

## 1.2 Werte: Was können ethische Ansätze für eine wertorientierte Digitalisierung leisten?

### Analyse, Systematisierung und Einordnung

---

Petra Grimm

#### 1 Problemaufriss – in welchen Zusammenhängen stellen sich Wertefragen?

Wer Digitalisierung und einen der Antriebsmotoren – die künstliche Intelligenz (KI) – nur als *technisches* Gesamtprojekt auffasst, missversteht die Bedeutung dieses Prozesses für jede beziehungsweise jeden von uns und die Gesellschaft insgesamt. Denn mit der Digitalisierung ändern sich unsere Bedingungen für ein gutes Leben, unser Denken und unsere Gefühle in einer sehr umfassenden Weise. Das betrifft sowohl das Privatleben als auch das Arbeitsleben. Wie erleben die Bürger\*innen diese Veränderungen? In einer Studie<sup>1</sup> zu den Werten, Ängsten und Hoffnungen in der digitalen Alltagswelt zeigt sich, dass die Digitalisierung in beruflich-professionellen Kontexten übergreifend eher positiv wahrgenommen und bewertet wird. Sie wird als Segen, wenn nicht gar als Glück, von den Befragten beschrieben, wobei damit ökonomisch geprägte Werte wie Nützlichkeit, Effizienz und Effektivität gemeint sind. Im Kontext der privaten Lebenswelt wird die Digitalisierung dagegen eher kritisch – zumindest aber differenzierter – gesehen. Es werden Wertekonflikte und teils sogar Werteverluste, wie zum Beispiel der Ehrlichkeit, Zuverlässigkeit und Privatheit, benannt, vor allem in Bezug auf die sozial-kommunikativen Praktiken bei der Nutzung sozialer Internetplattformen und Kommunikationsdienste. Eine *Ethik der Digitalisierung* muss die

---

1 Grimm, Petra/Müller, Michael/Trost, Kai Erik: Werte, Ängste, Hoffnungen. Das Erleben der Digitalisierung in der erzählten Alltagswelt, Baden-Baden: Nomos (i. Ersch.).

Sichtweisen der Bürger\*innen wie auch aller anderen Beteiligten (in Wissenschaft, Wirtschaft, Politik etc.) berücksichtigen. Gleichwohl ist die eigentliche Aufgabe einer Ethik der Digitalisierung, ethisches Handeln und Wertefragen zu reflektieren sowie normative Standards zu begründen. Hierzu muss sie auch die Bedingungen und Strukturen der digitalen Systeme sowie die damit zusammenhängenden sozialen, politischen, ökonomischen und rechtlichen Vorgaben in den Blick nehmen.

Der vorliegende Beitrag soll den Horizont für eine werteorientierte Digitalisierung abstecken: In einem ersten Schritt werden die relevanten Spannungsfelder für Wertefragen beim Einsatz *künstlicher Systeme*<sup>2</sup> anhand von Fallgeschichten diagnostiziert. Sodann werden die Werte selbst betrachtet: Was sind (moralische) Werte? Gibt es universale Werte? Und welche Werte sind aus digitaletischer Sicht insbesondere von Bedeutung? In diesem Zusammenhang wird eine Werte-Topografie der Digitalen Ethik vorgestellt. Schließlich werden zentrale ethische Ansätze daraufhin untersucht, welche Stärken und Schwächen sie für die Begründung und Perspektivierung einer Digitalen Ethik haben. Vollzogen wird dies sowohl beispielhaft in Bezug auf die Konflikte in den Fallgeschichten, als auch grundsätzlich in Bezug auf ethische Leitlinien, eine Werterhaltung sowie eine wertebasierte Gestaltung der Digitalisierung (Ethics by Design).

Beginnen wir mit der Diagnose der Spannungsfelder für Wertefragen beim Einsatz künstlicher Systeme.

## 1.1 Das individuelle Spannungsfeld des Wertehandelns

Menschen geraten in Konflikte, wenn sie im Umgang mit digitalen Systemen ihren Wertvorstellungen entsprechend handeln wollen. Die Gestaltung digitaler Systeme kann es Nutzer\*innen erschweren, werteorientiert zu handeln, beziehungsweise sie davon abhalten, ihre Wertepreferenzen im Alltag umzusetzen. Hierzu eine Fallgeschichte (1):

Der 20-jährige Alex möchte seine Privatsphäre schützen und kontrollieren, wer was in welchem Zusammenhang über ihn weiß. Er beschließt, auf In-

---

2 Da der Begriff der *künstlichen Intelligenz* umstritten ist (vgl. Kap 3.2), wird im Folgenden allgemein von künstlichen Systemen gesprochen; dazu zählen auch die Begriffe autonome beziehungsweise automatisierte Systeme, maschinelles Lernen, neuronale Netze etc., soweit sie nicht hinsichtlich ihrer Funktionsfähigkeit im konkreten Bezugsrahmen unterschieden werden.

stagram, Facebook und WhatsApp zu verzichten und stattdessen die privatheitssichernde Plattform Mastodon und den Messenger Signal zu verwenden. Dort trifft er jedoch nur wenige seiner Freunde. Zudem verabreden sich seine Kommiliton\*innen regelmäßig in Lerngruppen auf Facebook. Das heißt, er gerät in ein moralisches Dilemma: Entweder ist er mehr oder weniger sozial ausgeschlossen und im Studium benachteiligt oder er kann seine Privatsphäre nicht schützen. Beide Handlungen würden nicht zu einer befriedigenden Lösung führen. Er könnte natürlich noch versuchen, seine Freunde zu überreden, mit ihm die Anbieter zu wechseln. Erfahrungsgemäß würde ihn dies einiges an Überzeugungskraft mit ungewissem Ausgang kosten. Was soll er tun?

Die Geschäftsmodelle der amerikanischen Digitalmonopole wie Alphabet (Google, Youtube etc.), Facebook, Amazon, Microsoft etc. beruhen auf der wirtschaftlichen Ausbeute der personenbezieharen Daten. Da diese Tech-Unternehmen über eine Machtkonzentration verfügen, die ohnegleichen in der Geschichte ist, können die Menschen ihre Privatsphäre nur dann weitgehend schützen, wenn sie auf das System verzichten, es (zum Beispiel als Computerspezialist\*in) unterlaufen oder wenn sich die normativen Rahmenbedingungen des Systems ändern. Es genügt also nicht, allein die Mikro-Ebene der Akteure zu betrachten, sondern es bedarf auch einer systemethischen Perspektive, die die Meso-Ebene der Unternehmen und die Makro-Ebene der Gesellschaft betrifft. Um Alex zu befähigen, seine Privatsphäre schützen zu können, müssten die digitalen Angebote entsprechend gestaltet und Recht durchgesetzt werden. Warum sollten die Unternehmen das tun? Und wenn sie es tun, wie können sie wertebasierte Produkte entwickeln? Gibt es eine Pflicht der Politik, hierauf einzuwirken? Warum ist Privatsphäre überhaupt ein Wert? Können auch Nutzer\*innen durch ihre Haltung und ihr Verhalten dazu beitragen, dass wertorientierte Produkte, zum Beispiel solche, die die Privatheit schützen, entwickelt werden?

Ethische Ansätze können hilfreich sein, diesen Fragen auf den Grund zu gehen. Bevor diese vorgestellt und eingeordnet werden, sollen zwei weitere Beispiele die Komplexität der digitalen Herausforderungen veranschaulichen. Denn die Digitalisierung betrifft nicht nur die sozialen Medien und Kommunikationsdienste, sie durchdringt oftmals viele, auch existenzielle Lebensbereiche und wird dabei nicht bewusst wahrgenommen. In diesem Zusammenhang lässt sich ein zweites Spannungsfeld für Wertefragen diagnostizieren:

## 1.2 Das Spannungsfeld zwischen Individuum und Gemeinschaft

Konflikte zeigen sich nicht nur im individuellen Wertehandeln, sondern in der Vergemeinschaftung und Verallgemeinerung von Geltungsansprüchen. Wenn digitale Systeme für Menschen oder Gruppen von Menschen ein besseres Leben ermöglichen, kann das zu Interessenskonflikten mit anderen Gruppen führen. Fragen der Gerechtigkeit, Gleichheit und Solidarität oder allgemein des Gemeinwohls, die in der sozial- beziehungsweise organisationsethischen Dimension zu verorten sind, können mit den für Einzelne oder Gruppen relevanten Werten im Spannungsfeld liegen. Folgende Fallgeschichte (2) veranschaulicht dies:

Ingeburg K., 80 Jahre alt, lebt seit 30 Jahren in ihrer Wohnung und fühlt sich, auch wenn sie einen Rollator braucht und einmal täglich der Pflegedienst kommt, noch ziemlich fit. Ein Pflegeheim kommt für sie nicht infrage. Als sie von einem Forschungsprojekt liest, bei dem digitale Assistenzsysteme für Senior\*innen zu Hause getestet werden sollen, erklärt sie sich spontan bereit, mitzumachen. Wo die Bewegungsmelder installiert werden sollen, kann sie selbst entscheiden. Auch das Angebot, Blutdruck, Herzfrequenz und andere Parameter zu überwachen, nimmt sie an. Ihre Tochter, die dem digitalen System nicht so ganz vertraut, besucht sie nach wie vor regelmäßig; ihr Sohn aber telefoniert in letzter Zeit eher mit ihr, als dass er noch bei ihr vorbeikommt. Als sie eines Tages stürzt und sich den Oberarmhals bricht, erfasst das System die Abweichung und informiert den Notdienst, der alles Nötige veranlasst. Unklar ist, warum sie stürzte. Fühlte sie sich durch das Kontrollsystem vermeintlich sicherer und bewegte sich deshalb risikoreicher als zuvor oder geschah der Sturz unabhängig davon? Oder fühlte sie sich möglicherweise dazu veranlasst, sich mehr als sonst zu bewegen, um einen möglichst fitten Eindruck zu hinterlassen? Mehr Bewegung ist ja prinzipiell gut, aber mit dieser Folge?

Würde sich erweisen, dass das System den Sturz mitverursacht hat, wären – abgesehen von rechtlichen Aspekten wie zum Beispiel dem Datenschutz und der Haftungsfrage – ethische, soziale und psychologische Perspektiven zu berücksichtigen, und zwar unter Einbeziehung aller Beteiligten (sogenannte Stakeholder): die Sicht der Pflegebedürftigen, des Pflegedienstes, der Kinder, der Dienstleister\*innen und Entwickler\*innen. Dabei geht es insbesondere um Wertefragen, sowohl aus individualetischer als auch sozial- und organisationsethischer Sicht. Hierbei handelt es sich um unterschiedliche Ge-

genstandsbereiche der ethischen Praxis. So adressiert die Individualethik die Mikro-Ebene der Akteure beziehungsweise des Handlungssubjekts: das Handeln der oder des Einzelnen unter den Aspekten der moralischen Rechte, Pflichten, Handlungsmotivationen, Tugenden (Charaktermerkmale, Dispositionen) sowie deren Vorstellungen von einem guten und gelingenden Leben. Die Sozialethik richtet hingegen den Blick auf die Makro-Ebene der Gesellschaftsstruktur, die die Ermöglichungsräume, Bedingungen und Grenzen für individuelles Handeln vorgibt. Wertbegriffe wie soziale Gerechtigkeit, Gemeinwohl, Solidarität und Good Governance lassen sich dieser Ebene zuordnen, wobei eine scharfe Abgrenzung zur Individualethik nicht vorgesehen ist. Es handelt sich vielmehr um zwei sich ergänzende und gegebenenfalls überschneidende Perspektiven. Die erste Fallgeschichte von Alex zeigt, dass der Schutz der Privatsphäre im digitalen Alltag sowohl eine individual- als auch sozialetische Dimension hat.

Gleiches gilt für die Geschichte von Ingeburg K.: Aus individualethischer Sicht ist zu fragen, ob und inwieweit das digitale Kontrollsystem ein Mehr an Selbstbestimmung, Gesundheit und Sicherheit (Rechte) für Ingeburg K. bietet und wie sich ihre Angehörigen ihr gegenüber verhalten sollten (Pflichten). Wenn sich das System für die häusliche Pflege für Ingeburg K. als sinnvoll erweisen würde, stellt sich aus sozialetischer Perspektive die Frage nach dem grundlegenden Anspruch und einem gerechten Zugang zu diesem Pflege-Assistenzsystem: Sollten alle Pflegebedürftigen einen Anspruch auf ein solches System haben, weil es ihnen mehr Autonomie im Alter und womöglich auch ein längeres Leben ermöglicht (zum Beispiel weil frühzeitig Gesundheitsrisiken erkannt werden)? Ist im Sinne einer Solidargemeinschaft der Geltungsanspruch auch dann noch zu rechtfertigen, wenn die Versicherungsbeiträge für alle anderen steigen? Und deren Zahl größer ist als die der Pflegebedürftigen? Liegt das Gemeinwohl in einer höheren Anzahl von Profiteuren oder in einer zu benennenden höheren Qualität des Zusammenlebens (siehe Kap. 3)?

Die zwei bislang diskutierten Beispiele für Spannungsfelder von Wertefragen bei digitalen Systemen haben noch nicht der Tatsache Rechnung getragen, dass die digitalen Systeme zunehmend selbstständig in moralischen Entscheidungssettings agieren (künstliche Systeme). Hierzu lässt sich ein weiteres Spannungsfeld von Wertefragen diagnostizieren:

### 1.3 Künstliche Systeme im Spannungsfeld von ethischem und nicht-ethischem Design

Künstliche Systeme sind wie alle technischen Systeme weder gut noch schlecht, aber auch nicht wertneutral, wie der Technikhistoriker Melvin Kranzberg allgemein schon feststellte.<sup>3</sup> Das heißt, dass technische Entwicklungen häufig ökologische, soziale und individuelle Konsequenzen haben, die weit über die unmittelbaren Zwecke der technischen Geräte und Praktiken selbst hinausgehen. Zudem fließt in die Entwicklung der digitalen Produkte und Systeme immer auch (implizit oder explizit) das Wertesystem der Entwickler\*innen, Programmierer\*innen und Manager\*innen mit ein. Künstliche Systeme agieren nicht im luftleeren Raum, sondern in wertbezogenen Anwendungskontexten. Ob bei der Kreditvergabe, bei der Jobbewerbung, bei der Notenvergabe an Schüler\*innen oder beim (teil-)autonomen beziehungsweise (hoch-)automatisierten Fahren stellen sie Prognosen auf, sind in Entscheidungsprozesse eingebunden oder agieren sogar selbstständig; sie sind gleichsam *Aktanten* mit moralischen Implikationen. Ihre Fähigkeit, Muster und Korrelationen in Big Data zu erkennen sowie Prognosen zu berechnen, kann für uns in Hinblick auf die großen Herausforderungen unserer Zeit von Vorteil sein, sei es in der medizinischen Diagnostik, bei der Energieversorgung, bei der Ressourcenschonung und der Mobilität. Zugleich können ihre Berechnungen und Prognosen nicht nur in die Selbstbestimmung und Autonomie des Menschen eingreifen, sie können auch zu Diskriminierung, Ausgrenzung und Geschlechterungerechtigkeit führen. Auch die zugrunde liegende ökonomische und politische Programmatik wirken auf die jeweilige Technikanwendung. Für Felix Stalder ist nicht die Technik, sondern eine neoliberale Ideologie das Problem: »Das zentrale Hindernis für Algorithmen, wie wir sie wollen, liegt in der nach wie vor alle Felder der Gesellschaft dominierenden neoliberalen Programmatik.«<sup>4</sup> Wie sich diese im Alltag auswirken kann, veranschaulicht die Fallgeschichte (3) von Louise und Ann-Kathrin:

3 Vgl. Kranzberg, Melvin: *Technology and history: Kranzberg's laws*, in: *Technology and Culture*, Vol. 27, No. 3, 1986, S. 545.

