VDW VEREINIGUNG DEUTSCHER WISSENSCHAFTLER

Stellungnahme zu den Asilomar-Prinzipien zu künstlicher Intelligenz

Studiengruppe Technikfolgenabschätzung der Digitalisierung

Autoren (Mitglieder der VDW Studiengruppe)

Prof. Dr. Ulrich Bartosch .

Prof. Dr. Stefan Bauberger SJ

Tile von Damm

Dr. Rainer Engels

Prof. Dr. Malte Rehbein

Frank Schmiedchen (Leiter)

Prof. Dr. Heinz Stapf-Finé

Angelika Sülzen

Impressum

Herausgeber: Vereinigung Deutscher Wissenschaftler e.V. (VDW) © April 2018 Geschäftsstelle der Vereinigung Deutscher Wissenschaftler e.V., Marienstraße 19/20, 10117 Berlin

Printed in Germany. Lizenz: CC BY-NC-ND

VDW VEREINIGUNG DEUTSCHER WISSENSCHAFTLER

Inhaltsverzeichnis

1. Einleitung

Die Vereinigung Deutscher Wissenschaftler e.V. (VDW) legt durch ihre Studiengruppe „Technikfolgenabschätzung der Digitalisierung“ ihre erste Stellungnahme zu ethischen, politischen und rechtlichen Fragen der Erforschung, Entwicklung und Anwendung künstlicher Intelligenz (Artificial Intelligence; K.I.) vor, wobei sie der allgemeinen oder starken künstlichen Intelligenz besonderes Gewicht beimisst. Dabei bezieht sie sich als Ausgangspunkt auf die am 06. Januar 2017 beschlossenen Asilomar-Prinzipien zu künstlicher Intelligenz, die den aktuellen Stand der von der K.I.-Entwickler Gemeinschaft begonnenen Diskussion zu den oben genannten Fragen gut abbildet.[1] Mit den Prinzipien identifiziert sich eine sehr große Gruppe der weltweit relevanten Akteure, die in Forschung und Entwicklung künstlicher Intelligenz eingebunden sind. Darunter finden sich viele der in den westlichen Industrieländern maßgeblichen und zum Teil in führenden Funktionen Handelnden.

Die Erforschung, Entwicklung und Anwendung künstlicher Intelligenz können eine fundamentale Gefährdung für die Menschen und ihr friedliches Zusammenleben bedeuten. Es ist im selbstverständlichen öffentlichen Interesse, solche Entwicklungen frühzeitig zu erkennen und vernünftige Handlungen zur Gefährdungsabwehr zu ermöglichen. Es gehört zum Prinzip von Gefährdungsabwehr, präventiv zu agieren, und zwar bevor Entwicklungen, die eine Gefährdung implizieren, in eine unumkehrbare Dynamik geraten sind, die eine wirksame Gegensteuerung nicht mehr zulassen. Die Fortschritte auf dem Gebiet der K.I. erzeugen eine solche fundamentale Gefährdung. Dies ist von führenden Spezialistinnen und Spezialisten dieses Forschungsfeldes erkannt worden und hat u.a. zur Formulierung der Asilomar-Prinzipien geführt. Das ist höchst anzuerkennen und zu würdigen.

Unsere Prüfung der Asilomar-Prinzipien hat ergeben, dass diese der Logik einer erfolgreichen Gefahrenabwehr nicht gerecht werden. Würde die Einhegung der K.I. lediglich der Expertise

[1] Future of Life Institute (2017).

der Asilomar-Prinzipien folgen, würde aus unserer Sicht, ein Risiko in existentieller Größenordnung in Kauf genommen, welches zu anderen Gefährdungen der Menschheit gleichrangig zu sehen ist und für die die Notwendigkeit und Logik unbedingter präventiver Gefahrenabwehr bereits akzeptiert ist.

Dieser Logik folgt z.B. bis heute die atomare militärische Strategie weltweit. Es ist bis auf wenige Ausnahmen bis heute anerkannt, dass es zu verhindern gilt, (auch taktische) atomare Waffen einzusetzen, weil eine unkontrollierbare Eskalation zur Selbstvernichtung der Menschheit, zu einem Auslöschungskrieg ohne Sieger, zu führen droht. Da gegenseitiges Vertrauen nicht als Garantie gegen eine feindliche Handlung vorausgesetzt werden konnte und kann, gilt es, die gegenseitige Kontrollierbarkeit mit dem gleichzeitigen gegenseitigen Vernichtungspotential zuverlässig zu koppeln und ein Konzept „Gemeinsamer Sicherheit“[2] umzusetzen.

Ebenfalls folgt dieser Logik die globale Klimapolitik. Hier gilt seit 1992 für die Staatengemeinschaft, dass eine weitere Erhöhung der Erdtemperatur zu begrenzen sei. Die Gefahrenlage wurde klar erkannt und es steht an, konzertiert zu handeln, um eine „Selbstverbrennung“[3] der Menschheit zu verhindern. Da die Interessenlagen im Bereich der Klimapolitik ungleich Akteurs-reicher und verzweigter sind als im Falle der kriegerischen Bedrohung durch Atomwaffen, ist eine Konsensfindung für globales gemeinschaftliches politisches Handeln weit schwieriger. Der naturwissenschaftlich zweifelsfrei diagnostizierte Sachverhalt anthropogenen Klimawandels wird vielfach überspielt von gegensätzlichen wirtschaftlichen, politischen und auch kulturellen Interessen. Ungeachtet dessen repräsentiert die Transformationsagenda des Übereinkommens von Paris zur Klimarahmenkonvention[4] und

[2] VDW-Mitglied Egon Bahr. Bahr/Lutz (1992) und Independent Commission on Disarmament and Security Issues (1982).
[3] VDW-Mitglied Hans Joachim Schellnhuber. Schellnhuber (2015).
[4] United Nations (2015a).

der Sustainable Development Goals (SDG)[5] die Logik unbedingter präventiver Gefahrenabwehr für diese Menschheitsgefährdung.

Auch auf dem Gebiet der Biotechnologie ergeben sich durch irreversible Eingriffe in das natürliche Erbgut von Mensch, Tier und Pflanze zunehmend Gefährdungen für die Menschheit insgesamt. Die Veränderung der menschlichen Natur und der Natur als Mitwelt des Menschen wird durch einen Prozess in Kauf genommen, dessen fortschreitende Entwicklung – nach begrenzten, kontrollierten Eingriffen – als unkontrollierbare weitere Auswirkung in Kauf genommen wird. Auch hier sind die Ursachen komplex und unübersichtlich. Vor allem wissenschaftliche Neugier, Fortschrittsoptimismus, Geschäftsinteressen und Machtansprüche überlagern sich vielfach. Neueste Entwicklungen machen es Bio-Hackern in „Garagen-Labors“ möglich, gefährliche Organismen nach „Kochrezepten“ zu produzieren und entziehen die technologische Anwendung jeglicher staatlicher Kontrolle. Mit Gene Drive wurden die „Gesetze der Evolution gebrochen“ und eine Dynamik ungeahnten Ausmaßes in Gang gesetzt. Führende Wissenschaftlerinnen und Wissenschaftler in diesem Bereich fordern: „Schützt die Gesellschaft vor unseren Erfindungen“.[6] „Es ist eine wirklich menschheitsbedrohende Gefährdung. Und die Zeitungen berichten darüber nicht.[7]

Diesen fundamentalen Gefährdungen ist gemeinsam, dass sie über eine Dynamik verfügen, die zum Kontrollverlust über ihre Entwicklungen führen. Atomare Kriegführung, Klimakatastrophe, gentechnologische Evolutionsänderung besitzen das Potential zur umfassenden Vernichtung menschlichen Lebens auf der Erde. Zugleich ist dieses Gefahrenpotential verborgen hinter der vermeintlichen Steuerungshoheit durch den Macher dieser Entwicklung: den Menschen. Es ist daher nötig, jenen Gefahren entgegenzutreten, die kaum oder noch nicht sichtbar sind (zumindest für die Mehrheit der Beobachter). Im Bereich der atomaren Bedrohung ist die

[5] United Nations (2015b).
[6] Oye, et.al. (2014).
[7] VDW-Mitglied Ernst-Ulrich von Weizsäcker. Weizsäcker/Wijkman (2017).

öffentliche Aufmerksamkeit am weitesten sensibilisiert. Auch ein gefährliches bis katastrophales Ausmaß des menschgemachten Klimawandels ist als Besorgnis in breiteren Kreisen präsent. Eher gering ist das Unbehagen gegenüber den Eingriffen in die Evolutionsprozesse verbreitet. Eine erst in den letzten drei Jahren realistisch gewordene neue Gefährdung von gleichartigem epochalem Ausmaß ist der Öffentlichkeit ebenfalls weitgehend unbekannt. Sie ist Gegenstand dieser VDW-Stellungnahme.

Mit der künstlichen Intelligenz öffnen die Menschen eine weitere Büchse der Pandora. Sie hat das Potential die Logik der Steuerung – auch der genannten Gefahren – zu unterlaufen. Im digitalen Zeitalter entfalten die Hilfsmittel für menschliches Denken womöglich eine autonome Position, die sich mächtig gegen den ohnmächtig werdenden Menschen richtet. Es ist höchste Zeit für eine reflektierte, aktive Gefahrenabwehr. Dies will der folgende Text anmahnen.

Die Vereinigung Deutscher Wissenschaftler sieht sich zu dieser Mahnung verpflichtet. Ihre Gründung geht auf die Warnung der „Göttinger Achtzehn“[8] von 1957 vor den Gefahren der Atomrüstung zurück.[9] Die Verantwortung der Wissenschaft bedeutet heute nicht weniger als damals, die unerkannten Gefahren für die Öffentlichkeit sichtbar zu machen und mit wissenschaftlicher Expertise der Gefahrenabwehr zu dienen. In diesem Sinne hat die VDW „Studiengruppe Technikfolgenabschätzung der Digitalisierung“ ihre Arbeit aufgenommen und legt hiermit ihre ersten Ergebnisse vor.

[8] Göttinger Achtzehn (1957).
[9] https://vdw-ev.de/ueber-uns/geschichte-und-ziele/

2. Zielsetzung, Definitionen und Ausgangsthesen

Ziel der Stellungnahme ist es, einen Beitrag zur europäischen und internationalen Diskussion über mögliche Folgen von K.I. und notwendigen Maßnahmen im Hinblick darauf zu leisten. Als Studiengruppe der VDW argumentieren wir aus Sicht des Schutzes der individuellen, wirtschaftlichen, sozialen und kulturellen Menschenrechte[10] und betrachten darüber hinaus mögliche Folgen einer allgemeinen oder starken K.I. für die Menschheit als Ganzes. Dabei ist das Vorsorgeprinzip der EU rechtlicher Kompass für die Betrachtung.[11]

Der Begriff Intelligenz ist umstritten. Wir nutzen deshalb ein allgemein anerkanntes Grundverständnis, welches Intelligenz als ein nicht direkt beobachtbares Phänomen beschreibt, das die kognitiven Fähigkeiten und Wissensbestände einer Person bezeichnet, die ihr zu einem gegebenen Zeitpunkt zur Verfügung stehen.[12]

Künstliche Intelligenz ist ein (empirisches) Teilgebiet der Informatik und beschäftigt sich mit Methoden, die es einem Computer ermöglichen, solche Aufgaben zu lösen, die, wenn sie vom Menschen gelöst werden, Intelligenz erfordern. Der Begriff „künstliche Intelligenz“ wurde erstmalig 1956 von dem US-Informatiker John McCarthy schriftlich verwendet.[13] Bereits 1950 beschrieb Alan Turing die Möglichkeit einer von Computern simulierten Intelligenz.[14] Definitionsmerkmale von K.I. sind die von Beginn an eingebaute Fähigkeit zu Lernen und mit Unsicherheit, Ungenauigkeit/Unschärfe und Wahrscheinlichkeiten umzugehen.[15]

[10] United Nations (1948) und United Nations (1966).
[11] Kommission der Europäischen Gemeinschaften (2000).
[12] Kognition wird hier als ‚situierte Kognition‘ verstanden, die neben internen Berechnungsprozessen im Gehirn auch und vor allem die reziproke Echtzeitinteraktion eines körperlich auf bestimmte Weise verfassten Systems mit seiner Umwelt meint (Walter, 2014).
[13] BITKOM (2017).
[14] EFI (2018), S. 69.
[15] BITKOM (2017) beschreibt eine Taxonomie der Automation des Entscheidens und entwickelt ein 5-Stufen-Modell hierzu.

