

Verantwortungslose Maschinen?

Digitaletische Herausforderung autonomer Systeme in der Medizin

Von *Robert Ranisch* und *Joschka Haltaufderheide*

I. Einführung

Verantwortungslücken entstehen, wenn positive oder negative Wirkungen von Technologien keinem Akteur zugeschrieben werden können, da eine Handlung oder Entscheidung beispielsweise durch eine Maschine bedingt war. In solchen Fällen drohen moralische Forderungen, Schuld oder auch Lob ins Leere zu laufen. Verantwortungszuschreibungen gehören zu den grundlegenden Phänomenen der moralischen und rechtlichen Praxis. Sie betreffen nicht nur das Verhältnis zwischen Menschen, sondern auch zwischen Menschen und den von ihnen geschaffenen Entitäten. Die Frage der Verantwortlichkeit ist in diesen Konstellationen besonders brisant, wenn Technologien oder Artefakte einen Entwicklungsgrad erreichen, der ihnen eine gewisse funktionale Autonomie ermöglicht, sie aber zugleich nicht über die Eigenschaften verfügen, als moralische Verantwortungssubjekte zu gelten. Mit funktionaler Autonomie ist gemeint, dass Maschinen oder Artefakte innerhalb einer bestimmten Handlungsdomäne eine Repräsentation ihrer Welt erzeugen können und auf diese im Rahmen ihrer Zielerreichung durch eigenständige Anpassung reagieren.¹ Bekannte Figuren der europäischen und US-amerikanischen Literatur wie Frankenstein's Monster, der Golem von Prag oder John Sladeks Roderick stehen beispielhaft für die literarische Reflexion dieser Fragen.² Beim Begriff Roboter handelt es sich nicht zuletzt um eine Übertragung des slawischen Wortstammes für Frohnarbeit, der sowohl auf eine gewisse Eingeschränktheit und Unterordnung des entsprechenden Subjektes als auch seine Fähigkeit zu selbstständigen Verrichtung verweist. Der Verdacht ist jedoch nicht unbegründet, dass insbesondere Entwicklungen der Künstlichen Intelligenz (KI) und hier besonders sogenannte selbstlernende Systeme aufgrund ihrer weitreichenden Möglichkeiten, ihrer Komplexität und ihrer Entwicklungspotenziale das Problem möglicher Verantwortungslücken erheblich verschärfen.

¹ S. Nyholm, *Humans and robots. Ethics, agency, and anthropomorphism*, 2020, S. 54.

² S. Cave/K. Dihal/S. Dillon, Introduction, in: *AI narratives. A history of imaginative thinking about intelligent machines*, Oxford 2020, S. 1 (4); P. March-Russell, *Machines Like Us? Modernism and the Question of the Robot*, in: *AI narratives. A history of imaginative thinking about intelligent machines*, 2020, S. 165 (177).

Als Anwendungsfall und Prüfstein dienen dieser Arbeit die Entwicklungen im Bereich der digitalen Medizin, im speziellen der Einsatz von robotischen Systemen in der Chirurgie. Diese wurden bereits in den 1980er Jahren erforscht und werden spätestens seit der Zulassung des bekannten Da-Vinci-Operationssystems zunehmend eingesetzt.³ Wenngleich roboter-assistierte Eingriffe bisher unter dem Vorbehalt einer vollständigen menschlichen Kontrolle stehen, so wird davon ausgegangen, dass es sich bei der derzeitigen Form der roboter-assistierten Chirurgie um eine Zwischenstufe zwischen manuellen chirurgischen Verfahren und weitaus autonomen Geräten handelt.⁴

Wir argumentieren, dass mit dem Einsatz von fortgeschrittenen Maschinen Konstellationen entstehen können, in denen weder dem Operateur noch dem Hersteller die moralische Verantwortung für Behandlungsfehler zugeschrieben werden können. Nach einer konzisen Einführung zu den Begriffen der Verantwortung sowie von Verantwortungslücken werden wir entsprechende Konstellationen am Beispiel von autonomen Systemen in der robotischen Chirurgie diskutieren. Hier zeigen sich Fragen von Verantwortungslücken zukünftig als unausweichlich. Zugleich argumentieren wir jedoch, dass der moderne Verantwortungsbegriff mehrdimensional ist und beispielsweise in individuelle Verantwortung sowie in Rollen- und Kollektivverantwortung unterschieden werden kann. Er changiert zudem zwischen einem Konzept von Verantwortung als Zuschreibung und Verantwortung als Übernahme. Insbesondere auf der Grundlage der wohlbekannten und ausgeprägten professionellen Rollenverantwortung von Ärztinnen und Ärzten gemeinsam mit Überlegungen zur Verantwortungsübernahme ergeben sich daher Möglichkeiten, zukünftige Verantwortungslücken zumindest zu reduzieren.

II. Was bedeutet „Verantwortung“?

In der Alltagspraxis der Moral ist „Verantwortung“ zwar eine allgegenwärtige Kategorie, gleichwohl handelt es sich weder um einen einheitlichen noch um einen einfachen Begriff.⁵ Alltagssprachlich drückt Verantwortung die Notwendigkeit aus, in einer Sache Rede und Antwort stehen zu müssen, gegebenenfalls auch für die positiven oder negativen Folgen dieser Sache eintreten zu müssen oder Verantwortung im Sinne von Schuld zu tragen.⁶ In einem philosophischen Sinn lässt sich Verantwortung in gleich mehreren Hinsichten spezifizieren. Wir orientieren uns hierbei an der

³ S. Kalan/S. Chauhan/R. Coelho et al., History of robotic surgery, *Journal of Robotic Surgery*, 3 (2010), S. 141 (142).

⁴ P. Fiorini/K. Goldberg/Y. Liu/R. Taylor, Concepts and Trends in Autonomy for Robot-Assisted Surgery, *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, 7 (2022), S. 993 (994).

⁵ L. Heidbrink, Definitionen und Voraussetzungen der Verantwortung, in: *Handbuch Verantwortung*, 2017, S. 3 (4); J. Loh, Strukturen und Relata der Verantwortung, in: *Handbuch Verantwortung*, 2017, S. 35 (36).

⁶ Heidbrink (Fn. 5), S. 4.

gängigsten Verwendung des Begriffes, wie er auch im Diskurs um die Verantwortungslücken am häufigsten in Erscheinung tritt.⁷

Die erste wichtige Differenz betrifft deskriptive und präskriptive Begriffsbedeutungen. In einem deskriptiven Sinn lässt sich von kausaler Verantwortung sprechen. So lässt sich ausdrücken, dass ein Sturm für das Abbrechen eines Baumes verantwortlich ist oder eine Operation für den Tod eines Patienten. Hier muss allerdings noch keine normative Aussage über Verantwortung vorliegen. Ein Sturm stellt schlichtweg ein Wetterereignis dar, welches kein Akteur im präskriptiven Sinn verantworten kann. Im Fall des tödlich verlaufenden medizinischen Eingriffs ist dieser zwar auf einen Akteur zurückzuführen. Wurde die Operation allerdings *lege artis* durchgeführt, kann die Rolle des Arztes allenfalls als kausal, aber nicht als moralisch (oder auch rechtlich) verantwortlich beschrieben werden. Die normative Forderung des Entstehen-müssens lässt sich weiter in mindestens eine rechtliche und eine moralische Ebene unterscheiden. Rechtlich ergibt sich die besondere Stellung von Personen als Verantwortliche aus ihrer grundsätzlich freien und letztinstanzlichen Urheberschaft von Ereignissen in der Welt.⁸ Diese Freiheit ist verknüpft mit der normativen Erwartung, Rechtsnormen und Befehle auch gegen möglicherweise anders liegende Anreize zu befolgen oder für das Nicht-befolgen antworten zu müssen.⁹ Davon zu unterscheiden ist moralische Verantwortung, um die es im Folgenden gehen soll. Sie meint das Abgeben von Rechenschaft vor Anderen auf der Grundlage moralischer Prinzipien. Während sich beide Dimensionen inhaltlich durchaus weit überlappen können, unterliegt die rechtliche Verantwortung zudem einer Reihe von rechtspragmatischen Bedingungen. Ist hingegen die Rede von moralischer Verantwortung, so sind die Bedingungen eine Frage des Arguments, das entlang objektivierbarer Kriterien und entsprechender Begriffe das Vorliegen einer bestimmten Verantwortung schlüssig darlegen kann.¹⁰ Moralische Verantwortung wird in diesem Sinn häufig (aber nicht immer) als ein Zuschreibungsbegriff verwendet. Eine kausale Verantwortung ist in der Regel eine notwendige Bedingung für die Zuschreibung von moralischer Verantwortung, sie ist aber nicht hinreichend. Üblicherweise werden für die Zuschreibung moralischer Verantwortung mindestens zwei Voraussetzungen angenommen:

⁷ P. Königs, Artificial intelligence and responsibility gaps: what is the problem?, Ethics and Information Technology, 3 (2022), S. 1 (2); D. Gunkel, Mind the gap: responsible robotics and the problem of responsibility, Ethics and Information Technology, 4 (2020), S. 307 (307); A. Matthias, The responsibility gap: Ascribing responsibility for the actions of learning automata, Ethics and Information Technology, 3 (2004), S. 175 (175); J. Danaher, Tragic Choices and the Virtue of Techno-Responsibility Gaps, Philosophy & technology, 2 (2022), S. 1 (5). Hier wird lediglich die These vertreten, dass der Diskurs um Verantwortungslücken sein Problem in einer bestimmten Weise strukturiert, um es zu adressieren, ohne damit anzunehmen, dass dies auch für die Lösungsvorschläge gilt.

