

Die Moral einer Maschine

„Maschinen sind im wahrsten Sinne des Wortes unmündig. Sie können sich nicht für ihr Handeln verantworten, weil sie die Fähigkeit zur Täterschaft ebenso wenig besitzen wie guten Geschmack: Maschinen sind ihrem Wesen nach apolitisch, weil sie weder Ungerechtigkeit erfahren können noch der Gerechtigkeit bedürfen.“

Lisa Hirn

Wenn ein Mensch einem anderen Menschen ein Unrecht zufügt, so ist die Vorgangsweise der Justiz recht simpel: Es wird betrachtet, in welchem Maße das Opfer verletzt wurde, egal ob körperlich, geistig oder finanziell, welche Gründe es für diese Tat gab und wie der geistige Zustand sowohl des Täters als auch des Opfers zur Zeit der Tat war. Natürlich ist dies eine sehr stark vereinfachte Beschreibung dieses Prozesses, und es gibt mehr Fälle der Justiz wo der Vorgang nicht so geradlinig ist, wie er hier beschrieben wird, aber ein Fakt steht immer klar fest: Wenn Menschen in einem Unrecht involviert sind, können, und in vielen Situationen müssen, jene Menschen auch bestraft werden. Für eine lange Zeit war dieser Fakt auch alles, was man für die reine Grundlage der Gerechtigkeit in solchen Unrechtsfällen gebraucht hat, bis eine neue Komponente ins Spiel gekommen ist: die Maschine. Die Maschine ist kein Mensch und wird auch nicht wie ein Mensch behandelt, kann teilweise aber anderen Menschen Schaden zufügen. Wie sollte also eine solche Situation gehandhabt werden? Wie viel Verantwortung kann man einer Maschine zuschreiben? Und basierend darauf, sollte man solche intelligenten, entscheidungsfähigen Maschinen überhaupt kreieren?

Maschine ist nicht gleich Maschine

Eine wichtige Unterscheidung, die man in dieser Problematik treffen muss, bevor man sich tiefer mit ihr befassen kann, ist sicherlich jene zwischen einer simplen Maschine und einer künstlichen Intelligenz. Eine traditionelle Maschine hat bestimmte Funktionen, für jene sie gebaut und programmiert wurde und kann dementsprechend nicht von ihnen abweichen. Solche Maschinen werden beispielsweise in Fabriken verwendet, um die Produktion zu beschleunigen und für den Betrieb Geld zu sparen, da sie mit mehr Maschinen weniger menschliche Arbeitskräfte benötigen. Eine KI ist zwar im Grunde auch eine Maschine, in dem Sinn, dass sie von uns Menschen gebaut und ins „Leben“ gerufen wurde, sie hat allerdings ein bestimmtes Maß an freier Entscheidungskraft. Der Grund, warum unsere Gesellschaft heute mehr und mehr auf die Entwicklung von künstlichen Intelligenzen fokussiert ist, ist einfach: Je mehr Arbeit von einer Maschine gemacht werden kann, desto mehr Geld spart man sich im Bezug auf Löhne für menschliche Arbeiter. Je komplexer diese Arbeiten jedoch werden, desto weiter entwickelt muss die den Menschen ersetzende Maschine sein, weshalb traditionelle Maschinen heute teilweise nicht mehr ausreichen. Sie müssen in der Lage sein, Inhalte einer Tabelle, eines Bildes, eines Videos oder einer rapide ablaufenden Situation eigenständig zu analysieren und entsprechende Entscheidungen zu treffen, wofür es im Endeffekt eine KI benötigt.

Die Macht der Entscheidung

Hier stellt sich nun allerdings die Frage, bis zu welchem Punkt es die Möglichkeit einer solchen Maschine sein sollte, Entscheidungen zu treffen, die sonst ein Mensch mit einem freien Willen und einem Gewissen treffen müsste. Um tiefer auf diese Frage einzugehen, ist ein bekanntes Gedankenexperiment hilfreich: das „Trolley-Problem“. Ein Mensch steht an einem Bahngleis, das sich in zwei unterschiedliche Schienen aufteilt. Auf der einen Schiene ist eine Person angebunden, auf der

anderen sind es fünf Personen. Der Zug, den der Mensch schon in der Ferne kommen sieht, würde die fünf Menschen überfahren und damit töten, allerdings hat der Mensch am Bahngleis die Möglichkeit, die Weichen umzustellen, um nur einen Menschen zu töten. Es gibt zahlreiche Versionen dieses Experiments, mit unterschiedlichen Anzahlen an Personen oder mit der Unterscheidung zwischen Fremden auf dem einen Gleis und einer Person, die man liebt, auf dem anderen. Die zentrale Fragestellung bleibt jedoch immer dieselbe: Würde man direkt den Tod einer Person verursachen, um fünf andere Personen zu retten? Dies ist eine Frage, die jeder für sich selbst beantworten muss, denn es gibt hier keine richtige und keine falsche Antwort.

