

Brauchen Roboter eine Ethik und handeln Menschen moralisch?

31. Oktober 2017, Florian Rötzer

Überlegungen dazu, wie autonome Systeme handeln sollten, um in die Welt der Menschen integriert werden zu können

Ob Maschinen oder Roboter eine Ethik brauchen, um verantwortungsvoll zu handeln, oder ob Regeln schlicht reichen, ist umstritten. Um moralisch handeln zu können, müssten KI-Systeme nach geläufigen philosophischen Überlegungen entweder Selbstbewusstsein und freien Willen besitzen und vernünftig oder emotional in uneindeutigen Situationen eine moralisch richtige Entscheidung treffen. Ob Menschen meist so handeln, sofern sie überhaupt moralisch verantwortlich entscheiden, oder auch nur meinen, sie würden moralisch richtig handeln, was andere mit gewichtigen Argumenten bestreiten können, sei dahingestellt, ganz abgesehen davon, ob es einen freien Willen gibt.

Man kann aber durchaus der Überzeugung sein, dass Menschen in der Regel etwa im Straßenverkehr, wenn schnelle Reaktionen erforderlich sind, instinktiv oder aus dem Bauch heraus entscheiden, also bestenfalls nach einer Moral handeln, die in der Evolution und durch Vererbung und eigene Erfahrung entstanden ist und das Verhalten vor jedem Selbstbewusstsein und freiem Willen determiniert. Ist es daher nicht naiv, wenn von Maschinen wie autonomen Fahrzeugen moralische Entscheidungen gefordert werden und nicht einfach Regeln oder eine "programmierte Ethik" eingebaut werden?

Wenn man besonders umstrittene autonome Kampfroboter betrachtet, die manche deswegen verhindern wollen, weil sie nicht moralisch handeln, kann man sich etwa fragen, ob Soldaten moralisch handeln, wenn sie auf Befehl andere Menschen auch dann töten, wenn sie selbst nicht unmittelbar betroffen sind, wozu sie schließlich trainiert oder abgerichtet wurden? Es gibt allerdings Einsatzregeln und Kriegsgesetze, die sich auch in Kampfroboter einprogrammieren ließen und die sie dann vielleicht genauer befolgen würden als Menschen, die schon allein aus Sicherheits- und Selbsterhaltungsgründen weniger regelgetreu handeln.

Häufig wird so getan, als wäre es eine besonders schreckliche Vorstellung, wenn autonome Systeme wie Kampfroboter aufgrund bestimmter Regeln oder Algorithmen selbst entscheiden, vor allem wenn es um Situationen geht, in denen es keine eindeutigen richtigen Entscheidungen gibt. Hinter den Ängsten steht der Glaube, es sei besser, wenn ein Mensch, der moralisch handeln können soll, die Entscheidungen trifft.

Man fragt sich angesichts all der Grausamkeiten, Zerstörungen, Egoismen und Profitgier, warum man von vorneherein glaubt, dass eine maschinelle programmierte Ethik schlechter oder furchtbarer sein muss, zumal ein Roboter schneller und "kühl" entscheiden kann. Ist es besser, von einem Menschen getötet oder verletzt zu

werden, weil er emotional handeln kann und sein Leben retten will oder wie ein Selbstmordattentäter aufs Spiel setzt, als von einer wie auch immer regelgeleiteten Maschine?

Leitbild: "Kraftfahrzeuge, die unter allen Umständen einen Unfall vermeiden können"

Interessant ist in diesem Kontext der Bericht der vom Bundesminister für Verkehr und digitale Infrastruktur eingesetzte Ethik-Kommission über "Automatisiertes und Vernetztes Fahren". Die Absicht war klar, das Ministerium wollte die Einführung von autonomen Straßenfahrzeugen beschleunigen und Barrieren abbauen. Schon im Vorwort heißt es, man könne nicht ausschließen, "dass am Ende der Entwicklung Kraftfahrzeuge stehen, die inhärent sicher sind, also unter allen Umständen einen Unfall vermeiden können".

Das muss man als interessengeleitete Fiktion bezeichnen, schließlich kann wohl niemals ausgeschlossen werden, dass einem technischen System unter allen Umständen keine Panne oder Störung unterlaufen wird, wodurch aus internen oder externen Ursachen ein Unfall verursacht werden könnte. Zur Zeit, so heißt es weiter, werde aber "eine vollständige Unfallvermeidung nicht möglich" sein. Das bringt die Ethik ins Spiel, also ob autonome Systeme überhaupt zugelassen werden sollen bzw. ob sie ethischen Regeln bei der Entscheidung über das Fahrverhalten folgen müssen.