⁸ J. Klement, Rechtliche Verantwortung, in: Handbuch Verantwortung, 2017, S. 559 (561).

⁹ Klement (Fn. 8), S. 562.

¹⁰ Heidbrink (Fn. 5), S. 4.

a) eine *Form von Kontrolle* über eine Handlung, dies schließt eine kausale Verursachung ein, aber auch z. B. die Handlungsfreiheit des Akteurs;

b) eine *epistemische Bedingung*, etwa bestimmte Formen des Wissens oder der Intentionalität des Handelnden.¹¹

Neben deskriptiven und präskriptiven Bedeutungsdimensionen ist es wichtig zu sehen, dass Verantwortung nicht schlechthin besteht, sondern ein relationaler Begriff ist. Alfred Schütz hat als einer der ersten darauf hingewiesen, dass Verantwortung sich mindestens auf ein Objekt und eine Instanz bezieht.¹² Während seitdem weitestgehend Einigkeit herrscht, dass die Berücksichtigung der jeweiligen Relata für eine angemessene Rekonstruktion notwendig sind, herrscht weniger Einigkeit darüber wie viele es sind. Es zeigen sich jedoch im mindesten drei, die zu berücksichtigen sind: Ein Verantwortungssubjekt (z. B. ein Arzt bzw. eine Ärztin) ist für etwas – einen Verantwortungsgegenstand (z. B. die Operation) – gegenüber einem bestimmten Verantwortungsobjekt (z. B. dem Patienten bzw. der Patientin) verantwortlich.

Darüber hinaus sind handelnde Personen in der Regel nicht nur für jede Form von intendierter Handlung verantwortlich, sondern können auch für nicht beabsichtigte Handlungen oder deren Folgen verantwortlich sein. So ist beispielsweise Ödipus für den Tod seines Vaters verantwortlich, auch wenn es nicht seiner Absicht entsprach, seinen Vater zu töten. Verantwortet werden kann darüber hinaus sowohl ein Tun als auch ein entsprechendes Unterlassen. So kann ein Arzt für einen verfehlten Eingriff, ebenso wie für die Unterlassung eines Eingriffes moralisch verantwortlich sein, sofern dieser medizinisch indiziert ist. Schließlich sind zwei Blickrichtungen zu unterscheiden: Verantwortung kann in einem retrospektiven Sinne bestehen. Geht es beispielsweise um Fragen der moralischen Schuld oder Anklage, z. B. für den verfehlten Eingriff, liegt das Augenmerk auf der retrospektiven Verantwortung. Spricht man dagegen von der Verantwortung des Arztes gegenüber seinen Patienten, ist damit auch ein prospektiver Sinn von Verantwortung gemeint, der bestimmte Pflichten umfasst.

III. Verantwortungslücken

Damit von Verantwortungslücken im hier diskutierten Sinne gesprochen werden kann, müssen drei Bedingungen erfüllt sein: Zum Ersten muss ein hinreichend komplexes maschinelles System (eine KI oder ein Roboter) einen Entwicklungsgrad erreichen, der ihm funktionale Autonomie ermöglicht. Zum zweiten muss das System im Rahmen dieser funktionalen Autonomie kausal für ein (positives oder negatives) Ergebnis verantwortlich sein. Drittens muss das System in einem relevanten Sinne nicht (vollständig) der Kontrolle durch Menschen oder Handelnde unterliegen.¹³ Sind diese Bedingungen erfüllt, kann aufgrund der fehlenden Kontrollbedingung

¹¹ Heidbrink (Fn. 5), S. 6; Königs (Fn. 7), S. 2.

¹² Loh (Fn. 5), S. 37.

¹³ Danaher (Fn. 7), S. 2; Königs (Fn. 7), S. 1.

die Verantwortung für das Ereignis womöglich keinem Verantwortungssubjekt im üblichen Sinne zugeschrieben werden.

Im Rahmen der entsprechenden Diskussionen um die Ethik automatisierter Maschinen ist mit Verantwortung in den meisten Fällen vorrangig ein spezifisches Problem hinsichtlich der Kontrollbedingung adressiert. So heißt es etwa in der Erklärung von Montreal für eine verantwortungsvolle KI-Entwicklung: „Für Entscheidungen, die auf von KI-Systemen getroffenen Empfehlungen basieren, und für die daraus folgenden Handlungen können ausschließlich Menschen verantwortlich gemacht werden.“¹⁴ Dies folgt der Erkenntnis, dass einer Maschine kausale Verantwortung zugeschrieben werden kann, während es zugleich schwer scheint, ihr auch moralische Verantwortung zuzuschreiben. Demgemäß müssten stets bestimmte andere Akteure für die möglichen nachteiligen Folgen verantwortlich sein. Fraglich ist allerdings, ob diese normative Forderung auch deskriptiv eingeholt werden kann. Mit zunehmend komplexer werdenden Systemen, die darüber hinaus ihre Agentivität zunehmend ausweiten, geht das Risiko einher, dass sich Verantwortungslücken auftun, d.h., dass keine Akteure – nicht die Maschine als kausal verantwortlich aber eben auch nicht Nutzer oder Hersteller – für mögliche Wirkungen moralisch verantwortlich gemacht werden können.

Verantwortungslücken wie oben beschrieben, können in allen Anwendungsfeldern der KI oder von autonomen Entscheidungssystemen auftreten, angefangen von selbstfahrenden Autos¹⁵ bis hin zu autonomen Waffensystemen.¹⁶ Die jeweils gleiche Frage lautet: Wer ist dafür verantwortlich, wenn ein Auto ohne menschliches Zutun einen Unfall oder Personenschäden verursacht, oder eine autonome Drohne eine bestimmte Person tötet? Umgekehrt können Verantwortungslücken aber auch dann auftreten, wenn wir es mit einem wünschenswerten Ergebnis zu tun haben: So kann gefragt werden, wer denn eigentlich die Anerkennung verdient, wenn Alpha-Go die weltbesten Go-Spieler schlägt.¹⁷

Um die Diskussion einzugrenzen, sind zwei Hinweise sinnvoll. Zum einen muss an den aktuellen technischen Entwicklungsstand erinnert werden. Auch wenn mittlerweile in vielen Bereichen von hochgradiger Automatisierung gesprochen wird, handelt es sich in der Regel allenfalls um eine teilweise Automatisierung. Wenn beispielsweise von sogenannten selbstfahrenden Autos oder einem „Autopiloten“ im Fahrzeug gesprochen wird, mag dies suggerieren, dass ein System vollumfänglich die Fahrt übernimmt. Folgt man jedoch der üblichen Taxonomie der *Society of Automotive Engineers*, lassen sich sechs Automatisierungsstufen unterscheiden, begon-

¹⁴ Erklärung von Montreal für eine verantwortungsvolle Entwicklung von künstlicher Intelligenz 2023, URL: declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/AL-UdeM_Declaration-IA-Resp_allemand_vf.pdf, zuletzt aufgerufen am; 30.06.2023, S. 16.

¹⁵ Nyholm (Fn. 1), S. 51 ff.

¹⁶ J. Himmelreich, Responsibility for Killer Robots, *Ethical Theory and Moral Practice*, 3 (2019), S. 731 (734).

¹⁷ Gunkel (Fn. 7), S. 312.

nen bei Stufe 0 bis Stufe 5.¹⁸ Viele bislang verfügbare „selbstfahrende“ Autos befinden sich hier nur auf einer Stufe der teilweisen Automatisierung (Stufe 2), bei der der Fahrer stets das Fahrverhalten überwachen muss und damit in definierter Weise die Kontrollbedingung erfüllt. Hier von möglichen Verantwortungslücken zu sprechen kann in die Irre führen und wird mitunter als strategisch motiviertes Argument vorgebracht, mit der sich bestimmte Instanzen (z. B. die Hersteller von entsprechenden Geräten) von jeglicher Verantwortung freizustellen versuchen. Die Frage der Verantwortungslücken ist in diesem Sinn eine auf die Zukunft gerichtete Frage, die in plausibler Weise antizipiert, dass echte Verantwortungslücken heutzutage wahrscheinlich noch eher selten sind, mit fortschreitender Entwicklung der entsprechenden Technologien aber durchaus zunehmend auftreten dürften.

