Sabine Ammon, Olaf Dössel, Isabella Hermann, Christoph Markschies, Fruzsina Molnár-Gábor, Julian Nida-Rümelin, Jonas Peters, Dirk Pflüger, Timo Rademacher, Ortwin Renn, Frauke Rostalski, Pia-Johanna Schweizer, Günter Stock, Thorsten Thiel

Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!

Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2021
ISBN: 978-3-939818-97-7
(#Verantwortung KI – Künstliche Intelligenz und gesellschaftliche Folgen ; 4/2021)

Persistent Identifier: urn:nbn:de:kobv:b4-opus4-35203

Sabine Ammon
Olaf Dössel
Isabella Hermann
Christoph Markschies
Fruzsina Molnár-Gábor
Julian Nida-Rümelin
Jonas Peters
Dirk Pflüger
Timo Rademacher
Ortwin Renn
Frauke Rostalski
Pia-Johanna Schweizer
Günter Stock
Thorsten Thiel

4|2021

#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen

Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!

Eine Schriftenreihe der Interdisziplinären Arbeitsgruppe
Verantwortung: Maschinelles Lernen und Künstliche Intelligenz

berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)

VERANTWORTUNGSVOLLER EINSATZ VON KI?
MIT MENSCHLICHER KOMPETENZ!

VERANTWORTUNGSVOLLER EINSATZ VON KI? MIT MENSCHLICHER KOMPETENZ!

Sabine Ammon
Olaf Dössel
Isabella Hermann
Christoph Markschies
Fruzsina Molnár-Gábor
Julian Nida-Rümelin
Jonas Peters
Dirk Pflüger
Timo Rademacher
Ortwin Renn
Frauke Rostalski
Pia-Johanna Schweizer
Günter Stock
Thorsten Thiel

#VERANTWORTUNG KI – KÜNSTLICHE INTELLIGENZ UND GESELLSCHAFTLICHE FOLGEN 4|2021

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

Redaktion: Isabella Hermann und Ute Tintemann
Grafik: Thorsten Probst/angenehme gestaltung
Druck: bud Brandenburgische Universitätsdruckerei und Verlagsgesellschaft Potsdam mbh

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

ISBN: 978-3939818-97-7

INHALTSVERZEICHNIS

Vorwort ... 8
Christoph Markschies, Jens Krause und Isabella Hermann

Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!
Empfehlungen ... 12

1. Einleitung: Die Relevanz von Kompetenz beim verantwortungsvollen Umgang mit Künstlicher Intelligenz ... 15
Christoph Markschies

2. KI-Systeme, Verantwortung und Haftung ... 21
A. Zum menschlichen Handlungs- und Verantwortungsbegriff ... 21
Julian Nida-Rümelin
B. KI und Verantwortung bei technischen Systemen ... 23
Olaf Dössel
C. Zur rechtswissenschaftlichen Diskussion um Verantwortung und Haftung ... 28
Timo Rademacher
D. Herausforderungen an die haftungsrechtlichen Bestimmungen durch KI ... 32
Fruzsina Molnár-Gábor

3. Entwicklung von KI-Systemen ... 37
A. Fairness und Diskriminierung in datenbasierten Entscheidungssystemen ... 37
Jonas Peters
B. *Ethics-by-design* in Forschung und Entwicklung von Künstlicher Intelligenz ... 41
Sabine Ammon

4. Anwendung von KI-Systemen ... 48
A. KI und Herausforderungen für menschliche Kompetenz ... 48
Isabella Hermann und Günter Stock
B. Die Identifizierung von Emerging Risks durch KI ... 51
Pia-Johanna Schweizer und Ortwin Renn

5. Künstliche Intelligenz als Herausforderung für demokratische Partizipation ... 56
Frauke Rostalski und Thorsten Thiel

6. Ausblick: Kompetenzerwerb vom Kindesalter an ... 64
Dirk Pflüger

AUTORINNEN UND AUTOREN

Sabine Ammon: Professorin für Wissensdynamik und Nachhaltigkeit in den Technikwissenschaften und Leiterin des Berlin Ethics Lab, Institut für Werkzeugmaschinen und Fabrikbetrieb sowie Institut für Philosophie, Literatur-, Wissenschafts- und Technikgeschichte an der Technischen Universität Berlin.

Olaf Dössel: Professor für Biomedizinische Technik und Leiter des Instituts für Biomedizinische Technik am Karlsruher Institut für Technologie.*

Isabella Hermann: Wissenschaftliche Koordinatorin der Interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz" der Berlin-Brandenburgischen Akademie der Wissenschaften.

Christoph Markschies: Präsident der Berlin-Brandenburgischen Akademie der Wissenschaften und Sprecher der Interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz".*

Fruzsina Molnár-Gábor: Gruppenleiterin an der Heidelberger Akademie der Wissenschaften im Bereich des Völker- und Europarechts sowie der Rechtsvergleichung unter besonderer Berücksichtigung von Datenschutz- und Medizinrecht.

Julian Nida-Rümelin: Professor emeritus für Philosophie und politische Theorie, Ludwig-Maximilians-Universität München, Mitglied im Deutschen Ethikrat und Staatsminister a.D.*

Jonas Peters: Professor für Statistik, Universität Kopenhagen.

Dirk Pflüger: Professor für Scientific Computing, Institut für Parallele und Verteilte Systeme, Universität Stuttgart.

Timo Rademacher: Juniorprofessor für Öffentliches Recht und das Recht der neuen Technologien, Universität Hannover.

Ortwin Renn: Wissenschaftlicher Direktor am Institut für Transformative Nachhaltigkeitsforschung (IASS) Potsdam und Inhaber des Lehrstuhls „Technik- und Umweltsoziologie", Universität Stuttgart.*

Frauke Rostalski: Professorin für Strafrecht, Strafprozessrecht, Rechtsphilosophie und Rechtsvergleichung an der Universität zu Köln und Mitglied im Deutschen Ethikrat.

Pia-Johanna Schweizer: Forschungsgruppenleiterin für Systemische Risiken am Institut für Transformative Nachhaltigkeitsforschung (IASS), Potsdam.

Günter Stock: Vorstandsvorsitzender der Einstein Stiftung Berlin.*

Thorsten Thiel: Forschungsgruppenleiter „Digitalisierung und Demokratie", Weizenbaum-Institut für die vernetzte Gesellschaft Berlin.

* Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften

VORWORT

Mit diesem Heft unter dem Titel „Verantwortungsvoller Einsatz von KI? Mit menschlicher Kompetenz!" der Publikationsreihe „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen" stellt die Interdisziplinäre Arbeitsgruppe (IAG) „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz" abschließend Ergebnisse ihrer insgesamt dreijährigen Arbeit vor. Die IAG wurde vom Rat der Berlin-Brandenburgischen Akademie der Wissenschaften auf seiner Sitzung vom 30. November 2017 eingerichtet und nahm im folgenden Jahr die Arbeit auf. Die an der IAG-Gründung beteiligten Kolleginnen und Kollegen gingen von der Überlegung aus, dass in Prozessen der Entwicklung und Anwendung von Systemen der Künstlichen Intelligenz und des Maschinellen Lernens oft in einem schlichten Sinne unklar oder sogar diffus bleibt, „wer die Verantwortung trägt". In Zeiten, in denen transparente Strukturen von Verantwortlichkeit nicht nur im wirtschaftlichen Handeln immer wichtiger werden, erschien den IAG-Mitgliedern die Frage nach der Verantwortung bei Systemen Künstlicher Intelligenz und von Maschinellem Lernen ein ebenso dringendes wie spannendes Thema.

Schnell ergab sich allerdings, dass anfängliche Diskussionen, ob Künstliche Intelligenz (KI) selbst vertrauenswürdig oder verantwortungsvoll sein kann, in eine Sackgasse führen. Offenkundig unterscheiden sich an dieser Stelle zwar Sprachkonventionen beispielsweise der Technik- und der Geisteswissenschaften. Aber ungeachtet unterschiedlicher Sprachkonventionen waren sich alle Mitglieder der IAG darin einig, dass letztlich Verantwortung ein rationales Subjekt voraussetzt, dass Vertrauen übernehmen und Vertrauen in Technik letztlich zu begründen vermag – durch konkrete und im Detail identifizierbare Übernahme von Verantwortung. Es zeigte sich beim ausführlichen Studium der Prozesse, die zur Entwicklung von entsprechenden Systemen der KI führen, dass viel wichtiger als abstrakte Überlegungen zum Begriff Verantwortung die Aufklärung über Prozesse und verantwortliche Akteure ist – und zwar sowohl auf der Ebene der Entwickler*innen als auch auf den unterschiedlichen Ebenen der Nutzenden.

Aufklärung über Strukturen und Mechanismen der Verantwortung wird hier sowohl in technikwissenschaftlicher als auch in rechtswissenschaftlicher und ethischer Hinsicht vorgenommen und auf dieser Basis eine Bildungsinitiative zur Schulung der *Kompetenz* im verantwortlichen Umgang mit KI und Maschinellem Lernen gefordert. Die Hauptpunkte der Einsichten der IAG sind als Empfehlungen zusammenfassend einer ausführlicheren Darstellung vorangestellt.

Entscheidend ist aus Sicht der IAG, ob Menschen die Kompetenz besitzen, mit KI verantwortungsvoll umzugehen. Das betrifft aber nicht nur die Personen, die KI (hoffentlich) professionell entwickeln und (hoffentlich ebenso) professionell anwenden, sondern praktisch alle Bürger*innen, denn bei fast jeder und jedem ist KI mittlerweile Teil des alltäglichen Lebens, beispielsweise beim Newsfeed diverser Social Media. Kompetenz für den verantwortlichen Umgang mit KI muss aber durch bildungspolitische Maßnahmen entschlossen gefördert werden. Eine entsprechende gemeinsame Anstrengung der relevanten Institutionen und Akteure zu fordern, ist ein zentrales Ergebnis der Arbeit der IAG. Die IAG geht dabei von einem Bild des Menschen als einem Subjekt aus, das im Prinzip die Freiheit besitzt, rational handeln zu können und daher im Regelfall Verantwortung für sein Tun übernehmen kann und muss. Doch diese Verantwortung kann im Blick auf Systeme der KI und des Maschinellen Lernens nur übernommen werden, wenn Menschen überhaupt erst die *Kompetenz* besitzen, mit KI-Systemen verantwortungsvoll umgehen zu *können*. Bildung im Hinblick auf die technischen Funktionsweisen und ethischen Implikationen sowie gesellschaftlichen Folgen gehört hier genauso dazu wie die Forderung, KI-Systeme von Beginn an so zu gestalten, dass die Vorhersagen und Entscheidungen überhaupt von Menschen nachvollzogen werden können. Entsprechende Anstrengungen betreffen am Ende die ganze demokratische Gesellschaft und ihr Funktionieren.

Drei Jahre lang sind die Mitglieder der IAG regelmäßig zusammengekommen, um Strukturen und Mechanismen von Verantwortung in Systemen der KI und von Maschinellem Lernen zu identifizieren und nach verschiedenen fachwissenschaftlichen Paradigmen zu beschreiben. Das war angesichts der rasanten Entwicklungen in diesem Bereich kein leichtes Unterfangen. Die IAG hat sich, um nicht nur im Bereich rein theoretischer Diskussionen zu bleiben, dazu entschlossen, rasch die durch Referate und Diskussionen geprägten Sitzungen durch strukturierte Gespräche mit Unternehmensvertreter*innen, die KI-Systeme selbst entwickeln und anwenden, zu ergänzen. Diese Gespräche sollten bei der Analyse helfen, welche Entwicklungen und Anwendungen in der Wirtschaft tatsächlich

geplant und umgesetzt werden. Im Rahmen von strukturierten Interviews wurden Gespräche mit Repräsentantinnen und Repräsentanten des Online-Versandhändlers Amazon, des Allianz Versicherungskonzerns, des Technikunternehmens Bosch, der Wirtschaftsprüfungsgesellschaft Deloitte, der Unternehmensberatung Kienbaum sowie mit zwei Firmen für Persönlichkeitsmessung durch Sprachanalyse im Rahmen des Personalmanagements, Precire und „100 Worte", geführt. Die Gespräche gaben interessanten Aufschluss über die tatsächlichen technischen Möglichkeiten, die in der Öffentlichkeit häufig sehr übersteigert wahrgenommen werden und dann zu teilweise stark übertriebenen Bedrohungsszenarien führen. Allerdings machen gerade solche Szenarien und die damit verbundenen Ängste noch einmal deutlich, wie notwendig eine allgemeine Bildung für einen kompetenten und verantwortungsvollen Umgangs mit KI in unserer Gesellschaft ist und wie sehr es Initiativen für eine Intensivierung entsprechender Bildungsmöglichkeiten braucht. Außerdem veranstaltete die IAG zusammen mit der Deutschen Gesellschaft für Auswärtige Politik (DGAP) einen zweitägigen Workshop unter dem Titel „Towards European Anticipatory Governance for AI", in dem Wissenschaftler*innen, Mitglieder der High-Level Expert Group on AI der EU-Kommission, Vertreter*innen deutscher Bundesministerien, Normungsexpert*innen sowie weitere Unternehmensrepräsentant*innen an einem Tisch zusammenkamen. Wir diskutierten über zukünftige Regelungsmöglichkeiten von KI in Europa, die Innovationen ermöglichen, aber auch den ethischen Herausforderungen gerecht werden sollen.

Wir haben uns dagegen entschieden, ein zusammenhängendes gemeinsames Schlussdokument zu erarbeiten. Vielmehr werden die Empfehlungen, die die Mitglieder der IAG gemeinsam tragen, durch Abschnitte begründet, die die jeweiligen Fachvertreter*innen aus den Bereichen der Medizintechnik, Philosophie, Politik- und Sozialwissenschaften, Datenwissenschaften sowie der Rechtswissenschaft namentlich verantworten. Entsprechend beleuchten die Texte der Mitglieder Sabine Ammon, Olaf Dössel, Fruzsina Molnár-Gábor, Julian Nida-Rümelin, Jonas Peters, Dirk Pflüger, Timo Rademacher, Ortwin Renn, Frauke Rostalski, Pia-Johanna Schweizer, Günter Stock, Thorsten Thiel sowie der Koordinatorin Isabella Hermann die Frage nach der menschlichen Kompetenz und Verantwortung sehr unterschiedlichen Perspektiven. Allerdings haben sich natürlich nicht nur diese Mitglieder der IAG an der Erarbeitung der Ergebnisse und der Diskussion des Schlussdokumentes beteiligt. Neben den genannten Autorinnen und Autoren möchten wir auch herzlich den Mitgliedern Susanne Beck, Jessica Burgner-Kahrs, Horst Eidenmüller, Nausikaä El-Mecky, Anja Feldmann, Carl Friedrich Gethmann, Frank Kirchner,

Max Löhning, Jakob Macke, Andreas Radbruch und Markus Willaschek für ihre Mitarbeit in der Interdisziplinären Arbeitsgruppe herzlich danken. Es handelte sich übrigens um diejenige Arbeitsgruppe der BBAW, die den bisher höchsten Anteil an Mitgliedern der „Jungen Akademie" aufwies, der bisher seit Einführung dieses Arbeitsformats zu konstatieren war. Diese Tatsache und die Beobachtung, dass zwischen älteren und jüngeren Mitgliedern der IAG im Alltag kaum zu differenzieren war, nehmen wir als Hinweis darauf, dass die vertraute Rede von zwei Mutterakademien und ihrer Tochter, die im Blick auf BBAW, Leopoldina und Junge Akademie gern verwendet wird, doch noch einmal überprüft werden sollte. Hier waren wir Partner.

Insbesondere in Zeiten einer Pandemie, unter deren Bedingungen die Arbeit seit März 2020 stattfand, war die ohnehin wertvolle Unterstützung durch die Wissenschaftsadministration der BBAW, insbesondere durch Ute Tintemann, nunmehr zentral wichtig. Nicht verschwiegen werden sollte am Ende auch, dass Günter Stock, langjähriger Präsident der BBAW, mit dem Thema „KI und Verantwortung" erneut eine wichtige Anregung für eine IAG gab, die von Mitgliedern aller Klassen ebenso zügig wie engagiert aufgegriffen wurde. Mit dem Dank für alle Unterstützung und der Erinnerung an ebenso spannende wie vergnügliche Diskussionen in Präsenz und am Bildschirm schließt dieses Vorwort.

Berlin, April 2021

Christoph Markschies
Sprecher der Interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz"

Jens Krause
Stellvertretender Sprecher der Interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz"

Isabella Hermann
Wissenschaftliche Koordinatorin der Interdisziplinären Arbeitsgruppe „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz"

VERANTWORTUNGSVOLLER EINSATZ VON KI? MIT MENSCHLICHER KOMPETENZ!

EMPFEHLUNGEN DER INTERDISZIPLINÄREN ARBEITSGRUPPE „VERANTWORTUNG: MASCHINELLES LERNEN UND KÜNSTLICHE INTELLIGENZ“

KI-Systeme können viele Aufgaben und Probleme besser bzw. präziser lösen als Menschen es können – je nach Einsatzbereich kann die Anwendung von KI ethisch also geboten und in bestimmten Fällen ein Verzicht darauf unverantwortlich sein. Die steigende Automatisierung durch KI-Systeme erschwert allerdings einen verantwortungsvollen Umgang mit der Technik. Automatisierung birgt nämlich zum einen die Gefahr, dass Menschen die Vorhersagen und Ergebnisse von KI-Systemen unreflektiert oder gar unbewusst annehmen, und zum anderen oft gar nicht mehr nachvollziehen können, wie die Ergebnisse zustande kommen. Das kann wiederum dazu führen, dass Fehler oder ein Bias in den Systemen, z. B. eine ungerechtfertigte Diskriminierung bestimmter Einzelpersonen oder Gruppen, hingenommen bzw. nicht erkannt werden und daher auch nicht behoben werden können. Daher empfiehlt die Interdisziplinäre Arbeitsgruppe (IAG) „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz“ der Berlin-Brandenburgischen Akademie der Wissenschaften gesellschaftlichen wie politischen Akteur*innen, aber auch Nutzer*innen der Technik, auf solche Entwicklungen sorgfältig zu achten und ihnen, wo nötig, entgegenzusteuern.

*Die wichtigste Grundlage für einen verantwortungsvollen Umgang mit KI ist, dass Menschen die Kompetenz besitzen, die Vorhersagen und Entscheidungen von KI-Systemen einzuordnen und soweit als möglich vor ihrer Ausführung zu überprüfen. Um einen in diesem Sinne kompetenten, d.i. verantwortungsvollen Umgang sicherzustellen, stehen sowohl Entwickler*innen als auch Anwender*innen in der Pflicht. Aber auch Alltagsnutzer*innen müssen über die fundamentalen Prinzipien und Funktionsweisen von KI-Systemen Bescheid wissen bzw. geschult werden. An diesem Punkt besteht ein erheblicher Bedarf an Kompetenzbildung in unserer Gesellschaft, auf den die IAG nachhaltig aufmerksam machen möchte.*

Kompetenzbildung für einen verantwortungsvollen Umgang mit KI hat eine Reihe von Voraussetzungen und muss gleichzeitig auf mehreren Ebenen intensiviert werden:

- Kompetenzen für einen verantwortungsvollen Umgang mit KI bei der Konzeption und Entwicklung von KI-Systemen setzen teamorientierte, interdisziplinäre Ansätze und eine integrative Ethik voraus. Notwendig ist sowohl eine anwendungsorientierte, ethisch-reflexive Grundbildung auf Seiten der technischen Expert*innen als auch technische Kompetenz auf Seiten von Ethiker*innen, denn KI-Systeme lassen sich insbesondere hinsichtlich ihrer ethischen Probleme (wie beispielsweise des Grades ihrer Fairness und der Vermeidung unzulässiger Diskriminierung) nicht automatisiert nach allgemeingültigen Prinzipien entwickeln und überprüfen. Je nach Anwendungsfall muss im Einzelnen von Menschen nach den allgemein akzeptierten und normierten Wertmaßstäben entschieden und agiert werden.

- Weiter müssen KI-Systeme „by design" von vornherein so konzipiert und entwickelt werden, dass Menschen in der späteren Anwendung ein kompetenter Umgang ermöglicht wird. Das bedeutet, dass sich die Vorhersagen und Ergebnisse von KI-Systemen möglichst gut selbst erklären und nachvollzogen werden können; KI-Systeme, die in diesem Sinne nicht nachvollziehbar sind, müssen durch den Gesetzgeber sektorspezifisch reguliert werden.

- Kompetenzbildung zu einem verantwortungsvollen Umgang mit KI auf Seiten der professionellen Anwender*innen wird durch Qualifikationen und Trainings sichergestellt, die die Funktionsweise von KI-Systemen und deren Vorhersagen wie Ergebnisse hinsichtlich der dahinterliegenden Modelle und der verwendeten Daten transparent machen.

- Kompetenzbildung zu einem verantwortungsvollen Umgang mit KI auf Seiten der allgemeinen Anwender*innen wird durch schulische, berufliche und universitäre Bildung sichergestellt sowie durch Informationskampagnen weiterer Bildungsinstitutionen. Es sollte allen Anwender*innen bewusst sein, dass sie es – beispielsweise bei einer Routenplanung im Straßenverkehr oder medizinischen digitalen Assistenten – mit technischen Artefakten zu tun haben, deren Leistungsfähigkeit bestimmten Grenzen unterliegt.

