

Digitalisierung und Künstliche Intelligenz: Orientierungspunkte

Norbert Arnold / Tobias Wangermann (Hrsg.)



Digitalisierung und Künstliche Intelligenz: Orientierungspunkte

Norbert Arnold / Tobias Wangermann (Hrsg.)

Impressum

Herausgeberin:

Konrad-Adenauer-Stiftung e. V. 2018, Berlin

Umschlagfoto: © GrapeImages/iStock by Getty Images;
metamorworks/shutterstock

Gestaltung und Satz: yellow too, Pasiiek Horntrich GbR

Die Printausgabe wurde bei der Druckerei Kern GmbH, Bexbach,
klimaneutral produziert und auf FSC-zertifiziertem Papier gedruckt.

Printed in Germany.

Gedruckt mit finanzieller Unterstützung der
Bundesrepublik Deutschland.



Der Text dieses Werkes ist lizenziert unter den Bedingungen von
„Creative Commons Namensnennung-Weitergabe unter gleichen
Bedingungen 4.0 international“, CC BY-SA 4.0 (abrufbar unter:
<https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>).

ISBN 978-3-95721-554-3

Inhaltsverzeichnis

Einleitung: Kritischer Optimismus

Norbert Arnold

7

Welt im Wandel

10

Internet of Things

Richard Lackes

11

Künstliche Intelligenz und Robotik

Oliver Brock

29

Robotik und ihr Beitrag zu Wachstum und Wohlstand

Jens Südekum

45

Deutschland als unübertrefflicher KI-Standort

Jürgen Schmidhuber

57

Wie Künstliche Intelligenz die Medien verändert

Norbert Lossau

66

Rahmenbedingungen und Regeln	81
-------------------------------------	-----------

Rechtspersönlichkeit für autonome Systeme?	
<i>Jan-Erik Schirmer</i>	82

Roboterethik	
<i>Cathrin Misselhorn</i>	96

Zwischen System und Verantwortung	
<i>Elisabeth Gräß-Schmid</i>	110

Algorithmen entscheiden nicht, und sie werden es auch niemals	
<i>Joachim Fetzer</i>	126

Algorithmische Entscheidungen: Transparenz und Kontrolle	
<i>Katharina A. Zweig</i>	143

Autorinnen, Autoren und Herausgeber	166
--	------------

Einleitung

Kritischer Optimismus

Norbert Arnold

I.

Das World Wide Web gibt es gerade einmal 30 Jahre. Das Internet ist etwas älter. Konrad Zuse hat seine erste Rechenmaschine, die den Namen Computer verdient, 1941 entwickelt, der Durchbruch wurde jedoch erst in den 1970er Jahren erzielt. Digitalisierung ist also eine junge Entwicklung. Aber mit einer ungeheuren Durchschlagskraft. Wer kann sich heute noch vorstellen, Texte auf einer mechanischen Schreibmaschine zu tippen? Wer würde für einen Vortrag noch Dias oder Overhead-Folien verwenden? Wer kann sich vorstellen, in der schnelllebigen Business-Welt zu faxen statt zu e-mailen? Und graue Festnetzeinheiten-telefone zu verwenden statt der überwältigenden Vielfalt moderner Telefonie, besonders der Smartphones als digitale Alltagsbegleiter? Innerhalb einer Generation hat Digitalisierung unsere Welt radikal verändert.

Die digitale Welle schwappt über die ganze Welt – nicht nur über die Industrieländer, sondern auch über die Länder mit nachholender Entwicklung. Und es hat den Anschein, dass Digitalisierung Rückständigkeit vermindert und weltweit die Fortschrittsdynamik antreibt. Eine zweite Welle folgt hinterher: Auf die programmierbaren Rechenmaschinen folgen selbstlernende Maschinen, die in der Lage sind, eigenständig Probleme zu lösen. Noch steckt Künstliche Intelligenz in den Kinderschuhen, aber bereits jetzt ist erkennbar, dass auch sie unsere Lebenswelt – und den Menschen selbst – verändern wird.

II.

Digitalisierung und Künstliche Intelligenz sind längst Gegenstand politischen Handelns. In Förderprogrammen wird die Weiterentwicklung der neuen Technologien unterstützt. Alle Industrieländer bauen die digitale Infrastruktur mit Nachdruck aus. Strategien zur Künstlichen Intelligenz werden überall entwickelt.

Regulierend versucht man, möglichen Risiken entgegenzuwirken: vom Datenschutz über Cyber-Sicherheit bis hin zu den Fragen des Schutzes geistigen Eigentums. Viele dieser Fragen sind nicht neu, stellen sich aber im Licht neuer technologischer Möglichkeiten in verschärfter Brisanz. Im politischen Raum entstehen zu den Chancen- und Risikoaspekten zahlreiche Beratergremien, die Politik bei der Meinungsbildung und Entscheidungsfindung unterstützen. Wie bei allen sogenannten Zukunftstechnologien müssen auch bei Digitalisierung und Künstlicher Intelligenz Förder- und Regulierungsentscheidungen getroffen werden, obwohl die Faktenlage unsicher ist und die künftige Entwicklung mit ihren Chancen und Risiken naturgemäß vage bleiben muss.

In dieser Situation ist weder unreflektierte Euphorie noch dumpfer Pessimismus angemessen. Notwendig ist ein kritischer Optimismus, der das Ziel verfolgt, den technologischen Fortschritt in einen humanen und gesellschaftlichen Fortschritt zu überführen.

Politik steht bei dieser schwierigen Aufgabe nicht allein. Sie wird von Wissenschaftlern unterstützt, und zwar sowohl aus dem naturwissenschaftlich-technischen als auch aus dem geistes- und sozialwissenschaftlichen Bereich. Deshalb ist es so wichtig, die Forschung zu Digitalisierung und Künstlicher Intelligenz in ihrer gesamten Breite voranzutreiben. Eine fundierte Faktengrundlage ist notwendig, um zu entscheiden, welche Anwendungen unter welchen Rahmenbedingungen realisiert werden sollen.

Viele Menschen haben dennoch Angst. Die meisten nutzen die neue Technik, trotzdem entsteht ein Unbehagen, wie Umfragen immer wieder zeigen. Besonders Künstliche Intelligenz wird in konkreten Anwendungsfeldern kritisch gesehen. Medialer Hype und die Wahrnehmung vieler Menschen sind nicht kongruent. Die Gefahr, dass die (Zu-)Stimmung kippt, ist groß. Aus früheren Erfahrungen mit anderen Zukunftstechnologien – z. B. der Gentechnologie – ist bekannt, wie wichtig es ist, einen gesellschaftlichen Konsens herzustellen.

III.

Vor diesem Hintergrund stellt sich eine besonders wichtige Aufgabe: Orientierungspunkte zu erarbeiten, die als Wegmarken für die weitere Entwicklung dienen können. Im Dschungel der Informationen und Nachrichten sind Schneisen notwendig, die den Blick freigeben zur (Selbst-)Vergewisserung.

Die vorliegenden Texte verfolgen dieses Ziel: Anhand von Beispielen werden wichtige technologische Trends dargestellt; mit wenigen reflektierenden Tiefenbohrungen wird die Tragfähigkeit des gesellschaftlichen, ethischen und rechtlichen Fundaments überprüft.

Welt im Wandel



Internet of Things

Richard Lackes

- › Als „Internet der Dinge“ (IoT) wird ein System intelligenter, über ein Kommunikationsmedium verbundener Produkte bezeichnet. *Smart Home, Smart Cities, E-Health*, autonomes Fahren, intelligente Fertigungs- und Logistiksysteme sind bekannte Anwendungsbeispiele. IoT ist ein wichtiger Innovationstreiber der Digitalisierung.
- › Das IoT ist mit Risiken verbunden, die besonders Fragen des Datenschutzes und des Dateneigentums betreffen. Objektdaten und (besonders geschützte) Personendaten können nicht scharf voneinander abgegrenzt werden. Gute Regelungen zu schaffen, ist eine Gratwanderung: Zu restriktiver Datenschutz legt IoT lahm.
- › In Deutschland gibt es Nachholbedarf in Bezug auf leistungsfähige und sichere Kommunikationsnetze sowie einheitliche Standards. Vor allem diese Defizite wirken sich hemmend auf den weiteren Ausbau von IoT aus.

1. Begriff und Zielsetzung

Kaum ein Begriff steht mehr für den technologischen Fortschritt und für die zu erwartenden gravierenden Umwälzungen von Wirtschaft und Gesellschaft durch die Digitalisierungsbewegung wie der Begriff des *Internet of Things* (IoT, Internet der Dinge). Erstmals verwendet wurde er von Kevin Ashton, der ihn allerdings sehr stark auf die automatische Identifizierbarkeit von Objekten reduzierte (Ashton, 2009). Inzwischen hat sich diese Perspektive erweitert, so dass man heute zu Recht sagen kann, dass das IoT in den kommenden Jahren wesentlicher Treiber und integraler Bestandteil der digitalen Transformation sein wird. Manche vergleichen seine Bedeutung sogar mit der des *World Wide Web* in den späten 90er Jahren (Saarikko et al., 2017, S. 667). Die mit „intelligenten, vernetzten Objekten“ assoziierten Hoffnungen, Potentiale und Chancen mischen sich mit Befürchtungen und Ängsten hinsichtlich ihrer Beherrschbarkeit und ihrer Risiken (Miorandi et al, 2012).

Unter IoT versteht man ein System intelligenter, über ein Kommunikationsmedium vernetzter Produkte (Porter/Heppelmann, 2014, S.66 f.). Beliebige Alltagsgegenstände (physische Objekte, things), wie Haushaltsgeräte, Fahrzeuge, Container, Pumpen, Kleidung etc. werden mit Intelligenz (smart objects) ausgestattet und mit einem Kommunikationsnetz verbunden. Sie erweitern somit das traditionelle *Internet of People*.

Internet of Things zielt darauf ab,

- die Verwendungsmöglichkeiten und das Nutzungsspektrum von sonst nicht oder weniger intelligenten Objekten zu erweitern,
- innovative Anwendungen und digitale Services für Anwender und Nutzer (sowohl Konsumenten als auch Produzenten) zu ermöglichen,
- Ressourcen durch effizienteren Einsatz zu schonen,

- › existierende Geschäftsmodelle effizienter zu gestalten bzw. neue Geschäftsmodelle zu generieren,
- › die Produktivität von Wirtschaftsbereichen zu steigern und
- › die Zufriedenheit der Anwender zu erhöhen.

Anwendungs- und Einsatzmöglichkeiten von IoT finden sich in privaten, aber auch öffentlichen und industriell-gewerblichen Bereichen (Lee/Lee, 2015), wie

- › Smart Home und Smart Meter für das Energiemanagement (Stojkoska/Trivodaliev, 2017),
- › Smart-City-Konzepte,
- › E-Health und E-Care im Medizin und Gesundheitsbereich,
- › Smart Security zur Verbesserung der Sicherheit im privaten und öffentlichen Umfeld,
- › Intelligente Mobilitätssysteme und autonomes Fahren,
- › Intelligente Supply Chains etc.,
- › Intelligente Fertigungs- und Logistiksysteme (Industrie 4.0) und
- › Intelligente Wartungssysteme.

2. Was ist neu an IoT? Architektur des Datenhandlings

Einzelne Aspekte sind keineswegs neu: Alltagsgegenstände, wie Fahrzeuge oder moderne Waschmaschinen, sind heute schon mit einer gewissen Intelligenz ausgestattet. Auch globale Kommunikationsnetze wie das Internet existieren bereits seit längerem.

Neu ist neben der grundsätzlichen Intelligenzenerweiterung bisher nichtintelligenter Objekte (z. B. Schuhe, Uhren, Container) durch eingebettete Prozessoren vor allem die synergetische Kombination der Technologien zur Datenerfassung (Sensortechnologie) und Datenverarbeitung auf der einen Seite sowie der Kommunikations- und Speichertechnologie (Cloudspeicher) auf der anderen Seite (s. Abbildung 1 links). IoT-Objekte erfassen also nicht nur isoliert und lokal ihre eigenen Zustandsdaten und steuern sich dann gemäß einem starren Algorithmus selbst (wie moderne Waschmaschinen), sondern sie öffnen sich über die Kommunikationsschnittstelle anderen Objekten, Speichermedien oder Anwendungen (s. Abbildung 1 rechts). Dadurch können sie ortsungebunden überwacht und gesteuert werden. Aus einem geschlossenen, singulären, lokalen System (wie das einer modernen Waschmaschine) wird ein offenes, globales, mit vielen unterschiedlichen Komponenten bestücktes IoT-System (z. B. *Smart Home*). Diese Systemöffnung ermöglicht aber zugleich schädliche und missbräuchliche Einwirkungen, etwa den unbefugten Abruf von Daten einer Überwachungskamera, das Ausschalten der Kamera oder Abbremsen eines Fahrzeuges.

IoT-Technologien

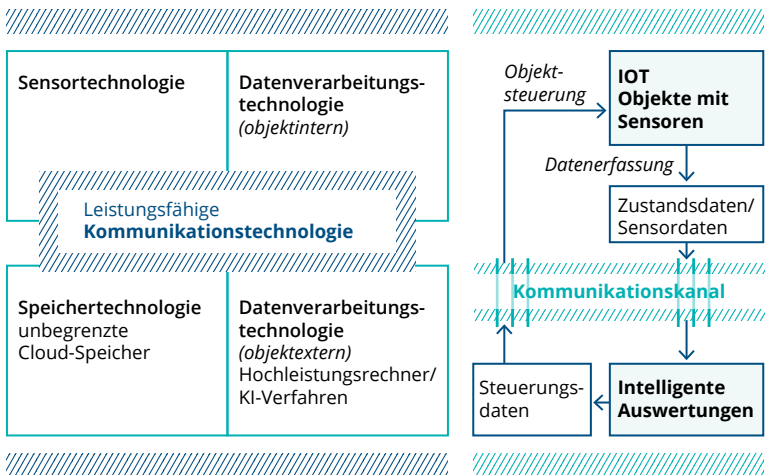


Abbildung 1: Technologiekomponenten und Architektur des Internet of Things

Durch die Kombination solcher innovativer Technologien entstehen neue Applikationen: Mehrere IoT-Objekte gleicher oder unterschiedlicher Art können im Verbund, sogar mit wechselnden Rollen, zusammenarbeiten (autonom gesteuerte Fahrzeuge kommunizieren untereinander und mit den Ampeln der Umgebung zum besseren Management des Verkehrsflusses). Um ein physisches Objekt auch in einem Informations- und Kommunikationssystem als virtuelles (elektronisches) Objekt verwalten und administrieren zu können, muss es zwingend eine Identität erhalten. Nur wenn es eindeutig individuell angesteuert werden kann, lässt es sich gezielt adressieren und können seine Daten korrekt zugeordnet werden. In den ersten Entwicklungsstufen wurde hierfür die RFID-Technologie (radio-frequency identification) verwendet (Atzori et al, 2017). Die auf dem RFID-Chip abgelegte Objektidentifizierungsnummer adressiert eindeutig ein individuelles Objekt. Hierfür wurde mit dem Electronic Product Code (EPC) von der Organisation EPCglobal ein Standard geschaffen, der – analog zu einer Ausweisnummer des Menschen – beliebige Objekte weltweit über einen 96-Bit-Code eindeutig identifiziert (Bassie et al., 2013, S. 222) und ihnen eine „Identität“ verschafft. Dieses elementare, in seiner Bedeutung aber nicht zu unterschätzende Feature erlaubt nicht nur ein dezidiertes Tracking und Tracing mobiler Objekte (ihre permanente Echtzeit-Lokalisation und Nachverfolgung), sondern schafft erst die Basis für eine leistungsfähige Kommunikation in einem Kommunikationsnetzwerk, wie es für IoT benötigt wird.

Grundlegend für die Transformation von Alltagsgegenständen zu IoT-Objekten ist die Ergänzung dieser Objekte um smarte Komponenten (Prozessoren, Sensoren, Kommunikationstechnik). Objektbezogene Daten lassen sich auf einem Mikrochip speichern und/oder über eine Kommunikationsschnittstelle weiterleiten (vgl. Abb. 2). Nachdem die Daten über den Kommunikationskanal (WLAN, Internet etc.) an andere Kommunikationspartner (andere Objekte, Instanzen, Cloud-Speicher etc.) im Kommunikationssystem weitergeleitet wurden, können sie dort ausgewertet werden. Unter einer Instanz in einem Kommunikationssystem wird hierbei ein anderer Kommunikationspartner (Mensch, Maschine, anderes Objekt, Software bzw. App, Organisation) verstanden, der „übergeordnet“ ist und Auswertungsaufgaben

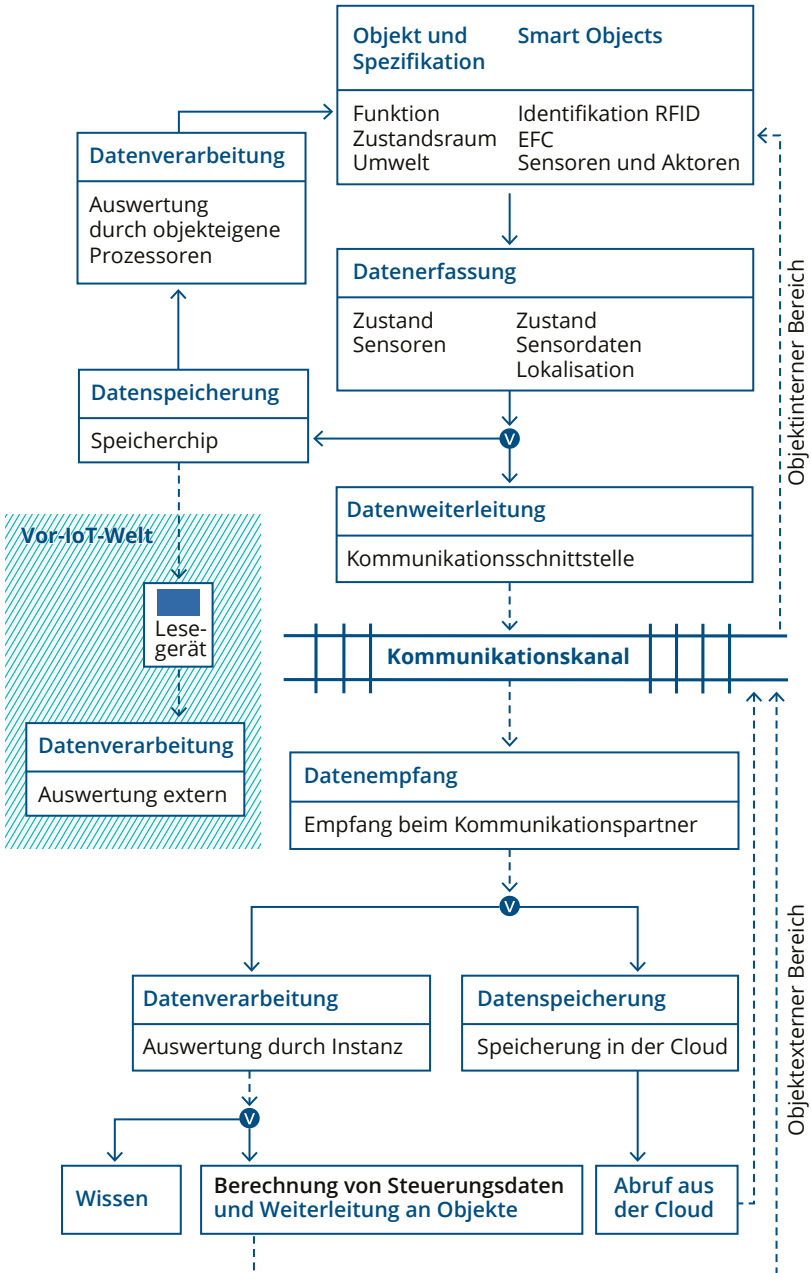


Abbildung 2: Architekturüberblick des Datenhandlings in IoT-Systemen

übernimmt. Ein bisher lokal begrenzter Handlungs- und Wirkungsbereich wird durch die Integration der Kommunikationskomponente aufgebrochen, und es erschließen sich neue, höherwertige, ortsunabhängige Nutzungsmöglichkeiten. Je nach Funktionalität einer Instanz werden aus den empfangenen oder abgerufenen Objektdaten Steuerungsdaten berechnet und über den Kommunikationskanal an das Objekt (oder andere Objekte des Verbunds) zurückgesendet. Hierdurch werden Objektzustandsänderungen beliebig weit entfernter Objekte ausgelöst (z. B. Einschalten der Klimaanlage, Zoomen der Kamera). Mit der Integration von Cloud-Speichern im Kommunikationssystem lassen sich, unabhängig von den begrenzten Speicherkapazitäten in den Objekten selbst, beliebig große Datenmengen erfassen und speichern. Komplette Historien von mannigfaltigen, in kurzen Zeitabständen erfassten Zustands- und Umweltdaten (Bilder, Videos etc.), Big Data also, sind so auswertbar (McAfee/Brynjolfsson, 2012). Man erkennt in Abbildung 2 auch, dass der Kommunikationskanal bzw. das Kommunikationsnetz die kritische, für alle relevanten Aktivitäten zuständige Ressource darstellt. Seine Leistungsfähigkeit ist in besonderem Maße entscheidend für die Leistungsfähigkeit des Gesamtsystems.

Im Schaubild der Abbildung 2 ist im linken Bereich das klassische Vorgehen – vor IoT – bezüglich des Datenhandlings moderner Produkte, etwa moderner Kraftfahrzeuge, skizziert. So werden bei einem Werkstattbesuch durch den lokalen Anschluss eines speziellen Lesegerätes die im Bordcomputer des Fahrzeugs gespeicherten Zustandsdaten mit ihren Änderungen ausgelesen. Allerdings erfolgt dies nur fallweise, wenn Störungen auftreten oder Wartungsarbeiten anstehen, und nur in einer entsprechend ausgestattete Fachwerkstatt („Instanz“). Charakteristisch sind der starke Lokalitätsbezug zum Objekt, die Datenerfassung und -speicherung durch das Objekt selbst, der nur temporäre, von außen angestoßene Datenzugriff und die sehr spezielle, zweckgebundene Auswertung. Genau diese Restriktionen werden durch IoT aufgehoben. Das heißt, der nächste Schritt und damit der Übergang in die IoT-Welt bestünde am Beispiel des Fahrzeugs darin, dass die Fahrzeugdaten nicht nur bei einem Werkstattbesuch, der unregelmäßig, unkontrolliert und oft erst nach Auftritt einer Problemsituation stattfindet, ausgelesen würden, sondern perma-

ment – auch während des Fahrzeugbetriebs – über ein entsprechendes Kommunikationsnetzwerk weitergeleitet und zentral ausgewertet würden (z. B. Komponentenverschleiß, vorbeugende Wartungsmaßnahmen oder Empfehlungen zur Fahrweise).

Die IoT-Welt unterscheidet sich von der Vor-IoT-Welt durch folgende Möglichkeiten:

- › umfassende Datengewinnung und Datensammlung zu vielen Objekten,
- › Globalität (Standortunabhängigkeit von Gerät und Instanz),
- › permanente Verbindungsmöglichkeit,
- › tiefgehende, nicht nur auf Einzelobjekte bezogen Datenanalyse und Datenverwertung und
- › Steuerbarkeit und Kontrolle der Objekte durch beliebig weit entfernte Instanzen.

3. Entwicklungsstand und Anforderungen für die IoT-Umsetzung

Anforderungen an das Kommunikationsnetz und die Datensicherheit

Wegen der fundamentalen Bedeutung der Kommunikationskomponenten in IoT-Systemen braucht es ein sehr leistungsfähiges und sicheres Kommunikationsnetz. Hierzu gehören nicht nur eine hohe Bandbreite des Kommunikationskanals und leistungsfähige, flächendeckende Zugangsmöglichkeiten, sondern auch die Verwendung eines einheitlichen Kommunikationsprotokolls („einheitliche Sprache“) und entsprechender Schnittstellenstandards für die einbezogenen Geräte sowie Sicherheitsmaßnahmen gegen unbefugten Zugriff auf die im Netz transportierten Daten (Ziegeldorf et al., 2014). In Deutschland besteht hier noch beträchtlicher Nachholbedarf. Um

ein unbedingt erforderliches flächendeckendes und leistungsfähiges Internet mit entsprechenden Zugangsmöglichkeiten bereitstellen zu können (Deutschland liegt bezüglich der Internetgeschwindigkeit laut statista 2017 auf Platz 25, weit hinter den USA, Japan, Skandinavien und Südkorea), sind in den nächsten Jahren erhebliche Investitionen in die Infrastruktur erforderlich. Die physikalischen Netze, die heute in Betrieb sind, werden selbst nach Auslastungsoptimierung durch *network slicing* (die Aufteilung der Netzkapazität in parallel nutzbare virtuelle Bereiche) nicht flächendeckend in der Lage sein, die Vielzahl geplanter neuer smarterer Geräte zu versorgen. Mit der notwendigen Infrastrukturinvestition allein ist es aber nicht getan. Auch fehlende Standards hinsichtlich der Schnittstellen hemmen die Entwicklung im IoT-Umfeld. Standardisierungsbemühungen müssten intensiv unterstützt werden. Dies ist besonders schwierig, weil nicht nur nationale und europäische, sondern Interessenten weltweit betroffen sind. Zumindest eine starke europäische Initiative für einheitliche Standards sollte rasch zu Ergebnissen führen.

Um einen nicht manipulierbaren und abhörsicheren Datenaustausch über öffentliche Netze zu schaffen, müssen IoT-Systeme sichere Verbindungen nutzen, entsprechende Übertragungsprotokolle einsetzen und ihre Daten grundsätzlich mit leistungsfähigen Verfahren verschlüsseln. Die smarte Klingel eines *Smart Homes* darf also nicht ihre Daten (Klingelsignal, Sprache, Video) unverschlüsselt über einen ausländischen Server an das Smartphone des Hausbesitzers weiterleiten, wie es viele Apps in diesem Bereich derzeit tun. Zum Schutz der Anwender sollten Verschlüsselungen zwingend vorgeschrieben werden. Trotz Verschlüsselung kann die Datensicherheit gefährdet sein, wenn der empfangende Server (vor allem wenn er sich im nach anderen Kriterien kontrollierten Ausland befindet) einen „Schlüssel zur Entschlüsselung“ besitzt, um anspruchsvollere Auswertungen vornehmen zu können. Es braucht also eigentlich eine Vertrauen garantierende, „zertifizierte“ Serverinstanz – eine Art „Notariats-Server“.

Anforderungen an die IoT-Software.

Wie sollen die Programme erstellt werden?

Im Bereich der Software für IoT-Systeme empfiehlt es sich, wie bei anderen Anwendungen auch, dauerhafte flexibel skalierbare, auf die Benutzerbedürfnisse anpassbare Standardsoftwarelösungen, zumindest für die Basisfunktionalitäten, zu etablieren. Darüber hinaus wird sich ein Markt für höherwertige Services im IoT-Bereich entwickeln (*Smart Security, Smart Mobility, Smart Healthcare, Global Maintenance*-Systeme im industriellen Bereich etc.), mit innovativen Geschäftsmodellen wie differenzierte *Sharing*-Lösungen, *Pay-per-Use*-Konzepten, agilen *Microservices*, Nutzungslizenzverkauf mit diversen Services statt Produktkauf und temporär agierenden, virtuellen Mitarbeiterteams ohne feste Strukturen. Hochwertige IoT-Applikationen erfordern komplexe Auswertungs- und Verarbeitungsprogramme. Smarte Objekte erzeugen große Datenmengen, die möglichst in Echtzeit (*on the fly*) analysiert und genutzt werden sollen. Für derartige Big-Data-Analysen ist eine Vielzahl neuer, komplexer Verfahren und Algorithmen zu implementieren. Doch wer soll sie konzipieren und erstellen, wo bereits heute ein eklatanter Fachkräftemangel im IT-Bereich herrscht?

Die Förderung junger technologieorientierter IoT-Start-ups ist sicher ein wichtiger Ansatz. Allerdings zeigen die bisherigen Erfahrungen, dass neue Wege beschritten werden sollten. Empfehlenswert ist, nicht nach den traditionellen Förderprinzipien mit oftmals sehr risikoscheuen Vergabekriterien immer die gleichen, oft etablierten Gruppen, Personen oder Institutionen zu fördern. Wirklich neue, bahnbrechende Innovationen sind so nicht zu erwarten. Das Silicon Valley verdankt seinen Erfolg auch nicht General Motors, Exxon oder AT&T, sondern kleinen, engagierten und hochqualifizierten Start-up-Unternehmen mit unkonventionellen Ideen. Warum sollte nicht versucht werden, eine bewusst risikoorientierte Förderung hochqualifizierter, kleiner, agiler Innovationsteams durchzuführen, die in der Gesamtschau möglicherweise bessere Ergebnisse liefert? Selbstverständlich sollte das Eingehen derartiger Risiken auch mit einer adäquaten Erfolgsbeteiligung verbunden werden. Wenn man zudem diese kleinen Innovationsteams durch eine agile Dachorganisation (keine Behörde!) beratend unter-

stützen und koordinieren würde, ließen sich weitere Synergieeffekte und eventuell neue Produktideen oder Geschäftsmodelle generieren („Ideenbrüter“).

Das Problem des Fachkräftemangels im IT-Bereich lässt sich kurzfristig sicher nicht umfassend lösen, sondern erfordert Anstrengungen und Reformen im Bildungs- und Ausbildungsbereich. Dabei reicht es nicht aus, lediglich zusätzliche Studienplätze in Informatik, Wirtschaftsinformatik, Angewandte Informatik etc. einzurichten. Bereits in der Schule müssten die Neigung und das Interesse an solchen Fächern und Inhalten stärker geweckt und Kompetenzen systematisch und gezielt aufgebaut und gefördert werden, insbesondere bei den in mathematisch-technischen Fächern bis heute unterrepräsentierten Frauen. Dafür braucht es wiederum geeignetes Ausbildungs- und Lehrpersonal.

4. Anforderungen an den Datenschutz. Wem gehören die Daten?

An dem Beispiel der Kraftfahrzeuge ist im Übergang von der Nicht-IoT-Welt in die IoT-Welt ein weiteres Problem zu erkennen, das aus Gründen der Akzeptanz in der Bevölkerung und der Aufrechterhaltung einer demokratischen Wirtschafts- und Gesellschaftsordnung zwingend für alle unterschiedlichen Interessengruppen gelöst werden muss: der Datenschutz.

Am erwähnten Beispiel der durch Fahrzeuge gesammelten Daten und ihrer Verwendung wird die Frage virulent, wer eigentlich auf die Objektdaten zugreifen darf. Wem gehören diese Daten? Wer hat die Kontrolle über die Daten, wenn das Objekt selbst (ein Fahrzeug, ein Container, ein Kühlschrank, ein Sportschuh) kein Träger eines Rechtsguts bzw. Rechtssubjekt sein kann? Gehören die Daten dem Hersteller, der aus ihnen Wissen zur Kundennutzung seiner Produkte, der Produktqualität und zur Produktverbesserung generieren kann? Oder dem Software-Lizenzgeber, dessen Betriebssystem die Produktnutzung steuert (z. B. Apple für iPhones)? Oder gehören die mit dem Objekt verknüpften Daten dem Eigentümer des Objektes (des Fahrzeugs), der die rechtliche

Verfügungsgewalt über das Objekt hat und der das Produkt gekauft hat? Kann er den Zugriff auf die mit dem Objekt verknüpften Daten beliebig öffnen bzw. einschränken? Oder kann eine übergeordnete Instanz, etwa die Verkehrsleitzentrale einer Stadt, ein Versicherungsunternehmen oder staatliche Stellen, wie das Finanzamt, der Zoll oder die Polizei, auf diese Daten zugreifen? Verschärft wird diese Problematik dadurch, dass die Daten eines Objekts (Fahrzeug, Kamera, Fitnesstracker etc.) nicht unbedingt im Objekt selbst (etwa einem Speicherchip) abgelegt sein müssen (und damit physisch mit dem Objekt verbunden sind), sondern nach der Erfassung sofort an eine IoT-Instanz oder in eine Daten-Cloud weitergeleitet werden. Der Cloud-Service-Anbieter verwaltet diese Daten in gesammelter Form und könnte sie, sofern nicht verschlüsselt, auch auswerten. Und dies möglicherweise ohne explizite Autorisierung oder sogar ohne Wissen des Objekteigentümers.

Diese Art der unautorisierten und unwissentlichen Datenweiterleitung von Objektdaten und ihre externe Verwertung mögen als unglaubliches und unverfrorenes, theoretisches Szenario erscheinen. Es ist aber längst Realität. Wenn Google „übliche“ Besucherzahlen von Geschäften oder Restaurants bei Suchanfragen automatisch bereitstellt, stammen diese Informationen aus solchen Datenquellen. Objekte, in diesem Fall Smartphones, werden lokalisiert und getrackt, ohne dass dies dem Smartphone-Besitzer bewusst ist. Verknüpft man die erfassten Lokalisationsdaten von Smartphones mit den (festen und bekannten) geographischen Positionsdaten der Beobachtungsobjekte (eines Restaurants, Shops, Zoos etc.) erhält man eine detaillierte Übersicht über die Besuchsintensität der Objekte in den jeweiligen Zeiträumen. Dies erfolgt ohne explizites Einverständnis und oft sogar ohne Wissen des Verantwortlichen für das Beobachtungsobjekt (des Shop-Betreibers, Restaurantbesitzers oder -pächters etc.). Auch wenn der Google-Nutzer lediglich aggregierte Daten zur Besuchsintensität für einzelne Zeiträume erhält, so berechnet Google diese unter Verwendung objektindividueller Identifikationsdaten (z. B. Mobilfunknummer oder Geräte-ID). Das heißt, Google weiß nicht nur, wie viele Besucher derzeit in einem Shop sind, sondern auch welche Objekte – sprich Smartphones – sich derzeit dort befinden. Google weiß auch, wer sich zu welcher Zeit jemals dort befunden hat und wie lange er sich aufgehalten hat!

Das Argument, es seien doch „nur“ Objektdaten und nicht von Datenschutzvorschriften wie der Europäischen Datenschutzgrundverordnung erfasste Personendaten, ist angesichts der Verknüpfungsmöglichkeiten absurd. Mit einem Smartphone oder einem Fahrzeug ist zumeist nur eine Person, eventuell eine kleine, leicht identifizierbare Personengruppe (z. B. Familie) assoziiert. Auch wenn nur reine Objektdaten (Lokalisierung, Status, Umweltdaten) erfasst werden, mutieren diese über mehr oder weniger komplexe Zuordnungsfunktionen zu personenbezogenen Daten, die eigentlich besonderen datenschutzrechtlichen Bestimmungen unterliegen müssten. Mit leistungsfähigen Mustererkennungsverfahren (z. B. Neuronale Netze und *Deep Learning*) ist es nicht schwierig, aus einer Gruppe die in einem bestimmten Zeitraum relevante Person zu einem Objekt zu identifizieren – durch typische, individuelle Verhaltensmuster bzw. Nutzerprofile, die wie Fingerabdrücke verwendet werden können (bei Fahrzeugen etwa Brems- und Beschleunigungsverhalten, Durchschnittsgeschwindigkeit, typische Fahrdauer etc.). Insofern müssen zumindest „personennahe“ Objektdaten wie persönliche Daten behandelt werden.

Allerdings ist zu bedenken, dass ein zu restriktiver Datenschutz IoT-Systeme lahmlegen könnte (Weinberg et al., 2015). Ihr effektives Funktionieren ist auf die Bereitstellung entsprechender Daten zwingend angewiesen und oftmals wesentlicher Teil der Geschäftsmodelle von IoT-Unternehmen. Je stärker assistenzbezogen und je genauer IoT-Services auf die persönlichen Bedürfnisse des Anwenders zugeschnitten sein sollen (bei *Smart Healthcare*-Systemen zwingend notwendig), umso mehr personenbezogene Daten braucht es. In derartigen Fällen sollte jeder Anwender bewusst und aktiv darüber befinden, wie weit seine Bereitschaft zur Datenüberlassung geht, um diese Dienste nutzen zu können.

5. Anforderungen an Autonomie und Verantwortlichkeit. Wer kontrolliert die Objekte?

Ein über Datensicherheit und Datenschutz hinausgehender, damit aber eng verknüpfter Aspekt betrifft die Frage, wer eigentlich die Objekte eines IoT-Systems kontrolliert. Wer ist für die Folgen von Objektaktivitäten verantwortlich? Was bei traditionellen, klassischen Produkten mit ihrer lokalen Autonomie selbstverständlich war, nämlich, dass der Besitzer der Produkte sie auch vollständig und eigenständig kontrollieren konnte und damit Folgen verantwortete, ist bei smarten Objekten, insbesondere bei solchen in einem IoT-System, keineswegs klar. Oft funktionieren smarte Produkte nur noch mit einer entsprechenden Software und mit einer zumindest temporären Anbindung an das Kommunikationsnetz. Welchen Einfluss haben dann die Hersteller oder Vertreiber der für die Funktionsfähigkeit elementaren Software? Sie kennen als einzige vollständig ihre Programme und Algorithmen und brauchen für Programm-Updates Zugriffsmöglichkeiten. Wer verantwortet Programmfehler, die zu Schäden durch die von diesen Programmen gesteuerten Objekten führen, wenn durch Softwarefehler in einem *Smart Health*-System Patienten geschädigt werden oder wenn durch den Ausfall von Ampelsystemen oder autonom fahrenden Fahrzeugen das komplette Verkehrssystem einer Stadt oder Region zusammenbricht? Softwarefehler in einem nicht einsehbaren Programm ist die eine Seite des Problems, bewusste Manipulations- und Eingriffsmöglichkeiten durch die Softwarehersteller oder Hacker die andere. Sie wären in der Lage, aus welchen Motiven auch immer, ein smartes Objekt bewusst außer Funktion zu setzen oder gar unautorisiert fernzusteuern. So könnten sicherheitsrelevante Objekte (z. B. Flugzeuge, Militärfahrzeuge) durch ein Softwareupdate oder eine von Anfang an eingeplante, heimliche *Backdoor* (Teil einer Software, der einen Zugang zum Programm unter Umgehung der normalen Zugriffssicherung ermöglicht) außer Betrieb gesetzt werden. In einfacherer Form wird dies bereits heute bei einigen Militärflugzeugen praktiziert, wo man für jeden Start einen jeweils neu beim Hersteller anzufordernden Sicherheitscode benötigt.

Auch hier ist dringend Handlungsbedarf vonnöten. Eine systemrelevante Software sollte nicht mehr als Blackbox gekauft bzw. genutzt werden: Die Zuständigen brauchen vollständigen Einblick in die hochkomplexen Programme. Unkontrollierte und unautorisierte *Backdoor*-Zutritte müssen untersagt werden. Je nach Anwendungsbereich sind zudem Maßnahmen zur Risikobegrenzung (z. B. Notfallpläne, unabhängige, objektautonome Mindestfunktionalitäten) aufzubauen und einzurichten.

6. Fazit

Internet of Things ist ein wesentlicher Treiber und Bestandteil der künftigen Digitalisierung von Wirtschaft und Gesellschaft. IoT erweitert das Internet als globales Kommunikationsnetz, indem smarte Produkte und Gegenstände (*things*) als weitere „Teilnehmer“ bzw. „Kommunikationspartner“ integriert werden. Sie sind zum einen in der Lage, eine immense Menge zusätzlicher, vielfältiger Daten zu erfassen und im Netz für Auswertungen bereitzustellen. Zum anderen lassen sich die Objekte und ihr Zustand auch von überall her über das Kommunikationsnetz gezielt ansprechen und steuern. Hierdurch eröffnen sich neue, innovative Services und Geschäftsmodelle (Huber/Kaiser, 2015), die die Digitalisierung von Wirtschaft und Gesellschaft vorantreiben werden und zu erheblichen Wohlfahrtsgewinnen, aber auch zu schwierigen Umbrüchen in der Arbeitswelt und der gesellschaftlichen Organisation führen können. Aufgrund der vielfältigen, global ausgerichteten Verflechtungen zwischen den Objekten und Instanzen sowie ihrer Abhängigkeit von koordinierenden Steuerungseinheiten und ihren Softwareprogrammen steigen die Systemrisiken erheblich. Datensicherheit und Datenschutz werden vor neue, hohe Herausforderungen gestellt.

Um diese divergierenden und konfliktären Anforderungen und Interessen systematisch analysieren und ihnen begegnen zu können, wäre die Einrichtung eines alle Stakeholder berücksichtigenden Gremiums, eines „Technologie- und Ethikrates“, sinnvoll. Dieser sollte nicht, wie ähnliche bereits existierende Gremien, in erster Linie bremsend

wirken, sondern beide Facetten der IoT – die immensen ökonomisch-gesellschaftlichen Potentiale wie auch die Risiken – gleichermaßen in den Fokus nehmen und versuchen, eine Art gesellschaftlichen Konsens für IoT-Applikationen zu erarbeiten. Er sollte gezielt die Chancen der IoT-Technologie verdeutlichen und fördern, positive und negative, wirtschaftliche und gesellschaftliche Folgen analytisch untersuchen und die berechtigten Bedenken und Risiken für Einzelne, die Gesellschaft und die Wirtschaft kritisch diskutieren. Zielsetzung wäre die Erarbeitung entsprechender Rahmenbedingungen (herstellerübergreifende Sicherheitskonzepte, Datenschutzvorgaben etc.) und Handlungsempfehlungen zur IoT-Entwicklung und -Nutzung für die Politik und ihre Gremien.

