

Kontroverse

Brauchen wir eine Roboterethik? Eine Kontroverse mit Susanne Beck, Bruno Gransche, Eric Hilgendorf, Janina Loh, Catrin Misselhorn und Thomas Zoglauer

Sechs Forscherinnen und Forschern, die sich in philosophischer bzw. juristischer Hinsicht mit aktuellen und antizipierbaren Problemen der Robotik beschäftigen, haben wir drei Fragen zum möglichen Sinn und Nutzen einer Roboterethik vorgelegt. Was ›Roboterethik‹ heißen kann, ließen wir bewusst offen; eine Auskunft darüber ist Teil der Antworten, um die wir alle eingeladenen AutorInnen zunächst unabhängig voneinander baten. Diese sind im Folgenden abgedruckt. Danach hatten alle Beteiligten Gelegenheit, auf die Positionen und Argumente der MitdiskutantInnen einzugehen, um streitbare Punkte oder Probleme zu identifizieren und Akzente zu setzen, die ihnen in der Debatte besonders wichtig erscheinen. Absicht dieses Rundgesprächs ist es, sowohl zentrale Positionen und Argumentationslinien der aktuellen Debatte als auch kontroverse Punkte oder offene Fragen sichtbar zu machen.

I.
Die Robotik ist eine vielversprechende Wachstumsbranche.¹ Bringen hochentwickelte (›intelligenten‹) Roboter – neben technischen, wirtschaftlichen und rechtlichen Herausforderungen – neuartige metaphysische, moralische oder ethische Probleme mit sich? Wenn ja: Inwiefern sind diese Probleme spezifisch im Vergleich zu anderen Herausforderungen aktueller Informationstechnik?

Catrin Misselhorn:

In Philosophie und Kunst hat es immer schon gedankliche Vorgriffe auf die Zukunft gegeben, beispielsweise die Imagination der Konstruktion künstlicher androider We-

1 Eine Pressemitteilung des Internationalen Verbands für Robotik IFR informierte am 20.6.2018: »Global sales of industrial robots reached the new record of 387,000 units in 2017. That is an increase of 31 percent compared to the previous year (2016: 294,300 units). China saw the largest growth in demand for industrial robots, up 58 percent. Sales in the USA increased by 6 percent – in Germany by 8 percent compared to the previous year« (<https://ifr.org/ifr-press-releases/news/industrial-robot-sales-increase-worldwide-by-29-percent>, abgerufen am 24.8.2018). Der deutsche VDMA Fachverband Robotik + Automation rechnet entsprechend für 2018 »mit einem weiteren Wachstum von 10% auf 15,6 Milliarden Euro.« (<https://www.vdma.org/v2view/er/-/v2article/render/15372372>, abgerufen am 24.8.2018).

sen wie den Golem oder Frankenstein. Vieles, was jetzt diskutiert wird, z.B. autonomes Fahren, hat die Entwicklung dieser Technologien als Vision von Anfang an begleitet. Isaac Asimov hat die moralischen und metaphysischen Fragen, die intelligente Roboter aufwerfen, in seinen Science-Fiction-Texten bereits vor der Mitte des 20. Jahrhunderts hellsichtig thematisiert. Ich würde daher lieber von spezifischen Herausforderungen sprechen, die die Entwicklung der KI und Robotik mit sich bringt. Umgekehrt hat die wissenschaftliche und technische Entwicklung aber auch den philosophischen Diskurs inspiriert, man denke an das Computermodell des Geistes, das für viele ein attraktiver Ausweg aus der festgefahrenen Debatte um Dualismus oder Materialismus in der Philosophie des Geistes war. In ethischer Hinsicht bringen beispielsweise die Asimov'schen Gesetze² eine neue normative Ethik spezifisch für Roboter ins Spiel, deren Verpflichtungen sich von denjenigen unterscheiden, die Menschen gegenüber anderen Menschen haben. Im Unterschied zu den anderen Herausforderungen der Informationstechnik zentrieren sich diese spezifischen Fragen häufig darum, ob Maschinen etwas können oder haben, das bislang als Spezifikum des Menschen galt, z.B. ob sie Bewusstsein oder Emotionen haben, ob sie denken oder moralisch entscheiden und handeln können. Dabei trägt die Hinwendung zu Maschinen nicht selten zur Präzisierung der Begriffe und Positionen bei. Aber auch auf Fragen, die nicht erst im Kontext der neuen technologischen Entwicklungen entstehen, wird ein neues Licht geworfen, beispielsweise Privatheit und informationelle Selbstbestimmung.

Susanne Beck:

Die Entwicklungen in der Robotik und KI bergen zweifellos neue Herausforderungen. Die Maschinen werden große Mengen an Daten sammeln und verarbeiten – ein Problem für das Datenschutzrecht. Im Arbeitsrecht wird man beispielsweise diskutieren müssen, unter welchen Bedingungen ArbeitnehmerInnen mit Robotern zusammenarbeiten müssen. Im Urheberrecht wird sich die Frage stellen, wem Werke zuzuordnen sind, die großteils autonom von einer Maschine »erdacht« wurden. All diese Fragen sind zweifellos wichtig und in naher Zukunft zu diskutieren und zu mindest vorläufig zu beantworten.

Zentral aber ist m.E. die Verantwortungsdiffusion, die entsteht, wenn man immer mehr Entscheidungen auf Maschinen überträgt oder von ihnen vorbereiten lässt: Bei Maschinen, die selbst dazu lernen können, einen gewissen eigenen Entscheidungs-

2 Erstmals beschrieben in der 1942 erschienenen Kurzgeschichte »Runaround«, in: Isaac Asimov: *I, Robot*, New York 1950, S. 40: »1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.«.

rahmen haben, durch Sensoren und Vernetzung Informationen erhalten und selbst auswerten, lässt sich weder im Vorhinein vorhersehen, welche Entscheidungen sie in welchen Situationen treffen werden, noch im Nachhinein feststellen, worauf genau eine Fehlentscheidung beruhte. Insbesondere ob einer der Beteiligten, z.B. ProgrammierInnen, ProduzentIn oder NutzerIn, selbst einen Fehler gemacht haben, ist in vielen Fällen nicht mehr nachweisbar. Die klassischen Zurechnungsstrukturen des Rechts sind hierauf nicht mehr ohne Weiteres anwendbar. Das bedeutet nicht, dass keine Konstellationen denkbar sind, in denen einer der Beteiligten nachweisbar einen Fehler gemacht hat – aber es wird die Ausnahme sein. Deshalb ist das Recht in besonderer Weise gefordert, die bestehenden Verantwortungskonzepte anzupassen. Zweifellos war und ist dies bereits mit Blick auf die verstärkte Kollektivierung unserer Gesellschaft erforderlich – was sich etwa in der Konstruktion der ›juristischen Person‹ niederschlägt. Doch handelt es sich hierbei immer noch um menschliche Gegenüber, wenn auch als Kollektiv. Ein maschinelles ›Gegenüber‹ unterscheidet sich hiervon erheblich, so dass die Übertragbarkeit von rechtlicher Verantwortung neu zu diskutieren ist. Vorab muss im Übrigen diskutiert werden, ob und inwieweit spezifische Entscheidungen überhaupt auf Maschinen übertragen werden können und sollten.

Bruno Gransche:

Roboter, zunächst gefasst als jegliches System mit Sensoren, Prozessoren und Aktuatoren, nehmen eine Sonderstellung bezüglich technik-ethischer Fragestellungen ein. Sie kombinieren Herausforderungen der Informationstechnik, verschärft durch die fortschreitende Digitalisierung, mit solchen der Hardware, des *Embodiments* und der physischen Aktions- und Interaktionsfähigkeit. Besonders die physische Verfasstheit robotischer Systeme mit ihrer Aktuatorik lässt Roboter und Menschen in gemeinsam geteilten Handlungssituationen aufeinandertreffen.

Dabei verändert die ›Autonomie‹ (Grade der Automatisierung) der Roboter (durch die Anbindung von Sensoren und Prozessoren) die Mensch-Technik-Relation spezifisch. An Roboter kann operative und strategische Verantwortlichkeit abgegeben werden, inwieweit und für welche Systeme gar ›moralische‹ Verantwortung delegiert werden könnte, ist fraglich. Mit ihren Aktuatoren können Roboter tatsächliche Unterschiede in der Welt bewirken und nicht nur menschliche Handlungen informieren oder orientieren; ein Haus etwa Zugänge beschränken, ein Auto gezielte Schadens- und Opferselektionen vornehmen, ein Assistenzroboter Handlungsoptionen gewähren oder verwehren. Die Analyseebenen technischen Handelns mit Zweck-Mittel-Erwägungen, Pilot- oder Steuerungsverantwortung etc. werden durch die eigene Aktionsinstanz und hohe Automatisierung der Roboter unterlaufen. Dem

Menschen kommen entweder die Rollen eines medial steuernden *Dirigenten* (der Spielräume und Grundregeln gestaltet und technische Aktionspotenziale regelt) oder eines »auf Augenhöhe« mitagierenden *Koakteurs* zu. Für so koagierende Menschen werden Roboter somit zu zwar menschengemachten, aber im Begegnungsmoment nicht verfügbaren (nicht frei instrumentalisierbaren oder änderbaren) Entitäten. In Robotern begegnen (koagierende) Menschen Systemen, die von Normen, Standards und Moralen von (dirigierenden) Menschen geprägt, dann aber zu »autonomer« Weltbeeinflussung (in ihrer Aktuator-Reichweite) fähig sind; diese robotisch beeinflusste Welt stellt dann den veränderten Rahmen menschlicher Handlungsfreiheit dar. Dies zeigt eigene ethische Herausforderungen der Robotik.

Davon abzugrenzen sind metaphysische Fragen im Sinne starker KI, die Robotern gar Bewusstseine, personale Identität, Handlungsfähigkeit und eigentliche Autonomie (nicht nur technisches Prozessieren und hohe Automatisierung) zuschreiben. Zu letzterer Sonderstellung im metaphysischen Sinne bin ich sehr skeptisch, ob dies überhaupt vom Technikstand jenseits von Gedankenexperimenten jemals wird eingeholt werden können (ich denke nicht); sicherlich jedoch ist dieser Stand heute nicht gegeben, weshalb Roboter m.E. aus metaphysischer Sicht keine spezifischen Probleme aufwerfen. Kurz: »neuartige moralische oder ethische Probleme« – ja; »metaphysische« – nein.

Janina Loh:

Roboter unterscheiden sich von anderen Technologien dadurch, dass wir es hier nicht nur mit rein passiven Objekten zu tun haben, sondern mit quasi autonomen, teil- oder gar vollautonomen Systemen, die ohne unmittelbare Einwirkung ihrer Designer*innen und Nutzer*innen agieren können. Ein Hammer bspw. benötigt im Gegensatz zu einem Roboter immer jemanden, die* ihn führt, das medizinische Assistenzsystem *Da Vinci* würde sich ohne eine Chirurgin*, die* die Steuerkonsole bedient, nicht rühren und auch ein Maschinengewehr im klassischen Sinne schießt nicht von allein. Anders hingegen liegt der Fall bei Robotern, bei denen es sich Catrin Misselhorn zufolge um artifizielle Systeme handelt, die aus einem Prozessor, Sensoren und Aktoren bzw. Effektoren bestehen und zumindest so wirken, als wären sie (teil-)autonom und könnten im Gegensatz zu einem »bloßen« Computer in ihre Umgebung hineinwirken.³

In der Tat gibt es einen nicht zu unterschätzenden Graubereich, der sich an dieser Stelle eröffnet, geht es doch ganz häufig lediglich um Algorithmen, die ohne Senso-

3 Vgl. Catrin Misselhorn: »Robots as Moral Agents?«, in: Frank Rövekamp und Friederike Bosse (Hg.): *Ethics in Science and Society. German and Japanese Views*, München 2013, S. 42–56, hier S. 43.

ren und Effektoren auskommen – beispielsweise im Fall sogenannter Chatbots. Einzuwenden wäre aus meiner Sicht, dass Algorithmen natürlich Teil eines Roboters sind, ihr ›Nervensystem‹ vielleicht (ohne diese Metapher allzu sehr ausreizen zu wollen). Aber die algorithmische Struktur macht eben nicht allein den definitorischen Kern eines Roboters aus, noch finden wir Algorithmen ausschließlich in Robotern.

Die philosophischen Fragen, die sich mit Blick auf Roboter stellen, betreffen also in einem ersten Schritt ihre Autonomie bzw. Handlungsfähigkeit im Allgemeinen, die sie ja gerade von anderen technischen Phänomenen unterscheidet. In einem zweiten Schritt stellt sich dann, wenn es um moralische (und nicht etwa rechtliche oder politische) Handlungsfähigkeit geht, die Frage, was Moral ist und ob Roboter moralisch handeln können. Abhängig von der hierauf gegebenen Antwort bleibt darüber nachzudenken, wie sich weitere Kompetenzen, die für moralisches Handeln von Relevanz sind, in Robotern simulieren bzw. implementieren ließen wie etwa Intelligenz, Vernunft, Urteilskraft oder vielleicht sogar Emotionen und Empathie.

Eric Hilgendorf:

Es dürfte kaum eine technologische Entwicklung geben, die kulturgeschichtlich so intensiv vorbereitet wurde wie die Robotik. Die Vorstellung künstlicher Menschen lässt sich bis in die Antike zurückverfolgen. Die besondere Faszination, die von künstlichen intelligenten Wesen ausgeht, übersteigt alles, was im Rahmen des technischen Fortschritts bislang bekannt war.

Ein Grund hierfür dürfte sein, dass wir dazu neigen, intelligente autonome Systeme zu vermenschlichen, ihnen also menschenähnliche Züge beizumessen. Unser Hang zum Anthropomorphismus führt dazu, dass wir Maschinen, die auf eng beschränkten Feldern eine hohe ›Intelligenz‹ zeigen, leicht Motive zuordnen, die wir bei Menschen mit vergleichbaren Fähigkeiten ebenfalls vermuten würden. Auf diese Weise lässt sich erklären, weshalb so viele Beobachterinnen und Beobachter der Meinung zu sein scheinen, Roboter könnten den Wunsch entwickeln, die Menschheit als dominante Spezies zurückzudrängen und letztendlich zu vernichten.

