

Einzelbesprechung

Künstliche Intelligenz

Anna Beckers / Gunther Teubner, Digitale Aktanten, Hybride, Schwärme: Drei Haftungsregime für künstliche Intelligenz. Berlin: Suhrkamp 2024, 345 S., br., 26,00€

Besprochen von **PD Dr. Carsten Ochs**: Universität Kassel, Fachgebiet Soziologische Theorie/Wissenschaftliches Zentrum für Informationstechnik-Gestaltung, E-Mail: carsten.ochs@uni-kassel.de

<https://doi.org/10.1515/srsr-2025-2047>

Schlüsselwörter: Künstliche Intelligenz, Rechtswissenschaft und -soziologie, Systemtheorie, Science & Technology Studies, Regulierung

Von nichtmenschlichen Entitäten und ihrer Rolle für die Konstitution von Gesellschaft hat man in den letzten Jahrzehnten in der Soziologie so einiges gehört und die empirische Wirklichkeit hat die Relevanz ihres analytischen Einbezugs zuletzt durch die Verbreitung sprachfähiger Systeme der Künstlichen Intelligenz (KI) wie ChatGPT noch einmal zusätzlich erhöht. Wenn KI nunmehr bei der Konstitution von Gesellschaft ein ganz alltagspraktisches Wörtchen mitredet, sieht sich dadurch jedoch nicht zuletzt auch das Recht herausgefordert, das *Anna Beckers* und *Gunther Teubner* zufolge auf „autonome Quasi-Subjekte – generative KI (etwa ChatGPT), Hochgeschwindigkeits-Handelsalgorithmen, Roboter, Softwareagenten, Cyborgs, Hybride, Computernetzwerke (...) nicht vorbereitet“ ist: „So absurd es klingt: Wenn autonom agierende Algorithmen Fehlentscheidungen treffen und Schäden verursachen, kann, sofern den beteiligten Menschen kein Fehlverhalten vorzuwerfen ist, nach dem zurzeit geltenden Recht niemand verantwortlich und haftbar gemacht werden“ (S. 18). Im normativen Gewebe der Digitalen Gesellschaft klafften eklatante „Haftungslücken“ (S. 21), die *Beckers/Teubner* mit ihrem dezidiert soziologisch orientierten Buch zu schließen gedenken.

Um das Vorhaben umzusetzen, arbeiten die Autor:innen mit der Methode der „vergleichende[n] soziologische[n] Jurisprudenz“, d.h. sie analysieren „sozio-digitale Institutionen und ihre Risiken, um die relevanten Rechtsprobleme identifizieren zu können“, und nutzen die gewonnenen Erkenntnisse dann zur Konzeption von Haftungsregimen (S. 12). So ausgerüstet und immer wieder auf Versatzstücke der Systemtheorie, *science & technology studies* und Technikphilosophie zurückgreifend, bestimmen *Beckers/Teubner* drei versch. sozio-digitale Institutionen, in denen zunehmend autonome (d.h. unvorherseh- und unerklärbar agierende) Algorithmen operieren, die jeweils spezifische Risiken erzeugen, denen mit ebenso spezi-

fischen Haftungsregimen begegnet werden soll: digitale Aktanten, Hybride und Schwärme.

Der Behandlung der Institutionen/Risiken/Regime vorgeschaltet ist ein einführendes Kapitel, in dem die juristisch-soziologische Problemstellung und Herangehensweise des Buches erläutert wird. Bestehende Lücken und die Unvorbereitetheit und Abwehrhaltung aktuellen Rechts bzgl. neuartiger KI-Konstellationen werden anhand von Beispielen diskutiert: Aktuell werde mit „humanistischem Pathos“ (S. 25) auf die exklusive Handlungsfähigkeit des Menschen insistiert, paradoxerweise führe genau dies zur „Haftungsimmunität der Schädiger“ und resultiere in „einem Überangebot der problematischen Aktivitäten“ (S. 26). Bislang vorgebrachte Reformvorschläge würden fehlgehen, zu kurz oder zu weit greifen (S. 34). Grundsätzlich müsse von der „Verflechtung digitaler und menschlicher Handlungen“ (S. 40) ausgegangen werden, was es wiederum nahelege, Algorithmen „eine begrenzte Rechtssubjektivität zuzuschreiben“ (S. 45).