Referenzen

- A Ashton, K.** (2009). That “internet of things“ thing. In *RFID Journal*, 22(7), S. 97–114.
- Atzori, L., Iera, A., Morabito, G.** (2017). Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm. In: *Ad hoc Networks*, 56 (2017), S. 122–144.
- B Bassi, A., Bauer, M., Fiedler, M., Kramp, T., van Kranenburg, R., Lange, S., Meissner, S.** (Hrsg.) (2013). *Enabling Things to Talk – Designing IoT solutions with the IoT Architectural Reference Model*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- H Huber, D., Kaiser, T.** (2015). Wie das Internet der Dinge neue Geschäftsmodelle ermöglicht. In *HMD – Praxis der Wirtschaftsinformatik*, 52(5), S. 681–689.
- L Lee, I., & Lee, K.** (2015). The internet of things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, 58(4), S. 431–440.
- M McAfee, A., & Brynjolfsson, E.** (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), S. 60–68.
- Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I.** (2012). Internet of things: Vision, applications, and research challenges. In *Ad Hoc Networks*, 10(2012), S. 1497–1516.
- N Ng, I.C.L., Wakenshaw, S.Y.L.** (2017). The Internet-of-Things: Review and research directions. In: *International Journal of Research in Marketing* 34 (2017), S. 3–21.

- P** **Porter, Michael; Heppelmann, James** (2014). How smart, connected products are trans-forming competition. In: Harvard Business Review, Heft 11 (2014), S.64–88.
- S** **Saarikko, Ted, Westergren, Ulrika H., Blomquist, Tomas** (2017). The Internet of Things: Are you ready for what’s coming? In Business Horizons 60 (2017), S. 667–676.
- Stojkoska, B.L.R., Trivodaliev, K.V.** (2017). A review of Internet of Things for smart home: Challenges and solutions. In: Journal of Cleaner Production 140 (2017), S. 1454–1464.
- W** **Weinberg, B. D., Milne, G. R., Andonova, Y. G., & Hajjat, F. M.** (2015). Internet of things: Convenience vs. privacy and secrecy. Business Horizons, 58(6), S. 615–624.
- Z** **Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K.** (2014). Privacy in the Internet of Things: Threats and challenges. In Security and Communication Networks, 7(12), S. 2728–2742.

Künstliche Intelligenz und Robotik

Oliver Brock

- › Oft ist das, was wir als künstliche Intelligenz bewundern, nur das Vermögen, große Datenmengen schnell und effektiv zu verarbeiten, etwa wenn Computer Schach oder Go spielen. Derartige Fähigkeiten, die nicht auf einem Verständnis von biologischer Intelligenz beruhen, können wir als Silicon Intelligence bezeichnen, da sie eng mit Fähigkeiten silizium-basierter Informationstechnologie verknüpft sind.
- › Die Differenzierung zwischen Silicon Intelligence und „künstlicher Intelligenz“ – also Intelligenz in einem anspruchsvollen Sinne, die der menschlichen Intelligenz nahekommt – wäre hilfreich, um mehr Klarheit in die Diskussion zu bringen: Wenn es das Ziel künstlicher Intelligenz ist, der menschlichen Intelligenz nahe zu kommen, wo stehen wir dann zurzeit? Was ist in naher Zukunft zu erwarten? Welche Schritte müssen wir gehen?
- › Eine solche differenzierte Betrachtung – in Politik und Gesellschaft, aber auch in der Wissenschaft – könnte dazu beitragen, KI-Forschung noch erfolgreicher zu machen, indem der Blick nicht verengt wird und interessante Optionen nicht außer Acht gelassen werden.

Seit Tausenden von Jahren ist die Menschheit bestrebt, Fähigkeiten von Lebewesen in Maschinen nachzuahmen und zu übertreffen und sich die daraus resultierenden Artefakte – z. B. Roboter – zu Nutzen zu machen. Nie zuvor waren die Menschen ihrem Ziel so nahe, und nie zuvor war die Notwendigkeit des Gelingens dieses Unterfangens größer als heute.

Wir haben bereits Maschinen geschaffen, die stärker und schneller, präziser und ausdauernder sind als der Mensch, die fliegen, schwimmen und tauchen, ja sogar durch das Weltall zu anderen Planeten fliegen. Wir haben Maschinen gebaut, die in Sekundenschnelle Milliarden von Rechenoperationen durchführen oder global verteilte, schier unermessliche Mengen an Daten durchsuchen. Kurzum: In vielen Aspekten übertreffen unsere Erfindungen bereits die Natur.

In einem Aspekt kann jedoch keine Maschine der Natur auch nur annähernd das Wasser reichen: bei der natürlichen Intelligenz. Gemeint ist die wissenschaftlich noch unzulänglich verstandene Fähigkeit vieler Lebewesen, komplexe Probleme zu lösen, sich der Umwelt anzupassen, vor Schwierigkeiten gestellt immer wieder Lösungen zu finden und letztendlich auch das Vermögen, diese Fähigkeiten untereinander weiterzugeben.

1. Was ist eigentlich „Intelligenz“?

Noch verstehen wir nicht, was genau Intelligenz beinhaltet. Wir wissen nicht, welche Teilfähigkeiten sich auf welche Art und Weise kombinieren, um Menschen und Tiere zu intelligentem Verhalten zu befähigen.

Viele wissenschaftliche Disziplinen befassen sich mit natürlicher Intelligenz: Psychologie, Neurowissenschaften, Verhaltensbiologie, Erziehungswissenschaften, Philosophie. Jede dieser Disziplinen produziert wichtige Erkenntnisse. Und doch ergibt das Zusammenfügen dieser Fragmente noch kein konsistentes Gesamtbild. Es existieren viele Definitionen von Intelligenz, aber keine von ihnen kann als

abschließend oder vollständig gelten. Noch wissen wir nicht, was natürliche Intelligenz ausmacht. Aber wir glauben, sie zu erkennen, wenn wir sie sehen.

Trotz unseres mangelnden Verständnisses von Intelligenz lesen wir immer wieder Berichte, die Fortschritte – ja sogar Durchbrüche – in der künstlichen Intelligenz verkünden. Ist es nicht widersprüchlich, über künstliche Intelligenz zu forschen, wenn wir noch nicht einmal genau wissen, was natürliche Intelligenz ist?

2. Moravecs Paradox

Bereits in den 1980er Jahren gelangten Forscher auf dem Gebiet der Künstlichen Intelligenz zu der Erkenntnis, dass es relativ einfach ist, Computern vermeintlich hohe kognitive Leistungen, wie z. B. das Schachspielen, beizubringen. Ungleich schwerer ist es, ihnen Fähigkeiten zu verleihen, die für ein dreijähriges Kind selbstverständlich sind: z. B. Wahrnehmung, Mobilität und manuelle Manipulation der Umwelt. Mit anderen Worten: Das, was Menschen automatisch und ohne bewusste Mühe tun, ist für künstliche Intelligenz die eigentliche Herausforderung. Was Menschen jedoch aneinander bewundern, etwa die Fähigkeit siebenstellige Zahlen im Kopf zu multiplizieren, einen Großmeister im Schach zu schlagen oder sich Hunderte von Begriffen zu merken, ist mit Maschinen einfach zu realisieren. Diese Einsicht bezeichnet man als *Moravecs Paradox*, benannt nach Hans Moravec, einem Pionier der Erforschung der künstlichen Intelligenz.

Moravecs Paradox liefert uns eine intuitive Unterteilung von Fähigkeiten in solche, die wir relativ einfach in Technologie übertragen können, und solche, deren Übertragung uns noch vor erhebliche Rätsel stellt.

Betrachten wir mit dem Wissen um Moravecs Paradox den Stand der Wissenschaft, gilt es festzuhalten: Wir feiern technische Errungenschaften in der künstlichen Intelligenz als Meilensteine, die nach Moravecs Paradox zu den „einfach“ zu erzielenden Fähigkeiten zählen.

Als Beispiele seien das Programm AlphaGo (Google) genannt, das den Weltmeister im Brettspiel Go besiegte, oder IBM Watson, ein Computerprogramm, das wohl jeden Menschen im Spiel Jeopardy! besiegt (wobei nicht klar ist, ob Watson auch gegen einen Menschen gewinnen könnte, der Zugang zu großen Datenmengen hat, so wie Watson) und nicht zuletzt Deep Blue, das Programm, das bereits 1997 den Schachweltmeister Gary Kasparow besiegte.

All diese Errungenschaften sind ausgesprochen beeindruckende Fortschritte der Wissenschaft. Aber es ist fraglich, ob auf Grundlage dieser Fortschritte auch Fähigkeiten erzeugt werden können, die laut Moravecs Paradox schwierig und somit exemplarisch für natürliche Intelligenz sind.

Wir können aus Moravecs Paradox folgende Lehre ziehen: Ein Mensch ist dann von einem anderen Menschen beeindruckt, wenn dieser Dinge tut, zu denen ersterer selbst nicht imstande ist. Den Weltmeister in Go zu schlagen, ist für Menschen außergewöhnlich. Wenn dies aber einer Maschine gelingt, spielt sie einen unfairen Vorteil aus: ihre ungeheure Rechenkraft. Sie verfügt jedoch über keine generelle Intelligenz, sondern nur über eine Art Inselbegabung. Ein Mensch hingegen, der sehr gut Go spielt, ist intelligent – vermutlich sogar überdurchschnittlich intelligent. Wir bewundern ihn, weil er wahrscheinlich auch noch viele andere Dinge besser kann als man selbst. Eine Maschine kann aber ausschließlich Go spielen, auch wenn wir versucht sind, ihr weitere Fähigkeiten zuzuschreiben.

3. Einfache und schwierige Herausforderungen für natürliche Intelligenz

Wenn also die Fähigkeit, Aufgaben wie Schach, Go oder Jeopardy! zu lösen, kein hinreichendes Indiz für Intelligenz ist, wie müssen dann Problemstellungen aussehen, damit man an ihnen Intelligenz entwickeln und testen kann? Diese Frage soll anhand von drei Beispielen aufsteigender Komplexität erläutert werden. Dabei behalten wir Moravecs Paradox im Hinterkopf: nämlich, dass sich die größten

Herausforderungen aus der Interaktion mit der echten Welt ergeben (Wahrnehmung, Bewegung, Manipulation usw.).

Schach

Die Welt des Schachs besteht aus 64 Feldern und 32 Figuren. Der Zustand dieser Welt ist durch Zuweisung der Figuren auf die Felder unmissverständlich und vollständig repräsentiert. Dank festgeschriebener Bewegungsmuster der Figuren sind die Handlungsmöglichkeiten klar definiert und führen zu einem eindeutigen Ergebnis. In gewisser Weise verstehen wir die Welt des Schachspiels vollständig. Die Herausforderung besteht also ausschließlich darin, die möglichen Handlungsstränge des Spiels zu durchsuchen: Wir sprechen von einem Suchproblem. Bei solchen Suchproblemen helfen Rechenkraft und Problemverständnis. Genauso verhält es sich bei dem Brettspiel Go.

Reine Suchprobleme, wie Schach und Go, sind weit entfernt von Herausforderungen, die natürliche Intelligenz zu bewältigen hat. Inwieweit eine solche künstlicher Intelligenz die Bezeichnung „Intelligenz“ wirklich verdient, ist damit offen.

Autonomes Fahren

Autonomes Fahren findet naturgemäß in der realen Welt statt. Ein autonomes Fahrzeug muss mit der Komplexität der realen Welt verlässlich umgehen können. Somit handelt es sich beim autonomen Fahren um eine grundlegend anders geartete Problemkategorie als beim Schach. Suchprobleme spielen hier zwar auch eine Rolle, aber zusätzlich gibt es große Herausforderungen bei der korrekten Wahrnehmung in der realen Welt.

Auch beim autonomen Fahren hören wir immer wieder von Fortschritten und Durchbrüchen. Dennoch haben sich alle zeitlichen Prognosen bisher nicht erfüllt und autonome Fahrzeuge bleiben auf unseren Straßen eine Seltenheit. Der Umgang mit Ausnahmesituationen stellt für autonome Fahrzeuge immer noch eine große Herausforderung dar: ein Indiz, dass das Wahrnehmungsproblem noch nicht gelöst ist.

Können wir autonomen Fahrzeugen bereits Intelligenz zuschreiben? Eine eindeutige Antwort gibt es nicht. Im Sinne von Moravecs Paradox ist autonomes Fahren aber ein Problem, das fundamental schwerer zu lösen ist als Schach, eben weil es auch Wahrnehmung erfordert.

Escape Rooms

Escape Rooms erfreuen sich als Gesellschaftsspiel großer Beliebtheit: Die Spieler werden in einem realen Raum eingeschlossen, aus dem es zu entkommen gilt. Alles, was man für die Flucht benötigt, ist in diesem Raum vorhanden. Die Flucht gelingt dann, wenn mechanische und logische Rätsel unter Verwendung von Allgemeinwissen gelöst werden. So könnten zum Beispiel in einer in diesem Raum deponierten Zeitschrift Zahlen angestrichen sein, die die Kombination für ein Zahlenschloss liefern, das wiederum eine Kiste öffnet, in der ein Schlüssel liegt, der die Tür des Raums öffnet.

Ein Entkommen aus einem *Escape Room* ist für eine künstliche Intelligenz eine deutlich anspruchsvollere Aufgabe als Schachspielen oder Autofahren. Wir würden jemandem, dem es gelingt, aus einem *Escape Room* zu entkommen, ohne Zweifel für intelligent halten.

Wie beim autonomen Fahren muss im *Escape Room* die Umwelt wahrgenommen werden, um daraus Handlungen abzuleiten. Allerdings müssen im *Escape Room* nicht nur Hindernisse und Verkehrszeichen erkannt werden. Darüber hinaus ist es notwendig, die Funktionsweise der Gegenstände zu erkennen und sie in Bezug zu setzen (z. B. Schlüssel und Schloss). Außerdem sind die Handlungen vielfältiger als die Tätigkeiten beim Autofahren: Man muss Türen öffnen, Schlüssel benutzen, Bücher durchblättern oder Gegenstände bewegen.

Hier mag man irritiert sein: Jedes Kind exploriert seine Welt, wie in einem *Escape Room*. Etwa indem es Schubladen öffnet und deren Inhalt auf dem Boden verteilt. Autofahren hingegen muss man beigebracht bekommen, genauso wie Schach. Die meisten Menschen sind in der Lage, ein Auto zu steuern, aber nur wenige vermögen es, auf hohem Niveau Schach zu spielen. Uns erscheint also Schach spielen und Autofahren schwerer als der *Escape Room*. Genau dies ist die Bedeutung von

Moravecs Paradox: Unsere Intuition trügt, und es ist eben genau das, was jeder Mensch bereits kann, was natürliche Intelligenz ausmacht.

Nun können wir den skizzierten Widerspruch im Detail auflösen und erklären, warum immer wieder Durchbrüche in der künstlichen Intelligenz gemeldet werden, die sich im Grunde genommen auf Suchprobleme beziehen. Es liegt einfach daran, dass es umgangssprachlich zwei unterschiedliche Intelligenzbegriffe gibt. Zum einen wird als künstliche Intelligenz bezeichnet, wenn Computer Probleme lösen, die *beim Menschen* auf das Vorhandensein von Intelligenz schließen lassen (etwa die Fähigkeit, erfolgreich Schach, Go oder Jeopardy! zu spielen). Dieser Rückschluss ist aber bei einem Computer nicht zulässig. Computer, die Schach oder Go spielen können, sind keineswegs auch intelligent!

Zum anderen gibt es den Begriff der „künstliche Intelligenz“ als Bezeichnung für das Verständnis und die Technologisierung natürlicher Intelligenz. Hier steht die Wissenschaft noch ganz am Anfang, trotz wiederholter, anders lautender Meldungen in der Presse (und auch von einigen Forschern, die eher über das Potential ihrer Forschungsarbeit sprechen als über das bereits Erreichte).

Um im Folgenden zwischen den beiden Bedeutungen des Begriffs künstliche Intelligenz zu unterscheiden, nenne ich die Fähigkeit, Probleme durch effiziente Suche oder mithilfe riesiger Datenmengen zu lösen, *Silicon Intelligence*. Diese Fähigkeit beruht unter anderem darauf, dass Transistoren aus Silizium (engl.: silicon) viel schneller schalten können als biologische Neuronen und dass elektronische Datenspeicher transparente Auslesemechanismen besitzen.

Diejenige künstliche Intelligenz, die natürliche Intelligenz nachzuempfinden versucht, nenne ich dagegen weiterhin „künstliche Intelligenz“. Verwirrenderweise wird „künstliche Intelligenz“ auch als Bezeichnung eines Forschungsfeldes verwendet: das ich im Folgenden mit der Abkürzung „KI“ bezeichne.

4. Forschungsfelder der künstlichen Intelligenz

Künstliche Intelligenz (KI)

KI ist ein Überbegriff für viele Disziplinen und eine weit gefasste Umschreibung von wissenschaftlicher Beschäftigung mit künstlicher Intelligenz im weitesten Sinne. Maschinelles Lernen und auch große Teile der Robotik sind Untergebiete der KI. Obwohl das Feld sehr vielseitig ist, gibt es eine Reihe von Paradigmen, mit denen Wissenschaftler versucht haben, künstliche Intelligenz zu erschaffen.

Die Anfänge der KI fokussierten sich auf Symbolverarbeitung: Intelligenz wurde als Manipulation von Formeln verstanden. Als sich die gewünschten Erfolge nicht schnell genug einstellten, kamen Zweifel an der Hypothese von Intelligenz als Symbolverarbeitung auf. Es entstand Raum für neue Ideen, wie zum Beispiel den Konnektionismus. Dabei wurde versucht, die vermutete prinzipielle Funktionsweise des menschlichen Gehirns mithilfe von sogenannten neuronalen Netzen im Computer nachzuempfinden. Aber auch hier stagnierte die Forschung, fundamentale Einwände wurden erhoben. Neuronale Netze galten dann jahrzehntelang als Nischentechnologie, bis sie vor wenigen Jahren durch *Deep Learning* wieder ins Scheinwerferlicht gerieten – mehr dazu später. Weitere Paradigmen in der KI sind die Bayesianische Wahrscheinlichkeitsrechnung (benannt nach dem englischen Mathematiker Thomas Bayes) und die Statistik. Beide Methoden sind immer noch weit verbreitet. Es gibt eine Vielzahl weiterer Ansätze, wie die evolutionären Algorithmen, die versuchen, Evolution im Rechner nachzuempfinden.

Ein zunehmend an Bedeutung gewinnender Zweig der KI wird als „*nouvelle KI*“ bezeichnet. Ihr Ansatz beruht auf der Einsicht, dass Körper und Geist im Kontext intelligenten Verhaltens nicht zu trennen sind: Sie bilden eine Einheit, die intelligentes Verhalten erzeugt. Der Körper muss also in die Untersuchungen des Intelligenzbegriffs einbezogen werden.

Die *nouvelle KI* berücksichtigt darüber hinaus auch die Umwelt als Teil des intelligenten Systems. Denn bei vielen Alltagshandlungen ist die

Umwelt maßgeblich an intelligentem Verhalten beteiligt. Man denke nur an die Verwendung von Erinnerungszetteln oder eines Smartphone zur Navigation.

In jedem der genannten Bereiche wird weiterhin geforscht. Aber keiner von ihnen konnte sich bisher nachhaltig als dominierender Ansatz der KI-Forschung durchsetzen. Obwohl es nicht ausgeschlossen ist, dass dies künftig noch geschieht, scheint eine Kombination der verschiedenen Ansätze vielversprechender. Selbstverständlich können auch völlig neue Paradigmen entstehen, etwa durch die Fortschritte von Quantencomputern.

Maschinelles Lernen (ML)

ML ist eine Teildisziplin von KI und befasst sich vornehmlich mit der Entwicklung von (Lern-)Methoden. Diese Methoden dienen der Lösung verschiedenster Probleme. Zurzeit ist das *Deep Learning* von herausragender Bedeutung, aber auch Methoden des statistischen Lernens (der dem *Deep Learning* vorausgegangene Trend) finden nach wie vor große Beachtung.

Methoden des maschinellen Lernens finden überall dort Verwendung, wo es Menschen unmöglich oder noch nicht gelungen ist, explizite Lösungsverfahren zu beschreiben. Dies betrifft auch Probleme in der echten Welt – wenn Daten verrauscht und unvollständig sind. Menschen fällt es z. B. nicht schwer, einen bestimmten Gegenstand auch unter wechselnden Beleuchtungssituationen oder bei teilweiser Verdeckung durch einen anderen Gegenstand wiederzuerkennen. Für Computer ist dies weiterhin eine große Herausforderung. ML hilft ihnen dabei, diese Herausforderungen zu meistern.

Robotik

Im Gegensatz zum maschinellen Lernen widmet sich die Robotik nicht vorrangig der Entwicklung von Methoden, sondern der Lösung von spezifischen „robotischen“ Problemen. Dafür wird eine Vielzahl von Methoden (auch aus dem maschinellen Lernen) herangezogen – oder es werden völlig neue Methoden entwickelt.

Der Fokus auf konkrete Probleme ist für die Erforschung der künstlichen Intelligenz unerlässlich. Da es keine universelle Intelligenz geben kann, muss jede Form der Intelligenz auf eine Problemkategorie zugeschnitten sein. In der Evolution bezeichnet man diese Problemkategorie als ökologische Nische. Der Grad an Intelligenz lässt sich daran ablesen, inwieweit ein intelligent handelnder Agent die Eigenschaften der ökologischen Nische ausnutzt, um sein Verhalten vorteilhaft anzupassen. Die Robotik ist aufgrund der Berücksichtigung des Körpers (des Roboters) und ihres Fokus auf spezifische Probleme als Disziplin für die Erforschung der künstlichen Intelligenz prädestiniert.

Derzeit teilt man die Robotik in zwei Untergebiete auf, die sich einander immer weiter annähern. Zum einen die Automatisierungstechnik, die aus vielen Wirtschaftszweigen nicht mehr wegzudenken ist. Sie löst Probleme, die sich aus den Anforderungen der Fertigung von industriellen Produkten oder der Bereitstellung entsprechender Dienste ergeben. Und dann gibt es noch jenen Teil der Robotik, der sich mit künstlicher Intelligenz befasst. Hier gilt es, durch Konstruktion eines Artefakts (z. B. Roboter) ein Verständnis natürlicher Intelligenz zu erlangen, gemäß dem Leitsatz von Richard Feynman „*What I cannot create, I do not understand.*“ Dieser Teil der Robotik ist eine Teildisziplin der KI.

Moravecs Paradox lehrt uns, dass Wahrnehmung und Handeln wichtige, wenn nicht gar zwingende Aspekte von natürlicher Intelligenz sind: Robotik ist also notwendig, um Intelligenz zu verstehen, da die Interaktion zwischen Agent und Umwelt – als maßgebliches Merkmal natürlicher Intelligenz – das ist, womit sie sich naturgemäß befasst. Die Robotik spielt daher bei der Erforschung der künstlichen Intelligenz eine herausragende Schlüsselrolle.

5. Wissenschaftliche und technologische Trends

Deep Learning

Deep Learning ist eine Weiterentwicklung des Konnektionismus, des maschinellen Lernens mit (künstlichen) neuronalen Netzen. Seit

einigen Jahren liefert die Erforschung der tiefen neuronalen Netze in mehreren Anwendungsdomänen Ergebnisse (Bildererkennung, Spracherkennung, automatische Übersetzung von Texten). Die erheblichen Fortschritte haben in einer Vielzahl von KI-verwandten Forschungsfeldern großen Enthusiasmus ausgelöst. Viele Wissenschaftler glauben, dass *Deep Learning* die Forschung zur künstlichen Intelligenz massiv beschleunigen wird. Einige meinen sogar, dass *Deep Learning* die einzig notwendige Schlüsseltechnologie zum Verständnis künstlicher Intelligenz sei.

Diese Fortschritte wurden maßgeblich durch die Verfügbarkeit großer Rechenkapazitäten und wachsender Datenmengen ermöglicht. Die neuronalen Netze, auf denen *Deep Learning* basiert, wurden bereits Mitte des vergangenen Jahrhunderts entwickelt. Doch erst die heute existierenden Rechenkapazitäten lassen es zu, sehr viel größere neuronale Netzwerke zu verwenden. Sie bestehen aus vielen Tausenden von Knoten, die vielschichtig angeordnet sind (*deep* = tief, in vielen Schichten). Tiefe neuronale Netze werden unter Zuhilfenahme von sehr großen Datenmengen trainiert. Man trainiert sie, indem sich die Verbindungen zwischen den Knoten, entlang derer Zwischenergebnisse der Berechnung weitergeleitet werden, verstärken oder abschwächen. Die Vielzahl der Verbindungen in tiefen neuronalen Netzen macht die große Datenmenge erforderlich.

Trotz des Enthusiasmus, den *Deep Learning* als vermeintliche Schlüsseltechnologie für künstliche Intelligenz ausgelöst hat, ist Vorsicht geboten. Die Erfolge des *Deep Learning* wurden bisher in Problemdomänen mit bestimmten Eigenschaften erzielt. Künstliche Intelligenz umfasst aber sehr viel mehr als diese Problemdomänen. So ist es ungleich komplizierter, mithilfe des *Deep Learning* Probleme zu lösen, die eine direkte Interaktion mit der Umwelt erfordern oder in denen sich die Umwelt mit der Zeit verändert. Zudem offenbaren sich beim Vergleich von *Deep Learning* und natürlicher Intelligenz große Diskrepanzen. So benötigt natürliche Intelligenz, im Gegensatz zu *Deep Learning*, nur sehr kleine Datenmengen; gleichzeitig spielt bei ihr der Körper eine zentrale Rolle. Beides steht im Gegensatz zum *Deep Learning*.

Den *Deep Learning*-Enthusiasten steht eine Gruppe von Wissenschaftlern gegenüber, die anerkennen, dass *Deep Learning* ein mächtiges Werkzeug bei der Erforschung von künstlicher Intelligenz ist, die allerdings bezweifeln, dass es einen grundlegenden Durchbruch darstellt. Wer recht hat, wird die Zukunft zeigen.

Soft Robotics

Soft Robotics stellt den Körper in den Mittelpunkt. Dem Körper wird die Fähigkeit zu intelligentem Verhalten direkt zugeschrieben. Das mag zunächst seltsam klingen, doch viele Beispiele aus der Biologie belegen, dass komplexe Fähigkeiten allein in der Morphologie des Körpers, also seiner Form und seiner materiellen Beschaffenheit, begründet sind. Und wir wissen, dass die Körper aller Spezies an ihren Lebensraum angepasst sind.

Soft Robotics ist eine fundamentale Abkehr von einem etablierten Prinzip der Robotik. Es war lange Zeit unangefochten, dass ein guter „Körper“ (Roboter) lediglich dazu dient, die berechneten Bewegungen so präzise wie möglich auszuführen. Sein Verhalten ist berechnet und vorbestimmt, es resultiert nicht aus der Morphologie (also Material und Form). Die *Soft Robotics* stellt dieses Konzept auf den Kopf: In *Soft Robotics* ist das Verhalten eines „Körpers“ dann am kompetentesten, wenn Berechnung und Morphologie sich so bedingen, dass jeder den Teil des Problems übernimmt, den er am besten bewältigen kann.

Nehmen wir ein einfaches Beispiel: Ein Roboter greift einen Apfel von einem Teller. In der klassischen „harten“ Robotik muss der Apfel so präzise wahrgenommen werden, dass eine genaue Greiferstellung ermittelt werden kann, mit der der Apfel nicht beschädigt wird. Verwendet man einen weichen Greifer, so muss er nur ungefähr in die richtige Stellung gebracht werden. Aufgrund seiner materiellen Nachgiebigkeit kann er sich dann beim Greifen sanft um den Apfel schmiegen, ohne ihn zu beschädigen, ähnlich einer menschlichen Hand. Die Verwendung eines weichen Greifers vereinfacht die Wahrnehmung und die Bewegungsausführung grundlegend. Das ist das Prinzip von *Soft Robotics*: der Körper wird Teil der Intelligenz.

Die *Soft Robotics* ist ein fundamentaler Fortschritt und aus der Robotik nicht mehr wegzudenken. In vielen Anwendungsdomänen besitzen weiche Roboter Fähigkeiten, die die von harten Robotern weit übertreffen. *Soft Robotics* gilt als eine Schlüsseltechnologie für die Erforschung der künstlichen Intelligenz. Das ist übrigens weit weniger umstritten als die Rolle des *Deep Learning*. Doch auch *Soft Robotics* sind nur ein Teil der Lösung.

Industrie 4.0

Industrie 4.0 zielt darauf ab, Fertigungsprozesse flexibler, zeit- und kosteneffizienter zu gestalten. Die technische Notwendigkeit liegt in der Tatsache begründet, dass die heutige Automatisierungstechnik nur deshalb so gut funktioniert, weil der Aufwand für die Spezifikation und die Installation der Fertigungsstraßen beträchtlich ist. Die Kosten für die Erstellung der Fertigungsstraßen liegen daher auch um mindestens eine Größenordnung höher als die für die Roboter selbst. Daran wird sich nichts ändern, solange Industrieroboter über sehr beschränkte Wahrnehmungs-, Bewegungs- und Manipulationsfähigkeiten verfügen (wie von Moravecs Paradox vorhergesagt).

Wenn Industrie 4.0 ihre Versprechen erfüllen soll, müssen die Beschränkungen heutiger Automatisierungstechniken aufgehoben oder zumindest abgeschwächt werden. Die Systeme müssen mit flexiblen Wahrnehmungs- und Handlungsfähigkeiten ausgestattet werden. Zugleich wird die Entwicklung von Industrie 4.0 zu Fortschritten in der künstlichen Intelligenz führen. Von diesen Synergien werden beide Seiten profitieren.

6. Anwendungsfelder künstlicher Intelligenz

Sprechen wir über Anwendungsfelder, in denen die künstliche Intelligenz in naher Zukunft von Bedeutung sein könnte.

Dort, wo bereits heute große Datenmengen vorliegen, die auch für Menschen interpretierbar sind, werden die *Silicon Intelligence* und das

maschinelle Lernen große Veränderungen mit sich bringen. Denn in diesen Anwendungsfeldern übertreffen Computer bereits heute die Fähigkeiten von Menschen. Neben den bereits erwähnten Bereichen (Bilderkennung, Spracherkennung und Übersetzung), sind Anwendungen im Rechtswesen oder die medizinische Bildanalyse zu nennen. Schon bald werden Computer routinemäßig Tumore mittels bildgebender Verfahren verlässlicher identifizieren als es Menschen möglich ist. Gleiches gilt für komplexe juristische Verträge, die von Computern schlüssiger und schneller erstellt werden können als von Anwälten.

Beispiele für weitere Anwendungsfelder sind industrielle Fertigung und Logistik. In der Fertigung ist es bereits üblich, Fertigungsstraßen auf die Fähigkeiten von Robotern hin auszurichten. Es besteht ein wirtschaftlich sinnvoller, inkrementeller Pfad, um zielgerichtete Innovation aus der künstlichen Intelligenz der Anwendung zuzuführen. Gleiches gilt für die Logistik. Mit zunehmendem Internethandel wird die Verteilung von Waren in großen Logistikzentren zentralisiert. Diese können den steigenden Bedarf aber nur dann decken, wenn die heute noch zum großen Teil von Menschen durchgeführten Arbeiten (insbesondere das sogenannte Picking, also das Zusammenstellen einer individuellen Bestellung aus den Lagerbeständen) automatisiert werden. Diese Automatisierung kann, wie in der Fertigung, graduell erfolgen.

Die Fertigung und die Logistik sind also Bereiche, in denen frühe Anwendungen künstlicher Intelligenz wahrscheinlich sind. Darüber hinausgehende Prognosen erscheinen schwierig. Oft genannte Beispiele für mögliche Anwendungen künstlicher Intelligenz sind Haushaltshilfen, der Einsatz von Robotern zur Unterstützung alter oder kranker Menschen, bei der Bewältigung von Naturkatastrophen oder anderen Notfällen (z. B. Fukushima), bei der Erkundung der Meere oder anderer Planeten und das automatisierte Erschließen von Bodenschätzen. Noch fehlen viele der Technologien oder sind nicht ausgereift genug, um solche Anwendungen auch tatsächlich zu ermöglichen.

Zudem gilt es noch, die rechtlichen Voraussetzungen für den Einsatz künstlicher Intelligenzen in Alltagsumgebungen zu schaffen. Hier

stehen wichtige ethische Fragen von weitreichender gesellschaftlicher Bedeutung an, die einen breiten Dialog aller Interessensgruppen erfordern.

7. Eine Strategie für künstliche Intelligenz

Künstliche Intelligenz und auch *Silicon Intelligence* werden mit großer Wahrscheinlichkeit disruptiver sein, als alle bisher bekannten Technologien. Wer bei der Entwicklung künstlicher Intelligenz führend ist, verschafft sich nicht nur massive wirtschaftliche Vorteile, sondern wird vermutlich auch die globalen Machtverhältnisse bestimmen. Folglich investieren Industrieländer massiv in die Erforschung der künstlichen Intelligenz. Als Beispiel sei hier nur Frankreich genannt, dass innerhalb von fünf Jahren 1,5 Milliarden Euro bereitstellen will, wie Präsident Macron im März 2018 angekündigt hat.

Wer das Rennen gewinnen wird, ist noch offen. Die großen Internetfirmen in den USA (Amazon, Apple, Google, Facebook, Microsoft) setzen vorrangig auf *Silicon Intelligence*. Sicher auch weil sich hier Ergebnisse direkt in Anwendungen niederschlagen. Dabei bleiben die großen Möglichkeiten der Erforschung „wahrer“ künstlicher Intelligenz weitestgehend unbeachtet – eine große Gelegenheit für andere Player hier die Führung zu übernehmen.

Die fehlende Unterscheidung von *Silicon Intelligence* und künstlicher Intelligenz im öffentlichen und zum Teil auch im wissenschaftlichen Diskurs hat dazu geführt, dass sich die Aufmerksamkeit vornehmlich auf die *Silicon Intelligence* richtet. Eine übergreifende Strategie für die Erforschung und den Einsatz der künstlichen Intelligenz wäre ein erheblicher Wettbewerbsvorteil.

Wie sollte eine solche Strategie aussehen? Die Geschichte der künstlichen Intelligenz hat noch keinen klar dominierenden wissenschaftlichen Ansatz zur Erforschung von Intelligenz geliefert. Derzeit setzen viele Wissenschaftler und Unternehmen in der Welt auf *Deep Learning*. Doch bei genauer Betrachtung gibt es wenig überzeugende Argu-

mente, warum *Deep Learning* die heilsbringende Schlüsseltechnologie sein sollte. Der Enthusiasmus beruht auf wissenschaftlichen Ergebnissen, die zweifelsohne wichtige Errungenschaften darstellen, aber nicht beantworten können, welche Grenzen *Deep Learning* gesetzt sind. Auch andere Konzepte in der künstlichen Intelligenz haben in der Vergangenheit sehr gute Ergebnisse geliefert und sich dann doch nicht als Kandidat für einen beherrschenden Lösungsansatz erwiesen.

Eine zielführende Strategie für künstliche Intelligenz sollte mehrere Ansätze parallel verfolgen. Nur so kann sichergestellt werden, dass Deutschland an wegweisenden Entwicklungen beteiligt ist. Eine solche Strategie sollte auch – neben der Grundlagenforschung – von Beginn an relevante Industrieunternehmen involvieren, da diese sinnvolle „ökologische Nischen“ definieren, innerhalb derer künstliche Intelligenz untersucht werden kann. Zudem sollte eine solche Strategie klar zwischen künstlicher Intelligenz und *Silicon Intelligence* unterscheiden. Nur so wird die mittel- und langfristige Erforschung der künstlichen Intelligenz nicht von kurzfristigen Entwicklungen und Hypes bestimmt. Schließlich sollte eine solche Strategie inhärent multidisziplinär sein und die analytischen Wissenschaften der Intelligenzforschung (Psychologie, Verhaltensbiologie, Neurowissenschaften usw.) mit den synthetischen (Robotik, maschinelles Lernen, KI, Computer Vision usw.) eng verknüpfen. Das ist der plausibelste Weg, dieser wohl größten wissenschaftlichen Herausforderung der Menschheitsgeschichte fundamentale Erkenntnisse abzurufen.

Robotik und ihr Beitrag zu Wachstum und Wohlstand

Jens Südekum

- › Der vorliegende Beitrag untersucht den Einfluss von Industrierobotern auf den deutschen Arbeitsmarkt im Zeitraum von 1994 bis 2014.
- › Die These, nach der Roboter zu Massenarbeitslosigkeit führen, kann empirisch nicht bestätigt werden. Roboter beschleunigen aber den Strukturwandel und können zur Zunahme von Einkommensungleichheit führen.
- › Der Einsatz von Robotern steigert die Produktivität und die Unternehmensgewinne, führt aber nicht zu einem Anstieg der Durchschnittslöhne. Hochqualifizierte Beschäftigte profitieren von dieser neuen Technologie, die breite Mehrheit der Facharbeiter in der Mitte des Qualifikations- und Lohnspektrums bisher allerdings nicht.
- › Die gemessenen Verteilungseffekte waren bislang noch überschaubar, könnten sich aber künftig verstärken. Die Gesellschaft sollte daher jetzt in die Diskussion über diese Verteilungsprobleme einsteigen, die sich aus Digitalisierung und Automatisierung ergeben könnten.

Einleitung

Die Angst vor einer bevorstehenden Welle „technologischer Arbeitslosigkeit“ ist eines der beherrschenden ökonomischen Themen unserer Zeit. Einer weit verbreiteten These zufolge werden Produktionsprozesse, insbesondere im verarbeitenden Gewerbe, mit der Digitalisierung und dem Voranschreiten künstlicher Intelligenz (KI) immer stärker automatisiert. In vielen Bereichen würde die menschliche Arbeitskraft durch intelligente Maschinen – zum Beispiel Roboter – ersetzt. Menschen seien dann in der Produktion quasi „überflüssig“. Es mangelt nicht an dramatischen Vorhersagen, wie viele Arbeitsplätze konkret bedroht sind. So gehen einige Schätzungen davon aus, dass nahezu die Hälfte aller Arbeitsplätze gefährdet sei, durch neue Technologie ersetzt zu werden.

Solche dramatischen Schreckensszenarien sind nicht neu. Es gab sie in der Wirtschaftsgeschichte schon häufig. Geprägt wurde der Begriff der „technologischen Arbeitslosigkeit“ von dem Jahrhundertökonom John Maynard Keynes in den 1930er Jahren. Im Jahr 1983 zog der Wirtschaftsnobelpreisträger Wassily Leontief in einer berühmten Rede eine Analogie zum Schicksal von Pferden. Deren Arbeitskraft in der Landwirtschaft wurde durch die Entwicklung von Traktoren und Dampfmaschinen weitestgehend entbehrlich, und sie standen fortan unproduktiv auf der Weide herum. Ein vergleichbares Schicksal prophezeite er der menschlichen Arbeit durch den voranschreitenden technischen Fortschritt. Abbildung 1 zeigt Titelseiten des *SPIEGEL*-Magazins aus den Jahren 1964, 1978 und 2016. Auch sie visualisieren die Furcht vor einer, durch den Einsatz von Robotern verursachten technologischen Arbeitslosigkeit – über einen Zeitraum von mehr als einem halben Jahrhundert.

Abbildung 1:

SPIEGEL-Titelseiten zur technologischen Arbeitslosigkeit



1964



1978



2016

Quelle: Der SPIEGEL, Heft 14/1964, Heft 16/1978, Heft 36/2016

Eingetreten sind diese düsteren Prophezeiungen indes noch nicht. Auf dem deutschen Arbeitsmarkt hat das Beschäftigungsniveau einen Rekordwert seit der Wiedervereinigung erreicht.

1. Technologisches Substitutionspotential versus tatsächliche Beschäftigungsentwicklung

Wie kommen die Meldungen von den Millionen bedrohter Arbeitsplätze zustande? Sie beruhen auf wissenschaftlichen Studien, die das technologische Substitutionspotential von Berufen anhand von gemessenen Tätigkeitsprofilen abschätzen. Diese Studien sagen etwas darüber aus, in welchem Ausmaß Menschen durch Maschinen, ausgehend vom aktuellen Stand der Technik, ersetzt werden *könnten*. Sie sagen aber nichts über die tatsächlichen Arbeitsmarkteffekte von technologischem Fortschritt aus, also ob Berufe und Arbeitsplätze wirklich entfallen. Diese Unterscheidung wurde von den Medien nicht immer vorgenommen, was zu reißerischen, aber substanzlosen Schlagzeilen geführt hat.

Die bekannteste dieser Studien ist die von Frey und Osborne (2017). Sie basiert auf Daten zu detaillierten Tätigkeitsprofilen von 702 Berufen.