Unter schwacher K.I. werden solche Formen von K.I. verstanden, bei denen Maschinen nur in einem spezifischen Anwendungsbereich intelligentes Verhalten mit Mitteln von Mathematik und Informatik simulieren und Lernfähigkeit besitzen. Demgegenüber meint allgemeine oder starke K.I. eine generelle Lernfähigkeit, einschließlich der Befähigung, sich selbst autonom weiterzuentwickeln. Eine Superintelligenz ist definiert als eine starke K.I., die dem menschlichen Gehirn zumindest in vielen Bereichen überlegen ist. Sie entsteht hypothetisch, wenn eine starke K.I. sich durch Rekursion selbst verbessert und erweitert.[16] Mögliche Entwicklungsrichtungen zur Realisation einer Superintelligenz sind Algorithmen (auch in Maschinen, Robotern, o.a.), Transhumanismus (z.B. gentechnische Verbesserungen von Menschen oder die Verschmelzung von beidem in Cyborgs) sowie der Einsatz von künstlichen, neuronalen Netzen.

Die künstliche Intelligenz hat eine wechselvolle Geschichte. Mittlerweile hat sie einen Entwicklungsstand erreicht, durch ihre Anwendung alle Lebensbereiche drastisch zu verändern. Ausgangspunkt unserer Arbeit ist die Sorge, dass die wahrscheinlichen Gefahren einer allgemeinen oder starken künstlichen Intelligenz nicht rechtzeitig angemessen diskutiert und damit adäquate Maßnahmen zur Gefahrenabwehr unmöglich werden. Dies liegt zum einen daran, dass die großen Durchbrüche bei schwacher K.I. ganz überwiegend erst in den letzten drei Jahren stattgefunden haben und zum anderen daran, dass ebenso wie bei anderen Risikotechnologien (z.B. Atomkraft, Klimawandel, Gentechnik) die möglichen Gefahren der noch nicht-existierenden starken K.I. abstrakt, unsichtbar und unbekannt sind. Daten und Algorithmen entfalten ihre Wirkung erst, wenn sie in die menschliche Lebenswirklichkeit eintreten und sich dort manifestieren.

[16] Dabei gelingt es der starken K.I. ihren eigenen Programmcode so anzupassen und zu verbessern, dass sie in der nächsten Stufe, bzw. ihrer nächsten Generation weitere Verbesserungen vornehmen kann, welche den vorherigen Versionen unmöglich gewesen wären. Diese rekursiven Lernzyklen werden von der starken K.I. so lange fortgesetzt, bis die menschliche Intelligenz übertroffen und sie zur Superintelligenz wird. Als Voraussetzung hierfür wird angesehen, dass die starke K.I. ihren Anwendungszweck und die Gestaltung des eigenen Programmcodes „versteht“.

Folgende Ausgangsthesen liegen der Stellungnahme zu Grunde:

1. Forschung, Entwicklung und Anwendung von K.I. entwickeln sich exponentiell.[17] Dabei ist Deutschland neben den USA und China in der Grundlagenforschung führend.[18]
2. Schon Anwendungen schwacher K.I. verändern das Sozial- und Kommunikationsverhalten[19] und unsere Alltagskultur (z.B. Social Scoring) und können gesellschaftliche Systeme bedrohen (z.B. Wahlbeeinflussungen).
3. Die Schaffung starker K.I. ist angesichts der exponentiellen Wachstumsgeschwindigkeit der K.I. Forschung wahrscheinlich, auch wenn der Realisationszeitpunkt schwer erkennbar und nicht exakt vorhersagbar ist.[20]
4. Der Übergang der Entwicklung von schwacher zu starker K.I. ist dabei ein kontinuierlicher Prozess, der vor allem dort stattfindet, wo die hierfür erforderlichen Ressourcen (Daten, Finanzen, Macht) vorliegen. Die Ergebnisse von Forschung und Entwicklung (F+E) in diesem Bereich werden aufgrund wirtschaftlicher und machtpolitischer Interessen nicht zwangsläufig veröffentlicht, was eine gesellschaftlich legitimierte Steuerung unmöglich macht oder zumindest erheblich erschwert.
5. Eine starke K.I. kann durch ihr autonomes Wirken für einzelne Menschen und für die Menschheit gefährlich werden:
    a. Eine K.I. trifft vom Menschen intendierte Entscheidungen zur Zielerreichung oder Selbsterhaltung, die als Nebeneffekt (kollateral) Menschen schaden, was

[17] BITKOM (2017); EFI (2018).
[18] EFI (2018), S. 68ff.
[19] Eberle (2015), Henk (2014). Der Einfluss von K.I. ist aber schwer unabhängig messbar von anderen Aspekten der Digitalisierung.
[20] Bostrom/Müller (2013): Diese gehen von 2022 (Median optimistisches Jahr), über 2040 (Median realistisches Jahr) bis 2075 (Median pessimistisches Jahr) aus. Zu ähnlichen Schätzungen kann man auch kommen, wenn man das Moore´sche Gesetz von 1965 zu Grunde legt und die explosionsartigen Fortschritte der letzten drei Jahre extrapoliert. Die Rechenleistung des menschlichen Gehirns liegt laut Raymond Kurzweil bei ca. 10.000 Teraflops. Diese Rechenleistung haben Großrechenanlagen bereits deutlich überschritten. Darüber hinaus sind sie miteinander direkt vernetzbar. Es gibt aber auch Stimmen die behaupten, dass die Entwicklung starker K.I. „in absehbarer Zeit noch nicht realisierbar“ sei (EFI (2018), S. 69).

auch bis zum Grad der völligen Unterwerfung oder Vernichtung der Menschheit führen kann, ohne dass es die K.I. „bewusst gewollt" hätte.[21]

b. Eine K.I. ist destruktiv intendiert (z.B. als tödliches autonomes Waffensystem) und steigert seine Effizienz auf eine Weise, die mehr Menschen/die Menschheit gefährdet.

c. Eine K.I. entwickelt nicht intendierte Kompetenzen und verfolgt selbst gesetzte Ziele, die einzelne Menschen oder die Menschheit gefährden.

6. Ab einem unbestimmten Zeitpunkt können Menschen dem Selbstverbesserungsprozess einer starken K.I. nicht mehr folgen (auch nicht als „zusammengeschaltetes" Kollektiv), da ihr Lernfortschritt immer langsamer als der der K.I. ist, womit eine Steuerung/Korrektur unmöglich wird.

7. Aus den Ausgangsthesen 3 bis 6 ergibt sich die Notwendigkeit zur Umsetzung des Vorsorgeprinzips.[22] Mit einem hinreichenden zeitlichen Sicherheitsabstand vor der Erschaffung einer starken K.I. müssen alle erforderlichen (v.a. normativen) Schritte erfolgreich abgeschlossen sein, die verhindern, dass eine starke K.I. einzelne Menschen und die Menschheit als Art insgesamt in irgendeiner Form, zu irgendeinem Zeitpunkt, unter irgendwelchen Umständen gefährden kann (antizipatorische Governance). Um dies zu erreichen bedarf es unverzüglich zielgerichteter Entscheidungen und entsprechender Maßnahmen.

[21] In Weiterentwicklung eines Beispiels von Stephan Hawking: Verantwortungsvolle Menschen planen und bauen ein Wasserkraftwerk. Dabei haben sie umfangreiche Umwelt- und Sozialverträglichkeitsprüfungen unternommen. Säugetiere und Vogelnester wurden evakuiert und Menschen großzügig für die erforderliche Umsiedlung vor Flutung des Stausees entschädigt. Niemand aber hat sich um die 50.000 Ameisenkolonien gekümmert, die nun am Grunde des Stausees ertrunken sind: 2.500.000.000 ermordete Ameisen sind das Ergebnis, „was niemand gewollt hat". Hawking sagt, dass wir jetzt sofort handeln müssen, wollen wir verhindern, eines Tages möglicherweise an Stelle der Ameisen zu sein.

[22] Gemäß der bei der UNO-Konferenz über Umwelt und Entwicklung in Rio (1992) verabschiedeten Agenda 21 (Kapitel 35, Abs. 3) kann das Vorsorgeprinzip definiert werden als zwingende Handlungsaufforderung, mögliche Risiken/Gefahren, die mit einer nicht (exakt) vorhersagbaren Eintrittswahrscheinlichkeit zu nicht (exakt) vorhersagbaren negativen Folgewirkungen für Menschen (v.a. Leben und Gesundheit) und Umwelt, jetzt oder in Zukunft, führen können, alle zielführenden Schutzmaßnahmen vorbeugend zu ergreifen. Auch dies knüpft an Hans Jonas Verantwortungsethik an.

8. Mensch und Maschine beginnen sich auf drei Arten zu entgrenzen:
    a. Technik agiert zunehmend selbständig und übernimmt damit Funktionen und Aufgaben von Menschen;
    b. neurologisch-kognitiv wirkende Komponenten werden in Menschen eingebaut (sog. „Enhancements“);[23]
    c. Kopplung von menschlichen Gehirnen und K.I.-Systemen.

    Es kann von einer fortschreitenden Entgrenzung gesprochen werden, die starke ethische Implikationen hat. Die mit dem Transhumanismus verbundenen Fragen werden in dieser Stellungnahme nur ansatzweise angesprochen.

3. Kritische Reflexion der Asilomar-Prinzipien

3.1 Einführung

Das in Boston (USA) gegründete Future of Life Institute (FLI) beschäftigt sich seit März 2014 mit möglichen existentiellen Gefahren weiterer Technikentwicklung für die Menschheit. Dabei stehen die Arbeiten zur Risikominderung von K.I. ausdrücklich im Mittelpunkt der Institutsarbeit.[24] Im Januar 2017 organisierte das FLI in Asilomar (an der kalifornischen Küste) die Tagung “Beneficial AI” mit knapp 1000 Teilnehmenden, darunter über 100 der weltweit führenden K.I. Forscher/-innen und Unternehmer/-innen, um über die Auswirkungen von K.I. zu diskutieren. Die “Asilomar AI Principles” sind ein Ergebnis dieser Tagung. Das Programmkomitee war besetzt mit Menschen, die direkt oder indirekt beruflich mit K.I. befasst sind.[25] Die hierbei verabschiedeten und von zahlreichen Wissenschaftlern unterzeichneten 23

[23] TAB (2016) und Vorwinkel (2017).
[24] Gründer, bzw. starke Unterstützer des Instituts sind Stephan Hawking, Elon Musk, Max Tegmark (MIT), Jaan Tallinn (Skype-Erfinder), Stuart J. Russell (Informatik), George Church (Biologie), Saul Perlmutter und Frank Wilczek (Physik) sowie Alan Alda und Morgan Freeman (Schauspieler). Elon Musk hat dem FLI im Januar 2015 ein mit 10 Mio. USD dotiertes Forschungsprogramm zu K.I. finanziert, das auf Sicherheitsfragen und die Entwicklung „nützlicher“ K.I. fokussiert. Mit den Finanzmitteln wurden 37 Forschungsprojekte gefördert.
[25] Future of Life Institute (2017).

Prinzipien stellen einen Vorschlag für eine freiwillige Selbstverpflichtung für die Forschung, Entwicklung und Anwendung von K.I. dar. Sie sind als Reaktion auf die beschleunigte technologische Entwicklung in diesem Bereich zu sehen, der von den Organisatoren des FLI zu Recht als "major change [...] across every segment of society"[26] bezeichnet wird. Bereits in den ersten sechs Wochen sind die Prinzipien von über 1000 unmittelbar an Forschung und Entwicklung von K.I. beteiligten Wissenschaftler/-innen sowie von weiteren knapp 2000 Personen unterzeichnet worden.

