



Quelle: © Yuya Shino, Reuters.

[Die digitale Zukunft](#)

# Regeln für Roboter

Warum wir eine digitale *Magna Carta* für das  
Zeitalter der intelligenten Maschinen brauchen

Olaf Groth / Mark Nitzberg / Mark Esposito

Wir stehen zurzeit an einem Wendepunkt der Menschheitsgeschichte – an der Schwelle zu einer unbekannteren digitalen Zukunft: Eine leistungsstarke neue Technologie – die Künstliche Intelligenz (KI) – durchdringt alle Lebensbereiche, was wir vor allem Fortschritten in der Entwicklung neuronaler, dem menschlichen Gehirn nachempfunder Netzwerke zu verdanken haben. Gesellschaft und Wirtschaft sind zunehmend abhängig vom Einsatz Künstlicher Intelligenz. Um auch künftig Freiheit, Inklusion und Wachstum sicherzustellen, bedarf es neuer Handlungsgrundsätze. Wir brauchen somit eine digitale *Magna Carta* für das Zeitalter kognitiver Maschinen.

---

### Dämmerung der kognitiven Ära

Künstliche Intelligenz vermag Muster in gigantischen ungeordneten Datensätzen zu erkennen.<sup>1</sup> Sie kann angesichts der wachsenden Verfügbarkeit von Daten die Leistungsfähigkeit von Unternehmen verbessern, Objekte schnell und genau identifizieren und noch schnellere Entscheidungsprozesse ermöglichen, wobei störende Einflüsse seitens komplexer, politischer und menschlicher Gegebenheiten auf ein Minimum beschränkt werden. Diese Konstellation wirft wesentliche Fragen zum Grad der menschlichen Entscheidungsfreiheit und Inklusion auf, deren Bedeutung in den kommenden Jahrzehnten wachsen wird. Hinzu kommt, dass sich sowohl bei der Einstellung zu diesen Fragen als auch beim Ansatz zu deren Lösung erhebliche Unterschiede zwischen den führenden Nationen herauskristalisieren. Die gegenwärtigen Differenzen im internationalen Wertgefüge werden sich verstärken, das Potenzial für gesellschafts- und geopolitische Konflikte ist erheblich.

Inwieweit werden wir Menschen – sowohl Eliten als auch Vertreter aller Machtpositionen und Einkommensklassen – künftig überhaupt noch in Entscheidungsprozesse einbezogen werden? Wie können wir diese schöne neue Welt der „Maschinenmeritokratie“ steuern?

Um eine Antwort auf diese Fragen zu finden, bedarf es zunächst eines Exkurses in die – acht Jahrhunderte zurückliegende – Vergangenheit: Im Januar 1215 hatte sich der soeben aus Frankreich zurückgekehrte König Johann von England mit aufgebrauchten englischen Baronen auseinandersetzen, die gegen sein unbeliebtes Herrschaftsprinzip *vis et voluntas* („Macht und Wille“) rebellierten. Um sie zu beschwichtigen, versammelten der König und der Erzbischof von Canterbury 25 dieser aufständischen Barone, um mit ihnen über eine *Charter of Liberties* zu verhandeln – einen Rechtekanon, der dem Adel mit dem Ziel eingeräumt werden sollte, die Entscheidungsgewalt des Königs einzuschränken. Nach längeren Verhandlungen kam man im Juni schließlich zu einer Einigung, durch die die königlichen Entscheidungsprozesse transparenter wurden und die den Adligen ein größeres Mitspracherecht zugestand. Darüber hinaus wurden Steuern und Lehenszahlungen eingeschränkt, und selbst Leibeigene erhielten gewisse Rechte. Natürlich war diese berühmte *Magna Carta* noch keineswegs perfekt: Viele darin enthaltene Klauseln dienten lediglich den Interessen einer bestimmten sozialen Klasse. Und doch tendieren wir heute dazu, die *Magna Carta* als große Wende in der Menschheitsgeschichte anzusehen – als substanziellen Fortschritt auf dem Weg zu einem gerechteren Verhältnis zwischen Mächtigen und Abhängigen. Letztlich hat sie den Boden

für die Aufklärung, die Renaissance und unsere heutige konstitutionelle Demokratie bereitet.

Eben dieses Gleichgewicht zwischen der wachsenden Macht des neuen Potentaten (der intelligenten Maschine) und der des Menschen steht heute auf dem Spiel – in einer Welt, in der Maschinen immer stärker an der Wertschöpfung beteiligt sind und immer mehr Alltagsprodukte herstellen, uns Menschen indes die Kontrolle über das Design und andere wichtige Aspekte zunehmend entgleitet. Dadurch verändern sich unsere bisherigen Lebens- und Arbeitsmuster auf Dauer. Denn diese von uns selbst geschaffene Technologie wird schon bald bestimmte kognitive Fähigkeiten des Menschen überholen und somit ihren Vorsprung uns gegenüber in Fragen der Produktivität und Effizienz in atemberaubendem Tempo vergrößern – ein Prozess, dessen Folgen sich nach Ablauf der nächsten ein bis zwei Jahrzehnte zeigen werden. Schließlich erkennen Computer innerhalb von Wochen, oft von Stunden, sowie mit einer dem Menschen weit überlegenen Präzision und Unermüdlichkeit komplizierte Muster aus Dutzenden, ja Hunderten von Datenströmen, deren Generierung Jahrhunderte wissenschaftlichen Arbeitens und Wirtschaftens gekostet hat. Hieraus Einsichten zu gewinnen und Entscheidungen zu kommunizieren ist der Kern der Kognition, also des Denkens.