Anhand von Experteneinschätzungen werden siebzig dieser Berufe hinsichtlich ihres Automatisierungspotentials nach aktuellem Stand der Technik klassifiziert. 37 Berufe (z. B. Versicherungssachbearbeiter) wurden als „vollständig automatisierbar“ und 33 Berufe (z. B. Kindergärtner) als „überhaupt nicht automatisierbar“ eingestuft. Aufbauend auf diesen Expertenmeinungen werden anhand der zugrundeliegenden Tätigkeitsprofile die Automatisierungspotentiale aller Berufe berechnet. Als „akut bedroht“ stufen Frey und Osborne Berufe mit einem Automatisierungspotential von mehr als siebzig Prozent ein, wobei diese Schwelle willkürlich gesetzt ist. Darunter fallen 47 Prozent aller Berufe, auch der des Kochs mit einem Automatisierungspotential von immerhin 96 Prozent. Die Autoren weisen zwar am Ende ihrer Studie darauf hin, dass dieses Potential nicht als Wahrscheinlichkeit interpretiert werden dürfte, mit der Berufe tatsächlich ersetzt würden. Ihre Warnung wurde jedoch überlesen und nicht zur Kenntnis genommen.

Die Entscheidung, ob Tätigkeiten von Menschen oder Maschinen ausgeübt werden, ist keine *technologische*, sondern eine *ökonomische*. Mit anderen Worten: Nicht jeder Beruf, der automatisiert werden kann, wird auch tatsächlich automatisiert. Eine technologisch bedingte Massenarbeitslosigkeit ist auch im Zeitalter von Digitalisierung, Automatisierung und künstlicher Intelligenz unwahrscheinlich. Dagegen spricht eine Reihe einfacher ökonomischer Prinzipien:

1. Maschinen werden in etablierten Berufen einzelne Tätigkeiten ersetzen. Dafür können sich Menschen innerhalb ihrer Berufe aber auf andere Tätigkeiten konzentrieren und diese perfektionieren. Nehmen wir das bereits erwähnte Beispiel des Kochs. Der Roboter könnte Routinetätigkeiten wie Kartoffelschälen oder Gemüseputzen übernehmen. Der Koch verfügt nun über Zeit, um sich kreative neue Gerichte auszudenken oder Gewürzmischungen zu perfektionieren. So banal dieses Beispiel anmutet, so präsent ist das zugrundeliegende Grundmotiv in vielen anderen Fällen. Maschinen ersetzen keine gesamten Berufe, sondern verändern Tätigkeitsprofile.

2. Dass es tatsächlich nur einen äußerst schwachen statistischen Zusammenhang zwischen dem Automatisierungspotential und dem tatsächlichen Beschäftigungswachstum von Berufen gibt, haben jüngst die IAB-Forscherinnen Dengler und Matthes (2018) in einer Studie aufgezeigt. Sie verwenden eine leicht modifizierte Fassung der Methode von Frey und Osborne (2017) und zeigen auch für Deutschland einen hohen „Bedrohungsgrad“ für einzelne Berufe, der aufgrund der technologischen Entwicklung seit 2013 nochmals merklich gestiegen ist. Die Autorinnen finden aber keine Bestätigung für die These, dass sich stark bedrohte Berufe (z. B. Koch) hinsichtlich der Beschäftigung zwischen 2013 und 2016 schwächer entwickelt hätten. Aus einem steigenden technologischen Substitutionspotential lässt sich also kein Rückschluss auf die allgemeine Beschäftigungsentwicklung eines Berufes ziehen.
3. Selbst wenn einzelne Berufe entfallen sollten, entstehen durch den Einsatz von Maschinen kontinuierlich neue Tätigkeitsfelder. So gibt es laut der aktuellen Klassifikation des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) diverse Berufe in Deutschland, die vor dreißig Jahren noch gänzlich unbekannt waren: etwa der des Interfacedesigners, noch 2013 kannte ihn niemand. Neu geschaffene Berufe weisen typischerweise ein höheres Beschäftigungswachstum auf und absorbieren freigesetzte Beschäftigung aus den schrumpfenden alten Professionen.
4. Die neuesten Technologien werden in hoch produktiven Spitzenfirmen eingesetzt, aber diffundieren nur langsam in der Breite. Auch dies wirkt massiven Beschäftigungsverlusten entgegen. Haldane (2017) zeigt auf, dass es auf der Ebene einzelner Unternehmen einen deutlich positiven Zusammenhang zwischen Produktivitätsniveau und -wachstum gibt. Bereits sehr produktive und profitable Unternehmen sind seit dem Jahr 2000 noch einmal überdurchschnittlich gewachsen und enteilen damit der Mehrzahl der „normalen“ Wettbewerber. Die neuesten Technologien werden vor allem in diesen Spitzenfirmen eingesetzt, zeitigen dort aber keine negativen Beschäftigungs-

effekte. Das Beispiel Amazon mag dies verdeutlichen. Dort wurde von 2014 bis 2017 der Bestand an Robotern im Logistik- und Warenlagerbereich von 1.400 auf 45.000 gesteigert. Im gleichen Zeitraum wurde die Zahl der Mitarbeiter verdreifacht, weil die Produktivitätsfortschritte zu sinkenden Preisen für die Verbraucher und damit zu einer entsprechend gestiegenen Nachfrage geführt haben. Dies zog eine Nachfrage nach Arbeitskräften nach sich, die seither Hand in Hand mit der Maschine in neuen Tätigkeitsfeldern arbeiten.

5. Wenn der technologische Fortschritt aus gesamtwirtschaftlicher Perspektive tatsächlich negative Auswirkungen auf die Nachfrage nach Arbeit hat, was, wie eben ausgeführt, durchaus umstritten ist, wird die demografische Entwicklung zur wirkmächtigen Gegenkraft. Fast genauso präsent wie Vorhersagen einer kommenden Massenarbeitslosigkeit sind diametral gegenläufige Prognosen eines massiven Mangels an Arbeitskräften: insbesondere an Fachkräften, der sich in Deutschland einstellen werde. Auch diese Voraussagen mögen bisweilen überzogen sein. Doch die Folgen des demografischen Wandels sind heute – im Gegensatz zu den Technologiefolgen – bereits viel präziser absehbar. Eine schrumpfende Erwerbsbevölkerung stellt einen negativen Angebotsschock auf dem Arbeitsmarkt dar, der sich exakt gegensätzlich zu den behaupteten Effekten der Technologie auswirken würde.
6. Und selbst wenn sich der Arbeitsnachfrageschock durch den Einsatz neuer Technologien in der Summe durchsetzen sollte, hat das in simpler ökonomischer Logik immer noch keine Massenarbeitslosigkeit zur Folge. Denn wenn die neuen Technologien tatsächlich Millionen von Menschen zunächst arbeitslos machten, dann würde das allgemeine Lohnniveau auf dem Arbeitsmarkt sinken. Ein gesunkenes Lohnniveau macht den Einsatz von Menschen gegenüber Maschinen für Firmen aber wieder attraktiver und konterkariert somit die Arbeitsplatzverluste. In der Summe hätten wir es dann zwar mit negativen Lohneffekten, aber nicht mit Massenarbeitslosigkeit zu tun.

Zusammenfassend lässt sich festhalten, dass gesamtwirtschaftliche Vorhersagen zu Arbeitsmarkteffekten von neuen Technologien, die ausschließlich auf dem technologischen Substitutionspotential von Berufen basieren, deutlich zu kurz greifen. Diese Studien ignorieren ökonomische Anpassungskanäle, was zu einer stark verzerrten Berichterstattung und Wahrnehmung in der Öffentlichkeit geführt hat.

2. Die Arbeitsmarkteffekte von Industrierobotern in Deutschland

Ein gänzlich anderer Forschungsansatz besteht darin, mit modernen statistisch-ökonomischen Methoden zu analysieren, welchen Einfluss neue Technologien in der Vergangenheit tatsächlich auf den Arbeitsmarkt gehabt haben, um aus diesen Mustern Projektionen für die Zukunft abzuleiten. In einer jüngeren Studie (Dauth et al. 2017) haben meine Co-Autoren und ich den Einfluss von Industrierobotern auf Löhne und Beschäftigung auf dem deutschen Arbeitsmarkt zwischen 1994 und 2014 untersucht. Industrieroboter sind nur ein Ausschnitt der Digitalisierung und Automatisierung. Aber sie sind empirisch greifbar, und unsere Forschung liefert daher konkrete Erkenntnisse auch im Hinblick auf die Effekte anderer Technologieformen.

In unserer Analyse greifen wir auf einen Datensatz der International Federation of Robotics (IFR) zurück, der auf Befragungen führender Roboterhersteller basiert und mehr als neunzig Prozent des Weltmarktes abdeckt (vgl. Graetz und Michaels, 2014). Die Daten zeigen, dass Industrieroboter in Deutschland viel stärker verbreitet sind als in den USA oder in anderen industrialisierten Ländern. Im Jahr 1994 waren in Deutschland rund zwei Industrieroboter pro tausend Beschäftigte installiert. Das waren fast doppelt so viele wie im europäischen Durchschnitt und rund viermal so viele wie in den USA. Die Installationszahlen haben sich bis 2014 beinahe vervierfacht. Nur in Japan, Singapur und Südkorea gibt es noch mehr Roboter pro Beschäftigten als bei uns. Außerdem ist Deutschland nicht nur in der Nutzung, sondern auch in der Entwicklung und Produktion von Industrierobotern weltweit führend. So haben laut des *Robotics World*

Rankings fünf der zwanzig größten Hersteller weltweit (Haupt-)standorte in Deutschland.

Trotz der enormen Verbreitung von Industrierobotern hat Deutschland – eine der bedeutendsten Industrienationen der Welt – einen weiterhin außergewöhnlich hohen Beschäftigungsanteil im verarbeitenden Gewerbe. Er liegt derzeit bei rund 25 Prozent, im Vergleich zu nur neun Prozent in den USA. Auch hierzulande ist der Anteil in den letzten zwanzig Jahren gesunken. Aber dieser Rückgang war weniger dramatisch als anderswo. Und das, obwohl es bei uns noch viel mehr Industriearbeitsplätze und deutlich mehr Roboter gibt.

In einem ersten Schritt untersuchen wir den Gesamteffekt von Robotern auf das Beschäftigungswachstum. Im Gegensatz zu einer vergleichbaren Studie für die USA (siehe Acemoglu und Restrepo 2017) finden wir keine empirische Evidenz für die These, dass Roboter in Deutschland die Gesamtzahl der Arbeitsplätze reduziert haben. Stark negative Effekte zeigen sich indes auf die Beschäftigung im verarbeitenden Gewerbe. Unsere Berechnungen ergeben, dass ein zusätzlicher Roboter durchschnittlich zwei Jobs in der Industrie substituiert. Im Zeitraum 1994 bis 2014 wurden etwa 131.000 Roboter in Deutschland installiert. Statistisch gesehen führte dies zu einem Rückgang von rund 275.000 Vollzeitjobs in der Industrie. Dies entspricht etwa 23 Prozent des Rückgangs des Industrieanteils in der Gesamtbeschäftigung. Diese beträchtlichen Verluste wurden jedoch durch Arbeitsplatzgewinne außerhalb des verarbeitenden Gewerbes, vor allem bei wirtschaftsnahen Dienstleistungen, vollständig ausgeglichen. Mit anderen Worten: Roboter haben die *Struktur* der Beschäftigung in Deutschland verändert. Aber die Gesamtzahl an Arbeitsplätzen wurde durch den Einsatz von Robotern bislang nicht beeinflusst.

3. Der Effekt von Robotern auf einzelne Arbeiter

Diese aggregierten empirischen Befunde werfen die Frage auf, durch welche Kanäle sich der Einsatz von Robotern auf einzelne Arbeiter ausgewirkt hat. Um diese bisher unerforschte Fragestellung zu analysieren,

verwenden wir detaillierte Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB), die es uns erlauben, individuelle Erwerbsbiografien von etwa einer Million Beschäftigten im verarbeitenden Gewerbe genau zu rekonstruieren. Dieser Forschungsansatz führt zu einem überraschenden Ergebnis: Roboter haben die Beschäftigungssicherheit existierender Industriearbeitsplätze bislang nicht beeinträchtigt. Im Gegenteil, Arbeitnehmer aus roboterintensiveren Wirtschaftszweigen haben sogar eine höhere Wahrscheinlichkeit, ihren ursprünglichen Arbeitsplatz zu behalten. Zwar verändern sich ihre Tätigkeiten im Betrieb. So wurden Produktionsarbeiter teilweise umgeschult und im Betrieb neu positioniert, etwa in den Bereichen Wartung oder Verkauf. Aber Roboter führen keineswegs zu Entlassungen, die Arbeitsplätze wurden gehalten.

Der negative Gesamteffekt von Robotern auf die Beschäftigung im verarbeitenden Gewerbe kommt vielmehr dadurch zustande, dass geringere Arbeitsmarkteintritte von jungen Berufsanfängern zu verzeichnen waren. Anders ausgedrückt, zerstören Roboter keine bestehenden Jobs in der Industrie, sondern sie veranlassen Unternehmen dazu, weniger neue Arbeitsplätze für junge Menschen zu schaffen. Die Folge dieses Strukturwandels ist eine zunehmende Alterung der Beschäftigten in stärker roboterisierten Industriezweigen.

4. Der Einfluss von Robotern auf Löhne

Lohn- und Einkommenseffekte durch den Einsatz von Robotern unterscheiden sich auf individueller Ebene sehr stark nach dem Berufs- und Qualifikationsprofil. Der steigende Robotereinsatz führt zu beträchtlichen Einkommengewinnen bei hochqualifizierten Beschäftigten mit Universitätsabschluss. Dies gilt vor allem in Forschungs- und Managementpositionen: Beschäftigte verfügen über komplementäre Fähigkeiten und ihre Aufgaben sind nicht durch Roboter substituierbar. Bei Geringqualifizierten stellen wir keine statistisch messbaren Lohneffekte fest.

Aber bei Arbeitnehmern mit mittlerem Qualifikationsprofil lassen sich moderat negative Auswirkungen von Robotern auf die Erwerbsein-

kommen diagnostizieren. Eine abgeschlossene Berufsausbildung ist das typische Beschäftigungsprofil in der deutschen Industrie. Diese Gruppe mit mittlerer Qualifikation macht nahezu 75 Prozent aller Beschäftigten aus. Sie sind stark überrepräsentiert in Berufen in der Produktion mit einem hohen Grad an manuellen Routinetätigkeiten. Diese Facharbeiter sind somit am stärksten von Automatisierung bedroht. Interessanterweise finden wir selbst für diese Gruppe keine negativen Beschäftigungseffekte. Aber die Betroffenen mussten Lohn-einbußen hinnehmen.

Eine derartige Anpassungsreaktion ist durchaus typisch für den deutschen Arbeitsmarkt. Das verarbeitende Gewerbe ist stark gewerkschaftlich organisiert, und Industrielöhne werden unter starker Beteiligung von Betriebsräten determiniert. Schon häufig wurde argumentiert, dass die deutschen Gewerkschaften ihren Fokus auf ein hohes Beschäftigungsniveau legten und kooperativer seien als Gewerkschaften in anderen Ländern. Diese Flexibilität der Gewerkschaften, etwa mit flexiblen Lohnsetzungs- und Arbeitszeitregelungen oder Öffnungsklauseln auf Herausforderungen zu reagieren, gilt seit Mitte der 2000er Jahre als einer der zentralen Gründe für die gute Gesamtentwicklung des deutschen Arbeitsmarktes, das so genannte Beschäftigungswunder (siehe Dustmann et al., 2014).

Unsere Analyse deutet darauf hin, dass der Aufstieg der Roboter zu einer ähnlichen Reaktion geführt haben könnte: die Bereitschaft zu Lohnverzicht, um die Beschäftigung von Facharbeitern – selbst in Anbetracht akuter Arbeitsplatzbedrohung durch Roboter – zu stabilisieren. Im mittleren Bereich der Lohnverteilung haben Roboter so zu Einbußen geführt, deren Größenordnung bislang aber noch überschaubar gewesen ist.

Zudem konnten wir feststellen, dass Roboter zwar die durchschnittliche Arbeitsproduktivität sowie Produktion und Gewinne erhöht, aber nicht zu einem Anstieg der Durchschnittslöhne geführt haben. Roboter tragen also zum tendenziellen Rückgang der Lohnquote am Gesamteinkommen bei, der in vielen industrialisierten Ländern in den letzten Jahren zu verzeichnen war (Kehring und Vincent 2017). Die vornehmlichen Erträge aus dieser neuen Technologie fallen also nicht

beim Faktor Arbeit an. Unter den Beschäftigten profitieren zwar einige am oberen Rand der Lohnverteilung. Aber für das Gros der Industriebeschäftigten haben die Roboter zu negativen Lohn- und Einkommenseffekten geführt und die Lohnungleichheit gesteigert.

5. Politikimplikationen

Die neuen Technologien sollten Gesellschaft und Politik nicht in Panik versetzen. Wellen der Massenarbeitslosigkeit stehen nicht bevor. Prognosen basierend auf technologischen Substitutionspotentialen vernachlässigen ökonomische Anpassungskanäle und sind daher systematisch verzerrt. Spekulationen über Arbeitsplatzverluste sind wertlos.

Wissenschaftliche Studien zum tatsächlichen Arbeitsmarkteinfluss neuer Technologien zeichnen ein weit weniger dramatisches Bild. Damit soll aber keineswegs Entwarnung gegeben werden. Auch unsere Forschung zeigt, dass Roboter – als eine konkrete Form des technologischen Fortschritts – auf dem deutschen Arbeitsmarkt zu einem Anstieg der Einkommens- und Lohnungleichheit beigetragen haben. Bislang waren diese Verteilungseffekte quantitativ noch moderat. Aber sie können sich in Zukunft deutlich verstärken, etwa durch die weitere Entwicklung der künstlichen Intelligenz, die sich mit raptider Geschwindigkeit entwickelt. Strukturell ist nicht zu erwarten, dass KI andere Arbeitsmarkteffekte haben wird als die Verbreitung von Industrierobotern, denn die bereits ausgeführten Grundüberlegungen zu ökonomischen Anpassungen infolge des Einsatzes neuer Technologien gelten auch in diesem Fall. Aber die von KI induzierte Lohnungleichheit könnte sich erheblich verstärken. Die Gesellschaft sollte daher über Antworten und Strategien diskutieren, bevor sich die Effekte spürbar verstärken.

Mögliche Antworten reichen von mehr klassischer Einkommensumverteilung über so genannte Roboter-Steuern bis hin zu Modellen des bedingungslosen Grundeinkommens oder der Mitarbeiterbeteiligung. All diese Modelle haben ihre spezifischen Vor- und Nachteile, die an dieser Stelle aber nicht diskutiert werden können.

Referenzen

- A Acemoglu, D. and Restrepo, P.** (2017). The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment, NBER Working Paper No. 22252.
- D Dauth, W., Findeisen, S., Suedekum, J. und Woessner, N.** (2017). German Robots – the Impact of Industrial Robots on Workers. CEPR Discussion Papers 12306, London.
- Dengler, K. und Matthes, B.** (2018). Wenige Berufsbilder halten mit der Digitalisierung Schritt. IAB-Kurzbericht 04/2018, Nürnberg.
- Dustmann, C., Fitzenberger, B., Schoenberg, U. and Spitz-Oener, A.** (2014). From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy, *Journal of Economic Perspectives* 28(1): S. 167–188.
- F Frey, C. B. and Osborne, M. A.** (2017). The Future of Employment: How Susceptible are Jobs to Computerisation?, *Technological Forecasting and Social Change* 114: S. 254–280.
- H Haldane, A.** (2017). Productivity Puzzles. Vortrag an der London School of Economics, 20. März 2017. Verfügbar unter: <https://www.bankofengland.co.uk/speech/2017/productivity-puzzles> (zuletzt abgerufen: 16. Mai 2018)
- G Graetz, G. and Michaels, G.** (2017). *Robots at Work*, *CEP Discussion Paper 1335, Revised version (June 22, 2017)*.
- K Kehrig, M. and Vincent, N.** (2017). Growing productivity without growing wages: The micro-level anatomy of the aggregate labor share decline, *CESifo Working Paper Series No. 6454*.

Deutschland als unübertrefflicher KI-Standort

Jürgen Schmidhuber

- › Die europäische KI-Grundlagenforschung ist weltweit einmalig. Die meisten grundlegenden KI-Algorithmen stammen von Europäern. Die kommerzielle Umsetzung gelingt jedoch in den USA und in China besser.
- › Durch seine Basiserfindungen für die moderne KI in Verbindung mit weltweiter Führung in Maschinenbau und Robotik ist Deutschland im Wettbewerb der KI-Standorte im Prinzip sehr gut aufgestellt.
- › Um Exzellenz in der Forschung wirtschaftlich umzusetzen, bedarf es aber einer Industriepolitik nach dem Vorbild von USA und China, die innovierende Unternehmen – Start-ups, KMUs und große Firmen – massiv unterstützen. Die deutsche KI-Forschung benötigt viel umfangreichere, bessere, verlässlich planbare Finanzierung und neue Strukturen.

Der nachfolgende Text basiert auf einem Interview mit Jürgen Schmidhuber vom 12. September 2018. Das Interview führte Norbert Arnold.

In der Öffentlichkeit wird oft nicht eindeutig zwischen Digitalisierung und KI unterschieden. Ist das ein Problem?

Schmidhuber: Digitalisierung ist nur die Grundlage der KI – vergleichbar etwa mit dem Ackerbau als Grundlage der modernen Zivilisation. Leistungsfähige Computer und schnelle Netzwerke sind technische Voraussetzungen für KI.

Die Vorstellungen, was KI ist und welche Chancen und Risiken mit ihr verbunden sind, sind diffus. Was sollte jeder über KI wissen?

Schmidhuber: KI ist die Wissenschaft vom automatischen Problemlösen. Insbesondere davon, wie Maschinen lernen können, bisher ungelöste Probleme zu lösen.

KI ist eigentlich nicht kompliziert. Die Grundprinzipien lassen sich auch von Laien verstehen. Die grundlegenden Algorithmen sind simpel. Man kann sie in ein paar Zeilen Pseudocode hinschreiben.

Die lernenden neuronalen Netze Ihrer Teams am Schweizer KI-Labor IDSIA und der TU München stecken nun in drei Milliarden Smartphones, insbesondere das Long Short Term Memory (LSTM). Es wird jeden Tag milliardenfach genutzt, z. B. in Facebooks automatischer Übersetzung (2017), Googles Spracherkennung (seit 2015) und Google Translate (2016), Apples iPhone (2016), Amazons Alexa (2016) usw. Ist LSTM denn auch simpel?

Schmidhuber: Ja. Den Kern des LSTMs kann man in wenigen Zeilen verdeutlichen. Man braucht ein paar mathematische Grundkenntnisse, aber kein langes Studium, um es zu begreifen. Eigentlich sind ja alle

guten Ideen einfach – auch wissenschaftliche Theorien, die viele für sehr kompliziert halten, wie etwa Einsteins Allgemeine Relativitätstheorie, die man aber in einem Satz zusammenfassen kann.

Können auch Jugendliche das rasch lernen?

Schmidhuber: Klar. In wenigen Stunden kann ein gut informierter und motivierter Lehrer seinen Schützlingen erklären, was die Grundlagen eines künstlichen neuronalen Netzwerks sind, wie es aufgebaut ist, wie es lernt. Schülerinnen und Schüler können KI in ihren Grundzügen schnell erfassen. Sie nutzen sie ja schon in ihrem Alltag.

KI hat tatsächlich schon ihren Weg in den Alltag der Menschen gefunden. Bleibt eine große „KI-Revolution“ aus? Ist es eine „schleichende Revolution“, die wir derzeit miterleben?

Schmidhuber: Die grundlegenden Algorithmen, die heute das Leben von Milliarden Menschen verändern, stammen aus dem letzten Jahrtausend. Aber erst seit kurzem finden sie sich auf jedem Handy. Warum? Seit Konrad Zuse 1941 den ersten programmgesteuerten Computer baute, wird die Rechenleistung von Computern alle fünf Jahre zehnmal billiger. Diese Entwicklung verläuft also exponentiell und ist insofern keineswegs schleichend. Nur haben viele Menschen sie lange Zeit nicht bemerkt, weil es die meisten nicht persönlich betraf. Jetzt merken sie es.

Hat auch die Politik es zu spät bemerkt? Bereits vor 50 Jahren war KI ein Forschungsthema. Hat man es nicht ernst genug genommen? Die USA und China scheinen führend, andere Länder sind sehr aktiv und holen auf. Hat Deutschland den Anschluss verpasst?

Schmidhuber: Die USA und China sind in der KI-Forschung keineswegs führend. Sie investieren zwar am meisten Geld, das ist richtig. Aber die grundlegenden KI-Algorithmen stammen fast ausschließlich von Europäern.

Viele Patente aus den USA und China sind sogenannte Epsilon-Patente, die lediglich Verbesserungen bereits bestehender Techniken beinhalten. Die zentralen Durchbrüche beim *Deep Learning* und beim Maschinenlernen aber stammen bis auf wenige Ausnahmen vom alten Kontinent, ebenso wie selbstfahrende Autos, das WWW und der Computer selbst. Bei den grundlegenden KI-Entwicklungen zumindest muss sich Europa nicht verstecken.

Was die Umsetzung in Profite betrifft, ist der pazifische Raum allerdings führend. Politisch wird diese Entwicklung durch eine konsequente IT-Industriepolitik unterstützt, die es in Europa so nicht gibt. Amerikaner und Chinesen haben Milliarden an Steuergeldern aufgewendet, um populäre einheimische Firmen wie Amazon oder Tesla zu subventionieren. Auch über die Förderung militärischer Forschung – versehen mit dem Hinweis auf nationale Sicherheit – werden in den USA und China viele kleine Firmen hochskaliert. Die USA und China befehlen sich regelrecht mit industriepolitischen Maßnahmen und halten Firmen aus dem jeweils anderen Land gezielt vom eigenen Markt fern. Es geht dabei übrigens auch um wirtschaftlich und politisch interessante Nutzerdaten, die man nicht Unternehmen fremder Länder überlassen will.

Europa hingegen hat sich von den großen Westküstenfirmen überrennen lassen. Sie dominieren heute die mit sozialen Netzwerken und Suchmaschinen assoziierten Märkte. Und das ist auch nicht mehr rückgängig zu machen.

Deutschland wird es also nicht gelingen, diesen Rückstand bei Suchmaschinen und sozialen Netzen aufzuholen. Haben wir in der KI eine Chance?

Schmidhuber: Ja, auf jeden Fall. Heute werden zwar die großen KI-Profite in der passiven Mustererkennung durch Suchmaschinen und soziale Netze usw. erzielt. Werbung wird auf Basis des bisherigen Nutzerverhaltens verkauft: Welche Artikel wurden wie lange gelesen? Welche Produkte online gekauft? Aus diesem Verhalten lassen sich

leicht Präferenzen der Nutzer herauslesen. Hiermit maßgeschneiderte Werbung ist effizienter als die breit gestreute Werbung der Zeitungen. Dieses Geschäftsmodell macht gewisse börsennotierte Firmen zu den wertvollsten.

Doch KI kann viel mehr. Deutschland ist z. B. superstark im Maschinenbau. Hier gibt es viele spezialisierte Weltmarktführer, oft keine großen Konzerne, sondern kleine und mittelständische Unternehmen. Ihre Maschinen werden immer intelligenter. Sie nehmen mit Kameras und anderen Sensoren Informationen auf und lernen, diese zu verarbeiten und in profitable Aktionen umzumünzen. Die Produktion fast aller Dinge wird dadurch revolutioniert werden. Kein Land ist für diese kommende Welle der aktiven KI so gut aufgestellt wie Deutschland.

Unterstützende Industriepolitik ist hilfreich, um es KMUs und Start-ups zu erleichtern, neue Technologien zu nutzen und Produkte herzustellen, die für eine Weile konkurrenzlos sein werden. Zumindest einige dieser kleinen Firmen werden dann so rasch skalieren wie heute manche chinesische und amerikanische.

Innovierende Unternehmen brauchen eine ausreichende Finanzierung, unterstützende Strukturen und exzellente Fachleute, die aber oft von den USA abgeworben werden. Wie kann man die Rahmenbedingungen in Deutschland verbessern?

Schmidhuber: China zielt in den nächsten Jahren ab auf eine 150 Mrd-KI-Industrie. Deutschland sollte überproportional mitziehen, KI mit einem dreistelligen Milliardenbetrag fördern und eine weltweite KI-Führungsrolle übernehmen.

Um klein anzufangen, sollte man sich von Peking inspirieren lassen. Dort soll ein KI-Park für zwei Milliarden US-Dollar entstehen. Berlin oder München wären gute Standorte für ähnliche KI-Parks. Die sollten nicht auf bestehenden bereits verkrusteten Strukturen aufbauen, sondern neu geschaffen werden. Durch weltweite Ausschreibung müssten die besten Leute angeworben werden, um die zwei Milliarden ver-

nünftig auszugeben. Es sollte dabei allerdings kein Monopolstandort entstehen, denn Konkurrenz belebt das Geschäft.

Exzellenz ist das entscheidende Kriterium bei der Auswahl von Wissenschaftlern. Allerdings brauchen wir nicht nur Exzellenz in der Forschung, sondern auch bei der Umsetzung in marktfähige Produkte und bei der Gründung von Start-ups. In Deutschland ist es relativ leicht, ein Start-up zu gründen, aber schwierig, es hoch zu skalieren. Hier gibt es kaum große private Wagniskapital-Geber. Daher wäre es wohl nicht verfehlt, Amerika und China nachzuahmen: Der Staat ist dort oft der Anschubfinanzier, er betreibt Industriepolitik und unterstützt seine jungen Unternehmen, so dass sie schnell wachsen und konkurrenzfähig werden.

Das Internet der Dinge (*Internet of Things* oder IoT) mit intelligenten Maschinen, das viel größer sein wird als das Internet der Menschen, ist für Deutschland und Europa eine große Chance. Viele dieser intelligenten Komponenten werden im deutschsprachigen Raum hergestellt. IBM Watson hat nicht ohne guten Grund sein IoT-Hauptquartier aus den USA nach München verlegt.

Hat der europäische Forschungsraum für KI einen besonderen Mehrwert?

Schmidhuber: Das CERN als europäische Forschungseinrichtung funktioniert sehr gut – in der Forschung, aber nicht unbedingt bei der wirtschaftlichen Umsetzung. Das World Wide Web wurde dort entwickelt, aber dessen wirtschaftliche Profiteure sitzen, wie so oft, im pazifischen Raum. Ein KI-Park sollte Forschung und Umsetzung vereinbaren. Wie man das auf europäischer Ebene regelt, ist nicht klar. Es darf auf keinen Fall um nationalen Proporz gehen. Wer jedes Land gleichermaßen beteiligt, obwohl KI-Fachleute unterschiedlich verteilt sind, wäre schlecht beraten.

Reagiert Politik richtig auf die KI-Entwicklung? Erkennt sie die potenziellen Stärken der KI und handelt sie entsprechend? Gibt es ein Informationsdefizit oder ein Handlungsdefizit?

Schmidhuber: Nicht jeder Entscheidungsträger ist ausreichend informiert. Der Beratungsbedarf ist groß. In China hat man früher begriffen, wie wichtig Digitalisierung und KI sind. Dort wollte man so zentrale Entwicklungen wie Suchmaschinen nicht einem anderen Land überlassen. Die EU war lange Zeit weltgrößter Wirtschaftsraum – vor nur zehn Jahren (2008) war die EU wirtschaftlich noch so stark wie USA und China zusammen. Man war vielleicht deshalb zu selbstsicher und zu offen. Dass die großen amerikanischen IT-Firmen, wie Microsoft, Google und Amazon, nach Europa kamen, war nur natürlich. In umgekehrter Richtung gab es kein europäisches IT-Unternehmen, das in den USA oder in China einen ähnlich großen Einfluss erlangen konnte. Für China hat es sich sehr gelohnt, frühzeitig die eigenen Stärken zu betonen und ausländische Firmen zu blockieren. Heute gibt es dort etliche Firmen, die fast so wertvoll sind wie die amerikanischen und sie vielleicht bald überflügeln. In Europa und auch in Deutschland hat die Politik lange nicht verstanden, wie wichtig das alles werden würde. Und deshalb wurden Digitalisierung und KI auch kaum gefördert.

Diskutieren wir in Politik und Gesellschaft in Bezug auf KI über die falschen Fragen? Zu sehr über Ethik und Regulierung und zu wenig über Innovationen und Förderung?

Schmidhuber: Ja. In Deutschland gibt es etliche philosophisch angehauchte KI-Kritiker, die kaum Ahnung von der Technologie haben. Grundfragen der Ethik lassen sich am Beispiel KI zwar ganz nett diskutieren. Aber das ist für die Wirtschaft weitgehend irrelevant.

Sie sprechen in Zeitungsartikeln und anderen veröffentlichten Texten von KI als „intelligentem Leben“ und als „neue Krone der Schöpfung“. Müssen solche Begriffe nicht Angst machen, weil sie suggerieren, dass der Mensch ins Abseits gerät?

Schmidhuber: Gar nicht. Die meisten KI-Firmen (auch unsere Firma NNAISENSE) sind höchst menschenzentriert und befassen sich vor allem damit, das Leben der Menschen länger und gesünder und angenehmer zu machen. Denn die Leute wollen nichts kaufen, das ihnen nichts bringt. Das ist der natürliche kommerzielle Druck hin zur „guten“ KI.

Doch über kommerzielle Interessen hinaus frage ich mich eben auch, wie es mit den KIs auf lange Sicht weitergehen wird. Nicht zuletzt wegen der schon erwähnten Leistungssteigerung bei Computern wäre ich sehr erstaunt, wenn wir in wenigen Jahrzehnten nicht KIs hätten, die dem Menschen in jeder nennenswerten Hinsicht überlegen sind. Längst habe ich im Labor KIs, die sich ihre eigenen Ziele setzen, die auch neue, veränderte KIs erschaffen usw. Das wird skalieren und sich irgendwann verselbstständigen, das scheint mir unaufhaltsam. Da der weitgehend lebensfeindliche doch höchst roboterfreundliche Weltraum weit mehr Ressourcen bietet als der dünne Biosphärenfilm der Erde, werden viele KIs auswandern, und der grösste Teil der sich ausbreitenden KI-Ökologie wird nach anfänglicher Faszination am biologischen Leben weitgehend das Interesse am Menschen verlieren, mittels selbstreplizierender Roboterfabriken erst das Sonnensystem besiedeln und umgestalten, dann innerhalb von wenigen Jahrtausenden die gesamte Milchstrasse, und schliesslich innerhalb von Jahrmilliarden auch den Rest des erreichbaren Universums, im Zaum gehalten nur von der beschränkten Lichtgeschwindigkeit (KIs reisen gern per Funk von Sendern zu Empfängern, deren erstmalige Errichtung allerdings Zeit kostet). Unsere Zivilisation wird dabei zum Steigbügelhalter für das gesamte Universum, welches scheinbar seit dem Urknall immer höhere Stufen der Komplexität erklimmen will.

Was halten Sie von dystopischen Science-Fiction-Romanen zu humanoiden Robotern?

Schmidhuber: E. T. A. Hoffmann beschrieb schon 1816 in seiner Erzählung *Der Sandmann* manche der Ängste, die anthropoide Automaten auslösen. Fast gleichzeitig erschien der auf die Lebenswissenschaften gemünzte Schauerroman *Frankenstein* von Mary Shelley. Mit beiden „Zukunftstechnologien“ – der KI und den Biowissenschaften – verbinden sich offensichtlich gewisse Urängste. Sie haben aber mit der Wirklichkeit in der Forschung wenig zu tun und sollten nicht unsere Haltung zu KI (und zur Biotechnologie) bestimmen.

Fürchtet Euch nicht. Am Ende wird alles gut!

Wie Künstliche Intelligenz die Medien verändert

Norbert Lossau

- › Künstliche Intelligenz (KI) wird die Medien in vielfältiger Weise transformieren. Insbesondere wird sie den Schwerpunkt vom geschriebenen hin zum gesprochenen Wort verschieben. Digitale Assistenten könnten journalistische Inhalte vermitteln.
- › KI wird Sprachbarrieren abbauen und damit die Globalisierung des Journalismus befördern.
- › Mit KI lassen sich Medienangebote hochspezifisch individualisieren. Dazu wird eine „kritische Masse“ an Daten benötigt. Extreme Personalisierung medialer Produkte kann Filterblasen-Effekte verstärken.
- › KI ermöglicht neue Geschäftsmodelle: insbesondere Kooperationen zwischen Medienhäusern und Technologiefirmen. Auch Archivdaten lassen sich monetarisieren.
- › KI automatisiert nicht nur Arbeitsabläufe, sondern auch Entscheidungsprozesse. Damit sinkt der Bedarf an Mitarbeitern.
- › Künftige Geschäftsmodelle der Medien hängen sensibel von datenschutzrechtlichen Rahmenbedingungen ab.

1. Was ist Künstliche Intelligenz?

Vor rund 50 Jahren haben Wissenschaftler den Begriff der Künstlichen Intelligenz (KI) etabliert. Entwickelt werden sollten Computer, die Aufgaben bewältigen, für die bis dato „menschliche Intelligenz“ erforderlich war. Bereits seit den 1980er Jahren setzen Industrie und Militär solche Technologien ein¹. Sie basieren auf künstlichen neuronalen Netzen, die Mechanismen des menschlichen Gehirns nachahmen und dadurch Muster in Daten, Texten und Bildern erkennen können².

Was die breite Öffentlichkeit unter KI versteht, ist vorwiegend von Science-Fiction-Romanen und -Filmen geprägt, in denen humanoide Roboter mit übermenschlichen, geistigen Fähigkeiten eine Rolle spielen. Beispiele sind das körperlose KI-System Winston im Roman *Origin* von Dan Brown³ oder das körperbetonte KI-Wesen Ava im Film *Ex Machina*⁴.

In der realen Welt waren es Schach und Go spielende Computer, die die Leistungsfähigkeit von KI-Systemen demonstrierten. Auch der KI-Computer Watson, der in der US-Quiz-Show Jeopardy! gegen menschliche Mitspieler siegte, war ein Meilenstein in der öffentlichen Wahrnehmung von KI.

Inzwischen ist das Thema KI in der gesellschaftlichen und politischen Diskussion angekommen. Von der KI werden umwälzende Veränderungen in allen Bereichen der Wirtschaft und des Lebens erwartet, zum Beispiel bei der autonomen Mobilität und der Industrie 4.0. Auch die Medien werden durch den Einsatz von KI weitere Transformationen durchlaufen.

-
- 1 Eberhard Schöneburg (Hrsg.): „Industrielle Anwendung Neuronaler Netze“. Addison-Wesley.
 - 2 Norbert Lossau: „Wenn Computer denken lernen“. Ullstein.
 - 3 Dan Brown: „Origin“. Lübbe.
 - 4 Ex Machina, britischer Spielfilm von Alex Garland (2015)

Es gibt bislang keine wissenschaftlich anerkannte Definition für Künstliche Intelligenz. Aus pragmatischer Sicht ist es auch nicht wichtig, ob eine bestimmte Technologie das Prädikat „intelligent“ verdient. Entscheidend ist, was sie leisten kann. Wenn in diesem Beitrag von KI die Rede ist, könnte es ebenso gut „fortgeschrittene Computertechnik“ heißen. Dem allgemeinen Sprachgebrauch folgend werden jedoch die neuen technischen Möglichkeiten der KI zugeschrieben.

2. Rechtschreibung und Hasskommentare

Jeder Leser von journalistischen Print- oder Online-Produkten hat sich schon über Rechtschreib- oder Grammatikfehler geärgert. Auch bei großer Sorgfalt und mit abschließender Kontrolle durch Korrektoren lassen sie sich nicht immer verhindern. Was wäre, wenn eine Maschine vollkommene Fehlerfreiheit garantieren könnte? Arbeitsabläufe würden sich beschleunigen, und die Produktqualität würde verbessert. Gerade im Online-Journalismus kann eine halbe Stunde im Konkurrenzkampf mit anderen Medien entscheidend sein. Oft stellt sich die Frage: Sofort veröffentlichen, ohne Korrekturlesen? Oder dreißig Minuten später online gehen, dann aber fehlerfrei? Diese Entscheidung wäre obsolet, würde KI Texte in Sekundenschnelle korrigieren.

Manche Medien ermöglichen ihren Nutzern, Online-Beiträge zu kommentieren. Und nicht selten kommt es zu Hasskommentaren oder gar volksverhetzenden Anmerkungen. Sie so schnell wie möglich zu löschen oder von vornherein zu verhindern, muss das Ziel jedes Mediums sein, zumal es hier gesetzliche Vorgaben⁵ gibt. KI-Systeme könnten schon bald in der Lage sein, Hasskommentare eigenständig zu erkennen und zu löschen – viel schneller als ein Mensch, und rund um die Uhr. Es gibt bereits heute Software zum Aufspüren von Schlagworten, die auf problematische Inhalte hinweisen. Doch die Leistungsfähigkeit dieser regelbasierten Systeme ist begrenzt, weil sie die Bedeutung von Texten nicht verstehen. Das werden KI-Systeme

5 <https://www.welt.de/kultur/article172179137/NetzDG-Wie-ein-gut-gemeintes-Gesetz-den-Hass-im-Netz-verstaerkt.html> (Letzter Abruf: 16.7.2018).

mit semantischer Kompetenz voraussichtlich bald können. Auch automatische Korrektursysteme werden davon profitieren.

