

Janina Loh

Roboterethik

Eine Einführung
suhrkamp taschenbuch
wissenschaft

suhrkamp taschenbuch
wissenschaft 2277

Die Philosophin Janina Loh befasst sich in ihrem grundlegenden Buch mit den moralischen Herausforderungen, die beim Bau von Robotern und im Umgang mit ihnen eine Rolle spielen: Sind Roboter autonom? Können sie gar moralisch handeln? Haben sie einen moralischen Wert? Sollten ihnen Rechte zuerkannt werden? Wer ist zur Rechenschaft zu ziehen, wenn ein Roboter einen Menschen schädigt? Kritisch diskutiert Loh diese und weitere ethische Fragen und stellt die wichtigsten Lösungsansätze vor.

Janina Loh ist Universitätsassistentin im Bereich Technik- und Medienphilosophie an der Universität Wien.

Janina Loh
Roboterethik
Eine Einführung

Suhrkamp

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet über <http://dnb.dnb.de> abrufbar.

Erste Auflage 2019

suhrkamp taschenbuch wissenschaft 2277

© Suhrkamp Verlag Berlin 2019

Alle Rechte vorbehalten, insbesondere das
der Übersetzung, des öffentlichen Vortrags sowie
der Übertragung durch Rundfunk und Fernsehen,
auch einzelner Teile.

Kein Teil des Werkes darf in irgendeiner Form
(durch Fotografie, Mikrofilm oder andere Verfahren)
ohne schriftliche Genehmigung des Verlages reproduziert
oder unter Verwendung elektronischer Systeme
verarbeitet, vervielfältigt oder verbreitet werden.

Umschlag nach Entwürfen

von Willy Fleckhaus und Rolf Staudt

Druck: Druckhaus Nomos, Sinzheim

Printed in Germany

ISBN 978-3-518-29877-0

*Meinen Gefährt*innen*

NORA TOBI
HANNAH WULF
FRANZI MIA
CHRISTIAN

Inhalt

Einleitung	9
1. Bereiche der Robotik und ihre ethischen Fragen	19
2. Die Arbeitsfelder der Roboterethik	35
2.1 Roboter als moralische Handlungsobjekte	48
2.2 Roboter als moralische Handlungsobjekte	72
2.3 Inklusive Ansätze	95
2.4 Kritische Zwischenbilanz	121
3. Verantwortungszuschreibung in der Mensch-Roboter- Interaktion	126
3.1 Roboter als Verantwortungsobjekte	138
3.2 Roboter als Verantwortungsobjekte	162
3.3 Inklusive Ansätze der Verantwortungszuschreibung .	187
3.4 Kritische Zwischenbilanz	198
4. Abschließende Bemerkungen – Plädoyer für einen inklusive und kritischen Diskurs	204
Literaturverzeichnis	213

Einleitung

Seit Mitte des 20. Jahrhunderts halten Roboter Einzug in immer mehr Bereiche des menschlichen Lebens. Angefangen mit der Industrie und ihrem ersten Roboter *Unimate*¹ über Militär und Kriegsführung bis hin zu Service, Pflege, Medizin und Haushalt werden entweder bereits heute oder aber in absehbarer Zukunft Roboter eingesetzt. Die moralischen Fragen, die beim Bau von und im Umgang mit Robotern aufgeworfen werden, sind Gegenstand der philosophischen Disziplin der Roboterethik. Gegen eine häufig anzutreffende Intuition, dass Technik neutral sei, lässt sich einwenden, dass Technik allgemein Produkt menschlichen Handelns und damit immer (ob bewusst oder unbewusst) durch Normen und Werte bestimmt wird. Denn gerade durch ihre Intention unterscheidet sich eine Handlung vom Instinkt oder bloßen Verhalten (Anscombe 1957; Bratman 1987; Davidson 1980). Durch die Intention werden Werte in eine Handlung eingeschrieben. Menschen wählen über Gründe zwischen unterschiedlichen Handlungsalternativen, die zuvor implizit oder explizit gegen andere Gründe abgewogen wurden. Mit Robotern als spezifischen Technologien,² den in ihnen implementierten Werten und den sich aus ihnen ergebenden (gesellschaftlichen) Konsequenzen gehen also immer moralische Fragen einher (siehe zur sogenannten Neutralitätsthese der Technik auch Kapitel 4).