## **Die kognitive Fähigkeit Künstlicher Intelligenz wird das menschliche Dasein in den nächsten 10 bis 20 Jahren grundlegend verändern.**

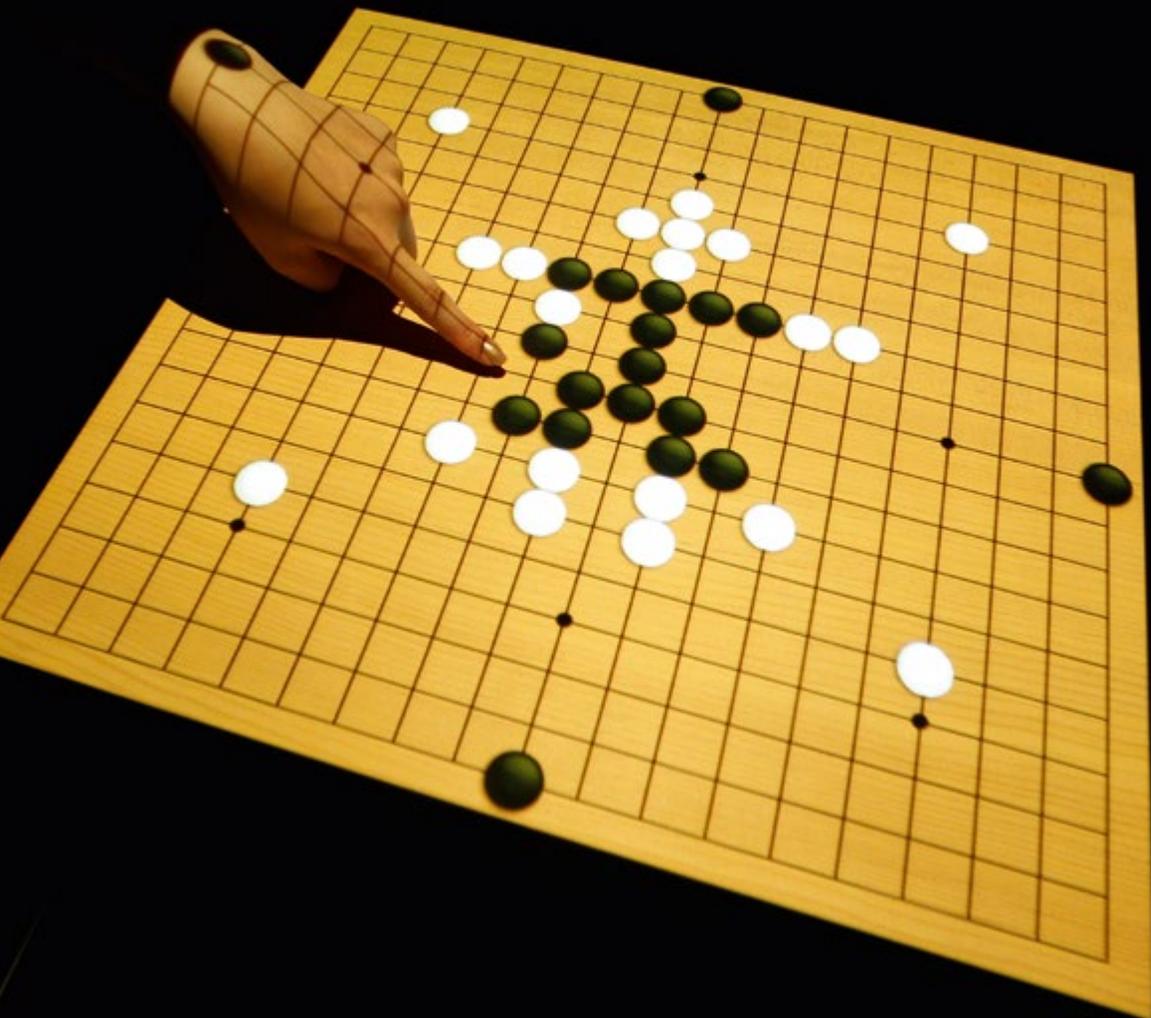
---

Zweifellos sind Maschinen noch Jahrzehnte davon entfernt, das intuitive Projektionsvermögen des menschlichen Gehirns zu replizieren, dessen Evolution einige Millionen Jahre in Anspruch genommen hat. Wir sind mittels weniger Daten, aber dafür einer in unserer DNA verankerten Fähigkeit des „Denkens um

Ecken“, des Erfahrens von Inspiration oder des Kreierens von Ideen bei gedanklichen Kollisionen den Maschinen in zahlreichen Situationen noch überlegen – ganz zu schweigen von der dem Menschen eigenen emotionalen Intelligenz, seiner Empathie, seinem Bewusstsein, seinem Moralverständnis sowie seiner Befähigung zu Intuition, zu Sinnggebung, zu seelischem Empfinden etc. Jedoch sollten wir uns nicht dem Trugschluss hingeben, denkende Maschinen könnten nicht auch parallele Fertigkeiten entwickeln, die zwar nicht genau unseren Fertigkeiten entsprechen, die diese jedoch in gewissem Maße umgehen oder substituieren könnten. Dies soll den Menschen im Ganzen keineswegs ersetzen, sondern, ganz im Gegenteil, sein Leben bereichern. Allerdings könnte es ihn auch Gefahren aussetzen. Das Potenzial dieser Disruption wird dadurch missachtet, dass Hollywood ihm zu viele attraktive Gesichter und Stimmen gegeben hat. Eine KI in Form von Scarlett Johansson im Film *Her* ist sicher fantasieanregend, liegt jedoch in der kritischen Reflektion realistisch denkender Menschen noch in ferner Zukunft. Diese kritische Reflektion und dieser Realismus haben allerdings auch ihre Schattenseiten und können dazu führen, den „Hype“ aus den USA und China lediglich zu belächeln anstatt ihn ernst zu nehmen und aktiv mitzugestalten.

