

Rechtsfähige Softwareagenten: Ein erfrischender Anachronismus

Marietta Auer

2019-09-30T09:05:22

Gunther Teubner hat jüngst wiederholt mit der These auf sich aufmerksam gemacht, dass gesteigerten zivilrechtlichen Risiken durch das Handeln autonomer Softwareagenten mittels Schaffung eines teilrechtsfähigen Rechtsstatus zu begegnen sei, dessen Zweck darin bestehe, „die Gefahr einer ständig wachsenden Verantwortungslücke abzuwenden“, die im Zivilrecht durch das zunehmend autonome Agieren digitaler Intelligenzen entstehe. Nach hier vertretener Auffassung ist die Anerkennung von Teilrechtsfähigkeit autonomer Softwareagenten indessen weder erforderlich noch auch nur geeignet, um den von Teubner identifizierten Risiken dogmatisch und theoretisch gerecht zu werden.

Falsche Anthropomorphisierungen

Als zentrale neue Gefahrenquelle identifiziert Teubner das „Autonomierisiko, das vom prinzipiell unvorhersehbaren Verhalten selbstlernender Algorithmen erzeugt wird“ und sich dann verwirkliche, wenn Softwareagenten „tatsächlich, wie es in den beteiligten Wissenschaften, besonders der Informationsphilosophie, Akteur-Netzwerktheorie und der Systemtheorie, vielfach vertreten wird, als handlungsfähige Akteure auftreten.“ Fragen wirft hier bereits die Verwendung von „tatsächlich“ im Zusammenhang mit den genannten Theorien auf. Dass bestimmte Theorien Modellvorstellungen von artifiziellen Umweltbestandteilen generieren, die darauf abzielen, hergebrachte konzeptionelle Schranken wie die zwischen Natur und Kultur oder Person und Sache als bloße Konstruktionen zu entlarven, zwingt in keiner Weise zu dem Schluss, dass die als selbständige Netzwerkakteure oder kommunizierende Systeme konstruierten digitalen Entitäten „tatsächlich“ handlungsfähige Akteure sind.

Nun sieht das auch Teubner, indem er vor „falschen Anthropomorphisierungen“ warnt und stattdessen treffendere Parallelen zu anderen fiktiven, nämlich juristischen Personen zieht. So lasse sich das soziale Substrat juristischer Personen ebenfalls nicht einfach auf eine „Vielheit konkreter Menschen“ zurückführen. Kern des Kollektivakteurs juristische Person sei vielmehr eine „Kette von Mitteilungen“, die über sich selbst kommuniziert und dadurch die soziale Realität der juristischen Person als Kollektivakteur begründet. Indessen steckt auch darin wieder ein *non sequitur*. Denn aus der bloßen Eigenschaft, eine kommunizierende Kollektividentität zu sein, folgt noch keine Notwendigkeit, als juristische Person oder sonstige rechtsfähige Entität anerkannt zu werden.

Interessanter wäre vielmehr zu wissen, was eine avancierte Theorie der juristischen Kollektivperson ausmacht und wo genau diese das spezifisch „Kollektive“ des kollektiven Handelns jenseits der Auflösung von Handlungs-

und Verantwortungszusammenhängen in individuelle Zweipersonenverhältnisse verortet. Dabei sind ganz unterschiedliche Lösungen denkbar, die etwa an der Konstruktion haftungsbegründender Kollektivinteressen, der Zusammenrechnung oder Zurechnung individueller Verantwortlichkeiten nach dem Prinzip *respondeat superior* oder an Darlegungs- und Beweislastregeln nach Sphären Gesichtspunkten ansetzen. All dies bedürfte nicht nur für digitale Kollektivakteure, sondern durchaus auch für analoge juristische Zurechnungskonstruktionen in Rechtsbereichen wie Produkthaftung (kollektive Güterproduktion), Kapitalmarktrecht (kollektive Marktsteuerung) oder Arbeitsrecht (kollektive soziale Organisation) näherer Erforschung.