Zugelassen werden können autonome Systeme nur, so die Kommission, wenn Risiken für Menschen durch diese vermindert werden. Das Prinzip müsse sein, dass die "freien Entfaltung und der Schutz des Menschen" gewahrt bleibe, wobei die Menschen ihre "eigenverantwortliche Entscheidung" an KI-Systeme delegieren können. Gefordert wird, dass die Technik nach ihrem jeweiligen Stand so ausgelegt sein müsse, "dass kritische Situationen gar nicht erst entstehen, dazu gehören auch Dilemma-Situationen, also eine Lage, in der ein automatisiertes Fahrzeug vor der 'Entscheidung' steht, eines von zwei nicht abwägbaren Übeln notwendig verwirklichen zu müssen".

Das ist schnell gesagt, unterläuft allerdings das Problem, denn automatisierte Fahrzeuge können nicht selbst Dilemma-Situationen ausschließen. Sie dürften eigentlich gar nicht erst vorkommen. Aber wie soll das geschehen?

Militärisch wäre dies vergleichbar mit einem Präzisionsgriff, der Kollateralschäden ausschließt. Voraussetzung wäre, dass es in der Welt keine diffusen Zonen gibt, die eine Abwägung erzwingen. Es dürfen keine Zivilisten in der Nähe von Gegnern sein, es darf nicht die Situation entstehen, dass ein Fahrzeug nur die Alternative hat, entweder die eine oder die andere Person zu gefährden, wozu auch noch kommt, wie hoch der Schutz des eigenen Lebens bzw. aus der Sicht des autonomen Fahrzeugs: das des Passagiers zu gelten hat.

Die Erwägung trifft jeder Mensch irgendwie und spontan in Unfallsituationen, die ein ausgiebiges Reflektieren nicht erlauben. Aber dem Menschen überlässt man letztlich die Eigenverantwortung, um ihn dann womöglich, wie auch immer abgestuft, zur Rechenschaft zu ziehen. Aber kann man dies bei KI-Systemen? Wie und zu

welchem Zweck könnten diese nachträglich zur Rechenschaft gezogen werden? Wäre bei Fehlentscheidungen der Hersteller verantwortlich? Und wie ist das beim Menschen? Ist Gott oder die Natur verantwortlich?

Das ethische Dilemma mit dem Dilemma

Die Kommission macht dann doch deutlich, dass es ein Dilemma mit dem Dilemma gibt: "Echte dilemmatische Entscheidungen, wie die Entscheidung Leben gegen Leben sind von der konkreten tatsächlichen Situation unter Einschluss 'unberechenbarer' Verhaltensweisen Betroffener abhängig. Sie sind deshalb nicht eindeutig normierbar und auch nicht ethisch zweifelsfrei programmierbar."

Gleichzeitig wird gefordert: "In Notsituationen muss das Fahrzeug autonom, d.h. ohne menschliche Unterstützung, in einen 'sicheren Zustand' gelangen."

Was aber als Prinzip festgelegt werden soll, offenbart die Unmöglichkeit, dies lösen zu können: "Bei unausweichlichen Unfallsituationen ist jede Qualifizierung nach persönlichen Merkmalen (Alter, Geschlecht, körperliche oder geistige Konstitution) strikt untersagt. Eine Aufrechnung von Opfern ist untersagt. Eine allgemeine Programmierung auf eine Minderung der Zahl von Personenschäden kann vertretbar sein. Die an der Erzeugung von Mobilitätsrisiken Beteiligten dürfen Unbeteiligte nicht opfern." Man scheint sich also geeinigt zu haben, dass man etwa Alte gegen Junge nicht ausspielen darf, wohl aber Mehr gegen Wenige. Statt Qualität zieht man sich auf die Quantität zurück, die offenbar als weniger moralisch anstößig gilt.

Würde man von Menschen absehen, die nicht mit autonomen Vehikeln unterwegs sind, könnte man Unfallfreiheit und Abwesenheit von Dilemmata durch eine zentrale Steuerung realisieren, so lange das Steuerungssystem fehler- und pannenfrei arbeitet. Aber dem würde die Freiheit der Menschen widersprechen: "Eine vollständige Vernetzung und zentrale Steuerung sämtlicher Fahrzeuge im Kontext einer digitalen Verkehrsinfrastruktur ist ethisch bedenklich, wenn und soweit sie Risiken einer totalen Überwachung der Verkehrsteilnehmer und der Manipulation der Fahrzeugsteuerung nicht sicher auszuschließen vermag."

Das aber kann man nicht, wie die Kommission selbst feststellt: "Funktionsstörungen vernetzter Systeme oder Angriffe von außen lassen sich nicht vollständig ausschließen. Solche Risiken sind, wenngleich selten vorkommend, komplexen Systemen immanent." Also bleiben autonome Systeme inhärent riskant. Zudem wird seltsam argumentiert. Entweder man will Sicherheit, dann geht dies auf Kosten der Freiheit, oder man bewahrt Freiheit, dann lassen sich Risiken nicht ausschließen. Die treten sowieso auf, weil man gelingende Cyberangriffe auf autonome Fahrzeuge nicht ausschließen kann. Bislang gibt es kein absolut sicheres vernetztes System, gäbe es dieses, wäre es wiederum ein Sicherheitsrisiko. Und wenn die Verkehrsteilnehmer Daten aufgrund der "Autonomie und Datenhoheit" nicht weitergeben müssen, ist eine zentrale Steuerung schlicht unmöglich.