### **Vom Homo oeconomicus zum Homo digitalis**

Kognitive Computer und intelligente Maschinen sind kein Science-Fiction mehr. Es gibt bereits jetzt gute Beispiele: Japan befindet sich am Rand einer Krise im Bereich des Pflegepersonals, weist doch die Bevölkerungsentwicklung eine ausgesprochene Überalterung der Gesellschaft aus. Es fehlen eine Million Krankenpfleger, Sozialarbeiter etc. Dementsprechend stark hat das Land in Roboter-Technologie investiert, um älteren Mitmenschen das Leben dennoch zu erleichtern. Dies gilt nicht nur für schwere Arbeiten wie das kraftaufwändige Heben von Patienten oder das Einkaufen für Ältere, sondern auch für Gesellschaftsdienste wie den simulierten Kontakt zu Haustieren und Mitmenschen, der durch mehr oder weniger echt wirkende, tierähnliche und



Go robot, go! Im Vergleich zum Schach stellt das Brettspiel Go für KI die ungleich größere Herausforderung dar. Inzwischen haben Computerprogramme den Menschen aber auch hier abgehängt. [Quelle: © Kim Kyung-Hoon, Reuters.](#)

humanoide Robotik angeboten wird (siehe hier auch den Beitrag zu Japan in dieser Ausgabe). Die Vereinigten Arabischen Emirate, wo soeben ein eigenes Ministerium für Künstliche Intelligenz geschaffen wurde, haben unterdessen den Einsatz simpler *Robocops* zur Durchführung einfacher Beobachtungs- und Informationsdienstleistungen mit Pilotprojekten gestartet. In Nigeria entstehen Applikationen wie Touchabl.com, die es u. a. selbst Analphabeten erlauben, an gesellschaftlichen Produktionsprozessen teilzuhaben oder ihre Rolle als Konsumenten

wahrzunehmen, kurz: besser am ökonomischen Leben teilzunehmen und eine digitale Stimme zu entwickeln.

Problematischer wird es allerdings, wenn Persönlichkeitsmerkmale, demografisches Profil und ein Großteil der sozialen Interaktionen von Menschen digital bewertet und öffentlich dargestellt werden und es somit zur Verletzung von Persönlichkeitsrechten und der Privatsphäre im westlichen Verständnis kommt.

In den USA, in England, in China und in Russland werden KI-Technologien wie etwa Gesichtserkennungs-, Sprachverarbeitungs- und Stimmungsanalyse Algorithmen mit dem Ziel einer präventiven Vorgehensweise gegen Kriminalität und Terrorismus eingesetzt. So können Polizeidirektionen in New York City und Los Angeles das Kriminalitätsrisiko bestimmter Personen und ganzer Stadtteile erfassen sowie rechtzeitig Beamte einsetzen. Inwieweit diese KI-Applikationen in ihrer Gesamtheit dem Persönlichkeits- und Datenschutz unterliegen, unterscheidet sich erheblich von Land zu Land. In Los Angeles zum Beispiel erfassen Kameras an Straßen in verschiedenen Stadtteilen Gesichter von Menschen, die zum Zeitpunkt begangener Straftaten

in Reichweite waren. Dadurch werden auch Personen in Datenbanken erfasst, die nicht in die Tat verwickelt waren, jedoch bereits durch Korrelation Punkte sammeln. Ob diese Vorgehensweise zulässig ist, wird sich durch eine von der *American Civil Liberties Union* angestrebte Klage entscheiden. Eine ähnliche Vorgehensweise gibt es in New York City, wo das NYPD mittels durch Technologie-Firmen angebotener Algorithmen erfasst, wo es in der Stadt mit Wahrscheinlichkeit zu Straftaten kommen wird, um dort präventiv Polizeikräfte einzusetzen. Und auch Gerichte in Wisconsin und Florida setzen bereits prädiktive Analytik ein, um Urteile darüber zu fällen, welche Art von Risiko ein Angeklagter darstellt und wie dementsprechend die Kautionsansatz



Von Alexa bis Barbie: Dass uns Gegenstände des täglichen Lebens „überwachen“, ist längst keine Zukunftsmusik mehr. [Quelle: © Mario Anzuoni, Reuters.](#)

werden sollte. Hier wie auch in New York können die Richter oder ihre „intelligenten“ Technologien nicht genau sagen, nach welcher Logik eine bestimmte Beurteilung zustandekam, denn die Algorithmen sind als neurale Netzwerke zusammengesetzt, die auf Schlussfolgerungen zwischen verschiedenen Levels des Netzwerks mit hoher Geschwindigkeit vor- und zurückpropagieren, ohne dass dies nachvollzogen werden könnte. Da dies selbst den sich auf die Transparenz ihres Rechtssystems berufenden US-Amerikanern zu weit geht, hat die DARPA-Forschungseinheit des Pentagon ein Projekt der „erklärbaren KI“ gestartet.

In China werden derzeit milliardenschwere KI-Systeme entwickelt, die für jeden Bürger ein öffentlich einsehbares „Vertrauens-Rating“ ermitteln.<sup>2</sup> Sowohl die großen chinesischen Internetkonzerne als auch der chinesische Staat entwickeln KI-betriebene Systeme wie das *Zhima Credits* („Sesam-Punkte“-) Programm, die anhand von Daten wie Lob oder Beschwerden durch Mitbürger und amtliche Stellen eine auf einem Punktesystem von bis zu 800 Punkten basierende Rangliste für Personen erstellen. In diese Kalkulation finden auch demografische Basisdaten Eingang. So etwa hat eine 28-jährige schwangere Frau ein besseres Ausgangs-„Rating“ als der 18-jährige Motorradkäufer. Der Inhaber eines Sesam-Kontostands von 700 Punkten wird als äußerst ehrbarer Mensch geachtet, während ein Stand von lediglich 300 Punkten Anlass zur gesellschaftlichen Benachteiligung ist, etwa bei der Bewilligung internationaler Reisen. Offizielles Ziel ist, der Korruption und vertrauensbehindernden Verhaltensmustern entgegenzuwirken und größere Verlässlichkeit in wirtschaftlichen und zwischenmenschlichen Transaktionen zu schaffen. Wer bereits einmal in China war, weiß um das Problem, authentische Produkte zu bekommen, wie z. B. beim Kauf von Markenwaren. Außerdem fehlte in China bisher ein für das Funktionieren einer Wirtschaft notwendiger Mechanismus, um die Kreditwürdigkeit von Konsumenten zu prüfen, wie dies etwa *credit rating agencies* in den USA oder die Schufa in Deutschland gewährleisten. Unklar ist jedoch, wie dieses System gegen Missbrauch geschützt

