

12. Neue Intelligenz, neue Ethik?

Julian Stubbe, Jan Wessels, Guido Zinke

Aus den vergangenen Jahren gibt es unzählige Beispiele für die beeindruckenden Entwicklungen im Bereich der Künstlichen Intelligenz (KI). Voran getrieben werden sie durch die jetzt verfügbaren Datenmengen, Rechenleistungen und besonders durch die großen Fortschritte im maschinellen Lernen (Machine Learning, ML), insbesondere dem tiefen Lernen (Deep Learning, DL). Dies wird erhebliche Einwirkungen auf sämtliche sozialen, politischen und ökonomischen Systeme haben.

Nicht allein die Zeitschrift *Economist* sieht einen fundamentalen Wandel nahen, durch den künftig nicht mehr die Maschine vom Menschen lerne, sondern umgekehrt. Und die Maschine lerne viel schneller und eigne sich dank ihrer wachsenden Leistungsfähigkeit menschliche Eigenschaften wie Kreativität, Intuition, Iteration und Impulsivität an (The Economist 2016). Einige Autoren sehen in der KI eine existenzielle Herausforderung der Menschheit, u. a. der kürzlich verstorbene Stephen Hawking oder der Tech-Unternehmer Elon Musk. Sie leugnen nicht den Nutzen der KI, erwarten aber einen Wettlauf mit einer möglichen „Superintelligenz“, den der Mensch u.a. deshalb verliere, weil er in evolutionären Prozessen nicht schritthalten könne (Handelsblatt 2014, FAZ 2017).

Vor diesem Hintergrund ist die Auseinandersetzung mit der ethischen Dimension von KI von einer rein akademisch-philosophischen zu einer gesellschaftlichen und innovationspolitischen Debatte gewachsen. Das zunehmende Interesse an der Ethik einer KI kommt nicht von ungefähr, schließlich greifen Algorithmen tief in gesellschaftliche Zusammenhänge ein und sind nicht auf spezifische Anwendungsdomains beschränkt. KI verändert Wertschöpfungsprozesse genauso wie die private Kommunikation und die Interaktion der Menschen. Und dieser Einfluss wird insbesondere dann deutlich wachsen, wenn der Schritt von der jetzigen schwachen KI (durch den Menschen vorgegebene bzw. programmierte Algorithmen) hin zu einer starken KI (die sich selbst gestaltet bzw. fortentwickelt, siehe Einleitung zu Teil A Technologie „Entwicklungswege zur KI“) gegangen wird. Umso drängender sind Antworten auf jene Fragen zu finden, was diese Schlüsseltechnologie mit unserer Gesellschaft macht und wie sie die uns bekannten Lebens- und Arbeitswelten verändern wird. Populäre Technologiema-gazine greifen das Thema vermehrt auf und bereiten es verständlich für eine breite Bevölkerungsgruppe auf. Öffentliche Institutionen reagieren parallel, z. B. durch den Deutschen Ethikrat oder in Frankreich durch die nationale Datenschutzbe-

hörde und eine breitangelegte gesellschaftliche Debatte. Ebenso wächst das private bzw. unternehmerische Engagement, wenn internationale Initiativen wie etwa OpenAI einen demokratischen und insbesondere weniger Eliten-zentrierten Zugriff auf KI zu ermöglichen und die Souveränität des Menschen im Umgang mit der Maschine sowie die Akzeptanz zu stärken suchen.

Solche Bestrebungen erfahren aktuell eine große Resonanz, auch ausgelöst durch eine wachsende KI-Skepsis in der Bevölkerung. Aktuelles Beispiel für ein neues Bewusstsein ist Facebook. Das Unternehmen geriet zunehmend unter öffentlichen Druck, seine gesellschaftliche Verantwortung wahrzunehmen und manipulative Praktiken zu unterbinden. Dies zeigt, wie der ausgeprägte Glaube an das enorme technologische Lösungspotenzial einer KI vielerorts zur Annahme führte, ihre Innovationen würden auf große Akzeptanz in der Gesellschaft stoßen. Ein solcher „technischer Imperativ“ findet aber gleichwohl nicht unbedingt das erhoffte positive Echo in der Gesellschaft. Stattdessen antizipieren Anbieter von KI-Technologien mittlerweile selbst Missbrauch und Akzeptanzverlust und leiten entsprechende Maßnahmen in Richtung eines ethischen Imperativs einer KI ein.

In der aktuellen Debatte über die ethische Dimension von KI steht deshalb eine Frage im Mittelpunkt: Wirft KI neue, über den bestehenden Diskurs hinausgehende ethische Fragen auf, oder verstärkt sie bereits im Vorfeld existierende Spannungsfelder?

Autonomes Entscheiden

Der zentrale Nutzen von KI besteht in ihrer prinzipiellen Fähigkeit, ohne menschliche Einwirkung selbstständig – autonom – Entscheidungen zu treffen. Zur Diskussion steht dabei die Frage, inwieweit und wann Maschinen dies tun sollten und können und welche Konsequenzen dies hat. Ethisch relevant sind vor allem negative Folgen, wenn also Menschen zu Schaden kommen oder Dinge beschädigt werden. In diesem Fall ist zu fragen, wer die Verantwortung dafür trägt bzw. wer dafür haftet. Ein Algorithmus ist schließlich keine juristische Person. Um die Koexistenz von Mensch und Maschine zu organisieren, definiert die Ethik für autonome Entscheidungssysteme also vor allem den Aspekt der Verantwortung.

Sehr eingängig lässt sich dieses Dilemma am Beispiel des autonomen Fahrens nachzeichnen. Da die Maschine keine juristische Person ist, sollte im Schadensfall geklärt werden können, wer letztlich verantwortlich ist. Typischerweise hat der menschliche „Fahrer“ im Moment der Schadensentstehung keinen Einfluss auf die Entscheidungsfindung gehabt, kann also nicht verantwortlich sein bzw. werden. Dies gilt auch für den Programmierer des autonomen Systems, der möglicherweise nur die Ausgangsversion einer Software geschrieben hat, die sich dann selbstlernend wei-

terentwickelte. Dennoch müssen Geschädigte eine Chance auf Entschädigung haben.

