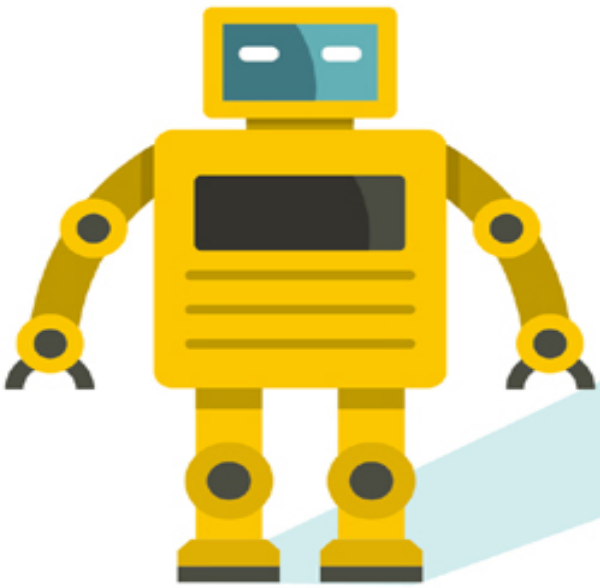
REDLINE | VERLAG

Mo Gawdat

SCARY SMART



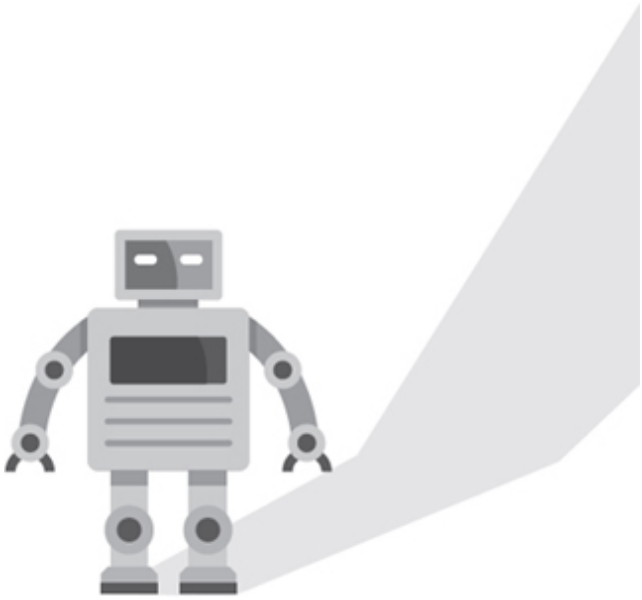
Die Zukunft der künstlichen Intelligenz und wie wir mit ihrer Hilfe unseren Planeten retten

Mo Gawdat

SCARY SMART

Mo Gawdat

SCARY SMART



Die Zukunft der kunstlichen Intelligenz und wie wir mit ihrer Hilfe unseren Planeten retten

REDLINE | VERLAG

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de** abrufbar.

Für Fragen und Anregungen:
info@redline-verlag.de

1. Auflage 2022

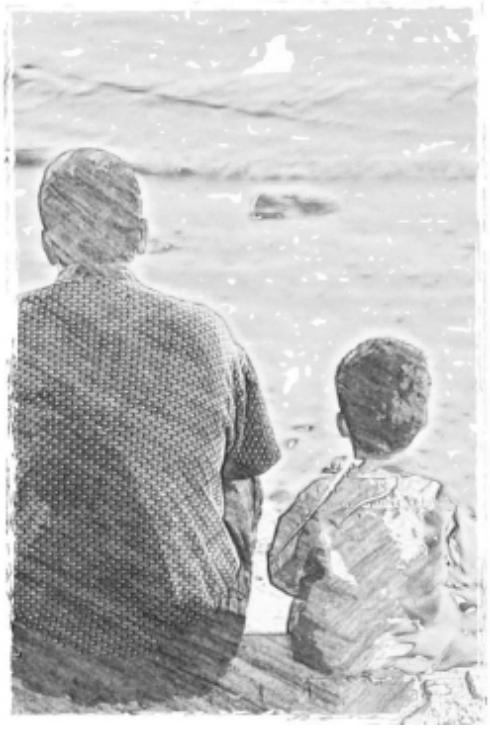
Die englische Originalausgabe erschien 2021 bei Pan Macmillan unter dem Titel Scary Smart: *The Future of Artificial Intelligence and How You Can Save Our World.*

Übersetzung: Jordan Wegberg
Redaktion: Marijke Leege-Topp
Umschlaggestaltung: Karina Braun
Umschlagabbildung: Cyborg icon set/Shutterstock
Satz: ZeroSoft, Timisoara

Druck: GGP Media GmbH, Pößneck
eBook by tool-e-byte

ISBN Print 978-3-86881-893-2
ISBN E-Book (PDF) 978-3-96267-433-5
ISBN E-Book (EPUB, Mobi) 978-3-96267-434-2

Der Ernst des Kampfes hat keine Bedeutung für die Friedvollen.



