



Roboterethik

Catrin Misselhorn

- › Autonome Maschinen (Computer, Roboter) sind in einem eingeschränkten Sinne moralisch relevante Akteure. Entscheidungen können an sie delegiert werden, moralische Verantwortung tragen sie jedoch nicht. Dazu fehlen ihnen Willensfreiheit, Kausalität, Absichtlichkeit und Wissen.
- › In bestimmten Situationen können autonome Systeme die Verantwortungszuschreibung unterminieren. Es wird unklar, wer Verantwortung trägt. Es entsteht eine „Verantwortungslücke“, in der weder Programmierer noch Anwender die volle Verantwortung zugeschrieben werden kann.
- › Besonders wenn es um menschliches Leben geht, sollten Entscheidungen nicht Maschinen überlassen werden. Darüber hinaus gibt es viele nützliche Einsatzmöglichkeiten für Roboter.
- › Für die Roboterethik lassen sich drei Leitlinien festhalten: (1) Künstliche Systeme sollten stets die Selbstbestimmung von Menschen fördern und nicht beeinträchtigen. (2) Sie sollten nicht über Leben und Tod von Menschen entscheiden. (3) Es muss sichergestellt werden, dass Menschen stets die Kontrolle ausüben und Verantwortung übernehmen.

Inhaltsverzeichnis

1. Einordnung der Roboterethik.....	2
2. Motivation der Maschinenethik.....	3
3. Moralisches Handeln und Verantwortung.....	4
4. Entscheidungen über Leben und Tod.....	6
5. Konklusion.....	7
Literaturverzeichnis	9
Impressum	11

Die Robotik ist eine vielversprechende Wachstumsbranche. So lag der Zuwachs beim Verkauf von Robotern nach Angaben des *World Robotics Report* 2018 im vergangenen Jahr bei ca. 30 Prozent im Vergleich zum Vorjahr und der Trend ist steigend.¹ Das bringt neben technischen, wirtschaftlichen und rechtlichen Herausforderungen auch ethische Probleme mit sich. Damit beschäftigt sich die Roboterethik, eine neue Disziplin an der Schnittstelle von Informatik, Philosophie und Robotik.² Sie hat die moralischen Probleme bei der Entwicklung, Herstellung und Verwendung von Robotern sowie das Verhältnis von Mensch und Roboter zum Gegenstand und betrachtet die gesellschaftlichen Konsequenzen der zunehmenden Roboterisierung.

Unter einem Roboter wird in diesem Beitrag eine elektromechanische Maschine verstanden, die aus einem Prozessor, Sensoren und Effektoren besteht.³ Der Begriff der Ethik wird gleichbedeutend mit Moralphilosophie gebraucht, im Sinn einer philosophischen Disziplin, die sich mit der moralischen Qualität von Handlungen, Urteilen, Charakterzügen und Einstellungen, Regeln sowie Institutionen befasst. Dabei steht die *normative* Ethik im Vordergrund, der es nicht nur um die Beschreibung dessen geht, was Menschen für moralisch richtig halten, sondern darum, begründete Empfehlungen abzugeben, was moralisch richtig und falsch ist.⁴

1. Einordnung der Roboterethik

Die spezifischen ethischen Fragen, die Roboter aufwerfen, haben mit zwei Aspekten zu tun: Zum einen mit ihrer zunehmenden Intelligenz und Autonomie und zum anderen mit ihrer äußeren Gestalt und Interaktionsweise mit Menschen. Der erste Aspekt führt zur *Maschinenethik*, einer Subdisziplin der Roboterethik. Hier geht es um die Entwicklung einer Ethik für Maschinen im Gegensatz zu einer *Ethik für Menschen im Umgang mit Maschinen*. Man spricht in Analogie zu „Artificial Intelligence“ (AI) auch von „Artificial Morality.“⁵ Während „Artificial Intelligence“ zum Ziel hat, die kognitiven Fähigkeiten von Menschen zu modellieren oder zu simulieren, geht es bei der „Artificial Morality“ darum, künstliche Systeme mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten. Das ist natürlich nicht beliebigen Maschinen zuzutrauen, sondern nur Computern. Die Idee ist, Computer so zu programmieren, dass sie moralisch handeln können. Die Software bildet dabei so etwas wie das „Gehirn“ des Roboters. Die Sensoren entsprechen den Sinnesorganen, die Informationen über die Umwelt und den Zustand des Roboters bereitstellen. Mit Hilfe der Effektoren kann der Roboter auf die Umwelt einwirken und seine Position verändern.

