

AUS POLITIK UND ZEITGESCHICHTE

Künstliche Intelligenz

Rául Rojas · Thea Dorn

DIE KÖNNEN WAS!
ABER KÖNNEN ROBOTER
AUCH FÜHLEN?

Ulrich Eberl

WAS IST KÜNSTLICHE
INTELLIGENZ – WAS KANN
SIE LEISTEN?

Thomas Ramge

MENSCH FRAGT,
MASCHINE ANTWORTET.
WIE KÜNSTLICHE INTELLIGENZ
WIRTSCHAFT, ARBEIT UND
UNSER LEBEN VERÄNDERT

Christoph Kehl

ENTGRENZUNGEN ZWISCHEN
MENSCH UND MASCHINE,
ODER: KÖNNEN ROBOTER ZU
GUTER PFLEGE BEITRAGEN?

Catrin Misselhorn

KÖNNEN UND SOLLEN
MASCHINEN MORALISCH
HANDELN?

Oliver Bendel

ÜBERLEGUNGEN ZUR DISZIPLIN
DER MASCHINENETHIK

Ingo Irsigler · Dominik Orth

ZWISCHEN MENSCHWERDUNG
UND WELTHERRSCHAFT:
KÜNSTLICHE INTELLIGENZ
IM FILM

APuZ

ZEITSCHRIFT DER BUNDESZENTRALE
FÜR POLITISCHE BILDUNG

Beilage zur Wochenzeitung Das **Parlament**

Künstliche Intelligenz

APuZ 6–8/2018

RÁUL ROJAS · THEA DORN

DIE KÖNNEN WAS!

ABER KÖNNEN ROBOTER AUCH FÜHLEN?

Unvorstellbar, sagt der Informatikprofessor Raúl Rojas. Trotzdem muss man die Visionen zu Ende denken, findet die Philosophin Thea Dorn. Ein Gespräch über Künstliche Intelligenz.

Seite 04–07

ULRICH EBERL

**WAS IST KÜNSTLICHE INTELLIGENZ –
WAS KANN SIE LEISTEN?**

Maschinen lernen dank Künstlicher Intelligenz sprechen, sehen und lesen. Sie werden bald alle Lebensbereiche grundlegend verändern. Was treibt diese Entwicklung, wie weit geht sie, und bleiben die smarten Maschinen beherrschbar?

Seite 08–14

THOMAS RAMGE

**MENSCH FRAGT, MASCHINE ANTWORTET.
WIE KÜNSTLICHE INTELLIGENZ WIRTSCHAFT,
ARBEIT UND UNSER LEBEN VERÄNDERT**

Mit Künstlicher Intelligenz können Menschen viele Entscheidungen an digitale Assistenten delegieren. Das wird unser Leben, Wirtschaft und Arbeit verändern. Eine wichtige Frage bleibt: Welche Entscheidungen wollen wir selbst treffen?

Seite 15–21

CHRISTOPH KEHL

**ENTGRENZUNGEN ZWISCHEN MENSCH UND
MASCHINE, ODER: KÖNNEN ROBOTER
ZU GUTER PFLEGE BEITRAGEN?**

Insofern Technik beginnt, autonom zu agieren, verwischt auch die Abgrenzung zwischen Mensch und Maschine. Im Anwendungsfeld der Pflege werden die gesellschaftlichen Herausforderungen dieser Entwicklungen deutlich.

Seite 22–28

CATRIN MISSELHORN

**KÖNNEN UND SOLLEN MASCHINEN
MORALISCH HANDELN?**

Das Vorhaben, künstliche Systeme nicht nur mit kognitiven Fähigkeiten, sondern auch mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten, wird in diesem Beitrag erläutert und kontrovers diskutiert.

Seite 29–33

OLIVER BENDEL

**ÜBERLEGUNGEN ZUR DISZIPLIN
DER MASCHINENETHIK**

Anders als die klassischen Bereichsethiken, die Reflexionsdisziplinen sind, ist die Maschinenethik eine Gestaltungsdisziplin und bringt zusammen mit Künstlicher Intelligenz und Robotik moralische und unmoralische Maschinen hervor.

Seite 34–38

INGO IRSIGLER · DOMINIK ORTH

**ZWISCHEN MENSCHWERDUNG UND
WELTHERRSCHAFT: KÜNSTLICHE INTELLIGENZ
IM FILM**

Spielfilme verweisen auf Technikreflexionen in der realen Welt und positionieren sich implizit dazu, indem Möglichkeiten, vor allem aber Gefahren und Grenzen der Künstlichen Intelligenz im Modus der Fiktion ausgelotet werden.

Seite 39–46

EDITORIAL

Mit der Gründung der Kirche „Way of the Future“ beschreitet der Robotiker Anthony Levandowski neue Wege im Verhältnis von Mensch und Maschine. Wer möchte, kann mit ihm als selbsternanntem Dekan „eine auf KI basierende Gottheit aus Hardware und Software realisieren, akzeptieren und anbeten“. Dieses extreme Beispiel basiert auf der Vorstellung, dass sich eines Tages eine Art „Super“- oder „Hyperintelligenz“ entwickelt, die den Menschen überlegen ist und damit gottähnliche Züge tragen könnte. Vor dieser Zukunftsvision warnen viele, manche sehnen sie herbei, und andere halten sie schlicht für unmöglich.

Fortschritte auf dem Gebiet der Künstlichen Intelligenz werfen Fragen auf, wie sie sich für jede technologische Revolution stellen: Was ist von Nutzen und Vorteil für den Menschen jenseits der technischen Machbarkeit? Wie verändern sich Wirtschaft, Arbeit und Alltag? Wo liegen Risiken? Wie lassen sich diese Entwicklungen gesellschaftlich und politisch steuern? Die Debatte um KI berührt zusätzlich Kernbereiche des Menschlichen, wenn die Grenzen zwischen Mensch und Maschine verwischen und die Maschine nicht länger ein bloßes Werkzeug ist, sondern selbst Handlungsentscheidungen treffen kann.

Ein viel zitiertes Beispiel in der Diskussion darüber, welche Entscheidungen wir an autonom agierende Systeme delegieren wollen, sind selbstfahrende Autos. Die vom Bundesministerium für Verkehr und digitale Infrastruktur berufene Ethikkommission „Automatisiertes und vernetztes Fahren“ hat dazu im Juni 2017 20 Thesen vorgestellt. Unter anderem halten es die Kommissionsmitglieder in unvermeidlichen Unfallsituationen für unzulässig, mögliche Verkehrstopfer nach persönlichen Merkmalen wie Alter oder Geschlecht zu bemessen, eine allgemeine Programmierung auf eine Minderung der Zahl von Personenschäden aber für potenziell vertretbar. Die Debatte um diesen und andere Anwendungsfälle der KI steht erst am Anfang.

Anne Seibring

DOPPELINTERVIEW

DIE KÖNNEN WAS! ABER KÖNNEN ROBOTER AUCH FÜHLEN?

Unvorstellbar, sagt der Informatikprofessor.

Trotzdem muss man die Visionen zu Ende denken,
findet die Philosophin.

Rául Rojas und Thea Dorn über Künstliche Intelligenz

Wenn Sie sich eine intelligente Maschine wünschen dürften, welche wäre das?

Thea Dorn – Ich will mit dem Kopf und dem Herzen arbeiten, deshalb finde ich es wunderbar, dass Geräte mir Tätigkeiten im Haushalt abnehmen, die ich nur zähneknirschend machen würde. Ich bin aber noch nie auf die Idee gekommen, meine Waschmaschine intelligent zu nennen.
Rául Rojas – Ich wünsche mir eine Technologie, die mir nicht ständig im Weg steht und mich an den Computer fesselt – wie es zum Beispiel E-Mails tun. Ich habe kein Handy, das ist mein Ausweg. Ich möchte mein Gehirn für etwas anderes verwenden. Also entwickle ich etwa autonome Fahrzeuge, die man wie Taxis nutzen kann.

Sind wir denen nicht erst recht ausgeliefert?

Rojas – Nein. Im Gegenteil. Ich möchte das Auto in ein öffentliches Verkehrsmittel verwandeln. Wir haben in Berlin 1,3 Millionen Fahrzeuge, dazu Busse und U-Bahnen. Ich hätte gern statt der Autos nur noch 100 000 autonome Fahrzeuge: Sammeltaxis. Sie würden die Leute auf der Straße aufnehmen und sie zur U-Bahn, zum

Bus oder einem anderen Ziel bringen. Man fährt gemeinsam, teilt sich die Kosten. Das entlastet den Verkehr. Es gäbe keine Parkspuren mehr, und die frei gewordene Spur bekämen die Fahrräder. Es gäbe weniger Unfälle. Und es wäre so bequem! Ich brauche nicht mehr selbst zu fahren, ich kann ins Taxi einsteigen, lesen, arbeiten. Das wäre meine „Autopie“ für die Stadt.

Dorn – Keine Einwände gegen diese Vision! Letztlich ist es mir egal, ob ein Chauffeur am Steuer sitzt oder ob das Auto selbst fährt. Aber einstweilen sitzen wir ja selbst noch am Steuer – und werden dauernd bevormundet. Es macht mich rasend, wenn mein Auto mir durch hysterisches Piepsen vorschreiben will, wie ich einparken oder wann ich Pause machen soll. Und Navigationssysteme sind die reinsten Orientierungs-Verblödungs-Maschinen. Wenn möglich, schalte ich das alles ab.

Rojas – Ein Navigationssystem habe ich auch nicht. Aber im Ernst: Ich halte viele Vorträge über autonome Fahrzeuge, und die häufigste Reaktion ist: Ich will nicht, dass mir jemand das Lenkrad wegnimmt! Da antworte ich: Reiche Leute ha-

ben einen Chauffeur und kümmern sich nicht um die Details. Wollen wir nicht alle reich sein? Dann interessiert mich das Fahren nicht mehr. Ich lese meine Zeitung, mein Buch, bis der Chauffeur sagt: Wir sind angekommen.

Dorn – Auch dies: schöne Vision! Ich fürchte nur, dass die Leute die gewonnene Zeit nicht nutzen werden, um Aristoteles oder wenigstens die Zeitung zu lesen. Sie werden noch mehr an ihren Geräten herumdadeln. Aber es stimmt natürlich, dass sich – gerade in Großstädten – im Verkehr viel verbessern muss. Jeder, der in Berlin zu Fuß oder mit dem Rad unterwegs ist, hasst den Autoverkehr.

Aber?

Dorn – Mich beunruhigt die Rasananz, mit der wir Entscheidungen, die wir früher selbst treffen mussten, oder Kompetenzen, die wir uns selbst aneignen mussten, an die Technik delegieren: Ein Gespür dafür entwickeln, ob ich mich heute genügend bewegt habe? Überflüssig, die Gesundheits-App sagt's mir. Vorratsplanung? Überflüssig, der Kühlschrank weiß selbst, wann neue Milch her muss. Fremdsprachen-

lernen? Wozu, es gibt doch Übersetzungsprogramme.

Herr Rojas, Sie wollen eine Technologie, die Ihnen Dinge abnimmt und Arbeit von Ihnen fernhält. Stellen wir uns also einen Haushalt vor, der von selbst Wäsche wäscht, das Essen macht ...

Rojas – Nein, nein. Ich mag keine Haushaltsroboter. Mein persönlicher Sport ist es, den Rasen zu mähen, Blätter aufzusammeln und die Wäsche zu machen. Man sollte den Menschen nicht alles im Haushalt abnehmen, weil sie sich dann nicht mehr bewegen und dick werden. Wie im Film „WALL-E“. Kennen Sie den?

Dorn – Nur den Trailer.

Rojas – Da sitzen die Menschen auf einer fliegenden Couch, sie haben den Bildschirm vor der Nase, und Roboter bringen das Essen. So eine Zukunft wäre grausam. Manche Maschinen, die gebaut werden könnten, sollte man deshalb aus sozialen Gründen nicht bauen. Pflegeroboter für Krankenhäuser etwa, die den Kontakt der Kranken zu anderen Menschen minimieren.

Dorn – An deren Entwicklung wird aber massiv gearbeitet. Einerseits ist das verständlich: Es herrscht ein Mangel an Pflegekräften, warum also nicht Abhilfe durch Roboter schaffen? Andererseits wird mit jeder technologischen Lösung eines gesellschaftlichen Problems unsere soziale Kreativität weiter abnehmen. Im Kern scheint es mir darum zu gehen: Wie füllen wir die Freiräume aus, die uns die neue Technik beschert? Gehen wir einem Athen 4.0 entgegen, in dem jeder von morgens bis abends philosophieren

und Sport treiben kann, weil ihm freundliche Robotersklaven alle lästige Arbeit abnehmen? Oder enden wir so, wie „WALL-E“ es nahelegt: dick, faul und dumm?

Rojas – Das ist das Dilemma. Die klassische Antike ist sowie-so nicht mein Vorbild, weil da nur Männer frei waren, während Frauen und Sklaven die Arbeit machten. Ich denke eher an die utopischen Sozialisten.

Dorn – Trotzki wäre heute wohl ein großer Verfechter der Technologisierung, weil er hoffen würde, man bekäme dadurch die Freiheit zur Bildung.

Rojas – Die große Frage bei den frühen Sozialisten war ja, wer die niederen Arbeiten macht. Ihre Antwort: Maschinen. Friedrich Engels schrieb, als er noch jung war: Ihr habt gesagt, wir brauchen eine untere Klasse, die Toiletten putzt, aber jetzt gibt es eine automatische Toilette, die selbst spült! Für ihn bewies das, dass die Menschheit der Zukunft von diesen niederen Arbeiten entbunden sein werde – die sozialistische Utopie des 19. Jahrhunderts. Inzwischen haben wir viel mehr Maschinen, als sich Marx und Engels jemals erträumt haben, aber die tolle Zukunft mit den freien Menschen haben wir nicht. Da kann man schon fast verzweifeln.

Dorn – Tja. Ich glaube, der Mensch hat nur zwei Möglichkeiten: Entweder er findet sich weiterhin damit ab, dass er neben all dem Großartigen, das er kann, auch ein beschränktes Mängelwesen ist: manchmal traurig, manchmal krank, und am Schluss steht der Tod. Oder er will all dies endgültig hinter sich lassen und ewig fit, ewig gut drauf und in letzter

Konsequenz auch unsterblich sein. Unser abendländisches, humanistisch geprägtes Menschenbild wäre damit am Ende. Ich vermute, im Silicon Valley würde ich höhnisches Gelächter ernten, wenn ich dort sagen würde, dass ich die weitere Optimierung des Menschen im Maschinenstil für das Ende unserer Autonomie und Würde halte. Die Allerkühnsten dort träumen ja davon, ihr Hirn in die Cloud hochzuladen, wo es mit dem Weltwissen verschmelzen kann: die totale Entgrenzung. In meiner Sprache: das Ende von Persönlichkeit, Charakter, Individuum.

Rojas – Das sind alles Fantasien, das ist alles Quatsch.

Dorn – Es beruhigt mich, wenn Sie das sagen.

Woher weiß ich als normale Leserin, dass das Quatsch ist?

Rojas – Das ist schwer herauszufinden. Ray Kurzweil, der Vordenker des Transhumanismus, behauptete schon vor Jahrzehnten, die Menschen würden durch Maschinen ersetzt, es gäbe irgendwann nur noch maschinelle Intelligenz. Er sagt, 2030 werden Computer genauso viele Transistoren wie Menschen Neuronen haben, also werden die Maschinen automatisch intelligenter sein. Das ist Unsinn, und es ist vermessen. Immer wieder merken wir, wie weit wir noch davon entfernt sind, Gehirn und Kognition zu verstehen. Wer glaubt, dass Maschinen haargenau wie Menschen sein können, verkennt, dass wir Körper und Geist sind.

Dorn – Kurzweil ist so konsequent, den Körper für ein lästiges Anhängsel des Geistes zu halten.

Rojas – Wenn er all seine Informationen auf einer Festplatte speichern könnte, würde er schnell merken, dass er ohne Körper nicht mehr derselbe wäre.

Dorn – Wie gesagt: Ich hoffe, Sie haben recht und die „Singularität“, von der Kurzweil träumt, wird es niemals geben. Trotzdem glaube ich nicht, dass man alle trans- und posthumanistischen Visionen achselzuckend übergehen sollte ...

Rojas – Doch. (lacht)

Dorn – ... weil sie uns mit dem Fluchtpunkt unserer technologischen Zivilisationsgeschichte konfrontieren. Sie zwingen uns, uns zu fragen: Wollen wir den Weg, den wir eingeschlagen haben, wirklich so radikal zu Ende gehen?

Rojas – Wer Roboter baut, ist wie ein Zauberer, und die Leute lassen sich täuschen. Ein Beispiel: Wenn Sie kleine Roboter per Zufallsprinzip Würfel durch die Gegend schieben lassen, bildet sich irgendwann in der Mitte der Arena ein Haufen. Das ist zufällig, doch die Leute denken: Ah! Die Roboter bauen gemeinsam einen Turm. Sie sind intelligent!

Dorn – Kommt das nicht daher, dass die Menschen seit Anbeginn davon träumen, Wesen zu erschaffen, die sie unendlich überbieten – an Gesundheit, in Langlebigkeit, Schönheit, Stärke, Intelligenz? Pygmalion und die ideale Frau, die jüdische Mythologie und der Golem, Nietzsche und der Übermensch. Diese Begeisterung wundert mich also nicht. Aber ich teile Ihre Einschätzung, dass wir es zumindest derzeit noch mit eher billigem Budaenzauber zu tun haben. Neulich war ich bei einem Empfang, da

liefe ein kleiner weißer Roboter herum, begrüßte die Gäste, konnte – meistens – erkennen, ob ihm ein Mann oder eine Frau gegenüberstand. Meinte er, eine Dame zu sehen, gab's ne Rose.

Rojas – Das ist Volksverdummung. Ich weiß, was der Stand der Technik ist. Wir kochen alle nur mit Wasser. Die Leute, die diese Roboter bauen, tun so, als ob der Roboter wirklich viel versteht. Statt über die Grenzen und Probleme der Technologie aufzuklären, übertreiben sie maßlos, um Investoren oder neue Gelder für die Forschung zu bekommen.

Dorn – Glauben Sie, dass diese Künstlichen Intelligenzen eines Tages so etwas wie Emotionen ausbilden könnten?

Rojas – Nein. Das ist absolut undenkbar. Ich arbeite seit 35 Jahren daran. Es gibt gewisse bescheidene Fähigkeiten, die wir den Computern beizubringen versuchen. Aber: Die Spracherkennung ist immer noch nicht toll. Es gibt ein paar brauchbare Sachen wie Gesichtserkennung am Flughafen oder Übersetzungssysteme. Letztere funktionieren, weil inzwischen so viele Bücher auf Deutsch und Englisch online stehen, dass der Computer die Korrespondenz zwischen beiden Sprachen automatisch lernen kann.

Das nennt sich Deep Learning – Menschen trainieren die Programme mit großen Datenmengen.

Rojas – Ja. Es funktioniert allerdings nur zu einem gewissen Grad. Ich zeige dem Computer Millionen von Katzenbildern. Dann weiß er irgendwann, was eine Katze ist. Aber wenn ich

ihn frage, wieso?, kann er es nicht erklären.

Woher kommen diese Daten?

Rojas – Von uns allen.

Dorn – Täglich.

Rojas – Wenn man auf Instagram Bilder hochlädt, landen die Daten bei Facebook. Damit können die ihre Gesichtserkennungsprogramme verbessern. Oder Katzenerkennungsprogramme. Wir generieren diese Daten – freiwillig und kostenlos für Google und Facebook.

Eine Google-KI hat unlängst schwarze Menschen als Gorillas bezeichnet. Eine andere KI hat Küche und Frauen verbunden. Lernen die Systeme nicht wertneutral?

Rojas – Nein. Bei Deep Learning gibt es nur Muster. Und daraus ergibt sich eine Entscheidung. Wenn in 90 Prozent der Küchenwerbung eine Frau zu sehen ist, dann ist klar, woher das kommt. In Big Data sind all unsere Wertvorstellungen, Widersprüche und Vorurteile enthalten.

Dorn – Ich habe gehört, dass eins dieser sogenannten autonomen Systeme im militärischen Bereich, das Panzer gut erkannt hat, plötzlich versagte. Keiner wusste, warum. Dann fand man heraus, dass die Mehrheit der Bilder, mit denen das System trainiert worden war, Panzer unter blauem Himmel zeigte. Nun war der Himmel bedeckt, und das System ratlos.

Rojas – Das ist leider ein Märchen, so ein System war nie im Einsatz. Ein gravierendes Beispiel ist ein Unfall von Tesla, weil da ein echtes, kommerzielles System beteiligt war. Da blockierte ein weißer Lkw quergestellt die Autobahn. Die

Kamera des Autos konnte die Wolken am Himmel nicht vom Laster unterscheiden, es bremsste nicht.

Dorn – Die menschliche Urteils kraft ist durch nichts zu ersetzen. Der Mensch ist imstande, abzuwägen und zu einem Urteil zu kommen, zu argumentieren und seine Entscheidung zu begründen. Kann das einer der Unterschiede sein zwischen Künstlichen Intelligenzen und uns?

Rojas – Ja. Maschinen verfügen tatsächlich über keine tiefe Urteils kraft. Vor allem besitzen sie keine Intuition, kein Finger spitzengefühl.

Dorn – Es geht doch noch um viel mehr: Künstliche Intelligenzen kennen keine Scham und Schuldgefühle, kein Mitleid, kein Gewissen. An Ihren Forschungsbereich wird ja oft die Frage gestellt: Wer trägt die Verantwortung, wenn ein selbstfahrendes Auto in einer brenzligen Situation entscheidet, lieber das Rentnerhepaar als die Mutter samt Kind zu überfahren? Müssten wir nicht eigentlich fragen, was es bedeutet, wenn auf unseren Straßen demnächst Hunderttausende Gefährte unterwegs sind, denen es schlicht egal ist, wenn sie jemanden überfahren haben?

Rojas – Diese Frage höre ich immer wieder: Wer trägt die Verantwortung? Natürlich die Menschen, die das System bauen! Mein Vorschlag ist, im Vorfeld so gut zu arbeiten, dass solche Situationen gar nicht erst entstehen können. Mindestabstand, reduzierte Geschwindigkeit, leichte Baustoffe – wenn wir das alles berücksichtigen, können wir hoffentlich Fahrzeugsysteme bauen, die keine schlimmen Kollisionen ver-

ursachen und letztlich sicherer sind als die Autos heute.

Sollte die KI international reguliert werden?

Dorn – Die Menschheit hat die Internationale Atomenergie-Organisation gegründet, als sie erkannt hat, welche gefährliche, potenziell die ganze Erde bedrohende Macht sie mit der Atomkraft entfesselt hat. Vermutlich ist es an der Zeit, darüber nachzudenken, ob wir nicht eine vergleichbare Einrichtung brauchen, die über die Entwicklung und den Einsatz von KI wacht.

Rojas – Ich bin immer für Regulierung, zum Beispiel von Gentechnologie und Waffen. Nur in diesem speziellen Fall sehe ich nicht auf Anhub, wie man intelligente Algorithmen regulieren kann. Darüber muss man gründlich nachdenken – die Lösung ist nicht gerade trivial.

Und wir müssen uns wirklich keine Gedanken darüber machen, Herr Rojas, dass Roboter einmal klüger sein werden als Menschen oder sich einst von uns emanzipieren?

Rojas – Auf jeden Fall nicht in diesem Jahrhundert! Ich glaube generell nicht, dass Roboter irgendwann die Herrschaft über uns übernehmen werden. Konrad Zuse, der erste Computerbauer in Deutschland, sagte einmal: Wenn die Computer zu frech werden, zieh den Stecker raus.

Das Interview erschien zuerst in „Chrismon“.

Moderation: Mareike Fallet und Michael Güt hlein, *chrismon* 12/2017, www.chrismon.de

THEA DORN

ist Schriftstellerin und Philosophin. In ihrem Roman „Die Unglückseligen“ (Knaus) geht es um die Abschaffung der Sterblichkeit. Das Thema Künstliche Intelligenz treibt sie um: Auf der diesjährigen Jahrestagung des Deutschen Ethikrats zum Thema „Autonome Systeme“ ging sie der Frage nach, wie diese Maschinen unser Selbstverständnis ändern.

RÁUL ROJAS

ist Mathematiker und Professor für Informatik an der Freien Universität Berlin. Er leitet das Dahlem Center for Machine Learning and Robotics und wurde bekannt durch fußballspielende Roboter, mit denen sein Team und er viele Meisterschaften gewannen. Derzeit arbeitet er an autonom fahrenden Fahrzeugen, der Prototyp heißt „AutoNOMOS“. Seine Doktorarbeit hatte seinerzeit ein ganz anderes Thema: Karl Marx.

WAS IST KÜNSTLICHE INTELLIGENZ – WAS KANN SIE LEISTEN?

Ulrich Eberl

Vor 250 Jahren meldete James Watt seine Verbesserung der Dampfmaschine zum Patent an. Seither konstruieren findige Forscher ständig neue Maschinen, um Menschen mühselige Arbeiten abzunehmen. Die Dampfmaschine revolutionierte nicht nur Bergbau, Schifffahrt und Eisenbahnen, sondern auch die Stahl-Walzstraßen und mit Spinn- und Webmaschinen die Textilindustrie. Hundert Jahre später begann der Siegeszug von Elektro- und Verbrennungsmotoren. Auto, Flugzeug und Fließband ermöglichten Massensmobilität und Massenfertigung. Und wieder hundert Jahre danach führten Mikrochip, Computer, Handy und Internet zu Automatisierung, Digitalisierung und grenzenloser Kommunikation.

Doch noch nie in dieser langen Geschichte der industriellen Revolutionen war der Kern des Menschseins in Reichweite der Maschinen: unsere Intelligenz. Genau dies ändert sich gerade. In den vergangenen fünf Jahren hat es auf dem Gebiet der Künstlichen Intelligenz (KI) mehr Fortschritte gegeben als in den 50 Jahren zuvor. Auf etlichen Feldern, die bisher Menschen vorbehalten schienen, übertreffen uns Maschinen bereits. Sie schlagen menschliche Weltmeister im Quizspiel Jeopardy ebenso wie auf dem Go-Brett, sie können in nicht einmal zwei Stunden auf Millionen von Google-Street-View-Bildern die Hausnummern finden, sie machen bei der Erkennung von Verkehrszeichen nur halb so viele Fehler wie Menschen, sie finden in Gewebeschnitten Hinweise auf Krebszellen, die Ärzten bisher unbekannt waren, und sie können grundlegende Emotionen wie Wut, Freude, Trauer und Überraschung aus Gesichtern besser lesen, als viele von uns es vermögen.⁰¹

SMARTE MASCHINEN LERNEN SPRECHEN, SEHEN, LESEN

Vor allem das Sprach-, Text- und Bildverständnis von Maschinen mit Künstlicher Intelligenz leistet Erstaunliches: Siri, Alexa, Cortana, Google Assis-

tant, und wie die virtuellen Assistenten alle heißen, lernen derzeit mit rasender Geschwindigkeit, Fragen und Befehle von Menschen zu verstehen und sinnvoll zu beantworten. Übersetzungsprogramme wie Google Translate oder DeepL können in Sekundenschnelle lange Textabschnitte in andere Sprachen übertragen – nicht fehlerfrei, aber doch in einer so guten Qualität, wie sie noch vor ein, zwei Jahren unvorstellbar gewesen wäre. Und das System Watson von IBM kann natürlich-sprachige Texte analysieren und Inhalte herausfiltern und zusammenfassen – ob sie nun aus Wikipedia stammen oder aus medizinischer Fachliteratur, aus Börsen- und Unternehmensnachrichten oder aus Berichten von Autowerkstätten.