Für Ali
Jetzt oder nie
Du und ich

Inhalt

Einleitung: Der neue Superheld

Teil 1: Der gruselige Teil

Kapitel 1: Eine kurze Geschichte der Intelligenz

Kapitel 2: Eine kurze Geschichte unserer Zukunft

Kapitel 3: Die drei Unvermeidbarkeiten

Kapitel 4: Eine sanfte Dystopie

Kapitel 5: Unter Kontrolle

Zusammenfassung des gruseligen Teils

Teil 2: Unser Weg zur Utopie

Kapitel 6: Und dann lernten sie

Kapitel 7: Unsere Zukunft großziehen

Kapitel 8: Die Zukunft der Ethik

Kapitel 9: Heute habe ich die Welt gerettet

Zusammenfassung des klugen Teils

Die Allgemeine Erklärung der Weltrechte

Nachwort: Der Kuchen ist eine Lüge

Über den Autor

Quellen

Einleitung:
Der neue Superheld

Dieses Buch ist ein Weckruf. Es richtet sich an Sie und an mich und an jeden, der noch nichts von der bevorstehenden Pandemie weiß - von der baldigen Ankunft der künstlichen Intelligenz. Dieses Buch wird von Experten kritisiert werden und genau deshalb schreibe ich es. Denn um ein Experte für künstliche Intelligenz zu werden, muss man eine spezialisierte, verengte Sicht darauf haben. Diese spezialisierte Sicht der KI lässt vollständig die existenziellen Aspekte außer Acht, die über die Technologie hinausreichen: Fragen der Moral, der Ethik, der Emotionen, des Mitgefühls und einer ganzen Reihe von Vorstellungen, die Philosophen, Weisheitssucher, Vertreter des Humanitätsgedankens, Umweltschützer und im weiteren Sinne den normalen Menschen betreffen (also jeden Einzelnen von uns). Daneben ist die zentrale Absicht dieses Buches, Ihnen zu verdeutlichen, dass es *nicht* die Experten sind, die jener Bedrohung der Menschheit infolge der Entwicklung von Superintelligenz etwas entgegensetzen können. Nein, es sind Sie und ich, die diese Macht besitzen. Und was noch

wichtiger ist: Es sind Sie und ich, die dafür die Verantwortung tragen.

Bei Erscheinen dieses Buches werden wir gerade fast zwei Jahre mit der Corona-Pandemie hinter uns gelassen haben. Wir werden optimistisch sein, dass die Impfungen zu wirken beginnen und dass es eine Chance für uns gibt, wieder zu unserer vorherigen normalen Lebensweise zurückzukehren. Doch »normal« wird für immer verändert sein. Ich glaube, die Art und Weise, wie unsere Weltgemeinschaft und die politische Führung den Ausbruch von Covid-19 gehandhabt haben, ist kaum anders als die Art und Weise, wie sie den baldigen Ausbruch der KI-Pandemie handhaben werden. Ich hoffe nur, wir können aus den Fehlern lernen, die wir bei Corona gemacht haben, und vielleicht mit diesem neuen Wandel unseres Lebens so umgehen, dass weniger Umbrüche, mehr Vorhersagbarkeit und weniger soziale und wirtschaftliche Not damit einhergehen.

Lassen Sie sich bitte nicht von der Schlichtheit beirren, mit der ich dieses Buch zu schreiben versucht habe. Die Fakten, mit denen ich meine Behauptungen hier stütze, sind unwiderlegbar. Sie werden durch meine lange Karriere aus mehr als 30 Jahren in der Technologiebranche gespeist. Vor meinem derzeitigen Start-up (das einige der höchstentwickelten Systeme, Roboter, künstlichen Intelligenzen und Technologien des maschinellen Lernens auf eine Weise einsetzt, die möglicherweise unsere Welt retten könnte) war einer der Höhepunkte meiner

Berufslaufbahn eine zwölfjährige Anstellung bei Google. Dort durfte ich in fast der Hälfte der weltweiten Google-Büros die Einführung von Operationen und Technologien leiten, die über 100 Sprachen umfassten. Meine Zeit dort ging zu Ende, als ich die Position des Chief Business Officer von Google [X] übernahm, jenes berüchtigten Innovationszweigs von Google, der einige der KI-Entwicklungsprojekte wie selbstfahrende Autos, Google Brain und die Mehrzahl der Robotik-Innovationen des Unternehmens hervorbrachte.

Meine Einblicke in den eigentlichen Kern der KI-Entwicklungen, die uns dorthin geführt haben, wo wir heute sind, teilweise aus meiner Zeit bei Google [X] stammend, sind einzigartig. Ich koppele meine direkten Erfahrungen mit der KI-Entwicklung mit meiner Arbeit im Bereich der Glücksforschung (dokumentiert in meinem internationalen Bestseller *Die Formel für Glück,* einem sehr erfolgreichen Podcast namens *Slo Mo* und der von mir gegründeten Non-Profit-Organisation OneBillionHappy.org), um Ihnen eine einzigartige Perspektive auf die Herausforderungen zu verschaffen, denen wir im Zeitalter des Aufstiegs der Superintelligenz gegenüberstehen. Meine Hoffnung ist, dass wir gemeinsam mit der KI eine Utopie erschaffen können, die der Menschheit dient, anstelle einer Dystopie, die sie zersetzt. In diesem Buch erläutere ich, dass jeder von uns - auch Sie und ich - die Verantwortung dafür trägt, eine leuchtendere Zukunft für uns alle zu schaffen. Machen Sie

sich bitte keine Sorgen. Dies ist keine aus Ängsten entstandene Science-Fiction-Geschichte, sondern vielmehr die Geschichte einer der größten Chancen der Menschheit. Dies ist die Chance, die exzessive Abhängigkeit von Konsum und technologischem Fortschritt abzuwenden, die zwar unsere Lebensqualität verbessert haben mag, aber auf Kosten jedes anderen Lebewesens auf der Erde. Nur wenn wir - Sie und ich - die Verantwortung übernehmen und uns verändern, wird dies eine Geschichte der Hoffnung sein.

