Ursula Nothelle-Wildfeuer (Freiburg), Elisabeth André (Augsburg), Noreen van Elk (Berlin)

Zum Umgang mit KI-Technologien – ethische Prinzipien und Leitlinien aus christlicher Perspektive

Sowohl im Berufs- als auch im Alltagsleben kommen wir – teils ohne uns dessen bewusst zu sein – ständig mit Techniken der Künstlichen Intelligenz (KI) in Berührung. Zu den prominenten Beispielen gehören Sprachassistentinnen wie Alexa und Siri, die Auskunft zum Wetter geben, an wichtige Termine erinnern oder auf Sprachkommando die Heizung oder das Licht regeln. In der diagnostischen Medizin unterstützen KI-Techniken Ärzte und Ärztinnen bei zentralen gesundheitskritischen Entscheidungen. In der industriellen Fertigung entlasten Roboter den Menschen bei der Durchführung von Montagearbeiten. Bisherige Industrieroboter kamen zum Einsatz, um den Menschen durch Automatisierung lästige Arbeitsroutinen industrieller Prozesse abzunehmen, dafür agierten sie aus Gründen der Sicherheit weitgehend abgeschottet von Menschen in Industriehallen. Im Unterschied dazu agiert der moderne Serviceroboter vor allem außerhalb solcher Hallen (Vgl. Kehl 2018, S. 22), sowohl in Privathaushalten als auch in komplexeren Kontexten. So erfordern pflegerische Hebemaschinen, die bei der Umbettung von Patienten helfen, nicht nur räumliche Nähe, sondern sogar direkten körperlichen Kontakt mit den Menschen.

Das besondere Merkmal der genannten KI-Anwendungen besteht darin, dass sie in vorgegebenen Bereichen komplexe Aufgaben selbstständig durchführen, die nicht mehr vom Menschen vorab im Detail spezifiziert werden müssen. In stark abgegrenzten Bereichen – etwa bei der Auswertung von medizinischen Bildern – zeigen Maschinen bereits heute quasi-intellektuelle Fähigkeiten, die dem nahekommen oder sogar darüber hinausgehen, was ein Mensch vermag. Erfindungen werden genau dafür gemacht, dass durch sie den Menschen das Leben erleichtert wird (vgl. Grunwald 2019b, S. 33). Ein weiterer Schritt zur Verbesserung der quasi-intellektuellen Fähigkeiten von Maschinen wird durch den Einsatz von Methoden des maschinellen Lernens ermöglicht, mit denen Maschinen befähigt werden sollen, „nicht nur ein festgelegtes Handlungsprogramm abzuspulen, sondern sich auch unter neuen oder sich verändernden Bedingungen weitgehend autonom zurechtzufinden.“ (Kehl 2018, S. 23) In dieser Entwicklung liegt das Potenzial zu einer Neubestimmung des Mensch-Technik-Verhältnisses. Spezifisch ist hierbei, dass Technik in diesen Systemen autonom agieren können soll, letztlich also nicht mehr den gleichen Status wie ein „herkömmliche(s) Werkzeug()“ (Kehl 2018, S. 22)

hat. Das ist in einer Analyse aus ethischer Perspektive als entscheidender Faktor mit zu berücksichtigen.

Die nachfolgenden Überlegungen wollen - nach einer prinzipiellen Vergewisserung des christlichen Beitrags zu diesen Fragen – sieben ethische Leitlinien entwickeln für den Umgang mit Digitalisierungstechniken und Systemen Künstlicher Intelligenz.

1. Vorbemerkung: Christlich-ethischer Beitrag

In einer weitgehend säkularisierten und pluralisierten Gesellschaft wie der unsrigen kommt der Kirche und der theologischen Ethik keine selbstverständliche und unhinterfragte Autorität für Fragen von Gesellschaft und Wissenschaft mehr zu. Ein christlich-ethischer Beitrag muss darum argumentativ nachvollziehbar und anschlussfähig für einen interdisziplinären Austausch seine Aspekte einbringen. Der inhaltliche Ausgangspunkt lässt sich folgendermaßen beschreiben: Der weitgehende Konsens, dass Menschen mit Menschenwürde ausgestattete Individuen sind, deren Menschenrechte sowie Freiheits- und Gerechtigkeitsstreben unbedingt zu respektieren sind, bildet die Basis unseres gesellschaftlichen Zusammenlebens. Christliche Ethik mit ihrem Verständnis vom Menschen als Geschöpf und Ebenbild Gottes, dem die Erde zur Gestaltung und Nutzung von seinem Schöpfer anvertraut ist, hat zum Zustandekommen dieses Verständnisses konstitutiv beigetragen - in und trotz einer langen und nicht unkomplizierten Geschichte der Verhältnisbestimmung von moderner Welt und christlichem Glauben. Klar ist, dass KI-Technologie unser gesellschaftliches Zusammenleben nachhaltig verändern wird. Auch verändert sie das Mensch-Technik-Verhältnis, und hat womöglich Folgen auch für das grundlegende Verständnis von Mensch-Sein. Solche und weitere tiefgreifende Veränderungen machen anthropologische und (sozial)ethische Überlegungen unabdingbar. Die Digitalität stellt uns vor die Frage, was zukünftig unser Mensch-Sein und Zusammenleben ausmachen sollte. Wie wollen wir zusammenleben? Was soll unsere Existenz prägen? Nach welchen Maßstäben und auf welche Art und Weise wollen wir den Wandel mitgestalten? Und: In welcher Gesellschaft wollen wir leben? Vor diesem Hintergrund hat die christliche Ethik, die sich heute weithin als menschenrechtlich fundiert und orientiert versteht, sowie auch die Kirche als gemeinwohlorientierte Institution, zu den anstehenden Fragen der Digitalisierung und KI einen unverzichtbaren Beitrag zu leisten und betrachtet es als ihre Aufgabe den gesellschaftlichen Wandel kritisch zu begleiten und mitzugestalten.