wird oder wie der chinesische Normalverbraucher Fehler und Missverständnisse korrigieren kann. Auch vermuten Experten, dass sich hierdurch neue soziale Klassen bilden werden, denn Menschen mit guten Bewertungen könnten solche mit schlechten scheuen. Die schwerwiegendste Konsequenz dieser Systeme ist allerdings die der staatlichen Kontrolle und Umerziehung zwecks politischer Unterwerfung. In einer Welt, in der Menschen nicht nur für kriminelle Akte, sondern auch für ihre demografischen Merkmale und zwischenmenschlichen „Ecken und Kanten“ umfassend durchleuchtet, quantitativ-statistisch ausgewertet und sodann zur Schau gestellt werden, ist zu fragen, wie weit man noch vom digitalen Brandmal entfernt ist. Derzeit ist die Teilnahme noch freiwillig, bis 2020 soll sie jedoch Pflicht werden. Wie wird sich dies auf die wirtschaftliche Existenz und die soziale Dynamik unter 1,5 Milliarden Chinesen auswirken? Werden daraus hervorgehende Verwerfungen auch international Wellen schlagen? Werden chinesische Konzerne diese Systeme auch international in Hilfe aus China beziehenden Entwicklungsländern, etwa im Rahmen der neuen *One Belt One Road*-Politik, einrichten?

Fakt ist in jedem Fall, dass prädiktive Analytik, KI und interaktive Robotik bereits heute als selbstverständliches Werkzeug von Regierungen und Unternehmen gelten müssen. Und Fakt ist auch, dass der Dialog in Politik und Gesellschaft, insbesondere der grenzübergreifende, diesbezüglich dem technologischen Fortschritt weit hinterherhinkt.

### **Was Maschinen wissen**

Dies alles beschränkt sich indes nicht nur auf den Sicherheitsbereich: Inzwischen beginnen Maschinen, uns in fast allen Lebensbereichen Entscheidungen abzunehmen. Sie erkennen unsere Denk- und Verhaltensmuster und diejenigen vermeintlich ähnlicher Menschen auf der ganzen Welt. Die Nachrichten, die wir erhalten und die unsere Meinung, unsere Sichtweise und unser Handeln prägen, beruhen auf Tendenzen, die wir (oder andere uns ähnliche Menschen) in unserem früheren Verhalten gezeigt haben.

Während wir am Steuer sitzen, sammeln Autohersteller und Versicherungsgesellschaften Informationen über unsere Verhaltensmuster, um uns noch bessere Navigationshilfen und eine immer autonomere Fahrzeugtechnologie anzubieten, die den Straßenverkehr bequemer und sicherer gestaltet. Wir genießen immer ausgefeiltere, immer maßgeschneidertere Unterhaltungs- und Videospiele, deren Hersteller immer besser über unsere sozioökonomischen Profile, Bewegungsmuster sowie unsere kognitiven und visuellen Präferenzen informiert sind – ein Wissen, das sie selbst wiederum für ihre Preisgestaltung nutzen können.

## **In China soll schon bald die Vertrauenswürdigkeit aller Bürgerinnen und Bürger anhand eines öffentlichen Punktesystems bewertet werden.**

---

Die neueste Entwicklungsstufe der Barbie-Puppe der Firma Mattel mag als gutes Beispiel dienen. Mittlerweile wieder vom Markt genommen, ist sie in der Lage, das Spielverhalten, die Reaktionen und die Sprache von Kindern aufzunehmen und auf Servern bei Mattel so zu verarbeiten, dass adaptive Dienstleistungen angeboten werden können. Dazu gehören z. B. das Beaufsichtigen von Kindern durch aktuell nicht in Reichweite befindliche Eltern, Einsichten in das soziale Verhalten sowie die Entwicklung neuer Produkte. All dies hat indes auch einen gewissen Flair des *big brother watching you*, insbesondere dann, wenn kommerzielle Interessen die Daten eines ungeschützten und unwissenden minderjährigen Kindes vermarkten. Die Kunden waren (noch!) nicht bereit, dieses Eindringen in ihre Privatsphäre zu tolerieren. Allerdings findet genau dies bereits auf vielfältige Weise statt: Jedes Smartphone sammelt heute derlei Daten, vertreibt sie zu Werbezwecken oder verwendet sie für die besagten

neuen, besseren Dienste und Produkte, die unser Leben noch gezielter bereichern sollen. Das Start-up *Scientific Revenue* südlich von San Francisco ermöglicht es den Entwicklern von Computer- und Handyspielen, das Spielverhalten, den Kontext sowie die Preis-Sensitivität von Spielern in der Absicht zu ermitteln und zu projizieren, Preise für ein Spiel oder einen In-Spiel-Kauf entsprechend individuell festzusetzen. Natürlich gibt es diese Verfahrensweise auch schon im analogen Leben. Vielerorts auf der Welt spielt der sozio-ökonomische Eindruck, den man seinem Gegenüber bei der Verhandlung von Preisen für Waren gibt, eine Rolle bei der Preisfindung. Die KI-getriebenen digitalen Plattformen kodifizieren, amplifizieren und skalieren dieses Verfahren – allerdings ohne die wünschenswerte Transparenz.