Deshalb stehen unter anderem auch ethische Aspekte auf der Tagesordnung der Politik. So wurden entsprechende Experten-Gremien etabliert oder parlamentarische Auseinandersetzungen initiiert. Für Deutschland besonders relevant ist die Einrichtung einer Ethikkommission zum autonomen Fahren durch das Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) im Jahr 2016 (BMVI 2016) sowie die Entschließung des Europäischen Parlaments zu Robotik und Künstlicher Intelligenz Anfang 2017. Diese fordert, für das autonome Fahren eine Pflichtversicherung einzuführen sowie einen Zusatzfonds, der die nötigen Mittel für eine hinreichende Entschädigung gewährleisten soll (Europäisches Parlament 2017). Auch der Deutsche Ethikrat hat sich wiederholt mit dem Thema KI auseinandergesetzt, zuletzt auf seiner Jahrestagung im Juni 2017 (Deutscher Ethikrat 2017). Schließlich sieht der Koalitionsvertrag zwischen CDU, CSU und SPD aus dem Februar 2018 eine Ethikkommission zu digitalen Fragen vor, die auch den Aspekt KI behandeln soll (Bundesregierung 2018).

Der Diskurs um die Verantwortlichkeit klärt also in erster Linie, welche Rollenverständnisse sich zwischen Mensch und Maschine ausprägen. Ebenso ethisch relevant und damit intensiv diskutiert ist die Frage, nach welchen Kriterien autonome Entscheidungen überhaupt zustande kommen. Dies betrifft in erster Linie die Würde des Menschen. Denn hier existieren klassische Dilemmata, die vorab durchdacht und auf ihre Relevanz geprüft werden müssen. So kann ein Entscheider vor der Alternative stehen, zwischen zwei Optionen wählen zu müssen, die beide negative Auswirkungen haben. In der Literatur ist dies als Trolley-Problem bekannt (Heise 2017).

Dazu hat z. B. die Ethikkommission zum autonomen Fahren des BMVI in ihrem Abschlussbericht festgehalten, dass grundsätzlich Sachschaden einem Personenschaden vorzuziehen ist, also Menschen auf jeden Fall geschützt werden müssen. Eine Einteilung von Menschen in verschiedene Gruppen mit unterschiedlichem Schutzstatus darf es aus Sicht der Ethikkommission nicht geben. Alter, Geschlecht oder ähnliche Faktoren dürfen also bei Entscheidungen eines autonomen Systems keine Rolle spielen, junge Menschen dürfen z. B. keinen höheren Schutz genießen als Alte oder Kranke.

Daran knüpft sich ein weiterer wichtiger Aspekt, der zweifelsohne einen erheblichen Einfluss darauf hat, ob KI sich durchsetzt und Akzeptanz findet. Denn die Frage nach den Entscheidungskriterien verlangt, dass die Entscheidungsprozesse selbst und die ihnen zugrunde liegenden Annahmen transparent sind – und zwar auf Seiten des Menschen und der Maschine. Es muss nachvollziehbar sein, auf welcher Basis Entscheidungen getroffen wurden.

Transparenz ist auf der technischen Ebene allerdings keine leicht zu lösende Aufgabe, denn selbstlernende Systeme lassen eine Nachvollziehbarkeit nicht ohne Weiteres zu. Damit sind auch mögliche diskriminierende Kriterien nicht sichtbar. So hat sich bei der Nutzung von KI im amerikanischen Justizsystem gezeigt, dass Minderheiten benachteiligt werden und damit das Gleichheitsgebot verletzt wird. Eine wichtige Forderung der Kritiker von KI lautet deshalb, die Transparenz der Entscheidungsfindung wirklich sicherzustellen oder zumindest Kontrollsysteme zu entwickeln, die im Nachhinein die Entscheidungsfindung rekonstruieren können (reverse engineering). Wie sich so etwas gestalten kann, zeigt New York City. Dort gewährleistet künftig eine eigens geschaffene Verordnung, dass die vielfältig eingesetzten Algorithmen in der Stadtverwaltung eingesehen werden können und Bürger das Recht auf umfangreiche Information erhalten. Damit soll eine Diskriminierung aufgrund unzureichender oder fehlerhafter KI verhindert werden, und die Bürger haben gegebenenfalls eine realistische Möglichkeit, dagegen juristisch vorzugehen (Netzpolitik 2018).

Im ethischen Diskurs, ob, inwieweit und entlang welcher Kriterien Maschinen im Zusammenspiel mit dem Menschen autonom entscheiden, geht es also in erster Linie um das Ausmaß der Verantwortlichkeit, die der Mensch der Maschine in der Organisation einer künftigen Koexistenz überträgt. Im Kern drückt sich damit aus, ob der Einsatz oder Nicht-Einsatz künstlicher Intelligenz einen Unterschied macht – oder besser: bis zu welchem Grad autonomen Entscheidens der Einsatz von KI einen Unterschied macht. Die ethischen Anforderungen an autonome entscheidende Systeme und den Grad der Verantwortungsübertragung von Mensch auf Maschine ergeben sich also aus den antizipierenden Abwägungen, inwieweit die durchaus erheblichen Vorteile für viele Menschen in einem verträglichen Verhältnis zu den in spezifischen Situationen auftretenden Nachteilen für einzelne Betroffene stehen. Dies illustrieren die Festlegungen der Ethikkommission zum autonomen Fahren sehr gut: In Anbetracht der Erwartungen, dass der Einsatz von KI beim autonomen Fahren die Zahl der Verkehrstoten voraussichtlich drastisch verringern wird, ist ihr Einsatz selbst bereits ethisch geboten, um Menschenleben zu schützen. Dies gilt auch, wenn im Einzelfall Menschen durch fehlerhafte KI zu Schaden kommen.