Auf den folgenden Seiten wird zunächst eine innerdisziplinäre Abgrenzung der Roboterethik von anderen Ethiken sowie weiteren philosophischen Disziplinen vorgenommen. Danach wird die außerdisziplinäre Verwandtschaft der Roboterethik mit nichtphilosophischen Fächern besprochen, bevor die eigentlichen Themen und

1 Ein Kompositum aus »Universal« und »Automation«; erbaut von Joseph Engelberger 1961 (de Miranda 2019a; Ichbiah 2005; 207).

2 Den folgenden Ausführungen liegt ein weiter Technikbegriff zugrunde, der Artefakte (Technologien, Sachtechnik), eingespielte, standardisierte Praktiken, Methoden und Verfahren (Techniken, Prozesstechnik), weitere Technikformen wie Werkstoffe (Realtechnik), Zeichen (Intellektualtechnik), Vollzüge (Sozialtechnik) sowie Attribute (bspw. technikaffin) umfasst (Hubig 2013, 1995; Kranz u. a. 1971-2007).

Fragen der Roboterethik und der sich daraus ergebende Aufbau der vorliegenden Studie vorzustellen sein werden. Eine (Arbeits-) Definition von »Roboter« erfolgt ebenfalls weiter unten im Text.

INNERDISZIPLINÄRE ABGRENZUNG: Die Roboterethik stellt innerhalb des sogenannten westlichen Kulturraums, auf den sich die folgenden Ausführungen beschränken, eine verhältnismäßig junge Bereichsethik dar, eine Teilbereichsethik der Maschinenethik, um ganz genau zu sein. Denn alle Roboter sind Maschinen, aber nicht umgekehrt alle Maschinen auch Roboter (Loh 2019a). Diese Einordnung der Roboterethik als (Teil-)Bereichsethik ruht auf einem Verständnis von Ethik als einer für Menschen spezifischen Kategorie des Handelns. So befasst sich die philosophische Disziplin der Ethik seit Aristoteles einerseits mit den menschlichen Sitten, Bräuchen und Gewohnheiten, andererseits mit dem guten Leben, nimmt darüber hinaus die wissenschaftliche Reflexion menschlicher Praxis vor und beurteilt die Kriterien guten und schlechten Handelns. Es wird vorausgesetzt, dass nur Menschen handelnde Wesen sind, deren Tun nicht blindem Instinkt und Trieb unterworfen ist, sondern durch Intentionen, Normen und Gründe geformt wird (Aristoteles 2006).

Die Rede von Bereichsethiken³ bezieht sich nun auf zwei Typen: 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-, Computer- und allgemein die Technikethik. Ähnlich wie in der Roboterethik nimmt etwa die Tierethik die moralischen normativen Kriterien in den Blick, die in der Züchtung, Domestikation und Haltung von Tieren, allgemein im Umgang mit Tieren und im Verhältnis von Menschen und Tieren eine Rolle spielen (Beauchamp/Frey 2011; Regan 1983; Schmitz 2014; Singer 2015). Bereits der Roboterethiker David Gunkel hat festgestellt, dass sich die Roboterethik insbesondere der Tierethik verwandt fühlen kann, insofern »[d]ie Frage nach der Maschine das Gegenstück zu der Frage nach dem Tier« (2012: 5)⁴ darstelle. Dementsprechend

3 In der akademischen deutschsprachigen Philosophie ist der Oberbegriff für gewöhnlich die Angewandte Ethik, die sich in unterschiedliche Bereichsethiken gliedert (Nida-Rümelin 2005).

4 Sofern nicht anders angegeben, stammen im Folgenden alle Übersetzungen von mir.

hatte schon René Descartes Tieren und Maschinen zunächst denselben ontologischen Status zugeschrieben (Gunkel 2012: 3). Erst im 20. Jahrhundert wurde diese ontologische Gleichstellung von Tier und Maschine zugunsten der Tiere aufgehoben.