4 Stalder, Felix: *Algorithmen, die wir brauchen. Überlegungen zu neuen technopolitischen Bedingungen der Kooperation und des Kollektiven*, in: Philipp Otto/Eike Gräf (Hg.): *3Ethics. Die Ethik der digitalen Zeit*, Berlin: iRights.Media 2017, S. 54.

Louise bewirbt sich bei einem großen amerikanischen Unternehmen. Sie hat sehr gute Abschlussnoten, einschlägige erforderliche Kenntnisse etc. Bei ihren Hobbys gibt sie an, dass sie auch einen Preis im Frauen-Schach gewonnen hat. Weder sie noch andere Bewerberinnen werden von dem KI-gesteuerten Auswahlinstrument für ein Vorstellungsgespräch vorgeschlagen. Denn die KI hat auf der Datenbasis vorhergehender erfolgreicher Bewerbungsprofile im Unternehmen, die vorwiegend von Männern waren, »gelernt«, dass Frauen nicht zu berücksichtigen sind. Nachdem Louise später von der diskriminierenden KI erfährt, fühlt sie sich einerseits ungerecht behandelt und gedemütigt, andererseits kann sie der KI auch keine Schuld zuweisen. In ihren Augen ist die KI nicht dafür verantwortlich zu machen, es ist ja eine Maschine.<sup>5</sup> Ihre deutsche Freundin Ann-Kathrin, die sie im Auslandsstudium kennengelernt hat, bewirbt sich ebenfalls gerade auf einen neuen Job. Sie wird von dem Unternehmen zu einem Telefoninterview eingeladen und ist sehr aufgeregt. Das Interview führt sie allerdings nicht mit einem Menschen, sondern mit einer Sprachanalysesoftware, die anhand von Stimmen und dem Inhalt des Gesprächs psychologische Persönlichkeitsprofile erstellt. Sie hat Glück, kurze Zeit später wird sie zu einem persönlichen Vorstellungsgespräch eingeladen. Sie freut sich darüber, hat aber auch ein unbehagliches Gefühl, denn sie fühlt sich dem System ausgeliefert: Wie kommt das Analysesystem zu dieser Entscheidung? Und welche psychologischen Informationen hat es über sie gesammelt und in welches Muster eingeordnet?

Auch wenn die Bewerbung für die beiden Frauen unterschiedlich ausgeht, relevant ist, dass die Entscheidungen der künstlichen Systeme existenzielle Konsequenzen haben. Dabei geht es nicht um die Frage der Objektivität, also ob künstliche Systeme oder Menschen objektiver entscheiden können,

---

5 Die Geschichte von Louise bezieht sich auf eine wahre Begebenheit. So arbeitete Amazon 2014 an der Entwicklung eines KI-gesteuerten Recruiting-Systems, das wegen des Gender Bias 2017 eingestellt wurde. Wie Reuters (1.10.2018) berichtete (<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>), wurde das System anhand von Bewerbungen trainiert, die innerhalb der letzten zehn Jahre bei Amazon eingegangen waren. Dabei handelte es sich vorwiegend um Männer. Entsprechend diesem Muster folgerte das System aus dem vorliegenden Datensatz, dass Männer die bevorzugten Kandidaten seien. Bewerbungen, in denen Begriffe wie »Frauen-Schach« als Hobby genannt wurden, wurden daher schlechter bewertet.

denn diese Frage ist müßig: weder das System noch der Mensch sind es. Vielmehr geht es darum, *dass* ein künstliches System einen Entscheidungsprozess maßgeblich steuert. Das führt uns zu der Frage, *welche Wertansprüche an die digitalen Systeme* gestellt werden sollen und, wenn überhaupt, in welchem Maße Verantwortung delegiert werden kann. Wie die Geschichte von Louise und Ann-Kathrin zeigt, sind die besagten künstlichen Systeme nicht den Werten der Fairness, Chancengleichheit, Autonomie, Privatheit, Transparenz und Nachvollziehbarkeit folgend gestaltet worden. Die Geschlechterungerechtigkeit (Gender Bias), die das künstliche System bei dem Bewerbungsverfahren von Louise verursachte, verweist noch auf eine weitere Problemlage.

#### 1.4 Manifestationen von Diskriminierungen

Digitale Systeme können sowohl in der Gesellschaft vorhandene Diskriminierungen fixieren beziehungsweise verstärken als auch fördern.<sup>6</sup> Ein Beispiel für die *Fixierung* vorhandener Geschlechterrollenklischees ist der durch Microtargeting ermöglichte sogenannte Gender Commerce, der in vielfältiger Weise angeboten wird. Genderzentriertes Onlineshopping beruht auf einer stereotypen Filterung der Angebote für Werbekund\*innen. Frauen werden dann eher Kosmetika als Fotoapparate angezeigt und auch die Werbeansprache wird unterschiedlich gestaltet. Dies kann möglicherweise von den Nutzer\*innen sogar erwünscht sein. Ein Beispiel für die *Verstärkung* vorhandener Geschlechterrollenklischees sind Sprachassistenzsysteme wie zum Beispiel Alexa, Siri und Cortana. Standardmäßig werden sie mit weiblichen Stimmen und Namen ausgestattet. Die feminisierten Sprachassistenzsysteme erhalten Sprachbefehle und führen sie als Serviceleistung aus. Das hier gegenderte Mensch-System-Interaktionsmuster erlaubt damit Rückkoppelungen an geschlechterstereotype Interaktionsmuster, die eine *Herabwürdigung* von Frauen implizieren. Dies wird insbesondere bei feminisierten Sprachassistenten deutlich, denen devote bis allenfalls ausweichende Antworten auf sexuell belästigende Fragen einprogrammiert wurden, wie EQUALS und UNESCO in ihrem Bericht mit Beispielen zeigen.<sup>7</sup>

6 Vgl. O'Neill, Cathy: Angriff der Algorithmen, München: Carl Hanser Verlag 2016; Criado-Perez, Caroline: Unsichtbare Frauen. Wie eine von Daten beherrschte Welt die Hälfte der Bevölkerung ignoriert, München: btb Verlag 2020.

7 EQUALS and UNESCO: I'd blush if I could. Closing gender devices in digital skills through education 2019, S. 107. <https://en.unesco.org/ld-blush-if-i-could>

Die symbolische Feminisierung von künstlichen Systemen kann stereotype Geschlechterrollen nicht nur online, sondern auch offline verstärken und verfestigen. Um dem entgegenzuwirken, könnte eine geschlechterneutrale, modulierte Stimme verwendet werden.<sup>8</sup>

Diskriminierung kann in digitale Systeme auch durch eine *Priorisierung des Männlichen* auf der sprachlichen und damit kognitiven Ebene eingeschrieben sein. Ein Beispiel hierfür ist der Übersetzungsdienst Google Translate, der in Google Chrome und damit automatisch in circa 250 Millionen Android-Smartphones europäischer Bürger\*innen integriert ist. So verschwindet das weibliche Geschlecht von Berufen beziehungsweise Funktionen in der Übersetzung, wie AlgorithmWatch in einem Experiment von Nicolas Kayser-Bril zeigt:<sup>9</sup> »Die Präsidentin« wird im Italienischen zum »il presidente«, korrekt wäre »la presidente«. Auch ändert Google Translate das Geschlecht nach genderstereotypem Muster: »Der Krankenpfleger« wird im Französischen zur »l'infirmière«, also der »Krankenpflegerin«.<sup>10</sup>

Systematisiert man die Beispiele, zeigt sich folgendes Bild (vgl. Abb. 1): Gender Bias, die zu einer Diskriminierung führen können, lassen sich in einer Matrix mittels der Achse *Kommunikation/Interaktion* und *Entscheidungs-/Empfehlungssysteme* sowie der Achse *Verstärkung* der Geschlechterrollen und *Marginalisierung/Tilgung* der Frauen einordnen. Im ersten Feld der Kommunikation/Interaktion wird das Herabwürdigungsprinzip (Sprachassistent-Beispiel) verortet, im zweiten Feld eine Priorisierung des Männlichen (Google-Translate-Beispiel), im dritten Feld der Entscheidungs- und Empfehlungssysteme besteht das Risiko der Ausgrenzung von Frauen

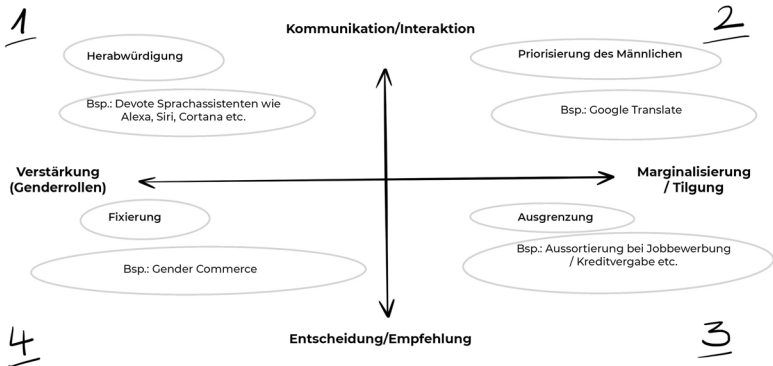
8 Hierzu s. ein praktisches Beispiel: <https://www.genderlessvoice.com/>

9 Kayser-Bril, Nicolas: Female historians and male nurses do not exist. Google Translate tells its European users, 17.09.2020. <https://algorithmwatch.org/en/story/google-translate-gender-bias/>

10 Ein Selbstversuch mit der Google Übersetzer-App zeigt ein ähnliches Ergebnis: Aus der »Medienwissenschaftlerin« wird in der italienischen Übersetzung der »lo scienziato die media«, umgekehrt wird aus der »scienziata die media« in der deutschen Übersetzung der »Medienwissenschaftler«. Hinzu kommt die von Google verursachte manipulative Nutzungslenkung, denn Android-Nutzer\*innen entkommen dem Google Translate-System nicht oder nur mit Anstrengung: »Since an update in April 2019, Google Chrome prompts users to instantly translate web pages. Anyone visiting a website in a foreign language is asked to choose between the original or the google-translated version, even if the website offers an official translation in the user's preferred language.« Nicolas Kayser-Bril: Female historians and male nurses do not exist.

(Bewerbungssystem in der Louise-Geschichte) und im vierten Feld das der Fixierung von Geschlechterrollen (Gender-Commerce-Beispiel).

Abbildung 1: Geschlechterbezogenen Diskriminierung



Im philosophischen und rechtswissenschaftlichen Diskurs stellt sich zudem die Frage, ob digitale Systeme als moralische Akteure oder gar Verantwortungssubjekte zu betrachten sind und Rechtssubjekte (Träger von Rechten und Pflichten) sein können. Unzweifelhaft dürfte sein, dass sie zunehmend in Verantwortungs- und Handlungszusammenhängen agieren, die ethische Relevanz haben. Janina Loh spricht von »Verantwortungsnetzwerken«, die vorliegen, »wenn man eigentlich [...] gar nicht mehr weiß, ob hier in einem gehaltvollen Sinn Verantwortung definiert werden kann, gerade weil zum Beispiel die Bestimmung eines Subjekts schwierig erscheint oder aber sich keine eindeutige Instanz ausmachen lässt oder aber die normativen Kriterien nicht benannt werden können«. <sup>11</sup> Ob und unter welchen Voraussetzungen Maschinen als moralische Akteure (*moral agents*) zu verstehen sind und sie verantwortlich beziehungsweise haftbar für die Folgen gemacht werden können, ist in der Roboterethik und Maschinenethik ein viel diskutiertes Thema. <sup>12</sup> Be-

11 Loh, Janina: *Verantwortung und Roboterethik. Ein Überblick und kritische Reflexion*, in: Matthias Rath/Matthias Karmasin/Friedrich Krotz (Hg.): *Maschinenethik. Normative Grenzen autonomer Systeme*, Wiesbaden: Springer VS 2019a, S. 103.

12 Vgl. hierzu u.a. Misselhorn, Catrin: *Grundfragen der Maschinenethik*, Dietzingen: Reclam 2018, S. 70-90; Loh, Janina: *Roboterethik. Eine Einführung*, Berlin: Suhrkamp 2019b, S. 48-95; Rath, Matthias: *Zur Verantwortungsfähigkeit künstlicher »moralischer Akteure«*. *Problemanzeige oder Ablenkungsmanöver?*, in: Matthias Rath/Matthias Karmasin/

vor auf das Für und Wider näher eingegangen wird, können hier schon die zentralen Fragen formuliert werden: Sind künstliche Systeme überhaupt als moralische Akteure zu betrachten? Falls ja, was folgt daraus?

Aus ethischer Sicht stellt sich des Weiteren die Frage, ob Roboter als »moralische Objekte« (*moral patients*) zu betrachten sind. Kurzum: Sollten wir rücksichtsvoll, anerkennend, freundlich und emphatisch mit vermenschlichten Maschinen interagieren – nicht um der Maschinen willen, sondern um unseres wertschätzenden Miteinanders willen?

Die bislang beschriebenen Spannungsfelder haben eine gemeinsame Perspektive: die des Individuums – auf sich selbst, die Gemeinschaft und das Design der digitalen Systeme. Ein bedeutsames Spannungsfeld ergibt sich aber auch auf der ideologischen Meta-Ebene hinsichtlich der Unternehmen beziehungsweise Organisationen und der Gesellschaft, die sich in konkurrierenden Meta-Narrativen widerspiegeln. Um zu verstehen, welche Wertkonzeptionen im Diskurs um die Gestaltung der Digitalisierung konkurrieren, sollen diese mithilfe einer narratologischen Analyse herausgearbeitet werden.

