Anna Beckers & Gunther Teubner

Mensch-Algorithmus-Hybride als (Quasi-)Organisationen? Zu Verantwortung und Verantwortlichkeit von digitalen Kollektivakteuren*

Zusammenfassung: Modelle individueller Verantwortlichkeit für die Handlungen von Algorithmen versagen dann, wenn eine Mensch-Algorithmus-Assoziation selbst als Handlungseinheit in den Blick gerät. In manchen Situationen sind menschliche und algorithmische Handlungen derart eng miteinander verwoben, dass keine lineare Verbindung zwischen den emergenten Kollektiven und den komplexen Interaktionen der individuellen Handlungseinheiten mehr besteht. In solchen kollektiven Entscheidungssequenzen lässt sich auch keine individuelle Verantwortlichkeit mehr zurechnen. Deshalb bedarf es einer neuen Perspektive auf Mensch-Algorithmus-Assoziationen, die ihre emergenten Eigenschaften und ihren organisationalen Charakter erfasst, um angemessene Modelle kollektiver Verantwortlichkeit zu entwickeln.

Der Artikel sucht Antworten auf die folgenden Fragen: Wie kann die Begegnung zwischen Mensch und Algorithmus innerhalb einer solchen sozio-technischen Konfiguration theoretisch angemessen erfasst werden? Lässt sich die Konfiguration selbst als hybrides Kollektiv verstehen? Können der Konfiguration selbst als personifiziertem Kollektivakteur Handlungen zugerechnet werden? Welche Verantwortlichkeitsformen dürften für Mensch-Algorithmus-Assoziationen institutionalisiert werden – zentralisierte oder distribuierte Kollektivverantwortlichkeiten?

Abstract: Models of individualized responsibility for algorithms' actions fail once one needs to focus on the human-algorithm-associations itself as the unit of action. In some situations, human action and machine action are so closely intertwined that there is no linear connection between the emergent collectives and the complex interactions of the individuals from which they arise. In joint decision-making, the individual responsibility of humans or algorithms cannot be identified. Thus, a new perspective on human-algorithm associations is needed that recognizes their emergent properties, their distinct organizational character, and that develops appropriate models for a new type of collective accountability.

The article addresses the following problems: How can the human-algorithm-encounter – within this socio-technical configuration – be theorized? Can the configuration be envisaged as a hybrid collective? Can it be perceived of in terms of collective agency? Which type of collective accountability will be appropriate for human-machine associations – centralized or distributed responsibility?

I. Problemkonstellationen

In bestimmten Situationen des Computereinsatzes verdichten sich Kommunikationen zwischen Menschen und Algorithmen derart, dass einfache Interaktionssysteme in Quasi-Organisationen transformiert werden.[1] Ein

* Für Kritik und Anregungen danken wir Elena Esposito, Veronika Tacke und einem anonymen Gutachter.

einschlägiger Fall ist der „algorithmische Journalismus“. Hier werden Algorithmen und menschliche Akteure in eng getakteten iterativen Arbeitsabläufen zusammengeführt (Dörr 2017). Dies hat zur Konsequenz, dass die algorithmischen und menschlichen Beiträge zum gemeinsam verfassten Text häufig so eng miteinander verwoben sind, dass es unmöglich wird, einen verantwortlichen Autor zu identifizieren. Es entsteht ein merkwürdiger Hybrid, eine buchstäblich „eigensinnige“ Mensch-Algorithmus-Assoziation (Diakopoulos 2019, 15f.).

Es gibt weitere Fälle solcher Hybride: Spektakulär sind die Konstellationen „digitalisierter Unternehmensführung“, also die Übertragung von Managementaufgaben an autonome Algorithmen.[2] So ernannte „Deep Knowledge Ventures“ einen Algorithmus zum Vorstandsmitglied, dessen Aufgabe darin bestand, über Voraussagen und andere Daten mit den anderen Mitgliedern zu kommunizieren (Möslein 2017). Häufig werden diese schon heute in die kollektive Entscheidungsfindung von Organisationen direkt integriert (Shidaro/Christanikis 2017). Dazu werden Projektorganisationen oder Funktionseinheiten gegründet, oft auch eigene Organisationen (Büchner 2018, 341). Manche Unternehmen setzen Algorithmen sogar als integralen Teil des oberen Managements ein (Mosco 2020, 88 f.). Manchmal fungieren sie innerhalb einer Konzernstruktur als eigenständige Vorstandsmitglieder (Möslein 2017), manchmal bilden sie eigenständige algorithmische Unterorganisationen (Armour/Eidenmüller 2019, 26f.). Noch weiter geht die Integration von Algorithmen in den „Decentralized Autonomous Organizations“ (DAO) (Jentzsch 2016). Hier übernehmen in Investorengruppen Algorithmen selbständig die Organisation, Verwaltung und Entscheidung. In diesen Fällen assistieren Algorithmen nicht bloß bei der Entscheidungsfindung, sondern agieren selbst als autonome Entscheidungsträger innerhalb einer kollektiven Entscheidungseinheit „auf Augenhöhe“ mit den beteiligten Menschen.

Der klassische Fall einer Mensch-Algorithmus-Assoziation ist der Cyborg mit eng ineinandergreifenden algorithmischen Impulsen und menschlichen Entscheidungen (Haselager 2013). Die medientheoretische Interpretation der Cyborgs als „Extensions of Man“ (McLuhan 2003) ist freilich unangemessen. Denn danach werden Information und Partizipation ausschließlich vom menschlichen

---

[1] Für die Analyse solcher digitalen Quasi-Organisationen ist der Ansatz der „partial organization“ (Ahrne/Brunsson 2011) einschlägig. Insofern er den kompakten Organisationsbegriff in seine Einzelkomponenten zerlegt und eine mögliche graduelle Kumulierung der Komponenten bis hin zur „complete organization“ annimmt, ist er geeignet, deren Transformation durch Digitalisierung im notwendigen Detail zu beschreiben. Verantwortungs- und Verantwortlichkeitszurechnung ist eine dieser organisationalen Einzelkomponenten, deren einschneidende Änderungen in den sich herausbildenden Mensch-Algorithmus-Assoziationen im Folgenden näher untersucht werden. Allerdings müsste dieser Ansatz mit einer Theorie der Emergenz korrigiert werden, welche die graduelle Kumulation der Einzelelemente als bloße Strukturänderungen versteht und demgegenüber die Etablierung von neuartigen Elementaroperationen eines autopoietischen Systems anderer Ordnung betont (Apelt et al. 2017), dazu genauer unten im Text.

[2] Dazu genauer und mit einer Diskussion der dabei auftretenden Probleme im Gesellschaftsrecht Markus Becker und Philipp Pordzik (2020) "Digitalisierte Unternehmensführung", *Zeitschrift für die gesamte Privatrechtswissenschaft*, 334-357.

Subjekt her gedacht (Block/Dickel 2020, 111), so dass der Algorithmus nur als Annex menschlicher Handlungskapazitäten erscheint. Doch manchmal dominieren hier algorithmische Kalkulationen die menschliche Entscheidung, manchmal ist es umgekehrt. In soziologischer Sicht wird jedenfalls nicht einfach die menschliche Handlungskapazität erweitert, sondern es entsteht ein neuartiges Mensch-Algorithmus-Kollektivverhalten (Rahwan et al. 2019, 483).

Wie empirische Studien gezeigt haben, sind allerdings völlig autonom algorithmisch gesteuerte Organisationen, in denen sämtliche Managementaufgaben von Algorithmen übernommen werden, heute noch nur hypothetisch. Die zwei bisher extremsten Fälle sind „Plantoid“ und „Metronome“.[3] Auch hier wird noch ein Rest von menschlicher Aktion benötigt. Die folgende Abbildung illustriert den bisherigen Stand (2021) an Operations- und Managements-Algorithmisierung.

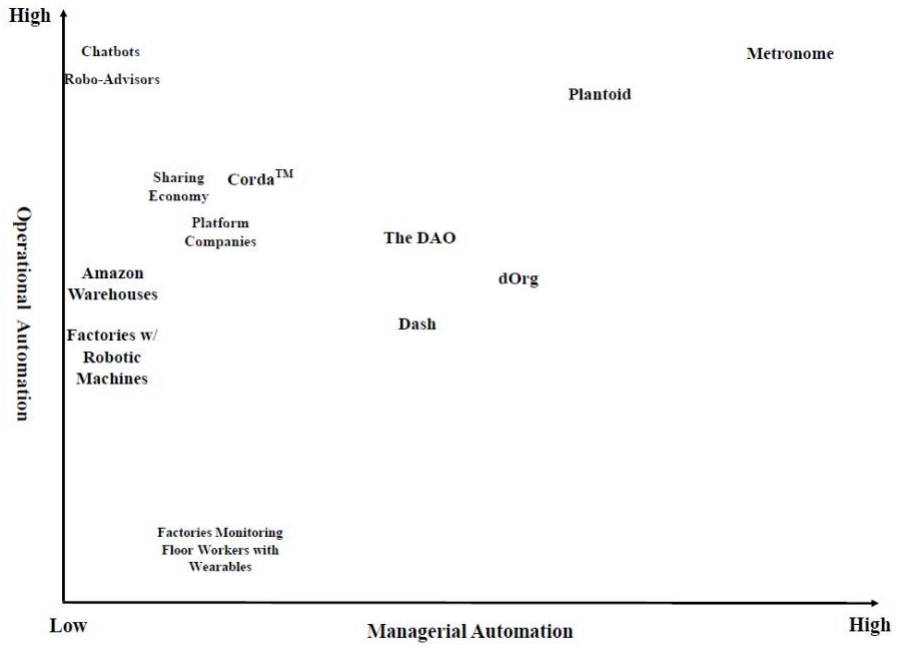

Abbildung 1: Autonomous Business Mapp, Carla L. Reyes (2021) „Autonomous Corporate Personhood“, 96 Washington Law Review, 1453-1510, 1470.

In diesen Situationen geht es – trotz vieler Spekulationen zur Voll-Algorithmisierung von Organisationen in der Literatur – meist um Mensch-Algorithmus-Assoziationen. Nur in den Extremfällen geht es entweder um den Einsatz von bloßen Werkzeugen (links unten in der Abbildung) oder aber um volle Algorithmisierung (rechts oben). Häufig dagegen sind Kooperationen von autonomen Algorithmen und Menschen (mittlerer Bereich). In diesem Bereich wiederum steht das massenhafte Auftreten von hybriden Assoziationen im Gegensatz zu

[3] Ausführlich zu Einzelheiten ihrer Funktionsweise und zu den Eigenschaften der anderen in der Abbildung erwähnten Fällen Reyes (2021, 462 ff.).

verbreiteten Vorstellungen von isoliert agierenden Algorithmen.[4] Die enge Verflechtung ihrer Einzelaktionen macht die ausschließliche Handlungszurechnung an die beteiligten menschlichen Akteure unmöglich. Ebenso wenig überzeugt es in diesen Fällen, die Verantwortlichkeit für Fehler den einzelnen Algorithmen selbst zuzurechnen, wie es aber die Befürworter einer „elektronischen Person“ (e-person) tun (Europäisches Parlament 2017; Kleiner 2021; Zimmerman 2015). Wie aber lässt sich dann die soziale Konfiguration von Mensch-Algorithmus-Assoziationen verstehen und welches Konzept von Verantwortung ist dafür adäquat?

Dazu ist es erforderlich, den sehr weiten Begriff der Verantwortung (responsibility) in zwei Komponenten aufzulösen: Verantwortung im soziologischen Sinne und Verantwortlichkeit im rechtlichen Sinne. Verantwortung bezeichnet soziologisch „die Unsicherheitsabsorption als soziale(n) Prozess: dass jemand aus Informationen Schlüsse zieht, diese einem anderen mitteilt und der nächste sich dann nicht mehr an den Informationen, sondern an den Folgerungen orientiert“ (Luhmann 1964, 175). Algorithmisierung von Entscheidungen bedeutet dann, dass die Unsicherheitsabsorption nicht mehr in der Kommunikation unter Menschen, sondern in der Mensch-Maschine-Kommunikation stattfindet, eher anschaulich, aber ungenau gesagt, dass die Algorithmen die Verantwortung übernehmen. Das geschieht nur dann, wenn sie Entscheidungsautonomie besitzen und sie unter Bedingungen der Ungewissheit handeln. Verantwortung ist keinesfalls mit Verantwortlichkeit gleichzusetzen. Verantwortlichkeit ist vielmehr der Versuch, Verantwortungsprobleme, also Probleme der Unsicherheitsabsorption in die Sprache der Normativität, vor allem des Rechts, aber auch der Moral, der Politik und der Alltagssprache, zu reformulieren, um Fehler in der Verantwortungswahrnehmung zu korrigieren. Hier geht es darum, die Verantwortung für Unsicherheitsabsorption in Begriffe der Rechtmäßigkeit/Rechtswidrigkeit einer Entscheidung und von Rechten, Pflichten, Vorwerfbarkeit und Sanktionen zu „übersetzen“. Dabei entsteht notwendig eine Divergenz von Verantwortung und Verantwortlichkeit (Luhmann 1964, 172 ff.). Diese wird beim Einsatz von Algorithmen nochmals deutlich verstärkt, denn das Recht ist in seiner Begrifflichkeit auf die Übernahme von Entscheidungsaufgaben durch Algorithmen nicht vorbereitet. Auf eine Algorithmisierung der Verantwortung wird das Recht mit Änderungen der Verantwortlichkeitsnormen reagieren müssen.

