



Autonome Systeme wie das selbstfahrende Auto nehmen dem Menschen wichtige Entscheidungen ab. Was aber, wenn es ausweglos um Leben oder Tod geht? Sind derartige Ereignisse überhaupt normier- und programmierbar?

# Moralisches Dilemma

Autonome Systeme erobern unsere Welt und treffen selbstständig Entscheidungen. Doch auf welcher Basis? Welchen moralisch-ethischen Grundregeln folgen sie? An dieser Mensch-Maschine-Schnittstelle forschen der Philosoph Mark Siebel und der Neuropsychologe Jochem Rieger

Wir leben in einer Welt, in der wir immer häufiger Entscheidungen an technische Systeme abgeben: Wer heute einparkt, wird von seinem Assistenzsystem in die Parklücke eingewiesen. Wer eine Autofahrt antritt, lässt sich von seinem Navigationsgerät führen und vertraut darauf, dass es den richtigen Weg findet. Im OP-Saal gehen Roboter dem Chirurgen zur Hand und operieren auf einen Zehntelmillimeter genau. Die technologische Weiterentwicklung – Jahrzehnte als science-fiction belächelt – ist längst Gegenwart. Für die einen stellt die Entwicklung dieser cyberphysischen

Systeme eine große Chance dar, andere befürchten, dass die Automatisierung schon bald ganze Berufszweige verdrängen wird: Wer braucht Lkw-Fahrer, wenn die „Brummis“ allein unterwegs sind? Oder Chirurgen, wenn Roboter präziser operieren?

## Neue Fragestellungen betreffen die gesamte Gesellschaft

Ob Befürworter oder Gegner – aus dem Technologie-Fortschritt ergeben sich Fragestellungen, die die gesamte

Gesellschaft betreffen: Wenn zunehmend Maschinen für uns entscheiden und lernen, auf unvorhergesehene Ereignisse zu reagieren, besteht da nicht die Gefahr der gesellschaftlichen Entmündigung? Können wir am Ende überhaupt noch nachvollziehen, warum eine Maschine eine Entscheidung getroffen hat? Und in welchem Ausmaß wollen wir Entscheidungen an technische Systeme abgeben? An der Universität suchen Wissenschaftler verschiedener Disziplinen nach Antworten. „Denn eines ist klar: Wenn sich Forschung und Politik nicht mit dem Thema auseinandersetzen, wird am

Ende die Industrie die Richtung vorgeben“, ist Prof. Dr. Jochem Rieger überzeugt. Ob die Gesellschaft dann noch nachvollziehen kann, wie und warum technische Systeme bestimmte Entscheidungen fällen, bezweifelt der kognitive Neuropsychologe, der unter anderem zur kooperativen Zusammenarbeit von Mensch und Maschine forscht. Sein Schwerpunkt liegt in dem Zyklus „Wahrnehmen, Informationen verarbeiten, Handeln“. „Über diesen Zyklus interagiert der Mensch mit seiner Umwelt. Wir wollen verstehen, wie das genau im Gehirn funktioniert. Entscheidungen sind in diesem Zyklus eine notwendige Voraussetzung für eine Interaktion mit der Welt.“ Sein Team versucht anhand von Gehirnaktivierungsmessungen psychologische Zustände des Menschen vorherzusagen: Ob jemand übermüdet am Steuer sitzt, gestresst oder kognitiv überbelastet ist. „Sind wir in der Lage, diesen Status zu messen und vorherzusagen, könnten wir dem System zusätzliche Informationen über den Menschen liefern, ob er ‚ansprechbar‘ für eine gemeinsame Entscheidungsfindung ist“, erklärt Rieger. Auch ließe sich möglicherweise feststellen, über welchen der fünf Sinne die Maschi-