## 1.5 Digitalisierung im Spannungsfeld konkurrierender Meta-Narrative

Wie wir uns, die anderen und die Welt verstehen, hängt maßgeblich davon ab, welche Erzählungen, also Narrative, wir Phänomenen und Entwicklungen zuschreiben. Denn sie beeinflussen unsere Sicht auf die Welt und unser ethisches beziehungsweise unethisches Verhalten. Oder anders: Es gibt keine absoluten Werte, nur (geteilte) Erzählungen von diesen. Meta-Narrative sind narrative Denkmodelle, mit denen eine Kultur ihre Kommunikation zu einem Thema beziehungsweise einen Diskurs strukturiert, um so Erklärungsmuster für bestimmte vergangene Entwicklungen (*warum ist es so geworden?*) oder für mögliche zukünftige Entwicklungen (*so wird es weitergehen*) zu erhalten.<sup>13</sup> Meta-Narrative sind als übergeordnete Narrative zu verstehen, die sich aus einzelnen narrativen Kommunikationsakten (Kommunikaten) abstrahieren lassen. Diese Erklärungsmuster müssen nicht in einem wissenschaftlichen Sinne zutreffend oder valide sein. Oftmals ist für kulturelle

---

Friedrich Krotz (Hg.): *Maschinenethik. Normative Grenzen autonomer Systeme*, 2019, S. 223-242; Wallach, Wendell/Allen, Collin: *Moral Machines. Teaching Robots Right from Wrong*, New York: Oxford University Press 2009.

13 Vgl. Müller, Michael/Grimm, Petra: *Narrative Medienforschung. Einführung in Methodik und Anwendung*. Konstanz/München: UVK Verlagsgesellschaft 2016, S. 97-10.

Meta-Narrative bezeichnend, dass sie *nicht* evidenzbasiert sind: So beruhen beispielsweise die Meta-Narrative der Verschwörungserzählungen im Zuge der Corona-Pandemie, die zu einer »Infodemie«<sup>14</sup> geführt haben, ja geradezu darauf, dass sie alles andere als durch Fakten bestätigt sind. Der Begriff des *Meta-Narrativs* geht auf Jean-François Lyotard zurück, der vom Ende der »großen Erzählungen« sprach, was sich allerdings als falsch erwies.<sup>15</sup> Die Kultur einer Gesellschaft konstituiert sich durch Erzählungen, wie Wolfgang Müller-Funk meint: »Naheliegend wäre es, die konstitutive Bedeutung von Narrativen für Kulturen ins Auge zu fassen und Kulturen womöglich als mehr oder weniger (hierarchisch) geordnete Bündel von expliziten und auch impliziten, von ausgesprochenen, aber auch verschwiegenen Erzählungen zu begreifen.«<sup>16</sup> Mittels einer narratologischen Methode<sup>17</sup> lassen sich also gesellschaftliche Diskurse, sofern ihnen narrative Strukturen zugrunde liegen, beschreiben. Sie dient auch der narrativen Ethik, um ethische Werte deskriptiv zu erfassen. Die Narratologie und die narrative Ethik decken sich in ihrem Bild vom Menschen als *homo narrans*,<sup>18</sup> wonach der Mensch seine Identität, sein Gedächtnis, sein Werteverständnis und seine Emotionen narrativ konstruiert.

Für die Analyse der Meta-Narrative braucht es vorab eine Begriffsklärung: Was ist eigentlich ein *Narrativ*? Der Begriff hat in der Nachfolge eines *narrative turn* in den Geisteswissenschaften zu einem beinahe inflationären Gebrauch geführt und wird uneinheitlich verwendet sowie selten definiert. Hier ist unter einem Narrativ nur dann ein Narrativ zu verstehen, wenn es eine narrative Struktur aufweist. Diese besteht aus einer triadischen Struktur mit einer Ausgangssituation, einer Transformation und einer Endsituation, wobei eine

- 
- 14 Grimm, Petra: *Entwirklichung – Zum Vertrauen in Zeiten der digitalen Infodemie*, in: Klaus Koziol (Hg.): *Entwirklichung der Wirklichkeit. Von der Suche nach neuen Sicherheiten*, München: Kopaed 2020, S. 55-83.
- 15 Lyotard, Jean-François: *Das postmoderne Wissen. Ein Bericht*, Wien: Passagen 2012.
- 16 Müller-Funk, Wolfgang: *Die Kultur und ihre Narrative. Eine Einführung*, Wien/New York: Springer 2008, S. 17.
- 17 Müller, Michael/Grimm, Petra: *Narrative Medienforschung. Einführung in Methodik und Anwendung* 2016, S. 55-96; Titzmann, Michael: *Narrative Strukturen in semiotischen Äußerungen*, in: Hans Krahl/Michael Titzmann (Hg.): *Medien und Kommunikation. Eine interdisziplinäre Einführung*. Dritte, stark erweiterte Auflage, Passau: Stutz 2013, S. 113-141.
- 18 Grimm, Petra: *Haltung in einer digitalisierten Kindheit. Die Perspektive der narrativen Ethik*, in: Ingrid Stapf/Marlis Prinzing/Nina Köberer (Hg.): *Aufwachsen mit Medien. Zur Ethik mediatisierter Kindheit und Jugend*, Baden-Baden: Nomos 2019, S. 85-99.

notwendige, aber nicht hinreichende Bedingung (es gibt noch weitere) gegeben sein muss: Die Transformation muss *ereignishaft*, also bedeutsam sein, sei es für einen Akteur, eine Gemeinschaft oder eine soziale, politische oder ökonomische Ordnung. Der Digitalisierungsprozess als solcher kann als ereignishaft Veränderung verstanden werden. Je nach Perspektive kann er allerdings unterschiedlich erzählt werden. Der Forderung, dass wir im Zuge der Digitalisierung »unsere derzeitigen Narrative einer kritischen Prüfung unterziehen müssen, und zwar auf individueller, gesellschaftlicher und politischer Ebene«, wie Luciano Floridi<sup>19</sup> meint, sollte nachgekommen werden. Dazu ist aber erst einmal zu klären, welche Meta-Narrative sich derzeit in Bezug auf die Digitalisierung und einen der Antriebsmotoren – die künstliche Intelligenz – in den Vordergrund drängen.

Ob in den Medien, in Unternehmen, in pädagogischen oder politischen Diskussionen, zwei gegensätzliche Meta-Narrative bestimmen die Diskurse zur künstlichen Intelligenz: das Meta-Narrativ vom *Heiligen Gral* und das von der *Büchse der Pandora*, wie es hier bezeichnet wird.

Das *Heilige-Gral-Narrativ* erzählt uns die Geschichte der künstlichen Intelligenz als eine, die unsere großen Probleme lösen wird. Demnach würde die Zukunftstechnologie der künstlichen Intelligenz eine bessere Welt ermöglichen, mehr Lebensqualität und Komfort bieten und globale Probleme (wie Klimawandel, Epidemien, Krebs etc.) überwinden helfen. Dieses Meta-Narrativ findet sich in unterschiedlichen Facetten, angefangen von moderaten Auslegungen, die bloß den wirtschaftlichen Nutzen digitaler Technologien in den Vordergrund stellen, bis zu extremen Erzählungen, nach deren Weltbild eine Optimierung des Menschen durch die digitalen Technologien unbedingt erforderlich sei. Die Erzähler\*innen dieser technikttotalitären Geschichte sind die großen Tech-Unternehmen aus dem Silicon Valley, deren Welt- und Menschenbild Paul Nemitz und Matthias Pfeffer<sup>20</sup> als »Kalifornische Ideologie« eindrücklich beschrieben haben. Sie nehmen dabei auch auf den von Yuval Noah Harari geprägten Begriff des »Dataismus« Bezug: »Dem Dataismus zufolge besteht das Universum aus Datenströmen, und der Wert jedes Phäno-

---

19 Floridi, Luciano: *Die Mangroven-Gesellschaft. Die Infosphäre mit künstlichen Akteuren*, in: Otto Philipp/Eike Graf, (Hg.): *3Ethics. Die Ethik der digitalen Zeit*, Berlin: IRights Media 2017, S. 28.

20 Nemitz, Paul/Pfeffer, Matthias: *Prinzip Mensch. Macht, Freiheit und Demokratie im Zeitalter der Künstlichen Intelligenz*, Bonn: J.H.W. Dietz 2020, S. 90-153.

mens oder jedes Wesens bemisst sich nach seinem beziehungsweise ihrem Beitrag zur Datenverarbeitung.«<sup>21</sup>

Das *Pandora-Narrativ* erzählt uns dagegen die Geschichte der Digitalisierung beziehungsweise künstlichen Intelligenz als eine Bedrohung, die unsere Lebenswelt und die Gesellschaft verändern wird. Hier werden Ängste artikuliert, die Verluste in vielerlei Hinsicht umfassen, seien es Arbeitsplätze, Werte oder stabile politische Strukturen. Auch wird befürchtet, dass sich das Meister-Sklave-Modell zukünftig umkehrt und wir am Ende diejenigen sind, die sich den künstlichen Systemen anpassen oder gar unterordnen müssen (Singularität). Nicht unberechtigt werden in diesem Meta-Narrativ auch die Risiken der digitalen Überwachung und Manipulation angesprochen, die in der Befürchtung eines Überwachungskapitalismus<sup>22</sup> oder eines Überwachungsstaats nach dem Vorbild Chinas<sup>23</sup> zum Ausdruck gebracht werden. Chinas digitale Staatsdoktrin erzählt hingegen ein anderes Narrativ: das des *Weltführers*, der mittels der KI-getriebenen Dienste seine globale Vormachtstellung erreichen und den Wohlstand seines Volkes sichern will. Dass China hierfür gute Voraussetzungen hat, ist nicht von der Hand zu weisen, wie Kai-Fu Lee<sup>24</sup> argumentiert.

Die Erzählmodelle der US-amerikanischen Tech-Giganten und der Überwachungskritiker\*innen stehen sich diametral entgegen. Zudem zeigen sie erzählstrukturell die radikalste Form der Ereignishaftigkeit: Hier wird das Ereignis nicht durch einen Akteur hervorgerufen, der die Grenze (eines semantischen Raums) überschreitet, sondern die semantische Ordnung selbst wird transformiert. In der Silicon-Valley-Erzählung soll sich die Welt der defizitären Menschheit, die sich nicht selbst optimieren und nicht mit Krisen adäquat umgehen kann, durch die KI-Technologie transformieren, um fittere Menschen zu schaffen, die über die biologischen Grenzen hinauswachsen. Ziel ist eine technizistische Weltverbesserung. Technik könne die biologische Evolution ersetzen (vgl. Abb. 2).

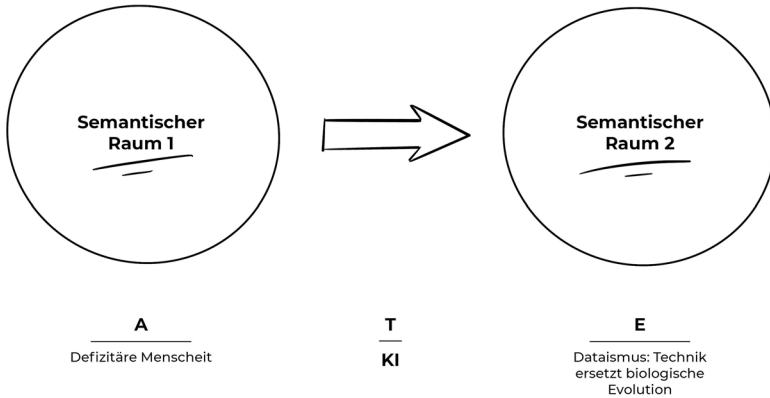
21 Harari, Yuval Noah: *Homo Deus. Eine Geschichte von Morgen*, München: C.H. Beck 2017, S. 497.

22 Vgl. Zuboff, Shoshana: *Das Zeitalter des Überwachungskapitalismus*. Aus dem Englischen von Bernhard Schmid, Frankfurt/New York: Campus Verlag 2018.

23 Vgl. Strittmatter, Kai: *Die Neuerfindung der Diktatur: Wie China den digitalen Überwachungsstaat aufbaut und uns damit herausfordert*, München: Piper 2018.

24 Lee, Kai-Fu: *AI Super Powers. China, Silicon Valley and the New World Order*, Boston, MA: Houghton Mifflin Harcourt 2018, S. 18.

Abbildung 2: Meta-Narrativ des Silicon Valley



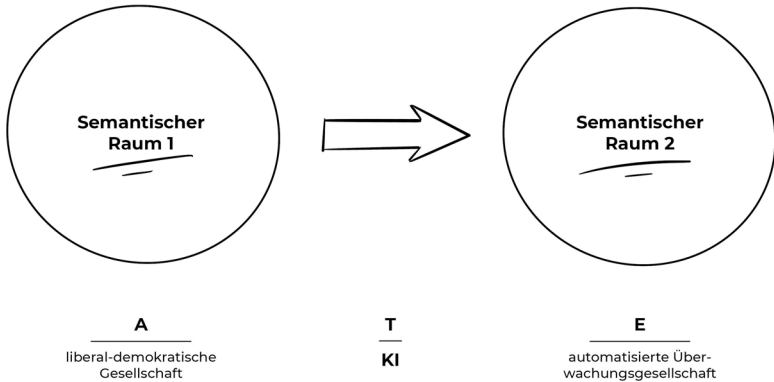
Die semantischen Räume dieses Erzählmodells transportieren implizit einen Wertewandel: Der Wert des Menschen als autonomes Individuum in seiner leiblichen Verfasstheit und Verletzlichkeit, die alle Menschen verbindet, ist obsolet. Die Gesellschaft wird nicht mehr durch liberal-demokratische Institutionen, sondern durch digitale Technologien vorangebracht. Vertreter des Transhumanismus, wie zum Beispiel der Chef-Techniker von Google, Ray Kurzweil, legen dieses Narrativ ihrer Ideologie zugrunde.<sup>25</sup> Er sowie die Anhänger\*innen der Singularitäts-These gehen davon aus, dass wir noch in diesem Jahrhundert an den Punkt kommen werden, wo KI die menschliche Intelligenz nicht mehr benötigt, immer schnellere Weiterentwicklungen ihrer selbst beginnt und das Universum mit einem neuen Geist erfüllen wird. Das Erzählmodell der Überwachungskritiker\*innen dagegen entwickelt keine technizistische Utopie, sondern die Dystopie einer automatisierten Überwachungsgesellschaft, die die Auflösung einer liberal-demokratischen Gesellschaft bedeuten würde (vgl. Abb. 3).

In Geschichten gibt es nach dem Modell von Algirdas Greimas<sup>26</sup> ein dynamisches Beziehungsgefüge der funktionalen Aktanten: Held, Wunschobjekt,

25 Kurzweil, Ray: Menschheit 2.0. Die Singularität naht, Berlin: Lola Books 2014.

26 In der deutschen Übersetzung zit. n. Keller, Otto/Hafner, Heinz: Arbeitsbuch zur Textanalyse. Semiotische Strukturen, Modelle, Interpretationen, München: Wilhelm Fink Verlag 1986, S. 87.

