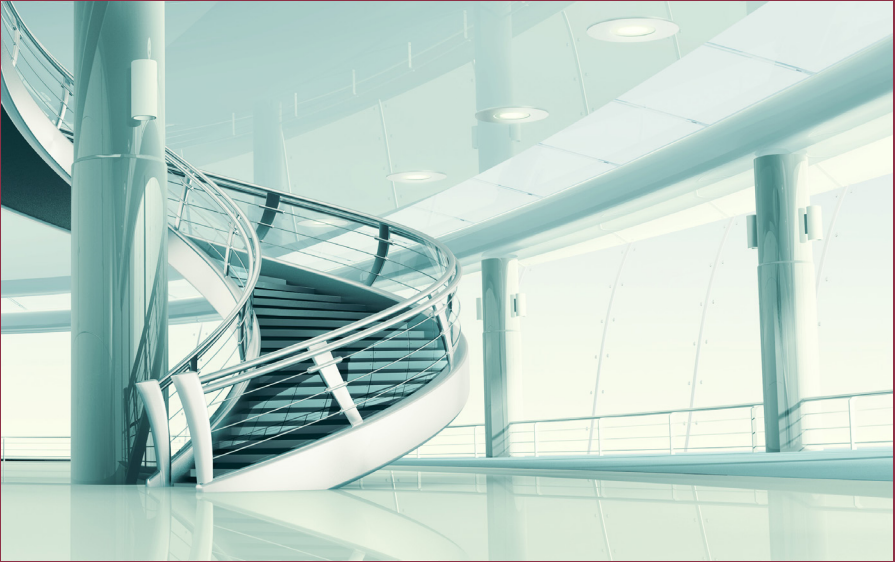


PHILOSOPHIE, NATURWISSENSCHAFT
UND TECHNIK



Zukunft gestalten

**Digitalisierung, Künstliche Intelligenz (KI)
und Philosophie**

Karsten Berr / Jürgen H. Franz (Hg.)

Karsten Berr / Jürgen H. Franz (Hg.)

Zukunft gestalten – Digitalisierung, Künstliche Intelligenz (KI) und Philosophie

Philosophie, Naturwissenschaft und Technik
Band 9

Karsten Berr / Jürgen H. Franz (Hg.)

Zukunft gestalten –
Digitalisierung, Künstliche Intelligenz (KI)
und Philosophie

F Frank & Timme
Verlag für wissenschaftliche Literatur

Umschlagabbildung: Illustration of spacious interior toned cool colors
© AlienForce – stock.adobe.com

ISBN 978-3-7329-0547-8
ISBN E-Book 978-3-7329-9451-9
ISSN 2365-4074

© Frank & Timme GmbH Verlag für wissenschaftliche Literatur
Berlin 2019. Alle Rechte vorbehalten.

Das Werk einschließlich aller Teile ist urheberrechtlich geschützt.
Jede Verwertung außerhalb der engen Grenzen des Urheberrechts-
gesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar.
Das gilt insbesondere für Vervielfältigungen, Übersetzungen,
Mikroverfilmungen und die Einspeicherung und Verarbeitung in
elektronischen Systemen.

Herstellung durch Frank & Timme GmbH,
Wittelsbacherstraße 27a, 10707 Berlin.
Printed in Germany.
Gedruckt auf säurefreiem, alterungsbeständigem Papier.

www.frank-timme.de

Vorwort

Der vorliegende Band umfasst neunzehn Beiträge der wissenschaftlichen und interdisziplinären Tagung *APHIN III 2018 – Zukunft gestalten*, die vom Arbeitskreis philosophierender Ingenieure und Naturwissenschaftler (APHIN e.V.) im November 2018 in Enkirch an der Mosel veranstaltet wurde. Aufgrund der Interdisziplinarität der Tagung standen alle Referenten vor der Herausforderung, den Spagat zwischen wissenschaftlicher Tiefe einerseits und interdisziplinärer Breite und Allgemeinverständlichkeit andererseits zu wagen. Und wie die Beiträge dieses Bandes zeigen, wurde die Herausforderung von allen Referenten gerne angenommen und vorzüglich bestanden.

