

# » Beobachtungen zur Zeit «

---

No 19



M.M. WARBURG & CO  
BANK



# **Algorithmen steuern die Welt – Plädoyer für eine positive Interpretation**

von Dr. Tobias Knobloch<sup>1</sup>



Dass die Delegation von Entscheidungen an Maschinen in bestimmten Situationen zu wünschenswerten gesellschaftlichen Ergebnissen führt, weiß jeder Autofahrer. Von Zeit zu Zeit fällt immer mal irgendwo eine Ampel aus, so dass Polizisten den Verkehr regeln müssen. Und es ist frapierend, um wie viel schlechter dann der Verkehr an besagten Kreuzungen fließt – unabhängig davon, welche Personen das machen und um welche Straßensituation und Verkehrslage es sich handelt. Immer entsteht bei händischer Verkehrsregelung in irgendeine Richtung ein Kreuzungstau. Offenbar ist die automatische Lichtzeichenverkehrsregelung dem Menschen also weit überlegen. Sie war es schon, als das noch über Relais-technik mit Zeitschaltautomatik funktionierte und es ist erst recht der Fall in der sogenannten Smart City, in der dies in Interdependenz mit anderen Steuerungsbereichen von Algorithmen auf Mikroprozessorbasis geregelt wird.

### **Experimente in einer undurchsichtigen Technikwelt**

Dieses Beispiel ist nur eines von vielen, die illustrieren, wie sehr in unserer Welt gesellschaftspolitische Entscheidungen (z. B. darüber, wie der Verkehr fließen soll und welche Verkehrsmittel dabei zum Einsatz kommen) bereits technisch ausgeführt, überwacht und optimiert werden. Hinzu kommt der in dieser Hinsicht schon viel weiter entwickelte private Bereich: Jede Bestellung bei Amazon, jede Nutzung eines Kartendienstes, jedes Facebook-Posting und jedes Öffnen eines Carsharing-Autos mit dem Smartphone involviert im Hintergrund komplexe algorithmische Systeme, die inzwischen selbst für die Hersteller und Betreiber manchmal schwer durchschaubar sind. Diese relative Undurchsichtigkeit aufgrund von (allmählich gewachsener) Komplexität ist übrigens nichts Neues: Schon Wettermodelle, deren Nutzung für uns seit etwa 50 Jahren selbstverständlich ist und den meisten von uns zudem vermutlich unverzichtbar erscheint, sind so komplex, dass selbst Fachleute nicht mehr genau sagen können, welcher Modellteil funktional für was verantwortlich ist

und was passieren würde, wenn man ihn herausnähme oder veränderte. Sie tun einfach das, was sie sollen und können optimiert werden, indem man den Input gegen die gewünschte Performanz optimiert. Mehr braucht es zunächst nicht, um über halbwegs funktionale Wetterprognosen zu verfügen. Die zugrundeliegenden Gleichungen und technischen Implementierungen können teilweise opak, die Gründe für ihre Wirksamkeit unbekannt sein, ohne in der Praxis unbrauchbar sein zu müssen. Ganz anders ist das bei sicherheitsrelevanten Systemen (wie z. B. Flugzeugautopiloten): Dort gelten strikte Standards für Transparenz und sogar für die retrospektive Nachvollziehbarkeit des Entwicklungsprozesses.

Das Thema Begründungszusammenhänge wird uns ebenso noch beschäftigen wie das Transparenzthema. Beide werden bedeutender, weil gesellschaftliche Prozesse, an denen algorithmische Datenverarbeitung beteiligt sind, immer auch mal zu anderen Ergebnissen führen, als sie sich die meisten von uns vorstellen und wünschen. Die Wahl Donald Trumps zum US-Präsidenten ist beispielsweise – jedenfalls nach allem, was wir spätestens seit dem neuerlichen Skandal um die Nutzung von Facebook-Daten durch die Analysefirma Cambridge Analytica darüber wissen – eben auch ein Technologie-induziertes Phänomen. Anders ist das Wahlergebnis angesichts der Millionen Stimmen insgesamt weniger und des hauchdünnen Vorsprungs Trumps in den Swing States gar nicht erklärbar. Die Trump-Wahl ist zumindest auch darauf zurückzuführen, dass sich ein politisches Lager die Meinungsbeeinflussungsmöglichkeiten der Online-Werbetechnologie perfekt zunutze gemacht hat. Mit anderen Worten wurden technische Mechanismen, die Internetkonzerne dazu entwickelt hatten, um mit Werbung Geld zu verdienen, in einen anderen Bereich transferiert und dazu benutzt, trotz einer gegenteiligen Gesamtstimmung im Land ein bestimmtes Wahlergebnis zu erzielen. Dadurch wird die technische Steuerung ganzer gesellschaftlicher Systeme, von der Kybernetiker und von der Kybernetik begeisterte Politiker schon in den 1960er Jahren geträumt haben, langsam aber sicher vorstellbar. Man könnte auch sagen, dass wir dank der heutigen technischen Möglichkeiten, deren Auswirkungen wir kaum verstehen, geschweige denn kontrollieren können, nolens volens am größten

anzunehmenden Gesellschaftsexperiment teilnehmen. Wohin es uns führt, hängt auch davon ab, wie schnell wir uns unsere Lage bewusst machen und entsprechend handeln können. Dieser Aufsatz soll dazu einen kleinen Beitrag leisten.

### **Der Mensch: Das automatisierende Wesen**

Die Automation von Prozessen findet nicht erst seit dem vorigen oder vorvorherigen Jahrhundert statt. Vielmehr automatisiert unsere Spezies, seit es Homo Sapiens gibt. Der Weg führt über die Nutzung von Werkzeugen im Bereich der physischen Welt, zum Beispiel durch die Nutzung von Jagd- und Landwirtschaftswerkzeugen, schließlich bis in die Welt der Ideen und des Sozialen hinein. Eine frühe signifikante Verbindung beider Seiten der Automation, der physikalischen und der kognitiven, stellt der Jacquard-Webstuhl aus dem Jahr 1804 dar. Ein automatischer mechanischer Webvorgang wird hier durch Lochkarten gesteuert; gewebt wird das auf den Lochkarten als Information gespeicherte Muster. Man kann so weit gehen und behaupten, die Fähigkeit und der Hang zur Automation sei das Distinktionsmerkmal des Menschen schlechthin (und nicht Intelligenz, Sprache, Kultur – alles Merkmale, die Tiere in einem lange von uns unterschätzten Grad ebenfalls aufweisen). Die Geschichte der menschlichen Naturbeherrschung ist eine Siegesgeschichte automatisiert in Wirklichkeitsveränderung überführter Ideen.

Was den Menschen evolutionär so erfolgreich, ja zum Herrscher der Welt gemacht hat, könnte man auch sagen, ist die Verbindung automatisch ablaufender physikalischer Vorgänge mit Ideen, wie diese am zweckmäßigsten (beziehungsweise am schönsten, am schnellsten, originellsten...) gestaltet sein sollten. Weit über die Kleidung, die wir tragen, und die buchstäblichen Werkzeuge, die wir verwenden, hinaus können wir unsere von Natur aus beschränkten Fähigkeiten auf diese Weise nahezu beliebig erweitern. Die Grundlage dafür ist die Fähigkeit zur Abstraktion, die wiederum auf mentaler Modellbildung beruht.

Evolutionär bestand der große Vorteil von Sapiens gegenüber Tieren und primitiven Hominiden darin, dass er mögliche zukünftige Weltverläufe modellartig antizipieren, miteinander vergleichen und den für ihn jeweils günstigsten zur Geltung bringen konnte.

Dieser Vorteil, der dem Menschen im Zuge der Evolution zuteil wurde, differenzierte sich in den zurückliegenden 5000 Jahren immer weiter aus. Abstraktion, Modellbildung und gedankliche Variation haben uns zur Mathematik, zur wissenschaftlichen Methode des Experiments und letztlich auch zu den Geistes- und Sozialwissenschaften verholfen. Ein unmittelbarer Nachfahre des strategischen Verhaltens der Jäger und Sammler vor zigtausend Jahren ist das Gedankenexperiment. Albert Einstein war ein Meister in der Anwendung dieser Methode. Gedankenexperimente wie der Weltraumfahrtstuhl oder das Zwillingssparadoxon waren Wegbereiter für Einsteins Relativitätstheorie. Nimmt man die zahlreichen quantenphysikalischen Gedankenexperimente wie etwa Schrödingers Katze hinzu, so kann man mit einigem Recht behaupten, dass die Entwicklung der gesamten neuzeitlichen Physik ohne die gezielte und kontrollierte gedankliche Variation tatsächlicher Umstände und Ideen nicht denkbar gewesen wäre. Noch weiter zugespitzt: Wir könnten uns heute nicht von Navigationsgeräten durch fremde Städte leiten lassen (wofür die Gültigkeit der Allgemeinen Relativitätstheorie die Voraussetzung ist), wenn ein Mann wie Einstein nicht ein so großer Meister in gedanklicher Modellvariation gewesen wäre.

Seit etwa 50 Jahren leben wir nun in einem Zeitalter, da wir unsere einzigartige Fähigkeit zur Abstraktion, Modellbildung und Variation zum Zwecke von Erkenntnis, Naturbeherrschung und Machterwerb immer mehr in Maschinen verlagert haben. Und das aus gutem Grund, denn diese können Möglichkeitsräume quantitativ in ganz andere Dimensionen explorieren und auf diese Weise qualitative Umschläge sichtbar machen, die dem menschlichen Geist ohne Computerhilfe unzugänglich sind.