Der zweite Aspekt ergibt sich eher aus der äußeren Gestalt von Robotern und ihrer Interaktion mit Menschen. Man kann auch von *Ethik der Mensch-Maschine-Interaktion* sprechen. Menschen neigen dazu, intelligente Systeme zu vermenschlichen. Wir nehmen Roboter, die

Maschinenethik

Ethik der Mensch-
Maschine-Interaktion

sich anscheinend autonom und intelligent verhalten und womöglich eine menschen- oder tierähnliche Gestalt aufweisen, allzu leicht als Wesen mit vergleichbaren Gedanken, Motiven und Empfindungen wie Menschen wahr. Damit sind verschiedene Fragen verbunden.

Zum einen geht es etwa darum, ob der Interaktion mit solchen Robotern eine problematische Form der Täuschung oder Manipulation zugrunde liegt, weil sie uns so anmuten, als hätten sie menschliche Eigenschaften, die ihnen gar nicht zukommen.⁶ Zum anderen ist zu fragen, ob Maschinen, die solche Reaktionen bei Menschen hervorrufen, uns auch gewisse moralische Einschränkungen im Umgang mit ihnen auferlegen, auch wenn sie nicht wirklich über die entsprechenden Gedanken, Motive oder Empfindungen verfügen.⁷ Maschinenethik und Ethik der Mensch-Maschine-Interaktion überlappen sich, wenn es darum geht, ob Maschinen, die die Fähigkeit zum moralischen Entscheiden und Handeln haben, ihrerseits auch Subjekt moralischer Ansprüche sind, die Menschen im Umgang mit ihnen zu berücksichtigen haben. Der vorliegende Beitrag konzentriert sich auf den Aspekt der Maschinenethik.⁸ Diese wird in einem offenen Brief, den eine beeindruckende Zahl von KI-Forschern und Wissenschaftlern unterzeichnet hat, als eines der wichtigsten und drängendsten Forschungsgebiete der KI hervorgehoben.⁹

2. Motivation der Maschinenethik

Es gibt viele Arbeiten, die wir gerne Maschinen überlassen würden, weil sie zu schwer, gefährlich oder einfach unangenehm sind. In manchen Fällen gibt es auch nicht genügend Menschen, die sie übernehmen können und wollen. Schließlich können Maschinen manche Dinge auch einfach besser und schneller erledigen als Menschen. In vielen Fällen erfordert das, dass die Maschinen möglichst eigenständig agieren, d. h. ohne dass ein Mensch direkt kausal eingreift. So stellt ein Roboterstaubsauger gerade deshalb eine Arbeiterleichterung dar, weil er nicht von einem Menschen geführt muss, sondern selbständig in der Wohnung herumfährt, wenn man gerade nicht vor Ort ist. Diese Fähigkeit, sich selbständig zu orientieren und zu agieren erfordert intelligente Systeme.¹⁰

Selbst ein so einfaches Modell wie der Staubsaugerroboter steht nun bereits einer ethischen Entscheidung gegenüber, nämlich: Soll es einen Marienkäfer einfach einsaugen oder soll er den Käfer verscheuchen bzw. umfahren? Und wie sieht es bei einer Spinne aus? Bei der Frage, ob man ein Insekt zu Reinigungszwecken töten darf, handelt es sich um eine basale moralische Frage. Gewöhnliche Staubsaugerroboter besitzen allerdings noch nicht die Fähigkeit, eine solche Entscheidung zu treffen. Es gibt jedoch erste Ansätze, eine um ein Ethikmodul erweiterte Version des populären Modells *Roomba* zu schaffen, die das Leben von Insekten berücksichtigt (der Prototyp ist mit einem optionalen „Kill-Button“ für Spinnen ausgestattet).¹¹

