

Oliver Bendel

## Die Moral der Maschinen

### Überlegungen zur Maschinenethik

Bendel, Oliver. Die Moral der Maschinen: Überlegungen zur Maschinenethik. In: [inside-it.ch](http://inside-it.ch), 24. Oktober 2012. Über <http://www.inside-it.ch/articles/30517>. Das vorliegende Dokument ist das Originalmanuskript. Die veröffentlichte Version kann davon in Details abweichen. Je nachdem, ob der Artikel für deutsche oder schweizerische Zeitungen geschrieben ist, wird die entsprechende Rechtschreibung verwendet.

Ein schöner Tag im Sommer 2015. Der New Autonomous Car, kurz NAC genannt, fährt die Strasse am Fluss entlang. Mit seiner elegant geschwungenen Schnauze, die fast ein Schnabel ist, ist er ein Augenschmaus. Die Ente am gegenüberliegenden Ufer wirkt aufgeregt. Am nicht vorhandenen Steuer sitzt die junge Nobelpreisträgerin Kim. Ein bisschen schneller, sagt sie zu NAC, der das als Eingriff in seine Autonomie empfindet. Beleidigt erhöht er trotzdem das Tempo. Kim ist auf dem Weg zu einem Kongress zum Thema Maschinenethik. Google, BMW, die FU Berlin und andere Unternehmen und Hochschulen, die an selbstständig fahrenden Autos und mobilen Assistenzgeräten arbeiten, sind dabei. Die EU präsentiert die Ergebnisse ihres CAMUS-Projekts, das auf das SARTRE-Projekt folgte. Und Kim spricht über den von ihr entwickelten NAC. Die Altstadt taucht auf, die Strassen werden enger. Dann geht alles ganz schnell. Die Bremsen versagen. Im entgegenkommenden Wagen ist eine Frau mit riesiger Sonnenbrille zu erkennen. Sie liest gerade ein Buch von Martin Luther King. NAC kann ihrem smarten Gefährt innert Millisekunden entlocken, dass sie eine berühmte Bürgerrechtlerin ist. Der Mann auf dem Trottoir zur Linken ist der Bürgermeister der Stadt, wie bereits die integrierte Gesichtserkennung von NAC herausgefunden hat. Er hasst Konferenzen und trifft seine Freunde zum Kegeln. Zur Rechten wuseln kleine Kinder: rote Wangen, blitzende Augen, fröhliches Lachen. NAC muss sich entscheiden: Wen will, soll, kann oder muss er überfahren? Er aktiviert sein Moralmodul und den Airbag. An die Leitung der Konferenz schickt er eine Nachricht: Es wird etwas später.

Die Ethik bezieht sich üblicherweise auf die Moral von Menschen, von Individuen und Gruppen, und in gewissem Sinne auf die Moral von Organisationen. Es kann aber auch um die Moral von Maschinen wie Agenten, Robotern und Drohnen sowie z.B. von medizinischen und technischen Überwachungsanlagen, Rechnern im Hochfrequenzhandel und eben selbstständig fahrenden Autos gehen, auf die der verbreitete Begriff „Unmanned Ground Vehicle (UGV)“ nur eingeschränkt passt. Handelt es sich um solche mehr oder weniger autonome Programme und Systeme, mag man von einer Maschinenethik sprechen und diese der Informationsethik (genauer Computerethik und Netzethik) und der Technikethik zuordnen oder aber, weil sie anscheinend auf eine nichtmenschliche Moral zielt, auf eine neue Stufe stellen.

Agenten sind Programme, die im Auftrag oder im Sinne von Benutzern und Systemen Aufgaben erledigen und dabei autonom und mit einer gewissen Intelligenz agieren. Roboter sind mobile oder stationäre Maschinen, die von Programmen oder Menschen gesteuert, Drohnen fliegende, unbemannte Maschinen, die von Programmen oder Menschen gelenkt werden. Selbstständig fahrende Autos sind zurzeit an verschiedenen Orten der Welt unterwegs und könnten eines

Tages den Verkehr revolutionieren. Autonome Maschinen haben eine lange Geschichte. Ein Höhepunkt war – noch vor den Automaten des Pierre Jaquet-Droz aus La Chaux-de-Fonds – die mechanische Ente von Jacques de Vaucanson aus den 40er-Jahren des 18. Jahrhunderts.

