

Ethische Aspekte des autonomen Fahrens

Von Martin Randelhoff, Münster (Westf.) am 25.08.2016

Wieso benötigen wir eine Debatte?

In der öffentlichen Diskussion um hoch- und vollautomatisiertes Fahren steht neben der technischen Machbarkeit auch eine ethisch-moralische Dimension im Fokus.

Im Rahmen diverser Testfahrten sind bereits heute zahlreiche Fahrzeuge¹ im automatisierten Fahrmodus auch im deutschen Straßennetz unterwegs. Diese Fahrten dienen bspw. der Erprobung von verschiedenen Sensoren, der Verarbeitung der Sensordaten mittels Algorithmen und der Prüfung von situationsabhängigen Entscheidungen des Fahrzeugrechners.

Vertreterinnen und Vertreter der Automobilindustrie gehen davon aus, dass voll- und hochautomatisierte Fahrzeuge spätestens in den frühen 2020er Jahren im Handel zu erwerben sein werden. Bereits heute sind bereits diverse Teilautomatisierungsfunktionen für den normalen Autokäufer erhältlich, angefangen über Einparkassistenten, Spurhalter, Abstandshalter, etc.

Die Entwicklung hoch- und vollautomatisierter Fahrzeuge stellt nochmals höhere Anforderungen an die Entwicklung und Implementierung entsprechender Algorithmen, welche aus den gesammelten Sensordaten Entscheidungen bezüglich Geschwindigkeit, Beschleunigung / Abbremsen und Fahrrichtung eines Fahrzeugs ableiten. Mit Blick auf die heute eingesetzte und auch in naher Zukunft zur Verfügung stehenden Technik sind die Basis für diese Entscheidungen bestimmte Kriterien, welche vorab in den Fahrzeugrechner implementiert werden.

Automatisierte Fahrzeuge werden mittels Sensoren und Software Fahrzeugbewegungen und ihr Umfeld in Sekundenbruchteilen erfassen und im Falle einer (möglichen) Kollision Ausweichmanöver einleiten. Falls ein Zusammenprall unausweichlich ist, werden sie den

¹ Der Begriff „Fahrzeug“ bezieht sich in diesem Pkw auf Pkw, Lkw, Busse, Motorräder, etc. Ebenfalls synonym sind die Begriffe Roboterfahrzeug, automatisiert, autonom, u.ä. zu verstehen.

optimalen Einschlagswinkel und die optimale Einschlagsgeschwindigkeit weitaus besser als ein Mensch berechnen können. Durch entsprechende Fahrmanöver wird das Fahrzeug in die optimale Position gebracht, um die Fahrzeuginsassen bestmöglich zu schützen.

Hierfür ist es jedoch notwendig, dass im Vorfeld festgelegt wird, wie sich ein Fahrzeug in einer bestimmten Situation verhalten soll. Im Gegensatz zu einem menschlichen Fahrzeugführer, der spontan auf die Situation reagieren kann, müssen einem Fahrzeug vorab alle möglichen Situationen und die entsprechenden Reaktionen mitgeteilt werden. Ein automatisiertes Fahrzeug kann Sensordaten interpretieren und eine Entscheidung treffen, aber die Entscheidung selbst ist ein Ergebnis einer Logik, welche Monate oder Jahre im Voraus implementiert wurde.

Doch wie lauten die dieser Logik zugrundeliegenden Kriterien und wer legt diese fest? Die Automobilindustrie, da es sich um ihr Produkt handelt? Entscheidet eine demokratisch legitimierte Regierung für die Bevölkerung? Hat der Fahrzeughalter die Möglichkeit, den moralischen Kompass seines Fahrzeugs an seine persönlichen ethischen Vorstellungen anzupassen und übernimmt dafür auch Verantwortung mit den damit verbundenen Konsequenzen? Welchen Stellenwert haben Reflexion, Empathie und Gewissen in diesem Prozess? Ist es überhaupt vertretbar zu definieren, dass bspw. zwei Menschen weniger wert seien als fünf, oder jüngere Menschen mehr wert seien als ältere? Ist eine von einer virtuellen Intelligenz korrekt bewertete Situation mit dem bestmöglichen Ergebnis auch richtig?

Zurzeit werden in der Automobilindustrie entsprechende Algorithmen entwickelt und programmiert. Hierbei wird ein rationaler Ansatz mit festen Entscheidungskriterien präferiert, da diese einem Computer einprogrammiert werden können. Gleichzeitig hat dieses Vorgehen den Nachteil, dass irgendein menschliches Wesen mit absoluter moralischer Autorität im Vorfeld festlegen muss, welche Reaktion in einer spezifischen Situation richtig oder falsch ist. In der maximalen Ausprägung verbunden mit der Frage: wer darf leben, wer muss sterben?