Neues Zusammenleben und -arbeiten

Im Mittelpunkt des Diskurses um die ethischen Anforderungen an das Zusammenleben und -arbeiten von Mensch und KI stehen Fragen um die Übertragung der Leistungsfähigkeiten des Menschen auf die Maschine. Hierdurch erweitert sich der Diskurs zu den ethischen Anforderungen autonomer Verantwortung um Aspekte einer sukzessiven Übertragung von Aufgaben des Menschen an die Maschine. Dies mündet letztlich in der Frage der Rolle des Menschen in einer Koexistenz mit der Maschine. Allein in den vergangenen fünf Jahren ist KI sehr sicht- und spürbar in die Lebens-

und Arbeitswelten vorgedrungen. Im Krankenhaus assistieren OP-Roboter, im Büro und im Alltag unterstützen smarte Assistenten den Menschen und im Verkehr ermöglichen intelligente Sensoren autonomes Fahren. Kontinuierlich wachsen dadurch die Interaktionen mit intelligenten Systemen, aus dem „artificial“ in AI, der englischen Abkürzung für KI, wird ein „augmented“. KI wird allgegenwärtig. Dabei zeigen sich Nutzen, aber auch Risiken immer deutlicher. Etwa, wenn Algorithmen diskriminieren, weil sie aus den Informationen über die Handlungen der Menschen lernen (Süddeutsche 2017b), wenn autonome Systeme sich nicht mehr steuern lassen, wenn Chatbots öffentliche Diskussionen dominieren oder wenn KI-Algorithmen zu kriminellen Zwecken eingesetzt werden. Nicht zu reden von den Gefahren intelligenter Waffensysteme (Kleinberg et al. 2017).

Dass KI immer mehr Aufgaben übernimmt, wird hier mitunter zum Problem. Und es erweist sich, dass ein Zusammenleben mit KI nicht zwangsläufig ein besseres sein muss. Das Selbstverständnis des Menschen prägt sich dabei entlang der Entscheidung aus, ob er in eine existentielle Konkurrenz mit der KI eintritt, oder aber ob er die Oberhand behält, indem die Position einer „Superintelligenz“ klar geregelt ist und KI ihm sehr kontrolliert assistiert und sukzessive Aufgaben abnimmt.

Dafür braucht es ethische und institutionelle Arrangements, die nicht die technologischen Möglichkeiten beschränken, wohl aber die Risiken benennen. Es geht dann nicht nur darum, wer Entscheider und wer verantwortlich ist, sondern auch darum, wer im Zusammenspiel Mensch-Maschine die Kontrolle behält und wie der Mensch einen eigenen Willen aufrechterhalten und schlussendlich auch durchsetzen kann. Die Beantwortung dieser Fragen ist aufgrund der enormen Komplexität von KI-Systemen, der Datenmenge und schieren Entwicklungsgeschwindigkeit alles andere als einfach. Inwieweit die Algorithmen fehlerfrei arbeiten und die Daten korrekt sind, kann der einzelne Nutzer kaum noch einschätzen. Und selbst Entwickler sind zum Teil überrascht von der sprunghaft steigenden Leistungsfähigkeit der KI. Der Aufbau von Transparenz und Überwachungsstrukturen, von Standards und Sanktionsmustern ist darum die Grundvoraussetzung für eine ethisch verantwortungsvolle Nutzung. Der kritische Umgang mit den KI-Systemen und ihren Ergebnissen schaffen schließlich erst einen zumindest gesellschaftlich verträglichen Durchsatz von KI. Dafür braucht es Vertrauen, basierend auf mehr menschlicher Souveränität und Kompetenz im Umgang mit KI.

Viele Akteure – nicht nur KI-Kritiker, sondern auch die Politik und nicht zuletzt starke Profiteure von KI wie Google oder IBM – sind in den vergangenen Jahren in dieser Frage aktiv geworden. So lassen die Vereinten Nationen die Potenziale von KI auch unter ethischen Gesichtspunkten prüfen (UN 2017). Initiativen wie AINOW oder OpenAI wollen globale Standards für KI etablieren, um sie zu demokratisieren und vor allem weniger auf Eliten zentriert zu gestalten. Alle sollen KI nutzen können und

KI soll allen Nutzen bringen (AINOW 2018, BBC 2015). Google, Apple, Facebook und Amazon (GAFA) stellen KI-Tools als Open Sources bereit, um KI-Souveränität zu fördern und selbst von der Nutzung im Sinne einer Schwarmintelligenz zu profitieren (Die Zeit 2016b). Im „Partnership on AI“ liefern GAFA und andere KI-Lösungen für globale Probleme, u. a. um fehlerhafte KI-Systeme zu stoppen (MIT TechRev 2016a, MIT TechRev 2016b, CBR 2017). Für Europa denkt man an die Einrichtung einer spezialisierten Forschungseinrichtung für KI, an der die europäischen Kompetenzen wirkmächtig zur Untersuchung von KI-Algorithmen gebündelt werden. Ein solches „KI-CERN“ soll aber nicht nur helfen, technologisch mit USA oder China mithalten, sondern auch sozio-ökonomische Implikationen besser abschätzen zu können (TechCrunch 2017). Der Europäische Wirtschafts- und Sozialausschuss fordert ein europäisches OpenAI (EESC 2017), während die EU-Kommission eine AI-on-Demand-Plattform für Europa prüft (AI Business 2017). In Deutschland diskutiert die Politik zentrale ethische Fragen ebenso wie die Wirtschaft, die sich u. a. ein System einer Corporate Digital Responsibility geben möchte (Süddeutsche 2018). Und in Frankreich findet eine breite öffentliche Debatte um die Gestaltung des Zusammenlebens und Zusammenarbeitens mit KI statt (CNIL 2018).

