

RAY KURZWEIL

DIE NÄCHSTE STUFE DER EVOLUTION

WENN MENSCH UND MASCHINE
EINS WERDEN



»Ich kenne
niemanden, der
besser geeignet ist,
die Zukunft von KI
vorherzusagen,
als Ray Kurzweil.«
Bill Gates

PIPER

Ray Kurzweil

Die nächste Stufe der Evolution

RAY KURZWEIL

DIE NÄCHSTE STUFE DER EVOLUTION

WENN MENSCH UND
MASCHINE EINS WERDEN

Aus dem amerikanischen Englisch von
Sigrid Schmid, Moritz Langer, Jens Hagededt

Mit 64 Schwarz-Weiß-Abbildungen

PIPER

Mehr über unsere Autorinnen, Autoren und Bücher:
www.piper.de

Inhalte fremder Webseiten, auf die in diesem Buch (etwa durch Links) hingewiesen wird, macht sich der Verlag nicht zu eigen. Eine Haftung dafür übernimmt der Verlag nicht. Wir behalten uns eine Nutzung des Werks für Text und Data Mining im Sinne von § 44b UrhG vor.

ISBN 978-3-492-07306-6

Die Originalausgabe erschien 2024

unter dem Titel *The Singularity is Nearer*

bei VIKING, einem Imprint von Penguin Random House LLC,
New York, penguinrandomhouse.com.

Alle Rechte vorbehalten, einschließlich des Rechts auf
vollständige oder teilweise Vervielfältigung in jeder Form.

Diese Ausgabe wurde in Absprache mit VIKING veröffentlicht.

Copyright © 2024 by Ray Kurzweil

Für die deutsche Ausgabe:

© Piper Verlag GmbH, München 2024

Bildnachweis: Abbildung 51, S. 206: © Valcenteu, 2010;

Abbildung 52, S. 209: © Simon Stone

Herstellung und Satz: Sieveking Agentur, München

Gesetzt aus der Minion

Litho: Lorenz & Zeller, Inning am Ammersee

Druck und Bindung: GGP Media GmbH, Pößneck

Printed in Germany

*Für Sonya Rosenwald Kurzweil.
Vor wenigen Tagen war es 50 Jahre her,
dass ich sie kennen-
(und lieben) gelernt habe!*

INHALT

	EINLEITUNG	9
KAPITEL 1:	WO STEHEN WIR IN DEN SECHS STADIEN?	15
KAPITEL 2:	DIE NEUERFINDUNG DER INTELLIGENZ	19
KAPITEL 3:	WER BIN ICH?	86
KAPITEL 4:	UNSER LEBEN WIRD EXPONENTIELL BESSER	124
KAPITEL 5:	DIE ZUKUNFT DER ARBEITSPLÄTZE: GUT ODER SCHLECHT?	222
KAPITEL 6:	GESUNDHEIT UND WOHLBEFINDEN: DIE NÄCHSTEN 30 JAHRE	269
KAPITEL 7:	GEFAHR	305
KAPITEL 8:	ZWIEGESPRÄCH MIT KASSANDRA	327
	DANK	333
	ANHANG	335
	ANMERKUNGEN	362
	REGISTER	482

EINLEITUNG

In meinem Buch *The Singularity is Near* aus dem Jahr 2005* stellte ich die These auf, dass konvergente, exponentielle technologische Trends zu einer umwälzenden Transformation der Menschheit führen werden. Mehrere wichtige Bereiche verändern sich gleichzeitig immer schneller: Rechenleistung wird billiger, wir wissen immer mehr über die menschliche Biologie, und Technik wird in immer kleinerem Maßstab möglich. Künstliche Intelligenz (KI) wird immer leistungsfähiger, und es wird immer einfacher, auf Informationen zuzugreifen. Im selben Maß integrieren wir diese neuen Möglichkeiten immer mehr in unsere natürliche biologische Intelligenz. Irgendwann werden wir dank Nanotechnologie unser Gehirn mit Schichten aus virtuellen Neuronen in der Cloud erweitern können. So werden wir mit der KI verschmelzen und unsere biologische Ausstattung um die millionenfache Rechenleistung aufstocken können. Dadurch werden sich unsere Intelligenz und unser Bewusstsein so grundlegend erweitern, wie wir es heute kaum zu begreifen vermögen. Dieses Ereignis meine ich, wenn ich von Singularität spreche.

Der Begriff »Singularität« ist aus der Mathematik entlehnt (wo er einen undefinierten Punkt in einer Funktion bezeichnet, wenn man zum Beispiel etwas durch null teilt) sowie aus der Physik (wo er einen Punkt mit unendlicher Dichte im Zentrum eines schwarzen Lochs

* Die deutsche Ausgabe erschien 2014 unter dem Titel *Menschheit 2.0. Die Singularität naht*. Im Folgenden wird der deutsche Haupttitel verwendet. (Anm. d. Übers.)

bezeichnet, an dem die normalen physikalischen Gesetze nicht mehr gelten). Ich verwende den Begriff allerdings metaphorisch. Meine Prognose der technologischen Singularität besagt nicht, dass sich Veränderungen irgendwann unendlich schnell vollziehen werden. Exponentielles Wachstum impliziert kein unendliches Wachstum, und ebenso wenig tut das eine physikalische Singularität. In einem schwarzen Loch ist die Gravitationskraft so groß, dass nicht einmal Licht entkommt, aber nichts in der Quantenmechanik weist darauf hin, dass Masse unendlich werden könnte. Ich verwende die Metapher der Singularität, weil sie unsere Unfähigkeit erfasst, eine derart radikale Veränderung unseres derzeitigen Intelligenzniveaus zu verstehen. Aber wir werden unsere Denkfähigkeit schnell genug erweitern, dass wir uns anpassen können, wenn diese Veränderung tatsächlich eintritt.

Wie ich in *Menschheit 2.0* ausgeführt habe, deuten langfristige Trends darauf hin, dass die Singularität um das Jahr 2045 eintreten wird. Bei Veröffentlichung jenes Buches lag dieses Datum 40 Jahre – zwei ganze Generationen – in der Zukunft. Auf diese Distanz konnte ich die Kräfte, die diese Transformation herbeiführen, nur grob skizzieren, aber für die meisten Leserinnen und Leser war das Thema recht weit von der Alltagsrealität des Jahres 2005 entfernt. Und viele Kritiker schrieben, mein Zeitrahmen sei überaus optimistisch, oder sogar, die Singularität sei völlig unmöglich.

Seit damals ist jedoch etwas Bemerkenswertes geschehen. Der Fortschritt hat sich allen Zweiflern widersetzt und weiter beschleunigt. Soziale Medien und Smartphones entwickelten sich von nicht existent zu Alltagsbegleitern, die heute den Großteil der Weltbevölkerung verbinden. Algorithmische Innovationen und die Verbreitung von Big Data ermöglichten KI erstaunliche Durchbrüche früher, als sogar Experten erwartet hatten – zunächst meisterte sie Spiele wie *Jeopardy!* und *Go*, dann fuhr sie Autos, schrieb Essays, bestand Juraexamen und diagnostizierte Krebs. Inzwischen können leistungsfähige und flexible große Sprachmodelle (Large Language Models, LLMs) wie GPT-4 und Gemini natürlichsprachige Anweisungen in Computercode übersetzen – was die Hürde zwischen Mensch und Maschine erheblich senkt. Wenn dieses Buch

erscheint, werden wahrscheinlich Millionen Menschen schon selbst Erfahrungen mit diesen Fähigkeiten gemacht haben. Unterdessen sind die Kosten für die Sequenzierung eines menschlichen Genoms um etwa 99,997 Prozent gesunken, und neuronale Netzwerke haben angefangen, Biologie digital zu simulieren und so erste große medizinische Erkenntnisse zu ermöglichen. Es wird sogar endlich möglich, Computer direkt mit dem Gehirn zu verbinden.

All diesen Entwicklungen liegt das zugrunde, was ich Gesetz vom steigenden Ertragszuwachs nenne: Informationstechnologien, wie die Rechenleistung, werden exponentiell günstiger, weil jeder Fortschritt die Planung des nächsten Evolutionsschritts vereinfacht. Das führte dazu, dass man zum Zeitpunkt, an dem ich dies schreibe, für einen US-Dollar inflationsbereinigt etwa die 11 200-fache Rechenkraft kaufen kann, wenn man sie mit der damals bestehenden vergleicht, als *Menschheit 2.0* in die Buchhandlungen kam.