Diese rationale Vorgehensweise äußert sich entweder im deontologischen Prinzip, in welchem das automatisierte System einer Reihe von Regeln gehorchen muss, oder dem konsequentialistischen Prinzip, bei dem ein gewisser Nutzen maximiert werden soll.

Hintergrund für diese Überlegungen ist das 1967 von der britischen Philosophin Philippa Foot in den Diskurs eingebrachte „Trolley-Problem“: Eine Straßenbahn ist außer Kontrolle geraten und droht, fünf Personen zu überrollen. Durch Umstellen einer Weiche kann die Straßenbahn auf ein anderes Gleis umgeleitet werden. Unglücklicherweise befindet sich dort eine weitere Person. Darf (durch Umlegen der Weiche) der Tod einer Person in Kauf genommen werden, um das Leben von fünf Personen zu retten? Das „Trolley-Problem“ kann mit dem „Fette-Mann-Problem“ und diversen weiteren Abwandlungen des Problems² ergänzt werden, wobei die Kern-Problemstellung identisch bleibt.

Das deontologische Prinzip stellt die Tat an sich weitestgehend unabhängig von den Konsequenzen und der Motivationslage in den Mittelpunkt. Entscheidend ist dabei, ob die Handlung einer verpflichtenden Regel gemäß. Deontologische Normen, in ihrer simpelsten Form, haben die Gestalt von direkten Handlungsregeln wie „Lügen ist falsch“ oder „Hilfeleistung ist geboten“ – gleichgültig aus welchem Antrieb oder mit welchen Folgen. Auf das Trolley-Problem bezogen, gilt das deontologische Verbot zu töten auch in einem Fall, in dem eine Person durch das Töten eines Unschuldigen die Tötung fünf Unschuldiger verhindern könnte. Es geht also darum, nicht selbst zu töten, und nicht darum, möglichst wenig Tötungen herbeizuführen.

Dem gegenüber steht das konsequentialistische Prinzip. Hierbei steht das Ergebnis des Handelns im Vordergrund. Der einzig moralisch relevante Faktor für die Richtigkeit oder Falschheit einer Handlung ist die Güte der Handlungskonsequenzen. Eine Handlung ist moralisch richtig genau dann, wenn ihre Konsequenzen mindestens so gut sind, wie die Konsequenzen jeder anderen Handlung, die man an ihrer Stelle ausführen könnte.

² Thomson, Judith (1985): The Trolley Problem. In: The Yale Law Journal, Vol. 94, Nr. 6 (Mai 1985), S. 1395-1415

Dem Konsequentialismus folgend würde man durch Umstellen der Weiche die fünf Leben auf Kosten des einen retten, da in der Summe weniger schlechte Konsequenzen auftreten. Der Gesamtnutzen aller ethischen Subjekte soll durch die Konsequenzen einer Handlung größtmöglich sein. Vielmehr ist sogar jede Handlung, deren Konsequenzen schlechter sind als diejenigen einer anderen Handlung moralisch falsch, d. h. verboten. Daraus folgt, dass nur diejenige Handlung mit den besten Konsequenzen erlaubt und geboten ist.

Die Anwendung des konsequentialistischen Prinzips im Bereich des automatisierten Fahrens bedeutet eine eindeutige Entscheidung, wenn eine Wahl besteht zwischen einem Zusammenstoß oder keinem Zusammenstoß, zwischen einer schweren Verletzung und Sachschaden oder zwischen zwei Sachschäden in eindeutiger Höhe. Eine Entscheidung, die ein Computer bei entsprechender Programmierung zu treffen vermag.

Jedoch ergeben sich auch bei dieser Vorgehensweise diverse Fragen: Wie werden Leid oder ein Schaden quantifiziert? Bei einer Bewertung in Geldeinheiten wäre die Minimierung des Schadens ein mögliches Ziel. Jedoch gehen auch damit einige Probleme einher:

So würde ein automatisiertes Fahrzeug im Falle eines Kontrollverlusts jenes Fahrzeug wählen, welches den geringeren Fahrzeugwert besitzt. Oder es wählt jenes Fahrzeug, welches den höheren Sicherheitsstandard erfüllt. Oder es zieht den behelmtten Motorradfahrer dem unbehelmtten Motorradfahrer vor. Spätestens an diesem Punkt zeigt sich, dass ein derartiges Vorgehen problembehaftet ist: Die Investition in einen höheren Schutz ginge mit einem größeren Risiko einher, die Internalisierung von möglichen externen Unfallkosten würde gar mit einer Strafe belegt werden.