All diese Initiativen beziehen sich in erster Linie auf den Umgang mit der aktuellen „schwachen“ KI, der Menschen die Muster vorgeben bzw. von vornherein einprogrammieren. Ethische Anforderungen richten sich hierbei zunächst konsequenterweise nicht an die Maschine, sondern an den Menschen. Anspruchsvoller wäre die Formulierung geeigneter ethischer Anforderungen für eine deutlich höher entwickelte „starke“ KI. „Einfach den Stecker zu ziehen“ (Die ZEIT 2017) würde hier nicht mehr gelingen, stattdessen bräuchte es eine echte Maschinenethik, eine artificial morality. Diese würde sich dann nicht mehr an den Menschen richten, sondern an die Maschine selbst.

Die einfachste, heute diskutierte Lösung ist, komplexere KI permanent zu überwachen, während ein menschlicher Benutzer die Aufgaben vorab definiert hat und immer noch sämtliche ethisch relevanten Entscheidungen trifft. Dies ist in einigen Fällen sicher auch künftig denkbar. Dort jedoch, wo KI extra deshalb eingesetzt wird, um menschliche Arbeit zu ersetzen, um schnelle Entscheidungen herbeizuführen, um Gefahren für den Menschen zu umgehen oder aber den Menschen selbst als Risikofaktor zu ersetzen, ist eine permanente Überwachung langfristig nicht realistisch (Ethik-Kommission AVF 2017). Denn es werden immer weniger Entscheidungssituationen vorhersehbar und damit Handlungsmuster programmierbar sein. Im Zusammenleben mit Menschen sollten Maschinen daher selbst ethisch handeln und ein solches Handeln aus dem Umgang mit ihnen erlernen (Arkin 2009).

Dafür müssen Maschinen grundsätzlich in die Lage versetzt werden, mit ihrer Umwelt zu interagieren, ihr Handeln an sich ändernde Bedingungen anzupassen und letztlich

eigenständig zu agieren. Solche Verhaltensmuster können auch heute schon einige KI-Systeme leisten. Entwickelt KI zusätzlich die Fähigkeit, aus der Verarbeitung von Informationen Gründe für ihr Verhalten – dank Belief-Desire-Intention-Anwendungen – abzuleiten, wird sie grundsätzlich einem ethisch handelnden Akteur ähnlicher (Floridi und Sanders 2004, Dennett 1987). Noch erreicht die KI auf dieser Entwicklungsstufe keine ganzheitliche Handlungsfähigkeit, die den Menschen auszeichnet, sondern lediglich die Fähigkeit, Ethik einzuschätzen und zu erkennen. Und dies ist auch nur auf einen bestimmten Handlungskontext beschränkt. Um ganzheitlich ethisch handlungsfähig zu sein, müsste sie Willen selbst ausprägen und Bewusstsein entwickeln können. Beides würde auch bedeuten, Emotionen zu empfinden – auch solche, die irrationales Handeln verursachen können (Scheutz 2011).

In einem anthropozentrischen Weltbild wird eine freundliche, weil im Umgang mit dem Menschen ethisch korrekt agierende KI benötigt. Noch besser: Sie sollte sogar korrekter als der Mensch handeln. Gerade hierin liegt auch ein großer Nutzen für die Koexistenz Mensch-Maschine: Jede Maschine agiert rational, weil ihr Handeln keiner emotional fundierten Willensbildung unterliegt und äußere Einflüsse die Rationalität kaum beeinträchtigen (Arkin 2009). Dies macht sie für das Zusammenleben berechenbarer, nachvollziehbarer und letztlich steuerbarer. Und: Die ethischen rationaleren Handlungsmuster werden in gewisser Weise zu Blaupausen für die Gesellschaft selbst. Eine so sich evolutionär ausprägende artificial morality ist dann jener Spiegel, den die KI den Menschen vorhält.

Veränderte Identitäten

Verändert die wachsende Ausprägung von KI bekannte Verantwortungsmuster und das Aufgabenverhältnis im Zusammenleben und -arbeiten, bleibt dies nicht folgenlos für das Rollen- und letztlich Selbstverständnis des Menschen. Mit dieser Frage dringt die KI-Ethik in die Bereiche individueller, menschlicher Identitätsbildung und Sozialisation vor.

Inwiefern KI Zugriff auf diese intimste, mithin wichtigste Ebene der Gesellschaft erlangen wird, hängt technisch gesehen davon ab, wie viele Daten wir über uns selbst erzeugen, wie wir mit diesen umgehen und in welchem Maß wir verstehen, wie KI diese Daten nutzt. Bei der Frage nach den Auswirkungen von KI auf unser Selbstverständnis verschmelzen somit die Diskurse um Verantwortung, Aufgaben und Rolle des Menschen in der Koexistenz mit KI. Die entstehenden Verantwortungsmuster in sich wandelnden Lebens- und Arbeitswelten münden also letztlich in neuen Sozialisationsformen, die auf neue Art identitätsstiftend bzw. verändernd wirken werden. Statt also KI als isolierten Faktor zu betrachten, gilt es vielmehr, ihre ethischen Implikationen in einem Netz verschiedener Veränderungen zu verstehen.

Welche Effekte KI auf die Bildung und Entfaltung von Identität hat, ist keine neue Frage. Bereits 1984 untersuchte Sherry Turkle aus mikro-soziologischer Perspektive die Wirkung der Auseinandersetzung mit KI auf die Persönlichkeitsentfaltung von Wissenschaftlern (Turkle 1984). Sie stellte fest, dass KI häufig zu einem „evozierenden Objekt“ wird, anhand dessen Menschen sich selbst und ihr Wesen hinterfragen. Intelligenz künstlich zu produzieren bzw. zu simulieren knüpft sich unter diesen Annahmen an die Frage, was Intelligenz denn eigentlich ist und welche Aspekte davon uns zum Menschen machen.

