ÜBERLEGUNGEN ZUR DISZIPLIN DER MASCHINENETHIK

Oliver Bendel

Mit dem rechtlichen und moralischen Status von Maschinen mit Chips beschäftigt man sich in der Wissenschaft schon seit den 1950er Jahren. Die Science-Fiction-Literatur war noch früher dran. Lange Zeit ging es vor allem darum, ob Roboter Objekte der Moral sind, sogenannte *moral patients*, ob man ihnen etwa Rechte zugestehen kann. Ich befasse mich seit den 1990er Jahren mit diesem Thema. Ich habe damals keinen Grund gesehen, Robotern Rechte zu geben, und bis heute nicht meinen Standpunkt geändert. Wenn sie eines Tages etwas empfinden oder wenn sie leiden können, oder wenn sie eine Art von Lebenswillen haben, lasse ich mich bestimmt überzeugen. Aber im Moment bemerke ich keine entsprechenden Tendenzen.

SUBJEKTE DER MORAL

Die Diskussion, ob Roboter Subjekte der Moral sind, das heißt, ob von ihnen moralische Handlungen ausgehen können oder sie bestimmte Pflichten haben, folgte in den 2000er Jahren. Ich tue mir in diesem Zusammenhang schwer mit dem letztgenannten Begriff. Ich habe von „Verpflichtungen" gesprochen, bin mir jedoch nicht sicher, ob das Problem dadurch wirklich gelöst wird. In der Monografie „Moral Machines" (2009) von Wendell Wallach und Colin Allen[01] wurden Roboter in systematischer Weise als *moral agents* besprochen und Einteilungen und Unterscheidungen der moralischen Subjekte angeboten. In dem zwei Jahre später folgenden Sammelband „Machine Ethics" von Michael Anderson und Susan Leigh Anderson wurde der Diskurs fortgeführt.[02]

Seit den 1980er Jahren spielt für mich die Tierethik (mithin Tierschutz) eine große Rolle.[03] Tiere sind für mich Objekte der Moral, keine Subjekte. Sicherlich haben Schimpansen oder Elefanten vormoralische Qualitäten, und weil wir von Tieren abstammen, eventuell solche sind, dürfte es fließende Übergänge geben. Jedenfalls können Tiere nicht gut oder böse sein, und sie können ebenso wenig unterschiedliche Handlungsmöglichkeiten aus moralischer Perspektive beurteilen und sich dann beispielsweise für das geringste Übel entscheiden. Wir können „Du böser Hund!" sagen, wenn er uns gebissen hat, aber wir meinen das nicht moralisch, höchstens pädagogisch.

GUTE UND BÖSE MASCHINEN

Ebenso wie Tiere halte ich auch Maschinen nicht für gut oder böse, zumindest nicht im Sinne der Philosophin Annemarie Pieper, die einen bösen oder guten Willen voraussetzt (ob Handlungen an sich gut oder böse sind, kann man dennoch diskutieren).[04] Aber offenbar können sie blitzschnell Optionen erfassen und dann nach einer vorgegebenen moralischen Regel entscheiden oder die moralischen Folgen abschätzen und dann auswählen. In diesem Sinne steht ein hochentwickelter Roboter (oder ein hochentwickeltes System der Künstlichen Intelligenz, kurz KI) zwischen Mensch und Tier. Das mag irritieren, und ich weise darauf hin, dass er in einem anderen Sinne ganz woanders steht. Als Objekt der Moral taucht er im Moment gar nicht auf. Man darf ihn behandeln, wie man will, ihn schlagen und zerstören, und wenn das jemand bedauert, ist eben der, der es bedauert, das Objekt der Moral, nicht der Roboter. Ein Spiel über Bande sozusagen, wie bei Teddybären und heiligen Steinen oder Bergen. Diese kommen nur in die Sphäre der Moral, weil sie jemandem gehören oder jemanden interessieren.