Möglich, wenn auch verstörend, wäre eine Bewertung der Insassen potenzieller Unfallgegner. Mittels Gesichtserkennung könnten Alter, Geschlecht und Fahrtüchtigkeit detektiert und auf Basis von Wahrscheinlichkeiten bewertet werden: eine gleichaltrige Frau hat im Gegensatz zu

einem Mann eine 28 % höhere Wahrscheinlichkeit bei einem Unfall ums Leben zu kommen.³ Ein 70-jähriger männlicher Autofahrer weist im Gegensatz zu einem 20-jährigen Autofahrer eine drei Mal so hohe Wahrscheinlichkeit auf, im Falle eines Unfalls zu sterben⁴. Betrunkene Autofahrer sterben doppelt so häufig wie nüchterne, da Alkohol die Körperorgane negativ beeinflusst. Ist ein Fahrzeug mit zwei Personen besetzt, liegt die Wahrscheinlichkeit eines Todesfalls um 14% niedriger, da das Fahrzeuggewicht größer ist als bei einem einzelnen Fahrzeuginsassen⁵. All diese Faktoren könnten bei einem rein nach dem konsequentialistischen Prinzip programmierten Fahrzeug mit einbezogen werden.

Es scheint jedoch kaum denkbar zu sein, einen derartigen Bewertungsmaßstab anzuwenden. Den Wert eines Lebens gegenüber einem anderen als schützenswert(er) einzustufen, widerspricht zudem dem deutschen Rechtsverständnis.

Prof. Eric Hilgendorf von der Würzburger Forschungsstelle für „RobotRecht“ sieht zwar in diesem Bereich eine Verschiebung vom deutschen Rechtsprinzip, welches eher dem deontologischen Prinzip folgend die Tat in den Mittelpunkt stellt und jede Gewichtung von Menschenleben verbietet, hin zu einer quantifizierenden Rechtsethik nach amerikanischem Modell, jedoch liegt der Grund vor allem in einer besseren Implementierbarkeit und weniger in einer Verschiebung ethischer Maßstäbe⁶. Es kristallisiert sich vielmehr in Ansätzen heraus, dass es keine global gültige Moral geben kann und wird. Hierfür sind die kulturellen Hintergründe, die gesellschaftlichen Prägungen und auch das Rechtsverständnis zu unterschiedlich. Eine

³ Evans, L. (2008): Death in Traffic: Why Are the Ethical Issues Ignored? Studies in Ethics, Law, and Technology, Vol. 2, No. 1, 2008.

⁴ Evans, L. (2004): Traffic Safety. Science Serving Society, Bloomfield, Michigan, 2004.

⁵ Evans, L. (2001): Causal Influence of Car Mass and Size on Driver Fatality Risk. American Journal of Public Health, Vol. 91, Nr. 7, 2001, S. 1076–1081

⁶ Thiel, Thomas (2015): Fahren nach Zahlen. Online abrufbar unter: www.faz.net/aktuell/feuilleton/geisteswissenschaften/ethische-prinzipien-fuer-autonome-autos-13722800-p2.html?printPagedArticle=true#pageIndex_2, zuletzt abgerufen am 22.08.2016

Übernahme der angelsächsischen Rechtsauffassung in Deutschland rein aus dem Grund, dass eine weitestgehende Autonomie von Maschinen möglich ist, scheint aus heutiger Perspektive mehr als zweifelhaft zu sein.

Ein deutsches Gesetz, welches es autonomen Fahrzeugen ermöglicht, den Wert von Menschenleben gegeneinander aufzurechnen, ist mit Artikel 1 des Grundgesetzes und der darin verankerten unantastbaren Würde des Menschen unvereinbar. Die Diskussion um das Luftsicherheitsgesetz, welches den Abschuss eines Passagierflugzeugs zum Schutz Dritter legitimierte, weist entsprechende Parallelen auf und verdeutlicht noch einmal die Nichtaufrechenbarkeit des Lebens⁷.

Gleichsam stellt auch dies wiederum ein Dilemma dar. Wenn eine Entscheidung zwischen zwei Leben nicht möglich ist, müssten entweder beide sterben oder niemand. In einer entsprechenden Unfallsituation ist jedoch die Physik Maß aller Dinge – ein Nullfall also eventuell gar nicht möglich. Ebenso stellt eine Entscheidung zum Nicht-Handeln eine eigene Form der Entscheidung dar, die ebenfalls getroffen werden muss. Eine paradoxe Situation.

Kann das Dilemma gelöst werden?