Ebenso unterschieden werden muss die Frage nach möglichen Verantwortungslücken zweitens von einem Fehlgebrauch oder Missbrauch einer Maschine, oder von Fällen technischen Versagens oder tragischer Fehlschläge. Im Zusammenhang mit dem Da-Vinci-Operationssystem sind weltweit tausendfache Gerätestörungen oder Fehlfunktionen bekannt.¹⁹ Derartige Fehlschläge, die im schlimmsten Fall auch zum Tode auf dem Operationstisch führen können, evozieren keine Verantwortungslücken im hier diskutierten Sinne und stellen auch keine grundsätzlich neuen Herausforderungen dar. Hier ist nicht eine Maschine in relevanter Hinsicht für einen Fehler kausal verantwortlich, sondern technisches Versagen oder Fehlbedienung, die gegebenenfalls auf einen menschlichen Akteur zurückgeführt werden können. Entscheidend für Verantwortungslücken ist, dass die entsprechende Aktion (sei es die Lenkbewegung eines Autos, der Abschuss eines ballistischen Geschosses oder die Schnittführung während einer Operation) auf Grundlage einer algorithmischen Entscheidung durch das jeweilige System erfolgte, allerdings ohne dass diese automatisierte Entscheidung auf die Intention oder Kontrolle eines dahinterliegenden Akteurs zurückgeht.

IV. Verantwortungslücken in der Medizin

Der Erste, der sich aus philosophischer Perspektive explizit mit dem Thema der Verantwortungslücken beschäftigt hat, war der Technikphilosoph Andreas Matthias, der vor etwa 20 Jahren in einem Aufsatz den Begriff der Verantwortungslücken (*responsibility gaps*) einführte.²⁰ Matthias fokussierte dabei auf den zunehmenden Einsatz von selbstlernenden Algorithmen, mit Ansätzen die heute unter dem Schlagwort vom Maschinellen Lernen diskutiert werden. Für Matthias wird aus dem Programmierer (*coder*) zunehmend der Kreateur, da die wachsende Autonomie von Maschi-

¹⁸ *On-Road Automated Driving (ORAD) committee*, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles 2021, 22736:2021.

¹⁹ H. Alemzadeh/J. Raman/N. Leveson/Z. Kalbarczyk/Iyer, Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data, *PloS one*, 4 (2016), e0151470 (7).

²⁰ Matthias (Fn. 7), S. 176.

nen das technische Geschöpf von ihrem Schöpfer entkoppelt. Verantwortungslücken als ein Ausdruck dieser Entkopplung führen für Matthias in ein Dilemma: Entweder müsse man auf hochpotente autonome Systeme verzichten oder akzeptieren, dass diese eben zu Verantwortungslücken führen.

Obwohl Medizin und Gesundheitswissenschaften zu den Vorreitern eines praktischen Einsatzes eben jener Technologien gehören und zugleich über ein ausgeprägtes und höchst ausdifferenziertes Verantwortungsethos verfügen, haben Fragen möglicher Verantwortungslücken hier überraschend wenig Aufmerksamkeit erfahren. In den vergangenen Jahren sind zunehmend sogenannte intelligente Systeme in der Klinik angekommen. Die US-amerikanische Food & Drug Administration (FDA) hat mittlerweile mehr als 500 Zulassungen für entsprechende Produkte ausgestellt.²¹ Während die meisten Entwicklungen in der Radiologie bzw. Bildgebung Einsatz finden, sind andere Systeme auch unmittelbar an der Diagnose oder Behandlung von Patienten beteiligt. Hier sind insbesondere Systeme des maschinellen Lernens zu nennen, die zugleich häufig in Form einer Black Box auftreten können.²² Für Anhänger der digitalen Medizin ist dies nur ein Auftakt: Ähnlich wie im Bereich der selbstfahrenden Autos, meinen viele, dass die Automatisierung in der Klinik weiter vorangetrieben werden sollte. Wo dies in Zukunft der Fall sein wird, stellt sich auch die Frage nach Verantwortungslücken. Moderne Diagnose und Screeningsysteme reichen zuweilen bereits an die Leistungen von erfahrenen Expertinnen und Experten heran. Sie könnten in der Zukunft außerhalb einer direkten menschlichen Kontrolle eingesetzt werden. Auch dort wo die menschlichen Entscheidungsfähigkeiten aufgrund von Zeitdruck oder Komplexität an ihre Grenze gelangen – etwa in der Notfallmedizin – sind autonome Entscheidungen von Maschinen vorstellbar. In all diesen Fällen ist davon auszugehen, dass große Anreize bestehen, die Potenziale autonom agierender Technologien auszunutzen und sie dazu der menschlichen Kontrolle zu entheben. Gleichzeitig ist jedoch auch von diesen Geräten keine Fehlerfreiheit zu erwarten.

V. Wer trägt nun Verantwortung?

Ein vielversprechendes Anwendungsfeld für KI und robotische Systeme stellt dabei die Chirurgie dar. Bekannte und existierende Systeme, wie das „Da-Vinci-Operationssystem“, sind derzeit Assistenzsysteme, die einen besseren Outcome versprechen und die Fähigkeiten des Operators unterstützen. Gemäß der Taxonomie für autonome Systeme wären diese derzeit im Bereich der Stufen 0 oder 1 zu verorten, da

²¹ *U.S. Food & Drug Administration*, Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices 2023, URL: [fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices](https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices) [zuletzt aufgerufen am: 30.06.2023].

²² *H. Smith*, Clinical AI: opacity, accountability, responsibility and liability, *AI & SOCIETY*, 2 (2021), S. 535 (540).

sie lediglich die Entscheidungen des Operators umsetzen.²³ Mittlerweile werden aber auch ausgereifere Systeme, etwa für die Implantation von Cochlea-Implantaten verwendet. Im Jahr 2022 wurde mit dem STAR-System (Smart Tissue Autonomous Robot) in einer präklinischen Studie eine erste teil-autonome Laparoskopie am Tiermodell durchgeführt.²⁴

Kommt es bei einer Operation durch (zukünftige) autonome Systeme zu einem „Kunstfehler“, wirft dies die Frage nach der Verantwortungszuschreibung auf. Stellen wir uns die Situation möglichst plastisch vor: Angenommen ein Chirurgieroboter erlangt durch die Kopplung mit einem KI-System eine spezifische Form von Agentivität, die im Sinne von funktionaler Autonomie als hinreichend erachtet werden kann. Dieses System könnte etwa für Operationen im Bauchraum eingesetzt werden. Es könnte durch die Analyse vorliegender Operationsdaten erlernt haben, dass der mechanische Ausgleich der Atembewegungen von Patientinnen und Patienten im Rahmen der vorgegebenen Zielstellung der Operation ein wichtiger Faktor ist. Auf diesem Weg könnte es beispielsweise möglich sein, die relative Lage von Operationsfeld und Instrument zueinander so zu stabilisieren, dass extrem präzise Eingriffe etwa in der Nähe großer Blutgefäße (z. B. im Rahmen einer Tumorentfernung) möglich werden. Sagen wir weiter, das System entschiede auf dieser Grundlage eigenständig über Zeitpunkt, Amplitude und Richtung der Ausgleichsbewegung. Die mechanische Bewegung der Instrumente, die das System daraufhin vollziehen würde, wäre – anders als das übergeordnete Ziel der Operation – nicht durch Anwender, Hersteller und Programmierer vorgegeben. Sie unterläge nicht ihrer Kontrolle und könnte durch die Beteiligten nicht vorhergesagt werden. Der begleitende Chirurg hätte ebenfalls keine Möglichkeit die Bewegung vorherzusagen, zu kontrollieren oder womöglich sogar zu korrigieren. Entstände bei einer solchen Operation, zum Beispiel durch die Verletzung einer Bauchraumarterie ein Schaden, so könnte zugleich auch eine Verantwortungslücke entstehen. Wie Sven Nyholm an einem anderen Beispiel argumentiert, sind Unkontrollierbarkeit und Unvorhersehbarkeit hier unmittelbare Folge der Agentivität und funktionalen Autonomie des Systems, die ausreichen, um es der Kontrolle (und damit der moralischen Verantwortung) des Operateurs zu entheben. Zugleich sind diese Agentivität und Autonomie jedoch nicht in einem solchen Maß vorhanden, dass man dem Gerät den Status eines moralischen Agenten – und damit eines Verantwortungsträgers – zuweisen könnte.²⁵

Wenn solche Verantwortungslücken tatsächlich entstehen, wäre dies revisionär hinsichtlich der sozialen Praxis der Verantwortungszuschreibung und würde Moral und Recht grundsätzlich in Frage stellen. Wie Char und Kollegen in einem vielbeachteten Aufsatz im *New England Journal of Medicine* konstatieren, kann

²³ A. Attanasio/B. Scaglioni/E. Momi et al., *Autonomy in Surgical Robotics*, Annual Review of Control, Robotics, and Autonomous Systems, 1 (2021), S. 651 (652).