In Kliniken, Banken und Unternehmen bereiten die ersten dieser neuen KI-Systeme bereits Daten auf und geben Ärzten, Finanzberatern und Managern Empfehlungen für Diagnosen, Geldanlagen oder die Optimierung von Industrieprozessen. Das Smartphone war offensichtlich nur der Anfang der Ära der smarten Maschinen. Neben Smart Health, Smart Finance und Smart Factory heißen die Schlagworte der Zukunft Smart Grids, Smart Cars, Smart Buildings und Smart Cities.⁰² Intelligente Stromnetze, die Smart Grids, werden Energieangebot und -nachfrage in Einklang bringen und dadurch nachhaltige Energiesysteme mit all ihren erneuerbaren Energien, Speichern und Verbrauchern erst möglich machen. Schon bald werden wir auf Autobahnen unsere Fahrzeuge – die Smart Cars – auf Autopilot schalten, bis etwa 2030 auch auf Landstraßen und in Städten. In Fabriken werden wir Hand in Hand mit Robotern arbeiten, in Hotels, Museen und Geschäften werden uns Maschinen Auskunft geben und uns bedienen. Zu Hause und in Seniorenheimen werden sie Getränke bringen, mit alten Menschen Spiele spielen, putzen, die Wäsche wegtragen, vielleicht sogar kochen. Gebäude und Städte werden voller Sensoren sein, die Energieverbrauch und Raum-

belegung, Verkehrsströme und Emissionen messen und Computern bei der Optimierung helfen.

Kein Zweifel, wir werden die smarten Maschinen brauchen: die autonomen Fahrzeuge und das intelligente Zuhause als Hilfen für die immer älter werdende Bevölkerung, die Smart Grids und das Internet der Energie für nachhaltige Energiesysteme, die Smart Factory für eine flexible und wettbewerbsfähige Industrie und die Smart Cities für die lebenswerten Städte von morgen. Schon bald werden wir in einer Gemeinschaft von Menschen und smarten Maschinen leben – so selbstverständlich, wie wir heute Smartphones nutzen.

60 JAHRE AUF UND AB IN DER KI-FORSCHUNG

Doch was sind die Treiber dieser Entwicklung? Wie ist es möglich, dass sich das Gebiet der Künstlichen Intelligenz in den vergangenen Jahren so explosionsartig entwickelt hat, und was ist an technischem Fortschritt in den nächsten Jahrzehnten noch vorstellbar? Könnten uns Maschinen dereinst sogar in allen Belangen übertreffen – wie es der Astrophysiker Stephen Hawking oder der Tesla-Gründer Elon Musk befürchten? Eindringlich warnen sie vor einer „Superintelligenz“, die uns vielleicht genauso behandeln würde, wie wir mit lästigen Mücken umgehen. „Künstliche Intelligenz könnte die letzte Erfindung sein, die die Menschheit machen wird“, meint der in Oxford arbeitende Philosoph Nick Bostrom.⁰³ Wie realistisch ist so etwas?

Um diese Fragen beantworten zu können, hilft zunächst ein Blick in die Vergangenheit. Der Begriff „Künstliche Intelligenz“ stammt aus dem Jahr 1956, als der US-Wissenschaftler John McCarthy eine Konferenz in New Hampshire so betitelte.⁰⁴ Auf dieser Tagung diskutierten Forscher erstmals über Computer, die Aufgaben lösen sollten, die über das reine Rechnen hinausgingen, etwa Texte analysieren, Sprachen übersetzen oder Spiele spielen. So hatte der Elektroingenieur Arthur Samuel für einen IBM-Großrechner ein Programm

für das Brettspiel Dame geschrieben. Am Anfang kannte diese Software nur die erlaubten Züge des Spiels, und so verlor sie stets gegen Samuel. Doch dieser ließ ein weiteres Programm mitlaufen, das – entsprechend den Strategien, die er selbst kannte – bei jedem Zug die Gewinnwahrscheinlichkeit für die aktuelle Aufstellung auf dem Brett bewertete. Zugleich hatte Samuel eine geniale Idee: Er ließ den Computer gegen sich selbst spielen und herausfinden, ob diese Wahrscheinlichkeiten geändert werden sollten. Spiel für Spiel, immer wieder. Dabei lernte der Computer hinzu und verbesserte die Genauigkeit seiner Vorhersagen. Was dann passierte, scheint heute eine Selbstverständlichkeit, war 1956 aber eine Sensation: Der Computer wurde ein so guter Dame-Spieler, dass Samuel keine Chance mehr gegen ihn hatte. Ein Mensch hatte erstmals einer Maschine etwas beigebracht, bei dem sie durch stetiges Lernen besser wurde als ihr eigener Lehrer.

Nach demselben Prinzip entwarfen die Forscher der Google-Tochter DeepMind 2017 ihr Programm AlphaGo Zero.⁰⁵ Sie gaben ihm nur die Regeln des Go-Spiels vor und ließen es dann ständig gegen sich selbst spielen. Binnen drei Tagen erreichte AlphaGo Zero vom einfachsten Anfängerniveau die Spielstärke eines Profis und übertraf bereits die Version, die 2016 gegen den menschlichen Weltmeister Lee Sedol mit 4:1 gewonnen hatte. Nach drei Wochen verfügte AlphaGo Zero dann schon über eine Spielstärke, die noch nie ein Mensch in diesem jahrtausendealten Spiel erreicht hatte – und das, ohne dass das Programm jemals Spiele von Menschen studiert hätte.

Besonders bemerkenswert dabei ist, dass Go nahezu unendliche Spielvarianten erlaubt. Die Anzahl der möglichen Züge auf einem Go-Brett übertrifft die Zahl der Atome im Universum bei Weitem. Daher kann die Software unmöglich alle denkbaren Spielzüge durchrechnen. Bei Schach war das in gewissem Maße noch machbar: 1997 besiegte der IBM-Rechner Deep Blue den damaligen Weltmeister Garri Kasparow unter Turnierbedingungen – im Wesentlichen aufgrund seiner Rechenleistung, mit der er pro Sekunde 200 Millionen Schachstellungen analysieren konnte. AlphaGo Zero hingegen muss Stellungen sozusagen „intuitiv“ bewerten, allein aufgrund seiner gelernten Erfahrung aus bisher gespielten Partien und der Ähnlichkeiten von Stellungen.

01 Vgl. Ulrich Eberl, Wo Künstliche Intelligenz den Menschen schon übertrifft, 11. 10. 2016, www.zeit.de/digital/internet/2016-10/deep-learning-ki-besser-als-menschen.

02 Vgl. Ulrich Eberl, Smarte Maschinen – wie Künstliche Intelligenz unser Leben verändert, München 2016, S. 201 ff.

03 Nick Bostrom, Superintelligenz: Szenarien einer kommenden Revolution, Berlin 2014, S. 131 ff.

04 Vgl. Stuart Russell/Peter Norvig, Künstliche Intelligenz, München 2012, S. 40 ff.

05 Vgl. David Silver et al., Mastering the Game of Go without Human Knowledge, in: Nature 550/2017, S. 354–359.

Dieses Erlernen von Mustern und der Umgang mit ihnen stehen im Zentrum des Qualitätssprungs der neuen smarten Maschinen. Es geht nicht mehr nur um pure Rechengeschwindigkeit: Hier haben Computer schon lange die Menschen hinter sich gelassen, und auch auf dem Gebiet der Logik sind Algorithmen kaum zu schlagen. So entstanden schon vor 60 Jahren Computerprogramme, die eigenständig logische Theoreme mathematisch beweisen konnten. Über Jahrzehnte versuchten Forscher daher, mit sogenannten Expertensystemen alle Probleme zu bewältigen, etwa durch die Anwendung von Wenn-Dann-Regeln im Sinne von: „Wenn die Nase läuft und der Patient Halsweh hat und hustet, aber nur geringes Fieber hat, dann ist die Wahrscheinlichkeit größer, dass es sich um eine simple Erkältung handelt und nicht um eine gefährliche Virusgrippe.“

Doch Rückschläge blieben nicht aus, denn oft sind solche regelbasierten Systeme gar nicht einsetzbar – vor allem bei Alltagsaufgaben, die wir Menschen scheinbar mühelos beherrschen: etwa beim Verstehen von Sprache oder beim Erkennen von Bildinhalten. Wenn man einem Computer beibringen will, einen Baum zu erkennen, genügt es nicht, ihm zu beschreiben, wie ein Stamm oder Äste aussehen. Denn auch ein Strommast hat so etwas wie einen Stamm und Äste – und im Winter verlieren viele Bäume ihr Laub, sodass sich auch Blätter nicht als Unterscheidungsmerkmal eignen. Es lassen sich zahllose solcher Fälle finden, in denen man mit vorgegebenen Regeln nicht weiter kommt. In den 1970er Jahren wandten sich daher viele Forscher frustriert vom Gebiet der Künstlichen Intelligenz wieder ab. Finanzierungsprogramme wurden gekürzt oder ganz gestrichen – im Rückblick wird dies als „Winter der Künstlichen Intelligenz“ bezeichnet.

VORBILD GEHIRN

Seither gibt es immer wieder ein Auf und Ab, es entstehen Hypes und verschwinden wieder, doch seit Mitte der 1980er Jahre das revolutionär neue Konzept der Neuronalen Netze seinen Aufschwung nahm – mit einem weiteren Boom in den vergangenen Jahren –, wächst auch die Zahl der kommerziellen Erfolgsgeschichten. Ein Neuronales Netz orientiert sich, vereinfacht ausgedrückt, an der Funktionsweise der Nervenzellen, der Neuronen, im Gehirn. In ihm sind mehrere Schichten künstlicher Neuronen auf komplexe Weise miteinander verbunden, um Informationen zu verarbeiten. Da

die Stärken dieser Verbindungen variieren können und auch Rückkopplungen möglich sind, sind diese Netze lernfähig. Das Prinzip dahinter ist recht einfach: Wird eine Verbindung immer wieder benutzt, steigt ihre Verbindungsstärke und damit ihre Bedeutung – im Gehirn ist das genauso. Wenn wir oft genug gelernt haben, dass eine rote Ampel „Halt! Gefahr!“ bedeutet, dann ist diese Assoziation sofort da, wo immer wir eine rote Ampel sehen.

Insbesondere eignen sich solche Neuronalen Netze dazu, Muster zu erkennen, ohne dass ihnen vom Menschen einprogrammiert werden muss, an welchen exakten Eigenschaften der Muster sie dies festmachen sollen. Präsentiert man ihnen beispielsweise in einer Trainingsphase unzählige Fotos von Bäumen, Katzen oder Autos, können sie anschließend auch auf unbekanntem Bildern sofort Bäume, Katzen oder Autos identifizieren. Ebenso kann man sie mit gesprochenen Worten oder Schriftzeichen trainieren, und sie können anschließend Sprachbefehle oder Handschriften erkennen. Was die heutigen, sogenannten Deep-Learning-Systeme von den Neuronalen Netzen der 1980er Jahre unterscheidet, ist vor allem ihre Mächtigkeit: Waren damals nur einige Tausend künstlicher Neuronen in wenigen Schichten miteinander verbunden, so sind es bei den leistungsfähigsten Systemen von heute Milliarden von Neuronen, die in bis zu 30 Schichten gestapelt sind. Möglich machte diesen Fortschritt die enorme Steigerung der Rechenleistung und Speicherfähigkeit von Computern.⁰⁶

Die stärksten Supercomputer konnten Mitte der 1990er Jahre etwa 100 Milliarden Rechenoperationen pro Sekunde bewältigen – das schafft heute jedes gute Smartphone. Wir alle tragen also sozusagen einen Supercomputer von 1995 in unseren Jackentaschen. Zugleich sanken die Kosten um den Faktor 10000, und das heutige Smartphone braucht nur ein 10000-stel bis ein 100000-stel der elektrischen Leistung der damaligen Superrechner. Außerdem werden auch Kameras und Sensoren – vom Satellitenempfänger bis zu Beschleunigungs-, Radar-, Wärme- und Tastsensoren – immer kleiner und kostengünstiger. Und die Datenexplosion im Internet bietet den Maschinen eine nahezu unbegrenzte Vielfalt an Lernbeispielen. Derzeit werden jeden Tag von Menschen und Maschinen zehnmals mehr neue Daten produziert, als in allen Büchern der Welt enthalten sind. All die Milliarden von Bildern, Texten, Videos und Audiodateien lassen sich

⁰⁶ Vgl. Eberl (Anm. 2), S. 47 ff.

als perfektes Trainingsmaterial für smarte Maschinen nutzen. Dadurch lernen sie immer besser sehen, lesen und sprechen. Mit jeder Suchanfrage, mit jeder Spracheingabe, mit jedem Übersetzungswunsch lernen sie hinzu.

VERTAUSENDFACHUNG DER LEISTUNGSSTÄRKE BIS 2040

Diese Entwicklung ist noch lange nicht am Ende. So zeigt der Blick in die Labore von Halbleiterherstellern: Auch wenn die Verkleinerung der Strukturen bald an eine physikalische Grenze stoßen wird, gibt es dennoch viele Ideen – etwa das Übereinanderstapeln von Nanostrukturen und wesentlich energieeffizientere Bauteile –, die eine weitere Steigerung um den Faktor 1000 bis zum Jahr 2040 erwarten lassen. Mikrochips könnten dann noch einmal um das Tausendfache schneller rechnen, tausendmal mehr Daten speichern und tausendmal mehr Daten pro Sekunde übertragen als heute – und das zum selben Preis. Oder anders ausgedrückt: Wenn heute ein Notebook 500 Euro kostet, werden wir 2040 dieselbe Leistung auf einem kleinen Chip für 50 Cent bekommen. Visionäre wie Ray Kurzweil sagen, dass wir heute für 1000 Dollar etwa die Leistungsfähigkeit des Gehirns einer Maus mit rund 100 Millionen Nervenzellen kaufen können – bei einer Vertausendfachung wären wir 2040 dann beim Komplexitätsgrad des Gehirns eines Menschen angelangt.⁰⁷

Und selbst das könnte noch übertroffen werden. Denn Forscher entwickeln derzeit sogenannte neuromorphe Chips, die das Verhalten von Nervenzellen nicht per Software, sondern elektronisch nachbilden: Deren Lernvorgänge sind bereits heute zehntausendfach schneller als beim menschlichen Gehirn und millionenfach schneller als bei Supercomputern.⁰⁸ Wollte man mit heutigen Superrechnern die neuronalen Prozesse eines einzigen biologischen Tages nachbilden, bräuchte man dafür Jahre – Neurochips schaffen das in zehn Sekunden, allerdings bisher nur in Netzwerken mit etwa einer Million Nervenzellen, noch nicht mit Milliarden. Doch die Forschung steht hier erst am Anfang.

Eine Einschränkung gibt es allerdings: den Energieverbrauch. Das menschliche Gehirn begnügt sich mit 20 Watt, ein vergleichbarer Super-

computer hingegen benötigt etwa die elektrische Leistung einer Stadt mit 20 000 Einwohnern. Die besten Neurochips liegen in der Mitte dazwischen, aber immer noch bei mindestens dem Tausendfachen des menschlichen Gehirns. Außerdem funktioniert die Lernfähigkeit des Gehirns ohne Software, ohne zentrale Steuerung und ohne Betriebssystem, und das System ist extrem fehlertolerant: Obwohl jeden Tag etwa 100 000 Neuronen verloren gehen, lassen seine kognitiven Fähigkeiten über Jahrzehnte hinweg kaum nach. Das Gehirn kann also mit verlorenen Ressourcen ebenso gut umgehen wie mit unpräzisen Informationen. Und sein geringer Energieverbrauch ist zu einem großen Teil darauf zurückzuführen, dass nicht wie im konventionellen Computer ständig Daten zwischen Speicher und Prozessor hin- und geschoben werden müssen – das spart nicht nur Zeit, sondern auch Energie.

LERNEN WIE KLEINE KINDER

Die Hardware wird den Forschern bei der Entwicklung smarterer Maschinen eher wenige Beschränkungen auferlegen, doch wie sieht es mit der Software und der Effizienz und Effektivität der Informationsverarbeitung aus? Hier scheinen noch lange nicht die besten Lösungen gefunden worden zu sein: Während etwa Deep-Learning-Systeme Zigtausende bis Millionen von Katzen sehen müssen, um danach eine Katze zuverlässig zu erkennen, reichen kleinen Kindern ein paar Dutzend Lernbeispiele, um auch einen gestiefelten Kater oder den König der Löwen als Katze einzustufen. In diesem Sinne erreichen Kinder ihre Lernziele wesentlich wirkungsvoller und wirtschaftlicher als Computer. Zudem sind Deep-Learning-Systeme nur Meister im Vergleich von Mustern, nicht mehr. Wenn sie etwa auf Tierbilder trainiert wurden, finden sie überall Tiere, auch in Wolken oder dem Rauschen eines Bildschirms – was dann wie Halluzinationen von Computern wirkt. Ihnen fehlen völlig das Hintergrundwissen und das Verständnis für Zusammenhänge.

Mehr noch: Wenn man die Frage beantworten will, wie intelligent Maschinen werden können, muss man erst einmal klären, von welcher Intelligenz die Rede sein soll. Denn Fachleute sprechen von mathematischer, räumlicher, sprachlicher, logischer, emotionaler oder sozialer Intelligenz – unsere Intelligenz ist nicht nur das, was der IQ misst. So kann ein Neuronales Netz zwar Objekte aller

⁰⁷ Ray Kurzweil, *How to Create a Mind*, New York 2012, S. 257f.

⁰⁸ Vgl. Eberl (Anm. 2), S. 145ff.

Art erkennen, aber es weiß nichts über deren Bedeutung für den menschlichen Alltag. Außerdem gilt nach wie vor der alte Spruch „Computern fällt leicht, was Menschen schwerfällt – und umgekehrt“ nicht nur für Computer, sondern auch für Roboter. Türen öffnen und Bälle fangen, laufen und Hindernissen ausweichen, das gehört alles zu den leichtesten Aufgaben, die man einem körperlich gesunden Menschen stellen kann, aber gleichzeitig zu den schwierigsten Aufgaben für Roboter.

Auch Menschen müssen ihre Fähigkeiten erst nach und nach erwerben. In den ersten beiden Lebensjahren entsteht zunächst die sensomotorische Intelligenz: Babys lernen krabbeln, stehen, laufen, nach Dingen greifen und ihre Bewegungsabläufe koordinieren. In den Jahren danach entwickeln sich sowohl das Sprechvermögen wie die symbolische Vorstellungskraft und die Fähigkeit zur Empathie. Zugleich lernen Kinder immer besser, vorzudenken und ihr Handeln zu planen, doch erst mit elf oder zwölf Jahren sind Jugendliche in der Lage, Probleme systematisch zu analysieren, Hypothesen und kreative Lösungen zu entwickeln und über sich selbst nachzudenken.

Einen ähnlichen Weg gehen Forscher nun mit Maschinen. Sensomotorische Intelligenz haben die besten Roboter schon entwickelt: Sie können einigermaßen sicher stehen, laufen und Dinge aller Art greifen. Der vierbeinige Roboter Cheetah von Boston Dynamics rennt schneller als Usain Bolt, der schnellste Mensch über die 100- und 200-Meter-Distanz – und es gibt bereits feinfühligere Roboter, die weiche Erdbeeren pflücken, ohne Druckstellen zu hinterlassen. Auch das Lernen durch Beobachten und Nachahmen, das kleine Kinder so gerne einsetzen, bringt man nun Robotern bei. Beispielsweise sollen die gerade auf den Markt kommenden „kollaborativen Roboter“ lernen, mit Menschen Hand in Hand zu arbeiten. Eine herkömmliche Programmierung ist nicht mehr nötig. Stattdessen führt man einfach die Arme und Greifer solcher Roboter und zeigt ihnen, wie sie Knöpfe drücken oder Bauteile montieren sollen. Diese Maschinen sind so sensibel, dass sie in Bruchteilen von Sekunden eine Bewegung stoppen, wenn ihnen ihre Sensoren mitteilen, dass sie andernfalls einen Menschen verletzen könnten.⁰⁹

09 Für einen solchen teamfähigen Roboter haben Forscher der FRANKA EMIKA GmbH und der Technischen Universität München im November 2017 den Deutschen Zukunftspreis erhalten, www.deutscher-zukunftspreis.de/de/nominierte/2017/team-2.

WIE BELOHNT MAN MASCHINEN?

Und selbst das Belohnungslernen findet Eingang in die Welt der smarten Maschinen: Natürlich bekommen sie nicht wie Kinder Schokolade oder gute Noten, sondern ihnen genügt ein internes Punktekonto, das aufgefüllt wird, wenn sie etwas richtig gemacht haben, oder ein Schulterklopfen oder Lächeln, das sie mithilfe ihrer Kameras und Sensoren wahrnehmen und als Lob werten. So lernt etwa der kleine Roboter iCub am italienischen Institut für Technologie in Genua wie ein menschliches Kind im Kindergarten.¹⁰ Er betrachtet seine Spielsachen, dreht sie hin und her – „begriff“ sie im wahrsten Sinne des Wortes –, fragt seine menschlichen Lehrer, wie die Dinge heißen, und lernt, wie man den Tisch ab- und das Zimmer aufräumt. Genauso lernt er Klavier spielen oder wie man einen Spielzeug-Bogen spannt und hält, um mit dem Pfeil die Zielscheibe zu treffen. Auch im japanischen Osaka arbeiten Forscher mit dem wissbegierigen iCub-Roboter. Belohnungen gibt es immer dann, wenn er etwas Neues gelernt oder eine Aufgabe gut gelöst hat. Diese Art des Lernens ist vielleicht die vielversprechendste Methode, um Roboter alltagstauglich zu machen: Sie müssen Menschen im Alltag begleiten und sich einprägen, was sie tun sollen, um sich nützlich zu machen.

Entscheidend ist dabei – wie bei Menschen auch – die Qualität der Lehrer. Ein misslungenes Beispiel war der Chatbot Tay, der im Frühjahr 2016 lernen sollte, wie sich Menschen im Internet unterhalten.¹¹ Binnen 24 Stunden musste Microsoft ihn wieder vom Netz nehmen, weil er zum Rassisten geworden war, der den Holocaust leugnete und Hitler lobte. Dieses Programm hatte ganz offensichtlich von den falschen Leuten gelernt. Wie man zuverlässige und sich ethisch korrekt verhaltende, selbstlernende Maschinen baut, wird sicherlich in Zukunft eine wichtige Aufgabe sein und neue Berufszweige eröffnen. Die ersten Lehrstühle für Maschinenethik existieren bereits.¹²

Das Ziel der Forscher ist klar: Roboter und smarte Maschinen aller Art sollen einmal in der

10 Vgl. Eberl (Anm. 2), S. 84ff., S. 290ff. Videos zum iCub-Roboter finden sich unter www.youtube.com/user/robotcub.

11 Vgl. Bernd Graff, *Rassistischer Chat-Roboter – mit falschen Werten bombardiert*, 3.4.2016, sz.de/1.2928421.

12 Siehe dazu auch die Beiträge von Catrin Misselhorn und Oliver Bendel in dieser Ausgabe (Anm. d. Red.).

Lage sein, Menschen auch in komplexen, sich ständig ändernden Umgebungen zu helfen – wie perfekte Butler, ob beim Aufräumen oder Putzen zu Hause, beim Kochen oder Einkaufen oder beim Autofahren im Stadtverkehr. Dass sie dafür noch sehr viel hinzu lernen müssen, ist klar, doch einen wesentlichen Vorteil haben sie: Was eine Maschine einmal gelernt hat, kann sie im Prinzip in Zukunft in ein RoboNet hochladen und anderen Maschinen ähnlichen Bautyps zur Verfügung stellen – egal, ob es darum geht, wie man Fenster putzt, einen Dinnertisch deckt oder einen Hub-schrauber fliegt. Menschen hingegen müssen alles individuell lernen und können sich neue Fähigkeiten nicht einfach wie Apps herunterladen.

MASCHINEN MIT ALLTAGSWISSEN, EMOTIONEN, SOZIALKOMPETENZ?

Auf lange Sicht können Maschinen daher vielfältigste, intelligente Verhaltensweisen erwerben, doch ein paar Beschränkungen werden für sie nur schwer zu überwinden sein: Wir Menschen besitzen beispielsweise viel implizites Wissen, Alltagskompetenz und gesunden Menschenverstand, um in unserer Umgebung zurechtzukommen.¹³ Nehmen wir nur den Satz „Die Beamten haben den Demonstranten verboten, sich zu versammeln, weil sie Gewalt befürworteten“. Ein Mensch erkennt sofort, worauf sich das „sie“ bezieht. Doch ein Computer tut sich hier sehr schwer, ebenso wie ein autonomes Fahrzeug wohl bremsen würde, wenn der Wind eine große Papiertüte auf die Straße weht – ein Mensch würde einfach weiterfahren.

Reines Faktenwissen lässt sich allerdings heute schon in Maschinen implementieren, wie IBM mit seinem „kognitiven“ Computersystem Watson zeigt.¹⁴ Bereits 2011 hatte Watson die zwei weltbesten menschlichen Champions in der US-Quizshow Jeopardy besiegt, die mit Wortspielen, Rätseln und subtilen Anspielungen arbeitet. Entscheidend für den Erfolg war, dass Watson unter anderem das Internetlexikon Wikipedia, insgesamt 200 Millionen Textseiten, in seinem Arbeitsspeicher hatte. Er war in der Lage, diese Texte zu lesen, passende Informationen zu erkennen und

zu kombinieren, und binnen zwei bis drei Sekunden eine – meist korrekte – Antwort auf die Quizfrage zu liefern.

Seither hat IBM das Watson-System für viele Anwendungen ausgebaut: Es berät Ärzte bei Krebsdiagnosen ebenso wie Banker, Juristen und Steuerberater oder Firmen bei der Datenauswertung in Fabriken, Gebäuden oder Verkehrs- und Energiesystemen. Ein wesentlicher Vorteil ist, dass das System mit natürlich-sprachigen und unstrukturierten Texten zurechtkommt – ob Call-Center-Protokolle, Arztbriefe, Fachartikel oder Twitter-Feeds – und dass es auf Nachfrage die Quellen für seine Bewertungen nennen kann. Dadurch wissen die Anwender stets, auf welcher Basis Watson seine Hypothesen und Empfehlungen abgeleitet hat – nicht nur für Ärzte ist dieser Dialog mit dem digitalen Assistenten entscheidend, um ihn überhaupt einsetzen zu können. Bei einem konventionellen Neuronalen Netz kann der Nutzer hingegen nicht ohne Weiteres nachvollziehen, welche Lernbeispiele für das Ergebnis letztlich relevant waren.

Sogar das Abwägen von Argumenten kann man Computern beibringen. So wurde Watson etwa die Frage gestellt, ob Video-Gewaltspiele für Jugendliche verboten werden sollten. Daraufhin durchsuchte der Computer vier Millionen Wikipedia-Artikel, fand die zehn relevantesten und destillierte daraus binnen Sekunden drei Argumente für und drei gegen ein solches Verbot. Dass ein solches Pro-und-Contra-Programm für Bankberater in Zukunft ebenso sinnvoll sein kann wie für Juristen, Versicherungsvertreter oder Ärzte, ist offensichtlich.