Je komplexer die Einsatzbereiche autonomer Systeme sind, desto anspruchsvoller werden die moralischen Entscheidungen, die sie treffen müssen. Ein wichtiger Anwendungsbereich für moralische Maschinen ist etwa die Altenpflege. Aufgrund des demographischen Wandels wird der Anteil pflegebedürftiger Menschen in den nächsten Jahrzehnten stark zunehmen. Künstliche Systeme werden immer wieder als eine Möglichkeit ins Spiel gebracht, um dem Pflegenotstand entgegenzutreten. Doch Systeme, die in diesem Kontext eingesetzt werden sollen, stehen vor moralischen Entscheidungen, beispielsweise: Wie häufig und eindringlich soll ein Pflegesystem an Essen und Trinken sowie die Einnahme von Medikamenten erinnern? Wann sollte ein Pflegesystem die Angehörigen verständigen oder den medizinischen Dienst rufen, wenn jemand sich eine Zeitlang nicht rührt? Soll das System den Nutzer rund um die Uhr überwachen und wie ist mit den dabei erhobenen Daten zu verfahren?

In all diesen Situationen muss ein künstliches System zwischen bestimmten moralischen Werten abwägen: Im ersten Fall beispielsweise zwischen der Selbstbestimmung des Nutzers und bestimmten gesundheitlichen Risiken, die entstehen, wenn er seine Medikamente nicht wie vorgeschrieben einnimmt. Im zweiten Fall zwischen der Selbstbestimmung des Nutzers, der Sorge der Angehörigen, die vielleicht gerne sofort informiert würden, und erneut der Gesundheit. Im dritten Fall geht es wiederum um die Selbstbestimmung des Nutzers, Gesundheit, Sorge der Angehörigen sowie um die Privatheit seiner Daten.

Abwägung zwischen
moralischen Werten

Ein weiteres viel diskutiertes Beispiel für die Notwendigkeit moralischer Maschinen ist das autonome Fahren. Auch vollautomatisierte Fahrzeuge stehen vor moralischen Entscheidungen. So gilt es beispielsweise, sie so zu programmieren, dass in unvermeidlichen Gefahrensituationen der Schutz menschlichen Lebens Vorrang vor Sach- und Tierschäden besitzt. Doch auch Tiere sollten nach Möglichkeit verschont werden. Eine besondere Schwierigkeit stellen die in diesem Anwendungsbereich unter Umständen auftretenden moralischen Dilemmata dar. Was ist, wenn ein autonomes Fahrzeug ausschließlich die Möglichkeit hat, entweder einen Menschen am Ende seines Lebens oder ein kleines Kind zu töten? Was, wenn es nur dadurch fünf Menschenleben retten kann, dass es eine auf dem Gehweg stehende Person anfährt? Ist ein besonderer Schutz für die Insassen moralisch legitim oder kommt den anderen Verkehrsteilnehmern vom moralischen Standpunkt mehr Gewicht zu?¹²

Nicht zu vergessen sind schließlich die militärischen Anwendungen. Der Traum besteht darin, dass keine Soldaten mehr auf dem Schlachtfeld ihr Leben aufs Spiel setzen müssen, sondern an ihrer Stelle autonome Maschinen in den Kampf geschickt werden.¹³ Diese sollen mit dem Kriegsvölkerrecht und kontextspezifischen Einsatzregeln ausgestattet werden, die ihren Handlungsspielraum begrenzen und sicherstellen, dass sie sich rechtlich und moralisch einwandfrei verhalten. So müssen sie entscheiden, wann eine Aktion militärisch notwendig und angemessen ist und wie sich Kombattanten von Zivilisten unterscheiden lassen.