Sie haben auch schon erste Wirkungen gezeigt:

- Die Universität von Montreal hat 2017 einen offenen Prozess zur Erarbeitung der Montreal Responsible AI Prinzipien begonnen, der 2018 abgeschlossen werden soll.[27]
- Etwa 60 Wissenschaftler/-innen aus 30 Staaten rufen zum Boykott des südkoreanischen KAIST-Instituts auf, da dieses mit dem südkoreanischen Rüstungskonzern Hanwha zusammenarbeitet.[28]

Das Motto der o.g. Tagung „Beneficial AI" hat die Stoßrichtung bereits vorgegeben, die sich auch in der kurzen Präambel der Prinzipien wiederfindet: "Artificial intelligence has already provided beneficial tools that are used every day by people around the world. Its continued development, guided by the following principles, will offer amazing opportunities to help and empower people in the decades and centuries ahead."[29] Mit dieser rein positivistischen und utilitaristischen Sichtweise auf den Einsatz von K.I.-Technologien bleibt (neben anderen) die große und vielleicht entscheidende Frage offen: Wie ist mit den Entwicklungen umzugehen, die nicht oder nicht für alle "beneficial" sind und vor allem, wie mit den Bedrohungen durch diese Entwicklungen, den "threats".

[26] Ebenda.
[27] Université de Montréal (2018).
[28] Beim Aufbau des „Research Centers for the Convergence of National Defense and Artificial Intelligence". Centre on Impact of AI and Robotics (2018).
[29] Future of Life Institute (2017).

Die Bewertungen in dieser Stellungnahme basieren auf dem Postulat der Einhaltung der Menschenrechtskonvention der Vereinten Nationen als absoluter, wenn auch nicht hinreichender Mindestbedingung und nehmen das EU-Rechtsverständnis des Vorsorgeprinzips als einzuhaltendes Werte- und Normengefüge. Außerdem beziehen sich die Überlegungen auf Hans Jonas' Verantwortungsethik[30] als philosophischer Referenz, die auch dem Vorsorgeprinzip zu Grunde liegt.

Nach Jonas' Ansatz der "Heuristik der Furcht" ist bei jeder menschlichen Entscheidung zunächst von den potentiellen Folgen für die Zukunft auszugehen, die diese Entscheidung nach sich ziehen könnte. Jonas' Motiv, "die Unversehrtheit seiner Welt [des Menschen] und seines Wesens gegen die Übergriffe seiner Macht zu bewahren"[31] und sein Imperativ "Handle so, dass die Wirkungen deiner Handlung verträglich sind mit der Permanenz echten menschlichen Lebens auf Erden"[32] sind ein hilfreicher Maßstab auch für die Bewertung von K.I. Da sich aus dieser Sichtweise aber lediglich ableiten lässt, wie wir nicht leben wollen, müssen darüberhinausgehend auch positiv-normative Vorschläge unterbreitet werden, wie der Umgang mit K.I. gestaltet werden soll.

Wie sieht also die Welt aus, in der wir zukünftig leben wollen? Diese Frage beantworten die Asilomar-Prinzipien nicht. Sie unterstellen eine allgemeingültige und breit akzeptierte Zustimmung zu einer technikoptimistischen Zukunftskonzeption, die einerseits unbestimmt bleibt und auf Grund fehlender sozio-ökonomischer Analyse Gefahr läuft, dass nur wenige sie bestimmen werden, andererseits aber, mit der weiteren Verfolgung des Technologiepfades, K.I. als unabwendbares Schicksal akzeptiert. Aus beidem zusammen genommen erwachsen schon heute gesellschaftliche Gefahren für Demokratie, Rechtsstaat und Menschenrechte.

[30] Jonas (1979).
[31] Ebenda, S.9.
[32] Ebenda, S.35.

- Wer bestimmt was “gut” ist, wenn die Technologie faktisch allumfassend ist und alle betrifft - nicht nur die, die ein bestimmtes K.I.-basiertes Produkt nutzen?
- Haben wir einen Konsens darüber, welche Gefahren/Risiken wir bereit sind zu akzeptieren, um in den Genuss von K.I. zu kommen?
- Wie ist ein solcher Konsens auf globaler Ebene herstellbar?
- Wird dieser einmal hergestellte Konsens auch in der Zukunft bestehen, wenn Dinge nicht mehr rückgängig zu machen sind, die aus dem Konsens resultieren?
- Wenn ein Regelwerk zur Kontrolle von starker K.I. festgelegt würde, wer garantiert wie, dass nicht die starke K.I. selbst sich über dieses Regelwerk hinwegsetzt und eigene Maßstäbe (auch ethische) festlegt – und dass sie dies mit einer Aktionsgeschwindigkeit tut, die effektive menschliche (Gegen-) Reaktionen unmöglich macht?

Die Asilomar-Prinzipien greifen eine Reihe ethischer Fragen in Bezug auf K.I. auf und beschreiben moralisch hergeleitete Best Practices in Bezug auf Forschung und Entwicklung (F+E) von K.I., wobei sie einen großen Interpretationsspielraum zulassen. Die Prinzipien verwenden dabei zahlreiche unbestimmte Rechtsbegriffe, die definiert werden müssten, sofern sie zu einem handhabbaren Instrument weiterentwickelt werden sollen. Damit verbunden ist die Frage nach dem Definitionsrecht.

Viele gewählte Formulierungen scheinen von einem herrschaftsfreien, kooperativen Zusammenarbeiten der mit F+E von K.I. betrauten Wissenschaftler/-innen auszugehen, bzw. davon, dass ein solches möglich ist, sofern der gute Wille hierzu bei den Forscher/-innen vorhanden ist. Bereits hier ist zu fragen:

- Ist das ein realistischer Ausgangspunkt?
- Inwiefern muss von Anfang an bedacht werden, dass F+E auch zu K.I. in Zusammenhängen stattfindet, in denen externe Zielformulierungen (bspw. unternehmerischer Gewinn, nationale Sicherheit) zumindest signifikant hineinwirken, bzw. die Forschungsagenda im Wesentlichen bestimmen?

- Ist es realistisch, dass der Einsatz von K.I. ohne formale Beteiligung existierender institutioneller Strukturen und Prozesse des demokratisch verfassten politischen Raums allein durch freiwillige Übereinkünfte der Forschenden kontrolliert werden kann?

Die Asilomar-Prinzipien sind in die Abschnitte Fragen der Forschung, Ethik und Werte, und längerfristige Probleme gegliedert.

3.2 Forschungsfragen

Ein wesentlicher Ausgangspunkt der Prinzipien ist der Anspruch, dass F+E zu K.I. so fokussiert werden, dass nur „nützliche“ K.I. entsteht. Unbeantwortet bleibt die Frage, wer wie definiert, ob und wann eine K.I. nützlich ist. Der im Prinzip Nr. 1 konstruierte Zusammenhang zwischen „directed" und „beneficial" ist unlogisch, da Beides unterschiedliche Kategorien sind.[33] Etwas kann ungeleitet oder unkontrolliert, aber dennoch nützlich sein, so wie es sehr Vieles gibt, das kontrolliert entsteht, aber unnütz oder schädlich ist. Ziel der K.I.-Forschung müsste es daher sein, die "Intelligenz" sowohl zu kontrollieren als auch dafür zu sorgen, dass sie grundsätzlich immer und überall ökologisch und sozial nachhaltige, sinnvolle Dinge tut.

Zu diskutieren wären darauf aufbauend die grundlegenden Probleme des Utilitarismus, da der Zweck nicht, bzw. nur in Situationen äußerster Gefahr alle Mittel rechtfertigt, schon gar nicht, wenn über den Zweck (noch) kein gemeinsames Grundverständnis besteht, sondern die Zweckbestimmung vielleicht nur von Wenigen beherrscht wird (z.B. sind aus Sicht der Produzenten tödlicher autonomer Waffensysteme sicherlich sehr „nützlich“).

Wer legt also fest, für welche Menschen, Gruppen oder Institutionen die K.I. mindestens nützlich sein muss, um insgesamt als nützlich zu gelten? So kann K.I. zu wirtschaftlichen Effizienzgewinnen führen, aber die Frage, ob wir diese Effizienzgewinne sozio-kulturell wollen,

[33] “The goal of AI research should be to create not undirected intelligence, but beneficial intelligence”.

kann nicht nur über die Summe von Konsumentscheidungen beantwortet werden. Wir haben einen Punkt technologischer Entwicklung erreicht, an dem im Einzelfall zu hinterfragen ist, ob ein Mehr an Bequemlichkeit nicht einen zu hohen Preis hinsichtlich des Verlustes menschlicher Fähigkeiten und Fertigkeiten verlangt.[34] Wenn wir entscheiden, dass die Effizienzgewinne „netto" gewollt sind, muss in einem zweiten Schritt über deren Verteilung gesellschaftlich entschieden werden. Mit diesen Fragen müssen sich vorrangig die kultur-, sozial- und wirtschafts-wissenschaftlichen Disziplinen beschäftigen.[35]

Generell bedarf es einer legitimierten Forschungspolitik, die nachvollziehbar definiert, was ethisch verantwortungsvolle Innovation auf dem Gebiet der K.I. im Detail bedeutet.

Prinzip Nr. 2 beschäftigt sich mit der Notwendigkeit begleitender Forschung. K.I.-Systeme wirken sich, wie andere technische Entwicklungen, auf gesellschaftliche Entwicklungsprozesse aus und beeinflussen sie fundamental. Aus diesem Grund ist es notwendig, Fragen der Irreversibilität und der Risikofolgenabschätzung in die Forschung einzubeziehen, um alle relevanten ethischen und gesellschaftlichen Herausforderungen besser verstehen und bewerten zu können. Dies muss auch für schwache K.I. und insbesondere vor Entwicklung einer starken K.I. erfolgen und die diesbezügliche Forschung muss mit ausreichend finanziellen Mitteln ausgestattet werden. Die EU legt im Augenblick hierzu erste Vorschläge vor, die genau zu betrachten sein werden.[36]

Prinzip Nr. 3 spricht von einem „konstruktiven" und „gesunden" Austausch zwischen K.I.-Forschern und politischen Entscheidungsträgern.[37] Demgegenüber sieht das Demokratie-

[34] Ein seit Jahrzehnten bekanntes Beispiel ist die gesellschaftlich anerkannte „Selbstfolter" im Fitness-Studio, weil unsere Körper im Alltag zu wenig gefordert sind.
[35] Es gibt bisher keine Hinweise, dass die Nutzung von K.I. zu einer Verringerung sozialer Ungleichheit führen wird. Demgegenüber gibt es aber erste Hinweise, dass sie genau das Gegenteil zumindest tendenziell befördert. Dies gilt generell für die Digitalisierung wirtschaftlicher Prozesse und damit auch für K.I.
[36] Im Rahmen der geplanten umfassenden europäischen Strategie zur künstlichen Intelligenz, die am 25. April 2018 veröffentlicht wird. EU-KOM (2018) und Krempl (2018).
[37] So auch: Brundage, et al (2018).

prinzip vor, dass Legislative und Exekutive nicht gleichrangig mit Gruppen von Industrievertretern oder Wissenschaftler/-innen ist. Vielmehr müssen Zielsetzungen demokratisch legitimiert und Verstöße gegen operative Regeln zur Umsetzung dieser Ziele staatlich und, im Fall grenzüberschreitender Aspekte, auch multilateral sanktioniert werden können.[38]

Die Prinzipien Nr. 4 und 5 gehen davon aus, dass es möglich ist, eine Kultur von Zusammenarbeit, Vertrauen und Transparenz bei Forschern und Entwicklern von K.I. zu schaffen.[39] Dies ist in dieser allgemeinen Formulierung lediglich auf persönlicher Ebene realistisch. Nur in wenigen (v.a. öffentlichen) Institutionen werden Forscher/-innen dafür bezahlt, vertrauensvoll zusammenzuarbeiten. Ansonsten wird das nur insoweit positiv sanktioniert, wie es der institutionellen Zielerreichung dient. Im Konfliktfall, wenn im Wettbewerb stehende oder sogar gegensätzliche institutionelle Interessen (z.B. Unternehmensgewinn, nationale Sicherheit[40]) aufeinanderstoßen, müssten Forscher/-innen, zur Einhaltung dieses Prinzips bereit sein, negative Sanktionen (Arbeitsplatz- und Statusverlust, Gefängnis) zu ertragen. In der aufsehenerregenden Studie „The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation“[41] fordern 26 führende K.I.-Entwickler von Forschern und Entwicklern als nicht delegierbare Aufgabe, die Folgen ihrer Arbeiten nicht nur vorzudenken, sondern auch alle relevanten Akteure aktiv vor negativen Konsequenzen zu warnen und den Kreis derer, die informiert mitagieren und entscheiden sollen, beständig zu erweitern. Ein Dilemma besteht darin, dass einerseits offen über die möglichen Folgen konkreter K.I. F+E und Anwendungen frühzeitig, formalisiert und legitimiert gesprochen werden muss, anderseits aber eine Offenlegung zu Grunde liegender Algorithmen die Gefahr für missbräuchliche Nutzung erhöht.[42]

[38] Die o.g. EU-Strategie geht davon aus, dass eine Überwachung des Fortschrittes bei der K.I.-Entwicklung durch ein K.I.-Observatorium notwendig ist. EU-KOM (2018).
[39] Ähnlich: Executive Office of the President (2016), S. 42, Empfehlung 19 und 21.
[40] NSTC (2016), S.3.
[41] Ebenda.
[42] Brundage, et al (2018).