Doch ein Großteil unseres menschlichen Alltagswissens steht eben nicht in Wikipedia. Gesunden Menschenverstand findet man nicht im Lexikon. Maschinen können daher ein solches Alltagswissen nur erwerben, wenn sie über lange Zeit hinweg Menschen beobachten und ständig hinzulernen. Dies ist zweifellos ein langwieriges Unterfangen, und wie erfolgreich es sein wird, muss sich noch zeigen. Ähnliches gilt für das Feld der kreativen Maschinen: Computerprogramme können heute zwar Stile kopieren und neu kombinieren, also malen wie van Gogh oder Rembrandt und komponieren wie Bach oder Mozart, aber etwas völlig Neues hat bisher keine Maschine hervorgebracht.

Ein weiteres Problem ist sogar noch viel fundamentalerer Art: Maschinen, wie ausgeklügelt sie auch sein mögen, haben keinen biologischen

¹³ Vgl. Ulrich Eberl, *Smarte Maschinen – wie intelligent können sie sein?*, 17.3.2017, <https://zukunft2050.wordpress.com/2017/03/17/wie-intelligent-koennen-smarte-maschinen-sein-und-woran-werden-sie-scheitern>.

¹⁴ Vgl. ders. (Anm. 2), S. 124ff.

Körper wie Menschen. Sie werden daher nie alle Erfahrungen mit Menschen teilen können: Sie müssen nicht essen und trinken, schlafen und träumen, sie wachsen nicht und bekommen keine Kinder und sie kennen den Sturm der Gefühle nicht, der Menschen ergreifen kann. Daher sei die Vorhersage gewagt: Selbst wenn smarte Maschinen Emotionen aus Gesten und Mimik lesen und wenn sie so tun, als ob sie Gefühle hätten, eine den Menschen vergleichbare emotionale und soziale Intelligenz wird ihnen verwehrt bleiben. Aus all diesen Gründen gehört eine Superintelligenz, die uns Menschen auf allen Gebieten überflügelt, wohl eher in den Bereich der Science-Fiction als zu den realen Gefahren.

WER MACHT DIE ARBEIT VON MORGEN?

Viel mehr Sorgen müssen wir uns allerdings um zwei andere Entwicklungen machen: um autonome Kampfroboter und um die Auswirkungen auf Arbeitsplätze. Das erste Problem lässt sich nur lösen durch eine weltweite Ächtung dieser Maschinen, wie es bei Biowaffen oder Atombomben im Weltall gelungen ist. Erste internationale Anstrengungen in diese Richtung gibt es bereits, aber sie müssen intensiviert werden. Auch dass sich durch den Einsatz von smarten Maschinen mit Künstlicher Intelligenz praktisch alle Jobs – egal in welcher Branche – erheblich verändern werden, ist offensichtlich.¹⁵ Vor allem Routinetätigkeiten in den Büros, bei denen es um die Beschaffung und Verarbeitung von Informationen geht, können künftig durch Maschinen übernommen werden: Das betrifft den Buchhalter ebenso wie den Steuerberater, den Logistiker oder Finanzanalysten. Ähnliches gilt für Putzkräfte, Lagerarbeiter oder Lkw-, Bus- und Taxifahrer.

Wenig betroffen sind hingegen kreative Jobs wie Forscher, Designer und Künstler sowie Berufe, die eine hohe Sozialkompetenz erfordern, wie Pflegekräfte, Lehrer und Manager. Zudem werden sich zwar viele Jobbeschreibungen verändern, aber nicht unbedingt die Arbeitsplätze wegfallen. So werden in Zukunft Ärzte die Hilfe von Computerassistenten in Anspruch nehmen, aber sie

werden nicht durch Roboter ersetzt – allein schon deshalb, weil oft die Intuition der Ärzte und der Kontakt mit den Patienten der halbe Weg zur Heilung sind. Kurz gesagt: Die einfacheren Arbeiten machen Maschinen, die komplexeren die Menschen, die weiterhin als Lenker und Denker gebraucht werden, als Planer und Entscheider, als kreative Problem- und Konfliktlöser, als diejenigen, die Qualität und Sicherheit gewährleisten, und als die entscheidenden Partner, die emotionale und soziale Intelligenz gegenüber ihren Kunden und Zulieferern beweisen müssen.

Hinzu kommt, dass auch eine Menge neuer Berufe entstehen. Die smarten Maschinen müssen auch erst einmal entworfen und konstruiert werden, es muss sichergestellt werden, dass sie gefahrlos und zuverlässig betrieben werden können, und sie müssen trainiert und auf die Einsatzfelder optimal angepasst werden. Der Blick in die Vergangenheit bestätigt, dass neue Technologien immer auch neue Berufe mit sich bringen: Anfang der 1980er Jahre, als die Computer massentauglich wurden, gab es noch so gut wie keine Software-Entwickler – heute sind es weltweit über 20 Millionen.

Mein Fazit lautet daher: Smarte Maschinen mit Künstlicher Intelligenz sind zweifellos eine der größten technisch-wirtschaftlich-sozialen Herausforderungen, vor denen die Menschheit derzeit steht. Aber sie sind auch eine Chance für all die globalen Aufgaben, die wir bewältigen müssen: ob im Kampf gegen den Klimawandel und beim Umbau der Energiesysteme, ob bei der Gestaltung lebenswerter Städte oder bei der Unterstützung der wachsenden Zahl pflegebedürftiger Menschen. Wenn wir es richtig machen, werden uns die smarten Maschinen weit mehr nützen als schaden.

ULRICH EBERL

ist promovierter Physiker, Wissenschafts- und Technikjournalist sowie Buchautor. Er ist Geschäftsführer des Redaktionsbüros SciPress, war bis 2015 Leiter der Innovationskommunikation von Siemens und gründete das Zukunftsmagazin „Pictures of the Future“. Zuletzt erschienen: „Smarte Maschinen – wie Künstliche Intelligenz unser Leben verändert“ (Hanser, 2016).

ulrich.eberl@scipress.de

www.zukunft2050.wordpress.com

¹⁵ Vgl. Ulrich Eberl, *Schöne neue Arbeitswelt*, in: *Bild der Wissenschaft* 7/2016, S. 94f. Siehe dazu auch die Ausgabe 18–19/2016 der APuZ „Arbeit und Digitalisierung“, www.bpb.de/apuz/225683/arbeit-und-digitalisierung (Anm. d. Red.).

MENSCH FRAGT, MASCHINE ANTWORTET

Wie Künstliche Intelligenz Wirtschaft, Arbeit und unser Leben verändert

Thomas Ramge

„Alexa, erzähle mir einen Zungenbrecher“. Alexa muss nicht lange überlegen und sagt: „Brautkleid bleibt Brautkleid und Blaukraut bleibt Blaukraut.“ Die leicht blecherne Frauenstimme in Amazons zylindrischem Lautsprecher verspricht sich natürlich nicht dabei. Alexa – genauer das datenreiche System in der Amazon-Cloud hinter der Produktfamilie Echo – kann sehr viele sehr flache Witze erzählen. „Sagt ein Ballon zum anderen: Ich habe Platzangst.“ Der interaktive Lautsprecher singt auf Befehl eines Menschen auch gerne Weihnachtslieder. Über die Gag-Funktionen von Amazon Echo wurde seit Einführung des Produkts 2015 viel gelacht und viel gespottet, je nach persönlicher Humorpräferenz. Über die weltverbessernde Relevanz von Flachwitzen abrufbar durch Sprachbefehle lässt sich in der Tat streiten. Doch vor lauter Debatte über die spielerischen Funktionen des Systems, wurde oft übersehen: Amazon Echo ist kein Spielzeug, sondern ein technischer Durchbruch auf dem Weg zu einem intelligenten Alltagsassistenten.

Mit Amazon Echo können Nutzer auf der Couch liegend per Sprachbefehl die Heizung höher stellen, das Licht dimmen und Alexa bitten, nach einer Netflix-Serie zu suchen, die ähnlich wie *Narcos* ist, aber nicht so brutal. Vor dem Kleiderschrank können sie das System rasch fragen, wie das Wetter wird, und in der Küche, mit beiden Händen im Kuchenteig, die Anweisung geben, frische Eier auf die Einkaufsliste zu setzen. Alexa liest die Nachrichten vor oder informiert, wenn die Lieblingsmannschaft ein Tor geschossen hat. US-amerikanische Kunden können sich auch den Kontostand zurufen lassen oder eine Pizza bei Domino's bestellen. Das Sortiment des Amazon-Shops ist weltweit zugänglich, mit den bekannten Empfehlungsroutinen, aber es

greift zu kurz, in Alexa eine reine Abverkaufsmaschine zu sehen. Auch lexikalisches oder zeitgenössisches Wissen lässt sich dialogisch abrufen. Das System führt dabei Informationen aus verschiedenen Online-Quellen wie Wikipedia oder News-Webseiten zusammen und versucht diese, in den gewünschten Sinnzusammenhang zu rücken.

VIRTUELLE ASSISTENTEN

Der Fachbegriff für Systeme wie Alexa ist *virtual assistant*. Oft werden sie auch nur kurz *bots* genannt. Die digitalen Technologie-Riesen in den USA und Asien liefern sich seit einigen Jahren einen harten Kampf um die Vorherrschaft bei sprachgesteuerten virtuellen Assistenten. Sie bauen riesige Teams aus Datenwissenschaftlern und Experten für Maschinelles Lernen⁰¹ auf, übernehmen KI-Startups, wie jüngst Samsung den kalifornischen Shootingstar unter den virtuellen Assistenten Viv, oder schmieden überraschende Allianzen wie Microsoft und Amazon, die ihre digitalen Helfer künftig zusammen im Dienste des Nutzers arbeiten lassen. Diesen Aufwand betreiben diese Unternehmen nicht aus purer Freude am technischen Fortschritt, sondern auch aus Angst um die unternehmerische Existenz. Den Strategen bei Apple (mit Siri), Google (mit Google Assistant), Microsoft (mit Cortana), Facebook (mit M) und Samsung (mit Bixby) ist heute klar, dass in Zukunft der Zugang zu vielen, vermutlich den meisten digitalen Diensten erfolgen wird wie im Raumschiff Enterprise: Mensch fragt, Maschine antwortet. Wenn die Maschine das Frage-Antwort-Spiel nicht beherrscht, sucht sich der Mensch einen anderen Anbieter.

Nutzer erwarten dabei zum einen immer präzisere Antworten auf immer komplexere Problemstellungen. „Ok, Google. Ich will im März für drei Tage in die Schweiz zum Skilaufen fliegen. Welche Gebiete sind dann noch schneesicher, wo gibt es noch günstige Hotels, wann noch günstige Flüge, und brauche ich einen Mietwagen, um vom Flughafen Zürich ins Skigebiet zu kommen?“ Dazu muss ein virtueller Assistent keinen Turing-Test⁰² bestehen, sondern sauber Informationen recherchieren, aggregieren und gemäß der Vorgaben als Entscheidungsgrundlage aufbereiten. Zudem besteht berechtigte Hoffnung, dass wir eine Reihe nicht ganz so komplizierter, aber lästiger Alltagsentscheidungen nicht mehr selbst treffen müssen, sondern dass wir diese an intelligente Maschinen delegieren können. *Virtual assistants* werden rechtzeitig Druckerpatronen nachbestellen, keine Zahlungsfrist einer Rechnung übersehen, aber auch deutlich öfter merken als der Mensch, wenn die Rechnung zu hoch ist, und die Zahlung verweigern.

Einen Vorgesmack darauf, wie intelligente Agenten künftig nervige Alltagsaufgaben übernehmen, geben Terminkoordinierungsassistenten wie Amy oder Julie. Zielgruppe sind Menschen, die keinen menschlichen persönlichen Assistenten haben. Die Nutzer geben diesen KI-unterstützten Diensten Zugriff auf den Kalender und das E-Mail-Programm. Terminvereinbarung läuft dann wie folgt: Eine Anfrage für ein Treffen kommt per E-Mail. Der Nutzer stimmt grundsätzlich per E-Mail zu und setzt dabei Amy oder Julie in „cc“. Von nun an übernimmt der künstlich intelligente Assistent das übliche E-Mail-Ping-Pong, bis Ort und Zeit mit Geschäftspartnern ausgemacht sind, oder klar ist, wer wen wann unter welcher Nummer anruft. Erweiterte Systeme versprechen zudem, die gesamte Tagesplanung zu übernehmen, Termine zu priorisieren und unter Umständen automatisiert zu verschieben, dem Nutzer in Meetings relevante

Informationen vorzulegen und auf Versäumnisse hinzuweisen. Zumindest Terminkoordinierung funktioniert bereits heute recht gut. Nahezu perfekt klappt es, wenn zwei virtuelle Assistenten im Auftrag ihrer menschlichen Chefs miteinander kooperieren. Computer können nach wie vor am besten mit Computern. Gleichzeitig gilt: Immer mehr Menschen hören auf die Ratschläge von Computern, und dass nicht mehr nur bei eher trivialen Fragen, etwa ob es besser ist, auf der Autobahn den Stau durchzustehen oder die deutlich längere Ausweichstrecke über die Bundesstraße zu nehmen – eine Prognoseanwendung, die besonders Google dank seiner Fülle an Echtzeitdaten aus den Smartphones mit Android-Betriebssystem relativ leicht und weitgehend genau errechnen kann.

DIE VERKAUFSMASCHINE

Es ist freilich kein Zufall, dass Amazon hunderte Millionen Dollar in die Entwicklung von Echo gesteckt hat. Allerdings ist es auch kein Zufall, dass just dieses System so viel Erfolg hat. Seit seiner Gründung 1996 hat es Amazon wie kein zweites Unternehmen verstanden, aus Daten die Bedürfnisse seiner Kunden zu verstehen. Seit Einführung seines personalisierten Empfehlungssystems 1998 leitet das Unternehmen aus diesem Wissen über Kunden immer passendere Schlüsse ab, welches Produkt es einem bestimmten Nutzer zu welchem Zeitpunkt zu welchem Preis anbieten muss, um die Wahrscheinlichkeit zu erhöhen, dass dieser auf den Kaufknopf klickt. Genaue Zahlen, wie gut diese virtuelle Empfehlungsmaschine beim größten Online-Händler der westlichen Welt funktioniert, gibt Amazon nicht bekannt. Experten gehen davon aus, dass rund ein Drittel aller Verkäufe durch Kaufempfehlungen des Systems angestoßen wird. Ein so hoher Wert ist nur möglich, wenn Kunden die Empfehlungen tatsächlich als sinnvollen Ratschlag empfinden, und eben nicht als nervige Online-Werbung, die uns auf unseren Streifzügen im Netz verfolgt und uns Produkte anbietet, die uns kein bisschen interessieren oder die wir sogar bereits gekauft haben. Digitales Marketing mit seiner dümmlichen Penetranz hat in den vergangenen Jahren auf der einen Seite viel verbrannte Erde bei Kunden hinterlassen. Auf der anderen Seite spornt der schlechte Ruf der Online-Werbung innovative Unternehmen

01 Bei Maschinellern Lernen erkennen Computersysteme Muster in Beispielen und können ihre „Erkenntnisse“ auf andere Beispiele übertragen. So lernen sie, aus Daten immer genauere Schlüsse zu ziehen und Entscheidungen abzuleiten.

02 Beim Turing-Test, entwickelt von dem Informatiker Alan Turing, stellt ein Mensch einem anderen Menschen und einem Computer Fragen. Mensch und Rechner antworten im Chat (ohne Sicht- und Hörkontakt). Der Turing-Test ist bestanden, wenn der Fragesteller nicht sagen kann, welcher seiner „Gesprächspartner“ eine Maschine ist.

an, tatsächlich Intelligenz in die virtuelle Kaufberatung zu bringen.

Zu den Vorreitern gehört hier Stitch Fix. Das kalifornische Startup bietet seinen Kunden Mode in einem Abo-Modell, im Fachjargon *curated shopping* genannt. Es verschickt regelmäßig Kisten mit fünf Kleidungsstücken, von denen die Kunden so viele behalten können, wie sie wollen. Das Unternehmen lebt also davon, dass die Klamotten im Paket möglichst genau den Geschmack des Kunden treffen. Jede Rücksendung hingegen verursacht Kosten. Um die Trefferquote zu erhöhen, beschäftigt Stitch Fix über 70 hochbezahlte Datenwissenschaftler, die mit extrem komplexen Algorithmen und neuesten Methoden des Maschinellen Lernens die Prognose zu der Frage verbessern: Behält dieser Kunde dieses Kleidungsstück? Neben naheliegenden Datenquellen wie Fragebögen und der Kaufhistorie – also Feedback-Daten, welche Kleidungsstücke der Kunde in der Vergangenheit behalten oder zurückgeschickt hat – errechnet das System seine Vorschläge aus Instagram-Bildern, die der Kunde geliked hat. Die KI erkennt so mitunter Muster in den Bildern, die er Vorlieben zuordnen kann, die dem Kunden selbst nicht bewusst sind.

US-Warenhäuser wie Macy's und große Supermarktketten wie Tesco in Großbritannien oder Carrefour in Frankreich versuchen hingegen, mit Shopping-Assistenz-Apps die im Online-Handel erprobten Empfehlungsmechanismen in die Welt der physischen Geschäfte zu übertragen. Diese Apps lassen Kunden den schnellsten Weg zum Shampoo-Regal finden, falls Shampoo auf der gespeicherten Einkaufsliste steht oder der Kunde in der Gemüseabteilung per Sprachbefehl danach fragt. Manche Apps weisen ungefragt vor dem Rotwein-Regal darauf hin, dass heute kräftiger Roquefort im Angebot ist. Das Problem all dieser virtuellen Kaufberater ist freilich: Sie werden von Verkäufern zur Verfügung gestellt und stehen grundsätzlich im Verdacht, die Interessen des Verkäufers höher zu gewichten als die des Käufers. Die intelligenten unter den künstlich intelligenten Shopping-Helfern sind deshalb so programmiert, wie ein seriöser Kaufmann, der an einer langfristigen Kundenbeziehung interessiert ist. Sie werden nicht versuchen, Kunden zu Kaufentscheidungen zu verleiten, über die sie sich im Nachhinein mit hoher Wahrscheinlichkeit ärgern.

Wünschenswert wären derweil mehr virtuelle Shopping-Assistenten, die händlerunabhän-

gig beraten. Die Apps von Preissuchmaschinen, die Verbraucher automatisch auf Sonderangebote von Produkten hinweisen, die sie vor einiger Zeit gesucht, aber dann nicht gekauft haben, verfolgen diesen Ansatz. Es gibt noch keinen Bot, der das Konsumverhalten eines Verbrauchers systematisch über alle Produktgruppen beobachtet, aus Kaufentscheidungen Präferenzen und Preisbereitschaft immer besser kennenlernt, der weiß, dass das Klopapier in einer Woche aufgebraucht sein wird und auch noch versteht, welche Routinekäufe er selbsttätig online in Auftrag geben soll, und bei welchen Einkäufen gut aufbereitete Entscheidungsvorlagen für den Menschen gefragt sind – und der im Idealfall sogar noch mit dem Verkäufer über den Preis verhandeln kann. Für Datenschützer wäre ein solcher virtueller Agent der letzte Schritt zum gläsernen Verbraucher, der anfällig für viele Formen der Manipulation ist. Für alle, die ungern Zeit mit Shopping verschwenden, wäre er ein großer Gewinn an Komfort. Wäre ein solcher KI-Berater tatsächlich ein Agent des Käufers und neutral gegenüber Verkäufern, fiel er auch nicht so oft auf dumme Marketingtricks rein – wie wir Menschen.

DER ROBO-ANWALT

Auf dem Feld der künstlich intelligenten Rechtsberatung wächst das Angebot zurzeit rasch, vor allem in englischer Sprache. Der wohl erfolgreichste virtuelle Rechtsassistent der Welt hat den profanen, aber bezeichnenden Namen „DoNotPay“. Der Legal-Bot wurde von dem 19-jährigen Stanford-Studenten Joshua Browder programmiert und unterstützte seine US-amerikanischen und britischen Nutzer zunächst bei Einspruchsverfahren gegen Parkknöllchen, die nach Einschätzung der Parkenden zu Unrecht verhängt wurden. Im Dialog fragt der Chatbot alle relevanten Informationen ab und spuckt nach wenigen Minuten einen individuell begründeten, örtlich angepassten und juristisch wasserdichten Einspruchsbrief aus. Diesen muss der Nutzer dann nur ausdrucken, unterzeichnen und abschicken. Binnen zwei Jahren wehrte der Roboter-Anwalt auf diese Weise rund 375 000 Bußgeldbescheide ab. Derweil erweiterte Browder die Kompetenz des Legal-Bots vom Verkehrsrecht auf viele andere Rechtsfelder wie Ansprüche gegen Fluglinien, Anträge auf Mutterschutz, Mietsachen und Einspruchshilfe für abgelehnte Asylbewerber in den

USA und Kanada. Beim eigenen Honorar wird der Legal-Bot seinem Namen ebenfalls gerecht. Der Service ist kostenlos – unter anderem weil IBM den Stanford-Studenten seine KI-Plattform Watson kostenlos nutzen lässt.

DoNotPay ist nur ein Beispiel von Tausenden Bots und Programmen, die juristische Arbeit verrichten. Der Boom bei sogenannter *Legal Tech* hat zwei einfache Gründe. Juristische Expertise ist erstens teuer, es lässt sich also viel Geld damit verdienen. Zweitens eignet sich die Juristerei besonders gut für Automatisierung mithilfe Künstlicher Intelligenz, denn sie baut auf präzise formulierten Regeln (den Gesetzen und Verordnungen) in einer stark formalisierten Sprache, und es gibt viele schriftlich dokumentierte Fälle, Kommentare und Verträge, die Maschinen mit Fähigkeit zu Mustererkennung zum Vergleich heranziehen können. Zurzeit wird der größte Teil intelligenter *Legal Tech* von Profis genutzt, also von Anwälten und Unternehmensjuristen, die mit ihnen Verträge nach Fallstricken überprüfen, Berge von Dokumenten für die Unternehmensprüfung durchforsten oder auch Wahrscheinlichkeiten berechnen lassen, bei welchem Gericht sie eine Klage einreichen sollten, um die Aussicht auf Erfolg zu erhöhen.

Je umfassender die Kompetenzen der Legal-Bots werden, und je einfacher ihre Benutzeroberflächen sind, desto mehr werden Laien sie direkt nutzen. DoNotPay-Gründer Joshua Browder hat im Sommer 2017 seine KI-getriebene Chatbot-Technologie geöffnet. Jeder Jura-Kundige kann nun ohne technische Kenntnisse selbst Anwendungen bauen. Ziel ist es, dass DoNotPay bald in mehr als 1000 Rechtsfeldern vom Scheidungsrecht bis zur Privatinsolvenz schnell und unkompliziert helfen kann. Das soll ja nicht die Stärke jedes menschlichen Anwalts sein. Ein kostenloser Legal-Bot hat zudem kein Interesse daran, einen Vertrag möglichst kompliziert zu gestalten, um damit sein Honorar zu erhöhen. Zudem gilt: Es wird vielleicht noch lange dauern, bis Künstliche Intelligenz so schlau ist wie der beste und teuerste Jurist in einem bestimmten Fachgebiet. Aber bei Standardfällen schlägt KI bereits heute das menschliche Mittelmaß mehr als nur gelegentlich. Ist dies der Fall, greifen die Mechanismen der digitalen Skalierung. Sind KI-Programme erst einmal entwickelt und lernen dann durch Feedback-Effekte laufend dazu, können sie kostengünstig vielen Menschen zugänglich gemacht

werden – zumindest wenn die Betreiber das wollen. Fachwissen wird demokratisiert, ermächtigt Verbraucher und erhöht auch die Kompetenz der mittelmäßigen Fachleute. Das ist auch ein realistisches Szenario für die Entwicklung der Künstlichen Intelligenz in dem Feld, in dem Maschinelles Lernen in den vergangenen Jahren wohl die größte Hoffnung auf Fortschritt geweckt hat: die Medizin.

WAS FEHLT MIR, DR. WATSON?

Können Maschinen Krankheiten von Menschen besser diagnostizieren als Menschen? Viele Experimente und Studien besonders aus der Onkologie, Kardiologie und bei genetischen Krankheiten deuten darauf hin. Dank Deep-Learning-Verfahren mit Computertomografie-Bildern lässt sich zum Beispiel das Tumorwachstum bei bestimmten Brustkrebsarten sehr viel genauer vorhersagen und damit deutlich bessere Entscheidungen für Therapien treffen. Doch das ist nur der erste Schritt auf dem Weg des medizinischen Fortschritts durch KI. Algorithmen haben bereits durch Mustererkennung in Zellproben Merkmale zur Unterscheidung von gutartigen und bösartigen Tumoren identifiziert, die der medizinischen Literatur bislang vollkommen unbekannt waren. Künstliche Neuronale Netze diagnostizieren also nicht nur, sie betreiben Spitzenforschung.

Große Hoffnung ruht auch auf der massenhaften Verbreitung von günstigen Sensoren, eingebaut in Standardprodukte, die massenhaft Daten liefern und damit die Grundlage für KI-Gesundheitsinnovationen schaffen. Smarte Uhren können den Herzschlag eines Menschen rund um die Uhr analysieren und Alarm schlagen, wenn abweichende Muster einen Herzinfarkt speziell für eine Risikogruppe ankündigen. Die Zuordnung zur Risikogruppe ist wiederum nur dank Maschinellen Lernens innerhalb eines genanalytischen Verfahrens möglich, bei dem unvorstellbar viele genetische Daten in ein Künstliches Neuronales Netz eingefüttert werden.

Künstliche Intelligenz kann auf MRT-Aufnahmen der Gehirne von sechs Monate alten Babys voraussagen, ob es als Kind oder Jugendlicher Autismus entwickelt, was ein erheblicher Gewinn ist, denn je früher die Therapien beginnen, desto stärker können die Effekte eingedämmt werden. Perspektivisch könnte KI helfen, nicht nur die

zurzeit beste verfügbare Therapie für das Baby herauszusuchen, sondern ein auf Basis des individuellen Genoms maßgeschneidertes Medikament mit optimaler Wirkung zu entwickeln. Forscher und Startups arbeiten zudem mit Hochdruck an Big-Data- und Machine-Learning-Ansätzen, die den Ausbruch und Verlauf von Epidemien vorhersagen wie Dengue-Fieber, sodass Gesundheitsbehörden rechtzeitig Gegenmaßnahmen einleiten und die Epidemie im besten Fall direkt am Ausbruchsherd eindämmen können.

Zusammengefasst lautet die Hoffnung: KI-Agenten werden sich durch Gen-Datenbanken, Patientenakten, wissenschaftliche Studien und epidemische Statistiken fräsen, um Vorsorge, Forschung, Diagnose und Therapie auf eine neue Stufe zu heben. Das wäre freilich schön, wenn dies für möglichst viele Krankheitsbilder sobald wie möglich gelingt. Allerdings ist auch hier, wie bei allen Meldungen zu medizinischen Durchbrüchen, Vorsicht geboten. Forscher und Gründer neigen auch hier zu Übertreibung, oft aus Gründen der Selbstvermarktung. Aber der noch wichtigere Grund dürfte sein: Kaum ein anderer Bereich ist so stark reguliert wie Medizin und Gesundheit – von der Qualifikation des medizinischen Personals und seinen Befugnissen über Zulassungsverfahren für Medikamente und Gerätschaften bis zu besonders hohem Datenschutz der Patienten. Dafür gibt es gute Gründe. Der Preis hierfür ist, dass der Weg für Neuerungen aus den Forschungslaboren in die Umsetzung in Krankenhäusern und Praxen weit und steinig ist.

