
1. Prinzipal-Agenten-Haftung: <i>Robo-advice</i> .....	277
2. Netzwerkhaftung: Hybrider Journalismus .....	279
3. Kollektivfonds: Flash Crash .....	282
4. Google <i>autocomplete</i> .....	285
5. Generative KI (ChatGPT) .....	289
V. Ausblick: Haftungsrecht und Digitalverfassung .....	293
Bibliographie .....	297
Sachindex .....	336



## Vorwort

In diesem Buch schlagen wir drei Haftungsregime für das Fehlverhalten künstlicher Intelligenz vor. Sie sollen die erheblichen Verantwortungslücken, welche die Ankunft autonomer Algorithmen in der gesellschaftlichen Praxis aufgerissen hat, beseitigen. Die drei Regime sind: eine Prinzipal-Agenten-Haftung für Handlungen autonomer Softwareagenten (»Aktanten«), eine Netzwerkhafung für verdichtete Mensch-KI-Interaktionen (»Hybride«) und eine Haftung kollektiver Fonds für vernetzte KI-Systeme (»Schwärme«). Die Haftungsregime sind sorgfältig abgestimmte Reaktionen auf Verantwortungslücken unterschiedlicher Qualität. Statt einer übergeneralisierten Einheitshaftung oder einer untergeneralisierten sektoralen Haftung schlagen wir vor, auf drei typische Risiken der künstlichen Intelligenz – autonome KI-Entscheidungen, Menschen-KI-Kollektive und systemische Vernetzung – mit je unterschiedlichen Haftungsregimen zu reagieren.

Methodisch gehen wir durchweg interdisziplinär vor, um die vielfältigen Wechselbeziehungen zwischen Informationstechnologie und Haftungsrecht zu erfassen. Den fatalen Technik-Recht-Kurzschluss, der voreilig IT-Strukturen direkt mit Haftungsregeln zusammenschließt, suchen wir zu vermeiden. Zu diesem Zweck führen wir das intermediäre Konzept der »sozio-digitalen Institutionen« ein. Denn als technische Artefakte haben Algorithmen nicht schon die Eigenschaften von Akteuren, die mit Menschen kommunizieren. Erst die Art und Weise, in der die gesellschaftliche Praxis die sogenannten Affordanzen, das heißt die Nutzungsangebote digitaler Technik, annimmt, entscheidet über den gesellschaftlichen Status von Algorithmen. Genauer: Sozio-digitale Institutionen konstituieren Algorithmen entweder als Akteure oder sie ordnen ihnen einen anderen Sozialstatus zu. Um diese neuen Institutionen genauer zu verstehen, rekurren wir auf sozialwissenschaftliche Analysen, welche die notwendigen Vermittlungsleistungen zwischen IT-Wissenschaften und Jurisprudenz erbringen können. Sie erst dürften in der Lage sein, die unterschiedlichen sozio-technischen Situationen, in denen KI-Systeme gesellschaftlich genutzt werden, zu identifizieren und zugleich die dabei auftretenden ge-

sellschaftlichen Risiken herauszuarbeiten, auf die das Haftungsrecht reagieren sollte.

Unser Ansatz unterscheidet sich daher von den heute verbreiteten rechtsökonomischen Analysen, die ihren Fokus ausschließlich auf monetäre Kosten und Nutzen von Haftungsnormen einstellen. Stattdessen suchen wir zugleich Erkenntnisse der Soziologie, der Moralphilosophie und der Technikphilosophie zu integrieren. Deren Einsichten sind besonders relevant für eine etwaige Personifizierung von Algorithmen, für emergente Eigenschaften von Mensch-Algorithmus-Assoziationen und für verteilte Kognition in vernetzten KI-Systemen.

Wir unterscheiden uns von den üblichen rechtsbereichsspezifischen Analysen, die im Ergebnis autonomen Algorithmen in der Rechtsgeschäftslehre, in der vertraglichen Haftung und in der außervertraglichen Haftung einen je unterschiedlichen Rechtsstatus zuweisen, was unter Gleichbehandlungsaspekten problematisch ist. Demgegenüber versuchen wir, den Rechtsstatus konsistent in allen drei Rechtsgebieten zu bestimmen und ihn zugleich in Übereinstimmung mit ihrem Sozialstatus zu halten.

