

Rahmen zu verlassen. Mit anderen Worten: Es ist schlichtweg nicht mehr so, dass diese Systeme nur das tun, was ihnen der Entwickler vorgegeben bzw. ein- oder anprogrammiert hat. Diese Systeme treffen aufgrund ihres Lernfortschrittes dann Entscheidungen, die ihr Entwickler unter Umständen nicht mehr nachvollziehen kann. So hat man beispielsweise ein Spiel entwickelt, das ohne Anfangsregeln ausgestattet, sozusagen im »nackten Zustand« im Rahmen einer Simulation gegen sich selbst gespielt und sich dabei die eigenen Regeln beigebracht hat. Am Ende dieses Lernvorganges war dieses System in der Lage, eigene Gewinnstrategien zu entwickeln, die bisher von keinem Menschen gekannt wurden. Spätestens an diesem Punkt muss die Frage erlaubt sein, ob wir solchen Systemen eine eigene Würde zusprechen wollen oder in einigen Anwendungsfällen sogar müssen. Man denke dabei an die Systeme, die irgendwann im Bereich des autonomen Fahrens die Entscheidung treffen müssen, ob im Rahmen einer Dilemma-Situation das Fahrzeug an die Wand fährt und damit den Insassen gefährdet oder lieber die Oma überrollt, deren Tage gemäß der Berechnung eh gezählt scheinen, oder die Gruppe von Kindern ansteuert, weil die Sensorik diesen Kindern eine gewisse jugendliche Reaktionsschnelligkeit zuschreibt. Wie dem auch sei, kann man diese Systeme als kognitive Entitäten auffassen, die neben rationalen Entscheidungen auch emotional und motivational beladene Entscheidungen treffen können. All dies geben die derzeitigen Kognitionsmodelle her, in die neben den klassischen rationalen Entscheidungsmodellen unter anderem auch Emotionsmodelle und Motivationsmodelle integriert sind. Es ist davon auszugehen, dass immer weitere kognitive Fähigkeiten des Menschen erschlossen und über die Vorlage von kognitiven Funktionsmodellen in Software gegossen und somit als artifiziell-kognitive Fähigkeiten in Maschinen naturalisiert werden.

*Gernot Meier:* Sie sprechen zwei für mich wichtige Bereiche an: Die Frage des freien Willens bzw. der eigenen Handlungsmacht und die Frage der Repräsentanz. Mein Eindruck ist, dass Argumentationsfiguren, die mir hier z.B. aus rechtlichen, philosophischen oder theologischen Bereichen begegnen, noch einem anderen Denken verpflichtet sind. Wenn ich richtig informiert bin, mögen Automobilistinnen und Automobilisten diese Dilemma-Entscheidungen so lange, bis der Algorithmus sagt: »Nicht die beiden Personengruppen außerhalb, sondern ich habe mich entschieden, dass Du, der Fahrer, gegen die Wand gefahren wirst.« Ich denke hier haben wir eine lange Tradition in der Theologie und Philosophie und diese Fragen werden beantwortet werden. Die zweite Frage finde ich aber weitaus brisanter: Was passiert, wenn ich Beziehungen zu einem Wesen aufbaue, dass digital mein Doppelgänger ist oder mit mir unterwegs ist. Was passiert, wenn dieser Doppelgänger oder diese Entität sich weiterentwickelt und andere Erfahrungen machen möchte, wie dies im neuen Film Blade Runner 2049 dargestellt wurde? Aus einem theologischen Blick heraus und als Kulturwissenschaftler sind solche Filme

immer sehr spannend. Hier werden Szenarien neu entwickelt und sie sind dann fast gleichzeitig wieder Quellen für neue Utopien und Dystopien. Eine KI, die eine Prostituierte einlädt, weil sie sagt, sie möchte gerne körperliches Empfinden lernen? Ist das für Sie aus dem Blick eines Informatikers Illusion und wie denken Sie als Philosoph darüber? Stehen wir an dieser Schwelle?