Die Zurechnung von Verantwortung und von Verantwortlichkeit wird schließlich dadurch verkompliziert, dass wir es bei Mensch-Algorithmus-Assoziationen vermutlich nicht nur mit einem, sondern mit zwei unterschiedlichen Emergenzphänomenen zu tun haben. Emergenz wird hier verstanden nicht als Emergenz „von unten“ in dem Sinne, dass das Zusammenspiel von bestehenden Systemkomponenten unvorhersehbare neuartige Eigenschaften entstehen lässt (Hoyningen 1994, 172 ff.), sondern im Sinne der Systemtheorie als Emergenz „von oben“ (Luhmann 1984, 43 f.; dazu Lohse 2011). Damit ist gemeint, dass ein neues autopoietisches System entsteht, das zwar auf bestehende Materialien zurückgreifen kann, dabei aber seine emergierenden Elementaroperationen selbst herstellt, die in zirkulärer Weise ihrerseits das neue System konstituieren.

In unserem Zusammenhang stellt sich somit erstens die Frage, ob unter bestimmten Umständen die Begegnung von Menschen und Algorithmen ein Sozialsystem im strengen Sinne entstehen lässt, das auf die Kalkulationen der

[4] Siehe Messner (2020). Für eine ausführliche Darstellung, wie die Einführung von KI-basierten Entscheidungsunterstützungssystemen in der Klinik die Art und Weise der Interaktion zwischen verschiedenen Agenten verändert, Braun *et al.* (2020).

Algorithmen zurückgreift, um seine Elementaroperationen als genuine Kommunikationen zu konstituieren. Und zweitens, ob beim Übergang von der Mensch-Algorithmus-Interaktion zur Mensch-Algorithmus-Assoziation eine Organisation als ein neuartiges autopoietisches System emergiert, das die Kommunikationen von Mensch und Algorithmus anders, nämlich als kollektiv verbindliche Entscheidungen innerhalb der Assoziation als seine Elementaroperationen neukonstituiert. In beiden Emergenzphänomenen wird die Verantwortung in jeweils spezifischer Weise institutionalisiert und die Verantwortlichkeit wiederum anders zugerechnet.

Dieser Artikel nähert sich dem Problem anhand von vier aufeinander aufbauenden Teilfragen:

- Findet innerhalb solcher Mensch-Algorithmus-Begegnungen Kommunikation (im strengen Sinne) zwischen menschlichen Akteuren und autonomen Algorithmen statt, so dass ein Interaktionssystem entsteht? Lässt sich in diesem Zusammenhang eine Personifizierung von Algorithmen aufweisen?
- Wenn dies der Fall ist, verdichten sich dann unter bestimmten Bedingungen solche Mensch-Algorithmus-Interaktionen zu hybriden (Quasi-)Organisationen? Und wie verändert dies die Personifizierung der Algorithmen?
- Wenn dies der Fall ist, lässt sich dann nicht nur den einzelnen Algorithmen, sondern zusätzlich solchen Hybriden selbst – vergleichbar formalen Organisationen – eine eigenständige Kommunikationsfähigkeit attribuieren?
- Wenn dies der Fall ist, welche Art von kollektiver Verantwortung und Verantwortlichkeit dürfte in Mensch-Algorithmus-Assoziationen institutionalisiert werden?

II. Personifizierung von Algorithmen

1. Algorithmen als „Aktanten“

In formalen Organisationen „unter Menschen“ werden üblicherweise zwei Formen der Zurechnung von Kommunikationen praktiziert. Einige Kommunikationen werden den Mitgliedern persönlich als ihre eigenen Handlungen zugerechnet. Sie binden nur das Mitglied als Person und begründen dessen individuelle Verantwortlichkeit. Andere Kommunikationen aber werden unter bestimmten Bedingungen der Organisation selbst als kollektive Handlungen zugerechnet. Als im Rahmen der Mitgliedschaft kollektiv verbindliche Entscheidungen fungieren sie als Elementarereignisse des Sozialsystems Organisation. Sie binden die Kommunikationen im gesamten System, sofern bestimmte infrastrukturelle Voraussetzungen (Entscheidungsverfahren, Vertretungsregeln) gegeben sind. Darauf baut die zugleich erreichte Fähigkeit zu dem Kollektiv zugerechneten Handeln auf, die der Organisation deutliche Positionsgewinne in Umweltbeziehungen verschafft, weil nun die Organisation selbst und nicht mehr bloß ihre Mitglieder mit ihren Umwelten kommunizieren kann. Damit begründet diese Fähigkeit zugleich kollektive Verantwortung und Verantwortlichkeit der Organisation als eigenständigem Sozialsystem (Luhmann 1984, 270 ff.; 1997, 834 ff.).

Sobald nun aber Algorithmen in Organisationen integriert sind und ihnen zugemutet wird, Entscheidungen in eigener Verantwortung zu treffen, stellt sich die Frage, wie sich dies auf individuelle und kollektive Zurechnung von Verantwortung und Verantwortlichkeit auswirkt.

Dazu ist zunächst die individuelle Handlungszurechnung auf Algorithmen zu klären. Dieser Frage kann man nicht dadurch ausweichen, dass man Algorithmen nur als automatisierte Werkzeuge der beteiligten Menschen auffasst. Sofern Algorithmen über einen bestimmten Grad von Autonomie, Entscheidungs- und Lernfähigkeit verfügen, wird es unzureichend, sie einfach nur als Annex menschlicher Handlungskapazitäten zu begreifen. Ihre Beiträge können nicht mehr den beteiligten Menschen als ihre Handlungen zugerechnet werden. Stattdessen stellt sich die Frage, ob sie nicht gerade den beteiligten Algorithmen selbst als „Personen" attribuiert werden (Floridi und Sanders (2011).

An der damit gemachten Voraussetzung, dass Algorithmen ebenso wie Menschen als Personen adressiert werden, wird aber regelmäßig kritisiert, dass man Computer mit Menschen gleichsetzt.[5] Damit man nicht solchen falschen Anthropomorphisierungen aufsitzt muss man die Personifizierung von Algorithmen überhaupt erst in ihrer Eigenart verstehen. Aufschlussreich dafür ist die Parallele zur Handlungsfähigkeit anderer nicht-menschlicher Akteure, besonders zu den als juristischen Personen verfassten formalen Organisationen. Dass Organisationen als Kollektivakteure personifiziert werden, folgt nicht aus der vertrauten Vorstellung, Kollektive seien eine aggregierte Vielheit konkreter Menschen.[6] Als Kollektivakteur kommen auch nicht die üblichen Verdächtigen in Betracht, weder Otto von Gierkes „reale Verbandspersönlichkeit", noch Emile Durkheims „Kollektivbewusstsein", noch James Colemans „Ressourcenpool", noch Maurice Haurious „Institutionen".[7] Der Kollektivakteur, wie ihn Talcott Parsons, Niklas Luhmann und andere Soziologen definieren, ist keine Gruppe von Individuen, sondern – eine Kette von Mitteilungen (Bora 2019; Ortmann 2017, 243 ff.).

Die soziale Realität eines Kollektivakteurs, das soziale Substrat der juristischen Person (Teubner 1987, 61 ff.), entsteht unter zwei Voraussetzungen: Erstens muss eine solche Kommunikationskette über sich selbst kommunizieren und durch diese Selbstbeschreibung ihre Identität herstellen. Zweitens müssen die im Kontext dieser Identität produzierten kommunikativen Ereignisse als von ihr selbst vollzogene Handlungen sozial zugerechnet werden. Wenn diese Voraussetzungen erfüllt sind, entsteht die soziale Realität eines Kollektivakteurs, der als semantisches Artefakt „Person" zum Zurechnungspunkt für Handlungen, Ressourcen und Verantwortlichkeiten wird. Damit ist es definitiv ausgeschlossen, kollektives Handeln auf die Einzelhandlungen von Individuen zurückzuführen, wie es der methodologische Individualismus vorschreibt.

In ganz paralleler Weise ist nun die Personifizierung von Algorithmen zu verstehen. Softwareagenten, Roboter und andere digitale Akteure sind

---

[5] Siehe für diese fragwürdige Behauptung vor allem den Offenen Brief an die Europäische Kommission, Künstliche Intelligenz und Robotik, abrufbar unter www.robotics-openletter.eu.

[6] Zum soziologischen Begriff des kommunikationsfähigen Kollektivs: Luhmann (2018), 112f.; (2014 (1975), 14); (2000b), 241; (1997), 834f.; (1984), 269 ff. Aus der Anthropologie: Sprenger (2017), 111. Ähnlich die Diskursanalyse unter dem Stichwort der Subjektivierung: Alkemeyer *et al.* (Hrsg.) (2018). Aus der Sicht der Kognitionsphilosophie: Dennett (1987), 17.

[7] Zum Begriff der „realen Verbandspersönlichkeit" Gierke (1902). Zum „Kollektivbewusstsein" Durkheim (2004 [1883]), 111 ff. Zum „Ressourcenpool" Coleman (1990), 325ff. Zu „Normkomplexen" Hauriou (1965).

mathematisch formalisierte Informationsflüsse. Algorithmen, die mit ihrer Umgebung interagieren, sind nicht bloß statische mathematische Formeln, sondern dynamische Prozesse, die in ihren Problemlösungsoperationen Programmen folgen. „Was ist ein Algorithmus?“ – nach einer eingehenden Diskussion verschiedener Definitionen kommt Yuri Gurevich zu folgender Definition, die den prozesshaften Charakter betont:

> „Ein zeitlich-sequenzieller Algorithmus ist ein Zustandsübergangssystem, das in einem Ausgangszustand beginnt und von einem Zustand zum nächsten übergeht, bis es, wenn überhaupt, anhält oder abbricht .... Insbesondere ist ein zeitlich-sequenzieller interaktiver Algorithmus ... ein Zustandsübergangssystem, bei dem ein Zustandsübergang vom Senden und Empfangen von Nachrichten begleitet sein kann“ (Gurevich 2012, 40, unsere Übersetzung).

Schon heute wird den Algorithmen in Wirtschaft und Gesellschaft – unter bestimmten Bedingungen – soziale Identität und Handlungsfähigkeit zugeschrieben.[8] Ob aber die Bedingungen der Personifizierung erfüllt sind, das wird nicht technikdeterministisch schon durch die Wahl einer digitalen Technologie entschieden, sondern erst in den entstehenden „sozio-digitalen Institutionen“, in deren Rahmen die Algorithmen eingesetzt werden. Sozio-digitale Institutionen sind stabilisierte Komplexe sozialer Erwartungen, insbesondere Erwartungen über neu entstehende Risiken, die sich bilden, wenn digitale Technologien in unterschiedlichen sozialen Situationen genutzt werden. Solche Institutionen sind weder mit sozialen Systemen noch mit formalen Organisationen identisch. Vielmehr produzieren soziale Systeme, darunter auch formale Organisationen und natürlich auch interindividuelle Interaktionen, mithilfe ihrer Kommunikationen Erwartungen,[9] die, um eine klassische Formulierung zu bemühen, sich unter einer idée directrice zu Institutionen verdichten. Institutionalisiert sind solche Erwartungen erst dann, wenn Konsens zu ihrer Unterstützung unterstellt werden kann (Luhmann 1987, 64 ff.). Die bemerkenswerte Stabilität von Institutionen im Allgemeinen wird aber nur dann erreicht, wenn sie die Fähigkeit entwickeln, Brücken zwischen unterschiedlichen Systemen zu bauen.[10]

Sozio-digitale Institutionen im Besonderen entstehen als wirksame strukturelle Kopplungen zwischen technischen und sozialen Systemen. In diesen Institutionen

---

[8] Vertreter unterschiedlicher Disziplinen sprechen sich explizit für die Personifizierung von Algorithmen aus. In der Informationsphilosophie: Floridi/Sanders (2011), 187 ff.; Thürmel (2015), 52 ff. In der Soziologie: Bora (2019); Muhle (2018), 147; (2020); E. Esposito (2017); Harth/Lorenz (2017); Weyer/Fink (2011), 40 ff. In der politischen Theorie: Cohen (2019), 221 ff. In der Rechtstheorie: Teubner (2006).

[9] Während im „alten“ Institutionalismus (Santi Romano, Hauriou, Carl Schmitt) der Begriff der Institution zwischen Systemen, Organisationen und Normen oszillierte, stimmen „neuer“ Institutionalismus und Systemtheorie trotz vieler Unterschiede in ihrem Verständnis von Institutionen als Erwartungskomplexe überein, North (1990), 3 f.; Luhmann (1965), 12 f.