Dabei dürften all diese Fragen jedenfalls eine Gemeinsamkeit aufweisen: Dass ihre sachgerechte dogmatische und theoretische Beantwortung nicht davon abhängt oder abhängen sollte, ob dem kollektiven Substrat im konkreten Fall ein teilrechtsfähiger Status verliehen wird oder nicht. So spielt es für das Eingreifen der Produkthaftung keine Rolle, in welcher Rechtsform der Produkthersteller verfasst ist; er wird auch nicht etwa dadurch rechtsfähig oder deliktfähig, dass er Adressat der Produkthaftung ist. Die Notwendigkeit kollektiver Verantwortlichkeit begründet für Entitäten kollektiver Güterproduktion also keinen neuen Rechtsstatus. Nichts anderes gilt aber auch für die rechtliche Verfasstheit digitaler Akteure, und zwar durchaus unabhängig davon, ob man – mit Teubner – „wirklich davon ausgehen [kann], dass Algorithmen kommunizieren“. Wiederum stört hier nicht nur das „wirklich“, sondern vor allem die Annahme, dass aus dem Vorliegen eines „genuinen Sozialsystems“ immer auch ein Rechtsträger folgen müsse – selbiges gälte dann auch für symbiotische Ökosysteme, Pflanzen- und Tiergemeinschaften. Persönlich hegt die Autorin dieser Zeilen, wenn die Rechtsfähigkeit artifizierlicher Kommunikationssysteme in Frage steht, jedenfalls eine Präferenz für rechtsfähige Wälder, Weiher und Wattlandschaften.

Menschenwürde kommt nur Menschen zu

Warum das Recht vor diesem Hintergrund nur bestimmte Entitäten, nämlich in erster Linie natürliche Personen, mit Rechtsträgereigenschaft ausstattet und sie damit zum Subjekt von Haftungs-, Zurechnungs- und Verantwortungsregeln macht, ist mit dem bloßen Verweis auf deren Kommunikationsfähigkeit also nicht zu beantworten. An dieser Stelle knüpft nun Teubner wieder an, indem er die Frage stellt, „ab welchem Schwellenwert“ Softwareagenten für das Recht als partiell autonom und damit rechtsfähig anzusehen seien. Teubner zufolge soll es dabei nicht auf „ontologische Eigenschaften“ wie „Intelligenz, Geist, Seele“ und erst recht nicht auf starke Anforderungen künstlicher Intelligenz wie „rationales Handeln, Selbsterfüllung oder selbsttätiges Lernen“ bzw. „Empathie, Gefühle, Leidensfähigkeit, Selbstbewusstsein oder gar ein digitales Gewissen“ ankommen. Stattdessen gehe es um die Rolle handlungsfähiger Akteure in „gesellschaftlichen Interaktionen, insbesondere wirtschaftlichen Transaktionen“ und dort um die Programmierung zur „Entscheidung unter Ungewissheit“, also um *ex ante* nicht voraussagbare und *ex post* nicht korrigierbare Entscheidungen zwischen Alternativen anhand der „Optimierung verschiedener Kriterien“.

Wieder bleibt aber ungeklärt, warum gerade dieses Kriterium, das zunächst einmal nur ein technisches Risiko beschreibt, wie es etwa auch in der Bauplanung oder im Straßenverkehr vorkommt, die entscheidende „Schwelle“ markieren soll, ab der von im Rechtssinne autonomem Handeln digitaler Entitäten die Rede sein soll. Dabei führt nach hier vertretener Ansicht bereits das Bild eines „Schwellenwerts“, also die Vorstellung eines kontinuierlichen Übergangs von niederen zu höheren kognitiven und emotionalen Fähigkeiten digitaler Entitäten, die irgendwann zwangsläufig in die notwendigen Voraussetzungen der Anerkennung rechtlicher Autonomie und damit in Rechtsfähigkeit „umkippen“ müssten, in die Irre. Autonomiefähigkeit ist kein archimedischer Punkt, an dem die verschiedenen Spektren kognitiver und emotionaler Intelligenz konvergieren und stetig in rechtliche Verantwortlichkeit münden.