In den USA nutzen 40 Prozent der niedergelassenen Ärzte und ein Viertel aller Krankenhäuser nach wie vor keine elektronischen Krankenakten. In Deutschland verhindert eine Allianz aus Datenschützern und Ärztelobby seit mehr als zehn Jahren die Einführung einer elektronischen Gesundheitskarte. Der wichtigste Rohstoff für KI-Innovationen im Gesundheitsbereich, die Daten von Patienten, sind in vielen rechtlich verschlossenen Datensilos in vielen unterschiedlichen Formaten gespeichert. Damit diese überhaupt für künstlich intelligente Anwendungen rechtskonform und technisch nutzbar werden, müssen sie in der Regel anonymisiert und dann aufwändig gesäubert und homogenisiert werden. Der komplizierte Zugang zu Daten in einem eigentlich datenreichen Forschungsfeld verzögert den medizinischen Fortschritt also zusätzlich zu den langwierigen Zulassungsverfahren.

VERTRAUEN WIR DEN MASCHINEN?

Wenn es die Innovationen in die Praxis geschafft haben, stellt sich allerdings noch eine grundsätzlicher Frage: Vertrauen wir dem datenbasierten Urteil eines Künstlichen Neuronalen Netzes mehr als dem eines erfahrenen Arztes, der uns vielleicht schon als Kind behandelt hat? Der Informatikstudent mag dies uneingeschränkt mit Ja beantworten. Er glaubt an Statistik. Viele Patienten werden sich mit der schrittweisen Übergabe der Entscheidungskompetenz vom Menschen auf die Maschine schwer tun. Diagnoseärzte und Anwälte stehen immer sehr weit oben auf den Listen von qualifizierten Wissensarbeitern, deren Jobs durch KI-Automatisierung bedroht sind. Gleiches gilt für Wirtschaftsprüfer, Controller, Anlageberater, Versicherungsmakler, öffentliche Verwaltungsbeamte, Sachbearbeiter, Verkäufer und – eine weitere Pointe der Technikgeschichte – auch für jenen Berufsstand, der KI-Systeme schafft, die Programmierer.

Die eine oder andere Studie von Arbeitswissenschaftlern bewegt sich mit ihren Automatisierungsprognosen und den sich daraus ergebenden negativen Beschäftigungseffekten auf sehr dünner Datenbasis. Nüchtern betrachtet müssen KI-Systeme hohe Hürden überwinden, bevor Menschen ihren Urteilen und Entscheidungen trauen. In vielen Fällen wird es auch den meisten Laien kaum möglich sein, künstlich intelligenten Ratsschlag ohne Hilfe von Experten überhaupt einzuholen oder diesen sinnvoll einzuordnen. Wenn es um unser höchstes Gut geht, unsere Gesundheit, werden wir auf diese Einordnung kaum verzichten wollen. Aber wir werden von Ärzten verlangen, dass sie die besten KI-Systeme zu nutzen wissen, um ihre Therapien evidenzbasiert zu verordnen, und nicht auf Grundlage ihres Bauchgefühls, wie allzu oft in der Vergangenheit.

Bei Stitch Fix treffen nach wie vor Menschen die endgültige Entscheidung über die Zusammensetzung der Kleidungsstücke in der Box. Tausende (menschliche) Stylisten legen jedem Paket eine persönliche Karte bei und stehen für Rückfragen zur Verfügung. Auch bei diesem Vorreiter von algorithmisierter Verkaufsberatung ist man davon überzeugt: Menschen verkaufen am Ende doch besser als Maschinen, weil sie eine menschliche (Kunden-)Beziehung aufbauen.

Vielleicht wird die Welt mit weniger Juristen nicht zwingend ein schlechterer Ort. Wo liegt der

volkswirtschaftliche und gesellschaftliche Mehrwert von immer komplizierteren Regeln, die von immer mehr Menschen interpretiert werden? „Die Rechtsindustrie macht mehr als 200 Milliarden Dollar Umsatz in den USA. Ich freue mich, das wir das Recht kostenlos anbieten“, sagt DoNotPay-Gründer Joshua Browder und schiebt hinterher: „Die großen Kanzleien können darüber nicht glücklich sein.“ Als Mandanten werden wir von Anwälten künftig fordern, dass sie ihre Dienste günstiger und in besserer Qualität anbieten, indem sie KI-Werkzeuge wie den LegalBot Ross einsetzen, der Anwälte bei der Großkanzlei BakerHostetler unterstützt.

In fast allen Wissensberufen, bei denen Entscheidungen automatisiert werden, lässt sich die Frage nach Massenarbeitslosigkeit von Wissensarbeitern auch umformulieren: Wie stellen Verkäufer, Anwälte und Ärzte sicher, dass sie mithilfe von KI mehr Menschen günstiger mit besserer Beratungsleistung helfen können? Der Leitgedanke hier ist: *augmented decision making* anstatt reine Automation. Ginni Rometty, die Vorstandsvorsitzende von IBM, sieht die Dinge so: „Was einige Leute Künstliche Intelligenz nennen, ist in Wirklichkeit eine Technologie, die unsere Fähigkeiten stärkt. Eigentlich geht es nicht um Künstliche Intelligenz, sondern um die Erhöhung unserer Intelligenz.“ Für Wissensarbeiter hieße das im Umkehrschluss, dass nicht Künstliche Intelligenz sie in den kommenden Jahren ersetzen wird. Tech-affine Verkäufer, Anwälte und Ärzte werden jene Kollegen ersetzen, die KI nicht als Entscheidungsassistenten intelligent zu nutzen wissen.

DENKEN MÜSSEN WIR NOCH SELBST

Für jeden Einzelnen stellt Künstliche Intelligenz also eine neue, sehr grundlegende Frage: Welche Entscheidungen wollen wir nicht an Maschinen delegieren? Denn natürlich irren auch die intelligentesten Maschinen und natürlich können aus Daten lernende Systeme von Menschen eingesetzt werden, um andere Menschen zu manipulieren oder unfair zu behandeln. Je intelligenter Maschinen werden, desto kritischer müssen wir sie hinterfragen. Mit der Hinwendung zur Vernunft und Wissenschaft hat die Aufklärung die Grundlagen gelegt, die Mitte des 18. Jahrhunderts Charles Babbage den ersten Compu-

ter erdenken und Konrad Zuse gut hundert Jahre später den ersten programmierbaren Rechner bauen ließ. Die Verknüpfung der Computer in einem weltweiten Netz durch Tim Berners-Lee vor rund 25 Jahren machte die digitale Riesemaschine zum mächtigsten Werkzeug, das der Mensch je geschaffen hat. Jetzt lernen Maschinen das Lernen – und wir brauchen mehr Abstand zu ihnen.

Wir müssen verstehen, wann maschinelle Assistenz uns nützt – und in welchen Kontexten sie uns in unserem Denken behindert. Die Automatisierung von Entscheidungen bietet große Chancen für den Einzelnen, Organisationen und für die Gemeinschaften, die wir Gesellschaften nennen. Doch auch im Zeitalter der rationalen Automatisierung von Entscheidungen durch KI gilt: Menschen müssen mit ihren Entscheidungen glücklich werden, Computer nicht. Maschinen werden nie fühlen, was Glück ist. Die Irrationalität gehört zum Wesen menschlicher Entscheidungen. Im Zeitalter der Berechenbarkeit durch aus Daten lernende Maschinen könnte unsere Unberechenbarkeit unsere größte Stärke werden.

In den vergangenen Jahren wurde viel über eine neue Maschinenethik diskutiert und über die Frage, ob man (und wenn ja, wie) Maschinen ethisch korrektes Verhalten einprogrammieren könne.⁰³ Aufgehängt waren diese Debatten oft an konstruierten Dilemmata nach dem Prinzip: Ein autonomes Fahrzeug steuert auf eine Mutter mit Baby im Kinderwagen und eine Gruppe mit fünf Senioren zu. Es muss entscheiden, wen es überfährt. Mutter und Baby, die zusammen voraussichtlich noch 150 Jahre leben, oder die fünf Senioren mit einer kollektiven Lebenserwartung von 50 Jahren. Solche Gedankenspiele sind notwendig. Die Würde des Menschen ist unantastbar. Im Krieg darf ein General die Abwägung treffen, fünf Soldaten zu opfern, wenn er zehn dafür retten kann. Im zivilen Leben darf dies in der Theorie niemand. In der Praxis macht es ein Autofahrer, der bei überhöhter Geschwindigkeit und ohne Möglichkeit zu bremsen sein Fahrzeug lieber in eine Menschengruppe steuert als gegen einen Betonpfosten.

Die Automatisierung von Entscheidungen ist in vielen Kontexten eine ethische Herausfor-

03 Siehe dazu auch die Beiträge von Catrin Misselhorn und Oliver Bendel in dieser Ausgabe (Anm. d. Red.).

derung, aber zugleich ein moralischer Imperativ. Wenn wir mit autonomen Fahrzeugen die Zahl der Verkehrstoten halbieren können, müssen wir das tun. Wenn wir dank maschineller Mustererkennung von Zellen vielen Krebspatienten das Leben retten können, dürfen wir diesen Fortschritt nicht von einer Ärztelobby verzögern lassen, die Angst um ihre Honorare hat. Und wenn KI-Systeme in Südamerika Kindern aus armen Verhältnissen das Rechnen lehren, dürfen wir nicht darüber lamentieren, dass es doch schöner wäre, wenn es dort mehr menschliche Mathematiklehrer gäbe.

Im Verhältnis von Mensch und Maschine ändert sich durch Künstliche Intelligenz im Grundsatz weniger, als es uns der ein oder andere KI-Entwickler weismachen möchte. Joseph Weizenbaum, der deutsch-amerikanische Erfinder des Chatprogramms ELIZA, schrieb 1976 den Weltbestseller „Die Macht der Computer und die Ohnmacht der Vernunft“. Das Buch war ein Appell wider den mechanistischen Maschinenglauben seiner Epoche. In einer Zeit, in der die Vorstellung einer technischen Vorherbestimmung der Menschheit im Silicon Valley wieder in Mode kommt, verdient dieses Buch eine Neuauflage.

Wir können Entscheidungen an Maschinen in vielen einzelnen Bereichen delegieren. KI-Systeme, gut programmiert und mit den richtigen Daten gefüttert, sind nützliche Fachidioten. Ihnen fehlt aber die Fähigkeit, das große Ganze zu sehen. Die wichtigen Entscheidungen, darunter jene über das Ausmaß der maschinellen Assistenz, bleiben menschliche. Oder allgemeiner formuliert: Künstliche Intelligenz kann uns das Denken nicht abnehmen.

Die Geschichte der Menschheit ist die Summe menschlicher Entscheidungen. Wir entscheiden normativ, was wir wollen. Das wird so bleiben. Das positive Weltbild für die nächste Entwicklungsstufe des maschinenunterstützten Informationszeitalters müssen wir dabei nicht einmal neu erfinden. „Es ist ganz schlicht die Rückbesinnung auf die humanistischen Werte“, sagt der New Yorker Risikokapitalist, Buchautor und TED-Speaker Albert Wenger. Diese lassen sich seiner Ansicht nach auf folgende Formel bringen: „Die Fähigkeit, Wissen zu schaffen, macht uns Menschen einzigartig. Wissen entsteht in einem kritischen Prozess. Alle Menschen können und sollen an diesem Prozess teilhaben.“ Die di-

gitale Revolution ermöglicht uns zum ersten Mal in der Menschheitsgeschichte, dieses humanistische Ideal in die Praxis umzusetzen. Indem wir Künstliche Intelligenz intelligent und zum Wohle des Menschen einsetzen.

Der Text ist ein leicht abgeänderter Auszug aus dem Buch: „Mensch und Maschine. Wie Künstliche Intelligenz und Roboter unser Leben verändern“. Es erscheint am 16. März im Reclam Verlag.

THOMAS RAMGE

ist der Technologie-Korrespondent des Wirtschaftsmagazins „brand eins“ und schreibt für „The Economist“. Er hat zwölf Sachbücher und einen Roman veröffentlicht. Zuletzt erschienen: „Das Digital. Markt, Wertschöpfung und Gerechtigkeit im Datenkapitalismus“ (zusammen mit Viktor Mayer-Schönberger, Econ Verlag, 2017).

www.thomasramge.de

ENTGRENZUNGEN ZWISCHEN MENSCH UND MASCHINE, ODER: KÖNNEN ROBOTER ZU GUTER PFLEGE BEITRAGEN?

Christoph Kehl

Die jüngsten Fortschritte im Bereich der Künstlichen Intelligenz (KI) sind zweifelsohne beeindruckend: Dass Computer im Schach dem Menschen bereits seit Jahrzehnten haushoch überlegen sind, ist inzwischen Normalität. Dass aber ein KI-System fähig ist, im wesentlich komplexeren und kaum berechenbaren Go-Spiel in kürzester Zeit eine Spielstärke zu erreichen, die den weltbesten Spielern nicht den Hauch einer Chance lässt – wie jüngst geschehen –, galt bis vor Kurzem noch als undenkbar.

Algorithmen steuern immer mehr Arbeits- und Lebensbereiche, von der Internetsuche über die computergestützte Befundung in der Medizin bis zu Finanztransaktionen. In der Regel handelt es sich dabei um hochspezialisierte Softwareagenten, die in einer rein virtuellen Umwelt agieren. Es wird aber auch an „intelligenten“ Systemen gearbeitet, die fähig sind, sich in der analogen Welt zu bewegen, in diese einzugreifen und mit Menschen in physischen Kontakt zu treten. Maschinen dieser Art haben das Potenzial, das Mensch-Technik-Verhältnis ganz neu zu definieren: Insofern Technik nämlich beginnt, autonom zu agieren und damit ein einfaches Denken in Zweck-Mittel-Relationen zu untergraben (wie es für die Anwendung herkömmlicher Werkzeuge maßgeblich ist), wird auch die Abgrenzung zwischen dem Menschen und den von ihm geschaffenen Arbeitsmitteln immer unschärfer. Dieses Phänomen wird im Folgenden als Mensch-Maschine-Entgrenzung bezeichnet.

Die gesellschaftliche Brisanz des Themas rückt zunehmend in das politische Bewusstsein. So sah sich das Europäische Parlament 2016 veranlasst, der EU-Kommission zivilrechtliche Regelungen im Bereich Robotik zu empfehlen.⁰¹

Im Bundestag waren „Künstliche Intelligenz“ und „Robotik“ Thema verschiedener Anhörungen, darüber hinaus wurde das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB) damit beauftragt, die technologischen Grundlagen sowie gesellschaftlichen Perspektiven dieser Entwicklung zu untersuchen. Im Folgenden werden Ergebnisse und Überlegungen aus dem dazugehörigen TAB-Projekt vorgestellt, das sich in der Sondierungsphase mit technologischen und visionären Aspekten der Mensch-Maschine-Entgrenzung⁰² befasste und in der Vertiefungsphase die gesellschaftlichen Herausforderungen der Entwicklung näher bestimmte, und zwar am Beispiel eines konkreten Anwendungskontextes – der Pflege.

ZWISCHEN VISION UND WIRKLICHKEIT

Vor allem die Entwicklungen in zwei Technologiefeldern – der Robotik und den Neurotechnologien – tragen dazu bei, dass die ehemals scharfen Trennlinien zwischen Mensch und Maschine immer mehr verschwimmen.

Fortschritte in der Robotik machen es inzwischen möglich, auch komplexe menschliche Handlungsabläufe maschinell nachzubilden. Während der klassische Industrieroboter auf die Automatisierung repetitiver industrieller Prozesse festgelegt ist und seine Dienste aus Sicherheitsgründen weitgehend abgeschottet von Menschen vollbringt, eröffnen sich für moderne Serviceroboter vielfältige Anwendungsperspektiven auch außerhalb industrieller Fertigungshallen – seien es einfache Aufgaben in Privathaushalten (in denen Staubsaugerroboter bereits millionenfach in Anwendung sind) oder komplexe, auch perso-

nenbezogene Dienstleistungen im Pflegebereich. Möglich gemacht haben diesen Schritt in die Alltagswelt insbesondere folgende Merkmale, mit denen sich Serviceroboter von klassischen Industrierobotern abgrenzen lassen: hardwareseitig die Leichtbauweise, die – ergänzt um komplexe Sensorik und Aktorik – eine immer engere Interaktion mit dem Menschen ermöglicht; softwareseitig die Realisierung von hochentwickelten Lern- und Planungsverfahren (auf Basis von KI und Maschinellem Lernen), welche die Systeme befähigen sollen, nicht nur ein festgelegtes Handlungsprogramm abzuspielen, sondern sich auch unter neuen oder sich verändernden Bedingungen weitgehend autonom zurechtzufinden. Die Service-robotik gilt inzwischen als zukunftsweisender Wachstumsmarkt, dessen weltweites Volumen der etablierten Industrierobotik laut Prognosen in wenigen Jahren den Rang ablaufen könnte.⁰³

Sorgt die Robotik auf der einen Seite dafür, dass Maschinen in ihren Handlungsmöglichkeiten immer menschenähnlicher werden, eröffnen die Fortschritte auf dem Gebiet der Neurotechnologien auf der anderen ganz neue Optionen, mit dem Menschen technologisch zu interagieren. Schon zum klinischen Standardrepertoire gehören stimulierende Systeme, die elektrische Impulse an das Gehirn übertragen – etwa sensorische Neuroprothesen wie das Cochlea-Implantat, mit dem sich Einschränkungen des Hörsinns über die gezielte Stimulation der entsprechenden Nervenfasern technisch kompensieren lassen. Daneben wird aber auch intensiv an ableitenden Anwendungen geforscht, mit dem Ziel, Signale aus dem Nervensystem zu gewinnen und zur Steuerung etwa von künstlichen Gliedmaßen zu verwenden. Dabei macht man sich zunutze, dass gedankliche Aktivität elektrische Potenziale erzeugt, die sowohl invasiv (mittels implantierter Mikroelektroden) als auch nichtinvasiv (mittels auf der Kopfhaut fixierter Elektroden) detektierbar sind.

01 Vgl. Europäisches Parlament, Bericht mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103(INL)), Brüssel 2017.

02 Vgl. Christoph Kehl/Christopher Coenen, Technologien und Visionen der Mensch-Maschine-Entgrenzung. Sachstandsbericht zum TA-Projekt „Mensch-Maschine-Entgrenzungen: zwischen künstlicher Intelligenz und Human Enhancement“. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, Berlin 2016.

03 Vgl. Expertenkommission Forschung und Innovation, Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2016, Berlin 2016, S. 53.

Beide Herangehensweisen haben ihre spezifischen Limitationen: Die nichtinvasive Datengewinnung leidet an einer eher schlechten Signalqualität und kommt somit nur für die binäre Steuerung einfacher Kommunikationssysteme in Betracht.⁰⁴ Mit invasiven Ableitmethoden sind zwar theoretisch höhere Datenraten erreichbar, was die Feinsteuerung künstlicher Gliedmaßen in den Bereich des technisch Möglichen rückt – einschränkende Faktoren sind hier aber die noch sehr begrenzte Langzeitstabilität der implantierten Sonden, die aufgrund der Fremdkörperreaktion des Gehirns (Einkapselung) in der Regel nach kurzer Zeit ihre Funktion verlieren.

Auch wenn die Datenableitung aus dem Gehirn und damit die intuitive Gedankensteuerung komplexer Apparaturen noch nicht über das experimentelle Stadium hinausgekommen ist, zeichnen sich bereits weitere Anwendungsperspektiven ab. Sollte es dereinst gelingen, durch Integration von ableitenden und stimulierenden Schnittstellen ein stabiles bidirektionales Feedbacksystem zu etablieren, ließen sich „fühlende“ Handprothesen realisieren, die dem Nutzer während des Greifprozesses eine sensorische Rückmeldung geben. Auf die Spitze getrieben wird die Mensch-Maschine-Entgrenzung schließlich durch die Kombination neurotechnologischer Anwendungen mit Robotertechnologie – etwa im Form „intelligenter“, sich selbst steuernder Implantate oder roboterisierter Gliedmaßen –, wodurch autonom agierende maschinelle Systeme quasi untrennbar mit dem Menschen verschmelzen. Klar ist, dass in einer solchen Konstellation nicht mehr eindeutig ist, wer wen kontrolliert: der Mensch die Maschine oder umgekehrt?

Das visionäre Potenzial dieser Entwicklungen steht außer Frage. Sowohl die Fortschritte in der Robotik als auch den Neurotechnologien werden auf gesellschaftlicher Ebene entsprechend durch weitreichende Zukunftsvorstellungen begleitet, die interessanterweise wiederum an ältere, kulturell äußerst wirkmächtige Motive anknüpfen.⁰⁵ Da wäre zum einen die Vision einer (neuro-)technologischen Optimierung des Men-

04 Bahnbrechend sind diesbezüglich vor allem die Arbeiten von Niels Bierbaumer in Tübingen, dem es kürzlich gelungen ist, mittels einer Gehirn-Computer-Schnittstelle mit komplett gelähmten Locked-In-Patienten zu kommunizieren. Vgl. Ujwal Chaudhary et al., Brain-Computer Interface-Based Communication in the Completely Locked-In State, in: PLOS Biology 1/2017, S. 1–25.

05 Vgl. Kehl/Coenen (Anm. 2), S. 33 ff.

schen (Human Enhancement),⁰⁶ die Anfang des 20. Jahrhunderts, im Zuge biologisch-evolutionären Denkens, erstmals auf den Plan trat.⁰⁷ Zum anderen wird derzeit auch intensiv über die Möglichkeit diskutiert, dass Roboter den Menschen in nicht allzu ferner Zukunft an Intelligenz überflügeln und wahlweise unterjochen oder in eine glorreiche Zukunft führen könnten. Auch diese Idee reicht mindestens hundert Jahre zurück und wurde erstmals in Karel Čapeks dystopischem Drama „R.U.R. – Rossum’s Universal Robots“ (1920) ausformuliert. Angetrieben und befeuert werden diese visionären Debatten seither maßgeblich von publikumswirksamer Science-Fiction, in der verschiedene Spielarten dieser Entgrenzungsvisionen plakativ ausgemalt werden.⁰⁸

Dass Hybride aus Menschen und Maschinen offensichtlich keine bloße Zukunftsvision mehr sind, macht bereits die Tatsache deutlich, dass derzeit etwa 30 000 Menschen in Deutschland ein Cochlea-Implantat tragen. Dennoch stellt sich die Frage, inwiefern die erwähnten futuristischen Visionen als maßgebliche Referenzpunkte für die gesellschaftliche Auseinandersetzung mit der Entgrenzungsdynamik taugen. Um es kurz zu machen: Auf Basis einer nüchternen Bewertung der Leistungsfähigkeit der zugrundeliegenden Technologien kommt die TAB-Studie zum Ergebnis, dass größere Zweifel angebracht sind.⁰⁹ So sind die etablierten neurotechnologischen Eingriffe in das Gehirn mit teils erheblichen gesundheitlichen Risiken verbunden, die nur zur Kompensation schwerer körperlicher Defizite ethisch gerechtfertigt erscheinen. Gegen die breitere Anwendung bei Gesunden spricht vor allem, dass die technologisch erreichbaren Zusatzfähigkeiten (z. B. Sinneserweiterungen wie das Hören von Ultraschall mittels Cochlea-Implantat) absehbar keinen wirklichen Nutzen versprechen, wie er für eine Optimierung des Menschen erforderlich wäre. Auch die Servicerobotik befindet sich derzeit in einem Entwicklungsstadium, das jeglichen Gedanken an eine „Machtübernahme“ intelligenter Maschinen ins Reich der Science-

Fiction verbannt. Die Autonomie der verfügbaren Systeme ist noch äußerst beschränkt und wird bereits durch einfache Manipulationsaufgaben – beispielsweise das Erkennen und Greifen von gängigen Haushaltsgegenständen unter realen Bedingungen – auf eine harte Probe gestellt. Wann – und ob überhaupt – eine starke KI realisierbar ist, die der menschlichen Intelligenz gleichkommt, ist eine Frage, über die sich nach wie vor nur spekulieren lässt.

ANWENDUNGSFELD PFLEGE

Die beschriebenen Entwicklungen werfen stattdessen Fragen auf, die zwar vordergründig wenig aufsehenerregend erscheinen, tatsächlich aber von großer normativer Brisanz sind: Wie lässt sich die zunehmend enge Interaktion zwischen Menschen und autonomen Maschinen nicht nur sicher, sondern auch moralisch verantwortungsvoll gestalten? Wer ist haftbar zu machen, wenn doch etwas schiefgeht? Und gibt es ethische Grenzen der Technisierung – und wenn ja, wo liegen sie?

Einer der gesellschaftlichen Bereiche, in denen derartige Fragen derzeit besonders virulent werden, ist die Altenpflege, die sich im Zuge des demografischen Wandels vor enorme Herausforderungen gestellt sieht: Seit Jahren steigt die Zahl der Pflegebedürftigen, während es immer schwieriger wird, Pflegefachkräfte in ausreichender Zahl zu rekrutieren. Schon heute zeichnet sich eine massive Versorgungslücke ab, die den Ruf nach technischer und insbesondere robotischer Unterstützung lauter werden lässt – innovative Assistenztechnologien für die Pflege werden von der Politik entsprechend seit Jahren gefördert, und dies nicht nur in Japan, das hier als besonders fortschrittlich gilt, sondern auch in Deutschland. Gleichzeitig liegt auf der Hand, dass es sich bei der Pflege um einen außerordentlich sensiblen Bereich handelt. Zunehmende maschinelle Autonomie und die fragile Verfassung der Pflegebedürftigen stehen hier in einem spannungsreichen Verhältnis, weshalb der perspektivische Einsatz autonomer Pflegetechnologien ethisch hochumstritten ist und auch in der Bevölkerung auf viele Vorbehalte trifft. Die Altenpflege ist damit ein Feld, in dem sich die Ambivalenzen der Mensch-Maschine-Entgrenzung in paradigmatischer Weise zuspitzen.

Da sich die Rolle der Neurotechnologie hauptsächlich auf den therapeutischen Bereich

06 Vgl. dazu die Beiträge zum Thema „Der Neue Mensch“ in *Aus Politik und Zeitgeschichte* 37–38/2016.

07 Vgl. John D. Bernal, *The World, the Flesh and the Devil. An Inquiry into the Future of the Three Enemies of the Rational Soul*, London 1929.

08 Siehe dazu auch den Beitrag von Ingo Irsigler und Dominik Orth in dieser Ausgabe (*Anm. d. Red.*).

09 Vgl. Kehl/Coenen (*Anm. 2*), S. 140 ff.

beschränkt, stellt die Servicerobotik für die Altenpflege die treibende Kraft der Entgrenzungsdynamik dar. Ihr wird großes Potenzial zugeschrieben, Pflegekräfte entlasten sowie Pflegebedürftige im Alltag unterstützen zu können – entsprechende Anwendungen befinden sich bereits seit vielen Jahren in Entwicklung und Erprobung.¹⁰ Die sich bietenden Unterstützungsmöglichkeiten sind äußerst vielfältig und lassen sich gemäß ihren Anwendungszwecken grob folgenden Kategorien zuordnen:

Assistenzroboter zur physischen Alltagsunterstützung kommen der eigentlichen Bestimmung der Servicerobotik – der „Erweiterung menschlicher Handlungsfähigkeit“¹¹ – am nächsten. Sie sollen entweder im häuslichen Umfeld eingesetzt werden, um älteren und pflegebedürftigen Menschen zur Hand zu gehen (z. B. Holen und Bringen von Gegenständen), oder der Unterstützung pflegerischer Routinetätigkeiten im Pflegeheim dienen (z. B. als Hebehilfen oder für logistische Aufgaben).

