

# Unverkäufliche Leseprobe

Alle Rechte vorbehalten. Die Verwendung von Text und Bildern, auch auszugsweise, ist ohne schriftliche Zustimmung des Verlags urheberrechtswidrig und strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.



Stephen Hawking warnt davor, dass die Weiterentwicklung der Künstlichen Intelligenz das Ende der Menschheit bedeute. Andere hingegen feiern das neue Zeitalter der Superintelligenz, da menschliche Kapazitäten mittels intelligenter Maschinen enorm ausgeweitet würden. Sicher ist, dass KI aus dem Bereich der Science-Fiction Einzug in die Realität, ja in unseren Alltag gefunden hat.

Der bekannte Visionär John Brockman hat daher die führenden Wissenschaftler, Philosophen und Künstler unserer Zeit gefragt, was sie von denkenden Maschinen halten.

Der bekannte Visionär *John Brockman*, ehemaliger Aktionskünstler, Herausgeber der Internetzeitschrift »Edge« und Begründer der »Dritten Kultur« (»Third Culture«), leitet eine Literaturagentur in New York und hat bereits zahlreiche Bücher veröffentlicht, u. a. ›Das Wissen von morgen. Was wir für wahr halten, aber nicht beweisen können: Die führenden Wissenschaftler unserer Zeit beschreiben ihre großen Ideen‹, ›Leben, was ist das? Ursprünge, Phänomene und die Zukunft unserer Wirklichkeit‹, ›Welche Idee wird alles verändern? Die führenden Wissenschaftler unserer Zeit über Entdeckungen, die unsere Zukunft bestimmen werden‹, ›Wie funktioniert die Welt? Die führenden Wissenschaftler unserer Zeit stellen die brilliantesten Theorien vor‹ und ›Welche wissenschaftliche Idee ist reif für den Ruhestand? Die führenden Köpfe unserer Zeit über die Ideen, die uns am Fortschritt hindern‹.

*Weitere Informationen finden Sie auf [www.fischerverlage.de](http://www.fischerverlage.de)*

John Brockman

# Was sollen wir von Künstlicher Intelligenz halten?

Die führenden Wissenschaftler  
unserer Zeit über intelligente Maschinen

Aus dem Englischen von Jürgen Schröder

FISCHER Taschenbuch



Erschienen bei FISCHER Taschenbuch  
Frankfurt am Main, Juli 2017

Die amerikanische Originalausgabe erschien 2015 unter dem Titel  
»What to Think About Machines That Think.  
Today's Leading Thinkers on the Age of Machine Intelligence«  
im Verlag Harper Collins Publishers, New York  
© 2015 by Edge Foundation, Inc.

Für die deutschsprachige Ausgabe:  
© 2017 S. Fischer Verlag GmbH,  
Hedderichstr. 114, D-60596 Frankfurt am Main

Satz: Dörlemann Satz, Lemförde  
Druck und Bindung: CPI books GmbH, Leck  
Printed in Germany  
ISBN 978-3-596-29705-4

