mitp

Die Intelligenz der Maschinen

Mit Koryphäen der Künstlichen Intelligenz im Gespräch

Innovationen, Chancen und Konsequenzen für die Zukunft der Gesellschaft

Financial Times
Best books of 2018: Technology

Yoshua Bengio
Stuart Russell
Geoffrey Hinton
Nick Bostrom
Yann LeCun
Fei-Fei Li
Demis Hassabis
Andrew Ng
Rana el Kaliouby
Ray Kurzweil
Daniela Rus
James Manyika
Gary Marcus
Barbara Grosz
Judea Pearl
Jeff Dean
Daphne Koller
David Ferrucci
Rodney Brooks
Cynthia Breazeal
Josh Tenenbaum
Oren Etzioni
Bryan Johnson

Im Interview mit
Martin Ford

Vorwort von
Jürgen Schmidhuber

INHALT

EINLEITUNG

MARTIN FORD

AUTOR, ZUKUNFTSFORSCHER

Künstliche Intelligenz (KI) ist zunehmend keine Science Fiction mehr, sondern wird schnell zur alltäglichen Realität. Unsere Geräte verstehen, was wir sagen, antworten uns und übersetzen Fremdsprachen immer flüssiger. KI-gestützte Bilderkennungsalgorithmen übertreffen Menschen und finden überall Anwendung, wie in selbstfahrenden Autos oder Systemen, die anhand von medizinischen Aufnahmen Krebs diagnostizieren. Große Medienkonzerne nutzen zunehmend einen »automatisierten Journalismus«, der Rohdaten in zusammenhängende Meldungen umwandelt, die von denen eines menschlichen Journalisten praktisch nicht zu unterscheiden sind.

Die Liste der Fähigkeiten wird immer länger, und es wird offensichtlich, dass KI eine der bedeutendsten Kräfte sein wird, die unsere Welt prägen. Im Gegensatz zu spezialisierteren Innovationen wird KI zu einer echten Allzwecktechnologie. Mit anderen Worten: KI entwickelt sich zu einem Bedarfsgut, ähnlich der Elektrizität, das in allen Industriezweigen und in allen Wirtschaftssektoren Verwendung findet und praktisch alle Bereiche der Wissenschaft, der Gesellschaft und der Kultur durchdringt.

Die in den letzten Jahren unter Beweis gestellte Leistungsfähigkeit der KI hat zu umfangreicher Medienberichterstattung und einer Vielzahl von Kommentaren geführt. Unzählige Zeitungsberichte, Bücher, Dokumentarfilme und Fernsehsendungen zählen die Erfolge der KI auf und verkünden den Anbruch eines neuen Zeitalters. Das Ergebnis war manchmal eine unverständliche Mischung aus sorgfältiger Analyse, zusammen mit Hype, Spekulationen und dem, was man fast schon als Panikmache bezeichnen kann. Es heißt, dass schon in ein paar Jahren selbstfahrende Autos auf unseren Straßen unterwegs sein werden – und dass Millionen Jobs für Fernfahrer, Taxifahrer und Uber-Fahrer bald verschwinden werden. Bestimmte Machine-Learning-Algorithmen zeigen ein rassistisches oder frauenfeindliches Verhalten. Bedenken, welchen Einfluss KI-gestützte Technologien wie Gesichtserkennung auf die Privatsphäre haben, erscheinen also wohlbegründet. Die Medien berichten regelmäßig über Warnungen, dass Roboter bald bewaffnet sein werden oder dass wirklich intelligente (oder superintelligente) Maschinen eine existenzielle Bedrohung für die Menschheit darstellen. Eine Reihe sehr bekannter Personen des öffentlichen Lebens – von denen keine tatsächlich ein KI-Experte ist – haben sich zu Wort gemeldet. Elon Musk hat sich besonders extrem dazu geäußert und erklärt, dass die KI-Forschung »einen Dämon heraufbeschwört« und dass »KI gefährlicher als Nuklearwaffen« sei. Auch weniger launenhafte Persönlichkeiten, wie Henry Kissinger oder zuletzt Stephen Hawking, haben düstere Warnungen ausgesprochen.

Dieses Buch möchte das Forschungsgebiet Künstliche Intelligenz (KI) sowie die damit verbundenen Chancen und Risiken beleuchten. Dazu dient eine Reihe ausführlicher, weitreichender Gespräche mit einigen der bekanntesten KI-Forschern und in diesem Bereich tätigen Unternehmern. Viele dieser Menschen haben wegweisende Beiträge geleistet, auf denen die überall stattfindenden Umwälzungen unmittelbar beruhen. Andere haben Unternehmen gegründet, die die Forschung in den Bereichen KI, Robotik und Machine Learning vorantreiben.

Das Erstellen einer Liste der bekanntesten und einflussreichsten Persönlichkeiten eines Fachgebiets ist natürlich ein subjektives Unterfangen, und zweifelsohne gibt es viele andere Leute, die entscheidende Beiträge für die Weiterentwicklung der KI geleistet haben oder leisten. Dessen ungeachtet bin ich mir sicher, dass jeder mit

fundierten Kenntnissen des Fachgebiets, der eine Liste der führenden Köpfe erstellen sollte, die die aktuelle KI-Forschung prägen, zu einer Namensliste gelangen würde, auf der sich die meisten der in diesem Buch interviewten Persönlichkeiten befinden. Die hier vorgestellten Männer und Frauen sind tatsächlich die Architekten der Künstlichen Intelligenz – und darüber hinaus der Revolution, die sie bald auslösen wird.

Die hier wiedergegebenen Gespräche waren grundsätzlich zeitlich nicht begrenzt, sind aber so angelegt, dass die dringendsten Fragen zur Sprache kommen, mit denen wir uns bei der Fortentwicklung der KI konfrontiert sehen: Welche Ansätze und Technologien sind am vielversprechendsten, und welche Durchbrüche sind in den kommenden Jahren zu erwarten? Stellen »denkende Maschinen«, also KI auf menschlichem Niveau, eine echte Möglichkeit dar, und wann könnte ein solcher Durchbruch erfolgen? Welchen Risiken oder Bedrohungen, die mit KI verbunden sind, sollten wir ernsthafte Beachtung schenken? Und wie können wir diese Probleme in Angriff nehmen? Welche Rolle spielt eine staatliche Regulierung? Wird KI in der Wirtschaft und auf dem Arbeitsmarkt drastische Umwälzungen verursachen, oder sind solche Bedenken übertrieben? Könnten superintelligente Maschinen sich eines Tages unserer Kontrolle entziehen und zu einer ernsthaften Bedrohung werden? Sollten wir uns Sorgen machen, dass es zu einem »KI-Rüstungswettlauf« kommt, oder dass andere Länder mit autoritären politischen Systemen, insbesondere China, irgendwann die Führung übernehmen?

Es versteht sich von selbst, dass niemand die Antworten auf diese Fragen kennt. Niemand kann die Zukunft vorhersagen. Aber die KI-Experten, mit denen ich gesprochen habe, wissen mehr über den Status quo der Technologie und über die am Horizont erkennbaren Innovationen, als irgendjemand sonst. Sie verfügen über oft jahrzehntelange Erfahrung und haben zur Entstehung der Revolution beigetragen, die sich jetzt anbahnt. Deshalb sollte man ihren Überlegungen und Ansichten Beachtung schenken. Neben den Fragen zum Fachgebiet KI und dessen Zukunft habe ich mich auch mit den Hintergründen, den beruflichen Laufbahnen und den derzeitigen Forschungsinteressen dieser Persönlichkeiten befasst, und ich glaube, dass die diversen Werdegänge und die vielfältigen Lebenswege, die zu ihrer Bekanntheit geführt haben, für eine faszinierende und anregende Lektüre sorgen.

Künstliche Intelligenz ist ein umfassendes Forschungsgebiet mit einer ganzen Reihe von Teilgebieten, und viele der interviewten Forscher waren auf mehreren davon tätig. Einige von ihnen haben auch fundierte Erfahrung in anderen Fachgebieten, wie etwa der Untersuchung der menschlichen kognitiven Fähigkeiten. Dennoch versuche ich im Folgenden kurz zu beschreiben, in welcher Beziehung die hier interviewten Persönlichkeiten zu den wichtigsten Innovationen der KI-For-

schung und den zukünftigen Herausforderungen stehen. Weitere Informationen über die Personen können Sie den Kurzbiografien entnehmen, die Sie jeweils nach dem Interview finden.