Die folgende Grafik, auf die ich später in diesem Buch weiter eingehen werde, fasst den wichtigsten Trend zusammen, der unsere

Preis-Leistungs-Verhältnis von Rechenkraft, 1939 bis 2023¹

Bestes Preis-Leistungs-Verhältnis bei Berechnungen pro Sekunde pro konstantem Dollar von 2023

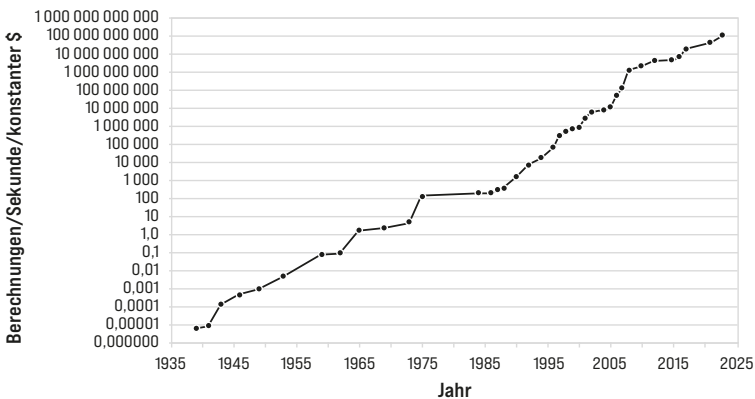


Abbildung 1: Um Maschinen maximal vergleichbar zu machen, konzentriert sich diese Grafik auf das Preis-Leistungs-Verhältnis im Zeitalter programmierbarer Computer, aber Näherungen für frühere elektromechanische Rechengeräte zeigen, dass dieser Trend mindestens in die 1880er-Jahre zurückreicht.²

technologische Zivilisation antreibt: das langfristige exponentielle Wachstum bei der Rechenleistung, die man für einen konstanten Dollar kaufen kann (dargestellt als mehr oder weniger gerade Linie in diesem logarithmischen Maßstab). Das Mooresche Gesetz besagt ja, dass Transistoren beständig kleiner werden, was Computern immer mehr Leistung ermöglicht – aber das ist nur eine Manifestation des Gesetzes vom steigenden Ertragszuwachs, das schon lange galt, bevor Transistoren überhaupt erfunden wurden, und wohl noch lange gelten wird, nachdem Transistoren ihre physikalischen Grenzen erreicht haben und durch neue Technologien ersetzt sein werden. Dieser Trend hat die moderne Welt geprägt und wird fast alle kommenden Durchbrüche, die in diesem Buch besprochen werden, direkt oder indirekt ermöglichen.

Wir sind also im Zeitplan für die Singularität. Die Dringlichkeit dieses Buches ergibt sich aus der Natur exponentieller Veränderung. Trends, die zu Beginn dieses Jahrhunderts noch kaum wahrnehmbar waren, beeinflussen jetzt aktiv das Leben von Milliarden Menschen. In den frühen 2020er-Jahren traten wir in den steil ansteigenden Teil der Exponentialkurve ein, und das Innovationstempo hat mehr Auswirkungen auf die Gesellschaft als je zuvor. Um das Ganze anschaulicher zu machen: Der Moment, in dem Sie dies lesen, liegt zeitlich wahrscheinlich näher an der Erschaffung der ersten übermenschlichen KI als am Veröffentlichungsdatum meines letzten Buches *How to Create a Mind* von 2012.* Und wir sind zeitlich wahrscheinlich näher an der Singularität als an der Veröffentlichung meines Buches *Die Intelligenz der Evolution* von 1999. Oder an der menschlichen Lebenszeit gemessen: Heute neu geborene Babys werden gerade das College abschließen, wenn sich die Singularität ereignet. Das ist auf sehr persönlicher Ebene eine andere Art von »nahe« als im Jahr 2005.

Aus diesem Grund habe ich dieses Buch jetzt geschrieben. Der jahrtausendealte Marsch der Menschheit hin zur Singularität ist zu einem Sprint geworden. In der Einführung zu *Menschheit 2.0* schrieb ich, wir befänden uns »in der frühen Übergangsphase«. Jetzt treten

* Die deutsche Ausgabe erschien 2014 unter dem Titel *Das Geheimnis des menschlichen Denkens*. (Anm. d. Übers.)

wir in die Endphase ein. In *Menschheit 2.0* ging es um etwas, das sich am fernen Horizont abzeichnete – in diesem Buch geht es um die letzten Kilometer auf dem Weg, der zum Ziel führt.

Zum Glück ist der Weg jetzt sehr viel deutlicher zu erkennen. Zwar wird es noch viele weitere technologische Veränderungen geben, bevor wir die Singularität erreichen, aber die wichtigsten Vorläufer verlassen bereits das Stadium der theoretischen Wissenschaft und treten in die Phase der aktiven Forschung und Entwicklung ein. In zehn Jahren werden Menschen mit KI interagieren, die überzeugend menschlich wirkt, und einfache Gehirn-Computer-Schnittstellen werden das tägliche Leben ebenso prägen, wie es heute die Smartphones tun. Dank einer digitalen Revolution in der Biotechnologie wird es möglich sein, Krankheiten zu heilen und die gesunde Lebenszeit der Menschen erheblich zu verlängern. Allerdings werden viele Berufstätige die negativen Folgen der wirtschaftlichen Verwerfungen spüren, und uns allen drohen neue Gefahren, wenn diese neuen Fähigkeiten aus Versehen falsch eingesetzt oder bewusst missbraucht werden. In den 2030er-Jahren wird die selbstlernende KI und eine sich entwickelnde Nanotechnologie Menschen und ihre maschinellen Schöpfungen vereinen wie nie zuvor – und dadurch Chancen und Gefahren weiter verstärken. Wenn wir die wissenschaftlichen, ethischen, gesellschaftlichen und politischen Probleme, die diese Fortschritte mit sich bringen, lösen können, werden wir bis 2045 das Leben auf der Erde grundlegend transformiert und verbessert haben. Wenn uns das nicht gelingt, könnten wir das nicht überleben. In diesem Buch geht es also um die letzten Meter zur Singularität – die Chancen und Gefahren, denen wir uns in der letzten Generation der Welt, wie wir sie kennen, gemeinsam stellen müssen.

Zunächst werden wir betrachten, wie sich die Singularität tatsächlich ereignen wird, und das dann in den Kontext der Bemühungen unserer Spezies, unsere eigene Intelligenz neu zu erfinden, stellen. Wenn man Empfindungsfähigkeit durch Technologie erschaffen will, wirft das wichtige philosophische Fragen auf. Daher wird ein Thema sein, wie dieser Übergang unsere eigene Identität und den Sinn unserer Existenz beeinflussen wird. Danach widmen wir uns den praktischen Trends, die die nächsten zehn Jahre prägen werden.

Wie ich zeigen werde, wird das Gesetz vom steigenden Ertragszuwachs exponentielle Fortschritte in vielen Bereichen vorantreiben, die das menschliche Wohl beeinflussen. Ein offensichtlicher Nachteil der Innovationen ist jedoch Arbeitslosigkeit, die durch Automatisierung in verschiedenen Formen entsteht. Diese negativen Auswirkungen sind sehr real, dennoch gibt es Anlass für langfristigen Optimismus, wie wir sehen werden – und letztendlich stehen wir nicht in Konkurrenz zur KI.

Diese Technologien werden enormen materiellen Überfluss für unsere Zivilisation ermöglichen, und danach werden wir uns darauf konzentrieren, das nächste Hindernis zu beseitigen, bevor wir unser volles Potenzial verwirklichen: die Schwächen unserer Biologie. Daher werfen wir auf den folgenden Seiten einen Blick auf die Werkzeuge, mit denen wir die Biologie selbst zunehmend beherrschen werden – indem wir zunächst den Alterungsprozess unserer Körper besiegen, dann unsere begrenzte Gehirnkapazität erweitern und schließlich die Singularität einläuten werden. Doch diese Durchbrüche könnten uns wie gesagt auch in Gefahr bringen. Revolutionäre biotechnologische, nanotechnologische oder KI-Systeme könnten zu einer existenziellen Katastrophe führen, etwa einer verheerenden Pandemie oder einer Kettenreaktion selbstreplizierender Maschinen. Am Ende dieses Buches werden wir eine Einschätzung dieser Bedrohungen vornehmen, die eine sorgfältige Planung verlangen. Dabei gibt es, so werde ich ausführen, vielversprechende Ansätze, wie sich diese Gefahren eingrenzen lassen.

Wir erleben die aufregendsten und bedeutsamsten Jahre der gesamten Menschheitsgeschichte. Niemand kann mit Sicherheit sagen, wie das Leben nach der Singularität sein wird. Aber indem wir die Übergänge zu dieser Singularität verstehen und uns darauf einstellen, können wir dazu beitragen, dass der letzte Endspurt der Menschheit sicher und erfolgreich ablaufen kann.

WO STEHEN WIR IN DEN SECHS STADIEN?

In *Menschheit 2.0* schrieb ich, die Grundlage des Bewusstseins sei Information, und ich teilte die Entwicklung von den Anfängen unseres Universums an in sechs Stadien oder Epochen ein. Jedes Stadium erschafft das darauffolgende mittels der Methoden der Informationsverarbeitung aus dem vorhergehenden Stadium. Die Evolution der Intelligenz funktioniert also durch eine indirekte Abfolge anderer Prozesse.

Die erste Epoche war die Geburt der physikalischen Gesetze und der Chemie, die diese Gesetze ermöglichen. Wenige Hunderttausend Jahre nach dem Urknall bildeten sich Atome aus Elektronen, die einen Kern aus Protonen und Neutronen umkreisen. Eigentlich sollte es gar nicht möglich sein, dass Protonen in einem Atomkern so dicht beieinander sind, weil die elektromagnetischen Kräfte sie auseinanderreiben. Doch es existiert noch eine andere Kraft, die sogenannte Nuklearkraft, die die Protonen zusammenhält. »Wer auch immer« die Regeln für das Universum aufgestellt hat, erfand auch diese zusätzliche Kraft, ohne die eine Evolution durch Atome unmöglich gewesen wäre.

