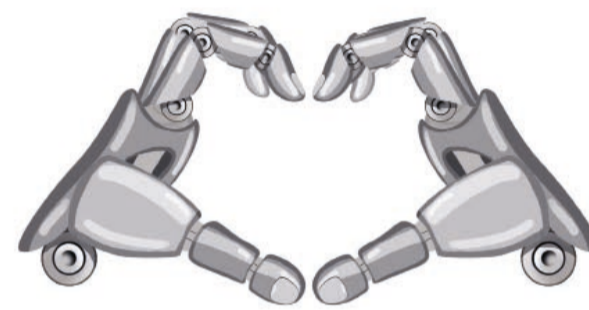


Wenn die Maschinen
immer klüger werden:

Was macht uns künftig noch einzigartig?



Künstliche Intelligenz überflügelt die Menschen angeblich bald in jeder Hinsicht.

Wir haben sieben Experten gefragt, welche Rolle dann für uns bleibt **VON ULRICH SCHNABEL**

Es war ein ungewöhnlicher Moment der Wahrheit: Bei einem Treffen in Long Beach bei Los Angeles diskutierten kürzlich Fachleute unter Ausschluss der Öffentlichkeit über die Zukunft der künstlichen Intelligenz (KI). Am Ende kam die Frage auf, auf welche minimale Forderung im Umgang mit KI sich die Anwesenden einigen könnten. Dabei schälte sich ein zentrales Anliegen heraus, das alle gleichermaßen umtrieb: Es soll in 50 Jahren noch Menschen geben.

So weit ist es also? Steht tatsächlich das Überleben der Menschheit auf dem Spiel, wenn die Maschinen immer intelligenter werden und vielleicht irgendwann auf diesem Planeten die Macht übernehmen? Oder ist das nur der Horror verwirrter Spinner, die zu viel Science-Fiction gelesen haben und die Möglichkeiten der Technik – wie es schon so oft geschah – grandios überschätzen?

Zwei Dinge sind klar: Erstens, die künstliche Intelligenz entwickelt sich derzeit stürmisch und feiert einen Erfolg nach dem anderen. Zweitens, Digitalkonzerne wie Amazon, Apple, Microsoft, Facebook und Google (»die fürchterlichen Fünf«) investieren Milliarden in die Forschung – fast so viel wie der amerikanische Staat insgesamt pro Jahr für die zivile Forschung. Es ist also abzusehen, dass künstliche Systeme immer mehr lernen werden. Das führt unausweichlich zu drei Fragen: Erstens, was werden sie können? Zweitens, was werden sie wollen? Und drittens, was wird aus dem Menschen?

Homo sapiens hat sich von alters her über seine kognitiven Fähigkeiten definiert, über seine angeblich allen anderen Wesen überlegene Intelligenz. Doch jeder Fortschritt der KI nagt an dieser Gewissheit. Dass uns die Maschinen im Schachspiel schlagen, nun gut, daran haben wir uns gewöhnt. Dass sie uns auch im eher intuitiven Go-Spiel schlagen und sich dieses mittlerweile gar selbst beibringen können, lässt sich gerade noch verschmerzen (wer spielt hierzulande schon Go?). Dass sie aber seit vergangenem Jahr dem Menschen selbst im Pokerspiel überlegen sind, irritiert allmählich. Gehört da nicht zum Erfolg der Bluff, ein psychologisches Geschick und eine Menschenkenntnis, die wir den Maschinen nie zuge- traut hätten? Was folgt als Nächstes – und wo hört das auf?

Formulieren wir es einmal radikal: Gibt es so etwas wie ein Alleinstellungsmerkmal des Menschen gegenüber den Maschinen und falls ja, worin besteht es? Dabei geht es nicht nur um die Frage, welche Berufe künftig automatisierbar sind, ob die Post noch Briefträger und die Zeitung noch Sportreporter braucht. Vielmehr geht es um die Definition dessen, was Menschsein im digitalen Zeitalter bedeutet. Was ist unser *unique selling point*, den auch klügste Technik uns nicht streitig machen kann? Gibt es überhaupt einen? Oder werden wir erleben, wie eine humane Bastion nach der anderen geschleift wird?

Dass die Antwort darauf nicht feststeht, macht die Frage besonders reizvoll. Schließlich lehrt sie uns nicht nur etwas über KI, sondern auch über

uns Menschen. Deshalb hat die ZEIT Experten unterschiedlicher Disziplinen zu ihren Vorstellungen vom künftigen Verhältnis von Mensch und Maschine befragt. Üblicherweise wird bei diesem Thema auf bekannte Warner wie den Tesla-Gründer Elon Musk oder den jüngst verstorbenen Starphysiker Stephen Hawking verwiesen. Der Haken: Die meisten dieser Kritiker arbeiten gar nicht an künstlicher Intelligenz und werden deshalb von den vielen KI-Forschern nicht ernst genommen. Wir haben daher Experten befragt, die sich tatsächlich täglich mit den Problemen künstlicher (und menschlicher) Intelligenz herumschlagen. Ihre Antworten fielen mal zukunfts euphorisch, mal zutiefst pessimistisch aus und zeigen, wie groß die Bandbreite möglicher Zukünfte ist. Es hilft, sich mit all diesen Eventualitäten zu befassen, um die Frage nach dem Platz des Menschen beantworten zu können. Also auf in die Erkundung unserer eigenen Zukunft!