# Inhalt

**Danksagung** 23

**Vorwort** von John Brockman: Die *Edge*-Frage 25

Murray Shannahan

**Bewusstsein in der KI auf menschlichem Niveau** 27

Steven Pinker

**Denken bedeutet nicht Unterjochen** 31

Martin Rees

**Die organische Intelligenz hat keine langfristige  
Zukunft** 36

Steve Omohundro

**Ein Wendepunkt in der Künstlichen Intelligenz** 40

Dimitar D. Sasselov

**KI ist I** 44

Frank Tipler

**Wenn du sie nicht unterkriegen kannst,  
schließ' dich ihnen an** 46

Mario Livio

**Intelligente Maschinen auf der Erde und anderswo** 49

Antony Garrett Lisi

**Ich heiße jedenfalls unsere maschinellen Gebiete  
willkommen** 52

8 Inhalt

John Markoff

**Unsere Gebieter, Sklaven oder Partner?** 56

Paul Davies

**Entworfenen Intelligenz** 60

Kevin P. Hand

**Der superintelligente Eigenbrötler** 63

John C. Mather

**Es wird eine stürmische Reise werden** 66

David Christian

**Trägt irgendjemand die Verantwortung  
für dieses Gebilde?** 69

Timo Hannay

**Zeugen des Universums** 72

Max Tegmark

**Machen wir uns bereit!** 75

Tomaso Poggio

**»Turing+«-Fragen** 79

Pamela McCorduck

**Ein epochales menschliches Ereignis** 84

Marcelo Gleiser

**Willkommen zu unserem transhumanen Selbst** 87

Sean Carroll

**Wir sind alle Denkmaschinen** 89

Nicholas G. Carr

**Die Krise der Kontrolle** 92

Jon Kleinberg & Sendhil Mullainathan

**Wir bauen sie zwar, aber wir verstehen sie nicht** 96

Jaan Tallinn

**Wir müssen unsere Hausaufgaben machen** 101

George Church

**Was kümmert es dich, was andere Maschinen denken?** 103

Arnold Trehub

**Maschinen können nicht denken** 107

Roy Baumeister

**Kein »Ich« und keine Fähigkeit zur Heimtücke** 108

Keith Devlin

**Menschliche Intelligenz mit Hebelwirkung** 110

Emanuel Derman

**Eine Maschine ist ein »Materie«-Ding** 114

Freeman Dyson

**Ich könnte mich auch irren** 116

David Gelernter

**Warum lässt sich »Sein« oder »Glück«  
nicht berechnen?** 117

Leo M. Chalupa

**Keine Maschine denkt über die ewigen Fragen nach** 120

Daniel C. Dennett

**Die Singularität – eine moderne Legende?** 123

W. Tecumseh Fitch

**Nanointentionalität** 128

Irene Pepperberg

**Ein schöner (visionärer) Geist** 132

Nicholas Humphrey

**Der Koloss ist ein GFR** 135



10 Inhalt

Rolf Dobelli

**Selbstbewusste KI? Nicht in 1000 Jahren!** 138

Cesar Hidalgo

**Maschinen denken nicht, aber Menschen auch nicht** 143

James J. O'Donnell

**Im Netz der Frage verstrickt** 147

Rodney A. Brooks

**Die Verwechslung von Performanz und Kompetenz** 149

Terrence J. Sejnowski

**Die KI wird uns klüger machen** 153

Seth Lloyd

**Seichtes Lernen** 157

Carlo Rovelli

**Natürliche Geschöpfe einer natürlichen Welt** 160

Frank Wilczek

**Drei Bemerkungen zur Künstlichen Intelligenz** 163

John Naughton

**Wenn ich »Bruno Latour« sage, meine ich nicht  
»Banana till«** 166

Nick Bostrom

**Es ist immer noch früh** 168

Donald D. Hoffman

**Evoluierende KI** 170

Roger Schank

**Denkmaschinen gibt es im Kino** 174

Juan Enriquez

**Kopftransplantate?** 179

Esther Dyson

**KI/KL** 182

Tom Griffiths

**Gehirne und andere Denkmachines** 184

Mark Pagel

**Ihr Nutzen übersteigt ihre Schädlichkeit** 188

Robert Provine

**Denkmachines an der Leine halten** 191

Susan Blackmore

**Der nächste Replikator** 193

Tim O'Reilly

**Was wäre, wenn wir das Mikrobiom der siliziumbasierten KI sind?** 196

Andy Clark

**Man ist, was man isst** 199

Moshe Hoffman

**Das der KI eigene System von Rechten und Regierung** 203

Brian Knutson

**Der Roboter mit Hintergedanken** 206

William Poundstone

**Können U-Boote schwimmen?** 210

Gregory Benford

**Keine Angst vor der KI** 212

Lawrence M. Krauss

**Worüber sollte ich mir Sorgen machen?** 216

Peter Norvig