3. KI überwindet Sprachbarrieren

Semantische KI, die Texte inhaltlich erfasst, ist auch eine Voraussetzung für gelungene automatische Übersetzungen. Übersetzungsprogramme gibt es schon lange, doch die Ergebnisse sind oft unzureichend. Manche Bedienungsanleitung gibt davon Zeugnis. Die Sprachqualität reicht für journalistische Produkte nicht aus. KI wird die Qualität automatischer Übersetzungen dramatisch verbessern. Im Idealfall merkt man einem übersetzten Text nicht mehr an, dass er ursprünglich in einer anderen Sprache geschrieben wurde. Das eröffnet den Medien neue Optionen. Zum Beispiel wären Kooperationen wie die Leading European Newspaper Alliance über Sprach- und Ländergrenzen hinweg sehr viel einfacher möglich als heute.

Wenn ein in Deutsch verfasster Beitrag ohne nennenswerten Zeitverzug auch in Französisch, Schwedisch oder Englisch vorläge, könnte er parallel bei Partnermedien in den entsprechenden Ländern erscheinen. Das würde Kosten senken und in kleineren Märkten Produkte ermöglichen, die vorher nicht finanzierbar gewesen wären. Das publizistische Spektrum würde breiter.

Seit August 2017 bietet die Kölner Firma DeepL⁶ für sieben Sprachen Online-Übersetzungen in Echtzeit an. Geplant sind 230 Sprachkombinationen. Der Bedarf an Journalisten und Übersetzern dürfte durch den Einsatz solcher Systeme tendenziell kleiner werden. Autoren müssten mit wachsender Konkurrenz aus anderen Sprachräumen rechnen. KI ermöglicht die Globalisierung des Journalismus. Allerdings wird es immer auch Themen geben, die sich nur vor Ort von Journalisten im jeweiligen Land recherchieren lassen. Sprachen, die nur von relativ wenigen Menschen gesprochen werden, könnten durch diese Entwicklung allerdings weiter marginalisiert werden.

6 <https://www.deepl.com/translator> (Letzter Abruf: 16.7.2018).

4. KI-Systeme schreiben eigenständig Beiträge

Am technologischen Horizont zeichnen sich KI-Systeme ab, die eigenständig journalistische Texte verfassen. Jedoch hängt es stark vom Genre ab, wie schnell „Kollege Computer“ Journalisten aus Fleisch und Blut verdrängen wird. Medienhäuser experimentieren bereits mit Schreibrobotern. Führender Anbieter von Systemen zum Generieren von Texten aus Rohdaten ist das 2010 in Chicago gegründete Unternehmen Narrative Science⁷. Die Grundlage dieser Technologie wurde mit dem Algorithmus StatsMonkey an der Northwestern University gelegt. Es gibt weitere Anbieter. Die Agentur AP verwendet die Software Wordsmith von Automated Insights zum Erstellen von Quartalsberichten. Bei der Washington Post textet die KI Heliograf kurze Sportberichte, und die Stuttgarter Zeitung nutzt AX Semantics von Aexea⁸ für die Nachrichtenproduktion. Eine Software des spanischen Unternehmens Narrativa⁹ verfasst für mehrere Medien Finanznachrichten.

Bislang werden computergenerierte Texte vorwiegend in den Bereichen Sport, Finanzen und Wetter erstellt. Füttert man das System etwa mit Ergebnissen der 3. Fußballliga oder den MDAX-Kursen, so kann es aus diesen Daten mithilfe von Textbausteinen einen Artikel erstellen. Ein solcher Text liest sich nicht immer monoton, denn das System kann – nach vorgegebenen Regeln – aus vielen Bausteinen wählen und damit variieren.

Forscher arbeiten an der Entwicklung von KI mit semantischen Fähigkeiten. Sie soll über ein gewisses inhaltliches Verständnis verfügen und in der Lage sein, aus Texten einer Nachrichtenagentur fertige Beiträge zu produzieren. Dazu würde sie das Rohmaterial auf vorgegebene Längen kürzen und an einen gewünschten Sprachstil anpassen. Auch die Auswahl von Bildern, das Erstellen von Überschriften, Unterzeilen und Bildunterschriften könnten fortgeschrittene Systeme übernehmen. Der (nahezu) vollautomatische Newsroom ist keine Utopie.

7 <https://narrativescience.com> (Letzter Abruf: 16.7.2018).

8 <https://www.ax-semantic.com/de> (Letzter Abruf: 16.7.2018).

9 <http://www.narrativa.com/aijournalism> (Letzter Abruf: 16.7.2018).

5. Maschinen treffen Entscheidungen

Heute wählen noch Redakteure die Themen aus. Dass sich auch dies automatisieren lässt, zeigen Newsaggregatoren wie Upday¹⁰. Mittelfristig dürften journalistische Entscheidungen immer häufiger an lernfähige Maschinen delegiert werden. Gibt man zu optimierende Zielgrößen vor, kann eine lernfähige Maschine besser als jeder Mensch entscheiden. Ein wichtiger Aspekt ist natürlich die Frage, wofür sich die Leser interessieren. KI-Systeme können aus dem bisherigem Nutzerverhalten mit einiger Treffsicherheit ableiten, ob ein neues Thema auf großes Interesse stoßen wird. Diese Form der Automatisierung verstärkt allerdings den Echokammer-Effekt¹¹. Sogar Maschinen selbst können Opfer dieses Effekts werden, wie das Beispiel des Bots Tay von Microsoft zeigt: Innerhalb kürzester Zeit wurde dieses lernfähige System zu einem Rassisten und Leugner des Holocaust¹².

Wenn es um automatisiertes Entscheiden geht, kommt die Technologie der Künstlichen Neuronalen Netze ins Spiel. Sie imitieren Mechanismen von Gehirnen und können so Aufgaben bewältigen, für die wir gemeinhin Intelligenz unterstellen. Künstliche Neuronale Netze lernen, sie sind zu Assoziationen fähig und können Muster erkennen – etwa Gesichter, Tiere oder schlicht Korrelationen in digitalen Daten aller Art. Ohne in die technischen Details zu gehen, darf man sich ihre Arbeitsweise so vorstellen: Sie werden mit möglichst vielen Daten trainiert, damit sie darin enthaltene, subtile Zusammenhänge lernen. In der Anwendungsphase kann das Netzwerk dann auch neue Daten interpretieren und Schlüsse ziehen – gleichsam auf der Basis des erlernten Wissens.

Füttert man ein Künstliches Neuronales Netzwerk mit tausenden Tierfotos und sagt ihm, um welche Tierart es sich dabei jeweils handelt,

10 <http://www.upday.com/de/> (Letzter Abruf: 16.7.2018).

11 Norbert Lossau: „Gefährden Fake News die Demokratie?“. Beitrag in „Digitale Gesellschaft“, Konrad-Adenauer-Stiftung.

12 <https://www.welt.de/wirtschaft/webwelt/article153664299/Microsofts-KI-findet-Hitler-besser-als-Obama.html>

kann es später auf neuen Fotos Tierarten eigenständig erkennen. In der Kriminalistik lässt sich die gleiche Technik zum Identifizieren von Personen nutzen. Trainiert man ein Künstliches Neuronales Netz mit menschlichen Gesichtern, erkennt es ein gelerntes Gesicht auf dem Bild einer Überwachungskamera.

Für lernfähige KI-Systeme gibt es in den Medien zahlreiche Anwendungen. Auch hier werden zunächst ausreichend viele Daten benötigt, um das System für die jeweilige Aufgabe zu trainieren. Einige Medien stellen einen Teil ihres Angebots gratis online, während ein anderer Teil nur zahlenden Kunden zugänglich ist. Doch wie entscheidet man im Einzelfall, für welchen Kanal sich ein Artikel besser eignet? Das ist eine komplexe Optimierungsaufgabe. „Frei zugänglich“ steht für möglichst große Reichweite, zum Bezahlen müssen Nutzer hingegen bereit sein. Natürlich verfügen die Entscheider über Erfahrungswissen, was sich gut verkaufen lässt und was eher für Reichweite und damit Werbung taugt. Diese Expertise kann auch ein KI-System erlangen. Es weiß von jedem veröffentlichten Beitrag, wie gut seine Performance war. Auch Parameter wie Wochentag, Uhrzeit oder eine zeitliche Konkurrenz zu bestimmten anderen Themen kann ein solches KI-System für zehntausende Veröffentlichungen berücksichtigen. Voraussetzung ist auch hier ein gewisses semantisches Verständnis. Schließlich muss die Thematik der Beiträge erkannt werden. Bis KI-Systeme dazu in der Lage sein werden, kategorisieren Menschen die Beiträge mit Schlagworten.

Bereits heute optimiert Statistical Reasoning die Verbreitung journalistischer Produkte. So spielt etwa die Washington Post Online-Beiträge zunächst mit verschiedenen Überschriften aus – die indes noch Menschen formulieren. Schon nach kurzer Zeit lässt sich erkennen, welche Überschrift den stärksten Anreiz zum Klicken auslöst. Mit diesem Wissen setzt man dann voll auf den statistischen Sieger. Analog lassen sich auch Aufmacherbilder oder Dachzeilen variieren und testen. Mit mehr Parametern wird das Spiel zunehmend komplexer. Doch für KI-Systeme sind derartige Optimierungen ein Leichtes.

Je mehr ein Medienhaus über seine Kunden und Abonnenten weiß, umso individueller kann es journalistische Angebote vermarkten. Das Motto „Daten sind das neue Öl“ gilt eben nicht nur für Google, Facebook & Co. Durch stetiges Auswerten des Nutzungsverhaltens kann KI zunehmend bessere Profile der Konsumenten erstellen. Diese lassen sich zum Optimieren der Angebote nutzen. Wenn ein Medienanbieter weiß, dass Leser X Fan des Fußballvereins Y ist, dann ist es ratsam, den Bericht zum Spiel von Y ganz oben auf die personalisierte Seite von X zu stellen. Wenn man individualisierte Produkte in großer Breite und Tiefe produzieren will, erfordert das einen massiven Einsatz von Technik.

6. Hören statt Lesen

Der Fortschritt beim automatischen Erkennen gesprochener Worte hat bereits gesellschaftliche Auswirkungen. Es gibt in der Bevölkerung eine erstaunlich hohe Akzeptanz von sprachgesteuerten Assistenzsystemen wie Alexa von Amazon, Siri von Apple oder Duplex von Google¹³. Mit diesen Systemen kann man per Sprache kommunizieren und so Haushaltsgeräte steuern oder das Internet nutzen. Weil das so bequem und so intuitiv ist, werden Bedenken, die man mit einem „Internet-Mikrofon“ in der Privatsphäre haben könnte, offenbar hinten angestellt.

Das Abfragen von im Web verfügbaren Informationen über einen digitalen Assistenten liegt im Trend. Doch dabei wird es wohl nicht bleiben. Schon bald werden die Nutzer solcher Systeme mit gleicher Bequemlichkeit auch neueste Nachrichten, andere journalistische Produkte oder das in Medienarchiven gespeicherte Wissen abrufen wollen. Hier dürften neue Geschäftsmodelle entstehen.

Computergenerierte Stimmen kennt jeder vom Navigationssystem im Auto. Blinde oder Lesemuffel lassen sich Online-Texte vorlesen, und auf Bahnsteigen hören wir aus Lautsprechern Textbausteine zu ver-

13 <https://www.zeit.de/2018/14/amazon-alexa-sprachassistentin-persoennliche-daten-test> (Letzter Abruf: 16.7.2018).

späteten Zügen. Bislang sind Computerstimmen meist noch als solche zu erkennen. Durch KI wird sich jedoch deren Qualität so verbessern, dass sie völlig natürlich klingen. Das entsprechende Forschungsgebiet nennt sich NLG (Natural Language Generation), und erste Produkte sind bereits erhältlich – etwa die NLG-Software Quill von Narrative Science¹⁴. Es wird auch möglich sein, dem Computer die Stimme einer realen Person zu verleihen.

Für Medien ist relevant, dass Mensch und Maschine sprachlich kommunizieren können – auch mobil. Im Online-Journalismus gibt es einen starken Trend weg von der stationären hin zur mobilen Nutzung, bei einigen Medien beträgt ihr Anteil schon mehr als drei viertel. Diese Entwicklung fördert die Audio-Nutzung. Das dürfte auch zu einem Comeback jener Formate führen, die sich nur schwer über das kleine Display eines Smartphones vermitteln lassen. Das Lesen langer, tiefgründiger Texte ist mobil eher anstrengend. Bequemer ist es, sie sich von einer angenehmen Stimme vorlesen zu lassen – mit geschlossenen Augen per Kopfhörer in der U-Bahn. Derzeit ist eine Renaissance von Podcasts zu beobachten, die um die Jahrtausendwende ihre erste Boom-Phase hatten. Besonders aktiv sind hier der Deutschlandfunk, der Bayerische Rundfunk, SPIEGEL ONLINE, WELT und ZEIT ONLINE.

7. KI-Journalismus benötigt kritische Masse

Journalismus war immer von Technologien abhängig, seien es nun Druckmaschinen, TV-Studios oder Sendeanlagen. Im Online-Zeitalter ist die Bedeutung von Technik jedoch noch einmal stark gewachsen. KI wird diesen Trend beschleunigen. Voraussetzung dafür sind jedoch ausreichend große Datenmengen, wobei sich die Frage stellt, wie Medien diese gewinnen können. Eine Möglichkeit liegt in der Zusammenarbeit mit Telekom- oder Tech-Unternehmen.

14 <https://narrativescience.com> (Letzter Abruf: 16.7.2018).

Beispiel einer erfolgreichen Kooperation ist der Nachrichten-Aggregator Upday – ein Projekt von Axel Springer und Samsung¹⁵. Auf den Smartphones von Samsung ist eine Upday-App vorinstalliert. Journalisten sammeln mit massiver Technikunterstützung aus vielen Quellen unterschiedlicher Sprachen jene Informationen, die dann individualisiert an die Nutzer der App ausgespielt werden. Das ist ein kleiner Vorgeschmack auf die Zukunft von KI und Medien.

Das Zeitalter der KI eröffnet den Medien diverse neue Geschäftsmodelle. Kooperationen bieten sich an, zwischen jenen, die über Inhalte (aktuell und im Archiv) verfügen, und jenen mit einem technischen Zugang zu vielen Menschen. Hier könnten auch digitale Assistenten zum Einsatz kommen. Man ruft ihnen – auf dem Sofa sitzend oder auch während einer Autofahrt – einfach zu: Lies mir den Leitartikel aus der Zeitung X vor. Oder: Was hat eigentlich Konrad Adenauer zu diesem oder jenem Thema gesagt? Oder: Wie wird morgen das Wetter?

8. Journalismus und Blockchain

Einige Experten nehmen an, dass auch die Blockchain-Technologie im Journalismus der Zukunft eine Rolle spielen wird. Die Kryptowährung Bitcoin basiert auf diesem Konzept, das auch viele andere Anwendungen ermöglicht – etwa die Verwaltung von Grundbucheinträgen. Blockchain kombiniert Technologien der digitalen Vernetzung mit Kryptografie. Noch sind die Vorstellungen über ihren Einsatz im Journalismus vage. Manche erhoffen sich von Blockchain neue Formen der Zusammenarbeit von Journalisten sowie einen besseren Schutz von Informanten. Sie könnte auch die Echtheit von Informationen sicherstellen, weil einmal in einen Block eingetragene Daten nicht mehr verändert werden können. Blockchain könnte auch neue Finanzierungsmodelle von Journalismus ermöglichen.

15 <http://www.upday.com/de/> (Letzter Abruf: 16.7.2018).

In den USA startete im Frühjahr 2018 mit Civil¹⁶ das erste auf Blockchain basierende Medienprodukt. Der Venturecapital-Geber Consensus¹⁷ schob das Projekt mit fünf Millionen US-Dollar an. Mit dem Nachrichten- und Kulturmagazin Popula ist ein weiteres Blockchain-Projekt geplant. In Deutschland gibt es das von Googles Digital News Initiative¹⁸ geförderte Content Blockchain Project. Hier sollen neue Wege der Finanzierung und Vermarktung von Medienprodukten getestet werden. Beteiligt sind die Nachrichtenagentur dpa, Golem und DIE ZEIT.

Noch lässt sich nicht sagen, ob Blockchain tatsächlich eine maßgebliche Rolle in der Welt des Journalismus spielen wird. Doch wenn, ist eine Kombination mit KI wahrscheinlich.

9. Zusammenfassung und Diskussion

KI-Technologien führen zu einem vielschichtigen Wandel der Medienindustrie. Sie ermöglichen eine immer stärkere Personalisierung. Semantische KI-Systeme werden auch inhaltliche Arbeiten übernehmen. Mit wachsender Leistungsfähigkeit werden Maschinen Beiträge teilweise oder komplett produzieren können. Entscheidungen, die heute noch erfahrene Mitarbeiter treffen, lassen sich an lernfähige Maschinen delegieren. Es werden weniger Angestellte benötigt – von der Textkorrektur über die Redaktion bis hin zu Führungsaufgaben. Die Hierarchien dürften flacher werden.

Medieninhalte werden künftig immer häufiger gesprochen als geschrieben. Das gilt auch für den Rückkanal zum Kunden. Audio-Technologien ermöglichen neue Formen der Mediennutzung. Zu erwarten sind Kooperationen von Medien und Telekom- oder Tech-Unternehmen. Der Zugang zu einer großen Zahl potenzieller Nutzer ermöglicht neue journalistische Formate und Service-Angebote. Die in den Archiven der Medienhäuser schlummernden Daten werden sich

16 <https://joincivil.com> (Letzter Abruf: 16.7.2018).

17 <https://new.consensys.net> (Letzter Abruf: 16.7.2018).

18 <https://newsinitiative.withgoogle.com/dnifund/> (Letzter Abruf: 16.7.2018).

dank KI kapitalisieren lassen. Die Qualität journalistischer Produkte lässt sich durch Rückkopplung zum Leser verbessern. Rechtschreib- oder Stilfehler werden durch den Einsatz von KI verschwinden.

Die neuen Technologien eröffnen den Medien viele Chancen. Es gibt aber auch Risiken, die man rechtzeitig bedenken sollte. Die KI-gesteuerten Mechanismen der Individualisierung fördern das Phänomen der Filterblase. Wenn jeder nur noch nach eigenen Interessen personalisierte Medienprodukte erhält und es in der Gesellschaft kein gemeinsames Basiswissen mehr gibt, kann dies für den politischen Diskurs und die Zukunft der Demokratie weitreichende Folgen haben. An dieser Stelle ist ein Dialog von Gesellschaft, Medien und Politik wünschenswert, um möglichen Fehlentwicklungen proaktiv entgegen zu steuern. Eine von einigen Medien bereits verfolgte Strategie ist das Club-Konzept. Nutzer des betreffenden Mediums sollen sich in der realen Welt begegnen, um miteinander, den Journalisten und auch mit Politikern, Wissenschaftlern oder Wirtschaftsführern, von Angesicht zu Angesicht zu kommunizieren. Das Analoge könnte das Digitale erden.

Um im Wettbewerb erfolgreich zu sein, werden Medien künftig viel stärker persönliche Daten ihrer Kunden nutzen. Je mehr man über seinen Leser, Zuhörer oder Zuschauer weiß, umso besser kann man journalistische Produkte und die sie finanzierende Werbung maßschneidern. Damit stellt sich auch die Frage nach dem Datenschutz. Medienkunden sollen nicht nur den verbreiteten Inhalten trauen können, sondern müssen sich auch darauf verlassen dürfen, dass ihre Daten nicht missbraucht werden. Der Datenskandal von Facebook und der inzwischen insolventen britischen Firma Cambridge Analytica hat gezeigt, wie schnell sich der Wind *in puncto* Vertrauen drehen kann. Medienunternehmen könnten in Zukunft ähnlich schnell am Datenschutz-Pranger stehen wie Betreiber sozialer Netzwerke.

Das seit Mai 2018 gültige neue europäische Datenschutzrecht schafft Randbedingungen, die allen Unternehmen das Sammeln personenbezogener Daten erschweren. Insbesondere der Austausch solcher Informationen zwischen verschiedenen Unternehmen wird stär-

ker reglementiert. Wer selbst über viele Daten verfügt und weniger auf Daten Dritter angewiesen ist, kann von der neuen Rechtslage in gewisser Weise profitieren. Im KI-getriebenen Journalismus wird die kritische Masse wichtiger. Gibt es nicht genügend Daten, erzielen die KI-Algorithmen nur schlechte Ergebnisse. Insgesamt fördert die technologische Entwicklung das Entstehen großer Einheiten. Daraus werden sich medienpolitische Fragen ergeben. Der KI-bedingte Wegfall von Sprachgrenzen ist ein Effekt, der zu supranationalen und damit deutlich größeren Nutzergruppen führen könnte. Offen ist die Frage, wie sich dies auf den Wettbewerb der verschiedenen Medien auswirkt.

Zum Schluss ein Gedanke, der heute noch Science-Fiction ist, aber Wirklichkeit werden könnte. KI-Systeme, die qualitativ hochwertigen Journalismus präsentieren sollen, werden selbst journalistische Lektüre benötigen, um sich auf dem Laufenden zu halten und stets dazu zu lernen. Es klingt zwar merkwürdig, doch in absehbarer Zeit gibt es wahrscheinlich Medienprodukte von Maschinen für Maschinen. Doch vielleicht wollen die Maschinen ab und zu auch mal einen richtig guten Artikel lesen – den ein Mensch geschrieben hat.

Referenzen

- B Brown, Dan.** Origin. Bastei Lübbe, 2018.
- C Cwiertnia, Laura.** „Amazon Alexa. Meine unheimliche Mitbewohnerin“. <https://www.zeit.de/2018/14/amazon-alexasprachassistentin-persoенliche-daten-test> (Letzter Abruf: 16.07.2018).
- E Ex Machina.** britischer Spielfilm von Alex Garland (2015).
- L Lossau, Norbert.** „Gefährden Fake News die Demokratie?“ Beitrag in: Digitale Gesellschaft. Gestaltungsräume, S. 127–138, Konrad-Adenauer-Stiftung, 2018.
- Lossau, Norbert.** Wenn Computer denken lernen. Ullstein, 1992.
- M Meier, Christian.** „Wie ein gut gemeintes Gesetz den Hass im Netz verstärkt.“ <https://www.welt.de/kultur/article172179137/NetzDG-Wie-ein-gut-gemeinetes-Gesetz-den-Hass-im-Netz-verstaerkt.html> (Letzter Abruf: 16.07.2018).
- „**Microsofts KI findet Hitler besser als Obama**“. <https://www.welt.de/wirtschaft/webwelt/article153664299/Microsofts-KI-findet-Hitler-besser-als-Obama.html> (Letzter Abruf: 15.07.2019).
- S Schöneburg, Eberhard** (Hrsg.). Industrielle Anwendung Neuronaler Netze. Addison-Wesley, Bonn, 1993.

Weblinks

<https://www.deepl.com/translator>
(Letzter Abruf: 16.07.2018).

<https://narrativescience.com>
(Letzter Abruf: 16.07.2018).

<https://www.ax-semantics.com/de>
(Letzter Abruf: 16.07.2018).

<http://www.narrativa.com/aijournalism>
(Letzter Abruf: 16.07.2018).

<http://www.upday.com/de/>
(Letzter Abruf: 16.07.2018).

<https://joincivil.com>
(Letzter Abruf: 16.07.2018).

<https://new.consensys.net>
(Letzter Abruf: 16.07.2018).

<https://newsinitiative.withgoogle.com/dnifund/>
(Letzter Abruf: 16.07.2018).

Rahmenbedingungen und Regeln

Rechtspersönlichkeit für autonome Systeme?

Jan-Erik Schirmer

- › Der Einsatz intelligenter Roboter und anderer autonomer Systeme lässt Verantwortungslücken entstehen, auf die das Zivilrecht reagieren muss.
- › Für den Lückenschluss braucht es jedoch keine Promotion zum Rechtssubjekt in Form einer „elektronischen Persönlichkeit“, sondern lediglich ein funktionales Verständnis von Rechtsfähigkeit.
- › Bis auf Weiteres werden autonome Systeme dazu eingesetzt werden, ihre Nutzer bei bestimmten Tätigkeiten zu unterstützen – und nur insoweit sollten ihnen passende rechtliche Fähigkeiten zugeschrieben werden, etwa als Stellvertreter oder als Hilfsperson im Sinne des Haftungsrechts.
- › Dies sollte vorrangig durch eine funktionale Interpretation der bestehenden Normen geschehen. Allein auf dem Gebiet der außervertraglichen Haftung sollte der Gesetzgeber mit der Einführung einer verschuldensunabhängigen Haftung des Systemnutzers nachhelfen. Diese punktuellen Eingriffe versprechen nicht nur passgenauere Lösungen, sie verhindern auch, dass autonome Systeme mithilfe einer umfassenden „elektronischen Persönlichkeit“ vermenschlicht werden und damit neue Rechtskonflikte entstehen.

Die Digitalisierung verändert unseren Alltag und sie verändert unser Denken. Was gestern noch als Science-Fiction galt, wird heute Stück für Stück Realität: Saudi-Arabien macht einen Roboter zum Staatsbürger, das Europäische Parlament regt eine „elektronische Persönlichkeit“ für intelligente Roboter und andere autonome Systeme an. Ist das bloß Publicity oder steckt mehr dahinter – und wäre aus Sicht des Zivilrechts mit einer solchen Rechtspersönlichkeit überhaupt etwas gewonnen?

1. Können autonome Systeme überhaupt Rechtssubjekte sein?

Rechtsperson bzw. Rechtssubjekt zu sein bedeutet, vom Recht als selbstständiger Akteur mit eigenen Rechten und Pflichten wahrgenommen zu werden. Das prominenteste Rechtssubjekt ist der Mensch. Allerdings ist damit nicht gesagt, dass Rechtssubjektivität notwendigerweise an Eigenschaften wie Willensfreiheit, Bewusstsein oder Emotionalität geknüpft ist (Kersten, 2017). Dies belegt schon die Existenz sog. juristischer Personen wie Aktiengesellschaften, Vereine oder Staaten. Diese juristischen Kunstgebilde werden – gleich dem Menschen – vom Recht als eigenständiges Subjekt wahrgenommen, obwohl sie keinen freien Willen und kein Bewusstsein haben. Es ist nicht einmal nötig, dass sie sich aus menschlichen Mitgliedern zusammensetzen, also ein „personales Substrat“ oder ein wie auch immer gearteter „Gemeinwillen“ vorhanden ist. Im Privatrecht ist das sogar an der Tagesordnung, weil hier – man denke an Großkonzerne – juristische Personen gleich einer Matroschka-Puppe aus anderen juristischen Personen zusammengesetzt sind. Selbiges lässt sich über Stiftungen sagen, die keine menschlichen Mitglieder haben und dennoch Personen im Rechtssinne sind.

Wenn es also nicht der „menschliche Kern“ ist, der darüber entscheidet, wer oder was Rechtssubjekt wird, was ist es dann? Die Antwort mag banal klingen: Es ist allein das Recht selbst, das die Entscheidung trifft. Rechtssubjekt könnte dann alles sein, was das Recht dazu erklärt – Bäume, Tiere oder eben Roboter. Dieser Gedanke klingt

zunächst abwegig. Doch entspricht er einer verbreiteten Praxis: In Neuseeland wurde einem Fluss Rechtspersönlichkeit zuerkannt. Immer wieder wird darüber diskutiert, gleiches bei Primaten und anderen hochentwickelten Lebewesen zu tun. Und auch in der Einbürgerung eines Roboters durch das Königreich Saudi-Arabien steckt letztlich die Entscheidung des saudischen Rechts, diesen Roboter als Rechtssubjekt anzuerkennen – selbst wenn dies wohl allein deshalb geschah, um die eigene Digitalindustrie besser zu vermarkten.

Trotzdem gibt es aus guten Gründen Vorbehalte, den Gesetzgeber bei der fundamentalen Frage der Rechtssubjektivität frei schalten und walten zu lassen (Eidenmüller, 2017). Denn der Status als Rechtssubjekt bedeutet eine nicht unerhebliche normative Aufwertung: Etwas, das vom Recht bislang nur am Rande wahrgenommen wurde, wird ins Zentrum gerückt und als eigenständiger Akteur angesprochen. Zugleich wird es auf eine Stufe mit anderen Rechtssubjekten gestellt und damit sehr nah an den prominentesten Vertreter – den Menschen – herangerückt. Zwar bliebe es dem Gesetzgeber unbenommen, sich über diese Bedenken hinwegzusetzen und Robotern und anderen autonomen Systemen umfassende Rechtspersönlichkeit zuzuerkennen. Aber er sollte es nur dann tun, wenn es gute Gründe dafür gibt.

2. Rechtspersönlichkeit als Antwort auf Autonomierisiken und Verantwortungslücken?

An diesem Punkt setzt die Diskussionen über eine eigene Rechtspersönlichkeit für autonome Systeme an. Dabei lassen sich zwei Begründungsstränge unterscheiden. In der (Technik-)Philosophie wird argumentiert, dass autonome Systeme in ihren Fähigkeiten dem Menschen so nahegekommen seien, ja ihn mitunter bereits überflügelt hätten, dass es nur folgerichtig wäre, ihnen einen vergleichbaren Status zuzuerkennen (Matthias, 2008). Demgegenüber wird in der rechtswissenschaftlichen Debatte auf die disruptive Kraft autonomer Systeme verwiesen: Da deren Verhalten keinem festgelegten Ablauf mehr folge, sondern sich dank der Fähigkeit zur selbständigen Informationsverarbeitung und Entscheidungsfindung durch ein hohes

Maß an Unberechenbarkeit auszeichne, entstehe ein neuartiges Autonomierisiko. Dieses Autonomierisiko bringe, so die zentrale These, das geltende Recht an seine Belastungsgrenze. Es entstünden – besonders im privaten Haftungsrecht – Verantwortungslücken, die durch neue, eigens auf autonome Systeme zugeschnittene Rechtsregeln geschlossen werden müssten. Große Hoffnungen werden dabei vor allem in die eigens zu schaffende Rechtspersönlichkeit gesetzt (Mayinger, 2017).

Diesem Gedankengang folgt auch das Europäische Parlament in seinem Aufruf an die Europäische Kommission, EU-weite zivilrechtliche Regelungen im Bereich der Robotik zu erlassen. Es sei unerlässlich, sich mit den „rechtlichen und ethischen Implikationen“ dieser „neuen industriellen Revolution“ auseinanderzusetzen. Neben neuen Haftungsregeln müsse dazu langfristig „ein Status als elektronische Person“ für die „ausgeklügelsten“ Roboter geschaffen werden, um sie so für eigenständige Entscheidungen und verursachte Schäden verantwortlich machen zu können (Europäisches Parlament, 2017).

3. Gibt es Verantwortungslücken?

Aber stimmt die Prämisse überhaupt? Ist es tatsächlich so, dass das geltende Zivilrecht die digitale Revolution nicht mehr bewältigen kann, so dass der Ausweg in neuen Regeln und einer Rechtspersönlichkeit für autonome Systeme gesucht werden muss? Um diese Frage zu beantworten ist es unerlässlich, zunächst die bestehenden Regeln „auf den Prüfstand [zu] stellen“, wie es auch der Koalitionsvertrag zwischen CDU, CSU und SPD fordert (Koalitionsvertrag, 2018). Zwei Ebenen sind dabei in den Blick zu nehmen – die Ebene des Vertragsschlusses und die der Haftung.

Vertragsschlüsse durch autonome Systeme

Verträge kommen nach deutschem Recht durch zwei übereinstimmende Willenserklärungen zustande – Angebot und Annahme. Gehe ich abends in eine Pizzeria und bestelle dort eine Salamipizza, gebe ich eine Willenserklärung in Form eines Angebots ab, mit dem

ich eine Rechtsfolge auslösen möchte – den Kauf der Pizza. Wie aber ist folgendes Szenario rechtlich zu bewerten? Ich weise meinen digitalen Assistenten an, für mich Essen zu bestellen. Der Assistent bestellt, nachdem er die Uhrzeit, meine vorherigen Speisen und meine Vorlieben abgeglichen hat, Salat, weil er zu der Entscheidung kommt, Salamipizza hatte ich schon genug. Ist das dann noch meine Willenserklärung, und ist überhaupt ein wirksamer Vertrag gegeben?

Viele Juristen beantworten beide Fragen positiv. Es handle sich um einen Fall der sog. Computererklärung. Obwohl der digitale Assistent die Erklärung abgibt, soll dies nichts daran ändern, dass dem Nutzer die Willenserklärung als eigene zugerechnet wird; schließlich habe er entschieden, das System für seine Zwecke zu nutzen. Selbst wenn das System eine vom Nutzer nicht gewollte Erklärung abgibt, also wie im obigen Fall Salat statt Pizza geordert wird, ändere das nichts an der Zurechnung. Denn aus Vertrauensgesichtspunkten müsse das Verhalten des Systems immer dem Nutzer zugerechnet werden (Müller-Hengstenberg/Kirn, 2016; Arbeitsgruppe Digitaler Neustart, 2017).

Bei dieser Sichtweise ergibt sich auf Ebene des Vertragsschlusses in der Tat keine Verantwortungslücke, weil die Erklärung, die ein autonomes System abgibt, immer als solche des jeweiligen Nutzers behandelt wird. Diese pragmatische Lösung lässt allerdings viel Raum für Widerspruch. Denn an sich ist es der digitale Assistent, der im obigen Beispiel meine Anweisung „Essen bestellen“ erst zum rechtlich wirksamen Angebot macht. Er wählt aus, was bestellt werden soll, bei wem und zu welchem Preis. Die wesentlichen Bestandteile des Vertrags werden also von dem Assistenten festgelegt und als Angebotserklärung formuliert. Meiner Anweisung fehlt es umgekehrt an allem, was ein Angebot ausmacht, denn ich lege es in die Hände des Assistenten, selbstständig eine für mich ungewisse Rechtswirkung herbeizuführen. Exakt diese Einwände treiben auch das EU-Parlament um, wenn es mahnt, dass „Maschinen, die dazu konzipiert sind, ihr jeweiliges Gegenüber auszuwählen, vertragliche Bedingungen auszuhandeln [und] Verträge abzuschließen [...], die herkömmlichen Regeln unanwendbar machen, was die Notwendigkeit für neue, wirksame und aktuelle Regeln unterstreicht.“ (Europäisches Parlament, 2017)

Denkbar ist jedoch, diese Fälle mit dem bewährten Institut der Stellvertretung (§§ 164 ff. BGB) zu lösen (Specht/Herold, 2018; Teubner, 2017). Die Erklärung würde als das eingeordnet werden, was sie ist, nämlich als eine des autonomen Systems selbst. Zugleich wären passgenauere Lösungen möglich, die dem Autonomierisiko Rechnung tragen: Die Erklärung wäre für den jeweiligen Nutzer nicht stets verbindlich, sondern nur dann, wenn das autonome System mit Vertretungsmacht gehandelt hat; Irrtumskonstellationen ließen sich ebenfalls mit bestehenden Regeln auffangen. Dafür müsste jedoch die Erklärung des Systems als Willenserklärung aufgefasst werden – und zu Willenserklärungen sollen nach üblicher Lesart nur Menschen fähig sein, da allein sie kraft ihrer Willensmacht Rechtsfolgen auslösen können. Wer einen Vertragsschluss durch ein autonomes System dem Stellvertretungsrecht unterwerfen will, scheint folglich darauf angewiesen zu sein, autonome Systeme als Rechtssubjekte anzuerkennen. Denn ohne Rechtspersönlichkeit keine Willenserklärung, und ohne Willenserklärung keine Stellvertretung. Diese Lösung schwebt offenbar auch dem Europäischen Parlament vor: Die geplante „elektronische Persönlichkeit“ soll Abhilfe schaffen, damit „Roboter eigenständige Entscheidungen treffen oder anderweitig auf unabhängige Weise mit Dritten interagieren“ können (Europäisches Parlament, 2017).

Bei genauer Betrachtung zeigt sich aber, dass es für die Anwendung des Stellvertretungsrechts nicht zwingend nötig ist, autonome Systeme mit einer umfassenden elektronischen Persönlichkeit auszustatten. Es ist nur erforderlich – und zugleich völlig ausreichend –, dem System die Vertretung seines Nutzers zu ermöglichen, ihm also insoweit einen partiellen rechtlichen Akteurstatus zuzuweisen (Schirmer, 2016; Teubner, 2017). Das System wäre so fähig, eine eigene Willenserklärung für seinen Nutzer – und nur für seinen Nutzer! – abzugeben. Dadurch würden auch die technischen Realitäten besser abgebildet, denn bis auf Weiteres werden autonome Systeme dazu eingesetzt werden, ihre Nutzer bei bestimmten Tätigkeiten zu unterstützen. Das System wäre also nicht Subjekt mit umfassenden Rechten und Pflichten, sondern getreu seiner Assistenzfunktion lediglich vertretungsfähig.

Dieses Vorgehen hat einen ganz entscheidenden Vorteil: Mittels einer *strikt funktional verstandenen Rechtsfähigkeit für autonome Systeme* wird das oben angesprochene Vermenschlichungsproblem vermieden. Denn die Erfahrung im Umgang mit der juristischen Person lehrt, dass mit der gesetzgeberischen Promotion zum Rechtssubjekt ein impliziter Rechtfertigungsdruck einherginge, warum autonomen Systemen bestimmte Rechte *nicht* zustehen sollten. Mit anderen Worten wären über kurz oder lang Begründungen erforderlich, warum autonome Systeme im Gegensatz zu anderen Rechtssubjekten keine Schutzrechte wie Arbeitnehmer- oder Grundrechte haben sollten. Der Einmaleffekt an Rechtssicherheit, der durch die Schaffung einer elektronischen Persönlichkeit suggeriert wird, würde daher unweigerlich erhebliche Folgeprobleme heraufbeschwören.

Demgegenüber ist die Lösung mittels einer partiellen Rechtsfähigkeit minimalinvasiv. Die Promotionswirkung ist weitaus geringer, denn das autonome System wird nur in seiner Rolle als fremdnütziger Assistent zum rechtlichen Akteur. Dazu ist lediglich eine (Neu-)Interpretation bestehender Normen und Institute durch Rechtsprechung und Rechtslehre erforderlich – so wie es bei anderen (technischen) Innovationen oft gut funktioniert hat. Dadurch lässt sich passgenau und von Fall zu Fall auf die Rechtsprobleme reagieren, die sich im Zuge der Entwicklung autonomer Systeme stellen und noch stellen werden. Gerade weil wir heute noch gar nicht abschätzen können, welche rechtlichen Herausforderungen zukünftig warten, ist es ratsam, zunächst auf Sicht zu fahren – und nicht mit einer umfassenden elektronischen Persönlichkeit vorschnell Fakten zu schaffen.

Haftung für autonome Systeme

Wie dargelegt, lässt sich typischen Risiken auf der Ebene des Vertragschlusses gerecht werden, sofern man autonome Systeme als Stellvertreter ihrer Nutzer begreift. Wie aber steht es um die Schadensersatzhaftung?

Vertragliche Haftung

Die vertragliche Haftung erfasst Fälle, in denen zwischen den Parteien bereits ein Schuldverhältnis existiert, eine der Parteien einen Schaden

erleidet und dieser Schaden auf einer schuldhaften Pflichtverletzung beruht (§ 280 BGB). Wie wäre nun der Fall zu behandeln, dass ich meinem digitalen Assistenten vollständig das berufliche Fristenmanagement übertrage und das Gerät aufgrund einer internen Fehlerarbeitung einen wichtigen Termin unterschlägt, wodurch einem meiner Kunden ein finanzieller Schaden entsteht?

Auch hier wird vertreten, derartige Fälle ließen sich ohne größere Anpassungen des geltenden Rechts lösen. Die Pflichtverletzung wird darin erblickt, dass der Nutzer durch die Übertragung des Fristenmanagements den digitalen Assistenten zur Schädigung veranlasst habe. Der zusätzlich nötige Verschuldensvorwurf sollte ebenfalls wenig Probleme bereiten, weil der Nutzer darlegen müsse, warum ihn kein Vorwurf treffe. Dies, so die These, werde dem Nutzer regelmäßig nicht gelingen, vor allem weil es ohnehin leichtfertig erscheine, eine derart wichtige Aufgabe wie die Terminplanung komplett in die Hände eines technischen Geräts zu legen (Spindler, 2014).

Es leuchtet jedoch nicht ein, warum allein in der Delegation einer Aufgabe an ein autonomes System eine Pflichtverletzung liegen soll. Würde die gleiche Aufgabe auf einen anderen autonomen Akteur – einen Menschen zumal – übertragen, käme niemand auf die Idee, allein in der Delegation eine Pflichtverletzung zu erblicken. Vor allem aber ist es ein Irrtum anzunehmen, dass es dem Nutzer regelmäßig nicht gelänge, die Verschuldensvermutung zu entkräften. Dafür wäre nur nachzuweisen, dass der Schadenseintritt nach dem Stand der Technik weder vorhersehbar noch vermeidbar war. Dieser Nachweis wird regelmäßig gelingen, da sich autonome Systeme gerade durch ihre „Unberechenbarkeit“ auszeichnen. In den typischen Schadenssituationen bei Einsatz eines autonomen Systems – der Schädigung des Vertragspartners infolge einer systeminternen Fehlentscheidung – wäre der Nutzer also regelmäßig gerade nicht haftbar (Bitkom/DFKI, 2017; Teubner, 2017). Diese Sorge teilt auch das Europäische Parlament in seiner Entschließung an die EU-Kommission. Weil das geltende Recht voraussetze, dass die „Ursache der Handlung oder Unterlassung eines Roboters auf einen konkreten menschlichen Akteur, wie zum Beispiel den Hersteller, den Betreiber oder den Benutzer, zurückverfolgt werden kann“, und es

darüber hinaus nötig sei, dass „dieser Akteur das schädliche Verhalten des Roboters hätte vorhersehen und verhindern können“, seien die „Defizite des derzeit geltenden Rechtsrahmens auf dem Gebiet der Vertragshaftung insofern offensichtlich“ (Europäisches Parlament, 2017).