Eine solche Individualisierung ist für uns sicherlich auf der einen Seite lebensbereichernd und angenehm. Auf der anderen Seite sollten wir uns darüber im Klaren sein, dass wir darauf bauen, dass die Maschine „schon wissen wird, was richtig für uns ist“. Und tatsächlich lernt die Maschine uns vielleicht noch besser kennen als wir uns selbst – zumindest aus einer rein rationalen und empirischen Perspektive. Kognitive Dissonanzen zwischen dem Menschen, den wir zu sein vorgeben, und demjenigen, der wir wirklich sind, kann die Maschine nicht ohne Weiteres erfassen. Sie nimmt nur die realen Daten unseres realen Handelns wahr und beschränkt uns damit auf die Person, die wir bisher gewesen sind, statt auch den Menschen zu erkennen, der wir gern wären oder eines Tages zu sein hoffen.

### **Persönliche Entscheidungsfreiheit**

Wird die Maschine unsere persönliche Entscheidungs- und Entfaltungsfreiheit einschränken? Wird sie die glücklichen Zufälle des Lebens eliminieren? Wird sie unsere Existenz flächendeckend durchplanen, sodass wir nur noch Personen begegnen, die genauso sind wie wir, und uns damit der zwischenmenschlichen Begegnungen und Reibungen berauben, die uns dazu zwingen, zu anderen, vielleicht besseren Menschen zu werden? Die KI birgt enormes

Verbesserungspotenzial. Denn hinter manchen unserer persönlichen Entscheidungen sollten tatsächlich objektivere Analysen stehen: So könnte beispielsweise eine rationale Synthese aus dem CO<sub>2</sub>-Fußabdruck verschiedener Verkehrsmittel, unseren Zeitplänen und sozioemotionalen Bedürfnissen zu umweltpolitisch sinnvollerem Entscheidungen führen. Ein Blick auf die Scheidungsraten in den meisten Industrieländern könnte zu der Erkenntnis führen, dass uns einige wenige objektive analytische Anhaltspunkte dafür, wie vernünftig wir bei unserer Partnerwahl vorgehen und wer wirklich zu wem passt, nicht schaden könnten. Schließlich decken sich unser Selbstbild und unsere Aspirationen nicht immer mit unseren realen Verhaltensmustern – ein Phänomen, das die Psychologie „kognitive Dissonanz“ nennt. Zudem müssten effektivere Lehrpläne und eine kognitiv-adaptive Vermittlung für unterschiedliche Gruppen von Schülern und Studenten mit unterschiedlichen Lernprofilen entwickelt werden. Amerikanische KI-Spezialisten arbeiten an Systemen, die helfen können, Lebensmittelknappheit und Hungerkatastrophen zu vermeiden, indem Veränderungen von Wetter-, Boden-, Infrastruktur- und Marktfaktoren in komplexe Modelle integriert werden, um so rechtzeitig Abhilfe schaffen zu können. Die Anzahl der nützlichen Anwendungen ist fast unendlich.

### **Polarisierung oder sozialer Ausgleich?**

Die Künstliche Intelligenz könnte unsere Gesellschaft indes auch polarisieren, indem sie uns in zunehmendem Maß in virtuelle Gruppen gleichgesinnter Menschen hineinzwängt und uns dadurch in unseren Überzeugungen und Wertvorstellungen bestärkt, ohne dass wir durch zufällige Konfrontation mit Andersdenkenden die Chance erhalten, diese Ansichten zu überprüfen, zu verteidigen oder womöglich zu revidieren. Nicht zuletzt könnte KI auch für „digitales Social Engineering“ und für die Schaffung von Mikro-Parallelgesellschaften missbraucht werden. So könnten etwa Zimmer- oder Apartment-Vermittlungen in bestimmten Bezirken Unterkünfte nur an Personen mit genau ausgewähltem soziopolitischem, ökonomischem

oder psychometrischem Profil vermieten bzw. nur Immobilien von solchen Anbietern anmieten.

## **Schon heute kommt KI bei der Personalrekrutierung zum Einsatz.**

---

Auch Unternehmen könnten ihre Mitarbeiter mithilfe von KI sehr viel gezielter bei Einstellungsgesprächen auswählen. Dies geschieht mithilfe von Algorithmen, die Video-Streams der verschiedenen Kandidaten entlang der für die Firmen wichtigen Verhaltens-Kriterien evaluieren. Dies passiert teilweise heute schon. Vielversprechende Start-ups wie *HireView* und *Koru* in den USA machen große Fortschritte und sind bereits gut in der Industrie verankert – Unilever, Urban Outfitters und Vodafone gehören zu ihren Kunden. Diese Unternehmen sind dadurch in der Lage, den oft mit persönlichen Vorurteilen belegten zwischenmenschlichen Gesprächen eine objektivere Analyse-Methode zur Seite zu stellen. Diese dient nicht zuletzt als ein Gegenpol zum Lebenslauf, der sich mit Blick auf die Vorhersehbarkeit des möglichen beruflichen Erfolgs eines potenziellen Mitarbeiters in neuen Situationen als relativ ineffektiv herausgestellt hat. Auch ist es teilweise einfacher, Kandidaten zu rekrutieren, die zwar vor dem Hintergrund ihrer beruflichen Laufbahn weniger „glamourös“ wirken, vom situativen Verhalten her jedoch besser zum Unternehmen passen. Gerade für Übergangene Minoritäten könnte dies wünschenswert sein. Letztendlich wird dadurch die kurzfristige Erfolgsrate eines Arbeitgebers steigen, doch ist auch zu fragen, ob ein allzu eng gesteckter Analyse-Parcours nicht auch den Mitarbeiterpool zu stark homogenisiert, was wiederum die längerfristigen strategischen Optionen der Unternehmen einschränken könnte.