²⁴ H. Saaidi/J. Opfermann/M. Kami/S. Wei et al., *Autonomous robotic laparoscopic surgery for intestinal anastomosis*, Science robotics, 62 (2022), eabj2908 (3).

²⁵ Nyholm (Fn. 1), S. 57; Königs (Fn. 7), S. 2.

sich unter anderem mit der Einführung von selbstlernenden Systemen in der Medizin nicht weniger als die Aufgabe von persönlicher Verantwortlichkeit im Rahmen der Arzt-Patienten-Beziehung ergeben.²⁶ Dann wäre beispielsweise ein Behandlungsfehler (wie im Rahmen des obigen Beispiels) in etwa vergleichbar mit einem Tornado, der durch den Operationssaal fegt. Er kann viel Schaden anrichten, aber verantwortlich ist für diese Katastrophe wohl niemand.

Eine Reihe von Autoren meinen, dass die Möglichkeit von Verantwortungslücken und damit verbundene Herausforderungen für das Moralsystem so frappierend sind, dass entsprechende Technologien nicht entwickelt oder eingesetzt werden dürften. Während eine solche Schlussfolgerung am Beispiel von autonomen Waffensystemen womöglich plausibel gemacht werden kann,²⁷ erscheint dies bei lebensrettenden Technologien in der Medizin unwahrscheinlich. Immerhin wäre der Verzicht mit weit höheren Opportunitätskosten verbunden. Mit Blick auf das obige Beispiel ließe sich darauf verweisen, dass die skizzierte Lagestabilisierung durch eine KI und die damit verbundene Präzision bestimmte Operationen überhaupt erst ermöglicht und damit große Vorteile hätte.

Es erscheint daher lohnenswert einen zweiten Blick auf Verantwortungslücken zu werfen und die Frage zu stellen, ob sich diese nicht zumindest teilweise füllen ließen.²⁸ In jüngerer Zeit wurden hierzu eine Reihe von Lösungen vorgeschlagen, die sich vor allen Dingen an der Vorstellung orientieren, dass es möglich sei, Verantwortungssubjekte zu identifizieren, wenn man nur genau genug nach ihnen sucht. Zu den Vorschlägen gehören der Versuch, Maschinen selbst Verantwortung zuzuweisen, der Versuch, relevante menschliche Akteure doch argumentativ zur Verantwortung zu ziehen, sowie der Versuch, Verantwortungssubjekte jenseits von Individuen zu identifizieren.

1. Maschinenverantwortung

Eine vielleicht naheliegende Antwort wäre in Fällen wie der oben genannten Operation, die Maschine bzw. das System verantwortlich zu machen. Immerhin zeigt sich ein unmittelbarer kausaler Zusammenhang zwischen der „Entscheidung“ der Maschine und dem unerwünschten Ergebnis. Dies schließt an Überlegungen hinsichtlich der Handlungsfähigkeit bzw. Agentenschaft von Robotern an, die über reine funktio-

²⁶ D. Char/N. Shah/D. Magnus, Implementing Machine Learning in Health Care – Addressing Ethical Challenges, *The New England journal of medicine*, 11 (2018), S. 981 (983).

²⁷ R. Sparrow, Killer Robots, *Journal of Applied Philosophy*, 1 (2007), S. 62 (74).

²⁸ Während Einigkeit herrscht, dass autonome Technologie Herausforderungen für Verantwortungszuschreibung bergen, hinterfragen einige Autoren, ob es Verantwortungslücken tatsächlich geben kann (verg. zum Beispiel D. Tigard, There Is No Techno-Responsibility Gap, *Philosophy & technology*, 3 (2021), S. 589 (590).) Andere Autoren heben zugleich auch mögliche entlastende Potenziale von Verantwortungslücken hervor (vgl. Danaher (Fn. 7), S. 14).

nale Autonomie hinausgehen. Der Philosoph James Moor²⁹ hat hierzu in einem vielbeachteten Essay vier verschiedenen Typen von sogenannten moralischen Maschinen (*ethical robots*) unterschieden: i) *Ethical-impact agents*, ii) *implicit ethical agents*, iii) *explicit ethical agents*, und iv) *full ethical agents*. Nur letzte könnten dabei beanspruchen, vollständig handlungsfähig und damit womöglich auch Träger von Verantwortung zu sein. Christian List³⁰ meint, dass Systeme der Künstlichen Intelligenz, zumindest wenn diese sehr fortgeschritten seien, durchaus verantwortungsfähige Agenten sein könnten. Dies gelte zumindest in einem Sinn, in dem auch menschlichen Kollektiven unter bestimmten Bedingungen Verantwortung zugeschrieben werden kann.

Ungeachtet der konkreten Ausgestaltung der jeweiligen Argumente, gehen diese Positionen davon aus, dass Maschinen in bestimmter Weise Verantwortung zugeschrieben werden kann. Sie tun dies entweder indem argumentiert wird, dass Maschinen einen Grad von Autonomie erreichen können, der sie nicht nur funktional befähigt, sondern darüber hinaus auch zu einer Art moralischer Handlungsfähigkeit führt. Dies könnte manche Maschine in den Status eines moralischen Agenten heben. Oder aber sie machen deutlich, dass sie zwar nicht die Eigenschaften einer moralischen Handlungsfähigkeit erfüllt sehen, es sich aber gleichwohl im Sinne einer sozialen Konvention als zielführend erweisen könnte, intelligenten Systemen Verantwortung zuzusprechen.³¹

Dem gegenüber steht der gegenwärtige Entwicklungsstand. Er gibt kaum Anhaltspunkte, die es erlauben würden, Maschinen als vollumfängliche Verantwortungssubjekte zu verstehen. Die Frage, ob Entwicklungen der KI jemals an einen solchen Punkt kommen können, ist ungeklärt und soll an dieser Stelle nicht weiter verfolgt werden. Selbst wenn es allerdings so wäre, bliebe offen, inwieweit die übliche soziale Praxis der Verantwortung hier auf künstliche Akteure übertragen werden kann. Zur Funktion von Verantwortung gehören schließlich auch noch weitere soziale Aspekte, von Rechenschaft auf Seiten der Verantwortungsträger bis hin zu Dynamiken aus Lob, Tadel oder auch Strafe. Selbst wenn also Maschinen verantwortlich sein könnten, wäre noch immer unklar, in welcher Weise diese beispielsweise sinnvoll Adressat von Strafen sein könnten.³² In gewisser Weise erinnert die Idee, Maschinen zur Verantwortung ziehen zu wollen, an mittelalterliche Strafprozesse gegen Tiere oder den Versuch ganz und gar menschliche Konzepte, die über lange Zeit an der Interaktions- und Sozialpraxis von Menschen geschärft worden sind, in einen anderen und eher unpassenden Kontext zu übertragen.

²⁹ J. Moor, The Nature, Importance, and Difficulty of Machine Ethics, IEEE Intelligent Systems, 4 (2006), S. 18 (19).

³⁰ C. List, Group Agency and Artificial Intelligence, Philosophy & technology, 4 (2021), S. 1213 (1217).

³¹ Gunkel (Fn. 7), S. 316.

³² Sparrow (Fn. 27), S. 71; Nyholm (Fn. 1), S. 56.

Zu bedenken ist zudem abschließend, dass selbst wenn Maschinen in einem gewissen Sinne autonom agieren und interagieren, diese noch immer einer festgelegten Zielstellung unterliegen, die auf ihr Design zurückzuführen ist, ggf. überwacht wird und vorher bestimmten Regeln folgt. Nyholm³³ spricht hier von verschiedenen Formen einer Domänen-spezifischen Handlungsfähigkeit. So mag zwar ein weiter ausgeklügeltes Da-Vinci-System einen Operationsfehler begehen, also für den vorgesehenen Zweck versagen. Es wird sich allerdings nicht verselbstständigen oder beispielsweise gezielt einen Mord begehen. Der Aktionsradius dieser Maschine ist in gewisser Weise *hard-wired*. Dies nimmt ihr nicht nur die für Verantwortung notwendige Freiheit (siehe hierzu die Ausführungen zum Rechtsbegriff der Verantwortung), sondern lässt es auch plausibel erscheinen, die Suche nach den Verantwortungssubjekten an der Stelle fortzusetzen, wo ein Akteur auf diese Zielstellungen Einfluss zu nehmen vermag.

