

Detlef Horster

Künstliche Intelligenz am Beispiel des autonomen Fahrens

1. Einführung.....	1
2. Der Handlungsbegriff.....	2
3. Das Trolley-Dilemma.....	5
4. Philippa Foots Diskussion des Trolley-Problems	6
5. Der Unterschied der beiden Fälle von Foot.....	8
6. Die Konsequenzen für den Algorithmus des autonomen Fahrens	9

1. Einführung

Am 26. Mai 2024 erschien in der Hannoverschen Allgemeinen Zeitung ein Artikel mit der Überschrift „Bedroht die KI die Menschheit“. Darin steht der Satz: „Während einige glauben, KI sei die größte Errungenschaft der Menschheit und werde das Leben in naher Zukunft entscheidend verbessern, malen andere seit Jahren Horrorszenarien an die Wand, eine superintelligente KI könne die Menschheit gar zerstören.“ Einig sind sich Kritiker und Befürworter darin, dass in naher Zukunft vieles durch die Künstliche Intelligenz bearbeitet werden wird. Dadurch soll alles schneller und effektiver werden: Zum Beispiel werden unsere Pakete von Robotern ausgetragen, und die Roboter werden am Fließband arbeiten und vieles mehr. Eine starke Entlastung des Menschen und seiner Arbeit wird somit in Aussicht gestellt. Bei der Rechtsanwaltskanzlei Ratis in Passau werden Rechtsfälle mithilfe von Algorithmen bearbeitet. Dazu werden unendliche viele Urteile eingespeist. Die Ergebnisse der KI können also nicht falsch sein. Die Anwälte arbeiten nur noch 30 Stunden in der Woche. Ohne die Arbeit, die heute Algorithmen übernehmen, haben sie das zwei- bis dreifache gearbeitet, also 60-90 Stunden in der Woche.

Dennoch gibt es große Befürchtungen und Ängste. In dieser Hinsicht stehen sich Skeptiker und Befürworter der Künstlichen Intelligenz unversöhnlich gegenüber. Die einen sagen, es entgleitet uns alles und wir werden demnächst von nicht mehr unter

Kontrolle zu bringenden Computerprogrammen beherrscht, die anderen stellen die großen Vorteile der menschlichen Arbeitsentlastung in den Vordergrund.

Wir Menschen sind es aber, die den Algorithmus entwickeln und bestimmen, wie das entsprechende Gerät zu agieren hat. Ich möchte Ihnen das am Beispiel des autonomen Fahrens verdeutlichen. Ich habe das ausgewählt, weil sich daran am besten die Problemlage verdeutlichen lässt.

Stellen Sie sich vor, Sie fahren Auto, lassen das Steuer los und sagen dem Bordcomputer, wo Sie hinwollen. Sie lesen Zeitung, Akten oder suchen sich einen Film aus, den Sie ansehen wollen oder Sie beginnen mit Ihrer Arbeit. Oder der Urlaub beginnt schon vor Ihrer Haustür. Die Bahn verspricht für Ihren Urlaub, dass er gleich am ersten Tag beginnt. Nur beim autonomen Fahren brauchen Sie nicht einmal mehr dafür zu sorgen, dass die Koffer zum Bahnhof kommen. Heute geht das schon so weit, dass Sie Ihr Auto automatisch eine passenden Parklücke finden lassen. Inzwischen sind die Autos standardmäßig so ausgerüstet. Wenn das Auto eine passende Parklücke gefunden hat, bekommen Sie die Aufforderung, den Rückwärtsgang einzulegen und das Steuer los zu lassen. Das Auto parkt sich dann selbsttätig ein. Was meinen Sie: Ist das noch Handeln?