Diese Auseinandersetzung ist nicht allein akademischer Natur, sondern ein häufiges Motiv in Literatur und Film für ein großes Publikum. Dass etwa die Unterscheidbarkeit von Mensch und Maschine zu einer gesellschaftlichen und exekutiven Herausforderung werden kann, war bereits im Jahr 1962 der Plot in Philip K. Dicks Roman „Träumen Androiden von elektrischen Schafen?“, den der Regisseur Ridley Scott zuerst 1982 in „Blade Runner“ und fortgesetzt 2017 in „Blade Runner 2049“ sehr erfolgreich inszenierte. Es gelang, das enorme Spannungsfeld zwischen KI und Menschen eindrücklich darzustellen. Die Auswirkungen der emotionalen Bindung eines Menschen zu einer KI ist wiederum Thema im Film „Her“. Und was passieren kann, wenn die Maschine vom Werkzeug zum Partner mit Intention und Persönlichkeitsmerkmalen wird, zeigt der Spielfilm „Ex Machina“. Das dort dargestellte manipulative Wesen der Maschine ist jedoch nicht vorprogrammiert, sondern Resultat und Reflex auf Informationen, mit der die KI durch Online-Suchanfragen gefüttert wird.

Die ethische Diskussion über das Verhältnis von Mensch und KI ist also auch Teil populärer Kultur, die so auch Reflexion beinhaltet. Dennoch ist der ethische Gehalt der Science Fiction begrenzt, denn sie liefert wenig Orientierung für das praktische Handeln. Die praktische Relevanz der Reflexion wird jedoch in zwei Dimensionen der Identitätsbildung ersichtlich: einer materiellen und einer sozialen.

Die materielle Dimension der Identität betrifft die Art, wie Menschen durch den Umgang mit Dingen ihre Persönlichkeit finden und ausdrücken. Der Anthropologe Daniel Miller versteht Identität und Materialität als zwei zusammengehörige Elemente: Erst durch Interaktion mit der materiellen Welt finden Menschen heraus, wer sie sind, was sie gestalten können und wo ihre Grenzen liegen (Miller 2014). Aus dieser Perspektive entstehen kulturelle Typen, wie z. B. Sari tragende Frauen, erst im Zusammenspiel von sozialen Werten und materiellen Dingen, in diesem Fall einem leichten Stoff, der verhüllt und gleichzeitig sanft fällt. Dieses Prinzip führen Menschen online in sozialen Netzen fort, wenn sie sich mit Fotografien inszenieren. KI jedoch beginnt, Identität zu entmaterialisieren, indem sie repräsentierte Gegenstände auf ihren informatorischen Gehalt reduziert – Haptik, Ecken, Kanten und Gebrauchsspuren kennt die KI nicht. Jedoch sind es gerade diese materiellen Spuren und Eigenarten, die aus profanen Dingen persönliche Gegenstände machen und

mehr Bedeutung für das Individuum haben als ein vorgegebener Zweck. Mit KI, so die Befürchtung, wird die körperlich-materielle Erfahrung der Welt und des eigenen Selbst von einer sterilen, austauschbaren und geglätteten Inszenierung eines „Quasi-Selbst“ abgelöst. Damit geht auch das materielle Erbe eines Menschen verloren: Persönliche Gegenstände, die Geschichten erzählen, verschwinden zugunsten von Datenprofilen, die sich aus beliebigen Punkten zusammensetzen (Unlimited World 2017).

Identität zeigt eine soziale Seite insbesondere dort, wo sie durch Interaktionen mit anderen Menschen entsteht. KI ist bereits heute sowohl Interaktionspartner als auch -filter. Dienstleistungen wie Spotify, Facebook oder Amazon wissen bereits mehr über persönliche Präferenzen als die engsten Freunde. Algorithmen erhärten Kaufentscheidungen und Geschmacksnuancen. Sie determinieren so die Präferenzen ihrer Nutzer. Nach der Kritik hieran entstanden sogenannte encoding-Algorithmen, die alternative Vorschläge machen sollen, um die Souveränität des Nutzers zu stärken. Nur lösen sie das Problem nicht, eher im Gegenteil: Sie wirken noch manipulativer, weil sie eine scheinbar objektivierte Präferenzfindung suggerieren. Das Ergebnis ist aber das Gleiche: Die Souveränität der Nutzer wird gewollt – oder auch nicht – drastisch reduziert. Ähnlich wirken Siri und Alexa, die sehr gefügig und wenig kontrovers einen emotionslosen Austausch simulieren. Kommunikation verliert hier ihre Ambivalenz, mit der wir durch soziale Erfahrungen umzugehen wissen und durch die wir herausfinden, wie wir auf andere wirken und wer wir für die andere Person sind. Aus ethischer Perspektive ist diese Art der Kommunikation mit KI, die eine vermeintlich partnerschaftliche soziale Interaktion simuliert, hintergründig aber strategische Zwecke verfolgt, hochgradig bedenklich.

Ein sehr eindrückliches Beispiel ist hierfür die „Hello-Barbie-Puppe“ von Mattel (NYT 2015). Diese sprechende Puppe verwickelt Kinder mit Hilfe von Spracherkennung und KI in ein Gespräch, das dem einer realen Freundschaft nachempfunden ist. Wenn Kinder der Puppe eigene Gefühle anvertrauen und eine enge Bindung eingehen, verlieren sie ihre Intimsphäre, da die Puppe diese Informationen an Dritte weitergibt, etwa die Eltern. Das Spielzeug war deshalb Gegenstand einer kontroversen Diskussion in amerikanischen Medien. Jedoch ist das Prinzip das gleiche wie auch bei Online-Kaufvorschlägen: Die soziale Identitätsbildung wird durch KI für strategische Interessen ausgenutzt. Dies betrifft auch vermeintlich positive Anwendungen von KI, wenn z. B. Algorithmen genutzt werden, um über WhatsApp-Chatverläufe depressive Züge von Kindern und Jugendlichen zu diagnostizieren (RP Online 2017). Während KI hier eine gesunde Identitätsbildung unterstützen soll und gleichzeitig dem gesellschaftlichen Wert – nämlich dem Schutz von Minderjährigen – Rechnung trägt, wird auch der Zweifel genährt, wie weit Schutz gehen darf und wann Persönlichkeitsrechte verletzt werden.