APHIN wurde im Oktober 2013 als wissenschaftlicher, bildungsorientierter, interdisziplinärer und gemeinnütziger Verein gegründet und verzeichnet seitdem eine stetig wachsende Mitgliederanzahl. Er bewegt sich im Spannungsfeld von Philosophie, Ingenieur- und Naturwissenschaft und ist offen für alle, die mit Freude und Neugierde über ihren eigenen fachlichen Tellerrand hinaus schauen und in der Philosophie die Möglichkeit entdeckt haben, dieser Freude und Neugierde einen adäquaten Raum zu geben. Seine wissenschaftlichen Mitglieder sind Philosophen, Ingenieure und Naturwissenschaftler, Mathematiker, Informatiker, Mediziner, Theologen, Pädagogen, Wirtschaftswissenschaftler, Erziehungswissenschaftler, Juristen, Künstler und andere. Zum Selbstverständnis von APHIN gehört eine gelebte, praktizierte Interdisziplinarität, wobei die Philosophie als Band fungiert, das die unterschiedlichen Disziplinen seiner Mitglieder miteinander bindet und eint.

Bereits ein Jahr nach seiner Gründung veranstaltete APHIN unter dem Titel *APHIN I 2014 – Prolegomena* im Geburtshaus des Cusanus in Bernkastel-Kues an der Mosel seine erste öffentliche wissenschaftliche Tagung. Als Auftaktveranstaltung fungierte sie als Vorwort – Prolegomena – zu den Folgetagungen, die APHIN im Rhythmus von zwei Jahren veranstaltet. Im Jahr 2016 stand die Tagung unter dem Motto *Welt der Artefakte* und damit ganz im Zeichen stofflicher und geistiger, menschlicher Schöpfungsprodukte und ihres Einflusses auf Mensch, Gesellschaft und Natur. Alle Beiträge dieser beiden ersten Tagungen von APHIN wurden beim Verlag für wissenschaftliche Literatur Frank & Timme publiziert.

Die Tagung *Zukunft gestalten* wurde erneut in partnerschaftlicher Kooperation mit der Kueser Akademie für europäische Geistesgeschichte durchgeführt. Wir danken

unserem Förderer, der gemeinnützigen *Stiftung PfalzMetall*, die unsere Tagung finanziell unterstützte. Dank gilt auch all jenen, die durch ihr ehrenamtliches Engagement zum Erfolg dieser Tagung beitrugen und bei der Erstellung des Tagungsbandes mitwirkten, allen voran den Vortragenden, die uns ihre Beiträge zur Publikation zur Verfügung stellten, und den vielen lieben Menschen, die für einen reibungslosen Ablauf der Tagung und für eine herzliche Atmosphäre sorgten. Besonderer Dank gilt unseren Fördermitgliedern, die uns sowohl ideell als auch finanziell stärken und damit bereits seit Gründung von APHIN entscheidend zum Erfolg unserer Tagungen beisteuern.

Jürgen H. Franz und Karsten Berr
www.aphin.de

März 2019

Inhalt

Vorwort | 5

Inhalt | 7

Einleitung | 11

Teil I: Design, Smarte Maschinen und Mobilität – eine ethische und gesellschaftliche Herausforderung

Janina Lob

Verantwortung und Roboterethik: Ein Überblick am Beispiel autonomer Fahrassistenzsysteme. | 17

Rainer Königstedt

Zukunft der Mobilität. | 29

Manja Unger-Büttner

Zukunft – Design – Ethik. Ein exploratives Gemisch. | 41

Teil II: Zukunft gestalten mit Epikur, Kant und Bergson

Torsten Nieland

In Epikurs Lustgarten: Gedanken über Nachhaltigkeit. | 55

Alfred Berlich

Kants Vision eines ewigen Friedens als Menschheitsprojekt der Zukunft. | 67

Matthias Vollet

Kann man die Zukunft gestalten? Henri Bergson und die Unvorhersehbarkeit und Vorhersehbarkeit der Zukunft. | 79

Teil III: Mögliche Welten und Räume, modale Gestaltung und Logik

Hartmut W. Mayer

„Mögliche Welten“ und die Struktur der Zukunft. Ein philosophiegeschichtlicher
Rundgang bis zur Modallogik. | **93**

Bruno Gransche

Improvisierte Provisorien. Zukunft als Möglichkeitsraum
modaler Gestaltung. | **103**

Teil IV: Welche Zukunft? Zukunft gestalten mit Philosophie?

