

Künstliche Intelligenz, Autonomie und Verantwortung

Erkundungen im maschinen- und roboterethischen Reflexionskontext

Johannes J. Frühbauer

Hashtags #Ethik #Anthropologie
#Steuerung #Hermeneutik #Autonomie
#Verantwortung #Intelligenz
#MenschMaschine

Abstract Maschinen- und Roboterethik sind als neuere Fortführungen der Technikethik zu sehen, die sich mit moralischen Fragen in der Anwendung von Technik und deren Folgen befasst. Sie befassen sich mit ethischen Fragen, die sich insbesondere aus konkreten KI-Anwendung ergeben. KI ist zu verstehen als vielseitige digitale Programmierung und Steuerung in Alltagsgeräten und insbesondere in maschinellen Systemen, die Roboter sein können und die ein selbstgesteuertes Agieren ermöglichen. KI lässt sich unter anderem in einen so genannten Konnektionismus und in eine verhaltensbasierte Künstliche Intelligenz unterscheiden. Zu den ethischen Kernfragen im Kontext von KI gehören etwa in den Ansätzen der Philosophin-

nen Catrin Misselhorn und Janina Loh vor allem die Frage nach dem moralischen Status von Maschinen beziehungsweise Robotern: Sind sie lediglich moralische Objekte oder lassen sie sich auch als moralische Subjekte qualifizieren, also als moralisch zurechenbare Entscheidungsakteure und Handlungsträger? Diese Zuordnung hängt wesentlich mit der Frage der Zuschreibung von Autonomie an Maschinen und Roboter zusammen. Um sachlich fundiert eine Antwort auf diese Frage formulieren zu können, wird im Rekurs auf die Philosophin Beate Rössler das philosophisch-ethische Verständnis von Autonomie näher beleuchtet und ausdifferenziert. Schlussendlich werden fünf Thesen formuliert, die sich unter anderem mit der Frage nach der Zurechnung von Verantwortung, dem Bedarf einer neuen Ethik und der Perspektiven im Verhältnis von Mensch und Maschine befassen.

1 Digitalisierung und Künstliche Intelligenz: Fortführung der Technikethik durch Maschinenethik und Roboterethik

In den Diskurskontext zur Digitalisierung gehört auch alles, was mit Künstlicher Intelligenz¹ zu tun hat.² Denn Digitalisierung stellt letztlich die elementare technische Voraussetzung für alle Entwicklungen und Anwendungen von KI dar. In diesen zur eigentlichen Fragestellung hinführenden Überlegungen werde ich alles zur Seite schieben, was ohnehin schon als bekannt vorausgesetzt werden kann, was andernorts zum Teil umfangreich darlegt ist und an dieser Stelle den weiteren Fortgang nur unnötig aufhalten würde. Das heißt ich werde mich hier nicht mit überaus detaillierten begrifflichen Klärungen zu Digitalisierung und allzu vertieft mit KI und ihren Entwicklungs- und Anwendungskontexten befassen. Und ich werde mich auch nicht mit dem immer wieder anzutreffenden skeptischen Einwand auseinandersetzen, ob es so etwas wie KI überhaupt gibt. Im Mittelpunkt steht die ethische Perspektive mit einem klar formulierten Erkenntnisinteresse. Dabei soll es weniger um die bereits seit geraumer Zeit unter ethischen Gesichtspunkten diskutierte anwendungsbezogenen Grundfragen und Problematisierungen gehen, etwa um die moralischen Bewertungen von KI im Einsatz bei autonomen Fahr- oder Waffensystemen oder bei Pflegerobotern,³ sondern es soll vorrangig um metaethische und moralhermeneutische Klärungen gehen, die sich mit dem Verständnis und der Verwendung des ethischen Grundvokabulars befassen und nach dem moralischen Subjekt, insbesondere aber nach der Angemessenheit des Autonomiebegriffs und der damit verbundenen Verantwortungszuschreibung im Vorfeld und im Kontext von KI-Anwendungen fragen.⁴

Einige sachliche Ausgangspunkte seien hier dennoch kurz notiert: Ich gehe vom Faktum dessen aus, was gemeinhin als KI bezeichnet wird. Dabei schließe ich mich folgendem Grundverständnis von KI in der zunächst weichen Beschreibung von Manuela Lenzen an: Künstliche Intelligenz steht heute für ein weites, interdisziplinäres Gebiet digitaler Technik. Forscher*innen mit ganz unterschiedlichen Fragestellungen und Methoden arbeiten in erster Linie daran, technische bzw. maschinelle Systeme zu entwickeln, die in der Lage sind,

1 Die Großschreibung von Künstlicher Intelligenz hat sich etabliert; fortan verwende ich im Text die gängige Abkürzung KI.

2 Siehe zur Thematik auch den Beitrag von Matthias Kettner in diesem Band.

3 Vgl. hierzu exemplarisch: Misselhorn 2018: 136–204, zu Waffensystemen vgl. Werkner/Hofheinz 2019.

4 Vgl. hierzu auch den Beitrag von Thorsten Moos in diesem Band. Der Beitrag von Matthias Kettner in diesem Band hingegen nimmt Paradigmen der KI-Anwendung in den Blick und differenziert diese aus.