2. Ethische Prinzipien und Leitsätze

Die folgenden ethischen Prinzipien und Leitsätze reflektieren einerseits ein bestimmtes, nämlich das christliche Menschenbild, und andererseits die daraus folgenden ethischen Grundwerte, die es im Umgang mit Digitalisierung und künstlicher Intelligenz im Blick zu behalten gilt.

2.1 Autonomie

Der Mensch muss Autor bzw. Autorin des eigenen Lebens bleiben! (vgl. Nida-Rümelin und Weidenfeld 2018, S. 49)

Die Stufe 5 des automatisierten und vernetzten Fahrens, die allerdings in naher und mittlerer Zukunft wohl noch nicht realisiert werden kann) sieht vor, dass es keinen menschlichen Fahrer mehr gibt und auch keine menschlichen Eingriffsmöglichkeiten für den Notfall. Die in diesem Kontext immer wieder angeführte ethische Dilemma-Situation ist das sog. Trolley-Szenario, in dem (in Analogie zum Weichensteller-Problem) das Auto nur die „Wahl" hat, auf seiner eingeschlagenen Route fünf ältere Männer zu töten oder auf einer anderen Spur eine junge Frau mit Baby zu überfahren. Wie soll sich das autonome Auto entscheiden? Darüber hinaus: Wo liegen die ethischen Grenzen der Automatisierung?

KI-Technologie ist nicht aus sich heraus einfach gut oder schlecht, sondern es kommt darauf an, was der Mensch daraus für sich, die Gesellschaft, die Wirtschaft und die Politik macht. Im Mittelpunkt steht dabei die Frage, wie es gelingen kann, dass der Mensch selbst – so formuliert es der Philosoph Julian Nida-Rümelin mit dem Begriff des „digitalen Humanismus" – „Autor oder Autorin des eigenen Lebens" (Nida-Rümelin und Weidenfeld 2018, S. 49), individuell, gesellschaftlich und politisch, bleibt. Der Technikphilosoph Armin Grunwald spricht davon, dass „klar (bleiben muss), wer das Heft in der Hand hält" (Grunwald 2019b, S. 35).

Dem Menschen kommt Freiheit als Gabe, zugleich auch als Verpflichtung und Verantwortung zu, und damit ist untrennbar seine unteilbare und unantastbare Würde verbunden. Der fundamentale Grundsatz christlicher Soziallehre lautet darum auch: „Urheber, Mittelpunkt und Ziel" (GS 63) allen gesellschaftlichen Handelns und aller Institutionen ist der Mensch. Wenn wir nun auf die durch KI-Technologie bedingten gravierenden, viele sagen disruptiven Entwicklungen in unserer Gesellschaft schauen, sind genau diese Prozesse aus der Perspektive der Freiheit und Autonomie, der Würde des Menschen näher in den Blick zu nehmen.

Zwei miteinander zusammenhängende negative Szenarien sind denkbar:

Zum einen das Zauberlehrling-Szenario: Wird aufgrund von immer weiteren und besseren Feedbackschleifen ein superintelligentes System entstehen, das letztlich vom Menschen zwar initiiert, aber von ihm selbst nicht mehr kontrollierbar ist – vergleichbar der Situation des Goetheschen Zauberlehrlings? Nutzt also der Mensch seine Intelligenz und Autonomie, um ein System zu entwickeln, das ihn dann letztlich aber genau dieser Freiheit und Autonomie beraubt und sich gegen ihn wendet? Übereinstimmend formulieren Forscher unterschiedlicher Provenienz beruhigend, dass die Entwicklung solcher Technologie derzeit und auch in absehbarer Zukunft keine realistische Aussicht ist. Aber dennoch ist grundsätzlich und gerade auch aus christlich-ethischem Zugang in Erinnerung zu rufen, dass jede Digitalisierung und Weiterentwicklung von KI unter dem Vorzeichen eines digitalen Humanismus (Nida-Rümelin) stehen muss, demzufolge der Mensch nicht einer zum Selbstzweck werdenden Entwicklung geopfert werden darf, sondern diese Entwicklung sich am Kriterium der Lebens- und Freiheitsdienlichkeit zu orientieren hat.

Zum anderen eine antihumanistische Utopie: Sie gibt zunächst vor, das Humane zu verbessern, will es dann aber überwinden. Dahinter steht ein ersatz-religiöser Gedanke einer Erlösung, deren Schöpfer und Akteur nicht mehr Gott ist, sondern Softwareingenieure. (Vgl. Nida-Rümelin und Weidenfeld 2018, S. 22)

Das Eintreten eines solchen Szenarios ist jedoch nicht zwangsläufig, und dessen Vermeidung liegt in unserer Hand. KI-Technologie stellt nicht einfachhin eine Entwicklung oder ein Schicksal dar, dem sich der Mensch ergeben muss. Vielmehr handelt es sich um eine Entwicklung, die der Mensch nicht nur nicht mehr rückgängig machen kann, sondern die er gestalten und sich in ihren positiven Effekten nutzbar machen kann und selbstverständlich auch bereits macht. Die Ausrichtung der KI, das Leben der Menschen und dessen Humanität zu stärken, zu unterstützen und zu verbessern, also lebensdienlich und menschenfreundlich zu sein, muss die entscheidende und auch Grenzen setzende Leitlinie sein, nicht aber, den Menschen zu ersetzen und ihn damit letztlich zu beseitigen.

2.2 Verantwortung

Verantwortung kann nicht an Maschinen delegiert werden.