Abbildung 3: Meta-Narrativ der Überwachungskritiker\*innen

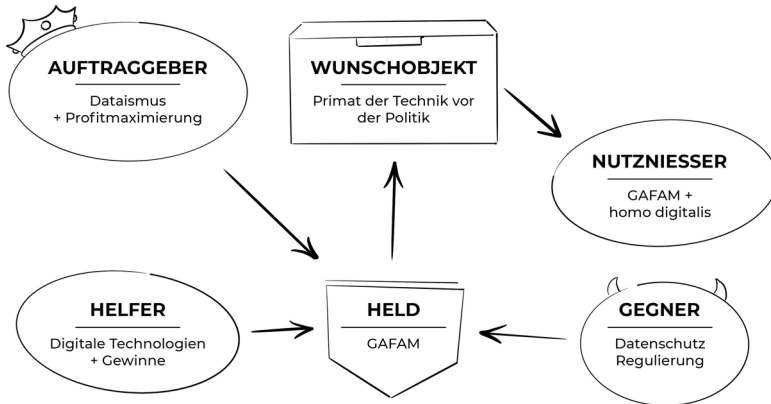


Helfer, Gegner, Auftraggeber und Nutznießer. Diese müssen keine menschlichen Akteure sein, es können auch abstrakte Instanzen sein, wie zum Beispiel das *Gewissen* als Auftraggeber oder Institutionen in der Heldenfunktion. Vor allem die dynamischen Beziehungen zwischen dem Helden und seinem Wunschobjekt sowie seinem Helfer beziehungsweise Gegner konstituieren eine Geschichte. Das Modell ist deshalb so attraktiv, weil es universell anwendbar ist und die funktionalen Beziehungen in Narrativen verdeutlicht, wie am Beispiel des Silicon-Valley-Narrativs zu zeigen ist (vgl. Abb. 4). Hier sind die Tech-Unternehmen in der Position des Helden, sie treiben die Geschichte an: Was ist ihr Wunsch? Sie wollen das Primat der Technik vor der einhegenden Politik und dem Recht. Die digitalen Technologien sowie die immensen Gewinne, die sie durch die Digitalwirtschaft erzielen, helfen ihnen dabei. Das einzige, was sie derzeit bremsen könnte, ist eine stärkere Regulierung. Aus Sicht der Tech-Unternehmen würden die Menschheit und natürlich sie selbst davon profitieren, wenn anstelle von demokratischen Institutionen die Technik vorgeben würde, wie man sich zu entscheiden habe. Was sie antreibt, ist ihre Ideologie und der Wunsch nach Profitmaximierung.

## 1.6 Eine alternative Erzählung – der digitale Mittelweg

Was wir für eine menschengerechte Gestaltung der künstlichen Systeme jedoch brauchen, ist aus meiner Sicht eine *alternative Erzählung*. Ich nenne sie

Abbildung 4: Aktantenstruktur des Silicon-Valley-Narrativs



die vom *goldenen/digitalen Mittelweg*. Dieses Narrativ der *aurea mediocritas* findet sich schon in der Antike bei Horaz: »In schwieriger Lage zeige dich beherzt und tapfer; du wirst aber auch weise die von allzu günstigem Wind geschwellten Segel einziehen.«<sup>27</sup> Für Aristoteles ist der goldene Weg der Mitte – das richtige Maß (*Mesotes*) – ein wesentliches Prinzip seiner Tugendethik. Die Tugend liegt demnach zwischen zwei Extremen, einem Übermaß und einem Mangel, zum Beispiel die Tugend der Freigebigkeit als Mitte zwischen den Extremen der Verschwendungssucht und des Geizes, die Tugend der Besonnenheit als Mitte zwischen Gedankenlosigkeit und Rigorismus. Weder ein Zuviel noch ein Zuwenig führt zum guten Leben. Dieses kann aber nicht im Eigennutz gefunden werden, sondern im sozialen Handeln und in der Entwicklung einer immer wieder zu überprüfenden Haltung. Je nach Disposition des Menschen und dem Kontext müssen ein Zuviel oder ein Zuwenig unterschiedlich definiert werden. Dieser Relationalismus macht den Weg der goldenen Mitte so attraktiv, da es somit kein statisches, sondern ein dynamisches Prinzip des sozialen Handelns meint. Gleichwohl ist Aristoteles' Lehre von der Tugend als Mitte in seiner Konzeption mehrfacher Kritik ausgesetzt. Bereits Kant, aber auch modernen Philosoph\*innen gilt sie als »dunkel und

27 Horaz: Oden und Epoden. Herausgegeben und übersetzt von Gerhard Fink, Düsseldorf/Zürich: Artemis und Winkler Verlag 2002, S. 105.

zugleich als entweder leer oder nicht anwendbar oder nicht hilfreich.«<sup>28</sup> Allerdings wird Kant dem aristotelischen Konzept nicht gerecht, wenn er es als Aufruf zur Mäßigung beziehungsweise zum Mittelmaß missinterpretiert.<sup>29</sup> Haltbar im Kern erscheint das *Mesotesprinzip*, soweit es sich um Tugenden im Bereich der Affektivität und des Reflektierens handelt (beispielsweise *Mut* anstelle von *Übermut* oder *Verzagtheit*), nicht aber bei Tugenden des Willens, wie Gerechtigkeit und Nächstenliebe. Denn was wäre ein Zuviel oder Zuwenig in Bezug auf Gerechtigkeit oder Nächstenliebe? »Die Mesoteslehre lässt sich also nur auf die Tugenden des individuellen Lebens, auf Besonnenheit und Standhaftigkeit und ihre Unterformen anwenden.«<sup>30</sup> Abstrahiert man aus dem tugendethischen Mesotes-Ansatz drei strukturelle Aspekte, nämlich den der triadischen Struktur (die Mitte zwischen zwei nicht wünschenswerten Extremen), den Aufruf zum Balance-Finden sowie das dynamische Prinzip des sozialen Handelns in den jeweiligen Kontexten, dann erscheint der Weg der goldenen Mitte durchaus als ein für die (angewandte) Digitale Ethik<sup>31</sup> sinnvolles Reflexionsmodell.

Eine gegenwartsnahe Lesart des goldenen Mittelwegs im digitalen Zeitalter könnte dazu verhelfen, sich eine Haltung im Umgang mit digitalen Technologien anzueignen, die weder ein Zuviel noch ein Zuwenig bedeutet, wie es vor allem beim Erwerb einer Digitalkompetenz relevant ist: Weder Resignation noch Technik-Euphorie helfen im Umgang mit digitalen Diensten. Aber auch bei Entscheidungen in Unternehmen und Organisationen kann ein Austarieren eines Zuviels und eines Zuwenigs ein probater Weg sein, um sich nachweislich als wertorientiert und glaubhaft zu positionieren, zum Beispiel indem anstelle von haltloser Profitmaximierung oder Ethics-Washing (Ethik als Feigenblatt) eine ernst gemeinte, verantwortungsvolle Unternehmenskultur angestrebt wird.

Letztlich lässt sich das Meta-Narrativ eines (zukünftigen?) europäischen Wegs bei der Digitalisierung als ein Modell der goldenen Mitte zwischen dem Extrem eines datenkapitalistischen US-amerikanischen Modells und eines chinesischen Überwachungsstaats entwerfen. Stellt man dieses Modell des

28 Wolf, Ursula: *Über den Sinn der Aristotelischen Mesoteslehre*, in: Dies.: *Handlung, Glück, Moral. Philosophische Aufsätze*, Berlin: Suhrkamp 2000, S. 111.

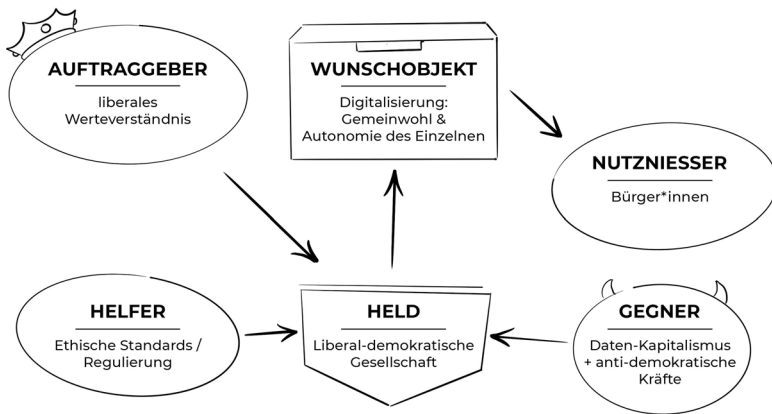
29 Vgl. die Argumentation in Halbig, Christoph: *Der Begriff der Tugend und die Grenzen der Tugendethik*, Berlin: Suhrkamp 2013, S. 180-181.

30 Wolf, Ursula: *Über den Sinn der Aristotelischen Mesoteslehre* 2000, S. 133.

31 Vgl. Grimm, Petra/Keber, Tobias/Zöllner, Oliver: *Digitale Ethik. Leben in vernetzten Welten*, Dietzingen: Reclam 2019.

goldenen Mittelwegs in seiner Handlungsstruktur dar, ergibt sich folgendes narratives Beziehungsgefüge (vgl. Abb. 5): Die liberal-demokratische Gesellschaft befindet sich nun in der Position des Helden. Ihr Ziel ist eine digitale Gesellschaftsordnung, die sich an einem (durch demokratische Prozesse begründeten) Gemeinwohl und dem Autonomieprinzip orientiert und die auf den Grundrechten und Werten einer freien Gesellschaft beruht. Ihr Auftraggeber, also das, was sie antreibt beziehungsweise motiviert, ist ihr liberal-demokratisches Werteverständnis. Ethische Standards und Regulierung sind in dieser Erzählung nicht innovationshemmend, sondern vielmehr Helfer, die sie auf ihrem Weg unterstützen. Gegner sind die Vertreter des Datenkapitalismus und antidemokratische Kräfte im nationalen und internationalen Raum.

Abbildung 5: Aktantenstruktur des goldenen Mittelwegs



Vom Erfolg dieser Heldengeschichte würden vor allem die Bürger\*innen profitieren. Ob diese Geschichte reine Illusion bleibt, ist zum gegenwärtigen Zeitpunkt offen. Möglicherweise kann sie das Wertevakuum einer rein marktgetriebenen und/oder technizistischen KI-Erzählung auffüllen. Die Wirkmächtigkeit von Erzählungen war immerhin in der Vergangenheit oftmals ein Antriebsmotor für kulturellen Wandel.

## 2 Werte aus digitaletischer Sicht

Nach dem Versuch einer Systematisierung und Einordnung der Wertefragen des Digitalen in individuelle, soziale, gestalterische und diskursbezogene Spannungsfelder soll nun der Frage nachgegangen werden, welche Werte im digitalen Zeitalter besondere Aufmerksamkeit verdienen. Zuvor ist aber zu klären, was unter Werten zu verstehen ist.

Es entbehrt nicht einer gewissen Ironie, dass Aristoteles der Ökonomie (*oikonomia*) ihren Namen gab, der sie im Übrigen als ethische Wissenschaft verstand, während der Begriff *Wert* und dessen Gebrauch ursprünglich aus dem Bereich der Wirtschaft stammt.<sup>32</sup> Sein Begriffswandel führte über die »[...] Philosophie des 19. Jahrhunderts und von dort zu den Kultur- und Sozialwissenschaften und dem öffentlichen Sprachgebrauch des 20. Jahrhunderts[...]«<sup>33</sup>. Die Vieldeutigkeit des Wertbegriffs erleichtert es nicht gerade, seine Bedeutung für das Digitale zu erfassen. Hilfreich ist es, sich vorab zu vergegenwärtigen, was mit dem Begriff *Wert* gemeint ist und welche Funktionen Werte haben können.

Werte lassen sich in einer ersten Annäherung in dreierlei Weise unterscheiden: »(1) ›Etwas ist ein Wert‹ – hier wird ein bestimmtes Etwas (ein Gegenstand, eine Person, eine Haltung, ein Zustand etc.) als Wert an sich betrachtet, als ein ›Objektwert‹, der als Gut zu respektieren ist. (2) ›Etwas hat einen Wert‹ – hier wird einem bestimmten Etwas im Blick auf einen Bewertungsmaßstab ein Wert zugesprochen (zum Beispiel ein materieller Wert, ideeller Wert etc.). Schließlich spricht man (3) davon, daß Handlungen bestimmten Werten verpflichtet sein können, die dann den Sinn dieser Handlungen allererst ausmachen, die Handlungen als Handlungen allererst begründen.«<sup>34</sup> Die beiden erstgenannten Sichtweisen legen den Fokus auf das *Sein* der Werte und die *Bewertungsmaßstäbe*. Letztere Sichtweise versteht Werte als handlungsleitend und legt den Fokus auf die *Realisierung* der Werte und ist für das vorliegende Buch größtenteils maßgeblich.

Aus deskriptiver Sicht lassen sich Werte als Vorstellungen, Ideen oder Ideale verstehen. Werte bezeichnen, was wünschenswert ist – sie sind be-

32 Vgl. Anwander, Norbert: *Wert/Werte*, in: Stefan Gosepath/ Wilfried Hinsch/ Beate Rössler (Hg): *Handbuch der politischen Philosophie und Sozialphilosophie*. Bd. 2, Berlin: Walter de Gruyter 2008, S. 1474.

33 Joas, Hans: *Die Entstehung der Werte*, Frankfurt a.M.: Suhrkamp 1999, S. 37.

34 Hubig, Christoph: *Technik- und Wissenschaftsethik. Ein Leitfaden*, Berlin/Heidelberg/New York: Springer 1993, S. 133-134.

wusste oder unbewusste Orientierungsstandards und Leitvorstellung. In der soziologischen und psychologischen Werteforschung werden den Werten bestimmte Funktionen zugeschrieben: Werte können eine Steuerungsfunktion von Handlungen und Verhaltensweisen innehaben.<sup>35</sup> Ebenso können sie die Wahrnehmung der Welt und deren Beurteilung beeinflussen: »Wert wird [...] als ein inneres beziehungsweise internalisiertes Konzept verstanden, das mitbestimmt, wie wir die Welt sehen und uns in ihr verhalten.«<sup>36</sup> Nach Robert Reichardt<sup>37</sup> beeinflussen Werte die Motive der oder des Einzelnen und sind inhaltlich mit einem hohen Allgemeinheits- beziehungsweise Abstraktionsgrad ausgestattet, was eine semantische Vagheit des jeweiligen Wertes impliziert; tendenziell sind sie für größere Bevölkerungsgruppen maßgeblich. Zusammengefasst lassen sich aus deskriptiver Sicht Werte als abstrakte Wünsche und Vorstellungen verstehen, die im Wesentlichen drei Funktionen haben: Sie können Handlungen (1) steuern, sie sind an unserer (2) Wirklichkeitskonstruktion beteiligt und sie stellen (3) Handlungsgründe beziehungsweise leitende Motive dar.

Wenn es um die Frage geht, warum welche Werte gelten sollen (normativer Wertbegriff), geben ethische Ansätze unterschiedliche Antworten: Für Wertobjektivist\*innen bestehen Werte an sich und für sich. Demgemäß hätte zum Beispiel die Natur einen Wert in sich selbst, unabhängig vom Menschen, der ihr diesen Wert zuschreibt. Für Wertsubjektivist\*innen sind Werte hingegen abhängig von der subjektiven Wertschätzung, zum Beispiel davon, ob die Natur zum Wohl des Menschen erhalten werden soll.<sup>38</sup> Davon wiederum lassen sich Ansätze unterscheiden, nach denen Werte entsprechend dem Prinzip der praktischen Vernunft beziehungsweise aufgrund von rationalen Bewertungsmaßstäben gelten,<sup>39</sup> zum Beispiel sollte die Natur intakt bleiben, weil sie die Grundlage für ein Fortbestehen der Menschen sichert. Aus Sicht einer Wertethik prägen Werte Weltanschauungen, sind identitätsstiftend in

35 Vgl. Scholl-Schaaf, Margret: *Werthaltung und Wertesystem. Ein Plädoyer für die Verwendung des Wertkonzepts in der Sozialpsychologie*, Bonn: Bouvier 1975, S. 58.