**Maschinen entwerfen, um die Komplexität der Welt zu bewältigen** 220

Jonathan Gottschall

**Der Aufstieg Geschichten erzählender Maschinen** 225

Michael Shermer

**Man sollte sich eine Protopie, und keine Utopie oder Dystopie vorstellen** 227

Chris Dibona

**Die Grenzen biologischer Intelligenz** 230

Joscha Bach

**Jede Gesellschaft bekommt die KI, die sie verdient** 233

Quentin Hardy

**Die Bestien der KI-Insel** 237

Clifford Pickover

**Wir werden eins werden** 241

Ernst Pöppel

**Eine außerirdische Beobachtung menschlicher Hybris** 244

Ross Anderson

**Wer die KI bezahlt, darf auch bestimmen** 248

W. Daniel Hillis

**Ich denke, also KI** 251

Paul Saffo

**Welchen Platz werden die Menschen einnehmen?** 253

Dylan Evans

**Der große KI-Schwindel** 256

- Anthony Aguirre  
**Die Wahrscheinlichkeiten der KI** 259
- Eric J. Topol  
**Eine neue Weisheit des Körpers** 263
- Roger Highfield  
**Von herkömmlicher Intelligenz zur KI** 266
- Gordon Kane  
**Wir brauchen mehr als Denken** 268
- Scott Atran  
**Gehen wir in die falsche Richtung?** 270
- Stanislas Dehaene  
**Zwei kognitive Funktionen, die Maschinen immer noch fehlen** 274
- Matt Ridley  
**Zwischen Maschinen, nicht in ihnen** 278
- Stephen M. Kosslyn  
**Eine andere Art von Vielfalt** 280
- Luca Di Biase  
**Erzählungen und unsere Kultur** 283
- Margaret Levi  
**Menschliche Verantwortung** 287
- D. A. Wallach  
**Verstärker/Implementationen menschlicher Entscheidungen** 290
- Rory Sutherland  
**Mach es unmöglich, es zu hassen** 293

Bruce Sterling

**Schauspielerinnen-Maschinen** 296

Kevin Kelly

**Nennen wir sie künstliche Außerirdische** 299

Martin Seligman

**Handeln Maschinen?** 302

Timothy Taylor

**Denkraumverlust** 305

George Dyson

**Das Analoge: die Revolution, die ihren Namen  
nicht auszusprechen wagt** 309

S. Abbas Raza

**Die Werte der Künstlichen Intelligenz** 311

Bruce Parker

**Künstliche Selektion und unsere Enkelkinder** 314

Neil Gershenfeld

**Wirklich gute Programmiertricks** 318

Daniel L. Everett

**Der Airbus und der Adler** 321

Douglas Coupland

**Menschartigkeit** 324

Josh Bongard

**Manipulateure und Manipulanda** 327

Ziyad Marar

**Denken wir mehr wie Maschinen?** 331

Brian Eno

**Einfach nur ein fraktales Detail im großen Ganzen** 335

Marti Hearst

**eGaia, ein verteiltes, technisch-soziales,  
geistiges System** 338

Chris Anderson

**Der Bienenstockgeist** 340

Alex (Sandy) Pentland

**Die globale Künstliche Intelligenz ist schon da** 343

Randolph Nesse

**Werden Computer sich zu denkenden,  
sprechenden Hunden entwickeln?** 347

Richard E. Nisbett

**Denkmaschinen und Langeweile** 350

Samuel Arbesman

**Naches gegenüber unseren Maschinen** 354

Gerald Smallberg

**Keine gemeinsame Theorie des Geistes** 357

Eldar Shafir

**Blind für das Zentrum menschlicher Erfahrung** 361

Christopher Chabris

**Eine intuitive Maschinentheorie** 364

Ursula Martin

**Denkende Salzsümpfe** 367

Kurt Gray

**Killer-Denkmaschinen halten unser Gewissen rein** 370

Bruce Schneier

**Wenn Denkmaschinen das Gesetz brechen** 374

Rebecca Mackinnon  
**Elektrogehirne** 378

Gerd Gigerenzer  
**Roboärzte** 382

Alison Gopnik  
**Können Maschinen je so klug sein wie Dreijährige?** 386

Kevin Slavin  
**Tic-Tac-Toe-Hühner** 390

Alun Anderson  
**Die KI wird uns klug und Roboter werden uns  
Angst machen** 393

Mary Catherine Bateson  
**Wenn Denkmachines kein Segen sind** 396

Steve Fuller  
**Gerechtigkeit für Maschinen in einer  
organizistischen Welt** 398

Tania Lombrozo  
**Im Hinblick auf das Denken sollte man nicht  
chauvinistisch sein** 402

Virginia Heffernan  
**Das klingt himmlisch** 405

Barbara Strauch  
**Maschinen, die funktionieren,  
bis sie es eben nicht mehr tun** 406