2. Menschenverantwortung

Dabei ist in erster Linie an die verschiedenen menschlichen Akteure zu denken, die sich in Interaktionszusammenhängen mit der Technologie befinden. Sie stehen im Zentrum der zweiten Gruppe von Vorschlägen, die jedoch ebenfalls mit einer Reihe von Schwierigkeiten versehen sind. Diesen Versuchen kann entgegengehalten werden, dass bei Fehlschlägen in autonomen Systemen kaum einem Menschen, sei es z. B. einem Ingenieur oder einem Arzt, die Verantwortung zugeschrieben werden kann, ohne damit weitere Widersprüche zu erzeugen. Aus entsprechenden Studien ist beispielsweise bekannt, dass Menschen dazu neigen, denjenigen die Schuld an Schäden durch Maschinen zu geben, die sie bedienen oder dicht an ihnen stehen.³⁴ Dies wäre in vielen Fällen allerdings nicht nur unangemessen, sondern auch unfair. Selbst bei einer menschlichen Aufsicht einer Maschine – dem *human-in-the-loop* – sind bei (teil)autonomen Systemen regelhaft Konstellationen denkbar, wo keine bestimmte Person einen relevanten Beitrag an negativen Ereignissen hat. Dies ist unter anderem durch die Komplexität entsprechender Systeme und Handlungszusammenhänge zu erklären, sowie durch die vielfältigen Akteure, die an Entwicklung und Verwendung von Systemen beteiligt sind. So kann ein chirurgischer Roboter von einem Arzt A bedient werden, wobei das System vom Hersteller B entwickelt und von C installiert worden ist, die Hardware von D produziert und durch ein KI-System von E unterstützt wird, welches mit Daten aus unterschiedlichen Quellen durch F trainiert wurde und so fort.³⁵ Wenn sich etwa im Fall des robotischen Kunstfehlers herausstellt, dass dieser auf Verzerrungen in den Trainingsdaten eines autonomen

³³ Nyholm (Fn. 7), S. 58 ff.

³⁴ A. Jamjoom/A. Jamjoom/J. Thomas et al., Autonomous surgical robotic systems and the liability dilemma, *Frontiers in Surgery*, 2022, S. 1 (4).

³⁵ F. Santoni de Sio/G. Mecacci, Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them, *Philosophy & technology*, 4 (2021), S. 1057 (1062).

Systems zurückzuführen ist,³⁶ scheint sich kaum ein bestimmter Akteur für die Zuschreibung von Verantwortung anzubieten. Erschwerend kommt die vielbeschriebene Opazität von KI hinzu. Regelmäßig ist auch für die Entwickler (und noch viel mehr für Anwender) von Systemen kaum nachvollziehbar, wie Entscheidungen oder Outputs zustande kommen.

3. Kollektive oder kollaborative Verantwortung

Im Anbetracht dieses Geflechts scheint eine weitere naheliegende Lösung zu sein, hier auf Kollektive als mögliche Verantwortungsträger von robotischen Kunstfehlern auszuweichen. Auch dies ist eine Lösungsstrategie, die im Zusammenhang mit anderen autonomen Systemen vielfach diskutiert wird. So meint etwa Nyholm,³⁷ dass sich Verantwortungslücken schließen lassen, wenn vermeintlich autonome Entscheidungen von Maschinen als das Ergebnis einer Mensch-Maschinen-Interaktion und Zusammenarbeit verstanden werden. Nyholm spricht in diesem Zusammenhang von einer kollaborativen Verantwortung, die sowohl einem Roboter, als auch dem interagierenden Menschen zukomme. Offen bleibt bei ihm allerdings die Frage, wie dann die Verantwortung zwischen Menschen und Maschine verteilt werden müsste. Hier scheint naheliegend, dass sich Verantwortungslücken entweder erneut auftun, da Maschinen eben doch keine adäquate Verantwortungssubjekte sind, oder alleine dem menschlichen Akteur die Last der Verantwortung auferlegt wird. Auch bei der kollaborativen Verantwortung wiederholt sich zudem die oben beschriebene Herausforderung, dass bei Entwicklung und Verwendung komplexer KI-Systeme vielfältige Akteure beteiligt sind. Folglich müssten Fragen der Verantwortlichkeit für Fehler von KI-Systemen womöglich ähnlich behandelt werden, wie die von Kollektiven – ein bekanntes Thema aus der Wirtschaft- oder Organisationsethik. Damit verbindet sich allerdings eine weitere Hürde, namentlich die hierbei drohende Verantwortungsdiffusion, die im Zusammenhang mit dem Versuch bekannt ist, Kollektive zu Verantwortungssubjekten zu erklären. Tatsächlich wurde unlängst von Bleher und Braun³⁸ vorgeschlagen, die Verantwortungslücken vielmehr als eine Verantwortungsdiffusion zu betrachten. Während diese Deutung einen Erklärwert besitzt, bleibt allerdings offen, wie eine Verantwortungslücke als diffuse Kollektivverantwortung gefüllt werden kann.³⁹ Im Kontext von hochkomplexen autonomen Maschinen verschärft sich schließlich das sogenannte „Problem der vielen Hände“, da hier

³⁶ H. Bleher/M. Braun, Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems, AI and ethics, 4 (2022), S. 747 (755).

³⁷ Nyholm (Fn. 7), S. 62 ff.

³⁸ Bleher/Braun (Fn. 36), S. 751.

³⁹ R. Jong, The Retribution-Gap and Responsibility-Loci Related to Robots and Automated Technologies: A Reply to Nyholm, Science and engineering ethics, 2 (2020), S. 727 (731).

nun nicht nur viele Hände, sondern auch viele Dinge und Artefakte zur Fragmentierung der Verantwortung beitragen.⁴⁰

VI. Verantwortung übernehmen?

Die skizzierten Vorschläge Verantwortung Menschen, Maschinen oder Kollektiven zuzuschreiben, erweisen sich vor diesen Überlegungen als ungeeignet, um die durch autonome Systeme entstehenden Verantwortungslücken zu schließen. Hier war bislang entscheidend, dass für eine moralische Verantwortung eine Kontrollbedingung sowie epistemische Bedingungen erfüllt sein müssen, damit einem Akteur Verantwortung zugeschrieben werden kann. Verantwortung wurde zudem durch gewisse Relata gekennzeichnet: Das Verantwortungssubjekt ist hinsichtlich eines Verantwortungsgegenstandes gegenüber einem bestimmten Verantwortungsobjekt verantwortlich, beispielsweise der Arzt gegenüber dem Patienten. In dieser Konstellation erschien es in der Tat eine Herausforderung im Hinblick auf die Kontrollbedingung, wenn ein autonomes System in diese Dyade eindringt.

Mindestens im Kontext des ärztlichen Handelns und des entsprechenden professionellen Ethos drängt sich jedoch noch eine weitere Möglichkeit auf, die in der Diskussion bisher weniger Beachtung gefunden hat. Neben dem Zuschreibungsbegriff von Verantwortung, kennt die moralische Praxis noch ein weiteres Phänomen, nämlich die *Verantwortungsübernahme*, das insbesondere im Rahmen der so genannten Professions- und Rollenverantwortung von Bedeutung ist. Damit ist eine spezifische Form der Verantwortung gemeint, die eine Person qua ihrer Funktion innerhalb eines bestimmten Kontextes übernimmt. So können Eltern die Verantwortung für ihre Kinder übernehmen oder Tierhalter für ihr Tier. Diese Form der Verantwortungsübernahme umfasst nicht nur einen prospektiven Sinn, nach dem beispielsweise Eltern gewisse Fürsorgepflichten für ihre Nachkommen übernehmen. Vielmehr schließt diese Verantwortungsübernahme auch ein, sich zuweilen für seine Kinder zur Verantwortung ziehen zu lassen, beispielsweise Tadel auf sich zu nehmen, oder sich für ihr Fehlverhalten zu entschuldigen. Markant an diesen Formen der Verantwortungsübernahme ist erstens, dass diese nicht oder nur indirekt (qua Rollenverantwortung) auf eine Zuschreibung angewiesen ist. Der Akt der Verantwortungsübernahme kann in der Hand des verantwortungsübernehmenden Subjekts liegen.⁴¹ Zweitens ist für eine Verantwortungsübernahme keine unmittelbare kausale Verantwortung notwendig, die für eine relevante Handlung vorliegen muss. Die oben genannte Kontrollbedingung der Verantwortung verliert damit an Bedeutung.