2. Der Handlungsbegriff

Der Handlungsbegriff ist ein zentraler Begriff der Ethik. An dieser Stelle vielleicht ein kleiner Einschub. Was ist der Unterschied zwischen Ethik und Moral? Was wir jetzt hier tun, ist der Ethik zuzuordnen. Wir reflektieren theoretisch über das, was in einer Situation zu tun ist. Das machen wir beispielsweise auch in Universitätsseminaren. Doch für das richtige Handeln im Alltag müssen wir wissen, was richtig und falsch ist. Das nennen wir dann Moral. Beim moralischen Handeln müssen wir wissen, welche Rechte und Pflichten wir haben. Wir kennen als handelnde Menschen immer schon

moralische Regeln.¹ Die einfachen Menschen (plain men) sind über die moralischen Regeln genauso gut informiert wie Moralphilosophen.²

Was nun das Handeln betrifft, halte ich die Definition von Max Weber für plausibel und praktikabel, mit der man bei der Bewertung, ob wir es mit einer Handlung zu tun haben, erst einmal ein gutes Stück weiter kommt. Weber schreibt: „Handeln’ soll [...] ein menschliches Verhalten (einerlei ob äußeres oder innerliches Tun, Unterlassen oder Dulden) heißen, wenn und insofern als der oder die Handelnden mit ihm einen subjektiven Sinn verbinden.“³ Wenn ich ein Glas umstoße, kann das ein Versehen sein. Wir definieren es dann als Verhalten. Stoße ich aber dieses Glas um, um den Gastgeber zu ärgern, dann ist es ein Handeln, weil meinem Verhalten ein Sinn unterlegt werden kann, wenn auch ein verwerflicher.

Nun muss man noch darüber nachdenken, wie es ist, wenn Kausalketten unterbrochen oder gestört werden, und wir zu ganz anderen als den intendierten Ergebnissen kommen, ob es dann auch noch Handeln ist. Die Störungen von Kausalketten werden im Zusammenhang mit dem autonomen Fahren eine ganz wesentliche Rolle spielen. Wenn wir uns in den Straßenverkehr begeben, haben wir die Absicht, niemanden zu gefährden und zu Schaden kommen zu lassen. Ein Unfall ist darum eine Unterbrechung oder Störung der Kausalkette. So lautet auch die Definition des Bundesgerichtshofes: Ein Unfall ist ein unvorhergesehenes, plötzlich eintretendes Ereignis.⁴ Der Unfall ist also nicht mit Vorsatz herbeigeführt worden. Folgen wir der weberschen Definition, so ist auch bei einem Unfall der ursprüngliche Sinn, niemanden zu gefährden und zu Schaden kommen zu lassen. Ein Unfall ist regelmäßig ein Abweichen von diesem Kausalverlauf.

Eine weitere Besonderheit, die zu berücksichtigen ist: Es geht beim autonomen Fahren nicht um Handeln im klassischen Sinne, aber um etwas, das auf das Handeln anderer,

¹ Vgl. William David Ross: *The Right and the Good* [1930], new edition, ed. by Philip Stratton-Lake, New York, 2002, S. 40.

² Vgl. ders.: *Foundations of Ethics. The Gifford Lectures 1935-36*, New York 2000, S. 311.

³ Max Weber: *Wirtschaft und Gesellschaft*, 5. revidierte Auflage, Tübingen 1976, S. 1.

⁴ Vgl. BGHSt 24, 382 ff.

neuerer Art hinausläuft und vor allem die Folgen von Handeln im klassischen Sinne hat. Im Moment des Fahrens handelt nicht ein Mensch, sondern ein Algorithmus. Bei einem Unfall zeigen sich dennoch dieselben Folgen wie beim klassischen Handeln. Allerdings wird nach allen Vorhersagen die Unfallgefahr um 90% minimiert, weil menschliches Versagen ausgeschlossen sein wird, denn die große Mehrzahl der Unfälle im Straßenverkehr geht bislang auf menschliches Versagen zurück. Wir hätten dann nicht mehr 3000 Verkehrstote jährlich in Deutschland zu beklagen, sondern 300. Das führt zu der vielleicht ungewöhnlichen Frage: „Wann verbietet die Politik endlich das Fahren von Menschen?“

Zusammengefasst heißt das also, dass wir es hier mit Handeln zu tun haben, denn wir Menschen sind es, die den Algorithmus programmieren. Es kann natürlich beim Programmieren des Algorithmus einen Fehler geben, wodurch ein Unfall verursacht wird. Wenn ein solcher Programmierungsfehler einmal erkannt ist, kann er gleich für alle Systeme bzw. für alle Autos behoben werden, denn er bezieht sich nicht nur auf eine einzige Abweichung vom Handeln.