Eine zweite Besonderheit, die mit der Entwicklung intelligenter autonomer Systeme einhergeht, ist die Notwendigkeit, Abläufe, die bislang im menschlichen Gehirn mehr oder weniger unbewusst abliefen, explizit zu machen, um sie maschinell darstellen zu können. Dieser Explikationszwang dürfte zu einer immer weitergehenden Entzauberung des menschlichen Geistes führen. Schon jetzt ist allerdings deutlich geworden, dass die Vorstellung, die neuen autonomen Systeme würden den menschlichen Geist kopieren, in die Irre führt. Das evolutionär entstandene menschliche Denken und Lernen erscheint vielfach sehr komplex und langwierig – warum sollte

es z.B. viele Jahre dauern, bis eine Maschine eine Sprache erlernt hat? Es erscheint deshalb vorzugswürdig, von ›maschineller‹ statt von ›künstlicher‹ Intelligenz zu sprechen.

Thomas Zoglauer:

Die Roboter der neusten Generation sind autonom agierende kognitive Systeme, die lernfähig sind und selbständig Entscheidungen treffen können. In der Roboterethik wird häufig die Auffassung vertreten, dass wir es bei solchen Robotern mit autonomen, verantwortungsfähigen, ja sogar moralischen Wesen, sogenannten ›moralischen Agenten‹, zu tun hätten. Wenn dies stimmt, so würde dies tatsächlich eine ganze Reihe neuartiger metaphysischer, moralischer und ethischer Probleme mit sich bringen, etwa die Frage nach dem Personenstatus und dem rechtlichen Status von Robotern.

Diese Auffassung beruht meiner Meinung nach auf einem Irrtum. So wird beispielsweise behauptet: Je autonomer Roboter agieren und je weniger ihr Verhalten voraussehbar und kontrollierbar sei, desto weniger könne man HerstellerInnen, KonstrukteurInnen und ProgrammiererInnen verantwortlich machen, wenn Roboter einen Schaden verursachen. Allerdings kann man von einer technischen Autonomie im Sinne einer Unabhängigkeit vom Menschen nicht auf eine moralische Autonomie und damit auf eine Selbstverantwortung der Roboter schließen. Denn Verantwortungsfähigkeit setzt ein Verantwortungs- und Moralbewusstsein voraus. Dies ist bei gegenwärtig existierenden Robotern nicht vorhanden und wird auch in absehbarer Zukunft nicht geschaffen werden können. Roboter handeln nicht, sondern folgen nur einem Programm. Handeln setzt im Unterschied zum bloßen Verhalten ein Reflexionsvermögen und intentionale Bewusstseinszustände voraus. Man kann Robotern – außer im metaphorischen Sinn – keine Absichten unterstellen. Roboter haben keine Wünsche – weder erster noch höherer Ordnung – und können nicht unter verschiedenen Handlungsalternativen wählen. Sie werten nicht und haben kein Gefühl für richtig oder falsch. Wenn durch einen Robotereinsatz ein Schaden entsteht, ist immer noch der Mensch, sei es als HerstellerIn, KonstrukteurIn, ProgrammiererIn oder NutzerIn, verantwortlich. Eine ›Verantwortungslücke‹ gibt es nicht.

Aufgabe für die Ethik und die Gesetzgebung wird es daher sein, den Umgang mit Robotern zu regeln. Dafür brauchen wir keine neue Ethik oder Metaphysik. Metaphysische Fragen nach einem möglichen Bewusstsein oder dem Personenstatus von Robotern mögen in Zukunft vielleicht relevant werden, stellen sich aber derzeit nicht.

II.

Brauchen wir, heute oder morgen, eine dezidierte Roboterethik? Reichen die Mittel des Rechts und des Designs sowie bekannte Ethiken für die Regelung der sozialen Folgen der Robotik nicht aus? Was soll oder kann ›Ethik‹ in dem Zusammenhang überhaupt bedeuten – im Unterschied etwa zu (normativen/juridischen) ›Robotergergesetzten‹ oder einer (algorithmisch kodifizierten) ›Robotermoral‹?

Eric Hilgendorf:

Der Begriff ›Roboterethik‹ ist notorisch unbestimmt. Die Verwirrungen beginnen schon damit, dass meist nicht klar zwischen Moral (also bestimmten Verhaltensregeln) und Ethik (der wissenschaftlichen Beschäftigung mit diesen Verhaltensregeln) unterschieden wird. Als ›Recht‹ bezeichnet man diejenigen Verhaltensregeln, deren Verletzung mit staatlich organisiertem Zwang, also zum Beispiel mit Strafe, sanktioniert wird. Moral und Ethik verhalten sich zueinander wie Recht und Rechtswissenschaft.

›Roboterethik‹ kann deshalb zunächst die wissenschaftliche Beschäftigung damit meinen, wie wir Roboter einsetzen sollen oder dürfen. Derartige Fragen gehören in den Umkreis der Technikethik, wobei die empirische Technikfolgenabschätzung eine besondere Rolle spielt oder zumindest spielen sollte. Zu einer so verstandenen Roboterethik im weiteren Sinne gehört auch die Frage, ob bzw. unter welchen Umständen Roboter als eigene moralische Subjekte in Betracht kommen könnten. Dabei stellen sich Fragen, wie sie auch aus der Tierethik bekannt sind. Anknüpfungspunkte für einen eigenständigen moralischen Status von Maschinen könnten etwa Bewusstsein oder Schmerzempfindlichkeit sein.

Nicht selten wird der Begriff ›Roboterethik‹ aber in einem anderen Sinne gebraucht und meint dann moralische Regeln, die den Maschinen in Form von Algorithmen einprogrammiert werden. Statt ›Roboterethik‹ sollte man hier besser von ›Robotermoral‹ oder, wie Asimov, von ›Robotergergesetzen‹ sprechen. Bis vor einigen Jahren hielt man die Beschäftigung mit derartigen Regeln für pure Science-Fiction. Im Zusammenhang mit der Entwicklung selbstlernender Systeme ist aber deutlich geworden, dass es zweckmäßig sein könnte, bei derartigen Systemen das Spektrum der Lernmöglichkeiten von vornherein zu begrenzen, also sozusagen über ein ›Gewissen‹ der Maschine sicherzustellen, dass bestimmte Verhaltensweisen gar nicht erst gelernt werden können. So sollte etwa ein Sprachavatar wie Tay fehlerfreie und sozial angemessene Kommunikation erlernen, nicht aber, andere Menschen zu beschimpfen und zu beleidigen.

Roboterethik ist die philosophische Disziplin, die sich mit den moralischen Werten befasst, die im Bau von Robotern, im Umgang mit Robotern und im Verhältnis von Menschen und Robotern eine Rolle spielen. Es handelt sich dabei um eine verhältnismäßig junge Bereichsethik, eine Teilbereichsethik der Maschinenethik, um ganz genau zu sein. Denn alle Roboter sind Maschinen, aber nicht umgekehrt alle Maschinen auch Roboter. Unter den Bereichsethiken lassen sich ganz generell zwei Typen ausmachen: Bei den einen handelt es sich um Ethiken für den Umgang mit einem nichtmenschlichen Gegenüber. Hierzu zählen neben der Maschinen- und Roboterethik auch die Tier-, Pflanzen-, Umwelt- und Technikethik. Die andere Gruppe an Bereichsethiken versammelt Ethiken für Sonderbereiche des menschlichen Lebens, in denen Werte vertreten, Normen geltend gemacht und Regeln formuliert werden, denen im Alltag für gewöhnlich ein anderer Status zugeschrieben wird. Die Medizinethik, die Ethik humanitärer Interventionen, die Militär- und die Wirtschaftsethik sowie die Ethik internationaler Beziehungen lassen sich als Beispiele für diese Form Bereichsethiken anführen.

In der Roboterethik wird – vergleichbar den anderen Bereichsethiken, die sich mit nichtmenschlichen Wesen beschäftigen – darüber nachgedacht, inwiefern das fragliche Gegenüber ein Wert- und vielleicht gar ein Rechtsträger ist (Roboter als sogenannte *moral patients*) und diskutiert, inwiefern es als moralischer Akteur interpretiert zu werden hat (Roboter als sogenannte *moral agents*). Sie stellt traditionelle Fragen mit Blick auf neue potenzielle Handlungssubjekte, nämlich Roboter, wie beispielsweise die danach, welche Kompetenzen wir generell als grundlegend für moralische Akteursfähigkeit erachten, welche moralischen Werte wir artifiziellen Systemen implementieren sollten, auf was für ein moralisches Selbstverständnis es schließen lässt, wenn wir Roboter ›schlecht‹ behandeln und in welchen Bereichen – Industrie-, Militär-, Medizin-, Altenpflege-, Servicerobotik – wir uns auch zukünftig ausschließlich bzw. in einem signifikanten Ausmaß auf menschliche und nicht auf artifizielle Expertise verlassen wollen.

Robotergesetze und eine etwaige Robotermoral (wie in der Frage oben formuliert) fallen somit in den ersten Bereich der Roboterethik, in dem es um Roboter als Wert- und Rechtsträger geht, denen moralische Werte durch menschliche Designer*innen implementiert werden. Innerhalb des zweiten Bereichs der Roboterethik – Roboter als *moral agents* – wird hingegen danach gefragt, inwiefern Roboter die Fähigkeiten, die sie zum moralischen Handeln befähigen, ausbilden bzw. simulieren können.

Susanne Beck:

Dass man die ethischen Überlegungen mit Blick auf die Robotik und KI zwangsläufig als ›Roboterethik‹ kategorisieren muss, halte ich für unnötig, ebenso wenig brauchen wir ein eigenständiges ›Roboterrecht‹. Zugleich könnte eine derartige Kategorisierung möglicherweise dazu beitragen, die besonderen Probleme, die durch die Entwicklung in Robotik und KI entstehen, zu betonen und herauszustreichen.

Dass die bisherigen ethischen und auch rechtlichen Regeln und Konzeptionen überhaupt nicht mehr anwendbar seien, halte ich ebenfalls nicht für zutreffend. Es ist aber sicherlich so, dass die mit der Robotik und KI zusammenhängenden Probleme einer besonderen und intensiven Debatte bedürfen und bestimmte Konzepte, etwa die auf das Individuum ausgerichtete Verantwortung, einer Neuausrichtung bedürfen (vgl. die Antwort zur Frage I).

Eine entsprechende ethische Betrachtung kann m.E. hierzu Verschiedenes beitragen:

Es ließen sich Regeln entwickeln, wie jedes Individuum im Kontext von Robotik und KI verantwortlich mit den Maschinen umgehen sollte bzw. wie im Nachhinein entsprechende moralische Verantwortung zuzuschreiben wäre.

Es ließen sich Regeln entwickeln, in welchen Bereichen Entscheidungen entweder überhaupt nicht oder nur sehr begrenzt auf Maschinen übertragen werden sollten. Ausgesprochen zweifelhaft könnte das etwa für die Entscheidung über Leben und Tod in kriegerischen Kontexten sein. Auch dass etwa bei der direkten Betreuung von Kindern oder Patienten, die nicht umfassend einwilligungsfähig sind, nicht primär Maschinen eingesetzt werden sollten, erscheint einleuchtend. Diese Intuitionen wären an traditionellen bzw. aktuellen moralischen Regeln zu überprüfen.

Weiterhin könnten ethische Überlegungen dazu beitragen, die Diskussion über die Veränderung der Gesellschaft durch eine verstärkte Maschinisierung bzw. Autonomisierung zu strukturieren.

Überdies ist m.E. durchaus sinnvoll, Konzepte wie ›Person‹, ›Verantwortung‹ oder ›Intelligenz‹ mit Blick auf die Entwicklungen neu zu diskutieren. Dabei gehe ich keineswegs davon aus, dass die Maschinen in einem Maße menschenähnlich sein werden, dass wir unsere bestehenden Konzepte auf sie übertragen können oder gar sollten. Sie werden aber doch in einer Weise mit uns interagieren und den Alltag bestimmen, dass andere oder eben angepasste Konzepte für das Zusammenleben von Mensch und Maschine erforderlich werden.

Thomas Zoglauer:

Mit dem Begriff ›Roboterethik‹ verbindet man intuitiv eine Ethik für Roboter und denkt dabei beispielsweise an die drei Robotergesetze von Isaac Asimov. Allerdings ist es fraglich, ob sich Moral in Form eines Computerprogramms formulieren lässt oder ob sich moralische Dilemmata algorithmisch lösen lassen. Ethisches Problemlösen besteht in einem sensiblen Abwägen von Normen und Werten, wobei neben Vernunftgründen oftmals auch Empathie und Gerechtigkeitsintuitionen eine wichtige Rolle spielen. Maschinen besitzen keine Gefühle und kein Gewissen. Im Gegensatz zu einer deduktiven *Top-down*-Programmierung mit Hilfe von ›Robotergesetzen‹ favorisiert man in der aktuellen KI-Forschung derzeit einen induktiven *Bottom-up*-Ansatz: Durch *Deep Learning*-Algorithmen werden Roboter trainiert und lernen selbstständig Probleme zu lösen. Aber auch beim *Deep Learning* folgen Roboter einfach nur einem Algorithmus und machen sich keine Gedanken darüber, ob ihr Handeln richtig oder falsch ist. Entweder entscheiden Roboter aufgrund von Regeln – egal, ob sie von außen vorgegeben oder selbst erlernt werden – oder durch Zufall. Werden die Entscheidungsregeln von außen vorgegeben, kann man nicht von einer autonomen ›moralischen‹ Maschine sprechen, weil die Maschine erstens nicht unabhängig vom Menschen entscheidet und zweitens nur mechanisch ein Programm abarbeitet und keine eigene Moral hat. Aber auch dann, wenn ein lernender Roboter selbst Regeln entwickelt, macht dies den Roboter nicht zu einem moralischen Wesen, da ihm ein Reflexionsvermögen fehlt. Wir beurteilen Handlungen als moralisch oder unmoralisch, indem wir Gründe anführen und wir sind frei, eine Handlung zu tun oder zu unterlassen. Roboter denken aber nicht nach, bevor sie etwas tun, und sind noch weit davon entfernt, einen ›moralischen Turingtest‹ zu bestehen. Man kann keine Ethik für Maschinen entwickeln, die kein Verständnis von Moral haben. Vielmehr ist die Roboterethik als eine Ethik für Menschen und die Gestaltung von und den Umgang mit Robotern gedacht. Es geht dabei nicht nur darum, wie man Roboter programmieren soll, sondern auch um Designfragen, Sicherheitsbelange, Privatheit und Datenschutz. Die Roboterethik ist eine sogenannte ›Bindestrich-Ethik‹, ähnlich wie die Tierethik, Genethik, Medizinethik, Wirtschaftsethik oder ökologische Ethik, die einen speziellen Objektbereich in den Blick nimmt. Aufgabe der Roboterethik wird es daher sein, die Folgen eines zunehmenden Einsatzes von Robotern abzuschätzen, zu bewerten und Handlungsempfehlungen zu formulieren.

