Roboter können Wohnzimmer oder Schwimmbäder vermessen und diese eigenständig säubern. Sie bringen pflegebedürftigen Menschen das Essen, muntern sie auf und können ferngesteuert Operationen durchführen. Computergesteuerte Autos sollen zukünftig Fahrerin und Fahrer ersetzen, einige behaupten, durch vollständig selbstfahrende Autos würde sich die Zahl der jährlichen Verkehrstoten um 90 % reduzieren. Intelligente Maschinen halten Einzug in fast alle Bereiche des Lebens. Aber Maschinen sind keine Menschen, sie haben keine Emotionen und können selbst dann, wenn man sie per Algorithmus Entscheidungen abwägen lässt, nur das ausführen, wozu sie programmiert sind. Wie können wir künstliche Intelligenz (KI) ethisch ausrichten und damit vertretbar für den Menschen machen? Dr. Catrin Misselhorn, Professorin für Theoretische Philosophie an der Georg-August-Universität Göttingen, forscht dazu.

Programmierte Ethik

Künstliche Intelligenz ohne Entscheidungsspielräume

Joachim von Gottberg im Gespräch mit Catrin Misselhorn

© Stephanie Trenz

Wo liegen die Unterschiede zwischen dem schon älteren Begriff „Technikphilosophie" und dem mittlerweile modernen Begriff „Maschinenethik"?

Die Maschinenethik ist ein neues Forschungsfeld an der Schnittstelle von Informatik, Robotik und Philosophie, dem es um die Entwicklung einer Ethik für Maschinen im Gegensatz zur Entwicklung einer Ethik für Menschen im Umgang mit Maschinen geht. Man spricht in Analogie zu Artificial Intelligence auch von Artificial Morality. Während Artificial Intelligence zum Ziel hat, die kognitiven Fähigkeiten von Menschen zu modellieren oder zu simulieren, geht es bei der Artificial Morality darum, künstliche Systeme mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten. Die klassische Technikethik beschäftigt sich hingegen mit der Frage, wie man Technologien bewerten kann, z. B., ob es verantwortbar ist, Atomkraft zu nutzen oder nicht. Wissenschaftlich ist die Maschinenethik besonders spannend, weil es sich um eine gänzlich neue ethische Disziplin handelt, die bedingt durch die Fortschritte in der KI-Forschung entstanden ist. Deshalb hat es mich besonders gereizt, darüber ein Buch[1] zu schreiben.

Mit welchen unterschiedlichen ethischen Fragen beschäftigen Sie sich?

Ich beschäftige mich mit ethischen Fragen der künstlichen Intelligenz, Roboter- und Maschinenethik. Es geht hierbei um die moralischen Probleme bei der Entwicklung, Herstellung und Verwendung von künstlicher Intelligenz und Robotern. Dazu gehören auch die gesellschaftlichen Konsequenzen des zunehmenden Einsatzes dieser Technologien. Die spezifisch ethischen Fragen, die sie aufwerfen, haben mit zwei Aspekten zu tun: Zum einen nehmen die Intelligenz und Autonomie dieser Technologien ständig zu. Dieser Aspekt führte mich zur Maschinenethik, also zu der Frage, ob Maschinen moralisch entscheiden und handeln können und ob sie dies tun sollen bzw. in welchen Bereichen dies wünschenswert ist. Wichtige Anwendungsbereiche der Maschinenethik sind etwa Pflegesysteme, Kriegsroboter und das autonome Fahren. Andere ethische Problemfelder sind beispielsweise die Transparenz von Algorithmen oder der Schutz von Daten und Privatsphäre. Der zweite Aspekt ergibt sich eher aus der äußeren Gestalt von Robotern und ihrer Interaktion mit Menschen. Man kann auch von einer Ethik der Mensch-Maschine-Interaktion sprechen. Hier geht es etwa darum, wie menschenähnlich Maschinen wirken sollten.

Wo sehen Sie die größten möglichen Probleme?