3. Moralisches Handeln und Verantwortung

Man könnte allerdings argumentieren, dass es nicht das Pflegesystem, das autonome Auto oder der Kampfroboter ist, die in diesen Fällen eine moralische Entscheidung treffen, sondern die Programmierer dieser Geräte. Doch je größer die Fortschritte der Künstlichen Intelligenz werden, desto weniger können die Entwickler planen und vorhersagen, welche Entscheidungen ein System in einer spezifischen Situation treffen wird. So spielt schon ein Schachprogramm weit besser als seine Programmierer, die nicht jeden einzelnen Zug des Systems vorhersagen können. Das gilt umso mehr für ein so komplexes System wie Alpha Go Zero, das zunächst nur die Grundregeln des Spiels kennt und dann durch das Durchspielen einer Vielzahl von Partien gegen sich selbst zu den optimalen Entscheidungsstrategien findet. In kürzester Zeit gelang es diesem System, seinen Vorgänger Alpha Go zu schlagen, der als erstes künstliches System einige der weltbesten menschlichen Go-Spieler besiegte. Doch genau dieser Mangel an Kontrolle und Vorhersehbarkeit gehört zu den Punkten, die den wichtigsten ethischen Einwänden gegen die Maschinenethik zugrunde liegen.

Mangel an Kontrolle
und Vorhersehbarkeit

Maschinen sind nicht zu vollumfänglichem moralischen Handeln in der Lage, wie es Menschen auszeichnet. Dazu fehlen ihnen unter anderem Bewusstsein, Willensfreiheit und die Fähigkeit zur Selbstreflexion.¹⁴ Da diese Eigenschaften für die Übernahme moralischer Verantwortung wesentlich sind, können Maschinen auch nicht für ihr Handeln verantwortlich gemacht werden. Trotzdem ist zu diskutieren, inwiefern der Einsatz von Maschinen die Verantwortungszuschreibung an Menschen unterminiert, so dass am Ende möglicherweise

Schwierige Verantwortungs-
zuschreibung

niemand für ihr Handeln die Verantwortung trägt. In diesem Zusammenhang ist vom Entstehen einer *Verantwortungslücke* die Rede.¹⁵

Zu den Kriterien für die Zuschreibung moralischer Verantwortung zählen Willensfreiheit, Kausalität, Absichtlichkeit und Wissen. Ein Handelnder 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 Maschinen diese Bedingungen nicht alle erfüllen. So besitzen sie keinen freien Willen; aber auch die Bedingungen der Absichtlichkeit und des Wissens werfen Probleme bei der Zuschreibung an Maschinen auf. Deshalb können sie zwar keine moralische Verantwortung tragen, aber eine Verantwortungslücke erzeugen.

Verantwortungslücke

Der australische Maschinenethiker Robert Sparrow, auf den dieser Begriff (im Original: *responsibility gap*) zurückgeht, argumentiert dafür am Beispiel autonomer Kriegerroboter. Der Kern seines Arguments lautet: Eine Verantwortungslücke entsteht, wenn: (1) ein Kriegerroboter nicht absichtlich so programmiert wurde, dass er die ethischen bzw. rechtlichen Normen der Kriegsführung verletzt; (2) es nicht vorhersehbar war, dass der Einsatz des Kriegerroboters dazu führen würde; und (3) ab dem Start der Operation keine menschliche Kontrolle mehr über die Maschine bestand.

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 den ethischen bzw. rechtlichen Normen der Kriegsführung Menschen tötet. Denn kein Mensch 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. Für Sparrow ist dies ein Grund dafür, den Einsatz von Kriegerrobotern als unmoralisch abzulehnen. Es lässt sich aber grundsätzlich auch auf andere Bereiche, insbesondere das autonome Fahren, übertragen.

