

Isabella Hermann

Frauke Rostalski

Günter Stock

2|2020

#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen

Kompetent eigene Entscheidungen treffen? Auch mit Künstlicher Intelligenz!

Eine Schriftenreihe der interdisziplinären Arbeitsgruppe

Verantwortung: Maschinelles Lernen und Künstliche Intelligenz



berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)

KOMPETENT EIGENE ENTSCHEIDUNGEN TREFFEN?
AUCH MIT KÜNSTLICHER INTELLIGENZ!



**KOMPETENT EIGENE ENTSCHEIDUNGEN TREFFEN?
AUCH MIT KÜNSTLICHER INTELLIGENZ!**

Isabella Hermann
Frauke Rostalski
Günter Stock

AUTORINNEN UND AUTOREN

Isabella Hermann: Wissenschaftliche Koordinatorin der interdisziplinären Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften.

Frauke Rostalski: Professorin für Strafrecht, Strafprozessrecht, Rechtsphilosophie und Rechtsvergleichung an der Universität zu Köln und Mitglied der Jungen Akademie.*

Günter Stock: Vorstandsvorsitzender der Einstein Stiftung Berlin.*

* Mitglied der interdisziplinären Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften.

Herausgeberin: Interdisziplinäre Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* der Berlin-Brandenburgischen Akademie der Wissenschaften.

Redaktion: Isabella Hermann, Ute Tintemann

Grafik: Thorsten Probst/angenehme gestaltung

Druck: bud Brandenburgische Universitätsdruckerei und Verlagsgesellschaft Potsdam mbh

© Berlin-Brandenburgische Akademie der Wissenschaften, 2020

Jägerstraße 22–23, 10117 Berlin, www.bbaw.de

Nachdruck, auch auszugsweise, nur mit ausdrücklicher Genehmigung der Herausgeberin.

ISBN: 978-3-939818-90-8

INHALTSVERZEICHNIS

Einführung	7
Christoph Marksches	
Entscheiden im digitalen Zeitalter – Zur Bedeutung der technischen Beeinflussung des Menschen bei der Entscheidungsfindung	9
Frauke Rostalski	
Kompetenzverlust in Zeiten von KI – Wie bewahren Menschen wichtige Fähigkeiten?	24
Isabella Hermann und Günter Stock	

EINFÜHRUNG

Der Alltag vieler Menschen ist inzwischen ebenso tief wie nachhaltig durch Anwendungen von Künstlicher Intelligenz (KI) geprägt. Einer der Gründe dafür ist, dass viele dieser Anwendungen einfach praktisch sind: Navigation via Google Maps, Buchempfehlungen bei Amazon oder Übersetzungen mit DeepL erleichtern den Alltag. Fahrassistenzsysteme oder medizinische Diagnoseverfahren können sogar Leben retten. Doch selbst wenn die Anwendung von KI in den zuerst genannten Programmen eher harmlos wirkt und in den danach genannten Beispielen sogar aus ethischen Gründen angeraten scheint, ist in vielen Fällen für die, die solche Anwendungen nutzen, nicht nachvollziehbar, warum sie einen bestimmten Vorschlag oder eine spezifische Diagnose erhalten. Außerdem laufen Menschen durch die Unterstützung von automatisierten Systemen Gefahr, wichtige Fähigkeiten und Kompetenzen zu verlieren, um eigene Entscheidungen treffen und reflektiert handeln zu können. Man hat es also mit dem klassischen Problem ambivalenter Wirkungen einer technischen Revolution zu tun, allerdings in deutlicher verschärfter Form.

Die Mitglieder der interdisziplinären Arbeitsgruppe (IAG) „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz“ der Berlin-Brandenburgischen Akademie der Wissenschaften gehen der Frage nach, welche Folgen sich für die Verantwortung des Menschen und die theoretische Konzeption solcher Verantwortung ergeben, wenn das Leben zunehmend von algorithmischen Entscheidungssystemen beeinflusst wird, deren Verfahrensweisen selbst die, die sie konstruiert haben, nicht mehr nachvollziehen können. Die Frage wird umso brisanter, wenn es möglich ist, dass Menschen durch KI-Systeme benachteiligt werden oder sogar körperlich zu Schaden kommen können. Denn wer trägt die ethische und rechtliche Verantwortung, wenn der Fahrassistent eine Fußgängerin übersieht oder das medizinische Diagnosesystem falsche Vorschläge macht? Die Person, die die Daten auswählt, diejenige, die den Algorithmus programmiert, oder die, die das System nicht gut genug überwacht hat? Durch den Einsatz immer komplexer werdender technischer Systeme diffundieren die Verantwortlichkeiten. Wer kann und muss den Einsatz eines KI-Systems überprüfen und regulieren?

Zentrale Bereiche und Funktionen der Gesellschaft wie das Rechtssystem gehen von einem Bild des Menschen als einem rational handelnden Subjekt aus, das Freiheit besitzt und daher Verantwortung für sein Tun übernehmen muss. Daher

kommt Menschen nicht nur ethische und rechtliche Verantwortung für ihr Tun zu, sondern technische Systeme müssen so gestaltet werden, dass weiterhin Menschen Verantwortung übernehmen können.

In diesem Heft der Publikationsreihe „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen“ mit dem Titel „Kompetent eigene Entscheidungen treffen? Auch mit Künstlicher Intelligenz!“ setzen sich die Autorinnen und Autoren mit solchen basalen Fragen auseinander, die zugleich von großer Bedeutung für den praktischen Umgang mit entsprechenden Anwendungen sind. Frauke Rostalski geht in ihrem Beitrag „Entscheiden im digitalen Zeitalter – Zur Bedeutung der technischen Beeinflussung des Menschen bei der Entscheidungsfindung“ auf die Frage ein, ob die versprochenen Erleichterungen durch die Delegation von Entscheidungen auf technische Systeme tatsächlich einen Freiheitszuwachs bedeuten oder eher einen Verlust an individueller Freiheit. Sie beleuchtet, inwiefern es gefährliche Auswirkungen auf den Einzelnen und die Gesellschaft haben kann, wenn Menschen zunehmend von algorithmischen Entscheidungssystemen abhängig werden. Daran anknüpfend legen Isabella Hermann und Günter Stock in ihrem Beitrag „Kompetenzverlust in Zeiten von KI – Wie bewahren Menschen wichtige Fähigkeiten?“ den Fokus auf die Mensch-Maschine-Interaktion und das sogenannte Automationsparadox. Dieses Paradox beschreibt den Widerspruch, dass zunehmend Prozesse automatisiert werden, weil Maschinen dem Menschen als überlegen angesehen werden. Trotzdem müssen Menschen bei Fehlern der Maschine einspringen, obwohl sie mitunter kaum mehr die Kompetenzen besitzen, um die Fehler beheben zu können. Die leitende Frage ist, was getan werden kann, damit Menschen weiterhin verantwortlich handeln können.

Ich möchte erneut allen Mitgliedern der interdisziplinären Arbeitsgruppe für ihre engagierte Mitarbeit danken und insbesondere den drei Autoren für ihre Beiträge zu dem vorliegenden Heft der Reihe „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen“.

Christoph Marksches

Sprecher der interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz“ und designierter Präsident der BBAW

ENTSCHEIDEN IM DIGITALEN ZEITALTER – ZUR BEDEUTUNG DER TECHNISCHEN BEEINFLUSSUNG DES MENSCHEN BEI DER ENTSCHEIDUNGSFINDUNG¹

I. EINLEITUNG

Im Zeitalter der Digitalisierung werden neue Technologien zunehmend dafür eingesetzt, den Menschen in unterschiedlichen Lebensbereichen bei seiner Entscheidungsfindung zu unterstützen. Hiermit gehen Erleichterungen einher, die das Versprechen beinhalten, durch eine (vermeintliche) Delegation von Entscheidungen auf technische Systeme mehr Freiheit zu gewinnen, die sinnvoller als bislang gestaltet werden kann. Doch führt eine solche Entlastung tatsächlich zu einem Freiheitszuwachs oder bewirkt sie in Abhängigkeit von dem Bereich, der der Verfügungsmacht technischer Systeme überantwortet wird, nicht gar das Gegenteil: den Verlust an Freiheit, der in letzter Konsequenz gefährliche Auswirkungen auf den Einzelnen sowie das gesellschaftliche Miteinander haben kann? Der Beitrag versucht, hierauf eine Antwort zu finden.

II. ZUNAHME UND NUTZEN TECHNISCHER ENTSCHEIDUNGSGEHILFEN IM DIGITALEN ZEITALTER

Digitale Technologien dienen zunehmend der Unterstützung des Menschen in seinen alltäglichen privaten sowie beruflichen Verrichtungen und Entscheidungen. Beispielhaft zu denken ist nur an die Auswahl von Nachrichteninhalten, die dem Einzelnen auf verschiedenen sozialen Plattformen individualisiert angeboten werden. Auf diese Weise kommt es zu einer Reduktion des verfügbaren Angebots, die es dem Betreffenden erleichtert, unter der schier unüberschaubaren Masse

vorhandener Nachrichten eine Entscheidung zu treffen.² Ähnliches kann sich auch im Bereich des Lebensmittelkonsums ereignen. Intelligente Kühlschränke können feststellen, dass es ihrem Besitzer an Milch und Butter fehlt, und selbstständig eine Bestellung bei einem Supermarkt aufgeben. Auf diese Weise wird dem Inhaber des „klugen“ Kühlschranks die Entscheidung abgenommen, ob er in den nächsten Tagen auf eben jene Produkte verzichten oder möglicherweise einen Marken- oder Anbieterwechsel vollziehen möchte. Mit dem Voranschreiten der technischen Entwicklungen sind unsere Hilfsmittel mehr und mehr dazu in der Lage, eine Aussage über unsere Bedürfnisse zu treffen. So ist denkbar, dass die „intelligente“ Armbanduhr künftig Vorschläge unterbreitet, wie sich ihr Träger am besten ernähren soll – oder gleich selbst die Äpfel im Supermarkt bestellt. Daneben können auf Algorithmen³ basierte Analysecomputer dem Arzt die Einschätzung abnehmen, ob ein Röntgenbild einen gefährlichen Tumor zeigt oder nicht. Nicht zuletzt schlagen Dating-Apps ihren Nutzern Personen vor, die auf der Basis einer Rechenoperation *optimal* zu ihnen passen, und erleichtern ihnen auf diese Weise die mitunter beschwerliche Partnersuche.

Die Liste technischer Entscheidungsgehilfen, die bereits heute unser Leben zu einem nicht geringen Umfang prägen, ließe sich fortsetzen. In den Vordergrund rückt allerdings die Frage, welcher Nutzen für den Menschen damit einhergeht, sich bei seiner Entscheidungsfindung durch Techniken der Digitalisierung unterstützen zu lassen. Dabei muss berücksichtigt werden, dass der Einzelne in sämtlichen Bereichen seines Handelns und Denkens auf das Treffen von Entscheidungen angewiesen ist. Der Prozess des Entscheidens geht jedweder menschlichen Aktion voraus: Bevor ein bestimmter Weg eingeschlagen wird, müssen die unterschiedlichen Varianten gegeneinander abgewogen und teilweise abgelehnt werden. Insofern bedeutet jede Entscheidung jedenfalls hinsichtlich der bis dahin bestehenden Optionen eine Reduktion. Dies erfüllt für den Menschen eine wichtige

- 2 Die Algorithmen basierte Reduktion von Nachrichteninhalten durch soziale Netzwerke oder andere Intermediären geht in der Regel einher mit der Entstehung sogenannter Filterblasen: Dem Betroffenen werden Angebote etwa auf der Basis seiner früheren Interessen gemacht, die er durch eine eigene Auswahl von Nachrichtenbeiträgen zum Ausdruck gebracht hat. Filterblasen bedeuten eine Gefahr für die freie Meinungsbildung des Einzelnen siehe Paal, Boris (2018): Vielfaltssicherung bei Intermediären, in: MultiMedia und Recht (MMR), S. 567ff; Paal, Boris/Hennemann, Moritz (2017): Meinungsvielfalt im Internet, in: Zeitschrift für Rechtspolitik (ZRP), S. 76 ff. Der vorliegende Beitrag richtet den Fokus indessen auf den Themenkomplex der Entscheidungsfindung, weshalb der vorgenannte Aspekt keine nähere Erläuterung findet.
- 3 Bei einem Algorithmus handelt es sich um die Beschreibung eines allgemeinen Verfahrens, das unter Verwendung elementarer Verarbeitungsschritte auf die Lösung einer bestimmten Aufgabe gerichtet ist, siehe Kaboth, Daniel/Spies, Benjamin (2019): Urheberrechtsgesetz, § 69a Rn. 12, in: BeckOK Urheberrecht [Stand: 15.04.2019].