Versuche ethischer und rechtlicher Regelungen für K.I. laufen ins Leere, solange sowohl die dystopische Sicht (unter Berücksichtigung von Jonas' Konzeption der "Fernwirkung der Technik": Kollateralschäden sowie Auswirkungen auf die Zukunft) nicht jedem klar ist und ein Zukunftsentwurf ("wie wir leben wollen") noch nicht verhandelt wurde. Symptomatisch hierfür ist der Grundsatz: "Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards". Denn die *Safety Standards* (oder allgemeiner gesetzliche Regelungen und Limitierungen) gibt es noch nicht, und es wird schwer sein, sie zu finden, insbesondere solange die technologische Entwicklung so viel schneller ist als eine Gesetzgebung reagieren und Rahmen setzen kann (sog. Problem des *Cultural Lag*).[43] [44]

3.3 Ethik und Werte

Die Prinzipien Nr. 6 und 7 behandeln den Anspruch von Fehlertransparenz und Betriebssicherheit und fordern umfassenden Sicherheitsschutz während der gesamten Betriebsdauer.[45] Hier stellen sich beispielsweise folgende erste Fragen:

1) Wie kann der geforderte Schutz technisch effektiv umgesetzt werden?
2) Was tun bei nicht eindeutigen Entscheidungen?
3) Was tun, wenn das Handeln der K.I. an komplexen, zahlreichen und/oder widersprüchlichen Werten gemessen werden muss, deren wirksame Berücksichtigung, wenn überhaupt, nur wiederum mit Hilfe einer K.I. möglich wäre?
4) K.I.-Systeme werden zumindest teilweise nicht eindeutig lokal konzentriert sein, sondern regional/global verstreut, über mehrere Staaten. In diesen Fällen sind lückenlose Überwachung, Sicherheit und Nachprüfbarkeit extrem schwierig zu gewährleisten.

[43] Ebenda.
[44] Bei aller Unsicherheit: Im Rahmen der geplanten umfassenden europäischen Initiative wird die Entwicklung einer Charta für K.I.-Ethik angekündigt. Diese soll ab Anfang 2019 in einer breiten Debatte entwickelt werden. Siehe EU-KOM (2018).
[45] Siehe auch: NSTC (2016).

Um Fehlfunktionen lückenlos zu dokumentieren und die Erfolgschancen für Reparaturen zu optimieren, müssen Konstruktionsangaben und Quellcodes in staatlichen Kontrollinstitutionen gespeichert werden, damit sie langfristig öffentlich zugänglich sind. Das umfasst auch Trainingsdaten und -kriterien. Der Zugang zu Quellcodes (und zu Algorithmen) ist bisher nur unzureichend geregelt. Wir wissen, dass eine Freigabe von urheberrechtlich geschützten Quellcodes durch wirtschaftlich mächtige Akteure relativ leicht umgangen werden kann. Es bedarf daher klarer und strenger gesetzlicher Regelungen sowie umfassender Forschung, wie das Risiko der Geheimhaltung von Quellcodes zumindest deutlich verringert werden kann. Andererseits ist nach heutiger Rechtslage die Rekonstruktion von Quellcodes aus dem öffentlich zugänglichen kompilierten Code (Reverse Engineering) grundsätzlich zulässig (wenn auch oft, wie in der EU, auf bestimmte Fälle beschränkt). Das erleichtert wiederum die nicht kontrollierte Entwicklung neuer oder verbesserter K.I. außerhalb bekannter Strukturen mit allen damit verbundenen, möglichen Folgen. Daher ist die Forderung nach Einführung von Auskunftsrechten sowie Kennzeichnungs- und Publikationspflichten, wie sie beispielsweise durch den Verbraucherzentrale Bundesverband (vzbv) gestellt wird, von zentraler Bedeutung.[46]

Zu den Wesenseigenschaften von starker K.I. sowie von weiter entwickelter schwacher K.I. gehört es, dass diese neues und unvorhersehbares „Verhalten“ zeigen, welches sich aufgrund seiner Komplexität nicht aus den Algorithmen rekonstruieren lässt. Die Systeme AlphaGo und AlphaGo Zero zeigen das bereits exemplarisch. Die Forderung nach Fehlertransparenz ist daher widersprüchlich, weil damit auch manche Fehler unerklärbar bleiben werden. Das verschärft die Ansprüche an Kontrolle und Haftungsregelungen.

Auf den im Prinzip Nr. 8 angesprochenen Einsatz von K.I. in gerichtlichen Verfahren wird in Zukunft umfassend einzugehen sein. Motivation des bereits stattfindenden Einsatzes in juristischen Entscheidungsprozessen sind angestrebte Effizienzgewinne, insbesondere bei

[46] Verbraucherzentrale Bundesverband (2017).

umfangreichen Verfahren sowie die Unterstützung „besserer“ Entscheidungen. Unmittelbar in Prozessen der juristischen Entscheidungsfindung kann K.I. eingesetzt werden, um den Sachverhalt zu strukturieren, einen Vorschlag für die Entscheidung zu erstellen, Rückfall- und Sozialprognosen zu geben oder im Extremfall eine Entscheidung autonom zu treffen. Dies ist in den meisten Staaten nach derzeitiger Rechtslage nur in sehr engen Grenzen möglich.

Wir sehen den Einsatz von K.I. bei Bewertung und Entscheidung in gerichtlichen Verfahren problematisch.[47] Insbesondere die Grenzziehung bei der Unterstellung von Kausalitäts-zusammenhängen sowie die Definition der Grenzen des freien Willens sind eine Herausforderung. In USA ist bereits eine Einstufungssoftware zur Rückfall- und Sozial-prognose im Einsatz, bei der erste Evaluationen zeigen, dass eine deutliche Gefahr besteht, dass sie zu Diskriminierung aufgrund der Hautfarbe führt.[48]

Ein sehr wahrscheinliches Einsatzgebiet der K.I. wird das Erstellen von Sachverständigen-Gutachten (z.B. zur Glaubwürdigkeit von Zeugen) oder der Nachweis zur Begründung des Bestehens oder Nichtbestehens eines Anspruchs werden. Überall im Bereich des Einsatzes als Beweismittel ist ein Mindestmaß an wissenschaftlich fundiertem Vorgehen, Transparenz und Nachvollziehbarkeit durch die K.I. einzuhalten und nachzuweisen. Dabei müssen die methodischen Mittel dem aktuellen wissenschaftlichen Kenntnisstand des Fachgebiets entsprechen. Generell ist die Frage der Transparenz im Hinblick auf die eingeflossenen Daten und der zugrunde gelegten Regeln ein wesentliches Problem. Im Rahmen eines Verfahrens ist es denkbar, dass bereits die Sachverhaltsfeststellung durch Erkenntnisse, die von einer K.I. gewonnen wurden, unmittelbar beeinflusst wird.

[47] Bereits die Nutzung von K.I. als Informationsquelle ist kontrovers, v.a. hinsichtlich der Aussagekraft. So hat sich das EU-Parlament für einen rechtlichen Status von Robotern als einer "elektronischen Person" ausgesprochen. EU-Parlament (2017). Darauf haben über 200 EU-Wissenschaftler/-innen und Unternehmer/-innen (v.a. Robotik-Experten) in einem offenen Brief reagiert: http://www.robotics-openletter.eu/

[48] Angwin/Larson/Mattu/ Kirchner (2016).

Im Einzelfall ist hier mit erheblichem Aufwand bei der Nachvollziehbarkeit der Ergebnisse zu rechnen, um die Beweisführung nachvollziehen zu können. Erforderlich sind vor allem gesetzliche Regelungen, die in allen Widerspruchs-/Berufungs-/Revisionsfällen zwingend eine Überprüfung durch höhere Instanzen, welche ausschließlich von Menschen besetzt sind, vorsieht. Zusätzlich sind die Richter/-innen (ebenso Justizmitarbeiter/-innen, Anwälte, Staatsanwälte, Notare) entsprechend aus- und fortzubilden und hinsichtlich der Gefahren einer „blinden" Technikgläubigkeit zu sensibilisieren. Wir neigen zum jetzigen Zeitpunkt zu der stärkeren Forderung, K.I. in richterlichen Verfahren generell zu verbieten.

Prinzip Nr. 9 beschreibt die Verantwortlichkeiten bei fortgeschrittener K.I. (*Advanced AI Systems*). Es wird ausgeführt, dass K.I-Konstrukteure für beabsichtigte und unbeabsichtigte Folgen (*Implications*) als „Beteiligte" („Stakeholder") Verantwortung übernehmen sollen. Die Formulierung impliziert, dass Systeme, die weniger *advanced* sind, dieser Verantwortung nicht unterliegen.[49] Die Frage nach Verantwortlichkeit (d.h. auch nach Kausalitäten) gestaltet sich bei komplexen Systemen schwierig. Dieses Problem wird bei globalisierten Systemen potenziert. Darüber hinaus zeigen analoge Fälle (z.B. Unfälle in Industrieanlagen), dass beteiligte finanzstarke und mächtige Interessensgruppen die Durchsetzung von Haftungsansprüchen (*Liabilities*) in der Regel engagiert zu verhindern versuchen. Die multilateralen Verhandlungen nach 1992, v.a. im Rahmen des Übereinkommens über die biologische Vielfalt (CBD) (z.B. beim Biosafety-Protokoll) haben gezeigt, dass Haftungsfragen (insbesondere die Regelung der Beweislast) ein entscheidendes und pragmatisches Regulierungsinstrument sind. Die Frage, wer wann für welchen Schaden haftet, hat, z.B. über Strategien des Risikomanagements und erforderlicher Due-Diligence-Prüfungen, Rückwirkungen bis auf die Ebene von Investitionsentscheidungen.

[49] Demgegenüber sieht der wissenschaftliche Dienst des EP grundsätzlich immer die Notwendigkeit, Verantwortlichkeiten ex-ante zu definieren: EPRS (2016).

Prinzip Nr. 12 fordert das Recht auf persönlichen Datenschutz, beschränkt es aber auf solche Daten, die von den Nutzern generiert werden. Demgegenüber bleiben solche Daten außen vor, die über diese Nutzer gesammelt werden. Damit bleibt der weitaus größte Teil der Daten außerhalb der Kontrolle derjenigen, über die diese Daten etwas aussagen. Mit Blick auf K.I. beinhaltet dies auch solche personenrelevanten Informationen, die die K.I. durch Assoziation/Verknüpfung/Kombination von Daten erzeugt.[50] Das barrierearme und unbeschränkte Recht auf Zugriff zu allen einen selbst betreffenden, personenbezogenen Daten muss gewährleistet werden. Dieses Recht muss die explizite, vorherige Zustimmung zu Erhebung und Vernetzung ebenso beinhalten wie Vollständigkeit, Transparenz und das Recht auf Löschung (das Recht, vergessen zu werden), so es hiergegen keine übergeordneten rechtlichen Gründe gibt (z.B. Verschleierung einer Straftat).[51]

Der Schutz der Privatsphäre und personenbezogener Daten steht im Mittelpunkt von Prinzip Nr. 13, das besagt, dass der Einsatz von K.I. menschliche Rechte beschneiden darf, dies aber nicht in einer Weise, die als "*unreasonably*" charakterisiert wird. Eine potentielle Einschränkung der individuellen und kollektiven Freiheit der Menschen muss immer im Zusammenhang mit einer potentiellen Gefährdung anderer Grundwerte bewertet werden. Grundsätzlich sind Datenschutz und Privatheit wesentliche, garantierte Grundwerte der Gesellschaft, die es gerade im Bereich von K.I. so zu gestalten gilt, dass diese weder teilweise noch als Ganzes ohne existentielle Not eingeschränkt werden. Hierfür sind folgende Grundlagen notwendig:

[50] Neben den Daten, die Menschen von sich direkt oder indirekt preisgeben (etwa durch soziale Medien, Nutzung von Suchmaschinen oder dem Internet der Dinge) und gegen die sich ein jeder wehren könnte, ist die Bedeutung einer Sammlung von Daten durch Sensoren (angefangen mit Überwachungskameras) nicht zu unterschätzen. Auch aktuelle Entwicklungen in Richtung "Neuro-Daten" sind zu beobachten.