Mehrere Milliarden Jahre später bildeten Atome erstmals Moleküle, die komplexe Informationen speichern konnten. Kohlenstoff war dabei der nützlichste Baustein, weil ein Kohlenstoffatom vier Bindungen eingehen kann, nicht nur eine, zwei oder drei wie

viele andere Atome. Die Wahrscheinlichkeit, in einem Universum zu leben, in dem komplexe chemische Verbindungen möglich sind, ist äußerst gering. Wenn zum Beispiel die Gravitationskraft nur ein wenig schwächer wäre, gäbe es keine Supernovae, bei denen die chemischen Elemente entstehen, aus denen Lebewesen bestehen. Wäre die Schwerkraft stärker, würden Sterne ausbrennen und sterben, bevor sich intelligente Lebewesen entwickeln könnten. Schon diese eine physische Konstante musste in einem extrem schmalen Bereich liegen, damit wir möglich wurden. Wir leben in einem sehr fein ausbalancierten Universum, das eine ausreichend hohe Ordnung aufweist, damit die Evolution ablaufen konnte.

Vor mehreren Milliarden Jahren begann die zweite Epoche: das Leben. Moleküle wurden so komplex, dass sich ein ganzer Organismus in einem einzigen Molekül definieren ließ. So konnten sich lebende Wesen mit jeweils eigener DNA entwickeln und verbreiten.

In der dritten Epoche bildeten durch DNA definierte Tiere Gehirne aus, die ihrerseits Informationen speichern und verarbeiten konnten. Diese Gehirne waren ein evolutionärer Vorteil, durch den sich im Verlauf von Millionen Jahren noch komplexere Gehirne ausbilden konnten.

In der vierten Epoche setzten gewisse Tiere dank ihrer hoch entwickelten kognitiven Fähigkeiten und ihrer Daumen Gedanken in komplexe Handlungen um: die Menschen. Unsere Art erschuf mit ihren Fähigkeiten verschiedene Technologien, mit denen sich Informationen speichern und bearbeiten ließen und lassen – von Papyrus bis Festplatten. Diese Technologien erweiterten die Fähigkeit unseres Gehirns, Informationsmuster wahrzunehmen, sich zu merken und auszuwerten. Das wurde zu einem Motor der Evolution, die weit größere Fortschritte ermöglichte als zuvor. Unsere Gehirne wuchsen um etwa 16 Kubikzentimeter alle 100 000 Jahre, während sich bei digitalen Computern das Preis-Leistungs-Verhältnis etwa alle 16 Monate verdoppelt.

In der fünften Epoche werden wir biologische menschliche Intelligenz unmittelbar über Gehirn-Computer-Schnittstellen mit der Geschwindigkeit und Rechenleistung unserer digitalen Technologie verbinden. Menschliche neuronale Informationsverarbeitung voll-

zieht sich mit mehreren Hundert Zyklen pro Sekunde, bei der digitalen Technologie sind es mehrere Milliarden Zyklen pro Sekunde. Neben Geschwindigkeit und Speicherkapazität können wir durch nicht biologische Computer auch unseren Neokortex um viele zusätzliche Schichten erweitern – und dadurch weit komplexere und abstrakte Denkleistungen ermöglichen, als wir uns derzeit vorstellen können.

In der sechsten Epoche wird sich unsere Intelligenz über das gesamte Universum ausbreiten, wobei normale Materie in Computonium umgewandelt wird, Materie mit höchstmöglicher Rechendichte.

In meinem Buch *Homo sapiens* von 1999* sagte ich voraus, dass im Jahr 2029 eine KI den Turing-Test bestehen werde – dass also eine KI per Text so kommunizieren könne, dass sie nicht mehr von einem Menschen zu unterscheiden sei. In *Menschheit 2.0* aus dem Jahr 2005 wiederholte ich diese Prognose. Wenn eine KI den Turing-Test besteht, bedeutet das, dass sie Sprache und vernünftige Argumentation so gut beherrscht wie ein Mensch. Alan Turing beschrieb diese Vorstellung im Jahr 1950,¹ führte aber nicht genau aus, wie der Test durchgeführt werden sollte. Ich habe mit dem PC-Pionier Mitch Kapor dazu eine Wette laufen, und wir haben dafür unsere eigenen Regeln definiert, die weitaus anspruchsvoller sind als andere Interpretationen.

Ich rechnete damit, dass wir, um einen gültigen Turing-Test bis spätestens 2029 zu bestehen, bis 2020 eine große Bandbreite an intellektuellen Leistungen durch KI ermöglichen müssten. Und tatsächlich hat seit dieser Prognose die KI viele der schwierigsten menschlichen Herausforderungen gemeistert – von Quizshows und Spielen wie *Jeopardy!* und *Go* bis zu ernsthaften Anwendungsgebieten wie etwa der Radiologie oder der Medikamentenentwicklung. Während ich dies schreibe, erweitern KI-Systeme wie Gemini und GPT-4 ihre Fähigkeiten auf viele verschiedene Leistungsgebiete – vielversprechende Schritte auf dem Weg zu allgemeiner Intelligenz.

* Die deutsche Ausgabe erschien 2000 unter dem Titel *Homo sapiens: Leben im 21. Jahrhundert – was bleibt vom Menschen?* Im Folgenden wird der Haupttitel verwendet. (Anm. d. Übers.)

Letztendlich wird ein Programm, wenn es den Turing-Test besteht, in vielen Bereichen dümmer erscheinen müssen, als es in Wahrheit ist, weil es sonst eindeutig als KI zu erkennen wäre. Wenn es zum Beispiel jede Matheaufgabe in Sekundenschnelle lösen könnte, würde es bei dem Test durchfallen. Auf dem Turing-Test-Niveau müsste eine KI also über Fähigkeiten verfügen, die jene der besten menschlichen Fachleute in den meisten Bereichen übersteigt.

Die Menschheit befindet sich aktuell in der vierten Epoche. Unsere Technologie zeitigt jetzt schon bei manchen Aufgaben Ergebnisse, die unser Verständnis übersteigen. Bei den Aspekten des Turing-Tests, die KI noch nicht gemeistert hat, machen wir immer schnellere Fortschritte. Wenn die KI den Turing-Test besteht, was ich wie gesagt für das Jahr 2029 erwarte, treten wir in die fünfte Epoche ein.

Eine entscheidende Entwicklung der 2030er-Jahre wird sein, dass wir die oberen Bereiche des Neokortex eines Menschen mit der Cloud verbinden und so unser Denkvermögen unmittelbar erweitern können. Wenn das möglich wird, werden die nicht biologischen Teile unseres Gehirns im Vergleich zu den biologischen die vieltausendfache kognitive Leistung liefern.

Dank der exponentiellen Beschleunigung der Fortschritte werden wir unser Gehirn bis zum Jahr 2045 um das Millionenfache erweitern können. Die Geschwindigkeit und der Umfang dieser Transformation sind unfassbar, und sie rechtfertigen, dass wir die Singularitätsmetapher aus der Physik entlehnen, um unsere Zukunft zu beschreiben.

DIE NEUERFINDUNG DER INTELLIGENZ

WAS BEDEUTET ES, WENN INTELLIGENZ NEU ERFUNDEN WIRD?

Wenn die Geschichte des Universums die Geschichte sich entwickelnder Paradigmen der Informationsverarbeitung ist, dann kommt die Menschheit erst in der zweiten Hälfte vor. Unser Kapitel in dieser umfassenden Geschichte erzählt, wie sich Tiere mit biologischen Gehirnen zu transzendenten Wesen entwickeln, deren Gedanken und Identitäten nicht mehr an genetische Vorgaben gebunden sind. In den 2020er-Jahren werden wir in die letzte Phase dieser Transformation eintreten – die Neuerfindung der naturgegebenen Intelligenz auf einem leistungsfähigeren digitalen Substrat, mit dem wir dann verschmelzen. Mit diesem Vorgang wird die vierte Epoche des Universums die fünfte Epoche hervorbringen.

Doch wie wird das konkret ablaufen? Um zu verstehen, was es heißt, die Intelligenz neu zu erfinden, müssen wir zunächst auf die Geburt von KI zurückblicken und auf die beiden großen Denkschulen, die daraus entstanden. Wir werden das mit den Erkenntnissen der Neurowissenschaft darüber, wie Kleinhirn und Neokortex die menschliche Intelligenz ermöglichten, in Verbindung setzen und so erkennen, warum die eine Denkschule über die andere siegte. Wir werden betrachten, wie Deep Learning derzeit die Leistungsfähig-

keit des Neokortex nachbildet, und danach einschätzen können, was KI noch leisten muss, um ein menschliches Niveau zu erreichen, und wie wir erkennen werden, wenn sie es erreicht hat. Schließlich werden wir uns damit beschäftigen, wie wir mithilfe der übermenschlichen KI Gehirn-Computer-Schnittstellen bauen werden, die unseren Neokortex durch Schichten virtueller Neuronen stark erweitern. Dadurch werden uns völlig neue Denkweisen eröffnet und unsere Intelligenz am Ende millionenfach ausgedehnt: Das ist die Singularität.

DIE GEBURT VON KI

Im Jahr 1950 veröffentlichte der britische Mathematiker Alan Turing (1912–1954) einen Artikel in der Zeitschrift *Mind* mit dem Titel »Computing Machinery and Intelligence«.¹ Darin stellte Turing die wichtigste Frage in der Geschichte der Wissenschaft: »Können Maschinen denken?« Die Vorstellung von denkenden Maschinen reicht mit dem bronzenen Riesen Talos bis in die antike griechische Mythologie zurück,² aber Turing reduzierte die Vorstellung auf etwas, das sich empirisch überprüfen ließ. Er schlug vor, anhand des »Imitationsspiels« – das wir heute als Turing-Test kennen – festzustellen, ob eine Maschine in der Lage war, dieselben kognitiven Aufgaben auszuführen wie unser Gehirn. Bei diesem Test kommunizieren menschliche Prüfer per Textnachrichten mit einer KI und menschlichen Probanden, ohne zu sehen, mit wem sie sich gerade unterhalten. Die Prüfer stellen Fragen zu beliebigen Themen oder Situationen. Wenn sie nach einer bestimmten Zeitspanne nicht sagen können, welcher Gesprächspartner die KI war und welcher menschlich, hat die KI den Test bestanden.