Die Kommission bleibt auch unentschieden, ob einem autonomen Fahrzeug in kritischen Situationen die Entscheidung überlassen werden soll. Als Szenario wird gewählt: "Der Fahrer eines Wagens fährt eine Straße am Hang entlang. Der vollautomatisierte Wagen erkennt, dass auf der Straße mehrere Kinder spielen. Ein eigenverantwortlicher Fahrer hätte jetzt die Wahl, sich selber das Leben zu nehmen,

indem er über die Klippe fährt oder den Tod der Kinder in Kauf zu nehmen, indem er auf die im Straßenraum spielenden Kinder zusteuert. Bei einem vollautomatisierten Auto müsste der Programmierer oder die selbstlernende Maschine entscheiden, wie diese Situation geregelt werden soll."

Vergessen wird aber bei der Überlegung das Szenario, das im öffentlichen Verkehr oder im Flugzeug normal ist, wo die Menschen Passagiere und der Entscheidung der Fahrer bzw. Piloten überlassen sind, also auch nicht eigenverantwortlich entscheiden können. Kaum bestreitbar ist, dass die Kommission empfiehlt, die Vermeidung von Personenschäden müsse Vorrang haben vor Sachschäden. Verletzung sei überdies einer Tötung vorzuziehen, und es müsse das Prinzip der geringstmöglichen Zahl von Verletzten und Getöteten gelten. Aber all das muss im vorneherein mit blitzschnellen Wahrscheinlichkeitsberechnungen entschieden werden, die selbstverständlich falsch sein können, weil in der Welt immer Unerwartetes, Singuläres geschehen kann.

Theoretisch hat die Forderung nach einer "Minimierung der Opfer" durch eine Programmierung etwa gegenüber der Regelung im Luftsicherheitsgesetz, dass die Opferung von unschuldigen Menschen zugunsten anderer Menschen unzulässig ist, und dem Weichenstellerdilemma, gewissermaßen einen Vorteil. In realen Verkehrssituationen müsse eine Wahrscheinlichkeitsprognose aus der Situation erfolgen, "bei der die Identität der Verletzten oder Getöteten (im Gegensatz zu den Trolley-Fällen) noch nicht feststeht. Eine Programmierung auf die Minimierung der Opfer (Sachschäden vor Personenschäden, Verletzung von Personen vor Tötung, geringstmögliche Zahl von Verletzten oder Getöteten) könnte insoweit jedenfalls ohne Verstoß gegen Art. 1 Abs. 1 GG gerechtfertigt werden, wenn die Programmierung das Risiko eines jeden einzelnen Verkehrsteilnehmers in gleichem Maße reduziert."

Man würde also nicht Menschenleben gegeneinander verrechnen, wie das bei dem Trolley-Dilemma gemacht wird, sondern die Programmierung müsse so erfolgen, dass das Risiko für alle Verkehrsteilnehmer, den Fahrer oder die Passagiere eingeschlossen, gleichermaßen minimiert wird, was auch heißen würde, dass es für Manche erhöht werden müsste, die faktisch bei irgendeinem Verhalten gefährdet sein könnten. Moralisch wären nach dem interessanten Ansatz die Gleichverteilung des Risikos und ein zufälliges Ergebnis. Allerdings muss trotzdem immer eine Entscheidung getroffen werden. Dazu kommt, dass nicht nur die jeweils eingesetzte KI, sondern eine Vielzahl von technischen Systemen unterschiedlicher Hersteller funktionieren muss. Eine Verantwortung tragen auch, wie der Bericht anmerkt, die Hersteller der Sensoren.

Die Folgerung der Kommission für die Hersteller setzt dabei die Schwelle tief und entspricht einem utilitaristischen Prinzip, wobei genau die entscheidende Abwägung offen bleibt, wo die Grenzen des Zumutbaren für die Sicherheit und die Minimierung der Opfer liegt:

Folgt man der hier vertretenen Position, so ergibt sich das Folgeproblem, ob bzw. inwieweit Hersteller für die als "Unrecht" qualifizierte Verletzung oder gar Tötung durch automatisierte Systeme zur Verantwortung gezogen werden können. Es ist deshalb darauf hinzuweisen, dass für Kollisionsvermeidungssysteme im Prinzip nichts

anderes gilt als für Airbags oder Gurte: Die Tötung durch einen fehlauslösenden Airbag bleibt Unrecht, der Hersteller wird aber nicht in Haftung genommen, wenn er alles Zumutbare unternommen hat, um derartige Risiken zu minimieren.

Der Einbau von automatisierten Systemen ist daher zulässig und führt nicht zu besonderen Haftungsrisiken, wenn die Hersteller alles Zumutbare tun, um ihre Systeme so sicher wie möglich zu machen und insbesondere die Gefahr von Personenschäden minimieren.