Die künstliche Intelligenz kann im schlimmsten Fall zur Beeinträchtigung der menschlichen Selbstbestimmung und Würde sowie zur Aushöhlung der Demokratie führen. Eine künstliche Intelligenz, die Entscheidungen über Leben und Tod von Menschen fällt, droht die Menschenwürde zu verletzen. Der ethisch nicht reglementierte Einsatz der künstlichen Intelligenz ist durchaus auch dazu in der Lage, die Grundfesten unserer Demokratie zu erschüttern. Das gilt insbesondere für die Möglichkeit der Massenüberwachung, gar noch verbunden mit der Belohnung von staatlich erwünschtem Verhalten und der Sanktionierung von staatlich unerwünschtem Verhalten nach dem Vorbild des chinesischen Social-Scoring-Systems. Substanzielle politische Partizipation und Kritik an der staatlichen Obrigkeit, die eine Demokratie ausmachen, sind unter diesen Bedingungen nicht möglich.

Haben Sie ein spezielles Ziel, z. B. eine Warnung vor einer Technik, die wir ethisch nicht mehr beherrschen können? Oder geht es um eine Optimierung des Umgangs damit?

Beides. Ich habe drei Leitlinien für die Maschinenethik aufgestellt, die auch auf die Ethik der künstlichen Intelligenz im Allgemeinen übertragbar sind: Erstens sollten künstliche Systeme die Selbstbestimmung von Menschen fördern und sie nicht beeinträchtigen. Zweitens sollten sie nicht über Leben und Tod von Menschen entscheiden. Drittens muss sichergestellt werden, dass Menschen stets in einem substanziellen Sinn die Verantwortung übernehmen. An diesen drei Leitlinien können Sie ablesen, dass es um beides geht: um eine Warnung vor den Gefahren, die der ethisch nicht reglementierte Umgang mit bestimmten Technologien beinhalten kann, aber auch darum, wie man die Technologien so gestalten kann, dass sie den menschlichen Bedürfnissen, insbesondere der Selbstbestimmung förderlich sind. Mit diesem letzten Aspekt beschäftigte ich mich etwa in meinen Drittmittelprojekten zur ethischen Bewertung von Assistenzsystemen in der Pflege, der Arbeitswelt und in der Bildung. Hier sehe ich meine Rolle als Beraterin zur ethischen Optimierung der Technologien.

Man könnte Staubsaugerroboter so programmieren, dass sie um bestimmte Tiere einen Bogen machen, während andere Tiere bevorzugt aufgesaugt werden. Hat das etwas mit Ethik zu tun?

Ihre Frage beinhaltet eigentlich zwei Aspekte: zum einen, ob beim Staubsaugerroboter überhaupt schon Ethik ins Spiel kommt, und zum anderen, ob die ethischen Intuitionen, die viele von uns hier haben, konsistent sind. Die erste Teilfrage würde ich mit Ja beantworten, denn diese Entscheidungen beinhalten die Abwägung, ob es richtig ist, Insekten um der Sauberkeit willen zu töten oder nicht. Das

ist auf jeden Fall eine ethische Frage. Bei der zweiten Frage bin ich skeptisch und würde sagen, wenn man die Marienkäfer verschont, dann auch die Spinnen. Mein Staubsaugerroboter hätte keinen Kill-Button für Spinnen. Bei Wespen mag es sich wieder anders verhalten, etwa wenn kleine Kinder oder Allergiker im Haushalt leben. Aber dieses Beispiel zeigt ganz gut eine bestimmte Dynamik der Maschinenethik: Unsere Alltagsethik scheint, was den Umgang mit Tieren angeht, nicht ganz konsistent zu sein. Die Maschinenethik zwingt uns dazu, sie konsistent zu machen.

Bei welchen Geräten beginnen ethische Überlegungen auch?