Nach dem, was bisher gesagt wurde, sind manche Maschinen neue, fremde, merkwürdige Subjekte der Moral. So habe ich es in den vergangenen Jahren vertreten. Sie sind nicht gut oder böse, zumindest nicht in Bezug auf einen Willen

(den sie nicht haben), und sie haben keine Pflichten im engeren Sinne (höchstens „Verpflichtungen", im Sinne von Aufgaben, die man an sie delegiert, und vielleicht nicht einmal die). Wenn sie eine Verantwortung tragen, dann allenfalls, weil ihnen etwas anvertraut oder übereignet wurde. Eine maschinelle Primärverantwortung könnte das sein, die noch genauer zu untersuchen wäre. Bei der Sekundärverantwortung wird es schon schwieriger, denn wie sollten wir die Maschine zur Rechenschaft ziehen? An den Ohren können wir sie nicht ziehen, und selbst wenn sie welche hat, wird es sie nicht stören. Die Tertiärverantwortung wird eventuell ein Thema, wenn sie ins Rechtliche übergeht. Ein Jurist kann alles konstruieren, sogar eine elektronische Person, die man verklagen und belangen kann. Als Ethiker bin ich hier vorsichtig.

MASCHINELLE MORAL

Den Begriff der maschinellen Moral verwende ich gerne ähnlich wie den der künstlichen Intelligenz. Beide Male meine ich den Gegenstand der Disziplinen. Die Künstliche Intelligenz hat die maschinelle oder künstliche Intelligenz zum Gegenstand. Sie simuliert menschliche Intelligenz oder strebt danach, diese eines Tages in wesentlichen Funktionen abzubilden. Die Maschinenethik hat die maschinelle Moral zum Gegenstand. Sie simuliert derzeit die menschliche Moral. Allerdings konzentriert sie sich in ihren aktuellen Ausprägungen auf gewisse Grundzüge. Die meisten moralischen Maschinen sind wie menschliche Fundamentalisten. Sie halten sich stur an Regeln, die man ihnen eingetrichtert hat. Einige moralische Maschinen vermögen immerhin die Folgen abzuschätzen, die ihre Handlungen nach sich ziehen würden, und unterschiedliche Entscheidungen treffen, die ihnen wiederum beigebracht wurden. Selbstlernende Maschinen könnten indes mehr.

Bisher scheint also festzustehen, dass Maschinen neuartige Subjekte der Moral sein können. Sie können unterschiedliche Optionen beurteilen und dann Entscheidungen treffen, die moralisch adäquat zu sein scheinen. Die moralischen Fähigkeiten werden ihnen von Menschen beigebracht. Dieser Transferprozess bleibt freilich nicht ohne Folgen. Wir haben es in der Regel mit teilautonomen und autonomen Maschinen zu tun, die alleingelassen sind, die nicht von uns beaufsichtigt werden, die in Situationen geraten, die wir vielleicht vorausgesehen haben, aber doch ein wenig anders sind. Es ergeben sich erste Unschärfen: Moral und Anwendungsfall der Moral passen nicht immer zueinander.

DIE DISZIPLIN DER MASCHINENETHIK

Die Disziplin, die sich mit Maschinen als Subjekten der Moral beschäftigt, die die maschinelle Moral untersucht und hervorbringt, ist die Maschinenethik.[05] Sie ist ein Pendant zur Menschenethik, die sich mit Menschen als Subjekten der Moral beschäftigt. Oder auch nur ein Ausnahmefall der angewandten Ethik. Sie scheint aber nicht ganz zu den klassischen Bereichsethiken zu passen, zu Informations-, Technik- oder Wirtschaftsethik.[06] Etwas ist anders, eben die Antwort auf die Frage nach dem Subjekt der Moral. Die Ethik ist für mich, wie gerade deutlich wurde, die Disziplin, die Moral der Gegenstand. Ein Student von mir hat es so ausgedrückt: Ethik treibt man, Moral hat man. Anders als die klassischen Bereichsethiken denkt die Maschinenethik nicht nur über Subjekte der Moral nach, sondern schafft sie im besten Falle auch. Sie bringt, zusammen mit Künstlicher Intelligenz und Robotik, moralische und manchmal unmoralische Maschinen hervor.