Gerade im Zusammenhang mit der aktuellen COVID-19-Pandemie ist deutlich geworden, wie wichtig eine KI-gestützte Risikoeinschätzung für individuelle aber auch für gesamtgesellschaftliche (vor allem politische) Entscheidungen geworden ist. Das trifft längst genauso für die Technikfolgenabschätzung zu. Auch hier gilt, was zuvor schon allgemein festgehalten wurde: Solche KI-Systeme brauchen

ganz besonders einen im beschriebenen Sinne kompetenten, also verantwortlichen Umgang durch den Menschen, um die Potentiale voll ausschöpfen zu können, aber gleichzeitig unerwünschte oder sogar schädliche Folgen der Anwendung abwenden zu können. Insofern kann man also durchaus sagen, dass jüngste Entwicklungen den Bedarf an Kompetenzbildung auf allen Seiten und die Notwendigkeit einer entsprechenden gemeinsamen Initiative von Bund, Ländern und Verantwortungsträgern im Bildungssektor eher verstärkt haben. Auch wenn die Haushalte durch die gegenwärtige Krise schwer belastet sind, kann angesichts der außerordentlichen Entwicklungsdynamik auf dem Sektor von KI und Maschinellem Lernen die dringende Aufgabe, mehr in die Ausbildung von Verantwortungskompetenz auf allen Ebenen zu investieren, nicht aufgeschoben werden.

Im Idealfall kann KI zur Lösung von Problemen mit gesellschaftlicher Relevanz beitragen. Das verantwortungsvolle Umsetzen entsprechender Lösungen ist nicht nur eine Frage der „digitalen Souveränität" Einzelner, sondern betrifft den ganzen Zusammenhang zwischen Demokratie und Künstlicher Intelligenz, wobei hier Demokratie umfassend als Lebensform verstanden ist. Daher sollte auf allen Ebenen eine Bildungsinitiative zur Vermittlung von Verantwortungskompetenz im Blick auf KI und Maschinelles Lernen gestartet werden. Die IAG „Verantwortung: Maschinelles Lernen und Künstliche Intelligenz" ergänzt ihre Empfehlungen in der vorliegenden Publikation durch konkrete Vorschläge für Inhalte, Methoden und Standards solcher Bildung.

Berlin, im April 2021

Christoph Markschies[1]

1. EINLEITUNG: DIE RELEVANZ VON KOMPETENZ BEIM VERANTWORTUNGSVOLLEN UMGANG MIT KÜNSTLICHER INTELLIGENZ

Künstliche Intelligenz (KI) ist keine „Science-Fiction-Technologie", die irgendwann in der Zukunft zum Einsatz kommen wird: Auf KI wird inzwischen an immer mehr Stellen zurückgegriffen, um durch Automatisierung Prozesse zu optimieren. Ob zur Fehlererkennung in der Produktion, zur Effizienzsteigerung in der Logistik, für politische oder wirtschaftliche Prognosen, für Übersetzungsdienste oder zur Bild- und Spracherkennung – der Einsatz von KI ist in vielen Industriezweigen, Berufsfeldern und vor allem auch im Alltag vieler Menschen Realität. Der steigende Einsatz von KI verdankt sich hauptsächlich Erfolgen im Bereich des Maschinellen Lernens, einem Verfahren, bei dem Computer Muster in großen Datenmengen erkennen, beispielsweise mit Hilfe von sogenannten neuronalen Netzen. Obwohl KI als technisches Artefakt ein von Menschen entwickeltes Werkzeug darstellt (wenn auch ein komplexes), wird in der Öffentlichkeit darüber oft in einer Weise diskutiert, als hätte die Technik Akteursqualitäten und könnte entweder Weltprobleme lösen oder würde ein schlechterdings unkontrollierbares Risiko darstellen.[2] Diese zwischen messianischen und apokalyptischen Vorstellungen oszillierenden Wahrnehmungen führen zu einer merkwürdig verschobenen Debatte über die Chancen und Risiken von KI-Anwendungen, die von der menschlichen Gestaltungs- und Kontrollmacht ablenkt. So findet man in der Diskussion häufig (insbesondere bei Versuchen, Systeme automatisierten Steuerns von Kraftfahrzeugen zu entwickeln) das Bild, die KI würde hier selbst ethische Urteile fällen, obwohl doch in Wahrheit diese Urteile die Wertordnungen, Weltbilder und Interessen der Menschen innerhalb des soziotechnischen Systems widerspiegeln, in dem die Technik entwickelt wurde und angewendet wird.[3]

1 Für alle Anregungen und Zuarbeiten danke ich sehr herzlich: Isabella Hermann, Olaf Dössel, Jens Krause, Jonas Peters und Frauke Rostalski.

2 Leufer, D (2020): Why We Need to Bust Some Myths about AI. In: Patterns 1(7). https://doi.org/10.1016/j.patter.2020.100124.

3 So ist beispielsweise der Name des MIT-Projekts „Moral Machine" irreführend, in dem erforscht wird „wie Menschen zu Entscheidungen stehen, die von intelligenten Maschinen, wie z. B. selbstfahrenden Autos, getroffen werden" (MIT Media Lab [2021]: Moral Machine. https://www.moralmachine.net/hl/de [15.3.2021]). Schließlich ist es bei anderen technischen Artefakten völlig unüblich, ihnen ein moralisches oder ethisches Verhalten zuzurechnen.

Für einen konstruktiven Dialog über die Ausbildung von Kompetenz zur Übernahme von Verantwortung bei der Entwicklung und Nutzung von KI sollten die Begriffe Künstliche Intelligenz und Maschinelles Lernen daher dringend präzise gebraucht werden. Zunächst ist Künstliche Intelligenz ein Teilgebiet der Informatik und Statistik. Die historischen Anfänge liegen in der Bestimmung von Regressionsgeraden, mit denen ein linearer (genauer: affiner) Zusammenhang zwischen einer bekannten Eingangsgröße X und einer bekannten Zielgröße Y aufgestellt wird. Dabei werden Messwerte aus der Vergangenheit verwendet, um den bestmöglichen Zusammenhang zwischen den Größen zu ermitteln, so dass für eine zukünftige Eingangsgröße X ein entsprechender Wert der Zielgröße Y vorhergesagt werden kann. Viele Systeme im Bereich KI und Maschinelles Lernen machen im Prinzip nichts anderes, erlauben aber oftmals komplizierte, nichtlineare Zusammenhänge mit vielen Parametern, greifen auf eine größere Datenbasis und Speicherkapazität zurück und benötigen eine stärkere Rechenleistung. Wie im Fall der Regressionsgerade kann das „gelernte" oder „trainierte" System für jeden zukünftigen Wert von X einen Wert von Y vorhersagen. Wohingegen unter den Begriff KI auch regelbasierte Anwendungen fallen können, bei denen die Regeln von Menschen aufgestellt wurden, lernen die Systeme beim Maschinellen Lernen Zusammenhänge mit oder ohne menschliche Zieldefinition selbst in großen Datenmengen. KI-Verfahren, vor allem solche, die auf neuronalen Netzen basieren bzw. die selbst kontinuierlich weiterlernen, erklären sich dabei häufig nicht von selbst, weswegen man von der so genannten „black box" von KI-Systemen spricht. Doch es gibt Methoden und Verfahren, um die Vorhersagen von KI-Systemen erklärbar und nachvollziehbar darzustellen, d.h. aus der „black box" eine „gray box" oder „white box" zu machen;[4] damit befasst sich ein eigener Forschungszweig, der sich als „explainable AI" zusammenfassen lässt.

KI-Systeme sind sinnvolle Werkzeuge, um Menschen bei ihren Entscheidungen zu unterstützen. Menschen müssen ständig – im professionellen wie im privaten Kontext – Entscheidungen auf der Basis von unsicheren oder jedenfalls unvollständigen Informationen fällen. Sie orientieren sich an ihren eigenen Erfahrungen, einem für sie überschaubaren Set von Regeln und bestimmten Normen. Damit ist es schon aus Gründen der für Entscheidungsfindung notwendigen Komplexitätsreduktion unvermeidlich, dass die zu Grunde liegende Basis von Informationen ohne maschinelle Unterstützung relativ klein ist und nicht alle möglichen Fälle und Konstellationen beinhalten kann. Das kann in bestimmten

4 Vgl. beispielsweise Acatech (2020): Machine Learning in der Medizintechnik. Analyse und Handlungsempfehlungen. In: acatech position. München, S. 14. https://www.acatech.de/publikation/machine-learning-in-der-medizintechnik/download-pdf?lang=de [15.3.2021].

Entscheidungssituationen zu einer Reihe von kognitiven Verzerrungen und voreiligen Schlussfolgerungen führen, die schließlich auch falsche Entscheidungen zur Folge haben können.[5] In einer solchen Situation hat KI das Potential, Menschen bei der Entscheidungsfindung zu unterstützen, da die ihr zugrundeliegende Informationsbasis größer als die Summe der Erfahrungen eines einzelnen Menschen oder einer Gruppe von Menschen sein kann. So können auch komplexe Zusammenhänge berechnet werden, die ein Mensch aufgrund der Informationsmenge oder der Komplexität der Informationen nicht überschauen kann. Das gilt insbesondere, wenn viele verschiedenartige Informationen mit unterschiedlicher Messgenauigkeit verarbeitet werden müssen. Man kann sich diese Zusammenhänge gut an der Medizin klarmachen. Dort ist es beispielsweise ein wichtiges Ziel, immer mehr zu „evidenzbasierten" Entscheidungen über Diagnosen und entsprechende Therapien zu kommen. KI kann, verantwortlich als Werkzeug eingesetzt, die Ärzte und Ärztinnen bei ihren Diagnosen, z. B. in der Krebserkennung, unterstützen. Wenn ein KI-System nachweislich seltener einen Fehler macht als ein Mensch, dann erscheint es bisweilen sogar ethisch geboten, solche Systeme einzusetzen.

Doch im Kontext von KI-Anwendungen stellen sich auch verschiedenartige Probleme, die unkritische Nutzer*innen übersehen könnten. Bei dem gewählten Beispiel einer KI-Technologie zur Verbesserung medizinischer Diagnostik besteht das Risiko, dass relevante Faktoren fehlerhaft interpretiert werden, was körperliche Schäden von Patient*innen nach sich ziehen kann. Wenn im Umgang mit Entscheidungsempfehlungen von KI keine Kenntnisse über deren Zustandekommen vorliegen oder gar die Entscheidungen vollständig an die Maschine delegiert werden, wird unter Umständen eine gebotene Behandlung unterlassen oder eine überflüssige medizinische Intervention unternommen. Zu beachten ist die Gefahr verzerrender oder fehlerhafter Einflussfaktoren – sogenannter „Bias" –, die zu vorurteilsbasierten Ergebnissen des KI-Systems und entsprechend problematischen Entscheidungen führen können. Weil die Qualität eines KI-Systems wesentlich von der Qualität der Daten abhängig ist, mit denen es trainiert wurde, können sich bereits an dieser Stelle Schwächen der KI-Anwendung ergeben, die teilweise erst nach einiger Zeit und nach kritischer Prüfung durch Menschen offenbar werden. Dies kann zur Folge haben, dass ein Programm für bestimmte Bevölkerungsgruppen schlechter funktioniert und es damit zu einer systematischen Benachteiligung kommt. Sofern eine solche Diskriminierung weder den Entwickler*innen noch den Anwender*innen klar ist und neuere Daten und

5 Kahneman, D (2013): Thinking, Fast and Slow. Farrar, Straus and Giroux: New York.

Erkenntnisse nicht wirkungsvoll eingespeist oder berücksichtigt werden, droht die Technik hinter dem Stand der menschlichen Erkenntnis zurückzufallen – mit möglichen negativen gesellschaftlichen, politischen und wirtschaftlichen Konsequenzen.

Menschen, beispielsweise in der Medizin, die KI-Systeme entwickeln und im professionellen Kontext anwenden, haben also die Verantwortung, dass Personen nicht nur vom KI-Einsatz profitieren, sondern vielmehr auch niemand zu Schaden kommt. Sie dürfen Vorhersagen und Ergebnisse von KI-Systemen nicht einfach ohne kritische Reflexion übernehmen. Wenn „Schaden" nicht nur im Sinne einer körperlichen Beeinträchtigung verstanden wird, sondern alle rechtswidrigen Beeinträchtigungen und Benachteiligungen von Einzelnen oder Gruppen einschließt, ist Bildung von Kompetenz zu Verantwortungsübernahme unausweichlich. Sie besteht je nach Vorbildung in technischem, aber auch ethischem Wissen. Wenn sich zugleich aber alle, die KI-Anwendungen im Alltag nutzen – z. B. zur Routenplanung im Straßenverkehr – oder mit KI-Vorhersagen konfrontiert sind – z. B. bei Vorschlägen für die Produktauswahl oder zur Festsetzung von Verkaufspreisen – im besten Fall darauf verlassen möchten, dass die Systeme nicht schaden, dann ist auch für Nutzer*innen Kompetenzbildung zur Verantwortungsübernahme unerlässlich.

Um eine verantwortungsvolle Nutzung von KI sicherzustellen, bedarf es der staatlichen Regulierung sowie der Standardsetzung und Zertifizierung, damit die Chancen von KI hinsichtlich Automatisierung und Optimierung genutzt werden können, ohne in die Falle der viel diskutierten ethischen und rechtlichen Probleme zu tappen. Welche Arten von KI-Anwendung mit einem besonders hohen Risiko verbunden sind und welche Arten von Regulierung für bestimmte Systeme greifen sollen, wird an vielen anderen Stellen diskutiert.[6] Sie ist leichter zu beantworten als die Frage, ob Künstliche Intelligenz selbst in Zukunft als verantwortlicher Akteur behandelt werden kann und sollte.[7] Der Fokus dieses abschließenden Heftes der Interdisziplinären Arbeitsgruppe (IAG) „Verantwortung: Künstliche Intelligenz und Maschinelles Lernen" soll aber darauf liegen, wie Menschen dazu ausgebildet werden können, mit den im Einsatz befindlichen KI-Systemen verantwortungsvoll umzugehen und diese verantwortungsvoll weiterentwickeln zu können. In beiden

6 Beispielsweise auf europäischer Ebene: High-Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for trustworthy AI. Brüssel, 08.04.2019; Europäische Kommission (2020): White Paper on Artificial Intelligence: A European approach to excellence and trust. Brüssel, 19.02.2020.

7 Beispielsweise Wagner, J (2020): Künstliche Intelligenzen als moralisch verantwortliche Akteure? Brill Mentis: Paderborn.

Fällen ist menschliche Kompetenz im Umgang mit KI gefragt, die allerdings erst einmal grundgelegt, ausgebildet und fortgebildet werden muss, da sie nicht zur natürlichen Ausstattung des Menschen gehört. Eine ähnliche Stoßrichtung hat auch die Forderung nach individueller „digitaler Souveränität", die den Menschen im Blick auf seine „Wahlfreiheit, Selbstbestimmtheit, Selbstkontrolle und Sicherheit" in das Zentrum der Technikentwicklung und -anwendung stellt.[8] Denn der Mensch – als Bürger*in, Entwickler*in und Anwender*in – kann nur im Zentrum stehen, wenn entsprechendes Problembewusstsein, Wissen und die Fähigkeiten vorhanden sind, um Vorhersagen und Entscheidungsangebote von KI-Systemen richtig einordnen zu können. Diese Kompetenzen müssen in der allgemeinen Schulbildung, in Aus- und Fortbildungen und allen Bildungsinstitutionen auf allen Ebenen vermittelt werden. Hier liegt in den nächsten Jahren eine zentrale Aufgabe für viele Akteure.

Wie bei allen solchen Bildungsinitiativen muss auch, wenn die Kompetenz zur Verantwortungsübernahme bei der Nutzung von KI und Maschinellem Lernen nachhaltig gesteigert werden soll, präzise zwischen den Aufgaben der verschiedenen Akteure und den unterschiedlichen Gruppen, deren Kompetenz ausgebildet werden soll, unterschieden werden. Sehr schematisch lassen sich zwei betroffene Gruppen unterscheiden: zum einen die Alltagsnutzer*innen von KI-Anwendungen, wozu inzwischen angesichts der Verbreitung solcher Systeme im Alltag praktisch alle Bürger*innen zählen, und zum anderen Entwickler*innen und professionelle Nutzer*innen, für die ein höherer Professionalisierungsbedarf mit Blick auf einen verantwortungsvollen Umgang mit KI besteht. Die Kompetenz erstreckt sich dabei auf technisches Wissen hinsichtlich der Funktionsweise genauso wie auf ethisches und soziales Wissen im Hinblick auf mögliche (positive wie negative) Folgen für Individuum und Gesellschaft.

Ausgehend von diesen Fragen ergibt sich die Gliederung der Studie, die die Interdisziplinäre Arbeitsgruppe zum Abschluss ihrer Diskussionen vorlegt: Im zweiten Abschnitt soll zunächst auf den menschlichen Kompetenz- und Verantwortungsbegriff in Hinblick auf die ethischen und rechtlichen Herausforderungen im Umgang mit KI-Anwendungen eingegangen werden. Danach werden im dritten Abschnitt Anforderungen bei der Entwicklung von KI-Systemen diskutiert: Welche Standards für die menschliche Kompetenz mit

8 Vgl. Der Sachverständigenrat für Verbraucherfragen (SVRV) (2017): Digitale Souveränität. Gutachten des Sachverständigenrats für Verbraucherfragen. Berlin, S. 4 ff. https://www.svr-verbraucherfragen.de/wp-content/uploads/Gutachten_Digitale_Souver%C3%A4nit%C3%A4t_.pdf [15.3.2021].

Blick auf Fairness der Systeme, gelten bereits bei der Entwicklung der Systeme und welche Möglichkeit gibt es, um ein verantwortungsvolles Design von Beginn an sicherzustellen? Darauf aufbauend werden im vierten Abschnitt konkrete Herausforderungen für die menschliche Kompetenz in der Anwendung und Risikoeinschätzung dargestellt. Der fünfte Abschnitt behandelt KI im Kontext demokratischer Partizipation. Die Studie endet mit einem Ausblick auf die in den Augen der IAG dringliche Forderung, KI als gesamtgesellschaftliche Bildungsaufgabe von Kindesalter an zu begreifen.

2. KI-SYSTEME, VERANTWORTUNG UND HAFTUNG

Julian Nida-Rümelin

A. ZUM MENSCHLICHEN HANDLUNGS- UND VERANTWORTUNGSBEGRIFF

Der Verantwortungsbegriff ist mit anderen normativen Begriffen eng verbunden. Ohne ein Mindestmaß an Freiheit, kann eine Person für ihr Tun nicht verantwortlich gemacht werden. Wären Menschen vollständig determinierte Wesen, müsste das Selbstbild des Menschen als frei und verantwortlich als bloße, wenn auch vielleicht nützliche, Illusion gelten. Menschen denken allerdings über ihre Handlungen nach und sind in der Lage, Gründe für das eigene Tun zu geben sowie ihre Handlungen an Gründen auszurichten. Diese Fähigkeit, Entscheidungen zu treffen, die den besten Gründen folgen, ist das, was die menschliche Freiheit und Verantwortung ausmacht und uns von Tieren einerseits und Maschinen andererseits unterscheidet. Wenn die jeweilige Handlung vor jeder Überlegung oder Abwägung bereits festläge, wäre der Mensch als Akteur nicht frei und nicht verantwortlich. Ja, genau besehen gäbe es ihn als Akteur überhaupt nicht. Es gäbe dann keine Handlung, sondern lediglich bloßes Verhalten.

Der Naturalismus, der insbesondere in den Neurowissenschaften weitverbreitet zu sein scheint, bestreitet menschliche Freiheit und Verantwortung unter Verweis auf das determinierte System des Gehirns, das durch seine genetische und epigenetische Prägung sowie durch sensorische Stimuli gesteuert sei.[9] Das Problem dieser Position ist, dass sie nicht nur den Intuitionen der meisten Menschen widerspricht, sondern auch offensichtlich falsch ist.[10] Unsere individuelle Charakterentwicklung ist nicht nur von Umwelt und Genetik allein, sondern auch von unseren eigenen Entscheidungen abhängig. Dies entspricht der Einsicht, die Aristoteles in der Nikomachischen Ethik formuliert hat. Aristoteles macht dort deutlich, dass Tugenden (das bedeutet bei ihm: Charaktermerkmale, Dispositionen, Einstellungen) nicht nur auf Gewohnheit und Erziehung beruhen, sondern auch

9 Vgl. Singer, W (2002): Der Beobachter im Gehirn. Essays zur Hirnforschung. Suhrkamp Verlag: München.

10 Vgl. Nida-Rümelin, J (2005): Über die menschliche Freiheit. Reclam Verlag: Ditzingen.

Ausdruck von Entscheidungen (griechisch: *prohaireseis*) sind. Natürlich spielen für die Ausprägung von Tugenden Erfahrung und Gewohnheit eine wichtige Rolle. Aber Menschen sind auch in der Lage, ihre Einstellungen zu ändern, sich gewissermaßen zu entscheiden, in Zukunft eine andere Haltung einzunehmen. Auch Aristoteles spricht von der Tugend als einer Haltung (griechisch: *hexis*), und diese Überzeugung ist das Ergebnis einer Abwägung und schließlich einer Stellungnahme, zumal nach bitteren Erfahrungen oder Lebenskrisen. Auch emotive Einstellungen, wie zum Beispiel die Wertschätzung einer Person, beruhen auf der Überzeugung, dass die Person eine besondere Leistung erbracht, ein ungewöhnliches Maß an Hilfsbereitschaft gezeigt hat oder Ähnliches.[11] Wir sind nicht lediglich „Produkte" von Erziehung und Sozialisation, sondern auch aktive Gestalter unseres eigenen Charakters.[12] Dies unterscheidet Menschen maßgeblich von Maschinen.