Die andere Gruppe an Bereichsethiken versammelt ethische Systeme für Sonderbereiche des menschlichen Lebens, in denen Werte vertreten, Normen geltend gemacht und Regeln formuliert werden, denen im Alltag der Menschen für gewöhnlich ein anderer Status zugeschrieben wird. Die Medizinethik, die Ethik humanitärer Interventionen, die Kriegs- und die Wirtschaftsethik sowie die Ethik internationaler Beziehungen lassen sich als Beispiele für diese Form von Bereichsethiken anführen. Dieser zweite Typ wird in der vorliegenden Studie lediglich am Rande angesprochen.

Nicht jede*r versteht die Roboterethik in der hier vorgeschlagenen Weise.⁵ Oliver Bendel etwa identifiziert schon die Maschinenethik nicht als Bereichsethik, sondern stellt sie »auf eine Stufe mit der Menschenethik«, obwohl er zuvor der traditionellen Auffassung zustimmt, der zufolge sich »Ethik üblicherweise auf die Moral von Menschen« richtet. Weiterhin sieht auch Bendel in der »Roboterethik eine Keimzelle und ein Spezialgebiet der Maschinenethik, wenn sie nicht – wozu man mehr und mehr tendiert – als Bereichsethik aufgefasst wird« (alle Zitate in 2017a: 5). Die vorliegende Untersuchung verpflichtet sich dagegen einer Einordnung der Roboterethik sowohl als Spezialgebiet der Maschinen- als auch als (Teil-)Bereichsethik derselben.⁶

Die *Roboterethik* ist noch einmal von der *Roboterphilosophie* zu unterscheiden. Roboterethik ist eine Disziplin der Roboterphilosophie, die zusätzlich beispielsweise epistemologische, ästhetische, politikphilosophische und rechtsphilosophische Themen behandelt, wobei es innerhalb der Roboterphilosophie natürlich zu einer Überschneidung zahlreicher Fragen – wie etwa von ethischen

5 Im Folgenden nutze ich sowohl das Gender-Sternchen als auch die ausgeschriebene weibliche und männliche Form möglichst im Wechsel. An manchen Stellen wird bewusst nur die weibliche oder männliche Form angegeben. Auch bei Übersetzungen wird, sofern im Original nicht eindeutig eines der Geschlechter gemeint ist, ebenfalls das Gendersternchen bzw. die ausgeschriebene Form verwendet.

6 Zur Maschinenethik siehe Allen u. a. 2006; Anderson/Anderson 2011, 2007; Bendel 2019a; Edgar 2003; Misselhorn 2018a; Moor 2006; Rath u. a. 2019.

und politikphilosophischen Fragen – kommt (Coeckelbergh 2016a; Coeckelbergh u. a. 2018; Ford u. a. 2006; Seibt 2018). Auch im Rahmen dieser Studie ergeben sich an mehreren Stellen Brückenschläge von der Roboterethik in etwa politikphilosophische Gebiete (insbesondere in den Kapiteln 3.2 und 3.3). Eine roboterspezifische Entsprechung der philosophischen Anthropologie, also der philosophischen Disziplin, die nach dem Wesen ›des‹ Menschen fragt, existiert hingegen nicht. Zumeist begnügt man sich auch in philosophischen Untersuchungen zu Robotern mit einer lediglich ein paar Zeilen umfassenden Definition, wie sie auch weiter unten in diesem Text vorgenommen wird (erst recht gilt das für nichtphilosophische Werke). Dieses Phänomen wird in Kapitel 2.3 noch einmal genauere Aufmerksamkeit erfahren, wenn es um Ansätze geht, die die klassische Differenzierung zwischen moralischen Subjekten und Objekten sowie die klare (und üblicherweise anthropologisch-essenzialistische) Grenzziehung zwischen Menschen als den im strengen Sinne einzigen moralischen Subjekten einerseits und nichtmenschlichen Wesen andererseits hinterfragen.