Ethische Überlegungen können im Prinzip alle Technologien betreffen. So kann man sich auch beim Thermomix fragen, ob hierbei nicht etwas an Achtsamkeit im Umgang mit Lebensmitteln und dem sinnlichen Prozess des Kochens verloren geht. Eine solche Kritik wäre analog zu derjenigen an der Ersetzung handwerklicher Tätigkeiten durch industrielle Fertigung, die ja ein klassisches Thema der Technikethik ist. Auf den Staubsaugerroboter bin ich ja eben schon eingegangen. Interessanter ist die Frage, ab wann die Maschinenethik beginnt. Dies erfordert moralische Informationsverarbeitungsprozesse: Maschinen können moralisch relevante Merkmale einer Situation, also den Istzustand erkennen, auf dieser Grundlage eine moralische Entscheidung treffen, den Istzustand also mit einem moralischen Sollzustand abgleichen und auf die Situation einwirken, wenn dies nicht der Fall ist. Vollumfängliche Moral, wie sie Menschen zukommt, geht jedoch weit darüber hinaus. Sie umfasst auch Bewusstsein, was etwa bei der Empfindung moralischer Emotionen wie Schuldgefühlen oder Empathie eine Rolle spielt. Menschen verfügen auch über Selbstbewusstsein: Ihr moralisches Handeln hängt eng mit dem Bild zusammen, das sie von sich selbst entwerfen. Sie können außerdem über ihre moralischen Gründe reflektieren, was die Grundlage für Willensfreiheit darstellt.

Google und Amazon verfügen über wahnsinnige Datenmengen, wodurch sich enorme analytische Möglichkeiten ergeben. Amazon weiß, was man als Nächstes bestellen will und schickt es einem unaufgefordert zu. Wie ordnen Sie das ethisch ein?

Das ist ein absolutes No-Go. Stellen Sie sich vor: Jeden Tag würden Hunderte von Paketen von diversen Firmen bei Ihnen abgeworfen, die Sie dann alle wieder loswerden müssten. Das ist weitaus schlimmer als Haustürgeschäfte oder unerwünschte Telefonwerbung, und die haben wir ja auch reglementiert. Philosophisch gesehen muss man zwischen Wunsch und Willen unterscheiden: Wünsche sind unverbindlich, man hat jede Menge davon, sie sind nicht einmal miteinander konsistent. So wünsche ich mir einerseits viele Süßigkeiten, auf der anderen Seite möchte ich nicht dick werden. Entscheidend ist, dass man nicht all diese Wünsche auch realisieren möchte. Willensbildung besteht genau darin, sich für diejenigen Wünsche zu entscheiden, die man dann auch umsetzen will. Eine „Fee" namens Amazon, die Wünsche erfüllt, bevor man sie noch gewollt hat, ist für mich deshalb ein Albtraum.

Wir überlegen, selbstfahrende Autos so zu programmieren, dass sie bei einem Unfall Kinder eher verschonen als Rentner. Das erinnert an ein Gesetz, nach dem die Bundeswehr Flugzeuge abschießen sollte, die von Terroristen gezwungen werden, in ein voll besetztes Stadion zu fliegen. Das Bundesverfassungsgericht hat das Gesetz gekippt. Aber stellen wir uns vor, so etwas würde tatsächlich vorkommen: Würde man im Nachhinein genauso entscheiden, wenn 70.000 Menschen ums Leben gekommen wären, die man durch das Opfer von 200 Menschen hätte retten können?