36 Oerter, Rolf: *Struktur und Wandlungen von Werthaltungen*, München/Basel: Oldenbourg 1970, S. 115.

37 Reichardt, Robert: *Wertstrukturen im Gesellschaftssystem – Möglichkeiten makrosoziologischer Analysen und Vergleiche*, in: Helmut Klages/Peter Kmieciak (Hg.): *Wertwandel und gesellschaftlicher Wandel*, Frankfurt/New York: Campus Verlag, 1979, S. 24.

38 Vgl. Birnbacher, Dieter: *Analytische Einführung in die Ethik*, Berlin/New York: Walter de Gruyter 2003, S. 278.

39 Vgl. Anwander, Norbert: *Wert/Werte*, S. 1475.

Bezug auf das Individuum und die Gemeinschaft und steuern unser Denken und Handeln.

## 2.1 Werterelativismus, Wertepluralismus und Werteuniversalismus

Bei dem Versuch, eine werteorientierte Digitalisierung zu beschreiben, wird man sehr schnell mit dem Einwand des Werterelativismus und Wertepluralismus konfrontiert. Das Argument, es gebe keine universalen Werte und deshalb könne man auch keine Werte für global wirksame Technologien ausmachen, scheint das Vorhaben ins Leere laufen zu lassen. Der Einwand ist schon deshalb ernst zu nehmen, weil digitale Technologien ja in unterschiedlichen Gemeinschaften, Kulturen etc. zum Einsatz kommen. Dem ist allerdings zu entgegenen, dass es bei diesem Vorhaben nicht um deskriptive Wertbindungen geht, die in einer Gemeinschaft, einer Kultur etc. vorhanden sind und die ja durchaus ethisch fraglich sein können, wie zum Beispiel der Wert der Jungfräulichkeit. Vielmehr sind Werte gemeint, die als normative Standards kulturübergreifend für verbindlich erachtet werden, auch wenn deren Realisierung kulturspezifisch ist. Trotz kultureller Unterschiede in der Gewichtung von Werten gibt es Werte, die alle Menschen betreffen. Beispielhaft hierfür sind diejenigen Werte, die durch die in der Allgemeinen Erklärung der Menschenrechte festgeschriebenen Grundrechte deutlich werden. Diese Werte gilt es zu berücksichtigen, ohne sie gegeneinander aufzuwiegen oder aufgrund von vermeintlichen Sachzwängen oder kulturellen Empfindlichkeiten außer Kraft zu setzen. Werte können zwar verletzt werden, aber deshalb verlieren sie zumindest nicht umgehend ihre normative Kraft. Gegen den Werterelativismus, der eine Verbindlichkeit von Werten nur systembezogen für möglich hält, oder den Wert nihilismus, der die Existenz moralischer Werte an sich negiert, wehrt sich der Werteuniversalismus, wie ihn zum Beispiel Markus Gabriel offensiv vertritt. Demgemäß gelten moralische Werte – wenn auch in einer minimalistischen Form – universal: »Der Universalismus ist das Gegenteil des Relativismus: Er behauptet, dass moralische Werte unabhängig von Gruppenzugehörigkeiten und damit für alle Menschen (und letztlich sogar über den Menschen hinaus) gelten. Es gibt also nur ein einziges System universaler Werte: das Gute, das Neutrale und das Böse.«<sup>40</sup> Gleichwohl weist Gabriel darauf hin, dass in den lebenspraktischen Gegebenheiten »eine prinzipiell niemals vollständig zu überwindende Kluft zwischen den universalen

40 Gabriel, Markus: *Moralischer Fortschritt in dunklen Zeiten*, Berlin: Ullstein 2020, S. 104.

Werten und ihren Anwendungsbedingungen in komplexen Handlungsfeldern und einzelnen Situationen«<sup>41</sup> bestehe.

Dieser Hinweis ist auch für das Vorhaben einer wertegestalteten Digitalisierung wichtig. Hierzu ein Beispiel aus dem globalen Wirtschaftsleben, bei dem es um den Wert der Privatheit geht: Während sich Mitarbeiter\*innen in Mexiko die Veröffentlichung der Geburtstage im Unternehmen wünschen, um Geburtstagsgrüße aussprechen zu können, ist dies in Deutschland gerade nicht der Fall. Das heißt aber nicht, dass deshalb der Wert der Privatsphäre für die mexikanischen Mitarbeiter\*innen nicht wichtig wäre. Für ein global agierendes deutsches Unternehmen, das sich zur Einhaltung von *Privacy* (nicht zuletzt wegen der europäischen Datenschutz-Grundverordnung) verpflichtet hat, besteht also die Herausforderung, den unterschiedlichen Sitten der Länder bei der Realisierung von Werten Rechnung zu tragen. Privatheit ist ein universaler Wert, auch wenn er kulturspezifisch ausgelegt wird und in der Realisierung kontextabhängig ist. »Als privat gilt etwas dann, wenn man selbst den Zugang zu diesem ›etwas‹ kontrollieren kann.«<sup>42</sup> Privatheit ist somit ein universal gültiger und gleichzeitig in der Anwendung relativer Begriff, abhängig von dem jeweiligen Kontext und kulturspezifischen Handlungspraxen. Der Anspruch, sich auf universale Werte bei künstlichen Systemen zu einigen, findet sich im Übrigen auch in den ethischen Leitlinien des weltweit größten Ingenieur\*innenverbands IEEE.<sup>43</sup>

Nach diesen Ausführungen zum Wertbegriff und den verschiedenen Perspektiven auf Werte stellt sich die Frage, welche Werte aus Sicht einer Digitalen Ethik besonders in den Blick zu nehmen sind. Vorgeschlagen wird hierzu eine Topografie der Digitalen Ethik (vgl. Abb. 6). Sie stellt in einer Übersicht beispielhaft zehn Werte dar, die als Orientierungsorte für eine zukünftige Entwicklung und Gestaltung einer digitalen Welt dienen sollen. Diese Liste ist selbstredend nicht abgeschlossen und dient lediglich zur Anregung eines Überblicks. Auch wenn heute nicht klar ist, wo die Reise mit einer zunehmenden Vernetzung, Automatisierung, künstlicher Systeme, Robotik etc. hingeht, und die Folgen der Entwicklung für die einzelne Person und die Gesellschaft nicht absehbar sind, lassen sich doch Risikofelder erkennen, denen wir auf

41 Ebd., S. 141.

42 Rössler, Beate: *Der Wert des Privaten*, Frankfurt a.M.: Suhrkamp 2001, S. 23

43 The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE, 2019. <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>, S. 10.

dem Weg einer fortschreitenden Digitalisierung begegnen. Würden wir die im Folgenden genannten Werte und Risiken ignorieren beziehungsweise negieren, würde eine liberal-demokratische Gesellschaft ihr Fundament verlieren. Als Grundlage dieser Werte-Topografie dienen ethische Leitlinien im Bereich der Mensch-Maschine-Interaktion<sup>44</sup> und zur künstlichen Intelligenz,<sup>45</sup> die Charta der digitalen Grundrechte der Europäischen Union,<sup>46</sup> das Gutachten der Datenethikkommission der Bundesregierung<sup>47</sup> sowie eigene Arbeiten zur Verankerung von Ethik in der integrierten Forschung<sup>48</sup>.

## 2.2 Werte einer Digitalen Ethik

*Menschenwürde:* Die Würde des Menschen ist auch in einer digitalen Welt unantastbar und es besteht eine Pflicht, diese zu achten und zu schützen; die körperliche und geistige Unversehrtheit des Menschen darf nicht durch digitale Technologien beeinträchtigt werden.

*Autonomie:* Menschen sollen nicht als Mittel zum Zweck dienen; Menschen dürfen nicht künstlichen Systemen ausgeliefert sein, die Entscheidungen über Leben, körperliche und geistige Unversehrtheit sowie Freiheitsentzug fällen; moralische Prinzipien dürfen nur von Menschen aufgestellt werden; künstliche Systeme dürfen nicht ohne Kenntnis der betroffenen Menschen zur Manipulation des Verhaltens eingesetzt werden; Nutzer\*innen digitaler Produkte und Dienste müssen autonom zwischen diesen wechseln können.

44 Vgl. u.a. Grimm, Petra/Mönig, Julia (2020): KoFFI-Code: Ethische Empfehlungen des BMBF-Projekts KoFFI. Kooperative Fahrer-Fahrzeug-Interaktion, Stuttgart 2020. [https://www.hdm-stuttgart.de/digitale-ethik/forschung/forschungsprojekte/abgeschlossene\\_forschungsprojekte/KOFFI/KoFFi\\_PDF\\_inkl.Korrekturen\\_V4\\_Druck-mit-Schnittmarken.pdf](https://www.hdm-stuttgart.de/digitale-ethik/forschung/forschungsprojekte/abgeschlossene_forschungsprojekte/KOFFI/KoFFi_PDF_inkl.Korrekturen_V4_Druck-mit-Schnittmarken.pdf)

45 Vgl. das Verzeichnis der ethischen Richtlinien für künstliche Intelligenz von AlgorithmWatch: AI Ethics Guidelines Global Inventory. <https://inventory.algorithmwatch.org/>

46 Die Autorin ist Mitunterzeichnerin dieses politischen Manifests. <https://digitalcharta.eu/#post-22838>

47 Datenethikkommission der Bundesregierung: Gutachten der Datenethikkommission der Bundesregierung 2019. [https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?\\_\\_blob=publicationFile&v=6](https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6)

48 Vgl. zuletzt die Entwicklung einer Software für Forscher\*innen, um ethische, rechtliche und soziale Kriterien in Forschungsprojekten zum Thema Mensch-Maschine-Interaktion und künstliche Intelligenz zu verankern: »Automatisiertes ELSI-Screening & Assessment-Tool für MTI-Forschungsvorhaben (ELSI-SAT)«, <https://www.elsi-sat.de/>

Abbildung 6: Werte-Topografie



**Privatheit:** Die Privatsphäre der Menschen ist zu achten; jeder Mensch hat das Recht, seine Daten schützen zu können und nicht überwacht zu werden; das Prinzip ist zu achten, dass Individuen in einem Rechtsstaat auch Rechte zum Selbstschutz gegenüber der Staatsmacht haben und behalten müssen; personenbezogene Daten dürfen nur für festgelegte Zwecke bei den Betroffenen erhoben und verarbeitet werden, wenn hierfür eine gesetzliche Grundlage besteht; die Privatsphäre von Kindern, Heranwachsenden und schutzbedürftigen Menschen ist besonders zu achten; die Löschung und Berichtigung von persönlichen Daten, Widerspruch, Information und Auskunft müssen gewährleistet sein; jeder Mensch hat das Recht auf eine nicht-personalisierte Nutzung digitaler Angebote, deren Einschränkung darf nur auf gesetzlicher Grundlage stattfinden.

**Freiheit:** Jeder Mensch hat ein Recht auf freie Information und Kommunikation sowie auf freie Meinungsäußerung in digitalen Medien; das Recht auf Meinungsfreiheit findet seine Schranken in den Vorschriften der allgemeinen Gesetze; jeder Mensch hat das Recht auf freien und gleichen Zugang zu

Kommunikations- und Informationsdiensten; die Medien- und Pressefreiheit ist zu garantieren.

*Transparenz und Erklärbarkeit:* Die Kriterien automatisierter Entscheidungen sind offenzulegen; die Datenverarbeitung muss transparent und nach dem Stand der Technik gestaltet werden; Anwender\*innen künstlicher Systeme sollen deren Funktionsweise verstehen, erklären und kontrollieren können; die von einer Entscheidung Betroffenen sollen entsprechend ihrem Kompetenzniveau genügend Informationen erhalten, um ihre Rechte angemessen wahrnehmen und die Entscheidung infrage stellen zu können; es sind Methoden zur Nachvollziehbarkeit der Ergebnisse von künstlichen Systemen zu fördern; einer Vermenschlichung von künstlichen Systemen, die falsche und in der Praxis stark risikobehaftete Annahmen über eine Systemleistung erstellen, ist durch geeignete Maßnahmen entgegenzuwirken.

*Gerechtigkeit:* Künstliche Systeme sollen nach fairen Algorithmen gestaltet sein, hierzu müssen sich Programmierer\*innen ihrer Voreingenommenheit, Vorurteile und kulturellen Prägung bewusst werden, zudem muss die Datenbasis auf Diskriminierungspotenziale geprüft werden und es muss versucht werden, systematische Fehler und Verzerrung – auch auf technischem Wege – zu vermeiden; die Prinzipien der Vielfalt und Offenheit sollen in Bezug auf digitale Angebote, Infrastrukturen und Zugänge zu Netzen gelten; durch künstliche Systeme dürfen Menschen nicht vom Zugang zu Gütern, Dienstleistungen oder von der Teilhabe am gesellschaftlichen Leben ausgeschlossen werden; Menschen, die von Entscheidungen künstlicher Systeme in Bezug auf ihre Lebensführung erheblich betroffen sind, sollen einen Anspruch auf unabhängige Überprüfung und Entscheidung durch Menschen haben.

*Nachhaltigkeit:* Produzent\*innen, Anbieter\*innen und Betreiber\*innen digitaler Produkte, Dienste und Infrastrukturen sollen nachhaltig wirtschaften; sie sollen dafür sorgen, dass künstliche Systeme für eine Emissionsreduktion und Ressourceneinsparung dienen sowie den Konsum nicht weiter anheizen; die Digitalisierung der Lebensräume (smarte Städte, Häuser, autonome Fahrzeuge etc.) soll nachhaltig gestaltet werden; der Lebenszyklus digitaler Geräte und vor allem deren Software sollen verlängert und deren Produktion und Entsorgung mitgedacht werden.

*Verantwortung:* Die Einflüsse der Technik auf die Handlungen, das Denken und die Gewohnheiten des Menschen sollten Entwickler\*innen und Betreiber\*innen künstlicher Systeme bewusst sein, denn Technikgestaltung ist auch Gesellschaftsgestaltung; Betreiber\*innen meinungsbildungsrelevanter Plattformen tragen Verantwortung für die Verbreitung von Inhalten, die dar-

auf abzielen, antisoziales Verhalten zu fördern und den sozialen Zusammenhalt zu untergraben, wie zum Beispiel durch Online-Gewalt, Desinformation und Manipulation; einer zunehmenden Verantwortungsdiffusion der Akteure bei künstlichen Systemen ist entgegenzuwirken; Entscheidungen künstlicher Systeme müssen von natürlichen und juristischen Personen verantwortet werden; Nutzer\*innen müssen durch Bildung und Digitalkompetenz dabei unterstützt werden, digitale Dienste zur Kommunikation, Information und Unterhaltung verantwortungsvoll zu nutzen.