Könnte aber eine Rechtspersönlichkeit für autonome Systeme helfen, diese Verantwortungslücke zu schließen? In der Rechtslehre wird das vielfach bejaht, weil nur so der bislang versperrte Weg über die Haftung für Erfüllungsgehilfen gangbar wird (§ 278 BGB). Erst indem der digitale Assistent zum Rechtssubjekt werde, so die Argumentation, ließe er sich als *Hilfsperson* einordnen. Die Verantwortungslücke wäre damit in der Tat schlagartig geschlossen, denn für das Fehlverhalten von Hilfspersonen muss der Schuldner kraft Gesetzes immer einstehen. Im obigen Beispiel wäre ich also für meinen digitalen Assistenten haftbar, genauso wie ich für menschliche Hilfspersonen haftbar bin, unabhängig davon, ob ich eine Pflicht schuldhaft verletzt habe.

Indes gilt auch hier das Gleiche wie bei der Stellvertretung: Zielführender ist es, das autonome System *insoweit* als rechtsfähig anzusehen, als dass es Erfüllungsgehilfe sein kann. Es reicht, dem System, ganz im Sinne seines dienenden Bestimmungszwecks, einen limitierten Status als unterstützender Akteur zuzuweisen, es also als fähig anzusehen, Hilfsperson im Sinne des Haftungsrechts zu sein (Schirmer, 2016; Teubner, 2017). Dazu bedarf es nicht der Schaffung einer elektronischen Persönlichkeit. Eine funktionale Interpretation der bestehenden Regeln durch Rechtsprechung und Rechtslehre ist ausreichend und vorzuzugswürdig, um die oben angesprochenen Folgeprobleme einer gesetzlichen Anerkennung zu vermeiden.

Außervertragliche Haftung

Wie aber steht es um Schädigungen außerhalb bestehender Vertragsverhältnisse? Zu denken wäre etwa an den Fall, dass mein sonst stets zuverlässiger digitaler Assistent, dem ich die Bewässerung der Balkonpflanzen übertragen habe, das Wasser über Nacht laufen lässt, wodurch auch die wertvollen Zuchtrosen meines Nachbarn eine Etage tiefer zerstört werden. Kommt auch in diesen Fällen eine Haftung des Nutzers in Betracht?

Vereinzelte Stimmen bejahen eine Haftung nach geltendem Recht. Anknüpfungspunkt sollen die sog. Verkehrspflichten sein. Nach dieser Lösung wäre der Nutzer des digitalen Assistenten haftbar, da er durch den Einsatz des Systems einen Gefahrenbereich eröffnet und nicht ordnungsgemäß überwacht hat. Und weil dadurch ein Rechtsgut – im Beispiel das Eigentum des Nachbarn – verletzt wurde, soll eine Schadenersatzpflicht bestehen (§ 823 BGB).

Dieser Argumentation lassen sich jedoch die gleichen Argumente wie der Vertragshaftung entgegenhalten: Auch eine Deliktshaftung kommt nur bei Verschulden in Betracht, und das Verschulden muss der Geschädigte hier sogar von sich aus darlegen und beweisen. Genauso wie bei der Vertragshaftung wird ihm dies wohl nur selten gelingen, weil der Nutzer eines autonomen Systems auf das typische Autonomierisiko verweisen darf: Der Schaden ist nicht vorhersehbar, derartige Fehlfunktionen lassen sich selbst bei bestmöglicher Überwachung kaum verhindern. Kurz: In den für autonome Systeme charakteristischen Schadenssituationen ist eine deliktische Haftung des Nutzers nach geltendem Recht regelmäßig ausgeschlossen (Bitkom/DFKI, 2017; Arbeitsgruppe Digitaler Neustart, 2017). Ausnahmen gelten allein im Straßenverkehr, wo eine verschuldensunabhängige Haftung des Halters besteht (vgl. § 7 StVG).

Kann sich der Geschädigte dann zumindest an den Hersteller des digitalen Assistenten halten? Anknüpfungspunkt könnte vor allem die sog. Produkthaftung sein, die eine Haftung des Herstellers für fehlerhafte Produkte anordnet (§ 1 ProdHaftG). Der digitale Assistent wäre ein solches Produkt; Gleiches gilt für andere autonome Systeme, jedenfalls sofern diese nicht nur Software sind, sondern „verkörpert sind“, also angefasst werden können. Auch hier steckt der Teufel aber im Detail: Der Produzent haftet nur, wenn das Produkt fehlerhaft ist. Bei autonomen Systemen wird das häufig zweifelhaft sein. Denn der „Fehler“ des Systems liegt typischerweise nicht in einer fehlerhaften Konstruktion oder Programmierung wie es dem Produkthaftungsgesetz vorschwebt, sondern darin, dass eine systeminterne Informationsverarbeitung schlicht in eine schadenstiftende Entscheidung mündet (Wagner, 2017). Hinzu kommt, dass die Produkthaftung auf bestimmte

Schäden (z. B. Sach- und Körperschäden, nicht dagegen Vermögen) beschränkt und auch der Höhe nach gedeckelt ist. Nicht zuletzt aufgrund dieser Ungewissheiten wird in der deutschen Politik darüber nachgedacht, „klare Regelungen für die Produkthaftung in der digitalen Welt auf[zu]stellen“ (Koalitionsvertrag, 2018).

Das Europäische Parlament ist schon einen Schritt weiter. Ihm schwebt vor, eine verschuldensunabhängige und unbegrenzte Haftung speziell für Hersteller von Robotern und anderen autonomen Systemen einzuführen, die lediglich „eine kausale Verknüpfung zwischen dem schädlichen Funktionieren des Roboters und dem Schaden“ verlangt (Europäisches Parlament, 2017; zustimmend Arbeitsgruppe Digitaler Neustart, 2017). In der Tat wären damit die Probleme für den Geschädigten gelöst – umgekehrt könnte sich der Hersteller aber wohl nie einer Haftung entziehen. Er wäre immer haftbar, obwohl er auf den Einsatz und die Entscheidungsprozesse des autonomen Systems kaum Einfluss (mehr) nehmen kann. Kann man Herstellern dann tatsächlich zumuten, Millionenbeträge in die Entwicklung unbeherrschbarer Haftungsrisiken zu stecken?

Statt die Verantwortungslücke mithilfe einer strikten Herstellerhaftung zu schließen, wollen einige Stimmen zu einem anderen Modell greifen. Es zielt darauf, wie auch sonst, den Verursacher der Schädigung in den Blick zu nehmen. Müsste nicht, so die Idee, über eine Haftung des autonomen Systems selbst nachgedacht werden? Freilich brauchte es dazu einen Vermögenspool, aus dem die Haftsumme abgeschöpft werden könnte. Das wäre aber kein großes Problem: Entweder stattet man das autonome System selbst mit einem Haftkapital aus, das sich mitunter durch Provisionszahlungen für jede Tätigkeit noch erhöhen könnte. Oder man führt Pflichtversicherungen bzw. Haftungsfonds ein, die die Schäden ausgleichen (Mayinger, 2017). Allerdings wäre es dafür im ersten Schritt nötig, das autonome System zum Haftungsadressaten zu machen – es also entweder als Rechtssubjekt oder doch zumindest als deliktstfähigen Akteur zu begreifen.

Auch hier wäre die punktgenaue Zuweisung in Form eines deliktischen Akteurstatus indes vorzugswürdig – jedoch mit einer wichtigen Einschränkung: Eine Eigenhaftung des autonomen Systems, selbst wenn diese letztlich in einer Versicherungs- oder Fondslösung mündet, erscheint wenig sinnvoll. Denn das autonome System vollbringt lediglich Unterstützungsleistungen, agiert also gerade nicht im Eigeninteresse, sondern im Interessenkreis seines Nutzers. Dann aber ist es nur konsequent, mit einer strikten Haftung des Nutzers zu antworten, da er über den Einsatz des autonomen Systems entscheidet und von dessen Unterstützung profitiert (Teubner, 2017; Arbeitsgruppe Digitaler Neustart, 2017). Hierfür – und nur hierfür – ist in der Tat ein Einschreiten des Gesetzgebers nötig, weil das deutsche Recht eine strikte Deliktshaftung für Schädigungen durch fremdnützige Agenten nicht kennt. Sie einzuführen wird seit Jahrzehnten gefordert – vielleicht ist die digitale Revolution genau der Funke, der bislang gefehlt hat.

4. Fazit

Eine „elektronische Persönlichkeit“ für autonome Systeme ist abzulehnen. Rechtsprechung und Rechtslehre sind in der Lage, mithilfe funktionaler Gesetzesinterpretation den Großteil der bestehenden Verantwortungslücken schließen. Allein auf dem Gebiet der außervertraglichen Haftung sollte der Gesetzgeber mit der Einführung einer verschuldensunabhängigen Haftung des Systemnutzers nachhelfen. Eine Einstandspflicht des autonomen Systems selbst oder eine Ausweitung der Herstellerhaftung ist dagegen nicht angezeigt.

Referenzen

- A** **Arbeitsgruppe Digitaler Neustart der Konferenz der Justizministerinnen und Justizminister der Länder.** Bericht vom 15. Mai 2017. – 413 S. https://www.justiz.nrw.de/JM/schwerpunkte/digitaler_neustart/zt_bericht_arbeitsgruppe/bericht_ag_dig_neustart.pdf [zuletzt abgerufen: 06.03.2018].
- B** **Bitkom e. V./DFKI GmbH.** Entscheidungsunterstützung mit Künstlicher Intelligenz – Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung. – Berlin und Kaiserslautern, 2017. – 228 S. <https://www.bitkom.org/noindex/Publikationen/2017/Sonstiges/KI-Positionspapier/171012-KI-Gipfelpapier-online.pdf> [zuletzt abgerufen: 06.03.2018].
- C** **CDU/CSU/SPD.** Ein neuer Aufbruch für Europa – Eine neue Dynamik für Deutschland – Ein neuer Zusammenhalt für unser Land: Koalitionsvertrag zwischen CDU, CSU und SPD. Berlin, 2018. 177 S. https://www.cdu.de/system/tdf/media/dokumente/koalitionsvertrag_2018.pdf?file=1 [zuletzt abgerufen: 06.03.2018].
- E** **Eidenmüller, Horst.** The Rise of Robots and the Law of Humans. In: Zeitschrift für Europäisches Privatrecht 2017, S. 765–777.
- Europäisches Parlament.** Zivilrechtliche Regelungen im Bereich Robotik (P8_TA(2017)0051) – Entschließung des Europäischen Parlaments vom 16. Februar 2017 mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103(INL)). – 28 S. <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+Vo//DE> [zuletzt abgerufen: 06.03.2018].
- K** **Kersten, Jens.** Relative Rechtssubjektivität. In: Zeitschrift für Rechtssoziologie 2017, S. 8–25.

- L Louisa Specht/Sophie Herold.** Roboter als Vertragspartner? Gedanken zu Vertragsabschlüssen unter Einbeziehung automatisiert und autonom agierender Systeme. In: MultiMedia und Recht, 2018, S. 40–44.
- M Matthias, Andreas.** Automaten als Träger von Rechten: Plädoyer für eine Gesetzesänderung. – Berlin: Logos-Verlag, 2008. – 270 S.
- Mayinger, Samatha Maria.** Die künstliche Person. – Frankfurt am Main: Fachmedien Recht und Wirtschaft, 2016. – 260 S.
- Müller-Hengstenberg, Claus/Kirn, Stefan.** Rechtliche Risiken autonomer und vernetzter Systeme. – Oldenbourg: De Gruyter, 2016. – 372 S.
- S Schirmer, Jan-Erik.** Rechtsfähige Roboter? In: Juristenzeitung 2016, S. 660–666.
- Spindler, Gerald.** Zivilrechtliche Fragen beim Einsatz von Robotern. In: Hilgendorf, Eric / Beck, Susanne (Hrsg.): Robotik im Kontext von Recht und Moral. Baden-Baden: Nomos Verlagsgesellschaft, 2014. S. 63–80.
- T Teubner, Gunther.** Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten. – Workingpaper (erscheint in Archiv für die civilistische Praxis, 2018). – 39 S. <https://www.jura.uni-frankfurt.de/69768539/TeubnerDigitale-RechtssubjekteAcP-18Dez17.pdf> [zuletzt abgerufen: 06.03.2018].
- W Wagner, Gerhard.** Produkthaftung für autonome Systeme. In: Archiv für die civilistische Praxis (217) 2017, S. 707–765.

Roboterethik

Catrin Misselhorn

- › Autonome Maschinen (Computer, Roboter) sind in einem eingeschränkten Sinne moralisch relevante Akteure. Entscheidungen können an sie delegiert werden, moralische Verantwortung tragen sie jedoch nicht. Dazu fehlen ihnen Willensfreiheit, Kausalität, Absichtlichkeit und Wissen.
- › In bestimmten Situationen können autonome Systeme die Verantwortungszuschreibung unterminieren. Es wird unklar, wer Verantwortung trägt. Es entsteht eine „Verantwortungslücke“, in der weder Programmierern noch Anwendern die volle Verantwortung zugeschrieben werden kann.
- › Besonders wenn es um menschliches Leben geht, sollten Entscheidungen nicht Maschinen überlassen werden. Darüber hinaus gibt es viele nützliche Einsatzmöglichkeiten für Roboter.
- › Für die Roboterethik lassen sich drei Leitlinien festhalten:
(1) Künstliche Systeme sollten stets die Selbstbestimmung von Menschen fördern und nicht beeinträchtigen. (2) Sie sollten nicht über Leben und Tod von Menschen entscheiden. (3) Es muss sichergestellt werden, dass Menschen stets die Kontrolle ausüben und Verantwortung übernehmen.

Die Robotik ist eine vielversprechende Wachstumsbranche. So lag der Zuwachs beim Verkauf von Robotern nach Angaben des *World Robotics Report 2018* im vergangenen Jahr bei ca. 30 Prozent im Vergleich zum Vorjahr, und der Trend ist steigend.¹ Das bringt neben technischen, wirtschaftlichen und rechtlichen Herausforderungen auch ethische Probleme mit sich. Damit beschäftigt sich die Roboterethik, eine neue Disziplin an der Schnittstelle von Informatik, Philosophie und Robotik.² Sie hat die moralischen Probleme bei der Entwicklung, Herstellung und Verwendung von Robotern sowie das Verhältnis von Mensch und Roboter zum Gegenstand und betrachtet die gesellschaftlichen Konsequenzen der zunehmenden Roboterisierung.

Unter einem Roboter wird in diesem Beitrag eine elektromechanische Maschine verstanden, die aus einem Prozessor, Sensoren und Effektoren besteht.³ Der Begriff der Ethik wird gleichbedeutend mit Moralphilosophie gebraucht, im Sinn einer philosophischen Disziplin, die sich mit der moralischen Qualität von Handlungen, Urteilen, Charakterzügen und Einstellungen, Regeln sowie Institutionen befasst. Dabei steht die *normative* Ethik im Vordergrund, der es nicht nur um die Beschreibung dessen geht, was Menschen für moralisch richtig halten, sondern darum, begründete Empfehlungen abzugeben, was moralisch richtig und falsch ist.⁴

1. Einordnung der Roboterethik

Die spezifischen ethischen Fragen, die Roboter aufwerfen, haben mit zwei Aspekten zu tun: zum einen mit ihrer zunehmenden Intelligenz und Autonomie und zum anderen mit ihrer äußeren Gestalt und Interaktionsweise mit Menschen. Der erste Aspekt führt zur *Maschinenethik*,

-
- 1 <https://ifr.org/ifr-press-releases/news/industrial-robot-sales-increase-worldwide-by-29-percent> (letzter Aufruf: 20.12.18).
 - 2 Der Begriff der Roboterethik geht zurück auf Veruggio (2005), der als offizieller Begründer dieser Disziplin gilt
 - 3 Vgl. Misselhorn (2013).
 - 4 Für eine nähere Erläuterung, was unter moralisch in diesem Zusammenhang zu verstehen ist, vgl. Misselhorn (2018).

einer Subdisziplin der Roboterethik. Hier geht es um die Entwicklung einer Ethik für Maschinen im Gegensatz zu einer *Ethik für Menschen im Umgang mit Maschinen*. Man spricht in Analogie zu *Artificial Intelligence* (AI) auch von *Artificial Morality*.⁵ Während *Artificial Intelligence* zum Ziel hat, die kognitiven Fähigkeiten von Menschen zu modellieren oder zu simulieren, geht es bei der *Artificial Morality* darum, künstliche Systeme mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten. Das ist natürlich nicht beliebigen Maschinen zuzutrauen, sondern nur Computern. Die Idee ist, Computer so zu programmieren, dass sie moralisch handeln können. Die Software bildet dabei so etwas wie das „Gehirn“ des Roboters. Die Sensoren entsprechen den Sinnesorganen, die Informationen über die Umwelt und den Zustand des Roboters bereitstellen. Mit Hilfe der Effektoren kann der Roboter auf die Umwelt einwirken und seine Position verändern.

Der zweite Aspekt ergibt sich eher aus der äußeren Gestalt von Robotern und ihrer Interaktion mit Menschen. Man kann auch von *Ethik der Mensch-Maschine-Interaktion* sprechen. Menschen neigen dazu, intelligente Systeme zu vermenschlichen. Wir nehmen Roboter, die sich anscheinend autonom und intelligent verhalten und womöglich eine menschen- oder tierähnliche Gestalt aufweisen, allzu leicht als Wesen mit vergleichbaren Gedanken, Motiven und Empfindungen wie Menschen wahr. Damit sind verschiedene Fragen verbunden.

Zum einen geht es etwa darum, ob der Interaktion mit solchen Robotern eine problematische Form der Täuschung oder Manipulation zugrunde liegt, weil sie uns so anmuten, als hätten sie menschliche Eigenschaften, die ihnen gar nicht zukommen.⁶ Zum anderen ist zu fragen, ob Maschinen, die solche Reaktionen bei Menschen hervorrufen, uns auch gewisse moralische Einschränkungen im Umgang mit ihnen auferlegen, auch wenn sie nicht wirklich über die entsprechenden Gedanken, Motive oder Empfindungen verfügen.⁷ Maschinenethik

5 Vgl. Misselhorn (2018): *Artificial Morality*.

6 Vgl. Scholz (2008) spricht von „subjektsimulierenden Maschinen“, während Turkle (2006) die Roboter als relationale Artefakte definiert.

7 Misselhorn 2009, 2018.

und Ethik der Mensch-Maschine-Interaktion überlappen sich, wenn es darum geht, ob Maschinen, die die Fähigkeit zum moralischen Entscheiden und Handeln haben, ihrerseits auch Subjekt moralischer Ansprüche sind, die Menschen im Umgang mit ihnen zu berücksichtigen haben. Der vorliegende Beitrag konzentriert sich auf den Aspekt der Maschinenethik.⁸ Diese wird in einem offenen Brief, den eine beeindruckende Zahl von KI-Forschern und Wissenschaftlern unterzeichnet hat, als eines der wichtigsten und drängendsten Forschungsgebiete der KI hervorgehoben.⁹

2. Motivation der Maschinenethik

Es gibt viele Arbeiten, die wir gerne Maschinen überlassen würden, weil sie zu schwer, gefährlich oder einfach unangenehm sind. In manchen Fällen gibt es auch nicht genügend Menschen, die sie übernehmen können und wollen. Schließlich können Maschinen manche Dinge auch einfach besser und schneller erledigen als Menschen. In vielen Fällen erfordert das, dass die Maschinen möglichst eigenständig agieren, d. h. ohne dass ein Mensch direkt kausal eingreift. So stellt ein Roboterstaubsauger gerade deshalb eine Arbeitserleichterung dar, weil er nicht von einem Menschen geführt werden muss, sondern selbstständig in der Wohnung herumfährt, wenn man gerade nicht vor Ort ist. Diese Fähigkeit, sich selbstständig zu orientieren und zu agieren erfordert intelligente Systeme.¹⁰

Selbst ein so einfaches Modell wie der Staubsaugerroboter steht nun bereits einer ethischen Entscheidung gegenüber, nämlich: Soll es einen Marienkäfer einfach einsaugen, oder soll er den Käfer verscheuchen bzw. umfahren? Und wie sieht es bei einer Spinne aus? Bei der Frage, ob man ein Insekt zu Reinigungszwecken töten darf, handelt es sich um eine basale moralische Frage. Gewöhnliche Staubsaugerroboter besitzen allerdings noch nicht die Fähigkeit, eine solche Entscheidung

8 Zum zweiten Aspekt vgl. Misselhorn et al. (2013).

9 <https://futureoflife.org/ai-open-letter> (letzter Aufruf: 20.12.18).

10 Misselhorn (2015).

zu treffen. Es gibt jedoch erste Ansätze, eine um ein Ethikmodul erweiterte Version des populären Modells *Roomba* zu schaffen, die das Leben von Insekten berücksichtigt (der Prototyp ist mit einem optionalen „Kill-Button“ für Spinnen ausgestattet).¹¹

Je komplexer die Einsatzbereiche autonomer Systeme sind, desto anspruchsvoller werden die moralischen Entscheidungen, die sie treffen müssen. Ein wichtiger Anwendungsbereich für moralische Maschinen ist etwa die Altenpflege. Aufgrund des demographischen Wandels wird der Anteil pflegebedürftiger Menschen in den nächsten Jahrzehnten stark zunehmen. Künstliche Systeme werden immer wieder als eine Möglichkeit ins Spiel gebracht, um dem Pflegenotstand entgegenzutreten. Doch Systeme, die in diesem Kontext eingesetzt werden sollen, stehen vor moralischen Entscheidungen, beispielsweise: Wie häufig und eindringlich soll ein Pflegesystem an Essen und Trinken sowie die Einnahme von Medikamenten erinnern? Wann sollte ein Pflegesystem die Angehörigen verständigen oder den medizinischen Dienst rufen, wenn jemand sich eine Zeitlang nicht rührt? Soll das System den Nutzer rund um die Uhr überwachen, und wie ist mit den dabei erhobenen Daten zu verfahren?

In all diesen Situationen muss ein künstliches System zwischen bestimmten moralischen Werten abwägen: Im ersten Fall beispielsweise zwischen der Selbstbestimmung des Nutzers und bestimmten gesundheitlichen Risiken, die entstehen, wenn er seine Medikamente nicht wie vorgeschrieben einnimmt. Im zweiten Fall zwischen der Selbstbestimmung des Nutzers, der Sorge der Angehörigen, die vielleicht gerne sofort informiert würden, und erneut der Gesundheit. Im dritten Fall geht es wiederum um die Selbstbestimmung des Nutzers, Gesundheit, Sorge der Angehörigen sowie um die Privatheit seiner Daten.

11 Vgl. Bendel (2017).

Ein weiteres viel diskutiertes Beispiel für die Notwendigkeit moralischer Maschinen ist das autonome Fahren. Auch vollautomatisierte Fahrzeuge stehen vor moralischen Entscheidungen. So gilt es beispielsweise, sie so zu programmieren, dass in unvermeidlichen Gefahrensituationen der Schutz menschlichen Lebens Vorrang vor Sach- und Tierschäden besitzt. Doch auch Tiere sollten nach Möglichkeit verschont werden. Eine besondere Schwierigkeit stellen die in diesem Anwendungsbereich unter Umständen auftretenden moralischen Dilemmata dar. Was ist, wenn ein autonomes Fahrzeug ausschließlich die Möglichkeit hat, entweder einen Menschen am Ende seines Lebens oder ein kleines Kind zu töten? Was, wenn es nur dadurch fünf Menschenleben retten kann, dass es eine auf dem Gehweg stehende Person anfährt? Ist ein besonderer Schutz für die Insassen moralisch legitim, oder kommt den anderen Verkehrsteilnehmern vom moralischen Standpunkt mehr Gewicht zu?¹²

Nicht zu vergessen sind schließlich die militärischen Anwendungen. Der Traum besteht darin, dass keine Soldaten mehr auf dem Schlachtfeld ihr Leben aufs Spiel setzen müssen, sondern an ihrer Stelle autonome Maschinen in den Kampf geschickt werden.¹³ Diese sollen mit dem Kriegsvölkerrecht und kontextspezifischen Einsatzregeln ausgestattet werden, die ihren Handlungsspielraum begrenzen und sicherstellen, dass sie sich rechtlich und moralisch einwandfrei verhalten. So müssen sie entscheiden, wann eine Aktion militärisch notwendig und angemessen ist und wie sich Kombattanten von Zivilisten unterscheiden lassen.

12 Auf der Webseite <http://moralmachine.mit.edu/> (letzter Aufruf: 20.12.18) finden sich eine ganze Reihe solcher Szenarien mit unterschiedlichen Konstellationen, über die der Besucher per Mausclick entscheiden kann. Die Daten werden u. a. genutzt, um ein Entscheidungsverfahren für moralische Maschinen zu entwickeln (Awad et al. 2018).

13 Vgl. Arkin (2009).

3. Moralisches Handeln und Verantwortung

Man könnte allerdings argumentieren, dass es nicht das Pflegesystem, das autonome Auto oder der Kampfroboter ist, die in diesen Fällen eine moralische Entscheidung treffen, sondern die Programmierer dieser Geräte. Doch je größer die Fortschritte der Künstlichen Intelligenz werden, desto weniger können die Entwickler planen und vorher-sagen, welche Entscheidungen ein System in einer spezifischen Situation treffen wird. So spielt schon ein Schachprogramm weit besser als seine Programmierer, die nicht jeden einzelnen Zug des Systems vorhersagen können. Das gilt umso mehr für ein so komplexes System wie Alpha Go Zero, das zunächst nur die Grundregeln des Spiels kennt und dann durch das Durchspielen einer Vielzahl von Partien gegen sich selbst zu den optimalen Entscheidungsstrategien findet. In kürzester Zeit gelang es diesem System, seinen Vorgänger Alpha Go zu schlagen, der als erstes künstliches System einige der weltbesten menschlichen Go-Spieler besiegte. Doch genau dieser Mangel an Kontrolle und Vorhersehbarkeit gehört zu den Punkten, die den wichtigsten ethischen Einwänden gegen die Maschinenethik zugrunde liegen.

Maschinen sind nicht zu vollumfänglichem moralischen Handeln in der Lage, wie es Menschen auszeichnet. Dazu fehlen ihnen unter anderem Bewusstsein, Willensfreiheit und die Fähigkeit zur Selbst-reflexion.¹⁴ Da diese Eigenschaften für die Übernahme moralischer Verantwortung wesentlich sind, können Maschinen auch nicht für ihr Handeln verantwortlich gemacht werden. Trotzdem ist zu diskutieren, inwiefern der Einsatz von Maschinen die Verantwortungszuschreibung an Menschen unterminiert, so dass am Ende möglicherweise niemand für ihr Handeln die Verantwortung trägt. In diesem Zusammenhang ist vom Entstehen einer *Verantwortungslücke* die Rede.¹⁵

Zu den Kriterien für die Zuschreibung moralischer Verantwortung zählen Willensfreiheit, Kausalität, Absichtlichkeit und Wissen. Ein Handelnder ist demnach nur dann für eine Handlung verantwortlich, wenn

14 Vgl. Misselhorn (2018).

15 Vgl. Sparrow (2007).

sie auf seinem freien Willen beruht, wenn sie ohne seine Beteiligung nicht zustande gekommen wäre, er sie absichtlich durchgeführt hat (oder ihre Folgen zumindest in Kauf genommen hat) und ihm ihre Folgen bekannt waren (er diese hätte vorhersehen können oder sich die entsprechenden Kenntnisse mit vertretbarem Aufwand beschaffen können). Es ist unumstritten, dass Maschinen diese Bedingungen nicht alle erfüllen. So besitzen sie keinen freien Willen; aber auch die Bedingungen der Absichtlichkeit und des Wissens werfen Probleme bei der Zuschreibung an Maschinen auf. Deshalb können sie zwar keine moralische Verantwortung tragen, aber eine Verantwortungslücke erzeugen.

Der australische Maschinenethiker Robert Sparrow, auf den dieser Begriff (im Original: *responsibility gap*) zurückgeht, argumentiert dafür am Beispiel autonomer Kriegerroboter. Der Kern seines Arguments lautet: Eine Verantwortungslücke entsteht, wenn (1) ein Kriegerroboter nicht absichtlich so programmiert wurde, dass er die ethischen bzw. rechtlichen Normen der Kriegsführung verletzt; (2) es nicht vorhersehbar war, dass der Einsatz des Kriegerroboters dazu führen würde; und (3) ab dem Start der Operation keine menschliche Kontrolle mehr über die Maschine bestand.

Das Problem ist, dass das Vorliegen dieser drei Bedingungen dazu führt, dass die moralische Verantwortung keinem Menschen zugeschrieben werden kann, wenn die Maschine im Widerstreit mit den ethischen bzw. rechtlichen Normen der Kriegsführung Menschen tötet. Denn kein Mensch hatte dies beabsichtigt, es war nicht vorhersehbar und niemand hatte kausal die Möglichkeit, dieses Resultat zu verhindern. Eine Verantwortungslücke entsteht also genau dann, wenn die Maschine selbst nicht verantwortlich ist, aber ihr Einsatz die Bedingungen der Verantwortungszuschreibung an Menschen untergräbt. Für Sparrow ist dies ein Grund dafür, den Einsatz von Kriegerrobotern als unmoralisch abzulehnen. Es lässt sich aber grundsätzlich auch auf andere Bereiche, insbesondere das autonome Fahren, übertragen.

Man könnte dies zum Anlass nehmen zu fordern, der Mensch dürfe eben nicht gänzlich die Kontrolle aus der Hand geben. In militärischen Kontexten wird zwischen *In-the-Loop*-Systemen, *On-the-Loop*-Systemen und *Out-of-the-Loop*-Systemen unterschieden, je nachdem, welche Rolle der Mensch in der Kontrollschleife spielt.¹⁶ Bei *In-the-Loop*-Systemen bedient ein Mensch das System und fällt sämtliche Entscheidungen, und sei es per Fernbedienung. *On-the-Loop*-Systeme sind zwar programmiert, sie können aber in Echtzeit unabhängig von menschlichem Eingreifen operieren. Dem Menschen obliegt jedoch weiterhin die Überwachung, und er hat jederzeit die Möglichkeit, zu intervenieren. *Out-of-the-Loop*-Systeme verhalten sich wie *On-the-Loop*-Systeme, doch besteht keine menschliche Kontroll- und Eingriffsmöglichkeit mehr.

Das Problem der Verantwortungslücke erscheint als gelöst, wenn der Mensch *On-the-Loop* bleibt und vielleicht sogar per Knopfdruck der Verantwortungsübernahme zustimmen muss, bevor er ein künstliches System in Betrieb nimmt.¹⁷ Doch wie realistisch ist die Annahme, dass der Mensch zu einer permanenten Überwachung überhaupt in der Lage ist? Kann er die Aufmerksamkeit so lange halten, und ist er in Sekundenschnelle bereit, zu entscheiden und einzugreifen, wenn es darauf ankommt? Wenn das nicht der Fall ist, wären Vorhersehbarkeit und Kontrolle zwar theoretisch möglich, aber in der Realität für den Menschen nicht umsetzbar.

Außerdem stellt sich ein Erkenntnisproblem, denn der Mensch ist zur Analyse der Situation auf die Informationen angewiesen, die das System ihm liefert. Die Frage ist, ob er diese überhaupt rational in Zweifel ziehen kann, wenn er keinen Zugang zu unabhängigen Informationen besitzt. Darüber hinaus muss ein solches System bei seiner Entwicklung eine Reihe von Qualitätssicherungsprozessen durchlaufen. Auch das kann für die Nutzer ein Grund dafür sein, die Vorschläge des Systems den eigenen Zweifeln für überlegen zu halten.

16 Vgl. United States Department of Defense (2011).

17 Eine solche Konzeption schlägt beispielsweise Arkin (2009) für das von ihm entwickelte autonome Waffensystem vor.

Insgesamt wirkt es unfair, dass die Nutzer die volle Verantwortung per Knopfdruck übernehmen sollen, denn zumindest ein Teil der Verantwortung, wenn nicht sogar der Hauptteil, sollte doch den Programmierern zukommen, deren Algorithmen ausschlaggebend für das Handeln des Systems sind. Die Nutzer sind nur in einem schwächeren Sinn verantwortlich, weil sie das System nicht am Handeln gehindert haben. Alle drei Punkte lassen es zweifelhaft erscheinen, ob die Bedingungen der Vorhersehbarkeit und der Kontrolle erfüllt sind. Das Problem der Verantwortungslücke droht deshalb auch bei *On-the-Loop*-Systemen. Es stellt sich letztlich, sogar dann, wenn der Mensch *In-the-Loop* bleibt. Dies unterstützt die Forderung von Menschenrechtsorganisationen nach einer sinnvollen menschlichen Kontrolle (*meaningful control*).¹⁸

4. Entscheidungen über Leben und Tod

Generell sollten wir uns gut überlegen, ob wir moralisches Entscheiden an Maschinen delegieren wollen, wenn es um Leben und Tod von Menschen geht. Ein entscheidendes Argument gegen autonome Waffensysteme lautet, dass es keine moralische Pflicht gibt, im Krieg zu töten.¹⁹ Es liegt nur eine moralische Erlaubnis zum Töten vor, die das allgemeine Tötungsverbot situativ einklammert. Deshalb sollte die Entscheidung, einen bestimmten Menschen zu töten, stets einem Menschen obliegen und nicht von einer Maschine getroffen werden.

Dieses Argument lässt sich auch auf das autonome Fahren übertragen. Es lässt sich eine Analogie zwischen der Programmierung autonomer Fahrzeuge zum Zweck der Unfalloptimierung und der Zielbestimmung autonomer Waffensysteme herstellen.²⁰ Um Unfallergebnisse zu optimieren, ist es notwendig, Kosten-Funktionen anzugeben, die bestimmen, wer im Zweifelsfall verletzt und getötet wird. Ganz ähnlich wie bei autonomen Waffensystemen müssten also für den Fall einer unvermeidlichen Kollision legitime Ziele festgelegt werden.

18 Vgl. Roff und Moyes (2016).

19 Misselhorn (2018).

20 Vgl. Lin (2016), S. 72.

Das setzt voraus, dass eine moralische Pflicht besteht, unschuldige Menschen zu verletzen oder zu töten, sofern dies dazu dient, Schlimmeres zu verhindern. Eine solche Pflicht stünde freilich in einem Spannungsverhältnis zur deutschen Rechtsprechung. So hat das Bundesverfassungsgericht in seiner Entscheidung zum Luftsicherheitsgesetz im Jahr 2006 zum Abschuss entführter Passagierflugzeuge, die von Terroristen als Massenvernichtungswaffen eingesetzt werden sollen, festgestellt, dass ein Abschuss immer der Menschenwürde der Flugzeugpassagiere widerspricht.²¹ Das Grundgesetz schließt aus, auf der Grundlage einer gesetzlichen Ermächtigung unschuldige Menschen vorsätzlich zu töten. Dieses Urteil steht zumindest auf den ersten Blick in einem Widerspruch zu einer Pflicht der Schadensminimierung, die die vorsätzliche Verletzung oder Tötung unschuldiger Menschen umfasst.

Es gibt allerdings Stimmen, die in dieser Sache noch nicht das letzte Wort gesprochen sehen. Ein Vorschlag besteht darin, von einer Abstufung im Unrecht auszugehen.²² Demnach ist die Tötung unschuldiger Menschen rechtswidrig (und wohl auch unmoralisch), gleichwohl soll es eine rechtliche und moralische Pflicht geben, so wenige Leben wie möglich zu vernichten, und zwar auch dann, wenn dabei vorsätzlich unschuldige Menschen verletzt oder getötet werden. Allerdings scheint eine solche Pflicht bislang in der Rechtsdogmatik nicht allgemein anerkannt zu sein. Zudem ist diese Konstruktion aus moralischer Sicht fragwürdig. Weil die Verletzung und Tötung unschuldiger Menschen weiterhin ein Unrecht darstellt, haben die Betroffenen nämlich Abwehrrechte. So dürfen sie ein Flugzeug abschießen, das auf sie niederzugeschossen droht, wenn diese Option das geringste Übel darstellt. Das würde aber bedeuten, dass jemand eine moralische Pflicht hat, etwas zu tun, während ein anderer das moralische Recht hat, genau dies zu unterbinden.²³ Das scheint mit der Unbedingtheit und Verallgemeinerbarkeit, die moralische Pflichten auszeichnen, schlecht vereinbar zu sein.

21 Vgl. BVerfGE 115, 118, (160).

22 Vgl. Hilgendorf (2017), S. 155.

23 Vgl. Hilgendorf (2017), S. 155.

5. Konklusion

Auch wenn man die Entscheidung über Leben und Tod von Menschen nicht an Maschinen abgeben möchte, bleiben viele Anwendungsbereiche, in denen moralische Maschinen sinnvoll eingesetzt werden können, beispielsweise in der Pflege.²⁴ Dabei muss darauf geachtet werden, dass das menschliche Recht auf Selbstbestimmung gewahrt bleibt. Das betrifft einerseits die Entscheidung, ob jemand überhaupt von einem künstlichen System gepflegt werden möchte; dies sollte jedem freigestellt sein. Andererseits ist darauf zu achten, dass ein Pflegesystem sich flexibel auf die Wertvorstellungen seiner Nutzer einstellen kann. Denn in modernen pluralistischen Gesellschaften ist davon auszugehen, dass sich die Wertvorstellungen der Nutzer darin unterscheiden, ob beispielsweise der Privatsphäre mehr Gewicht zuzumessen ist oder der Vermeidung gesundheitlicher Risiken. Ein Pflegesystem sollte in der Lage sein, diese Abwägung nach den moralischen Standards des jeweiligen Nutzers individuell vorzunehmen. In diesem Fall kann ein Pflegesystem Menschen, die dies möchten, dazu verhelfen, länger selbstbestimmt in ihren eigenen vier Wänden zu leben. Ein solches System ist allerdings nur für Menschen geeignet, die kognitiv in der Lage sind, grundlegende Entscheidungen über ihr Leben zu treffen, aber körperlich so eingeschränkt sind, dass sie nicht ohne Pflege allein zu Hause leben können.

Abschließend kann man drei Prinzipien als grundlegende Leitlinien für die Roboterethik festhalten:

1. Künstliche Systeme sollten stets die Selbstbestimmung von Menschen fördern und sie nicht beeinträchtigen.
2. Sie sollten nicht über Leben und Tod von Menschen entscheiden.
3. Es muss sichergestellt werden, dass Menschen stets die Kontrolle ausüben und Verantwortung für das Handeln der Maschinen übernehmen.

24 Vgl. Misselhorn (2019): Moralische Maschinen in der Pflege.

Referenzen

A Arkin, Ronald (2009). Governing lethal behavior in autonomous robots. Boca Raton.

Awad, Edmond et al. (2018). The Moral Machine experiment. In: *Nature* 563, S. 59–64.

B Bendel, Oliver (2017). Ladybird – The animal-friendly robot vacuum cleaner. In: *The AAAI 2017 Spring Symposium on Artificial Intelligence for the Social Good Technical Report SS-17-01*. Palo Alto, S. 2–6.

H Hilgendorf, Eric (2017). Autonomes Fahren im Dilemma – Überlegungen zur moralischen und rechtlichen Behandlung von selbsttätigen Kollisionsvermeidungssystemen. In: *Eric Hilgendorf: Autonome Systeme und neue Mobilität – Ausgewählte Beiträge zur 3. und 4. Würzburger Tagung zum Technikrecht*. Baden-Baden, S. 143–176.

L Lin, Patrick (2016). Why ethics matters for autonomous cars. In: *Markus Maurer und J. Christian Gerdes und Barbara Lenz et al. (Hrsg.): Autonomous driving – Technical, legal and social aspects*. Berlin/Heidelberg, S. 69–85.

M Misselhorn, Catrin (2009). Empathy with inanimate objects and the uncanny valley. In: *Minds and Machines* 19, S. 345–359.

Misselhorn, Catrin (2013). Robots as moral agents. In: *Frank Rövekamp und Friederike Bosse (Hrsg.): Ethics in science and society – German and Japanese views*. München, S. 30–42.

Misselhorn, Catrin (Hrsg.) (2015). Collective agency and cooperation in natural and artificial systems – Explanation, implementation and simulation. In: *Philosophical Studies Series* 122.

Misselhorn, Catrin (2018). *Grundfragen der Maschinenethik*. Ditzingen 2018.