Wie sollte nun der Algorithmus für eine Situation programmiert werden, in der in einer Verkehrssituation die Frage aufgeworfen wird, ob ein Rentner überfahren wird oder drei Schulkinder? Das nun ist ein Fall für Moralphilosophen. Das ist inzwischen unbestritten. In der Zeitschrift „Nature“ vom 25. April 2019 wird bedauert, dass sich die momentane Forschung „nur ad hoc um die Auswirkungen der Algorithmen auf Menschen und Gesellschaft“ kümmere. Es seien vor allem Mathematiker und Ingenieure, die sich mit künstlicher Intelligenz beschäftigten. „Derzeit sind die Wissenschaftler, die das Verhalten von Maschinen untersuchen, dieselben Wissenschaftler, welche die KIs geschaffen haben. Diese konzentrieren sich aber vor allem auf die Funktionstüchtigkeit ihrer künstlichen Intelligenz.“ Dies sei zu bedauern.⁵

⁵ Andrian Kreye: Moralische Maschinen, in: Süddeutsche Zeitung vom 25. April 2019

Doch die Beamten des Bundesverkehrsministeriums haben begriffen, dass man sich in diesem Zusammenhang mit ethischen Problemen zu beschäftigen habe. Mit dem oben genannten Fall beschäftigte sich die Ethik-Kommission zum autonomen Fahren beim Bundesverkehrsministerium. Es gibt in der angenommenen Situation keine Alternative zum Überfahren von mindestens einem Menschen. Kommt man in einer solchen oder ähnlichen Einrichtung nicht zu einem Ergebnis, wird die Gründung einer Arbeitsgruppe beschlossen, die sich dann weiter damit beschäftigen soll. Auch die Ethik-Kommission beim Bundesverkehrsministerium kam zu keinem Ergebnis. Sie schlug eine öffentliche Einrichtung vor, um Erfahrungen systematisch zu verarbeiten.⁶

3. Das Trolley-Dilemma

In dem Bericht dieser Ethik-Kommission wurde das Trolley-Problem angeführt, das von Philippa Foot 1967 in die moralphilosophische Diskussion eingebracht wurde und inzwischen unabdingbarer Bestandteil gängiger moralphilosophischer Lehrbücher ist. Ein Fahrer – schreibt Foot – „einer außer Kontrolle geratenen Straßenbahn [...], die er nur von einem Gleis auf ein anderes steuern kann: fünf Männer arbeiten auf dem einen Gleis, einer auf dem anderen; jeder auf dem Gleis, das er ansteuert, wird getötet werden.“⁷ Was soll der Fahrer tun? Der Kommentator der Ethik-Kommission beim Bundesverkehrsministerium schreibt dazu: „Darauf gibt es zwei Antworten, je nachdem welcher philosophischen Ethik man anhängt. Entweder rechtfertigt das Glück der größeren Zahl den Schaden eines unschuldigen Einzelnen – oder nicht.“⁸ So wird es in der Regel in den philosophischen Lehrbüchern behandelt. Man löst den Fall nach der deontologischen oder nach der utilitaristischen Ethik. Ist die Straßenbahn auf dem Gleis mit fünf Männern, darf der Fahrer nach deontologischer Auffassung die Bahn nicht umlenken auf das Gleis mit nur einem Mann. Menschenleben dürfen nicht gegeneinander aufgerechnet werden. Das ist auch im deutschen Recht so. Unter dem Eindruck von 9/11 wurde am 11. Januar 2005 ein Bundesgesetz zur Luftsicherheit erlassen, in dem es hieß: „Die unmittelbare Einwirkung mit Waffengewalt ist nur

⁶ Christian Weber: Wer muss sterben?, in: Süddeutsche Zeitung vom 25. Juni 2017

⁷ Philippa Foot: Das Abtreibungsproblem und die Doktrin der Doppelwirkung, in: Anton Leist (Hg.): Um Leben und Tod. Moralische Probleme bei Abtreibung, Künstliche Befruchtung, Euthanasie und Selbstmord, Frankfurt/M. 1990, S. 196-211 [hier: S. 200].