Der größte Teil der bahnbrechenden Fortschritte, die im letzten Jahrzehnt erzielt wurden (Bild- und Gesichtserkennung, Übersetzung von Fremdsprachen, AlphaGos Erfolge beim Go-Spiel), beruht auf einer Technologie, die als *Deep Learning* bezeichnet wird. Der Begriff *tiefe neuronale Netze* ist ebenfalls gebräuchlich. Künstliche neuronale Netze, in denen, vereinfacht gesagt, Software die Strukturen und Interaktionen biologischer Neuronen im Gehirn emuliert, gibt es schon seit den 1950er-Jahren. Einfache Versionen dieser Netze können rudimentäre Aufgaben der Mustererkennung erledigen und sorgten damals bei den Forschern für ziemlichen Enthusiasmus. In den 1960er-Jahren verloren die Forscher jedoch das Interesse an neuronalen Netzen (was zumindest teilweise auf Kritik an der Technologie von Marvin Minsky, einem der ersten KI-Pioniere, zurückzuführen ist), und sie wurden kaum noch verwendet, während sich die Forscher mit anderen Ansätzen befassten.

Während eines etwa 20-jährigen Zeitraums, der in den 1980er-Jahren begann, befasste sich eine kleine Gruppe von Forschern weiterhin mit neuronalen Netzen und konnte Fortschritte erzielen. Zu dieser Gruppe gehörten insbesondere Geoffrey Hinton, Yoshua Bengio und Yann LeCun. Diese drei Männer leisteten nicht nur maßgebliche Beiträge zur mathematischen Theorie, auf der Deep Learning beruht, sie waren auch die bedeutendsten Fürsprecher der Technologie. Zusammen gelang es ihnen, sehr viel ausgeklügeltere – tiefe – neuronale Netze zu entwickeln, die aus vielen Schichten künstlicher Neuronen bestanden. Fast wie die mittelalterlichen Mönche, die klassische antike Texte verwahrten und kopierten, führten Hinton, Bengio und LeCun neuronale Netze durch ein dunkles Zeitalter – bis nach jahrzehntelangem exponentiellen Wachstum der Rechenleistung und einer nahezu unfassbaren Zunahme der Menge verfügbarer Daten eine Renaissance des Deep Learnings möglich wurde.[1] Aus dem Fortschritt wurde 2012 eine Revolution, als ein Team von Hintons Doktoranden an der University of Toronto an einem bedeutenden Bilderkennungswettbewerb teilnahm und die Konkurrenz dank des Einsatzes von Deep Learning weit hinter sich ließ.

In den darauffolgenden Jahren wurde Deep Learning allgegenwärtig. Alle bedeutenden Technologiekonzerne – Google, Facebook, Microsoft, Amazon, Apple sowie

1 Für ihre herausragenden konzeptionellen und technischen Arbeiten im Bereich Deep Learning erhielten Yoshua Bengio, Geoffrey Hinton und Yann LeCun Ende März 2019 den Turing Award für 2018.

führende chinesische Firmen wie Baidu und Tencent – investierten in die Technologie und machten sie sich zunutze. Die Unternehmen, die Mikroprozessoren und Grafikchips (GPUs) entwickeln, wie NVIDIA und Intel, mussten ihr Geschäft anpassen und für neuronale Netze optimierte Hardware produzieren. Deep Learning ist, zumindest bis jetzt, die wichtigste Technologie, die für die KI-Revolution verantwortlich ist.

In dieses Buch sind Gespräche mit den drei Deep-Learning-Pionieren Hinton, LeCun und Bengio sowie mit einigen anderen sehr bekannten Spitzenforschern wiedergegeben. Andrew Ng, Fei-Fei Li, Jeff Dean und Demis Hassabis haben Fortschritte bei neuronalen Netzen ermöglicht, die in Bereichen wie Websuche, Computer Vision, selbstfahrenden Autos und eher allgemein intelligenten Systemen Verwendung finden. Sie sind auch als Lehrende, Leiter von Forschungsorganisationen und Unternehmer im Bereich Deep-Learning-Technologie tätig.

In den übrigen Interviews in diesem Buch kommen Personen zu Wort, die man nicht gerade als Deep-Learning-Experten bezeichnen würde, sondern vielleicht sogar als Kritiker. Sie erkennen zwar die im letzten Jahrzehnt erzielten Erfolge an, sind jedoch der Ansicht, dass es nur ein Verfahren unter vielen sei und dass weitere Fortschritte die Integration von Ideen aus anderen Bereichen der KI erfordern. Einige von ihnen, wie Barbara Grosz und David Ferrucci, haben sich stark auf das Problem des Verstehens natürlicher Sprache konzentriert. Gary Marcus und Josh Tenenbaum haben der Untersuchung der kognitiven Fähigkeit des menschlichen Gehirns viel Zeit gewidmet. Andere, wie Oren Etzioni, Stuart Russell und Daphne Koller, sind KI-Generalisten oder haben sich auf probabilistische Verfahren konzentriert. In der letzten Gruppe ist besonders Judea Pearl hervorzuheben, dem 2012 der Turing Award verliehen wurde – die höchste Auszeichnung in der Informatik, vergleichbar dem Nobelpreis –, hauptsächlich für seine Arbeit an probabilistischen Ansätzen in der KI und beim Machine Learning. Von dieser sehr groben Unterscheidung aufgrund ihrer Haltung zum Deep Learning einmal abgesehen, haben sich einige der Forscher, mit denen ich gesprochen habe, auf bestimmte Bereiche konzentriert. Rodney Brooks, Daniela Rus und Cynthia Breazeal sind in der Robotik die führenden Köpfe. Breazeal und Rana el Kaliouby sind bei der Entwicklung von Systemen führend, die Emotionen verstehen und darauf reagieren können und dadurch in der Lage sind, auf sozialer Ebene mit Menschen zu interagieren. Bryan Johnson hat ein Start-up namens Kernel gegründet und hofft, früher oder später Technologie zur Verbesserung der kognitiven Fähigkeiten des menschlichen Gehirns einzusetzen.

Es gibt drei allgemeine Themenbereiche, die ich für so interessant halte, dass ich sie in allen Interviews angesprochen habe. Da wären zunächst einmal die potenziellen Auswirkungen von KI und Robotik auf den Arbeitsmarkt und die Wirt-

schaft. Ich persönlich glaube Folgendes: Wenn KI allmählich unter Beweis stellt, dass sie alle sich wiederholenden, vorhersagbaren Aufgaben automatisiert lösen kann (wobei es keine Rolle spielt, ob es sich um Tätigkeiten von Angestellten oder Arbeitern handelt), wird es unweigerlich zu Ungerechtigkeiten und womöglich sogar zu Arbeitslosigkeit kommen, zumindest bei bestimmten Gruppen von Arbeitern. In meinem Buch *Rise of the Robots: Technology and the Threat of a Jobless Future* aus dem Jahr 2015 (deutscher Titel: *Aufstieg der Roboter*) habe ich das ausführlich dargelegt.

Meine Gesprächspartner hatten vielfältige Ansichten zu dieser potenziellen wirtschaftlichen Umwälzung und den möglichen politischen Gegenmaßnahmen. Ich habe mich an James Manyika gewendet, den Vorsitzenden des McKinsey Global Institute, um diesem Thema auf den Grund zu gehen. Als erfahrener KI- und Robotik-Forscher, der sich seit Kurzem mit den Auswirkungen dieser Technologien auf Organisationen und Arbeitsplätze befasst, hat Manyika eine einzigartige Sichtweise. Das McKinsey Global Institute ist bei Forschungen auf diesem Gebiet führend, und das Interview enthält viele wichtige Erkenntnisse über die Art der sich anbahnenden Umwälzungen.