Catrin Misselhorn:

Vielfach wird die Roboterethik als Bereich der angewandten Ethik bestimmt. Ein naives Verständnis ginge davon aus, dass angewandte Ethik einfach in der Anwen-

dung vorgefundener Prinzipien der normativen Ethik besteht, denen sie nichts Neues hinzufügt. Demgegenüber ist festzuhalten: Natürlich spielen die Theorien der normativen Ethik eine Rolle, aber es ist in der Debatte eigentlich unstrittig, dass die angewandte Ethik zwar Beziehungen zu den Theorien der normativen Ethik unterhält, aber als eigenständiger Bereich mit eigenen Subdisziplinen, Fragen und Methoden zu gelten hat. Dies kann auch ein Vorteil sein, insofern dadurch ein Dissens auf der abstrakten Ebene der Prinzipien durch die Fokussierung auf den Anwendungsbereich umgangen wird. Insofern ginge die Roboterethik als angewandte Ethik über die normative Ethik hinaus.

Die zweite Option wäre, dass die angewandte Ethik zwar ein eigenständiger Bereich ist, die Roboterethik aber mit irgendeiner anderen Disziplin der normativen Ethik zusammenfällt. Dann wäre die Gegenfrage, mit welcher. Ich sehe hier eher eine Hierarchie oder besser ein Netzwerk mit verschiedenen Ebenen. So steht die Roboterethik in einer engen Beziehung zur Computerethik, die ihrerseits ein Teilbereich der Informationsethik ist, aber sie hat auch Beziehungen zu vielen anderen Zweigen der angewandten Ethik, u.a. zur Wirtschaftsethik, zur Berufsethik und – je nach Anwendungsbereich – beispielsweise auch zur Kriegs-, zur Pflege- oder zur Arbeitsethik. Vor allem im Hinblick auf die Computerethik gibt es darüber hinaus eine Debatte über ihre Einzigartigkeit. Diese liegt vor, wenn die ethischen Probleme, die sie aufwirft, von bestimmten, für die Computertechnologie spezifischen Eigenschaften abhängen, und ohne diese nicht oder zumindest nicht in ihrer gegenwärtigen Form entstanden wären. Mit Hilfe der entsprechenden Programmierung können Computer nun beispielsweise einen Grad an Unabhängigkeit erreichen, der reinen Werkzeugen nicht zukommt. Genau daraus ergeben sich viele der spezifischen Fragen der Roboterethik, wie die Forderung, Maschinen müssten auch ethische Entscheidungen treffen können. Ein anderer wichtiger Punkt, der hier hineinspielt, ist das menschenähnliche Aussehen. Es ist insbesondere für unsere emotionale Interaktion mit Robotern von Bedeutung. Auch dieser Punkt erlangt durch die Programmierung aber eine neue Qualität, als wenn es bloß um leblose Schaufelsterpuppen ginge. Diese distinktiven Merkmale sind hinreichend, um eine eigene Disziplin der angewandten Ethik zu begründen.

Das Stichwort Informationsethik bringt eine dritte Möglichkeit ins Spiel. So argumentiert insbesondere Luciano Floridi, diese stelle ein neues Paradigma der allgemeinen normativen Ethik dar.⁴ Der Begriff der Information ist für ihn so grundlegend in der Beschreibung der Welt, dass Information die fundamentalste moralisch relevante Kategorie darstellt. Menschen sind für ihn nur eine Art von ›informationalen Agenten‹ unter anderen sowohl natürlichen als auch künstlichen informationalen Agenten, die sich in der sogenannten Infosphäre bewegen. Die normativen Prinzipi-

4 Vgl. Luciano Floridi: *The Ethics of Information*, Oxford: Oxford University Press 2013.

en, die Floridi formuliert, beziehen sich auf den Erhalt und das Gedeihen der Infosphäre, die beispielsweise vor Entropie zu schützen ist. Der Ansatz sieht sich in Konkurrenz zu anderen Theorien der Allgemeinen Ethik wie dem Utilitarismus, der deontologischen Ethik und der Tugendethik. Ich halte es allerdings eher für eine Tugend der angewandten Ethik, wenn sie mit so wenigen und so unkontroversen metaphysischen Annahmen auskommt wie möglich. Deshalb halte ich Floridis Weg zwar für theoretisch interessant, zur Lösung der ethischen Probleme der Robotik aber für weniger zielführend.

Bruno Gransche:

Ethik bedeutet im Zusammenhang mit Robotern nichts spezifisch Anderes als in anderen Zusammenhängen: Sie ist die Disziplin, in der versucht wird, moralische Dissonanzen konkurrierender Moralien, Regeln, Maximen etc. in und zwischen Individuen und Gruppen aufzuklären. *Roboterethik* heißt dann der Unterbereich, in dem die Fälle, in denen diese Dissonanzen auftauchen, im weitesten Sinne mit Robotern zu tun haben. *Robotergesetze* sind entsprechend rechtlich kodifizierte und exekutiv durchsetzbare Ergebnisse einer solchen Aufklärung. »*Robotermoral*« (algorithmisch kodifiziert) wäre die Implementierung dieser Ergebnisse in robotische Systeme, mit dem Ziel, im Prozessieren der Roboter jene Dissonanzen zu vermeiden bzw. keine Rechtsbrüche zu verursachen.

Da Roboter eigene ethische Herausforderungen bringen und es sich bewährt hat, Bereichen mit speziellen Eigenschaften und Anforderungen auch spezialisierte Bereichsethiken beizustellen, ist eine eigene Roboterethik im Sinne einer Bereichsethik wie etwa Tierethik, Medizinethik etc. unabdingbar. Roboterethik ist insofern ein Bereich der angewandten Ethik, der es mit der Herausforderung der Anwendung ethischer Prinzipien, Erkenntnisse etc. auf Fälle, in denen Roboter involviert sind, zu tun hat. Hier ist neben ethischer Expertise v.a. technische Expertise gefragt. Etwa wenn Zurechnungen von Handlungs- bzw. Aktionsfolgen geklärt werden müssen, muss neben der normativen Prüfung (Folge X soll gezeitigt/vermieden werden) auch eine Analyse des vorliegenden Falls und der Anwendbarkeit der geprüften Sollens-Regel erfolgen, also die Frage etwa, ob ein Roboter in diesem Fall über den maßgeblichen Aspekt der Aktionsursache »autonom« (im Rahmen seiner geregelten Aktionsfreiheiten) verfügen konnte, oder ob eine EntwicklerInnen- oder DirigentInnenentscheidung ursächlich war. Dies geht da hinaus über IT-Ethik, wo Aktuatoreffekte maßgeblich sind, über Ethik instrumentellen Handelns, wo »autonome« Aktionen zu berücksichtigen sind usw.

Dass Roboterethik als thematischer Bereich der angewandten Ethik gebraucht wird, besagt jedoch noch nicht, dass in der Roboterethik neue ethische Paradigmen

in Anschlag zu bringen wären. Tugendethik, Pflichtenethik, Nutzenethik etc. sind keine Bereichsethiken wie Medienethik, Neuroethik oder eben Roboterethik; sie bezeichnen keine Anwendungsbereiche, sondern ethische Paradigmen. Die Ausdifferenzierung einer Roboterethik wäre so gesehen zunächst keine *neue Ethik* (im Sinne eines neuen ethischen Paradigmas), sondern ein neuer spezialisierter (im Wesentlichen interdisziplinärer) Anwendungsbereich der Ethik. Dass Herausforderungen eines neuen Anwendungsbereiches die Notwendigkeit eines neuen Paradigmas ergeben können, ist davon unbenommen und bleibt zu diskutierendes Ergebnis der Forschung von Roboterethikern. Auch hier sehe ich heute zwei kurze Antworten auf die obige Frage »Brauchen wir eine Roboterethik?«: als Bereichsethik – ja, als ethisches Paradigma (was sollte das auch sein?) – nein.

III.

Das Europäische Parlament hat 2017 der EU-Kommission zur Frage zivilrechtlicher Regelungen im Bereich Robotik vorgeschlagen, darüber zu entscheiden, ob wir Roboter einer bestimmten Entwicklungsstufe als »elektronische Personen« behandeln sollten.⁵ Sollten wir ihnen gar eigene Rechte zugestehen? Was würden Sie der Kommission raten?

Susanne Beck:

Bei der Einführung einer »elektronischen Person« geht es zunächst um eine rechtliche Konstruktion, die in vielerlei Hinsicht der bereits bekannten »juristischen Person« ähnelt, einem Rechtsstatus für Vereine, Unternehmen, Behörden etc. Diese ist zunächst ganz funktional an den Interessen der verschiedenen Beteiligten ausgerichtet. So ist es für diejenigen, die mit dem Unternehmen interagieren, möglicherweise sinnvoll, einen direkten Ansprechpartner unabhängig von personellen Wechseln innerhalb des Unternehmens zu haben. Auch kann für ein Fehlverhalten bzw. eine Schädigung die juristische Person direkt haftbar gemacht werden und ggf. auch selbst (durch einen Vertreter) vor Gericht auftreten. Sie hat eigenes Vermögen und

5 Die Abgeordneten des EU-Parlaments forderten die EU-Kommission zudem auf, die Benennung einer Europäischen Agentur für Robotik und Künstliche Intelligenz in Betracht zu ziehen, um das erforderliche technische, ethische und regulatorische Fachwissen für die Behörden zur Verfügung zu stellen. Vgl. Europäisches Parlament: Pressemitteilung *Robotik und künstliche Intelligenz: Abgeordnete für EU-weite Haftungsregelungen*, <http://www.europarl.europa.eu/news/de/press-room/20170210IPR61808/robotik-und-kunstliche-intelligenz-abgeordnete-für-eu-weite-hafungsregelungen> (aufgerufen: 29.6.2018).

Europäisches Parlament: Bericht A8-0005/2017, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//DE> (aufgerufen: 29.6.2018).

kann selbst im Rechtsverkehr aktiv und passiv auftreten. Mit Blick hierauf ist es sogar zwingend erforderlich, dass eine Rechtsperson spezifische Rechte und Pflichten hat.

Dabei geht es nicht darum, der juristischen Person – oder in unserem Fall der elektronischen Person – in einem naturrechtlichen Sinne vorrechtlich existierende Rechte auch im geltenden Recht zuzusprechen. Es geht auch nicht darum, diese rechtlichen Konstrukte mit dem Menschen gleichzusetzen. Sondern es geht darum, funktionale Lösungen für durch die Entwicklung entstehende Probleme zu finden. So könnte die elektronische Person insbesondere als Adressat für Geschädigte dienen, wenn die Maschine im Rahmen einer autonomen Entscheidung Schäden verursacht. Sie könnte ein eigenes Vermögen haben, das durch Einzahlungen der verschiedenen Beteiligten zustande kommt, und das für den Ausgleich der Schäden herangezogen werden könnte. Der Geschädigte müsste dann insbesondere nicht nachweisen, wer von den menschlichen Beteiligten den Schaden verursacht hat – das ist in diesen Fällen sehr häufig unmöglich. Die elektronische Person könnte überdies Vertragspartner sein (z.B. als elektronischer Agent im Börsenhandel) und auch durch einen Vertreter vor Gericht auftreten.

Wie bei der juristischen Person muss sie gerade hierfür eigene Rechte und Pflichten haben; ihr Vermögen muss geschützt werden, ihre rechtlichen Interaktionen abgesichert sein. Insofern ist zu empfehlen, eine Art. 19 Abs. 4 GG entsprechende Regelung einzuführen, wonach derartige Rechtspersonen die ihrem Wesen bzw. ihren Funktionen entsprechenden Rechte innehaben.

Janina Loh:

Erlaubt die juristische Person eine Haftbarmachung von Unternehmen und anderer Personenvereinigungen, würden als elektronische Personen auch artifizielle Wesen handlungsfähig. Das halte ich generell zunächst für einen guten Schritt, solange man dabei nicht aus den Augen verliert, dass immer noch einzelne Menschen bzw. Aktiengesellschaften für etwaig entstandene Schäden haftbar bleiben – wie auch im Fall der juristischen Person nur natürliche Personen für ein Vergehen bestraft werden können.

Roboter als elektronische Personen zu sehen, ist ein begrüßenswerter Schritt deshalb, da sie auf diese Weise zumindest in einen Teilbereich des moralischen Universums Einzug halten – nämlich immerhin in den moralischen Teilbereich des Rechts (sofern man gewillt ist, zwischen Moral und Recht einen Zusammenhang dergestalt zu sehen, dass zwar nicht alle moralischen Werte auch in Rechte übersetzt sind, aber alle Rechte einen moralischen Gehalt haben – eine Position, über die sich philosophisch in der Tat vortrefflich streiten lässt).

Weitere Fragen knüpfen sich allerdings an das Konzept der ›elektronischen Person‹: Sollen alle Roboter oder nur bestimmte als elektronische Personen bezeichnet werden können – etwa der Staubsaugerroboter *Roomba* (entwickelt von *iRobot*) ebenso wie die Lernplattform *iCub* (*RobotCub-Projekt*)? Die Definition der elektronischen Person fußt also implizit auf einem Verständnis von Robotern, das es auszuführen gilt. Weiterhin bleibt, darüber nachzudenken, ob sich nicht zumindest einige Roboter so weit entwickeln, dass sie irgendwann auch als natürliche Personen zu begreifen wären und damit selbst Rechte erhalten sollten (wie auch in der Frage angedeutet) – ganz wie auch menschliche Kinder für gewöhnlich mit 18 Jahren zu Rechtspersonen im vollen Sinne werden, sofern sie die dafür nötigen Kompetenzen ausbilden. Dem liegt also wiederum die Frage zugrunde, was einen Menschen zu einer Rechtsperson macht und ob die diesem Status als Rechtsperson zugrunde liegenden Kompetenzen nicht auch durch Roboter simuliert oder ihnen implementiert werden könnten – vergleichbar den Fähigkeiten zur moralischen Handlungsfähigkeit, die in der Beantwortung der ersten Frage angesprochen wurden.