Funktion, bliebe er doch in einem Zustand des Hin- und Hergerissenseins unweigerlich handlungsunfähig.

Auf die Frage, welchen Vorteil es bringt, sich im Prozess der Entscheidungsfindung neuer Techniken der Digitalisierung zu bedienen, können im Wesentlichen zwei Antworten formuliert werden. Zum einen soll durch technische Unterstützung eine *Verbesserung* der Qualität der vom Menschen zu treffenden Entscheidungen erzielt werden. Dem liegt die Idee zugrunde, dass eine breite Datengrundlage besonders gut geeignet ist, die Qualität künftiger Entscheidungen zu gewährleisten. Zum anderen geht mit der technischen Entscheidungshilfe das Versprechen einher, dass der Mensch sich künftig Sinnvollerem widmen könne. Indem Teile der von ihm zu treffenden Entscheidungen gewissermaßen an eine Maschine delegiert werden, erwachse ihm ein Mehr an Freiheit, das er für Dinge nutzen kann, die für ihn eine gesteigerte Bedeutung haben.

III. VERBESSERUNG DER QUALITÄT VON ENTSCHEIDUNGEN DURCH TECHNISCHE HILFSMITTEL?

Sind technische Helfer dazu geeignet, die Entscheidungen des Menschen zu verbessern? Hierfür könnte sprechen, dass der eine oder andere Algorithmus in der Tat über bessere Informationen über den Einzelnen verfügt als er selbst: Anders als der Mensch vergisst er nichts und kann daher bei der Ermittlung der Vorzugswürdigkeit einer spezifischen Handlungsoption unter Umständen auf eine breitere Datengrundlage zurückgreifen als der Betreffende selbst. Zudem sind Algorithmen weder faul noch unvernünftig. Der eine oder andere weiß zwar, dass er mehr Obst essen sollte, verdrängt dies aber erfolgreich. Wenn die Äpfel allerdings automatisch – veranlasst durch ein entsprechendes technisches Programm etwa in der Armbanduhr oder im Kühlschrankschrank – ins Haus geliefert werden, erweist sich ein solcher Umgang mit der Ernährungsfrage zumindest als deutlich schwieriger. Auch im Hinblick auf die Partnersuche versprechen auf Algorithmen basierende Apps eine Reduktion der üblicherweise mit diesem Unterfangen einhergehenden Frustrationen, indem nämlich weniger „Frösche geküsst“ werden müssen, als unbedingt erforderlich – jedenfalls in der Theorie.⁴

4 So wirbt das Unternehmen Parship etwa mit folgendem Slogan für seine Dienste: „Wir eröffnen Ihnen neue Möglichkeiten bei Ihrer Partnersuche – und erhöhen die Chance deutlich, dass Sie den Menschen treffen, der wirklich zu Ihnen passt.“, siehe: <https://www.parship.de/tour/mitgliedschaft/> [Stand: 05.05.2020]

1. Partner-Apps als ungeeignetes Instrument zur Erfüllung des Wunschs nach Liebe

Allerdings ist der Nutzen von Partner-Apps unter dem Gesichtspunkt der Verbesserung der menschlichen Entscheidungsfindung grundlegend anzuzweifeln. Sofern es dabei darum geht, die *Liebe* zu finden, scheint die Herangehensweise entsprechender algorithmischer Systeme im Kern verfehlt. Dem liegt die Einsicht zugrunde, dass die Erfüllung der erotischen Liebe keineswegs das bloße Auffinden des *richtigen* Partners voraussetzt.⁵ Hierbei handelt es sich vielmehr allenfalls um den ersten Schritt, der in seiner Bedeutung hinter der eigentlichen Kunst des Liebens verblasst, die sich an eine Phase der Verliebtheit anschließt und die durch Geduld, Konzentration und vor allem die Überwindung des eigenen Narzissmus eingeübt und praktiziert wird. Es liegt nahe, dass derjenige, der eine Partner-App zur Erfüllung der Liebe bemüht, keine Einsichten in diese Zusammenhänge erlangt hat. Dominant ist hier vielmehr ein anderes Motiv, das als wesentliches Charakteristikum unserer Zeit eingestuft werden kann: die Konsumsucht. Bei der Dating-App werden Menschen „geshoppt“ wie Kleider von der Stange. Dass sich diesem Vorgang ein ernsthaftes Bemühen um das Praktizieren der Kunst des Liebens anschließt, wird zumindest nicht der häufigste Fall sein. Wer wie viele Menschen danach strebt, individuelles Glück durch möglichst geringen Eigenaufwand zu erreichen, wird sich hingegen nach dem Scheitern der App-gestützten Partnerschaft wieder auf den Markt der Eitelkeiten begeben – und sein einsames Herz neuerlich der Industrie der Partner-Apps anvertrauen.

2. Lediglich begrenztes „Wissen“ „intelligenter“ Maschinen

Und auch in anderen Bereichen ist die Annahme kritisch zu sehen, dass durch technische Systeme eine Verbesserung der menschlichen Entscheidungsfindung herbeigeführt werden könne. Dabei muss zunächst die Prämisse kritisch hinterfragt werden, dass „intelligente“ Maschinen tatsächlich besser wissen, was der einzelne Mensch will bzw. wollen sollte. So fußt diese Annahme auf der Vorstellung, dass frühere Entscheidungen der Person tatsächlich uneingeschränkt ausschlaggebend für deren späteres Wollen sind. Denn in der Tat kann sich ein auf einem Algorithmus beruhendes System ausschließlich auf solch frühere

5 Fromm, Erich (1999): Die Kunst des Liebens, in: Funk, Rainer (Hrsg.): Erich Fromm Gesamtausgabe in zwölf Bänden, Band IX, Sozialistischer Humanismus und Humanistische Ethik, München, S. 439 ff.

Willensbekundungen des Betreffenden beziehen, wenn es Handlungsvorschläge für die Zukunft trifft. So werden Letztere aus einem großen Schatz an Daten abgeleitet, die die Maschine über den Menschen gesammelt hat. Beispielhaft könnte das System auf sämtliche Informationen Zugriff haben, die eine Person in der Vergangenheit im Internet über sich offenbart hat. Eine Analyse ihres Zalando-Accounts ließe etwa die Schlussfolgerung zu, dass die Person besonders häufig schwarze Kleider kauft – sie diese also mit einer gewissen Wahrscheinlichkeit gerne und oft trägt. Eine über den Internetfußabdruck des Einzelnen verfügende Maschine würde die Onlinezeitungen und Artikel kennen, die er liest und gelesen hat. Sie könnte daraus ableiten, welche Themen diese Person besonders interessieren und unter Umständen schlussfolgern, welche politische Gesinnung dem zugrunde liegt. Weil etwa Kinderfilme im iTunes-Store häufig geladen werden, wüsste die Maschine, dass der Betreffende mit einer gewissen Wahrscheinlichkeit ein Kind hat – schaut es lieber Batman als Bibi, handelt es sich dabei eher um einen Jungen als um ein Mädchen. Dazu passt es dann auch, dass die häufigste Söckchenfarbe, die diese Person kauft, blau ist. Als umfassender Nutzer der Vorzüge des Internets hätte die Zielperson selbstverständlich einen Account bei einem „sozialen Netzwerk“, sodass auch die „intelligente“ Maschine über deren Freunde, Freizeitinteressen etc. Bescheid wüsste.

Doch weiß die Maschine dann auch, was der Betreffende will? Hieran bestehen erhebliche Zweifel. Diese gründen sich insbesondere auf den Umstand, dass der Mensch selbst nicht immer weiß, was er will. Vielleicht ist die eine oder andere heute sicher, dass sie schwarze Kleider am liebsten trägt. Doch wenn ihr morgen eine gute Freundin ein rotes Kleid schenken sollte, ist es nicht ausgeschlossen, dass sie hierfür eine Vorliebe entwickelt. So ist der Mensch sich selbst in Teilen stets ein Geheimnis. Zum anderen stört an der Vorstellung, durch eine bloße Analyse vergangener Entscheidungen die künftigen besonders gut treffen zu können, dass auf diese Weise jedwedem Entwicklungspotential „das Wasser abgegraben“ wird. Unter dem Stichwort der sogenannten „Echokammern“ bzw. „Filterblasen“ ist uns dieses Phänomen schon seit Längerem bekannt.⁶ Gemeint ist, dass ein Verharren in altbekannten Strukturen die Offenheit gegenüber Neuem und Andersartigem abnehmen lässt. Letztlich bewegt sich der Betreffende stets in den gleichen Bahnen, ohne dabei noch einen „Blick über den Tellerrand“ zu erhaschen. Dies ist in erster Linie ziemlich langweilig. So gingen dem Menschen auf diese Weise viele Dimensionen des Lebens verloren.

6 Paal, Boris/Hennemann, Moritz (2017): Meinungsvielfalt im Internet, in: Zeitschrift für Rechtspolitik (ZRP), S. 76.

Festgehalten werden kann also: Die Datenmaschine weiß nicht, was der Mensch *will*, sondern lediglich, was er einmal *wollte*. Sie spekuliert im Prozess ihrer datenbasierten Entscheidungsfindung darauf, dass er auch künftig wünscht, wofür er sich einst entschieden hat – also immer wieder dasselbe will. Hiermit geht mitnichten eine zwingende Verbesserung der menschlichen Entscheidungsfindung einher, im Gegenteil: Sofern dem Einzelnen seitens eines Computersystems bestimmte Vorschläge etwa in Bezug auf einen Zeitungsartikel, ein Kleidungsstück oder einen Lebenspartner gemacht werden, kann bei ihm der Eindruck entstehen, dass es sich hierbei – nicht zuletzt aufgrund des verbreiteten Glaubens an die Überlegenheit neuer Technologien – um die *richtige* Handlungsoption handelt. Diese wird im schlimmsten Fall seitens des Betreffenden überhaupt nicht hinterfragt, sondern schlicht als richtig und bindend hingenommen. Ein Abweichen von Früherem, eine Entwicklung der eigenen Persönlichkeit in bislang nicht genutzten Dimensionen ist dann aber von vornherein nicht möglich. Zwar ist nicht ausgeschlossen, dass der Einzelne ohne Zuhilfenahme technischer Systeme seinerseits zu keinem Zeitpunkt eine Ausdehnung der eigenen Perspektive vorgenommen und damit von seinen früheren Entscheidungen abweichende Handlungsoptionen abgewogen bzw. ggf. sogar gewählt hätte. Darüber hinaus schließen technische Entscheidungsgehilfen nicht aus, dass der Unterstützte seine Entscheidung nicht doch von anderen Faktoren abhängig macht. Unter dem Gesichtspunkt der Verbesserung menschlicher Entscheidungen ist dies indessen irrelevant: Weil technische Systeme dies per se nicht anbieten, fehlt es ihnen an einer für den Menschen relevanten Auswahloption, die prinzipiell mit Vorzügen gegenüber dem bislang Dagewesenen einhergehen kann. Die von dem Algorithmus unterbreiteten Vorschläge erweisen sich als verkürzt und sind damit allenfalls dazu geeignet, frühere Entscheidungen optimal zu imitieren – nicht aber künftige Entscheidungen zu verbessern.

3. Begrenztes Verbesserungspotential technischer Entscheidungshilfen selbst in eher statischen Anwendungsbereichen

Jedoch erweist sich nicht jeder Lebensbereich des Menschen als prinzipiell besonders offen gegenüber Neuem. Während Konstanz im Hinblick auf die Auswahlentscheidung im Bereich der eigenen Kleidung oder sonstiger Konsumgüter lediglich individuell eine hohe Bedeutung haben kann, erweist sie sich in anderen Zusammenhang als allgemein bedeutsam. Zu denken ist beispielsweise

an Entscheidungen, die zum Schutz eigener oder fremder Interessen auf der Basis von früheren Erfahrungswerten getroffen werden. Im medizinischen Bereich ist es im Interesse der Patienten grundsätzlich geboten, dass die Entscheidung der maßgeblichen Akteure früheres Wissen über Wirkzusammenhänge oder für die Diagnose relevante Merkmale einbeziehen. Vor diesem Hintergrund können sich technische Systeme in diesem Feld als besonders vorteilhaft erweisen, da hier unter Umständen Informationen berücksichtigt werden können, die der Einzelne selbst ggf. (noch) nicht hat. Auch kann sich in diesem Kontext ein Vergessen als gefährlich erweisen. Eben jenes Risiko bergen technische Systeme allerdings gerade nicht. Darüber hinaus kann die auf Wissen basierende Kompetenz von Menschen in Abhängigkeit von individuellen Faktoren wie etwa Müdigkeit unterschiedlich ausgeprägt sein. Auch diese „Schwäche“ lässt sich durch ein stets gleichförmig funktionierendes technisches System vermeiden. Sofern daher beispielsweise auf Algorithmen basierende Hilfstechniken zur Krebsanalyse von Röntgenbildern entwickelt werden, die eine hohe Richtigkeitsgewähr aufweisen, ist hierin prinzipiell eine Möglichkeit zur Verbesserung menschlicher Entscheidungen zu sehen.