Um diese nicht technikdeterministisch nur aus der algorithmischen Operationsweise als solcher abzuleiten, schlagen die Autor:innen das Konzept der „sozio-digitalen Institution“ vor: „(temporär) stabilisierte Komplexe sozialer Erwartungen, (...) besonders Risikoerwartungen, die entstehen, wenn digitale Technologien in sozialen Situationen genutzt werden“ (S. 47). In sozio-digitalen Institutionen erzeugten Nutzungs- und Wirkweisen der Technik „techno-digitale Normativität“ (S. 49). *Beckers/Teubner* betonen die zentrale Rolle der Sozialwissenschaften bei der Bestimmung dieser Normativität (S. 51) und unterscheiden heuristisch drei Typen von Maschinenverhalten, aus denen sie dann drei versch. sozio-digitale Institutionen ableiten:

- individuelles Maschinenverhalten: eigenständiges Operieren identifizierbarer Einzel-Systeme;
- hybrides Maschinenverhalten: enge Kopplung menschlicher und algorithmischer Handlungen; und
- kollektives Maschinenverhalten: gemeinsam-wechselwirkendes Operieren einer Vielzahl algorithmischer Systeme jenseits menschlicher Intervention (S. 54–55).

Beckers/Teubner zufolge erzeugen die drei sozio-digitalen Institutionen jeweils spezifische Risiken: Individuell agierende algorithmische *Aktanten* entsprächen der Prinzipal-Agenten-Beziehung *digitaler Assistenz*, die ein „Autonomierisiko“ heraufbeschwöre. *Hybride Mensch-Algorithmus-Assoziationen* würden ob ihres Verflechtungsgrads gar keine individuellen Operationsbeiträge mehr zu identifizieren erlauben und erzeugten ein „Verbundrisiko“. Schließlich könnten Algorithmen (z.B. im Hochfrequenzhandel) wie *Schwärme* „in enger Vernetzung mit anderen Algorithmen agieren“ (S. 61); sie täten dies unterhalb menschlicher Wahrnehmungs-

schwellen und beschwüren so ein „Vernetzungsrisiko“ herauf (vgl. die „Flash Crashes“ an US-amerikanischen Börsen).

Nachdem das konzeptionelle Fundament gelegt ist, widmen sich die Autor:innen im Weiteren der Ausarbeitung jedes der drei angeführten Institutionen-Risiken-Haftungsregime-Komplexe. Die umfassendste Beschäftigung erfolgt mit dem „Autonomierisiko“ durch „digitale Assistenz“ (S. 64–169). Dabei werden algorithmische Aktanten als Stellvertreter behandelt und ähnlich wie Organisationen konzipiert, um zu einer (rechtlichen) „Personifizierung von Algorithmen“ zu gelangen (S. 67). Digitale Assistenten würden an Kommunikation teilnehmen, autonom Entscheidungen in Situationen der Ungewissheit treffen (S. 88) und so versch. Risiken heraufbeschwören (Einzelentscheidungen schwer zu identifizieren, Konsens Nutzer:in-Assistent kaum explizierbar, große Handlungsspielräume, „Abweichungsrisiko“ des Assistenten, kontextunabhängige Sturheit algorithmischer Prozessierung, S. 99–103). Daraus ergäben sich juristische Konsequenzen für das Recht algorithmischer Verträge (Kap. 2 III.), vertraglicher Haftung (2. IV.) und außervertraglicher Haftung (2. V.). Die Autor:innen diskutieren in diesem Zusammenhang im Detail bestehende Haftungsregime und -vorschläge, und bringen dagegen jeweils einen eigenen Lösungsvorschlag in Stellung, der sich am Prinzip der Prinzipal-Agenten-Haftung (PAH) orientiert. Digitale Assistenten gelten somit als Agenten eines menschlichen Prinzipals (einer Nutzer:in), d.h. als Stellvertreter (algorithmische Verträge), Erfüllungsgehilfen (vertragliche Haftung) oder Verrichtungsgehilfen (außervertragliche Haftung). Die PAH der Nutzer:innen digitaler Assistenten werde immer dann wirksam, wenn letztere „deliktisches Verhalten“ an den Tag legen, das auf die Autonomie der Algorithmen, nicht auf Fehlverhalten von Programmierer:innen, Hersteller:innen, Vertrieben oder Betreibern zurückzuführen sei: „Für das Versagen eines algorithmischen Assistenten (...) haftet allein der Nutzer, und zwar in gleichem Maße wie für das Versagen eines menschlichen Assistenten“ (S. 169).

Vergleichsweise kurz (jeweils um die 40 Seiten) nehmen sich die beiden Folgekapitel zum „Verbundrisiko“ (Kap. 3) und „Vernetzungsrisiko“ (Kap. 4) aus. Ersteres gehe aus der sozio-digitalen Institution der Mensch-Algorithmus-Assoziation hervor, wie man sie etwa im „Computerjournalismus“ antreffe (S. 172). *Beckers/Teubner* bestimmen Assoziationen unter Rückgriff auf die „Organisationsanalogie“ (S. 180–182). Demnach könne, ähnlich wie bei Organisationen, ein Mensch-Algorithmus-Kollektiv „verantwortlich für eine Handlung sein, auch wenn seine Mitglieder es nicht sind“ (S. 182). Denn, wenn eine Rückführung von Handlungsfolgen auf Einzelhandlungen des Verbundes nicht mehr möglich ist, bestehe die einzige Möglichkeit, Haftungslücken zu vermeiden „darin, den Hybriden selbst als verantwortlichen kollektiven Akteur zu behandeln“ (S. 184). Diesbezüglich könne die Rechtsprechung dahingehend erneuert werden, Assoziationen volle Rechtspersönlichkeit zu gewäh-