Der Philosoph Dr. Patrick Lin von der California Polytechnic State University plädiert daher dafür, dem Zufall einen gewissen Raum einzuräumen⁸. Im Leben eines Menschen spielt der Zufall eine große Rolle – im Guten wie im Schlechten. Ein Dilemma bei der Entscheidungsfindung könnte dadurch umgangen werden, dass keine endgültige Entscheidung getroffen wird. Vielmehr sollte die nächste Aktion des Fahrzeugs durch einen Zufallsgenerator bestimmt werden: bei einer ungeraden Zahl wählt es Möglichkeit 1 (z. B. ein Ausweichen nach links), bei einer geraden Zahl Möglichkeit 2 (z. B. ein Ausweichen nach rechts). Eine derartige Entscheidung wäre weder richtig

⁷ siehe BVerfG, 1 BvR 357/05 vom 15.02.2006, Rn. (1-156)

⁸ Lin, Patrick (2014): The Robot Car of Tomorrow May Just Be Programmed to Hit You. Wired. Online im Internet: www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you. Zuletzt abgerufen am 22.08.2016

noch falsch. Auch aus dem Grund, da es mangels fehlendem bewertbaren Ergebnis kein richtig oder falsch geben kann. Problematisch ist jedoch, dass in diesem Moment auch keine Verantwortung entstehen kann. Denn für den Zufall kann man niemanden verantwortlich machen.

Lin weist an dieser Stelle zurecht darauf hin, dass bei ethischen Entscheidungen der Weg zum Ergebnis genauso wichtig ist wie das Ergebnis an sich. Der Gedankengang hinter einer Entscheidung hilft zum einen dabei, diese argumentativ zu verteidigen und stellt zum anderen den moralischen Kompass eines Individuums dar und setzt diesen in einen gesellschaftlichen Kontext. Bei einer Zufallsentscheidung entfällt dies. Statt durchdachter Entscheidungen sind sie gedankenlos, und dies kann schlimmer sein, als reflexive menschliche Urteile, die zu schlechten Ergebnissen führen, so Lin.

Da das moralische Dilemma keine abschließend richtige Lösung erwarten lässt, könnte eine Zufallsentscheidung in dieser Situation trotz ihrer Schwächen gerechtfertigt sein. Denn auch der Mensch vermag es nicht, absolut richtige Entscheidungen zu treffen.

Wird ein automatisiertes Fahrzeug überhaupt in Unfälle verwickelt sein?

Die Technologie des automatisierten Fahrens ist eng mit der Hoffnung auf einen signifikant sichereren Straßenverkehr verbunden. Durch die Kombination von Sensordatenfusion und leistungsfähigen Fahrzeugrechnern sollen in Sekundenbruchteilen komplexe Situationen erfasst und optimale Entscheidungen getroffen werden. So kündigt Toyota an, mithilfe künstlicher Intelligenz selbstfahrende Autos entwickeln zu wollen, „die unfähig sind, einen Unfall zu verursachen“. Automatisierte Fahrzeuge besäßen so etwas wie eine technische Unfehlbarkeit. Die Komplexität der Welt, die Unmöglichkeit absolute Sicherheit zu erreichen, mögliche Eingriffe von außen und die Möglichkeit des technischen Versagens einzelner Komponenten oder Komponentengruppen machen dieses Unterfangen jedoch unmöglich. Automatisierte

Fahrzeuge werden in Unfälle verwickelt sein und in entsprechenden Situationen Entscheidungen treffen müssen.

Bislang fehlt zudem der empirische Beleg, dass automatisierte Fahrzeuge einen signifikanten Sicherheitsvorteil gegenüber vom Menschen gesteuerten Fahrzeugen aufweisen. Laut Smith⁹ muss ein automatisierter Pkw mindestens 1,2 Millionen Kilometer unfallfrei unter normalen Bedingungen zurückgelegt haben [Poissonverteilung, Signifikanzniveau von 0,01], um als so sicher wie vom Menschen gefahrene Pkw zu gelten. Für den Beweis, dass automatisierte Pkw die Zahl der Verkehrstoten senken, steigt die notwendige Fahrleistung auf 480 Millionen Kilometer.

Schwere automatisierte Lkw müssen eine Fahrleistung von 4,2 Millionen km bzw. 390 Millionen km aufweisen, um ein gleiches Sicherheitsniveau wie vom Menschen gesteuerte Lkw zu bieten. Bislang konnte noch kein automatisiert fahrender Lkw ein höheres Sicherheitsniveau empirisch belegen.