*Matthias Haun:* Als Philosoph und Informatiker schlagen da in der Tat sozusagen »zwei Herzen« in einer Brust. Als Informatiker und damit aus technologischer Perspektive sehe ich es mit Freude, dass immer mehr an kognitiven Fähigkeiten des Menschen erschlossen und dann in Software abgebildet werden können. Das wird uns Maschinen bescheren, die sich in Zukunft in Problem- und Anwendungsräumen bewegen werden, die ihnen in der Vergangenheit eher noch verschlossen waren. Beispielsweise gibt es bereits Unternehmen, die ein intelligentes Entscheidungssystem als gleich- und abstimmungsberechtigtes Mitglied der Geschäftsleitung installiert haben. Bei anonymisierten Entscheidungssituationen ist dann gar nicht mehr von außen erkennbar, welcher Mensch bzw. welche Maschine wie entschieden hat. Man traut diesen Systemen eine hohe Entscheidungskompetenz zu und vertraut somit diesen Systemen im gewissen Sinne. Es wird damit eine neue Richtung eingeschlagen, in der man diesen Systemen eine gewisse Form der Identität zugesteht. Neu ist diese Richtung deshalb, weil man, jetzt kommt der Philosoph in mir ins Spiel, solch eine **Identitätszuschreibung** an eine Maschine eigentlich bisher größtenteils kategorisch abgelehnt hat. Insofern muss man sich in Sachen Identitätszuschreibung meines Erachtens von den traditionellen Denk- und Argumentationsfiguren lösen. In diesem Zusammenhang sei an die Fortschritte der Implantate erinnert, deren Anwendungsfelder auch zunehmend breiter werden. Immer mehr pathologische Fehlfunktionen des Menschen bzw. körperliche Ausfallerscheinungen können durch Implantate korrigiert bzw. übernommen werden. Immer mehr wird aus der ehemaligen Koexistenz von Mensch und Implantat eine Vermischung beider stattfinden. Auch gilt es der Frage nachzugehen, wie sich der Mensch denn fühlt, wenn er immer mehr an Implantate »einverleibt« bekommt. Merkt er das Implantat? Ist der Mensch mit seinem Implantat eigentlich noch derselbe, wie er das ohne das Implantat war? Nimmt er dieses Implantat als Fremdkörper irgendwann mal an? Und wenn, dann wann? Bin ich noch derselbe, wenn ich plötzlich durch diese Technologie über Wahrnehmungsfähigkeiten verfüge, die vorher nicht mein Eigen waren? Bin ich noch der gleiche Mensch, wenn ich plötzlich nicht mehr über sieben natürliche, sondern über zwölf Sinneswahrnehmungen verfüge? An dieser Stelle sei an die Erfahrungen aus dem Bereich der Demenzerkrankung erinnert, wo sich zeigen lässt, dass bereits eine kleine Maschine, die etwas wackelt, surrt, singt oder Farben annimmt, positive Auswirkungen auf das Verhalten der dementen Person zeigt. Vielleicht neigt der Mensch dazu, solche künstlichen Geräte wesentlich früher anzunehmen, sozusagen in sich zu adoptieren, als man bis-

her eigentlich gedacht hatte. Man denke auch an den damaligen Tamagotchi-Hype, wo es sogar zu Selbstmorden kam, als sich das kleine Ding batteriebedingt verabschiedete.

*Gernot Meier:* Ich denke, in vielen Bereichen des Lebens ist uns die Technologie sehr nahegekommen und Jean-Luc Nancy hat das in seinem Eindringling treffend beschrieben. Jedes Cochlea-Implantat ist ein Segen für den Menschen, jeder Herzschrittmacher bedeutet Leben, jeden Tag neu. Und ich freue mich auf den Moment, wenn endlich eine kleine Technologie bei diabeteskranken Menschen den Blutzuckerspiegel permanent überprüft und angemessen handelt. Die Technologie hat dann hoffentlich auch die Fähigkeit, von den Lebensgewohnheiten des Menschen zu lernen: Was er isst, wie er isst, wie er sich bewegt, um sich wirklich in den Organismus integrieren zu können. Selbstlernende und zusätzlich personalisierte Technologien können ein großer Segen sein. Der Punkt ist natürlich, wenn Sie sich eine künstliche Retina einpflanzen lassen, mit der Sie ultraviolettes Licht sehen und auf 500 Meter eine Zeitung lesen können, wird Sie jeder Arbeitgeber nehmen. Ich werde die Arbeitsbiene bleiben, die wahrscheinlich ohne Arbeit sein wird.