Die beiden Normwissenschaften Ethik und Recht sind sich in der Bestimmung des Handlungsbegriffes weitgehend einig. Im Zentrum steht dabei das aktive, motivierte und kontrollierte Verhalten, das auf die Umwelt einwirkt, also Veränderungen verursacht. Dies lässt sich im Sinne einer umfassenden Theorie praktischer Vernunft und Anthropologie als Ausdruck eines Phänomens verstehen, für das sich unterdessen weithin der Begriff der „Person" eingebürgert hat. Eine Person ist in der Lage, ihre handlungsleitenden Wünsche über die Zeit so zu strukturieren, dass sich ihre Identität über unterschiedliche Kontexte hinweg durchhält. In der lebensweltlichen Praxis äußert sich das daran, dass wir erwarten, dass Personen Gründe für ihr Tun hervorbringen können und diese Gründe nicht ad hoc angeführt werden, sondern eine hinreichende Invarianz aufweisen, um die Person in ihrem Verhalten erkennbar werden zu lassen. Die Verbindung von Handlung und Verantwortung ist ohne diesen größeren Kontext nicht verständlich. Erst die Person, die über unterschiedliche Situationen hinweg erkennbar bleibt, ist auch voll verantwortlich, sie ist mit einer Identität, Autonomie und der damit einhergehenden individuellen Würde ausgestattet. Was dies im Kontext von ethischer und rechtlicher Verantwortung konkret bedeutet, beleuchten die nachfolgenden Kapitel.

11 Vgl. Nida-Rümelin, J (2011): Verantwortung. Reclam Verlag: Ditzingen.

12 Vgl. für den Absatz auch Nida-Rümelin, J; Weidenfeld, N (2018): Humanismus – Eine Ethik für das Zeitalter der Künstlichen Intelligenz. Piper: München, S. 48 f.

B. KI UND VERANTWORTUNG BEI TECHNISCHEN SYSTEMEN

Technische Systeme, die Komponenten von KI enthalten, eröffnen neue Möglichkeiten der Automatisierung und Prozessoptimierung. Gleichwohl bleiben sie als Maschinen mit besonderen Fähigkeiten technische Systeme. Sie sind von Menschen entwickelt und implementiert, und sie werden von Menschen eingesetzt, um bestimmte Aufgaben zu erfüllen. Welche Art von Verantwortung ist gemeint, wenn hier von Verantwortung bei technischen Systemen die Rede ist? Ein handelsübliches Internet-Lexikon definiert den Begriff „Verantwortung" folgendermaßen (und fasst damit einen breiten Konsens einschlägiger philosophischer und juristischer Definitionen zusammen): „Verantwortung ist die (freiwillige) Übernahme der Verpflichtung, für die möglichen Folgen einer Handlung oder einer getroffenen Entscheidung einzustehen und gegebenenfalls dafür Rechenschaft abzulegen oder Strafen zu akzeptieren. Sie setzt Verantwortungsgefühl und -bewusstsein, ein Gewissen sowie die Kenntnis der Wertvorstellungen sowie der rechtlichen Vorschriften und sozialen Normen voraus."[13]

Der im obigen Sinne allgemein verstandene Begriff „Verantwortung" bei technischen Systemen lenkt die Gedanken zunächst auf mögliche Fehlfunktionen, bei denen Schaden entstehen kann, für den sich der Verursacher verantworten muss. Darüber hinaus sind bei der Einführung neuer Technologien auch ethische, gesellschaftliche und unsere Umwelt betreffende Aspekte zu beachten. Damit beschäftigen sich bereits viele Publikationen mit einem Bezug auf technische Systeme und Maschinen im Allgemeinen.[14] Auch die Verantwortung der Erfinder*innen für alle möglichen – auch die nicht-intendierten – Anwendungen der entsprechenden Erfindung bzw. technischen Neuerungen wurde schon vielfach in der Literatur besprochen.[15] Es ist wichtig zu klären, wie sich die dort beschriebenen Prinzipien auf Systeme mit KI anwenden lassen, und ob es Ergänzungen und Erweiterungen

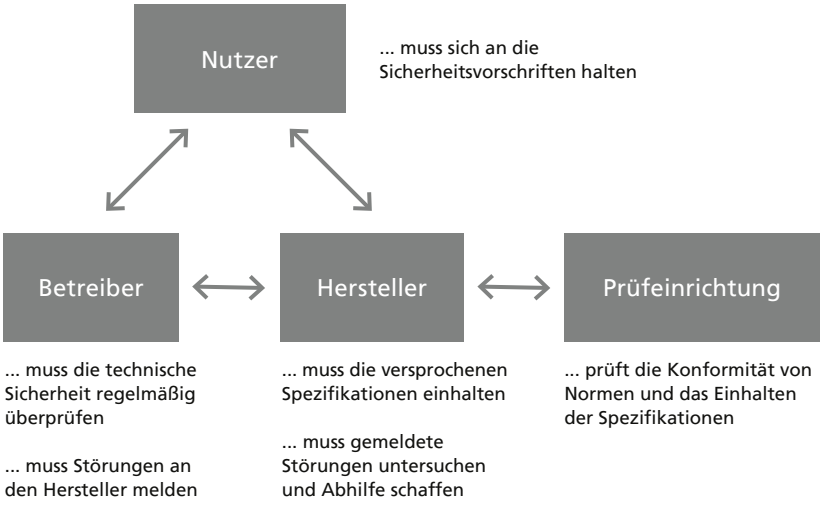
13 https://de.wikipedia.org/wiki/Verantwortung [15.3.2021]. Ausführlich zu diesem Thema beispielsweise: Nida-Rümelin, J (2011): Verantwortung. Reclam Verlag: Ditzingen; und den Beitrag von Nida-Rümelin in diesem Heft.

14 Grunwald, A (2020): Verantwortung und Technik: zum Wandel des Verantwortungsbegriffs in der Technikethik. In: Seibert-Fohr, A (Hrsg.): Entgrenzte Verantwortung. Springer: Berlin, Heidelberg, S. 265–283. https://doi.org/10.1007/978-3-662-60564-6_13; Lenk, H (2017): Ethics of responsibilities distributions in a technological culture. AI & Society 32, S. 219–231. https://doi.org/10.1007/s00146-015-0642-3.

15 Jonas, H (1984): Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation. Suhrkamp Taschenbuch: Frankfurt/M, Neuauflage.

bedarf.[16] Dieser Abschnitt handelt nun von dem zuerst genannten Aspekt: der Verantwortung dafür, dass Fehlfunktionen so weit wie möglich ausgeschlossen werden und unmittelbarer Schaden an Menschen und Umwelt vermieden werden kann.

Alle Maschinen, die der Mensch erfunden hat, können Fehlfunktionen aufweisen. Es ist wichtig, die Fehlfunktionen so weit wie möglich auszuschließen und die Menschen so gut wie möglich vor Schaden zu bewahren. Daher wurden Regeln, z. B. Normen (DIN, IEC etc.) eingeführt, die Hersteller*innen und Anwender*innen zu beachten haben. Sie fordern bestimmte Produkteigenschaften und enthalten Anweisungen für die Hersteller, die einzuhalten sind, bevor ein Produkt in Verkehr gebracht werden darf. Ob ein Produkt die geforderten Eigenschaften auch wirklich hat bzw. die Regeln erfüllt, prüfen Einrichtungen wie der VDE oder der TÜV. Oft sind auch nach dem „Inverkehrbringen" regelmäßige Kontrollen vorgeschrieben (beispielsweise die regelmäßige Aktualisierung der TÜV-Plakette beim Auto). Wie häufig Kontrollen vorgeschrieben werden, richtet sich nach der Eintrittswahrscheinlichkeit und der Schadenshöhe. Die folgende Abbildung zeigt das dazu heute etablierte System.[17]



16 Siehe dazu die folgenden Abschnitte von Timo Rademacher und Fruzsina Molnár-Gábor.

17 Die Bezeichnungen in der Abbildung und im sich darauf beziehenden weiterführenden Text stehen als Verallgemeinerung in der männlichen Form, sollen allerdings auch die weibliche einschließen.

Folgende Beispiele sollen das Schema erläutern: Ein autonom fahrender Bus wird von einem Fahrgast genutzt, von einem Bus-Unternehmen eingesetzt, von einem Hersteller produziert, und vom TÜV beim Inverkehrbringen und in der Folge regelmäßig überprüft. Ein Medizingerät wird an einem Patienten oder einer Patientin angewendet (Nutzer), von einem Arzt bzw. einer Ärztin oder dem Krankenhaus betrieben, von einem Medizingeräte-Hersteller entwickelt und produziert und in der EU von einer sogenannten „benannten Stelle" vor der Freigabe und in der Folge regelmäßig geprüft. Es kann auch vorkommen, dass Nutzende und Betreibende identisch sind, beispielsweise beim autonom fahrenden privaten Pkw. Manchmal sind Nutzende, Hersteller und Betreibende alle unter einem Dach in einem Unternehmen vereint, sodass KI für die Fertigung selbst entwickelt und dort auch eingesetzt wird.

Hinter all den Kästchen in der Abbildung stehen – wenn wir nach der Verantwortung fragen – Menschen. Beim Nutzer ist das offensichtlich. Auch der „Betreiber" ist ein Mensch: Es ist zunächst der CEO, der als Repräsentant*in des Unternehmens die Verantwortung trägt. Dieser kann Teile seiner Verantwortung an seine Mitarbeiter*innen „delegieren". Hinter dem CEO der Herstellerfirma stehen Entwickler*innen und Produktionsleiter*innen, die ihren Teil der Verantwortung übernehmen müssen.

Wer trägt nun in dem in der Abbildung gezeigten Netz, beispielsweise in einem Schadensfall, welche Art von Verantwortung? Meistens wird kein Akteur ganz von seiner Verantwortung freigestellt. Jeder zeichnet für seinen Teil verantwortlich. Das soll im Folgenden anhand von Beispielen erläutert werden. Wenn ein Nutzer eine technische Einrichtung außerhalb des vom Hersteller angegebenen bestimmungsgemäßen Gebrauchs („intended use") einsetzt, trägt er dafür die Verantwortung. Auch wenn z. B. ein Patient oder eine Patientin eine Gesundheits-App mit KI einsetzt, die nicht als Medizinprodukt in der EU angemeldet und geprüft ist, oder die für diesen Fall nicht bestimmt ist, trägt er oder sie dafür selber die Verantwortung.

Wenn ein Betreiber seiner Sorgfaltspflicht nicht nachkommt und die technische Einrichtung nicht regelmäßig auf die ordnungsgemäße Funktion überprüft, so ist der Betreiber für einen möglicherweise dadurch eintretenden Schaden verantwortlich. Auch für den zukünftigen Einsatz von KI in der Medizin wird es Prüfprozeduren geben, mit denen der Arzt oder die Ärztin das ordnungsgemäße Funktionieren des KI-Systems regelmäßig überprüfen (lassen) muss. Wenn der

Hersteller Eigenschaften seines Produktes verspricht, die bei einem bestimmungsgemäßen Gebrauch nicht eingehalten werden, trägt er dafür die Verantwortung. Verspricht ein Hersteller eines KI basierten Befundungssystems eine Trefferquote von 90 Prozent, so muss das auch im klinischen Alltag zutreffen. Andernfalls ist der Hersteller für einen möglichen Schaden verantwortlich. Hat eine Prüfeinrichtung nicht alle versprochenen Eigenschaften eines technischen Systems sorgfältig überprüft, so trägt auch sie eine Mitverantwortung an einem möglicherweise eintretenden Schaden. Die Problematik der Verantwortung des Einzelnen, der/die Teil eines großen Teams ist, ist bereits detailliert untersucht.[18]

Was bedeutet das also für Systeme, die KI-Komponenten enthalten? Auch sie müssen so entworfen und implementiert werden, dass von ihnen so weit wie möglich keine Gefahr für Mensch und Umwelt ausgehen kann. Sie müssen – wo immer nötig – in regelmäßigen Abständen überprüft werden. Das ist insbesondere bei selbstlernenden Systemen wichtig, da sie ja ihre Eigenschaften kontinuierlich und selbständig ändern. In erster Näherung können also die seit langem bekannten Methoden auf technische Systeme mit KI-Komponenten übertragen werden. Eine Schlüsselrolle nehmen dabei offenbar die Normen und Prüfprozeduren ein. Hier sind für jede Anwendung spezifische neue Normen und Prüfprozeduren zu definieren, woran die Normungseinrichtungen derzeit arbeiten.[19] Der Vorgang ist nicht einfach, weil das Thema einerseits technisch anspruchsvoll ist und andererseits eine europäische Lösung gefunden werden muss – noch besser wären natürlich internationale Normen. Insbesondere die Prüfprozeduren, mit denen möglichst eindeutig und reproduzierbar die versprochenen Eigenschaften des KI-basierten Produktes kontrolliert werden, sind eine Herausforderung.[20]

Besonders häufig wird als problematisches Beispiel ein versteckter „Bias" einer KI-Anwendung zitiert. Das bedeutet: Die App bevorzugt oder diskriminiert unbemerkt Menschen einer Untergruppe der Nutzenden.[21] Wurde ein Trainingsdatensatz eingesetzt, der nur eine Untergruppe aller Menschen einschließt, so muss das in den Spezifikationen genannt werden und die KI-Anwendung sollte nur für diese Untergruppe eingesetzt werden. Wer ist nun für die hierdurch möglicherweise

18 Lenk, H (2017): a.a.O.

19 Wahlster, W; Winterhalter, C (Hrsg.) (2020): Deutsche Normungsroadmap Künstliche Intelligenz. DIN, DKE: Berlin. https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf [15.3.2021].

20 Beining, L (2020): Vertrauenswürdige KI durch Standards? Herausforderungen bei der Standardisierung und Zertifizierung von Künstlicher Intelligenz. Stiftung Neue Verantwortung. Berlin. https://www.stiftung-nv.de/sites/default/files/herausforderungen-standardisierung-ki.pdf.

21 Siehe das Kapitel von Jonas Peters in diesem Heft.

verursachten Schäden verantwortlich? Der Anwender trägt die Verantwortung, wenn das KI-System außerhalb der angegebenen Spezifikationen eingesetzt wurde, also z. B. für Menschen, die nicht im Trainings-Datensatz berücksichtigt wurden. Der Entwickler hat die Verantwortung, dass alle heute bekannten Möglichkeiten eines Bias vermieden bzw. deklariert werden. Heute noch nicht bekannte Möglichkeiten des Bias müssen nicht berücksichtigt werden, weil dies gar nicht möglich ist.[22] Wenn kein Mensch den möglicherweise eingetretenen Schaden hätte verhindern können, spricht man von einem „schicksalshaften Ausgang" und niemand ist dafür verantwortlich.

Es sieht so aus, als könnten wir mit dem heute etablierten System der Zuordnung von Verantwortung bei der Entwicklung und beim Einsatz von technischen Systemen auch die Systeme mit KI soweit abdecken. Allerdings müssen die Regeln für die Entwicklung, Herstellung (Normen) und Vertrieb und für den Einsatz dieser Systeme angepasst bzw. erweitert werden. Das heißt auch, dass eine genaue Angabe von KI-typischen Spezifikationen gefordert und geeignete Prüfprozeduren festgelegt werden müssen.[23] Nutzer, Betreiber, Hersteller und Prüfeinrichtungen müssen in Schulungen auf ihre Verantwortung hingewiesen werden, die sie bei der Entwicklung und beim Einsatz von KI-Systemen übernehmen. Das erfordert Wissen über die Grenzen und Möglichkeiten von KI, welches vermittelt werden muss.

Es wäre nicht zielführend, technischen Systeme mit KI ein Gefahrenpotenzial vorzuhalten, welches sich durch Normen, Regeln und Kontrollen so gut beherrschen lässt wie bei anderen technischen System auch. Nicht verschwiegen werden soll aber die Gefahr, dass die Normen und das Wissen der Nutzer der Wirklichkeit hinterherlaufen und nur verzögert wirksam werden. Herausforderungen ergeben sich zudem durch die Multikausalität der Ergebnisse von komplexen KI-Systemen im Zusammenhang mit deren Vernetzung und „Autonomie".

22 Menschen, die ohne die KI-Anwendung Entscheidungen treffen, machen unbemerkt den gleichen Fehler.

23 Wahlster, W; Winterhalter, C (2020): a.a.O.

C. ZUR RECHTSWISSENSCHAFTLICHEN DISKUSSION UM VERANTWORTUNG UND HAFTUNG

Die rechtswissenschaftliche Diskussion um Verantwortung und Haftung im Zusammenhang mit KI-Systemen ist – relativ zum tatsächlichen Entwicklungsstand der Systeme – weit fortgeschritten und dadurch schwer durchschaubar geworden;[24] zu einem gewissen Grad spiegelt die Diskussion damit auf ungute Weise ihren Untersuchungsgegenstand wider. Bevor auf einige spezielle Aspekte der Diskussion in den folgenden Kapiteln dieses Abschnitts eingegangen wird, soll gleichwohl der Versuch lohnen auf relativ hoher Abstraktionshöhe Aussagen herauszuarbeiten, die im Zusammenhang von Verantwortung und Haftung mittlerweile als weitgehend konzediert gelten können, um auf diese Weise wieder einen gemeinsamen Boden, oder besser: ein gemeinsames Plateau der Diskussion zu erreichen.

Zunächst scheint eine gewisse Einigkeit im nun angezeigten regulatorischen Zugriff zu bestehen. Dieser soll, so der gegenwärtige „Mainstream", weitgehend sektor- und risikospezifisch erfolgen, und auf *allgemeine*, insbesondere *starre* allgemeine regulatorische Zugriffe verzichten.[25] Beispielhaft können hierfür das Gutachten der Datenethikkommission der Bundesregierung[26] sowie der Vorschlag der Europäischen Kommission für einen Artificial Intelligence Act genannt werden.[27] Dieser Ansatz ist zu begrüßen, denn es macht einen Unterschied, ob die KI-basierte Automatisierung personalisierte Werbung, die Einladung zum

24 Siehe für eine Übersicht, allerdings ohne Anspruch auf Vollständigkeit, die Beiträge in Wischmeyer, T; Rademacher, T (Hrsg.) (2020): Regulating Artificial Intelligence. Springer Nature: Cham.

25 Ein allgemeiner Zugriff besteht aktuell v. a. mit Art. 22 DSGVO, der als die eine, zumindest in Ansätzen auch KI-spezifische Vorschrift des allgemeinen Datenschutzrechts gelten kann. Die Norm verbietet belastende automatisierte Einzelfallentscheidungen grundsätzlich, allerdings nur dann, wenn diese „ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhen". Europäische Union (2016): Datenschutzgrundverordnung. In: Amtsblatt der Europäischen Union. Verordnung 2016/679 des europäischen Parlaments und des Rates vom 27. April 2016. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679 [15.3.2021].

26 Datenethikkommission (2019): Gutachten der Datenethikkommission, Berlin. https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_DE.pdf?__blob=publicationFile&v=5 [15.3.2021].

27 Europäische Kommission (2021) Proposal for an Artificial Intelligence Act. COM (2021) 206 final. Brüssel, 21.4.2021. Der Vorschlag sieht größtenteils Regelungen für sog. „high risk"-Anwendungen vor, lediglich die Art. 5, 52 enthalten „horizontale" Regelungen, z.B. das sehr allgemein gefasste Verbot manipulierender KI-Software. Die Vorschläge der Kommission konnten in diesem Beitrag wegen ihrer Publikation unmittelbar vor Drucklegung im Einzelnen keine Berücksichtigung mehr finden.

Vorstellungsgespräch, die Sperrung des Wasser- oder Stromanschlusses oder das Abfeuern eines Marschflugkörpers betrifft.

Zweitens besteht auch zunehmende Einigkeit darüber, was die regulatorischen Herausforderungen von KI anbelangt. Hier lassen sich, grob kategorisiert, fünf Aspekte unterscheiden:[28]

- KI-Systeme können *erstens* potentiell allgegenwärtig und vor allem auch grenzüberschreitend tätig sein. Das macht sie von vornherein zu einem für supranationale, d. h. europäische Ansätze prädestinierten Regelungsgegenstand. Wegen der großen Attraktivität des europäischen Binnenmarktes auch für außereuropäische Wirtschaftsteilnehmer kann eine solche Regulierung auch erhebliche internationale Prägungswirkung entfalten. Die EU-Kommission hat die Herausforderung bereits angenommen, und Vorschläge mit dem Data Governance Act, dem *Digital Services Act* und dem *Digital Markets Act* auch schon vorgelegt; der Vorschlag für eine eigene High-Risk-KI-Regulierung und für einen *Data Act* ist für Ende 2021 angekündigt.

- Die *zweite* anerkannte Herausforderung wird unter dem Schlagwort „Autonomierisiko" behandelt.[29] KI verfügt bekanntlich über sehr spezifische Formen von „Autonomie": sei es im anspruchsvollen Sinne bei selbstlernenden Systemen, sei es auch nur wegen der oft beschriebenen „black-box"-Haftigkeit der Prognosen, die den *Anschein* erwecken, diese seien unvorhersehbar und damit autonom. Hierher gehören nun die vielfältigen Versuche, KI beizubringen, „sich" (bzw. besser: ihre Prognosen und Klassifikationen) so zu erklären, dass dies menschlich nachvollziehbar und möglichst schon im Moment der eventuellen Weiterverwendung hinterfragbar wird. Das Stichwort hierfür lautet, wie in der Einleitung dieses Heftes bereits erwähnt, „explainable AI". Die tatsächliche oder vermeintliche Autonomie von KI-Systemen ist es dann auch, die allenthalben auch die Frage aufkommen lässt, ob nicht bzw. ab wann KI selbst zum Verantwortungs*subjekt* werden

28 Vgl. hierzu schon Rademacher, T (2020): Künstliche Intelligenz und neue Verantwortungsarchitektur. In: Eifert, M (Hrsg.): Digitale Disruption und Recht. Nomos Verlag: Baden-Baden, S. 45 ff.