Sheizaf Rafaeli  
**Die beweglichen Torpfosten** 408

Edward Slingerland  
**Richtungslose Intelligenz** 411

- Nicholas A. Christakis  
**Die menschliche Kultur als erste KI** 413
- Joichi Ito  
**Jenseits des unheimlichen Tals** 416
- Douglas Rushkoff  
**Figur oder Hintergrund?** 420
- Helen Fisher  
**Schnell, präzise und dumm** 422
- Stuart Russell  
**Werden sie uns zu besseren Menschen machen?** 425
- Eliezer S. Yudkowsky  
**Das Wertladeproblem** 429
- Kate Jeffery  
**Nach unserem Bilde** 433
- Maria Popova  
**Die Umwelt des Unbeantwortbaren** 438
- Jessica L. Tracy & Kristin Laurin  
**Werden sie über sich selbst nachdenken?** 440
- June Gruber & Raul Saucedo  
**Organisches versus künstliches Denken** 444
- Paul Dolan  
**Es kommt gewiss auf den Kontext an** 447
- Thomas G. Dietterich  
**Wie man eine Intelligenzexplosion verhindert** 449
- Matthew D. Lieberman  
**Denken von innen oder von außen?** 453



Michael Vassar

**Sanfter Autoritarismus** 458

Gregory Paul

**Was werden künstliche Intelligenzen über uns denken?** 461

Andrian Kreye

**Ein John-Henry-Moment** 464

N. J. Enfield

**Maschinen kennen sich mit Beziehungen nicht aus** 467

Nina Jablonski

**Die nächste Phase der Evolution des Menschen** 469

Gary Klein

**Herrschaft versus Domestizierung** 472

Gary Marcus

**Maschinen werden in nächster Zeit nicht denken** 475

Sam Harris

**Können wir eine digitale Apokalypse vermeiden?** 478

Molly Crockett

**Könnten Denkmaschinen die Empathielücke schließen?** 482

Abigail Marsh

**Fürsorgliche Maschinen** 485

Alexander Wissner-Gross

**Motoren der Freiheit** 488

Sarah Demers

**Noch Fragen?** 492

Bart Kosko

**Denkmaschinen = alte Algorithmen auf  
schnelleren Computern** 494

Julia Clarke

**Die Nachteile von Metaphern** 498

Michael McCullough

**Eine universale Grundlage für Menschenwürde** 501

Haim Harari

**Über Menschen nachdenken,  
die wie Maschinen denken** 505

Hans Halvorson

**Metadenken** 509

Christine Finn

**Der Wert der Antizipation** 512

Dirk Helbing

**Ein Ökosystem von Ideen** 515

John Tooby

**Das eiserne Gesetz der Intelligenz** 517

Maximilian Schich

**Gedanken stehlende Maschinen** 522

Satyajit Das

**Unbeabsichtigte Folgen** 525

Robert Sapolsky

**Es kommt darauf an** 529

Athena Vouloumanos

**Werden Maschinen unser Denken für uns erledigen?** 530

Brian Christian

**Tut mir leid, wenn ich Sie störe** 532

Benjamin K. Bergen

**Moralische Maschinen** 534

Laurence C. Smith

**Nachdem der Stecker gezogen wurde** 536

Guilio Boccaletti

**Die Erde überwachen und verwalten** 538

Ian Bogost

**Panexperientialismus** 541

Aubrey De Grey

**Wann ist ein Untergebener kein Untergebener?** 545

Michael I. Norton

**Nicht fehlerhaft genug** 549

Thomas A. Bass

**Mehr Funk, mehr Soul, mehr Poesie und Kunst** 551

Hans Ulrich Obrist

**Die Zukunft ist uns versperrt** 553

Koo Jeong-A

**Eine nichtmaterielle denkbare Maschine** 556

Richard Foreman

**Verdutzt und besessen** 557

Richard H. Thaler

**Wer hat Angst vor Künstlicher Intelligenz?** 560

Scott Draves

**Ich sehe die Entwicklung einer Symbiose** 564

- Matthew Ritchie  
**Die Neukonzeption des Selbst in einer verteilten Welt** 567
- Raphael Bousso  
**Es ist leicht, die Zukunft vorherzusagen** 572
- James Croak  
**Die wiederbelebte Furcht vor einem Gott** 575
- Andrés Roemer  
**Tulpen auf das Grab meines Roboters** 577
- Lee Smolin  
**Für eine naturalistische Theorie des Geistes** 580
- Stuart A. Kauffman  
**Denkende Maschinen? Quatsch!** 584
- Melanie Swan  
**Der zukünftige Möglichkeitsraum der Intelligenz** 587
- Tor Nørretranders  
**Liebe** 591
- Kai Krause  
**Ein frappierender Drei-Ringe-Test für *Machina sapiens*** 595
- Georg Diez  
**Frei von uns** 600
- Eduardo Salcedo-Albarán  
**Makellose KI scheint Science-Fiction zu sein** 603
- Maria Spiropulu  
**Emergente hybride Mensch/Maschinen-Chimären** 606
- Thomas Metzinger  
**Was ist, wenn sie leiden können müssen?** 609