Selbstverständlich darf man ganz anders sprechen, darf man Ethik und Moral in eins setzen, wie es im Englischen oft gemacht wird, darf man die Ethik oder die Moral weiter fassen. Wichtig ist mir, dass deutlich wird, was ich meine. Der Rest ist Übersetzungsarbeit. Die Ethik ist also die Disziplin (eine Disziplin der Philosophie), gerne auch eine Theorie oder eine Lehre. Als Moralphi-

01 Vgl. Wendell Wallach/Colin Allen, Moral Machines: Teaching Robots Right from Wrong, New York 2009.
02 Vgl. Michael Anderson/Susan Leigh Anderson (Hrsg.), Machine Ethics, Cambridge 2011.
03 Vgl. Ursula Wolf, Ethik der Mensch-Tier-Beziehung, Frankfurt/M. 2012.
04 Vgl. Annemarie Pieper, Einführung in die Ethik, Tübingen–Basel 2007.
05 Vgl. Oliver Bendel, Wirtschaftliche und technische Implikationen der Maschinenethik, in: Die Betriebswirtschaft 4/2014, S. 237–248.
06 Vgl. Oliver Bendel, Maschinenethik, in: Gabler Wirtschaftslexikon, 2012, http://wirtschaftslexikon.gabler.de/Definition/maschinenethik.html.

losophie westlicher Prägung ist sie für mich, Annemarie Pieper folgend, Wissenschaft. Was wir in der Ethik, genauer: in der normativen Ethik, auch finden, sind Modelle wie die Pflicht- oder Pflichtenethik und die Folgenethik. Was wir der Maschine einpflanzen, ist eine maschinelle Moral, eine Moral, die in ihr, der Maschine, funktioniert. Diese Moral können wir einbetten in ein Modell der normativen Ethik.

ARTEFAKTE DER MASCHINENETHIK

Wer sich heute als Moralphilosoph der Maschinenethik verschließt, verkennt deren historische Bedeutung. Zum ersten Mal bauen wir in der Ethik etwas, bringen Artefakte hervor, treiben nicht nur Gedankenexperimente, sondern „Tatsachenexperimente". Natürlich ist die Frage, ob man jede moralische oder unmoralische Maschine umzusetzen hat. In einem Labor sollte man vielleicht die Atombombe der Maschinenethik bauen, dann aber dort lassen. Vielleicht wäre es zu gewagt, sie zu bauen, weil man sie kaum unter Verschluss halten kann. Aber man sollte der Maschinenethik, wie der Physik, möglichst wenig verbieten. Auch solche Erkenntnisse, die problematisch zu sein scheinen, können sich als nützlich erweisen.

Ich entwickle unter anderem spezielle Chatbots, also Vertreter der Softwareroboter. Zuletzt haben wir einen Hardwareroboter auf die Welt gebracht. Er zeigt den Fokus meiner Forschung: Ich will im Kontext von Maschinenethik und Tierethik (und Tierschutz) tierfreundliche Maschinen erfinden. Saugroboter, Mähmaschinen, Windkraftanlagen, selbstständig fahrende Autos – es existieren zahlreiche Maschinen, die man „moralisieren" kann. Nicht mit allen sollte man das tun, und bei selbstständig fahrenden Autos bin ich sehr skeptisch, was ihre Entscheidungen und Handlungen uns gegenüber anbelangt, vor allem dann, wenn menschliche Unfallopfer auszuwählen sind.