Turing machte aus der philosophischen Vorstellung also eine wissenschaftliche und löste dadurch große Begeisterung unter Forschern aus. Im Jahr 1956 organisierte der Mathematikprofessor John McCarthy (1927–2011) eine zweimonatige Studie mit zehn Teilnehmern am Dartmouth College in Hannover, New Hampshire, mit der folgenden Zielsetzung:³

Die Studie wird von der Annahme ausgehen, dass alle Aspekte des Lernens und jedes andere Merkmal von Intelligenz in Grundzügen so präzise beschrieben werden können, dass man eine Maschine dazu bringen kann, sie zu simulieren. Es soll untersucht werden, wie Maschinen dazu gebracht werden können, dass sie Sprache verwenden, Abstraktionen und Konzepte bilden und Probleme lösen können, die bisher Menschen vorbehalten sind, und wie sie sich selbst verbessern können.⁴

Bei den Vorbereitungen zur Studie schlug McCarthy für dieses Arbeitsgebiet, das letztendlich alle anderen Arbeitsgebiete automatisieren würde, den Namen »künstliche Intelligenz« vor.⁵ Mir selbst gefällt diese Bezeichnung nicht, weil »künstlich« andeutet, diese Form der Intelligenz sei »nicht real«, aber der Begriff hat sich durchgesetzt.

Die Studie wurde durchgeführt, aber ihre Zielsetzung – insbesondere Maschinen dazu zu bringen, Aufgaben zu verstehen, die in natürlicher Sprache formuliert waren – wurde in den zwei Monaten nicht erreicht. An diesem Problem wird heute immer noch gearbeitet – inzwischen natürlich mit weit mehr als zehn Leuten. Nach Angaben des chinesischen Branchenriesen Tencent gab es im Jahr 2017 bereits mehr als 300 000 »KI-Forscher und -Anwender« weltweit,⁶ und der *Global AI Talent Report* von Jean-François Gagné, Grace Kiser und Yoan Mantha aus dem Jahr 2019 führt 22 400 KI-Experten auf, die eigene Forschungsergebnisse veröffentlicht haben – von denen etwa 4000 als äußerst einflussreich eingeschätzt wurden.⁷ Laut dem Institute for Human-Centered Artificial Intelligence der Stanford University veröffentlichten KI-Forscher im Jahr 2021 mehr als 496 000 Fachschriften und reichten mehr als 141 000 Patentanträge ein.⁸ Im Jahr 2022 investierten Unternehmen weltweit 189 Milliarden Dollar in KI. Damit hatte sich die Investitionssumme innerhalb von zehn Jahren um das 13-Fache erhöht.⁹ Wenn Sie dies lesen, wird die Zahl noch höher liegen.

1968 wäre all das kaum vorstellbar gewesen. Und dennoch entsprach die Zielsetzung der Studie in Dartmouth der Erschaffung einer KI, die den Turing-Test bestehen konnte. Seit 1999, seit mei-

nem Buch *Homo sapiens*, prognostiziere ich wie erwähnt, dass wir dies bis zum Jahr 2029 schaffen werden. Zum Zeitpunkt seiner Veröffentlichung glaubten viele Beobachter, dieser Meilenstein könne *niemals* erreicht werden.¹⁰ Bis vor Kurzem galt diese Prognose unter Fachleuten als äußerst optimistisch. Eine Umfrage aus dem Jahr 2018 ergab, dass die Mehrheit der KI-Experten nicht vor 2060 damit rechnete, dass eine Maschine menschenähnliche Intelligenz erlangen würde.¹¹ Aber die jüngsten Fortschritte bei den Large Language Models haben die Zielerwartung stark verschoben. Während ich an der ersten Fassung dieses Buches arbeitete, lag der Durchschnitt bei Metaculus, der wichtigsten Prognose-Website der Welt, zwischen 2040 und 2050. Aber der überraschende Fortschritt bei KI in den letzten beiden Jahren hat die Erwartungen erschüttert, und im Mai 2022 stimmte die Mehrheit bei Metaculus mit meiner Prognose für 2029 exakt überein.¹² Seitdem lag der Konsens zeitweise sogar bei 2026 und hatte damit meine Prognose sogar überholt!¹³

Viele der jüngsten Durchbrüche bei KI haben sogar Fachleute überrascht. Sie gelingen nicht nur schneller, als die meisten erwartet haben, sondern treten vor allem scheinbar plötzlich auf, ohne Vorwarnung, dass ein großer Sprung nach vorn bevorsteht. So sagte Tomaso Poggio, am Massachusetts Institute for Technology (MIT) Experte für KI und Kognitionswissenschaft, im Oktober 2014 noch: »Die Fähigkeit, den Inhalt eines Bildes zu beschreiben, wäre eine der größten intellektuellen Herausforderungen für eine Maschine. Bevor wir dieses Problem lösen können, werden wir noch einen weiteren Forschungszyklus brauchen.«¹⁴ Poggio schätzte, mit diesem Durchbruch sei erst in zwanzig Jahren zu rechnen. Bereits einen Monat später stellte Google eine KI zur Objekterkennung vor, die ebendas konnte. Von Raffi Khatchadourian von *The New Yorker* dazu befragt, flüchtete sich Poggio in philosophische Skepsis, ob diese Fähigkeit für echte Intelligenz stehe. Damit will ich keineswegs Poggio kritisieren, sondern nur etwas beschreiben, zu dem wir alle neigen. Bevor eine KI ein Ziel erreicht, scheint dieses Ziel ungemein schwer zu erreichen und einzigartig für uns Menschen zu sein. Nachdem die KI es erreicht hat, wird dieses Ziel aus menschlicher Sicht abgewertet. Oder anders ausgedrückt: Unsere Fortschritte sind in Wahrheit be-

deutsamer, als es rückblickend erscheint. Auch aus diesem Grund halte ich optimistisch an meiner Prognose für 2029 fest.

Warum kam es zu diesen überraschenden Durchbrüchen? Die Antwort liegt in einem theoretischen Problem, das bis zu den Anfängen des Forschungsbereichs zurückreicht. Im Jahr 1964, als ich noch die Highschool besuchte, lernte ich zwei Pioniere der künstlichen Intelligenz kennen: Marvin Minsky (1927–2016), der den Dartmouth Workshop mitorganisiert hatte, und Frank Rosenblatt (1928–1971). Im Jahr 1965 schrieb ich mich am MIT ein und studierte bei Minsky, der damals Grundlagenforschung betrieb, auf der die dramatischen Durchbrüche bei der KI, die wir heute beobachten, basieren. Minsky lehrte, es gebe zwei Methoden, wie Automaten Aufgaben lösen könnten: die symbolische und die konnektionistische.

Bei der symbolischen Methode wird mit regelbasierten Ausdrücken beschrieben, wie ein menschlicher Experte ein Problem lösen würde. In manchen Fällen war ein solches System erfolgreich. So führte die RAND Corporation im Jahr 1959 den General Problem Solver (GPS) vor – ein Computerprogramm, das einfache mathematische Axiome kombinierte, um logische Probleme zu lösen.¹⁵ Herbert A. Simon, J. C. Shaw und Allen Newell programmierten den General Problem Solver so, dass er theoretisch jedes Problem lösen konnte, das sich in einer Reihe von wohlgeformten Formeln (WWF) ausdrücken ließ. Dazu musste der General Problem Solver (GPS) in jeder Phase des Prozesses eine WWF (entsprechend einem Axiom) verwenden und die WWF dann methodisch zu einem mathematischen Beweis der Antwort zusammenfügen.

Das entspricht im Prinzip dem Vorgehen in der Algebra. Wenn man weiß, dass $2 + 7 = 9$ ist und eine unbekannte Zahl x 10 ergibt, wenn man sie zu 7 addiert, dann kann man beweisen, dass $x = 3$ ist. Doch lässt sich diese Art Logik noch auf weit mehr anwenden als das Lösen von Gleichungen. Wir wenden sie auch an (ohne darüber nachzudenken), wenn wir uns fragen, ob etwas einer bestimmten Definition entspricht. Wenn man weiß, dass eine Primzahl nur durch 1 und sich selbst geteilt werden kann, dass 11 ein Teiler von 22 ist und 1 nicht gleich 11 ist, kann man daraus schließen, dass 22 keine Primzahl ist. Der GPS basierte auf den einfachsten und grundlegendsten

Axiomen, die es gibt, und konnte so derartige Berechnungen bei weit komplexeren Aufgaben durchführen. Im Prinzip tun menschliche Mathematiker nichts anderes – nur kann eine Maschine (zumindest theoretisch) alle Kombinationsmöglichkeiten der grundlegenden Axiome durchprobieren, um die Lösung zu finden.