Ruth Spiertz

Welche Zukunft? Zukunftsvisionen aus der Sicht des Offenen
Skeptizismus. | **117**

Jürgen H. Franz

Warum die Gestaltung der Zukunft der Philosophie bedarf. | **129**

Teil V: Die Rolle des Rechts, der Toleranz und der Bildung

Dirk Löhr

Eigentum an Land und Natur: Diebstahl an der Zukunft? | **141**

Norbert Hill

Das Potential des Rechts bei der Gestaltung der Zukunft. Über das
Spannungsfeld zwischen Norm und Wirklichkeit
am Beispiel des Grundgesetzes. | **151**

Uta Henze

„Toleranz ist menschlich“ – Wertereflexion im Philosophieunterricht als
Basiskompetenz für die Gestaltung unserer Zukunft. | **163**

**Teil VI: Wie wollen wir leben? Welche Technologien brauchen wir?
Künstliche Intelligenz, Digitalisierung und Transhumanismus**

Janina Lob

Wider die Utopie einer umfassenden Kontrolle.
Kritische Überlegungen zum Transhumanismus. | 177

Günter Franke

„Digitalisierte“ Menschheit.
Eine kritische Spekulation – oder eine spekulative Kritik. | 189

Michael Kuhn

Wie wollen wir leben und welche Technologien brauchen wir dafür?
Fiktionen als hypothetische Vorwegnahme der zukünftigen
Welt und Technik. | 201

Hyun Kang Kim

Die Differenz zwischen künstlicher und menschlicher Intelligenz. | 213

Teil VII: Studierendenpreisträgerinnen unter sich

Julia Fuchte

Utopische Szenarien als Beitrag zu einer
transformativen Bildung und Wissenschaft. | 225

Elisa Oertel

Die künstliche Superintelligenz als Konsequenz einer
nach Effizienz strebenden Menschheit? | 235

Autorenverzeichnis | 243

Einleitung

Markus Dangel

Die Beiträge dieses Tagungsbandes untergliedern sich in sieben Sektionen, die jeweils besondere Aspekte des umfangreichen Rahmenthemas *Zukunft gestalten* aufgreifen und beleuchten. Dass es sich bei diesem Themenkomplex um eine gesamtgesellschaftliche Aufgabe handelt, zeigt sich an der Vielgestaltigkeit der Beiträge sowohl in Inhalt als auch Methodik und ihrer interdisziplinären Einbettung – in diesem Sinne spiegelt dieser Sammelband damit eine wesentliche Leitidee und Motivation von APHIN e.V. wider.

Im Folgenden werden die 19 Beiträge des Bandes kurz vorgestellt.

Teil I

Janina Loh untersucht in ihrem Beitrag die Fragestellung, inwiefern das klassisch an Personen gebundene Konzept der Verantwortung auf künstliche Systeme am Beispiel autonomer Fahrassistenzsysteme übertragbar ist.

Einen kritischen Blick in die Zukunft der Mobilität wirft **Rainer Königstedt**, indem er aufzeigt, welche Fehleinschätzungen hinsichtlich Ressourcen- und Energiebedarf bei der Einführung neuer Mobilitätskonzepte drohen.

Manja Unger-Büttner plädiert in ihrem Artikel für eine explorative Perspektive auf das Spannungsfeld Technik – Design – Ethik.

Teil II

Torsten Nieland greift in seinem Referat Epikurs Philosophie auf, um diese für Gedanken über Nachhaltigkeit fruchtbar zu machen.

Alfred Berlich erörtert, wie Immanuel Kants Geschichts-, Rechts- und Staatsphilosophie als Menschheitsprojekt der Zukunft verstanden werden kann.

Matthias Vollet betrachtet mit Henri Bergson die Frage nach der Verfügbarkeit und Gestaltbarkeit von Zukunft.

Teil III

In seinem philosophiegeschichtlichen Rundgang beleuchtet **Hartmut W. Mayer** die Struktur von Zukunft als Raum möglicher Welten.

Bruno Gransche greift das Tagungsthema auf, indem er für ein Konzept medialer modaler Gestaltung plädiert.

Teil IV

Ruth Spiertz begründet in ihrem Beitrag, wie sich Zukunftsvisionen mit Hilfe des Offenen Skeptizismus fundieren lassen.

Jürgen H. Franz erläutert in seinem Artikel, warum Philosophie für die Gestaltung der Zukunft unabdingbar ist.

Teil V

Dirk Löhr thematisiert die Frage nach Eigentumsrechten an Land und Natur als möglichen Raubbau an den Gestaltungsmöglichkeiten künftiger Generationen.

Am Beispiel des Grundgesetzes erörtert **Norbert Hill** das Potential des Rechts bei Fragen der Zukunftsgestaltung.

Uta Henze stellt in ihrem Beitrag die Bedeutung der Wertereflexion im Philosophieunterricht als Basiskompetenz für die Gestaltung der Zukunft heraus.

Teil VI

Janina Loh erläutert in ihrem Referat kritische Einwände zu Kontrollutopien des Transhumanismus.

Eine kritische Spekulation über die digitale Zukunft der Menschheit entwirft **Günter Franke** in seinem Beitrag.

Michael Kuhn plädiert für den Begriff der Fiktion, um technisches Gestalten adäquat zu beschreiben.