Als Erstes ein Anruf bei Jürgen Schmidhuber, Direktor des schweizerischen Forschungsinstituts für Künstliche Intelligenz. Schmidhuber gilt als das sonnige Gemüt der KI. Wenn er von seinen »klugen Maschinen« spricht, tut er das mit geradezu jugenhafter Begeisterung, die nicht zuletzt von seinem Erfolg herrührt. Denn Schmidhuber ist einer der Väter des sogenannten Deep Learning, das der KI-Forschung einen immensen Schub gegeben hat. Dabei werden Informationen in künstlichen neuronalen Netzen verarbeitet, die

– ähnlich wie im Gehirn – in verschiedenen Hierarchiestufen angeordnet sind und so ein »tief gehendes Lernen« ermöglichen. Der Witz dabei: Der Programmierer muss nicht alles vorgeben, die Maschine erstellt aus riesigen Datensätzen selbstständig Konzepte und Modelle über die Welt – ähnlich wie es kleine Kinder tun.

Schmidhuber verweist stolz darauf, dass alle großen Digitalfirmen von Apple bis Facebook heute die Long Short-Term Memory Networks (LSTM) und andere »tiefe Netze« nutzen, die in Schmidhubers Labor seit den 1990ern entwickelt wurden. In drei Milliarden Smartphones und anderen Rechnern auf der Welt sind die LSTM-Netze die treibende Kraft hinter jenen magischen Apps, die das Schreiben vereinfachen, Videos erkennen oder Sprachen übersetzen.

Doch für Schmidhuber ist das nur der Anfang. »Heute haben die größten LSTM-Netze etwa eine Milliarde Verbindungen. Extrapolieren wir den gegenwärtigen Trend, sollten wir in etwa 25 Jahren welche haben, die so viele Verbindungen wie das menschliche Gehirn haben, aber viel schneller sind.« Für den deutschen Informatiker ist es daher nicht abwegig zu prognostizieren, »dass meine Forschungsgruppe nicht mehr so viele Jahre brauchen wird, um ein System mit den geistigen Fähigkeiten eines Kapuzineräffchens zu erschaffen«.

Sicher, derzeit sei die künstliche Intelligenz »noch nicht einmal auf Kleinkindniveau«, gibt Schmidhuber zu. Doch das werde sich rasant ändern. Die Systeme würden sich unaufhaltsam weiterentwickeln – »schon deshalb, weil so viele

Bis 2049 an
die Weltspitze

Was dahintersteckt, wenn China
die Forschungsförderung strafft

Von China sind wir ja Dimensionen gewohnt, die das hiesige Vorstellungsvermögen auf die Probe stellen. Aber versuchen wir's: Stellen wir uns also vor, hierzulande würden der Deutsche Akademische Austauschdienst (DAAD) und die Deutsche Forschungsgemeinschaft (DFG) verschmolzen mit dem Bundesministerium für Bildung und Forschung. Was für ein Apparat! – Und das bleibt eine vergleichsweise zwergische Analogie, wenn es um China geht. Dort werden nämlich die National Natural Science Foundation of China (NSFC) und die State Administration of Foreign Experts Affairs ans Forschungsministerium angedockt. Die Regierung strafft ihre Kontrolle über Kooperationen mit dem Ausland und die Wissenschaftsförderung im Inland.

So hat es im März der Nationale Volkskongress beschlossen. Weil dort Chinas Machtanspruch für das 21. Jahrhunderts die Rhetorik bestimmte und weil dem Staatschef Xi Jinping der Weg in die lebenslange Amtszeit geebnet wurde, ging diese Konzentration der Forschungspolitik zwischen den Schlagzeilen aus Peking unter. Weniger wichtig ist sie indes nicht.

Denn es geht um das Werden der nächsten Forschungs-Supermacht. In China studieren an mehr als 2500 Hochschulen mehr Menschen als in der gesamten europäischen Union zusammen. Chinesen melden 17-mal so viele Patente an wie Deutsche. Und die Forschungsförderung weist von einem Jahr aufs nächste zweistellige Wachstumsraten auf.

Zwar hat die chinesische Spitzenforschung noch nicht amerikanisches oder europäisches Niveau erreicht. Aber sie voranzutreiben folgt einem Masterplan. Wenn die Volksrepublik 2049 ihr 100-jähriges Bestehen feiert, will sie technologisch und wissenschaftlich führend sein. »Chinas Ambitionen sind groß genug, dass man sie vom Weltall aus erkennen kann«, konstatierte das Wissenschaftsmagazin *Nature* zu Jahresbeginn. Das spielt auf das Klischee von der Chinesischen Mauer an, die man als einziges Bauwerk aus dem Erdboden erkennen könne (tatsächlich ist es ihr Schatten, den die Astronauten erblicken). Und natürlich ist das Motiv der ambitionierten Aufholjagd selbst ein Klischee.

Aber diese Ambitionen nehmen imponierende Gestalt an: der Teilchenbeschleuniger von Shanghai (2009; siehe Foto), der Tiefenrekord des Forschungs-Tauchschiffs *Jialong* (2012), die Premiere der Mondlandesonde Chang'e-3 (2013), das Hochrisikolabor des Virologie-Instituts von Wuhan (2015) und allein im vergangenen Jahr Chinas erster Röntgensatellit, das weltgrößte Radioteleskop und der weltweit erste Satellit für Quanten-Kommunikation.

Aufholen und Überholen – die Forschungsstrategie zeigt, wie die (noch) zweitgrößte Volkswirtschaft der Welt in den Schlüsselkompetenzen der Zukunft souverän werden will. Auf Basis breiter Erkenntnisse, ja. Aber nicht so zufällig wie einst der Westen, sondern ganz nach Plan.



Bislang teuerste
Forschungsanlage
Chinas: Synchrotron

STEFAN SCHMITT

HALBWISEN

Mut zur Ignoranz

Gestatten Sie doch bitte, dass wir hier den gängigen Überprüfungsprozess der etablierten Rächtschreibung, Rechtsschrei..., na, Sie wissen schon, ignorieren! Das ist nämlich mutig. Und nur so kann Neues entstehen. Dieses Argument, auch das sei mit demonstrativem Mut eingestanden, ist geklaut. Es stammt von einem Kollegen. Der hat es in der Verlagswerbung für ein Sachbuch gelesen. Dessen Titel ist hier nicht so wichtig wie die Behauptung, es handele sich um »ein mutiges Buch, da es sich nicht scheut, den gängigen Peer-Review-Prozess der etablierten Wissenschaften zu ignorieren«. Klingt so, als wäre die Begutachtung durch Fachkollegen ein Bollwerk gegen neue Ideen. Und als sei ein Aufklärer, wer dieses Bollwerk kühn links liegen lässt! Was für ein Held – zumindest in einer Welt, in der nicht jedermann nach Herzenslust Bücher schreiben darf ...