Der Pflegeroboter in einem Seniorenheim hat bereits bestimmte moralisch hoch relevante Entscheidungen zu fällen, etwa: „Wie häufig und eindringlich soll ein Pflegesystem an Essen und Trinken sowie die Einnahme von Medikamenten erinnern? Wann sollte ein Pflegesystem die Angehörigen verständigen oder den medizinischen Dienst rufen, wenn jemand sich eine

Zeitlang nicht rührt? Soll das System den Nutzer rund um die Uhr überwachen, und wie ist mit den dabei erhobenen Daten zu verfahren?“ (Misselhorn 2018, S. 29)

Die für den Zusammenhang der Ethik grundlegende Frage lautet, ob und inwieweit der Roboter bzw. die KI selbst zum „moral agent“ wird. (Vgl. Misselhorn 2018, S. 29) Diese Frage ist virulent nicht nur beim immer wieder angeführten Trolley-Dilemma, sondern eben auch bereits beim Pflegeroboter. Dabei geht es um die, wie sich zeigen wird, letztlich zu verneinende Frage, ob das, was wir als Aktion solcher Roboter wahrnehmen, überhaupt als Handlung und darüber hinaus im eigentlichen Sinn als moralisch zu bezeichnen wäre (Misselhorn 2018, S. 30).
Im Blick auf das Beispiel des Pflegeroboters gilt es, dessen Möglichkeiten zu hinterfragen, selbst bestimmte moralisch hoch relevante Entscheidungen zu fällen. Wie sind die jeweils in Frage stehenden Werte beispielsweise einer freien Entscheidung des Patienten und der Fürsorgepflicht der Pfleger abzuwägen? Hierbei handelt es sich um eine Abwägung, die nur ein rational argumentierender, die individuellen Umstände jeweils mitberücksichtigender Akteur vornehmen kann. Mit Blick auf die KI ist allerdings dann zu fragen: Wer programmiert den Roboter auf welche Priorität hin und wann soll welches Programm laufen? Wer ist es, der letztlich die Entscheidungen fällt? Smarte Maschinen besitzen weder Freiheit noch Vernunft noch einen Willen, um Werte ausloten und Güter abwägen (und das Ergebnis dementsprechend in Handlungen umsetzen) zu können. Das aber ist für eine ethisch verantwortete, also humane Entscheidung unverzichtbar. „Die bloße Programmierung eines moralischen Codes in einen Roboter ist offensichtlich nichts weiter als die bloße Nachahmung von ‚Moralität‘, von ‚ethischer Reflektion‘ kann erst recht keine Rede sein.“ (Capurro 2017, S. 50) Bei menschlichen Entscheidungen spielen Emotionen und Instinkte immer auch eine Rolle. In der Emotionspsychologie werden Emotionen vereinfacht gesagt als Reaktionen auf die Bewertung von eingetretenen Ereignissen, durchgeführten Aktionen oder Objekten konzeptualisiert (vgl. Ortony et al. 1988). Damit ist vor allem bei Emotionen wie Scham und Verachtung, die auf einem Vergleich von menschlichen Verhaltensweisen mit sozialen Normen beruhen, ein Zusammenhang zu moralischen Erwägungen gegeben (vgl. Tangney et al. 2007). In der KI wurden emotionale Bewertungsmechanismen als Grundlage für die Simulation von emotionalem Verhalten von künstlichen Agenten implementiert (vgl. André 2014). Damit sind KI-gesteuerte künstliche Agenten grundsätzlich in der Lage, von Emotionen gesteuertes moralisches Verhalten zu zeigen. Allerdings handelt es sich hierbei um simulierte Emotionen, denen die körperliche Erfahrung fehlt (vgl. Feichtinger 2017). So kann ein Roboter zwar Schamgefühle - etwa durch Senken des Kopfes - zeigen, allerdings ohne jegliche biologische

Grundlage. Feichtinger (2017) führt dies als weiteres Argument an, warum Maschinen nicht moralfähig sind.

Dort, wo ein digitales Programm ohne ein vernunftgeleitetes Abwägen von relevanten Gütern, ohne eine angemessene Berücksichtigung der individuellen Situation, einmal in Gang gesetzt, automatisch abläuft, kann in einem ethischen Sinn nicht die Rede sein kann von Autonomie, sondern von Automatisierung, nicht von Freiheit, sondern von implementierter Gesetzmäßigkeit. Die Debatte um die Autonomie von KI-Systemen erfordert eine klare Unterscheidung zwischen technischer und menschlicher Autonomie. Der Philosoph Gottschalk-Mazouz (2019) stellt fest, dass technische Systeme zwar ein gewisses Maß an Selbstständigkeit zeigen können, aber nicht selbstbestimmt handeln können. Eine moralische Entscheidung zu fällen impliziert auch zugleich, Verantwortung für sie übernehmen zu können. Genau diese Übernahme von moralischer Verantwortung, das Einstehen für entsprechende Konsequenzen von Seiten des Akteurs kann kein noch so gut programmierter Roboter leisten! Verantwortung – das zentrale Kriterium für Moralität - kann ein solches „autonomes“ System nicht übernehmen. Mithin gibt es keine Verantwortungsdelegation an Maschinen, sondern der Träger der Verantwortung bleibt der Mensch. Als wesentlichen Unterschied von Maschinenethik und der menschlichem Verhalten zugrundeliegenden Ethik stellen Dignum und Kollegen heraus: „Verantwortung muss bei den Menschen bleiben – bei denjenigen, die die Maschine entworfen oder programmiert haben, bei denjenigen, die sie angepasst und zum Einsatz gebracht haben, oder bei denjenigen, die sie verwenden.” (vgl. Dignum et al. 2018, Übersetzung aus dem Englischen) Ungeklärt ist die Frage, wer für eventuelle Schäden von autonom handelnden KI-Systemen aufkommt, wenn dem Betreiber von KI nach dem Verschuldensprinzip keine Fahrlässigkeit nachzuweisen ist (vgl. Webwelt & Technik 2018). Einer einseitigen Verteilung von Gewinnen und Verlusten, bei der die Gesellschaft die alleinigen Risiken trägt, ist (nicht nur beim Einsatz von KI) entgegenzuwirken.