Sprache zu verwenden, Begriffe zu bilden, Probleme zu lösen. Etwas flapsig formuliert könnte man sagen: Künstliche Intelligenz zu entwickeln, bedeutet Maschinen zu konstruieren, die das können, was die Maschinen im Science-Fiction-Film bzw. in der Literatur bereits können. Sprachverarbeitung, Darstellung von Wissen in Datenverarbeitungssystemen, automatisches Schlussfolgern, Wahrnehmung und Bildanalyse, Robotik und maschinelles Lernen gehören unter anderem zu den konkreten Anwendungsgebieten der KI.⁵ Als Basisdisziplin für KI steht die Informatik im Mittelpunkt. In Alltagsanwendungen begegnet uns KI inzwischen in vielen Bereichen: Sie steckt in Apps, Suchalgorithmen, Robotern, Fahrassistenten oder Smart Watches. Diese digitale Technik wird immer intelligenter – was die einen mit Euphorie gutheißen und andere wiederum mit Sorge betrachten.⁶ Die Philosophin Catrin Misselhorn beschreibt in ihren Darlegungen zur KI den KI-Ansatz des *Konnektionismus*. Dieser unternimmt den Versuch mit Mitteln der Informatik ein Analogon der Neuronenverbände, die das menschliche Gehirn ausmachen, mit den Mitteln der Informatik zu konstruieren.⁷ Und Misselhorn betont:

»Es wäre allerdings ein Missverständnis davon auszugehen, dass künstliche neuronale Netze Gehirnstrukturen eins zu eins abbilden. Vielmehr handelt es sich um mathematische Modelle für Computerprogramme, die bestimmten Organisationsprinzipien biologischer neuronaler Netze nacheifern. Die computationalen Neurowissenschaften streben zwar an, die Funktionsweise des Gehirns mit Hilfe künstlicher neuronaler Netze zu verstehen und auf dem Computer zu simulieren. Doch für viele KI-Anwendungen kommt es nicht auf die biologischen Details an, sondern darauf, praktikable Lösungen für bestimmte Probleme zu finden.«⁸

Eine gegenüber dem Konnektionismus anders gelagerte Form ist die verhaltensbasierte KI. Diese bezieht nun auch den Körper und seine Umwelt in die Entwicklung und Anwendung von KI mit ein. *Verkörperung* und *Situiertheit der Kognition* lauten hier die zentralen Stichworte. Misselhorn zufolge stehen nicht mehr anspruchsvolle geistige Leistungen im Vordergrund,

»sondern sensomotorische Fähigkeiten, über die auch verhältnismäßig einfache Organismen wie Insekten verfügen. Dieser Ansatz bleibt folgerichtig nicht bei Computermodellen stehen, sondern führt zum Bau von Robotern, die über einen Körper, Sensoren und Aktoren verfügen. Ziel ist es, Systeme zu schaffen, die die Fähigkeit haben, ihre

5 Lenzen 2018: 23.

6 Misselhorn 2018: 17.

7 Misselhorn 2018: 22.

8 Misselhorn 2018: 22.

Umwelt wahrzunehmen, sich in ihr zu bewegen und Objekte zu manipulieren, so dass sie sich erhalten und unter Umständen sogar reproduzieren können.«⁹

Diese Form der Einbettung einer verhaltensbasierten KI widerfährt jedoch, wie Misselhorn ausführlich, grundlegende Kritik. Denn sie könne der Einbettung der menschlichen Intelligenz in eine Welt, die sich durch bestimmte Sinnstrukturen auszeichnet, nicht gerecht werden. Dieser Kritik zufolge sei menschliche Intelligenz nicht formalisierter, sondern hochgradig kontextabhängig und sie werde auch dadurch geprägt, dass der Mensch über einen Körper verfüge.¹⁰ Nun, Misselhorn zufolge kann man mit Blick auf die praktischen Ergebnisse des verhaltensbasierten KI-Ansatzes jedenfalls festhalten,

»dass dieser insbesondere im Hinblick auf die Ausstattung von Robotern mit sensorischen Fähigkeiten überzeugt. Betrachtet man hingegen das, was wir für gewöhnlich Denken nennen, als paradigmatischen Fall der Intelligenz, so kann der verhaltensbasierte Ansatz bislang keine überzeugenden Resultate vorweisen, und es bleibt weiterhin schwierig, sich den Weg dorthin auszumalen.«¹¹

Eine weiterführende und vertiefte Auseinandersetzung mit KI müsste sich über die skizzierten Aspekte hinaus mit der Bedeutung von Bewusstsein, Emotionen sowie mit der Frage nach künstlichen Emotionen befassen.¹² An dieser Stelle soll der Hinweis auf diese Begriffe und auf ihre Bedeutung für die ethische Reflexion der Problematik und für entsprechende Schlussfolgerungen genügen.¹³ Ganz allgemein lässt sich feststellen, dass die ethische Reflexion bereits seit einigen Jahrzehnten das Themenfeld der KI mit ihren Möglichkeiten, Verheißungen, ihren Risiken und Gefahren im Blick hat. Doch je mehr faktisch bereits möglich ist und in absehbarer Zeit möglich sein wird, desto mehr beschleunigt und intensiviert sich auch der ethische Diskurs und er differenziert sich weiter aus. Das offenbart nicht zuletzt der Blick auf eine ganze Reihe an Publikationen, die gerade in den zurückliegenden Jahren erschienen sind.¹⁴ Offenkundig wirft KI eine Vielzahl an ethischen Fragen auf

9 Misselhorn 2018: 27.

10 Misselhorn 2018: 29.

11 Misselhorn 2018: 29. Vgl. zur Abgrenzung von menschlicher und Künstlicher Intelligenz den Beitrag von Kettner in diesem Band.

12 Vgl. zu dieser Fragestellung: Misselhorn 2021.

13 Vgl. Misselhorn 2019: 35–45. Misselhorn wirft unter anderem die Frage auf, ob sich für »emotionale Maschinen« gewisse moralische Ansprüche ergeben könnten und infolgedessen ihr moralischer Status dem von Tieren vergleichbar wäre. Misselhorn 2019: 45.