⁴⁰ M. Coeckelbergh, Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability, Science and engineering ethics, 4 (2020), S. 2051 (2056).

⁴¹ I. van de Poell/M. Sand, Varieties of responsibility: two problems of responsible innovation, Synthese, S19 (2021), S. 4769 (4774).

1. Warum sollte jemand Verantwortung für Maschinen übernehmen wollen?

Die sich aufdrängende Frage ist daher, ob auf diesem Weg Verantwortungslücken geschlossen werden können, indem eine Person die Verantwortung für eine Maschine *übernimmt*. Der erste kritische Einwand, der in dieser Hinsicht zu berücksichtigen ist, ist sicherlich, warum eine Person dies überhaupt tun sollte. Diese berechnete Frage hat einige Autoren dazu bewegt, die Idee von Verantwortungsübernahmen im Kontext von autonomen Systemen abzulehnen. So arbeitet Nyholm eine verbreitete Asymmetrie in unserer sozialen Praxis heraus: Einen Akteur zu finden, der bereit ist, eine sogenannte positive Verantwortung zu übernehmen, der also bereit ist, berechtigtes Lob entgegen zu nehmen, wenn etwas gelingt, ist leicht.⁴² Dagegen einen Akteur zu finden, der die Last einer negativen Verantwortung auf sich nimmt, wenn etwas schiefgegangen ist, ist häufig unmöglich. Diese Feststellung kann als kritische Diagnose Bestand haben: Man denke in diesem Fall nur an das oben entwickelte Beispiel des chirurgischen Roboters oder des Herstellerunternehmens von selbstfahrenden Autos. Es dürfte dem beteiligten Chirurgen keine Schwierigkeiten bereiten, das Lob für die außergewöhnliche und lebensrettende Präzision der Operation entgegenzunehmen und in diesem Sinn die Verantwortung für ein gutes Gelingen zu übernehmen. Ähnliches dürfte für den Hersteller gelten. Kommt es im Falle des selbstfahrenden Autos zu einem Unfall im „Autopilot“ oder setzt der Chirurgieroboter den Schnitt an der falschen Stelle, dürften Hersteller und Personen weit weniger geneigt sein, die Verantwortung auf sich zu nehmen.

Die verbreitete Skepsis gegenüber einem volitionalen und gutwilligen Akt der Verantwortungsübernahme ist durchaus verständlich. Sie sollte jedoch nicht dazu verleiten, diese Idee zu früh als naiv abzutun. Dies gilt insbesondere dann, wenn sie im Kontext einer bestimmten Rollenmoral gebraucht wird, die es erlaubt, die Übernahme von Verantwortung nicht allein als freiwilligen Akt, sondern als Teil einer moralischen Pflicht zu präskribieren. In diesen Bereich fallen nicht nur die eingangs genannten Eltern, die für ihre Kinder verantwortlich sind, sondern auch Träger professioneller Rollen. So kann es Teil der Rollenverantwortung einer guten Führungsperson sein, die Verantwortung für seine Mitarbeiter zu übernehmen, oder es kann Teil einer politischen Rollenverantwortung ausdrücken, Verantwortung für Vorgänge im eigenen Dienstbereich zu tragen. In gleicher Weise kann es Teil der professionsethischen Verantwortung der Rolle von Ärztinnen und Ärzten sein, Verantwortung für Patientinnen und Patienten zu übernehmen.

Im Zusammenhang mit autonomen Systemen wurde diese Form von Verantwortungsübernahme als Teil einer professionellen Rollenmoral unlängst eingeführt. So haben Champagne und Tonkens⁴³ argumentiert, dass sich beim Einsatz von autono-

⁴² S. Nyholm, Artificial Intelligence, Responsibility Gaps, and Asymmetries between the Good and the Bad, im Erscheinen.

⁴³ M. Champagne/R. Tonkens, Bridging the Responsibility Gap in Automated Warfare, Philosophy & technology, 1 (2015), S. 125 (126). Vergl. auch M. Kiener, Can we Bridge AI's responsibility gap at Will?, Ethical Theory and Moral Practice, 4 (2022), S. 575 (582).

men Waffensystemen mögliche Verantwortungslücken schließen ließen, indem z. B. ein Heerführer qua professioneller Rolle die Übernahme der prospektiven Verantwortung für tödliche Systeme trägt. Der Kommandeur könne sich damit sozusagen zum Subjekt der Verantwortung erklären und sich somit auch zum gerechtfertigten Adressaten von moralischer Anklage machen, falls es auf dem Kampfplatz zu Verletzungen des Kriegsrechts kommt. Auf diese Weise, so Champagne und Tonkens, ließen sich manche Verantwortungslücken reduzieren.

Dieser Vorschlag kann als Vorlage für einen Versuch dienen, Verantwortungslücken in der digitalen Medizin zu schließen. Denn gerade das ärztliche Berufsethos – die Rollenmoral, die mit professionsspezifischen Verpflichtungsdimensionen einhergeht – erscheint ein geeigneter Ausgangspunkt für Verantwortungsübernahmen. Das professionsspezifische Ethos von Ärztinnen und Ärzten ist einerseits stark ausdifferenziert und wirkmächtig. Vor allen Dingen aber bindet es die Güte des ärztlichen Handelns durch die Pflicht zu Sorge und Fürsorge stets an das Wohl von Patientinnen und Patienten. Es verlangt damit, das Ärztinnen und Ärzte qua ihrer Fähigkeiten und Expertise Verantwortung für Andere übernehmen. Mindestens graduell unterscheidet dies die ärztliche Profession damit auch von anderen beteiligten Gruppen wie Ingenieuren oder Programmierern, bei denen trotz verschiedener Bemühungen eine analoge Rollenmoral bislang allenfalls in Ansätzen erkennbar ist.⁴⁴

2. Wofür können Ärzte Verantwortung übernehmen?

Die Frage ob Ärztinnen und Ärzte einspringen *können*, wenn es darum geht Verantwortung für autonome technische Systeme zu übernehmen, lässt sich so prima facie bejahen. Die ärztliche Rollenverantwortung mit ihrem starken Fokus auf Sorge und Wohlergehen von Patientinnen und Patienten eröffnet darüber hinaus durchaus Möglichkeiten, diese Übernahme als Teil der Rolle eines guten Arztes oder einer guten Ärztin zu empfehlen. Ärztinnen und Ärzte können nicht nur – sie sollten unter bestimmten Umständen Verantwortung für funktional autonome Maschinen übernehmen.

Auf diese Weise scheint es möglich einige Verantwortungslücken zu schließen. Zugleich muss dieser Befund jedoch mit Einschränkungen versehen werden, die eine absurde Entgrenzung des Verantwortungsbegriffs vermeiden. Um sinnvoll von einer Verantwortungsübernahme sprechen zu können, müssen zunächst einmal bestimmte Voraussetzungen erfüllt sein. Selbst wenn für eine Verantwortungsübernahme die genannte Kontrollbedingung an Bedeutung verliert, ist es keineswegs sinnvoll anzunehmen, dass es möglich ist, Verantwortung für jeden beliebigen Zustand zu übernehmen. Sonst wäre es auch möglich, willkürlich Verantwortung für Vorgänge fernab des ärztlichen Handlungsbereiches zu übernehmen, ebenso wie

⁴⁴ Vergl. mit einem Vorschlag in dieser Richtung: T. Goetze, Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement, in: FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, S. 390 (395).

für Zustände, die sich prinzipiell jedem Einfluss entziehen. Ebenso wenig wie Personen, die Verantwortung für das Wetter übernehmen können, kann ein Arzt Verantwortung für robotische Behandlungsfehler übernehmen, die vollständig abseits seiner Handlungssphären stattfinden.

Maximilian Kiener⁴⁵ hat im Zusammenhang mit Verantwortungslücken kürzlich zwei Aspekte vorgeschlagen, die Minimalbedingungen für gerechtfertigte Verantwortungsübernahmen auszeichnen: Die *räumliche und zeitliche Nähe* des Subjekts der Verantwortungsübernahme zum Verantwortungsgegenstand sowie die *Relevanz des Subjekts* im kausalen Geflecht des Verantwortungsgegenstandes. So können beispielsweise Programmierer eine hohe Relevanz für die Entwicklung entsprechender Systeme in der Medizin aufweisen. Zugleich fehlt es ihnen an räumlicher und zeitlicher Nähe zur Behandlungssituation. Ärztinnen und Ärzte dagegen sind unmittelbar am Patienten. Zugleich haben sie eine hohe Relevanz. Denn selbst wenn sie manche Wirkungen von autonomen Systemen nicht kontrollieren könnten, sind sie doch wesentlich für die Indikationsstellung und somit relevante Akteure für die Mittelwahl.