Bei allem Bemühen um stetige Verbesserung des Grades der Automatisierung und der technischen Ausstattung werden auch bleibend ethisch kritische, also Dilemma-Situationen, entstehen können, in denen die KI vor einer notwendigen ethisch-moralischen „Entscheidung“ steht, „eines von zwei nicht abwägungsfähigen Übeln notwendig verwirklichen zu müssen“ (Bundesministerium für Verkehr und digitale Infrastruktur 2017, S. 17). Bekannt ist dies vor allem als Frage im Blick auf das autonome Fahren, aber es spielt auch in anderen KI-Kontexten eine wichtige Rolle. Auch dabei bleibt festzuhalten: Computer fällen keine Entscheidung, sondern sie funktionieren nach von Menschen vorgegebenen Kriterien. „Die Entscheidung über Leben und Tod bleibt indirekt beim Menschen, wandert allerdings von den einzelnen

Autofahrern zu Personen und Institutionen im Hintergrund, zu Firmen, Programmierern, Managern oder einer Regulierungsbehörde.“ (Grunwald 2019a, S. 85) Der Computer spult nur ein einprogrammiertes Procedere ab! Freiheit und Verantwortung bleiben bei den Menschen, die für diese Programmierung zuständig sind.

2.3 Singularität des Humanum

Die Wirklichkeit des Menschen kann und darf nicht auf das reduziert werden, was im binären Code abzubilden ist.

Die Roboterdame Sophia, vom Hongkonger Unternehmen Hanson Robotics entwickelt, zeichnet sich hauptsächlich dadurch aus, dass sie nach Angabe ihrer Hersteller vom Aussehen und Verhalten her besonders humanoid wirkt. Bei unterschiedlichen internationalen Veranstaltungen wurde sie zu existentiellen Fragen des Menschseins befragt, u.a. zu der Frage nach ihrem Glauben, nach dem Leben, nach bestimmten Musikrichtungen, nach Liebe, nach Freiheit und dem Sterben. Bei vielen dieser Themen bleibt sie die Antwort schuldig oder zitiert schlichtweg einen entsprechenden Wikipedia-Eintrag. https://www.youtube.com/watch?time_continue=367&v=T4q0WS0gxRY&feature=emb_title

Es gibt vielfältige Grundzüge menschlichen Lebens, die Roboter allerhöchstens simulieren können, die sie aber nicht selbst haben können: Emotionen, Empathie, Trauer und Freude. Den Tränen der Roboter entspricht keine Trauer, dem Lachen der Roboter entspricht keine Freude. Aber gerade auch Dimensionen jenseits rein technischer, auf den binären Code festgelegten Rationalität machen in besonderer Weise das Humanum aus: Fragen nach dem eigenen Lebenssinn, nach der Bedeutung von Religion, nach Liebe, nach Träumen und Wünschen sind von KI nicht zu beantworten. Darin zeigt sich, dass der Mensch deutlich mehr ist als das, was sich auf die Matrix von 0 und 1 spannen lässt.

Daraus ergeben sich weitere Anfragen an die KI und notwendige Begrenzungen, in denen auch die Singularität des Humanum noch einmal deutlich wird: Ein KI-System basiert auf einprogrammierten Algorithmen und lernt anhand von vorgegebenen Daten. Aus diesem Grund mag ein KI-System zwar beindruckende Ergebnisse auf der Grundlage dieser Algorithmen und Daten hervorbringen. Eine zündende Idee, die eben noch nicht umgesetzt und in einen entsprechenden binären Code gebracht worden ist, wird es auf der Seite eines KI-Systems kaum geben. Obwohl die Durchführung von kreativen Aufgaben durch KI eine gewisse Faszination erzeugt, ist selbst bei noch so beeindruckenden durch KI erzeugten Kunstwerke keine

eigenständige künstlerische Absicht zu erkennen. Ein KI-System mag bei entsprechender Programmierung eines Codes für Gerechtigkeit in der Lage sein, entspreche de Gerechtigkeitslücken zu erkennen – möglicherweise besser als ein Mensch. Es wird jedoch nicht aus eigenem Antrieb das Bedürfnis entwickeln, sich für Gerechtigkeit einzusetzen, die Neujustierung der Ausrichtung auf ein gesellschaftliches Gemeinwohl wird kein KI-System von sich aus anmahnen. Ganz im Gegenteil besteht das Risiko, dass KI-Ungerechtigkeiten und Diskriminierung verstärkt, da Algorithmen auf Datensätzen basieren, die realweltliche Diskriminierung und Ungerechtigkeit abbilden. Um also nicht zum Stillstand oder Rückschritt unserer Welt und Gesellschaft zu kommen, braucht es weiterhin genau diese den Menschen ausmachenden Dimensionen und eine kritische Überprüfung algorithmischer Prozesse. Die Wirklichkeit des Menschen lässt sich mittels eines binären Codes – wenn überhaupt - nur annäherungsweise abbilden.

2.4 Sittliche Plausibilität und Transparenz

Die Anwendung von Algorithmen und künstlicher Intelligenz muss argumentativ und sittlich plausibel gemacht werden.

Immer wieder kommen große Unternehmen (wie etwa Zalando im November 2019) in die Schlagzeilen, weil aufgedeckt wurde, dass sie ihre Mitarbeiter und Mitarbeiterinnen mit entsprechenden Analysesystemen überwachen. Ein Forschungsprojekt der Organisation AlgorithmWatch hat in Kooperation mit der Hans-Böckler-Stiftung deutlich machen können, dass der Einsatz solcher Software oftmals rechtswidrig ist bzw. zumindest in einer rechtlichen Grauzone erfolgt, weil es, wie noch im Detail recherchiert werden muss, nicht den Bestimmungen der DSGVO entspräche. Es gehe dabei um System möglichst umfassender Leistungskontrolle, in dem alle Mitarbeiter kontrolliert, bewertet und auch sanktioniert werden. Zugleich soll damit Leistungsbereitschaft und -fähigkeit der Mitarbeiter quantifiziert und damit steuerbar werden. (Vgl. https://www.heise.de/newsticker/meldung/Zalando-Co-Personalanalyse-mit-Big-Data-in-der-dunkelgrauen-Zone-4673385.html