## Im Datenzeitalter

Seit der Entstehung des sogenannten Web 2.0 und dem unmittelbar darauffolgenden Siegeszug des Smartphones, also seit erst etwa 15 Jahren, leben wir in einem regelrechten Datenzeitalter. Die Datenmenge ist vor allem mit dem Aufkommen der Sozialen Medien exponentiell gestiegen. Nachrichten, Fotos, Texte, Musik und Videos werden seither in unvorstellbarer Menge erstellt, bearbeitet, zerteilt und neu zusammengesetzt („remixed“), vervielfältigt und geteilt. Bei all diesen Vorgängen fallen weitere digitale Daten an, die angeben, wann all das passiert, wo und durch wen; welche Hardware daran beteiligt ist und über welche Internetverbindungen die Vorgänge laufen. Solche Metadaten (ein einfaches Beispiel ist der Browserverlauf, also die Historie der von einem Nutzer besuchten Websites) sind unglaublich aufschlussreich, was wirtschaftliche und kriminelle Vorgänge sowie persönliche Informationen angeht. Der nächste Schritt in der Geschichte der Datenerzeugung und Datenverarbeitung hat mit der massenhaften Verbreitung von mit dem Internet verbundenen Geräten (sogenannte „embedded devices“) längst begonnen. Damit erzeugt heute auch die Bräunung eines Toasts und das Waschen schmutziger Wäsche zunehmend jene um die Welt vagabundierenden Datenströme.

Da die Zahlen, mit denen Datenmengen angegeben werden, ohnehin längst von jeder Vorstellungskraft entkoppelt sind, soll den zahlreichen, mehr oder weniger aufschlussreichen Vergleichen hier kein weiterer hinzugefügt werden. Stattdessen mag es genügen, das für das Jahr 2025 prognostizierte weltweite Datenvolumen mit über 160 Zettabyte anzugeben und darüber zu sagen, dass das einer sehr großen Informationsmenge entspricht und viel mehr ist als das, was alle Bibliotheken der Erde zusammen enthalten.

Dieses Wachstum der Datenerzeugung beruht auf der gleichen Tatsache wie unser Vermögen, bei der Verarbeitung und Analyse mit diesem Wachstum Schritt zu halten: Die Halbleitertechnologie (also die Herstellung materieller Computerrecheneinheiten, sogenannter Chips) war jedenfalls bislang in der Lage, die Leistungsfähigkeit dieser Recheneinheiten in etwa alle zwei Jahre zu verdoppeln, sie gleichzeitig entsprechend zu verkleinern und den Preis zu

halbieren. Doch diese Regel beginnt allmählich ihre Gültigkeit zu verlieren. Das liegt schlicht daran, dass man an physikalische Grenzen stößt: Im Nanometerbereich können durch immer mehr Schaltkreise auf immer kleinerer Fläche bald keine Fortschritte mehr erzielt werden – im Gegenteil: Ab einer bestimmten Größe gelten quantenmechanische Gesetze wie die Unschärfere-lation des deutschen Physikers Werner Heisenberg. Das Verhalten von Silizium-Teilchen wird im Grenzbereich von 7 bis 5 Nanometern unkontrollierbar; die fortschrittlichsten heutigen Mikroprozessoren arbeiten mit einer Schaltkreisgröße von 14 Nanometern. Mit dem Erreichen physikalischer Grenzen der Rechenbeschleunigung von Prozessoren ist die Geschichte des Fortschritts jedoch nicht vorzeitig beendet. Bereits heute kommen vermehrt spezifische, eigens auf Aufgabentypen zugeschnittene Chip-Architekturen, bei denen die Transistorendichte nicht der entscheidende Faktor ist, zum Einsatz; außerdem wird intensiv am sogenannten Quantencomputer und an anderen Alternativen zum Silizium und zur Von-Neumann-Computerarchitektur geforscht.

### **Algorithmen, Big Data, Künstliche Intelligenz – eine Begriffsentwerrung**

Mittlerweile redet alle Welt von Algorithmen, genauer von Big Data, Algorithmen und Künstlicher Intelligenz. Aber was sind Algorithmen eigentlich, und sind die drei genannten Begriffe mehr oder weniger austauschbar? Sie werden ahnen, dass das natürlich nicht der Fall ist, weshalb zunächst ein wenig Arbeit in begriffliche Hygiene zu investieren ist.

Ein Algorithmus ist zunächst nichts anderes als eine logisch verknüpfte Reihe von Anweisungen zum Durchlaufen eines Prozesses, an dessen Ende ein bestimmtes Ziel erreicht werden soll, ganz ähnlich einem Kochrezept. Die Ausführung dieser Anweisungen ist zunächst unabhängig vom Medium der Ausführung. Deshalb ist eine mündliche Wegbeschreibung auch ein einfacher Algorithmus, der Menschen zum Finden eines Ortes dient. Wenn heute in der medialen Debatte von Algorithmen die Rede ist (beispielsweise in Bezug auf

Facebooks Nachrichtenauswahl, Googles Suche oder selbstfahrende Autos), dann sind meistens umfangreiche Computerprogramme gemeint, also komplexe algorithmische Systeme, die von mächtigen Maschinen mit viel Rechenkraft realisiert werden. Leichter veranschaulichen kann man sich das an einem einfacheren Beispiel: Im oben erwähnten Fall der Wegbeschreibung könnte man auch sagen, die Beschreibung ist der Algorithmus, der Mensch ist die ausführende Rechenmaschine und die Karte, die er dabei zu Hilfe nimmt (sei es im Kopf, auf Papier oder dem Smartphone), ist die Datenbasis. Jeder Algorithmus, der durch ein Computerprogramm realisiert wird, verarbeitet Daten. Rechen-einheit, Schritt-für-Schritt-Anweisungen und Daten sind die zentralen Elemente der automatisierten Welt, die wir uns geschaffen haben, indem wir unsere geistige Fähigkeit zu Abstraktion, Modellbildung und Variation auf Maschinen übertragen haben.

Die Debatte um Künstliche Intelligenz (kurz: KI) kreist im Grunde darum, wie weit diese Übertragung gediehen ist und als wie selbständig man sie zu betrachten hat. Unterschieden wird inzwischen die sogenannte schwache KI von der starken KI.

Starke KI gibt es in Wirklichkeit (noch) nicht, sie existiert nur als Idee.

Als schwache KI, unter die demnach sämtliche heute gebräuchliche und bekannte KI-Systeme fallen, bezeichnet man Systeme, die konkrete Anwendungsprobleme lösen, sich dabei jedoch nur in engen Grenzen selbst optimieren können; in Wahrheit werden Neujustierungen nämlich fast immer durch den Menschen veranlasst und gesteuert. Sie greift dabei auf Methoden zurück, die ihr für die Problemlösung vorgegeben wurden und erlangt kein tieferes Problemverständnis. Das dominierende System-Paradigma ist das Prinzip der automatischen Mustererkennung. Beispiele für schwache KI sind demnach: Text- und Spracherkennung, Bilderkennung, Übersetzungssysteme, Navigationsprogramme, Autovervollständigung und Autokorrektur bei Texteingaben, Chatbots und Avatare in der Kundenbetreuung sowie überhaupt sämtliche sogenannte Expertensysteme (das sind Systeme, die auf Basis von Wissensdatenbanken Handlungsempfehlungen ableiten und ausgeben). Diese Aufzählung verdeutlicht, dass wir bereits an

vielen Stellen unseres alltäglichen Lebens mit schwacher KI interagieren, meist ohne es wahrzunehmen.

Starke KI muss dagegen viel weiterreichende Fähigkeiten aufweisen, namentlich logisches Schlussfolgern, Entscheidungsfähigkeit unter Unsicherheit, Planungs- und Lernfähigkeit, die Fähigkeit zu natürlich-sprachlicher Kommunikation sowie das Kombinieren all dieser Fähigkeiten zur Erreichung eines übergeordneten Ziels. Letztlich müssten also die eingangs für den Menschen als entscheidende Abgrenzungsmerkmale beschriebenen Fähigkeiten zur Abstrahierung, Modellbildung und Antizipation erreicht werden. Eine in diesem Sinne starke KI müsste sich im Grunde also auch selbst weiterentwickeln können – so wie der Mensch mit eben diesen Fähigkeiten KI erschaffen hat.

Die Latte hängt aus Sicht der Algorithmenentwicklung also hoch – allerdings nicht höher als das, wozu ein durchschnittlich intelligenter Mensch in der Lage ist.

Um den Übergang von schwacher zu starker KI zu erkennen, sind verschiedene Testverfahren vorgeschlagen worden, angefangen beim sogenannten Turing-Test, den der britische Mathematiker Alan Turing bereits 1950 formuliert hat. Dabei führt ein Mensch über eine Tastatur ohne Sicht- und Hörkontakt gleichzeitig ein Gespräch mit einem Menschen und einem Computerprogramm. Wenn der Mensch im Verlauf des Gesprächs nicht sagen kann, welcher Gesprächspartner Mensch und welcher Computer ist, dann hat das Computerprogramm den Test bestanden und ihm kann ein dem Menschen gleichrangiges kognitives Vermögen attestiert werden. Kritiker haben darauf hingewiesen, dass der Turing-Test menschliche Fähigkeiten nur unvollständig prüft und entsprechende Erweiterungen vorgeschlagen. Der nach der Programmiererin Ada Lovelace benannte Lovelace-Test verlangt etwa vom Computer, ein kreatives Werk (z.B. eine Kurzgeschichte) nach menschlichen Vorgaben wie Sujet, Handlung etc. zu erschaffen. Dem deutschen Kognitionswissenschaftler Thomas Metzinger ist auch das nicht genug: Er ist der Meinung, ein Computerprogramm müsse sinnvoll an intellektuellen Diskursen wie der Frage, was Intelligenz eigentlich ausmacht, teilnehmen können, damit man es im starken Sinne intelligent nennen kann. Aber auch an diesen Vorschlägen zur Erweiterung des

Turing-Tests zeigt sich vor allem, wie schwer es uns Menschen fällt, von der Art der Intelligenz, die für uns charakteristisch ist, zu abstrahieren und uns vorzustellen, dass andere Entitäten eines Tages vielleicht über sehr mächtige Fähigkeiten verfügen könnten, die der menschlichen Intelligenz strukturell überhaupt nicht gleichen. Außerdem ist der Intelligenz-Begriff unscharf definiert; die meisten erkennen sie, wenn sie sie vor sich haben, können aber nicht präzise sagen, was alles notwendig dazu gehört. Klar scheint heute zu sein: Falls es überhaupt jemals zu so etwas wie einer starken oder allgemeinen KI kommen wird (die Mehrzahl der Experten geht heute davon aus, dass das in einem Zeitfenster von 20 bis 50 Jahren passieren könnte), dann wird sie mit recht hoher Wahrscheinlichkeit nicht die menschliche Intelligenz eins zu eins imitieren, sondern eine völlig neue Art Intelligenz darstellen.