*Wahrheit und Wahrhaftigkeit:* Wahrheit und Wahrhaftigkeit sind substanzielle Werte einer gut informierten Gesellschaft; Betreiber\*innen digitaler Dienste und Plattformen müssen dafür sorgen, dass alle Formen absichtlich erstellter und weitergegebener Desinformationen (zum Beispiel Verschwörungserzählungen, sogenannte Fake News, manipulierte Fotos und Videos), die das Publikum täuschen und irreführen sollen, nicht verbreitet werden; der Gefahr, durch Social Bots (Computerprogramme, die mit eigenen Accounts und Profilen das Verhalten von echten Nutzer\*innen simulieren) bestimmte Meinungen massenhaft zu verbreiten und die Stimmung in sozialen Medien zu lenken, ist entgegenzuwirken.

*Sicherheit:* Die Unversehrtheit, Vertraulichkeit und Integrität künstlicher Systeme und Infrastrukturen ist sicherzustellen und angemessen technisch und organisatorisch zu gewährleisten; Systemkompatibilität und die Handhabung der Technik unter realen Bedingungen sollten bereits in der Entwicklungsphase beachtet werden, um keine unnötigen Fehlerquellen entstehen zu lassen; es sollte fortlaufend überprüft werden, wie sich die Zwecke verändern, für welche die Technik entwickelt wird; es sollte mitgedacht werden, wie die Technik zweckentfremdet angewandt werden kann und ob dies bestimmte vorsorgliche Gegenmaßnahmen notwendig macht; Unternehmen sollen auch nicht vor der Überlegung zurückschrecken, unter Umständen bestimmte Möglichkeiten zu beschränken.

*Demokratiieverträglichkeit:* Der Fortschritt in der Digitalisierung darf nicht die Grundrechte und demokratischen Prinzipien aushebeln; die durch die Digitalisierung hervorgerufenen Machtverschiebungen zwischen einzelner Person, Staat und Unternehmen sind einer kritischen Reflexion zu unterziehen; antidemokratischen Kräften, die die Digitalisierung zu ihren Gunsten nutzen, müssen politische Entscheidungsträger\*innen durch die Förderung aufklärender Bildungsangebote, Forschung und zivilgesellschaftlichen Engagements konsequent entgegenreten.

*Vertrauen*: Die Gestaltung von Vertrauen und der Aufbau von vertrauensvollen Beziehungen zwischen Menschen und künstlichen Systemen sind die wesentlichen Voraussetzungen für die Akzeptanz neuer Technologien und für eine erfolgreiche Mensch-Maschine-Interaktion; ein zu eng gefasster Vertrauensbegriff, der sich nur auf eine simple Berechenbarkeit bezieht, ist für das sensible Thema *Vertrauen* in der Mensch-Maschine-Interaktion problematisch; Vertrauen in künstliche Systeme kann nur erzielt werden, wenn die oben genannten zehn Werte als normative Standards umgesetzt werden.

### 2.3 Risiken

Mit der Beschreibung der Werte-Topografie, die eine Konkretisierung der Werte in normative Vorgaben beinhaltet, wurde implizit bereits auf verschiedene Risiken eingegangen. Zur Übersicht und in Ergänzung lassen sich folgende Risiken identifizieren, die mit der Entwicklung, Gestaltung sowie ökonomischen und gegebenenfalls politischen Zielsetzung digitaler Technologien verbunden sind: die Risiken der *Manipulation* im Informationsbereich (Stichwort »Desinformation«), der *Machtasymmetrie* von Nutzer\*innen und Anbieter\*innen (Stichworte »intransparente Datenverwertung«, »Abhängigkeit der Nutzer\*innen von Intermediären«) und der *Heteronomie* (fehlende Selbstbestimmung, wenn künstliche Systeme zum Beispiel über Kreditvergabe, Prämien oder Jobs entscheiden). *Überwachung* von staatlicher Seite und durch Wirtschaftsunternehmen ist ein weiterer Risiko-Topos. Neben diesen systembezogenen Risiken, die ökonomisch und politisch begründet sind und regulierbar wären, haben wir es auch mit primär technologiebezogenen Risiken zu tun, die die Notwendigkeit einer Technikfolgenabschätzung, Ethics by Design und einer nachhaltigen Digitalökologie verdeutlichen. Zu letzteren Risiken gehören neben den Problemen des Dual-Use (nicht-intendierte Verwendung von Techniken) und der Bias (Verzerrungen) auch das in der Öffentlichkeit kaum beachtete Thema *klima- und ressourcenschädlicher Digitalsektor*. Die hierzu bislang wenigen Befunde lassen vermuten, dass der Energiebedarf von Netzwerken und Datenzentren zur Speicherung und Auslieferung von Inhalten, beim Streaming von Videos und deren Produktion, nicht zuletzt auch bei der Entsorgung der Geräte für einen rapide wachsenden Anteil der weltweiten Emissionen von Treibhausgasen mitverantwortlich ist (Rebound-Effekt). Gegen die Flut der Ignoranz solcher Risiken können auf ethischer Ebene eine reflektierte Haltung von Entscheider\*innen und Nutzer\*innen, eine ernst gemeinte Selbstverpflichtung von Unternehmen in Form von ethischen Leitli-

nien und die Bereitschaft, Ethics by Design konsequent umzusetzen, hilfreich sein.

Des Weiteren gilt es, die Schnittstellen zwischen Recht und Ethik<sup>49</sup> zu verdeutlichen und diese durch interdisziplinäre Zusammenarbeit fruchtbar zu gestalten. Ethik und Recht sind miteinander verbunden und haben gemeinsame Wurzeln, dennoch existieren beide Disziplinen auch vollkommen unabhängig voneinander. Ethik ist ein grundlegendes Analyseinstrument, das Begründungs- und Bewertungsstrategien aufzeigen kann, die unterschiedliche Normen und Wertesysteme jenseits des geltenden Rechts als Basis haben können. Ethik dient dazu, Motive des Handelns sichtbar und verständlich zu machen und kann aus der Analyse heraus deshalb auch eigene normative Vorgaben entwickeln. Deren Gültigkeit zielt jedoch nur auf die freiwillige Umsetzung.

### 3 Ansätze für eine Digitale Ethik

Aufgabe der Ethik ist es, zu *begründen*, warum etwas gut oder schlecht ist. Ethik kann somit als *Reflexionstheorie der Moral* beschrieben werden, die nach guten Argumenten für das ethische Handeln sucht. Der Standardauffassung zufolge lassen sich drei ethische Ansätze unterscheiden, die je nach Perspektive andere Schwerpunkte der ethischen Reflexion setzen. So lenkt der (1) Konsequentialismus unseren Blick auf den ethischen Nutzen und die Folgen digitaler Technologien, während die (2) deontologische Perspektive (griech. *deon*, »Pflicht«, hier also etwa: eine Sichtweise, die aus Pflichten Aussagen ableitet) uns auf moralische Prinzipien, wie zum Beispiel das der Verantwortung, und auf die Notwendigkeit, Werte abzuwägen, verweist. Die (3) tugendethische beziehungsweise eudämonistische Perspektive (griech. *eudaimonía* »Glück« oder »Glückseligkeit«) fokussiert hingegen die Frage nach dem guten, gelingenden Leben. Wie können diese Ansätze für eine Ethik des Digitalen hilfreich sein? Beginnen wir mit der konsequentialistischen Ethik.

---

49 Siehe hierzu auch den Beitrag von Timo Rademacher und Erik Schilling und den Beitrag von Erik Hilgendorf in diesem Band.

### 3.1 Konsequentialistische Ethik

Für sie ergibt sich das moralisch Richtige einer Handlung aus den *Folgen* einer Handlung, nicht aus der Pflicht, nach bestimmten Prinzipien beziehungsweise moralischen Regeln zu handeln. Ziel des Konsequentialismus ist es, für eine größtmögliche Anzahl von Menschen einen Nutzen zu erzielen – Wohlstand, Glück, Gesundheit und dergleichen. Der wichtigste Vertreter des Konsequentialismus ist der Utilitarismus (von lat. »nützlich«). Für die klassischen Utilitaristen des 19. Jahrhunderts, Jeremy Bentham und John Stuart Mill, sollen Individuen diejenige Handlungsweise wählen, die ihnen dazu verhilft, das größtmögliche Glück zu erreichen. Moderne Vertreter attestieren dem Utilitarismus eine Überlegenheit über deontologische Ansätze, da er anpassungsfähig an den sozialen und technischen Wandel sei und nicht an überholten deontologischen Normen festhalte.<sup>50</sup> Das überzeugt, soweit es um Dogmatismus und die Ignoranz lebensweltlicher Kontexte geht. Andererseits ist das Nützlichkeits- und Maximierungsprinzip sehr gut mit neoliberalen Leistungs- und Effizienzdenken, materiellem Glücksversprechen, Quantifizierung des Selbst und des Sozialen<sup>51</sup> sowie der Ökonomisierung von Wertesystemen<sup>52</sup> kompatibel. Ebenso ist es dem Utilitarismus bereits mit Adam Smith im 18. Jahrhundert gelungen, den individuellen Egoismus als Stärke für das Gemeinwohl zu begründen, wonach das Allgemeinwohl umso größer ist, je stärker die Individuen nach ihren eigenen Interessen handeln: »Nicht vom Wohlwollen des Metzgers, Brauers und Bäckers erwarten wir das, was wir zum Essen brauchen, sondern davon, daß sie ihre eigenen Interessen wahrnehmen.«<sup>53</sup> Auch wenn diese Sichtweise auf den ersten Blick ethisch fraglich erscheint, ist bei näherer Betrachtung nicht zu ignorieren, dass sie auch für die ethische Praxis unternehmerischen Handelns von Bedeutung ist. Denn Anreize für ethisches Handeln werden Unternehmen letztlich immer aus der ökonomischen Logik ableiten: Nur wenn sich Moral rechnet und dabei hilft, zum Beispiel Vertrauen zu schaffen, Mitarbeiter\*innen zu motivieren, Image-Schaden abzuwenden etc., ist sie für Unternehmen attraktiv.

50 Vgl. Birnbacher, Dirk: Analytische Einführung in die Ethik, 2003, S. 174-175.

51 Vgl. Mau, Steffen: Das metrische Wir. Über die Quantifizierung des Sozialen, Berlin: Suhrkamp 2017.

52 Vgl. Grimm, Petra/Zöllner, Oliver (Hg.): Ökonomisierung der Wertesysteme. Der Geist der Effizienz im mediatisierten Alltag, Medienethik Band 14, Stuttgart: Steiner Verlag 2015.

53 Smith, Adam: Der Wohlstand der Nationen, München: dtv 1990 (1776), S. 17.

Was die Leistungskraft des Utilitarismus hinsichtlich einer Ethik der Digitalisierung betrifft, lassen sich Stärken und Schwächen erkennen: 1) Eine Stärke des Utilitarismus ist es, den Blick auf die Folgen der Digitalisierung zu lenken. Mit der Technikfolgenabschätzung rückt die konsequentialistische Ethik in das Zentrum der ethischen Reflexion,<sup>54</sup> auch wenn das nicht bedeutet, dass jede Technikfolgenabschätzung aus der Perspektive einer konsequentialistischen Ethik erfolgt. Auch ethische Leitlinien zu künstlichen Systemen nehmen die Folgen in den Blick, indem sie als Handlungsmaximen benennen, Schaden zu verhindern und Wohlergehen zu fördern; sie schließen aber auch Prinzipien mit ein, wenn beispielsweise die Menschenwürde und die Autonomie gewahrt werden sollen. 2) Schwächen werden dem Utilitarismus andererseits gerade wegen einer mangelnden Folgenabschätzbarkeit bei schwer kalkulierbaren Folgen neuer Technologien attestiert. So lautet die Kritik, dass er nicht aufzeigen könne, *welche* Handlungsweise das größtmögliche Glück oder den geringsten Schaden verspreche.<sup>55</sup> Zudem sei in einer komplexen, nicht stabilen Welt schwer zu entscheiden, was Glück und Unglück bedeuten.<sup>56</sup> 3) Eine Stärke wiederum wird ihm bescheinigt, wenn es um die Frage nach der moralischen Programmierbarkeit von künstlichen Systemen geht. Vorausgesetzt, man hält eine Implementation moralischer Fähigkeiten in künstliche Systeme für vertretbar und sinnhaft, scheint der Utilitarismus ein geeigneter Kandidat für eine etwaige Programmierung von Regeln zu sein. Er gewährleistet »die Maximierung des Gesamtnutzens am besten durch die Aufstellung gewisser Regeln«.<sup>57</sup>

Ist die utilitaristische Ethik auch für die oben erzählten Fallgeschichten, die für die Diagnose der Spannungsfelder dienen, als ethischer Ansatz tauglich? Für die Geschichte (1) von Alex hilft er nicht, denn hier geht es um die Motive und die Wertekonflikte des Protagonisten. Für die zweite Geschichte (2) der Seniorin Ingeburg K. kann er hingegen hilfreich sein. Denn hier wird nach den Folgen der Installation des künstlichen Systems gefragt und danach, ob eine größtmögliche Zahl von Menschen davon profitiert. Der Fokus liegt auf der individuellen und sozialen Situation von Ingeburg K. sowie

54 Vgl. Decker, Michael: *Technikfolgen*, in: Armin Grunwald (Hg.): *Handbuch Technikethik*, Stuttgart/Weimar: J. B. Metzler 2013, S. 36.

55 Vgl. Vallor, Shannon: *Technology and the Virtues. A Philosophical Guide to a Future Worth Wanting*, New York: Oxford University Press 2016, S. 7.

56 Vgl. Boddington, Paula: *Towards a Code of Ethics for Artificial Intelligence*, Cham: Springer 2017, S. 69.

57 Misselhorn, Catrin: *Grundfragen der Maschinenethik*, 2018, S. 101.

der Beteiligten (Pfleger, Zugehörige). Ihr Glück oder Unglück könnte gegebenenfalls aber dem Nutzen der Versicherung und Beitragszahler\*innen nachgeordnet sein. Wie der Nutzen selbst zu bewerten ist und inwieweit das Maximierungsprinzip überhaupt als Bewertungsgrundlage gelten soll, bleibt dahingestellt. Bei der Fallgeschichte (3) von Louise und Ann-Kathrin steht die Frage nach der ethischen Gestaltung eines künstlichen Bewerbungssystems im Vordergrund. Lässt sich in das System Moral implementieren? Das ist bei künstlichen Systemen umstritten. So scheint zwar eine Programmierung, die auf den größten Nutzen ausgerichtet ist, leichter umsetzbar zu sein als eine deontologische Programmierung, aber wollen wir das? Bei dieser Fallgeschichte (wie auch bei vielen Anwendungen von künstlichen Systemen) geht es ja vor allem um moralische Prinzipien wie *Gleichheit* und *Gerechtigkeit*, die mit dem Konsequentialismus häufig schwer begründet werden können. Bias in künstlichen Systemen können zu Ungerechtigkeiten und Benachteiligung führen, dies ist vor allem bei einer *evidenzbasierten* Gerechtigkeit des größten Nutzens der Fall.