⁸ Christian Weber, a.a.O.

zulässig, wenn nach den Umständen davon auszugehen ist, dass das Luftfahrzeug gegen das Leben von Menschen eingesetzt werden soll, und sie das einzige Mittel zur Abwehr dieser gegenwärtigen Gefahr ist.“ (§ 14, Abs. 3 LuftSiG) Am 15. Februar 2006 entschied das Bundesverfassungsgericht in einem Normenkontrollverfahren, in dessen Urteil es heißt: „Der Staat, der zur Abwehrmaßnahme des § 14 Abs. 3 LuftSiG greift, missachtet die entführten Betroffenen als Subjekte mit Würde und unveräußerlichen Rechten. Sie werden dadurch, dass ihre Tötung als Mittel zur Rettung anderer benutzt wird, verdinglicht und zugleich entrechtlicht; indem über ihr Leben von Staats wegen einseitig verfügt wird, wird den als Opfern selbst schutzbedürftigen Flugzeuginsassen der Wert abgesprochen, der dem Menschen um seiner selbst willen zukommt.“⁹ Daraufhin wurde das Gesetz geändert.

Folgen wir dieser Entscheidung, dann sieht das in der eingangs geschilderten Situation der Ethik-Kommission so aus, dass die Schulkinder keinen höheren Wert als der Rentner haben. Menschenleben ist Menschenleben, egal welche Qualität es hat. Nach utilitaristischer Auffassung hingegen muss man die größere Zahl an Menschen retten.

In der Ethik-Kommission war man der Auffassung, dass das Trolley-Problem für die Programmierung des Algorithmus für das autonome Fahren nicht weiter hilft. Wenn man sich das Trolley-Problem so ansieht, stimmt das. Aber das ist eine verkürzte Sichtweise der Mitglieder der Ethikkommission beim Bundesverkehrsministerium. Ich habe mir das Trolley-Dilemma daraufhin noch einmal genauer angesehen. Lesen wir also Philippa Foot neu. Ich bin überzeugt, dass wir dann die Situation nicht auf die Differenz von Utilitarismus und Deontologie verkürzen können, denn wir haben mehr an der Hand für eine Situation, die relevant für die Programmierung des Algorithmus ist.

4. Philippa Foots Diskussion des Trolley-Problems

Foot hat das Trolley-Dilemma bei dessen Einführung mit einem weiteren Beispiel konfrontiert und verglichen. Das konfrontierende Beispiel lautet: „Angenommen, ein

⁹ BVerfGE 115, 118.

Richter oder Beamter sieht sich Aufrührern gegenüber, die fordern, einen Angeklagten für ein bestimmtes Verbrechen zu bestrafen, und damit drohen, blutige Rache an einem bestimmten Teil der Gemeinschaft zu üben. Da der wirkliche Verbrecher unbekannt ist, sieht sich der Richter nur so in der Lage, das Blutvergießen zu verhindern, daß er eine unschuldige Person fälschlicherweise anklagt und hinrichten läßt.“¹⁰ Dann stellt die Autorin die Frage, „warum wir ohne Zögern sagen sollten, daß der Fahrer auf das weniger frequentierte Gleis steuern sollte, während die meisten von uns die Vorstellung empörend fänden, daß der unschuldige Mann zu Unrecht verurteilt wird.“¹¹