Ein Beispiel: Wenn an jedem Punkt 10 Axiome zur Auswahl stünden und man 20 Axiome für die Lösung bräuchte, ergäbe das 10^{20} oder 100 Milliarden Milliarden mögliche Lösungen. Heute kann man dank Computern mit derart großen Zahlen umgehen, aber im Jahr 1959 lag das noch weit jenseits der Leistungsfähigkeit von Computern. Damals konnte der DEC PDP-1-Computer circa 100 000 Operationen pro Sekunde ausführen.¹⁶ Im Jahr 2023 schaffte eine Google Cloud A3 virtuelle Maschine rund 26 000 000 000 000 000 000 Rechenoperationen pro Sekunde.¹⁷ Für einen Dollar bekommt man heute die 1,6-billionenfache Rechenleistung wie zu Zeiten des GPS.¹⁸ Aufgaben, die mit der Technologie von 1959 Zehntausende Jahre erfordert hätten, sind heute in wenigen Minuten mit einem handelsüblichen Computer lösbar. Um die begrenzten Rechenfähigkeiten auszugleichen, wurde der GPS mit Heuristiken programmiert, die mögliche Lösungen nach Priorität sortierten. Die Heuristiken funktionierten in manchen Fällen, und ihre Erfolge stützten die Vorstellung, dass sich mit Computern letztlich jedes präzise formulierte Problem lösen ließ.

Ein weiteres Beispiel war das MYCIN-System, das in den 1970ern entwickelt wurde, um Infektionskrankheiten zu diagnostizieren und Behandlungen zu empfehlen. Im Jahr 1979 verglich ein Team aus Fachleuten seine Leistung mit dem Können eines menschlichen Arztes und stellte fest, dass MYCIN genauso gut oder sogar besser abschnitt als jeder Arzt.¹⁹

Eine typische MYCIN-»Regel« sah folgendermaßen aus:

- WENN**
- 1) die Infektion, die therapiert werden soll, Meningitis ist und
 - 2) es sich um eine Pilzinfektion handelt und
 - 3) in der Kultur aus dem Abstrich keine Organismen zu erkennen sind und
 - 4) der Patient keine Vorerkrankungen hat und

- 5) der Patient sich in einer Region aufgehalten hat, in der Kokzidioidomykosen vorkommen, und
- 6) der Patient entweder [s]chwarz, [a]siatisch oder [i]ndio ist und
- 7) im Hirnwasser keine Antigene für Kryptokokken nachweisbar sind,

DANN liegt es nahe (.5), dass Kryptokokken nicht zu den Organismen gehören, die die Infektion verursachen (und die nicht in Kulturen oder Abstrichen zu erkennen sind).²⁰

Ende der 1980er-Jahre verwendeten diese »Expertensysteme« Wahrscheinlichkeitsmodelle und konnten viele Datenquellen kombinieren, um zu einer Entscheidung zu gelangen.²¹ Eine einzelne Wenn-dann-Regel reichte nicht aus, aber indem das System viele Tausende solcher Regeln kombinierte, konnte es verlässliche Entscheidungen für ein beschränktes Problem treffen.

Der symbolische Ansatz wird seit mehr als fünfzig Jahren angewendet, doch gibt es dabei eine »Komplexitätsgrenze«.²² Wenn MYCIN und andere derartige Systeme einen Fehler machten, konnte man dieses spezielle Problem durch eine Korrektur lösen, doch löste das wiederum drei andere Fehler in anderen Situationen aus. Offenbar war mit diesen Systemen nur eine begrenzte Komplexität möglich, was die Bandbreite der Probleme in der realen Welt, auf die sie anwendbar waren, stark einschränkte.

Man kann die Komplexität eines regelbasierten Systems auch als Menge möglicher Fehlerpunkte betrachten. Mathematisch hat eine Gruppe aus n Objekten 2^{n-1} Untermengen (die leere Menge nicht mitgezählt). Wenn eine KI also eine Regelmenge verwendet, die nur eine Regel enthält, gibt es nur einen Fehlerpunkt: Funktioniert diese Regel für sich genommen korrekt oder nicht? Bei zwei Regeln gibt es bereits drei Fehlerpunkte: jede Regel für sich und Interaktionen, bei denen diese beiden Regeln kombiniert werden. Die Anzahl wächst exponentiell. Fünf Regeln bedeuten 31 mögliche Fehlerpunkte, bei 10 Regeln sind es 1023, bei 100 Regeln mehr als 1000 Milliarden Milliarden und bei 1000 Regeln gibt es mehr als ein Googol

Googol Googol mögliche Fehlerpunkte. Je mehr Regeln es bereits gibt, umso mehr mögliche Untermengen kommen mit jeder weiteren Regel dazu. Selbst wenn nur ein winziger Teil möglicher Regelkombinationen zu einem neuen Problem führt, erreicht man irgendwann den Punkt (wo genau er liegt, unterscheidet sich von Situation zu Situation), an dem eine neue Regel, die man einführt, um ein Problem zu beheben, wahrscheinlich mehr als ein neues Problem auslöst. Das ist die Komplexitätsgrenze.

Das wahrscheinlich langlebigste Expertensystemprojekt ist CYC (vom englischen Wort *encyclopedic*), das Douglas Lenat und Kollegen bei Cycorp 1984 ins Leben riefen.²³ Ziel von CYC ist, alles »allgemeine Weltwissen« zu codieren – also weithin bekannte Fakten wie *Ein Ei, das fallen gelassen wird, zerbricht* oder *Ein Kind, das mit schmutzigen Schuhen durch die Küche läuft, verärgert seine Eltern*. Diese vielen Millionen kleiner Ideen sind nirgendwo deutlich niedergeschrieben. Es sind unausgesprochene Annahmen, die dem Verhalten und den Überlegungen von Menschen zugrunde liegen und die notwendig sind, um zu verstehen, was der Durchschnittsmensch zu verschiedenen Themen weiß. Das CYC-System stellt dieses Wissen durch symbolische Regeln dar und ist daher ebenfalls mit einer Komplexitätsgrenze konfrontiert.

In den 1960ern, als Minsky mir die Vor- und Nachteile des symbolischen Ansatzes erläuterte, erkannte ich den Mehrwert der konnektionistischen Methode. Dabei erschaffen Netzwerke aus Knoten Intelligenz durch Struktur, nicht Inhalte. Statt cleverer Regeln werden dumme Knoten eingesetzt, die durch ihre Anordnung Erkenntnisse direkt aus den Daten extrahieren können. Dadurch sind sie in der Lage, kaum wahrnehmbare Muster aufzuspüren, die einem menschlichen Programmierer, der symbolische Regeln aufstellt, nie in den Sinn kämen. Ein wichtiger Vorteil des konnektionistischen Ansatzes besteht darin, dass man damit Probleme lösen kann, ohne sie verstehen zu müssen. Selbst wenn wir fehlerfreie Regeln für einen symbolischen KI-Problemlöser aufstellen könnten (was wir nicht können), wüssten wir immer noch nicht genau, welche Regeln optimal wären.

Die konnektionistische Methode bringt gute Ergebnisse bei komplexen Problemen, aber sie ist auch ein zweischneidiges Schwert.

Konnektionistische KI wird leicht zur »Blackbox« – sie spuckt die richtige Antwort aus, kann aber nicht erklären, wie sie sie gefunden hat.²⁴ Das kann sehr problematisch werden, weil die Gedankengänge für Entscheidungen, bei denen viel auf dem Spiel steht – etwa bezüglich einer medizinischen Behandlung, im Strafvollzug, der Epidemiologie oder im Risikomanagement –, nachvollziehbar sein sollten. Aus diesem Grund arbeiten inzwischen viele KI-Experten daran, die »Transparenz« (oder »mechanistische Interpretierbarkeit«) von Entscheidungen, die auf maschinellem Lernen basieren, zu verbessern.²⁵ Wie effektiv die Transparenz sein wird, wenn Deep-Learning-Systeme komplexer und leistungsfähiger werden, bleibt abzuwarten.

Als ich begann, mich mit dem konnektionistischen Ansatz zu beschäftigen, waren die Systeme noch erheblich einfacher. Die Grundidee war, ein Computermodell nach dem Vorbild menschlicher neuronaler Netzwerke zu erstellen. Zunächst waren das sehr abstrakte Überlegungen, weil die Methode entwickelt wurde, bevor man genau verstand, wie biologische neuronale Netzwerke organisiert sind.

DIAGRAMM EINES EINFACHEN NEURONALEN NETZWERKS

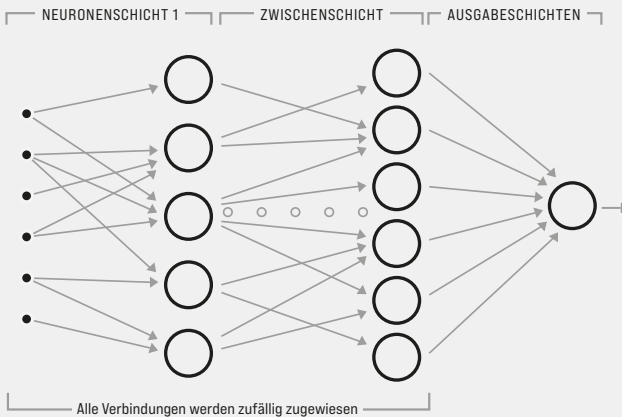


Abbildung 2

Dies ist das Grundschemata eines Neuronales-Netzwerk-Algorithmus. Es gibt zahlreiche mögliche Variationen, und wer solch ein System entwirft, muss gewisse unverzichtbare Parameter und Methoden bereitstellen (Details siehe unten).

Wenn man ein neuronales Netz entwickeln möchte, um ein Problem zu lösen, braucht es dazu folgende Schritte:

- Eingabe definieren.
- Topologie des neuronalen Netzwerks definieren (d. h. die Neuronschichten und die Verbindungen zwischen den Neuronen).
- Das neuronale Netzwerk mit Beispielen des Problems trainieren.
- Das trainierte neuronale Netzwerk neue Beispiele des Problems lösen lassen.
- Das neuronale Netzwerk der Öffentlichkeit präsentieren.