*Matthias Haun:* Da bin ich bei Ihnen, zumal ich in einem Alter bin, wo es so langsam an einigen Stellen im Körper zu knacken und zu krachen beginnt, man hier und da nicht mehr ganz so geschmeidig und spritzig ist wie früher. Aber es gibt auch einige Krankheiten, wo die Ergebnisse aus der Künstlichen-Intelligenz-Forschung Früchte tragen, im Sinne von Heilung bringen. Ich erwarte auch von der Technologisierung der Kognitionswissenschaften im Allgemeinen und dort unter anderem durch Themen wie **Big Data** oder **Cognitive Analytics** im Speziellen, dass neben dem bisherigen, eher empirischen Zugang auch Erkenntnismöglichkeiten in den unglaublichen, nunmehr zugänglichen Datenmengen schlummern und sich diese Potenziale mit den inzwischen ausgereiften Techniken verarbeiten und heben lassen. Schon jetzt kann beispielsweise WATSON zu seiner Diagnosefindung auf einen weltweiten Wissensfundus zugreifen, der einem einzelnen Mediziner verschlossen bleibt, weil seine kognitiven Fähigkeiten gar nicht ausreichen, um diese Datenmengen in seiner endlichen Zeit zu verarbeiten.

Solche kognitiven Systeme werden uns Menschen aber auch in die Lage versetzen, auch im hohen Alter unter Umständen noch Dinge tun zu können, die man eigentlich lassen sollte. Man denke nur an den aktuellen Fall, wo eine ältere Dame mit ihrem Auto in eine Motorradgruppe fuhr. Da diskutieren wir doch derzeit, ob man ab einem gewissen Alter noch fahren darf oder nicht. Hier würde ich mir für meine Zukunft schon wünschen, dass ich auch weiterhin am Straßenverkehr teilnehmen kann, wenn gewisse altersbedingte Fehlfunktionen dann von artifiziellen Systemen übernommen werden können. Dieser Verwendungszweck wäre sicherlich ein weiteres positives Beispiel.

Das darf aber nicht darüber hinwegtäuschen, dass Technologie niemals zum **Nulltarif** zu haben sein wird. An diesem Faktum wird auch die künstliche Intelligenz nichts ändern. Die Ergebnisse der Forschung werden in Zukunft höchstens den Abstand zu diesem Nulltarif verringern, sich somit diesem Nulltarif annähern. Es wird also auch weiterhin zu Fehlentscheidungen, Personenschaden oder gar Todesfällen trotz oder aber gerade wegen des Einsatzes von Technologie kommen. So intelligent oder kognitiv diese Technologie auch jemals sein mag. Umso mehr plädiere ich an dieser Stelle dazu, die klassische **Technologiefolgenabschätzung** und dort die Aspekte der funktionalen Sicherheit methodisch und technisch so aufzurüsten, dass diese sich auch den Bereichen der KI-Systeme, wie beispielsweise der autonomen Systeme, adäquat zuwenden können.

Ich möchte aber neben dem technologischen Aspekt an dieser Stelle auch einen gesellschaftlichen Aspekt ins Spiel, sprich in die Diskussion einbringen. Schon heute zeichnet es sich ab, dass sich diese artifizialen kognitiven Funktionen, wie beispielsweise die eben angesprochenen Implantate, nicht jeder leisten können wird. Nicht jeder wird in deren Genuss kommen, weil sie entweder zu teuer sind und/oder die Krankenversicherung die Kosten nicht übernimmt.