Beatrice Golomb

**Werden wir es erkennen, wenn es geschieht?** 614

Noga Arikha

**Metarepräsentation** 618

Demis Hassanis, Shane Legg & Mustafa Suleyman

**Envoi: eine kurze Strecke vor uns –  
und noch viel zu tun** 620

Murray Shannahan  
**Bewusstsein in der KI auf  
menschlichem Niveau**

Professor für kognitive Robotik am Imperial College, London;  
Autor von *Embodiment and the Inner Life*

Nehmen wir einmal an, wir könnten eine Maschine mit einer Intelligenz menschlichen Niveaus ausstatten, das heißt mit der Fähigkeit, einem typischen Menschen in jedem (oder fast jedem) Bereich intellektueller Bemühungen gleichzukommen und vielleicht jeden Menschen in einigen Bereichen zu übertreffen. Hätte eine solche Maschine zwangsläufig Bewusstsein? Das ist eine wichtige Frage, weil eine positive Antwort uns stutzen lassen würde. Wie würden wir mit so etwas umgehen, wenn wir es konstruierten? Wäre es in der Lage, Leid oder Freude zu empfinden? Würde es dieselben Rechte wie ein Mensch verdienen? Sollten wir überhaupt Maschinenbewusstsein erzeugen?

Die Frage, ob eine KI auf Menschenniveau zwangsläufig Bewusstsein hätte, ist außerdem schwierig. Eine Quelle der Schwierigkeit ist die Tatsache, dass mit Bewusstsein bei Menschen und anderen Tieren zahlreiche Attribute verbunden sind. Alle Tiere weisen ein Gefühl für Zwecke auf. Alle (wachen) Tiere sind sich der Welt, in der sie leben und der darin enthaltenden Gegenstände mehr oder weniger bewusst. Alle Tiere zeigen in einem gewissen Grad kognitive Integration, was bedeutet, dass sie alle ihre geistigen Ressourcen – Wahrnehmungen, Erinnerungen und Fertigkeiten – zur Verfolgung ihrer Ziele in der aktuellen Situation in Anschlag bringen können. In dieser Hinsicht weist jedes Tier eine gewisse Art von

Einheit, eine Art von Selbstsein auf. Manche Tiere, einschließlich des Menschen, sind sich auch ihrer selbst bewusst – ihrer Körper und ihres Gedankenflusses. Schließlich sind die meisten, wenn nicht alle Tiere fähig zu leiden, und manche sind in der Lage, Mitgefühl mit dem Leiden anderer zu haben.

Bei (gesunden) Menschen kommen alle diese Attribute als Paket zusammen. Aber in einer KI können sie möglicherweise voneinander getrennt werden. Unsere Frage muss daher verfeinert werden. Welches der Attribute, die wir bei Menschen mit Bewusstsein verbinden, ist, wenn überhaupt, eine notwendige Begleiterscheinung von Intelligenz auf Menschenniveau? Nun, jedes der aufgeführten Attribute (und die Liste ist gewiss nicht erschöpfend) verdient jeweils längere Ausführungen. Ich möchte daher nur zwei herausgreifen – nämlich Bewusstsein der Welt und die Fähigkeit zu leiden. Bewusstsein der Welt, würde ich behaupten, ist tatsächlich eine notwendige Eigenschaft von Intelligenz menschlichen Niveaus.

Sicherlich würde man nicht eine Intelligenz auf menschlichem Niveau zuschreiben, wenn es keine Sprache hätte, und die Hauptverwendung menschlicher Sprache besteht im Sprechen über die Welt. In diesem Sinne ist Intelligenz unzertrennlich mit dem verknüpft, was Philosophen *Intentionalität* nennen. Darüber hinaus ist die Sprache ein soziales Phänomen, und eine grundlegende Verwendung von Sprache innerhalb einer Menschengruppe besteht im Sprechen über Dinge, die alle wahrnehmen können (wie beispielsweise dieses Werkzeug oder jenes Stück Holz), oder wahrgenommen haben (das Stück Holz von gestern) oder wahrnehmen könnten (vielleicht das Stück Holz von morgen). Kurz, die Sprache gründet im Bewusstsein der Welt. Bei einem verkörperten Wesen oder einem Roboter wäre ein solches Bewusstsein aufgrund seiner Interaktionen mit der Umgebung offenbar (Hindernisse vermeiden, Dinge aufnehmen usw.). Aber wir könnten den Begriff auch erweitern, so dass auch eine verteilte, entkörperlichte Künstliche Intelligenz, die mit geeigneten Sensoren ausgestattet ist, darunter fällt.