Viel nüchterner lässt sich der letzte der hier zu erläuternden Begriffe erklären: „Big Data“ ist, wenn man nicht bloß eine große digitale Datenmenge, sondern das entsprechende Auswertungsverfahren für große Datenmengen meint (was meistens der Falls ist), eigentlich eine Abkürzung von „Big Data Analytics“. Was sich dahinter verbirgt, ist in seiner Bedeutung für die Gegenwart kaum zu überschätzen, denn oft verbirgt sich auch hinter der Rede von „Algorithmen“ oder „Künstlicher Intelligenz“ nichts anderes als diese zeitgemäße Form von Business Intelligence, also statistische Auswertung von Daten. Hier geht es darum, in kurzer Zeit statistische Erkenntnisse und Korrelationen in großen und womöglich unstrukturierten Datenmengen zu finden, um daraus entsprechende Erkenntnisse abzuleiten, die einem staatlichen, wirtschaftlichen oder privaten Zweck dienen können. Ein Beispiel: Wenn ich sämtliche Gespräche, die in sozialen Netzwerken öffentlich über die Verkehrssituation einer Stadt geführt werden, auswerte, dann kann ich darüber bereits ziemlich viel über die aktuelle Verkehrssituation dieser Stadt erfahren. Das gilt erst recht dann, wenn ich diese Informationen mit anderen Datenquellen – Staumachrichten, offene Fahrplandaten, ÖPNV-Störungsmeldungen, Baustelleninformationen, Satellitenbilder etc. – kombiniere. Aber auch wenn mir als Nutzer sozialer Netzwerke Werbung angezeigt wird, wenn mir Musik-Apps Songs vorschlagen oder die Online-Partnerbörse einen Menschen zum Kennenlernen,

dann läuft im Hintergrund stets und für den Nutzer meistens unmerklich Big Data Analytics ab. Aber was können all diese Algorithmen, Big-Data- und KI-Systeme inzwischen eigentlich? Wozu sind sie imstande und wie nützlich oder gefährlich ist das?

### **Von der Intelligenz und Dummheit algorithmischer Systeme**

Auf der Website der Robotik-Firma *Boston Dynamics* kann man ein Video ansehen, in dem der Roboter Atlas – der zurzeit am weitesten entwickelte humanoide Roboter weltweit – einen Rückwärtssalto macht. Auch kann man dort sehen, wie Roboter untereinander koordiniert gemeinsam Türen öffnen, Werkzeug zur Problemlösung nutzen und einige weitere erstaunliche Dinge vollbringen. Allein das Gehen im unwegsamen und sogar gleichzeitig verschneiten Gelände, das Atlas inzwischen problemlos beherrscht, hat Vorgängergenerationen von Robotern vor scheinbar unüberwindliche sensomotorische Koordinierungsschwierigkeiten gestellt.

Noch als ich um das Jahr 2005 herum Doktorand an der Universität Bielefeld war, und die dortige Robotikforschungseinrichtung CITEC zwei Roboter der Baureihe ASIMO von der japanischen Firma *Honda* – damals ihrerseits die am weitesten entwickelten menschenähnlichen Roboter – zur Verfügung gestellt bekam, da konnten diese kaum zwei, drei gerade Treppenstufen steigen, ohne dass man sich sorgen musste, dass sie jeden Moment umfallen. Und nicht selten sind sie, auch bei öffentlichen Demonstrationen, tatsächlich umgefallen. Soll heißen: Die Entwicklung hat seither beeindruckende Fortschritte erzielt und sich durch die Verfügbarkeit großer Datenmengen, die Nutzung (alter) neuer informationstechnischer Architekturen wie neuronaler Netze, Verfahren wie Machine Learning und das stetige Anwachsen maschineller Rechenkraft nach dem Mooreschen Gesetz sehr beschleunigt.

Das hat Erstaunliches ermöglicht. Manch Leser hat vielleicht gehört, dass das Programm AlphaGo inzwischen das Brettspiel „Go“, das um ein Vielfaches

komplexer ist als Schach, besser beherrscht als menschliche Go-Meister. Go ist zwar komplex, aber es verfügt wie Schach doch über ein diskretes Regelwerk, das man erlernen kann und aus dem sich in Kombination mit einer Spielstellung eine endliche Zahl möglicher Züge ergibt. Da es bei derlei Spielen darum geht, den Raum möglicher weiterer Spielverläufe möglichst vollständig zu überblicken, leuchtet es ein, warum der Computer dem Menschen hier überlegen ist. Vor diesem Hintergrund ebenfalls wenig überraschend ist, dass Computer das Spiel Jeopardy mit ihrer ungleich größeren Menge an verfügbarem Expertenwissen gewinnen. Überraschter sind wir dann schon, wenn ein Computer Menschen beim Poker schlägt oder wohlschmeckende Kochrezepte kreiert. Hier werden wir unsicher, weil wir davon ausgehen, dass technische Systeme dabei kreative Fähigkeiten entfalten, von denen wir bisher glaubten, nur Menschen seien dazu in der Lage. Dabei handelt es sich in allen genannten Fällen lediglich um bloßes Number Crunching (frei übersetzt: Datenverarbeitung), wodurch algorithmische Systeme diese uns in Erstaunen versetzenden Fähigkeiten entfaltet haben.

Die Emergenz uns zunächst überraschender Fähigkeiten, die bei näherem Hinsehen jedoch aus trivialen Routinen bestehen, lässt sich gut am Beispiel automatisierter Sprachübersetzungen beleuchten. Das im Internet frei verfügbare und sicher von einigen Lesern bereits genutzte Programm *Google Translate*, das auf der Neuronale-Netz-Architektur „Tensor Flow“ beruht, übersetzt 96 Sprachen. Die Qualität der Übersetzung hängt sehr davon ab, was zur Übersetzung ansteht. Die Ergebnisse reichen von erstaunlich gut bis sinnlos. Der Kognitionswissenschaftler und professionelle Übersetzer Douglas Hofstadter, manchem vielleicht aufgrund seines Bestsellers „Gödel, Escher, Bach“ bekannt, hat das kürzlich in dem bemerkenswerten Aufsatz „The Shallowness of Google Translate“ näher analysiert. Zu Beginn seiner Untersuchung weist er zu Recht darauf hin, dass jede Übersetzung mehr ist als die Decodierung und Recodierung von Zeichen, sondern vielmehr ein kreativer Akt des Verstehens und ‚Nachschaffens‘ originärer Gedanken.

Oft werden zwar Worte zusammengestellt, die einen gewissen Sinn ergeben, die aber bestenfalls eine näherungsweise, keineswegs aber eine adäquate Übersetzung des Originals darstellen.

Hofstadter kommt zu dem Schluss, dass *Google Translate* und ähnliche Übersetzungsprogramme ebenso wenig Sprache verstehen wie eine *Microsoft Xbox* oder eine *Sony Playstation* Fußball spielen können. Völlig zu Recht weist Hofstadter allerdings auch darauf hin, dass maschinelle Übersetzungsprogramme trotz ihres Nichtverstehens enorm nützlich sein können und in bestimmten Anwendungsbereichen durchaus ihre Berechtigung haben. Dem schließe ich mich hier an, gebe aber sogleich zu bedenken, dass uns Technik, die in eingegrenzten Bereichen gut funktioniert, oftmals suggeriert, sie könne mehr und sei – zumindest in diesem einen Anwendungsbereich – ein Universalwerkzeug. Durch seine erstaunlich adäquate Übersetzung bestimmter Texte, beispielsweise wenn sie von Technik und Politik handeln (an massenhaften Texten vor allem solcher Art wurde das System nämlich trainiert) suggeriert *Google Translate* nicht nur, dass es Sprache im Vollsinn des Wortes versteht, sondern legt uns dadurch auch nahe, dass Maschinen generell sprachbegabt und so weit ausgereift sind, dass wir sie verständlich nennen können. Und an dieser Erwartung richten wir unseren Umgang mit Maschinen fortan aus. Diese Erwartung ist aber falsch und die Ausrichtung unseres Umgangs mit Maschinen ist es folglich auch. Denn die im Ergebnis bisweilen erstaunlichen Fähigkeiten von Übersetzungsprogrammen beruhen nur teilweise auf dem Algorithmus, ein wesentlicher anderer Teil ist die digitale Verfügbarkeit großer Textmengen, in denen der Algorithmus im Grunde nur ‚nachschauf‘.

Normalerweise werden Algorithmen zunächst an einem Pool historischer Daten „darauf trainiert“, wie man im Fachjargon sagt, mit einer bestimmten Aufgabenklasse fertig zu werden. Bilderkennungssoftware wird beispielsweise mit großen Mengen an Bildern gefüttert, die sie später klassifizieren soll. Diese Trainingsbilder sind „gelabelt“, wie man sagt, das heißt, dass sie bereits der jeweils richtigen Bildkategorie, die der Algorithmus später selbst finden soll, zugeordnet sind. An diesen Trainingsdaten ‚lernt‘ der Algorithmus also die gewünschten Ergebnisse und wie er zu ihnen gelangt (ganz grob vereinfacht ausgedrückt). Auf diese Weise war auch AlphaGo mit unzähligen historischen Daten von Go-Spielen zwischen Menschen trainiert worden. Das hat das Programm zu einem so guten Go-Spieler gemacht, dass es im März 2016 den



südkoreanischen Go-Weltmeister Lee Sedol in fünf Partien 4:1 besiegte; lediglich das vierte Match konnte Sedol für sich entscheiden. Er gab später zu Protokoll, dass er sich zu keiner Zeit während der fünf Spiele als der das Spiel kontrollierende Spieler gefühlt habe – eine wirklich bemerkenswerte Aussage. Doch so weit, so gut, wir haben diese Erfahrung schon beim Schach gemacht. Vielleicht kann der eine oder andere Leser, der sich gegen einen Schachcomputer versucht hat, sogar gut nachvollziehen, wie Lee Sedol sich in den Partien gegen AlphaGo gefühlt haben mag.