Da das Haftungsrecht nach wie vor entlang nationaler Grenzen fragmentiert ist, gehen wir in unserer Untersuchung auch rechtsvergleichend vor. Wir konzentrieren uns auf die kontinentaleuropäischen Rechtsordnungen, besonders auf das deutsche Recht, sowie auf die Welt des Common Law, besonders auf das US-amerikanische und das englische Recht. Wo immer relevant, beziehen wir zugleich die europarechtliche Dimension des Themas ein. Wir folgen einer Methode, die Hugh Collins als »vergleichende soziologische Jurisprudenz«<sup>1</sup> bezeichnet hat. Soziologische Jurisprudenz analysiert sozio-digitale Institutionen und ihre Risiken, um die relevanten Rechtsprobleme identifizieren zu können. Rechtsvergleichung nutzt diese Qualifizierung mit Blick auf unterschiedliche Rechtssysteme und berücksichtigt die Besonderheiten der nationalen Rechtsdogmatik.

Das Buch hat von intensiven Diskussionen mit vielen Kollegen profitiert. Unser Dank gilt insbesondere Marc Amstutz, Alfons Bora, Carmela Camardi, Ricardo Campos, Elena Esposito, Pasquale Femia, Andreas Fischer-Lescano, Malte Gruber, Albert Ingold,

1 Collins (2009), »Comparative Sociological Jurisprudence«, S. 253 ff.

Günter Küppers, Dimitrios Linardatos, Martin Schmidt-Kessel, Juliano Maranhão, Marc Mölders, Michael Monterossi, Daniel On, Oren Perez, Valentin Rauer, Jan-Erik Schirmer, Thomas Vesting, Gerhard Wagner, Dan Wielsch und Rudolf Wiethölter. Anna Huber und Dirk Hildebrandt haben uns ihr kunsthistorisches Fachwissen über Max Ernst und über das Bild der *Figure ambiguë*, das wir zum Einstieg in das Buch gewählt haben, zur Verfügung gestellt. Jan-Erik Strasser und Nikolai Ballast danken wir für die sorgfältige Lektorierung.

Anna Beckers  
Gunther Teubner



# Kapitel 1 – Digitalisierung: Emergente Risiken

## I. Verantwortungslücken im geltenden Recht

*Figure ambiguë* – Max Ernsts surrealistischer (Alb-)Traum aus der Zeit unmittelbar nach dem Ende des Ersten Weltkriegs scheint heute Realität zu werden. In dem 1919 entstandenen Kunstwerk – abgebildet ganz zu Anfang dieses Buchs – versucht Ernst offenbar, den exzessiven Ambivalenzen der modernen Technik Ausdruck zu verleihen. Als einer der Protagonisten des Dadaismus und Surrealismus war er fasziniert von der Dynamik der Maschinenutopie und zugleich abgestoßen von ihren inhumanen Folgen. Auf dem rechten Bildteil findet sich eine leichte, fröhliche Stimmung, die die genialen Erfindungen der modernen Wissenschaft symbolisieren dürfte. Buchstaben sind in komplexen Anordnungen miteinander verschlungen und scheinen sich zunächst in seltsame Maschinen zu transformieren. Nach Ralph Ubls Interpretation verwandeln sich bei Ernst solche Figuren durch Metamorphose oder Verdopplung ihrer Identität dann wiederum in menschliche Körper, die zu springen, zu tanzen und zu fliegen scheinen. Diese *homines ex machina* »vollführen [...] geradezu einen Triumph der Mobilität: Rotation, Verdopplung, Verschiebung, Spiegelung und optische Täuschung«.<sup>1</sup>

Ganz anders ist die Atmosphäre auf dem linken Teil des Bildes. Die Symbole verändern ihre Farbe ins Düstere, wirken nun brutal und bedrohlich. In der linken oberen Ecke wirft eine schwarze Sonne, die ihrerseits aus seltsamen Symbolen zusammengesetzt ist, aus ihrem unheimlichen Antlitz heraus ein dunkles Licht über die Welt. Mit diesem und vielen anderen Bildern drückte Ernst unmittelbar nach dem Ersten Weltkrieg seine gespaltene Haltung zur modernen perfekten Maschinenwelt aus, deren heitere Logik, Rationalität und Ästhetik ins Absurde, Irrationale und Brutale umzuschlagen droht.<sup>2</sup> Ernst suchte

nach Wegen, gesellschaftliche Mechanismen und Wahrheiten zu registrieren und auch deren tieferliegende Strukturen »einzufangen« oder bildlich zu