[51] Die generelle Forderung nach dem Recht auf Vergessen wirft rechtliche Probleme auf. Beispielsweise versuchten ehemalige Mitarbeiter des Staatssicherheitsdienstes der DDR die Löschung ihrer Daten durchzusetzen, womit sie vor Gericht nachvollziehbarerweise scheiterten.

- Privacy-by-Design bzw. auf diesem Konzept basierende Gestaltungsvarianten, die bereits in den technischen Designprozess inhärent Privatheit und Datenschutz effektiv berücksichtigen und technisch umsetzen bzw. ermöglichen.
- Opt-Out- und Opt-In-Optionen sollten grundsätzlich verpflichtend sein. Dies muss auch gerade die Verknüpfung von Daten beinhalten.
- Die grundlegende Entscheidung über die Verwendung von Daten sollte beim einzelnen Menschen liegen. Dazu bedarf es einer klaren Regelung, die auch BigData transparent für den Einzelnen nachvollziehbar gestaltet.
- Freedom to Infomation (Informationsfreiheit, -zugangsfreiheit und -transparenz) sollte gesetzlich so ausgestaltet werden, dass auch verknüpfende Daten einschließlich der dahinterliegenden Algorithmen von den Sammlern der Daten bereitgestellt werden müssen.
- Mittels des Haftungs- und Strafrechts sind ggf. Überwachungs- und Datensammlungen durch Dritte so auszugestalten, dass der Einzelne wirksamen, rechtlichen Schutz davor bekommt.

Prinzip Nr. 14 möchte Menschen „empowern“ und bekräftigt das Ziel, dass K.I. so vielen Menschen wie möglich nutzen solle. „Empowerment“ dürfte hier so zu verstehen sein, dass Menschen in ihrer Handlungsfähigkeit gestärkt werden sollen. Zunächst ist grundlegend, dass der Mensch Subjekt seines Handelns bleibt - mit einem Höchstmaß an Selbstbestimmung, und sein Zugang zu wirtschaftlicher, gesellschaftlicher und politischer Teilhabe durch K.I. nicht geschwächt, sondern tendenziell gestärkt wird. Um zu gewährleisten, dass so viele Menschen wie möglich von K.I.-Technologien profitieren, sind entsprechende wirtschaftliche, soziale und politische, aber auch kulturelle Weichenstellungen und Entscheidungen erforderlich.

Angesichts der großen Verheißungen von K.I. ist die politische Forderung legitim, dass K.I. u.a. sichtbar dazu beiträgt, wirtschaftliche und soziale Ungleichheit zwischen Menschen zu verringern. Die gewählte Formulierung des Prinzips birgt jedoch die Gefahr, dass das

Allgemeinwohl lediglich als Summe maximierten Eigennutzes verstanden wird. Stattdessen muss gefordert werden, dass K.I. grundsätzlich dem Allgemeinwohl dienen muss.[52]

Prinzip Nr. 15 fordert, dass "economic prosperity created by AI" breit geteilt werden soll. Als politisches Postulat setzt das in letzter Konsequenz ein anderes Wirtschaftssystem voraus, was jedoch im Text nicht gefordert wird und daher wohl auch nicht intendiert ist. Bei aller Vorsicht in den Voraussagen, zeigen vorliegende erste Studien (z.B. der Universität von Oxford), dass die durch Vernetzung und K.I. ermöglichte vierte Automatisierungswelle der Produktion auch unter Berücksichtigung vieler neu entstehender Arbeitsplätze zur „Netto"-Vernichtung von Millionen Arbeitsplätzen führen wird.[53] Davon werden in erster Linie weniger qualifizierte Menschen und Frauen betroffen sein, für die keine Alternativen entstehen.[54]

Prinzip Nr. 18 befasst sich mit der Gefahr eines Wettrüstens von tödlichen autonomen Waffensystemen. Diese werden nicht generell geächtet. Lediglich vor der Gefahr eines Wettrüstens wird gewarnt. Die USA, das Vereinigte Königreich, Russland, China, Israel und Südkorea entwickeln bereits solche Waffensysteme und an der Grenze zu Nordkorea sind sie bereits stationär im Einsatz. Angesichts der dynamischen Beschleunigung bei der Entwicklung von tödlichen autonomen Waffensystemen fehlen klare Aussagen, wie ein Wettrüsten verhindert werden soll. Der Einsatz autonomer Einheiten in Kriegs- oder Kampfsituationen ist wegen des Fehlens ethischen Verhaltens und möglicher Fehlsteuerung ein Risiko, auch für die Anwender.[55] Darüberhinaus zeigt die Formulierung von Prinzip 18, dass auch militärische Anwendungen von K.I. keinen „Verbotsreflex" auslösen.[56]

[52] Im Rahmen der geplanten umfassenden europäischen Initiative (EU-KOM (2018), Heise(2018)) wird eine „K.I. on demand"-Plattform vorgeschlagen, die neben neuen K.I.-Exzellenz-Zentren und „digitalen Innovationshubs" eine Beteiligung von kleinen und mittleren Unternehmen an der Entwicklung ermöglichen sollen.
[53] McKinsey (2017): Im Mittelwert droht bis 2030 brutto der Wegfall von rund 15% (maximal bis zu 30%) aller Arbeitsplätze weltweit (rund 400 Mio. Jobs; maximal 800 Mio. Jobs. Frey/Osborne (2013).
[54] Frey/Osborne (2013), S.36ff. WEF (2018).
[55] Scott, et al (2018) und Brundage, et al (2018).
[56] Dennoch ist das Future of Life Institut engagiert. Sie haben bei den Verhandlungen über die Convention on Certain Conventional Weapons / Group of Governmental Experts on Lethal Autonomous Weapons Systems, im November 2017 ein "Schock"-Video vorgestellt, das die Gefahren autonomer Waffensysteme drastisch illustriert.

Führende K.I.-Forscher sehen darüber hinaus die Gefahr aggressiver Angriffe bereits weit unter der Schwelle von Waffensystemen schnell anwachsen (z.B. durch Sabotage von Fahrzeugen oder Großanlagen, sowie die gezielte Unterhöhlung von Demokratien und staatlicher Autorität), da die Kosten für solche Einsätze beständig sinken und der Angreifer nur äußerst schwer ermittelbar ist.[57] [58]

3.4 Längerfristige Probleme

Prinzip Nr. 19 fordert, angesichts eines fehlenden Konsenses, den Verzicht auf Annahmen hinsichtlich der Frage, ob und welche technischen Grenzen es für die weitere K.I.-Entwicklung gibt, und vermeidet so das Definieren absoluter „roter Linien" für die weitere K.I.-Entwicklung. Demgegenüber verlangt das Vorsorgeprinzip genau das. Wenn wir keine Kenntnis darüber haben, ob und wann eine starke K.I. entsteht, und auch nicht wissen, ob und mit welchen Folgen diese gefährlich sein könnte, müssen Gesellschaften und Staaten sich vorab über rote Linien verständigen und diese wirksam durchsetzen.

Prinzip Nr. 20 führt aus, dass fortgeschrittene K.I. die Zukunft des Lebens auf dem Planeten Erde tiefgreifend verändern kann.[59] Einer solchen fundamentalen Möglichkeit wird die Forderung entgegengestellt, „angemessen" zu planen und zu entwickeln. „Angemessen" ist ein unbestimmter Rechtsbegriff - wer bestimmt, was „angemessen" ist und was die Maßstäbe sind? Das Hauptproblem des Prinzips ist, dass nicht gefragt wird, ob es ohne Vorlage zwingender Gründe überhaupt zulässig ist, tiefgreifende Veränderungen in der Geschichte des Lebens auf der Erde in Kauf zu nehmen, deren Richtung und Ausmaß Menschen nicht kennen und voraussehen können. Dem Prinzip liegt offenbar das Verständnis zu Grunde, dass alles was

[57] Brundage, et al (2018).
[58] Das Thema wird auch adressiert in: Executive Office of the President (2016), S. 42, Empfehlung 22 und 23.
[59] Das verwendete Wort „history" ist dabei missverständlich und passt hier nicht, da Geschichte rückwärtig ist. "Course of life" wäre geeigneter. Zumindest diskussionswürdig ist auch die zu Grunde liegende teleologische Vorstellung eines Ziels menschlicher Existenz.

möglich ist, auch gemacht wird. Wenn also der Mensch fähig ist, starke K.I. zu entwickeln, dann wird er das auch tun, einfach weil er es kann.[60]

Die Geschichte zeigt demgegenüber (wenn auch wenige) Ausnahmen, in denen sich Gesellschaften, Staaten oder Staatengemeinschaften dafür entschieden haben, bestimmte Technologiepfade nicht weiter zu verfolgen, da die mit ihnen verbundenen Risiken als nicht akzeptabel angesehen wurden (z.B. Kernspaltung oder fossile Brennstoffe als Energiequelle, biologische und chemische Waffen, FCKW u.a. als Kühlmittel). Dies passierte jedoch nur dann, wenn es weite Teile der Bevölkerungen (v.a. die Eliten) mit deutlicher Mehrheit ablehnten, bestimmte Gefahren (weiterhin) in Kauf zu nehmen, die mit der Anwendung dieser Technologien verbunden waren/sind. Das geschah in der Regel erst nach Eintritt unerwünschter Folgen. Auch gibt es Beispiele dafür, dass der „Technologieverzicht“ eines Akteurs nicht dazu geführt hat, dass andere dem Beispiel gefolgt sind, sondern im Gegenteil ihren Nutzen daraus gezogen haben (derzeit: friedliche Nutzung der Atomkraft). Dies kann aber keine Rechtfertigung für grobe Fahrlässigkeit im Hinblick auf das Vorsorgeprinzip sein.

Prinzip Nr. 21 fordert Planungs- und Risikominderungsbemühungen, die den zu erwartenden Auswirkungen existenzieller Risiken und möglicher Katastrophen angemessen sind. Wichtig ist festzuhalten, dass das Prinzip impliziert, dass es existenzielle Risiken gibt und Katastrophen möglich sind. Denen soll mit besten „Bemühungen“ begegnet werden. „Bemühungen“ sind aber nicht gleichbedeutend mit Erfolg. D.h. mit der Formulierung des Prinzips wird in Kauf genommen, dass diese Bemühungen, zwar im Umfang dem Risiko angemessen sein müssen, aber in letzter Konsequenz scheitern können.

Im Gegensatz zu Prinzip Nr. 21 sind die Formulierungen in Prinzip Nr. 22 dem Gegenstand angemessen. Dennoch lassen auch sie zu, dass F+E und Anwendung „starker“ K.I. fortgesetzt werden können, sofern sie „strengen Sicherheits- und Kontrollmaßnahmen unterliegen“. Die

[60] Vgl. Beck (1986).