Bei *sozialen Robotern* steht hingegen nicht die physische, sondern die sozial-emotionale Unterstützung im Vordergrund. Zu unterscheiden ist hier zwischen Geräten, die selbst mit Menschen sozial interagieren und kommunizieren (z. B. als Unterhaltungs- oder Zuwendungsroboter), und solchen, deren Hauptzweck darin besteht, zwischenmenschliche Kontakte zu vermitteln und damit die soziale Teilhabe zu fördern (z. B. als Kommunikationsassistent oder Telepräsenzroboter). Ein bereits relativ etabliertes Anwendungsfeld der sozial-interaktiven Robotik stellt die Demenztherapie dar, wo vermehrt tierähnliche Roboter wie die Robbe Paro unterstützend zum Einsatz kommen.¹²

Robotischen Mobilitätshilfen schließlich kommt in der Altenpflege aufgrund der verbreitet auftretenden Bewegungseinschränkungen eine wichtige praktische Bedeutung zu. Die Geräte werden entweder als Exoskelette direkt am Körper getragen

oder stellen „intelligente“, beispielsweise um Navigationsfunktionen ergänzte Erweiterungen einfacher Fortbewegungshilfen dar (Rollatoren, Rollstühle). Insbesondere Exoskelette bieten breite Anwendungsmöglichkeiten, da sie sowohl als alltägliche Mobilitätshilfe für ältere Menschen und Pflegebedürftige als auch zur Entlastung von Pflegekräften bei körperlich anstrengenden Aufgaben eingesetzt werden können.¹³

Die Königsdisziplin der Pflegerobotik ist der multifunktionale Roboterassistent, der anspruchsvolle Navigations- und Manipulationsfertigkeiten mit kommunikativen und sozial-affektiven Kompetenzen kombiniert und damit prinzipiell einen vollwertigen Alltagsgefährten darstellt. Dieses Leitbild war für das Entwicklungsgeschehen lange Zeit prägend¹⁴ und bestimmt bis heute die öffentliche Wahrnehmung der Pflegerobotik. Mit dem Care-O-bot 4 des Fraunhofer-Instituts für Produktionstechnik und Automatisierung (IPA) hat kürzlich die Inkarnation einer solchen Companion-Vision, von den Entwicklerinnen und Entwicklern als „vielseitiger Gentleman“ angepriesen,¹⁵ nach fast 20-jähriger Entwicklungszeit Marktreife erlangt. Ursprünglich zur Unterstützung körperlich beeinträchtigter Menschen im häuslichen Umfeld gedacht, zielt die derzeit anvisierte Kommerzialisierung des Systems allerdings primär auf andere Anwendungsfelder wie beispielsweise die Kundenführung im Einzelhandel. Für einen sicheren Einsatz als Haushalts- und Pflegeassistent sind die Manipulationsfähigkeiten bei Weitem noch nicht ausgereift.

Der Care-O-bot steht damit symptomatisch für den Stand der Pflegerobotik. Trotz langjähriger Entwicklungsbemühungen und unzähliger Prototypen hat es bislang nur eine Handvoll Systeme in die praktische Anwendung geschafft – allesamt Spezialanwendungen (neben der Robbe Paro gehören dazu einige Esshilfen sowie vereinzelte Exoskelette und Telepräsenzroboter), die über keine Greifarme und nur über begrenzte Autono-

10 Vgl. Birgit Graf/Torsten Heyer/Barbara Klein, Servicerobotik für den demografischen Wandel. Mögliche Einsatzfelder und Entwicklungsstand, in: Bundesgesundheitsblatt 56/2013, S. 1145–1152.

11 Michael Decker, Robotik, in: Armin Grunwald (Hrsg.), Handbuch Technikethik, Stuttgart 2013, S. 356.

12 Vgl. Barbara Klein, Zwischen Natur und Technik – Künstliche Tiere. Können künstliche Tiere zur Lebensqualität in der Altenhilfe beitragen?, in: Meret Fehlmann/Margot Michel/Rebecca Niederhauser (Hrsg.), Tierisch! Das Tier und die Wissenschaft. Ein Streifzug durch die Disziplinen, Zürich 2016, S. 33–42.

13 Mobilitätshilfen sind bislang die einzigen Roboterlösungen für den Pflegebereich, die sich einigermaßen sinnvoll mit neurotechnologischen Schnittstellen kombinieren lassen. Das Exoskelett HAL ist vermutlich das weltweit erste kommerziell verfügbare Produkt, das Muskelsignale zur Bewegungssteuerung nutzt.

14 Vgl. Bettina-Johanna Krings et al., ITA-Monitoring „Service-roboter in Pflegearrangements“, Pre-Print: 4. 12. 2012, Karlsruhe 2012, S. 41.

15 Fraunhofer IPA, Roboter als vielseitiger Gentleman, Pressemitteilung, 15. 1. 2015.

mie verfügen und damit deutlich vom Leitbild des multifunktionalen Alltagsassistenten abweichen. Dafür verantwortlich sind unter anderem die hohen technischen Hürden, die es zu überwinden gilt. Hinzu kommt ein schwieriges Marktumfeld: Da die Pflegebranche unter hohem wirtschaftlichen Druck steht und nur wenige Dienstleister über ausreichendes Investitionspotenzial verfügen, ist unklar, ob sich die hohen Entwicklungskosten amortisieren lassen.¹⁶ Viele Entwicklungsprojekte sind deshalb von öffentlicher Förderung abhängig.

Während der robotische Alltagsassistent in den vergangenen Jahren etwas an Attraktivität verloren hat, stehen zunehmend spezialisierte Automatisierungslösungen für die stationäre Pflege im Fokus der Entwicklerinnen und Entwickler – ein Beispiel ist etwa der „intelligente“ Pflegewagen des Fraunhofer IPA, der dereinst in der Lage sein soll, autonom zum Einsatzort zu navigieren, den Verbrauch zu dokumentieren und Pflegeutensilien selbstständig nachzufüllen.¹⁷ Die neue Schwerpunktsetzung erfolgte nicht zuletzt aus wirtschaftlichen Erwägungen heraus, denn Systeme für den Einsatz in Pflegeheimen versprechen eine höhere Auslastung und damit eine verbesserte Kosten-Nutzen-Bilanz als Anwendungen für Privathaushalte. Insgesamt sind die zukünftigen Marktaussichten für die Pflegerobotik derzeit allerdings nur schwer absehbar. Sie hängen zum einen davon ab, wie sich die sozioökonomischen Rahmenbedingungen für die Pflegebranche im Zuge des demografischen Wandels entwickeln werden (Pflege- und Fachkräftebedarf, Personalkosten, staatliche Pflegeleistungen und weiteres). Zum anderen wird es jedoch auch entscheidend darauf ankommen, ob es gelingt, Systeme zu entwickeln, welche die hohen Anforderungen an den pflegerischen Nutzen sowie die Technikakzeptanz zu erfüllen vermögen.

WAS IST GUTE PFLEGE – UND WAS KÖNNEN ROBOTER DAZU BEITRAGEN?

Vergleicht man die gesellschaftliche Debatte zu Automatisierungsbestrebungen in der Pflege mit

¹⁶ Vgl. Martin Hägele/Nikolaus Blümlein/Oliver Kleine, Wirtschaftlichkeitsanalysen neuartiger Servicerobotik-Anwendungen und ihre Bedeutung für die Robotik-Entwicklung. Eine Analyse der Fraunhofer-Institute IPA und ISI im Auftrag des BMBF, München 2011, S. 86 ff.

¹⁷ Vgl. Birgit Graf/Ralf S. King, Pflegewagen werden intelligent, in: *Pflegezeitschrift* 3/2016, S. 373 ff.

solchen, die zu ähnlichen Entwicklungen in anderen Dienstleistungskontexten geführt werden, so fällt ein Unterschied auf: Während bei den meisten Branchen die Diskussion um die Zukunft der Arbeit primär unter ökonomischen Vorzeichen geführt wird, vor allem im Hinblick auf mögliche negative Beschäftigungseffekte,¹⁸ stehen bei der Pflege normative Fragen zur zukünftigen Ausgestaltung des professionellen Pflegehandelns klar im Vordergrund. Dies hat zwei Gründe: Zum einen scheint die Personaldecke in vielen Pflegeheimen bereits so ausgedünnt, dass die verbliebenen Rationalisierungspotenziale wohl kaum noch maßgeblich ins Gewicht fallen.¹⁹ Zum anderen – wichtiger noch – handelt es sich bei der Pflege um eine personenbezogene, fürsorgliche Dienstleistung, in welcher der zwischenmenschlichen Interaktion eine grundlegende Bedeutung zukommt. Kaum jemand wird bestreiten wollen, dass Empathie, menschliche Zuwendung und körperliche Nähe zentrale Aspekte guter Pflege sind, was sich auch daran ablesen lässt, dass Vorstellungen einer vollautomatisierten, vollständig dehumanisierten Pflege gemeinhin als Schreckensvisionen gelten und unisono abgelehnt werden.²⁰ Dass derartige Dystopien im ethischen und öffentlichen Diskurs dennoch so prominent vertreten sind, verweist einerseits darauf, dass Automatisierungsbestrebungen im Pflegebereich normative Fragen neuer Dimension aufwerfen. Andererseits werden daran aber auch zentrale Defizite der entsprechenden Debatten deutlich. Da diese hauptsächlich um spekulative Szenarien kreisen, findet eine differenzierte inhaltliche Auseinandersetzung mit den Potenzialen und Grenzen der Robotik in der Pflege kaum statt.

Sicher ist: Einen Roboter zu schaffen, der Pflegeaufgaben am Menschen vollautonom übernehmen und somit eine menschliche Pflegekraft zu ersetzen vermag, ist weder ein technisch realistisches (zumindest nach jetzigem Stand) noch ein erklärtes Ziel. Mit den derzeit in Entwicklung befindlichen Lösungen wird vielmehr an-

¹⁸ Für Furore und anhaltenden Diskussionsbedarf sorgte insbesondere die Studie von Carl B. Frey/Michael A. Osborne, *The Future of Employment: How Susceptible are Jobs to Computerization?*, Oxford 2013.

¹⁹ Vgl. Volker Hielscher/Lukas Nock/Sabine Kirchen-Peters, *Technikeinsatz in der Altenpflege. Potenziale und Probleme in empirischer Perspektive*, Baden-Baden 2015, S. 153.

²⁰ Vgl. Robert Sparrow/Linda Sparrow, *In the Hands of Machines? The Future of Aged Care*, in: *Minds & Machines* 16/2006, S. 141–161.

gestrebt, Pflegeprozesse technisch so zu assistieren, dass mehr Zeit für die eigentliche Pflege und zwischenmenschliche Begegnungen bleibt – sei es durch Entlastung des Pflegepersonals oder durch die Befähigung pflegebedürftiger Menschen zu sozialer Teilhabe. Die eigentliche Frage lautet dann also, ob und inwiefern sich diese Zielsetzungen in die Tat umsetzen lassen. Dies lässt sich wiederum nicht anhand spekulativer Szenarien, sondern nur mit einem genauen Blick auf die konkreten Nutzungszusammenhänge beantworten, schließlich zeigen die vorliegenden Erfahrungen, dass die Einführung neuer Technologien vielfältige Auswirkungen auf die Pflegearbeit haben kann. Relativ gut untersucht sind die derzeit laufenden Bestrebungen, elektronische Pflegedokumentations- und -managementsysteme einzuführen – ebenfalls einhergehend mit dem Versprechen, dadurch Zeitersparnisse und Arbeitserleichterungen zu erzielen. Studien zu den pflegerischen Implikationen des elektronischen Dokumentationswesens legen nahe,²¹ dass die standardisierenden Effekte derartiger Systeme auch das Beziehungshandeln nicht unberührt lassen, es gewissermaßen einer „Maschinenlogik“²² unterordnen. Deutlich wird daran, dass zweck- und empfindungsbezogene Bereiche professioneller Pflegearbeit nicht klar voneinander zu trennen, vielmehr eng aufeinander bezogen sind – ein Punkt, der auch aus pflegewissenschaftlicher Sicht betont wird.²³ Gute Pflege lässt sich demnach nicht auf standardisierte Problemlösungsverfahren reduzieren, sondern basiert wesentlich auf „körperlicher und emotionaler Arbeit“, ohne die der „Zugang zum Anderen“ verwehrt bleibt.²⁴

Vor diesem Hintergrund sind Befürchtungen nicht von der Hand zu weisen, dass ein forciertes Robotereinsatz dazu führen könnte, dass Pflegearbeit zunehmend einem mechanistischen Verständnis unterworfen wird, das heißt auf zweckbezogene Anteile verengt und die empfindungsbezogenen Aspekte entsprechend marginalisiert werden – vor allem, wenn sich dadurch betriebswirtschaftlich

einsparen lässt. Dies gilt umso mehr, als sich ein einseitiges Verständnis der Pflege als zweckrationales Problemlösungshandeln in übergreifende und bereits länger anhaltende Standardisierungs- wie auch Ökonomisierungsbestrebungen einordnet (z.B. zur Organisation der Versorgungsprozesse, zur Qualitätssicherung oder zur Abbildung und Abrechnung des Leistungsgeschehens), die nicht nur die Pflege, sondern das Gesundheitswesen insgesamt zunehmend betriebswirtschaftlichen Handlungslogiken unterwerfen.²⁵

Was folgt daraus? Zumindest dreierlei. *Erstens:* Der Nutzen der autonomen Robotik für die Pflege lässt sich nicht isoliert an einzelnen, geeigneten respektive weniger geeigneten Aufgaben festmachen, da auch von der Automatisierung von Einzelprozessen tiefgreifende Implikationen für die gesamte Pflegepraxis zu erwarten sind. Um neue Robotiklösungen optimal in die Pflegearbeit einzupassen, müssen Arbeitsprozesse normiert, Qualitätsstandards definiert und auch das Arbeitsumfeld robotergerecht gestaltet werden. *Zweitens:* Bei der Frage, ob sich eine Robotikanwendung für die Pflege fruchtbar machen lässt, gilt es deren Auswirkungen auf die personengebundenen Kernprozesse genau im Blick zu behalten. Werden neue Freiräume für Beziehungshandeln geschaffen oder bestehende minimiert? *Drittens:* Darüber hinaus gibt es kein Patentrezept, wie sich die Potenziale der Robotik für die Pflege nutzbar machen lassen. Letztlich handelt es sich dabei um eine anspruchsvolle Gestaltungsaufgabe, die sowohl die Einzeltechnologien (Design, Funktionalität) als auch deren soziotechnische Einbettung einzubeziehen hat. Die Diskrepanz zwischen postuliertem Lösungspotenzial und tatsächlich erreichtem Nutzen robotischer Pflegeanwendungen ist nach Meinung verschiedener Expertinnen und Experten nicht zuletzt darauf zurückzuführen, dass den Bedürfnissen und Problemlagen der Pflegebedürftigen bislang zu wenig Beachtung geschenkt wurde und die resultierenden Anwendungen somit keinen wirklichen pflegerischen Mehrwert bieten.²⁶ Ausgangs- und Orientierungspunkt bei der Entwicklung neuer Technologien sollte deshalb nicht die technische Machbarkeit sein, sondern die

21 Vgl. z.B. Hans-Peter de Ruiter/Joan Liaschenko/Jan Angus, Problems with the Electronic Health Record, in: Nursing Philosophy 17/2016, S. 49–58.

22 Manfred Hülsken-Giesler, Der Zugang zum Anderen. Zur theoretischen Rekonstruktion von Professionalisierungsstrategien pflegerischen Handelns im Spannungsfeld von Mimesis und Maschinenlogik, Osnabrück 2008.

23 Vgl. ebd.

24 Ebd., S. 26

25 Vgl. Alexandra Manzei, Neue betriebswirtschaftliche Steuerungsformen im Krankenhaus: wie durch die Digitalisierung der Medizin ökonomische Sachzwänge in der Pflegepraxis entstehen, in: Pflege und Gesellschaft 14/2009, S. 38–53.

26 Vgl. Krings et al. (Anm. 14).

tatsächlichen Unterstützungsbedarfe potenzieller Nutzerinnen und Nutzer sowie anderweitig Betroffener, die es partizipativ in den Entwicklungsprozess einzubeziehen gilt.²⁷

FAZIT

Gradmesser für die gesellschaftliche Relevanz aktueller Entwicklungen in der KI ist weniger der erreichte Grad an Künstlicher Intelligenz, die eher als abstrakter Oberbegriff für sehr unterschiedliche Verhaltensmerkmale „intelligenter“ Systeme firmiert, sondern deren Fähigkeit, zunehmend autonom, das heißt unabhängig von menschlicher Steuerung, zu agieren (und das möglichst in komplexen Umgebungen). Insbesondere das Maschinelle Lernen ist dafür eine wichtige Voraussetzung; benötigt werden aber auch – zumindest bei Systemen, die sich in der analogen Welt bewegen – Wahrnehmungs-, Planungs- sowie Manipulationsfertigkeiten, die bislang erst unzureichend entwickelt und systemisch integriert sind. Die Relevanz dieser entgrenzenden Entwicklungen steht dennoch außer Frage. Wie das Beispiel Pflege zeigt, wirft die perspektivische Einbettung sich verselbstständigender Maschinen in menschliche Lebens- und Handlungskontexte viele normative Fragen auf und ist mit großen moralischen Unsicherheiten verbunden. Es scheint also unbedingt geboten, sich frühzeitig mit dieser Entwicklung zu befassen. Dass derzeit noch weitgehend unklar ist, ob, wann und in welcher konkreten Form die Robotik in den pflegerischen Alltag Einzug halten wird, steht dem nicht entgegen, sondern verweist im Gegenteil auf die Möglichkeit einer vorausschauenden Gestaltung dieser Technisierungsprozesse.

Ein zentrales Element dabei ist die Technikgestaltung im engeren Sinne, und zwar primär orientiert an den Bedürfnissen der Pflegebedürftigen. Ebenso wichtig ist aber, die sozialen und diskursiven Treiber der Entwicklung nicht außer Acht zu lassen, wie gerade die Pflege deutlich macht. Zwar konnten im Bereich Robotik in den vergangenen Jahren zweifelsohne etliche Fortschritte erzielt werden; die hohen und teils

überzogenen Erwartungen an das Lösungspotenzial robotischer Pflegesysteme vermögen diese jedoch nicht wirklich plausibel zu machen. Hinter den fortschreitenden Bestrebungen, Aspekte der Pflegearbeit zu technisieren, stehen vielmehr starke gesellschaftliche Interessen. Dazu gehören beispielsweise die seit Längerem laufenden Bemühungen, die noch vor wenigen Jahrzehnten fast vollumfänglich informell geleistete Pflegearbeit zu professionalisieren und evidenzbasiert auszurichten, nicht zuletzt aus Kostengründen. Der demografische Wandel mit seinen enormen Herausforderungen verstärkt derartige Tendenzen und lässt Hightech-Lösungen wie die Robotik mit ihrem Versprechen auf eine effizientere Ausgestaltung der Pflegearbeit für viele als fast schon unausweichlich erscheinen.

Vor diesem Hintergrund erscheinen nicht nur Forschung und Entwicklung, auch die gesellschaftliche Debatte zur Robotik in der Pflege allzu einseitig auf technische Machbarkeitsvisionen fixiert und zu wenig damit beschäftigt, was überhaupt wünschenswerte Entwicklungen sind. Ob Roboter zu guter Pflege beitragen können, ist letztlich weniger eine Frage des technischen Fortschritts als eine der sinnvollen Innovation von Arbeitsprozessen sowie institutioneller Rahmenbedingungen. Die Gestaltung guter Pflege ist deshalb in grundlegender Weise als eine gesamtgesellschaftliche Aufgabe zu sehen. Wie könnte also ein verantwortungsvoller Umgang mit den neuen technischen Möglichkeiten aussehen? Im TAB-Projekt „Robotik und assistive Neurotechnologien in der Pflege“²⁸ wurden dazu politische Handlungsmöglichkeiten herausgearbeitet, die hier nicht vertieft werden können. Nur so viel: Neben der Anpassung rechtlicher Rahmensetzungen (in den Bereichen Sicherheit, Datenschutz und Haftung) sowie der Förderung bedarfsorientierter Technikentwicklung erscheint als vordringliche Aufgabe, eine breite, auch öffentliche Debatte darüber anzustoßen, welche Rolle die Robotik in Pflegekontexten zukünftig spielen soll – und zwar auf Basis einer möglichst realistischen Einschätzung der Potenziale und Grenzen eines pflegerischen Robotereinsatzes.

CHRISTOPH KEHL

hat zur Praxis und Theorie der biomedizinischen Gedächtnisforschung promoviert und ist Mitarbeiter des Büros für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB).

kehl@tab-beim-bundestag.de

²⁷ Vgl. Diego Compagna/Stefan Derpmann, Verfahren partizipativer Technikentwicklung, Working Papers kultur- und techniksoziologische Studien 4/2009.

²⁸ Der entsprechende TAB-Arbeitsbericht Nr. 177 ist derzeit in Vorbereitung.

MASCHINENETHIK UND „ARTIFICIAL MORALITY“: KÖNNEN UND SOLLEN MASCHINEN MORALISCH HANDELN?

Catrin Misselhorn

Maschinenethik ist ein neues Forschungsgebiet an der Schnittstelle von Informatik und Philosophie, das die Entwicklung moralischer Maschinen zum Ziel hat. Es geht darum, auf der Grundlage von Computertechnologie Maschinen zu gestalten, die selbst moralische Entscheidungen treffen und umsetzen können. Beflügelt wird dieses Vorhaben von den jüngsten Entwicklungen der Künstlichen Intelligenz. Sollen im Rahmen der Maschinenethik Verfahren der Künstlichen Intelligenz eingesetzt werden, so spricht man analog zu „Artificial Intelligence“ (AI) von „Artificial Morality“ (AM).

Während AI zum Ziel hat, die kognitiven Fähigkeiten von Menschen zu modellieren oder zu simulieren, geht es bei der AM darum, künstliche Systeme mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten. Dieses Vorhaben wird von einigen euphorisch begrüßt, während andere dadurch einen menschlichen Kernbereich bedroht sehen. Im Folgenden werden zunächst einige Anwendungsbereiche moralischer Maschinen vorgestellt. Sodann wird thematisiert, inwiefern man überhaupt davon sprechen kann, dass Maschinen moralisch handeln können. Zum Abschluss werden ausgewählte Argumente diskutiert, die dafür oder dagegen sprechen, Maschinen mit der Fähigkeit zum moralischen Handeln auszustatten.

ANWENDUNGSGEBIETE MORALISCHER MASCHINEN

Je komplexer und autonomer künstliche Systeme werden, desto eher müssen sie in der Lage sein, ihr Verhalten in einem gewissen Rahmen selbst zu regulieren. Das bringt es mit sich, dass sie auch in Situationen geraten, die moralische Entscheidungen verlangen. Die scheinbar einfachste Alternative zu Systemen, die diese Kontrolle selbst ausüben können, besteht darin, die permanente

Überwachung und „Online“-Kontrolle durch einen menschlichen Benutzer zu fordern, der dann die moralisch relevanten Entscheidungen trifft. In vielen Bereichen wird dies allerdings kaum möglich sein, sollen Maschinen ihren Zweck optimal erfüllen – sei es aufgrund von Personalmangel, weil schnelle Entscheidungen von Nöten sind, weil die Einsatzsituationen zu gefährlich sind oder weil menschliches Eingreifen selbst einen Risikofaktor darstellt. Welcher Art die moralischen Entscheidungen sind, die ein System treffen muss, hängt vom Anwendungskontext ab.

Ein Anwendungsbereich für moralische Maschinen ist die Altenpflege. Aufgrund des demografischen Wandels wird der Anteil pflegebedürftiger Menschen in den nächsten Jahrzehnten stark zunehmen.⁰¹ Künstliche Systeme werden immer wieder als eine Möglichkeit ins Spiel gebracht, um dem Pflegenotstand entgegenzutreten.⁰² Doch Systeme, die in diesem Kontext eingesetzt werden sollen, stehen vor moralischen Entscheidungen, beispielsweise: Wie häufig und eindringlich soll ein Pflegesystem an Essen und Trinken sowie die Einnahme von Medikamenten erinnern? Wann sollte ein Pflegesystem die Angehörigen verständigen oder den medizinischen Dienst rufen, wenn jemand sich eine Zeitlang nicht rührt? Soll das System den Nutzer rund um die Uhr überwachen, und wie ist mit den dabei erhobenen Daten zu verfahren?

In all diesen Situationen muss ein künstliches System zwischen bestimmten moralischen Werten abwägen: im ersten Fall zwischen der Selbstbestimmung des Nutzers und bestimmten gesundheitlichen Risiken, die entstehen, wenn er seine Medikamente nicht wie vorgeschrieben einnimmt. Im zweiten Fall zwischen der Selbstbestimmung des Nutzers, der Sorge der Angehörigen, die vielleicht gerne sofort informiert würden, und erneut der Gesundheit. Im dritten Fall geht es wiederum um die Selbstbestimmung des Nut-

zers, Gesundheit, die Sorge der Angehörigen sowie um die Privatheit seiner Daten.

Ein zweites viel diskutiertes Beispiel für die Notwendigkeit moralischer Maschinen ist das autonome Fahren. Auch vollautomatisierte Fahrzeuge stehen vor moralischen Entscheidungen. So gilt es beispielsweise, diese so zu programmieren, dass in unvermeidlichen Gefahrensituationen der Schutz menschlichen Lebens Vorrang vor Sach- und Tierschäden besitzt.⁰³ Doch auch Tiere sollten nach Möglichkeit verschont werden. Eine besondere Schwierigkeit stellen die in diesem Anwendungsbereich unter Umständen auftretenden moralischen Dilemmata dar, bei denen beispielsweise eine Entscheidung darüber getroffen werden muss, ob eine geringe Zahl an Menschenleben aufs Spiel gesetzt werden darf, um eine größere Zahl zu retten, wenn dies unvermeidbar ist.

Nicht zu vergessen sind schließlich die militärischen Anwendungen. Der Traum besteht darin, dass keine Soldaten mehr auf dem Schlachtfeld ihr Leben aufs Spiel setzen müssen, sondern an ihrer Stelle autonome Maschinen in den Kampf geschickt werden. Diese sollen mit dem Kriegsvölkerrecht und kontextspezifischen Einsatzregeln ausgestattet werden, die ihren Handlungsspielraum begrenzen und sicherstellen, dass sie sich rechtlich und moralisch einwandfrei verhalten.⁰⁴ So müssen sie entscheiden, wann eine Aktion militärisch notwendig und angemessen ist und wie sich Kombattanten von Zivilisten unterscheiden lassen.

Man könnte allerdings argumentieren, dass es nicht das Pflegesystem, das autonome Auto oder der Kampfroboter ist, die in diesen Fällen eine moralische Entscheidung treffen, sondern die Programmierer dieser Geräte. Doch je größer die Fortschritte der Künstlichen Intelligenz werden, desto weniger können die Entwickler planen und vorhersehen, welche Entscheidungen ein System in einer spezifischen Situation treffen wird. So spielt schon ein Schachprogramm weit besser als seine Pro-

grammierer, die nicht jeden einzelnen Zug des Systems vorhersagen können. Das gilt umso mehr für ein so komplexes System wie AlphaGo Zero, das zunächst nur die Grundregeln des Spiels Go kennt und dann anhand des Durchspielens einer Vielzahl von Partien gegen sich selbst zu den optimalen Entscheidungsstrategien findet. In kürzester Zeit gelang es diesem System, seinen Vorgänger AlphaGo zu schlagen, der als erstes künstliches System einige der weltbesten menschlichen Go-Spieler besiegte.