14 Vgl. hierzu insbesondere Misselhorn 2018; Loh 2018; Bendel 2019; Woopen/Jannes 2019; Brand 2018; Ramge 2018.

und nötigt zu ethischer Reflexion. Die damit verbundenen Analysen und Reflexionen sind in erster Linie in den Bereich der angewandten Ethik einzuordnen. Aus der Basisdisziplin einer Technikethik, die sich in den 1980er Jahren etabliert hat, haben sich inzwischen verschiedene neue Bereichsethiken entwickelt und ausdifferenziert. Dazu zählen in erster Linie Maschinenethik sowie Roboterethik. Was sie jeweils kennzeichnet, soll im weiteren Fortgang zumindest elementar verdeutlicht werden. Doch zunächst stellt sich die Frage, was eigentlich neu ist an diesen noch jungen Bereichsethiken gegenüber einer seit Jahrzehnten etablierten Technikethik. Die Technikethik setzt bei ihrem Referenzobjekt an. Dabei ist Technik zu verstehen als ein allgemeines Produkt menschlichen Handelns; infolgedessen wird Technik immer auch – ob bewusst oder unbewusst – durch Normen und Werte bestimmt.¹⁵ Angesichts der Herausforderungen einer Technikethik ließ sich vor geraumer Zeit noch folgende Einschätzung formulieren:

»Sowohl die technisch induzierten Naturbezüge als auch die Nutzenorientierung müssen normativ beurteilt und es muss problematisiert werden, ob sie dem technikimmanenten Prinzip der Sicherung und unseren Vorstellungen von einem guten und rechtfertigbaren Lebensvollzug entsprechen. Denn mit zunehmender Eingriffstiefe und Langfristigkeit der Folgen verändern sich nicht nur die Möglichkeitsräume fortsetzbaren Handelns und wird nicht nur unser Wissen um diese Veränderungen unsicher, sondern auch traditionelle Orientierungen, Welt- und Menschenbilder sehen sich verschwindenden oder neu eröffneten Bezugsbereichen gegenüber.«¹⁶

Diese technikethische Charakterisierung hat weiterhin Gültigkeit, hinzu kommen jedoch neue Perspektiven, insofern Maschinen oder Roboter nicht allein mehr als Referenzobjekte ethischer Reflexion mit Fokus auf ihre bloße Instrumentalität für menschliche Weltgestaltung betrachtet werden, sondern aufgrund ihrer spezifischen Akteurseigenschaft die Frage aufkommt, ob und inwiefern diese moralische Subjekte sein können. Dieser Aspekt führt nun unmittelbar zum Selbstverständnis und zur Aufgabenstellung der Maschinenethik bzw. Roboterethik.

15 Vgl. Loh 2019b: 9.

16 Hubig 2011: 170 f.

2 Grundzüge und Grundfragen der Maschinen- und Roboterethik

Maschinenethik und Roboterethik liegen inhaltlich sehr nahe beieinander und haben weithin dieselben Sachbezüge. Grundsätzlich gilt: Jeder Roboter ist eine Maschine, aber nicht jede Maschine ist ein Roboter. Diese sachlich-technische Unterscheidung markiert letztlich auch eine gewisse Trennlinie zwischen einer Maschinenethik zum einen und einer Roboterethik zum anderen.¹⁷ Dennoch lässt sich beobachten, dass oftmals beide Bezeichnungen nebeneinander oder miteinander gebraucht werden, und es nicht immer eindeutig ersichtlich ist, warum in dem einen Fall von Maschinenethik und im anderen Fall von Roboterethik die Rede ist. In jedem Fall sind beides Typen angewandter Ethik und lassen sich als spezialisierte Fortführung der Technikethik sowie als Subdisziplinen einer digitalen Ethik, die aufgrund der Vielzahl ihrer Themen und Fragestellungen deutlich weiter gefasst ist, verstehen.¹⁸ Der Sachgegenstand der Maschinenethik sind Maschinen, die sich wiederum als technische Apparaturen oder Geräte mit einem wesentlichen Bezug zur Digitalisierung und somit zur Informatik und zur Computertechnologie verstehen lassen. Maschinen herkömmlicher Art – wie sie etwa im Baugewerbe, in der Landwirtschaft, bei Transport und Verkehr zum Einsatz kommen – sind somit nicht im Blick der Maschinenethik. Als Form der angewandten Ethik fragt die Maschinenethik unter anderem, wie es zu bewerten ist, wenn Maschinen mit moralischen Fähigkeiten ausgestattet werden (könnten) und was die technischen Voraussetzungen für eine solche Ausstattung wären. Dabei gilt die Reflexion nicht nur der eigentlichen Maschine, sondern sie besitzt auch eine weiter gefasste gesellschaftliche Dimension, insofern Menschen und Gesellschaften von den Auswirkungen dieser technischen Entwicklungen und Möglichkeiten betroffen sind.¹⁹

Der Sachgegenstand der Roboterethik sind Roboter. Vermutlich werden unsere alltäglichen Vorstellung von einem Roboter durch deren tatsächliche

17 Vgl. zu den Unterscheidungen: Bendel 2019a; Loh 2019a; Loh 2019b; Misselhorn 2018: 47.

18 Das *Handbuch Angewandte Ethik* aus dem Jahr 2011 kennt bemerkenswerterweise weder die eine noch die andere Disziplin. Einschlägige Fragestellung werden allenfalls unter den Einträgen zur ›Technikethik‹ oder zur ›Informationsethik‹ behandelt. – Vgl. Stoecker et al. 2011: 170–175 bzw. 210–115. Vgl. zu Konzeptionen einer ›Digitalen Ethik‹ unter anderem: Grimm et al. 2019, Spiekermann 2019, Nida-Rümelin/Weidenfeld 2020.