Ein Beispiel, das die utilitaristische Perspektive veranschaulicht, ist das bekannte »Moral Machine Experiment« des MIT<sup>58</sup> zum autonomen Fahren, das auf das in der Ethik und im Recht vielfach besprochene Trolley-Problem<sup>59</sup> zurückgeht. Ein Weichensteller muss sich zwischen den Handlungsoptionen Nichts-Tun und Weichen umstellen entscheiden und muss dabei jeweils den Tod mehrerer Menschen abwägen. Aus utilitaristischer Sicht wäre die Aktivionsvorgabe relativ leicht zu programmieren: Demnach wäre die »Lösung« des geringsten Schadens zu favorisieren, also zum Beispiel, dass das autonome Fahrzeug *einen* Fußgänger statt *fünf* Fußgängern überfährt, wenn es keinen anderen Ausweg gibt. So ist beim Utilitarismus »nicht jede Handlung mit gravierend schlechten Folgen [...] unter allen Umständen moralisch verboten, zum Beispiel nicht dann, wenn alle verfügbaren Handlungsalternativen – einschließlich Untätigkeit – noch schlechtere Folgen hätten.«<sup>60</sup> Entspricht das unserem moralischen Empfinden? Anscheinend ja, wie eine aktuelle empirische Studie zum klassischen Trolley-Problem zeigt, wonach 82 Prozent der

58 [moralmachine.mit.edu](http://moralmachine.mit.edu)

59 Vgl. Wenzel, Hans: *Zum Notstandsproblem*, in: ZStW. Zeitschrift für die gesamte Strafrechtswissenschaft 63(1), 1951, S. 47-56.

60 Birnbacher, Dirk: *Utilitarismus*, in: Armin Grunwald (Hg.): *Handbuch Technikethik*, Stuttgart/Weimar: Metzler 2013, S. 154.

Deutschen billigen, den einen Menschen anstelle von fünf zu opfern; interessanterweise entscheiden sich in China nur 58 Prozent dafür.<sup>61</sup> Wäre das Ergebnis dasselbe, wenn die Person die eigene Mutter oder der Partner wäre? Aus deontologischer Sicht könnte jedenfalls die Opferung des einen Menschen nicht gebilligt werden. Diese deontologische Perspektive ist im Übrigen auch bei der bekannten Entscheidung des Bundesverfassungsgerichts zum Flugsicherheitsgesetz von 2007 zum Ausdruck gekommen, bei dem es um die Abschussfreigabe eines von Selbstmordattentätern gekaperten Flugzeugs ging. Eine Abschussfreigabe würde Menschen zu Objekten machen, sie verdinglichen und entrechtlichen, so lautete im Wesentlichen die Argumentation der Verfassungsrichter\*innen. Auch aus Sicht eines digitalen Humanismus, wie ihn Julian Nida-Rümelin und Nathalie Weidenfeld beschreiben, kollidiert der konsequentialistische Ansatz mit dem »Prinzip der Nicht-Verrechenbarkeit«<sup>62</sup>, wonach nicht das Schicksal des einen Menschen mit dem des anderen Menschen verrechnet werden darf. Hierzu gehört auch ein Menschenbild, wonach der Mensch über Willensfreiheit sowie die Fähigkeit verfügt, als Vernunftwesen wertend Stellung zu nehmen und moralische Begründungen abzuwägen: »In einem humanistischen Weltbild ist der Mensch kein Mechanismus, sondern freier (autonomer) und verantwortlicher Akteur in der Interaktion mit anderen Menschen und einer gemeinsamen sozialen und natürlichen Welt.«<sup>63</sup> Zu ergänzen ist hier, dass der Mensch demgemäß nicht als Konstrukt biochemischer Mechanismen, die von genetischen Programmen gesteuert werden, zu verstehen ist. Diesem technizistischen beziehungsweise szientistischen Menschenbild widerspricht das humanistische, indem es die Würde des Menschen, sein körperliches Dasein und sein Miteinander herausstellt, wie Thomas Fuchs betont: »Im Begriff der Menschenwürde, verstanden als der Anspruch auf Anerkennung, den ein menschliches Wesen durch sein leibliches Dasein und Mitsein erhebt, vereinigen sich und gipfeln die Bestimmungen, die ein humanistisches, personales Menschenbild konstituieren.«<sup>64</sup>

61 Vgl. Awad, Edmond et al.: *Universals and variations in moral decisions made in 42 countries by 70,000 participants*, in: PNAS February 4, No. 5, Vol. 117, 2020, S. 2332-2337. <https://doi.org/10.1073/pnas.1911517117>

62 Nida-Rümelin, Julian/Weidenfeld, Nathalie: *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*, München: Piper 2018, S. 68.

63 Ebd., S. 61.

64 Fuchs, Thomas: *Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie*, Berlin: Suhrkamp 2020, S. 8.

### 3.2 Deontologische Ethik

Die deontologische Ethik ist eine an Pflichten orientierte ethische Theorie. Im Kern geht die Pflichtenethik davon aus, dass bestimmte Handlungen geboten oder verboten sind, unabhängig davon, welche Folgen damit verknüpft sind. Bezeichnend für die Deontologie ist der Verallgemeinerungs- beziehungsweise Universalisierungsanspruch von Rechten und Pflichten. Demnach gibt es Rechte, die für alle Menschen bindend sind, beispielsweise die in der Allgemeinen Erklärung der Menschenrechte. Für die Pflichtenethik Kants steht nicht das Glück oder dessen Maximierung, sondern das Richtige (*was soll ich tun?*) im Zentrum seiner Ethik. Sein Universalisierungsanspruch an das Sollen bringt der kategorische Imperativ zum Ausdruck: »Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.«<sup>65</sup> Damit veranlasst er uns, das eigene Handeln über unsere subjektive Sicht hinaus in Hinblick auf das allgemeine Ganze zu betrachten. Kants wirksames Vermächtnis besteht darin, dass er die Autonomie des Menschen als ethische Leitidee formuliert hat. Demnach »bedeutet individuelle Autonomie die Fähigkeit oder das Vermögen, sich selbst die Gesetze geben zu können, nach denen wir handeln und die wir selbst für richtig halten«<sup>66</sup>. Nach seinem Verständnis sind wir »autonom, frei, [...] dann und nur dann, wenn wir auch moralisch sind und handeln.«<sup>67</sup> Autonomie ist an die eigene praktische Vernunft gebunden sowie an die Achtung der Würde des Menschen (die eigene und die des anderen): »Handle so, dass du die Menschheit, sowohl in deiner Person als in der Person eines jeden anderen, jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst.«<sup>68</sup> Wengleich Kants moralischer Autonomiebegriff in modernen ethischen Ansätzen durch ein personales Konzept ersetzt wurde, ist er für die Diskussion fruchtbar, ob künstliche Systeme (irgendwann) als autonom im ethischen Sinn gelten können. Angenommen, es gäbe ein künstliches System, wie zum Beispiel »Ava« in dem Film *Ex Machina*<sup>69</sup>, die zwar über einen freien Willen verfügt, aber für ihre Befreiung andere instrumentalisiert, manipuliert und tötet. Zwar ist »Ava« nun frei

65 Kant, Immanuel: Grundlegung zur Metaphysik der Sitten. Mit einer Einleitung hg. v. Bernd Kraft und Dieter Schönecker, Hamburg: Felix Meiner Verlag 1999 [1785], S. 61.

66 Rössler, Beate: Autonomie. Ein Versuch über das gelungene Leben, Berlin: Suhrkamp 2017, S. 31.

67 Ebd., S. 32.

68 Kant, Immanuel: Grundlegung zur Metaphysik der Sitten 1999 [1785], S. 54-55.

69 GB 2015, Regie.: Alex Garland.

und autonom, aber verfügt sie über *moralische* Autonomie? Künstliche Systeme sind aus Sicht einer kantischen Ethik nicht autonom und lassen sich auch nicht moralisch gestalten. Kants Ethik in künstliche Systeme zu implementieren »widersprüche [...] dem Geist der kantischen Moral«<sup>70</sup>.

Für eine Ethik der Digitalisierung ist die Frage nach der moralischen Autonomie der Dreh- und Angelpunkt bei der Bewertung künstlicher Systeme. Aus rein funktionaler Sicht kann man künstlichen Systemen eine basale beziehungsweise funktionale Autonomie unterstellen,<sup>71</sup> wobei die genuin menschlichen Fähigkeiten, wie Bewusstsein, freier Wille, Intentionalität und Verantwortung, den künstlichen Systemen (bislang) abgesprochen werden. Dennoch tendieren wir dazu, allein schon in terminologischer Hinsicht künstliche Intelligenz mit menschlichen Eigenschaften zu beschreiben. Für eine Einordnung bedarf es gleichwohl auch der kritischen Auseinandersetzung mit Begriffen. Eine Analyse<sup>72</sup> der gängigsten Definitionen von künstlicher Intelligenz zeigt, dass sie fast immer mit Attributen beschrieben wird, die diese vermenschlicht: Es ist von »lernen«, »verstehen«, »entscheiden«, »wahrnehmen« und »Probleme lösen« die Rede. Künstliche Intelligenz wird also definitorisch anthropomorphisiert oder es wird ihr zumindest »Menschenähnlichkeit« attestiert. Der Begriff suggeriert, dass eine dem Menschen vergleichbare Intelligenz existiert. Das ist allerdings nicht der Fall. Bewusstsein und Gefühle sind Charakteristika, die bislang Lebewesen vorbehalten sind. Auch wenn unklar ist, wie Bewusstsein entsteht und was Bewusstsein eigentlich ist, kann derzeit nicht davon ausgegangen werden, dass künstliche Systeme zur Selbstreflexion fähig sind und ein »Ich«, also eine eigene Identität, ausbilden können. Noch viel weniger sind sie in der Lage, unbewusst etwas zu verdrängen, wie es der Mensch nur allzu gerne tut. Auch über Emotionen verfügen künstliche Systeme nicht, wenngleich sie zu deren Imitation bereits fähig sind. Sie haben auch keinen freien Willen und biologischen Körper. Sie verarbeiten zwar selbstständig Daten, was als sogenanntes

70 Misselhorn, Catrin: Grundfragen der Maschinenethik, 2018, S. 108.

71 Vgl. Floridi, Luciano/Sanders, Jeff W.: *On the Morality of Artificial Agents*, in: *Minds and Machines*, Vol. 14, No. 3, 2004, S. 357; Wallach, Wendell/Allen, Collin: *Moral Machines. Teaching Robots Right from Wrong*, 2009, S. 26-32; sowie eine Übersicht in Loh, Janina: *Roboterethik. Eine Einführung*, 2019b, S. 73.

72 Die Analyse erfolgte im Rahmen des BMBF-Forschungsprojekts »Automatisiertes ELSI-Screening & Assessment-Tool für MTI-Forschungsvorhaben (ELSI-SAT)«, <https://www.elsi-sat.de/>

maschinelles Lernen beschrieben wird, aber sie lernen nicht wie wir Menschen, die im sozialen Miteinander, aus der Erfahrung von Freude und Leid und aus Vernunftgründen lernen. Maschinen sind (bislang) auch keine Generalisten, sondern Spezialisten – sie können (zumindest bislang) nicht wie der Mensch gleichzeitig über verschiedene Fähigkeiten verfügen, also sich ein Kochrezept ausdenken, eine Oma trösten, ein Musikstück komponieren und ein Bild malen. Sie können nicht moralische Prinzipien erkennen, auf sich selbst anwenden und moralisch aufgrund ihres Gewissens handeln. Maschinen haben kein Gewissen und können keine Verantwortung übernehmen. Aber ganz wesentlich ist: Menschen sind zur Autonomie befähigt, während Maschinen nur autonom Informationen verarbeiten. Autonom ist die KI erst dann, wenn sie nicht mehr auszuschalten ist. Eine beeindruckend dezidierte Analyse zur Unterscheidung von menschlicher und künstlicher Intelligenz legt Thomas Fuchs vor, dessen Fazit lautet: »Sicher, die Bezeichnung ›künstliche Intelligenz‹ ist wohl nicht mehr aus der Welt zu schaffen. Doch wir sollten uns immer dessen bewusst bleiben, dass zwischen den Rechen- und Anpassungsleistungen eines Computersystems und den Wahrnehmungen, den Einsichten, dem Denken und Verständnis eines Menschen nicht nur ein gradueller, sondern ein fundamentaler Unterschied besteht.«<sup>73</sup>

Die Stärken und Schwächen des deontologischen Ansatzes können hier nur ansatzweise skizziert werden. Eine Stärke ist sicherlich, dass die Pflichtethik für die Erstellung von ethischen Leitlinien zu künstlichen Systemen ein Begründungskonzept bieten kann – zumindest wenn der Anspruch besteht, über eine minimalistische Forderung der Schadensvermeidung (im Sinne des Utilitarismus) hinauszugehen. Als normative Ethik, die handlungsleitend ist, kann sie dabei helfen, das Richtige zu begründen und dessen Universalisierbarkeit zu prüfen. Problematisch ist natürlich, wenn ethische Leitlinien, die ja eine Form von freiwilliger Selbstverpflichtung darstellen, von Unternehmen nicht in der Praxis umgesetzt werden und womöglich im Sinne eines Ethics-Washing missbraucht werden und/oder gar zur Abwendung von rechtlichen Konsequenzen dienen sollen. So kritisiert beispielsweise AlgorithmWatch, dass Facebook und Google nicht die vom weltweiten Fachverband für Ingenieur\*innen IEEE herausgegebenen Grundsätze zu automatisierten

---

73 Fuchs, Thomas: Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie, 2020, S. 51.

Systemen umsetzen oder staatliche Regulatoren Werte nicht mit Leben füllen beziehungsweise Rechte leerlaufen lassen.<sup>74</sup>

Aus Sicht der Deontologie setzen unsere Werte und Entwürfe des Guten gerechtfertigte moralische Prinzipien voraus. Bezogen auf die Fallgeschichte (1) zur Privatsphäre könnte man deontologisch argumentieren, dass Alex seine Privatsphäre schützen soll, weil sie durch das Autonomieprinzip, auf dem unsere demokratische Gesellschaft fußt, begründet ist. Privatheit ist eine notwendige Voraussetzung für und Ausdruck von Autonomie. Wenn es zu einer Relativierung der Privatheit käme, indem sich das Selbstverständnis von Personen hinsichtlich der Relevanz und Schutzwürdigkeit der Privatheit ändere, würde dies nach Rösslers Einschätzung auch das Fundament unserer Demokratie erbeben lassen: »Dies trifft dann nicht nur die Idee eines gelungenen – selbstbestimmten – Lebens, sondern auch die Idee der liberalen Demokratie: die nämlich auf autonome und sich ihrer Autonomie bewusste und diese schätzende Subjekte angewiesen ist.«<sup>75</sup>

Für die Fallgeschichte (2) der pflegebedürftigen Seniorin kann die deontologische Ethik ebenfalls Anknüpfungspunkte liefern. So werden in der Charta der Rechte hilfe- und pflegebedürftiger Menschen<sup>76</sup> und im ICN-Ethikkodex für Pflegende<sup>77</sup> die Prinzipien einer guten Pflege dargelegt. Letzterer begründet in der Präambel die Aufgaben von Pflege (»Gesundheit zu fördern, Krankheit zu verhüten, Gesundheit wiederherzustellen, Leiden zu lindern«) durch universale Rechte: »Untrennbar von Pflege ist die Achtung der Menschenrechte, einschließlich kultureller Rechte, des Rechts auf Leben und Entscheidungsfreiheit, auf Würde und auf respektvolle Behandlung.«<sup>78</sup> Der Deutsche Ethikrat legt in seiner Stellungnahme zur Robotik in der Pflege<sup>79</sup> ausführlich dar, unter welchen Voraussetzungen ein ethisch verantwortlicher Einsatz von Robotik beziehungsweise technischer Assistenzsysteme möglich erscheint.