ne den Partner aktuell am besten erreicht: „Ist der auditive Kanal belegt, weil sich der Mensch unterhält, würde das System versuchen, über den visuellen Kanal Kontakt aufzunehmen“, so der Wissenschaftler. Damit eine Zusammenarbeit funktioniert, müssen Forscher Wege finden, wie Mensch und Maschine in ihrer Zielverfolgung, in ihren Handlungen so miteinander integriert werden, dass sie ihre wechselseitigen Schwächen ausgleichen und ihre Stärken bündeln. „Möglicherweise können wir dann Aufgaben lösen, die wir derzeit gar nicht überschauen können“, so Rieger. Jemand, der beispielsweise einen Produktionsplan erstellt, bekommt durch die Interaktion mit einem technischen System einen neuen Überblick von den Produktionsmaschinen, verbessert daraufhin Arbeits- und Produktionsprozesse und spart auch noch Ressourcen. „Ein abstraktes Beispiel, das zeigt, wie Mensch-Maschine-Systeme die Fähigkeiten und Ziele, die Menschen erreichen können, erweitern können. Die Gesellschaft muss weg von der eher pessimistischen Sichtweise und sollte die Chancen nutzen, um die Kooperation zwischen Mensch und Maschine zu optimieren“, sagt Rieger.

Was einleuchtend klingt, bedarf einer breiten gesellschaftlichen Akzeptanz, genauer Systemvorgaben und der Berücksichtigung ethischer Wertvorstellungen. „Einfach ausgedrückt: Jemand muss dem System im Vorfeld sagen, welche Entscheidungen in verschiedenen Situationen gut oder schlecht sind“, bringt es Prof. Dr. Mark Siebel vom Institut für Philosophie auf den Punkt. An dieser Schnittstelle treffen die Kollegen Rieger und Siebel aufeinander und profitieren vom Know-how des jeweils anderen. „Uns Neurowissenschaftler interessiert vor allem die technische Seite: Wir nehmen ein System, implementieren es und fragen die Menschen, die mit dem System interagieren, welche Handlungen ihnen gefallen und welche sie schlecht finden. Aber daraus können wir keine ethischen Richtlinien ziehen. Dazu bedarf es der Philosophen, die geübt sind in der ethischen Reflexion“, erklärt Rieger. „Wir wiederum profitieren von den Daten der Neurowissenschaftler“, ergänzt Siebel.

### Entscheiden Algorithmen über Leben und Tod?

Ihn interessiert in diesem Zusammenhang vor allem die Moralphilosophie: Was sind moralisch gute und schlechte Entscheidungen? „Da passen autonome Systeme und wie sie lernen, richtige Entscheidungen zu treffen, gut ins Bild“, so Siebel. Allerdings sei längst nicht immer klar, was in bestimmten Situationen moralisch richtig oder falsch ist – „da streiten sich selbst die Philosophen“, gibt Siebel zu bedenken. An dieser Stelle gern zitiert: das als „Weichenstellerfall“ bekannte Trolley-Problem, also die Frage, was man tun soll und wen man opfern würde, um auf Gleisen eingeklemmte Personengruppen vor einer Straßenbahn zu retten. Es wirft eine alte philosophische Frage rund um ein schweres moralisches Dilemma auf. Der Betroffene befindet sich dabei immer in einer schier unlösbaren

Zwickmühle: Ist es gerechtfertigt, den Zug auf ein anderes Gleis zu lenken, wo er zwar auch einen Menschen überfahren würde, aber eben nur einen? Es ist ein Dilemma, in das auch ein autonomes Fahrzeug geraten könnte: Wie soll der Algorithmus da entscheiden, ob das von ihm gesteuerte Auto einen 80-jährigen tötet, um fünf Kindern das Leben zu retten? Argumentiert man im Sinne des Utilitarismus, einer der drei großen Denkrichtungen in der Ethik, würde die „Netto-Rettung“ von fünf Personen das Umbringen einer Person rechtfertigen. „Schwierig, denn da würden Menschenleben gegeneinander aufgewogen werden. Andererseits: Wie sieht eine Alternative aus? Ein Zufallsgenerator, der entscheidet, kann kaum die Lösung sein“, sagt Siebel. Kein Wunder also, dass die von Verkehrsminister Alexander Dobrindt eingesetzte „Ethikkommission für automatisiertes und vernetztes Fahren“ zu dem Ergebnis kommt, dass „Kollege Computer“ am Steuer schwere moralische Dilemmata nicht lösen kann: „Echte dilemmatische Entscheidungen“ wie eine über „Leben gegen Leben“ seien von der konkreten tatsächlichen Situation „unter Einschluss ‚unberechenbarer‘ Verhaltensweisen Betroffener abhängig. Sie sind deshalb nicht eindeutig normierbar und auch nicht ethisch zweifelsfrei programmierbar“. Die Begründung zeigt, wie die Automatisierung die Menschen zwingt, sich mit den Graubereichen der Ethik zu beschäftigen, die sie früher der akademischen Philosophie und dem Zufall auf der Straße überlassen konnten. Denn welcher Autofahrer ist schon in der Lage, in einer Extremsituation binnen Millisekunden eine rationale ethische Entscheidung zu treffen? Da bleibt der Mensch unberechenbar.