19 Ob man formulieren kann, wie Oliver Bendel dies tut, dass die Maschinenethik ein Gestaltungsprinzip sei und die maschinelle Moral bzw. moralische und unmoralische Maschinen hervorbringe, wage ich doch zu bezweifeln. Überhaupt wirft seine »Grundlegung« der Maschinenethik nicht nur die ein oder andere Frage auf, sondern lässt auch intellektuelles Unbehagen aufgrund fehlender Stringenz und Systematik – etwa im Vergleich zu den systematisch durchkomponierten Darstellungen von Loh oder Misselhorn – aufkommen (Bendel 2019b).

Vielfalt in unserer Zeit übertroffen. Und viele unserer Vorstellungen lehnen sich entweder an ferngesteuerte Spielzeugroboter an, beziehen sich auf elektronische Haushalts- und Gartengeräte (Staubsauger- oder Mähroboter) oder sind geprägt durch filmische Figuren mit humanoiden Erscheinungsformen wie beispielweise in »Westworld« (1973), als der »Terminator« (1984), als »RoboCop« (1987), in »I, Robot« (2004) oder in den im Phänotyp deutlich technischeren Variante wie etwa in »Nummer 5 lebt« (1986), als »Transformers« (2007) oder als »WALL·E« (2008). Doch was genau ist eigentlich gemeint, wenn von »Roboter« die Rede ist? Ein Roboter ist zunächst als technische Einheit bzw. System zu kennzeichnen, das bestimmte Aktionen ausführen kann. Somit lässt sich ein Roboter als Akteur:in kennzeichnen. In einem präzisierten Verständnis ist ein Roboter eine sensomotorische Maschine, die über mindestens drei unterscheidbare Ebenen freier Beweglichkeit verfügt.²⁰ Das Aussehen eines Roboters kann sich als Humanoid an Statur und Antlitz eines Menschen anlehnen. Sein Aussehen bzw. seine äußere Form kann aber auch sehr maschinenhaft sein, dann etwa wenn Roboter in der Industrieproduktion oder als Haushaltsassistenten eingesetzt werden. Aus ethischer Sicht werden Roboter dann zu einem bedeutsamen Reflexionsgegenstand, wenn sie aufgrund ihrer technischen Ausstattung, ihrer Funktionseigenschaften und den Möglichkeiten der KI weithin selbstständig agieren können und somit als »autonom« handelnde Subjekte bezeichnet werden. Mit Robotern sind ganz allgemein Hoffnungen, Erwartungen, aber auch Befürchtungen verbunden. Nimmt man übergeordnete gesellschaftliche Zwecke in den Blick, so verbindet sich mit Robotern die Erwartung auf eine weitere Humanisierung der Arbeitswelt oder auf Erleichterungen und Entlastungen bei Tätigkeiten im Alltag (z. B. durch Rasenroboter, Staubsaugerroboter). Vieles davon ist bereits Alltagsrealität, nicht immer ist hier jedoch eine ausgetüftelte KI im Spiel. Die angedeuteten allgemeinen Befürchtungen verbinden sich mit Bedrohungs- und Gefährdungspotenzialen, die für den Menschen von KI und von selbstständig agierenden Maschinen oder Robotern ausgehen.

Die Philosophin Janina Loh stellt nun ihre Studien dezidiert unter die Überschrift der Roboterethik. Dabei unterscheidet sie in dieser zwei zentrale Arbeitsfelder: In dem einen Arbeitsfeld geht es um die Frage, inwiefern Roboter selbst als moralische Akteur*innen und damit als *Subjekte moralischen Handelns* begriffen werden müssen. Mit anderen Worten: Können sie sogenannte *moral agents* sein? In dem anderen Arbeitsfeld geht es darum zu klären, inwieweit sich Roboter als *moral patients* verstehen lassen: Sind sie Wert- oder möglicherweise sogar Rechtsträger, also in einem ganz allgemeinen Sinne *Objekte moralischen Handelns*? Und wenn ihnen ein Wert zukommt: Welche mo-

20 Vgl. diesbezüglich den Beitrag von Thorsten Moos in diesem Band.

ralischen Implikationen können aus dieser Wertzuschreibung hervorgehen? Eine vergleichbare Frage stellt sich beispielsweise mit Blick auf nichtmenschliche Entitäten der Natur bzw. der natürlichen Mitwelt, also etwa hinsichtlich Pflanzen und Tieren. Die Grundeinschränkung lautet, dass diese Entitäten nicht zum moralischen Handeln im herkömmlichen Sinne in der Lage sind. Diese beiden zentralen Arbeitsfelder sind nun nicht exklusiv, sondern komplementär zu verstehen.²¹