Nun wird die Angelegenheit aber doch noch ein wenig verrückter: 2017 hat das Team von Google DeepMind in der Fachzeitschrift Nature eine neue Version von AlphaGo vorgestellt: AlphaGo Zero. Der Name ist durchaus Programm, denn anders als die Vorgängerversion wurde die neue Version mit dem Zusatz „Zero“ an keinerlei historischen Spieldaten trainiert. Sie wusste anfangs nichts anderes über Go als die Spielregeln. Trotzdem ist sie in nur drei Tagen noch stärker gewesen als AlphaGo. Wie das, werden Sie zu Recht fragen. Nun, schlicht und ergreifend, indem das System gegen sich selbst gespielt hat, allein in den ersten drei Tagen 4,9 Millionen Partien. Dadurch war AlphaGo Zero schon nach wenigen Tagen in der Lage, es mit menschlichen Go-Meistern aufzunehmen. Die Vorgängerversion AlphaGo hatte dafür noch Monate des Trainings mit historischen Spieldaten gebraucht. Der Mitgründer von Google DeepMind meinte dazu lapidar, AlphaGo Zero sei deshalb so mächtig geworden, weil es durch den Wegfall der Trainingsdatenbasis nicht länger durch die Grenzen menschlichen Wissens und Könnens begrenzt gewesen sei. Dieses Prinzip ist aber natürlich nur dann anwendbar, wenn die zu modellierende Welt so überschaubar ist, dass sie vollständig simuliert werden kann – eben wie bei Go oder auch beim Schachspiel. Beim Poker und erst recht in allen komplexen Zusammenhängen, in denen mit der Irrationalität menschlichen Verhaltens zu rechnen ist, sieht die Sache schon anders aus. Vielleicht wird auch deshalb immer mehr in Richtung maschineller Selbstoptimierung geforscht. All das ist zugegebenermaßen nicht nur faszinierend, sondern auch ein wenig beängstigend, weil wir durch solche Ansätze über kurz oder lang die Kontrolle über die Technikentwicklung verlieren könnten.

Doch bei allem gebotenen Staunen, das auch mich dabei erfasst: Bei Schach, Go und Jeopardy, also den Beispielen, bei denen sich der Mensch bislang von Algorithmen geschlagen geben muss, handelt es sich um Spiele. Meiner Ansicht nach tun wir einstweilen gut daran, die Entwicklungen auf dem Gebiet der algorithmischen Entscheidungssysteme und KI nicht als Bedrohung zu begreifen. Wohl aber kommen sie – nach der kosmologischen, der biologischen und der psychologischen – in Zukunft oder vielleicht auch jetzt schon als mögliche vierte (nämlich technologische) Kränkung des Menschen in Betracht. Nur ist Gekränktsein keine zweckdienliche Haltung, wenn es darum geht, einen adäquaten Umgang mit etwas, das in der Welt ist und auch nicht mehr weggehen wird, zu finden. Angst vor KI ist ein ebenso schlechter Ratgeber.

Immer wieder sprechen manche von einer existentiellen Bedrohung der Menschheit durch KI, Tesla-Gründer Elon Musk und der kürzlich verstorbene britische Star-Physiker Steven Hawking zum Beispiel. Aufgrund der Fortschritte in Informatik und Robotik haben sie wiederholt öffentlich vor der Gefahr gewarnt, Maschinen könnten irgendwann, wenn ihre Fähigkeiten unsere erst einmal übersteigen, die Oberhand gewinnen und uns Menschen versklaven oder auslöschen. Musk, dem man kaum unterstellen möchte, dass er verrückt ist, denn immerhin ist er drauf und dran, sowohl die Personenbeförderung auf der Erde (Tesla, Hyperloop) als auch die Raumfahrt (SpaceX) zu revolutionieren, geht sogar davon aus, dass die Wahrscheinlichkeit, dass wir aktuell nicht in einer Computersimulation leben, sehr gering ist. Sein Argument besagt folgendes: Selbst wenn wir vom heutigen Stand der Dinge aus lediglich irgendeinen technischen Fortschritt in der Zukunft annehmen, dann wird die digitaltechnologische Fiktion durch Weiterentwicklungen in Virtual und Augmented Reality schon bald ununterscheidbar von der Realität sein. Und das, nachdem wir noch vor 40 Jahren lediglich triviale pixelige Spiele wie Ping Pong am Bildschirm spielen konnten. Angesichts dieser Geschwindigkeit sei es nicht sehr plausibel anzunehmen, dass eine solche Entwicklung im Laufe der Evolution nicht schon längst stattgefunden habe und uns seither all das, was wir für wirklich halten, was aber in Wahrheit virtuelle Realität ist, von mittlerweile

herrschenden Maschinen nur vorgegaukelt wird. Die Kinofilm-Trilogie „Matrix“ hat zwischen 1999 und 2003 eben diese Möglichkeit, dass wir bereits von Maschinen verklavt worden sind und dies alles nur träumen, aufwendig in Szene gesetzt. Bezeichnenderweise basiert der Film jedoch auf dem Gedankenexperiment „Gehirne im Tank“ des US-amerikanischen Philosophen Hilary Putnam, mit dem dieser gerade ausschließt, dass ein solcher Traum logisch überhaupt möglich ist. Das zeigt die Beliebtheit der Vereinnahmungen von Ideen durch die populäre Kultur. Sie macht sich alles zu eigen, was die Phantasie anregt und darauf hoffen lässt vermarktungsfähig zu sein. Und am Ende sind es dann tatsächlich die Bilder von Kinofilmen wie „Neuromancer“, „Terminator“ und „2001 – Odyssee im Weltraum“ oder auch Spike Jonez’ Meisterwerk „Her“, die unsere – letztlich falschen – Einschätzungen der Möglichkeiten und Gefahren von Technologie bestimmen.

### **Algorithmen zwischen Heilmittel und Teufelswerk**

Ein gutes Beispiel für unsere kulturell verursachte, aber realitätsferne Angstprägung in Bezug auf Technologie ist auch der Kinofilm „Minority Report“ aus dem Jahr 2002. Darin sagen durch den Konsum verunreinigter Drogen hirngeschädigte Personen, die in einer Art Trance-Zustand gehalten werden („Precogs“), zukünftige Verbrechen in der US-Hauptstadt Washington voraus, die von der Precrime-Einheit der dortigen Polizei dann tatsächlich seit sechs Jahren erfolgreich vereitelt werden. Der Film, bei dem Steven Spielberg Regie geführt und Tom Cruise die Hauptrolle gespielt hat, geht auf eine Science-Fiction-Kurzgeschichte von Philip K. Dick zurück. Heute, da sich algorithmische Entscheidungssysteme zur Unterstützung einer vorausschauenden Polizeiarbeit nicht nur in den angelsächsischen Ländern, sondern auch in sechs deutschen Bundesländern bereits im Einsatz befinden, wird in diesem Zusammenhang gerne auf diesen Film Bezug genommen. Dabei gibt es zwischen der Fiktion von „Minority Report“ und der heutigen algorithmisch

unterstützten Polizeiarbeit keinerlei sinnvollen Bezug außer den, dass die Polizei im Film regelmäßig und in der heutigen Wirklichkeit dort, wo solche Systeme im Einsatz sind, manchmal am Ort des Verbrechens ist, bevor das Verbrechen überhaupt geschieht.

Dieser Zusammenhang verweist, wenn man ihn so nüchtern darstellt, zugleich auf eines der zentralen Probleme prognostischer algorithmischer Systeme. Ein geflügeltes Wort besagt, dass Prognosen schwierig seien, insbesondere wenn sie die Zukunft betreffen. Das liegt schlicht daran, dass sich Prognosen auf statistische Zusammenhänge stützen, aber streng genommen keine Aussagen über Einzelfälle zulassen. Im hier diskutierten Zusammenhang führt diese Schwierigkeit zu folgendem Paradoxon: Wenn Verbrechen nicht mehr geschehen, weil sie mithilfe präziser Prädiktion und durch bloße Polizeipräsenz am vermeintlichen Ort des Geschehens beziehungsweise Nicht-Geschehens verhindert werden, dann haben wir eine Welt ohne Verbrechen, das heißt, was wir bekämpfen wollen, kommt nicht mehr vor, ist nicht mehr sichtbar. Angesichts dessen wird es nach einer Weile sehr schwierig sein, Ressourcen für etwas Unsichtbares, nicht Vorkommendes bereitzustellen. Dass dies keine bloße intellektuelle Spitzfindigkeit ist, zeigt die Tatsache, dass solche Debatten um das sogenannte Predictive Policing tatsächlich geführt werden, und zwar unter dem Stichwort Wirkungsmessung bei der Einführung und Evaluierung solcher Systeme. Es ist nämlich schwierig bis unmöglich, die Wirksamkeit einer Maßnahme zu untersuchen, die – wenn sie das bewirkt, was sie soll – dazu führt, dass der Grund für ihr Ergreifen im Grunde entfällt. Ein ähnliches Problem hat man übrigens bereits bei der Kalibrierung (d. h. funktionalen Optimierung) solcher Systeme. Dazu sind nämlich immer historische Deliktdaten erforderlich, doch diese sind insofern einseitig unvollständig, dass Delikte sich nur dort in Daten abgebildet haben können, wo es auch einen Beobachter (z. B. eine Polizeistreife) gab.