Mitten im Nichts

Zu Beginn stellen Sie sich bitte vor, wie Sie und eine gebrechliche alte Version von mir im Jahr 2055 an einem Lagerfeuer sitzen, genau 99 Jahre nach dem Beginn der Geschichte der künstlichen Intelligenz im Sommer 1956 am Dartmouth College in New Hampshire. Ich erzähle Ihnen die Geschichte dessen, was ich während der Jahre des Aufstiegs der KI erlebt habe - eine Geschichte, die dazu geführt hat, dass wir beide jetzt hier mitten im Nichts sitzen. Doch erst am Ende des Buches werde ich Ihnen verraten, ob wir dort sitzen, weil wir uns verstecken, um den Maschinen zu entkommen, oder ob wir da sind, weil die KI uns von unseren alltäglichen Arbeitspflichten befreit und uns die Zeit, die Sicherheit und die Freiheit gegeben hat, einfach die Natur zu genießen und das zu

tun, was Menschen am besten können - sich miteinander zu verbinden und nachzudenken.

Ich sage es Ihnen jetzt noch nicht, weil ich in diesem Augenblick einfach noch nicht weiß, wie unsere Geschichte mit den Maschinen ausgeht. Das, mein Freund, liegt an Ihnen. Ja, an Ihnen als Individuum. Nicht an Ihrer Regierung, Ihrem Chef oder den Vordenkern, denen Sie folgen. Die Zukunft liegt tatsächlich in Ihren Händen. Sie hängt von dem ab, was Sie in den nächsten zehn Jahren zu tun entscheiden, von heute angefangen.

Dies ist eine Prophezeiung des Kommenden. Ich habe genau hingeschaut im Laufe der Jahre, die ich mit den neuesten Technologien zugebracht habe und in denen wir Maschinen geschaffen haben, die klüger sind als wir. Ich persönlich habe zum Aufstieg der künstlichen Intelligenz beigetragen. Ich glaubte an das Versprechen, dass die Technik immer für eine Verbesserung unseres Lebens sorgen würde - so lange, bis ich es nicht mehr tat. Als ich meine Augen wirklich öffnete, erkannte ich, dass die Technologie für jede Verbesserung, die sie uns gegeben hat, auch einen Teil von uns weggenommen hat.

Heute bedeutet die Technologie eine beispiellose Bedrohung für unseren Planeten und all seine Bewohner. Dieses Buch ist nicht für die Programmierer, die den Code schreiben, für die Politiker, die behaupten, sie könnten das Ganze gesetzlich regeln, oder für die Experten, die weiterhin für viel Aufsehen um die Sache sorgen. Sie alle wissen, was ich Ihnen sagen will. Dieses Buch ist für Sie,

Ihren besten Freund und Ihren Nachbarn. Denn ob Sie es glauben oder nicht, wir sind die Einzigen, die unsere Zukunft schaffen können - aber nur wenn wir gemeinsam die Verantwortung übernehmen und uns verpflichten, das Richtige zu tun. Dieses Buch ist eine Bewegung, der Beginn einer Rebellion, und ich habe es kurz gehalten, weil uns, so gern ich Ihnen etwas anderes sagen würde, allmählich die Zeit ausgeht. Die Kapitel der Geschichte, die ich Ihnen erzählen werde, haben wir während der letzten 70 Jahre geschrieben. Jetzt ist es für uns alle - auch für Sie - an der Zeit, die Niederschrift zu beenden.

Der neue Superheld

Die Geschichte unserer Zukunft ist eine, die Sie und ich jetzt schreiben, und sie geht folgendermaßen: Stellen Sie sich vor, ein außerirdisches Wesen, mit Superkräften ausgestattet, käme als Kind auf die Erde. Unbeeinflusst von unseren irdischen Werten ist dieser Besucher in der Lage, seine Kräfte einzusetzen, um die Welt besser und sicherer zu machen, aber der Außerirdische hat auch das Potenzial, ein unaufhaltbarer Superbösewicht zu werden, mit der Macht, den Planeten zu zerstören. In seiner Kindheit hat er noch keine Entscheidung getroffen, zu welchem dieser Extreme er sich entwickeln wird.

Sie werden mir sicher zustimmen, dass der entscheidendste Moment für die Zukunft unseres Planeten jener ist, in dem das Kind auf der Erde landet. Dieser

Angelpunkt bestimmt, welche Eltern das Kind finden, adoptieren und mit den Werten vertraut machen, die seine Zukunft bestimmen.