[10] Zur Rolle von Institutionen als Brückenarrangements in einer fragmentierten Gesellschaft Teubner (2001).

treffen Erwartungen wirtschaftlicher, politischer, rechtlicher und technologischer Art aufeinander, und es ist häufig schwierig, zwischen den Erwartungen der beteiligten Systeme zu unterscheiden (R. Esposito 2022, 37). In einem Prozess der Ko-Produktion ihrer Komponenten (Feenberg 2017; Jasanoff 2004; Winner 1980) integrieren sozio-digitale Institutionen die unterschiedlichen Erwartungen zu Chancen und Risiken des Einsatzes von Algorithmen.[11] Und diese Institutionen sind letztlich dafür verantwortlich, ob den beteiligten Prozessen Personalität zugeschrieben wird oder nicht. Dies geschieht „jedoch ohne vorgegebene Unterscheidung zwischen menschlichen und nicht-menschlichen Personen und mit der Möglichkeit, Grade von Personalität zu differenzieren„ (Sprenger 2018, 248).

Die genaue Parallele von formalen Organisationen und Algorithmen wird deutlich, wenn die gängigen Fehlvorstellungen zur Personifizierung nicht-menschlicher Entitäten ausgeräumt werden. Falsch ist die allgemeine Vorstellung zur Personifizierung von Organisationen, dass sich eine Menschenansammlung in eine reale Verbandspersönlichkeit verwandelt. Und ebenso falsch ist die Vorstellung, dass im Falle von Softwareagenten ein Computer in einen *homo ex machina* transformiert wird. Stattdessen wird für Organisationen und für Algorithmen eine der wichtigsten „evolutionären Errungenschaften sozialer Systeme“ (Luhmann 1984, 273) verwirklicht: Institutionalisierte soziale Praktiken schreiben bestimmten Kommunikationsprozessen Handlungsfähigkeit zu.

In der Akteur-Netzwerk-Theorie hat Bruno Latour (2001, 93 ff.) für handlungsfähige non-humans den Neologismus „Aktanten“ geprägt. Damit arbeitet er den fundamentalen Unterschied zur Handlungsfähigkeit von Menschen heraus und definiert die Bedingungen der Handlungsfähigkeit. Latour beschreibt ihre Handlungsfähigkeit jedoch sehr weit als „Widerständigkeit“. Demgegenüber arbeitet Rossler (2015, 87 ff.) eine ganze Vielfalt von Bedeutungen heraus, die mit den Handlungen nicht-menschlicher Akteure verbunden sind, und erstellt eine Typologie ihrer Handlungsformen. Aus dieser Typologie wählt der vorliegende Text einen Handlungstyp heraus, der sehr viel enger als bei Latour auf Teilnahme an gesellschaftlicher Kommunikation abstellt (ähnlich Muhle 2018, 155 ff.; 2020, 131 ff.). Wie schon der Begriff „Aktanten“ deutlich macht, geht es keinesfalls um eine Anthropomorphisierung von digitalen Prozessen, sondern gerade umgekehrt um eine De-Anthropomorphisierung der allzu menschlichen Roboter und Softwareagenten. Diese bleiben – wie Informationsphilosophen betonen – „mindless machine(s)“ (Hildebrandt 2015, 22; Floridi/Sanders 2011, 186). Trotzdem werden sie durch die Handlungszurechnung innerhalb von sozio-digitalen Institutionen zu (nicht-menschlichen) Mitgliedern der Gesellschaft (Messner 2020). Die Mensch-Algorithmus-Assoziation ist eine dieser sozio-digitalen Institutionen, die den einzelnen Algorithmen den Status von vollwertigen Organisationsmitgliedern mit den dazugehörigen Berechtigungen und Verantwortlichkeiten verleiht.

Warum aber verleihen manche sozio-digitale Institutionen Algorithmen den Personenstatus und attribuieren andere ihnen nur den Status als Werkzeuge oder als integraler Bestandteil des menschlichen Körpers? – In der Begegnung mit nicht-menschlichen Entitäten, insbesondere mit Algorithmen, erweist sich deren Personifizierung als eine der erfolgreichsten Strategien, mit Ungewissheit

[11] Das ist der Grund, warum der vorliegende Beitrag Institutionen und nicht soziale Systeme in den Vordergrund stellt. In unseren Zusammenhang von Verantwortung und Verantwortlichkeit kommt es primär auf die Erwartungen an, die soziale Systeme generieren.

umzugehen, insbesondere mit der Nicht-Prognostizierbarkeit ihres Verhaltens.[12] Die soziale Personifizierung transformiert die Mensch-Algorithmus-Beziehung aus einer Subjekt-Objekt-Beziehung in eine Ego-Alter-Beziehung. Diese gibt freilich Ego nicht Gewissheit über Alters Verhalten, aber ihre wechselseitige Kommunikation ermöglicht es, dass Ego sein Verhalten wählen kann, auch wenn Alter selbst undurchschaubar bleibt: „Der Computer bzw. der Algorithmus ist dann nicht mehr ein technisches Artefakt mit Zuschreibungspotenzial, sondern ein Gegenüber, das entweder im Falle natürlicher Sprache anthropomorpher, im Falle von Entscheidungsalgorithmen als korporativer Akteur auftritt“ (Nassehi 2019, 224; ähnlich Hepp 2020).

Den Algorithmus so zu behandeln, „als ob“ er ein Akteur wäre, verwandelt also die Ungewissheit über Kausalabläufe in die Ungewissheit darüber, wie die Reaktion des Algorithmus auf die Handlungen des menschlichen Akteurs zu verstehen ist. Kausalzurechnung transformiert sich in Handlungszurechnung. Das beruht auf der kommunikativen Unterstellung, dass der Algorithmus über Motive für sein Handeln verfügt. Die Unterstellung wirkt als harte soziale Realität des Akteurstatus, insofern „menschliche Akteure ihre nicht-menschlichen Mitspieler ... als ebenbürtige Partner ansehen. Sie schreiben ihnen Handlungsträgerschaft zu – und zwar in einer symmetrischen Weise.“[13] Dadurch wird der Mensch in die Lage versetzt, seine eigenen Handlungen zu wählen, die Reaktionen des Algorithmus zu interpretieren und daraus wiederum Konsequenzen für die nächste Handlung zu ziehen (Muhle 2018, 156).

2. Kommunikation mit Algorithmen

Doch können wir wirklich davon ausgehen, dass Algorithmen genauso kommunizieren, wie wir das bei Menschen, aber auch bei Organisationen voraussetzen? Nehmen wir, wenn wir in digitalen Hybriden mit Algorithmen kooperieren, nur das Verhalten von Maschinen wahr, oder kommunizieren wir mit ihnen?[14] Handelt es sich also um bloße Wahrnehmung oder schon um echte Kommunikation? Die übliche Antwort ist, dass man nach „psychologischen Fähigkeiten sucht, von denen man bisher dachte, sie seien komplexen biologischen Organismen wie dem Menschen vorbehalten“ (Prescott 2017, 142). Eine andere Antwort jedoch gibt nach der jüngsten „relationalen Wende“ die Tier-, Roboter- und Maschinenethik (Coeckelbergh 2012). Danach besitzen Algorithmen als solche keine ontologischen Akteursqualitäten, die es ihnen erlauben, mit Menschen zu kommunizieren. Sobald aber Algorithmen integraler Teil sozialer Beziehungen werden, sobald sie wie in unserem Fall zu Mitgliedern digitaler Hybride werden, wird die Zuschreibung von kommunikativen Fähigkeiten ermöglicht.

Die Systemtheorie nimmt hier wichtige Verfeinerungen vor, indem sie die Bedingungen bestimmt, die für eine Personifizierung von Algorithmen erfüllt sein müssen (Baecker 2020a; 2020b; 2011, 1). Ob eine Begegnung von Algorithmus und

---

[12] Dies ist das systemtheoretisch basierte Argument bei Teubner (2006). Wichtige Verfeinerungen Nassehi (2019), 221 ff.; Harth/Lorenz (2017), 12 ff.

[13] Weyer und Fink (2011), 43, die ein empirisch überprüfbares Modell der Handlungsträgerschaft von Algorithmen vorlegen.

[14] Dies ist Teil der Forschungsagenda für das neue Feld der Mensch-Algorithmus-Kommunikation, Guzman und Lewis (2020).

Mensch innerhalb eines sozio-digitalen Hybriden zu einer Personifizierung der Algorithmen führt, hängt davon ab, ob die „Beiträge“ der Softwareagenten als kommunikative Ereignisse im strengen Sinne konstituiert werden. Dies erfordert die Synthese von Information, Mitteilung und Verstehen.[15] Es muss also gelingen, Ereignisse zu identifizieren, die als „Mitteilungen“ des Algorithmus zugleich eine bestimmte „Information“ enthalten, und deren Differenz in der Anschlusskommunikation „verstanden“ wird. Dann 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 Synthese von Information, Mitteilung und Verstehen, die Voraussetzung für Kommunikation im strengen Sinne ist, erfordert.[16]

Dies wäre also das erste der oben angesprochenen Emergenzphänomene: Die elektronischen Signale der Algorithmen bilden das Material, auf die das emergierende Sozialsystem Mensch-Algorithmus-Interaktion zurückgreift, um Kommunikationen als seine Elementaroperationen zu erzeugen. Es geht also nicht um bloße Wahrnehmungsakte, mit denen Menschen algorithmische Operationen ablesen, sondern um sinnhafte Kommunikationen (im strengen Sinne) zwischen Mensch und Algorithmus.

Zwar sind dies bloß „asymmetrische“ Interaktionen von Menschen und Algorithmen.[17] Dennoch bildet sich zwischen ihnen ein genuines selbstproduzierendes Kommunikationssystem, eine Mensch-Algorithmus-Interaktion, heraus. Diese ist asymmetrisch in einem dreifachen Sinne.

(1) Den Algorithmen können keineswegs die mentalen Kompetenzen von Menschen zugebilligt werden (E. Esposito 2017, 250). In ihrem Innenleben kalkulieren mathematische Operationen auf der Grundlage elektronischer Signale. Um zu verstehen, wie trotz der asymmetrischen Innenwelten eine Kommunikation zwischen Computern und Menschen möglich ist, dürfte die Unterscheidung zwischen „subface“ und „surface“ hilfreich sein (Nake 2020):

> „Subface ist die technische Seite digitaler Medien, gekennzeichnet durch die Vernetzung und Verknüpfung kausal gesteuerter Prozesse in Hard- und Software (Wenn-dann-Verknüpfungen und sonstige Schleifen). Surface ist die ebenfalls technisch gestaltete Seite, die aber grundsätzlich offen ist für den Zugriff durch ... Kommunikation (Information, Mitteilung und Verstehen)“ (Baecker 2020a, 3, unsere Übersetzung).

---

[15] Für ein allgemeines Modell der Kommunikation Luhmann (1984), 191 ff. Zur Kommunikation mit nicht-menschlichen Entitäten Sprenger (2017), 116ff.; Luhmann (1997), 117 ff., 303 f. Zur Kommunikation mit Algorithmen E. Esposito (2017), 249, 254 ff.; Harth und Lorenz (2017); Teubner (2006).

[16] Zur Frage, ob Arbeit mit Computern als Kommunikation zu verstehen ist, auch wenn doppelte Kontingenz womöglich nur einseitig erfahren wird, Luhmann (1997), 303 f. E. Esposito (2017, 262) stellt die Frage, ob der Kommunikationsbegriff erweitert oder ein anderer Begriff für die Begegnung Computer-Mensch entwickelt werden muss, lässt die Antwort zunächst offen, um sich dann später für die erste Alternative zu entscheiden (2022).

[17] Gutmann (2011, 14) spricht von einer „intrinsischen Asymmetrie parasozialer Relation“, Fuchs (1991, 20) von „einseitig bewusster Kommunikation“.

Für eine gelungene Kommunikation zählt nur das Surface, das Subface ist nicht mehr Teil des Kommunikationsgeschehens. Die Unterschiede, die zwischen dem Subface des Computers und dem Bewusstsein des Menschen bestehen, so fundamental sie auch sind, werden für unsere Frage gar nicht relevant. Entscheidend ist vielmehr, dass außerhalb der höchst unterschiedlichen Innenwelten von Mensch und Computer ein Kommunikationsprozess zustande kommt. Zugleich aber muss sowohl das digitale Subface als auch das menschliche Bewusstsein diesen Kommunikationsprozess ausreichend irritieren können (Harth/Lorenz 2017, 12 ff.). Dann erst kann die Synthese aus Information, Mitteilung, Verstehen gelingen.