3. Kann Verantwortungsübernahme eingefordert werden?

Die zweite Einschränkung betrifft die Frage in welchem Rahmen es für Ärzte möglich ist, Verantwortung für robotische Behandlungsfehler zu übernehmen, beziehungsweise in welchem Rahmen eine Verantwortungsübernahme als Teil des Professionsethos gefordert werden kann. Das Phänomen von Verantwortungsübernahmen abseits von Rollenverantwortung wird häufig als eine Form von übergebührlichem Handeln interpretiert.⁴⁶ Leihst sich Hans das Auto seiner Bekannten Aisha und hat der Wagen auf den ersten Metern ohne Hans' Zutun einen Motorschaden, ist es sicherlich lobenswert, wenn Hans sich an den Reparaturkosten beteiligt und damit in gewisser Weise auch Verantwortung für die Wiedergutmachung übernimmt. Zugleich könnte Aisha diese Form der Verantwortungsübernahme wahrscheinlich nicht von Hans einfordern, weil dieser keinen relevanten Beitrag am Motorschaden hatte. Hans' Übernahme von Verantwortung könnte maximal gelobt werden, bliebe jedoch als supererogatorisches Verhalten außerhalb seiner Pflichten.

Im Rahmen der ärztlichen Rollenverantwortung lägen die Dinge anders: Hier kann die Übernahme als integraler Bestandteil der Verantwortung qua professioneller Funktion verstanden werden, die sich auch moralisch einfordern lässt. Wenn ein Arzt nach einem robotischen Kunstfehler durch einen hinreichend autonomen Roboter untätig und mit einem Schulterzucken den OP verlässt, wird er seiner Rolle nicht gerecht. Vielmehr würde er dann zum legitimen Adressat von moralischem Tadel werden, wenn er es in Folge des robotischen Fehlers versäumt, Verantwortung zu übernehmen. Damit würde er nicht für den Fehler der Maschine zur Verantwortung

⁴⁵ Kiener (Fn. 43), S. 588.

⁴⁶ van de Poel/Sand (Fn. 41), S. 4779.

gezogen. Vielmehr würde man ihm anlasten, auf den robotischen Kunstfehler eben nicht adäquat reagiert zu haben, in dem er sich einer Verantwortungsübernahme entzogen hat.

4. Was folgt aus der übernommenen Verantwortung?

Im Hinblick auf die Frage, was aus der Übernahme von Verantwortung erfolgt, müssen weitere Einschränkungen gemacht werden. Sicher ist, dass Ärztinnen und Ärzte nicht zum gerechtfertigten Gegenstand von Empörung und Schuld werden sollten, wenn es um den beschriebenen Fall einer Verantwortungslücke geht, die durch die Übernahme von Verantwortung geschlossen wird.

In der alltagsmoralischen Sprechweise von Verantwortung fällt die Zuweisung von Schuld und die Zuweisung von Verantwortung als das oben beschriebene Entstehen-müssen in der Regel zusammen. Für die Frage der gerechtfertigten Schuldzuweisung ist insbesondere entscheidend, ob die Zurechnung der Handlung zum jeweiligen Akteur möglich ist. Hierfür ist kausale Verantwortung ein zentrales Kriterium. In diesem Sinn bleibt auch bei der Übernahme von Verantwortung der Eindruck einer Lücke bestehen, da die entsprechende Kontrollbedingung auch bei übernommener Verantwortung unerfüllt bleibt. Der beteiligte Operateur ist nicht kausal für den entstandenen Schaden eines autonomen Systems verantwortlich. Er kann deswegen auch keine Schuld tragen.

Man könnte an dieser Stelle erneut einwenden, dass die retrospektiven Dimensionen von Schuld und Tadel, die sich intuitiv mit dem Verantwortungsbegriff verbinden, in psychologischer und sozialer Hinsicht durchaus von Bedeutung sind. In ihnen manifestiert sich ein bekannter und universeller Kern von Gerechtigkeitsvorstellungen, der auf den Ausgleich und die Wiederherstellung eines imaginierten Gleichgewichtes nach einem geschehenen Unrecht zielt. Dennoch scheint es fraglich, inwieweit dies als Einwand gegenüber einem Konzept von Verantwortungsübernahme im Rahmen von professioneller Rollenverantwortung taugt. Richtig ist, dass das hier dargelegte Konzept deutlich stärker in Richtung der Frage weist, wer zur Verantwortung gezogen werden kann, also von wem gerechtfertigt verlangt werden kann, für ein Ereignis „Rede und Antwort“ zu stehen. Gesucht ist weniger ein Akteur, der Schuld auf sich nimmt, als jemand der antworten kann – wie sich auch im Begriff der *Verantwortung* oder auch *responsability* ausdrückt. Allerdings kann dieses Antworten-müssen eine ähnliche sozialregulative Funktion übernehmen wie die Idee von Schuld. Mit ihr kann beispielsweise auch verbunden sein, prospektiv für die Heilung vergangener Ereignisse bereitzustehen, selbst wenn diese nicht der Kontrolle des entsprechenden Akteurs unterlagen und ohne dabei auf den Begriff der Schuld zurückzugreifen.⁴⁷ Im Unterschied zu kollektiver und kollaborativer Verantwortung

⁴⁷ Diese Dimensionen von Verantwortung im Umgang mit Technologien weisen eine interessante Parallele zu Debatten im Rahmen von globaler Armut auf. Hier wird zuweilen von einer Beseitigungsverantwortung gesprochen („remedial responsibility“ vergl. D. Miller, Na-

steht mit professioneller Rollenverantwortung zudem ein klar definiertes Verantwortungssubjekt zur Verfügung, das im Unterschied zu Konzepten maschineller Verantwortung zur Übernahme von Verantwortung tatsächlich fähig ist.

Es scheint in diesem Sinne angebracht, dass Ärzte eine ähnliche Verantwortung für robotische Kunstfehler übernehmen sollten. Dies bedeutet nicht, dass Behandler moralisch für den robotischen Kunstfehler Schuld tragen müssen oder sogar rechtlich für Maschinen haften. Es bedeutet aber, dass sie für mögliche entstandene Schäden „Rede und Antwort“ stehen und nicht zuletzt qua Rollenmoral eine prospektive Verantwortung für die Behebungen von möglichen Fehlern tragen.

Welche konkreten Handlungsdimensionen sich aus dieser postulierten Verantwortungsübernahme ergeben, lässt sich erst in spezifischen Kontexten erschließen. Neben den genannten Bemühungen der Wiedergutmachung gehören hierzu sicherlich weitere Verantwortlichkeiten: angefangen von Entschuldigen für den Fehler, über Bemühungen um Aufklärung des Ergebnisses, der Ableitung mögliche Erkenntnisse, die Antizipationen von vergleichbaren Situationen oder auch die Kommunikation entsprechender Ereignisse in der Fachgemeinschaft.⁴⁸ Mit einer Verantwortungsübernahme in diesem, hier nur angedeuteten Sinn, ließen sich zumindest einige der befürchteten Verantwortungslücken durch autonome Systeme auf eine nicht-triviale Weise schließen.

VII. Fazit

Auf Grundlage dieser Überlegungen gewinnt eine Idee von Verantwortung Kontur, die im Hinblick auf das Problem der Verantwortungslücken etwas beizutragen hat. Entwicklungen der Künstlichen Intelligenz und Robotik erlauben die Schaffung von teilautonomen oder zukünftig sogar von autonomen Systemen. Diese können, wie das Beispiel der robotischen Chirurgie zeigt, in verschiedener Hinsicht äußerst dienlich sein. Wenn es beim Einsatz von solchen Systemen allerdings zu automatisierten „Fehlentscheidungen“ oder „Kunstfehlern“ kommt, erscheint es mitunter unmöglich, ein Verantwortungssubjekt zu identifizieren. Die Sorge um hier entstehende Verantwortungslücken hat einige veranlasst, sich für ein Verbot von autonomen Systemen stark zu machen oder zumindest ein Dilemma zu konstatieren: entweder wir müssen entsprechende Verantwortungslücken hinnehmen oder auf die Potenziale entsprechender Technologien verzichten.

tional Responsibility and Global Justice, 2007, S. 83.), die beispielsweise Nationen haben, wenn es um die Bekämpfung der globalen Armut geht. Entscheidend an dieser Form der prospektiven Verantwortung ist, dass diese von einer retrospektiven Ergebnisverantwortung entkoppelt sein kann. Es ist gerade eine Kennzeichen von globalen Herausforderungen, dass die Akteure, die etwas zur Armut beigetragen haben, nicht diejenigen sind, die zukünftig für die Beseitigung globaler Armut zuständig sind. Vielmehr scheinen Faktoren wie Möglichkeiten der Akteure – im Falle von Nationen z. B. ihr wirtschaftlicher und technischer Entwicklungsgrad – für ihre Beseitigungsverantwortung ausschlaggebend.