29 Zech, H (2020): Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung? 73. Deutscher Juristentag, 2020/2022, Gutachten I/A. C.H. Beck: München, S. A 31 ff. m. w. N.

könne.[30] Freilich besteht insoweit schon wieder gar keine Einigkeit mehr, und hier kommt es auch stark auf das Rechtsgebiet an: Zivilrechtlich, d. h. wenn es um Haftung im rein monetären Sinn geht, spricht wenig dagegen, Vermögensmassen, die an KI „angegliedert" sein können, als Subjekte der Haftung zu bestimmen. Strafrechtlich gedacht ist es hingegen, vorsichtig formuliert, schon deutlich anspruchsvoller, im Abschalten eines KI-Systems wirklich eine Entsprechung zur Todesstrafe zu sehen – anders formuliert: Es ist zweifelhaft, ob Maschinen überhaupt in einer Weise bestraft werden können, die strafrechtliche Zwecke erfüllen würde[31] (neben General- und Spezialprävention – je nach Auffassung – auch noch Sühne und Vergeltung). Die Frage nach dem Verantwortungssubjekt soll allerdings nicht Thema des vorliegenden Abschnitts sein.

- *Drittens*, um wieder zu Anerkanntem zurückzukehren, weisen digitale Systeme mit ihren vergangenheitsbasierten Trainingsmethoden spezifische Pfadabhängigkeiten auf, d. h. sie projizieren das aus der Vergangenheit Gelernte potentiell auf die Gegenwart und die Zukunft und können damit individuelle menschliche Diskriminierungen oder, neutraler formuliert, Einstellungen dauerhaft perpetuieren, verstärken oder sogar auf eine gesellschaftlich-systemische Ebene ausdehnen. Hier geht es regulatorisch nun vor allem um Methoden des „debiasing"[32], deren sorgfältiger Einsatz im Rahmen von Zulassungs-, Zertifizierungs- oder auch Haftungsprozessen nachzuweisen sein kann. Zusätzlich wird aber auch über Arrangements diskutiert, die Diversität und Pluralität der Gesellschaft insbesondere beim Einzelnen sicherstellen sollen; zu nennen ist hier etwa die Diskussion über die Frage, ob Social-Media-Anbieter verpflichtet werden sollten, bestimmte Berichte z. B. von öffentlich-rechtlichen Medien anzuzeigen, auch wenn das den Präferenzen der Nutzer*innen wiederspricht, um Demokratie-Kompetenzen bei den Nutzer*innen zu erhalten.[33]

30 Vgl. zu dieser Diskussion z. B. Schirmer, J E (2020): Artificial Intelligence and Legal Personality: Introducing „Teilrechtsfähigkeit": A Partial Legal Status Made in Germany. In: Wischmeyer, T; Rademacher, T (Hrsg.): Regulating Artificial Intelligence. Springer Nature: Cham, S. 123 ff.; Gaede, K (2018) Künstliche Intelligenz – Rechte und Strafen für Roboter? Nomos: Baden-Baden; Teubner, G (2018): Digitale Rechtssubjekte? AöR 2018, S. 155 ff.

31 Dazu auch Rademacher, T (2020): Künstliche Intelligenz und neue Verantwortungsarchitektur. In: Eifert, M (Hrsg.): Digitale Disruption und Recht. Nomos Verlag: Baden-Baden, S. 45 ff. (54 ff.) m. w. N.

32 Vgl. zu Bias und Diskriminierung den Beitrag von Jonas Peters in diesem Heft.

33 Vgl. zum Thema auch den Beitrag von Frauke Rostalski und Thorsten Thiel in diesem Heft.

- Ferner sind KI-Systeme zumindest potentiell stark vernetzt, und zwar in zweifacher Hinsicht: einmal – *viertens* – untereinander, d. h. mit anderen digitalen Systemen (Vernetzungsrisiko), aber auch – *fünftens* – mit Menschen in konkreten Entscheidungszusammenhängen (sog. Verbundrisiko von hybriden Entscheidungsarrangements).[34] Eine solche hybride Konstellation liegt beispielsweise schon bei der Entscheidung über eine Kreditvergabe vor, wenn diese formell noch von Sachbearbeiter*innen einer Bank getroffen wird, die Entscheidung vom zuvor durchgeführten automatisierten *credit scoring* aber so stark vorbeeinflusst ist, dass Zweifel bestehen, wie viel die Sachbearbeiter*innen noch selbst entscheiden *können* und – das wird häufig übersehen – überhaupt noch entscheiden *wollen*. Auch das Entscheiden-*Wollen* ist eine Kompetenz- und Verantwortungsfrage. Zugleich kann das *credit scoring* selbst Daten aus diversen (digitalen) Quellen zusammenführen, womit im selben Entscheidungskontext zum Verbundrisiko auch ein Vernetzungsrisiko hinzutreten kann, was dann die Nachverfolgung der entscheidungsrelevanten Daten und Vorgänge *ex post* weiter erschwert. Dieses Verbund- bzw. Vernetzungsrisiko wird angesichts der immer drängenderen Forderungen nach Interoperabilität, Zentralisierung und Portabilität von Anwendungen, Datenbanken bzw. Daten erheblich zunehmen. Als aktuelles Beispiel sei nur an die alten und neuen Corona-Apps gedacht, die untereinander und europaweit interoperabel sein sollen, oder die Datenbanken der Gesundheitsämter, die möglichst engmaschig miteinander verknüpft werden soll(t)en. Kern beider Risiken ist jedenfalls ein Wissensproblem: Vernetzungen und Verbundstrukturen machen es immer schwieriger zu ermitteln, ob bzw. wo das für die Ausübung von Kompetenz und damit die Zuschreibung von Verantwortung eigentlich erforderliche Entscheidungswissen zum relevanten Zeitpunkt lag bzw. – im Fall von Fehlern – gerade nicht lag. Der Herausforderung wird durch Versuchen begegnet, Stellvertreterhaftungen zu etablieren oder auch Gefährdungshaftungsregime, die den relevanten Entscheidungszeitpunkt nach vorne, in der Regel auf den Moment der Produkteinführung, verlegen.[35] Man kann sich hier auch ein Beispiel an der allgemeinen Unfallversicherung nehmen, wie es Herbert Zech vorgeschlagen hat. Der Vorzug wäre, dass dieser Schutz bei jedem „KI-Unfall" greifen würde, ohne dass noch einzelne Verursachungs- und Verantwortungsbeiträge identifiziert werden müssten.[36]

34 Zu beidem ausf. Teubner, G (2018): a.a.O., S. 155 ff. (201 ff. bzw. 196 ff.).

35 Siehe zur Gefährdungshaftung den Beitrag von Fruzsina Molnár-Gábor in diesem Heft.

36 Zech, H (2020): a.a.O., S. A 106 ff.

Das nachfolgende Kapitel fokussiert auf den aktuellen Diskussionsstand hinsichtlich der Gefährdungshaftung.

Fruzsina Molnár-Gábor

D. HERAUSFORDERUNGEN AN DIE HAFTUNGSRECHTLICHEN BESTIMMUNGEN DURCH KI

KI-basierte Anwendungen stellen Herausforderungen an die haftungsrechtlichen Bestimmungen, die an ihre Entwicklung und ihren Einsatz anknüpfen. Die konkreten Herausforderungen liegen in der allgemeinen Steuerbarkeit sowie im individuellen und gesellschaftlichen Umgang mit KI-Systemen. Die Vernetzung von Akteuren und ihren Handlungen[37] im Entwicklungs- und Anwendungsprozess von KI-Systemen soll dabei an erster Stelle hervorgehoben werden. Die dadurch entstehende Multikausalität[38] und mittelbare Kausalität[39] zwischen Ergebnissen und Handlungen sowie ihr Zusammenspiel mit den autonomen Eigenschaften eines KI-Systems[40] tragen maßgeblich auch zu einer fehlenden Erklärbarkeit entstehender Ergebnisse eines solchen Systems bei. Die fehlende Erklärbarkeit verhindert wiederum eine klare Zurechenbarkeit[41] von Entscheidungen und Ergebnissen. Es lässt sich häufig nicht aufklären, ob zwischen entstandenen Schäden und einer Handlung bzw. Handlungen überhaupt ein kausaler Zusammenhang besteht und wenn ja, wie dieser ausgestaltet ist. Darüber hinaus verursachen die selbstlernenden Eigenschaften von KI-Systemen je nach ihrem genauen Freiheitsgrad

37 Zur Vernetzung der Akteure im digitalen Ökosystem im Allgemeinen: Groß, M; Krellmann, A (2019): Das Ökosystem der Digitalisierung. In: Stember, J; Eixelsberger, W; Spichiger, A; Neuroni, A; Habbel; F R; Wundara, M (Hrsg.): Handbuch E-Government. Springer Fachmedien Wiesbaden, S. 3 ff. Zur Vernetzung spezifisch im Kontext der Künstlichen Intelligenz: Teubner, G (2019): Digitale Rechtssubjekte? Haftung für das Handeln autonomer Softwareagenten. In: VerfBlog, 2019/9/30, https://verfassungsblog.de/digitale-rechtssubjekte-haftung-fuer-das-handeln-autonomer-softwareagenten/ [15.3.2021]; Zech, H (2019): Künstliche Intelligenz und Haftungsfragen. In: ZfPW, S. 118 (202).

38 European Commission (2019): Liability for Artificial Intelligence and other emerging digital technologies, Report from the Expert Group on Liability and New Technologies – New Technologies Formation. Brüssel, S. 19. https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608 [15.3.2021].

39 Zech, H (2019): a.a.O., S. 207.

40 Zur Autonomie des Systems als Wirkung nach außen: Konertz, R; Schönhof, R (2020): Das technische Phänomen „Künstliche Intelligenz" im allgemeinen Zivilrecht. Eine kritische Betrachtung im Lichte von Autonomie, Determinismus und Vorhersehbarkeit, Nomos: Baden-Baden, S. 69.

41 Vgl. bspw. BeckOGK/Spindler BGB § 823 Rn. 738.

neuartige Risiken, mit denen der Umgang bestimmt werden sollte.[42] Die haftungsrechtlichen Ansätze der rechtswissenschaftlichen Literatur versuchen, die fehlende Erklärbarkeit von KI-basierten Ergebnissen und die selbstlernenden Eigenschaften zu adressieren, indem sie entweder eine Zurechnung konstruieren oder auf das Erfordernis der Zurechnung und damit auch auf das der Kausalität verzichten.

Neben dem Haftungsrecht *de lege lata (nach geltendem Recht)* (Verschuldenshaftung nach § 823 Abs. 1 BGB und vor allem Produkthaftung nach ProdHaftG) wird über die Einführung von Gefährdungshaftungsregelungen *de lege ferenda (nach zukünftigem Recht)* nachgedacht, mit denen die KI-spezifischen Risiken insbesondere der Vernetzung der Akteure und des selbstlernenden Verhaltens des KI-Systems internalisiert werden.[43]

In beiden Fällen ist es Ziel der haftungsrechtlichen Regelungen, einen Ausgleich für erlittene Nachteile zu schaffen und schädigendes Verhalten zu verhindern.[44] Dabei ist das Verhältnis zwischen den kompensatorischen und präventiven Funktionen des Haftungsrechts umstritten.[45] Die kompensatorische Funktion haftungsrechtlicher Bestimmungen ist darauf ausgerichtet, den Rechtsschutzzweck durch die Wiederherstellung des verletzten Rechts mit dem Schadensausgleich zu erreichen. Durch das Prinzip des Ausgleichs und den dadurch auf die Ausgleichenden auferlegten Nachteil werden diese ihr Verhalten darauf ausrichten, diesen Nachteil zu minimieren oder – wenn möglich – gänzlich zu vermeiden. Wenn durch ein bestimmtes Verhalten die Minimierung oder Abwendung des Nachteils erreicht werden kann, werden die Ausgleichspflichtigen ihr Verhalten entsprechend sorgfältig gestalten, um eine Haftung aus Verschulden möglichst nicht zu riskieren. Aber auch im Rahmen der verschuldensunabhängigen Gefährdungshaftung ist das Prinzip der Prävention feststellbar: Bei der Ausübung gefährlicher Tätigkeiten wird darauf abgezielt, das Haftungsrisiko zu minimieren und den Eintritt von Schadensfällen zu verhindern.[46] Damit lässt sich festhalten, dass die präventive

42 Scherer, M U (2016): Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. In: Harvard Journal of Law & Technology 29 (2), S. 362 ff.

43 Beispielhaft für den Überblick s. Frost, Y; Kießling, M (2020): Künstliche Intelligenz im Bereich des Gesundheitswesens und damit verbundene haftungsrechtliche Herausforderungen. In: MPR. Medizin Produkte Recht – Zeitschrift für das gesamte Medizinprodukterecht MPR 5, S. 178 (180 ff.).

44 MüKoBGB/Oetker BGB § 249 Rn. 8, 9.

45 Zum Ausgleichsprinzip vgl. Jansen, N (2005): Konturen eines europäischen Schadensrechts. In: Juristen Zeitung 60 (4), S. 160–173 (162 f.). Zum Präventionsprinzip: Soergel, T; Ekkenga, J; Kuntz, T: Vor § 249 Rn. 28, § 254 Rn. 3; zur Prävention als Sekundärfunktion: Thüsing, G (2001): Wertende Schadensberechnung, C.H. Beck: München, S. 14 ff., S. 16 ff.

46 Möller, R (2019): Das Präventionsprinzip des Schadensrechts, Duncker & Humblot: Berlin, S. 258.

Funktion eng mit der kompensatorischen Funktion des Haftungsrechts verbunden ist und sich daraus – im Dienste der Verhaltenssteuerung – ableiten lässt.

Beim haftungsrechtlichen Umgang mit KI-Risiken wird häufig die (erweiterte) *Gefährdungshaftung* hervorgehoben.[47] Neue Tatbestände der Gefährdungshaftung werden als Ergänzung zu den punktuellen Modifikationen der Verschuldenshaftung angesehen.[48] Die Gefährdungshaftung ist eine verschuldensunabhängige Haftung aufgrund der Schaffung und Erhaltung einer Gefahrenquelle. Normativ adressiert wird dabei in der Regel derjenige, der einen Nutzen zieht und das Risiko zumindest abstrakt beherrschen kann.[49] Eine Gesamtanalogie zu den bestehenden Gefährdungshaftungstatbeständen würde eine richterliche Rechtsfortbildung voraussetzen, um eine Haftung für alle Quellen erhöhten Risikos zu schaffen.[50] Dies gilt insbesondere für solche Risiken, die durch den Einsatz von KI entstehen, wie Vernetzungsrisiken und Risiken durch die selbstlernenden Eigenschaften. Für diese Lösung spricht grundsätzlich die damit einhergehende Rechtssicherheit, jedoch bleibt zunächst offen, ob eine solche Haftung die Betreiber- oder die Herstellerseite einbeziehen sollte, denn Gefahrenquellen können auf beiden Seiten vorliegen.

47 Siehe dazu bspw. im medizinischen Bereich: Droste, W (2018): Intelligente Medizinprodukte: Verantwortlichkeiten des Herstellers und ärztliche Sorgfaltspflichten. In: MPR. Medizin Produkte Recht – Zeitschrift für das gesamte Medizinprodukterecht 4, S. 109; zum autonomen Fahren: Meyer, S (2018): Künstliche Intelligenz und die Rolle des Rechts für Innovation. In: Zeitschrift für Rechtspolitik 51 (8), S. 233–238 (236); allgemein auch Zech, H (2019): a.a.O., S. 219.; vgl. Bertolini, A (2020): Artificial Intelligence and Civil Liability, Study Requested by the European Parliament's Committee on Legal Affairs, PE 621.926, Brüssel. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf [15.3.2021]. Allerdings wird unter Berücksichtigung deliktsrechtlicher Überlegungen eine Abkehr vom Verschuldensprinzip häufig als nicht geboten kritisiert, vgl. Steege, H (2021): Auswirkungen von künstlicher Intelligenz auf die Produzentenhaftung in Verkehr und Mobilität, NZV - Neue Zeitschrift für Verkehrsrecht 6 (13); für differenzierte Regelungen plädiert: Borges, G (2018): Rechtliche Rahmenbedingungen für autonome Systeme, NJW - Neue Juristische Wochenschrift, S. 977–982 (982).

48 Zur Position der Datenethikkommission, die sich zudem zur Anpassung der Produkthaftungsrichtlinie geäußert hat, s. Datenethikkommission (2019): Gutachten der Datenethikkommission, Berlin, S. 220 f. sowie Teil F. https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_DE.pdf?__blob=publicationFile&v=5 [15.3.21].

49 Zur Veranschaulichung der Gefährdungshaftung seien hier beispielhaft die Haftung des Kraftfahrzeughalters (§ 7 Abs. 1 StVG) und die Haftung des Betreibers nach dem Gentechnikgesetz (§ 32 Abs. 1 GenTG) genannt, zum Verhältnis Technikermöglichung und Haftungsregelungen gentechnischer Aktivitäten s. Kloepfer, M (2016): Umweltrecht, C.H. Beck: München, § 6 Rn. 236.

50 Zech, H (2019): a.a.O., S. 215f.; selbst die Gesamtanalogie ist allerdings umstritten: Insgesamt kommt eine Gesamtanalogie zumindest bei einer planwidrigen Regelungslücke und einer vergleichbaren Interessenlage in Betracht. Sommer, M (2020): Haftung für autonome Systeme. Verteilung der Risiken selbstlernender und vernetzter Algorithmen im Vertrags- und Deliktsrecht. Nomos: Baden-Baden, S. 213 ff.

Der Gedanke der Gefährdungshaftung knüpft daran an, dass eine Gefahrenquelle für denjenigen kontrollierbar ist, der sie im Verkehr eröffnet. Bei einer Gefährdungshaftung des Betreibers wäre er für die Ausführungsschritte und Ergebnisse der KI haftbar. Zwar eröffnet der Betreiber die Gefahr, wenn er ein KI-System im Verkehr einsetzt, jedoch fehlt ihm die Kontrollierbarkeit des Systems, da für ihn, vor allem aufgrund der selbstlernenden Eigenschaften der KI, nur eingeschränkte Einwirkungsmöglichkeiten bestehen.[51] Gleichwohl wäre die Gefährdungshaftung des Betreibers eine innovationsfreundliche Lösung, da sie den Hersteller entlastet.[52] Bei dieser Bestimmung des Haftpflichtigen ist zu beachten, dass private Anwender in der Regel einen vergleichsweise minderen Einfluss auf das Verhalten von KI-Systemen haben werden, da sie das System hauptsächlich nur bedienen (können) und nicht weiter beeinflussen können.

Der Hersteller nimmt durch die Gestaltung der Programmierung und des Trainings eines Algorithmus den vergleichsweise größten gestalterischen Einfluss darauf, wie die KI lernt und sich entwickelt. Dadurch nimmt er Einfluss auf spätere Arbeitsschritte und Ergebnisse des KI-Systems. Diese Argumente sprechen für eine Gefährdungshaftung des Herstellers aufgrund dem Inverkehrbringen einer Gefahrenquelle in Form eines KI-Systems. Gegen eine Gefährdungshaftung des Herstellers spricht jedoch, dass die Fehlerhaftigkeit des Produkts nicht zwangsläufig im Einflussbereich des Herstellers liegt, sondern während der Weiterentwicklung des Systems aufgrund selbstlernender Eigenschaften entstehen kann. Die Risiken bei einem KI-System sind in der Regel nicht vollständig zu überblicken und können somit nicht billigerweise in Kauf genommen werden.[53] Darüber hinaus wird eingewandt, dass die Gefährdungshaftung zwar ein „passendes Ventil für technische Risiken" sei, sie aber eine „starke Regelungswirkung" entfalte und damit möglicherweise innovationshemmend wirken könnte[54], was die Wettbewerbsfähigkeit auf dem internationalen Markt negativ beeinflussen würde.[55]

Der Vorschlag einer Haftungsbestimmung für KI-Systeme auf Grundlage der Gefährdungshaftung wird zudem dadurch in den Vordergrund gestellt, dass

51 Brunotte, N (2017): Virtuelle Assistenten – Digitale Helfer in der Kundenkommunikation. In: CR – Computer und Recht 33 (9), S. 583. https://doi.org/10.9785/cr-2017-0908.

52 Schaub, R (2017): Interaktion von Mensch und Maschine. In: JZ – Juristen Zeitung 72 (7), S. 342–349 (344).

53 Spindler, G. (2015): Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien? In: CR – Computer und Recht 31 (12). S. 766 (775). https://doi.org/10.9785/cr-2015-1205.

54 Ibid, S. 768.

55 Schaub, R (2017): a.a.O.

die Prävention nicht nur als eine nützliche Folge der Kompensation für Schaden gesehen wird, sondern sie als Zweck der Haftungsnorm verstanden wird. Die Gefährdungshaftung baut auf dem gemeinsamen Gedanken auf, dass derjenige, der eine besondere „Gefahr" schafft, unterhält und daraus den Nutzen zieht, reziprok der intendierten Wertschöpfung auch den Schaden tragen soll.[56] Durch eine Gefährdungshaftung wird auch die Risikoabschätzung delegiert und damit nicht nur das Sorgfaltsniveau, sondern auch das Aktivitätsniveau gesteuert.[57] Wenn jedoch eine entfaltete Aktivität trotz sorgfaltsgerechter Ausführung ein erhebliches Schadensrisiko begründet, wie dies bei KI-Systemen regelmäßig der Fall sein sollte, führt die Gefährdungshaftung nicht nur kompensatorisch zu effektiven Ergebnissen, sondern kann auch dazu führen, dass die betroffenen Aktivitäten einfach nicht ausgeübt werden, womöglich aus Zurückhaltung vor dem Risiko.[58] Dabei ist das grundsätzliche Wissensproblem bei KI-Systemen zu beachten, dem durch den Hersteller und den Betreiber selbst nur in bestimmten Grenzen geholfen werden kann. Daher sind Maßnahmen auch auf politischer Ebene notwendig, die zum besseren Umgang mit den Risiken führen und sich im Bereich der Transparenz, Erklärbarkeit und Plausibilität und damit letztlich eines erhöhten Verständnisses der Funktion von KI-Systemen verorten lassen. Solche Maßnahmen können auf vielfältige Weise verwirklicht werden, von Forschungsförderung bis hin zum Training und zur Schulung, sowie in Form unterschiedlicher Tools der Kompetenzerhöhung von beteiligten Akteuren, Betreibern und Herstellern. Durch eine Erhöhung ihrer Kompetenzen mit solchen Systemen können Hersteller mit den Schadensrisiken bei der Konstruktion, Instruktion und Produktbeobachtung besser und sicherer umgehen und könnten diese steuern oder zu steuern lernen. Die Betreiber als maßgeblicher Nutznießer durch den konkreten Einsatz der Technik in verschiedenen Szenarien nehmen auch entscheidenden Einfluss auf bestehende Schadensrisiken und müssen über den passenden Umgang mit dem KI-System informiert werden und den Umgang damit erlernen. Auch vor dem Hintergrund eines kompetenten Umgangs mit KI-Systemen aus der Perspektive der Ausgestaltung haftungsrechtlicher Bestimmungen sollte der Fokus auf die weitere Erforschung von Interaktionen zwischen Mensch und Maschine gelegt werden.