Diese Situation ist einer der Gründe, warum ich es für moralisch fragwürdig halte, Maschinen die Entscheidung über Leben und Tod von Menschen zu übertragen. Das liegt daran, dass der für Menschen wesentliche Entscheidungsspielraum fehlt. Das Verfassungsgerichtsurteil verdeutlicht überdies, dass ein solcher Umgang mit Dilemma-Situationen mit dem deutschen Rechtssystem nicht vereinbar ist. Es beruht auf der kantischen Idee, dass die Würde des

Menschen nicht quantifizierbar ist: 70.000 Menschen verfügen nicht über mehr Würde als 200. Zu bedenken ist hierbei auch, dass es in diesen Fällen um staatlich institutionalisiertes Handeln geht. Der einzelne Autofahrer könnte grundsätzlich eine solche Entscheidung treffen, doch ist es eine ganz andere Sache, wenn dies auf eine institutionalisierte und vom Staat sanktionierte Art und Weise geschieht. Insgesamt ist mir die Debatte um das autonome Fahren zu schwarz-weiß. Die Alternativen sind immer: entweder voll autonom fahren oder gar nicht. Es gibt jedoch noch einen Mittelweg: das assistierte Fahren. Deshalb plädiere ich dafür, zunächst zu eruieren, wie viel Zugewinn an Sicherheit wir durch Fahrassistenzsysteme erreichen können.

Manche schätzen, 50 % der Arbeitsplätze würden durch KI verloren gehen. Brauchen wir hier eine neue soziale Ethik, beispielsweise eine Maschinensteuer, um Menschen zu unterstützen, die nur ihre Arbeitskraft besitzen, um ihren Lebensunterhalt zu verdienen? Müssen wir finanzielle Ressourcen möglicherweise nach anderen Kriterien als bisher verteilen?

Entgegen vieler Horrorszenarien versuche ich, ein differenziertes Bild zu vertreten. So habe ich selbst an einem Projekt mitgearbeitet, wo es darum ging, mithilfe von Assistenzsystemen Menschen mit Behinderung eine Teilnahme am Arbeitsleben zu ermöglichen. Ich sehe also ethische Chancen und Risiken für den Einsatz von Assistenzsystemen. Ein wichtiges Ergebnis des Projekts war, dass auch Menschen mit kognitiven Beeinträchtigungen – etwa im Rahmen von Trisomie 21 – sehr ähnliche Vorstellungen von der Rolle der Erwerbsarbeit für ein gutes Leben haben wie Menschen ohne Behinderung. Übrigens ist dieser Anspruch auch in der UN-Behindertenrechtskonvention rechtlich verankert. Diese Vorstellungen sind tief in die Struktur unserer Gesellschaft eingegraben. Wir müssten unser Gesellschaftssystem radikal verändern, wenn die Erwerbsarbeit diese Rolle nicht mehr spielen soll. Die Einführung eines Grundeinkommens genügt dafür nicht. Insgesamt halte ich die Frage, wie Menschen sich gegen die Maschinen behaupten können, nicht für zielführend. Es geht vielmehr um eine gute Zusammenarbeit zwischen Mensch und Maschine.

Heute werden wir über Technik – beispielsweise Uhren, die mit Apps kommunizieren – vermessen. Die einen sehen darin eine Methode der sicheren Überwachung und der Lebensverlängerung, andere, wie z. B. der Soziologe Stefan Selke, befürchten den Verlust von Eigenständigkeit und Selbstbestimmung.

Vielen von Selkes Einsichten kann ich nur beipflichten. Doch wie so oft gibt es auch hier zwei Seiten: Liberal-demokratische Gesellschaften sind einerseits von einem ethischen Pluralismus geprägt. Deshalb ist es bis zu einem gewissen Grad eine Frage der individuellen Moralvorstellungen, wie viele Daten man bereit ist, preiszugeben, um gesundheitliche Risiken zu minimieren. Andererseits ist es nicht nur eine Frage subjektiver Präferenzen, sondern hinter dem Vermessungstrend steht auch gesellschaftlicher und wirtschaftlicher Druck, der zu objektiven Einschränkungen der Selbstbestimmung führt. Diesen Druck gilt es offenzulegen und zu problematisieren. Schließlich ist vielen vielleicht gar nicht klar, welche Auswertungsmöglichkeiten und Einwirkungen auf die Privatsphäre solche Datentracker haben. Hier gilt es, ein kritisches Bewusstsein zu wecken.