### **Wo bleiben die Werte?**

Eine Maschine beurteilt uns lediglich aufgrund unseres *expliziten* Verhaltens und unserer Wertvorstellungen, vor allem jener, die sich aus

unseren kommerziellen Transaktionen ergeben – liefern diese doch handfeste formelle Datenpunkte. Sie ignoriert dabei jedoch andere tiefverwurzelte Überzeugungen, die wir in unserem Handeln zum betreffenden Zeitpunkt nicht unbedingt zum Ausdruck bringen und für die es

noch keine digitalisierten Datenpunkte gibt. Neu entstandene Überzeugungen oder Änderungen unserer Wertvorstellungen, die sich nur schwer systematisch erfassen lassen, können der Künstlichen Intelligenz leicht entgehen. So könnte KI etwa aufgrund früherer Daten Entscheidungen



zum Wohle unserer Sicherheit treffen, die das Wohlbefinden anderer Menschen gefährden – Entscheidungen, die wir in dem entsprechenden Moment möglicherweise verwerflich finden. Wir sind komplexe Wesen, die stets zwischen verschiedenen Wertvorstellungen und anderen

wichtigen Prioritäten abwägen. Wofür wir uns im Einzelfall entscheiden, hängt von der jeweiligen Situation ab. Manchmal gibt es zu solchen Situationen nur wenige oder gar keine systematisch erfassten Präzedenzfälle, auf die ein KI-Programm zurückgreifen könnte. So wäre denkbar, dass sich die Entscheidung eines tierliebenden Menschen, sein autonomes Fahrzeug einem Tier auf der Straße ausweichen zu lassen und damit sein eigenes Verletzungsrisiko zu erhöhen, verändert, sobald er selber Kinder bekommt. Wird die Maschine unser Recht auf Willensfreiheit und Werte-Evolution – und das Privileg, uns selbst gelegentlich neu zu erfinden – respektieren?

### Diskriminierung und Vorurteile

Eine Maschine könnte ferner Menschen mit schlechterem Gesundheitszustand oder niedrigerem sozialen Status diskriminieren, weil ihre Algorithmen auf Mustererkennung und groben statistischen Durchschnittswerten basieren. Als *Uber* bei der Entwicklung seiner Algorithmen Postleitzahlen heranzog, um herauszufinden, aus welchen Siedlungen die Mitfahrer mit größter Wahrscheinlichkeit stammten, kam es umgehend zu einem Aufschrei gegen die vermeintliche Rasendiskriminierung. Welchen Menschenschlag wird die Künstliche Intelligenz begünstigen? Wird sie sich für das Überleben der Stärksten, der Beliebtesten oder der Produktivsten einsetzen? Wird sie diese Entscheidungen transparent machen? Und welche Regressmöglichkeiten haben wir, um uns dagegen zu wehren?

Des Weiteren kann die Entwicklung von Algorithmen und die Beschaffung von Datensätzen ungewollten Einflüssen unterliegen – zum Beispiel der persönlichen Lebensgeschichte, den Präferenzen und unbewussten Vorurteilen des Programmierers – oder Motiven und Anreizen vonseiten seines Arbeitgebers. Können wir davon ausgehen, dass KI bei ihrer Arbeit stets objektiv sein wird? Welche Art von KI-Systemen werden

Auf in den Kampf: Wie jede Technologie lassen sich auch Roboter und KI zum Guten wie zum Schlechten einsetzen. Quelle: © Michael Mauldin, MegaBots Inc., Reuters.



Unternehmen wohl entwickeln? Solche, die im Interesse ihrer Kunden, Partner, Topmanager oder Aktionäre agieren? Wird eine von Technologieunternehmen, Klinikketten und Versicherungsgesellschaften gemeinsam entwickelte KI für das Gesundheitswesen tatsächlich in erster Linie das Interesse der Patienten wahrnehmen oder wird hierbei nicht vielleicht doch eher die Rendite die Prioritätenliste anführen?

Wir können die Geister, die wir gerufen haben, nicht wieder loswerden und das sollen wir auch nicht. Schließlich bietet die KI Vorteile, die unser Leben verändern und uns eine enorme Weiterentwicklung ermöglichen – sowohl menschlich und gesellschaftlich als auch wirtschaftlich. Man muss kein Anhänger utopischer oder dystopischer Science-Fiction sein, um zu erkennen, dass wir uns möglicherweise an der Schwelle zu einer faszinierenden, radikalen Veränderung in der Evolution der Menschheit befinden, wie es sie seit einem Jahrtausend nicht mehr gegeben hat. Revolutionen dieser Art verlaufen niemals reibungslos. Sie sind fast immer chaotisch, undurchsichtig und voller ethischer Fußangeln.

### **Eine neue digitale Ethik**

Der Bedarf an mehr Ethik und Verantwortung im digitalen Bereich wurde auch während eines dreitägigen Workshops zum Thema „Zukunft der Arbeit“ zwischen der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) und Vertretern aus Wirtschaft und Forschung an der *University of California Berkeley* Ende November klar artikuliert. Einige professionelle Gruppierungen nehmen sich dieser Aufgabe bereits an: Das renommierte *Institute for Electrical and Electronics Engineers* (IEEE) hat bereits einen *Code of Conduct* für diesen Berufsstand erstellt. Das *Institute for the Future of Life*, gegründet von MIT-Physiker Max Tegmark, ein Zusammenschluss führender Wissenschaftler und Unternehmer im Bereich KI, dem neben einigen deutschen und österreichischen Professoren unter anderem auch Elon Musk, Stephen Hawking, Ray Kurzweil, und Jaan Tallinn (Gründer von *Skype*) angehören, hat die *Asi-lomar Principles* (benannt nach einem bekannten