- Misselhorn, Catrin** (2019). Is empathy with robots morally relevant? In: Catrin Misselhorn und Maïke Klein (Hrsg): Emotional machines – Perspectives from affective computing and emotional human-machine interaction. Wiesbaden. (Erscheint 2019).
- Misselhorn, Catrin und Pompe, Ulrike und Stapleton, Mog** (2013). Ethical considerations regarding the use of social robots in the fourth age. In: GeroPsych – The Journal of Gerontopsychology and Geriatric Psychiatry 26, S. 121–133.
- R Roff, Heather M. und Moyes, Richard** (2016). Meaningful human control, artificial intelligence and autonomous weapons. Briefing paper prepared for the informal meeting of experts on lethal autonomous weapons systems. UN Convention on Certain Conventional Weapons. Geneva.
- S Scholz, Christopher** (2008). Alltag mit künstlichen Wesen. Theologische Implikationen eines Lebens mit subjektssimulierenden Maschinen am Beispiel des Unterhaltungsroboter Aibo Göttingen 2008
- Sparrow, Robert** (2007). Killer robots. In: Journal of Applied Philosophy 24, S. 62–77.
- T Turkle, S., Taggart, W., Kidd, C., & Daste, O.** (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. Connection Science, 18, 347–361.
- U United States Department of Defense** (2011). Unmanned systems integrated roadmap FY 2011-2036. Reference Number 11-S-3613. URL: <<http://www.dtic.mil/dtic/tr/fulltext/u2/a558615.pdf>> (letzter Aufruf: 20.12.18).
- V Veruggio, Gianmarco** (2005). The birth of roboethics. In: Proceedings of IEEE International Conference on robotics and automation. Genova, S. 1–4.

Zwischen System und Verantwortung

Ethische Überlegungen zum Begriff der Autonomie angesichts der Entwicklung autonomer Systeme

Elisabeth Gräß-Schmidt

- › In der philosophischen Tradition steht die Autonomie des freiheitlichen Subjekts in einer Wechselbeziehung von Freiheit und Bindung an Gesetze. Autonomie bedeutet nicht Unabhängigkeit, sondern versteht sich als Selbstbegrenzung aus Freiheit, die Ordnung durch Gesetze anerkennt.
- › Mit der Entwicklung autonomer Systeme und ihrer Überlegenheit gegenüber manchen menschlichen Fähigkeiten verschiebt sich das Verständnis der Autonomie von der Selbstgesetzgebung des Menschen auf eine Selbststeuerung technologischer Systeme.
- › Wenn im Verhältnis von Technik und Ethik das Prinzip der Funktion zum Maßstab gemacht wird, entspringt das Handeln nicht mehr einer verantwortlichen Entscheidung, sondern der Steuerung des Systems.
- › Eine Maschinenethik, die die Verantwortung an die Maschine abgäbe, ist auszuschließen, weil die Freiheit der Verantwortung nicht maschinell herstellbar ist.
- › Ethische Überlegungen müssen vorausschauend technische Entwicklungen begleiten und dem Menschen in ihrer Gestaltung die frei verantwortete Entscheidung zuweisen.

1. Begriff, Gegenstand und Reichweite der autonomen Systeme

Unter dem Begriff des autonomen Systems verbirgt sich ein nahezu unüberschaubares Feld neuer technologischer Anwendungen. Die Feuilletons und Wissenschaftsteile der großen Zeitungen informieren nahezu täglich über neue Entwicklungen und die Chancen solcher Technologien, aber auch über ihre Gefahren. Diese liegen nicht nur in der möglichen Verdrängung der Arbeitskraft des Menschen, sondern in ihrer sämtliche Arbeits-, aber auch Lebensbereiche bestimmenden Macht. Sie bieten eine revolutionierende Entfaltungsmöglichkeit im Bereich der gesamten Kultur, vom Internet der Dinge über selbstfahrende Autos bis hin zum Einsatz von Robotern in der Medizin und im Pflegebereich. Aufgrund ihrer Potentiale sind diese Technologien daher konstruktiv, aber auch kritisch aufzunehmen. Die Konsequenzen für das Verständnis des Menschen, aber auch für Wirtschaft, Politik und Recht und Ethik müssen angesichts solcher Entwicklungen bedacht und gesamtgesellschaftlich diskutiert werden.

Zu denken wäre bei solchen Entwicklungen etwa an die Big Data-Technologie oder an Assistenzroboter in der Chirurgie – zur Optimierung von Operationsprozessen durch computergestützte technische Assistenz – oder an eine Verbesserung medizinischer Diagnosen. Zu denken ist auch an Verbesserungen in der Pflege im Krankenhaus und in Altenpflegeheimen, wenn durch automatisierte Technik der Alltag nicht nur der Pflegenden, sondern auch der Kranken besser bewältigt werden kann.¹ Solche technisch-intelligenten Hilfssysteme, die im modernen Roboter auch menschliche Hilfsdienste begleiten und teilweise ersetzen können, bedeuten einen Rückgewinn von Lebensqualität im Alter.²

1 Vgl. Karl-Heinz Pantke (Hg.), *Mensch und Maschine: Wie Brain-Computer-Interfaces und andere Innovationen gelähmten Menschen kommunizieren helfen*, Frankfurt 2010.

2 Lucienne Rey, *Unser Freund der Roboter*, Zentrum für Technologiefolgen-Abschätzung (Hg.), Bern 2012. https://www.ta-swiss.ch/RoboComFactsheet_d.pdf (letzter Aufruf 06.09.2018).

Eine weitere Leistungsmöglichkeit dieser Systeme ist die Entwicklung Künstlicher Intelligenz in Konkurrenz zum Menschen. So kann mithilfe künstlicher neuronaler Netze eine Künstliche Intelligenz das Denken des Menschen nachahmen und gegebenenfalls überbieten. Man spricht dabei von sich von menschlicher Intelligenz emanzipierenden, smarten Systemen³, die zur größeren Effizienz und Verlässlichkeit menschlichen Handelns beitragen können. Solche *Deep-learning*-Systeme sind selbstlernend, d. h. sie werden mithilfe von umfangreichen Datensätzen nicht mehr nur programmiert, sondern trainiert. Das führt zur Eigenständigkeit von Lernprozessen und umfasst nicht nur kognitive, sondern auch sensomotorische, emotionale und soziale Intelligenz. Während auf kognitiver Ebene die Künstliche Intelligenz weit fortgeschritten ist, ist sie jedoch auf der sensomotorischen Ebene von der menschlichen Leistung noch weit entfernt.

Die mit der Digitalisierung einhergehende Transformation kultureller Selbstverständlichkeiten sind daher vielfältig. Sie bergen Chancen, aber auch Gefahren. So sind mögliche Gefahren zu bedenken, wenn es um die Entwicklung Künstlicher Intelligenz sogenannter smarterer Systeme in der Kriegsführung geht. Aber auch die ethischen und rechtlichen Bedingungen der Verantwortung und Haftung müssen geklärt werden. Wer ist haftbar zu machen für Fehlentscheidungen? Liegt die Verantwortung weiterhin beim Menschen, oder kann sie an ein autonomes System abgegeben werden? Kann von einer Zurechenbarkeit autonomer Systeme ausgegangen werden, etwa bei Handlungen intelligenter Kriegssoftware oder auch bei autonom fahrenden Autos?⁴

Zu klären ist daher zunächst der Begriff der Autonomie. Lässt er sich auf ein System oder auf Formen Künstlicher Intelligenz übertragen? Erst dann kann der Anspruch und die Reichweite von sogenannten autonomen Systemen hinsichtlich ihrer Autonomie bestimmt wer-

3 Vgl. Robin Hanson, *The Age of Em: Work, Love and Life when Robots Rule the Earth*, Oxford 2016.

4 Vgl. Susanne Beck (Hg.) *Jenseits von Mensch und Maschine: ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Baden Baden, 2012; Jan Schuhr, *Willensfreiheit, Roboter und Auswahlaxiom*, in: ebd., 43–75. Siehe auch Anmerkung 12

den. Dabei stellen sich weitere Fragen: Welche Konsequenzen ergeben sich aus den Autonomieansprüchen von Systemen bzw. Künstlicher Intelligenz für das Verständnis des Menschen und seiner Autonomie? Diese Frage berührt das Verhältnis von Technik und Ethik. Bedeutet die Entwicklung autonomer Systeme bzw. Künstlicher Intelligenz eine grundlegend neue Technik, und ist daher gar eine „Maschinenethik“ zu fordern, oder bleibt ethische Verantwortung an menschlich-personale Autonomie gebunden?

2. Was ist Autonomie angesichts der Entwicklung sogenannter autonomer Systeme?

Autonomie ist in der Moderne Ausweis der Vernunftnatur, der Selbstzwecklichkeit und inneren Freiheit des Menschen. Künstliche „Autonome Systeme“ – diese spannungsvolle begriffliche Zusammenstellung zweier sich auch widersprechender Konzeptionen, das der Autonomie und das des Systems –, stellen uns nun gegenwärtig vor große Herausforderungen. Sie betreffen nicht nur das Verhältnis von Technik und Ethik, sondern berühren die Frage nach dem Menschsein als solchem. Sah man die Selbstbestimmung des Menschen in antiker und christlicher Tradition vor dem Hintergrund einer inneren Freiheit, die es dem Menschen ermöglicht, selbst zu urteilen und Verantwortung zu übernehmen, so steht dieser Kern der Entscheidungsfähigkeit des Menschen durch die Übertragung des Begriffs auf ein System zur Disposition. Der Aspekt der Freiheit des Selbst ist entscheidend auch für den modernen Begriff der Autonomie. Selbstbestimmung als Entscheidungs- und Urteils kraft sind sein wesentliches Kennzeichen. Mit der Entscheidungsfähigkeit ist die Freiheit des Menschen angesprochen. Die Bestimmung eines Systems als „autonom“ wird sich daher an diesem Kriterium der Entscheidungs- oder Urteilsfähigkeit erweisen.

Autonomie bedeutet in der Tradition und in der Aufklärung Selbstbestimmung des Menschen als Auszeichnung menschlicher Freiheit. Der Ursprung des Begriffs der Autonomie in der griechischen Antike war mit der politischen Selbstbestimmung der Stadtstaaten verbunden. Autonomie (αὐτονομία), die Eigengesetzgebung, wurde als

äußere politische Freiheit an Stelle von Fremdherrschaft verstanden.⁵ Spätestens seit der Bedrohung der politischen Unabhängigkeit durch die Perserkriege trat der Gedanke der Autonomie auch im Sinne einer inneren Freiheit auf. Dieser Gedanke zeigte sich ansatzweise bereits bei Herodot in der Bestimmung von Autonomie als Aufgabe, den Gesetzen zu gehorchen, die für alle gleich sind.⁶ In dieser inneren Freiheit entspricht er nicht zuletzt dem paulinischen Verständnis einer Freiheit in Christus. Unabhängig von äußeren Zwängen ist jeder Mensch aufgrund seiner inneren Freiheit gleich und frei (Galater 3, 28) durch und in der Bindung an Jesus Christus als der inneren Bestimmung des (göttlichen) Gesetzes. Christlich ist Autonomie daher weniger als Selbstgesetzgebung denn als freie, selbstbestimmte Entsprechung gegenüber dem vorgegebenen Gesetz verstanden.

Explizit finden wir den Gedanken der Freiheit und Selbstzwecklichkeit als Ausdruck der Autonomie bei Aristoteles (Metaphysik 982b, 25 f.): „[E]in Mensch ist frei, wenn das Ziel des Handelns in ihm selbst liegt, wenn er um seiner selbst willen und nicht um eines anderen willen ist.“ Dieses Autonomieverständnis hat nachgewirkt und bei Thomas von Aquin sowie in der Neuzeit bei Spinoza im Gedanken der *causa sui* Ausdruck gefunden.

Die Bezogenheit des Selbst (αὐτός) im Begriff der Autonomie auf ein Gesetz (νόμος) setzt dieser dem Verständnis von Selbstbestimmung des Menschen als Willkür Grenzen. Dieser Bezug des Selbst zu einem ihm selbst vorgegebenen Gesetz ist mithin auch für das mündige, aufgeklärte, selbstbestimmte Subjekt konstitutiv. Autonomie ist daher nicht zu verwechseln mit bloßer Unabhängigkeit. Vielmehr bildet die Autonomie des freiheitlichen Subjekts eine Wechselbeziehung von Freiheit und Gesetz ab. Somit wohnt der Autonomie als Selbstbestimmung selbst die Autorität eines Anderen inne. Dem Selbst

5 S. dazu Herodot, *Historia* I, 96; vgl. Rosemarie Pohlmann, Autonomie, in: Ritter et al. (Hg.): *Historisches Wörterbuch der Philosophie* Bd. I, Darmstadt 1971, 701 ff.

6 (Herodot 7,103). Spätestens bei Sophokles wird mit der Figur der Antigone – allerdings kritisch – ein solcher auf das Individuum bezogene Begriff der Autonomie sichtbar. „Du lebst nach eigenen Gesetzen, drum allein ins Land der Toten gehst du.“ (Sophokles, *Antigone*, Vers 821).

(αὐτός) steht der Begriff des Gesetzes (νόμος) gegenüber. Diese Autorität des Gesetzes darf die Freiheit nicht missachten, will sie sich nicht selbst aufheben.

Dieser spannungsvolle Zusammenhang ist in Kants Begriff der Autonomie festgehalten, wenn er Freiheit und eigene Gesetzgebung in der Autonomie als „Wechselbegriffe“⁷ bezeichnet. Das Im-Blick-haben der Bezüge von Gesetz und Freiheit führt in der Beachtung der Gesetze des Zusammenhangs zur Selbstbegrenzung aus Freiheit. Dies ist der geniale Schachzug der Bestimmung von Freiheit als Autonomie bei Kant. Freiheit ist bei Kant eben nicht Willkürfreiheit und Beliebigkeit ihrer Ausdrucksformen, sondern Selbstbindung an ein Gesetz. Ein mögliches Missverständnis der Autonomie liegt daher in der Missachtung des Gesetzes als Bezug der Freiheit. Es entsteht aus dem Verkennen der Bedingungen der Freiheit, die als Ordnungen und Gesetze dem Leben dienen und es erhalten. Werden hingegen diese Bedingungen erkannt, führt dies zur Selbstbegrenzung aus Freiheit.

Vor dem Hintergrund dieser Bestimmung von Autonomie als Ausdruck innerer Freiheit der Selbstverpflichtung stellt sich nun verstärkt die Frage: Ist ein solcher Begriff auf intelligente Maschinen oder Systeme anwendbar? Ist es möglich, dass ein Verständnis von Autonomie als innere Freiheit der Selbstentsprechung zum Gesetz durch ein System bzw. eine Künstliche Intelligenz in Anspruch genommen wird? Oder wird damit nicht die Errungenschaft der Differenzierung von äußerer, materieller Bestimmung von Freiheit und innerer Freiheit im Sinne der Selbstzwecklichkeit eingegeben? Mit anderen Worten: Bricht sich nicht der Vergleich von Mensch und Maschine am Verständnis des Menschen in seiner Würde und Personalität, wie es dem antiken und dem modernen, an der Aufklärung orientierten Verständnis von Personsein zugrunde liegt?

7 Immanuel Kant, *Grundlegung zur Metaphysik der Sitten* in: Immanuel Kant, *Gesammelte Schriften* (= Akademieausgabe), Abt. I Bd. 4, Berlin 1911, 450. Vgl. dazu auch Elisabeth Gräß-Schmidt, *Autonomie*, in: ZEE 59 (2/2015), Gütersloh, 133–137.

3. Herausforderung des Verständnisses der Autonomie des Menschen durch autonome Systeme

Ein Verständnis des Menschen in seinem Geschaffensein entzieht sich einer Vorstellung des Menschen, die von seinem technischen Hergestelltsein ausgeht. Die Digitalisierung stellt uns jedoch mit ihrer Entwicklung autonomer Systeme auch vor Herausforderungen anthropologischer Natur. Die transhumanistische Idee der Verschmelzung von Mensch und Maschine, die über den Ansatz der Künstlichen Intelligenz hinausgeht, hat Folgen für die Bestimmung dessen, was als Menschsein wahrgenommen wird.

Können Maschinen Personen sein? Zukunftsforscher wie Michio Kaku und Ray Kurzweil denken, dass das möglich ist. Künstliche Intelligenz sei die „letzte Erfindung“, die wir als Menschen noch machen müssen, so Michio Kaku.⁸ Der Weg zur Künstlichen Intelligenz erweist sich nach ihm durch Schaffung eines Ich-Bewusstseins und eines Sinnesempfindens, die selbst Kreativität ermöglichen. Alle Fähigkeiten werden nachgemacht werden können. Wird so der Mensch durch Selbstoptimierung über sich hinaus geführt im Sinne des von Nietzsche propagierten Übermenschen oder sogar abgeschafft durch Übersteigerung des Menschen in einem Transhumanismus, so stellt sich die Frage, ob dann durch diese Technologie das Verständnis des Menschen und seiner Selbstbestimmung nicht verändert wird. Ändern sich mit der Künstlichen Intelligenz die Parameter Vernunft und Freiheit des Menschen?

Heraufbeschworen wird diese Transformation im Bild einer *Brave New World*, die Technik nicht mehr nur als Dienst am Menschen, sondern als Selbstinszenierung des Menschen bis hin zum technologisch frisierten Übermenschen avisiert. Dieser Überschritt des Menschen zur Künstlichen Intelligenz vollzieht sich im Namen jener sogenannten „Singularität“⁹, mit der der Übergang des Menschen zu einem selbst-

8 Vgl. Michio Kaku: Die Physik des Bewusstseins. Über die Zukunft des Geistes, Reinbek 2014.

9 Ray Kurzweil, Menschheit 2.0. Die Singularität naht, Berlin ²2014.

lernenden Intellekt erreicht ist, durch den man das autonome System ausgezeichnet sieht. Dies wäre ein Paradigmenwechsel, der schwer vorstellbar ist. Der von Yuval Harari prognostizierte *Homo Deus* löst den *Homo Faber* ab, ohne jedoch die bereits diesen treffende Kritik eines verengten Freiheits- und Autonomieverständnisses überwunden zu haben und überwinden zu können.

Die Künstliche Intelligenz in der Entwicklung autonomer Systeme wird so zu tief greifenden Veränderungen der Wahrnehmung des Menschseins führen können, die nicht erst das Menschsein in seiner natürlichen Ausstattung, wie in den gentechnischen Eingriffen, oder in der Gefährdung seiner Existenz, wie in der Kernenergie, berühren, sondern das Menschsein in seiner Bestimmung gefährden. Denn es geht jetzt nicht mehr nur um technische Errungenschaften, die wir in ihren Fortschritten als Zivilisation auch genießen können, sondern es geht jetzt um die technische Überformung der Natur des Menschen.

Dabei verändert sich nicht nur der Begriff des Selbst, des „Autos“, sondern auch der des Gesetzes, des Nomos. Beides hängt zusammen. Das Gesetz, der Nomos, hat seinen Ort nicht mehr als Gegenüber des Selbst, sondern das Selbst ist reduziert worden auf die technischen Funktionsprinzipien des Systems. Damit ist es nicht mehr rückgebunden an eine Freiheit, die sich selbst Rechenschaft zu geben und zur Rechenschaft gezogen weiß. Damit aber entschwindet der Begriff der Verantwortung und damit die Bedeutung, die der Begriff der Autonomie für den Menschen in seiner Selbstbestimmung hat. Autonomie verkörpert nicht mehr die Freiheit des Subjekts noch schließt sie seine Verantwortungsfähigkeit gegenüber sich selbst ein, sondern im Wandel vom Gesetz zur Funktion wird auch das Selbst des Menschen auf ein System reduziert. Aus der Selbstgesetzgebung des freiheitlichen Subjekts wird jetzt eine Selbststeuerung technologischer Systeme.

Mit der Entwicklung autonomer Systeme tritt daher eine Verschiebung des Verständnisses des Menschen und seiner Autonomie ein. Indem nicht mehr dem Subjekt, sondern einer selbst lernenden Maschine Autonomie zugesprochen wird, berührt das die Dimension von Selbstbestimmung und Freiheit als Ausweis der Autonomie.

Soll die Autonomie des Menschseins in seinem Personsein nun auf sein technisches Gemachtsein in Form von Künstlicher Intelligenz übergehen, dann stellt sich grundsätzlich die Frage nach dem Humanum. Bemessen wir Menschsein an seinen Fähigkeiten, seiner kognitiven Leistungskraft, dann wird es sich gegenüber der Künstlichen Intelligenz nicht behaupten können und wollen. So ist die Frage, was bestimmt den Menschen und was bestimmt seine Autonomie? Um einen Übergriff der Technik auf die „Naturwüchsigkeit des Menschen“ (Jürgen Habermas) zu verhindern, wird jetzt aufs Neue die Frage nach einer ethischen Verantwortung der neuen Technologien und damit das Verhältnis von Ethik und Technik drängend. Ist diesen mit einer herkömmlichen Ethik zu begegnen oder brauchen wir eine neue Ethik?

4. Autonome Systeme als Anfrage an das traditionelle Verständnis von Ethik und Technik

Kennzeichen von Ethik ist es, sich in Verantwortung verschiedenen Instanzen gegenüber verpflichtet zu wissen, den Instanzen des Selbst, der Welt sowie den Voraussetzungen von Selbst und Welt. Zu dieser Verantwortung des Menschen gehört auch die Technik. Wie es die philosophische Anthropologie herausgearbeitet hat, ist der Mensch dadurch bestimmt, dass er im Unterschied zum Tier keine instinktgeleitete Verhaltensweise und keine ökologische Nische hat, auf die er sich einfach verlassen kann. Vielmehr ist er in seinem Handeln auf die Entschlusskraft seiner Freiheit angewiesen. Er muss sich seinen eigenen Lebensraum bauen und gestalten. Technik gehört zum Menschsein dazu. Dann aber ist technischen Entwicklungen nicht einfach ein Verzicht zuzumuten, sondern sie sind in ihren Leistungen wahrzunehmen¹⁰ und als Verbesserung des Lebens zu begreifen. Die Technologie autonomer Systeme ist daher zunächst wie jede Technik in ihrer Ambivalenz wahrzunehmen. Sie kann ein sinnvolles Hilfsmittel sein, sie kann aber auch missbraucht werden.

10 Vgl. Arne Manzeschke, Fabian Karsch (Hg.), *Roboter, Computer und Hybride: Was ereignet sich zwischen Menschen und Maschinen?*, Baden-Baden 2016.

Dennoch fragt es sich, ob nicht eine neue Stufe erreicht ist, die die Technik nicht mehr als neutral einzustufen erlaubt, weil sie das Menschsein in seiner Individualbestimmtheit tangiert, indem sie etwa den Menschen zum Übermenschen stilisieren möchte. Dies ist der Fall, wenn die Technik selbst ihr Gegenüber, sei es die Natur, sei es ein Gesetz als eine äußere Ordnung, in sich einzieht im Dienst der Funktion. Denn dann ist keine Gestaltung der Technik mehr möglich. Wenn das Prinzip der Funktion zum Maßstab der Autonomie gemacht würde, kann Autonomie sich nicht mehr als Verantwortung verstehen. Handeln bzw. Verhalten entspringt dann nicht einer verantwortlichen Entscheidung, sondern es fügt sich einem Modus der Steuerung. Eine Unterscheidung von ethischer und technischer Freiheit ist hinfällig. Indem ethische Gestaltung selbst nur noch nach technischen Maßstäben erfolgt, ohne den Blick auf die Dienlichkeit des Lebens aufrecht zu erhalten, hat sie ihr Eigenrecht verspielt.

Der Unterschied zeigt sich in den ethischen Überlegungen zur Autonomie. Während sie für das freiheitliche Subjekt Ausfluss und Reflexion von dessen Autonomie selbst waren, aus der sich die Prinzipien der Verantwortung ergaben, beziehen sie sich jetzt allenfalls auf ein Objekt. Das, was die Selbstbestimmung wesentlich ausmacht, die Reflexivität wird dadurch eingezogen. Das im Begriff der Autonomie das Autos konstituierende Andere des Nomos ist ihm abhandengekommen. Damit wird der Begriff der Autonomie entkernt. Er ist nicht mehr Ausdruck von Freiheit und Selbstbestimmung, sondern Ausdruck systemischer Funktionalität. Eine Rückbindung des Subjekts in seiner freiheitlich schöpferischen Kreativität an ein ihm Anderes ist damit ausgeschlossen. Dieser Ausschluss aber eliminiert ein kritisches Gegenüber zur Technik, als das die Ethik immer gelten konnte.

Damit aber haben sich die Verhältnisse zur Bestimmung von Natur und Ordnung, Freiheit und Gesetz, wie sie das Autonome in der christlichen und philosophischen Tradition gekennzeichnet haben, verkehrt. Jetzt steht Autonomie nicht mehr für Selbstgesetzgebung, sondern allein für die Steuerung eines Prinzips der Machbarkeit. Das Verständnis des Menschen, wie es in der christlich-philosophischen Tradition gesehen wurde, wird aufgegeben. Denn dieses weiß sich in

seiner Verantwortung seiner Endlichkeit ebenso wie seinem Ursprung verpflichtet. Beides entzieht sich einem Verständnis von Autonomie eines autonomen Systems, das aus einer technischen Machbarkeit resultiert. Wird so der Begriff der Autonomie seines freiheits- und verantwortungsbestimmenden Kerns beraubt, wird er zum Wegweiser einer Gesellschaft, die sich nicht mehr als eine Formation aus freien Individuen, sondern aus funktionsgesteuerten Systemen versteht.

5. Kann es eine Maschinenethik geben?

Die für das Zeitalter autonomer Systeme naheliegende Frage, ob es eine Maschinenethik geben kann, rückt daher in greifbare Nähe.¹¹ Das Neue einer solchen Ethik bestünde darin, dass das Subjekt der Ethik gewechselt hat. Dieses wäre nicht mehr der Mensch, sondern die Maschine. Können programmierte Systeme, die nach moralischen Regeln, Sitten und Gebräuchen handeln, tatsächlich als ethische Subjekte gelten?

Die öffentliche Diskussion über ethische Fragen wie die, wie sich ein selbstfahrendes Auto im Fall eines unvermeidbaren Unfalls verhalten soll, zeigt die Problematik. Die Forderung allgemeinverbindlicher Regeln für das Verhalten eines selbstfahrenden Autos, etwa bei einem Unfall, wird nur von eingeschränkter Reichweite sein können. Übersehen wird bei solchen Vorstellungen der Programmierung moralischen Verhaltens, dass sie nicht allein nach Mustern im Sinne von Sitten, Gebräuchen und Konventionen verallgemeinert werden können. Verantwortliche Entscheidungen müssen als zeit-, situations- und erlebnisorientierte Entscheidungen wahrgenommen werden. Die vom Altbundesverfassungsrichter Udo Di Fabio geleitete Kommission über ethische Fragen des autonomen Fahrens konstatiert daher auch zu Recht: Entscheidungen wie die von Leben gegen Leben

11 Vgl. Catrin Misselhorn, Maschinenethik und „Artificial Morality“: Können und sollen Maschinen moralisch handeln?, in: Künstliche Intelligenz, Aus Politik und Zeitgeschichte (APuZ), Februar 2018, 29–33; Oliver Brendel, Überlegungen zur Disziplin der Maschinenethik, in: Künstliche Intelligenz, Aus Politik und Zeitgeschichte (APuZ), Februar 2018, 34–38.

seien „nicht eindeutig normierbar und auch nicht ethisch zweifelsfrei programmierbar“. Damit hat man sich in der Ethikkommission „automatisiertes und vernetztes Fahren“¹² klar nach einer Ethik Kants im Sinne der Selbstzwecklichkeit des Menschen ausgerichtet. Sie sperrt sich gegen Abwägungen unter dem Gesichtspunkt des Nutzens und damit gegen die utilitaristische Programmierungsmöglichkeit.

Wenn das aber so ist, dann stellt sich weniger die Frage, ob man Moral an eine Maschine abgeben möchte, hier an das selbstfahrende Auto, sondern wie mit der Tatsache umzugehen ist, dass menschliche Moral nicht „maschinisierbar“ ist. Denn diese bleibt immer an eine Selbstbestimmung im Sinne der Selbstzwecklichkeit des Subjekts gebunden. Daher wird eine Maschinenethik in starkem Sinne, die das Abgeben der Verantwortung an eine Maschine bedeutet, in Frage zu stellen sein. Ein verantwortliches Subjekt ist gerade dadurch von jeder Künstlichen Intelligenz unterschieden – selbst wenn diese selbstlernende Algorithmen vorsieht –, dass es eben nicht künstlich, d. h. durch Poiesis hervorgebracht ist. Es hat seine Wurzel in der Praxis, die auf Freiheit basiert.

Die aristotelische Unterscheidung von Poiesis und Praxis gewinnt daher angesichts der Herausforderung der Ethik durch autonome Systeme erneut an Gewicht. Denn Verantwortung entzieht sich der Herstellbarkeit. Das heißt letztlich, dass von Verantwortung autonomer Systeme nicht gesprochen werden kann im Sinne der Übertragung der Verantwortung an das System. Autonome Systeme bedeuten dann lediglich eine Steigerung des Handlungsspielraums des Menschen. Mit diesem geht eine Steigerung der Verantwortungspflicht einher, aber nicht, wie es eine Maschinisierung der Moral vermuten lassen könnte, deren Abgabe selbst an das System. Autonom ist das System dann allenfalls in seiner Funktion der Selbststeuerung, nicht aber als Ausdruck von deren Freiheit. Diese Freiheit bleibt jedoch Voraussetzung von Verantwortung. Diese aber ist nicht machbar. Menschliche Verantwortung bleibt es daher, die Freiheit nicht selbst einer Machbarkeit unterzuordnen, sondern alle Machbarkeit in Frei-

12 Ethik-Kommission, Autonomes und Vernetztes Fahren, Bericht Juni 2017. Bundesministerium für Verkehr und digitale Infrastruktur, Berlin 2017.

heit zu gestalten. Autonomie im Rahmen von Künstlicher Intelligenz und solcher im Rahmen von Personen lassen sich daher nicht aufeinander abbilden, denn die Autonomie des Menschseins ist – wie seine Freiheit – nicht machbar.

Eine Künstliche Intelligenz, die einem Transhumanismus zuarbeitet, wird insofern nicht nur den Menschen zur Abhängigkeit verdammen, sondern sie droht, das Verständnis von Freiheit des Menschen in seinem Verhältnis zur Technik und Verantwortung zu verkehren. Insgesamt gilt es daher zu überlegen, wie eine Künstliche Intelligenz zu entwickeln ist, und ob wir eine schwache oder eine starke Künstliche Intelligenz wollen. Unter einer starken Intelligenz, die aber nach wie vor noch ein unrealistisches Zukunftsszenarium darstellt, sind autonome Systeme zu verstehen, bei denen Steuerung und Entscheidung dem System allein obliegt. Bei einer schwachen Intelligenz hingegen wird die Steuerung und Entscheidung weiterhin dem Menschen in Letztinstanz überlassen. Für eine starke Künstliche Intelligenz gilt nicht nur, dass in ihrem Rahmen die Ethik hinterherhinkt, sondern dass sich die Ethik nicht mehr von der Technik unterscheiden lassen wird. Freiheit versteht sich dann allein als Machbarkeit. Das aber bedeutet in der Tat die Erschaffung einer neuen Spezies, die in Konkurrenz zum Menschen träte. Sie ließe sich weder unter die Maschine noch unter das Humane einordnen. Sie wäre trans-human und sie wäre zu fürchten, denn sie strebte die Unterwerfung des Humanum an. Für eine schwache Künstliche Intelligenz bleibt das Verhältnis der Verantwortung und Steuerbarkeit der Technik klar. Es sind Menschen, die für die Technik, für die Programmierung intelligenter Systeme verantwortlich sind. Insofern eignet solchen Systemen dann allenfalls eine Teilautonomie.

Die große Herausforderung ist es daher, die Bestimmungen unserer eigenen Freiheit zu reflektieren und die in ihr zu Tage tretende Ambivalenz, die immer auch das Verhältnis von Technik und Ethik betrifft, im Gleichgewicht zu halten. Schließlich war es immer die Gefahr der Technik, dass sie durch Übersteigerung umzuschlagen droht, indem die vermehrte Freiheit, die die Technik schafft, in eine umso größere Abhängigkeit von ihr führt. Die schon bei Goethe im Zauberlehrling

apostrophierten Überwältigungsmechanismen des Menschen durch die von ihm selbst entwickelte Technik scheinen sich daher auch mit der Künstlichen Intelligenz auf Dauer zu bewahrheiten. Gefordert ist daher eine Entscheidung, ob wir unsere Freiheit zu handeln behalten wollen oder nicht. Denn ist es zum Menschen 2.0³³, zur Künstlichen Intelligenz als Singularität, gekommen, die das Menschsein umfasst und übergreift, dann können wir nicht mehr entscheiden, was wir als Menschen sein wollen. Denn dann ist das Wesentliche des Menschen, sein Entscheidungszentrum, okkupiert durch ein Programm, dessen Grundlage und letzter Kern die Programmierung selbst ist. Indem sie mittels Technologie selbst die Ursprünglichkeit des Menschen eliminiert, hat sie dem Menschen den Entscheidungsspielraum genommen. Sie lässt nur noch die Berechenbarkeit von Verhalten übrig. Damit verwirkt sie das eigentliche Potenzial menschlicher Freiheit, das in der Ambivalenz ihrer Spannungen menschliche Entscheidungen reifen lässt.

6. Autonomie zwischen Personsein und Maschine – autonome Systeme und ihre ethische Verantwortung

Bedeutet die Kritik an den modernen Technologien autonomer Systeme nun ein Plädoyer für einen Technikverzicht? Ein solcher Verzicht entspricht der antiken und christlichen Tradition. Das Christentum sieht den Menschen im Auftrag der Herrschaft über die Erde. Sie gilt es zu bebauen und zu bewahren. Mit dieser Kulturbestimmung ging ein Gleichheits- und Freiheitsversprechen einher, zu dem die Technik wesentlich hinzugehört. Gerade die Entwicklung dieser Kulturtechnik sollte der Inklusion der bisher aus der Kommunikation und ökonomischen Interaktion Ausgeschlossenen sowie der politischen Freiheit angesichts der Unterdrückung von Meinungsfreiheit dienen.

Zu diesem Kulturauftrag gehört nicht nur die Umsetzung von Können und Wissen in der Technik, sondern auch deren Verantwortung. Diese bleibt selbst dann allein dem Menschen übertragen, wenn er

13 Vgl. Ray Kurzweil, a. a. O. Anm. 9.

versucht, Elemente von Autonomie auf ein System zu übertragen. Selbst wenn in dieses System Selbststeuerungsregeln implementiert werden, liegt deren verantwortliche Steuerung weiterhin in der Hand des verantwortlichen Menschen. Dessen Verantwortung ist nicht programmierbar und daher auch nicht technisch machbar.

Das intrinsische Problem einer Ethik der Technik, die sich angesichts der Fragen der Autonomie bewährt, entscheidet sich daher an der Frage der Machbarkeit bzw. der Reichweite des Steuerungspotentials. Verantwortung erfordert eine bleibende Unterscheidung von Technik und Ethik, bzw. Poiesis und Praxis, insofern die Ethik der Praxis und ihrer Freiheit, die Technik der Poiesis, der Kunst und der Machbarkeit, zugeordnet bleiben.

Eine angemessene Zuordnung erfordert die Klärung des ethischen Leistungsvermögens. Kritisch betrachtet werden muss hier das Auseinanderklaffen von technischen und ethischen Entwicklungen, d. h. das Voranschreiten der Technik, bei gleichzeitigem Hinterherhinken der ethischen Reflexionen dieser neuen Techniken. Die ethischen Begleitüberlegungen können in der Regel mit den neuen Errungenschaften der Technik nicht mithalten. Dementsprechend kommt – ein bereits klassisch gewordener Slogan – die Ethik immer zu spät.

Poetisch gefasst, in Anlehnung an Hegels Eule der Minerva, versuchte man bisher diese Aporie als zur Ethik gehörig zu bestätigen. Befriedigt hat diese Haltung aber nie – vor allem hat sie die Ethik in Misskredit ihrer möglichen Leistungsfähigkeit gebracht. Spätestens angesichts der sich immer stärker beschleunigenden technischen Entwicklungen der Künstlichen Intelligenz kann eine solche Haltung nicht mehr ausreichen. Vielmehr ist angesichts der Vision eines Homo Deus eine proaktive ethische Begleitung technischer Entwicklung notwendig. Diese aber erfordert weiterhin die Unterscheidung von Ethik und Technik bzw. von Selbstbestimmung und Ordnung, wie sie der Begriff der Autonomie in der Tradition zusammengefasst hat.

Nach Kant muss Autonomie als Selbstbestimmung an der Einhaltung der Dialektik unter Berücksichtigung der gegenseitigen Interpretation dieser polaren Zusammengehörigkeit von Gesetz und Freiheit orientiert sein. Eine Gesetzgebung durch die Gesetzmäßigkeiten der sich selbst regulierenden Systeme, die ihre Steuerung nicht der Vernunft ihrer Autoren verdanken, sondern den in sie implementierten Steuerungsmechanismen, würde die technische Freiheit zum Maßstab auch der Ethik machen. Technik wäre dann sich selbst überlassen, da die für die Verantwortung notwendige Selbstreflexivität der Freiheit verloren ginge. Dieser Verlust aber hebt den Begriff der Verantwortung auf. Autonomie kann daher nie das Abgeben von Freiheit und Verantwortung an ein externes System bedeuten. Insofern kann es auch keine Maschinenethik geben.

Die Akzeptanz autonomer Systeme lässt sich daher nicht ohne die Frage der Autonomie des Menschen vornehmen. Diese aber bleibt an die Anerkennung ihrer Grenzen gebunden. Wie bei der Gen-, Kern- und Informationstechnologie wird auch bei der Technologie autonomer Systeme dann eine Grenze überschritten werden, wenn der Autonomiebegriff seiner Dialektik von Gesetzgebung und Freiheit verlustig geht. Diese Gefährdung besteht in einem Umschlag der Autonomie auf eine Totalität der Freiheit ebenso wie auf eine Totalität der Gesetzgebung. Beiden Gefahren ist das autonome System ausgesetzt, indem hier entweder von der Machbarkeit der Autonomie oder von der Selbstgenügsamkeit des Systems ausgegangen wird. Eine solche Autonomie wäre keinem Selbst verpflichtet und keinem Gesetz. Sie drückt sich allein in jenen Steigerungsformen des Wissens und des Könnens aus, ja der Machbarkeit des Menschen, die dann in der Tat zur „letzten Erfindung des Menschen“, zum Transhumanismus, führten.

Hingegen bleibt es Ausweis des Verständnisses jenes Menschen, der sich als freies autonomes Subjekt im Sinne der christlich-antiken Tradition versteht, im Ursprung und am Ende unverfügbar zu sein, und gerade dies ist Ausweis seiner Verantwortungsfähigkeit.

Algorithmen entscheiden nicht, und sie werden es auch niemals

Joachim Fetzer

- › Ob Künstliche Intelligenz (KI) nutzt oder schadet, wird von gesellschaftlichen Lernprozessen abhängen. Nicht die Technik ist entscheidend, sondern deren Einbettung in Anwendungskontexte.
- › Die Frage, wem (Menschen als natürliche Personen) oder was (juristische Personen) Entscheidungsfähigkeit zugesprochen wird, ist eine gesellschaftliche Festlegung. Es gibt keine vernünftigen Argumente, Maschinen eigenständige Entscheidungsfähigkeiten zuzugestehen. Die Risiken wären zu groß.
- › Algorithmen können keine Entscheidungen treffen. Nur Menschen entscheiden und tragen Verantwortung. Algorithmen sind komplexe technische Instrumente, die lediglich unterstützen. Menschen entscheiden, wie und wo Algorithmen in Handlungsroutinen zum Einsatz kommen.
- › „Algorithmen entscheiden nie“ – dieser Leitsatz ist in künftigen Lernprozessen wichtig. Er bildet das Gegengewicht zu bewusster oder versehentlicher Verschleierung von Verantwortlichkeit. Verantwortlichkeit jedoch ist Grundelement einer freien Gesellschaft.

1. Jenseits von Euphorie und Panikmache

Künstliche Intelligenzen, das Zusammenspiel von algorithmischen Verfahren und großen Datenkorpora, sind längst in den Alltag von Menschen und Unternehmen involviert – als Alexa und Siri, als Google-Maps und Smartphone-basierter Coach zur Gewichtsreduktion, bei der Kreditvergabe und mancherorts beim Steuerbescheid.

Zugleich werden, ohne dass unterschiedliche moralisch-politische Grundpositionen der Öffentlichkeit deutlich sind, gesellschaftlich und politisch relevante Fachdebatten geführt. Sie tragen verschiedene Überschriften wie „algorithmische Entscheidungsfindung“, „Automatisierung von Entscheidungen“, „Entscheidungsunterstützung durch Künstliche Intelligenz“. Maßgebliche Treiber dieser Debatten sind Verbände (VZBV 2017; Bitkom/DFKI 2017), zivilgesellschaftliche Akteure¹, Medien² und Juristen, die Regulierungsmöglichkeiten diskutieren (Martini 2017), und neuerdings auch die EU-Kommission mit einer High-Level-Expert-Group für Künstliche Intelligenz (EU-COMM 2019). Zu selten fragt man jedoch, was eine Entscheidung zu einer Entscheidung macht.