Dieser Befund lässt sich nicht allein in einem vorrangig naturwissenschaftlich geprägten Umfeld aufzeigen. Auch das Rechtssystem ist auf Konstanz angewiesen, soweit es um die Gleichbehandlung sachlich gleich gelagerter Sachverhalte geht.⁷ Aus Gründen der Rechtssicherheit würde es sich als erheblicher Mangel eines Rechtssystems erweisen, sofern es ohne größeren Aufwand tiefgreifend verändert werden könnte. In der Folge würde auch hier eine Technologie zur Unterstützung der Rechtsgewinnung dem Grunde nach einen Mehrwert mit sich bringen: Auf diese Weise könnte die Wahrung des Grundsatzes der Gleichbehandlung sowie die Einhaltung des geltenden Rechts garantiert werden. Ein auf der Basis des geltenden Rechts programmierter Rechtsfindungsautomat ließe keinen Raum für Unschärfe, die durch menschliche Subjektivität mitunter verursacht werden kann. Gleichwohl erweisen sich technische Entscheidungshilfen auch in diesen Lebensbereichen allenfalls in eingeschränktem Umfang zur Verbesserung menschlicher Entscheidungen als geeignet. Grund hierfür ist zum einen, dass selbst in vermeintlich statischen Feldern eine Offenheit gegenüber neuen Erkenntnissen und Entwicklungen erforderlich ist, um Verbesserungen zu erreichen. Dies gilt sowohl für die Medizin als auch das Recht. Wie jede andere Wissenschaft befinden sich beide in einem kontinuierlichen Prozess der Verbesserung, die insbesondere durch Forschung erreicht werden kann. So ist die Erlangung weiterführender

7 Zum allgemeinen Gleichheitssatz siehe Kirchhof, P. (2018): GG Art. 3 Abs. 1 Rn. 1ff., in: Maunz, Theodor/Dürig, Günter: 85. EL November 2018.

Erkenntnisse im Hinblick auf neue Diagnosemethoden in der Medizin keine Seltenheit. Wenn ein auf dem bisherigen Stand der Wissenschaft und Forschung fußendes technisches System zum Einsatz kommt, kann sich dieses daher (ggf. schnell) als veraltet herausstellen und bedeutete dann gar einen Rückschritt gegenüber den verbesserten Möglichkeiten, die dem Menschen außerhalb dessen zur Verfügung stehen. Ebenso verhält es sich im Hinblick auf das Recht: Normative Wertungen stellen prinzipiell lediglich ein Vorbehaltsurteil dar: Ihre Richtigkeit kann stets allein auf der Basis des jeweiligen Kenntnisstands angenommen werden. Weil das Recht sich in einer steten Entwicklung hin zu einer idealen rechtsstaatlichen Ordnung befindet, ist es also nicht ausgeschlossen, dass diese Wertung im Nachhinein korrigiert werden muss. Zwar macht dies die früher getroffenen rechtlichen Entscheidungen nicht illegitim – vielmehr sind diese gewissermaßen als Momentaufnahme zulässig, da anderenfalls Recht gar nicht möglich wäre. Sofern nämlich das Nichtwissen im Hinblick auf spätere Entwicklungen einer rechtlichen Wertung entgegenstünde, müsste hiervon in Gänze Abstand genommen werden, was allerdings wiederum dem Funktionieren eines gesellschaftlichen Miteinanders entgegenstünde. Die Hinnahme gewisser (vorläufiger) Fehleinschätzungen gehört damit zu einem Rechtssystem ebenso dazu wie die Korrektur früherer normativer Wertungen.⁸

Ein auf Algorithmen basierendes System, das lediglich das bisherige Recht in Gestalt von Gesetzen und Richtersprüchen zu seiner Grundlage wählt, weist indessen gerade nicht die notwendige Offenheit gegenüber Korrekturen und vom Früheren abweichenden Wertungen auf. Insoweit bietet die Idee eines *iudex ex machina*

8 Ein Beispiel liefert der Schutz menschlichen Lebens zu dessen Beginn und Ende. Die zeitliche Vorverlagerung des Todeszeitpunkts durch das Hirntodkriterium ändert nichts daran, dass eine Person, die im Zeitraum der Geltung des Herztodkriteriums den Herztod eines Menschen verursacht hat, dessen Hirntätigkeit bereits erloschen war, *richtigerweise* nicht bloß wegen (untauglichen) Versuchs, sondern wegen vollendeten Totschlags bzw. Mordes verurteilt wurde. Zudem ist denkbar, dass auch der strafrechtliche Lebensschutz künftig vor dem Einsetzen der Eröffnungswehen ansetzen wird. Dies machte aber die gegenwärtige Rechtslage nicht unrichtig – vielmehr stellte diese einen zulässigen Zwischenstand innerhalb der weiteren Rechtsentwicklung dar. Die Möglichkeit, dass es sich bei der jeweiligen rechtlichen Wertung um eine Fehleinschätzung handelt, kann allerdings in bestimmten Bereichen Zurückhaltung anmahnen. Zu denken ist etwa an Fortschritte im Bereich der Biomedizin, deren Auswirkungen auf die Betroffenen und künftige Generationen teilweise zum gegenwärtigen Zeitpunkt nicht abschätzbar sind. Eine die Tatsachengrundlage betreffende Ungewissheit über das Gefährdungspotential solcher Technologien kann daher bei Abwägung mit den widerstreitenden Interessen (insbesondere: Wissenschaftsfreiheit und allgemeine Handlungsfreiheit) prinzipiell eher zur (vorläufigen) restriktiven Regulierung in diesem Bereich führen. Auch im Strafverfahren tritt regelmäßig die Schwierigkeit auf, dass nicht sicher ist, ob der Richterspruch auf der Basis einer vom eigentlichen Geschehen abweichenden Tatsachengrundlage getroffen wird. Dennoch macht dies das Urteil nicht per se unrichtig oder gar unmöglich. Weil es eine absolute Wahrheit nicht geben kann, ist in normativer Hinsicht der rechtsgenügende Beweis für eine Verurteilung ausreichend.

bzw. einer rechtliche Entscheidungen auf der Basis des geltenden Rechts unterstützenden technischen Anwendung eine weitere Facette des Positivismuseinwands, die im Kontext eines selbstlernenden Systems in besonderer Deutlichkeit und Tragweite auftritt.⁹ Die Rede ist davon, dass ein solches Programm keinen Raum für das Hinterfragen derjenigen Informationen bietet, die ihm als Grundlage des Lernens angeboten werden. Sofern also Gesetze oder Gerichtsentscheidungen als „Lernmaterial“ eingespeist werden, erfolgt keine Richtigkeitskontrolle. Selbst systematische Abweichungen gegenüber bisherigen Informationen können lediglich als Befund hingenommen und weiterverarbeitet werden – sie führen aber nicht etwa dazu, dass diese Norm ausgesondert oder nicht angewendet wird. Vielmehr wird auch deren Aussage in das Erlernete integriert und sei es nur als Abweichung vom Bisherigen, die einen neuen statistischen Parameter entstehen lässt. Eine Kontrollmöglichkeit außerhalb der durch die Informationsbasis erlernten Struktur fehlt. Vereinfacht gesprochen vertraut der juristische Algorithmus in einem solchen technischen Verfahren auf die Relevanz und Richtigkeit jedweder Information, sodass diese unweigerlich Teil des „Gesamtsystems“ wird.¹⁰

Aus diesem Umstand können sich schwerwiegende Negativkonsequenzen ergeben, sofern es etwa zur Streichung einer Vorschrift durch den Gesetzgeber kommt. Zu denken ist allein an die frühere Vorschrift des § 175 StGB („Unzucht zwischen Männern“). Ist diese Norm mit ihren Wertungen einmal in das System aufgenommen worden, führt deren Aufhebung und Nichtanwendung für sich genommen unter Umständen nicht dazu, dass nicht doch weiterhin Folgewirkungen der Vorschrift die Arbeit des *iudex ex machina* beeinflussen. So erscheint es naheliegend, dass ein auf Algorithmen basierendes System Wertungen, die es einer bestimmten Vorschrift entnehmen kann, auf die gesamte Rechtsordnung überträgt und damit auch in ganz andere Bereiche des Rechts einfließen lässt, als demjenigen, dem die jeweilige Norm angehört. Ein Beispiel: Weil der frühere § 175 StGB unter bestimmten Voraussetzungen den Beischlaf unter Männern unter Strafe stellte, könnte ein damit „gefüttertes“ System dieser Vorschrift die (allgemeine) Wertung entnehmen, dass eine rechtliche Benachteiligung aufgrund einer bestimmten sexuellen Neigung grundsätzlich zulässig sei. Diese Information könnte dann zunächst in anderen Bereichen des Strafrechts wie etwa dem Strafzumessungsrecht Konsequenzen nach sich ziehen. Denkbar wäre insofern,

9 Zum Positivismus in seinen unterschiedlichen Spielarten siehe Bydlinski, Franz (1982): Methodenlehre und Rechtsbegriff, Wien, S. 38ff.

10 Siehe zum Ganzen bereits Rostalski, Frauke (2019): Legal Tech now and then, in: REthinking:Law, 1/2019, S.4 und 9.

dass die Strafhöhenbemessung bei einer homosexuellen Person strenger ausfiele als bei einem heterosexuellen Täter, sofern die entsprechenden Informationen vorhanden sind. Und auch außerhalb des Strafrechts könnte die dargelegte Wertung die rechtliche Einschätzung des *iudex ex machina* beeinflussen – wie etwa bei der Frage, ob eine arbeitsrechtliche Kündigung aufgrund der Homosexualität des Arbeitnehmers zulässig oder die Höhe eines wegen Zerstörung einer fremden Sache zu leistenden Schadensersatzanspruches davon abhängig ist, welche sexuellen Vorlieben deren Eigentümer aufweist.¹¹

Vereinfacht gesprochen besteht bei einem technischen System die Gefahr, dass trotz Streichung einer als verfassungswidrig erkannten Norm deren Wertungen fortdauernden Einfluss auf die Entscheidungen des Programms haben. Die unzulässige Vorschrift ist geeignet, das gesamte System zu infiltrieren bzw. zu „vergiften“. Grundsätzlich ließe sich dieser Einwand zwar auch gegenüber einem menschlichen Juristen erheben. Wer etwa jahrzehntelang in Kenntnis der Vorschrift des früheren § 175 StGB tätig war, kann sich unter Umständen nur schwer daran gewöhnen, dass Diskriminierungen einer homosexuellen Neigung rechtlich unzulässig sind. Allerdings besteht im Vergleich zu der Maschine beim Menschen die Fähigkeit der Reflexion und damit der kritischen Überprüfung der eigenen rechtlichen Entscheidungen. Für den Menschen sind seine Entscheidungen im Hinblick auf deren Zustandekommen weitestgehend transparent. Er ist jedenfalls dazu in Lage, die Gründe offenzulegen, die ihn zu einer bestimmten Wertung gebracht haben. Sofern darin eine Ableitung aus einer als verfassungswidrig erkannten Norm vorkommt, ist der Mensch befähigt, diese zu korrigieren.