Die zweite Frage, die ich allen Gesprächspartnern gestellt habe, betrifft die KI auf menschlichem Niveau, die auch als *Artificial General Intelligence* (AGI) bezeichnet wird. AGI ist von Anfang an der Heilige Gral des Fachgebiets KI gewesen. Ich wollte von meinen Gesprächspartner wissen, was sie von der Aussicht auf eine »denkende Maschine« halten, welche Hürden es zu überwinden gilt und wie lange es dauern wird, bis dieses Ziel erreicht ist. Hier hatten alle wichtige Einsichten beizutragen, aber drei Gespräche waren besonders interessant: Demis Hassabis erläutert die gegenwärtig bei DeepMind unternommenen Anstrengungen, der größten und am besten finanzierten Initiative, die darauf ausgerichtet ist, AGI zu erreichen. David Ferrucci, der das Team leitete, das IBMs Watson entwickelt hat, ist jetzt Geschäftsführer bei Elemental Cognition, einem Start-up, das allgemeine Intelligenz durch das Verständnis von Sprache erreichen möchte. Ray Kurzweil, der jetzt bei Google ein Forschungsprojekt über natürliche Sprache leitet, hat zu diesem Thema wichtige Ideen beigetragen (wie so viele andere auch). Kurzweil ist insbesondere durch sein Buch *The Singularity is Near* aus dem Jahr 2005 (deutscher Titel: *Menschheit 2.0: Die Singularität naht*) bekannt. 2012 veröffentlichte er ein Buch über maschinelle Intelligenz, *How to Create a Mind* (deutscher Titel: *Das Geheimnis des menschlichen Denkens*), das die Aufmerksamkeit von Larry Page erlangte, was zu seiner Anstellung bei Google führte.

Ich hatte bei den Gesprächen die Möglichkeit, diese außerordentlich qualifizierte Gruppe von KI-Forschern um eine Einschätzung zu bitten, wann AGI zur Realität wird. Ich stellte die Frage: »In welchem Jahr wird Ihrer Meinung nach mit einer

Wahrscheinlichkeit von 50 % KI auf menschlichem Niveau erreicht?«. Die meisten Teilnehmer zogen es vor, ihre Einschätzung anonym abzugeben. Ich habe das Ergebnis dieser formlosen Umfrage in einem Abschnitt am Ende des Buchs zusammengefasst. Zwei Teilnehmer waren bereit, ihre Einschätzung nicht anonym abzugeben, und diese lassen schon erahnen, wie sehr die Meinungen voneinander abweichen. Ray Kurzweil glaubt, wie er zuvor schon des Öfteren gesagt hat, dass KI auf menschlichem Niveau etwa im Jahr 2029 erreicht wird, also in rund zehn Jahren (vom jetzigen Zeitpunkt aus betrachtet). Rodney Brooks hingegen tippt auf das Jahr 2200 – also erst in 180 Jahren. Zu den faszinierendsten Aspekten der Gespräche gehören sicherlich die äußerst unterschiedlichen Ansichten zu einem breiten Spektrum wichtiger Themen.

Das dritte Diskussionsthema betrifft die verschiedenen Risiken, die bei der Weiterentwicklung der KI auftreten werden, sowohl in naher Zukunft als auch langfristig. Die Anfälligkeit von mit dem Internet verbundenen autonomen Systemen für Cyberangriffe oder Hacking gehört zu den Bedrohungen, die schon heute offensichtlich sind. Da KI immer stärker in Wirtschaft und Gesellschaft integriert wird, ist die Lösung dieses Problems eine der wichtigsten Herausforderungen, denen wir gegenüberstehen. Zu den schon jetzt bekannten Problemen gehört auch die Verzerrung (oder Bias) mancher Machine-Learning-Algorithmen, beispielsweise bezüglich der Rasse oder des Geschlechts. Viele meiner Gesprächspartner betonten, wie wichtig es sei, dieses Problem in Angriff zu nehmen und erzählten mir, dass entsprechende Forschungsarbeiten bereits im Gange sind. Einige waren auch durchaus optimistisch, weil sich KI eines Tages als leistungsstarkes Werkzeug im Kampf gegen systemische Verzerrung oder Diskriminierung erweisen könnte.

Eine Gefahr, die viele Forscher thematisieren, ist das Schreckgespenst vollständig autonomer Waffen. Viele Mitglieder der KI-Gemeinschaft glauben, dass KI-gestützte Roboter und Drohnen mit der Fähigkeit, ohne eine Freigabe durch einen Menschen zu töten, früher oder später genauso gefährlich und destabilisierend sind wie biologische oder chemische Waffen. Im Juli 2018 haben mehr als 160 KI-Unternehmen und 2.400 Forscher aus aller Welt, zu denen auch viele meiner Gesprächspartner gehören, eine öffentlich Erklärung unterzeichnet, in der sie versprechen, niemals derartige Waffen zu entwickeln (siehe *https://futureoflife.org/lethal-autonomous-weapons-pledge/*). Die von bewaffneter KI ausgehenden Gefahren kommen in mehreren der Interviews zur Sprache.

Eine sehr viel futuristischere und spekulativere Gefahr ist das sogenannte »Kontrollproblem«. Dabei handelt es sich um die Befürchtung, dass eine intelligente oder vielleicht superintelligente Maschine sich unserer Kontrolle entzieht oder Entscheidungen trifft, die negative Folgen für die Menschheit haben. Diese Be-

fürchtung ist es offenbar, die zu so übertriebenen Aussagen wie der von Elon Musk führen. Fast alle Gesprächspartner hatten zu diesem Thema etwas zu sagen. Um zu gewährleisten, dass ich diesen Bedenken angemessen und ausgewogen Raum gebe, habe ich mit Nick Bostrom vom Future of Humanity Institute der University of Oxford gesprochen. Bostrom ist der Autor des Bestsellers *Superintelligence: Paths, Dangers, Strategies* (deutscher Titel: *Superintelligenz: Szenarien einer kommenden Revolution*), das sich sorgfältig mit den potenziellen Risiken auseinandersetzt, die mit Maschinen verbunden sind, die womöglich viel schlauer als irgendein Mensch sind.

Alle Gespräche wurden zwischen Februar und August 2018 geführt und haben jeweils mindestens eine Stunde gedauert, manchmal auch erheblich länger. Sie wurden aufgezeichnet, transkribiert und von Verlagsmitarbeitern überarbeitet. Der Text wurde schließlich meinen Gesprächspartnern vorgelegt, um ihnen Gelegenheit zu geben, ihn zu korrigieren oder zu ergänzen. Deshalb habe ich vollstes Vertrauen, dass die hier wiedergegebenen Worte die Ansichten der von mir interviewten Personen korrekt widerspiegeln.

Die KI-Experten, mit denen ich gesprochen habe, sind von ganz unterschiedlicher Herkunft und den verschiedensten Organisationen zugehörig. Schon ein kurzes Durchsehen des Buchs zeigt den übergroßen Einfluss, den Google auf die KI-Gemeinschaft hat. Von den 23 Gesprächspartnern sind oder waren sieben bei Google oder dem Mutterkonzern Alphabet tätig. Weitere Schwerpunkte der KI-Forschung sind das MIT und Stanford. Geoff Hinton und Yoshua Bengio sind an der University of Toronto bzw. an der University of Montreal tätig, und die kanadische Regierung hat den guten Ruf ihrer Forschungsinstitute genutzt, um einen strategischen Schwerpunkt auf Deep Learning zu legen. 19 meiner 23 Gesprächspartner arbeiten in den USA. Von diesen 19 wurden allerdings mehr als die Hälfte nicht in den USA geboren. Sie stammen aus Australien, China, Ägypten, Frankreich, Israel, Rhodesien (jetzt Simbabwe), Rumänien und Großbritannien. Ich würde sagen, dass es sich hier um einen ziemlich eindeutigen Beweis dafür handelt, wie groß die Bedeutung der Einwanderung für die technologische Führung der USA ist.

Bei der Durchführung der Interviews hatte ich immer die Vielfalt potenzieller Leser vor Augen, vom professionellen Informatiker über Manager und Investoren bis hin zu praktisch jedem, der ein Interesse an KI und ihrer Auswirkung auf die Gesellschaft hat. Ein besonders wichtiger Teil der Leserschaft besteht aus jungen Menschen, die in Betracht ziehen, eine berufliche Laufbahn im Bereich der KI einzuschlagen. Es fehlt derzeit an qualifiziertem Nachwuchs, insbesondere was die Fähigkeiten bezüglich des Deep Learnings betrifft. Eine berufliche Laufbahn im Bereich KI oder Machine Learning dürfte spannend, lukrativ und bedeutend sein.