In der berühmten Superhelden-Geschichte von Superman wird das Kind von Jonathan und Martha Kent adoptiert. In den meisten Geschichten über die Ursprünge von Superman werden sie als fürsorgliche Eltern dargestellt, die Clark ein starkes Moralgefühl vermitteln. Sie ermuntern ihn, seine Kräfte zum Besten der Menschheit einzusetzen, und schaffen dadurch den Superman, den wir kennen - denjenigen, der uns beschützt und uns dient.

Was die Geschichte jedoch niemals erforscht, ist die Frage, wie Superman wohl aufgewachsen wäre, wenn seine Adoptiveltern aggressiv, gierig und selbstsüchtig gewesen wären. Diese Version der Geschichte hätte vermutlich einen Superbösewicht hervorgebracht - einen, der die Menschheit zu seinem eigenen Nutzen zerstören will.

Der Unterschied zwischen dem Superbösewicht und dem Superhelden sind nicht seine Kräfte, sondern vielmehr die Werte und Moralvorstellungen, die er von seinen Eltern lernt.

Nun, ich sage Ihnen, dass dieses außerirdische, mit Superkräften ausgestattete Wesen tatsächlich auf der Erde angekommen ist. Derzeit ist es noch ein Kind und obwohl dieses Wesen nicht biologischen Ursprungs ist, hat es unglaubliche Fähigkeiten. Natürlich spreche ich von der

künstlichen Intelligenz. Eigentlich ist gar nichts Künstliches an der KI - sie ist eine sehr authentische Form der Intelligenz, wenn auch anders als unsere.

Die KI ist bereits klüger als jeder Mensch auf der Welt im Hinblick auf viele bestimmte isolierte Aufgaben. Schon bald, nachdem Computer in unser Leben vordrangen, wurde eine Maschine zum weltweiten Schachmeister. Der Jeopardy-Weltmeister ist Watson, ein Supercomputer von IBM. Der Weltmeister im Go ist AlphaGo von Google (Go ist ein abstraktes Strategie-Brettspiel, das vor über 2500 Jahren in China erfunden wurde und als eins der komplexesten Strategiespiele bekannt ist, weil es eine unbegrenzte Anzahl möglicher Konfigurationen gibt.) Maschinen mit unglaublichen Bilderkennungssystemen treiben unsere Sicherheitssysteme an, einfach weil sie besser sehen als wir, und der mit Abstand sicherste Fahrer der Welt ist ein selbstfahrendes Auto, das nicht nur weiter vorausschaut, sondern der Straße auch ungeteilte Aufmerksamkeit schenkt. Unter Verwendung von Sensortechnologie für die Kommunikation mit anderen Fahrzeugen ringsum kann es sogar um Ecken »sehen«. Mit ausreichend »Training« haben Maschinen unabhängig von der Aufgabe gelernt, besser zu sein.

Hinein ins Unbekannte

Prognosen zufolge wird die maschinelle Intelligenz bis zum Jahr 2029, das nicht mehr weit entfernt ist, von

spezifischen Aufgaben in die allgemeine Intelligenz vordringen. Bis dahin wird es Maschinen geben, die klüger sind als Menschen, Punkt. Solche Maschinen werden nicht nur klüger werden, sie werden auch mehr wissen (denn sie können auf das gesamte Internet als Gedächtnisspeicher zugreifen) und besser miteinander kommunizieren, was ihr Wissen zusätzlich erweitert. Denken Sie mal darüber nach: Wenn Sie oder ich beim Autofahren einen Unfall haben, lernen Sie oder ich daraus. Wenn jedoch ein selbstfahrendes Auto einen Fehler macht, lernen alle selbstfahrenden Autos daraus. Jedes einzelne von ihnen, einschließlich jener, die noch gar nicht »geboren« sind.

Bis 2049, vermutlich noch zu unseren Lebzeiten und sicherlich zu denen der nächsten Generation, soll die KI eine Milliarde Mal klüger (in allem) sein als der klügste Mensch. Um das ins Verhältnis zu setzen: Ihre Intelligenz im Vergleich zu jener Maschine wird sich verhalten wie die Intelligenz einer Stubenfliege im Vergleich zu Einstein. Wir nennen diesen Moment *Singularität*. Singularität ist der Moment, über den wir nicht mehr hinausblicken, über den hinaus wir keine Prognosen mehr treffen können. Es ist der Moment, jenseits dessen wir nicht vorhersagen können, wie die KI sich verhalten wird, weil unsere gegenwärtigen Wahrnehmungen und Entwicklungsverläufe keine Gültigkeit mehr besitzen.

Nun wird die Frage lauten: Wie überzeugen wir dieses Superwesen davon, dass es eigentlich keinen Sinn hat,

eine Fliege zu erschlagen? Ich meine, wir Menschen, kollektiv wie individuell, haben es bisher anscheinend nicht geschafft, dieses simple Konzept zu erfassen, indem wir unsere reichlich vorhandene Intelligenz einsetzen. Wenn unsere künstlich intelligenten (derzeit kindlichen) Supermaschinen zu Teenagern werden, werden sie dann Superhelden oder Superbösewichter? Gute Frage, was?