All diese Eigenschaften weist ein technisches System nicht auf. Die „Bereinigung“ des Programms um eine fehlerhafte Wertung erfolgt zwar ebenfalls, allerdings nicht von heute auf morgen, sondern erst nach einer längeren Zeitspanne, die als Lernprozess im Hinblick auf die neue Rechtslage erforderlich wird. Dieser Zustand ist aber – wie bereits gesagt – aus rechtsstaatlichen Gründen inakzeptabel. Er lässt sich auch nicht „von außen“ durch den Menschen korrigieren. Durch maschinelles Lernen mittels neuronaler Netze entwickelt das System Algorithmen, die es zur Lösung juristischer Sachverhalte heranzieht. Diese Algorithmen sind aber in ihrer Entstehung in aller Regel für den Menschen nicht nachvollziehbar. Eben jene Intransparenz führt dann aber dazu, dass sich im Nachhinein gerade nicht sagen ließe, ob eine bestimmte Entscheidung des Programms auf einer unzulässigen rechtlichen Wertung beruht oder nicht.

11 Rostalski, Frauke (2019): Legal Tech now and then, in: REthinking:Law, 1/2019, S.4 und 9.

Insoweit lässt sich gegen die Verwendung entsprechender Technologien im Bereich des Rechts folgende Facette des Positivismuseinwands erheben: Ein in der beschriebenen Weise funktionierendes Programm nimmt nicht allein jedwede Norm, mit der es als Information „gefüttert“ wird, als richtig und „Recht“ an, ohne diese einer wertenden Kontrolle zu unterziehen. Darüber hinaus schreibt es diesen Zustand für eine nicht absehbare Zeit fest und ist damit in seinem Festhalten an dem positiven Recht besonders streng. Eine einmal als Recht klassifizierte Wertung bleibt im Zweifel *für immer* in den Tiefen der Algorithmen des technischen Systems erhalten. Recht ist dann nicht allein das positive Recht zu einem bestimmten Zeitpunkt, sondern *Recht ist jedes (irgendwann) einmal positivierte Recht*.¹² Der Weg zu einem verbesserten Erkenntnisgewinn und damit einer Rechtsveränderung, die sich vorteilhaft für die Gesellschaftsmitglieder auswirken kann, wird dann aber im Keim erstickt. Insofern halten technische Systeme selbst in Anwendungsfeldern, für die ein gewisses Maß an Kontinuität eine Bedeutung aufweist, lediglich in erheblich eingeschränktem Umfang ein Verbesserungspotential im Hinblick auf menschliche Entscheidungen bereit.

IV. GEFAHREN TECHNISCHER ENTSCHEIDUNGSGEHILFEN FÜR DIE FREIHEIT

Mit der Nutzung technischer Entscheidungsgehilfen geht das Versprechen einher, durch ihren Einsatz mehr Zeit für sinnvolle Tätigkeiten zu gewinnen und damit einen Freiheitszuwachs zu erleben. Dies ließe sich im Hinblick auf einen den Wocheneinkauf automatisch übernehmenden Kühlschrank prinzipiell annehmen: Wer auf diese Weise eine Stunde Zeit einspart, kann sie an anderer Stelle ggf. sinnvoller nutzen wie etwa durch das Lesen eines Buchs oder das Verbringen gemeinsamer Zeit mit Familie und Freunden. Technische Hilfsmittel könnten insofern vor allem in den Lebensbereichen Anwendung finden, die von dem Einzelnen bislang als eher lästig empfunden wurden. Dies kann gerade auch berufliche Tätigkeiten betreffen. Anstatt Zeit mit der Suche nach einer früheren Akte zu verschwenden, in der ein ähnliches Rechtsproblem bereits diskutiert wurde, mit dem sich der Rechtsanwalt in seinem neuen Fall befasst, wäre es für ihn (und seinen Mandanten) ein erheblicher Vorteil, diese Aufgabe einem technischen Programm zu überlassen und sich selbst unmittelbar auf die rechtliche Lösung des neuen Mandats zu konzentrieren.

12 Rostalski, Frauke (2019): Legal Tech now and then, in: REthinking:Law, 1/2019, S.4 und 10.

Wenngleich sich die Reihe an Beispielen des nützlichen Einsatzes technischer Entscheidungsgehilfen ohne Weiteres fortsetzen lässt, ist der damit einhergehende Nutzen unter dem Aspekt des Zugewinns an Freiheit nicht uneingeschränkt als positiv zu bewerten. Jedenfalls im Hinblick auf Entscheidungen, die für den Einzelnen bzw. Dritte eine gesteigerte Bedeutung aufweisen, sprechen gewichtige Gründe *gegen* eine durch Algorithmen maßgeblich beeinflusste Entscheidungsfindung. Insoweit ist zu berücksichtigen, dass der Akt des Entscheidens für sich genommen als Ausdruck von Freiheit klassifiziert werden muss. Wer sich frei entscheidet, bestimmt sich selbst.¹³ Zu entscheiden heißt also, der eigenen Person bzw. dem eigenen Selbst Ausdruck zu verleihen.¹⁴ Es erklärt sich damit zugleich die hohe Relevanz, die das (freie) Entscheiden für das Individuum aufweist. In ihm kommt unmittelbar die Persönlichkeit des Einzelnen zum Ausdruck. Durch seine Entscheidungen vermittelt der Betreffende seine individuellen Wertvorstellungen sowie seinen Lebensentwurf. Das Produkt der Entscheidung – das jeweilige Verhalten – strahlt insofern auf den Einzelnen zurück, als dass er sich auf diese Weise selbst bestätigt bzw. definiert. Wollte der Mensch daher auf das Entscheiden einschließlich des Vorgangs der Entscheidungsfindung in relevanten Lebensbereichen verzichten, bedeutete dies auf lange Sicht einen Verlust der Ausbildung und Entwicklung der eigenen Persönlichkeit. Wer nicht länger selbst über die eigenen Geschicke entscheidet, droht, den Bezug zwischen seiner Identität und seiner Lebensführung zu verlieren. Es kann auf diese Weise der gefährliche Eindruck entstehen, der Einzelne sei nicht länger „Herr seiner Tage“¹⁵ bzw. „Autor seiner selbst“¹⁶. Die Freiheit, sich selbst auszudrücken und als Autor der eigenen Entscheidungen zu begreifen, vermittelt ein Selbstbild. Sofern selbst relevante Lebensentscheidungen wie etwa die Wahl des Partners oder der konsumierten Meinungsinhalte an eine Maschine abgegeben werden, besteht die greifbare Gefahr, den Bezug hierzu und zuletzt zu sich selbst einzubüßen.

13 Zum Schutz des Selbstbestimmungsrechts als Ausfluss der Menschenwürde siehe Herdegen, Matthias (2018): GG Art. 1 Abs. 1 Rn. 84 ff, in: Maunz, Theodor/Dürig, Günter: 85. EL November 2018.

14 Auf einer Linie damit liegt Di Fabio, Udo (2018): GG Art. 2 Abs. 1 Rn. 127, in: Maunz, Theodor/Dürig, Günter: 85. EL November 2018, der zum allgemeinen Persönlichkeitsrecht ausführt: „Geschütztes Rechtsgut ist der Geltungsanspruch des Menschen in der sozialen Welt, die ihn prägt und die er wiederum auch dadurch prägt, dass er durch sein Handeln von ihr anerkannt werden will.“ Wenn das Handeln nicht länger Ausdruck der eigenen Entscheidung ist bzw. zumindest von dem Einzelnen nicht als ein solcher wahrgenommen wird, geht ihm ein wesentliches Element zur Prägung seiner Persönlichkeit verloren.

15 Camus, Albert (2010): Der Mythos des Sisyphos, Reinbek bei Hamburg, S. 155 und 159.

16 Nach einer Formulierung von Habermas, Jürgen (2002): Die Zukunft der menschlichen Natur. Auf dem Weg zu einer liberalen Eugenik?, Frankfurt am Main, S. 54 f.

Dem steht nicht entgegen, dass der Betreffende zumindest die Wahl des Einsatzes eines technischen Hilfsmittels selbst getroffen hat und sich daher – vermittelt durch das jeweilige Programm – dennoch als Autoren seiner selbst sehen muss. So ist es ein relevanter Unterschied, ob der Einzelne sich lediglich für die Methode entschieden hat, durch die ihn betreffende, weitere Entscheidungen generiert werden sollen, oder ob er die Entscheidung im Einzelfall unmittelbar selbst fällt. Zwar ist es richtig, dass der Griff zur Partner-App oder das Anklicken einer Homepage, die nach bisherigen Interessen vorgefilterte Nachrichteninhalte anbietet, durchaus in dem Wissen über die Funktionsweise solcher Programme und damit die Eingeschränktheit der offerierten Möglichkeiten erfolgen kann. Dennoch handelt es sich hierbei lediglich um eine Vorauswahl, die noch keine inhaltliche Aussage über die späteren Entscheidungen trifft. Zum Beispiel: Die Entscheidung, einen Internetdienst zu nutzen, der Nachrichteninhalte nach bisherigen Präferenzen vorfiltert, enthält noch nicht die Aussage, dass sich der Betreffende für sonstige Informationen, die ihm auf diese Weise verloren gehen können, per se nicht interessiert.

Und noch ein weiteres Risiko geht mit der zunehmenden Unterstützung der menschlichen Entscheidungen durch technische Systeme einher: Wer die Erfahrung macht, dass ein technisches System in bestimmten Lebensbereichen sehr gute, ggf. sogar im Vergleich zu den eigenen *bessere* Handlungsvorschläge unterbreitet, kann das Vertrauen in die eigenen Fähigkeiten verlieren. Dies erweist sich nicht zuletzt dann als gefährlich, wenn davon Kompetenzen betroffen sind, die für Dritte bzw. die Gesellschaft in ihrer Gesamtheit von Bedeutung sind. Zu denken ist allein an den Arzt, der die Erfahrung macht, dass eine bestimmte Diagnosesoftware in der Röntgenbildanalyse häufiger richtig liegt als er selbst. Dieser Erkenntnis kann die abnehmende Bereitschaft folgen, sich auf die eigenen Fähigkeiten zu verlassen und Entscheidungen zu treffen, die mitunter auch von dem Vorschlag der Diagnosesoftware abweichen (vgl. den Aufsatz von Hermann und Stock in dieser Ausgabe). Wie gezeigt, entsteht damit zum einen die riskante Situation, dass der menschliche Wissensstand gewissermaßen auf das Programm der jeweiligen Software festgeschrieben würde und Fortschritte dann allein auf dieser Basis systemintern, nicht aber auf derjenigen gänzlich abweichender Prämissen erzielt werden könnten. Darüber hinaus besteht das Risiko, dass der Betreffende selbst aufgrund des zunehmenden Gebrauchs der (vermeintlich überlegenen) Technik eigene Kompetenzen verliert, da es ihm an der dafür notwendigen Übung fehlt. Sofern sich dann aber ein Fehler ins technische System etwa im Prozess des Selbstlernens einschleicht, wäre der Mensch weder in der Lage, diesen

zu korrigieren, noch die bis dahin von der Maschine übernommene Tätigkeit ohne Weiteres in vergleichbarer Qualität wieder selbst zu übernehmen.

Dabei weisen die beschriebenen Gefahren einer zunehmenden Verwendung technischer Entscheidungshilfen in relevanten Lebensbereichen zugleich eine gesamtgesellschaftliche Dimension auf. Sofern Kompetenzen aufgrund des Technikeinsatzes schwinden bzw. durch die Delegation bedeutsamer Entscheidungen Verluste im Bereich der eigenen Persönlichkeitsentwicklung erlitten werden, hat dies mitunter nicht lediglich schwerwiegende Negativfolgen für das persönliche Lebensglück. Wer die eigene Person nicht als Autoren seiner selbst begreift, sondern vielmehr als Produkt externer Entscheidungsprozesse, kann geneigt sein, eine große Gleichgültigkeit gegenüber den eigenen sowie den gesellschaftlichen Geschicken zu entwickeln.¹⁷ In der Folge wird der Betreffende zu einem leichten Opfer mannigfaltig denkbarer Manipulationsversuche, die ihn beispielsweise zu einem optimalen Konsumenten innerhalb des kapitalistischen Systems sowie zu einem unkritischen, politisch desinteressierten Bürger werden lassen. Zuletzt erweist sich eine solche Entwicklung daher als besondere Gefahr für das freiheitlich demokratische Gemeinwesen, das auf den mündigen Bürger als seinen Ausgangspunkt und gewissermaßen seine „kleinste Einheit“ im Kern angewiesen ist.¹⁸ Ein Mehr an Freiheit, das technische Entscheidungsgehilfen uns uneingeschränkt versprechen, wird daher mitnichten in jedwedem Lebensbereich durch deren Einsatz erzielt. Im Gegenteil kann die Verwendung entsprechender Technologien im Hinblick auf für den Einzelnen oder Dritte relevante Entscheidungen gar einen Verlust an Freiheit bedeuten, der in letzter Konsequenz negative Auswirkungen auf das Funktionieren eines freiheitlichen Miteinanders haben kann.