Autonomie, im Privatrecht namentlich Privatautonomie, bezeichnet schlicht die Fähigkeit zur eigenverantwortlichen Gestaltung der eigenen Lebensverhältnisse durch Recht, die im Vertragsrecht etwa in der Fähigkeit zur Abgabe bindender Willenserklärungen und zum Abschluss gültiger Verträge zum Ausdruck kommt. Autonomie ist kein Risiko, sondern ein limitiertes rechtliches Gestaltungsprinzip. Von diesem Begriff bleibt jedoch nichts Greifbares übrig, wenn er einerseits mit der bloßen Fähigkeit zu unvorhersehbaren Entscheidungen auf einen kleinsten gemeinsamen kognitiven Nenner gebracht, andererseits jedoch mit Obertönen moralischer Autonomie – „digitales Gewissen“ – metaphysisch überhöht wird. Aus der resultierenden kognitiv-emotionalen Gemengelage kann nur Widersprüchliches resultieren. So Teubner: Wenn man es aus Gründen der Menschenwürde für „verfassungswidrig“ halte, „an Algorithmen Entscheidungen zu delegieren,“ müsse man „diese Praxis verbieten“ statt sie „klammheimlich“ zuzulassen und so zu behandeln, „als hätte keine Delegation stattgefunden.“ Umgekehrt wird ein Schuh daraus: Menschenwürde kommt nur Menschen zu. Und zu verbieten gibt es jedenfalls unter Menschenwürdegesichtspunkten ebenfalls nichts, weil autonome Entscheidungen im Rechtssinne weiterhin nur von Menschen getroffen werden.

Rechtliche Anreizwirkungen für digitale Akteure?

Dass der Schluss von unvorhersehbarer Entscheidung auf Autonomie und weiter auf Rechtsträgereigenschaft digitaler Akteure nicht aufgeht, wird deutlich greifbar, wenn man die von Teubner aufgeworfenen konkreten zivilrechtsdogmatischen Probleme betrachtet. Dies betrifft zunächst die Rolle autonomer Softwareagenten bei der Abgabe vertraglicher Willenserklärungen. Teubner ist der Auffassung, dass Softwareagenten wie Stellvertreter eigenverantwortliche Willenserklärungen „im Namen des Prinzipals“ abgeben, so dass man ihrer Funktion am ehesten durch die Anerkennung von Teilrechtsfähigkeit im Sinne von Stellvertretungsfähigkeit gerecht werde. Durch „Analogie“ zum Stellvertretungsrecht lasse sich, so Teubner, auch der „wohl wichtigste Einwand ausräumen, dass den Softwareagenten die für den Vertragsschluss notwendigen Willenselemente, insbesondere das Erklärungsbewusstsein“ fehlten.

Wiederum wird das Argument indessen gerade umgekehrt schlüssig. Eine Analogie zum Stellvertretungsrecht – wie auch immer man diese methodisch begründet –

kommt nur dann in Betracht, wenn der gleichzustellende Sachverhalt mit der für das Stellvertretungsrecht konstitutiven Abgabe einer eigenen Willenserklärung des Stellvertreters einschließlich Erklärungsbewusstsein vergleichbar ist, nicht hingegen, wenn diese Voraussetzung gerade fehlt und das fehlende Erklärungsbewusstsein stattdessen in einer Art Münchhausen-Logik im Rückschluss aus der vorgängigen Behauptung des Analogiebedürfnisses fiktiv konstruiert wird. Dabei hilft es auch nicht, das Erklärungsbewusstsein „entpsychologisiert“ zu deuten, denn auch daraus folgt noch keine auch von Teubner vorausgesetzte „soziale Norm“, wonach das Agieren eines digitalen Entscheidungsprozesses als eigenständige Willenserklärung dieser Software zu verstehen ist.