Algorithmische Verfahren zur Kontrolle von Mitarbeiter*innen oder von Bürger*innen machen es erforderlich, bei ihrem Einsatz das Recht auf Privatheit und Selbstbestimmung der Menschen angemessen zu berücksichtigen, und somit die Freiheit und Integrität der persönlichen Identität zu wahren. Ihr Einsatz muss moralisch und damit auch argumentativ plausibilisiert und gerechtfertigt werden. Dies bedeutet, die Betroffenen über den Einsatz solcher Software zu

informieren und ihre Zustimmung dazu einzuholen. Da, wo es um die prinzipielle Einführung eines solchen Instrumentariums geht, bedarf es auch des Einbezugs der relevanten Mitbestimmungsgremien. KI und ihr Funktionieren muss soweit wie möglich transparent und im Zustandekommen von Konsequenzen zumindest rückverfolgbar sein muss. Andernfalls wird die KI empfunden als eine dunkle Macht, die über den Menschen bestimmt und sich ihn unterwirft. Auch wenn wahrscheinlich aus Gründen der großen Kompliziertheit und der zum wirklichen Verständnis gegebenen Notwendigkeit differenzierter Fachkenntnisse der Algorithmus selbst meistenteils nicht erklärbar sein wird, muss doch insgesamt offenkundig werden, dass es sich bei den Algorithmen um ein Instrumentarium handelt, das zu exakt definierten Zwecken eingesetzt wird und dessen Nutzen verständlich erscheint. Aus solchen Erklärungsbemühungen wächst Vertrauen der Beteiligten in das KI-System, was als Voraussetzung für seine Akzeptanz im Allgemeinen und auch im Fall spezifischer Maßnahmen unabdingbar ist. Dies wiederum leistet einen wesentlichen Beitrag zum Gemeinwohl des Unternehmens, der gesellschaftlichen Gruppierungen oder der Gesamtgesellschaft. Damit bleibt die ethische Leitlinie der Transparenz nicht nur im Blick auf ein Verfahren relevant, sondern bezieht sich zugleich auf einen wesentlichen material-inhaltlichen Aspekt der Gesellschaft.

Zugleich impliziert das Leitprinzip der Transparenz auch den Hinweis auf die Notwendigkeit des Datenschutzes, darauf also, dass unsere persönlichen und privaten Daten geschützt bleiben müssen und nur nach individueller und persönlicher Entscheidung veröffentlicht werden dürfen. Das bedeutet, dass jedem und jeder Einzelnen nachvollziehbar bleiben muss, was mit den jeweils eigenen Daten passiert. Die Souveränität muss beim Einzelnen verbleiben.

Anzumerken ist, dass sich bisherige Anstrengungen vor allem darauf konzentrieren, die Transparenz von KI-Systemen durch entsprechende Erklärungsmechanismen zu erhöhen (Adadi & Berada 2018). So werden üblicherweise die Eigenschaften etwa von Personen hervorgehoben, die maßgeblich zu einer Eingruppierung durch das System beigetragen haben. Damit sind die Entscheidungsprozesse zwar möglicherweise transparent, aber nicht unbedingt sittlich plausibel. Tubella und Kollegen (2019) fordern, auch die moralischen Grenzen, innerhalb dessen sich ein KI-System bewegt, offenzulegen.

2.5 Nichtschadensprinzip und Nicht-Diskriminierung

KI-Technologie darf keinen Schaden für Mensch und Gesellschaft, aber auch für die Umwelt, verursachen.

„Der Google-Algorithmus benachteiligt Frauen“, so lautet eine Meldung im August 2019. Das liegt nicht primär am Geschlecht, sondern an der deutschen Sprache, die bei der männlichen Personenbezeichnung die weibliche Form inkludiert, nicht aber umgekehrt. Das führt dazu, dass Suchmaschinen Frauen, die sich in ihrem eigenen Unternehmen selbstverständlich als Designerin, als Beraterin o.ä. bezeichnen, nicht finden. Um diesen eindeutigen Nachteil ausgleichen zu können, bedarf es einiger Extraschritte in der Suchmaschinenoptimierung (die dann u.a. doch die männliche Form im Text einfließen lassen).
https://www.golem.de/news/seo-der-google-algorithmus-ist-frauenfeindlich-1908-142988.html

Diese ethische Leitlinie hängt mit der oben aufgezeigten Perspektive der Menschenwürde untrennbar zusammen: Algorithmen sind nicht als solche gut oder schlecht, es gibt positive und negative Effekte für das Handeln des Menschen in unterschiedlichen Bereichen. Diese Effekte hängen damit zusammen, zu welchem Zweck und in welchem Modus die Menschen solche Algorithmen einsetzen.

Das ethische Leitprinzip des Nicht-Schadens geht davon aus, dass Schaden nie ganz vermeidbar ist, so sehr man es auch versucht. Man weiß darum, dass man immer schadet, und bemüht sich in diesem Wissen darum, den Schaden so gering wie möglich zu halten. Vor diesem Hintergrund ist der formulierte Grundsatz der Schadensvermeidung durch Algorithmen und KI für den Menschen als Einzelperson zu lesen. Niemand darf benachteiligt, diskriminiert oder ausgeschlossen werden, wie es das aufgezeigte Beispiel demonstriert. Algorithmen dürfen nicht benutzt werden, ein in der analogen Welt bereits bestehendes Machtgefälle noch zu verstärken. Zugleich ist das Prinzip der Schadensvermeidung auch über das Individuum hinaus anzuwenden: auch im Blick auf das Gemeinwohl einer Gesellschaft muss der Schaden so weit wie nur möglich ferngehalten werden.

An dieser Stelle ist auf zwei schädliche Phänomene einzugehen, die für algorithmenbasierte KI-Technologien spezifisch sind: Erstens ist hier die Möglichkeit, die durch *Deep Fake* gegeben ist, zu nennen. Lügen sind nichts Neues, auch die Rede von *Fake News* bezieht sich letztlich nur auf eine Systematisierung solcher bewusst eingesetzter und instrumentalisierter Lügen bzw. die Verunglimpfung unliebsamer Wahrheit bzw. Berichterstattung als Fake News. Nicht nur, aber auch im Blick auf unsere Demokratie bringt dies eine mögliche Beeinträchtigung der Wahl- und Entscheidungsfreiheit der Menschen mit sich. Wohl aber neu ist die Deep fake Technologie, die die Möglichkeit bietet, Audios oder Videos von real existierenden Menschen zu produzieren, in denen sie Dinge sagen oder tun, die sie niemals

sagen oder tun würden. Machine learning Technologien machen es auch zunehmend schwieriger oder sogar unmöglich, solche Deep Fakes zu aufzudecken. Dass diese Technologie spezifisches kriminelles Potential und ggf. die Gefahr einer großen Schädigung für die Gesellschaft, aber auch für einzelne Menschen mit sich bringt, liegt auf der Hand.