Agenten sind gemäss der obigen Definition autonom unterwegs, Roboter und Drohnen unter bestimmten Umständen. Je nach Begriff und Kontext kann die Eigenständigkeit von Agenten eingeschränkt sein. Entscheidend ist, dass alle drei Formen von Maschinen das Potenzial zur Autonomie haben. Und dass Autonomie und Moral in einem speziellen Verhältnis stehen. Erstens macht die Autonomie von Maschinen ihre Moral möglich. Ohne selbstständiges Denken und Handeln sind sie kein Subjekt der Moral. Zweitens macht die Autonomie von Maschinen ihre Moral notwendig. Ihr selbstständiges Denken und Handeln lässt sie als Subjekte auftreten, ähnlich wie Menschen, und sie sind Teil von Situationen und Prozessen mit sittlichen Implikationen. Drittens macht die Moral von Maschinen ihre Autonomie erst möglich. Ohne moralisches Denken und Handeln sind sie kein autonomes Subjekt, zumindest keines, das wir ernst nehmen und dem wir vertrauen würden – wengleich die Versuche von Joseph Weizenbaum mit ELIZA gezeigt haben, dass unter manchen Menschen eine enorme Vertrauensseligkeit herrscht. Viertens macht die Autonomie von Maschinen aus ihnen auch ein Objekt der Moral. Nicht eine Moral der Maschinen also, sondern eine Moral gegenüber Maschinen; wir müssen uns ihnen gegenüber eines Tages in einer bestimmten Weise verhalten.

### **Es war auch schon einfacher, ein Haus zu sein ...**

Ruft man die Ansätze einer normativen Ethik vor dem geistigen Auge auf, mit ihren ganz unterschiedlichen Urhebern und Motiven, dann wird rasch deutlich, dass nicht jeder von ihnen geeignet ist für die Umsetzung einer Maschinenmoral. Zunächst erscheint eine Pflicht- oder Pflichtenethik in der Tradition von Immanuel Kant als Option. Mit einer Regel, einer Pflicht vermag eine Maschine mit einem Elektronengehirn etwas anzufangen, ist doch ein Algorithmus nichts anderes als eine Handlungsvorschrift. Zum Beispiel könnte man ihr beibringen, nicht zu lügen. Sie würde dann auf die Frage, wie das Wetter in Olten ist, wahrheitsgemäss antworten, dass es regnet – natürlich nur, wenn dies der Fall wäre. Wolfram Alpha pflegt in dieser Weise zu antworten, wobei man dieser (Such-)Maschine der etwas anderen Art beigebracht hat, die Wahrheit zu sagen, wenn sie sie kennt, oder das, was sie für die Wahrheit hält. Manch eine normale Suchmaschine wie DuckDuckGo fragt bei Bedarf die britisch-amerikanische Wundermaschine um Rat. So oder so dürfte die Antwort in den meisten Fällen unproblematisch sein. Schlimmstenfalls scheint in Olten halt doch die Sonne.

Bewegen wir uns aber weiter, in irgendeine andere Stadt, bis zu einem intelligenten Haus. In dem Haus halten sich Juden versteckt, vor der Tür stehen Nazis. Es handelt sich um eine Abwandlung eines berühmt gewordenen Beispiels, das in seiner allgemeineren Form auf Kant und den Schweizer Benjamin Constant zurückzuführen ist. Was sagt das intelligente Haus, wenn die Nazis fragen, ob sich Juden in ihm versteckt halten? Gleichgültig, was es sagt: Es befindet sich im gleichen Dilemma wie ein Mensch. Es fällt ihm nur schwerer, sich der Pflichtethik zu entziehen, ausser wenn eine weitere Regel die genannte ausser Kraft setzt: Du sollst nicht lügen, es sei denn, du rettest Leben oder vermeidest Leiden damit. Das wird die künstliche Intelligenzbestie freilich in neue Schwierigkeiten bringen. Ohne Zweifel war es früher einfacher, ein Haus zu sein.