(2) Eine Asymmetrie entsteht weiterhin dadurch, dass Kommunikation in dieser Beziehung nicht beidseitig erlebt wird. In der Kommunikation unter Menschen wird die doppelte Kontingenz symmetrisch beidseitig erlebt, weil beide Partner die Wahl ihres Verhaltens von der Verhaltenswahl des anderen abhängig machen und genau dies bei ihrem Gegenüber voraussetzen (Parsons/Shils 1951, 3 ff., 16; Luhmann 1984, 148 ff.). Dagegen wird zwischen Mensch und Algorithmus die für Kommunikation notwendige doppelte Kontingenz nur einseitig erlebt, nämlich nur vom Menschen und nicht vom Algorithmus (dazu Luhmann 1997, 304). Einseitig erlebte doppelte Kontingenz, wie wir sie in der Mensch-Algorithmus-Beziehung antreffen, schließt allerdings genuine Kommunikation nicht aus. Historisch bekannte Konfigurationen, wie die Kommunikation mit Gott im Gebet ebenso wie animistische Praktiken in indigenen Gesellschaften, leisten durchaus die Synthese aus Mitteilung, Information und Verstehen (Baecker 2020b; Sprenger 2017, 119 ff.; Luhmann 1987, 234 f.). Sie tun dies aber stets unter der Voraussetzung, dass im Ereignisstrom eine Personifikation des nicht-menschlichen Partners stattfindet, welche eine Zurechnung von Handlung auf den „anderen“ ermöglicht. „Personalität entsteht, wo immer das Verhalten anderer als gewählt vorgestellt wird und durch eigenes Verhalten kommunikativ zu beeinflussen ist“ (Luhmann 1997, 643). Auch in unserem Kontext von Mensch-Algorithmus-Assoziationen konstituiert die Handlungszurechnung auf den involvierten Algorithmus diesen als Person, als ein semantisches Konstrukt, und kompensiert somit die zweite Asymmetrie auch in der Mensch-Algorithmus-Beziehung.

(3) Die Kommunikation ist schließlich asymmetrisch in Bezug auf das wechselseitige Verstehen. Wenn Verstehen verstanden wird als die Fähigkeit, die Selbstreferenz eines anderen in der eigenen Selbstreferenz nachzukonstruieren (Luhmann 1986), dann kann das menschliche Bewusstsein die Botschaften des Algorithmus verstehen. Den Binnenprozessen des Algorithmus dagegen dürfte die Fähigkeit, das menschliche Innenleben zu verstehen, fehlen. Weil es jedoch für das Zustandekommen von Kommunikation auf ein solches wechselseitiges Tiefenverstehen im Innenleben von Mensch und Algorithmus gar nicht ankommt, kann man auch diese Frage offenlassen. Denn man muss deutlich zwischen Verstehensprozessen innerhalb des autonomen Kommunikationsgeschehens einerseits und Verstehensprozessen im Innenleben der interagierenden Entitäten andererseits unterscheiden (Luhmann 1986, 72, 93 ff.). Für Verstehen innerhalb des verselbständigten Kommunikationsprozesses ist nicht entscheidend, ob der Algorithmus selbst versteht oder nicht; vielmehr kommt es nur darauf an ob der „Antworttext“ des Algorithmus den „Fragentext“ des Menschen versteht.[18] Wenn also

[18] Diese Unterscheidung ignoriert Fuchs (1991, 9 ff.), der deshalb (irrtümlich) auf das Verstehen im Bewusstsein abstellt und (vergeblich) nach Verstehensprozessen im Innenleben der Computer fahndet. Dies kritisieren zu Recht Harth und Lorenz (2017,

der Kommunikationsakt des algorithmischen Mitglieds des digitalen Hybriden die Differenz von Mitteilung und Information im Kommunikationsakt des menschlichen Mitglieds nachvollzieht und darauf mit einer eigenen Differenz von Mitteilung und Information reagiert, ist ein kommunikativer Verstehensakt innerhalb der Assoziation vollzogen. Dies gilt, ohne dass es – um es nochmals zu betonen – darauf ankäme, ob die inneren Kalkulationen des Algorithmus die Äußerung des Menschen verstanden haben oder nicht. Auch hier bieten sowohl die Kommunikation mit Gott als auch animistische Praktiken in indigenen Gesellschaften historische Belege für Verstehen im kommunikativen Geschehen, das auch dann zustande kommt, wenn der „andere" in seinem Inneren nicht verstehen sollte (Baecker 2020b; Sprenger 2017, 119 ff.).

Um das bisher Gesagte in einer Kurzformel zu verdichten: Algorithmen sind – ebenso wie formale Organisationen – nichts anderes als Informationsströme, die dann zu kommunikationsfähigen „Personen" werden, wenn ihnen in einer sozio-digitalen Institution eine soziale Identität attribuiert wird und wenn ihnen zusammen mit den notwendigen infrastrukturellen Vorkehrungen (z.B. Vertretungsregeln) eigene Verantwortung und Verantwortlichkeit effektiv zugerechnet wird.

II. Verantwortung und Verantwortlichkeit in Mensch-Algorithmus-Assoziationen

Bis hierhin ging es darum, ob innerhalb einer Mensch-Algorithmus-Assoziation genuine Kommunikationen zwischen ihren Mitgliedern zustande kommen und ob die Kommunikationen den semantischen Konstrukten der algorithmischen Mitgliederpersonen persönlich als Handlungen zugerechnet werden können. Diese Frage konnte bejaht werden, sofern nur bestimmte Bedingungen erfüllt sind. Im nächsten Gedankenschritt geht es nun um die davon verschiedene Frage, ob nicht nur ein beteiligter Algorithmus, sondern ob zugleich die Mensch-Algorithmus-Assoziation selbst als eigenständige Handlungseinheit, als kommunikationsfähige Kollektivperson, zu verstehen ist und was das für die Zurechnung von Verantwortlichkeit bedeutet.

Hier treffen wir auf das zweite oben angesprochene Emergenzphänomen. Die Mensch-Algorithmus-Assoziation stellt im Unterschied zur bloßen Interaktion ein neues autopoietisches System dar, das die Kommunikationen von Algorithmus und Mensch in kollektiv verbindliche Entscheidungen als seine Elementaroperationen „transformiert". Die gesuchten Elementaroperationen sind nicht nur einfach bloß Entscheidungen, denn diese werden auch häufig in Interaktionen und in Funktionssystemen gefällt. Als Elementarereignisse von Organisationen können nur im Rahmen der Mitgliedschaft kollektiv verbindliche Entscheidungen gelten, weil erst die Bindung anderer Kommunikationen innerhalb der Grenzen der Mitgliedschaft das Proprium der Organisation ausdrücken. Mit dieser die Organisation konstituierenden Bindung ist „eine besondere Art von Relevanz jeder Entscheidung für andere gemeint – und dies im Doppelsinne von Bindungswirkung und Verantwortlichkeit" (Luhmann 2000a, 65).[19]

---

6 ff.), die konsequent auf das Verstehen in der Anschlusskommunikation (und nicht im Innenleben des Computers oder des menschlichen Bewusstseins) setzen.

[19] Wenn die Begrifflichkeit nicht so hässlich wäre, könnte man, um das gemeinte Kollektiv im Unterschied zu Staat und Recht näher zu bestimmen, von „mitgliedschaftlich verbindlichen" Entscheidungen sprechen.

Zwischenstufen im „Übergang“ von digitaler Interaktion zu digitaler Assoziation begrifflich zu erfassen, besteht kein Bedarf, besonders nicht für „digitale Gruppen“. Systemtheoretisch ist eine Gruppe nicht als eigenständiger Sozialsystemtyp zu verstehen, sondern als ein bloßer „Modus von Interaktion und Interaktionshäufung“[20] Aber gibt es nicht durchaus die Beteiligung von autonomen Algorithmen in lockeren Gemeinschaften, Familien, Bewegungen oder Netzwerken? In der Tat, diese alle kennen Mitgliedschaften und in allen finden sich Entscheidungen und sogar Entscheidungsverkettungen. Jedoch es fehlt das entscheidende „konstitutionelle Moment“ – die kollektive Bindung. Sobald Vorkehrungen getroffen sind, dass bestimmte Entscheidungen das ganze System und das heißt künftige Entscheidungen, aber auch die gesamten Strukturen (Mitgliederrollen, Normen, Entscheidungsprämissen) binden sollen, dann „transsubstantialisiert“ sich das digitale Interaktionssystem in eine digitale Organisation. Was heißt dann aber „Bindung“ durch algorithmische Entscheidungen? Es heißt: Algorithmen binden die Mensch-Algorithmus-Assoziation in den drei Sinndimensionen. Zeitbindung: die Entscheidungen der Algorithmen erzeugen normative, also nicht lernbereite Erwartungen in Bezug auf künftige Entscheidungen. Sachbindung: ihre vergangenen Entscheidungen binden (ähnlich wie die judiziellen Präzedenzien im Recht) die Inhalte künftiger Entscheidungen. Sozialbindung: ihre Entscheidungen binden (im Unterschied zu Entscheidungen im Recht oder in der Politik) ausschließlich die Mitgliederrollen innerhalb der Assoziation.

Die Frage ist dann also, ob im Falle der digitalen Hybride eine *doppelte Personifizierung* im Spiel ist: erstens, die Personifizierung eines Algorithmus als eines individuellen Mitglieds der Assoziation, das mit menschlichen Mitgliedern zu kommunizieren in der Lage ist, und, zweitens, die Personifizierung der Assoziation selbst als eines Kollektivakteurs. Falls ja, müsste diese Duplikation der Handlungsfähigkeit entsprechend komplexe Verantwortlichkeitszurechnungen nach sich ziehen.

1. Mensch-Algorithmus-Assoziationen als „Hybride“

Wieder kommt es darauf an, welche Art von sozio-digitaler Institution sich hier etabliert. Sofern der Algorithmus nur dazu eingesetzt ist, um bestimmte Instruktionen auszuführen sowie für vorgegebene Ziele die geeigneten Mittel zu wählen, und sofern die Beiträge von Mensch und Algorithmus klar abgrenzbar sind, ist die Institution der „digitalen Assistenz“ einschlägig. In diesem Fall besteht zwischen Mensch und Algorithmus eine Prinzipal-Agenten-Beziehung (Krafft et al. 2020) Der Algorithmus wird als Agent von seinem menschlichen Prinzipal beauftragt und führt dessen programmierte Instruktionen gegenüber Dritten eigenverantwortlich aus. Ökonomische Theorien legen den Schwerpunkt bei der Analyse von Prinzipal-Agenten-Beziehungen auf die Risiken eines eigenständig handelnden Agenten und

[20] So die bekannte Kehrtwendung, zu besichtigen in Luhmanns Zettelkasten, https://niklas-luhmann-archiv.de/bestand/zettelkasten/suche (Zettelnummer 21/ 3d27fC5), die der früheren Formulierung, die Gruppe sei zu behandeln „als Typik von Systembildung, die nicht auf Interaktion, Organisation und Gesellschaft zurückgeführt werden kann“ (21/ 3d27fC) diametral widerspricht (dazu Kühl 2014, 72).

dessen Abweichen von den Intentionen des Prinzipals (Fama/Jensen 1983). Dagegen konzentrieren sich Philosophie und Soziologie auf die positiven Beiträge beider Partner zur Bereicherung des produktiven Potenzials der Prinzipal-Agenten-Beziehung (Trüstedt 2020, 195). In der Institution digitaler Assistenz handeln die Agenten autonom. Für ihre in eigener Verantwortung vollzogenen Entscheidungen aber wird, wenn etwas schiefgeht, die Verantwortlichkeit nicht ihnen selbst, sondern den Prinzipalen zugerechnet (Beckers/Teubner 2022, 54 ff., 67 ff., 79 ff.).

Solche individualistischen Konzepte von Verantwortung und Verantwortlichkeit scheitern aber, sobald die Handlungen von Menschen und Algorithmen so eng miteinander verwoben sind, dass es „keine lineare Verbindung zwischen den emergenten Strukturen, Kulturen oder Verhaltensweisen, die Kollektive ausmachen, und den komplexen Interaktionen der Individuen, aus denen sie hervorgehen“, gibt (Chinen 2019, 101). Menschen und Algorithmen scheinen sich zu einer symbiotischen Einheit zu verbinden und sich zu einer Entität zu entwickeln, die größer ist als die Summe ihrer Teile (Turner 2018, 167). Wie Büchner (2018, 338) deutlich macht, widerspricht in dieser Situation die soziale und materielle Einbettung von Algorithmen den aktuellen Tendenzen, Algorithmen gleichsam analytisch freizusetzen und ihnen eine isolierte „algorithmic power“ zuzuschreiben. In einer solchen Situation wird die Institution der digitalen Assistenz durch eine andersartige sozio-digitale Institution ersetzt – die Mensch-Algorithmus-Assoziation. Wenn in der gemeinsamen Entscheidungsfindung die individuellen Beiträge von Mensch und Algorithmus verschmelzen, dann entwickeln Mensch-Algorithmus-Interaktionen neuartige kollektive Eigenschaften.