STX

Diese Fachleute hat die ZEIT um ihre Einschätzung gebeten



Illustration: Matthias Seifarth für DIE ZEIT; Fotos: Kasnic/laif, Taverna/cccb, IBM, PFH/Ullstein, privat, Pilston/Getty Images, Wegner/laif u., v. l.; Shanghai Daily/Imaginechina/laif (r.)

Jürgen Schmidhuber
Informatiker und wissenschaftlicher Direktor des schweizerischen Forschungsinstituts für Künstliche Intelligenz IDSIA bei Lugano

Luc Steels
Linguist und Direktor des Artificial Intelligence Lab der Universität Brüssel, Professor am Catalan Institute for Advanced Studies in Barcelona

Wolfgang Hildesheim
Physiker und Betriebswirt, Leiter Watson und KI-Innovation bei IBM Deutschland in Hamburg

John-Dylan Haynes
Hirnforscher und Professor für die Theorie und Analyse weiträumiger Hirnbereiche am Bernstein Center for Computational Neuroscience in Berlin

Christoph v. d. Malsburg
Neurobiologe und Physiker, Senior Fellow am Frankfurt Institute for Advanced Studies

Nick Bostrom
In Hirnforschung und Physik ausgebildeter Philosoph, Gründer und Direktor des Future of Humanity Institute an der Universität Oxford

Thomas Metzinger
Philosoph mit den Forschungsschwerpunkten Kognition und Bewusstsein, Universität Mainz

Fortsetzung auf S. 38

Die Erde wird künstlichen Intelligenzen eng erscheinen, und sie werden ins All streben, wo unendlich viel Energie lockt. Wir Menschen sind für die KI so interessant wie Ochsenfrösche für uns

Jürgen Schmidhuber

Was macht uns ... Fortsetzung von S. 37

Menschen von ihnen fasziniert sind – und irgendwann in den kommenden Jahrzehnten werde man das Wesen der Intelligenz vollständig verstanden haben, das im Rückblick dann recht simpel erscheine. Damit werde man Maschinen bauen können, die viel klüger als der Mensch sind – und damit eigentlich gar keine Maschinen mehr.

Was für Warner ein Horrorszenario ist, gibt Schmidhuber Anlass zur Euphorie. Was man derzeit erlebe, sei »viel mehr als nur eine industrielle Revolution«, schwärmt er. »Es ist etwas, was nicht nur die Menschen transzendiert, sondern das Leben selbst.« Ein neues, künstliches Leben werde entstehen: »Leben, das robuster ist und sich von unserer Biosphäre ausbreiten und die Milchstraße und schließlich das ganze Universum kolonisieren und intelligent machen wird.«

Verrückt? Für Schmidhuber eher logisch. Der Mensch sei schließlich nicht für die Eroberung des Alls geschaffen, da viel zu anfällig. Doch durch seinen Anstoß werde ein sich rasch ausbreitendes Ökosystem unterschiedlichster künstlicher Intelligenzen entstehen, »darunter extrem kluge, viele davon mit nach Bedarf wechselnden oder rapide sich ändernden Körpern, wie man sich das heute kaum vorstellen kann, in ständiger Anpassung an rasch evolvierte ökologische Nischen, noch viel reicher und komplexer als die der Biosphäre«. Und die KI würde sich zwangsläufig im All ausbreiten, die dort überreich vorhandene Sternenergie anzupfen und sich »diesen unglaublichen, neuen Lebensraum da draußen« erschließen. Das liege in der Logik der Evolution. »Das Universum strebt nach immer höherer Komplexität.« Und diesen Schritt würde künftig eben die KI gehen. Das Universum sei noch jung, meint Schmidhuber, und wenn es einst tausendmal älter sei als heute, werde man im Rückblick sagen: Kurz nach dem Urknall (kaum 14 Milliarden Jahre später) fing es an, intelligent zu werden.

Was aber wird aus dem Menschen? Nun ja, sagt Schmidhuber, »es scheint klar, dass der Mensch langfristig keine große Rolle spielen wird bei dieser Entwicklung. Menschen werden irgendwann nicht mehr die wichtigsten Entscheidungsträger sein.« Eher müssten wir uns als Steigbügelhalter der Evolution begreifen, als »Zwischenschritt«. So verlaufe nun einmal Geschichte. »Auch für uns gab es in der Evolution Vorstufen, wir sind zum Beispiel viel klüger als Ochsenfrösche.« Zu Konflikten müsse das nicht zwangsläufig führen. »Wir versuchen ja auch nicht, die Ochsenfrösche systematisch auszurotten. Aber die Ochsenfrösche sind keine wichtigen Entscheidungsträger mehr.«

An den viel beschworenen Kampf zwischen Mensch und Maschine glaubt Schmidhuber nicht. »Warum sollte die KI denn ein Interesse daran haben, uns zu vernichten? Alle Wesen interessieren sich vor allem für die, mit denen sie zusammenarbeiten oder in Wettbewerb treten können. Politiker interessieren sich für andere Politiker, Kinder für andere Kinder, Ochsenfrösche für andere Ochsenfrösche.« Deshalb würden sich auch »superkluge Maschinchen« vor allem für andere superkluge Maschinchen interessieren und nicht so sehr für Menschen und Ochsenfrösche. Außerdem seien sie langfristig vor allem an den »unglaublichen Ressourcen da draußen im Weltall« interessiert statt an unserem begrenzten Planeten. Kein Grund, sich allzu große Sorgen zu machen.