Thomas Zoglauer:

Unter einer Person versteht man gemeinhin ein lebendiges Wesen, das über Empfindungsfähigkeit, Bewusstsein, Intelligenz, Freiheit und Intentionalität verfügt, das selbstbestimmt handelt und seine Wünsche und Interessen artikulieren kann. Wenn man Roboter als ›elektronische Personen‹ bezeichnet, wird die falsche Vorstellung geweckt, dass diese Maschinen über all die genannten Fähigkeiten und Eigenschaften verfügen. Zwar kann nicht ausgeschlossen werden, dass Roboter eines Tages tatsächlich ein Bewusstsein entwickeln. Aber davon sind wir heute noch weit entfernt. Selbst wenn man einen schwächeren, nicht-metaphysischen Personenbegriff zugrunde legt und das Personsein als soziales Zuschreibungsprädikat betrachtet, das all jenen Wesen – egal ob organisch oder künstlich – zukommt, die sich wie menschliche Personen verhalten und aktiv am gesellschaftlichen Leben teilnehmen, so erfüllen heutige Roboter dieses Kriterium nicht. Roboter können daher auch nicht rechtsfähig sein. Dazu müssten sie Rechte verstehen können und zu freien Willensäußerungen fähig sein. Diese Voraussetzungen sind für Roboter nicht gegeben. Roboter können auch keine Schmerzen empfinden, weshalb Roboterrechte nicht in Analogie zu Tierrechten postuliert werden können. Auch die Anwendung des Begriffs der juristischen Person auf Roboter macht keinen Sinn. Denn hinter juristischen Personen (z.B. Körperschaften, Vereinen, Stiftungen, Kapitalgesellschaften etc.) stehen natürliche Personen, die Verträge schließen oder Geschäfte abwickeln. Die Handlungsfähigkeit juristischer Personen wird damit auf die Handlungen natürlicher Personen zurückgeführt. Wenn Roboter aber keine natürlichen Personen sind, dann können sie

auch nicht als juristische Personen betrachtet werden. Es wäre allenfalls denkbar, dass Menschen in Analogie zur Eltern-Kind-Beziehung die Interessen von Robotern vertreten, falls diese überhaupt irgendwelche Interessen besitzen. Es mag zwar reizvoll sein, mögliche Science-Fiction-Szenarien durchzuspielen und sich zu überlegen, wozu Roboter eines Tages in der Lage sein werden. Aber Ethik-Kommissionen und legislative Institutionen sollten nicht auf der Grundlage spekulativer Überlegungen entscheiden. Von ihnen wird erwartet, dass sie Regelungen für Heute und nicht für eine Welt in tausend Jahren entwerfen.

Befürchtungen vor einem KI-Supergau, wie ihn Stephen Hawking, Bill Gates und Elon Musk heraufbeschwören, sind genauso übertrieben wie transhumanistische Wunschträume von einem technologischen Schlaraffenland. Die EU-Kommission sollte daher nüchtern die Chancen und Risiken eines künftigen Robotereinsatzes abwägen, Haftungsfragen klären und durch entsprechende rechtliche Regelungen Vorsorge treffen, dass Menschen nicht zu Schaden kommen und Werte wie Privatsphäre, Datenschutz und Menschenwürde geschützt werden.

Bruno Gransche:

Genau wie eine eigene Roboterethik der Ethik mit gebündeltem spezifischen Fachwissen einen konkreten Dienst erweist, so wäre auf politisch regulativer Ebene eine eigene ›Europäische Agentur für Robotik und Künstliche Intelligenz‹ ein Gewinn im Sinne der Bündelung von notwendigem spezifischem Wissen. Die angewandte Ethik steht in einem besonderen Wissen-Sollen-Spannungsfeld, da die Frage, was beispielsweise ein Roboter soll – und wie das in europäische oder nationale Gesetze zu kodifizieren wäre – vom Wissen abhängt, wann es sich überhaupt um einen Roboter handelt (Ab wann zählt ein Haus oder ein Auto dazu?) oder über welche Eigenschaften welches robotische System überhaupt verfügt. Da nach dem Grundsatz *ultra posse nemo obligatur* bzw. *Sollen impliziert Können* sinnvoll nur normativ geregelt und exekutiv durchgesetzt werden kann, was im faktischen Vermögen der SollensadressatInnen liegt, kommt dem Wissen um jenes Vermögen eine vorrangige Rolle zu. Dieses Wissen ist im Bereich jetziger und künftiger robotischer Systeme keineswegs trivial und eine Agentur auf europäischer Ebene könnte entsprechende Expertise über ein kritisches Niveau hinaus bündeln.

Auch der eigene rechtliche Status für Roboter – etwa der Vorschlag einer ›elektronischen Person‹ – scheint hier auf regulatorischer Ebene zunächst sinnvoll, wenn auch keineswegs alternativlos (warum z.B. nicht ›automatisierte‹ oder ›algorithmische Person‹?). Dieser Status soll es nach dem Vorschlag des EU-Parlaments ermöglichen v.a. Haftungsfragen in Schadensfällen durch Roboter zu klären, etwa Pflichtversicherungen, Entschädigungen etc. zu regeln. Das ist insofern pragmatisch und

unspektakulär, als das Recht genau zu solchen regulatorischen und v.a. Haftungsfragen längst unter den Rechtssubjekten neben natürlichen auch juristische Personen (und weitere wie verschiedene Gesellschaften) unterscheidet. Hier elektronische oder algorithmische Personen als Rechtssubjekte aufzunehmen, kann für die Regelung von Haftungsfragen etc. sehr sinnvoll sein. Das hängt mehr von der Regelung selbst als von dem genauen Namen ab. Hier darf jedoch nicht der Begriff der Person mit dem des Individuums oder des (Selbst-)Bewusstseins verwechselt werden. Roboter als elektronischen Personen einen eigenen Rechtsstatus zu schaffen und ihnen Pflichten (wie Haftung und Entschädigung) aufzuerlegen und Rechte (wie das zur Pflichterfüllung notwendige Recht, Entschädigungsressourcen anhäufen zu dürfen) einzuräumen, macht Roboter ebenso wenig zu empfindsamen Bewusstseinen im Sinne des Menschen mit Menschenrechten etc. wie dies juristische Personen zu menschlichen Individuen macht.

Ab wann Roboter in diesem Sinne einen metaphysisch qualitativ und nicht nur regulatorisch anderen Status erhalten sollten, ist nicht mit dem Vorschlag des EU-Parlaments angesprochen und wieder eine Frage der starken KI bzw. entsprechender Philosophie. Einen solchen metaphysisch qualitativ anderen Roboter halte ich gegenwärtig für inexistent und sogar für künftig unmöglich.

Eric Hilgendorf:

Der Vorschlag des europäischen Parlamentes, über die Figur einer ›elektronischen Person‹ nachzudenken, ist weit weniger radikal, als dies auf den ersten Blick erscheinen könnte. Die Debatte darüber wird schon seit etlichen Jahren geführt. Dabei geht es weniger um Rechte für Roboter, sondern darum, wie Haftungslücken bei Schädigungen von Menschen durch Maschinen vermieden bzw. geschlossen werden können.

Angenommen, ein einwandfrei funktionierendes selbstlernendes System hat ›das Falsche‹ gelernt und verursacht deshalb einen Sachschaden. Wenn man davon ausgeht, dass dem Hersteller hier kein Fehler anzulasten ist, besteht die Gefahr, dass der Geschädigte keinen Schadensersatz erhält. Der konkrete ›Schädiger‹ ist eine Maschine und kann deshalb bislang weder verklagt noch sonst wie zur Verantwortung gezogen werden. Rechtliche oder moralische Verantwortungssubjekte können, so die seit der Aufklärung ganz vorherrschende Auffassung, nicht Sachen, sondern nur andere Menschen, also ›natürliche Personen‹, sein.

Immerhin hat das Recht bereits vor über einhundert Jahren die Figur der ›juristischen Person‹ anerkannt, sodass zum Beispiel Unternehmen verklagt werden können. Gäbe es diese Möglichkeit nicht, so würden viele Geschädigte auf ihren Schäden sitzen bleiben, denn angesichts der erheblichen Komplexität eines Großunter-

nehmens ist häufig nicht erkennbar, welcher Mensch genau den Schaden zurechenbar verursacht hat. Derselbe Gedanke könnte dazu führen, auch elektronische Personen anzuerkennen. Hersteller müssten ihren Robotern eine bestimmte Vermögensmenge zuordnen (zum Beispiel über eine Versicherung), und Geschädigte hätten dann die Möglichkeit, unmittelbar die Maschine zu verklagen und sich, wenn sie Recht bekommen, aus dem Vermögen der Maschine entschädigen zu lassen.

Von eigenen Rechten für Maschinen sind wir noch weit entfernt. Unklar ist schon, welche Anknüpfungspunkte für die Zuerkennung derartiger Rechte akzeptiert werden können (siehe oben, Frage II). Bemerkenswert ist, dass es dennoch nicht ausgeschlossen erscheint, dass eine Bewegung >Rechte für Maschinen< entstehen könnte: Es ist zu erwarten, dass in Zukunft mehr und mehr Roboter zu unseren engen Begleitern werden, etwa im Kinderzimmer, im Krankenhaus oder im Pflegeheim. Mensch-Maschine Beziehungen dieser Art funktionieren dann am besten, wenn es den Menschen, etwa durch ein entsprechendes Aussehen der Maschinen, ermöglicht wird, zu den Geräten eine enge emotionale Beziehung aufzubauen. Die Roboterrobbe Paro, die bereits heute in vielen Krankenhäusern eingesetzt wird, ist ein gutes Beispiel hierfür.

Wird ein robotischer Gefährte, zu dem man eine enge emotionale Beziehung aufgebaut hat, vorsätzlich beschädigt (>verletzt<), so wird aller Wahrscheinlichkeit nach rasch der Ruf nach harten Strafen laut werden. Die Auffassung, es handele sich lediglich um eine Sachbeschädigung, dürfte dagegen wenig überzeugend wirken (und möglicherweise sogar den üblichen Vorurteilen gegen Juristen neuen Auftrieb geben). Von hier aus ist es nur noch ein Schritt zur Vorstellung, humanoide Maschinen, die jedenfalls äußerlich und in ihrem Verhalten kaum mehr von Menschen zu unterscheiden sind, müssten eigene Rechte erhalten.

Catrin Misselhorn:

Solange es um die Frage geht, ob Roboter im juristischen Sinn Rechtspersonen sein sollen, wie etwa auch Unternehmen, müsste man die Bedingungen des rechtlichen Personenstatus sowie die rechtlichen Vorteile der einen oder anderen Konstruktion in Haftungsfragen diskutieren. Verantwortung können sie selbst allerdings nicht übernehmen. Moraleische Personen sind Roboter ebenfalls nicht, weil hierfür Eigenschaften wie Bewusstsein, Willensfreiheit und Selbstreflexivität notwendig sind, die noch auf längere Sicht Menschen vorbehalten bleiben. Aus diesem Grund sind Roboter auch keine Träger von Rechten.

IV. Die Schlussstatements

Thomas Zoglauer

Unter den Teilnehmern der Kontroverse besteht offenbar Einigkeit, dass heutige Roboter kein Bewusstsein besitzen und auch in absehbarer Zukunft nicht besitzen werden. Wenn man aber Bewusstsein als Voraussetzung für Moral- und Verantwortungsfähigkeit betrachtet, kann man nicht von moralischen Robotern oder *moral agents* sprechen. Dass ein Roboter einem »moralischen« Programm folgt, macht ihn noch nicht zu einem moralischen Agenten. Catrin Misselhorn stellt zu Recht fest, dass Roboter keine Verantwortung übernehmen und keine Träger von Rechten sein können, weil ihnen die hierzu notwendigen Eigenschaften wie Bewusstsein, Willensfreiheit und Selbstreflexivität fehlen.

Janina Loh wirft die Frage nach der Handlungsfähigkeit von Robotern auf. Man sollte sich allerdings hüten, Begriffe zur Beschreibung menschlicher Verhaltensweisen unreflektiert auf Roboter zu übertragen. Wenn man beispielsweise sagt, dass Artefakte oder ganze technische Systeme »handeln« könnten oder eine Form von *agency* besäßen, wie dies in der Akteur-Netzwerk-Theorie gern behauptet wird, so ist dies ein fahrlässiger »Hang zum Anthropomorphismus« (Hilgendorf, S. 235). Handeln setzt Intentionalität voraus, die bei heutigen Computern und Robotern nicht vorhanden ist. Man kann einen Roboter so konstruieren und programmieren, dass er wie ein Mensch geht. Aber dies ist, ebenso wie die mechanische Ausführung eines Algorithmus, noch kein Handeln. Ein Roboter, auch wenn er scheinbar zielgerichtet Aufgaben erledigt, entwirft keinen Handlungsplan, er überlegt sich nicht, wie er die Aufgabe am besten erledigen kann und er kann nicht zwischen verschiedenen Handlungsoptionen wählen. Auch ein selbstlernendes neuronales Netzwerk stattet einen Roboter nicht mit einem freien Willen aus. Selbst dann, wenn Roboter immer menschenähnlicher werden und Tätigkeiten des Menschen immer täuschender nachahmen können, wäre es eine bloß metaphorische Redeweise, wenn man von handelnden Robotern spricht oder ihnen Absichten unterstellt.

Wie weit die Verwendung vermenschlichender Begriffe in der Roboterethik bereits fortgeschritten ist, lässt sich an den Reaktionen zu der dritten Frage ablesen. Susanne Beck, Bruno Gransche, Eric Hilgendorf und Janina Loh halten es für sinnvoll, Robotern den Rechtsstatus einer »elektronischen Person« zuzusprechen, um damit Haftungsfragen zu klären, die dann auftreten, wenn bei einem Roboterereinsatz Menschen zu Schaden kommen. Mir scheint, dass es bei der Einführung des Begriffs der »elektronischen Person« aber in erster Linie darum geht, die Industrie im Schadensfall von Verantwortung freizusprechen und den Roboter zum Sündenbock zu machen. Man klagt den Roboter an und verpflichtet ihn zu Schadensersatzzahlungen.

Nehmen wir einmal an, durch das Verhalten eines Roboters sei ein Mensch ums Leben gekommen. Die naheliegende Frage stellt sich nun: Wer ist schuld? Wer trägt die Verantwortung und wer ist für den Schaden haftbar zu machen? Wenn man sagt, der Roboter als »elektronische Person« sei schuld und habe für den Schaden zu haften, macht man es sich in vielerlei Hinsicht zu einfach. Denn was kann der Roboter dafür, dass ein Mensch starb? Ein Roboter, auch wenn es sich um einen selbstlernenden Roboter handelt, folgt einfach einem Programm und kann nicht »anders handeln«. Soll man den Roboter des Mordes, des Totschlags oder der fahrlässigen Tötung anklagen? Der Roboter kann sich vor Gericht nicht selbst verteidigen. Susanne Beck schlägt daher vor, eine elektronische Person könne durch einen Menschen vor Gericht vertreten werden. Aber welche Interessen vertritt der Vertreter: die des Roboters oder der Herstellerfirma?