In einem dynamischen System wie dem Straßenverkehr treten sowohl fixe wie auch sich bewegende Kollisionsobjekte auf. Während Vorhersagen bezüglich fixen Objekten vergleichsweise einfach erscheinen, steigt die Entscheidungskomplexität bei sich bewegenden Objekten stark an¹⁰. Die Vorhersage menschlicher Aktionen im Straßenverkehr ist schwierig. Hinzu kommt, dass Bewegungen und Fahrmanöver ausreichend früh in diesem dynamischen Umfeld erkannt werden müssen, sodass Gegenmaßnahmen und Fahrmanöver zur Kollisionsvermeidung noch möglich sind. Wissenschaftler kamen daher zu dem Schluss, dass auch perfekte Sensoren und Algorithmen keine Garantie für absolute und vollumfassende

⁹ Smith, B. W. (2012): Driving at Perfection. The Center for Internet and Society at Stanford Law School, März 2012

¹⁰ Laugier, C., und R. Chatila (Hrsg.): Autonomous Navigation in Dynamic Environments. Springer-Verlag. Berlin. 2007

Sicherheit bieten können¹¹ – entsprechende Aussagen von Ingenieuren, Entwicklern und Unternehmen sind mit Vorsicht zu bewerten. Stattdessen wird versucht, Wahrscheinlichkeitsmodelle bezüglich des erwarteten Verhaltens menschlicher Autofahrer zu entwickeln, um Situationen mit einem mittleren oder hohen Risiko zu umgehen. Weichen menschliche Fahrer von dem erwarteten Verhalten ab, können sich Unfälle ereignen.

Vonseiten der Industrie wird oftmals kommuniziert, dass automatisierte Fahrzeuge in kritischen Situationen entweder die Fahrstrategie „Bremsen und Ausweichen“ anwenden oder die Kontrolle an den menschlichen Fahrer zurück übertragen. Auf diese Weise sollen Entscheidungen mit ethischem Belang vermieden werden. Jedoch weisen beide Strategien Schwächen auf.

Eine Vollbremsung ist nicht immer die sicherste Option. Vielmehr hängt der Erfolg des Fahrmanövers von mehreren Faktoren ab: In welchem Abstand folgen Fahrzeuge? Ist der Abstand nach hinten bei einer Vollbremsung ausreichend? Wie ist das Fahrzeug beladen? Wie ist der Straßenzustand? Theoretisch sollte der Fahrzeugrechner in der Lage sein, die gefahrene Geschwindigkeit dem Umfeld anzupassen und immer die Option einer Vollbremsung mit positivem Ausgang zu haben. In der Praxis gibt es jedoch ein Problem: andere Fahrzeuge mit menschlichen Fahrern. Diese sind den Umgang mit Fahrzeugen, welche sich strikt an Regeln halten, nicht gewohnt und fahren aggressiver und mit weniger Aufmerksamkeit, als die Situation es eigentlich erfordern würde. Insbesondere das Einhalten des richtigen Abstands ist oftmals problematisch. Jedoch ist dies eine Notwendigkeit für eine unfallfreie Vollbremsung.

Die Übergabe der Fahrzeugkontrolle vom Bordrechner an den Menschen scheitert ebenfalls an menschlichen Unzulänglichkeiten. Je größer die automatisierten Fahranteile sind, desto weniger Fahrpraxis steht zur Verfügung. Bei komplexen und dynamischen Entscheidungen kann somit die

¹¹ Benenson, R., T. Fraichard und M. Parent: Achievable Safety of Driverless Ground Vehicles. Proc., 10th International Conference on Control, Automation, Robotics and Vision, Hong Kong, 2008, S. 515 – 521.

notwendige Erfahrung zur korrekten Erfassung und Bewertung der Situation fehlen.

Insbesondere Fahranfänger sind überproportional betroffen.

Erste wissenschaftliche Untersuchungen haben gezeigt, dass die Aufmerksamkeitsspanne in teilautomatisierten Fahrzeugen (Level 2) nachließ. So drehten sich Textpersonen in der Versuchsgruppe häufiger nach hinten um und beschäftigten sich mit anderen Aktivitäten wie bspw. lesen¹². Fahrzeugführer haben in Fahrzeugen mit Automatisierungslevel 2 die Aufgabe, stets und ohne vorherigen Hinweis die Kontrolle über das Fahrzeug übernehmen zu können. In Automatisierungsstufe 3 steht ihnen ein gewisser Zeitraum für den Übergang zur Verfügung.

Der Zeitraum zwischen Benachrichtigung und Übernahme der Fahrzeugkontrolle ist bislang nicht spezifiziert. Die American Association of State Highway and Transportation Officials (AASHTO) empfiehlt bei 100 km/h eine Minimumdistanz zwischen 200 und 400 Metern, um die Umgebung wahrnehmen, Situationen erkennen und Entscheidungen treffen zu können¹³. Dies entspricht einem Zeitraum von 7,2 – 14,5 Sekunden. Bei zwei sich aufeinander zu bewegenden Objekten, wie bspw. zwei Fahrzeugen, wächst die empfohlene Distanz auf 800 Meter. Der Zeitraum für eine Reaktion beträgt somit etwa 30 Sekunden. Jedoch ist die Reichweite aller sensorbasierten Systeme auf wenige Hundert Meter begrenzt: Velodyne HDL-64E LiDAR: 120 Meter, Kamera bis zu 500 Meter, Fernradar bis zu 250 Meter.