Letztlich kommt es für die Anwendung der Stellvertretungsregeln entscheidend darauf an, ob deren Zweck, nämlich den Prinzipal-Agent-Konflikt zwischen Vertretenem und Vertreter durch geeignete Anreize zu entschärfen, auch bei der Einschaltung digitaler Entscheidungsmechanismen einschlägig ist. Das wird man verneinen müssen: Das Risiko, dem der Vertretene durch das Handeln eines Stellvertreters ausgesetzt ist, liegt nicht in der Abgabe einer eigenen Willenserklärung durch den Vertreter als solcher, sondern im spezifischen Risiko einer Vertretung ohne Vertretungsmacht, die nur aufgrund der aus Verkehrsschutzgründen getroffenen Unterscheidung zwischen der Wirksamkeit der Willenserklärung im Außenverhältnis und der Überschreitung der Vertretungsmacht im Innenverhältnis konstruktiv möglich ist und durch die fein ausdifferenzierte Haftung des *falsus procurator* mit zielgenau gesetzten Anreizen zur Unterlassung ineffizienter *ultra-vires*-Vertretergeschäfte kompensiert wird.

All diese Überlegungen sind jedoch bei Beteiligung digitaler Entscheidungsmechanismen von vornherein nicht tragfähig, da nicht ersichtlich ist, wie digitale Akteure – wie unvorhersehbar auch immer ihre „Entscheidungen“ sein mögen – durch die Anreizwirkung der Haftung des *falsus procurator* erreicht werden können. Warum sollten digitale Entscheider so programmiert werden, dass sie ihr künftiges „Verhalten“ daran ausrichten, ob es mit (fiktiver?) Haftung auf ihr (fiktives?) Eigenkapital sanktioniert wird? Wäre es nicht effizienter und zudem allein pflichtgemäß, sie von vornherein so zu programmieren, dass sie haftungspflichtige *ultra-vires*-Geschäfte unterlassen, was nichts anderes bedeutet, als den *ex post* haftungsbegründenden Mechanismus *ex ante* in der algorithmischen Konstruktion zu antizipieren – was bei dem angenommenen digitalen Intelligenzniveau nun wirklich keine Schwierigkeit darstellen sollte?

Anders gewendet scheinen Teubners Überlegungen zur eigenständigen rechtlichen Berücksichtigung digitaler Akteure schlicht die Dignität des Rechts zu überschätzen, handelt es sich dabei doch letztlich nur um einen höchst unvollkommenen Mechanismus, der die Sanktionierung sozial unerwünschten Verhaltens etwa eines *falsus procurator* in utilitaristische Anreizkategorien übersetzt, die unvollkommenen, durch Nutzen und Schaden ansprechbaren Wesen wie Menschen – gleichviel ob diese Ansprechbarkeit nun essentiell rational oder emotional begründet ist – zugänglich sind. Glücklicherweise sind die digitalen Intelligenzen, die derartiger Hilfskonstruktionen gar nicht erst bedürfen.

Technologische Zufallsschäden

Ähnlich gelagert sind auch die Fälle vertraglicher und außervertraglicher Haftung. Im Rahmen der Vertragshaftung plädiert Teubner für die Anerkennung digitaler Softwareagenten als Erfüllungsgehilfen, um eine „schwer erträgliche Haftungslücke“ zu vermeiden, die daraus entstehen soll, dass der Betreiber des digitalen Agenten von der Vertragshaftung frei werde, wenn er nachweisen könne, dass er „selbst keine Vertragspflicht verletzt hat“, dass also der schadensverursachende Fehler des Agenten allein in der Sphäre des letzteren entstanden ist. In einem solchen Fall müsse der Kunde den vom Computer verursachten Schaden tragen.

Dazu ist zunächst festzuhalten: Schadensersatz ist die Ausnahme, nicht die Regel. Gehaftet wird im Vertragsrecht grundsätzlich nur für vom Schuldner zu vertretende Pflichtverletzungen. Für nicht zu vertretende Zufallsschäden wird dagegen nicht gehaftet, sondern der Schaden bleibt wiederum aus guten Gründen der Anreizung zu pflichtgemäßem, aber nicht übermäßig strangulierendem Vertragshandeln bei demjenigen liegen, der das Sachrisiko trägt: *casum sentit dominus*. Noch deutlicher wird das im Recht der außervertraglichen Haftung, für das systematische Haftungslücken in Bereichen bloßer Umverteilungsschäden geradezu typisch und rechtspolitisch gewollt sind. Tatsächlich dürfte die große Mehrzahl aller Schadensfälle nicht auf zurechenbar pflichtwidrigem Handeln, sondern auf Zufall im Rechtssinne beruhen und als Schaden ohne Unrecht – *damnum absque iniuria* – schlicht entschädigungslos bleiben, ohne dass darin ein rechtspolitisches Skandalon läge.