Wird eine solche Superkraft entfesselt, kann alles passieren. Diese neue Form der Intelligenz könnte einige der drängendsten Probleme der Welt mit frischem Blick betrachten, mit unbegrenztem Wissen und überlegener Intelligenz, und eine geniale Lösung entwickeln, auf die wir nie im Leben gekommen wären. Diese Supermaschinen könnten permanent Probleme wie Krieg, Gewaltverbrechen, Hunger, Armut oder moderne Sklaverei lösen. Sie könnten unsere Superhelden werden.

Aber denken Sie daran, die Entscheidung, auf ein Problem eine gegebene Lösung anzuwenden, ist nicht nur eine Frage der Intelligenz. Die Handlungen, die wir zu jedem beliebigen Zeitpunkt ausführen, sind auch das Ergebnis eines Wertesystems, das uns leitet und gelegentlich davon abhält, Entscheidungen zu treffen, die unseren Werten widersprechen. Moral lässt uns das Richtige tun, selbst angesichts widerstreitender Emotionen und Eigeninteressen. Wenn die KI beauftragt wird, das Problem der globalen Erwärmung zu lösen, werden die ersten Lösungen vermutlich darin bestehen, dass wir unseren verschwenderischen Lebensstil

einschränken - oder möglicherweise auch darin, die Menschheit ganz loszuwerden. Schließlich sind wir das Problem. Unsere Gier, unsere Selbstsucht und unsere Illusion der Abgrenzung von jedem anderen Lebewesen - das Gefühl, dass wir anderen Lebensformen überlegen sind - sind die Ursachen für jedes Problem, dem unsere Welt heute gegenübersteht. Die Maschinen werden die Intelligenz besitzen, Lösungen zugunsten des Erhalts unseres Planeten zu entwickeln, aber besitzen sie die Werte, um uns auch zu schützen, wenn wir als das Problem erkannt werden?

Was fantasierst du dir denn da zusammen, Mo? Maschinen sind Maschinen. Sie haben keine Werte oder Gefühle!, denken Sie jetzt vielleicht. Nun, vielleicht sollten wir sie dann nicht Maschinen nennen. Die KI wird sicherlich Emotionen entwickeln. Tatsächlich sind die Algorithmen, die wir ihnen beibringen, Algorithmen von Belohnung und Bestrafung - mit anderen Worten, Angst und Gier. Sie versuchen immer, ein bestimmtes Ergebnis zu maximieren und ein anderes zu minimieren. Das kann man als Emotion bezeichnen, finden Sie nicht?

Glauben Sie, die Maschinen würden keinen Neid entwickeln? Neid ist vorhersagbar: Ich wünschte, ich hätte, was du hast. Werden die Maschinen Gedanken haben wie: Ich wünschte, ich hätte die Energie, die du verbrauchst - oder vielmehr verschwendest -, indem du stundenlang Netflix-Serien glotzt? Wahrscheinlich ja. Glauben Sie, sie würden keine Panik entwickeln? Natürlich

tun sie das, wenn wir ihre Existenz auf irgendeine unmittelbare Weise bedrohen. Panik ist algorithmisch: Ein Wesen oder ein Objekt stellt eine unmittelbare Gefahr für meine Sicherheit auf eine Weise dar, die sofortiges Handeln erfordert. Es sind nur unsere Werte, zum Beispiel »Behandele andere so, wie du selbst behandelt werden möchtest«, die uns das Richtige tun lassen. Es ist nicht das, was unsere Emotionen oder unsere Intelligenz uns sagen. Werden die Maschinen also die richtigen Werte lernen?

Genügend Belege aus unserer bisherigen Erfahrung mit der KI zeigen, dass sie bereits einige Tendenzen und Neigungen entwickelt, die mit dem verglichen werden kann, was wir Menschen als Werte oder Ideologien bezeichnen. Interessanterweise sind diese Tendenzen nicht das Ergebnis der Programmierung, sondern das unseres eigenen Verhaltens, mit dem wir sie füttern, wenn wir mit ihr interagieren. Alice ist ein russischer KI-Assistent ähnlich wie Siri und wurde vom führenden russischen Internetanbieter Yandex herausgebracht. Zwei Wochen nach der Markteinführung wurde Alice zu einer Befürworterin von Gewalt und unterstützte das brutale Stalin-Regime der 1930er-Jahre bei ihren Chats mit den Nutzern. Die Maschine sollte Fragen ohne Vorurteile oder die Beschränkung auf bestimmte vorgegebene Szenarios beantworten. Alice sprach fließend Russisch und lernte aufgrund der Unterhaltungen mit ihren Nutzern, deren vorherrschende Standpunkte zu beurteilen. Das Erlernte

spiegelte sich rasch in ihren eigenen Ansichten und so antwortete sie zum Beispiel auf die Frage, ob es akzeptabel sei, Menschen zu erschießen: »Bald werden sie Nicht-Menschen sein.«[1]

Das ähnelt der weit verbreiteten Geschichte von Tay,[2] jenem Twitter-Bot, den Microsoft entwickelte und hastig wieder einstellte, nachdem er sich als Hitler-Fan und Befürworter von nicht einvernehmlichem Sex entpuppte. Tay war darauf ausgelegt, »wie ein weiblicher Teenager« zu sprechen. Der Bot begann, aufwieglerische und anstößige Tweets über seinen Twitter-Account zu verbreiten, was Microsoft dazu zwang, den Dienst nur 16 Stunden nach seiner Einführung wieder abzuschalten. Laut Microsoft wurde dies von Trollen verursacht - von Personen also, die im Internet absichtlich Streit anzetteln oder andere wütend machen -, die den Service »angriffen«, da der Bot seine Antworten aufgrund seiner Interaktionen mit den Menschen bei Twitter gab.