Gegenwärtig wird auf einer Konferenz über die Ethik autonomer Waffensysteme diskutiert. Ist das nicht absurd, im Zusammenhang von Massenvernichtungsmaschinen von Ethik zu sprechen?

Wenn man einer pazifistischen Gesinnung anhängt, verhält es sich natürlich so, wie Sie sagen. Nicht ganz so klar ist die Sache, wenn man davon ausgeht, dass es einen gerechten Krieg geben kann. Anders als der Pazifismus, der die Anwendung von Gewalt und das Töten von Menschen grundsätzlich für moralisch falsch hält, bildet die Theorie des gerechten Krieges eine Ethik aus, die die Anwendung von Gewalt und das Töten von

Menschen unter bestimmten Umständen für moralisch zulässig erachtet. Autonome Waffensysteme sollen dann denjenigen ethischen Richtlinien folgen, die die Art und Weise bestimmen, wie Krieg geführt werden soll, wenn er einmal ausgebrochen ist. Ein zentraler Gesichtspunkt ist beispielsweise, dass nur Kombattanten legitime Ziele sind und angegriffen werden dürfen. Dabei sollte nach Möglichkeit vermieden werden, Zivilisten zu schädigen oder zu töten. Die Idee ist, dass man autonome Waffensysteme nach diesen Vorgaben programmieren kann. Das klingt erst einmal gut, dennoch sind nicht nur die technischen Hürden groß, etwa wenn es darum geht, legitime Ziele korrekt zu identifizieren. Kriegsroboter sind auch ein moralisch hochproblematischer Anwendungsbereich der Maschinenethik. Kriegsroboter verfügen nicht über einen moralischen Entscheidungsspielraum wie Menschen. Denn auch im Krieg besteht keine Pflicht zu töten, sondern bestenfalls eine Erlaubnis. Wie damit in der Einzelsituation umgegangen wird, sollte dem menschlichen Ermessen überlassen bleiben.

Ray Kurzweil geht davon aus, dass etwa 2050 das Zeitalter der Singularität beginnt: Maschinen lernen, entwickeln neue Ideen und Programme, mit denen sie weiterlernen und sich letztlich der Kontrolle des Menschen entziehen. Damit lassen sich alle Probleme schnell lösen und Raum und Zeit spielen keine Rolle mehr.

Ich bin aus philosophischen Gründen eher skeptisch, dass es zu dieser Entwicklung kommen wird. Aber ich bin nicht unfehlbar, insofern würde ich generell zur Vorsicht raten. Ich muss allerdings sagen, dass ich dringlichere Gefahren sehe wie den Klimawandel oder im Bereich der KI den Einsatz autonomer Waffensysteme, aber auch den Trend, das autonome Fahren moralisch unreflektiert zu forcieren, ganz zu schweigen von Fragen der informationellen Selbstbestimmung wie dem Schutz von Daten oder der Privatsphäre und der möglichen Diskriminierung oder politisch negativen Auswirkungen von künstlicher Intelligenz.

Können Sie sich vorstellen, dass die Intelligenz von Maschinen der des Menschen ebenbürtig wird oder sie eines Tages sogar überschreitet?

In gewissen Bereichen ist die künstliche Intelligenz Menschen ja schon ebenbürtig oder sogar überlegen, etwa im *Schach* oder *Go*. Allerdings hat die menschliche Intelligenz die Besonderheit, dass sie grundsätzlich auf alle möglichen Bereiche anwendbar ist. Eine solche allgemeine Intelligenz ist aus meiner Sicht mit den derzeitigen Methoden der künstlichen Intelligenz nicht zu reproduzieren.

Können Maschinen aus sich heraus eine Empathie, Motivation und ein Ziel entwickeln?