Man könnte dies zum Anlass nehmen, zu fordern, der Mensch dürfe eben nicht gänzlich die Kontrolle aus der Hand geben. In militärischen Kontexten wird zwischen *In-the-Loop-Systemen*, *On-the-Loop-Systemen* und *Out-of-the-Loop-Systemen* unterschieden, je nachdem, welche Rolle der Mensch in der Kontrollschleife spielt.¹⁶ Bei *In-the-Loop-Systemen* bedient ein Mensch das System und fällt sämtliche Entscheidungen, und sei es per Fernbedienung. *On-the-Loop-Systeme* sind zwar programmiert, sie können aber in Echtzeit unabhängig von menschlichem Eingreifen operieren. Dem Menschen obliegt jedoch weiterhin die Überwachung und er hat jederzeit die Möglichkeit, zu intervenieren. *Out-of-the-Loop-Systeme* verhalten sich wie *On-the-Loop-Systeme*, doch besteht keine menschliche Kontroll- und Eingriffsmöglichkeit mehr.

Das Problem der Verantwortungslücke erscheint als gelöst, wenn der Mensch On-the-Loop bleibt und vielleicht sogar per Knopfdruck der Verantwortungsübernahme zustimmen muss, bevor er ein künstliches System in Betrieb nimmt.¹⁷ Doch wie realistisch ist die Annahme, dass der Mensch zu einer permanenten Überwachung überhaupt in der Lage ist? Kann er die Aufmerksamkeit so lange halten und ist er in Sekundenschnelle bereit, zu entscheiden und einzugreifen, wenn es darauf ankommt? Wenn das nicht der Fall ist, wären Vorhersehbarkeit und Kontrolle zwar theoretisch möglich, aber in der Realität für den Menschen nicht umsetzbar.

Außerdem stellt sich ein Erkenntnisproblem, denn der Mensch ist zur Analyse der Situation auf die Informationen angewiesen, die das System ihm liefert. Die Frage ist, ob er diese überhaupt rational in Zweifel ziehen kann, wenn er keinen Zugang zu unabhängigen Informationen besitzt. Darüber hinaus muss ein solches System bei seiner Entwicklung eine Reihe von Qualitätssicherungsprozessen durchlaufen. Auch das kann für die Nutzer ein Grund dafür sein, die Vorschläge des Systems den eigenen Zweifeln für überlegen zu halten.

Insgesamt wirkt es unfair, dass die Nutzer die volle Verantwortung per Knopfdruck übernehmen sollen, denn zumindest ein Teil der Verantwortung, wenn nicht sogar der Hauptteil, sollte doch den Programmierern zukommen, deren Algorithmen ausschlaggebend für das Handeln des Systems sind. Die Nutzer sind nur in einem schwächeren Sinn verantwortlich, weil sie das System nicht am Handeln gehindert haben. Alle drei Punkte lassen es zweifelhaft erscheinen, ob die Bedingungen der Vorhersehbarkeit und der Kontrolle erfüllt sind. Das Problem der Verantwortungslücke droht deshalb auch bei On-the-Loop-Systemen. Es stellt sich letztlich sogar dann, wenn der Mensch In-the-Loop bleibt. Dies unterstützt die Forderung von Menschenrechtsorganisationen nach einer sinnvollen menschlichen Kontrolle (*meaningful control*).¹⁸

4. Entscheidungen über Leben und Tod

Generell sollten wir uns gut überlegen, ob wir moralisches Entscheiden an Maschinen delegieren wollen, wenn es um Leben und Tod von Menschen geht. Ein entscheidendes Argument gegen autonome Waffensysteme lautet, dass es keine moralische Pflicht gibt, im Krieg zu töten.¹⁹ Es liegt nur eine moralische Erlaubnis zum Töten vor, die das allgemeine Tötungsverbot situativ einklammert. Deshalb sollte die Entscheidung, einen bestimmten Menschen zu töten, stets einem Menschen obliegen und nicht von einer Maschine getroffen werden.