Die Debatte um den manipulativen Charakter von KI ist jedoch kein Urteil über die Technologie als solche. Vielmehr wirft KI Fragen nach Persönlichkeitsrechten und informationeller Selbstbestimmung auf. Sie setzt keine Normen um, die nicht sowieso bereits in unserer Gesellschaft existieren. Jedoch ist ihr verstärkender Charakter ein relevanter Anlass, um persönliche Identität und wie sie in unserer Gesellschaft durch Technologie entsteht und sich entfaltet neu zu diskutieren.

Fazit und Ausblick

KI besitzt wie keine andere Technologie zuvor höchst vielfältige gesellschaftliche und ethische Bezüge. Diese ergeben sich vor allem aus der Präsenz in den uns bekannten Teilsystemen – von Wertschöpfung über politische Meinungsbildung bis hinein in die Freizeit. In all diesen Bereichen bleibt KI nicht neutral und hinterlässt Spuren.

Wirft KI daher neue ethische Fragen auf? Nein. Die wesentlichen Themen der Debatte sind nicht grundsätzlich neue ethische Themen. Mit den Fragen nach Verantwortung, dem sozialen Miteinander oder der individuellen Persönlichkeitsentfaltung setzt sich der Kerndiskurs der Ethik fort. Das bedeutet jedoch nicht, dass KI die Natur dieser Themen und den Umgang mit ihnen nicht auch verändert, nämlich durch zwei Faktoren:

Erstens erzeugt KI zwar keine grundsätzlich neuen ethischen Fragen, sie verleiht ihnen jedoch neues Gewicht und trägt sie in neue Bereiche hinein. So wurden z. B. Autonomie und Kontrolle bislang nicht im Kontext von Verkehr und Mobilität diskutiert, ebenso wenig führte der Abschluss eines Abonnements bislang zu der Frage, ob sich damit die materielle Identität des Menschen verflüchtigt. Alte ethische Fragen werden durch die KI also mit völlig neuen An- und Herausforderungen verknüpft.

Zweitens und vor allen Dingen potenziert KI ethische Fragestellungen. Das verdeutlicht schon die schiere Menge an ethischen Kommentaren sowohl in wissenschaftlichen und institutionellen als auch populären Medien. In diesem Prozess eilt die ethische Debatte der technischen Realität oftmals voraus, und so können ethische Implikationen von KI zum Gegenstand aktiver politischer und gesellschaftlicher Gestaltung werden.

Die Debatte zur Ethik einer KI ist also bei Weitem noch nicht erschöpft – nicht in qualitativer und erst recht nicht in quantitativer Hinsicht. KI wird den Menschen künftig immer wieder mit seinen nur allzu oft als selbstverständlich erachteten Wertestrukturen konfrontieren und ihn veranlassen, sie zu hinterfragen. Dabei geht es eben nicht darum, die technologischen Potenziale der KI zu beschränken, sondern ihre potenziellen negativen Auswirkungen auf den Menschen zu reflektieren. Das aktuell häufig genannte Argument, KI könne sehr viele, wenn nicht sogar alle Prob-

leme irgendwie lösen, kann sich mit Blick auf die möglichen negativen Aus- und Einwirkungen auf den Menschen folglich langfristig nicht halten.

Stattdessen bedarf es einer auf dem ethischen Diskurs aufbauenden proaktiven Gestaltung der ethischen und gesellschaftlichen Implikationen von KI, die den aktuellen technischen in den langfristig erforderlichen ethischen Imperativ einer KI überführt. Gerade mit Blick auf eine innovationspolitisch orientierte Gestaltung gilt es dabei folgende Elemente einzubeziehen:

Antizipation

KI ist eine Schlüsseltechnologie, die technische Innovationen potenziert, und das in unterschiedlichen Gesellschaftsbereichen. Durch strategische und methodisch angeleitete Antizipation (Foresight) kann eine entsprechende Governance technischen Realitäten vorgreifen und ethische und soziale Faktoren proaktiv berücksichtigen.

Reflexion

KI hat relevante Auswirkungen auf die Gesellschaft – und diese reflektiert dies immer kritischer. Sie hält sich den Spiegel vor, indem etwa Medien die Frage aufwerfen, ob KI mit bestehenden Werten und Normen vereinbar ist. Ähnlich sollte sich auch die Governance der KI-Forschung den Spiegel vorhalten und regelmäßig fragen, welche ethischen und gesellschaftlichen Folgen KI-Förderung impliziert und wie ihre Governance gestaltet werden sollte. Ziel sollte es sein, eine gleiche Ausprägung relevanter Normen und Werte zu wahren.

Transparenz

Die Mechanismen der KI bleiben für zu viele Außenstehende allzu oft verdeckt. Daher brauchen Nutzer mehr Souveränität, die es ihnen ermöglicht, die Funktionsweise der KI zu verstehen und Handlungsmuster nachzuvollziehen. Dies betrifft auch den Umgang mit und die Nutzung von Daten. Und Gleiches sollte auch für die Governance auf systemischer Ebene gelten. Hier ist mehr Kontroverse und Partizipation im Vorfeld notwendig als die bloße Konfrontation mit den neuen technologischen Möglichkeiten und eine anschließende Behebung der Probleme. Zudem können so die Akzeptanz gestärkt und ein möglicher Durchsatz von KI letztlich ebenfalls gesichert werden.