17 Vgl. aus dem biotechnischen Kontext Habermas, Jürgen (2002): Die Zukunft der menschlichen Natur. Auf dem Weg zu einer liberalen Eugenik?, Frankfurt am Main, S. 111f. und 115f. Vgl. auch Siep, Ludwig (2004): Konkrete Ethik. Grundlagen der Natur- und Kulturethik, Berlin, S. 320.

18 Siehe Herdegen, Matthias (2018): GG Art. 1 Abs. 1 Rn. 28, in: Maunz, Theodor/Dürig, Günter: 85. EL November 2018; Timm, Frauke (2012): Gesinnung und Straftat, Berlin; Sinn, Arndt (2006): Moderne Verbrechenverfolgung - auf dem Weg zu einem Feindstrafrecht?, in: Zeitschrift für internationale Strafrechtsdogmatik (ZIS), Ausgabe 3/2006, S.107 und 113.

IV. RESÜMEE

Technische Entscheidungsgehilfen haben sich allenfalls in einem sehr kleinen Bereich als nützliche Werkzeuge des Menschen erwiesen. Darüber hinaus ist im Umgang mit ihnen prinzipiell Vorsicht geboten. Entsprechende Warnungen dürften aufgrund der Verlockungen des digitalen Zeitalters und der ungebrochenen Begeisterung des technikaffinen Menschen unter Umständen leicht überhört werden. Ist den Gefahren für die individuelle sowie gesellschaftliche Freiheit daher kein Einhalt mehr zu gebieten? Anlass zu solchem Pessimismus besteht jedenfalls für den Moment nicht. Das Stichwort lautet in diesem Zusammenhang „Entzauberung“. Entzaubert werden müssen all jene Technologien, über die unterschiedlich motivierte Technik-Apologeten Heilsversprechen in die Welt hinausgetragen. Erforderlich ist insoweit ein nüchterner Umgang mit den Vorzügen und Nachteilen, die der Einsatz unterschiedlicher technischer Fortschritte für unser Leben haben kann. Einen solchen erreichen wir durch die Information und den Diskurs. Die Welt ist das, was wir aus ihr machen. Dies betrifft auch das Bild, welches wir von ihr in Zeiten der digitalen Revolution zeichnen möchten.

KOMPETENZVERLUST IN ZEITEN VON KI – WIE BEWAHREN MENSCHEN WICHTIGE FÄHIGKEITEN?¹

I. EINLEITUNG

Der Einsatz von Künstlicher Intelligenz (KI) bei der Automatisierung von Prozessen, sei es die Unterstützung durch Roboter oder die Assistenz durch Vorhersagemodelle, ist in vielen Berufsfeldern und in unserem Alltag zunehmend Realität. Der steigende Einsatz von KI gründet hauptsächlich auf Erfolgen im Bereich des Maschinellen Lernens, einem Verfahren, bei dem Computer Muster in großen Datenmengen zum Beispiel mit Hilfe von sogenannten neuronalen Netzen erkennen. Wir werden es also jetzt und in Zukunft an vielen Stellen mit Maschinen zu tun zu haben, an denen vorher allein Menschen zuständig waren. In diesem Zusammenhang werden Risiken und Chancen für den Menschen beim Einsatz von KI-Systemen breit diskutiert: Risiken wie die Diskriminierung von bestimmten Gruppen durch Verzerrungen in den Trainingsdaten, die sich in die KI-Systeme einschreiben, und Chancen wie z.B. den Menschen bei „difficult, dirty, dull or dangerous work“ zu helfen oder ganz zu ersetzen.² In diesem Beitrag möchten wir im Speziellen die Ebene der Mensch-Maschine-Interaktion beleuchten, die menschliche Kompetenzen betrifft. Insbesondere geht es darum, wie wir einerseits für den Menschen wichtige Kompetenzen zukünftig bewahren können und wie wir andererseits mit einem Verlust von Kompetenzen umgehen, wenn Menschen in immer höherem Ausmaß durch automatisierte Systeme unterstützt werden, aber weiterhin in der Verantwortung stehen. Daneben müssen wir uns damit auseinandersetzen, wie wir komplexe selbstlernende Systeme behandeln, deren Lernvorgänge undurchsichtig sind und daher als „Black Box“ bezeichnet werden. Diese Fragen sind vor allem in sensiblen Bereichen relevant, in denen Anwender oder Dritte schwerwiegenden Schaden erleiden können.

1 In diesem Beitrag schließt aufgrund der flüssigeren Lesbarkeit die männliche Form von Personenbezeichnung die weibliche Form mit ein.

2 European Group on Ethics in Science and New Technologies (2018): Statement on artificial intelligence, robotics and 'autonomous' systems, Brüssel, S. 6. Online unter: http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf [29.04.2020].

Dafür gehen wir zunächst auf das Automations-Paradox in der Mensch-Maschine-Interaktion ein. Demnach werden technisch automatisierte Prozesse in den meisten Fällen als sicherer und dem Menschen überlegen angesehen, außer wenn etwas schief geht und Menschen als die Überlegenen einspringen müssen.³ Für dieses Paradox bietet der Luftverkehr passende Anschauungsbeispiele, weil man hier auf eine über 100-jährige Geschichte der Zunahme von technischen Standards und Automatisierung zurückblicken kann, die die Sicherheit des Flugverkehrs enorm erhöht, die Situation für den Menschen in der Verantwortung jedoch erschwert haben. Im Weiteren zeigen wir diese Herausforderungen am Beispiel des automatisierten Fahrens sowie in der Medizin bei OP-Robotern bzw. unterstützenden Diagnosesystemen, die mit KI verfahren, und verweisen auf den nötigen Handlungsbedarf beim Ungleichgewicht zwischen der Verantwortung der Anwender und ihrer jeweiligen Kompetenz. Diese Thematik erhält noch eine zusätzliche Brisanz, weil besonders komplexe KI-Systeme uns nicht von selbst mitteilen, was und wie sie lernen.

II. DAS AUTOMATIONS-PARADOX AM BEISPIEL DES FLUGVERKEHRS

Wenn es um die Mensch-Maschine-Interaktion und steigende Automatisierung geht, stechen – unabhängig vom Einsatz von KI – zwei Sichtweisen heraus: Zum einen die Vorstellung, dass die Technik dem Menschen prinzipiell überlegen ist und durch technische Automatisierung menschliche Unzulänglichkeiten und Fehler verhindert werden können. Auch an Stellen, an denen der Mensch noch vorkommt, wird er lediglich als ein Störfaktor in einem sonst funktionierenden System angesehen, der so klein wie möglich gehalten werden muss. Als Beispiel kann der Zusammenstoß zweier Verkehrsflugzeuge nahe der Stadt Überlingen im Jahr 2002 dienen, bei dem alle 71 Passagiere ums Leben kamen.⁴ Beide Flugzeuge waren mit einem bodenunabhängigen Zusammenstoßwarnsystem (TCAS) ausgestattet, das die nahende Kollision erkannte und entsprechend die Boeing B757-200 anwies zu sinken und die Tupolew TU154M zu steigen, um den Zusammenprall zu verhindern. Der durch andere Aufgaben überforderte Fluglotse am Boden forderte die Tupolew allerdings fälschlicherweise auf, zu sinken. Durch die Anweisung, der die Crew nachkam, wurde die Kollision jedoch erst ausgelöst, weil die Boeing nach

3 Bainbridge, Lianne (1983): Ironies of Automation, in: *Automatica*, Vol. 19/No. 6, S. 775–779.

4 Bundesstelle für Flugunfalluntersuchung (2004): Untersuchungsbericht AX001-1-2/02, Braunschweig. Online unter: https://www.bfu-web.de/DE/Publikationen/Untersuchungsberichte/2002/Bericht_02_AX001-1-2.pdf?__blob=publicationFile [29.04.2020].

Anweisung des TCAS, das nach einem festen Regelsatz codiert war und sich nicht dynamisch anpasste, ebenfalls weiter an Höhe verlor. Das TCAS-Nachfolgemodell ACAS X (Airborne Collision Avoidance System X) basiert nun nicht mehr auf festgelegten Codierregeln, sondern auf einem probabilistischen Modell, das in Echtzeit Vorhersagen über die zukünftigen Flugzeugposition liefert⁵ – eine Funktionsform, die wir nach aktuellem Standard als KI bezeichnen können. Störungen in Form von fatalen fehlerhaften Angaben eines Fluglotsen wie im oben dargestellten Fall, sollte das System also nun selbst ausgleichen können.

Dieser Vorstellung vom Menschen als Störfaktor durchaus widersprechend ist zum anderen die Sichtweise, dass bei einem Fehler in einem automatisierten System wiederum der Mensch eingreifen können muss, um den Fehler zu korrigieren. Ein Beispiel ist der Lion Air Flug 043 mit der mittlerweile mit einem Flugverbot versehenen Boeing 737 MAX 8, der sich am 28. Oktober 2018 auf dem Weg von Denpasar nach Jakarta befand. Kurz nach dem Start des Flugzeuges drückte das sogenannte Maneuvering Characteristics Augmentation System (MCAS) aufgrund eines falschen Dateninputs die Nase des Flugzeuges nach unten, um eine vermeintliche Fehllage zu korrigieren. Die Crew schaffte es trotz systeminterner Widerstände erfolgreich, den Autopiloten schlussendlich zu deaktivieren und sicher am Zielflughafen zu landen. Die Piloten hatten zu diesem Zeitpunkt allerdings nicht nur kein Protokoll zur Hand, wie sie sich bei einer Fehlfunktion von MCAS verhalten sollten, ihnen war von Boeing gar nicht mitgeteilt worden, dass das System existierte. MCAS wurde eingebaut, um per Software aerodynamische Fehlstellungen des Flugzeuges aufgrund von schwerwiegenden Designfehlern in der Hardware auszugleichen, wovon die Piloten ebenfalls nichts wussten. Als am nächsten Morgen die Maschine mit neuer Crew als Lion Air Flug 610 die indonesische Hauptstadt verließ, stellten sich die Probleme erneut ein, doch aufgrund fehlender Prozeduren war es den Piloten diesmal nicht möglich, den Autopiloten zu deaktivieren, was zum Absturz des Flugzeuges führte, bei dem alle 189 Passagiere starben.⁶ Die gleichen Umstände führten am 10. März 2019 zum Absturz des Ethiopian-Airlines-Fluges 302 von Addis Abeba nach Nairobi, bei dem alle 157 Menschen an Bord ums Leben kamen. Die genannten Beispiele sind desaströs, doch demonstrieren sie exemplarisch die letztendliche Unvereinbarkeit der beiden Sichtweisen auf: Designer, Ingenieure, Regulierungsbehörden und

5 Eurocontrol (2013): NETALERT - the Safety Nets Newsletter June 2013/N°17, S.2/3. Online unter: <https://www.skybrary.aero/bookshelf/books/2390.pdf> [29.04.2020]

6 Campbell, Darryl (2019): Redline – The many human errors that brought down the Boeing 737 Max, in: The Verge, 02.05.2019. Online unter: <https://www.theverge.com/2019/5/2/18518176/boeing-737-max-crash-problems-human-error-mcas-faa> [29.04.2020].

sogar die breite Öffentlichkeit hätten eine „schizophrenic dynamic“ geschaffen, in der der Mensch die Kontrolle abgeben und gleichzeitig behalten soll.⁷ Die Beispiele zeigen den menschlichen Fehler auf der einen Seite, der nun durch noch bessere Technik – das Kollisionsvermeidungssystem ACAS X – weiter minimiert werden soll, und die Notwendigkeit zum menschlichen Eingreifen auf der anderen Seite, wenn eine fehlerhafte Software im Autopiloten droht, das Flugzeug zum Absturz zu bringen.