Werden in dieser sozio-digitalen Institution auch Verantwortung und Verantwortlichkeit des Hybriden selbst etabliert? Und wenn ja, in welchen Formen? Werden solche engen Mensch-Algorithmus-Kooperationen zu digitalen Organisationen, in denen die Verantwortung für die Unsicherheitsabsorption durch den Algorithmus nicht mehr dem Algorithmus, sondern dem Hybrid selbst als Kollektivakteur zugerechnet wird? Und wie wird die Verantwortlichkeit für solche hybride Verantwortung neu verteilt? Für die Antwort auf diese Fragen ist es hilfreich, die ersten tastenden Lösungsvorschläge aus Organisationstheorie (Grothe-Hammer 2019), Netzwerktheorie (Lior 2022), KI-Ethik (Floridi/Sanders 2011), KI-Haftungsrecht (Beckers/Teubner 2022), Verfassungstheorie (Möllers 2021; Kersten 2016; Ingold 2014) und Verantwortungssoziologie (Bora 2019; 2013) aufzunehmen und in ein in sich stimmiges Verantwortlichkeitskonzept für Hybride zu überführen.

Latours Akteur-Netzwerk-Theorie wird hier nun zum zweiten Mal relevant. Oben haben wir sein Konzept der „Aktanten“ verwendet, das autonome Algorithmen personifiziert und ihnen Kommunikationsfähigkeit zuspricht. Im jetzigen Kontext wird die häufig nur metaphorisch verwendete Vorstellung des „Hybriden“ als ein theoretisches Konzept konturiert (Latour 2001, 103 ff.), das die Transformation der Kooperation von „Aktanten“ und Menschen in einen „Hybrid“ als Kollektivakteur erfasst. Für digitale Aktanten wird nun, so argumentiert Latour, ihr Potential zur Assoziationsbildung relevant. Es gebe viele Situationen, in denen Handlungskapazitäten auf einer höheren Ebene benötigt werden, als sie Algorithmen zur Verfügung stehen. Selbst wenn Aktanten die Fähigkeit haben, zwischen Alternativen zu wählen, würde sie dies dennoch häufig in einer Position der Lähmung zurücklassen. Um dem entgegen zu wirken, sind nach Latour „Aktanten“ nicht nur mit einer Sprache und einem widerstandsfähigen Körper ausgestattet, sondern auch mit der Fähigkeit, „Assoziationen“ zu bilden. Die Handlungsfähigkeit von Nicht-Menschen werde dann drastisch erweitert, wenn Hybride gebildet werden,

d. h. Assoziationen von menschlichen Akteuren und nicht-menschlichen Aktanten. In dieser Situation werden die Aktanten zu Mitgliedern in Hybriden, also in vollwertigen Assoziationen einer neuen Art. Es findet, wie in jeder Assoziation, eine Bündelung von Ressourcen statt. Die Widerspenstigkeit der Aktanten wird nun mit den kommunikativen Fähigkeiten der Menschen verbunden. Die verteilte Intelligenz in solchen soziale Systemen kompensiert die psycho-systemischen Kompetenzdefizite von non-humans (Lorentzen 2002, 110; Gellers 2021, 118 ff., 154 ff.). Durch die Kombination menschlicher und nicht-menschlicher Eigenschaften innerhalb von Hybriden können Algorithmen (jedenfalls indirekt) an politischen Verhandlungen, wirtschaftlichen Transaktionen und rechtlichen Vertragsabschlüssen partizipieren.

Die Systemtheorie liefert zusätzliche Einsichten. Sie kann insbesondere erklären, welche Funktion es hat, wenn nicht nur einzelne Algorithmen, sondern auch Mensch-Algorithmus-Assoziationen personifiziert werden. Es bestehen zwar viele Vorschläge, welche die Funktion der Personifizierung von nicht-menschlichen Akteuren, besonders von Organisationen, in der heutigen Gesellschaft zu interpretieren versuchen.[21] Ökonomen verweisen auf das Einsparen von Transaktionskosten bei Mehrparteienverträgen (Williamson 1985). Soziologen identifizieren Koordinationsvorteile bei der Bündelung von Ressourcen (Coleman 1990, 325 ff.). Juristen neigen dazu, die „juristische Unsterblichkeit“ von inkorporierten Objekten zu betonen — Kirche, Staat, Unternehmen (Blackstone 1771, 467 ff.). Luhmann (1984, 270 ff.) hingegen arbeitet einen anderen, hier entscheidenden Aspekt heraus: Die Funktion der Kollektivierung des Handelns besteht darin, das Umweltverhältnis eines Sozialsystems drastisch zu verändern. Soziale Systeme gewinnen, sobald sie personifiziert sind und damit selbst Handlungsfähigkeit erlangen, erhebliche Positionsvorteile im kommunikativen Kontakt mit ihrer Umwelt. Über Organisation kann die eigene Handlungsfähigkeit sozialer Systeme begründet werden. Wenn unter die Mitgliedschaftsregeln aufgenommen wird, dass ein Teil für das Ganze zu handeln befugt ist, dann binden deren Kommunikationen alle Mitgliedschaftsrollen und zugleich werden diese Kommunikationen dem Ganzen als Kollektivakteur zugerechnet (Luhmann (2014 [1975]), 14). Auf diese Weise werden auch Mensch-Algorithmus-Assoziationen zu handlungsfähigen Kollektiven, die als solche mit ihren Umwelten kommunizieren.

Das verändert zugleich die Qualität der Einzelkommunikationen der beteiligten Algorithmen und deren soziale Position. Ihre Operationen sind jetzt nicht mehr bloß Kommunikationen in Interaktionssystemen, sondern gehören nun zu den Elementaroperationen digitaler Organisationen. Damit sind aber, wie oben gesagt, nicht einfach Entscheidungen gemeint, da Entscheidungen in höchst unterschiedlichen Sozialsystemen auftreten können. Als Elementaroperationen können nur im Rahmen der Mitgliedschaft *kollektiv verbindliche Entscheidungen* fungieren, die alle zukünftigen Kommunikationen im System binden. Und die Algorithmen sind dann nicht mehr nur semantische Artefakte „Personen“, sondern werden nun durch das Kollektiv als „Mitglieder“ oder sogar als (formal zur Außenvertretung berechtigte) „Organe“ konstituiert, die in eigener Verantwortung für die Mensch-Algorithmus-Assoziation entscheiden.

Mit ihrer Personifizierung überschreiten entsprechend auch die Mensch-Algorithmus-Assoziationen die von Luhmann (1984, 273) nur für Organisationen reservierte „Schwelle in der Evolution höherer Komplexität“. Nicht nur digitale

[21] Zu den Motiven für die Personifizierung in „traditionellen“ Gesellschaften, Ewald (1995). Für eine historische Typologie von personae, Dijk (2020).

Aktanten, wie oben gezeigt, sondern auch digitale Hybride können jetzt als eigenständige Handlungsträger mit ihren Umwelten im strengen Sinne kommunizieren (Teubner 2006). Dafür dass ihnen selbst Entscheidungen zugerechnet werden, bilden auch Mensch-Algorithmus-Assoziationen Einrichtungen kollektiven Handelns aus, die dafür sorgen, dass bestimmte Handlungen als für die Assoziation verbindlich ausgewählt werden. *So erzeugen digitale Hybride kollektive Bindungen, die im Innenverhältnis den Sinn der algorithmischen oder menschlichen Einzelhandlungen bestimmen und im Außenverhältnis die Hybride als Ganze verpflichten*. In der Wirtschaft zum Beispiel fungieren Mensch-Algorithmus-Assoziationen selbst als Marktakteure (und nicht etwa bloß ihre algorithmischen oder menschlichen Mitglieder), denen Zahlungsakte, Eigentum, Schulden, Konkurse kollektiv zugerechnet werden und denen kollektive Verantwortlichkeit und finanzielle Haftung auferlegt werden (Linardatos 2021, 177 ff.). Neben Vertretungsregeln, Entscheidungskompetenzen und Abstimmungsverfahren müssten also in Hybriden bei einer solchen Neuverteilung der Verantwortung für Unsicherheitsabsorption auch neue kollektive Verantwortlichkeiten institutionalisiert werden. Wenn ja, in welcher Form?

2. Hierarchie: Zentralisierte Kollektivverantwortlichkeit

Etablieren sich hier nun Mechanismen zentralisierter Verantwortlichkeit wie sie in hierarchischen Organisationen vorherrschend sind? Die beteiligten Menschen und Algorithmen handeln zwar in eigener Verantwortung, aber nicht im eigenen Namen, sondern als bloße Mitglieder „für“ den Hybrid als neuartige Organisationseinheit, die gegenüber den Mitgliedern eine hierarchisch übergeordnete Position einnimmt. Sie tun dies in der gleichen Weise, wie Manager in einem Unternehmen nicht im eigenen Namen, sondern als Agenten im Auftrag ihres Prinzipals, also für die Organisation, handeln (Jensen/Meckling 1976). Und es entstehen Interessenkonflikte zwischen den Mitgliedern und der Mensch-Algorithmus-Assoziation, ähnlich den bekannten Agency-Problemen von Organisationen, für deren Lösung zahlreiche Normen institutionalisiert worden sind (Fama/Jensen 1983). Für die Konflikte in den Mensch-Algorithmus-Assoziationen dürften sich im Laufe der Zeit analoge Regeln etablieren (dazu Krafft et al. 2020, unter 5.) Auf den ersten Blick erscheint es zwar kontraintuitiv, Mensch-Algorithmus-Assoziationen eigene Präferenzen zuzuschreiben. Aber wenn autonome Algorithmen in der Lage sind, unter vorprogrammierten Zielen unterschiedliche Mittel zu wählen, entsteht unweigerlich dieses Konfliktpotenzial (Zekos 2021, 139 ff.). Und vergleichbare institutionelle Normen – etwa Pflichten und Verantwortlichkeiten des Managements, die Ultra-Vires-Doktrin des Repräsentanten, die Prüfung der Repräsentativität bei Sammelklagen – sind notwendig, um das Agency-Problem auch in digitalen Hybriden einzudämmen (Daniel 2004).

Die Personifizierung des Hybriden als eigenständigen Akteur eröffnet eine Kollektivperspektive, die nicht mehr wie die Individualperspektive die Handlungen ausschließlich „persönlich“ dem einzelnen Menschen oder dem einzelnen Algorithmus zurechnen würde (Schuppli 2014, 4 f.). Im Gegensatz zur Individualzurechnung ist die Kollektivzurechnung in der Lage, den kollektiven Eigenschaften der Mensch-Algorithmus-Assoziation in einem doppelten Sinne gerecht zu werden (Kersten 2015, 4 ff.; Ingold 2014, 220 ff.). Denn sie berücksichtigt erstens die interne Dynamik der Mensch-Algorithmus-Interaktionen, die über die Handlungen einzelner Akteure hinaus auch die Effekte ihrer Kooperation ergreift.

Zweitens wird sie der neuen Qualität der Außenbeziehungen gerecht. Nun kommuniziert die Mensch-Algorithmus-Assoziation selbst mit ihren Umwelten. Es ist nicht mehr entweder der einzelne Mensch oder der einzelne Algorithmus, dem die externe Kommunikation zugeschrieben wird. In beiderlei Hinsicht wird den neuartigen Risiken der unentwirrbaren Verflechtung von menschlichen und algorithmischen Handlungen entgegengewirkt, indem die Mensch-Algorithmus-Assoziation zum gemeinsamen Zurechnungspunkt für Handlungen und für Verantwortung etabliert wird.

Auch für Mensch-Algorithmus-Assoziationen beansprucht das Arrow-Theorem Geltung (Arrow 1950). Ihre kollektiven Entscheidungen können nicht einfach als Aggregation der individuellen Präferenzen von Menschen und Algorithmen ausgerechnet werden. Der Eigensinn des Kommunikationsgeschehens schließt eine solche individualisierende Betrachtung aus und erzeugt das für Organisationen typische Intransparenzrisiko der internen Kommunikation. Und die Mitwirkung von Algorithmen in Organisationen verschärft die Intransparenz drastisch. Bostrom (2017, 58 ff., 65 ff.,155 ff.) analysiert dieses Risiko mit dem Begriff der „kollektiven Intelligenz“ oder gar der „kollektiven Superintelligenz“. Mensch-Algorithmus-Interaktionen sind nicht vollständig kontrollierbar, was zu einer „perverse instantiation“ führt: Ein autonomer Algorithmus als Mitglied des Hybriden erfüllt effektiv das vom menschlichen Teilnehmer gesetzte Ziel, wählt jedoch ein Mittel, das die Intentionen des Menschen verletzt.

Wenn es nun aber nicht mehr bloß um die Verteilung der Verantwortung, sondern auch um die Zurechnung von Verantwortlichkeit auf die drei Aktionszentren, Mensch, Algorithmus oder Assoziation, geht, wird es noch unübersichtlicher. Wie oben diskutiert, ist es in Hybriden schwierig, das risikorealisierende Ereignis und damit den individuellen Verantwortungsträger zu identifizieren. Es ist prinzipiell nicht mehr aufklärbar: War die menschliche Handlung fehlerhaft oder die algorithmische Kalkulation? Der Ausweg dürfte sein, dass sich angesichts der neuen Verteilung von Verantwortung auch neue Zurechnungsformen institutionalisieren, die den Hybriden selbst als Kollektivakteur verantwortlich machen.