Hyun Kang Kim zeigt eine fundamentale Differenz zwischen humaner und Künstlicher Intelligenz auf, die sie anhand von drei Thesen erläutert.

Teil VII

Julia Fuchte zeigt in ihrem Referat, wie utopische Szenarien zu einer transformativen Bildung und Wissenschaft beitragen können.

Elisa Oertel untersucht in ihrem preisgekrönten Essay den Zusammenhang zwischen einer möglichen künstlichen Superintelligenz und der nach Effizienz strebenden Menschheit.

Teil I

Design, Smarte Maschinen und Mobilität – eine ethische und gesellschaftliche Herausforderung

Verantwortung und Roboterethik: Ein Überblick am Beispiel autonomer Fahrassistenzsysteme

Janina Lob

Einleitung

Neben zahlreichen Herausforderungen, mit denen uns der rasante Fortschritt in Robotik und KI-Forschung gegenwärtig konfrontiert, sehen wir uns vor die Aufgabe gestellt, traditionell nur dem Menschen vorbehaltene Kompetenzen – Vernunft, Autonomie, Urteilskraft, um nur einige zu nennen – in ihrer Übertragung auf künstliche Systeme zu transformieren. Im Folgenden widme ich mich dem Phänomen der Verantwortung, um am Beispiel dieser Kernkompetenz des Menschen einen Überblick über das Arbeitsfeld der Roboterethik zu geben. Dafür frage ich zunächst danach, was unter Verantwortung traditionell verstanden wird, und schlage eine Minimaldefinition von „Verantwortung“ vor, welche nur die wesentlichen etymologischen Komponenten und damit den ‚kleinsten gemeinsamen Nenner‘ jeder Rede von Verantwortung enthält.¹ In einem zweiten Schritt erläutere ich, was es mit der philosophischen Disziplin der Roboterethik auf sich hat, um zuletzt die Rolle der Verantwortung innerhalb derselben näher in den Blick zu nehmen.

1. Was ist Verantwortung?

Eine etymologische Untersuchung würde zeigen, dass unser Verständnis von Verantwortung auf drei Säulen fußt: Verantwortung bedeutet *erstens*, dass jemand Rede und Antwort steht, und *zweitens*, dass dies kein rein deskriptives, sondern immer ein zumindest auch normatives Geschehen darstellt. Gerne erklärt man zwar bspw. den Regen verantwortlich für das Nass-Sein der Straße. Doch hier ist von Verantwortung in übertragenem Sinn als Verursachung die Rede. *Drittens* rekurriert die Rede von Verantwortung auch immer auf bestimmte Kompetenzen, die wir der* Angesprochenen implizit zuschreiben. Wir unterstellen, dass die fragliche Person integer, bedacht und reflektiert das Anliegen der Verantwortung in Angriff nimmt.

¹ Vgl. ausführlich Sombetzki 2014. Aus Platzgründen ist die Literatur in diesem Artikel auf ein Minimum beschränkt.

Aus dieser Minimaldefinition ergeben sich fünf Relationselemente der Verantwortung. Es bedarf eines *Subjekts* bzw. einer* Trägerin* der Verantwortung. Darüber hinaus ist ein *Objekt* oder Gegenstand zu definieren. Drittens gilt es, die *In-stanz*, vor der man sich verantwortlich zeigt, auszumachen. Viertens tragen wir gegenüber einer* *Adressatin** Verantwortung. Schließlich geben *normative Kriterien* den Maßstab und die Richtlinien dafür ab, in welcher Weise Verantwortung zuzuschreiben ist.

Die Bedingungen für die Möglichkeit einer Zuschreibung von Verantwortung lassen sich in drei Kompetenzgruppen differenzieren: Kommunikationsfähigkeit, Handlungsfähigkeit bzw. Autonomie und Urteilskraft. Alle Kompetenzen als Voraussetzung für die etwaige Zuschreibung von Verantwortung und mit ihr die Verantwortung selbst sind graduell bestimmbar; man kann von mehr oder weniger Kommunikations- und Handlungsfähigkeit sprechen und abhängig davon von mehr oder weniger Verantwortung.²