Worum geht es im ersten roboterethischen Arbeitsfeld genauer betrachtet? Werden Roboter als potentielle *moral agents* betrachtet, so wird danach gefragt, ob und in welcher Weise sie zu einem eigenständigen moralischen Handeln in der Lage sind und welche Voraussetzungen hierfür erfüllt sein müssen. Über welche Fähigkeiten müssen sie verfügen, damit ihnen die Qualität eines moralischen Subjektes zugeschrieben werden kann? Moralischen Subjekten lassen sich normalerweise Eigenschaften wie Freiheit oder Autonomie als Handlungsbedingung zuschreiben, hinzu kommen kognitive und emotionale Kompetenzen.²² Es scheint zumindest auf den ersten Blick fragwürdig, ob Robotern solche Eigenschaften zuzuschreiben sind. Doch erst wenn diese Eigenschaften vorhanden und infolgedessen die Voraussetzungen überhaupt erfüllt sind, lässt sich letztlich ein Handeln oder Funktionieren im engeren Sinne moralisch bewerten, erst dann wäre es gerechtfertigt von moralischen Subjekten im Sinne von Handlungsträgern zu sprechen. Wendell Wallach und Colin Allen schlagen zum Beispiel vor, allen Wesen dann Moralfähigkeit zuzuschreiben, wenn es naheliegend oder ersichtlich ist, dass sie in Situationen geraten können, in denen moralische Entscheidungen zu treffen sind.²³ Werden Systeme jedoch gezielt dazu programmiert, in bestimmten Situationen Entscheidungen zu treffen, ist damit noch kein im genuinen Sinne moralisches Handeln gemeint, selbst wenn die Entscheidung moralische Implikationen bzw. Folgen mit moralischer Tragweite mit sich bringt. Wenngleich die Akteursfähigkeit in ihrer moralischen Dimension alles andere als sicher gilt, so geben Szenarien und aus heutiger Sicht zumindest denkbare Möglichkeiten, Anlass genug, um ein ethisches Nachdenken über Roboter als *moral agents* bzw. moralische Subjekte plausibel und nachvollziehbar erscheinen zu lassen,

»ohne dass man sich gleich zu schließen gezwungen fühlen müsste, dass autonome artifizielle Systeme in derselben Weise wie Menschen zu moralischen Handeln befähigt seien, da sie mit den dafür nötigen Kompetenzen ausgerüstet wären.«²⁴

21 Loh 2019b: 35.

22 Loh 2019b: 35.

23 Vgl. Wallach/Allen 2009.

24 Loh 2019a: 79.

Kennzeichnend ist für den roboterethischen Diskurs, dass Autonomie für zahlreiche Ansätze zur moralischen Akteursfähigkeit von Robotern zentral ist. Loh zufolge könne Autonomie auch negativ definiert werden, insofern die Abwesenheit von äußerem Zwang oder direkter äußerer Kontrolle hervorgehoben werde. Dies sei für manche bereits hinreichend, um bestimmten Maschinen, auf die diese Abwesenheit zutrefte, rudimentäre Freiheit zuschreiben zu können. Lässt sich Autonomie in dieser und in ähnlichen Verständnisweisen als ein graduelles Konzept kennzeichnen? Maschinen bzw. Roboter könnten demzufolge mehr oder weniger »autonom« sein und somit auch in einem mehr oder weniger ausgeprägten Maße handlungsfähig. Im Vergleich zum Menschen als genuin moralischem Akteur*in wären Maschinen und Roboter als sehr viel schwächere Handlungssubjekte zu begreifen. Um einer weitergehenden Autonomie und moralischer Sensitivität eine Perspektive zu eröffnen, führen Wallach und Allen in diesem Denkkontext den Begriff der *funktionalen Moralität* im Sinne einer graduellen Zuschreibung von Kompetenzen und Fähigkeiten von Maschinen und Robotern ein. Diese funktionale Moralität beruht auf dem Prinzip der funktionalen Äquivalenz:

»Funktionale Äquivalenz bedeutet, dass spezifische Phänomene verstanden werden ›als ob‹ sie kognitiven, emotionalen oder anderen Kompetenzen und Fähigkeiten entsprechen, und beruht auf der Unterscheidung zwischen starker und schwacher Künstlicher Intelligenz. Starke KI [...] meint Maschinen, die im genuinen Sinne des Wortes mit Intelligenz, Bewusstsein und Autonomie ausgerüstet sind. Schwache KI ist lediglich an der Simulation spezifischer Kompetenzen in artifiziellen Systemen gelegen.«²⁵

Wallach und Allen denken sich den Übergang von operationaler über funktionaler bis hin zu voller Moralzuschreibung grundsätzlich graduell – in Abhängigkeit der tatsächlich vorhandenen »Autonomie« und moralischen Sensitivität. Völlig offen ist und bleibt jedoch die Frage, inwiefern Roboter (oder Maschinen) irgendwann tatsächlich intelligent, bewusst oder »autonom« im Sinne der starken KI-These genannt werden können. Für Loh ist es derzeit schwer vorstellbar, »wie ein artifizielles System ein funktionales Äquivalent zu der menschlichen Fähigkeit, höherstufige Wünsche bilden zu können, entwickeln könnte.«²⁶ Dies gilt ebenso für die selbstbestimmte Angabe von Entscheidungs- und Handlungsgründen vor dem Hintergrund eines selbstgewählten moralischen Überzeugungssystems. Loh verweist noch auf weitere Differenzierungen, die es in Verbindung mit dem »Autonomie-Begriff« gebe; so unterscheidet der Philosoph Stephen L. Darwall etwa zwischen

25 Loh 2019a: 80.

26 Loh 2019a: 81.

personaler, moralischer, rationaler oder handlungsbezogener »Autonomie«.²⁷ Dies legt nun unmittelbar die Spur, um im Folgenden den Autonomiebegriff in einer philosophisch-ethischen Perspektive genauer zu beleuchten.