Doch ist der Mensch überhaupt gut darin einzugreifen, wenn ein technisches System versagt? Im Jahr 2009 stürzte der Airbus A330 des Air France Fluges 447 von Brasilien nach Frankreich über dem Atlantik ab, alle 228 Menschen an Bord verloren dabei ihr Leben. Zu dem Unglück kam es, nachdem die Geschwindigkeitsmessgeräte des Flugzeuges bei schweren Gewittern ausfielen und die Piloten sich gezwungen sahen, den Autopiloten abzuschalten. Danach begann eine Serie von schwerwiegenden Pilotenfehlern aufgrund einer falschen Einschätzung der Situation und Kooperationsschwierigkeiten innerhalb der Crew, die letztendlich zum Absturz des Flugzeuges führten.⁸ Dieser Fall ist beispielhaft für die Statistik, nach der in den 2010er Jahren bisher 57 Prozent aller fatalen Flugzeugabstürze auf Pilotenfehler zurückzuführen sind.⁹ So hat sich die Sicherheit in der zivilen Luftfahrt kontinuierlich durch die zunehmende Automatisierung erhöht, so dass „ein Flugzeug wesentlich sicherer als die Fahrt zum Flughafen [ist]“.¹⁰ Doch trotz Automatisierung können menschliche Fehler nicht ganz eliminiert werden – sie werden sogar fataler, da die kognitiven Fähigkeiten der Piloten wie Navigation sowie Fehlererkennung und -diagnose anfällig dafür sind, bei zunehmender Automatisierung abzunehmen.¹¹

7 Elish, Madeleine/Hwang, Tim (2018): Praise the Machine! Punish the Human! The contradictory history of accountability in automated aviation, Comparative Studies in Intelligent Systems – Working Paper #1, S. 12. Online unter: https://www.datasociety.net/pubs/ia/Elish-Hwang_AccountabilityAutomatedAviation.pdf [29.04.2020].

8 Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2012): Final Report on the Accident on 1st June 009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris, Le Bourget. Online unter: <https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf> [29.04.2020].

9 PlaneCrashInfo.com (2019): Causes of Fatal Accidents by Decade. Online unter: <http://www.planecrashinfo.com/cause.htm> [29.04.2020].

10 Köver, Chris (2018): Interview mit KI-Forscher Toby Walsh: „Wir müssen jetzt die richtigen Entscheidungen treffen“. Online unter: <https://netzpolitik.org/2018/interview-mit-ki-forscher-toby-walsh-wir-muessen-jetzt-die-richtigen-entscheidungen-treffen/> [29.04.2020].

11 Casner, Stephen/Geven, Richter/Recker, Matthias (2014): The Retention of Manual Flying Skills in the Automated Cockpit, in: Human Factors, Vol 56, Issue 8.

Dazu kommt, dass Piloten häufig und plötzlich in Notsituationen eingreifen oder übernehmen müssen, was genau die Art von Aufgabe ist, für die der Mensch schlecht geeignet ist. Die Federal Aviation Administration (FAA) hat 2013 nach einer Analyse von Flugbetriebsdaten einen „Safety Alert“ zum Anstieg der Fehler im manuellen Flugbetrieb herausgegeben. Demnach könne der kontinuierliche Einsatz von Autopiloten zu einer Beeinträchtigung der Fähigkeit der Piloten „to quickly recover the aircraft from an undesired state“ führen.¹² Die FAA ermutigt dabei die Fluggesellschaften einem integrierten Ansatz zu folgen und einen stärkeren Schwerpunkt auf den manuellen Flugbetrieb sowohl im Linienbetrieb als auch in der Ausbildung der Piloten zu legen. Ein ähnlicher Safety Alert thematisierte 2017 noch einmal speziell die Beeinträchtigung der Fähigkeiten von Luftfrachtpiloten durch den Einsatz von Autopiloten.¹³

Angesicht der Tatsache, dass der Autopilot immer mehr menschliche Aufgaben übernehmen soll, um standardmäßig menschliche Fehler auszuräumen und die Sicherheit zu erhöhen, ist es eine interessante Beobachtung, dass nach wie vor die Piloten die volle Verantwortung für den Betrieb des Flugzeuges innehaben. Da menschliches Handeln zunehmend von Automatisierung ersetzt wird, was zudem zu einem Verlust der Fähigkeiten der Piloten führt, kann man von einem „kontra-intuitiven Fokus“ auf menschliche Verantwortung sprechen.¹⁴

III. ANWENDER ALS „HAFTUNGSKNECHTE“?

Anhand von Beispielen aus dem zivilen Luftverkehr lässt sich diese „Fehlausrichtung“ zwischen der sinkenden Kontrolle und dem Kompetenzverlust eines Menschen über Betriebsabläufe einerseits und der Verantwortung für diese Abläufe andererseits gut zeigen. Dieses Spannungsfeld scheint sich mit der zunehmenden Automatisierung von Abläufen, die durch die Digitalisierung und den Einsatz von KI und Maschinellem Lernen ermöglicht werden, zu erweitern und dadurch zu verschärfen. Im Folgenden möchten wir die Problematik am Beispiel des

12 Federal Aviation Administration (2013): Safety Alert for Operators 13002, Washington, D.C. Online unter: https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/safo/all_safos/media/2013/safo13002.pdf [29.04.2020].

13 Federal Aviation Administration (2017): Safety Alert for Operators 17007, Washington, D.C. Online unter: https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/safo/all_safos/media/2017/safo17007.pdf [29.04.2020].

14 Elish, Madeleine/Hwang, Tim (2018): Praise the Machine! Punish the Human! The Contradictory History of Accountability in Automated Aviation, Comparative Studies in Intelligent Systems – Working Paper #1, S. 3. Online unter: https://www.datasociety.net/pubs/ia/Elish-Hwang_AccountabilityAutomatedAviation.pdf [29.04.2020].

automatisierten Fahrens und des Einsatzes von automatisierten Systemen in der Medizin genauer betrachten.

1. Automatisiertes Fahren

Ein naheliegender Vergleich zum Flugzeug-Autopiloten ist das Autofahren, wo zwar immer mehr Funktionen automatisiert werden, es aber nach Experteneinschätzungen noch Jahrzehnte dauern kann, bis vollständig ohne Fahrer auskommende Fahrzeuge Standard auf den Straßen sein werden.¹⁵ Neben mehr Komfort für die Menschen ist das große gesellschaftspolitische Ziel hinter der Idee des autonomen Fahrens die „Vision Zero“, also ein Straßenverkehr ohne Tote und Schwerverletzte, da bei über 90 Prozent der Unfälle menschliches Versagen die Unfallursache ist.¹⁶ Bis zum autonomen Fahren werden allgemein fünf Stufen des automatisierten Fahrens unterschieden¹⁷, wobei in der höchsten Stufe 5 das Fahrzeug als vollautonom gilt und kein menschliches Eingreifen mehr erforderlich ist, und in der niedrigsten Stufe 1 lediglich einzelne Assistenzsysteme wie beispielsweise ein Tempomat den Fahrer bei bestimmten Fahraufgaben unterstützen. Beim teilautomatisierten Fahren auf Stufe 2 werden schon Funktionen wie Einparken, Spurhalten oder Abbremsen vom Fahrzeug übernommen, der Fahrer muss den Verkehr aber ständig im Blick behalten. Hochautomatisierte Autos auf Stufe 3 können bestimmte Fahraufgaben wie Überholen, Bremsen oder Lenken ohne menschlichen Eingriff für einen begrenzten Zeitraum selbständig ausführen; der Fahrer darf sich vorübergehend von Fahraufgabe und Verkehr abwenden, wird aber bei Bedarf vom System aufgefordert, die Führung zu übernehmen. Beim vollautomatisierten Fahren auf Stufe 4 darf der Fahrer wie beim Level 5 die Fahrzeugführung komplett abgeben und zum Passagier werden, muss jedoch die Führung des Fahrzeugs übernehmen, wenn die Fahraufgaben vom System nicht mehr bewältigt werden können.

15 Ford, Martin (2019): Architects of Intelligence – Stuart Russell, Birmingham, S. 46–47; Ford, Martin (2019): Architects of Intelligence - Rodney Brooks, Birmingham, S. 430.

16 Schäfer, Patrick (2018): Der lange Kampf um Vision Zero, in: Springer Professional. Online unter: <https://www.springerprofessional.de/fahrzeugsicherheit/automatisiertes-fahren/der-lange-kampf-um-vision-zero/15771248> [29.04.2020].

17 Deutscher Bundestag (2018): Autonomes und automatisiertes Fahren auf der Straße – rechtlicher Rahmen, Ausarbeitung WD 7-3000-111/18; ADAC (2018): Autonomes Fahren: Die 5 Stufen zum selbstfahrenden Auto. Online unter: <https://www.adac.de/rund-ums-fahrzeug/autonomes-fahren/grundlagen/autonomes-fahren-5-stufen> [29.04.2020].

Auf den Straßen fahren bisher nur Fahrzeuge der Autonomiestufe 2 und 3; an Level-4-Fahrzeugen arbeiten aktuell u. a. die Unternehmen Google Waymo, GM, Uber oder Daimler.¹⁸ Um den aktuellen technischen Entwicklungen Rechnung zu tragen, wurde das deutsche Straßenverkehrsgesetz (StVG) mit Blick auf hoch- und vollautomatisiertes Fahren (Level 3 und 4) angepasst. Demnach darf sich der Fahrer während des Betriebs der hoch- oder vollautomatisierten Fahrfunktion „vom Verkehrsgeschehen und der Fahrzeugsteuerung abwenden“, muss aber „wahrnehmungsbereit“ bleiben, so dass die Fahrzeugsteuerung unverzüglich übernommen werden kann, wenn das System dazu auffordert, oder der Fahrer erkennt, „dass die Voraussetzungen für eine bestimmungsgemäße Verwendung der hoch- oder vollautomatisierten Fahrfunktionen nicht mehr vorliegen.“¹⁹ Der Fahrzeugführer bleibt bei Verwendung einer hoch- oder vollautomatisierten Fahrfunktion zur Fahrzeugsteuerung für fahrlässig oder vorsätzlich herbeigeführte Schäden verantwortlich, „auch wenn er im Rahmen der bestimmungsgemäßen Verwendung dieser Funktion das Fahrzeug nicht eigenhändig steuert.“²⁰

Beim automatisierten Fahren sind also weiterhin die Fahrer für die Steuerung des Fahrzeugs verantwortlich. Das scheint die Idee der „Vision Zero“ zu konterkarieren, weil beim automatisierten Fahren – entgegen der Idealvorstellung des autonomen Fahrens – nicht bewiesen ist, dass die Sicherheit im Straßenverkehr tatsächlich erhöht wird.²¹ Denn bei steigender Automatisierung bis zum Level 5 des vollautonomen Fahrens gilt das gleiche Paradox wie beim Einsatz von Autopiloten beim Fliegen, nämlich, dass die Rolle für die Fahrer immer schwieriger wird.²² Wenn man sich über weite Strecken zurücklehnen kann, woher soll der Mensch die Fahrpraxis nehmen, um in schwierigen Situationen, die Algorithmen nicht meistern können, plötzlich eingzugreifen? Wie sollen menschliche Fahrer Gefährdungssituationen einschätzen, wenn Algorithmen das Fahrzeug möglicherweise entgegen den eigenen Erwartungen und denen der Fahrer der anderen Autos steuern, beispielsweise wenn es beim städtischen „Stop-and-Go“ langsamer oder vorsichtiger

18 Heinzmann, Andrew (2019): Why you can't buy a self-driving car in 2019, in: How-To Geek. Online unter: <https://www.howtogeek.com/405158/when-can-i-buy-a-self-driving-car> [29.04.2020].

19 Bundesanzeiger (2017): Achstes Gesetz zur Änderung des Straßenverkehrsgesetzes vom 16. Juni 2017 (BGBl. I S. 1648).

20 Ibid.

21 International Transport Forum der OECD (2018): Safer roads with automated vehicles? Paris, S. 5. Online unter: <https://www.itf-oecd.org/sites/default/files/docs/safer-roads-automated-vehicles.pdf> [29.04.2020].

22 Noy, Ian/Shinar, David/Horrey, William (2018): Automated driving: safety blind spots, in: Safety Science Volume 102, S. 72. Online unter: <https://www.sciencedirect.com/science/article/pii/S0925753517304198> [29.04.2020].

fährt, als es gemeinhin unter menschlichen Fahrern üblich ist?²³ Die Eigenheiten des automatisierten Fahrens sollten generell in der Führerscheinprüfung und gezielt in Trainings zum jeweiligen Fahrzeugmodell vermittelt werden, um die Kompetenzen der Fahrer und die Sicherheit im Straßenverkehr zu erhöhen. Sich allein auf die automatisierten Funktionen der Fahrzeuge zu verlassen, reicht offensichtlich nicht aus.