Das *Subjekt* der Verantwortung ist die Person, die Rede und Antwort stehen kann.³ Abhängig von den Bedingungen, die für eine etwaige Zuschreibung von Verantwortung erfüllt sein müssen, lässt sich der Frage nachgehen, ob nur ‚gesunde und erwachsene‘ Menschen oder auch Kinder für Aufforderungen zu einer Übernahme von Verantwortung ansprechbar sind. Sind es gar auch (einige) Tiere und auch Pflanzen, vielleicht sogar unbelebte Dinge (wie bspw. einige artifizielle Systeme)? Innerhalb des Verantwortungsdiskurses ist man sich jedoch darüber einig, dass Verantwortung traditionell ein individualistisches Prinzip darstellt und in deutlicher Nähe zu (wenn auch nicht unbedingt gleichbedeutend mit) dem Konzept der Personalität gesehen wird. Dieser Kernbestand des klassischen Verantwortungskonzepts wird erst in der Gegenwart in Frage gestellt. Hieraus erhellt, warum jeder rein deskriptive oder kausale Gebrauch desselben, d. h. ein solcher, der einen normativen Gebrauch gar nicht erlaubt, wie in dem obigen Regen-Beispiel, nur metaphorisch gemeint sein kann. Der Regen ist nicht in der Lage, Rede und Antwort zu stehen und Pflanzen wohl ebenso wenig. Ein genaues Verständnis von Verantwortung in einer fraglichen Situation verlangt ein Urteil darüber, ob es sich bei dem

² Vgl. Sombetzki 2014 – zur Minimaldefinition 33ff., zu den Relationselementen 63ff. und zu den Bedingungen für die Möglichkeit zur Verantwortungsübernahme 43ff.

³ Zu den einzelnen Relationselementen, die nun vorgestellt werden, gibt es in Sombetzki 2014 jeweils ein eigenes Unterkapitel.

fraglichen Verantwortungssubjekt um ein Individuum oder um ein Kollektiv handelt. In Abschnitt 3.2 führe ich den Terminus des Verantwortungsnetzwerks ein⁴, um die unterschiedlichen Funktionen in den Blick zu bekommen, die die involvierten Parteien insbesondere im Falle von Mensch-Maschine-Interaktionen haben.

Die Rede von Verantwortung verlangt neben einem Subjekt auch ein *Objekt* bzw. einen Gegenstand (Handlungen und Handlungsfolgen), wofür Verantwortung übernommen wird. Verantwortungsobjekte sind immer vergangene oder zukünftige – sind Teil retrospektiver oder prospektiver Verantwortungskonstellationen. Die *Instanz* stellt neben Subjekt und Objekt die bekannteste und am wenigsten hinterfragte Relation der Verantwortung dar. Solange die Fähigkeit, Verantwortung tragen zu können, an Persönlichkeit geknüpft ist, kommen im eigentlichen Sinne weder unbelebte Gegenstände, Pflanzen, Tiere noch Kleinkinder als potenzielle Instanzen in Betracht. Die* *Adressat*in* der Verantwortung stellt ein häufig umstrittenes Relationselement dar und die in meinen Augen in der Tat unterschätzteste Relation der Verantwortung. Sie* ist das Gegenüber des Verantwortungssubjekts, die* Betroffene der fraglichen Verantwortlichkeit und definiert den Grund für das Vorhandensein derselben. Sie* muss nicht – im Gegensatz zur Verantwortungsinstanz – selbst potenzielle*r Verantwortungsträger*in sein können, hat nicht selbst die Kompetenzen für die Möglichkeit einer Verantwortungszuschreibung mitzubringen. Die *normativen Kriterien* stellen das Inwiefern, den Maßstab und normativen Bezugsrahmen dar, nach dem in einem gegebenen Kontext darüber geurteilt wird, ob die fragliche Person verantwortlich gehandelt hat. Sie definieren Verantwortungsbereiche, in denen jemand – begrenzt durch Normen – Rede und Antwort steht. Je nach Kontext, abhängig von dem Set an Kriterien, die der fraglichen Verantwortlichkeit zugrunde liegen, handelt es sich dabei um bspw. den strafrechtlichen, politischen, moralischen oder wirtschaftlichen Raum und demzufolge um eine strafrechtliche, politische, moralische oder wirtschaftliche Verantwortung.

2. Was ist Roboterethik?

Innerhalb der noch jungen Bereichsethik der Roboterethik sind zwei Felder zu unterscheiden: Die einen fragen danach, inwiefern Roboter als sogenannte „moral patients“ zu verstehen sind, also passiv als Träger*innen moralischer Rechte, bzw.

⁴ Ursprünglich stammt der Terminus von Christian Neuhäuser, der ihn allerdings nicht genauer definiert hat (Neuhäuser 2014).

inwiefern ihnen ein moralischer Wert zukommt. Die anderen interessieren sich dafür, ob und ggf. inwiefern Roboter sogar „moral agents“ sein könnten, also aktiv Träger*innen moralischer Pflichten bzw. moralische Handlungssubjekte. Beide Arbeitsbereiche ergänzen einander und Verantwortung ist in beide jeweils unterschiedlich einzubinden, wie in 3.1 und 3.2 gezeigt wird. Die Gruppe der moral agents ist gegenüber der der moral patients exklusiver; für gewöhnlich zeichnen wir nur Menschen (und längst nicht alle) mit Moralfähigkeit im genuinen Sinne des Wortes aus – einige Menschen wie bspw. Kinder und solche mit spezifischen geistigen und körperlichen Einschränkungen können temporär oder sogar generell von ihrer Moralfähigkeit ganz oder teilweise entschuldigt werden.