Diese Schritte (mit Ausnahme des letzten) werden im Folgenden im Detail beschrieben:

DIE EINGABE DES PROBLEMS

Ein Problem wird durch eine Reihe von Zahlen in das neuronale Netz eingespeist. Eine solche Eingabe kann folgendermaßen erfolgen:

- Bei einem visuellen Mustererkennungssystem als zweidimensionales Zahlenarray, das die Pixel eines Bildes repräsentiert.
- Bei einem akustischen Erkennungssystem (z. B. für Sprache) als zweidimensionales Zahlenarray, das einen Laut repräsentiert. Dabei steht die erste Dimension für die Parameter des Lauts (z. B. Frequenzkomponenten), die zweite für verschiedene Zeitpunkte.
- Bei einem beliebigen Mustererkennungssystem als n -dimensionales Zahlenarray, das das Eingabemuster repräsentiert.

DIE TOPOLOGIE DEFINIEREN

Bei einem neuronalen Netzwerk besteht die Architektur jedes Neurons aus:

- mehreren Inputs (oder Eingabestellen), wobei jeder Input entweder mit dem Output (der Ausgabestelle) eines anderen Neurons oder einer der Eingabezahlen »verbunden« ist, und
- im Allgemeinen einem Output, der entweder mit dem Input eines anderen Neurons (in der Regel in einer höheren Schicht) oder dem finalen Output verbunden ist.

Die erste Neuronenschicht einrichten

- In der ersten Schicht werden N_0 Neuronen angelegt. Alle Inputs dieser Neuronen sind mit »Punkten« (z. B. Zahlen) in der Problemeingabe »verbunden«. Diese Verbindungen können zufällig oder anhand eines evolutionären Algorithmus festgelegt werden (siehe unten).
- Jeder Verbindung wird ein initiales synaptisches Gewicht zugewiesen. Diese Gewichte können anfangs alle gleich sein, zufällig zugewiesen oder auf andere Art bestimmt werden (siehe unten).

Weitere Neuronenschichten anlegen

Man legt M Neuronenschichten an. In jeder Schicht werden die Neuronen eingerichtet.

Für Schicht _{i} bedeutet das:

- N_i Neuronen in Schicht _{i} einrichten. Alle Inputs dieser Neuronen sind mit den Outputs von Neuronen in Schicht _{$i-1$} »verbunden« (siehe Variationen unten).
- Jeder Verbindung wird ein initiales synaptisches Gewicht zugewiesen. Diese Gewichte können anfangs alle gleich sein, zufällig zugewiesen oder auf andere Art bestimmt werden (siehe unten).
- Die Ausgaben der Neuronen in Schicht _{M} sind die Ausgaben des neuronalen Netzes (siehe Variationen unten).

DIE ERKENNUNGSDURCHGÄNGE

Funktionsweise der Neuronen

Wenn das Neuron eingerichtet ist, geschieht bei jedem Erkennungsdurchgang Folgendes:

- Jede gewichtete Eingabe an das Neuron wird berechnet, indem die Ausgabe des Neurons, von dem diese Eingabe stammt, (oder die initiale Eingabe) mit dem synaptischen Gewicht der Verbindung zu diesem vorgeschalteten Neuron multipliziert wird.
- All diese gewichteten Eingaben, die im Neuron ankommen, werden addiert.
- Wenn diese Summe eine bestimmte Schwelle übersteigt, »feuert« das Neuron, wobei die Ausgabe 1 ist. Liegt diese Summe unter dem Schwellenwert, ist die Ausgabe 0 (siehe Variationen unten).

Ablauf eines Erkennungsdurchgangs

In jeder Schicht, von Schicht₀ bis Schicht_M, und bei jedem Neuron in der jeweiligen Schicht geschieht Folgendes:

- Die gewichteten Eingaben werden addiert. (Jede gewichtete Eingabe = Ausgabe des anderen Neurons, mit dem das aktuelle Neuron verbunden ist, [oder die initiale Eingabe] multipliziert mit dem synaptischen Gewicht dieser Verbindung.)
- Wenn diese Summe der gewichteten Eingaben den für das jeweilige Neuron festgelegten Schwellenwert übersteigt, wird der Ausgabewert dieses Neurons auf 1 gesetzt, sonst auf 0.

TRAINING DES NEURONALEN NETZES

- Mehrere Erkennungsdurchgänge werden mit dem Beispielpapier durchgeführt.
- Nach jedem Durchgang wird das synaptische Gewicht aller Verbindungen zwischen den Neuronen angepasst, um die Leistung des neuronalen Netzes zu verbessern (wie das geht, siehe unten).

- Dieses Training wird durchgeführt, bis sich die Genauigkeitsquote des neuronalen Netzes nicht weiter verbessert (d. h., bis sie eine Asymptote erreicht).

WICHTIGE DESIGNENTSCHEIDUNGEN

In dem einfachen Schema oben muss der Entwickler dieses neuronalen Netzalgorithmus zu Anfang Folgendes festlegen:

- wofür die Eingabezahlen stehen
- die Anzahl der Neuronenschichten
- die Anzahl der Neuronen pro Schicht (nicht alle Schichten müssen dieselbe Anzahl Neuronen enthalten)
- die Anzahl der Inputs (d. h. der Verbindungen zwischen den Neuronen) an jedem Neuron pro Schicht (auch diese Zahl kann von Neuron zu Neuron und Schicht zu Schicht variieren)
- die eigentliche »Vernetzung« (also die Verbindungen). Für jedes Neuron in jeder Schicht gibt es eine Liste anderer Neuronen, deren Outputs die Inputs dieses Neurons darstellen. Ein wichtiger Aspekt. Dies kann auf verschiedene Weisen umgesetzt werden:
 1. Die Verbindungen im neuronalen Netz werden zufällig festgelegt.
 2. Man bestimmt durch einen evolutionären Algorithmus (siehe unten) die optimalen Verbindungen.
 3. Der Systementwickler entscheidet nach bestem Wissen selbst, wo Verbindungen angelegt werden.
- das initiale synaptische Gewicht jeder Verbindung. Auch dies kann man auf verschiedene Weise umsetzen:
 1. Alle synaptischen Gewichte bekommen denselben Wert.
 2. Die synaptischen Gewichte werden zufällig festgelegt.
 3. Ein evolutionärer Algorithmus bestimmt die optimalen Initialwerte.
 4. Der Systementwickler setzt die Initialwerte selbst nach bestem Wissen fest.
- die Schwellenwerte, bei dem die einzelnen Neuronen feuern

- die Ausgabe, die Folgendem entsprechen kann:
 1. der Ausgabe der Neuronenschicht_M
 2. der Ausgabe eines einzelnen Neurons, deren Eingabe die Ausgaben der Neuronen in Schicht_M sind
 3. einer Funktion (z. B. der Summe) der Ausgaben der Neuronen in Schicht_M
 4. einer anderen Funktion von Neuronenausgaben in mehreren Schichten
- das synaptische Gewicht aller Verbindungen, die während des Trainings des neuronalen Netzes angepasst werden müssen. Eine wichtige Entscheidung, zu der viel geforscht und über die viel diskutiert wird. Man kann sie auf verschiedene Weisen umsetzen:
 1. Bei jedem Erkennungsdurchgang wird das synaptische Gewicht um einen (in der Regel kleinen) festen Betrag so erhöht oder verringert, dass die Ausgabe des neuronalen Netzes der korrekten Antwort näher kommt. Man kann dazu ausprobieren, ob der höhere oder der niedrigere Wert den gewünschten Effekt hat. Das kann zeitaufwendig sein, daher gibt es andere Methoden, um jeweils festzulegen, ob das synaptische Gewicht erhöht oder verringert werden sollte.
 2. Es gibt andere statistische Methoden, um die synaptischen Gewichte nach jedem Erkennungsdurchgang so zu verändern, dass die Ausgabe des neuronalen Netzes näher an der korrekten Antwort liegt.
 3. Man kann ein neuronales Netz auch dann trainieren, wenn die Antworten auf die Erkennungsdurchgänge nicht alle korrekt sind. Man kann also reale Trainingsdaten verwenden, die eine naturgemäße Fehlerquote aufweisen. Entscheidend für den Erfolg eines Erkennungssystems auf Basis eines neuronalen Netzes ist die Datenmenge, mit der trainiert wird. Meist braucht es erhebliche Datenmengen, um zufriedenstellende Ergebnisse zu erzielen. Wie bei menschlichen Schülern hat auch die Zeit, die ein neuronales Netz mit dem Lernen bringt, entscheidenden Einfluss auf seine Leistung.

VARIATIONEN

Von den oben genannten Punkten sind viele Variationen umsetzbar:

- Die Topologie kann auf verschiedene Arten festgelegt werden. Insbesondere können die Verbindungen zwischen den Neuronen entweder zufällig festgelegt werden oder durch einen evolutionären Algorithmus, der die Wirkung von Mutationen und natürlicher Auslese bei einem Netzwerkentwurf nachahmt.
- Es gibt verschiedene Möglichkeiten, die initialen synaptischen Gewichte festzulegen.
- Die Eingabe für die Neuronen in Schicht_i muss nicht unbedingt von den Outputs der Neuronen in Schicht_{i-1} stammen. Sie können auch von jeder höheren oder niedrigeren Schicht kommen.
- Die finale Ausgabe lässt sich mit verschiedenen Methoden bestimmen.
- Bei der oben beschriebenen Methode gibt es nur »alles oder nichts« (1 oder 0) bei der Ausgabe eines Neurons. Dies wird als Nichtlinearität bezeichnet. Auch andere nicht lineare Funktionen können verwendet werden. Oft wird eine Funktion eingesetzt, die in rascher, aber eher gradueller Folge von 0 bis 1 führt. Auch andere Zahlen als 0 oder 1 können ausgegeben werden.
- Welche der unterschiedlichen Methoden, um die synaptischen Gewichte beim Training anzupassen, eingesetzt werden, ist entscheidend für das Design des Netzwerks.