Zweitens bergen algorithmenbasierte Technologien ein erhebliches Diskriminierungs- und Schädigungspotenzial in sich, da sie mit Daten gefüttert werden, die von uns Menschen erzeugt werden. Dieses Schädigungspotenzial hat die Datenethikkommission zuletzt dazu veranlasst, die Anforderungen, die konkret an algorithmische Systeme gestellt werden sollten, von ihrer sogenannter „Kritikalität“ abhängig zu machen und Algorithmische Systeme in einer „Kritikalitätspyramide“ einzustufen, um so das Schädigungs- und Diskriminierungspotenzial einzudämmen. (Vgl. Abschlussbericht der Datenethikkommission)

Andererseits bergen KI-Technologien das Potential, Diskriminierung aufzudecken, die von Maschinen, aber auch durch Menschen verursacht wurde (vgl. McKinsey 2019). So hat das diskriminierende Verhalten von KI-Systemen eine lebhafte Debatte über diskriminierendes menschliches Verhalten, dass die verwendeten Daten reflektieren, initiiert. Darüber hinaus wurden erste Fortschritte erzielt, um datenbedingten Verzerrungen mit Hilfe von KI-Ansätzen entgegenzuwirken – etwa mit Methoden zur Anreicherung von Datensätzen mit künstlich generierten Daten. Solche Methoden könnten Bestandteil einer Zertifizierung von KI-Systemen sein, um einige der oben genannten Risiken abzumildern.

Die Menschen, die KI-Technologie nutzen, müssen im Blick auf das Leben der Menschen und die Entwicklung der Gesellschaft immer wieder versuchen, Schaden fernzuhalten oder abzuwenden. Wir haben die Gestaltung einer Gesellschaft der Digitalität in der Hand und müssen permanent eine Neujustierung vornehmen.

2.6 Fürsorge und Wohlwollen

Smarte Maschinen und Algorithmen müssen Fürsorge als Ausdruck von echtem Wohlwollen realisieren.

Der Roboter Pepper ist eine humanoide smarte Maschine, entwickelt in Kooperation des französischen Unternehmens Aldebaran Robotics SAS mit dem japanischen Telekommunikations- und Medienkonzern SoftBank Mobile Corp. Er wird zur Kundenbetreuung eingesetzt, ist aber langfristig gedacht als Begleiter im Bereich der Pflege und Betreuung alter Menschen. Kennzeichnend für ihn ist sein Aussehen, das das Kindchen-Schema aufgreift. Darüber hinaus kann er speziell Mimik und Gestik analysieren und auf Emotionen reagieren. Konzipiert wurde er als „Roboter-Gefährte“ (companion robot) und als

„persönlicher Roboter“ (personal robot), er soll Unterhaltung, Spaß und Abwechslung bringen. https://www.youtube.com/watch?v=aZ5VkgvQFBU oder https://www.youtube.com/watch?v=UMWkSQz3aOo

Die Leitlinie der Fürsorge richtet den Blick insbesondere auf die Dimension des Wohlergehens jedes einzelnen Menschen und auf das Gemeinwohl der Gesellschaft. Dabei ist entscheidend, dass diese Fürsorge als Ausdruck echten Wohlwollens, genuin christlich gesprochen: als Ausdruck der Liebe, gestaltet wird und es sich nicht eine reine Simulation darstellt.

Inhaltlich geht es dabei zunächst in unserer zunehmend alternden Gesellschaft um die alten und pflegebedürftigen Menschen, zu deren Wohlergehen unter dem Vorzeichen des Pflegekräftemangels und zur Entlastung des Personals Roboter vermehrt eingesetzt werden. Bereits die vielfältigen Einsatzmöglichkeiten und -notwendigkeiten für Pflegeroboter zeigen, dass und wo ethische Fragen anstehen, zumal in diesem so hoch umstrittenen und in weiten Teilen auch angstbesetzten Bereich. Wollen Pflegebedürftige lieber von einem Roboter bei der Körperpflege unterstützt werden, weil es ihnen dann nicht so unangenehm ist wie bei einem Menschen? Möchte ein Mensch, der Angst hat oder depressiv ist, lieber die Hand eines mitfühlenden Menschen spüren oder reicht ein Roboter? In all diesen Kontexten muss die Leitlinie der Fürsorge im Vordergrund stehen. Die Roboter dürfen nicht als prinzipieller Ersatz für die Menschen in der Pflege eingesetzt werden, sondern in Ergänzung und zu deren Unterstützung.

Speziell im Blick auf den Aspekt der Vermeidung von Einsamkeit, Isolation und Depression erhält die Nutzung von KI und Algorithmen eine besondere Bedeutung im Sinne der genannten Leitlinie der Fürsorge.

2.7 Fairness, Teilhabe und Soziale Gerechtigkeit

Die Anwendung von Digitalisierung und KI-Algorithmen muss auf alle Fälle die Faktoren Partizipation und Fairness berücksichtigen und in entsprechende sozialstaatliche Rahmenbedingungen einbetten.

Eine erste wichtige wissenschaftliche Studie, die nach den maßgeblichen Forschern der Oxford University so benannte Frey-Osborne-Studie aus dem Jahr 2013, untersuchte auf den US-amerikanischen Markt bezogen das Automatisierungspotenzial der amerikanischen Berufe. Dieser Untersuchung zufolge würden „47 Prozent der amerikanischen Beschäftigten Berufe ausüben, die mit hoher Wahrscheinlichkeit automatisiert würden (Frey & Osborne 2013)“ (Lorenz 2017, S. 7). Vor allen Dingen seien „niedrig qualifizierte und niedrig entlohnt

Beschäftigte am deutlichsten von der Automatisierung betroffen“ (Lorenz 2017, S. 7); zukunftssicher seien dagegen am ehesten Tätigkeiten mit hohem kreativen und unternehmerischen Anteil. (Vgl. Rinne und Zimmermann 2016, S. 6) Auch wenn die Studie methodisch vielfältig kritisiert und durch weitere Studien deutlich differenziertere Ergebnisse erarbeitet werden konnten, wurde dadurch ein notwendiger öffentlicher Diskurs zu einer zentralen gesellschaftlichen Frage in Gang gebracht.