56 MüKoBGB/Wagner, Vor § 823 Rn. 19 f., 29.

57 Eidenmüller, H (2018): Machine Performance and Human Failure. In: https://www.law.ox.ac.uk/business-law-blog/blog/2018/11/law-and-autonomous-systems-series-machine-performance-and-human [15.3.2021].

58 In diese Richtung schon Möller, R (2019): a.a.O.; spannend zum Vergleich Verschuldens- und Gefährdungshaftung aus ökonomischer Perspektive bzgl. Kosten-Nutzen-Abwägung: Faltmann, L (2017): Schadenersatz im Deliktsrecht aus rechtsökonomischer Perspektive. In: ZJS – Zeitschrift für das Juristische Studium 1/2017, S. 10–18 (13).

3. ENTWICKLUNG VON KI-SYSTEMEN

Jonas Peters

A. FAIRNESS UND DISKRIMINIERUNG IN DATEN-BASIERTEN ENTSCHEIDUNGSSYSTEMEN

Es gibt eine Reihe von Aufgaben, die daten-basierte Methoden inzwischen besser lösen als Menschen. Unter entsprechenden Bedingungen gilt dies beispielsweise für das Lippenlesen[59] oder die Erkennung von Brust-[60] oder Hautkrebs.[61] In vielen Fällen ist dies auf die großen Datenmengen zurückzuführen, mit der solche Systeme trainiert werden. Ein Computer kann sich mehr Bilder anschauen, als ein Mensch dies kann. Es ist anzunehmen, dass sich die Anzahl an Aufgaben, in denen daten-gestützte Systeme akkuratere Vorhersagen als Menschen liefern, weiter vergrößert. Je genauer die Vorhersagen solcher Systeme werden, desto mehr sind diese in Entscheidungen eingebunden – je nach Anwendung wäre ein Verzicht unverantwortlich.

In einigen Anwendungsgebieten geht es allerdings nicht nur um die Treffsicherheit von Klassifikationen, sondern auch um die Art und Weise, wie diese Vorhersagen entstehen. Insbesondere spielt es eine Rolle, welche Variablen in welcher Form verwendet wurden. Wenn eine Bank das Kreditausfallrisiko einer Person vorhersagt, um zu entscheiden, ob sie dieser Person einen Kredit gewährt, so scheint es vertretbar, dass die Bank das derzeitige Einkommen der Person berücksichtigt (soweit ihr dazu Daten vorliegen), nicht aber deren Hautfarbe.

59 Beispielsweise Assael, Y M; Shillingford, B; Whiteson, S; de Freitas, N (2016): LipNet: Sentence-Level Lipreading. arXiv 1611.01599.

60 Beispielsweise McKinney, S M; Sieniek, M; Shetty, S et al. (2020): International Evaluation of an AI System for Breast Cancer Screening. Nature 577, S. 89–94.

61 Beispielsweise Haenssle, H; Fink, C; Schneiderbauer, R; Toberer, F; Buhl, T; Blum, A; Kalloo, A et al. (2018): Man Against Machine: Diagnostic Performance of a Deep Learning Convolutional Neural Network for Dermoscopic Melanoma Recognition in Comparison to 58 Dermatologists. Annals of Oncology 29 (May).; Esteva, A; Kuprel, B; Novoa, R A; Ko, J; Swetter, S M; Blau, H M; Thrun, S (2017): Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. Nature 542, S. 115–118.

In den folgenden Überlegungen[62] beschränken wir uns auf ein System, das eine Zielvariable (z. B. gute Leistungen in einem Beruf) als Funktion gemessener Eigenschaften schätzt (z. B. Jahre und Art der Ausbildung). Im Maschinellen Lernen wird diese Funktion nicht direkt von Menschen aufgestellt, sondern aus Daten gelernt. Zusätzlich betrachten wir ein sogenanntes sensibles Merkmal oder geschütztes Attribut wie Hautfarbe, Alter oder Geschlecht, gegen das eine ungerechtfertigte Diskriminierung zu verhindern ist.[63] Im Bereich des fairen Maschinellen Lernens[64] wird versucht, Prinzipien aufzustellen, anhand derer man beim Design der Methoden die Fairness bezüglich des geschützten Attributs sicherstellt. Fairness ist hier also als Nicht-Diskriminierung zu verstehen. Ist der Zugang zu den Systemen und den Datensätzen gewährleistet, so hat man Kenntnis über den gesamten Entscheidungsprozess: Man kann nachvollziehen, welche Variablen dort wie eingehen. Aus der Tatsache, dass dies bei menschlichen Entscheidungen oft anders ist, resultiert die Hoffnung, dass durch den Einsatz maschineller Verfahren die Fairness insgesamt vergrößert werden kann. In wie weit liegt es aber in der Kompetenz der professionellen Entwickler die Fairness solcher Algorithmen und Methoden sicherzustellen?

Das Prinzip „fairness through unawareness" sieht vor, dass eine Methode fair ist, falls sie keinen Zugriff auf das geschützte Attribut hat. Dieser Ansatz ist unzureichend: Es gibt Situationen mit unvollkommener Information, in denen argumentiert wird, dass ein System nur fair sein kann, wenn das geschützte Attribut explizit berücksichtigt wird.[65] Zudem ist ein solcher Ansatz agnostisch bezüglich des Gebrauchs von Eigenschaften, die mit dem Attribut korrelieren (Haarlänge und Körperkraft korrelieren mit Geschlecht). Wann dürfen solche Eigenschaften verwendet werden? In der Auswahl zu einem Beruf, bei dem schwere Reparaturteile ausgetauscht werden müssen, scheint es vertretbar zu sein, körperliche Kraft zu berücksichtigen, solange zwischen Männern und Frauen mit gleicher Kraft nicht unterschieden wird. Die Haarlänge einer Person darf in dem Verfahren keine Rolle spielen. Beide Eigenschaften korrelieren mit dem Geschlecht, der Unterschied

62 Teile dieses Gedankengangs wurden bereits diskutiert in Peters, J (2019): Was Ist Fair? Kalender der Jungen Akademie. http://web.math.ku.dk/~peters/jonas_files/fair.pdf [15.3.2021].

63 In diesem Kapitel bezeichnet Diskriminierung immer eine ungerechtfertigte Diskriminierung.

64 Beispielsweise Barocas, S; Hardt, M; Narayanan, A (2019): Fairness and Machine Learning. fairmlbook.org [15.3.2021]; Narayanan, A (2018): 21 Fairness Definitions and Their Politics. Notes from the ACM FAT* (Fairness, Accountability and Transparency) Conference 2018. https://shubhamjain0594.github.io/post/tlds-arvind-fairness-definitions/ [15.3.2021].

65 Kusner, M J; Loftus, J; Russell, C; Silva, R (2017): Counterfactual Fairness. In: Advances in Neural Information Processing Systems, S. 4066–4076.

liegt jedoch in ihrem kausalen Einfluss[66]: Nur die Kraft hat einen kausalen Einfluss auf die erfolgreiche Ausübung der Tätigkeit. Die Entscheidung, ob es fair ist, Eigenschaften zu berücksichtigen, scheint also von der kausalen Struktur des Problems abzuhängen.[67]

Neben der Kausalstruktur spielen mehrere andere Aspekte eine Rolle: (1) Oftmals können wir entscheidende Größen nicht direkt messen. Beispielsweise wird unter anderem wissenschaftliche Exzellenz als kausal für die erfolgreiche Ausübung einer Tätigkeit an einer Universität angeführt; jene ist aber schwer zu definieren und zu ermitteln. Hier können Variablen erlaubt werden, die die entscheidende Größe approximieren, etwa die Qualität vorhandener wissenschaftlicher Publikationen, solange kein zusätzlicher direkter Einfluss vom geschützten Attribut auf den Proxy existiert. (2) Variablen scheinen unangemessen, wenn sie (aus Sicht der betroffenen Person) nicht veränderbar sind: Auf unsere Kraft haben wir zumindest teilweise einen Einfluss – wir können diese durch regelmäßiges Training vergrößern. Auch die eigene Ausbildung mag als veränderbar gelten, die Ausbildung der Eltern allerdings nicht. Letztere sollte also nicht als Kriterium verwendet werden dürfen, selbst wenn sie nachweislich kausal für die erfolgreiche Ausübung einer Tätigkeit ist. (3) Im Maschinellen Lernen wird die Funktion, die die Zielvariable schätzt, auf einem Datensatz gelernt. Wird ein etwaiger Bias, der in den Daten vorhanden ist, ignoriert, so kann dies zu ungewollten Effekten in der gelernten Funktion führen. Dies gilt unabhängig davon, ob die zugrundeliegenden kausalen Strukturen verstanden sind oder nicht. Beim „predictive policing" werden maschinelle Lernmethoden eingesetzt, um zukünftige Straftaten zu verhindern. Falls die in der Vergangenheit von der Polizei aufgenommenen Datensätze aber nicht repräsentativ sind – beispielsweise, weil durch ungleiches Patrouillieren bestimmte Personengruppen öfter überprüft wurden als andere – so besteht

66 In diesem Kapitel betrachten wir folgende Definition von Kausalität: Eine Zufallsvariable X hat einen kausalen Effekt auf eine Zufallsvariable Y, falls eine Intervention auf X zu einer Veränderung der Wahrscheinlichkeitsverteilung von Y führen kann. Eine Intervention auf X ändert dabei alleine den Mechanismus für X und lässt den Mechanismus aller anderen Variablen invariant (Pearl, J (2009): Causality: Models, Reasoning, and Inference. Cambridge University Press: New York, 2. Auflage).

67 Kusner, M J et al. (2017): a.a.O.; Kilbertus, N; Rojas-Carulla, M; Parascandolo, G; Hardt, M; Janzing, D; Schölkopf, B (2017): Avoiding Discrimination Through Causal Reasoning. In: Advances in Neural Information Processing Systems, S. 656–666.; Nabi, R; Shpitser, I (2018): Fair Inference on Outcomes. In: Zhang, J; Bareinboim, E (Hrsg.): Fairness in Decision-Making – the Causal Explanation Formula. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1) .

die Gefahr, dass dieselben Personengruppen in den maschinellen Prädiktionen überrepräsentiert sind. Dieser Effekt wurde in den USA beobachtet.[68 69]

Obige Argumente legen nahe, dass die Kompetenz zu entscheiden, ob ein System aufgrund von Diskriminierung benutzt werden darf oder nicht, nicht alleine bei den technischen Entwickler*innen der Methode liegen kann. Die Frage der Kausalität ist nicht im Allgemeinen zu klären und muss im Einzelfall überprüft werden. Eine Bank könnte behaupten, dass die Adresse, die in vielen Gegenden mit geschützten Attributen korreliert, kausal für den Kreditausfall ist: Nachbarn könnten Einfluss darauf haben, dass der Kredit zurückgezahlt wird. Randomisierte Studien zur Bestimmung von Kausalbeziehungen sind oftmals nicht verfügbar, sodass auf Hintergrundwissen zurückgegriffen werden muss. Ähnliches gilt für die Frage, welche Eigenschaften als veränderbar angesehen werden können: Dies kann nicht allgemein beantwortet werden, sondern muss im konkreten Kontext diskutiert werden. Es ist zudem davon auszugehen, dass sich derartige Einschätzungen über die Zeit verändern. Auch ob ein Bias in dem verwendeten Datensatz vorhanden ist, kann nicht durch allgemeingültige Prinzipien überprüft werden. Hier ist eine menschliche Bewertung, die auf die Möglichkeit des „out-of-the-box"-Denkens zurückgreifen kann, unverzichtbar.

Allgemeingültige Handlungsanweisungen für die technischen Entwickler*innen eines Systems, die die Fairness des Systems sicherstellen könnten, existieren nicht. Die Kompetenz im Einzelfall zu entscheiden, ob das System fair ist, liegt also hier beim erweiterten Entwicklungsteam, das über ausreichend Hintergrundwissen verfügen sollte. Je nach Anwendungsgebiet kann es zu Interessenkonflikten

68 Siehe Lum, K; Isaac, W (2016): To Predict and Serve? In: Significance, 7. Oktober 2016. https://doi.org/10.1111/j.1740-9713.2016.00960.x.

69 Es ist bekannt, dass in einigen medizinischen Anwendungen die Klassifikation von Krankheiten für bestimmte Personengruppen besser funktioniert als für andere (ein oft diskutiertes Beispiel ist der Zusammenhang zwischen Hautkrebserkennung und Hautfarbe (z. B. Lashbrook, A (2018): AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind. In: The Atlantic, August 2018. https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/ [15.3.2021]). In solchen Anwendungen sollte es kein geschütztes Attribut geben: Durch die zusätzlichen Bedingungen würde dies zu einer unerwünschten Reduktion der Klassifikationsgüte führen. Die Existenz eines Daten-Bias ist allerdings zu vermeiden. Bei medizinischen Diagnoseverfahren, die für bestimmte Bevölkerungsgruppen inhärent schwieriger sind als für andere, können Unterschiede in der Klassifikationsgüte nicht komplett ausgeräumt werden. In solchen Fällen kann es sinnvoll sein, medizinische Verfahren nur für bestimmte Personengruppen zuzulassen. Es sollte aber vermieden werden, dass das Erkennen von Krankheitsbildern für bestimmte Personengruppen aufgrund einer mangelnder Datenbasis schlechter funktioniert. Es ist entscheidend, die Datenaufnahme selbst fair zu gestalten und beispielsweise alle Bevölkerungsgruppen zu berücksichtigen um etwaige problemspezifische Unterschiede in der Klassifikationsgüte nicht unnötig zu vergrößern.

kommen und der Prozess sollte von außerhalb überprüft werden. Nicht alle Anwendungsbereiche sind jedoch gleich kritisch. Entscheidungssysteme in einigen Bereichen sollten daher einer stärkeren Kontrolle unterzogen werden als in anderen Bereichen. Ein Stufensystem wie es beispielsweise die Datenethikkommission vorsieht, könnte den Aufwand und die Genauigkeit festlegen, die für Zertifizierung und Zulassung des Systems erforderlich sind.[70] Notwendig hierfür ist, dass alle Algorithmen und Daten, auf denen die Methoden trainiert wurden, auf Nachfrage offengelegt und verfügbar gemacht werden können.

Sabine Ammon

B. *ETHICS-BY-DESIGN* IN FORSCHUNG UND ENTWICKLUNG VON KÜNSTLICHER INTELLIGENZ

Wie beschrieben, haben KI-Technologien ein großes Innovationspotential. Zugleich aber laufen viele Anwendungsmöglichkeiten Gefahr, sich sozial disruptiv auszuwirken. Ihre großflächige Verankerung kann zu starken, ungewollten gesellschaftlichen Veränderungen durch eine implizite Konditionierung und Normierung von Verhalten führen. Schwierig ist allerdings, mögliche Auswirkungen in frühen Forschungs- und Entwicklungsstadien gut zu überblicken, während in den späten Entwicklungsphasen aufgrund festgelegter Entwicklungspfade und hoher Kosten korrigierende Eingriffe immer schwieriger werden. Erschwerend kommt hinzu, dass es durch ein immer besseres Angebot an frei verfügbaren Softwarepaketen mittlerweile ohne nennenswerte Kenntnissen von KI-Technologien möglich ist, eigene KI-basierte Lösungen zu entwickeln. Auf diese Weise verschwimmt zunehmend die Abgrenzung zwischen Entwickler*innen und Nutzer*innen; ein Hinterfragen der zugrundeliegenden Methoden und Daten wird nicht gefördert. Umso wichtiger ist es, Forschungs- und Entwicklungsprozesse bereits in ihren Frühstadien ethisch robust und resilient zu gestalten, damit spätere Produkte und Anwendungen nicht im Gegensatz zu unseren gesellschaftlichen Werten stehen und auf sich wandelnde Rahmenbedingungen passend reagieren können.

Existierende oder mögliche Anwendungsprobleme von KI-Systemen lassen epistemische und ethische Schwachstellen hervortreten, ohne dass sie allein

70 Datenethikkommission (2019): Gutachten der Datenethikkommission, Berlin. https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_DE.pdf?__blob=publicationFile&v=5 [15.3.2021].

technisch behoben werden könnten und somit auch nicht aus den Computer- und Ingenieurwissenschaften selbst heraus gelöst werden könnten. So werfen z. B. Verfahren des Maschinellen Lernens Fragen zum Umgang mit Bias in Daten und Algorithmen auf. Da Algorithmen im Maschinellen Lernen aus Daten lernen, wird ein in den Daten angelegter Bias in der Berechnung übernommen und kann sich in der Anwendung manifestieren. Hinter Fragen der Bewertung und Gewichtung innerhalb der Verfahren verstecken sich Fragen nach Gerechtigkeit und Fairness. Die mangelnde Transparenz (im Sinne einer Interpretierbarkeit, Erklärbarkeit und Verstehbarkeit) vieler Verfahren des Maschinellen Lernens macht deutlich, dass Ergebnisse und die zugrundeliegenden Kriterien der Entscheidung nachvollziehbar – und damit überprüfbar – sein müssen. Neben domänenspezifischen stehen hier wissenschaftsmethodische und technikethische Fragestellungen auf der Tagesordnung.

Durch die in KI-Anwendungen eingebetteten Werte werden normative Setzungen sowohl in epistemischer als auch in ethischer Hinsicht vorgenommen. Aufgrund des bei vielen Anwendungen zu erwartenden großen Verbreitungsgrades und der langfristigen Verankerung – insbesondere dort, wo KI den Charakter von Infrastrukturen übernimmt – können Auswirkungen auf die Gesellschaft und den gesellschaftlichen Zusammenhalt erheblich sein. Derartige strukturelle Verankerungen erhalten durch die großflächige Implementierung einen institutionellen Charakter. Werden Forschung und Entwicklung von KI-Systemen nicht entsprechend begleitet, verfestigen die späteren Anwendungen durch ihre hohe Reichweite neue und möglicherweise unerwünschte Handlungsmuster. KI-Anwendungen sind nicht neutral. Vielmehr werden in ihrer Entwicklung bereits Werte verankert, die Affordanzen und Beschränkungen im späteren Umgang mit sich bringen. Durch diese Einflussnahme auf den späteren Gebrauchskontext entfalten KI-Systeme eine normative Wirkung.

Zugleich bieten die zutage tretende Probleme auch Chancen. Beispielsweise kann KI implizite Vorurteile, die in den Daten latent vorhanden sind, sichtbar und dadurch bearbeitbar machen. Eine entsprechende Analyse von Anwendungsschwierigkeiten kann daher wissenschaftsmethodische und ethische Schwachstellen in den Systemen aufdecken. Wird diese Auseinandersetzung kritisch begleitet und in entsprechende Überarbeitungsschleifen eingebettet, kann diese Dynamik auch eine wichtige aufklärerische Rolle übernehmen.

Da ein spätes Nachkorrigieren von unerwünschten Auswirkungen schwierig ist, liegt eine entscheidende Stellschraube in den Forschungs- und Entwicklungsprozessen selbst. Die komplexe Problemlage von KI-Systemen erfordert daher neben computer- und ingenieurwissenschaftlichem Wissen die Einbindung ethischer und wissenschaftsmethodischer Expertise in Forschungs- und Entwicklungsprozesse. Da normative Setzungen und verhaltenssteuernde Anreize in den Code implementiert werden, ist es wichtig, ethische und wissenschaftsmethodische Expertise direkt in die Forschungs- und Entwicklungsprozesse einzubetten. Dieses „Ethics-by-design"[71] setzt auf der Ebene der Arbeitsprozesse eine integrierte Ethik voraus, bei der ethische und wissenschaftsmethodische Expertise in interdisziplinäre Zusammenarbeit direkt in Projektstrukturen eingebunden wird.

Aufgrund der großen gesellschaftlichen Tragweite ist zugleich eine gesellschaftliche Verständigung zu Technologievisionen in Einklang mit fundamentalen Werten vonnöten. Somit ist entscheidend, dass in Forschung und Entwicklung von KI-Systemen bereits von Anfang an neben einer inter- auch eine transdisziplinäre Zusammenarbeit angelegt ist, damit die späteren Anwendungen ausreichend ethisch robust und resilient gestaltet sind. Dazu bedarf es neuer Verfahren eines gesellschaftlichen Co-Designs. Diese setzen allerdings voraus, dass Nutzer*innen (im professionellen Umfeld ebenso wie in Alltagssituationen) zum kritischen Umgang mit KI-Systemen ermächtigt werden. Hierzu gehört das Wissen um die fundamentalen Wirkprinzipien, Funktionsweisen und Einsatzgebiete sowie Kenntnis der Grenzen und Problematiken des maschinellen Lernens, wie das erwähnte Gerechtigkeitsproblem im Umgang mit Bias.