AUSSERDISZIPLINÄRE VERWANDTSCHAFT: Aufgrund ihres Untersuchungsgegenstands tendieren Roboterethikerinnen und -ethiker zur Interdisziplinarität, denn sie erfahren erst aus der Kooperation mit der Informatik, Robotik, KI-Forschung, Kybernetik und weiteren Technik- beziehungsweise Computer- und Ingenieurwissenschaften, wie Roboter konstruiert, programmiert und designt werden. Umgekehrt zeigt sich in den vergangenen Jahren von Seiten der Technikwissenschaften sowie der Industrie ein wachsendes Bewusstsein für die ethischen Herausforderungen, die mit dem Bau und Vertrieb von Robotern einhergehen. Gleichwohl reichen diese Entwicklungen zumindest im deutschsprachigen Raum noch nicht so weit, verpflichtende Ethikkurse in der Ausbildung der Robotiker*innen von morgen zu etablieren oder Weiterbildungskurse für Unternehmen beziehungsweise in der Produktion von Robotern anzubieten (siehe hierzu auch Kapitel 4).

NÄHERE BESTIMMUNG DER ROBOTERETHIK UND AUFBAU DIESER STUDIE: Im deutschsprachigen Raum stellt die Roboterethik noch keine allgemein anerkannte Disziplin innerhalb der akademischen Philosophie dar, auch wenn das Interesse an teildisziplinübergreifenden Kollaborationen wächst. Im Vergleich mit dem englischsprachigen Raum, wo die ethische Auseinandersetzung

mit artifiziellen Systemen seit Mitte des vergangenen Jahrhunderts einen Kanon klassischer Literatur hervorgebracht hat, präsentiert sich der deutschsprachige Diskurs überschaubar.⁷ Hierzulande muss sich die Roboterethik aus den Reihen akademischer Philosophinnen und Philosophen ab und an noch den Vorwurf gefallen lassen, sie sei gar keine richtige Ethik beziehungsweise sie habe keinen spezifischen Gegenstand, da sich Ethik nur mit dem menschlichen Handeln beschäftigt. Doch selbst wenn sich nach vorheriger Prüfung herausstellen sollte, dass Roboter keine moralischen Handlungssubjekte sein können und damit zu moralischem Handeln selbst nicht befähigt wären, sollte man ihnen einen Platz im moralischen Universum zuweisen. Schließlich sind alle möglichen Wesen und Entitäten – wie etwa Tiere, Pflanzen, Häuser, Autos, Smartphones, Landschaften oder ganze Ökosysteme – Objekte moralischen Handelns, und wir sprechen einer ganzen Reihe von nichtmenschlichen und zum Teil auch unbelebten Entitäten einen Wert, ja, in manchen Fällen sogar Rechte zu. Um was für eine Art von Wert es sich im Falle von Robotern handelt, bleibt freilich zu diskutieren und hängt von dem zugrunde liegenden Ansatz ab. Doch wo, wenn nicht in der Ethik, wäre der angemessene Raum für eine solche Diskussion? Nachdem im ersten Kapitel ein Überblick über die ethischen Fragen, die sich in einigen Bereichen der Robotik stellen, gegeben wurde, behandelt das gesamte zweite Kapitel ebendie Frage, welchen Wert wir einem jeweiligen Roboter zuschreiben sollten.

Vergleichbar den anderen Bereichsethiken, die sich mit nichtmenschlichen Wesen und Entitäten beschäftigen, wird in der Roboterethik darüber nachgedacht, inwiefern das fragliche Gegenüber selbst als moralischer Akteur und damit als Subjekt moralischen Handelns interpretiert werden muss, inwiefern Roboter also *moral agents* sind, »Entitäten, die Handlungen ausführen können« (Floridi/Sanders 2004: 349). Roboter als *Subjekte* moralischen Handelns, denen potenziell ein intrinsischer (das heißt ein absoluter oder Eigenwert) zukommt, sind Gegenstand von Kapitel 2.1.