2018 wollen wir unser viertes Artefakt der Maschinenethik bauen. Nach dem GOODBOT und dem LIEBOT aka LÜGENBOT, zwei Chatbots, nach dem LADYBIRD, einem tierfreundlichen Staubsaugerroboter (der seine Arbeit einstellt, sobald er Marienkäfer erkennt), wird hoffentlich der BESTBOT geboren. Unsere Maschinen halten sich entweder an vorgegebene Regeln, die sie unmittelbar umsetzen, oder versuchen die Folgen abzuschätzen (um dann wieder vorgegebene Regeln anzuwenden). Hier sind Pflicht- und Folgenethik gefragt. Danach würde ich gerne selbstlernende moralische Maschinen bauen.[07] Diese wären in der Lage, eine eigene, genuin maschinelle Moral zu entwickeln. Das könnte ausgesprochen gefährlich sein, für Körper und Geist, und es ist nicht mein Bestreben, alle unsere Prototypen zum Vorbild von Produkten zu erklären. Es interessiert mich, was möglich ist, die neue Unschärfe, die entsteht, die maschinelle Moral, die nicht auf die menschliche passt. Unser LIEBOT war zwar nicht lernfähig, aber stark vernetzt, und er hatte genuin maschinelle Strategien des Lügens, wenn er Suchmaschinen und Klassifikationen benutzte, um die Wahrheit herauszufinden, und diese dann nach allen Regeln der Kunst verdrehte. Mich interessiert also auch die Maschine, die sich anders als der Mensch verhält. Der GOODBOT war lernfähig, dies am Rande, ohne seine Moral weiterzuentwickeln. Vielmehr hat er den Benutzer bewertet und eingeordnet und sein Verhalten angepasst, innerhalb eines vorgegebenen Rahmens.

Damit wären wir bei der Frage, was die Maschine kann und soll. Die immer wieder auftauchende Frage, ob man Maschinenethik treiben soll, halte ich für müßig. Soll man Teilchenphysik treiben oder Mikrobiologie? Wenn man damit eine Disziplin so ergänzt oder unterteilt, dass man gezielter forschen kann, und wenn man neue wissenschaftliche Erkenntnisse hervorzubringen vermag, durchaus.

AUTOMATISIERTES FAHREN

Es ist schade, dass man Maschinenethik vor allem mit theoretischen Dilemmata in Verbindung bringt, mit philosophischen Gedankenexperimenten, die ungemein wichtig und interessant sind, aber nicht immer zum entscheidenden Schritt führen. Auch praktische Dilemmata sind einzubeziehen, die tatsächlichen Situationen, in die eine Maschine kommt. Wenn man sagt, dass theoretischen Dilemmata der Realitätsbezug fehlt, hat man nicht verstanden, was Gedankenexperimente sind. Das Trolley-Problem ist nicht

07 Vgl. Luís Moniz Pereira/Ari Saptawijaya, Programming Machine Ethics, Cham 2016.

dafür da, dass es in der Wirklichkeit auftritt. In der Wirklichkeit sind selten fünf Personen auf einem Gleis, die man durch eine Weichenstellung retten kann, wobei man aber einen Unbeteiligten auf einem anderen Gleis opfern muss. Das Gedankenexperiment ist dafür da, unsere moralische Haltung offenzulegen und Modelle der normativen Ethik zu veranschaulichen. Es kann unterschiedliche Ansätze aufzeigen, an eine Sache heranzugehen.

Da wäre zum Beispiel das Qualifizieren und Quantifizieren. Ich habe mir 2012 auf der Grundlage des Trolley-Problems das etwas komplexere Roboterauto-Problem ausgedacht (auf dem Weg und der Straße befinden sich zwei voneinander unabhängige Erwachsene und drei Kinder) und dieses Anfang 2013 auf einer Konferenz zur Technologiefolgenabschätzung in Prag vorgestellt.[08] Einer meiner Studenten hatte für das autonome Auto des Gedankenexperiments (NAC, New Autonomous Car) eine Formel entwickelt, die qualifizieren und quantifizieren konnte. Es konnte also Menschen anhand ihres Geschlechts, Alters, Aussehens und so weiter beurteilen, sie in diesem Sinne klassifizieren, und sie durchzählen, also beispielsweise potenzielle Unfallopfer berechnen und gegeneinander aufrechnen. Wir haben damals theoretisch festgestellt, dass beide Verfahren problematisch sind. Wenn man qualifiziert, diskriminiert man meist, und wenn man quantifiziert, muss man die Frage beantworten, warum drei Menschen unbedingt mehr wert sein sollen als zwei.