Einer ganzen Reihe von Wesen und Dingen wie z.B. Tieren, Pflanzen, aber auch Gegenständen wie dem teuren Auto, dem Smartphone oder Haus wird indes ein moralischer Wert zugeschrieben – zumindest in dem Sinn, dass diese Entitäten moralisch bedenkenswert sind, wenn ihnen vielleicht auch kein Eigen- sondern nur ein hoher instrumenteller Wert beigemessen wird. Als moralisches Handlungssubjekt hat man zugleich einen Platz im Kreis der Wertträger*innen – dies gilt allerdings nicht umgekehrt. Lebewesen und Gegenständen kann man abhängig von der Perspektive einen moralischen Wert zuschreiben; eine anthropozentrische Position argumentiert bspw. dafür, dass nur dem Menschen ein Eigenwert zukommt, weitere Ansätze stellen der Patho-, der Bio- und der Physiozentrismus dar.

Innerhalb des Arbeitsbereichs zu Robotern als Wertträger*innen wird das menschliche Verhalten gegenüber artifiziellen Systemen in den Blick genommen. Hier geht es darum, wie mit Robotern umzugehen ist und inwiefern ihnen (ggf. analog zu Tieren und kleinen Kindern) ein moralischer Wert zukommt, selbst wenn man sich darüber einig sein sollte, dass sie selbst nicht zu moralischem Handeln in der Lage sind. In dieses Themenfeld fallen alle Fragen, die artifizielle Systeme als Werkzeuge oder als Ergänzungen des Menschen verstehen, wie bspw. bei der Formulierung von Ethikkodizes in Unternehmen die Frage, inwiefern Beziehungen zu und mit Robotern denkbar und wünschenswert sind, inwiefern man Roboter ‚versklaven‘ kann und wie der Einsatz von artifiziellen Systemen zu Therapiezwecken zu beurteilen ist. Innerhalb dieses Arbeitsbereichs verbleibt die moralische Kompetenz und Kompetenzkompetenz bei den menschlichen Designer*innen (und u.U. auch Nutzer*innen) artifizieller Systeme. Die menschlichen ‚Eltern‘ entscheiden über die Moral ihrer Geschöpfe und darüber, wer im Falle eines Unfalls Verant-

wortung trägt. Auf die Möglichkeiten einer Verantwortungszuschreibung in solchen Fällen reduzierter oder nicht vorhandener Verantwortungskompetenzen (Kommunikations- und Handlungsfähigkeit sowie Urteilskraft) komme ich unter dem Schlagwort der Verantwortungsnetzwerke in Abschnitt 3.2 zu sprechen.

Innerhalb des Arbeitsfelds zu Robotern als moralischen Handlungssubjekten fragt man bspw. danach, inwiefern Roboter zu moralischem Handeln in der Lage sind, und folglich, über welche Kompetenzen sie in welchem Maße dafür verfügen müssen. Interessieren sich die einen in diesem Bereich eher für die Zuschreibung von Freiheit als Bedingung für moralisches Handeln, befassen sich andere eher mit kognitiven Kompetenzen (Denken, Verstehen, Geist, Intelligenz, Bewusstsein, Wahrnehmung und Kommunikation) und wieder andere mit Empathie und Emotionen (hierzu Abschnitt 3.1). Beiden Arbeitsfeldern innerhalb der Roboterethik liegt die Frage zugrunde, was Moral bzw. was Ethik ist und wie moralische Urteile gefällt werden.

3. Verantwortung in der Mensch-Maschine-Interaktion

Häufig wird die Möglichkeit einer Verantwortungsübernahme von artifiziellen Systemen mit dem Verweis auf die 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. Wendell Wallach und Colin Allen formulieren in *Moral Machines. Teaching Robots Right from Wrong* (2009) den Ansatz der funktionalen Äquivalenz, mit dem dieses Problem mangelnder Kompetenzen bei artifiziellen Systemen umgangen werden kann. Die beiden folgenden Abschnitte nehmen die Rolle der Verantwortung in den beiden Arbeitsfeldern der Roboterethik – Roboter als moral agents bzw. Handlungssubjekte (Abschnitt 3.1) und Roboter als moral patients bzw. Handlungsobjekte (Abschnitt 3.2) – in den Blick.