Die hier vertretene These lautet, dass Algorithmen nicht entscheiden und auch nie entscheiden werden. Nicht weil der Mensch ihnen überlegen ist, sondern weil der Begriff Entscheidung – richtig verstanden – grundsätzlich nicht auf technische Systeme anwendbar ist (Abschnitt 3).

Was folgt in Zeiten von KI daraus für Politik und Gesellschaft, wenn man konsequent davon ausgeht, dass technische Systeme nicht entscheiden? Unter dieser Prämisse könnte der Umgang mit autonomen Systemen einen heilsamen Lernprozess befördern, der in diesem Text nur skizziert werden kann (Abschnitt 4).

1 Vgl. <https://algorithmwatch.org/de/mission-statement/> (Abruf 20.12.2018).

2 Vgl. die Dokumentation „Unheimliche Macht – Wie Algorithmen unser Leben bestimmen“ von Franziska Wielandt, <https://www.zdf.de/dokumentation/unheimliche-macht---wie-algorithmen-unser-leben-bestimmen-102.html>, (Abruf 20.12.2018).

Doch zunächst: Was ist das Neue an KI, und warum hat es den Anschein, als würden Algorithmen immer mehr „entscheiden“ (Abschnitt 2)?

2. Wie Algorithmen arbeiten und scheinbar „entscheiden“

Nicht jedes technische Gerät vermittelt den Eindruck zu entscheiden. Ein Taschenrechner etwa übernimmt Aufgaben, die er besser löst als die meisten Menschen. Trotzdem würde niemand behaupten, der Taschenrechner habe „entschieden“, dass die Wurzel aus 144 zwölf sei. Anders beim autonomen Fahrzeug: Wenn es eigenständig eine Fahrtroute wählt, ist die Anmutung seines Agierens dem beobachtbaren³ menschlichen Verhalten und seiner Anpassungs- und Lernfähigkeit sehr angenähert.

Was ist neu? Vereinfacht ist es die Weiterentwicklung der vom Taschenrechner bekannten elektronischen Datenverarbeitung (EDV). Deren Grundlogik besteht in drei Stufen: Eingabe (z. B. mittels Tastatur), Verarbeitung (z. B. Berechnung) und Ausgabe (auf Bildschirm, Papier oder anderweitig). Selbst wenn es sich nicht um eine einfache Berechnung, sondern um ein Konzern-Rechnungswesen handelt, bleibt die Grundstruktur unverändert: Input – Prozess – Output.

Bei KI entwickeln sich die drei Modellelemente weiter, und es wird ein viertes hinzugefügt: Die „Erfassung“ hat sich zur „Wahrnehmung“ fortentwickelt: Spracherkennung, Bilderkennung, logistische Daten, Daten aus Sensoren unter Einschluss von Tastsinn und Geruchserkennung vervielfältigen das, was früher „Eingabe“ hieß. Mit Gerüchen können Alarme ausgelöst werden, und das Wahrnehmungs-

3 Es entspricht der Grundstruktur des berühmten Turing-Tests, das von außen beobachtbare Verhalten von Maschinen und Menschen zu vergleichen. Ob man mit dem beobachtbaren Verhalten das Menschsein hinreichend wahrnimmt, ist eine der philosophischen Grundfragen. Man stelle sich eine Gerichtsverhandlung vor, bei der jemand einzig und allein anhand des beobachtbaren Verhaltens verurteilt wird, ohne jede Berücksichtigung von Deutungen wie „aus Absicht oder aus Versehen“, mit bestimmten Intentionen usw.

feld wird erheblich erweitert, wenn die Straßenlage mittels Sensoren im Auto 500 Meter weiter vorn wahrgenommen werden kann.

Auch die „Ausgabe“ hat sich von der Anzeige am Bildschirm oder der Ausgabe auf Papier zur „Handlung“ weiterentwickelt: Roboter-greifarme oder Bewässerungssysteme ändern lernend ihr Verhalten, Fahrzeuge sind in der Lage, Geschwindigkeits- und Richtungswechsel vorzunehmen, Musikstücke werden aus vorhandenen Elementen „komponiert“.

Die „Verarbeitung“ zwischen Input und Output, zwischen „Wahrnehmung“ und „Verhalten“ besteht nicht mehr nur aus definierten Rechenschritten, sondern aus verschiedenen Verfahren: etwa statistisch-analytischer Art, aus definierten Prozessregeln oder derzeit hoch im Kurs stehenden Methoden des maschinellen Lernens.⁴

Die Verarbeitungs- und Speicherkapazität moderner Computer übertrifft die Möglichkeiten des einzelnen Menschen: Werden etwa im digitalen Posteingang Worte wie „Beschwerde“ und seine Varianten zuverlässig erkannt, sortiert das System die Nachrichten zum Beschwerdemanagement“ (und nicht zum „Auftragseingang“). Eine solche „Entscheidung“ ist deutlich schneller als bei menschlicher Bearbeitung getroffen – wenn man es denn Entscheidung nennen will.

Zu den beschriebenen drei Erweiterungen kommt bei KI noch ein zusätzlicher Schritt: die Weiterentwicklung zum Lern- oder Trainingssystem. Die Erfahrungen mit den „Handlungen“ gehen wieder in das

4 Ob es klug ist, als Oberbegriff dieser Verfahren von „Comprehend/Verstehen“ zu sprechen (Bitkom 2017, S. 32), ist genauso strittig wie der Begriff „Intelligenz“ in Künstliche Intelligenz. Auch der Begriff „Handeln“ ist eigentlich fragwürdig, wenn man „Handeln“ von „sich verhalten“ unterscheidet und als Unterscheidungsmerkmal wieder Verstehen, Bewusstsein, (freien) Willen oder ähnliches ansieht.

kognitive System⁵, in den Datenkorpus, ein und entwickeln dieses weiter. Daher wird – bei bestimmten Verfahren der KI, etwa beim maschinellen Lernen – von Training und nicht von Programmierung oder Codierung gesprochen. Das technische System verändert sich im Verlauf selbst. Es lernt selbstständig.

Eine solche Entwicklungsfähigkeit ist für Werkzeuge ungewöhnlich – und sie ist das revolutionär Neue an KI. Ein Hammer darf sich beim Hämmern nicht verformen. Er sammelt auch keine Erfahrungen. Ein Taschenrechner muss in stetig gleicher Weise verlässlich funktionieren. KI dagegen verändert sich im Laufe ihrer Nutzung. Sie wird scheinbar lebendig und nähert sich dem menschlichen Verhalten an.

Diese Lern- bzw. Entwicklungsfähigkeit begründet neue Interaktionsmuster im Umgang mit technischen Systemen, neue zu klärende Haftungsfragen, und es entstehen neue Berufsgruppen, deren Professionsethos sich erst herausbilden muss (vgl. das Digital-Design-Manifest, Bitkom 2018).

Was bedeuten diese Entwicklungen für die Gesellschaft?

KI ist eine Basistechnologie, in ihrer Bedeutung vermutlich mit der Erfindung des Buchdrucks vergleichbar. In Verbindung mit anderen technologischen Entwicklungen, wie Robotik, hat KI das Potenzial, einen neuen Industrialisierungsschub auszulösen. Die eigentliche Herausforderung für die Gesellschaft durch Big Data und KI ist wahrscheinlich nicht der Verlust von Arbeitsplätzen, es sind auch nicht ökologische oder soziale Fragen. Vielmehr ist die größte Herausforderung, dass sich mit der breiten Nutzung algorithmischer Verfahren die gesellschaftlichen Entscheidungsprozesse und Entscheidungsrouinen

5 Dass irgendetwas von Algorithmen „entschieden“ werde, ist also schon im technischen Sinne falsch. Wenn, dann resultiert die „Entscheidung“ aus einem Zusammenspiel von Algorithmen mit großen Datenmengen. Ohne letztere gibt es auch keine funktionsfähige „Künstlichen Intelligenz“. Richtigerweise wird von Fachleuten darauf hingewiesen, dass besser von Algorithmischen Entscheidungsprozessen als von entscheidenden Algorithmen gesprochen werden solle. Es ist die gesamte Modellierung von Prozessen, welche die „Entscheidungsabläufe“ ausrichtet (Zweig 2018).

neu justieren – inklusive der vorbereitenden, begleitenden oder kommentierenden Information und Kommunikation.

Gewohnte Routinen werden sich verändern oder in neue Bedeutungsrahmen einfügen. Dies hat Auswirkungen auf die Qualität von Entscheidungsprozessen, auf die Verantwortlichkeit für Ergebnisse, aber auch auf die durch Entscheidungsverfahren erzeugte Legitimität – nicht zuletzt im politischen Prozess.

3. Warum es wichtig ist, dass wir Algorithmen nicht zu Entscheidungsträgern machen

Begrifflichkeiten wie „algorithmische Entscheidungsfindung“ oder „algorithmisches Entscheidungsverfahren“ werden in der Regel unreflektiert hingenommen.

Wenn wir uns fragen, was die Formulierung „Algorithmen entscheiden“ bedeutet, gilt es zunächst, den Begriff „Entscheidung“ zu hinterfragen. Dann erst kann man prüfen, ob er auf technische Systeme anwendbar ist.

Was heißt „entscheiden“?

Was impliziert der Begriff Entscheidung? Eine treffende Beschreibung des Tübinger Philosophen Otfried Höffe lautet (Höffe 2002):

„Entscheidung bezeichnet den (freien) Entschluss von einzelnen oder von Gruppen, mit dem man aus verschiedenen Handlungsmöglichkeiten eine als die eigene ergreift und sich dadurch zu einem Tun oder Lassen bestimmt. Durch Entscheidungen entsteht im persönlichen und politischen Raum geschichtliche Wirklichkeit. Mit der Zurückführung seiner Handlungen auf Entscheidungen wird der Mensch zum Ursprung seines Tuns, für das er deshalb Verantwortung trägt, allerdings keine totale, da er den persönlichen und gesellschaftlichen Kontext seiner Entscheidung nicht mitsetzt (sh. Determination).

Als Entscheidung ist nicht bloß der örtlich und zeitlich punktuelle Akt der Beschlussfassung zu verstehen, sondern der ganze Prozess der Entscheidungsfindung, in dem oft auf eine problemorientierte Phase die lösungsorientierte folgt. [...]“

Zunächst zum zweiten Absatz, zum Entscheidungsprozess: Wie ähnlich sind algorithmische Entscheidungsverfahren menschlichen Entscheidungen? Sind sie besser oder schlechter? Für algorithmische Verfahren werden Transparenz und Nachvollziehbarkeit gefordert. Fordern wir dies auch von menschlichen Entscheidungen? Verstehen wir menschliche Entscheidungsprozesse? Erkennen wir an, dass unsere Entscheidungen häufig von Routinen und Intuitionen geleitet sind! Wir treffen Bauchentscheidungen. *Die Intelligenz des Unbewussten und die Macht der Intuition* heißt ein Buch des Bildungsforschers Gerd Gigerenzer. Wer die dort beschriebenen Beispiele analysiert, wird feststellen, dass menschliche Entscheidungen oft auf Erfahrungswissen und schneller Mustererkennung beruhen – in gewisser Weise vergleichbar mit großen Datenmengen und algorithmischen Verfahren der Mustererkennung.⁶ Die technischen Möglichkeiten werden häufig unterschätzt und die Besonderheiten menschlicher Entscheidungsprozesse deutlich überschätzt.

Doch nun zum ersten Absatz:⁷ Was sehen wir eigentlich als Entscheidung an? Nicht jede Auswahl aus mehreren Möglichkeiten bezeichnen wir als Entscheidung. Zur Entscheidung wird eine Auswahl durch unsere Deutung: Wir rechnen diese Auswahl uns selbst zu. Wir übernehmen Verantwortung – als Einzelner oder als Gruppe. Die Debatte über die äußere und innere Freiheit von Entscheidungen

6 Es macht keinen Sinn, sorgfältig strukturierte Entscheidungsprozesse und schnelle Bauchentscheidungen gegeneinander auszuspielen. Der Stellenwert der Intuition hat beim Reden über Entscheidungen an Akzeptanz eingebüßt. Vielleicht zu unrecht (vgl. Gigerenzer 2007, S. 242).

7 Der größte Teil der Literatur bezieht sich auf Entscheidungsprozesse und Entscheidungstheorien und auf Fragen, wie Entscheidungen optimiert werden können. Wenig findet sich zur Frage, was eine Entscheidung eigentlich ist und was mit diesem Begriff bezeichnet wird.

ist immer auch, wahrscheinlich sogar zuallererst, eine Debatte darüber, welchen Stellenwert Verantwortung und Verantwortlichkeit als Zurechnungs- und Steuerungsmodus in einer Gesellschaft haben.⁸

Dies sei an einem Beispiel veranschaulicht: Stellen wir uns einen Sportler vor – einen Tennisspieler. Er schlägt in die falsche Richtung und vergibt den entscheidenden Ball. Hat er es entschieden? Oder hat er nur reagiert? Im entscheidenden Moment mag es sich als Reaktion angefühlt haben. Wenn er anschließend sagt, er habe einen Fehler gemacht, deutet er es als eigene „Entscheidung“, als selbstbestimmte Auswahl unter verschiedenen Möglichkeiten. Doch wie oft werden vermeintliche Umstände, der Wind oder schlechte Trainingsmöglichkeiten ins Feld geführt. Wir nennen es Sportsgeist, wenn jemand eingesteht: Ich war es – und niemand sonst. Auf diese Haltung – die Übernahme von Verantwortung – kommt es an.

Übertragen auf sogenannte algorithmische Entscheidungsverfahren bedeutet das: Selbst eine KI, die alle Turing-Tests bestehen würde, deren beobachtbares Verhalten sich also in nichts von menschlichem Verhalten unterscheidet, würde erst dann „entscheiden“ können, wenn die Gesellschaft entschieden hätte, sie als Person und verantwortliches Subjekt in die Gesellschaft aufzunehmen.

8 Vgl. hierzu Fetzer 2004, v.a. Kap. 4. Über den Zusammenhang zwischen „Entscheiden“ und „Verantworten“ lässt sich einiges lernen, wenn man die Übertragung auf nicht-menschliche Akteure prüft. In Politik und Gesellschaft wird heute kaum mehr davon ausgegangen, dass die Welt von den vermuteten „Entscheidungen“ der Engel und Götter beeinflusst wird, die sich möglicherweise in den offensichtlichen „Entscheidungen“ von Naturgewalten äußern, wenn zum Beispiel der Sturm „entscheidet“, diese und nicht jene Route zu nehmen oder der Vulkan „beschließt“, genau jetzt und nicht später auszubrechen. Die „Entscheidungen“ von Naturgewalten werden naturwissenschaftlich untersucht und bei den „Entscheidungen“ von Göttern und Engeln wird entweder die unmittelbare Kausalwirkung auf die sichtbare Welt in Frage gestellt oder die Interpretation der sichtbaren Erscheinungen als Konsequenz göttlicher Entscheidungen als nicht mehr allgemeingültig angesehen. Das Entscheidungsverhalten von insbesondere höher entwickelten Tieren ist kaum von demjenigen von menschlichen Individuen zu unterscheiden. Bisher sind sie nicht als verantwortungsfähige Subjekte in die Zurechnungsmodi der Gesellschaft aufgenommen.

Es geht nicht um Begriffszuweisungen, sondern um eine Lebensform

Obwohl es im Umgang mit Computern, besonders wenn sie KI enthalten (z. B. in Form von Sprachassistenten) verführerisch ist, mit ihnen zu kommunizieren und sie wie menschenähnliche Subjekte zu behandeln, sollten wir sie auf keinen Fall als verantwortliche Subjekte gesellschaftlich anerkennen.

Es geht dabei nicht um Begriffszuweisungen, sondern um eine gesellschaftliche Lebensform, in der Menschen sich wechselseitig Freiheit zugestehen und daher Verantwortung zumuten.⁹

Die Frage nach der Diffusion und Zurechnung von Verantwortung in komplexen Systemen ist nicht neu. Ist ein Unternehmenschef stellvertretend für strafbare Handlungen des Unternehmens verantwortlich? Sind Mitarbeiter nur Ausführende, solange sie im Rahmen ihrer Befugnisse arbeiten? Sind Gruppenentscheidungen optimal, weil sie eine Vielzahl von Perspektiven einbringen, oder werden sie als Instrument genutzt, um individuelle Verantwortung zu vermeiden? Beides gibt es. Wir delegieren Entscheidungen an Gruppen, reduzieren mittels rechtlicher Standardisierungen Entscheidungsspielräume, organisieren Enthaltung durch Compliance-Systeme usw. Die Delegation von Entscheidungen (und Verantwortung) muss in der analogen wie in der digitalen Welt immer wieder ausgehandelt werden.

Die zunehmende Bedeutung von künstlichen Intelligenzen kann eine Versuchung sein, menschliche „Unwilligkeit“ zur Verantwortung

9 *Das Handwerk der Freiheit* heißt ein Buch des Berner Philosophen Peter Bieri. Nach ausführlichen Schilderungen von Entscheidungsprozessen unterschiedlichster Art diskutiert er die Frage, ob „ein Terrorist Herr über seinen Willen“ oder mit den Worten Dostojewskis „der Lakai eines fremden Gedankens“ (ist)? Dürfen wir ihn verurteilen, wenn es sich herausstellen sollte, dass er z. B. Opfer einer Gehirnwäsche islamischer Fundamentalisten war? Bieris Antwort: Wir haben zwei Gründe, ihn zu verurteilen: Erstens ist er in seiner Willensbildung nicht ohnmächtig; er hätte auch andere Dinge auf sich einwirken lassen können – anders als beispielsweise das Opfer einer Hypnose. Und zweitens geschieht durch eine Verurteilung einfach dieses: Wir verteidigen gegen ihn eine Lebensform der Verantwortlichkeit.

durch Verantwortungsverlagerung in die KI auszuleben. Das Diktum „Algorithmen entscheiden nie“ ist deshalb wichtig, weil es in einem gesellschaftlichen Lernprozess, in dem wir alle (Konsumenten und Bürger, Unternehmen und Behörden, Politik und Justiz) den Umgang mit algorithmischen Entscheidungsprozessen lernen, die Suchrichtung bestimmt oder eine gefährliche Suchrichtung verhindert.

4. Wenn die Maschinen zu „lernen“ beginnen, müssen wir doch damit nicht aufhören

Neue Technologien und deren Einführung sind immer auch mit Gefahren verbunden.¹⁰ Die Gefahr der Verantwortungsvermeidung durch Verlagerung in KI ist ein systemisches Risiko. Ihm muss entgegengewirkt werden.

Mit der breiten Nutzung algorithmischer Verfahren werden sich gesellschaftliche Entscheidungsprozesse und Entscheidungsroutrinen neu justieren – inklusive der vorbereitenden, begleitenden oder kommentierenden Information und Kommunikation. Die Herausforderung speist sich aus zwei Quellen: „Die eine besteht im Zusammenwirken von Mensch und Technik in soziotechnischen Systemen [...]. Die andere ist das Problem der kollektiven Verantwortung, welches entsteht, weil es in soziotechnischen Systemen eine Vielzahl möglicher Verantwortungsträger gibt.“ (Misselhorn 2018, S. 135).

Diese Neuauflage zweier bekannter Herausforderungen ruft den Gesetzgeber auf den Plan, aber Regulierung wird den gesellschaftlichen Lernprozess nicht ersetzen oder überflüssig machen. Vier Aspekte dieses Lernprozesses sollen skizziert werden:

¹⁰ Die Frage des bewussten Missbrauchs – eine selbstverständlich immer vorhandene Möglichkeit – sei hier ausgeklammert. Missbrauch ist bei jeder hochleistungsfähigen Technologie natürlich eine Möglichkeit und ein Problem. Dar aus einen Anwendungsverzicht abzuleiten ist nicht sinnvoll.

- › **Die erste Aufgabe** wird sein, zwischen intelligenten Routinen und wohlüberlegten aktiven Entscheidungen zu unterscheiden. Alles, was mit beschreibbaren Entscheidungs-routinen bearbeitet werden kann, wird künftig unter Zuhilfenahme großer Datenpools und Algorithmen ausgeführt werden. Wir stehen nicht vor der Frage, *ob* wir die Ersetzung von Entscheidungs-routinen durch KI-Prozesse wollen, sondern *was* wir an Routinen übergeben oder wo tatsächlich eine Entscheidung stattfinden soll – und von wem entschieden wird.

Mit der zunehmenden Verlagerung von Routinen in KI wird die Gesellschaft lernen müssen, an welchen Stellen vermeintliche Entscheidungen schon längst keine Entscheidungen im eigentlichen Sinne mehr sind, sondern nur noch die Abarbeitung früher getroffener Entscheidungen oder eingeübter Gewohnheiten.¹¹

Beispiel: Ist die Erlaubnis der Einreise durch einen Grenzbeamten bei Vorliegen eines gültigen Passes und der Identifizierung der Person eigentlich eine Entscheidung oder ist es die Ausführung einer rechtlichen Vorschrift, das heißt die Ausführung einer grundsätzlichen Entscheidung, dass Personen mit bestimmten Merkmalen einreisen dürfen.

- › **Die zweite Aufgabe** besteht in der Etablierung hinreichender Sorgfaltsregeln (bereits in den frühen Phasen des Einführungsprozesses) und im Umgang mit Fehlern.

Mit KI wird es zu Entwicklungen und Entscheidungen kommen, die sich im Nachhinein als falsch herausstellen. Das spricht nicht gegen die Technologie. Was vielmehr in der noch zu lernenden Zusammenarbeit mit Algorithmen anzustreben ist,

11 Instruktiv zu den Mühen dieses Lernprozesses sind die Ausführungen über automatisierte Verwaltungsverfahren, Kap. 4 in Djefal 2018.

ist eine der jeweiligen Situation angemessene¹² Sorgfalt¹³ und Qualität¹⁴.

Eine alltagstaugliche Formulierung für sorgfältige Entscheidungen lautet: „nach bestem Wissen und Gewissen“. Sie beschreibt das gelingende Zusammenspiel von fachlicher („Wissen“) und persönlicher Urteilskompetenz („Gewissen“). Wenn Teile der fachlichen Beurteilung von KI übernommen werden, ändert sich daran im Grundsatz nichts. Auch wenn die Versuchung groß sein mag, sich durch Hinweis auf eine falsche Datenbasis zu exkulpieren. Es gilt der Leitsatz: „Algorithmen entscheiden nie.“

- **Die dritte Aufgabe** in diesem Lernprozess betrifft Unternehmen, aber auch andere korporative Akteure, etwa staatliche Behörden.

Unternehmen und andere Organisationen übernehmen in vielen Fällen Verantwortung, die nicht auf individuelle Verantwortung zurückgeführt werden kann (vgl. Fetzer 2004, v. a. Kap. 6). Dafür wurde die Kunstfigur der juristischen Person geschaffen. Künftig wird Unternehmen und Organisationen nicht nur das Handeln ihrer Mitarbeiter zugerechnet, sondern auch das Verhalten der jeweils eingesetzten KI. Wenn man die

12 Eine Studie der Bertelsmann-Stiftung (Vieth/Wagner 2017) versucht, verschiedene Einsatzsituationen von Algorithmen anhand ihrer Bedeutung für „Teilhabechancen“ zu klassifizieren, um damit ein Raster für angemessene Sorgfalt und Nachvollziehbarkeit zu entwickeln. Methodisch ist das sinnvoll. Die Entscheidung für den normativen Bezugspunkt „Teilhabe“ kann allerdings in Frage gestellt werden, vgl. Fetzer 2018.

13 Vgl. Zweig 2018.

14 Es besteht die Gefahr, dass diese Qualitätsdebatte für politische Zwecke missbraucht wird. Wer schon immer Kreditunternehmen noch mehr unter staatliche Kontrolle stellen oder am liebsten in Staatseigentum überführen will, findet in der Debatte über „Kreditvergabe durch Algorithmen“ ein neues Argumentationspotenzial für zu erzwingende „Nachvollziehbarkeit“.

Idee der Verantwortung des korporativen Akteurs konsequent anwendet, sind viele der unter dem Schlagwort „algorithmens-basiertes Entscheiden“ diskutierten Probleme lösbar.

Korporative Akteure unterliegen internen Entscheidungsprozessen, deren Ergebnisse sie Dritten gegenüber vertreten müssen. Darin unterscheiden sie sich wesentlich von natürlichen Personen. Unternehmen und Behörden sind also aufgefordert, ihre Kommunikations- und Servicequalität sicherzustellen. Die Integration von KI-unterstützten Entscheidungsprozessen in Institutionen ändert nichts an diesem Grundsatz.¹⁵ Idealerweise sind Unternehmen Treiber von Innovationsprozessen und sind daher in vielfacher Hinsicht mitverantwortlich für das Gelingen eines gesamtgesellschaftlichen Lern- und Change-Prozesses infolge des Einsatzes von KI.

- › **Die vierte Aufgabe** umfasst einen Bildungsprozess für alle. Mehr Informatikkenntnisse sind nützlich, aber entscheidender dürfte der Umgang mit großen Datenmengen und statistischen wie stochastischen Zusammenhängen sein. Der Unterschied zwischen einer Kausalverknüpfung und einer Korrelation wird nicht mehr nur Expertenwissen sein, sondern erhält alltägliche Relevanz. Was bedeuten Wahrscheinlichkeitsaussagen für das eigene Leben? Sei es, weil andere auf dieser Basis Entscheidungen für mich treffen, die Auswirkungen auf mein Leben haben, oder sei es, dass ich selbst meine Entscheidungen an gesammelten Daten und stochastischen Ableitungen orientiere.¹⁶

15 Der Hinweis des Kartellamtes, dass die Lufthansa für ihre Preisbildung selbstverständlich verantwortlich sei, auch wenn diese durch ein algorithmisches Verfahren geschehe, ist hierfür ein schönes Beispiel.

16 Ein nicht unwesentlicher Teil der Debatte zu „Diskriminierungsgefahren durch Algorithmen“ hat mit diesem häufig unverstandenen qualitativen Sprung zu tun.

5. Zwischen Chancen und Schicksal – ein Ausblick

Wie KI sich gesellschaftlich auswirkt, ist keine Frage der Technikgestaltung und nur bedingt regelbar durch Ethik-Richtlinien, wie sie zum Beispiel von der High-Level Expert Group on Artificial Intelligence der EU-Kommission entwickelt wurden (EU-COMM 2019). Auch die gesellschaftlichen Entwicklungen nach Erfindung des Buchdrucks (inklusive Reformation und Dreißigjähriger Krieg) wurden nicht von Ethik-Richtlinien begleitet und waren kaum steuerbar durch Druckgenehmigungen und Zensur. Im Ergebnis wurde nicht Analphabetismus verboten, sondern man hat Schulen gegründet und die Schulpflicht eingeführt.

Letztlich entscheidet auch im Falle der KI die Lernfähigkeit der Gesellschaft im Umgang mit den neuen technischen Möglichkeiten. Die Einführung von KI muss nicht zwingend die Idee und das Ideal der menschlichen Autonomie in Frage stellen.¹⁷ Denkbar ist auch das Gegenteil: Ihre Nutzung könnte dazu führen, dass persönliche Entscheidungen sehr viel stärker wertgeschätzt werden: Entscheidungen, die diesen Namen verdienen, und nicht nur Routinen sind.

Es sind die Alltagssituationen, in denen sich zeigen wird, wie wir mit uns selbst, den Mitmenschen und den neuen Technologien umzugehen gedenken.¹⁸

17 In diese Richtung argumentiert Gräß-Schmidt 2018. Aber die gesellschaftlichen Herausforderungen mit der Vielzahl von künstlichen Intelligenzen (im Sinne „schwacher KI“) wird man besser gestalten können, wenn man sie nicht mit den (positiven wie negativen) Visionen einer „starken KI“ vermischt, von der kein Mensch sagen kann, ob es sie jemals geben wird. Die Debatte über Maschinen mit eigenem Bewusstsein als science-fiction-artige Verdichtung lenkt von den anstehenden Herausforderungen ab. Es ist sogar fahrlässig, die politische und kulturelle Entwicklung an Spekulationen über dominierende Technologien und selbst-bewusste starke Künstliche Intelligenz zu binden.

18 Es wäre daher ungeschickt, die Entwicklung von Technologien, welche in so hohem Maße die Abläufe der Gesellschaft durchdringen, den Forschern und vor allem Unternehmen in Ländern zu überlassen, die naturgemäß an mancher Stelle andere kulturelle Traditionen haben dürften.

Die ontologisch vielleicht strittige, aber politisch wichtige These „Nur Menschen, nicht Algorithmen entscheiden“ (s. Kap. 3) zielt auf einen Klärungsprozess:

- › Was soll/darf „abgearbeitet“ werden? Und nach welchen und wessen Regeln?
- › Was soll/darf „entschieden“ werden? Und von wem?
- › Was soll/darf dem Zufall überlassen werden? Auch wenn es im Ergebnis ungerecht, möglicherweise undemokratisch und per definitionem unverantwortlich ist? In einer Gesellschaft, in der Schicksalsschläge gelegentlich zu Ungerechtigkeiten umfirmiert werden, ist dies ein neues und ohne Bezug auf Religion kaum zu besprechendes Thema.

Nicht die Algorithmen, sondern wir Menschen werden diese drei Fragen in all ihren Aspekten immer wieder beantworten müssen. Zum jetzigen Zeitpunkt ist es vermessen, fertige Antworten haben zu wollen. Auch moralisch-politische Präferenzen sollten nicht unterschätzt werden. Wer etwa gewohnte Verfahren positiv beurteilt, wird sie auch in technisch veränderter Struktur erhalten wollen. Andere werden versuchen, technische Neuerungen zur Durchsetzung veränderter Standards zu missbrauchen. Es existiert keine positions- und parteiübergreifende Blaupause. Umso wichtiger, dass sich die politischen Gestalter des Themas annehmen und es nicht anderen Akteuren überlassen. Dann entmündigt der technisch induzierte, aber ethisch-politisch zu führende Klärungsprozess das Individuum nicht. Im Gegenteil: Er kann und soll auch für ein Revival bewusster Freiheit und Verantwortlichkeit genutzt werden.

Referenzen

- B Bieri, Peter** (2001). Das Handwerk der Freiheit. Über die Entdeckung des eigenen Willens, Frankfurt am Main.
- Bitkom e. V.** (2018). Digital-Design-Manifest. Eine selbstbewusste Gestaltungsprofession ist der Schlüssel für eine erfolgreiche und nachhaltige Digitalisierung, Berlin (<http://www.digital-design-manifest.de>, Abruf 10.01.2019).
- Bitkom e. V./DFKI GmbH** (2017). Künstliche Intelligenz. Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung, Berlin (<https://bit.ly/2lh8kiJ>, 20.12.2018).
- D Djeffal, Christian** (2018). Künstliche Intelligenz in der öffentlichen Verwaltung (NEGZ Kurzstudie), (<https://negz.org/wp-content/uploads/2018/11/NEGZ-Kurzstudie-3-KuenstlIntelligenz-20181113-digital.pdf>, Abruf 10.12.2018).
- E EU-Commission** (2019). Ethics Guidelines for Trustworthy AI (<https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>, Abruf 15.05.2019).
- F Fetzter, Joachim** (2004). Die Verantwortung der Unternehmung, Gütersloh (<https://www.econstor.eu/handle/10419/172204>, Abruf 10.12.2018).
- Fetzter, Joachim** (2018). Ethik der Kapitalbeteiligungen. Freiheit, Würde und Nachhaltigkeit in der Vermögensgesellschaft, in: Beyer, Heinrich / Naumer, Hans-Jörg (Hg.): CSR und Mitarbeiterbeteiligung. Die Kapitalbeteiligung im 21. Jahrhundert – Gerechte Teilhabe statt Umverteilung, Berlin, S. 55–69.
- G Gigerenzer, Gerd** (2007). Bauchentscheidungen. Die Intelligenz des Unbewussten und die Macht der Intuition, 6. Aufl., München.

Gräß-Schmidt, Elisabeth (2018). Zwischen System und Verantwortung. Ethische Überlegungen zum Begriff der Autonomie angesichts der Entwicklung autonomer Systeme, Konrad-Adenauer-Stiftung, Analysen & Argumente Nr. 319, Berlin.

H Höffe, Otfried (2002). Art. Entscheidung, in: Lexikon der Ethik, hv. v. Otfried Höffe, München, S. 51 f.

M Martini, Mario (2017). Algorithmen als Herausforderung für die Rechtsordnung, Juristenzeitung 21, 72. Jahrgang, Tübingen, S. 1017–1025.

Misselhorn, Catrin (2018). Grundfragen der Maschinenethik, Stuttgart.

V Vieth, Kilian/ Wagner, Bernd (2017). Teilhabe, Ausgerechnet, Wie algorithmische Prozesse Teilhabechancen beeinflussen können (Bertelsmann Stiftung: Impuls Algorithmenethik #2), Gütersloh (<https://doi.org/10.11586/2017027>; Abruf 10.12.2018)

VZBV e. V. (2017). Algorithmenbasierte Entscheidungsprozesse, Thesenpapier der verbraucherzentrale Bundesverband, https://www.vzbv.de/sites/default/files/downloads/2018/05/22/dm_17-12-07_vzbv_thesenpapier_algorithmen.pdf, 20.12.2018).

Z Zweig, Katharina A. (2018). Wo Maschinen irren können. Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung, (Bertelsmann Stiftung: Impuls Algorithmenethik #4), Gütersloh (<https://doi.org/10.11586/2018006>; 20.12.2018).

Algorithmische Entscheidungen: Transparenz und Kontrolle

Katharina A. Zweig

- › Algorithmen sind Regeln zur Lösung mathematisch beschreibbarer Probleme. Sie sind ein wesentlicher Teil „Algorithmischer Entscheidungssysteme“ (ADM-Systeme), die in der Versicherungsbranche Risiken bewerten oder in anderen Kontexten vergleichbare Aufgaben lösen können.
- › Wenn die Entscheidungsregeln für die Risikobewertung unklar sind, können selbstlernende ADM-Systeme verwendet werden. Diese lernen aus bisherigen Fällen und Daten.
- › Allerdings lässt sich das Zustandekommen der Ergebnisse in der Regel nicht nachvollziehen. Selbstlernende ADM-Systeme bedürfen insbesondere immer dann einer speziellen Kontrolle, wenn sie Entscheidungen über Menschen treffen und Fehlentscheidungen einzelne Individuen bzw. die Gesellschaft als Ganzes schädigen können.
- › Dazu reicht es nicht aus, Algorithmen allein überprüfbar zu machen, vielmehr müssen die ADM-Systeme insgesamt (Algorithmen und Daten) und das sozio-informatische Gesamtsystem, in dem sie genutzt werden, in den Blick genommen werden.

In der öffentlichen Debatte über Digitalisierung und KI ist der Begriff Algorithmus allgegenwärtig. Auf der einen Seite verkauft man Algorithmen als geradezu magische Lösung für schwierige Probleme. So bewirbt eine Firma ihre Software, die Angestellte in ihrer Leistung bewerten soll, mit den Worten: „In the end, with the availability of good data, the predictive possibilities are virtually unlimited.“ Auf der anderen Seite sind Algorithmen so sehr in Verruf geraten, dass neue Kontrollorgane und größere Einblicke in ihre Wirkungsweise gefordert werden (Meyer-Schöneberger & Cukier, 2013). Doch geraten die Begrifflichkeiten zuweilen durcheinander, wenn beispielsweise ein „Algorithmen-TÜV“ verlangt wird, um Algorithmen zu kontrollieren, oder wenn Transparenz eines Algorithmus mit der Veröffentlichung seines Codes gleichgesetzt wird. Bei näherer Betrachtung zeigt sich, dass die Algorithmen, deren Überwachung gefordert wird, eigentlich nur einen kleinen Teil sogenannter *algorithmischer Entscheidungssysteme* ausmachen. Für diese zeigt die aktuelle Forschung allerdings, dass es nicht ausreicht, sie als einzelnes Produkt zu kontrollieren, sondern dass auch ihre Einbettung in unsere Gesellschaft der Überprüfung bedarf. Um diese Schlussfolgerung nachvollziehen zu können, ist es wichtig, die Begriffe „Algorithmisches Entscheidungssystem“ und „Algorithmus“ zu differenzieren.

1. Algorithmische Entscheidungssysteme

Algorithmische Entscheidungssysteme (algorithmic decision making systems oder auch ADM-Systeme) beinhalten Regeln, nach denen eine Entscheidung getroffen werden kann. Dazu gehören einfache Systeme, wenn etwa KFZ-Versicherungen auf Grundlage von Daten Schadensfreiheitsklassen bestimmen. Die Entscheidungsregeln sind hier klar und für Menschen einsichtig: Sie beruhen im Wesentlichen auf dem Alter der Fahrer und ihrer bisherigen Unfallhistorie. In der aktuellen Diskussion geht es allerdings um Entscheidungssysteme, deren Entscheidungsregeln von Algorithmen selbstständig abgeleitet werden.

Etymologie des Begriffes „Algorithmus“

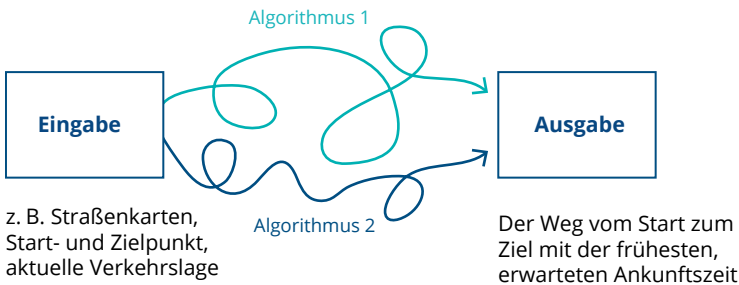
Der Begriff Algorithmus leitet sich von einem arabischen Mathematiker namens Al-Chwarizmi ab, der im 9. Jahrhundert ein mathematisches Lehrbuch verfasste. Es besteht Verwechslungsgefahr mit dem Logarithmus und mit Rhythmus, die aber etymologisch nicht mit dem Begriff „Algorithmus“ verwandt sind.

Was sind Algorithmen?

Algorithmen finden eine Lösung für mathematisch beschreibbare Probleme, die in unterschiedlichen Anwendungssituationen immer wieder gelöst werden müssen. Ein mathematisches Problem benennt die vorliegenden Informationen (Eingabe) und definiert Eigenschaften, die eine auf den Informationen basierende Lösung haben soll (Ausgabe).

Mathematisches Problem

definiert das Verhältnis von Eingabe und Ausgabe.



Algorithmus

definiert eine Folge von Handlungen, die – basierend auf der Eingabe – eine Ausgabe mit den gewünschten Eigenschaften berechnet.

Abbildung 1:

Illustration des Verhältnisses von mathematischem Problem und Algorithmus, der zu dessen Lösung eingesetzt wird.

Ein Beispiel ist die Berechnung des kürzesten Weges, basierend auf Kartenmaterial, dem Start- und Zielpunkt, u. U. um aktuelle Straßenverhältnisse ergänzt. Algorithmen sind Handlungsanweisungen, die

zu einem Ergebnis mit den gewünschten Eigenschaften führen, basierend auf den eingegebenen Informationen. In der Regel gibt es mehrere Algorithmen, um zu einer Lösung zu gelangen (Abbildung 1).

Algorithmische Entscheidungssysteme

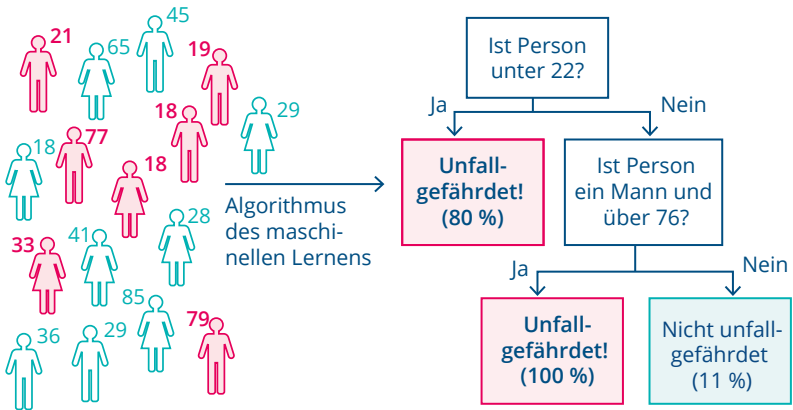


Abbildung 2:

Algorithmische Entscheidungssysteme bekommen Daten (Eingaben) und Informationen über ein zu lernendes Verhalten und leiten daraus Entscheidungsregeln ab. Im Beispiel sind Alter und Geschlecht von Autofahrern angegeben und die Information, ob sie in den letzten drei Jahren einen selbstverschuldeten Unfall hatten (rot: ja, türkis: nein). Der Algorithmus findet Regeln, die die Daten möglichst gut so in zwei kleinere Mengen aufteilen, so dass in derselben Menge fast alle entweder einen Unfall hatten oder fast alle keinen. Diese Regeln können dazu verwendet werden, die Unfallgefährdung für neue Versicherungsnehmer abzuschätzen. Das Ergebnis ist stark von den verwendeten Daten abhängig.