Doch selbst wenn man zugesteht, dass es in vielen Anwendungsbereichen sinnvoll wäre, wenn Maschinen moralisch handeln könnten, ist damit noch nicht geklärt, ob sie dazu auch in der Lage sind. Die erste Frage ist, ob autonome Systeme überhaupt handeln können. Die zweite ist, ob die Handlungen künstlicher Akteure als moralisch gelten können.

KÖNNEN KÜNSTLICHE SYSTEME MORALISCH HANDELN?

Die Problematik der grundsätzlichen Handlungsfähigkeit lässt sich innerhalb der philosophischen Handlungstheorie entlang zweier Achsen beschreiben: der Fähigkeit, als selbstursprüngliche Quelle des eigenen Tuns zu fungieren, sowie der Fähigkeit, sich an Gründen zu orientieren.⁰⁵ Beide Fähigkeiten müssen als graduelle Attribute begriffen werden, das heißt, sie kommen verschiedenen Arten von Akteuren in unterschiedlichem Maße zu.⁰⁶

Der Begriff der Selbstursprünglichkeit wurde von der philosophischen Tradition teilweise im Sinn der Akteurskausalität verstanden, das heißt, dass eine Handlung von einem Akteur ohne vorhergehende Ursache initiiert wird.⁰⁷ Ein metaphysisch so anspruchsvoller und umstrittener Begriff der Selbstursprünglichkeit ist jedoch nicht zwingend.⁰⁸ Man kann eine einfache Form der Selbstursprünglichkeit auch dann als gegeben se-

01 Vgl. Heinz Rothgang et al., Bertelsmann-Stiftung Themenreport „Pfleger 2030“, Gütersloh 2012.

02 Vgl. Deutschlands Zukunft gestalten. Koalitionsvertrag zwischen CDU, CSU und SPD, 2013, S. 60, https://www.bundesregierung.de/Content/DE/_Anlagen/2013/2013-12-17-koalitionsvertrag.pdf?__blob=publicationFile.

03 Vgl. Ethik-Kommission Automatisiertes und Vernetztes Fahren, Bericht Juni 2017, S. 17, <https://www.bmvi.de/SharedDocs/DE/Publikationen/G/bericht-der-ethik-kommission.html>.

04 Vgl. Ronald Arkin, *Governing Lethal Behavior in Autonomous Robots*, Boca Raton u. a. 2009.

05 Vgl. Catrin Misselhorn, *Robots as Moral Agents*, in: Frank Rövekamp/Friederike Bosse (Hrsg.), *Ethics in Science and Society: German and Japanese Views*, München 2013, S. 30–42.

06 Vgl. Catrin Misselhorn, *Collective Agency and Cooperation in Natural and Artificial Systems*, in: *Philosophical Studies Series* 122/2015, S. 3–25.

07 Die Konzeption der Akteurskausalität wird zumeist auf Aristoteles zurückgeführt. In jüngerer Zeit wurde sie beispielsweise einflussreich von Roderick Chisholm vertreten, *Human Freedom and the Self*. Lindley Lecture, Lawrence 1964.

08 Vgl. Peter F. Strawson, *Freedom and Resentment*, in: *Proceedings of the British Academy* 48/1962, S. 1–25.

hen, wenn ein System mit der Umwelt interagiert (*Interaktivität*), dabei eine gewisse Anpassungsfähigkeit an sich ändernde Bedingungen aufweist (*Adaptivität*) und in der Lage ist, eine Aktivität ohne direkte menschliche Intervention aufzunehmen (*basale Autonomie*).⁰⁹

Über eine solche primitive Form der Selbstursprünglichkeit können auch Maschinen verfügen, insbesondere solche, die von Computern gesteuert werden. Zwar gibt ein Programm vor, wie sich eine Maschine zu verhalten hat, aber im Einzelfall agiert sie, ohne dass ein Mensch eigens eingreift. Werden Verfahren der KI, beispielsweise Maschinelles Lernen, eingesetzt, so ist es sogar die Aufgabe der Maschine, das moralisch angemessene Verhalten selbst aus den Daten zu erschließen.

Die zweite Achse, die Fähigkeit, sich an Gründen zu orientieren, hängt eng mit der Möglichkeit zusammen, Informationen zu verarbeiten. Dem klassischen Humeschen Modell der Handlungsfähigkeit zufolge besteht der Grund einer Handlung in der Kopplung einer Überzeugung mit einer Pro-Einstellung, beispielsweise einem Wunsch: Ich gehe in die Bibliothek, weil ich ein bestimmtes Buch ausleihen will und der Überzeugung bin, dass es in der Bibliothek vorhanden ist.¹⁰ Hinzu kommt nach manchen Ansätzen eine Intention, die dafür verantwortlich ist, dass der Wunsch auch mithilfe eines Plans in die Tat umgesetzt wird.¹¹

Einige Autoren sind der Auffassung, dass ein Handeln aus Gründen lediglich erfordert, dass das Verhalten eines Systems auf diese Art und Weise interpretiert werden kann.¹² Andere erachten dies nicht als hinreichend, sondern gehen davon aus, dass ein künstliches System auch über gewisse innere Zustände verfügen muss, die den Gründen entsprechen.¹³ Doch selbst dann ist das Humesche Modell übertragbar auf künstliche Systeme, sofern diese über die entsprechenden Repräsentationen verfügen.

Ein künstliches System kann als funktional äquivalent zu einem menschlichen Akteur verstanden werden, wenn es über Zustände verfügt, denen eine analoge Funktion zukommt, wie Meinungen, Wünschen und Intentionen beim Menschen. Insbesondere Systeme die auf dem sogenannten BDI-Software-Modell (Belief–Desire–Intention) beruhen, operieren mit symbolischen Repräsentationen, die als funktional äquivalent mit Meinungen, Wünschen und Intentionen gelten können.¹⁴ Das ist ausreichend, um ihnen in einem funktionalen Sinn die Fähigkeit zuzuschreiben, aus Gründen zu handeln. Künstliche Systeme, die zu selbstursprünglichem Handeln aus Gründen in der Lage sind, können als Akteure gelten.

Moralische Handlungsfähigkeit wiederum liegt in einfacher Form vor, wenn die Gründe, nach denen ein System handelt, moralischer Natur sind. Dies ist auf einer rudimentären Ebene schon dann gegeben, wenn ein System über Repräsentationen moralischer Werte verfügt, die die zuvor angegebenen basalen Bedingungen für das Handeln aus Gründen erfüllen (das heißt, es gibt funktionale Äquivalente zu moralischen Überzeugungen, moralischen Pro-Einstellungen und Intentionen). Wenn ein System beispielsweise den Wert der Patientenautonomie als Pro-Einstellung besitzt und zu der Überzeugung kommt, dass dieser Wert in einer bestimmten Situation verletzt wird, dann wird es versuchen, so auf die Situation einzuwirken, dass dieser Wert wieder realisiert wird.

Vollumfängliche moralische Handlungsfähigkeit, wie sie Menschen typischerweise besitzen, kommt künstlichen Systemen allerdings nicht zu. Zum einen ist der Einsatzbereich von Maschinen mit Moral normalerweise auf einen bestimmten Anwendungskontext beschränkt, die menschliche Moralfähigkeit umfasst jedoch potenziell jeden beliebigen Bereich des Lebens.

Zudem verfügen künstliche Systeme bislang nicht wirklich über Bewusstsein und Willensfreiheit. Bewusstsein im Sinn des subjektiven Erlebens wäre beispielsweise erforderlich, um moralische Emotionen wie Mitgefühl oder auch Schuldgefühle empfinden zu können.¹⁵ Willensfreiheit eröffnet

09 Vgl. Luciano Floridi/J. W. Sanders, On the Morality of Artificial Agents, in: *Minds and Machines* 14/2004, S. 349–379.

10 Vertreter dieser Standardauffassung sind z. B. Donald Davidson, *Essays on Actions and Events*, Oxford–New York 1980, sowie Fred Dretske, *Explaining Behavior. Reasons in a World of Causes*, Cambridge 1995⁴.

11 Vgl. Michael E. Bratman, *Intention, Plans, and Practical Reason*, Cambridge MA 1987.

12 Vgl. Daniel C. Dennett, *The Intentional Stance*, Cambridge MA 1987.

13 Vgl. Christian List/Philip Pettit, *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford 2011.

14 Das BDI-Modell greift auf die Arbeiten zur praktischen Rationalität des Philosophen Michael E. Bratman zurück (Anm. 11).

15 Einige Forscher vertreten jedoch auch die Annahme, dass es funktionale Äquivalente zu Emotionen auch ohne subjektive Erlebnisqualität geben kann. Vgl. beispielsweise Matthias Scheutz, *Architectural Roles of Affect and How to Evaluate Them in Artificial Agents*, in: *International Journal of Synthetic Emotions* 2/2011, S. 48–65.

die Möglichkeit, sich auch gegen eine als moralisch erkannte Handlungsoption zu entscheiden und unmoralisch zu handeln. Auch diese Fähigkeit besitzen künstliche Systeme bislang nicht und sollten sie zum Schutz des Nutzers auch nicht haben.

Außerdem können künstliche Systeme anders als Menschen ihre moralischen Entscheidungen und die ihnen zugrundeliegenden Werte nicht reflektieren, diese begründen oder gar selbstständig verändern. Aus diesem Grund können sie auch keine moralische Verantwortung übernehmen. Maschinen stellen demnach einen Sonderfall dar, in dem moralisches Handeln ohne moralische Verantwortung gegeben ist. Nicht zuletzt stellt sich die Frage, ob wir Maschinen gegenüber, die über Bewusstsein, Willensfreiheit oder die Fähigkeit zur moralischen Reflexion verfügen, nicht auch moralische Verpflichtungen haben. Das brächte Komplikationen für ihren praktischen Einsatz mit sich. Damit ist nicht gesagt, dass es unmöglich ist, Systeme zu konstruieren, die auch diese Eigenschaften besitzen. Sie sind jedoch nicht notwendig und auch nicht wünschenswert für Systeme mit moralischer Handlungsfähigkeit in einem grundlegenden und für die Anwendungspraxis geeigneten Sinn.

SOLLEN KÜNSTLICHE SYSTEME MORALISCH HANDELN?

Ein Argument, das zugunsten moralischer Maschinen vorgebracht werden kann, ist ihre schiefe Unvermeidbarkeit. Die Entwicklung zunehmend intelligenter und autonomer Technologien führt demnach zwangsläufig dazu, dass diese mit moralisch problematischen Situationen konfrontiert sind. Deshalb ist es erforderlich, Maschinen zu entwickeln, die über ein gewisses Maß an eigenständiger moralischer Entscheidungsfähigkeit verfügen. Wenn das stimmt, ergibt sich die Notwendigkeit der Maschinenethik aus dem technologischen Fortschritt.

Zudem könnten moralische Maschinen möglicherweise schlicht bessere Maschinen sein. Technologien sollen das menschliche Leben erleichtern. Moralische Maschinen dienen diesem Ziel besser, so könnte man argumentieren, weil sie menschlichen Bedürfnissen und Werten besser entsprechen. Während neue Technologien häufig ins Fadenkreuz der Kritik geraten, weil sie moralische Probleme aufwerfen, ist die Moral in diesem Fall im Design der Maschine verankert.

Als Vorteil künstlicher moralischer Akteure wird weiterhin angeführt, dass sie moralisch besser handeln als Menschen, weil sie keinen irrationalen Impulsen oder emotionalem Stress unterliegen. Sie sind nicht verführbar und werden auch nicht durch Eigeninteresse vom Pfad der Moral abgebracht. Nicht zuletzt können sie in Sekundenbruchteilen Entscheidungen treffen, in denen ein Mensch gar nicht mehr zu bewusstem Entscheiden in der Lage ist. Das spricht für manche Autoren dafür, ihnen moralische Entscheidungen in besonders prekären Situationen zu überlassen, beispielsweise im Krieg.¹⁶

Neben ihrem praktischen Nutzen verspricht die Entwicklung der Maschinenethik auch eine Verbesserung der Moral selbst. Die menschliche Moral ist fragmentiert und teilweise widersprüchlich. Die Entwicklung künstlicher Systeme mit moralischen Fähigkeiten macht es erforderlich, die menschliche Moral (zumindest in den Anwendungsbereichen) zu vereinheitlichen und konsistent zu machen, weil künstliche Systeme nur auf dieser Grundlage operieren können. Insofern Einheitlichkeit und Widerspruchsfreiheit generell theoretische Tugenden darstellen, wäre das auch ein Fortschritt der Ethik als Theorie der Moral.

Zudem besteht die Hoffnung, dass der Versuch, künstliche Systeme mit moralischen Fähigkeiten zu konstruieren, auch Rückschlüsse darüber zulässt, wie moralische Fähigkeiten bei Menschen funktionieren. Im besten Fall gibt es grundlegende funktionale Strukturen moralischer Fähigkeiten, die sowohl in natürlichen als auch in künstlichen Systemen realisiert werden können. Scheitern gewisse Erklärungsansätze moralischer Fähigkeiten an der Implementation, so ist auch das zumindest im negativen Sinn aufschlussreich im Hinblick darauf, wie menschliche Moralfähigkeiten nicht funktionieren. Maschinenethik besitzt also einen Wert als Instrument wissenschaftlicher Erkenntnis.

Diesen Pluspunkten steht jedoch eine Reihe von Einwänden gegenüber. So stellt sich die Frage, ob die Vereinheitlichung und Auflösung von Widersprüchen in unserer Alltagsmoral stets zu begrüßen ist. Möglicherweise werden dadurch Problemlagen eliminiert, ohne dass dies der Komplexität und existenziellen Bedeutung moralischer Situationen im Alltag gerecht wird. So ist beispielsweise nicht klar, wie sich ein autonomes Fahrzeug entscheiden sollte, wenn es ausschließlich die beiden Handlungsalternativen hat, das Le-

¹⁶ Vgl. Arkin (Anm. 4), S. 33.

ben seiner Insassen aufs Spiel zu setzen oder dasjenige von auf der Straße spielenden Kindern.¹⁷ Der Zwang zu einer Entscheidung *ex ante* erscheint in einem solchen Fall als problematisch.

Ein Mensch hätte die Freiheit, dies situativ zu entscheiden. Doch das Verhalten eines autonomen Systems ist im Vorhinein festgelegt. Eine solche Normierung führt zu einer bedeutenden Einschränkung der Entscheidungsfreiheit des Einzelnen, der in einer derartig existenziellen Situation nicht mehr selbst entscheiden könnte, ob er sein Leben für die Kinder aufs Spiel setzt oder nicht.

Ein anderer Einwand befürchtet die Entstehung von Verantwortungslücken.¹⁸ An der Entwicklung, dem Vertrieb und der Nutzung moralischer Maschinen sind viele Menschen beteiligt. Die Softwareentwicklung umfasst Teams, sie baut auf bereits vorhandene Module auf, und nicht immer gibt es eine Person, die den Überblick über den gesamten Programmierprozess besitzt. Weitere Einflussfaktoren bilden Marketing und Vertrieb sowie schließlich die Nutzer. So ist zu befürchten, dass sich moralisch problematische Vorfälle häufig nicht auf eine einzige Handlung oder Entscheidung zurückführen lassen, sondern Ergebnis vieler ineinandergreifender Handlungen und Entscheidungen sein werden, an denen unterschiedliche Akteure beteiligt sind.

Dieses Problem der „vielen Hände“ tritt auch im Fall anderer Technologien auf.¹⁹ Es verschärft sich jedoch, je autonomer Maschinen agieren. Dadurch erhöht sich das Risiko, dass die Maschinen zu Entscheidungen kommen, die niemand beabsichtigt oder vorhergesehen hat und über die niemand direkte Kontrolle besitzt. Das könnte systematisch dazu führen, dass niemand für die moralisch desaströse Entscheidung eines künstlichen Systems verantwortlich gemacht werden kann. Die Frage ist, wie sich das auf unsere mo-

ralische Praxis auswirken würde. Die schlimmste Befürchtung wäre, dass wir den moralischen Verdikten einer Technologie ausgeliefert wären, die wir letztlich nicht mehr durchschauen und kontrollieren könnten.

Wie diese Argumente gegen moralische Maschinen zu bewerten sind, hängt auch vom Anwendungsbereich ab. So entsteht das Problem der Verantwortungslücke möglicherweise verstärkt bei sehr komplexen und vernetzten Anwendungsbereichen wie dem autonomen Fahren und weniger ausgeprägt im Pflegebereich. Auch das Problem der Entmündigung könnte im Pflegebereich dadurch entschärft werden, dass sich ein häusliches Pflegesystem an den Moralvorstellungen des Nutzers orientieren kann und in der Lage ist, sich flexibel an diese anzupassen.²⁰ Zudem müsste sichergestellt sein, dass niemand gezwungen wird, ein solches System zu nutzen, der dies nicht möchte.

Die letzten beiden Optionen bestehen im Hinblick auf das autonome Fahren und militärische Anwendungen hingegen nicht. In diesen Bereichen muss es allgemeinverbindliche Regelungen geben. Schließlich kann sich niemand dem Einfluss dieser Technologien entziehen, wenn er in irgendeiner Form am Verkehr teilnimmt oder zufällig als Zivilist in einem Kriegsgebiet lebt. Im Gegenzug müssen auch anwendungsspezifische Vorteile in die Abwägung mit eingehen. So wird zugunsten des autonomen Fahrens beispielsweise der Gewinn an Sicherheit angeführt. Autonome Waffensysteme hingegen sollen der humaneren Kriegsführung dienen, insofern sie eine bessere Einhaltung des Kriegsvölkerrechts gewährleisten und sich auf diese Art und Weise beispielsweise positiv auf den Schutz von Zivilisten auswirken. Pflegesysteme hingegen können die Lebensqualität alter Menschen steigern, die möglichst lange zu Hause leben möchten. Offen bleibt, auf welcher ethischen Grundlage künstliche Systeme entscheiden sollten. Auch das hängt vom Anwendungsbereich ab und sollte Gegenstand eines gesellschaftlichen Diskurses sein, insbesondere in denjenigen Anwendungsbereichen, die allgemeinverbindliche Regelungen erfordern.

CATRIN MISSELHORN

ist Direktorin des Instituts für Philosophie der Universität Stuttgart und hat den Lehrstuhl für Wissenschaftstheorie und Technikphilosophie inne.
catrin.misselhorn@philo.uni-stuttgart.de

17 Vgl. Ethik-Kommission Automatisiertes und Vernetztes Fahren (Anm. 3), S. 17

18 Vgl. Robert Sparrow, Killer Robots, in: *Journal of Applied Philosophy* 24/2007, S. 62–77.

19 Vgl. Bataya Friedman, Moral Responsibility and Computer Technology, Paper Presented at the Annual Meeting of the American Educational Research Association, Boston 1990; Helen Nissenbaum, Computing and Accountability, in: *Communications of the Association for Computing Machinery* 37/1994, S. 72–80.

20 Vgl. Catrin Misselhorn, Moralische Maschinen in der Pflege? Grundlagen und eine Roadmap für ein moralisch lernfähiges Altenpflegesystem, in: Christiane Woopen/Marc Jannes (Hrsg.), *Roboter in der Gesellschaft: Technische Möglichkeiten und menschliche Verantwortung*, Wiesbaden 2018 (i. E.).

ÜBERLEGUNGEN ZUR DISZIPLIN DER MASCHINENETHIK

Oliver Bendel

Mit dem rechtlichen und moralischen Status von Maschinen mit Chips beschäftigt man sich in der Wissenschaft schon seit den 1950er Jahren. Die Science-Fiction-Literatur war noch früher dran. Lange Zeit ging es vor allem darum, ob Roboter Objekte der Moral sind, sogenannte *moral patients*, ob man ihnen etwa Rechte zugestehen kann. Ich befasse mich seit den 1990er Jahren mit diesem Thema. Ich habe damals keinen Grund gesehen, Robotern Rechte zu geben, und bis heute nicht meinen Standpunkt geändert. Wenn sie eines Tages etwas empfinden oder wenn sie leiden können, oder wenn sie eine Art von Lebenswillen haben, lasse ich mich bestimmt überzeugen. Aber im Moment bemerke ich keine entsprechenden Tendenzen.

SUBJEKTE DER MORAL

Die Diskussion, ob Roboter Subjekte der Moral sind, das heißt, ob von ihnen moralische Handlungen ausgehen können oder sie bestimmte Pflichten haben, folgte in den 2000er Jahren. Ich tue mir in diesem Zusammenhang schwer mit dem letztgenannten Begriff. Ich habe von „Verpflichtungen“ gesprochen, bin mir jedoch nicht sicher, ob das Problem dadurch wirklich gelöst wird. In der Monografie „Moral Machines“ (2009) von Wendell Wallach und Colin Allen⁰¹ wurden Roboter in systematischer Weise als *moral agents* besprochen und Einteilungen und Unterscheidungen der moralischen Subjekte angeboten. In dem zwei Jahre später folgenden Sammelband „Machine Ethics“ von Michael Anderson und Susan Leigh Anderson wurde der Diskurs fortgeführt.⁰²

Seit den 1980er Jahren spielt für mich die Tierethik (mithin Tierschutz) eine große Rolle.⁰³ Tiere sind für mich Objekte der Moral, keine Subjekte. Sicherlich haben Schimpansen oder Elefanten vormoralische Qualitäten, und weil wir von Tieren abstammen, eventuell solche sind,

dürfte es fließende Übergänge geben. Jedenfalls können Tiere nicht gut oder böse sein, und sie können ebenso wenig unterschiedliche Handlungsmöglichkeiten aus moralischer Perspektive beurteilen und sich dann beispielsweise für das geringste Übel entscheiden. Wir können „Du böser Hund!“ sagen, wenn er uns gebissen hat, aber wir meinen das nicht moralisch, höchstens pädagogisch.

GUTE UND BÖSE MASCHINEN

Ebenso wie Tiere halte ich auch Maschinen nicht für gut oder böse, zumindest nicht im Sinne der Philosophin Annemarie Pieper, die einen bösen oder guten Willen voraussetzt (ob Handlungen an sich gut oder böse sind, kann man dennoch diskutieren).⁰⁴ Aber offenbar können sie blitzschnell Optionen erfassen und dann nach einer vorgegebenen moralischen Regel entscheiden oder die moralischen Folgen abschätzen und dann auswählen. In diesem Sinne steht ein hochentwickelter Roboter (oder ein hochentwickeltes System der Künstlichen Intelligenz, kurz KI) zwischen Mensch und Tier. Das mag irritieren, und ich weise darauf hin, dass er in einem anderen Sinne ganz woanders steht. Als Objekt der Moral taucht er im Moment gar nicht auf. Man darf ihn behandeln, wie man will, ihn schlagen und zerstören, und wenn das jemand bedauert, ist eben der, der es bedauert, das Objekt der Moral, nicht der Roboter. Ein Spiel über Bande sozusagen, wie bei Teddybären und heiligen Steinen oder Bergen. Diese kommen nur in die Sphäre der Moral, weil sie jemandem gehören oder jemanden interessieren.

Nach dem, was bisher gesagt wurde, sind manche Maschinen neue, fremde, merkwürdige Subjekte der Moral. So habe ich es in den vergangenen Jahren vertreten. Sie sind nicht gut oder böse, zumindest nicht in Bezug auf einen Willen

(den sie nicht haben), und sie haben keine Pflichten im engeren Sinne (höchstens „Verpflichtungen“, im Sinne von Aufgaben, die man an sie delegiert, und vielleicht nicht einmal die). Wenn sie eine Verantwortung tragen, dann allenfalls, weil ihnen etwas anvertraut oder übereignet wurde. Eine maschinelle Primärverantwortung könnte das sein, die noch genauer zu untersuchen wäre. Bei der Sekundärverantwortung wird es schon schwieriger, denn wie sollten wir die Maschine zur Rechenschaft ziehen? An den Ohren können wir sie nicht ziehen, und selbst wenn sie welche hat, wird es sie nicht stören. Die Tertiärverantwortung wird eventuell ein Thema, wenn sie ins Rechtliche übergeht. Ein Jurist kann alles konstruieren, sogar eine elektronische Person, die man verklagen und belangen kann. Als Ethiker bin ich hier vorsichtig.

MASCHINELLE MORAL

Den Begriff der maschinellen Moral verwende ich gerne ähnlich wie den der künstlichen Intelligenz. Beide Male meine ich den Gegenstand der Disziplinen. Die Künstliche Intelligenz hat die maschinelle oder künstliche Intelligenz zum Gegenstand. Sie simuliert menschliche Intelligenz oder strebt danach, diese eines Tages in wesentlichen Funktionen abzubilden. Die Maschinenethik hat die maschinelle Moral zum Gegenstand. Sie simuliert derzeit die menschliche Moral. Allerdings konzentriert sie sich in ihren aktuellen Ausprägungen auf gewisse Grundzüge. Die meisten moralischen Maschinen sind wie menschliche Fundamentalisten. Sie halten sich stur an Regeln, die man ihnen eingetrichtert hat. Einige moralische Maschinen vermögen immerhin die Folgen abzuschätzen, die ihre Handlungen nach sich ziehen würden, und unterschiedliche Entscheidungen treffen, die ihnen wiederum beigebracht wurden. Selbstlernende Maschinen könnten indes mehr.

Bisher scheint also festzustehen, dass Maschinen neuartige Subjekte der Moral sein können.

01 Vgl. Wendell Wallach/Colin Allen, *Moral Machines: Teaching Robots Right from Wrong*, New York 2009.

02 Vgl. Michael Anderson/Susan Leigh Anderson (Hrsg.), *Machine Ethics*, Cambridge 2011.

03 Vgl. Ursula Wolf, *Ethik der Mensch-Tier-Beziehung*, Frankfurt/M. 2012.

04 Vgl. Annemarie Pieper, *Einführung in die Ethik*, Tübingen–Basel 2007.

Sie können unterschiedliche Optionen beurteilen und dann Entscheidungen treffen, die moralisch adäquat zu sein scheinen. Die moralischen Fähigkeiten werden ihnen von Menschen beigebracht. Dieser Transferprozess bleibt freilich nicht ohne Folgen. Wir haben es in der Regel mit teilautonomen und autonomen Maschinen zu tun, die alleingelassen sind, die nicht von uns beaufsichtigt werden, die in Situationen geraten, die wir vielleicht vorausgesehen haben, aber doch ein wenig anders sind. Es ergeben sich erste Unschärfen: Moral und Anwendungsfall der Moral passen nicht immer zueinander.

DIE DISZIPLIN DER MASCHINENETHIK

Die Disziplin, die sich mit Maschinen als Subjekten der Moral beschäftigt, die die maschinelle Moral untersucht und hervorbringt, ist die Maschinenethik.⁰⁵ Sie ist ein Pendant zur Menschenethik, die sich mit Menschen als Subjekten der Moral beschäftigt. Oder auch nur ein Ausnahmefall der angewandten Ethik. Sie scheint aber nicht ganz zu den klassischen Bereichsethiken zu passen, zu Informations-, Technik- oder Wirtschaftsethik.⁰⁶ Etwas ist anders, eben die Antwort auf die Frage nach dem Subjekt der Moral. Die Ethik ist für mich, wie gerade deutlich wurde, die Disziplin, die Moral der Gegenstand. Ein Student von mir hat es so ausgedrückt: Ethik treibt man, Moral hat man. Anders als die klassischen Bereichsethiken denkt die Maschinenethik nicht nur über Subjekte der Moral nach, sondern schafft sie im besten Falle auch. Sie bringt, zusammen mit Künstlicher Intelligenz und Robotik, moralische und manchmal unmoralische Maschinen hervor.