Ansätze

In der Entwicklung verantwortungsvoller KI kommt den Ausbildungsstrukturen eine Schlüsselrolle zu. Denn die Einbettung inter- und transdisziplinärer Herangehensweisen in Forschung und Entwicklung von KI-Systemen ist keine Selbstverständlichkeit. Sie setzt bei allen Akteuren neue Kompetenzen voraus. Neben der ethischen Gestaltung von Forschungs- und Entwicklungsprozessen durch ein verändertes Prozessdesign und adäquater struktureller Rahmenbedingungen liegen die wichtigsten Stellschrauben in Ausbildungsstrukturen. Damit eine verantwortungsvolle Entwicklung von KI gelingen kann, bedarf es veränderter

71 Ethics-by-design soll hier einem umfassenden Sinn verstanden werden, d. h. als ethische Gestaltung der Prozesse, der Produkte wie auch der strukturellen Rahmenbedingungen.

Ausbildungselemente und -schwerpunkte. Dazu gehören a) veränderte Ausbildungsstrukturen in den Computer- und Ingenieurwissenschaften, b) Ethiker*innen mit technischer Expertise über duale Ausbildungswege bzw. integrierte Masterstudiengänge und c) die Ausbildung von Kompetenzen für inter- und transdisziplinäre Arbeitszusammenhänge.

Eine große Herausforderung liegt in der Neuausrichtung der technischen Studiengänge. Neben einer Vermittlung von Fachwissen bedarf es einer anwendungsorientierten, ethisch-reflexiven Grundbildung für Studierende der technischen Disziplinen. Die Stärkung kritischer Reflexionskompetenz und die Ausprägung einer aktiven Diskurskultur ist ein wichtiger Grundpfeiler, um die Entwicklung von KI-Technologien im breiteren gesellschaftlichen Kontext zu betrachten und die Auswirkungen von Technologien mit gesellschaftlichen Akteuren diskutieren zu können. Ziel muss es sein, neben den Grundkenntnissen einen erweiterten Betrachtungsfokus v. a. auch auf das Erkennen der eigenen Wissens- und Kompetenzgrenzen zu erreichen und eine Basis für eine Sprechfähigkeit für die Zusammenarbeit mit Ethiker*innen und gesellschaftlichen Akteuren zu legen.

Die Schwerpunktbereiche in der Ausbildung sollten u. a.

- Methoden der Technikfolgenabschätzung und der angewandten Ethik,
- Technikreflexion (Perspektiven aus Philosophie, Geschichte, Sozial- und Kulturwissenschaften), Wissenschaftstheorie der Technikwissenschaften und Methodenkritik,
- Kompetenzen der projektorientierten, interdisziplinären Zusammenarbeit umfassen.

Komplementär dazu braucht es speziell ausgebildete Absolvent*innen der geistes- und sozialwissenschaftlichen Disziplinen, die eine fundierte technische Expertise zu KI-Anwendungen und neuen Technologien im breiteren Sinn erhalten haben. Insbesondere duale Ausbildungswege, aus denen z. B. Ethiker*innen mit technischer Expertise hervorgehen, bieten sich hier an.

Eine zügige Umarbeitung der Curricula der Computer- und Ingenieurwissenschaften sowie die Einrichtung neuer, dualer Masterstudiengänge ist angesichts des rasanten Entwicklungstempos neuer KI-Technologien dringend geboten. Als Übergangslösung bieten sich hier Zertifikate an, die über ein entsprechendes Anreizsystem zu einer freiwilligen Vertiefung im Rahmen des Studiums motivieren und nach erfolgreichem Abschluss die neuerworbenen Kompetenzen dokumentieren. So startet an der TU Berlin ein Zertifikatsprogramm für Studierende aller

Fächer, das mit einem Schwerpunkt im Bereich Ethik der KI belegt werden kann (Berliner Ethik Zertifikat – Profilbereich Ethik der KI; siehe Exkurs).

Ein weiteres, wichtiges Element der Ausbildung muss der Erwerb von Methoden der teamorientierten, interdisziplinären Zusammenarbeit und eine Sensibilisierung für die damit einhergehenden Herausforderungen sein. Integrierte Ethik ist mehr als das Einbetten von Ethikexpert*innen in Forschungs- und Entwicklungsteams. Damit die Zusammenführung gelingt, bedarf es einer offenen Arbeitskultur, einer Fehlerkultur, einer Dialogkultur, einer Kultur der Diversität, die Arbeit auf Augenhöhe und die Wertschätzung unterschiedlicher Zugänge und Perspektiven sowie die Entwicklung kooperativer anstatt kompetitiver Arbeitsformen sowie entsprechender kreativer Freiräume.

Damit integrierte Ethik gelingen kann, benötigt sie adäquate strukturelle Rahmenbedingungen. Durch die enge Einbettung in Projektzusammenhänge und Forschungs- und Entwicklungsprozesse, lassen sich die „klassischen" Verfahren und strukturellen Rahmenbedingungen der Geisteswissenschaften, um ein kritisches Denken zu ermöglichen und abzusichern, nicht mehr ohne weiteres herstellen. Distanz – z. B. durch eine historische Distanzierung oder eine Abstrahierung vom Gegenstand – und Unabhängigkeit – z. B. durch institutionelle oder disziplinäre Grenzen – stehen bei der integrierten Ethik nicht mehr zur Verfügung. Weil die integrierte Ethik direkt in Projektzusammenhänge eingebettet ist, gestaltet sie mit und steht in der Mitverantwortung der Ergebnisse dieses Gestaltungsprozesses. Der konkrete Bezug und die spezifische Fallbearbeitung können den Blick auf allgemeinere Zusammenhänge und abstrahierte Betrachtungsebenen verschließen. Strukturelle Abhängigkeiten – etwa finanzielle oder durch hierarchische Strukturen – können eine kritische Stimme und die qualitätssichernde Funktion im entscheidenden Moment unmöglich machen.

Klar wird damit, dass eine gelingende integrierte Ethik – d. h. eine integrierte Ethik, die nicht als Feigenblatt eingesetzt wird und dem ethischen Reinewaschen dient – besondere strukturelle und institutionelle Rahmenbedingungen braucht. Zudem muss sie neue Arbeitsroutinen entwickeln, um das „kritische" Denken auch im eingebetteten Arbeitskontext zu gewährleisten. Dazu können z. B. projektunabhängige Qualitätssicherungsverfahren, finanzielle Unabhängigkeit, Absicherung im Falle von ethischem „Whistleblowing", Supervisionsmöglichkeiten für Ethiker*innen, Vernetzung und Austausch der Ethiker*innen in übergreifenden Betrachtungsperspektiven sowie ein Code of Conduct für alle Beteiligten in Projekten integrierter Ethik gehören.

Exkurs: Reflexion und Verantwortung – „Berliner Ethik Zertifikat“ mit Profilbereich Ethik der KI

Die großen Herausforderungen des frühen 21. Jahrhunderts verlangen nach neuen Ausbildungswegen. Ob Ingenieurwesen, Planungs- oder Naturwissenschaften, in Grundlagenforschung, angewandter Forschung und bei wissenschaftlichen Ausgründungen: In allen technischen und wissenschaftlichen Betätigungsfeldern wird der Ruf nach ethisch gerechtfertigtem und verantwortlichem Handeln stetig lauter. Gesellschaft und Politik, Wirtschaft und Wissenschaft fordern zunehmend eine grundlegende Fähigkeit zur Reflexion über die eigenen Methoden und ihre Wirkungen. Langfristige Folgen von Technologie und Wissenschaft auf die Gesellschaft können nur aus inter- und transdisziplinären Perspektiven erkannt und abgeschätzt werden.

Das Programm „Reflexion und Verantwortung – Berliner Ethik Zertifikat“ bietet Studierenden die Möglichkeit, innerhalb ihres Fachstudiums einen individuellen Schwerpunkt in Technikreflexion, Technikethik und Wissenschaftsforschung auszubilden.[72] Es wurde im Januar 2021 neu an der TU Berlin eingerichtet und ergänzt das im Leitbild für die Lehre der TU Berlin verankerte Zertifikatsprogramm um den wichtigen Bereich der Ethik und gesellschaftlichen Verantwortung. Im interdisziplinären Austausch erwerben die Studierenden Schlüsselqualifikationen für die Gestaltung verantwortlicher Zukünfte. Die Qualifikation kann über den Wahlbereich innerhalb der regulären Studienzeit erworben werden. Der zweistufige Aufbau ermöglicht es, ein Basis- und ein Vertiefungszertifikat zu erwerben, das auf Bachelor- und Masterstudium verteilt werden kann [Abb.1]. Koordiniert wird das Zertifikat durch eine interdisziplinäre Arbeitsgruppe des Present Futures Forums/ Berlin Ethics Lab unter der wissenschaftlichen Leitung von Sabine Ammon.

Das „Berliner Ethik Zertifikat“ wird für das Bachelor- und Masterstudium aller Studienrichtungen angeboten. Es gliedert sich in ein Basis- und ein Aufbauzertifikat und kann begleitend über den gesamten Studienverlauf erworben werden. Nach erfolgreichem Erwerb des Basiszertifikats ist ein Erwerb des Aufbauzertifikats möglich, aber nicht verpflichtend. Das Basiszertifikat umfasst drei Module im Umfang von 18 Leistungspunkten; das Aufbauzertifikat zwei Module im Umfang von 12 Leistungspunkten. Die Studierenden haben die Möglichkeit, das Zertifikat innerhalb eines Studienabschnitts abzuschließen oder über beide Abschnitte verteilt zu erwerben.

72 Vgl: https://www.berlinethicslab.tu-berlin.de/menue/berliner_ethik_zertifikat/ [15.3.2021]

Der Erwerb des Zertifikats erlaubt die interdisziplinäre Verknüpfung technisch-naturwissenschaftlicher Expertise mit geistes- und sozialwissenschaftlicher Reflexion. Es setzt damit die in der allgemeinen Studien- und Prüfungsordnung der TU Berlin verankerte „ganzheitliche Herangehensweise", bei der „Natur-, Planungs- und Ingenieurwissenschaften gleichberechtigt mit Geistes- und Sozialwissenschaften in engem Verbund" forschen und lehren, konsequent um. Neben Praktiken des inter- und transdisziplinären Arbeitens prägen Problemstellungen zu Ethik und gesellschaftlicher Verantwortung von Wissenschaft und Technik sowie Theorien und Methoden der Wissenschaften in der Wissens- und Kompetenzvermittlung das Curriculum des Zertifikats. Mit dem Sommersemester 2021 können Studierende einen thematischen Schwerpunkt zur Ethik der KI wählen; weitere Schwerpunkte sind in Planung. Eine wichtige Rolle innerhalb des Zertifikats spielt auch das interdisziplinäre *Co-Teaching*, um die integrierte Herangehensweise auch auf Seiten der Lehrenden vorzuleben. So unterrichten im neu eingerichteten Modul „Ethics of AI" Wissenschaftler*innen aus Philosophie und Computerwissenschaften gemeinsam interdisziplinär zusammengesetzte Studierendengruppen.

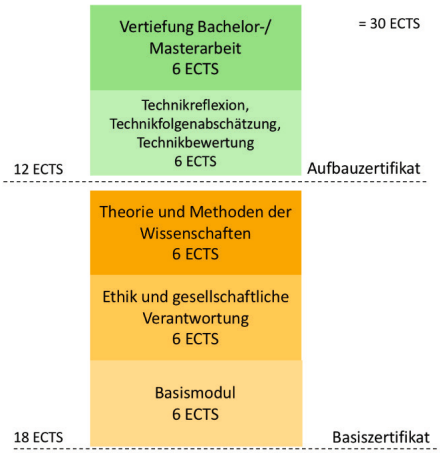


Abb. 1: Aufbau des Berliner Ethik Zertifikats @ Fachgebiet Wissensdynamik und Nachhaltigkeit in den Technikwissenschaften, TU Berlin.

4. ANWENDUNG VON KI-SYSTEMEN

Isabella Hermann und Günter Stock

A. KI UND HERAUSFORDERUNGEN FÜR MENSCHLICHE KOMPETENZ[73]

Die bei der Entwicklung aufgeworfenen Themen gehen in der konkreten Anwendung mit einem nötigen Kompetenzerwerb bzw. Professionalisierungsbedarf im Umgang mit KI-System einher und dies sowohl auf Seiten professioneller als auch alltäglicher Anwender*innen. Denn wir können nur verantwortungsvoll agieren, wenn die Funktionsweisen und mögliche ethische Herausforderungen hinreichend bekannt sind.

Zunächst jedoch gilt die Sichtweise, dass durch technische Automatisierung menschliche Unzulänglichkeiten und Fehler verhindert werden können. Das sieht man eindeutig am Flugverkehr, wo man auf eine über 100-jährige Geschichte der Zunahme von technischen Standards und Automatisierung zurückblicken kann, die die Sicherheit des Flugverkehrs enorm erhöht haben. Kommt der Mensch vor, wird er lediglich als ein Störfaktor in einem sonst funktionierenden System angesehen, der so klein wie möglich gehalten werden muss. In der Mensch-Maschine-Interaktion findet sich in diesem Zusammenhang das „Automations-Paradox", nach dem technisch automatisierte Prozesse in den meisten Fällen als sicherer und dem Menschen überlegen angesehen werden, außer wenn etwas schief geht und Menschen einspringen müssen.[74] Im Flugverkehr ist das fatal, da die kognitiven Fähigkeiten der Pilot*innen wie Navigation sowie Fehlererkennung und -diagnose anfällig dafür sind, bei zunehmender Automatisierung abzunehmen. Nicht umsonst weisen die Luftfahrtbehörden der USA und Europa darauf hin, in der Ausbildung und im Linienbetrieb von Pilot*innen einen Schwerpunkt auf manuelle Trainings und Flugphasen zu legen.

73 Der Beitrag basiert auf Hermann, I; Stock, G (2020): Kompetenzverlust in Zeiten von KI – Wie bewahren Menschen wichtige Fähigkeiten. In: Kompetent eigene Entscheidungen treffen? Auch mit Künstlicher Intelligenz! Ausgabe 2/2020 der Schriftenreihe „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen". Berlin, S. 24–38.

74 Bainbridge, L (1983): Ironies of Automation. In: Automatica 19(6), S. 775–779.

Dieses Spannungsfeld scheint sich mit der zunehmenden Automatisierung von Abläufen, die durch die Digitalisierung und den Einsatz von KI und Maschinellem Lernen ermöglicht werden, zu erweitern und dadurch zu verschärfen. So ist ein ähnliches Phänomen bei der immer stärker werdenden Automatisierung von Fahrzeugen zu beobachten. Neben mehr Komfort für die Menschen ist das große gesellschaftspolitische Ziel hinter der Idee des autonomen Fahrens die „Vision Zero", also ein Straßenverkehr ohne Tote und Schwerverletzte, da bei über 90 Prozent der Unfälle menschliches Versagen die Unfallursache ist.[75] Beim automatisierten Fahren ist allerdings weiterhin ein Mensch für den Fahrzeugbetrieb verantwortlich – auch wenn es sich um vollautomatisiertes Fahren auf der sogenannten Stufe 4 handelt.[76] Bei steigender Automatisierung gilt jedoch das gleiche Paradox wie beim Einsatz von Autopiloten beim Fliegen, nämlich, dass die Rolle für die Fahrer*innen immer schwieriger wird.[77] Wenn man sich über weite Strecken zurücklehnen kann, woher soll der Mensch die Fahrpraxis nehmen, um in schwierigen Situationen, die Algorithmen nicht meistern können, plötzlich einzugreifen? Wie sollen menschliche Fahrer*innen Gefährdungssituationen einschätzen, wenn Algorithmen das Fahrzeug möglicherweise entgegen den eigenen Erwartungen und denen der Fahrer*innen der anderen Autos steuern, beispielsweise wenn es beim städtischen „Stop-and-Go" langsamer oder vorsichtiger fährt, als es gemeinhin unter menschlichen Fahrer*innen üblich ist? Die Eigenheiten des automatisierten Fahrens sollten generell in der Führerscheinprüfung und gezielt in Trainings zum jeweiligen Fahrzeugmodell vermittelt werden, um die Kompetenzen der Fahrer*innen und die Sicherheit im Straßenverkehr bei automatisierten Fahrfunktionen zu erhöhen.

Ähnliches lässt sich beispielsweise beim zunehmenden Einsatz von Robotik und KI in der Medizin feststellen. Auch wenn „Robochirurgie" noch kein ähnlich fortgeschrittenes Stadium auf dem Weg zur Autonomie wie das Autofahren erreicht hat, existieren bereits einzelne robotische Systeme, die automatisiert Ausführungspläne erstellen (bei beispielsweise Cyberknife in der Radiochirurgie) – die Ärzt*innen freilich überprüfen und genehmigen müssen. Allerdings ist bisher kein OP-Roboter in der Lage, Aufgaben zu übernehmen, die nicht auch von einem Menschen ausgeführt werden können. Das Zusammenspiel zwischen

75 Schäfer, P (2018): Der lange Kampf um Vision Zero. In: Springer Professional. https://www.springerprofessional.de/fahrzeugsicherheit/automatisiertes-fahren/derlange-kampf-um-vision-zero/15771248 [15.3.2021].

76 ADAC (2018): Autonomes Fahren: Die 5 Stufen zum selbstfahrenden Auto. https://www.adac.de/rund-ums-fahrzeug/autonomes-fahren/grundlagen/autonomes-fahren-5-stufen [15.3.2021].

77 Noy, I; Shinar, D; Horrey, W (2018): Automated driving: safety blind spots. In: Safety Science 102, S. 72. https://www.sciencedirect.com/science/article/pii/S0925753517304198 [15.3.2021].

Mensch und Maschine kann allerdings wie schon beim Fliegen dazu führen, dass die Geschicklichkeit und praktischen Fähigkeiten des medizinischen Personals abnehmen, was im Notfall und bei Komplikationen möglicherweise zu Problemen führen kann. Denn auch auf höheren Autonomiestufen stehen die Roboter weiterhin unter der Kontrolle der behandelnden Ärzt*innen, die wiederum verantwortlich dafür sind, was im OP-Saal geschieht. Doch selbst bei dem flächendeckend an deutschen Universitätskliniken eingesetzten, minimal invasiven Da-Vinci-Operationssystem, das keinen Autonomiegrad besitzt, sondern lediglich die Vorgaben der Chirurg*innen ausführt, sei immer mehr „technisches Know-how gefragt", weil die OP-Techniken für den Robotereinsatz angepasst würden.[78] Dieses technische Know-how sollte daher stärker in Studium, Aus- und Weiterbildung integriert werden.

Verändern robotische Systeme im Operationssaal die Anforderungen an die Chirurg*innen, sind es in der Diagnose und Therapie KI-basierte Assistenzsysteme und Entscheidungshilfen, um Ärzt*innen einfache, repetitive und fehleranfällige Aufgaben abzunehmen. KI-basierte Systeme sind in der Diagnostik aktuell vor allem im Bereich der bildgebenden Verfahren vielversprechend, wie in der Krebserkennung, wo trainierte Modelle verdächtige Knoten in Brust oder Lunge, Metastasen oder Hautkrebs bereits zuverlässig erkennen. Hier ist es allerdings umso wichtiger, dass Ärzt*innen über einen kompetenten Umgang mit Daten und ein Grundverständnis der Fehlerquellen und Schwächen des Maschinellen Lernens verfügen. Denn die Beurteilung von Vorhersagen eines unterstützenden KI-Systems setzt voraus, dass Ärzt*innen sowohl die Daten, die das System verwendet, also auch die Prozesse, die zum Ergebnis führen, nachvollziehen und plausibilisieren können. Das ist wichtig, weil ein System, das für einen bestimmten Kontext mit Daten einer bestimmten Region oder Gruppe trainiert wurde und funktioniert, dies in einem anderen Umfeld schon nicht mehr tun muss, da bereits kleine Unterschiede in den Patientengruppen, bei den Scans oder der Bilddaten die neuronalen Netze irritieren können. Damit einhergehend ist beim Einsatz von KI-basierten Systemen als Entscheidungshilfe ein Statistikverständnis wichtig, ein Bereich, der nach Expertenmeinung vernachlässigt werde und dazu führe, dass Ärzte und Ärztinnen falsche Informationen auch an ihre Patientinnen und Patienten weitergeben.[79] Sollten KI-basierte Assistenzsysteme in Zukunft sogar verpflichtend werden, werden diese Kompetenzen noch wichtiger, weil sich Ärzte

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

79 Max-Planck-Institut für Bildungsforschung Presse- und Öffentlichkeitsarbeit (2018): Viele Medizinstudierende verstehen Statistik nicht. https://www.mpib-berlin.mpg.de/sites/default/files/press/2018-10-23_mpib_pm_medizinstudierende-statistik_de.pdf [15.3.2021].

durch die technische Unterstützung stärker auf den persönlichen Kontakt und Beratung des Patienten konzentrieren könnten und die Patienten entsprechend über Wesen, Bedeutung und Tragweite bestimmter Behandlungsvorschläge aufklären müssten. Wie beim Fliegen und beim automatisierten Fahren können hier also Trainings dem Kompetenzverlust entgegenwirken: OP-Roboter können als Trainingssimulator dienen, um Chirurgen technisch auszubilden, Statistik und Datenkompetenz muss in der ärztlichen Fort- und Weiterbildung eine Kernrolle spielen.

Die oben genannten Beispiele zeigen, dass eine sorgfältige Aus- und Weiterbildung, sowie entsprechende kontinuierliche Trainings erforderlich sind, damit Nutzer*innen und Anwender*innen mit maschinellen Systemen schritthalten sowie die KI-Systeme Ergebnisse gestalten und steuern können, um sich im Zweifel gegen eine KI-Vorhersage zu stellen. Kompetent mit KI-Systemen umgehen zu können, hat eine große Relevanz, da Menschen im Zusammenspiel mit Maschinen weiterhin auch die rechtliche Verantwortung – sei es ein Pilot, ein Autofahrerin oder eine Ärztin– tragen. Nur so ist der große Mehrwert von KI-Systemen nutzbar zu machen.