*Gernot Meier:* Sind wir dann an einem Punkt, wie er in der Literatur beispielsweise bei Isaac Asimov im Buch »Ich, der Roboter« beschrieben wird. Irgendwann sagt die Technologie: »Also die Menschen auf der Erde, das ist ja wirklich schrecklich, also ich glaub, ich muss die domestizieren.« Isaac Asimov ist ja in der Literatur bei Fragen von Utopien und Dystopien bis heute eine wichtige Quelle sogar bis hin zur Rezeption in der Politik oder der IT-Welt. Inklusive seiner drei grundlegenden Gesetze der Robotik: Ein Roboter darf keinem Menschen schaden oder durch Untätigkeit einen Schaden an Menschen zulassen. Ein Roboter muss jeden von einem Menschen gegebenen Befehl ausführen, aber nur, wenn dabei das erste Gesetz nicht gebrochen wird. Ein Roboter muss seine eigene Existenz bewahren, es sei denn, dies spricht gegen das erste oder zweite Gesetz. Zu ihren vorherigen Gedanken möchte ich gerne etwas anfügen: Nehmen wir an, es wird zunehmend personalisierte Preise in Supermärkten geben und nehmen wir weiter an, diese beruhen auf einer Technologie, die das Kaufverhalten bis zu jedem Duschgel analysiert. Wer wird diese Preise nutzen? Sparsame oder geizige Menschen vielleicht. Aber sicher Menschen, die auf jeden Euro achten müssen. Sie werden notgedrungen diese Technologie ganz nah an sich herankommen lassen. Sie werden notgedrungen gleichsam technomorph werden, weil sie wirtschaftlich dazu gezwungen sind. Leute, die es sich finanziell leisten können, werden datensparsam bleiben können und auch weiterhin zum Einzelhändler gehen. Neoliberale technomorphe Lebensformen werden dann für einen Teil der Menschen zwingend werden.

*Matthias Haun:* Wir haben schon heute den technischen Stand erreicht, wo Systeme wichtige Entscheidungen dem Menschen sozusagen »aus der Hand nehmen«. So wissen wir beispielsweise seit Langem, dass in 90 % der tödlichen Verkehrsunfälle der Mensch der Schuldige ist. Es ist also nur die logische und praktische Konsequenz, wenn ein autonomes Fahrzeug den Fahrer in Gefahrensituationen als »Bug« aus der Entscheidungskette herausnimmt und die Entscheidungen trifft. Das System »Fahrzeug« entscheidet, da der menschliche Fahrer als »Fehlfunktion« die Situation nur verschlimmern kann.

Oder nehmen wir den Fall, dass der Mensch sich als Wesen entpuppt, der mit seinen beschränkten kognitiven Fähigkeiten gar nicht in der Lage ist, die anstehenden Weltprobleme nachhaltig zu lösen. Auch das könnte ein legitimer Grund dafür sein, den Menschen, zumal er diese Problemlagen eigentlich selbst erschaffen hat, aus dem ganzen Dilemma zu eliminieren. Technologisch sind also die angesprochenen Dystopien durchaus in Reichweite. Was davor sich abspielen könnte, resultiert aus der immer größeren Komplexität der Lebenswelt durch die zunehmende Vernetzung der Dinge. Ich bezweifle, dass wir Menschen mit unseren limitierten kognitiven Fähigkeiten diese Komplexität noch durchdringen, geschweige denn beherrschen. Denken Sie nur an den schwarzen Freitag. Oder man führe sich unsere Abhängigkeit vom Energienetz vor Augen, indem ein Ausfall durch die Koppelwirkungen keinen einzigen Bereich der Lebenswelt verschonen wird. Ich glaube, dass diese Szenarien in greifbarer Nähe sind. Wir sollten also damit »rechnen«, uns darauf einstellen und im Rahmen einer nüchternen und fundierten Technologiefolgenabschätzung Lösungsszenarien für solche Fälle entwickeln. Vielleicht zeigt sich dann, dass die kognitiven Systeme in diesen kritischen Fällen die Entscheidungs- und Handlungshoheit zum Wohle des Menschen einfach übernehmen müssen. Und wenn man sich umschaute, muss man erkennen, dass wir diese Reise in das Land der Abhängigkeiten eigentlich bereits angetreten sind. Wir verlassen uns auch heute schon in einigen Lebensbereichen auf komplexe Technologie, ohne diese umfassend zu verstehen oder deren nachhaltige Wirksamkeit vollständig zu begreifen.