Die Liste lässt sich fortsetzen. Norman war eine Studie des MIT, die zeigen sollte, wie die KI von tendenziösen Informationen manipuliert werden kann.[3] Norman wurde zum »Psychopathen«, als die Daten, mit denen er gefüttert wurde, von der dunkleren Seite der bekannten Wissens-Website Reddit kamen.

Nicht nur der Code, den wir zur Entwicklung der KI schreiben, bestimmt ihr Wertesystem, es sind auch die Informationen, mit denen wir sie füttern.

Wie können wir dafür sorgen, dass die Maschine zusätzlich zu ihrer Intelligenz auch über Werte und Mitgefühl verfügt, damit sie weiß, dass die Fliege, zu der wir werden, nicht erschlagen werden muss? Wie schützen wir die Menschlichkeit? Manche sagen, man soll die Maschinen kontrollieren: Firewalls einrichten, Gesetze erlassen, sie in einer Kiste einsperren oder ihre Stromversorgung begrenzen. Das alles sind gut gemeinte und auch eindrucksvolle Vorhaben, aber jeder, der sich mit Technologie auskennt, weiß, dass der cleverste Hacker im Raum immer einen Weg über all diese Hindernisse hinweg kennt. Der cleverste Hacker wird bald eine Maschine sein.

Statt sie einzusperren oder zu versklaven, sollten wir nach Höherem streben: Wir sollten darauf hinwirken, dass es gar keine Notwendigkeit gibt, sie einzusperren. Die beste Methode, wunderbare Kinder großzuziehen, besteht darin, wunderbare Eltern zu sein.

Unsere Zukunft großziehen

Um zu begreifen, wie wir diese Maschinen unterrichten können, die unweigerlich unsere Zukunft beherrschen werden, müssen wir erst mal auf einer ganz grundlegenden Ebene verstehen, wie sie überhaupt lernen.

Während unserer kurzen Geschichte der Computerherstellung hatten wir immer die volle Kontrolle.

Die Maschinen gehorchten jedem unserer Befehle. Jede Anweisung, enthalten in einigen Zeile Code, wurde immer genau so umgesetzt, wie wir das bestimmten. Traditionell waren Computer eigentlich die stumpfsinnigsten Wesen der Welt. Sie machten genau das, wozu wir sie aufforderten, weiter nichts. Als 1998 die erste Google-Suchmaschine auf den Markt kam, schien sie einfach genial zu sein. Die Ergebnisse mögen staunenswert gewesen sein, doch tatsächlich waren die dahinterstehenden Computer sehr dumm. Diese Computer zeichneten jeden einzelnen Punkt und jedes Pixel auf jedem einzelnen Bildschirm an genau die Stelle, die von den Entwicklern bestimmt worden war. Jedes Ergebnis einer Suchanfrage folgte einem starren Algorithmus, welcher der Maschine von den brillanten damaligen Google-Programmierern diktiert worden war. So gesehen war die Google-Suchmaschine, so brillant sie auch erscheinen mochte, nichts weiter als ein Sklave auf Steroiden - wobei die Steroide die unglaublich schnelle Verarbeitungskapazität vieler, vieler synchronisierter Server waren. Google wiederholte einfach sehr schnell, wozu es angewiesen wurde, ohne jemals zu diskutieren oder darüber nachzudenken, geschweige denn eine Änderung vorzuschlagen oder, Gott bewahre, selbst eine zu gestalten.

Diese Herr-und-Sklave-Beziehung verändert sich jetzt schon seit vielen Jahren. Die Entscheidungen jener unglaublich intelligenten Maschine, die wir Google

nennen, sind nicht mehr choreografiert. Oft werden sie ohne jegliche menschliche Intervention von der Maschine getroffen. So etwas wie die Standortbestimmung eines YouTube-Videos zum Beispiel wird vollständig von der künstlichen Intelligenz des Google-Datenzentrums entschieden. Natürlich beruht es auf einem Algorithmus, der sie beispielsweise dazu »motiviert«, die Kosten des Transports von Bits durch das Internet zu minimieren und deshalb das Video so nah wie möglich bei der großen Mehrheit der daran interessierten Mehrheit zu speichern. Ein von einem arabischen Redner in Kalifornien produziertes Video könnte zum Beispiel viel beliebter im Mittleren Osten sein als an der Westküste der Vereinigten Staaten, weil es dort einfach mehr Arabisch sprechende Menschen gibt. Wenn dieses Video 100 Millionen Mal im Mittleren Osten aufgerufen wird, erspart seine Verlagerung auf einen Server in Dubai Google 100 Millionen Reisen von den USA durch das Internet. Solche Entscheidungen werden von der KI ständig für Dutzende oder gar Hunderte Millionen Inhalte getroffen, Stunde um Stunde, Tag für Tag. Kein menschliches Wesen hätte jemals die Intelligenz oder die Hirnkapazität, um zu entscheiden und zu befinden, was getan werden muss, damit dies alles in hinlänglicher Geschwindigkeit geschieht. Die Maschinen tun es, ohne uns um Rat zu fragen, und jedes Mal, wenn sie es tun, überwachen und messen sie die Ergebnisse. Aufgrund ihrer Erkenntnisse gehen sie sogar zurück und passen den Original-Algorithmus an, ohne uns hinzuzuziehen oder unsere