Es stellt sich daher die Frage, ob eine Übertragung der Fahrzeugkontrolle vom Fahrzeugrechner an den Menschen generell zu empfehlen ist. Je nach Situation (Tageszeit, Wetter, Ablenkungsgrad) verlängern sich Wahrnehmungs- und Entscheidungszeiträume. In Gefahrensituationen erscheint es daher sinnvoll, dass der Fahrzeugrechner eigenständig

¹² Llaneras, R.E., J.A. Salinger und C.A. Green (2013): Human Factors Issues Associated with Limited Ability Autonomous Driving Systems: Drivers' Allocation of Visual Attention to the Forward Roadway. Proc., 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, N.Y., 2013

¹³ AASHTO (2011): A Policy on Geometric Design of Highways and Streets, 6th ed. Washington, D.C., 2011.

zeitkritische Entscheidungen trifft – wiederum verbunden mit einer ethisch-moralischen Problemstellung. Und der Notwendigkeit für diese Entscheidung Verantwortung übernehmen zu können.

Muss ein Fahrzeug überhaupt moralisch handeln?

Fahrzeug und Moral – passt dies zusammen? Muss ein automatisiertes Fahrzeug überhaupt in der Lage komplexe ethische Situationen aufzulösen, die äußerst selten auftreten und zudem sehr konstruiert sind? Wieso soll ein Fahrzeug moralisch richtige Entscheidungen treffen (müssen), wenn auch Menschen dies nicht schaffen? Was ist „richtig“ und „falsch“? Wie soll ein Fahrzeug den Ausgang einer Situation und die Folgen einer spezifischen Entscheidung vorab beurteilen können? Der Mensch kann dies auch nicht.

Ist es notwendig, dass Fahrzeuge oder Maschinen moralisch richtige Entscheidungen treffen oder ist es nicht ausreichend, wenn keine moralisch falschen Entscheidungen getroffen werden? Sollte der Zufall eine größere Rolle spielen oder nicht?

Auf all diese Fragen gibt es keine abschließende Antwort.

Sie sollten aber auch nicht als Begründung genutzt werden, dass diese Fragen per se nicht zu beantworten wären. Bei ethischen Fragen handelt es in den allermeisten Fällen um abstrakte Gedankenspiele, die nicht abschließend beantwortet werden können. Dies ist auch nicht zwingend notwendig. Vielmehr sollen sie eine Art moralischen Kompass oder ethische Leitplanken bieten, an denen sich der Mensch zu orientieren versucht.

Letztendlich ist es Aufgabe der Gesellschaft, zu definieren, welche Regeln angewendet werden sollen und welche Risiken annehmbar sind. Auch heute akzeptieren wir, dass Menschen im Verkehr ums Leben kommen. Wir akzeptieren, dass durch die persönliche Ortsveränderung und das Fahren eines Fahrzeugs Risiken in die Gesellschaft eingebracht werden. Diese gehen aber mit einer entsprechenden Verantwortung einher.

Verletzt ein Mensch mit dem von ihm gesteuerten Fahrzeug einen anderen Menschen, so steht er in der Pflicht den entstandenen Schaden zu tragen – sowohl finanziell wie auch moralisch. Diese Verpflichtung sollte nicht an andere Stelle oder gar die Gesellschaft ausgelagert werden, wenn gleichzeitig der persönliche Nutzen der Ortsveränderung einem persönlich zugutekommt.

Rationalität und künstliche Intelligenz – Auf dem Weg zum Fahrzeug mit Moral

Der technische Status quo lässt derzeit nur unvollkommene Lösungen zu. Es steht zu vermuten, dass für die Akzeptanz des automatisierten Fahrens vorab ein gesellschaftlicher Konsens gefunden werden muss, der diese (ethische) Unvollkommenheit für eine Übergangsphase toleriert. Es ist geboten, dass diese Übereinkunft im Rahmen einer breiten gesellschaftlichen Diskussion entsteht und nicht aus dem Faktischen bestimmt wird. Ein „Durchsickern“ des automatisierten Fahrens aus einer vergleichsweise einfachen und geschlossenen Umgebung wie einer Autobahn in ein komplexeres System wie den Stadtverkehr ohne eine vorherige breite, fundierte und ergebnisneutrale öffentliche Diskussion dürfte Widerstände hervorrufen und eine Protestbewegung ähnlich der Maschinenstürmer im 19. Jahrhundert formen.