Adaptabilität

KI findet nur dann breite Verwendung und gesellschaftliche Akzeptanz, wenn sie anpassungsfähig gegenüber dem Menschen ist. Den Menschen einfach mit den vorgegebenen technologischen Anforderungen der Systeme zu konfrontieren, reicht nicht aus. Dass Deutschland sich so schwer tut bei der Digitalisierung macht nur allzu

deutlich, dass es eben nicht genügt, entsprechende Kompetenzen nur auszubilden. Es gilt: „structure follows user“ – und nicht umgekehrt. Was schon vielerorts in der Digitalisierung nicht zum Ziele führte, sollte bei einem wachsenden Einsatz von KI erst recht vermieden werden. Ein intelligentes System sollte eben gerade durch seine Intelligenz in der Lage sein, sich selbstständig auf Situationen oder Menschen einzustellen und entsprechend zu interagieren. Wie der Mensch Adaptabilität entfalten kann, sollte es ein KI-System auch leisten müssen. Insofern muss es sich stetig weiterentwickeln.

Literatur

- AI Business (2017): Europe is Looking Into Developing an AI-on-Demand Platform. Online verfügbar unter <https://aibusiness.com/europe-is-looking-into-developing-an-ai-on-demand-platform/>, zuletzt geprüft am 22.06.2018.
- AINOW-Institut (2018): about us. Online verfügbar unter <https://ainowinstitute.org/>, zuletzt geprüft am 22.06.2018.
- Arkin, R. (2009): *Governing Lethal Behavior in Autonomous Robots*, Chapman and Hall/CRC
- BBC (2015): Tech giants pledge \$1bn for 'altruistic AI' venture, OpenAI. Online verfügbar unter <http://www.bbc.com/news/technology-35082344>, zuletzt geprüft am 22.06.2018.
- Bundesregierung (2018): Koalitionsvertrag vom 14. März 2018. Online verfügbar unter <https://www.bundesregierung.de/Content/DE/StatischeSeiten/Breg/koalitionsvertrag-inhaltsverzeichnis.html>, zuletzt geprüft am 22.06.2018.
- Commission nationale Informatique & Libertés [CNIL] (2018): Comment Permettre à L'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Online verfügbar unter https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf, zuletzt geprüft am 22.06.2018.
- Computer Business Review BRJ (2017): OpenAI and Google DeepMind develop machine learning algorithm for a safer AI. Online verfügbar unter <https://www.cbronline.com/news/internet-of-things/cognitive-computing/openai-and-google-deepmind-develop-machine-learning-algorithm-for-a-safer-ai/>, zuletzt geprüft am 22.06.2018.
- Dennett, D. C. (1987): *The Intentional Stance*, Cambridge MA.
- Deutscher Ethikrat (2017): *Autonome Systeme: Wie intelligente Maschinen uns verändern*, Jahrestagung am 21. Juni 2017. Online verfügbar unter <http://www.ethikrat.org/veranstaltungen/jahrestagungen/autonome-systeme>, zuletzt geprüft am 22.06.2018.
- Die Zeit (2016a): Ich kann nicht mehr erkennen, wer Mensch und wer Maschine ist, Interview mit Christoph Gerlach. Online verfügbar unter <http://www.zeit.de/sport/2016-03/go-alphago-lee-sedol-google-kuenstliche-intelligenz>, zuletzt geprüft am 22.06.2018.
- Die Zeit (2016b): Selbst ist die Maschine. Online verfügbar unter <http://www.zeit.de/digital/internet/2016-05/kuenstliche-intelligenz-open-source-amazon-google-facebook>, zuletzt geprüft am 22.06.2018.
- Die Zeit (2017): Der Todesalgorithmus. Online verfügbar unter <http://www.zeit.de/kultur/2017-09/kuenstliche-intelligenz-algorithmus-spam-autonomes-fahren/komplettansicht>, zuletzt geprüft am 22.06.2018.
- Ethik-Kommission Automatisiertes und vernetztes Fahren des BMVI [BMVI] (2016): Bericht. Juni 2016. Online verfügbar unter https://www.bmvi.de/SharedDocs/DE/Publikationen/G/bericht-der-ethik-kommission.pdf?__blob=publicationFile, zuletzt geprüft am 22.06.2018.