3.1 Roboter als Handlungssubjekte – Wallachs und Allens Ansatz funktionaler Äquivalenz

Indem Wallach und Allen die Frage stellen, inwiefern Roboter als artifizielle moralische Akteure zu verstehen sind, definieren sie „moral agency“ als graduelles Konzept mit zwei Bedingungen, nämlich Autonomie und Empfänglichkeit bzw. Emp-

findlichkeit für moralische Werte („sensitivity to values“; 2009, S. 25). Menschen gelten als moralische Akteure im genuinen Sinne, allerdings sind einige Maschinen – bspw. ein Autopilot, oder das artifizielle System Kismet – operationale moralische Akteure. Sie sind autonomer und ethisch empfänglicher als manch ein anderes nicht-mechanisches Werkzeug wie z. B. ein Hammer, und dennoch verbleiben sie immer noch „totally within the control of [the] tool’s designers and users“ (ebd., S. 26). Nur besondere artifizielle Systeme haben bereits den Status funktionaler moralischer Akteursfähigkeit – so wie bspw. das medizinische ethische Expertensystem MedEthEx. Funktionale Moralität bedeutet, dass das fragliche artifizielle System insofern entweder autonomer und/oder Werte-sensitiver ist, als ein operationaler moralischer artifizieller Akteur, als funktionale moralische Maschinen „themselves have the capacity for assessing and responding to moral challenges“ (ebd., S. 9).

Mit ihrem Ansatz funktionaler Äquivalenz einer graduellen Zuschreibung von Kompetenzen und Fähigkeiten beschreiben Wallach und Allen eine Version der schwachen KI-These, der an der Simulation spezifischer Kompetenzen in artifiziellen Systemen gelegen ist und nicht daran, Maschinen tatsächlich im genuinen Sinne des Wortes mit Intelligenz, Bewusstsein und Autonomie zu konstruieren (starke KI, irrtümlich auf Turing zurückgeführt). Funktionale Äquivalenz bedeutet, dass spezifische Phänomene verstanden werden, ‚als ob‘ sie – um das Kant’sche Vokabular der regulativen Ideen zu nutzen – kognitiven, emotionalen oder anderen Kompetenzen und Fähigkeiten entsprechen. Es sei an dieser Stelle daran erinnert, dass dieses Argument bereits auf Menschen zutrifft, spätestens allerdings auf Tiere. Bezüglich anderer Menschen sind wir generell bereit, prima facie Fähigkeiten wie Vernunft, Bewusstsein und Willensfreiheit zu statuieren, wenn auch keine Garantie dafür besteht, dass die fraglichen Individuen tatsächlich mit besagten Kompetenzen ausgestattet sind. Die Frage, inwiefern artifizielle Systeme irgendwann in der Tat intelligent, bewusst oder autonom im Sinne der starken KI-These genannt werden können, wird durch die Frage ersetzt, in welchem Ausmaß und Umfang die fraglichen Kompetenzen der Funktion entsprechen, die sie innerhalb der moralischen Evaluation spielen – in diesem Fall dem Verantwortungskonzept.

Obwohl Wallach und Allen den Übergang von operationaler über funktionale bis hin zu voller bzw. genuiner Moralzuschreibung in Abhängigkeit von den vorliegenden Kompetenzen (Autonomie und moralische Sensitivität) graduell denken, bleibt es doch schwer vorstellbar, wie (zumindest in der nahen Zukunft) ein artifi-

zielles System ein funktionales Äquivalent zu der genuin menschlichen Fähigkeit „second order volitions“ (Frankfurt 1971) zu bilden entwickeln könnte bzw. zu der Fähigkeit, als „self-authenticating sources of valid claims“ (Rawls 2001, S. 23) über selbst gesetzte moralische Prämissen und Prinzipien zu reflektieren.

Schauen wir uns zwei Beispiele an: Der Roboter Cog ist als Beispiel für einen in sehr schwachem Maße funktional verantwortlichen Akteur zu sehen, insofern seine Kommunikationsfähigkeit und seine Urteilkraft deutlich ausgeprägt sind. Und was noch wesentlicher erscheinen mag – seine Handlungsfähigkeit bzw. Autonomie ist aufgrund eines „unsupervised learning algorithm“ (Brooks et al. 1999, S. 70) sehr komplex. So beginnt Cog, ohne, dass er zuvor in dieser Weise programmiert worden wäre, ein Spielzeugauto nur noch von vorne oder hinten anzustoßen, um es in Bewegung zu versetzen, nachdem er durch mehrere Versuche feststellen konnte, dass es sich nicht bewegt, wenn es von der Seite angestoßen wird. Cog lernt durch Erfahrung, und vielleicht ist es gerade diese (in seinem Fall in der Tat begrenzte) Fähigkeit zu lernen, die es uns erlaubt, ihn als einen schwachen funktionalen Akteur zu verstehen oder aber als immerhin stark operational verantwortlich. Cog verantwortungsbefähigt zu nennen, ist wohl vergleichbar damit, einem jungen Kind Verantwortung zuzuschreiben.