Da die Branche daran arbeitet, mehr qualifizierte Menschen für das Fachgebiet zu interessieren, wird zunehmend deutlich, dass mehr unternommen werden muss, um die Diversität dieser Menschen zu gewährleisten. Wenn KI unserer Welt tatsächlich ein neues Gesicht geben soll, dann ist es unverzichtbar, dass diejenigen, die sich am besten mit der Technologie auskennen – und somit auch am besten beeinflussen können, in welche Richtung sie sich bewegt –, die Gesellschaft als Ganzes repräsentieren.

Etwa ein Viertel der von mir interviewten Personen sind Frauen – und dieser Wert ist vermutlich schon höher als der Wert, den man im gesamten Fachgebiet der KI oder des Machine Learnings ermitteln würde. Eine kürzlich durchgeführte Untersuchung ergab, dass nur etwa 12% der führenden Forscher im Bereich Machine Learning Frauen sind (siehe *https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance*). Viele meiner Gesprächspartner betonten die Notwendigkeit, dass sowohl Frauen als auch Mitglieder von Minderheiten stärker repräsentiert werden.

Wie Sie dem Interview mit ihr entnehmen können, liegt einer der führenden im Bereich KI tätigen Frauen die Erhöhung der Diversität besonders am Herzen. Fei-Fei Li von der Stanford University ist Mitbegründerin einer Organisation, die jetzt den Namen AI4ALL trägt (*http://ai-4-all.org/*). Sie veranstaltet Ferienlager, die sich speziell an unterrepräsentierte Highschool-Schüler richten. AI4ALL hat von der Branche beträchtliche Unterstützung erhalten, wie beispielsweise von Google. Mittlerweile gibt es quer durch die USA sechs Universitäten, die Sommerlager veranstalten. Es ist zwar noch viel zu tun, aber es gibt gute Gründe, optimistisch zu sein, dass sich die Diversität der KI-Forscher in den kommenden Jahren und Jahrzehnten beträchtlich erhöhen wird.

Dieses Buch setzt zwar keine technischen Vorkenntnisse voraus, Sie werden jedoch einigen Konzepten und Begriffen des Fachgebiets begegnen. Wenn Sie vorher noch nie mit KI zu tun hatten, bietet sich eine gute Gelegenheit, direkt von den führenden Köpfen des Fachgebiets etwas über die Technologie zu lernen. Um den weniger erfahreneren Lesern den Einstieg zu erleichtern, folgt dieser Einleitung eine kurze Übersicht über das in der KI gebräuchliche Vokabular. Ich empfehle Ihnen, sich die Zeit zu nehmen, diesen Abschnitt zu lesen, bevor Sie mit der Lektüre der Interviews fortfahren. Darüber hinaus enthält das Interview mit Stuart Russell, der Koautor des führenden KI-Lehrbuchs *Artificial Intelligence: A Modern Approach* (deutscher Titel: *Künstliche Intelligenz. Ein moderner Ansatz*) ist, eine Erklärung vieler wichtiger Konzepte des Fachgebiets.

Es war mir eine besondere Ehre, die in diesem Buch wiedergegebenen Gespräche führen zu dürfen. Ich denke, Sie werden feststellen, dass alle meine Gesprächspartner sich wohlüberlegt und verständlich ausdrücken und sich sehr dafür engagieren, zu gewährleisten, dass die Technologie, an der sie arbeiten, zum Vorteil

der Menschheit eingesetzt wird. Eher selten werden Sie umfassenden Konsens vorfinden. Das Buch enthält eine Vielzahl von Erkenntnissen, Meinungen und Vorhersagen, die oftmals miteinander in Konflikt stehen. Das muss Ihnen klar sein: KI ist ein umfassendes, offenes und breit gefächertes Fachgebiet. Die Art kommender Innovationen, die Geschwindigkeit, mit der sie erreicht werden und ihre Anwendungsgebiete liegen noch völlig im Dunkeln. Diese Kombination aus potenziellen Umwälzungen und der grundsätzlichen Ungewissheit macht es unausweichlich, dass wir eine sinnvolle und umfassende Debatte über die Zukunft der KI führen und welchen Einfluss sie auf unser Leben hat. Ich kann mir nur wünschen, dass dieses Buch einen Beitrag zu dieser Debatte leisten wird.

Übersicht über das in der KI gebräuchliche Vokabular

Die Gespräche umfassen ein breites Themenspektrum und gehen manchmal auf bestimmte Verfahren ein, die in der KI verwendet werden. Für das Verständnis sind keine technischen Vorkenntnisse erforderlich, aber hin und wieder werden Sie auf gebräuchliche Fachbegriffe stoßen. Im Folgenden finden Sie eine kurze Übersicht der wichtigsten Begriffe, die Ihnen in den Interviews begegnen werden. Wenn Sie sich die Zeit nehmen, den folgenden Abschnitt zu lesen, sind Sie bestens für die Lektüre des Buchs gewappnet. Wenn Ihnen ein Abschnitt zu detailliert oder zu technisch erscheint, schlage ich vor, ihn einfach zu überspringen.

Machine Learning ist ein Teilgebiet der KI, das sich damit befasst, Algorithmen zu erstellen, die aus Daten lernen können. Man könnte auch sagen, dass Machine-Learning-Algorithmen Programme sind, die sich im Wesentlichen selbst programmieren, indem sie auf Informationen zurückgreifen. Es heißt oft, »Computer erledigen nur die Aufgaben, für die sie programmiert werden«, aber durch den Aufstieg des Machine Learnings trifft diese Aussage immer weniger zu. Es gibt viele verschiedene Arten von Machine-Learning-Algorithmen, aber Deep Learning hat zu den größten Umwälzungen geführt (und die größte Aufmerksamkeit der Medien auf sich gezogen).

Deep Learning ist ein Machine-Learning-Verfahren, das tiefe (aus vielen Schichten bestehende) **künstliche neuronale Netze** verwendet – also Software, die, vereinfacht gesagt, die Funktionsweise von Neuronen im Gehirn emuliert. Für die Revolution der KI, die wir im vergangenen Jahrzehnt erlebt haben, ist vor allem Deep Learning verantwortlich.

Es gibt noch einige andere Begriffe, die weniger technisch interessierte Leser einfach als »irgendetwas hinter den Kulissen des Deep Learnings« interpretieren können. Es steht Ihnen völlig frei, einen Blick hinter die Kulissen zu werfen und die Fachbegriffe genauer in Augenschein zu nehmen:

Backpropagation (oder kurz **Backprop**) ist der Lernalgorithmus, den Deep-Learning-Systeme verwenden. Beim Trainieren eines neuronalen Netzes (siehe nachfolgend »überwachtes Lernen«) breiten sich Informationen rückwärts durch die Neuronenschichten aus, die das Netz bilden und bewirken eine Neukalibrierung der Gewichte der einzelnen Neuronen. Das führt dazu, dass das Netz insgesamt der richtigen Lösung allmählich näher kommt. Geoff Hinton ist Koautor des wegweisenden Papers über Backpropagation aus dem Jahr 1986. In seinem Interview geht er ausführlicher auf Backpropagation ein.

Der Begriff **Gradientenabstiegsverfahren** ist noch undurchsichtiger. Er bezieht sich auf ein spezielles mathematisches Verfahren, das der Backpropagation-Algorithmus beim Training des Netzes verwendet, um den Fehler zu verringern.

Möglicherweise begegnen Ihnen auch Begriffe, die verschiedene Arten bezeichnen, wie etwa **rekurrente neuronale Netze (RNNs)** und **Convolutional Neural Networks (CNNs)** oder **Boltzmann-Maschinen**. Die Unterschiede betreffen im Allgemeinen die Art und Weise, wie die Neuronen miteinander verknüpft sind. Die Details sind sehr technisch und gehen über den Rahmen dieses Buchs hinaus. Ich habe Yann LeCun, der als Erfinder der CNNs gilt, die bei Anwendungen der Computer Vision häufig zum Einsatz kommen, dennoch gebeten, das Konzept kurz zu erklären.