Im süddeutschen Raum wurde beispielsweise festgestellt, dass die Anzahl der Wohnungseinbrüche seit der Einführung einer Software, welche die Wahrscheinlichkeit von Wohnungseinbrüchen in einem Bezirk berechnet (woraufhin die Polizeipräsenz eben dort jeweils erhöht wurde), stellenweise signifikant

rückläufig ist. Doch ohne eine Isolierung und Messung anderer nachgewiesener und möglicher Einflussfaktoren und ohne einen Kontrollbezirk, in dem alles gleich ist, jedoch keine solche Verbrechensvorhersagesoftware eingesetzt wird, ist eine belastbare Aussage über die Wirksamkeit vorausschauender Polizeiarbeit auf Software-Basis nicht möglich. In Deutschland jedoch sind solche algorithmischen Prognosesysteme zwar bereits in sechs Bundesländern im Einsatz, doch nur in einem wurde bislang der Versuch einer kontrollierten Wirkungsmessung vorgenommen. Dieser Umgang mit Technik ist leider verbreitet: Ein neues technisches System bewirkt irgendetwas, und weil wir mit dem Status quo in irgendeinem Bereich (z. B. steigenden Wohnungseinbruchszahlen) nicht zufrieden sind, sind wir bereit, diese Technik einigermaßen unreflektiert in diesem Bereich anzuwenden. Dadurch kann es aber noch schlimmer kommen, denn wenn die Briten, die ebenfalls mit steigenden Einbruchszahlen unzufrieden sind, Predictive-Policing-Software einsetzen und damit dann zufriedener sind als zuvor, dann können deutsche Politiker der deutschen Öffentlichkeit nicht leicht über längere Zeit erklären, in Deutschland könne man aus diesen oder jenen Gründen nicht das Gleiche tun.

Das liegt auch daran, dass uns der Sinn für etwas, das über Jahrhunderte ein sicherer Kompass für wissenschaftlich begründbare Wahrheit gewesen ist, im Datenzeitalter, in dem wir ja mittlerweile schon eine Weile leben, allmählich abhanden kommt. Indem uns beinahe allgegenwärtige Datenanalysen und Statistiken (ganz absurd ist diese Entwicklung im Profifußball) ständig Korrelationen aufzeigen, legen sie Kausalzusammenhänge nahe, wo keine sind. Der menschliche Geist neigt nun einmal zu dem klassischen Fehlschluss „Post hoc, ergo propter hoc“ (Es ist unmittelbar danach passiert, also ist es deswegen passiert). Unbedarfte Menschen sehen also plötzlich vielleicht lauter Kausalzusammenhänge, wo keine sind. Menschen, die zwar kritisch, aber auch lösungsorientiert sind, könnten sich hingegen immer mehr mit Korrelationen zufrieden geben, ohne den tatsächlichen (technischen, sozialen, natürlichen) Wirkmechanismen auf den Grund zu gehen. Oft reicht das für erfolgreiche punktuelle Interventionen aus, doch sobald Problemlagen so beschaffen sind, dass sie eine systemische Betrachtungsweise erfordern, wird es mit der reinen

Korrelationsanalyse schwierig. Zur Hypothesengenerierung ist die Betrachtung von Korrelationen dagegen ein durchaus probates Mittel.

Der entscheidende Wandel besteht darin, dass die algorithmische Datenanalyse an vielen Stellen, an denen wir früher die Wirkzusammenhänge von technischen, sozialen oder natürlichen Systemen zunächst durchdringen mussten, um sie beeinflussen zu können, diese verstehende Durchdringung vermeintlich oder tatsächlich erübrigt. Dies wiederum verändert unser Verhältnis zur Welt. Wenn ich es nicht mehr gewohnt bin, einfachen Dingen auf den Grund zu gehen, weil Datenanalysen immer schon die zur erfolgreichen Einflussnahme nötigen Korrelationen aufzeigen, dann fällt es mir irgendwann natürlich schwerer, wirklich harte Probleme zu lösen, bei denen ich um die Aufdeckung von Kausalzusammenhängen und das Malen eines großen Gesamtbildes nicht umhinkomme. Die Möglichkeiten beispielsweise, mittels sozialer Netzwerke wie Facebook politische Meinungen und Wahlen gezielt zu beeinflussen, wird man nicht dadurch beseitigen können, dass man den Code, aus dem Facebook besteht, ein bisschen modifiziert. Vielmehr muss man das Gesamtsystem der werbefinanzierten Online-Netzwerke analysieren, um seriös Abhilfe vorschlagen zu können.

Wahr ist auf der anderen Seite aber auch, dass wir zahlreichen Herausforderungen unserer Zeit nur werden begegnen können, indem wir elaborierte technische Möglichkeiten der Analyse von Daten und des Schlussfolgerns daraus (einschließlich des Vorbereitens und Treffens von Entscheidungen) nutzen. Zu diesen Herausforderungen gehört übrigens die algorithmische Datenverarbeitung selbst. Denn auch die Frage, ob Algorithmen aufgrund der Zugehörigkeit zu einem Geschlecht, einer ethnischen Gruppe, einer Altersklasse oder aufgrund anderer Merkmale diskriminieren, lässt sich aufgrund der großen Datenmengen, um deren Verarbeitung es beispielsweise in Einstellungsverfahren geht, eigentlich nur mit maschineller Hilfe zuverlässig beantworten und eventuell beheben. Der Begriff der Diskriminierung schillert hier, denn die rasche Unterscheidung von Einzeldingen nach Klassen ist gerade einer der Hauptvorteile maschineller Analysesysteme. Hier wird deutlich, dass wir oftmals der Technik zu Unrecht Resultate anlasten, die wir als nicht wünschens-

wert erachten, während vielmehr unser Gebrauch von der Technik fragwürdig ist. Das lässt sich an einem weiteren Beispiel illustrieren:

Die automatische maschinelle Auswertung von Satellitenbildern lässt sich einerseits dazu nutzen, vormals kompliziert herbeizuführende entwicklungs- politische Entscheidungen über die Verteilung von Ressourcen zu vereinfachen, denn die Muster auf diesen Bildern erlauben recht verlässliche Rückschlüsse auf soziodemographische Verhältnisse. Aus dem gleichen Grund kann diese Musteranalyse von Satellitenbildern allerdings zur schönen Gewinnmaximierung im Investmentbanking angewendet werden; bei Goldman Sachs ist dafür der Bereich „Quantitative Investment Strategies“ zuständig. Es ist also ein Dual Use, wie es in der Fachsprache heißt, eine Nutzung derselben Technik auf zwei ganz unterschiedliche Arten und zu unterschiedlichen Zwecken möglich. Diese Zwecke können noch verschiedenartiger sein als in den angeführten Beispielen, z. B. die Zustellung von Paketen und die Tötung von Menschen durch Drohnen (was im Übrigen dem klassischen Verständnis des Begriffs Dual Use entspricht). Doch der springende Punkt wird auch an einem Fall wie diesem deutlich: Von diesem Monat (April 2018) an plant die Bundesregierung den standardisierten Einsatz einer Software zur automatisierten Dialektanalyse bei Asylbewerbern, auch Sprachbiometrie genannt. Damit soll verifiziert werden, ob jemand tatsächlich aus der Erdgegend stammt, aus der er zu stammen vorgibt. Vergessen werden sollte dabei allerdings nicht, dass die Gründe für mögliche Abweichungen auch andere als Falschangaben oder Täuschungsversuche der Geflüchteten sein können. Außerdem ließe sich ein solches System auch zu einem ganz anderen, positiven Zweck einsetzen, nämlich für die Verbesserung von (sprachlichen) Integrationsmaßnahmen.

Vor dem Hintergrund, dass Technik zunächst neutral ist und es daher stark auf ihren Gebrauch durch uns ankommt, ist es vielleicht wenig verwunderlich, dass es Algorithmen in der öffentlichen Debatte nicht leicht haben. Einerseits sollen sie Probleme wie die Digitalisierung der Verwaltung oder die Jagd auf Einbrecher und Terroristen lösen helfen, andererseits stehen sie unter Verdacht, die menschliche Souveränität schlechthin zu untergraben. Dass die deutsche Grundhaltung eher Skepsis ist, wird etwa deutlich an der deutschen

Übersetzung des – wichtigen und guten – Buches „Weapons of math destruction“ von Cathy O’Neill, die „Angriff der Algorithmen“ lautet. Diese Titelgebung verkennt nicht nur, dass Algorithmen – jedenfalls bis auf weiteres – nicht vom Himmel fallen, sondern sie suggeriert auch, dass sie eine grundsätzliche Gefahr für die Menschheit darstellen, was in dieser pauschalen Form natürlich Unsinn ist. Dagegen brauchen wir einen unaufgeregten, arbeitsteilig organisierten Diskurs über die Möglichkeiten und Grenzen des Einsatzes von Algorithmen. Diese Nüchternheit ist nicht zuletzt deshalb angebracht, weil es letzten Endes um nicht mehr und nicht weniger geht als die Frage, wie wir im sogenannten Datenzeitalter eine lebenswerte Gesellschaft sein und bleiben können.

In diesem Zusammenhang ist es sicher angebracht, offenkundige Schwächen und Schwierigkeiten zu untersuchen und möglichst zu beheben. Ebenso notwendig ist allerdings die Frage, was durch den verstärkten Algorithmeninsatz neu möglich ist, was vorher nicht möglich war. Verschärft könnte man auch sagen: Abgesehen vom Einsatz guter Software, die so programmiert ist, dass sie laufend dazu lernt, haben wir überhaupt keine Chance, aus den gewaltigen Datenmassen, die uns heute zur Verfügung stehen, als Gesellschaft Nutzen zu ziehen. Dass Algorithmen bislang vor allem überall dort, wo sich mit ihrer Hilfe (viel) Geld verdienen lässt, eine herausgehobene und nicht immer rühmliche, gesellschaftsdienliche Rolle spielen, spricht nicht gegen Algorithmen per se, sondern illustriert, dass wir vielleicht nicht alles, aber doch sehr vieles wirtschaftlichen Interessen unterzuordnen bereit sind. Das ist nicht erst bei Algorithmen der Fall, sondern es war bei früheren technologischen Innovationen nicht anders. Die bedeutendsten Innovationstreiber waren fast immer das Militär und diejenigen mit den größten wirtschaftlichen Gewinnaussichten. Die Quintessenz daraus: Wir müssen aufpassen, dass wir nicht an Stellen, an denen wir eigentlich gesellschaftspolitische Debatten führen müssten, einen Technologiediskurs anzetteln – und das oft genug auch noch auf einer dürftigen Wissensgrundlage.