Nun kann man dies praktisch anwenden, man kann im automatisierten Fahren praktische Dilemmata und Gefahrensituationen aller Art voraussehen und die beiden Ansätze implementieren. Damit gelangt man von der Theorie in die Praxis, und da faktisch alle möglichen Gefahrensituationen auftreten, in denen man abwägen und urteilen muss, kann man hier nicht mehr sagen, dass nichts geschehen wird. Natürlich wird etwas geschehen und eine bestimmte Entscheidung getroffen werden müssen. Es ist sogar so: Ein autonomes Auto kann sich, frei nach Paul Watzlawick, nicht nicht entscheiden (der berühmte Kommunikationswissenschaftler hat vom Kommunizieren gesprochen). Selbst wenn es in einer Unfallsituation unbeirrt geradeaus fährt, ist das eben eine Entscheidung, die der Entwickler oder Programmierer der Maschine mitgegeben hat. Einige lehnen es ab, bei Maschinen von Entscheidungen zu sprechen. Es wird freilich schwierig bei einer solch extremen Position, überhaupt über sie zu sprechen. Gibt es Roboter, die Fußball spielen? In einem gewissen Sinne nicht.

Die Lösung des praktischen Problems lautet für mich: Man sollte beim hoch- und vollautomatisierten beziehungsweise beim autonomen Fahren vorsichtig sein mit moralischen Regeln, die man dem Auto beibringt. Man sollte zudem vorsichtig sein mit moralischen Fähigkeiten, zu denen das Auto selbst gelangt. Es existieren viele interessante Ansätze, die Maschine dazulernen zu lassen. Man kann sie aufziehen wie ein Kind, ihr in den ersten Monaten einen Fahrlehrer aufzwingen, sie nur in solche Gegenden und Situationen schicken, wo sie sich vorbildliches Verhalten abschaut. Ob das alles zum gewünschten Ziel führt, zu einem tier- und menschenfreundlichen Verhalten, ist die Frage. Ich denke, in den nächsten 20 Jahren ist man gut beraten, wenn man das selbstständig fahrende Auto auf die Autobahnen schickt und ansonsten uns steuern lässt. Sowohl die Stadt als auch die Landstraßen sind hochkomplexe Umgebungen. Ich finde autonome Autos als Maschinenethiker, wie vermutlich klar wurde, vor allem mit Blick auf Tiere interessant.

PFLEGEROBOTER, ZIVILE UND MILITÄRISCHE ROBOTER

Die Maschinenethik beschäftigt sich also vor allem mit (teil-)autonomen Systemen. Dazu gehören auch Pflege- und Therapieroboter oder zivile und militärische Drohnen. Aber selbst 3D-Drucker könnte man moralisieren. Man könnte ihnen beibringen, keine Waffen auszudrucken. Sie müssten wissen, was eine Waffe ist, die Objekte, Vorlagen und Dateien beurteilen und sich dann entsprechend entscheiden können. Bei Pflegerobotern ist zum Beispiel von Bedeutung, ob sie den Patienten töten können sollen, also Sterbehilfe leisten können. Die meisten Krankenhäuser und Pflegeheime dürften dies ablehnen. Der Maschinenethiker interessiert sich weniger für die Frage an sich, denn er ist kein Medizin- und kein Sterbeethiker. Er interessiert sich vielmehr dafür, wie er die Fähigkeit, einen Befehl aus mo-

08 Vgl. Oliver Bendel, Towards Machine Ethics, in: Tomáš Michalek/Lenka Hebáková/Leonhard Hennen et al. (Hrsg.), Technology Assessment and Policy Areas of Great Transitions, 1st PACITA Project Conference, March 13–15, 2013, Prague 2014, S. 321–326.