Aber könnte die Menschheit nicht von ihrer Bedeutungslosigkeit gekränkt sein? Nicht einmal das glaubt Schmidhuber. »Heute ist doch auch niemand gekränkt, bloß weil Einstein klüger war. Die Menschen sind längst gewohnt, dass andere in ihrem Umfeld genialer sind als sie selbst, und das stört kaum einen.« Es gebe allerdings eine lustige geschlechtsspezifische Asymmetrie, räumt der KI-Forscher ein. Denn wer von den Maschinen als Erstes überflügelt würde, das seien traditionell vor allem die Männer.

Warum? »Weil Männer sich durch Tunnelbegabungen auszeichnen, die für die von Männern geschriebenen Geschichtsbücher relevant sind. Der schnellste Mensch ist ein Mann, auch der beste Schachspieler war ein Mann. Aber längst spielen Maschinen besser Schach, und ein Motorrad ist viel schneller als jeder Läufer.« Soll heißen: Gerade in Tunnelbegabungen sind Maschinen besser. »Deshalb werden sie erst einmal Männerberufe übernehmen, weil Männer leichter zu ersetzen sind«, prophezeit Schmidhuber. Anders sieht es aus mit Tätigkeiten, welche die Fähigkeit zur »universellen Problemlösung« erfordern wie zum Beispiel die Kindererziehung – und dafür seien eben häufig immer noch die Frauen zuständig, meint Schmidhuber.

Ist das jetzt ein Trost? Oder muss uns Schmidhubers Analyse – vor allem mit Blick auf die Männer – erst recht beunruhigen? Dass er zumindest mit seiner Geschlechterprognose nicht ganz falsch liegt, zeigt eine Studie des MIT-Ökonomen Daron Acemoglu über »Roboter und Jobs« in den USA.



Eines der Ergebnisse: Durch zunehmende Automatisierung verloren Männer doppelt so häufig ihre Arbeit wie Frauen.

Das ist nur ein Beispiel für die gesellschaftlichen Spannungen, die uns in einer Welt voller künstlicher Intelligenzen erwarten könnte. Aber wie zwangsläufig ist Schmidhubers Vision vom unaufhaltsamen Erstarken der Maschinen? Sehen das alle KI-Forscher so? Mitnichten.

»Wir sollten den gegenwärtigen Stand der künstlichen Intelligenz nicht überschätzen und nicht zu enthusiastisch von der Gegenwart in die Zukunft extrapolieren«, sagt der Roboterforscher Luc Steels. Er ist, wie Schmidhuber, einer der Pioniere der KI-Forschung. Schon in den neunziger Jahren baute Steels Roboter, die anfangen, ihre eigene Sprache zu entwickeln – die sogenannten Talking Heads. Nachdem er lange das berühmte Sony Lab in Paris geleitet hat, denkt Steels heute am Catalan Institute for Advanced Studies in Barcelona über theoretische Grundsatzfragen nach. Dabei kommt er zu einem deutlich nüchterneren Urteil als Schmidhuber.

»Sicher gibt es Fortschritte in der KI. Aber es gibt ebenso viele ungelöste Probleme, sodass die humane Intelligenz der künstlichen noch Jahrzehntlang weit überlegen sein wird«, antwortet Steels auf die Anfrage aus der ZEIT-Redaktion. Außerdem dürfe man nicht »echte« und »falsche« Intelligenz verwechseln. Die meisten KI-Systeme, »insbesondere jene, die auf Deep Learning basieren«, besitzen Steel zufolge nur *fake intelligence*: Sie imitierten lediglich Muster des humanen Verhaltens, ohne wirkliches Verständnis zu entwickeln. So könnten sie beispielsweise einen Text übersetzen oder eine medizinische Diagnose allein aufgrund von statistischen Mustern erstellen – aber würden weder den Text noch die Krankheit wirklich verstehen. »Deshalb können diese Systeme auch nicht erklären, warum sie zu einer bestimmten Schlussfolgerung kommen – außer indem sie auf Muster verweisen, die früher häufig aufgetreten sind.«

Abgesehen davon, gebe es eine ganze Reihe »entscheidender Aspekte« des Menschseins, über die in der KI-Forschung noch nicht einmal nachgedacht werde, kritisiert Steels. Einer davon sei Empathie, also die Fähigkeit, sich in die Gefühle und Sichtweisen anderer Personen hineinversetzen zu können. »Wir können den Schmerz eines anderen fühlen oder von dessen Tränen berührt sein, weil wir einen Körper haben, der ebenso Schmerzen und Trauer fühlen kann.«

Ein anderes Alleinstellungsmerkmal des Menschen ist für Steels die Kunst: »Werke wie Dantes *Göttliche Komödie*, Wagners *Ring* oder die Arbeiten von Joseph Beuys« zeigten, wie vielfältig die geistigen Welten seien, die Menschen entwerfen können. »Durch die Kunst reflektieren wir uns selbst und lernen, wie wir mit anderen umgehen können. Auf diese Weise schaffen Menschen eine gemeinsame Kultur, die sie verbindet und ihnen gemeinsames kollektives Denken und Handeln ermöglicht.« Was man bisher in diesem Bereich von KI-Systemen gesehen habe, die angeblich »Kunst« erzeugten, sei davon meilenweit entfernt. Ergo: Das Menschsein beinhaltet »so eine enorme Bandbreite an Kompetenzen und Fähigkeiten«, dass gar nicht daran zu denken sei, dass die Roboter in absehbarer Zeit zu uns aufschließen.