1 Ubl (2004), *Prähistorische Zukunft*, S. 30.

2 Becchetti (2020), »Max Ernst: Il surrealista psicoanalitico«; Adamowicz (2019), *Dada Bodies*, Kap. 4 und 8.



verdeutlichen. Hier geht es wahrscheinlich im Kern um das Ertasten eines gesellschaftlichen Unbewussten im historischen Moment, in dem das totalitäre Potenzial der Technik als gesellschaftlichem Motor erahnbar wird.<sup>3</sup>

Ernsts surrealistischer (Alb-)Traum scheint wie gesagt heute zur Realität zu werden. Autonome Algorithmen sind die emblematischen, mehrdeutigen Figuren unserer Zeit, die als rätselhafte autonome künstliche Intelligenz die Ambivalenz der Automaten noch radikalisieren. Wie die Buchstaben in Ernsts Bild sind Algorithmen auf den ersten Blick nichts als unschuldige Symbolverkettungen. Aber in ihrer Metamorphose in elektronische Impulse beginnen auch die Algorithmen zu leben, zu springen, zu tanzen, zu fliegen. Mehr noch: Sie bringen mit künstlicher Intelligenz eine neue Sinn-dimension in die Welt. Ihre *creatio ex nihilo* verspricht eine bessere Zukunft für die Menschen. Selbstlernende Algorithmen, Big Data und generative Künstliche Intelligenz (etwa ChatGPT) nähren die Hoffnung auf algorithmische Kreativität, welche die kognitiven Fähigkeiten des menschlichen Geistes immens zu erweitern verspricht.

Doch ist dies nur die helle Seite ihrer exzessiven Ambivalenz. Es gibt eine bedrohliche dunkle Seite der schönen neuen Welt der Algorithmen. Nach der ersten Phase der Euphorie werden Algorithmen heute häufig als albtraumhafte Monster wahrgenommen. Nick Bostrom beschwört eine »perverse Instantiierung«, die entsteht, wenn intelligente Maschinen menschlicher Kontrolle entgleiten: Schon ein einzelner Algorithmus, der mit äußerster Effizienz das ihm von Menschen gesetzte Ziel verfolgt, ist in der Lage, in der Mittelwahl die menschlichen Intentionen zu pervertieren.<sup>4</sup> Darüber hinaus entsteht eine seltsame Hybridität, wenn Menschen und Algorithmen nicht nur miteinander kommunizieren, sondern sich zu neuartigen Kollektivakteuren, zu Quasi-Organisationen mit ungeahnten, potenziell schädlichen Eigenschaften zusammenschließen.<sup>5</sup> Schließlich entsteht eine besonders bedrohliche Situation, wie es Ernsts schwarze Sonne symbolisieren könnte, wenn Menschen einer undurchsichtigen Umwelt algorithmischer Vernetzung ausgesetzt sind, die für sie unkontrollierbar bleibt.<sup>6</sup>