Um auf überzeugende Weise als eine Facette des Bewusstseins zu gelten, müsste diese Art von Bewusstsein vielleicht Hand in Hand mit einem offenbaren Sinn für Zwecke und einem gewissen Grad kognitiver Integration gehen. Dieses Trio von Attributen wird daher auch bei einer KI als Paket vorhanden sein. Aber lassen wir die Frage einen Augenblick beiseite und kehren wir zur Fähigkeit zurück, Leid und Freude zu empfinden. Im Unterschied zum Bewusstsein der Welt gibt es keinen offensichtlichen Grund für die Annahme, dass Intelligenz menschlichen Niveaus dieses Attribut haben muss, auch wenn es bei Menschen innig mit Bewusstsein verbunden ist. Wir können uns eine Maschine vorstellen, die kalt und gefühllos das ganze Spektrum von Aufgaben ausführt, die beim Menschen Verstand erfordern. Einer solchen Maschine würde dasjenige Attribut des Bewusstseins fehlen, das am meisten zählt, wenn es um das Zugeständnis von Rechten geht. Wie Jeremy Bentham bei der Frage bemerkte, wie man nichtmenschliche Tiere behandeln sollte, geht es dabei nicht darum, ob sie rasonieren oder sprechen, sondern ob sie leiden können.

Ich lege hier nicht nahe, dass eine »bloße« Maschine nie fähig sein könnte, Leid oder Freude zu empfinden – dass es in dieser Hinsicht eine biologische Besonderheit gibt. Der Punkt ist vielmehr, dass die Fähigkeit des Empfindens von Leid und Freude von anderen psychologischen Attributen abgekoppelt werden kann, die im menschlichen Bewusstsein zusammengebündelt sind. Aber untersuchen wir diese scheinbare Abkoppelung genauer. Ich stellte bereits die Vorstellung zur Debatte, dass Bewusstsein der Welt mit einem offenkundigen Sinn für Zwecke Hand in Hand gehen könnte. Das Bewusstsein, das ein Tier von der Welt hat, von dem, was ihm die Welt an Gutem oder Schlechtem bietet (in der Ausdrucksweise J. J. Gibsons), dient seinen Zwecken. Ein Tier zeigt das Bewusstsein eines Raubtiers, indem es sich von ihm wegbewegt, und das Bewusstsein einer potentiellen Beute, indem es sich auf sie zubewegt. Vor dem Hintergrund einer Reihe von Zielen und Bedürfnissen ergibt das Verhalten des Tieres einen Sinn. Und



mit Bezug auf einen solchen Hintergrund kann man einem Tier einen Strich durch die Rechnung machen, können seine Ziele unerreicht und seine Bedürfnisse unbefriedigt bleiben. Das ist gewiss die Grundlage für einen Aspekt von Leid.

Wie steht es nun mit Künstlicher Intelligenz menschlichen Niveaus? Würde eine solche KI notwendigerweise eine komplexe Menge von Zielen haben? Könnten ihre Versuche der Zielerreichung nicht vereitelt und an jeder Ecke hintertrieben werden? Wäre es unter diesen scharfen Bedingungen angemessen zu sagen, dass die KI litte, auch wenn ihre Konstitution sie vielleicht immun gegen die Art von Schmerz oder körperlichem Unbehagen machen würde, mit der Menschen vertraut sind?

Hier stößt die Kombination von Einbildungskraft und Intuition an ihre Grenzen. Ich vermute, dass wir nicht herausfinden werden, wie diese Frage zu beantworten sei, bevor wir es nicht mit der echten Sache zu tun haben. Erst wenn eine raffiniertere KI zu einem vertrauten Teil unseres Lebens geworden ist, werden sich unsere Sprachspiele solchen fremdartigen Wesen anpassen. Aber natürlich könnte es bis dorthin zu spät sein, um unsere Meinung darüber zu ändern, ob sie verwirklicht werden sollen. Sie werden wohl oder übel schon da sein.