Wenn ein Begriff den Namen **Bayes** enthält, geht es im Allgemeinen um Wahrscheinlichkeiten oder um die Anwendung der Regeln der Wahrscheinlichkeitsrechnung. Ihnen könnten beispielsweise Begriffe wie Bayes'sches Machine Learning oder Bayes-Netze begegnen, die Algorithmen bezeichnen, die auf den Regeln der Wahrscheinlichkeitsrechnung beruhen. Die Bezeichnungen gehen auf den Namen des Pfarrers Thomas Bayes (1701–1761) zurück, der eine Methode entwickelte, die Wahrscheinlichkeit eines Ereignisses anhand neu vorliegender Informationen vorherzusagen. Bayes'sche Verfahren erfreuen sich bei Informatikern und Forschern, die versuchen, die menschlichen kognitiven Fähigkeiten zu modellieren, großer Beliebtheit. Judea Pearl, der auch zu meinen Gesprächspartnern gehörte, erhielt die höchste Auszeichnung in der Informatik, den Turing Award, zum Teil für seine Arbeit über Bayes'sche Verfahren.

Wie KI-Systeme lernen

Es gibt verschiedene Möglichkeiten, Machine-Learning-Systeme zu trainieren. Innovationen auf diesem Gebiet, also neue Methoden zum Trainieren von KI-Systemen zu entwickeln, ist für zukünftige Fortschritte unverzichtbar.

Überwachtes Lernen bedeutet, einem Lernalgorithmus sorgfältig vorbereitete Trainingsdaten bereitzustellen, die kategorisiert oder gekennzeichnet sind. Sie könnten beispielsweise einem Deep-Learning-System beibringen, einen Hund auf

einem Foto zu erkennen, indem Sie es mit vielen Tausend (oder sogar Millionen) Hundefotos füttern. Diese Bilder wären als »Hund« gekennzeichnet. Darüber hinaus müssten Sie auch eine Menge Bilder bereitstellen, auf denen kein Hund zu sehen ist und diese mit »Kein Hund« kennzeichnen. Nachdem das System trainiert wurde, können Sie völlig neue Bilder als Eingabe verwenden, und es wird als Ausgabe »Hund« oder »Kein Hund« liefern. Dabei kann es eine Leistungsstärke erreichen, die diejenige eines typischen Menschen übertrifft.

Bei aktuellen KI-Systemen ist überwachtes Lernen das weitaus am häufigsten verwendete Verfahren. Es wird in rund 95% der Anwendungen eingesetzt, etwa bei der Übersetzung von Fremdsprachen (das Training erfolgt mit Millionen Dokumenten, die zweisprachig vorliegen) oder bei KI-gestützten Radiologiesystemen (das Training erfolgt mit Millionen medizinischer Bilder, die mit »Krebs« oder »Kein Krebs« gekennzeichnet sind). Beim überwachten Lernen gibt es das Problem, dass Unmengen gekennzeichneter Daten benötigt werden. Das erklärt, weshalb Unternehmen wie Google, Amazon oder Facebook, die über gigantische Datenmengen verfügen, bei der Deep-Learning-Technologie eine so dominierende Stellung einnehmen.

Reinforcement Learning (dt. auch **bestärkendes oder verstärkendes Lernen**) bedeutet im Wesentlichen, durch Üben zu lernen oder durch die Trial-and-Error-Methode. Anstatt einen Algorithmus durch die Bereitstellung korrekt gekennzeichneter Daten zu trainieren, überlässt man es dem System, selbst eine Lösung zu finden, und wenn es erfolgreich ist, gibt es eine »Belohnung«. Stellen Sie sich vor, dass Sie Ihrem Hund beibringen wollen, auf das Kommando »Sitz!« zu hören. Wenn er gehorcht, gibt es ein Leckerli. Reinforcement Learning hat sich bei der Entwicklung von KI-System, die Spiele spielen, als besonders leistungsstark erwiesen. Aus dem Interview mit Demis Hassabis werden Sie erfahren, dass DeepMind ein klarer Befürworter des Reinforcement Learnings ist und es bei der Entwicklung des AlphaGo-Systems einsetzte.

Beim Reinforcement Learning gibt es das Problem, dass der Algorithmus sehr viele Übungsdurchgänge benötigt, bis er erfolgreich ist. Aus diesem Grund wird er vornehmlich für Spiele verwendet oder für Aufgaben, die auf einem Computer sehr schnell simuliert werden können. Auch bei der Entwicklung selbstfahrender Autos kann Reinforcement Learning eingesetzt werden, aber natürlich nicht, indem echte Autos auf der Straße üben, sondern durch das Trainieren virtueller Autos in simulierten Umgebungen. Nachdem die Software trainiert wurde, kann sie in echten Autos eingesetzt werden.

Unüberwachtes Lernen bedeutet, dass Maschinen direkt aus unstrukturierten Daten lernen, die ihrer Umgebung entstammen. Auf diese Weise lernen Menschen. Kleine Kinder lernen beispielsweise das Sprechen hauptsächlich dadurch, dass sie ihren Eltern zuhören. Überwachtes Lernen und Reinforcement Learning

spielen auch eine Rolle, aber das menschliche Gehirn verfügt über die erstaunliche Fähigkeit, allein durch Beobachtung und unüberwachte Interaktion mit der Umgebung zu lernen.

Unüberwachtes Lernen stellt eines der vielversprechendsten Verfahren dar, um in der KI Fortschritte zu erzielen. So sind etwa Systeme denkbar, die selbstständig lernen, ohne dass große Mengen gekennzeichneter Trainingsdaten erforderlich sind. Gleichzeitig ist unüberwachtes Lernen allerdings auch eine der schwierigsten Herausforderungen. Ein Durchbruch, der es ermöglichen würde, dass Maschinen tatsächlich unüberwacht effizient lernen, wäre eines der bedeutendsten Ereignisse in der Geschichte der KI und ein Meilenstein auf dem Weg zur KI auf menschlichem Niveau.

Artificial General Intelligence (AGI) bezeichnet eine denkende Maschine. AGI wird im Allgemeinen mehr oder weniger als Synonym für die Begriffe **KI auf menschlichem Niveau, generelle/allgemeine (Künstliche) Intelligenz** oder **starke KI** betrachtet. Sie kennen sicherlich verschiedene Beispiele für eine AGI – die aber alle der Science Fiction entstammen: HAL aus dem Film *2001: A Space Oddyssey* (*2001: Odyssee im Weltraum*), der Hauptcomputer der Enterprise (oder Mr. Data) aus *Star Trek*, C3PO aus *Star Wars* (*Krieg der Sterne*) oder Agent Smith aus *Matrix*. Diese erdachten Systeme wären in der Lage, den **Turing-Test** zu bestehen – sie könnten also ein Gespräch führen, das sich nicht von einem Gespräch mit einem Menschen unterscheiden lässt. Alan Turing schlug diesen Test in seinem Papier *Computing Machinery and Intelligence* aus dem Jahr 1950 vor, das wohl KI als Forschungsgebiet begründet hat. Mit anderen Worten: AGI war von Anfang an das Ziel.

Wenn wir es eines Tages geschafft haben, eine AGI zu erreichen, wird das smarte System wahrscheinlich schnell noch smarter werden. Wir erleben dann das Entstehen einer **Superintelligenz**, einer Maschine, deren allgemeinen intellektuellen Fähigkeiten diejenigen irgendeines Menschen übertreffen. Das könnte sich einfach durch leistungsfähigere Hardware ergeben, das Ganze würde sich aber deutlich beschleunigen, wenn eine intelligente Maschine sich darauf konzentriert, noch smartere Versionen von sich selbst zu entwickeln. Das könnte zu dem führen, was als »rekursiver Verbesserungszyklus« oder »schneller Intelligenz-Takeoff« bezeichnet wurde. Dieses Szenario hat zu den Bedenken bezüglich des Außer-Kontrolle-Geratens und zum KI-Ausrichtungsproblem geführt – dass ein superintelligentes System auf eine Weise handelt, die nicht im Interesse der Menschheit liegt.