3 Die philosophisch-ethische Konzeptualisierung von Autonomie

Wie bereits dargelegt findet die Verwendung des Attributs ›autonom‹ bzw. des Substantivs ›Autonomie‹ in den maschinenethischen und roboterethischen Diskursen eine regelmäßige und, wie ich kritisch anmerken möchte, oftmals eine allzu selbstverständliche Verwendung, die sich offenkundig etabliert hat. Da ich gerade diese Selbstverständlichkeit aufgrund semantisch-konzeptioneller Bedenken infragestellen und schlussendlich diese Verwendung des Autonomiebegriffs zurückweisen möchte, auch und gerade wegen der weiteren ethischen Implikationen wie etwa die Anbindung eines Verantwortungskonzeptes an den Begriff der Autonomie, werde ich mich in diesem Abschnitt mit einer systematischen Konzeptualisierung von Autonomie befassen. Dabei geht es mir vor allem um ein gehaltvolles normatives Verständnis von Autonomie. In den nachfolgenden Ausführungen hierzu werde ich mich in erster Linie auf die instruktiv-systematische Darstellung des ethischen Konzepts von Autonomie beziehen, wie es die Philosophin Beate Rössler in einer ihrer neueren Schriften fundiert und differenzierend ausgearbeitet hat.²⁸

Im Rahmen ihrer begrifflichen Klärung betont Rössler, dass Autonomie für uns Menschen deshalb Bedeutung hat,

»weil wir für unser Leben und für einzelne Handlungen nur Verantwortung übernehmen können, wenn wir sie – zumeist – selbst bestimmt haben und es tatsächlich in einem emphatischen Sinne die eigenen Handlungen sind, die wir vollziehen, die eigenen Pläne, die wir verfolgen, und Vorhaben, die wir umzusetzen trachten. Würden wir manipuliert oder gezwungen, dann könnten wir nicht aus eigenen Gründen handeln, wären es nicht unsere eigenen Werte und Überzeugungen, die den Rahmen unserer Handlungen und Vorhaben darstellen. Mehr noch, wir könnten uns nicht als verantwortlich verstehen für unser Leben als unser eigenes, und wir fühlten uns gegebenenfalls entfremdet von uns selbst.«²⁹

27 Vgl. Loh 2019a: 81.

28 Rössler 2019.

29 Rössler 2019: 29 f.

Rössler nimmt bereits hier vorweg, worauf ihre Studie unter anderem hinausläuft: Nämlich auf die logisch-korrespondierende Verbindung zwischen Autonomie zum einen und Verantwortung zum anderen. Autonomie ist eine notwendige Voraussetzung für die Zurechnung bzw. Übernahme moralischer Verantwortung. Einem Individuum kann in einem moralischen Sinne nur Verantwortung zukommen, wenn es in seinem Entscheiden und Handeln in einem ethisch normativen Sinne autonom ist. Bevor jedoch die Frage nach der moralischen Verantwortung noch etwas genauer in den Blick genommen wird, gilt unsere Aufmerksamkeit zunächst dem grundlegenden Verständnis von Autonomie.

Ganz elementar lässt sich Autonomie als individuelle Selbstbestimmung beschreiben. Autonomie im Sinne der Selbstbestimmung ist insbesondere deshalb von Wert und Bedeutung für uns, weil sie sich konstitutiv erweist sowohl für die Selbst- als auch für die Weltaneignung. Zu dieser Selbstbestimmung gehört, »dass wir darüber nachdenken können, was wir wirklich wollen im Leben, dass wir uns reflektierend zu unseren Wünschen und Überzeugungen verhalten können.«³⁰ Die individuelle Selbstbestimmung wichtiger Dimensionen des Lebens ist die leitende Idee für ein ethisches Verständnis von Autonomie.³¹ Diese ist als normatives Ideal nicht nur konstitutiv für unser eigenes Selbstverständnis, sondern – zumindest im Kontext westlich-liberaler Gesellschaften – auch für unsere grundlegende Vorstellung von Recht und Politik.³² Und nur wenn wir eine klare Vorstellung von Selbstbestimmung haben, können wir auch das Misslingen und Scheitern von Selbstbestimmung beschreiben: »Nur in ihrem Kontrast zur normativen Idee von Autonomie können Kontingenzen, Verpflichtungen, psychologische Unfähigkeit und strukturelle Hindernisse als solche beschrieben werden.«³³ Die Möglichkeit zu einem autonomen Leben schließt nicht aus, dass es Situationen in unserem Leben gibt, die wir gerade nicht gewählt haben und die von Schicksal, Zufall oder Unvorsichtigkeit geprägt sind.³⁴ Rössler betont zudem, dass »Ambivalenzen, Selbstentfremdung, Intransparenz des eigenen Selbst, autonomieerschwerende oder -verhindernde Strukturen« zu unserem autonom gelebten Alltag dazugehören und uns mit Spannungen konfrontieren.³⁵ Die Konfrontation mit den Kontingenzen und sozialen Komplikationen des eigenen Lebens kann letztlich

30 Rössler 2019: 21, 15.

31 Rössler 2019: 21.

32 Rössler 2019: 13, 21.

33 Rössler 2019: 21.

34 Rössler 2019: 14.

35 Rössler 2019: 21.

soweit führen, dass Zweifel an dessen Bestimmbarkeit aufkommen können.³⁶ Ein autonomes, selbstbestimmtes Leben bildet Rössler zufolge die Voraussetzung zu einem gelingenden Leben. Ein heteronomes, fremdbestimmtes Leben könnte niemals zu einem gelungenem, guten Leben führen.³⁷