- Beide Fälle enthalten den sogenannten Doppelleffekt, d.h., die Handlung kann in jedem der beiden Fälle zwei Wirkungen hervorrufen, die eine ist die beabsichtigte und gewollte, die andere die vorhergesehene, aber keineswegs gewollte. – Die beabsichtigte und gewollte ist in beiden Fällen, möglichst wenig Menschen zu töten und möglichst viele zu retten. – Die vorhergesehene, aber keineswegs gewollte ist im einen Fall, viele Menschen auf dem Gleis zu töten und im Richterfall, einen Unschuldigen umzubringen.
- In beiden Fällen ist es unvermeidlich, dass jemand stirbt.
- In beiden Fällen ist es so, dass dann, wenn wir uns weigern zu handeln, die größere Anzahl von Menschen getötet werden wird.
- Der Unterschied liegt laut Foot aber darin, dass der Richter im einen Fall etwas tut und im anderen der Fahrer etwas zulässt.
- In dem einen Fall kann der Richter jemanden überlegt sterben lassen und im anderen tut der Fahrer etwas, was er nicht will, nämlich den Mann auf dem Gleis töten.
- Philippa Foot schreibt von direkter und indirekter Absicht. Der Mensch mit indirekter Absicht lässt etwas zu. Das ist bei dem Fahrer der Bahn ein Abweichen vom Kausalverlauf, denn er will ja niemanden gefährden und zu Schaden kommen lassen. Jemand mit direkter Absicht handelt gewollt.¹² Letzteres ist das, was Max Weber Sinn nennt.

¹⁰ Philippa Foot, ebenda.

¹¹ Philippa Foot, a.a.O., S. 201.

¹² Philippa Foot, ebenda.

- Eine weitere Unterscheidung wird von Foot eingeführt; die Unterscheidung von negativen und positiven Pflichten. Negativ heißt, dass es ein moralisches Gebot ist, dass man niemanden verletzen darf. Positiv heißt, dass wir jemandem helfen sollen, beispielsweise den Hungernden in der Welt.¹³

5. Der Unterschied der beiden Fälle von Foot

Was machen wir nun mit dieser Reihe von Unterscheidungen, die Philippa Foot trifft? Beginnen wir mit dem Fahrer. Er ist konfrontiert mit zwei negativen Pflichten. Er darf weder die fünf Menschen auf dem einen Gleis verletzen, noch den einen auf dem anderen. Der Richter hingegen steht vor einer positiven Pflicht, nämlich diesem bestimmten Teil der Gemeinschaft zu helfen, den die Aufrührer meinen und töten wollen. Daneben hat er die negative Pflicht, den Unschuldigen nicht zu verletzen. Darum ist es wegen dieses gravierenden Unterschieds nicht möglich, vom Fall des Fahrers der Straßenbahn auf den des Richters zu schließen. „Interessanterweise wiegt sogar das Vorliegen einer stärksten Pflicht zu aktiver Hilfe nicht so schwer wie eine bestehende negative Pflicht. [...] Deshalb ist es nicht widersprüchlich von uns, zu denken, daß der Fahrer auf das Gleis steuern muß, auf dem nur ein Mensch steht, während der Richter [...] nicht töten darf, um den Aufruhr zu verhindern.“¹⁴

Die Unterscheidung ist also die zwischen der negativen Pflicht, Schaden zu vermeiden und der positiven, Hilfe zu bringen. In ähnlicher Weise unterscheidet auch William David Ross. Beim Konflikt zwischen zwei Pflichten sei es für jedermann evident, dass man zuerst den negativen Pflichten nachkomme. Ross bezeichnet das nicht so, aber benennt die vorrangigen Pflichten. Beispielsweise müsse man erst seine Schulden bezahlen, bevor man für wohltätige Zwecke spende.¹⁵ Das sei selbst dem einfachen Mann (plain man) evident, meint Ross. Auch für ihn haben die negativen Pflichten Vorrang. Ebenso ist das für Foot evident, denn wir würden es „empörend“ finden, wenn der Richter den unschuldigen Mann verurteilen und hinrichten lassen würde. Er

¹³ Philippa Foot, a.a.O., S. 205.