Steven Pinker

## Denken bedeutet nicht Unterjochen

Johnstone-Family-Professor, Abteilung für Psychologie an der Harvard University; Autor von *The Sense of Style: The Thinking Person's Guide to Writing in the Twenty-First Century*

Thomas Hobbes' markige Gleichsetzung von Denken mit »nichts als Rechnen« ist eine der großartigen Ideen der Menschheitsgeschichte. Die Vorstellung, dass Rationalität durch den physischen Prozess des Rechnens erreicht werden kann, wurde im 20. Jahrhundert durch Alan Turings These bestätigt, dass einfache Maschinen jede berechenbare Funktion implementieren können, und durch Modelle von D. O. Hebb, Warren McCulloch und Walter Pitts und ihren wissenschaftlichen Erben, die zeigten, dass Netzwerke aus vereinfachten Neuronen vergleichbare Leistungen erzielen können. Die kognitiven Leistungen des Gehirns lassen sich in physischen Begriffen erklären: Um es grob (und Kritikern ungeachtet) zu formulieren, können wir sagen, dass Überzeugungen eine Art von Informationen sind, das Denken eine Art von Berechnung und Motivation eine Art von Rückkoppelung und Kontrolle.

Aus zwei Gründen ist das eine großartige Idee. Erstens vervollständigt sie ein naturalistisches Verständnis des Universums, indem sie geheimnisvolle Seelen, Geister und Gespenster in der Maschine austreibt. Ebenso wie Darwin es einem nachdenklichen Beobachter der natürlichen Welt ermöglichte, ohne Kreationismus auszukommen, ermöglichten es Turing und andere einem nachdenklichen Beobachter der kognitiven Welt, ohne Spiritualismus auszukommen.

Zweitens öffnet die komputationale Theorie der Vernunft der Künstlichen Intelligenz die Tür – den Denkmaschinen. Ein von Menschen geschaffener Informationsprozessor könnte die Kräfte des menschlichen Geistes im Prinzip kopieren und übertreffen. Nicht, dass das wahrscheinlich auch in der Praxis geschehen wird, da wir wahrscheinlich nie die anhaltende technische und wirtschaftliche Motivation erleben werden, die dafür notwendig ist. Ebenso wie die Erfindung des Autos keine Verdopplung des Pferdes erforderte, würde die Entwicklung eines KI-Systems, das sich selbst unterhalten könnte, keine Verdopplung eines Exemplars von *Homo sapiens* erfordern. Ein Gerät, das dazu entworfen wurde, Auto zu fahren oder eine Epidemie vorherzusagen, muss nicht dazu entworfen sein, einen Paarungspartner anzuziehen oder verfaultes Aas zu meiden.

Dennoch haben vor kurzem erzielte Minischritte in Richtung auf intelligentere Maschinen zu einer Wiederbelebung der wiederkehrenden Angst geführt, dass unsere Erkenntnis für unseren Untergang verantwortlich sein wird. Meine eigene Ansicht ist, dass gegenwärtige Ängste mit Bezug auf amoklaufende Computer eine Verschwendung emotionaler Energie sind – dass das Szenario mehr Ähnlichkeit mit dem Jahr-2000-Problem als mit dem Manhattan-Projekt hat.

Denn zum einen steht uns eine lange Zeit zur Verfügung, um uns darauf vorzubereiten. KI auf menschlichem Niveau ist immer noch der fünfzehn bis zwanzig Jahre entfernte Standard, so wie sie es immer gewesen ist, und viele ihrer in jüngster Zeit angepriesenen Fortschritte haben flachgründige Wurzeln. Es stimmt zwar, dass »Experten« in der Vergangenheit auf groteske Weise die Möglichkeit technischer Fortschritte, die sich schnell vollzogen, von der Hand gewiesen haben. Aber das ist zweischneidig: »Experten« haben auch unmittelbar bevorstehende Fortschritte angekündigt (oder sind darüber in Panik geraten), die nie eintraten, wie zum Beispiel mit Kernenergie betriebene Autos, Unterwasserstädte, Kolonien auf dem Mars, Designerbabys und Lagerhäuser mit Zombies, die

am Leben erhalten werden, um Menschen mit Ersatzorganen zu versorgen.