Möglichkeit, eines absoluten Banns, also eines generellen Verbots „starker“ K.I. wird nicht erwogen. Selbst wenn ein solcher Bann durchsetzbar wäre, bestünde immer noch die Gefahr, dass irgendwo auf der Welt Menschen, um des Vorteils willen die Verbote ignorieren würden, um z.B. politische und/oder weltanschauliche/religiöse Macht zu erlangen. Nicht nur der russische Präsident Wladimir Putin hat bereits öffentlich geäußert, dass diejenige Macht, die als erste über eine „starke“ K.I. verfügt, die Welt beherrschen wird.[61] Hierbei ist insbesondere die globale Vernetzung zu berücksichtigen, mit denen solche Systeme Zugriff auf „Exekutivmittel“ wie Waffen (unmittelbare Bedrohung), als Waffen nutzbare Einrichtungen, wie z.B. Atomkraftwerke und kritische Infrastrukturen, wie z.B. Stromnetze (mittelbare Bedrohung) erlangen können.

In Prinzip Nr. 23 wird direkt von einer „Superintelligenz“ gesprochen, die „im Dienste weit verbreiteter ethischer Ideale der gesamten Menschheit und nicht nur einem Staat oder einer Organisation dienen soll“. Zunächst gilt jedoch: Wir Menschen leben nicht in einer weltweiten, idealen (paradiesischen) Gesellschaft, die es vermutlich, ob des permanenten Wandels allen Lebens auch niemals geben wird. Hier stellt sich die Frage: Warum sollen wir die Erschaffung einer Superintelligenz zulassen wollen? Würden wir dann nicht zu Gläubigen einer Religion, die den Menschen als eine notwendige Zwischenstufe der Evolution hin zu höheren, digitalen Wesen begreift?[62] Das können und dürfen wir nicht wollen. Eine Superintelligenz könnte im Rahmen ihrer selbständig weiterentwickelten Grundprogrammierung eventuell über das weitere Schicksal der Menschheit entscheiden. Das müsste nicht bedeuten, dass die Menschheit vernichtet oder versklavt würde, aber wir Menschen hätten dann nur noch „geliehene/gewährte“ Mitwirkungsmöglichkeiten. Ein gewaltsames Durchsetzen menschlicher Interessen gegen die Superintelligenz wäre mit großer Wahrscheinlichkeit zum Scheitern verurteilt.

[61] RT.com (2017).
[62] Vgl. Hariri (2017), S. 497ff. und Dotzauer (2017).

Dabei bleibt weiterhin offen, was diese „weit verbreiteten Ideale“ überhaupt sind. Beispielsweise sind in Teilen der Welt „reich sein“ oder „konsumieren“ oder „Genuss“ oder „Spaß“ weit verbreitete Ideale. Aber auch wortgetreu religiöse Vorschriften umzusetzen, ist ein weit verbreitetes Ideal in Teilen der Welt. Dann gibt es auch sich weitgehend ausschließende weltweit verbreitete Ideale, wie „das Selbstbestimmungsrecht der Frau abzutreiben“ und „der Schutz des ungeborenen Lebens ab der Empfängnis“. Für beides finden wir über eine Milliarde Anhänger/-innen. Selbst die konsequente Durchsetzung der UNO Menschenrechtserklärung wird von einer großen Zahl von Menschen immer noch zumindest in Teilen abgelehnt. Was also ist die legitimierende Basis des Begriffs „weit verbreitete ethische Ideale“?

3.5 Fazit

Die Verfasser der Asilomar-Prinzipien und die VDW-Studiengruppe sind sich einig, dass K.I. das Leben auf der Erde tiefgreifend verändern wird und dass von der Schaffung einer starken K.I. auszugehen ist. Zumindest ansatzweise sind sie sich auch einig, dass sie ein wesentliches Gefahrenpotential in Bezug auf den weiteren Verlauf der Menschheitsgeschichte birgt. Die Asilomar-Prinzipien sind ein hervorragend geeigneter Ansatzpunkt für Diskussionen darüber, wie das Potential der künstlichen Intelligenz in den kommenden Jahren genutzt werden kann. Die Prinzipien sind jedoch weder ein geeigneter normativer Rahmen für die erforderliche Definition absoluter Grenzen hinsichtlich Erforschung, Entwicklung und Anwendung von K.I. noch für die Durchsetzung solcher Grenzen zur Gefahrenabwehr.

Als problematisch wird von der VDW-Studiengruppe vor allem angesehen, dass die Prinzipien auf Basis einer (impliziten) exklusiven Utopie-Prämisse entstanden sind, die davon ausgeht, dass grundsätzlich eine grenzenlose Technologieentwicklung bzw. -anwendung möglich ist, die nur in Einzelfällen einer Regulierung bedarf. So war in der Gestaltung der Asilomar-Prinzipien 90% Zustimmung erforderlich, um ein regulierendes Prinzip aufzunehmen. Bei einer Prämisse auf Gefährdungspotential, die dem EU-Vorsorgeprinzip entspräche, würde demgegenüber die Zulassung von technologischer Entwicklung bzw. Anwendung 90% Zustimmung erfordern. Das

in der EU geltende Vorsorgeprinzip sollte daher als Leitschnur für alle weiteren Diskussionen dienen.[63]

4. Handlungsempfehlungen

Die VDW Studiengruppe Technikfolgenabschätzung der Digitalisierung ist von der Notwendigkeit überzeugt, dass eine breite, zielgerichtete gesellschaftliche und wissenschaftliche Diskussion über die Bewältigung der Herausforderungen, vor die uns F+E und Anwendungen künstlicher Intelligenz stellen, erforderlich ist. Hierzu gibt es erste Ansätze, die jedoch unseres Erachtens noch zu sehr von technischer Begeisterung und den bereits sichtbaren und zukünftig noch deutlich wachsenden wirtschaftlichen Produktivitätsgewinnen dominiert wird. Die grundsätzlich positive Sichtweise vieler Stimmen ist oftmals von der Überzeugung getragen, dass es ohnehin nur darum gehen kann, unerwünschte Folgen von K.I. zu vermeiden oder zumindest zu verringern, nicht aber darum, ob zumindest bestimmte Anwendungen schwacher K.I. und die Erzeugung einer starken K.I. grundsätzlich verboten werden soll. Insgesamt überwiegt der Eindruck, dass die Versprechungen der K.I. nicht durch zweifelnde oder pessimistische Betrachtungen getrübt werden sollen.

Gründe hierfür liegen aber auch darin, dass die Mehrzahl der denkbaren negativen Folgen noch nicht eingetreten oder bekannt ist und diese im Fall starker K.I. vermutlich in einer noch Jahre oder Jahrzehnte dauernden Zukunft liegen. Die Gefahren sind abstrakt, unsichtbar und größtenteils unbekannt. Unsere Vorstellungskraft über diese Zukunft ist geprägt durch Science-Fiction-Literatur und -Filme. Warnende Stimmen werden daher im Augenblick noch nicht ernst genommen. Die Zukunft ist scheinbar weit weg - zu weit für politische,

[63] Wenn ein Risiko bedeutet, dass katastrophale oder existenziell bedrohliche Ereignisse in Folge von F+E oder Anwendung von K.I.-Systemen eintreten können und diese Risiken nicht vernachlässigbar sind, dann können die potentiellen Risiken nicht Gegenstand von „Planungs- und Risikominderungsbemühungen“ sein, sondern müssen unmittelbar und unverzüglich alle erforderlichen Maßnahmen zur Gefahrenabwehr auslösen.

gesellschaftliche Debatten, die sich um die Probleme von heute und naher Zukunft drehen. Wir beleuchten im Weiteren die wichtigsten Anknüpfungspunkte für das, was zu tun ist.

Wissenschaftlicher Diskurs und Forschung zur Technikfolgenabschätzung von K.I.

Technikfolgen und Sicherheitsfragen adressierende Grundlagen-Forschung, vor allem im Hinblick auf die Mensch-Maschine-Interaktion im Speziellen und die Maschine-Umwelt-Interaktion im Allgemeinen, gibt es bisher nur in Ansätzen. Es besteht jedoch ein erheblicher öffentlicher Forschungsbedarf, Forschungsprozesse müssen hinsichtlich Risikofolgen-abschätzung, gesellschaftlichen Interdependenzen und zukünftigen Entwicklungsschritten anwendungsorientiert und begleitend zur technischen Entwicklung gestaltet und intensiv vorangetrieben werden.

Die hierfür benötigte Forschungsförderung (aus öffentlichen Mitteln) muss inter- und transdisziplinäre Grundlagenforschung in den Bereichen Recht, Ethik, Sozial- und Wirtschafts-wissenschaften, Informatik, aber auch Medienwissenschaften, Technik und Psychologie vor allem hinsichtlich möglicher technischer Designs sowohl spezifisch als auch übergreifend (auch zur Durchsetzung des Schutzes der Grundrechte) auflegen.[64] [65] Ziel muss u.a. sein, für den technischen Designprozess konkrete Disziplinen-übergreifende Richtlinien, inkl. der Programmierung und Technik, zu entwickeln.

Kooperationen und formalisierte, transdisziplinäre wissenschaftliche Kommunikations-prozesse in Forschung und Lehre (z.B. durch gemeinsame Veranstaltungen und Publikationen) werden dabei eine rasche Durchdringung, gegenseitige Befruchtung und öffentliche Verbreitung des gewonnenen Wissens befördern.

[64] Executive Office of the President (2016), S. 42, Empfehlung 18 sieht auch die Schule in der Pflicht.

[65] Zum Beispiel sollten Modellversuche, die in einem klaren abgegrenzten Raum stattfinden und somit eine Überprüfung von Wirkungsmechanismen unter Begrenzung potenziell irreversibler Folgen, unter Einbeziehung unabhängiger Prüfungsmechanismen ermöglicht werden.

Gesellschaftlicher und politischer Diskurs

Bürgerinnen und Bürger, Entscheidungsträger und Multiplikatoren müssen mit relevanten, wissenschaftlich fundierten Informationen zu K.I. vertraut gemacht werden. Es bedarf erheblicher Anstrengungen, um die erforderlichen Informationen aufzuarbeiten und in den einschlägigen Foren zu diskutieren. Insbesondere politische, juristische und wirtschaftliche Entscheidungsträger müssen in die Lage versetzt werden, informierte Entscheidungen zu treffen. Dies gilt auch für Medien und gesellschaftliche Institutionen sowie relevante, nachgeordnete staatliche Bereiche (v.a. Sicherheitsbereich, Verbraucherschutz). Auch sollen Konferenzen und Veranstaltungen, die sich hierfür anbieten, genutzt werden, um auf die wesentlichen Fragestellungen zum Thema K.I. hinzuweisen.

Wir wollen dazu beitragen, die notwendigen Diskussionen und Verhandlungen anzustoßen, um denkbare Gefahren von K.I. (insbesondere durch starke K.I.) effektiv zu verhindern oder wo das nicht möglich/nötig ist, zu minimieren. Dabei ist zu berücksichtigen, dass ein Großteil der F+E-Aktivitäten zu K.I. nicht unter staatlicher Kontrolle, in einem globalen Wettbewerb stattfindet und zudem militärisch orientierte K.I.-Forschung zumindest nur begrenzt demokratischer Kontrolle unterliegt. Insbesondere die Einbindung der Zivilgesellschaft und des Einzelnen bei der Mit- und Ausgestaltung dieser bedeutenden Fragen kann deshalb ein entscheidender Schritt sein, um die unten skizzierten erforderlichen Maßnahmen zum Erfolg zu führen.

Regulierung

Wie für alle Forschungsgebiete gilt auch für die F+E von K.I., dass sie normativen Prinzipien (ethischen und rechtlichen) folgen muss.[66] Grundlage müssen aus der Sicht der VDW die Allgemeine Erklärung der Menschenrechte, bzw. deren Kodifizierungen in je nationalem Recht sein. Im Weiteren sind bereits geltende Rechtsnormen wie das Vorsorgeprinzip explizit auf technische Entwicklungen auszuweiten und rechtlich verbindlich zu verankern

[66] NSTC (2016).

(Vorsorgeprinzip 2.0).[67] Risiko- und Technikfolgenabschätzung sind in diesem Sinne verbindlich zu gestalten. Es bedarf kodifizierter Regeln, die alle einschlägigen Rechtsfragen auf allen erforderlichen, nationalen und internationalen Rechtsebenen umfassen,[68] einschließlich Verboten oder Moratorien. Normierungen müssen einerseits innerhalb der jeweiligen Anwendungskontexte, anderseits darüber hinaus gesamtgesellschaftlich erfolgen. Das beinhaltet auch die Forderung, das Vorsorgeprinzip dort rechtlich weiterzuentwickeln, wo dies im Sinne einer umfassenden Gefahrenabwehr erforderlich ist. Da die Ausarbeitung international gültiger und mit Durchsetzungsinstrumenten ausgestatteter Rechtssysteme eher Jahrzehnte als Jahre braucht, müssen Arbeiten hierzu unverzüglich begonnen werden.