In ihrer begrifflichen Ermittlung und konzeptionellen Entwicklung des Autonomiebegriffs geht Rössler auf verschiedene Theorieansätze im Autonomiediskurs ein. Sie unterscheidet dabei zwischen einer prozeduralen, internalistischen, historischen und relationalen Theorie der Autonomie im Rekurs auf unterschiedliche philosophische Autor:innen. Überdies sind weitere Unterscheidungen festzuhalten und mit Hinweischarakter kurz zu notieren: Nämlich die Unterscheidung zwischen einem lokalen und globalen Begriff von Autonomie, zwischen einer kategorischen und graduellen Zuschreibung von Autonomie sowie zwischen moralischer und personaler Autonomie. Das sind zumindest Fährten, die es sich im Kontext der Roboterethik bzw. Maschinenethik weiterzuerfolgen lohnen könnte. Aus der Konzeption des Autonomiebegriffs bei Rössler lassen sich meines Erachtens die folgenden Voraussetzungen, die für einen ethisch gehaltvollen Autonomiebegriff gegeben sein müssen, herausgreifen: Im Anschluss an Immanuel Kant die Zuschreibung von Würde und Vernunft an das moralische Subjekt, sodann die Fähigkeit einen eigenen Willen auszubilden, das Selbstverständnis sich als autonom zu begreifen, die Fähigkeit Gründe für autonomes Entscheiden und Handeln anzugeben und eine selbstevaluative Haltung einzunehmen sowie dialogisch-kontextuell moralische Überzeugungen zu konstituieren. Diese Merkmale lassen sich noch erweitern und ergänzen durch Begriffe wie Selbstachtung, Authentizität, Rationalität und Kohärenz sowie durch die Fähigkeit zur Planung von Projekten und Vorhaben.³⁸ Für Rössler ist autonomes Handeln stets ein Handeln auf der Basis von Gründen und nicht einfach nur auf der Basis von wie auch immer entstandenen Wünschen bzw. Wahloptionen.³⁹

Mit Blick auf Maschinen und Roboter lässt sich meines Erachtens kein authentisch selbstbestimmtes Entscheiden und Handeln im Sinne von Rösslers Autonomiekonzeption erkennen – entgegen aller klugen und weitsichtigen Reflexionen, die es zu dieser zumindest denkbaren Möglichkeit gibt. Was wir bei Robotern und Maschinen bis in die Gegenwart hinein antreffen, ist kein selbstbestimmtes Handeln, sondern letztlich ein programmiertes und damit manipuliertes »Entscheiden«, »Wählen« und »Agieren«. Kommen wir zu abschließenden Überlegungen.

36 Rössler 2019: 20.

37 Rössler 2019: 13.

38 Rössler 2019: 43–57.

39 Rössler 2019: 50.

4 Künstliche Intelligenz und die Notwendigkeit einer neuen Ethik? – Ein Fazit in fünf Thesen

Wie deutlich geworden sein dürfte, werden mit großer Selbstverständlichkeit in den Diskursen sowohl der Maschinenethik als auch der Roboterethik ethische Kernbegriffe wie Moral, Verantwortung, Freiheit oder Autonomie verwendet. Die vorausgehenden Ausführungen haben sich vor diesem Hintergrund in besonderer Weise mit dem Autonomiebegriff und damit verbunden auch andeutungsweise mit der Konzeption moralischer Verantwortung befasst. In den folgenden fünf Thesen werde ich versuchen, sowohl ein Fazit als auch Perspektiven für die weitere Auseinandersetzung zu formulieren.

These 1: Die Verwendung des Begriffs ›autonom‹ in der Rede von KI-gesteuerten Systemen, die gemeinhin als Maschinen oder Roboter bezeichnet werden, ist aus philosophischer Perspektive inadäquat und damit irreführend.⁴⁰ Der Begriff suggeriert ein moralisch-qualitatives selbstbestimmtes Handeln, das zu weiteren ethischen Implikationen wie etwa der Zuschreibung moralischer Verantwortung führt, das de facto aber nicht gegeben ist.⁴¹ Problematisch ist, dass sich diese Verwendung von ›autonom‹ bzw. ›Autonomie‹ fest etabliert hat und zur Selbstverständlichkeit geworden ist. Offenkundig besteht hier ein semantischer Dissens. Angezeigt wäre etwa gut begründet der Rede von »hochautomatisierten autoregulativen« Systemen zu folgen, wie dies in der jüngst erschienen EKD-Denkschrift »Freiheit digital« in bewusst-kritischer Abgrenzung zum Begriff »autonom« praktiziert wird.⁴²

These 2: Mit der Relevanz eines ethisch gehaltvollem Autonomieverständnis steht und fällt die Bedeutung bzw. Zuschreibung von Verantwortung – auch und insbesondere der Ausdifferenzierungen in Konzeptionen der Verantwortung wie Zuschreibbarkeit und Zurechnungsfähigkeit. Erst wenn ein moralisches Subjekt wirklich autonom entscheidet und handelt, erst und nur dann, ist es für sein Entscheiden und Tun wirklich moralisch verantwortlich.

These 3: Auch bei nicht vorhandener Autonomie von Maschinen und Robotern, macht es Sinn über differenzierende Verantwortungskonzeptionen nachzudenken und Perspektiven zu entwickeln, da für die Aktionswirkungen von

40 Arkin, der für »autonome« Waffensysteme plädiert, betont, dass er sein Autonomieverständnis der Robotik entnommen habe und nicht auf dessen philosophisch-ethischer Bedeutung abhebe. Arkin 2014, 11 Fn. 1.

41 Vgl. hierzu Rössler 2019: 387 ff.

42 Evangelische Kirche in Deutschland 2021: 133 f.

Maschinen und Robotern letztlich von einem bestimmbareren Verantwortungssubjekt, das zu wirklichen, moralisch qualifizierbaren Handlungen in der Lage ist, moralische Verantwortung übernommen werden muss. Interessant und zu weiterer Reflexion einladend ist der Gedanke, Verantwortungnetzwerke zu entwickeln, die den Beteiligten je nach Grad ihrer Mitwirkung graduell Verantwortung zuschreiben lassen.⁴³

These 4: Wie sich zeigt, ist es nicht erforderlich eine »neue Ethik«, die eine Zäsur zur bisherigen Prinzipien- und angewandten Ethik darstellen würde, zu etablieren. Was jedoch erkennbar geworden sein dürfte, ist, dass technische Entwicklungen nach wie vor zu neuen Herausforderungen für Mensch und Gesellschaft führen, die nach ethischen Erkundungen, Klärungen und letztlich Orientierungen verlangen. Offenkundig bedarf es grundsätzlich bei moralischen Herausforderungen durch neue Techniken in der Auseinandersetzung und Reflexion einer hohen Aufmerksamkeit für die Semantik ethischer Begriffe und ihrer präzisen Verwendung, um Missverständnisse, Ungenauigkeiten und ungewollte Folgeprobleme zu vermeiden. Die ethischen Erkundungen und Reflexionen der Gegenwart beziehen auch zukünftig Denkbare und Mögliches mit ein; das ist keine überflüssige Übung, sondern von wichtigem heuristischen und prospektiven Wert, vergleichbar mit dem Denken in Extremen.