Das obige Schema beschreibt ein »synchrones« neuronales Netz, bei dem in jedem Erkennungsdurchlauf nacheinander die Ausgaben jeder Schicht berechnet werden, von Schicht₀ bis Schicht_M. In einem echten parallelen System, in dem jedes Neuron unabhängig von den anderen funktioniert, können Neuronen auch »asynchron« arbeiten. Bei einem asynchronen Ansatz scannt jedes Neuron ständig seine Inputs und feuert, sobald die Summe der gewichteten Eingaben seine spezifische Schwelle (oder den Wert, den seine Ausgabefunktion angibt) übersteigt.

Danach braucht man echte Beispiele, anhand derer das System lernen kann, wie sich ein Problem lösen lässt. Zu Anfang werden die Verbindungen und synaptischen Gewichte im neuronalen Netz üblicherweise zufällig festgelegt, daher sind die Antworten, die dieses untrainierte neuronale Netz ausgibt, ebenfalls zufällig. Ein neuronales Netzwerk muss sein Sachgebiet erst lernen wie das Säugetiergehirn, dem es (zumindest grob) nachgebildet ist. Ein neuronales Netz weiß zu Beginn gar nichts, ist aber darauf programmiert, eine »Belohnungsfunktion« zu maximieren. Dann wird es mit Trainingsdaten gefüttert (z. B. Fotos von Corgis und Fotos ohne Corgis, die jeweils von Menschen vorher beschriftet wurden). Wenn das neuronale Netz zu einem korrekten Ergebnis gelangt (wenn es z. B. korrekt feststellt, ob in dem Bild ein Corgi zu sehen ist), bekommt es eine Rückmeldung, anhand der dann die Gewichte der einzelnen Verbindungen zwischen den Neuronen angepasst werden können. Bei Verbindungen, die zur richtigen Antwort passen, wird das Gewicht erhöht, bei anderen verringert.

Mit der Zeit ordnet sich das neuronale Netz so an, dass es ohne Anleitung die richtige Antwort ausgibt. Experimente haben gezeigt, dass neuronale Netze ihr Sachgebiet auch mit unzuverlässigen Lehrern lernen können. Wenn die Trainingsdaten zu 60 Prozent richtig ausgezeichnet sind, kann ein neuronales Netz seine Lektion immer noch mit einer Treffsicherheit von mehr als 90 Prozent umsetzen. Unter manchen Bedingungen reicht sogar ein noch geringerer Anteil korrekter Auszeichnungen.²⁶

Dass ein Lehrer einen Schüler so anleiten kann, dass der Schüler am Ende den Lehrer übertrifft, ist nicht leicht zu verstehen. Ebenfalls verwirrend ist, dass unzuverlässige Trainingsdaten zu exzellenten Leistungen führen können. Das geht, weil sich Fehler gegenseitig aufheben können. Nehmen wir als Beispiel ein neuronales Netz, dem beigebracht werden soll, die Ziffer 8 in handschriftlichen Beispielen der Ziffern 0 bis 9 zu erkennen. Und nehmen wir weiterhin an, ein Drittel der Auszeichnungen sei fehlerhaft – eine zufällige Mischung aus Achten, die als Vierer, Fünfer oder Achter ausgezeichnet sind usw. Wenn der Datensatz groß genug ist, werden sich diese Ungenauigkeiten ausgleichen und das Training nicht groß beein-

trächtigen. Die entscheidende Information im Datensatz, wie Achter aussehen, bleibt erhalten, und das neuronale Netz kann durch Training immer noch ein hohes Niveau erreichen.

Trotz dieser Stärken hatten frühe konnektionistische Systeme eine schwerwiegende Einschränkung. Neuronale Netze mit nur einer Schicht waren mathematisch nicht in der Lage, derartige Probleme zu lösen.²⁷ Als ich 1964 Professor Frank Rosenblatt an der Cornell University besuchte, führte er mir ein einschichtiges neuronales Netzwerk namens Perzeptron vor, das gedruckte Buchstaben erkannte. Ich wandelte die Eingaben versuchsweise leicht ab. Bei Autoassoziationen (wenn der Buchstabe teilweise verdeckt war) funktionierte das System recht gut, bei der Invarianz (wenn Größe und Schriftart anders waren) weniger gut.

Im Jahr 1969 kritisierte Minsky das stark gestiegene Interesse für dieses Gebiet, obwohl er 1953 Pionierarbeit für neuronale Netzwerke geleistet hatte. Er und Seymour Papert, die beiden Gründer des MIT Artificial Intelligence Laboratory, führten damals in ihrem Buch *Perceptrons* formal aus, warum ein Perzeptron niemals feststellen könne, ob ein gedrucktes Bild zusammenhängend sei oder nicht. Die beiden folgenden Bilder stammen vom Cover von *Perceptrons*. Das obere Bild zeigt eine nicht zusammenhängende Figur (die aus keiner ununterbrochenen schwarzen Linie besteht), während im unteren Bild eine zusammenhängende Figur zu sehen ist (die aus einer ununterbrochenen schwarzen Linie besteht). Ein Mensch kann das ebenso feststellen wie ein einfaches Softwareprogramm. Ein vorwärtsgerichtetes Perzeptron (bei dem die Verbindungen zwischen den Knoten keine Schleifen bilden) wie Rosenblatts Mark 1 Perzeptron kann das nicht erkennen.

Der Grund dafür ist, dass man die XOR-Computerfunktion (exklusives Oder) benötigt, um das Problem zu lösen. Diese Funktion bestimmt, ob eine Linie Teil einer durchgehenden Figur im Bild ist, aber nicht Teil einer anderen Figur. Doch bei einer einzelnen Knotenschicht ohne Rückkopplung ist es mathematisch unmöglich, XOR zu implementieren, weil in einem solchen Netzwerk alle Daten in einem Durchgang anhand einer linearen Regel eingeordnet werden müssen (z. B. »Wenn diese beiden Knoten feuern, ist die Ausgabe-

funktion wahr«). XOR braucht jedoch eine Rückkopplung («Wenn einer dieser beiden Knoten feuert, *aber nicht beide*, ist die Ausgabefunktion wahr«).

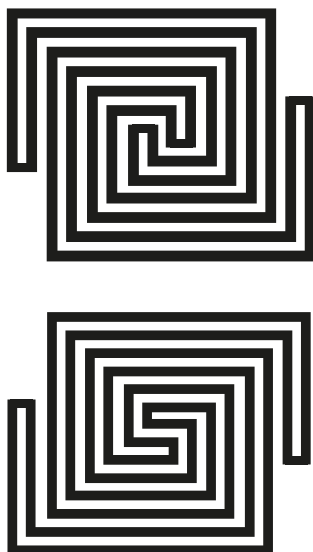


Abbildung 3

Nachdem Minsky und Papert zu diesem Schluss gelangt waren, versiegte die Finanzierung für die Konnektionismusforschung fast vollständig und erhöhte sich erst Jahrzehnte später wieder. Tatsächlich hatte mir Rosenblatt bereits 1964 erklärt, das Perzeptron könne nicht mit Invarianz umgehen, weil es zu wenige Schichten habe. Wenn man die Ausgabe eines Perzeptrons in eine weitere Schicht eingebe, die genauso aussehe, sei die Ausgabe allgemeingültiger. Wenn man diesen Vorgang wiederholte, würden die Ergebnisse bezüglich Invarianzen immer besser werden. Bei ausreichend vielen Schichten und genügend Trainingsdaten könnte ein solches Netzwerk erstaunliche Komplexität erreichen. Ich fragte ihn, ob er das schon ausprobiert habe. Er verneinte, sagte aber, es stehe oben auf seiner Forschungsliste. Das war eine erstaunliche Erkenntnis, aber Rosenblatt starb nur sieben Jahre später, 1971, bevor er die Gelegenheit hatte, seine Ideen

zu überprüfen. Erst zehn Jahre später kamen mehrfache Schichten verbreitet in Gebrauch, und auch dann benötigten mehrschichtige Netzwerke noch mehr Rechenkraft und Trainingsdaten, als verfügbar waren. Die riesigen Fortschritte bei der KI-Entwicklung der letzten Jahre wurden durch den Einsatz multipler neuronaler Netzschichten möglich, die Rosenblatt mehr als fünfzig Jahre zuvor erträumt hatte.

Der konnektionistische Ansatz bei KI wurde bis Mitte der 2010er-Jahre also weitgehend ignoriert. Erst dann erschlossen neue Entwicklungen bei der Hardware sein schlummerndes Potenzial. Jetzt endlich waren ausreichende Mengen an Rechenkraft und Trainingsbeispielen erschwinglich genug, dass diese Methode ihre Stärke beweisen konnte. Zwischen der Veröffentlichung von *Perceptrons* im Jahr 1969 und Minskys Tod 2016 verbesserte sich das Preis-Leistungs-Verhältnis bei der Rechenleistung (inflationsbereinigt) um einen Faktor von circa 2,8 *Milliarden*.²⁸ Dadurch wurden neue Ansätze bei KI möglich. Ich sprach gegen Ende seines Lebens noch einmal mit Minsky, und er bedauerte den großen Einfluss von *Perceptrons*, denn zu dem Zeitpunkt feierte der Konnektionismus erste größere Erfolge.