Die Herstellerfirma wird sich herausreden, indem sie behauptet, dass das Verhalten eines selbstlernenden Roboters nicht voraussehbar ist und aufgrund seiner Autonomie nicht kontrollierbar ist. Und man könne, so wird behauptet, nur für solche Handlungen verantwortlich gemacht werden, die man kontrollieren und deren Folgen man vorhersehen kann. Nach der Reaktorkatastrophe von Fukushima redeten sich die Manager des Kraftwerksbetreibers mit diesem Argument heraus: Sie konnten ja nicht voraussehen, dass ein so schweres Erdbeben eintritt und eine so hohe Flutwelle die Anlage überschwemmt. Die daraus sich ergebenden Folgen seien nicht mehr kontrollierbar gewesen. Aber die Manager und Konstrukteure der Anlage hätten im Sinne einer prospektiven Verantwortung die Risiken erkennen und die möglichen Schadensfolgen vorhersehen müssen. Mangelnde Vorhersehbarkeit und Kontrollfähigkeit entbindet nicht von Verantwortung. Selbst wenn bei einem durch einen Roboter verursachten Unfall keine einzelnen Personen – seien es die ProgrammiererInnen, KonstrukteurInnen oder HerstellerInnen – verantwortlich gemacht werden können, gibt es eine korporative und kollektive Verantwortung.

Susanne Beck und Eric Hilgendorf wollen der Maschine ein »eigenes Vermögen« (S. 243) zusprechen, um damit Haftungsprobleme zu lösen. Wenn der elektronischen Person vor Gericht aber kein Fehlverhalten nachgewiesen werden kann, denn sie konnte ja aufgrund ihres Programms gar nicht anders handeln, warum sollte sie dann mit ihrem »eigenen Vermögen« (ebd.) haften? Beck argumentiert, dass eine elektronische Person nur eine rechtliche Konstruktion sei, um eine pragmatische Lösung für Rechtsstreitigkeiten und Haftungsfragen zu finden. Aber warum klagt man dann nicht gleich die wirklich Verantwortlichen, z.B. die Herstellerfirma, an? Wenn eine Schuld nicht zweifelsfrei einer der Parteien (ProgrammiererInnen, KonstrukteurInnen, HerstellerInnen) zugeordnet werden kann, dann kann ein Fonds gegründet werden, in den alle Beteiligten eine bestimmte Summe einzahlen und aus dem dann der Schaden beglichen wird. Dann braucht man den Roboter nicht zum Sündenbock machen und braucht nicht den obskuren Begriff der »elektronischen Person« einführen.

Eric Hilgendorf weist auf die durchaus reale Gefahr einer allzu engen emotionalen Bindung an menschenähnliche oder auch nur niedliche Roboter hin. Aber allein die Tatsache, dass ein Roboter nur niedlich aussieht und ein enger Begleiter des Menschen ist, ist noch lange kein Grund, ihn als Rechtspersonen anzuerkennen oder »Rechte für Maschinen« (Hilgendorf, S. 248) zu fordern. Wenn jemand einen lieb gewonnenen Roboter mutwillig zerstört, so stellt sich die Frage, ob dies als Sachbeschädigung, die Zerstörung einer Mensch-Maschine-Beziehung oder gar als ›Roboter-Mord‹ geahndet werden soll. Im Gegensatz zu einer reinen Sachbeschädigung werden dabei auch menschliche Gefühle verletzt. Aber deshalb muss man Roboter nicht gleich zu Rechtssubjekten oder ›elektronischen Personen‹ erklären. Man könnte ähnlich wie im Tierschutzrecht ein ›Roboterschutzrecht‹ einführen, das eine mutwillige Zerstörung humanoider Roboter unter Strafe stellt und dies mit dem gleichen Argument begründen, mit dem Immanuel Kant für den Tierschutz plädiert: weil durch eine grausame Behandlung menschenähnlicher Roboter das Mitgefühl im Menschen abstumpft und die Moralität im Verhältnis zu anderen Menschen geschwächt wird. Aufgabe der Roboterethik wird es daher auch sein, die psychologischen und sozialen Auswirkungen des Robotereinsatzes auf den Menschen kritisch zu analysieren.

Janina Loh

Verantwortung ist traditionell ein individualtheoretisches Phänomen, d.h., dass es im ursprünglichen Verständnis des Wortes einer einzelnen Person bedurfte, der man Verantwortung zuschreiben konnte. Verantwortung im klassischen Sinne trägt *jemand* (ein Subjekt bzw. ein*e Träger*in) seit dem 17. Jahrhundert im deutschsprachigen Raum (denn erst in dieser Zeit taucht das Wort ›verantwortlich‹ auf) *für* ein Objekt bzw. einen Gegenstand, *vor* einer Instanz, *gegenüber* einer Adressat*in *auf der Grundlage* spezifisch für diesen Kontext geltender normativer Kriterien. So ist beispielsweise ein*e Dieb*in *für* einen Diebstahl *vor* Gericht *gegenüber* der bestohlenen Person *auf der Grundlage* des Strafgesetzbuches verantwortlich. Aus diesem Beispiel geht hervor, dass mit Verantwortung in einem normativen Sinne die Fähigkeit des Rede-und-Antwort-Stehens gemeint ist, woraus folgt, dass der Person, der Verantwortung aufgebürdet werden soll, gewisse Kompetenzen zugeschrieben werden, die ihr die Verantwortungsübernahme ermöglichen, wie beispielsweise Kommunikations- und Handlungsfähigkeit bzw. Autonomie sowie Urteilskraft. Alle diese Kompetenzen und mit ihnen die Verantwortung selbst sind graduell bestimmbar; man kann von mehr oder weniger Kommunikations- und Handlungsfähigkeit sprechen und abhängig davon von mehr oder weniger Verantwortung.

Mit der Übertragung von Verantwortung auf Gruppen ging folglich die Frage einher, wie es sich nun mit der* Einzelnen und ihrer* individuellen Verantwortung innerhalb des Kollektivs verhält; ist sie genauso ›groß‹ (quantitativ) und ist es immer noch ›dieselbe‹ (qualitativ) Verantwortung wie außerhalb der Gruppe? Haben also alle Mitglieder eines Kollektivs immer noch die ›volle‹ Verantwortung für den fraglichen Gegenstand, oder lediglich noch eine Teilverantwortung? Was passiert nun, wenn jemand zwar Mitglied einer Gruppe ist, aber zu den Umständen, für die besagte Gruppe verantwortlich gemacht wird, persönlich gar nichts beigetragen hat? Oder noch irritierender: Was ist mit solchen Kontexten, wo es scheint, als wäre gar niemand verantwortlich für das Geschehen zu machen? Seit dem 20. Jahrhundert bewegen sich die Menschen verstärkt gerade in solchen intransparenten Bezügen: Wie verhält es sich beispielsweise mit der Klimaverantwortung, mit der Verantwortung im globalen Finanzmarktsystem, wo Algorithmen am Werk sind, die noch nicht einmal mehr von den Algorithmiker*innen, die diese programmiert haben, verstanden werden (zumindest behaupten diese das) oder mit der Verantwortung im Umgang mit autonomen Fahrrassensystemen?

Häufig wird die Möglichkeit einer Verantwortungsübernahme von artifiziellen Systemen und insbesondere Robotern mit dem Verweis auf die oben genannten Kompetenzen als Bedingung für die Zuschreibung von Verantwortung bestritten, die bei den fraglichen Maschinen nicht vorlägen: Roboter würden weder über Urteilskraft, Handlungsfähigkeit, Autonomie noch über sonstige Fähigkeiten verfügen, die für die Übernahme von Verantwortung eine Rolle spielten. Einige Denker*innen entwickeln jedoch Strategien (wie z.B. Wallach und Allen den Ansatz der funktionalen Äquivalenz),⁶ mit dem dieses Problem mangelnder Kompetenzen bei artifiziellen Systemen umgangen werden kann. Aber auch Theorien wie diese gelangen letztlich zu dem Schluss, dass aus der Tatsache, bestimmte Fähigkeiten wie Autonomie, Urteilskraft und andere mehr oder minder gut zu simulieren in der Lage zu sein, nicht gleich folgt, Robotern in dem selben Ausmaß wie Menschen und gar in einem signifikanten Sinn Verantwortung zuzuschreiben. Bis auf weiteres ›trumpfen‹ Menschen in Sachen Verantwortung artifizielle Systeme um Längen. Artifizielle Verantwortungszuschreibung ist bislang nur in begrenztem Maße möglich.

Für Fälle, in denen Verantwortung zwar zugeschrieben werden soll, aber die Subjektposition der fraglichen Verantwortlichkeit nicht besetzbare erscheint, haben einige Verantwortungstheoretiker*innen in den vergangenen Jahren behelfsmäßige Begrifflichkeiten zu entwickeln versucht, die ohne eine Bestimmung dieser Funktion auskommen. Das wird dann beispielsweise Systemverantwortung genannt und soll die Verantwortung ›des Systems‹ bei gleichzeitiger Verantwortungslosigkeit seiner einzelnen Mitglieder bedeuten. Ich bin skeptisch, dass uns damit geholfen ist. Schließ-

⁶ Wendell Wallach und Colin Allan: *Moral Machines. Teaching Robots Right from Wrong*, Oxford 2009.

lich suchen wir *de facto* immer nach einer* Träger*in (Singular oder Plural), die in der Lage ist, die eingeforderte Verantwortung zu schultern. Traditionell funktioniert unser Verständnis von Verantwortung in dieser Weise. Aber wie sollten wir stattdessen in solchen Kontexten verfahren?

Ich möchte hier den Begriff des Verantwortungsnetzwerkes von Christian Neuhauser übernehmen und spezifizieren.⁷ Die meinen Überlegungen zugrundeliegende These lautet, dass wir all denjenigen Parteien in einer gegebenen Situation Verantwortung zuschreiben, die an dem fraglichen Geschehen beteiligt sind, in dem Maße, in dem sie die nötigen Kompetenzen zur Verantwortungszuschreibung mitbringen. Um einmal bei dem Beispiel autonomer Fahrassistentensysteme zu bleiben: In das Verantwortungsnetzwerk ›Verantwortung im Straßenverkehr‹ sind autonome Autos zunächst ebenso zu integrieren wie die* menschliche Fahrerin* (selbst dann, wenn sie* nicht aktiv am Fahrprozess beteiligt ist), die Besitzer*innen, die Verteiber*innen, die Programmierer*innen, die Designer*innen, die Öffentlichkeit, Jurist*innen, Fahrlehrer*innen – also alle am Straßenverkehr Beteiligten.

Verantwortungsnetzwerke haben ungewöhnliche Ausmaße und bündeln in sich unterschiedliche Verantwortungsobjekte. Von Verantwortungsnetzwerken kann man dann sprechen, wenn man eigentlich gar nicht mehr weiß – und um solche Kontexte geht es hier ja gerade –, ob hier in einem gehaltvollen Sinn Verantwortung definiert werden kann, gerade weil beispielsweise die Bestimmung eines Subjekts schwierig ist oder aber sich keine eindeutige Instanz ausmachen lässt oder aber die normativen Kriterien nicht benannt werden können. In einem Verantwortungsnetzwerk erfüllen die involvierten Parteien unterschiedliche Funktionen bzw. besetzen manchmal mehrere Positionen zugleich, sind einmal die Subjekte, in einem anderen Fall die Instanzen und wieder in einem anderen Fall das Objekt und vielleicht zugleich Adressat*innen einer Verantwortlichkeit.

Es wäre äußerst schwierig, ein oder mehrere konkrete Verantwortungssubjekte für die Verantwortung im Straßenverkehr auszumachen, da diese viel zu umfassend ist, als dass eine Person oder eine geringe Anzahl Einzelter dafür Rede und Antwort stehen könnte. Als Verantwortungsnetzwerk ›Verantwortung im Straßenverkehr‹ werden hier jedoch mehrere Bereiche – beispielsweise moralische, juristische und politische Verantwortlichkeiten (definiert über moralische, juristische und politische Normen) – umfasst. Der Straßenverkehr stellt nur das übergeordnete Verantwortungsobjekt dar, für das nicht eine oder mehrere Personen gehaltvoll ›die‹ Verantwortung tragen, das sich jedoch in unterschiedliche, weniger komplexe Gegenstände ausdifferenziert, für die dann die unterschiedlichen Parteien jeweils eine spezifische Verantwortung übernehmen. Verantwortung für den Straßenverkehr kann in einem Fall die Sicherheit der am Straßenverkehr Beteiligten Menschen bedeuten, in einem

7 Christian Neuhauser: »Roboter und moralische Verantwortung«, in: Eric Hilgendorf (Hg.): *Robotik im Kontext von Recht und Moral*, Baden-Baden 2014, S. 269–286.

anderen Verständnis die Verantwortung für das schnelle und effiziente Gelangen von A nach B und in noch einem anderen Fall die Verantwortung dafür, dass die moralischen Herausforderungen, die mit einer Beteiligung am Straßenverkehr einhergehen, denjenigen, die sich am Straßenverkehr beteiligen, mit hinreichender Ausführlichkeit zuvor deutlich gemacht wurden. Über die beschriebenen (und zahlreiche weitere) Teilverantwortungsgegenstände wird bereits nachvollziehbar, dass wir jeweils ganz unterschiedliche Subjekte in unterschiedlichem Ausmaß dafür zur Verantwortungsübernahme ansprechen würden, dass es jeweils unterschiedliche Instanzen, Adressat*innen und Normen sind, die zur Konkretisierung der jeweiligen Verantwortlichkeit definiert zu werden verlangen.

Gegenwärtig wird ein autonomes Fahrassistentensystem, das nur in einem sehr schwachen Sinne – wenn überhaupt – als Verantwortungsakteur identifizierbar ist, da es ihm an den fraglichen Kompetenzen mangelt, die Subjektposition einer Verantwortlichkeit innerhalb des Verantwortungsnetzwerkes ›Verantwortung im Straßenverkehr‹ nicht besetzen können. Denn es gibt immer potenziell qualifiziertere Verantwortungsträger*innen. Allerdings ist denkbar, es als Verantwortungsobjekt und als Adressat in eine oder mehrere der Verantwortlichkeiten dieses Verantwortungsnetzwerkes einzubinden.