Wenn man dieses Gedankenexperiment nun etwas umdenkt, um es zu einer für die heutige Zeit realistischeren Situation zu machen, so gelangt man auf einen Punkt der Diskussion, über den sich die Autoindustrie in der kommenden Zeit der selbstfahrenden Autos anfangen muss, Gedanken zu machen: Wie würde sich eine Maschine in einer solchen Situation entscheiden? Entfernen wir uns für dieses Gedankenexperiment nun von den Zuggleisen und wenden wir uns dem Straßenverkehr zu. Ein Mensch fährt in einem selbstfahrenden Auto, welches von einer KI gesteuert wird, und kommt auf einen Zebrastreifen zu, über den gerade eine alte Frau geht. Zufälligerweise haben die Bremsen des Autos aufgegeben, und die KI muss nun entscheiden, ob sie weiter geradeausfährt und damit die alte Frau überfährt, was zu ihrem Tod führen würde, oder ob sie das Auto in die nächstgelegene Wand fährt, damit die alte Frau rettet, den Passagier des Autos jedoch direkt umbringt. Auch hier ist es möglich, die involvierten Personen beliebig anzupassen: Im Auto können vier Personen sitzen statt einer, die alte Frau könnte ein junger Mann sein oder eine Gruppe Kindergartenkinder. Egal, welche Entscheidung die KI hier treffen würde, es würde immer zum öffentlichen Skandal kommen. Im ersteren Fall würden Tausende protestieren, um Fahrzeuge, die wissentlich Passanten verletzen, von der Straße zu bringen, im zweiten Fall würde sich niemand mehr ein Auto kaufen, das den eigenen Tod im Fall eines Problems nicht nur in Kauf nimmt, sondern auch aktiv herbeiführt. Müsste ein Mensch diese Entscheidung treffen, so könnte er vom Gesetz bestraft werden, da er aus freiem und vernünftigem Willen heraus gehandelt hat. Eine Maschine hat diesen freien Willen allerdings nicht, denn sie hat zwar die Möglichkeit, Entscheidungen zu treffen, im Kern ihrer Entscheidungen liegt jedoch noch immer die Programmierung, welche auf einen Menschen zurückzuführen ist. Müsste also der Mensch, der die KI in eine bestimmte Richtung entwickelt und programmiert hat, der also indirekt für die in einem wie oben beschriebenen Unglücksfall getroffenen Entscheidungen verantwortlich ist, für die Maschine haften und an ihrer Stelle bestraft werden? Und wenn dies der Fall wäre, wer würde diese Arbeit dann freiwillig durchführen? Auf die Frage nach der Täterschaft gibt es in diesem Gedankenexperiment gleich wenig eine Antwort wie beim traditionellen „Trolley-Problem“, allerdings müssen die Industrie und all jene, die entscheidungsfähige Maschinen auf die Öffentlichkeit loslassen wollen, etwas finden, was einer Antwort nahekommt, sonst wird das Zeitalter der Maschinen ebenso schnell enden, wie es begonnen haben wird.

Mensch oder Maschine?

Ich möchte mich nun erneut an ein weiteres Gedankenexperiment wagen: Man stelle sich vor, man hätte die Möglichkeit, sein Bewusstsein mit all seinen Erinnerungen, Gefühlen, Meinungen und Gedanken zu duplizieren und in eine Maschine zu übertragen. Diese Maschine wäre nun also wie ein Roboter, ausgestattet mit einer Kopie von unserem „Selbst“. Zunächst würde sich die Frage stellen, ob man diesen Roboter aufgrund seines Bewusstseins nun als Person betrachten könnte, worauf die meisten sicherlich mit „Nein“ antworten würden. Er sei ja nur eine Kopie, er habe nichts von seinen „Erinnerungen“ wirklich selbst erfahren, keine seiner Meinungen seien wirklich die seinigen, so könnte man sich die Argumentation dafür vorstellen.

Dementsprechend würde für diese Maschine nun dasselbe gelten wie für alle anderen künstlichen Intelligenzen: Die Maschine könnte nicht als Person von vollem und gesundem Menschenverstand

angesehen werden, obwohl sie eigentlich genau damit ausgestattet ist, und könnte dementsprechend nicht strafrechtlich verfolgt werden. Also würde sich auch hier wieder die Frage nach der Schuldfähigkeit der Maschine entwickeln: Wer würde für ein Vergehen der Maschine bestraft werden, die Maschine selbst oder die Person, deren Bewusstsein im Endeffekt die Entscheidung für jenes Vergehen getroffen hat? Es stellt sich also eine ähnliche Problematik wie beim selbstfahrenden Auto, da auch im Fall des Roboters mit dem kopierten Bewusstsein eine andere Person indirekt hinter „seinen“ Entscheidungen steht.