## Das ausgezeichnete Potential von Algorithmen

Tatsächlich kann das Zusammenwirken von Daten und Algorithmen viele Aufgaben, für die Menschen schlussfolgerndes Denken und implizites Können brauchen, fehlerfreier, effizienter und effektiver lösen, darunter zahlreiche Allokationsaufgaben, auch und gerade in der öffentlichen Verwaltung. Bei anderen Herausforderungen, beispielsweise im medizinisch-diagnostischen Bereich, etwa der Krebserkennung und -therapie oder bei der Erkennung und Behandlung seltener Krankheiten, haben sich Computerprogramme als hervorragende Ergänzung menschlichen Wissens und ärztlicher Intuition herausgestellt. Zuletzt gibt es Dinge, die nur Maschinen leisten können, weil unsere kognitiven Kapazitäten bei weitem nicht ausreichen. So ist eine inhaltliche Auswertung und Klassifizierung sämtlicher Forschungsergebnisse, die alleine in kleinen Teilbereichen der medizinischen Forschung heute publiziert werden, nur mit maschineller Unterstützung überhaupt möglich.

In der sogenannten vorausschauenden Wartung (Predictive Maintenance) werden durch die Erfassung, Zusammenführung und Analyse riesiger Datenmengen in Echtzeit Fehler oder Ermüdungen am Material erkannt, bevor es zu Katastrophen kommt. Im Flugverkehr beispielsweise ist es ein nicht von der Hand zu weisender Vorzug, dass Teile gezielt ausgewechselt werden können, bevor sie versagen. Doch auch weit unterhalb dieser Ebene, in den Komfortzonen unseres Lebens, übernehmen algorithmische Systeme mittlerweile Funktionen, auf die viele nicht mehr gerne verzichten würden. Maschinen geben Menschen etwa Empfehlungen, die auf andere Art nicht zu haben wären: Ich habe nicht genug Freunde und Bekannte, um all die Literatur und all die Musik über persönliche Empfehlungen zu entdecken, die mir Amazon und Spotify schon empfohlen haben.

Was zeichnet Algorithmen ganz allgemein aus, welche Vorzüge haben sie gegenüber der datenverarbeitenden Biomasse unseres Gehirns? Nun, sie arbeiten konsistent das ab, wozu sie erstellt wurden. Sie machen keine Flüchtigkeitsfehler, die Menschen regelmäßig unterlaufen – was nicht heißen soll, dass sie nicht katastrophal falsche Entscheidungen treffen können, wenn sie entsprechend

schlecht konzipiert wurden. Was gibt es noch an übergeordnet Positivem anzumerken? Das Wertefundament von Algorithmen ist, anders als das von menschlichen Entscheidungsträgern, programmierbar und damit grundsätzlich kontrollierbar. Die Ergebnisse algorithmischer Entscheidungen lassen sich, zumindest im Prinzip, im Nachhinein überprüfen und korrigieren. Durch die Automatisierung von Prozessen werden Menschen von Routinetätigkeiten entlastet, was Freiräume für die Beschäftigung mit drängenderen Problemen und sogar mehr Zeit für menschliche Interaktion (zum Beispiel im Verhältnis zwischen Arzt und Patient oder zwischen Verwaltung und Bürgern) schaffen kann. Angebote und Dienste lassen sich leichter personalisieren oder auf Gruppen zuschneiden.

Eine Algorithmenentwicklung, die ihren essenziell sozialen Charakter nicht verkennt (ja, Programmierer sind Schöpfer *soziotechnischer* Systeme!), hat den weiteren Vorzug, dass im Zuge dessen viele bereits zuvor bestehende Probleme und Schieflagen erkannt und adressiert werden können, weil Daten Grundlagen, implizite Annahmen und Vorgehensweisen explizit gemacht werden müssen, bevor sie Gegenstand von Programmcode werden können. Wenn ich etwa Menschen mit dunkler Hautfarbe vom Zutritt zu einem Bus mit automatisch öffnenden und schließenden Türen ausschließen wollte, so müsste ich dem System einen entsprechenden, explizit programmierten Hautfarbendetektor zugeben, während solcherlei Dinge in den USA und in Südafrika zum Teil noch lange nach dem Ende der entsprechenden offiziellen Staatsraison eine mehr oder weniger offensichtliche soziale Praxis war, der nicht leicht beizukommen ist. Dem Code eines nach Hautfarbe selektierenden automatischen Busses sehe ich den Rassismus in der Regel dagegen sofort an; der entsprechend funktionale Teil dieses Codes lässt sich dann isolieren und anfechten.

## Transparenz und Nachvollziehbarkeit algorithmischer Systeme

Schwieriger wird es, wenn beim Zutritt zu jenem Bus nicht nur nach Menschen mit verschiedener Hautfarbe, sondern auch nach den Kategorien Mensch und Tier unterschieden werden soll. Das mag zunächst widersinnig erscheinen, dem liegt aber ein tatsächliches algorithmisches Problem zugrunde: Die Bilderkategorisierung des kostenlosen Bilderverwaltungsprogramms *Google Photos* vermochte es auch im Jahr 2017 nicht, farbigen Menschen mit an Sicherheit grenzender Wahrscheinlichkeit von Affen zu unterscheiden und ordnete in manchen Fällen Fotos, auf denen Farbige abgebildet waren, der Kategorie „Gorilla“ zu. Die entsprechende Funktion musste kurzerhand außer Betrieb gesetzt werden, weil der potenzielle Reputationsschaden für Google gewaltig zu werden drohte.

Es ist ein vielfach in der Geschichte der Automatisierung zu beobachtendes Phänomen, dass die ersten 90 Prozent relativ leicht zu erzielen sind, während die letzten zehn Prozent eine große Herausforderung darstellen. Um sagen zu können, warum das so ist, bräuchten wir genaue Kenntnisse von der Funktionsweise der entsprechenden Algorithmen. Aber die sind bei den heute gebräuchlichen Systemen wie *Google Translate* oder *Google Photos* nicht leicht zu erzielen.

Der Durchbruch in der Bilderkennung kam mit der Wiederentdeckung neuronaler Netze als algorithmische Funktionsgrundlage, in Verbindung mit der Verfügbarkeit großer Mengen annotierter Bilddaten durch das Internet. Ganz grob vereinfacht kann man sagen, dass sich neuronale Netze das informationsverarbeitende Prinzip des Gehirns zunutze machen, ohne es allerdings nachbilden zu wollen. Bei neuronalen Netzen handelt es sich zumindest teilweise immer um implizite Programmierung. Das bedeutet, dass der korrekte Lösungsweg nicht explizit vom Programmcode vorgegeben ist, sondern im Zuge der Kalibration erst gefunden wird. Gefunden wird er in Testläufen an sogenannten Trainingsdaten. Algorithmen ‚lernen‘ an bereits entsprechend ausgewiesenen Daten und können dieses ‚Wissen‘ dann auf neue, noch unbekannte Fälle anwenden. Die Verwendung der Begriffe lernen und Wissen ist hier

streng genommen falsch, die Anführungszeichen müssten deshalb größer sein, als sie es sind. Technisch korrekt müsste man sagen, dass Mustererkennungsverfahren auf annotierten historischen Daten kalibriert werden – was sehr viel weniger nach mystischer, dem Menschen ähnlicher Maschinenleistung klingt.

Ein solches Maschinenlernen hat sich bei der Lösung mancher Klassen von Aufgaben, beispielsweise der Bilderkennung, als sehr wirksam erwiesen. Diese größere Wirksamkeit gegenüber klassischen Programmieransätzen hat jedoch einen Preis, und zwar die größere Intransparenz dessen, was im Zuge der Problemlösung im Innern der Maschine eigentlich passiert. Je nach Anwendungsbereich (z. B. bei selbstfahrenden Autos, die ihre Umwelt erkennen und klassifizieren können müssen) kann das ein Problem darstellen. Unabhängig davon können uns avancierte selbstlernende algorithmische Systeme mit ihrem Lösungsweg überraschen, weil ihre Beurteilungskriterien stark von menschlichen abweichen. Beispielsweise ist für einen Bilderkennungsalgorithmus zur Einschätzung des Geschlechts der Personen, die auf den zu klassifizierenden Bildern zu sehen sind, auch der Faltenwurf der Kleidung dieser Personen von Bedeutung, während kein Mensch, den ich zumindest kenne, dies als ein nennenswertes Kriterium seiner Einschätzung anführen würde. Das erstreckt sich auf alle Bereiche, in denen wir Maschinen menschliche Probleme lösen lassen. Bei den Spielen der Google-KI „AlphaGoZero“, die sich, wie oben beschrieben, durch Trainingsspiele gegen sich selbst erfolgreiche Strategien beigebracht hatte, haben die menschlichen Go-Meister zu Protokoll gegeben, dabei seien Züge zu beobachten gewesen, die ihnen völlig unorthodox und geradezu unmenschlich vorkamen.

Wir sollten uns beim verstärkten Einsatz von KI nicht darüber wundern, dass selbstlernende Systeme mit einer anderen Logik erfolgreich sind als wir. Schließlich sind es keine Nachbauten unserer selbst oder unseres Gehirns, sondern Maschinen, die mit dem Ziel erschaffen wurden, uns auf ihre eigene Art in bestimmten Teilbereichen funktional überlegen zu sein. Aber eben nur dort! Unsere Logik erschließt sich ein tatsächliches Verständnis der Welt, kann also eine viel höhere Allgemeingültigkeit als heutige KI-Systeme für sich beanspruchen. Den Anspruch, prinzipiell immer nachvollziehen zu können, was

im Innern der Maschinen vor sich geht, dürfen wir aber nicht um möglicher Leistungssprünge in einzelnen kleinen, funktional differenzierten Bereichen willen aufgeben. Denn ein solches Eigenleben von Maschinen, die in Lebensbereichen von Menschen Entscheidungen vorbereiten oder sogar autonom treffen, ist nicht mit unseren basalen Regeln des Zusammenlebens, nicht mit unserem Strafrecht, nicht mit unserer Ethik und nicht mit unserer freiheitlichen demokratischen Grundordnung vereinbar. Aber noch weit davor – und das weiß niemand besser als der klassische Ingenieur, und damit wissen es auch die heutigen Designer autonomer Fahrzeuge in den Schmieden der deutschen Automobilhersteller – ist Nachvollziehbarkeit ein Gebot sowohl der technischen als auch der ökonomischen Vernunft: Opake Systeme können nicht gewartet werden, das Wissen über sie kann nicht weitergegeben werden und sie lassen sich aus eben diesen Gründen am Ende auch nicht verkaufen.