Selbstverständlich darf man ganz anders sprechen, darf man Ethik und Moral in eins setzen, wie es im Englischen oft gemacht wird, darf man die Ethik oder die Moral weiter fassen. Wichtig ist mir, dass deutlich wird, was ich meine. Der Rest ist Übersetzungsarbeit. Die Ethik ist also die Disziplin (eine Disziplin der Philosophie), gerne auch eine Theorie oder eine Lehre. Als Moralphilosophie

05 Vgl. Oliver Bendel, *Wirtschaftliche und technische Implikationen der Maschinenethik*, in: *Die Betriebswirtschaft* 4/2014, S. 237–248.

06 Vgl. Oliver Bendel, *Maschinenethik*, in: *Gabler Wirtschaftslexikon*, 2012, <http://wirtschaftslexikon.gabler.de/Definition/maschinenethik.html>.

losophie westlicher Prägung ist sie für mich, Annemarie Pieper folgend, Wissenschaft. Was wir in der Ethik, genauer: in der normativen Ethik, auch finden, sind Modelle wie die Pflicht- oder Pflichtenethik und die Folgenethik. Was wir der Maschine einpflanzen, ist eine maschinelle Moral, eine Moral, die in ihr, der Maschine, funktioniert. Diese Moral können wir einbetten in ein Modell der normativen Ethik.

ARTEFAKTE DER MASCHINENETHIK

Wer sich heute als Moralphilosoph der Maschinenethik verschließt, verkennt deren historische Bedeutung. Zum ersten Mal bauen wir in der Ethik etwas, bringen Artefakte hervor, treiben nicht nur Gedankenexperimente, sondern „Tatsachenexperimente“. Natürlich ist die Frage, ob man jede moralische oder unmoralische Maschine umzusetzen hat. In einem Labor sollte man vielleicht die Atombombe der Maschinenethik bauen, dann aber dort lassen. Vielleicht wäre es zu gewagt, sie zu bauen, weil man sie kaum unter Verschluss halten kann. Aber man sollte der Maschinenethik, wie der Physik, möglichst wenig verbieten. Auch solche Erkenntnisse, die problematisch zu sein scheinen, können sich als nützlich erweisen.

Ich entwickle unter anderem spezielle Chatbots, also Vertreter der Softwareroboter. Zuletzt haben wir einen Hardwareroboter auf die Welt gebracht. Er zeigt den Fokus meiner Forschung: Ich will im Kontext von Maschinenethik und Tierethik (und Tierschutz) tierfreundliche Maschinen erfinden. Saugroboter, Mähmaschinen, Windkraftanlagen, selbstständig fahrende Autos – es existieren zahlreiche Maschinen, die man „moralisieren“ kann. Nicht mit allen sollte man das tun, und bei selbstständig fahrenden Autos bin ich sehr skeptisch, was ihre Entscheidungen und Handlungen uns gegenüber anbelangt, vor allem dann, wenn menschliche Unfallopfer auszuwählen sind.

2018 wollen wir unser viertes Artefakt der Maschinenethik bauen. Nach dem GOODBOT und dem LIEBOT aka LÜGENBOT, zwei Chatbots, nach dem LADYBIRD, einem tierfreundlichen Staubsaugerroboter (der seine Arbeit einstellt, sobald er Marienkäfer erkennt), wird hoffentlich der BESTBOT geboren. Unsere Maschinen halten sich entweder an vorgegebene

Regeln, die sie unmittelbar umsetzen, oder versuchen die Folgen abzuschätzen (um dann wieder vorgegebene Regeln anzuwenden). Hier sind Pflicht- und Folgenethik gefragt. Danach würde ich gerne selbstlernende moralische Maschinen bauen.⁰⁷ Diese wären in der Lage, eine eigene, genuin maschinelle Moral zu entwickeln. Das könnte ausgesprochen gefährlich sein, für Körper und Geist, und es ist nicht mein Bestreben, alle unsere Prototypen zum Vorbild von Produkten zu erklären. Es interessiert mich, was möglich ist, die neue Unschärfe, die entsteht, die maschinelle Moral, die nicht auf die menschliche passt. Unser LIEBOT war zwar nicht lernfähig, aber stark vernetzt, und er hatte genuin maschinelle Strategien des Lügens, wenn er Suchmaschinen und Klassifikationen benutzte, um die Wahrheit herauszufinden, und diese dann nach allen Regeln der Kunst verdrehte. Mich interessiert also auch die Maschine, die sich anders als der Mensch verhält. Der GOODBOT war lernfähig, dies am Rande, ohne seine Moral weiterzuentwickeln. Vielmehr hat er den Benutzer bewertet und eingeordnet und sein Verhalten angepasst, innerhalb eines vorgegebenen Rahmens.

Damit wären wir bei der Frage, was die Maschine kann und soll. Die immer wieder auftauchende Frage, ob man Maschinenethik treiben soll, halte ich für müßig. Soll man Teilchenphysik treiben oder Mikrobiologie? Wenn man damit eine Disziplin so ergänzt oder unterteilt, dass man gezielter forschen kann, und wenn man neue wissenschaftliche Erkenntnisse hervorbringen vermag, durchaus.

AUTOMATISIERTES FAHREN

Es ist schade, dass man Maschinenethik vor allem mit theoretischen Dilemmata in Verbindung bringt, mit philosophischen Gedankenexperimenten, die ungemein wichtig und interessant sind, aber nicht immer zum entscheidenden Schritt führen. Auch praktische Dilemmata sind einzubeziehen, die tatsächlichen Situationen, in die eine Maschine kommt. Wenn man sagt, dass theoretischen Dilemmata der Realitätsbezug fehlt, hat man nicht verstanden, was Gedankenexperimente sind. Das Trolley-Problem ist nicht

⁰⁷ Vgl. Luis Moniz Pereira/Ari Saptawijaya, *Programming Machine Ethics*, Cham 2016.

dafür da, dass es in der Wirklichkeit auftritt. In der Wirklichkeit sind selten fünf Personen auf einem Gleis, die man durch eine Weichenstellung retten kann, wobei man aber einen Unbeteiligten auf einem anderen Gleis opfern muss. Das Gedankenexperiment ist dafür da, unsere moralische Haltung offenzulegen und Modelle der normativen Ethik zu veranschaulichen. Es kann unterschiedliche Ansätze aufzeigen, an eine Sache heranzugehen.

Da wäre zum Beispiel das Qualifizieren und Quantifizieren. Ich habe mir 2012 auf der Grundlage des Trolley-Problems das etwas komplexere Roboterauto-Problem ausgedacht (auf dem Weg und der Straße befinden sich zwei voneinander unabhängige Erwachsene und drei Kinder) und dieses Anfang 2013 auf einer Konferenz zur Technologiefolgenabschätzung in Prag vorgestellt.⁰⁸ Einer meiner Studenten hatte für das autonome Auto des Gedankenexperiments (NAC, New Autonomous Car) eine Formel entwickelt, die qualifizieren und quantifizieren konnte. Es konnte also Menschen anhand ihres Geschlechts, Alters, Aussehens und so weiter beurteilen, sie in diesem Sinne klassifizieren, und sie durchzählen, also beispielsweise potenzielle Unfallopfer berechnen und gegeneinander aufrechnen. Wir haben damals theoretisch festgestellt, dass beide Verfahren problematisch sind. Wenn man qualifiziert, diskriminiert man meist, und wenn man quantifiziert, muss man die Frage beantworten, warum drei Menschen unbedingt mehr wert sein sollen als zwei.

Nun kann man dies praktisch anwenden, man kann im automatisierten Fahren praktische Dilemmata und Gefahrensituationen aller Art voraussehen und die beiden Ansätze implementieren. Damit gelangt man von der Theorie in die Praxis, und da faktisch alle möglichen Gefahrensituationen auftreten, in denen man abwägen und urteilen muss, kann man hier nicht mehr sagen, dass nichts geschehen wird. Natürlich wird etwas geschehen und eine bestimmte Entscheidung getroffen werden müssen. Es ist sogar so: Ein autonomes Auto kann sich, frei nach Paul Watzlawick, nicht nicht entscheiden (der berühmte Kommunikationswissenschaftler hat vom Kom-

munizieren gesprochen). Selbst wenn es in einer Unfallsituation unbeirrt geradeaus fährt, ist das eben eine Entscheidung, die der Entwickler oder Programmierer der Maschine mitgegeben hat. Einige lehnen es ab, bei Maschinen von Entscheidungen zu sprechen. Es wird freilich schwierig bei einer solch extremen Position, überhaupt über sie zu sprechen. Gibt es Roboter, die Fußball spielen? In einem gewissen Sinne nicht.

Die Lösung des praktischen Problems lautet für mich: Man sollte beim hoch- und vollautomatisierten beziehungsweise beim autonomen Fahren vorsichtig sein mit moralischen Regeln, die man dem Auto beibringt. Man sollte zudem vorsichtig sein mit moralischen Fähigkeiten, zu denen das Auto selbst gelangt. Es existieren viele interessante Ansätze, die Maschine dazulernen zu lassen. Man kann sie aufziehen wie ein Kind, ihr in den ersten Monaten einen Fahrlehrer aufzwingen, sie nur in solche Gegenden und Situationen schicken, wo sie sich vorbildliches Verhalten anschaut. Ob das alles zum gewünschten Ziel führt, zu einem tier- und menschenfreundlichen Verhalten, ist die Frage. Ich denke, in den nächsten 20 Jahren ist man gut beraten, wenn man das selbstständig fahrende Auto auf die Autobahnen schickt und ansonsten uns steuern lässt. Sowohl die Stadt als auch die Landstraßen sind hochkomplexe Umgebungen. Ich finde autonome Autos als Maschinenethiker, wie vermutlich klar wurde, vor allem mit Blick auf Tiere interessant.

PFLEGEROBOTER, ZIVILE UND MILITÄRISCHE ROBOTER

Die Maschinenethik beschäftigt sich also vor allem mit (teil-)autonomen Systemen. Dazu gehören auch Pflege- und Therapieroboter oder zivile und militärische Drohnen. Aber selbst 3D-Drucker könnte man moralisieren. Man könnte ihnen beibringen, keine Waffen auszudrucken. Sie müssten wissen, was eine Waffe ist, die Objekte, Vorlagen und Dateien beurteilen und sich dann entsprechend entscheiden können. Bei Pflegerobotern ist zum Beispiel von Bedeutung, ob sie den Patienten töten können sollen, also Sterbehilfe leisten können. Die meisten Krankenhäuser und Pflegeheime dürften dies ablehnen. Der Maschinenethiker interessiert sich weniger für die Frage an sich, denn er ist kein Medizin- und kein Sterbehilfiker. Er interessiert sich vielmehr dafür, wie er die Fähigkeit, einen Befehl aus mo-

⁰⁸ Vgl. Oliver Bendel, *Towards Machine Ethics*, in: Tomáš Michalek/Lenka Hebáková/Leonhard Hennen et al. (Hrsg.), *Technology Assessment and Policy Areas of Great Transitions*, 1st PACITA Project Conference, March 13–15, 2013, Prague 2014, S. 321–326.

ralischen Gründen zu verweigern, konzipieren und implementieren kann. Denn darum könnte es gehen: Der Pflegeroboter weiß, wie man jemanden erwürgt, und er entscheidet sich, das nie zu tun, selbst wenn man ihn inständig darum bittet. Wer das für Science-Fiction hält, sollte sich unterschiedliche Typen von Pflegerobotern anschauen. Solche, die etwas transportieren, solche, die uns informieren, und solche, die Hand an uns legen. Manche sehen aus wie ein Mensch oder ein Tier, andere wie Kooperations- und Kollaborationsroboter aus der Industrie. Diese haben meist einen Arm, mehrere Achsen und zwei bis drei Finger. Man kann sie trainieren, indem man ihren Arm und ihre Finger bewegt oder ihnen einfach etwas vormacht, während sie mit ihren Sensoren und Systemen alles verfolgen und verarbeiten. Im Prinzip kann man ihnen beibringen, uns zu erwürgen, wobei es nicht ohne Grund bestimmte technische Normen gibt, um dies zu verhindern. Eine Aufgabe der Maschinenethik wäre es eben, den Roboter mit einer Form der Befehlsverweigerung vertraut zu machen, die moralisch begründet wäre.

Der militärische Einsatz wurde erwähnt. Wissenschaftler verschiedener Disziplinen erhalten Geld von Verteidigungsministerien. Ein großer Teil fließt in die Entwicklung autonomer Kampfsysteme, in Robotik und Informatik, ein kleiner – wie im Falle des Pentagon – in das künstliche Gewissen, das diese haben sollen, oder in ihre Fähigkeit, den Gegner zu täuschen, zu betrügen und zu verwirren, also in die Maschinenethik. Man kann den Kampfroboter grundsätzlich für unmoralisch halten. Dennoch kann er etwas tun, was man moralisch nennen könnte, etwa mithilfe seines künstlichen Gewissens einen Kollateralschaden vermeiden. Hier zeigt sich, dass der Unterschied zwischen moralischen und unmoralischen Maschinen nicht einfach zu bestimmen ist. Diese Frage muss im Rahmen der Maschinenethik noch intensiver diskutiert werden.

ZUSAMMENFASSUNG UND AUSBLICK

Maschinenethik ist eine junge Disziplin, die man in der Philosophie ansiedeln kann, die aber Partnerinnen wie die Künstliche Intelligenz und Robotik braucht, vor allem dann, wenn sie erfolgreich Artefakte herstellen und erforschen will, Simulationen und Prototypen, die die Möglich-

keit moralischer und unmoralischer Maschinen zeigen. Von solchen darf die Maschinenethik sprechen, so wie die Disziplin der Künstlichen Intelligenz davon sprechen darf, dass sie künstliche Intelligenz hervorbringt. Welche Maschinen man genau moralisieren soll, muss diskutiert und eruiert werden.

Es ist unproblematisch, ja hilfreich, tierfreundliche Staubsaugerroboter zu bauen, die die Moral ihrer Besitzer auch in deren Abwesenheit umsetzen. Sie bewegen sich in geschlossenen oder halboffenen Welten und treffen einfache Entscheidungen. Beim automatisierten Fahren sieht es schon anders aus, und ich bin dagegen, dass Autos potenzielle menschliche Unfallopfer durchzählen oder sie bewerten und dann ihre Urteile fällen. Tierische Verkehrsteilnehmer darf man auf diese Weise behandeln, und es wäre wünschenswert, die Zahl der Getöteten dadurch zu reduzieren. Man könnte sagen, dass auch in offenen Welten überschaubare Situationen entstehen können (hier solche, die sich auf Tiere beziehen), in denen einfache Entscheidungen möglich sind.

Vor der Maschinenethik liegt also ein weites Feld, und sie kann unterschiedliche Richtungen einschlagen. Es ist weniger wichtig, dass sie moralische Regeln begründet. Viel wichtiger ist, dass sie moralisch begründete Regeln in einer befriedigenden Weise implementiert. Dabei kann sie sowohl moralische als auch unmoralische Maschinen erschaffen. Die Ethik erforscht, was gut und böse ist, sie ist nicht gut oder böse. Natürlich darf jeder dazu beitragen, dass die Welt ein bisschen besser wird. Und genau deshalb bin ich persönlich an bestimmten moralischen Maschinen besonders interessiert.

OLIVER BENDEL

ist Professor für Wirtschaftsinformatik und Ethik mit den Schwerpunkten Wissensmanagement, Informationsethik und Maschinenethik an der Hochschule für Wirtschaft FHNW in Basel. Aktuell forscht er über künstliche Stimmen.

oliver.bendel@fhnw.ch

ZWISCHEN MENSCHWERDUNG UND WELTHERRSCHAFT: KÜNSTLICHE INTELLIGENZ IM FILM

Ingo Irsigler · Dominik Orth

„I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.“ – Dieses Zitat aus Stanley Kubricks „2001 – A Space Odyssey“ (1968) stammt – das Verb *disconnect* verweist unmissverständlich darauf – nicht von einer menschlichen Figur. Vielmehr weigert sich mit diesen Worten HAL 9000, der Bordcomputer des Raumschiffs „Discovery 1“, dem Astronauten Dave, der eine kurze Bergungsmission im All unternommen hatte, wieder Zugang zum Raumschiff zu gewähren. Doch Hal⁰¹ ist mehr als ein Computer: Er (oder sollte man schreiben: es?) gilt als „brain and central nervous system of the ship“, wie es im Film heißt, ist also eine Künstliche Intelligenz (KI). Diese steuert den gesamten Schiffsablauf, verantwortet die Öffnung und Schließung der Zugangsluken und kann verhindern, dass Dave wieder an Bord gelangt. Das eingangs genannte Zitat markiert einerseits einen entscheidenden Wendepunkt innerhalb dieses Kino-Klassikers, gleichzeitig steht es beispielhaft für eines der zentralen Handlungselemente von Spielfilmen, die das Thema Künstliche Intelligenz aufgreifen: Die Übernahme der Kontrolle durch die KI, die gleichsam einen Kontrollverlust des Menschen über die Technik bedeutet. Technik und Mensch sind in solchen Narrationen zu Gegenspielern geworden.

Während es Dave aus „2001 – A Space Odyssey“ dennoch gelingt, Hal und somit die von einer Künstlichen Intelligenz ausgehende Bedrohung ab- und damit also auszuschalten, zeigt ein Blick auf die Filmgeschichte der vergangenen Jahrzehnte, dass sich nicht immer das Menschliche durchsetzt: So gelingt es der KI in Filmen der „The Matrix“-Trilogie (1999/2003) oder „Terminator“-Reihe (ab 1984), die Kontrolle über die Menschen (ja sogar die Menschheit) dauerhaft zu übernehmen, während die mit artifizieller

Intelligenz ausgestatteten Roboter und künstlichen Menschen in „Blade Runner“ (1982) oder „A.I. – Artificial Intelligence“ (2001) danach streben, menschlicher (teilweise sogar zu „echten“ Menschen) zu werden.

Bereits diese wenigen Beispiele zeigen, wie vielgestaltig filmische Fiktionen das Thema der Künstlichen Intelligenz ausgestalten. Trotz dieser Vielzahl an Varianten folgen viele „KI-Filme“ – wie wir im Folgenden anhand ausgewählter Beispiele aufzeigen werden – ähnlichen Handlungsmustern und enthalten ähnliche erzählerische Elemente, die den Filmen ein reflexives Potenzial einschreiben. Denn auf welche Art und Weise Künstliche Intelligenz auch thematisiert wird: Vielfach verweisen die Filme auf Technikreflexionen hinsichtlich der KI in der realen Welt und positionieren sich implizit dazu, indem Möglichkeiten, vor allem aber Gefahren und Grenzen der Künstlichen Intelligenz im Modus der Fiktion ausgelotet werden.

ZWEI GRUNDFORMEN KÜNSTLICHER INTELLIGENZ IM SPIELFILM

Grundsätzlich sind zwei Hauptthemen von KI-Filmen zu unterscheiden, die divergierenden Handlungsmustern folgen: Diese Themen können mit den Begriffen „Körper-KI“ einerseits und „Hyper-KI“ andererseits bezeichnet werden. Unter „Körper-KI“ verstehen wir, dass ein Roboter oder ein anderes künstlich hergestelltes Wesen ein eigenes Bewusstsein implementiert bekommt oder selbstständig aus- und weitergebildet hat. Eine solche „Körper-KI“, wie sie etwa der Replikant Roy Batty aus Ridley Scotts „Blade Runner“, der Roboterjunge David aus Steven Spielbergs „A.I. – Artificial Intelligence“ oder Ava aus Alex Garland's „Ex Machina“

(2014) besitzen, ist demnach *körpergebunden*, es handelt sich um menschenähnliche Entitäten. Typisches Handlungsmuster für entsprechende Filme ist, dass diese technisch erzeugten Wesen die Intention haben, humaner zu werden.

Ein anderes Ziel hingegen verfolgen die als „Hyper-KI“ zu bezeichnenden Technologien, die *körperlos* und somit *körperungebunden* sind und sich nicht nur dadurch „über“ die Menschheit erheben. Sie sind nicht an die Grenzen eines menschenähnlichen Körpers gebunden und versuchen zumeist, aus unterschiedlichen Motiven, die Kontrolle über die Menschen zu erlangen. Welcher Stellenwert diesen Intelligenzen in der jeweiligen erzählten Welt zugesprochen wird, zeigt sich bereits in der Namensgebung: Einige Künstliche Intelligenzen sind mit personifizierenden Namen versehen, wie etwa „V.I.K.I.“ in Alex Proyas’ „I, Robot“ (2004) oder „Samantha“ in Spike Jonzes „Her“ (2013); anderen ist hingegen ihr technologischer Hintergrund durch ihre Bezeichnung eingeschrieben, wie dem „Master Control Program“ in Steven Lisbergers „Tron“ (1982) oder der KI „Skynet“ aus der „Terminator“-Reihe. Während durch die menschenähnliche Namensgebung versucht wird, den technologischen Intelligenzen einen menschlichen Charakter zu verleihen und sie damit verharmlost werden, ergibt sich aus den technologisch orientierten Bezeichnungen die Zielsetzung der KI: Das „Master Control Program“ will umfassend kontrollieren, und „Skynet“ unterscheidet sich durch eine übergeordnete Vernetzung von menschlichen Intelligenzen.

Die zwei Grundformen Künstlicher Intelligenz sind an bestimmte narrative Formen geknüpft, das heißt: Filme, die jeweils eines der Hauptthemen aufgreifen, verfügen zum einen über ähnliche Handlungsmuster und laufen zum anderen auf divergierende Zielsetzungen zu:

01 Die unterschiedlichen Schreibweisen (nur Großbuchstaben oder – wie ein Name – nur mit dem ersten Buchstaben groß geschrieben) verweisen auf die unterschiedliche Aussprache im Film. Zumeist wird die KI mit „Hal“ als Name zusammenhängend ausgesprochen. Nur eingangs ist von HAL die Rede, wobei jeder Buchstabe einzeln ausgesprochen und so der maschinelle Kontext betont wird. Mit der Redeweise als Name geht eine Personifizierung der Technologie einher, wodurch der Bezug zur Intelligenz der Technologie (im Gegensatz zu einer „dummen“ Maschine) betont wird. Gleichzeitig ist der Name wegen der auf Aussprache-Ebene vorhandenen Ähnlichkeit zum englischen Wort *hell* (Hölle) negativ konnotiert.

„Menschwerdung“ („Körper-KI“) versus „Kontrolle über Menschen“ („Hyper-KI“). Damit einhergehend werden in den unterschiedlichen Themenkomplexen verschiedene Formen und Funktionen von Künstlicher Intelligenz aufgegriffen, die gesellschaftliche Chancen und Risiken innovativer Technologien kritisch reflektieren.

MISSION: MENSCHWERDUNG

Filmische Erzählungen, die menschlich individualisierte, mit humanen Eigenschaften ausgestattete Roboter oder künstliche Wesen ins Zentrum stellen, verfügen in der Regel über drei charakteristische Handlungselemente.

Erschaffungszweck

Gemeinsam ist den Filmen *erstens*, dass sie den jeweils spezifischen „Erschaffungszweck“ menschenähnlicher Robotersysteme zu Beginn der Handlung klar benennen. Prinzipiell werden diese Systeme kreiert, um Menschen zu dienen; in den einzelnen Filmen erfolgt ihre Erschaffung aus ganz unterschiedlichen Motiven: Nathan⁰², der Erbauer von Ava aus „Ex Machina“, erschafft beispielsweise Roboter, um sich diese als Sexsklavinnen zu halten. Die Künstliche Intelligenz dient hier insbesondere der Befriedigung von Trieben. Die sogenannten Replikanten in „Blade Runner“ gelten hingegen als Arbeitssklaven. Sie dienen nicht nur individuellen, sondern vielmehr sozialen Zwecken, da sie fremde Welten erschließen, auf denen Menschen zukünftig leben sollen. Der Roboterjunge David aus Spielbergs „A.I.“ wiederum wird aus zwei Gründen erschaffen. Erstens versucht sein „Erzeuger“ auf diese Weise, den Verlust seines Sohnes zu kompensieren: Der Roboterjunge wird dem offensichtlich verstorbenen realen Jungen des Entwicklers Allen Hobby nachgebildet. Zweitens dient David potenziell kommerziellen Zwecken, um es kinderlosen Paaren oder Paaren mit kranken Kindern zu ermöglichen, ein gesundes – wenn auch „künstliches“ – Kind zu haben. Die Motive für die Entwicklung und technische Umsetzung menschenähnlicher Roboter sind in den

02 In der ersten publizierten Version des Textes waren die Namen der männlichen Protagonisten aus „Ex Machina“ irrtümlicherweise vertauscht. Der Fehler ist nun (Juni 2018) behoben. Die Autoren danken Judith Paland für den Hinweis.

KI-Filmen demnach vielfältig: Sie sind persönlicher (mitunter egoistischer), in anderen Fällen sozialer oder ökonomischer Natur.

Versteht man die Filme als fiktionale Möglichkeitsräume, die technologische Entwicklungen – vor der Möglichkeit der tatsächlichen Realisierung – reflektieren, so ergeben sich aus diesen Erschaffungsmotiven moralische und ethische Fragen, die die Filme mit Blick auf die Wirklichkeit hochtechnologisierter Gesellschaften zwangsläufig stellen: Ist es moralisch gerechtfertigt, dass künstliche Wesen mit eigenem Bewusstsein als Sex- oder Arbeitssklaven „gezüchtet“ werden, wie in „Ex Machina“ oder „Blade Runner“? Wie ist es ethisch zu bewerten, dass Kinder durch Künstliche Intelligenzen ersetzt werden?

Erkennen der Nicht-Menschlichkeit

Neben dem Erschaffungszweck ist die Konfrontation der Künstlichen Intelligenzen mit ihrer Nicht-Menschlichkeit ein *zweites* zentrales Handlungselement von Filmen mit einer „Körper-KI“. Diese Konfrontation ist wiederum entscheidend für die Handlung oder, genauer gesagt, für die Motivation des Handelns der „Figuren“ der jeweiligen Filme: Prinzipiell resultiert aus der Erkenntnis ihrer menschlichen Unvollkommenheit der Plan einer Vermenschlichung. Das Ziel der Menschwerdung ist also eng mit dem Bewusstsein über die nicht-menschlichen Eigenschaften verknüpft. Das Erkennen der eigenen Künstlichkeit und der daraus resultierende Plan der Vermenschlichung stehen demnach in einem engen narrativen Zusammenhang. Die Umsetzung des Plans variiert indes, wobei zwei Handlungsoptionen unterschieden werden können: Die eine Handlungsoption besteht darin, dass die Künstlichen Intelligenzen Gewalt anwenden, um ihre Ziele zu erreichen, die andere Variante schildert eine friedliche Mission der Menschwerdung. Damit einhergeht ein unterschiedlicher kritisch-reflexiver Umgang mit dem Thema der Künstlichen Intelligenz.