Der Konnektionismus ist damit ein bisschen wie die Flugmaschinen, die Leonardo da Vinci erfand – eine zukunftsweisende Idee, aber erst umsetzbar, nachdem leichtere und stabilere Materialien entwickelt worden waren.²⁹ Nachdem die Hardware aufgeholt hatte, wurde Konnektionismus im großen Maßstab, etwa in Netzwerken mit 100 Schichten, realisierbar. Solche Systeme konnten dann Probleme lösen, die nie zuvor in Angriff genommen worden waren. Dieses Paradigma steht hinter allen spektakulären Fortschritten der vergangenen Jahre.

DAS KLEINHIRN: EINE MODULARE STRUKTUR

Um neuronale Netze im Kontext menschlicher Intelligenz besser zu verstehen, schlage ich einen kleinen Umweg vor: Gehen wir zurück an den Anfang des Universums. Die anfängliche Entwicklung von Materie hin zu höherer Ordnung verlief sehr langsam, ohne Anleitung durch ein Gehirn. (Siehe den Abschnitt »Die unglaubliche Unwahr-

scheinlichkeit des Seins« in Kapitel 3 zur Frage, wie wahrscheinlich es ist, dass das Universum überhaupt über die Fähigkeit verfügt, nützliche Informationen zu encodieren.) Es dauerte Hunderte Millionen bis Milliarden Jahre, einen neuen Detaillierungsgrad zu erreichen.³⁰

Tatsächlich vergingen Milliarden von Jahren, bevor ein erstes Molekül codierte Anweisungen formulieren konnte, um ein Lebewesen zu erschaffen. Aktuell wird noch diskutiert, wann das Leben auf der Erde begann, aber die meisten Wissenschaftler gehen von einer Zeit vor 3,5 bis 4 Milliarden Jahren aus.³¹ Das Universum selbst ist schätzungsweise 13,8 Milliarden Jahre alt (oder genauer gesagt ist so viel Zeit seit dem Urknall vergangen), und die Erde entstand wahrscheinlich vor etwa 4,5 Milliarden Jahren.³² Zwischen dem ersten Atom im Universum und dem ersten selbstreplizierenden Molekül (auf der Erde) vergingen also etwa 10 Milliarden Jahre. Grund für diese Verzögerung ist möglicherweise der Zufall – wie unwahrscheinlich es war, dass Moleküle in der »Ursuppe« auf der Erde sich zufällig richtig kombinierten, ist nicht bekannt. Vielleicht könnte das Leben früher begonnen haben, vielleicht war es auch wahrscheinlicher, dass es sehr viel später beginnt. Bevor diese notwendigen Bedingungen erfüllt werden konnten, mussten ganze Sterne, die Wasserstoff zu schwereren Elementen fusionierten, ohne die komplexes Leben unmöglich war, geboren werden und sterben.

Wissenschaftler schätzen, dass danach 2,9 Milliarden Jahre zwischen dem ersten Leben und dem ersten mehrzelligen Leben auf der Erde vergingen.³³ Weitere 500 Millionen Jahre verstrichen bis zu den ersten Landtieren, und 200 Millionen Jahre danach tauchten die ersten Säugetiere auf.³⁴ Bei der Entwicklung des Gehirns lagen mehr als 100 Millionen Jahre zwischen den ersten primitiven Nervennetzen und den frühesten zentralisierten, dreiteiligen Gehirnen.³⁵ Der erste einfache Neokortex tauchte erst weitere 350 bis 400 Millionen Jahre später auf, und es dauerte noch weitere 200 Millionen Jahre, bis sich das Gehirn des modernen Menschen entwickelt hatte.³⁶

Hochentwickelte Gehirne waren schon immer ein evolutionärer Vorteil. Beim Wettbewerb um Ressourcen hatten die klügeren Tiere häufig die Nase vorn.³⁷ Intelligenz entwickelte sich in sehr viel kürzerer Zeit als die vorherigen Entwicklungsschritte. Die bemerkens-

werteste Veränderung im Gehirn der Vorfahren der Säugetiere war das sogenannte Kleinhirn. Im Gehirn heutiger Menschen besteht das Kleinhirn aus mehr Neuronen als der Neokortex, der für die höheren Hirnfunktionen wichtig ist.³⁸ Im Kleinhirn können zahlreiche Abläufe für motorische Aufgaben gespeichert und aktiviert werden, für das Ausführen einer Unterschrift zum Beispiel. (Diese Abläufe werden umgangssprachlich auch »Muskelgedächtnis« genannt. Das Phänomen hat aber weniger mit Muskeln als mit dem Kleinhirn zu tun. Wenn eine Bewegung häufig wiederholt wird, passt sich das Gehirn an, um die Bewegung einfacher und unbewusster zu machen – wie bei einem Weg, in den viele Wagenräder eine Furche eingraben.)³⁹

Man kann einen geworfenen Ball fangen, indem man die Flugbahn des Balls und die eigenen Bewegungen anhand vieler Differenzialgleichungen berechnet und den Körper entsprechend dieser Rechenergebnisse in Position bringt. Leider haben wir im Kopf keine Maschine, die Differenzialgleichungen lösen kann, daher lösen wir ein einfacheres Problem: Wie bringe ich den Baseballhandschuh am effektivsten zwischen den Ball und meinen Körper? Das Kleinhirn geht davon aus, dass die Hand und der Ball ähnliche relative Positionen zueinander einnehmen müssen, damit man den Ball fangen kann. Wenn die Flugkurve des Balls also zu steil ist und die Hand sich zu langsam bewegt, wird das Kleinhirn die Hand anweisen, sich schneller an die gewohnte relative Position zu bewegen.

Diese einfachen Handlungen des Kleinhirns, um den sensorischen Input passenden Muskelbewegungen zuzuordnen, entsprechen dem mathematischen Konzept der »Basisfunktionen« und ermöglichen uns, den Ball zu fangen, ohne Differenzialgleichungen lösen zu müssen.⁴⁰ Wir können mit dem Kleinhirn auch überlegen, wie wir uns bewegen würden, selbst wenn wir die Bewegung gar nicht wirklich ausführen. Das Kleinhirn kann uns sagen, dass wir den Ball zwar fangen, dabei aber mit einem anderen Spieler zusammenstoßen könnten, und dass wir es deshalb bleiben lassen sollten. All das geschieht instinktiv.

Beim Tanzen lenkt oft das Kleinhirn unsere Bewegungen, ohne dass wir uns dessen bewusst wären. Menschen, deren Kleinhirnfunktion durch Krankheit oder Verletzung eingeschränkt ist, können willentliche Bewegungen immer noch über den Neokortex steuern.

Allerdings braucht es dazu eine bewusste Anstrengung, und häufig kommt es zu Koordinationsproblemen, der sogenannten Ataxie.⁴¹

Wenn man eine körperliche Tätigkeit meisterlich beherrschen will, muss man die entsprechenden Teilbewegungen so oft ausführen, dass sie sich in das Muskelgedächtnis einprägen. Man verschiebt die Kontrolle über die Bewegungen dadurch vom Motorcortex an das Kleinhirn. Ob man nun Fußball spielt, einen Zauberwürfel löst oder Klavier spielt, je weniger bewusste geistige Anstrengung für die Aufgabe benötigt wird, umso bessere Leistungen kann man bringen. Die Bewegungen werden schneller und flüssiger, und man kann sich auf andere Aspekte des Erfolgs konzentrieren. Wenn Musiker ihr Instrument beherrschen, können sie jede Note so mühelos und intuitiv spielen, wie normale Menschen »Happy Birthday« singen. Wenn ich Sie fragen würde, wie Sie Ihre Stimmbänder dazu veranlassen, die richtige Note hervorzubringen, könnten Sie den Vorgang wahrscheinlich nicht in Worte fassen. Dies bezeichnen Psychologen und Trainer als »unbewusste Kompetenz«, weil das entsprechende Können weitgehend im nicht bewussten Denken abläuft.⁴²

Doch resultiert die Leistungskraft des Kleinhirns nicht aus einer hochkomplexen Architektur. Zwar befindet sich dort der Großteil der Neuronen in einem ausgewachsenen menschlichen Gehirn (und auch jeder anderen Tierart), doch enthält das Genom kaum Informationen über seinen Aufbau – es besteht überwiegend aus kleinen und einfachen Modulen.⁴³ In der Neurowissenschaft wird immer noch dazu geforscht, wie genau das Kleinhirn funktioniert, doch man weiß bereits, dass es aus vielen Tausend kleinen Verarbeitungsmodulen besteht, die in einer vorwärtsgerichteten Struktur angeordnet sind.⁴⁴ Das hilft uns zu erkennen, was neuronale Architekturen benötigen, um die Funktionen des Kleinhirns zu erfüllen. Neue Forschungsergebnisse zum Kleinhirn könnten weitere Erkenntnisse liefern, die für die KI-Entwicklung nützlich sind.

Die meisten Module des Kleinhirns haben eine genau definierte Funktion – die Module, die die Fingerbewegungen beim Klavierspielen steuern, haben nichts mit den Beinbewegungen beim Gehen zu tun. Seit vielen Hundert Millionen Jahren ist das Kleinhirn eine wichtige Gehirnregion, doch die Menschen nutzen es immer weniger