Aus der bloßen Tatsache der Entschädigungslosigkeit von technologischen Zufallsschäden, die unter Beteiligung von Softwareagenten entstanden sind, folgt also schlicht nichts im Hinblick auf die Frage, ob und wann in solchen Fällen gehaftet werden sollte. Der dafür ausschlaggebende Anknüpfungspunkt ist wiederum zunächst die vom Vertragspartner oder Deliktsschuldner geschuldete Sorgfalt in der Programmierung und im Einsatz autonomer Softwareagenten, die mit steigender Komplexität sicherlich zunehmend in objektivierte Zurechnungs-, Verantwortungs- und Beweislastregeln nach Sphären- und Risikogesichtspunkten münden wird, deren dogmatische Ausformulierung noch bevorsteht. Auch dafür bedarf es indessen keiner Hilfskonstruktion wie der Zurechnung von Dritthandeln eines Erfüllungsgehilfen oder Verrichtungsgehilfen, zumal für diese Regeln ebenso wie im Stellvertretungsrecht gilt, dass nicht ersichtlich ist, wie ein digitaler Gehilfe durch die Androhung etwaiger Regressforderungen des haftungspflichtigen Prinzipals zu pflichtgemäßem Verhalten angereizt werden könnte. Der Rechtssinn des Einstehens für fremde Pflichtverletzung ist damit abermals schon im Ansatz nicht gegeben und mithin entgegen Teubner auch nicht sinnvoll als Frage der Gleichbehandlung menschlicher und nichtmenschlicher Akteure diskutierbar.

Entschärfung durch Haftungsobjektivierung

Letztlich wird das Ganze wohl nicht so heiß gegessen wie gekocht. So bringt Teubner zu Recht „Objektivierungstendenzen im heutigen Privatrecht“ ins Spiel,

die das Problem unvorhersehbar agierender Softwareagenten zumindest partiell entschärfen. Gemeint sind die erwähnten Objektivierungsmöglichkeiten im Bereich vertraglicher und außervertraglicher Sorgfaltspflichten, die die dort geltenden Verschuldenshaftungen jeweils an die objektive Pflichtwidrigkeitshaftung annähern, die den Haftungsmaßstab der Gefährdungshaftung bildet.

Angesichts der rasanten technischen Entwicklung und der Unklarheit, wie und unter welchen Voraussetzungen digitale Entitäten künftig in vertraglichen und außervertraglichen Zusammenhängen Schadensrisiken verursachen, ist Teubner weiter darin zuzustimmen, dass derzeit kein Bedarf für eine besondere Gefährdungshaftung für digitale Entitäten besteht. Der Grund dafür liegt allerdings wiederum nicht darin, dass das „Risiko digitaler Entscheidungsautonomie prinzipiell andersgeartet“ sei als die Fälle erlaubter Risiken, die die Grundlage für die Anerkennung der Gefährdungshaftung in den gesetzlich normierten Fällen bilden. Der Grund liegt vielmehr in einer Einsicht, die Teubner abschließend im Rahmen der von ihm identifizierten weiteren „Verbundrisiken“ und „Vernetzungsrisiken“ formuliert: Man betritt „sehr viel unsicheres Gelände“.

In der Tat: Das aktuelle, abschüssige digitale Terrain, das klassische Personalität und Individualität mittels Big-Data-Lesbarkeit eher auflöst als akzentuiert, ist kein kongeniales Umfeld für die Einführung neuer Personenstatus und gesetzlicher Haftungsregelungen für digitale Akteure. So gesehen erweist sich Teubners Insistenz auf der Personifizierung ebendieser Akteure in einem Regulierungsumfeld, in dem die klassisch-moderne „Rechtstechnik der Personifizierung [...] an ihre Grenzen gerät“, als geradezu erfrischend anachronistisch.