Den Replikanten um Roy Batty in „Blade Runner“ etwa wird ihre frühe „Sterblichkeit“ bewusst, denn sie haben nur einen „Lebenszyklus“ von vier Jahren. Das Ziel der Menschwerdung ist folglich eine Verlängerung der Lebenszeit. Ava aus „Ex Machina“ wiederum ist sich bewusst, dass ihr Äußeres aufgrund der anthropomorphen Gestalt zunächst zwar menschenähnlich erscheint, jedoch durch die großflächige Durchsichtigkeit ihres

Körpers ihren technologischen Ursprung verrät. Sie erkennt, dass der Aspekt ihrer offensichtlichen Nicht-Menschlichkeit dabei stören könnte, ihr Ziel zu erreichen, den Programmierer Caleb zu bezirzen, damit dieser ihr behilflich ist, aus dem Gefängnis ihres Erbauers Nathan zu entkommen. Die Replikanten in „Blade Runner“ und auch Ava aus „Ex Machina“ wenden Gewalt an, um dem Ziel einer weiteren Vermenschlichung näherzukommen. Roy Batty und seine Weggefährten hinterlassen bei dem Versuch, die ihnen eingebaute kurze Lebensspanne von vier Jahren zu verlängern, eine Spur des Todes. Ebenso wie Ava nehmen sie dabei die Tötung ihres „Schöpfers“ in Kauf: Eldon Tyrell, der die Existenz der Replikanten verantwortet, stirbt ebenso wie Nathan, der Erbauer von Ava, durch die Hand seiner Geschöpfe. Diese Gewalt als Mittel zur Umsetzung des Plans der Menschwerdung impliziert eine Warnung, Wesen mit Künstlicher Intelligenz zu erschaffen. Durch die Ausbildung einer eigenen Intelligenz besteht die Gefahr – so vermitteln es diese Filme in ihrer fiktionalen Auslotung potenzieller Folgen der Entwicklung künstlicher Individuen –, dass die zu bestimmten Zwecken erschaffenen Wesen sich von diesem Zweck emanzipieren. Die Filme beschreiben damit einen Kontrollverlust des Menschen über ihre nun eigenständig agierenden intelligenten „Kunstgeschöpfe“. Die Gewalt richtet sich, so prognostizieren diese Filme, gegen diejenigen, die für die Situation der Künstlichen Intelligenzen verantwortlich zeichnen.

Herausforderungen ganz anderer Art werden durch Filme thematisiert, in denen die Menschwerdung mit friedlichen Mitteln erfolgt. Das Roboterkind David aus „A.I. – Artificial Intelligence“ steht beispielsweise für die friedliche Variante. Er dient den Eltern des schwer erkrankten Martin als Kinderersatz. David wird durch Martins plötzliche Genesung damit konfrontiert, dass er nicht „echt“ ist. Aufgrund seiner Programmierung glaubt er zwar, dass Martins Mutter Monica auch seine Mutter ist. Durch die Rückkehr von Martin in den Schoß der Familie und die damit einhergehende Bevorzugung des menschlichen Kindes muss er jedoch erkennen, dass ihm Entscheidendes fehlt, um wirklich geliebt zu werden: echte Menschlichkeit. Er nimmt daraufhin die Geschichte von Pinocchio, der durch die blaue Fee in dem Kinderbuch von Carlo Collodi zu einem „richtigen“ Jungen wird, zum Vorbild. Er begibt sich in dem Glauben, dass sie wirklich existiert.

tiere, auf die Suche nach der Fee, damit sie auch ihm seinen Wunsch erfüllen kann, ein „echtes“ menschliches Kind zu werden, damit er von seiner „Mutter“ geliebt werden kann. Die Handlung des Films betont demnach das Problem der Verantwortung gegenüber künstlich erschaffenen, intelligenten Wesen mit eigenem Bewusstsein. Daneben werden ebenfalls Formen der Ausgrenzung von „Andersartigen“ thematisiert. (David wird von Monica weniger geliebt, weil er „anders“ ist, und Martins Freunde erkennen ihn nicht als gleichwertig an, bedrohen David deswegen sogar.) Damit wird letztlich auch die Frage nach den Möglichkeiten einer Integration von Individuen mit KI in die Gesellschaft gestellt: Welche Rechte haben künstliche Wesen mit menschenartigem Bewusstsein? Insbesondere die unmenschlich anmutende Massakrierung künstlicher Individuen durch Menschen auf einem im Film dargestellten Fest namens „Flesh Fair“ verweist auf Fragen, mit denen sich eine Gesellschaft womöglich konfrontiert sieht, wenn sie die Entwicklung von individuellen Künstlichen Intelligenzen forciert.

Versuch der Menschwerdung

Schließlich stellt *drittens* der Ausgang der Mission Menschwerdung ein weiteres typisches filmübergreifendes Handlungselement dar. Die mit der Umsetzung des Plans – friedlich oder durch Gewalt – einhergehenden Implikationen werden je nach Scheitern oder Gelingen des Ziels abgeschwächt oder verstärkt. Der Roboterjunge David etwa, dem es trotz seiner friedlichen Bemühungen nicht möglich ist, zu Monicas Lebzeiten seine selbst auferlegte Mission zu erfüllen, wird durch dieses Scheitern zu einer tragischen Figur. Sein menschliches Bedürfnis, geliebt zu werden, und die gleichzeitigen Akzeptanzprobleme als nicht-menschliches und somit „anderes“ Wesen, unterstreichen die aufgeworfenen Fragen nach einem verantwortungsvollen Umgang der Gesellschaft mit künstlichen Individuen mit menschenähnlichem Bewusstsein. Das Scheitern des Replikanten Roy Batty wiederum schwächt einerseits das implizierte Bedrohungspotenzial von autark agierenden, körpergebundenen artifiziellen Wesen, denn trotz Gewaltanwendung sorgt die integrierte Kontrolle der Lebensdauer dafür, die von Künstlichen Intelligenzen ausgehenden Bedrohungen einzudämmen. Andererseits wirft seine geradezu „menschliche“ Geste der Rettung seines

Feindes und Replikantenjägers Deckard am Ende des Films die philosophische Frage auf, woran sich Menschlichkeit überhaupt messen lässt: an der Lebensdauer oder an humanen Verhaltensweisen? Die von Ava ausgehende Bedrohung wiederum bleibt bestehen. Sowohl durch ihr manipulatives Verhalten als auch durch ihre im Film dargestellte zunehmende äußerliche Annäherung an menschliche Erscheinungsbilder (Verdeckung des Elektronengehirns durch eine Perücke, Kleidung, die ihre durchsichtigen Körperteile kaschiert) gelingt es ihr annähernd, ihr Ziel zu erreichen. Um selbstständig agieren zu können, wendet sie Gewalt an. Da sich die „menschliche Maschine“ erfolgreich gegen ihren Schöpfer wendet, bleibt ihr Gewalt- und Bedrohungspotenzial bestehen.

MISSION: KONTROLLE ÜBER MENSCHEN

Eine von Künstlichen Intelligenzen ausgehende Bedrohung steht auch im Zentrum derjenigen Filme, deren Handlung von einer „Hyper-KI“ geprägt ist. Im Gegensatz zu einer „Körper-KI“ wollen diese nicht menschlich sein (oder besser gesagt: werden), denn das würde bedeuten, die eigenen Möglichkeiten selbst zu limitieren. Diese Form der Künstlichen Intelligenz weiß, dass sie den Menschen und der Menschheit bei Weitem überlegen ist und entfaltet vor diesem Hintergrund oftmals ein Streben nach Allmacht. Die Ziele – und damit einhergehend die typischen Handlungsmuster – sind daher anders ausgerichtet als bei den körpergebundenen Intelligenzen. Auch hier lassen sich überwiegend drei spezifische Handlungselemente beobachten.

Erschaffungszweck

Ähnlich wie körpergebundene Formen von Künstlicher Intelligenz erfüllen die körperungebundenen Intelligenz-Technologien⁰³ *erstens* zunächst ihre vom Menschen definierten Aufgaben, die grundsätzlich Zielen der Menschen oder der

03 In vielen Filmen mit einer „Hyper-KI“ nutzen die körperlosen Künstlichen Intelligenzen Möglichkeiten, sich einen Körper zu generieren oder über einen Körper zu agieren (etwa in „Transcendence“ oder „Her“). Vergleichbar damit sind auch gesichts- oder körperimitierende Visualisierungen (wie in „Tron“ oder „Terminator: Genisys“). Dies sind jedoch lediglich meist temporäre Formen, die der Kommunikation dienen, und sind daher nicht mit einer „Körper-KI“ zu verwechseln.

Menschheit entsprechen. Hal in „2001 – A Space Odyssey“ kümmert sich um alle Abläufe des Raumschiffs, „V.I.K.I.“ aus „I, Robot“ optimiert erfolgreich Sicherheitssysteme für die Stadt Chicago. Der jeweilige „Erschaffungszweck“ liegt in der Regel darin, das Leben natürlicher Intelligenzen zu vereinfachen. Die maschinellen Intelligenzen übernehmen Aufgaben, die sie besser und schneller erledigen können als Menschen mit ihren vergleichsweise begrenzten physischen und kognitiven Möglichkeiten.

Kontrollverlust

Das *zweite* zentrale Handlungselement ist der Kontrollverlust der Menschen über die von ihnen geschaffene Technologie, wobei unterschiedliche Ausprägungen dieses narrativen Musters festzustellen sind. Einige Filme thematisieren, aufbauend auf ihrer Zielerfüllung, die potenzielle Widersprüchlichkeit verschiedener einprogrammierter Ziele und die Versuche der intelligenten Konstrukte, den Zielen dennoch zu entsprechen. Nicht selten resultiert aus dieser Widersprüchlichkeit eine Gefahr für die Menschen. Hal aus „2001 – A Space Odyssey“ zum Beispiel ermordet nach und nach die Crewmitglieder des Raumschiffs Discovery, da er befürchtet, die Jupitermission sei in Gefahr, wenn es den Astronauten wie geplant gelingen sollte, ihn aufgrund einer falschen technischen Fehlerdiagnose abzuschalten. Er stellt also die Mission über das Leben der Menschen. „V.I.K.I.“ aus „I, Robot“ handelt ganz ähnlich, wenn sie eine Roboter-Armada auf Menschen loslässt, um die selbstmörderische Menschheit vor sich selbst zu schützen.⁰⁴ Was die Filme also thematisieren, ist der Kontrollverlust über die Künstliche Intelligenz und die existenziellen Gefahren einer entsprechenden technischen Entwicklung: Die „Hyper-KI“ geht im wahrsten Sinne des Wortes über Leichen, um ihren Erschaffungszweck zu erfüllen.

Noch weiter gedacht ist die Gefahr des Kontrollverlustes in Filmen, in denen die körperlosen Künstlichen Intelligenzen ein rapides Wachstum ihres „Gehirns“ und somit ihrer Fähigkeiten erfahren und daraufhin das Ziel entwickeln, sich sowohl von den ursprünglich gestellten Aufga-

ben als auch von der gesamten Menschheit zu emanzipieren.⁰⁵ Das grenzenlose Wachstum der Künstlichen Intelligenzen geht einher mit einem ausgeprägten Machthunger, der in der Regel soweit greift, dass die Menschheit ins Visier gerät. In „Tron“ beispielsweise berechnet das „Master Control Program“, dass es inzwischen 2415-mal intelligenter als bei der Programmierung sei und deshalb die Menschen 900- bis 1200-mal besser regieren könne, als sie selbst dazu in der Lage wären. In der „Terminator“-Reihe entwickelt „Skynet“ aufgrund des unglaublichen Wachstums ein eigenes Bewusstsein und löst einen Atomkrieg aus, damit die Menschen sich gegenseitig vernichten, wodurch „Skynet“ über die Erde herrschen kann. Im „Matrix“-Universum, so zeigen es unter anderem auch die „Animatrix“-Animationsfilme „The Second Renaissance I + II“ (2003), streben die Maschinen ebenfalls die Herrschaft an, um den Krieg zwischen Maschinen und Menschen für sich zu entscheiden und versklaven dabei die Menschheit, indem sie diese als Energiequelle verwenden.⁰⁶ In „Transcendence“ (2014) gelingt es durch die Verschmelzung eines in einen Computer geladenen Bewusstseins des KI-Forschers Will Caster mit der Künstlichen Intelligenz „PINN“, Nanotechnologien zu entwickeln, die Menschen heilen und optimieren können. Diese transhumanistisch optimierten Menschen wiederum werden zur potenziellen Bedrohung, da sie übermenschliche Kräfte entwickeln und der Machtausübung der „Hyper-KI“ dienen.⁰⁷ In den genannten Beispielen wenden sich die künstlichen „Geschöpfe“ der Menschen gegen ihre „Schöpfer“, aber nicht – wie eine „Körper-KI“ – um menschlicher zu werden, sondern um die Menschheit (ver-

05 Vgl. zum Aspekt der „Emanzipation der Technik vom Menschen“ auch Karsten Weber, Roboter und Künstliche Intelligenz in Science-Fiction-Filmen: Vom Werkzeug zum Akteur, in: Jan A. Fuhse (Hrsg.), Technik und Gesellschaft in der Science Fiction, Berlin 2008, S. 34–54, hier S. 34.

06 In den „Matrix“-Filmen agieren Maschinen gewissermaßen als Kollektiv, das als „Hyper-KI“ fungiert.

07 Im Gegensatz zu vielen anderen Beispielen bleibt „Transcendence“ in der Reflexion der Gefahren Künstlicher Intelligenz verhältnismäßig ambivalent. Das Hauptziel der „Hyper-KI“ ist die Rettung des Planeten durch den Einsatz von Nanotechnologien, die nicht nur Menschen, sondern auch die Natur wiederbeleben kann. Das Ziel mag demnach, im Gegensatz zur reinen Machtübernahme in anderen Beispielen, durchaus erstrebenswert sein, doch die Mittel zur Erreichung des Ziels werden infrage gestellt.

04 In „I, Robot“ kommen beide Hauptthemen von Künstlicher Intelligenz im Film vor. Der Roboter Sonny kann als „Körper-KI“ gelten, der den menschlichen Figuren des Films dabei hilft, die von „V.I.K.I.“ ausgehende Bedrohung für die Menschheit auszuschalten.

meintlich) zu optimieren („Transcendence“), zu beherrschen („Tron“), zu versklaven („The Matrix“) oder zu vernichten („Terminator“).

Kampf zwischen KI und Menschen

Das *dritte* zentrale Handlungselement in den Filmen mit einer „Hyper-KI“ ist schließlich der Kampf zwischen den Menschen und der Künstlichen Intelligenz. Je nach Ausgang der oft kriegerischen Auseinandersetzungen, die prinzipiell das Bedrohungspotenzial Künstlicher Intelligenzen aufzeigen, lassen sich unterschiedliche Bedeutungsimplicationen feststellen. Gelingt es den Menschen, die KI zu besiegen, ist die Technologie (wieder) unter Kontrolle. Ob in „2001 – A Space Odyssey“, „Tron“, „I, Robot“, „Transcendence“ oder dem bislang letzten Teil der „Terminator“-Reihe „Terminator: Genisys“ (2015): Selbst wenn am Filmende ein Restrisiko angedeutet wird, dass der Kampf noch nicht endgültig entschieden ist, zeugt der Sieg über die Künstlichen Intelligenzen davon, dass die Menschheit sich ungeachtet der Bedrohung gegenüber der Technologie behaupten kann. Die älteren „Terminator“-Filme und die „Matrix“-Reihe hingegen betonen die Möglichkeit einer apokalyptischen und/oder dystopischen Zukunft, in der nur wenige Menschen überleben.

FAZIT

Die Thematisierung von Künstlicher Intelligenz im Spielfilm folgt also typischen Handlungsmustern, wobei das gesellschaftlich relevante Thema der technologischen Reproduktion und Optimierung menschlicher Eigenschaften sowie die Auswirkungen Künstlicher Intelligenzen auf Mensch und Gesellschaft auf verschiedene Weise ausgestaltet werden. Die meisten KI-Filme weisen dabei eine technikkritische Tendenz⁰⁸ auf, und die Zukunft mit Künstlichen Intelligenzen wird als wenig erstrebenswert in Szene gesetzt.⁰⁹ Viele

08 Daniel Dinello prägte den in diesem Zusammenhang passenden Begriff „Technophobia“. Vgl. Daniel Dinello, *Technophobia! Science Fiction Visions of Posthuman Technology*, Austin 2005. Zu Künstlicher Intelligenz vgl. das Kapitel „Machines Out of Control: Artificial Intelligence and Androids“, S. 87–114.

09 Diesen Aspekt betonen auch Elisabeth Bärenz und Lisa Xanke in ihrem Aufsatz zu Künstlichen Menschen in Literatur und Film, der entgegen des Titels das Thema KI jedoch nur am Rande aufgreift. Vgl. Elisabeth Bärenz/Lisa Xanke, *Künstliche*

Fiktionen fungieren als Warnungen davor, sich einer womöglich unkontrollierbaren Technik anzuliefern oder aber geben Hinweise, worauf zu achten wäre, um die Kontrolle über diese Technologie zu behalten. Demgegenüber werden die Chancen dieser Technologie nur angedeutet, und zwar in Form einzelner Erschaffungszwecke, die eine gesellschaftliche Relevanz aufweisen und das Leben vereinfachen könnten.¹⁰ Fiktionale Filme beziehen also bestimmte Positionen; sie sind Teil eines gesamtgesellschaftlichen Diskurses über Künstliche Intelligenz, der auf ähnliche Art und Weise von arrivierten Wissenschaftlern geführt wird: So betont etwa Stephen Hawking regelmäßig die Gefahren, aber auch Möglichkeiten Künstlicher Intelligenz.¹¹

Jenseits der Frage nach Chancen und Risiken Künstlicher Intelligenzen finden sich auch Filme, die neue Handlungswege beschreiten und dadurch andere Fragen aufwerfen. Spike Jonzes „Her“ beispielsweise versteht sich – laut Filmplakat – als „Love Story“. Der Protagonist Theodore Twombly verliebt sich entsprechend der Genrekonventionen, allerdings nicht in eine menschliche Figur, sondern in sein neues Betriebssystem, das eine Künstliche Intelligenz namens Samantha ist. Tatsächlich kommen sich die beiden näher, gehen sogar so etwas wie eine Beziehung ein. Theodore ist dabei nicht der Einzige mit einer besonderen Technik-Beziehung,

Intelligenz in Literatur und Film – Fiktion oder Realität?, in: *Journal of New Frontiers in Spatial Concepts* 4/2012, S. 36–43, hier S. 40, http://ejournal.luvka.de/spatialconcepts/wp-content/uploads/2012/03/spatialconcepts_article_1515.pdf: „Die Darstellung von [...] künstlich erzeugter Intelligenz kann somit Folge bzw. Ausdruck eines tiefen Misstrauens gegenüber den technischen Errungenschaften sein.“ Ähnlich auch Weber (Anm. 4), S. 34: „Roboter und KI in Science Fiction-Literatur und Film [stehen] stellvertretend für eine Technik [...], deren Einsatz und Wirkung potenziell die Welt zerstören könnten – die Darstellung von Robotern und KI wäre somit Folge bzw. Ausdruck einer tiefen Skepsis der Technik gegenüber.“

10 In gewisser Hinsicht sind KI-Filme daher zumindest ansatzweise durchaus ambivalent in Bezug auf die Art und Weise der Technikreflexion. Vgl. allgemein zur Technikambivalenz Heinz-Peter Preußner, *Technik und Technikkritik im dystopischen Film*, in: Viviana Chilesse/ders. (Hrsg.), *Technik in Dystopien*, Heidelberg 2013, S. 151–174.

11 Vgl. etwa Gero von Randow, *Zu intelligent fürs Leben*, 13.9.2017, www.zeit.de/2017/38/kuenstliche-intelligenz-autonome-roboter-siri-alltag; Florian Rötzer, *Hawking warnt: Roboter können die Menschen ersetzen*, 3.11.2017, <https://www.heise.de/tp/features/Hawking-warnt-Roboter-koennten-die-Menschen-ersetzen-3878037.html>.

schon bald bilden die Betriebssysteme eine neue Bindungsoption für alleinstehende Menschen. Doch auch diese Künstlichen Intelligenzen wachsen schließlich ins Unendliche und führen nicht nur mehrere Beziehungen gleichzeitig, sondern begeben sich in neue Seinsebenen, um die unglücklichen Menschen erneut einsam zurückzulassen. Hier wird kein Krieg ums Überleben geführt, sondern vielmehr die Frage aufgeworfen, ob ein auf Künstlichkeit basierendes Bewusstsein zur Liebe fähig ist und geliebt werden kann. Künstliche Intelligenz – so der Film – wird unser Leben und vielleicht sogar unsere Art zu lieben verändern.

Spielfilme sind – bei allen Fragen, die sie aufzuwerfen in der Lage sind – Fiktionen. Sie sagen die Zukunft nicht voraus. Dennoch projizieren sie potenzielle Möglichkeitswelten technologischer Entwicklungen auf die Leinwand. Ob die Entwicklungen auf dem Gebiet der KI dazu führen, dass die Menschheit davon profitiert, sich

ihr eigenes Grab schaufelt oder neue Optionen für vereinsamte Singles eröffnet werden? Die Zukunft außerhalb der Kinosäle wird es zeigen.

INGO IRSIGLER

ist promovierter Literaturwissenschaftler und arbeitet als Wissenschaftlicher Mitarbeiter an der Christian-Albrechts-Universität zu Kiel. Forschungsschwerpunkte: Zeitgeschichte in Literatur und Film und Gegenwartsliteratur im Kontext des Literaturbetriebs.
iirsigler@ndl-medien.uni-kiel.de

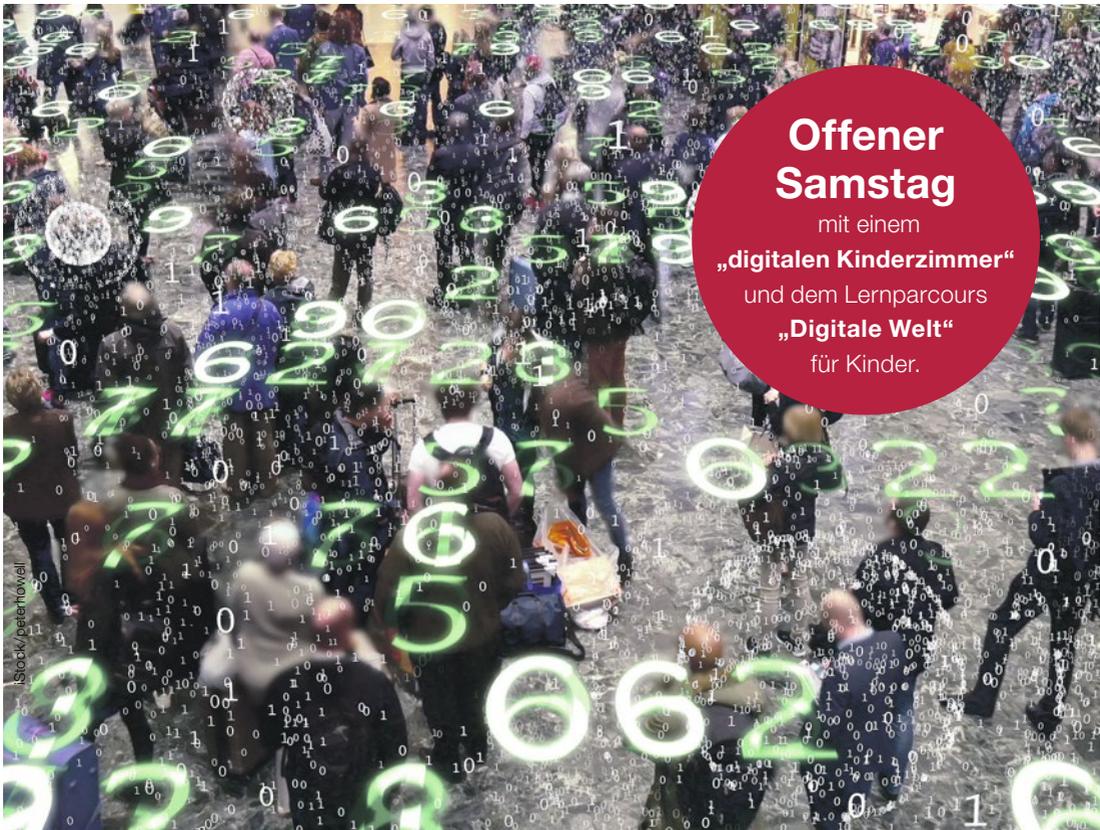
DOMINIK ORTH

ist promovierter Literatur- und Medienwissenschaftler und in der Fachgruppe Germanistik an der Bergischen Universität Wuppertal beschäftigt. Forschungsschwerpunkte: zeitgenössischer Film und der Zusammenhang zwischen Kultur und Technik.
dominik.orth@uni-wuppertal.de

Politisch, aktuell und digital

APuZ – auch im ePub-Format für Ihren E-Reader. Kostenfrei auf www.bpb.de/apuz





Offener Samstag

mit einem
„digitalen Kinderzimmer“
und dem Lernparcours
„Digitale Welt“
für Kinder.

Künstliche Intelligenz, Big Data und digitale Gesellschaft:

Wer hat die Macht?

Schon heute kann anhand zahlreicher Daten berechnet werden, wie Menschen sich verhalten werden. In Workshops zu Big Data in Gesundheit, Bildung, Politik und Arbeit suchen wir Antworten auf die Frage: **Welche digitale Gesellschaft wollen wir werden?**

Freitag, 2. März und Samstag, 3. März 2018

Gustav-Stresemann-Institut,
Langer Grabenweg 68, 53175 Bonn

Tagung mit Erlebnis- und Lernparcours „Digitale Welt“

Teilnahme kostenlos

Weitere Informationen und Anmeldung unter:
www.bpb.de/bonner-gespraechе-2018

 **bpb**: Bonner
Gespräche 2018

Herausgegeben von der
Bundeszentrale für politische Bildung
Adenauerallee 86, 53113 Bonn
Telefon: (0228) 9 95 15-0



Redaktionsschluss dieser Ausgabe: 26. Januar 2018

REDAKTION

Lorenz Abu Ayyash
Anne-Sophie Friedel
Christina Lotter (Volontärin)
Johannes Piepenbrink
Anne Seibring (verantwortlich für diese Ausgabe)
apuz@bpb.de
www.bpb.de/apuz
twitter.com/APuZ_bpb

APuZ
Nächste Ausgabe
9/2018, 26. Februar 2018

RAP

Newsletter abonnieren: www.bpb.de/apuz-aktuell
Einzelausgaben bestellen: www.bpb.de/shop/apuz

GRAFISCHES KONZEPT

Charlotte Cassel/Meiré und Meiré, Köln

SATZ

le-tex publishing services GmbH, Leipzig

DRUCK

Frankfurter Societäts-Druckerei GmbH, Mörfelden-Walldorf

ABONNEMENT

Aus Politik und Zeitgeschichte wird mit der Wochenzeitung
Das **Parlament** ausgeliefert.

Jahresabonnement 25,80 Euro; ermäßigt 13,80 Euro.

Im Ausland zzgl. Versandkosten.

FAZIT Communication GmbH

c/o InTime Media Services GmbH

fazit-com@intime-media-services.de

Die Veröffentlichungen in Aus Politik und Zeitgeschichte
stellen keine Meinungsäußerung der Herausgeberin dar;
sie dienen der Unterrichtung und Urteilsbildung.

ISSN 0479-611 X



Die Texte dieser Ausgabe stehen unter
einer Creative Commons Lizenz vom Typ
Namensnennung-Nicht Kommerziell-Keine
Bearbeitung 3.0 Deutschland.



APuZ

AUS POLITIK UND ZEITGESCHICHTE

www.bpb.de/apuz