7 Zum englischsprachigen Diskurs siehe Asaro 2006; Bekey 2005; Brey u. a. 2008; Capurro/Nagenborg 2009; Lin u. a. 2017, 2012, 2011; Stahl 2004; Sullins 2006; Turing 1950; Versenyi 1974; Wallach/Allen 2009. Zum deutschsprachigen Diskurs siehe Hilgendorf 2014; Loh 2017a; Mainzer 2010; Misselhorn 2018a; Remmers 2018; Sombetzki 2016.

In einem zweiten Bereich wird danach gefragt, inwiefern man sich Robotern gegenüber in einer bestimmten Weise verhalten sollte beziehungsweise ob ihnen ein spezifischer Umgang zusteht und inwiefern sie damit in die Kategorie der sogenannten *moral patients* fallen, »als Entitäten, die man gut oder schlecht behandeln kann« (Floridi/Sanders 2004: 349). Mit Robotern als *Objekten* moralischen Handelns, denen ein intrinsischer oder ein instrumenteller Wert zukommt, befasst sich das Kapitel 2.2.⁸ Weiterhin, in einem dritten Bereich, diskutiert die Roboterethik Alternativen zu dieser klassischen Unterscheidung zwischen Subjekten und Objekten moralischen Handelns. In Kapitel 2.3 wird sich zeigen, inwiefern einige Roboterethikerinnen und -ethiker die aristotelische Differenzierung zwischen moralischen Handlungssubjekten und -objekten sowie insbesondere das gängige Verständnis vom moralischen Handlungssubjekt problematisieren und welche Alternativen sie dafür vorschlagen.⁹

Die Roboterethik stellt also traditionelle ethische Fragen mit Blick auf Roboter – etwa danach, welche Kompetenzen wir generell als grundlegend für moralische Akteursschaft erachten, welche moralischen (und anderen) Werte wir artifiziellen Systemen implementieren sollten, auf was für ein moralisches Selbstverständnis es schließen lässt, wenn wir Roboter »schlecht« behandeln, und in welchen Bereichen – zum Beispiel Industrie-, Militär-, Medizin-, Altenpflege-, Servicerobotik – wir uns auch zukünftig ausschließlich beziehungsweise in einem signifikanten Ausmaß auf menschliche und nicht auf artifizielle Expertise verlassen sollten.

Das dritte Kapitel und damit der zweite Hauptteil dieser Studie ist Fragen der Verantwortungszuschreibung in der Mensch-Roboter-Interaktion gewidmet – vorrangig aus ethischer Perspektive, bietet es dabei jedoch auch einige Ausblicke auf weitere für die Zuschreibung von Verantwortung relevante Handlungsräume wie etwa rechtliche und gesellschaftliche Verantwortlichkeiten, die mit Robotern einhergehen. Denn die Verantwortung ist

8 Zur in der Philosophie geläufigen Unterscheidung zwischen intrinsischem/absolutem Eigenwert (den etwas um seiner selbst willen hat) und instrumentellem/extrinsischem Wert (den etwas für etwas anderes hat) vgl. Korsgaard 1983 sowie Zimmerman/Bradley 2019.

9 Damit orientiert sich diese Studie in Kap. 2 grob an Gunkels Werk *The Machine Question* (2012), in dem der Autor ähnlich vorgeht.

nicht nur ein grundsätzliches Phänomen in nahezu jeder Sphäre menschlichen Handelns, sie wird auch von vielen Menschen als anthropologische Konstante gesehen und ist eine fundamentale ethische Kategorie. Daher ist die Verantwortung auch grundlegend für ein ethisches Nachdenken über Roboter. Eine Untersuchung von Verantwortungszuschreibungen in der Roboterethik sowie eine Analyse der mit dem Bau von Robotern einhergehenden Verantwortlichkeiten bietet zudem einen Brückenschlag zu weiteren Bereichen menschlichen Handelns, in denen Entscheidungen über den Einsatz von sowie den Umgang mit Robotern zu treffen sind, wie etwa in der Politik (politische Verantwortung), der Ökonomie (Konsument*innenverantwortung, Verantwortung von Unternehmen) und dem Recht (strafrechtliche Verantwortung). In Orientierung am zweiten Kapitel wird in Kapitel 3.1 danach gefragt, inwiefern Roboter Verantwortungssubjekte sein können, in Kapitel 3.2 diskutiert, inwiefern Menschen für Roboter als Verantwortungsobjekte Rede und Antwort zu stehen haben, und in Kapitel 3.3 besprochen, welche alternativen Zuschreibungsmodelle von Verantwortung es in der Mensch-Roboter-Interaktion gibt.