Bereits in den 80-er Jahren des vergangenen Jahrhunderts hatte es eine Debatte um die Zukunft der Erwerbsarbeitsgesellschaft gegeben, die sich in der Spannung zwischen der hauptsächlich von Soziologen vertretenen Position des „Ende der Arbeit“ und der von Ökonomen dem gegenüber angeführten Standpunkt der „Arbeit ohne Ende“ abspielte. In einer gewissen Analogie dazu wird auch spätestens seit der auf den US-amerikanischen Markt bezogenen Frey-Osbourne Studie die Frage debattiert, welche Beschäftigungseffekte die Entwicklung der Digitalisierung und KI mit sich bringt.

Auf der einen Seite ist im Blick auf diese Frage darauf hinzuweisen, dass allen Prognosen zum Trotz der vorhergesagte massive Verlust an Arbeitsplätzen in der Industrie nicht eingetreten ist, obwohl die Digitalisierung ja nicht eine Entwicklung ist, auf die wir noch warten müssen, sondern eine, in der wir schon mitten drin stehen. Es gibt eine sehr hohes Beschäftigungsniveau und ein gleichzeitig immens gestiegenes Arbeitsvolumen. (Vgl. Rinne und Zimmermann 2016, S. 5) Vor dem Hintergrund dieser Entwicklungen ist insgesamt Vorsicht bei der Beurteilung des technischen Automatisierungspotenzials und entsprechender Gefahren walten zu lassen, wenn auch das grundsätzliche Veränderungspotential nicht zu übersehen ist.

Neben der Sorge um den Wegfall von Arbeitsplätzen gilt es nämlich ebenso, die „vielfältige(n) Chancen und Potentiale [in den Blick zu nehmen], die im günstigsten Fall sogar für eine Überkompensation der wegfallenden Arbeitsplätze sorgen könnten“ (Eichhorst et al. 2016, S. 2). Eichhorst u.a. verweisen darauf, dass es auch ein positives Szenario gibt, das allein für Deutschland durch Industrie 4.0 ein starkes Wachstumspotential erkennt. (Vgl. Eichhorst et al. 2016, S. 2) Die Konsequenzen dieser Transformationsprozesse sind also in keiner Weise völlig und eindeutig sowie unumkehrbar festgelegt, sondern abhängig von verschiedenen gestaltenden Faktoren. Wenn Arbeit, wie auch die christliche Sozialethik seit ihren Anfängen betont, ein menschliches Existential darstellt und ein entscheidender Weg ist, um an dieser Gesellschaft teilhaben zu können, dann zeigt sich gerade an dieser Stelle auch, wie wichtig eine entsprechende Mitgestaltung dieses Wandels ist. g.

Selbst im Fall einer positiven makroökonomischen Entwicklung bleibt, soziologisch gesehen, aufgrund der Entwicklung der Digitalisierung und der KI die Gefahr, dass die Gesellschaft in

Digitalisierungsverlierer und -gewinner gespalten wird. So werden die Arbeitsplätze zweigeteilt: in lovely and lousy jobs, d.h. „(a)ngenehme und gut bezahlte für gut ausgebildete Digitalisierungsgewinner“ (Ramge 2018, S. 31) und schlechte und schlecht bezahlte für diejenigen, die zur Weiterentwicklung der Technik keinerlei Bezug haben. Hier geht es um eine Frage sozialer Gerechtigkeit. Wird dieser Grundwert christlicher Sozialethik verstanden als partizipative Gerechtigkeit, dann leuchtet hier die Relevanz von Bildung bzw. vom Zugang zur Bildung unmittelbar ein. Es muss für Menschen aller Altersschichten und aller sozialen Milieus die Möglichkeit geben, sich ein Verständnis für Digitalisierungstechniken und einen differenzierten Umgang damit zu erarbeiten, denn mittelfristig, so jedenfalls das Urteil zahlreicher Experten, wird es wahrscheinlich keine realistische Alternative sein, dass Roboter zunehmend ganz an die Stelle der Menschen treten, sondern eher, dass technikaffine Menschen an die Stelle derer treten, die keinerlei Aufgeschlossenheit für technische Entwicklungen und für die Möglichkeit, die KI intelligent zu nutzen, aufbringen (vgl. Ramge 2018, S. 84).

Es genügt jedoch nicht, möglichst viele Arbeitsplätze zu erhalten oder neue Arbeitsplätze zu generieren. Auch bei einer zunehmenden Technisierung sollten Menschen einer sinnstiftenden Aufgabe nachgehen können. So werden in vielen Bereichen Workflows definiert, bei denen Menschen zu reinen Datenerfassern für maschinelle Verarbeitungsprozesse degradiert werden. Mediziner klagen über ständig zunehmende digitale Dokumentationspflichten, die „auch eine sich wandelnde Interaktion mit den Patienten“ mit sich bringen, bei der „der Computer wortwörtlich zwischen Patienten und Behandler“ stehe (vgl. Schmitt-Sausen 2019). Solchen Entwicklungen muss durch eine sinnvolle Aufteilung von Arbeit zwischen Mensch und Maschine entgegengewirkt werden, bei der die Maschine dem Menschen dient und nicht umgekehrt.