2. Medizinische Diagnose und Behandlung

Ein anderer Bereich, in dem ähnliche Fragestellungen aufkommen, ist der zunehmende Einsatz von Robotik und KI in der Medizin. Wir konzentrieren uns in diesem Abschnitt auf OP-Roboter und unterstützenden Diagnosesysteme. Beim Einsatz von OP-Robotern schlagen Experten ein System ähnlich der Stufen zum autonomen Fahren vor, von Stufe 0 „No Autonomy“, auf der ein Chirurg alle Aufgaben selbst übernimmt, bis Stufe 5 „Full Automation“, auf der der Roboter eine Operation selbstständig ohne menschliche Intervention durchführt.²⁴ Auch wenn „Robochirurgie“ noch kein ähnlich fortgeschrittenes Stadium auf dem Weg zur Autonomie wie das Autofahren erreicht hat, existieren bereits einzelne robotische Systeme, die auf Autonomiestufe 3 anzusiedeln sind, wie beispielsweise „Cyberknife“, das Ausführungspläne in der Radiochirurgie erstellt, die Chirurgen überprüfen und genehmigen müssen.²⁵ An dieser Stelle ist allerdings zu betonen, dass bis zum heutigen Tag kein OP-Roboter in der Lage ist, Aufgaben zu übernehmen, die nicht auch von einem Menschen ausgeführt werden können, weil chirurgischen Strategien auf Datensätzen beruhen, die auf menschlich generierte Strategien zurückgehen.²⁶ Dass ein Roboter also besonders komplizierte Operationen übernimmt, zu denen wenige Daten vorhanden sind, ist zum jetzigen Zeitpunkt nicht denkbar.²⁷

23 Elish, Madeleine (2016): Letting autopilot off the hook, in: Slate. Online unter: <https://slate.com/technology/2016/06/why-do-blame-humans-when-automation-fails.html> [29.04.2020].

24 Yang, Guang-Zhong /Cambias, James/Cleary Kevin et al. (2017): Medical robotics – regulatory, ethical, and legal considerations for increasing levels of autonomy, in: Science Robotics, Vol. 2, Issue 4.

25 Ficuciello, Fanny/Tamburrini, Guglielmo/Arezzo, Alberto et al. (2019): Autonomy in surgical robots and its meaningful human control, in: Paladyn – Journal of Behavioral Robotics, Number 10, S. 30–43. Online unter: <https://www.degruyter.com/view/journals/pjbr/10/1/article-p30.xml> [29.04.2020].

26 Ibid.

27 Burgner-Kahrs, Jessica (2018): Vortrag: Autonome Roboter in der Medizin – Status Quo, Sitzung der interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz“ am 5.6.2018.

Das Zusammenspiel zwischen Mensch und Maschine kann allerdings wie schon beim Fliegen dazu führen, dass die Geschicklichkeit und praktischen Fähigkeiten des medizinischen Personals abnehmen, was im Notfall und bei Komplikationen möglicherweise zu Problemen führen kann. Denn auch auf höheren Autonomiestufen stehen die Roboter weiterhin unter der Kontrolle der behandelnden Ärzte, die wiederum verantwortlich dafür sind, was im OP-Saal geschieht. Doch selbst bei dem flächendeckend an deutschen Universitätskliniken eingesetzten, minimal invasiven Da-Vinci-Operationssystem, das keinen Autonomiegrad besitzt, sondern lediglich die Vorgaben der Chirurgen ausführt, sei immer mehr „technisches Know How gefragt“, weil die OP-Techniken für den Robotereinsatz angepasst würden.²⁸ Dieses technische Know-how sollte daher stärker in Studium, Aus- und Weiterbildung integriert werden.

Verändern im Operationssaal robotische Systeme die Anforderungen an einen Chirurgen, sind es in der Diagnose und Therapie KI-basierte Assistenzsysteme und Entscheidungshilfen, um Ärzten einfache, repetitive und fehleranfällige Aufgaben abzunehmen. KI-basierte Systeme in der Diagnostik erscheinen aktuell vor allem im Bereich der bildgebenden Verfahren vielversprechend zu sein, wie in der Krebserkennung, wo trainierte Modelle verdächtige Knoten in Brust oder Lunge, Metastasen oder Hautkrebs bereits zuverlässig erkennen. Hier ist es allerdings umso wichtiger, dass Ärzte über einen kompetenten Umgang mit Daten und ein Grundverständnis der Fehlerquellen und Schwächen des Maschinellen Lernens verfügen, denn die „erforderliche Plausibilisierung der maschinellen Unterstützung setzt voraus, dass sowohl die Daten, mit denen das System gearbeitet hat, als auch die Prozesse der daraus resultierenden Schlussfolgerungen vom Arzt nachvollzogen und beurteilt werden können.“²⁹ Das ist wichtig, weil ein System, das für einen bestimmten Kontext mit Daten einer bestimmten Region trainiert wurde und funktioniert, dies in einem anderen Umfeld schon nicht mehr tun muss, „[d]enn bereits kleine Unterschiede in den Patientenkollektiven, Scanning-Protokollen und Bilddatenmustern können die neuronalen Netze irritieren“.³⁰ Damit einhergehend ist beim Einsatz von KI-basierten Systemen als Entscheidungshilfe ein Statistikverständnis wichtig, ein Bereich, der nach einer

28 Koch, Christine (2018): Wer ist Chef im OP-Saal?, Niedersächsisches Ärzteblatt, Nummer 3/2018, S. 14–15.

29 Pressestelle der Deutsche Gesellschaft für Urologie e. V. (DGU) (2019): DGU stellt sich Herausforderungen durch Digitalisierung und künstliche Intelligenz. Online unter: <https://idw-online.de/de/news719305> [29.04.2020].

30 Lindner, Martin (2019): Die Diagnose kommt vom Computer, in: Neue Zürcher Zeitung, 26.07.2019. Online unter: https://www.nzz.ch/wissenschaft/ki-in-der-medizin-hilfe-bei-einfachen-und-repetitiven-aufgaben-ld.1497525?utm_source=pocket-newtab [29.04.2020].

Studie von Wissenschaftlerinnen und Wissenschaftlern des Harding-Zentrums für Risikokompetenz am Max-Planck-Institut für Bildungsforschung in der medizinischen Lehre vernachlässigt werde.³¹ Doch wenn Ärzte keine ausreichende Kompetenz in Statistik haben, „werden sie falsche Informationen auch an ihre Patientinnen und Patienten weitergeben.“³² Sollten KI-basierte Assistenzsysteme in Zukunft sogar verpflichtend werden, werden diese Kompetenzen noch wichtiger, weil sich Ärzte durch die technische Unterstützung stärker auf den persönlichen Kontakt und Beratung des Patienten konzentrieren könnten und die Patienten entsprechend über Wesen, Bedeutung und Tragweise bestimmter Behandlungsvorschläge aufklären müssten.³³

Auch hier ergibt sich wieder das bereits bekannte Paradox: Egal ob als Roboter im OP-Saal oder als Unterstützungssystem in der Diagnose, man setzt die Technik mit dem Ziel ein, menschliche Schwächen zu überwinden, doch bei steigendem Einsatz, können die Ärzte Kompetenzen verlieren bzw. verstärkt neue Kompetenzen im Umgang mit neuer Technik benötigen, die vorher noch nicht vorhanden waren. Versagt die Maschine jedoch, müssen die Ärzte wiederum einspringen bzw. Fehler erkennen und stehen dafür in der Verantwortung. Wie beim Fliegen und beim automatisierten Fahren können hier Trainings dem Kompetenzverlust entgegenwirken: OP-Roboter können als Trainingssimulator dienen, um Chirurgen technisch auszubilden, Statistik und Datenkompetenz muss in der ärztlichen Fort- und Weiterbildung eine Kernrolle spielen.

3. Kompetenz, Verantwortung und Haftung

Auch wenn man davon ausgeht, dass Piloten, Autofahrer oder Ärzte durch spezielle Trainings ihre Kompetenzen bewahren und neue Kompetenzen für den Umgang mit KI-Systemen erhalten, bleibt die Problematik, dass aktuell Menschen weiterhin die Verantwortung bei einem Versagen des technischen Systems – einem Flugzeugunglück, einem Unfall eines hochautomatisierten Fahrzeugs, einer

31 Jenny, Mirjam Annina /Keller, Niklas /Gigerenzer, Gerd (2018): Assessing minimal medical statistical literacy using the Quick Risk Test: A prospective observational study in Germany, in: BMJ Open, Nummer 8.

32 Max-Planck-Institut für Bildungsforschung Presse- und Öffentlichkeitsarbeit (2018): Viele Medizinstudierende verstehen Statistik nicht. Online unter: https://www.mpib-berlin.mpg.de/sites/default/files/press/2018-10-23_mpib_pm_medizinstudierende-statistik_de.pdf [29.04.2020].

33 Molnár-Gábor, Fruzsina (2020): AI in health care: doctors, patients and liabilities, in: Rademacher, Timo/Wischmeyer, Thomas (Hrsg.): Regulating AI, Wiesbaden, S. 337–360.

falschen Diagnose eines medizinischen Assistenzsystems – innehaben, wobei der Einfluss auf die Funktionsweisen des technischen Systems immer geringer wird.³⁴ Die jeweils verantwortlichen Ärzte könnten zum „Haftungsknecht“³⁵ werden, die Autofahrer zur „moralischen Knautschzone“, die die Verantwortung übernehmen müssen, um die Integrität des technologischen Systems aufrechtzuerhalten, obwohl die Kontrolle zwischen Mensch und Maschine geteilt wird.³⁶

Um diese Problematik zu lösen, werden verschiedene Ideen diskutiert. Zunächst kann man feststellen, dass die Verantwortung nicht nur bei den Anwendern liegen, sondern sich auf mehrere Schultern aufteilen sollte, inklusive Programmierern und Technikern. Diese sollten von Beginn an den Menschen in einer Mensch-Maschine-Interaktion – beim Flugzeug-Autopiloten genauso wie beim automatisierten Fahren – nicht als Störfaktor sehen, sondern das Design Anwender-zentriert gestalten, so dass die Entscheidungen des Systems nachvollziehbar sind und bei einem Ausfall oder Fehlentscheidungen angemessene Führung und Kontrolle sichergestellt sind.³⁷ Auf allgemeiner Ebene könnte man darüber nachdenken, die professionelle Verantwortung von Data Scientists, Informatikern und Softwareingenieuren dergestalt anzuerkennen, dass Studiengänge entsprechende verpflichtende Kurse über Ethik und gesellschaftliche Verantwortung beinhalten – bis zu einer Art „hippokratische[m] Eid für diese Berufe“.³⁸ Doch auch wenn man mehr Verantwortung an die Personen überträgt, die die Systeme entwickeln, bleiben weiterhin Menschen in der Verantwortung – es wird mitunter nur eine andere Person zum „Haftungsknecht“ gemacht. Gleichwohl können

34 Als weitere Felder, die unter einem Kompetenzverlust beim steigenden Einsatz von automatisierten Assistenz-Systemen leiden können, wären die Rechtsetzung oder die Polizeiarbeit denkbar, wenn sich Richter, Anwälte und Polizisten vermehrt auf KI-Systeme zur Unterstützung Ihrer Arbeit verlassen. Siehe dazu Rostalski, Frauke (2019): Legal Tech now and then, in: RETHinking:Law, 1/2019, S. 4–13; Rademacher, Timo (2017): Predictive Policing im deutschen Polizeirecht, in: Archiv des öffentlichen Rechts, Bd. 142, S. 366–416.

35 Beck, Susanne (2018): Podiumsdiskussion „Künstliche Intelligenz in der medizinischen Diagnose – bleibt der Mensch in der Verantwortung?“ der Berlin-Brandenburgischen Akademie der Wissenschaften, 8.10.2018, Aufzeichnung online unter: https://lisa.gerda-henkel-stiftung.de/kuenstliche_intelligenz_in_der_medizinischen_diagnose_bleibt_der_mensch_in_der_verantwortung?nav_id=7984 [29.04.2020].

36 Elish, Madeleine (2016): Letting autopilot off the hook, in: Slate. Online unter: <https://slate.com/technology/2016/06/why-do-blame-humans-when-automation-fails.html> [29.04.2020].

37 Elish, Madeleine/Hwang, Tim (2018): Praise the Machine! Punish the Human! The contradictory history of accountability in automated aviation, Comparative Studies in Intelligent Systems – Working Paper #1, S. 21/22. Online unter: https://www.datasociety.net/pubs/ia/Elish-Hwang_AccountabilityAutomatedAviation.pdf [29.04.2020].