Der perfekte Soldat

Ich habe nun schon mehrmals postuliert, dass es nicht möglich ist, eine Maschine für eine Tat zu bestrafen, die sie ausgeführt hat, egal wie viel Schaden diese Tat mit sich gebracht hat. Allerdings komme ich jetzt zum Hauptgrund, warum die Schuldzuweisung an eine Maschine nicht möglich ist: Neben dem Fakt, dass eine Maschine, egal wie intelligent oder selbstständig sie ist, keinen wahren freien Willen hat, hat sie ebenso wenig die Fähigkeit, die eigenen Taten als „gute Taten“ oder „schlechte Taten“ einzuschätzen. Maschinen haben keinen moralischen Kompass, sie haben keine Meinungen und keine Erfahrungen, die eine solche Moral ausbilden könnten. Selbstverständlich können Maschinen nicht mit Menschen gleichgestellt werden und dementsprechend auch nicht mit denselben juristischen Mitteln beurteilt werden, allerdings kann man im Rahmen dieser Argumentation den Willen einer Maschine wegen der vorher benannten Eigenschaften mit dem eines Menschen vergleichen, der nicht die mentalen Kapazitäten besitzt, um seine Entscheidungen wissentlich und bei klarem Verstand zu treffen und zu beurteilen. Auch solche Personen werden strafgerichtlich nicht auf dieselbe Weise und im selben Ausmaß bestraft wie Personen, die bei klarem Verstand sind.

Nun möchte ich tiefer auf die Unfähigkeit einer Maschine eingehen, eigene moralische Vorstellungen und Überzeugungen zu haben, da vor allem dieser Punkt ein enormes Risiko bei der vermehrten Utilisierung von Maschinen darstellen wird. Man stelle sich vor, eine Regierung habe einem anderen Staat den Krieg erklärt und brauche nun Soldaten, um diesen Krieg zu führen. Diese Soldaten sollten dem eigenen Staat treu sein, keine Entscheidungen und Anweisungen ihres Staatsoberhauptes hinterfragen und vor allem alle Befehle durchführen, die sie erhalten, egal ob sie mit ihren eigenen moralischen Grundsätzen vereinbar sind oder nicht. Diese Beschreibung wird nie, in der gesamten Geschichte der Menschheit, vollkommen und ohne Ausnahme auch nur auf einen einzigen Menschen zutreffen. Auf wen diese Beschreibung allerdings perfekt zutrifft, ist eine Maschine: Sie ist durch ihre Fähigkeiten als KI intelligent genug, um alle noch so grässlichen und unmenschlichen Befehle auszuführen und hat „dank“ ihrer totalen Unfähigkeit zu moralischen Überzeugungen auch keine Skrupel, ihnen ohne weitere Fragen zu folgen. In einer Zeit, in der sich die Kriege auf unserer Erde wieder anfangen zu häufen und in der es diverse Staaten gibt, die zur totalen gegenseitigen Zerstörung fähig wären, ist der Gedanke eines „perfekten Soldaten“ sicherlich kein beruhigender.

Die Zukunft der Maschinen-Heit

Wenn man all die oben beschriebenen Argumente bezüglich der Fähigkeit zur Täterschaft von Maschinen betrachtet, sei es nun die Frage nach der Schuld bei selbstfahrenden Autos und Robotern mit einem menschlichen Bewusstsein oder die Möglichkeit einer skrupel- und morallosen Armee im Fall eines Krieges, könnte man sich nun durchaus die Frage stellen, ob es nicht vielleicht besser wäre, solche Maschinen schlicht und einfach nicht zu kreieren. Dadurch würden solche Diskussionen über Schuld und Verantwortung sicherlich vermieden werden, allerdings ergibt es heute kaum mehr Sinn, über eine solche Welt ohne Maschinen mit einem gewissen Maß an Bewusstsein zu sprechen. Dafür ist es leider schon zu spät, denn es gibt mittlerweile schon zu viele Firmen und Betriebe, die all ihre

Ressourcen dafür einsetzen, immer bessere und schnellere künstliche Intelligenzen und die damit verbundenen Maschinen zu programmieren. Diese Entwicklung kann nicht mehr aufgehalten werden, deshalb ist es heute wichtiger denn je, sich über die oben beschriebenen Fragen und Situationen Gedanken zu machen. In einer idealen Welt hätten diese Diskussionen stattgefunden, bevor man mit der Entwicklung dieser Maschinen überhaupt begonnen hat, aber in dieser idealen Welt leben wir leider nicht. Jetzt geht es also darum, das Beste aus der Situation zu machen und alles dafür zu geben, dass es noch vor dem endgültigen Einsatz von wirklich hoch entwickelten, intelligenten und entscheidungsfähigen Maschinen zur Festsetzung von gewissen Richtlinien kommt, welche endgültig die Frage beantworten, wer an der Tat einer Maschine schuld ist: die Maschine selbst oder der Mensch.