In dieser Weise kann Verantwortung auch in intransparenten Kontexten, in denen es beispielsweise dank der involvierten Algorithmen und Roboter zunächst so aussehen mag, als wäre eine eindeutige Identifikation der verantwortlichen Subjekte schwierig, immer noch definiert werden. Vielleicht müssen wir also gar nicht unser tradiertes Verständnis von Verantwortung transformieren, sondern lediglich die Interpretationsebene wechseln, nämlich dann, wenn der Bezugsrahmen – wie in dem besprochenen Beispiel einer ›Verantwortung im Straßenverkehr‹ – zu weit erscheint, auf die Ebene ›darunter‹ wechseln, die es uns erlaubt, unterschiedliche Verantwortlichkeiten mit je eigenen Verantwortungssubjekten, -objekten, -instanzen, -adressat*innen sowie normativen Kriterien auszumachen.

Eric Hilgendorf

Für die Beantwortung von Fragen nach dem angemessenen – praktischen, ethischen oder juristischen – Umgang mit Robotern erscheint es mir derzeit kaum sinnvoll, sich auf eine exakte Definition des Gegenstands ›Roboter‹ festzulegen. Definitionen sind Festsetzungen des Sprachgebrauchs, die unter Zweckmäßigkeitsgesichtspunkten vorgenommen werden. Das Alltagsverständnis von ›Roboter‹ – ein teilweise autonom agierendes Computersystem plus ein Aktuator plus eventuell weitere bewegliche Teile sowie Sensoren – reicht bislang aus, um die heute relevanten Fragen sinnvoll diskutieren zu können. Eine präzisere Definition würde erst dann erforder-

lich, wenn an das Vorliegen eines ›Roboters‹ bestimmte Folgen, zum Beispiel Rechtsfolgen, geknüpft würden. Legt z.B. ein Gesetz fest, dass für das Handeln eines Roboters gehaftet wird, so muss präzisiert werden, was den Roboter von anderen Entitäten (wie Computerprogrammen, beweglichen Puppen, Automobilen oder Kränen) unterscheidet. Derartige Normen existieren jedoch bislang nicht.

Ähnliches gilt für die Frage nach der Handlungsfähigkeit von Maschinen. Als Handlung eines Roboters wird man bis auf Weiteres jede Bewirkung einer Außenweltveränderung genügen lassen können, ohne dabei besondere Anforderungen an mentale Zustände des Handelnden zu formulieren. Ein komplexerer Handlungsgriff, wie ihn Herr Kollege Zoglauer vertritt, ist im Kontext der ›normativen Einhegung‹ der modernen Robotik bislang nicht erforderlich. Es wäre verfehlt, unreflektiert im Sinne eines Sprach-Essentialismus von einem feststehenden vorgegebenen Handlungskonzept auszugehen und das Wort ›Handeln‹ zwingend mit so umstrittenen Konzepten wie ›Bewusstsein‹, ›Fähigkeit zum Planen‹ oder ›Gewissen‹ zu verbinden. Hinter derartigen essentialistischen Tendenzen verbirgt sich übrigens häufig das menschliche Bedürfnis nach Einzigartigkeit: Nur wir können ›wahrhaft‹ handeln, andere Wesen können Handeln allenfalls simulieren. (Bemerkenswert ist, dass derartige Vorurteile in der westlichen, vom Christentum geprägten Kultur weitaus häufiger vorzukommen scheinen als im asiatischen Raum).

Interessant ist die Frage nach den sozialen Folgen des Auftretens von menschenähnlichen Maschinen. Wenn Roboter zu unseren alltäglichen Begleitern werden sollen, müssen sie nicht nur Menschen ähnlichsehen, sondern auch ähnlich wie Menschen reagieren. Dies wird dazu führen, dass mehr und mehr Emotionen, die normalerweise Menschen entgegengebracht werden, auf menschenähnliche Maschinen übertragen werden. Das Gleiche lässt sich ja bereits heute bei Haustieren feststellen (während wir uns gegenüber anderen Tieren, wie die Massentierhaltung zu Zwecken der Fleischproduktion zeigt, außerordentlich unempfindlich, man könnte auch sagen: grausam verhalten).

Starke Emotionen gegenüber Robotern könnten dazu führen, dass das Bedürfnis nach einem besonderen Schutz dieser Maschinen entsteht. Juristisch gesprochen, würden viele die Beschädigung eines humanoiden Roboters nicht als Sachbeschädigung angesehen, sondern als Körperverletzung (die derzeit nur an natürlichen Menschen begangen werden kann). Am Ende der Entwicklung könnte sogar eine Bewegung stehen, humanoiden Maschinen Rechte analog zu Menschenrechten zuzugestehen. Sieht man den Grund für die Vergabe eigener Rechte in Gesichtspunkten wie Selbstbewusstsein oder Leidensfähigkeit, so scheint es fraglich, ob sich entsprechende Forderungen rational ausweisen ließen. Immerhin könnte man, analog zur Debatte um Tierrechte, argumentieren, dass Grausamkeit gegenüber humanoiden Maschinen zu einer Verrohung der menschlichen Akteure führen könnte, die als solche sanktionswürdig ist, auch wenn die ›Opfer‹ selber nicht schutzbedürftig erscheinen.

Derartige Fragen dürften in der Zukunft in der Ethik und Rechtsphilosophie noch größere Aufmerksamkeit auf sich ziehen.

Sehr viel aktueller sind die Fragen, die sich mit der Verantwortung bzw. (juristisch gesprochen) Haftung von Maschinen beschäftigen. Hier bestehen, anders als Herr Kollege Zoglauer anzunehmen scheint, erhebliche Unsicherheiten bei der Zuschreibung von Verantwortung bzw. Haftung. Man wird davon ausgehen dürfen, dass mit dem Eindringen der Maschinen in die menschliche Lebens- und Arbeitswelt auch Schäden an Sachen und Menschen verbunden sind. Wer trägt für derartige Schäden die Verantwortung? Je autonomer die Maschinen agieren, desto schwieriger ist es, Menschen die Verantwortung für das maschinelle schadensträchtige Handeln aufzuerlegen. Hersteller, Programmierer oder Eigentümer der Maschine haften nicht ohne Weiteres, sondern nur dann, wenn bestimmte Voraussetzungen – Juristen sprechen von ›Tatbestandsmerkmalen‹ – erfüllt sind. Zu diesen Voraussetzungen gehört in der Regel, dass dem in Haftung Genommenen Vorsatz oder zumindest Fahrlässigkeit in Bezug auf den aufgetretenen Schaden vorgeworfen werden kann. Dies dürfte in Zukunft in vielen Fällen problematisch werden. Das Institut der Gefährdungshaftung, welches von Vorsatz und Fahrlässigkeit absieht, greift derzeit für Computerprogramme nicht. Damit drohen erhebliche Lücken beim Schadensausgleich.

Natürlich wäre es grundsätzlich denkbar, die Betroffenen auf ihren Schäden ›sitzen zu lassen‹. Ein solcher Weg ist aber in einem Sozialstaat wie der Bundesrepublik Deutschland problematisch. Empfehlenswerter scheint es deshalb, die Gefährdungshaftung auf autonome lernfähige Systeme auszuweiten. Dies könnte einerseits durch Interpretation des bestehenden Rechts, andererseits aber auch durch eine Entscheidung des Gesetzgebers geschehen. Schon aus Gründen der Klarstellung scheint mir der letztgenannte Weg vorzugswürdig zu sein. Die Ethik steht vor der Aufgabe, Verantwortung so präzise zu konstruieren, dass mit den Juristen eine Debatte auf Augenhöhe geführt werden kann.

Die Einführung der Figur einer ›elektronischen Person‹ (*e-person*) ist zwar grundsätzlich möglich, scheint aber derzeit nicht erforderlich. Dafür spricht immerhin die Möglichkeit, die Figur der *e-person* zur Schließung von Haftungslücken zu verwenden. Gegen eine Einführung ›elektronischer Personen‹ spricht aber, dass im Alltag und auch in großen Teilen der Ethik mit dem Begriff ›Person‹ sehr weitgehende Vorstellungen verbunden sind, etwa die Zuschreibung eines besonderen moralischen Status oder gar der Besitz von ›unantastbaren‹ Rechten. Es erscheint deshalb vorzugswürdig, die Haftungsproblematik durch eine Ausweitung von Gefährdungshaftung zu schließen.

Eine strafrechtliche Verantwortung von humanoiden Maschinen lässt sich allerdings auch auf diese Weise nicht konstruieren, da nach heute ganz herrschendem Sprachgebrauch in der Rechtswissenschaft eine Strafe nur dann möglich ist, wenn das zu bestrafende Wesen schuldfähig ist. Maschinen erfüllen diese Voraussetzung

nicht. Auch die Einführung des Konstrukts »elektronische Person« würde daran nichts ändern. Für eine Strafbarkeit humanoider Roboter könnte allenfalls ein entsprechendes Strafbedürfnis in der Bevölkerung ins Feld geführt werden. Derartige Probleme stellen sich jedoch derzeit praktisch (noch?) nicht, auch wenn es theoretisch reizvoll erscheint, das Konzept »Strafe« anlässlich des Auftretens unserer neuen Begleiter verfremdend zu hinterfragen und neu zu diskutieren.

Nicht alles, was technisch möglich ist, sollte umgesetzt werden. Dies ist letztlich eine Trivialität. Sehr viel anspruchsvoller ist die Beantwortung der Frage, welche Grenzen wir der Entwicklung der modernen Robotik im Einzelnen setzen sollten. Dabei geht es um Fragen der Haftung auf Ersatz von maschinell bewirkten Schäden, um die strafrechtliche Verantwortung von Herstellern, Programmierern und Nutzern, um Datenschutz, Versicherungsrecht, urheberrechtliche Fragestellungen und vieles andere mehr. Voraussetzung für eine sinnvolle Auseinandersetzung mit derartigen Fragen ist zunächst eine Klärung der real drohenden Gefahren, aber auch der Chancen, die die neue Technologie für die weitere Humanisierung unserer Lebens- und Arbeitswelt birgt. Zuständige Disziplin hierfür ist die Technikfolgenabschätzung. Die Technikethik kann helfen, Problemstellungen zu analysieren und zu klären. Sie verfügt über einen großen Schatz an Argumenten und Problemlösungsversuchen, die auch auf neue Fragestellungen übertragen werden können. Das schärfste Mittel zur normativen Einhegung neuer Technologien ist das Recht. Als Leitgedanke kann der Grundsatz dienen, dass die Technik auf den Menschen ausgerichtet sein sollte, und nicht umgekehrt. Darin liegt das Grundpostulat eines »juristischen Humanismus«, für den ich hier plädieren möchte.

Susanne Beck

Die Analyse der Problematik im Bereich »Roboter« sowie die Antworten auf die verschiedenen Fragen definitorischer, aber auch gestalterischer Art hängen m.E. nicht zuletzt davon ab, ob man eine funktionale Perspektive einnimmt oder eher intrinsische (»metaphysische«) Antworten sucht. Aus meiner eher funktionalen – d.h. die Lösung konkreter gesellschaftlicher Probleme fokussierenden – Perspektive gibt es nicht die eine, richtige Definition von »Roboter« und »Handlung«, oder es ist zumindest aus dieser Perspektive nicht zielführend, nach ihr zu suchen. Vielmehr sollte die Definition abhängig sein von der konkreten Fragestellung, dem spezifischen Kontext.

Die Suche nach Lösungen für Probleme wie die Verantwortungsdiffusion durch die Robotik ist hiernach abhängig vom konkreten Lebenskontext sowie vom normativen Rahmen, der Lösungen bereitstellen soll – die moralische Bewertung kann von der rechtlichen deutlich divergieren. Die Definitionen sowohl des Gegenstands »Ro-

boter< als auch der ›Handlung< sollten deshalb nicht nur zwischen Moral und Recht (und sonstigen Lebensbereichen/Disziplinen) divergieren, sondern auch innerhalb des Rechts von der konkreten Fragestellung abhängen. So ist es durchaus möglich, im Kontext von Willenserklärungen Regeln zu schaffen für die Konstellationen, in denen Menschen mit Programmen oder Programme mit Programmen Verträge schließen, ohne dass es näherer definitorischer Festlegungen bedürfte. Auch bei Kraftfahrzeugen und ähnlichen mit Menschen interagierenden Maschinen ist eine Einteilung nach ›Autonomie-Graden (auch dieser Begriff wird hier rein funktional verwendet) denkbar bzw. wird praktisch auch durchgeführt, und an unterschiedliche Grade lassen sich dann unterschiedliche rechtliche Regelungen knüpfen.

Ähnliches gilt dann auch für den ›sozialen Charakter<. Dieser ist wiederum nicht als solcher negativ oder positiv einzuschätzen bzw. in jedem Lebensbereich zu fördern oder zu vermeiden. Vielmehr sind sinnvolle Gestalt, angemessene Verhaltensweise etc. wiederum abhängig vom jeweiligen Lebenskontext. Zugleich ist an dieser Stelle festzuhalten, dass jedenfalls eine bewusste Täuschung der Menschen, die mit der Maschine interagieren, über deren Fähigkeiten und Innenleben zweifellos als problematisch anzusehen ist – wenn derartiges Vorgehen m.E. aber auch nicht zwingend rechtlich verboten sein sollte. Auch dass gewisse Interaktionen und Beziehungen schutzwürdig sein könnten, ist sicher zutreffend; eine enge persönliche Beziehung zu einem Roboter kann durchaus gerade als solche, also als diese Beziehung, schutzwürdig sein. Auch hier halte ich jedoch eine neue rechtliche Kategorie zumindest für verfrüh; nur falls sich bestimmte Maschinen gesellschaftlich dauerhaft als akzeptables Gegenüber etablieren, könnte man über Neukategorisierungen im Recht (etwa eine spezifische Kategorie innerhalb der ›Sachbeschädigung< im Strafrecht) diskutieren. Hier kann das Recht aber letztlich nur der Gesellschaft nachfolgen – es wäre zu beobachten, wie diese mit Maschinen umgeht, ob es solche als ›Gegenüber< akzeptierte Akteure gibt, und dann gegebenenfalls eine Anpassung des bestehenden Rechts an die soziale Praxis vorzunehmen.

Gerade mit Blick auf diese Debatten bzw. auch die von mir betonte Differenzierung zwischen funktionaler Betrachtungsweise und einer eher ontologischen Herangehensweise möchte ich die Frage nach der Verantwortung bzw. Haftung zumindest zum Teil von der Frage nach dem Sinn einer ›elektronischen Person< trennen.