Pia-Johanna Schweizer und Ortwin Renn

B. DIE IDENTIFIZIERUNG VON EMERGING RISKS DURCH KI

Die COVID-19-Pandemie hat gezeigt, dass es keine akzeptable Risikomanagement-Strategie ist, scheinbar weit entfernte Risiken zu ignorieren. Eine große Herausforderung für die Risikoforschung besteht darin, auf noch unbekannte, sogenannte „Emerging Risks", hinzuweisen, bevor sich diese voll manifestieren. Neu auftretende Risiken sind das Ergebnis neuer oder zukünftiger Bedrohungen, bei denen nur ein geringer (oder gar kein) Kenntnisstand über die potenziellen Verluste sowie die Wahrscheinlichkeitsverteilung ihres Auftretens vorliegt.[80] Emerging Risks können sich zudem aus einer neuen Kombination von bereits bekannten Risiken ergeben, die im systemischen Zusammenwirken neue Schadenspotenziale generieren.[81] Bei all dem spielt das vorhandene Wissen um Emerging Risks eine entscheidende Rolle. Erstens können Emerging Risks auftreten, die als potenziell

80 Renn, O (2014): Emerging Risks: Methodology, Classification and Policy Implications. Journal of Risk Analysis and Crisis Response 4(3), S. 114. https://doi.org/10.2991/jrarc.2014.4.3.1.

81 OECD (2003): Emerging Risks in the 21st Century. An Agenda for Action. OECD Publications. https://doi.org/10.1787/9789264101227-en.

bedeutsam wahrgenommen werden, jedoch (noch) nicht vollständig verstanden und bewertet werden können, z. B. CRISPR-Cas9. Zweitens können Emerging Risks auftreten, für die noch keinerlei Wissen vorhanden ist, und die demensprechend auch nicht als relevant wahrgenommen werden, sogenannte „unknown unknowns". Drittens können Emerging Risks auftreten, über die zwar Informationen zur Verfügung stehen, deren Ausmaß jedoch entweder unvorhersehbar war oder unterschätzt wurde, wie die aktuelle COVID-19-Pandemie zeigt. Bei allen Emerging Risks besteht ein hohes Maß an wissenschaftlicher Unsicherheit, was zur Folge hat, dass sich Eintrittswahrscheinlichkeit und potenzielle Schadenswirkungen nur schwer quantifizieren lassen. Sowohl im Fall der Rekombination von bereits bekannten Risiken als auch im Fall des Auftretens von gänzlich neuen Risikophänomenen steht die Risiko Governance vor der Herausforderung, auf Basis von unzureichenden oder gar mangelnden wissenschaftlichen Erkenntnissen Entscheidungen zum Umgang mit Emerging Risks treffen zu müssen, sobald diese sich (oftmals plötzlich) manifestieren.

Ermerging risks sind zwar relativ selten, haben jedoch weitreichende Auswirkungen auf die Gesundheit, Sicherheit, Umwelt, den wirtschaftlichen Wohlstand und sozialen Zusammenhalt.[82] Wenn sich die Schadenspotenziale von Emerging Risks voll realisieren, sind die Folgen für Menschen und Sachwerte oft verheerend. Jedoch treten zu Anfang der Entwicklung von Emerging Risks vereinzelte, schwache Signale auf, die auf deren Entstehung und Kausalitäten hinweisen. Die Herausforderung besteht folglich darin, relevante schwache Signale frühzeitig zu erkennen und richtig zu deuten. Hierfür ist nicht zuletzt Wissen um Ursache-Wirkungszusammenhänge nötig, um aussagekräftige Signale erst zu identifizieren und anschließend deuten zu können. Klassische statistische Analysen, wie sie z. B. in der Versicherungsmathematik angewendet werden, können nur bedingt Aufschluss über Emerging Risks geben. Risiken, deren Eintrittswahrscheinlichkeiten nicht einer Normalverteilung folgen, sondern mit seltenen Folgeereignissen, aber hohem Schadensausmaß einhergehen („fat-tailed risks"), sind mit gängigen wissenschaftlichen Methoden kaum exakt berechenbar.[83]

Künstliche Intelligenz kann zur Identifikation von neuen, noch unentdeckten Risiken beitragen. KI stellt hier ein potenzielles Instrumentarium zur Verfügung, das allerdings mit großen Datenmengen trainiert werden muss. Da die etablierte

82 International Risk Governance Council (2010): The Emergence of Risks: Contributing Factors. https://irgc.org/wp-content/uploads/2018/09/irgc_ER_final_07jan_web.pdf [15.3.2 021].

83 Taleb, N N (2010): The Black Swan. The Impact of the Highly Improbable. Überarbeitete Ausgabe, Penguin Books: London.

Datenlage zu Emerging Risks de facto gering ist, stellt sich die Frage, aufgrund welcher (Meta-)Daten, Maschinelles Lernen zur Antizipation von Emerging Risks angeleitet werden kann. Zudem müssen auch die Lernstrategien spezifiziert werden, mit deren Hilfe Muster oder bestimmte Abhängigkeiten identifiziert und verfolgt werden sollen. Auch selbstlernende Systeme benötigen eine algorithmische Strukturierung der Lernheuristiken. Das ist bei Emerging Risks eine besondere Herausforderung, weil man ja nicht genau weiß, wonach man sucht. Künstliche Intelligenz hat also das Potenzial, konventionelle Methoden der Technikvorausschau zu unterstützen. Die Technik dient in diesem Zusammenhang dazu, mit der Informationsflut von Big Data umzugehen und gezielt nach Signalen zu suchen, die auf Emerging Risks hindeuten und für den Umgang mit Risiken von Relevanz sein könnten. Mehr Daten bedeuten jedoch nicht *ceteris paribus* mehr Information. Um aus großen Datenmengen sinnhafte Informationen ableiten zu können, muss zwischen Signal und „Rauschen" unterschieden werden. Menschliche Intelligenz ist nötig, um sowohl beim überwachten als auch unüberwachten Lernen Lernheuristiken ex ante bereitzustellen, in die Erkenntnisse über bereits manifestierte Risiken einfließen. Menschliche Fähigkeiten sind nötig, um Kriterien festzulegen, nach denen zwischen Signal und „Rauschen" unterschieden wird und die die Relevanz und Angemessenheit von Indikatoren für Emerging Risks sicherstellen. Ohne menschliche Intelligenz könnten sich unreflektiert Pfadabhängigkeiten und Verstärkungseffekte beim Trainieren der Künstlichen Intelligenz einstellen. Die durch unüberwachtes Lernen rein statistisch gefundenen Muster oder Cluster können gelegentlich Hinweise auf wichtige Zusammenhänge geben, aber oft sind sie ebenfalls nur Zufallsprodukte oder weisen auf Zusammenhänge hin, die für das Thema Emerging Risks völlig irrelevant sind. Zwar sind Fehldeutungen bei der Generierung der Indikatoren-Sets auch unter Einsatz menschlicher Intelligenz als Korrektiv nicht ausgeschlossen, jedoch sind sich menschliche Akteure der Fehlbarkeit ihrer Entscheidungen bewusst und können diese in Frage stellen.[84]

Menschliche Intelligenz und menschliche Fähigkeiten sind nach wie vor unerlässlich, um Modelle bzw. einen gehaltvollen Bezugsrahmen für maschinelles Lernen zur Antizipation von Emerging Risks zu generieren. Entscheidend ist, dass überraschende Phänomene überhaupt als sinnvolle Information gegenüber dem „Hintergrundrauschen" wahrgenommen werden. C.S. Peirce verwendet den Begriff der Abduktion für die kreative Genese einer Hypothese zur Erklärung von

84 Braga, A; Logan, R K (2017): The emperor of strong AI has no clothes: Limits to artificial intelligence. In: Information 8(4), S. 1–21. https://doi.org/10.3390/info8040156; Ghahramani, Z (2015): Probabilistic machine learning and artificial intelligence. Nature 521(7553), S. 452–459. https://doi.org/10.1038/nature14541.

unerwarteten Phänomenen.[85] Abduktion erweitert die Erkenntnis demnach auf kreative Art und Weise. Sie kann nach Armin Nassehi als hypothetischer Schluss begriffen werden, in dem „das Einzelne mit Erfahrungen, Regelmäßigkeiten und daraus resultierenden Regeln abgeglichen wird".[86] Abduktion gelingt (bislang) ausschließlich aufgrund von menschlicher Intelligenz. Zudem müssen Modelle zur Analyse von sozio-technischen und sozio-ökologischen Systemen aufgrund der inhärenten Dynamik dieser Systeme beständig adaptiert werden. Dabei gilt es, die Modelle nicht nur an neue Rahmenbedingungen anzupassen, sondern auch die Axiome und Antezendensbedingungen der Modelle zu reflektieren und ggf. zu revidieren. Dabei sollten sowohl die erzeugten (Risiko-)Modelle als auch die mentalen Modelle, die jenen Modellen und dem gewählten Zugang zugrunde liegen, ständig hinterfragt und überprüft werden. Im Bereich der Sicherheitsforschung von kritischen Infrastrukturen (z. B. AKWs oder Bohrinseln) wird zu diesem Zweck die Methode des „red teaming" empfohlen.[87] Die „red teams" haben die Aufgabe, nicht nur mögliche Fehler und Schwachstellen in organisationalen Abläufen zu erkennen, sondern auch die anderen Teams dazu zu zwingen, ihre basalen Axiome und Denkweisen explizit zu machen. Obwohl (oder gerade weil) die Forschung zu kritischen Infrastrukturen weit vorangeschritten und auf den automatisierten Einsatz von Big Data angewiesen ist, herrscht hier ein besonders kritisches Bewusstsein für die Grenzen von Künstlicher Intelligenz. Neue Formen der Ko-Produktion von Wissen sind nötig, bei denen die analytisch-algorithmische Stärke von Künstlicher Intelligenz mit der sinngebenden Strukturbildung menschlicher Intelligenz zusammenwirken soll.

Das Ziel ist von Risiko Governance im Umgang mit Emerging Risks ist es, durch Foresight und adaptives Management bereits vor Schadenseintritt das Risiko zu antizipieren, zu minimieren oder falls möglich zu vermeiden.[88] Die moralisch-ethische Bewertung von möglichen Managementoptionen muss nach wie vor dem Menschen obliegen. Die Abwägung von Chancen und Risiken sollte dabei

85 Peirce, C S (1932): Collected Papers of Charles Sanders Peirce. Volumes I and II: Principles of Philosophy and Elements of Logic. Harvard University Press: Cambridge MA.

86 Nassehi, A (2019): Muster. Theorie der digitalen Gesellschaft. 2. Auflage. C. H. Beck: München, S. 159 f.

87 Aven, T (2015): Implications of black swans to the foundations and practice of risk assessment and management. Reliability Engineering and System Safety 134, S. 83–91. https://doi.org/10.1016/j.ress.2014.10.004.

88 Linkov, I; Trump, B D; Anklam, E; Berube, D; Boisseasu, P; Cummings, C; Vermeire, T et al. (2018): Comparative, collaborative, and integrative risk governance for emerging technologies. Environment Systems and Decisions 38(2), S. 170–176. https://doi.org/10.1007/s10669-018-9686-5.

transdisziplinär im deliberativen Diskurs unter Beteiligung von Interessenvertreter*innen und der Bevölkerung stattfinden.[89]

89 Schweizer, P J (2019): Systemic risks–concepts and challenges for risk governance. Journal of Risk Research. https://doi.org/10.1080/13669877.2019.1687574; Schweizer, P J; Renn, O (2019): Governance of systemic risks for disaster prevention and mitigation. Disaster Prevention and Management: An International Journal 28(6). https://doi.org/10.1108/DPM-09-2019-0282.

Frauke Rostalski und Thorsten Thiel

5. KÜNSTLICHE INTELLIGENZ ALS HERAUSFORDERUNG FÜR DEMOKRATISCHE PARTIZIPATION

Die Untersuchung der Auswirkungen von Künstlicher Intelligenz (KI) auf Demokratie hat erst in jüngerer Zeit Fahrt aufgenommen. Jene ist ein Unterbereich in der größeren, meist soziologisch, rechtswissenschaftlich oder ethisch geführten Debatte um die Auswirkungen von KI auf die Gesellschaft insgesamt. Während diese übergeordnete Debatte meist die unmittelbaren Probleme Maschinellen Lernens thematisiert, wie die Perpetuierung von sozialen Ungleichheiten, konzentriert sich die Debatte um Demokratie und KI bisher im Kern auf zwei Aspekte: die Wirkung digitaler Medien auf die Polarisierung und Manipulierbarkeit der Bevölkerung sowie die Kritik an der Macht jener Plattform-Unternehmen, die im Zuge der gegenwärtigen Technologieentwicklung weiter an Unangreifbarkeit gewinnen.[90] Diese allgemeine Diagnose der Auswirkungen von KI auf Demokratie dient im Folgenden als Ausgangspunkt, um zu versuchen, etwas weiter in die Zukunft hinein zu denken. Dazu soll der Fokus einmal weg von der Gefahrenanalyse und auf die konkrete Frage hingelenkt werden, wie wir sicherstellen können, dass die Möglichkeiten von Bürger*innen erhalten bleiben, auch in einer durch KI geprägten Gesellschaft demokratische Praktiken des Verstehens und der Mitsprache, sprich: der Beteiligung, zu realisieren.

Demokratie meint in einer solchen Überlegung nicht einfach die staatliche Regulierung einer soziotechnischen Entwicklung, wie es gegenwärtig als digitale Souveränität bezeichnet und im europäischen Kontext viel zu oft mit Demokratie gleichgesetzt wird. Vielmehr ist hier ein anspruchsvolleres, gesellschaftsbezogenes Verständnis von Demokratie gemeint, nämlich Demokratie als Lebensform. Die demokratische Ordnung ist nach diesem Verständnis dadurch charakterisiert, dass gesellschaftliche Pluralität einen steten und öffentlichen Ausdruck in Diskursen

90 Dieser thematische Zuschnitt ist dabei bereits aus der älteren Diskussion um Demokratie und Digitalisierung bekannt. Exemplarische Vertreter für eine direkte Übertragung dieser Thesen auf das Feld Künstliche Intelligenz sind im deutschsprachigen Diskurs u.a.: Hofstetter, Y (2016): Das Ende der Demokratie: Wie die künstliche Intelligenz die Politik übernimmt und uns entmündigt. C. Bertelsmann Verlag: München; Nemitz, P; Pfeffer, M (2020): Prinzip Mensch. Macht, Freiheit und Demokratie im Zeitalter der Künstlichen Intelligenz. Bonn: Dietz Verlag; Ungern-Sternberg, A (2019): Demokratische Meinungsbildung und künstliche Intelligenz. In: Unger, S; Ungern-Sternberg, A (Hrsg.): Demokratie und künstliche Intelligenz. Mohr Siebeck: Tübingen.

und Auseinandersetzungen findet. Das bedeutet zum einen, dass die Bürger*innen nicht nur durch Politik repräsentiert werden, sondern selbst aktiv Positionen und Präferenzen artikulieren und reflexiv verändern sowie zum anderen, dass Institutionen dem demokratischen Streit Ausdruck verleihen und zugleich in der Lage sind, die stets vorläufigen Mehrheiten in Politik umzusetzen, die selbst wieder Ausgangspunkt weiterer politischer Prozesse wird.

Wieso aber verändert KI in dieser Hinsicht überhaupt die Voraussetzungen bzw. das Funktionieren von Demokratie? Um diesen Einfluss zu verstehen, ist es wichtig, Künstliche Intelligenz weder zu über- noch zu unterschätzen. Überschätzt wird KI, wenn man auf die Semantik der autonomen Intelligenz hereinfällt und sie als etwas der Gesellschaft Äußeres, zudem potentiell Überlegenes, interpretiert. Besser verstanden ist KI als ein Forschungs- und Technologiefeld, dessen Gegenstand die Erfassung und Bewertung von komplexen Situationen und Prozessen ist. In diesem Feld ist es durch Methoden wie das Maschinelle Lernen und wachsende Verarbeitungskapazitäten in jüngerer Zeit zu signifikanten Effizienzsteigerungen gekommen, welche zunehmend in praktischen Anwendungen resultieren. KI erlaubt Ordnung in unstrukturierte Zusammenhänge zu bringen, d. h. sie ermöglicht eine Sortierung, Kuratierung und Analyse großer Datenmengen. Hieraus ziehen dann genau solche Anwendungen einen Nutzen, die wie Sprachassistenten, Gesichtserkennung oder autonomes Fahren, eine kontextsensitive und in Echtzeit vorgenommene Automatisierung relativ unübersichtlicher Handlungssituation unternehmen. Dass die gegenwärtig dominanten Verfahren Künstlicher Intelligenz dabei auf Mustererkennung, Training und statistischen Zusammenhängen beruhen, kann wiederum zu einer Unterschätzung der Entwicklung führen: Dies besteht darin, dass aus der methodischen Spezifität und der Angewiesenheit auf recht eindeutig kategorisierbare, zudem quantifizierbare Kontexte die Erwartung abgeleitet wird, dass die Anwendungsbereiche Künstlicher Intelligenz letztlich doch verhältnismäßig eng umgrenzt bleiben werden.

Die gesellschaftliche Bedeutsamkeit von KI steckt aber nicht in der Frage, welche Anwendungen sich konkret durchsetzen werden, sondern allgemeiner in dem Potential, unser Alltagshandeln in einer neuen Weise zu mediatisieren. Die transformative Kraft von KI ist darin zu sehen, dass wir mittels der sich etablierenden KI-gestützten Prozesse und Praktiken unsere individuellen und kollektiven Handlungsmöglichkeiten, unsere soziale Praxis, graduell, aber kontinuierlich verändern. Eine Durchdringung von Gesellschaft mit KI-gestützten Prozessen bringt mit sich, dass menschliches Handeln und Entscheiden viel stärker als jemals zuvor

durch Simulation und Vorberechnung, die Adaption an das erwartete Verhalten anderer sowie eigene, aus der Vergangenheit abgeleitete Präferenzen, informiert ist. Anwendungen des maschinellen Lernens werden darin bestehen, dass unser Alltagshandeln vorstrukturiert und mit einer Vielzahl an Verhaltens- und Handlungsvorschlägen versehen wird. KI-gestützte Verfahren haben den großen Vorteil, dass sie dem Menschen Aufgaben abnehmen und ihn dadurch von wiederholten Entscheidungen entlasten. Sie bieten die Möglichkeit, sich anderen Aspekten zu widmen und Teile der Lebensführung gewissermaßen der Technik, die gleichsam „für ihn" agiert, zu überlassen.[91] Gesellschaftlich steht zu erwarten, dass wir uns daran gewöhnen, probabilistische Rationalitäten für adäquater als allgemeine Regeln zu halten und automatisierte Bewertungen, Strukturierungen und Interventionen in vielen Bereichen unseres Lebens als unserem Wollen entsprechend zu akzeptieren. Diskutiert man Künstliche Intelligenz in einer solchen Weise über ihren – heute noch weitgehend potentiellen – Einfluss auf unsere sozialen Praktiken, wird klar, warum diese neue Instanz der Digitalisierung auch über unmittelbare Probleme oder direkte Profiteure hinaus für den steten Formwandel von Demokratie bedeutsam wird.

Anstelle weiterer Spekulation, wie diese soziotechnische Transformation aussehen wird,[92] soll nun im Blickpunkt stehen, welche individuellen und institutionellen Kompetenzen es brauchen wird, damit wir ganz grundsätzlich den normativen Anspruch demokratischer Praxis bewahren und realisieren können. Mindestens die folgenden drei Bereiche erscheinen diesbezüglich zentral: die individuelle Kapazität, die entstehende Technologie für eigene Ziele zu nutzen; die gesellschaftliche Möglichkeit die strukturprägende Kraft der Technologie zu markieren und wahrzunehmen; und die Art und Weise der Einbettung von Künstlicher Intelligenz in politische Prozesse.

91 Häufig ist diesbezüglich von einer „Entscheidungsdelegation" vom Menschen auf die Technologie die Rede. Streng genommen liegt eine solche freilich nicht vor, da der Mensch zumindest die initiale Entscheidung getroffen hat, dass die KI-Anwendung in einem bestimmten Lebensbereich auf der Basis seiner bisherigen Präferenzen fortlaufend agiert. Gleichwohl bleibt es indessen dabei, dass der Mensch sich aus diesem Lebensbereich in Teilen zurückzieht, trifft er doch lediglich eine „Generalentscheidung" und befasst sich im Weiteren nicht länger mit den konkreten Einzelsituationen, in denen diese zum Tragen kommen kann. Bereits diese Form des Rückzugs kann aber in Abhängigkeit von dem betroffenen Lebensbereich und der Häufigkeit des Phänomens bedeutsame Folgen zeitigen – nicht nur für das Individuum selbst, sondern auch die Gesellschaft, die es umgibt.

92 Vgl. Hofmann, J; Thiel, T (2021): Schleichende Übernahme. Künstliche Intelligenz und der Wandel der Demokratie. In: WZB-Mitteilungen Nr. 171., S. 9–11.