3 So interpretiert in einem Brief an uns die Kunsthistorikerin Anna Huber das Werk von Max Ernst.

4 Bostrom (2014), *Superintelligenz*, S. 171 ff.

5 Hendrycks u. a. (2023), »Catastrophic AI Risks«, S. 25 ff.

6 Coeckelbergh (2012), »Experiences of the Tragic«, S. 37.

Wie geht das heutige Recht mit dieser Ambivalenz von Algorithmen um? Dieser Frage gehen wir am Beispiel der Haftung für algorithmisches Versagen nach. Das Recht spiegelt durchaus die Ambivalenz der Algorithmenwelt wider. Auf ihrer hellen Seite behandelt das Recht Algorithmen als willkommene Instrumente, die in den Dienst menschlicher Bedürfnisse gestellt werden. Der Versuchung, angesichts ihrer beträchtlichen Schädigungsrisiken autonome Algorithmen einfach zu verbieten, hat es bisher erfolgreich widerstanden. Im Gegenteil, das Privatrecht stärkt sogar die Macht der Algorithmen, indem es ihnen eine quasi magische »*potestas vicaria*«<sup>7</sup> verleiht, sodass sie als autonome Agenten Verträge mit Dritten mit zugleich bindender Kraft für ihre Prinzipale abschließen und eigenständig durchführen können. Auf ihre Schattenseite jedoch reagiert das bis heute geltende Recht nur defizitär. Das aktuelle Haftungsrecht ist nicht darauf vorbereitet, die neuen Gefahren autonomer Algorithmen zu identifizieren, geschweige denn ihnen zu begegnen. Der weit überwiegende Teil der rechtswissenschaftlichen Literatur behandelt autonome Algorithmen mit kaum zu überbietender Schlichtheit als mechanische Werkzeuge, Maschinen, Objekte oder Produkte. Wenn sie Schäden verursachen, verlässt man sich darauf, dass das geltende Recht, besonders das Recht der Produkthaftung, die angemessenen Reaktionen schon bereithalte.

Doch das ist zu einfach gedacht. Im Vergleich zu den vertrauten Situationen der Produkthaftung steigert sich das Ausmaß potenzieller Schäden, sobald das Produkt eine neue Qualität aufweist – Intelligenz.<sup>8</sup> Genau an dieser Stelle reißen aber die neuen Verantwortungslücken auf.<sup>9</sup> Die *figures ambiguës*, die in die Räume

7 Damit ist die Macht gemeint, die eine normsetzende Instanz dem Repräsentanten über die Repräsentierten verleiht. Zu den theologischen Ursprüngen des Vikariatsverhältnisses vgl. Agamben (2011), *The Kingdom and the Glory*, S.138f. Zu einer eingehenden interdisziplinären Analyse dieser *potestas vicaria* siehe Trüstedt (2022), *Stellvertretung*, insbesondere für algorithmisches Handeln, S.375ff. Dass die herrschende Lehre im Privatrecht diese realexistierende *potestas vicaria* der Algorithmen hinter der Vorstellung eines bloßen Werkzeuges versteckt, steht auf einem anderen Blatt.

8 Rachum-Twaig (2020), »Whose Robot«, S.1149.

9 Verantwortung wird hier vorläufig in einem sehr weiten Sinne verstanden. Zur Unterscheidung des soziologischen Begriffs der Verantwortung und des juristischen Begriffs der Verantwortlichkeit siehe Kap. 2, I.6.

des Privatrechts eindringen, sind nicht einfach Objekte, sondern autonome Quasi-Subjekte – generative KI (etwa ChatGPT), Hochgeschwindigkeits-Handelsalgorithmen, Roboter, Softwareagenten, Cyborgs, Hybride, Computernetzwerke. Einige davon sind mit einem hohen Maß an Autonomie und der Fähigkeit zu lernen ausgestattet. Mit ihrer rastlosen Energie erzeugen sie bisher unbekannte Gefahren für Mensch und Gesellschaft. Auf diese jedoch ist das Privatrecht, auch das Recht der Produkthaftung, nicht vorbereitet.

So absurd es klingt: Wenn autonom agierende Algorithmen Fehlentscheidungen treffen und Schäden verursachen, kann, sofern den beteiligten Menschen kein Fehlverhalten vorzuwerfen ist, nach dem zurzeit geltenden Recht niemand verantwortlich und haftbar gemacht werden! Dieser verblüffende Befund ist auf die traditionellen Zurechnungstechniken des Privatrechts zurückzuführen, die das Computerverhalten stets als Verhalten der dahinterstehenden Menschen ausgeben müssen. Für den Fall, dass intelligente Maschinen selbst folgenreiche Entscheidungen treffen, hat das Recht keine Begriffe entwickelt. Softwareagenten können rechtlich nur als bloße Maschinen, als willige Werkzeuge in den Händen ihrer menschlichen Herren, behandelt werden.<sup>10</sup> Für reines Maschinenversagen aber wird, wenn den beteiligten Menschen keine Pflichtverletzung nachzuweisen ist, nach geltendem Recht nicht gehaftet.<sup>11</sup> Das gilt auch für das Versagen von Algorithmen: »Soweit dem Unternehmen kein eigenes Auswahl-, Wartungs- und Überwachungsver schulden nachgewiesen werden kann, haftet für Fehlfunktionen des digitalen Systems – niemand.«<sup>12</sup>

Autonome Algorithmen jedoch wollen sich den jeweiligen strikten Alternativen von Mensch oder Maschine, Subjekt oder Objekt, Person oder Sache nicht fügen.<sup>13</sup> Die »*active digital agency*« der Al-

10 So dezidiert der Bundesgerichtshof BGHZ 195, 126, Rn. 17; für das US-Recht Restatement (Third) of Agency Law § 1.04 cmt. e. (2006).