ren. Bis das Recht so weit sei, solle zunächst die Netzwerk-Haftung in Anschlag gebracht werden: Assoziationen sollten zunächst als Ganzes in Haftung genommen werden, indem sich Geschädigte an irgendeine Instanz der Assoziation als „one stop shop“ (S. 204–205) wendeten, welche dann erst einmal für den Schaden eintreten müsse. Im Anschluss könne die Instanz dann die Haftung je nach Netzwerkanteil kanalisieren, d. h. die beteiligten „Betreiber, Eigentümer, Hersteller und Lieferanten der digitalen Technologie“ (S. 199) anteilig in Regress nehmen (S. 207).

Damit verbleibt als dritte behandelte Institution-Risiko-Haftungs-Konstellation die der algorithmischen „Schwärme“ (Kap. 4), die das „Vernetzungsrisiko“ der Entfaltung von kausal nicht-zurechen-, unvorherseh- und in puncto Verantwortung nicht attribulierbarer Eigendynamik heraufbeschwören. Die Autor:innen argumentieren systemtheoretisch, dass nicht-kommunikative, aber Informations-prozessierende algorithmische Black Boxes „durch die Vernetzung als technische Konfiguration (...) ‚außerhalb‘ der Gesellschaft“ lägen (S. 210). Schwärme seien eigenmächtige Infrastrukturen, die algorithmische Operationen jenseits menschlichen Verstehens/Eingreifens kausal verketteten. Um Fehlleistungen trotz unmöglicher Handlungsattribution beizukommen, schlagen sie eine „Vergesellschaftung des Vernetzungsrisikos“ vor (S. 232): an der Erzeugung des Vernetzungsrisikos beteiligte Instanzen sollten zu „Zwangsmitgliedern“ (S. 235) von Risikopools bestimmt werden, die dann *ex ante* Pflichtbeiträge in einen Fonds einzahlen und beim Katastrophen-Eintritt *ex post* Haftungsbeiträge entrichten müssten. Verwalten solle die Fonds eine aufgrund der Globalität des Problems auf EU-Ebene anzusiedelnde Behörde: „So könnte sich eine europäische Fondslösung für das Vernetzungsrisiko auch global durchsetzen“ (S. 250).

Im letzten Teil des Buches (Kap. 5) runden die Autor:innen ihren Argumentationsgang dann ab, indem sie eine Synopse, Unterschiede zwischen den Haftungsregimen, Haftungsträger und Verfahrensregeln präsentieren, bevor eine Anwendung der Regime auf fünf empirische Fälle erfolgt. Dabei wird die Nützlichkeit der vorgelegten Konzeption und der mögliche Beitrag herausgekehrt, den „das Haftungsrecht (...) zu einer künftigen Digitalverfassung“ leisten könne (S. 294).

Insgesamt legen *Beckers/Teubner* ein spannendes und überzeugendes Buch vor, das den großen Nutzen soziologischer Analysen für die normative Gestaltung digitaler Gesellschaften aufzeigt. Ein wenig juristischen Appetit sollte man bei der Lektüre ob der detaillierten rechtswissenschaftlichen Diskussionen allerdings schon mitbringen, denn die Autor:innen betreiben letztlich keine Sozialtheorie, sondern wenden diese an. Deshalb fällt es auch nur bedingt ins Gewicht, dass ihre Nutzung sozialwissenschaftlichen Wissens mitunter ein wenig selektiv oder gar inkonsistent zu werden droht. Wenn etwa Technik und Gesellschaft mit der Systemtheorie als kategorisch zu unterscheidende Welten bezeichnet werden (S. 210), dann wird damit einer zentralen Prämisse der Akteur-Netzwerk-Theorie widersprochen – wäh-

rend das umgekehrte Abstellen letzterer auf Hybridität für die Konzeption der sozio-digitalen Institution der Assoziation von den Autor:innen gerade stark gemacht wird. Aus dem selektiven Wissensbezug scheint dennoch ein stimmiger Haftungsrahmen hervorzugehen, juristische Leser:innen müssen sich daran also nicht stören. Soziologische Rezipient:innen mögen die sozialtheoretische Selektivität indessen als Aufruf interpretieren, ebenso großen Mut bei der Reformulierung hergebrachter Theorien mitzubringen, wie sie *Beckers/Teubner* mit ihrer Behandlung von KI-Haftungsfragen demonstrieren.