Dieses Argument lässt sich auch auf das autonome Fahren übertragen. Es lässt sich eine Analogie zwischen der Programmierung autonomer Fahrzeuge zum Zweck der Unfalloptimierung und der Zielbestimmung autonomer Waffensysteme herstellen.²⁰ Um Unfallergebnisse zu optimieren, ist es notwendig, Kosten-Funktionen anzugeben, die bestimmen, wer im Zweifelsfall verletzt und getötet wird. Ganz ähnlich wie bei autonomen Waffensystemen müssten also für den Fall einer unvermeidlichen Kollision legitime Ziele festgelegt werden, die dann vorsätzlich verletzt oder womöglich sogar getötet würden.

Das setzt voraus, dass eine moralische Pflicht besteht, unschuldige Menschen zu verletzen oder zu töten, sofern dies dazu dient, Schlimmeres zu verhindern. Eine solche Pflicht stünde freilich in einem Spannungsverhältnis zur deutschen Rechtsprechung. So hat das Bundesverfassungsgericht in seiner Entscheidung zum Luftsicherheitsgesetz im Jahr 2006 zum Abschuss entführter Passagierflugzeuge, die von Terroristen als Massenvernichtungswaffen eingesetzt werden sollen, festgestellt, dass ein Abschuss immer der Menschenwürde der Flugzeugpassagiere widerspricht.²¹ Das Grundgesetz schließt aus, auf der Grundlage einer gesetzlichen Ermächtigung unschuldige Menschen vorsätzlich zu töten. Dieses Urteil steht zumindest auf den ersten Blick in einem Widerspruch zu einer Pflicht der Schadensminimierung, die die vorsätzliche Verletzung oder Tötung unschuldiger Menschen umfasst.

Es gibt allerdings Stimmen, die in dieser Sache noch nicht das letzte Wort gesprochen sehen. Ein Vorschlag besteht darin, von einer Abstufung im Unrecht auszugehen.²² Demnach ist die Tötung unschuldiger Menschen rechtswidrig (und wohl auch unmoralisch), gleichwohl soll es eine rechtliche und moralische Pflicht geben, so wenige Leben wie möglich zu vernichten, und zwar auch dann, wenn dabei vorsätzlich unschuldige Menschen verletzt oder getötet

Grenzen für das
Delegieren
moralischer
Entscheidungen

werden. Allerdings scheint eine solche Pflicht bislang in der Rechtsdogmatik nicht allgemein anerkannt zu sein. Zudem ist diese Konstruktion aus moralischer Sicht fragwürdig. Weil die Verletzung und Tötung unschuldiger Menschen weiterhin ein Unrecht darstellt, haben die Betroffenen nämlich Abwehrrechte. So dürfen sie ein Flugzeug abschießen, das auf sie niederzugehen droht, wenn diese Option das geringste Übel darstellt. Das würde aber bedeuten, dass jemand eine moralische Pflicht hat, etwas zu tun, während ein anderer das moralische Recht hat, genau dies zu unterbinden.²³ Das scheint mit der Unbedingtheit und Verallgemeinerbarkeit, die moralische Pflichten auszeichnen, schlecht vereinbar zu sein.

5. Konklusion

Auch wenn man die Entscheidung über Leben und Tod von Menschen nicht an Maschinen abgeben möchte, bleiben viele Anwendungsbereiche, in denen moralische Maschinen sinnvoll eingesetzt werden können, beispielsweise in der Pflege.²⁴ Dabei muss darauf geachtet werden, dass das menschliche Recht auf Selbstbestimmung gewahrt bleibt. Das betrifft einerseits die Entscheidung, ob jemand überhaupt von einem künstlichen System gepflegt werden möchte; dies sollte jedem freigestellt sein. Andererseits ist darauf zu achten, dass ein Pflegesystem sich flexibel auf die Wertvorstellungen seiner Nutzer einstellen kann. Denn in modernen pluralistischen Gesellschaften ist davon auszugehen, dass sich die Wertvorstellungen der Nutzer darin unterscheiden, ob beispielsweise der Privatsphäre mehr Gewicht zuzumessen ist oder der Vermeidung gesundheitlicher Risiken. Ein Pflegesystem sollte in der Lage sein, diese Abwägung nach den moralischen Standards des jeweiligen Nutzers individuell vorzunehmen. In diesem Fall kann ein Pflegesystem Menschen, die dies möchten, dazu verhelfen, länger selbstbestimmt in ihren eigenen vier Wänden zu leben. Ein solches System ist allerdings nur für Menschen geeignet, die kognitiv in der Lage sind, grundlegende Entscheidungen über ihr Leben zu treffen, aber körperlich so eingeschränkt sind, dass sie nicht ohne Pflege allein zu Hause leben können.