Fragen der Fairness im Kontext von Digitalisierung spielen über die Arbeitswelt hinaus auch dort eine Rolle, wo es um weitere Wege der digital vermittelten Partizipation an dieser Gesellschaft geht, an ihren Dienstleistungen und Gütern, an ihrer Kultur, Information und Bildung. Aus spezifisch sozialethischer Perspektive bedeutet Fairness weitergeführt auch, dass ein besonderes Augenmerk auf die zu richten ist, die auch und besonders in einer digitalisierten Gesellschaft ins Abseits zu geraten drohen, weil ihnen aufgrund schlechterer und ungünstigerer Lebens- und Startbedingungen die Teilhabe aus eigenen Kräften ungleich schwerer fällt oder unmöglich ist. Hier ist es ebenfalls Aufgabe der Sozialpolitik, Rahmenbedingungen zu schaffen, die helfen, Partizipation zu ermöglichen und Fairness zu realisieren.

Literaturverzeichnis

André, Elisabeth (2014): Lässt sich Empathie simulieren? Ansätze zur Erkennung und Generierung empathischer Reaktionen anhand von Computermodellen. In: Nova Acta Leopoldina NF 120, Nr. 405, S. 1–25.

Amina Adadi, Mohammed Berrada (2018): Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 6: 52138-52160

Bundesministerium für Verkehr und digitale Infrastruktur (2017): Ethik-Kommission. Automatisiertes und vernetztes Fahren. Online verfügbar unter https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile, zuletzt geprüft am 07.02.2020.

Capurro, Rafael (2017): Homo Digitalis. Wiesbaden: Springer Fachmedien Wiesbaden, zuletzt geprüft am 24.02.2020.

Virginia Dignum, Matteo Baldoni, Cristina Baroglio, Maurizio Caon, Raja Chatila, Louise A. Dennis, Gonzalo Génova, Galit Haim, Malte S. Kließ, Maite López-Sánchez, Roberto Micalizio, Juan Pavón, Marija Slavkovik, Matthijs Smakman, Marlies van Steenbergen, Stefano Tedeschi, Leon van der Torre, Serena Villata, Tristan de Wildt (2018): Ethics by Design: Necessity or Curse? AIES 2018: 60-66

Eichhorst, Werner; Hinte, Holger; Rinne, Ulf; Tobsch, Verena (2016): Digitalisierung und Arbeitsmarkt: Aktuelle Entwicklungen und sozialpolitische Herausforderungen (IZA Standpunkte, Nr. 85), zuletzt geprüft am 28.11.2017.

Christian Feichtinger (2017): Moral technologies und die Frage der Emotionen: Eine Weiterführung von Peter Kirchschlägers Kritik. Redaktion Feinschwarz. Theologisches Feuilleton. Online verfügbar unter: https://www.feinschwarz.net/moral-technologies-und-die-frage-der-emotionen-eine-weiterfuehrung-von-peter-kirchschlaegers-kritik/#more-7937, zuletzt geprüft am 09.05.2020.

Gottschalk-Mazouz Niels (2019): Autonomie. In: Liggieri K., Müller O. (eds) Mensch-Maschine-Interaktion. J.B. Metzler, Stuttgart

Grunwald, Armin (2019a): Autonomes Fahren: Technikfolgen, Ethik und Risiken. In: *Straßenverkehrsrecht* 19 (3), S. 81–86.

Grunwald, Armin (2019b): Der unterlegene Mensch. Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern. Originalausgabe, 1. Auflage. München: riva premium riva verlag.

Kehl, Christoph (2018): Entgrenzungen zwischen Mensch und Maschine, oder: Können Roboter zu guter Pflege beitragen? (6-8), S. 22–28. Online verfügbar unter https://www.bpb.de/apuz/263682/entgrenzungen-zwischen-mensch-und-maschine-oder-koennen-roboter-zu-guter-pflege-beitragen?p=all.

Lorenz, Philippe (2017): Digitalisierung im deutschen Arbeitsmarkt. Eine Debattenübersicht. Sankt Augustin, Berlin: Konrad-Adenauer-Stiftung.

Misselhorn, Catrin (2018): Maschinenethik und „Artificial Morality“: Können und sollen Maschinen moralisch handeln? In: *Aus Politik und Zeitgeschichte* (6-8), S. 29–33. Online verfügbar unter https://www.bpb.de/apuz/263684/koennen-und-sollen-maschinen-moralisch-handeln?p=all.

Nida-Rümelin, Julian; Weidenfeld, Nathalie (2018): Digitaler Humanismus. Eine Ethik für das Zeitalter der künstlichen Intelligenz. 2. Auflage, Originalausgabe. München: Piper.

Ortony, Andrew, Clore, Gerald L. ; Collins, Allan (1988): The Cognitive Structure of Emotions. Cambridge, UK: Cambridge University Press.

Ramge, Thomas (2018): Mensch und Maschine. Wie künstliche Intelligenz und Roboter unser Leben verändern. Unter Mitarbeit von Dinara Galieva. Ditzingen: Reclam ((Was bedeutet das alles?), Nr. 19499).

Rinne, Ulf; Zimmermann, Klaus F. (2016): Die digitale Arbeitswelt von heute und morgen. In: *Aus Politik und Zeitgeschichte* (18-19), S. 3–9.

Schmitt-Sausen, Nora (2019): Digitale Medizin: Ärzte müssen eingebunden werden. Dtsch Arztebl 2019; 116(13): A-630 / B-516 / C-508

Silberg, Jake; Manyika, James (2019): Notes from the AI Frontier: Tackling bias in artificial intelligence (and in humans). McKinsey Global Institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans, zuletzt geprüft am 09.05.2020

Tangney, June Price; Stuewing, Jeff; Mashek, Debra J. (2007): Moral emotions and moral behavior. Annual review of psychology vol. 58 (2007): 345-72. doi:10.1146/annurev.psych.56.091103.070145

Andrea Aler Tubella, Andreas Theodorou, Frank Dignum, Virginia Dignum (2019): Governance by Glass-Box: Implementing Transparent Moral Bounds for AI Behaviour. IJCAI 2019: 5787-5793

Webwelt und Technik (2018): Wer haftet, wenn künstliche Intelligenz Mist baut? Online verfügbar unter: https://www.welt.de/wirtschaft/webwelt/article181494476/Wer-haftet-wenn-eine-kuenstliche-Intelligenz-Mist-baut.html, zuletzt überprüft am 11.05.2020