Im Rahmen dieser gesellschaftlichen Diskussion sollten Wege erkennbar sein, welche auf technischer Ebene aus der ethischen Unvollkommenheit führen. Noah Goodall hat ein mögliches dreistufiges Phasenmodell entwickelt, welches bei Verfügbarkeit der entsprechenden notwendigen Technologie greift¹⁴:

Phase 1 – Rationalität

Die erste Phase basiert auf heute verfügbarer Technologie und stützt sich weitestgehend auf die obigen Ausführungen. Basis sind fixe und transparente Grundregeln und Standards, welche in einem gesellschaftlichen sowie fachlichen Prozess festgelegt werden, und zwingend

¹⁴ Goodall, N. (2014): Ethical Decision Making During Automated Vehicle Crashes. In: Transportation Research Record: Journal of the Transportation Research Board, No. 2424, Transportation Research Board of the National Academies, Washington, D.C., 2014, S. 58–65.

eingehalten werden müssen. Jede in ein Fahrzeug implementierte Regel muss der Öffentlichkeit zugänglich und verständlich sein.

Die Regeln basieren auf einem breiten gesellschaftlichen Konsens und den bereits heute gültigen Moralkodex (Verletzungen sind Todesfällen vorzuziehen, schwache Verkehrsteilnehmer sind besonders zu schützen, Sachschaden vor Personenschaden, etc.).

Zur Auflösung einer Situation, in denen das auf Basis der Regeln entstandene Ergebnis nicht eindeutig ist, ist eine Sicherheitsmetrik zu implementieren. Diese könnte beispielsweise auf dem Wert eines statistischen Lebens (WSL) oder ggf. auch einem Zufallsgenerator basieren.

Da es unwahrscheinlich ist, dass alle möglichen Szenarien durch vorab festgelegte Regeln abgedeckt werden, greift eine weitere Ebene. Sollte eine Situation nicht definiert sein, Regeln miteinander kollidieren oder das ethische Handeln ungewiss sein, wendet das Fahrzeug immer eine „Bremsen und Ausweichen“-Strategie an.

Phase 2 – Ein Hybrid aus Rationalität und künstlicher Intelligenz

Für die zweite Phase wird anspruchsvolle Software benötigt, welche heute noch nicht existiert. Hierbei verwendet der Fahrzeugrechner Techniken des maschinellen Lernens, um menschliche Entscheidungen und die damit einhergehende ethische Bewertung nachzuvollziehen. Die lernfähige künstliche Intelligenz ist hierbei weiterhin an die Regeln der Phase 1 gebunden.

Ein künstliches neuronales Netzwerk würde aus einer Kombination von Simulationen und Aufnahmen von Unfällen sowie Beinahe-Unfällen trainiert werden. Menschen würden die jeweiligen Entscheidungen und Ergebnisse als mehr oder weniger ethisch vertretbar bewerten. Alternativ könnte ein inkrementeller Ansatz verfolgt werden, bei dem ein Computer sich selbst trainiert. Es wäre möglich, dass ein Roboter eine moralisch bessere Vorgehensweise wählt, da er

möglicherweise mehr Handlungsoptionen als der Mensch erkennen und bewerten kann¹⁵.

Entsprechende Forschung steht jedoch noch am Anfang.

Um ein angemessenes Verhalten des Fahrzeugs zu gewährleisten, sollten Grenzen definiert werden. In Phase 2 sollten die Regeln aus Phase 1 als Randbedingungen gesetzt sein. Der maschinelle Lernansatz sollte sich auf Szenarien konzentrieren, die nicht durch die erste Phase abgedeckt sind.

Phase 3 – Regelbasierte Erklärungen neuronaler Netzwerke

Ein großes Manko eines neuronalen Netzes ist seine Unfähigkeit, Entscheidung zu erklären. Im Gegensatz zu einem Entscheidungsbaum, in welchem Entscheidungen in mehreren Schritten bis zum Ursprung zurückverfolgt werden können, kann es schwierig sein zu bestimmen, wie eine künstliche Intelligenz zu ihrer Entscheidung gekommen ist. In einem automatisierten Fahrzeugunfall ist jedoch das Verständnis für die Logik hinter einer Entscheidung eines automatisierten Fahrzeugs von entscheidender Bedeutung. Ohne entsprechende Daten können unerwünschte Entscheidungen nicht nachvollzogen und unterbunden werden.

Computerwissenschaftler haben daher mit der Entwicklung von Techniken begonnen, welche das Auslesen von regelbasierten Erklärungen aus neuronalen Netzwerken ermöglichen¹⁶. Diese sind für den Menschen verständlich.

Neue Technologie der Phase 3 sollte schrittweise in künstliche Intelligenzen der Phase 2 implementiert werden.

¹⁵ Wallach, W., und C. Allen (2008): *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, New York, 2008.