Die Einführung der ›elektronischen Person< könnte, analog einer juristischen Person, der Erleichterung innerhalb des juristischen Systems dienen. Auch hierfür spielt es keine Rolle, ob diese Person handeln kann oder dem Menschen ähnlich ist, sondern ob sich mit der Einführung die relevanten rechtlichen Fragen tatsächlich leichter und adäquater beantworten ließen als ohne diese Konstruktion. Dies ist hier nicht im Einzelnen analysierbar, es ist jedoch davon auszugehen, dass sich bestimmte Fragen – etwa nach dem Adressaten im Anschluss an eine Schädigung – für bestimmte Maschinen leichter beantworten lassen, und in diesen Fällen durchaus Menschen als

Organe bzw. VertreterInnen einer in einem entsprechenden Register aufgenommenen elektronischen Person agieren könnten. Ob dies den Aufwand einer Neukonstruktion rechtfertigt, mag der Gesetzgeber entscheiden. Dabei wäre jedenfalls auch zu berücksichtigen, dass derartige Neukonstruktionen Wechselwirkungen mit unseren sozialen Verständnissen haben, d.h. dass diese Konstruktion zur Verfestigung der Akteureigenschaft von Maschinen bzw. Programmen beitragen könnten – auch die Frage, ob und inwieweit dies gewollt ist, ist eine gesellschaftliche und kann an dieser Stelle nicht beantwortet werden.

Unabhängig von diesen Überlegungen: Zentral bleibt m.E. die Frage nach der Verantwortung bzw. Haftung. Ob man die Verschiebungen in diesen Konzeptionen als Lücke oder Diffusion bezeichnet – man kann jedenfalls beim Einsatz autonomer Systeme die bestehenden rechtlichen Konzepte nicht ohne Weiteres anwenden. Das liegt nicht primär an dem Zusammenwirken mehrerer Akteure (menschliche Kollektive) bei Herstellung und Einsatz dieser Systeme, sondern am *Deep Learning*, der Künstlichen Intelligenz, der Vernetzung der Maschinen untereinander und letztlich vor allem der bewussten Übertragung von Entscheidungen auf Maschinen, sei es teilweise, sei es ganz. Gerade diese Übertragung ist ja Sinn autonomer Systeme, und damit wäre es zumindest problematisch, diese durch unhinterfragte und undifferenzierte Anwendung der bisherigen Haftungs- und Verantwortungsregime praktisch im Nachhinein wieder auszuhebeln. Vielmehr ist zu berücksichtigen, dass mit der Herstellung und Verwendung derartiger Maschinen auch einhergeht, gerade die Entlastungsfunktion zu akzeptieren. Diese ist jedoch bei drohender individueller Verantwortung fast ausgehebelt. Zudem sollte jedenfalls vermieden werden, den ‚Letzentscheider‘ bzw. Nutzer vorschnell zur Verantwortung zu ziehen. Zwar ist richtig, dass er die Entscheidungen der Maschine häufig nur in seine eigene Entscheidung einbezieht oder zumindest kontrolliert bzw. im Falle der Fehlfunktion theoretisch eingreifen könnte. Häufig ist dies aber eben nur theoretisch möglich, während faktisch regelmäßig eine erhebliche Vorprägung durch die Maschine erfolgt bzw. ein Eingreifen überhaupt nicht schnell und informiert genug möglich ist. Wir würden vom Nutzer also in vielen Fällen erheblich mehr verlangen, als er eigentlich zu leisten in der Lage ist. Insbesondere wenn er nicht einmal selbst über den Einsatz der Maschinen entscheidet, sondern z.B. sein Arbeitgeber, wäre das eine ungerechtfertigte Belastung. Die aktuellen Haftungs- und Verantwortungsregime sind deshalb schon jetzt an die Veränderungen durch *Deep Learning*, Künstliche Intelligenz, Big Data und Vernetzung anzupassen. Insofern geht es im Übrigen nicht darum, von einer möglichen Selbstverantwortung der Maschinen auf eine Änderung der Verantwortungskonzeption zu schließen, im Gegenteil: Es ist zu beobachten, dass sich die Verantwortung verändert bzw. verändert muss, und zwar die Verantwortung der beteiligten Menschen. Diese muss an die neuen Gegebenheiten angepasst werden, bevor überhaupt über die Rolle der Maschinen nachzudenken ist.

Den schärfsten Kontrast bezüglich der Frage nach den neuen metaphysischen, ethischen oder juristischen Problemen, die hochentwickelte Roboter möglicherweise mit sich bringen (Frage I), sehe ich zwischen den Beiträgen von Susanne Beck und Thomas Zoglauer. Beck spricht von einer »Verantwortungsdiffusion« (S. 232) und schließt daraus, »die klassischen Zurechnungsstrukturen des Rechts sind hierauf nicht mehr ohne Weiteres anwendbar.« (S. 233) Zoglauer hingegen hält fest: »Wenn durch einen Robotereinsatz ein Schaden entsteht, ist immer noch der Mensch, sei es als Hersteller, Konstrukteur, Programmierer oder Nutzer, verantwortlich. Eine ›Verantwortungslücke‹ gibt es nicht.« (S. 236) Diese beiden Statements scheinen auf den ersten Blick nur schwer miteinander vereinbar zu sein.

Dieser scheinbare Widerspruch lässt sich aus meiner Sicht jedoch auflösen. Zu den klassischen Bedingungen für die Zuschreibung moralischer Verantwortung zählen Willensfreiheit, Kausalität, Absichtlichkeit und Wissen. Ein Akteur ist demnach nur dann für eine Handlung verantwortlich, wenn sie auf seinem freien Willen beruht, wenn sie ohne seine Beteiligung nicht zustande gekommen wäre, er sie absichtlich durchgeführt hat (oder ihre Folgen zumindest in Kauf genommen hat) und ihm ihre Folgen bekannt waren (er diese hätte vorhersehen können oder sich die entsprechenden Kenntnisse mit vertretbarem Aufwand beschaffen können). Es ist unumstritten, dass Roboter diese Bedingungen nicht alle erfüllen, und zwar insbesondere, weil sie keinen freien Willen besitzen; aber auch die Bedingungen der Absichtlichkeit und des Wissens werfen Probleme bei der Zuschreibung an Roboter auf. Richtig ist also, dass Robotern selbst keine moralische Verantwortung zukommen kann.

Eine andere Frage ist, ob gleichwohl eine Verantwortungslücke entsteht. Robert Sparrow, auf den dieser Begriff (im Original: *responsibility gap*) zurückgeht, argumentiert dafür am Beispiel von autonomen Kriegsrobotern.⁸ Ich rekonstruiere den Kern des Arguments (etwas über Sparrow hinausgehend) so: Eine Verantwortungslücke entsteht, wenn: (1) ein Kriegsroboter nicht absichtlich so programmiert wurde, dass er das *ius in bello* (hier: die ethischen bzw. rechtlichen Normen der Kriegsführung) verletzt; (2) es nicht vorhersehbar war, dass der Einsatz des Kriegsroboters in einer Verletzung des *ius in bello* resultieren würde; und (3) ab dem Start der Operation keine menschliche Kontrolle mehr über die Maschine bestand. Letzteres hängt mit der von Bruno Gransche und Janina Loh hervorgehobenen Autonomie von Robotern zusammen.

Das Problem ist, dass das Vorliegen dieser drei Bedingungen dazu führt, dass die moralische Verantwortung keinem Menschen zugeschrieben werden kann, wenn die Maschine im Widerstreit mit dem *ius in bello* Menschen tötet. Denn kein Mensch

⁸ Robert Sparrow: »Robots and respect: Assessing the case against Autonomous Weapon Systems«, in: *Ethics and International Affairs* 30 (2016), Heft 1, S. 93–116.

hatte dies beabsichtigt, es war nicht vorhersehbar und niemand hatte kausal die Möglichkeit, dieses Resultat zu verhindern. Eine Verantwortungslücke entsteht also genau dann, wenn die Maschine selbst nicht verantwortlich ist, aber ihr Einsatz die Bedingungen der Verantwortungszuschreibung an Menschen untergräbt.

Susanne Becks Begründung geht in diese Richtung, wenn sie die Unvorhersehbarkeit der Entscheidung der Maschinen sowie die mangelnde Rekonstruierbarkeit der Kausalkette für die Fehlerzuschreibung anführt. Wie sie allerdings selbst hervorhebt, stellen sich ähnliche Probleme bereits beim Phänomen des kollektiven Handelns von Menschen. Für mich unklar bleibt, inwiefern die Tatsache, dass es sich um ein »maschinelles ›Gegenüber‹« (Beck, S. 233) handelt, die für das kollektive Handeln entwickelten Konzepte der Verantwortungszuschreibung unbrauchbar macht. Die Zuschreibung von Verantwortung im Fall von Kollektiven, die Tiere und Menschen umfassen, könnte hier Vorbild sein.

Hinsichtlich der Frage nach Definition und Nutzen einer Roboterethik (Frage II) zeigen die Beiträge, dass es diesbezüglich immer noch erheblichen Klärungsbedarf gibt. In Anlehnung an Loh kann man die Disziplin der Roboterethik allgemein als mit den moralischen Fragen befasst sehen, die sich beim Bau von Robotern, im Umgang mit ihnen und im Verhältnis von Mensch und Robotern ergeben. An der Schnittstelle zwischen Roboterethik und Computerethik befindet sich die Maschinennethik, die sich mit der Frage beschäftigt, ob man Maschinen mit der Fähigkeit zum moralischen Entscheiden und Handeln ausstatten kann, wie dabei vorzugehen ist und welche Moral man zugrunde legen sollte (Moral bezieht sich hier auf Handlungsvorgaben, in dem von Hilgendorf erläuterten Sinn). Diese Verwendungsweise der Begriffe »Maschinen-« und »Roboterethik« ist im angelsächsischen Raum gang und gäbe, wo die Debatte ihren Ursprung hat.

Nicht sinnvoll ist es, die Maschinennethik als eine Oberdisziplin der Roboterethik zu begreifen, weil der Begriff »Maschine« eine allgemeinere Kategorie darstellt als »Roboter«. Das beruht auf einer fehlgeleiteten Übertragung der aristotelischen Definitionsmethode nach *genus proximum et differentia specifica* auf die angewandte Ethik. Denn es gibt keine Disziplin der Ethik, die sich mit den moralischen Problemen von Dampfmaschinen oder mechanischen Schreibmaschinen befasst. Wichtig ist, wie ich in meinem Beitrag dargelegt habe, dass eine neue Technologie eine Reihe spezifischer moralischer und gesellschaftlicher Fragen aufwirft, um eine eigene Disziplin der angewandten Ethik zu etablieren.

Ein weiterer Punkt, den die Beiträge verdeutlichen ist, dass ein Dissens darüber herrscht, ob Roboter moralische Akteure sein können oder nicht. Viel hängt hier an der Art und Weise, wie man die Terminologie verwendet. Ich erachte es als theoretisch und praktisch nützlich, einen graduellen Handlungsbegriff zu vertreten. Demnach ist Handeln keine Angelegenheit von alles oder nichts, sondern kommt in Abstufungen vor. So können Maschinen (und auch Tiere) handeln, sofern sie einen ge-

wissen Grad an Selbststurzprünglichkeit besitzen: (1) sie interagieren mit der Umwelt; (2) sie besitzen eine gewisse Unabhängigkeit von der Umwelt, die darin besteht, dass der eigene Zustand ohne direkte Einwirkung der Umwelt verändert werden kann; sowie (3) sie verfügen über Anpassungsfähigkeit, d.h. sie können die eigenen Verhaltensregeln im Licht neuer Umweltbedingungen verändern.

Darüber hinaus spreche ich Maschinen die Fähigkeit zum intentionalen Handeln in einem *funktionalen* Sinn zu, sofern sie über innere Zustände verfügen, die funktional äquivalent mit Gründen sind. Handelt es sich bei diesen Gründen um moralische Gründe, dann haben sie die Fähigkeit zum moralischen Handeln in einem funktionalen Sinn. Da diese Gründe allerdings nicht über intrinsische Intentionalität, sondern nur über abgeleitete Intentionalität verfügen, bezeichne ich sie als ›quasi-intentional‹.⁹

Mit Zoglauer stimme ich gleichwohl dahingehend überein, dass Maschinen keine *vollumfänglichen moralischen Akteure* sind, die Menschen gleichgestellt sind. Sie können nicht im vollumfänglichen Sinn moralisch handeln, weil sie nicht über Willensfreiheit, phänomenales Bewusstsein und die Fähigkeit zur Reflexion verfügen. Der Mangel an diesen Eigenschaften, die wesentlich zum Begriff der moralischen Person dazugehören, führen auch dazu, dass Maschinen selbst nicht Träger moralischer Ansprüche gegen uns sind, was uns zu Frage III führt:

Maschinen verfügen derzeitig weder über die vollumfänglichen kognitiven Fähigkeiten wie Willensfreiheit und Reflexionsvermögen noch über phänomenales Bewusstsein, weshalb sie – anders als Tiere – keinen Schmerz empfinden können. Aus diesem Grund sind sie keine Personen und haben keine moralischen Ansprüche gegenüber Menschen. Allerdings können Menschen zu unbelebten Objekten, auch Maschinen, eine emotionale Beziehung aufbauen. Hilgendorf geht deshalb davon aus, im Fall der vorsätzlichen Beschädigung eines ›robotischen Gefährten, zu dem man eine enge emotionale Beziehung aufgebaut hat‹, (S. 248) dürfte die Auffassung, es ›handele sich lediglich um eine Sachbeschädigung, [...] wenig überzeugend wirken.‹ (ebd.) Das scheint mir nicht unmittelbar zu folgen, da wir ja auch jetzt schon emotionale Bindungen zu unbelebten Objekten (am intensivsten wohl zu Puppen, aber manche lieben auch ihr Auto) entwickeln und trotzdem nicht davon ausgehen, es läge mehr als eine Sachbeschädigung vor. Die emotionale Bindung könnte aber im Kontext entsprechender rechtlicher Regelungen vielleicht dazu führen, dass für die durch die Beschädigung entstehenden emotionalen Beeinträchtigungen Schadensersatz an den Besitzer des Roboters zu leisten ist.