### Gefragt: kulturadaptive Systeme

Nicht unwesentlich für die Entscheidungen künftiger autonomer Systeme ist der kulturelle Hinter-



Mark Siebel: „Jemand muss dem System im Vorfeld sagen, welche Entscheidungen in verschiedenen Situationen gut oder schlecht sind.“

grund des Einsatzgebiets: „Die Kulturadaption dürfen wir nicht vergessen“, so Siebel. Während hierzulande das Wohlergehen eines Kindes Priorität habe, gebe es Kulturen, in denen alte Menschen besonderen Schutz genießen, Kühe heilig sind oder Ziegen wertvoller als Hunde – Details, die beim Programmieren berücksichtigt werden müssen. Rieger geht noch einen Schritt weiter: „Autonome Fahrssysteme, so wie wir sie hier in Deutschland diskutieren, werden in einer Stadt wie Neu-Delhi nicht funktionieren!“ Der Verkehr funktioniert dort mit sehr wenigen Normen, die sich aufgrund kultureller Unterschiede auch noch stark von den unseren unterscheiden. Jedes autonome Fahrzeug, ausgestattet mit „westeuropäischen Sensoren“, wäre da überfordert. Welche Algorithmen sollten da noch richtige Entscheidungen treffen?

### Entscheidungsfindung muss nachvollziehbar sein

Eines steht für die beiden Wissenschaftler fest: Autonome Systeme müssen so funktionieren, dass ihre Entscheidungsfindung nachvollziehbar ist. „Wir wissen aus der Rechtsprechung, dass die Motivation für eine Handlung auch immer in die Bewertung des Ergebnisses der Handlung mit eingeht“, erklärt Rieger. Stirbt ein Mensch, ist es ein Unterschied, ob jemand ihn mutwillig, geplant mut-

willig oder böswillig umgebracht hat oder ob sich die Tat durch unglückliche Umstände ereignet hat. „Für die Bewertung spielt somit die Begründung eine große Rolle, deshalb muss sie auch für die Bewertung von Ergebnissen autonomer Systeme eine Rolle spielen“, fordert Rieger. Für ihn gibt es noch einen weiteren Grund, die Entscheidungsfindung im Blick zu behalten: Eine selbstlernende Maschine liefert in Testsituationen Ergebnisse, die im Einklang damit stehen, wie andere Menschen entschieden hätten. „Wenn Sie aber ins Betriebsinnere des Systems gucken, stellen Sie eventuell fest, dass es völlig andere Begründungen für seine Handlungen anbringt als erwartet, beispielsweise rassistische Motive. So etwas lässt sich am Ergebnis aber nicht erkennen.“ Ob sich in der Realität die Beweggründe aus den Algorithmen extrahieren lassen? Rieger und Siebel: „Da sind unsere Kollegen aus dem Bereich des maschinellen Lernens gefragt.“ Das Beispiel zeigt in jedem Fall: Die optimale Entscheidung ist nicht das Ziel, vielmehr eine gut begründete.

Für Rieger und Siebel bieten autonome Systeme viele Chancen, um gesellschaftliche und wirtschaftliche Herausforderungen zu bewältigen. Den Prozess, bis es so weit ist, wollen sie mit richtungsweisenden Inhalten füllen. „Wir sind lange noch nicht so weit, Antworten zu geben, wir müssen erst einmal die richtigen Fragen stellen.“ (kl)



Jochem Rieger: „Die Gesellschaft muss weg von der eher pessimistischen Sichtweise und sollte die Chancen nutzen.“