Darüber hinaus schlägt das dritte Kapitel mehrere Antworten vor, wie mit einer gegenwärtig vermuteten Transformation dieser wichtigen menschlichen Kompetenz umgegangen werden kann (Loh 2019b). Die Vermutung einer Transformation der Verantwortung speist sich *zum einen* aus den fundamentalen Umwälzungen im Wesen *des* Menschen, die einer Entwicklung autonomer, selbstlernender Roboter zugeschrieben werden. *Zum anderen* spricht man von radikalen Paradigmenwechseln und einer damit einhergehenden Transformation unseres Verantwortungsverständnisses in den Organisationsformen unserer gesellschaftlichen, politischen und ökonomischen Systeme aufgrund der Herausforderungen durch Robotisierung, Automatisierung, Digitalisierung und Industrie 4.0. Allenthalben wird auch der Sorge Ausdruck verliehen, dass dank dieser Umstände unserer modernen technisierten Massengesellschaft der Verantwortung ultimative Grenzen gesetzt sind, sich gar gefährliche Lücken in den Möglichkeiten, Verantwortung zuzuschreiben, auftun (siehe etwa Lenk 1994 und Matthias 2004 in Kapitel 3.2). Dennoch scheint der Ruf nach Verantwortung zugleich mit ungebrochener Vehemenz zu erklingen. Die Frage, die das dritte Kapitel beantwortet, lautet, ob wir trotz

all dieser Veränderungen auf das traditionelle Konzept der Verantwortung weiterhin bauen können (die Kapitel 3.1. und 3.2) oder ob sich unser tradiertes Verständnis von Verantwortung tatsächlich wandelt beziehungsweise wandeln sollte (Kapitel 3.3).

(ARBEITS-)DEFINITION VON »ROBOTER«: Einleitend wurde die Roboterethik als Teilbereichsethik der Maschinenethik bezeichnet, weil alle Roboter Maschinen, nicht aber alle Maschinen Roboter sind. Eine Maschine ist ein künstliches Gebilde, das aus durch ein Antriebssystem (Motor, Wind, Wasser) bewegten Teilen besteht und Energie umsetzt (Canguilhem 2012; Strandh 1980). Roboter sind spezielle Maschinen. Historisch geht der Ausdruck »Roboter« auf das tschechische Wort »robota« für Arbeit, Frondienst und Zwangsarbeit zurück, das 1920 von dem Künstler Josef Čapek geprägt wurde (vom slawischen Wortstamm »rab« für »Sklave«; Jordan 2017: 50; siehe auch deMiranda 2019b). Sein Bruder Karel Čapek nutzte es in dem Theaterstück *R.U.R. Rossum's Universal Robots* (1921) als Bezeichnung für humanoide Apparaturen, die den Menschen zu Diensten stehen. Der erste etablierte Bereich der Robotik, nämlich die Industrie, spiegelt die Vision, die Čapek in dem genannten Theaterstück entwirft (siehe hierzu Kapitel 1). Auch ist sie Kernbestand der sogenannten Industrie 4.0, der technologischen Transformation der menschlichen Arbeitswelt durch Digitalisierung und Automatisierung. Čapek ist allerdings kein uneingeschränkter Technikenthusiast, sondern geht in *R.U.R.* auf zahlreiche Herausforderungen ein, die die Erschaffung von Robotern mit sich bringt. Der Plot seines Stücks läuft letztlich auf eine Revolte der Roboter hinaus, die die Weltherrschaft anstreben. Čapek wirft darüber hinaus zahlreiche weitere philosophische Fragen auf, etwa die nach dem Wesen »des« Menschen, nach der Verantwortung der Wissenschaftlerinnen und Wissenschaftler für ihre künstlichen Kreaturen sowie danach, was es heißt, eine emotionale Bindung zu einem anderen Wesen einzugehen. So endet sein Stück beispielsweise mit einer Liebesbeziehung, die sich zwischen zwei Robotern anbahnt. Damit ist in dem durch Čapek begründeten historischen Verständnis vom Roboter eine breite Grundlage für die Diskussionen angelegt, die in den sich anschließenden Jahrzehnten aufkommen sollten.