*Gernot Meier:* Die Verbindung zwischen digitaler Gedankenlosigkeit und einer positiv-utopischen Technologievorstellung ist hier eine gleichsam wirkungsvolle Allianz. Da diese Utopie eine berechenbare Welt verheißt inklusive der Frage der Gerechtigkeit und der Idee: Endlich wird der Gute belohnt und der Böse bestraft. Die Technologie wird es richten. Diese Entwicklung finde ich beispielsweise bezüglich des Einsatzes von militärischen Drohnen besorgniserregend. Hat die sogenannte »Killchain« noch einen Menschen gleichsam dazwischen? Wie ist es mit autonomen Systemen an Grenzen. Sind die Drohnen von heute die Selbstschussanlagen von gestern? Autonome Systeme, die dann entscheiden: Ist das ein illegaler Grenzübertritt oder ist es ein Bauer, der sich verfahren hat?

*Matthias Haun:* Wenn man dem Thema »Drohne im Kriegseinsatz« überhaupt einen positiven Aspekt abringen kann, dann ist es der, dass durch die Kognitivierung dieser Killermaschine eventuell Kollateralschäden minimiert werden können. Durch die erweiterten Wahrnehmungsmöglichkeiten, die schnellere Verarbeitung von riesigen Datenmengen kann das System wesentlich trennschärfer zwischen Freund und Feind unterscheiden. Handelt es sich bei der Ansammlung von Menschen an dem Tankwagen um militante Terroristen oder spielende Kinder, die dort mit dem Wasser spielen? Aber auch hier gilt, dass es wiederum nicht die einzelne Drohne ist, die Entscheidungen trifft. Vielmehr sind diese Drohnen in Zukunft miteinander vernetzt. Das bedeutet, die Drohnen treffen Entscheidungen nicht mehr alleine, sondern im Kollektiv beteiligter Drohnen. Ein weiterer Aspekt, den es zu berücksichtigen gilt, ist der, dass die fatalen Wirkungen von Entscheidungen sich in der Ferne abspielen. Der Pilot als Entscheider ist sozusagen mit den Wirkungen nicht mehr frontal und unmittelbar konfrontiert. Alles spielt sich weit weg ab. Er kann diese Drohne von seiner Kommandozentrale über Tausende Kilometer entfernt in sein Ziel fernsteuern. Sarkastisch könnte er die Entscheidungen auch schon heute der autonomen Drohne überlassen, sozusagen den Monitor und damit den Rest an seelischem Konflikt einfach ausschalten. Kognitionstechnologisch lassen sich diese Drohnen mit Kognitionsmodellen und damit mit Emotionsmodellen oder Motivationsmodellen ausstatten. Das bedeutet, dass diese Drohnen mit einem Überlebensmodus unterwegs sind, dass das Überleben der eigenen System- und Vitalfunktionen mit der höchsten Priorität versehen ist. Auch die Motivation, unbedingt als Sieger vom Feld zu gehen, lässt sich entsprechend modellieren und programmiertechnisch abbilden. Auch in diesem Anwendungsfeld könnte es passieren, dass man ganz bewusst das Töten einer autonomen Maschine überlässt, um sich als Mensch nicht mehr selbst verantworten zu müssen. Man ist dann nur an Erfolgsmeldungen interessiert. Ziel erreicht! 20 Kinder haben gelitten. So what? Die Misserfolgsmeldung wird ignoriert. Ist ja alles weit weg. Kein schmerz erfülltes Schreien, keine verzerrten Gesichter, keine weinenden Mütter. All das weiche Menschliche wird dann im digitalen Rauschen der Informationen untergehen.

*Gernot Meier:* Felder, in welche diese Technologie der abgeschobenen Agency schon subtil Einzug gehalten hat, sind beispielsweise das Bankenwesen, Stichwort Kreditvergabe, oder auch bei Personalkontakten und bei Einstellungen. Weil oftmals maßgebende Personen aus der Wirtschaft oder der Politik einer Technologie, und das möchte ich schon nochmal verstärken, ein Stück weit technische Objektivität zuschreiben. Technologie wird gedanklich in das Feld von Mathematik, Kaffee- oder Drehmaschinen und Smartphones bewegt. Fehler oder Probleme werden hier in der Regel so gefixt, dass die Anwendung smarter wird und die Technologie entscheidet weiter. Und dann denke ich: Freunde, so funktioniert das nicht.