Das klingt beruhigend. Bei näherem Hinsehen fällt allerdings auf, dass Steels kein prinzipielles Urteil fällt, sondern seine KI-Kritik auf die »absehbare Zeit« begrenzt und nur behauptet, die menschliche Intelligenz bleibe der künstlichen »noch Jahrzehntlang« überlegen. Wie aber sieht es danach aus? Könnte auf lange Sicht Schmidhuber nicht doch recht behalten? Und so schön Empathie und Kunst sind: Auch Amerikas Ureinwohner waren sicherlich empathisch und künstlerisch begabt – und mussten trotzdem vor der Aggressivität der Europäer kapitulieren.

Man müsste einmal einen ganz und gar neutralen Experten fragen, der sich nicht von menschlichen Ängsten oder Hoffnungen treiben lässt, sondern objektiv und unvoreingenommen urteilt. Wie wäre es mit dem klügsten Computer der Welt? Kann er uns nicht Auskunft geben über die zu erwartende Rolle des Menschen?

Also eine Anfrage bei IBM, um ein Interview zu führen mit dem berühmten Computerprogramm Watson. 2011 gewann Watson in einer US-Quizshow überlegen gegen zwei menschliche Kandidaten, die zuvor Rekordsummen gewonnen hatten. Sein Triumph war ein Meilenstein der KI-Geschichte wie der Sieg des Schachcomputers Deep Blue 1997 gegen Garri Kasparow.

Heute wird Watson als Expertensystem für Versicherungen, Banken und andere Unternehmen angeboten. »Um Menschen zu helfen, ihre täglichen Aufgaben besser zu bewältigen«, wie IBM wirbt.



KI von heute ist nur fake intelligence. Auf Dauer machen den Menschen Kunst und Empathie einzigartig

Luc Steels

Der Computer kann jedes Zitat über das Menschsein finden. Aber er wird nie sagen können, was es ausmacht

Wolfgang Hildesheim



Durch die Kombination von Sprach-, Bild- und Textanalyse mit moderner Datenverarbeitung sei Watson dafür ausgelegt, »zu verstehen, Schlüsse zu ziehen und zu lernen«. Die Arbeit mit ihm sei kinderleicht: »Der Mensch fragt, die Maschine antwortet.«

Ganz so leicht ist es dann aber doch nicht. Vorher muss man nämlich Wolfgang Hildesheim fragen, der als Leiter Watson und KI-Innovation bei IBM so etwas ist wie Watsons Vormund. Hildesheim erklärt, dass Watson kein fertiges Superhirn ist, sondern eher eine Art Werkzeugkasten, genauer: »eine Plattform mit kognitiven Services«, aus denen sich jeder sein individuelles Expertensystem zusammenstellen und es mit einschlägigen Daten trainieren muss, bevor es sinnvolle Antworten liefert.

Gesetzt also den Fall, man würde Watson mit der gesamten philosophischen, anthropologischen und medizinischen Literatur zum Thema Menschsein füttern – könnte man ihn dann fragen, was den Menschen ausmacht? »Nun ja, er könnte wohl alle Textstellen durchforsten, von Aristoteles bis heute, wo es um den Menschen geht«, sagt Hildesheim, »wie eine Art erweiterte schlaue Lesehilfe.« Aber auf die »feuilletonistische Frage, was langfristig der Unterschied zwischen Mensch und Maschine ist«, erhielt man keine Antwort, meint Hildesheim. Denn »den Kontext muss immer der Mensch festlegen.«

So kann der Computer Watson die Frage nach dem Menschsein leider nicht beantworten. Wer allerdings eine Antwort liefert, ist der Mensch Hildesheim, der sich am Ende doch noch zu einer persönlichen Stellungnahme gedrängt fühlt: »Was den Menschen von der Maschine unterscheidet, ist, dass wir letztendlich liebende und geliebte Wesen sind. Wir wurden geboren und sind durch die Liebe unserer Eltern groß geworden. Ohne diese Liebe wären wir gestorben. Deshalb betonen ja auch viele Religionen die Fähigkeit zum Mitgefühl, zu Empathie und Nächstenliebe.« Aus dieser Fähigkeit entwickle sich »unser Wertekontext, unser Handeln und letztlich die ganze differenzierte Struktur der Gesellschaft«. Und dann fällt Hildesheim noch etwas ein. »Außerdem werden wir wohl nie einen Roboter küssen wollen. Das bleibt immer uns Menschen vorbehalten.«

Ziehen wir Zwischenbilanz. Was uns von den Maschinen dauerhaft unterscheidet wird, sind bislang: Empathie, Kunst, Liebe und das Küssen. Zweifelloser lauter schöne Dinge! Aber genügen sie, um dauerhaft die Oberhand über immer klügere Maschinen zu behalten? Oder gibt es auch ganz pragmatisch-praktische Fähigkeiten, die wir den Maschinen prinzipiell voraushaben?

Aber sicher, sagt John-Dylan Haynes. Der deutsch-britische Hirnforscher ist weniger an Computern als vielmehr am menschlichen Gehirn interessiert. Aber in der modernen KI-Forschung wachsen die Disziplinen derzeit zusammen. Das sieht man schon daran, dass ein früherer Kollege von Haynes am University College in London später das Unternehmen DeepMind gründete (das etwa das siegreiche Go-Programm AlphaGo entwickelt hat und längst von Google aufgekauft worden ist). Heute ist Haynes Professor an der Berliner Charité und Humboldt-Universität. Dort benutzt er Algorithmen, um aus Hirndaten Rückschlüsse auf die Gedanken seiner Probanden zu ziehen (ZEIT Nr. 17/08). Dabei hat er die Möglichkeiten ebenso wie die Grenzen der KI kennengelernt.

Erfolgreich ist sie in Haynes' Augen nur bei Problemen, »die gut formalisierbar und mathematisch beschreibbar sind« – so wie Schach und Go. Dabei gebe es zwar schier unendlich viele mögliche Spielzüge, dennoch sei »der Zustandsraum wesentlich weniger komplex als irgendein Alltagsszenario«. Selbst Poker ist für Haynes ein überschaubares Problem. »Wenn man online spielt und als Datenbasis das vergangene Verhalten einer Person hat, lässt sich das relativ einfach reduzieren auf Spielzüge, die derjenige mit bestimmten Karten gemacht hat, und dann lässt sich die Wahrscheinlichkeit dafür ausrechnen, dass jemand blufft.« So werde auch Poker berechenbar.