These 5: Über die konkreten ethischen Herausforderungen und metaethischen Klärungen im Kontext von Maschinenethik und Roboterethik hinaus, rückt eine Frage immer mehr in den Mittelpunkt, nämlich das Verhältnis zwischen Mensch und Maschine – hier ist ein grundlegender und vielseitiger interdisziplinärer Diskurs gefordert, der allenfalls am Anfang steht. Kaum verwunderlich daher, dass der Deutsche Ethikrat erst jüngst eine Arbeitsgruppe unter diesem Titel eingerichtet hat. Auf deren Erkenntnisse und Ergebnisse darf man schon jetzt gespannt sein.⁴⁴ Dies soll jedoch keinesfalls zur Passivität und zum Abwarten einladen. Vielmehr ist Mit- und Vordenken gefordert.

43 Loh 2019a: 88.

44 Siehe: <https://www.ethikrat.org/anhoerungen/kuenstliche-intelligenz-und-mensch-maschine-schnittstellen/>.

Literaturverzeichnis

- Arkin, Ronald C. 2014: Vollautonome Waffensysteme und Kollateralofer. In: Ethik und Militär. Kontroversen der Militäretik und Sicherheitskultur, hg. v. Bock, Veronika u. a. 2014/1, 3–12.
- Bendel, Oliver 2019a: Einleitung. In: Bendel, Oliver (Hg.): Handbuch Maschinen Ethik. Wiesbaden, Springer VS: 1–10.
- Bendel, Oliver 2019b: Wozu brauchen wir die Maschinenethik? In: Bendel, Oliver (Hg.): Handbuch Maschinen Ethik. Wiesbaden, Springer VS: 13–32.
- Brand, Lukas 2018: Künstliche Tugend. Roboter als moralische Akteurin. Regensburg, Pustet.
- Evangelischen Kirche in Deutschland (Hg.) 2021: Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland. Leipzig, EVA. <https://www.ekd.de/freiheit-digital-63984.htm> (aufgerufen am 30.06.2021).
- Grimm, Petra/Keber Tobias O./Zöllner, Oliver 2019: Digitale Ethik. Leben in vernetzten Welten. Stuttgart, Reclam.
- Hubig, Christoph 2011: Technikethik. In: Stoecker, Ralf/Neuhäuser, Christian/Raters, Marie-Luise (Hg.) 2011: Handbuch Angewandte Ethik. Stuttgart, Metzler: 170–175.
- Lenzen, Manuela 2018: Künstliche Intelligenz. Was sie kann & Was uns erwartet. München, Beck.
- Lenzen, Manuela 2020: Künstliche Intelligenz. Fakten, Chancen, Risiken. München, Beck.
- Loh, Janina 2019a: Maschinenethik und Roboterethik. In: Bendel, Oliver (Hg.): Handbuch Maschinen Ethik. Wiesbaden, Springer VS: 1–10. 75–93.
- Loh, Janina 2019b: Roboterethik. Eine Einführung. Berlin, Suhrkamp.
- Mainzer, Klaus 2019: Künstliche Intelligenz – Wann übernehmen die Maschinen? 2. Aufl. Wiesbaden, Springer VS.
- McEwan, Ian 2019: Maschinen wie ich. Zürich, Diogenes.
- Misselhorn, Katrin 2018: Grundfragen der Maschinenethik. 2. Aufl. Stuttgart, Reclam.
- Misselhorn, Katrin 2021: Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern und Co. Stuttgart, Reclam.
- Nida-Rümelin, Julian/Weidenfeld Nathalie 2020: Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz. München, Piper.
- Precht, Richard David 2020: Künstliche Intelligenz und der Sinn des Lebens. München, Goldmann.
- Ramge, Thomas 2018: Mensch und Maschine. Wie Künstliche Intelligenz und Roboter unser Leben verändern. Stuttgart, Reclam.

- Rössler, Beate 2019: *Autonomie. Ein Versuch über das gelungene Leben*. Berlin, Suhrkamp.
- Simanowski, Roberto 2020: *Todesalgorithmus. Das Dilemma der künstlichen Intelligenz*. Wien, Passagen.
- Spiekermann, Sarah 2019: *Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert*. München, Droemer.
- Stoecker, Ralf/Neuhäuser, Christian/Raters, Marie-Luise (Hg.) 2011: *Handbuch Angewandte Ethik*. Stuttgart, Metzler.
- Wallach, Wendell/Allen, Colin 2009: *Moral machines. Teaching robots right from wrong*. New York/Oxford, University Press.
- Werkner, Ines-Jacqueline/Hofheinz, Marco (Hg.) 2019: *Unbemannte Waffen und ihre ethische Legitimierung*. Wiesbaden, Springer VS.
- Woopan, Christiane/Jannes, Mark (Hg.) 2019: *Roboter in der Gesellschaft. Technische Möglichkeiten und menschliche Verantwortung*. Wiesbaden, Springer VS.