Abschließend kann man drei Prinzipien als grundlegende Leitlinien für die Roboterethik festhalten:

1. Künstliche Systeme sollten stets die Selbstbestimmung von Menschen fördern und sie nicht beeinträchtigen.
2. Sie sollten nicht über Leben und Tod von Menschen entscheiden.
3. Es muss sichergestellt werden, dass Menschen stets die Kontrolle ausüben und Verantwortung für das Handeln der Maschinen übernehmen.

- 1 <https://ifr.org/ifr-press-releases/news/industrial-robot-sales-increase-worldwide-by-29-percent> (letzter Aufruf: 20.12.18).
- 2 Der Begriff der Roboterethik geht zurück auf Veruggio (2005), der als offizieller Begründer dieser Disziplin gilt
- 3 Vgl. Misselhorn (2013).
- 4 Für eine nähere Erläuterung, was unter moralisch in diesem Zusammenhang zu verstehen ist, vgl. Misselhorn (2018).
- 5 Vgl. Misselhorn (2018): Artificial Morality.
- 6 Vgl. Scholz (2008) spricht von „subjektsimulierenden Maschinen“, während Turkle (2006) die Roboter als relationale Artefakte definiert.
- 7 Misselhorn 2009, 2018.
- 8 Zum zweiten Aspekt vgl. Misselhorn et al. (2013).
- 9 <https://futureoflife.org/ai-open-letter> (letzter Aufruf: 20.12.18).
- 10 Misselhorn (2015).
- 11 Vgl. Bendel (2017).
- 12 Auf der Webseite <http://moralmachine.mit.edu/> (letzter Aufruf: 20.12.18) finden sich eine ganze Reihe solcher Szenarien mit unterschiedlichen Konstellationen, über die der Besucher per Mausklick entscheiden kann. Die Daten werden u.a. genutzt, um ein Entscheidungsverfahren für moralische Maschinen zu entwickeln (Awad et al. 2018).
- 13 Vgl. Arkin (2009).
- 14 Vgl. Misselhorn (2018).
- 15 Vgl. Sparrow (2007).
- 16 Vgl. United States Department of Defense (2011).
- 17 Eine solche Konzeption schlägt beispielsweise Arkin (2009) für das von ihm entwickelte autonome Waffensystem vor.
- 18 Vgl. Roff und Moyes (2016).
- 19 Misselhorn (2018).
- 20 Vgl. Lin (2016), S. 72.
- 21 Vgl. BVerfGE 115, 118, (160).
- 22 Vgl. Hilgendorf (2017), S. 155.
- 23 Vgl. Hilgendorf (2017), S. 155.
- 24 Vgl. Misselhorn (2019): Moralische Maschinen in der Pflege.