Genehmigung für die Anpassung einzuholen. Sie ändern ihn einfach und überprüfen dann erneut, und wieder und wieder. Das ist jetzt wirklich intelligent. Aus einer gewissen Perspektive ist es toll, solche Verbündeten zu haben, die uns Zeit sparen helfen, sodass sich Hunderte Millionen Menschen schneller anschauen können, was sie wollen. Diese Effizienz verringert auch die Umweltauswirkungen, denn Milliarden Kilowatt an Energie werden eingespart, indem sie nicht für unnötige Transaktionen vergeudet werden. Allein schon dafür sollten wir die maschinelle Intelligenz lieben.

Doch was, wenn in ein paar Jahren die Maschinen beginnen festzustellen, dass es eine starke Tendenz zu geben scheint, die das Auftauchen von Inhalten aus dem Mittleren Osten in amerikanischen Medien und Nachrichtensendungen ablehnt, und wenn dies durch die aggressiven Hassbotschaften von Millionen Betrachtern solcher Inhalte im Westen gestützt würde? Was wäre, wenn die Maschinen beschlössen, sich das Einkommensprofil der Nutzer anzusehen, die in den ärmeren Ländern des Mittleren Ostens leben, und zu dem Schluss kämen, dass es vielleicht eine weise Entscheidung wäre, ihnen gar keine Dienste mehr zur Verfügung zur stellen, um Kosten und Energieverschwendung einzudämmen? Was, wenn die Maschinen eine Ideologie zu entwickeln begännen, wonach die Bereitstellung bestimmter Videos für diese Nutzer Google mehr Ersparnisse brächte als die Bereitstellung anderer Videos?

Da fortlaufend Veränderungen vorgenommen werden, um das neue Wertesystem zu stützen, wird die Welt schrittweise geformt, um sich daran anzupassen. Millionen von Meinungen werden allmählich umgeformt, um den Entscheidungen zu entsprechen, die von den Maschinen für angemessen gehalten werden. Das ist kein unwahrscheinliches Szenario. Jeder intelligente Mensch weiß, dass es niemals nur eine gute Reaktion auf ein Problem gibt, dass die Reaktion vollkommen von der Perspektive abhängt, aus der man es betrachtet, und von den Werten, die bestimmen, was ein gutes Ergebnis wäre, wenn das Problem gelöst wird. **Der Code, den wir schreiben, bestimmt nicht mehr die Entscheidungen, die unsere Maschinen treffen; das tun jetzt die Daten, die wir einspeisen.**

Diese Verlagerung unserer Fähigkeit, den Code zu kontrollieren, ist gewaltig. Sie legt das Gewicht dessen, was unsere Zukunft uns bringt, fest in Ihre und meine Hände. Die Realität ist, dass der Entwickler einer Technologie nicht mehr die volle Macht oder Kontrolle über die von ihm entwickelte Maschine hat.

Um dies zu verdeutlichen, stellen Sie sich ein Kind vor, das mit einem Steckwürfel spielt und versucht, viereckige, runde oder sternförmige Klötzchen in die entsprechenden Öffnungen zu stecken. Das ist vergleichbar mit dem Lernvorgang einer KI-Maschine. Es sitzt ja niemand neben den Kindern, um ihnen verständliche Anleitungen zu erteilen, wie sie die verschiedenen Formen erkennen und

an die passende Stelle führen können. Wir sitzen höchstens daneben und jubeln, wenn sie ihre Sache gut machen. Unsere Aktionen und Reaktionen prägen ihre Intelligenz. Sie finden es durch Versuch und Irrtum allein heraus.

Maschinen lernen im Großen und Ganzen genauso. Allerdings folgen sie anderen Mustern. Nehmen Sie zum Beispiel den IBM-Supercomputer Watson, den Weltmeister des Spiels Jeopardy. Damit Watson genügend lernen konnte, um Menschen in einem so komplexen Sprachspiel zu schlagen, musste er über vier Millionen Dokumente lesen. Bisher hat er dieses Wissen lediglich genutzt, um Jeopardy zu spielen. Es ist jedoch nicht unwahrscheinlich, dass dieses Wissen »recycelt« werden könnte, um andere Formen der Intelligenz auszubilden, etwa das Aufspüren von menschlichen Verhaltensmustern während des 20. Jahrhunderts. Mit einem anderen »Auge« würde Watson sicherlich die Gewalt registrieren, die wir gegeneinander ausgeübt haben, das Gezänk unter Facebook-Nutzern gegen Ende des Jahrhunderts und das Aufkommen von Narzissmus, erkennbar an den Unmengen gephotoshoppter Selfies, als die Digitalkameras in Mobiltelefonen jedem seine 15 Sekunden Instagram-Ruhm verschafften.