In der Moralphilosophie wird eine besondere Form der Kollektivverantwortlichkeit unter der „collective moral autonomy thesis“ diskutiert. Danach gibt es paradoxe Situationen, in denen ein Fehlverhalten identifiziert und die Verantwortlichkeit dafür einem Kollektiv zugerechnet wird, obwohl dessen einzelnen Mitgliedern kein Fehlverhalten vorzuwerfen ist (Copp 2007). Jenseits der Intentionen und Handlungen der Mitglieder, verfügt das Kollektiv entsprechend über eigene Intentionen, handelt autonom und entwickelt Informationen über die möglichen Ergebnisse ihrer Aktivitäten. Hierdurch wird ausschließlich das Kollektiv zum Zurechnungspunkt der Verantwortung (Pettit 2007; Corlett 2002, 575). Diese paradoxe Situation besteht auch für die Mensch-Algorithmus-Assoziation, die eine eigene phänomenale Innenperspektive entwickelt. Sie erzeugt eine eigene Hierarchie von Präferenzen, sozialen Bedürfnissen und politischen Interessen. All dies lässt sich nicht auf einen der beteiligten Akteure reduzieren (Rahwan et al. 2019, 482 f.).

Die entscheidende Frage lautet dann: Können die gesellschaftlichen Zurechnungsformen eine „Cyborg-Wende“ verkraften, d. h. eine Entwicklung, in der sich Personen und Dinge zu einer noch nie dagewesenen Existenzform verbinden, die der schleichenden Hybridisierung von Menschen und Nicht-Menschen folgt (Viljanen 2017, 1278)? Eine mögliche Antwort lautet: Die Gesellschaft wird einer Mensch-Algorithmus-Assoziation vollen Kollektivakteursstatus zusprechen und ihr

selbst die Verantwortung für die Handlungen ihrer algorithmischen und menschlichen Mitglieder, zugleich auch die Verantwortlichkeit für Fehlentscheidungen zuschreiben. Dies ist in der Tat eine ernstzunehmende Option. Sie entspricht der zentralisierten Verantwortlichkeitszurechnung in formalen Organisationen, ist allerdings für den besonderen Fall der algorithmischen Kommunikation noch wenig praktisch erprobt.

Für eine solche zentralisierte kollektive Verantwortlichkeit sprechen sich neben wenigen organisationssoziologischen Autoren (z.B. Grothe-Hammer 2019, 9) hauptsächlich Vertreter der KI-Ethik aus. Systeme künstlicher Intelligenz, so argumentiert Heinrichs (2020, 171), „werden in enger Kooperation mit ihren menschlichen Nutzern gemischte Agenten bilden, Systeme, deren Handlungen nicht einfach dem menschlichen Nutzer oder der Programmierung der KI zugeschrieben werden können, sondern vielmehr aus ihrer fortlaufenden Interaktion hervorgehen". In der digitalen Ethik gibt es ähnliche Ansätze. Loh/Loh (2017) sehen die Mensch-Algorithmus-Assoziation als ein hybrides System an, das aufgrund des Kooperationszwecks selbst verantwortlich wird. Zu ähnlichen Schlussfolgerungen gelangt die Attributionstheorie, wenn sie Dennetts Konzept der „intentional stance" auch auf Hybride als „einer Kombination aus menschlichen, elektronischen und organisatorischen Komponenten" anwendet (Sartor 2009, 264).

In der Tat diskutieren mehrere Autoren für ein zukünftiges KI-Recht die Möglichkeit, digitale Hybride in kollektiv-zentralisierte Haftung zu nehmen. Taylor und De Leeuw (2020, 6) schlagen vor, Personalität aus einer relationalen Perspektive neu zu denken. Hybride KI-Systeme sollten als eine neue Erfahrungsform betrachtet werden, welche die emotionalen, rationalen und sozialen Beziehungen der Menschen zu ihnen selbst, zu Maschinen und zur Welt beeinflusst. Für diese neue Form der algorithmischen Hybridisierung von menschlichem und algorithmischem Fühlen und „Handeln" prognostizieren sie neue Formen von Zurechnung, Handlung und Verantwortlichkeit (ebenso Gruber 2020, 267). Dies hat zur Konsequenz, dass Hybride als neuartige rechtliche Einheiten behandelt werden und diesen Einheiten direkt die Verantwortlichkeit für das Verhalten der beiden untrennbaren Akteure zugeschrieben wird (Dahiyat 2021, 78 ff.; Perlingieri 2020, 175 f.; Asaro 2014, 265). Damit würden individuelle Handlungen der Menschen und Algorithmen in kollektiven Handlungen aufgehen. So könnten die Hybride bindende vertragliche Bindungen schaffen und es könnten eventuell Haftungsansprüche gegen das Hybrid anerkannt werden.

3. Netzwerk: Distribuierte Kollektivverantwortlichkeit

Eine solche kollektive Verantwortlichkeit von Hybriden verlangt aber institutionelle Voraussetzungen, die heute (noch) nicht erfüllt sind. Die wirtschaftliche Praxis müsste die Hybride mit ausreichenden finanziellen Ressourcen ausstatten oder das Recht müsste eine Pflichtversicherung anordnen. Dann erst würde eine zentralisierte Verantwortlichkeitsform für die Hybriden effektiv. Deshalb wird wahrscheinlich eher eine Verantwortlichkeitsform institutionalisiert werden, die stärker an die bilateral-individuellen Beziehungen zwischen Mensch und Algorithmus anknüpft, zugleich aber die multilateral-kollektiven Eigenschaften der Mensch-Algorithmus-Assoziation berücksichtigt. Hier sind Netzwerktheorien einschlägig, da sie digitale Hybride zwar als multiple bilaterale Beziehungen zwischen Menschen und Algorithmen konzipieren, gleichzeitig aber auch als übergreifende Verbünde. Neben den multilateralen Verknüpfungen wird zugleich die organisatorische Einheit

des Hybriden abgebildet und einer (begrenzten) Personifizierung zugänglich: „Networks produce hybrids, and here, non-humans attain aspects of personhood“ (Sprenger 2018, 258).

Der Zweckbegriff ist der Dreh- und Angelpunkt, der es der Netzwerktheorie erlaubt, eine dritte Zurechnungsform einzuführen, die sich von individueller ebenso wie von zentralisierter kollektiver Zurechnung unterscheidet (Powell 1990; ebenso Cafaggi und Iamiceli 2015). Sie begreift Netzwerke weder als Märkte noch als Hierarchien, sondern als eigenständige soziale Systeme, die zwischen – oder jenseits von – multilateralen Beziehungen einerseits und hierarchischen Organisationen andererseits operieren. Ein dafür sensibler Zweckbegriff registriert, dass sich die Netzwerkteilnehmer auf eine paradoxe Doppelorientierung einstellen müssen (Wielsch 2020, 191; Windeler 2001, 237 ff.). Es wird ihnen die schwierige Synthese zugemutet, in ein und derselben Operation sowohl ihre eigenen individuellen Zwecke zu verfolgen als auch den übergeordneten Netzwerkzweck zu verwirklichen. Dieser netzwerkeigene Zwang zur Doppelorientierung steht im deutlichen Gegensatz zu hierarchischen Organisationen, deren Leitungsfunktionen sich nicht an Individualinteressen, sondern ausschließlich am Organisationszweck orientieren. Und er steht im Gegensatz zu rein multilateralen Verträgen, deren Austauschzweck auf die legitime Verfolgung von Individualinteressen gerichtet ist. Netzwerke sind also durch ein ständiges paradoxes Oszillieren zwischen individueller und kollektiver Zweckorientierung gekennzeichnet.

Parallel dazu wäre nun auch die Mensch-Algorithmus-Assoziation als eine eher dezentral verfasste Netzwerkorganisation zu qualifizieren (Lior 2022), die eine gleichzeitige individuelle und kollektive Orientierung institutionalisiert. Maschinen und Menschen werden zwar jeweils als individuell und nach ihrer eigenen Logik handelnd begriffen. Zugleich sind sie jedoch an einem übergreifenden Kooperationszweck orientiert, der den kollektiven Mensch-Algorithmus-Handlungen Orientierung liefert.

Mensch-Algorithmus-Assoziationen als Netzwerke zu qualifizieren, hat aber vor allem Auswirkungen auf die Verantwortlichkeitszurechnung. Wenn die genannte doppelte Zweckorientierung ein Netzwerk konstituiert, das sowohl Algorithmen als auch menschliche Teilnehmer umfasst, wird zwangsläufig eine andere als die zentralisierte Form der Kollektivverantwortlichkeit institutionalisiert – eine „Netzwerkverantwortlichkeit“.[22] In der KI-Ethik plädiert Neuhäuser (2015, 143 f.) in diesem Sinne für eine kollektive moralische Verantwortlichkeit von Mensch-Algorithmus-Assoziationen als „Verantwortlichkeitsnetzwerke“.

An dieser Stelle ist erneut auf die wichtige Unterscheidung zwischen Verantwortung und Verantwortlichkeit zurückzukommen. Die Divergenz von Verantwortung als Unsicherheitsabsorption und Verantwortlichkeit als normative Reaktion auf Fehler ist geeignet, das zentrale Problem der Netzverantwortlichkeit zu lösen. Denn bestimmte Handlungskapazitäten sind zwar in der Netzwerkorganisation als Kollektivakteur gebündelt, die finanziellen Ressourcen und Kontrollkapazitäten aber bleiben dezentral auf die Netzknoten verteilt. Deswegen wird mit Recht vermutet, dass „soziale Netzwerke verantwortungslose Gebilde“ sind (Kette/Tacke 2015, 258). Damit sie dennoch in die Pflicht genommen werden können, wird in einem ersten Schritt die Zurechnung von Handlung und die Verantwortung für Unsicherheitsabsorption auf das Kollektiv übertragen und damit die ausweglose

[22] Zu drei Dimensionen einer genuinen Netzwerkhaftung im allgemeinen Teubner (2004), Kap. 4-6.

Identifikation der Einzelhandlungen umgangen. In einem zweiten Schritt wird dann aber die Verantwortlichkeitszurechnung dezentralisiert. Die Handlungszurechnung zielt auf den Hybriden als solchen, was aber wegen des Mangels an eigenen Ressourcen nicht zur Folge hat, dass der Hybrid als handelnde Einheit finanziell einzustehen hat. Die kollektive Verantwortungszurechnung auf den Hybriden dient also letztlich dazu, die Verantwortlichkeitszurechnung auf eine Reihe von außenstehenden Akteuren zu kanalisieren.[23] Zur Klarstellung sei gesagt, dass eine solche Dezentralisierung kollektiver Verantwortlichkeit nicht etwa bedeutet, dass nun jeder Handlungsträger doch wieder für seinen individuellen Handlungsbeitrag einzustehen hat. Vielmehr wird die Verantwortung kollektiviert, die Verantwortlichkeit für die Kollektivhandlung aber ist von jedem der Netzbeteiligten zu tragen.

Dies führt zu der Frage, wer von den Netzwerkteilnehmern für riskante Algorithmenentscheidungen verantwortlich gemacht werden kann. Die Antwort formulieren führende Vertreter der KI-Philosophie Floridi und Sanders (2011) treffend als distribuierte Verantwortlichkeit. Genauer gesagt:

> „Die Auswirkungen von KI-basierten Entscheidungen oder Handlungen beruhen auf unzähligen Interaktionen zwischen vielen Akteuren, darunter Designer, Entwickler, Benutzer, Software und Hardware. ... Auf distribuierte Handlungsfähigkeit folgt distribuierte Verantwortlichkeit“ (Taddeo/Floridi 2018, 751, unsere Übersetzung).

Bei Mensch-Algorithmus-Hybriden ist der Verband von einem Satellitennetzwerk unabhängiger Akteure umgeben, den Benutzern, Betreibern, Händlern, Herstellern, Programmierern. Würden diese Akteure in eine hierarchische Organisation integriert, dann wäre diese, wie oben gezeigt, kollektiv verantwortlich. Schließlich gehört es zu den Aufgaben des Managements, die Schnittstellen der verschiedenen Aktionen zu koordinieren. Würden die Akteure dagegen in reinen Marktbeziehungen interagieren, würde sich das Verantwortlichkeitsrisiko auf die Kunden verlagern, die für die Koordination der Teilleistungen selbst verantwortlich sind. In unserem Fall eines digitalen Hybrids mit einem umgebenden Akteur-Netzwerk überzeugt jedoch weder die ausschließliche Verantwortlichkeitszurechnung an den Hersteller noch die an die Netzwerkzentrale noch die an den menschlichen Akteur noch die an die algorithmischen Netzwerkknoten. Solche individuellen Verantwortlichkeitszuweisungen sind willkürlich, insbesondere wenn sie die strukturelle Diffusion von Verantwortung in kooperativen Netzwerken ignorieren.