- Ethik-Kommission Automatisiertes und Vernetztes Fahren (2017): Bericht Juni 2017. Online verfügbar unter <https://www.bmvi.de/SharedDocs/DE/Publikationen/G/bericht-der-ethik-kommission.html>, zuletzt geprüft am 22.06.2018.
- Europäisches Parlament (2017): Robotik und künstliche Intelligenz: Abgeordnete für EU-weite Haftungsregelungen. Online verfügbar unter <http://www.europarl.europa.eu/news/de/press-room/20170210IPR61808/robotik-und-kunstliche-intelligenz-abgeordnete-fur-eu-weite-haftungsregelungen>, zuletzt geprüft am 22.06.2018.
- European Economic and Social Committee [EESC] (2017): Artificial intelligence. Online verfügbar unter <http://www.eesc.europa.eu/our-work/opinions-information-reports/opinions/artificial-intelligence>, zuletzt geprüft am 22.06.2018.
- Floridi, L.; Sanders, (2004): On the Morality of Artificial Agents. In: *Minds and Machines* 14/2004. S. 349–379.
- Frankfurter Allgemeine Zeitung [FAZ] (2017): Elon Musk warnt vor 3. Weltkrieg durch Künstliche Intelligenz. Online verfügbar unter <http://www.faz.net/aktuell/wirtschaft/kuenstliche-intelligenz/elon-musk-tesla-chef-warnt-vor-kuenstlicher-intelligenz-15182958.html>, zuletzt geprüft am 22.06.2018.
- Handelsblatt (2014): Physiker warnt vor künstlicher Intelligenz. Online verfügbar unter <http://www.handelsblatt.com/technik/forschung-innovation/stephen-hawking-physiker-warnt-vor-kuenstlicher-intelligenz/11067072.html?ticket=ST-1923859-cYaBmAqjzKgfbUY5tFE-ap1>, zuletzt geprüft am 22.06.2018.
- Heise (2017): Ethik bei autonomen Autos und das Trolley-Problem: Was tut der Weichensteller? Online verfügbar unter <https://www.heise.de/newsticker/meldung/Ethik-bei-autonomen-Autos-und-das-Trolley-Problem-Was-tut-der-Weichensteller-3766885.html>, zuletzt geprüft am 08.03.2018.
- Kleinberg, J.; Lakkaraju, H.; Leskovec, J.; Ludwig, J.; Mullainathan, S. (2017): Human Decisions and Machine Predictions, NBER Working Paper No. 23180 Issued in February 2017. Online verfügbar unter <http://nber.org/papers/w23180>, zuletzt geprüft am 22.06.2018.
- MIT Technology Review (2016a): Facebook's New Map of World Population Could Help Get Billions Online. Online verfügbar unter <https://www.technologyreview.com/s/600852/facebook-s-new-map-of-world-population-could-help-get-billions-online/>, zuletzt geprüft am 22.06.2018.
- MIT Technology Review (2016b): Can Machine Learning Help Lift China's Smog? Online verfügbar unter <https://www.technologyreview.com/s/600993/can-machine-learning-help-lift-chinas-smog/>, zuletzt geprüft am 22.06.2018.
- Netzpolitik (2018): New York City plant Arbeitsgruppe zur Überprüfung von städtischen Algorithmen. Online verfügbar unter <https://netzpolitik.org/2018/new-york-city-plant-arbeitsgruppe-zur-ueberpruefung-von-staedtischen-algorithmen/>, zuletzt geprüft am 22.06.2018.

- New York Times [NYT] (2018): Barbie Wants to Get to Know Your Child. Online verfügbar unter <https://www.nytimes.com/2015/09/20/magazine/barbie-wants-to-get-to-know-your-child.html>, zuletzt geprüft am 22.6.2018.
- OpenAI (2017): About OpenAI. Online verfügbar unter <https://openai.com/about/>, zuletzt geprüft am 22.06.2018.
- Partnership on AI (2017): Goals. Online verfügbar unter <https://www.partnershiponai.org/#s-goals>, zuletzt geprüft am 22.06.2018.
- RP Online (2017): Forscher wollen Depressionen bei Whatsapp erkennen. Online verfügbar unter <http://www.rp-online.de/panorama/wissen/forscher-wollen-depressionen-bei-whatsapp-erkennen-aid-1.7263186>, zuletzt geprüft am 22.06.2018.
- Scheutz, M. (2011): Architectural Roles of Affect and How to Evaluate Them in Artificial Agents, in: *International Journal of Synthetic Emotions* 2/2011, S. 48–65.
- Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, Ioannis; Huang, A.; Guez, A.; Hubert, Th.; Baker, L.; Lai, M.; Bolton, A.; Chen, Y.; Lillicrap, T.; Hui, F.; Sifre, L.; van den Driessche, G.; Graepel, Th.; Hassabis, D. (2017): Mastering the Game of Go without Human Knowledge. Online verfügbar unter https://www.nature.com/articles/nature24270.epdf?author_access_token=VjXbVjaSHxFoCtQQ4p2k4tRgN0jAjWel9jnR3ZoTv0PVW4gB-86EEpGqTRDtplz-2rmo8-KG06gqVobU5NSCFeHILHcVFUeMsbvws-lxjqQGg98faovwjxe-TUgZAUMnRQ, zuletzt geprüft am 22.06.2018.
- Süddeutsche Zeitung (2018): Mehr Schutz vor Algorithmen. Online verfügbar unter <http://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-mehr-schutz-vor-algorithmen-1.3856354>, zuletzt geprüft am 22.06.2018.
- Süddeutsche Zeitung (2017a): Computer spielt Go gegen sich selbst - und wird unschlagbar. Online verfügbar unter <http://www.sueddeutsche.de/digital/kuenstliche-intelligenz-champion-aus-dem-nichts-1.3713570>, zuletzt geprüft am 22.06.2018.
- Süddeutsche Zeitung (2017b): Mit Daten werden Maschinen intelligent – und rassistisch. Online verfügbar unter <http://www.sueddeutsche.de/digital/kuenstliche-intelligenz-mit-vorurteilen-daten-machen-maschinen-intelligent-und-rassistisch-1.3494546>, zuletzt geprüft am 22.06.2018.
- TechCrunch (2017): Discussing the limits of artificial intelligence. Online verfügbar unter <https://techcrunch.com/2017/04/01/discussing-the-limits-of-artificial-intelligence/>, zuletzt geprüft am 22.06.2018.
- The Economist (2016): Artificial intelligence and Go - A game-changing result. Online verfügbar unter <https://www.economist.com/news/science-and-technology/21694883-alphagos-masters-taught-it-game-electrifying-match-shows-what>, zuletzt geprüft am 22.06.2018.
- Turkle, Sherry (1984): *The Second Self: Computers and the Human Spirit*. MIT University Press Group Ltd: Cambridge, Massachusetts.
- United Nations [Un] (2017): UN artificial intelligence summit aims to tackle poverty, humanity's 'grand challenges', UN News. Online verfügbar unter <https://news.un.org/en/>

story/2017/06/558962-un-artificial-intelligence-summit-aims-tackle-poverty-humanity-grand#.WUpXHdKJG70, zuletzt geprüft am 22.06.2018.

Unlimited World (2017): Digital legacy: Identity multiplied, The Evolution of The Self. Online verfügbar unter <https://www.unlimited.world/ubs/digital-legacy-identity-multiplied>, zuletzt geprüft am 22.06.2018.



Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz <http://creativecommons.org/licenses/by/4.0/deed.de> veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.