Das ist allerdings kein Plädoyer gegen technischen Fortschritt. An dieser Stelle muss mit einer weit verbreiteten falschen Annahme aufgeräumt werden, nämlich dass eigenständig lernfähige Systeme, die auf neuronalen Netzen basieren können (aber nicht müssen), notwendig undurchschaubar sind. Dies ist ein Märchen, das im Wesentlichen dadurch entstanden ist, dass sich zu viele Nicht-Computerwissenschaftler (sprich Menschen, die technisch gar keine Ahnung haben) zu diesem Thema äußern. Richtig ist dies: Selbstlernende Systeme nachvollziehbar zu machen, erfordert zusätzlichen technischen Aufwand, der nicht gescheut oder aus Kostengründen gemieden werden darf. Das bedeutet, dass man gerade Systemen mit implizit programmierten Lösungswegen (etwa neuronalen Netzen) explizit den Zusatzauftrag erteilen muss, über die inneren Abläufe in einer für den Menschen verständlichen Weise zu berichten. Das ist natürlich nicht trivial, aber daran wird – auch in Deutschland – erfolgreich geforscht. Der dazu nötige Aufwand müsste vielleicht sogar durch politische Entscheidungen verbindlich gemacht werden, damit wir einerseits technischen Fortschritt und andererseits technische Strukturen haben können, die mit unseren Grundüberzeugungen und Institutionen vereinbar sind.

Dass viele Systeme, darunter die am meisten verbreiteten, dennoch intransparent sind, hat eher mit wirtschaftlichen Anreizstrukturen als mit inhärenten

technischen Bedingungen zu tun, wie man an der Online-Werbeindustrie gut sehen kann. Das Verschleierungsprinzip von Facebook basiert auf einem System, das in kurzer Zeit sowohl funktional als auch quantitativ sehr stark angewachsen ist und mit dem jedes Quartal ein Milliarden Gewinn eingefahren wird; weniger hat es mit übelmeinenden Verantwortlichen oder Programmierern zu tun. Der Code ist nicht böse und wird deshalb im Dunkeln gehalten, sondern er ist komplex und profitabel. Die inneren Mechanismen des schnell gewachsenen, gewaltigen Softwaresystems mit unzähligen angeschlossenen Drittanwendungen sind undurchsichtig, und Facebook hat aus wirtschaftlichen Gründen kein Interesse daran, sie zu transparent zu machen. Allerdings könnte der gesellschaftspolitische Druck in der Folge der im März 2018 neuerlich entfachten Diskussion um die Beeinflussung der Wahl zum US-Präsidenten im Jahr 2016 so sehr zunehmen, dass die Offenlegung des technischen Hintergrundes und eventuell sogar eine Beschneidung von Funktionen zulasten der Profitabilität unausweichlich werden wird.

Algorithmen sind nichts Mystisches. In einfache Bestandteile zerlegt, lassen sie sich recht leicht erklären, denn sie bringen letztlich ja nur unsere Werte und Entscheidungen in eine von Computern interpretierbare Sprache. Natürlich sind für jeden verständliche Erklärungen komplexer Softwaresysteme (und die einflussreichsten Programme unserer Tage sind sehr komplex) schwierig zu erzielen. Der springende Punkt ist aber, dass Intransparenz dem Code nicht per se innewohnt. Eines Tages wird der Code vielleicht ausschließlich von Maschinen geschrieben oder repliziert sich selbst. Aber solange das nicht so ist und es die Singularität noch nicht gibt, geht die Intransparenz von Algorithmen letzten Endes immer auf menschliches Verhalten zurück. Und dieses Verhalten muss bestimmte Kriterien erfüllen und seine Resultate müssen bestimmten Anforderungen genügen, falls es uns wichtig genug ist, unser Leben nur von solchen Algorithmen durchweben zu lassen, die wir zumindest im Prinzip nachvollziehen können.

Ein Beispiel: Kürzlich ist im US-Bundesstaat Arizona eine Frau von einem selbstfahrenden SUV des Fahrdienstes Uber getötet worden. Gerade weil es sich um eine noch im Teststatus befindliche Technologie handelt, von der aus-

zugehen ist, dass sie in recht naher Zukunft massenhaft zum Einsatz kommen wird, möchte man natürlich gerne wissen, wie es zu diesem tragischen Unfall kommen konnte. Wir gehen mit Recht davon aus, dass selbstfahrende Autos so programmiert sind, dass sie keine Menschen töten, und zwar egal, wie ungeschickt oder auch fahrlässig sich Personen auch verhalten mögen. Schließlich wird die Entwicklung (und Einführung) dieser Technologie ganz wesentlich damit begründet, dass man das Leben der vielen Menschen, die jährlich durch menschliches Versagen im Straßenverkehr sterben, retten möchte. Was hat im Arizona-Fall also dazu geführt, dass diese Rechnung nicht aufgegangen ist? Bereits heute treffen Algorithmen ganz oder teilweise automatisch für Menschen verbindliche Entscheidungen in folgenden Bereichen: Kreditvergabe, Zuweisung von Schul- und Studienplätzen, Bewerberauswahl in Einstellungsverfahren, Ausstellung von Steuerbescheiden, Erstellung von Krankheitsdiagnosen, Auswertung wissenschaftlicher Forschungsliteratur, Rechtsberatung in Standardfällen, Kundensupport, Steuerung von Waffensystemen, Entscheidungen über die Aussetzung von Haftstrafen zur Bewährung. Natürlich ist diese Liste unvollständig, aber sie mag dennoch einen Eindruck davon vermitteln, welchen Stellenwert algorithmische Analyse- und Entscheidungssysteme in unserer Welt bereits haben, und wie wichtig es ist, dass wir uns spätestens jetzt auf gesellschaftskompatible Designprinzipien und eventuell auch auf Stoppregeln verständigen müssen. Auf beides komme ich am Ende zurück.

### **Einsatz von Technik für das Erreichen menschlicher Ziele**

Ein Algorithmus knüpft unmittelbar an Zusammenhänge in der physischen Welt an. Um zu funktionieren, braucht er keine von und für Menschen gemachten Gründe – Korrelationen reichen aus. Im Vergleich zur menschlichen Wahrnehmung und Informationsverarbeitung erfüllt er seinen Zweck also mit anderen, eigenen Mitteln. Das ist in der Geschichte der Technikentwicklung ein durchaus altbekanntes Phänomen: Auch Flugzeuge etwa sind nach eigenen

Regeln konstruiert worden und nicht mit der Absicht, naturgetreu Vögel nachzuzahlen. Überhaupt sind viele Phänomene des Berechenbar-Machens und der Automatisierung, mit denen wir uns heute zu Recht intensiv befassen, weil sie immer mehr an Bedeutung gewinnen, qualitativ nicht wirklich neu.

Der vermehrte Einsatz von Algorithmen ist – wie wissenschaftliche Experimente, die seit Jahrhunderten durchgeführt werden – eine Folge unseres Erkenntniswillens und des Strebens, uns die Welt untertan zu machen. Wir wenden dieses, uns aus der analogen Zeit bereits vertraute und nun digitale Werkzeug auf die Welt an, die wir vorfinden. Und diese Welt ist eine von Daten geflutete Welt. Daher ist es kaum verwunderlich, dass der verbreitete Einsatz von Algorithmen in dieser unübersichtlichen Datenwelt experimentelle Züge aufweist. Allerdings gleicht er nicht kontrollierten Experimenten, wie sie in geschützten Labor-Räumen durchgeführt werden, vielmehr ähnelt er großen Gesellschaftsexperimenten wie der Nutzung der Atomkraft, dem sozial-ökologischen Design von Landschaften oder der durch massive Subventionen angefeuerten deutschen Energiewende, bei denen man unmöglich alle Variablen isolieren und nicht alle Parameter kontrollieren kann. Das kann zwar gutgehen, ist aber grundsätzlich gefährlich. Big Data Analytics (also der Einsatz von Algorithmen, um aus einem riesigen Datenmeer Erkenntnisse zu fischen), das für das Leben der meisten von uns heute zentrale Anwendungsfeld von Algorithmen, hat zwar sehr viele Annehmlichkeiten gebracht, uns aber auch zu Objekten des sogenannten „Überwachungskapitalismus“ von Google, Facebook, Amazon und vielen anderen gemacht. Big Data Analytics hat zwar das Zeug dazu, etliche Leistungen der öffentlichen Hand wesentlich zu verbessern oder überhaupt erst zu ermöglichen, doch es hat auch zu staatlicher Überwachung geführt, deren wahres Ausmaß uns Edward Snowden 2013 durch seine Enthüllungen vor Augen geführt hat.

Nun sind Innovationen aus Daten im Grunde immer offene Entdeckungsverfahren. Das heißt, wir wissen gar nicht, was wir suchen, bis wir es gefunden haben. Eben deshalb wird vielfach ja ein möglichst vollständiges Datenspektrum angestrebt, werden sogenannte Datenseen (Data Lakes) angelegt. Doch wollen wir uns diesbezüglich wirklich damit begnügen, „dass Innovationen im



Wesentlichen Innovationen einer kleinen reichen Minderheit für eine andere kleine reiche Minderheit sind“, wie es der Wissenschaftsjournalist Rangar Yogeshwar einmal ausdrückte? Dieses Thema, das aber in der öffentlichen und politischen Debatte mittlerweile zu Recht einen gewissen Raum einnimmt, ist kein genuin technisches: Natürlich müssen wir als Gesellschaft Wege finden, wie nicht nur wenige, sondern viele, im Idealfall ganze Gesellschaften an den durch immer umfangreichere algorithmische Datenanalysen geschaffenen Mehrwerten teilhaben können. Es geht dabei um nicht weniger als die Frage, welche von Daten befeuerte und durch Algorithmen angetriebene Gesellschaft wir sein wollen. Datensouveränität möglichst vieler sowie digitale Aufklärung und Bildung sind sicher ein Schlüssel. Ein blindes Sich-Ausliefern an das, was der israelische Philosoph Yuval Harari treffend als „Dataismus“ (das ist die unbedingte Überzeugung, dass die ständige Erzeugung und Verarbeitung von Daten etwas an sich Wertvolles und zu Förderndes sei) genannt hat, führt dagegen mit Sicherheit auf gefährliche Abwege, wie wir vielerorts sehen können.