Ich bin zu dem Schluss gekommen, dass der Weg zur AGI und die Aussicht auf eine Superintelligenz so interessante Themen sind, dass ich sie mit allen Gesprächspartnern diskutiert habe.

Martin Ford *ist Zukunftsforscher und Autor zweier Bücher, dem Bestseller der New York Times* Rise of the Robots: Technology and the Threat of a Jobless Future *(Gewinner des Financial Times/McKinsey Business Book of the Year Award 2015, übersetzt in mehr als 20 Sprachen, deutscher Titel:* Aufstieg der Roboter*) und* The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future. *Er ist außerdem Gründer einer im Silicon Valley ansässigen Softwareentwicklungsfirma. Seinen TED-Talk über die Auswirkungen von KI und Robotik auf Wirtschaft und Gesellschaft haben sich mehr als 2 Millionen Zuschauer angesehen.*

Ford ist beratender KI-Experte des neuen »Rise of the Robots Index« der Société Générale, die zur Lyxor Robotics & AI ETF gehört, die in Unternehmen investiert, die bei der KI- und Robotik-Revolution eine bedeutende Rolle spielen. Er besitzt einen Abschluss in Informatik der University of Michigan, Ann Arbour, und einen betriebswirtschaftlichen Studienabschluss der University of California, Los Angeles.

Er befasst sich mit zukünftigen Technologien und ihren Auswirkungen und hat schon für The New York Times, Fortune, Forbes, The Atlantic, The Washington Post, Harvard Business Review, The Guardian *und* The Financial Times *geschrieben. Er ist in zahllosen Radio- und Fernsehsendungen zu Gast gewesen, unter anderem bei NPR, CNBC, CNN, MSNBC und PBS. Ford hält regelmäßig Vorträge über die zunehmenden Fortschritte der Robotik und der KI und was diese Fortschritte für die Wirtschaft, den Arbeitsmarkt und die Gesellschaft der Zukunft bedeuten.*

Ford ist weiter als Unternehmer tätig und engagiert sich aktiv als Vorstandsmitglied von Genesis Systems, einem Start-up, das eine revolutionäre Technologie zur Gewinnung von Wasser aus der Atmosphäre (Atmospheric Water Generator, AWG) entwickelt hat. Genesis beabsichtigt, schon bald in den trockensten Regionen der Erde automatisierte Systeme mit eigener Energieversorgung einzurichten, die in industriellem Maßstab Wasser aus der Luft gewinnen.

> ***Ein moralisches Bewusstsein oder eine Moralvorstellung dafür, was richtig und was falsch ist, besitzt die gegenwärtige – und die in absehbarer Zeit verfügbare – KI nicht und wird so etwas auch nie besitzen.***

YOSHUA BENGIO

WISSENSCHAFTLICHER LEITER AM MONTREAL INSTITUTE FOR LEARNING ALGORITHMS UND PROFESSOR FÜR INFORMATIK UND UNTERNEHMENSFORSCHUNG AN DER UNIVERSITY OF MONTREAL

Yoshua Bengio ist Professor für Informatik und Unternehmensforschung an der University of Montreal und als einer der Pioniere auf dem Forschungsgebiet Deep Learning bekannt. Bengio hat maßgeblich zu den Fortschritten bei der Erforschung neuronaler Netze beigetragen Hier ist insbesondere das unüberwachte Lernen zu nennen, bei dem neuronale Netze lernen können, ohne dass große Trainingsdatenmengen erforderlich sind.

MARTIN FORD: Sie sind an vorderster Front in der KI-Forschung, deshalb möchte ich zunächst fragen, bei welchen derzeitigen Forschungsaufgaben Ihrer Ansicht nach in den nächsten Jahren Durchbrüche zu erwarten sind und wie sie uns dabei helfen können, das Ziel einer AGI zu erreichen.

YOSHUA BENGIO: Ich weiß nicht genau, was zu erwarten ist, aber ich kann Ihnen sagen, dass wir einige wirklich schwierige Probleme vor uns haben und von einer KI auf menschlichem Niveau noch weit entfernt sind. Die Forscher versuchen zu verstehen, was eigentlich Probleme bereitet, wie etwa: Warum können wir keine Maschinen bauen, die die Welt genauso gut verstehen wie wir? Liegt es nur daran, dass wir nicht genug Trainingsdaten haben, oder liegt es an mangelnder Rechenleistung? Viele von uns denken, dass uns noch grundlegende Bestandteile fehlen, etwa die Fähigkeit, kausale Beziehungen in den Daten zu erkennen – eine Fähigkeit, die es uns ermöglicht, zu generalisieren und die richtigen Antworten zu finden unter Bedingungen, die sich sehr von denen unterscheiden, unter denen wir geübt haben.

Ein Mensch kann sich vorstellen, eine Erfahrung zu machen, die für ihn völlig neu ist. Sie waren vielleicht noch nie in einen Autounfall verwickelt, aber Sie können sich das vorstellen, und aufgrund all der Dinge, die Sie bereits darüber wissen, können Sie diese Rolle einnehmen und die richtigen Entscheidungen treffen – zumindest in Gedanken. Das gegenwärtige Machine Learning beruht auf überwachtem Lernen, bei dem ein Computer im Wesentlichen aus den statistischen Daten lernt, die ihm bereitgestellt werden. Und dieser Vorgang muss von Hand veranlasst werden. Mit anderen Worten: Menschen müssen alle Kennzeichnungen vorgegeben, möglicherweise hunderte Millionen richtiger Antworten, aus denen der Computer lernt.

Ein Großteil der aktuellen Forschung befasst sich mit Gebieten, auf denen wir nur wenige Fortschritte erzielt haben, wie etwa beim unüberwachten Lernen. Hier kann der Computer autonomer vorgehen, indem er Wissen über die reale Welt sammelt. Ein weiteres Forschungsgebiet betrifft die Kausalität. Der Computer kann die Daten, wie Bilder oder Videos, nicht nur beobachten, sondern selbst handeln und die Auswirkungen erfassen, um Rückschlüsse auf kausale Beziehungen in der realen Welt zu ziehen. Das, was beispielsweise DeepMind, OpenAI oder Berkeley mit virtuellen Agenten anstellen, weist in die richtige Richtung, um solche Fragestellungen beantworten zu können, und wir machen so etwas in Montreal ebenfalls.

MARTIN FORD: Gibt es bestimmte Projekte, auf die Sie verweisen würden, die derzeit beim Deep Learning ganz im Mittelpunkt stehen? AlphaZero liegt auf der Hand, aber welche anderen Projekte gehören noch zur Spitzentechnologie?

YOSHUA BENGIO: Es gibt eine Reihe interessanter Projekte, aber diejenigen, von denen ich glaube, dass sie langfristig große Auswirkungen haben, sind solche, in denen ein Agent in einer virtuellen Welt versucht, Aufgaben zu erledigen und seine Umgebung zu erkunden. Wir arbeiten bei MILA daran, und bei DeepMind, OpenAI, Berkeley, Facebook und Google Brain sind Projekte in Arbeit. Das ist jetzt die neue Pioniergrenze.

Man darf aber nicht vergessen, dass es sich hier nicht um kurzfristige Forschung handelt. Wir arbeiten nicht an einer bestimmten Anwendung des Deep Learnings, sondern untersuchen, wie ein lernender Agent zukünftig seine Umgebung erkennen und wie er sprechen lernen oder eine Sprache verstehen kann, also insbesondere das, was wir als »Grounded Language Learning« (Realitäts-basierter Spracherwerb) bezeichnen.

MARTIN FORD: Können Sie diesen Begriff erläutern?

YOSHUA BENGIO: Aber gern. Bei den früheren Versuchen, Computer dazu zu bringen, Sprache zu verstehen, ließ man den Computer einfach jede Menge Text lesen. Das ist auch schön und gut, aber für den Computer ist es tatsächlich schwierig, die Bedeutung der Wörter zu erfassen, wenn den Sätzen nicht auch reale Objekte zugeordnet sind. Sie können Wörter beispielsweise mit Bildern oder Videos verknüpfen oder im Fall von Robotern mit Objekten in der realen Welt.