⁴⁸ Goetze (Fn. 44), S. 395.

Demgegenüber wurde hier argumentiert, dass es tatsächlich mit einem geläufigen Begriff von Verantwortung schwer bis kaum möglich ist, Verantwortungslücken zu schließen. Zugleich haben wir die Idee von Verantwortungsübernahmen ins Spiel gebracht, die ein natürliches Element der Rollenmoral in verschiedenen Professionen darstellt. In manchen Konstellationen, so das Argument, sollten Ärzte oder Behandler die Verantwortung für robotische Kunstfehler übernehmen. Das bedeutet nicht, dass sie in einem retrospektiven Sinne Schuld an den Fehlern des Systems tragen. Es kann aber bedeuten, dass ihnen dadurch gerechtfertigte Erwartungen entstehen, sich in bestimmter Weise zuständig für die Folgen zu erklären, ansprechbar zu sein oder auch Vorkehrung gegen zukünftige Fehler zu treffen.

Diese Form der Verantwortungsübernahme macht Ärzte und Ärztinnen nicht zu Sündenböcken für Maschinen, sondern stellt einen Ausdruck ihrer professionalen Rolle und eines ärztlichen Ethos dar, welcher das Wohlergehen der Patienten zum obersten Anliegen macht. Zugleich lassen sich auf diese Weise zumindest einige der befürchteten Verantwortungslücken schließen und damit zukünftige (teil)autonome Systeme widerstandslos in die moralische Praxis integrieren.

Literatur

- Alemzadeh, Homa/Raman, Jaishankar/Leveson, Nancy/Kalbarczyk, Zbigniew/Iyer, Ravishankar K.*: Adverse Events in Robotic Surgery: A Retrospective Study of 14 Years of FDA Data, in: *PloS one*, Heft 4, 2016, e0151470.
- Association for Computing Machinery* (Hg.): *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, New York 2022.
- Attanasio, Aleks/Scaglioni, Bruno/Momi, Elena de/Fiorini, Paolo/Valdastri, Pietro*: Autonomy in Surgical Robotics, in: *Annual Review of Control, Robotics, and Autonomous Systems*, Heft 1, 2021, S. 651 ff.
- Bleher, Hannah/Braun, Matthias*: Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems, in: *AI and ethics*, Heft 4, 2022, S. 747 ff.
- Cave, Stephen/Dihal, Kanta/Dillon, Sarah* (Hg.): *AI narratives. A history of imaginative thinking about intelligent machines*, Oxford 2020.
- Champagne, Marc/Tonkens, Ryan*: Bridging the Responsibility Gap in Automated Warfare, in: *Philosophy & technology*, Heft 1, 2015, S. 125 ff.
- Char, Danton S./Shah, Nigam H./Magnus, David*: Implementing Machine Learning in Health Care – Addressing Ethical Challenges, in: *The New England journal of medicine*, Heft 11, 2018, S. 981 ff.
- Coeckelbergh, Mark*: Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability, in: *Science and engineering ethics*, Heft 4, 2020, S. 2051 ff.
- Danaher, John*: Tragic Choices and the Virtue of Techno-Responsibility Gaps, in: *Philosophy & technology*, Heft 2, 2022, S. 26.

- Fiorini, Paolo/Goldberg, Ken Y./Liu, Yunhui/Taylor, Russell H.:* Concepts and Trends in Autonomy for Robot-Assisted Surgery, in: *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, Heft 7, 2022, S. 993 ff.
- Goetze, Trystan S.:* Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement, in: *Association for Computing Machinery (Hrsg.): FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA 2022, S. 390 ff.
- Gunkel, David J.:* Mind the gap: responsible robotics and the problem of responsibility, in: *Ethics and Information Technology*, Heft 4, 2020, S. 307 ff.
- Heidbrink, Ludger:* Definitionen und Voraussetzungen der Verantwortung, in: *Heidbrink, Ludger/Langbehn, Claus/Loh, Janina (Hrsg.): Handbuch Verantwortung*, Wiesbaden 2017, S. 3 ff.
- Heidbrink, Ludger/Langbehn, Claus/Loh, Janina (Hg.):* Handbuch Verantwortung, Wiesbaden 2017.
- Himmelreich, Johannes:* Responsibility for Killer Robots, in: *Ethical Theory and Moral Practice*, Heft 3, 2019, S. 731 ff.
- Jamjoom, Aimun A.B./Jamjoom, Ammer M.A./Thomas, Jeffrey P./Palmisciano, Paolo/Kerr, Karen/Collins, Justin W./Vayena, Effy/Stoyanov, Danail/Marcus, Hani J.:* Autonomous surgical robotic systems and the liability dilemma, in: *Frontiers in Surgery*, 2022.
- Jong, Roos de:* The Retribution-Gap and Responsibility-Loci Related to Robots and Automated Technologies: A Reply to Nyholm, in: *Science and engineering ethics*, Heft 2, 2020, S. 727 ff.
- Kalan, Satyam/Chauhan, Sanket/Coelho, Rafael F./Orvieto, Marcelo A./Camacho, Ignacio R./Palmer, Kenneth J./Patel, Vipul R.:* History of robotic surgery, in: *Journal of Robotic Surgery*, Heft 3, 2010, S. 141 ff.
- Kiener, Maximilian:* Can we Bridge AI's responsibility gap at Will?, in: *Ethical Theory and Moral Practice*, Heft 4, 2022, S. 575 ff.
- Klement, Jan Henrik:* Rechtliche Verantwortung, in: *Heidbrink, Ludger/Langbehn, Claus/Loh, Janina (Hrsg.): Handbuch Verantwortung*, Wiesbaden 2017, S. 559 ff.
- Königs, Peter:* Artificial intelligence and responsibility gaps: what is the problem?, in: *Ethics and Information Technology*, Heft 3, 2022.
- List, Christian:* Group Agency and Artificial Intelligence, in: *Philosophy & technology*, Heft 4, 2021, S. 1213 ff.
- Loh, Janina:* Strukturen und Relata der Verantwortung, in: *Heidbrink, Ludger/Langbehn, Claus/Loh, Janina (Hrsg.): Handbuch Verantwortung*, Wiesbaden 2017, S. 35 ff.
- March-Russell, Paul:* Machines Like Us? Modernism and the Question of the Robot, in: *Cave, Stephen/Dihal, Kanta/Dillon, Sarah (Hrsg.): AI narratives. A history of imaginative thinking about intelligent machines*, Oxford 2020, S. 165 ff.
- Matthias, Andreas:* The responsibility gap: Ascribing responsibility for the actions of learning automata, in: *Ethics and Information Technology*, Heft 3, 2004, S. 175 ff.
- Miller, David:* *National Responsibility and Global Justice*, Oxford 2007.

- Montreal Deklaration*: Erklärung von Montreal für eine verantwortungsvolle Entwicklung von künstlicher Intelligenz 2023, URL: declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/AL-UdeM_Declaration-IA-Resp_allemand_vf.pdf, zuletzt aufgerufen am: 30.06. 2023.
- Moor, J. H.*: The Nature, Importance, and Difficulty of Machine Ethics, in: IEEE Intelligent Systems, Heft 4, 2006, S. 18 ff.
- Nyholm, Sven*: Humans and robots. Ethics, agency, and anthropomorphism, London/New York 2020.
- On-Road Automated Driving (ORAD) committee*: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles 2021, 22736:2021.
- Saeidi, H./Opfermann, J. D./Kam, M./Wei, S./Leonard, S./Hsieh, M. H./Kang, J. U./Krieger, A.*: Autonomous robotic laparoscopic surgery for intestinal anastomosis, in: Science robotics, Heft 62, 2022, eabj2908.
- Santoni de Sio, Filippo/Mecacci, Giulio*: Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them, in: Philosophy & technology, Heft 4, 2021, S. 1057 ff.
- Smith, Helen*: Clinical AI: opacity, accountability, responsibility and liability, in: AI & SOCIETY, Heft 2, 2021, S. 535 ff.
- Sparrow, Robert*: Killer Robots, in: Journal of Applied Philosophy, Heft 1, 2007, S. 62 ff.
- Tigard, Daniel W.*: There Is No Techno-Responsibility Gap, in: Philosophy & technology, Heft 3, 2021, S. 589 ff.
- U.S. Food & Drug Administration*: Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices 2023, URL: [fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices](https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices), zuletzt geprüft am: 30.06.2023.
- van de Poel, Ibo/Sand, Martin*: Varieties of responsibility: two problems of responsible innovation, in: Synthese, S19, 2021, S. 4769 ff.