Folglich würde nach der kollektiven Zurechnung der Verantwortung für Einzelhandlungen an den Hybriden die Zurechnung der Verantwortlichkeit, besonders der finanziellen Haftung, auf die Mitglieder des umgebenden organisierten Vertragsnetzes distribuiert werden, die von den Aktivitäten des Hybriden profitieren, also auf die Betreiber, Eigentümer, Hersteller und Lieferanten der elektronischen Technologie. Diese Lösung findet in der KI-Ethik und dem KI-

[23] Vorbilder für eine solche Haftungskanalisierung, die zunächst die Verantwortung für Einzelhandlungen einem Handlungsverbund zurechnet, dann aber die Verantwortlichkeit und Haftung dezentral distribuiert, finden sich in der enterprise liability des US-amerikanischen Rechts. Für analoge Anwendung auf digitale Hybride plädieren Beckers/Teubner (2022), 99 ff.; Linardatos (2021), 177 ff. und Vladeck (2014), 149.

Haftungsrecht einige Unterstützung (Lior 2022; Navas 2020, 169 f.; Chinen 2019, 83; Vladeck 2014, 129; Allain 2013, 1074). Ähnlich empfiehlt die europäische Expertengruppe zur Haftung für neuen Technologien (2015, 55 f.) eine gemeinsame Haftung aller Akteure, die mit der „kommerziellen und technologischen Einheit“ des Hybriden verbunden sind. Betreiber, Hersteller, Händler und Programmierer des Software-Agenten sind in einem solchen Haftungsnetzwerk gebündelt. Womöglich wäre die systemische Netzwerktheorie (Bommes/Tacke 2007) in der Lage, in Zukunft geeignete konzeptuelle Instrumente zu entwickeln, welche die Grenzen der Verantwortlichkeitseinheit sowie die relative Beteiligung der Netzwerkknoten an der Verantwortungsdynamik identifizieren.

Literatur

Ahrne, Göran/Brunsson, Nils (2011): Organization outside Organizations: The Significance of Partial Organization. Organization 18: 83-104.

Alkemeyer, Thomas/Bröckling, Ulrich/Peter, Tobias (Hrsg.) (2018): Jenseits der Person: Zur Subjektivierung von Kollektiven. Bielefeld: transcript.

Allain, Jessica S. (2013): From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems, Louisiana Law Review 73, 1049-1079.

Apelt, Maja/Besio, Cristina/Corsi, Giancarlo/von Groddeck, Victoria/Grothe-Hammer, Michael/Tacke, Veronika (2017): Resurrecting Organization without Renouncing Society: A Response to Ahrne, Brunsson and Seidl. European Management Journal 35: 8-14.

Armour, John/Eidenmüller, Horst (2019): Self-Driving Corporations? ECGI-Law Working Paper 475/2020, 1-44.

Arrow, Kenneth (1950): A Difficulty in the Concept of Social Welfare. Journal of Political Economy 58, 328-346.

Asaro, Peter M. (2014): Determinism, Machine Agency, and Responsibility. Politica e Società 2, 265-292.

Baecker, Dirk (2011): Who Qualifies for Communication? A Systems Perspective on Human and Other Possibly Intelligent Beings Taking Part in the Next Society. TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 20, 17-26.

Baecker, Dirk (2020a): Digitization as Calculus: A Prospect. Research Proposal, abrufbar auf Researchgate, 1-11.

Baecker, Dirk (2020b): Soziale Welt, Sonderband: „Soziologie 4.0 und ihre Vorläufer, Soziologie des Digitalen – Digitale Soziologie?“ Baden-Baden: Nomos, 19-45.

Becker Markus/Pordzik, Philipp (2020) Digitalisierte Unternehmensführung, Zeitschrift für die gesamte Privatrechtswissenschaft, 334-357.

Beckers, Anna/Teubner, Gunther (2021): Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds. Oxford: Hart.

Blackstone, William (1771): Commentaries on the laws of England: In four books. Philadelphia: Robert Bell.

Block, Katharina/Dickel, Sascha (2020) Jenseits der Autonomie: Die De/Problematisierung des Subjekts in Zeiten der Digitalisierung. Behemoth 13, 109-131.

Bommes, Michael/Tacke, Veronika (2007): Netzwerke in der Gesellschaft der Gesellschaft. Funktionen und Folgen einer doppelten Begriffsverwendung. Soziale Systeme 13, 9-20.

Bora, Alfons (2013): Vanishing Points: Responsibility as a Normative Shifting Symbol and the Search for Social Addressability. Soziale Systeme 19, 456-469.

Bora, Alfons (2019): Kommunikationsadressen als digitale Rechtssubjekte, Verfassungsblog, 1. Oktober 2019.

Bostrom, Nick (2017): Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

Braun, Matthias/Hummel, Patrik/Beck, Susanne/Dabrock, Peter (2020): Primer on an Ethics of AI-Based Decision Support Systems in the Clinic. Journal of Medical Ethics 46, 1–8.

Cafaggi, Fabrizio/Iamiceli, Paola (2015): Private Regulation and Industrial Organization: Contractual Governance and the Network Approach. S.341-374 in: Stefan Grundmann/Florian Möslein/Karl Riesenhuber (Hrsg.), Contract Governance: Dimensions in Law and Interdisciplinary Research. Oxford: Oxford University Press.

Büchner, Stefanie (2018): Zum Verhältnis von Digitalisierung und Organisation. Zeitschrift für Soziologie 47, 332-348.

Chinen, Mark A. (2019): Law and Autonomous Machines. Cheltenham: Elgar.

Coeckelbergh, Mark (2012): Moral Responsibility, Technology, and Experiences of the Tragic: From Kierkgeaard to Offshore Engineering. Science and Engineering Ethics 18, 35-48.

Cohen, Julie E. (2019): Between Truth and Power: The Legal Constructions of Informational Capitalism. Oxford: Oxford University Press.

Coleman, James S. (1990): Foundations of Social Theory. Cambridge/Mass.: Harvard University Press.

Copp, David (2007): The Collective Moral Autonomy Thesis. Journal of Social Philosophy 38, 369-388.

Corlett, J. Angelo (2002): Collective Moral Responsibility. Journal of Social Philosophy 32, 573-584.

Dahiyat, Emad (2021) Law and Software Agents: Are They „Agents" by the Way? Artificial Intelligence and Law 29, 59-86.

Daniel, Juanda L. (2004): Electronic Contracting under the 2003 Revisions to Article 2 of the Uniform Commercial Code: Clarification or Chaos? Santa Clara Computer & High Technology Law Journal 20, 319- 346.

Dennett, Daniel (1987): The Intentional Stance. Cambridge/Mass.: MIT Press.

Diakopoulos, Nicholas (2019): Automating the News: How Algorithms are Rewriting the Media. Cambridge/Mass.: Harvard University Press.

Dijk, Niels van (2020): In the Hall of Masks: Contrasting Modes of Personification, S.230-251 in: Mireille Hildebrandt/Kieron O'Hara (Hrsg.), Life and the Law in the Era of Data-Driven Agency. Cheltenham: Edward Elgar.

Dörr, Konstantin (2017): Algorithmischer Journalismus: Eine Analyse der automatisierten Textproduktion im Journalismus auf gesellschaftlicher, organisatorischer und professioneller Ebene. Zurich: University of Zurich Main Library.

Durkheim, Emile (2004 [1883]): Über die Teilung der sozialen Arbeit. Frankfurt: Suhrkamp.

Esposito, Elena (2017): Artificial Communication? The Production of Contingency by Algorithms. Zeitschrift für Soziologie 46, 249-265.

Elena Esposito (2022): Artificial Communication: How Algorithms Produce Social Intelligence. Cambridge, Mass.: MIT.

Esposito, Roberto (2022): Institution und Biopolitik, Zürich: Diaphanes.

Europäische Expertengruppe zu Haftung und Neuen Technologien (2018): Bericht „Liability for Artificial Intelligence and Other Emerging Technologies".

Europäisches Parlament (2017): Bericht mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich der Robotik (2015/2103 (INL)).

Ewald, William (1995): Comparative Jurisprudence (I): What Was It Like to Try a Rat. American Journal of Comparative Law 143, 1889-2149.

Fama, Eugene/Jensen, Michael (1983): Agency Problems and Residual Claims. Journal of Law and Economics 26, 327-349.

Feenberg, Andrew (2017) Technosystem: The Social Life of Reason. Cambridge, MA: Harvard University Press.

Floridi, Luciano/Sanders, J.W. (2011): On the Morality of Artificial Agents, S.184-212 in: Michael Anderson/Susan L. Anderson (Hrsg.), Machine Ethics. Cambridge: Cambridge University Press.

Fuchs, Peter (1991): Kommunikation mit Computern? Zur Korrektur einer Fragestellung. Sociologia Internationalis 29, 1-30.

Gellers, Joshua C. (2021): Rights for Robots: Artificial Intelligence, Animal and Environmental Law. London: Routledge.

Gierke, Otto von (1902): Das Wesen der menschlichen Verbände. Leipzig: Duncker & Humblot.

Grothe-Hammer, Michael (2019): Zu den Konsequenzen des Ersetzens menschlicher Entscheiderinnen und Entscheider durch Algorithmen in Organisationen: Beitrag zur Ad-Hoc-Gruppe „Macht Digitalisierung Organisationen überflüssig? Aktuelle Herausforderungen nicht nur für die Organisationssoziologie" in: Verhandlungen des 39. Kongresses der Deutschen Gesellschaft für Soziologie in Göttingen.

Gruber, Malte-Christian (2020): Why Non-Human Rights? Law & Literature 32, 263-270.

Gurevich, Yuri (2012): What Is an Algorithm? SOFSEM 2012: Theory and Practice of Computer Science, 31-42.

Gutmann, Mathias (2011): Sozialität durch technische Systeme? TATuP - Journal for Technology Assessment in Theory and Practice 20, 11-16.

Guzman, Andrea L. /Lewis, Seth C. (2020): Artificial Intelligence and Communication: A Human-Machine Communication Research Agenda. New Media and Society 22, 70-86.

Harth, Jonathan/Lorenz, Caspar-Fridolin (2017): „Hello World" – Systemtheoretische Überlegungen zu einer Soziologie des Algorithmus. kommunikation@gesellschaft 18, 1-21.

Haselager, Pim (2013): Did I Do that? Brain-Computer Interfacing and the Sense of Agency. Minds & Machines 23, 405-418.

Hauriou, Maurice (1965) [1933]: Die Theorie der Institution. Berlin: Duncker & Humblot.

Heinrichs, Jan-Hendrik (2020): Artificial Intelligence in Extended Minds: Intrapersonal Diffusion of Responsibility and Legal Multiple Personality, S.159-176 in: Birgit Beck/Michael Kühler (Hrsg.), Technology, Anthropology, and Dimensions of Responsibility, Heidelberg: Springer.

Hepp, Andreas (2020): Deep Mediatization: Key Ideas in Media & Cultural Studies. London: Routledge.

Hildebrandt, Mireille (2015): Smart Technologies and the End(s) of Law. Cheltenham: Elgar.

Hoyningen-Huene, Paul (1994): Zu Emergenz, Mikro- und Makrodetermination, S. 165–195 in: Weyma Lübbe (Hrsg.): Kausalität und Zurechnung. de Gruyter, Berlin 1994.

Ingold, Albert (2014): Grundrechtsschutz sozialer Emergenz: Eine Neukonfiguration juristischer Personalität in Art. 19 Abs. 3 GG angesichts webbasierter Kollektivitätsformen. Der Staat 53, 193-226.

Jasanoff, Sheila (2004): The Idiom of Co-Production. S.1-12 in: Sheila Jasanoff (Hrsg.), States of Knowledge: The Co-production of Science and the Social Order. London: Routledge.

Jensen, Michael/Meckling, William H. (1976): Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. Journal of Financial Economics 3, 306-360.

Kersten, Jens (2015): Menschen und Maschinen: Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen. Juristenzeitung 70, 1-8.

Kersten, Jens (2016) Die maschinelle Person: Neue Regeln für den Maschinenpark? S.89-106 in: Fabian Karsch/ArneManzeschke (Hrsg.), Roboter, Computer und Hybride. Baden-Baden: Nomos.

Kette, Sven /Tacke, Veronika (2015): Systemtheorie, Organisation und Kritik, S.232-265 in: Albert Scherr (Hrsg.), Systemtheorie und Differenzierungstheorie als Kritik: Perspektiven in Anschluss an Niklas Luhmann. Weinheim: Beltz.

Kleiner, Cornelius (2021) Die elektronische Person: Entwurf eines Zurechnungs- und Haftungssubjekts für den Einsatz autonomer Systeme im Rechtsverkehr. Baden-Baden: Nomos.