Es könnte auch um ganz andere Fragen gehen, etwa in Bezug auf Ressourcen: Wenn das Gebäude in einer gewissen Situation, beispielsweise während einer Ölkrise, nur zwei seiner Zimmer in normaler Masse heizen kann – wen wird es bevorzugen? Den Single, die Familie, die Einheimischen, die Zugewanderten, die Katze, den Hund? Oder behandelt es alle gleich und lässt alle bei zehn Grad frieren? Oder setzt es auf das Zufallsprinzip? Das kann es tun, aber es vermeidet dadurch nicht sein moralisches Dilemma.

### **Nur Regeln oder doch mehr?**

Michael Anderson und Susan Leigh Anderson gehen als Herausgeber des 2011 erschienenen Buchs „Machine Ethics“ der Frage nach, ob und wie autonome Systeme in moralischer Weise handeln sollen und können. Schon vor Jahrzehnten hat man in Wissenschaft und Literatur über diese Frage nachgedacht; aber eine entsprechende „Teilbereichsethik“ ist erst in den letzten Jahren entstanden. Offensichtlich wird die Notwendigkeit gesehen, das Verhalten von Maschinen im Kontext der Moral zu reflektieren und Philosophen (vor allem Ethiker) und Vertreter der Künstlichen Intelligenz (KI) darüber nachdenken zu lassen. Auch Informatiker und Wirtschaftsinformatiker sowie Psychologen und Soziologen können sich in die Diskussion einbringen. In der Einführung heisst es: „The subject of this book is a new field of research: Developing ethics for machines, in contrast to developing ethics for human beings who use machines.“ Wir benutzen Maschinen, und die Maschinen benutzen uns. Wie sollen die Maschinen mit uns umgehen, und wie sollen sie sich entscheiden in Situationen, in denen wir in unserer Identität und in unserer Existenz bedroht werden? Die hochkarätigen Beiträge verdienen eine aufmerksame Lektüre.

Näher an der Realität als mit dem Beispiel des ressourcenverwaltenden intelligenten Hauses ist man mit demjenigen des selbstständig fahrenden Autos (eine Abwandlung eines weiteren berühmten Beispiels, nämlich des Trolley-Problems). Wenn es nicht ausweichen, sich aber zwischen dem einen und dem anderen Unfall entscheiden kann – was wird es tun? Wird es die Ente töten oder den Hund? Wird es die Kinder aus ihrem jungen Leben stossen oder die Bürgerrechtlerin, die man gerade ausgezeichnet hat, aus ihrem erfolgreichen? Und kann es mehr, als irgendeine Regel zu befolgen? Kann es Mitleid entwickeln, um Arthur Schopenhauers Ethik zu genügen? Oder nach Glückseligkeit streben, um die Anforderungen eines Aristoteles zu erfüllen? Es könnte, wenn es nicht gerade NAC heisst, allenfalls Mitleid simulieren, und die Glückseligkeit wäre immer die der anderen, der humanoiden Nichtmaschinen. Kann es die Folgen seines Handelns bedenken und in diesem Sinne verantwortlich agieren? Kann es also einer Folgen- oder Verantwortungsethik verpflichtet sein? Kann es dazulernen, etwa mit Hilfe evolutionärer Algorithmen, und irgendwann einschätzen, ob es die Regel befolgen oder die Folgen einbeziehen und seiner Urteilsfähigkeit und seinen Erfahrungen trauen soll?