Im Bereich des Grounded Language Learnings gibt es viele Forschungsbemühungen, um ein Sprachverständnis zu entwickeln – auch wenn es nur eine kleine Teilmenge des gesamten Wortschatzes betrifft – sodass ein Computer tatsächlich versteht, was die Wörter bedeuten und darauf basierend handeln kann. Er kann dann entsprechend dieser Wörter handeln. Das ist eine sehr interessante Entwicklung, die praktische Auswirkungen auf das Sprachverständnis von Dialogsystemen, persönlichen Assistenten und dergleichen hat.

MARTIN FORD: Die Idee besteht also im Grunde genommen darin, einen Agenten in einer simulierten Umgebung auszusetzen und ihn wie ein Kind lernen zu lassen?

YOSHUA BENGIO: Genau, tatsächlich wollen wir uns von Experten für Kindesentwicklung inspirieren lassen, die untersuchen, wie ein Neugeborenes in den ersten Lebensmonaten eine Reihe von Entwicklungsstufen durchläuft und dabei allmählich zu einem besseren Verständnis der Welt gelangt. Wir haben noch nicht richtig begriffen, welche Teile dieses Verständnisses angeboren sind und welche tatsächlich erlernt werden, und ich denke, dass das Wissen, welche Phasen Babys durchlaufen, bei der Entwicklung unserer Systeme hilfreich sein kann.

Ich habe vor einigen Jahren beim Machine Learning ein Konzept eingeführt, das *Curriculum Learning* (Lernen nach Lehrplan), das beim Training von Tieren häufig eingesetzt wird. Die Idee besteht darin, dass wir nicht nur alle Trainingsbeispiele als einen großen Haufen in willkürlicher Reihenfolge bereitstellen. Stattdessen gehen wir die Beispiele in einer Reihenfolge durch, die für den Lernenden sinnvoll ist. Wir fangen mit einfachen Dingen an, und wenn diese beherrscht werden, können wir sie wie Bausteine verwenden, um etwas kompliziertere Dinge zu erlernen. Deshalb gehen wir im Alter von sechs Jahren zunächst einmal zur Schule und besuchen nicht gleich die Universität. Diese Art des Lernens gewinnt auch beim Training von Computern zunehmend an Bedeutung.

MARTIN FORD: Ich möchte auf den Weg zur AGI zu sprechen kommen. Sie denken offenbar, dass unüberwachtes Lernen, also dass ein System wie ein Mensch lernt, ein wichtiger Bestandteil davon ist. Reicht das für eine AGI aus, oder gibt es weitere unverzichtbare Bestandteile oder noch fehlende Durchbrüche, um so weit zu kommen?

YOSHUA BENGIO: Mein Freund Yann LeCun verwendet eine schöne Metapher, um das zu beschreiben. Wir besteigen gerade einen Hügel und sind alle ganz aufgeregt, weil wir dabei viele Fortschritte erzielt haben. Aber während wir uns der Spitze des Hügels nähern, wird erkennbar, dass sich eine Reihe weiterer Hügel vor uns erhebt. Wir erkennen gerade, dass unsere Ansätze bei der Entwicklung einer AGI gewissen Einschränkungen unterliegen. Als wir beispielsweise beim Besteigen des ersten Hügels entdeckten, wie wir tiefere Netze trainieren können, waren uns die Grenzen der von uns entwickelten Systeme nicht klar, weil wir lediglich herausgefunden hatten, wie wir ein paar Schritte weiter aufsteigen.

Wenn wir mit unseren Verfahren zufriedenstellende Verbesserungen erzielen, also die Spitze des ersten Hügels erreichen, erkennen wir auch die Beschränkungen und können den nächsten Hügel sehen, den wir besteigen müssen. Und wenn wir diesen Hügel bestiegen haben, sehen wir einen weiteren Hügel und so weiter. Es ist unmöglich zu sagen, wie viele Durchbrüche oder bedeutsame Fortschritte noch nötig sein werden, um KI auf menschlichem Niveau zu erreichen.

MARTIN FORD: Wie viele Hügel gibt es? Wie sieht der Zeitrahmen für eine AGI aus? Wie schätzen Sie das ein?

YOSHUA BENGIO: Ich werde das nicht einschätzen, denn es wäre sinnlos, ein Datum anzugeben, weil wir keine Ahnung haben. Ich kann nur sagen, dass es nicht in den nächsten paar Jahren der Fall sein wird.

MARTIN FORD: Glauben Sie, dass Deep Learning und neuronale Netze grundsätzlich der richtige Weg sind?

YOSHUA BENGIO: Ja, die dem Deep Learning zugrunde liegenden wissenschaftlichen Konzepte, die wir entdeckt haben, und die Fortschritte, die wir in all den Jahren auf diesem Gebiet gemacht haben, bedeuten, dass viele der dem Deep Learning und den neuronalen Netzen zugrunde liegenden Konzepte fortbestehen werden. Vereinfacht gesagt: Sie sind unglaublich leistungsfähig. Tatsächlich werden sie uns vermutlich dabei helfen, besser zu verstehen, wie die Gehirne von Tieren und Menschen komplexe Dinge erlernen. Aber sie reichen wie gesagt nicht aus, um eine AGI zu erreichen. Wir befinden uns an einem Punkt, an dem wir einige Beschränkungen unserer derzeitigen Möglichkeiten erkennen, und wir werden sie verbessern und darauf aufbauen.

MARTIN FORD: Das Allen Institute for AI arbeitet am Projekt Mosaic, bei dem es darum geht, Computern eine Art »gesunden Menschenverstand« zu verleihen. Halten Sie das für wichtig, oder glauben Sie, dass sich ein gesunder Menschenverstand während des Lernvorgangs entwickelt?

YOSHUA BENGIO: Ich bin mir sicher, dass beim Lernvorgang ein gesunder Menschenverstand entsteht. Er wird jedenfalls nicht dadurch entstehen, dass man kleine Wissenshäppchen im Kopf deponiert. So funktioniert das beim Menschen nicht.

MARTIN FORD: Ist vor allem Deep Learning wichtig, um eine AGI zu erreichen, oder denken Sie, dass eine Art Hybridsystem erforderlich ist?

YOSHUA BENGIO: Klassische KI war rein symbolisch, und es gab kein Lernen. Sie konzentrierte sich auf einen wirklich interessanten Aspekt der kognitiven Fähigkeiten, nämlich wie wir Informationen der Reihe nach erfassen und miteinander kombinieren. Andererseits lag bei den neuronalen Netzen des Deep Learnings der Fokus schon immer auf einer Art Bottom-up-Vorstellung der kognitiven Fähigkeiten. Wir fangen mit der sinnlichen Wahrnehmung an und verankern darin das Verständnis der Welt, das die Maschine entwickelt. Davon ausgehend erzeugen wir verteilte Repräsentationen und können die Beziehungen zwischen vielen Variablen erfassen.

Um 1999 habe ich zusammen mit meinem Bruder die Beziehungen zwischen solchen Variablen untersucht. Das war der Ausgangspunkt für viele der kürzlich erzielten Fortschritte bei der Verarbeitung natürlicher Sprache, wie etwa Worteinbettungen oder verteilte Repräsentationen für Wörter und Sätze. Dabei wird ein Wort durch ein Aktivitätsmuster im Gehirn repräsentiert – oder durch eine Zahlenmenge. Wörtern mit ähnlicher Bedeutung wird dann eine ähnliche Zahlenmenge zugeordnet. Im Forschungsgebiet Deep Learning geschieht gerade Folgendes: Die Forscher bauen auf diesen Deep-Learning-Konzepten auf und versuchen, damit die

klassischen KI-Probleme, nämlich schlussfolgern, verstehen und planen, zu lösen. Sie versuchen, die von uns anhand der Wahrnehmung entwickelten Bausteine zu verwenden und sie um diese allgemeineren Aufgaben der Wahrnehmung zu erweitern (Psychologen bezeichnen das mitunter als »System 2«). Ich glaube, dass wir zumindest teilweise auf diesem Weg der KI auf menschlichem Niveau näher kommen können. Es ist kein Hybridsystem; wir versuchen dieselben Probleme wie die klassische KI zu lösen, bedienen uns dabei aber der Bausteine, die das Deep Learning liefert. Die Methoden unterscheiden sich sehr, aber die Ziele sind sehr ähnlich.