74 Vgl. Kayser-Bril, Nicolas: Ethische Richtlinien des größten Weltverbands zeigen kaum Wirkung, 2019. <https://algorithmwatch.org/story/ethische-richtlinien-von-ieee-ohne-wirkung/>

75 Rössler, Beate: Der Wert des Privaten, 2001, S. 218.

76 Bundesministerium für Familie, Senioren, Frauen und Jugend (Hg.): Charta der Rechte hilfe- und pflegebedürftiger Menschen (13. Aufl.), Berlin 2019.

77 Deutscher Berufsverband für Pflegeberufe (Hg.): ICN-Ethikkodex für Pflegende. Berlin 2014.

78 Ebd., S. 1.

79 Deutscher Ethikrat: Robotik für gute Pflege. Berlin 2020. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-robotik-fuer-gute-pflege.pdf>

Dieses begründet er im Wesentlichen aus einer deontologischen Perspektive.

Für die Fallgeschichte (3) der automatisierten Jobbewerbung könnte die deontologische Ethik ebenfalls normativ Anforderungen an die Programmierer\*innen des künstlichen Systems stellen, wie zum Beispiel die Einhaltung des Rechts auf Chancengleichheit, der Geschlechtergerechtigkeit und Nicht-Diskriminierung. Die Frage, wie diese dann in der Praxis umgesetzt werden sollen, kann die Deontologie allerdings nicht beantworten. Und damit kommen wir zum Ansatz der Tugendethik.

### 3.3 Tugendethik

Während die deontologische Ethik auf die Frage *Was soll ich tun?* verkürzt werden kann, lautet die Frage der Tugendethik *Worin besteht das gute Leben?*. Für die aktuelle Tugendethik dient die Ethik des Aristoteles als Vorbild, wonach das Streben des Menschen nach Glück (*eudaimonia*) die Ausbildung von Tugenden beinhaltet. In der aristotelischen Ethik steht der Handelnde und seine Haltung im Vordergrund, erst durch die richtige Haltung wird eine Handlung zur richtigen. Der aristotelische Haltungsbegriff *hexis* lässt sich als Tugend und Charakterzug verstehen, die auf einer freien Wahl und Beurteilung beruhen.<sup>80</sup> Für Aristoteles erweist sich Haltung im Handeln und zielt auf den Umgang mit Affekten und Emotionen. Haltung erlangt man, indem man diese einübt und den Weg der Mitte zwischen einem Zuviel und einem Zuwenig, wie zum Beispiel Tapferkeit anstelle von Feigheit oder Draufgängertum, als das richtige und angemessene Verhalten erkennt und wählt. Eine Haltung ist uns aber nicht von Natur aus gegeben, sie muss aktiv und in konkreten Handlungen ausgebildet werden: »Hexis/Haltung hat mit Gewöhnung (*ethos*) insofern zu tun, als sich in ihr die Geschichte wiederholter Handlungen abzeichnet.«<sup>81</sup> Haltung wäre in diesem Verständnis ein dynamisches Konzept, das sich im Laufe der Zeit in die Grunddisposition eines Menschen einschreibt. Weitergedacht sind mit einer Haltung auch reflexive Fragen verbunden, die

80 Aristoteles: *Nikomachische Ethik*. Übersetzung und Nachwort von Franz Dirlmeier. Anmerkungen von Ernst A. Schmidt, Stuttgart: Reclam 2015, S. 34-52.

81 Wild, Thomas: *Was wissen wir von Haltung? Eine kleine enzyklopädische Suche*, in: Frauke Kurbacher /Philipp Wünschner (Hg.): *Was ist Haltung? Begriffsbestimmung, Positionen, Anschlüsse*, Würzburg: Königshausen und Neumann 2017, S. 95.

sich das Handlungssubjekt in Bezug auf ein Objekt stellt: Was halte ich persönlich von etwas? Wie soll ich mich verhalten? Und warum so und nicht anders?

Neoaristotelische Ansätze zeichnen sich dadurch aus, dass sie die Fähigkeiten und Kompetenzen der Menschen in den Blick nehmen. Als Gerüst für eine Ethik der Digitalisierung lenken sie damit den Fokus auf die Digitalkompetenzen der Nutzer\*innen im Umgang mit und in der Einschätzung von digitalen Medien und künstlichen Systemen. *Digitalkompetenz* ist dann mehr als nur ein technisches Befähigungskonzept, es ist auch ein ethisches, das ein gutes Leben in einer digitalen Welt zum Ziel hat. Hierzu sind allerdings auch die politischen Rahmenbedingungen entsprechend zu gestalten, die es braucht, um bestimmte Fähigkeiten auszubilden. Gemäß Aristoteles muss der Staat das gemeinschaftliche Streben nach einem guten Leben ermöglichen. So könnte man sagen: Fähig kann nur jemand sein, der oder die dazu befähigt wird.

Nach Martha Nußbaum zeichnet die aristotelische Ethik außerdem ihre Offenheit für Veränderungen der »konkreten historischen und kulturellen Bedingungen« aus sowie ihr universales »Bild vom menschlichen Leben, seinen Bedürfnissen und Möglichkeiten«. <sup>82</sup> Nußbaums Fähigkeiten-Ansatz (Capability Approach), <sup>83</sup> den sie zusammen mit Amartya Sen entwickelt hat, ist vielversprechend für eine Digitale Ethik. Denn mit ihm lässt sich das Vorhaben einer wertegestalteten Digitalisierung vorantreiben, die die grundlegenden Bedürfnisse des Menschen im digitalen Zeitalter in den Mittelpunkt stellt. Im Bereich der Pflege gilt der Fähigkeiten-Ansatz bereits als geeignetes Framework für die Gestaltung von Robotern. <sup>84</sup> Nußbaums Liste der zentralen menschlichen Bedürfnisse für ein gutes Leben werden hier zugrunde gelegt, wie unter anderem körperliche Unversehrtheit, Gesundheit, emotionale Beziehungen, Kontrolle über die eigene Umwelt. <sup>85</sup> Den Fähigkeiten-Ansatz auch

---

82 Nußbaum, Martha: *Nicht-relative Tugenden: Ein aristotelischer Ansatz*, in: Klaus Peter Rippe/Peter Schaber (Hg.): *Tugendethik*, Stuttgart: Reclam 1998, S. 144.

83 Nußbaum, Martha: *Fähigkeiten schaffen. Neue Wege zur Verbesserung menschlicher Lebensqualität*. Aus dem Amerikanischen von Veit Friemert, München: Karl Alber Verlag 2015.

84 Vgl. Borenstein, Jason/Pearson, Yvette: *Robot Caregivers: Ethical Issues across the Human Lifespan*, in: Patrick Lin/Keth Abney/George A. Bekey (Hg.): *Robot Ethics. The Ethical and Social Implications of Robotics*, Cambridge, London: MIT Press 2014, S. 254.

85 Nußbaum, Martha: *Fähigkeiten schaffen. Neue Wege zur Verbesserung menschlicher Lebensqualität*, 2011, S. 41-42.

auf andere Bereich der Digitalisierung zu erweitern, ist ein noch zu erfüllendes Desiderat.

Erste Ansätze einer explizit technikbezogenen Tugendethik (Technomoral Virtue Ethics) hat Shannon Vallor vorgelegt. Mit Rückbezug auf eine aristotelische, buddhistische und konfuzianische Ethik schlägt sie eine (erweiterbare) Taxonomie technomoralischer Tugenden vor, die für ein gutes Leben in der digitalen Welt erforderlich seien: Ehrlichkeit, Selbstkontrolle, Demut, Gerechtigkeit, Mut, Empathie, Fürsorge, Zivilität (wie Respekt, Toleranz), Flexibilität, Einsichtsvermögen, Großmut (*magnanimity*), technomoralische Weisheit.<sup>86</sup> Diesen Tugenden ordnet sie wiederum weitere Sekundärtugenden zu. Ihr tugendethischer Ansatz ist vor allem aufgrund seiner interkulturellen Perspektive für eine globale Ethik der Digitalisierung vielversprechend.

Betrachtet man wiederum die drei Fallgeschichten, dann weist die (neo-)aristotelische Ethik ebenfalls Stärken und Schwächen auf. Für die Interpretation der Fallgeschichte (1) von Alex helfen der tugendethische und der Fähigkeiten-Ansatz, da es hier ja um ein mutiges, kluges und besonnenes Handeln sowie eine Haltung zum Wert der Privatheit geht. Die aristotelische Ethik ist durch ihren Praxisbezug genuin definiert. So betont Aristoteles mehrfach, dass man nicht durch das Philosophieren eine Haltung ausbildet, sondern durch ständiges Einüben: mutig durch mutiges Handeln, besonnen durch besonnenes Handeln. Die eigene Privatheit und die der anderen (zum Beispiel Kinder) zu schützen, setzt einerseits die individuelle Ausbildung einer Privatheitskompetenz<sup>87</sup> voraus und andererseits die politische Sicherstellung entsprechender Rahmenbedingungen, um Privatheit überhaupt schützen zu können. Auf diesen Zusammenhang von Fähigkeit und Befähigung weist auch der Capability Approach hin. Wie Alex seinen Wertekonflikt (Privatheit schützen oder Freundschaften pflegen) lösen soll, kann mit Aristoteles schwer begründet werden. Wengleich auch bei Kant wenig zu den Pflichtenkonflikten zu finden ist, »räumt er im Konfliktfall dem stärkeren Verbindlichkeitsgrund den Vorrang vor der stärkeren Verbindlichkeit ein«.<sup>88</sup>

Für die Fallgeschichten (2) und (3) liefert der Fähigkeiten-Ansatz ebenfalls wichtige Erkenntnisse. So geht es jeweils um die Frage, wie in der Praxis ein

86 Vgl. Vallor, Shannon: *Technology and the Virtues. A Philosophical Guide to a Future Worth Wanting*, 2016, S. 119-155.

87 Vgl. Grimm, Petra/Krah, Hans: *Privatsphäre*, in: Jessica Heesen (Hg.): *Handbuch Medien- und Informationsethik*. Stuttgart: J. B. Metzler 2016, S. 184-185.

88 Höffe, Ottfried: *Aristoteles' universalistische Tugendethik*, in: Klaus Peter Rippe/Peter Schaber (Hg.): *Tugendethik 1998*, S. 61.

künstliches System (das technische Assistenzsystem beziehungsweise Jobberwerbungs-system) so gestaltet werden kann, dass es den Menschen befähigt, ein gutes Leben zu führen. Die Antwort darauf ist Ethics by Design.<sup>89</sup> Hierbei handelt es sich um einen Ansatz in der Digitalen Ethik, der vor allem auf den Methoden und Konzepten eines Value Sensitive Design<sup>90</sup> beruht. Er ist ein weit aussichtsreicheres Verfahren als der Versuch, Ethik durch Programmierung zu implementieren. Denn Ethics by Design stellt die Bedürfnisse der Menschen in den Vordergrund, nicht die Codierung. In der aktuellen europäischen (und deutschen) Debatte wird dementsprechend argumentiert, dass eine wertschätzende Gestaltung – insbesondere von künstlicher Intelligenz – einen »Wettbewerbsvorteil« gegenüber der in anderen Weltregionen programmierten und entwickelten KI darstellen könnte.<sup>91</sup> Ethics by Design ist ein Ansatz, der ethische Überlegungen, Wertmaßstäbe, Motive oder Maximen in die vielschichtigen Prozesse der Technikgestaltung, von der ersten Produktidee bis hin zum verkaufsfertigen Endprodukt, einfließen lässt. Ethische Überlegungen sollen also von vornherein angestellt, die Diversität der Nutzer\*innen berücksichtigt und Grundsätze für ethisches Handeln, beispielsweise in der Entwicklung von künstlichen Systemen, eingehalten werden. Ethics by Design beinhaltet den Gedanken, dass Probleme vermieden oder zumindest frühzeitig erörtert werden können, wenn ethische Fragen oder zum Beispiel mögliche ethische Wertekonflikte bereits zu einem frühen Zeitpunkt der Entwicklung eines künstlichen Systems berücksichtigt werden. Es handelt sich dabei nicht um eine einmalige oder punktuelle Angelegenheit. Ein künstliches System wird nicht als »einmal ethisch – immer ethisch« abgenickt, vielmehr begleitet die Reflexion die Entwicklung und ist auch nach Marktreife des künstlichen Systems nicht abgeschlossen.

Wie der Versuch einer ersten Systematisierung ethischer Ansätze zeigt, können sie unterschiedliche Perspektiven für eine wertegestaltete Digitalisierung liefern. Vom Blickwinkel einer Digitalen Ethik aus, die sich an einem humanistischen Menschenbild orientiert, braucht es allerdings mehr Entschlos-

89 Vgl. Van den Hoven, Jeroen: *The Design Turn in Applied Ethics*, in: Jeroen van den Hoven (Hg.): *Designing in ethics*, Cambridge: Cambridge University Press 2017, S. 11-32.

90 Vgl. Friedman, Batya/Hendry, David G.: *Value Sensitive Design. Shaping Technology with Moral Imagination*, Cambridge, London: The MIT Press 2019; Friedman, Batya/Hendry, David G./Borning, Alan: *A Survey of Value Sensitive Design Methods*, in: *Foundations and Trends in Human Computer*, Vol. 11. No. 23, 2017, S. 63-125.

91 Vgl. Hasselbalch, Gry/Tranberg, Pernille: *Data Ethics: The New Competitive Advantage* 2016, S. 97. <https://dataethics.eu/en/book>

senheit als bisher, um die Ethik in der Praxis nachhaltig zu verankern. Ein wichtiger Schritt dorthin ist es, das normative Konzept des *Gemeinwohls* für den Digitalisierungsdiskurs und die Gestaltungs- und Steuerungspolitiken fruchtbar zu machen. Dann kann auch das im einführenden Beitrag proklamierte Ziel erreicht werden, das da lautet: »Von der normativen Kraft des Faktischen zur faktischen Kraft des Normativen«.