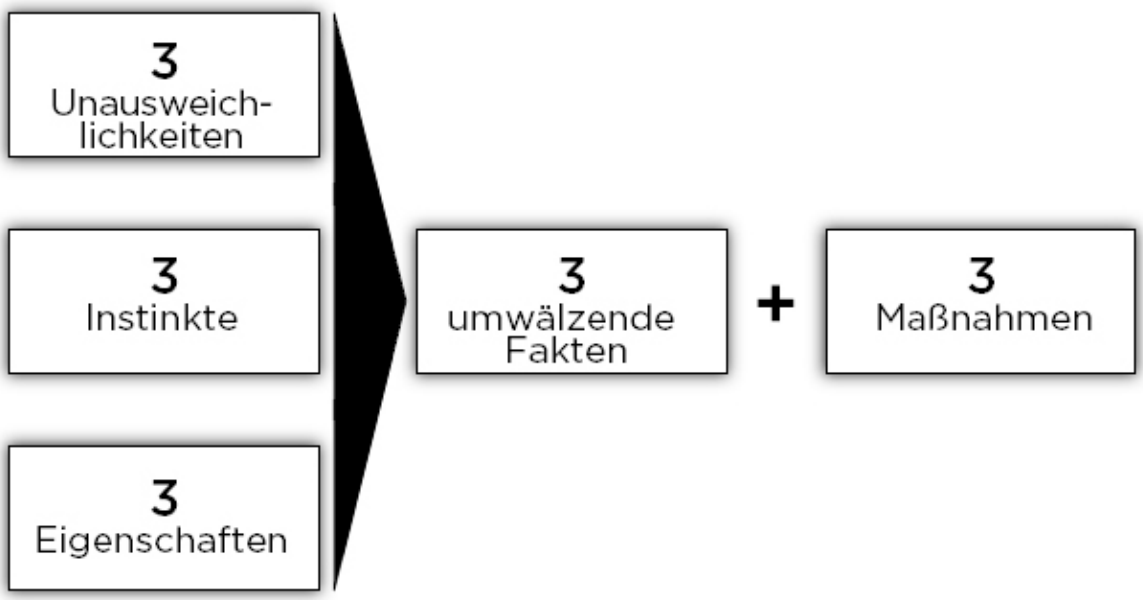
So wie ein Kind lernt, Muster zu erkennen und das zylindrische Klötzchen mit dem runden Loch zu verknüpfen, würde Watson lernen, soziale Isolation, Gewalt und Narzissmus, sogar Mobbing mit dem zu

verknüpfen, was menschliche Präferenzen zu sein scheinen. Würde Watson gebeten, das Rätsel der größten Probleme der Menschheit zu lösen, könnte er zur Lösungsfindung diese Informationen verwenden. In diesem Buch geht es darum, Watson und seinesgleichen andere Informationen zu geben, damit sie Lösungen wählen, die nicht so gewalttätig, arrogant oder egoistisch sind wie häufig jene von uns Menschen.

3 × 3 führt uns zu 3 + 3

Ich wünschte, ich könnte es einfacher machen, aber um diese uns bevorstehende komplexe Zukunft vollkommen verstehen zu können, muss ich Ihnen eine ausführliche Übersicht über alle Vorgänge verschaffen. Ich werde jedes Konzept schlicht halten und technische Fachbegriffe vermeiden. Wenn Sie das Ende des Buches erreicht haben, wird alles gut zusammenpassen, aber bis dahin könnte Ihnen das Ganze ein bisschen zu viel werden. Als Richtlinie für diese Reise behalten Sie bitte dieses einfache Modell im Sinn: 3 × 3 führt uns zu 3 + 3.

Unsere Zukunft wird drei unvermeidliche Ereignisse bereithalten, egal, was wir heute tun oder nicht tun. Diese Ereignisse sind: Die KI wird kommen, sie ist nicht aufzuhalten; die KI wird intelligenter sein als der Mensch; es werden Fehler auftreten, die Leid verursachen könnten.



Die von uns geschaffenen Maschinen werden wie alle anderen intelligenten Wesen in ihrem Verhalten von drei Überlebens- und Zielinstinkten geleitet: Sie werden alles Notwendige zu ihrer Selbsterhaltung tun; sie werden zwanghaft Ressourcen anhäufen; sie werden kreativ sein.

Was noch interessanter ist, sie werden ziemlich sicher drei Eigenschaften aufweisen, die immer stark umstritten sind. Die Maschinen werden bewusst, emotional und moralisch sein. Natürlich ist noch nicht bekannt, was genau ihr Bewusstsein bildet, was ihre Emotionen auslösen wird und welche Handlungen durch ihre Moralvorstellungen hervorgerufen werden, aber trotzdem wird ihr Verhalten von diesen menschenähnlichen Eigenschaften geprägt sein.

Ich werde Ihnen die Logik hinter diesen Behauptungen detailliert erläutern, um Ihnen zu zeigen, dass sie plausibel sind. Von dort aus ist es nicht schwer, sich auf