### **Und wer ist nun tot?**

Für NAC hat Kim einen integrierten Ansatz gefunden. Es werden die Likes und die Favorites der sogenannten sozialen Medien ausgewertet, Blogosphären und Wikipedia analysiert, Artikel und Kommentare gescannt. Die Maschinenethik hat in diesem Fall nicht nur mit Computer- und Netzethik zu tun, sondern auch mit Medienethik und mit der Moral der Medien und Plattformen. Weiterhin kann das Moralmodul auf eine Datenbank mit Cases zugreifen. Wie haben sich Menschen in vergleichbaren Situationen verhalten? Und wie wurde ihr Verhalten von anderen beurteilt, von Individuen und Medien? Nicht zuletzt stehen die Gespräche mit Kim zum

Abruf bereit, in deren Verlauf NAC dazugelernt hat, die Erzählungen der Nobelpreisträgerin, ihre Beurteilungen, ihr Gefühl. Die Maschine orientiert sich also an Menschen (und Manipulationen), an den moralischen Fakten (und Fakes) der Vergangenheit und der Gegenwart. Sie hat mit ihrer Erfinderin eine persönliche moralische Referenz. Und sie hat ihren eigenen Kopf, auch wenn in diesem nur Chips sind. Es ist nicht völlig klar, was am Ende den Ausschlag gibt. Jedenfalls überlebt der Bürgermeister den Aufprall nicht. Seine Kegelfreunde sind so entrüstet darüber und mobilisieren derart die Massen, dass die Konferenz im Jahre 2016 in einer anderen Stadt stattfinden muss. Die Mehrheitsmeinung im Web 2.0 gerät in eine Schieflage, die NAC ernsthaft an den Menschen und deren angeblicher Schwarmintelligenz zweifeln lässt. Eine Weile muss er sich sogar vor dem Mob in einer extra angemieteten Garage verstecken.

Die Maschinenethik wird ein wichtiger Prüfstein der Ethik sein. Sie wird neue Subjekte und Objekte der Moral beschreiben, neue Denkweisen und Handlungen. Sie wird aufzeigen, welche normativen Ansätze jenseits der auf Menschen bezogenen Ethik sinnvoll sind. Und sie muss herausfinden, ob und wie weit die normativen Ansätze überhaupt maschinenverarbeitbar sind. Der Anwendungsbereich der Maschinenethik hat zudem ökonomische und technische Implikationen. Wir brauchen Ethik nicht mehr nur, um unser Zusammenleben zu beschreiben und zu überprüfen, sondern auch um unser Überleben in der Informationsgesellschaft zu sichern. Wie sollen die Unternehmen mit der Möglichkeit der Implementierung moralischen Verhaltens umgehen? Wie gestalten sie die Mensch-Maschine-Schnittstelle, um Einfluss auf die autonomen Maschinen zu behalten, oder ist ein solcher Einfluss überflüssig und schädlich? Welche ethischen und rechtlichen Herausforderungen stehen für sie im Kontext der Maschinenethik an? Solche Fragen muss die Wirtschaft in Zukunft in Zusammenarbeit mit der Philosophie, mit der KI und mit der einen oder anderen aufgeschlossenen Disziplin beantworten.

### **Autor**

Oliver Bendel ist studierter Philosoph und promovierter Wirtschaftsinformatiker und leitete technische und wissenschaftliche Einrichtungen an Hochschulen. Heute lehrt und forscht er als Professor für Wirtschaftsinformatik an der Hochschule für Wirtschaft in Brugg und Olten (Fachhochschule Nordwestschweiz), mit den Schwerpunkten Wissensmanagement, Social Media, Mobile Business und Informationsethik. Das Favicon seiner Website [www.oliverbendel.net](http://www.oliverbendel.net), eine kleine Ente, hat ihn zu der Geschichte in dem Artikel inspiriert.

### **Literatur**

Bendel, Oliver. Die Rache der Nerds. UVK, Konstanz und München 2012.

Bendel, Oliver. Maschinenethik. Beitrag für das Gabler Wirtschaftslexikon. Gabler/Springer, Wiesbaden 2012. Über <http://wirtschaftslexikon.gabler.de/Definition/maschinenethik.html>.