Es bedarf effektiver staatlicher (und multilateraler) Strukturen und Sanktionsmechanismen, um sicherzustellen, dass K.I. jederzeit kontrollierbar ist. Dies gilt für alle Phasen von F+E und Anwendung. Vor allem Markteinführungen dürfen erst nach Abschluss hinreichender Sicherheitsbewertungen erfolgen. Hierbei sind beispielsweise umfangreiche Prüfungen in lebensnahen Szenarien notwendig. Technische Expertise ist dabei unterstützend erforderlich, darf aber ebenso wenig wie wirtschaftliche Interessen maßgeblichen Einfluss in Regulierungsfragen haben.[69] Auch bedarf es einer technisch informierten, umfassenden, weltweit funktionierenden, demokratischen Kontrolle von Forschung und Entwicklung in diesem Bereich.

Die oben beschriebene Forschungsförderung muss auch die Identifikation und wissenschaftliche Ausarbeitung eventuell erforderlicher, ergänzender rechtlicher Modelle der

[67] Das Vorsorgeprinzip muss weiterentwickelt werden. Vorsorgeprinzip und Innovationen schließen sich nicht aus. Im Gegenteil: Gerade die Zukunftsvorsorge sollte Anlass zu Innovationen sein, die die Nachhaltigkeit voranbringen. Das Offenhalten und die Förderung risikoärmerer Alternativen sowie die Schaffung von multiplen Optionen für die Zukunft kommt viel zu kurz. Forschung und Innovation sollten sich hier ihrer Verantwortung und Werthaltungen bewusst werden und früh mit der Gesellschaft in den Dialog über Technikfolgenabschätzung treten. Bei Risiken und Chancen sollten auch die unterlassenen Alternativen betrachtet werden.

[68] Erste Überlegungen hierzu: EPRS (2016).

[69] Weniger kritisch: Executive Office of the President (2016), S. 40, Empfehlung 5 und 6.

effektiven Regulierung adressieren, die helfen, die Wirkungen von K.I. frühestmöglich zu erkennen und alle erforderlichen Reaktionen zeitnah zu erzwingen.

Gremien aus Expert/-innen und Vertreter/-innen der Zivilgesellschaft sollten dies unterstützen. Darüber hinaus sind ethische Selbstverpflichtungen für den Umgang mit K.I. in F+E und Anwendung erforderlich.

Die EU-Kommission beginnt zeitgleich mit der Veröffentlichung dieses Papiers mit ihrer umfassenden europäischen Initiative zu K.I., einschließlich Plänen für eine europäische K.I.-Allianz und eines ehrgeizigen Ansatzes zu K.I., der die EU zu einem „Führer der K.I.-Revolution" machen soll.[70] Dagegen klingen die Hinweise der Strategie auf mögliche Verwerfungen wie erzwungene Zugeständnisse.[71] Es gibt auch erste Forderungen, sowohl des EU-Parlaments (Empfehlungen an die KOM vom 27. 1. 2017) als auch aus der EU-Kommission selbst, unverzüglich „alle erforderlichen Arbeiten zu beginnen, um ein international anerkanntes legales und ethisches Rahmenwerk für Design, Produktion, Nutzung und Governance von K.I. zu entwickeln."[72] Dies wird auch von wissenschaftlicher Seite konkret gefordert: Die EU soll unverzüglich ein K.I.-Charta entwickeln und auf die Entwicklung einer globalen Charta hinarbeiten.[73]

K.I.-inhärente Sicherheitsmechanismen

Für die Erschaffung jeglicher K.I. gilt darüber hinaus: Grundvoraussetzung ist die unwiderrufliche Programmierung ethischer Prinzipien, die in allen vorstellbaren Operationsmodi funktional bleiben. In Situationen, in denen dies dennoch nicht (mehr) gewährleistet ist und menschliches Eingreifen gefordert ist, um Schaden zu verhindern, müssen alle erforderlichen Aktionen jederzeit zeitgerecht und möglichst vorbeugend möglich

[70] EU-KOM (2018), S. 14
[71] Ebenda.
[72] EU-DG R+I (2018)
[73] Vgl. hierzu im einzelnen Metzinger (2018).

sein. Die wichtigste Leitlinie muss dabei sein, dass K.I. unter keinen vorstellbaren Umständen einem Menschen Schaden zufügen kann. Dies entspricht den Robotergesetzen von Asimov und ist unbedingte Voraussetzung für „Nützlichkeit" der K.I. Asimov selbst charakterisiert seine Gesetze als notwendig aber nicht hinreichend.[74] Ethische Prinzipien für Algorithmen sind auch deshalb zumindest problematisch, da diese keine „Ich-Persönlichkeit" haben, die die Erfahrung von Geburt, Freude, Schmerz, Krankheit und Tod haben kann. Sollte dies eines Tages doch geschehen, würden wir ganz anderen Herausforderungen gegenüberstehen.

Sonderfall tödlicher autonomer Waffensysteme

Hinsichtlich der militärischen Nutzung von tödlichen autonomen Waffensysteme muss auf dem deutsch-französischen (Non-)Working Paper zu den ersten förmlichen UN-Verhandlungen über tödliche autonome Waffensysteme aufgebaut werden.[75] Darin schlagen beide EU-Mitgliedstaaten gemeinsam eine politische Erklärung auf UN-Ebene vor, die erste Schritte zu einem internationalen Protokoll vorsehen, im Rahmen des „Übereinkommens über das Verbot oder die Beschränkung des Einsatzes bestimmter konventioneller Waffen, die übermäßige Leiden verursachen oder unterschiedslos wirken können."[76] Dies würde faktisch einen Bann bedeuten. Auch hier wird der Erfolg v.a. davon abhängen, dass alle Staaten zu der Überzeugung gelangen, dass tödliche autonome Waffen in letzter Konsequenz auch nicht durch ihre Nutzer sicher steuerbar sind. Dann würden nur noch Produzenten von tödlichen autonomen Waffensystemen ein Interesse haben, ein wirksames Verbot zu verhindern. Wie weit dieser Weg ist, zeigt das Papier des Leiters der Network Science Division of the Army Research Laboratory, Dr. Alexander Kott, der für die USA massive Forschungsanstrengungen auf dem Gebiet der Entwicklung von autonomen Waffensystemen fordert, da nur autonome

[74] Asimov (1950).

[75] Die Verhandlungen fanden im November 2017 im Rahmen der Convention on Certain Conventional Weapons statt: Group of Governmental Experts (GGE) on lethal autonomous weapons systems (LAWS). Group of Governmental Experts of the High Contracting Parties (2017) und International Committee of the Red Cross (2004). Zur Position der V.R. China siehe: Kania (2018).

[76] United Nations (1980).

Waffensysteme in der Lage wären, angemessen auf das zukünftige Zusammenwirken des „autonomen Internet der Kampf Dinge“ auf dem Schlachtfeld zu reagieren.[77]

Einladung zum Dialog

Die vor uns liegenden Aufgaben lassen sich nur gemeinsam bewältigen. Wir bieten deshalb allen unsere Unterstützung an, die bereit sind, die Chancen und Möglichkeiten von K.I. nur so weit zu nutzen, wie es nicht menschliche Gesundheit, Leben oder die Umwelt gefährdet und dem Gemeinwohl nicht schadet. Unterhalb der Schwelle existentieller Risiken wird es berechtigterweise eine intensive, gesellschaftliche und politische Auseinandersetzung darüber geben, was dem Gemeinwohl nutzt bzw. was ihm zumindest nicht schadet oder als schädlich anzusehen ist. Wir freuen uns darauf, hierzu in einen breiten Dialog einzutreten.

[77] Kott (2018).

5. Literatur

Angwin, Julia/ Larson, Jeff/ Mattu, Surya/ Kirchner, Lauren (2016): Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. In: ProPublica, May 23, 2016; https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Asimov, Isaac (1950): I, Robot. Gnome Press.

Bahr, Egon/ Lutz, Dieter S. (Hrsg.) (1992): Gemeinsame Sicherheit. Idee und Konzept. Bd. I: Zu den Ausgangsüberlegungen, Grundlagen und Strukturmerkmale Gemeinsamer Sicherheit, Baden-Baden.

Beck, Ulrich (1986): Risikogesellschaft. Auf dem Weg in eine andere Moderne. Erstausgabe, 1. Auflage. Suhrkamp, Frankfurt am Main.

BITKOM (2017): Künstliche Intelligenz verstehen als Automation des Entscheidens – Leitfaden, Berlin.

Brundage, Miles, et al (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation;https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf.

Centre on Impact of AI and Robotics (der UNSW Sydney) (2018): Open Letter to Professor Sung-Chul Shin, president of KAIST from some leading AI researchers in 30 different countries; https://www.cse.unsw.edu.au/~tw/ciair/kaist.html.

Dotzauer, Gregor (2017): Näher, mein Bot, zu dir. In: Tagesspiegel 12.12.2017.

Eberle, Ute (2015): Sprachassistenten verändern unser Leben. In: Wirtschaftswoche 28.7.15.

EFI (Expertenkommission Forschung und Innovation) (2018): Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit 2018, Berlin.

EPRS (2016): European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 563.501: Ethical Aspects of Cyber-Physical Systems, Scientific Foresight study, Brüssel.

EU-DG R+I (2018): European Commission, Directorate-General for Research and Innovation; European Group on ethics in Science and New Technologies: Statement on Artificial Intelligence, Robotics and "Autonomous" Systems, Brüssel.

EU-KOM (2018): Maximising the benefits of Artificial Intelligence (Version 15 – 27/02/2018). Unveröffentlichtes Arbeitsdokument, Brüssel.

EU-Parlament (2017): European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics; http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0051&language=EN&ring=A8-2017-0005.

Executive Office of the President (2016): Preparing for the Future of Artificial Intelligence. Washington, D.C.

Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, CCW/GGE.1/2017/WP.4; http://undocs.org/ccw/gge.1/2017/WP.4.

Frey, Carl Benedikt/ Osborne, Michael A. (2013): The Future of Employment: How Susceptible are Jobs to Computerisation, Oxford University.

Future of Life Institute (2017): Asilomar AI Principles; https://futureoflife.org/ai-principles/ und Beneficial AI 2017. Conference Schedule; https://futureoflife.org/bai-2017/.

Göttinger Achtzehn (1957): Göttinger Manifest; https://www.uni-goettingen.de/de/text+des+g%c3%b6ttinger+manifests/54320.html.

Group of Governmental Experts of the High Contracting Parties (2017): To the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects. For consideration by the Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS). Submitted by France and Germany, CCW/GGE.1/2017/WP.4, Geneva.

Harari, Yuval Noah (2017): Homo Deus – Eine Geschichte von Morgen, München.

Henk, Malte (2014): Jugend ohne Sex. In: Zeit online vom 15.6.14.

Independent Commission On Disarmament and Security Issues (1982): Common security: A blueprint for survival, Simon and Schuster, New York.

International Committee of the Red Cross (2004): Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Geneva.

Jonas, Hans (1979): Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation. 1. Auflage, Insel-Verlag. Frankfurt am Main.

Kania, Else (2018): Artificial Intelligence: China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems. In: Lawfare, April 17, 2018; https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

Kommission der Europäischen Gemeinschaften (2000): Mitteilung der Kommission: die Anwendbarkeit des Vorsorgeprinzips, KOM (2000) 1 endgültig, Brüssel; https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52000DC0001&from=DE.

Kott, Alexander (2018): Challenges and Characteristics of Intelligent autonomy for Internet of Battle Things in Highly Adversarial environments, Adelphi, MD; https://arxiv.org/ftp/arxiv/papers/1803/1803.11256.pdf.

Krempl, Stefan (2018): Künstliche Intelligenz: EU-Kommission plant umfassende europäische Initiative. In: heise online 26.3.2018.