Den folgenden Ausführungen liegt eine erweiterte Version einer von Catrin Misselhorn (2013: 43) vorgeschlagenen Definition zu-

grunde, der zufolge ein Roboter eine elektro-mechanische Maschine ist, die a) über einen *eigenständigen Körper* und b) über mindestens einen *Prozessor* verfügt, c) *Sensoren* hat, die Informationen über die Welt sammeln, d) sowie über *Effektoren oder Aktoren* verfügt, die Signale in mechanische Abläufe übersetzen. Das Verhalten eines Roboters e) ist oder erscheint zumindest *autonom*, und er kann f) in seine Umgebung hineinwirken beziehungsweise physisch auf sie *Einfluss* nehmen. Dieses Verständnis ist nicht unproblematisch, sind doch einige der angeführten Bedingungen (wie beispielsweise Verkörperung, Autonomie und Einflussmöglichkeit) mehrdeutig und daher mindestens erklärungsbedürftig. Auch schließt es häufig als Roboter bezeichnete künstliche Systeme wie etwa Computer, Chatbots, medizinische Assistenzsysteme und ferngesteuerte Drohnen aus. Damit eröffnet es den großen Graubereich, der von Technikphilosoph*innen ausgeleuchtet zu werden verdient und in dem der Science-Fiction-Schriftsteller Isaac Asimov bereits die roboternahe Sphäre vermutet hat, in der wir auch weiteren Verwandten und Bekannten der Roboter begegnen (1982: 53).

Nach der hier vorgeschlagenen Definition handelt es sich also um keinen Roboter im eigentlichen Sinne, wenn eine der genannten Bedingungen a) bis f) nicht vorliegt: So erfüllen Computer etwa nicht die Bedingung f). Die Art und Weise der Verkörperung spielt hingegen keine Rolle, Roboter gibt es in jeder denkbaren Gestalt, die humanoiden unter ihnen werden Androidinnen und Androiden genannt (Drux 1988). Computer sind im übertragenen Sinn eher das ›Gehirn‹ eines Roboters, so wie Algorithmen metaphorisch gesprochen etwa deren ›mentale Verhaltensmuster‹ und ›gelernte Abläufe‹ darstellen, nicht aber den eigentlichen Roboter selbst. Künstlichen Systemen wie beispielsweise dem chirurgischen Assistenzsystem *DaVinci* und ebenso Drohnen mangelt es an der Autonomiebedingung e) (Jordan 2017: 31, 59, 262-266). Die Bedingungen a) bis f) sind im Einzelnen notwendig und gemeinsam hinreichend für die in dieser Studie verwendete enge Definition von »Roboter«. Auf den folgenden Seiten werden insbesondere die philosophisch herausfordernden Facetten dieser Definition betrachtet, wie etwa das Autonomieverständnis (Kapitel 2.1) sowie das, was mit Einfluss gemeint ist (Kapitel 3.1 und 3.2).

An der einen oder anderen Stelle (insbesondere in Kapitel 1) wird diese Einführung auch Roboter in einem weiteren Sinne (also

Entitäten, bei denen eine oder mehrere der Bedingungen nicht erfüllt sind) zumindest streifen. Es geht weniger darum, strikt zwischen Robotern und Nichtrobotern zu differenzieren. Würden wir den Begriff »Roboter« jedoch generell in einem weiten Sinn verstehen, wozu auch die Alltagssprache neigt, dann kämen wir vermutlich dahin, ihn als generell deckungsgleich mit »Maschine« anzusehen.¹⁰