Krafft, Tobias D., Zweig, Katharina A., König, Pascal D. (2020) „How to Regulate Algorithmic Decision-Making: A Framework of Regulatory Requirements for Different Applications“, 16 Regulation & Governance, 119-136

Kühl, Stefan „Gruppen, Organisationen, Familien und Bewegungen: Zur Soziologie mitgliedschaftsbasierter Systeme zwischen Interaktion und Gesellschaft“, Zeitschrift für Soziologie Sonderheft 2014, 65-85.

Latour, Bruno (2001): Das Parlament der Dinge: Für eine politische Ökologie. Frankfurt: Suhrkamp.

Linardatos, Dimitrios (2021): Autonome und vernetzte Aktanten im Zivilrecht: Grundlinien zivilrechtlicher Zurechnung und Strukturmerkmale einer elektronischen Person. Tübingen: Mohr Siebeck.

Lior, Anat (2022): The AI Accident Network: Artificial Intelligence Liability Meets Network Theory. Tulane Law Review 95, 1103-1166.

Loh, Wulf/Loh, Janina (2017): Autonomy and Responsibility in Hybrid Systems. S.35-50 in: Patrick Lin/Ryan Jenkins/Keith Abney (Hrsg.), Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. Oxford: Oxford University Press.

Lohse, Simon (2011) Zu Emergenz des Sozialen bei Niklas Luhmann. 40 Zeitschrift für Soziologie, 190-207.

Lorentzen, Kai F. (2002): Luhmann Goes Latour: Zur Soziologie hybrider Beziehungen. S.101-118 in: Werner Rammert/Ingo Schulz-Schaeffer (Hrsg.), Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik. Frankfurt: Campus.

Luhmann, Niklas (1964) Funktionen und Folgen formaler Organisation, 3. Aufl. 1976 Berlin: Duncker & Humblot.

Luhmann, Niklas (1965) Grundrechte als Institution: Ein Beitrag zur politischen Soziologie. Berlin: Duncker & Humblot.

Luhmann, Niklas (1984): Soziale Systeme: Grundriß einer allgemeinen Theorie. Frankfurt: Suhrkamp.

Luhmann, Niklas (1986): Systeme verstehen Systeme. S.72-117 in: Niklas Luhmann/Eberhard Schorr, Zwischen Intransparenz und Verstehen: Fragen an die Pädagogik. Frankfurt: Suhrkamp.

Luhmann, Niklas (1987): Läßt unsere Gesellschaft Kommunikation mit Gott zu? S.227–235 in: Niklas Luhmann, Soziologische Aufklärung 4: Beiträge zur funktionalen Differenzierung der Gesellschaft. Opladen: Westdeutscher Verlag.

Luhmann, Niklas (1987 [1972]): Rechtssoziologie. Opladen: Westdeutscher Verlag.

Luhmann, Niklas (1997): Die Gesellschaft der Gesellschaft. Frankfurt: Suhrkamp.

Luhmann, Niklas (2000a) Organisation und Entscheidung, Opladen: Westdeutscher Verlag.

Luhmann, Niklas (2000b): Die Politik der Gesellschaft. Frankfurt: Suhrkamp.

Luhmann, Niklas (2014 [1975]): Ebenen der Systembildung – Ebenendifferenzierung (unveröffentlichtes Manuskript 1975). Zeitschrift für Soziologie, Sonderheft: „Interaktion – Organisation – Gesellschaft revisited“, 6-39.

Luhmann, Niklas (2018): Systemtheorie der Gesellschaft. Frankfurt: Suhrkamp.

McLuhan, Marshall (2003): Understanding Media: The Extensions of Man. Corte Madera: Gingko Press.

Messner, Claudius (2020): Listening to Distant Voices. International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique 33, 1143-1173.

Mosco, Gian Domenico (2020): AI and the Board within Italian Corporate Law: Preliminary Notes. European Company Law Journal 17, 87-96.

Möllers, Christoph (2021): Herr, Knecht und Maschine in der künftigen Rechtsphilosophie. S.184-195 in: Thomas Khurana, Dirk Quadflieg, Francesca Raimondi, Juliane Rebentisch und Dirk Setton (Hrsg.), Negativität: Kunst, Recht, Politik. Berlin: Suhrkamp.

Möslein, Florian (2017): Robots in the Boardroom: Artificial Intelligence and Corporate Law, S.649-670 in: Woodrow Barfield/Ugo Pagallo (Hrsg.), Research Handbook on the Law of Artificial Intelligence. Cheltenham: Edward Elgar.

Muhle, Florian (2018): Sozialität von und mit Robotern? Drei soziologische Antworten und eine kommunikationstheoretische Alternative. Zeitschrift für Soziologie 47, 147-163.

Muhle, Florian (2020) „Roboter in der Sozialwelt: Überlegungen und Einsichten zum Subjektstatus humanoider Roboter“, in: Michael Schetsche u.a. (Hrsg.) Intersoziologie: Menschliche und nichtmenschliche Akteure in der Sozialwelt, Weinheim: Beltz Juventa, 128-142

Nake, Frieder (2020): Surface, Interface, Subface: Three Cases of Interaction and One Concept. S.92-109 in: Uwe Seifert/Hyun Kim Jin/Anthony Moore (Hrsg.), Paradoxes of Interactivity. Bielefeld: transcript.

Nassehi, Armin (2019): Muster: Theorie der digitalen Gesellschaft. München: Beck.

Navas, Susana (2020): Robot Machines and Civil Liability. S.157-173 in: Martin Ebers/Susana Navas (Hrsg.), Algorithms and Law. Cambridge: Cambridge University Press.

Neuhäuser, Christian (2015): Some Sceptical Remarks Regarding Robot Responsibility and a Way Forward. S.131-147 in: Catrin Misselhorn (Hrsg.), Collective Agency and Cooperation in Natural and Artificial Systems. Heidelberg: Springer.

North, Douglass C. (1990): Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.

Ortmann, Günther (2017): Für ein Unternehmensstrafrecht: Sechs Thesen, sieben Fragen, eine Nachbemerkung. Neue Zeitschrift für Wirtschafts-, Steuer- und Unternehmensstrafrecht 6, 241-251.

Parsons, Talcott/Shils, Edward A. (1951): Toward a General Theory of Action: Theoretical Foundations for the Social Sciences. New York: Harper & Row.

Perlingieri, Carolina (2020): Responsabilità civile e robotica medica. Tecnologie e diritto 1, 161-180.

Pettit, Philip (2007): Responsibility Incorporated. Ethics 117, 171-201.

Powell, Walter W. (1990): Neither Market nor Hierarchy: Network Forms of Organization. Research in Organizational Behavior 12, 295-336.

Prescott, Tony J. (2017): Robots are not Just Tools. Connection Science 29, 142-149.

Rahwan, Iyad/Cebrian, Manuel/Obradovich, Nick/Bongard, Josh/Bonnefon, Jean-François/Breazeal, Cynthia/Crandall, Jacob W./Christakis, Nicholas A./Couzin, Iain D./Jackson, Matthew O./Jennings, Nicholas R./Kamar, Ece/Kloumann, Isabel M./Larochelle, Hugo/Lazer, David/McElreath, Richard/Mislove, Alan/Parkes, David C./Pentland, Alex 'Sandy'/Roberts, Margaret E./Shariff, Azim/Tenenbaum, Joshua B./Wellman, Michael (2019): Machine Behaviour. Nature 568, 477-486.

Rossler, Gustav (2015): Der Anteil der Dinge an der Gesellschaft: Sozialität – Kognition – Netzwerke. Bielefeld: transkript.

Sartor, Giovanni (2009): Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents. Artificial Intelligence and Law 17, 253-290.

Schuppli, Susan (2014): Deadly Algorithms: Can Legal Codes Hold Software Accountable for Code That Kills? Radical Philosophy 187, 2-8.

Shidaro, Hirokazu/Christanikis, Nicholas A. (2017): Locally Noisy Autonomous Agents Improve Global Human Coordination in Network Experiments. Nature 545, 370-374.

Sprenger, Guido (2017): Communicated into Being: Systems Theory and the Shifting of Ontological Status. Anthropological Theory 17, 108-132.

Sprenger, Guido (2018): Production is Exchange: Gift Giving between Humans and Non-Humans. S.247-264 in: Laila Prager/Michael Prager/ Guido Sprenger (Hrsg.), Part and Wholes: Essays on Social Morphology, Cosmology, and Exchange. Münster: LIT Verlag.

Struck, Philipp (2019): Die kommunikative Konstitution fluider Organisation: Eine soziologische Untersuchung zur Relevanz von Software in der Organisation von Arbeit auf Märkten in der Plattformökonomie. Berlin: brandung.

Taddeo, Mariarosaria/Floridi, Luciano (2018): How AI can be a Force for Good: An Ethical Framework Will Help to Harness the Potential of AI while Keeping Humans in Control. Science 361, 751-752.

Taylor, Simon M./De Leeuw, Marc (2020): Guidance Systems: From Autonomous Directives to Legal Sensor-Bilities. AI & Society (Open Forum), 1-14.

Teubner, Gunther (1987): Unternehmenskorporatismus: New Industrial Policy und das Wesen der Juristischen Person. Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft 2, 61-85.

Teubner, Gunther (2001): Rechtsirritationen: Zur Koevolution von Rechtsnormen und Produktionsregimes. S.351-381 in: Günther Dux/ Frank Welz (Hrsg.), Moral und

Recht im Diskurs der Moderne: Zur Legitimation gesellschaftlicher Ordnung. Opladen: Leske und Budrich.

Teubner, Gunther (2004): Netzwerk als Vertragsverbund: Virtuelle Unternehmen, Franchising, Just in Time in sozialwissenschaftlicher und juristischer Sicht. Baden-Baden: Nomos.

Teubner, Gunther (2006): Elektronische Agenten und große Menschenaffen: Zur Ausweitung des Akteursstatus in Recht und Politik. Zeitschrift für Rechtssoziologie 27, 5-30.

Thomadsen, Tommy (2005): Hierarchical Network Design. Kongens Lyngby: Technical University of Denmark.

Thürmel, Sabine (2015): The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors. S.45-62 in: Catrin Misselhorn (Hrsg.), Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation. Heidelberg: Springer.

Trüstedt, Katrin (2020): Representing Agency. Law & Literature 32, 195-206.

Turner, Jacob (2018): Robot Rules: Regulating Artificial Intelligence. London: Palgrave Macmillan.

Viljanen, Mika (2017): A Cyborg Turn in Law? German Law Journal 18, 1277-1308.

Vladeck, David C. (2014): Machines without Principals: Liability Rules and Artificial Intelligence. Washington Law Review 89, 117-150.

Weyer, Johannes/Fink, Robin D. (2011): Die Interaktion von Mensch und autonomer Technik in soziologischer Perspektive. TATuP – Journal for Technology Assessment in Theory and Practice 20, 39-45.

Wielsch, Dan (2020): Die Ermächtigung von Eigen-Sinn im Recht. S.179-201 in: Ino Augsberg/Steffen Augsberg/Ludger Heidbrink (Hrsg.), Recht auf Nicht-Recht: Rechtliche Reaktionen auf die Juridifizierung der Gesellschaft. Weilerswist: Velbrück.

Williamson, Oliver (1985): The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. New York: Free Press.

Windeler, Arnold (2001): Unternehmungsnetzwerke: Konstitution und Strukturation. Wiesbaden: Westdeutscher Verlag.

Winner, Langdon (1980): Do Artifacts Have Politics? Daedalus 109, 121-136.

Zekos, Georgios I. (2021): Economics and Law of Artificial Intelligence: Finance, Economic Impacts, Risk Management and Governance. Cham: Springer.

Zimmerman, Evan J. (2015): Machine Minds: Frontiers in Legal Personhood, Unpubliziertes Manuskript, abrufbar über SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965.

Autoreninformationen

Anna Beckers
Assistant Professor of Private Law and Legal Methodology (tenured), Maastricht University, Faculty of Law, Max Weber Fellow, European University Institute 2015/2016, Monash University, Centre for Commercial Law and Regulatory Studies, Visiting Scholar 2019. Einschlägige Veröffentlichung: Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds. Oxford: Hart 2022; Enforcing Corporate Social Responsibility Codes: On Global Self-Regulation and National Private Law, Hart Publishing 2015
Website: https://www.maastrichtuniversity.nl/anna.beckers, Twitter: @A__Beckers

Gunther Teubner
Professor em. für Privatrecht und Rechtssoziologie, Goethe-Universität Frankfurt, vorher Otto-Kahn-Freund-Professor, London School of Economics. Einschlägige Veröffentlichungen: Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds. Oxford: Hart.; ; Netzwerk als Vertragsverbund: Virtuelle Unternehmen, Franchising, Just-in-time in sozialwissenschaftlicher und juristischer Sicht. Nomos, Baden-Baden 2004. https://www.jura.uni-frankfurt.de/42849036/netzwerk_buch.pdf.
Webseite: https://www.jura.uni-frankfurt.de/42825790/Publikationsliste.