McKinsey (McKinsey Global Institute), (2017): Jobs Lost, Jobs Gained: Workforce Transitions in a time of Automation.

Metzinger, Thomas (2018): Towards a Global Artificial Intelligence Charter. In: European Parliamentary Research Service: Should we fear artificial intelligence? Brüssel.

Müller, Vincent/ Bostrom, Nick (2013): Future progress in artificial intelligence: A Survey of Expert Opinion. In: Vincent C. Müller (Hrsg): Fundamental Issues of Artificial Intelligence, Berlin.

NSTC (National Science and Technology Council) (2016): The National Artificial Intelligence Research and Development Plan, Washington, D.C.

Oye, Kenneth A., et.al. (2014): Regulating gene drives. Regulatory gaps must be filled before gene drives could be used in the wild. In: Science 17 Jul 2014; http://science.sciencemag.org/content/early/2014/07/16/science.1254287.full.

RT.com (2017): 'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day, 1 Sep, 2017 14:08; https://www.rt.com/news/401731-ai-rule-world-putin/.

Schellnhuber, Hans Joachim (2015): Selbstverbrennung: Die fatale Dreiecksbeziehung zwischen Klima, Mensch und Kohlenstoff, C. Bertelsmann Verlag, München.

Scott, Ben/ Heumann, Stefan/Lorenz, Philippe (2018): Artificial Intelligence and Foreign Policy. Stiftung Neue Verantwortung, Berlin.
TAB (Büro für Technikfolgenabschätzung beim Deutschen Bundestag) (2016): Technologien und Visionen der Mensch-Maschine-Entgrenzung. Sachstandbericht zum TA-Projekt „Mensch-Maschine-Entgrenzungen: zwischen künstlicher Intelligenz und Human Enhancements. Arbeitsbericht Nr. 167, Berlin.

United Nations (1948): Universal Declaration of Human Rights, Paris; http://www.un.org/en/universal-declaration-human-rights/.

United Nations (1966): International Covenant on Economic, Social and Cultural Rights. Adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966. Entry into force 3 January 1976, in accordance with article 27; http://www.institut-fuer-menschenrechte.de/fileadmin/user_upload/PDF-Dateien/Pakte_Konventionen/ICESCR/icescr_en.pdf.

United Nations (1980): Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects (with Protocols I, II and III), Geneva, 10 October 1980; http://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVI-2&chapter=26&lang=en.

United Nations (1992): Agenda 21. Konferenz der Vereinten Nationen für Umwelt und Entwicklung, Rio de Janeiro; http://www.un.org/Depts/german/conf/agenda21/agenda_21.pdf.

United Nations (2015a): Paris Agreement, Paris; https://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreement_english_.pdf.

United Nations (2015b): Transforming our world: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. A/RES/70/1; http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

Université de Montréal (2018): The Declaration: https://www.montrealdeclaration-responsibleai.com/the-declaration.

Verbraucherzentrale Bundesverband (2017): Algorithmenbasierte Entscheidungsprozesse - Thesenpapier des vzbv, Berlin.

Vowinkel, Bernd (2017): Ist der Mensch eine Maschine; https://transhumanismus.wordpress.com/2017/06/14/ist-der-mensch-eine-maschine/.

Walter, Sven (2014): Situierte Kognition. In: Information Philosophie; Heft 2/2014, S. 28-32, Lörrach.

Weizsäcker, Ernst Ulrich von/ Wijkman, Anders (2017): Wir sind dran. Club of Rome: Der große Bericht. Was wir ändern müssen, wenn wir bleiben wollen. Eine neue Aufklärung für eine volle Welt, Gütersloh.

WEF (World Economic Forum in collaboration with The Boston Consulting Group), (2018): Towards a Reskilling Revolution: A Future of Jobs for All, Köln/Genf.

6. Anhänge

6.1 Asilomar AI Principles

Research Issues

1) Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.

2) Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:

- How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?
- How can we grow our prosperity through automation while maintaining people's resources and purpose?
- How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI?
- What set of values should AI be aligned with, and what legal and ethical status should it have?

3) Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.

4) Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.

5) Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

Ethics and Values

6) Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.

7) Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.

8) Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.

9) Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.

10) Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.

11) Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.

12) Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.

13) Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.

14) Shared Benefit: AI technologies should benefit and empower as many people as possible.

15) Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.

16) Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

17) Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.

18) AI Arms Race: An arms race in lethal autonomous weapons should be avoided.

Longer-term Issues

19) Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.

20) Importance: Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.

21) Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.

22) Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.

23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.

6.2 Autoren (Mitglieder der VDW Studiengruppe)

Prof. Dr. Ulrich Bartosch (Politikwissenschaftler) ist Professor für Pädagogik an der Katholischen Universität Eichstätt-Ingolstadt, wo er zu Pädagogischer Theorie und Politischer Ideengeschichte, Hochschulreform und -bildung, Kompetenzbeschreibung und –entwicklung, Inklusion, Partizipation, Schulsozialarbeit sowie Weltinnenpolitik lehrt und forscht. Von 2005-2011 war er Vorsitzender des Fachbereichstages Soziale Arbeit. Er ist Mitglied der Ad-hoc-AG „Anrechnung und Anerkennung digitaler Lernformate“ des Hochschulforums Digitalisierung der Hochschulrektorenkonferenz (HRK) und Leiter des Kooperationsprojekts von Katholischer Universität Eichstätt-Ingolstadt (KU) und VDW „Laudato Si‘ – Die päpstliche Enzyklika im Diskurs der Großen Transformation“. Von 2009 bis 2015 war er Vorsitzender der Vereinigung Deutscher Wissenschaftler und ist seither Vorsitzender des Beirats der VDW.

Prof. Dr. Stefan Bauberger SJ (Physiker, Philosoph) ist Professor für Naturphilosophie und Wissenschaftstheorie an der Hochschule für Philosophie in München. Er ist Angehöriger des Jesuitenordens und Theologe. Er hat in theoretischer Physik promoviert und ist in Philosophie habilitiert. Er hat einige Jahre in der theoretischen Elementarteilchenphysik gearbeitet. Er forscht und lehrt über Grenzfragen zwischen Philosophie und Naturwissenschaft, insbesondere der Physik, im Bereich des Dialogs zwischen Naturwissenschaft und Religion sowie über die Philosophie des Buddhismus, und in den Bereichen Technikphilosophie und Wissenschaftstheorie. Zuvor war er Leiter der Ausbildung des Jesuitenordens in Deutschland. Er ist ZEN-Meister und leitet ein Meditationszentrum.

Tile von Damm (Politikwissenschaftler) ist Leiter des urbanen Forschungsinstituts MOD und Associate Researcher an der TU Berlin. Er ist Urban Expert im EU-Projekt „Orfeo&Majnun“. Zudem ist er Mitgründer und Geschäftsführer von DiMed zur Sicherstellung der ruralen medizinischen Grundversorgung. Seine Forschung und Arbeit fokussiert auf inklusiver ruraler und urbaner Entwicklung, Partizipation und Global Governance, Open Source und Open Data und Forschungsentwicklung und -transfer. Er ist Mitglied des europäischen Netzwerks zu Kultur und Creative Industries. Er war Forschungsmanager am Zentrum für Literatur- und Kulturforschung (ZfL), Leiter des Forschungsinstituts PerGlobal und Koordinator der Exzellenzinitiative an der Humboldt-Universität zu Berlin. Beim UN-Weltgipfel zur nachhaltigen Entwicklung und in den UNO-Verhandlungen zur Informationsgesellschaft war er Teil Verhandlungsdelegation der Zivilgesellschaft.

Dr. Rainer Engels (Agrarwissenschaftler) ist wirtschaftspolitischer Experte in der Entwicklungszusammenarbeit und beschäftigt sich seit vielen Jahren mit entwicklungsökonomischen Fragen. Sein Schwerpunkt liegt dabei auf dem Gebiet der entwicklungsförderlichen Ausgestaltung der Handels-, Investitions- und Industriepolitik, insbesondere auch geistiger Eigentumsrechte und technischer Standards. Seit 2015 arbeitet er zur Automatisierung und Digitalisierung der Industrieproduktion (Industrie 4.0) und Elektromobilität. Er ist Berater für nachhaltige Wirtschaftspolitik und Privatwirtschaftsförderung bei der GIZ. Zuvor war er langjährig Geschäftsführer von Germanwatch.

Prof. Dr. Malte Rehbein (Historiker) ist Inhaber des Lehrstuhls für Digital Humanities an der Universität Passau, wo er formale und computergestützte Methoden einschließlich K.I.-basierter Verfahren und ihre Anwendungsmöglichkeiten für geistes- und kulturwissenschaftliche Aufgaben- und Fragestellungen mit besonderem Schwerpunkt auf den Geschichtswissenschaften erforscht und lehrt. Er publiziert zu Historical Data Studies, Kulturgutdigitalisierung und Datenmodellierung sowie zu Fragen der Ethik und Wissenschafts- und Gesellschaftskritik. Berufliche Erfahrungen umfassen die IT- und Consultingbranche; akademische Stationen waren Göttingen, Würzburg, Galway/Irland, Victoria/Kanada und Lincoln-NE/USA. Er ist Mitglied der Historischen Kommission bei der Bayerischen Akademie der Wissenschaften. Seit 2017 ist er Mitglied der Vereinigung Deutscher Wissenschaftler**.**

Frank Schmiedchen (Volkswirt, Betriebswirt) ist Regierungsdirektor im BMZ; u.a. zuständig für den IWF und internationale Wirtschafts- und Finanzfragen. Zuvor war er im BMZ bzw. der Ständigen Vertretung der Bundesrepublik Deutschland bei der EU verantwortlich für biologische Vielfalt und Biosafety, außen- und sicherheitspolitische Fragen Afrikas, AKP, Industriepolitik, UNIDO, geistige Eigentumsrechte und den Aufbau lokaler Pharmaproduktion. Zuvor war er Dekan des Fachbereichs "KMU-Management" an der Päpstlich-Katholischen Universität Ecuadors und koordinierte für die Vereinigung Jesuitischer Universitäten in Lateinamerika (AUSJAL) die entsprechenden Fachbereiche. 2002-2009 und seit 2016 war/ist er Mitglied des Beirates der Vereinigung Deutscher Wissenschaftler. Seit Oktober 2017 leitet er die VDW Studiengruppe Technikfolgenabschätzung der Digitalisierung.

Prof. Dr. Heinz Stapf-Finé (Soziologe, Volkswirt) ist Professor für Sozialpolitik an der Alice Salomon-Hochschule Berlin und Akademischer Leiter der Paritätischen Akademie Berlin. Er hat zum Thema „Alterssicherung in Spanien“ promoviert und ist ein internationaler Experte im Bereich Arbeits- und Sozialpolitik. Vor seiner Berufung als Hochschullehrer war er Bereichsleiter Sozialpolitik beim Bundesvorstand des Deutschen Gewerkschaftsbundes (DGB). Erste Berufserfahrung sammelte er als Operations Manager der Luxembourg Income Study und als wissenschaftlicher Mitarbeiter des Instituts für Gesundheits- und Sozialforschung (IGES) Berlin. Er arbeitete dann als Referent für Politik der Deutschen Krankenhausgesellschaft. Seit 2006 ist er Mitglied der Vereinigung Deutscher Wissenschaftler.

Angelika Sülzen (Betriebswirtin) ist Regierungsdirektorin im Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) und dort aktuell für die Bereiche Gleichstellung, Vereinbarkeit Beruf und Familie sowie Gesundheitsmanagement zuständig. Zuvor war sie als Länderreferentin für Zentralafrika u.a. für die bilaterale Zusammenarbeit der Bundesrepublik Deutschland mit Burundi und der Zentralafrikanischen Republik zuständig. Davor war sie mit Budget - und Finanzfragen betraut und hat ein großes IT-Projekt im BMZ geleitet. Von 2003 bis 2007 war sie für den Deutschen Entwicklungsdienst in Südafrika und Lesotho tätig.

Wir danken für ihre Textbeiträge und Anregungen:

Lucas Bartosch, Judith Buttenmüller, Prof. Dr. Hartmut Graßl, Prof. Dr. Regine Kollek, Dr. Hans-Jochen Luhmann, Dr. Michael Marhöfer und Christine von Weizsäcker.