11 BGHZ 54, 332, 333 ff. (Versagen einer Ampel); OLG Frankfurt NJW 2018, 637, Rn. 17; Körber/König (2020), »Haftungsrecht 4.0«, S. 271; Spindler (2016), »Digitale Wirtschaft«, S. 816; Medicus/Lorenz (2015), *Schuldrecht AT*, Rn. 386, S. 170 f.; Caspers in: Staudinger-BGB (2019), § 278, Rn. 5; Grundmann, in: MüKo-BGB (9. Auflage, 2022), § 278, Rn. 46; differenzierend Pfeiffer, in: Soergel-BGB (13. Auflage, 2014), § 278, Rn. 25.

12 Wagner (2020), »Digitale Techniken«, S. 736; ebenso Wagner/Luyken (2020), »Robo Advice«, S. 173

13 Hier beziehen wir uns auf digitale Autonomie in einem eher umgangssprach-

gorithmen verursacht Probleme, auf die das Privatrecht mit neuer Begrifflichkeit reagieren muss:

Je autonomer die Roboter werden, desto weniger verstehen sie sich als bloße Werkzeuge in der Hand des Menschen und desto mehr gewinnen sie aktive digitale Handlungsfähigkeit. In diesem Zusammenhang werden Fragen der Verantwortung und Haftung für ihr Verhalten und mögliche Schäden, die ihr Verhalten verursacht, relevant.<sup>14</sup>

Letztlich ausschlaggebend dafür, dass das Privatrecht unter massivem Änderungsdruck steht, sind also inakzeptable Verantwortungslücken, welche die rasanten digitalen Entwicklungen schon heute aufreißen.<sup>15</sup> Wie der Soziologe Sven Kette konstatiert, läuft die (teilweise) Ersetzung personaler Entscheidungsbeiträge durch Algorithmen auf einen Adressausfall bezüglich der Zurechenbarkeit von Verantwortlichkeiten hinaus.<sup>16</sup> Softwareagenten und andere KI-Systeme verursachen zwangsläufig diese Verantwortungslücken, sofern ihre Handlungen unvorhersehbar sind. Das zieht einen massiven Kontrollverlust menschlicher Akteure nach sich. Nach geltendem Recht brauchen die Verursacher, wenn sie ihre Pflichten erfüllt haben, die Schäden nicht zu ersetzen; sie müssen von den Opfern getragen werden. Gleichzeitig aber wird die Gesellschaft in immer größerem Umfang von autonomen Algorithmen abhängig, sodass ein Verzicht auf ihren Einsatz äußerst unwahrscheinlich ist.<sup>17</sup>

Nicht umsonst warnt der Philosoph Christian List vor solchen Verantwortungslücken, wenn die Gesellschaft es ohne weiteres zulässt, künstliche Intelligenz einzusetzen. Denn dies schaffe

eine Situation, in der sich Einzelpersonen oder Unternehmen der Haftung für hochriskante Entscheidungen entziehen können, indem sie diese Entscheidungen an KI-Systeme delegieren und sich dann hinter der Auto-

lichen Sinne. Unten werden wir ihre genaue Bedeutung im Haftungsrecht ausführlich diskutieren, insbesondere in Kap. 2, I.4-6.

14 van Dijk (2020), »Hall of Masks«, S. 231. Den Begriff »*active digital agency*« hat Clarke (1994), »Digital Persona« eingeführt.

15 Zu den verschiedenen Dimensionen digitaler Verantwortungslücken Duffourc (2023), »Autonomous AI Physician«, S.17 ff.; De Sio/Mecacci (2021), »Four Responsibility Gaps with Artificial Intelligence«; Dyrkolbotn (2017), »Robot Harms«, S. 121 f.

16 Kette (2021), »Computer says No?«, S. 175 f.

17 Matthias (2010), *Automaten als Träger von Rechten*, S. 15.