¹⁶ Tickle, A. B., R. Andrews, M. Golea, J. Diederich (1998): *The Truth Will Come to Light: Directions and Challenges in Extracting the Knowledge Embedded Within Trained Artificial Neural Networks*. *IEEE Transactions on Neural Networks*, Vol. 9, Nr. 6, 1998, S. 1057–1068.

Bis diese Technologie jedoch existiert, wird das automatisierte Fahrzeug noch in vielen Bereichen Fragen aufwerfen – im ethisch-moralischen Bereich sind die Antworten mit am schwersten zu finden. Eine gesellschaftliche Debatte ist längst überfällig und wird sich neben dem Fahrzeugbereich auch auf andere Bereiche der Robotik erstrecken.

Polizeiliche Ermittlungsarbeit, Strafverfolgung und automatisiertes Fahren

Verkehrsunfälle mit Beteiligung eines hoch- oder vollautomatisierten Fahrzeugs stellen die Polizei vor neue Herausforderungen. Das Aufgabenfeld der Verkehrspolizei verschiebt sich von der Durchsetzung von Verkehrsregeln stärker in den Bereich IT-Forensik. Statt Geschwindigkeits- und Abstandskontrollen, die Verfolgung von Rotlicht- und Abbiegeverstößen sowie Parksündern dürften in Zukunft die Auswertung von Umfelddaten, Fahrzeugzuständen und Fahrentscheidungen in den Fokus rücken. Diese Verschiebung bringt vermutlich einen Rückgang der Personalstärke bei gleichzeitiger Spezialisierung mit sich. Insbesondere transformative Zwischenzustände, d.h. der Übergang zwischen manuellem und automatisiertem Fahren mit entsprechendem Mischverkehr, werden verkehrspolizeiliche Ermittlungen vor große Herausforderungen stellen.

Im Rahmen eines Verkehrsunfalls mit zwei manuell gesteuerten Fahrzeugen können die Fahrzeugführer im Rahmen der Unfallaufnahme direkt oder im Nachgang befragt werden. In Kombination mit einer Aufnahme diverser Umweltzustände, der Fahrzeugpositionen und Sicherung entsprechender Unfallsuren ergibt sich ein Lagebild, welches die Zuschreibung von Verantwortung in Teilen oder Gänze zulässt. Dies ist unter Umständen auch bei Nutzung teilautomatisierten Assistenzsysteme möglich. Bei Beteiligung eines hoch- bzw. vollautomatisierten Fahrzeugs steigt die Komplexität unter Umständen jedoch um ein Vielfaches an. Zudem ergeben sich diverse Fragestellungen, welche heute noch nicht beantwortet werden können:

- Ist es ausreichend den Fahrzeugspeicher auszulesen und ist die Verlässlichkeit und Vollständigkeit der gesammelten Daten zu Umweltzuständen, Fahrentscheidungen, etc. gegeben? Wie können Detektionsfehler erkannt werden?
- Wer nimmt die Auswertung der entsprechenden Daten vor? Die Komplexität und der Umfang dürften um ein Vielfaches höher liegen als bei Unfällen mit Beteiligung zweier oder mehrerer manuell gesteuerter Fahrzeuge.
- Wer haftet für die Nutzung eines Fahrzeugs mit einem defekten Sensor bzw. mehreren defekten Sensoren? Welche (haftungsrechtlichen / rechtlichen) Folgen ergeben sich, wenn ein Sensor während der Fahrt ausfällt?
- Greift das Auslesen eines entsprechenden Fahrzeugrechners in das Grundrecht auf informationelle Selbstbestimmung ein? In welchem Umfang ist dies noch verhältnismäßig?
- Wie können Entscheidungen neuronaler Netze nachvollzogen werden? Ist das Strafrecht hierfür ausgelegt?
- Inwieweit können Manipulationen von außen erkannt und verfolgt werden? Ist eine derartige Untersuchung bei jedem Unfall geboten bzw. notwendig?
- Inwieweit kann im Rahmen polizeilicher Ermittlungen auf die der Entscheidungen zugrundeliegenden Algorithmen zugegriffen werden? Können diese in ihrer Komplexität voll erfasst werden? Ist ein Zugriff auf Betriebsgeheimnisse von z.B. Automobilherstellern zur Aufklärung eines Verkehrsunfalls verhältnismäßig?
- Dürfen verunfallte automatisierte Fahrzeuge bei unklarer Ursachenlage wieder in den Verkehr gebracht werden? Sind entsprechende Fahrzeuge im Rahmen der Gefahrenabwehr aus dem Verkehr zu ziehen? Sind diese Eingriffe verhältnismäßig?
- Sind Polizei und Justiz überhaupt in der Lage Recht in diesen Bereichen durchzusetzen oder entsteht aufgrund der Komplexität ein Durchsetzungsvakuum?