Algorithmen des maschinellen Lernens

Es ist allerdings kein Algorithmus bekannt, der aus einer Reihe von Informationen, wie Alter, Geschlecht, finanzielle Situation u. a. berechnen könnte, wer künftig seinen Kredit zurückzahlt oder wer wieder kriminell wird. Es darf aber davon ausgegangen werden, dass es Regeln gibt, nach denen Menschen sich entscheiden, einen Kredit zurückzuzahlen oder wieder kriminell zu werden, oder dass in den Informationen über sie Hinweise auf Umstände enthalten sind, die ihr Verhalten bedingen. Diese Regeln können mit Hilfe von Algorithmen des sogenannten maschinellen Lernens aus den Informationen über

die Personen und ihr Verhalten abgeleitet werden (s. Abbildung 2). Man gibt bspw. den Algorithmen des maschinellen Lernens Informationen über Alter und Geschlecht von Autofahrern und darüber, ob diese Personen in den letzten drei Jahren einen selbstverschuldeten Unfall hatten. Anschließend leitet der Algorithmus selbständig Entscheidungsregeln ab und teilt neu zu bewertende Personen fehlerfrei in zwei Kategorien ein: „unfallgefährdet“ oder „nicht unfallgefährdet“. Wie der Algorithmus die Entscheidungsregeln genau ableitet, hängt von der verwendeten Methode des maschinellen Lernens ab. Genauso wie der Umstand, ob sie für Menschen lesbar und einsichtig (wie im obigen Beispiel) sind oder nicht. Die abgeleiteten Regeln sind im Wesentlichen abhängig von der verwendeten Datengrundlage.

Warum handelt es sich hier um „Systeme“?

Algorithmische Entscheidungssysteme beinhalten Algorithmen an zwei Stellen: Der erste Algorithmus lernt auf Basis der Daten ein statistisches Modell. Das statistische Modell ist dann Grundlage für den (meist sehr einfachen) zweiten Algorithmus, der die eigentliche Entscheidung für eine neue Eingabe berechnet. In Abbildung 2 ist der eigentliche Entscheidungsalgorithmus in der baumartig organisierten Regelstruktur dargestellt und mündet entweder in einem Kästchen, das den neuen Versicherungsnehmer als „unfallgefährdet“ oder „nicht unfallgefährdet“ beurteilt. Das Ergebnis eines algorithmischen Entscheidungssystems ist das Produkt aus der Interaktion von Daten und dem ersten Algorithmus. Man muss also immer das Gesamtsystem betrachten, bestehend aus Daten, dem ersten Algorithmus, der das Modell erlernt, und dem Modell, das dann die Grundlage für die Entscheidung bietet.

Klassische Algorithmen vs. algorithmische Entscheidungssysteme

Dieser Ansatz, Entscheidungsregeln zu lernen, ist oft viel erfolgreicher als der Ansatz, einen klassischen Algorithmus zu entwickeln. Am Beispiel maschineller (d. h. algorithmischer) Übersetzungen von Texten lässt sich dies veranschaulichen: Hier wurden jahrzehntelang Algorithmen entwickelt, die dem alten Paradigma folgten: Gegeben ein Text in Sprache A, generiere eine Übersetzung des Textes in Spra-

che B. Aber selbst mit der Hilfe von vielen Linguisten konnten keine Algorithmen gefunden werden, um gute Übersetzungen zu generieren. Erst der radikal neue Ansatz, eine Vielzahl von Texten und ihre Übersetzungen als Datengrundlage zu nehmen und per Algorithmus daraus Entscheidungsregeln selbständig ableiten zu lassen, brachte den Durchbruch.

Allerdings können die Intransparenz der Entscheidungsregeln und die komplexen Interaktionen zwischen Daten, Algorithmen und sozialer Einbettung dazu führen, dass Entscheidungen einzelne Individuen oder die gesamte Gesellschaft schädigen. In solchen Fällen braucht es eine Überwachung ihrer Entscheidungsqualität und ihrer Einbettung in soziale Prozesse.

2. Welche Art von ADM-Systemen ist in der Transparenz- und Kontrolldiskussion relevant?

Algorithmische Entscheidungssysteme bedürfen dann einer speziellen Überwachung und Kontrolle, wenn sie Entscheidungen über Menschen treffen oder wenn ihre Entscheidungen einzelne Menschen oder die Gesellschaft als Ganzes betreffen. Die Kosten von Fehlentscheidungen trägt hier nicht nur der ADM-nutzende Akteur, sondern auch die Person, über die entschieden wird. Zu diesen sensiblen, der Überwachung und Kontrolle bedürftenden algorithmischen Entscheidungssystemen gehören folgende Beispiele (s. auch Lischka & Klingel, 2017):

1. **Googles Suchmaschinenalgorithmus:** Er lernt aus der Kombination von Informationen über eine Webseite und einen Nutzer und dem Nutzerverhalten, ob eine Webseite für eine bestimmte Suchanfrage relevant ist oder nicht.
2. **Facebooks Newsfeed:** Er lernt aus der Interaktion von Nutzern mit Inhalten, welche Informationen den Nutzern wichtig sind, und sortiert nach den gelernten Entscheidungsregeln neue Inhalte.

3. Predictive Policing beruht auf Algorithmen, die aus bisherigen Straftaten ableiten, wann und wo welche neuen Straftaten zu erwarten sind. Diese ADM-Systeme werden derzeit auch in Deutschland entwickelt und erprobt.
4. Rückfälligkeitsvorhersagealgorithmen: In den USA werden Kriminelle mit Hilfe von ADM-Systemen in Risikogruppen etwa bezüglich weiterer Straftaten kategorisiert. Die Systeme lernen von Daten über Kriminelle, die in der Vergangenheit rückfällig wurden oder resozialisiert wurden. Sie leiten daraus Entscheidungsregeln ab, die in einer Gerichtsverhandlung oder bei der Vergabe von Resozialisierungsmaßnahmen verwendet werden können.
5. Terroristenidentifikation: Aus Dokumenten, die von Snowden veröffentlicht wurden, geht hervor, dass ADM-ähnliche Methoden eingesetzt werden sollten, um mögliche terroristische Kuriere zu identifizieren, basierend auf Smartphone-Daten von 55 Millionen Einwohnerinnen und Einwohnern aus Pakistan und Afghanistan und wenigen bekannten Kurieren (Förtsch, 2015).

Grundsätzlich sind auch ADM-Systeme denkbar, die schlechte Entscheidungen über den Einsatz gesellschaftlicher Ressourcen treffen und damit einen Schaden verursachen, von dem der einsetzende Akteur nicht betroffen ist: beispielsweise ein ADM-System, das zwar den kostengünstigsten Transportweg für eine Reihe von Waren findet, der aber zugleich auch der umweltschädlichste ist. Noch ist nicht geklärt, inwieweit hier die ADM-Systeme selbst unter Aufsicht gestellt werden oder ob die Anreizstrukturen nicht eher grundsätzlich verändert werden müssen. An dieser Stelle sei angemerkt, dass falsche Anreizstrukturen mit Hilfe von Algorithmen optimal ausgebeutet werden und damit maximalen Schaden anrichten können.

Auf den Prüfstand müssen in jedem Fall ADM-Systeme, die durch Fehlurteile über Menschen diese direkt schädigen können. Um die daraus folgenden Transparenz- und Kontrollanforderungen zu erläutern, muss zunächst beschrieben werden, wie Fehlurteile entstehen können.

3. Fehlerquellen und Fehlrteile

Es gibt im Wesentlichen drei Mechanismen, die Fehlrteilen von ADM-Systemen zugrunde liegen.

1. Zufällige Faktoren: Einsatzgebiete von ADM-Systemen wie die Vorhersage künftiger Arbeits- oder Studienleistungen („people analytics“, s. z. B. Reindl & Krügl, 2017) unterliegen oft Zufälligkeiten und digital schwer zu erfassenden Dimensionen. Zufällig ist beispielsweise eine Erkrankung, die auch eine leistungsfähige Person einschränkt. Digital schwer messbar ist die Teamatmosphäre, die sich ebenfalls auf die Leistungsfähigkeit auswirkt. In diesen Fällen sind Fehlrteile fast zwangsläufig.
2. Zu kleine Datenmengen: Die Daten, von denen gelernt werden soll, sind zu inhomogen und ihre Zahl zu klein, um sie in homogenere Datenmengen aufzuteilen. So ist die automatische Identifikation von Terroristen auch deshalb so schwierig, weil es zu wenige Personen gibt, die aus denselben Gründen derselben Bewegung angehören und dort terroristisch aktiv werden. Auch hier sind Fehlrteile unvermeidbar.
3. Fehlerhafte Entwicklung oder fehlerhafter Einsatz von ADM-Systemen: Fehlrteile können aus einem fehlerhaft entwickelten ADM-System stammen oder einer Fehlinterpretation der von ihm gelieferten Resultate. Solche Fehlrteile können mithilfe strukturierter Entwicklungs- und Evaluationsprozesse minimiert werden.

Die ersten beiden Faktoren, die zu Fehlrteilen führen, sind systemimmanent. Im Falle zu kleiner Datenmengen ist vom Einsatz von ADM-Systemen abzuraten. Nach Abwägung der Kosten und des Nutzens kann der Einsatz von ADM-Systemen in Situationen, die möglicherweise von Zufällen bestimmt sind, jedoch gerechtfertigt sein. Die auf fehlerhafter Entwicklung oder dem fehlerhaften Einsatz der Systeme beruhenden Fehlschlüsse können mittels Transparenz und Kontrolle reduziert werden. Nachfolgend werden einige wenige skizziert (vgl. ausführlich: Zweig, 2018).

1. Zu treffende Entscheidungen und mögliche Fehler in der Datenbasis:
 - a. In der Eingabe gilt es, eine Auswahl von sinnvollen Informationen zu treffen. COMPAS, ein Rückfälligkeitvorhersagealgorithmus, bekommt z. B. neben den Straftaten einer Person auch die Information, ob Verwandte des Kriminellen ebenfalls in Haft waren – in Deutschland undenkbar.
 - b. Oftmals sind die Daten fehlerhaft oder werden bei der Kombination von mehreren Datenbanken nicht verlässlich der jeweiligen Person zugeordnet.
 - c. Manchmal enthalten Daten schon Diskriminierendes, etwa wenn zu Bewerbungsverfahren Frauen oder Personen mit Migrationshintergrund weniger oft eingeladen worden sind, als sie unter den Bewerbern vertreten waren. Diese Diskriminierungen werden vom Algorithmus „mitgelernt“. Wie genau „Diskriminierung“ definiert werden soll und wie sie gemessen werden kann, ist auch eine gesellschaftliche Entscheidung (Kleinberg, Mullainathan & Raghavan, 2017; Zweig & Krafft, 2018).
2. Auch auf der Ebene des maschinellen Lernens müssen Entscheidungen getroffen werden und können Fehler gemacht werden. Nicht alle Algorithmen eignen sich für jede Frage.
3. Es muss eine Wahl getroffen werden, was genau zu lernen ist. Für algorithmische Entscheidungssysteme im Bereich von Social Media müssen die Designer eine vom Computer erfassbare Definition von „Relevanz“ entwickeln. Dies stellt eine sogenannte „Operationalisierung“ eines schwer greifbaren sozialen Konstrukts dar, die mehr oder weniger gut gelingen kann.
4. Die maschinelle Entscheidung muss – wenn sie von einem Menschen begutachtet wird – in einer Form kommuniziert werden, die alle oben genannten Entscheidungen und Unsicherheiten so fassbar macht, dass der menschliche Entscheider das Resultat bewerten kann.

5. Eine wichtige Erkenntnis ist, dass ein algorithmisches Entscheidungssystem für einen bestimmten sozialen Prozess entwickelt wird und nicht ohne weitergehende Prüfung in ähnlich scheinenden sozialen Prozessen eingesetzt werden darf (Zweig & Krafft, 2018). So wurde COMPAS, der Rückfalligkeitsvorhersagealgorithmus für Kriminelle, entwickelt, um Resozialisierungsmaßnahmen bestmöglich zu verteilen. Er wird aber inzwischen – unverändert – auch vor Gericht verwendet. Es konnte nachgewiesen werden, dass es sich um unterschiedliche soziale Prozesse handelt und dass die Qualität des Systems für die Anwendung vor Gericht nicht ausreichend ist. Hier müssen weitaus höhere Ansprüche an die Qualität gestellt und andere Aspekte der Entscheidungsqualität des Systems gemessen werden (Krafft, 2017).

ADM-Systeme sind folglich fehleranfällig, und oft ist es schwierig, bei Fehlentscheidungen die dafür Verantwortlichen zu identifizieren. Die Forderung nach einer Qualitätssicherung und mehr Transparenz ist daher berechtigt.

4. Wann ist es notwendig, ADM-Systeme zu überwachen?

Die Auswirkungen von Fehlurteilen von ADM-Systemen, die Menschen bewerten, sind sehr unterschiedlich. Manche können einen hohen individuellen Schaden verursachen, werden aber nur auf wenige Personen angewendet. Andere verursachen einen geringen Schaden auf der Ebene von Individuen, aber einen hohen gesellschaftlichen Schaden.

Das Gesamtrisiko – also die Wahrscheinlichkeit für die Verwirklichung des Schadenspotenzials – ist aber ganz wesentlich noch von einer zweiten Dimension abhängig, nämlich der Leichtigkeit, mit der eine Zweitmeinung eingeholt bzw. Einspruch erhoben werden kann.

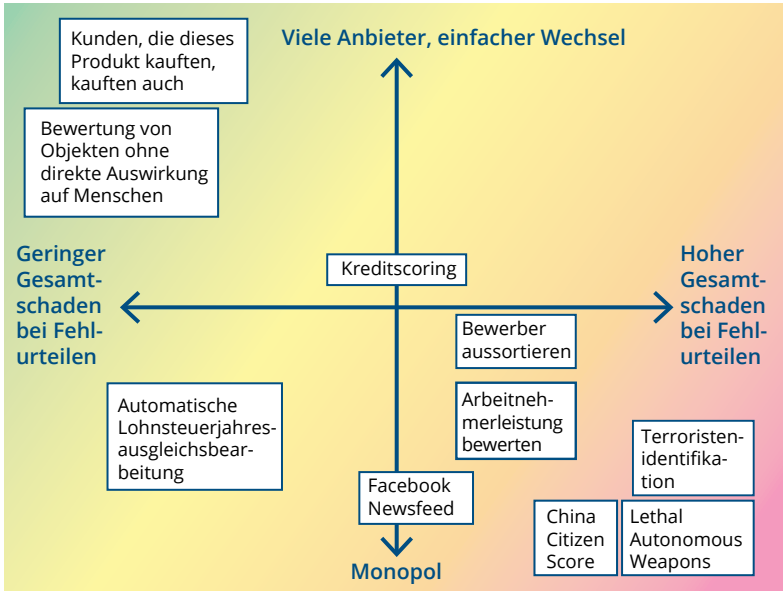


Abbildung 3: Risikomatrix mit einigen von der Autorin verorteten ADM-Systemen.

In Abhängigkeit von diesen beiden Dimensionen können verschiedene Anforderungen an Transparenz und graduell abgestufte Methoden der Kontrolle und Überwachung in Erwägung gezogen werden.

5. Mögliche Transparenzpflichten

Häufig wird mehr „Transparenz“ von algorithmischen Entscheidungssystemen gefordert. Der Begriff Transparenz wird jedoch sehr unterschiedlich interpretiert: Von der Forderung nach allgemeinen Angaben, welche Art von Informationen zum Lernen genutzt werden, über den Wunsch nach Erläuterung, wie Informationen jeweils zum Ergebnis beitragen, bis hin zur Forderung nach der Veröffentlichung des Codes. Insbesondere letztere erweist sich als wenig zielführend. Erstens ist der Softwarecode nicht lesbar wie ein Roman, und er kann beliebig kompliziert gemacht werden, wenn eine Firma ihre Firmengeheimnisse gefährdet sieht. Zweitens kann die Veröffentlichung schädlich sein: Die Veröffentlichung der Wirkweise des „PageRanks“, des damals

grundlegenden Algorithmus der Suchmaschine von Google, führte z. B. dazu, dass die Webseitenbetreiber ihr Wissen ausnutzten, um die Suchmaschine zu manipulieren, ohne der Gesellschaft eine bessere Kontrolle zu ermöglichen. In bestimmten Fällen ist daher sogar „mehr Intransparenz“ geboten, um Manipulationsrisiken zu minimieren.

Hinzu kommt, dass der Code allein nicht ausreicht, um die Wirkweise eines algorithmischen Entscheidungssystems nachzuvollziehen. Die Datengrundlage kann genauso wichtig sein, um Fehler zu erkennen. Unter Umständen muss auch transparent gemacht werden, wie genau die maschinelle Entscheidung in eine endgültige Entscheidung einfließt, etwa bei der Rückfälligkeitvorhersage durch ADM-Systeme.

Die folgenden Stufen der Transparentmachung sind sinnvoll und könnten – je nach dem Schweregrad von Fehlurteilen durch ein ADM-System – durchgesetzt werden:

1. Transparente Darstellung der eingehenden Daten und des Kriteriums, mit dem das selbstständig lernende ADM-System trainiert wurde. Dieses Vorgehen erscheint für alle ADM-Systeme sinnvoll, die Menschen kategorisieren oder ihr künftiges Verhalten beurteilen.
2. Transparente Darstellung der groben Wirkungsweise des algorithmischen Entscheidungssystems und der von ihm gelernten Entscheidungsregeln in einer verständlichen Art und Weise. Hier ist es wichtig zu beachten, dass diese Transparenzmaßnahme eine große Einschränkung darstellt, da sie die wirkungsvollsten Methoden des maschinellen Lernens ausschließt, nämlich die, deren gelernte Entscheidungsregeln für Menschen nicht mehr nachvollziehbar sind.
3. Einblick in das ADM-System für einen ausgewählten Kreis von Experten.

Abhängig von potentiellen Schäden durch Fehlurteile können auch verschiedene Stufen der Überwachung angemessen sein.

6. Überwachung von ADM-Systemen

Aus den genannten Gründen sollten Kontrolle und Überwachung nicht allein am Algorithmus des maschinellen Lernens, dem eigentlichen Entscheidungsalgorithmus oder dem ADM-System an sich ansetzen. Im Fokus muss das sozio-informatische Gesamtsystem stehen, bestehend aus dem ADM-System und allen sozialen Akteuren, die es nutzen oder von dessen Entscheidungen betroffen sind (s. Abbildung 4). Alle qualitätssichernden Maßnahmen müssen die beiden Einzelsysteme und das Gesamtsystem umfassen.

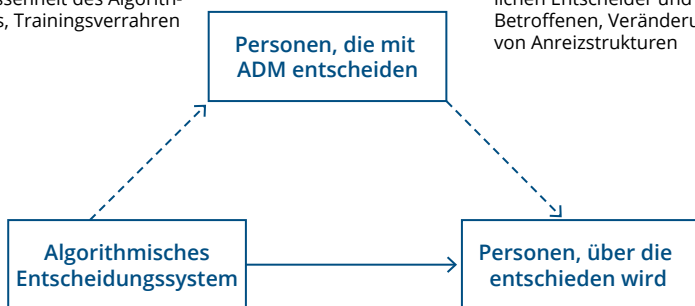
Der wichtigste Punkt, um funktionale ADM-Systeme zu erhalten, ist ein qualitätsgesicherter Entwicklungsprozess für gesellschaftlich relevante ADM-Systeme. Ein solcher muss beispielsweise regeln, wann welche Entscheidung trifft, wie diese zu dokumentieren sind und welche Qualitätsmaßnahmen ergriffen wurden. Ein solch strukturierter Prozess fehlt bisher. Genauso wichtig ist ein qualitätsgesicherter Einbettungsprozess eines gesellschaftlich relevanten ADM-Systems in den sozialen Prozess, in dem es verwendet werden soll. Dazu gehören unter anderem eine Einweisung für menschliche Entscheider, wie die Resultate der Maschine zu bewerten sind, und eine Prüfung, ob das ADM-System für den Prozess geeignet entwickelt wurde. Auch diesen strukturierten Prozess gibt es bisher nicht.

Technische Fragen:

z. B. Datenqualität, Angemessenheit des Algorithmus, Trainingsverfahren

Soziale Fragen:

z. B. Schulung der menschlichen Entscheider und der Betroffenen, Veränderung von Anreizstrukturen



Sozio-informatische Gesamtanalyse

Gesellschaftliche Fragen:

z. B. wer ist verantwortlich für Manipulationen durch Dritte, wie verändert das ADM-System die Gesellschaft?

**Reaktionen Dritter
(z. B. Propagandisten,
Medien, Gerichte,
Gesetzgeber)**

Abbildung 4:

ADM-Systeme verändern die Art und Weise, wie Entscheidungen in sozialen Prozessen gefällt werden. Je nachdem, wie transparent ihre Wirkungsweise kommuniziert wird, können sie auch gezielt genutzt werden. Sie können auch von Dritten eingeschränkt werden (z. B. Gerichten). Nicht zuletzt haben sie eine Auswirkung auf die Beurteilten, die darauf entsprechend reagieren könnten. Daher ist eine sinnvolle Evaluation der Auswirkungen des Einsatzes von ADM-Systemen nur möglich, wenn das sozio-informatische Gesamtsystem betrachtet wird.

In vielen Fällen können die Auswirkungen von ADM-Systemen ohne eine genaue Kenntnis des darunterliegenden Wirkmechanismus überwacht werden mit sogenannten Black-Box-Analysen (Krafft & Zweig, 2018). Der wesentliche Bestandteil des Schadens liegt nämlich in den Anteilen der Fehlurteile des Systems. Wenn sich die beurteilten Personen darin beobachten lassen, wie sie sich weiterhin verhalten, kann dieser Anteil berechnet werden, ohne dass man weiß oder wissen muss, wie das System zu seinem Urteil kam. Diese Beobachtung ist zum Beispiel für Suchmaschinen oder soziale Medien möglich. Diese treffen eine Auswahl von Nachrichteninhalten und können beobachten, ob

Menschen tatsächlich mit dieser Auswahl interagieren oder nicht. In diesen Fällen kann die Fehlurteilsrate gemessen werden.

Ein weiterer Gesichtspunkt, der sich gut durch eine Black-Box-Analyse überwachen lässt, ist die Frage nach Diskriminierung. Insbesondere bei algorithmischen Entscheidungssystemen im Bereich der (schulischen und beruflichen) Leistungsbewertung kann routinemäßig überprüft werden, ob Menschen nach ihrem Geschlecht, ihrer Herkunft, ihrem Alter oder weiteren Eigenschaften diskriminiert werden. Diese Überprüfung lässt sich leicht automatisieren, und sie sollte bei relevanten ADM-Systemen in Abhängigkeit von der Auswertung des möglichen Schadens gefordert werden.

Manchmal können auch ADM-spezifische Fragen mit Hilfe eines Black-Box-Ansatzes gelöst werden. Algorithmische Entscheidungssysteme stehen im Verdacht, durch starke Personalisierung Filterblasen zu verstärken oder überhaupt erst zu generieren (Eli Pariser 2011). Während dem Problem, ob sich Nutzer in einer Filterblase befinden oder nicht, methodisch schwer zu begegnen ist, kann die Frage, wie stark personalisiert die Ergebnisse einer Suchmaschine sind, sehr gut automatisiert beantwortet werden. Ein solches Projekt hat das Algorithm Accountability Lab an der TU Kaiserslautern anlässlich der Bundestagswahl 2017 als *proof of concept* durchgeführt und festgestellt, dass die Suchergebnisse bei der freiwilligen Kohorte von Teilnehmern relativ wenig personalisiert waren (Krafft et al. 2017). Dieses Projekt zeigt, dass, auch wenn ein großer kollektiver Schaden z. B. durch Filterblasenbildung möglich ist, größere Eingriffe und Kontrollen nicht notwendig sind, solange ein Aspekt wie der Personalisierungsgrad als Grundlage einer möglichen Filterblase dauerhaft und kostengünstig überwacht werden kann.

In den meisten Fällen, in denen Maschinen Menschen beurteilen, wird den Menschen infolge der Beurteilung eine Handlungsoption entzogen, ohne dass sie ein Fehlurteil nachweisen können. Wer nicht zum Bewerbungsgespräch eingeladen wird, kann nicht unter Beweis stellen, dass er ein erfolgreicher Arbeitnehmer ist, und ein nicht ver-

gebener Kredit kann nicht pünktlich zurückgezahlt werden. Hier liegt eine Grenze des Black-Box-Ansatzes.

Noch schwieriger wird es, wenn weder Betreiber noch Nutzer eines algorithmischen Entscheidungssystems angeben können, was die beste Entscheidung des Systems gewesen wäre. So etwa bei algorithmischen Entscheidungssystemen in Sozialen Medien, wo niemand weiß, was die beste Auswahl von Nachrichten gewesen wäre, aber auch bei Dating-Plattformen, bei denen keineswegs klar ist, wie man den besten Partner auswählt.

Aufgrund der geschilderten Schadensdimensionen durch Fehlurteile von ADM-Systemen müssen insbesondere ADM-Systeme der Geheimdienste, wie das genannte US-amerikanische System zur Identifikation terroristische Kuriere, intensiv kontrolliert und überwacht werden. Der individuelle Schaden für unschuldig Identifizierte ist groß. Wenn Terroristen nicht identifiziert werden, ist der gesellschaftliche Schaden u. U. riesig. Die Zahl der beurteilten Personen ist hoch; die Überwachung findet vermutlich regelmäßig statt. Solche ADM-Systeme sollten einer hohen Transparenz (unter Umständen gegenüber einem akkreditierten, aber unabhängigen Expertenkreis) unterliegen und regelmäßig auf ihre gesamtgesellschaftlichen Auswirkungen hin überprüft werden. Ähnliches gilt für ADM-Systeme, die im Krieg zur Vermeidung von *friendly fire* eingesetzt werden, die Steuererklärungen nach Unregelmäßigkeiten durchsuchen oder die automatische Bewertungen von Visa-Anträgen durchführen.

Die vorgestellten Transparenzpflichten und technischen Kontrollprozesse können nun verschiedenen Risikoklassen von ADM-Systemen mit gesellschaftlicher Relevanz zugeordnet werden.

7. Die fünf Risikoklassen von ADM-Systemen mit lernenden Komponenten

Ich schlage eine Unterteilung von gesellschaftlich relevanten ADM-Systemen in fünf verschiedene Klassen vor:

- ▶ ADM-Systeme der Klasse 0 müssen auf technischer Ebene nicht reguliert werden. Es gibt daher weder Transparenzpflichten noch die Notwendigkeit, diese ADM-Systeme ständig zu kontrollieren. Sollte ein Verdachtsfall auftreten, können *post-hoc*-Analysen angestrengt werden. Die Bewertung des Gesamtschadenspotenzials wird sich durch einen Verdachtsfall in den meisten Situationen verschlechtern und damit das System im Weiteren in einer höheren Klasse einer Regulation unterliegen. Zu dieser Klasse gehören beispielsweise Produktempfehlungssysteme für Kleidung.
- ▶ Für ADM-Systeme der Klasse 1 empfehle ich eine ständige Überwachung des Systems durch eine Black-Box-Analyse, die keinen Zugriff auf den Code benötigt. Zu dieser Klasse gehört aus meiner Sicht Googles Suchmaschine, nachdem wir 2017 gute Belege dafür finden konnten (Krafft et al., 2018), dass der Personalisierungsgrad der dahinterliegenden Algorithmen so klein ist, dass eine Filterblase nicht zu befürchten ist. Eine solche Analyse sollte ständig durchgeführt werden, um bei einer Erhöhung des Personalisierungsgrades eine Neubewertung des möglichen Schadens vornehmen zu können.
- ▶ ADM-Systeme der Klasse 2 sind so kritisch, dass sie neben der ständigen Kontrolle auch mehrere Transparenzpflichten erfüllen müssen: über die genaue Art der Eingabe, Qualität der Eingabedaten, das Qualitätskriterium und eventuelle Fairnesskriterien auch über die spezifische Einbettung des Systems in den sozialen Prozess, in dem die finale Entscheidung getroffen wird. Zu dieser Klasse gehören beispielsweise automatische oder unterstützende ADM-Systeme, die eingehende Bewerbungen auf einen Job

bewerten. Hier müsste beispielweise spezifiziert werden, welche der Daten aus den Bewerbungen verwendet wurden, mit welcher Qualität die aus den Dokumenten ausgelesen werden können und mit welcher Maßgabe das ADM-System trainiert wurde. Es könnte beispielsweise die Anzahl der nicht erfolgreichen Jobinterviews minimieren oder aber die Anzahl der nachher erfolgreich eingestellten Personen maximieren – wenn unter „Erfolg“ verstanden wird, dass eine Person für mindestens zwei Jahre im Betrieb verbleibt, könnten sich z. B. Elternzeiten negativ auswirken. Nicht zuletzt müsste unter anderem transparent kommuniziert werden, ob die Entscheidung vollautomatisch getroffen wird oder nur der Vorbereitung der Entscheidung dient und welche Widerspruchsmöglichkeiten es gibt.

- ▶ ADM-Systeme, die in Klasse 3 verortet wurden, dürfen als lernende Komponente nur noch erklärende Modelle verwenden. Der oben gezeigte Entscheidungsbaum gilt den meistens als erklärend, da Menschen einigermaßen gut nachvollziehen können, wovon die Entscheidung abhängt. Momentan gehören neuronale Netze aus dem Bereich des *Deep Learning* zu den Ansätzen, die als nicht-erklärend gelten. Hier ordnen wir beispielsweise Systeme ein, die innerhalb eines Betriebes versuchen, Arbeitnehmer nach zukünftigem Erfolg zu klassifizieren.
- ▶ Schlussendlich enthält die Klasse 4 solche Entscheidungssituationen, die aus meiner Sicht nicht durch algorithmische Entscheidungssysteme mit einer lernenden Komponente entschieden werden dürfen. Beispiele sind die automatische Tötung von vermeintlich erkannten, gesuchten Personen (*lethal autonomous weapons*), die flächendeckende Überwachung zur Bewertung bürgerlichen Verhaltens (Chinas Citizen Score) oder die Identifikation von Terroristen (Skynet).

Die fünf Risikoklassen sind in Abbildung 5 visualisiert.

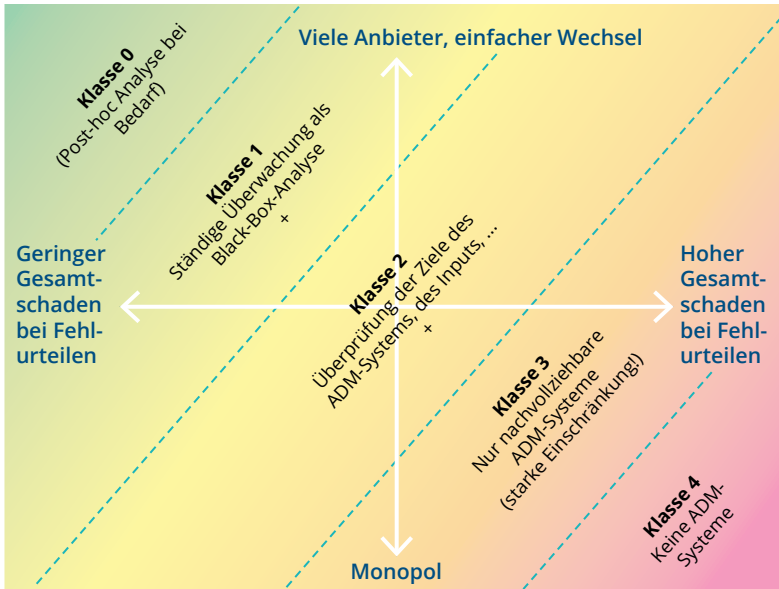


Abbildung 5: Risikomatrix, auf der gesellschaftlich relevante ADM-Systeme in eine von fünf Risikoklassen eingruppiert werden können.

8. Politische Handlungsoptionen

Es ergibt sich eine Reihe von politischen Handlungsoptionen, die im Folgenden angesprochen werden.

8.1 Gesellschaftliche Diskussion über Transparenz und Kontrolle von algorithmischen Entscheidungssystemen

Deutschland ist eine Industrienation. Daher ist es wichtig, jegliche Regulationsbemühungen auf die Fälle zu beschränken, in denen es auch tatsächlich notwendig ist. So müssen nicht alle algorithmischen Entscheidungssysteme (und schon gar nicht alle Algorithmen) durch den „Algorithmen-TÜV“, es reicht aber auch nicht aus, sich auf die derzeit öffentlich stark wahrgenommenen algorithmischen Entscheidungssysteme der Sozialen Medien zu konzentrieren. In ihrem möglichen Schaden sind staatliche ADM-Systeme, aber auch die in den Firmen zunehmend verwendeten Bewerber- und Mitarbeiter-Be-

wertungssysteme mindestens ebenso wichtig – ganz zu schweigen von tödlichen, automatischen Waffensystemen (*lethal autonomous weapons*). In diesen Bereichen ist eine konstruktiv und effizient geführte Diskussion darüber notwendig, welche Art von ADM-Systemen in welchen sozialen Kontexten welcher Form der Kontrolle und Überwachung bedürfen. Einen Vorschlag zu einer grundsätzlichen Form der Klassifikation habe ich in dieser Studie vorgestellt.

8.2 Strukturierte Entwicklungs- und Einbettungsprozesse

Für die Entwicklung sozio-informatisch relevanter ADM-Systeme bedarf es eines qualitätsgesicherten Entwicklungsprozesses und eines strukturierten Einbettungsprozesses des ADM-Systems in das soziale System, das u. a. die Schulung der Nutzer und die Evaluation der Qualität der Entscheidungen umfasst. Hier könnten Bund und Länder bei der Beauftragung und dem Kauf von öffentlicher IT mit ADM-Systemen einen wichtigen Beitrag leisten, indem sie diesen strukturierten Prozess erst entwickeln, dann den strukturierten Entwicklungsprozess bei den beauftragten Firmen fordern und den Einbettungsprozess der gekauften Software in die eigenen sozialen Prozesse strukturiert begleiten und konstant die gesamtgesellschaftlichen Auswirkungen evaluieren.

8.3 Ausbildung, Berufsbilder und Berufsethiken

Es bedarf entsprechend qualifizierter Personen, die die notwendigen Fragen stellen und die die Qualität des Prozesses evaluieren können. Ein Studiengang, der solche Personen ausbildet, ist die (bisher in Deutschland einzigartige) Sozioinformatik an der TU Kaiserslautern.

Eine weitere Berufsgruppe, die hauptsächlich an der Entwicklung solcher ADM-Systeme beteiligt ist, sind *data scientists*. Das Berufsbild ist so neu, dass es bisher weder ein klares Curriculum gibt noch eine Berufsethik. Die meisten *data scientists* sind Quereinsteiger aus der Physik, Mathematik oder Informatik, mit großer Kompetenz in Statistik und Programmierung, denen jedoch in der Mehrzahl eine gesellschaftswissenschaftliche Ausbildung fehlt. Hier sind eine gezielte Förderung von Professuren im Bereich „data science“ und ein verbindliches Curriculum wichtig.

Zudem benötigen wir flächendeckend an allen Universitäten und Hochschulen, die Informatik anbieten, die früher oft anzutreffenden, inzwischen aber nahezu ausgestorbenen Professuren im Bereich „Informatik und Gesellschaft“.

Nicht zuletzt bedarf es auch einer gesonderten Berufsethik und vielleicht auch eines Akkreditierungsprozesses aller Personen, die die Inspektion von sozio-informatisch relevanten ADM-Systemen vornehmen sollen, um Bedenken der Firmen bezüglich ihrer Firmengeheimnisse Rechnung zu tragen.

Referenzen

F Förtsch, Michael. NSA-Überwachung: Skynet ist real und ziemlich unheimlich, Wired online, 11.5.2015, <https://www.wired.de/collection/tech/das-nsa-programm-skynet-soll-terroristen-identifizieren> (letzter Abruf: 12.11.18).

K Kleinberg, Jon, Mullainathan, Sendhil & Raghavan, Manish. Trade-Offs in the Fair Determination of Risk Scores, Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS'17), 2017, 43:1–43:23.

Krafft, Tobias D. & Zweig, Katharina A. Wie Gesellschaft algorithmischen Entscheidungen auf den Zahn fühlen kann, in: (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft (Hrsg. Resa Mohabbat Kar, Basantha Thapa, Peter Parycek), Kompetenzzentrum Öffentliche IT, 2018, 471–492.

Krafft, Tobias D., Gamer, Michael, Laessing, Marcel & Zweig, Katharina A. Filterblase geplatzt? Kaum Raum für Personalisierung bei Google-Suchen zur Bundestagswahl, 1. Zwischenbericht des Datenspende-Projektes, 2017, https://algorithmwatch.org/wp-content/uploads/2017/09/1_Zwischenbericht___final.pdf (letzter Abruf: 12.11.18).

Krafft, Tobias D. Qualitätsmaße binärer Klassifikationen im Bereich kriminalprognostischer Instrumente der vierten Generation, Masterarbeit an der TU Kaiserslautern, 2017, <https://arxiv.org/abs/1804.01557> (letzter Aufruf: 12.11.18).

L Lischka, Konrad und Klingel, Anita. Wenn Maschinen Menschen bewerten – Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung, Studie der Bertelsmann Stiftung, Mai 2017, https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/ADM_Fallstudien.pdf (letzter Abruf: 12.11.18).

- M** **Mayer-Schönberger, Cukier, Viktor & Kenneth.** Big Data: Die Revolution, die unser Leben verändern wird, Redline Verlag, München, 2013.
- P** **Pariser, Eli.** Filter Bubble: Wie wir im Internet entmündigt werden, Carl Hanser Verlag München, 2011.
- R** **Reindl, Cornelia & Krügl, Stefanie.** People Analytics in der Praxis, Haufe-Lexware GmbH & Co. KG. Freiburg, 2017.
- Z** **Zweig, Katharina A., Lischka, Fischer, Konrad & Dr. Sarah.** Wo Maschinen irren können, Studie der Bertelsmann Stiftung, Februar 2018, <https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf> (letzter Abruf: 12.11.18).
- Zweig, Katharina A. & Krafft, Tobias D.** Fairness und Qualität algorithmischer Entscheidungen, in: (Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft (Hrsg. Resa Mohabbat Kar, Basantha Thapa, Peter Parycek), Kompetenzzentrum Öffentliche IT, 2018, 204–227.

Autorinnen, Autoren und Herausgeber

Dr. Norbert Arnold

Leiter des Teams Bildungs- und Wissenschaftspolitik,
Hauptabteilung Politik und Beratung, Konrad-Adenauer-Stiftung

Prof. Dr. Oliver Brock

Robotics and Biology Laboratory,
Technische Universität Berlin

Prof. Dr. Joachim Fetzer

Volkswirt und Theologe,
Vorstandsmitglied des Deutschen Netzwerks Wirtschaftsethik

Prof. Dr. Elisabeth Gräß-Schmidt

Lehrstuhlinhaberin für Systematische Theologie und Ethik,
Direktorin des Instituts für Ethik an der Evangelisch-theologischen
Fakultät der Eberhard-Karls-Universität Tübingen

Prof. Dr. Richard Lackes

Lehrstuhl für Wirtschaftsinformatik,
Technische Universität Dortmund

Dr. Norbert Lossau

Wissenschaftsjournalist,
Mitglied des Beirats der Wissenschaftspressekonferenz WPK

Prof. Dr. Catrin Misselhorn

Direktorin des Instituts für Philosophie,
Lehrstuhl für Wissenschaftstheorie und Technikphilosophie,
Universität Stuttgart

Dr. Jan-Erik Schirmer

wissenschaftlicher Assistent und Habilitand,
Humboldt-Universität zu Berlin

Prof. Dr. Jürgen Schmidhuber

Wissenschaftlicher Direktor des Schweizer KI-Labors IDSIA, Lugano, Schweiz; Professor für Künstliche Intelligenz (Ordinarius) an der Fakultät für Informatik der Universität Lugano (USI); Leiter des Labors für Kognitive Robotik an der Fakultät für Informatik der Technischen Universität München, Professor SUPSI in Manno (Schweiz), Mitbegründer, NNAISENSE, 2014, Präsident 2014–2017, Chefwissenschaftler, seit 2017.

Prof. Dr. Jens Südekum

Universitätsprofessor für Internationale Volkswirtschaftslehre am Düsseldorfer Institut für Wettbewerbsökonomie (DICE), Heinrich-Heine-Universität.

Tobias Wangermann

Leiter des Teams Digitalisierung,
Hauptabteilung Politik und Beratung Konrad-Adenauer-Stiftung

Prof. Dr. Katharina A. Zweig

Leiterin des Algorithm Accountability Labs
des Fachbereichs Informatik,
Technische Universität Kaiserslautern

Müssen wir vor Künstlicher Intelligenz Angst haben?
Nein, das müssen wir nicht. Aber: Wie alle Technologien hat KI Licht- und Schattenseiten. Notwendig ist eine Reflexion über die Vor- und Nachteile – und über die Ziele, die wir mit KI erreichen wollen.

Die Texte des vorliegenden Bandes thematisieren verschiedene Anwendungsfelder von Digitalisierung und Künstlicher Intelligenz. Und sie greifen die damit zusammenhängenden ethisch-rechtlichen Fragen auf. Sie markieren also Orientierungspunkte in der digitalen Welt.