### **Automatisierung mit oder ohne Algorithmen: nicht um jeden Preis**

Kehren wir zum Anfang der Überlegungen zurück und fragen erneut nach dem Wert unserer Fähigkeit, Prozesse – mentale ebenso wie physische – zu automatisieren. Wie mehrfach betont, hat uns diese Fähigkeit die Welt untertan gemacht und uns an die Spitze der Schöpfung gestellt. Nun hat diese Geschichte nicht immer und für alle nur positive Auswirkungen gehabt, wenn wir etwa an den Übergang von Manufakturen zur standardisierten Fließbandproduktion denken. Das wird auch bei der derzeit anstehenden Automatisierung kognitiver Arbeit („white collar automation“) gelten. Zudem kann Automatisierung seltsame Blüten treiben und regelrecht schiefgehen. Dafür gibt es viele Beispiele in den einfachsten Bereichen unserer automatisierten physischen Umwelt: Kaffeemaschinen, die keinen Startschalter mehr haben, sondern mit dem Kaffeemachen beginnen, sobald eine Kapsel eingelegt und der Schacht

geschlossen ist, lassen sich ebenso kaum noch adäquat nutzen wie Kofferraumklappen, die sich nur per Knopfdruck elektrisch öffnen und schließen lassen oder Staubsaugerschnüre, für deren Einfahren es eben keinen Aktivierungsknopf mehr gibt; stattdessen machen sie es, wenn man kurz ruckartig an ihnen zieht, was man leider beim Staubsaugen ständig tut, so dass der ‚Schnureinfahralgorithmus‘ ständig ungewollt aktiviert ist – mit allen der Wohnungssauberkeit abträglichen Folgen. Das Büro, in dem ich arbeite, ‚profitiert‘ von einer ‚intelligenten‘ Heizungssteuerung für das gesamte Haus; das führt dazu, dass in unserem Konferenzraum, der sich in drei Teile aufteilen lässt, der hintere regelmäßig sehr heiß ist, während es im mittleren sehr kalt ist, weil das Zentralsystem nur jeden zweiten Heizkörper aktiviert. Wie gerne würden wir die Thermostate von Hand bedienen, was sich aber leider nicht für ein Büro einzeln so einstellen lässt. Der langen Beispielrede kurzer Sinn ist hier: Wie alles, so kann der Mensch auch die Automatisierung (und damit das, was ihn zum Herrscher der Welt gemacht hat) übertreiben und sich damit selbst schaden.

Deshalb müssen wir uns endlich trauen, an bestimmten Stellen Stoppregeln zu definieren und auch im negativen, verneinenden Sinne visionär zu sein. Die Europäische Union hat in der Datenschutzgrundverordnung, die weitreichende Folgen für die Datenverarbeitungspraktiken von Unternehmen haben und im Mai 2018 in Kraft treten wird, beispielsweise vorausschauend bereits festgelegt, dass kein Unionsbürger zum Objekt einer reinen Maschinenentscheidung werden darf. Das heißt, dass jeder Unionsbürger das Recht hat, sich von einem Menschen die Zusammenhänge einer maschinell getroffenen Entscheidung erläutern zu lassen und diese im Zweifelsfall anzufechten. Diese Regelung löst noch nicht alle Probleme, die der Vormarsch algorithmischer Entscheidungssysteme mit sich bringt, aber es ist gleichwohl ein Schritt in die richtige Richtung. Nur wenn wir es schaffen, uns im Beurteilen und Entscheiden sowohl vom Zauber affirmativer Technikbegeisterung als auch vom Schrecken weltstürzender Technikphobie zu befreien, können wir im Datenzeitalter souverän agieren. Wir brauchen mehr Nüchternheit im Umgang mit neuen technischen Möglichkeiten – und wir brauchen:

## Sozialinklusive Technikentwicklung mit ethischem Bewusstsein

Derzeit stehen sich zwei verschiedene Herangehensweisen derer, die möchten, dass Technik die Welt verbessert, einander gegenüber. Erstens: Man schafft gut durchdachte, richtig gut funktionierende Technik (iPhone!) und wendet sie auf alles an, was auch nur im Verdacht steht, dadurch verbessert oder gelöst werden zu können. Zweitens: Man schafft, wo es sinnvoll ist, spezielle Technik zur Lösung ganz spezifischer Probleme zusammen mit den besonders davon betroffenen Personen. Den ersten Ansatz finden wir heute meistens vor. Das Problem daran ist, dass hier der Hammer-und-Nagel-Effekt droht: Wer einen anständigen Hammer in der Hand hält, dem scheint die Welt voller Nägel zu sein. Die Technik ist hier immer schon gesetzt, sie steht an erster Stelle, weil sie ja nunmal da ist und fortwährend neu entsteht. Man sucht nach dem passenden Problem für die Technik (Kritiker nennen diesen Ansatz „solutionism“).

Beim zweiten Ansatz steht gar nicht von vornherein fest, ob ein Problem den Einsatz von (neuer oder schon etablierter) Technik erfordert oder nicht. Schließlich hat unsere Welt auch im 21. Jahrhundert noch viele weitere Dimensionen. Und wenn es sich doch herausstellen sollte, dass Technik einen Beitrag leisten kann, dann wird sie von Technik-Experten (z. B. Programmierern) zusammen mit denjenigen geschaffen, die vom Einsatz dieser Technik mutmaßlich unmittelbar betroffen sind (z. B. ÖPNV-Nutzer von einem neuen, nachfrageorientierten dynamischen Fahrplansystem, genannt „ÖPNV on demand“). Allein dieser Ansatz, Ko-Design und Ko-Kreation in transdisziplinären Teams, trägt dem essentiell sozialen Charakter heutiger Technikentwicklung Rechnung. Das soziale Element bei der Entwicklung soziotechnischer Systeme darf, wie es heute leider allzu oft geschieht, nicht vernachlässigt werden. Das ist zunächst kein ethisch relevantes Feld, sondern eine Frage der Vernunft und des Willens, nachhaltige gesellschaftliche Strukturen zu schaffen, beziehungsweise zu erhalten. Doch auch die Ethik kann angesichts der Reichweite und Potenz von Algorithmen und KI mittlerweile schnell ins Spiel kommen (siehe zum Beispiel das selbstfahrende Auto). Deshalb darf die Vermittlung von Fertigkeiten zum Schaffen solcher Systeme heute nicht mehr

ohne die gleichzeitige Vermittlung entsprechender soziologischer und philosophischer Grundlagen auskommen. Was soll das heißen, werden Sie jetzt vielleicht fragen: Soziologie- und Philosophie-Curriculum für Programmierer? Genau das, und zwar von Semester eins an. Denn ein „Code first, fix later“ (Erst Programmieren, später Fehler beheben), das wir heute allenthalben sehen, wird nicht mehr lange gutgehen und eines unbestimmten zukünftigen Tages mit einiger Sicherheit in die Katastrophe führen.

Der Untertitel dieses Aufsatzes ist aus Überzeugung positiv formuliert, und so sollen die Gedanken auch ausklingen: Der Mensch hat die Welt erobert und sich beeindruckende Werkzeuge geschaffen, um Erkenntnisse zu generieren und in seinem Sinne auf die Welt einzuwirken. Warum sollte er es nicht schaffen, die Mächtigkeit der von ihm zuletzt geschaffenen Instrumente zu erkennen, ihrem Gebrauch im eigenen Interesse klare Regeln und auch Grenzen zu setzen? Ich bin überzeugt davon, dass Homo Sapiens das kann. Aber er muss sich beeilen. Schon einem Zeitreisenden aus den 1980er Jahren käme unsere heutige Welt wie sehr krude Science Fiction vor. Und die Zeitspanne von der Vorstellung des iPhone (2007) bis zur weltweiten Massenverbreitung des Smartphone (fünf Jahre später) wird uns vielleicht schon bald wie eine romantisch langsame Technikausbreitungsgeschichte vorkommen. Gerade für das bei solchen Aufgaben bisweilen zögerliche und allzu theoretisch agierende Deutschland gilt: Die Aufnahme von Geschwindigkeit bei der zielorientierten Erörterung und Lösung der in diesem Beitrag angesprochenen Desiderata tut not. Unsere Zukunft bestimmen wir jetzt.

Dr. phil. Tobias Knobloch leitet das Projekt „Algorithmen fürs Gemeinwohl“ der Stiftung Neue Verantwortung (SNV) in Berlin. Die SNV arbeitet u. a. an Fragen der Datenökonomie, der IT-Sicherheit, der künstlichen Intelligenz und der staatlichen Überwachung im Internet und entwickelt Ideen, wie die deutsche Politik den technologischen Wandel im Sinne des Gemeinwohls gestalten kann.

Knobloch war von 2011 bis 2015 stellvertretender Leiter des Referats für Öffentlichkeitsarbeit, digitale Kommunikation und Besucherdienst im Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ). Dort war er für zahlreiche Digitalisierungsthemen verantwortlich, hat die Online-Kommunikation verantwortet und ein entwicklungspolitisches Datenportal aufgebaut. Von 2008 bis 2010 hat Knobloch ein vom High-Tech Gründerfonds investiertes Software-Startup mitgegründet und war an mehreren wissenschaftlichen Forschungs- und Evaluationsprojekten beteiligt. Nach dem Staatsexamen in Latein und Philosophie für das Lehramt (Bochum) war er von 2005 bis 2008 Stipendiat im DFG-Graduiertenkolleg „Auf dem Weg in die Wissensgesellschaft“ (Bielefeld) und hat dort über Computersimulationen promoviert. Von 1999 bis 2004 war Knobloch, anfangs parallel zum Studium, Projektmanager in einer Essener IT-Agentur.