Konferenz-Zentrum in Kalifornien) erarbeitet. Die *Partnership on AI* (PAI), ein Verband einiger bekannter Internet-Konzerne in den USA wie Google und *Microsoft*, der auch *Amnesty International* und die *American Civil Liberties Union* angehören, bereitet ebenfalls Richtlinien vor, vermutlich nicht zuletzt in der Absicht, durch eine Selbstregulierung möglichen regierungsseits lancierten Restriktionsmaßnahmen zuvorzukommen. Das Weltwirtschaftsforum hat gerade das erste einer globalen Reihe von Zentren für die vierte industrielle Revolution in San Francisco ins Leben gerufen – samt einem KI-Programm zur Bearbeitung dieser Themen. Auch Informatik- und ingenieurwissenschaftliche Fakultäten sollten sich überlegen, wie sie ihre Studenten und Wissenschaftler auf verantwortungsvolle Designs der KI einschwören können. Joi Ito, Direktor des berühmten *MIT Media Lab*, gab in seinem Interview mit Präsident Obama im *WIRED Magazin*<sup>3</sup> zu bedenken, dass viele der brilliantesten KI-Köpfe in Laboren wie dem seinen noch nicht genug auf die Bedürfnisse der Menschen eingespielt seien, was zum Teil auch damit zu tun hat, dass solchen technisch brillanten „Nerds“ die Geduld für menschliche Komplexität, Emotionalität und zwischenmenschliche Politik fehlt und sie diese lieber gleich aus dem Spiel lassen.

Was diese bestehenden Initiativen insgesamt allerdings vermissen lassen, ist zum einen ein wirklich globaler Ansatz, der sich der schwierigen Gemengelage aus unterschiedlichen Werten und Definitionen der Ethik annimmt, zum anderen die richtige Zusammensetzung aus Teilnehmern verschiedener Sektoren der Gesellschaft, also ein *multi-stakeholder approach*.

### **Eine Magna Carta für das digitale Zeitalter**

Kognitive Technologien werden nicht nur unsere ökonomische Zukunft bestimmen, sondern auch unsere gesellschaftliche. Sie beeinflussen, was in unseren Kopf kommt, wer unseren Kopf kennt, mit wem unser Kopf kooperiert und wieviel Gedankengut unser Kopf gegenüber der Maschine generiert. Dies beeinflusst das gesamte gesellschaftliche Gefüge.

KI muss sowohl das Wachstum des Menschen als Ganzes als auch seine Persönlichkeitsentwicklung unterstützen, sodass er sein volles gesellschaftliches und ökonomisches Potenzial ausschöpfen kann. Es gilt, den Menschen zu helfen, mit den KI-Unsicherheiten umzugehen. Sie müssen wissen, dass die Politik ihnen hilft, sich vorzubereiten. Des Weiteren müssen Anreize für Unternehmen und Wissenschaft gegeben werden, damit sie beide Facetten gezielt umzusetzen suchen. Ein Mittel dieser Umsetzung ist die *Digitale Magna Carta* für die durch KI getriebene vierte industrielle Revolution. Hierbei handelt es sich um eine von verschiedenen Interessenvertretern gemeinsam entwickelte inklusive *Carta* der Rechte und der Werte, von der wir uns bei der Entwicklung Künstlicher Intelligenz künftig leiten lassen sollten. Sie sollte den Grundstein für die kommende Koexistenz zwischen Mensch und Maschine sowie für eine inklusivere Weiterentwicklung der Menschheit legen. Ob im wirtschaftlichen, sozialen oder politischen Bereich – wir als Gesellschaft müssen beginnen, uns über Rechte, Pflichten und Rechenschaftsregeln für Inklusion und Fairness an der Schnittstelle zwischen KI und unserem Leben als menschliche Wesen Gedanken zu machen.

### **Die KI-Branche arbeitet an eigenen Richtlinien, wohl auch um restriktiven Maßnahmen vonseiten der Regierungen zuvorzukommen.**

---

Die *Carta*-Initiative sollte idealerweise die Konstitutionalisierung einer globalen Multi-Stakeholder Institution für *KI-Governance* anstreben, die über eine Stabsfunktion globale Entwicklungen im Bereich KI verfolgt und analysiert (Thinktank-Funktion), und sie in einem öffentlichen Plenum thematisiert (Kongress-Funktion). Dies müsste die Verankerung eines *good faith*-Kollaborations-Modus zwischen Vertretern aus den Bereichen Privatwirtschaft,

Regierungen und Nichtregierungsorganisationen im Bereich KI einschließen.<sup>4</sup>

Die Verhandlung zwischen den verschiedenen Interessengemeinschaften wirtschaftlicher, wissenschaftlicher, politischer und sozialer Art muss über einen modernen, offenen Kongress geführt werden. Dieser Kongress sollte eine internationale, multi-sektorale Teilnehmerschaft zulassen, überschreitet doch auch die KI Grenzen und greift in alle Bereiche des gesellschaftlichen Lebens ein. Dies erfordert das Zusammenkommen nicht nur von Regierungen, sondern auch von Nichtregierungsorganisationen, akademischen Institutionen und Wirtschaftsvertretern an einem Tisch, um die grenzübergreifenden Konsequenzen der KI mit dem Ziel zu erörtern, nicht aneinander vorbeizugieren. Um die Anreize für alle Seiten zu erhalten, sollte der Kongress sowohl innovations- als auch gerechtigkeitsfördernde Regeln zum Ziel haben.