MARTIN FORD: Ihre Vorhersage ist also, dass neuronale Netze zum Einsatz kommen, lediglich mit anderen Architekturen?

YOSHUA BENGIO: Ja. Ihr Gehirn ist auch ein neuronales Netz. Wir benötigen andere Architekturen und andere Trainingsbedingungen, die das leisten können, was die klassische KI zu leisten versuchte, wie schlussfolgern, Erklärungen für Beobachtungen herleiten und Planung.

MARTIN FORD: Glauben Sie, dass diese Aufgaben mit Lernen und Training lösbar sind, oder ist hier irgendeine Struktur erforderlich?

YOSHUA BENGIO: Es gibt eine Struktur, nur nicht die Art Struktur, die wir beim Schreiben einer Enzyklopädie oder einer mathematischen Formel zur Repräsentation von Wissen verwenden. Die verwendete Struktur entspricht der Architektur des neuronalen Netzes sowie den ziemlich umfassenden Annahmen über die Welt und der Art der Aufgaben, die wir zu lösen versuchen. Wenn wir eine spezielle Struktur und Architektur einsetzen, die dem Netz einen Aufmerksamkeitsmechanismus bietet, wird ihm schon eine ganze Menge Vorwissen bereitgestellt. Wie sich gezeigt hat, ist das für bestimmte Aufgaben, wie etwa maschinelle Übersetzungen, von großer Bedeutung.

Man braucht ein solches Werkzeug, um einige der Probleme zu lösen, so wie man bei der Verarbeitung von Bildern eine CNN-Struktur benötigt, um eine Aufgabe gut zu erledigen. Wenn diese Struktur nicht vorgegeben wird, ist die Leistung viel schlechter. In den beim Deep Learning verwendeten Architekturen und Trainingszielen stecken implizit bereits viele bereichsspezifische Annahmen über die Welt und die Funktionen, die erlernt werden sollen. Darum geht es in den meisten Arbeiten, die derzeit veröffentlicht werden.

MARTIN FORD: Was ich mit der Frage nach der Struktur zu erreichen versuchte, ist, dass beispielsweise ein Baby sofort nach der Geburt menschliche Gesichter erkennen kann. Dann muss es doch im menschlichen Gehirn eindeutig irgend-

eine Struktur geben, die einem Baby das ermöglicht. Da gibt es ja keine Eingabeneuronen, die Pixel verarbeiten.

YOSHUA BENGIO: Da liegen Sie falsch! Es gibt tatsächlich Eingabeneuronen, die Pixel verarbeiten, allerdings gibt es im Gehirn des Babys eine bestimmte Architektur, die etwas Rundes mit zwei Punkten darin erkennt.

MARTIN FORD: Mir kommt es darauf an, dass die Struktur bereits vorhanden ist.

YOSHUA BENGIO: Das ist sie natürlich, aber all die Dinge, die wir bei der Entwicklung neuronaler Netze vorgeben, sind ja auch schon vorhanden. Das Vorgehen der Deep-Learning-Forscher entspricht der Funktionsweise der Evolution. Wir geben Wissen in Form der Architektur und des Trainingsverfahrens vor.

Wenn wir wollten, könnten wir etwas fest vorgeben, das es dem Netz ermöglicht, ein Gesicht zu erkennen, aber für eine KI ist das nutzlos, weil sie das sehr schnell erlernen kann. Stattdessen geben wir Dinge vor, die wirklich nützlich sind, um schwierigere Aufgaben zu lösen, mit denen wir versuchen, umzugehen.

Niemand behauptet, dass Menschen, Babys und Tiere kein angeborenes Wissen besitzen. Tatsächlich besitzen die meisten Tiere nur angeborenes Wissen. Eine Ameise lernt nicht viel. Ihr Verhalten ist wie ein großes, festgelegtes Programm. Aber wenn man in der Hierarchie der Intelligenz weiter nach oben blickt, nimmt der Anteil des Lernens immer weiter zu. Wir Menschen unterscheiden uns von vielen anderen Tieren dadurch, wie viel wir lernen und wie viel Wissen angeboren ist.

MARTIN FORD: Holen wir etwas weiter aus und definieren einige dieser Konzepte. In den 1980er-Jahren waren neuronale Netze eine Randerscheinung. Sie besaßen nur eine Schicht, von Tiefe konnte also keine Rede sein. Sie waren daran beteiligt, daraus das zu machen, was wir heute als Deep Learning bezeichnen. Könnten Sie, ohne technisch zu sehr ins Detail zu gehen, definieren, was Deep Learning eigentlich ist?

YOSHUA BENGIO: Deep Learning ist ein Ansatz des Machine Learnings. Beim Machine Learning versucht man, Computern Wissen zu vermitteln, indem man ihnen ermöglicht, anhand von Beispielen zu lernen. Beim Deep Learning geschieht das auf eine Weise, die durch das Gehirn inspiriert wurde.

Deep Learning und Machine Learning sind lediglich eine Fortsetzung der früheren Arbeiten über neuronale Netze. Man spricht von »tief«, weil es jetzt die Möglichkeit gibt, »tiefere« Netze zu trainieren, also Netze mit mehr Schichten, wobei jede Schicht eine andere Repräsentationsebene darstellt. Wir hoffen, dass tiefere

Netze abstraktere Dinge repräsentieren können, und bislang scheint das auch der Fall zu sein.

MARTIN FORD: Wenn Sie von Schichten sprechen, meinen Sie damit Schichten der Abstraktion? Dass also bei einem Bild die erste Schicht Pixel wären, die zweite Kanten, gefolgt von Ecken, bis dann allmählich ganze Objekte erreicht werden?

YOSHUA BENGIO: Ja, das stimmt.

MARTIN FORD: Wenn ich das richtig verstehe, hat der Computer aber noch keine Vorstellung davon, was das Objekt ist, oder?

YOSHUA BENGIO: Der Computer hat eine gewisse Vorstellung, das ist keine Entweder-oder-Frage. Eine Katze versteht, wie eine Tür funktioniert, aber nicht so gut wie Sie. Verschiedene Leute verstehen die vielen Dinge, von denen sie umgeben sind, auf ganz unterschiedlichen Ebenen, und in der Wissenschaft geht es darum, unser Verständnis dieser vielen Dinge zu vertiefen. Diese Netze besitzen eine gewisse Vorstellung von Bildern, wenn sie mit Bildern trainiert wurden, die aber nicht so abstrakt und allgemein ist wie die unsrige. Das liegt unter anderem daran, dass wir Bilder im Kontext unseres dreidimensionalen Verständnisses der Welt interpretieren, das wir dank räumlichen Sehens und unserer Bewegungen und Handlungen in der Welt erlangt haben. Auf diese Weise erwerben wir eine Vorstellung, die viel mehr als lediglich ein visuelles Modell umfasst, nämlich ein physisches Modell von Objekten. Die Vorstellung, die ein Computer derzeit von Bildern hat, ist noch immer primitiv, aber dennoch gut genug, um sich bei vielen Anwendungen als äußerst nützlich zu erweisen.

MARTIN FORD: Stimmt es, dass Deep Learning erst durch Backpropagation möglich wurde? Also der Idee, die Fehler-Information an die Schichten zurückzugeben und sie anhand des Endergebnisses anzupassen?

YOSHUA BENGIO: Backpropagation ist tatsächlich mitverantwortlich für den Erfolg des Deep Learnings in den letzten Jahren. Es handelt sich um eine Methode zur Zuweisung von Gewichten, das heißt, herauszufinden, wie die internen Neuronen geändert werden müssen, damit sich das Netz als Ganzes ordentlich verhält. Backpropagation wurde, zumindest im Kontext neuronaler Netze, Anfang der 1980er-Jahre entdeckt, zu dem Zeitpunkt, als ich mit meiner eigenen Arbeit anfing. Yann LeCun entdeckte es unabhängig etwa zum gleichen Zeitpunkt wie Geoffrey Hinton und David Rumelhart. Die Idee ist schon älter, aber wir konnten diese tieferen Netze in der Praxis bis etwa 2006, mehr als ein Vierteljahrhundert später, nicht erfolgreich trainieren.

INDEX

C

D

E

L

M

N

O

P

R

S

T

U