¹⁴ Philippa Foot, a.a.O., S. 206.

¹⁵ William David Ross: The Right and the Good [1930], a.a.O., S. 30.

hat primär die Pflicht den unschuldigen Mann zu retten, also seiner negativen Pflicht nachzukommen.

6. Die Konsequenzen für den Algorithmus des autonomen Fahrens

Das Trolley-Problem bringt also im Gegensatz zur Meinung der Mitglieder der Ethik-Kommission beim Bundesverkehrsministerium doch etwas, wenn wir die Einführung des Trolley-Problems durch Philippa Foot **genauer** lesen. Die Überlegungen von Foot führen dazu, zu sehen, dass für Unfallknotenpunkte, die negativen Pflichten, Schaden zu vermeiden, Vorrang haben. D.h. wir dürfen auf keinen Fall jemandem Schaden zufügen. Bei der Konfrontation mit zwei negativen Pflichten, die der Fahrer der Straßenbahn hat, kann man nicht beide negativen Pflichten gleichzeitig erfüllen. Da gilt es abzuwägen und den größeren Schaden abzuwenden, d.h., die größere Anzahl von Menschen nicht zu verletzen, also die fünf Männer auf dem einen Gleis. Er muss die Straßenbahn auf das Gleis mit einer Person lenken. Das führt bei dem eingangs geschilderten Fall dazu, bei dem auch zwei negative Pflichten gegenüber stehen, dass so programmiert werden muss, dass die geringste Anzahl von Menschen geschädigt wird. In dem Fall also der Rentner. Dabei ist nun nicht entscheidend, ob es sich um einen Rentner oder um Schulkinder handelt. Wichtig dabei ist, dass die geringste Anzahl von Menschen zu Schaden kommt. Die Qualität der Menschen darf hier nicht den Ausschlag geben. Jeder Mensch ist in einer solchen Situation als Mensch zu betrachten, unabhängig von seinen Eigenschaften. So hatten es auch die Richter des Bundesverfassungsgerichts in der oben genannten Entscheidung gesehen. Die Programmierung des Algorithmus wird also genauso vorgenommen, wie bei einer Entscheidung beim Handeln im klassischen Sinne. Negative Pflichten haben Vorrang. Menschen, die nicht in das Unfallgeschehen involviert sind, müssen besonders geschützt werden. Nehmen wir an, dass der Rentner und die drei Schulkinder gerade die Straße überqueren und das herannahende Auto nicht gesehen haben. Aber dann gibt es noch gänzlich am Verkehrsgeschehen Unbeteiligte, z.B. den Fußgänger auf dem Trottoir. Der muss ebenfalls geschützt werden. Darum darf der Algorithmus nicht so programmiert werden, dass das Auto ausweicht und auf das Trottoir gelenkt wird, wenn sich dort Menschen befinden, die in der Situation zu Schaden kommen können.

Der Algorithmus verkörpert nach der weberschen Definition den Sinn des Handelns. Aber auch hier kann es ein Abweichen vom Kausalverlauf geben, das dann im Sinne der Definition des Bundesgerichtshofes ein Unfall ist, den man auch beim autonomen Fahren nicht ausschließen kann. – Sie sehen nun – um auf den Anfang zurückzukommen –, dass dadurch, dass wir die Programmierung in der Hand haben, wir Menschen nicht entmündigt und von der Künstlichen Intelligenz beherrscht werden.

Probleme:

- Um ein Kind nicht zu überfahren, wurde eine Frau mit Kinderwagen in den Straßengraben gedrängt.
- In den USA gab es einen Zusammenstoß mit einem weißen Lieferwagen, weil die Farbe weiß für einen Lieferwagen nicht einprogrammiert war.
- In Südkorea starb Ende 2023 ein Arbeiter in einer Paprikaverpackungsanlage, weil ein Roboter ihn für Gemüse hielt und versuchte, ihn auf ein Förderband zu laden.
- Das Problem mit dem schlechten Gewissen, das es bei Robotern nicht gibt: Gebrannte Mandeln klauen.