Außerdem ist es absonderlich zu meinen, dass die Robotik-Ingenieure bei ihrer Arbeit keine Schutzmaßnahmen gegen Unheil einbauen würden. Dafür würden sie keine schwergewichtigen »Regeln der Robotik« oder irgendeine neumodische Moralphilosophie brauchen, sondern nur denselben gesunden Menschenverstand, der in den Entwurf von Nahrungsprozessoren, Tischsägen, Heizgeräten und Autos einfluss. Die Sorge, dass ein KI-System so klug mit Bezug auf das Erreichen eines seiner programmierten Ziele (wie die Beschaffung von Energie) werden würde, dass es die anderen (wie menschliche Sicherheit) rücksichtslos überginge, nimmt an, dass die KI schneller über uns hereinbrechen wird, als wir ausfallsichere Vorsichtsmaßnahmen entwerfen können. Die Wirklichkeit ist, dass der Fortschritt in der KI dem Medienrummel zum Trotz langsam ist, und es wird viel Zeit für Feedback von schrittweisen Implementationen geben, wobei die Menschen in jedem Stadium den Schraubenzieher in der Hand halten.

Würde ein künstlich intelligentes System diese Schutzmaßnahmen *vorsätzlich* deaktivieren? Warum würde es das wollen? Die Dystopien der KI projizieren eine engstirnige Alpha-Männchen-Psychologie auf den Begriff der Intelligenz. Sie nehmen an, dass übermenschlich intelligente Roboter Ziele entwickeln würden wie die Absetzung ihrer Herren oder die Übernahme der Weltmacht. Aber Intelligenz ist die Fähigkeit, neue Mittel zu ersinnen, um ein Ziel zu erreichen; die Ziele sind der Intelligenz selbst äußerlich. Klug zu sein ist nicht dasselbe wie etwas zu wollen. Die Geschichte wartet mit dem gelegentlichen größtenwahnsinnigen Despoten oder dem psychopathischen Serienmörder auf, aber diese sind die Erzeugnisse einer Geschichte natürlicher Selektion, die testosteronsensitive Schaltkreise bei einer bestimmten Primatenspezies formt, und kein unvermeidliches Merkmal intelligenter Systeme. Es ist bezeichnend, dass viele unserer Technopropheten gar nicht die Möglichkeit erwägen, dass die KI sich entlang

weiblicher Vorgaben entwickeln wird – indem sie völlig in der Lage ist, Probleme zu lösen, aber ohne den Wunsch, Unschuldige zu vernichten oder die Kultur zu beherrschen.

Wir können uns einen maliziösen *Menschen* vorstellen, der ein Bataillon von Robotern entwirft und freisetzt, um Massenvernichtung zu säen. Aber Katastrophenszenarios lassen sich in der Vorstellung billig durchspielen, und wir sollten die Kette von Wahrscheinlichkeiten im Auge behalten, die sich entfalten müsste, bevor dieses Szenario Wirklichkeit würde. Es müsste ein böser Geist aufsteigen, der sowohl einen Hunger nach sinnlosem Massenmord als auch Brillanz in technischen Erfindungen besitzt. Er müsste ein Team von Mitverschwörern rekrutieren und verwalten, die vollkommene Geheimhaltung, Loyalität und Kompetenz ausübten. Und die Operation müsste die Gefahren des Entdecktwerdens, des Verrats, von Sting-Operationen, grober Fehler und Missgeschicke überleben. Theoretisch könnte das passieren, aber es gibt dringlichere Dinge, über die wir uns Sorgen machen sollten.

Wenn wir die Science-Fiction-Katastrophenszenarien beiseitelegen, ist die Möglichkeit einer hochentwickelten Künstlichen Intelligenz erhebend – nicht nur im Hinblick auf die praktischen Vorteile, wie beispielsweise die phantastischen Gewinne an Sicherheit, Freizeit und Umweltfreundlichkeit selbstgesteuerter Autos, sondern auch im Hinblick auf die philosophischen Möglichkeiten. Die komputationale Theorie des Geistes hat niemals die Existenz von Bewusstsein im Sinne der Subjektivität in der ersten Person erklärt (obwohl sie vollkommen in der Lage ist, die Existenz von Bewusstsein im Sinne von zugänglicher und berichtbarer Information zu erklären). Ein Vorschlag lautet, dass Subjektivität von Natur aus jedem hinreichend komplizierten kybernetischen System zukommt. Ich pflegte zu meinen, dass diese Hypothese (ebenso wie ihre Alternativen) sich beständig nicht überprüfen ließe. Aber stellen wir uns einen intelligenten Roboter vor, der darauf programmiert ist, seine eigenen Systeme zu überwa-

chen und wissenschaftliche Fragen zu stellen. Wenn er, ohne dass er dazu aufgefordert wird, die Frage stellen würde, warum er selbst subjektive Erlebnisse hat, würde ich den Gedanken ernst nehmen.