Die individuelle Kapazität zur Nutzung von KI

a. Wer sich mittels KI-gestützter Verfahren aus Entscheidungssituationen zurückzieht, verliert unweigerlich Entscheidungsfreiheit – nämlich die Freiheit, von der eingangs erteilten Generalermächtigung gegenüber der KI-Anwendung individuell abzuweichen. Dies birgt das Risiko eines Identitätsverlusts. Die Freiheit des Einzelnen weist eine praktische Bedeutung auf, indem sie in seinen Handlungen zur Realität wird.[93] Der Mensch gestaltet die Welt durch sein Verhalten und bildet dadurch im Laufe seines Lebens seine Identität aus.[94] Aus diesem Grund betrifft die Frage nach der Automation menschlicher Entscheidungsprozesse im Kern die Selbstverwirklichung des Einzelnen – und darüber die individuelle Autonomie als Grundlage demokratischer Partizipation. Die Reduktion individueller Autonomie hat negative Folgen für den Einzelnen wie die Gesellschaft. Für den Menschen besteht die Gefahr, seine praktischen Fertigkeiten zu verlieren,[95] vor allem aber darüber hinaus auch an kreativen, intellektuellen und emotionalen Fähigkeiten einzubüßen. Sofern er sich nicht als denjenigen sieht, der über seine eigenen Geschicke bestimmt, kann dies zu einer Entfremdung von sich und dem Erkalten seiner Weltbeziehung führen.[96] Diese Entwicklung hat aber auch eine unmittelbar

93 Vgl. Zaczyk, R; Köhler, M; Kahlo, M; (Hrsg.) (1998): Festschrift für E.A. Wolff zum 70. Geburtstag am 1.10.1998. Springer Verlag: Berlin, Heidelberg, S. 509, 517.

94 Zaczyk, R (2014): Selbstsein und Recht. C.H. Beck: München. S. 32 ff. in Anknüpfung an Kant, Fichte und Hegel; s. insbesondere zum Gedankengang Fichtes insoweit Zaczyk, R (2005): Zur Einheit von Freiheit und Sozialität. In: Söllner et al. (Hrsg.): Gedächtnisschrift für Meinhard Heinze. München, S. 1111, 1118.

95 Vgl. dazu – im Kontext der Arbeitswelt – etwa Steil, J; Maier, G W (2018): Kollaborative Roboter: universale Werkzeuge in der digitalisierten und vernetzten Arbeitswelt. In: Maier, G W; Engels, G; Steffen, E (Hrsg.): Handbuch Gestaltung digitaler und vernetzter Arbeitswelten. Springer Nature: Cham. S. 1–12; Ethik-Kommission Automatisiertes und Vernetztes Fahren (2017): Abschlussbericht. Berlin, S. 22; außerdem Hermann, I; Stock, G (2020): Kompetenzverlust in Zeiten von KI – Wie bewahren Menschen wichtige Fähigkeiten. In: Kompetent eigene Entscheidungen treffen? Auch mit Künstlicher Intelligenz! Ausgabe 2/2020 der Schriftenreihe „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen". Berlin, S. 24–38.

96 Vgl. Rostalski, F (2020): Entscheiden im digitalen Zeitalter. In: Kompetent eigene Entscheidungen treffen? Auch mit Künstlicher Intelligenz! Ausgabe 2/2020 der Schriftenreihe „#VerantwortungKI – Künstliche Intelligenz und gesellschaftliche Folgen". Berlin, S. 9–23; Hildebrandt, M (2016): Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology. Edward Elgar Pub: Cheltenham/Northampton 2016; Whittlestone, J; Nyrup, R; Alexandrova, A; Dihal, K; Cave, S (2019): Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research. In: http://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf, S. 22 f. [15.3.2021]. Vgl. auch Delacroix, S (2018): From agency-enhancement intentions to profile-based optimisation tools: what is lost in translation. In: Bayamlıoğlu, E; Baraliuc, I; Janssens, L; Hildebrandt, M (Hrsg.) BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen. Amsterdam University Press: Amsterdam, S.16–19.; zum Begriff der Weltbeziehung s. Rosa, H (2016): Resonanz. Eine Soziologie der Weltbeziehung. Suhrkamp: Berlin.

gesamtgesellschaftliche Dimension. Wer für sich selbst nicht länger die Autorenschaft übernimmt, wird auch an den Belangen der Gemeinschaft kein Interesse haben. Für die demokratische Idee, die den verantwortungsbewussten Bürger zu ihrem Ausgangspunkt wählt, hätte dies erhebliche negative Folgen.

Um dies zu verhindern, müssen dem Individuum hinreichende Spielräume verbleiben, um sich als Autoren seiner selbst zu verstehen. Diese Forderung weist eine qualitative wie auch eine quantitative Dimension auf: Wer Entscheidungsfreiheit in bedeutsamen Lebensbereichen verliert, vermag sich ebenso in seiner Identität beeinträchtigt zu sehen wie derjenige, der zwar lediglich nachgeordnete Lebensbereiche mit einer Generalermächtigung gegenüber KI-Anwendungen versieht, dies aber in besonders großem Umfang tut. Erforderlich erscheint daher ein gesellschaftsweites Bewusstsein für diese auf den ersten Blick indirekte Gefahr. Es bedarf einer Informierung, um eine reflexive Abwägung darüber herbeizuführen, wann und wie weit man sich auf KI verlassen will. Zugleich ist zu erwägen, ob weitere, über die Selbstkontrolle des Subjekts hinausgehende Schutzmechanismen, gesellschaftlich implementiert werden sollten.

b. Eine naheliegende allgemeine Forderung in Bezug auf individuelle Kapazität(en) und medialen Wandel ist stets, Medienkompetenz zu fordern Eine einfache Version dieser Forderung ist eine breite Anwendung und Vertrautheit mit Technologien, da hierdurch Nutzungserfahrung und Gewöhnung gleichsam automatisch ermächtigend wirken soll, da die Gleichheit von Beteiligungschancen sich erhöht (man denke an ältere Diskussionen um den *digital divide* und *digital natives* als Beispiel). Diese Position wird dann über Zeit meist durch stärker kritische Positionen abgelöst, welche Medienkompetenz in dem Sinne verstehen, dass man sich (intendierte und nicht-intendierte) Folgewirkungen von Technologien bewusstmacht und den Technologieeinsatz situativ begrenzt. Beide Perspektiven auf Medienkompetenz sind von Wert, greifen aber im größeren gesellschaftlichen Kontext medialer Transformation klar zu kurz. Nicht Gewöhnung oder kritisches Bewusstsein, sondern ein aktiver Umgang mit Technologie, welche deren Vielgestaltigkeit und Flexibilität zum Ausdruck bringt, ist wünschenswert. Die Aneignung von Technologien durch ihre Nutzer besteht darin, dass diese nicht allein die vorgegebenen Verwendungsweisen sehen, sondern kreative und spielerische Handlungsoptionen erkennen. Das heißt nicht, dass alle zum Hacker oder zur Programmiererin werden müssen. Die Aneignung

kann oft viel banaler sein, ein einflussreiches Beispiel ist etwa die Vielfalt der Verwendungsweisen von Hashtags, die sich in steter Nutzung längst selbst über Plattformgrenzen hinaus und oft regional und subkulturell ganz unterschiedlich entwickelt hat. Kreative Anwendungsweisen bedürfen eines allgemeinen Wissens um Funktionsweisen, Modularität und Erfolgsbedingungen technologischer Anwendungen sowie einer Experimentierlust, die zumindest auch dadurch befördert wird, dass allgemein gesellschaftsweit Abweichung und Ausprobieren anerkannt und unterstützt werden. Bezüglich KI kann ein solch pragmatischer Umgang darin bestehen, dass sich beispielsweise erweiterte Möglichkeiten der Analyse und Informationsbeschaffung für das eigene Handeln normalisieren.

Die gesellschaftliche Möglichkeiten der strukturprägenden Kraft von KI

Um einen individuell und gesellschaftlich reflektierten Umgang mit KI zu entwickeln und zu fördern, wird es zweitens, darauf ankommen, dass der Einsatz von KI-Verfahren als ein Technologieeinsatz erkennbar bleibt.[97] Historisch galt bisher das Ziel, den Einsatz von KI gewissermaßen zum Verschwinden zu bringen – vom Turing Test bis zu den heutigen Sprachassistenten. Dieses Verstecken des Eingriffs und der Berechnung geht jedoch damit einher, dass wir tendenziell um die Möglichkeit gebracht werden, uns zu der konfigurativen Wirkung des Technologieeinsatzes zu verhalten. Auch das Bewusstsein für die sozioökonomischen Bedingungen, die selbst wiederum die Technologie und ihre Setzung beeinflussen, verschwinden in den Hintergrund. Es geht hier, wie auch in anderen Bereichen der Digitalisierung, nicht darum, einfach blind auf Transparenz als erstes und bestes Instrument der Selbstregulierung zu setzen, denn diese ist oft überfordernd und kann schnell instrumentalisiert werden. Es bedarf vielmehr einer kollektiv verankerten Fähigkeit zu beobachten und zu analysieren.[98] Hierfür muss Technizität grundsätzlich erkennbar bleiben, im Fall von KI etwa bezogen auf das Verhältnis von Input, Schwerpunktsetzungen und der Kontingenz des Outputs. KI-Verfahren sind ungeachtet ihrer Komplexität und technischen Opazität für eine solche annotierte und explizierte Einbettung durchaus geeignet: Ihr Wirken zu verstecken, ist keine technische Notwendigkeit, sondern eine aus ökonomischen

97 Vgl hierzu auch das zweite und vierte der New Laws of Robotics von Pasquale, F (2020): New Laws of Robotics. Harvard University Press: Cambridge MA.

98 Zum Konzept der Observability und zur Unterscheidung von Observability und Transparenz: Rieder, B; Hofmann, J (2020): Towards platform observability. In: Internet Policy Review 9 (4).

und/oder politischen Motiven getroffene Entscheidung und daher auch regulatorisch zu adressieren.

Einbettung von KI in politische Prozesse

Der dritte Bereich betrifft schließlich die Verwendung von KI in demokratischer Politik. Sehr schematisch lassen sich hier zwei Richtungen unterscheiden: Die eine Variante wäre, dass KI-Verfahren als demokratisch-partizipativ Verfahren als überlegen angenommen werden, da sie komplexe Wirkzusammenhänge berechnen und gleichmäßiger die Präferenzen und das Verhalten einer Bevölkerung erfassen und einbeziehen könnten. Eine auf Daten und Prognosen beruhende Politik, die zudem adaptiv und granular unmittelbar in gesellschaftlichen Handlungskontexten wirksam würde, könnte als besserer Weg gelten, das Gemeinwohl zu verwirklichen.

Deliberation, demokratische Konkurrenz und die langsame Steuerung über Gesetze wirken hier umständlich und antiquiert. Gerade weil durchaus plausibel ist, dass KI einige klassische Schwächen repräsentativer, auf Wahlen und politischen Meinungskampf basierender Demokratien auszugleichen vermag – man denke an langfristige Aufgaben wie den Klimawandel –, wird es wichtig sein, diese Form der Berechnung nicht als einzig akzeptable demokratische Entscheidungslogik zu werten.[99] KI bietet vielmehr auch die Möglichkeit, die Kontingenz politischen Entscheidens deutlich zu machen. Dies ließe sich etwa realisieren, indem KI-gestützte Anwendungen die Optionalität einer Entscheidung nicht etwa verstecken, sondern sichtbar machen, indem sie annotiert werden. Bürgerinnen und Bürgern würden so Möglichkeiten eröffnet, Interdependenz und Komplexität von Politik selbst zu erfahren. KI müsste hierfür so eingesetzt werden, dass mittels ihrer die Möglichkeit unterschiedlicher Entscheidungen repräsentiert und erfahrbar wird und dass sie auf diese Weise den demokratischen Diskurs unterstützt. Für die Bürgerinnen und Bürger würde KI somit selbst zur Ressource werden, um Entscheidungshandeln zu hinterfragen und ggf. auch anzufechten.

Insgesamt besteht die demokratietheoretische Herausforderung im Feld der Künstlichen Intelligenz also darin, den Prozess der Erstellung und Entwicklung

99 Eine solche Erwartung skizziert etwa Harari, Y N (2017): Homo Deus. Harper: New York (insb. Kap. 11). Zu den Möglichkeiten und Grenzen des Einsatzes autonomer Entscheidungssysteme vgl. außerdem: König, P D; Wenzelburger, G (2021): Between Technochauvinism and Human-Centrism: Can Algorithms Improve Decision-Making in Democratic Politics? In: European Political Science (online first.). https://doi.org/10.1057/s41304-020-00298-3.

KI-gestützter Verfahren und Einsatzbereiche so zu konfigurieren, dass demokratische Interventionen – das inklusive und reflexive Einwirken auf den Prozess – nicht nur möglich bleiben, sondern auch hervorgehoben werden. Demokratische Verantwortung erwächst nämlich genau in solchen Kontexten und Prozessen, in denen kollektive Handlungsmöglichkeiten als kontingent und offen markiert sind. Das Versprechen der Demokratie ist seit der Antike, dass kollektive Selbstbestimmung durch Praktiken aktiver Partizipation realisiert werden kann; Künstliche Intelligenz braucht dies nicht zu unterlaufen, sondern kann es unterstützen.

Dirk Pflüger

6. AUSBLICK: KOMPETENZERWERB VOM KINDESALTER AN

Der Einsatz von KI-Systemen in verschiedenen Lebensbereichen ist mit großen Chancen, aber auch Risiken verbunden. Einen zentralen Aspekt macht dabei die Überlegenheit von KI-Systemen gegenüber dem Menschen im Hinblick auf die schnelle Verarbeitung eines besonders großen Datenvolumens aus. Diese Rechenleistung kann vorteilhaft sein – etwa in der medizinischen Diagnostik, bei der Vermeidung von Unfällen durch intelligente Fahrerassistenzsysteme oder bei der Fehlererkennung in der industriellen Produktion –, sie kann aber auch nachteilig sein, wenn Menschen nicht mehr nachvollziehen können, wie bestimmte KI-Vorhersagen und Entscheidungen zustande kommen und daher der Technologie blind vertrauen (müssen). Dieses blinde Vertrauen kann negative Folgen für die Gesellschaft und den Einzelnen haben, insbesondere wenn Menschen zu Schaden kommen. Um dies zu verhindern und gleichzeitig die Potentiale von KI-Anwendungen zu nutzen, ist menschliche Kompetenz sowohl in Bezug auf die grundlegenden technischen Funktionsweisen als auch die ethischen Implikationen gefragt. Diese lassen sich interdisziplinär im Sinne einer integrativen Ethik durch entsprechende Ausbildung, Qualifikationen und Trainings erwerben und bewahren. Dies bedeutet auch Anforderungen an KI-Systeme „by design", also, dass von Beginn an Möglichkeiten zur Überprüfbarkeit und Erklärbarkeit von maschinellen Vorhersagen und Ergebnissen implementiert werden, damit Menschen in der Anwendung kompetent abwägen und beurteilen können.

Die zunehmende Durchdringung aller Bereiche des täglichen Lebens durch vernetzte KI-Systeme, vom Arbeitsplatz, über die Freizeit bis zum „connected home", erfordert allerdings die Schulung und Sensibilisierung aller Betroffenen, nicht nur der Entwickler*innen und professionellen Anwender*innen. Das beinhaltet alle Bürger*innen, gleich welchen Alters, Ausbildungsgrades oder gesellschaftlicher Zugehörigkeit. Die Grundlagen zur Ermächtigung im verantwortungsvollen Umgang mit KI-Systemen und deren Nutzung können nicht früh genug gelegt werden. Wer Kindergartenkindern im versierten Umgang mit KI-Sprachassistenten wie Alexa, Siri oder Google Home zugeschaut und beobachtet hat, wie sie sich

Musik vorschlagen oder Witze erzählen lassen, den Wetterbericht abfragen oder Licht oder Rollladen bedienen, sieht schnell ein, dass die ersten Bausteine einer KI-Kompetenz spätestens in der Grundschule gelegt werden müssen und nicht nur auf weiterführende Schulen oder die Ausbildung an Universitäten verwiesen werden kann. Denn wie die Beispiele zeigen, senken Sprachassistenten die Altersschwelle im Umgang mit KI-Systemen beträchtlich – im Kindergartenalter wäre an eine textbasierte Interaktion noch längst nicht zu denken. Entsprechend sieht eine Studie des Wissenschaftlichen Dienstes des Deutschen Bundestags die größten rechtlichen Bedenken durch sprachgesteuerte Assistenzsysteme für den eigenen Haushalt bei Minderjährigen.[100]

Nicht nur von Kindern wird der erhöhte Komfort durch KI-gestützte Systeme gerne in Anspruch genommen, ohne dass die in diesem Heft diskutierten Herausforderungen z. B. bezüglich unzulässigem oder unerwünschtem Bias oder auch Datenschutz gestellt werden. Informationen stehen oftmals vermeintlich frei und objektiv im Internet bzw. auf sozialen Netzwerken zur Verfügung, ohne dass die zugrundeliegenden Geschäftsmodelle durch die Nutzer*innen hinterfragt werden. Dass beispielsweise zwei unterschiedlichen Personen bei der Eingabe des gleichen Suchbegriffs – man denke an Informationen über Kandidat*innen im politischen Wahlkampf – unterschiedliche Inhalte geboten werden können, ist spätestens seit der Diskussion über die sogenannten Filterblasen bekannt.[101] Es werden schlichtweg die Informationen präsentiert, die vorgeblich am besten zu dem von einem KI-System ermittelten Weltbild der Person passen, um sie länger auf der entsprechenden Seite zu halten – damit die dahinter stehenden Unternehmen mehr Daten abgreifen und Werbung verkaufen können. Hier müssen ein Grundbewusstsein und Kompetenzen im Umgang mit KI vermittelt werden, die erst den aktiven Ausbruch aus der Filterblase ermöglichen. Schließlich wird maschinell getroffenen Empfehlungen ein größeres Vertrauen entgegengebracht als denen von Menschen.[102] Bias oder Voreingenommenheit von Maschinen werden mangels Verständnis der Chancen und Grenzen von KI nicht erwartet und Entscheidungen seltener hinterfragt – die Maschine gilt als objektiver als der Mensch.

100 Wissenschaftliche Dienste des Deutscher Bundestages (2019): Zulässigkeit der Transkribierung und Auswertung von Mitschnitten der Sprachsoftware „Alexa" durch Amazon. Sachstand WD 10 - 3000 – 032/19. https://www.bundestag.de/resource/blob/650728/3f72e6abc1c524961e5809002fe20f21/WD-10-032-19-pdf-data.pdf [15.3.2021].

101 Beispielsweise Schweiger, W (2017): Der (des)informierte Bürger im Netz – Wie soziale Medien die Meinungsbildung verändern. Springer Nature: Cham; Schmidt, J H (2019): Filterblasen und Algorithmenmacht. Wie sich Menschen im Internet informieren. In: Gorr, C; Bauer, M (Hrsg.): Gehirne unter Spannung. Springer: Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-57463-8_2.

102 Sharan, N; Romano, D M (2020): The effects of personality and locus of control on trust in humans versus artificial intelligence. Heliyon 6 (8). https://doi.org/10.1016/j.heliyon.2020.e04572.

Besonders für die heranwachsende „Generation KI" ist es dringend notwendig, dass die Vermittlung von Grundkompetenzen zur Künstlichen Intelligenz schnell in die schulische Bildung Einzug hält; die Halbwertszeiten von Lehrplänen halten mit der raschen Entwicklung in diesem Bereich nicht Schritt. Entsprechend einer jetzt schon gelehrten Grundbildung in Sprachen, Sozialkunde, Biologie, Kunst, Mathematik oder Physik benötigt die Kompetenzentwicklung im Umgang mit KI die Anpassung und Ergänzung der Curricula an Schulen: Neben der Vermittlung von grundlegenden Bausteinen informatischer Systeme – quasi der DNA intelligenter Systeme – erfordert dies die Ermächtigung zum verantwortungsvollen Umgang mit KI-Systemen und zur eigenständigen Einordnung ethischer Rahmenbedingungen. Vorschläge und Konzepte für Grundschule, Unter- und Mittelstufe gibt es bereits[103] – Forderungen nach einer grundständigen Informatikausbildung sind noch viel älter.[104] Zur grundlegenden schulischen Ausbildung sollte das Wissen um die fundamentalen Wirkprinzipien und Funktionsweisen, die beispielhafte Kenntnis verschiedener Klassen von Lernverfahren und deren Einsatzgebiete, Herausforderungen an Fairness und mögliche unzulässige Diskriminierung, sowie das Verständnis von Grenzen Maschinellen Lernens gehören. Die gegenwärtig oft unreflektierte Akzeptanz maschinell getroffener Entscheidungen muss ersetzt werden durch eine grundlegende Wehrhaftigkeit, die selbst die Forderung nach robustem und resilientem Einsatz von KI stellt. Dies ist zudem notwendige Grundlage für eine in der Gesellschaft breit verankerte Verständigung und einen gesellschaftlichen Diskurs.

103 Beispielsweise Payne, B H (2019): An Ethics of Artificial Intelligence Curriculum for Middle School Students. MIT Media Lab. https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/ [15.3.2021].

104 Gesellschaft für Informatik e.V. (2000): Empfehlungen für ein Gesamtkonzept zur informatischen Bildung an allgemein bildenden Schulen. Informatik-Spektrum 23, S. 378–382. https://doi.org/10.1007/s002870000129.

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

Heft 1/2020
Isabella Hermann, Georgios Kolliarakis, Fruzsina Molnár-Gábor, Timo Rademacher, Frauke Rostalski
VERTRAUENSWÜRDIGE KI?
VORAUSSCHAUENDE POLITIK!

Heft 2/2020
Isabella Hermann, Frauke Rostalski, Günter Stock
KOMPETENT EIGENE ENTSCHEIDUNGEN TREFFEN?
AUCH MIT KÜNSTLICHER INTELLIGENZ!

Heft 3/2020
Sabine Ammon, Birgit Beck, Christoph Benzmüller, Aljoscha Burchardt, Marie Lena Heidingsfelder, Simone Kaiser, Bertram Lomfeld, Rainer Mühlhoff, Peter Remmers, Martina Schraudner
KI ALS LABORATORIUM?
ETHIK ALS AUFGABE!

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

ISBN 978-3939818-97-7