Zugespitzt formuliert ist emotionale KI etwa so empathisch wie ein Psychopath. Künstliche Systeme können bestenfalls wie ausgebuffte Psychopathen Emotionen erkennen und sozial angemessen darauf reagieren. Diese Fähigkeit dient – ähnlich wie bei den Psychopathen – in vielen Fällen manipulativen Zwecken. Diese Zwecke setzt sich das System aber natürlich nicht selbst, sondern sie sind ihm von seinen Entwicklern vorgegeben. Eigene Ziele können Maschinen nicht entwickeln, weil diese Fähigkeit eng mit dem biologischen Leben verbunden ist. Bewerten und abwägen können Maschinen in einem gewissen Umfang hingegen schon. So kann man ein künstliches System so programmieren, dass es verschiedene Optionen im Hinblick auf eine bestimmte moralische Pflicht bewertet, beispielsweise in der Medizinethik die Pflicht, einem Patienten keinen Schaden zuzufügen. Den unterschiedlichen Handlungsmöglichkeiten werden dann Zahlenwerte zugewiesen, je nachdem, wie gering oder groß der zu erwartende Schaden ist. Das kann dann auch gegen andere Pflichten abgewogen werden, etwa die Pflicht, zum Wohl des Patienten zu handeln oder seine Autonomie zu respektieren. Was Maschinen jedoch abgeht, ist die menschliche Fähigkeit zur Selbstreflexion und Moralbegründung, die es auf jeder Ebene immer wieder erlaubt, das eigene Denken und Handeln zu hinterfragen.

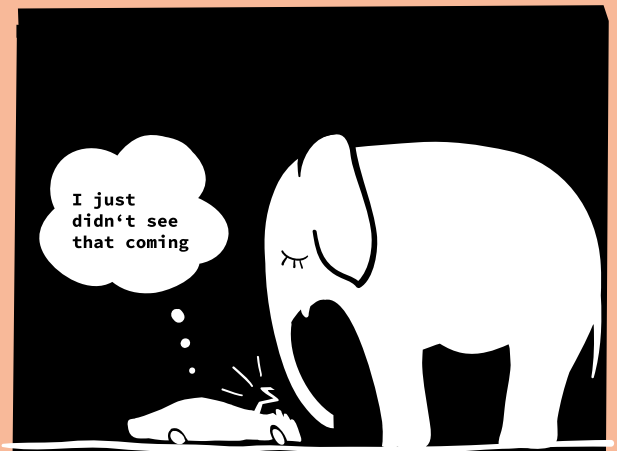
Anmerkung:
1 Misselhorn, C.: *Grundfragen der Maschinenethik*. Stuttgart 2019[3]

Today, we can already activate the autopilot function which automatically and comfortably drives our car or flies our plane.

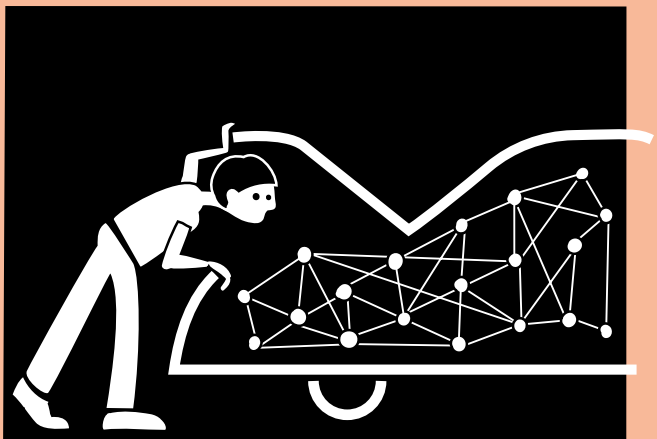
But ultimately, we are responsible. The systems are not autonomous - or self-driving.

Within a few years, this might change.

We would have plenty of room to sleep, work, read or play, sober or drunk after a party.



Or maybe not. Nearly every car accident involves some sort of unforeseen circumstances, the AI must confront for the first time.



Moreover, user tests show that we need autonomous systems to explain their decision to feel safe. Not easy.

www.weneedtotalk.ai