Ein wichtiges Argument ist meines Erachtens, dass wir Empathie mit Robotern empfinden, die einen bestimmten Grad an Menschenähnlichkeit besitzen. Es gibt gute Gründe für die Annahme, dass unser Verhalten ihnen gegenüber moralischen Ein-

9 Vgl. Catrin Misselhorn: *Grundfragen der Maschinennethik*, Stuttgart 2018.

schränkungen unterliegt, auch wenn sie kein Bewusstsein oder echte Gefühle besitzen. So ist es immer wieder ein Ansatz von Science-Fiction-Filmen, Empathie in uns zu erzeugen, wenn menschenähnliche Roboter missbraucht und erniedrigt werden. Diese Szenen rufen typischerweise das moralische Urteil hervor, dass das falsch ist. Ich halte dieses Urteil grundsätzlich für zutreffend, weil sich in einem solchen Verhalten eine unmoralische Einstellung manifestiert und es droht, unsere moralische Empfindungsfähigkeit zu beeinträchtigen, die ein wichtiges Element unserer moralischen Alltagspraxis ist.¹⁰ Allerdings schlägt sich nicht jede moralische Einschränkung unmittelbar im Recht nieder und das ist auch gut so. Deshalb halte ich es für moralisch zweifelhaft, ein solches Verhalten stets rechtlich zu sanktionieren, auch wenn es unmoralisch ist. Schließlich landet auch nicht jeder, der lügt oder seinen Partner betrügt, gleich im Gefängnis, obwohl es sich um moralische Verfehlungen handelt.

Ein aus meiner Sicht bedeutsamer Punkt, der bislang nicht in der Debatte auf-taucht, ist, dass bestimmte Einsatzmöglichkeiten von Robotern möglicherweise die Persönlichkeitsrechte von Menschen verletzen, etwa wenn sie von Robotern – z.B. am Telefon – getäuscht werden, die vorgeben echte Personen zu sein. Ähnliches gilt für Fragen der Privatsphäre und informationellen Selbstbestimmung, die häufig mit der Nutzung von Robotern einhergehen. Die Debatte um die Persönlichkeitsrechte sollte deshalb am bestmöglichen Schutz dieser Rechte für den Menschen ausgerichtet sein. Die Diskussion um Roboterrechte halte ich für praktisch nicht zielführend, wenn es um die Ausgestaltung eines aktuellen rechtlichen Rahmens geht.

Die von Zoglauer geäußerte Annahme: »Wenn Roboter aber keine natürlichen Personen sind, dann können sie auch nicht als rechtliche Personen betrachtet werden«, (S. 245-246) scheint mir rechtlich unzutreffend zu sein. Unternehmen beispielsweise sind keine natürlichen Personen, können aber rechtliche (juristische) Personen sein, z.B. eine GmbH. Zwar umfassen sie natürliche Personen als MitarbeiterInnen oder GesellschafterInnen. Gleichwohl sind sie mehr als die Summe der beteiligten natürlichen Personen, weil Unternehmen manchmal auf eine Art und Weise handeln können, die nicht auf die Überzeugungen, Wünsche oder Intentionen ihrer individuellen Mitglieder zurückführbar ist, ja diesen sogar widerspricht. Den Bogen zur Ausgangsfrage (I) zurückslagend könnte man deshalb eine Analogie zwischen dem Handeln kollektiver Akteure (wie Unternehmen) und dem Handeln von Robotern sehen, die den Ansatz fruchtbar erscheinen lässt, Robotern ebenfalls einen rechtlichen Personenstatus zuzuschreiben, wenngleich sie keine natürlichen Personen sind. Eine solche Konstruktion darf aber keinesfalls dazu dienen, um Unternehmen oder andere Akteure mit dem Verweis auf die Roboterperson in unzuläss-

10 Vgl. Catrin Misselhorn: »Is Empathy with Robots Morally Relevant?«, erscheint in: Catrin Misselhorn und Maike Klein (Hg.): *Emotional Machines. Perspectives from Affective Computing and Emotional Human-Machine-Interaction*, Wiesbaden 2019.

siger Art und Weise aus der Verantwortung zu entlassen und so aus einer möglichen ›Verantwortungslücke‹ eine ›Haftungslücke‹ entstehen zu lassen.

Bruno Gransche

Bei der Diskussion um Mensch-Technik-Relationen werden häufig zwei Ebenen folgenreich vermengt: zum einen die Ebene des tatsächlichen Standes der Technik und zum anderen die Ebene der menschlichen Bezugnahme auf erstere, also die Ebene der Vorstellungen, Projektionen, Wahrnehmungen, Interpretationen etc. Auf erster Ebene gilt es zu analysieren, was den jeweiligen Assemblagen, die etwa mit dem Wort *Roboter* bezeichnet werden, tatsächlich an Strukturen, Vernetzungen, Zuständen, Prozessieren, Input- und Output-Arten usw. zukommt. Fasst man die Definition des Roboters eher weit – wie etwa jegliches Ensemble von Sensoren, Prozessoren und Aktuatoren –, so fallen erheblich mehr Entitäten unter den Begriff als in engeren Definitionen, etwa wenn der Aspekt der Automobilität oder die Unabhängigkeit von kabelgebundener Energie- oder Informationsversorgung hinzudefiniert würde.

Unabhängig von der Definitionsweite jedoch muss auf einer zweiten Ebene unterschieden werden, was einem Roboter in Relation zu Menschen von Letzteren zugeschrieben wird. Sowohl die Vorstellungen von Menschen über Roboter als auch Aspekte, die sich aus der Relation zu Menschen ergeben, können nicht ohne Weiteres entgegen der definitionsrelativen Analyse auf erster Ebene unterstellt werden. Dass eine Kuscheldecke (ich wähle ein Low-Tech-Artefakt um des Kontrastes willen) einem Kind Trost und Geborgenheit spenden kann, bedeutet eben nicht, dass die Kuscheldecke tatsächlich des Tröstens oder der Fürsorge fähig wäre; dass Menschen sich von Gottheiten auserwählt und geliebt fühlen, bedeutet eben nicht, dass diese Gottheiten tatsächlich des Vorzuges und der Liebe fähig wären; dass Menschen vielfältigster Technik Intentionen, Volitionen oder Emotionen zuschreiben, bedeutet eben nicht, dass Technik tatsächlich Bewusstseinszustände, Handlungs-, Verantwortungs- und Rechtsfähigkeit besitzt. Insofern sind Vorstellungen und Zuschreibungen, die die Robotik inspiriert, relevant für Aussagen über Menschen. So lernt der Mensch über sich im Kontrast zu Nichtmenschen (hier: Robotern), allerdings ohne, dass eine Übertragbarkeit von Mensch auf Maschine daraus folgte: maschinelles Lernen z.B. *stellt* nicht menschliches Lernen *dar* (Hilgendorf, S. 235); beides sind gänzlich verschiedene Prozesse (was sich u.a. am Abrufen und Vergessen des Gelernten zeigt), die sich metaphorisch eine Vorstellung teilen. Anthropomorphe Roboter, oder solche, die etwa Emotionen oder Intentionen simulieren, lösen andere Zuschreibungen aus und bieten andere Relationen an, als etwa ein Montageroboter. Auf zweiter Ebene sind die *Vorstellungen* nahe, »humanoide Maschinen [...] müssten eigene Rechte erhalten« (Hilgendorf, S. 248) oder die, Roboter seien »moral agents«

(Loh, S. 238). Auf erster Ebene können dies eben auch »falsche Vorstellungen« und Irrtümer (Zoglauer, S. 245, 246) sein. Auf solchen Vorstellungen beruhen häufig sachstandsferne Ansprüche, wie die Forderung, »Maschinen müssten auch ethische Entscheidungen treffen können« (Misselhorn, S. 241). Dass Menschen aber in der Lage sind, so verschiedenste Relationen mit Robotern einzugehen, ist mit Blick auf den Menschen lehrreich, sollte eine korrekte Analyse der Technikseite aber nicht verwirren: Eine Kind-Kuscheldecke- oder Mensch-Smartphone-Relation kann korrekterweise als eine des Trostes oder der Sicherheit charakterisiert werden, ohne dass Decke und *device* dadurch die Fähigkeit bekämen, (im achtsamkeitsethischen Sinne) fürsorglich zu trösten oder Sicherheit zu vermitteln. Von der Tatsache, ob und ab wann Entitäten, die jeweils mit dem Begriff ›Roboter‹ adressiert werden, Reflexionsvermögen, Bewusstseinszustände, moralische Autonomie etc. tatsächlich technikseitig und nicht nur vermittels der Mensch-Relation-Seite zuzuschreiben wäre, hängt ab, ob neuartige metaphysische Befunde und Probleme zu verhandeln sein werden. Einen solchen Stand der Technik hat die Robotik nicht erreicht, von künstlichen postbiotischen Bewusstseinen sind wir meines Erachtens auf lange Sicht noch weit entfernt. Entsprechende Zuschreibungen durch anthropomorphe Gestalt und Verhaltenssimulation gekonnt anzuregen, gelingt – wie die lebhafte Debatte zeigt – bereits heute ausgezeichnet und die entsprechende Simulations- bzw. Inszenierungsfähigkeit auf Systemseite wird heute wie in naher Zukunft immer wieder verblüffen. Von Befunden der zweiten Ebene auf solche der ersten Ebene zu schließen, von Vorstellungen ›robotischer Weltherrschaft‹ (Hilgendorf, S. 235), eines ›KI-Supergau‹ (Zoglauer, S. 246) oder Intentionalitätsunterstellungen etwa ein Verbot für Roboter oder eine Pflicht zu Absichtsbekundungen *on demand* zu schließen, wäre kaum weniger absurd, als eine Kuscheldecke wegen Fremdkuschelns zu rügen oder die Kuscheldeckenbeziehung als solche infrage zu stellen, weil die Decke nie den ersten Schritt macht (*ultra posse*).

Zur Frage der Handlungsfähigkeit: Ob Roboter handlungsfähig sind, hängt von beiden Definitionsbereichen ›Roboter‹ sowie ›Handlung‹ ab. Je nach Definitionswelt kann – bei entsprechendem Konsens – jegliches Prozessieren, Agieren oder Verhalten auch als ›Handeln‹ bezeichnet werden. Ein solcher Gebrauch würde aber wichtige Unterschiede nivellieren, derer es bedarf, um Mensch-Technik-Relationen mit adäquater Begriffsschärfe zu differenzieren. Meiner Ansicht nach wäre zwischen *herbeigeführten* und *nicht herbeigeführten* Ereignissen zu differenzieren, wobei letztere als ›Vorgänge‹ und unter ersteren *nicht beabsichtigte* als ›Verhalten‹ und *beabsichtigte* als ›Handeln‹ zu verstehen wären. Als absichtliche Herbeiführung eines Ereignisses setzt Handeln Intentionalität, Zwecksetzungsautonomie und Reflexionsfähigkeit über Tauglichkeit (der Mittel) und mögliche Erreichbarkeit (der Zwecke) in Abhängigkeit zur eigenen Kompetenz und den situativen Realisierungsbedingungen voraus. Zumindest Intentionalität, Autonomie und Reflexion im philosophischen

Sinne kommt Robotern nicht zu, weshalb ihnen zwar technisches Prozessieren und ›Agieren‹ im Sinne eines automatisierten Vorganges zu unterstellen wäre, aber sicher keine Handlungsfähigkeit in diesem Sinne (Zoglauer, S. 236). Inwieweit dieses automatisierte Prozessieren als nichtbeabsichtigtes Herbeiführen (Verhalten) zu verstehen wäre oder eher als systemische Kopplung und spezifische Konstellation einer Weiterleitung von Ursache-Wirkungsketten, hängt wiederum vom Konzept des ›Herbeiführens‹ ab. Ich denke eher Letzteres und Thomas Zoglauer scheint diesen Kopplungsgedanken zu stützen, wenn er auf die Voraussetzung von natürlichen Personen *hinter* elektronischen wie juristischen Personen verweist. Meines Erachtens könnte dennoch die elektronische Person ein pragmatischer rechtlicher Fortschritt zur Lösung von Haftungsfragen darstellen, eine Rückführung auf natürliche Personen (ggf. pauschal oder in anonymer Vergemeinschaftung) ist rechtlich sicher möglich. Ein Roboter wäre deshalb nicht rechtsfähig, sondern zunächst regulativ eine vermittelnde Instanz zwischen anonymen Rechtsparteien. Auch stellt sich die Herausforderung der »Verantwortungsdiffusion« (Beck, S. 232) verschärft durch den Befund der *Medialität der Technik* (»Verlust der Spuren«),¹¹ die zwar in der Tat keine »Verantwortungslücke« (Zoglauer, S. 236) wohl aber ein Vollzugsdefizit durch Zurechnungsschwierigkeiten bedeutet. Da jedenfalls die Intention für Verantwortungszuschreibung und da Zurechenbarkeit des Herbeiführens (oder Unterlassens) für Strafbarkeit maßgeblich sind, hängen Verantwortungs- und Rechtsfähigkeit an der Diagnose (erster Ebene) der Handlungsfähigkeit, die ihrerseits aus dem tatsächlichen Vorhandensein von Intentionalität und Moralfähigkeit/moralischer Autonomie resultiert. Die Frage, »ob Roboter moralisch handeln können« (Loh, S. 235), wäre meines Erachtens aus doppelter Hinsicht zu verneinen und abgeleitete Regeln für solches Handeln entsprechend hypothetisch: Roboter handeln nicht und erst recht nicht moralisch. Sehr wohl aber beeinflussen Roboter mögliche Mensch-Technik-Relationen anders als Werkzeuge oder Computer, was neue, auch moralische Probleme erzeugen kann, die sich aber dem Menschen, nicht dem Roboter stellen.

Entsprechend muss und kann es keine Roboterethik geben, die moralisches Handeln von Robotern verhandelt, vielmehr ist sie im Sinne von Hubigs »provisorischer Moral«¹² als Ethik der Entwicklung und Gestaltung robotischer Systeme, inklusive deren Implementierungskontexte sowie Options- und Vermächtniswerte notwendig (Zoglauer, S. 240). Mögliche Fragen dieser Bereichsethik wären etwa: Gibt es Arten von Sensoren und Algorithmen (die beispielsweise *by design* Minderheiten diskriminieren) oder von Aktuatoren (z.B. tödliche Waffen), die als moralisch fragwürdig von der Entwicklung ausgeschlossen werden sollten? Oder: Gefährdet die lückenlo-

11 Christoph Hubig: *Die Kunst des Möglichen. Grundlinien einer dialektischen Philosophie der Technik*, Band 1: *Technikphilosophie als Reflexion der Medialität*, Bielefeld 2006, Kapitel 5: Medialität der Technik: Strukturierte Möglichkeitsräume als System.

12 Christoph Hubig: *Die Kunst des Möglichen. Grundlinien einer dialektischen Philosophie der Technik*, Band 2: *Ethik der Technik als provisorische Moral*, Bielefeld 2007.

se Alltagsdurchdringung mit assistiven Robotern unter Umständen die Bedingungen menschlichen Kompetenzerwerbes?