Es handelt sich hierbei um eine neue multilaterale Institution, die gegebenenfalls eigenständig oder aber unter dem Mantel der Vereinten Nationen (VN) ihren Platz einnimmt. Denn ausschlaggebend ist, dass sowohl der Stab als auch die Prozesse dieser Institution digitale Kompetenzen auf höchstem Niveau haben, um mit den wissenschaftlich-technischen Kompetenzen von Konzernen, Unternehmern und Forschungslaboratorien Schritt halten zu können. Dies ist kein triviales Unterfangen, denn KI-Talent ist teuer in einem Umfeld, in dem globale Internetkonzerne Saläre von mehreren Millionen Dollar pro Jahr zahlen. Des Weiteren sollte der Kongress inklusiv und somit nicht nur physisch, sondern auch digital konstituiert werden, damit die Barrieren und Kosten der Partizipation niedrig bleiben und die Dialoge zügig vorangetrieben werden können. Die Mechanismen der traditionellen Institutionen des Bretton-Woods-Systems lassen dies nicht zu. Darüber hinaus sind sie ungeeignet, um wichtigen digitalen Akteuren wie China, Indien, Russland oder Nigeria, die unsere digitale Zukunft entscheidend gestalten, eine formative Stimme zu geben und sie sodann konstruktiv-kritisch einzubinden. Wir sollten vor dieser Offenheit keine Angst haben, zumal die Alternative unattraktiv

ist: Wie im Falle der von Beijing aus geführten *Asian Infrastructure Development Bank* könnte es zu Parallel-Institutionen kommen, in denen Länder Clubs bilden. Dies steht jedoch im direkten Gegensatz zur globalen Verbreitung von KI-Technologie durch Konzerne und Regierungsstellen sowie zum digitalen, Grenzen mit Leichtigkeit überwindenden Datenfluss.

Die *Carta* und der Kongress sollten sich mit Blick auf die Schließung eines mittelfristigen Abkommens unter anderem mit folgenden Fragen befassen:

1. Welche Rolle soll die menschliche Entscheidungsfreiheit beim Einsatz von KI spielen? Wie können Entscheidungsfreiheit und Privatssphäre des Individuums geschützt werden? Wie kann dies mit den Bedürfnissen der Gesellschaft in Einklang gebracht werden?
2. Wie soll mit Akteuren, die sich gegen die Nutzung von KI-Applikationen entscheiden, umgegangen werden (z. B. Einräumen eines *Opt-Out*)?
3. Inwiefern können gesellschaftlich-politische Prozesse – etwa Wahlen, Meinungsbildung, Ausbildung und Erziehung – durch KI unterstützt werden und wie kann der schädigende Einsatz von KI verhindert werden?
4. Wie kann der Korrumpierung oder Fälschung von Datensätzen sowie der Diskriminierung von Individuen oder Gruppen in Datensätzen effektiv entgegengewirkt werden?
5. Inwiefern soll die menschen- und naturgerechte Nutzung und /oder Eindämmung von KI durch Richtlinien gesichert werden?
6. Welchen Stellenwert soll der gesellschaftlich-soziale Nutzen bei der Erforschung, Entwicklung, Förderung und Evaluierung von KI-Projekten haben?
7. Wie können die Förderung und Weiterbildung von Arbeitnehmern für neue Beschäftigungsmöglichkeiten und persönliches Wachstum in

die KI-getriebene Automatisierung von Produktions- und Arbeitsprozessen einbezogen werden?

8. Wie kann ein effektiver, kontinuierlicher Austausch verschiedener Stakeholder über KI ermöglicht werden?
9. Wie könnte eine ständige internationale Institution aussehen, die sich der Frühaufklärung sowie der Moderierung der – für Mensch und Wirtschaft – wachstumsorientierten, verantwortungsvollen Nutzung von KI annimmt?

Ohne einen sicher nicht einfachen Dialog zu diesen Fragen wird es uns nicht gelingen, in der globalen Gesellschaft eine ausreichende Vertrauensbasis für die Künstliche Intelligenz zu schaffen und die enormen, durch sie eröffneten Chancen zu nutzen. Nur wenn wir uns auf Regeln verständigen, können wir gemeinsam die KI-Zukunft steuern und sicherstellen, dass die Sozialverträglichkeit revolutionärer, neuer Technologien international garantiert wird und nicht nur im Dienste profit-, macht- und geopolitischer Interessen steht. Angesichts der signifikanten weltweiten Kapazitäten, wie den wissenschaftlichen und unternehmerischen Talenten in diesem Bereich, wäre es eine vertane einmalige Chance, würden wir diesen nächsten Schritt in Richtung eines gerechten, wertegebundenen menschlichen Wachstums in der vierten industriellen Revolution – einer kognitiven Revolution – nicht wagen.

---

**Dr. Olaf Groth, Ph.D.** ist CEO bei Cambrian.ai und Professor für Strategie, Innovation und Wirtschaftswissenschaften sowie Direktor für Digitale Zukunft an der Hult International Business School und Fellow am UC Berkeley Roundtable on the International Economy.

**Dr. Mark Nitzberg, Ph.D.** ist geschäftsführender Direktor des Center for Human-Compatible AI an der University of California in Berkeley sowie wissenschaftlicher Leiter und Direktor bei Cambrian.ai.

**Dr. Mark Esposito, Ph.D.** ist Mitgründer von Nexus FrontierTech und Professor für Betriebswirtschaft und Ökonomie an der Hult International Business School.

- 1 Teile dieses Artikels sind bereits im Onlineblog des Harvard Business Manager Deutschland (<http://harvardbusinessmanager.de>, 10/2017) erschienen.
- 2 Botsman, Rachel 2017: Big data meets Big Brother as China moves to rate its citizens, *Wired*, 21.10.2017, in: <http://buff.ly/2l4rzMj> [23.02.2018].
- 3 Dadich, Scott 2016: Barack Obama, Neural Nets, Self-Driving Cars, and the Future of the World, *Wired*, 11/2016, in: <http://buff.ly/2dC2AXY> [23.02.2018].
- 4 Die Autoren dieses Beitrages arbeiten bereits an einem „Cambrian Congress“-Konzept, welches sich sowohl der möglichen kambrischen Explosion positiver Möglichkeiten als auch der Risiken annimmt.