Ich bin mehreren Personen zu großem Dank verpflichtet, die mich bei der Entstehung dieses Buches unterstützt und mir mit Rat und ihrer Einschätzung zur Seite gestanden haben: John-Stewart Gordon, Isabella Hermann, Eric Hilgendorf, Tanja Kubes, Lara Lammer, Wulf Loh, Susanne Steigler und Mia Steinfeldt danke ich für ihre ausnehmend kritische Lektüre des Manuskripts, die unwunden vorgebrachten Verbesserungsvorschläge, zahlreiche Literaturtipps, ihre Hinweise und Korrekturen, die meine Argumentation geschärft haben und insbesondere für die Diskussion meines Ansatzes sowie ihre wiederholten Aufforderungen zu mehr Genauigkeit und Ausführlichkeit. Durch das Team der Technik- und Medienphilosophie an der Universität Wien um Mark Coeckelbergh wurde mir ein sehr produktiver Raum zur Arbeit am und zum Austausch über das Buch gegeben, der nicht zuletzt auch durch meine Student*innen und die mit ihnen geführten Gespräche belebt und bereichert wurde. Auch möchte ich ausdrücklich meinem Lektor, Jan-Erik Strasser, für die hervorragende Zusammenarbeit im Allgemeinen sowie die sensible Lektüre des Manuskripts im Besonderen danken.

10 Ähnliche Definitionsvorschläge für »Roboter« finden sich bei Bekey 2005: 2; Breazeal 2004; Jordan 2017: 41-59; Mataric 2007: 2; Remmers 2018.

I. Bereiche der Robotik und ihre ethischen Fragen

Diese Studie fokussiert die drei Hauptarbeitsfelder der Roboterethik sowie Rolle und Funktion der Verantwortung in diesen. Zuvor wird nun ein Einblick in fünf Sektoren der Robotik und die dort aufgeworfenen ethischen Fragen gegeben, nämlich Robotik und Arbeit (paradigmatisch an der Industrierobotik diskutiert), autonomes Fahren, Medizin-, Therapie- und Pflegerobotik (als exemplarisch für den Großbereich der *Social Robotics*), Sexrobotik sowie Militärrobotik. Denn es existiert nahezu kein Bereich des menschlichen Alltags, in den Roboter (mit mehr oder minder ausgeprägter Autonomie sowie unterschiedlich differenzierten Lern- und Interaktionsfähigkeiten) noch nicht Einzug gehalten hätten. In den Kapiteln 2 und 3 werde ich wiederholt auf diese fünf exemplarischen Bereiche der Robotik zurückkommen. Weitere Sektoren der Robotik, die hier nicht in den Blick genommen werden können, sind beispielsweise Haushaltsrobotik, Robotik in der Bildung, Robotik als Spielzeug oder Robotik in Kunst, Literatur, Film und Musik (Ichbiah 2005; Jordan 2017: 61-98; Ceppi 2019; McNulty 2019; Verdicchio 2019).

Ausgehend von der seit Karel Čapek und dessen Theaterstück *R.U.R. Rossum's Universal Robots* (1921) ersten Dimension des menschlichen Alltags, in die Roboter signifikant eingebunden worden sind, der *Industrierobotik*, geht das Thema *Robotik und Arbeit* inzwischen weit über die Industrie im engen Sinne hinaus. *Autonome Fahrassistenzsysteme* finden in der öffentlichen Diskussion breite Resonanz, wird doch seit einiger Zeit auch im deutschsprachigen Raum verstärkt über ihre serienmäßige Produktion und Vermarktung nachgedacht. Bereits in der Gegenwart werden Roboter-Assistenzsysteme für *Medizin, Therapie und Pflege* entwickelt, wo sie die Pfleger*innen etwa bei körperlich anstrengenden Tätigkeiten unterstützen sollen. Darüber hinaus arbeitet man gerade in diesem Sektor an autonomen artifiziellen Systemen, die eigenständig in Privathäusern zum Einsatz gebracht werden, dort körperlich beziehungsweise geistig eingeschränkten Menschen in ihrem Alltag zur Hand gehen und unter Umständen sogar soziale Interaktionspart-