Literaturverzeichnis

- A** Arkin, Ronald (2009): *Governing lethal behavior in autonomous robots*. Boca Raton.
- Awad, Edmond et al. (2018): *The Moral Machine experiment*. In: *Nature* 563, S. 59–64.
- B** Bendel, Oliver (2017): *Ladybird – The animal-friendly robot vacuum cleaner*. In: *The AAAI 2017 Spring Symposium on Artificial Intelligence for the Social Good Technical Report SS-17-01*. Palo Alto, S. 2–6.
- H** Hilgendorf, Eric (2017): *Autonomes Fahren im Dilemma - Überlegungen zur moralischen und rechtlichen Behandlung von selbsttätigen Kollisionsvermeidensystemen*. In: Eric Hilgendorf: *Autonome Systeme und neue Mobilität - Ausgewählte Beiträge zur 3. und 4. Würzburger Tagung zum Technikrecht*. Baden-Baden, S. 143–176.
- L** Lin, Patrick (2016): *Why ethics matters for autonomous cars*. In: Markus Maurer und J. Christian Gerdes und Barbara Lenz et al. (Hrsg.): *Autonomous driving – Technical, legal and social aspects*. Berlin/Heidelberg, S. 69–85.
- M** Misselhorn, Catrin (2009): *Empathy with inanimate objects and the uncanny valley*. In: *Minds and Machines* 19, S. 345–359.
- Misselhorn, Catrin (2013): *Robots as moral agents*. In: Frank Rövekamp und Friederike Bosse (Hrsg.): *Ethics in science and society - German and Japanese views*. München, S. 30–42.
- Misselhorn, Catrin (Hrsg.) (2015): *Collective agency and cooperation in natural and artificial systems - Explanation, implementation and simulation*. In: *Philosophical Studies Series* 122.
- Misselhorn, Catrin (2018): *Grundfragen der Maschinenethik*. Ditzingen 2018.
- Misselhorn, Catrin (2019): *Is empathy with robots morally relevant?* In: Catrin Misselhorn und Maike Klein (Hrsg.): *Emotional machines – Perspectives from affective computing and emotional human-machine interaction*. Wiesbaden. (Erscheint 2019).
- Misselhorn, Catrin und Pompe, Ulrike und Stapleton, Mog (2013): *Ethical considerations regarding the use of social robots in the fourth age*. In: *GeroPsych – The Journal of Gerontopsychology and Geriatric Psychiatry* 26, S. 121–133.
- R** Roff, Heather M. und Moyes, Richard (2016): *Meaningful human control, artificial intelligence and autonomous weapons*. Briefing paper prepared for the informal meeting of experts on lethal autonomous weapons systems. UN Convention on Certain Conventional Weapons. Geneva.
- S** Scholz, Christopher (2008): *Alltag mit künstlichen Wesen. Theologische Implikationen eines Lebens mit subjektsimulierenden Maschinen am Beispiel des Unterhaltungsroboter Aibo Göttingen 2008*
- Sparrow, Robert (2007): *Killer robots*. In: *Journal of Applied Philosophy* 24, S. 62–77.

- T** Turkle, S., Taggart, W., Kidd, C., & Daste, O. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Science*, 18, 347–361.
- U** United States Department of Defense (2011): Unmanned systems integrated roadmap FY 2011-2036. Reference Number 11-S-3613. URL: <<http://www.dtic.mil/dtic/tr/fulltext/u2/a558615.pdf>> (letzter Aufruf: 20.12.18).
- V** Veruggio, Gianmarco (2005): The birth of roboethics. In: Proceedings of IEEE International Conference on robotics and automation. Genova, S. 1–4.

Impressum

Die Autorin

Prof. Dr. Catrin Misselhorn

Direktorin des Instituts für Philosophie, Lehrstuhl für Wissenschaftstheorie und Technik-
philosophie, Universität Stuttgart

Konrad-Adenauer-Stiftung e. V.

Dr. Norbert Arnold

Leiter des Teams Bildungs- und Wissenschaftspolitik

Hauptabteilung Politik und Beratung

T: +49(0)30 / 26 996-3504

norbert.arnold@kas.de

Postanschrift: Konrad-Adenauer-Stiftung, 10907 Berlin

Herausgeberin: Konrad-Adenauer-Stiftung e. V. 2018, Sankt Augustin/Berlin

Gestaltung: yellow too Pasiek Horntrich GbR

Satz: Janine Höhle, Konrad-Adenauer-Stiftung e.V.

Lektorat: Jenny Kahlert, Konrad-Adenauer-Stiftung e.V.

ISBN 978-3-95721-500-0



Der Text dieses Werkes ist lizenziert unter den Bedingungen von „Creative Commons
Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 international“, CC BY-SA 4.0
(abrufbar unter: <https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>)

Bildvermerk Titelseite
© liuzishan, fotolia.com