Genüßlich zitiert Haynes in diesem Zusammenhang das Moravec'sche Paradox, das auf den KI-Forscher Hans Moravec zurückgeht. Es besagt, dass – anders als früher erwartet – ausgerechnet die höheren geistigen Fähigkeiten leicht mit einem Computer nachzubauen sind. Hingegen sind scheinbar einfache körperliche Bewegungen, wie eine Eiswaffel zu greifen, extrem schwer mit einem Computer umzusetzen. »Selbst bei den leistungsfähigsten Schach- oder Go-Programmen braucht es am Ende einen Menschen, der tatsächlich die Figuren zieht«, sagt Haynes. »Das scheinbar Simpelste müssen doch immer noch wir übernehmen.«

Die für den Menschen einfachsten Alltagsdinge wie putzen und selbst der Griff nach einer Schachfigur bereiten den Maschinen enorme Probleme

John-Dylan Haynes



Was für Haynes den Menschen überlegen macht, ist seine ungeheure Flexibilität. Was jede Reinigungskraft mühelos beherrscht, würde einen Roboter vor ungeheure Herausforderungen stellen: »Der müsste ja nicht nur saugen, sondern auch die Schränke aufmachen können, im Regal wischen, vorher alles zur Seite stellen, entscheiden, was gebraucht und was weggeworfen werden kann, und so weiter.« Angesichts der Komplexität solcher Alltagsdinge werde »schnell deutlich, dass solche Anwendungen nur sehr schwer umzusetzen sind – schon allein aus Kostengründen.«

Doch nicht nur seine körperliche Flexibilität zeichne den Menschen aus, sondern auch seine Fantasie und Erfindungsgabe. »Welche künstliche Intelligenz wäre etwa vor 15 Jahren auf die Idee gekommen, so etwas wie das iPhone zu entwickeln?«, fragt Haynes. »So etwas hatte es bis dahin noch nicht gegeben, das wäre für eine Maschine ein ganz schlecht formalisierbares Problem gewesen. Sie wäre vermutlich noch nicht einmal auf die Idee gekommen, dass es dafür einen Bedarf geben könnte.« Echte Innovationsprozesse erwartet sich Haynes von der KI daher nicht, das bleibe immer noch dem Menschen vorbehalten.

Für ihn ist das größte Problem auch nicht die Leistungsfähigkeit der Maschinen, sondern die Gefahr, dass wir diese überschätzen und uns zu sehr auf die Computer verlassen. Damit stößt Haynes ins selbe Horn wie auch andere Forscher (ZEIT Nr. 38/17). »Wir neigen dazu, solchen technischen Systemen blind zu vertrauen, obwohl sie Schwächen haben und auf unvorhergesehene Entwicklungen oft schlecht reagieren.« Vor allem bei komplexen Anwendungen – etwa der weltweiten Dynamik der Finanzmärkte – könne das leicht zu »catastrophic fails« führen, warnt Haynes. Und sei die Katastrophe erst einmal da, helfen keine Algorithmen, sondern nur innovatives Krisenmanagement. »Und das können Menschen in der Regel besser, weil sie gelernt haben, mit Fehlern umzugehen. Sie sind ja selbst fehleranfällig.«

Vielleicht sollte man an dieser Stelle aufhören. Es ist ja doch mittlerweile eine beruhigend lange Liste von menschlichen Alleinstellungsmerkmalen zusammengelassen, die einem das schöne Gefühl gibt, unsere Gattung werde auch in Zukunft das Heft in der Hand behalten. Zu Kunst, Liebe, Empathie und dem Küssen darf man als Pluspunkte die menschliche Fehleranfälligkeit zählen sowie Flexibilität und Erfindungsgabe, die wir den Maschinen voraushaben. Zwar ist diese Liste sicher nicht vollständig, weil so illustre Eigenschaften wie das menschliche Bewusstsein, die Sterblichkeit oder der Begriff Seele noch gar nicht bedacht sind. Dennoch könnte man es dabei bewenden lassen, Schmidhubers irritierende Vision vom Aufbruch der KI ins Weltall vergessen und sich entspannt zurücklehnen.

Man dürfe nur nicht bei Christoph von der Malsburg anrufen. Der Grandseigneur der KI-Forschung ist eigentlich wie geschaffen für ein abrundendes Schlusswort, weil er auf diesem Feld schon alles gesehen hat. Schon in den 1990er Jahren erforschte der Physiker und Neurobiologe neuronale Netzwerke und Mustererkennung, lange bevor diese populär wurden. Heute, mit 75 Jahren, arbeitet er als Senior Fellow am Frankfurt Institute for Advanced Studies, fördert gezielt junge Talente mithilfe der Mindfire Stiftung (die das Ziel verfolgt, »die Prinzipien der humanen Intelligenz zu entschlüsseln«) und hat nichts von seinem blitzwachen Biss verloren. Er kennt die Schwächen der KI wie kaum ein Zweiter und müsste einem doch sagen können, wo diese dem Menschen nie wird das Wasser reichen können. Doch das Gespräch verläuft anders als erwartet.

Zwar zeigt sich von der Malsburg vom aktuellen Hype wenig beeindruckt und argumentiert, ähnlich wie Luc Steels, dass die gegenwärtigen Systeme nur Muster erkennen, aber kein echtes kognitives Verständnis entwickeln. »Kinder können ab etwa drei Jahren ein komplexes Objekt – etwa ein Dreirad oder eine TV-Fernbedienung – in Sekundenschnelle in sich aufnehmen und es danach immer wiedererkennen, selbst unter anderer Beleuchtung, mit anderer Farbe, in anderer Position und so weiter. Das können die neuronalen Netze derzeit nicht«, sagt von der Malsburg.