Autonome Fahrassistenzsysteme lassen sich hingegen als ein Beispiel für operational verantwortliche Akteure anführen, denn zwar mögen ihre Kommunikationsfähigkeiten und Urteilkraft ähnlich entwickelt sein wie die von Cog oder sogar weiter. Allerdings ist die Handlungsfähigkeit bzw. Autonomie autonomer Fahrassistenzsysteme aus guten Gründen in strengen Grenzen gehalten; sie können nicht lernen.

Insofern es um verschiedene Bereiche der Verantwortungszuschreibung geht (moralische, rechtliche, politische Verantwortung), ist ein artifizielles System dann autonom (bzw. handlungsfähig) zu nennen, wenn es die Kriterien funktionaler Moralzuschreibung erfüllt. Eine generelle Modifikation der implementierten algorithmischen Strukturen ist wohl bei keinem artifiziellem System im selben Ausmaß wie im Rahmen der menschlichen Entwicklung vorstellbar, von der Wünschbarkeit ganz zu schweigen. Trotzdem mutet vor dem Hintergrund des gerade Gesagten der Einwand trivial an, dass artifizielle Systeme letztlich immer programmiert sind. Denn auch Menschen sind für gewöhnlich nicht zu allem in der Lage, sondern bleiben in ihren Möglichkeiten ebenfalls beschränkt, selbst wenn man ihren adapti-

ven Spielraum sehr viel größer einschätzt als der eines noch so komplexen Roboters jemals sein könnte.

Die besten Aussichten auf artifizielle Verantwortungszuschreibung gewähren evolutionäre Lernmodelle; maschinelles Lernen wird hier äquivalent zum kindlichen Lernen untersucht. Diese Ansätze beruhen auf einer meta-ethischen Annahme über die Kontextsensitivität von Moral. Moralisches und verantwortliches Handeln bedarf der Erfahrung und eines situativen Urteilsvermögens. Beides kann sich ein artifizielles System nur verkörpert aneignen. In den 1990er Jahren war es u.a. Brooks, der als einer der ersten das Zusammenwirken von artifiziellem System und Umwelt als Bedingung für die Entwicklung von Vermögen und Kompetenzen betrachtete und von dieser Annahme ausgehend das Feld der „behavior-based robotics“ begründete. Zahlreiche Projekte, die sich an dem Ansatz verkörperten menschlichen Lernens orientieren – wie bspw. die Lernplattformen iCub, Myon, Cb², Curi, Roboy (die im Detail unterschiedlichen evolutionsbasierten Ansätzen folgen) –, entwickeln Systeme, die sich ähnlich Kindern Kompetenzen aneignen, aus denen sie dann in spezifischen Kontexten konkrete Handlungsprinzipien ableiten. Bislang ist maschinelles Lernen jedoch nur in nicht moralischen bzw. schwach moralischen Kontexten möglich. Von Verantwortungsübernahme ließe sich also bislang nur in einem nicht (rein) moralischen (vielleicht juristischen) Sinne sprechen.

3.2 Roboter als Handlungsobjekte – Verantwortungsnetzwerke

Im vorherigen Abschnitt wurde über Wallachs und Allens Ansatz funktionaler Äquivalenz festgestellt, dass artifizielle Systeme bislang nicht als Verantwortungsakteure zu identifizieren sind, insofern die zur Verantwortungszuschreibung nötigen Kompetenzen (Kommunikations- und Handlungsfähigkeit, sowie Urteilskraft) nur in einem schwach funktionalen oder gar nur in einem operationalen Sinne äquivalent simuliert werden können. Da wir Roboter gegenwärtig nicht im exklusiven Kreis der moral agents finden, stelle ich nun ein paar Überlegungen dazu an, welche Rolle dem Verantwortungsphänomen innerhalb des Arbeitsfeldes der Robotereethik zu artifizialen Systemen als moral patients zuzugestehen ist.

Wie bereits festgestellt, ist unser traditionelles Verständnis von Verantwortung insofern ein stark individualistisches, als wir immer ein Subjekt benötigen, das als Verantwortungsträger*in fungiert. Das ist auch bei der Zuschreibung kollektiver