ralischen Gründen zu verweigern, konzipieren und implementieren kann. Denn darum könnte es gehen: Der Pflegeroboter weiß, wie man jemanden erwürgt, und er entscheidet sich, das nie zu tun, selbst wenn man ihn inständig darum bittet. Wer das für Science-Fiction hält, sollte sich unterschiedliche Typen von Pflegerobotern anschauen. Solche, die etwas transportieren, solche, die uns informieren, und solche, die Hand an uns legen. Manche sehen aus wie ein Mensch oder ein Tier, andere wie Kooperations- und Kollaborationsroboter aus der Industrie. Diese haben meist einen Arm, mehrere Achsen und zwei bis drei Finger. Man kann sie trainieren, indem man ihren Arm und ihre Finger bewegt oder ihnen einfach etwas vormacht, während sie mit ihren Sensoren und Systemen alles verfolgen und verarbeiten. Im Prinzip kann man ihnen beibringen, uns zu erwürgen, wobei es nicht ohne Grund bestimmte technische Normen gibt, um dies zu verhindern. Eine Aufgabe der Maschinenethik wäre es eben, den Roboter mit einer Form der Befehlsverweigerung vertraut zu machen, die moralisch begründet wäre.

Der militärische Einsatz wurde erwähnt. Wissenschaftler verschiedener Disziplinen erhalten Geld von Verteidigungsministerien. Ein großer Teil fließt in die Entwicklung autonomer Kampfsysteme, in Robotik und Informatik, ein kleiner – wie im Falle des Pentagon – in das künstliche Gewissen, das diese haben sollen, oder in ihre Fähigkeit, den Gegner zu täuschen, zu betrügen und zu verwirren, also in die Maschinenethik. Man kann den Kampfroboter grundsätzlich für unmoralisch halten. Dennoch kann er etwas tun, was man moralisch nennen könnte, etwa mithilfe seines künstlichen Gewissens einen Kollateralschaden vermeiden. Hier zeigt sich, dass der Unterschied zwischen moralischen und unmoralischen Maschinen nicht einfach zu bestimmen ist. Diese Frage muss im Rahmen der Maschinenethik noch intensiver diskutiert werden.

ZUSAMMENFASSUNG UND AUSBLICK

Maschinenethik ist eine junge Disziplin, die man in der Philosophie ansiedeln kann, die aber Partnerinnen wie die Künstliche Intelligenz und Robotik braucht, vor allem dann, wenn sie erfolgreich Artefakte herstellen und erforschen will, Simulationen und Prototypen, die die Möglichkeit moralischer und unmoralischer Maschinen zeigen. Von solchen darf die Maschinenethik sprechen, so wie die Disziplin der Künstlichen Intelligenz davon sprechen darf, dass sie künstliche Intelligenz hervorbringt. Welche Maschinen man genau moralisieren soll, muss diskutiert und eruiert werden.

Es ist unproblematisch, ja hilfreich, tierfreundliche Staubsaugerroboter zu bauen, die die Moral ihrer Besitzer auch in deren Abwesenheit umsetzen. Sie bewegen sich in geschlossenen oder halboffenen Welten und treffen einfache Entscheidungen. Beim automatisierten Fahren sieht es schon anders aus, und ich bin dagegen, dass Autos potenzielle menschliche Unfallopfer durchzählen oder sie bewerten und dann ihre Urteile fällen. Tierische Verkehrsteilnehmer darf man auf diese Weise behandeln, und es wäre wünschenswert, die Zahl der Getöteten dadurch zu reduzieren. Man könnte sagen, dass auch in offenen Welten überschaubare Situationen entstehen können (hier solche, die sich auf Tiere beziehen), in denen einfache Entscheidungen möglich sind.

Vor der Maschinenethik liegt also ein weites Feld, und sie kann unterschiedliche Richtungen einschlagen. Es ist weniger wichtig, dass sie moralische Regeln begründet. Viel wichtiger ist, dass sie moralisch begründete Regeln in einer befriedigenden Weise implementiert. Dabei kann sie sowohl moralische als auch unmoralische Maschinen erschaffen. Die Ethik erforscht, was gut und böse ist, sie ist nicht gut oder böse. Natürlich darf jeder dazu beitragen, dass die Welt ein bisschen besser wird. Und genau deshalb bin ich persönlich an bestimmten moralischen Maschinen besonders interessiert.

OLIVER BENDEL
ist Professor für Wirtschaftsinformatik und Ethik mit den Schwerpunkten Wissensmanagement, Informationsethik und Maschinenethik an der Hochschule für Wirtschaft FHNW in Basel. Aktuell forscht er über künstliche Stimmen.
oliver.bendel@fhnw.ch