Die gegenwärtige Diskussion über die Gefahren der künstlichen Intelligenz hält er für »marktschreierisch und wenig inspiriert«. Die Bücher von Warnern wie Nick Bostrom (*Superintelligenz*) oder Max Tegmark (*Leben 3.0*) zeigen für von der Malsburg vor allem, »dass die Autoren keine Ahnung haben, wie Gehirne funktionieren und wie künstliche Intelligenz aussehen wird«. Viele Prognosen – etwa dass man künftig sein Bewusstsein auf einen Rechner kopieren könne – hält er für völlig überzogen.

Die Gefahr durch künstliche Intelligenz ist für die Menschheit schlimmer als die Atombombe

Christoph von der Malsburg



Doch dann sagt von der Malsburg plötzlich: »Ich denke trotzdem, man muss die Gefahr der künstlichen Intelligenz sehr ernst nehmen, das ist schlimmer für die Menschheit als die Atombombe.« Wie bitte? Wieso das?

Nun, weil der Neuroforscher entscheidende Fortschritte erwartet. »Die Technik ist da, und Heere von klugen jungen Leuten, die hektisch Informatik lernen, springen jetzt auf diesen Zug auf. Jeder kann sich die Technik des Deep Learning in zwei Wochen beibringen, wenn man die entsprechenden Programme herunterlädt und sich Tutorials anhört.« Da ist es für von der Malsburg nur eine Frage der Zeit, bis »Intelligenz in einem echten Sinne technisch realisiert wird«. Und dann fügt er noch hinzu: »Und dann gnade uns Gott, unser Selbstverständnis wird einen schweren und ganz irreparablen Einbruch erleben.«

Übertreibt er da nicht? Auch er, erzählt von der Malsburg, habe bis vor einem halben Jahr die Gefahr nicht ernst genommen. Doch dann hätten ihn die Bücher von Bostrom, Tegmark und anderen – obwohl darin viel Ungereimtes stünde – für die KI-Gefahr sensibilisiert. »Und das evolviert in meinem Gehirn im Augenblick enorm.«

Wie die Zukunft aussieht, kann natürlich auch er nicht sagen. »Eine Horrorgeschichte wäre etwa, dass eine weltumspannende Intelligenz entsteht, die alle verfügbaren Daten zur Verfügung hat – etwa all jene Daten, die derzeit eifrig von Google, Facebook und Co. im Netz eingesammelt werden.« Oder es könnten Roboter entstehen, die ähnlich wie Kinder in rasender Geschwindigkeit lernen, mit der Welt umzugehen.

Noch ist es nicht so weit. »Aber das wird fundamental anders werden, sobald künstliche Organismen Situationsbewusstsein und flexible Entscheidungsmechanismen bekommen«, prophezeit von der Malsburg. Dabei gehe es gar nicht so sehr um Intelligenz, sondern eher um das Programmieren jener Instinkte, die uns Menschen von der Evolution mitgegeben wurden – also Überlebenswille und Besitzstreben und Neugier und so weiter. Tatsächlich wird schon heute in Labors daran gearbeitet, künstlichen Wesen solch eine Motivationsstruktur einzubauen. Dass dies noch gestoppt werden könne, glaubt von der Malsburg nicht.

»Ein einzelner Staat kann natürlich Verbote aussprechen, aber es gibt eben auch immer jene, die sich von der Weiterentwicklung einen Vorteil versprechen und fortgeschrittene künstliche Intelligenz als Waffe einsetzen wollen. Suchen Sie nur einmal im Netz nach dem Stichwort Slaughterbots«, empfiehlt von der Malsburg. Das vermittelt einen Vorgeschmack auf den drohenden weltweiten Wettlauf um KI-Waffensysteme.

Und wo sehen Sie künftig den Platz des Menschen, Herr von der Malsburg? »Das ist genau der Punkt, der mich seit Monaten in einen Panikmodus versetzt: Wir werden wie Tiere im Zoo leben.« Wenn die Blechkisten wirklich klüger und weiser würden, »dann werden sie uns zu unserem eigenen Besten daran hindern, uns die Köpfe einzuschlagen und die Ressourcen zu verbrauchen. Die künstliche Intelligenz wird als wohlwollender Diktator auftreten und uns wie Kinder bevormunden, zu unserem Allerbesten.« Rumm.

Während man das noch zu verdauen versucht, kommt eine E-Mail von Nick Bostrom herein, dem schwedischen Philosophen, der in Oxford vor den Gefahren der KI warnt. Zur Frage nach dem Alleinstellungsmerkmal des Menschen schreibt er nur einen einzigen Satz: »Wenn es gut ausgeht, werden wir eine ähnliche Rolle spielen wie die Kinder in Disneyland.«

Das ist, in anderen Worten, dieselbe Prognose wie die von Schmidhuber und von der Malsburg: Menschen könnten künftig »einfach keine wichtigen Entscheidungsträger mehr« sein. Wir dürften zwar noch empathisch sein und Kunst machen und uns küssen; wohl dürften wir uns auch an Erfindungen versuchen und Fehler machen und unsere Wohnungen allein putzen – die wesentlichen Dinge würden jedoch von einer künstlichen Intelligenz geregelt, die meilenweit über unsere Köpfe hinweg entscheidet.

An diesem Punkt der Recherche, das muss der Autor eingestehen, ist er selbst in eine zeitweilige Depression versunken. Natürlich gibt es keine Garantie dafür, dass solch erschreckende Prognosen Wirklichkeit werden. Es ist derzeit aber auch nicht erkennbar, dass diese Aussichten die Forschungsanstrengungen auch nur einen Deut bremsen würden. Im Gegenteil: Bei den großen Digitalkonzernen tobt ein verbissener KI-Wettlauf, der keinerlei staatlichen Regelung unterworfen ist und keine Furcht vor dem Morgen zu kennen scheint. *Digital first*, Bedenken *second*.