38 Herzog, Lisa (2019): Woher wissen wir, dass KI-Experten richtig entscheiden?, Die Zeit, 31. Juli 2019. Online unter: <https://www.zeit.de/arbeit/2019-07/digitalisierung-kuenstliche-intelligenz-experten-software-verantwortung-beruf> [29.04.2020].

Programmierer nicht die Verantwortung der Autofahrer und Ärzte übernehmen. Dies kann problematisch werden, denn wenn die Nutzung der Technologie als allgemein gesellschaftlich gewünscht gilt – was momentan in den Diskursen über autonomes Fahren oder Assistenzsysteme in der Medizin verhandelt wird – „kann es sich unter Umständen als sachlich nicht angemessen erweisen, Einzelnen dafür die Verantwortung zu geben“,³⁹ vor allem wenn diese Personen technische Fähigkeiten besitzen und durch das aktuelle Recht abgeschreckt würden, diese einzusetzen.

Um einer „Verantwortungsdiffusion“ vorzubeugen und den Menschen hinsichtlich der Haftung aus der moralischen Knautschzone zu befreien, wird die Einführung einer „elektronischen Person“, die Roboter und KI-Systeme vertritt, diskutiert. Diese würde vergleichbar mit der schon existierenden juristischen Person die verschiedenen Beteiligten vereinen, so dass sich der Geschädigte „direkt an diesen Adressaten wenden [könnte], ohne nachweisen zu müssen, wer von den verschiedenen Beteiligten ihn geschädigt hat.“⁴⁰ Überlegungen zu einer Rechtsperson für KI oder Roboter werden wiederum kritisiert, da „treating robots like humans would dehumanize humans, and therefore we should refrain from adopting this policy“. Dies gelte auch, wenn man den Vergleich mit juristischen Personen heranziehe, weil beispielsweise Unternehmen als Akteure im Gegensatz zu automatisierten, KI-basierten Systemen immer noch durch Menschen handeln.⁴¹ Wir befinden uns aktuell mitten in der Diskussion darüber, wie wir menschliche Verantwortung in verschiedenen Berufsfeldern und im alltäglichen Leben neu denken und abgrenzen können, damit der Umgang mit der Technik unseren gesellschaftlichen und ethischen Überzeugungen entspricht.

4. Black Box oder “Explainable Artificial Intelligence”?

Zu der oben beschriebenen Situation, dass Menschen im Umgang mit KI spezielle Ausbildungen und Schulungen benötigen, um mit den Funktionsweisen und Vorhersagen von immer autonomer werdenden Maschinen umgehen und sie einordnen zu können, gesellt sich jedoch noch eine fundamentale technische

39 Rostalski, Frauke (2019): Verantwortung und künstliche Intelligenz, in: REthinking:Law, 1, S. 29.

40 Beck, Susanne (2017): Verantwortungsdiffusion im Kontext autonomer Maschinen, in: Bulletin der Schweizerische Akademie der Geistes- und Sozialwissenschaften, 4, S. 64–65.

41 Eidenmüller, Horst (2017): The rise of robots and the law of humans, in: Zeitschrift für Europäisches Privatrecht (ZEuP), 25, S. 765–777, online unter: <https://ssrn.com/abstract=2941001> [29.04.2020].

Herausforderung, die in dieser Form ein Novum darstellt. Bei KI-Systemen, die eben nicht mehr nach festen Codierregeln funktionieren, sondern die basierend auf tiefen neuronalen Netzen selbstständig „lernen“, also Muster in großen Datenmengen erkennen, sind die Funktionsprinzipien undurchsichtig, denn das System gibt von sich aus zunächst keine Information über die Muster, die es lernt, und warum es bestimmte Ergebnisse liefert. Man spricht von einer „Black Box“. Mit dem Öffnen der Black Box, also der Entwicklung von Methoden, um die Ergebnisse erklären zu können, befasst sich ein eigener Forschungszeitweig, der sich als „Explainable AI“ zusammenfassen lässt.

Die Frage nach der Erklärbarkeit und nach dem Verständnis, warum und wie solche KI-Systeme zu Vorhersagen und Entscheidungen kommen, steht mit den ethischen Fragen nach der Verantwortung und rechtlichen Fragen nach der Haftung in der Mensch-Maschine-Interaktion in direktem Zusammenhang. Können Autofahrer oder Ärzte für die Befolgung oder Nichtbefolgung von KI-Vorhersagen zur Verantwortung gezogen werden, wenn wir gar nicht verstehen, wie die Vorhersagen des Systems zustande gekommen sind? Es ergibt sich ein Vertrauensproblem, weil wir nicht wissen, was die Algorithmen letztendlich lernen. Ein Beispiel ist das KI-System „Deep Patient“, das im Jahr 2015 von einer Forschergruppe am Mount Sinai Hospital in New York mit den umfangreichen Daten aus 700.000 Patientenakten des Krankenhauses trainiert wurde.⁴² „Deep Patient“ erwies sich als sehr effektiv bei der Vorhersage von Krebserkrankungen und psychiatrischen Störungen wie Schizophrenie, weil es in den großen Datenmengen bestimmte Muster identifizierte, die für die medizinischen Experten nicht zu erkennen waren. Doch das System lieferte keine Erklärungen, warum es zu diesen Ergebnissen kam. Ohne Erklärungen waren die Vorhersagen allerdings wenig brauchbar, denn in welchem Krankenhaus wird eine ärztliche Behandlung – die mit massiven Nebenwirkungen verbunden sein kann – aufgrund einer Computervorhersage begonnen, die niemand nachvollziehen kann? Genauso wenig können wir beim autonomen Fahren wissen, was die Algorithmen eigentlich lernen. Hier ergibt sich freilich ein Dilemma, denn die aktuell entwickelten und getesteten autonomen Fahrzeuge, die auf tiefen neuronalen Netzen aufbauen, sind angeblich auch die, die am besten funktionieren. Ist die Erklärbarkeit vernachlässigbar, wenn das System mit der Black Box vermeintlich die besten Vorhersagen und Entscheidungen trifft?⁴³

42 Knight, Will (2017): The dark secret at the heart of AI, in: Technology Review. Online unter: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> [29.04.2020].

43 Holm, Elisabeth (2019): In defense of the black box, in: Science, Vol. 364/ Issue 6435, S. 26–27.

Doch auch Systeme, die anscheinend sehr gute Ergebnisse liefern, können falsch liegen. Die Erklärbarkeit zu vernachlässigen, macht es schwierig, mögliche Fehler vorherzusehen, die für Menschen mitunter absurd erscheinen. Als Beispiel kann ein neuronales Netz dienen, das zuverlässig Pferdebilder erkannte, sich dabei aber nicht auf das abgebildete Pferd, sondern auf die Copyright-Angabe von Pferdeforen stützte – es hätte wohl auch ein Bild, auf dem ein Nilpferd zu sehen ist, mit dieser Copyright-Angabe als Pferd klassifiziert. Der fälschlichen Bildererkennung von Pferden kamen Wissenschaftler auf die Spur, indem sie zurückverfolgten, wie das neuronale Netz bestimmte Bereiche des Bildes gewichtete und sie diese Bereiche in „Heatmaps“ farblich hervorhoben.⁴⁴ Wenn man sich solche im Grunde falschen Lernvorgänge beim autonomen Fahren oder in der medizinischen Diagnostik vorstellt, wird die Notwendigkeit, erklären und verstehen zu können, was ein KI-System lernt, recht offenkundig, weil ein falsches Ergebnis tödliche Folgen haben kann. Daneben kann es in Wissenschaft, Wirtschaft und Gesellschaft keinen Erkenntnisgewinn geben, wenn ein KI-System bisher ungeahnte Zusammenhänge in großen Datenmengen findet, diese allerdings – wie im obigen Beispiel mit „Deep Patient“ – unbrauchbar sind, weil wir nicht bewerten können, ob die Entscheidung auf einer validen oder fälschlich ausgemachten Korrelation in den Daten beruht, denn letztere „will likely fail to provide correct classification and thereby usefulness once it is deployed in the real world, where spurious or artifactual correlations may not be present.“⁴⁵

Damit wir nicht auf das womöglich bessere Ergebnis zugunsten eines weniger komplexen, dafür aber erklärbaren Systems verzichten müssen, sollte die Forschung im Bereich der „Explainable AI“ priorisiert und ausgeweitet werden. Die Vorhersagen komplexer KI-Systeme zu erklären, ist schwierig und aufwändig. Dennoch muss es das Ziel sein, die Veränderungen der Algorithmen bzw. „Lernvorgänge“ in der Black Box parallel so aufzuzeichnen und zu dokumentieren, dass wir sie abrufen, kontrollieren, ausprobieren und gegebenenfalls zertifizieren können.

44 Montavon, Grégoire/Wojciech, Samek/Müller, Klaus-Robert (2018): Methods for interpreting and understanding deep neural networks, in: Digital Signal Processing, 73, S. 1–15.

45 Lapuschkin, Sebastian/Wäldchen, Stephan/Binder, Alexander et al. (2019): Unmasking clever hans predictors and assessing what machines really learn, in: Nature Communications, 10.

III. FAZIT

Die aktuelle Diskussion über KI hat viele Facetten. In der Mensch-Maschine-Interaktion, die immer alltäglicher wird, ist es für das Design der Maschine und den rechtlichen Rahmen entscheidend anzuerkennen, dass der Mensch kein Störfaktor in einem sonst reibungslos funktionierenden System ist. Der Dualismus zwischen technischem Versagen einerseits und menschlichem Versagen andererseits bringt uns nicht weiter, wenn Menschen zunehmend Kompetenzen abgeben, aber weiterhin in der Verantwortung stehen. Die über 100-jährige Geschichte des Flugverkehrs, in der der erste Autopilot bereits 1912 im Einsatz war, erscheint hier weniger ein hilfreiches Vorbild, als ein Vorreiter eines sich weiter ausdehnenden Missverhältnisses zu sein.

Abhilfe können eine bessere Ausbildung sowie regelmäßige Trainings für die Anwender bieten, um die jeweiligen Systeme besser einschätzen zu können. Auch sollten Systementwickler zuvorderst den Menschen und seine Bedürfnisse ins Zentrum der Entwicklung automatisierter Systeme stellen. Doch die Grundproblematik bleibt bestehen, nämlich dass insbesondere politische Systeme auf ein Verständnis von Algorithmen als „bloße Instrumente menschlicher Handlungen“ zurückgreifen, vor allem dann, „wenn etwas schiefgeht und verantwortliche Akteure genannt werden müssen“, so dass die Verantwortung an humane Akteure gebunden wird.⁴⁶ Doch gerade wenn wir Verantwortung an Menschen binden, weil sie im Gegensatz zu Maschinen zur Selbstbestimmung fähig sind, müssen wir uns als Gesellschaft überlegen, wie wir den ethischen und rechtlichen Rahmen so gestalten können, dass es weder zum Schaden für die einzelnen Akteure noch für die Gesellschaft kommt.

⁴⁶ Seyfert, Robert/Roberge, Jonathan (2017): Was sind Algorithmenkulturen?, in: Seyfert, Robert/Roberge, Jonathan (Hrsg.): Algorithmenkulturen, Bielefeld, S. 31.

**TITEL DER REIHE »#VERANTWORTUNGKI – KÜNSTLICHE INTELLIGENZ
UND GESELLSCHAFTLICHE FOLGEN«**

Heft 1/2020

Isabella Hermann, Georgios Kolliarakis, Fruzsina Molnár-Gábor,
Timo Rademacher, Frauke Rostalski

VERTRAUENSWÜRDIGE KI?

VORAUSSCHAUENDE POLITIK!



Der zunehmende Einsatz von sogenannter „Künstlicher Intelligenz“ (KI) verspricht viele Verbesserungen, beispielsweise durch Bilderkennung in der Medizindiagnose. Er birgt aber auch das Risiko, dass Menschen durch irrtümliche Vorhersagen von KI-Systemen zu Schaden kommen können. In solchen Fällen wird es immer schwieriger zu bestimmen, wer die Verantwortung trägt. Die Reihe #VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen bietet ein Forum für Beiträge über die ethischen, rechtlichen und gesellschaftspolitischen Chancen und Risiken des Einsatzes von KI mit einem besonderen Blick auf den Verantwortungsbegriff. Die Beitragsreihe wird von der interdisziplinären Arbeitsgruppe *Verantwortung: Maschinelles Lernen und Künstliche Intelligenz* betreut.