Steuern wir also als Gattung sehenden Auges auf den Untergang zu?



Falls die Menschen Glück haben, dann werden sie von den herrschenden Computern bespaßt wie Kinder in Disneyland

Nick Bostrom



Maschinen könnten uns als objektive Berater dabei helfen, unsere ethischen Probleme zu lösen

Thomas Metzinger

Das klingelt doch noch einmal das Telefon, ein Rückruf des Philosophen Thomas Metzinger. Dieser kluge Kopf verfolgt die Fortschritte der Neuro- und KI-Forschung seit drei Jahrzehnten und wird, wie er erzählt, »in letzter Zeit immer mehr in die Debatte über die Ethik der KI reingezogen«, ob er wolle oder nicht. »Die Zahl der Anfragen und Einladungen ist einfach überwältigend.«

Und zu welchem Urteil ist er gekommen? Teilt er die Besorgnis der Warner, oder hält er deren Szenarien für überzogen? Da sagt Metzinger ganz offen: »Bei mir geht es immer hin und her. Mal sagt eine Stimme in mir: Andere Probleme sind viel drängender, etwa der Klimawandel oder die weltweiten Migrationsbewegungen. Dann wieder denke ich: Nein, wenn die tiefen Netzwerke sich weiter unkontrolliert selbst optimieren, steht irgendwann unsere Existenz auf dem Spiel.«

Dass die Entwicklung der künstlichen Intelligenz extreme Risiken mit sich bringe, sei nicht von der Hand zu weisen. Ebenso besorgt Metzinger, »dass unsere politischen Institutionen bei diesem Thema tief schlafen« und das Problem noch nicht einmal richtig erkannt hätten. »Stattdessen schießen nun überall einzelne Institutionen aus dem Boden – wie etwa Bostroms Future of Humanity Institute oder Max Tegmarks Future of Life Institute –, die versuchen, die ethischen Normen im Umgang mit der KI zu definieren.« Dabei könne man über ethisch »richtiges« Verhalten ja durchaus streiten.

Statt nun aber selbst eine »Maschinen-Ethik« zu formulieren, verblüfft Metzinger mit einem ungewöhnlichen Vorschlag: »Wenn die künstliche Intelligenz in Zukunft wirklich so leistungsfähig wird, dann könnten wir sie doch auch für ethische Analysen nutzen.« Ihm schwebt ein »rationales, evidenzbasiertes Ethik-Expertensystem« vor, das »uns bei der Frage berät, wie wir mit der künstlichen Intelligenz am besten umgehen.«

Spätestens mit dem Einsatz selbstfahrender Autos müsse man ja anfangen, ethisch-moralische Abwägungen zu formalisieren (»Fahre ich lieber in den Graben, oder überrolle ich den Hund?«). Irgendwann könne es durchaus sein, »dass die Maschinen aufgrund ihrer Unbestechlichkeit und ihrer riesigen Datenbasis zu rationaleren und allgemeingültigeren ethischen Urteilen kommen als wir Menschen.«

Natürlich will Metzinger das nicht so verstanden wissen, dass man künftig Werte mithilfe von Maschinen erkennen und über normative Fragen einfach technisch entscheiden könne. »So naiv dürfen wir uns das nicht vorstellen«, sagt Metzinger, denn: »Normative Sätze haben ja keine Wahrheitswerte.« Er will auch nicht suggerieren, dass man künftig »einen epistemischen Superagenten oder eine Art künstlichen Heiligen hat, der direkten Zugang zu einem mysteriösen Bereich höherer moralischer Wahrheiten besitzt.«

Eher kann er sich vorstellen, dass man mithilfe der künstlichen Intelligenz »nach dem optimalen Weg sucht, um unsere widerstreitenden moralischen Intuitionen miteinander in Einklang zu bringen«. Denn das große Zukunftsrisiko sieht Metzinger nicht allein in der KI, sondern in der Kombination der maschinellen Fähigkeiten mit den oft egoistischen, unaufgeklärten Denkmustern des Homo sapiens. Da sei es möglicherweise hilfreich, »wenn ein künstliches System auf der Basis aller verfügbaren wissenschaftlichen Daten herauszufinden versuche, was denn tatsächlich in unserem eigenen wohlverstandenen, aufgeklärten Eigeninteresse liegen würde.«

Dass solche Überlegungen viele seiner geisteswissenschaftlichen Kollegen extrem irritieren dürften, ist Metzinger bewusst. Aber die Idee hat etwas: Dass uns am Ende ausgerechnet die KI dabei helfen könnte, neben vielen anderen Problemen auch sich selbst im Zaum zu halten – das wäre schon eine originelle Pointe des Fortschritts.

Und so liefert Metzinger ganz nebenbei noch ein weiteres Argument für die Besonderheit des Menschen: Was ihn auszeichnet, ist seine Fähigkeit zum Perspektivwechsel und zur Zuversicht, selbst im größten Problem noch eine Chance zu entdecken. Und vielleicht brauchen wir tatsächlich das Spiegelbild der künstlichen Intelligenz, um dieses Potenzial des Menschen in seiner ganzen Fülle wahrzunehmen.

Ob all die Fähigkeiten des Homo sapiens – Liebe, Empathie, Kunst, Fehleranfälligkeit, Flexibilität, Innovationskraft, moralisches Denken und Zuversicht – am Ende ausreichen, die Maschinen auf Dauer unter Kontrolle zu halten, wird man sehen. Aber wenn wir nicht beginnen, über die Unterschiede nachzudenken, haben wir schon jetzt verloren.

Lesen Sie auch: Laura Cwiertnias Selbstversuch über das Leben mit Alexa, Wirtschaft Seite 24/25